CORVINUS ECONOMICS WORKING PAPERS

04/2024

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

Parabolic Target-Space Interior-Point Algorithm for Weighted Monotone Linear Complementarity Problem

https://unipub.lib.uni-corvinus.hu/10311

CORVINUS UNIVERSITY OF BUDAPEST

Parabolic Target-Space Interior-Point Algorithm for Weighted Monotone Linear Complementarity Problem

Marianna E.-Nagy[*], Tibor Illés[†], Yurii Nesterov[‡], Petra Renáta Rigó[§]

Abstract

In this paper, we revisit the main principles for constructing polynomial-time primal-dual interior-point algorithms (IPAs). Starting from the break-through paper by Gonzaga (1989), their development was related to the *barrier methods*, where the objective function was added to the barrier for the feasible set. With this construction, using the theory of self-concordant functions proposed by Nesterov and Nemirovski (1994), it was possible to develop different variants of IPAs for a large variety of convex problems. However, in order to solve the initial problem, the most efficient *primal-dual methods* need to follow several *central paths* (up to three), which correspond to different *stages* of the solution process. This multistage structure of the methods significantly reduces their efficiency.

In this paper, we come back to the initial idea by Renegar (1988) of using the *methods of centers*. We implement it for the *weighted* Linear Complementarity Problem (WLCP), by extending the framework of Parabolic Target Space (PTS), proposed by Nesterov (2008) for primal-dual Linear Programming Problems. This approach has several advantages. It starts from an arbitrary strictly feasible primal-dual pair and travels directly to the solution of the problem in one stage. It has the best known worst-case complexity bound. Finally, it works in a large neighborhood of the deviated central path, allowing very large steps. The latter ability results in a significant acceleration in the end of the process, confirmed by our preliminary computational experiments.

Keywords: interior-point algorithm, parabolic target-space, monotone linear complementarity problems, bisymmetric matrices, polynomial complexity

JEL code: C61

1 Introduction

In this paper we deal with the *weighted linear complementarity problem (WLCP)*

$$-M\mathbf{u}+\mathbf{v}=\mathbf{q}, \quad \mathbf{u},\mathbf{v}\geq \mathbf{0}, \quad \mathbf{u}\,\mathbf{v}=\mathbf{p}, \tag{WLCP}$$

where $M\in\mathbb{R}^{n\times n}$ is a given matrix, $\mathbf{q},\mathbf{p}\in\mathbb{R}^n$ and $\mathbf{p}\geq\mathbf{0}$ are given vectors, and n is a natural number.

If we consider ($WLCP$) with $\mathbf{p}=\mathbf{0}$, then we get the class of *linear complementarity problem (LCP)*. The most important classical results about the theory, applications, and methods to solve LCPs are summarized in the monographs written by Cottle et al. [4] and Kojima et al. [18]. The largest class of matrices, called *sufficient matrices* that guarantees important properties of LCPs (e.g. the solution set of the LCP is convex) has been defined by Cottle et al. [3]. Kojima et al. [18] showed that for a new class of matrices, called $P_*(\kappa)$-matrices (P_*-matrices), IPA has polynomial iteration complexity in the size of the problem, in starting point's duality gap, in the accuracy

*Corvinus Centre for Operations Research at Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Budapest, Hungary, marianna.eisenberg-nagy@uni-corvinus.hu

†Corvinus Centre for Operations Research at Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Budapest, Hungary, tibor.illes@uni-corvinus.hu

‡Corvinus Centre for Operations Research at Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Budapest, Hungary, yurii.nesterov@uni-corvinus.hu

§Corvinus Centre for Operations Research at Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Budapest, Hungary, petra.rigo@uni-corvinus.hu

parameter and in the parameter κ. Later, Väliaho [33] showed that the sufficient matrix class and the class of P_*-matrices coincide. Interestingly enough, finite pivot algorithms for sufficient LCPs have been introduced, for details see [5, 6] and references there.

In the last few decades many IPAs have been introduced for $P_*(\kappa)$-LCPs, for details see [7, 8, 16]. Illés et al. [15] introduced an IPA for (general) LCPs that in polynomial time either gives a solution of the original problem or detects the lack of property $P_*(\tilde{\kappa})$, with arbitrary large, but apriori fixed $\tilde{\kappa}$. In the latter case, the IPA gives a polynomial size certificate depending on parameter $\tilde{\kappa}$, the initial interior point and the input size of the LCP.

Moreover, for IPAs two interesting questions arise: (i) which property of the matrix M ensures the existence and uniqueness of the central path, and (ii) what is needed to ensure to compute exact solutions instead of computing ε-optimal solutions. Illés et al. [13] proved that for sufficient LCPs central path exists and it is unique. Illés et al. [14] proved that for sufficient-LCPs from an ε-optimal solution, with small enough ε, an exact solution can be computed with a strongly polynomial algorithm called *rounding procedure*.

Ye [34] for the *weighted analytic center problem* with nonnegative weights, derived the first order optimality conditions that led to a special WLCP (see [34], problem (3) on page 319). Ye introduced a modified primal-dual path-following algorithm for solving his WLCP and derived the polynomial complexity result for the WLCP problem. During the same year, in 2008, starting from a different idea, Nesterov [25] derived exactly the same WLCP (see [25] problem (2.1) on page 2081) and used it to introduce a new IPA for solving primal-dual pair of linear programming problems.

Anstreicher [1] introduced a common generalization of the linear programming and the weighted analytic center problem, called LPWC. The dual problem (DPWC) of LPWC has been derived and both weak and strong duality results for problem pairs have been proved, as well. Anstreicher studied complexity results for several different IPAs for LPWC and DPWC. As an application, Anstreicher studied Fischer equilibrium problem, with linear utility functions, in the form of the Eisenberg-Gale formulation (for details see [34]) and using volumetric and logarithmic barriers obtained an improved complexity result.

Potra [27] observed that both problems of Ye [34] and Anstreicher [1] related to the Eisenberg-Gale formulation of the Fischer equilibrium problem, with linear utility functions led to WLCP [1]. Potra pointed out that the WLCPs of Anstreicher [1] and Ye [34] are monotone, since the matrices in the linear constraints are skew symmetric. This observation inspired him to introduce a more general class of convex optimization problems that generalise the LPWC, called *quadratic programming and weighted centering (QPWC)* problem (see subsection 2.3 in [27]). Potra [27] defined the dual problem of QPWC and derived the duality theory for this special convex programming problem class (Theorem 2.1, page 1640 in [27]). The optimality conditions of QPWC rise to a monotone WLCP. Potra [27] proposed two IPAs with polynomial iteration complexity for solving these monotone WLCPs.

Potra [28] introduced the sufficient WLCPs, studied the properties of this problem class. Most of these results (Theorem 1-4 in [28]) generalize the similar statements known in the literature of sufficient LCPs. Potra defined a predictor-corrector IPA for solving sufficient WLCPs. The complexity result depends on the initial, strictly feasible solution, its distance to the weight parameter $\mathbf{p}$, accuracy parameter and κ, the handicap of the problem's matrix.

In this paper we consider WLCPs with positive semidefinite matrices (special cases: skew symmetric and bisymmetric matrices). Such WLCPs are called *monotone WLCPs* and they can be derived from linear programming problems, linearly constrained convex quadratic programming problems and some other optimization problems, see [1, 27, 34].

We generalize the result of Nesterov [25], the primal–dual *interior-point algorithm* (IPA) for linear programming problems, which are based on the concept of *parabolic target space* (PTS) for monotone WLCPs.

The concept of *weighted central path* (WCP) in the literature of sufficient LCPs, first occurs in a paper of Illés, Roos, Terlaky [13]. Following the idea of Nesterov [25], first we introduce a relaxation of WCP, and show that the solution set of the relaxed problem is convex. Later, an additional (convex) constraint on the duality gap of the monotone WLCP has been added. Finally, we arrived at a *convex feasibility problem* (CFP) that has original variables of the monotone WLCP, and those related to the relaxation and the additional constraint (Section 3). The new variables, naturally satisfiy an extra condition leading to the observation of a PTS. The use of the PTS allows us to discuss the new IPA for both the weighted and classical monotne WLCP at the same time.

The solution of a monotone WLCP reduces to the solution of a sequence of CFPs. The driving force of our new IPA lies in the structure of CFPs and the assigned *self-concordant barrier function* F to he CFP, and its properties (Section 4).

[1]Potra was the first, who talked about *weighted complementarity problems.*

Nevertheless, the new, adaptive *parabolic target-space interior-point algorithm* (PTS IPA) for monotone WLCPs possesses the best known complexity result (Section 5). The computational efficiency of the new method has been illustrated on a test set problems (Section 6).

Throughout the paper we use the following notations. We use $\mathbb{R}^n_\oplus$ and $\mathbb{R}^n_+$ for the positive orthant and its interior. We denote by $\mathbf{e}_i$, $i=1,\dots,n$, the coordinate vectors in $\mathbb{R}^n$, and $\mathbf{e}$ is the vector of all ones. In general, with boldface small letters we denote finite dimensional vectors, while real numbers, coordinates of vectors are denoted by small letters. All arithmetic operations and relations involving vectors, like $\mathbf{xs}$, $\mathbf{x}/\mathbf{s}$ for $\mathbf{x},\mathbf{s}\in\mathbb{R}^n$, are understood in the component-wise sense. The scalar products and the norms are defined in the standard way: $\mathbf{s}^T\mathbf{x}=\sum_{i=1}^n x_i s_i$, $\quad \|\mathbf{x}\|^2=\mathbf{x}^T\mathbf{x}$, $\quad \mathbf{x},\mathbf{s}\in\mathbb{R}^n$.

Let $f:\mathbb{R}^n\to\mathbb{R}$ be a function, denote by $\nabla f(\mathbf{x})\in\mathbb{R}^n$ its gradient, where $\mathbf{x}\in\mathbb{R}^n$. For the function $F:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$, notations $\nabla_1 F(\mathbf{u},\mathbf{t})\in\mathbb{R}^n$ and $\nabla_2 F(\mathbf{u},\mathbf{t})\in\mathbb{R}^m$ are used for its partial gradients related to variables $\mathbf{u}\in\mathbb{R}^n$ and $\mathbf{t}\in\mathbb{R}^m$, respectively. A similar notation is applied to the partial Hessians, too

$$\nabla^2_{11}F(\mathbf{u},\mathbf{t})\in\mathbb{R}^{n\times n},\quad \nabla^2_{12}F(\mathbf{u},\mathbf{t})=\nabla^2_{21}F(\mathbf{u},\mathbf{t})^T\in\mathbb{R}^{n\times m},\quad \nabla^2_{22}F(\mathbf{u},\mathbf{t})\in\mathbb{R}^{m\times m}.$$

In this paper, we often use different facts from the general theory of self-concordant functions. For the reader's convenience, we summarize most of the important and useful notations and results in the Appendix.

2 Some optimization problems leading to monotone LCPs and WLCPs

In this section we introduce two, different forms of the linearly constrained convex quadratic programming problems. The first classical model contains only sign restricted variables, while the second one has free variables, as well. Due to the fact that the objective function is quadratic, the elimination of free variables can not be done in the similar way as in linear programming.

Let us consider the following linearly constrained, primal convex quadratic programming problem

$$\left.\begin{array}{c}\min \frac{1}{2}\mathbf{x}^T Q\mathbf{x}+\mathbf{c}^T\mathbf{x}\\ A\mathbf{x}\le\mathbf{b},\quad \mathbf{x}\ge\mathbf{0}\end{array}\right\}\quad (P-QP)$$

where $Q\in\mathbb{R}^{\bar{n}\times\bar{n}}$ is a given positive semidefinite matrix, and $A\in\mathbb{R}^{\bar{m}\times\bar{n}}$ is a given matrix. Furthermore, $\mathbf{c}\in\mathbb{R}^{\bar{n}}$ and $\mathbf{b}\in\mathbb{R}^{\bar{m}}$ are given vectors. Vector $\mathbf{x}\in\mathbb{R}^{\bar{n}}$ is the vector of the (primal) decision variables.

Let us consider the Lagrange function, $L:\mathbb{R}^{\bar{n}+\bar{m}+\bar{n}}_\oplus\to\mathbb{R}$ assigned to the $(P-QP)$ problem as

$$L(\mathbf{x},\mathbf{y},\mathbf{s})=\frac{1}{2}\mathbf{x}^T Q\mathbf{x}+\mathbf{c}^T\mathbf{x}+\mathbf{y}^T(A\mathbf{x}-\mathbf{b})-\mathbf{s}^T\mathbf{x}.$$

The first order optimality conditions, the Karush-Kuhn-Tucker constraints [2, 19] can be derived as

$$\begin{array}{rcl} A\mathbf{x}+\quad \mathbf{z} & = \mathbf{b}, & \mathbf{x}\ge\mathbf{0},\ \mathbf{z}\ge\mathbf{0},\\ -A^T\mathbf{y}-\ Q\mathbf{x}+\quad\quad \mathbf{s} & = \mathbf{c}, & \mathbf{y}\ge\mathbf{0},\ \mathbf{s}\ge\mathbf{0},\\ \mathbf{x}^T\mathbf{s}+\ \mathbf{y}^T\mathbf{z} & = 0, & \end{array}\tag{1}$$

where $\mathbf{z}=A\mathbf{x}-\mathbf{b}\in\mathbb{R}^{\bar{m}}_\oplus$. By introducing the following notations

$$M=\begin{bmatrix} Q & A^T\\ -A & O\end{bmatrix}\in\mathbb{R}^{(\bar{n}+\bar{m})\times(\bar{n}+\bar{m})},\quad \mathbf{q}=\begin{pmatrix}\mathbf{c}\\ \mathbf{b}\end{pmatrix},\ \mathbf{u}=\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix},\ \mathbf{v}=\begin{pmatrix}\mathbf{s}\\ \mathbf{z}\end{pmatrix}\in\mathbb{R}^{\bar{n}+\bar{m}},$$

and $n=\bar{n}+\bar{m}$. We can define a special LCP, as the *first order optimality criteria* of the $(P-QP)$ problem.

The second linearly constrained convex quadratic programming model was introduced by Klafszky and Terlaky [17] as

$$\left.\begin{array}{rl}\min\ \mathbf{c}^T\mathbf{x}\ +\ \frac{1}{2}\mathbf{x}^T C^T C\,\mathbf{x}\ +\ \frac{1}{2}\mathbf{z}^T\mathbf{z} & \\ A\mathbf{x}\ +\quad B\mathbf{z} & \ge\mathbf{b}\\ \mathbf{x} & \ge\mathbf{0}\end{array}\right\}\quad (P-QP_{KT}),$$

$$\left.\begin{array}{rl}\max\ \mathbf{y}^T\mathbf{b}\ -\ \frac{1}{2}\mathbf{y}^T BB^T\mathbf{y}\ -\ \frac{1}{2}\mathbf{w}^T\mathbf{w} & \\ \mathbf{y}^T A\ -\quad \mathbf{w}^T C & \le\mathbf{c}\\ \mathbf{y} & \ge\mathbf{0}\end{array}\right\}\quad (D-QP_{KT}),$$

where $A \in \mathbb{R}^{\bar{m}\times\bar{n}}$, $B \in \mathbb{R}^{\bar{m}\times\bar{k}}$, $C \in \mathbb{R}^{\bar{l}\times\bar{n}}$ are given matrices, and $\mathbf{c},\mathbf{x} \in \mathbb{R}^{\bar{n}}$, $\mathbf{b},\mathbf{y} \in \mathbb{R}^{\bar{m}}$, $\mathbf{z} \in \mathbb{R}^{\bar{k}}$, $\mathbf{w} \in \mathbb{R}^{\bar{l}}$ are vectors used in the problem description. It is easy to derive the *weak duality theorem* in the following form.

Proposition 2.1. *Let $(\mathbf{x},\mathbf{z})$ and $(\mathbf{y},\mathbf{w})$ be arbitrary primal and dual feasible solution. Then*

$$\mathbf{c}^T\mathbf{x} + \frac{1}{2}\,\mathbf{x}^T C^T C\,\mathbf{x} + \frac{1}{2}\,\mathbf{z}^T\mathbf{z} \ \geq\ \mathbf{y}^T\mathbf{b} - \frac{1}{2}\,\mathbf{y}^T BB^T\mathbf{y} - \frac{1}{2}\,\mathbf{w}^T\mathbf{w},$$

holds. The previous inequality is satisfied with equality, if and only if

$$\mathbf{w} = C\mathbf{x},\ \ \mathbf{z} = B^T\mathbf{y}, \qquad \text{and} \qquad \mathbf{r}^T\mathbf{y} = 0,\ \ \mathbf{s}^T\mathbf{x} = 0,$$

are fulfilled, where $\mathbf{r} = A\mathbf{x} + B\mathbf{z} - \mathbf{b}$ and $\mathbf{s} = \mathbf{c} - \mathbf{y}^T A + \mathbf{w}^T C$, are primal and dual slack variables.

Those primal- and dual feasible solutions that satisfies the weak duality inequality with equality are called *optimal primal- and dual solutions*. From those constraints that ensure the equality of the previous inequality can be derived the following *linear complementarity problem, (BLCP)*

$$\left.\begin{array}{rcl} -P\mathbf{y} - A\mathbf{x} + \bar{\mathbf{y}} & = & -\mathbf{b} \\ A^T\mathbf{y} - Q\mathbf{x} + \bar{\mathbf{x}} & = & \mathbf{c} \\ \mathbf{x},\ \mathbf{y},\ \bar{\mathbf{x}},\ \bar{\mathbf{y}} & \geq & \mathbf{0} \\ \mathbf{x}\bar{\mathbf{x}} = \mathbf{0}, \quad \mathbf{y}\bar{\mathbf{y}} & = & \mathbf{0} \end{array}\right\} \quad (BLCP),$$

where $P = BB^T$ and $Q = C^T C$ positive semidefinite matrices. The $(BLCP)$ is the corresponding Karush–Kuhn–Tucker system to $(P - QP_{KT})$ and $(D - QP_{KT})$ problems [2, 17]. Let us denote by M the matrix of the linear system of $(BLCP)$, then the matrix M has the following structure

$$M = \begin{bmatrix} P & A \\ -A^T & Q \end{bmatrix}$$

and introduce the following notations $n := \bar{m} + \bar{n}$, $\mathbf{q} := \binom{-\mathbf{b}}{\mathbf{c}}$, $\mathbf{u} := \binom{\mathbf{x}}{\mathbf{y}}$ and $\mathbf{v} := \binom{\bar{\mathbf{y}}}{\bar{\mathbf{x}}}$. The $(BLCP)$ problem with the given bisymmetric matrix M[2] is an (LCP) with a special structure.

Let us note here that we can define a corresponding (WLCP) problem by changing the right-hand side of the last equations in (1) or in (BLCP) from $\mathbf{0}$ to a nonnegative vector $\mathbf{p}$.

3 From weighted central path problem to a sequence of convex feasibility problems

3.1 Central path and weighted central path

In this paper, we assume that the matrix M is positive semidefinite. It is easy to show that this matrix class coincides with the class of $P_*(0)$-matrices, that is a subclass of $P_*(\kappa)$-matrices introduced by Kojima et al. [18], where $\kappa \geq 0$.

We denote by $\mathscr{F} = \{(\mathbf{u},\mathbf{v}) \in \mathbb{R}^n_\oplus \times \mathbb{R}^n_\oplus : -M\mathbf{u} + \mathbf{v} = \mathbf{q}\}$ and $\mathscr{F}^+ = \{(\mathbf{u},\mathbf{v}) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ : -M\mathbf{u} + \mathbf{v} = \mathbf{q}\}$ the set of feasible and strictly feasible solutions of the WLCP. Note that these sets are the same for the LCP, namely when $\mathbf{p} = \mathbf{0}$. In contrast, the solution set of the WLCP depends on $\mathbf{p}$, let us denote it by $\mathscr{F}^*_{\mathbf{p}} = \{(\mathbf{u},\mathbf{v}) \in \mathscr{F} : \mathbf{u}\mathbf{v} = \mathbf{p}\}$.

For any $(\mathbf{u},\mathbf{v}) \in \mathscr{F}^+$ the complementarity condition of LCP could not be satisfied, since $\mathbf{u}\,\mathbf{v} > \mathbf{0}$. Thus, the complementarity condition needs to be relaxed. Now, we are ready to introduce the corresponding *central path problem (CPP)*

$$-M\mathbf{u} + \mathbf{v} = \mathbf{q}, \quad \mathbf{u},\mathbf{v} > \mathbf{0}, \quad \mathbf{u}\,\mathbf{v} = \mu\,\mathbf{e}, \tag{2}$$

for a given $\mu > 0$ and to define the *central path*

$$\mathscr{C} = \{(\mathbf{u},\mathbf{v}) \in \mathscr{F}^+ : \mathbf{u}\,\mathbf{v} = \mu\,\mathbf{e} \text{ for some } \mu > 0\}$$

[2] The

$$M = \begin{bmatrix} P & A \\ -A^T & Q \end{bmatrix} = \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} + \begin{bmatrix} 0 & A \\ -A^T & 0 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

is *bisymmetric matrix*, if P and Q are symmetric positive semidefinite matrices.

that contains all strictly feasible solutions of the LCP, that solves CPP for some $\mu > 0$. The central path for linear optimization problem has been introduced by Sonnevend [32] and Megiddo [20] independently. Illés et al. [13] gave an elementary proof for the following theorem:

Theorem 3.1. *Let the matrix M of the LCP be a given $\mathscr{P}_*(\kappa)$-matrix. Then the following statements are equivalent:*

i) $\mathscr{F}^+ \neq \emptyset$,

ii) $\forall\, \mathbf{w} \in \mathbb{R}^n_+,\ \exists!(\mathbf{u},\mathbf{v}) \in \mathscr{F}^+ :\ \mathbf{u}\,\mathbf{v} = \mathbf{w}$,

iii) $\exists!(\mathbf{u},\mathbf{v}) \in \mathscr{F}^+ :\ \mathbf{u}\,\mathbf{v} = \mu\,\mathbf{e}$.

Statement i) is called the *interior point condition (IPC)* of the LCP.

Statement ii) of the previous theorem defines a very important problem[3]:

$$-M\mathbf{u} + \mathbf{v} = \mathbf{q}, \quad \mathbf{u},\mathbf{v} > \mathbf{0}, \quad \mathbf{u}\,\mathbf{v} = \mathbf{w}, \tag{$WCPP$}$$

that we call the *weighted central path problem (WCPP)* for a given $\mathbf{w} \in \mathbb{R}^n_+$. The unique solution of the (WCPP) can be denoted by $(\mathbf{u}(w),\mathbf{v}(w)) \in \mathscr{F}^+$.

Statement iii) of the previous theorem says that the central path of an LCP with $\mathscr{P}_*(\kappa)$-matrix is unique.[4] In our case (M is $P_*(0)$-matrix), assuming that LCP satisfies the IPC, namely $\mathscr{F}^+ \neq \emptyset$, it can be proved that $\mathscr{F}^* \neq \emptyset$ is a compact (see Corollary 3.4. in [21]) and convex set (see Theorem 5 in [3] or Corollary 3.3. in [14]).

In the classical IPAs we consider WCPPs with $\mathbf{w} = \mu\,\mathbf{e}$, $\mu \to 0+$. By solving these approximately we tend to a solution of the LCP, namely we change the entries of $\mathbf{w}$ proportionally. This somehow restricts the possible optimization strategies. Moreover, as we will see later, the limiting value $\mathbf{w} = \mathbf{0}$ is not the only interesting target. In this paper, we show that the general theory of self-concordant functions gives the opportunity to justify different strategies for updating approximations to $(\mathbf{u}(w),\mathbf{v}(w))$ with unbalanced weights. However, for that, we need to introduce the trajectory $(\mathbf{u}(w),\mathbf{v}(w))$ not by the weighted barriers, but by a kind of *method of centers* [10, 30] (as it was done in [25] for the linear programming problems). This is the subject of the next section.

3.2 Building the model for parabolic target space IPA

Since we assume that the matrix M is positive semidefinite, the (WCPP) has a unique solution based on the previous subsection. Now following the idea presented in [25], we are ready to define the *relaxed weighted central path problem (RWCPP)* with slight modification of (WCPP) as follows

$$-M\mathbf{u} + \mathbf{v} = \mathbf{q}, \quad \mathbf{u},\mathbf{v} \geq \mathbf{0}, \quad \mathbf{u}\,\mathbf{v} \geq \mathbf{w}^2, \tag{$RWCPP$}$$

where $\mathbf{w} \in \mathbb{R}^n$ thus $\mathbf{w}^2 \geq \mathbf{0}$ serves as a possible vector $\mathbf{w}$ in ($WCPP$). Clearly, the solution set of ($RWCPP$) contains the solution of the corresponding ($WCPP$), and it is non-empty if the IPC holds. Furthermore, it is easy to show that the nonlinear inequalities

$$u_i\,v_i - w_i^2 \geq 0,$$

define a convex cone in $\mathbb{R}^3$ [25], thus the solution set of (RWCPP) is a convex set. Since M is a positive semidefinite matrix, $\mathbf{u}^T(M\mathbf{u} + \mathbf{q})$ is a convex function of $\mathbf{u}$, thus the level set

$$\mathscr{L}_{w_0} = \{(\mathbf{u},\mathbf{v}) \in \mathscr{F} : \mathbf{u}^T\mathbf{v} \leq w_0\},$$

for all $w_0 \in \mathbb{R}$ is a convex set. Furthermore, if $w_0 \geq 0$, then it is nonempty and compact. Taking into consideration the definition of (RWCPP) and the level set $\mathscr{L}_{w_0}$, we can define the following *convex feasibility problem (CFP)*

$$-M\mathbf{u} + \mathbf{v} = \mathbf{q}, \quad \mathbf{u},\mathbf{v} \geq \mathbf{0}, \quad \mathbf{u}\,\mathbf{v} \geq \mathbf{w}^2, \quad \text{and} \quad w_0 \geq \mathbf{u}^T\mathbf{v}. \tag{CFP}$$

The next statement follows from our construction and the unique solvability of (WCPP).

Proposition 3.1. *Let us assume that $\mathscr{F}_+ \neq \emptyset$. For a given pair of $(w_0,\mathbf{w})$ (CFP) has feasible solution if and only if $w_0 \geq \|\mathbf{w}\|^2$.*

From the linear constraint of the monotone WLCP, we can express the variable $\mathbf{v}$ as $\mathbf{v} = \mathbf{q} + M\,\mathbf{u}$ and we can reformulate (CFP) in the following way

$$\mathbf{q} + M\mathbf{u} \geq \mathbf{0}, \quad \mathbf{u} \geq \mathbf{0}, \quad \mathbf{u}\,(\mathbf{q} + M\mathbf{u}) \geq \mathbf{w}^2, \quad \text{and} \quad w_0 \geq \mathbf{q}^T\mathbf{u} + \mathbf{u}^T M\mathbf{u}. \tag{3}$$

[3] This problem is a ($WLCP$) with $\mathbf{w} > \mathbf{0}$.

[4] Since Illés et al. [13] did not publish their approach, the details of the proof can be found in the thesis of his former PhD student [21].

When $\mathbf{u}$ is a solution of the (RWCPP), then $w_0 \geq \|\mathbf{w}\|^2$ follows. The $w_0 \geq \|\mathbf{w}\|^2$ is an important inequality defined during our relaxation process for the new variables $\mathbf{t} = (w_0, \mathbf{w})$, that we call the variables in the *parabolic target space*

$$\mathscr{T} = \{\mathbf{t} = (w_0, \mathbf{w}) \in \mathbb{R}^{1+n} : w_0 \geq \|\mathbf{w}\|^2\},$$

to distinguish from the original variables $(\mathbf{u}, \mathbf{v})$ of the LCP. Clearly, $\mathscr{T}$ is a convex set.

Now we shall define all those vectors that satisfy the system of convex inequalities (3) as

$$\mathscr{F}_z = \{\mathbf{z} = (\mathbf{u}, \mathbf{t}) \in \mathbb{R}^n \times \mathbb{R}^{1+n} : \ \mathbf{q} + M\mathbf{u} \geq \mathbf{0},\ \mathbf{u} \geq \mathbf{0},\ \mathbf{u}\,(\mathbf{q} + M\mathbf{u}) \geq \mathbf{w}^2,\ w_0 \geq \mathbf{q}^T\mathbf{u} + \mathbf{u}^T M\mathbf{u}\}.$$

Clearly, from $\mathscr{F}^+ \neq \emptyset$ follows that $\mathscr{F}_z$, has an interior point solution, too. The convex set $\mathscr{F}_z$ admits a standard self-concordant barrier ($M_F = 1$)

$$F(\mathbf{z}) = F(\mathbf{u}, \mathbf{t}) = -\ln\left(w_0 - \mathbf{u}^T(\mathbf{q} + M\mathbf{u})\right) - \sum_{i=1}^{n} \ln\left(u_i\,(\mathbf{q} + M\mathbf{u})_i - w_i^2\right),$$

with barrier parameter $\nu_F = 2n + 1$.

We will use the restriction of the function F on the $\mathbf{u}$-space defined as $F_{\mathbf{t}} = F(.,\mathbf{t}) : \mathbb{R}^n \to \mathbb{R}$ for a fixed vector $\mathbf{t} \in \mathbb{R}^{1+n}$. Since F is a self-concordant function with $M_F = 1$, therefore using Theorem 8.1 the function $F_{\mathbf{t}}$ is a self-concordant function, with $M_{F_{\mathbf{t}}} = 1$, as well.

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

$$\phi(\mathbf{t}) = \min_{\mathbf{u} : (\mathbf{u}, \mathbf{t}) \in \mathscr{F}_z} F(\mathbf{u}, \mathbf{t}). \tag{4}$$

In what follows, we use the notation $\mathbf{z}(\mathbf{t}) = (\mathbf{u}(\mathbf{t}), \mathbf{t})$ for the optimal solution of (4), and $\mathbf{v}(\mathbf{t}) = M\mathbf{u}(\mathbf{t}) + \mathbf{q}$. Note that

$$\nabla\phi(\mathbf{t}) = \nabla_2 F(\mathbf{u}(\mathbf{t}), \mathbf{t}), \qquad \mathbf{t} \in \operatorname{dom}\phi, \tag{5}$$

$$\nabla^2\phi(\mathbf{t}) = \nabla^2_{22} F(\mathbf{u}(\mathbf{t}), \mathbf{t}) - \nabla^2_{21} F(\mathbf{u}(\mathbf{t}), \mathbf{t}) \left[\nabla^2_{11} F(\mathbf{u}(\mathbf{t}), \mathbf{t})\right]^{-1} \nabla^2_{12} F(\mathbf{u}(\mathbf{t}), \mathbf{t}). \tag{6}$$

As in [25], function $\phi(\cdot)$ has a closed form representation.

Theorem 3.2. *Let us assume that $\mathscr{F}_+ \neq \emptyset$, then $\operatorname{dom}\phi = \operatorname{int}(\mathscr{T}) \neq \emptyset$ and the optimization problem (4) has a unique solution, and the corresponding optimal vector $\mathbf{u}(\mathbf{t})$ satisfies the following equation*

$$\mathbf{u}(\mathbf{t})\,(\mathbf{q} + M\mathbf{u}(\mathbf{t})) = \mathbf{u}(\mathbf{t})\,\mathbf{v}(\mathbf{t}) = \mathbf{w}^2 + \frac{w_0 - \|\mathbf{w}\|^2}{n+1}\,\mathbf{e}. \tag{7}$$

Moreover, for all $\mathbf{t} \in \operatorname{dom}\phi$, we have

$$\phi(w_0, \mathbf{w}) = -(n+1)\ln\frac{w_0 - \|\mathbf{w}\|^2}{n+1}. \tag{8}$$

Proof. We follow the proof of Lemma 1 in [25]. Let $\mathbf{t} = (w_0, \mathbf{w}) \in \operatorname{int}\mathscr{T}$, $\varepsilon = \frac{1}{2n}(w_0 - \|\mathbf{w}\|^2)$ and $\bar{\mathbf{w}} = \mathbf{w}^2 + \varepsilon\mathbf{e}$. Using Theorem 3.1 there exists $\bar{\mathbf{u}} > \mathbf{0}$ and $\bar{\mathbf{v}} = M\bar{\mathbf{u}} + \mathbf{q} > \mathbf{0}$ such that $\bar{\mathbf{u}}\,\bar{\mathbf{v}} = \bar{\mathbf{w}}$. Note that for $(\bar{\mathbf{u}}, \mathbf{t})$ the function F is well defined. Moreover, from the unique correspondence between $\bar{\mathbf{u}}$ and $\bar{\mathbf{w}}$, the optimization problem (4) gets the form of

$$\phi(\mathbf{t}) = \min_{\bar{\mathbf{w}} > \mathbf{0}} \left[-\ln\left(w_0 - \sum_{i=1}^{n} \bar{w}_i\right) - \sum_{i=1}^{n} \ln\left(\bar{w}_i - w_i^2\right)\right]. \tag{9}$$

Note that using the first order optimality condition for the previous optimization problem,

$$\frac{1}{w_0 - \sum\limits_{i=1}^{n} \bar{w}_i} - \frac{1}{\bar{w}_i - w_i^2} = 0, \qquad i = 1, 2, \ldots, n$$

hold, and the optimal vector $\bar{\mathbf{w}}^*$ can be found from the following equations

$$w_0 - \sum_{i=1}^{n} \bar{w}_i^* = \bar{w}_i^* - w_i^2, \qquad i = 1, 2, \ldots, n.$$

That is, $\bar{\mathbf{w}}^* = \mathbf{w}^2 + \frac{w_0 - \|\mathbf{w}\|^2}{n+1}\,\mathbf{e}$, proving (7). Furthermore,

$$\bar{w}_i^* - w_i^2 = \frac{w_0 - \|\mathbf{w}\|^2}{n+1}, \qquad i = 1,2,\ldots,n,$$

and then using (9), we obtain (8). □

It can be observed that in (7) and (8) the difference of w_0 and $\|\mathbf{w}\|^2$ appears. For this reason, let us introduce the function $\rho(\mathbf{t}) = \rho(w_0,\mathbf{w}) = w_0 - \|\mathbf{w}\|^2$, which in some sense could serve as a measure of the distance from the boundary of the parabolic target space $\mathscr{T}$.

Note that the relation (7) works in two ways. Indeed, for $\mathbf{t} \in \operatorname{dom}\phi$, we can easily compute the right-hand side of this equality, which gives us the exact value of the product of unknown vectors $\mathbf{u}(\mathbf{t})$ and $\mathbf{v}(\mathbf{t}) = \mathbf{q} + M\mathbf{u}(\mathbf{t})$.

On the other hand, if we have $\mathbf{u} > \mathbf{0}$ with $\mathbf{v} = M\mathbf{u} + \mathbf{q} > \mathbf{0}$, that are (strictly) feasible solutions of the WLCP, then it is always possible to find a vector $\mathbf{t}(\mathbf{u}) \in \operatorname{dom}\phi$, such that $\mathbf{u} = \mathbf{u}(\mathbf{t}(\mathbf{u}))$ and $\mathbf{v} = \mathbf{v}(\mathbf{t}(\mathbf{u}))$. Indeed, define

$$\xi(\mathbf{u}) = \min_{1\le i\le n} u_i\,(\mathbf{q} + M\mathbf{u})_i > 0.$$

Then we can define

$$\mathbf{w}(\mathbf{u}) = \Big[\mathbf{u}\,(\mathbf{q} + M\mathbf{u}) - \xi(\mathbf{u})\,\mathbf{e}\Big]^{1/2}, \qquad w_0(\mathbf{u}) = \mathbf{u}^T(\mathbf{q} + M\mathbf{u}) + \xi(\mathbf{u}). \tag{10}$$

It is easy to see that

$$\xi(\mathbf{u}) = \frac{w_0(\mathbf{u}) - \|\mathbf{w}(\mathbf{u})\|^2}{n+1}. \tag{11}$$

Therefore, denoting $\mathbf{t}(\mathbf{u}) = (w_0(\mathbf{u}), \mathbf{w}(\mathbf{u}))$, we have $\mathbf{u} = \mathbf{u}(\mathbf{t}(\mathbf{u}))$ and $\mathbf{v} = \mathbf{v}(\mathbf{t}(\mathbf{u}))$.

4 Parabolic Target-Space Interior Point Algorithm

In this section, we are interested in tracing the surface $\mathbf{u}(\mathbf{t})$ for $\mathbf{t} \in \mathscr{T}$. In fact, we cannot compute the point $\mathbf{u}(\mathbf{t})$ exactly. However, for our goals, it is sufficient to give the update strategies only for an approximation $\mathbf{u}$ to this surface. The closeness of $\mathbf{u}$ to $\mathbf{u}(\mathbf{t})$ can be measured by the (dual) local norm of the partial gradient

$$\lambda_{\mathbf{t}}(\mathbf{u}) := \|\nabla_1 F(\mathbf{u},\mathbf{t})\|_{\mathbf{u}}^* = \sqrt{\nabla_1 F(\mathbf{u},\mathbf{t})^T [\nabla_{11}^2 F(\mathbf{u},\mathbf{t})]^{-1} \nabla_1 F(\mathbf{u},\mathbf{t})}$$

based on the self-concordant property of function F. The (dual) local norm $\lambda_{\mathbf{t}}(\mathbf{u})$ is called the partial Newton decrement of the function F. Observe that $\lambda_{\mathbf{t}}(\mathbf{u})$ is the Newton decrement of the function $F_{\mathbf{t}} = F(.,\mathbf{t})$.[5]

Using $\lambda_{\mathbf{t}}(\mathbf{u})$ for a given solution $\mathbf{z} = (\mathbf{u},\mathbf{t}) \in \mathscr{F}_z$, we can define a neighbourhood of $\mathbf{u}(\mathbf{t})$:

$$\mathscr{N}_\lambda(\beta) = \{\mathbf{z} = (\mathbf{u},\mathbf{t}) \in \mathscr{F}_z :\ \lambda_{\mathbf{t}}(\mathbf{u}) \le \beta\}$$

which is called the λ*-neighbourhood* of the point $\mathbf{u}(\mathbf{t})$, where $0 < \beta < 1$.

However, since the exact value $F(\mathbf{u}(\mathbf{t}),\mathbf{t}) = \phi(\mathbf{t})$ is known, it is possible to use a *functional proximity measure*

$$\Psi(\mathbf{z}) = F(\mathbf{u},\mathbf{t}) - \phi(\mathbf{t}) \ge 0, \quad \mathbf{z} = (\mathbf{u},\mathbf{t}) \in \mathscr{F}_z.$$

Based on the functional proximity measure $\Psi(\mathbf{z})$ we define the following set:

$$\mathscr{W}(\gamma_1,\gamma_2) = \{\mathbf{z} = (\mathbf{u},\mathbf{t}) \in \mathscr{F}_z :\ \gamma_1 \le \Psi(\mathbf{z}) \le \gamma_2\},$$

where $0 \le \gamma_1 \le \gamma_2$.

Lemma 4.1. *Let* $0 < \beta < 1$ *and* $\mathbf{z} = (\mathbf{u},\mathbf{t}) \in \mathscr{N}_\lambda(\beta)$ *holds. Then*

$$\omega\big(\lambda_{\mathbf{t}}(\mathbf{u})\big) \le \Psi(\mathbf{z}) \ \le\ \omega_*\big(\lambda_{\mathbf{t}}(\mathbf{u})\big). \tag{12}$$

[5] The dual norm is computed similarly to the local norm, except that the inverse Hessian is used instead of the Hessian, namely $\lambda_{\mathbf{t}}(\mathbf{u}) := \|\nabla_1 F(\mathbf{u},\mathbf{t})\|_{[\nabla_{11}^2 F(\mathbf{u},\mathbf{t})]^{-1}}$. For more details, see the Appendix.

Proof. The function ω and its Fenchel conjugate ω_* are defined in the Appendix. Feasible solutions $\mathbf{z}(\mathbf{t}) = (\mathbf{u}(\mathbf{t}),\mathbf{t})$ and $\mathbf{z} = (\mathbf{u},\mathbf{t})$ are given. Since $\mathbf{z}(\mathbf{t}) = (\mathbf{u}(\mathbf{t}),\mathbf{t})$ is optimal solution of the problem (4) thus $\phi(\mathbf{t}) = F(\mathbf{z}(\mathbf{t})) = F(\mathbf{u}(\mathbf{t}),\mathbf{t})$ and $\nabla_1 F(\mathbf{u}(\mathbf{t}),\mathbf{t}) = \mathbf{0}$.

Now let us apply Theorem 8.13 to the self-concordant function $F_{\mathbf{t}}$ at points $\mathbf{u},\mathbf{u}(\mathbf{t}) \in \operatorname{dom} F_{\mathbf{t}}$. From the inequality (51), with $M_{F_{\mathbf{t}}} = 1$, follows that

$$F_{\mathbf{t}}(\mathbf{u}) \geq F_{\mathbf{t}}(\mathbf{u}(\mathbf{t})) + (\mathbf{u}-\mathbf{u}(\mathbf{t}))^T \nabla F_{\mathbf{t}}(\mathbf{u}(\mathbf{t})) + \omega(\|\nabla F_{\mathbf{t}}(\mathbf{u}) - \nabla F_{\mathbf{t}}(\mathbf{u}(\mathbf{t}))\|_{\mathbf{u}}^*).$$

Clearly, $\nabla F_{\mathbf{t}}(\mathbf{u}(\mathbf{t})) = \nabla_1 F(\mathbf{u}(\mathbf{t}),\mathbf{t}) = \mathbf{0}$, thus the second term on the right side of the inequality is 0. Rearranging the inequality, we get

$$F(\mathbf{u},\mathbf{t}) - F(\mathbf{u}(\mathbf{t}),\mathbf{t}) = F_{\mathbf{t}}(\mathbf{u}) - F_{\mathbf{t}}(\mathbf{u}(\mathbf{t})) \geq \omega(\|\nabla F_{\mathbf{t}}(\mathbf{u}) - \nabla F_{\mathbf{t}}(\mathbf{u}(\mathbf{t}))\|_{\mathbf{u}}^*) = \omega(\|\nabla_1 F(\mathbf{u},\mathbf{t})\|_{\mathbf{z}}^*).$$

Therefore

$$\Psi(\mathbf{z}) = F(\mathbf{u},\mathbf{t}) - \phi(\mathbf{t}) \geq \omega(\|\nabla_1 F(\mathbf{u},\mathbf{t})\|_{\mathbf{z}}^*) = \omega\big(\lambda_{\mathbf{t}}(\mathbf{u})\big),$$

proving the lower bound in (12).

Taking into consideration that $\mathbf{z} = (\mathbf{u},\mathbf{t}) \in \mathscr{N}_\lambda(\beta)$, we have $\lambda_{\mathbf{t}}(\mathbf{u}) \leq \beta < 1$, so the second inequality of Theorem 8.13 can be used with similar computations as before, getting the

$$\Psi(\mathbf{z}) \leq \omega_*\big(\lambda_{\mathbf{t}}(\mathbf{u})\big)$$

upper bound in (12).

□

Now we are ready to present our, new *Adaptive Parabolic Target-Space Interior Point Algorithm (PTS IPA)* for monotone WLCPs.

4.1 PTS IPA for monotone WLCPs

We define a vector $\mathbf{w}^*$ to the weight vector $\mathbf{p}$ of the WLCP such that $\mathbf{p} = (\mathbf{w}^*)^2$. The vector $(\mathbf{w}^*)^2$ is called the target vector. The algorithm starts with a given point $\mathbf{z}^{(0)} = (\mathbf{u}^{(0)},\mathbf{t}^{(0)}) \in \operatorname{int}\mathscr{F}_z$, so it satisfies all inequalities in (3) as strict inequalities. The algorithm consists of two different types of iterations. In the outer loop, the goal is to use such a direction in the parabolic target space that ensures a large enough decrease in the stopping criteria, namely the new iterates come closer to the target vector $(\mathbf{w}^*)^2 \in \mathbb{R}^n_\oplus$ in some measure. Although, the result of such computation achieves its goal of better approximating the target value, but it may end up in a vector that is not well centered in the sense of the Newton decrement, $\lambda_{\mathbf{t}}(\mathbf{u})$. The goal of the inner iteration is to restore this "well centered" property of the computed new solution, namely to ensure that the computed solution belongs to the λ-neighbourhood $\mathscr{N}_\lambda(\beta)$ of the current point $\mathbf{u}(\mathbf{t})$. During the inner loop the parameter vector $\mathbf{t}$ of the parabolic target space does not change.

Adaptive Parabolic Target-Space Interior Point Algorithm for Monotone WLCPs

---

Input: the initial point $\mathbf{z}^{(0)} = (\mathbf{u}^{(0)},\mathbf{t}^{(0)}) \in \operatorname{int}\mathscr{F}_z$, the target vector $(\mathbf{w}^*)^2 \in \mathbb{R}^n_\oplus$ such that $(\mathbf{w}^*)^2 = \mathbf{p}$, the accuracy parameter $\varepsilon > 0$, the margins $\delta_u \geq \delta_l$, and the proximity level $\beta \in \left(0,\frac{1}{2}\right)$: $0 < \frac{2\beta^2}{1-2\beta} \leq \delta_l$.

Begin

  $\mathbf{z} := \mathbf{z}^{(0)}$.

  While $\|\mathbf{u}\,\mathbf{v} - (\mathbf{w}^*)^2\| > \varepsilon$ **do**

   Choose a target direction $\Delta\mathbf{t} \in \mathbb{R}^{n+1} \setminus \{\mathbf{0}\}$.

   Compute direction $\mathbf{d} = (\Delta\mathbf{u},\Delta\mathbf{t}) \in \mathbb{R}^{2n+1}$ as

$$\mathbf{d} = \begin{pmatrix} \Delta\mathbf{u} \\ \Delta\mathbf{t} \end{pmatrix} = \begin{pmatrix} -[\nabla^2_{11}F(\mathbf{z})]^{-1}\nabla^2_{12}F(\mathbf{z})\,\Delta\mathbf{t} \\ \Delta\mathbf{t} \end{pmatrix}.$$

   Find the step length $\alpha > 0$ such that $\hat{\mathbf{z}} := \mathbf{z} + \alpha\,\mathbf{d} \in \mathscr{W}(\delta_l,\delta_u)$.

   Set $\hat{\mathbf{t}} := \mathbf{t} + \alpha\,\Delta\mathbf{t}$ and $\hat{\mathbf{u}} := \mathbf{u} + \alpha\,\Delta\mathbf{u}$.

   While $\lambda_{\hat{\mathbf{t}}}(\hat{\mathbf{u}}) > \beta$ **do**

$\Delta\hat{\mathbf{u}} = -\left[\nabla^2_{11} F(\hat{\mathbf{u}},\hat{\mathbf{t}})\right]^{-1} \nabla_1 F(\hat{\mathbf{u}},\hat{\mathbf{t}})$ and $\hat{\alpha} = \frac{1}{1+\lambda_{\hat{\mathbf{t}}}(\hat{\mathbf{u}})}$.

$\hat{\mathbf{u}} := \hat{\mathbf{u}} + \hat{\alpha}\Delta\hat{\mathbf{u}}$.

End While

$\mathbf{z} := (\hat{\mathbf{u}},\hat{\mathbf{t}})$.

End While

End

In the analysis of our predictor-corrector strategy, we use both measurements $\lambda_{\mathbf{t}}(\mathbf{u})$ and $\Psi(\mathbf{z})$, as well. Let us analyze the performance of this scheme, the PTS IPA for monotone WLCPs.

It is clear that after the inner loop of the PTS IPA the solution $\mathbf{z} = (\mathbf{u},\mathbf{t}) \in \mathcal{N}_\lambda(\beta)$.

Analysing the step of the outer loop of the PTS IPA, we find that the solution $\hat{\mathbf{z}} = (\hat{\mathbf{u}},\hat{\mathbf{t}}) \in \mathcal{W}(\delta_l,\delta_u)$, due to the fact that the function F is a self-concordant barrier, thus the step length $\alpha > 0$ can be chosen to satisfy the inequality $\delta_l \le \Psi(\hat{\mathbf{z}}) \le \delta_u$.

4.2 Analysis of the corrector step

First of all, note that the process at the inner while loop (corrector step) is a standard Damped Newton Method (DNM) (see Appendix or pages 348-349 in [26]).

Lemma 4.2. *Let us assume that $\hat{\mathbf{z}} \in \mathcal{W}(\delta_l,\delta_u) \setminus \mathcal{N}_\lambda(\beta)$, where $\beta \in (0,\frac{1}{2})$: $0 < \frac{2\beta^2}{1-2\beta} \le \delta_l$. Let k be the number of steps in an inner loop. Then we need*

$$k \le \left\lfloor \frac{\delta_u}{\omega(\beta)} \right\rfloor + 1$$

corrector steps in order to have $\mathbf{z} \in \mathcal{N}_\lambda(\beta)$.

Proof. In the inner while loop we minimizing the self-concordant function $F_{\hat{\mathbf{t}}}$, where $\hat{\mathbf{z}} = (\hat{\mathbf{u}},\hat{\mathbf{t}}) \in \mathcal{F}_z$.

Based on Theorem 8.14, the full Newton step is feasible. Since $\operatorname{dom} F_{\hat{\mathbf{t}}}$ is convex, a damped Newton step also gives a feasible solution. For two consecutive solutions of the inner loop $\bar{\mathbf{z}} = (\bar{\mathbf{u}},\hat{\mathbf{t}})$ and $\mathbf{z}^+ = (\mathbf{u}^+,\hat{\mathbf{t}})$ from Theorem 8.15 and using the monotone increasing property of the function ω follows that

$$F_{\hat{\mathbf{t}}}(\bar{\mathbf{u}}) - F_{\hat{\mathbf{t}}}(\mathbf{u}^+) = F(\bar{\mathbf{u}},\hat{\mathbf{t}}) - F(\mathbf{u}^+,\hat{\mathbf{t}}) \ge \omega\big(\lambda_{\hat{\mathbf{t}}}(\bar{\mathbf{u}})\big) \ge \omega(\beta), \tag{13}$$

hence the decrease of the self-concordant function F after each step in the inner loop is at least $\omega(\beta) > 0$, until we reach the neighborhood $\mathcal{N}_\lambda(\beta)$, thus k is finite. Since k is the first index when $\hat{\mathbf{z}}^{(k)} \in \mathcal{N}_\lambda(\beta)$, the iterates $\hat{\mathbf{z}} = \hat{\mathbf{z}}^{(0)} = (\hat{\mathbf{u}}^{(0)},\hat{\mathbf{t}})$, $\hat{\mathbf{z}}^{(1)} = (\hat{\mathbf{u}}^{(1)},\hat{\mathbf{t}}), \ldots, \hat{\mathbf{z}}^{(k-1)} = (\hat{\mathbf{u}}^{(k-1)},\hat{\mathbf{t}}) \notin \mathcal{N}_\lambda(\beta)$ and $\hat{\mathbf{z}}^{(k)} = (\hat{\mathbf{u}}^{(k)},\hat{\mathbf{t}}) \in \mathcal{N}_\lambda(\beta)$. On the other hand, we know that the starting point of the inner loop $\hat{\mathbf{z}}^{(0)} = (\hat{\mathbf{u}}^{(0)},\hat{\mathbf{t}}) \in \mathcal{W}(\delta_l,\delta_u)$, thus

$$\delta_u \ge \Psi(\hat{\mathbf{z}}^{(0)}) = F(\hat{\mathbf{u}}^{(0)},\hat{\mathbf{t}}) - \phi(\hat{\mathbf{t}}) \ge F(\hat{\mathbf{u}}^{(0)},\hat{\mathbf{t}}) - F(\hat{\mathbf{u}}^{(i)},\hat{\mathbf{t}}), \qquad i = 1,\ldots,k-1, \tag{14}$$

since $\phi(\hat{\mathbf{t}}) \le F(\hat{\mathbf{u}}^{(i)},\hat{\mathbf{t}})$ for any $\hat{\mathbf{z}}^{(i)} = (\hat{\mathbf{u}}^{(i)},\hat{\mathbf{t}}) \in \mathcal{F}_z$ $(i = 1,2,\ldots,k-1)$. Using (13) and (14), we have

$$\delta_u \ge F(\hat{\mathbf{u}}^{(0)},\hat{\mathbf{t}}) - F(\hat{\mathbf{u}}^{(k-1)},\hat{\mathbf{t}}) \ge (k-1)\,\omega(\beta). \tag{15}$$

Now the iteration bound of the inner loop follows. □

4.3 Analysis of the predictor step

Let us show now that the step length $\alpha_k > 0$ computed at the k^{th} predictor iteration is sufficiently big. For that, we need to compare the size of the steps in z-space $\left(\mathbb{R}^{2n+1}\right)$ with that in t-space $\left(\mathbb{R}^{n+1}\right)$. The local norm in z-space is defined using the self-concordant function F and the set $\operatorname{int}\mathcal{F}_z$, while in t-space using ϕ and the set $\operatorname{int}\mathcal{T}$.

First, we show that if $\mathbf{z}$ is in the neighborhood $\mathcal{N}_\lambda(\beta)$ then $\mathbf{z}(\mathbf{t})$ is close to $\mathbf{z}$ and it is also true for the gradient vectors.

Lemma 4.3. *Let $\mathbf{z} = (\mathbf{u},\mathbf{t}) \in \mathcal{N}_\lambda(\beta)$, where $\beta \in (0,\frac{1}{2})$. Then*

$$\|\mathbf{z} - \mathbf{z}(\mathbf{t})\|_{\nabla^2 F(\mathbf{z})} \le \frac{\lambda_{\mathbf{t}}(\mathbf{u})}{1-\lambda_{\mathbf{t}}(\mathbf{u})} \le \frac{\beta}{1-\beta} < 1, \tag{16}$$

$$\|\nabla F(\mathbf{z}) - \nabla F(\mathbf{z}(\mathbf{t}))\|_{[\nabla^2 F(\mathbf{z})]^{-1}} \;\leq\; \frac{\lambda_{\mathbf{t}}(\mathbf{u})}{1-2\lambda_{\mathbf{t}}(\mathbf{u})} \leq \frac{\beta}{1-2\beta}. \tag{17}$$

Proof. By definition, $\|\mathbf{z}-\mathbf{z}(\mathbf{t})\|_{\nabla^2 F(\mathbf{z})} = \|\mathbf{u}-\mathbf{u}(\mathbf{t})\|_{\nabla^2 F_t(\mathbf{u})}$. Using Corollary 8.7 with self-concordant function $F_{\mathbf{t}}$, we get

$$\frac{\|\mathbf{u}-\mathbf{u}(\mathbf{t})\|_{\nabla^2 F_t(\mathbf{u})}}{1+\|\mathbf{u}-\mathbf{u}(\mathbf{t})\|_{\nabla^2 F_t(\mathbf{u})}} \leq \lambda_{\mathbf{t}}(\mathbf{u}).$$

Since $\mathbf{z} \in \mathscr{N}_\lambda(\beta)$, $\lambda_{\mathbf{t}}(\mathbf{u}) \leq \beta < 1$, therefore (16) is proved by rearranging the previous inequality and then substituting the upper bound β on $\lambda_{\mathbf{t}}(\mathbf{u})$.

In a similar way, we can give an estimation for the gradient. Applying Corollary 8.11, we get

$$\frac{\|\nabla F(\mathbf{z}) - \nabla F(\mathbf{z}(\mathbf{t}))\|_{[\nabla^2 F(\mathbf{z})]^{-1}}}{1+\|\nabla F(\mathbf{z}) - \nabla F(\mathbf{z}(\mathbf{t}))\|_{[\nabla^2 F(\mathbf{z})]^{-1}}} \leq \|\mathbf{z}-\mathbf{z}(\mathbf{t})\|_{\nabla^2 F(\mathbf{z})}.$$

Since $\|\mathbf{z}-\mathbf{z}(\mathbf{t})\|_{\nabla^2 F(\mathbf{z})} < 1$, we can reformulate it as (17) and again substitute the upper bound β. □

To give an estimation on the length of the predictor direction $\mathbf{d}$ by the length of $\Delta\mathbf{t}$ (so considering only its t-space part), we need one simple fact on the monotonicity of Schur complement.

Lemma 4.4 (Schur monotonicity lemma)**.** *Consider two symmetric matrices*

$$Q_i = \begin{bmatrix} A_i & B_i \\ B_i^T & C_i \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}, \qquad i=1,2,$$

such that $0 \prec Q_1 \preceq Q_2$. *Then,* $C_1 - B_1^T A_1^{-1} B_1 \;\preceq\; \left[C_2 - B_2^T A_2^{-1} B_2\right]$.

Proof. Let $X_i = A_i^{-1} B_i$ and $S_i = C_i - B_i^T A_i^{-1} B_i$ $(i=1,2)$, then the well-known identity says

$$Q_i = \begin{bmatrix} I & O \\ X_i^T & I \end{bmatrix} \begin{bmatrix} A_i & O \\ O & S_i \end{bmatrix} \begin{bmatrix} I & X_i \\ O & I \end{bmatrix}, \qquad i=1,2,$$

where O is the zero matrix and I is the identity matrix with appropriate size. This means by $A_i \succ 0$ that

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}^T Q_i \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = (\mathbf{u}+X_i\mathbf{v})^T A_i (\mathbf{u}+X_i\mathbf{v}) + \mathbf{v}^T S_i \mathbf{v} \geq \mathbf{v}^T S_i \mathbf{v}, \qquad i=1,2,$$

and the quadratic form is minimal in $\mathbf{u} = -X_i\mathbf{v}$. Therefore, for any $\mathbf{v} \in \mathbb{R}^m$,

$$\mathbf{v}^T\left[S_2 - S_1\right]\mathbf{v} \;=\; \begin{bmatrix} -X_2\mathbf{v} \\ \mathbf{v} \end{bmatrix}^T Q_2 \begin{bmatrix} -X_2\mathbf{v} \\ \mathbf{v} \end{bmatrix} - \begin{bmatrix} -X_1\mathbf{v} \\ \mathbf{v} \end{bmatrix}^T Q_1 \begin{bmatrix} -X_1\mathbf{v} \\ \mathbf{v} \end{bmatrix} \geq \begin{bmatrix} -X_2\mathbf{v} \\ \mathbf{v} \end{bmatrix}^T (Q_2 - Q_1) \begin{bmatrix} -X_2\mathbf{v} \\ \mathbf{v} \end{bmatrix} \geq 0,$$

since $Q_2 - Q_1 \succeq 0$. □

Lemma 4.5. *Let* $\mathbf{z} = (\mathbf{u},\mathbf{t}) \in \mathscr{N}_\lambda(\beta)$, *where* $\beta \in (0,\frac{1}{2})$ *and* $\mathbf{d}$ *be the predictor direction of PTS IPA. Then*

$$\|\mathbf{d}\|_{\nabla^2 F(\mathbf{z})} \leq \frac{1-\beta}{1-2\beta}\|\Delta\mathbf{t}\|_{\nabla^2\phi(\mathbf{t})}. \tag{18}$$

Proof. We know that the Hessian $\nabla^2 F(\mathbf{z})$ is nondegenerate for any $\mathbf{z} \in \operatorname{int}\mathscr{F}_z$ (see Theorem 8.3) and based on the result of Theorem 8.5 and (16), we have

$$\nabla^2 F(\mathbf{z}) \preceq \frac{1}{(1-\|\mathbf{z}-\mathbf{z}(\mathbf{t})\|_{\nabla^2 F(\mathbf{z})})^2}\nabla^2 F(\mathbf{z}(\mathbf{t})) \preceq \left(\frac{1-\beta}{1-2\beta}\right)^2 \nabla^2 F(\mathbf{z}(\mathbf{t})).$$

Therefore using the monotonicity of Schur complement (Lemma 4.4), we complete the proof:

$$\|\mathbf{d}\|^2_{\nabla^2 F(\mathbf{z})} \;=\; \mathbf{d}^T \nabla^2 F(\mathbf{z}) \mathbf{d} = \begin{pmatrix} \Delta\mathbf{u} \\ \Delta\mathbf{t} \end{pmatrix}^T \begin{pmatrix} \nabla^2_{11} F(\mathbf{z}) & \nabla^2_{12} F(\mathbf{z}) \\ \nabla^2_{21} F(\mathbf{z}) & \nabla^2_{22} F(\mathbf{z}) \end{pmatrix} \begin{pmatrix} \Delta\mathbf{u} \\ \Delta\mathbf{t} \end{pmatrix}$$

$$= \Delta\mathbf{t}^T\left(\nabla^2_{22}F(\mathbf{z}) - \nabla^2_{21}F(\mathbf{z})\,[\nabla^2_{11}F(\mathbf{z})]^{-1}\,\nabla^2_{12}F(\mathbf{z})\right)\Delta\mathbf{t}$$
$$\le \left(\frac{1-\beta}{1-2\beta}\right)^2 \Delta\mathbf{t}^T\left(\nabla^2_{22}F(\mathbf{z}(\mathbf{t})) - \nabla^2_{21}F(\mathbf{z}(\mathbf{t}))\,[\nabla^2_{11}F(\mathbf{z}(\mathbf{t}))]^{-1}\,\nabla^2_{12}F(\mathbf{z}(\mathbf{t}))\right)\Delta\mathbf{t}$$
$$= \left(\frac{1-\beta}{1-2\beta}\right)^2 \Delta\mathbf{t}^T\,\nabla^2\phi(\mathbf{t})\,\Delta\mathbf{t} = \left(\frac{1-\beta}{1-2\beta}\right)^2 \|\Delta\mathbf{t}\|^2_{\nabla^2\phi(\mathbf{t})}.$$

□

Now we are ready to give a lower bound on the displacement in the t-space.

Theorem 4.6. *In the predictor step of the PTS IPA, if* $\mathbf{z} = (\mathbf{u},\mathbf{t}) \in \mathscr{N}_\lambda(\beta)$*, with* $\beta \in (0, \frac{1}{2})$*, and* $\alpha > 0$ *is such that* $\mathbf{z} + \alpha\,\mathbf{d} \in \mathscr{W}(\delta_l, \delta_u)$ *holds, then*

$$\alpha\,\|\Delta\mathbf{t}\|_{\nabla^2\phi(\mathbf{t})} \;\ge\; \frac{1-2\beta}{1-\beta}\,\omega_*^{-1}\left(\frac{1}{2}\left[\delta_l - \frac{2\beta^2}{1-2\beta}\right]\right) \;>\; 0, \tag{19}$$

where $\Delta\mathbf{t}$ *is the chosen target direction.*

Proof. On the contrary, assume that

$$\alpha\,\|\Delta\mathbf{t}\|_{\nabla^2\phi(\mathbf{t})} \;<\; \frac{1-2\beta}{1-\beta}\,\omega_*^{-1}\left(\frac{1}{2}\left[\delta_l - \frac{2\beta^2}{1-2\beta}\right]\right).$$

Using that $\omega_*^{-1}(x) < 1$ for $x > 0$, it means that $\alpha\,\|\Delta\mathbf{t}\|_{\nabla^2\phi(\mathbf{t})} < \frac{1-2\beta}{1-\beta}$. Hence, based on (18), this assumption means that $\alpha\,\|\mathbf{d}\|_{\nabla^2F(\mathbf{z})} < 1$.

Now using this fact and that F is self-concordant function, from Theorem 8.12, inequality (50) (see the Appendix) for solutions $\mathbf{z}$ and $\mathbf{z}+\alpha\,\mathbf{d}$, we have the following inequality

$$F(\mathbf{z}+\alpha\,\mathbf{d}) \le F(\mathbf{z}) + \alpha\,\mathbf{d}^T\nabla F(\mathbf{z}) + \omega_*\big(\alpha\,\|\mathbf{d}\|_{\nabla^2F(\mathbf{z})}\big). \tag{20}$$

Due to the fact that ϕ defined by (4) is a convex function,

$$\phi(\mathbf{t}+\alpha\,\Delta\mathbf{t}) \ge \phi(\mathbf{t}) + \alpha\,\Delta\mathbf{t}^T\nabla\phi(\mathbf{t}) = \phi(\mathbf{t}) + \alpha\,\mathbf{d}^T\nabla F(\mathbf{z}(\mathbf{t})). \tag{21}$$

After subtracting (21) from (20), we get

$$\Psi(\mathbf{z}+\alpha\,\mathbf{d}) = F(\mathbf{z}+\alpha\,\mathbf{d}) - \phi(\mathbf{t}+\alpha\,\Delta\mathbf{t}) \le \Psi(\mathbf{z}) + \alpha\,\mathbf{d}^T\big(\nabla F(\mathbf{z}) - \nabla F(\mathbf{z}(\mathbf{t}))\big) + \omega_*\big(\alpha\,\|\mathbf{d}\|_{\nabla^2F(\mathbf{z})}\big).$$

Remember that in the predictor step of PTS IPA the step length α has been chosen to satisfy the following condition $\mathbf{z}+\alpha\,\mathbf{d} \in \mathscr{W}(\delta_l, \delta_u)$, namely $\delta_l \le \Psi(\mathbf{z}+\alpha\,\mathbf{d})$. On the other hand, $\mathbf{z} \in \mathscr{N}_\lambda(\beta)$. From Lemma 4.1, we have $\Psi(\mathbf{z}) \le \omega_*(\lambda_\mathbf{t}) \le \omega_*(\beta)$ using the monotone increasing property of ω_*. Combining these estimations, we have

$$\delta_l \le \omega_*(\beta) + \alpha\,\|\mathbf{d}\|_{\nabla^2F(\mathbf{z})}\,\|\nabla F(\mathbf{z}) - \nabla F(\mathbf{z}(\mathbf{t}))\|_{[\nabla^2F(\mathbf{z})]^{-1}} + \omega_*\left(\alpha\,\|\mathbf{d}\|_{\nabla^2F(\mathbf{z})}\right).$$

Now taking into consideration that $\alpha\,\|\mathbf{d}\|_{\nabla^2F(\mathbf{z})} < 1$, we can apply Theorem 8.4 and get

$$\delta_l \le \omega_*(\beta) + \omega\left(\|\nabla F(\mathbf{z}) - \nabla F(\mathbf{z}(\mathbf{t}))\|_{[\nabla^2F(\mathbf{z})]^{-1}}\right) + 2\,\omega_*\left(\alpha\,\|\mathbf{d}\|_{\nabla^2F(\mathbf{z})}\right).$$

Functions ω and ω_* are monotone increasing, so we can use the upper bounds given in (17) and (18), which yields

$$\delta_l \le \omega_*(\beta) + \omega\left(\frac{\beta}{1-2\beta}\right) + 2\omega_*\left(\frac{1-\beta}{1-2\beta}\alpha\,\|\Delta\mathbf{t}\|_{\nabla^2\phi(\mathbf{t})}\right). \tag{22}$$

We provide an upper bound on the first two terms in (22) using the nonnegativity of the logarithm,

$$\omega_*(\beta) + \omega\left(\frac{\beta}{1-2\beta}\right) \;=\; -\beta - \ln(1-\beta) + \frac{\beta}{1-2\beta} - \ln\left(1 + \frac{\beta}{1-2\beta}\right) \le \frac{2\beta^2}{1-2\beta}.$$

We conclude from (22), that

$$0 < \delta_l - \frac{2\beta^2}{1-2\beta} \leq 2\,\omega_* \left(\frac{1-\beta}{1-2\beta}\alpha \,\|\Delta \mathbf{t}\|_{\nabla^2\phi(\mathbf{t})} \right), \tag{23}$$

which contradicts our initial assumption. □

Corollary 4.7. *In the predictor step of the PTS IPA, if* $\mathbf{z} = (\mathbf{u}, \mathbf{t}) \in \mathscr{N}_\lambda(\beta)$*, with* $\beta \in (0, \frac{1}{2})$*, and* $\alpha > 0$ *is such that* $\mathbf{z} + \alpha\,\mathbf{d} \in \mathscr{W}(\delta_l, \delta_u)$*, where* $\delta_l = (2+\kappa^2)\frac{\beta^2}{1-2\beta}$ *with some* $\kappa > 0$*, then*

$$\alpha \,\|\Delta \mathbf{t}\|_{\nabla^2\phi(\mathbf{t})} \;\geq \frac{\beta(1-2\beta)\kappa}{(1-\beta)(1-\beta+\beta\kappa)}, \tag{24}$$

where $\Delta \mathbf{t}$ *is the chosen target direction.*

Proof. Based on (19), with the given special value of δ and properties of the function ω_*, we have

$$\frac{1-\beta}{1-2\beta}\alpha \,\|\Delta \mathbf{t}\|_{\nabla^2\phi(\mathbf{t})} \;\geq \omega_*^{-1}\left(\frac{\kappa^2\beta^2}{2\,(1-2\beta)}\right) \geq \omega_*^{-1}\left(\frac{\kappa^2\beta^2}{2\,(1-\beta)^2}\right) \geq \omega_*^{-1}\left(\frac{\kappa^2\beta^2}{2\,(1-\beta)(1-\beta+\kappa\beta)}\right). \tag{25}$$

By Lemma 8.4, we get

$$\frac{\kappa^2\beta^2}{2\,(1-\beta)(1-\beta+\kappa\beta)} = \frac{\frac{\kappa^2\beta^2}{(1-\beta+\kappa\beta)^2}}{2\left(1-\frac{\kappa\beta}{1-\beta+\kappa\beta}\right)} \geq \omega_*\left(\frac{\kappa\beta}{1-\beta+\kappa\beta}\right).$$

Combining it with (25), we get (24). □

Remark 4.1. If $\beta = \frac{1}{4}, \delta_l = \frac{2+\kappa^2}{8}$, where $\kappa \geq 0$, then in each predictor step of the PTS IPA for monotone WLCPs, we have

$$\alpha\|\Delta \mathbf{t}\|_{\nabla^2\phi(\mathbf{t})} \geq \frac{2\,\kappa}{3\,(3+\kappa)}. \tag{26}$$

5 Complexity analysis of a PTS-IPA for monotone WLCPs

The goal of this section is to bound the number of outer iterations. As we have seen in Lemma 4.2 after finite number of corrector steps the solution satisfies $\mathbf{z} \in \mathscr{N}_\lambda(\beta)$. Thus the sequence of points $\mathbf{z}^{(0)}, \mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(k)}$ computed by the algorithm starting from the initial, feasible solution $\mathbf{z}^{(0)}$, all belongs to $\mathscr{N}_\lambda(\beta) \subset \mathscr{F}_z$. In particular, this means that

$$\mathbf{u}^{(k)} \geq \mathbf{0},\;\; M\mathbf{u}^{(k)} + \mathbf{q} \geq \mathbf{0},\;\; \mathbf{u}^{(k)}(M\mathbf{u}^{(k)} + \mathbf{q}) \geq \left(\mathbf{w}^{(k)}\right)^2,\;\; \left(\mathbf{u}^{(k)}\right)^T(M\mathbf{u}^{(k)} + \mathbf{q}) \leq w_0^{(k)}, \qquad k \geq 0.$$

Thus, if we have a target point $\mathbf{t}^* = (w_0^*, \mathbf{w}^*) \in \operatorname{dom}\phi$, it is enough to study the rate of convergence of $\mathbf{w}^{(k)} \to \mathbf{w}^*$. For obtaining the complexity result we focus on a special type of updating strategy in the target space, $\mathscr{T}$, namely the *Greedy Step*:

$$\mathbf{t}(\alpha) = \mathbf{t} + \alpha\,(\mathbf{t}^* - \mathbf{t}) = \alpha\,\mathbf{t}^* + (1-\alpha)\,\mathbf{t}, \tag{27}$$

where $\alpha \in (0, 1]$ and $\mathbf{t} = (w_0, \mathbf{w})$ is the current iterate, while $\Delta\mathbf{t} := \mathbf{t}^* - \mathbf{t}$.

Let $w_0^* := \|\mathbf{w}^*\|^2$, then $\rho(\mathbf{t}^*) = 0$,

namely the target point $\mathbf{t}^*$ is on the boundary of $\mathscr{T}$. We introduce the function $\ell^*(\mathbf{t}) := w_0 - w_0^* - 2\,\mathbf{w}^{*T}(\mathbf{w} - \mathbf{w}^*)$. Simple computations show that

$$\ell^*(\mathbf{t}) = w_0 - \|\mathbf{w}\|^2 + \|\mathbf{w} - \mathbf{w}^*\|^2 = \rho(\mathbf{t}) + \|\mathbf{w} - \mathbf{w}^*\|^2.$$

Thus, we have

$$\ell^*(\mathbf{t}) \geq \max\{w_0 - \|\mathbf{w}\|^2, \|\mathbf{w} - \mathbf{w}^*\|^2\} = \max\{\rho(\mathbf{t}), \|\mathbf{w} - \mathbf{w}^*\|^2\}, \qquad \mathbf{t} \in \operatorname{dom}\phi,$$

and clearly $\ell^*(\mathbf{t}^*) = 0$. Hence, we can use $\ell^*(\mathbf{t}^{(k)})$ as a natural measure for the quality of our approximate solutions, since it gives an upper bound for the distance of the point from the boundary and the target point, as well.

Now we can measure the distance of vector $\mathbf{t} \in \operatorname{dom}\phi$ from the boundary of $\operatorname{dom}\phi$ in the opposite direction (to the greedy one) by defining

$$\underline{\alpha}(\mathbf{t}) = \max\{\alpha \geq 0 : \mathbf{t}(1-\alpha) \in \operatorname{dom}\phi\}.$$

Note that the maximum is achieved when $\mathbf{t}(1-\alpha)$ is on the boundary of $\operatorname{dom}\phi$, so $w_0(1-\alpha) = \|\mathbf{w}(1-\alpha)\|^2$, therefore

$$\underline{\alpha}(\mathbf{t}) = \frac{\ell^*(\mathbf{t})}{\|\mathbf{w}-\mathbf{w}^*\|^2} = \frac{\rho(\mathbf{t})}{\|\mathbf{w}-\mathbf{w}^*\|^2} + 1 > 1. \tag{28}$$

Using the function $\underline{\alpha}$, we can estimate the local norm of the search direction $\Delta\mathbf{t}$.

Lemma 5.1. *Let $\mathbf{t} \in \operatorname{dom}\Phi$ and $\Delta\mathbf{t}$ be the target direction. Then, we have*

$$\sqrt{\frac{n+1}{2}} \leq \frac{\underline{\alpha}(\mathbf{t})-1}{\underline{\alpha}(\mathbf{t})}\,\|\Delta\mathbf{t}\|_{\nabla^2\phi(\mathbf{t})} \leq \sqrt{n+1}.$$

Proof. Consider the function $\psi(\mathbf{t}) := -\ln\rho(\mathbf{t})$. Let us compute its derivatives at point $\mathbf{t} \in \operatorname{int}\operatorname{dom}\phi$ along the direction $\Delta\mathbf{t} = (w_0^* - w_0, \mathbf{w}^* - \mathbf{w}) = (\|\mathbf{w}^*\|^2 - w_0, \mathbf{w}^* - \mathbf{w})$, namely compute the directional derivatives of $\psi(\mathbf{t} + \alpha\Delta\mathbf{t}) = -\ln(w_0 + \alpha\Delta w_0 - \|\mathbf{w} + \alpha\Delta\mathbf{w}\|^2)$. Note that

$$\begin{aligned}
D\psi(\mathbf{t})[\Delta\mathbf{t}] &= -\frac{1}{\rho(\mathbf{t})}(\Delta w_0 - 2\mathbf{w}^T\Delta\mathbf{w}) = -\frac{1}{\rho(\mathbf{t})}\left(\|\mathbf{w}-\mathbf{w}^*\|^2 - (w_0 - \|\mathbf{w}\|^2)\right) = 1 - \frac{\|\mathbf{w}-\mathbf{w}^*\|^2}{\rho(\mathbf{t})},\\
D^2\psi(\mathbf{t})[\Delta\mathbf{t}]^2 &= \frac{1}{\rho^2(\mathbf{t})}\left(\|\mathbf{w}-\mathbf{w}^*\|^2 - \rho(\mathbf{t})\right)^2 + \frac{2}{\rho(\mathbf{t})}\|\mathbf{w}-\mathbf{w}^*\|^2 = \frac{\|\mathbf{w}-\mathbf{w}^*\|^4}{\rho^2(\mathbf{t})} + 1 = \frac{1}{(\underline{\alpha}(\mathbf{t})-1)^2} + 1,
\end{aligned}$$

since the identity (28).

Using that $D^2\psi(\mathbf{t})[\Delta\mathbf{t}]^2 = \|\Delta\mathbf{t}\|^2_{\nabla^2\psi(\mathbf{t})}$ and the previous equation, we get

$$\left(\frac{\underline{\alpha}(\mathbf{t})-1}{\underline{\alpha}(\mathbf{t})}\right)^2 \|\Delta\mathbf{t}\|^2_{\nabla^2\psi(\mathbf{t})} = \frac{1}{\underline{\alpha}^2(\mathbf{t})} + \left(1 - \frac{1}{\underline{\alpha}(\mathbf{t})}\right)^2.$$

From the previous expression, we obtained the following bounds

$$1 \geq \frac{\underline{\alpha}(\mathbf{t})-1}{\underline{\alpha}(\mathbf{t})}\,\|\Delta\mathbf{t}\|_{\nabla^2\psi(\mathbf{t})} \geq \frac{1}{\sqrt{2}}. \tag{29}$$

In Theorem 4.6 the bound is given on $\|\Delta\mathbf{t}\|_{\nabla^2\phi(\mathbf{t})} = D^2\phi(\mathbf{t})[\Delta\mathbf{t}]^2$, where the closed form of the function $\phi(\mathbf{t}) = -(n+1)\,\ln\frac{\rho(\mathbf{t})}{n+1}$ is stated at (8). Similarities between the functions ψ and ϕ are clear, thus we need to understand the connections between $\|\Delta\mathbf{t}\|_{\nabla^2\psi(\mathbf{t})}$ and $\|\Delta\mathbf{t}\|_{\nabla^2\phi(\mathbf{t})}$. It can be shown that

$$D^2\phi(\mathbf{t})[\Delta\mathbf{t}]^2 = (n+1)\,D^2\psi(\mathbf{t})[\Delta\mathbf{t}]^2, \qquad \text{so} \qquad \|\Delta\mathbf{t}\|_{\nabla^2\phi(\mathbf{t})} = \sqrt{n+1}\,\|\Delta\mathbf{t}\|_{\nabla^2\psi(\mathbf{t})},$$

which combining with (29), we complete the proof. □

The expression of the distance function $\underline{\alpha}(\mathbf{t})$ appearing in Lemma 5.1 gives the idea to define the following *merit function* as follows

$$\mu^*(\mathbf{t}) := \frac{\underline{\alpha}(\mathbf{t})}{\underline{\alpha}(\mathbf{t})-1}\,\ell^*(\mathbf{t}) = \frac{\ell^*(\mathbf{t})^2}{\ell^*(\mathbf{t}) - \|\mathbf{w}-\mathbf{w}^*\|^2} \geq \ell^*(\mathbf{t}). \tag{30}$$

The following lemma plays an important role in the analysis of PTS IPAs. It shows that if the step-length is big enough, then the decrease of the merit function value $\mu^*(\mathbf{t}(\alpha))$ is large enough.

Lemma 5.2. *Let $\alpha \in (0,1)$ be a feasible step length in the predictor step of the PTS IPA. If $\alpha \geq \gamma\frac{\underline{\alpha}(\mathbf{t})-1}{\underline{\alpha}(\mathbf{t})}$, then $\mu^*(\mathbf{t}(\alpha)) \leq \frac{1}{1+\gamma}\mu^*(\mathbf{t})$ with some $\gamma \in (0,1)$.*

Proof. Note that $\ell^*(\mathbf{t}(\alpha)) = (1-\alpha)\ell^*(\mathbf{t})$ and $\|\mathbf{w}(\alpha)-\mathbf{w}^*\| = (1-\alpha)\|\mathbf{w}-\mathbf{w}^*\|$, thus $\underline{\alpha}(\mathbf{t}(\alpha)) = \frac{\underline{\alpha}(\mathbf{t})}{1-\alpha} > \underline{\alpha}(\mathbf{t})$. Using the definition of the merit function μ^*, we have

$$\mu^*(\mathbf{t}(\alpha)) = \frac{\underline{\alpha}(\mathbf{t}(\alpha))}{\underline{\alpha}(\mathbf{t}(\alpha))-1}\,\ell^*(\mathbf{t}(\alpha)). \tag{31}$$

We estimate the two factors from (31) separately. Using the lower bound on the step length α, we get

$$\frac{\underline{\alpha}(\mathbf{t}(\alpha))-1}{\underline{\alpha}(\mathbf{t}(\alpha))} = 1-\frac{1-\alpha}{\underline{\alpha}(\mathbf{t})} \geq 1-\frac{1}{\underline{\alpha}(\mathbf{t})}+\gamma\frac{\underline{\alpha}(\mathbf{t})-1}{\underline{\alpha}^2(\mathbf{t})} = \left(1-\frac{1}{\underline{\alpha}(\mathbf{t})}\right)\left(1+\frac{\gamma}{\underline{\alpha}(\mathbf{t})}\right), \tag{32}$$

and

$$\ell^*(\mathbf{t}(\alpha)) = (1-\alpha)\ell^*(\mathbf{t}) \leq \left(1-\gamma\frac{\underline{\alpha}(\mathbf{t})-1}{\underline{\alpha}(\mathbf{t})}\right)\ell^*(\mathbf{t}). \tag{33}$$

Substituting the bound obtained in (32) and (33) into (31), we get

$$\begin{aligned}\mu^*(\mathbf{t}(\alpha)) &\leq \frac{1-\gamma\frac{\underline{\alpha}(\mathbf{t})-1}{\underline{\alpha}(\mathbf{t})}}{\left(1-\frac{1}{\underline{\alpha}(\mathbf{t})}\right)\left(1+\frac{\gamma}{\underline{\alpha}(\mathbf{t})}\right)}\,\ell^*(\mathbf{t}) = \frac{\underline{\alpha}(\mathbf{t})-\gamma(\underline{\alpha}(\mathbf{t})-1)}{\underline{\alpha}(\mathbf{t})+\gamma}\,\mu^*(\mathbf{t})\\ &= \left(1-\frac{\underline{\alpha}(\mathbf{t})\gamma}{\underline{\alpha}(\mathbf{t})+\gamma}\right)\mu^*(\mathbf{t}) \leq \frac{1}{1+\gamma}\mu^*(\mathbf{t}),\end{aligned} \tag{34}$$

since $1-x\gamma/(x+\gamma)$ is a monotone decresing function for $x \geq 1$. ☐

Let us show now that the condition of Lemma 5.2 can be derived from the inequality (26). Indeed, we have

$$\alpha_k \geq \frac{1}{\|\Delta\mathbf{t}_k\|_{\nabla^2\phi(\mathbf{t}_k)}}\frac{2\,\kappa}{3\,(3+\kappa)} \geq \frac{\underline{\alpha}(\mathbf{t})-1}{\underline{\alpha}(\mathbf{t})}\gamma, \tag{35}$$

with $\gamma = \frac{2\,\kappa}{3\,(3+\kappa)\,\sqrt{n+1}}$.

In the following theorem, we give a bound on the number of iterations.

Theorem 5.3. *Let $\beta=\frac{1}{4}$, $\delta_u=\eta\,\delta_l$, where $\delta_l=\frac{2+\kappa^2}{8}$ and $\eta\geq 1$, $\kappa>0$, $\gamma=\frac{2\kappa}{3(3+\kappa)\sqrt{n+1}}$ and let $\mathbf{z}^{(0)}=(\mathbf{u}^{(0)},\mathbf{t}^{(0)})\in\mathscr{F}_z$ be the starting point. Then PTS IPA performs at most*

$$\frac{1}{\gamma}\log_2\frac{\mu^*(\mathbf{t}^{(0)})}{\varepsilon} = O\left(\sqrt{n}\ln\frac{\mu^*(\mathbf{t}^{(0)})}{\varepsilon}\right)$$

iterations using greedy step in $\mathscr{T}$.

Proof. Using (35) and Lemma 5.2, we get that after each iteration $\mu^*(\mathbf{t}(\alpha)) \leq \frac{1}{1+\gamma}\mu^*(\mathbf{t})$. Hence, $\mu^*(\mathbf{t}) \leq \varepsilon$ if

$$\left(\frac{1}{1+\gamma}\right)^k \mu^*(\mathbf{t}^{(0)}) \leq \varepsilon.$$

Rearranging the inequality, we obtain that $\mu^*(\mathbf{t}) \leq \varepsilon$ holds if

$$k \geq \frac{1}{\ln(1+\gamma)}\ln\left(\frac{\mu^*(\mathbf{t}^{(0)})}{\varepsilon}\right).$$

Using $\ln(1+\gamma) \geq \gamma\ln 2$ for $\gamma\in(0,1)$, we obtain the result. ☐

6 Numerical results

We tested the presented PTS IPA on two sets of instances: in the first case, we used matrices given on the website [31], while in the second case, we randomly generated the coefficient matrix of (WLCP). We considered $\beta = \frac{1}{4}$, $\delta_l = 0.9$, $\delta_u = 1$ and $\varepsilon = 10^{-8}$.

The website [31] gives a detailed description of how the collected matrices were generated. All the matrices are either bisymmetric or symmetric PSD. For these instances, we generated the vector $\mathbf{q}$ as $-M\mathbf{e}+\mathbf{e}$ and used the vector $\mathbf{e}$ as the initial solution.

A total of 66 problems (21 symmetric PSD and 45 bisymmetric) were considered, 47 of which were solved by the algorithm with an accuracy of at least 10^{-6} (see Tables 1 and 2), while in case of 7 problems we got a solution with accuracy worse than 10^{-3} (see Table 3). The required accuracy in all these runs was 10^{-8}. The earlier terminations are related with difficulties in Linear Algebra, which can be eliminated by a more accurate programming. As we can see, the total number of solutions of Newton systems is always quite moderate.

Problem	Dim	Predictor	Corrector	Total
TR_PSD_1	27	18	35	56
TR_PSD_2	32	17	32	52
TR_PSD_5	48	20	39	62
TR_PSD_6	47	20	40	61
TR_PSD_7	46	20	40	63
TR_PSD_8	50	34	59	96
TR_PSD_9	50	29	58	90
TR_PSD_10	49	29	58	90
TR_PSD_11	48	34	59	96
TR_PSD_12	74	18	36	57
TR_PSD_13	77	28	48	79
TR_PSD_14	83	21	42	64
TR_PSD_15	91	47	63	110
TR_PSD_16	96	16	33	51
TR_PSD_17	97	19	38	58
TR_PSD_20	50	29	58	90

Table 1 Numerical results for symmetric PSD matrices with reached accuracy at least 10^{-6}.

Problem	Dim	Predictor	Corrector	Total
TR_BS_1,2	59	19	38	60
TR_BS_4	68	19	39	61
TR_BS_5,6	73	17	34	54
TR_BS_7,8	74	24	48	75
TR_BS_9	75	20	84	106
TR_BS_10	75	20	40	60
TR_BS_11,12	80	25	49	77
TR_BS_13,14	81	26	51	79
TR_BS_15	84	16	34	50
TR_BS_16	84	16	33	53
TR_BS_17,18	88	24	46	73
TR_BS_19	89	24	70	94
TR_BS_20	89	19	39	59
TR_BS_21	89	19	39	58
TR_BS_22	95	23	45	71
TR_BS_23	96	24	47	74
TR_BS_24,25	95	23	46	72
TR_BS_28,29	102	29	57	89
TR_BS_30,31	103	24	47	75
TR_BS_42	90	20	42	64
TR_BS_43	90	19	38	70
TR_BS_45	90	24	49	73

Table 2 Numerical results for bisymmetric matrices with reached accuracy at least 10^{-6}.

Problem	Dim	Predictor	Corrector	Total
TR_PSD_3	41	11	23	34
TR_PSD_4	43	7	14	21
TR_PSD_22	79	6	12	18
TR_BS_3	68	14	28	42
TR_BS_26	97	15	30	45
TR_BS_27	97	17	35	52
TR_BS_34	140	23	46	69

Table 3 Numerical results for difficult problems (reached accuracy is worse than 10^{-3}).

Now let us present the computational results for randomly generated problems. Recall that we considered the following problem: Find $\mathbf{u},\mathbf{v} \in \mathbb{R}^n_+ : \quad \mathbf{uv} = \mathbf{p} \quad$ with $\mathbf{v} = M\mathbf{u} + \mathbf{q}$, where $\mathbf{p} \in \mathbb{R}^n_\oplus$ and $M + M^T \succeq 0$. A necessary and sufficient condition for the solvability of this problem is the existence of the strictly feasible primal-dual pair $(\hat{\mathbf{u}}, \hat{\mathbf{v}})$:

$$\hat{\mathbf{u}}, \hat{\mathbf{v}} \in \mathbb{R}^n_+ : \quad \hat{\mathbf{v}} = M\hat{\mathbf{u}} + \mathbf{q}.$$

Hence, this condition serves as a starting element for our random generator. It performs the following four steps.

1. Choose randomly $\hat{\mathbf{u}}, \hat{\mathbf{v}} \in \mathbb{R}^n_+$, which components are uniformly distributed in $(0,1)$.

2. Form a random matrix $A \in \mathbb{R}^{n\times n}$ and a lower triangular random matrix $L \in \mathbb{R}^{n\times n}$ with elements uniformly distributed in $(-1,1)$.

3. Set $M = AA^T + \xi(L - L^T)$, where $\xi \geq 0$ is a parameter. Define $\mathbf{q} = \hat{\mathbf{v}} - M\hat{\mathbf{u}}$.

4. With probability $\pi \in [0,1]$, decide if the element p_i is positive. Positive p_i are chosen uniformly distributed in $(0,1]$.

Thus, our generator has three parameters: n, $\xi \geq 0$, and $\pi \in [0,1]$. For our experiments, they are chosen as follows:

$$\pi = \tfrac{1}{2}, \quad \xi = 10.$$

In this way, we generated 100-100 problems of each size indicated in Table 4. In the table below, the columns present the average number of predictor and corrector steps for different dimensions in the series of one hundred random test problems.

Dimension	16	32	64	128	256	512
Predictions	11.3	12.5	14.7	18.5	22.9	29.0
Corrections	22.8	24.9	30.7	50.1	58.0	74.9

Table 4 Average iteration numbers for random problems

As we see, for one predictor step, we need approximately 2-2.5 corrector steps. The number of predictor steps grows a little bit slower than $\sqrt{n}$ (look at the pairs 64/256 and 128/512). With the growth of the dimension, the Newton system inside the algorithm becomes more and more degenerate. Hence, for dimension $n = 512$ our algorithm reached only the accuracy $\varepsilon = 10^{-7}$.

The most interesting observation from our preliminary experiments is the significant acceleration of convergence in the end of the process. We can see it from Table 2, which relates the progress in the proximity measure $\mu^*(\cdot)$, achieved at a random problem of dimension $n = 512$, with the number of predictor and corrector steps. In the last line there, we can see the relative size of the predictor step as compared with the maximal step preserving feasibility.

$\mu^*(\mathbf{t})$	$5\cdot 10^2$	$4\cdot 10^1$	$3\cdot 10^0$	$3\cdot 10^{-1}$	$2\cdot 10^{-2}$	$2\cdot 10^{-3}$	$3\cdot 10^{-4}$	$2\cdot 10^{-5}$	$2\cdot 10^{-6}$	$1\cdot 10^{-7}$
N_{pred}	0	15	21	24	26	27	28	29	30	31
N_{cor}	0	32	45	51	55	57	59	61	63	65
$\alpha_{\max}$	0.84	0.87	0.92	0.95	0.96	0.97	0.97	0.97	0.96	0.96

Table 5 Acceleration of convergence for a random problem

In the end of the process, the algorithm often demonstrates a good linear rate. However, the possibility to achieve a local super-linear rate of convergence for methods of this type remains an interesting open question.

7 Conclusion

In this paper, we intorduced an IPA for monotone WLCPs by extending the PTS framework proposed by Nesterov (2008) in [25] for primal-dual linear programming problems. The method performs two types of steps. In the predictor stage, the goal is to use a search direction in the PTS which ensures a large enough decrease in the stopping criteria. The goal of the corrector stage is to ensure that the computed solution belongs to the neighbourhood $\mathcal{N}_\lambda(\beta)$ of the current point $\mathbf{u}(\mathbf{t})$. The algorithm works in a wide neighborhood, allowing very large steps. This results in a significant acceleration in the end of the process, which was also illustrated in our preliminary numerical experiments. Furthermore, we showed that the proposed PTS IPA has the best known worst-case complexity bound. As further research, the algorithm could be extended to more general problems, such as $P_*(\kappa)$-WLCPs.

Funding

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] Anstreicher, K.M.: Interior-point algorithms for a generalization of linear programming and weighted centering. Optim. Methods Softw. 27(4-5), 605–612 (2012)

[2] Bazaraa, M.S., Sherali, H.D., Shetty., C.M.: Nonlinear programming: Theory and algorithms. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition (2006)

[3] Cottle, R.W., Pang, J.-S., Venkateswaran, V.: Sufficient matrices and the linear complementarity problem. Linear Algebra Appl. 114-115, 231-249 (1989)

[4] Cottle, R.W., Pang, J.-S., Stone., R.E.: The Linear Complementarity Problem. Computer Science and Scientific Computing. Academic Press, Inc., Boston (1992)

[5] Csizmadia, Zs., Illés, T.: New criss-cross type algorithms for linear complementarity problems with sufficient matrices. Optim. Methods Softw. 21(2), 247–266 (2006)

[6] Csizmadia, Zs., Illés, T., Nagy, A.: The s-Monotone Index Selection Rule for Criss-Cross Algorithms of Linear Complementarity Problems. Acta Univ. Sapientiae Inform. 5(1), 103-139 (2013)

[7] Darvay, Zs., Illés, T., Povh, J., Rigó, P.R.: Feasible Corrector-Predictor Interior-Point Algorithm for $P_*(\kappa)$-Linear Complementarity Problems Based on a New Search Direction. SIAM J. Optim. 30(3), 2628-2658 (2020)

[8] Darvay, Zs., Illés, T. Rigó, P.R.: Predictor-corrector interior-point algorithm for $P_*(\kappa)$-linear complementarity problems based on a new type of algebraic equivalent transformation technique. Eur. J. Oper. Res. 299(1), 25–35 (2022)

[9] de Klerk, E., Roos, C., Terlaky, T.: Nonlinear optimization [Hungarian translation]. In the series on Operations Research, Aula Kiadó, Budapest (2004)

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

[11] Glineur, F.: Topics in Convex Optimization: Interior-Point Methods, Conic Duality and Approximations. PhD Thesis, Faculté Polytechnique de Mons, Mons, Belgium (2001)

[12] Gonzaga, C.C.: An Algorithm for Solving Linear Programming Problems in $O(n^3L)$ Operations. In: Megiddo, N. (eds.) Progress in Mathematical Programming. Springer, New York, NY (1989)

[13] Illés, T., Roos, C., Terlaky, T.: Simple approach for the interior point theory of linear complementarity problems with P_* matrices. Unpublished manuscript (1997)

[14] Illés, T., Peng, J., Roos, C., Terlaky, T.: A Strongly Polynomial Rounding Procedure Yielding a Maximally Complementary Solution for $P_*(\kappa)$ Linear Complementarity Problems. SIAM J. Optim. 11(2), 320-340 (2000)

[15] Illés, T., Nagy, M., Terlaky, T.: A polynomial path-following interior point algorithm for general linear complementarity problems. J. Global Optim. 47(3), 329-342 (2010)

[16] Illés, T., Rigó, P.R., Török, R.: Unified Approach of Interior-Point Algorithms for $P_*(\kappa)$-LCPs Using a New Class of Algebraically Equivalent Transformations. J. Optim. Theory Appl. 202(1), 27-49 (2024)

[17] Klafszky, E., Terlaky, T.: Some generalizations of the criss-cross method for quadratic programming. Optimization. 24(1-2), 127-139 (1992)

[18] Kojima, M., Megiddo, N., Noma, T., Yoshise, A.: A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems. Lecture Notes in Comput. Sci. 538, Springer Verlag, Berlin, Germany (1991)

[19] Martos, B.: Nonlinear Programming: Theory and Methods. Akadémiai Kiadó, Budapest (1975)

[20] Megiddo, N.: Pathways to the optimal set in linear programming. In: N. Megiddo (ed.) Progress in Mathematical Programming: Interior Point and Related Methods, pp. 131-158. Springer Verlag, New York (1989)

[21] Nagy, M.: Interior point algorithms for general linear complementarity problems. PhD Thesis, Eötvös Loránd University of Sciences, Institute of Mathematics, Budapest, Hungary (2009)

[22] Nemirovski, A.: Interior point polynomial time methods in convex programming. Lecture Notes, Spring Semester, Georgia Institute of Technology, Atlanta, GA (1996)

[23] Nesterov, Yu., Nemirovski, A.: Interior point polynomial methods in convex programming: Theory and Applications, SIAM, Philadelphia (1994)

[24] Nesterov, Yu.: Introductury Lectures on Convex Optimization. Kluwer Academic Publishers, Dordrecht (2004)

[25] Nesterov, Yu: Parabolic target space and primal–dual interior-point methods. Discrete Appl. Math. 156, 2079-2100 (2008)

[26] Nesterov, Yu.: Lectures on Convex Optimization. Springer Optimization and Its Applications, Springer, Switzerland (2018)

[27] Potra, F.A.: Weighted Complementarity Problems-A New Paradigm for Computing Equilibria. SIAM J. Optim. 22(4), 1634-1654 (2012)

[28] Potra, F.A.: Sufficient weighted complementarity problems. Comput. Optim. Appl. 64, 467–488 (2016)

[29] Renegar, J.: A polynomial-time algorithm, based on Newton's method, for linear programming. Math. Program. 40, 59–93 (1988)

[30] Renegar, J.: A mathematical view of interior-point methods in convex optimization. MPS/SIAM Ser. Optim, vol. 3, SIAM, New York (2001)

[31] Török, R.: Test examples about linear complementarity problems. https://sites.google.com/view/lcp-test-examples-roland-trk/mainpage (2023)

[32] Sonnevend, Gy.: An "analytic center" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In: A. Prékopa, J. Szelezsán, B. Strazicky (eds.) System Modelling and Optimization: Proceedings of the 12th IFIP-Conference, Budapest, Hungary, Lecture Notes in Control and Information Sciences 84, pp. 866-876. Springer Verlag, Berlin (1986)

[33] Väliaho, H.: P_*-matrices are just sufficient. Linear Algebra Appl. 239, 103-108 (1996)

[34] Ye, Y.: A path to the Arrow-Debreu competitive market equilibrium. Math. Program. 111, 315–348 (2008)

8 Appendix: Self-concordant functions and some useful results

We need to introduce the definitions of the *barrier function*, *self-concordant function* and *self-concordant barrier function* and some important related results that will be used in this paper. This part of the paper is based on the unification of the approaches of self-concordant functions presented in [9, 11, 22, 26].

Definition 8.1. Let $Q \subseteq \mathbb{R}^n$ be a closed convex set with a nonempty interior and $f : \operatorname{int} Q \to \mathbb{R}$ be a continuous function. The function f is a *barrier function* for the convex set Q if and only if it satisfies the following assumptions:

1. f is a smooth function, for our purposes f is three times continuously differentiable,
2. f is convex function, i.e. $\nabla^2 f(\mathbf{x})$ is positive semidefinite matrix, for all $\mathbf{x} \in \operatorname{int} Q$, and
3. $\lim_{\mathbf{x} \to \partial Q} f(\mathbf{x}) = +\infty$, where ∂Q is the boundary of Q.

By fixing a point $\mathbf{x} \in \operatorname{int} Q$ and direction $\mathbf{h} \in \mathbb{R}^n$, we define a function $\varphi : \mathbb{R} \to \mathbb{R}$ in the following way

$$\varphi(t) = f(\mathbf{x} + t\,\mathbf{h}),$$

where the domain of φ is given as $\operatorname{dom} \varphi = \{t \in \mathbb{R} : \mathbf{x} + t\,\mathbf{h} \in \operatorname{int} Q\}$. It is obvious that φ is strictly convex, and three times differentiable function. Furthermore, we can define the first, second and third order differentials of function f taken at point $\mathbf{x}$ in the direction of $\mathbf{h}$, computed at $t = 0$ as follows

$$\begin{aligned} Df(\mathbf{x})[\mathbf{h}] &= \varphi'(0) = \sum_{i=1}^{n} h_i \frac{\partial f(\mathbf{x})}{\partial x_i} = \mathbf{h}^T \nabla f(\mathbf{x}), \\ D^2 f(\mathbf{x})[\mathbf{h},\mathbf{h}] &= \varphi''(0) = \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \mathbf{h}^T \nabla^2 f(\mathbf{x}) \mathbf{h} = \|\mathbf{h}\|^2_{\nabla^2 f(\mathbf{x})}, \\ D^3 f(\mathbf{x})[\mathbf{h},\mathbf{h},\mathbf{h}] &= \varphi'''(0) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} h_i h_j h_k \frac{\partial^3 f(\mathbf{x})}{\partial x_i \partial x_j \partial x_k} = \mathbf{h}^T \nabla^3 f(\mathbf{x})[\mathbf{h}]\mathbf{h}, \end{aligned}$$

where $\nabla^3 f(\mathbf{x})[\mathbf{h}] \in \mathbb{R}^{n \times n}$ and $\|\mathbf{h}\|_{\nabla^2 f(\mathbf{x})}$ is a *local norm* induced by the positive definite matrix $\nabla^2 f(\mathbf{x})$. If it does not cause misunderstanding, then the local norm is denoted as $\|\mathbf{h}\|_{\mathbf{x}}$, referring only to the point where the Hessian matrix of the function f has been computed.

Recall that the third degree Taylor polynomial of the function φ around 0 can be given as follows

$$\varphi(0) + \varphi'(0)\,t + \frac{1}{2}\,\varphi''(0)\,t^2 + \frac{1}{6}\,\varphi'''(0)\,t^3.$$

Thus, it will be clear that the definition of *self-concordance* of function f specifies that the cubic term of the Taylor polynomial of f can be bounded with the quadratic term. This property plays an important role in the complexity analysis of the Newton-method.

Definition 8.2. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *self-concordant* if it is a closed convex function with open domain and there exists a constant $M_f \geq 0$ such that the inequality

$$|D^3 f(\mathbf{x})[\mathbf{h},\mathbf{h},\mathbf{h}]| \leq 2M_f \, \|\mathbf{h}\|^{\frac{3}{2}}{}_{\nabla^2 f(\mathbf{x})} \tag{36}$$

holds for all $\mathbf{x} \in \operatorname{dom} f$ and $\mathbf{h} \in \mathbb{R}^n$. If $M_f = 1$, the function is called *standard self-concordant*.

To complete developing the technical tools, we should investigate a specific subfamily of self-concordant functions, the family of *self-concordant barriers*.

Definition 8.3. Let f be a standard self-concordant function. We call it a *ν-self-concordant barrier* for the set of $\operatorname{dom} f$, if

$$(Df(\mathbf{x})[\mathbf{h}])^2 = (\mathbf{h}^T \nabla f(\mathbf{x}))^2 \le \nu \|\mathbf{h}\|^2_{\nabla^2 f(\mathbf{x})} = \nu D^2 f(\mathbf{x})[\mathbf{h},\mathbf{h}] \tag{37}$$

for all $\mathbf{x} \in \operatorname{dom} f$. The value ν is called the parameter of the barrier.

Let us now state that self-concordance is an affine-invariant property.

Theorem 8.1 ([26], Theorem 5.1.2)**.** *Let a function f be self-concordant with constant M_f, and $\mathscr{A} : \mathbb{R}^n \to \mathbb{R}^m$ be a linear operator of the form $\mathscr{A}(\mathbf{x}) := A\mathbf{x} + \mathbf{b}$, where $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then the function*

$$\phi(\mathbf{x}) = f(\mathscr{A}(\mathbf{x}))$$

is also self-concordant and $M_\phi = M_f$.

Let us consider the following ellipsoids:

$$\begin{aligned} W^0(\mathbf{x}, r) &= \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})} < r\}, \\ W(\mathbf{x}, r) &= cl\left(W^0(\mathbf{x}, r)\right) = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})} \le r\}. \end{aligned}$$

This set is called the *Dikin ellipsoid* of the function f at $\mathbf{x}$.

Theorem 8.2 ([26], Theorem 5.1.5)**.** *1. For any $\mathbf{x} \in \operatorname{dom} f$, we have $W^0\left(\mathbf{x}, \frac{1}{M_f}\right) \subseteq \operatorname{dom} f$.*

2. For all $\mathbf{x}, \mathbf{y} \in \operatorname{dom} f$, the following inequality holds:

$$\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{y})} \ge \frac{\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}}{1 + M_f \|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}} \tag{38}$$

3. If $\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})} < \frac{1}{M_f}$, then

$$\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{y})} \le \frac{\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}}{1 - M_f \|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}}. \tag{39}$$

Theorem 8.3 ([26], Theorem 5.1.6)**.** *Let a function f be self-concordant and $\operatorname{dom} f$ contains no straight lines. Then the Hessian $\nabla^2 f(\mathbf{x})$ is nondegenerate at all points $\mathbf{x} \in \operatorname{dom} f$.*

Under the assumption that $\operatorname{dom} f$ contains no straight lines, we can define *local dual norm* for any $\mathbf{x} \in \operatorname{dom} f \subset \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^n$ as follows

$$\|\mathbf{g}\|^*_{\mathbf{x}} = \|\mathbf{g}\|_{[\nabla^2 f(\mathbf{x})]^{-1}} = \sqrt{\mathbf{g}^T [\nabla^2 f(\mathbf{x})]^{-1} \mathbf{g}}.$$

It can be shown that for any $\mathbf{g}, \mathbf{h} \in \mathbb{R}^n$: $|\mathbf{h}^T \mathbf{g}| \le \|\mathbf{h}\|_{\mathbf{x}} \|\mathbf{g}\|^*_{\mathbf{x}}$ holds.

Using the local dual norm we can introduce the *Newton decrement* of the function f at $\mathbf{x} \in \operatorname{dom} f$ in the following way

$$\lambda_f(\mathbf{x}) = \|\nabla f(\mathbf{x})\|^*_{\mathbf{x}} = \sqrt{\nabla f(\mathbf{x})^T [\nabla^2 f(\mathbf{x})]^{-1} \nabla f(\mathbf{x})}.$$

Let us define the Fenchel conjugate pair of the function

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

as

$$\omega_*(\tau) = -\tau - \ln(1-\tau), \quad \tau \in [0,1).$$

Note that ω and ω_* are monotone increasing convex functions.

Lemmas 5.1.4 and 5.1.5 from [26] are useful results, thus we summarize those.

Lemma 8.4. *For any $t \ge 0$ and $\tau \in [0,1)$, we have*

$$\omega'(\omega'_*(\tau)) = \tau, \qquad \omega'_*(\omega'(t)) = t,$$

$$\omega(t) = \max_{0\le\xi<1} [\xi t - \omega_*(\xi)], \qquad \omega_*(\tau) = \max_{\xi\ge 0} [\xi \tau - \omega(\xi)],$$

$$\omega(t)+\omega_*(\tau)\geq t\,\tau, \tag{40}$$

$$\omega_*(\tau)=\tau\,\omega_*'(\tau)-\omega(\omega_*'(\tau)), \qquad \omega(t)=t\,\omega'(t)-\omega_*(\omega'(t)),$$

$$\frac{t^2}{2\,(1+t)}\leq\frac{t^2}{2\,(1+\frac{2}{3}\,t)}\leq\omega(t)\leq\frac{t^2}{2+t},$$

$$\frac{\tau^2}{2-\tau}\leq\omega_*(\tau)\leq\frac{\tau^2}{2\,(1-\tau)}. \tag{41}$$

Theorem 8.5 ([26], Theorem 5.1.7)**.** *Let $\mathbf{x}\in\operatorname{dom}f$. Then for any $\mathbf{y}\in W^0\left(\mathbf{x},\frac{1}{M_f}\right)$ we have*

$$(1-M_f\,r)^2\,\nabla^2 f(\mathbf{x})\preceq\nabla^2 f(\mathbf{y})\preceq\frac{1}{(1-M_f\,r)^2}\,\nabla^2 f(\mathbf{x}), \tag{42}$$

where $r=\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}$.

We need additional inequalities that characterize self-concordant functions.

Theorem 8.6 ([26], Theorem 5.1.8)**.** *Let a function f be self-concordant. For any $\mathbf{x},\mathbf{y}\in\operatorname{dom}f$, we have*

$$(\mathbf{y}-\mathbf{x})^T\left(\nabla f(\mathbf{y})-\nabla f(\mathbf{x})\right)\geq\frac{\|\mathbf{y}-\mathbf{x}\|^2_{\nabla^2 f(\mathbf{x})}}{1+M_f\,\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}}, \tag{43}$$

$$f(\mathbf{y})\geq f(\mathbf{x})+(\mathbf{y}-\mathbf{x})^T\nabla f(\mathbf{x})+\frac{1}{M_f^2}\,\omega(M_f\,\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}), \tag{44}$$

where $\omega(t)=t-\ln(1+t)$.

An easy consequence of the previous theorem, especially the inequality (43) is the following statement.

Corollary 8.7. *Let a function f be self-concordant. For any $\mathbf{x},\mathbf{y}\in\operatorname{dom}f$, we have*

$$\frac{\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}}{1+M_f\,\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}}\leq\|\nabla f(\mathbf{y})-\nabla f(\mathbf{x})\|_{[\nabla^2 f(\mathbf{x})]^{-1}}. \tag{45}$$

Define the *Fenchel conjugate function* (or *Fenchel dual function*) f_* of the self-concordant function f for $\mathbf{s}\in\mathbb{R}^n$, the value of this function is defined as follows:

$$f_*(\mathbf{s})=\sup_{\mathbf{x}\in\operatorname{dom}f}\left(\mathbf{s}^T\mathbf{x}-f(\mathbf{x})\right). \tag{46}$$

Clearly, $\operatorname{dom}f_*=\{\mathbf{s}\in\mathbb{R}^n : f(\mathbf{x})-\mathbf{s}^T\mathbf{x}\quad\text{is bounded below on }\operatorname{dom}f\}$.

Now we are ready to list some results related to the Fenchel conjugate function f_* of the self-concordant function f.

Lemma 8.8 ([26], Lemma 5.1.6)**.** *Let a self-concordant function f be given and denote by f_* its Fenchel conjugate function. Then f_* is a closed convex function with nonempty open domain. Moreover, $\operatorname{dom}f_*=\{\nabla f(\mathbf{x}) : \mathbf{x}\in\operatorname{dom}f\}$.*

Using the previous lemma we can state the following corollary.

Corollary 8.9. *Let a self-concordant function f be given and denote by f_* its Fenchel conjugate function. If $\mathbf{s}=\nabla f(\mathbf{x})\in\operatorname{dom}f_*$, then*

$$\nabla^2 f_*(\mathbf{s})=[\nabla^2 f(\mathbf{x})]^{-1}, \tag{47}$$

where $\mathbf{x}\in\operatorname{dom}f$.

The next theorem states that f_* is a self-concordant function.

Theorem 8.10 ([26], Theorem 5.1.17 and)**.** *Let a self-concordant function f be given and denote by f_* its Fenchel conjugate function. Then f_* is self-concordant with $M_{f_*}=M_f$.*

Using Theorems 8.6 and 8.10, Lemma 8.8 and Corollary 8.9 we can obtain the following statement.

Corollary 8.11. *Let a self-concordant function f be given and denote by f_* its Fenchel conjugate function. For any $\mathbf{x},\mathbf{y}\in\operatorname{dom}f$ we have*

$$\frac{\|\nabla f(\mathbf{x})-\nabla f(\mathbf{y})\|_{[\nabla^2 f(\mathbf{x})]^{-1}}}{1+M_f\,\|\nabla f(\mathbf{x})-\nabla f(\mathbf{y})\|_{[\nabla^2 f(\mathbf{x})]^{-1}}}\leq\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}. \tag{48}$$

Similarly to lower bounds (43) and (44), under mild, natural assumptions upper bounds can be provided for the same expressions. Interestingly enough the function ω has been replaced by its Fenchel conjugate ω_*.

Theorem 8.12 ([26], Theorem 5.1.9)**.** *Let a function f be self-concordant. For any $\mathbf{x}, \in \operatorname{dom} f$ with $\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})} < \frac{1}{M_f}$, we have*

$$(\mathbf{y}-\mathbf{x})^T\left(\nabla f(\mathbf{y})-\nabla f(\mathbf{x})\right) \leq \frac{\|\mathbf{y}-\mathbf{x}\|^2_{\nabla^2 f(\mathbf{x})}}{1-M_f\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}}, \tag{49}$$

$$f(\mathbf{y}) \leq f(\mathbf{x})+(\mathbf{y}-\mathbf{x})^T\nabla f(\mathbf{x})+\frac{1}{M_f^2}\,\omega_*(M_f\|\mathbf{y}-\mathbf{x}\|_{\nabla^2 f(\mathbf{x})}), \tag{50}$$

where $\omega_*(t) = -t-\ln(1-t), \quad t\in[0,1)$.

Furthermore, the following important inequalities hold, as well.

Theorem 8.13 ([26], Theorem 5.1.12)**.** *Let a function f be self-concordant. For any $\mathbf{x},\mathbf{y} \in \operatorname{dom} f$, we have*

$$f(\mathbf{y}) \geq f(\mathbf{x})+(\mathbf{y}-\mathbf{x})^T\nabla f(\mathbf{x})+\frac{1}{M_f^2}\,\omega(M_f\|\nabla f(\mathbf{y})-\nabla f(\mathbf{x})\|^*_{\mathbf{y}}). \tag{51}$$

If in addition $\|\nabla f(\mathbf{y})-\nabla f(\mathbf{x})\|^*_{\mathbf{y}} < \frac{1}{M_f}$, then

$$f(\mathbf{y}) \leq f(\mathbf{x})+(\mathbf{y}-\mathbf{x})^T\nabla f(\mathbf{x})+\frac{1}{M_f^2}\,\omega_*(M_f\|\nabla f(\mathbf{y})-\nabla f(\mathbf{x})\|^*_{\mathbf{y}}). \tag{52}$$

The following theorem estimates the local convergence of the standard Newton Method. (For details see pages 190-191, [24].)

Theorem 8.14 ([24], Theorem 4.1.14)**.** *Let a function f be self-concordant. Assume that $\operatorname{dom} f$ contains no straight line, and $\mathbf{x}\in\operatorname{dom} f$ with $\lambda_f(\mathbf{x}) < 1$. Then the point*

$$\mathbf{x}^+ = \mathbf{x}-[\nabla^2 f(\mathbf{x})]^{-1}\nabla f(\mathbf{x})$$

belongs to $\operatorname{dom} f$ and we have

$$\lambda_f(\mathbf{x}^+) \leq \left(\frac{\lambda_f(\mathbf{x})}{1-\lambda_f(\mathbf{x})}\right)^2.$$

Let us consider now the scheme of the Damped Newton Method, namely the new iterate $\mathbf{x}^+$ is computed as follows

$$\mathbf{x}^+ = \mathbf{x}-\frac{1}{1+M_f\lambda_f(\mathbf{x})}[\nabla^2 f(\mathbf{x})]^{-1}\nabla f(\mathbf{x}).$$

Theorem 8.15 ([26], Theorem 5.1.15)**.** *Let a function f be self-concordant and apply the Damped Newton's method for minimizing the function f. Then we have*

$$f(\mathbf{x}^+) \leq f(\mathbf{x})-\frac{1}{M_f^2}\,\omega(M_f\lambda_f(\mathbf{x})).$$