SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT NO. 952

December 1990

Revised May 1991

TODD'S LOW-COMPLEXITY ALGORITHM IS A PREDICTOR-CORRECTOR PATH-FOLLOWING METHOD

by

M.J. Todd[1], J.-P. Vial[2]

[1]Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550.

[2]Département d'Economie Commerciale et Industrielle, Université de Genève. Research supported by FNRS through Grant 12-26434.89.

TODD'S LOW-COMPLEXITY ALGORITHM IS A PREDICTOR-CORRECTOR PATH-FOLLOWING METHOD

M.J. Todd[1]

School of Operations Research and Industrial Engineering,

Cornell University

J.-P. Vial[2]

Département d'Economie Commerciale et Industrielle

Université de Genève

May 22, 1991

[1]Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550.

[2]Research supported by FNRS through Grant 12–26434.89.

Abstract

We show that Todd's low-complexity algorithm for linear programming takes affine-scaling steps when and only when the current point is nearly centered, and, after such a step, requires at most three more iterations before taking another affine-scaling step. The main tool is Roos and Vial's measure of centrality.

Key Words: linear programming, interior-point method, path-following method.

1 Introduction

Interior-point methods for linear programming can be roughly divided into affine-scaling methods (e.g. Dikin [5], Barnes [3]) projective methods based on Karmarkar [8], path-following methods (e.g. Roos and Vial [10]) and potential reduction methods (e.g. Ye [13]). The path-following methods try to stay close to the so-called central path or central trajectory in either primal, dual, or primal-dual space, while potential reduction methods attempt only to guarantee a suitable reduction in a primal or primal-dual potential function.

Todd's "low-complexity algorithm" [11] at each iteration takes a step in either the affine-scaling direction of Dikin [5] or the constant-cost centering direction, with a very simple rule to determine which to take. The step length is fixed in the scaled space, i.e., when the current point is scaled so that each component is one. This fixed step length is characteristic of affine-scaling and projective methods, at least in their basic form. The low complexity algorithm requires only $O(\sqrt{n}\,t)$ iteration to attain precision t for a problem in standard form with n variables, and thus attains the current best theoretical complexity bound. The complexity analysis uses the primal-dual potential function.

In this paper we show that, after an initial centering phase, the low-complexity algorithm is a predictor-corrector path-following method. Affine-scaling steps are taken only when the current point is approximately centered. After such a step, at most three constant-cost centering steps are required before the new iterate is approximately centered and another affine-scaling step is taken.

Section 2 provides the tools to relate algorithms taking fixed step lengths in the scaled space to path-following methods. First, we show that if a primal feasible point is nearly centered according to the Roos-Vial measure of centrality [10] then such a step will yield a point that is also approximately centered (in a precisely quantified sense). This result seems to be of independent interest, and may relate to recent work showing that slight variants of Karmarkar's projective method [8] can be viewed as path-following methods (Xiao and Goldfarb [12], Anstreicher [2], Goffin and Vial [7]). We also give a complementary result for a primal-dual framework. Next, we relate the Roos-Vial measure of centrality to the quantity in Todd's algorithm that determines which kind of step to take. Finally, we adapt the Roos-Vial quadratic convergence result [10] on centering steps to account for partial projected Newton steps (the resulting convergence is, as expected, only linear).

Section 3 applies these results to the low-complexity algorithm and proves the main result. This suggests a number of variations to the algorithm which are discussed. The section concludes by mentioning results of very limited computational testing which indicates that far fewer centering steps seem to be required in practice (at most one every ten iterations). Finally, section 4 considers whether this new view of the algorithm can provide alternative proofs of the convergence results of [11].

2 Preliminary Results

We consider the linear programming problem in standard form :

$$(\mathrm{P}) \qquad \begin{aligned} \min \quad & c^T x \\ & Ax = b \\ & x \geq 0, \end{aligned}$$

where A is $m \times n$, and b, c and x are suitably dimensioned, and write its dual as

$$(\mathrm{D}) \qquad \begin{aligned} \max \quad & b^T y \\ & A^T y + s = c \\ & s \geq 0. \end{aligned}$$

We assume that A has full rank m and that both (P) and (D) have feasible solutions with strictly positive x and s respectively. This implies that both have optimal solutions with bounded sets of optimal solutions. In this case, for any $\mu > 0$, there is a unique optimal solution to the barrier problem.

$$\min\left\{\frac{c^T x}{\mu} - \sum_j \ln x_j : Ax = b,\ x > 0\right\}.$$

If we write this solution as $x(\mu)$, then it, together with some $y(\mu)$, $s(\mu)$, is the unique solution to

$$\begin{aligned} Ax &= b, \quad x > 0 \\ A^T y + s &= c \\ XSe &= \mu e, \end{aligned} \tag{1}$$

where X and S are the diagonal matrices containing the components of x and s and e denotes the vector of ones in $\mathbb{R}^n$. The set of all $x(\mu)$ is called the central path or central trajectory.

Clearly $s(\mu) > 0$ so that $(y(\mu), s(\mu))$ is feasible in (D), and (1) is sufficient for $(y(\mu), s(\mu))$ to be the unique optimal solution to the barrier problem:

$$\max\left\{\frac{b^T y}{\mu} + \sum_j \ln s_j : A^T y + s = c,\ s > 0\right\}.$$

If the feasible solution set for (P) is bounded, then the primal barrier problem above also has a unique optimal solution for $\mu < 0$, and this is also given by a solution to (1), except that now $s(\mu)$ is strictly negative.

Henceforth we suppose we are given some x that is strictly feasible in (P), i.e., $Ax = b$ and $x > 0$. We can then seek a pair (y, s) that satisfies (1) as nearly as possible. Roos and Vial [10] define

$$\delta(x,\mu) := \min_{y,s}\left\{\left\|\frac{Xs}{\mu} - e\right\| : A^T y + s = c\right\} \tag{2}$$

as a measure of proximity of x to $x(\mu)$.

Let $\gamma \in (0,1)$ be fixed, and suppose $AXd = 0$, $\|d\| = 1$. Let

$$x^+ := x + \gamma X d. \tag{3}$$

We see that x^+ is the result of a step of fixed length in the scaled space, since

$$\left\|X^{-1}(x^+ - x)\right\| = \|\gamma d\| = \gamma.$$

Our first result shows that such a step cannot degrade the measure of proximity to $x(\mu)$ by too much.

Proposition 1 *Suppose x is strictly feasible in (P) with $\delta(x,\mu) \leq \Delta$. Then x^+ given by (3) is also strictly feasible in (P) with*

$$\delta(x^+,\mu) \leq \Delta + \gamma(1+\Delta). \tag{4}$$

Proof: Clearly $Ax^+ = Ax = b$, since $AXd = 0$, and $X^{-1}x^+ = e + \gamma d > 0$, since $\|d\| = 1$ and $0 < \gamma < 1$. So x^+ is strictly feasible.

Now let y and s achieve the minimum in (2). Then $\delta(x^+,\mu) \leq \|\frac{X^+s}{\mu} - e\|$ and the maximum component of $\frac{Xs}{\mu} - e$ is at most Δ in absolute value, so that $\left\|\frac{XS}{\mu}\right\| \leq 1 + \Delta$. Hence

$$\begin{aligned}
\delta(x^+,\mu) &\leq \left\|\frac{X^+s}{\mu} - e\right\| \\
&= \left\|\frac{Xs}{\mu} - e + \gamma\frac{XS}{\mu}d\right\| \\
&\leq \left\|\frac{Xs}{\mu} - e\right\| + \gamma\left\|\frac{XS}{\mu}\right\| \|d\| \\
&\leq \Delta + \gamma(1+\Delta).
\end{aligned}$$

as required. $\square$

To complement this result we prove a result for a primal-dual framework. Let s be strictly feasible in (D), i.e., $A^T y + s = c$ for some y and $s > 0$. We now define our measure of proximity to $(x(\mu), s(\mu))$ as

$$\delta_{PD}(x,s,\mu) := \left\| \frac{Xs}{\mu} - e \right\|. \tag{5}$$

Let $\gamma_x, \gamma_s \in (0,1)$ be fixed, and suppose $\|d_x\| = \|d_s\| = 1$ with $AXd_x = 0$ and d_s in the range space of $S^{-1}A^T$. Let

$$\begin{aligned} x^+ &= x + \gamma_x X d_x, \\ s^+ &= s + \gamma_s S d_s. \end{aligned} \tag{6}$$

Then we have

Proposition 2 *Suppose x and s are strictly feasible in (P) and (D) respectively and $\delta_{PD}(x,s,\mu) \leq \Delta$. Then x^+ and s^+ given by (6) are also strictly feasible in (P) and (D) respectively and*

$$\delta_{PD}(x^+, s^+, \mu) \leq \Delta + (\gamma_x + \gamma_s + \gamma_x\gamma_s)(1+\Delta).$$

Proof: The first part follows easily from the definitions of x^+ and s^+. Also,

$$\begin{aligned} \delta_{PD}(x^+, s^+, \mu) &= \left\| \frac{X^+S^+}{\mu} e - e \right\| \\ &= \left\| \frac{XS}{\mu} e - e + \gamma_x \frac{XS}{\mu} d_x + \gamma_s \frac{XS}{\mu} d_s + \gamma_x\gamma_s \frac{XS}{\mu} D_x d_s \right\| \\ &\leq \Delta + (1+\Delta)(\gamma_x + \gamma_s + \gamma_x\gamma_s \|D_x\|) \\ &\leq \Delta + (1+\Delta)(\gamma_x + \gamma_s + \gamma_x\gamma_s), \end{aligned} \tag{7}$$

where $D_x = \operatorname{diag}(d_x)$ and $\left\| \frac{XS}{\mu} \right\| \leq 1 + \Delta$ as before. □

Similar results hold for

$$\delta'_{PD}(x,s) := \delta_{PD}\left(x, s, \frac{x^T s}{n}\right), \tag{8}$$

(since $\delta'_{PD}(x,s) = \frac{n}{x^T s} \inf_{\mu} \|Xs - \mu e\|$, we find then

$$\delta'_{PD}(x^+, s^+) \leq \frac{\Delta + (\gamma_x + \gamma_s + \gamma_x\gamma_s)(1+\Delta)}{1 - \frac{(\gamma_x+\gamma_s)(1+\Delta)}{\sqrt{n}} - \frac{\gamma_x\gamma_s(1+\Delta)}{n}}) \tag{9}$$

or if (6) is replaced by

$$\begin{aligned} x^+ &= x + \gamma_x \left(\tfrac{x^T s}{n}\right)^{\frac{1}{2}} \left(XS^{-1}\right)^{\frac{1}{2}} d_x \\ s^+ &= s + \gamma_s \left(\tfrac{x^T s}{n}\right)^{\frac{1}{2}} \left(XS^{-1}\right)^{\frac{1}{2}} d_s, \end{aligned} \tag{10}$$

but we omit the details. The measure $\delta'_{PD}(x,s)$ arises in many primal-dual path-following algorithms, and (10) is appropriate when symmetric primal-dual scaling is needed.

Next we relate $\delta(x,\mu)$ to an important parameter in the low-complexity algorithm [11]. For any vector $u \in \mathbb{R}^n$, let u_p denote its projection into the null space of AX. Let $\bar{c} = Xc$, so that $-\bar{c}_p$ is the search direction in the affine-scaling method of Dikin [5].

Let

$$\bar{d}_\beta := -\beta\bar{c}_p + e_p$$

and

$$\alpha := \arg\min\left\{\|\bar{d}_\beta\| : \beta \in \mathbb{R}\right\}. \tag{11}$$

We also let

$$\alpha_\geq := \arg\min\left\{\|\bar{d}_\beta\| : \beta \in \mathbb{R}_+\right\}, \tag{12}$$

i.e., where β is now restricted to be nonnegative.

Proposition 3

$$\text{a)} \qquad \|\bar{d}_\alpha\| = \inf_{\mu\in\mathbb{R}} \delta(x,\mu) \tag{13}$$

$$\text{b)} \qquad \|\bar{d}_{\alpha_\geq}\| = \inf_{\mu\in\mathbb{R}_+} \delta(x,\mu). \tag{14}$$

The infimum in case (a) (case (b)) is attained iff α ($\alpha_\geq$) is nonzero, and then the minimizing μ ($\mu_\geq$) is given by $1/\alpha$ ($1/\alpha_\geq$), and if s ($s^\geq$) attains the corresponding minimum in (2), then

$$\bar{d}_\alpha = -\frac{Xs}{\mu} + e \tag{15}$$

$$(\bar{d}_{\alpha_\geq} = -\frac{Xs^\geq}{\mu_\geq} + e). \tag{16}$$

Proof: For any fixed μ we can decompose $\frac{Xs}{\mu} - e$ into its component in the null space of AX and its component in the range space of XA^T:

$$\frac{Xs}{\mu} - e = \left[\frac{(Xs)_p}{\mu} - e_p\right] + \left[\frac{Xs-(Xs)_p}{\mu} - (e-e_p)\right].$$

The constraint that $A^Ty + s = c$ (or $XA^Ty + Xs = Xc$) for some y can be written as $(Xs)_p = \bar{c}_p$. Thus to minimize $\left\|\frac{Xs}{\mu} - e\right\|$ over such s's, we choose s so that $Xs - (Xs)_p = \mu(e_p - e)$, and then

$$\frac{Xs}{\mu} - e = \frac{\bar{c}_p}{\mu} - e_p = -\bar{d}_{1/\mu}. \tag{17}$$

The proposition then follows immediately — note that we need to take the infimum over μ since μ corresponds to $1/\beta$ and α (or $\alpha_{\geq}$) might be zero. □

The proof of the proposition shows that, if s achieves the minimum in (2) for any μ, then $-\frac{Xs}{\mu}+e$ lies in the null space of AX – see (17).

Now let

$$x^+ = x + \bar{\gamma} X(-\frac{Xs}{\mu} + e). \tag{18}$$

If $\bar{\gamma}=1$, then this is the step considered by Roos and Vial [10], which coincides with the projected Newton step for the primal barrier problem considered above (Gill et al. [6]). Roos and Vial have shown ([10, Theorem 2.1]) that, if $\bar{\gamma}=1$,

$$\delta(x^+,\mu) \leq \delta(x,\mu)^2. \tag{19}$$

In order to treat also the low-complexity algorithm, we now consider the case where

$$\bar{\gamma} = \gamma \Big/ \Big\| -\frac{Xs}{\mu} + e \Big\|, \tag{20}$$

where $\gamma \in (0,1)$ is fixed, so that x^+ is the result of a step of fixed length in the scaled space.

Proposition 4 *Let x^+ be given by (18), where $\bar{\gamma}$ satisfies (20) and $0 < \gamma \leq \delta(x,\mu) \leq 1$. Then*

$$\delta(x^+,\mu) \leq (1+\gamma)\delta(x,\mu) - \gamma. \tag{21}$$

Proof: Let us denote $\delta(x,\mu)$ by Δ. Then

$$\begin{aligned}
\delta(x^+,\mu) &\leq \Big\| \frac{X^+ s}{\mu} - e \Big\| \\
&= \Big\| \frac{Xs}{\mu} - e + \bar{\gamma}\frac{XS}{\mu}(-\frac{Xs}{\mu} + e) \Big\| \\
&\leq \Big\| \frac{Xs}{\mu} - e \Big\| \, \Big\| I - \bar{\gamma}\frac{XS}{\mu} \Big\|.
\end{aligned}$$

Now, for each j,

$$1 - \frac{\gamma}{\Delta}(1+\Delta) \leq 1 - \frac{\gamma}{\Delta}\frac{x_j s_j}{\mu} \leq 1 - \frac{\gamma}{\Delta}(1-\Delta).$$

Since $\frac{\gamma}{\Delta} \leq 1$, the term $1 - \frac{\gamma}{\Delta}\frac{x_j s_j}{\mu}$ is at most $1 - \frac{\gamma}{\Delta}(1-\Delta)$ in absolute value, so that

$$\delta(x^+,\mu) \leq \Big(1 - \frac{\gamma}{\Delta}(1-\Delta)\Big)\Delta = (1+\gamma)\Delta - \gamma.$$

□

Note that if $\gamma = \delta(x,\mu)$ then (21) yields (19).

Usually, propositions 1 and 4 would be enough to analyze the low-complexity algorithm. However, it may happen at some iteration that α in (11) is zero so that the infimum in (13) is not attained and (15) does not hold. In order to cover all possible cases we wish to analyze the step

$$x^+ = x + \bar{\gamma} X \bar{d}_\alpha \tag{22}$$

where

$$\bar{\gamma} = \gamma / \|\bar{d}_\alpha\| \tag{23}$$

and $\gamma \in (0,1)$ is fixed.

Proposition 5 *Let x^+ be given by (22), where $\bar{\gamma}$ satisfies (23) and $0 < \gamma \leq \|\bar{d}_\alpha\| < 1$. Then, if $\bar{d}^+_{\alpha+}$ denotes the $\bar{d}_\alpha$ calculated at x^+, we have*

$$\|\bar{d}^+_{\alpha+}\| \leq (1+\gamma)\|\bar{d}_\alpha\| - \gamma. \tag{24}$$

Proof: If $\alpha \neq 0$, we can choose $\mu = 1/\alpha$ and then s attaining the minimum in (2) so that (15) holds. Then x^+ is also given by (18), so that by proposition 4 we have

$$\|\bar{d}^+_{\alpha+}\| \leq \delta(x^+, \mu) \leq (1+\gamma)\delta(x,\mu) - \gamma = (1+\gamma)\|\bar{d}_\alpha\| - \gamma$$

as desired.

Suppose now $\alpha = 0$. Then for any $0 < \epsilon \leq 1 - \|\bar{d}_\alpha\|$ we can choose μ so that $\delta(x,\mu) \leq \|\bar{d}_\alpha\| + \epsilon \leq 1$ and for $\beta = 1/\mu$ and

$$\bar{e}_{\alpha\beta} := \frac{\bar{d}_\alpha}{\|\bar{d}_\alpha\|} - \frac{\bar{d}_\beta}{\|\bar{d}_\beta\|}$$

we have $\|\bar{e}_{\alpha\beta}\| \leq \epsilon$. Then

$$\begin{aligned}
\|\bar{d}^+_{\alpha+}\| &\leq \delta(x^+, \mu) \\
&\leq \left\| \frac{X^+ s}{\mu} - e \right\| \\
&= \left\| \frac{Xs}{\mu} - e + \gamma \frac{XS}{\mu} \frac{\bar{d}_\alpha}{\|\bar{d}_\alpha\|} \right\| \\
&= \left\| \frac{Xs}{\mu} - e + \gamma \frac{XS}{\mu} \left(\frac{\bar{d}_\beta}{\|\bar{d}_\beta\|} + \bar{e}_{\alpha\beta} \right) \right\| \\
&\leq \left\| \frac{Xs}{\mu} - e + \gamma \frac{XS}{\mu} \frac{\bar{d}_\beta}{\|\bar{d}_\beta\|} \right\| + \gamma \left\| \frac{XS}{\mu} \right\| \|\bar{e}_{\alpha\beta}\|.
\end{aligned}$$

The first term is at most $(1+\gamma)(\|\bar{d}_\alpha\| + \epsilon) - \gamma$ by proposition 4 and the second at most $\gamma(1 + \|\bar{d}_\alpha\| + \epsilon)\epsilon$ using $\|XS/\mu\| \leq 1 + \delta(x,\mu) + \epsilon$ as before. Since ϵ can be arbitrarily small, the result follows. $\square$

3 Application to Todd's Low-Complexity Algorithm

We now show how the results in section 2 demonstrate that, after an initial centering phase, Todd's algorithm is a path-following method.

First we describe the algorithm. Given a strictly feasible x compute $\bar{c}_p$ and e_p and hence $\bar{d}_\alpha$ (see the paragraph containing(11)). If $||\bar{d}_\alpha||$ is at least a threshold Δ, set $d = \bar{d}_\alpha/||\bar{d}_\alpha||$; else set $d = -\bar{c}_p/||\bar{c}_p||$. The first case is called a constant-cost centering step, while the second is an affine-scaling step. Then define x^+ by (3), so that in the scaled space a step of fixed length γ is taken in the direction d. Todd [11] chooses $\Delta = .3$ and $\gamma = .2$, but it is easy to check that his analysis requires only trivial changes if $\Delta = .25$ instead. An affine-scaling step is only taken if $||\bar{d}_\alpha||$ is sufficiently small; by proposition 3 this means that x is close to some central point $x(\mu)$ according to the Roos-Vial measure.

Our main result is:

Theorem 1 *Each affine-scaling step in Todd's low-complexity algorithm is followed by at most three constant-cost centering steps before another affine-scaling step is taken, for $\Delta = .3$ or $.25$ and $\gamma = .2$.*

Proof: We first assume $\Delta = .25$ and $\gamma = .2$. Suppose an affine-scaling step is taken at x, so that $||\bar{d}_\alpha|| < .25$. By proposition 3 there is a μ so that $\delta(x, \mu) < .25$. Let the result of the step be x^+. Then x^+ is given by (3) for $d = -\bar{c}_p/||\bar{c}_p||$. Proposition 1 then implies that

$$\delta(x^+, \mu) < .25 + .2(1 + .25) = .5.$$

Let $\bar{d}_\alpha^+$ denote $\bar{d}_\alpha$ computed at x^+ (and similarly $\bar{d}_\alpha^{++}$ denote $\bar{d}_\alpha$ computed at x^{++}, etc.); note that the corresponding α's are all different, but our notation ignores this for typographic ease. Then (13) shows that $||\bar{d}_\alpha^+||$ is at most .5. If $||\bar{d}_\alpha^+||$ is less than .25, then another affine-scaling step is taken immediately, and there is nothing to prove. Otherwise, certainly $||\bar{d}_\alpha^+|| \geq .25 > \gamma$. Then the step is of the form (22), with $\bar{\gamma}$ given by (23), and proposition 5 implies that the resulting iterate, x^{++}, satisfies

$$||\bar{d}_\alpha^{++}|| < .5(1 + .2) - .2 = .4.$$

Again, either an affine-scaling step is taken at x^{++} or the next iterate x^{+++} satisfies

$$||\bar{d}_\alpha^{+++}|| < .4(1 + .2) - .2 = .28.$$

Finally, either an affine-scaling step is taken at x^{+++} or at the next iterate x^{++++}, since in the latter case $||\bar{d}_\gamma^{++++}|| < .28(1 + .2) - .2 < .25$. This concludes the proof for $\Delta = .25$. If $\Delta = .3$

the proof only differs in the bounds: we have $||\bar{d}_\alpha^+|| < .56$, $||\bar{d}_\alpha^{++}|| < .48$, $||\bar{d}_\alpha^{+++}|| < .38$ and $||\bar{d}_\alpha^{++++}|| < .26 < .3$. □

The theorem suggests several variations of the basic low-complexity algorithm. (Note that Todd [11] discusses several variations, most of which involve some form of line search, which immediately precludes close path-following behaviour.)

First, Roos-Vial's quadratic convergence result [10] (see (19)) implies that, when taking constant-cost centering steps, a full projected Newton step ($\bar{\gamma} = 1$ in (18)) is generally preferable to a partial one ($\bar{\gamma} < 1$). Indeed, the proof above shows that, if $\Delta = .25$ is chosen as the threshold, at most one such centering step is necessary after each affine-scaling step, since $(.5)^2 = 25$. (If $\Delta = .3$, at most two such steps are necessary.) The next result shows that these steps do not invalidate the complexity analysis in [11].

Theorem 2 *If the basic algorithm is modified so that full projected Newton steps are taken whenever $||\bar{d}_\alpha||$ lies between .25 and .9 (and the threshold is set at $\Delta = .25$), then it still only requires $O(\sqrt{n}\,t)$ iterations to attain precision t given a suitable initial point x^0. Moreover at most one such step is required after each affine-scaling step before another affine-scaling step is taken.*

Proof: We only need to prove the first part. The complexity analysis in [11] is based on showing that the potential function

$$\varphi_{n+\sqrt{n}}(x, s) := (n + \sqrt{n}) \ln x^T s - \sum_j \ln x_j - \sum_j \ln s_j$$

decreases by a constant (at least .02) at each iteration.

In the initial centering phase of the algorithm (until $||\bar{d}_\alpha||$ drops below .25), the quadratic convergence result shows that there are at most four steps where the full projected Newton step is taken. Each of these can only increase the potential function by a constant, using lemma 2 of [11]. Thereafter, every modified step is immediately preceded and followed by an affine-scaling step, at which $\varphi_{n+\sqrt{n}}$ is decreased by at least .02. Hence it suffices to show that these modified steps do not increase the potential.

But in a modified step, we have chosen γ equal to $\delta := ||\bar{d}_\alpha||$. Then lemma 2 of [11] shows that the potential function decrease is at least $\delta^2 - \delta^2/2(1-\delta)$. But, after the initial centering phase, δ is at most .5 so that the decrease is nonnegative. □

Next, we prove that longer steps, with $\gamma = .4$, can be taken if $||\bar{d}_\alpha||$ is very small, and this is easy to obtain with quadratic convergence. Hence we can modify the algorithm as follows :

- if $||\bar{d}_\alpha|| \geq .9$, take a constant-cost centering step with $\gamma = .4$ (partial projected Newton step);
- if $||\bar{d}_\alpha|| \in [.0625, .9)$, take a full projected Newton step (constant cost centering step);
- if $||\bar{d}_\alpha|| < .0625$, take an affine-scaling step with $\gamma = .4$.

After the inital centering phase, each affine-scaling step will be followed by at most two full projected Newton steps before another affine-scaling, step, since after such a step, $||\bar{d}_\alpha|| < .0625 + .4(1 + .0625) < .5$ by proposition 1 and $\left((.5)^2\right)^2 = .0625$. Moreover, these full projected Newton steps will not increase the potential function, by the argument in the proof above, while the argument in [11] shows that the affine-scaling steps decreases it by at least

$$.4\sqrt{(.4)^2 - (0.625)^2} - (.4)^2/2(1 - .4) > .02.$$

Finally, we know that path-following algorithms usually only consider positive values of μ. Here, however, α may be negative. Since α corresponds to $1/\mu$, this seems unatural. Indeed, it is possible to replace $\bar{d}_\alpha$ by $\bar{d}_{\alpha_\geq}$ in the basic algorithm of [11] and in all the variants considered above without sacrificing anything. Instead of $\bar{d}_\alpha^T \bar{c}_p = 0$, we have $\bar{d}_{\alpha_\geq}^T \bar{c}_p \leq 0$, so that the centering direction taken is a direction of nonincreasing cost. Since $\bar{d}_{\alpha_\geq}$ makes a non-obtuse angle with any $\bar{d}_\beta$, $\beta \geq 0$, the argument in [11] still shows that a constant decrease in the potential function can be achieved. And the proofs of proposition 5 and theorem 1 carry over with obvious changes; for instance, $||\bar{d}_{\alpha_\geq}|| < .25$ shows that there is a positive μ with $\delta(x, \mu) < .25$.

We have done some computational testing of the basic algorithm in [11] as well the variants described above, using MATLAB 3.5e [9]. The problems, of sizes 50×100 up to 300×600, were randomly generated as described in [11]. The testing has been extremely limited, because these short-step algorithms require several hundred steps to converge; we have run at most two problems of each size. Nevertheless, in these runs, after the inital centering phase the algorithms using threshold .25 or .3 *never* require further centering steps, while those with threshold .0625 require at most one centering step every ten iterations, with none needed after the first one hundred iterations.

Our analysis does not anticipate this behaviour. Since our proposition 1 holds for steps of fixed length in the scaled space in *any* direction, while the affine-scaling step is in fact close to being parallel to the central trajectory, it is not surprising that fewer centering steps are required than indicated by the theory. However, there seems to be in addition some automatic self-centering, rather than the slow drift away from centrality that we would expect, and we do not yet understand this.

4 Convergence Rate

From the results of section 3 we can view the basic low-complexity algorithm as a predictor-corrector path-following method. The question then arises as to whether we can recover the convergence results of [11] using tools familiar in path-following methods. That is, can we show directly that each affine-scaling (predictor) step decreases the primal gap (from the current primal objective value to the optimal value), the duality gap, or the barrier parameter μ, by a factor $(1 - \text{constant}/\sqrt{n})$? We have not been able to do so; the analysis in [11] using the primal-dual function appears to be the only way to show that $O(\sqrt{n}\,t)$ iterations suffice to attain precision t, i.e., a primal gap no more than 2^{-t}. In this section we give the results we have been able to obtain.

First, we note that Dikin [5] for $\gamma = 1$ and Barnes [3] for $\gamma < 1$ showed that the affine-scaling method with fixed step length γ in the scaled space had an asymptotic convergence rate for the primal gap that is linear, with convergence ratio $1 - \gamma/\sqrt{n-m}$, in the nondegenerate case. The same result holds here. If x^{k+1} is the result of an affine-scaling step from x^k we have

$$c^T(x^{k+1} - x^k) = -\gamma\|\bar{c}_p\| \tag{25}$$

and

$$\begin{aligned} c^T x^k - c^T x^* &= (Xc)^T(X^{-1}x^k - X^{-1}x^*) \\ &= \bar{c}_p^T(e - X^{-1}x^*) \\ &\leq \|\bar{c}_p\|\|e - X^{-1}x^*\|, \end{aligned} \tag{26}$$

where $X = \mathrm{diag}(x^k)$. Since $x^k \to x^*$, and x^* has $n-m$ zero and m positive components, the last norm converges to $\sqrt{n-m}$, so

$$\frac{c^T x^{k+1} - c^T x^*}{c^T x^k - c^T x^*} = 1 - \gamma\frac{\|\bar{c}_p\|}{c^T x^k - c^T x^*} \leq 1 - \frac{\gamma}{\sqrt{n-m} + \varepsilon_k}, \tag{27}$$

where $\varepsilon_k \to 0$. In the low-complexity algorithm, we expect that this inequality is in fact close to tight, since x^k is close to the central path and we would expect that the direction towards the optimal solution in the scaled space, $X^{-1}x^* - e$, is close to the affine-scaling direction in this space, $-\bar{c}_p$. Indeed, from results in Adler and Monteiro [1], it is easy to see that for x^k on the central path, these directions converge as $x^k \to x^*$.

A similar result holds for the duality gap with respect to s, where s attains the minimum in (2) for $\mu = 1/\alpha$, and we assume $\alpha > 0$. We have

Proposition 6 *Suppose* $\left\|\frac{Xs}{\mu} - e\right\| = \|\bar{d}_\alpha\| \leq \Delta < 1$. *Then*

$$\frac{(x^{k+1})^T s}{(x^k)^T s} \geq 1 - \gamma\frac{\sqrt{n} + \Delta}{n - \Delta\sqrt{n}}. \tag{28}$$

Proof: From (25) we have

$$\frac{(x^{k+1})^T s}{(x^k)^T s} = 1 - \gamma \frac{\|\bar{c}_p\|}{(x^k)^T s} = 1 - \gamma \frac{\|\alpha \bar{c}_p\|}{\alpha (x^k)^T s}. \tag{29}$$

Now $\|\alpha \bar{c}_p\|$ lies in $\left[\|e_p\| - \Delta,\ \|e_p\| + \Delta\right]$, since

$$\|-\alpha \bar{c}_p + e_p\| = \|\alpha X s - e\| \leq \Delta,$$

and $\alpha (x^k)^T s = e^T(\alpha X s)$ lies in $\left[n - \Delta\sqrt{n},\ n + \Delta\sqrt{n}\right]$ for the same reason. Since $\|e_p\| \leq \|e\| = \sqrt{n}$, the result follows. □

As long as $\|e_p\| \geq 1 > \Delta$, (29) implies

$$\frac{(x^{k+1})^T s}{(x^k)^T s} \leq 1 - \gamma \frac{1 - \Delta}{n + \Delta\sqrt{n}},$$

so that we might expect at least convergence with ratio $(1 - \text{constant}/n)$. In fact, we can prove this for the primal gap, using the following extension of theorem 2.1 of Barnes, Chopra and Jensen [4]:

Proposition 7 *Suppose* $\|\bar{d}_\alpha\| = \|\alpha \bar{c}_p + e_p\| \leq \Delta \leq .5$ *and* $\alpha > 0$. *Then*

$$\frac{c^T x^{k+1} - c^T x^*}{c^T x^k - c^T x^*} \leq 1 - \frac{\gamma}{4n}. \tag{30}$$

Proof: From (25) and (26), we only need to show that $\|e - X^{-1} x^*\| \leq 4n$. Now let

$$\bar{s} := \frac{1}{\alpha} X^{-1}(e - e_p + \alpha \bar{c}_p) \geq \frac{1}{\alpha}(1 - \Delta) X^{-1} e > 0$$

and note that $(X\bar{s})_p = \bar{c}_p$ so $\bar{s}$ is a feasible dual slack. It follows that the duality gap between x^* and $\bar{s}$ is no greater than that between x^k and $\bar{s}$, so that

$$\begin{aligned} \frac{1}{\alpha}(1 - \Delta) e^T X^{-1} x^* \leq \bar{s}^T x^* &\leq \bar{s}^T x^k = \frac{1}{\alpha} e^T (e - e_p + \alpha \bar{c}_p) \\ &\leq \frac{1}{\alpha}(n + \Delta\sqrt{n}). \end{aligned}$$

Since $X^{-1} x^* \geq 0$, this shows that $\|X^{-1} x^*\| \leq \frac{n + \Delta\sqrt{n}}{1 - \Delta} \leq 3n$; since $\|e\| = \sqrt{n} \leq n$, this gives the desired inequality. □

Unfortunately, (30), together with theorem 1, if we assume we start with a point x^0 that gives $\|\bar{d}_\alpha\| < .5$ and that all α's are positive, only yields a bound of $O(nt)$ iterations to attain precision t, as in the results of Barnes, Chopra and Jensen [4].

References

[1] I. Adler and R. D. C. Monteiro, "Limiting behaviour of the affine scaling continuous trajectories for linear programming problems," *Mathematical Programming* **50** (1991), 29–51.

[2] K. Anstreicher, "Strict Monotonicity and Improved Complexity in the Standard Form Projective Algorithm for Linear Programming," CORE discussion paper, CORE, Université Catholique de Louvain (Louvain-la-Neuve, Belgium) (1990).

[3] E. R. Barnes, "A variation on Karmarkar's algorithm for solving linear programming problems," *Mathematical Programming* **36** (1986), 174–182.

[4] E. R. Barnes, S. Chopra and D. L. Jensen, "A polynomial time version of the affine algorithm," Manuscript, IBM T.J. Watson Research Center, Yorktown Heights, NY (1988).

[5] I. I. Dikin, "Iterative solution of problems of linear and quadratic programming," *Doklady Academiia Nauk SSSR* **174** (1967), 747–748.

[6] P. E. Gill, W. Murray, M. A. Saunders, J. A. Tomlin and M. H. Wright, "On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method," *Mathematical Programming* **36** (1986), 183–209.

[7] J.-L. Goffin and J.-P. Vial, "Short steps with Karmarkar's projective algorithm for linear programming," Manuscript, Département d'Economie Commerciale et Industrielle, University of Geneva (Geneva) (1990).

[8] N. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica* **4** (1984), 373-395.

[9] C. B. Moler, J. Little, S. Bangert and S. Kleiman, *Pro-Matlab User's Guide*, (MathWorks, Sherborn, MA., 1987).

[10] C. Roos and J.-P. Vial, "A polynomial method of approximate centers for linear programming," Report, Delft University of Technology (1988), to appear in *Mathematical Programming*.

[11] M. J. Todd, "A low complexity interior point algorithm for linear programming," Technical report 903, School of Operations Research and Industrial Engineering, Cornell University (Ithaca, NY) (1990), to appear in *SIAM Journal on Optimization*.

[12] D. Xiao and D. Goldfarb, "A path-following projective interior point method for linear programming," Manuscript, Department of IE/OR, Columbia University, New York (1990).

[13] Y. Ye, "An O(n^3L) potential reduction method for linear programming," *Mathematical Programming* **50** (1991), 239–258.