The primal power affine scaling method

Romesh Saigal*

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

E-mail: rsaigal@engin.umich.edu

In this paper, we present a variant of the primal affine scaling method, which we call the primal power affine scaling method. This method 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 Terlaky and Sheu and Fang. Here, we analyze the methods 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 method converges for any choice of the step size α. To analyze 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 method to generalize the 2/3rd result of Tsuchiya and Muramatsu. We show that with a constant step size α such that $\alpha/(1 - \alpha)^{2r} < 2/(2r - 1)$ and with a variable asymptotic step size α_k uniformly bounded away from $2/(2r + 1)$, 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 method. We show that the two-step superlinear convergence rate of the method is $1 + r/(r + 1)$, while the three-step convergence rate is $1 + 3r/(r + 2)$. 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.

Keywords: Linear programming, affine scaling methods, interior point methods, power barrier method, power center, merit function, superlinear convergence, three-step quadratic convergence, efficient acceleration.

1. Introduction

We consider here the linear programming problem:

$$\begin{aligned}
\text{minimize} \quad & c^T x \\
& Ax = b, \\
& x \geq 0,
\end{aligned} \tag{1}$$

with its dual

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

where A is an $m \times n$ 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 application of the primal affine scaling method for solving this problem. The primal method was proposed by Dikin [6] in 1967, who subsequently proved its convergence under the primal non-degeneracy assumption (Dikin [7]). His proof also appears in Vanderbei and Lagarias [27]. This method was rediscovered by Barnes [3], 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 et al. [1]) and the primal-dual (Monteiro et al. [14]) were generated in the process of implementing the projective transformation method of Karmarkar [12]. See also Adler and Monteiro [2] for an analysis of the limiting trajectories generated by these methods.

The convergence behavior of the affine scaling method without the non-degeneracy assumption is now known. For example, Mascarenhas [13] 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 [22], who introduced a local potential function to analyze the convergence of this method, significant developments have occurred. Dikin [8], 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 $\alpha \leq 1/2$. Tsuchiya and Muramatsu [25] subsequently proved the same convergence behavior when $\alpha \leq 2/3$. Simpler proofs of this result have been developed by Monteiro et al. [15] and Saigal [17]. It is also known that the dual sequence may not converge when $\alpha > 2/3$. Hall and Vanderbei [11] have produced an example where this happens. Saigal [17] and Gonzaga [10] 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 $\alpha \leq 2q/(3q-1)$, 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.

Using the connection between the steps of Newton's method for computing an analytic center of a certain polytope, and the affine scaling method interpreted on this polytope, Saigal [18] and Tsuchiya and Monteiro [24] have devised a variable step size selection strategy which produces super-linearly convergent sequences. This strategy makes the affine scaling method, asymptotically, behave like a predictor-corrector method. Reference [18] shows that a two-step method, i.e., one corrector step taken between each pair of predictor steps, attains a convergence rate of 1.5 ([24] shows a rate of 1.3) and a three-step method, i.e., two corrector steps taken between each pair of predictor steps, converges quadratically. Using the measure of Ostrowski [16], it can be shown that the three-step method is more "efficient" than the two-step method.

For each $r > 0.5$, we will consider in this paper the primal power affine scaling method based on the following approximating problem:

$$\begin{array}{ll} \text{minimize} & c^T x \\ & Ax = b, \\ & \| X_0^{-r}(x - x^0) \| \leq 1, \end{array} \qquad (3)$$

where $x^0 > 0$ is a given interior point of the linear program (1). Problem (3) is well defined for all $r > 0$. We note here that the sequences may not converge for values $0 < r \leq 1/2$, and thus these methods are not considered in this paper. When $r = 1$, the above approximating problem generates the primal affine scaling method, see for example, Barnes [3]. The method thus generated by choosing $r > 0.5$ is analogous to the power barrier method of primal-dual homotopy (barrier) method of den Hertog et al. [5] and Sheu and Fang [20]. Under the non-degeneracy assumption on the primal, we prove convergence of primal and dual sequences to respective optimal solutions for any step size $\alpha < 1$. To prove convergence under degeneracy, we introduce the concept of a power center of a polytope. We define two polytopes, with their power centers defined by maximization of concave functions. These functions are related to each other in the same manner as "dual norms" are. By using the connection of iterates of Newton's method applied to computing a power center of 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 $\alpha/(1-\alpha)^{2r} < 2/(2r-1)$, the primal sequence converges to the relative interior of the primal optimal face, and the dual sequence to the power center of the optimal dual face. In the second result, which gives the same convergence behavior of sequences as the first, we consider the case of variable step

size, and implement step size α_k at iteration k. We choose sequence $\{\alpha_k\}$ such that it is, asymptotically, defined by $\alpha_k/(1 - \alpha_k) < 2/(2r - 1)$ with α_k uniformly bounded away from $2/(2r + 1)$, which is $2/3$ for $r = 1$. Our result can thus be considered a generalization of the 2/3rd result of Tsuchiya and Muramatsu [25]. In both these cases, the proof is obtained by considering a merit function (which is the objective function of the center problem related to primal), which plays the same role as the local potential function used in analysis of affine scaling method [8,22,25].

By exploiting the relationship between Newton and affine scaling iterates, we present an accelerated version of the method, which generalizes the accelerated version of Saigal [18]. 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 $1 + r/(r + 1)$ and the three-step rate of $1 + 3r/(r + 2)$. These rates are 1.5 and 2 when $r = 1$, and thus generalize the two-step convergence rate and the three-step quadratic convergence rate of Saigal [18]. Using the measure of Ostrowski [16], 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.

Considering convergence and convergence rate results of the methods for $r \neq 1$, we note that they specialize to the corresponding results for the affine scaling method when the value of one is substituted for r. But analysis for the the affine scaling method is different, and thus its results are not obtainable as a corollary, by setting $r = 1$ in results for the power primal affine scaling method. When r is set to one, the objective function of power center problems must be changed to one defining an analytic center, and the merit function must be changed to a local potential function (which is also the objective function of analytic center problem). However, the power center approaches the analytic center as r approaches 1, even though the objective function of the power center problem does not approach the objective function of the analytic center problem. In this sense, the power affine scaling method is a proper generalization of the classical affine scaling method of Dikin [6].

In this paper, we restrict our attention to 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. In addition to the introduction, this paper has five other sections. In section 2, we present the primal power affine scaling method and obtain some properties of the sequences. In section 3, we prove convergence of the sequences to optimality with the primal non-degeneracy assumption, and in section 4, using a merit function, we prove convergence to optimality when the step size is a constant and when the step size is variable. In section 5, we present the accelerated primal affine scaling method, prove its convergence and derive its convergence rate. Finally, we end the paper with concluding remarks in section 6.

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, $k = 1, 2, \ldots$, 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, $V_{*,N}^p$ represents the diagonal matrix generated by v_N^* raised to the power p.

2.   The primal power affine scaling method

We now present the primal power affine scaling method generated when $r > 0.5$.

Step 0. Let x^0 be an interior point solution, $0 < \alpha < 1$, 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:

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

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

Step 3. Min-ratio test:

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

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

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.

We now comment on our choice of r. The method presented above is also defined for $0 < r \leq 1/2$. As can be seen from step 3, when the sequence $\{x^k\}$ converges to the boundary of the primal polyhedron and the sequence $\{s^k\}$ to a non-negative vector (this will happen when convergence is to optimality), $\phi(X_k^{2r-1}s^k) \to \infty$ if there is at least one pair of variables satisfying the strict complementarity condition. This makes the analysis of these methods different, and we do not consider them in this paper.

We can prove the following theorem related to this algorithm:

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 5. 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 4, we note that

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

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

$$c^T X_k^{2r} s^k = c^T X_k^{2r} (c - A^T y^k)$$

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

$$= \| P_k X_k^r c \|^2,$$

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.

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 4 we see that

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

and thus

$$1 = \alpha \frac{(x_l^k)^{2r-1} s_l^k}{\phi(X_k^{2r-1} s^k)}.$$

So $\alpha = 1$ and $(x_l^k)^{2r-1} s_l^k = \phi(X_k^{2r-1} s^k) \geq 0$. It then follows that $s_l^k \geq 0$ and, from step 3, that $\theta_k = 1$. Thus, $(x_l^k)^{2r-1} s_l^k = \| X_k^{2r-1} s^k \|$. Hence, for every $j \neq l$, $(x_j^k)^{2r-1} s_j^k = 0$, 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. $\qquad \square$

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.

Under assumption 4, we now establish some important properties of the sequences $\{x^k\}$, $\{y^k\}$, $\{s^k\}$ and $\{X_k s^k\}$ and show that the first and the fourth converge. For this purpose, consider the approximating problem defined for $k = 0$ by (3). Setting $p = x^k - x$, we obtain the equivalent problem:

$$\text{maximize} \quad c^T p$$

$$Ap = 0, \tag{5}$$

$$\| X_k^{-r} p \| \leq 1.$$

It is readily confirmed that the solution to problem (5) is

$$\hat{p}^k = \frac{X_k^{2r} s^k}{\| X_k^r s^k \|}. \tag{6}$$

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

THEOREM 2

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

$$c^T \hat{p}^k \geq \rho \| \hat{p}^k \|.$$

Proof

See corollary 6 of Saigal [17].   □

We can now prove the convergence of the primal sequence.

COROLLARY 3

Let assumption 4 hold. Then

1.   The sequence $\{x^k\}$ converges, say, to x^*.
2.   There is a $\rho > 0$ such that for every $k = 1, 2, \ldots$

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

Proof

To see part 1, let $\hat{p}^k$ be the solution to the approximating problem (5) given by (6). For each k, define $\gamma_k > 0$ such that $x^{k+1} - x^k = \gamma_k \hat{p}^k$. From theorem 2, we obtain the relation

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

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

To see part 2, 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

$$B = \{ j : x_j^* > 0 \},$$
$$N = \{ j : x_j^* = 0 \}.$$

The next theorem relates to the dual sequences $\{y^k\}$ and $\{s^k\}$, and is well known. We state it here without proof.

THEOREM 4

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

Proof

Follows from step 1 and theorem 4 of Saigal [17]. $\qquad\square$

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

THEOREM 5

Let assumption 4 hold, and $r \geq 1/2$. Then $X_k s^k \to 0$.

Proof

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

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

where $\alpha > 0$ and $\theta_k \geq 1$. Since $\{x^k\}$ converges and $\{y^k\}$ and $\{s^k\}$ are bounded when $r \geq 1/2$, the denominator of the above expression is bounded; thus

$$X_k^r s^k \to 0.$$

But this is only possible if $X_k s^k \to 0$, and we are done. $\qquad\square$

3. Optimality under non-degeneracy

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

THEOREM 6

Let the assumptions 1–4 hold, and let the primal be non-degenerate. Then there exist vectors x^*, y^* and s^* such that

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

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 4, let (y^*, s^*) be a cluster point of the sequence $\{(y^k, s^k)\}$. From theorem 5, $s_B^* = 0$. 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

$$y^k \to y^*$$

and thus $s^k \to s^*$. 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,

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

and thus $x_j^k \nrightarrow 0$, 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. $\square$

4. Optimality without non-degeneracy

In this section, we investigate the convergence to optimality without the non-degeneracy assumption. We will first develop some results on sequences, then introduce the power center of two polytopes and establish their relationship. We then establish a relationship between the Newton step for finding the power center and the affine scaling step, and then use this relationship to establish convergence to optimality. We use a merit function to establish this result.

4.1. MORE ON SEQUENCES

In this subsection, we will derive some important properties of sequences generated by the method. Consider the translated sequences:

$$v^k = \frac{x^k - x^*}{c^T x^k - c^*},$$

$$\tilde{u}^k = \frac{X_k s^k}{c^T x^k - c^*},$$

$$u^k = \frac{X_k^r s^k}{(c^T x^k - c^*)^r},$$

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

The following are simple consequences of the results already established.

PROPOSITION 7

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

Proof

Follows from corollary 3, part 2. ☐

PROPOSITION 8

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

Proof

Follows from the definition of u^k, theorem 4 and corollary 3, part 2. ☐

PROPOSITION 9

For every $k = 1, 2, \ldots, Ap^k = 0$.

Proof

Readily follows by substitution of definitions. ☐

Given B and N as defined by relations (7), we define the set of all possible dual estimates that are complementary to x^* as the polyhedron:

$$F_{\mathcal{D}} = \{(y, s) : A^T y + s = c, s_B = 0\}. \tag{9}$$

We can then prove:

PROPOSITION 10

$F_{\mathcal{D}} \neq \varnothing$.

Proof

From theorem 4, 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 5, it is readily established that $(y^*, s^*) \in F_{\mathcal{D}}$. ☐

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

LEMMA 11

For each $k = 1, 2, \ldots$:

1. $c^T x^k - c^* = \bar{s}_N^T x_N^k$.
2. There are constants $\rho_1 > 0$ and $r_2 > 0$ such that

$$\rho_1(c^T x^k - c^*) \leq \|x_N^k\| \leq \rho_2(c^T x^k - c^*).$$

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

Proof

Part 1 of the theorem follows from the following identity:

$$c^T x^k - c^* = \bar{s}^T(x^k - x^*) = \bar{s}_N^T x_N^k. \tag{10}$$

The upper bound of part 2 follows from corollary 3 part 2, and the lower bound from part 1. Part 3 readily follows from the identity (10).   □

As a consequence of lemma 11, we can define the polyhedron

$$\mathcal{V} = \{v : A_N v_N + A_B v_B = 0, \bar{s}_N^T v_N = 1, v_N \geq 0\} \tag{11}$$

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

We are now ready to prove two important results.

LEMMA 12

There exist $\rho_3 > 0$ and $\rho_4 > 0$ such that for every $k = 1, 2, \ldots,$

1. $\|X_k^r s^k\| \leq \rho_3 \phi(x_N^k)^r$.
2. $\|p_B^k\| \leq \rho_4 \phi(x_N^k)^r \|X_{k,N}^r s_N^k\|$.

Proof

Using the argument of the proof of theorem 1 and proposition 9, for $(\bar{y}, \bar{s}) \in F_{\mathcal{D}}$, we obtain:

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

and part 1 follows with $\rho_3 = \|\bar{s}_N\|$. To see part 2, note that from theorem 2 for some $\rho > 0$, $\|p_B^k\| \leq \|p^k\| \leq \frac{1}{\rho} c^T p^k = \frac{1}{\rho} \bar{s}_N^T p_N^k \leq \frac{1}{\rho} \|\bar{s}_N\| \phi(x_N^k)^r \|X_{k,N}^r s_N^k\|$ and the result follows from part 1 with $\rho_4 = \frac{1}{\rho} \rho_1$.   □

LEMMA 13

There are constants $\rho_5 > 0$, $\rho_6 > 0$ and $L \geq 1$ such that for all $k \geq L$,

1. $\|u_B^k\| \leq \rho_5 (c^T x^k - c^*)^r \|u_N^k\|$.

2. $\|s_B^k\| \leq \rho_6 (c^T x^k - c^*)^{2r}$.

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

4.
$$\frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} = 1 - \alpha \delta(u_N^k) \geq 0,$$

where

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

Proof

Since $x^k \to x^*$ and $x_B^* > 0$, there is a $\rho > 0$ and an $L \geq 1$ such that for all $k \geq L$, $\|X_{k,B}^{-r}\| \leq \rho$. Let $k \geq L$. Part 1 follows from corollary 3, part 2, lemma 12 and

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

To see part 2, note that

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

Substituting part 1, part 2 follows from proposition 8. Part 3 follows from proposition 7 and part 1. Part 4 follows readily by substitution, results of proposition 7 and part 1. $\qquad \square$

4.2. TWO POWER CENTERS AND THEIR RELATION

In this subsection, we consider the situation when $r > 1$, and the polyhedrons $\mathcal{V}$ and $F_{\mathcal{D}} \cap \{s : s_N \geq 0\}$ defined by (11) and (9), respectively. We define the power center of $\mathcal{V}$ as the solution to the following concave maximization problem:

$$\begin{aligned}
\text{maximize} \quad & -\sum_{j \in N} v_j^{-2(r-1)} \\
& A_N v_N + A_B v_B = 0, \\
& \bar{s}_N^T v_N = 1, \\
& v_N > 0,
\end{aligned} \tag{12}$$

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

$$2(r-1)V_N^{-(2r-1)}e - A_N^T z - \theta \bar{s}_N = 0, \tag{13}$$

$$- A_B^T z = 0, \tag{14}$$

$$A_N v_N + A_B v_B = 0, \tag{15}$$

$$\bar{s}_N^T v_N = 1. \tag{16}$$

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

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

$$A_N^T y + s_N = c_N, \tag{17}$$

$$A_B^T y = c_B,$$

$$s_N > 0,$$

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

$$\frac{2(r-1)}{2r-1} S_N^{-1/(2r-1)}e - u_N = 0, \tag{18}$$

$$A_N u_N + A_B u_B = 0, \tag{19}$$

$$A_N^T(y - \bar{y}) + (s_N - \bar{s}_N) = 0, \tag{20}$$

$$A_B^T(y - \bar{y}) = 0, \tag{21}$$

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

By modifying the objective function of power center problem (17) to

$$\frac{2r-1}{2(r-1)} \sum_{j \in N} s_j^{2(r-1)/(2r-1)},$$

we note that equation (18) is modified to

$$S_N^{-1/(2r-1)}e - u_N = 0.$$

And, as $r \to 1$, this approaches the analogous equation for K.K.T. conditions defining the analytic center of $F_{\mathcal{D}} \cap \{s : s_N \geq 0\}$, which uses $\sum_{j \in N} \log(s_j)$ as its objective function. Thus, the power center of $F_{\mathcal{D}} \cap \{s : s_N \geq 0\}$ approaches its analytic center as r approaches 1. This is curious, since the objective functions do not share this property.

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

THEOREM 14

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

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

Proof

Let (v^*, z^*, θ^*) be the solution to system (13)–(16). Then it can be verified that

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

$$y - \bar{y} = -\frac{1}{\theta^*} z^*,$$

$$u = \frac{2(r-1)}{2r-1} \left(\frac{2r-1}{\theta^*}\right)^{-1/(2r-1)} v^*$$

will solve the system (18)–(21).

Now let the center of $F_{\mathcal{D}} \cap \{s : s_N \geq 0\}$ exist and the system (18)–(21) have the solution (s^*, u^*, y^*). Then it can be verified that the transformation

$$\gamma = e^T S_{*,N}^{2(r-1)/(2r-1)} e,$$

$$\beta = \frac{2r-1}{2\gamma(r-1)},$$

$$\theta = 2(r-1)\gamma^{2r-1},$$

$$z = -\theta(y^* - \bar{y}),$$

$$v = \beta u^*$$

will solve the system (13)–(16). Here we have used $\bar{s}_N^T u_N = s_N^T u_N$, which is readily established using equations (19)–(21). $\qquad\square$

Please note that these power centers may not exist, since the concave functions involved are not bounded above, and we have not shown that the polyhedrons involved are bounded.

4.3. NEWTON'S METHOD AND POWER CENTER OF $\mathcal{V}$

In the previous subsection, we have seen that the power center of $\mathcal{V}$ is determined by solving the system of equations (13)–(16). This is a nonlinear system of equations, to which we can apply Newton's method to find its zero. The purpose of this subsection is to investigate this application.

Given v, z, θ, with $v \in \mathcal{V}$, the Newton direction Δv, Δz, $\Delta\theta$ is given by:

$$-2(r-1)(2r-1)V_N^{-2r}\Delta v - A_N^T\Delta z - \Delta\theta\bar{s}_N$$

$$= -2(r-1)V_N^{-2r+1}e + A_N^T z + \theta\bar{s}_N, \tag{22}$$

$$-A_B^T\Delta z = A_B^T z, \tag{23}$$

$$A_N\Delta v_N + A_B\Delta v_B = 0, \tag{24}$$

$$\bar{s}_N^T\Delta v_N = 0. \tag{25}$$

Consider the change of variables:

$$\hat{z} = \frac{z + \Delta z}{2(r-1)},$$

$$\hat{\theta} = \frac{\theta + \Delta\theta}{2(r-1)}, \tag{26}$$

$$\Delta v_B' = (2r-1)\Delta v_B,$$

$$w_N = (2r-1)V_N^{-r}\Delta v_N.$$

Substituting this change in the system (22)–(25), we can derive the following equivalent system with $\hat{A}_N = A_N V_N^r$ and $\hat{s}_N = V_N^r\bar{s}_N$,

$$w_N + \hat{A}_N^T\hat{z} + \hat{s}_N\hat{\theta} = V_N^{-(r-1)}e, \tag{27}$$

$$-A_B^T\hat{z} = 0, \tag{28}$$

$$\hat{A}_N w_N + A_B\Delta v_B' = 0, \tag{29}$$

$$\hat{s}_N^T w_N = 0. \tag{30}$$

We can then prove the following result:

LEMMA 15

Up to a choice of $\Delta v_B'$, the solution to the system (27)–(30) is unique. Also,

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

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 (27) by w_N^T, we obtain

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

The second and the third terms on the left-hand side of the above expression, from equations (28)–(30), 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 $x = V_N^{-(r-1)} e$. $\qquad\qquad \square$

For every $v > 0$, by an appropriate choice of a submatrix of A_B, we can guarantee that the system (27)–(30) is defined by a non-singular matrix. By a simple argument, it can be shown that different submatrices of A_B affect only the value of $\Delta v_B'$. The following is a well-known result relating to the rate and the convergence of Newton's method, not involving $\Delta v_B'$. We state this result without proof.

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

THEOREM 16

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

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

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

4.4.   AFFINE SCALING AND NEWTON STEPS IN $\mathcal{V}$

In this subsection, we will investigate the relationship between the affine scaling step interpreted in the polyhedron $\mathcal{V}$, and the Newton step for computing the power center of this polyhedron. We will show that there is a close connection between these steps, and this connection will be used in the sections that follow. We first investigate the affine scaling step.

The affine scaling step is defined by solving the optimization problem (5). Using part 1 of lemma 11 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

$$\bar{s}_N - A_N^T y - 2\theta X_{k,N}^{-2r} p_N = 0, \tag{31}$$

$$-A_B^T y - 2\theta X_{k,B}^{-2r} p_B = 0, \tag{32}$$

$$A_N p_N + A_B p_B = 0, \tag{33}$$

$$\| X_k^{-r} p \| = 1. \tag{34}$$

Using the definitions of u^k and v^k given in (8), and setting $u = u^k$, $v = v^k$ and

$$p_B' = \frac{p_B^k}{(c^T x^k - c^*)^r \| u \|},$$

$$\tilde{y} = \frac{-(c^T x^k - c^*)^r y^k}{2\theta \| u \|}, \tag{35}$$

$$\tilde{\theta} = \frac{(c^T x^k - c^*)^r \| u \|}{2\theta}$$

$$= 1,$$

we can rewrite the system (31)–(34) as:

$$\frac{u_N}{\| u \|^2} - \hat{A}_N^T \tilde{y} - \hat{s}_N \frac{\tilde{\theta}}{\| u \|^2} = 0, \tag{36}$$

$$-A_B^T \tilde{y} = \frac{-s_B^k}{\| u \|^2}, \tag{37}$$

$$\hat{A}_N \frac{u_N}{\| u \|^2} + A_B p_B' = 0, \tag{38}$$

$$\hat{s}_N^T \frac{u_N}{\| u \|^2} = 1, \tag{39}$$

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

The next proposition relates to the system that defines the affine scaling step.

PROPOSITION 17

Consider the systems represented by the equations (31)–(34) and (36)–(39).

1. (31)–(34) have a unique solution which generates a solution to (36)–(39).

2. The solution to (36)–(39) is unique up to a choice of p_B', and there is a value for p_B' for which the resulting solution also solves (31)–(34).

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

Proof

Since equations (31)–(34) represent the solution to a strictly convex problem, they have a unique solution. It is now readily confirmed, by simple algebra, that $u, \tilde{y}, \rho$ and p'_B, defined by the change of variables (35), solve the system (36)–(39). Thus we have proved part 1.

From part 1, it follows that (36)–(39) have a solution. Considering $q_N = u_N / \|u\|^2$, $\tilde{\rho} = \rho / \|u\|^2$, $\tilde{y}$ and p'_B as variables, this system is linear in these variables. If A_B has full column rank, the solution to (36)–(39) is unique, and part 3 follows. Otherwise, since (36)–(39) 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 (37) and (38) by

$$- A_C^T \tilde{y} = \frac{-s_B^k}{\|u\|^2} \tag{40}$$

and

$$\hat{A}_N \frac{u_N}{\|u\|^2} + A_C p'_C = 0, \tag{41}$$

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 (41) generates a solution to (38). Now, let $(q_N, \tilde{y}, \tilde{\rho}, p'_B)$ be any solution to (36)–(39). This then generates the unique solution $(q_N, \tilde{y}, \tilde{\rho}, p'_C - \Lambda p'_D)$ to (36) and (39)–(41). Since only the vector p'_B is modified in any solution to (36)–(39), part 2 is established with the required $p'_B = p_B / (c^T x^k - c^*) \|u\|$.   □

As a byproduct of the representation of the affine scaling step by the equations (36)–(39), we prove the following important property of the sequences.

LEMMA 18

There exist 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 (36) by $e^T V_{k,N}^{-(r-1)}$, we obtain

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

Substituting $A_N v_N = -A_B v_B$, and equation (37), we obtain

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

The theorem now follows from proposition 7 and lemma 13. □

Now consider the system (27)–(30), defining the Newton step in $\mathcal{V}$. 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 (27)–(30), we obtain the equivalent system:

$$w_N' - \hat{A}_N^T \hat{z} - \hat{s}_N \hat{\theta} = 0, \tag{42}$$

$$- A_B^T \hat{z} = 0, \tag{43}$$

$$\hat{A}_N w_N' + A_B \Delta \hat{v}_B = 0, \tag{44}$$

$$\hat{s}_N^T w_N' = 1. \tag{45}$$

Comparing systems (42)–(45) and (36)–(39), we note that if $\tilde{p}_N = u_N / \|u\|^2$,

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

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

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

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

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

$$M(v)a = b, \tag{46}$$

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 to the solutions of the two systems.

PROPOSITION 19

There are 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 $\|\Delta^k\| \leq \beta \|s_B^k\| / \|u\|^2$.

Proof

It is readily confirmed that system (46) represents the K.K.T. conditions of the following quadratic programming problem:

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

$$A_N a_1 + A_B a_4 = 0,$$

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

where we can assume, because of lemma 15 and proposition 17, that the columns of A_B are linearly independent. Thus, we can substitute

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

in the above optimization problem to obtain the equivalent problem

$$\text{minimize} \quad a_1^T V_{k,N}^r a_1 - a_4^T \tilde{s}_B^k$$

$$\tilde{A}_N a_1 = 0,$$

$$\bar{s}_N a_1 = 0,$$

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

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

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

Since

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

by defining

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

we obtain our result.   □

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

Consider the sequence $\{v_N^k\} \in \mathcal{V}_N$ generated by the affine scaling method. We now prove the following corollary to proposition 19 which establishes the connection between an affine scaling step and the Newton step.

COROLLARY 20

Let $v_N^k \in \mathcal{V}_N$. Then

1. The affine scaling direction at v_N^k is

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

where

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

2. 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

To see the proof of part 1, by simple substitutions, we note that

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

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

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

Part 2 follows from the change of variables (26) and proposition 19.   □

4.5. CONVERGENCE TO OPTIMALITY

We now investigate the convergence to optimality of primal and dual sequences, without the non-degeneracy assumption. This proof closely relates to the

proof of the 2/3rd result for $r = 1$. In that case, a local potential function is analyzed to establish the convergence of the dual sequence. Here, we will instead use a merit function, which also defines the power center of $\mathcal{V}$. We will introduce this function after proving a technical lemma.

LEMMA 21

There are an $L \geq 1$, $\beta > 0$, $\gamma > 0$, $\delta > 0$ and $\delta_1 > 0$ such that for every $k \geq L$,

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

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

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

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

4. $(c^T x^{k+1} - c^*) \leq \delta(c^T x^k - c^*)$ and thus $\sum_{k=0}^{\infty} (c^T x^k - c^*) < \infty$.

Proof

From lemma 18 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, $\|u_N^k\| \to 0$. Then, for $k \in K$,

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

or

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

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

$$v_l^k \to 0.$$

But $\tilde{u}_l^k = v_l^k s_l^k$. Thus, $s_l^k \to \infty$, a contradiction, as theorem 4 implies that the dual variables s are bounded, and part 1 follows.

To see part 2, from proposition 19,

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

$$= \phi(e - w_N^k + \Delta^k)$$

$$\leq 1 + \|w_N^k\| + \delta_k, \tag{47}$$

where $\|\Delta^k\| = \delta_k$; and the lower bound follows from lemma 18 and part 1 for each real $a > -1$, $1/(1 + a) \geq 1 - a$.. The upper bound follows from the fact that lemma 13 part 4 holds for every $\alpha \leq 1$. Part 3 follows from part 1 and propositions 7 and 8. Part 4 follows from part 3 and lemma 13 part 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 it is the negative of the objective function of the power center problem on $\mathcal{V}$.

We now prove a simple lemma related to this function:

LEMMA 22

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

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

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

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

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

Proof

Part 1 follows from the following identity, obtained by using the mean value version of Taylor's formula and noting that $w_j \geq 0$ $\hat{w}_j \geq 0$. 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 $\hat{w}_j \geq -\phi(-w)$:

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

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

and we have our result.   □

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

PROPOSITION 23

Let
$$\tilde{\alpha} = \alpha \, \frac{\|u^k\|^2}{\phi(V_k^{r-1}u^k)} \quad \text{and} \quad \hat{w}_N^k = (2r-1)V_{k,N}^{-1}\Delta v_N^k.$$

Then, there are an $L \geq 1$ and a $\beta > 0$ such that for every $k \geq L$

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

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

Proof

From step 4 and proposition 19, for each $j \in N$,

$$\frac{x_j^{k+1}}{x_j^k} = 1 - \alpha \, \frac{\|u^k\|^2}{\phi(V_k^{r-1}u^k)} \frac{(v_j^k)^{r-1}u_j^k}{\|u^k\|^2}$$

$$= 1 - \tilde{\alpha} \, \frac{(v_j^k)^{r-1}u_j^k}{\|u^k\|^2}$$

$$= 1 - \tilde{\alpha}(1 - \hat{w}_j^k + \hat{\Delta}_j^k),$$

where $\hat{\Delta}^k = V_{k,N}^{r-1}\Delta^k$. From the above, definitions and lemma 13 parts 3 and 4, we obtain

$$\frac{v_j^{k+1}}{v_j^k} = \frac{x_j^{k+1}}{x_j^k} \frac{c^T x^k - c^*}{c^T x^{k+1} - c^*}$$

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

$$> 0.$$

Let $\theta_j^k = v_j^{k+1}/v_j^k - 1 = \tilde{\alpha}/(1 - \tilde{\alpha})(\hat{w}_j^k - \hat{\Delta}_j^k)$, and note that $\theta_j^k > -1$. First, consider the case when $1 > \phi(-\theta^k) > 0$. Then, from lemma 22 part 2, we obtain

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

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

$$
\left. -\frac{(2r-1)\tilde{\alpha}}{2(1-\tilde{\alpha})(1-\phi(-\theta^k))^{2r}}(\hat{w}_N^k - \hat{\Delta}^k)^T V_{k,N}^{-2(r-1)}(\hat{w}_N^k - \hat{\Delta}^k) \right).
$$

Using proposition 19 and lemma 13 part 3, we obtain

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

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

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

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

Thus, from lemma 15, as $1 - \tilde{\alpha} < 1$, we obtain

$$
F_N(x^{k+1}) - F_N(x^k)
$$

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

$$
\leq -\frac{2\tilde{\alpha}(r-1)}{1-\tilde{\alpha}}\left(\left(1 - \frac{(2r-1)\tilde{\alpha}}{2(1-\alpha)^{2r}}\right)(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)}\hat{w}_N^k + \mu_k\right), \tag{48}
$$

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 22 part 1 must be used to calculate this change. By an identical analysis as above, in this case we obtain

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

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 22 is larger. As $-\phi(-v_N^k) = v_{j_k}^k = \min_{j \in N} v_j^k$ we see, using lemma 15 part 2

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

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

and our result follows from proposition 19 and lemma 21 part 1. □

We are now ready to prove the main convergence theorem.

THEOREM 24

Let $\alpha/(1-\alpha)^{2r} < 2/(2r-1)$, and assumptions 1–4 hold. Then there exist vectors x^*, y^* and s^* such that

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

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

From lemma 21 part 3, we see that there is a $\gamma > 0$ such that $\tilde{\alpha}_k \ge \gamma\alpha$. 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 lemma 21 upper bound in part 2, $\alpha \ge \tilde{\alpha}_k$. Thus,

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

Let $v_{j_k}^k = \min_j v_j^k$, K be the subsequence such that for each $k \in K$, $v_{j_k}^k \le \delta$ with $\delta > 0$ and $1 - 2Mq\delta^{r-1} > 0$ when $M > 0$ is an upper bound on $\|u_N^k\|$. Any cluster

point π of the subsequence satisfies $\pi \leq \delta$. From proposition 19, there is an $\hat{L} \geq 1$ such that for all $k \geq \hat{L}$,

$$w_{j_k}^k = 1 - \frac{u_{j_k}^k (v_{j_k}^k)^{r-1}}{\|u^k\|^2} \geq \frac{1 - 2qM\delta^{r-1}}{2} > 0,$$

$$(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)} \hat{w}_N^k \geq \mu_2,$$

where $\mu_2 = 0.25(1 - 2Mq\delta^{r-1})^2 (v_{j_k}^k)^{-2(r-1)}$. Thus, from proposition 23,

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

where

$$|\varepsilon_k| = \frac{|\mu_k|}{(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)} \hat{w}_N^k} \leq \mu_2 \beta (v_{j_k}^k)^{r-1} (c^T x^k - c^*)^{2r},$$

and from proposition 7, $\varepsilon_k \to 0$. 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 3 part 2, there is a $\rho > 0$ such that

$$(\hat{w}_N^k)^T V_{k,N}^{-2(r-1)} \hat{w}_N^k \geq \frac{\|\hat{w}_N^k\|^2}{\phi(v_N^k)^{2(r-1)}}$$

$$\geq \|\hat{w}_N^k\|^2 \left(\frac{c^T x^k - c^*}{\|x_N^k\|} \right)^{2(r-1)}$$

$$\geq \rho^{2(r-1)} \|\hat{w}_N^k\|^2.$$

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, for each $k \notin K$,

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

Thus, from proposition 23 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

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

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

From lemma 21 part 4,

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

From corollary 3 part 2, $F_N(x^k)$ is bounded below by zero on this sequence. Thus, $\|\hat{w}_N^k\| \to 0$. As $\hat{w}_N^k = (2r-1)V_{k,N}^{-1}\Delta v_N^k$, from proposition 7,

$$\Delta v_N^k \to 0$$

and, from proposition 19,

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

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

$$v_N^k \to v_N^*.$$

Let $\overline{K}$ be such that for $k \in \overline{K}$ (it exists since all these sequences are bounded) for some vectors x^*, y^* and s^*,

$$s^k \to s^*,$$
$$y^k \to y^*,$$
$$u^k \to u^*,$$

and so

$$\frac{V_{k,N}^{r-1}u_N^k}{\|u^k\|^2} \to \frac{V_{*,N}^{r-1}u_N^*}{\|u^*\|^2} = e. \tag{49}$$

Using theorem 14, 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. $\square$

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

PROPOSITION 25

Let $\alpha/(1-\alpha)^{2r} < 2/(2r-1)$. Then

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

Proof

Follows readily from proposition 19, lemma 21 part 2 and equation (49).

$\square$

Theorem 24 proves the convergence to optimality for a constant step size $\alpha > 0$ determined such that

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

For $r = 1$, this requires $\alpha < 1/2$. To obtain a result analogous to the 2/3rd for the standard affine scaling method, we introduce a variable step strategy, i.e., 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

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

which for $r = 1$ gives the required 2/3rd. We obtain this increase by using the following estimate for $\delta(u^k)$ (of lemma 21)

$$\tau_k = \frac{\|X_k^r s^k\|^2}{\phi(X_k^{2r-1}s^k)(x^k)^T s^k}. \tag{50}$$

We now establish the goodness of the estimate (50).

LEMMA 26

Let τ_k be defined by equation (50). There are an $L \geq 1$ and a $\beta > 0$ such that for all $k \geq L$

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

where $|\delta_k| \leq \beta(c^T x^k - c^*)$ and $\delta(u_N^k) = \|u^k\|^2/\phi(V_{k,N}^{r-1}u_N^k)$.

Proof

This result follows readily from lemmas 13 part 2, 18 and the upper bound of part 2 of lemma 21.

$\square$

Since τ_k is, asymptotically, a very good estimate of $\delta(u_N^k)$, in place of using $\alpha/(1-\alpha)^{2r}$ as an estimate of $\tilde{\alpha}(1-\tilde{\alpha})^{2r-1}/(1-\alpha)^{2r}$ in the relations (48), we will use the estimate $\tau_k \alpha$ of $\tilde{\alpha}$ in the above formula. Also, from lemma 21 part 2, whenever $\|w_N^k\| \to 0$, $\delta(u_N^k) \to 1$. Thus, in this case $\tau_k \to 1$, and the new estimate approaches $\alpha/(1-\alpha)$. 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

$$\frac{\tau_k\alpha'(1-\tau_k\alpha')^{2r-1}}{(1-\alpha')^{2r}} = \frac{2}{(2r-1)} - \theta. \tag{51}$$

Step 2. Define

$$\alpha_k = \begin{cases} \alpha' & \text{if } \alpha' \geq \alpha^*, \\ \alpha^* & \text{otherwise.} \end{cases} \tag{52}$$

The above choice guarantees that the estimate is not smaller than one obtained from theorem 24. The next lemma establishes a relationship between the α' computed in (51) and the related expression in system (48). Note that the non-linear system (51) has to be solved to obtain α'.

LEMMA 27

Let α' be computed by (51) and let $\tilde{\alpha} = \delta(u_N^k)\alpha'$. Then, there are 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 $|\delta_k| \leq \beta(c^Tx^k - c^*)^{2r}$.

Proof

Follows readily from lemma 26, and the fact that for small $\varepsilon > 0$, $(1+\varepsilon)^{2r-1} < 1 + 4^r\varepsilon$. $\qquad\square$

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

THEOREM 28

For each k, let α_k be generated by the above rules and let the assumptions 1–4 hold. Then, there exist vectors x^*, y^* and s^* such that

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

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 24. Otherwise, for each k for which $\alpha_k = \alpha' > \alpha^*$, using the same argument of theorem 24, it is readily shown that for some $\beta > 0$,

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

where $\mu_k \le \gamma(c^T x^k - c^*)^{2r}$. The proof is completed in the same way as in theorem 24. $\square$

We now obtain the asymptotic behavior of α_k.

COROLLARY 29

$\alpha_k \to \hat\alpha$, where

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

Proof

As a consequence of theorem 28, we obtain the fact that $\|w_N^k\| \to 0$. From lemma 21, $\delta(u_N^k) \to 1$, from lemma 26, $\tau_k \to 1$, and we have our result. $\square$

5.   Accelerated primal power affine scaling method

We will now use the connection between the Newton and the Affine Scaling step, developed in section 4, to accelerate the convergence of the method. We first present the accelerated version of the method and then investigate its convergence and convergence rate. This accelerated primal power affine scaling method is generated by replacing steps 3 and 4 of the method described in section 2 by the following three steps, where $0 < \delta < 1$ is a constant whose value will be specified later:

Step 3.1. Min-ratio test:

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

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

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

Step 3.2. Step size selection: If $e^T X_k s^k \geq 1$, set $\alpha_k = \alpha$ and go to step 4. Otherwise, define

$$\alpha'_k \quad \text{defined by (52)},$$

$$N_k = \left\{ j : x^k_j \leq \sqrt{(x^k)^T s^k} \right\},$$

$$\gamma_k = e^T X_{k,N_k} s^k_{N_k},$$

$$h^k_{N_k} = \frac{1}{2r-1} \left(\frac{X_{k,N_k}}{\gamma_k} - \frac{X^{2r-1}_{k,N_k} s^k_{N_k}}{\|X^r_k s^k\|^2} \right),$$

$$\varepsilon_k = \|h^k_{N_k}\|,$$

$$\rho_k = \min \left\{ 3r + 2, \frac{\log(\varepsilon_k)}{\log(\gamma_k)} \right\}.$$

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

$$\alpha_k = 1 - \max \left\{ \varepsilon^\delta_k, \gamma^{\delta \rho_k}_k \right\}.$$

2. Corrector step: Otherwise,

$$\alpha_k = \min \left\{ \frac{\gamma_k \phi(X^{2r-1}_{k,N_k} s^k_{N_k})}{2r \|X^r_k s^k\|^2}, \alpha'_k \right\}.$$

Step 4. Next interior point:

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

Some comments are in order here. This acceleration scheme is identical to the two-step acceleration scheme of Saigal [18]. h^k_N computed in step 4 is a very good estimate of the Newton step Δv^k_N, and its magnitude ε_k is used to estimate the distance to the power center v^*_N of $\mathcal{V}_N \cap \{v_N : v_N \geq 0\}$. 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, $\|\Delta v^k_N\|$ is a very good estimate of $\|v^k_N - v^*_N\|$. During the corrector step, α_k is chosen so that

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

and thus the affine scaling step behaves, asymptotically, like a Newton step. We now establish three straightforward results and then prove the convergence of this accelerated method.

LEMMA 30

Let N be defined by the system (7). There is an $\hat{L} \geq 1$ such that for all $k \geq \hat{L}$, $N_k = N$, where N_k is defined in step 3.2.

Proof

From lemmas 13 part 2 and 18, there are $\rho_1 > 0$, $\pi_2 > 0$ and $L \geq 1$ such that for all $k \geq L$ we have $\|\tilde{u}_B^k\| = \|X_{k,B}s_B^k\|/(c^T x^k - c^*) \leq \rho_1(c^T x^k - c^*)^{2r-1}$ and $e^T \tilde{u}_N^k = 1 + \delta_k$ with $|\delta_k| \leq \rho_2(c^T x^k - c^*)^{2r}$. Thus, for some $\delta_3 > 0$, $\hat{L} \geq L$ and all $k \geq \hat{L}$, we have

$$e^T X_k s^k = (c^T x^k - c^*)(e^T \tilde{u}_B^k + e^T \tilde{u}_N^k)$$

$$\geq 0.50(c^T x^k - c^*),$$

$$e^T x_N^k \leq \rho_3(c^T x^k - c^*),$$

$$x_j^k > \sqrt{e^T X_k s^k} \qquad \text{for all } j \in B,$$

$$x_j^k < \sqrt{e^T X_k s^k} \qquad \text{for all } j \in N,$$

where the second inequality follows from lemma 11 part 2. The third inequality follows from the fact that $x_j^* > 0$ for each $j \in B$ and the fourth from first and second inequalities.   ☐

And now a corollary to theorem 16.

COROLLARY 31

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 16, part 1.   ☐

Another lemma follows:

LEMMA 32

Let $\hat{L}$ be generated by lemma 30, and $M > 0$ be large. There is an $\bar{L} \geq \hat{L}$, $\theta_1 > 0$ and $\theta_2 > 0$ such that for all $k \geq \bar{L}$,

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

2. $0.50(c^T x^k - c^*) \leq \gamma_k \leq 1.50(c^T x^k - c^*)$.

3.   Whenever $\| v_N^k - v_N^* \| \le \beta (c^T x^k - c^*)^p$, with $M > \beta > 0$ and $p \ge 1.5r$, then

$$\| w_N^k \| \le \bar{\theta} (c^T x^k - c^*)^p .$$

Proof

We note that by substitution of the results of lemma 18 and corollary 20 part 2, and definitions, we get

$$h_N^k = \frac{1}{2r-1} \left(\frac{v_N^k}{e^T \tilde{u}_N^k} - \frac{V_{k,N}^r u_N^k}{\|u^k\|^2} \right)$$

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

$$- \frac{1}{2r-1} \left(V_{k,N}^r \Delta^k + \frac{\delta_k}{1+\delta_k} v_N^k \right)$$

$$= \Delta v_N^k - t^k .$$

As v_N^k is bounded by proposition 7, part 1 follows. Part 2 readily follows from lemma 18. To see part 3, since $v_N^* > 0$, there is an $\bar{L} \ge \hat{L}$ such that for all $k \ge \bar{L}$,

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

Thus, under the hypothesis of the lemma,

$$v_j^k \ge v_j^* - |v_j^* - v_j^k| \ge a$$

and our result now follows by observing from the change of variables (26), that $\| w_N^k \| \le (2r-1)(\min_j v_j^k)^{-r} \| \Delta v_N^k \|$, and the result of corollary 31.   □

We now use the connection of the power affine scaling method to Newton's method computing the power center v_N^* of $\mathcal{V}_N$ and establish convergence and convergence rate of the accelerated algorithm. We will give a choice of the constant $\delta > 0$ for both the two-step and the three-step cases, and obtain a rate which is a function of r. This rate approaches two-step quadratic as r approaches infinity. We now investigate the predictor and the corrector steps under the assumption that power center v_N^* of $\mathcal{V}_N$ exists. This will be implicit in the hypothesis of the propositions.

We are now ready to investigate the predictor step.

PROPOSITION 33

Let $\bar{L}$ be as in lemma 32. Assume that $0 < \delta < 1$, $k \ge \bar{L}$, $M > \beta > 0$, $1.5r \le p$ and $\| v_N^k - v_N^* \| \le \beta (c^T x^k - c^*)^p$. Then

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

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

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

$$\| v_N^{k+1} - v_N^* \| \leq \theta_3 (c^T x^{k+1} - c^*)^{\rho_k'/(1 + \delta \rho_k)},$$

where $\rho_k^* = 1 + \min\{p, \delta \rho_k\}$ and $\rho_k' = \min\{2p, (1 - \delta)\rho_k\}$.

Proof

From corollary 31, $\| \Delta v_N^k \| \leq 1.5 \beta (c^T x^k - c^*)^p$. Also, by lemma 32,

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

We note the following sequence of inequalities which follow from lemmas 13, 21, 32 part 2, 32 part 3 and some $\bar{\theta} > 0$:

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

$$\leq 1 - \alpha_k$$

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

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

$$\leq 1 - \left(1 - \max\left\{\varepsilon_k^\delta, \gamma_k^{\delta \rho_k}\right\}\right)(1 - \| w_N^k \| - |\delta_k|)$$

$$\leq \max\left\{\varepsilon_k^\delta, \gamma_k^{\delta \rho_k}\right\} + \| w_N^k \| + |\delta_k|$$

$$\leq \bar{\theta}(c^T x^k - c^*)^{\min\{p, \delta \rho_k\}},$$

and we have part 1. From corollary 20 parts 1 and 2, and some simple algebra, we obtain

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

where

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

Using the facts that $\delta(u_N^k) \le 1$, $\alpha_k \le 1 - \varepsilon_k^\delta$, we obtain, for some $\hat\theta > 0$,

$$\|t^k\| \le \hat\theta(c^T x^k - c^*)^{(1-\delta)\rho_k}.$$

Now for appropriate positive constants, from theorem 16 we see that

$$\|v_N^{k+1} - v_N^*\| \le \|v_N^k + \Delta v_N^k - v_N^*\| + \|t^k\|$$

$$\le \rho_2 \|v_N^k - v_N^*\|^2 + \hat\theta(c^T x^k - c^*)^{(1-\delta)\rho_k}$$

$$\le \rho_2'(c^T x^k - c^*)^{2p} + \hat\theta(c^T x^k - c^*)^{(1-\delta)\rho_k}$$

and part 2 follows from the lower bound of part 1. $\qquad\square$

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

PROPOSITION 34

Let $\bar{L}$ be as in lemma 32. Assume that for some $k \ge \bar{L}$ and $\beta > 0$, $\|v_N^k - v_N^*\| \le \beta(c^T x^k - c^*)^p$. In the case

1. $1.5r/2 \le p \le r$, one corrector step will be taken, after which, for some $\theta_1 > 0$,

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

2. $1.5r/4 \le p \le r/2$, at least one corrector step will be taken, and after at most two steps, for some $\theta_2 > 0$,

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

Proof

Let

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

From lemmas 18 and 21 part 2, $2r\mu_k \ge e^T \tilde{u}_k = 1 + \delta_k$. From equation (47), $2r\mu_k \le (1 + \delta_k)(1 + \|w_N^k\| + \|\Delta_k\|)$. Thus,

$$\frac{1}{2} - |\delta_k'| \le \mu_k \le \frac{1}{2r} + |\rho_k'|,$$

where, using lemma 32 part 3, we note that $|\delta_k'| \le \theta^*(c^T x^k - c^*)^{2r}$ and $|\rho_k'| \le \theta'(c^T x^k - c^*)^p$. From corollary 29, as α_k' approaches $2/(2r+1) - \varepsilon$ for some small $\varepsilon > 0$, for all sufficiently large k, at step 3.2,

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

Thus, from lemma 13 part 1, corollary 20 parts 1, 2, and simple algebra, we see that

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

$$= \Delta v_N^k - t^k,$$

where

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

Using propositions 7, 8 and lemma 21 part 2, we can show that

$$\|t^k\| \le \beta'(c^T x^k - c^*)^{2r}.$$

Thus, after one step with α_k, we see that (using theorem 16),

$$\|v_N^{k+1} - v_N^*\| \le \|v_N^k + \Delta v_N^k - v_N^*\| + \|t^k\|$$

$$\le \rho'\|v_N^k - v_N^*\|^2 + \beta'(c^T x^k - c^*)^{2r} \qquad (53)$$

$$\le \beta^*(c^T x^k - c^*)^{2p}.$$

From, lemma 13 part 4,

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

$$= 1 - \frac{e^T \tilde{u}_N^k}{2r},$$

and using lemma 18, we see that for all sufficiently large k,

$$1 - \frac{1.5}{2r} \le \frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} \le 1 - \frac{0.5}{2r}. \qquad (54)$$

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

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. $\qquad \square$

We are now ready to prove the main convergence theorem.

THEOREM 35

Let the sequences $\{x^k\}$, $\{y^k\}$ and $\{s^k\}$ be generated by the accelerated method with $r > 1$, and let assumptions 1–4 hold. Then, there exist vectors x^*, y^* and s^* such that

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

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 $\delta = 1/2(r + 1)$, the convergence is two-step superlinear at the rate $1 + r/(r + 1)$.

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

Proof

We now show that asymptotically a predictor step must be taken. Assume that there is an $L \geq 1$ such that for every $k \geq L$, a corrector step is taken. Then α_k is selected by the variable step selection strategy, from theorem 28, $\{v_N^k\}$ converges to the power center v_N^*.

As in the proof of proposition 34, from equations (53) and (54) we obtain

$$\|v_N^{k+1} - v_N^*\| \leq \rho \|v_N^k - v_N^*\|^2 + \hat{\beta}(c^T x^{k+1} - c^*)^{2r}.$$

Thus, after several such corrector iterations, $\|v_N^{k+l} - v_N^*\| \leq \rho'(c^T x^{k+l} - c^*)^p$ for $p \geq 1.5r$ and $l \geq 1$. From corollary 31 and lemma 32, $\rho_{k+l} \geq 1.5r$ and a predictor step must be taken, and we have a contradiction. We note that the constant $\rho' > 0$ is independent of k, and is the required M in proposition 33.

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 $\delta = (\rho_k - r)/(\rho_k(r + 1))$. By the choice of ρ_k, at step 3.2, for $p \geq 1.5r$, $\rho_k' = \min\{2p, (1 - \delta)\rho_k\} = (1 - \delta)\rho_k$ after one corrector step, from propositions 33 and 34, we obtain

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

where $2(1 - \delta)\rho_k/(1 + \delta\rho_k) = 2r$. Thus, part 1 follows as

$$\delta = \frac{1}{2(r - 1)} \quad \text{when } \rho_k = 2r,$$

and the convergence rate obtained is

$$\rho^* = 1 + \min\{2r, \delta\rho_k\} = 1 + \delta\rho_k = 1 + \frac{r}{r + 1}.$$

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

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

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

We now investigate the efficiency of the asymptotic convergence rates obtained, and thus get some measure of the relative effectiveness of the acceleration. For this purpose, we will use the measure introduced by Ostrowski [16, section 6.11] to compare algorithms achieving different asymptotic rates of convergence, and requiring different amounts of work per iteration. He suggested the following measure:

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

where ρ is the asymptotic convergence rate of the acceleration, and w is a measure of the work per iteration. The larger this measure, the more efficient the acceleration. This measure has been used by Brent [4], who investigated the hybrids of Newton's method proposed by Shamanskii, and by Saigal and Todd [19], who investigated the hybrids of fixed point computing methods with variants of Newton's method.

The asymptotic convergence rate of the accelerated power primal affine scaling method depends on the choice of r and the two-step or three-step method. Table 1 shows these calculations for several choices.

Table 1

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

[1] Tsuchiya and Monteiro [24] obtain a rate of 1.3. This can be established by the method of Saigal [18].

[2] This is obtained in Saigal [18].

[3] The efficiency of the three-step cubic is greater than the two-step quadratic, which is $0.3466/w$.

6. Concluding remarks

In this paper, we have shown that for every $r > 0.50$, there is a variant of the affine scaling method, which we call the primal power affine scaling method. The usual method is generated when $r = 1$. We have analyzed the convergence of these methods 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 center associated with the optimal primal face and the power center of the optimal dual face 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 $\alpha/(1 - \alpha)^{2r} < 2/(2r - 1)$, 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 $\alpha_k/(1 - \alpha_k) < 2/(2r - 1)$. This sequence is required to stay uniformly away from from $2/(2r - 1)$. Thus, $\alpha_k < 2/(2r + 1)$, and hence this result is a generalization of the 2/3rd result of Tsuchiya and Muramatsu [25] for $r = 1$.

An accelerated primal power affine scaling method is also presented. This method achieves superlinear convergence, and the rate is higher for larger values of $r > 1$. This generalizes the work of Saigal [18] and Tsuchiya and Monteiro [24].

This work opens up the study of hybrid variants of the power affine scaling method in which different values of r are implemented at different stages of the method. From lemmas 11 and 13, 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 Programming 44(1989)297–335.

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

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

[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, Recherche Opérationelle 28(1994)135–163.

[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 (1991).

[9] I.I. Dikin, Determination of interior point of one system of linear inequalities, Kibernetica and System Analysis 1(1992)74–96.

[10] C.C. Gonzaga, private communication (1993).

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

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

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

[14] 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.

[15] 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.

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

[17] 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 (1992), Annals of Operations Research, this volume.

[18] 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 (1993).

[19] R. Saigal and M.J. Todd, Efficient acceleration techniques for fixed point algorithms, SIAM Journal on Numerical Analysis 2(1978)997–1007.

[20] R.L. Sheu and S.C. Fang, On the generalized path-following methods for linear programming, Technical Report, O.R. Technical Report No. 261, North Carolina State University, Raleigh, NC 27695 (1992), to appear in Optimization.

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

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

[23] T. Tsuchiya, Global convergence property of the affine scaling method for primal degenerate linear programming problems, Mathematics of Operations Research 17(1992)527–557.

[24] T. Tsuchiya and R.D.C. Monteiro, Superlinear convergence of the Affine Scaling Algorithm, Research Memorandum, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Tokyo, Japan (1992).

[25] 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), to appear in SIAM Journal on Optimization.

[26] R.J. Vanderbei, M.J. Meketon and B.A. Freedman, A modification of Karmarkar's linear programming algorithm, Algorithmica (1986)395–407.

[27] R.J. Vanderbei and J.C. Lagarias, I.I. Dikin's convergence result for the affine-scaling algorithm, in: *Mathematical Developments Arising from Linear Programming: Proceedings of a Joint Summer Research Conference*, ed. J.C. Lagarias and M.J. Todd, Bowdoin College, Brunswick, Maine, USA (1988), vol. 114 of Contemporary Mathematics (American Mathematical Society, Providence, RI, USA, 1990) pp. 109–119.