Journal of the Operations Research Society of Japan
Vol. 40, No. 3, September 1997

A THREE STEP QUADRATICALLY CONVERGENT VERSION OF PRIMAL AFFINE SCALING METHOD

Romesh Saigal
The University of Michigan

(Received February 16, 1995; Final August 19, 1996)

Abstract In this paper we consider the primal affine scaling method and show that, asymptotically, a step selection strategy exists which can be viewed as a predictor-corrector method. We investigate two step selection strategies. In the first, one corrector step is taken between each pair of predictor steps, and we call this the 2-step method. In the other two such steps are taken, which we call the 3-step method. We show that the 2-step method attains a superlinear rate of 1.5 while the 3-step method attains a quadratic convergence rate. This improves upon the work of Tsuchiya and Monteiro, who obtain a 2-step rate of 1.3.

1. Introduction

We consider here the linear programming problem:

$$\begin{array}{rcl} \text{minimize} & c^T x & \\ & Ax = b & \\ & x \geq 0 & \end{array} \tag{1.1}$$

with its dual

$$\begin{array}{rcl} \text{maximize} & b^T y & \\ & A^T y + s = c & \\ & s \geq 0 & \end{array} \tag{1.2}$$

where A is an $m \times n$ matrix and b and c are appropriate vectors. We assume that

Assumption 1. *The primal linear program has an interior solution.*

Assumption 2. *The objective function is not constant on the primal feasible region.*

Assumption 3. *The matrix A has rank m.*

In this paper we consider application of the primal affine scaling method for solving this problem. The primal method was proposed by Dikin [3] in 1967, who subsequently proved its convergence under the primal non-degeneracy assumption, Dikin [4]. His proof also appears in Vanderbei and Lagarias [21]. This method was re-discovered, Barnes [2], who proved its convergence under the non-degeneracy assumption on both the primal and the dual linear programs. In addition, several of its variants like the dual, Adler, Karmarkar, Resende and Veiga [1], and the primal-dual, Monteiro, Adler and Resende [11], were generated in the process of implementing the projective transformation method of Karmarkar [9].

The convergence behavior of the affine scaling method without the non-degeneracy assumption is now known. For example, Mascarenhas [10] has recently produced an example on which the method fails when α, the step size to the boundary in the affine scaling direction, is 0.999. Starting with the work of Tsuchiya [17] who introduced a local potential function to analyze the convergence of this method, significant developments have occurred. Dikin [5], using the local potential function, has shown the convergence of the primal sequence to the interior of the optimal primal face and the dual sequence to the analytic center of the optimal dual face for $\alpha \leq \frac{1}{2}$. Tsuchiya and Muramatsu [19] subsequently proved the same convergence behavior when $\alpha \leq \frac{2}{3}$. Simpler proofs of this result have been developed by Monteiro, Tsuchiya and Wang [12] and Saigal [15]. It is also known that the dual sequence may not converge when $\alpha > \frac{2}{3}$. Hall and Vanderbei [8] have produced an example where this happens. Saigal [15] and Gonzaga [7] have shown the convergence to optimality of the limit of the primal sequence and a cluster point of the dual sequence for a slightly larger step size of $\alpha \leq \frac{2q}{3q-1}$, where q is the number of zero components in the limit of primal sequence. It appears that this may be the largest step size for which convergence to optimality can be proved.

By establishing a connection between the affine scaling step and the Newton step, Tsuchiya and Monteiro [20] devise a strategy of adjusting step sizes under which the dual sequence converges to the analytic center for the optimal face and the primal sequence to the interior of the primal optimal face. Their method, asymptotically, attains a two step super-linear convergence rate of 1.3, and can be viewed as a predictor-corrector method. This paper builds up on their work and generates a different step selection strategy, which also can be viewed as a predictor-corrector method. This step selection strategy, asymptotically, attains a two step superlinear rate of 1.5 and a three step quadratic convergence rate. In the 3-step method, we take two corrector steps between each pair of predictor steps. Also, the primal converges to the interior of the optimal primal face and the dual to the analytic center of the optimal dual face.

This paper is organized as follows. Besides the introduction it has 3 other sections. In Section 2, we introduce the affine scaling method, and state some of its known properties. In Section 3 we relate the affine scaling step to Newton step generated when computing the analytic center of a certain polyhedron. In Section 4, we introduce the accelerated version and establish its convergence and convergence rate.

We now present the notation. Given a vector v, the largest component of v is denoted by $\phi(v)$, i.e., $\phi(v) = \max_i v_i$ and $\|v\|$ represents its 2-norm. e is a vector of appropriate size with each component equal to 1. Given a matrix A and a subset N we represent by

1. v_N the subvector of v composed of components indexed in N.
2. A_N the submatrix of A with columns indexed in N.

V represents the diagonal matrix generated by the corresponding components of v. k is the iteration counter. v^k, $k = 1, 2, \cdots$ is a sequence of vectors, which is also denoted by $\{v^k\}$. K denotes a subsequence and is a subset of the positive integers. Thus $\{v^k\}_{k\in K}$ is the subsequence of $\{v^k\}$ generated by K. $\{V_k\}$ is a sequence of matrices. If v^* is the limit of $\{v^k\}$, V_* represents the diagonal matrix generated by v^*. Thus $V^p_{*,N}$ represents the diagonal matrix generated by v^*_N raised to the power p.

2. The Affine Scaling Method

In this section we present the primal affine scaling method and known results (without proof) about the sequences generated by this method. We now present the method we will deal with in this paper:

Step 0 Let x^0 be an interior point solution, $1 > \sigma > 0$ and let $k = 0$.

Step 1 Tentative Solution to the Dual:

$$y^k = (AX_k^2A^T)^{-1}AX_k^2c$$

Step 2 Tentative Dual Slack:

$$s^k = c - A^T y^k$$

If $s^k \leq 0$ then STOP. The solution is unbounded.

Step 3 Min-Ratio test:

$$\begin{aligned}\theta_k &= \min\{\frac{\|X_k s^k\|}{x_j^k s_j^k} : s_j^k > 0\} \\ &= \frac{\|X_k s^k\|}{\phi(X_k s^k)}\end{aligned}$$

where $\phi(x) = \max_j x_j$.

Step 4 Step Size Selection: Choose, by some rule, the next step size $\sigma < \alpha_k < 1$. Also, if $\theta_k = 1$ set $\alpha_k = 1$.

Step 5 Next Interior Point:

$$x^{k+1} = x^k - \alpha_k \theta_k \frac{X_k^2 s^k}{\|X_k s^k\|}$$

Step 6 Iterative Step: If $x_j^{k+1} = 0$ for some j, then STOP. x^{k+1} is an optimal solution to the primal, y^k the optimum solution to the dual. Otherwise set $k = k + 1$ and go to step 1.

The method presented above is called the long step affine scaling method. This method can stop at steps 2 or 6. We now show that in this case, the problem is solved in a finite number of iterations.

Theorem 1. *$\{c^T x^k\}$ is strictly decreasing. Also, exactly one of the following holds:*

1. *The algorithm stops at Step 2. Then the linear program has an unbounded solution, i.e., its dual is infeasible.*
2. *The algorithm stops at Step 6. Then x^{k+1} is an optimal solution of the primal and y^k is an optimal solution of the dual.*
3. *The sequence $\{x^k\}$ is infinite and $\{c^T x^k\}$ is not bounded below. Then the linear program has an unbounded solution.*
4. *The sequence $\{x^k\}$ is infinite and $\{c^T x^k\}$ is bounded below. Then $\{c^T x^k\}$ converges to, say c^*.*

Proof: To see the first part, from Step 5, we note that

$$c^T x^{k+1} = c^T x^k - \alpha_k \theta_k \frac{c^T X_k^2 s^k}{\|X_k s^k\|}.$$

As can be established from the definitions, $x^k > 0$ and $\theta_k \geq 1$. Also,

$$\begin{aligned}c^T X_k^2 s^k &= c^T X_k^2 (c - A^T y^k) \\ &= c^T X_k (I - X_k A^T (A X_k^2 A^T)^{-1} A X_k) X_k c \\ &= \|P_k X_k c\|^2\end{aligned}$$

where $P_k = I - X_k A^T (A X_k^2 A^T)^{-1} A X_k$ is the projection matrix into the null space $\mathcal{N}(AX_k)$ of the matrix AX_k. Now, by a simple calculation, we see that $\|P_k X_k c\| = \|X_k s^k\|$ and thus we have

$$c^T x^{k+1} = c^T x^k - \alpha_k \theta_k \|X_k s^k\|. \tag{2.1}$$

From Assumption (2) the subtracted term in the above formula is non-zero.

To see part 1, we note that for $s^k \leq 0$, x^{k+1} remains strictly positive for every $\alpha > 0$, and thus $c^T x^{k+1} \to -\infty$ as $\alpha \to \infty$.

To see part 2, let $x_l^{k+1} = 0$. Then, from Step 5 we see that

$$0 = x_l^{k+1} = x_l^k - \alpha_k \frac{(x_l^k)^2 s_l^k}{\phi(X_k s^k)}$$

and thus $1 = \alpha_k \frac{x_l^k s_l^k}{\phi(X_k s^k)}$. So $\alpha_k = 1$ and $x_l^k s_l^k = \phi(X_k s^k) \geq 0$. It then follows that $s_l^k \geq 0$ and, from Step 4, that $\theta_k = 1$. Thus $x_l^k s_l^k = \|X_k s^k\|$. Hence, for every $j \neq l$ $x_j^k s_j^k = 0$, and so $s_j^k = 0$ and $x_j^{k+1} = x_j^k > 0$. Thus $s^k \geq 0$ and $x^{k+1} \geq 0$ satisfy the conditions of the complementary slackness theorem.

Part 3 follows from the monotonicity of $\{c^T x^k\}$, and part 4 from the fact that every bounded monotone sequence converges. ■

We will henceforth make the following assumption:

Assumption 4. *The sequence $\{x^k\}$ is infinite and the sequence $\{c^T x^k\}$ is bounded below.*

We now state two results whose proofs can be found in the cited references:

Proposition 2. *Let Assumption (4) hold. Then*

1. $X_k s^k \to 0$.
2. $\{x^k\}$ *converges, say to* x^*.
3. *There is a* $\delta > 0$ *such that for each* $k = 1, 2, \cdots$

$$\frac{c^T x^k - c^*}{\|x^k - x^*\|} \geq \delta.$$

Proof: The proof can be found in Tsuchiya [17], Monteiro, Tsuchiya and Wang [12] and Saigal [15]. ■

Given that the sequence $\{x^k\}$ converges, let the limit of this sequence be x^*. Define:

$$\begin{aligned} N &= \{j : x_j^* = 0\} \\ B &= \{j : x_j^* > 0\} \\ q &= |N| \\ u^k &= \frac{X_k s^k}{c^T x^k - c^*} \\ v^k &= \frac{x^k - x^*}{c^T x^k - c^*}. \end{aligned} \tag{2.2}$$

We now state known properties of the sequences $\{y^k\}$, $\{s^k\}$, $\{u^k\}$ and $\{v^k\}$. Their proof can be found in cited references.

Proposition 3. *Consider the sequences $\{y^k\}$, $\{s^k\}$, $\{u^k\}$ and $\{v^k\}$.*

1. *They are bounded.*
2. *There is an $L \geq 1$, $\rho_1 > 0$ and $\rho_2 > 0$ such that for all $k \geq L$*
   - (a) $e^T u_N^k = 1 + \delta_k$ *where* $|\delta_k| \leq \rho_1 (c^T x^k - c^*)^2$.
   - (b) $\|s_B^k\| \leq \rho_2 (c^T x^k - c^*)^2$.
   - (c) $\phi(u^k) = \phi(u_N^k) \geq \frac{1}{2q}$.
   - (d) $\|u^k\|^2 > \frac{1}{2q}$.

Proof: The proof can be found in [15]. ■

3. Affine Scaling and Newton's Method

In this section we will establish a relationship between Newton step for computing the analytic center of a certain polyhedron and affine scaling step. We will use this to show that step size α can be chosen so that, asymptotically, the affine scaling step gets arbitrarily close to the Newton step to analytic center. This then allows development of a step selection strategy in which this analogy generates a "corrector step" towards the analytic center of this polyhedron. Higher order convergence is obtained by taking large values (close to 1) for α when sufficiently close to the analytic center of this polyhedron; otherwise value determined by the analogy to Newton's method, to get close to the analytic center again. In this section, we first introduce this polyhedron, then investigate Newton's method for computing the analytic center of the polyhedron. We then establish the required connection.

3.1. Two Polyhedrons

Consider B and N defined by (2.2). Define the affine set

$$\mathcal{F}_{\mathcal{D}} = \{(y,s) : A^T y + s = c, s_B = 0\}$$

which represents all possible dual solutions (not necessarily feasible) which are complementary to x^*. First observe that since it contains all cluster points of bounded sequence $\{(y^k, s^k)\}$, $\mathcal{F}_{\mathcal{D}} \neq \emptyset$.

Let $(\bar{y}, \bar{s}) \in \mathcal{F}_{\mathcal{D}}$ be arbitrary, but fixed. Then for any pair of primal feasible solutions x and $\hat{x}$,

$$\begin{aligned} c^T(x-\hat{x}) &= (\bar{s} + A^T\bar{y})^T(x-\hat{x}) \\ &= \bar{s}^T(x-\hat{x}) \\ &= \bar{s}_N^T(x_N - \hat{x}_N). \end{aligned}$$

Consider $\{v^k\}$ defined by (2.2). From Proposition 3, part 1, sequence $\{v^k\}$ is bounded. Also, from above identity, it belongs to the polyhedron $\mathcal{V} = \{v : Av = 0, \bar{s}_N^T v_N = 1, v_N \geq 0\}$.

Let $\mathcal{V}_N = \{v_N : v \in \mathcal{V}\}$. Starting at a given v_N^k, asymptotically, Newton's step for computing the analytic center of this polyhedron is related to the affine scaling step $v_N^{k+1} - v_N^k$. The latter is determined by α_k. We explore this relationship in subsections that follow. There is also a close relationship between the analytic centers of the two polyhedrons $\mathcal{F}_{\mathcal{D}} \cap \{s : s_N \geq 0\}$ and $\mathcal{V}_N$, which we now explore.

The analytic center of $\mathcal{F}_{\mathcal{D}} \cap \{s : s_N \geq 0\}$ is defined by the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & \textstyle\sum_{j\in N} \log(s_j) \\ & A_N^T y + s_N = c_N \\ & A_B^T y = c_B \\ & s_N > 0 \end{aligned}$$

with its K.K.T. conditions:

$$S_N^{-1} e - v_N = 0 \quad (3.1)$$

$$A_N v_N + A_B v_B = 0 \quad (3.2)$$

$$A_N^T y + s_N = c_N \quad (3.3)$$

$$A_B^T y = c_B \quad (3.4)$$

$$s_N > 0$$

while the analytic center of $\mathcal{V}_N$ is defined by:

$$\begin{array}{lrcl} \text{maximize} & \sum_{j\in N}\log(v_j) & & \\ & A_N v_N + A_B v_B & = & 0 \\ & \bar{s}_N^T v_N & = & 1 \\ & v_N & > & 0 \end{array}$$

with its K.K.T. conditions:

$$V_N^{-1}e - A_N^T y - \bar{s}_N\theta = 0 \tag{3.5}$$

$$-A_B^T y = 0 \tag{3.6}$$

$$A_N v_N + A_B v_B = 0 \tag{3.7}$$

$$\bar{s}_N^T v_N = 1 \tag{3.8}$$

$$v_N > 0.$$

The following relates the two analytic centers defined by systems (3.5)–(3.8) and (3.1)–(3.4).

Proposition 4. *Analytic center (y^*, s^*) of $F_\mathcal{D}\cap\{s : s_N \geq 0\}$ exists if and only if analytic center v_N^* of $\mathcal{V}_N$ exists. Also, for some $\rho^* > 0$*

$$\rho^* v_N^* = S_{*,N}^{-1}e$$

Proof: If v_N^* is the analytic center of $\mathcal{V}_N$, then for some $\theta = \theta^* > 0$, $v_B = v_B^*$ and $y = u^*$, equations (3.5)–(3.8) are satisfied. Thus

$$s_N = \frac{1}{\theta^*}A_N^T u^* + \bar{s}_N, v = \theta^* v^*, y = \bar{y} - \frac{1}{\theta^*}u^*$$

satisfy the equations (3.1)–(3.4). Also, if (y^*, s^*) is the analytic center of $F_\mathcal{D}\cap\{s : s_N \geq 0\}$, then for some v^* equations (3.1)–(3.4) are satisfied. Thus

$$\theta = \bar{s}_N^T s_N^*, v = \frac{1}{\theta}v^*, y = -\theta(y^* - \bar{y})$$

solve equations (3.5)–(3.8). The result follows from equation (3.1). ∎

3.2. The Affine Scaling Step

We will study the affine scaling step with a view to relating it to a Newton step for computing an analytic center of $\mathcal{V}_N$. As is now well known (see for example, Barnes [2]) the affine scaling direction used in Step 5 of the method, is generated by solving the following Ellipsoidal Approximating Problem:

$$\begin{array}{lrcl} \text{minimize} & c^T x & & \\ & Ax & = & b \\ & \|X_k^{-1}(x - x^k)\| & \leq & 1. \end{array}$$

Substituting $p = x^k - x$ we obtain the equivalent problem:

$$\begin{array}{lrcl} \text{maximize} & c^T p & & \\ & Ap & = & 0 \\ & \|X_k^{-1}p\| & \leq & 1. \end{array}$$

Now $c^Tp = \bar{s}^Tp = \bar{s}_N^T p_N$ for $(\bar{y}, \bar{s}) \in F_{\mathcal{D}}$. Thus, for fixed $(\bar{y}, \bar{s})$, the above problem is equivalent to

$$\begin{array}{lrcrcl} \text{maximize} & \bar{s}_N^T p_N & & & & \\ & A_N p_N & + & A_B p_B & = & 0 \\ & p_N^T X_{k,N}^{-2} p_N & + & p_B^T X_{k,B}^{-2} p_B & \leq & 1 \end{array}$$

By noting that the solution of this problem is on the boundary of the ellipsoid (i.e., the second constraint is at equality), the K.K.T. conditions for this problem (with $\theta > 0$) are

$$\bar{s}_N - A_N^T y - 2\theta X_{k,N}^{-2} p_N = 0 \tag{3.9}$$

$$-A_B^T y - 2\theta X_{k,B}^{-2} p_B = 0 \tag{3.10}$$

$$A_N p_N + A_B p_B = 0 \tag{3.11}$$

$$\|X_k^{-1} p\| = 1 \tag{3.12}$$

The solution to the system (3.9)–(3.12) is the following: $p^k = \frac{X_k^2 s^k}{\|X_k s^k\|}$, $\bar{y}^k = (AX_k^2A^T)^{-1}AX_k^2\bar{s}$ $= y^k - \bar{y}$ and $2\theta^k = \|X_k s^k\|$. $\bar{y}^k$ is obtained by multiplying (3.9) by $A_N X_{k,N}$, (3.10) by $A_B X_{k,B}$, adding and substituting (3.11). p^k and θ_k are then obtained by substituting $\bar{y}^k$ into (3.9), (3.10) and (3.13).

Now, let the sequences $\{u^k\}$ and $\{v^k\}$ be as defined in (2.2). Letting, for some given k, u and v represent u^k and v^k respectively, we define $\hat{A}_N = A_N V_N$ and $\hat{s}_N = V_N \bar{s}_N$, where V_N is the diagonal matrix whose jth diagonal entry is v_j for each $j \in N$. Consider the system:

$$\frac{u_N}{\|u\|^2} - \hat{A}_N^T \tilde{y} - \frac{\rho}{\|u\|^2}\hat{s}_N = 0 \tag{3.13}$$

$$-A_B^T \tilde{y} = -\frac{s_B^k}{\|u\|^2} \tag{3.14}$$

$$\hat{A}_N \frac{u_N}{\|u\|^2} + A_B p_B' = 0 \tag{3.15}$$

$$\frac{\hat{s}_N^T u_N}{\|u\|^2} = 1 \tag{3.16}$$

The following proposition establishes a connection between the conditions represented by the above systems (3.9)–(3.12) and (3.13)–(3.16).

Proposition 5. *Consider the systems represented by the equations (3.9)–(3.12) and (3.13)–(3.16).*

1. *(3.9)–(3.12) have a unique solution which generates a solution to (3.13)–(3.16).*
2. *The solution to (3.13)–(3.16) is unique up to a choice of p_B'; and, there is a value for p_B' for which the resulting solution also solves (3.9)–(3.12).*
3. *When A_B has full column rank, the two systems are equivalent.*

Proof: Since the Equations (3.9) - (3.12) represent the solution to a strictly convex problem, they have a unique solution. Using this solution, define the vectors

$$\tilde{y} = \frac{-y(c^Tx^k - c^*)}{2\theta\|u\|}$$

$$\rho = \frac{(c^Tx^k - c^*)\|u\|}{2\theta}$$

$$p_B' = \frac{p_B}{(c^Tx^k - c^*)\|u\|}.$$

It is confirmed, by simple algebra, that u, $\tilde{y}$ ρ and p'_B solve the System (3.13)–(3.16). Thus we have proved Part 1.

From part 1, it follows that (3.13)–(3.16) have a solution. Considering $q_N = \frac{u_N}{\|u\|^2}$, $\tilde{\rho} = \frac{\rho}{\|u\|^2}$, $\tilde{y}$ and p'_B as variables, this system is linear in these variables. If A_B has full column rank, the solution to (3.13)–(3.16) is unique, and part 3 follows. Otherwise, since (3.13)–(3.16) can have a solution only if s^k_B lies in the row space $\mathcal{R}(A^T_B)$ of A_B, the third condition must have redundant constraints which are identified by choosing any full column rank submatrix of A_B.

To see part 2, let $A_B = (A_C, A_D)$ where A_C has full column rank and spans the range (or column space) $\mathcal{R}(A_B)$ of A_B. Thus, for some unique matrix Λ, $A_D = A_C\Lambda$. Replacing equations (3.14) and (3.15) by

$$-A^T_C\tilde{y} = \frac{s^k_C}{\|u\|^2} \tag{3.17}$$

and

$$\hat{A}_N\frac{u_N}{\|u\|^2} + A_C p'_C = 0 \tag{3.18}$$

respectively, we obtain a new system that has a unique solution. By setting $p'_B = (p'_C, p'_D)$, and letting $p'_D = 0$, the solution to Equation (3.18) generates a solution to (3.15). Now, let $(q_N, \tilde{y}, \tilde{\rho}, p'_B)$ be any solution to (3.13)–(3.16). This then generates the unique solution $(q_N, \tilde{y}, \tilde{\rho}, p'_C - \Lambda p'_D)$ to (3.13) and (3.16)–(3.18). Since only the vector p'_B is modified in any solution to (3.13)–(3.16), part 2 is established with the required $p'_B = \frac{p_B}{(c^Tx^k - c^*)\|u\|^2}$. ■

A consequence of Proposition 5 is that though p_B is determined uniquely by affine scaling method, when a full column rank submatrix A_C of A_B is substituted in its place, the resulting solution only changes the value of p'_B. It turns out that p'_B plays no role in the asymptotic analysis of the affine scaling method.

3.3. Analytic Center and Newton's Method

In this section we consider the application of Newton's method for finding the analytic center of $\mathcal{V}_N$.

Solution to the K.K.T. conditions (3.5)–(3.8) is unique, if it exists. Analytic center problem has a feasible solution, but the set may not be bounded, and thus the center may not exist. Indeed, it can be shown that, for a given N, the center exists if and only if x^* is an optimal solution of the primal linear program. Since we have not shown this fact, we cannot claim the K.K.T. conditions have a solution. None the less Newton's method can be applied to these conditions, and its convergence properties investigated. In case the analytic center of $\mathcal{V}_N$ does not exist, Newton's method will not converge. In each of the results we derive, whenever needed, we will make the explicit assumption about the existence of the analytic center in the hypothesis of the result. However, such results can only be used after this existence has been established.

We now apply Newton's method to the system of equations (3.5)–(3.8) to determine its zero. The Newton direction $(\Delta v, \Delta y, \Delta\theta)$ at (v, y, θ) is obtained by solving the following system:

$$\begin{aligned}
-V_N^{-2}\Delta v_N - A^T_N\Delta y - \bar{s}_N\Delta\theta &= -V_N^{-1}e + A^T_N y + \bar{s}_N\theta && (3.19)\\
-A^T_B\Delta y &= A^T_B y && (3.20)\\
A_N\Delta v_N + A_B\Delta v_B &= 0 && (3.21)\\
\bar{s}^T_N\Delta v_N &= 0 && (3.22)
\end{aligned}$$

Defining $\hat{y} = y + \Delta y$, $\hat{\theta} = \theta + \Delta\theta$ and substituting $w_N = V_N^{-1}\Delta v_N$, $\hat{A}_N = A_N V_N$ and $\hat{s}_N = V_N \bar{s}_N$ we can rewrite the system (3.19)–(3.22) as:

$$\begin{aligned} w_N + \hat{A}_N^T\hat{y} + \hat{s}_N\hat{\theta} &= e & (3.23)\\ A_B^T\hat{y} &= 0 & (3.24)\\ \hat{A}_N w_N + A_B\Delta v_B &= 0 & (3.25)\\ \hat{s}_N^T w_N &= 0 & (3.26)\end{aligned}$$

We substitute A_C, a full column rank submatrix of A_B, in the system (3.23)–(3.26) and note that it is linear in the variables w_N, Δv_C, $\hat{y}$ and $\hat{\theta}$, with the underlying matrix:

$$M(v_N) = \begin{bmatrix} I & \hat{A}_N^T & \hat{s}_N & 0 \\ 0 & A_C^T & 0 & 0 \\ \hat{A}_N & 0 & 0 & A_C \\ \hat{s}_N^T & 0 & 0 & 0 \end{bmatrix}.$$

This matrix is non-singular for every $v_N > 0$. As seen in the proof of Proposition 5, solution to the system (3.23) - (3.26) is unique up to a choice of Δv_B. The result below, derived from this analysis, relates to the sequence $\{v_N^k\}$ in $\mathcal{V}_N$, and is thus applicable to the sequence generated by the affine scaling method.

We state, without proof, the following standard result on the convergence and convergence rate of Newton's method.

Lemma 6. *Let $\{v_N^k\}$ be a sequence in $\mathcal{V}_N$ that converges to v_N^*, the analytic center of $\mathcal{V}_N$. Then there is an $L \geq 1$ and $\rho_1 > 0$, $\rho_2 > 0$ such that for all $k \geq L$*

1. $\frac{\|\Delta v_N^k\|}{\|v_N^k - v_N^*\|} = 1 + \delta_k$ *where* $|\delta_k| \leq \rho_1\|v_N^k - v_N^*\|$.
2. $\|v_N^k + \Delta v_N^k - v_N^*\| \leq \rho_2\|v_N^k - v_N^*\|^2$.

Now consider the affine scaling step as determined by the system (3.13)–(3.16). We note that if we consider $\frac{u_N}{\|u\|^2}$, p'_C, $\tilde{y}$ and $\frac{\rho}{\|u\|^2}$ as variables, this system is also linear with the underlying matrix $M(v_N)$. Thus we can prove the following proposition:

Proposition 7. *There exist $L \geq 1$ and $\rho > 0$ such that for every choice of C and all $k \geq L$,*

$$e - w_N^k - \frac{u_N^k}{\|u^k\|^2} = \Delta^k$$

with $\|\Delta^k\| \leq \rho\|c^T x^k - c^*\|^2$.

Proof: Consider the systems (3.13)–(3.16) and (3.23)–(3.26). In the latter system, make the change of variable $w'_N = e - w_N$. Then $\hat{s}_N^T w'_N = 1$. Now, by defining $v'_C = v_C - \Delta v_C$ and

$$\begin{aligned} a_1 &= w'_N - \frac{u_N}{\|u\|^2}\\ a_2 &= \tilde{y} + \hat{y}\\ a_3 &= \frac{\rho}{\|u\|^2} + \hat{\theta}\\ a_4 &= p'_C + v'_C\end{aligned}$$

generate the system $M(v_N)a = (0, \frac{s_C}{\|u\|^2}, 0, 0)^T$. This system is seen, with $\bar{a}_1 = V_N^{-1}a_1$ and $s'_C = \frac{s_C}{\|u\|^2}$, as the K. K. T. conditions of the following optimization problem:

$$\begin{array}{llcll} \text{minimize} & \frac{1}{2}\bar{a}_1^T V_N^{-2}\bar{a}_1 & + & (s'_C)^T a_4 & \\ & A_N\bar{a}_1 & + & A_C a_4 & = 0 \\ & \bar{s}_N^T\bar{a}_1 & & & = 0. \end{array}$$

Here, a_2 and a_3 are the respective Lagrange multipliers of two constraints. Consider the first constraint. Since A_C has full column rank, by using the formula (generated by the first constraint)

$$a_4 = -(A_C^T A_C)^{-1} A_C^T A_N \bar{a}_1$$

we can eliminate a_4 to generate the following equivalent problem:

$$\begin{array}{lll} \text{minimize} & \frac{1}{2}\bar{a}_1^T V_N^{-2} \bar{a}_1 + s_N^{*T} \bar{a}_1 & \\ & A_N^* \bar{a}_1 & = 0 \\ & \bar{s}_N^T \bar{a}_1 & = 0 \end{array}$$

where $s_N^* = -A_N^T A_C (A_C^T A_C)^{-1} s_C'$ and $A_N^* = A_N - A_C(A_C^T A_C)^{-1} A_C^T A_N$. The K.K.T. conditions of this problem are:

$$\begin{aligned} V_N^{-2}\bar{a}_1 + s_N^* - (A_N^*)^T a_2' - \bar{s}_N a_3' &= 0 \\ A_N^* \bar{a}_1 &= 0 \\ \bar{s}_N^T \bar{a}_1 &= 0 \end{aligned}$$

and we obtain the solution $(a_2', a_3')^T = (\bar{A}_N V_N^2 \bar{A}_N)^{-1} \bar{A}_N V_N^2 s_N^*$ where $\bar{A}_N^T = ((A_N^*)^T, \bar{s}_N)$. From Theorem 4, Saigal [15], $\|(a_2', a_3')\|$ is bounded above by $q(A_N^*, \bar{s}_N)\|s_N^*\|$ where $q(A_N^*, \bar{s}_N)$ is a positive constant independent of the diagonal matrix V_N. Thus, for some $\bar{q}(A, \bar{s}_N, C) > 0$,

$$\|(a_2', a_3')\| \leq \frac{\bar{q}(A, \bar{s}_N, C)\|s_C\|}{\|u\|^2}.$$

Thus, for $\beta' = \|(A_N^T, \bar{s}_N)\|\bar{q}(A_N, \bar{s}_N) + \|A_N^T A_C (A_C^T A_C)^{-1}\|$, from the first relation of the K.K.T. conditions, we see that $\|a_1\| = \|V_N \bar{a}_1\| \leq \|V_N^3((A_N^*)^T a_2' + \bar{s}_N a_3' - s_N^*)\| \leq \frac{\beta'\|s_C\|\|v_N\|^3}{\|u\|^2}$. Since $\|s_C\| \leq \|s_B\|$ for every choice of C, our result follows from Proposition 3 parts 1, 2(b) and 2(d) by choosing C that gives β' its largest value. ■

3.4. Affine Scaling and Newton Directions in $\mathcal{V}_N$

In this section, we show the connection between the Newton direction $\Delta v_N^k = V_{k,N} w_N^k$ at v_N^k studied in Proposition 7, and the affine scaling direction as interpreted in $\mathcal{V}_N$. Consider the sequence $\{v_N^k\}$ in $\mathcal{V}_N$ generated by the affine scaling algorithm. Then we can show the following result:

Proposition 8. *The affine scaling direction at v_N^k in $\mathcal{V}_N$ is*

$$v_N^{k+1} - v_N^k = \frac{\alpha_k \delta(u^k)}{1 - \alpha_k \delta(u^k)} (v_N^k - V_{k,N} \frac{u_N^k}{\|u^k\|^2})$$

where $\delta(u^k) = \frac{\|u^k\|^2}{\phi(u^k)}$. Also, the Newton direction Δv_N^k at v_N^k in $\mathcal{V}_N$ is:

$$\Delta v_N^k = v_N^k - V_{k,N} \frac{u_N^k}{\|u^k\|^2} - V_{k,N} \Delta^k$$

where Δ^k is as in Proposition 7.

Proof: Since $w_N^k = V_{k,N}^{-1} \Delta v_N^k$ the formula for the Newton direction follows from Proposition 7. To see the affine direction note that using step 5 and definitions we obtain:

$$v_N^{k+1} - v_N^k = \frac{x_N^{k+1}}{c^T x^{k+1} - c^*} - \frac{x_N^k}{c^T x^k - c^*}$$

$$= \frac{v_N^k - \frac{\alpha_k V_{k,N} u_N^k}{\phi(u^k)}}{1 - \frac{\alpha_k \|u^k\|^2}{\phi(u^k)}} - v_N^k$$

$$= \frac{\frac{\alpha_k \|u^k\|^2}{\phi(u^k)}}{1 - \frac{\alpha_k \|u^k\|^2}{\phi(u^k)}} (v_N^k - V_{N,k} \frac{u_N^k}{\|u^k\|^2}) \qquad (3.27)$$

and we are done. ■

If one is able to choose $\alpha_k = \frac{1}{2\delta(u^k)}$, the scalar term in the formula for the affine scaling direction becomes equal to 1, and thus the direction taken by Newton's method for computing the analytic center of $\mathcal{V}_N$ becomes, asymptotically, close to the affine scaling direction. Using Proposition 8, we will show that a sensible choice for α_k exists so that the scalar term in formula (3.27) becomes close to 1. It is this observation that allows the interpretation of this step as a corrector step.

4. Accelerated Primal Affine Scaling Method

Sequence $\{v_N^k\}$, generated by affine scaling method, lies in the polyhedron $\mathcal{V}_N$. We see from Proposition 8 that, by a choice of an appropriate value of α_k (which is a good estimate of $\frac{1}{2\delta(u^k)}$), constant $\frac{\alpha_k \delta(u^k)}{1-\alpha_k \delta(u^k)}$ can be made close to one. Then, the affine scaling step, asymptotically, gets close to the Newton step for computing the analytic center v_N^* of $\mathcal{V}_N$. The basic idea now is to take aggressive, or large, values of α_k when v_N^k is determined to be close to v_N^*. We call this the predictor step. And, when v_N^k is 'far' from v_N^*, take smaller values (which approach 0.50) that make the constant in affine scaling step approach one. We call this the corrector step. Another variation we treat here is in the number of corrector steps taken between each pair of predictor steps. We investigate situations when either one (which we call the two step method) or two such steps (which we call the three step method) are taken, and show that the asymptotic rates of convergence attained are respectively, superlinear at 1.5 and quadratic. We now introduce the accelerated method.

4.1. The Method

The accelerated primal affine scaling method is generated by replacing steps 3 and 4 of the method described in Section 2 by the following steps: let $0 < \alpha < 1$, typically close to 0.95, and $\sigma = \frac{1}{3}$,

Step 3′ Min-Ratio Test:

$$\theta_k = \min\{\frac{\|X_k s^k\|}{x_j^k s_j^k} : s_j^k > 0\}$$
$$= \frac{\|X_k s^k\|}{\phi(X_k s^k)}.$$

If $\theta_k = 1$ set $\alpha_k = 1$, and go to Step 5.

Step 4′ Step Size Selection: If $e^T X_k s^k \geq 1$, set $\alpha_k = \alpha$ and go to Step 5. Otherwise, define

$$N_k = \{j : x_j^k \leq \sqrt{(x^k)^T s^k}\}$$
$$\gamma_k = e^T X_{k,N_k} s_{N_k}^k$$
$$h_{N_k}^k = \frac{x_{N_k}^k}{\gamma_k} - \frac{X_{k,N_k}^2 s_{N_k}^k}{\|X_k s^k\|^2}$$
$$\epsilon_k = \|h_{N_k}^k\|$$

$$\rho_k = \frac{\log(\epsilon_k)}{\log(\gamma_k)}.$$

1. Predictor Step: If $\rho_k \geq 1.5$ then define

$$\tau_k = \begin{cases} \frac{\rho_k - 1}{2\rho_k} & \text{for 2 step method} \\ \frac{2\rho_k - 1}{3\rho_k} & \text{for 3 step method} \end{cases}$$

and

$$\alpha_k = \max\{\frac{1}{3}, 1 - \epsilon_k^{\tau_k}\} = \max\{\frac{1}{3}, 1 - \gamma_k^{\tau_k \rho_k}\}.$$

2. Corrector Step: Otherwise,

$$\alpha_k = \max\{\frac{1}{3}, \min\{\frac{\gamma_k \phi(X_{k,N_k} s_{N_k}^k)}{2\|X_k s^k\|^2}, \frac{2}{3}\}\}.$$

Some comments are in order here. Note that from Theorem 1, $c^T X_k^2 s^k = \|X_k s^k\|^2$. The step size implemented during the acceleration phase of the method is never less than $\frac{1}{3}$, and thus the propositions of section 2 apply to the generated sequences. Also, we see from Proposition 3 part 2(a) that $e^T u_N^k = 1 + \delta_k$ and thus

$$h_N^k = \frac{1}{1 + \delta_k} v_N^k - V_{k,N} \frac{u_N^k}{\|u_N\|^2}.$$

As we shall subsequently see (Lemma 10 part 4) γ_k is of the order $O(c^T x^k - c^*)$. Using Proposition 8, it can be seen that (see proof of part 1 Lemma 10) h_N^k computed in step 4$'$, is a very good estimate of the Newton step Δv_N^k, and its magnitude ϵ_k can be used to estimate the distance from v_N^k to the analytic center v_N^* of $\mathcal{V}_N$. Asymptotically, we apply a predictor step when the size, ϵ_k of the Newton step is of the order $O(c^T x^k - c^*)^2$; and the corrector step otherwise. Since γ_k is of the order $O(c^T x^k - c^*)$, ρ_k is an estimate of p where $O(c^T x^k - c^*)^p$ is a measure of this distance. τ_k is computed in such a way that after the appropriate number of corrector steps, subsequent ρ_k becomes close to 2. As is well known about the Newton step, $\|\Delta v_N^k\|$ is a very good estimate of $\|v_N^k - v_N^*\|$. During the corrector step, α_k is chosen so that

$$v_N^{k+1} - v_N^k = \Delta v_N^k + O(c^T x^k - c^*)^2$$

and thus the affine scaling step behaves, asymptotically, like a Newton step. Because $v_N^{k+1} - v_N^k$ and Δv_N^k can only be estimated to within $O(c^T x^k - c^*)^2$, ρ_k cannot be guaranteed to be greater than 2, and thus higher rates of convergence cannot be guaranteed. We now state the main theorem we will prove about the convergence properties of this accelerated method:

Main Theorem *Let the sequences $\{x^k\}$, $\{y^k\}$ and $\{s^k\}$ be generated by the accelerated method, and let assumptions (1)-(4) hold. Then, there exist vectors x^*, y^* and s^* such that*

1. $x^k \longrightarrow x^*$
2. $y^k \longrightarrow y^*$
3. $s^k \longrightarrow s^*$

where x^* lies in the relative interior of the optimal face of the primal, and (y^*, s^*) is the analytic center of the optimal face of the dual. In addition, asymptotically, the sequence $\{c^T x^k - c^T x^*\}$ converges to zero as follows:

1. *The rate of convergence of the two step method is superlinear with a rate of* 1.5.
2. *The rate of convergence of the three step method is quadratic.*

We now prove a sequence of lemmas that will be used in the proof of this theorem.

Lemma 9. *There exists an $\hat{L} \geq 1$ such that for every $k \geq \hat{L}$, $N_k = N$.*

Proof: From Proposition 3, there is an $L \geq 1$ such that for all $k \geq L$, from part 2(b) $|(x_B^k)^T s_B^k| \leq \rho_2 \|x_B^k\|(c^T x^k - c^*)^2$, and from part 2(a) $(x_N^k)^T s_N^k = (c^T x^k - c^*)(1+\delta_k)$ where $|\delta_k| \leq \rho_1 (c^T x^k - c^*)^2$. Thus there is an $\bar{L} \geq L$ such that for all $k \geq \bar{L}$, $(x^k)^T s^k = (x_B^k)^T s_B^k + (x_N^k)^T s_N^k \geq 0.50(c^T x^k - c^*)$. Also from Proposition 2 part 3, $\delta x_j^k \leq (c^T x^k - c^*)$ for each $j \in N$. Since $x_B^* > 0$ and $X_k s^k \longrightarrow 0$, there is an $\hat{L} \geq \bar{L}$ such that for all $k \geq \hat{L}$, and $j \in B$

$$x_j^k > \sqrt{(x^k)^T s^k}$$

and, for each $j \in N$,

$$\begin{aligned} \sqrt{(x^k)^T s^k} &\geq \sqrt{0.50}(c^T x^k - c^*)^{0.50} \\ &\geq \tfrac{1}{\delta}(c^T x^k - c^*) \\ &\geq x_j^k \end{aligned}$$

and we are done. ∎

Lemma 10. *Let $\hat{L}$ be as in Lemma 9. There exists an $L \geq \hat{L}$, $\theta_1 > 0$ and $\theta_2 > 0$ such that for all $k \geq L$*

1. $\|h_N^k - \Delta v_N^k\| \leq \theta_1 (c^T x^k - c^*)^2$.
2. $\frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} = 1 - \alpha_k \delta(u_N^k)$ *where* $\delta(u_N^k) = \frac{\|u^k\|^2}{\phi(u_N^k)}$.
3. $1 - \|w_N^k\| - |\Delta^k| \leq \delta(u_N^k) \leq 1$.
4. $0.50(c^T x^k - c^*) \leq (x_N^k)^T s_N^k \leq 1.5(c^T x^k - c^*)$.

Proof: The following identity is derived from Proposition 8 and definitions:

$$h_N^k - \Delta v_N^k = \frac{-\delta_k v_N^k}{1+\delta_k} - V_{k,N}\Delta^k$$

and part 1 follows from Proposition 3 parts 1 and 2(a), where δ_k is as in Proposition 3 part 2(a).

From steps 3 and 5 and definitions, we obtain

$$\frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} = 1 - \alpha_k \frac{\|u^k\|^2}{\phi(u^k)} \tag{4.1}$$

and part 2 follows from Proposition 3 part 2(c).

Since part 2 holds for every $\alpha_k \leq 1$, choosing $\alpha_k = 1$ gives the upper bound of part 3. The lower bound follows from Proposition 7,

$$\begin{aligned} \frac{\phi(u_N^k)}{\|u^k\|^2} &= \phi(\frac{u_N^k}{\|u^k\|^2}) \\ &= \phi(e + w_N^k - \Delta^k) \\ &\leq 1 + \|w_N^k\| + \|\Delta^k\| \end{aligned} \tag{4.2}$$

and the fact that $\frac{1}{1+x} \geq 1 - x$. Part 4 follows from the fact that $(x_N^k)^T s_N^k = (c^T x^k - c^*) e^T u_N^k$, and Proposition 3, part 2(a). ∎

The next proposition shows that the step chosen during the corrector step is converging to $\frac{1}{2}$, when the sequence $\{v_N^k\}$ is converging to the analytic center v_N^* of $\mathcal{V}_N$.

Proposition 11. *Let $w_N^k \to 0$ on some subsequence K. Then, on K*

$$a_k = \frac{\gamma_k \phi(X_{k,N} s_N^k)}{2\|X_k s^k\|^2} = \frac{e^T u_N^k \phi(u_N^k)}{2\|u^k\|^2} \to \frac{1}{2}$$

Proof: From Proposition 3 part 2(a) and upper bound of part 3 of Lemma 10, $2a_k \geq e^T u_N^k \geq 1 - |\delta_k|$. From equation (4.2)

$$a_k \leq \frac{1+\delta_k}{2}(1 + \|w_N^k\| + \|\Delta^k\|)$$

and we get the result as δ_k and $\|\Delta^k\|$ go to zero. ■

4.2. The Predictor and Corrector Steps

We now investigate the predictor step under the assumption that the analytic center v_N^* exists and the iterate v_N^k is close to it.

Lemma 12. *Let the analytic center $v_N^* > 0$ of $\mathcal{V}_N$ exist and $\hat{L}$ be as in Lemma 9. There exist $L \geq \hat{L}$ and $\pi > 0$ such that for all $k \geq L$ for which $\|v_N^k - v_N^*\| \leq \beta(c^T x^k - c^*)^p$ with $p \geq 0.25$*

$$\|w_N^k\| \leq \frac{1.5\beta}{\pi}(c^T x^k - c^*)^p.$$

Proof Let $\pi = \frac{\min_{j\in N} v_j^*}{2} > 0$ and define $L \geq \hat{L}$ such that for all $k \geq L$, $\pi \geq \beta(c^T x^k - c^*)^{0.25}$ and

$$0.50\|v_N^k - v_N^*\| \leq \|\Delta v_N^k\| \leq 1.5\|v_N^k - v_N^*\|. \tag{4.3}$$

The inequality (4.3) follows from part 1 of Lemma 6. Now, for each $j \in N$,

$$v_j^k \geq v_j^* - |v_j^* - v_j^k| \geq 2\pi - \beta(c^T x^k - c^*)^p \geq \pi.$$

Thus $\|w_N^k\| = \|V_{k,N}^{-1} \Delta v_N^k\| \leq \frac{1.5}{\pi}\|v_N^k - v_N^*\|$ and we are done. ■

The next proposition investigates the predictor step (Step 4, part 1) of the accelerated method. This result also assumes that the analytic center v_N^* exists, and some iterates get close to it. This is implicit in the hypothesis of the proposition.

Proposition 13. *Let $L \geq 1$ be as in Lemma 12. Assume that for some $k \geq L$ and $\beta > 0$, $\|v_N^k - v_N^*\| \leq \beta(c^T x^k - c^*)^p$ for some $1 < p \leq 2$, and $\alpha_k > \frac{1}{3}$. Then*

1. *There are constants $\theta_1 > 0$ and $\theta_2 > 0$ such that*

$$\theta_1(c^T x^k - c^*)^{1+\tau_k\rho_k} \leq c^T x^{k+1} - c^* \leq \theta_2(c^T x^k - c^*)^{1+min\{p,\tau_k\rho_k\}}.$$

2. *There is a $\theta_3 > 0$ such that*

$$\|v_N^{k+1} - v_N^*\| \leq \theta_3(c^T x^{k+1} - c^*)^{\frac{(1-\tau_k)\rho_k}{1+\tau_k\rho_k}}.$$

Proof To see the first part, note that from step 4′, part 1 and Lemmas 10 and 12 we get

$$\begin{aligned}
0.50^{\tau_k\rho_k}(c^T x^k - c^*)^{\tau_k\rho_k} &\leq ((x_N^k)^T s_N^k)^{\tau_k\rho_k} \\
&\leq 1 - \alpha_k \\
&\leq 1 - \alpha_k \frac{\|u^k\|^2}{\phi(u_N^k)} \\
&= \frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} \\
&\leq 1 - (1 - \gamma_k^{\tau_k\rho_k})(1 - \|w_N^k\| - |\delta_k|) \\
&\leq \gamma_k^{\tau_k\rho_k} + \|w_N^k\| + |\delta_k| \\
&\leq \theta_2(c^T x^k - c^*)^{\min\{p,\tau_k\rho_k\}}
\end{aligned}$$

and the first part follows. To see the second part, from Proposition 8, definitions and some straightforward manipulation, we see that

$$h_N^k = \Delta v_N^k - \frac{\delta_k}{1+\delta_k} v_N^k + V_{k,N}\Delta_k$$

and

$$v_N^{k+1} - v_N^k = \Delta v_N^k + t^k$$

where

$$t^k = \frac{2\alpha_k \delta(u^k) - 1}{1-\alpha_k\delta(u^k)}(h_N^k + \frac{v_N^k \delta_k}{1+\delta_k}) + V_{k,N}\Delta_k$$

where δ_k is as in Proposition 3. From Lemma 10 part 3, $\delta(u_N^k) \leq 1$, $\|h_N^k\| = \epsilon_k = \gamma_k^{\rho_k} \leq (1.5)^{\rho_k}(c^Tx^k - c^*)^{\rho_k}$ and $1 - \alpha_k = \gamma_k^{\tau_k\rho_k}$. Thus, for $|\delta_k| \leq 0.50$, and some $\theta > 0$

$$\begin{aligned} \|t^k\| &\leq \frac{1}{\gamma_k^{\tau_k\rho_k}}(\gamma_k^{\rho_k} + 2\|v_N^k\||\delta_k|) + \|v_N^k\|\|\Delta^k\| \\ &\leq \theta(c^Tx^k - c^*)^{(1-\tau_k)\rho_k}. \end{aligned}$$

Thus $v_N^{k+1} - v_N^* = v_N^k + \Delta v_N^k - v_N^* + t^k$. For some $\mu_1 > 0$ and $2p > (1-\tau_k)\rho_k$ from part 2 of Lemma 6 we have

$$\begin{aligned} \|v_N^{k+1} - v_N^*\| &\leq \|v_N^k + \Delta v_N^k - v_N^*\| + \|t^k\| \\ &\leq \mu_1\|v_N^k - v_N^*\|^2 + \theta(c^Tx^k - c^*)^{(1-\tau_k)\rho_k} \\ &\leq \beta\mu_1(c^Tx^k - c^*)^{2p} + \theta(c^Tx^k - c^*)^{(1-\tau_k)\rho_k} \\ &\leq \theta_3(c^Tx^{k+1} - c^*)^{\frac{(1-\tau_k)\rho_k}{1+\tau_k\rho_k}} \end{aligned}$$

the last inequality of which follows from Proposition 13 part 1, and we are done. ■

We now investigate the corrector step, again with the assumption that the analytic center exists.

Proposition 14. *Let L be as in Lemma 10, and assume that for some $k \geq L$ and $\beta > 0$, $\|v_N^k - v_N^*\| \geq \beta(c^Tx^k - c^*)^p$ with $p \leq 1$ and $\alpha_k > \frac{1}{3}$. In case*

1. $0.50 \leq p \leq 1$, *then a corrector step will be taken, after which, for some $\theta_1 > 0$*

$$\|v_N^{k+1} - v_N^*\| \leq \theta_1(c^Tx^{k+1} - c^*)^{2p}.$$

2. $0.25 \leq p < 0.5$, *then at least one corrector step will be taken, and after at most two steps, for some $\theta_2 > 0$*

$$\|v_N^{k+2} - v_N^*\| \leq \theta_2(c^Tx^{k+2} - c^*)^{4p}.$$

Proof: From parts 1 of Lemmas 6 and 10, we note that for some $\sigma_1 > 0$, $\epsilon_k \geq \sigma_1(c^Tx^k - c^*)^p$. Also, from part 4 of Lemma 10, $\gamma_k \leq 1.50(c^Tx^k - c^*)$. Thus ρ_k is close to p, and so a corrector step will be taken. Let

$$a_k = \frac{\gamma_k\phi(X_{k,N}s_N^k)}{2\|X_ks^k\|^2} = \frac{e^Tu_N^k\phi(u_N^k)}{2\|u^k\|^2}.$$

From part 2 of Proposition 3 and Equation 4.2, $a_k \leq \frac{1+\delta_k}{2}(1 + \|w_N^k\| + \|\Delta^k\|)$. Thus from Lemma 12,

$$\frac{1}{2} - |\delta_k'| \leq a_k \leq \frac{1}{2} + |\rho_k'|$$

where, for some $\sigma_2 > 0$ and $\sigma_3 > 0$, $|\delta'_k| \leq \sigma_2(c^Tx^k - c^*)^2$ and $|\rho'_k| \leq \sigma_3(c^Tx^k - c^*)^p$. Thus, for all large k, $\frac{1}{3} < a_k < \frac{2}{3}$, and so $\alpha_k = a_k$. Now, from part 2(a) of Proposition 3,

$$\begin{aligned}\alpha_k\delta(u^k) &= \frac{(e^Tu_N^k)\phi(u_N^k)}{2\|u^k\|^2}\frac{\|u^k\|^2}{\phi(u^k)} \\ &= \frac{1+\delta_k}{2}.\end{aligned}$$

Using Proposition 8,

$$\begin{aligned}v_N^{k+1} - v_N^k &= \frac{1+\delta_k}{1-\delta_k}(v_N^k - V_{k,N}\frac{u_N^k}{\|u^k\|^2}) \\ &= \Delta v_N^k + t^k\end{aligned}$$

where

$$t^k = \frac{2\delta_k}{1-\delta_k}(v_N^k + V_{k,N}\frac{u_N^k}{\|u^k\|^2}) + V_{k,N}\Delta^k.$$

From Proposition 3 parts 1, 2(a), 2(c) and 2(d) and for some $\sigma_4 > 0$, $\|t^k\| \leq \sigma_4(c^Tx^k - c^*)^2$. Thus after one step with this α_k, we see from part 2 of Lemma 6

$$\begin{aligned}\|v_N^{k+1} - v_N^*\| &\leq \|v_N^k + \Delta v_N^k - v_N^*\| + \|t^k\| \\ &\leq \rho'\|v_N^k - v_N^*\|^2 + \sigma_4(c^Tx^k - c^*)^2 \\ &\leq \beta^*(c^Tx^k - c^*)^{2p}.\end{aligned}$$

From parts 2 and 3 of Lemma 10 we see that

$$\begin{aligned}\frac{c^Tx^{k+1} - c^*}{c^Tx^k - c^*} &= 1 - \alpha_k\delta(u^k) \\ &= \frac{1-\delta_k}{2}\end{aligned}$$

and, for all sufficiently large k such that $|\delta_k| \leq 0.50$, we obtain

$$0.25 \leq \frac{c^Tx^{k+1} - c^*}{c^Tx^k - c^*} \leq 0.75. \tag{4.4}$$

Part 1 follows with $\theta_1 = (0.25)^{2p}\beta^*$. Part 2 follows by taking one more corrector step, and using the same analysis as for part 1. ■

4.3. Proof of Main Theorem

We now give the proof of Main Theorem.

Proof: We first show that an infinite number of predictor steps are taken. Assume that only a finite number are taken. Then, there is an $L \geq 1$ such that for all $k \geq L$, a corrector step is taken. But by the definition of this step, the value assigned to α_k is less or equal to $\frac{2}{3}$ and greater than or equal to $\frac{1}{3}$. Thus the global convergence follows from the theorem of Tsuchiya and Muramatsu [19]. And the primal sequence converges to the interior of the optimal primal face and the dual sequence to the analytic center of the optimal dual face, $F_\mathcal{D} \cap \{s : s_N \geq 0\}$. Thus, from Proposition 4, v_N^* the analytic center of $\mathcal{V}_N$ exists, and $v_N^k \longrightarrow v_N^*$. Also, from Lemma 6, $\{v_N^k\}$ converges quadratically to v_N^*, and from Equation 4.4, $\{c^Tx^k\}$ converges no faster than linearly to c^*. Thus, for every $2 > p > 0$, there exists an $L \geq 1$ such that for every $k \geq L$, $\|v_N^k - v_N^*\| \leq (c^Tx^k - c^*)^p$. Now, from part 1 of Lemma

6 and parts 1 and 4 of Lemma 10, ρ_k will approach 2. This contradicts our assumption, since a predictor step will be taken.

For each $k \in K'$, let a predictor step be taken, and let $k \in K'$. rom Step 4′ we see that $\epsilon_k \leq \gamma_k^{1.5}$ and, from Proposition 3 part 2(a), since $\gamma_k = (1+\delta_k)(c^T x^k - c^*)$, we see that $\epsilon_k \longrightarrow 0$ for $k \in K'$. Thus

$$h_N^k = \frac{v_N^k}{1+\delta_k} - V_{k,N}\frac{u_N^k}{\|u^k\|^2} \longrightarrow 0 \text{ for } k \in K'.$$

Thus, from Proposition 8, $\Delta_N^k \longrightarrow 0$ for $k \in K'$ and thus the sequence $\{v_N^k\}_{k\in K'}$ converges to the analytic center v_N^*, and thus it exists.

Let k be sufficiently large at which a predictor step is performed. Since $\alpha_k \geq \frac{1}{3}$ for every iteration, Propositions 2 and 3 hold. Thus, during the predictor steps $\alpha_k \longrightarrow 1$ and a predictor step with $\alpha_k > \frac{1}{3}$ will be taken. From Proposition 13, we see that, for some $\theta_3 > 0$,

$$\|v_N^{k+1} - v_N^*\| \leq \theta_3(c^T x^{k+1} - c^*)^{\frac{(1-\tau_k)\rho_k}{1+\tau_k\rho_k}}.$$

By substituting $\tau_k = \frac{\rho_k - 1}{2\rho_k}$, we obtain

$$\|v_N^{k+1} - v_N^*\| \leq \theta_3(c^T x^{k+1} - c^*)$$

and after one corrector step will attain conditions of the predictor step. Also, by substituting $\tau_k = \frac{2\rho_k - 1}{3\rho_k}$, we obtain

$$\|v_N^{k+1} - v_N^*\| \leq \theta_3(c^T x^{k+1} - c^*)^{0.50}.$$

So, after at most two corrector steps, the conditions for the predictor step are satisfied. Now, the convergence rate of $\{c^T x^k - c^*\}$ is given by Proposition 13 as $1 + \tau_k\rho_k$. Thus, asymptotically, the convergence rates for the two and three step methods are $\frac{1+\rho_k}{2}$ and $\frac{2(1+\rho_k)}{3}$. From Lemma 6, the rates of convergence of $\{v_N^k\}$ and $\{c^T x^k\}$, Equation (4.4), part 1 of Lemma 10 and the definition of ρ_k, we note that ρ_k will become greater than or equal to 2, and our theorem follows. ■

4.4. Asymptotic efficiency of Acceleration

We now investigate the asymptotic efficiency of this acceleration scheme and show that the three step method maximizes a measure introduced by Ostrowski [14] section 6.11. This measure balances the greater work done to achieve the higher asymptotic rate of convergence. The simplest way to get order four convergence from a sequence generated by quadratically convergent Newton's method is to drop each odd element of the sequence. The convergence rate has increased but so has the work per iteration. Ostrowski's measure is invariant under such manipulations. For a method which requires w units of work per iteration and achieves a convergence rate of p, the measure of efficiency is defined as

$$\frac{\log(p)}{w}.$$

By choosing $\tau_k = \frac{\rho_k - a}{(1+a)\rho_k}$ for $a = \frac{1}{2^r}$, we obtain a convergence rate of $\frac{3.2^r}{1+2^r}$ for a $(r+2)$-step method, i.e., where $(r+1)$ corrector steps are taken between each pair of predictor steps. The table below gives the calculation of this efficiency:

Algorithm	Rate	Work/Iter	Efficiency	Factor
2 Step	1.5	$2w$	$\frac{0.2027}{w}$	0.8773
3 Step	2	$3w$	$\frac{0.23105}{w}$	1.00
4 Step	2.4	$4w$	$\frac{0.2188}{w}$	0.9471
Two step Quadratic	2	$2w$	$\frac{0.34657}{w}$	1.50

We note that, under this measure, the three step method is most efficient.

References

[1] I. Adler, N. K. Karmarkar, M. G. C. Resende and G. Veiga, "An implementation of Karmarkar's algorithm for linear programming," *Mathematical Programming* 44 (1989) 297–335.

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

[3] I. I. Dikin, "Iterative solution of problems of linear and quadratic programming," *Soviet Mathematics Doklady* 8 (1967) 674–675.

[4] I. I. Dikin, "On the convergence of an iterative process," *Upravlyaemye Sistemi* 12 (1974) 54–60. (In Russian) .

[5] I. I. Dikin, "The convergence of dual variables," Technical Report, Siberian Energy Institute, Irkutsk, Russia, December 1991.

[6] I. I. Dikin, "Determination of interior point of one system of linear inequalities," *Kibernetica and system analysis* 1 (1992) 74–96.

[7] C. C. Gonzaga, Private Communication, 1993.

[8] L. A. Hall and R. J. Vanderbei, "Two-thirds is sharp for affine scaling," *O. R. Letters* 13(1993) 197–201.

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

[10] W. F. Mascarenhas, "The affine scaling algorithm fails for $\lambda = 0.999$," Technical Report, Universidade Estadual de Campinas, Campinas S. P., Brazil, October 19, 1993.

[11] R. D. C. Monteiro, I. Adler and M. G. C. Resende, "A polynomial time primal-dual affine scaling algorithm for linear and convex quadratic programming," *Mathematics of Operations Research* 15 (1990) 191–214.

[12] R. D. C. Monteiro, T. Tsuchiya, and Y. Wang, "A simplified global convergence proof of the affine scaling algorithm," *Annals of Operations Research* 47 (1993) 443–482.

[13] M. Muramatsu and T. Tsuchiya, "A convergence analysis of a long-step variant of the projective scaling algorithm," Research Memorandum 454, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan, November 1992.

[14] A. M. Ostrowski, *Solution of Equations and Systems of Equations* , Academic Press, New York, (1960).

[15] R. Saigal "A simple proof of the primal affine scaling method," Technical Report No. 92-60, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, December 1992. (to appear *Annals of Operations Research*)

[16] P. Tseng and Z. Q. Luo,"On the convergence of the affine-scaling algorithm," *Mathematical Programming* 56 (1992) 301–319.

[17] T. Tsuchiya, "Global convergence of the affine-scaling methods for degenerate linear programming problems," *Mathematical Programming* 52 (1991) 377–404.

[18] T. Tsuchiya, "Global convergence property of the affine scaling method for primal

degenerate linear programming problems," *Mathematics of Operations Research* 17 (1992) 527–557.

[19] T. Tsuchiya and M. Muramatsu, "Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems," *SIAM Journal on Optimization*, 5(1995) 525-551.

[20] T. Tsuchiya and R. D. C. Monteiro "Superlinear Convergence of the Affine Scaling Algorithm," Research Memorandum, November 28, 1992.

[21] R. J. Vanderbei and J. C. Lagarias, "I. I. Dikin's convergence result for the affine-scaling algorithm," In J. C. Lagarias and M. J. Todd, editors, *Mathematical Developments Arising from Linear Programming: Proceedings of a Joint Summer Research Conference held at Bowdoin College, Brunswick, Maine, USA, June/July 1988* volume 114 of Contemporary Mathematics, pages 109–119. American Mathematical Society, Providence, Rhode Island, USA, 1990.

Romesh Saigal
Department of Industrial and Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109-2117, USA