No. **648**

Polynomiality of primal-dual affine scaling algorithms for nonlinear complementarity problems

by

Benjamin Jansen, Kees Roos, Tamás Terlaky
and
Akiko Yoshise

October 1995

Polynomiality of primal-dual affine scaling algorithms for nonlinear complementarity problems *

Benjamin Jansen†, Kees Roos, Tamás Terlaky

Faculty of Technical Mathematics and Informatics
Delft University of Technology
P.O. Box 5031, 2600 GA Delft, The Netherlands

and

Akiko Yoshise‡

Institute of Socio-Economic Planning
University of Tsukuba
Tsukuba, Ibaraki 305, Japan

September 5, 1995

Abstract This paper provides an analysis of the polynomiality of primal-dual interior point algorithms for nonlinear complementarity problems using a wide neighborhood. A condition for the smoothness of the mapping is used, which is related to Zhu's scaled Lipschitz condition, but is also applicable to mappings that are not monotone. We show that a family of primal-dual affine scaling algorithms generates an approximate solution (given a precision ϵ) of the nonlinear complementarity problem in a finite number of iterations whose order is a polynomial of n, $\ln(1/\epsilon)$ and a condition number. If the mapping is linear then the results in this paper coincide with the ones in [13].

*This paper is entirely the same as the following report: B. Jansen, C. Roos, T. Terlaky and A. Yoshise, "Polynomiality of primal-dual affine scaling algorithms for nonlinear complementarity problems", Technical Report 95-83, Faculty of Technical Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands, 1995.

†Research supported by Dutch Organization for Scientific Research (NWO), grant 611-304-028

‡Research supported in part by Grant-in-Aids for Encouragement of Young Scientists (06750066) from the Ministry of Education, Science and Culture, Japan.

1 Introduction

Nonlinear complementarity problems (NCP) form a fairly general class of mathematical programming problems with a large number of applications. For instance, any convex programming problem can be modelled as a monotone nonlinear complementarity problem (MNCP). Also, the problem has a close connection with variational inequalities, which play an important role in the study of equilibria in e.g. economics, transportation planning and game–theory. For a good introduction in CP and traditional solution methods we refer the reader to the book of Cottle et al. [2]. A survey on variational inequalities is provided by Harker and Pang [9].

The study of interior point methods for linear programming that has flourished since 1984, also led to the use of barrier methods for nonlinear convex problems, see Nesterov and Nemirovsky [40], Jarre [15], Den Hertog et al. [12, 11, 10], Kortanek and Zhu [29], Monteiro and Adler [36], Zhu [49], etc. An important aspect of these algorithms is that their convergence rate has been established for classes of problems which satisfy certain Lipschitz conditions. A general and unifying analysis was provided by Nesterov and Nemirovskii [40], who managed to give a framework for the study of central path–following methods for nonlinear programming problems satisfying the so-called self-concordance condition. Recently, Nesterov and Todd [41] analyzed primal–dual potential reduction algorithms in a similar framework by considering a symmetric primal–dual cone representation of convex programming problems. Jarre [15] and Den Hertog et al. [12] used the *relative Lipschitz* condition, which was later shown (see e.g. [17]) to be essentially equivalent to self-concordance. Zhu et al. [49, 29, 44]) used the *scaled Lipschitz* condition to analyze path-following methods.

The fundamental work of McLinden [31] brought us a lot of ideas to develop interior point algorithms for linear complementarity problems (LCP) and NCPs ([6, 8, 7, 13, 24, 26, 21, 22, 23, 20, 18, 27, 33, 38, 42, 46, 48, 47] etc.). The global convergence of these algorithms has been shown by using the existence of the central path, which can be seen as a minimum requirement for interior point methods to be applicable. Similarly to the case of nonlinear convex problems, the study of the convergence rate has also become active for NCPs satisfying a smoothness condition on the mapping. In particular, Chapter 7 in [40] is completely devoted to the study of variational inequalities and their solution. Some drawbacks of the analysis in [40] are that it focuses on central path–following methods, that the type of search–direction to be used is prescribed (Newton's direction w.r.t. a self-concordant function), that the algorithm is in essence a pure primal algorithm, and that small neighborhoods of the central path are used. The last aspect has been handled by Nesterov in [39], but we are not aware of results concerning other search-directions. Potra and Ye [43] dealt with primal–dual interior point algorithms for solving MNCPs, and studied the global and local convergence rate and the complexity of the algorithms. Their smoothness condition (a scaled Lipschitz condition) can be regarded as a generalization of Zhu's condition [49].

In this paper we will focus on the complexity analysis for a family of primal–dual affine scaling algorithms for NCP. This serves as an extension to the analysis of the same family for linear CP by Jansen et al. [13]. The family contains the classical primal–dual affine scaling algorithm of Monteiro et al. [37] and the primal–dual Dikin affine scaling algorithm of Jansen et al. [14]

as special cases. In the analysis we make use of wide neighborhoods of the central path. The definition of the neighborhood is equivalent to the so-called infinity-norm neighborhood as used for linear programming in [25], [30] and [1], among others. The introduction of such a type of neighborhood has two important consequences. The first is that we may not hope for a complexity bound that is better than $\mathcal{O}(n \ln 1/\epsilon)$, where n is the number of variables and ϵ the required accuracy. The second is that we need to consider separate components of the vector of complementarity products instead of its norm. The latter aspect needs us to reconsider the use of several smoothness conditions (as self-concordance) for this type of analysis. The analysis of our algorithms nor the algorithms themselves use any barrier functions. This leads us to make assumptions on the smoothness of the mappings involved. A version of the scaled Lipschitz condition appears to be a natural choice. We introduce a new condition, Condition 3.2, which is trivial in linear cases. The advantage of this condition is that it does not require the monotonicity of the mapping while it seems to be indispensable for the scaled Lipschitz condition even if the mapping is linear. Therefore, we consider in this paper mappings for which the Jacobian matrix is a so-called P_*-matrix (see, Kojima et al. [20]). A different approach to the analysis of primal-dual large neighborhood algorithms was taken in Nesterov and Todd [41]. They reformulate a given convex programming problem in a primal-dual self-scaled conic reformulation and assume the existence of a logarithmically homogeneous barrier for the cones involved. However, although theoretically the reformulation is possible, it is far from clear what its implications are for the existence of logarithmically homogeneous barriers.

This paper is built up as follows. In Section 2 we introduce the mathematical formulation of the CP and define some notation. In Section 3 we introduce a search mapping and derive some general results. Section 4 introduces the family of algorithms we consider and provides a complete complexity analysis. In Section 5 we consider several smoothness conditions, their relationships, and their usability for our purposes.

2 Problem statement

Let us consider the NCP:

$$(CP) \quad \text{Find } (x, s) \in \mathbb{R}^{2n} \text{ such that } s = f(x),\ (x, s) \geq 0 \text{ and } x^T s = 0.$$

Here f is a C^1 mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$. We denote the sets of feasible and interior-feasible points of *(CP)* as follows:

$$\begin{aligned} \mathcal{F} &= \{(x, s) \in \mathbb{R}^{2n} : s = f(x),\ (x, s) \geq 0\}, \\ \mathcal{F}^0 &= \{(x, s) \in \mathbb{R}^{2n} : s = f(x),\ (x, s) > 0\}. \end{aligned}$$

We assume that $\mathcal{F}^0$ is not empty, or stated otherwise, that an interior point exists.

In the literature the linear complementarity problem (LCP) has gained much attention, see e.g., Cottle et al. [2]. In this special case the mapping f is given by

$$f(x) = Mx + q,$$

for $M \in \mathbb{R}^{n\times n}$ and $q \in \mathbb{R}^n$. Special conditions on the matrix M have been developed to guarantee the existence of a solution. These classes include PSD (M positive semi–definite), P (M has positive principal minors), P_* (see below) and CS and RS (column–sufficient resp. row–sufficient). Some known implications are

$$PSD \subset P_* \subset CS, \quad P \subset P_*, \quad P_* = CS \cap RS,$$

see e.g., Cottle [2], Väliaho [45] and Kojima et al. [20]. In this paper we will be interested in the class P_*.

Definition 2.1 *Let $\kappa \geq 0$. The matrix $M \in \mathbb{R}^{n\times n}$ is in $P_*(\kappa)$ if*

$$(1+4\kappa)\sum_{i\in I_+(\xi)} \xi_i(M\xi)_i + \sum_{i\in I_-(\xi)} \xi_i(M\xi)_i \geq 0, \quad \forall\, \xi \in \mathbb{R}^n,$$

where

$$I_+(\xi) = \{\, i \;:\; \xi_i(M\xi)_i > 0 \,\}, \quad I_-(\xi) = \{\, i \;:\; \xi_i(M\xi)_i < 0 \,\}.$$

The matrix M is in P_* if it is in $P_*(\kappa)$ for some κ.

Throughout this paper, we impose the following condition on the mapping f.

Condition 2.2 *There exists a constant $\kappa \geq 0$ such that the Jacobian $\nabla f(x)$ of the mapping f is a $P_*(\kappa)$ matrix for all $x \geq 0$.*

For ease of notation we allow ourselves to perform componentwise operations (like multiplication and taking powers) on vectors. For instance, by $d = \sqrt{x/s}$ we mean the vector d obtained from $d_i = \sqrt{x_i/s_i}$. Wherever this abuse of notation might be inconvenient, we use capital syllables to denote the diagonal matrix obtained from a vector; for instance $D = \text{diag}\,(d)$. We also define the mappings v, $v_{\min}$, $v_{\max}$ and ω for every $(x,s) \in \mathcal{F}^0$, which are continuous with respect to $(x,s) \in \mathcal{F}^0$:

$$\begin{aligned} v(x,s) &= (xs)^{1/2}, \\ v_{\min}(x,s) &= \min\{v_i : i = 1,2,\ldots,n\}, \\ v_{\max}(x,s) &= \max\{v_i : i = 1,2,\ldots,n\}, \\ \omega(x,s) &= v_{\min}/v_{\max} \leq 1. \end{aligned} \tag{1}$$

We often use the symbols v, $v_{\min}$, $v_{\max}$ and $\omega(v)$ to denote $v(x,s)$, $v_{\min}(x,s)$, $v_{\max}(x,s)$ and $\omega(x,s)$, respectively. For a given $\rho \in (0,1)$, we employ the following set as a neighborhood of the central path, which plays a key role in our analysis:

$$\mathcal{N}(\rho) \;=\; \left\{(x,s) \in \mathcal{F}^0 : \omega(x,s) \geq \rho\right\}.$$

3 The search mapping and its properties

Suppose that we have an interior-feasible point $(x, s) \in \mathcal{F}^0$, i.e., $s = f(x)$ and $(x, s) > 0$. Given displacements Δx and Δs, let us define

$$\begin{aligned} x(\theta) &= x + \theta\Delta x, & (2)\\ \Delta s(\theta) &= \Delta s + g(\theta)/\theta, \\ s(\theta) &= s + \theta\Delta s + g(\theta) = s + \theta\Delta s(\theta), & (3)\\ g(\theta) &= f(x + \theta\Delta x) - f(x) - \theta\nabla f(x)\Delta x. & (4) \end{aligned}$$

The mapping $s(\theta)$ above was introduced in [27] as a modification of the one in [36] for the convex programming problem. The mapping $g(\theta)$ contains the second order effect introduced by the displacement. Obviously, we have $(x(0), s(0)) = (x, s)$. We require our search–directions to satisfy the following equation

$$-\nabla f(x)\Delta x + \Delta s = 0. \qquad (5)$$

Then we see that

$$\begin{aligned} s(\theta) - f(x(\theta)) &= (s + \theta\Delta s) + (f(x + \theta\Delta x) - f(x) - \theta\nabla f(x)\Delta x) - f(x + \theta\Delta x) \\ &= (s - f(x)) + \theta(-\nabla f(x)\Delta x + \Delta s) \qquad (6)\\ &= 0 \end{aligned}$$

for every $\theta \geq 0$, i.e., feasibility is preserved by construction. Consequently, if we find θ such that $(x(\theta), s(\theta)) > 0$ then $(x(\theta), s(\theta))$ is also an interior–feasible point.

The term $g(\theta)$ is continuous and higher order in θ, i.e., $\lim_{\theta\to 0} \|g(\theta)\|/\theta = 0$; hence we have

$$\left.\frac{ds(\theta)}{d\theta}\right|_{\theta=0} = \Delta s.$$

When the mapping f is linear, the term $g(\theta)$ vanishes and we obtain

$$\begin{aligned} x(\theta) &= x + \theta\Delta x, \\ s(\theta) &= s + \theta\Delta s, \end{aligned}$$

which means a usual line search mapping.

Our analysis starts from representing the componentwise complementarity product $x(\theta)s(\theta)$ where $x(\theta)$ and $s(\theta)$ are given by (2) and (3):

$$\begin{aligned} x(\theta)s(\theta) &= (x + \theta\Delta x)(s + \Delta s(\theta)) \\ &= xs + \theta\,(s\Delta x + x\Delta s(\theta)) + \theta^2\Delta x\Delta s(\theta). \qquad (7) \end{aligned}$$

We introduce the primal-dual scaling which is usual in the analysis of interior point methods (see e.g., Gonzaga [5]):

$$d = (xs^{-1})^{1/2}.$$

The important property of this scaling is that it maps both x and s to the same vector v; this makes it possible to express the scaled search–directions as orthogonal components of a vector. In scaled space the search directions are denoted by

$$p_x = d^{-1}\Delta x, \quad p_s = d\Delta s = D\nabla f(x)\Delta x. \tag{8}$$

Let us define the vector p_v as

$$p_v = p_x + p_s.$$

Part of the following lemma is similar to [13, Lemma 4.1].

Lemma 3.1 *Let p_x, p_s, and p_v be as defined above. Then, we have*

(i) $-\kappa\|p_v\|^2 \le \Delta x^T \Delta s = p_x^T p_s \le \frac{\|p_v\|^2}{4}$,

(ii) $\|\Delta x \Delta s\|_\infty = \|p_x p_s\|_\infty \le \frac{1}{4}(1+4\kappa)\|p_v\|^2$.

Proof:

The vectors p_x and p_s satisfy

$$\begin{aligned} -D\nabla f(x) D p_x + p_s &= 0 \\ p_x + p_s &= p_v. \end{aligned}$$

Applying Lemma 3.4 in Kojima et al. [20] gives

$$p_x^T p_s \ge -\kappa\|p_v\|^2. \tag{9}$$

Note that the lemma applies since the $P_*(\kappa)$ property is preserved by pre– and post–multiplication with a positive diagonal matrix (cf. [20, Th. 3.5]). Defining $q_v = p_x - p_s$, it holds

$$p_x^T p_s = \frac{1}{4}(\|p_v\|^2 - \|q_v\|^2) \le \frac{1}{4}\|p_v\|^2.$$

Using (9) we further obtain

$$\|q_v\|^2 \le (1+4\kappa)\|p_v\|^2.$$

Finally, since $p_x p_s = (p_v^2 - q_v^2)/4$

$$\begin{aligned} \|p_x p_s\|_\infty &\le \frac{1}{4}\max\{\|p_v\|_\infty^2, \|q_v\|_\infty^2\} \\ &\le \frac{1}{4}\max\{\|p_v\|^2, \|q_v\|^2\} \\ &\le \frac{1}{4}(1+4\kappa)\|p_v\|^2. \end{aligned}$$

This completes the proof. ∎

For convenience in further discussions we define the mappings:

$$\begin{aligned} p_s(\theta) &= d\Delta s(\theta) = p_s + d\left(\frac{g(\theta)}{\theta}\right), \\ p_v(\theta) &= p_x + p_s(\theta) = p_v + d\left(\frac{g(\theta)}{\theta}\right), \\ v(\theta) &= (x(\theta)s(\theta))^{1/2}. \end{aligned} \tag{10}$$

Using these definitions we may write

$$\begin{aligned} s\Delta x + x\Delta s(\theta) &= v p_v(\theta), \\ \Delta x \Delta s(\theta) &= p_x p_s(\theta). \end{aligned}$$

Hence equation (7) can be rewritten as

$$v(\theta)^2 = x(\theta)s(\theta) = v^2 + \theta v p_v(\theta) + \theta^2 p_x p_s(\theta). \tag{11}$$

The reader may verify that (11) is very similar to the corresponding equation in Jansen et al. [13] for the linear case. Our next task will be to prove results analogous to Lemma 3.1 for the vectors Δx and $\Delta s(\theta)$. To be able to do this, we impose the following condition on the mapping f.

Condition 3.2 *There exists a $\rho \in (0,1)$, $\Theta > 0$ and a $\gamma \geq 0$ such that if $(x,s) \in \mathcal{N}(\rho) := \{(x,s) \in \mathcal{F}^0 : \ \omega(x,s) \geq \rho\}$, and if $(\Delta x, \Delta s) \in \mathbb{R}^{2n}$ satisfies (5) and $\|x^{-1}\Delta x + s^{-1}\Delta s\| \leq 1$, then*

$$\|p_s(\theta) - p_s\| \leq \gamma\theta \|p_s\|$$

for every $\theta \in (0, \Theta]$. Here $\omega(x,s)$, p_s and $p_s(\theta)$ are defined by (1) and (10).

The following observation is useful in the analysis and follows by definition:

$$p_v(\theta) - p_v = p_s(\theta) - p_s.$$

Also, notice that the inequality $\|p_s(\theta) - p_s\| \leq \gamma\theta\|p_s\|$ is equivalent to

$$\left\| d\left(\frac{f(x+\theta\Delta x) - f(x)}{\theta} - \nabla f(x)\Delta x\right)\right\| \leq \gamma\theta \|d\nabla f(x)\Delta x\|, \tag{12}$$

or stated otherwise

$$\left\| d\frac{g(\theta)}{\theta}\right\| \leq \gamma\theta\|d\Delta s\|.$$

Note that the condition depends not only on the mapping f but also on the displacement $(\Delta x, \Delta s)$ used in an algorithm. If the mapping f is linear, however, the above condition holds with $\Theta = +\infty$, $\gamma = 0$ and with every $\rho \in (0,1)$, independent of the search–directions. In Section 5, we will show how Condition 3.2 is related to smoothness conditions on the mapping f and certain primal–dual interior point algorithms.

Using Condition 3.2 we derive the following lemma.

Lemma 3.3 *Let p_x, $p_s(\theta)$, p_v and $p_v(\theta)$ be as defined above and let Condition 3.2 hold. Then*

(i) $\|p_v(\theta)\| \leq (1 + \gamma\theta\sqrt{1+2\kappa})\|p_v\|$.

(ii) $|p_v(\theta) - p_v| \leq \gamma\theta\sqrt{1+2\kappa}\|p_v\|e \leq \frac{\gamma\sqrt{n(1+2\kappa)}}{\omega(|p_v|)}\theta|p_v|$

for any $(x, s) \in \mathcal{N}(\rho)$ and $\theta \in (0, \Theta]$. Here the function ω is defined by (1).

Proof:

The relation $p_v = p_x + p_s$ and (i) of Lemma 3.1 imply

$$\|p_s\|^2 = \|p_v\|^2 - 2p_x^T p_s - \|p_x\|^2 \leq (1 + 2\kappa)\|p_v\|^2. \tag{13}$$

Since Condition 3.2 holds, we have

$$\|p_v(\theta)\| \leq \|p_v\| + \|p_s(\theta) - p_s\| \leq \|p_v\| + \gamma\theta\|p_s\| \leq (1 + \gamma\theta\sqrt{1+2\kappa})\|p_v\|$$

which is the assertion of (i). Similarly,

$$|p_v(\theta) - p_v| = |p_s(\theta) - p_s| \leq \gamma\theta\|p_s\|e \leq \gamma\theta\sqrt{1+2\kappa}\|p_v\|e.$$

Finally, we see that

$$\|p_v\|e = \|p_v\|p_v^{-1}p_v \leq \|p_v\|\,|p_v^{-1}|\,|p_v| \leq \frac{\sqrt{n}|p_v|_{\max}}{|p_v|_{\min}}|p_v| = \frac{\sqrt{n}}{\omega(|p_v|)}|p_v|.$$

This completes the proof of (ii). ∎

We may now prove the following lemma.

Lemma 3.4 *Let p_x, p_s, $p_s(\theta)$, p_v and $p_v(\theta)$ be as defined above and let Condition 3.2 hold. Then, for every $(x, s) \in \mathcal{N}(\rho)$ and $\theta \in (0, \Theta]$, we have*

(i) $-(1 + \gamma\theta)(1 + 2\kappa)\|p_v\|^2 \leq \Delta x^T\Delta s(\theta) = p_x^T p_s(\theta) \leq \frac{1}{4}(1 + \theta\gamma\sqrt{1+2\kappa})^2\|p_v\|^2$.

(ii) $\|\Delta x\Delta s(\theta)\|_\infty = \|p_x p_s(\theta)\|_\infty \leq \left(\frac{1}{4}(1 + \theta\gamma\sqrt{1+2\kappa})^2 + (1 + \gamma\theta)(1 + 2\kappa)\right)\|p_v\|^2$.

Proof:

(i): By (4), (5) and (10), we have

$$p_s(\theta) = d\Delta s(\theta) = d\left(\frac{f(x + \theta\Delta x) - f(x)}{\theta}\right).$$

Using also definition (8) we obtain

$$(\Delta x)^T\Delta s(\theta) = p_x^T p_s(\theta) = \frac{1}{\theta}(\Delta x)^T(f(x + \theta\Delta x) - f(x)).$$

¿From (12) we derive

$$\|d(f(x+\theta\Delta x)-f(x))\| \leq (1+\gamma\theta)\theta\|d\nabla f(x)\Delta x\|.$$

Just as in (13) it holds

$$\|p_x\|^2 = \|p_v\|^2 - 2p_x^T p_s - \|p_s\|^2 \leq (1+2\kappa)\|p_v\|^2. \tag{14}$$

Consequently,

$$\begin{aligned} |\Delta x^T \Delta s(\theta)| &= \frac{1}{\theta}|(\Delta x)^T(f(x+\theta\Delta x)-f(x))| \\ &\leq \frac{1}{\theta}\|d^{-1}\Delta x\|\,\|d(f(x+\theta\Delta x)-f(x))\| \\ &\leq \frac{1}{\theta}\|p_x\|\,(1+\gamma\theta)\theta\|p_s\| \\ &\leq (1+\gamma\theta)(1+2\kappa)\|p_v\|^2, \end{aligned} \tag{15}$$

where the last inequality follows from (13) and (14); this proves the left inequality in (i). For the right we can be better. We proceed as in the proof of Lemma 3.1. By (10) we have

$$p_v(\theta) = p_x + p_s(\theta).$$

Letting $q_v(\theta) = p_x - p_s(\theta)$, we obtain the following bound:

$$p_x^T p_s(\theta) = \frac{1}{4}(\|p_v(\theta)\|^2 - \|q_v(\theta)\|^2) \leq \frac{1}{4}\|p_v(\theta)\|^2 \leq \frac{1}{4}(1+\theta\gamma\sqrt{1+2\kappa})^2\|p_v\|^2, \tag{16}$$

where the last inequality follows from Lemma 3.3. For (ii) observe that combining (15) and (16) leads to the bound

$$\frac{1}{4}\|q_v(\theta)\|^2 \leq \frac{1}{4}(1+\gamma\theta\sqrt{1+2\kappa})^2\|p_v\|^2 + (1+\gamma\theta)(1+2\kappa)\|p_v\|^2.$$

Hence, using

$$p_x p_s(\theta) = \frac{1}{4}(p_v(\theta)^2 - q_v(\theta)^2)$$

we have

$$\begin{aligned} \|p_x p_s(\theta)\|_\infty &\leq \frac{1}{4}\max\{\|p_v(\theta)\|_\infty^2, \|q_v(\theta)\|_\infty^2\} \\ &\leq \frac{1}{4}\max\{\|p_v(\theta)\|^2, \|q_v(\theta)\|^2\} \\ &\leq \frac{1}{4}(1+\gamma\theta\sqrt{1+2\kappa})^2\|p_v\|^2 + (1+\gamma\theta)(1+2\kappa)\|p_v\|^2. \end{aligned}$$

This completes the proof of the lemma. ∎

The above lemmas give us some tools for analyzing primal–dual algorithms applying to NCP. In the linear case. Lemma 3.1 is important to provide the polynomiality of many primal–dual algorithms ([25, 24, 34, 35, 26, 32, 33, 20, 19, 18], etc.). On the other hand, Lemma 3.4 suggests us that these analyses may be extended to nonlinear cases.

Before proceeding we mention that for the monotone NCP the bounds in Lemma 3.4 can be improved by using

$$\Delta x^T \Delta s(\theta) = \frac{1}{\theta^2}(\theta \Delta x)^T (f(x + \theta \Delta x) - f(x)) \geq 0.$$

4 Primal–dual affine scaling algorithms

4.1 Development of the algorithm

Up to this point, our analysis was general, in the sense that we didn't specify our search–directions. The conditions imposed till this point are feasibility (5) and Condition 3.2. We will now derive a family of affine scaling directions as in Jansen et al. [13]. The directions are obtained by minimizing the complementarity (suitably scaled) over an ellipsoid, which is the idea of Dikin's primal affine scaling algorithm [3].

Consider the problem

$$\text{minimize} \;\; x^T s \;\; \text{subject to} \;\; (x, s) \in \mathcal{F}.$$

The NCP is equivalent to the above problem in the sense that (x, s) is a solution of the NCP if and only if it is a minimum solution of the above problem with objective value zero. According to the search mapping defined by (2) and (3), the complementarity $x(\theta)^T s(\theta)$ is obtained by

$$x(\theta)^T s(\theta) = x^T s + \theta \left(s^T \Delta x + x^T \Delta s(\theta) \right) + \theta^2 \Delta x^T \Delta s(\theta).$$

It follows from definition (4) of g that

$$\left. \frac{d(x(\theta)^T s(\theta))}{d\theta} \right|_{\theta=0} \equiv \lim_{\theta \to 0} \frac{x(\theta)^T s(\theta) - x^T s}{\theta} = s^T \Delta x + x^T \Delta s.$$

The above relation gives us an idea for the determination of the search–direction $(\Delta x, \Delta s)$. Let r be a fixed nonnegative constant (the order of scaling in the algorithm). Taking account of the equation $-\nabla f(x) \Delta x + \Delta s = 0$, we consider the following problem, which is essentially the same as the one given in [13]:

$$\begin{aligned} \text{Minimize} \quad & ((xs)^r)^T (x^{-1} \Delta x + s^{-1} \Delta s) \\ \text{subject to} \quad & -\nabla f(x) \Delta x + \Delta s = 0, \\ & \| x^{-1} \Delta x + s^{-1} \Delta s \| \leq 1. \end{aligned}$$

Obviously, the solution of the above system satisfies the assumption imposed on $(\Delta x, \Delta s)$ in Condition 3.2. If we take $r = 1$ then the solution of this subproblem $(\Delta x, \Delta s)$ minimizes the derivative

$$\left. \frac{d(x(\theta)^T s(\theta))}{d\theta} \right|_{\theta=0}.$$

It is not difficult to find that the solution of the KKT (Karush-Kuhn-Tucker) optimality conditions for the above problem satisfies the following system:

$$-\nabla f(x)\Delta x + \Delta s = 0, \tag{17}$$

$$s\Delta x + x\Delta s = -\frac{v^{2r+2}}{\|v^{2r}\|}. \tag{18}$$

The reader may observe that in case of linear or quadratic programming for $r = 0$ this system exactly determines the well-known (classical) affine scaling direction of Monteiro et al. [37]. From this moment on we let p_v have the following definition:

$$p_v := \frac{v^{2r+1}}{\|v^{2r}\|}. \tag{19}$$

Using the definitions in (8), the above optimality system can be rewritten as

$$-D\nabla f(x)Dp_x + p_s = 0, \tag{20}$$

$$p_x + p_s = p_v. \tag{21}$$

Since the Jacobian $\nabla f(x)$ is a $P_*(\kappa)$ matrix the system has a unique solution for every $x \in \mathbb{R}^n$ (cf. [20, Lemma 4.1]).

We can now state our algorithm which is based on [13].

Input
  (x^0, s^0): the initial pair of interior–feasible solutions;
  $r \geq 0$: the degree of scaling;
Parameters
  ε is the accuracy parameter;
  θ is the step size;
begin
  $x := x^0$; $s := s^0$;
  while $x^T s > \varepsilon$ **do**
    Calculate Δx and Δs from (17) and (18);
    Compute the search mapping $(x(\theta), s(\theta))$ by (2) and (3);
    Find $\bar{\theta}$ such that $(x(\bar{\theta}), s(\bar{\theta})) > 0$;
    $x := x(\bar{\theta})$;
    $s := s(\bar{\theta})$;
  end
end.

Figure 1: Primal–dual affine scaling algorithm.

4.2 Convergence results

We will analyze the behavior of the family of primal–dual affine scaling algorithms as follows. First, we will give some general results for the case $r \geq 0$. These regard the complementarity and the feasibility of the iterates after a step. Then, we will analyze the complexity of algorithms with $r > 0$; finally, we consider the classical primal–dual affine scaling algorithm (with $r = 0$) of Monteiro et. al. [37]. Naturally, we will impose Condition 3.2 for those p_s and $p_s(\theta)$ generated in the algorithm under consideration. Hence, in this section we will further assume that p_v is given by (19) for certain constant $r \geq 0$, and that $p_x, p_s, p_s(\theta)$ etc., are obtained from solving (20)–(21).

¿From Lemma 3.3 and Lemma 3.4 we obtain the following result, which is a key for observing the behavior of $v(\theta)^2 = x(\theta)s(\theta)$.

Lemma 4.1 *Suppose that Condition 3.2 holds. Then, for every $(x, s) \in \mathcal{N}(\rho)$ and $\theta \in (0, \Theta]$, we have*

(i) $\|p_v(\theta)\| \leq (1 + \theta\gamma\sqrt{1+2\kappa})\|v\|_\infty \leq (1 + \theta\gamma\sqrt{1+2\kappa})\|v\|$,

(ii) $-\left(1 + \theta\frac{\gamma\sqrt{n(1+2\kappa)}}{\omega(v)^{2r+1}}\right)\frac{v^{2r+2}}{\|v^{2r}\|} \leq vp_v(\theta) \leq -\left(1 - \theta\frac{\gamma\sqrt{n(1+2\kappa)}}{\omega(v)^{2r+1}}\right)\frac{v^{2r+2}}{\|v^{2r}\|}$,

(iii) $p_x^T p_s(\theta) \leq \frac{1}{4}\left(1 + \theta\gamma\sqrt{1+2\kappa}\right)\|v\|^2$,

(iv) $\|p_x p_s(\theta)\|_\infty \leq \left(\frac{1}{4}(1 + \theta\gamma\sqrt{1+2\kappa})^2 + (1+\theta\gamma)(1+2\kappa)\right)\|v\|_\infty^2$.

Proof:

The vector p_v is given by $p_v = -\frac{v^{2r+1}}{\|v^{2r}\|}$. Hence we have

$$\|p_v\| = \frac{\|v^{2r+1}\|}{\|v^{2r}\|} \leq \|v\|_\infty \leq \|v\|,$$
$$\omega(|p_v|) = \omega(v)^{2r+1},$$
$$|p_v| = -p_v.$$

Thus the lemma follows from Lemmas 3.3 and 3.4. ∎

The following lemma is completely equivalent to Lemma 4.2 of [13]. It concerns a technical result used in estimating the new complementarity.

Lemma 4.2 (Lemma 4.2 of [13]) *Let $v \in \mathbb{R}^n_+$ be an arbitrary vector.*

(i) *If $0 \leq r \leq 1$, then* $-\dfrac{e^T v^{2r+2}}{\|v^{2r}\|} \leq -\dfrac{\|v\|^2}{\sqrt{n}}$.

(ii) *If* $1 \le r$, *then* $-\frac{e^T v^{2r+2}}{\|v^{2r}\|} \le -\frac{\omega(v)^{2r-2}}{\sqrt{n}}\|v\|^2$.

Let us introduce some notation:

$$\begin{aligned} \overline{\gamma} &:= \gamma\sqrt{1+2\kappa}, \\ \overline{\vartheta} &:= \frac{\omega(v)^{2r+1}}{2\sqrt{n}}, \\ \overline{\pi} &:= \frac{\gamma\sqrt{n(1+2\kappa)}}{\omega(v)^{2r+1}} = \frac{\overline{\gamma}}{2\overline{\vartheta}}. \end{aligned} \tag{22}$$

Notice that $\overline{\vartheta}$ and $\overline{\gamma}$ depend on v; however, in this section we are concerned with the behavior in one iteration so v can be considered to be fixed. Later we will derive uniform bounds for these quantities. We are now ready to show how the error in complementarity can be reduced by taking a suitable step size θ. Combining Lemma 3.4, Lemma 4.1 and Lemma 4.2 with equation (11) we obtain the following lemma (cf. [13, Lemma 4.3]).

Lemma 4.3 *Let* $\overline{\gamma}, \overline{\vartheta}, \overline{\pi}$ *be as defined in (22). Then*

(i) *If* $0 \le r \le 1$ *and* $\theta \le \min\left\{\Theta, \frac{1}{2\overline{\pi}}, \frac{1}{\sqrt{n}(1+\overline{\vartheta})^2}\right\}$ *then*

$$x(\theta)^T s(\theta) = \|v(\theta)\|^2 \le \left(1 - \frac{\theta}{4\sqrt{n}}\right)\|v\|^2.$$

(ii) *If* $1 \le r$ *and* $\theta \le \min\left\{\Theta, \frac{1}{2\overline{\pi}}, \frac{\omega(v)^{2r-2}}{\sqrt{n}(1+\overline{\vartheta})^2}\right\}$ *then*

$$x(\theta)^T s(\theta) = \|v(\theta)\|^2 \le \left(1 - \frac{\omega(v)^{2r-2}}{4\sqrt{n}}\theta\right)\|v\|^2.$$

Proof:

Since $\theta \le 1/(2\overline{\pi})$ we have $1 + \theta\overline{\gamma} \le 1 + \overline{\vartheta}$. The new complementarity is

$$\begin{aligned} e^T v(\theta)^2 &= e^T v^2 + \theta v^T p_v(\theta) + \theta^2 p_x^T p_s(\theta) \\ &\le \|v\|^2 - \theta(1-\theta\overline{\pi})\frac{e^T v^{2r+2}}{\|v^{2r}\|} + \theta^2\frac{1}{4}(1+\theta\overline{\gamma})^2\|v\|^2 \\ &\le \|v\|^2 - \theta\frac{\|v\|^2}{2\sqrt{n}} + \theta\frac{1}{\sqrt{n}(1+\overline{\vartheta})^2}\frac{1}{4}(1+\overline{\vartheta})^2\|v\|^2 \\ &= \left(1 - \frac{\theta}{4\sqrt{n}}\right)\|v\|^2. \end{aligned}$$

The proof of (ii) is similar. ∎

The following lemma gives us a bound $\bar{\theta}$ such that $(x(\theta), s(\theta)) \in \mathcal{F}^0$ for every $\theta \le \bar{\theta}$, i.e., the new iterate is interior-feasible.

Lemma 4.4 *Let $r \geq 0$ be a given constant and let $\overline{\pi}, \overline{\vartheta}, \overline{\gamma}$ be as defined in (22). Suppose that Condition 3.2 holds. Also let*

$$\overline{\eta}^2 := \frac{1}{4}(1+\overline{\vartheta})^2 + \left(1 + \frac{\overline{\vartheta}}{\sqrt{1+2\kappa}}\right)(1+2\kappa). \tag{23}$$

If $(x,s) \in \mathcal{N}(\rho)$ and

$$0 \leq \theta < \min\left\{\Theta, \frac{1}{2\overline{\pi}}, \frac{3\sqrt{n}\,\omega(v)^{2r}}{2(1+r)}, \frac{\omega(v)}{\overline{\eta}}\left(\sqrt{1+\frac{9\omega(v)^2}{16n\overline{\eta}^2}} - \frac{3\omega(v)}{4\sqrt{n}\overline{\eta}}\right)\right\} \tag{24}$$

then $(x(\theta), s(\theta)) \in \mathcal{F}^0$.

Proof:

The proof is essentially the same as a part of the proof of Theorem 5.1 in [13]. From the fact that the search direction satisfies (5) we have feasibility from (6). We still need to show that $(x(\theta), s(\theta))$ is interior–feasible. The first upper bound $\theta \leq \Theta$ follows from Condition 3.2. Due to the second bound on θ it holds $1 + \theta\overline{\gamma} \leq 1 + \overline{\vartheta}$. ¿From the second bound on θ and (iv) of Lemma 4.1 we get

$$\|p_x p_s(\theta)\|_\infty \leq \overline{\eta}^2 \|v\|_\infty^2.$$

Relation (11) and Lemma 4.1 imply then

$$\begin{aligned} v(\theta)^2 &\leq v^2 - \theta(1-\theta\overline{\pi})\frac{v^{2r+2}}{\|v^{2r}\|} + \theta^2\overline{\eta}^2\|v\|_\infty^2 e, \\ v(\theta)^2 &\geq v^2 - \theta(1+\theta\overline{\pi})\frac{v^{2r+2}}{\|v^{2r}\|} - \theta^2\overline{\eta}^2\|v\|_\infty^2 e, \end{aligned}$$

for every $\theta \in (0, \Theta]$. Now let α be a given positive number and consider the function

$$\phi(t) = t - \theta\alpha\frac{t^{r+1}}{\|v^{2r}\|}. \tag{25}$$

Then one easily verifies that ϕ is monotonically increasing on the interval $[0, v_{\max}^2]$ if $\theta \leq \frac{\|v^{2r}\|}{(1+r)\alpha v_{\max}^{2r}}$ for every $\alpha > 0$. Note that

$$\frac{\|v^{2r}\|}{(1+r)(1-\theta\overline{\pi})v_{\max}^{2r}} \geq \frac{\|v^{2r}\|}{(1+r)(1+\theta\overline{\pi})v_{\max}^{2r}} \geq \frac{2\sqrt{n}\,v_{\min}^{2r}}{3(1+r)v_{\max}^{2r}} = \frac{2\sqrt{n}\,\omega(v)^{2r}}{3(1+r)},$$

hence if we enforce the third upperbound in (24) the largest coordinate $v(\theta)_{\max}$ of $v(\theta)$ and the smallest coordinate $v(\theta)_{\min}$ can be estimated as follows:

$$v(\theta)_{\max}^2 \leq v_{\max}^2 - \theta(1-\theta\overline{\pi})\frac{v_{\max}^{2r+2}}{\|v^{2r}\|} + \theta^2\overline{\eta}^2 v_{\max}^2, \tag{26}$$

$$v(\theta)_{\min}^2 \geq v_{\min}^2 - \theta(1+\theta\overline{\pi})\frac{v_{\min}^{2r+2}}{\|v^{2r}\|} - \theta^2\overline{\eta}^2 v_{\max}^2. \tag{27}$$

By the continuity of the mapping v, if $v(\theta)^2_{\min} > 0$ for every $\theta \in [0, \bar{\theta}]$, then $(x(\theta), s(\theta)) > 0$ and consequently $(x(\theta), s(\theta)) \in \mathcal{F}^0$ for this interval. Hence it suffices to show that $v(\theta)^2_{\min} > 0$ if θ satisfies (24).

Dividing the relation (27) by $v^2_{\min}$ gives us the following condition to ensure that $v(\theta)_{\min} > 0$:

$$1 - \theta(1 + \theta\overline{\pi})\frac{v^{2r}_{\min}}{\|v^{2r}\|} - \theta^2 \frac{\overline{\eta}^2}{\omega(v)^2} > 0$$

Since

$$\frac{v^{2r}_{\min}}{\|v^{2r}\|} \leq \frac{v^{2r}_{\min}}{\sqrt{n}\, v^{2r}_{\min}} = \frac{1}{\sqrt{n}}$$

and $1 + \theta\overline{\pi} \leq 3/2$ the condition certainly holds if

$$1 - \theta\frac{3}{2\sqrt{n}} - \theta^2 \frac{\overline{\eta}^2}{\omega(v)^2} > 0.$$

It is easy to check that the last upper bound in the lemma ensures the inequality above. This completes the proof of the lemma. ∎

By putting the above together, the following bound on θ enjoys both the results in Lemma 4.3 and Lemma 4.4:

$$0 \leq \theta < \min\left\{\Theta, \frac{1}{2\overline{\pi}}, \frac{1}{\sqrt{n}(1+\overline{\vartheta})^2}, \frac{2\sqrt{n}\,\omega(v)^{2r}}{3(1+r)}, \frac{\omega(v)}{\overline{\eta}}\left(\sqrt{1 + \frac{9\omega(v)^2}{16n\overline{\eta}^2}} - \frac{3\omega(v)}{4\sqrt{n}\overline{\eta}}\right)\right\}. \tag{28}$$

If the mapping f is linear, we can take $\Theta = +\infty$ and $\gamma = 0$ in Condition 3.2, and consequently $\overline{\gamma} = \overline{\pi} = 0$. Furthermore, if f has the monotonicity property, Condition 2.2 holds with $\kappa = 0$. In this case, we obtain the bound $5/4 \leq \overline{\eta}^2 \leq 3$, since $0 < \overline{\vartheta} \leq 1$. Thus, Lemma 4.3 and Lemma 4.4 above almost coincide with Lemma 4.3 and a part of Theorem 5.1 of [13] for linear monotone complementarity problems.

4.3 The polynomial convergence for $r > 0$

This section is devoted to analyzing the polynomiality of the class of primal–dual affine scaling algorithms for $r > 0$. For a given parameter $\rho \in (0, 1)$, each algorithm in this class generates a sequence of iterates $\{(x^k, s^k) : k = 1, 2, \ldots\}$ satisfying $\omega(v^k) \geq \rho$ for every $k \geq 1$ where the function ω is given by (1). This condition on the iterates is in fact equivalent to requiring that the iterates are in the wide neighborhood defined by infinity norms that has been used in the analysis of interior point methods in e.g. [25] and [1]; the wide neighborhood is given by

$$\beta \in (0, 1), \;\; \mu := \frac{x^T s}{n}, \;\; (1 - \beta)\mu \leq x_i s_i \leq (1 + \beta)\mu.$$

Suppose that the current point belongs to $(x, s) \in \mathcal{N}(\rho)$, i.e., $(x, s) \in \mathcal{F}^0$ and $\omega(v) \geq \rho$. Our algorithm determines the next point along the curve $(x(\theta), s(\theta))$ given by (2) and (3) by choosing a step size θ. The following theorem ensures the existence of $\bar{\theta} > 0$ for which $(x(\theta), s(\theta)) \in \mathcal{F}^0$ and $\omega(v(\theta)) \geq \rho$ for every $\theta \in (0, \bar{\theta}]$.

Theorem 4.5 *Let $r > 0$ be a given constant and let $\overline{\pi}, \overline{\vartheta}, \overline{\gamma}$ be as in (22). Suppose that Condition 3.2 holds. If $(x, s) \in \mathcal{N}(\rho)$, θ satisfies (28) and*

$$0 \leq \theta \leq \min \left\{ \frac{1 - \rho^{2r}}{2\overline{\pi}(1 + \rho^{2r})}, \frac{\rho^2(1 - \rho^{2r})}{2\overline{\eta}^2(1 + \rho^2)\sqrt{n}} \right\}, \tag{29}$$

then $(x(\theta), s(\theta)) \in \mathcal{N}(\rho)$.

Proof:

The part $(x(\theta), s(\theta)) \in \mathcal{F}^0$ is obvious from Lemma 4.4. Hence we need only show that $\omega(v(\theta)) \geq \rho$, i.e., $\rho^2 v(\theta)^2_{\max} \leq v(\theta)^2_{\min}$.

Utilizing the relation $v_{\min} = \omega(v)v_{\max} \geq \rho v_{\max}$, the same discussion for finding the bounds (26) and (27) leads to the following relation:

$$v(\theta)^2_{\min} \geq \rho^2 \left(v^2_{\max} - \theta(1 + \theta\overline{\pi}) \frac{\rho^{2r} v^{2r+2}_{\max}}{\|v^{2r}\|} - \theta^2 \frac{\overline{\eta}^2}{\rho^2} v^2_{\max} \right). \tag{30}$$

Hence, from (26) and (30) we derive a sufficient condition for θ as follows:

$$\rho^2 \left(v^2_{\max} - \theta(1 - \theta\overline{\pi}) \frac{v^{2r+2}_{\max}}{\|v^{2r}\|} + \theta^2 \overline{\eta}^2 v^2_{\max} \right) \leq \rho^2 \left(v^2_{\max} - \theta(1 + \theta\overline{\pi}) \frac{\rho^{2r} v^{2r+2}_{\max}}{\|v^{2r}\|} - \theta^2 \frac{\overline{\eta}^2}{\rho^2} v^2_{\max} \right).$$

By rearranging this inequality, we have

$$\theta \frac{\overline{\eta}^2(1 + \rho^2)}{\rho^2} \leq \frac{v^{2r}_{\max}}{\|v^{2r}\|} \left((1 - \theta\overline{\pi}) - (1 + \theta\overline{\pi})\rho^{2r} \right).$$

Since $\|v^{2r}\| \leq \sqrt{n}\, v^{2r}_{\max}$, we obtain the bound

$$\theta \leq \frac{\rho^2 \left((1 - \theta\overline{\pi}) - (1 + \theta\overline{\pi})\rho^{2r} \right)}{\overline{\eta}^2(1 + \rho^2)\sqrt{n}}$$

Using the first bound in (29) we find that θ will certainly satisfy this inequality if

$$\theta \leq \frac{\rho^2(1 - \rho^{2r})}{2\overline{\eta}^2(1 + \rho^2)\sqrt{n}}$$

Thus we obtain the theorem. ∎

We are now in the position to derive the complexity of our algorithms. The following theorem can be derived from Lemma 4.3 and the above theorem.

Theorem 4.6 (Theorem 5.2 of [13]) *Let $r > 0$ be a given constant and let $0 < \rho < 1$ be given. Suppose that Condition 3.2 is satisfied. Let $\epsilon > 0$, $(x^0, s^0) \in \mathcal{N}(\rho)$ be given and let θ satisfy the conditions in (28) and (29). Then the primal-dual affine scaling algorithm with order of scaling r stops with a solution (x^*, s^*) for which $(x^*)^T s^* \leq \epsilon$ and $\omega(v^*) \geq \rho$ holds, after at most*

(i) $\frac{4\sqrt{n}}{\theta} \ln \frac{(x^0)^T s^0}{\epsilon}$ *iterations if* $0 < r \leq 1$.

(ii) $\frac{4\sqrt{n}}{\rho^{2r-2}\theta} \ln \frac{(x^0)^T s^0}{\epsilon}$ *iterations if* $1 < r$.

To be more specific about the complexity, we have to check which of the various conditions on the step size θ is strongest and how θ depends on the input parameters. This will be done below.

It is easy to verify the following bounds

$$\begin{array}{rcccl} \frac{\rho^{2r+1}}{2\sqrt{n}} & \leq & \overline{\vartheta} & \leq & \frac{1}{2\sqrt{n}}, \\ \gamma\sqrt{n(1+2\kappa)} & \leq & \overline{\pi} & \leq & \frac{\gamma\sqrt{n(1+2\kappa)}}{\rho^{2r+1}}. \end{array}$$

Using also $n \geq 2$, we get from (23)

$$\frac{5}{4} \leq \overline{\eta}^2 \leq \frac{1}{4}\left(1+\frac{1}{2\sqrt{2}}\right)^2 + \left(1+\frac{1}{2\sqrt{2}}\right)(1+2\kappa) < 3(1+\kappa). \tag{31}$$

We analyze the bounds in (28) consecutively. Notice,

$$\frac{1}{2\overline{\pi}} \geq \frac{\rho^{2r+1}}{2\gamma\sqrt{n(1+2\kappa)}}.$$

$$\frac{1}{\sqrt{n}(1+\overline{\vartheta})^2} \geq \frac{1}{\sqrt{n}(1+1/(2\sqrt{2}))^2} \geq \frac{1}{2\sqrt{n}}.$$

$$\frac{2\sqrt{n}\omega(v)^{2r}}{3(1+r)} \geq \frac{2\sqrt{n}\rho^{2r}}{3(1+r)}.$$

We have

$$\frac{3\omega(v)}{4\sqrt{n}\overline{\eta}} \leq \frac{3}{4\sqrt{2}\sqrt{5/4}} < 1,$$

so using the fact that the function $\phi(t) = \sqrt{1+t^2} - t$ is monotonically decreasing we obtain

$$\frac{\omega(v)}{\overline{\eta}}\left(\sqrt{1+\frac{9\omega(v)^2}{16n\overline{\eta}^2}} - \frac{3\omega(v)}{4\sqrt{n}\overline{\eta}}\right) \geq \frac{\omega(v)}{\overline{\eta}}(\sqrt{2}-1) \geq \frac{\rho}{5\sqrt{1+\kappa}}.$$

For the bounds in (29) we obtain

$$\frac{1-\rho^{2r}}{2\overline{\pi}(1+\rho^{2r})} \geq \frac{(1-\rho^{2r})\rho^{2r+1}}{2\gamma\sqrt{n(1+2\kappa)}(1+\rho^{2r}} \geq \frac{(1-\rho^{2r})\rho^{2r+1}}{4\gamma\sqrt{n(1+\kappa)}(1+\rho^{2r}}$$

$$\frac{\rho^2(1-\rho^{2r})}{2\bar{\eta}^2(1+\rho^2)\sqrt{n}} \geq \frac{\rho^2(1-\rho^{2r})}{6(1+\kappa)(1+\rho^2)\sqrt{n}}.$$

Thus we obtain the following result as a corollary of Theorem 4.6.

Corollary 4.7 *Let us take the situation as in Theorem 4.6, besides that we specify $\rho = 1/\sqrt{2}$.*

(i) *If $0 < r \leq 1$ and $n \geq 2$ we may choose*

$$\theta = \min\left\{\Theta, \frac{1-2^{-r}}{12\gamma\sqrt{n(1+\kappa)}}, \frac{1-2^{-r}}{18\sqrt{n}(1+\kappa)}\right\},$$

hence the complexity of the algorithm is

$$\mathcal{O}\left(\sqrt{n}\max\left\{\frac{1}{\Theta}, \frac{\sqrt{n(1+\kappa)}}{1-2^{-r}}\max\left\{\gamma, \sqrt{1+\kappa}\right\}\right\}\ln\frac{(x^0)^T s^0}{\epsilon}\right).$$

(ii) *If $r = 1$ and $n \geq 2$ then we may choose*

$$\theta = \min\left\{\Theta, \frac{1}{24\gamma\sqrt{n(1+\kappa)}}, \frac{1}{36\sqrt{n}(1+\kappa)}\right\},$$

hence the complexity of the algorithm is

$$\mathcal{O}\left(\sqrt{n}\max\left\{\frac{1}{\Theta}, \sqrt{n(1+\kappa)}\max\left\{\gamma, \sqrt{1+\kappa}\right\}\right\}\ln\frac{(x^0)^T s^0}{\epsilon}\right).$$

(iii) *If $1 < r$ and n sufficiently large we may choose*

$$\theta = \min\left\{\Theta, \frac{1}{2^{r+2}\gamma\sqrt{n(1+\kappa)}}, \frac{1}{36\sqrt{n}(1+\kappa)}\right\}$$

hence the complexity of the algorithm is

$$\mathcal{O}\left(2^r\sqrt{n}\max\left\{\frac{1}{\Theta}, \sqrt{n(1+\kappa)}\max\left\{2^r\gamma, \sqrt{1+\kappa}\right\}\right\}\ln\frac{(x^0)^T s^0}{\epsilon}\right).$$

Remark 4.8 Notice that the neighborhood $\mathcal{N}(\rho)$ fulfills two roles in the above analysis. The first is as the admissible region in which the algorithm generates the sequence as in Section 5 of [13]. In this case, the initial point (x^0, s^0) prescribes the possible choices of ρ so that $\rho \leq \omega(x^0, s^0)$. The second is as the region where the nonlinear mapping f can be approximated suitably as in Condition 3.2. If the mapping f is linear, however, Condition 3.2 holds with every $\rho \in (0,1)$; hence we can start from an arbitrary initial point $(x^0, s^0) \in \mathcal{F}^0$. In addition, $\Theta = +\infty$ and $\gamma = 0$ in those cases, and $\kappa = 0$ if f is monotone; hence the above corollary corresponds to Corollary 5.1 of [13] if f is linear and monotone. Observe also, that for $r \geq 1$ the complexity bound gets worse as r increases, and similarly for $r < 1$ as r decreases to zero. A final remark to be made is that the actual value of κ (which might be hard to compute) is not required in practice, although the theoretical step size and complexity depend on it. In an implementation we can just compute the maximal step that ensures $\omega(x^0, s^0) = \rho$, which is guaranteed to be as large as the theoretical step.

4.4 Polynomial complexity if $r = 0$

In this section we show that, with suitable step size, the classical primal–dual affine scaling algorithm of Monteiro et al. [37] can be applied to NCPs satisfying Conditions 2.2 and 3.2 with a polynomial complexity bound. We believe that this is the first proof of polynomial convergence of the affine scaling algorithm for NCPs.

So from now on we assume that $r = 0$. It is easily verified that Lemma 4.3 and Lemma 4.4 still apply in the present case. Theorem 4.5, however, is not valid for $r = 0$. In fact, by taking the limit in (29) as r tends to zero one obtains that the step size θ becomes zero. Below we show that by making a positive step, (i.e., $\theta > 0$) $\omega(v)$ may well decrease, but the decrease can be bounded from below. This is the contents of the next lemma.

Lemma 4.9 *Let $\overline{\pi}, \overline{\vartheta}$ be as defined in (22) and $\overline{\eta}$ as in (23). If $(x,s) \in \mathcal{F}^0$ and*

$$0 \le \theta < \min\left\{\Theta, \frac{1}{2\overline{\pi}}, \frac{1}{\sqrt{n}(1+\overline{\vartheta})^2}, \frac{2\sqrt{n}}{3}, \frac{\omega(v)}{\overline{\eta}}\left(\sqrt{1+\frac{9\omega(v)^2}{16n\overline{\eta}^2}} - \frac{3\omega(v)}{4\sqrt{n}\overline{\eta}}\right)\right\} \tag{32}$$

then $(x(\theta), s(\theta)) \in \mathcal{F}^0$ and

$$1 + \omega(v(\theta)^2) \ge \frac{1+\omega(v^2)}{1 + \frac{\theta^2(\overline{\eta}^2\sqrt{n}+2\overline{\pi})}{\sqrt{n}-\theta(1+\theta\overline{\pi})}}. \tag{33}$$

Proof:

It may be clear from Lemma 4.4 that the given bounds (32) on θ guarantee the feasibility of the new iterate $(x(\theta), s(\theta))$. So it remains to show that (33) holds. First observe that (26) and (27) also hold for $r = 0$. Hence, by using the notation $\omega^2 = \omega(v^2) = \alpha/\beta$ with α and β such that $\alpha e \le v^2 \le \beta e$, one has

$$\begin{aligned}
\omega(v(\theta)^2) &\ge \frac{(1-\theta(1+\theta\overline{\pi})/\sqrt{n})\alpha - \theta^2\overline{\eta}^2\beta}{(1-\theta(1-\theta\overline{\pi})/\sqrt{n})\beta + \theta^2\overline{\eta}^2\beta} \\
&= \frac{\omega^2(\sqrt{n}-\theta(1+\theta\overline{\pi})) - \theta^2\overline{\eta}^2\sqrt{n}}{(\sqrt{n}-\theta(1+\theta\overline{\pi})) + \theta^2\overline{\eta}^2\sqrt{n} + 2\theta^2\overline{\pi}} \\
&\ge \frac{\omega^2(\sqrt{n}-\theta(1+\theta\overline{\pi})) - \theta^2\overline{\eta}^2\sqrt{n} - 2\theta^2\overline{\pi}}{(\sqrt{n}-\theta(1+\theta\overline{\pi})) + \theta^2\overline{\eta}^2\sqrt{n} + 2\theta^2\overline{\pi}}.
\end{aligned}$$

After adding one to both sides and rearranging the terms the inequality (33) follows. ∎

Now we are ready to prove the polynomial complexity. We will denote by $(x^{(k)}, s^{(k)})$ the iterate after k iterations and for simplicity we use the notation $\omega_k := \omega(x^{(k)}, s^{(k)})$.

Theorem 4.10 *Let an initial interior point $(x^{(0)}, s^{(0)}) \in \mathcal{F}^0$, with $\sqrt{2\rho} \le \omega_0 \le 1$ and $0 < \epsilon < (x^{(0)})^T s^{(0)}$ be given. We define parameters L and τ as follows.*

$$L := \ln\frac{(x^{(0)})^T s^{(0)}}{\epsilon}, \quad \tau := \frac{64}{\omega_0^2}\left\{(1+\kappa) + \frac{\gamma}{\omega_0}\sqrt{n(1+\kappa)}\right\} + \frac{2}{nL},$$

We assume that $L \geq 1$ and $n \geq 2$. Let t be the smallest real number in the interval $(\tau, \tau+1/4nL^2]$ such that $K := 2tnL^2$ is integral. If $\theta := 1/(t\sqrt{n}L) \leq \Theta$, then after $K = O(nL^2/(\omega_0^2))$ iterations the algorithm yields a solution (x^*, s^*) such that $(x^*)^T s^* \leq \epsilon$ and $\omega(x^*s^*) \geq \frac{\omega_0^2}{2} \geq \rho$.

Proof:

For the moment we make the assumption that in each iteration of the algorithm the step size $\theta = \frac{1}{t\sqrt{n}L}$ satisfies the conditions of Lemma 4.9. Later on we will justify this assumption. Taking logarithms in (33), and substituting the given value of θ, we obtain

$$\begin{aligned}
\ln \frac{1+\omega_0^2}{1+\omega_k^2} &\leq k \ln \left(1 + \frac{\theta^2(\overline{\eta}^2\sqrt{n} + 2\overline{\pi})}{\sqrt{n} - \theta(1+\theta\overline{\pi})}\right) \\
&\leq k \frac{\theta^2(\overline{\eta}^2\sqrt{n} + 2\overline{\pi})}{\sqrt{n} - \theta(1+\theta\overline{\pi})} \\
&\leq k \frac{1/(t^2nL^2)(\overline{\eta}^2 + 2\overline{\pi})\sqrt{n}}{\sqrt{n} - 2/(t\sqrt{n}L)} \\
&= k \frac{\overline{\eta}^2 + 2\overline{\pi}}{tL(tnL-2)},
\end{aligned}$$

where we used $\overline{\pi} < \overline{\pi}\sqrt{n}$ and $\theta\overline{\pi} < 1$ in the third inequality. Hence we certainly have $\omega_k^2 \geq \omega_0^2/2 \geq \rho$ as long as

$$k \frac{\overline{\eta}^2 + 2\overline{\pi}}{tL(ntL-2)} \leq \ln \frac{1+\omega_0^2}{1+\frac{\omega_0^2}{2}}. \tag{34}$$

Since $\phi(\sigma) := \ln((1+\sigma)/(1+\sigma/2))$ is a concave function, and $\phi(0) = 0$, $\phi(1) \geq \frac{1}{4}$, it holds

$$\ln \frac{1+\omega_0^2}{1+\frac{\omega_0^2}{2}} \geq \frac{\omega_0^2}{4}.$$

As a consequence, inequality (34) is certainly satisfied if

$$k \leq \frac{\omega_0^2 tL(tnL-2)}{4(\overline{\eta}^2 + 2\overline{\pi})}. \tag{35}$$

We conclude that if the total number of iterations satisfies (35) then the inequality $\omega_k^2 \geq \frac{\omega_0^2}{2}$ holds.

Since Lemma 4.3 is valid and we employ a fixed step parameter θ, the algorithm stops after at most k iterations, where

$$k \geq \frac{4\sqrt{n}}{\theta} \ln \frac{(x^{(0)})^T s^{(0)}}{\epsilon} = 4tnL^2,$$

and then we have $(x^{(k)})^T s^{(k)} \leq \epsilon$. Note that the definition of t guarantees that $4tnL^2$ is integral. So, as far as the reduction in complementarity is concerned, the algorithm needs not more than $4tnL^2$ iterations. This number of iterations will respect the bound (35) if

$$4tnL^2 \leq \omega_0^2 tL \frac{tnL-2}{4(\overline{\eta}^2 + 2\overline{\pi})}.$$

Dividing by $\omega_0^2 tL$ and rearranging terms gives the condition

$$t \geq \frac{16(\overline{\eta}^2 + 2\overline{\pi})}{\omega_0^2} + \frac{2}{nL}. \tag{36}$$

which is satisfied by the value assigned to t in the theorem, since $\overline{\eta}^2 \leq 2(1+\kappa)$ and

$$\overline{\pi} = \frac{\gamma\sqrt{n(1+2\kappa)}}{\omega_k} \leq \frac{\gamma\sqrt{2n(1+2\kappa)}}{\omega_0} \leq \frac{2\gamma\sqrt{n(1+\kappa)}}{\omega_0}$$

It remains to show that in each iteration of the algorithm the specified step size θ satisfies condition (32) of Lemma 4.9. First, observe that $\theta < \Theta$ by assumption. ¿From (36) we have

$$t \geq \frac{32\overline{\pi}}{\omega_0^2} \geq 32\overline{\pi}.$$

The condition $\theta \leq 1/(2\overline{\pi})$ is equivalent to

$$t \geq \frac{2\overline{\pi}}{\sqrt{n}L},$$

which is guaranteed by the assumption on L. Since $\overline{\vartheta} \leq \frac{1}{2\sqrt{n}}$, a sufficient condition for the third bound in (32) is

$$\theta \leq \frac{1}{\sqrt{n}(1+1/(2\sqrt{n}))},$$

which is satisfied since $t \geq 2$ from (36) and $L \geq 1$. The fourth condition is trivially guaranteed, so it remains to deal with the condition that for each k

$$\theta < \frac{\omega_k}{\overline{\eta}}\left(\sqrt{1+\frac{9\omega_k^2}{16n\overline{\eta}^2}} - \frac{3\omega_k}{4\sqrt{n}\overline{\eta}}\right).$$

Using $n \geq 2$ and $\overline{\eta} \geq \sqrt{5/4}$ (see (31)), we have

$$\frac{3\omega_k}{4\sqrt{n}\overline{\eta}} \leq \frac{\omega_k}{2} \leq \frac{1}{2}.$$

Therefore, since

$$\sqrt{1+\sigma^2} - \sigma > \frac{1}{2} \quad \text{if} \quad 0 \leq \sigma < \frac{3}{4},$$

it is sufficient to show that $\theta \leq \omega_k/(2\overline{\eta})$, for each k. As we have seen before, for the given step size we have $\omega_k \geq \omega_0/(\sqrt{2})$ for each k. So is it sufficient that θ satisfies $\theta \leq \omega_0/(2\sqrt{2}\overline{\eta})$ or even $\theta \leq \omega_0/(2\sqrt{6}\sqrt{1+\kappa})$. This amounts to $\omega_0 t\sqrt{n}L \geq 2\sqrt{6}\sqrt{1+\kappa}$. Due to the assumption that $L \geq 1$ and $n \geq 2$, this certainly holds if t satisfies $t \geq \frac{\sqrt{1+\kappa}}{\omega_0}$. Using $\omega_0 \leq 1$, $t \geq \tau$ and the definition of τ, it is obvious that t satisfies this inequality. Hence the proof of the theorem is complete. ∎

As the proof of the above theorem shows, we need the assumption $\omega_0 \geq \sqrt{2\rho}$ on the initial point (x^0, s^0) so that the generated sequence $\{(x^k, s^k)\}$ lies in $\mathcal{N}(\rho)$ for which Condition 3.2 is effective. However, as we have mentioned in Remark 4.8, if f is linear then we can take any $\rho \in (0,1)$, and the assumption does not affect any choice of the initial point.

5 Smoothness conditions on the mapping f

In the literature on interior point methods for nonlinear programming problems a few smoothness conditions on the functions involved have been proposed. These conditions serve the purpose of bounding the second order effects not taken into account by Newton's method. Since these conditions are only applicable to *monotone* mappings, we confine ourselves in this section to mappings f satisfying the monotonicity property. In this section we show how these conditions are generalized to the setting of NCP and indicate their use for analyzing algorithms in wide neighborhoods.

We first introduce some notation regarding a trilinear form $N \in \mathbb{R}^{n \times n \times n}$. We use the notation

$$N[x, y, z] = \sum_i \sum_j \sum_k x_i y_j z_k N^i_{jk},$$

where N^i, $i = 1, \ldots, n$, are matrices.

5.1 Zhu's scaled Lipschitz condition

In this section we will show that the (modified) scaled Lipschitz condition as introduced by Zhu [49] also guarantees Condition 3.2 to hold. The scaled Lipschitz condition was also used by Potra and Ye [43] for the analysis of Newton's method for monotone NCP, by Kortanek et al. [28] for an analysis of a primal–dual method for entropy optimization problems and by Sun et al. [44] for min–max saddle point problems.

Definition 5.1 *Let G be a closed convex domain in $\mathbb{R}^n$, with nonempty interior $Q := \text{int}(G)$. A single–valued monotone operator $f : Q \to \mathbb{R}^n$ satisfies the scaled Lipschitz condition if there is a nondecreasing function $\psi(\alpha)$ such that*

$$\|X(f(x+h) - f(x) - \nabla f(x)h)\| \leq \psi(\alpha) h^T \nabla f(x) h \tag{37}$$

for all $x > 0$ and h satisfying $\|x^{-1}h\| \leq \alpha$.

We will show that any mapping satisfying the scaled Lipschitz condition also satisfies Condition 3.2.

Theorem 5.2 *Let the mapping f satisfy the scaled Lipschitz condition with $G = \{x \in R^n : x \geq 0\}$. Then, for every $\rho \in (0, 1)$, there exist values for γ and Θ such that f satisfies Condition 3.2.*

Proof:

Suppose that $(x, s) \in \mathcal{N}(\rho)$ and $(\Delta x, \Delta s) \in \mathbb{R}^{2n}$ satisfy (5) and $\|x^{-1}\Delta x + s^{-1}\Delta s\| \leq 1$. Since f is monotone, Lemma 3.1 holds with $\kappa = 0$. Hence

$$\begin{aligned}
\|x^{-1}\Delta x\|^2 &= \|v^{-1}p_s\|^2 \\
&\leq \|v^{-1}\|^2(\|p_x\|^2 + 2p_x^T p_s + \|p_s\|^2) \\
&\leq \frac{1}{v_{\min}^2}\|p_x + p_s\|^2 \\
&\leq \frac{v_{\max}^2}{v_{\min}^2}\|x^{-1}\Delta x + s^{-1}\Delta s\|^2 \\
&\leq \frac{1}{\rho^2},
\end{aligned}$$

where the last inequality follows from the fact that $\omega(v) = \frac{v_{\min}}{v_{\max}} \geq \rho$ and $\|x^{-1}\Delta x + s^{-1}\Delta s\| \leq 1$. Thus, if $\Theta \leq \alpha\rho$ then $\|\theta x^{-1}\Delta x\| \leq \alpha$ for every $\theta \in (0, \Theta]$, and we have

$$\begin{aligned}
\|p_s(\theta) - p_s\| &= \left\| d\left(\frac{f(x+\theta\Delta x) - f(x)}{\theta} - \nabla f(x)\Delta x\right)\right\| \\
&\leq \frac{1}{\theta}\|v^{-1}\|_\infty \|x(f(x+\theta\Delta x) - f(x) - \theta\nabla f(x)\Delta x)\| \\
&\leq \frac{1}{\theta}\|v^{-1}\|_\infty \psi(\alpha)\theta^2 \Delta x^T \nabla f(x)\Delta x \\
&\leq \theta\|v^{-1}\|_\infty \psi(\alpha)\|d^{-1}\Delta x\| \, \|d\nabla f(x)\Delta x\| \\
&\leq \theta\|v^{-1}\|_\infty \psi(\alpha)\|v\|_\infty \|x^{-1}\Delta x\| \, \|d\nabla f(x)\Delta x\| \\
&\leq \theta\frac{1}{\omega(v)}\psi(\alpha)\|x^{-1}\Delta x\| \, \|p_s\|, \\
&\leq \theta\frac{\psi(\alpha)}{\rho^2}\|p_s\|.
\end{aligned}$$

Consequently, we obtain Θ and γ as

$$\Theta = \alpha\rho \quad \text{and} \quad \gamma = \frac{\psi(\alpha)}{\rho^2}.$$

This completes the proof. ∎

Definition 5.1 of the scaled Lipschitz condition implies that $h^T\nabla f(x)h \geq 0$ for every $x > 0$ and h with $\|x^{-1}h\| \leq \alpha$, a priori. Hence, using this condition seems not to be suitable when we analyze the case where the mapping f has no monotonicity. In fact, even if f is linear, i.e. given by $f(x) = Mx + q$, the condition does not necessarily hold for the matrices M in P_*. On the other hand, Condition 3.2 does not need the monotonicity and holds for any linear mapping, which may be an advantage of the condition.

5.2 Self–concordance and relative Lipschitz condition

The most important (and most general) condition is *self–concordance* introduced by Nesterov and Nemirovskii [40], later also used by, e.g., Jarre [16], Den Hertog [10] and Den Hertog et al. [11], Freund and Todd [4], Nesterov and Todd [41]. The crux of the condition is that it bounds the third order derivative of a function in its second order derivative. In [40, Chapter 7] it is shown how the condition can be generalized to nonlinear mappings. A main difference with the Lipschitz conditions is that the self-concordance does not immediately apply to the mapping itself; what is needed is a so-called self-concordant barrier for the domain (in our case $\mathbb{R}^n_+$, and we may use the function $-\sum_{i=1}^n \ln x_i$) and a mapping that is β-compatible with this barrier function. We have the following definition. Since these conditions only apply to monotone mappings, we assume f to be monotone in this section.

Definition 5.3 *A C^2 smooth monotone operator $f : \mathbb{R}^n_+ \to \mathbb{R}^n$ is called β-compatible with $F(x) = -\sum_{i=1}^n \ln x_i$ if for all $x > 0$ and $h_i \in \mathbb{R}^n$, $i = 1, 2, 3$, the inequality*

$$|f''(x)[h_1, h_2, h_3]| \leq 3^{3/2}\beta \prod_{i=1}^{3} \{f'(x)[h_i, h_i]^{1/3} \|x^{-1}h_i\|^{1/3}\}$$

holds.

If the mapping f is β compatible with the function $-\sum_{i=1}^n \ln x_i$ then the barrier function $f_t(x) := (1+\beta)^2\{tf(x) + x^{-1}\}$ is strongly self-concordant for all $t > 0$ (see Prop. 7.3.2 of [40]). A similar property can be obtained if the mapping f is strongly self-concordant itself. For all $\alpha > 0$ and $\beta > 0$, if the mappings ϕ and ψ are self-concordant then the mapping $\alpha\phi + \beta\psi$ is also self-concordant with some parameter. Self-concordant mappings satisfy the following condition, which may be called the 'relative Lipschitz condition' (see Prop. 7.2.1 of [40] and Section 2.1.4 of [17]).

Definition 5.4 *Let G be a closed convex domain in $\mathbb{R}^n$, with nonempty interior $Q := \text{int}(G)$. A single-valued monotone operator $h : Q \to \mathbb{R}^n$ satisfies the relative Lipschitz condition if for all $x \in Q$, $y \in Q$ for which $r := \sqrt{(y-x)^T \nabla h(x)(y-x)} \leq 1$ the inequality*

$$|q^T(\nabla h(y) - \nabla h(x))q| \leq \left(\frac{1}{(1-r)^2} - 1\right) q^T \nabla h(x) q.$$

holds for all $q \in \mathbb{R}^n$.

In the following lemma, we will show how β-compatibility and the relative Lipschitz condition can be used to bound the inner product $p_x^T p_s(\theta)$, which plays an important role in the complexity analysis, see Lemma 3.3 and Lemma 3.4.

Lemma 5.5 *Let the mapping f be β-compatible with $F(x) = -\sum \ln x_i$ and $\rho \in (0, 1)$. If $(x, s) \in \mathcal{N}(\rho) := \{(x, s) \in \mathcal{F}^0 : \ \omega(x, s) \geq \rho\}$, and if $(\Delta x, \Delta s) \in \mathbb{R}^{2n}$ satisfy (5) and $\|x^{-1}\Delta x + s^{-1}\Delta s\| \leq 1$, then*

$$p_x^T p_s(\theta) \leq \frac{1}{4}\left(1 + \theta \frac{15\beta v_{\min}}{\rho^2 \|p_v\|}\right) \|p_v\|^2.$$

for every $\theta > 0$. Here $\omega(x, s)$, p_s and $p_s(\theta)$ are defined by (1) and (10).

Proof:

Using Taylor's expansion, we have

$$\begin{aligned} g(\theta) &= f(x+\theta\Delta x) - f(x) - \theta\nabla f(x)\Delta x \\ &= \frac{1}{2}\theta^2\ f''(y)[\Delta x, \Delta x, \cdot], \end{aligned}$$

where $y = x + \theta\lambda\Delta x$ for some $\lambda \in \mathbb{R}^n_+$ with $\lambda_i \leq 1$. Then, for every $t > 0$, it holds

$$\begin{aligned} p_x^T p_s(\theta) &= \Delta x^T\left(\Delta s + \frac{g(\theta)}{\theta}\right) \\ &= \Delta x^T\Delta s + \left(\frac{1}{2}\theta f''(y)[\Delta x, \Delta x, \Delta x]\right) \\ &\leq \Delta x^T\Delta s + \frac{3^{3/2}}{2}\beta\theta f'(y)[\Delta x, \Delta x]\ \|y^{-1}\Delta x\| \\ &\leq \Delta x^T\Delta s + \frac{3^{3/2}}{2}\beta\theta\left(tf'(y) + Y^{-2}\right)[\Delta x, \Delta x]\ \|y^{-1}\Delta x\|\frac{1}{t}. \end{aligned} \tag{38}$$

We will apply the relative Lipschitz condition with y and x to the self-concordant mapping

$$h(x) = tf(x) - x^{-1}.$$

Since $\Delta x^T\nabla f(x)\Delta x = \Delta x^T\Delta s$, the relative Lipschitz condition gives

$$\begin{aligned} \left(tf'(y) + Y^{-2}\right)[\Delta x, \Delta x] &\leq \left(1 + \frac{1}{(1-1/\sqrt{2})^2} - 1\right)\left(\Delta x^T\left(tf'(x) + X^{-2}\right)\Delta x\right) \\ &\leq 12\left(t\Delta x^T\Delta s + \|x^{-1}\Delta x\|^2\right). \end{aligned}$$

By the monotonicity of the mapping f and Lemma 3.1, we have

$$0 \leq \Delta x^T\Delta s \leq \frac{1}{4}\|p_v\|^2.$$

Also, the assumption guarantees

$$\begin{aligned} \frac{1}{v_{\min}^2}\left(\Delta x^T\Delta s + v_{\min}^2\|x^{-1}\Delta x\|^2\right) &= \frac{1}{v_{\min}^2}\left(p_x^T p_s + v_{\min}^2\|v^{-1}p_x\|^2\right) \\ &\leq \frac{1}{v_{\min}^2}\left(p_x^T p_s + \|p_x\|^2\right) \\ &\leq \frac{1}{v_{\min}^2}\|p_x + p_s\|^2 \\ &\leq \frac{v_{\max}^2}{v_{\min}^2}\|x^{-1}\Delta x + s^{-1}\Delta s\|^2 \\ &\leq \frac{1}{\rho^2}. \end{aligned}$$

It is easily seen that

$$\|y^{-1}\Delta x\| \leq \frac{1}{1-\theta\|x^{-1}\Delta x\|}\|x^{-1}\Delta x\| \leq 2\|x^{-1}\Delta x\| \leq \frac{2}{v_{\min}}\|p_x + p_s\| = \frac{2}{v_{\min}}\|p_v\|.$$

Substituting the latter results in (38) with $t = 1/v_{\min}^2$ gives

$$\begin{aligned} p_x^T p_s(\theta) &\leq \frac{1}{4}\|p_v\|^2 + \frac{3^{3/2}}{2}\beta\theta\frac{12}{\rho^2}\frac{2\|p_v\|}{v_{\min}}v_{\min}^2 \\ &\leq \frac{1}{4}\left(1+\theta\frac{250\beta v_{\min}}{\rho^2\|p_v\|}\right)\|p_v\|^2 \end{aligned}$$

which completes the proof. ∎

¿From the lemma we can derive the following corollary in the case of the primal–dual affine scaling algorithms considered in Section 4.

Corollary 5.6 *If Δx and Δs are determined with a primal–dual affine scaling algorithm as in Section 4, then*

$$p_x^T p_s(\theta) \leq \frac{1}{4}\left(1+\theta\frac{250\beta}{\rho^3}\right)\|p_v\|^2.$$

Proof: In the algorithm, the vector p_v is given by $-\frac{v^{2r+1}}{\|v^{2r}\|}$. The result immediately follows from the inequality

$$\frac{v_{\min}}{\|p_v\|} = \frac{v_{\min}\|v^{2r}\|}{\|v^{2r+1}\|} \leq \frac{v_{\min}\|v^{2r}\|}{v_{\max}\|v^{2r}\|} \leq \frac{1}{\rho}.$$

∎

The result of this corollary gives an alternative proof for a bound of the type as in Lemma 4.1(i). Unfortunately, our analysis of the primal–dual affine scaling algorithms is based on large neighborhoods. This implies that also a bound (cf. Lemma 4.1(ii)) on

$$\|p_x p_s(\theta)\|_\infty$$

and on $\|g(\theta)\|_\infty$ is required. We have not been able to derive such a bound using self–concordance and relative Lipschitz. However, to have more insights into the relationship between the self concordance and our condition, we will need to enforce more strict conditions on the mapping. Recall that Condition 3.2 is equivalently written as

$$\|D\left(f(x+\theta\Delta x) - f(x) - \theta\nabla f(x)\Delta x\right)\| \leq \gamma\theta\|D\nabla f(x)\Delta x\|. \tag{39}$$

Using Taylor's expansion we obtain

$$\|D\left(f(x+h) - f(x) - \nabla f(x)h\right)\| = \|D\left(\nabla f(y) - \nabla f(x)\right)h\|$$

where $y = x + \lambda h$ for some vector λ satisfying $\lambda_i < 1$, $\forall i$. For matrices A and B we mean by $A \preceq B$ that $B - A$ is positive semidefinite. It is not difficult to derive the following results from the above observations:

Theorem 5.7 *Let G be a closed convex domain in $\mathbb{R}^n$, with nonempty interior $Q := \text{int}(G)$. Let f be a single valued monotone operator $f : Q \rightarrow \mathbb{R}^n$, and let*

$$M = \nabla f(x) \text{ and } N = \nabla f(y) - \nabla f(x)$$

for each $x, y \in Q$. Then

(i) *if $-\left(\frac{1}{(1-r)^2} - 1\right) M \preceq N \preceq \left(\frac{1}{(1-r)^2} - 1\right) M$ for all $x \in Q$, $y \in Q$ such that $(y-x)^T M(y-x) \leq r^2$ then f satisfies the relative Lipschitz condition,*

(ii) *if $N^T D^2 N \preceq \theta^2 \gamma^2 M^T D^2 M$ for all $x \in Q$, $y := x + \theta \Delta x \in Q$ such that $\|x^{-1}(y-x)\| \leq \nu$ for every $\theta \in (0, \Theta]$, and $D = \text{diag}(d)$ with $d > 0$, then f satisfies Condition 3.2 with Θ.*

While the above theorem seems to be quite trivial, it may be somewhat practical since the conditions in the theorem can be more easily checked than the original conditions. In fact, we can find some examples to satisfy the relative Lipschitz condition and/or Condition 3.2 by using the above theorem.

Remark 5.8 If the Jacobian $\nabla f(x)$ is symmetric for every $x \in Q$, the assumption $N^T D^2 N \preceq \gamma^2 M^T D^2 M$ in (ii) of the above theorem can be replaced by the condition that there exist positive number γ and a nonsingular matrix P such that

(i) $N \preceq \gamma M$ and

(ii) $P^{-1} D(\sqrt{\gamma} M - N) DP$ and $P^{-1} D(\sqrt{\gamma} M + N) DP$ are diagonal matrices.

Example 5.9 [LCP]
Consider the linear complementarity problem with $f(x) = Mx + q$; here M is a symmetric positive semidefinite matrix and $q \in \mathbb{R}^n$. Then $\nabla f(x) = M$ and it is easy to see that all of the smoothness conditions are satisfied. Specifically, Condition 3.2 is satisfied with $\Theta = +\infty$ and $\pi = 0$.

Example 5.10 [Entropy function]
Let $u \in \mathbb{R}^n$ and $u > 0$ and let $\phi(x)$ be an entropy function of the form

$$\phi(x) = \sum_{i=1}^{n} x_i \log \left(\frac{x_i}{u_i} \right).$$

Let us define $f(x) = \nabla \phi(x)$, that is $f_i(x) = \log x_i - \log u_i + 1$ for all $i = 1, 2, \ldots, n$. Then f satisfies all of the smoothness conditions (c.f. Theorem 4.1 of [49]).

It should be noted that the above two examples also satisfy the conditions in Theorem 5.7.

References

[1] K.M. Anstreicher and R. Bosch. A new infinity–norm path following algorithm for linear programming. Working Paper, Department of Management Sciences, University of Iowa, Iowa City, USA, 1993.

[2] R.W. Cottle, J.S. Pang, and R.E. Stone. *The Linear Complementarity Problem, Computer Science and Scientific Computing*. Academic Press Inc., San Diego, USA, 1992.

[3] I.I. Dikin. Iterative solution of problems of linear and quadratic programming. *Doklady Akademii Nauk SSSR*, 174:747–748, 1967. (Translated in : *Soviet Mathematics Doklady*, 8:674–675, 1967).

[4] R.M. Freund and M.J. Todd. Barrier functions and interior–point algorithms for linear programming with zero–, one–, or two–sided bounds on the variables. *Mathematics of Operations Research*, 20:415–440, 1995.

[5] C.C. Gonzaga. Path following methods for linear programming. *SIAM Review*, 34:167–227, 1992.

[6] O. Güler. Path following and potential reduction algorithms for nonlinear monotone complementarity problems. Working Paper, Dept. of Management Science, University of Iowa, Iowa City, USA, 1991.

[7] O. Güler. Existence of interior points and interior paths in nonlinear monotone complementarity problems. *Mathematics of Operations Research*, 18:128–147, 1993.

[8] O. Güler. Generalized linear complementarity problems. *Mathematics of Operations Research*, 20:441–448, 1995.

[9] P.T. Harker and J.S. Pang. Finite–dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. *Mathematical Programming*, 48:161–220, 1990.

[10] D. den Hertog. *Interior Point Approach to Linear, Quadratic and Convex Programming, Algorithms and Complexity*. Kluwer Publishers, Dordrecht, The Netherlands, 1994.

[11] D. den Hertog, F. Jarre, C. Roos, and T. Terlaky. A sufficient condition for self–concordance, with application to some classes of structured convex programming. *Mathematical Programming*, 69:75–88, 1995.

[12] D. den Hertog, C. Roos, and T. Terlaky. On the classical logarithmic barrier function method for a class of smooth convex programming problems. *Journal of Optimization Theory and Applications*, 73:1–25, 1992.

[13] B. Jansen, C. Roos, and T. Terlaky. A family of polynomial affine scaling algorithms for positive semi–definite linear complementarity problems. Technical Report 93–112, Faculty of Technical Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands, 1993.

[14] B. Jansen, C. Roos, and T. Terlaky. A polynomial primal–dual Dikin-type algorithm for linear programming. Technical Report 93–36, Faculty of Technical Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands, 1993. (To appear in Mathematics of Operations Research).

[15] F. Jarre. *The method of analytic centers for smooth convex programs.* PhD thesis, Institut für Angewandte Mathematik und Statistik, Universität Würzburg, Würzburg, Germany, 1989.

[16] F. Jarre. Interior-point methods for convex programming. *Applied Mathematics & Optimization*, 26:287–311, 1992.

[17] F. Jarre. Interior-point methods via self-concordance or relative Lipschitz condition. *Optimization Methods and Software*, 5:75–104, 1995.

[18] M. Kojima, Y. Kurita, and S. Mizuno. Large-step interior point algorithms for linear complementarity problems. *SIAM Journal on Optimization*, 3:398–412, 1993.

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

[20] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise. *A unified approach to interior point algorithms for linear complementarity problems*, volume 538 of *Lecture Notes in Computer Science*. Springer Verlag, Berlin, Germany, 1991.

[21] M. Kojima, N. Megiddo, and Y. Ye. An interior point potential reduction algorithm for the linear complementarity problem. *Mathematical Programming*, 54:267–279, 1992.

[22] M. Kojima, S. Mizuno, and T. Noma. A new continuation method for complementarity problems with uniform P-functions. *Mathematical Programming*, 43:107–113, 1989.

[23] M. Kojima, S. Mizuno, and T. Noma. Limiting behavior of trajectories by a continuation method for monotone complementarity problems. *Mathematics of Operations Research*, 15:662–675, 1990.

[24] M. Kojima, S. Mizuno, and A. Yoshise. A polynomial-time algorithm for a class of linear complementarity problems. *Mathematical Programming*, 44:1–26, 1989.

[25] M. Kojima, S. Mizuno, and A. Yoshise. A primal–dual interior point algorithm for linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming : Interior Point and Related Methods*, pages 29–47. Springer Verlag, New York, 1989.

[26] M. Kojima, S. Mizuno, and A. Yoshise. An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems. *Mathematical Programming*, 50:331–342, 1991.

[27] M. Kojima, T. Noma, and A. Yoshise. Global convergence in infeasible-interior-point algorithms. *Mathematical Programming*, 65:43–72, 1994.

[28] K.O. Kortanek, F. Potra, and Y. Ye. On some efficient interior point methods for nonlinear convex programming. *Linear Algebra and Its Applications*, 152:169–189, 1991.

[29] K.O. Kortanek and J. Zhu. A polynomial barrier algorithm for linearly constrained convex programming problems. *Mathematics of Operations Research*, 18:116–127, 1993.

[30] P. Ling. A new proof of convergence for the new primal-dual affine scaling interior-point algorithm of Jansen, Roos and Terlaky. Technical Report SYS-C93-09, School of Information Systems, University of East-Anglia, Norwich, Great Britain, 1993.

[31] L. McLinden. The analogue of Moreau's proximation theorem, with applications to the nonlinear complementarity problem. *Pacific Journal of Mathematics*, 88:101–161, 1980.

[32] S. Mizuno. $\mathcal{O}(n^\rho L)$ iteration $\mathcal{O}(n^3 L)$ potential reduction algorithms for linear programming. *Linear Algebra and Its Applications*, 152:155–168, 1991.

[33] S. Mizuno and M.J. Todd. An $\mathcal{O}(n^3 L)$ adaptive path following algorithm for a linear complementarity problem. *Mathematical Programming*, 52:587–595, 1991.

[34] R.D.C. Monteiro and I. Adler. Interior path following primal–dual algorithms : Part I : Linear programming. *Mathematical Programming*, 44:27–41, 1989.

[35] R.D.C. Monteiro and I. Adler. Interior path following primal–dual algorithms : Part II : Convex quadratic programming. *Mathematical Programming*, 44:43–66, 1989.

[36] R.D.C. Monteiro and I. Adler. An extension of Karmarkar–type algorithm to a class of convex separable programming problems with global linear rate of convergence. *Mathematics of Operations Research*, 15:408–422, 1990.

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

[38] R.D.C. Monteiro, J.-S. Pang, and T. Wang. A positive algorithm for the nonlinear complementarity problem. *SIAM Journal on Optimization*, 5:129–148, 1995.

[39] Y. Nesterov. Long–step strategies in interior point potential reduction methods. Technical Report, University of Geneva, Geneva, Switzerland, 1993.

[40] Y. Nesterov and A.S. Nemirovsky. *Interior Point Polynomial Algorithms in Convex Programming*. SIAM Studies in Applied Mathematics, Vol. 13. SIAM, Philadelphia, USA, 1994.

[41] Y. Nesterov and M.J. Todd. Self–scaled barriers and interior–point methods for convex programming. Technical Report 1091, School of OR and IE, Cornell University, Ithaca, New York, USA, 1994. (To appear in Mathematics of Operations Research).

[42] J.S. Pang and S.A. Gabriel. NE/SQP: A robust algorithm for the nonlinear complementarity problem. *Mathematical Programming*, 60:295–337, 1993.

[43] F.A. Potra and Y. Ye. Interior point methods for nonlinear complementarity problems. Reports on Computational Mathematics 15, Dept. of Mathematics, The University of Iowa, Iowa City, USA, 1991. (To appear in Journal of Optimization Theory and Applications).

[44] J. Sun, J. Zhu, and G. Zhao. A predictor-corrector algorithm for a class of nonlinear saddle point problems. Technical Report, National University of Singapore, Singapore, 1994.

[45] H. Väliaho. P_* matrices are just sufficient. Research Report, Department of Mathematics, University of Helsinki, Helsinki, Finland, 1994. (To appear in Linear Algebra and Its Applications).

[46] S.J. Wright. A path-following infeasible-interior-point algorithm for linear complementarity problems. *Optimization Methods and Software*, 2:79–106, 1993.

[47] Y. Ye and K.M. Anstreicher. On quadratic and $O(\sqrt{n}L)$ convergence of predictor–corrector algorithms for LCP. *Mathematical Programming*, 62:537–551, 1993.

[48] Y. Ye and P.M. Pardalos. A class of linear complementarity problems solvable in polynomial time. *Linear Algebra and Its Applications*, 152:3–17, 1991.

[49] J. Zhu. A path following algorithm for a class of convex programming problems. *Zeitschrift für Operations Research*, 36:359–377, 1992.