On the Convergence of the Mizuno-Todd-Ye Algorithm to the Analytic Center of the Solution Set

Clovis C. Gonzaga
Richard A. Tapia

CRPC-TR92292
June 1992

Center for Research on Parallel Computation
Rice University
P.O. Box 1892
Houston, TX 77251-1892

Revised November 1992.

On the Convergence of the Mizuno-Todd-Ye Algorithm to the Analytic Center of the Solution Set*

Clovis C. Gonzaga[†] Richard A. Tapia [‡]

June, 1992
(revised November 1992)

Keywords: Linear programming, primal-dual interior-point algorithm, predictor-corrector algorithm, analytic center

Abbreviated title: 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

In this work we demonstrate that the Mizuno-Todd-Ye predictor-corrector primal-dual interior-point method for linear programming generates iteration sequences that converge to the analytic center of the solution set.

1 Introduction and Preliminaries

The basic primal-dual interior-point method for linear programming was originally proposed by Kojima, Mizuno, and Yoshise [6] based on earlier work of Megiddo [10]. This algorithm can be viewed as perturbed (centered) and damped Newton's method applied to the first order conditions for a particular standard form linear program. They established linear convergence of the duality gap sequence to zero and an iteration complexity of $O(nL)$ for their basic algorithm. Soon after Mizuno, Todd and Ye [13] 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. Mizuno, Todd, and Ye also established linear convergence of the duality gap sequence to zero; however they established a superior iteration complexity of $O(\sqrt{n}L)$ for their predictor-corrector algorithm.

The literature now abounds with papers concerned with issues related to primal-dual interior-point methods. Moreover, when we discuss convergence or convergence attributes (including complexity) of one of these algorithms we are in general discussing convergence of the duality gap (or some other measure of residual error) to zero. This interpretation has become standard in the 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. Indeed, the earlier works on fast (superlinear) convergence of the duality gap sequence to zero, i.e., Zhang, Tapia, and Dennis [25], Zhang, Tapia and Potra [26], Zhang and Tapia [22], Ye, Tapia, and Zhang [20], and McShane [9], all made the assumption that the iteration sequence converged.

In some applications, e.g. see Charnes, Cooper, and Thrall [2], it is important to obtain a solution that is not near the boundary of the solution set. Hence there is significant value in designing a primal-dual interior-point method for linear programming that converges to the analytic center of the solution set.

Tapia, Zhang, and Ye [16] derived conditions under which the iteration sequence generated by the Kojima-Mizuno-Yoshise primal-dual interior-point method converged. These conditions were essentially the conditions for fast (superlinear) convergence established by Zhang, Tapia, and Dennis [25] (see also Zhang and Tapia [23]). Zhang and Tapia [24] derived conditions under which this iteration sequence converged to the analytic center, assuming that the sequence converged. However, these conditions are not completely compatible with the Tapia-Zhang-Ye conditions for the convergence of the iteration sequence.

Ye, Güler, Tapia, and Zhang [19], and independently Mehrotra [12], based on the work of Ye, Tapia, and Zhang [20], demonstrated that the Mizuno-Todd-Ye predictor-corrector algorithm in all cases gives quadratic convergence of the duality gap sequence to zero. A highlight of this contribution was that the assumption of iteration sequence convergence was not needed (for the first time). Soon after Zhang and Tapia [23] removed this assumption from the Zhang-Tapia-Dennis theory for superlinear convergence. Quite recently Zhang and El-Bakry [21] were able to show that a modified version of the Mizuno-Todd-Ye predictor-corrector algorithm had the property that the iteration sequence that it generated converged to the analytic center. Their modified algorithm dynamically chose the steplength in the Newton predictor step so that the corrector step would asymptotically enforce arbitrary close proximity to the central path.

In this paper we show that the predictor-corrector algorithm as originally stated by Mizuno, Todd, and Ye has the property that the iteration sequences (predictor-step sequence and corrector-step sequence) it generates converge to the analytic center of the solution set.

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 establish some properties concerning this step. Some mathematical tools concerning projections and scalings are derived in Section 3. Central path issues are discussed in Section 4. The Mizuno-Todd-Ye predictor-corrector algorithm and some of its properties are presented in Section 5. In Section 6 we combine all our previous discussion and in Theorem 6.1 demonstrate that the Mizuno-Todd-Ye algorithm generates sequences that converge to the analytic center of the solution set.

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 have non-empty relative interiors.

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

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

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 $F = (F_P, F_D)$ 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 $\{N, B\}$ 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, $x_B > 0$ 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.

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 and only 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 $I\!R^{2n}$ defined on the positive reals by

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

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

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

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's method applied to the perturbed KKT conditions (1) or (2).

2 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 subscript 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.

Given a strictly feasible pair (x, s), we shall define three parameters:

$$\begin{aligned} \mu(x,s) &= s^T x/n, \\ w(x,s) &= sx/\mu(x,s), \\ \phi(x,s) &= 1/\sqrt{w(x,s)}. \end{aligned}$$

The first two parameters will be extensively studied below. The parameter ϕ has no special meaning, and is introduced because it will simplify many formulas in the text. When no confusion can arise, we drop the reference to the variables, and continue to use other symbols in a consistent manner. For example $\bar{w} = w(\bar{x}, \bar{s})$ or $\phi^0 = \phi(x^0, s^0)$.

Given a strictly feasible pair (x, s), we are interested in finding $(x^+, s^+) = (x, s) + (u, v)$ that solves (1) or (2) with $\mu = \gamma\mu(x, s)$, where $\gamma \in [0, 1]$. The Newton equation for (1) at (x, s) with μ replaced by $\gamma\mu$ can be written

$$\begin{aligned} xv + su &= -xs + \gamma\mu(x, s)e \\ u &\in \mathcal{N}(A) \\ v &\in \mathcal{R}(A^T). \end{aligned} \tag{3}$$

where as usual $\mathcal{N}$ denotes null space and $\mathcal{R}$ denotes range space. The solution of (3) is obtained by scaling the equations. Define the scaling matrix by $d = \sqrt{x/s}$, $D = \mathrm{diag}(d_1, \ldots, d_n)$, and the scaling

$$(\bar{x}, \bar{s}) = (d^{-1}x, ds).$$

The relationship between d and the vector ϕ defined above is

$$d = \sqrt{\frac{x}{s}} = \frac{x\phi}{\sqrt{\mu}} = \frac{\sqrt{\mu}}{s\phi}. \tag{4}$$

When applied to the original pair (x, s), the resulting scaled pair will be

$$(\bar{x}, \bar{s}) = (\sqrt{xs}, \sqrt{xs}). \tag{5}$$

After scaling, the system (3) becomes

$$\begin{aligned} \bar{x}\bar{v} + \bar{s}\bar{u} &= -\bar{x}\bar{s} + \gamma\mu e \\ \bar{u} &\in \mathcal{N}(AD) \\ \bar{v} &\in \mathcal{R}(DA^T). \end{aligned} \tag{6}$$

Since $\bar{x} > 0$, the first equation can be multiplied by $\bar{x}^{-1}$, leading to

$$\bar{v} + \bar{u} = -\bar{s} + \gamma\mu\bar{x}^{-1},$$

and the solution is simply the orthogonal decomposition of the vector $-\bar{s} + \gamma\mu\bar{x}^{-1}$ along $\mathcal{N}(AD)$ and its orthogonal complement. Let P_{AD} be the projection matrix into $\mathcal{N}(AD)$, and $\tilde{P}_{AD} = I - P_{AD}$:

$$\begin{aligned} \bar{u} &= P_{AD}(-\bar{s} + \gamma\mu\bar{x}^{-1}) \\ \bar{v} &= \tilde{P}_{AD}(-\bar{s} + \gamma\mu\bar{x}^{-1}). \end{aligned} \tag{7}$$

The Newton step in original coordinates is given by $v = d^{-1}\bar{v}$ and $u = d\bar{u}$.

A convenient formulation is obtained by substituting $d = \frac{1}{\sqrt{\mu}} x\phi$ and $d^{-1} = \frac{1}{\sqrt{\mu}}\frac{s}{\phi}$.

$$\begin{aligned} u &= x\phi P_{AX\Phi}\phi\left(-\frac{xs}{\mu} + \gamma e\right) \\ v &= \frac{s}{\phi}\tilde{P}_{AX\Phi}\phi\left(-\frac{xs}{\mu} + \gamma e\right) \end{aligned} \tag{8}$$

We now describe two alternative ways of writing the expression for u (the expressions for v are similar).

Using the definition of w,

$$u = -x\phi P_{AX\Phi}\phi(w - \gamma e), \tag{9}$$

Observing the symmetrical formulation of (LD), we see that for any two feasible dual slacks s^1, s^2, $P_{AD}ds^1 = P_{AD}ds^2 = P_{AD}dc$. In particular, we can choose a fixed dual slack and use it in (8). We shall choose s^*, the analytic center of the dual optimal face, and write

$$u = -dP_{AD}d(s^* - \gamma\mu x^{-1}).$$

By the same process as above,

$$u = -x\phi P_{AX\Phi}\phi\left(\frac{xs^*}{\mu} - \gamma e\right). \tag{10}$$

The original pair satisfies $x^T s = n\mu$, from the definition of $\mu = \mu(x,s)$. The new duality gap is

$$(x+u)^T(s+v) = x^T s + x^T v + s^T u + v^T u.$$

But $v^T u = 0$, and multiplying the first equation in (3) by e^T, we get

$$x^T s + x^T v + s^T u = n\gamma\mu.$$

It follows that

$$(x+u)^T(s+v) = n\gamma\mu, \tag{11}$$

or still,

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

Two special cases of problem (3) have been studied extensively in the literature. They are
(i) $\gamma = 0$: The resulting directions (h_x^1, h_s^1) are called the primal-dual affine scaling directions.
(ii) $\gamma = 1$: The resulting directions (h_x^2, h_s^2) are called the constant gap centering directions.

The first equation of the Newton system (3) can be rewritten as

$$xu + su = -(1-\gamma)xs + \gamma(-xs + \mu e). \tag{12}$$

This is a combination of the solutions of two systems with

$$\begin{aligned} xu^1 + su^1 &= -xs \\ xu^2 + su^2 &= -xs + \mu e, \end{aligned} \tag{13}$$

where $\mu = \mu(x,s)$. The complete solution is given by

$$(u,v) = (1-\gamma)(u^1, v^1) + \gamma(u^2, v^2). \tag{14}$$

It is quite common to use these two directions separately, possibly as a way to simplify the analysis. This is done by the predictor-corrector algorithms that we study in this paper.

3 Mathematical Tools

In this section we state some lemmas on projections and scalings that will be useful in the analysis below.

3.1 Properties of Scaled Projections

In this subsection we slightly extend results published by Megiddo and Shub [11].

Consider the primal feasible set for (LP),

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

and the map

$$d \in I\!R^n_+, d \neq 0; \rho \in I\!R^n \mapsto h(d,\rho) = P_{AD}\rho, \tag{15}$$

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 other 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 these results conveniently extended to fit our needs.

By definition of projection, $h(d, \rho)$ solves the problem

$$\begin{array}{rl} \text{minimize} & \|h_N - \rho_N\|^2 + \|h_B - \rho_B\|^2 \\ \text{subject to} & A_B D_B h_B = -A_N D_N h_N. \end{array} \tag{16}$$

Assume now that $h_N(d, \rho)$ is given. Then $h_B(d, \rho)$ solves

$$\begin{array}{rl} \text{minimize} & \|h_B - \rho_B\| \\ \text{subject to} & A_B D_B h_B = -A_N D_N h_N(d, \rho). \end{array} \tag{17}$$

We shall study the point to set mapping

$$d \in I\!R^n_+, \rho \in I\!R^n \mapsto \theta(d, \rho) = \{h_B \in I\!R^{|B|} \mid A_B D_B h_B = -A_N D_N h_N(d, \rho)\}, \tag{18}$$

near a pair $\bar{d}, \bar{\rho}$ as described above. Note that at this point, $\theta(\bar{d}, \bar{\rho}) = \mathcal{N}(A_B \bar{D}_B)$.

Lemma 3.1 *The map defined above is continuous at $(\bar{d}, \bar{\rho})$*

Proof.

(i) Upper semi-continuity: consider sequences $d^k \to \bar{d}$, $\rho^k \to \bar{\rho}$ and (h_B^k) such that $A_B D_B^k h_B^k = -A_N D_N^k h_N(d^k, \rho^k)$ and (h_B^k) converges to some point $\bar{h}_B$. We must prove that $A_B \bar{D}_B \bar{h}_B = 0$.

The sequence $h_N(d^k, \rho^k)$ is bounded, because $d_N \to 0$ and $\|h_N(d^k, \rho^k)\| \leq \|\rho^k\|$, since $h(d^k, \rho^k)$ is a projection. Hence $A_B D_B^k h_B^k \to 0$ and consequently $A_B \bar{D}_B \bar{h}_B = 0$, completing this part of the proof.

(ii) Lower semi-continuity: Consider now an arbitrary point $\bar{h}_B \in \mathcal{N}(A_B \bar{D}_B)$. Given arbitrary feasible sequences $d^k \to \bar{d}$, $\rho^k \to \bar{\rho}$, we must construct (h_B^k) such that $A_B D_B^k h_B^k = -A_N D_N^k h_N(d^k, \rho^k)$ and $h_B^k \to \bar{h}_B$.

Consider arbitrary sequences $d^k \to \bar{d}$, $\rho^k \to \bar{\rho}$ such that $d^k > 0$ for $k = 1, 2, \ldots$ and define $h_N^k = h_N(d^k, \rho^k)$. For each k let $\tilde{h}_B^k$ be a minimum-norm solution of $A_B D_B^k h_B = -A_N D_N^k h_N^k$. Then $\tilde{h}_B^k \to 0$, since $D_N^k h_N^k \to 0$. Construct

$$h_B^k = (D_B^k)^{-1} \bar{D}_B \bar{h}_B + \tilde{h}_B^k. \tag{19}$$

Then

$$A_B D_B^k h_B^k = A_B \bar{D}_B \bar{h}_B + A_B D_B^k \tilde{h}_B^k = -A_N D_N^k h_N^k,$$

since $\bar{h}_B \in \mathcal{N}(A_B \bar{D}_B)$. Thus $h_B^k \in \theta(d^k, \rho^k)$. Since $D_B^k \to \bar{D}_B > 0$ and $\tilde{h}_B^k \to 0$, it follows that $h_B^k \to \bar{h}_B$, completing the proof. ∎

Lemma 3.2 *Consider the map $d, \rho \mapsto h(d, \rho)$ and the points $\bar{d}$ and $\bar{\rho}$ defined above. Then when $d \in I\!R_+^n$, $d \to \bar{d}$ and $\rho \to \bar{\rho}$,*
(i) $h_B(d, \rho) \to P_{A_B \bar{D}_B} \bar{\rho}_B$.
(ii) If $\bar{\rho}_N = 0$ then $h_N(d, \rho) \to 0$.

Proof.

(i) The map $d, \rho \mapsto \min\{\|h_B - \rho_B\| \mid h_B \in \theta(d, \rho)\}$ is continuous at $\bar{d}, \bar{\rho}$ as a consequence of the continuity of the map θ (see for example Hogan [4]). It follows that $\|h_B(d, \rho) - \rho_B\| \to \|\bar{h}_B - \bar{\rho}_B\|$, where $\bar{h}_B = h_B(\bar{d}, \bar{\rho}) = P_{A_B \bar{D}_B} \bar{\rho}_B$. Since $\bar{h}_B$ is the unique minimizer of $\|h_B - \bar{\rho}_B\|$ in $\theta(\bar{d}, \bar{\rho})$, we must have (i).

(ii) Here we follow a similar proof in Megiddo and Shub [11]. Assume that $\bar{\rho}_N = 0$ and by contradiction that for some sequence $d^k \to \bar{d}$, $\rho^k \to \bar{\rho}$ we have $h_N(d^k, \rho^k) \to \bar{h}_N \neq 0$. Define $\epsilon = \|\bar{h}_N\|^2 > 0$. We have:

$$\|h(d^k, \rho^k) - \rho^k\|^2 = \|h_B(d^k, \rho^k) - \rho_B^k\|^2 + \|h_N(d^k, \rho^k) - \rho_N^k\|^2.$$

By (i), $h_B(d^k, \rho^k) \to \bar{h}_B$, where $\bar{h}_B = P_{A_B \bar{D}_B} \bar{\rho}_B$. For sufficiently large k,

$$\|h_B(d^k, \rho^k) - \rho_B^k\|^2 < \|\bar{h}_B - \bar{\rho}_B\|^2 + \epsilon/2. \tag{20}$$

Now construct the following sequence:

$$\tilde{h}_B^k = (D_B^k)^{-1} \bar{D}_B \bar{h}_B \quad , \quad \tilde{h}_N = 0.$$

It follows that $\tilde{h}_B^k \to \bar{h}_B$, and $\tilde{h}^k \in \mathcal{N}(AD^k)$, since $AD^k \tilde{h}^k = A_B \bar{D}_B \bar{h}_B = 0$.

Comparing this with (20), for k sufficiently large $\|\tilde{h}^k - \rho^k\| < \|h(d^k, \rho^k) - \rho^k\|$ and $\tilde{h}^k \in \mathcal{N}(AD^k)$, contradicting the definition of $h(d^k, \rho^k) = P_{AD^k} \rho^k$ and completing the proof. ∎

3.2 Shifted Scalings

This subsection contains some useful consequences of scalings on projections and norms. The first lemma concerns projections and slightly shifted scalings.

Lemma 3.3 *Let $q \in \mathbb{R}^n$ be such that $\|q - e\|_\infty \leq \alpha$, where $\alpha \in (0, 0.25)$, and consider the projections $\hat{h} = P_A \rho$, $h = qP_{AQ}q\rho$. Then $\|h - \hat{h}\| \leq 3\alpha\|\hat{h}\|$.*

Proof. Note that since $\rho = \hat{h} + A^T w$ for some $w \in \mathbb{R}^m$,

$$q\rho = q\hat{h} + (AQ)^T w$$

and thus

$$P_{AQ}q\rho = P_{AQ}q\hat{h}$$

It follows that

$$q^{-1}h = P_{AQ}q\hat{h}$$

On the other hand, by definition of projection,

$$q\hat{h} = P_{AQ}q\hat{h} + y,$$

where $y \in \mathcal{R}(QA^T)$. Merging the last expressions,

$$q\hat{h} = q^{-1}h + y,$$

where $q^{-1}h \in \mathcal{N}(AQ)$ and $y \in \mathcal{R}(QA^T)$. Subtracting $q^{-1}\hat{h} \in \mathcal{N}(AQ)$ from both sides,

$$(q^{-1} - q)\hat{h} = q^{-1}(h - \hat{h}) + y,$$

and from the orthogonality of the right-hand side terms,

$$\|(q^{-1} - q)\hat{h}\| \geq \|q^{-1}(h - \hat{h})\|.$$

Now use the following facts: $\|(h - \hat{h})\| \leq \|q\|_\infty \|q^{-1}(h - \hat{h})\|$ and $\|(q^{-1} - q)\hat{h}\| \leq \|(q^{-1} - q)\|_\infty \|\hat{h}\|$. Combining these three expressions leads to

$$\|h - \hat{h}\| \leq \|q\|_\infty \|q^{-1} - q\|_\infty \|\hat{h}\|.$$

But $\|q\|_\infty \|q^{-1} - q\|_\infty \leq (1 + \alpha)\left(\frac{1}{1-\alpha} - (1 - \alpha)\right) \leq 3\alpha$ which is easily verified for $\alpha \in (0, 0.25)$, completing the proof. ∎

Our second lemma concerns scaled norms. Given a vector $x \in \mathbb{R}^n_{++}$, the following map defines a norm:

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

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, because it characterizes the proximity from a point to a central point in the following sense: let $x(\mu)$ be the primal central point associated with the parameter $\mu > 0$. If $\|x - x(\mu)\|_x \leq \delta < 1$ then a Newton centering iteration from x produces an efficient centering step (which is usually imprecisely stated as being in the region of quadratic convergence of Newton's method).

The following lemma relates the scaled norms for different reference points.

Lemma 3.4 *Consider $x, y \in \mathbb{R}^n_{++}$, $h \in \mathbb{R}^n$, $\alpha \in (0,1)$. If either $\|x-y\|_x^\infty \leq \alpha$ or $\|x - y\|_y^\infty \leq \alpha$, then*

$$\|h\|_x \leq \frac{1}{1-\alpha}\|h\|_y$$

Proof. To begin with

$$\|h\|_x = \left\|\frac{h}{x}\right\| = \left\|\frac{y}{x}\frac{h}{y}\right\| \leq \left\|\frac{y}{x}\right\|_\infty \|h\|_y.$$

If $\|x - y\|_x^\infty \leq \alpha$, then $|(x_i - y_i)/x_i| \leq \alpha$, or $1 - y_i/x_i \geq \alpha$, which implies $y_i/x_i \leq 1+\alpha \leq 1/(1-\alpha)$. In the other case, $|(x_i-y_i)/y_i| \leq \alpha$, or $x_i/y_i \geq 1-\alpha$, which implies $y_i/x_i \leq 1/(1-\alpha)$, completing the proof. ∎

4 Trajectories, Centrality and Proximity

The primal-dual central path defined above crosses the set of interior points and ends at a point (x^*, s^*) in the relative interior of the optimal face. This point is the analytic center of the face. See problem (24) for an equivalent characterization.

In this section we study (primal-dual) proximity criteria that describe how far a pair (x, s) is from the primal-dual central path, then study (primal) proximity criteria to evaluate how far a point in the optimal face is from its analytic center.

4.1 Primal-Dual Proximity

Given an interior pair (x, s) and a parameter $\mu > 0$ (not necessarily equal to $\mu(x, s)$), the proximity of (x, s) in relation to $(x(\mu), s(\mu))$ is measured by

$$\delta(x, s, \mu) = \left\| \frac{xs}{\mu} - e \right\| . \tag{21}$$

When $\mu = \mu(x, s)$, this is the proximity with relation to the central path,

$$\delta(x, s) = \left\| \frac{xs}{\mu(x, s)} - e \right\| = \|w(x, s) - e\| . \tag{22}$$

Let us compute the proximity at the pair (x^+, s^+) resulting from the Newton step described in (3), with $\mu = \mu(x, s)$. We have

$$\begin{aligned} x^+ s^+ &= (x + u)(s + v) \\ &= xs + xv + su + uv \\ &= \gamma\mu e + uv. \end{aligned}$$

But $\mu(x^+, s^+) = \gamma\mu$ from (11), and thus

$$\frac{x^+ s^+}{\mu(x^+, s^+)} - e = \frac{uv}{\mu(x^+, s^+)},$$

or

$$\delta(x^+, s^+) = \left\| \frac{uv}{\gamma\mu} \right\| = \left\| \frac{uv}{\mu(x^+, s^+)} \right\| . \tag{23}$$

A fundamental result on the effect of the Newton step on proximity is given in the following lemma. This result is due to Mizuno, Todd, and Ye and can be found in [13].

Lemma 4.1 *Consider an interior pair (x, s) and a parameter $\mu^+ > 0$. If $\delta(x, s, \mu^+) = \delta \leq 0.5$, then $\delta(x^+, s^+) \leq \delta^2/\sqrt{2}$.*

The primal-dual affine-scaling directions are the solution of (3) with $\gamma = 0$. These directions associated with each interior feasible pair (x, s) generate a continuous vector field, which extends continuously to the boundary.

This vector field was thoroughly studied by Adler and Monteiro [1], who describe the trajectories generated by it and the derivatives of these trajectories. The trajectories are parameterized bu μ, and there is one trajectory passing through each interior pair (x, s).

For each interior pair (x, s), we defined the vector $w(x, s) = xs/\mu(x, s)$. Each trajectory is associated with this vector in the following two ways:

(i) The trajectory associated with $w > 0$ is composed of the pairs (x, s) such that

$$\frac{xs}{\mu(x,s)} = w.$$

In particular, the central path is the trajectory associated with $w = e$.

(ii) The trajectory associated with $w > 0$ is composed of the minimizer pairs of the parameterized primal-dual penalized function

$$x^T s - \mu \sum_{i=1}^{n} w_i \ln x_i - \mu \sum_{i=1}^{n} w_i \ln s_i.$$

Each trajectory is composed of interior points, and ends in the relative interior of the optimal face.

¿From here on, we assume that the vectors $w(x, s)$ are always in a compact set defined by

$$\|w(x,s) - e\| \leq \alpha,$$

where $\alpha \in (0, 1)$.

When the weight vectors w are in a compact set bounded away from the boundary of the positive orthant, the trajectories end in the relative interior of the optimal face. Specifically at the minimizers of the parameterized barrier function,

$$\begin{aligned} x^*(w) &= \text{argmin } \{-\sum_{i \in B} w_i \ln x_i \mid x \in F_P\} \\ s^*(w) &= \text{argmin } \{-\sum_{i \in N} w_i \ln s_i \mid x \in F_D\}. \end{aligned}$$

In particular, the central path ends at the analytic center of the optimal face $(x^*, s^*) = (x^*(e), s^*(e))$.

The sets of end points of all trajectories for such weights w are sets of minimizers of parameterized continuously differentiable functions, and are compact. It is easy to see that the nonzero variables are all bounded away from zero, because the compact sets are in the relative interior of the optimal faces. This is also clear from the fact that the barrier functions become arbitrarily large as the boundaries of the faces are approached.

Similarly, all the trajectories in the bundle associated with this compact set of parameter vectors are in the relative interior of the feasible set, and bounded away from the non-optimal faces.

4.2 Primal Proximity

We shall summarize some facts about the analytic center of a polytope, and derive properties of descent methods for finding the center.

Consider the primal centering problem

$$\begin{array}{lrcl} \text{minimize} & p(x) & = & -\sum_{i=1}^{n} \ln(x_i) \\ \text{subject to} & Ax & = & b \\ & x & > & 0, \end{array} \tag{24}$$

where $b \in I\!R^m$, $A \in I\!R^{m\times n}$, such that S, the closure of the feasible set is compact with a nonempty set S^0 of interior points. The analytic center of S is the unique optimal solution of (24),

$$\chi = \operatorname*{argmin}_{x\in S^0} p(x).$$

The analytic center was defined by Sonnevend [15]; see also McLinden [8]. Its properties and the description of the Newton primal centering algorithm (SSD algorithm) are described in Gonzaga [3]. The following facts come from this latter reference.

Given a point $x \in S^0$, the Newton centering direction from x is given by $h(x) = x\bar{h}(x)$, where

$$\bar{h}(x) = -P_{AX}e$$

is the centering direction after scaling the problem so that the point x is taken to e.

The (primal) proximity of x in relation to χ, defined above, is given by

$$\delta(x) = \|\bar{h}(x)\| = \|h(x)\|_x, \tag{25}$$

where $\|\cdot\|_x$ is the norm relative to x.

The following important results are described for example in [3]. Let $x \in S^0$ be such that $\delta(x) = \delta < 1$, then

$$\begin{aligned} \|x - \chi\|_x &\leq \frac{\delta}{1-\delta}. \\ \delta(x + h(x)) &\leq \delta^2. \end{aligned} \tag{26}$$

The first result above gives an upper bound for $\|x - \chi\|_x$. We shall also need a lower bound for this distance, and this will be provided by the next lemma.

Lemma 4.2 *If $\delta(x) = \delta < 0.5$, then*

$$\|x - \chi\|_x \geq \frac{1-2\delta}{1-\delta}\delta.$$

In particular, if $\delta \leq 0.09$, then $\|x - \chi\|_x \in [0.9\delta, 1.1\delta]$.

Proof. Let $x^+ = x + h(x)$. We know that $\|h(x)\|_x = \delta$, and that $\delta(x^+) \leq \delta^2$. It follows from (26) that

$$\|x^+ - \chi\|_{x^+} \leq \frac{\delta^2}{1-\delta^2},$$

and hence

$$\|x^+ - \chi\|_x \leq \left\|\frac{x^+}{x}\right\|_\infty \frac{\delta^2}{1-\delta^2}.$$

But $x^+/x = e + h(x)/x$, and thus

$$\left\|\frac{x^+}{x}\right\|_\infty \leq 1 + \left\|\frac{h(x)}{x}\right\| \leq 1 + \delta.$$

It follows that

$$\|x^+ - \chi\|_x \leq (1+\delta)\frac{\delta^2}{1-\delta^2} = \frac{\delta^2}{1-\delta}.$$

Finally,

$$\begin{aligned} \|x - \chi\|_x &= \|x - x^+ + x^+ - \chi\|_x \\ &\geq \|x - x^+\|_x - \|x^+ - \chi\|_x \\ &\geq \delta - \frac{\delta^2}{1-\delta} \\ &= \frac{1-2\delta}{1-\delta}\delta. \end{aligned}$$

The numeric values are obtained by substitution, completing the proof. ∎

This lemma shows that when the proximity measure is small, it is indeed a good approximation to the actual scaled distance to the center. The values $\delta \le 0.09$ will be quite reasonable for our analysis below.

One final technical result also will be useful below. It reproduces the bounds above using the norm relative to χ.

Lemma 4.3 *If $\delta(x) = \delta \le 0.1$, then for $x^+ = x + h(x)$,*

$$\begin{aligned} \|x^+ - \chi\|_\chi &\le 1.05\delta^2 \\ \|x - \chi\|_\chi &\ge 0.75\delta. \end{aligned}$$

Proof.

¿From (26), $\|x^+ - \chi\|_{x^+} \le \delta^2/(1-\delta^2)$, since $\delta(x^+) \le \delta^2$. Using Lemma 3.4 with $\alpha = \delta^2/(1-\delta^2)$, we obtain $\|x^+ - \chi\|_\chi \le \delta^2/(1-2\delta^2)$. The first result in the lemma follows from this with $\delta = 0.1$.

¿From Lemma 4.2, $\|x - \chi\|_x \ge \delta(1-2\delta)/(1-\delta)$. From (26), $\|x - \chi\|_x \le 1/(1-\delta)$. Using Lemma 3.4 with $\alpha = 1/(1-\delta)$, we get $\|x - \chi\|_\chi \ge (1-\alpha)\|x - \chi\|_x$. Manipulating these expressions, we arrive at

$$\|x - \chi\|_\chi \ge \left(\frac{1-2\delta}{1-\delta}\right)^2 \delta.$$

Substituting $\delta = 0.1$, we obtain the second result, therefore completing the proof. ∎

The primal centering direction $h(x)$ is the Newton direction for $p(\cdot)$ from x, and it coincides with the steepest descent direction for $x = e$, i.e., $h(x)$ is the Cauchy direction from e. To see this notice that $h(x) = -xP_{AX}x\nabla p(x) = xP_{AX}xx^{-1}$.

Other scalings give rise to descent directions that are in general not as efficient as this one. We shall apply Lemma 3.3 to study the effect of slightly shifted scalings on the descent directions.

Lemma 4.4 *Let $q \in \mathbb{R}^n$ be such that $\|q - e\| \le \alpha \in (0, 0.25)$, define for $x \in \mathcal{P}^0$, $h(x) = xP_{AX}e$ and $h(x,q) = xqP_{AXQ}q$. Then*

$$\|h(x,q) - h(x)\|_x \le 3\alpha.$$

Proof. By Lemma 3.3, $\|P_{AX}e - qP_{AXQ}q\| \le 3\alpha$. But this is by definition equal to $\|h(x) - h(x,q)\|_x$, completing the proof. ∎

5 The Mizuno-Todd-Ye Algorithm

The 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

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

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

Algorithm 5.1 *Given* $\alpha \leq 0.3$, (x^{0^1}, s^{0^1}) *such that* $\delta(x^{0^1}, s^{0^1}) \leq \alpha^2/\sqrt{2}$, *k=1.*

REPEAT

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

Predictor: Compute the (affine-scaling) step u^0, v^0, $x := x^0 + u^0$, $s := s^0 + v^0$ such that

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

where γ is such that (x, s) is feasible and $\delta(x, s) = \alpha$.

Corrector: Set $\mu := \mu(x, s)$. Compute the (centering) step (u, v) such that

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

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.

We now list some properties of this algorithm. Some proofs are presented here for the sake of completeness. The proofs that are not given here can be found in Mizuno, Todd, and Ye [13]. Mizuno, Todd, and Ye proved that the algorithm is well defined in the sense that the centering step produces (x^+, s^+) such that $\delta(x^+, s^+) \leq \alpha^2/\sqrt{2}$.

Bounds on the quantities appearing in the algorithm are given in the lemmas below. Let $\{B, N\}$ be the optimal partition for the linear programming problem, i.e., the index partition associated with the optimal face. As we described in Subsection 4.1, the central path ends at the analytic center of the optimal face, and the pairs (x,s) such that $\|w(x,s)-e\| \leq \alpha$ constitute a neighborhood of the central path bounded away from the non-optimal faces of the feasible polyhedron and correspond to a bundle of w-weighted affine-scaling trajectories. For α small, the bundle of trajectories ends in a compact neighborhood of the analytic center of the optimal face, and so all the sequences generated by the algorithm are in compact sets.

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 stay in small neighborhoods of x_B^*, s_N^*, the analytic centers of the primal and dual optimal faces.

Lemma 5.1 *Consider quantities generated by the MTY algorithm. Then*

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

Proof. (i): Since (x,s) approaches a compact set contained in the relative interior of the optimal face depending on the problem data, $s_N = \Omega(1)$. At all points visited by the algorithm, $\|w(x,s)-e\| \leq \alpha \in (0,0.5)$, and consequently $x_i s_i \leq \mu(x,s)(1+\alpha)$. It follows that for $i=1,\dots,n$,

$$x_i \leq (1+\alpha)\,\frac{\mu}{\Omega(1)} = O(\mu).$$

The proof for all four equalities is similar.

(ii) Using (9) and (10) with $\gamma = 0$, we have

$$u^0 = -x^0\phi^0\bar{h},$$

where $\bar{h}$ can be represented by the two equivalent expressions:

$$\bar{h} = P_{AX^0\Phi^0}\phi^0 w^0, \tag{27}$$

$$\bar{h} = P_{AX^0\Phi^0}\phi^0 x\frac{s^*}{\mu^0}, \tag{28}$$

where s^* will be chosen as an arbitrary optimal dual slack, with $s^*_B = 0$ and $s^*_N = O(1)$.

¿From (27), $\bar{h} = O(1)$ and consequently, since $x^0_N = O(\mu^0)$,

$$u^0_N = x^0_N \phi^0_N \bar{h}_N = O(\mu^0)$$

We must prove that $u^0_B = O(\mu^0)$. Using the definition of projection and problem (17), $\bar{h}_B$ must solve the problem derived from (28) (with $\rho_B = 0$):

$$\begin{array}{ll} \text{minimize} & \|\bar{h}_B\| \\ \text{subject to} & A_B X^0_B \Phi^0_B h_B = -A_N X^0_N \Phi^0_N \bar{h}_N. \end{array}$$

The right-hand side of the constraint satisfies $A_N X^0_N \Phi^0_N \bar{h}_N = O(\mu^0)$, since $x^0_N = O(\mu^0)$ by (i). Choosing any non-singular subsystem, we can compute a solution $\tilde{h}_B = O(\mu^0)$, since $x^0_B = \Omega(1)$ and $\phi^0_B = \Omega(1)$. It follows that $\|\bar{h}_B\| \leq \|\tilde{h}_B\| = O(\mu^0)$, proving that $u^0_B = -x^0_B \phi^0_B \bar{h}_B = O(\mu^0)$. The proof for v^0 is similar.

(iii) Now for the corrector step. From (9),

$$u = -x\phi P_{AX\Phi}\phi(w - e)$$

As in the proof of (ii), $\phi(w - e) = O(1)$ and $x_N = O(\mu)$, resulting in $u_N = O(\mu)$ and similarly $s_B = O(\mu)$, completing the proof. ∎

The lemma above shows that all the variations in (x, s) due to a MTY step are bounded by $O(\mu)$, with exception of u_B and v_N. These are the variations in the large variables due to the corrector step.

6 Convergence of the MTY Algorithm

In this section we establish the main result of the paper: the points generated by the MTY algorithm always converge to the analytic center of the optimal face. We shall assume that the optimal face is not a single point. Our convergence proofs will be carried out for primal solutions. The symmetric results for dual slacks can always be proved by the same methods using the complete symmetry of conditions (1).

We begin by studying the map that results from the algorithm. Towards this end we describe the relationship between primal-dual pairs (x^0, s^0) and the result (x^+, s^+) of a MTY step originating at (x^0, s^0). It is essential to keep in mind that at this point we are not studying sequences generated by the algorithm. We derive a lemma (a main result of the paper) on the boundary behaviour of the algorithmic map for sequences with strong convergence properties; a second lemma extends the result to nonconvergent sequences, and provides the main convergence property of the algorithmic map*. We then consider a sequence generated by the algorithm, and prove in Theorem 6.3 that it converges to the analytic center of the optimal face.

Consider a sequence of interior primal-dual pairs (x^{0^k}, s^{0^k}), and all the quantities that would be generated by applying one MTY step from each of these points, namely (u^{0^k}, v^{0^k}), (x^k, s^k), (u^k, v^k), (u^{+^k}, v^{+^k}), μ^{0^k}, $\mu^k = \gamma^k \mu^{0^k}$, w^{0^k}, w^k, ϕ^{0^k}, ϕ^k. Again, we stress the fact that presently $(x^0, s^0)^{k+1}$ is not necessarily related to $(x^+, s^+)^k$. Our main interest is in measuring how the large variables approach x^*_B. A good metric for measuring this is given by the norm $\|\cdot\|_{x^*_B}$, defined on $I\!R^{|B|}$. To simplify notation, we write

$$\|\cdot\|_* \equiv \|\cdot\|_{x^*_B}.$$

Lemma 6.1 *Let (x^{0^k}, s^{0^k}) be such that $\delta(x^{0^k}, s^{0^k}) \leq 0.1$, and assume that $\mu^{0^k} \to 0$, $(x^{0^k}, s^{0^k}) \to (\bar{x}, \bar{s})$, $w^{0^k} \to \bar{w}^0$, $w^k \to \bar{w}$. Then*
(i): If $\bar{x} = x^*$, then $u^k \to 0$ and $x^{+^k} \to x^*$.
(ii): If $\bar{x} \neq x^*$, then for sufficiently large k,

$$\|x_B^{+^k} - x^*_B\|_* \leq 0.8\|x_B^{0^k} - x^*_B\|_*.$$

Proof. The proof consists of two technical parts and a conclusion. In the first part we analyse the boundary behaviour of the MTY steps; in the second part we describe the centering direction from $\bar{x}$ in the optimal face. Finally, the conclusion is reached from the comparison of the results of the first two parts.

We begin by considering MTY steps. From Lemma 5.1, $u^{0^k} \to 0$ and consequently $x^k \to \bar{x}$. ¿From the same lemma, $u^k_N \to 0$. We must describe the behaviour of u^k_B. ¿From (10),

$$u^k = -x^k \phi^k P_{AX^k\Phi^k} \phi^k \left(\frac{x^k s^*}{\mu^k} - e \right).$$

*The reader might consider Lemma 6.2 before going through the technical proof of Lemma 6.1.

Using Lemma 3.2 with $s_B^* = 0$, $x_B^k \to \bar{x}_B > 0$, $\phi_B^k \to \bar{\phi}_B > 0$,

$$u_B^k \to \bar{u}_B = \bar{x}_B \bar{\phi}_B P_{A\bar{X}_B\bar{\Phi}_B} \bar{\phi}_B. \tag{29}$$

Since $x^{+^k} = x^{0^k} + u^{0^k} + u^k$ and $u^{0^k} \to 0$, $u_N^k \to 0$,

$$x^{+^k} \to \bar{x}^+ = \bar{x} + \bar{u},$$

where $\bar{u}_N = 0$.

Our attention now goes to centering in the optimal face. Consider the following primal centering direction associated with each (x^{0^k}, s^{0^k}):

$$h^k = -x^{0^k} P_{AX^{0^k}} \left(\frac{x^{0^k} s}{\mu^{0^k}} - e \right), \tag{30}$$

where s is an arbitrary dual slack (remember that $dP_{AD}ds = dP_{AD}ds'$ for any dual slacks s, s' and any scaling $d > 0$.)

With $s = s^{0^k}$, we see that $h^k = -x^{0^k} P_{AX^{0^k}}(w^{0^k} - e)$. It follows that $\bar{h}_N = 0$ and

$$\|h^k\|_{x^{0^k}} \leq \|w^{0^k} - e\| = \delta(x^{0^k}, s^{0^k}) \leq 0.1.$$

With $s = s^*$, we use Lemma 3.2 with $s_B^* = 0$ and conclude that $h^k \to \bar{h}$, where

$$\bar{h}_N = 0 \quad , \quad \bar{h}_B = \bar{x}_B P_{A_B \bar{X}_B} e_B \quad , \quad \|\bar{h}_B\|_{\bar{x}_B} \leq 0.1.$$

We conclude that $\bar{h}$ is the Newton centering direction in the optimal face, and that the proximity measure of $\bar{x}$ is

$$\delta(\bar{x}_B) = \|\bar{h}_B\|_{\bar{x}_B} \leq 0.1.$$

Let $y = \bar{x} + \bar{h}$ be the result of a primal centering step. Then by Lemma 4.3,

$$\begin{aligned} \|\bar{x}_B - x_B^*\|_* &\geq 0.75\delta(\bar{x}_B) \\ \|y_B - x_B^*\|_* &\leq 1.05\delta^2(\bar{x}_B). \end{aligned} \tag{31}$$

Our attention now turns to shifted scaling. We study the effect of the direction $\bar{u}_B$ defined in (29), when it is used for primal centering instead of $\bar{h}$. The quantity

$$\bar{u}_B = \bar{x}_B \bar{\phi}_B P_{A\bar{X}_B\bar{\Phi}_B} \bar{\phi}_B$$

corresponds to $\bar{h}_B$ by way of a shifted scaling. Here $\bar{\phi} = 1/\sqrt{\bar{w}}$, as usual. Since $\|\bar{w} - e\| \leq 0.1$, it follows that for $i = 1, \ldots, n$ $\bar{w}_i \in [0.9, 1.1]$ and it is trivial to check that $\bar{\phi}_i \in [0.9, 1.1]$. Hence $\left\|\bar{\phi} - e\right\|_\infty \leq 0.1$, and by Lemma 3.3,

$$\|\bar{h}_B - \bar{u}_B\|_{\bar{x}_B} \leq 0.3\|\bar{h}_B\|_{\bar{x}_B} = 0.3\delta(\bar{x}_B). \tag{32}$$

If $\bar{x} = x^*$, then $\delta(\bar{x}_B) = 0$ and it follows that $\bar{h}_B = \bar{u}_B = 0$. This proves part (i) of the lemma. Assume from here on that $\|\bar{x}_B - x^*_B\| \neq 0$.

We need (32) in the norm $\|\cdot\|_*$. Using (26), define

$$\alpha = \|\bar{x}_B - x^*_B\|_{\bar{x}_B} \leq \frac{\delta(\bar{x}_B)}{1 - \delta(\bar{x}_B)} \leq \frac{0.1}{0.9}.$$

Using Lemma 3.4,

$$\|\bar{h}_B - \bar{u}_B\|_* \leq \frac{1}{1-\alpha}\|\bar{h}_B - \bar{u}_B\|_{\bar{x}_B}$$

Merging this and (32) with $1/(1-\alpha) \leq 1.2$ we obtain

$$\|\bar{h}_B - \bar{u}_B\|_* \leq 0.4\delta(\bar{x}_B). \tag{33}$$

And now we compare the points $y_B = \bar{x}_B + \bar{h}_B$ and $\bar{x}^+_B = \bar{x}_B + \bar{u}_B$, using (31). Specifically

$$\begin{aligned}
\|\bar{x}^+_B - x^*_B\|_* &\leq \|y_B - x^*_B\|_* + \|\bar{x}^+_B - y_B\|_* \\
&= \|y_B - x^*_B\|_* + \|\bar{u}_B - \bar{h}_B\|_* \\
&\leq 1.05\delta^2(\bar{x}_B) + 0.4\delta(\bar{x}_B) \\
&\leq 0.51\delta(\bar{x}_B).
\end{aligned}$$

Using (31), we conclude that

$$\frac{\|\bar{x}^+_B - x^*_B\|_*}{\|\bar{x}_B - x^*_B\|_*} \leq \frac{0.51}{0.75} \leq 0.7.$$

Finally, we conclude from this expression that since $x^{0^k} \to \bar{x}$ and $x^{+^k} \to \bar{x}^+$, for sufficiently large k,

$$\|x^{+^k}_B - x^*_B\|_* \leq 0.8\|x^{0^k}_B - x^*_B\|_*,$$

completing the proof. ∎

The lemma above studies convergent sequences (x^{0^k}, s^{0^k}). The next lemma shows that the reduction in distance from x^* can be extended uniformly for nonconvergent sequences.

Lemma 6.2 *Let (x^{0^k}, s^{0^k}) be 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 ϵ^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. Assume by contradiction that there exists $\epsilon > 0$ and a subsequence of (x^{0^k}, s^{0^k}) with indices $\mathcal{K}^0 \subset I\!N$ such that for $k \in \mathcal{K}^0$,

$$\|x_B^{+^k} - x_B^*\|_* > \epsilon \quad , \quad \|x_B^{+^k} - x_B^*\|_* > 0.8\|x_B^{0^k} - x_B^*\|_*. \tag{34}$$

The sequences (x^{0^k}, s^{0^k}), (w^{0^k}), (w^k) are all in compact sets by construction, and thus there must exist a subsequence with indices $\mathcal{K} \subset \mathcal{K}^0$ such that these three sequences are convergent in $\mathcal{K}$.

In particular, $(x_B^{+^k})_{\mathcal{K}}$ does not converge to x_B^*, due to (34). Applying Lemma 6.1(i), we see that $(x^{0^k})_{\mathcal{K}}$ does not converge to x^*, and thus (ii) must hold for this subsequence. This contradicts (34), completing the proof. ∎

Finally we are ready to establish our convegence result.

Theorem 6.3 *Consider sequences (x^{0^k}, s^{0^k}), (x^k, s^k) generated by the MTY algorithm. Then $(x^{0^k}, s^{0^k}) \to (x^*, s^*)$ and $(x^k, s^k) \to (x^*, s^*)$, where (x^*, s^*) is the analytic center of the solution set.*

Proof. We 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 $\{\epsilon^k\}$ be the sequence guaranteed by Lemma 6.2, and let $\bar{k}$ be such that the conclusions of that lemma are valid for $k \geq \bar{k}$. Choose an index $j \geq \bar{k}$

such that $\|x_B^{0^j} - x_B^*\|_* < 1.1\sigma$, and such that for $k \geq j$, $\epsilon^k < 0.5\sigma$. This index exists because $\epsilon^k \to 0$ and $\bar{x}_B$ is an accumulation point of $\{x_B^{0^k}\}$.

We prove by induction that for any $k > j$, $\|x_B^{0^k} - x_B^*\|_* < 0.9\sigma$.

(a) $\|x_B^{0^{j+1}} - x_B^*\|_* < 0.8 \times 1.1\sigma < 0.9\sigma$ by Lemma 6.2.

(b) Assume that for an index $k > j$, $\|x_B^{0^k} - x_B^*\|_* < 0.9\sigma$. Then by Lemma 6.2, $\|x_B^{0^{k+1}} - x_B^*\|_* \leq \max\{\epsilon^k, 0.8\|x_B^{0^k} - x_B^*\|_*\} < 0.9\sigma$.

(a) and (b) prove that for all $k > j$, $\|x_B^{0^k} - x_B^*\|_* < 0.9\sigma$, contradicting the fact that σ is an accumulation point of the sequence $(\|x_B^{0^k} - x_B^*\|_*)$, and completing the proof. ∎

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] A. CHARNES, W. W. COOPER, AND R. M. THRALL, *A structure for classifying and characterizing efficiency and inefficiency in data envelopment analysis*, The Journal of Productivity Analysis 2 (1991), pp. 197-237.

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

[4] W. W. HOGAN, *Point-to-set maps in mathematical programming*, SIAM Review 3 (1973), pp. 591-603.

[5] P. HUARD, *Point-to-set maps in mathematical programming*, Math. Programming Study 10, North Holland Publishing, Amsterdam (1979).

[6] 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).

[7] 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.

[8] L. McLinden, *An analogue of Moreau's proximation theorem, with application to the nonlinear complementarity problem*, Pacific Journal of Math. 88 (1980), pp. 101-161.

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

[10] 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).

[11] N. Megiddo and M. Shub, *Boundary behaviour of interior point algorithms in linear programming*, Mathematics of Operations Research 14 (1989), pp. 97–146.

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

[13] 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).

[14] R. C. Monteiro and I. Adler, *Interior path-following primal-dual algorithm. Part I: linear programming*, Mathematical Programming 44 (1989), pp. 27-41.

[15] G. Sonnevend, *An analytical centre for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming*, in Lecture Notes in Control and Information Sciences 84, Springer Verlag, New York, NY, (1985), pp. 866-876.

[16] 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).

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

[18] Y. YE, *On the Q-order convergence of interior-point algorithms for linear programming*, Department of Management Sciences, The University of Iowa (1991).

[19] Y. YE, O. GÜLER, R. A. TAPIA AND Y. ZHANG, *A quadratically convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming*, Technical Report TR91-26, Department of Computational and Applied Mathematics, Rice University (1991). To appear in Mathematical Programming.

[20] 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).

[21] Y. ZHANG AND A. EL-BAKRY, *A modified predictor-corrector algorithm for locating weighted centers in linear programming*, Technical Report TR92-19, Dept. of Computational and Applied Mathematics, Rice University (1992).

[22] Y. ZHANG AND R. A. TAPIA, *A superlinearly convergent polynomial primal-dual interior-point algorithm for linear programming*, Technical Report TR90-40, Department of Computational and Applied Mathematics, Rice University (1990). To appear in SIAM Journal on Optimization.

[23] 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.

[24] Y. ZHANG AND R. A. TAPIA, *On the convergence of interior-point methods to the center of the solution set in linear programming*, Technical Report TR91-30, Dept. of Computational and Applied Mathematics, Rice University (1991).

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

[26] Y. ZHANG, R. A. TAPIA, AND F. POTRA, *On the superlinear convergence of interior-point algorithms for a general class of problems*, Technical Report TR90-9, Dept. of Computational and Applied Mathematics, Rice University (1990). To appear in SIAM Journal on Optimization.