EFFICIENT VARIANTS OF THE
AFFINE SCALING METHOD

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

Efficient Variants of the Affine Scaling Method

Romesh SAIGAL

Department of Industrial and Operations Engineering,

The University of Michigan,

Ann Arbor, Michigan 48109-2117, USA

August 20, 1993 (Revised October 22, 1993)

Abstract

In this paper we present a variant of the affine scaling method. This variant is defined by choosing a real $r > 0.5$, and is similar to the power barrier variant of the primal-dual homotopy methods, considered by den Hertog, Roos and Tarlaky and Sheu and Fang.

Here we analyse the variants for $r > 1$. The analysis for $0.50 < r < 1$ is similar, and can be readily carried out with minor modifications. Under the non-degeneracy assumption, we show that the variant converges for any choice of the step size α. To analyse the convergence without the non-degeneracy assumption we define a power center of a polytope. We use the connection of the computation of the power center by Newton's method and the steps of the variant, to generalize the 2/3rd result of Tsuchiya and Muramatsu. We show that with a constant step size α such that $\frac{\alpha}{(1-\alpha)^{2r}} < \frac{2}{2r+1}$; and with a variable asymptotic step size α_k uniformly bounded away from $\frac{2}{2r+1}$, the primal sequence converges to the relative interior of the optimal primal face, and the dual sequence converges to the power center of the optimal dual face.

We also present an accelerated version of the variant. We show that the two step superlinear convergence rate of the variant is $1 + \frac{r}{r+1}$, while the three step convergence rate is $1 + \frac{3r}{r+2}$. Using the measure of Ostrowski, we note that the three step method

for $r = 4$ is more efficient than the two step quadratically convergent method, which is the limit of the two step method as r approaches infinity.

Abbreviated title: Variants of affine scaling method

1 Introduction

We consider here the linear programming problem:

$$\text{minimize} \quad c^T x$$
$$Ax \;=\; b \tag{1}$$
$$x \;\geq\; 0$$

with its dual

$$\text{maximize} \quad b^T y$$
$$A^T y \;+\; s \;=\; b \tag{2}$$
$$s \;\geq\; 0$$

where A is a $m \times n$ matrix and b and c are appropriate vectors. We also 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 the application of the affine scaling method and its variants for solving this problem. The first (the primal method) was proposed by Dikin [6] in 1967, who subsequently proved its convergence under the non-degeneracy assumption, Dikin [7]. This method was rediscovered, Barnes [2], Karmarkar and Ramakrishna [12]; and, several of its variants like the dual and primal-dual, Vanderbei, Meketon and Freedman [25], were generated in the process of implementing the projective transformation method of Karmarkar [11]. The convergence of this method has been shown when the step size $\alpha \leq \frac{2}{3}$, Tsuchiya and Muramatsu [24], and simpler proofs of this result have been developed by Monterio, Tsuchiya and Wang [13], and Saigal [16]. In addition accelerated versions of this method, which attain superlinear convergence, have also been developed by Tsuchiya and Monterio [23] and Saigal [17]. In the convergence rate analysis, a connection between the iterates of Newton's method applied to computing an analytic center of certain polyhedron is exploited. A local potential function, which is the objective function of the analytic center problem, is investigated in the proof of convergence. And it is shown that the primal

sequence converges to the relative interior of the optimal primal face while the dual sequence to the analytic center of the optimal dual face.

For each $r > 0.5$, we will consider here the variant based on the following approximating problem:

$$\text{minimize} \quad c^T x$$
$$Ax \quad = \quad b \qquad\qquad (3)$$
$$\|X_0^{-r}(x - x^0)\| \quad \leq \quad 1$$

where $x^0 > 0$ is a given interior point of the linear program (1). When $r = 1$, the above approximating problem generates the (primal) affine scaling method. The variant thus generated by choosing $r > 0.5$ is analogous to the power barrier variant of the primal-dual homotopy (barrier) method of den Hertog, Roos and Terlaky [5] and Sheu and Fang [19]. Under the non-degeneracy assumption on the primal, we prove the convergence of the primal and the dual sequences to respective optimal solutions for any step size $\alpha < 1$. To prove convergence under degeneracy, we introduce the concept of the power center of a polytope. We define two polytopes, and their power centers are defined by the maximization of concave functions. These functions are related to each other in the same manner as the "dual norms" are. By using the connection of the iterates of Newton's method applied to computing the power center of the polytope associated with the primal problem, we prove two results. In the first result we consider the case of constant step size, and prove that if the step size α satisfies $\frac{\alpha}{(1-\alpha)^{2r}} < \frac{2}{2r-1}$, the primal sequence converges to the relative interior of the primal optimal face, and the dual to the power center of the optimal dual face. In the second result , which gives the same convergence behavior of the sequences as the first, we consider the case of the variable step size, and implement the step size α_k at iteration k. We choose the sequence $\{\alpha_k\}$ such that it is, asymptotically defined by $\frac{\alpha_k}{1-\alpha_k} < \frac{2}{2r-1}$; and the sequence is required to be uniformly bounded away from $\frac{2}{2r+1}$, which is 2/3 for $r = 1$. Our result can thus be considered as a generalization of the 2/3rd result of Tsuchiya and Muramatsu [24]. In both these cases, the proof is obtained by considering a merit function (which is the objective function of the center problem related to the primal), which plays the same role as the local potential function used in the analysis of the affine scaling method, Tsuchiya [21], Dikin [9] and Tsuchiya and Muramatsu [24].

By exploiting the relationship between the Newton and the affine iterates, we present here an accelerated version of the variant, which generalizes the accelerated version of Saigal [17]. We prove its convergence, and show that the primal sequence converges to the relative interior of the optimal primal face while the dual sequence converges to the power center of the optimal dual face. In addition, for each $r > 1$, we obtain the two-step superlinear convergence rate of $1 + \frac{r}{r+1}$; and the three-step rate of $1 + \frac{3r}{r+2}$. Using the measure of Ostrowski [15], we investigate the efficiency of each of these methods, and note that for each $r > 1$, the three-step version is more efficient than the two step version.

In this paper we restrict our attention to the values of $r > 1$, and note that with minor modifications in several formulae as well as the objective functions of the power center problems, our analysis carries over to the values of $0.50 < r < 1$ as well. Besides the introduction, this paper has 15 other sections. In section 2 we present the variant, in section 3 we prove the convergence of the primal sequence, in section 4 we prove some properties of the dual and the primal-dual sequences. In section 5, we prove the convergence of the sequences to optimality with the primal non-degeneracy assumption. In section 6 we introduce some translated sequences and prove some of their important properties. In section 7 we introduce the two power centers, and establish a connection between them. In section 8 we investigate the properties of Newton's method when computing the power center, and in section 9 we re-interpret the affine scaling steps in the same polytope. In section 10 we relate the Newton and affine iterates. In section 11 we introduce the merit function and prove the convergence to optimality without the primal non-degeneracy assumption and with a constant step size, while in section 12, we prove the same result with a variable step size. In section 13 we present the accelerated variant, and prove its convergence. In section 14 we derive the convergence rate of the accelerated variant while in section 15 we investigate its efficiency. Finally, we end the paper with concluding remarks in section 16.

2   The Variant

We now present the variant generated when $r > 0.5$.

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

4

Step 1 Tentative Solution to the Dual:

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

Step 2 Tentative Dual Slack:

$$s^k = c - A^Ty^k$$

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

Step 3 Min-Ratio test:

$$\theta_k = \min\{\frac{\|X_k^{2r-1}s^k\|}{(x_j^k)^{2r-1}s_j^k} : s_j^k > 0\}$$

$$= \frac{\|X_k^{2r-1}s^k\|}{\phi(X_k^{2r-1}s^k)}$$

where $\phi(t) = \max_j t_j$. If $\theta_k = 1$ set $\alpha = 1$.

Step 4 Next Interior Point:

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

Step 5 Iterative Step: If $x_j^{k+1} = 0$ for some j, then STOP. x^{k+1} is an optimal solution. Otherwise set $k = k + 1$ and go to step 1.

Let the above variant generate the infinite sequence $\{x^k\}$, and let the sequence $\{c^Tx^k\}$ be bounded below. In the contrary case, we can conclude that the dual of the linear program (1) is not feasible. We can then prove:

Theorem 1 *The sequence $\{c^Tx^k\}$ is monotone strictly decreasing; and thus converges, say to c^*.*

Proof: From Step 4, we note that

$$c^Tx^{k+1} = c^Tx^k - \alpha\theta_k\frac{c^TX_k^{2r}s^k}{\|X_k^{2r-1}s^k\|}.$$

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

$$c^TX_k^{2r}s^k = c^TX_k^{2r}(c - A^Ty^k)$$

$$= c^TX_k^r(I - X_k^rA^T(AX_k^{2r}A^T)^{-1}AX_k^r)X_k^rc$$

$$= \|P_kX_k^rc\|^2$$

5

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

$$c^T x^{k+1} = c^T x^k - \alpha \theta_k \frac{\| X_k^r s^k \|^2}{\| X_k^{2r-1} s^k \|}. \tag{4}$$

From Assumption (2) the subtracted term in the above formula is non-zero, and we are done since every bounded monotone sequence converges. ∎

3 Primal Sequence

In this section we consider some important properties of the primal sequence $\{x^k\}$ and also establish its convergence.

Consider the approximating problem defined, for $k = 0$, by (3). Setting $p = x^k - x$, we obtain the equivalent problem:

$$
\begin{aligned}
\text{maximize} \quad & c^T p \\
& Ap = 0 \\
& \| X_k^{-r} p \| \leq 1.
\end{aligned}
\tag{5}
$$

The following result is stated without proof, which can be found in the cited reference.

Theorem 2 *Let p^k solve the approximating problem (5). There is a $\rho > 0$ such that for each $k = 1, 2, \cdots,$*

$$c^T p^k \geq \rho \| p^k \|$$

Proof: See Corollary 6, Saigal [16]. ∎

We can now prove the convergence of the primal sequence.

Theorem 3 *Let $\{c^T x^k\}$ be bounded below. Then the sequence $\{x^k\}$ converges, say, to x^*.*

Proof: It is readily seen that the solution to the approximating problem (5) is

$$p^k = \frac{X_k^{2r} s^k}{\| X_k^r s^k \|}.$$

For each k, define $\gamma_k > 0$ such that $x^{k+1} - x^k = \gamma_k p^k$. From Theorem 2 we obtain the relation

$$\infty > c^T x^1 - c^* = \sum_{k=1}^{\infty} c^T (x^k - x^{k+1}) \geq \rho \sum_{k=1}^{\infty} \gamma_k \|p^k\| = \rho \sum_{k=1}^{\infty} \|x^{k+1} - x^k\|.$$

From the above relation, we see that the sequence $\{x^k\}$ is a Cauchy sequence, and thus converges to some vector x^*, and we are done. ∎

The following simple corollary is a consequence of Theorem 2.

Corollary 4 *There is a $\rho > 0$ such that for all $k = 1, 2, \cdots$*

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

Proof: Let k be arbitrary. The following relation is a consequence of Theorem 2 and the triangular inequality:

$$c^T x^k - c^* = \sum_{j=0}^{\infty} c^T (x^{k+j} - x^{k+j+1}) \geq \rho \sum_{j=0}^{\infty} \|x^{k+j+1} - x^{k+j}\| \geq \rho \|x^k - x^*\|$$

and we are done. ∎

Given that the sequence $\{x^k\}$ converges to x^*, define

$$\begin{aligned} B &= \{j : x_j^* > 0\} \\ N &= \{j : x_j^* = 0\}. \end{aligned} \tag{6}$$

4   Dual and Primal-Dual Sequences

In this section we consider the dual sequences $\{y^k\}$ and $\{s^k\}$ and the primal-dual sequence $\{X_k s^k\}$. We now investigate their properties.

The following result is well known and its proof can be found in the cited reference.

Theorem 5 *The sequences $\{y^k\}$ and $\{s^k\}$ are bounded.*

Proof: Follows from Step 1 and Theorem 4 of Saigal [16]. ∎

Now consider the sequence $\{X_k s^k\}$. We now prove that it converges to 0.

Theorem 6 $X_k s^k \longrightarrow 0$.

Proof: From Theorem 1, $\{c^T x^k\}$ converges, and thus from relation (4), we note that

$$\alpha \theta_k \frac{\|X_k^r s^k\|^2}{\|X_k^{2r-1} s^k\|} \longrightarrow 0,$$

where $\alpha > 0$ and $\theta_k \geq 1$. Since $\{x^k\}$ converges and $\{y^k\}$ and $\{s^k\}$ are bounded, the denominator of the above expression is bounded, thus

$$X_k^T s^k \longrightarrow 0.$$

But this is only possible if $X_k s^k \longrightarrow 0$, and we are done. ∎

5   Convergence of Dual Sequence

We now show that if the primal is non-degenerate, the dual sequences also converge, and the limit points are optimal for the respective problems. We do this in the next theorem.

Theorem 7 *Let the Assumptions (1) - (3) hold, and let the primal be non-degenerate. Then*

1. $x^k \longrightarrow x^*$,

2. $y^k \longrightarrow y^*$,

3. $s^k \longrightarrow s^*$,

where x^* is an optimal solution to the primal linear program, (y^*, s^*) is an optimal solution to the dual linear program.

Proof: Using Theorem 5, let (y^*, s^*) be a cluster point of the sequence $\{(y^k, s^k)\}$. From Theorem 6, $s_B^* = 0$. Thus

$$A_B^T y^* = c_B.$$

From the non-degeneracy assumption, A_B has full row rank m, and thus the above system has at most one solution. But each cluster point y^* of $\{y^k\}$ solves this system, thus the sequence has only one cluster point y^*, and so

$$y^k \longrightarrow y^*$$

and thus $s^k \longrightarrow s^*$. Now assume that for some $j \in N$, $s_j^* < 0$. Then there is an $L \geq 1$ such that for all $k \geq L$, $s_j^k < 0$. Thus, from Step 4

$$\begin{aligned}
x_j^{k+1} &= x_j^k - \alpha \frac{(x_j^k)^{2r} s_j^k}{\phi(X_k^{2r-1} s^k)} \\
&> x_j^k
\end{aligned}$$

and thus $x_j^k \not\to 0$, and we have a contradiction. Thus $s^* \geq 0$ and so (y^*, s^*) is dual feasible, and the theorem follows from the complementary slackness theorem. ∎

6 More on Sequences

Our aim is to investigate the convergence without the non-degeneracy assumption, as well as the convergence rate of the sequences generated by the variants, and, if possible, to accelerate the asymptotic convergence rate by controlling the step size selection, as per Tsuchiya and Monterio [23] and Saigal [17]. For this purpose we derive some important properties of the sequences generated by the variant. Consider the translated sequences:

$$\begin{aligned}
v^k &= \frac{x^k - x^*}{c^T x^k - c^*} \\
\tilde{u}^k &= \frac{X_k s^k}{c^T x^k - c^*} \\
u^k &= \frac{X_k^r s^k}{(c^T x^k - c^*)^r} \\
p^k &= X_k^{2r} s^k.
\end{aligned} \tag{7}$$

The following are simple consequences of the results already established.

Proposition 8 *The sequence $\{v^k\}$ is bounded.*

Proof: Follows from Corollary 4. ∎

Proposition 9 *The sequence $\{u_N^k\}$ is bounded.*

Proof: Follows from definition of u^k, and Corollary 4. ∎

Proposition 10 *For every $k = 1, 2, \cdots$, $Ap^k = 0$.*

Proof: Readily follows by substitution of definitions. ∎

Given B and N as defined by relations (6), we define the set of all possible dual estimates that are complementary to x^* as the polyhedron:

$$\mathcal{D} = \{(y, s) : A^T y + s = c, s_B = 0\}. \tag{8}$$

We can then prove:

Proposition 11 $\mathcal{D} \neq \emptyset$.

Proof From Theorem 5, the sequences $\{y^k\}$ and $\{s^k\}$ are bounded, thus on some common subsequence K, $y^k \to y^*$ and $s^k \to s^*$. Using Theorem 6, it is readily established that $(y^*, s^*) \in \mathcal{D}$. $\blacksquare$

Consider $(\bar{y}, \bar{s}) \in \mathcal{D}$. We can show that

Theorem 12 *For each* $k = 1, 2, \cdots$

1. $c^T x^k - c^* = \bar{s}_N^T x_N^k$.

2. There are constants $\alpha > 0$ *and* $\beta > 0$ *such that*

$$\alpha(c^T x^k - c^*) \leq \|x_N^k\| \leq \beta(c^T x^k - c^*).$$

3. $\bar{s}_N^T v_N^k = 1$

Proof Part 1 of the theorem follows from the following identity:

$$c^T x^k - c^* = \bar{s}^T(x^k - x^*) = \bar{s}_N^T x_N^k. \tag{9}$$

The upper bound of Part 2 follows from Corollary 4, and the lower bound from Part 1. Part 3 readily follows from the identity (9). $\blacksquare$

As a consequence of Theorem 12, we can define the polyhedron

$$\mathcal{P} = \{v : A_N v_N + A_B v_B = 0, \bar{s}_N^T v_N = 1, v_N \geq 0\} \tag{10}$$

and we note that the sequence $\{v^k\} \subset \mathcal{P}$.

We are now ready to prove two important results.

Proposition 13 *There exist* $\rho_1 > 0$ *and* $\rho_2 > 0$ *such that for every* $k = 1, 2, \cdots$

1. $\|X_k^r s^k\| \le \rho_1 \phi(x_N^k)^r$

2. $\|p_B^k\| \le \rho_2 \phi(x_N^k)^r \|X_{k,N}^r s_N^k\|.$

Proof: Using the argument of the proof of Theorem 1 and Proposition 10 , for $(\bar{y}, \bar{s}) \in \mathcal{D}$, we obtain:

$$
\begin{aligned}
\|X_k^r s^k\|^2 &= c^T p^k \\
&= (A^T \bar{y} + \bar{s})^T p^k \\
&= \bar{s}_N^T p_N^k \\
&= (X_{k,N}^r \bar{s}_N)^T (X_{k,N}^{-r} p_N^k) \\
&\le \|\bar{s}_N\| \phi(x_N^k)^r \|X_{k,N}^r s_N^k\|
\end{aligned}
$$

and the Part 1 follows with $\rho_1 = \|\bar{s}_N\|$. To see Part 2, note that from Theorem 2 for some $\rho > 0$, $\|p_B^k\| \le \|p^k\| \le \rho c^T p^k = \rho \bar{s}_N^T p_N^k \le \rho \|\bar{s}_N\| \phi(x_N^k)^r \|X_{k,N}^r s_N^k\|$ and the result follows from Part 1 with $\rho_2 = \rho \rho_1$ ∎

Proposition 14 *There are constants $\rho_1 > 0$, $\rho_2 > 0$ and $L \ge 1$ such that for all $k \ge L$*

1. $\|u_B^k\| \le \rho_1 (c^T x^k - c^*)^r \|u_N^k\|$

2. $\|s_B^k\| \le \rho_2 (c^T x^k - c^*)^{2r}.$

Proof: Since $x^k \to x^*$ and $x_B^* > 0$, there is a $\rho > 0$ and an $L \ge 1$ such that for all $k \ge L$, $\|X_{k,B}^{-r}\| \le \rho$. Let $k \ge L$. Part 1 follows from Corollary 4, Proposition 13 and

$$
\|u_B^k\| \le \frac{\|X_{k,B}^{-r}\| \|p_B^k\|}{(c^T x^k - c^*)^r}.
$$

To see part 2, not that . . .

$$
\|s_B^k\| \le \|X_{k,B}^{-r}\| \frac{\|X_{k,B}^r s_B^k\|}{(c^T x^k - c^*)^r} (c^T x^k - c^*)^r = \|X_{k,B}^{-r}\| \|u_B^k\| (c^T x^k - c^*)^r
$$

Substituting Part 1, our result follows from Proposition 9. ∎

11

7 Two Power Centers and their Relation

In this section we consider the situation when $r > 1$, and the polytopes $\mathcal{P}$ and $\mathcal{D} \cap \{s : s_N \geq 0\}$ defined by (10) and (8) respectively. We define the power center of $\mathcal{P}$ as the solution to the following concave maximization problem:

$$\text{maximize} \quad -\sum_{j \in N} v_j^{-2(r-1)}$$

$$
\begin{aligned}
A_N v_N + A_B v_B &= 0 \\
\bar{s}_N^T v_N &= 1 \\
v_N &> 0
\end{aligned}
\tag{11}
$$

where the K.K.T. conditions defining the center are:

$$2(r-1)V_N^{-(2r-1)} e - A_N^T z - \theta \bar{s}_N = 0 \tag{12}$$

$$-A_B^T z = 0 \tag{13}$$

$$A_N v_N + A_B v_B = 0 \tag{14}$$

$$\bar{s}_N^T v_N = 1. \tag{15}$$

We also define the power center of $\mathcal{D} \cap \{s : s_N \geq 0\}$ as the solution to the concave maximization problem:

$$\text{maximize} \quad \sum_{j \in N} s_j^{\frac{2(r-1)}{2r-1}}$$

$$
\begin{aligned}
A_N^T y + s_N &= c_N \\
A_B^T y &= c_B \\
s_N &> 0
\end{aligned}
\tag{16}
$$

where the K.K.T. conditions defining the center are:

$$\frac{2(r-1)}{2r-1} S_N^{-\frac{1}{2r-1}} e - u_N = 0 \tag{17}$$

$$A_N u_N + A_B u_B = 0 \tag{18}$$

$$A_N^T(y - \bar{y}) + (s_N - \bar{s}_N) = 0 \tag{19}$$

$$A_B^T(y - \bar{y}) = 0 \tag{20}$$

where $(\bar{y}, \bar{s})$ is an arbitrary element of $\mathcal{D}$, with $\bar{s}_B = 0$.

There is an intimate relationship between these two centers, and we explore this in the next theorem.

Theorem 15 (v^*, z^*, θ^*) *is a center of* $\mathcal{P}$ *if and only if* (s^*, u^*, y^*) *is a center of* $\mathcal{D} \cap \{s : s_N \geq 0\}$, *and*

$$s_N^* = \frac{2r-1}{\theta^*} V_{*,N}^{-2r+1} e.$$

Proof: Let (v^*, z^*, θ^*) be the solution to system (12) - (15). Then it can be verified that

$$
\begin{aligned}
s_N &= \frac{2r-1}{\theta^*} V_{*,N}^{-2r+1} e \\
y - \bar{y} &= -\frac{1}{\theta^*} z^* \\
u &= \frac{2(r-1)}{2r-1} \left(\frac{2r-1}{\theta^*}\right)^{-\frac{1}{2r-1}} v^*
\end{aligned}
$$

will solve the system (17) - (20).

Now let the center of $\mathcal{D} \cap \{s : s_N \geq 0\}$ exist and the system (17) - (20) have the solution (s^*, u^*, y^*). Then it can be verified that the transformation

$$
\begin{aligned}
\gamma &= e^T S_{*,N}^{\frac{2(r-1)}{2r-1}} e \\
\beta &= \frac{2r-1}{2\gamma(r-1)} \\
\theta &= 2(r-1)\gamma^{2r-1} \\
z &= -\theta(y^* - \bar{y}) \\
v &= \beta u^*
\end{aligned}
$$

will solve the system (12) - (15). Here we have used $\bar{s}_N^T u_N = s_N^T u_N$, which is readily established using equations (18) - (20). ∎

8 Newton's Method and Power Center of $\mathcal{P}$

In the previous section we have seen that the power center of $\mathcal{P}$ is determined by solving the system of equations (12) - (15). This is a non-linear system of equations, to which we can apply Newton's method to find its zero. The purpose of this section is to investigate this application.

Given v, y, θ, with $v \in \mathcal{P}$, the Newton direction $\Delta v, \Delta y, \Delta \theta$ is given by:

$$-2(r-1)(2r-1)V_N^{-2r}\Delta v - A_N^T \Delta y - \Delta \theta \bar{s}_N = -2(r-1)V_N^{-2r+1}e + A_N^T y + \theta \bar{s}_N \quad (21)$$

$$-A_B^T \Delta y = A_B^T y \tag{22}$$

$$A_N \Delta v_N + A_B \Delta v_B = 0 \tag{23}$$

$$\bar{s}_N^T \Delta v_N = 0. \tag{24}$$

Consider the change of variables:

$$
\begin{aligned}
\hat{y} &= \frac{y + \Delta y}{2(r-1)} \\
\hat{\theta} &= \frac{\theta + \Delta\theta}{2(r-1)} \\
\Delta v_B' &= (2r-1)\Delta v_B \\
w_N &= (2r-1)V_N^{-r}\Delta v_N.
\end{aligned}
\tag{25}
$$

Substituting this change in the system (21) - (24), we can derive the following equivalent system with $\hat{A}_N = A_N V_N^r$ and $\hat{s}_N = V_N^r \bar{s}_N$,

$$w_N + \hat{A}_N^T \hat{y} + \hat{s}_N \hat{\theta} = V_N^{-(r-1)} e \tag{26}$$

$$-A_B^T \hat{y} = 0 \tag{27}$$

$$\hat{A}_N w_N + A_B \Delta v_B' = 0 \tag{28}$$

$$\hat{s}_N^T w_N = 0 \tag{29}$$

We can then prove the following result:

Theorem 16 *Upto a choice of $\Delta v_B'$, the solution to the system (26) - (29) is unique. Also,*

1. $e^T V_N^{-(r-1)} w_N = w_N^T w_N.$

2. $\|w_N\|^2 \le e^T V_N^{-2(r-1)} e.$

Proof: The uniqueness follows from the fact that when the columns of the matrix A_B are linearly independent, the system has a unique solution. Otherwise, it is readily confirmed that a unique solution can be found by replacing A_B with a submatrix spanning the column space $\mathcal{R}(A_B^T)$ of A_B. This only changes the value of $\Delta v_B'$.

Multiplying equation (26) by w_N^T we obtain

$$w_N^T w_N + w_N^T \hat{A}_N^T \hat{y} + w_N^T \hat{s}_N \hat{\theta} = w_N^T V_N^{-(r-1)} e.$$

The second and the third terms of left hand side of the above expression, from equations (27) - (29), are readily seen as zero, and we have Part 1. To see Part 2, consider the optimization problem:

$$\text{maximize} \quad x^T x$$

$$x^T x \;=\; e^T V_N^{-(r-1)} x$$

whose solution is $x = V_N^{-(r-1)} e.$ $\blacksquare$

By an appropriate choice of a submatrix of A_B, we can guarantee that the system (26) - (29) is defined by a non-singular matrix, for every $v > 0$. In that case, the following is a well known result relating to the rate and the convergence of Newton's method, which we state without proof.

Assume that the sequence $\{v^k\}$ in $\mathcal{P}$ is converging to the power center v^*. Then the following can be proved about the Newton steps taken at the iterates v^k:

Theorem 17 *There is an $L \geq 1$ and constants $\rho_1 > 0$, $\rho_2 > 0$ such that for all $k \geq L$,*

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

2. $\|v_N^k + \Delta v_N^k - v_N^*\| \leq \rho_2 \|v_N^k - v_N^*\|^2$.

9 Affine Scaling Step in $\mathcal{P}$

The affine scaling step is defined by solving the optimization problem (5). Using Part 1 of Theorem 12 we can restate the problem (5) as:

$$\text{minimize} \quad \bar{s}_N^T p_N$$

$$A_N p_N \;+\; A_B p_B \;=\; 0$$

$$p_N^T X_{k,N}^{-2r} p_N \;+\; p_B^T X_{k,B}^{-2r} p_B \;\leq\; 1$$

The K.K.T. conditions for the above problem are

$$\bar{s}_N - A_N^T y - 2\theta X_{k,N}^{-2r} p_N = 0 \tag{30}$$

$$-A_B^T y - 2\theta X_{k,B}^{-2r} p_B = 0 \tag{31}$$

$$A_N p_N + A_B p_B = 0 \tag{32}$$

$$\|X_k^{-r} p\| = 1. \tag{33}$$

Using the definitions of u^k and v^k given in (7), and setting $u = u^k$, $v = v^k$ and

$$
\begin{aligned}
p'_B &= \frac{p_B^k}{(c^T x^k - c^*)^r \|u\|} \\
\tilde{y} &= \frac{-(c^T x^k - c^*)^r y^k}{2\theta \|u\|} \\
\tilde{\theta} &= \frac{(c^T x^k - c^*)^r \|u\|}{2\theta} \\
&= 1
\end{aligned}
\tag{34}
$$

we can rewrite the system (30) - (33) as:

$$
\frac{u_N}{\|u\|^2} - \hat{A}_N^T \tilde{y} - \hat{s}_N \frac{\tilde{\theta}}{\|u\|^2} = 0
\tag{35}
$$

$$
-A_B^T \tilde{y} = \frac{-s_B^k}{\|u\|^2}
\tag{36}
$$

$$
\hat{A}_N \frac{u_N}{\|u\|^2} + A_B p'_B = 0
\tag{37}
$$

$$
\hat{s}_N^T \frac{u_N}{\|u\|^2} = 1
\tag{38}
$$

where $\hat{A}_N = A_N V_N^r$ and $\hat{s}_N = V_N^r \bar{s}_N$.

Theorem 18 *Consider the systems represented by the equations (30) - (33) and (35) - (38).*

1. *(30) - (33) have a unique solution which generates a solution to (35) - (38).*

2. *The solution to (35) - (38) is unique upto a choice of p'_B; and, there is a value for p'_B for which the resulting solution also solves (30) - (33).*

3. *When A_B has full column rank, the two systems of equations are equivalent.*

Proof: Since the Equations (30) - (33) represent the solution to a strictly convex problem, they have a unique solution. It is now readily confirmed, by simple algebra, that u, $\tilde{y}$ ρ and p'_B, defined by the change of variables (34), solve the System (35) - (38). Thus we have proved Part 1.

From Part 1, it follows that (35) - (38) 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 (35) - (38) is unique, and Part 3 follows. Otherwise, since (35) - (38) can have a solution only if s_B lies in the row space $\mathcal{R}(A_B^T)$ of A_B, the third condition must have redundant constraints readily 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 (36) and (37) by

$$- A_C^T \tilde{y} = \frac{-s_B^k}{\|u\|^2} \tag{39}$$

and

$$\hat{A}_N \frac{u_N}{\|u\|^2} + A_C p_C' = 0 \tag{40}$$

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 (40) generates a solution to (37). Now, let $(q_N, \tilde{y}, \tilde{\rho}, p_B')$ be any solution to (35) - (38). This then generates the unique solution $(q_N, \tilde{y}, \tilde{\rho}, p_C' - \Lambda p_D')$ to (35) and (38) - (40). Since only the vector p_B' is modified in any solution to (35) - (38), Part 2 is established with the required $p_B' = \frac{p_B}{(c^T x^k - c^*)\|u\|}$. ∎

We can also prove the following important property of the sequences.

Theorem 19 *There exists a $\rho > 0$ and an $L \geq 1$ such that for all $k \geq L$*

$$e^T \tilde{u}_N^k = 1 + \delta_k$$

where $\delta_k \leq \rho(c^T x^k - c^*)^{2r}$.

Proof Multiplying equation (35) by $e^T V_{k,N}^{-(r-1)}$, we obtain

$$\frac{e^T \tilde{u}_N^k}{\|u^k\|^2} - e^T V_{k,N} A_N^T \tilde{y} - \frac{e^T V_{k,N} \bar{s}_N}{\|u^k\|^2} = 0.$$

Substituting $A_N v_N = -A_B v_B$, and equation (36), we obtain

$$e^T \tilde{u}_N^k + v_B^T s_B^k - 1 = 0.$$

The theorem now follows from Propositions 8 and 14. ∎

By making the change of variables

$$w_N' = V_N^{-(r-1)} e - w_N$$

$$\Delta \hat{v}_B = v_B - \Delta v_B'$$

in the system (26) - (29), we obtain the equivalent system:

$$w'_N - \hat{A}_N^T \hat{y} - \hat{s}_N \hat{\theta} = 0 \tag{41}$$

$$-A_B^T \hat{y} = 0 \tag{42}$$

$$\hat{A}_N w'_N + A_B \Delta \hat{v}_B = 0 \tag{43}$$

$$\hat{s}_N^T w'_N = 1. \tag{44}$$

Comparing systems (41) - (44) and (35) - (38), we note that if $\tilde{p}_N = \frac{u_N}{\|u\|^2}$,

$$a_1 = V_{k,N}^r (w'_N - \tilde{p}_N)$$

$$a_2 = \hat{y} - \tilde{y}$$

$$a_3 = \hat{\theta} - \frac{1}{\|u\|^2}$$

$$a_4 = \Delta \hat{v}_B - p'_B$$

and $b = (0, \frac{-s_B^k}{\|u\|^2}, 0, 0)^T$

$$M(v)a = b \tag{45}$$

where

$$M(v) = \begin{bmatrix} V_{k,N}^{-2r} & -A_N^T & \bar{s}_N & 0 \\ 0 & -A_B^T & 0 & 0 \\ A_N & 0 & 0 & A_B \\ \bar{s}_N^T & 0 & 0 & 0 \end{bmatrix}.$$

We now prove an important result relating solutions of the two systems.

Theorem 20 *There is an $L \geq 1$ and $\beta > 0$ such that for all $k \geq L$*

$$\frac{u_N^k}{\|u^k\|^2} = V_{k,N}^{-(r-1)} e - w_N^k + \Delta^k$$

with $\|\Delta^k\| \leq \beta \frac{\|s_B^k\|}{\|u\|^2}$.

Proof It is readily confirmed that the system (45) represents the K.K.T. conditions of the following quadratic programming problem:

$$\text{minimize} \quad a_1^T V_{k,N}^r a_1 \quad + \quad a_4^T \frac{s_B^k}{\|u\|^2}$$

$$A_N a_1 \quad + \quad A_B a_4 = 0$$

$$\bar{s}_N^T a_1 \qquad\qquad = 0.$$

where we can assume, because of Theorems 16 and 18, that the columns of A_B are linearly independent. Thus, we can substitute

$$a_4 = -(A_B^T A_B)^{-1} A_B^T A_N a_1$$

in the above optimization problem to obtain the equivalent problem

$$\text{maximize} \quad a_1^T V_{k,N}^r a_1 + a_1^T \tilde{s}_B^k$$
$$\tilde{A}_N a_1 \quad = 0$$
$$\tilde{s}_N a_1 \quad = 0$$

where $\tilde{A}_N = A_N - A_B(A_B^T A_B)^{-1} A_B^T A_N$ and $\tilde{s}_B^k = -A_N^T A_B (A_B^T A_B)^{-1} \frac{s_B^k}{\|u\|^2}$.

The above problem is a quadratic programming problem, and thus, its KKT multipliers are independent of the diagonal matrix $V_{k,N}$. Thus

$$\|(a_2, a_3)\| \leq q(A) \|\tilde{s}_B^k\|.$$

Since

$$V_{k,N}^{-r}(V_{k,N}^{-(r-1)} e - w_N^k - \frac{u_N^k}{\|u^k\|^2}) = A_N^T a_2 + \tilde{s}_N a_3$$

by defining

$$\Delta^k = -V_{k,N}^r (A_N^T a_2 + \tilde{s}_N a_3).$$

we obtain out result. ∎

10 Affine Scaling Step and Newton Step

We now investigate the Newton step and the affine scaling step in the projected polyhedron $\mathcal{P}_N = \{v_N : v \in \mathcal{P}\}$.

Consider the sequence $\{v_N^k\} \in \mathcal{P}_N$ generated by the affine scaling method. Then

Theorem 21 *The affine scaling direction at v_N^k is*

$$v_N^{k+1} - v_N^k = \frac{\tilde{\alpha}_k}{1 - \tilde{\alpha}_k}(v_N^k - V_{k,N} \frac{V_{k,N}^{r-1} u_N^k}{\|u^k\|^2})$$

where $\tilde{\alpha}_k = \alpha_k \frac{\|u^k\|^2}{\phi(V_k^{(r-1)} u^k)}$.

Proof: We note that

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

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

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

and we are done. ∎

We can also define the Newton step at v_N^k.

Theorem 22 *The newton step at v_N^k is*

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

Proof: The result follows from the change of variables (25), and Theorem 20. ∎

11 On Convergence of Dual Sequence

We now investigate the convergence to optimality of the primal sequence and the convergence of the dual sequence, without the non-degeneracy assumption. We now prove a lemma and a proposition:

Lemma 23 *There is an $L \geq 1$ and $\delta_1 > 0$ such that for all $k \geq L$*

$$\|u_N^k\| \geq \delta_1.$$

Proof: From Theorem 19 and definitions, $u_N^k = V_{k,N}^{r-1} \tilde{u}_N^k$, and for some $L' \geq 1$ and all $k \geq L'$, $e^T \tilde{u}_N^k = 1 + \delta_k$ with $\delta_k \leq \rho(c^T x^k - c^*)^{2r}$. Thus, there is an $L \geq L'$ such that for all $k \geq L$

$$\phi(\tilde{u}_N^k) \geq \frac{1 + \delta_k}{q} \geq \frac{1}{2q}.$$

Now assume that, on some subsequence K, $\|u_N^k\| \longrightarrow 0$. Then, for $k \in K$

$$V_{k,N}^{r-1} \tilde{u}_N^k \longrightarrow 0$$

20

or

$$v_j^{r-1}\tilde{u}_j^k \longrightarrow 0 \text{ for all } j \in N.$$

Let $\tilde{u}_{l_k}^k = \phi(\tilde{u}_N^k) \geq \frac{1}{2q}$. Thus on some subsequence $K' \subset K$, $l_k = l$. Now, for $k \in K'$, we note that $\tilde{u}_l^k \geq \frac{1}{2q}$. Thus

$$v_l^k \longrightarrow 0.$$

But $\tilde{u}_l^k = v_l^k s_l^k$. Thus $s_l^k \longrightarrow \infty$, a contradiction, as the Theorem 5 implies that the dual variables s are bounded, and we have our result. ∎

Proposition 24 *There is an $L \geq \hat{L}$ $\beta > 0$ and $\gamma > 0$ such that for every $k \geq L$*

1. $\frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} = 1 - \alpha\delta(u_N^k) \geq 0$ *where* $\delta(u_N^k) = \frac{\|u^k\|^2}{\phi(V_{k,N}^{r-1} u_N^k)}$.

2. $1 - \|w_N^k\| - |\delta_k| \leq \delta(u_N^k) \leq 1$ *with* $|\delta_k| \leq \beta(c^T x^k - c^*)^{2r}$.

3. $\phi(V_k^{r-1} u^k) = \phi(V_{k,N}^{r-1} u_N^k)$.

4. $\delta(u_N^k) \geq \gamma > 0$.

Proof: Part 1 follows readily by substitution, results of Theorem 8 and Proposition 14, Part 1. Now, to see Part 2, from Theorem 20,

$$\begin{aligned}
\frac{\phi(V_{k,N}^{r-1} u_N^k)}{\|u^k\|^2} &= \phi(\frac{V_{k,N}^{r-1} u_N^k}{\|u^k\|^2}) \\
&= \phi(e - w_N^k + \Delta^k) \\
&\leq 1 + \|w_N^k\| + \delta_k \quad\quad (46)
\end{aligned}$$

where $\|\Delta^k\| = \delta_k$; and the lower bound follows from Theorem 19, Lemma 23 for each real $a > -1$, $\frac{1}{1+a} \geq 1 - a$. The upper bound follows from the fact that part 1 holds for every $\alpha \leq 1$. Part 3 readily follows from Propositions 8 and 14 and Part 4 from Lemma 23 and the Propositions 8 and 9. ∎

We now show that, at least asymptotically, the sequence $\{c^T x^k\}$ converges linearly. We do this in the next lemma.

Lemma 25 *There is an $L \geq 1$ and $0 < \beta < 1$ such that for each $k \geq L$*

$$(c^T x^{k+1} - c^*) \leq \beta(c^T x^k - c^*),$$

and thus

$$\sum_{k=L}^{\infty} (c^T x^k - c^*) < \infty.$$

Proof: Follows from Proposition 24 Parts 1 and 4. ∎

When $r = 1$, the convergence to optimality is proved by using a merit function, called a local potential function, which is shown to decrease locally. This merit function also defines the analytic center. Here, we will use

$$F_N(x) = \sum_{j \in N} (v_j)^{-2(r-1)}$$

as the merit function in our analysis, and is the negative of the objective function of the power center problem on $\mathcal{P}$.

We now prove a simple lemma related to this function:

Lemma 26 *Let w and $v > 0$ be arbitrary p vectors with V the diagonal matrix generated by v.*

1. Let $\phi(-w) \le 0$. Then

$$\sum_{j=1}^{p} v_j((1 + w_j)^{-2(r-1)} - 1) \le -2(r-1)(e^T V w - \frac{(2r-1)}{2} w^T V w).$$

2. Let $1 > \phi(-w) > 0$. Then

$$\sum_{j=1}^{p} v_j((1 + w_j)^{-2(r-1)} - 1) \le -2(r-1)(e^T V w - \frac{(2r-1)}{2(1 - \phi(-w))^{2r}} w^T V w).$$

Proof: Part 1 follows from the following identity obtained by using the mean value version of Taylor's formula, and noting that $w_j \ge 0$ so $\hat{w}_j \ge 0$. This implies that the last term in the expression below is non-negative.

$$(1 + w_j)^{-2(r-1)} = 1 - 2(r-1)w_j + \frac{2(r-1)(2r-1)}{2} w_j^2 - \frac{2(r-1)(2r-1)2r}{3!}(1 + \hat{w}_j)^{-(2r+1)} w_j^3$$

Part 2 follows from the following relations, obtained by again applying the mean value version of Taylor's formula, where $\hat{w}_j \ge -\phi(-w)$:

$$(1 + w_j)^{-2(r-1)} = 1 - 2(r-1)w_j + \frac{2(r-1)(2r-1)}{2(1 + \hat{w}_j)^{2r}} w_j^2$$

$$\le 1 - 2(r-1)w_j + \frac{2(r-1)(2r-1)}{2(1 - \phi(-w))^{2r}} w_j^2$$

22

and we have our result. ∎

We are now ready to prove the main theorem related to the decrease in the merit function:

Theorem 27 *Let* $\tilde{\alpha} = \alpha \frac{\|u^k\|^2}{\phi(V_k^{r-1}u^k)}$ *and* $\hat{w}_N^k = (2r-1)V_{k,N}^{-1}\Delta v_N^k$. *Then, there is an* $L \geq 1$ *and an* $\beta > 0$ *such that for every* $k \geq L$

$$F_N(x^{k+1}) - F_N(x^k) \leq -\frac{2(r-1)\tilde{\alpha}}{1-\tilde{\alpha}}(1 - \frac{(2r-1)\tilde{\alpha}}{2(1-\alpha)^{2r}})(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)}\hat{w}_N^k + \mu_k).$$

where $|\mu_k| \leq \beta(-\phi(-v_N^k))^{-(r-1)}(c^Tx^k - c^*)^{2r}$.

Proof: From Step 4 and Theorem 20, for each $j \in N$,

$$
\begin{aligned}
\frac{x_j^{k+1}}{x_j^k} &= 1 - \alpha \frac{\|u^k\|^2}{\phi(V_k^{r-1}u^k)} \frac{(v_j^k)^{r-1}u_j^k}{\|u^k\|^2} \\
&= 1 - \tilde{\alpha}\frac{(v_j^k)^{r-1}u_j^k}{\|u^k\|^2} \\
&= 1 - \tilde{\alpha}(1 - \hat{w}_j^k + \hat{\Delta}_j^k),
\end{aligned}
$$

where $\hat{\Delta}^k = V_{k,N}^{r-1}\Delta^k$. From the above, definitions and Proposition 24 Parts 1 and 3 we obtain

$$
\begin{aligned}
\frac{v_j^{k+1}}{v_j^k} &= \frac{x_j^{k+1}}{x_j^k}\frac{c^Tx^k - c^*}{c^Tx^{k+1} - c^*} \\
&= 1 + \frac{\tilde{\alpha}}{1-\tilde{\alpha}}(\hat{w}_j^k - \hat{\Delta}_j^k) \\
&> 0.
\end{aligned}
$$

Let $\theta_j^k = \frac{v_j^{k+1}}{v_j^k} - 1 = \frac{\tilde{\alpha}}{1-\tilde{\alpha}}(\hat{w}_j^k - \hat{\Delta}_j^k)$, and note that $\theta_j^k > -1$. First, consider the case when $1 > \phi(-\theta^k) > 0$. Then, from Lemma 26, Part 2 we obtain

$$
\begin{aligned}
F_N(x^{k+1}) - F_N(x^k) &= \sum_{j \in N}((v_j^{k+1})^{-2(r-1)} - (v_j^k)^{-2(r-1)}) \\
&\leq -\frac{2\tilde{\alpha}(r-1)}{1-\tilde{\alpha}}(e^T V_{k,N}^{-2(r-1)}(\hat{w}_N^k - \hat{\Delta}^k) \\
&\quad - \frac{(2r-1)\tilde{\alpha}}{2(1-\tilde{\alpha})(1-\phi(-\theta^k))^{2r}}(\hat{w}_N^k - \hat{\Delta}^k)^T V_{k,N}^{-2(r-1)}(\hat{w}_N^k - \hat{\Delta}^k)).
\end{aligned}
$$

Using Theorem 20 and Proposition 24 Part 3, we obtain

$$1 - \phi(-\theta^k) = 1 - \frac{\tilde{\alpha}}{1-\tilde{\alpha}}\phi(-\hat{w}_N^k + \hat{\Delta}^k)$$

23

$$= \frac{1}{1 - \tilde{\alpha}}(1 - \tilde{\alpha}\phi(e - \hat{w}_N^k + \hat{\Delta}^k))$$

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

$$= \frac{1 - \alpha}{1 - \tilde{\alpha}}.$$

Thus, from Theorem 16, as $1 - \tilde{\alpha} < 1$, we obtain

$$F_N(x^{k+1}) - F_N(x^k) \leq -\frac{2\tilde{\alpha}(r-1)}{1 - \tilde{\alpha}}(1 - \frac{(2r-1)\tilde{\alpha}(1-\tilde{\alpha})^{2r-1}}{2(1-\alpha)^{2r}}(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)}\hat{w}_N^k + \mu_k)$$

$$\leq -\frac{2\tilde{\alpha}(r-1)}{1 - \tilde{\alpha}}(1 - \frac{(2r-1)\tilde{\alpha}}{2(1-\alpha)^{2r}}(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)}\hat{w}_N^k + \mu_k) \qquad (47)$$

where

$$\mu_k = e^T V_{k,N}^{-2(r-1)}\hat{\Delta}^k - \frac{(2r-1)\tilde{\alpha}(1-\tilde{\alpha})^{2r-1}}{2(1-\alpha)^{2r}}(2(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)}\hat{\Delta}^k - (\hat{\Delta}^k)^T V_{k,N}^{-2(r-1)}\hat{\Delta}^k)$$

$$= e^T V_{k,N}^{-(r-1)}\Delta^k - \frac{(2r-1)\tilde{\alpha}(1-\tilde{\alpha})^{2r-1}}{2(1-\alpha)^{2r}}(2(w_N^k)^T \Delta^k - (\Delta^k)^T \Delta^k).$$

Now consider the case when $\phi(-\theta^k) \leq 0$. Then Lemma 26 Part 1 must be used to calculate this change. By an identical analysis as above, in this case we obtain

$$F_N(x^{k+1}) - F_N(x^k) \leq -\frac{2\tilde{\alpha}(r-1)}{1 - \tilde{\alpha}}(1 - \frac{(2r-1)\tilde{\alpha}}{2(1-\tilde{\alpha})}(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)}\hat{w}_N^k + \mu_k')$$

where

$$\mu_k' = e^T V_{k,N}^{-(r-1)}\Delta^k - \frac{(2r-1)\tilde{\alpha}}{2(1-\tilde{\alpha})}(2(w_N^k)^T \Delta^k - (\Delta^k)^T \Delta^k).$$

As $\alpha < 1$, $\tilde{\alpha} < \alpha$ it is readily seen that the bound obtained by Part 2 of Lemma 26 is larger. As $-\phi(-v_N^k) = v_{j_k}^k = \min_{j \in N} v_j^k$ we see using Theorem 16, Part 2

$$|e^T V_{k,N}^{-(r-1)}\Delta^k| \leq q(v_{j_k}^k)^{-(r-1)}\|\Delta^k\|$$

$$|(w_N^k)^T \Delta^k| \leq \sqrt{q}(v_{j_k}^k)^{-(r-1)}\|\Delta^k\|$$

and our result follows from Theorem 20 and Lemma 23. ∎

We are now ready to prove the main convergence theorem.

Theorem 28 Let $\frac{\alpha}{(1-\alpha)^{2r}} < \frac{2}{2r-1}$, and assumptions (1) - (3) hold. Then,

1. $x^k \longrightarrow x^*$

2. $y^k \longrightarrow y^*$

3. $s^k \longrightarrow s^*$

where x^* is an optimum solution of the primal, and (y^*, s^*) is an optimum solution of the dual. In addition, (y^*, s^*) is the power center of the optimal dual face, and thus strict complementarity holds between this pair of optimum solutions.

Proof: We see from Proposition 24 Part 4 we see that there is a $\gamma > 0$ such that $\tilde{\alpha}_k \geq \gamma\alpha$. Thus

$$\frac{2(r-1)\tilde{\alpha}_k}{1-\tilde{\alpha}_k} > \frac{2(r-1)\alpha\gamma}{1-\alpha\gamma} = \theta_1 > 0$$

and from Proposition 24 upper bound in Part 2, $\alpha \geq \tilde{\alpha}_k$. Thus

$$1 - \frac{(2r-1)\tilde{\alpha}_k}{2(1-\alpha)^{2r}} \geq 1 - \frac{(2r-1)\alpha}{2(1-\alpha)^{2r}} = \theta_2 > 0.$$

Assume that on some subsequence K,

$$v_{j_k}^k = \min_{j \in N} v_j^k \longrightarrow 0$$

and let this be the largest and thus the only such subsequence. Then, from Theorem 20, $\hat{w}_{j_k}^k \longrightarrow 1$; and so there is an $\hat{L} \geq 1$ such that for all $k \geq \hat{L}$, $\hat{w}_{j_k}^k \geq 0.50$. Thus

$$(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)} \hat{w}_N^k \geq 0.25(v_{j_k}^k)^{-2(r-1)}.$$

Thus, from Theorem 27

$$F_N(x^{k+1}) - F_N(x^k) \leq -\theta_1(\theta_2 + \epsilon_k)(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)} \hat{w}_N^k$$

where

$$|\epsilon_k| = \frac{|\mu_k|}{(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)} \hat{w}_N^k} \leq 4\beta(v_{j_k}^k)^{r-1}(c^T x^k - c^*)^{2r}$$

and from Proposition 8, $\epsilon_k \longrightarrow 0$. Thus there is an $L \geq \hat{L}$ such that for all $k \geq L$

$$F_N(x^{k+1}) - F_N(x^k) \leq -\frac{\theta_1\theta_2}{2}(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)} \hat{w}_N^k$$

Now, from Corollary 4, there is a $\rho > 0$ such that

$$\begin{aligned}
(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)} \hat{w}_N^k &\geq \frac{\|\hat{w}_N^k\|^2}{\phi(v_N^k)^{2(r-1)}} \\
&\geq \|\hat{w}_N^k\|^2 \left(\frac{c^T x^k - c^*}{\|x_N^k\|}\right)^{2(r-1)} \\
&\geq \rho^{2(r-1)}\|\hat{w}_N^k\|^2.
\end{aligned}$$

Thus for each $k \geq L$ in K

$$F_N(x^{k+1}) - F_N(x^k) \leq -\frac{\theta_1 \theta_2}{2} \rho^{2(r-1)} \|\hat{w}_N^k\|^2.$$

From the definition of K, there is a $\delta > 0$ such that for each $k \notin K$

$$v_{j_k}^k > \frac{1}{\delta}.$$

Thus, from Theorem 27 there is a $\beta' > 0$ such that $|\mu_k| \leq \beta(c^T x^k - c^*)^{2r}$ where $\beta = \beta' \delta^{r-1}$. and

$$F_N(x^{k+1}) - F_N(x^k) \leq -\theta_1 (\rho^{2(r-1)} \theta_2 \|\hat{w}_N^k\|^2 + \mu_k).$$

Thus for each $k \geq L$ there is a $\sigma_k > 0$ such that

$$\sum_{k=L}^{\infty} (F_N(x^{k+1}) - F_N(x^k)) \leq -\theta_1 \theta_2 \rho^{2(r-1)} (\sum_{k \in K} \frac{1}{2} \|\hat{w}_N^k\|^2 + \sum_{k \notin K} \|\hat{w}_N^k\|^2) - \theta_1 \sum_{k \notin K} \mu_k$$

$$\leq -\sum_{k=L}^{\infty} \sigma_k \|\hat{w}_N^k\|^2 - \theta_1 \sum_{k \notin K} \mu_k.$$

From Lemma 25,

$$|\sum_{k \notin K} \mu_k| \leq \sum_{k \notin K} |\mu_k| \leq \sum_{k \notin K} \beta(c^T x^k - c^*)^{2r} < \infty.$$

From Corollary 4 $F_N(x^k)$ is bounded below by zero on this sequence. Thus $\|\hat{w}_N^k\| \longrightarrow 0$. As $\hat{w}_N^k = (2r-1) V_{k,N}^{-1} \Delta v_N^k$, from Proposition 8,

$$\Delta v_N^k \longrightarrow 0$$

and, from Theorem 20

$$\frac{V_{k,N}^{r-1} u_N^k}{\|u^k\|^2} \longrightarrow e.$$

Since the only vector in $\mathcal{P}_N$ for which the Newton step $\Delta v_N = 0$ is its power center v_N^*, we must have,

$$v_N^k \longrightarrow v_N^*.$$

Let $\bar{K}$ be such that for $k \in \bar{K}$ (it exists since all these sequences are bounded)

$$s^k \longrightarrow s^*$$

$$y^k \longrightarrow y^*$$

$$u^k \longrightarrow u^*,$$

26

and, so

$$\frac{V_{k,N}^{r-1} u_N^k}{\|u^k\|^2} \longrightarrow \frac{V_{*,N}^{r-1} u_N^*}{\|u^*\|^2} = e. \tag{48}$$

Using Theorem 15, it is readily seen that (y^*, s^*) is the power center of the optimal dual face; and our result follows from the complementary slackness theorem. ∎

We get the following sharp bound on the linear convergence rate of $\{c^T x^k\}$.

Theorem 29 *Let* $\frac{\alpha}{(1-\alpha)^{2r}} < \frac{2}{2r-1}$. *Then*

$$\lim_{k \to \infty} \frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} = 1 - \alpha.$$

Proof: Follows readily from Proposition 24 Part 2 and Equation (48). ∎

12 Convergence of Dual with Variable Step Size

In the previous section, the step size implemented at each iteration was a constant α. In this section we will allow the stepsize to vary, and in iteration k, we will choose the stepsize α_k by rules described below. This increases the step size implemented, which will be shown, asymptotically to be given by the formula

$$\frac{\alpha_k}{1 - \alpha_k} < \frac{2}{2r-1}.$$

We obtain this increase by using the following estimate of $\delta(u^k)$ (of Proposition 24)

$$\tau_k = \frac{\|X_k^r s^k\|^2 (x^k)^T s^k}{\phi(X_k^{2r-1} s^k)}. \tag{49}$$

The lemma that follows is useful in understanding the above estimate.

Lemma 30 *There exists an* $L \geq 1$ *and a* $\beta > 0$ *such that for all* $k \geq L$

$$\frac{(x^k)^T s^k}{c^T x^k - c^*} = 1 + \delta_k$$

where $|\delta_k| \leq \beta (c^T x^k - c^*)^{2r-1}$.

Proof: Since $(x^k)^T s^k = (x_B^k)^T s_B^k + (x_N^k)^T s_N^k$, from Proposition 14 Part 1, and the fact that $x_B^k \longrightarrow x_B^* > 0$, there is an $L \geq 1$ such that the first term is bounded by $\beta_1 (c^T x^k - c^*)^{2r}$ for some $\beta_1 > 0$. Our result now follows from Theorem 19. ∎

We now establish the goodness of the estimate (49).

Proposition 31 *Let τ_k be defined by Equation (49). There is an $L \geq 1$ and a $\beta > 0$ such that for all $k \geq L$*

$$\delta(u_N^k) = \tau_k + \delta_k$$

where $|\delta_k| \leq \beta(c^T x^k - c^*)^{2r-1}$.

Proof: This result follows readily from Lemma 30, Proposition 14, Part 2 and the upper bound of Part 2 of Proposition 24. ∎

Since τ_k is, asymptotically, a very good estimate of $\delta(u_N^k)$, in place of using $\frac{\alpha}{(1-\alpha)^{2r}}$ as an estimate of $\frac{\tilde{\alpha}(1-\tilde{\alpha})^{2r-1}}{(1-\alpha)^{2r}}$ in the relations (47), we will use the estimate $\delta_k \alpha$ of $\tilde{\alpha}$ in the above formula. Also, from Proposition 24, Part 2 whenever $\|w_N^k\| \longrightarrow 0$, $\delta(u_N^k) \longrightarrow 1$. Thus, in this case $\tau_k \longrightarrow 1$, and the new estimate approaches $\frac{\alpha}{1-\alpha}$. We now present this step selection strategy.

Define α^* such that

$$\frac{\alpha^*}{(1 - \alpha^*)^{2r}} < \frac{2}{2r - 1}.$$

Let $\theta > 0$ be very small and at iteration k choose the step size α_k by the following strategy:

Step 1 Let α' be such that

$$\frac{\tau_k \alpha'(1 - \tau_k \alpha')^{2r-1}}{(1 - \alpha')^{2r}} = \frac{2}{(2r - 1)} - \theta. \tag{50}$$

Step 2 Define

$$\alpha_k = \begin{cases} \alpha' & \text{if} \quad \alpha' \geq \alpha^* \\ \alpha^* & \text{otherwise} \end{cases} \tag{51}$$

The above choice guarantees that the estimate is not smaller than one obtained from Theorem 28. The next lemma establishes a relationship between the α' computed in (50) and the related expression in system (47).

Lemma 32 *Let α' be computed by (50) and let $\tilde{\alpha} = \delta(u_N^k)\alpha'$. Then, there is an $L \geq 1$ and a $\beta > 0$ such that for all $k \geq L$*

$$\tilde{\alpha}(1 - \tilde{\alpha})^{2r-1} = \tau_k \alpha'(1 - \tau_k \alpha')^{2r-1} + \delta_k$$

where $|\delta_k| \leq \beta(c^T x^k - c^*)^{2r-1}$.

Proof: Follows readily from Proposition 31, and the fact that for small $\epsilon > 0$, $(1 + \epsilon)^{2r-1} < 1 + 4^r\epsilon$. ∎

We are now ready to prove the main theorem of this section.

Theorem 33 *For each k, let α_k, be generated by the above rules and let the assumptions (1) - (3) hold. Then,*

1. $x^k \longrightarrow x^*$

2. $y^k \longrightarrow y^*$

3. $s^k \longrightarrow s^*$

where x^* is an optimum solution of the primal, and (y^*, s^*) is an optimum solution of the dual. In addition, (y^*, s^*) is the power center of the optimal dual face, and thus strict complementarity holds between this pair of optimum solutions.

Proof: Assume that $\alpha_k = \alpha^*$ for each k. Then this theorem follows from Theorem 28. Otherwise for each k for which $\alpha_k = \alpha' > \alpha^*$, using the same argument of Theorem 28 , it is readily shown that for some $\beta > 0$

$$F_N(x^{k+1}) - F_N(x^k) \leq -\beta \|w_N^k\|^2 + \mu_k$$

where $\mu_k \leq \gamma (c^T x^k - c^*)^{2r}$. The proof is completed in the same was as in the Theorem 28. ∎

We now obtain the asymptotic behavior of α_k.

Corollary 34 $\alpha_k \longrightarrow \alpha$ *where*

$$\frac{\alpha}{1 - \alpha} = \frac{2}{2r - 1} - \theta.$$

Proof: As a consequence of Theorem 33, we obtain the fact that $\|w_N^k\| \longrightarrow 0$. From Proposition 24, $\delta(u_N^k) \longrightarrow 1$, from Proposition 31, $\tau_k \longrightarrow 1$, and we have our result. ∎

13 Accelerated Variant

We will now use the connection between the Newton and the Affine Scaling step to accelerate the convergence of the algorithm. We now present the accelerated version of the method and then investigate its convergence rate.

Step 0 Let x^0 be an interior point solution, $\frac{\alpha}{(1-\alpha)^{2r}} < \frac{2}{2r-1}$, $\delta > 0$, $r > 1$; and, let $k = 0$.

Step 1 Tentative Solution to the Dual:

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

Step 2 Tentative Dual Slack:

$$s^k = c - A^Ty^k$$

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

Step 3 Min-Ratio test:

$$
\begin{aligned}
\theta_k &= \min\{\frac{\|X_k^{2r-1}s^k\|}{(x_j^k)^{2r-1}s_j^k} : s_j^k > 0\} \\
&= \frac{\|X_k^{2r-1}s^k\|}{\phi(X_k^{2r-1}s^k)}.
\end{aligned}
$$

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

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

$$
\begin{aligned}
\alpha_k' &\quad \text{defined by (51)} \\
N_k &= \{j : x_j^k \leq e^TX_ks^k\} \\
\gamma_k &= e^TX_{k,N_k}s_{N_k}^k \\
h_{N_k}^k &= \frac{1}{2r-1}(\frac{X_{k,N_k}}{\gamma_k} - \frac{X_{k,N_k}^{2r-1}s_{N_k}^k}{\|X_k^Ts^k\|^2}) \\
\epsilon_k &= \|h_{N_k}^k\| \\
\rho_k &= \frac{\log(\epsilon_k)}{\log(\gamma_k)}.
\end{aligned}
$$

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

$$\alpha_k = 1 - \max\{\epsilon_k^\delta, \gamma_k^{\delta\rho_k}\}.$$

30

2. Corrector Step: Otherwise,

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

Step 5 Next Interior Point:

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

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

Some comments are in order here. This acceleration scheme is identical to the three step acceleration scheme of Saigal [17]. h_N^k computed in step 4 is a very good estimate of the Newton step Δv_N^k, and its magnitude ϵ_k is used to estimate the distance to the power center v_N^* of $\mathcal{P}_N \cap \{v_N : v_N \geq 0\}$. Asymptotically, we apply a predictor step when the size, ϵ_k of the Newton step is of the order $O(c^T x^k - c^*)^{2r}$; and the corrector step otherwise. 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

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

and thus the affine scaling step behaves like a Newton step. We now establish some preliminary results and prove the convergence of this accelerated method.

Proposition 35 *There is an $L \geq 1$ and a subset N of $\{1, \cdots, n\}$ such that for all $k \geq L$, $N_k = N$.*

Proof: From Theorem 6, $\gamma_k \longrightarrow 0$; and by definition $x_j^* > 0$ for each $j \in B$. Thus for some $L \geq 1$ and all $k \geq L$

$$x_j^k > \gamma_k \text{ for all } j \in B$$

and we are done. ∎

Let $\hat{L}$ be generated by Proposition 35. We now prove a series of results using the property of $\hat{L}$.

Proposition 36 *There is an $L \geq \hat{L}$ and an $\alpha > 0$ such that for all $k \geq L$*

$$\|h_N^k - \Delta v_N^k\| \leq \alpha(c^T x^k - c^*)^{2r}.$$

31

Proof: We note that by substitution of the results of Theorems 19 and 22 and definitions, we get

$$
\begin{aligned}
h_N^k &= \frac{1}{2r-1}\left(\frac{v_N^k}{e^T \tilde{u}_N^k} - \frac{V_{k,N}^r u_N^k}{\|u^k\|^2}\right) \\
&= \frac{1}{2r-1}\left(v_N^k - \frac{V_{k,N}^r u_N^k}{\|u^k\|^2} + V_{k,N}^r \Delta^k\right) \\
&\quad - \frac{1}{2r-1}\left(V_{k,N}^r \Delta^k + \frac{\delta_k}{1+\delta_k} v_N^k\right) \\
&= \Delta v_N^k - E^k.
\end{aligned}
$$

As v_N^k is bounded by Theorem 8, our result follows from Theorems 19 and 22. ∎

Proposition 37 *Let $w_N^k \to 0$ on some subsequence K. Then, on K*

$$
\rho_k = \frac{\gamma_k \phi(X_{k,N}^{2r-1} s_N^k)}{2r\|X_k^r s^k\|^2} \longrightarrow \frac{1}{2r}
$$

Proof: From Theorem 19 and Proposition 24 Part 2, $2r\rho_k \geq e^T \tilde{u}_N^k = 1 + \delta_k$. From equation (46)

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

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

Proposition 38 *There is an $L \geq \hat{L}$ such that for all $k \geq L$*

$$
0.50(c^T x^k - c^*) \leq \gamma_k \leq 1.5(c^T x^k - c^*)
$$

Proof: Note that $e^T X_{k,N_k} s_{N_k}^k = (c^T x^k - c^*)e^T \tilde{u}_N^k$, and so the result follows from Theorem 19. ∎

We are now ready to prove the convergence to optimality of the limit point of the primal sequence generated by the accelerated algorithm.

Theorem 39 *Let assumptions (1) - (3) hold. There exists a subsequence K such that for $k \in K$*

1. $x^k \longrightarrow x^*$

2. $y^k \longrightarrow y^*$

3. $s^k \longrightarrow s^*$,

where x^* and (y^*, s^*) are optimal solutions of the primal and the dual problems respectively. Also, (y^*, s^*) is the power center of the optimal dual face.

Proof: Let L be as in Proposition 35. In case the predictor step is taken only a finite number of times, then α_k is selected by the variable step size selection strategy, and our result follows from Theorem 33. Thus, let there exist a subsequence K such that for each $k \in K$, a predictor step is taken. Then, for $k \in K$

$$\|h_N^k\| \leq \gamma_k^{1.5r}.$$

But, from Theorem 6, $\gamma_k \longrightarrow 0$. Thus, as

$$\|\Delta v_N^k\| \leq \|h_N^k\| + \|\Delta v_N^k - h_N^k\|$$

$\Delta v_N^k \longrightarrow 0$; and the result follows by an argument identical to the proof of Theorem 28. ∎

14 Convergence Rate of Accelerated Variant

We will now investigate the convergence rate of the variant. This rate is a function of the choice of r, and approaches 2 step quadratic as r approaches infinity. From Theorem 39, we can conclude that the power center of $\mathcal{P}$ exists. We investigate now the application of Newton's method to compute this center.

Proposition 40 *There is a $\theta > 0$ such that for all $\|v_N - v_N^*\| < \theta$,*

$$0.50\|v_N - v_N^*\| \leq \|\Delta v_N\| \leq 1.50\|v_N - v_N^*\|$$

Proof Follows from Theorem 17, Part 1. ∎

The following is a consequence of the above proposition:

Proposition 41 *There is an $\alpha > 0$ such that for every $0 < \delta < 1$, there is an $L_\delta \geq 1$, and; if $k \geq L_\delta$, $\beta > 0$, $p \geq 1.5r$ and $\|v_N^k - v_N^*\| \leq \beta(c^T x^k - c^*)^p$ then*

$$\|w_N^k\| \leq \alpha(c^T x^k - c^*)^p.$$

Proof Since $v_N^* > 0$, for every $0 < \delta < 1$, there is an $L_\delta \geq 1$ such that for all $k \geq L_\delta$,

$$a = \frac{\min_j v_j^*}{2} \geq (c^T x^k - c^*)^{1.5r}.$$

Thus for some $\gamma > 0$,

$$v_j^k \geq v_j^* - |v_j^* - v_j^k| \geq a.$$

and our result now follows by observing that $\|w_N^k\| \leq (\min_j v_j^k)^{-1}\|\Delta v_N^k\|$, and the result of Proposition 40. ■

Lemma 42 *Assume that for some $k \geq L_\delta$ and $1.5r \leq p \leq 2r$, $\|v_N^k - v_N^*\| \leq \beta(c^T x^k - c^*)^p$. Then, there is an $\alpha > 0$ such that*

$$\|h_N^k\| \leq \alpha(c^T x^k - c^*)^p.$$

Proof: From Proposition 40 $\|\Delta v_N^k\| \leq 1.5\beta(c^T x^k - c^*)^p$. Also, by Theorem 36

$$\|h_N^k\| \leq \|\Delta v_N^k\| + \|h_N^k - \Delta v_N^k\| \leq 1.5\beta(c^T x^k - c^*)^p + \gamma(c^T x^k - c^*)^{2r}$$

and we are done. ■

We are now ready to investigate the predictor step.

Proposition 43 *Assume that $0 < \delta < 1$, $k \geq L_\delta$, $\beta > 0$, $1.5r \leq p$ and $\|v_N^k - v_N^*\| \leq \beta(c^T x^k - c^*)^p$. Then*

1. *There are $0 < \theta_1 < \theta_2$ such that*

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

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

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

Proof: From Lemma 42, we obtain a $\beta > 0$ such that

$$\epsilon_k \leq \beta(c^T x^{k+1} - c^*)^p.$$

We note the following sequence of inequalities which follow from the Propositions 38 and 24:

$$(0.50)^{\delta\rho_k}(c^T x^k - c^*)^{\delta\rho_k} \;\leq\; \gamma_k^{\delta\rho_k}$$

$$\leq\; 1 - \alpha_k$$

$$\leq\; 1 - \alpha_k\frac{\|u^k\|^2}{\phi(V_k^{r-1}u^k)}$$

$$=\; \frac{(c^T x^{k+1} - c^*)}{(c^T x^k - c^*)}$$

$$\leq\; 1 - (1 - \max\{\epsilon_k^\delta, \gamma_k^{\delta\rho_k}\})(1 - \|w_N^k\| - |\delta_k|)$$

$$\leq\; \max\{\epsilon_k^\delta, \gamma_k^{\delta\rho_k}\} + \|w_N^k\| + |\delta_k|$$

$$\leq\; (1.5)^{\delta\rho_k}(c^T x^k - c^*)^{\delta\rho_k},$$

and we have part 1. Now, from Theorems 21 and 22, and some simple algebra, we can write

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

where

$$E^k = \frac{2r\alpha_k\delta(u_N^k) - 1}{1 - \alpha_k\delta(u_N^k)}\frac{1}{2r-1}(h_N^k + \frac{1}{2r-1}\frac{\delta_k}{1+\delta_k}v_N^k) - \frac{1}{2r-1}V_{k,N}^r\Delta^k.$$

Using the facts that $\delta(u_N^k) \leq 1$, $\alpha_k \leq 1 - \epsilon_k^\delta$, we obtain, for some $\theta_3 > 0$

$$\|E^k\| \leq \theta_3(c^T x^k - c^*)^{(1-\delta)\rho_k}$$

and the Part 2 follows. ∎

We are now ready to investigate the corrector step of the algorithm.

Proposition 44 *Assume that for some $k \geq L_\delta$ and $\beta > 0$, $\|v_N^k - v_N^*\| \leq \beta(c^T x^k - c^*)^p$. Then*

1. *$\frac{1.5r}{2} \leq p \leq r$. One iteration of the corrector step will be taken, and for some $\theta_1 > 0$*

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

2. *$\frac{1.5r}{4} \leq p \leq \frac{r}{2}$. Then at least one corrector step will be taken, and after at most two steps, for some $\theta_2 > 0$*

$$\|v_N^{k+2} - v_N^*\| \leq \theta_1(c^T x^{k+2} - c^*)^{4p}.$$

35

Proof: From Theorems 19 and 20 and Proposition 41, we can write

$$\frac{1}{2r} - |\delta_k| \leq \frac{\gamma_k \phi(X_{k,N_k}^{2r-1} s_{N_k}^k)}{2r \|X_k^r s^k\|^2} \leq \frac{1}{2r} + |\rho_k|$$

where $|\delta_k| \leq \theta(c^T x^k - c^*)^{2r}$ and $|\rho_k| \leq \theta'(c^T x^k - c^*)^p$. Thus, as α'_k approaches $\frac{2}{2r+1} - \epsilon$ for some small $\epsilon > 0$, for all sufficiently large k,

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

Thus, from Proposition 14 Part 1, Theorems 21, 22 and simple algebra we see that

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

where

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

Using Propositions 8, 9 and Lemma 23 we can show that

$$\|E^k\| \leq \beta(c^T x^k - c^*)^{2r}.$$

Thus, after one step with α_k, we see that (using Theorem 17),

$$\begin{aligned} \|v_N^{k+1} - v_N^*\| &\leq \|v_N^k + \Delta v_N^k - v_N^*\| + \|E^k\| \\ &\leq \rho \|v_N^k - v_N^*\|^2 + \beta(c^T x^k - c^*)^{2r} \\ &\leq \alpha(c^T x^k - c^*)^{2p}. \end{aligned}$$

From, Proposition 24

$$\begin{aligned} \frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} &= 1 - \alpha_k \delta(u_N^k) \\ &= 1 - \frac{e^T \tilde{u}_N^k}{2r}. \end{aligned}$$

and using Theorem 19 we see that for all sufficiently large k

$$1 - \frac{1.5}{2r} \leq \frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} \leq 1 - \frac{0.5}{2r}.$$

Substituting the above inequality, we obtain part 1 of the theorem for $\theta_1 = (1 - \frac{1.5}{2r})^{2p} \alpha$.

To see part 2, we note that after one corrector step, either $2p$ becomes greater than $1.5r$ and we stop the corrector iterates and go to the predictor step; or, after one more corrector step the desired result is obtained. ∎

We are now ready to prove the main convergence theorem.

Theorem 45 *Let the sequences $\{x^k\}$, $\{y^k\}$ and $\{s^k\}$ be generated by the accelerated variant with $r > 1$, and let assumptions (1) - (3). Then*

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 power 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. For $\delta = \frac{1}{2(r+1)}$, the convergence is two-step superlinear at the rate $1 + \frac{r}{r+1}$.

2. For $\delta = \frac{3}{2(r+2)}$, the convergence is three-step superlinear at the rate $1 + \frac{3r}{r+2}$.

Proof: Assume that only a finite number of predictor steps are taken. Using the argument of Theorem 28 it can be shown that there is a subsequence K along which

$$\Delta v_N^k \longrightarrow 0.$$

Using Theorem 17 Part 2, we can show that there is an $L \geq 1$ and a $\theta > 0$ such that for all $k \geq L$ and $k \in K$

$$\|v_N^k + \Delta v_N^k - v_N^*\| \leq \theta \|v_N^k - v_N^*\|^2.$$

For each $k \in K$, sufficiently large, using an argument similar to that of Proposition 44, we can show that for some $\theta_1 > 0$

$$\|v_N^k - v_N^*\| \leq \theta_1 \|v_N^k - v_N^*\|^2.$$

and from Proposition 40, $\|\Delta v_N^{k+1}\| \leq \theta_2 \|\Delta v_N^k\|^2$, and a corrector step can also be performed at the step $k + 1$, and we will obtain $\|\Delta v_N^{k+2}\| \leq \theta_2^2 \|\Delta v_N^k\|^4$. Repeating this argument we contradict the fact that only a finite number of predictor steps are taken.

Let k be an index, sufficiently large at which a predictor step is performed. To investigate the convergence rate of the two-step method, assume $\rho_k \geq 1.5r$, and let $\delta = \frac{\rho_k - r}{\rho_k(r+1)}$. After one corrector step, from Propositions 43 and 44, we obtain

$$\|v_N^{k+2} - v_N^*\| \leq \theta_3 (c^T x^{k+2} - c^*)^{\frac{2(1-\delta)\rho_k}{1+\delta\rho_k}}.$$

where $\frac{2(1-\delta)\rho_k}{1+\delta\rho_k} = 2r$, we obtain part 1 as

$$\delta = \frac{1}{2(r+1)} \text{ when } \rho_k = 2r$$

and the convergence rate

$$1 + \delta\rho_k = 1 + \frac{r}{r+1}.$$

For the three step method, let $\delta = \frac{2\rho_k - r}{\rho_k(r+2)}$. After two corrector steps

$$\|v_N^{k+3} - v_N^*\| \leq \theta_4 (c^T x^{k+3} - c^*)^{\frac{4(1-\delta)\rho_k}{1+\delta\rho_k}}$$

where $\frac{4(1-\delta)\rho_k}{1+\delta\rho_k} = 2r$. Part 2 now follows since $\delta = \frac{3}{2(r+2)}$ if $\rho_k = 2r$ and the convergence rate $1 + \delta\rho_k = 1 + \frac{3r}{r+2}$. $\blacksquare$

15 Efficiency of Acceleration

Ostrowski [15], section 6.11 introduced the following measure of efficiency for algorithms achieving different rates of convergence, involving different amount of work per iteration. He suggested the following measure

$$\frac{\log(\rho)}{w}$$

where ρ is the convergence rate of the algorithm, and w is a measure of the work per iteration. The larger this measure, the more efficient the algorithm. This measure has been used by Brent [4] who investigated the hybrids of Newton's method proposed by Shamanskii; and, Saigal and Todd [18] who investigated the hybrids of fixed point computing methods with variants of Newton's method.

The convergence of the accelerated variant depends on the choice of r and the two-step or three-step method. The table below shows these calculations for several choices.

	Two - Step		Three - Step	
	rate	efficiency	rate	efficiency
$r = 1.0$	1.5^1	$\frac{0.2027}{w}$	2.0^2	$\frac{0.2310}{w}$
$r = 1.5$	1.6	$\frac{0.2350}{w}$	2.2857	$\frac{0.2756}{w}$
$r = 2.0$	1.67	$\frac{0.2564}{w}$	2.50	$\frac{0.3054}{w}$
$r = 4.0$	1.80	$\frac{0.2939}{w}$	3.0^3	$\frac{0.3662}{w}$

1. Tsuchiya and Monterio [23] obtain a rate of 1.3.
   This can be established by the method of Saigal [17].

2. This is obtained in Saigal [17].

3. The efficiency of three step cubic is larger than the two-step quadratic, which is $\frac{0.3466}{w}$.

16 Concluding Remarks

In this paper we have shown that for every $r > 0.50$, there is a variant of the affine scaling method. The usual method is generated when $r = 1$. We have analysed the convergence of these variants for $r > 1$. The analysis for $0.50 < r < 1$ is analogous, with a few changes in the formulae to account for the sign changes; and, the objective functions of the power center problems.

Under the assumption of non-degeneracy, convergence to optimality of the primal sequence is shown for any step size less than 1. To investigate the convergence without the non-degeneracy assumption, the concept of a power center is introduced. The power centers of the primal and the dual optimal faces are related in an intimate way, and the objective functions defining these centers are related in the same sense as the "dual norms" are. In this case, it is shown that if the step size α is chosen such that $\frac{\alpha}{(1-\alpha)^{2r}} < \frac{2}{2r-1}$, for $r > 1$, the primal sequence converges to the relative interior of the optimal primal face and the dual sequence converges to the power center of the optimal dual face. Also, a variable step selection strategy is presented where the sequence $\{\alpha_k\}$ of step sizes, asymptotically is selected by $\frac{\alpha_k}{1-\alpha_k} < \frac{2}{2r-1}$. This sequence is required to stay uniformly away from from $\frac{2}{2r-1}$. Thus, $\alpha_k < \frac{2}{2r+1}$, and thus this result is a generalization of the 2/3rd result of Tsuchiya and

Muramatsu [24] for $r = 1$.

An accelerated variant is also presented. This variant achieves superlinear convergence, and the rate is higher for larger values of $r > 1$. This generalize the work of Saigal [17] and Tsuchiya and Monterio [23].

This work opens up the study of hybrid variants of the affine scaling method in which different values of r are implemented at different stages of the method. From Theorems 12 and 14, it is evident that the rate of convergence of $\|x_N^k\|$ is $O(c^T x^k - c^*)$ while that of $\|s_B^k\|$ is $O(c^T x^k - c^*)^{2r}$. Implementing $0.50 < r < 1$ in the early iterates will reduce this disparity between the accuracy of the primal and the dual sequence, and thus make the method behave more like the primal-dual methods where the accuracy of the two sequences is similar. In the later iterations (when $\gamma_k < 1$), a value of $r > 1$ (say $r = 1.5$ or 2.0) can be implemented to get a higher rate of convergence. These hybrids have not been studied yet, and we expect to report computational experience on them at a later date.

References

[1] I. Adler, N. K. Karmarkar, M. G. C. Resende and G. Veiga, " An implementation of Karmarkar's algorithm for linear programming," *Mathematical Progamming* 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] J. R. Birge and C. Rosa, " A simplified proof of the convergence of affine scaling," Technical Report, Department of Industrial and Operations Engineering, University of Michigan, (Ann Arbor, Michigan 1991).

[4] R. P. Brent, " Some efficient algorithms for solving systems of non-linear equations," *SIAM Journal on Numerical Analysis* 10 (1973) 327-344.

[5] D. den Hertog, C. Roos and T. Terlaky, " Inverse barrier methods for linear programming," Report No 91-27, Delft University of Technology, Netherlands (1991).

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

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

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

[9] I. I. Dikin, " Determination of interior point of one system of linear inequalities" *Kibernetica and system analysis* ˙1 (1992).

[10] L. A. Hall and R. J. Vanderbei, " Two-thirds is sharp for affine scaling," Technical Report SOR-92/9, Department of Civil Engineering and Operations Research, Princeton University, Princeton, New Jersey, 08544, September 1992.

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

[12] N. K. Karmarkar and R. G. Ramakrishna, " Further developments in the new polynomial-time algorithm for linear programming " Talk given at ORSA/TIMS National Meeting, Boston, MA, April 1985.

[13] R. D. C. Monteiro, T. Tsuchiya, and Y. Wang, " A simplified global convergence proof of the affine scaling algorithm," Technical Report, Dept. of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721, USA, October 1992.

[14] 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.

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

[16] 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.

[17] R. Saigal, " A three step quadratically convergent implementation of the primal affine scaling method," Technical Report No 93-9, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, February 26, 1993.

[18] R. Saigal and M. J. Todd, " Efficient acceleration techniques for fixes point algorithms," *SIAM Journal on Numerical Analysis* 2 (1978) 997-1007.

[19] R. L. Sheu and S. C. Fang, " On the generalized path-following methods for linear programming," Technical report, 1992.

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

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

[22] T. Tsuchiya, " Global convergence property of the affine scaling method for primal degenerate linear programming problems'" *Mathematics of Operations Research* 17 (1992) 527- 557.

[23] T. Tsuchiya and R. D. C. Monterio, " Superlinear convergence of the Affine Scaling Algorithm," Research Memorandum, November 28, 1992.

[24] T. Tsuchiya and M. Muramatsu, " Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems'" Research Memorandum 423, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan, January 1992 (revised, September, 1992).

[25] R. J. Vanderbei, M. J. Meketon and B. A. Freedman, " A Modification of Karmarkar's linear programming algorithm," *Algorithmica* 1 (1986) 395 - 407.

[26] 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.