4 The Affine Scaling Algorithm

In this Chapter, we present one of the most efficient yet simple interior point algorithm for solving the linear programming

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

and its dual

$$\begin{aligned} \max \quad & b^T p \\ \text{s.t.} \quad & A^T p \leq c\,, \end{aligned}$$

where A is an $m \times n$ matrix.

Let $P = \{x \mid Ax = b, x \geq 0\}$ be the **feasible set**. We will call

$$P^o = \{x \in P \mid x > 0\}$$

be the **interior** of P and its elements **interior points**. The main geometric idea of the affine scaling algorithm is instead of minimizing $c^T x$ directly over P (may be difficult), we solve a series of optimization problems over ellipsoids (the solution can be obtained in the closed form).

Lemma 4.1 *Let $\beta \in (0,1)$ be a scalar, let $y \in \Re^n$ satisfy $y > 0$, and let*

$$S = \left\{ x \in \Re^n \;\Big|\; \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{y_i^2} \le \beta^2 \right\}.$$

Then, $x > 0$ for all $x \in S$.

Proof. Let $x \in S$. Then, for each i, we have

$$(x_i - y_i)^2 \le \beta^2 y_i^2 < y_i^2 .$$

Therefore,

$$|\, x_i - y_i \,| < y_i .$$

In particular,

$$-x_i + y_i < y_i$$

and it follows that $x_i > 0$. Q.E.D

Fix some $y \in \Re^n$ satisfying $y > 0$ and $Ay = b$. Let

$$Y = \mathrm{diag}(y_1, \ldots, y_n)$$

denote the $n \times n$ diagonal matrix whose entries are $y_1, \ldots, y_n$. Then $x \in S$ can be written as

$$\|\, Y^{-1}(x - y) \,\| \le \beta ,$$

where $\| \cdot \|$ stands for the Euclidean norm (2-norm). The set S is an ellipsoid centered at y. The set

$$S_0 = S \cap \{ x \,|\, Ax = b \}$$

is a section of the ellipsoid S, and is itself an ellipsoid contained in the feasible set.

Next, we replace the original linear programming problem with the problem of minimizing over the ellipsoid S_0:

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & \| Y^{-1}(x - y) \| \le \beta . \end{aligned} \tag{4.1}$$

We introduce a new variable $d = x - y$. Since y is feasible, we have $Ay = b$. Furthermore, for every $x \in S_0$, we have $Ax = b$. It follows that $Ad = 0$. If we optimize over d, instead of x, problem (4.1) becomes

$$\begin{aligned} \min \quad & c^T d \\ \text{s.t.} \quad & Ad = 0 \\ & \| Y^{-1} d \| \le \beta . \end{aligned} \tag{4.2}$$

Lemma 4.2 *Assume that the rows of A are linearly independent and that c is not a linear combination of the rows of A. Let y be a positive vector in P. Then an optimal solution d^* to problem (4.2) is given by*

$$d^* = -\beta \frac{Y^2(c - A^T p)}{\| Y(c - A^T p) \|},$$

where

$$p = (AY^2A^T)^{-1} AY^2 c .$$

Furthermore, the vector $x = y + d^*$ belongs to P and

$$c^T x = c^T y - \beta \| Y(c - A^T p) \| < c^T y .$$

Proof. We first argue that the matrix AY^2A^T is invertible so that p is well defined. If it is not invertible, there exists some $z \neq 0$ such that

$$z^T AY^2 A^T z = 0 .$$

Let $w = YA^T z$. We then have $w^T w = 0$ which implies $w = 0$. Since Y has positive diagonal entries, it follows that $A^T z = 0$, which means that the rows of A are linearly dependent, contradicting our assumption. Also, since c is not a linear combination of the row of A, it follows that $c - A^T p \neq 0$ and d^* is well defined.

We now show that d^* is a feasible solution to problem (4.2). We have

$$Y^{-1} d^* = -\beta \frac{Y(c - A^T p)}{\| Y(c - A^T p) \|},$$

which shows that $\|Y^{-1} d^*\| = \beta$. In order to show that $Ad^* = 0$, it suffices to show that

$$AY^2(c - A^T p) = \left(- \frac{\| Y(c - A^T p) \|}{\beta} \right) Ad^* = 0 ,$$

and this follows from the definition of p because

$$\begin{aligned} AY^2(c - A^Tp) &= AY^2c - AY^2A^T(AY^2A^T)^{-1}AY^2c \\ &= AY^2c - AY^2c = 0\,. \end{aligned}$$

We next show the optimality of d^*. For any feasible solution d to problem (4.2), we have $Ad = 0$ and $\|Y^{-1}d\| \leq \beta$. Using these facts and the Schwartz inequality, we obtain

$$\begin{aligned} c^Td &= (c^T - p^TA)d \\ &= (c^T - p^TA)YY^{-1}d \\ &\geq -\|\,Y(c - A^Tp)\,\|\,\|\,Y^{-1}d\,\| \\ &\geq -\beta\|\,Y(c - A^Tp)\,\|\,. \end{aligned}$$

On the other hand,

$$\begin{aligned} c^Td^* &= (c^T - p^TA)d^* \\ &= -(c^T - p^TA)\beta\frac{Y^2(c - A^Tp)}{\|\,Y(c - A^Tp)\,\|} \\ &= -\beta\frac{(Y\,(c - A^Tp\,))^T\,(Y\,(\,c - A^Tp)\,)}{\|\,Y\,(\,c - A^Tp)\,\|} \\ &= -\beta\|\,Y(c - A^Tp)\,\|\,. \end{aligned}$$

It follows that d^* is optimal. In addition,

$$c^Tx = c^Ty + c^Td^* = c^Ty - \beta\,\|Y(c - A^Tp\,)\|\,,$$

which is strictly smaller than c^Ty because $c - A^Tp \neq 0$. Finally, $x \in P$, because the feasible set S_0 of the problem (4.1) is contained in P. Q.E.D

Exercise: Use the KKT theory in Nonlinear Programming to prove Lemma 4.2 [Note that $\|\,Y^{-1}d\,\| \leq \beta$ if and only if $\|\,Y^{-1}d\,\|^2 \leq \beta^2$.]

Note that if $d^* \geq 0$, the feasible set of the original linear programming problem is unbounded, since $x + \alpha d^* \geq 0$ for all $\alpha \geq 0$, and $Ad^* = 0$. Given that $c^Td^* < 0$ (Lemma 4.2), it follows that the optimal cost in the original problem is $-\infty$.

We next provide an interpretation of the formula for p. Let y be a nondegenerate basic feasible solution and apply the same formula in Lemma 4.2 to define a vector p. Let B be the corresponding basis matrix. We assume without loss of generality that the first m variables are basic, so that $A = [B\ N]$ for some matrix N of dimension $m \times (n-m)$. If

$$Y = \operatorname{diag}(y_1, \ldots, y_m, 0, \ldots, 0)$$

and

$$Y_0 = \operatorname{diag}(y_1, \ldots, y_m)\,,$$

then $AY = [BY_0\ 0]$ and

$$\begin{aligned} p &= (AY^2A^T)^{-1}AY^2c \\ &= (B^T)^{-1}Y_0^{-2}B^{-1}BY_0^2c_{\mathcal{B}} \\ &= (B^T)^{-1}c_{\mathcal{B}}\,, \end{aligned}$$

which is the corresponding dual basic solution. Hence, we call the vectors p, corresponding to the primal solution y, the **dual estimates** even if y is not a basic feasible solution. Moreover, the vector $r = c - A^Tp$ becomes

$$r = c - A^T(B^T)^{-1}c_{\mathcal{B}}\,,$$

i.e.,

$$r^T = c^T - c_{\mathcal{B}}^TB^{-1}A\,.$$

This implies that r is the reduced cost vector in the simplex method. Note that if y is degenerate, then the matrix AY^2A^T is not invertible, and this interpretation breaks down. We will assume that all primal basic feasible solutions are nondegenerate.

Suppose that $r = c - A^Tp$ is nonnegative. In this case, the vector p is a dual feasible solution. Note, in addition, that

$$r^Ty = (c - A^Tp)^Ty = c^Ty - p^TAy = c^Ty - p^Tb\,,$$

i.e., the difference in objective values between the primal solution y and the dual solution p is simply r^Ty. We call this difference the **duality gap**. By weak duality, the duality gap r^Ty is always nonnegative. If $r^Ty = 0$, then the complementary slackness conditions hold and the vectors y and p are primal and dual optimal solutions, respectively.

The next lemma shows that if the duality gap satisfies $r^Ty < \varepsilon$, where $\varepsilon > 0$ is small, then the primal and dual solutions are near-optimal.

Lemma 4.3 *Let y and p be a primal and a dual feasible solution, respectively, such that*

$$c^Ty - b^Tp < \varepsilon\,.$$

Let y^* and p^* be optimal primal and dual solutions, respectively. Then

$$c^Ty^* \le c^Ty < c^Ty^* + \varepsilon\,,$$

$$b^Tp^* - \varepsilon < b^Tp \le b^Tp^*\,.$$

Proof. Since y is a primal feasible solution and y^* is an optimal primal solution, $c^T y^* \leq c^T y$. By weak duality, $b^T p \leq c^T y^*$. Since $c^T y - b^T p < \varepsilon$, we have

$$c^T y \,<\, b^T p \,+\, \varepsilon \,\leq\, c^T y^* \,+\, \varepsilon\,.$$

Similarly, we obtain

$$b^T p^* \,=\, c^T y^* \,\leq\, c^T y \,<\, b^T p \,+\, \varepsilon\,.$$

Q.E.D.

The previous lemma suggests a termination criterion. The algorithm terminates when $r \geq 0$ (dual feasibility) and the duality gap

$$r^T y \,=\, y^T r \,=\, e^T Y r$$

is small, where $e = (1, 1, \ldots, 1)^T$. Then the primal and dual solutions y and p are "near-optimal" in the sense that their cost is within ε from optimality (ε-optimal).

The affine scaling algorithm uses the following inputs:

(a) the data of the problem (A, b, c);

(b) an initial primal feasible solution $x^0 > 0$;

(c) the optimality tolerance $\varepsilon > 0$;

(d) the parameter $\beta \in (0, 1)$.

The Affine Scaling Algorithm

1. (Initialization) Start with some feasible $x^0 > 0$; let $k = 0$.

2. (Computation of dual estimates and reduced costs) Given some feasible $x^k > 0$, let

$$X_k \,=\, \mathrm{diag}(x_1^k, \ldots, x_n^k)\,,$$
$$p^k \,=\, (\, A X_k^2 A^T\,)^{-1} A X_k^2 c\,,$$
$$r^k \,=\, c \,-\, A^T p^k\,.$$

3. (Optimality check) Let $e = (1, 1, \ldots, 1)^T$. If $r^k \geq 0$ and $e^T X_k r^k < \varepsilon$, then stop; the current solution x^k is primal ε-optimal and p^k is dual ε-optimal.

4. (Unboundedness check) If $-X_k^2 r^k \geq 0$ then stop; the optimal cost is $-\infty$.

5. (Update of the primal solution) Let

$$x^{k+1} \,=\, x^k \,-\, \beta \frac{X_k^2 r^k}{\|\, X_k r^k \,\|}\,. \tag{4.3}$$

Replace k by $k + 1$ and go to Step **2.**

Example: Consider the following problem

$$\begin{array}{ll} \max & x_1 + 2x_2 \\ \text{s.t.} & x_1 + x_2 \le 2 \\ & -x_1 + x_2 \le 1 \\ & x_1, x_2 \ge 0\,. \end{array}$$

By introducing slack variables x_3 and x_4 we transform the problem to

$$\begin{array}{ll} \min & -x_1 - 2x_2 \\ \text{s.t.} & x_1 + x_2 + x_3 = 2 \\ & -x_1 + x_2 + x_4 = 1 \\ & x_1, x_2, x_3, x_4 \ge 0\,. \end{array}$$

Starting with the vector

$$x^0 = (0.1, 0.1, 1.8, 1)^T$$

and $\beta = 0.995$, we can calculate

$$p^0 = (-0.0281, 0.0191)^T\,,$$

$$r^0 = (-0.9528, -1.9909, 0.0091, -0.0191)^T$$

and

$$x^1 = (0.1427, 0.1892, 1.6681, 1.0854)^T\,.$$

The above process can be repeated until a satisfied approximate solution is found.

There are other variants of the affine scaling algorithm, which differ in the choice of stepsize. Given a vector u, we introduce the notation

$$\|\, u \,\|_\infty = \max_i |\, u_i \,|\,,$$

$$\gamma(u) = \max\{u_i \,|\, u_i > 0\,\}\,.$$

It is easy to check that

$$\gamma(u) \le \|\, u \,\|_\infty \le \|\, u \,\|\,.$$

The version of affine scaling we presented is called the **short-step** method. In contrast, in **long-step** variants, we update in the same direction $-X_k^2 r_k$ but take a larger step. More specifically, equation (4.3) is replaced by

$$x^{k+1} = x^k - \beta \frac{X_k^2 r^k}{\|\, X_k r^k \,\|_\infty} \tag{4.4}$$

or

$$x^{k+1} = x^k - \beta \frac{X_k^2 r^k}{\gamma(\, X_k r^k\,)}\,. \tag{4.5}$$

Note that the new vector x^{k+1}, as determined by long-step versions, is also feasible and positive. This is because equation (4.4) yields

$$\| X_k^{-1} (x^{k+1} - x^k) \|_\infty = \beta \frac{\| X_k r^k \|_\infty}{\| X_k r^k \|_\infty} = \beta < 1 .$$

In particular, we have for all i, $|x_i^{k+1} - x_i^k| / x_i^k \leq \beta < 1$, which implies that $x_i^{k+1} > 0$. [The argument under the stepsize in equation (4.5) is similar]. In addition, since long-step methods make a larger step along a direction of cost decrease the resulting reduction in the objective function is the largest for the stepsize in equation (4.5) and smallest for the step-size in equation (4.3). As a result, the stepsize in equation (4.5) is more popular in practice.

Assumption 4.1 *(a) The rows of matrix A are linearly independent.*

(b) The vector c is not a linear combination of the rows of A.

(c) There exists an optimal solution.

(d) There exists a positive feasible solution.

Note that if Assumption 4.1 (b) fails to hold, then there exists some vector p such that $c^T = p^T A$. In that case, the cost of every feasible vector x is the same, namely $c^T x = p^T A x = p^T b$, and every feasible solution is optimal.

Assumption 4.2 *(a) Every basic feasible solution to the primal problem is nondegenerate.*

(b) At every basic feasible solution to the primal problem, the reduced cost of every nonbasic variable is nonzero.

Assumption 4.2 (b) implies that for every basic feasible solution to the primal, the corresponding dual basic nondegenerate. Moreover, Assumptions 4.2 (a)-(b) and 4.1 (c) imply that the primal and the dual problems have a unique optimal solution [prove it as an exercise].

Next, we present the principal convergence results for the long-step affine scaling variants. The results for the short step algorithms are similar and will be discussed shortly.

Theorem 4.1 *If we apply the long-step affine scaling algorithm with $\varepsilon = 0$, the following hold:*

(a) *For the stepsize in equation (4.4) or (4.5), under Assumptions 4.1 or 4.2, and if* $0 < \beta < 1$*, the sequences* $\{x^k\}$ *and* $\{p^k\}$ *converge to the optimal primal and dual solutions, respectively.*

(b) *For the stepsize in equation(4.5), under Assumption 4.1 , and if* $0 < \beta < 2/3$*, the sequences* $\{x^k\}$ *and* $\{p^k\}$ *converge to the optimal primal and dual solutions, respectively.*

Initialization: In order to start the affine scaling algorithm, we need an interior feasible solution. Such a feasible solution can be constructed as follows. Let $e \in \Re^n$ be the vector with all components equal to 1. We introduce a new variable x_{n+1}. We create a new column $A_{n+1} = b - Ae$ and consider the problem

$$\begin{array}{ll} \min & c^T x + \ M x_{n+1} \\ \text{s.t.} & Ax + (b - Ae)x_{n+1} = b \\ & (x,\, x_{n+1}) \geq 0\,, \end{array}$$

where M is a large positive scalar. Notice that $(x,\, x_{n+1}) = (e, 1)$ is a positive feasible solution to the augmented problem and the affine scaling algorithm can be applied. If M is very large, and as long as the original problem has an optimal solution, it can be shown that an optimal solution to the augmented problem will have $x_{n+1} = 0$, and will therefore provide an optimal solution to the original problem.

Next, we prove the convergence of the short-step affine scaling algorithm, in the absence of degeneracy, for every $\beta \in (0, 1)$. For the long-step version with the step-size in equation (4.4) the proof is similar.

Lemma 4.4 *Consider the standard form polyhedron* $P = \{x \in \Re^n \,|\, Ax = b, x \geq 0\}$*. Suppose that* A *has* m *rows and that they are linearly independent. Assume also that every basic feasible solution is nondegenerate. Let* $\bar{x}$ *be an element of* P *and let* $\mathcal{B} = \{i \,|\, \bar{x}_i > 0\}$*. Then:*

(a) *There exist indices* $\mathcal{B}(1), \ldots, \mathcal{B}(m) \in \mathcal{B}$ *such that the vectors* $A_{\mathcal{B}(1)}, \ldots, A_{\mathcal{B}(m)}$ *are linearly independent and the corresponding basic solution is feasible. In particular,* $|\mathcal{B}| \geq m$.

(b) *if* $|\mathcal{B}| = m$*, then* $\bar{x}$ *is a basic feasible solution.*

Proof. (a) Consider some $\bar{x} \in P$ and let $\mathcal{B} = \{i \,|\, \bar{x}_i > 0\}$. If the columns $A_i, i \in \mathcal{B}$, are linearly dependent, there exist coefficients $\lambda_i, i \in \mathcal{B}$, not all of them zero, such that

$$\sum_{i \in \mathcal{B}} \lambda_i A_i = 0$$

Note that we have

$$\sum_{i \in \mathcal{B}} A_i(\bar{x}_i + \theta\lambda_i) = b, \ \forall\, \theta .$$

Let us choose θ such that $\bar{x}_i + \theta\lambda_i \geq 0$ for all $i \in \mathcal{B}$, and $\bar{x}_k + \theta\lambda_k = 0$ for some $k \in \mathcal{B}$. We then obtain an element $\hat{x}$ of P such that $\{i \,|\, \hat{x}_i > 0\}$ is a proper subset of $\mathcal{B}$. By repeating this procedure as many times as needed, we arrive a some $\tilde{x} \in P$ such that $I = \{i \,|\, \tilde{x}_i > 0\}$ is a proper subset of $\mathcal{B}$, and the columns $A_i, i \in I$, are linearly independent. If $|I| < m$, we can augment the columns $A_i, i \in I$ to obtain m linearly independent columns (a basis). The corresponding basic feasible solution is $\tilde{x}$, which is degenerate, and we have arrived at a contradiction. Hence $|I| = m$. This implies that $|\mathcal{B}| \geq m$. Furthermore, since $I \subset \mathcal{B}$, the columns $A_i, i \in I$, are m linearly independent columns associated with positive components of $\bar{x}$.

(b) Using the result of part (a), the m vectors $A_i, i \in \mathcal{B}$, are linearly independent and, therefore, $\bar{x}$ is a basic feasible solution. Q.E.D.

Theorem 4.2 *Consider the short-step affine scaling algorithm, with* $\varepsilon = 0$. *If Assumptions 4.1 and 4.2 hold, and if* $0 < \beta < 1$, *then* $\{x^k\}$ *and* $\{p^k\}$ *converge to the optimal solutions of the primal and dual, respectively.*

Proof. Let x^* be an optimal basic feasible solution to the primal problem. By Assumption 4.2, the reduced costs of all nonbasic variables are positive. This implies that x^* is the unique solution to the primal problem.

Using the optimality of x^*, and Lemma 4.2 with $y = x^k$, $x = x^{k+1}$, we have

$$c^T x^* \leq c^T x^{k+1} = c^T x^k - \beta \| X_k(c - A^T p^k) \| < c^T x^k .$$

Thus, the sequence $\{c^T x^k\}$ is monotonically decreasing and bounded below, which implies that it converges. In particular, $c^T x^{k+1} - c^T x^k$ converges to zero. Using the definition of X_k, we obtain

$$\lim_{k \to \infty} x_i^k (c_i - A^T p^k) = 0, \quad i = 1, \ldots, n, \tag{4.6}$$

which shows that complementary slackness holds in the limit.

We will now show that the sequence $\{x^k\}$ is bounded. Consider the problem

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} x_i \\ \text{s.t.} \quad & Ax = b \\ & c^T x \leq c^T x^* \\ & x \geq 0 \,. \end{aligned}$$

The above problem has finite optimal cost, because x^* is the only feasible solution. Therefore, using the Minkowski Theorem, we know that for every extreme direction d^j of the polyhedral set P, $c^T d^j > 0$. Thus, the following problem has finite optimal cost

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} x_i \\ \text{s.t.} \quad & Ax = b \\ & c^T x \leq c^T x^0 \\ & x \geq 0 \,. \end{aligned}$$

It follows that the set

$$\{x \,|\, Ax = b, \, x \geq 0, \, c^T x \leq c^T x^0\}$$

is bounded. Since $c^T x^k \leq c^T x^0$, the sequence $\{x^k\}$ is bounded.

Recall that equation (4.6) establishes complementary slackness, in the limit. Given Assumptions 4.1 and 4.2, it can be shown that if a pair (x, p), with x primal feasible, satisfies complementary slackness, then x is a basic feasible solution [prove it as an exercise]. Using similar arguments, we will show next that complementary slackness, in the limit, implies that $\{x^k\}$ approaches the set of basic feasible solutions.

Since $\{x^k\}$ is bounded, there exists a sequence $\{k_j\}$ of positive integers such that $\{x^{k_j}\}$ converges to some $\bar{x}$. It is easily seen that $\bar{x}$ is a feasible solution. Let

$$\mathcal{B} = \{i \,|\, \bar{x}_i > 0\}.$$

Using Lemma 4.4 (a), the set $\mathcal{B}$ has at least m elements and there exist

$$\mathcal{B}(1), \ldots, \mathcal{B}(m) \in \mathcal{B}$$

such that the columns

$$A_{\mathcal{B}(1)}, \ldots, A_{\mathcal{B}(m)}$$

are linearly independent, and such that the corresponding basic solution $\hat{x}$ is feasible. Let B be the corresponding basis matrix and let

$$c_{\mathcal{B}} = (c_{\mathcal{B}(1)}, \ldots, c_{\mathcal{B}(m)})^T \,.$$

Using equation (4.6),

$$x_i^{k_j}(c_i - A_i^T p^{k_j}) \rightarrow 0 \text{ as } j \rightarrow \infty$$

for all i. For $i = 1, \ldots, m$, we have

$$\mathcal{B}(1), \ldots, \mathcal{B}(m) \in \mathcal{B}$$

and $x^{k_j}_{\mathcal{B}(i)}$ converges to $\bar{x}_{\mathcal{B}(i)}$, which is positive because $\mathcal{B}(i) \in \mathcal{B}$. It follows that $c_i - A^T_{\mathcal{B}(i)} p^{k_j}$ converges to zero. In matrix notation, we have that $c_{\mathcal{B}} - B^T p^{k_j}$ converges to zero. Hence p^{k_j} converges to $(B^T)^{-1} c_{\mathcal{B}}$. Consequently, for every i, $c_i - A_i^T p^{k_j}$ converges to $c_i - A_i^T (B^T)^{-1} c_{\mathcal{B}}$, which is the reduced cost of the ith variable, at the basic feasible solution $\hat{x}$. By Assumption 4.2 (b), we have

$$c_i - A_i^T (B^T)^{-1} c_{\mathcal{B}} \neq 0 \text{ for } i \neq \mathcal{B}(1), \ldots, \mathcal{B}(m).$$

Thus, for $i \neq \mathcal{B}(1), \ldots, \mathcal{B}(m)$, $c_i - A_i^T p^{k_j}$ converges to a nonzero value. We use equation (4.6) once more to conclude that

$$\bar{x}_i = 0 \text{ for } i \neq \mathcal{B}(1), \ldots, \mathcal{B}(m).$$

In particular,

$$\mathcal{B} = \{ \mathcal{B}(1), \ldots, \mathcal{B}(m) \}$$

and $\bar{x}$ is the basic feasible solution associated with the basis matrix B.

We have proved so far that the sequence $\{x^k\}$ has at least one limit point and that every limit point must be a basic feasible solution. We now argue that there can only be a single limit point. Let $\delta > 0$ be such that every basic variable at every basic feasible solution is larger than δ. Because of the nondegeneracy assumption, basic variables are always positive and, therefore, such a δ exists. Let $\varepsilon = \delta/3$. Since every possible limit point of the sequence $\{x^k\}$ is a basic feasible solution, it follows that there exists some K such that for all k larger than K, x^k is within ε of some basic feasible solution.

Suppose that there exists some $k \geq K$ and two different basic feasible solutions $\bar{x}$ and $\hat{x}$ such that

$$\| x^k - \bar{x} \| \leq \varepsilon, \quad \| x^{k+1} - \hat{x} \| \leq \varepsilon.$$

We assume that the ith variable is nonbasic at $\bar{x}$ (in particular, $\bar{x}_i = 0$) and basic at $\hat{x}$ (in particular, $\hat{x}_i \geq \delta$). It then follows that

$$x_i^k \leq \varepsilon$$

and

$$x_i^{k+1} \geq \delta - \varepsilon = 2\varepsilon.$$

On the other hand, equation (4.3) yields

$$\begin{aligned} x_i^{k+1} &= x_i^k \left(1 - \beta \frac{x_i^k r_i^k}{\| X_k r^k \|} \right) \\ &\leq x_i^k (1 + \beta) \\ &< 2 x_i^k \\ &\leq 2\varepsilon \\ &\leq x_i^{k+1}. \end{aligned}$$

This is a contradiction. It establishes that if $k \geq K$ and if x^k is within ε of $\bar{x}$, then the same must be true for x^{k+1}. We conclude that no basic feasible solution other than $\bar{x}$ can be a limit point and therefore the sequence $\{x^k\}$ converges to $\bar{x}$. As shown earlier, $\{p^k\}$ also converges to the associated dual basic solution.

It remains to show that the limit $\bar{x}$ must be optimal. If $\bar{x}$ is not optimal, then there exists some nonbasic variable $\bar{x}_i$ whose reduced cost is negative. As shown earlier,

$$r_i^k = c_i - A_i^T p^k$$

converges to the reduced cost of $\bar{x}_i$ and, therefore, r_i^k eventually becomes and stays negative. It follows from equation (4.3) that eventually x_i^k becomes a strictly increasing sequence. On the other hand, x_i^k converges to $\bar{x}_i$, which is zero because $\bar{x}_i$ is a nonbasic variable, and we obtain a contradiction. Q.E.D.

Notes: The affine scaling algorithm was introduced by Dikin in 1967, but remained mostly unnoticed. After the appearance of Karmarkar's seminal work (1984), which gave the first interior point algorithm with polynomial time complexity, the affine scaling algorithm was rediscovered by Barnes (1986), and Vanderbei, Meketon, and Freeman (1986). Dikin (1974) analyzed the convergence of short-step affine scaling algorithm under a primal non-degeneracy assumption. Convergence result for different variants of affine scaling, without any nondegeneracy assumptions, have been developed by Tsuchiya (1991), Tseng and Luo (1992), and Tsuchiya and Maramatsu (1995), and etc. The proof of convergence of the affine scaling algorithm given here is from E. R. Barnes, "A variation on Karmarkar's algorithm for solving linear programming problems", *Mathematical Programming*, 36 (1986), 174-182.