Mathematical Programming 76 (1996) 95–115

Fast convergence of the simplified largest step path following algorithm

Clovis C. Gonzaga [a,*], J. Frédéric Bonnans [b]

[a] *Department of Mathematics, Federal University of Santa Catarina, Cx. Postal 5210, 88040-970 Florianópolis, SC, Brazil*

[b] *INRIA-Rocquencourt, B.P. 105, 78153 Rocquencourt, France*

Received 1 January 1995; revised manuscript received 1 November 1995

Abstract

Each master iteration of a simplified Newton algorithm for solving a system of equations starts by computing the Jacobian matrix and then uses this matrix in the computation of p Newton steps: the first of these steps is exact, and the other are called "simplified".

In this paper we apply this approach to a large step path following algorithm for monotone linear complementarity problems. The resulting method generates sequences of objective values (duality gaps) that converge to zero with Q-order $p+1$ in the number of master iterations, and with a complexity of $\mathrm{O}(\sqrt{n}L)$ iterations.

Keywords: Linear complementarity problem; Primal-dual interior-point algorithm; Convergence of algorithms; Simplified Newton method

1. Introduction

The monotone horizontal linear complementarity problem is stated as follows[1]: Find $(x,s) \in \mathbb{R}^{2n}$ such that

$$\text{(P)} \qquad \begin{aligned} xs &= 0, \\ Qx + Rs &= b, \\ x, s &\geqslant 0 \end{aligned}$$

[*] Corresponding author. Research done while visiting the Delft Technical University, and supported in part by CAPES – Brazil.

[1] The notation and the problem will be formally discussed ahead.

Published by Elsevier Science B.V.
PII S0025-5610(96)00035-4

where $b \in \mathbb{R}^n$, and $Q, R \in \mathbb{R}^{n\times n}$ are such that for any $u, v \in \mathbb{R}^n$,

$$\text{if } Qu + Rv = 0 \text{ then } u^{\mathrm{T}}v \geqslant 0.$$

This problem is a natural extension of linear and convex quadratic programming problems. If we add to the problem a strict complementarity hypothesis, then most of the primal–dual interior point algorithms developed for linear programming can be directly adapted to (P), preserving the same complexity and asymptotic convergence properties.

Presently the best convergence properties are shown for path following algorithms, which generate points in a neighborhood of the curve known as central path. The primal–dual central path was initially described by Megiddo [14], and the first algorithms to achieve the low complexity of $\mathrm{O}(\sqrt{n}L)$ iterations following this path are due to Kojima et al. [11] and Monteiro and Adler [20,21].

The evolution of these methods is surveyed in the papers by Gonzaga [5] and by Kojima et al. [9]. The complexity bound of $\mathrm{O}(\sqrt{n}L)$ iterations seems to be the lowest obtainable by the existing techniques, and it is common to all methods that follow the central path using an Euclidean proximity measure. The asymptotic convergence rates of these methods has been the object of much effort in the last few years, and is also the object of this paper.

The predictor-corrector algorithm introduced by Mizuno et al. [19] for linear programming was the first method for which a high convergence rate was proved. Ye et al. [29] and independently Mehrotra [15] proved Q-quadratic convergence for this algorithm, when considering the predictor and corrector steps as a single iteration. Ye and Anstreicher [28] proved the same rate for linear complementarity problems. But since the method requires two complete Newton steps per iteration, the actual rate of convergence is $\sqrt{2}$, and not quadratic.

This fact motivated a vivid search for an algorithm with genuine quadratic convergence. Ye [27] proposed an algorithm in which the need for corrector steps is less frequent as the sequence approaches the optimal face, thus achieving sub-quadratic convergence, i.e, convergence of order two but not quadratic. Very recently, Luo and Ye [12] showed how corrector steps can be restrained to a finite number, thus achieving genuine quadratic convergence. The same rate was also obtained by Gonzaga [3], for an algorithm that follows the central path with the largest possible pure Newton steps.

The largest step path following algorithm in a formulation based on the the proximity measure given by the primal–dual potential function was studied by McShane [13] for linear programming and for linear complementarity. He proves superlinear convergence under the hypothesis that the iterate sequence converges. Bonnans and Gonzaga [2] show that the sequence always converges, thus confirming the superlinear convergence. This algorithm will be the basis for the present paper, and will be thoroughly discussed ahead.

An advancement to predictor corrector algorithms that proved very efficient in practice was proposed by Mehrotra [16]: he computed the corrector step by a simplified Newton step, based on the Jacobian matrix already factored in the predictor step. Gonzaga and

Tapia [7] added a safeguard to this algorithm to guarantee the efficiency of the corrector steps when far from the optimal face, and proved that the quadratic rate of convergence is preserved.

In [17], Mehrotra describes a version of the predictor-corrector algorithm for linear programming in which each iteration takes one predictor step followed by r simplified corrections and a centering step. The algorithm follows the central path in a narrow neighborhood, and attains a Q-order of convergence of $r+2$. No complexity results are proved.

Similar asymptotic convergence results are obtained by Wright and Zhang [26] for linear complementarity problems using a different approach, based on a different characterization of proximity to the central path, and allowing infeasible iterates.

In the present paper we apply the simplified Newton approach to the feasible largest step path following algorithm, and show that it achieves the best possible results that can be expected from a simplified Newton algorithm with p steps per iteration: a Q-order of convergence of p+1 (the same as for the methods in the last paragraph), with a complexity of $\mathrm{O}(\sqrt{n}L)$ iterations, and a single matrix factorization per master iteration.

Conventions. Vectors may be denoted with sub- or superscripts. The scalars γ and μ will be denoted only with subscripts. Hence μ^j denotes μ to the power j.

Given a vector x, d, the corresponding upper-case symbol denotes as usual the diagonal matrix X, D defined by the vector. The symbol e will represent the vector of all ones, with dimension given by the context.

Given a matrix A, its null space and column range space are denoted respectively by $\mathcal{N}(A)$ and $\mathcal{R}(A)$. The projection matrix onto $\mathcal{N}(A)$ is P_A, and its complement is $\tilde{P}_A = I - P_A$.

We shall denote component-wise operations on vectors by the usual notations for real numbers. Thus, given two vectors u, v of the same dimension, uv, u/v, etc. will denote the vectors with components $u_i v_i$, u_i/v_i, etc. This notation is consistent as long as component-wise operations always have precedence in relation to matrix operations. Note that $uv \equiv Uv$ and if A is a matrix, then $Auv \equiv AUv$, but in general $Auv \neq (Au)v$.

We shall frequently use the $\mathrm{O}(\cdot)$ and $\Omega(\cdot)$ notation to express the relationship between functions. Our most common usage will be associated with a sequence $\{x^k\}$ of vectors and a sequence $\{\mu_k\}$ of positive real numbers. In this case $x^k = \mathrm{O}(\mu_k)$ means that there is a constant K (dependent on problem data) such that for every $k \in \mathbb{N}$, $\|x^k\| \leqslant K\mu_k$. Similarly, if $x^k > 0$, $x^k = \Omega(\mu_k)$ means that $(x^k)^{-1} = \mathrm{O}(1/\mu_k)$. Finally, $x^k \approx \mu_k$ means that $x^k = \mathrm{O}(\mu_k)$ and $x^k = \Omega(\mu_k)$.

2. The simplified largest step algorithm

Consider the monotone linear complementarity problem (P).

The feasible set for (P) and the set of interior solutions are respectively

$$F = \{(x,s) \in \mathbb{R}^{2n} \mid Qx + Rs = b,\ x,s \geqslant 0\},$$
$$F^0 = \{(x,s) \in F \mid x > 0,\ s > 0\}.$$

We say that respectively x or s is feasible if there exists s or x such that $(x,s) \in F$.

The set of optimal solutions and the set of strictly complementary optimal solutions are respectively

$$\mathcal{F} = \{(x,s) \in F \mid xs = 0\},$$
$$\mathcal{F}^0 = \{(x,s) \in \mathcal{F} \mid x + s > 0\}.$$

Centrality. The centrality measure is a map that associates with each $(x,s) \in F$ and $\mu \in \mathbb{R}$, $\mu > 0$, the value

$$\delta(x,s,\mu) = \left\| \frac{xs}{\mu} - e \right\|.$$

If $\delta(x,s,\mu) = 0$, then (x,s) is the *central point* associated with the parameter value μ, and the set of such points composes the well-known *central path* (see for instance Kojima et al. [9]).

Assumptions. In this paper we study only feasible problems, with the following assumptions:

Assumption 1. $F^0 \neq \emptyset$, and initial data $(x^0, s^0) \in F^0$, $\bar{\mu}_0 > 0$ are given such that $\delta(x^0, s^0, \bar{\mu}_0) < 0.5$.

Assumption 2. There exists a strictly complementary solution, i.e. $\mathcal{F}^0 \neq \emptyset$.

The strict complementarity hypothesis is discussed in Monteiro and Wright [23], where it is shown to be necessary for obtaining algorithms with fast convergence rates.

With these hypotheses the central path is a well defined differentiable curve, which ends at a strictly complementary solution (x^*, s^*).

The set $\mathcal{F}$ is a face of the polyhedral feasible set F, and $\mathcal{F}^0$ is its relative interior. This face is characterized by a partition $\{B, N\}$ of the set of indices $\{1, \dots, n\}$, such that if $(x,s) \in \mathcal{F}^0$ then $x_N = 0$, $s_B = 0$, $x_B > 0$, $s_N > 0$. The variables x_N, s_B are called *small variables*, and the variables x_B, s_N are the *large variables*. The analytic center of the optimal face (also called the central optimal solution) is characterized by $x_N^* = 0$, $s_B^* = 0$ and

$$(x_B^*, s_N^*) = \operatorname{argmin} \left\{ -\sum_{i \in B} \log x_i - \sum_{i \in N} \log s_i \mid (x,s) \in \mathcal{F}^0 \right\},$$

and it coincides with the end point of the central path (see [9]).

Notation. The symbol w will be used to denote triplets $w = (x, s, \mu)$, where $(x, s) \in F$ and $\mu > 0$. The notation will be consistent, e.g. $w^k = (x^k, s^k, \mu_k)$. We say that $w = (x, s, \mu)$ is feasible if $(x, s) \in F$ and $\mu \in [0, \bar{\mu}_0]$.

Our algorithm will work in an Euclidean norm neighborhood of the central path, defined by

$$\mathcal{N} = \{w = (x, s, \mu) \mid w \text{ is feasible, } \delta(w) \leqslant 0.5\}.$$

If $w = (x, s, \mu) \in \mathcal{N}$, then Monteiro and Tsuchiya [22] show that the large and small variables satisfy respectively

$$(x_B, s_N) \approx 1 \quad \text{and} \quad (x_N, s_B) \approx \mu. \tag{1}$$

They also show that the central path does not approach (x^*, s^*) tangentially, with the following result: if (x_μ, s_μ) is the central point associated with each $\mu > 0$, then

$$x_\mu - x^* = \mathrm{O}(\mu) \quad \text{and} \quad s_\mu - s^* = \mathrm{O}(\mu). \tag{2}$$

The path following algorithms start each iteration from data $w \in \mathcal{N}$, usually coming from a former iteration, and then take Newton steps for solving the system below, which corresponds to the search for a point in the central path:

$$x_\natural s_\natural = \gamma\mu e, \qquad Qx + Rs = b,$$

where $\gamma \in [0, 1]$.

A Newton step for solving this system computes a point

$$x_\natural := x + u, \qquad s_\natural := s + v, \tag{3}$$

where u, v are obtained by solving a linearization of the system above. The system to be solved is

$$\begin{aligned} su + xv &= -xs + \gamma\mu e, \\ Qu + Rv &= 0. \end{aligned} \tag{4}$$

A simplified Newton step does this linearization around a point $w_\flat$, possibly different from w. The reason for using simplified steps is that the same linearization can be used in several steps, with great computational savings. The system to be solved is

$$\begin{aligned} s_\flat u + x_\flat v &= -xs + \gamma\mu e, \\ Qu + Rv &= 0. \end{aligned} \tag{5}$$

This system can be rewritten as

$$\begin{aligned} s_\flat u + x_\flat v &= -(1 - \gamma)\, xs + \gamma\,(-xs + \mu e), \\ Qu + Rv &= 0. \end{aligned} \tag{6}$$

By superposition, we see that

$$(u, v) = \gamma(u_c, v_c) + (1 - \gamma)(u_a, v_a), \tag{7}$$

where (u_c, v_c) is the solution corresponding to $\gamma = 1$, called the *centering step*, and (u_a, v_a) is the solution corresponding to $\gamma = 0$, called the *affine-scaling step*.

Defining now

$$\begin{aligned} w_c &= (x_c, s_c, \mu_c) = (x + u_c, s + v_c, \mu), \\ w_a &= (x_a, s_a, \mu_a) = (x + u_a, s + v_a, 0), \end{aligned} \tag{8}$$

the result of the simplified Newton step will be

$$w_\sharp = \gamma w_c + (1 - \gamma) w_a. \tag{9}$$

Now we present the algorithm, beginning by a formal statement of the simplified Newton step procedure. This procedure computes the centering and affine-scaling steps from given $w_\flat, w$.

Procedure 2.1 (*Computation of centering and affine-scaling steps*).

Data: $w_\flat = (x_\flat, s_\flat, \mu_\flat)$, $w = (x, s, \mu)$ such that $w_\flat, w \in \mathcal{N}$, $\mu \leqslant \mu_\flat$.

Compute w_a and w_c as in (8) by solving the system below twice, for $\gamma = 0$ and $\gamma = 1$:

$$\begin{aligned} s_\flat u + x_\flat v &= -xs + \gamma \mu e, \\ Qu + Rv &= 0. \end{aligned}$$

Now we state the main algorithm. Each master iteration begins by computing an exact Newton step, i.e, by using the procedure above with $w_\flat = w$, followed by at most $p - 1$ simplified steps which keep $w_\flat$ constant.

In the beginning of the iteration there is a safeguard (to be extensively commented below,) which decides whether a centering step should be taken: if this is the case, then the centering step is accepted and the iteration is restarted.

Algorithm 2.2.

Data: $\epsilon > 0$, $w^0 = (x^0, s^0, \bar{\mu}_0) \in \mathcal{N}$, $p \in \mathbb{N}, p > 0$.

```
$k := 0$
repeat (master iteration)
    Exact step: Set $w := w^k, w_\flat := w^k$ and compute $w_c, w_a$ by Procedure 2.1.
    Safeguard: If $\delta(0.1 w_c + 0.9 w_a) \in [0.42, 1]$ then set $w := w_c, w_\flat := w_c$ and
    compute $w_c, w_a$ by Procedure 2.1.
    Steplength: Compute $\gamma \in (0,1)$ such that $\delta(\gamma w_c + (1-\gamma) w_a) = 0.5$, and set
    $w_\sharp := \gamma w_c + (1-\gamma) w_a$.
    for $i := 1$ to $p - 1$ do (simplified steps)
        Set $w := w_\sharp$ and compute $w_c, w_a$ by Procedure 2.1.
        If $\delta(w_c) \geqslant 0.5$ then exit loop.
        Steplength: Compute $\gamma \in (0,1)$ such that $\delta(\gamma w_c + (1-\gamma) w_a) = 0.5$, and
        set $w_\sharp := \gamma w_c + (1-\gamma) w_a$.
        end for
        $w^{k+1} := w_\sharp$.
        $k := k + 1$.
    until $\mu_k < \epsilon$.
```

Remarks. Each iteration of the algorithm computes either two or one exact Newton steps, depending on whether the safeguard is triggered or not, followed by a maximum of $p-1$ simplified steps. The simplified steps do not reset $w_\flat$, and therefore do not need to compute a factorization.

In the case $p = 1$, the algorithm essentially reduces to the largest step path following algorithm [3]. With $p = 1$ and no safeguard, it reduces to the algorithm described by McShane [13].

The safeguard computes the proximity of the point resulting from the Newton step with $\gamma = 0.1$. Any smaller fixed number can be used instead of 0.1. If this proximity is in a narrow interval, then the iteration is restarted from w_c, i.e, a centralization step is performed. Note that the safeguard can only be triggered when the step $\gamma = 0.1$ is acceptable, which only occurs in the final stages of the algorithm. We will explain below why it is reasonable to expect that the safeguard will be triggered only a very small number of times. We shall very soon comment further on the safeguard.

In each inner iteration (simplified step) the centering step is tested. If $\delta(w_c) \geqslant 0.5$, then it is inefficient, and cannot be used. In this case the remaining simplified steps are aborted.

The rest of the paper will be dedicated to proving the following theorem:

Theorem 2.3. *Assume that Assumptions* 1 *and* 2 *hold, and let* $L = \log_2(\bar{\mu}_0/\epsilon)$. *Then*

(i) *The algorithm finds a feasible point that satisfies* $x^{\mathrm{T}}s \leqslant \epsilon$ *in* $\mathrm{O}(\sqrt{n}L)$ *iterations.*

(ii) *There exists* $\bar{\mu} > 0$ *depending only on problem data such that the safeguard is activated at most once whenever* $\mu_k \leqslant \bar{\mu}$.

(iii) *For all* $k \in \mathbb{N}$, $\mu_{k+1} = \mathrm{O}((\mu_k)^{p+1})$, *and hence the sequence* μ_k *converges to* 0 *with Q-order* $p+1$.

The analysis will be done by proving the following facts:

- (i) is an immediate consequence of the polynomiality of the short step algorithm, proved by Kojima et al. [11], since the steplength here is always at least as large as the one in that reference.
- The safeguard detects the situation in which the Newton step is very efficient ($\gamma < 0.1$) but (x, s) is far from the central solution (x^*, s^*), and corrects this by a centering step. If this happens for μ_k small, then the iterates will stay forever near the central optimum. Consequently, the safeguard will not be activated again.
- If (x^k, s^k) is near the central optimum, then the exact Newton step is very efficient, producing $\gamma = \mathrm{O}(\mu)$. If besides this, $(x_\flat, s_\flat) - (x, s) = \mathrm{O}(\mu_\flat)$, then the simplified step is also efficient, producing $\gamma = \mathrm{O}(\mu_\flat)$. These conditions will be true at all iterations:

the first one because of the safeguard; the second one because large steps (steps with small γ) have little influence on the large variables.

3. Analysis of the simplified Newton step

In this section we shall use general results about the linear complementarity problem, fully described in [2]. The notation is greatly simplified if we use the following observation from that reference: some of the variable pairs (x_i, s_i) can be permuted for the analysis of the algorithm, in such a way that $x \leftarrow (x_B, s_N)$ and $s \leftarrow (x_N, s_B)$. Renaming the variables in this way (and rearranging the matrices in (P)) obviously does not affect the algorithms, but simplifies the treatment: it is the same as assuming that $N = \emptyset$, x is the vector of large variables, and s is the vector of small variables. So, for the rest of the paper we use the following assumption:

Assumption 3. Assume that x is the vector of large variables and s is the vector of small variables.

Remark. This change of variables can only be done in the analysis: it cannot be used by algorithms, since B and N are unknown. The Newton steps are obviously not affected by renaming the variables.

Notation. All entities related to the reference point $w_\flat$ will be indicated by the symbol '$\flat$'; the entities related to the result of a Newton step will be indicated by '$\sharp$'. We also define

$$\gamma_\flat = \mu / \mu_\flat.$$

With this permutation, the optimal face is characterized simply by

$$\mathcal{F} = \left\{ (x, s) \in \mathbb{R}^{2n} \mid s = 0, Qx = b, x \geqslant 0 \right\}. \tag{10}$$

We assume for the remainder of this section that $w_\flat, w \in \mathcal{N}$ are given, and study the Newton step from these data. The Newton equations remain unchanged:

$$\begin{aligned} s_\flat u + x_\flat v &= -xs + \gamma \mu e, \\ Qu + Rv &= 0. \end{aligned} \tag{11}$$

Scaled equations. Let us define the scaling vector

$$d_\flat := \sqrt{\mu_\flat x_\flat / s_\flat},$$

and scale the variables by

$$\bar{x} = x/d_\flat, \qquad \bar{u} = u/d_\flat, \qquad \bar{s} = d_\flat s_\flat / \mu_\flat, \qquad \bar{v} = d_\flat v / \mu_\flat. \tag{12}$$

Note that this scaling differs from the usual one by the presence of $\mu_\flat$ in the expressions. This makes sense because x and s are now the large and small variables.

This scaling transfers the given $x_\flat$ and $s_\flat$ to the same vector

$$\phi_\flat = x_\flat / d_\flat = \frac{d_\flat s_\flat}{\mu_\flat} = \sqrt{\frac{x_\flat s_\flat}{\mu_\flat}}. \tag{13}$$

We define similarly $\phi = \sqrt{xs/\mu}$. The proximity measure is given by

$$\delta(x, s, \mu) = \left\| \phi^2 - e \right\|. \tag{14}$$

Substituting the scaled variables into (11), we obtain

$$\begin{aligned} \mu_\flat \phi_\flat \bar{u} + \mu_\flat \phi_\flat \bar{v} &= -\mu_\flat \bar{x}\bar{s} + \gamma\mu e, \\ QD_\flat \bar{u} + \mu_\flat R D_\flat^{-1} \bar{v} &= 0. \end{aligned}$$

Dividing the first equation by $\mu_\flat \phi_\flat$, we get

$$\begin{aligned} \bar{u} + \bar{v} &= -\phi_\flat^{-1} \bar{x}\bar{s} + \gamma\gamma_\flat \phi_\flat^{-1}, \\ QD_\flat \bar{u} &= -\mu_\flat R D_\flat^{-1} \bar{v}, \end{aligned} \tag{15}$$

where $\gamma_\flat := \mu/\mu_\flat$.

We note that $\bar{u}^{\mathrm{T}}\bar{v} \geqslant 0$ whenever $\bar{u}, \bar{v}$ satisfy the second relation in (15).

Sizes. The following size limitations hold: if $w_\flat, w \in \mathcal{N}$, then

$$\begin{array}{llll} x \approx 1, & s \approx \mu, & \phi \approx 1, & d \approx 1; \\ x_\flat \approx 1, & s_\flat \approx \mu_\flat, & \phi_\flat \approx 1, & d_\flat \approx 1; \\ \bar{x} \approx 1, & \bar{s} \approx \gamma_\flat, & \bar{v} = \mathrm{O}(\gamma_\flat). & \end{array} \tag{16}$$

That $x \approx 1$ follows from (1). That $\phi \approx 1$ is a consequence of the definition of $\mathcal{N}$, as $\left\|\phi^2 - e\right\| \leqslant 0.5$. It follows that $s = \mu\phi^2/x \approx \mu$, and hence $d = \sqrt{\mu x/s} \approx 1$. Similar relations hold for $x_\flat$, $x_\flat$, $\phi_\flat$, $d_\flat$ and $\bar{x}$. The size of $\bar{s}$ stems from $\bar{s} = d_\flat s/\mu_\flat = \gamma_\flat d_\flat \mathrm{O}(\mu)/\mu$. Finally, from $\bar{u}^{\mathrm{T}}\bar{v} = u^{\mathrm{T}}v \geqslant 0$, we deduce that $\bar{u}$ and $\bar{v}$ are of the order of the right hand side of (15), i.e. of order $\gamma_\flat$.

Solution of the scaled equations. To derive the most useful properties of the scaled equations, we shall use Lemma 5.3 of reference [2], which proves the following result: Consider a system with the format

$$\begin{aligned} u + v &= f, \\ Au &= g, \end{aligned}$$

with $v \perp \mathcal{N}(A)$. Then u is the orthogonal projection of f into the affine space given by $Au = g$.

If $g = \mathrm{O}(\mu)$, then the lemma proves that

$$u = P_A f + \mathrm{O}(\mu).$$

If follows that $v = f - P_A f + \mathrm{O}(\mu)$, or

$$v = \tilde{P}_A f + \mathrm{O}(\mu).$$

Now we apply this result to the scaled system (15), for the case in which x is near $x_\flat$.

Lemma 3.1. *Assume that $w, w_\flat \in \mathcal{N}$ are such that $\mu = \gamma_\flat \mu_\flat \leqslant \mu_\flat$, and $x - x_\flat = \mathrm{O}(\mu_\flat)$. Then the solution of the simplified Newton step* (11) *gives*

$$u = \gamma\gamma_\flat d_\flat P_\flat \phi_\flat^{-1} + \mathrm{O}(\mu),$$
$$v = -s + \gamma\gamma_\flat \mu_\flat d_\flat^{-1} \tilde{P}_\flat \phi_\flat^{-1} + \mathrm{O}(\mu\mu_\flat).$$

Proof. Using the assumption that $x = x_\flat + \mathrm{O}(\mu_\flat)$ and remembering that $\bar{x}_\flat = \phi_\flat$ and $d_\flat \approx 1$, we deduce that $\bar{x} = \phi_\flat + \mathrm{O}(\mu_\flat)$, and it follows from (16) that

$$\phi_\flat^{-1} \bar{x} = e + \mathrm{O}(\mu_\flat).$$

It also follows from (16) that $\bar{v} = \mathrm{O}(\gamma_\flat)$. Applying these two results to (15),

$$\bar{u} + \bar{v} = -\bar{s} + \mathrm{O}(\gamma_\flat \mu_\flat) + \gamma\gamma_\flat \phi_\flat^{-1},$$
$$Q D_\flat \bar{u} = \mathrm{O}(\gamma_\flat \mu_\flat).$$

Following [2], we know that with the monotonicity condition, any dual feasible direction $\bar{v}$ for the scaled LCP problem (15) satisfies

$$\bar{v} \perp \mathcal{N}(Q D_\flat).$$

Since $d_\flat \approx 1$ and $\mu = \gamma_\flat \mu_\flat$, it follows from the result on projections described above that

$$\bar{u} = -P_\flat \bar{s} + \gamma\gamma_\flat P_\flat \phi_\flat^{-1} + \mathrm{O}(\mu).$$

But s is the vector of small variables, and the following interesting fact holds: since any optimal solution (x^*, s^*) satisfies $s^* = 0$, any feasible vector of small variables s is also a feasible direction $s - s^*$. In particular for the scaled problem, $\bar{s}$ is a dual feasible direction, and thus

$$\bar{s} \perp Q D_\flat, \qquad P_\flat \bar{s} = 0, \qquad \tilde{P}_\flat \bar{s} = \bar{s}.$$

It follows that

$$\bar{u} = \gamma\gamma_\flat P_\flat \phi_\flat^{-1} + \mathrm{O}(\mu),$$
$$\bar{v} = \bar{s} + \gamma\gamma_\flat \tilde{P}_\flat \phi_\flat^{-1} + \mathrm{O}(\mu).$$

The results for the original coordinates are obtained by multiplying the first and second equations respectively by $d_\flat \approx 1$ and $\mu_\flat d_\flat^{-1} \approx \mu_\flat$, completing the proof. □

This lemma shows a very interesting fact: if both $w_\flat$ and w are in the region $\mathcal{N}$ and the large variables have a small variation between $w_\flat$ and w, then the result of the simplified Newton step suffers very little influence from the actual values of the small variables s. In fact, the only contribution of the data from w in the expressions in the lemma is the value of $\gamma_\flat$: the influence of x, s are hidden in the residuals $O(\mu)$ and $O(\mu\mu_\flat)$ in the formulae. This will become clearer in the lemma below.

Let $w_c = (x_c, s_c, \mu)$ and $w_a = (x_a, s_a, 0)$ be defined as usual, as the result of the simplified Newton step from w, with $(x_c, s_c) = (x, s) + (u_c, v_c)$ and $(x_a, s_a) = (x, s) + (u_a, v_a)$. Define the exact Newton centering and affine-scaling steps from $w_\flat$ respectively as $x_c^\flat = (x_c^\flat, s_c^\flat, \mu_\flat)$ and $w_a^\flat = (x_a^\flat, s_a^\flat, 0)$, with $(x_c^\flat, s_c^\flat) = (x_\flat, s_\flat) + (u_c^\flat, v_c^\flat)$.

Lemma 3.2. *Let $w_\flat$ and w be as in Lemma* 3.1. *Then*

$$u_c = \gamma_\flat u_c^\flat + O(\mu_\flat), \qquad s_c = \gamma_\flat(s_c^\flat + O(\mu_\flat^2)).$$

and

$$u_a = O(\mu), \qquad s_a = \gamma_\flat O(\mu_\flat^2)$$

Proof. The centering step from $w_\flat$ is obtained from Lemma 3.1 with $w = w_\flat$, $\gamma = \gamma_\flat = 1$, and gives with $s_c^\flat = s_\flat + v_c^\flat$

$$u_c^\flat = d_\flat P_\flat \phi_\flat^{-1} + O(\mu_\flat),$$
$$s_c^\flat = \mu_\flat d_\flat^{-1} \tilde{P}_\flat \phi_\flat^{-1} + O(\mu_\flat^2).$$

The simplified centering step from w is obtained with $\gamma = 1$:

$$u_c = \gamma_\flat d_\flat P_\flat \phi_\flat^{-1} + O(\mu_\flat),$$
$$s_c = \gamma_\flat \mu_\flat d_\flat^{-1} \tilde{P}_\flat \phi_\flat^{-1} + \gamma_\flat O(\mu_\flat^2).$$

The results for the centering step follow immediately from the comparison of these two systems. The results for u_a and s_a follow from Lemma 3.1 with $\gamma = 0$, completing the proof. □

This is then the behaviour of a centering step when $x - x_\flat = O(\mu_\flat)$ and μ is small: the large variables approach the center by a step u_c approximately equal to $\gamma_\flat u_c^\flat$. Noting that $u_c^\flat$ approximates the exact centering step from x, we see that the simplified centering step looses efficiency. If $\gamma_\flat$ is small (we will show that $\gamma_\flat = O(\mu_\flat)$), then we conclude that the simplified centering step does not affect the large variables significantly.

The effect of simplified centering on the small variables for small $\mu_\flat$ independs on the actual value of s, and results in s_c proportional to $s_c^\flat$. If the central path approaches a straight line, than the simplified centering step becomes as efficient as the exact centering step from w.

4. Proximity measures

In this section we describe some results relating proximity measures and Euclidean distances between points.

We start by describing the effect of an exact centering step on the proximity measure. This is a well known result, but we shall specialize it for our case. We use the following lemma due to Mizuno [18]:

Lemma 4.1. *Let $y, z \in \mathbb{R}^n$ be such that $y^{\mathrm{T}} z \geqslant 0$. Then*

$$\|yz\| \leqslant \|y+z\|^2 .$$

Lemma 4.2. *Consider $w \in \mathcal{N}$ and let w_c be the result of an exact centering step from w. Then*

$$\delta(w_c) \leqslant \frac{1}{\sqrt{8}} \frac{\delta^2(w)}{1-\delta(w)}.$$

Proof. The exact centering step corresponds to $w_\natural = w$, $\gamma = \gamma_\flat = 1$. Using scaled variables, $\bar{x} = \bar{s} = \phi$, and the first equation of (15) becomes

$$\bar{u} + \bar{v} = \phi^{-1}(\phi^2 - e). \tag{17}$$

The proximity at w is given by $\delta(w) = \left\|\phi^2 - e\right\|$ by (14). The proximity after centering is $\delta(w_c) = \|(\phi+\bar{u})(\phi+\bar{v}) - e\|$. Developing this and using (17), we get $\delta(w_c) = \|\bar{u}\bar{v}\|$. Using Lemma 4.1,

$$\delta(w_c) \leqslant \frac{1}{\sqrt{8}} \|\bar{u}+\bar{v}\|^2 .$$

From (17), $\|\bar{u}+\bar{v}\|^2 = \left\|\phi^{-1}(\phi^2 - e)\right\|^2 \leqslant \left\|\phi^{-1}\right\|_\infty^2 \delta^2(w)$. As $\left\|\phi^2 - e\right\| = \delta(w)$, $\phi_i^2 \geqslant 1 - \delta(w)$ for $i = 1, \ldots, n$. It follows that $\left\|\phi^{-1}\right\|_\infty^2 \leqslant 1/(1-\delta(w))$. Putting together these results,

$$\delta(w_c) \leqslant \frac{1}{\sqrt{8}} \frac{\delta^2(w)}{1-\delta(w)},$$

completing the proof. □

The primal scaled distance from x to x_μ

Remark. In this paragraph we do not assume that x and s were reordered according to the optimal partition. We are arbitrarily calling x 'primal variables', and this abuse of language will be often used ahead, but all the results in this paragraph are true if we switch x and s.

Consider problem (P), and denote the central point associated with each $\mu > 0$ by $w_\mu = (x_\mu, s_\mu, \mu)$.

For any feasible (x, s) we define the scaled distance from x to x_μ by

$$\operatorname{dist}(x,\mu) = \left\| \frac{x - x_\mu}{x_\mu} \right\|. \tag{18}$$

As we remarked above, $\operatorname{dist}(s,\mu)$ is defined similarly.

By definition, $x_\mu s_\mu = \mu e$. Substituting this into (18), we obtain

$$\operatorname{dist}(x,\mu) = \left\| \frac{x s_\mu}{\mu} - e \right\|. \tag{19}$$

The scaled Euclidean distance between x and x_μ has been studied by Renegar and Shub [25] for LP problems, using scaling about x rather than about x_μ. Our result below is valid for the monotone LCP, and is sharper than theirs when applied to linear programming.

Lemma 4.3. *Consider a feasible $w = (x, s, \mu)$ such that $\delta(w) \leqslant \sqrt{2}/2$. Then*

$$\operatorname{dist}(x,\mu) \leqslant \sqrt{2}\left(1 - \sqrt{1 - \sqrt{2}\delta(w)}\right).$$

Proof. Without loss of generality, assume that $x_\mu = e$ and $s_\mu = \mu e$. This can be achieved by a change of scale, since $x_\mu s_\mu = \mu e$. Denote

$$x = e + dx, \qquad s = \mu(e + ds), \qquad \delta = \delta(w).$$

The primal scaled distance is given by

$$\operatorname{dist}(x,\mu) = \|dx\| \leqslant \|dx + ds\|,$$

because $dx^{\mathrm{T}} ds \geqslant 0$. The primal–dual proximity is given by

$$\delta = \left\| \frac{xs}{\mu} - e \right\| = \|dx + ds + dx ds\|.$$

Using Lemma 4.1, $\|dx ds\| \leqslant \|dx + ds\|^2 / \sqrt{8}$, and hence

$$\delta \geqslant \|dx + ds\| - \|dx + ds\|^2 / \sqrt{8}.$$

Denoting $z = \|dx + ds\|$, we have the inequality

$$z^2 - \sqrt{8} z + \sqrt{8}\delta \geqslant 0$$

Solving the equation, the roots are $z = \sqrt{2}\left(1 \pm \sqrt{1 - \sqrt{2}\delta}\right)$. Since $\delta \leqslant \sqrt{2}/2$, the roots are real. The inequality is satisfied for

$$z = \sqrt{2}\left(1 - \sqrt{1 - \sqrt{2}\delta}\right) \leqslant \sqrt{2} \quad \text{or} \quad z = \sqrt{2}\left(1 + \sqrt{1 - \sqrt{2}\delta}\right) > \sqrt{2}. \tag{20}$$

Since $\operatorname{dist}(x,\mu) = \|dx\| \leqslant z$, all we need is to prove that the second case cannot happen, i.e. that whenever $\delta(w) \leqslant \sqrt{2}/2$, we have $\|dx + ds\| \leqslant \sqrt{2}$. Assume by contradiction that for some feasible $\tilde{w}$, $\delta(\tilde{w}) \leqslant \sqrt{2}/2$ and $\|\tilde{d}x + \tilde{d}s\| > \sqrt{2}$.

The map $x, s \in F^0 \mapsto \delta^2(x, s, \tilde{\mu})$ is smooth and has a unique minimizer at $(x_{\tilde{\mu}}, s_{\tilde{\mu}})$. Hence there exists a continuous curve $\alpha \mapsto (x(\alpha), s(\alpha))$ in F^0 that joins $(\tilde{x}, \tilde{s})$ and $(x_{\tilde{\mu}}, s_{\tilde{\mu}})$, along which $\delta(x(\alpha), s(\alpha), \tilde{\mu})$ decreases strictly. Since $\|\tilde{d}x + \tilde{d}s\| > \sqrt{2}$ and $\|dx_{\tilde{\mu}} + ds_{\tilde{\mu}}\| = 0$, there must be a point $(\hat{x}, \hat{s})$ on the curve such that $\delta(\hat{x}, \hat{s}, \tilde{\mu}) < \sqrt{2}/2$ and $z = \|\hat{d}x + \hat{d}s\| = \sqrt{2}$. This contradicts (20), completing the proof. □

The following lemma does not make much sense at this point, but it will soon be very useful. The lemma relates the primal distance given by $\|xs_\mu/\mu - e\|$ to an approximation $\|xs_c/\mu - e\|$ which will often appear.

Lemma 4.4. *Consider $w \in \mathcal{N}$ and let w_c be the result of an exact Newton centering step from w. Then*

$$\left\| \frac{xs_c}{\mu} - e \right\| \leqslant \operatorname{dist}(s_c, \mu) + (1 + \operatorname{dist}(s_c, \mu)) \operatorname{dist}(x, \mu).$$

Proof. As in the proof of the previous lemma, let $x_\mu = e$, $s_\mu = \mu e$, $x = e + dx$ and $s_c = \mu(e + ds_c)$. Then

$$\left\| \frac{xs_c}{\mu} - e \right\| = \|dx + ds_c + dx ds_c\| \leqslant \|ds_c\| + (1 + \|ds_c\|) \|dx\| .$$

The result follows from $\operatorname{dist}(x, \mu) = \|dx\|$ and $\operatorname{dist}(s_c, \mu) = \|ds_c\|$, completing the proof. □

Numerical values. The useful features of the lemmas in this section can be summarized by applying them in sequence to compute some numerical bounds. The facts below do this. Their proofs are trivial.

The first fact applies to any $w \in \mathcal{N}$.

Fact 4.5. *If $\delta(w) \leqslant 0.5$ then*

$$\operatorname{dist}(x, \mu) \leqslant 0.65, \qquad \delta(w_c) \leqslant 0.177,$$
$$\operatorname{dist}(s_c, \mu) \leqslant 0.1895, \qquad \|xs_c/\mu - e\| \leqslant 0.97.$$

The second one applies to a point w resulting from an exact centering (note that from the fact above such points have $\delta(x) \leqslant 0.177$).

Fact 4.6. *If $\delta(w) \leqslant 0.177$ then*

$$\operatorname{dist}(x, \mu) \leqslant 0.1897, \qquad \delta(w_c) \leqslant 0.014,$$
$$\operatorname{dist}(s_c, \mu) \leqslant 0.014, \qquad \|xs_c/\mu - e\| \leqslant 0.206.$$

The last fact applies to a point in $\mathcal{N}$ for which we know that the primal (large) variables are near their center.

Fact 4.7. *If $\delta(w) \leqslant 0.5$ and $\operatorname{dist}(x, \mu) \leqslant 0.2$ then $\|xs_c/\mu - e\| \leqslant 0.206$.*

Primal distance to the central optimum. Now assume that x is the vector of large variables. The central path approaches the central optimum $(x^*, s^*) = (x^*, 0)$ when μ decreases, and we extend (18) by defining

$$\operatorname{dist}(x, 0) = \left\| \frac{x - x^*}{x^*} \right\|. \tag{21}$$

Using (2),

$$\operatorname{dist}(x, 0) = \left\| \frac{x - x_\mu + \mathrm{O}(\mu)}{x_\mu + \mathrm{O}(\mu)} \right\|,$$

and since $x_\mu \approx 1$,

$$\operatorname{dist}(x, 0) = \operatorname{dist}(x, \mu) + \mathrm{O}(\mu). \tag{22}$$

5. Acceptation of very large steps

In this section we prove Theorem 2.3. We start by deriving properties of the proximity at the point $w_\natural$ resulting from an exact or simplified Newton step from $w \in \mathcal{N}$.

Lemma 5.1. *Consider $w_\flat, w \in \mathcal{N}$, $\gamma \in (0,1]$, $\gamma_\flat = \mu/\mu_\flat \in (0,1]$ and the result $w_\natural$ of the simplified Newton step from w. Then*

$$\frac{x_\natural s_\natural}{\gamma\mu} - e = \gamma\left(\frac{x_c s_c}{\mu} - e\right) + (1-\gamma)\left(\frac{x s_c}{\mu} - e\right) + \frac{\mathrm{O}(\mu_\flat)}{\gamma}. \tag{23}$$

If $x - x_\flat = \mathrm{O}(\mu_\flat)$ and $\gamma_\flat = \mathrm{O}(\mu_\flat)$ then

$$\frac{x_\natural s_\natural}{\gamma\mu} - e = \frac{x s_c}{\mu} - e + \frac{\mathrm{O}(\mu_\flat)}{\gamma}. \tag{24}$$

Proof. Using Lemma 3.2, $x_a = x + \mathrm{O}(\mu_\flat)$, $s_a = \mathrm{O}(\mu\mu_\flat)$. It follows that

$$\begin{aligned} x_\natural &= \gamma x_c + (1-\gamma) x_a = \gamma x_c + (1-\gamma) x + \mathrm{O}(\mu_\flat) = \mathrm{O}(1), \\ s_\natural &= \gamma s_c + (1-\gamma) s_a = \gamma s_c + \mathrm{O}(\mu\mu_\flat). \end{aligned}$$

Multiplying these expressions and remembering that $s_c = \mathrm{O}(\mu)$,

$$\frac{x_\natural s_\natural}{\gamma\mu} = \gamma \frac{x_c s_c}{\mu} + (1-\gamma)\frac{x s_c}{\mu} + \frac{\mathrm{O}(\mu_\flat)}{\gamma}. \tag{25}$$

The first result follows by subtracting $e = \gamma e + (1-\gamma) e$.

If $x = x_\flat + \mathrm{O}(\mu_\flat)$ and $\gamma_\flat = \mathrm{O}(\mu_\flat)$, then by Lemma 3.2 $u_c = \mathrm{O}(\mu_\flat)$ and $s_c = \mathrm{O}(\mu_\flat)$. From (25),

$$\frac{x_\natural s_\natural}{\gamma\mu} = \frac{x s_c}{\mu} + \frac{\mathrm{O}(\mu_\flat)}{\gamma} + \gamma \frac{s_c \mathrm{O}(\mu_\flat)}{\mu}.$$

The last term equals $\gamma\gamma_\flat \mathrm{O}(\mu_\flat^2)/\mu_\flat = \mathrm{O}(\mu_\flat)$, completing the proof. □

The expression in the lemma will appear several times ahead. The safeguard in particular computes $\delta(w_\sharp)$ for $\gamma = 0.1$ and $w = w_\flat$, an exact Newton step. Let us give numerical values for this case. The values follow directly from the application of Facts 4.5 to 4.7 and (22) to the norm of the expression (23),

$$\delta(w_\sharp) \leqslant 0.9\,\delta(w_c) + 0.1\left\|\frac{x}{\mu} - e\right\| + \mathrm{O}(\mu_\flat).$$

Fact 5.2. *Assume that $w = w_\flat$, $\gamma = 0.1$. Then the result $\tilde{w}$ of the (exact) Newton step satisfies:*

(i) *If $\delta(w) \leqslant 0.5$ then $\delta(\tilde{w}) \leqslant 0.89 + \mathrm{O}(\mu_\flat)$.*

(ii) *If $\delta(w) \leqslant 0.177$ then $\delta(\tilde{w}) \leqslant 0.21 + \mathrm{O}(\mu_\flat)$.*

(iii) *If $\delta(w) \leqslant 0.5$ and $\mathrm{dist}(x, 0) \leqslant 0.2$ then $\delta(\tilde{w}) \leqslant 0.403 + \mathrm{O}(\mu_\flat)$.*

The exact Newton step and safeguard. Each iteration of Algorithm 2.2 starts by an exact Newton step and an application of the safeguard, and then proceeds to the simplified steps. We shall begin the study of the algorithm by describing the effect of the exact step and safeguard on a generic point w.

Assume then that a point $w \in \mathcal{N}$ is given. We shall study the following procedures extracted from Algorithm 2.2:

Exact step: Set $w_\flat := w$ and compute w_c, w_a by Procedure 2.1.

Safeguard: If $\delta(0.1w_c + 0.9w_a) \in [0.42, 1]$ then set $w := w_c$, $w_\flat := w_c$ and compute w_c, w_a by Procedure 2.1.

Steplength: Compute $\gamma \in (0, 1)$ such that $\delta(\gamma w_c + (1-\gamma)w_a) = 0.5$, and set $w_\sharp := \gamma w_c + (1-\gamma)w_a$.

Remark. In the next lemmas we shall introduce several bounds for μ, e.g. $\bar{\mu}_1, \tilde{\mu}_j$. These numbers are constants that depend only on the problem data, and not on the initial point.

Lemma 5.3. *Consider the procedure above applied to $w \in \mathcal{N}$. There exists $\bar{\mu}_1 \leqslant \bar{\mu}_0$ such that for $\mu \leqslant \bar{\mu}_1$,*

(a) *At the safeguard step, $\delta(0.1w_c + 0.9w_a) < 1$.*

(b) *After the safeguard (whether or not it was triggered), $\delta(0.1w_c + 0.9w_a) < 0.42$ and $\|xs_c/\mu - e\| \leqslant 0.49 + \mathrm{O}(\mu)$.*

(c) *At the end of the computation, $x_\sharp - x = \mathrm{O}(\mu)$ and $\gamma = \mathrm{O}(\mu)$.*

Proof. Let $\tilde{w} = 0.1w_c + 0.9w_a$.

(a) follows from Fact 5.2(i) for μ sufficiently small (say $\mu \leqslant \tilde{\mu}_1$).

If the safeguard is not triggered, then $\delta(\tilde{w}) < 0.42$ by construction; otherwise a centering step produces $\delta(w) \leqslant 0.177$ by Fact 4.5 and then $\delta(\tilde{w}) \leqslant 0.21 + \mathrm{O}(\mu)$

by Fact 5.3. In any case, $\delta(\tilde{w}) < 0.42$ for μ sufficiently small (say $\mu \leqslant \tilde{\mu}_2 \leqslant \tilde{\mu}_1$). Applying this to (23) for $\mu \leqslant \tilde{\mu}_2$,

$$0.42 \geqslant -0.1\,\delta(w_c) + 0.9\left\|\frac{xs_c}{\mu} - e\right\| + \mathrm{O}(\mu).$$

Since $\delta(w_c) \leqslant 0.177$ by Fact 4.5, it follows that $\|xs_c/\mu - e\| \leqslant 0.487 + \mathrm{O}(\mu)$, proving (b). Now we prove (c). At the steplength computation, using (b),

$$\begin{aligned} 0.5 = \delta(w_\sharp) &\leqslant \gamma\delta w_c + (1-\gamma)\left\|\frac{xs_c}{\mu} - e\right\| + \frac{\mathrm{O}(\mu)}{\gamma} \\ &\leqslant 0.177\gamma + 0.49(1-\gamma) + \mathrm{O}(\mu)/\gamma \\ &\leqslant 0.49 + \mathrm{O}(\mu)/\gamma. \end{aligned}$$

This implies that $\gamma = \mathrm{O}(\mu)$, completing the proof. □

The simplified Newton steps. Now we extend these results to the simplified Newton steps. Since the safeguard is already well understood, we shall study the behaviour of the algorithmic step at a generic point w satisfying the conditions (b) in the last lemma, i.e. a point as the Algorithm 2.2 would generate after the safeguard. To do this, we extract from the algorithm the internal iteration, composed of one exact Newton step followed by $p-1$ simplified steps:

Algorithm 5.4.

```
Data: $w_\flat = w^0 \in \mathcal{N}$, $p \geqslant 1$.
    for $j := 0$ to $p - 1$ do
        Set $w := w^j$, $\gamma_\flat = \mu/\mu_\flat$ and compute $w_c, w_a$ by Procedure 2.1.
        If $\delta(w_c) \geqslant 0.5$ then exit loop.
        Steplength: Compute $\gamma \in (0,1)$ such that $\delta(\gamma w_c + (1-\gamma)w_a) = 0.5$, and set
        $w^{j+1} = w_\sharp := \gamma w_c + (1-\gamma)w_a$.
    end for
```

Lemma 5.5. *Consider an application of Algorithm* 5.4 *to a point* $w_\flat = w^0 \in \mathcal{N}$ *satisfying the conditions in Lemma* 4.3(b). *There exists* $\bar{\mu}_2 \leqslant \bar{\mu}_1$ *such that for* $\mu_\flat \leqslant \bar{\mu}_2$, $j = 0, 1, \ldots, p-1$

(i) $\delta(w_c^j) < 0.5$, *and hence the 'exit loop' is not triggered.*
(ii) $\mu_{j+1} = \mathrm{O}(\mu_\flat^{j+2})$, *or equivalently,* $\gamma_0\gamma_1\ldots\gamma_j = \mathrm{O}(\mu_\flat^{j+1})$. [2]
(iii) $x^{j+1} = x_\flat + \mathrm{O}(\mu_\flat)$.

Proof. For $j = 0$ we have the exact Newton step studied in Lemma 4.3. The results are immediate for $\mu_\flat \leqslant \tilde{\mu}_0 = \bar{\mu}_1$, since by that lemma $\delta(w_c^0) \leqslant 0.177$, $x^1 = x^0 + \mathrm{O}(\mu_\flat)$, $\gamma_0 = \mathrm{O}(\mu_\flat)$.

[2] Remember that a superscript on a scalar always means power.

Consider now $j \in \{1, \dots, p-1\}$ and assume that the results are true for $i \in \{0, 1, \dots, j-1\}$ for $\mu_\flat \leqslant \tilde{\mu}_{j-1}$. We shall prove them for iteration j.

Note initially that $x^j = x_\flat + \mathrm{O}(\mu_\flat)$ because (ii) holds at the former iterations, and that $\gamma_\flat = \gamma_0\gamma_1 \dots \gamma_{j-1} = \mathrm{O}(\mu_\flat^j)$ by (iii). Also, by Lemma 3.2, $x_c^j = x^j + \gamma_\flat u_c^\flat + \mathrm{O}(\mu_\flat) = x^j + \mathrm{O}(\mu_\flat)$ and $s_c^j = \gamma_\flat(s_c^\flat + \mathrm{O}(\mu_\flat^2))$. Since $s_c^\flat \approx \mu_\flat$, $s_c^j = \gamma_\flat \mathrm{O}(\mu_\flat)$. Using these results we can calculate the following:

$$\frac{x^j s_c^j}{\mu_j} = \frac{\gamma_\flat(x_\flat + \mathrm{O}(\mu_\flat))(s_c^\flat + \mathrm{O}(\mu_\flat^2))}{\gamma_\flat \mu_\flat} = \frac{x_\flat s_c^\flat}{\mu_\flat} + x_\flat \mathrm{O}(\mu_\flat) + s_c^\flat \mathrm{O}(\mu_\flat).$$

Since $x_\flat \approx 1$ and $s_c^\flat \approx \mu_\flat$,

$$\frac{x^j s_c^j}{\mu_j} = \frac{x_\flat s_c^\flat}{\mu_\flat} + \mathrm{O}(\mu_\flat). \tag{26}$$

Using the facts established above that $x_c^j = x^j + \mathrm{O}(\mu_\flat)$ and $s_c^j = \gamma_\flat \mathrm{O}(\mu_\flat)$,

$$\frac{x_c^j s_c^j}{\mu_j} = \frac{x^j s_c^j}{\mu_j} + \frac{\gamma_\flat \mathrm{O}(\mu_\flat^2)}{\mu_j} = \frac{x^j s_c^j}{\mu_j} + \mathrm{O}(\mu_\flat).$$

Merging this with (26),

$$\frac{x_c^j s_c^j}{\mu_j} = \frac{x_\flat s_c^\flat}{\mu_\flat} + \mathrm{O}(\mu_\flat). \tag{27}$$

Now we use the condition in Lemma 5.3(b), $\|x_\flat s_c^\flat / \mu_\flat - e\| \leqslant 0.49$ to conclude from (26) and (27) that

$$\left\| \frac{x^j s_c^j}{\mu_j} - e \right\| \leqslant 0.49 + \mathrm{O}(\mu_\flat), \qquad \left\| \frac{x_c^j s_c^j}{\mu_j} - e \right\| \leqslant 0.49 + \mathrm{O}(\mu_\flat). \tag{28}$$

For $\mu_\flat$ sufficiently small, say $\mu \leqslant \tilde{\mu}_j \leqslant \tilde{\mu}_{j-1}$, the second relation above proves (i). To prove the other relations, we use (24): in the steplength computation,

$$0.5 = \left\| \frac{x^{j+1} s^{j+1}}{\gamma_j \mu_j} - e \right\| \leqslant \left\| \frac{x^j s_c^j}{\mu_j} - e \right\| + \frac{\mathrm{O}(\mu_\flat)}{\gamma_j} \leqslant 0.49 + \frac{\mathrm{O}(\mu_\flat)}{\gamma_j}.$$

This leads directly into $\gamma_j = \mathrm{O}(\mu_\flat)$ and hence $\mu_{j+1} = \mathrm{O}(\mu)\mathrm{O}(\mu^j) = \mathrm{O}(\mu_\flat^{j+1})$, proving (ii).

To prove (iii), we use Lemma 3.1 and the assumption that $x^j = x_\flat + \mathrm{O}(\mu_\flat)$ to get

$$x^{j+1} = x^j + \gamma_j \gamma_\flat \mathrm{O}(1) + \mathrm{O}(\mu_j) = x_\flat + \mathrm{O}(\mu_\flat).$$

This proves the relations in the lemma for $j \in \{1, \dots, p-1\}$ with $\mu \leqslant \tilde{\mu}_{p-1}$. Setting $\bar{\mu}_2 = \tilde{\mu}_{p-1}$ completes the proof. □

The application of the lemma above to the sequences generated by Algorithm 2.2 is summarized in the corollary below:

Corollary 5.6. *Consider an iteration k of Algorithm 2.2, starting at $w^k \in \mathcal{N}$, and assume that $\mu_k \leqslant \bar{\mu}_2$. Let $w_\flat^k$ be the point obtained after the safeguard. Then all $p-1$ simplified steps are completed, $\mu_{k+1} = \mathrm{O}(\mu_k^{p+1})$ and $x^{k+1} - x_\flat^k = \mathrm{O}(\mu_k)$.*

Proof. Immediate from Lemma 5.5 for $j = p-1$. □

This proves that in Algorithm 2.2, μ_k converges to zero with Q-order $p+1$. To complete the proof of the main theorem, we must still prove that the safeguard must stop being triggered beyond a value of μ that depends only on problem data.

Triggering of the safeguard

Lemma 5.7. *Consider a sequence (x^k) generated by Algorithm 2.2. There exists $\bar{\mu} > 0$ dependent only on problem data such that the safeguard can only be triggered once for $\mu_k \leqslant \bar{\mu}$.*

Proof. The proof will be done by showing that after triggering the safeguard for μ_k small, the condition in Fact 5.2(iii) will be forever satisfied. The statement of that condition will be reworded as follows: If μ_k is sufficiently small, say $\mu_k \leqslant \tilde{\mu}_1$, then $\delta(w^k) \leqslant 0.5$ and $\operatorname{dist}(x^k, 0) \leqslant 0.2$ imply $\delta(0.1 w_c^k + 0.9 w_a^k) < 0.42$ (the safeguard is not triggered).

By Corollary 5.6, $x^{k+1} = x_\flat^k + \mathrm{O}(\mu_k)$ and hence

$$\operatorname{dist}(x^{k+1}, 0) = \left\| \frac{x_\flat^k + \mathrm{O}(\mu_k)}{x^*} - e \right\| = \left\| \frac{x_\flat^k}{x^*} - e \right\| + \mathrm{O}(\mu_k) = \operatorname{dist}(x_\flat^k, 0) + \mathrm{O}(\mu_k).$$

Thus there exists $K > 0$ dependent on problem data such that

$$\operatorname{dist}(x^{k+1}, 0) \leqslant \operatorname{dist}(x_\flat^k, 0) + K\mu_k. \tag{29}$$

Consider now an iteration k in which the safeguard is triggered, and let $\bar{k} > k$ be the first iteration in which it is triggered again. By Fact 4.5 and (22), we have after centering

$$\operatorname{dist}(x_\flat^k, 0) \leqslant 0.19 + \mathrm{O}(\mu_k).$$

Merging this and (29), we get

$$\operatorname{dist}(x^{\bar{k}}, 0) \leqslant 0.19 + \mathrm{O}(\mu_k) + K \sum_{j=k}^{\bar{k}-1} \mu_j.$$

But the sequence (μ_k) converges to zero at least linearly because the algorithm is polynomial, and hence

$$K \sum_{j=k}^{\infty} \mu_j = \mathrm{O}(\mu_k),$$

and we arrive at

$$\operatorname{dist}(x^{\bar{k}},0) \leqslant 0.19 + \mathrm{O}(\mu_k).$$

If μ_k is sufficiently small, say $\mu_k \leqslant \bar{\mu} \leqslant \bar{\mu}_1$, then $\operatorname{dist}(x^{\bar{k}},0) < 0.2$. As we saw above, the safeguard cannot be triggered in iteration $\bar{k}$. This prevents the safeguard from being triggered twice for $\mu_k \leqslant \bar{\mu}$, completing the proof. □

This completes the proof of Theorem 2.3.

Acknowledgement

Gonzaga thanks Kees Roos and Tamas Terlaky, his hosts in Delft Technical University, for the fruitful discussions and for the opportunity of working in their dynamical research group.

References

[1] E.R. Barnes, S. Chopra and D.J. Jensen, "The affine scaling method with centering," Technical Report, Dept. of Mathematical Sciences, IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, 1988.

[2] J.F. Bonnans and C.C. Gonzaga, "Convergence of interior point algorithms for the monotone linear complementarity problem," *Mathematics of Operations Research* 21 (1996) 1–25.

[3] C.C. Gonzaga, "The largest step path following algorithm for monotone linear complementarity problems," *Mathematical Programming*, to appear.

[4] C.C. Gonzaga, "Large steps path-following methods for linear programming, Part I: Barrier function method," *SIAM Journal on Optimization* 1 (1991) 268–279.

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

[6] C.C. Gonzaga and R.A. Tapia, "On the convergence of the Mizuno–Todd–Ye algorithm to the analytic center of the solution set," *SIAM Journal on Optimization*, to appear.

[7] C.C. Gonzaga and R.A. Tapia, "On the quadratic convergence of the simplified Mizuno–Todd–Ye algorithm for linear programming," *SIAM Journal on Optimization*, to appear.

[8] O. Güler, "Generalized linear complementarity problems and interior point algorithms for their solutions," Internal Report, Dept. of Industrial Engineering and Operations Research, University of California, 1993.

[9] M. Kojima, N. Megiddo, T. Noma and A. Yoshise, *A unified approach to interior point algorithms for linear complementarity problems*, Lecture Notes in Computer Science, Vol. 538 (Springer, Berlin, 1991).

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

[11] M. Kojima, S. Mizuno and A. Yoshise, "A primal–dual interior point algorithm for linear programming," in: N. Megiddo, ed., *Progress in Mathematical Programming: Interior Point and Related Methods* (Springer, New York, 1989) pp. 29–47.

[12] Z.Q. Luo and Y. Ye, "A genuine quadratically convergent polynomial interior point algorithm for linear programming," Technical Report, Dept. of Management Science, University of Iowa, Iowa City, IA 52242, 1993.

[13] K. McShane, "Superlinearly convergent $\mathrm{O}(\sqrt{n}L)$-iteration interior-point algorithms for LP and the monotone LCP," *SIAM Journal on Optimization* 4 (1994) 247–261.

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

[15] S. Mehrotra, "Quadratic convergence in a primal–dual method," Technical Report 91-15, Dept. of Industrial Engineering and Management Science, Northwestern University, Evanston, IL 60208, 1991.

[16] S. Mehrotra, "On the implementation of a primal–dual interior point method," *SIAM Journal on Optimization* 2 (1992) 575–601.

[17] S. Mehrotra, "Asymptotic convergence in a generalized predictor-corrector method," *Mathematical Programming*, to appear.

[18] S. Mizuno, "A new polynomial time method for a linear complementarity problem," *Mathematical Programming* 56 (1992) 31–43.

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

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

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

[22] R.D.C. Monteiro and T. Tsuchiya, "Limiting behavior of the derivatives of certain trajectories associated with a monotone horizontal linear complementarity problem," Working Paper 92-28, Dept. of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721, Dec. 1992.

[23] R.D.C. Monteiro and S. Wright, "Local convergence of interior-point algorithms for degenerate monotone LCP," Preprint MSC-P357-0493, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, April 1993.

[24] C. Roos and J.P. Vial, "A polynomial method of approximate centers for linear programming," *Mathematical Programming* 54 (1992) 295–305.

[25] J. Renegar and M. Shub, "Unified complexity analysis for Newton LP methods," *Mathematical Programming* 53 (1992) 1–16.

[26] S. Wright and Y. Zhang, "A superquadratic infeasible-interior-point algorithm for linear complementarity problems," Preprint MSC-P418-0294, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, 1994.

[27] Y. Ye, "Improving the asymptotic convergence of interior-point algorithms for linear programming," Working Paper 91-15, Dept. of Management Science, University of Iowa, Iowa City, IA 52242, 1991.

[28] Y. Ye and K.M. Anstreicher, "On quadratic and $O(\sqrt{n}L)$ convergence of a predictor–corrector algorithm for the linear complementary problem," *Mathematical Programming* 62 (1993) 537–551.

[29] Y. Ye, O. Güler, R.A. Tapia and Y. Zhang, "A quadratically convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming," *Mathematical Programming* 59 (1993) 151–162.