A SIMPLIFIED PROOF OF THE GENERAL CONVERGENCE OF AFFINE SCALING

John R. Birge
Charles Rosa

Department of
Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109-2117, USA

Technical Report 91-7

March 1991
Revised September 1991
A Simplified Proof of the Convergence of Affine Scaling

John R. Birge*, Charles Rosa

Abstract: This paper analyzes the convergence of the (dual) affine scaling interior point algorithm for linear programs. The analysis includes a primal nondegeneracy assumption and allows full steps in the normalized search direction. The proof is motivated within the geometry of the problem and does not rely on reference to the projective scaling step.

* Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109, USA. His work was supported by Office of Naval Research Grant N0014-86-K-0628 and the National Science Foundation under Grant ECS-885101.
1. Introduction

Following Karmarkar's [8] projective scaling algorithm, several authors proposed affine scaling methods (Vanderbei, Meketon and Freedman [12], Barnes [3]) and showed convergence under primal and dual nondegeneracy. The method had actually been proposed earlier by Dikin [5] who also gave a proof of convergence only using primal nondegeneracy([6]). Vanderbei and Lagarias [13] later refined this proof and answered some open questions concerning it.

All convergence proofs assumed some form of nondegeneracy until Tsushiya's [11] proof that affine scaling (in the "dual" form also proposed by Adler, et al.[1] and Monma and Morton[9] for affine scaling and by Freund[7] for projective scaling) is convergent without any degeneracy assumptions but with a maximum step size of 1/8 the maximum distance within an inscribed ellipsoid. We also use this inequality form of affine scaling (which is equivalent to the equality primal form as noted by Tsushiya and explained in Birge and Freund[4]). We derive a proof with full step sizes in this form but again include a primal degeneracy assumption. A key to the result is a lemma from Vanderbei and Lagarias about the boundedness of a fundamental transformation.

2. General Problems and Affine Scaling

We consider a linear program of the form:

$$\max \quad c^Tz$$
subject to
$$Ax \leq b.$$  

(1)

where $z \in \mathbb{R}^n$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, and a finite optimal value $c^Tz^* = z^*$. We assume the following:

(A1) The feasible region has an interior point, the set of optima is nonempty, and $c \neq 0$;

(A2) $A$ has full rank;

(A3) The set of optima in (1) is bounded;
(A4) Every extreme point of the feasible region is nondegenerate.

We also suppose that a strict interior feasible point is available (\( z^0 \) such that \( Az^0 < b \)). We may obtain such a starting value using phase one types of procedures as, for example, in Anstreicher [2], Karmarkar [8].

We use the form in (1) because of its generality and intuitive geometry. It is often called the dual form of a linear program with nonnegative variables and equality constraints, but it may be derived directly. Equality constraints may also be added to (1) which would complicate the following analysis through an additional projection, although the results would remain the same.

We now give the affine scaling method. Our approach is similar to the original procedure in Dikin [5,6]. Given an initial interior point, the procedure creates a sequence of points \( z^0, z^1, ..., z^k \) by following the basic steps of finding a search direction, \( u^k \), choosing a step size, \( \sigma^k \), and updating the iterate, \( z^{k+1} = z^k - \sigma^k u^k \). We describe these steps below.

**Affine Scaling Algorithm**

**Step 1. Search direction.** Consider an iterate \( z^k \) and a positive diagonal matrix, \( W^k = \text{diag}(w^k) \). Often, the matrix \( W^k \) is composed of the slack variables, \( s^k = b - Az^k > 0 \), so that \( w^k = s^k \). This choice is not necessary (as long the ratios between similar elements remain bounded), but we will use it to simplify the analysis.

In any ascendant optimization procedure, we would like to find a step \( v \) that improves the objective and is feasible, i.e., such that

\[
A(z^k + v) \leq b, \text{or } (W^k)^{-1}Av \leq (W^k)^{-1}(b - Az^k) = e, \quad (\text{if } s^k = w^k),
\]

and \( c^Tv > 0 \). Let this set of feasible directions be \( V^k \), which we note depends on \( z^k \).

An intuitive simplifying transformation is to let \( y = (W^k)^{-1}Av \). Its inverse transformation is

\[
v = (A^T(W^k)^{-2}A)^{-1}A^T(W^k)^{-1}y.
\]
The objective, $c^Tv$, becomes $d^Ty$, where $d^T = c^T(A^T(W^k)^{-2}A)^{-1}A^T(W^k)^{-1}$. The region in (2) becomes $Y = \{y|y \leq \varepsilon\}$ plus a restriction that $y$ must lie in the affine space spanned by $(W^k)^{-1}Av$. We let this subspace be $F^k = \{y|(I-(W^k)^{-1}A(A^T(W^k)^{-2}A)^{-1}A^T(W^k)^{-1})y = 0\}$. Problem 1 becomes

$$\max\{d^Ty(+c^Tz^k)|y \in Y \cap F^k\}. \quad (3)$$

Since the unit ball in completely contained in the feasible region in $Y$, it seems reasonable to choose a search direction by optimizing $d^Ty$ over the intersection of $F^k$ and the unit sphere, $S = \{y|y^Ty \leq 1\}$. The affine scaling search direction is, therefore, chosen by solving

$$\max\{d^Ty|y \in S \cap F^k\}. \quad (4)$$

Since $d^k \in F^k$, the optimal solution of (4) is $y^k = d^k/||d^k||$. The search direction in the original problem variables is $v^k$, where

$$v^k = (A^T(W^k)^{-2}A)^{-1}c/\sqrt{c^T(A^T(W^k)^{-2}A)^{-1}c}.$$  

This is often given as $u^k = -(A^T(W^k)^{-2}A)^{-1}c$, where $x^{k+1} = x^k - \sigma^k u^k$ for some step size $\sigma^k$.

**Step 2. Step size:** We already note that we can choose a feasible step size ($\sigma^k$) as large as

$$\sigma^* = 1/\sqrt{c^T(A^T(W^k)^{-2}A)^{-1}c},$$

as in the convergence proof in Dikin [6] and Vanderbei and Lagarias [13], although larger step sizes appear to work better in practice. Other procedures take a step as some fraction $\gamma$ of the maximum feasible step size $\tilde{\sigma} = \min\{\frac{b-a^Tz^k}{-a^T u^k}|a^T u^k < 0\}$ or in Tsushiya's approach as a fraction $(1/8)$ of $\sigma^*$. If $\sigma^* = \tilde{\sigma}$, then a face has been hit and the algorithm stops. Otherwise, the algorithm proceeds to Step 3.

**Step 3. Update:** Let $z^{k+1} = z^k + v^k$ and $u^{k+1} = b - Ax^{k+1}$. We continue with a full update to $W^{k+1}$, although rank one updating procedures can be used here as in projective scaling methods (see Karmarkar [8] and Shanno [10]).
3. Proof of Convergence

Assuming nondegeneracy of the constraints in (1), the affine scaling algorithm obtains an optimal point as a limiting point of the sequence \( z^k \) (see Vanderbei and Lagarias [13] for the proof for the dual to (2.1)). Tsushiya has proved that this method also converges with degeneracy if a step of \( \sigma^k = \frac{1}{2} \sigma^* \) is used. We will show that convergence is generally obtained with the full \( \sigma^k = \sigma^* \) step.

Our proof applies directly to the problem with inequality constraints. It relies on the basic geometry of the problem and is stated in the following theorem.

**Theorem 1.** Under the above assumptions, the affine scaling algorithm converges to an optimal solution of (1).

Before we proceed directly with the proof of this main theorem, we prove several useful lemmas. The first concerns the termination of the algorithm. In each lemma, we make only assumptions A1-A3.

**Lemma 1.** The affine scaling algorithm either terminates finitely with an optimal solution of (1) or obtains an infinite sequence of iterates.

**Proof:** The algorithm terminates if \( \sigma^* = \bar{\sigma} \) in Step 2. In this case, \( \exists i \) such that \( a_i^T (x^k + v^k) = b_i \) (or \( a_i^T v^k = b_i - a_i^T x^k \)) and \( \forall j \neq i \ a_j^T v^k \leq b_j - a_j^T x^k \). (i.e., a face has been hit). All movement occurs within the ellipsoid:

\[ \{ v \mid g(v) = v^T (A^T (W^k)^{-2} A) v \leq 1 \} \]

The \( v^k \) we choose sets us on the boundary of the ellipsoid. At this point, the gradient of \( g(v) \) is simply a multiple of \( c \). By definition of \( v^k \), \( g(v^k) = \sum^n_{i=1} (\frac{1}{w_i})^2 (a_i^T v^k)^2 = 1 \), and for \( \sigma^* = \bar{\sigma} \), \( (\frac{1}{w_i}) a_i^T v^k = 1 \), which implies \( a_i^T v^k = 0 \) for all \( j \neq i \) or \( a_i^T v^k = 0 \) for all \( j \neq i \).

Hence, for \( \nabla g(v) = 2 A^T (W^k)^{-2} A v \), \( \nabla g(v^k) = \frac{2\epsilon}{\sqrt{\epsilon^2 (A^T (W^k)^{-2} A)^{-1}}} \) and \( \nabla g(v^k) = \frac{2\epsilon}{w_i a_i} \). So, \( c = \frac{2\epsilon}{w_i a_i} \), for an active constraint, \( a_i^T (x^k + v^k) = b_i \). This implies that the K.K.T. conditions are met at \( x^k + v^k \) implying optimality. Thus, the algorithm terminates finitely with \( x^k + v^k \) or an infinite sequence of feasible iterates occurs.
Lemma 2. If the affine scaling algorithm produces an infinite sequence of iterates, then there exists some convergent subsequence of iterates in $x^k$ and $w^k$.

Proof: Since the set of optimal solutions is bounded, there exists no nonzero $v$ such that $c^Tv = 0$ and $Av \leq 0$. Hence, every level set of the algorithm is bounded. Since the algorithm starts with a finite objective value and always improves the objective, the iterates are within a bounded set. ■

Lemma 3. Along any sequence $\{w^k | w^k > 0\}$, the corresponding price vector, $\pi(w^k) = (W^k)^{-2}A(A^T(W^k)^{-2}A)^{-1}c$ is bounded.

Proof: Using Cramer's rule and the Cauchy-Binet Theorem, Vanderbei and Lagarias [13] prove that $f^T(X)^2A(A^T(X)^2A)^{-1}$ is bounded for any positive, diagonal matrix $X$ and fixed $f$ and $A$. Letting $X = (W^k)^{-1}$ and $f_i = 1$ if $\pi(w^k)_i \geq 0$ and $f_i = -1$ if $\pi(w^k)_i < 0$, we obtain that $\pi(w^k)_i$ is bounded for every $i$ and any $w^k > 0$. ■

Lemma 4. If the affine scaling algorithm produces a convergent sequence of iterates in $\{w^k\}$, then the algorithm converges to an optimal solution of (1).

Proof: To show that an optimal solution occurs requires that we show that complementary slackness holds and that the dual variables converge to a nonnegative vector in the limit. To show complementary slackness, observe that $c^Tv^k = -(c^Tw^k)^{1/2}$ where $-c^Tw^k = c^T(A^T(W^k)^{-2}A)^{-1}c = \sum_{i=1}^{n}(c^T(A^T(W^k)^{-2}A)^{-1}a_i(w^k)_i)^2$, so that $c^Tw^k \to 0$ implies that $c^T(A^T(W^k)^{-2}A)^{-1}a_i(w^k)_i^{-1} \to 0$. We write this as $w^k_i \pi_i^k \to 0$ where $\pi_i^k = (1/w^k_i)^2a_i^T(A^T(W^k)^{-2}A)^{-1}c$. Hence, complimentary slackness holds in the limit.

Let $\bar{w} = \lim w^k$. Since $\pi^k$ is a bounded, continuous function of $w^k$, the sequence $\{\pi^k\}$ has some limit point. For any such limit point, $\bar{\pi}$, we must have $\bar{\pi}_j = 0$ for all $j \notin I = \{i|\bar{w}_i = 0\}$. However, we also have $\bar{\pi}^T A = c^T$, so $\bar{\pi}_I^T A_I = c^T$. By the nondegeneracy assumption, the rows of $A_I$ are linearly independent. Thus, $\bar{\pi}_I^T A_I = c^T$ has a unique solution, $\bar{\pi}$, which is then the unique limit of $\{\pi^k\}$. Hence, the sequence $\{\pi^k\}$ converges.
To show that $\pi^k$ is nonnegative in the limit note that:

$$w^{k+1}_i = (b - Ax_{k+1})_i$$

$$= w^k_i + (Ax_k)_i - (Ax_{k+1})_i$$

$$= w^k_i - (A(x_{k+1} - x_k))_i$$

$$= w^k_i - (A \frac{[A^T(W^k)^{-2}A]^{-1}c}{\sqrt{c^T[A^T(W^k)^{-2}A]^{-1}c}})_i$$

$$= w^k_i - \frac{(w^k_i)^2 x^k_i}{\sqrt{c^T[A^T(W^k)^{-2}A]^{-1}c}}.$$

Assume now that $\exists K$ such that $\forall k \geq K$, $\pi^k_i < 0$. This would imply that $w_i^{k+1} \geq w_i^k$ which means that $w^k \pi^k \not\to 0$. This contradicts complimentary slackness. Thus, $\pi^k_i \geq 0$ in the limit, and the algorithm converges to an optimal solution if the iterates converge. ■

**Proof of Theorem 1.** From Lemma 1, we need only consider the case where the algorithm produces an infinite sequence of iterates. In this case, since the objective is bounded above we must have:

$$\sum_{k=1}^{\infty} c^T v^k < \infty. \quad (5)$$

Next, we wish to show that the sequence of iterates $z^k$ also converges. If so, then $w^k = b - Ax^k$ converges and Lemma 4 gives the result. We, therefore, wish to show that

$$\sum_{k=1}^{\infty} \|z^{k+1} - z^k\| = \sum_{k=1}^{\infty} \|v^k\| < +\infty.$$ This result follows if, we have $c^T v^k \geq (1/\alpha) \|v^k\|$ for some $\alpha < \infty$.

To show this result, let $B = [\beta_i^T]$ be composed of a maximal independent set of vectors such that $\beta_i^T c = 0$ (i.e., $B$ is a basis for the orthogonal subspace to the subspace spanned by $c$). Let $a_i = \gamma_i c + B^T \mu^i$ for each $i = 1, \ldots, m$. For $M = [\mu^i]$, we then have $A = \gamma c^T + M^T B$.

Now, we rewrite the optimization problem in (4) by substituting for $v = c\lambda + B^T \nu$. Since $c^T B = 0$, we obtain:

$$\max c^T (c\lambda)$$

subject to

$$(\lambda c^T + (\nu)^T B)(c\gamma^T + B^T M)(W^k)^{-2}(c\gamma^T + B^T B)(c\lambda + B^T \nu) \leq 1. \quad (6)$$
By the construction, the feasible region is equivalent to

$$
(\lambda)^2 c^T c \gamma^T (W^k)^{-2} \gamma c^T c + 2 \lambda c^T c \gamma^T (W^k)^{-2} (M^T B) B^T \nu
$$

$$
+ (\nu)^T B (B^T M) (W^k)^{-2} (M^T B) B^T \nu \leq 1.
$$

(7)

The optimality conditions for $(\lambda^k, \nu^k)$ optimal in (6) state that there exists some $\zeta^k > 0$ such that

$$
c^T c - 2 \zeta^k (\lambda^k c^T c \gamma^T (W^k)^{-2} \gamma c^T c + c^T c \gamma^T (W^k)^{-2} (M^T B) B^T \nu^k) = 0
$$

$$
\lambda^k c^T c \gamma^T (W^k)^{-2} (M^T B) B^T + (\nu^k)^T B (B^T M) (W^k)^{-2} (M^T B) B^T = 0.
$$

(8)

Since $A$ has full rank by assumption, $B (B^T M) (W^k)^{-2} (M^T B) B^T$ is full rank $(n - 1)$ and invertible. The second set of conditions in (8) therefore implies that $(\nu^k)^T = \lambda^k c^T c \gamma^T (W^k)^{-2} (M^T B) B^T (B^T M) (W^k)^{-2} (M^T B) B^T$.

Letting $\tilde{A} = M^T B B^T$, we have $(\nu^k)^T = -(\lambda^k c^T c \gamma^k) (W^k)^{-2} \tilde{A} (\tilde{A}^T (W^k)^{-2} \tilde{A})^{-1}$. Again applying the result from Vanderbei and Lagarias as in Lemma 3, we have that $|\nu_i^k| \leq \eta \left( \frac{\|c\|}{\|\beta_i\|} \right) |\lambda^k|$ for every $i = 1, \ldots, n - 1$, and some $\eta < \infty$ that does not depend on $W^k$. Hence,

$$
\|\nu^k\| = \|c \lambda^k + B^T \nu^k\|
$$

$$
\leq \|c\| |\lambda^k| + \sum_{i=1}^{n-1} (|\beta_i|) |\nu_i^k|
$$

$$
\leq \|c\| |\lambda^k| (1 + (n - 1) \eta) = c^T \nu^k(\alpha),
$$

(9)

for $\alpha = \frac{1 + (n - 1) \eta}{\|c\|} < \infty$ since $c^T \nu^k = c^T (\lambda^k c + B^T \nu^k) = \lambda^k \|c\|^2$. This gives the desired result and completes the proof. □
References


