On the Quadratic Convergence of the Simplified Mizuno-Todd-Ye Algorithm for Linear Programming

Clovis C. Gonzaga
Richard A. Tapia

June 1992
(revised November 1992)
(revised September 1994)

TR92-42

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
SEP 1994		**00-00-1994 to 00-00-1994**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
On the Quadratic Convergence of the Simplified Mizuno-Todd-Ye Algorithm for Linear Programming	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**34**	

On the Quadratic Convergence of the Simplified Mizuno-Todd-Ye Algorithm for Linear Programming*

Clovis C. Gonzaga† Richard A. Tapia ‡

June, 1992
(revised November 1992)
(revised September 1994)

Keywords: Linear programming, primal-dual interior-point algorithm, predictor-corrector algorithm, quadratic convergence

Abbreviated title: Quadratic Convergence of Mizuno-Todd-Ye

AMS (MOS) subject classification: 49M, 65K, 90C

---

*This research was initiated in February 1992 while the first author was visiting the Center for Research on Parallel Computation, Rice University. A preliminary version of this paper was presented at the Fourth SIAM Meeting on Optimization, Chicago, Illinois, May 1992.

†Department of Systems Engineering and Computer Science, COPPE-Federal University of Rio de Janeiro. Cx. Postal 68811 21945 Rio de Janiero, RJ Brazil. Research supported in part by NSF Grant DMS-8920550 while the author was visiting the Center for Applied Mathematics, Cornell University.

‡Department of Computational and Applied Mathematics and Center for Research on Parallel Computation, Rice University, Houston, Texas 77251-1892. Research supported in part by NSF Coop. Agr. No. CCR-8809615, AFOSR 89-0363, DOE DEFG05-86ER25017 and ARO 9DAAL03-90-G-0093.

Abstract

It is known that the Mizuno-Todd-Ye predictor-corrector primal-dual Newton interior-point method generates a duality-gap sequence which converges quadratically to zero, and this is accomplished with an iteration complexity of $O(\sqrt{n}L)$. Very recently the present authors demonstrated that the iteration sequence generated by this method converges, and this convergence is to the analytic center of the solution set. In the current work we show that within a finite number of iterations the Newton corrector step can be replaced with a simplified Newton corrector step and the resulting algorithm maintains $O(\sqrt{n}L)$ iteration complexity, quadratic convergence of the duality-gap sequence to zero, and convergence of the iteration sequence (however not necessarily to the analytic center). The simplified predictor-corrector algorithm requires only one linear solve per iteration in contrast to the two linear solves per iteration required by the original predictor-corrector algorithm.

1 Introduction and Preliminaries

The basic primal-dual interior-point method for linear programming was originally proposed by Kojima, Mizuno, and Yoshise [4] based on earlier work of Megiddo [8]. This method can be viewed as perturbed and damped Newton's method applied to the first-order conditions for a particular standard form linear program. They established linear convergence and an iteration complexity bound of $O(nL)$ for this basic algorithm. Soon after Mizuno, Todd, and Ye [11] considered a predictor-corrector variant of the Kojima-Mizuno-Yoshise basic algorithm. In their algorithm the predictor step is a damped Newton step and the corrector step is a perturbed (centered) Newton step. Hence one iteration of the predictor-corrector algorithm requires the solution of two linear systems; essentially two Newton steps. Hence when comparing convergence rate results they should technically be considered to be two-step results. Mizuno, Todd, and Ye established linear convergence for their predictor-corrector algorithm and a superior iteration complexity bound of $O(\sqrt{n}L)$.

We now briefly give a chronological account of the development of fast (superlinear) convergence for these primal-dual interior-point methods. We refer to the Kojima-Mizuno-Yoshise method as the basic method, and to

the Mizuno-Todd-Ye method as the predictor-corrector method. When we discuss convergence or convergence attributes of one of these methods we are describing the convergence of the duality-gap to zero. This interpretation has become standard in this area, even though convergence of the duality-gap sequence does not imply convergence of the iteration sequence. The convergence of the iteration sequence is certainly an important issue in its own right and to some extent has been neglected. For an interesting result concerning the convergence of the iteration sequence generated by the basic method see Tapia, Zhang, and Ye [12]. For a definitive result concerning the convergence of the iteration sequence for the predictor-corrector method see Gonzaga and Tapia [3].

Zhang, Tapia, and Dennis [19] demonstrated that under certain assumptions the algorithmic parameters in the basic method could be chosen so that superlinear convergence was obtained for degenerate problems and quadratic convergence was obtained for nondegenerate problems. However, they did not demonstrate that polynomial complexity would be retained. Zhang and Tapia [18] demonstrated that the algorithmic parameters in the basic algorithm could be chosen so that the polynomial complexity bound was maintained and superlinear convergence was obtained for degenerate problems, while quadratic convergence was obtained for nondegenerate problems. Ye, Tapia and Zhang [16] demonstrated that the predictor-corrector algorithm was superlinearly convergent for degenerate problems and quadratically convergent for nondegenerate problems while maintaining its $O(\sqrt{n}L)$ iteration complexity. McShane [6] independently obtained a similar result. Up to this point all superlinear convergence results assumed that the iteration sequence converged. Ye, Güler, Tapia, and Zhang [15], and independently Mehrotra [9], based on Ye, Tapia, and Zhang [16] demonstrated the surprising result that neither the nondegeneracy assumption nor the assumption of iteration sequence convergence was needed for the quadratic convergence of the predictor-corrector algorithm.

In this paper we add to the literature on the predictor-corrector algorithm by demonstrating that its quadratic convergence and $O(\sqrt{n}L)$ complexity are retained if one replaces the Newton corrector step with a simplified Newton step, i.e., the Jacobian from the Newton predictor step is used also in the computation of the corrector step. Hence the corrector step only requires a back-solve, and the complete iteration only requires the solution of one linear system. Actually the Newton corrector step cannot be replaced with a sim-

plified Newton corrector step at the beginning of the iterative process, but only after a particular criterion is satisfied. We demonstrate that this criterion will be satisfied within a finite number of iterations. We also show that the simplified algorithm generates an iteration sequence which is convergent, but not necessarily to the analytic center.

Recently Ye [14] was able to show that a variant of the Mizuno-Todd-Ye predictor-corrector algorithm could be given that eventually did not require the corrector step. He demonstrated that this variant algorithm gave subquadratic convergence (the Q-rate is two, but the Q_2-factor may be unbounded). Hence Ye attains a convergence rate of two with an algorithm which (eventually) only requires one linear solve per iteration. Our simplified Mizuno-Todd-Ye algorithm gives Q-quadratic convergence but requires the solution of one linear system and an additional back solve per iteration. It should be clear that any convergence rate analysis based on total number of arithmetic operations per iteration will favor the Ye variant. It should also be clear that numerical efficiency of an algorithm is determined by effective number of iterations needed for numerical convergence and not convergence rate alone.

The paper is organized as follows. In the remainder of this section we introduce our notation and several fundamental background notions. In Section 2 we discuss the primal-dual Newton step and the primal-dual simplified Newton step and derive several properties concerning these two steps. Some results on scaled projections from Gonzaga and Tapia will be collected in Section 3. These results will be used in Section 5. The Mizuno-Todd-Ye predictor-corrector algorithm is presented in Section 4. Section 5 begins with the presentation of the simplified predictor-corrector algorithm and then turns to establishing our convergence theory for the simplified predictor-corrector algorithm. In Section 6 we make some observations that imply that quadratic convergence is optimal for both the predictor-corrector method and its simplified variant. We indicate that cubic convergence might be obtained by appropriately modifying the corrector step.

Given a vector x, d, ϕ, the corresponding upper case symbol denotes (as usual) the diagonal matrix X, D, Φ defined by the vector.

We 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. denotes the vectors with components $u_i v_i$, u_i/v_i, etc. This notation is consistent as long as component-wise operations are given precedence over

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 frequently use the $O(\cdot)$ and $\Omega(\cdot)$ notation to express a 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 $x = O(\mu)$, or $x^k = O(\mu^k)$, means that there is a constant K (dependent on problem data) such that for every $k \in I\!N$, $\|x^k\| \leq K\mu^k$. Similarly, $x = \Omega(\mu)$, or $x^k = \Omega(\mu^k)$, means that there is $\epsilon > 0$ such that for every $k \in I\!N$, $\|x^k\| \geq \epsilon\mu^k$.

The primal and dual linear programming problems are:

$$(LP) \qquad \begin{array}{lrcl} \text{minimize} & c^T x & & \\ \text{subject to} & Ax & = & b \\ & x & \geq & 0, \end{array}$$

and

$$(LD) \qquad \begin{array}{lrcl} \text{maximize} & b^T y & & \\ \text{subject to} & A^T y + s & = & c \\ & s & \geq & 0, \end{array}$$

where $c \in I\!R^n$, $b \in I\!R^m$, $A \in I\!R^{m \times n}$. We assume that both problems have optimal solutions, and that the sets of optimal solutions are bounded. This is equivalent to the requirement that both feasible sets contain points satisfying all inequality constraints strictly.

Given any feasible primal-dual pair $(\tilde{x}, \tilde{s})$, the problems can be rewritten as

$$(LP) \qquad \begin{array}{lrcl} \text{minimize} & \tilde{s}^T x & & \\ \text{subject to} & Ax & = & b \\ & x & \geq & 0, \end{array}$$

and

$$(LD) \qquad \begin{array}{lrcl} \text{minimize} & \tilde{x}^T s & & \\ \text{subject to} & Bs & = & Bc \\ & s & \geq & 0, \end{array}$$

where B^T is a matrix whose columns span the null space of A. Popular choices for B^T are an orthonormal basis for the null space of A and $B = P_A$, the projection matrix into the null space of A.

The feasible sets for (LP) and (LD) will be denoted respectively by $\mathcal{P}$ and $\mathcal{D}$. Their relative interiors will be respectively $\mathcal{P}^0$ and $\mathcal{D}^0$.

The set of optimal solutions for the primal-dual pair of problems constitutes a face $F = F_P \times F_D$ of the polyhedron of feasible solutions, where F_P and F_D are respectively the primal and dual optimal faces. By hypothesis, this face is a compact set. It is well known that this face is characterized by a partition $\{B, N\}$ of the set of indices $\{1, \ldots, n\}$ such that $F_P = \{x \in \mathcal{P} \mid x_N = 0\}$ and $F_D = \{s \in \mathcal{D} \mid s_B = 0\}$. In the relative interior of the face F, $x_B > 0$ and $s_N > 0$.

We study algorithms that converge to the optimal face. Our main concern is with the behaviour of the iterates as they approach the optimal face. We want this to happen in such a manner that all limit points are in the relative interior of the optimal face. We shall see later on how this condition can be enforced by requiring some adherence to the central path. For detail on the central path see Gonzaga [2].

Given $\mu > 0$, $\mu \in I\!R$, the pair (x, s) of feasible primal and dual solutions is the central point $(x(\mu), s(\mu))$ associated with μ if

$$xs = \mu e,$$

where e stands for the vector of all ones, with dimension given by the context.

The central path is the curve in $I\!R^{2n}$ parametrized by the positive real μ, i.e.,

$$\mu \mapsto (x(\mu), s(\mu)).$$

Thus (x, s) is a central point if and only if

$$\begin{array}{rcl} xs & = & \mu e \\ Ax & = & b \\ Bs & = & Bc \\ x, s & \geq & 0, \end{array} \tag{1}$$

where the columns of B^T span the null space of A.

The first-order or Karush-Kuhn-Tucker (KKT) conditions for problem (LP) (or (LD)) are

$$\begin{array}{rcl} xs & = & 0 \\ Ax & = & b \\ A^T y + s & = & c \\ x, s & \geq & 0. \end{array}$$

The perturbed KKT conditions, for perturbation parameter $\mu > 0$, are

$$\begin{aligned} xs &= \mu e \\ Ax &= b \\ A^T y + s &= c \\ x, s &\geq 0. \end{aligned} \tag{2}$$

Observe that the perturbed KKT conditions are merely the defining relations for the central path and (2) can equivalently be written as (1). Essentially all primal-dual interior-point methods for problem (LP) consist of some variant of the damped Newton method applied to the perturbed KKT conditions (1) or (2).

2 The Newton and Simplified Newton Steps

When dealing with an iterative procedure we will use the superscript 0 to denote the previous iterate, no superscript to denote the current iterate, a superscript of + to denote the subsequent iterate. In two-step algorithms like the Mizuno-Todd-Ye algorithm described in Section 4 this notation will apply to the current iterate, the intermediate iterate, and the final iterate.

Suppose that (x^0, s^0) and (x, s) have been obtained from a form of Newton's method and are both feasible pairs. The Newton step (or correction) for (1) at (x, s) is given by (u, v) the solution of

$$\begin{aligned} xv + su &= -xs + \mu e \\ Au &= 0 \\ Bv &= 0, \end{aligned} \tag{3}$$

and the simplified Newton step for (1) at (x, s) is given by (u, v) the solution of

$$\begin{aligned} x^0 v + s^0 u &= -xs + \mu e \\ Au &= 0 \\ Bv &= 0. \end{aligned} \tag{4}$$

It should be clear that the difference between (3) and (4) is that (3) uses the Jacobian of (1) at (x, s) and (4) uses the Jacobian of (1) at (x^0, s^0).

We introduce some additional notation that will be used throughout the paper. Given a pair (x, s), we define

$$\begin{aligned} \mu(x,s) &= x^T s/n \\ w(x,s) &= xs/\mu(x,s) \\ \delta(x,s) &= \|w(x,s)-e\| \\ \phi(x,s) &= (\sqrt{w(x,s)})^{-1}. \end{aligned} \tag{5}$$

When no confusion can arise, we drop the reference to the variables, and continue to use other symbols in a consistent manner. For instance, given a pair $(\bar{x},\bar{s})$, the parameters above will be denoted simply $\bar{\mu},\bar{w}$ and $\bar{\phi}$.

Given a pair (x,s), $\mu(x,s)$ is the penalty parameter associated to (x,s), in the following sense: if (x,s) is a central point, then $xs=\mu e$; otherwise μ is the penalty parameter associated with the central point that is nearest the pair (x,s), in terms of a certain proximity measure. The vector w consists of logarithmic barrier weights associated with (x,s). It characterizes the weighted primal-dual affine scaling trajectory through (x,s), as studied by Monteiro and Adler [11]. The scalar δ is a measure of proximity from (x,s) to the central point $(x(\mu),s(\mu))$. The definition of ϕ was made merely for convenience; it will simplify expressions below.

At this point we are interested in obtaining usable closed form solutions for the simplified Newton step and the Newton step. We also derive an interesting property of the simplified Newton step. In what follows it is important not to confuse μ in (3) and (4) with $\mu(x,s)$ given in (5), because they are not necessarily the same. Hence μ denotes the μ in (3) and (4) and $\mu(x,s)$ means the $\mu(x,s)$ given in (5). Since no confusion will arise in the case of μ^0, we use μ^0 to denote $\mu(x^0,s^0)$.

Proposition 2.1 *The simplified Newton step (u,v) given by (4) can be written*

$$\begin{aligned} u &= x^0\phi^0 P_{AX^0\Phi^0\phi^0}\left(-\frac{xs}{\mu^0}+\frac{\mu}{\mu^0}e\right) \\ v &= s^0\phi^0 \tilde{P}_{AX^0\Phi^0\phi^0}\left(-\frac{xs}{\mu^0}+\frac{\mu}{\mu^0}e\right) \end{aligned} \tag{6}$$

where $\tilde{P}=I-P$.

Proof. Assume that instead of (4), the simplified Newton equations are written as

$$x^0 v + s^0 u = -xs+\mu e \ , \ u\in\mathcal{N}(A), \ v\in\mathcal{R}(A^T) \tag{7}$$

where as usual $\mathcal{N}$ denotes null space and $\mathcal{R}$ denotes range space.

The solution is obtained by associating a scaling vector

$$d(x,s) = \sqrt{\frac{x}{s}}$$

to each pair (x,s).

Using the definitions in (5) and dropping argument references when no confusion will arise

$$d = \sqrt{\frac{x}{s}} = \frac{x\phi}{\sqrt{\mu(x,s)}} = \frac{\sqrt{\mu(x,s)}}{\phi s} \tag{8}$$

The solution of (7) is obtained by scaling the problems by $\bar{x} = (d^0)^{-1}x$, $\bar{s} = d^0 s$:

$$\begin{aligned} \bar{x}^0\bar{v} + \bar{s}^0\bar{u} &= -\bar{x}\bar{s} + \mu e \\ \bar{u} &\in \mathcal{N}(AD^0) \\ \bar{v} &\in \mathcal{R}(D^0A^T) \end{aligned}$$

The choice of this scaling becomes clear when we notice that by direct substitution,

$$\bar{x}^0 = \bar{s}^0 = \sqrt{x^0 s^0} \tag{9}$$

Dividing the equation by $\bar{s}^0$ and using the definitions of scaled variables,

$$\bar{u} + \bar{v} = -\frac{\bar{x}}{\bar{x}^0}\bar{s} + \mu(\bar{x}^0)^{-1} = \frac{d^0}{x^0}(-xs + \mu e).$$

Hence $\bar{u}$ and $\bar{v}$ are the components of the right-hand side in the complementary subspaces, the null space and row space of AD^0, and are given by

$$\begin{aligned} \bar{u} &= P_{AD^0}\frac{d^0}{x^0}(-xs + \mu e) \\ \bar{v} &= \tilde{P}_{AD^0}\frac{d^0}{x^0}(-xs + \mu e), \end{aligned} \tag{10}$$

where $\tilde{P}_{AD^0} = I - P_{AD^0}$. Finally, $u = d^0\bar{u}$ and $v = (d^0)^{-1}\bar{v}$.

A convenient formulation is obtained by substituting $d^0 = \frac{1}{\sqrt{\mu^0}}x^0\phi^0$ and $(d^0)^{-1} = \frac{1}{\sqrt{\mu^0}}s^0\phi^0$, and this leads to (6). ∎

The simplified Newton step and the Newton step satisfy an interesting property. This property will turn out to be fundamental to the analysis presented in Section 5. Hence we derive this property in a form which covers both the simplified Newton step and the Newton step.

Proposition 2.2 *Let* $(\hat{x}, \hat{s})$ *and* (x, s) *be feasible pairs. Consider* $x^+ = x + u$ *and* $s^+ = s + v$ *where* (u, v) *satisfies*

$$\begin{gathered} \hat{x}v + \hat{s}u = -(1 - \hat{\gamma})xs + \hat{\mu}e \\ u \in \mathcal{N}(A) \\ v \in \mathcal{R}(A^T) \ . \end{gathered}$$

Then

$$\mu(x^+, v^+) = \hat{\gamma}\mu(x, s) + \hat{\mu} \ . \tag{11}$$

Proof. Left multiplying by e^T, we obtain

$$\hat{x}^T v + \hat{s}^T u = -(1 - \hat{\gamma})x^+ s + n\hat{\mu} \ .$$

From the definition

$$x^{+T} s^+ = x^T s + x^+ v + s^T u \ ,$$

since $u^T v = 0$. But $\hat{x}^T v = x^T v$, because $\hat{x} - x \in \mathcal{N}(A)$ and $v \in \mathcal{R}(A^T)$, and similarly $\hat{s}^T u = s^T u$. Substituting in the expressions above we immediately obtain (11). ∎

3 Scaled Projections

In this section we collect some results on scaled projections from Gonzaga and Tapia [3]. These results are extensions of results published by Megiddo and Shub [8]. We use $\mathbb{R}_+$ to denote the nonnegative reals, and $\mathbb{R}_{++}$ to denote the positive reals.

Consider the primal feasible set for (LP),

$$\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$$

and the map h defined for $d \in \mathbb{R}^n_+$, $d \neq 0$ and $\rho \in \mathbb{R}^n$ by

$$(d, \rho) \mapsto h(d, \rho) = P_{AD}\rho, \tag{12}$$

where P_{AD} represents the projection matrix into the null space of AD.

We study the behaviour of this map when $d > 0, d \to \bar{d}$ and $\rho \to \bar{\rho}$, where $\bar{d} \geq 0, \bar{d} \neq 0$, and $\bar{\rho} \in I\!R^n$.

Given $\bar{d}$, we define the index sets $B = \{i = 1, \ldots, n \mid \bar{d}_i > 0\}$ and $N = \{i = 1, \ldots, n \mid \bar{d}_i = 0\}$. The variables with indices in B are called the "large variables," and the others are called the "small variables." It is difficult to describe the behaviour of the small variables $h_N(d, \rho)$ of the scaled projection defined above. The theory of Megiddo and Shub concerns the large variables $h_B(d, \rho)$. We shall describe this theory conveniently extended to fit our needs. The following proposition is Lemma 3.2 of Gonzaga and Tapia[3]. We refer the reader to that paper for the proof.

Proposition 3.1 *Let $h(d, \rho)$ be given by (12). Consider $(\bar{d}, \bar{\rho}) \in I\!R^n_+ \times I\!R^n, \bar{d} \neq 0$, and $(d^k, \rho^k) \in I\!R^n_+ \times I\!R^n$ such that $(d^k, \rho^k) \to (\bar{d}, \bar{\rho})$. Then*
(i) $h_B(d^k, \rho^k) \to h_B(\bar{d}, \bar{\rho}) = P_{A_B \bar{D}} \bar{\rho}_B$.
(ii) If $\bar{\rho}_N = 0$, then $h_N(d^k, \rho^k) \to 0$.

Consider compact sets $\Gamma \subset I\!R^n$ and $\Delta \subset I\!R^n_+$, such that for any $d \in \Delta$, $d_B > 0$ and $d_N = 0$, where $\{B, N\}$ is a partition of $\{1, \ldots, n\}$. We now extend the proposition above for the case of sequences $\{d^k\}$ in $I\!R^n_{++}$ and $\{\rho^k\} \in I\!R^n$ such that $d^k \to \Delta$ and $\rho^k \to \Gamma$ *.

Proposition 3.2 *For the situation described above we have the following:*
(i) If $d^k \to \Delta$ and $\rho^k \to \Gamma$, then

$$h_B(d^k, \rho^k) - P_{A_B D^k_B} \rho^k_B \to 0.$$

(ii) If $d^k \to \bar{d} \in \Delta$ and $\rho^k \to \bar{\rho} \in \Gamma$, then

$$h_B(d^k, \rho^k) - P_{A_B \bar{D}_B} \bar{\rho}_B \to 0.$$

Proof. Implication (ii) follows from (i), since for convergent sequences $P_{A_B D^k_B} \rho^k_B \to P_{A_B \bar{D}_B} \bar{\rho}_B$.

To prove (i), assume by contradiction that there exists $\epsilon > 0$ and sequences $\{d^k\}$ in $I\!R^n_{++}$ and $\{\rho^k\}$ in $I\!R^n$ such that for $k = 1, 2, \ldots$

$$\| h_B(d^k, \rho^k) - P_{A_B D^k_B} \rho^k_B \| > \epsilon. \tag{13}$$

*A sequence $\{z^k\}$ converges to a set Z if $d(z^k, Z) \to 0$, where $d(z^k, Z) = \inf_{z \in Z} \|z^k - z\|$.

Since the sequences $\{d^k\}$ and $\{\rho^k\}$ converge to compact sets they must be bounded. Hence they have accumulation points $\bar{d}, \bar{\rho}$, such that for some $\mathcal{K} \subset I\!N$, $d^k \xrightarrow{\mathcal{K}} \bar{d}$ and $\rho^k \xrightarrow{\mathcal{K}} \bar{\rho}$. From the fact that d^k converges to Δ and ρ^k converges to Γ and the compactness of Δ and Γ, $\bar{d} \in \Delta$ and $\bar{\rho} \in \Gamma$. From Proposition 3.1,

$$h_B(d^k, \rho^k) \xrightarrow{\mathcal{K}} P_{A_B \bar{D}_B \bar{\rho}_B},$$

and since $\bar{D}_B > 0$,

$$P_{A_B D_B^k \rho_B^k} \xrightarrow{\mathcal{K}} P_{A_B \bar{D}_B \bar{\rho}_B}.$$

Subtracting these last expressions we see that

$$h_B(d^k, \rho^k) - P_{A_B D_B^k \rho^k} \xrightarrow{\mathcal{K}} 0,$$

contradicting (13) and completing the proof. ∎

Now we present two facts related to projections and slightly shifted scalings.

Proposition 3.3 *Let $q \in I\!R^N$ be such that $\|q - e\|_\infty \leq \alpha$, $\alpha \in (0, 0.25)$, and consider the projections $\hat{h} = P_A \rho$, $h = q P_{AQ} q \rho$. Then $\|h - \hat{h}\| \leq 3\alpha \|\hat{h}\|$.*

Proof. See [3]. ∎

Given a vector $x \in I\!R^n_{++}$, the following map defines a norm

$$h \in I\!R^n \mapsto \|h\|_x = \|x^{-1} h\|.$$

This is the Euclidean norm of the vector corresponding to h after a scaling $\bar{h} = x^{-1} h$. This norm is very usual in interior-point methods.

The following result shows that all scaled norms for x in a compact set in the interior of the positive orthant are uniformly equivalent.

Proposition 3.4 *Let $\Delta \subset I\!R^n_{++}$ be a compact set. Then there is a number $\Gamma > 0$ such that for any $h \in I\!R^n$, $x \in \Delta$,*

$$\frac{1}{\Gamma} \|h\| \leq \|h\|_x \leq \Gamma \|h\|.$$

Proof. By definition, given $x \in \Delta$, $\|h\|_x = \|x^{-1} h\|$. We immediately obtain

$$\min_{i=1,\ldots,n} x_i^{-1} \|h\| \leq \|h\|_x \leq \max_{i=1,\ldots,n} x_i^{-1} \|h\|$$

Since x_i , $i = 1, \ldots, n$ are bounded and bounded away from zero for $x \in \Delta$, the scalar Γ must exist, completing the proof. ∎

4 The Mizuno-Todd-Ye Algorithm

The Mizuno-Todd-Ye (MTY) algorithm is a path-following predictor-corrector algorithm. All activity is restricted to a region near the central path, i.e., all points (x, s) generated by the algorithm satisfy

$$\delta(x,s) = \|w(x,s) - e\| = \|\frac{xs}{\mu(x,s)} - e\| \leq \alpha,$$

where $\alpha \in (0, 0.5)$.

We shall describe a typical iteration of the algorithm and list its properties. Complete proofs can be found in Mizuno, Todd, and Ye [10].

Given $\alpha = 0.1$*, a typical iteration begins with feasible (x^0, s^0) such that $\delta(x^0, s^0) = \|w^0 - e\| \leq \alpha^2/\sqrt{2}$.

Predictor step: Given (x^0, s^0) compute the (affine-scaling) step (u^0, v^0) and let $x = x^0 + u^0$, $s = s^0 + v^0$, where (u^0, v^0) is defined by

$$x^0 v^0 + s^0 u^0 = -(1-\gamma)x^0 s^0, \quad u^0 \in \mathcal{N}(A), \quad v^0 \in \mathcal{R}(A^T),$$

with $\gamma \in [0, 1)$ such that $\delta(x, s) = \|w(x, s) - e\| \leq \alpha$. (The specific choice of γ will be discussed below.)

Corrector step: Given (x, s) compute the (centering) step (u, v) and let $x^+ = x + u$, $s^+ = s + v$, where (u, v) is defined by

$$xv + su = -xs + \mu e, \quad u \in \mathcal{N}(A), v \in \mathcal{R}(A^T)$$

with $\mu = \mu(x, s)$.

Observe that our γ in the predictor step is effectively a steplength parameter. To see this let us denote the predictor step by $(u^0(\gamma), v^0(\gamma))$ and let $\theta = 1 - \gamma$. Then

$$\theta(u^0(0), v^0(0)) = (u^0(\gamma), v_0(\gamma))$$

*The original paper uses $\alpha = 0.5$. We shall use a convenient value of 0.1, since this simplifies some formulas ahead.

and

$$(x, s) = (x^0, s^0) + \theta(u^0(0), \gamma^0(0)) \ ;$$

which is the usual way of writing the MTY predictor step. The usual choice for θ is θ^k, the largest $\theta \in (0, 1]$ such that $\delta(x(\theta), s(\theta)) \leq \alpha$ for all $0 \leq \theta \leq \theta^k$. For further detail see, for example, Section 2 of Ye, Güler, Tapia, and Zhang [15]. Hence our choice for γ in the predictor step is $\gamma = 1 - \theta^k$, and can be viewed as the smallest $\gamma \in [0, 1)$ in the sense just described.

Mizuno, Todd, and Ye [10] prove that the algorithm is well defined, in the sense that the centering step produces (x^+, s^+) such that $\delta(x^+, s^+) \leq \alpha^2/\sqrt{2}$. Ye, Güler, Tapia, and Zhang [15] (and independently Mehrotra [9]) prove that the duality-gap (or equivalently the parameter μ) converges to zero Q-quadratically, i.e.,

$$\mu^+ = \mu(x^+, s^+) = O(\mu^{0^2}).$$

Using Proposition 2.2 with $(\hat{x}, \hat{s}) = (x^0, s^0)$, $\hat{\gamma} = \gamma$, and $\hat{\mu} = 0$, we see that for the corrector step

$$\mu(x, s) = \gamma\mu(x^0, s^0) \ .$$

Using Proposition 2.2 with $(\hat{x}, \hat{s}) = (x, s)$, $\hat{\gamma} = 0$, and $\hat{\mu} = \gamma\mu(x^0, s^0)$, we see that for the corrector step

$$\mu(x^+, s^+) = \gamma\mu(x^0, s^0) \ .$$

So, on one hand we have $\mu^+ = O(\mu^{0^2})$ and on the other hand we have $\mu^+ = \gamma\mu^0$. It follows that

$$\gamma = O(\mu^0).$$

Bounds on the quantities appearing in the algorithm are given in the propositions below. Let $\{B, N\}$ be the optimal partition for the linear programming problem, i.e., the index partition associated to the optimal face. It is well known (see Adler and Monteiro [1]) that the central path ends at the analytic center of the optimal face, and that the pairs (x, s) such that $\|w(x, s) - e\| \leq \alpha$ constitute a neighborhood of the central path corresponding to the bundle of w-weighted affine-scaling trajectories for w such that $\|w - e\| \leq \alpha$. For α small, the bundle of trajectories ends in a compact neighborhood of the analytic center of the optimal face, contained in the

relative interior of the face. Namely, the end points in the primal optimal face are the w-weighted centers given by

$$x^*(w) = \text{argmin} \left\{ \sum_{i \in B} w_i \log x_i \mid A_B x_B = b \right\} .$$

Hence, the algorithm behaves as follows. As the optimal face is approached (and this happens in polynomial time), $x_N^k \to 0$, $s_B^k \to 0$ and x_B^k, s_N^k remain in small neighborhoods of x_B^* and s_N^*, the analytic centers of the primal and dual optimal faces.

Actually, it is always true that $x^k \to x^*$, $s^k \to s^*$, due to the results proved in Gonzaga and Tapia [3], which we describe.

As was stressed in the beginning of Section 6 of Gonzaga and Tapia [3], it is important to realize that our estimates do not require $(x^{0^{k+1}}, s^{0^{k+1}})$ to be related to (x^{+^k}, s^{+^k}), i.e. (x^{0^k}, s^{0^k}) does not have to be generated by the MTY algorithm. All that is required is that (x^{0^k}, s^{0^k}) satisfy the condition $w\|(x^{0^k}, s^{0^k}) - e\| \leq \alpha$, for the appropriate choice of α. Hence in what follows in this section and in Section 5 we employ this broad interpretation when discussing quantities generated by the MTY algorithm or the simplified MTY algorithm for only one iteration.

Proposition 4.1 *Consider quantities generated by the MTY algorithm. Then*

(i) $x_N = O(\mu)$, $s_B = O(\mu)$, $x_N^0 = O(\mu^0)$, $s_B^0 = O(\mu^0)$
(ii) $u^0 = O(\mu^0)$, $v^0 = O(\mu^0)$
(iii) $u_N = O(\mu)$, $v_B = O(\mu)$

Proof. See Lemma 5.1 of [3]. ∎

The proposition above shows that the variations in (x, s) due to either an MTY predictor or corrector step are bounded by $O(\mu^0)$, with exception of u_B and v_N. These are the variations in the large variables due to the corrector step.

The following proposition is the main result in Gonzaga and Tapia [3]. It is related to the map that associates to a pair (x^0, s^0) the pair (x^+, s^+) resulting from a MTY iteration. The proposition says that near the optimal face, a MTY iteration causes the large variables to approach the large variables of the analytic center (x^*, s^*) of the optimal face. The proposition describes

only the behaviour of the primal variables; the dual variables behave in a similar fashion, due to the symmetry of the optimality conditions (1).

The approach to the center is measured in the norm relative to x_B^*, defined for $h \in \mathbb{R}^n$ by $\|h_B\|_* = \|(x_B^*)^{-1} h_B\|$.

Proposition 4.2 *Consider a sequence (not necessarily generated by the algorithm) (x^{0^k}, s^{0^k}) of primal-dual pairs such that $\delta(x^{0^k}, s^{0^k}) \leq 0.1$ and $\mu^{0^k} \to 0$. Then there exists a sequence of positive reals $\{\epsilon^k\}$ such that $\epsilon^k \to 0$ and for sufficiently large k,*

$$\|x_B^{+^k} - x_B^*\|_* \leq \max\{\epsilon^k, 0.8\|x_B^{0^k} - x_B^*\|_*\}.$$

Proof. See Lemma 6.2 of [3]. ∎

This result implies that the iterates approach (x^*, s^*) and thus the sequence generated by the algorithm converges to the central optimum.

We are now concerned with bounding the sum of the variations (corrections) made to either the x-variable or the s-variable in either the predictor step or the corrector step in all iterations. The variation in x due to a predictor step is u^0. By the total variation in x due to predictor steps we mean $\sum_k \|u^{0^k}\|$. If we do not mention predictor steps or corrector steps we mean both steps. Analogous terminology is used for corresponding situations.

Proposition 4.3 *Consider quantities x^{0^k}, s^{0^k}, x^k, etc. generated by the MTY algorithm starting at (x^{0^1}, s^{0^1}).*

Then

(i) $\sum_{k=1}^{\infty} \mu^{0^k} = O(\mu^{0^1})$.

(ii) The total variation in x_N and in s_B is bounded by $O(\mu^{0^1})$.

(iii) The total variation in x_B and in s_N due to predictor steps is bounded by $O(\mu^{0^1})$.

Proof. To prove (i), it is enough to show that for some constant $\beta \in (0, 1)$, $\mu^{k+1} \leq \beta \mu^k$. This was shown by Mizuno, Todd, and Ye [10] when proving the polynomiality of the algorithm. Now (ii) and (iii) are direct consequences of Proposition 4.1, completing the proof. ∎

5 The Simplified Mizuno-Todd-Ye Algorithm

The simplified MTY algorithm is the MTY algorithm with the Newton corrector step replaced by a simplified Newton step. This means that the computation of the projections in (6) for the corrector step are reduced to a back substitution, instead of a complete solution of the system.

We now state the complete algorithm.

Algorithm 5.1 *Given* $\alpha \leq 0.1$, *and feasible* (x^{01}, s^{01}) *such that* $\delta(x^{01}, s^{01}) \leq \frac{\alpha^2}{\sqrt{2}}$, *set* $k = 1$.

REPEAT

$x^0 := x^{0^k}$, $s^0 := s^{0^k}$, $\mu^0 := \mu(x^0, s^0)$.

Predictor: Given (x^0, s^0) compute (u^0, v^0), and let $x := x^0 + u^0$, $s := s^0 + v^0$ where (u^0, v^0) satisfies
$x^0 v^0 + s^0 u^0 = -(1-\gamma)x^0 s^0, \quad u^0 \in \mathcal{N}(A)$, $v^0 \in \mathcal{R}(A^T)$,
and γ is as in the MTY predictor step.

Simplified Corrector: Given (x, s) set $\mu := \mu(x, s)$. Compute $(\hat{u}, \hat{v})$ satisfying
$x^0 \hat{v} + s^0 \hat{u} = -xs + \mu e, \quad \hat{u} \in \mathcal{N}(A)$, $\hat{v} \in \mathcal{R}(A^T)$,
and set $x^+ := x + \hat{u}$, $s^+ := s + \hat{v}$.

Safeguard: If $\delta(x^+, s^+) > \alpha/2$, then discard (x^+, s^+) and compute the Newton corrector step
$xv + su = -xs + \mu e, \quad u \in \mathcal{N}(A)$, $v \in \mathcal{R}(A^T)$,
and set $x^+ := x + u$, $s^+ := s + v$.

Subsequent iterate:
$x^{0^{k+1}} := x^+$, $s^{0^{k+1}} := s^+$.

$k := k + 1$

UNTIL convergence.

Before we formally state the convergence properties that we have derived for the simplified predictor-corrector algorithm, there is value in collecting some fundamental observations. In what follows all quantities should be indexed by k; however as we have been doing above we will not always write the index k.

Proposition 5.2 *Let* $\{(x^0, s^0)^k, (x, s)^k, (x^+, s^+)^k\}$ *be generated by the simplified MTY predictor-corrector algorithm. Then*

(i) $x^{+T} s^+ = x^T s$

(ii) $x^T s = \gamma x^{0T} s^0$

(iii) $\gamma = O(x^{0T} s^0)$

(iv) $x^T s \leq (1 - \frac{\delta}{\sqrt{n}}) x^{0T} s^0$ *for some* $\delta > 0$ *that does not depend on* k.

Proof. The proof of (i) follows from Proposition 2.2 with $(\hat{x}, \hat{s}) = (x, s)$, $\hat{\gamma} = 0$, and $\hat{\mu} = \mu(x, s)$. The proof of (ii) follows from Proposition 2.2 with $(\hat{x}, \hat{s}) = (x^0, s^0)$, $(x, s) = (x^0, s^0)$, $\hat{\gamma} = \gamma$ and $\hat{\mu} = 0$. Both (iii) and (iv) follow from Theorem 4.1 of Ye, Güler, Tapia, and Zhang [15], once we observe that their β is related to our α by the relationship $\beta = \frac{\alpha}{2}$ and their steplength θ is related to our γ by the relationship $\theta = 1 - \gamma$. ∎

The algorithm uses a simplified Newton iteration in the corrector step. If the simplified corrector produces the reduction in the proximity δ that ensures the quadratic convergence of the algorithm, i.e., if $\delta(x^+, s^+) \leq \alpha/2$, then the step is accepted. Otherwise the simplified step is discarded and the algorithm performs a Newton corrector step.

Two things must be proved: first that the iterates are still convergent, not necessarily to the analytic center of the optimal face, and second, that the safeguard cannot be activated more than a finite number of times.

The predictor step is the same as that for the MTY algorithm. Our analysis will be based on a comparison of the simplified and exact corrector steps. The conclusions will be the following: For points near the optimal face

(i) The simplified corrector step does not center the large variables. The variation in x_B and s_N due to simplified steps will be bounded by $O(\mu^0)$.

(ii) The behaviour of the small variables x_N and s_B tends to be identical in both methods.

These two facts will be proved and then used to contradict the hypothesis that the safeguard is activated an infinite number of times.

We begin by studying the behaviour of the large variables.

Proposition 5.3 *Consider the corrector directions (u^k, v^k) and $(\hat{u}^k, \hat{v}^k)$ generated at iteration k of Algorithm 5.1 (independently of which one is actually accepted by the algorithm). Then there exist a number $K > 0$ and sequences $\{\theta_x^k\}$, $\{\theta_s^k\}$ in $\mathbb{R}_+$ such that $\theta_x^k \to 0$, $\theta_s^k \to 0$ and*

$$\|\hat{u}_B^k\| \leq \gamma^k K(\|u_B^k\| + \theta_x^k) \ , \ \|\hat{v}_N^k\| \leq \gamma^k K(\|v_N^k\| + \theta_s^k).$$

Hence $\hat{u}_B = O(\mu^0)$ and $\hat{v}_N = O(\mu^0)$.

Proof. We shall prove the result for $\hat{u}_B^k$. The proof of the other result is similar.

Dropping the index k for notational simplicity, the primal directions are computed from (6):

$$\begin{aligned} \hat{u} &= x^0\phi^0 P_{AX^0\Phi^0\phi^0}\left(-\frac{xs}{\mu^0} + \frac{\mu}{\mu^0}e\right), \\ u &= x\phi P_{AX\Phi\phi}\left(-\frac{xs}{\mu} + e\right). \end{aligned}$$

Substituting $\mu = \gamma\mu^0$, we obtain for $\rho = \left(-\frac{xs}{\mu} + e\right)$,

$$\begin{aligned} \frac{\hat{u}}{\gamma} &= x^0\phi^0 P_{AX^0\Phi^0\phi^0}\rho, \\ u &= x\phi P_{AX\Phi\phi}\rho. \end{aligned}$$

The points x^k and x^{0^k} approach the relative interior of the optimal face, converging to a small compact neighborhood of the central optimum x^*. The vectors ϕ and ϕ^0 have the following bounds.

By construction, $w_i^0 \in [0.95, 1.05]$, $w_i \in [0.9, 1.1]$. Since $\phi_i = 1/\sqrt{w_i}$ by definition, the following bounds can be easily checked:

$$\phi_i^0 \in [0.97, 1.03] \ , \ \phi_i \in [0.95, 1.06] \ , \ \frac{\phi_i^0}{\phi_i} \in [0.92, 1.08]. \tag{14}$$

Thus $x^0\phi^0$ and $x\phi$ also converge to compact sets. Since $\|\rho\| = \delta(x,s) \leq 0.1$, the vectors $\phi\rho$ and $\phi^0\rho$ are also in compact sets, and we can use Proposition 3.2 to obtain

$$\begin{aligned} \tfrac{\hat{u}_B}{\gamma} - x_B^0 \phi_B^0 P_{A_B X_B^0 \Phi_B^0} \phi_B^0 \rho_B &\rightarrow 0, \\ u_B - x_B \phi_B P_{A_B X_B \Phi_B} \phi_B \rho_B &\rightarrow 0. \end{aligned} \tag{15}$$

The scaled projections above are almost in the format required by Proposition 3.3, on slightly shifted scalings. To put them in the desired format, let us write

$$\rho_B = x_B^0 (x_B^0)^{-1} \rho_B.$$

Due to Proposition 4.1, since $x_B = \Omega(1)$, we have

$$x_B = x_B^0 + u_B^0 = x_B^0 (e + O(\mu^0)) \,. \tag{16}$$

It follows that $(x_B^0)^{-1} = x_B^{-1}(e + O(\mu^0))$. Thus

$$\rho_B = x_B^0 x_B^{-1} \rho_B (e + O(\mu^0)) = x_B^0 x_B^{-1} \rho_B + O(\mu^0).$$

Since $O(\mu^0) \rightarrow 0$, (15) can be written as

$$\frac{\hat{u}_B}{\gamma} - x_B^0 \phi_B^0 P_{A_B X_B^0 \Phi_B^0} x_B^0 \phi_B^0 x_B^{-1} \rho_B \rightarrow 0, \tag{17}$$

$$u_B - x_B \phi_B P_{A_B X_B \Phi_B} x_B \phi_B x_B^{-1} \rho_B \rightarrow 0. \tag{18}$$

Defining $q = \dfrac{x_B^0 \phi_B^0}{x_B \phi_B}$, we see from (14) and (16) that for μ sufficiently small, $q_i \in [0.9, 1.1]$, and thus $\|q - e\|_\infty \leq 0.1$. Now (17) can be written as

$$\frac{\hat{u}_B}{\gamma} - x_B \phi_B q P_{A_B X_B \Phi_B Q} x_B \phi_B q \; x_B^{-1} \rho_B \rightarrow 0. \tag{19}$$

Defining $\hat{h}_B = q P_{A_B X_B \Phi_B Q} x_B \phi_B q \; x_B^{-1} \rho_B$, $h_B = P_{A_B X_B \Phi_B} x_B \phi_B \; x_B^{-1} \rho_B$, we see from Proposition 3.3 that

$$\|h_B - \hat{h}_B\| \leq 0.3 \|h_B\|.$$

Dividing (18) by $x_B \phi_B$, and using scaled norms, it follows that

$$\|u_B\|_{x_B \phi_B} - \|h_B\| \rightarrow 0. \tag{20}$$

Subtracting (18) from (19) establishes that

$$\frac{\frac{\hat{u}_B}{\gamma} - u_B}{x_B \phi_B} + \hat{h}_B - h_B \rightarrow 0, \tag{21}$$

or (making the iteration indices explicit),

$$\begin{aligned} \|\frac{\hat{u}_B^k}{\gamma^k} - u_B^k\|_{x_B^k \phi_B^k} &\leq \|h_B^k - \hat{h}_B^k\| + \sigma_1^k \quad , \quad \sigma_1^k \to 0 \\ &\leq 0.3\|h_B^k\| + \sigma_1^k \\ &\leq 0.3\|u_B^k\|_{x_B^k \phi_B^k} + \sigma_2^k, \end{aligned}$$

where the last inequality comes from (20), with $\sigma_2^k \to 0$.

Using Proposition 3.4 twice to relate $\|\cdot\|_{x_B^k \phi_B^k}$ and $\|\cdot\|$, we see that there exists a constant $K_1 > 0$ such that

$$\|\frac{\hat{u}_B^k}{\gamma^k} - u_B^k\| \leq K_1\|u_B^k\| + \theta_x^k \ ,$$

where $\theta_x^k \to 0$. Finally,

$$\|\frac{\hat{u}_B^k}{\gamma^k}\| \leq \|u_B^k\| + K_1\|u_B^k\| + \theta_x^k.$$

The final statement follows from the fact that $\{u_B^k\}$ and $\{v_B^k\}$ are bounded, and $\gamma = O(\mu^0)$ from (iii) of Proposition 5.2.

∎

Proposition 5.4 *Consider the quantities* x^{0^k}, s^{0^k}, x^k, *etc. generated by Algorithm 5.1, starting at* (x^{0^1}, s^{0^1}). *Then*

(i) The total variation in (x, s) *due to simplified Newton steps is bounded by* $O(\mu^{0^1})$.

(ii) The sequences $\{(x^{0^k}, s^{0^k})\}$ *and* $\{(x^k, s^k)\}$ *converge to a pair* $(\bar{x}, \bar{s})$ *in the optimal face.*

If the safeguard is activated an infinite number of times *, *then* $(\bar{x}, \bar{s}) = (x^*, s^*)$, *the central optimal pair. Otherwise* $(\bar{x}, \bar{s})$ *is not necessarily equal to* (x^*, s^*).

Proof. (i): Recall that $\mu(x, s) = \gamma\mu(x^0, s^0)$ and apply Proposition 2.1 with $\mu = \gamma\mu^0$ to obtain

$$\hat{u} = \gamma x^0 \phi^0 P_{AX^0\Phi^0} \phi^0 \rho$$

*We shall prove below that this hypothesis is vacuous, but it will be needed to establish a contradiction.

and

$$\hat{v} = \gamma s^0 \phi^0 \tilde{P}_{AX^0\Phi^0} \phi^0 \rho$$

where $\rho = \left(-\frac{xs}{\mu} + e\right)$. Since $\delta(x,s) \leq \frac{1}{2}$, we see that $\|\rho\| = \delta(x,s) \leq \frac{1}{2}$. Moreover, since $\delta(x^0,s^0) = \|w(x^0,s^0) - e\| \leq \frac{1}{2}$ we see that the components of $w(x^0,s^0)$ are contained in $[\frac{1}{2}, \frac{3}{2}]$; hence the components of ϕ^0 are contained in $\left[\sqrt{\frac{2}{3}}, \sqrt{2}\right]$. Also the sequence $\{(x^{0^k}, s^{0^k}\}$ is bounded, and projection operators are bounded. It follows from the above expressions and the fact that all quantities are bounded that $\hat{u} = O(\gamma) = O(\mu^0)$ and $\hat{v} = O(\gamma) = O(\mu^0)$.

(ii): If the safeguard is activated a finite number of times, the conclusion follows from (i), because then the sequences generated by the algorithm are Cauchy sequences. Otherwise, the convergence proof is similar to the proof for the MTY algorithm, presented in Gonzaga and Tapia [3].

We shall prove the result for the primal variables. The proof for the dual slacks is similar. Also, it is enough to prove that $x^{0^k} \to x^*$, since $u^{0^k} = O(\mu^{0^k}) \to 0$.

Assume by contradiction that the sequence $\{x^{0^k}\}$ has an accumulation point $\bar{x} \neq x^*$. Since $\bar{x}_N = x^*_N = 0$, we have

$$\sigma \equiv \|\bar{x}_B - x^*_B\|_* > 0.$$

Let $\mathcal{K} \subset I\!N$ be the set of iterations in which the safeguard is activated (MTY iterations). Our first step is to show that $\bar{x}$ must also be an accumulation point of $(x^{0^k})_{k \in \mathcal{K}}$.

Let $\mathcal{K}_1 \subset I\!N$ be a subsequence such that $x^{0^k} \xrightarrow{\mathcal{K}_1} \bar{x}$, and let $j(k)$ be the first index in $\mathcal{K}$ greater than or equal to k. Then for any $k \in \mathcal{K}_1$, $\|x^{0^{j(k)}} - x^{0^k}\| = O(\mu^{0^k})$ by (i), and thus $x^{0^{j(k)}} \xrightarrow{\mathcal{K}_1} \bar{x}$. Thus it is enough to consider in our assumption subsequences in $\mathcal{K}$.

Let $\{\epsilon^k\}$ be the sequence given by Proposition 4.2, and let $\bar{k}$ be such that for $k \geq \bar{k}$ the conclusions of that proposition are valid and $\epsilon^k < 0.5\sigma$.

Choose an index $j \geq \bar{k}$ with the following characteristics: $j \in \mathcal{K}$, $\|x^{0\,j}_B - x^*_B\|_* < 1.1\sigma$, and the total variation of x due to simplified steps after j satisfies

$$\sum_{\substack{k \notin \mathcal{K} \\ k \geq j}} \|x^{0^{k+1}} - x^{0^k}\|_* < 0.05\sigma. \tag{22}$$

Such an index exists by definition of σ and by (i). We shall prove by induction that for $k \in \mathcal{K}$, $k > j$, $\|x_B^{0^k} - x_B^*\|_* < 0.95\sigma$.

(a) $\|x_B^{0^{j+1}} - x_B^*\|_* < 0.8 \times 1.1\sigma < 0.9\sigma$ by Proposition 4.2. Let $k' = j(j+1)$ be the next index in $\mathcal{K}$. Using (22),

$$\|x_B^{0^{k'}} - x_B^*\|_* \leq \|x_B^{0^{j+1}} - x_B^*\|_* + \|x_B^{0^{k'}} - x_B^{0^{j+1}}\|_* < 0.95\sigma.$$

(b) Assume that for an index $k \in \mathcal{K}$, $k > j$, $\|x_B^{0^k} - x_B^*\|_* < 0.95\sigma$. Then by Proposition 4.2, $\|x_B^{0^{k+1}} - x_B^*\|_* \leq \max\{\epsilon^k, 0.8\|x_B^{0^k} - x_B^*\|_*\} < 0.9\sigma$. As in (a), using(22), let $k' = j(k+1)$ be the next index in $\mathcal{K}$:

$$\|x_B^{0^{k'}} - x^*\|_* \leq \|x_B^{0^{k+1}} - x^*\|_* + \|x_B^{0^{k'}} - x_B^{0^{k+1}}\|_* \leq 0.95\sigma.$$

(a) and (b) prove that for all $k \in \mathcal{K}$, $k > j$, $\|x_B^{0^k} - x_B^*\|_* < 0.95\sigma$, contradicting the fact that σ is an accumulation point of the sequence $(\|x_B^{0^k} - x_B^*\|_*)_{k \in \mathcal{K}}$, and completing the proof. ∎

Having described the behaviour of the large variables, we can now compare the small variables for the exact and simplified Newton corrector steps.

At a typical iteration, the simplified step (u, v) and the exact step $(\hat{u}, \hat{v})$ satisfy the equations below:

$$\begin{aligned} x_B^0 v_B + s_B^0 u_B &= -x_B s_B + \mu e_B \\ x_N^0 v_N + s_N^0 u_N &= -x_N s_N + \mu e_N \end{aligned} \tag{23}$$

$$\begin{aligned} x_B \hat{v}_B + s_B \hat{u}_B &= -x_B s_B + \mu e_B \\ x_N \hat{v}_N + s_N \hat{u}_N &= -x_N s_N + \mu e_N \end{aligned} \tag{24}$$

where $\mu = \gamma\mu^0$, $\gamma = O(\mu^0)$.

Before we state the main result, we establish some relationships within a typical iteration:

(i) (Large variables) Since $u^0 = O(\mu^0)$, $v^0 = O(\mu^0)$ and all components of x_B and s_N are bounded away from zero,

$$x_B^0 = x_B(e + O(\mu^0)) \ , \ s_N^0 = s_N(e + O(\mu^0)) \tag{25}$$

(ii) (Small variables) By construction,

$$\begin{aligned} x^0 s^0 &= \mu^0 w^0 \\ x\ s &= \mu\ w, \end{aligned}$$

where $w_i^0 \in [0.95, 1.05]$, $w_i \in [0.9, 1.1]$, $i = 1, \ldots, n$. Dividing these expressions,

$$\frac{x_N^0}{x_N} = \frac{1}{\gamma}\frac{s_N}{s_N^0}\frac{w_N^0}{w_N} \quad , \quad \frac{s_B^0}{s_B} = \frac{1}{\gamma}\frac{x_B}{x_B^0}\frac{w_B^0}{w_B}.$$

From (25), it is immediate that $s_N/s_N^0 = (e + O(\mu^0))$, and $x_B/x_B^0 = (e + O(\mu^0))$. By a simple calculation, $w_i^0/w_i \in [0.85, 1.17], i = 1, \ldots, n$.

Defining

$$\sigma_N = \frac{s_N}{s_N^0}\frac{w_N^0}{w_N} \quad , \quad \sigma_B = \frac{x_B}{x_B^0}\frac{w_B^0}{w_B},$$

it follows that for sufficiently small μ^0,

$$\sigma_i \in [0.8, 1.2]$$

and we can write

$$x_N^0 = \frac{1}{\gamma}\sigma_N x_N \ , \ s_B^0 = \frac{1}{\gamma}\sigma_B s_B. \tag{26}$$

Proposition 5.5 *Consider an application of Algorithm 5.1. Then the safeguard cannot be activated an infinite number of times.*

Proof. Assume by contradiction that the safeguard is activated at the iterations with indices in an infinite set $\mathcal{K}$.

From Proposition 5.4, the sequences $(x^0, s^0)^k$ and $(x, s)^k$ converge to the analytic center (x^*, s^*) of the optimal face. It follows that

$$\hat{u}^k \to 0 \quad , \quad \hat{v}^k \to 0. \tag{27}$$

Let us substitute the relations (25) and (26) into the Newton equations (23). We shall analyse the first equation (indices in B); the analysis for the other one is similar. Our approach is to compare the behaviour of the small variables in the simplified and exact corrector steps. To begin with

$$(e + O(\mu^0))x_B v_B + \frac{1}{\gamma}\sigma_B s_B u_B \ = \ -x_B s_B + \mu e_B. \tag{28}$$

Subtracting (24) from (28), and restoring the iteration indices,

$$((e + O(\mu^{0^k}))v_B^k - \hat{v}_B^k)x_B^k \;=\; -\left(\frac{1}{\gamma^k}\sigma_B u_B^k - \hat{u}_B^k\right) s_B^k.$$

Taking norms,

$$\|(e + O(\mu^{0^k}))v_B^k - \hat{v}_B^k)x_B^k\| \;\leq\; \|s_B^k\|_\infty \;\left(\|\sigma_B\|_\infty \frac{1}{\gamma^k}\|u_B^k\| \;+\; \|\hat{u}_B^k\|\right).$$

From Proposition 5.3, $\|u_B^k\|/\gamma^k \leq K(\|\hat{u}_B^k\| + \theta_x^k)$, where $\theta_x^k \to 0$. Since $\|\sigma_B\|_\infty \leq 1.2$ for sufficiently large k, and $\|s_B^k\|_\infty = O(\mu^k)$ by Proposition 4.1, the inequality becomes

$$\begin{aligned}\|(e + O(\mu^{0^k}))v_B^k - \hat{v}_B^k)x_B^k\| \;&\leq\; O(\mu^k)(1.2K(\|\hat{u}_B^k\| + \theta_x^k) + \|\hat{u}_B^k\|) \\ &\leq\; K_1\mu^k(\|\hat{u}_B^k\| + \theta_x^k),\end{aligned}$$

where K_1 is a constant that depends on the problem data. Since $\hat{u}_B^k \to 0$ by (27), and since x_B^k has all components bounded away from zero, we conclude that

$$(e + O(\mu^k))\frac{v_B^k - \hat{v}_B^k}{\mu^k} + \frac{O(\mu^k)}{\mu^k}\hat{v}_B^K \to 0$$

and since $\mu^k \to 0$,

$$\frac{v_B^k - \hat{v}_B^k}{\mu^k} \longrightarrow 0 \;\;,\;\; \frac{u_N^k - \hat{u}_N^k}{\mu^k} \longrightarrow 0. \tag{29}$$

The second expression is obtained by a similar process, using the second equation in (23).

Now we shall establish a contradiction. At a typical iteration, let

$$w^+ = \frac{(x+u)(s+v)}{\mu} \;\;,\;\; \hat{w} = \frac{(x+\hat{u})(s+\hat{v})}{\mu}$$

From the analysis of the MTY algorithm presented in Section 4, we see that

$$\|\hat{w} - e\| \leq \frac{\alpha^2}{\sqrt{2}} < 0.01\ .$$

At any iteration $k \in \mathcal{K}$,

$$\|w^+ - e\| > \frac{\alpha}{2} \geq 0.05.$$

At such an iteration, either $\|w_N^+ - e_N\| > 0.02$ or $\|w_B^+ - e_B\| > 0.02$. Assume that at an infinite number of iterations $\mathcal{K}_1 \subset \mathcal{K}$, $\|w_N^+ - e_N\| > 0.02$ (the analysis for the other case is completely similar).

Then for $k \in \mathcal{K}_1$,

$$\|w_N^+ - e_N\| > 0.02 \ , \ \|\hat{w}_N - e_N\| < 0.01$$

This implies that in these iterations.

$$\|w_N^+ - \hat{w}_N\| \ = \ \|(w_N^+ - e_N) - (\hat{w}_N - e_N)\| \geq 0.01 \tag{30}$$

On the other hand, we have by definition,

$$\begin{aligned} \mu w_N^+ &= (x_N + u_N)\,(s_N + v_N) \\ \mu \hat{w}_N &= (x_N + \hat{u}_N)\,(s_N + \hat{v}_N) \end{aligned}$$

Subtracting,

$$\mu(w_N^+ - \hat{w}_N) = (x_N + u_N)\,(s_N + v_N) - (x_N + \hat{u}_N)\,(s_N + \hat{v}_N).$$

Reordering terms in this expression, we obtain

$$w_N^+ - \hat{w}_N = \frac{u_N - \hat{u}_N}{\mu}(s_N + v_N) + \frac{x_N + \hat{u}_N}{\mu}(v_N - \hat{v}_N).$$

Let us analyse the terms in the right-hand side (restoring the index k):

(i) By (29), $\dfrac{u_N^k - \hat{u}_N^k}{\mu^k}(s_N^k + v_N^k) \longrightarrow 0.$

(ii) By Proposition 4.1, $x_N^k = O(\mu^k)$ and $\hat{u}_N^k = O(\mu^k)$.

From (27), $\hat{v}_N^k \to 0$. From Proposition 5.3, $v_N^k \to 0$. Hence

$$\|\frac{x_N^k + u_N^k}{\mu^k}(v_N^k - \hat{v}_N^k)\| \leq K_2 \|v_N^k - \hat{v}_N^k\|,$$

where K_2 depends on problem data, and so this term converges to zero.

We conclude that $(w_N^+)^k - \hat{w}_N^k \longrightarrow 0$, contradicting (30), and completing the proof. ∎

We are now ready to formally state our convergence results.

Theorem 5.1 *Let $\{(x^0, s^0)^k\}$ and $\{(x, s)^k\}$ denote the sequences generated by the simplified MTY predictor-corrector algorithm. Then*

(i) The safeguard in the corrector step is activated only a finite number of times.

(ii) The algorithm has iteration complexity $O(\sqrt{n}L)$.

(iii) The duality-gap sequence $\{x^{0T} s^0\}$ converges quadratically to zero.

(iv) Both sequences $\{(x^0, s^0)\}$ and $\{(x, s)\}$ converge to a point $(\bar{x}, \bar{s})$ in the optimal face.

Proof. Property (i) follows from Proposition 5.5. Also (ii) follows from (iv) of Proposition 5.2 in a standard manner. See Mizuno, Todd, and Ye [10] for details. Property (iii) is a combination of (i), (ii), and (iii) of Proposition 5.2. Finally (iv) is (ii) of Proposition 5.4. ∎

6 Concluding Remarks

The fact that so much of Theorem 5.1 follows from Proposition 5.2 and Proposition 5.2 depends so little on the corrector step leads us to take a closer look at the role of the corrector step in our convergence theory.

Consider a typical simplified MTY predictor-corrector iteration represented by $\{(x^0, s^0), (x, s), (x^+, s^+)\}$. The predictor step takes (x^0, s^0) to (x, s) and the corrector step takes (x, s) to (x^+, s^+). A close look at the derivation

of our theory shows that for the establishment of $O(\sqrt{n}L)$ complexity and quadratic convergence we only used the fact that the corrector step satisfies

$$\begin{array}{lll} & \text{(i)} & x^{+T}s^{+} \leq x^{T}s \\ \text{and} & & \\ & \text{(ii)} & \delta(x^{+}, s^{+}) \leq \alpha/2. \end{array} \tag{31}$$

Hence any corrector step satisfying (34) will lead to $O(\sqrt{n}L)$ complexity and quadratic convergence, but not necessarily iteration sequence convergence. It follows that quadratic convergence is the best that should be expected from either the MTY algorithm or the simplified MTY predictor-corrector algorithm. This is because for both these algorithms the corrector step does not improve the duality-gap, i.e. $x^{+T}s^{+} = x^{T}s$, and therefore the quadratic decrease is obtained entirely from the damped Newton predictor step, and quadratic decrease (in general) is optimal for a (damped) Newton method. Clearly the same is true for any corrector step that does not decrease the duality-gap.

We are accustomed to expect cubic decrease from the pair consisting of a Newton step and a simplified Newton step and quartic decrease from the pair consisting of two Newton steps. In order to attain these objectives along with $O(\sqrt{n}L)$ complexity the predictor-corrector approach will have to be modified so that the corrector step still satisfies (34) but also gives the appropriate decrease in the duality-gap. For example if in the simplified corrector step of Algorithm 5.1 we replace μ with $\gamma\mu$ and the safeguard is activated only a finite number of times, then we would obtain cubic convergence from the simplified MTY algorithm. We did not pursue this issue in the present work.

The contribution of this paper is the demonstration that in the MTY predictor-corrector algorithm the Newton corrector step can be replaced with a safeguarded simplified Newton corrector step and all the algorithmic properties are maintained, except that the convergence of the iteration sequence is no longer to the analytic center. Whether this loss is important or not clearly depends on the application.

Acknowledgement. The authors acknowledge two anonymous referees and Mike Todd for finding mistakes and making numerous comments and suggestions that led to improvements in the original presentation. They are particularly grateful to one of these referees for an unusual level of detail in reading and understanding.

References

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

[2] C. GONZAGA, *Path following algorithms for linear programming*, SIAM Review 34 (1992), pp.167-224.

[3] C. GONZAGA AND R. A. TAPIA, *On the convergence of the Mizuno-Todd-Ye algorithm to the analytic center of the solution set.* Technical Report TR92-41, Department of Computational and Applied Mathematics, Rice University (1992).

[4] M. KOJIMA, S. MIZUNO AND A. YOSHISE, A primal-dual interior-point method for linear programming. In Nimrod Megiddo, editor, *Progress in Mathematical Programming, Interior-point and Related Methods*, pages 29-47, Springer-Verlag, New York (1989).

[5] I. J. LUSTIG, R. E. MARSTEN AND D. F. SHANNO, *Computational experience with a primal-dual interior-point method for linear programming*, Linear Algebra and Its Applications 152 (1991), pp. 191-222.

[6] K. MCSHANE, *A superlinearly convergent $O(\sqrt{n}L)$ iteration primal-dual linear programming algorithm.* Manuscript, 2537 Villanova Drive, Vienna, VA (1991).

[7] N. MEGIDDO, Pathways to the optimal set in linear programming. In Nimrod Megiddo, editor, *Progress in Mathematical Programming, Interior-point and Related Methods*, pages 131-158, Springer-Verlag, New York (1989).

[8] N. MEGIDDO AND M. SHUB, *Boundary behaviour of interior point algorithms in linear programming*, Mathematics of Operations Research 14 (1989), pp. 97–146.

[9] S. MEHROTRA, *Quadratic convergence in a primal-dual method.* Technical Report 91-15, Department of Industrial Engineering and Management Science, Northwestern University (1991).

[10] S. MIZUNO, M. J. TODD AND Y. YE, *On adaptive-step primal-dual interior-point algorithms for linear programming*, Technical Report No. 944, School of ORIE, Cornell University, Ithaca, New York (1990). To appear in Mathematics of Operations Research.

[11] R. C. MONTEIRO AND I. ADLER, *Interior path-following primal-dual algorithm. Part I: linear programming*, Mathematical Programming 44 (1989), pp. 27-41.

[12] R. A. TAPIA, Y. ZHANG AND Y. YE, *On the convergence of the iteration sequence in primal-dual interior point methods*, Technical Report TR91-24, Department of Computational and Applied Mathematics, Rice University (1991).

[13] M. J. TODD AND Y. YE, *A centered projective algorithm for linear programming*, Math. of O.R. 15 (1990), pp. 508-529.

[14] Y. YE, *On the Q-order of convergence of interior-point algorithms for linear programming*, Proceedings of the 1992 Symposium on Applied Mathematics, Institute of Applied Mathematics, Chinese Academy of Sciences, pp. 534-543, 1992.

[15] 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), pp. 151-162.

[16] Y. YE, R. A. TAPIA, AND Y. ZHANG, *A superlinearly convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming*, Technical Report TR91-22, Department of Computational and Applied Mathematics, Rice University (1991).

[17] Y. ZHANG AND R. A. TAPIA, *A superlinearly convergent polynomial primal-dual interior-point algorithm for linear programming*, SIAM Journal on Optimization 3(1993), pp. 118-131.

[18] Y. ZHANG AND R. A. TAPIA, *Superlinear and quadratic convergence of primal-dual interior-point methods for linear programming revisited*, Journal of Optimization Theory and Applications 73 (1992), pp. 229-242.

[19] Y. ZHANG, R. A. TAPIA, AND J. E. DENNIS, *On the superlinear and quadratic convergence of primal-dual interior-point linear point algorithms*, SIAM Journal on Optimization 2 (1992), pp. 303-324.