MATHEMATICS OF OPERATIONS RESEARCH
Vol. 14, No. 1, February 1989
Printed in U.S.A.

BOUNDARY BEHAVIOR OF INTERIOR POINT ALGORITHMS IN LINEAR PROGRAMMING*†

NIMROD MEGIDDO‡ AND MICHAEL SHUB§

This paper studies the boundary behavior of some interior point algorithms for linear programming. The algorithms considered are Karmarkar's projective rescaling algorithm, the linear rescaling algorithm which was proposed as a variation on Karmarkar's algorithm, and the logarithmic barrier technique. The study includes both the continuous trajectories of the vector fields induced by these algorithms and also the discrete orbits. It is shown that, although the algorithms are defined on the interior of the feasible polyhedron, they actually determine differentiable vector fields on the closed polyhedron. Conditions are given under which a vector field gives rise to trajectories that each visit the neighborhoods of all the vertices of the Klee-Minty cube. The linear rescaling algorithm satisfies these conditions. Thus, limits of such trajectories, obtained when a starting point is pushed to the boundary, may have an exponential number of breakpoints. It is shown that limits of projective rescaling trajectories may have a linear number of such breakpoints. However, projective rescaling trajectories may visit the neighborhoods of linearly many vertices. The behavior of the linear rescaling algorithm near vertices is analyzed. It is proved that all the trajectories have a unique asymptotic direction of convergence to the optimum.

1. Introduction. Interest in interior point algorithms for linear programming was revived by the work of Karmarkar [Kar]. In this paper we sometimes refer to Karmarkar's algorithm as the *projective rescaling algorithm*. This reflects the property that the algorithm moves in the direction of the gradient of the objective function after a projective scaling transformation has been applied. A variation on this algorithm, which was proposed in various forms by many people[1] (e.g., [Bar, CaS, VMF]), is called the *linear rescaling algorithm*, reflecting the property that a linear scaling transformation is applied before the gradient step is taken. The projective and the linear rescaling algorithms were shown in [GMSTW] to be related to the logarithmic barrier function technique using Newton's method. In this paper we study the behavior of all these algorithms. We consider both continuous and discrete versions of the algorithms. Our main interest is in the boundary behavior of these algorithms. We study the differences among the different algorithms through their behavior near boundaries. We first introduce the algorithms and the notation to be used later.

Interior point algorithms for linear programming usually update a point x, interior to the feasible polyhedron P, by moving along a straight line in the direction of a vector $V(x)$ defined at x. The new point depends of course not only on the direction of $V(x)$ but also on the step size which is assigned at x. Thus, the new point can be

*Received September 29, 1986; revised September 28, 1987.
AMS 1980 subject classification. Primary: 90C05.
IAOR 1973 subject classification. Main: Programming: Linear.
OR/MS Index 1978 subject classification. Primary: 643 Programming/Linear/Algorithms.
Key words. Interior point algorithms, boundary behavior.
†This work was done in part while the authors were members at the Mathematical Sciences Research Institute, Berkeley, California. Partially supported by NSF Grants.
‡IBM Almaden Research Center and Tel Aviv University.
§IBM T. J. Watson Research Center.
[1] *Postscript.* It recently came to our attention that this algorithm was proposed by I. I. Dikin in 1967 (see *Math. Programming* **41** (1988), p. 393).

0364-765X/89/1401/0097$01.25

represented in the form $x' = x + \alpha(x)V(x)$, where $\alpha(x)$ denotes a real number that determines the step size. The iteration formula defines a transformation of the polyhedron P into itself. We are concerned with the properties of this transformation, or the vector field itself, near the boundary of P. We denote the boundary of P by ∂P. In this paper we usually consider the linear programming problem in standard form:

$$(SF) \qquad \begin{aligned} &\text{Minimize} && c^T x \\ &\text{subject to} && Ax = b, \\ & && x \geqslant 0 \end{aligned}$$

where $A \in R^{m \times n}$ $(m \leqslant n)$, $b \in R^m$ and $c, x \in R^n$.

1.1. *The linear rescaling algorithm.* Following the description of [VMF], the algorithm is stated with respect to the linear programming problem in standard form. It is assumed that a point x^0 is known such that $Ax^0 = b$ and $x^0 > 0$. Given a point $x \in R^n$, we denote by $D = D(x)$ a diagonal matrix of order n whose diagonal entries are the components of x. We frequently denote $D = D_x$ to emphasize the dependence on x. Let $x \in R^n$ be any point such that $Ax = b$ and $x > 0$. The algorithm assigns to the point x a "search direction", that is, a vector ξ (whose norm is not necessarily equal to 1) which is computed as follows. Consider a transformation of space T_x: $R^n \to R^n$ given by $T_x(y) = D^{-1}y$. In the transformed space, the direction $\eta = T_x(\xi)$ is obtained by projecting the vector Dc orthogonally into the linear subspace $\{\eta\colon AD\eta = 0\}$. Thus, η is the solution of the following least-squares problem:

$$\begin{aligned} &\text{Minimize} && \|Dc - \eta\|^2 \\ &\text{subject to} && AD\eta = 0. \end{aligned}$$

Assuming A is of full rank, the solution is

$$\eta_l = \left[I - DA^T(AD^2A^T)^{-1}AD\right] Dc.$$

In the original space, the linear rescaling algorithm assigns to a point x the vector

$$\xi_l = \xi_l(x) = D\left[I - DA^T(AD^2A^T)^{-1}AD\right] Dc,$$

to define a search direction. We note that since the problem is in the minimization form, the new point has the form $x - \alpha(x)\xi_l(x)$ where $\alpha(x)$ is positive.

1.2. *The projective rescaling algorithm.* Following [Kar], the algorithm is stated with respect to the linear programming problem given in the following form ("Karmarkar's standard form"):

$$(KSF) \qquad \begin{aligned} &\text{Minimize} && c^T x \\ &\text{subject to} && Ax = 0, \\ & && e^T x = 1, \\ & && x \geqslant 0, \end{aligned}$$

where $A \in R^{(m-1) \times n}$ $(1 \leqslant m \leqslant n)$, $x, c \in R^n$ and $e = (1, \dots, 1)^T \in R^n$. In the original statement of the algorithm it was assumed that $Ae = 0$ so the point $x^0 = e/n$ is interior relative to the linear subspace $\{Ax = 0\}$. We do not use this assumption in our analysis. It was assumed that the optimal value of the objective function is zero, but the algorithm is well defined without this assumption. Let $\tilde{A}$ denote the matrix

$$\tilde{A} = \begin{pmatrix} A \\ e^T \end{pmatrix}.$$

Let $x \in R^n$ be such that $Ax = 0$, $e^T x = 1$, and $x > 0$ and continue to denote $D = D(x) = \mathrm{Diag}(x)$. The new point is computed as a function of x as follows. Consider a transformation of space

$$T_x\colon R^n \to R^n$$

given by

$$T_x(y) = \frac{1}{e^T D^{-1} y} D^{-1} y.$$

Thus, $T_x(x) = e/n$. In the transformed space, the direction η_p is obtained by projecting the vector Dc into the nullspace of the matrix

$$\bar{A} = \begin{pmatrix} AD \\ e^T \end{pmatrix}.$$

Thus,

$$\eta_p = \left[I - \bar{A}^T \left(\bar{A}\bar{A}^T\right)^{-1} \bar{A}\right] Dc.$$

The nullspace of the matrix $\bar{A}$ equals the intersection of the nullspaces of the matrices AD and e^T. However, e is orthogonal to every row of AD since $ADe = Ax = 0$. This property implies that η_p can be obtained by projecting on the nullspace of AD and then projecting the projection on the null space of e^T (see Appendix C). It follows that the search direction in the transformed space is given by

$$\eta_p = \left[I - \bar{A}^T \left(\bar{A}\bar{A}^T\right)^{-1} \bar{A}\right] Dc = \left[I - \frac{1}{n} ee^T\right]\left[I - DA^T(AD^2A^T)^{-1} AD\right] Dc.$$

The search direction ξ_p in the original space is obtained as follows. The algorithm moves in the transformed space from the point e/n to a point of the form

$$q = \frac{1}{n} e - \frac{\rho}{\|\eta_p\|} \eta_p,$$

where ρ is a certain positive constant. The step in the original space is thus given by the vector $u = T_x^{-1}(q) - x$. The inverse transformation is given by

$$T_x^{-1}(q) = \frac{1}{e^T Dq} Dq.$$

Letting

$$v = \frac{\rho}{\|\eta_p\|} \eta_p,$$

we have

$$u = \frac{D\frac{1}{n}e - Dv}{e^T\left(D\frac{1}{n}e - Dv\right)} - x = \frac{-Dv + (x^Tv)x}{\frac{1}{n} - x^Tv}.$$

Let us ignore the size of the step, and consider just a vector ξ_p in the (opposite) direction of u:

$$\xi_p = D\eta_p - \left(x^T\eta_p\right)x = [D - xx^T]\eta_p.$$

Note that $x^Te = 1$, so we have

$$[D - xx^T]\left[I - \frac{1}{n}ee^T\right] = D - \frac{1}{n}xe^T - xx^T + \frac{1}{n}xx^Tee^T = D - xx^T.$$

Thus, the algorithm assigns to the point x the vector

$$\xi_p = \xi_p(x) = [D - xx^T]\left[I - DA^T(AD^2A^T)^{-1}AD\right] Dc,$$

to define the search direction. As in the case of the linear rescaling algorithm, the new point has the form $x - \alpha(x)\xi_p(x)$ where $\alpha(x)$ is positive. Note that ξ_p is well defined even without the hypothesis that the minimum of the objective function is equal to zero.

1.3. *The barrier function technique.* The logarithmic barrier technique considers the nonlinear optimization problem

$$(SF(\mu))\qquad \begin{aligned} &\text{Minimize} \quad F_\mu(x) = c^Tx - \mu\sum_j \ln x_j \\ &\text{subject to} \quad Ax = b, \\ &\qquad\qquad\qquad x > 0, \end{aligned}$$

where $\mu > 0$ is a scalar. If $x^*(\mu)$ is an optimal solution for $SF(\mu)$, and if $x^*(\mu)$ tends to a point x^* as μ tends to zero, then it follows that x^* is an optimal solution for the linear programming problem (SF). Consider the problem $(SF(\mu))$ where μ is fixed. As explained in [GMSTW], the Newton search direction v_μ at a point x is obtained by solving the following quadratic optimization problem:

$$\begin{aligned} &\text{Minimize} \quad \tfrac{1}{2}v^T\nabla^2F(x)v + (\nabla F(x))^Tv \\ &\text{subject to} \quad Av = 0, \quad \text{where} \end{aligned}$$

$$\nabla F(x) = c - \mu D_x^{-1}e \quad \text{and} \quad \nabla^2F(x) = \mu D_x^{-2}.$$

Let w_μ denote the vector of Lagrange multipliers. The vectors v_μ and w_μ must satisfy the following system of equations

$$\begin{pmatrix} \mu D_x^{-2} & A^T \\ A & O \end{pmatrix}\begin{pmatrix} v_\mu \\ w_\mu \end{pmatrix} = \begin{pmatrix} c - \mu D_x^{-1}e \\ 0 \end{pmatrix}.$$

Let $\eta_\mu = D_x^{-1}\mu v_\mu$. Thus,

$$\begin{pmatrix} I & D_x A^T \\ AD_x & O \end{pmatrix}\begin{pmatrix} \eta_\mu \\ w_\mu \end{pmatrix} = \begin{pmatrix} D_x c - \mu e \\ 0 \end{pmatrix}.$$

It follows that

$$\eta_\mu = \left(I - D_x A^T\left(AD_x^2A^T\right)^{-1}AD_x\right)(D_x c - \mu e) \quad \text{and}$$

$$V_\mu(x) = \mu v_\mu = D_x\eta_\mu = D_x\left(I - D_x A^T\left(AD_x^2A^T\right)^{-1}AD_x\right)(D_x c - \mu e)$$

is the vector field corresponding to the fixed value of μ. It was noted in [GMSTW] that $\xi_l(x) = \lim_{\mu \to 0} V_\mu(x)$.

In this paper we study the boundary behavior of the above interior point algorithms for linear programming. We study both the continuous trajectories of the vector fields induced by these algorithms and the discrete sequences of iterates of a point given by transformations of the polytope to itself. (In both cases we sometimes refer to these trajectories as orbits.) In §2 we show that, although the algorithms are defined on the interior of the feasible polyhedron, the vector fields actually extend continuously to the whole closed polyhedron. This is true even when the problem is degenerate. In §3 we provide conditions under which a vector field gives rise to trajectories that visit the neighborhoods of all the vertices of the Klee-Minty cube. The linear rescaling algorithm satisfies these conditions. Thus, limits of such trajectories obtained when a starting point is pushed to the boundary may have an exponential number of breakpoints. It is shown that limits of projective rescaling trajectories may have a linear number of such breakpoints. Projective rescaling trajectories may visit the neighborhoods of linearly many vertices. In §§4 and 5 we consider the behavior of the linear rescaling trajectories near vertices. We show that all the trajectories have a unique direction of convergence to the optimum. This direction is given by the vector of the reciprocal values of the reduced costs of the nonbasic variables at the vertex. In §6 we prove the differentiability (over the closed polytope) of the vector field underlying the logarithmic barrier technique with a fixed parameter, assuming nondegeneracy. The linear rescaling algorithm is a special case. In §7 a similar result is proven for the projective rescaling vector field. §8 analyzes the boundary behavior of the discrete linear rescaling algorithm. The unique direction of convergence is proven for this case too. In §9 the boundary behavior of the discrete version of the projective rescaling algorithm is studied. The limiting behavior is characterized in terms of reduced problems where the feasible domains are faces of the given polyhedron. In Appendix A we describe the behavior of the linear rescaling algorithm on the unit hypercube. We show that each ascending sequence of adjacent vertices can be approximated by a trajectory. In Appendix B we consider the projective rescaling trajectories on the unit simplex. We show that certain trajectories visit all the vertices. Also, there are trajectories starting from the center and visiting the centers of linearly many faces of the simplex. Appendix C proves a lemma on orthogonal projections. In Appendices D and E we present similar results on the barrier function technique in inequality form. In Appendix F we include an extension of §2, proving the differentiability of the linear rescaling vector field on the closed feasible polyhedron. We also represent this derivative in terms of projections on nullspaces.

2. Interior point algorithms continuously extend to the boundary. As seen in §1, the central feature of the interior point algorithms under consideration is a projection of a certain vector on a certain subspace. In this section we study the behavior of the resulting vector as the current point of the algorithm tends to a boundary point.

Let A denote any fixed matrix of order $m \times n$. Let $N = \{1, \ldots, n\}$ and let I_1 and I_2 define a partition of N, i.e., $N = I_1 \cup I_2$ and $I_1 \cap I_2 = \varnothing$. Let A_i denote a submatrix of A consisting of the columns of A with indices in I_i ($i = 1, 2$). Similarly, for any n-vector v, let v_i denote the subvectors of v consisting of the components of v corresponding to the sets I_i ($i = 1, 2$). Let $D(v)$ denote a diagonal matrix whose diagonal consists of the components of v. Let c denote any fixed n-vector and let c_1 and c_2 denote its subvectors as defined above.

Given x, a step of the linear rescaling algorithm amounts to the evaluation of the orthogonal projection of a vector $D(x)c$ on a linear subspace $L(x) = L(x; A) = \{y\colon AD(x)y = 0\}$. We are interested here in the behavior of this projection when x tends to a limit point $\bar{x}$. The interesting case is when some of the components of $\bar{x}$ are zero. Let I_1 denote the set of indices j for which $\bar{x}_j \neq 0$. For simplicity of notation, we assume $\bar{x}_j > 0$ ($j \in I_1$) but this is not really necessary for the argument. If $\bar{x}$ is a feasible point then of course this condition holds.

The orthogonal projection of $D(x)c$ on $L(x)$ is equal to the point in $L(x)$ which is closest to $D(x)c$. Thus, it is the solution of the following optimization problem (where the decision variables are the components of y):

$$\begin{aligned} &\text{Minimize} && \|D(x)c - y\|^2 \\ &\text{subject to} && AD(x)y = 0. \end{aligned}$$

With the notation introduced above, the latter is equivalent to

$$\begin{aligned} &\text{Minimize} && \|D(x_1)c_1 - y_1\|^2 + \|D(x_2)c_2 - y_2\|^2 \\ &\text{subject to} && A_1D(x_1)y_1 + A_2D(x_2)y_2 = 0. \end{aligned}$$

Let us denote this projection by $y(x)$, and also let $y_1(x)$ and $y_2(x)$ denote the restrictions to the sets of indices I_1 and I_2, respectively. Obviously, if x tends to $\bar{x}$ then the point $D(x)c$ tends to the point $D(\bar{x})c$. The distance between $D(x)c$ and $y(x)$ is always less than or equal to $\|D(x)c\|$ since the origin is in the linear subspace. It follows that the point $y(x)$ is bounded while x tends to $\bar{x}$. Since x_2 tends to zero, the vector $A_2D(x_2)y_2(x)$ also tends to zero (since $y_2(x)$ is bounded). Observe that the point $y_1(x)$ is the orthogonal projection of the point $D(x)c$ on the affine subspace

$$\Phi(x) = \{u\colon A_1D(x_1)u = -A_2D(x_2)y_2(x)\}.$$

Consider the point-to-set mapping Φ that takes every $x \in R^n$ to $\Phi(x)$. First, recall the definition of a continuous point-to-set mapping:

DEFINITION 2.1. Let Ψ be a point-to-set mapping that takes points $x \in R^n$ to subsets $\Psi(x)$ of R^m. The mapping ψ is *continuous* at $\bar{x}$ if for any sequence $\{x^k\}$, converging to a point $\bar{x}$, the following is true

(i) for any convergent sequence $\{z^k\}$, where $z^k \in \Psi(x^k)$, necessarily $\bar{z} = \lim z^k \in \Psi(\bar{x})$.

(ii) for any point $z' \in \Psi(\bar{x})$, there exists a sequence $\{z^k\}$ converging to z' where $z^k \in \Psi(x^k)$.

PROPOSITION 2.2. *The mapping $\Phi(v)$ is continuous at $\bar{x}$.*

PROOF. Let $\{x^k\}$ be any sequence converging to $\bar{x}$. By assumption, $\bar{x}_1 > 0$ and $\bar{x}_2 = 0$. Notice that $\Phi(\bar{x}) = \{u: A_1 D(x_1)u = 0\}$. Obviously, condition (i) is satisfied since $A_2 D(x_2^k) y_2(x^k)$ tends to zero. In other words, the set Ω of all limits of sequences $\{u^k\}$, such that $u^k \in \Phi(x^k)$, is contained in the subspace $\Phi(\bar{x})$. It is easy to check that Ω is a linear subspace, which is in a sense the limit of the affine subspaces $\Phi(x^k)$. The dimension of Ω is the same as the common dimension of all the $\Phi(x^k)$'s for k sufficiently large. This dimension is obviously equal to $|I_1| - \operatorname{rank}(A_1)$. On the other hand, $\Phi(\bar{x})$ is a linear subspace of the same dimension (since $\bar{x}_1 > 0$). It follows that $\Phi(\bar{x}) = \Omega$ and this completes the proof.

PROPOSITION 2.3. *If x tends to $\bar{x}$ (where $\bar{x}_1 > 0$ and $\bar{x}_2 = 0$) then the point $y_1(x)$ tends to the projection of $D(\bar{x}_1)c_1$ on the linear subspace $\Phi(\bar{x})$.*

PROOF. Given the interpretation of the orthogonal projection as the closest point, the proof is immediate.

COROLLARY 2.4. *The limit of the orthogonal projection of $D(x)c$ on the subspace $\{z: AD(x)z = 0\}$ is equal to the orthogonal projection of $D(\bar{x})c$ on the subspace $\{z: AD(\bar{x})z = 0\}$.*

PROOF. It suffices to show that $y_2(x)$ tends to zero since the orthogonal projection of $D(\bar{x})c$ onto $\{z: D(\bar{x})c = 0\}$ is of the form $(y_1(\bar{x}), 0)$ and since $y_1(x)$ tends to $y_1(\bar{x})$ by Proposition 2.3. Assume that $y_2(x)$ has a limit point $\tilde{y}_2 \neq 0$. (It obviously has some limit point by the boundedness of $\|y(x)\|$.) Then $\|D(x)c - (y_1(x), y_2(x))\|^2$ tends to $\|D(\bar{x})c - (y_1(x), 0)\|^2 + \|\tilde{y}_2\|^2$. On the other hand, letting $\check{y}_1(x) = D^{-1}(x_1)D(\bar{x}_1)y_1(x)$, we have that $\|D(x)c - (\check{y}_1(x), 0)\|$ tends to $\|D(\bar{x})c - (y_1(\bar{x}), 0)\|$. Together these imply that for x sufficiently near to $\bar{x}$,

$$\|D(x)c - (y_1(x), y_2(x))\| > \|D(x)c - (\check{y}_1(x), 0)\|.$$

However, $(\check{y}_1(x), 0) \in \{z: AD(x)z = 0\}$. Thus we reach a contradiction to the fact that $(y_1(x), y_2(x))$ is the orthogonal projection of $D(x)c$ onto $\{z: AD(x)z = 0\}$.

The vector $\xi = \xi(x)$ assigned by the linear rescaling algorithm to a point x can be described as $\xi(x) = D(x)y$ where y is the projection of the vector $D(x)c$ on the subspace $\{y: AD(x)y = 0\}$. Thus, we have the following proposition:

PROPOSITION 2.5. *Suppose $x \in R^n$ satisfies $Ax = b$, has positive components, and tends to a point $\bar{x}$ such that $\bar{x}_1 > 0$ and $\bar{x}_2 = 0$. Then the vector $\xi(x)$ of the linear rescaling algorithm at $x > 0$ in the problem (SF) tends to the vector $\xi(\bar{x}_1)$ assigned by this algorithm at $\bar{x}_1$ in the problem*

$$\begin{aligned} &\textit{Minimize} \quad c_1^T z \\ &\textit{subject to} \quad A_1 z = b, \\ &\qquad\qquad\qquad z \geqslant 0. \end{aligned}$$

The argument for similar results about the projective rescaling algorithm and the barrier function technique are essentially the same:

PROPOSITION 2.6. *Suppose $x \in R^n$ satisfies $Ax = 0$ and $e^T x = 1$, has positive components, and tends to a point $\bar{x}$ such that $\bar{x}_1 > 0$ and $\bar{x}_2 = 0$. Then the vector $\xi_p(x)$*

assigned by the projective algorithm at a point $x > 0$ in the problem (KSF) *tends to the vector* $\xi_p(\bar{x}_1)$ *assigned by this algorithm at* $\bar{x}_1$ *in the problem*

$$\begin{aligned} \text{Minimize}\quad & c_1^T z \\ \text{subject to}\quad & A_1 z = 0, \\ & e_1^T z = 1, \\ & z \geqslant 0. \end{aligned}$$

PROOF. We have $\xi_p = \xi_p(x) = [D - xx^T]y$ where y is the projection of $D(x)c$ on the subspace $\{z\colon AD(x)z = 0\}$ and the proof follows easily.

PROPOSITION 2.7. *Suppose $x \in R^n$ satisfies $Ax = b$ and has positive components, and tends to a point $\bar{x}$ such that $\bar{x}_1 > 0$ and $\bar{x}_2 = 0$. Let $\mu > 0$ be fixed. Then the vector $V_\mu(x)$ assigned by the Newton logarithmic barrier function method at $x > 0$ in the problem $(SF(\mu))$ tends to the vector $V_\mu(\bar{x}_1)$ assigned by this algorithm at $\bar{x}_1$ in the problem*

$$\begin{aligned} \text{Minimize}\quad & c_1^T z - \mu \sum_{j \in I_1} \ln z_j \\ \text{subject to}\quad & A_1 z = b, \\ & z > 0. \end{aligned}$$

PROOF. The vector $V_\mu(x)$ assigned to x can be represented as $D(x)(y' - y'')$ where y' and y'' are the projections of $D(x)c$ and μe, respectively, on the subspace $\{z\colon AD(x)z = 0\}$. The argument about the vector y' is the same as in Proposition 2.5. The argument about the vector y'' is similar. The vector e is a sum $e = e' + e''$ of vectors where $e_j' = 1$ for $j \in I_1$ and $e_j'' = 1$ for $j \in I_2$. The projection of the vector e'' on the subspace $\{z\colon AD(x)z = 0\}$ is bounded so $D(x)$ times it tends to zero.

3. Interior point algorithms and the Klee-Minty cube. Some variants of the simplex method require exponential numbers of pivot steps in the worst case. The first examples of such behavior were provided by Klee and Minty [KM]. The "tilted cube" described in their paper is a very useful construct which we also use here.

The n-dimensional Klee-Minty cube is defined by the following inequalities:

$$(KM)\qquad \begin{aligned} & \nu \leqslant x_1 \leqslant 1 - \nu, \\ & \nu x_{j-1} \leqslant x_j \leqslant 1 - \nu x_{j-1} \qquad (j = 2, \ldots, n), \end{aligned}$$

where ν is any positive number less than $\frac{1}{2}$. The associated linear programming problem is to maximize the value of x_n subject to the set of inequalities (KM). It can be verified that the maximum is attained at a unique point, namely, the vertex $(\nu, \nu^2, \ldots, \nu^{n-1}, 1 - \nu^n)$.

If x is a vertex of the (KM) cube then obviously each x_j equals either the lower or the upper bound implied by the values of the other components of x. This suggests a correspondence between vertices of the (KM) cube and vertices of the unit cube. Thus, we use a $(0,1)$-vector $v = (v_1, \ldots, v_n)$ to describe the vertex x of (KM) where $x_1 = (1 - v_1)\nu + v_1(1 - \nu)$, and for every $j \geqslant 2$, $x_j = (1 - v_j)\nu x_{j-1} + v_j(1 - \nu x_{j-1})$. We say that v is the characteristic vector of the vertex x. Some simplex variants visit all the vertices of (KM) (or an analogous construct) in a nice order which can be described, inductively, as follows. The case $n = 1$ is trivial (the two vertices are the numbers ν and $1 - \nu$). Let $v^1, \ldots, v^m$ be the sequence of characteristic vectors of the vertices of the $(n-1)$-dimensional (KM) cube in the order they are visited $(m = 2^{n-1})$. Then the 2^m vertices of the n-dimensional (KM) cube $((v^j, 0), (v^j, 1), j = 1, \ldots, m)$ are visited in the following order:

$$(v^1, 0), (v^2, 0), \ldots, (v^m, 0), (v^m, 1), (v^{m-1}, 1), \ldots, (v^1, 1).$$

Faces of the (KM) cube can be easily described by the characteristic vectors. Thus, a d-dimensional face Φ is described by $n - d$ equations of the form $v^j = e^j$, where $e^j \in \{0, 1\}$. We denote the relative interior of a face Φ by $\dot{\Phi}$. It is interesting to note that every face Φ has a unique point $x^*(\Phi)$ where the value of x_n is maximized over Φ. We call this point the *optimal* point of Φ.

We shall later consider the vector field induced by the linear rescaling algorithm on the (KM) cube. However, we first discuss the subject in a more general context. Let us identify the linear programming problem

$$(\tilde{P}) \qquad \begin{aligned} &\text{Maximize} \quad c^T x \\ &\text{subject to} \quad Ax \geqslant b \end{aligned}$$

with the triple (A, b, c). We are interested here in algorithms that can be described by vector fields as follows. The underlying vector field $\mathbf{A}$ is defined for quadruples $(x; A, b, c)$ where $A \in R^{m \times n}$, $b \in R^m$ and $c, x \in R^n$, such that $Ax \geqslant b$. The vector field assigns a vector $y = \mathbf{A}(x; A, b, c) \in R^n$ such that $A(x + y) \geqslant b$. The vector field describes an iterative algorithm defined by $x^{k+1} = x^k + \mathbf{A}(x; A, b, c)$.

We need our algorithms to be defined in a slightly more general context. First, the algorithms extend to minimization problems in the obvious way that the direction assigned in the "minimize $c^T x$" problem is the same as the direction assigned in the "maximize $-c^T x$" problem. Also, we assume the algorithm is defined for affine objective functions $c^T x + c_0$ and the vector field is independent of the constant c_0. Similarly, if an inequality is given in a more general form, $d^T x + \delta \geqslant g^T x + \gamma$, then the algorithm converts it into $(d - g)^T x \geqslant \gamma - \delta$.

The vector field $\mathbf{A}$ and the algorithm $\mathbf{A}$ will be referred to interchangeably. Conceptually, the discrete iterates of the algorithm approximate the solution curves of the vector field $\mathbf{A}$. We now state conditions on the algorithm $\mathbf{A}$ which are needed for establishing "long" paths in the (KM) cube. The corresponding linear programming problem is nondegenerate. Thus we need these conditions to hold only for nondegenerate problems.

1. *Reversibility*. The algorithm is called reversible if, when the objective function vector is multiplied by -1, the direction of movement from x is reversed:

$$\frac{\mathbf{A}(x; A, b, -c)}{\|\mathbf{A}(x; A, b, -c)\|} = -\frac{\mathbf{A}(x; A, b, c)}{\|\mathbf{A}(x; A, b, c)\|}.$$

In other words, the directions computed by the algorithm in the minimization and the maximization problems (with the same data) are precisely opposed to each other.

2. *Independence of the representation*. First, this condition includes all the assumptions listed above with respect to the extensions of the algorithm to problems in the minimization form and inequalities in nonstandard form. In addition, we require the

following: (i) The vector field is invariant under permutations of the set of inequalities. In other words, if Q is a permutation matrix then

$$\mathbf{A}(x; QA, Qb, c) = \mathbf{A}(x; A, b, c).$$

(ii) The vector field is invariant under "affine scaling automorphisms" in a sense as follows. Consider an affine transformation of R^n, $T(x) = Mx + q$, where M is *diagonal*. Denoting the new variable $y = T(x)$ (so $x = M^{-1}(y - q)$), the problem $(\tilde{P})$ is transformed into

$$\text{Maximize} \quad c^T M^{-1} y$$

$$\text{subject to} \quad AM^{-1}y \geqslant b + AM^{-1}q.$$

Thus, the quadruple $(x; A, b, c)$ is transformed into

$$(x'; A', b', c') = (Mx + q; AM^{-1}, b + AM^{-1}q, M^{-T}c).$$

A translation Δx maps to a translation $\Delta y = M\Delta x$ (since $y + \Delta y = M(x + \Delta x) + q$). Suppose the new problem (A', b', c') is the same as (A, b, c) up to permutation of the set of inequalities (that is, there exists a permutation matrix Q such that $A' = QA$ and $b' = Qb$), and up to changing the sense of the optimization from maximization to minimization or vice versa, that is, c' is in the direction of $\pm c$. In this case our condition requires that the direction assigned in the transformed problem to the transformed point be equal to the transformed direction assigned to the original point in the original problem. In other words,

$$\mathbf{A}(x'; A', b', c') = M\mathbf{A}(x; A, b, c).$$

3. *Continuity*. The vector field $\mathbf{A}$ is continuous at every x such that $Ax = b$.

4. *Invariance of faces*. The vector $\mathbf{A}(x; A, b, c)$ is tangent to any face Φ of the feasible polyhedron such that $x \in \Phi$, is equal to the vector field of the problem restricted to the face, and satisfies (1)–(3) on the face. Note that this condition necessitates that $\mathbf{A}(x; A, b, c) = 0$ if x is a vertex.

5. *Convergence*. For every bounded face Φ of the (nondegenerate) feasible polyhedron which contains the optimal vertex and every $x^0 \in \dot{\Phi}$, the orbit induced by the vector field $\mathbf{A}$ at x converges to this optimal vertex.

DEFINITION 3.1. A vector field $\mathbf{A}$ (or, equivalently, an algorithm subject to the interpretation given above) that satisfies the conditions of reversibility, independence of the representation, continuity, invariance and convergence, defined above (in nondegenerate problems), will be called *proper*.

Note that by the reversibility assumption, the orbit induced at a point $x^0 \in \dot{\Phi}$ (where Φ is bounded) by a proper algorithm, converges at one end to a maximum point and at the other end to a minimum point of the face Φ. Also, the restriction of a proper algorithm to any face of the feasible polyhedron is itself a proper algorithm.

LEMMA 3.2. *If* $\mathbf{A}$ *is a proper algorithm then all the orbits induced by* $\mathbf{A}$ *on the* (KM) *cube are symmetric with respect to the hyperplane* $H = \{x \in R^n \colon x_n = \frac{1}{2}\}$. *More precisely, if* π *is one such orbit then a point* $x = (x_1, \ldots, x_n)^T$ *is on* π *if and only if the point* $\bar{x} = (x_1, \ldots, x_{n-1}, 1 - x_n)^T$ *is on* π.

PROOF. The given problem is

$$(P_1)\qquad \begin{array}{lll} \text{Maximize} & x_n & \\ \nu & \leqslant x_1 \leqslant 1-\nu & \\ \nu x_{j-1} & \leqslant x_j \leqslant 1-\nu x_{j-1} & (j=2,\dots,n). \end{array}$$

Consider the transformation of reflection with respect to the hyperplane H, that is, $\bar{x}_n = 1 - x_n$. Let $\bar{x} = (x_1, \dots, x_{n-1}, 1 - x_n)^T$. The affine transformation is given by the matrix

$$M = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & -1 \end{pmatrix}$$

and the vector $q = (0, \dots, 0, 1)^T$. The substitution $x_n = 1 - \bar{x}_n$ transforms the original problem into the following:

$$(P_2)\qquad \begin{array}{lll} \text{Maximize} & 1-\bar{x}_n & \\ \nu & \leqslant x_1 \leqslant 1-\nu & \\ \nu x_{j-1} & \leqslant x_j \leqslant 1-\nu x_{j-1} & (j=2,\dots,n-1) \\ \nu x_{n-1} & \leqslant 1-\bar{x}_n \leqslant 1-\nu x_{n-1} & \end{array}$$

which is equivalent to

$$(P_3)\qquad \begin{array}{lll} \text{Minimize} & \bar{x}_n & \\ \nu & \leqslant x_1 \leqslant 1-\nu & \\ \nu x_{j-1} & \leqslant x_j \leqslant 1-\nu x_{j-1} & (j=2,\dots,n-1). \\ \nu x_{n-1} & \leqslant \bar{x}_n \leqslant 1-\nu x_{n-1} & \end{array}$$

The latter is simply the minimization problem with the same data as in (P_1). Let $\mathbf{A}_i(x)$ denote the direction assigned at any point x in the problem (P_i) $(i = 1, 2, 3)$. By the properties of independence of the representation and reversibility,

$$\mathbf{A}_1(x) = M^{-1}\mathbf{A}_2(\bar{x}).$$

Also,

$$\mathbf{A}_2(\bar{x}) = \mathbf{A}_3(\bar{x}).$$

By reversibility,

$$\mathbf{A}_3(\bar{x}) = -\mathbf{A}_1(\bar{x}).$$

Thus,

$$\mathbf{A}_1(x) = -M^{-1}\mathbf{A}_1(\bar{x}).$$

Note that $M^{-1} = M$ so

$$(\mathbf{A}_1(x))_j = -(\mathbf{A}_1(\bar{x}))_j, \qquad (j = 1, \dots, n-1)$$

$$(\mathbf{A}_1(x))_n = (\mathbf{A}_1(\bar{x}))_n.$$

In particular, if $x_n = \frac{1}{2}$ then $x = \bar{x}$ and we get $(\mathbf{A}_1(x))_j = 0$ for $j = 1, \ldots, n-1$. It follows that the point sets of the orbits through x and $\bar{x}$ coincide, and also if time is reversed in the upper half of the cube, then the orbit starting at x and the one starting at $\bar{x}$ reach the hyperplane H at the same time, hitting it perpendicularly.

We are now ready to state a theorem on long paths.

THEOREM 3.3. *If* $\mathbf{A}$ *is a proper algorithm then for every* $\epsilon > 0$, *there exists an orbit* π, *which is induced by* $\mathbf{A}$ *on the* (KM) *cube, such that for every vertex* v *of the cube, the distance between* v *and the orbit* π *is less than* ϵ.

PROOF. We prove the theorem by induction on the dimension of the cube. The theorem is trivial for $n = 1$. Consider the general case $n > 2$. Consider the restriction of the (KM) problem to the "base" of the cube, that is, the face Φ characterized by the equality $x_n = \nu x_{n-1}$. Thus, the problem of maximizing x_n on Φ is equivalent to the problem of maximizing x_{n-1} on Φ, that is, the (KM) problem in dimension $n-1$. It follows by the induction hypothesis that for every ϵ there exists an orbit π', that lies completely within the base Φ, such that the distance between any vertex of Φ and π' is less than ϵ. Given $\epsilon > 0$, let y denote a point in Φ such that for every vertex v of the base Φ, the distance between v and the orbit through y is less than ϵ. If x is an interior point of the (KM) cube which is sufficiently close to y then, by continuity, the distances between the orbit through x and all the vertices of the base are each less than ϵ. Moreover, by the symmetry proved in Lemma 3.2, the point $\bar{x} = (x_1, \ldots, x_{n-1}, 1 - x_n)$ also has the property that the distances between the orbit through $\bar{x}$ and all the vertices of the "ceiling" (that is, the face characterized by the equality $x_n = 1 - \nu x_{n-1}$) are each less than ϵ. However, these two orbits are actually the same by Lemma 3.2 and this completes the proof.

It is easy to see that, for ϵ sufficiently small, the path (whose existence was proven in Theorem 3.3) visits the ϵ-neighborhoods of the vertices of the cube in ascending order with respect to the nth coordinate, so in a certain sense it approximates the behavior of the simplex method. It is also interesting to note that not every ascending sequence of adjacent vertices can be approximated by an orbit of the algorithm. The latter follows from the symmetry property since the sequence of visited vertices of the base determines the sequence of visited vertices of the ceiling. Interestingly, on the regular unit hypercube every ascending sequence of adjacent vertices can be approximated by an orbit of the algorithm. This is shown in Appendix A.

We now show that the linear rescaling algorithm is proper. The linear rescaling algorithm was stated originally for problems in standard form. We now recast it in inequality form and prove it is proper. For problems in the form $(\tilde{P})$ we can do one of two things:

(i) We can introduce surplus variables $s = Ax - b$ constrained to be nonnegative. We then eliminate the x variables. Assume without loss of generality that

$$A = \begin{pmatrix} B \\ N \end{pmatrix}$$

where $B \in R^{n \times n}$ is nonsingular and $N \in R^{(m-n) \times n}$. Represent $s = (s_B, s_N)$ and $b = (b_B, b_N)$ accordingly. Thus, $x = B^{-1}(b_B + s_B)$ and the problem is

$$(\tilde{P}_s) \qquad \begin{aligned} &\text{Maximize} && c^T B^{-1} s_B \\ &\text{subject to} && N B^{-1} s_B - s_N = b_N - N B^{-1} b_B, \\ & && s \geqslant 0. \end{aligned}$$

It can be verified that if $(x'; A', b', c')$ is obtained from $(x; A, b, c)$ by a *general* affine transformation as above, then both these problems have the same representation in the form $(\tilde{P}_s)$.

(ii) We can develop an analogous algorithm, for problems in inequality form, based on similar principles. This is included in Appendix D. The search direction is then given by the vector

$$v = v(x) = \left(A^T D_s^{-2} A\right)^{-1} c \quad \text{where}$$

$$D_s = D_s(x) = \mathrm{Diag}(A_1 x - b_1, \ldots, A_m x - b_m).$$

We now prove that the algorithms outlined in (i) and (ii) above are actually the same.

PROPOSITION 3.4. *The vector $v = (A^T D_s^{-2} A)^{-1} c$ is equal to the vector u assigned at x by applying the affine rescaling algorithm in standard form to the corresponding problem $(\tilde{P}_s)$.*

PROOF. Let D_B and D_N denote, respectively, the diagonal submatrices of D of the orders $n \times n$ and $(m - n) \times (m - n)$ corresponding to B and N. The direction Δs in the space of the s variables is obtained by projecting the vector $(D_B B^{-T} c, 0) \in R^m$ orthogonally into the nullspace of the matrix $(NB^{-1} D_B, -D_N) \in R^{(m-n)\times m}$, and then multiplying the result by D_s. Thus $\Delta s = (\Delta s_B, \Delta s_N)$ is the solution of the following problem

$$\begin{aligned} &\text{Minimize} \quad \|D_B B^{-T} c - D_B^{-1}\, \Delta s_B\|^2 + \|D_N^{-1}\, \Delta s_N\|^2 \\ &\text{subject to} \quad NB^{-1}\Delta s_B - \Delta s_N = 0. \end{aligned}$$

This is equivalent to

$$\text{Minimize} \quad \|D_B B^{-T} c - D_B^{-1}\, \Delta s_B\|^2 + \|D_N^{-1} N B^{-1}\, \Delta s_B\|^2.$$

Thus

$$-D_B^{-1}\left(D_B B^{-T} c - D_B^{-1}\, \Delta s_B\right) + B^{-T} N^T D_N^{-2} N B^{-1}\, \Delta s_B = 0 \quad \text{or}$$

$$\left(D_B^{-2} + B^{-T} N^T D_N^{-2} N B^{-1}\right) \Delta s_B = B^{-T} c.$$

Since

$$A^T D_s^{-2} A = B^T D_B^{-2} B + N^T D_N^{-2} N,$$

it follows that

$$\left(A^T D_s^{-2} A\right) B^{-1}\, \Delta s_B = c$$

which completes the proof.

PROPOSITION 3.5. *The linear rescaling algorithm, applied to problems in the form $(\tilde{P})$, is proper in the sense of Definition* 3.1.

PROOF. In view of Proposition 3.4, we can rely on either form of the algorithm for proving the required conditions. Reversibility is trivial to verify. Independence of the representation follows from the fact that the vector $s = Ax - b$ is invariant; thus,

$$\left((AM^{-1})^T D_s^{-2}(AM^{-1})\right)^{-1}(M^{-T}c) = \left(M^{-T}A^T D_s^{-2} A M^{-1}\right)^{-1} M^{-T}c$$

$$= M\left(A^T D_s^{-2} A\right)^{-1} c.$$

Continuity and invariance of faces were proven in §2. Appendix D contains analogous proofs for inequality form. We now consider the convergence of linear rescaling trajectories. First, note that the objective function is monotone increasing along trajectories. Thus, all the accumulation points of a trajectory must have the same objective function value. Moreover, if x is neither a vertex nor an optimal point, the objective function strictly increases along any trajectory in a neighborhood of x. By continuity, this implies that the only candidates for accumulation points are vertices of the feasible polyhedron and optimal solutions. In §§4 and 5 we analyze the behavior of the trajectories near vertices (see also Appendix E). It follows from our analysis that interior trajectories cannot accumulate in nonoptimal vertices, and, therefore, if there is a unique optimal solution, all the interior trajectories converge to it.

Interestingly, Theorem 3.3 does not apply to the projective rescaling algorithm. Two requirements of Definition 3.1 are not satisfied. First the reversibility requirement fails. Recall that the algorithm has to be applied to the problem in the form (KSF) with the additional requirement that the optimal value be equal to zero. The transformation that takes a problem into this form when we wish to reverse the sense of the optimization causes a change in the direction of search which is, in general, not the reverse direction. Second, although the invariance of faces holds, convergence within a face is not necessarily to the optimum of the face, unless the face contains the global optimum of the problem. The reason is that the projective rescaling algorithm induces paths that converge within faces to optima of a "reduced" potential function. More precisely, let

$$P = \left\{x \in R^n \colon Ax = 0, e^T x = 1, x \geqslant 0\right\}$$

denote the feasible polytope and for $J \subset N = \{1, \ldots, n\}$ let

$$\Phi_J = P \cap \left\{x \in R^n \colon x_j = 0, j \notin J\right\}$$

denote a face of P. Every nonempty face Φ_J of P contains a *center*, namely, a point q_J where the reduced potential function $\psi_J(x) = |J| c^T x - \sum_{j \in J} \ln x_j$ is minimized over Φ_J. If the minimum of $c^T x$ over Φ_J is zero then paths through the interior of Φ_J converge to such a minimum of $c^T x$. The latter lies on the relative boundary of the face unless the linear function is constant on the face. If the minimum is not zero, the point q_J is interior. A detailed discussion of these issues is given in §7.

We now consider the vector field $V_\mu(x)$ given by the Newton logarithmic barrier function method with a fixed μ. This vector field is initially defined for $\mu > 0$. It is obviously not proper since convergence is to the optimum of the nonlinear approximate objective function rather than the given linear objective function. Recall that $V_\mu(x)$ has a limit as μ tends to zero and, moreover, the direction of the limit $V_0(x)$ coincides with the direction assigned by the linear rescaling algorithm $\xi_l(x)$ (see [GMSTW]). Thus, the vector field $V_0(x)$ is proper. It follows that although $V_\mu(x)$ is not proper, it has "long" paths if μ is sufficiently small. More precisely, PROPOSITION 3.6. *For every $\epsilon > 0$, there exists a $\mu_0 > 0$ such that for every fixed μ, $0 < \mu < \mu_0$, the vector field $V_\mu(x)$ on the (KM) cube has solution paths that visit the ε-neighborhoods of all the vertices.*

4. The behavior of the linear rescaling algorithm near vertices. Consider the linear programming problem in standard form (SF). Let B denote the square matrix of order m, consisting of the first m columns of A. We assume B is nonsingular and $B^{-1}b > 0$. In other words, B is a nondegenerate feasible basis. Let N denote the matrix of order $m \times (n-m)$ consisting of the last $n-m$ columns of A. We denote the restriction of any n-vector v to the first m coordinates by v_B, and its restriction to the last $n-m$ coordinates by v_N. Thus, the objects c_B, c_N, x_B and x_N are defined with respect to the vectors c and x. Recall that $D = D(x)$ is a diagonal matrix (of order n) whose diagonal entries are the components of the vector x. Also, D_B and D_N are diagonal matrices of orders m and $n-m$, respectively, corresponding to the vectors ξ_B and ξ_N.

In the transformed space, the direction $\eta = T_x(\xi)$ is the solution of the following least-squares problem:

$$\begin{aligned}&\text{Minimize} \quad \|Dc - \eta\|^2\\ &\text{subject to} \quad AD\eta = 0.\end{aligned}$$

This is equivalent to

$$\begin{aligned}&\text{Minimize} \quad \|D_Bc_B - \eta_B\|^2 + \|D_Nc_N - \eta_N\|^2\\ &\text{subject to} \quad BD_B\eta_B + ND_N\eta_N = 0.\end{aligned}$$

In the original space, $\xi = D\eta$, so the problem is

$$\begin{aligned}&\text{Minimize} \quad \|D_Bc_B - D_B^{-1}\xi_B\|^2 + \|D_Nc_N - D_N^{-1}\xi_N\|^2\\ &\text{subject to} \quad B\xi_B + N\xi_N = 0.\end{aligned}$$

Eliminating ξ_B by the substitution $\xi_B = -B^{-1}N\xi_N$, we obtain an equivalent problem:

$$\text{Minimize} \quad \|D_Bc_B + D_B^{-1}B^{-1}N\xi_N\|^2 + \|D_Nc_N - D_N^{-1}\xi_N\|^2.$$

A vector ξ_N is an optimal solution for the latter if and only if the gradient of the objective function vanishes; that is,

$$N^TB^{-T}D_B^{-1}\left(D_Bc_B + D_B^{-1}B^{-1}N\xi_N\right) - D_N^{-1}\left(D_Nc_N - D_N^{-1}\xi_N\right) = 0.$$

Equivalently,

$$\left(I + D_N^2N^TB^{-T}D_B^{-2}B^{-1}N\right)\xi_N = D_N^2\left(c_N - N^TB^{-T}c_B\right).$$

We now consider points x in the neighborhood of the vertex $v(B)$ determined by B, that is, $v_j(B) = (B^{-1}b)_j$ for $j = 1, \ldots, m$, and $v_j(B) = 0$ for $j = m+1, \ldots, n$. Obviously, if x tends to $v(B)$ then x_B tends to the positive vector $B^{-1}b$ and x_N tends to 0. Note that the coefficient matrix of the latter system is

$$I + D_N^2N^TB^{-T}D_B^{-2}B^{-1}N = I + O\left(\|x_N\|^2\right),$$

which tends to the identity matrix as x approaches $v(B)$. We thus have

PROPOSITION 4.1. *The nonbasic part of the search direction is*

$$(\xi_N)_i = \left(D_N^2\left(c_N - N^T B^{-T} c_B\right)\right)_i + O\left(\|x_N\|^2 |x_i|^2\right).$$

The basic part of the search direction satisfies $\|\xi_B\| \leqslant O(\|x_N\|^2)$.

Notice that the vector $\tilde{c}_N = c_N - N^T B^{-T} c_B$ is precisely the "reduced-cost" vector associated with the basis B.

We first provide some intuition about the behavior of trajectories near vertices based on the description of the asymptotic vector field. More rigorous arguments will be given later. Consider the orbit induced by the (asymptotic) vector field ξ at a point x in the neighborhood of $v(B)$. The underlying differential equations are $\dot{x}_j = -\tilde{c}_j x_j^2$ $(j = m+1, \dots, n)$, whose solution obviously is

$$x_j(t) = \frac{1}{\dfrac{1}{x_j(0)} + \tilde{c}_j t}.$$

Recall that $x_B(t)$ is determined by $x_N(t)$, namely, $x_B(t) = B^{-1}(b - Nx_N(t))$. Notice that $v(B)$ is the unique optimal solution of the linear programming problem if and only if for every j, $j = m+1, \dots, n$, $\tilde{c}_j > 0$. If this is the case then the trajectory $x(t)$ converges to $v(B)$. Moreover, as t tends to infinity, the direction of $x_N(t)$ tends to the direction of the vector

$$\left(\frac{1}{\tilde{c}_{m+1}}, \dots, \frac{1}{\tilde{c}_n}\right).$$

Note that we obtain a unique asymptotic direction near a vertex corresponding to each face containing the vertex.

5. More on the trajectories near vertices. Here we rigorously carry out the analysis suggested in the last section of trajectories near the optimal vertex. It is convenient in this section to assume the vector field is real analytic even though this is stronger than what is actually required. We now examine the behavior of the solution curves of the equation $\dot{x} = -x^2 + o(\|x\|^2)$ where $x \in R^n$ and we denote $x^2 = (x_1^2, \dots, x_n^2)^T$. It is convenient to express x in polar coordinates. We start with a slightly more general problem and follow Gomory [G].

Let F: $R^n \to R^n$ be a real analytic vector field defined in the neighborhood of the origin. Consider the differential equation $\dot{x} = F(x)$. Let $S^{n-1} = (x \in R^n\colon \|x\| = 1)$ denote, as usual, the unit sphere in R^n. A nonzero vector $x \in R^n$ is represented in polar coordinates by a pair (σ, u) where $\sigma = \sigma(x) = \|x\|$ and $u = u(x) = x/\sigma$. Thus, the vector x can be expressed as a product $x = \sigma u$ where $\sigma \in R_+$ and $u \in R^n$ with $\|u\| = 1$.

Consider a solution path $x = x(t)$ of the equation $\dot{x} = F(x)$. The polar coordinates of a point along the path are also functions of t, so we denote in short $\sigma = \sigma(t)$ and $u = u(t)$. We shall represent the equation $\dot{x} = F(x)$ in polar coordinates. The polar coordinates pairs (σ, u) are of course points in $R_+ \times S^{n-1}$. We shall obtain an equivalent vector field on a neighborhood of $\{0\} \times S^{n-1}$ relative to $R_+ \times S^{n-1}$.

First,

$$\sigma(t) = \sqrt{(x(t))^T x(t)}$$

so

$$\frac{d\sigma}{dt} = \frac{x^T\dot{x}}{\sqrt{x^Tx}} = u^TF(x) = u^TF(\sigma u).$$

Also,

$$u(t) = \frac{1}{\sigma(t)}x(t)$$

so

$$\frac{du}{dt} = \frac{1}{\sigma}\dot{x} - \frac{1}{\sigma^2}\frac{d\sigma}{dt}x = \frac{1}{\sigma}\left(F(\sigma u) - \left[u^TF(\sigma u)\right]u\right).$$

Since F is real analytic, it follows that $F(x) = \sum_{i=0}^{\infty} F_i(x)$ where for every i $(i = 0, 1, \dots)$, $F_i(x)$ is a homogeneous polynomial of degree i. In our case, $F(x) = -x^2 + o(\|x\|^2)$ where F is real analytic so F_0 and F_1 are identically zero. Whenever there exists an $m \geqslant 2$ such that for every $i < m$, F_i is identically zero, we have

$$\frac{d\sigma}{dt} = u^TF(\sigma u) = u^T\sum_{i=m}^{\infty} F_i(\sigma u) = u^T\sum_{i=m}^{\infty} \sigma^i F_i(u).$$

Similarly,

$$\frac{du}{dt} = \frac{1}{\sigma}\left(F(\sigma u) - \left[u^T(\sigma u)\right]u\right) = \sum_{i=m}^{\infty} \sigma^{i-1}\left(F_i(u) - \left[u^TF_i(u)\right]u\right).$$

We have obtained a vector field which is well defined in a neighborhood of $\{0\} \times S^{n-1}$ relative to $R_+ \times S^{n-1}$. In fact, if we divide by σ^{m-1} we still obtain a vector field on the same neighborhood and the orbits of the new vector field are the same as those of the old one (in $R_+ \times S^{n-1}$). Note that the sphere $\{0\} \times S^{n-1}$ is invariant in the sense that the flow induced by the field on this sphere remains in the sphere. Thus, we may consider, instead, the following equations:

$$\frac{d\sigma}{dt} = u^TF(\sigma u) = u^T\sum_{i=m}^{\infty} \sigma^{i-m+1}F_i(u), \quad \text{and}$$

$$\frac{du}{dt} = \sum_{i=m}^{\infty} \sigma^{i-m}\left(F_i(u) - \left[u^TF_i(u)\right]u\right).$$

As a vector field this can be written in the form

$$\tilde{V}(\sigma, u) = \left(u^T\sum_{i=m}^{\infty} \sigma^{i-m+1}F_i(u), \sum_{i=m}^{\infty} \sigma^{i-m}\left(F_i(u) - \left[u^TF_i(u)\right]u\right)\right)$$

(for $\sigma > 0$) and

$$\tilde{V}(0, u) = \left(0, F_m(u) - \left[u^TF_m(u)\right]u\right).$$

The latter is just the projection of the homogeneous equation $\dot{x} = F_m(x)$ into the unit sphere. The projections of the solution curves of the homogeneous equation are solution curves as computed above. The derivative of $\tilde{V}$ at a point $(0, u)$, where

$\tilde{V}(0, u) = 0$, is the following:

$$D\tilde{V}(0, u) = \begin{pmatrix} u^T F_m(u) & 0 \\ F_{m+1}(u) - [u^T F_{m+1}(u)]u & D_u(F_m(u) - [u^T F_m(u)]u) \end{pmatrix}.$$

We now return to our special case where $m = 2$ and

$$F_2(x) = -x^2 = -(x_1^2, \dots, x_n^2)^T$$

and will study the behavior of orbits on the sphere $\{0\} \times S^{n-1}$. Consider the equation $\dot{x} = -x^2$, that is, $\dot{x}_i = -x_i^2$ $(i = 1, \dots, n)$. If $x_i^0 > 0$ then by integration, the solution is

$$x_i(t) = \frac{1}{t + C_i}$$

and the ith component of the curve through x^0 is given by

$$x_i(t) = \frac{x_i^0}{x_i^0 t + 1}.$$

If $x_i^0 = 0$ then obviously $x_i(t) = 0$. For every i and j $(1 \leqslant i, j \leqslant n)$ if $x_i^0, x_j^0 > 0$ then

$$\lim_{t \to \infty} \frac{x_i(t)}{x_j(t)} = 1.$$

These orbits project to orbits on the sphere.

We now study the zeros of the vector field

$$\tilde{V}(0, u) = (0, F_2(u) - [u^T F_2(u)]u) = (0, -u^2 + [u^T u^2]u).$$

The solutions of the system

$$\tilde{V}(0, u) = 0, \qquad u \in S_+^{n-1}, \quad u \geqslant 0,$$

are the nonnegative solutions of the system

$$-u_k^2 + \left(\sum_{i=1}^{n} u_i^3\right) u_k = 0 \qquad (k = 1, \dots, n).$$

It can easily be verified that a solution u of this system is characterized as follows. There exists a j $(1 \leqslant j \leqslant n)$ such that j components of the vector u equal $1/\sqrt{j}$ while the rest of the components are zero.

The *forward* or *ω-limit points* of an orbit $u(t)$ are those points u^0 such that there is a sequence of reals t_i tending to infinity with $u(t_i)$ tending to u^0. *Backward* or *α-limit points* are defined by letting t_i tend to $-\infty$. We see from the discussion above that the ω-limit points of $\tilde{V}(0, u)$ are precisely the $2^n - 1$ zeros of $\tilde{V}(0, u)$. The same is true for α-limits as well. If we let t be negative then $x_i(t)$ becomes infinite at $t = 1/x_i^0$, thus the projection of this orbit to the unit sphere kills any coordinate x_j with $x_j^0 x_i^0$. If the maximum of the coordinates of u^0 is achieved by j components then the α-limit of the orbit $u(t)$ through u^0 has the corresponding j components equal to $1/\sqrt{j}$ and the rest of the components are zero. This establishes the following proposition:

PROPOSITION 5.1. *The only α- and ω-limit points of the vector field $\tilde{V}(0, u)$ on S_+^{n-1} are the zeros.*

We have

$$D\tilde{V}(0, u) = \begin{pmatrix} -u^T u^2 & 0 \\ F_3(u) - [u^T F_3(u)]u & D_u(-u^2 + [u^T u^2]u) \end{pmatrix}.$$

To understand the stability properties of the zeros on the sphere, we calculate the eigenvalues of $D\tilde{V}(0, u)$. Since it is lower triangular, the eigenvalues of $D\tilde{V}(0, u)$ are of two kinds: (i) the number $-u^T u^2$, corresponding to σ, and (ii) the eigenvalues of the matrix

$$D_u(-u^2 + [u^T u^2]u)$$
$$= \begin{pmatrix} -2u_1 + \Sigma u_i^3 + 3u_1^3 & 3u_1 u_2^2 & \cdots & 3u_1 u_n^2 \\ 3u_2 u_1^2 & -2u_2 + \Sigma u_i^3 + 3u_2^3 & \cdots & 3u_2 u_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ 3u_n u_1^2 & 3u_n u_2^2 & \cdots & -2u_n + \Sigma u_i^3 + 3u_n^3 \end{pmatrix},$$

corresponding to u. The first eigenvalue is then

$$-u^T u^2 = -\Sigma u_i^3 = \frac{1}{\sqrt{j}}.$$

The other eigenvalues are those of the operator $D_u(-u^2 + (\Sigma u_i^3)u)$ defined on the tangent space to the sphere. Suppose that $u \in S^{n-1}$ is such that $u_k = \gamma$ for all k such that $u_k \neq 0$. Then for every v tangent to the sphere at $(0, u)$,

$$\Sigma(3u_i^2 v_i)u = 3\gamma\Sigma(u_i v_i)u = 0.$$

Thus for an eigenvector v

$$(D_u\tilde{V})v = \begin{pmatrix} \pm\frac{1}{\sqrt{j}} & & \\ & \ddots & \\ & & \pm\frac{1}{\sqrt{j}} \end{pmatrix} v,$$

where the sign is positive if the corresponding component of u is zero.

Thus, $D_u\tilde{V}$ has a component repelling from each facet of the positive orthant in which u lies. Each vertex of S_+^{n-1} is a source. Each zero of $D_u\tilde{V}$ that lies on an edge has one stable eigenvalue and the corresponding eigenvector is tangent to that edge. Each zero that lies on a two-dimensional face has two stable eigenvalues and their corresponding eigenvectors are tangent to that face, and so on.

For each zero u' of $\tilde{V}(0, u)$, define $W^u(u')$ as the set of those points u whose α-limit is equal to u', and define $W^s(u')$ as the set of those points u whose ω-limit is equal to u'. Note that $W^s(u')$ is the interior of the face in which u' lies. We now define a pre-order on the set of zeros of $\tilde{V}(0, u)$. We write $u' > u''$ if there is a nonstationary orbit whose α-limit is u' and whose ω-limit is u''. This pre-order has no cycles since the dimension of the set $W^s(u')$ is strictly increasing along a chain in the pre-order. For any fixed time t, let $\tilde{\phi}_t(u)$ denote the point on the orbit at time t, assuming it starts at u at time 0. The transformation $\tilde{\phi}_t$ is called the *time t map* of the flow. For the proof of the following proposition the reader is referred to Chapter 2 in [Sh]:

PROPOSITION 5.2. *There is a time $t_0 > 0$ and compact sets*

$$\varnothing = M_0 \subset M_1 \subset \cdots \subset M_{2^n-1} = S_+^{n-1}$$

such that

(i) *For every i, M_i is the closure of its interior.*
(ii) *The difference $M_i \setminus M_{i-1}$ contains one zero, denoted z^i, of $\tilde{V}(0, u)$.*
(iii) *The image $\tilde{\phi}_{t_0}(M_i)$ is contained in the interior of the set M_i.*
(iv) *The intersection of the iterates $\tilde{\phi}_{t_0}^q(M_i)$ (that is, q applications of $\tilde{\phi}$), for $q \geqslant 0$, is equal to the union of the sets $W^u(z^j)$ over all $j \leqslant i$.*

The construction described in Proposition 5.2 is called a *filtration*. We are now ready for the proof of the following proposition.

PROPOSITION 5.3. *Suppose $\dot{x} = V(x) = -x^2 + o(\|x\|^2)$ is a real analytic vector field defined on a neighborhood of the origin in R^n. Suppose that for every $x \geqslant 0$ and every i such that $x_i = 0$, also $(V(x))_i = 0$. Under these conditions, there exists an $\epsilon > 0$ such that if $x^0 > 0$ and $\|x^0\| < \epsilon$ then the solution curve $\phi(t) = \phi_{x^0}(t)$ of the equation $\dot{x} = V(x)$ is defined for all nonnegative values of t. Moreover, as t tends to infinity, $\phi(t)$ tends to the origin tangent to the line $\{x_1 = \cdots = x_n\}$.*

PROOF. If $x > 0$ is sufficiently close to the origin then $d\sigma/dt < 0$. This implies that $\phi_{x^0}(t)$ is defined for all nonnegative t and $\phi_{x^0}(t) \to 0$. Consider the vector field $\tilde{V}$ and the corresponding $\tilde{\phi}_{x^0}(t)$. The filtration described above can be fattened to a filtration of a neighborhood of $\{0\} \times S_+^{n-1}$ in $[0, t] \times S_+^{n-1}$ since $d\sigma/dt < 0$. Thus every point tends to a zero. The stable sets of zeros in the boundary stay in the boundary since the boundary is invariant. Thus the orbit of any interior point tends to the point $(0, (1/\sqrt{n})e)$. It does so with a definite limiting direction (see [H] on C^1 linearization for contractions). This implies that the projected curve in the x-variable is tangent to the ray through e at the origin.

Note that throughout this section we used differentiability only up to second order. In the context of linear programming, Proposition 5.3 translates to the following:

PROPOSITION 5.4. *Given a nondegenerate linear programming problem in standard form, suppose we express the linear rescaling search direction vector field ξ_l in terms of the nonbasic variables at the optimal vertex as in Proposition 4.1. Then any interior solution curve is tangent to the vector*

$$\left(\frac{1}{\tilde{c}_{m+1}}, \ldots, \frac{1}{\tilde{c}_n}\right)$$

at the origin where the vector $(\tilde{c}_{m+1}, \ldots, \tilde{c}_n)$ is the reduced cost vector.

The discrete analog of this fact was observed experimentally by Earl Barnes. Subsequent to this analysis Megiddo [Me2] found different behavior for a class of differential equations related to the barrier method.

6. Differentiability of the Newton barrier function method. We continue to consider the problem in standard form (SF) and represent the new point given by the algorithm at a point x,

$$x' = x + \alpha(x)V(x).$$

We say that the system (A, b) is nondegenerate if for every x such that $Ax = b$, the submatrix of A, consisting of the columns with indices j for which $x_j \neq 0$, has rank m. The feasible polyhedron P is the set of all the solutions of the system $\{Ax = b, x \geqslant 0\}$. We denote the interior of P by $\dot{P}$.

PROPOSITION 6.1. *For a nondegenerate system (A, b), the matrix $(AD_x^2A^T)^{-1}$ constitutes a well-defined real analytic mapping from the affine flat $Ax = b$ into $R^{m\times m}$.*

PROOF. The mapping that takes a nonsingular matrix to its inverse is real analytic by Cramer's rule. Thus, we need only show that the matrix $AD_x^2A^T$ is invertible at $x \in P$ even when $x_j = 0$ for some j's. Suppose, without loss of generality, that $x_1, \dots, x_p \neq 0$ and $x_{p+1} = \cdots = x_n = 0$ $(m \leqslant p \leqslant n)$. Write $A = (B, N)$, where $B \in R^{m\times p}$ and $N \in R^{m\times(n-p)}$. Let $\check{x} = (x_1, \dots, x_p)$. Since

$$D_x^2 = \begin{pmatrix} D_{\check{x}}^2 & O \\ O & O \end{pmatrix},$$

it follows that

$$AD_x^2A^T = BD_{\check{x}}^2B^T$$

which is invertible since BD_x has maximal rank by the nondegeneracy assumption. Obviously,

$$\left(AD_x^2A^T\right)^{-1} = \left(BD_{\check{x}}^2B^T\right)^{-1}.$$

This completes the proof.

We now recall that the Newton barrier vector field corresponding to a fixed value of μ is

$$V_\mu(x) = \mu v_\mu = D_x r_\mu = D_x\left(I - D_xA^T\left(AD_x^2A^T\right)^{-1}AD_x\right)(D_xc - \mu e).$$

PROPOSITION 6.2. *For a nondegenerate system (A, b),*

(i) *The Newton barrier vector field $V_\mu(x)$ is well defined for every $x \in P$ and $\mu > 0$ and, moreover, at every such point it is real analytic.*

(ii) *If x is on a face Φ of the polytope P then the vector $V_\mu(x)$ is tangent to Φ. In particular, if x is a vertex then $V_\mu(x) = 0$.*

(iii) *If x is on a face Φ then $V_\mu(x)$ coincides with the Newton barrier vector field (with the same μ) which is associated with the restricted problem on the face*:

$$\text{Minimize} \quad c^Tx$$
$$\text{subject to} \quad x \in \Phi.$$

PROOF. Claim (i) is obvious in light of Proposition 6.1 and the formula for V_μ. For claims (ii) and (iii), suppose (without loss of generality), as in the proof of Proposition 6.1, that $x_1, \dots, x_p \neq 0$ and $x_{p+1} = \cdots = x_n = 0$ $(m \leqslant p \leqslant n)$, and let $\check{x}$ also be as there. We have

$$AD_x = (BD_{\check{x}} \quad O).$$

Denote $\check{c} = (c_1, \dots, c_p)^T$ and $\check{e} = (e_1, \dots, e_p)^T$. We now have

$$D_x c = \begin{pmatrix} D_{\check{x}}\check{c} \\ O \end{pmatrix} \quad \text{and} \quad D_x \mu e = \begin{pmatrix} D_{\check{x}}\mu\check{e} \\ O \end{pmatrix}.$$

On the other hand,

$$AD_x^2 A^T = BD_{\check{x}}^2 B^T.$$

Substituting the right-hand sides of these equalities into the formula for V_μ we prove (ii) and (iii).

REMARK 6.3. The nondegeneracy hypothesis implies that at a vertex x of the polytope the matrix B is invertible. Thus

$$\left(BD_{\check{x}}^2 B^T\right)^{-1} = B^{-T} D_{\check{x}}^{-2} B^{-1}$$

and the matrix

$$D_x\left(I - D_x A^T\left(AD_x^2 A^T\right)^{-1} AD_x\right)$$

is the zero matrix.

We now compute the derivative of V_μ at a vertex x.

LEMMA 6.4. *Let*

$$M_x = D_x\left(I - D_x A^T\left(AD_x^2 A^T\right)^{-1} AD_x\right).$$

Then

$$M_x'(h) = D_h - 2M_x D_h A^T\left(AD_x^2 A^T\right)^{-1} AD_x - D_x^2 A^T\left(AD_x^2 A^T\right)^{-1} AD_h.$$

PROOF.

$$\begin{aligned} M_x'(h) = {} & D_h\left(I - D_x A^T\left(AD_x^2 A^T\right)^{-1} AD_x\right) \\ & + D_x\left(-D_h A^T\left(AD_x^2 A^T\right)^{-1} AD_x - D_x A^T\left(AD_x^2 A^T\right)^{-1} AD_h\right) \\ & + D_x\left(-D_x A^T\left(\left(AD_x^2 A^T\right)^{-1}\right)'(h) AD_x\right). \end{aligned}$$

Since $D_x D_h = D_h D_x$, we have

$$\begin{aligned} M_x'(h) = {} & D_h - 2D_x D_h A^T\left(AD_x^2 A^T\right)^{-1} AD_x \\ & + 2D_x^2 A^T\left(AD_x^2 A^T\right)^{-1} AD_x D_h A^T\left(AD_x^2 A^T\right)^{-1} AD_x - D_x^2 A^T\left(AD_x^2 A^T\right)^{-1} AD_h \\ = {} & D_h - 2D_x\left(I - D_x A^T\left(AD_x^2 A^T\right)^{-1} AD_x\right) D_h A^T\left(AD_x^2 A^T\right)^{-1} AD_x \\ & - D_x^2 A^T\left(AD_x^2 A^T\right)^{-1} AD_h. \end{aligned}$$

PROPOSITION 6.5. *At a vertex x, $V_\mu'(x) = -\mu I$.*

PROOF.

$$V_\mu' = \left(M_x(D_x c - \mu e)\right)'(h) = M_x'(h)(D_x c - \mu e) + M_x D_h c.$$

Since $M_x = 0$ (see Remark 6.3), and h is tangent to the polytope (so $Ah = 0$), it follows that

$$\begin{aligned} V_x'(h) &= D_h(D_x c - \mu e) - D_x^2 A^T\left(AD_x^2A^T\right)^{-1} AD_h(D_x c - \mu e) \\ &= -\mu h + M_x D_h c = -\mu h. \end{aligned}$$

7. **Differentiability of the projective rescaling vector field.** In this section we develop results analogous to those of the preceding sections. We work in Karmarkar's standard form (KSF). We assume nondegeneracy of the matrix A (not the entire matrix $\tilde{A}$ of the linear system of constraints) in the sense that for every x in the affine subspace $L = \{Ax = 0, e^T x = 1\}$, the submatrix of A, consisting of the columns with indices j for which $x_j \neq 0$, has rank m. The polyhedron P is the set of all the solutions of the system $\{Ax = 0, e^T x = 1, x \geqslant 0\}$. The following proposition is essentially the same as Proposition 6.1.

PROPOSITION 7.1. *For a nondegenerate problem, the matrix $(AD_x^2A^T)^{-1}$ constitutes a well-defined real analytic mapping from the affine flat L into $R^{m\times m}$.*

Recall that

$$\xi_p = \xi_p(x) = [D - xx^T]\left[I - DA^T(AD^2A^T)^{-1}AD\right]Dc.$$

Analogously, we have

PROPOSITION 7.2. *For a nondegenerate problem (A, b),*

(i) *The direction ξ_p is well defined for every $x \in P$ and, moreover, at every such point it is real analytic.*

(ii) *If x is on a face Φ of the polytope P then the vector ξ_p is tangent to Φ. In particular, if x is a vertex then $\xi_p(x) = 0$.*

(iii) *If x is on a face Φ then $\xi_p(x)$ coincides with the vector ξ_p which is associated with the restricted problem on the face*:

$$\begin{aligned} &\textit{Minimize} \quad c^T x \\ &\textit{subject to} \quad x \in \Phi. \end{aligned}$$

REMARK 7.3. Recall that the vector field $\xi_p(x)$ is well defined even without the assumption that the optimal value of the linear objective function equals zero. Thus, the ξ_p vector for the restricted problem on the face is defined this way and the restricted problem is not transformed into the form KSF with optimal value zero.

Let

$$\tau = \tau(x) = \left[I - DA^T(AD^2A^T)^{-1}AD\right]Dc.$$

Thus,

$$\xi_p = D\tau(x) - (x^T\tau(x))x.$$

Notice that the logarithmic barrier vector field $V_\mu(x)$ is well defined even when μ is negative. The following proposition was first pointed out in [GMSTW].

PROPOSITION 7.4. *If the barrier parameter μ is chosen as a function of the point x, $\mu(x) = x^T\tau(x)$, then $\xi_p(x) = V_{\mu(x)}(x)$.*

PROOF. If $\mu(x) = x^T\tau(x)$ then (see Remark 7.3)

$$V_\mu(x) = D\tau(x) - \mu De = D\tau(x) - \mu x = \xi_p(x).$$

PROPOSITION 7.5. *For a nondegenerate problem, at any vertex x,*

$$\frac{d\xi_p(x)}{dx} = -\mu(x)I.$$

PROOF. Since $V_\mu(x)$ is differentiable in (x, μ), the vector $V_{\mu(x)}(x)$ is differentiable and

$$\frac{d\xi_p(x)}{dx} = \frac{\partial V}{\partial \mu}\frac{d\mu}{dx} + \frac{\partial V_\mu}{\partial x}.$$

Since $V_\mu(x)$ is identically zero (as a function of μ) at any vertex,

$$\frac{\partial V_\mu}{\partial \mu} = 0 \quad \text{and} \quad \frac{d\xi_p}{dx} = \frac{\partial V_\mu}{\partial x} = -\mu(x)I$$

by Proposition 6.5.

8. The discrete version of the linear rescaling algorithm. In this section we consider a specific choice of a step size in the linear rescaling algorithm (as in [Bar]). Given an interior point x, the algorithm determines a new point $X_l(x)$ as follows

$$X(x) = x - \rho\frac{\xi_l(x)}{\|D_x^{-1}\xi_l(x)\|}$$

where $0 < \rho < 1$ is a constant. The choice of ρ guarantees that $X(x)$ is in the interior of the polytope (see [Bar]). It has been proven [Bar, VMF] that for nondegenerate problems, for any interior point x, $X^q(x)$ converges to the optimal solution. In this section we study the asymptotic behavior and extensions to the boundary of this discrete algorithm. Nondegeneracy in this section means (i) for every feasible solution x, the submatrix of the matrix A, consisting of the columns corresponding to nonzero coordinates, is of rank m, and (ii) every face of the feasible polyhedron has a unique optimum with respect to the objective function vector c.

LEMMA 8.1. *Suppose the problem is nondegenerate. Let $\{x^k\}$ be a sequence of interior points converging to a point $\bar{x}$ on the boundary of P. Let J denote the set of j's with $\bar{x}_j \neq 0$. Under these conditions, for every $j \notin J$, the ratio $\xi_l(x^k)_j/(x_j^k)^2$ converges* *to a finite limit. In particular, for every $j \notin J$,*

$$\lim_{k \to \infty} \frac{\xi_l(x^k)_j}{x_j^k} = 0.$$

PROOF. Recall that

$$\xi_l(x) = D_x^2\left(I - A^T\left(AD_{x^k}^2 A^T\right)^{-1} AD_{x^k}^2\right)c.$$

Thus,

$$\frac{\xi_l(x^k)_j}{\left(x_j^k\right)^2} = \left(c - A^T\left(AD_{x^k}^2 A^T\right)^{-1} AD_{x^k}^2 c\right)_j$$

and the lemma follows from the fact that the matrix $(AD_{x^k}^2 A^T)^{-1}$ extends continuously to the closed polytope.

Let's denote

$$V_l(x) = \frac{\xi_l(x)}{\|D_x^{-1}\xi_l(x)\|}.$$

THEOREM 8.2. *Suppose the problem is nondegenerate. Then*

(i) *The vector field $V_l(x)$ extends continuously to the boundary of the polytope and real analytically at any point which is not a vertex.*

(ii) *On any face Φ of the polytope P, the vector field V_l coincides with the vector field associated with the restricted problem on the face Φ. In particular, $V_l(x)$ is tangent to each face that contains the point x and vanishes at vertices.*

(iii) *The iteration $x' = X(x) = x - \rho V_l(x)$ extends continuously to the boundary of the polytope.*

(iv) *The iteration X takes any face of the polytope into itself and is in fact the iteration of the problem restricted to the face.*

(v) *If x lies in the face Φ then, as q tends to infinity, $X^q(x)$ converges to the minimum of the linear objective function relative to the face Φ.*

PROOF. By Proposition 2.5, ξ_l extends continuously to the closed polytope and vanishes only at vertices (see §2). By Lemma 8.1 $\|D_x^{-1}\xi_l\|$ has the appropriate limiting value on any face. This establishes (i)–(iv) except at vertices. The difference in the values of the objective function is

$$c^T(X(x) - x) = -\frac{\rho}{\|D_x^{-1}\xi_l\|}(\xi_l(x))^T\xi_l(x),$$

and this is negative provided x is not a vertex. Therefore, the only possible accumulation points of iterates are vertices. The remaining claims will follow from local analysis of the linear rescaling algorithm at vertices which is discussed below. It is convenient to represent the points in the neighborhood of a nondegenerate vertex in terms of their nonbasic components at a nondegenerate vertex. Let us also use N to denote the set of indices of nonbasic variables. If $j \in N$, that is, $x_j = 0$ at the vertex then by Proposition 4.1

$$(X(x))_j = x_j - \rho\frac{\tilde{c}_j x_j^2 + O\left(\|x_N\|^4\right)}{\sqrt{\sum_{i \in N} \tilde{c}_i^2 x_i^2 + O\left(\|x_N\|^4\right)}}.$$

Continuity at vertices follows from this formula. It is clear that if x is in the relative interior of a face Φ and the iterates converge to a vertex of Φ then all the $\tilde{c}_i$ $(i \in N)$ must be positive (assuming nondegeneracy). Thus, the vertex is a minimum relative to Φ.

To study the asymptotic behavior of iterates near a nondegenerate optimal vertex, we express the mapping $X(x)$ in terms of the nonbasic variables alone. Recall that x_N denotes the restriction of the vector x to the nonbasic variables. For every $j \in N$, we have

$$(X_n(x_N))_j = x_j - \rho \frac{\tilde{c}_j x_j^2 + O(\|x_N\|^4)}{\sqrt{\sum_{i \in N} \tilde{c}_i^2 x_i^2 + O(\|x_N\|^4)}}.$$

Since the basic variables are affine functions of the nonbasic ones, we are reduced to studying the asymptotic behavior of X_N near the origin. We first change variables. Let $y_i = \tilde{c}_i x_i$ for $i \in N$, and let $y = (y_i)_{i \in N}$.

PROPOSITION 8.3. *The change of variables $y_i = \tilde{c}_i x_i$ conjugates X_N to $\tilde{X}_N$ where for $j \in N$,*

$$(\tilde{X}_N(y))_j = y_j - \rho y_j^2 \frac{1 + O(\|y\|^2)}{\sqrt{\sum_{i \in N} y_i^2 + O(\|y\|^4)}}$$

in a neighborhood of the origin.

The mapping $\tilde{X}_N$ is not differentiable at the origin, but its directional derivatives along rays exist. As in the case of the vector field (see §5), it is now convenient to study $\tilde{X}_N$ in "polar coordinates". For convenience, assume without loss of generality that $N = \{1, \ldots, n - m\}$ so $\tilde{X}_N$ is a mapping from a neighborhood U of the origin in R^{n-m} into R^{n-m}. A vector $y \in R^{n-m}$ can be expressed as a product $y = \sigma u$ where $\sigma \geqslant 0$ is a scalar and $u \in R^{n-m}$ is a unit vector. Of course, σ is just $\|y\|$ and if $y \neq 0$ then $u = y/\|y\|$. If $y \geqslant 0$ then $u \in S_+^{n-m-1}$. In these polar coordinates $\tilde{X}_N$ is expressed as

$$\begin{aligned} W(\sigma, u) = W_\rho(\sigma, u) &= \left(\|\tilde{X}_N(\sigma u)\|, \frac{1}{\|\tilde{X}_N(\sigma u)\|} \tilde{X}_N(\sigma u) \right) \\ &= \Bigg(\sigma \|u - \rho(1 + O(\sigma^2)) u^2\|, \\ &\qquad \frac{1}{\|u - \rho(1 + O(\sigma^2)) u^2\|} \left[u - \rho(1 + O(\sigma^2)) u^2 \right] \Bigg) \end{aligned}$$

where $u^2 = (u_1^2, \ldots, u_{n-m}^2)^T$ and σ is sufficiently small. Let

$$Z \colon R_+ \times S^{n-m-1} \to R_+ \times S^{n-m-1}$$

be defined by

$$Z(\sigma, u) = \left(\sigma \|u - \rho u^2\|, \frac{1}{\|u - \rho u^2\|} \left[u - \rho u^2 \right] \right).$$

PROPOSITION 8.4. *For any ρ, $0 < \rho < 1$, there is an $r > 0$ such that W_ρ is defined on $[0, r] \times S_+^{n-m-1}$ and*

$$W_\rho\left([0, r] \times S_+^{n-m-1}\right) \subset [0, r] \times S_+^{n-m-1}.$$

Moreover, the function values and the derivative of W and Z coincide at $(0, u)$, that is, $Z(0, u) = W(0, u)$ and $Z'(0, u) = W'(0, u)$.

PROOF. All the assertions follow from the expression for W above and the chain rule for differentiation.

The map Z seems simple as it maps rays to rays. The ray determined by a unit vector u is contracted by the constant factor $\|u - \rho u^2\|$. This contraction constant is minimized at the unit vectors e^i $(i = 1, \dots, n - m)$ and maximized at $e/\sqrt{n - m}$ where its value is $1 - \rho/\sqrt{n - m}$. We will show that $(0, e/\sqrt{n - m})$ is an attractor for both Z and W. However, we do not know yet the precise domain of attraction of this point even for Z. Let $G\colon S_+^{m-1} \to S_+^{m-1}$ be defined by

$$G(u) = \frac{1}{\|u - \rho u^2\|}(u - \rho u^2)$$

so G is just the second coordinate of Z.

It is not known whether every interior point of S_+^{m-1} tends to $e/\sqrt{n - m}$ under the iteration of G.

PROPOSITION 8.5.

$$Z'\left(0, \frac{1}{\sqrt{n - m}} e\right) = \begin{pmatrix} 1 - \dfrac{\rho}{\sqrt{n - m}} & O \\ O & \left(1 - \dfrac{\rho}{\sqrt{n - m - \rho}}\right) I \end{pmatrix}$$

and hence $(0, e/\sqrt{n - m})$ is an attractor for both Z and W.

PROOF.

$$Z'\left(0, \frac{1}{\sqrt{n - m}} e\right) = \begin{pmatrix} \sigma'\left(0, \dfrac{1}{\sqrt{n - m}} e\right) & O \\ O & G'\left(\dfrac{1}{\sqrt{n - m}} e\right) \end{pmatrix}$$

$$= \begin{pmatrix} 1 - \dfrac{\rho}{\sqrt{n - m}} & O \\ O & G'\left(\dfrac{1}{\sqrt{n - m}} e\right) \end{pmatrix} \quad \text{and}$$

$$G'(u)v = \frac{1}{\|u - \rho u^2\|}(I - 2\rho D_u)v - \frac{[(I - 2\rho D_u)v]^T[u - \rho u^2]}{\|u - \rho u^2\|^2}(u - \rho u^2).$$

Thus, at $e/\sqrt{n - m}$ with w tangent to S_+^{n-m-1},

$$G'\left(\frac{1}{\sqrt{n - m}} e\right)(w) = \left(1 - \frac{\rho}{\sqrt{n - m - \rho}}\right) w.$$

PROPOSITION 8.6. *There are a neighborhood U_1 of the origin in R_+^{n-m} and a neighborhood $U_2 \subset U_1$ of the intersection of U_1 with the line $\{\tilde{c}_1 x_1 = \cdots = \tilde{c}_{n-m} x_{n-m}\}$ such that*

(i) *The set U_2 contains a definite angle at the origin.*

(ii) *For every $x \in U_2$,*

$$\lim_{q \to \infty} \frac{\tilde{c}_j(X_N^q(x))_j}{\tilde{c}_i(X_N^q(x))_i} = 1$$

for all i and j.

(iii) *There exist constants $K_1, K_2 > 0$ such that*

$$K_1\left(1 - \frac{\rho}{\sqrt{n-m}}\right)^q \leqslant \|X_N^q(x)\| \leqslant K_2\left(1 - \frac{\rho}{\sqrt{n-m}}\right)^q$$

for all x and $q > 0$.

PROOF. We need only prove the comparable facts for $\tilde{X}_N^q(y)$ and a definite angular wedge around the diagonal $\{y_1 = \cdots = y_{n-m}\}$. Let U_3 be a neighborhood of $(0, e/\sqrt{n-m})$ consisting of points attracted to $(0, e/\sqrt{n-m})$ under W, that is, for $v \in U_3$,

$$\lim_{q \to \infty} W^q(v) = \left(0, \frac{1}{\sqrt{n-m}} e\right).$$

The set of points in R^{n-m} corresponding to U_3 contains an angular wedge about a small piece of the line $\{y_1 = \cdots = y_{n-m}\}$. Moreover, since the contraction rates given by the eigenvalues of $W'(0, e/\sqrt{n-m})$ are stronger along the sphere $\{0\} \times S^{n-m-1}$ than along the line through $e/\sqrt{n-m}$, any orbit $W^q(v)$ becomes tangent to the ray and the asymptotic rate of convergence to zero is the rate along the ray. This type of argument can be found in center-unstable manifold theory in [Sh].

Given a point x in the interior of the polytope and an optimal point x^*, let

$$a_l(x) = \limsup_{q \to \infty} \frac{1}{q} \log \|X^q(x) - x^*\|$$

be the asymptotic rate of convergence to the optimum. We have shown

COROLLARY 8.7. *For a nondegenerate problem,*

$$a_l = \log\left(1 - \frac{\rho}{\sqrt{n-m}}\right)$$

for a nonempty open set interior to the polytope.

Barnes [Bar] shows that

$$a_l \leqslant \log\left(1 - \frac{\rho}{\sqrt{n-m}}\right)$$

for any interior point x. It is still an open question whether

$$a_l = \log\left(1 - \frac{\rho}{\sqrt{n-m}}\right)$$

for all x interior to the polytope.

9. The discrete version of the projective rescaling algorithm. In this section we analyze the boundary behavior of the discrete iteration of the projective rescaling algorithm. The linear programming problem is considered in Karmarkar's standard form (KSF):

$$(KSF) \qquad \begin{aligned} \text{Minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = 0 \\ & e^T x = 1, \\ & x \geqslant 0, \end{aligned}$$

where $c, x \in R^n$, $A \in R^{m \times n}$ and $e = (1, \dots, 1)^T \in R^n$. Let

$$\Omega = \{x \in R^n : Ax = 0\},$$

$$S = \{x \in R^n : e^T x = 1, x \geqslant 0\} \quad \text{and}$$

$$P = \Omega \cap S.$$

We denote the interior of the feasible domain by $\dot{P}$, that is, $\dot{P} = P \cap \{x \in R^n : x > 0\}$. The algorithm assigns to any point $x \in \dot{P}$ a new point $Y(x)$ defined as follows. Recall the matrix $\bar{A}$ from §1 and the vector

$$\eta_p(x) = \left[I - \bar{A}^T \left(\bar{A}\bar{A}^T\right)^{-1} \bar{A}\right] D_x c.$$

For simplicity let us denote $\eta(x) = \eta_p(x)$. A unit vector, $u(x)$, in the direction of $\eta(x)$ is given by

$$u(x) = \frac{1}{\|\eta(x)\|} \eta(x).$$

The underlying projective transformation at an interior point x is the following:

$$T_x(y) = \frac{1}{e^T D_x^{-1} y} D_x^{-1} y.$$

Obviously, $T_x(x) = e/n$. In the transformed space the algorithm moves from the point e/n to the point

$$Y'(x) = \frac{1}{n} e - \gamma r u(x) \quad \text{where}$$

$$r = \frac{1}{\sqrt{n(n-1)}}$$

and γ is a constant which was originally chosen in [Kar] as $\frac{1}{4}$. In the source space the new point $Y(x)$ is equal to the inverse image of the point $Y'(x)$ under the transformation T_x:

$$Y(x) = \frac{1}{e^T D_x Y'(x)} D_x Y'(x).$$

For the sake of simplicity we replace $r = 1/\sqrt{n(n-1)}$ by $r = 1/n$ and call the resulting vector $Y(x) = Y_\gamma(x)$. It follows that

$$Y(x) = \frac{D_x\left(\frac{1}{n}e - \frac{\gamma}{n\|\eta(x)\|}\eta(x)\right)}{e^T D_x\left(\frac{1}{n}e - \frac{\gamma}{n\|\eta(x)\|}\eta(x)\right)}$$

$$= \frac{D_x\left(e - \frac{\gamma}{\|\eta(x)\|}\eta(x)\right)}{e^T D_x\left(e - \frac{\gamma}{\|\eta(x)\|}\eta(x)\right)}.$$

In this section we assume nondegeneracy in the following sense. For any feasible point x, the submatrix of A, consisting of the columns with indices j such that $x_j > 0$, has rank m. We also denote by $\tau(x)$ the orthogonal projection of the vector $D_x c$ into the nullspace of the matrix AD_x, that is,

$$\tau(x) = \left(I - D_x A^T (A D_x^2 A^T)^{-1} A D_x\right) D_x c.$$

LEMMA 9.1. *Suppose $\{x^\nu\}$ is a sequence of interior points converging to a boundary point $x^0 \in \partial P$. Let J denote the set of indices j such that $x_j^0 \neq 0$ ($J \neq N$). Under these conditions, if the problem is nondegenerate, then*

$$\lim_{\nu \to \infty} (\tau(x^\nu))_j = 0$$

for every $j \notin J$ and, moreover, the limit

$$\lim_{\nu \to \infty} \frac{(\tau(x^\nu))_j}{x_j^\nu}$$

exists for every $j \notin J$.

PROOF. First, note that by the nondegeneracy assumption the matrix $(AD_{x^\nu}^2 A^T)^{-1}$ tends to the matrix $(AD_{x^0}^2 A^T)^{-1}$ as ν tends to infinity. Now, we have

$$\tau(x^\nu) = D_{x^\nu}\left[I - A^T (A D_{x^\nu}^2 A^T)^{-1} A D_{x^\nu}^2\right] c$$

so

$$\frac{(\tau(x^\nu))_j}{x_j^\nu} = \left(\left[I - A^T (A D_{x^\nu}^2 A^T)^{-1} A D_{x^\nu}^2\right] c\right)_j.$$

The latter tends to

$$\left(\left[I - A^T\left(AD_{x^0}^2A^T\right)^{-1}AD_{x^0}^2\right]c\right)_j.$$

COROLLARY 9.2. *Suppose the problem is nondegenerate and the optimal objective function value equals zero. Under these conditions, if $x \in P$ is such that $c^Tx > 0$ then $\eta(x) \neq 0$ and hence the mapping $Y(x)$ is smooth at x.*

PROOF. Suppose x is a feasible point such that $\eta(x) = 0$. Recall that $\eta(x)$ is the orthogonal projection of the vector D_xc into the nullspace of $\bar{A}$. This implies that $\eta(x)$ is also the orthogonal projection of the vector $D_xc - (c^Tx/n)e$ into the nullspace of the matrix AD_x. Thus, there exists a vector v such that

$$D_xc - \frac{c^Tx}{n}e = (AD_x)^Tv.$$

Suppose first that $x \in \dot{P}$. In this case we have

$$c - \frac{c^Tx}{n}D_x^{-1}e = A^Tv$$

and therefore both c and $(c^Tx/n)D_x^{-1}e$ induce the same linear functional on P. The vector $D_x^{-1}e$ is positive which contradicts the assumption that the optimal value is 0. Now suppose that x is a boundary point and let J denote the set of j's such that $x_j > 0$. For every $j \notin J$ we have

$$(\eta(x))_j = \left(D_xc - \frac{c^Tx}{n}e - D_xA^Tv\right)_j = \left(-\frac{c^Tx}{n}e\right)_j = -\frac{c^Tx}{n} \neq 0$$

which is a contradiction.

LEMMA 9.3. *The vector e is orthogonal to $\eta(x)$.*

PROOF. By definitions, $\eta(x)$ is in the nullspace of e^T if x is in the interior, and hence for every x in the polytope.

For every $J \subset N = \{1, \dots, n\}$, let us denote $L_J = \{x \in R^n\colon x_j = 0 \text{ for } j \notin J\}$. Thus, L_J is a linear subspace of dimension $|J|$. The set J also determines a face of the polytope: $\Phi_J = L_J \cap P$. Let e^J denote the vector consisting of 1's in the positions corresponding to J and 0's in all the other positions. We are interested in restrictions of the linear programming problem to faces of the polytope P. Specifically, the restricted problem corresponding to the set J is the following:

$$\begin{aligned} &\text{Minimize} && c^Tx \\ &\text{subject to} && x \in \Phi_J. \end{aligned}$$

This restricted problem gives rise to a new vector field $\eta_J(x)$ on the face Φ_J by ignoring the vanishing coordinates (that is, those with indices not in J). Thus, $\eta_J(x) \in L_J$. The τ vector for the restricted problem is the same as the τ vector for the original by Proposition 7.1 and §6:

$$\eta_J(x) = \tau(x) - \frac{c^Tx}{|J|}e^{N \setminus J}, \quad \eta(x) = \tau(x) - \frac{c^Tx}{n}e^N$$

and $\eta(x)$ may be written as the sum

$$\eta(x) = \eta_J(x) + c^T x\left(\frac{1}{|J|} - \frac{1}{n}\right)e^J - \frac{c^T x}{n}e^{N \setminus J}$$

of mutually orthogonal vectors since e^J is orthogonal to $\eta_J(x)$ by Lemma 9.3 and e^J and $\eta_J(x)$ lie in L_J which is orthogonal to $e^{N \setminus J}$. So we have

LEMMA 9.4.

$$\|\eta(x)\|^2 - \|\eta_J(x)\|^2 = \left(\frac{1}{|J|} - \frac{1}{n}\right)(c^T x)^2.$$

LEMMA 9.5.

$$D_x \eta(x) = D_x \eta_J(x) + c^T x\left(\frac{1}{|J|} - \frac{1}{n}\right)x.$$

LEMMA 9.6. *If the problem (KSF) has a unique minimum, with nonnegative minimum value, then for a nonoptimal point x,*
(i) $\|\eta(x)\| \leqslant c^T x$
(ii) $\|\eta(x)\| + \|\eta_J(x)\| < 2c^T x$, *and*
(iii) $\|\eta(x)\| - \|\eta_J(x)\| > \frac{1}{2}(1/(n - |J|) - 1/n)c^T x$.

PROOF. Inequality (i) implies inequality (ii) by Lemma 9.4. The equality of Lemma 9.4 divided by the inequality (ii) implies inequality (iii). Inequality (i) was proved in [Blu] and we provide here another proof. The point

$$\frac{1}{n}e - \frac{1}{n\|\eta(x)\|}\eta(x)$$

is in the interior of the polytope defined by

$$AD_x y = 0,$$

$$e^T y = 0,$$

$$y \geqslant 0.$$

The objective function $(D_x c)^T y$ is nonnegative on this polytope with equality only possible at the optimal vertex since every point in this polytope is a positive multiple of a point in $D_x^{-1}P$. Thus we have

$$(D_x c)^T\left(\frac{1}{n}e - \frac{1}{n\|\eta(x)\|}\eta(x)\right) > 0.$$

Now, $\eta(x)$ equals the projection of the vector $D_x c$ into the intersection of the nullspaces of the matrix AD_x and the vector e^T. Thus

$$(D_x c)^T \frac{1}{n\|\eta(x)\|}\eta(x) = \frac{(\eta(x))^T \eta(x)}{n\|\eta(x)\|} = \frac{\|\eta(x)\|^2}{n\|\eta(x)\|}.$$

This implies the claim at interior points x. By continuity the claim (i) holds on the closed polytope and hence (ii) and (iii) follow.

Suppose still that the problem has a unique optimal solution. Consider the vector

$$D_x Y'(x) = \frac{1}{n}x - \frac{\gamma}{n\|\eta(x)\|} D_x \eta(x).$$

Obviously, for any real $M \neq 0$,

$$\frac{D_x Y'(x)}{e^T D_x Y'(x)} = \frac{M D_x Y'(x)}{e^T M D_x Y'(x)}.$$

It follows from Lemma 9.5 that there is a real number M such that

$$M D_x Y'(x) = x - \frac{\gamma}{\|\eta(x)\| - \gamma c^T x\left(\frac{1}{|J|} - \frac{1}{n}\right)} D_x \eta_J(x).$$

For $\gamma \leqslant \frac{1}{2}$, it follows from part (iii) of Lemma 9.6 that

$$\|\eta(x)\| - \gamma c^T x\left(\frac{1}{|J|} - \frac{1}{n}\right) > \|\eta_J(x)\| + \left(\frac{1}{2} - \gamma\right) c^T x\left(\frac{1}{|J|} - \frac{1}{n}\right) > \|\eta_J(x)\|.$$

It thus follows that the point $MD_x Y'(x)$ lies on the line segment between x and $x - (\gamma/\|\eta_J(x)\|) D_x \eta(x)$ and thus $Y(x)$ lies on the line segment between x and $Y_J(x)$, where $Y_J(x)$ is the point assigned by the algorithm when the problem is restricted to the face Φ_J.

Let us denote by $Y^q(x)$ the transformation resulting from q iterations of Y, that is, $Y^1(x) = Y(x)$ and $Y^{q+1}(x) = Y(Y^q(x))$.

Given a face Φ_J that does not contain the global optimum, let A' denote the submatrix of A consisting of the columns j such that $j \in J$, and similarly let c', x' and e' denote the corresponding subvectors of c, x and e, respectively. The problem (KSF) restricted to the face Φ_J is the following:

$$(KSF_J) \qquad \begin{array}{ll} \text{Minimize} & (c')^T x' \\ \text{subject to} & A'x' = 0, \\ & (e')^T x' = 1, \\ & x' \geqslant 0. \end{array}$$

The vector $\eta(x)$ defined above for problem (KSF) is well defined for the problem (KSF_J), where we denote it by $\eta_J(x)$. We associate with the face Φ_J a *reduced potential function*

$$\psi_J(x) = |J| \ln c^T x - \sum_{j \in J} \ln x_j,$$

defined only for interior points of the face. We denote the interior of the face Φ_J by $\mathring{\Phi}_J$.

THEOREM 9.7. *Suppose the linear programming problem (KSF) is nondegenerate and the optimal objective function value is* 0. *Under these conditions*

(i) *The transformation $Y(x)$ extends continuously to the boundary of P, leaving each face invariant.*

(ii) *If x lies on a face Φ_J then $Y(x)$ lies on the line segment between x and the point $Y_J(x)$ which is assigned by the algorithm when the problem is restricted to the face Φ_J.*

(iii) *For every $x \in P$, the limit $\lambda(x) = \lim_{q\to\infty} Y^q(x)$ exists. Moreover, if Φ_J is the smallest face that contains x then $\lambda(x)$ is precisely the minimum of the reduced potential function with respect to the face Φ_J.*

(iv) *The mapping $Y(x)$ is smooth at every $x \in P$ except, perhaps, at the optimal vertex of (KSF).*

(v) *Every nonoptimal vertex is a local repeller.*

PROOF. The continuity of $Y(x)$ at the optimal vertex follows from the convergence of Karmarkar's algorithm. It remains to analyze the vertex behavior and the iterates $Y^q(x)$. Let x be a nonoptimal vertex. Let

$$\phi(x) = \frac{\gamma}{\|\eta(x)\| - \gamma x^T\eta(x)}.$$

Since $e^Tx = 1$ and $x \in R^n$, it follows that $\|x\| \leqslant 1$ and hence by the Cauchy-Schwartz inequality that $|x^T\eta(x)| \leqslant \|\eta(x)\|$ so $\phi(x)$ is well defined. Moreover, by Corollary 9.2, $\phi(x)$ is positive and differentiable away from the optimal vertex. Thus,

$$Y(x) = x - \phi(x)\xi_p(x).$$

Consider the derivative dY/dx. Since $\xi_p(x) = 0$ at any vertex, we have

$$\frac{dY(x)}{dx} = I - \frac{d\phi(x)}{dx}\xi_p(x) - \phi(x)\frac{d\xi_p(x)}{dx} = I - \phi(x)\frac{d\xi_p(x)}{dx}.$$

By Proposition 7.5, for any vertex x,

$$\frac{d\xi_p(x)}{dx} = V_\mu'(x) = -\mu(x)I.$$

By Lemma 9.8 below, $\mu(x) > 0$ at a vertex. Thus,

$$\frac{dY(x)}{dx} = (1 + \phi(x)\mu(x))I$$

and hence x is a repeller (a source; see [Sh]).

LEMMA 9.8. *If $x \in P$ is such that $\xi_p(x) = 0$ then $\mu(x) = x^T\tau(x) \geqslant 0$ with equality holding only at the optimal vertex.*

PROOF. We have from

$$0 = \xi_p(x) = D_x\tau(x) - (c^T\tau(x))x$$

that

$$D_x\tau(x) = \mu(x)x.$$

If $x_j = 0$, by Lemma 9.1, $(\tau(x))_j = 0$. Thus, all the nonzero components of $\tau(x)$ have the same sign as $\mu(x)$. by Lemma 9.3,

$$0 = e^T\eta(x) = e^T\tau(x) - c^Tx$$

so $e^T\tau(x) = c^Tx \geqslant 0$ and hence $\mu(x) \geqslant 0$.

We now return to an analysis of the iterates $Y^q(x)$. In the following lemma we consider faces of the polytope P which do not contain the optimal vertex.

LEMMA 9.9. *The zeros of the vector field $\eta_J(x)$ in $\dot{\Phi}_J$ are zeros of the gradient vector field of the reduced potential function $\psi_J(x)$ in $\dot{\Phi}_J$.*

PROOF. As in Corollary 9.2, $\eta_J(x) = 0$ if and only if the vector

$$c' - \frac{(c')^Tx'}{|J|} D_{x'}^{-1}e'$$

is orthogonal to the nullspace of A'. The gradient of the reduced potential function is

$$\nabla\left(|J|\ln(c')^Tx' - \sum_{j\in J} \ln x_j\right) = \frac{|J|}{(c')^Tx'}c' - D_{x'}^{-1}e',$$

so the gradient is zero if this vector is orthogonal to Φ_J. Since the two vectors are multiplies of each other, they are simultaneously orthogonal to Φ_J.

Now it is not hard to see (as in [Kar]) that if $\eta_J(x) \neq 0$, then the value of the potential function at any point $y \neq x$ in the line segment between x and $Y_J(x)$ is strictly less than its value at x, i.e., $\psi_J(y) < \psi_J(x)$. This is true because Karmarkar's proof is valid for all constants strictly between 0 and γ and these points generate the line segment. Suppose $x \in \dot{\Phi}_J$ where Φ_J does not contain the minimizing vertex, so $c^Tx > 0$ on Φ_J. The function $\psi_J(Y_J^q(x))$ decreases in value and by compactness the sequence $\{Y_J^q(x)\}$ has limit points in $\dot{\Phi}_J$. Any limit point must be a zero of η_J; for if $q_i \to \infty$ and $Y_J^{q_i}(x) \to x^0$ and $\eta_J(x^0) \neq 0$ then $\psi_J(Y(x^0)) < \psi_J(x^0)$. For q_i sufficiently large $\psi_J(Y_J^{q_i+1}(x)) < \psi_J(x^0)$, but then the subsequent iterates of $Y_J^q(x)$ $(q > q_i)$ cannot return to a small neighborhood of x^0 where ψ takes on values greater than $\psi_J(Y_J^{q_i+1}(x))$. Since ψ_J has only one critical point in $\dot{\Phi}_J$, any x in $\dot{\Phi}_J$ tends to this point. If Φ_J contains the minimizing vertex, $\dot{\Phi}_J$ has no critical points of ψ_J and by a similar argument as above $Y_J^q(x)$ tends to the minimizing vertex as q tends to infinity.

Appendix A. The linear rescaling algorithm applied to the hypercube. Consider the general linear programming problem on the unit cube

$$\text{Maximize} \quad \sum_j^n c_jx_j$$

$$\text{subject to} \quad 0 \leqslant x_j \leqslant 1 \qquad (j = 1, \dots, n).$$

Let $c = (c_1, \dots, c_n, 0, 0, \dots, 0)^T \in R^{2n}$. The standard form of the problem is to maximize c^Tx $(x \in R^{2n})$ subject to $Ax = e$, where the underlying matrix is

$$A = \begin{pmatrix} 1 & & & 1 & & \\ & \ddots & & & \ddots & \\ & & 1 & & & 1 \end{pmatrix}$$

and $e = (1, \dots, 1) \in R^n$. Let $\bar{x}_j = 1 - x_j$ ($j = 1, \dots, n$) and let us restrict attention to vectors of the form $x = (x_1, \dots, x_n, \bar{x}_1, \dots, \bar{x}_n)^T \in R^{2n}$. We denote by D a diagonal matrix of order $2n$ whose diagonal entries are $x_1, \dots, x_n, \bar{x}_1, \dots, \bar{x}_n$. The vector field associated with the linear rescaling algorithm assigns to a point x the vector

$$\xi_l(x) = D^2c - D^2A^T(AD^2A^T)^{-1}AD^2c.$$

Now,

$$AD^2A^T = \begin{pmatrix} x_1^2 + \bar{x}_1^2 & & \\ & \ddots & \\ & & x_n^2 + \bar{x}_n^2 \end{pmatrix} \quad \text{and}$$

$$AD^2c = \left(x_1^2c_1, \dots, x_n^2c_n\right)^T.$$

It follows that

$$D^2A^T(AD^2A^T)^{-1}AD^2c = \left(\frac{x_1^4c_1}{x_1^2 + \bar{x}_1^2}, \dots, \frac{x_n^4c_n}{x_n^2 + \bar{x}_n^2}, \frac{x_1^2\bar{x}_1^2c_1}{x_1^2 + \bar{x}_1^2}, \dots, \frac{x_n^2\bar{x}_n^2c_n}{x_n^2 + \bar{x}_n^2}\right)^T.$$

We now have the expression for the vector field:

$$(\xi_l(x))_j = \left(x_j^2 - \frac{x_j^4}{x_j^2 + \bar{x}_j^2}\right)c_j = \frac{x_j^2\bar{x}_j^2c_j}{x_j^2 + \bar{x}_j^2} \qquad (j = 1, \dots, n).$$

It is interesting to examine the orbits in this vector field. Fortunately, the underlying differential equations are *separable*. For every j ($j = 1, \dots, n$),

$$\dot{x}_j = \frac{x_j^2\bar{x}_j^2c_j}{x_j^2 + \bar{x}_j^2}.$$

It follows that

$$\frac{\dot{x}_j}{(1 - x_j)^2} + \frac{\dot{x}_j}{x_j^2} = c_j.$$

The solution is given implicitly by

$$\frac{1}{1 - x_j(t)} - \frac{1}{x_j(t)} = c_jt + C_j \quad \text{where}$$

$$C_j = \frac{1}{1 - x_j(0)} - \frac{1}{x_j(0)}.$$

It follows that for every j ($j = 1, \dots, n$) the function $x_j = x_j(t)$ is monotone increasing or decreasing as c_j is positive or negative, with $x_j(-\infty) = 0$ or $x_j(-\infty) = 1$ and $x_j(\infty) = 1$ or and $x_j(\infty) = 0$, respectively. Suppose, for simplicity, that $c = (1, \dots, 1)^T$. It is easy to verify that for every $\epsilon > 0$, if $C_i - C_j > 2/\epsilon$ then there exists a time t such that $x_i(t) > 1 - \epsilon$ while $x_j(t) < \epsilon$. Consider any permutation $(i_1, \dots, i_n)$ of the indices $(1, \dots, n)$. For simplicity of notation, let us assume though that $(i_1, \dots, i_n) =$

$(1, \ldots, n)$. Suppose we choose the initial point $x(0)$ to be of the form $x^0 = x^0(\delta) = (\delta, \delta^2, \ldots, \delta^n)^T$ where $\delta > 0$. For every $\epsilon > 0$, there exists a $\delta > 0$ such that the orbit through the point $x^0(\delta)$ visits the vertices $(0, \ldots, 0)^T, (1, 0, \ldots, 0)^T, (1, 1, 0, \ldots, 0)^T, \ldots, (1, \ldots, 1)^T$ in this order. Note that there is a one-to-one correspondence between such permutations of the set of indices and ascending paths of vertices of the hypercube. Thus, the following is true:

PROPOSITION A.1. *For any linear programming problems on the unit hypercube, every ascending path of adjacent vertices can be approximated by an orbit in the vector field induced by the linear rescaling algorithm.*

Appendix B. Projective rescaling trajectories on the unit simplex. We consider linear programming problems on the unit simplex Δ, that is, problems of the form

$$(S) \qquad \begin{aligned} \text{Minimize} \quad & c^T x \\ \text{subject to} \quad & e^T x = 1, \\ & x \geqslant 0 \end{aligned}$$

where $e = (1, \ldots, 1)^T$, $c, x \in R^n$. Furthermore, to simplify the statement of the projective rescaling algorithm, we restrict attention to those problems in which the optimal value of the objective function is zero. Let x be any interior point, that is, $x \in R^n$, $x > 0$ and $e^T x = 1$ and let $D = D_x = \mathrm{Diag}(x_1, \ldots, x_n)$. Obviously, $De = x$.

The projection of any vector $v \in R^n$ on the subspace $\{z\colon e^T z = 0\}$ is equal to

$$v - \frac{1}{n}(e^T v)e.$$

The search direction is derived as follows. The interior point x determines a projective transformation T_x defined by

$$T_x(y) = \frac{1}{e^T D^{-1} y} D^{-1} y.$$

Thus, $T_x(x) = e/n$. The search direction in the image space is computed by projecting the vector Dc on the subspace $\{z\colon e^T z = 0\}$. This projection equals

$$Dc - \frac{1}{n}(e^T Dc)e$$

and also

$$Dc - \frac{1}{n}(c^T x)e.$$

Thus, in the image space the algorithm moves from the point e/n to a point of the form

$$\frac{1}{n}e - t\left[Dc - \frac{1}{n}(c^T x)e\right],$$

where $t > 0$ is a certain scalar. The inverse image of such a point is equal to

$$\frac{\frac{1}{n}x - t\left[D^2 c - \frac{1}{n}(c^T x)x\right]}{\frac{1}{n} - t\left[e^T D^2 c - \frac{1}{n}(c^T x)\right]}.$$

Now subtract x as in §1 to find a negative multiple of ξ_p, which is proportional to $-D^2c + (e^TD^2c)x$. In other words,

PROPOSITION B.1. *The search direction is at x is a multiple of*

$$\pi(x) = x - \frac{1}{e^TD_x^2c}D_x^2c.$$

A useful interpretation of the search direction is as follows. Imagine the vertices of the simplex (that is, the unit vectors $e^1, \ldots, e^n$) are repelling. Suppose the force at x that pushes away from e^j is proportional to $x_j^2c_j$. Then the direction of the resultant of these forces is the search direction at x. In particular, if $c = e^1 = (1, 0, \ldots, 0)^T$ then for every x ($x > 0$, $e^Tx = 1$), the movement at x is away from the point e^1. This means that all the trajectories are straight lines, namely, starting at an interior point x, we move along the line determined by x and e^1, away from e^1, until we hit the face where the first coordinate vanishes. The following proposition generalizes this observation.

PROPOSITION B.2. *Suppose the objective function c has the form $c = (c_1, \ldots, c_k, 0, \ldots, 0)^T$ where $c_1, \ldots, c_k > 0$ ($k < n$). Let x^0 be an interior point of the simplex. Under these conditions, the trajectory induced by c, starting at x^0, has the following properties*:

(i) *For every i, $j > k$, for any x along the path,*

$$\frac{x_i}{x_j} = \frac{x_i^0}{x_j^0}$$

and, moreover, the path hits the point

$$\frac{1}{\sum_{i=k+1}^n x_i^0}\left(0, \ldots, 0, x_{k+1}^0, \ldots, x_n^0\right)^T.$$

(ii) *The projection of the path on the set of the first k coordinates is the same as the projection on the first k coordinates of the path starting at $(x_1^0, \ldots, x_k^0, \sum_{i=k+1}^n x_i^0)^T$ where the problem is*

$$\begin{aligned} \textit{Minimize} \quad & c_1x_1 + \cdots + c_kx_k \\ \textit{subject to} \quad & x_1 + \cdots + x_{k+1} = 1, \\ & x_i \geqslant 0. \end{aligned}$$

PROOF. It is easy to verify the claims by looking at the differential equations defining the path:

$$\dot{x}_i = x_i - \frac{c_ix_i^2}{\sum_{j=1}^k c_jx_j^2} \qquad (i = 1, \ldots, k)$$

$$\dot{x}_i = x_i \qquad (i = k+1, \ldots, n).$$

Let us denote by z^k an n-vector in the unit simplex, consisting of 0's in the first k

positions followed by equal coordinates in the last $n-k$ positions. Thus,

$$z^0 = \frac{1}{n}(1,1,1,\dots,1)^T, \quad z^1 = \frac{1}{n-1}(0,1,1,\dots,1)^T, \quad z^2 = \frac{1}{n-2}(0,0,1,\dots,1)^T,$$

and so on. The next proposition asserts that there exist objective functions that induce on the simplex trajectories that visit the neighborhoods of all the points z^i ($i = 0,1,\dots,n-1$).

PROPOSITION B.3. *For any $\delta > 0$, there exists an objective function vector c ($c_i > 0$, $i = 1,\dots,n-1$, $c_n = 0$), such that in the problem (S), the trajectory starting at the center z^0 visits the δ-neighborhoods of the points $z^1,\dots,z^{n-2}$ and then hits the optimal point $z^{n-1} = e^n$.*

PROOF. Let $\delta > 0$ be any number. Consider first the problem (S) with the objective function vector $c^0 = e^1 = (1,0,\dots,0)^T$. With this vector, starting at z^0, the trajectory hits the point z^1. Let us now consider an objective function vector of the form $c^1 = (1,\epsilon,0,\dots,0)^T$, where $\epsilon > 0$. With the vector c^1, starting at z^0, the trajectory hits the point z^2. However, by continuity, there exists an $\bar{\epsilon}_1$ such that the trajectory will also visit the δ-neighborhood of z^1, provided $\epsilon < \bar{\epsilon}_1$. Let us now set $\epsilon = \epsilon_1$ where $0 < \epsilon_1 < \bar{\epsilon}_1$, so $c^1 = (1,\epsilon_1,0,\dots,0)^T$. Suppose, by induction, we have defined

$$c^k = (1,\epsilon_1,\epsilon_2,\dots,\epsilon_k,0,\dots,0)^T \qquad (k < nA - 2)$$

as an objective function vector such that with c^k, the trajectory starting at z^0, visits the δ-neighborhoods of the points $z^1,\dots,z^k$ and then hits the point z^{k+1}. Consider now an objective function vector of the form $c^{k+1} = (1,\epsilon_1,\dots,\epsilon_k,\epsilon,0,\dots,0)^T$, where $\epsilon > 0$. With the vector c^{k+1}, starting at z^0, the trajectory hits the point z^k. However, by continuity, there exists an $\bar{\epsilon}_{k+1}$ such that the trajectory will also visit the δ-neighborhoods of the points $z^1, z^2,\dots,z^{k+1}$, provided $\epsilon < \bar{\epsilon}_{k+1}$. We now set $\epsilon = \epsilon_{k+1}$ where $0 < \epsilon_{k+1} < \bar{\epsilon}_{k+1}$, so

$$c^{k+1} = (1,\epsilon_1,\dots,\epsilon_{k+1},0,\dots,0)^T.$$

Our proposition follows with $k = n-2$.

COROLLARY B.4. *For every $\delta > 0$, there exist an objective function vector c and an interior point of the simplex, x, such that the projective rescaling trajectory induced by c, starting at x, visits the δ-neighborhoods of all the vertices of the simplex.*

PROOF. First, for any $\epsilon > 0$, consider a projective scaling transformation, T_ϵ defined by

$$T_\epsilon(y) = \frac{1}{\sum_{i=1}^n \epsilon^i y_i}\left(\epsilon y_1, \epsilon^2 y_2,\dots,\epsilon^n y_n\right)^T.$$

Every face of the unit simplex is invariant under T_ϵ. It is easy to verify that when ϵ tends to zero, the point z^i tends to the vertex e^{i+1} ($i = 0,1,\dots,n-1$). It follows from Proposition B.3 that for every ϵ there exist objective functions inducing trajectories that visit the neighborhoods of the points $T_\epsilon(z^i)$ ($i = 0,1,\dots,n$). This implies our claim.

It is easy to see that the arguments used in this appendix actually suffice for proving a stronger result:

PROPOSITION B.5. *Let $P \in R^n$ be any convex polyhedral set of dimension n and suppose $x \in P$ is any nondegenerate vertex. Under these conditions, there exist n pairwise distinct points $x^1, \ldots, x^n \in P$, belonging to faces of decreasing dimensions which contain x, such that for every $\delta > 0$, there exists an objective function vector c that satisfies the following:*

(i) *the vertex x maximizes the function $c^T x$ over P, and*

(ii) *the projective rescaling trajectory through x^1 visits the δ-neighborhoods of all the points $x^1, \ldots, x^n$.*

Appendix C. A lemma on orthogonal projections. The following lemma is a special case of Corollary 1 in [P]. We thank L. D. Pyle for giving us this reference.

LEMMA C.1. *Let $A \in R^{m_1 \times n}$ and $B \in R^{m_2 \times n}$ be matrices such that $AB^T = 0$. Under these conditions, the orthogonal projection of any vector $v \in R^n$ on the intersection of the nullspaces of A and B can be obtained as follows. First, project v orthogonally into the nullspace of B, and then project this projection orthogonally into the nullspace of A.*

PROOF. Without loss of generality, assume A and B are of full rank. Let

$$C = \begin{pmatrix} A \\ B \end{pmatrix}.$$

Since $AB^T = 0$, it follows that

$$CC^T = \begin{pmatrix} AA^T & 0 \\ 0 & BB^T \end{pmatrix}.$$

Also, since AA^T and BB^T are nonsingular (even positive-definite), CC^T is nonsingular. The orthogonal projection of v into the nullspace of C is given by $[I - C^T(CC^T)^{-1}C]v$. It follows that

$$C^T(CC^T)^{-1}C = A^T(AA^T)^{-1}A + B^T(BB^T)^{-1}B.$$

On the other hand, the sequence of projections stated in the lemma results in the vector $[I - A^T(AA^T)^{-1}A][I - B^T(BB^T)^{-1}B]v$. The lemma now follows since $AB^T = 0$.

Appendix D. On the general barrier method in inequality form. In this appendix we consider a more general barrier function technique, where the barrier function is not necessarily the logarithm function. We also work here with the linear programming problem in the inequality form

$$\begin{aligned} &\text{Minimize} \quad c^T x \\ &\text{subject to} \quad Ax \geqslant b \end{aligned}$$

where $A \in R^{m \times n}$ is of full rank. We assume the feasible domain is of full dimension. The barrier function method works with a related function

$$f(x) = c^T x + \mu \sum_{i=1}^{m} g(A_i x - b_i)$$

where A_i denotes the ith row of the matrix A, μ is a positive parameter which is driven by the algorithm to zero, and $g(\xi)$ is a strictly convex function over the positive reals ($g''(x) > 0$ for $x > 0$) such that $g(\xi)$ tends to infinity as ξ tends to zero. The common choice, which was discussed throughout this paper, is $g(\xi) = -\ln \xi$.

The Newton barrier function technique amounts to taking a Newton step with respect to the problem of minimizing $f(x)$, followed by an update of the value of μ. Let x be a point such that $Ax > b$ and let D'_x denote a diagonal matrix of order m:

$$D'_x = \mathrm{Diag}\big(g'(A_1x - b_1), \dots, g'(A_mx - b_m)\big).$$

It is easy to check that the gradient of $f(x)$ is

$$\nabla f(x) = c + \mu A^T D'_x e$$

where $e = (1, \dots, 1)^T \in R^m$. Let

$$D''_x = \mathrm{Diag}\big(g''(A_1x - b_1), \dots, g''(A_mx - b_m)\big).$$

The Hessian matrix is thus

$$H_f(x) = \mu A^T D''_x A.$$

The direction given by Newton's method is the same as the direction of the vector

$$v = H_f^{-1} \nabla f = \left(A^T D''_x A\right)^{-1}\left(c + \mu A^T D'_x e\right).$$

So far we have not specified the choice and update rule of μ. Consider first the case where μ is taken at its limit, that is, we set μ to zero after the Newton direction has been computed. In other words, $v = (A^T D''_x A)^{-1} c$.

Remark D.1. We note that with $g(\xi) = -\ln \xi$ this choice of μ yields the analogue of the linear rescaling method for the problem in inequality form (see also [GMSTW]). The latter can be seen as follows. Given an interior point x, consider the ellipsoid

$$E = \left\{ y \colon \sum_{i=1}^{m} \left(\frac{A_i y - b_i}{A_i x - b_i} - 1 \right)^2 \leqslant 1 \right\}.$$

In other words,

$$E = \left\{ y \colon \|D_x^{-1}(Ay - b) - e\| \leqslant 1 \right\}.$$

The direction v corresponds to moving towards the minimum of the function $c^T y$ over E. Thus, consider the following optimization problem with respect to v:

$$\begin{aligned} &\text{Minimize} \quad c^T(x + v) \\ &\text{subject to} \quad \|D_x^{-1}[A(x + v) - b] - e\| = 1. \end{aligned}$$

Since $D_x^{-1}(Ax - b) = e$, it follows that this problem is equivalent to

$$\begin{aligned} &\text{Minimize} \quad c^T v \\ &\text{subject to} \quad \|D_x^{-1} A v\| = 1. \end{aligned}$$

However, we are interested only in the direction of the vector v, so we can write the following set of equations for the optimality conditions: $(A^T D_x^{-2} A)v = c$. This implies our claim that the choice $\mu = 0$ yields the analogue of the linear rescaling algorithm. Notice how the linear rescaling algorithm is simplified when the problem is posed in the inequality form rather than the standard form.

We now return to general barrier functions g and consider the limiting behavior of the direction

$$v = v(x) = \left(A^T D_x'' A\right)^{-1} c$$

as the point x approaches the boundary of the feasible domain. Recall that the matrix A is assumed to be of full rank and the diagonal entries of D_x'' are positive. The vector $v = v(x)$ is the solution of the system $(A^T D_x'' A)v = c$. An equivalent system is obtained by defining $w = D_x'' A v$:

$$(D_x'')^{-1} w - Av = 0,$$

$$A^T w = c.$$

Let $R = R(x)$ denote the following diagonal matrix:

$$R = \mathrm{Diag}\left(\frac{1}{\sqrt{g''(A_1 x - b_1)}}, \ldots, \frac{1}{\sqrt{g''(A_m x - b_m)}} \right).$$

The equations that determine v and w are the optimality conditions of the problem

$$(O(x)) \qquad \begin{array}{ll} \text{Minimize} & \|Rw\|^2 \\ \text{subject to} & A^T w = c. \end{array}$$

For an interior point x, the optimal solution is unique, $w = w(x)$. The optimization problem $((O(\bar{x}))$ is well-defined but the solution is not necessarily unique. Suppose x tends to a boundary point $\bar{x}$. Let us assume that the function $g(\xi)$ that underlies the barrier method is convex, twice continuously differentiable, and $g''(\xi)$ tends to infinity as ξ tends to zero. Since $g''(0) = \infty$, the matrix R tends to a finite limit $\bar{R}$ (with some diagonal entries equal to zero), which we denote by $R(\bar{x})$. Assume, without loss of generality, that for $i = 1, \ldots, l$, $A_i \bar{x} > b_i$, whereas for $i = l + 1, \ldots, m$, $A_i \bar{x} = b_i$. Let us rewrite the optimization problem in the form

$$\begin{array}{ll} \text{Minimize} & \|R^1 w^1\|^2 + \|R^2 w^2\|^2 \\ \text{subject to} & A_1^T w^1 + A_2^T w^2 = c, \end{array}$$

where the indices 1 and 2 correspond to the first l and the last $m - l$ rows of A, respectively, and describe submatrices accordingly. Denote the optimal value of the optimization problem $(O(x))$ by $f(x)$.

PROPOSITION D.2. $\lim_{x \to \bar{x}} f(x) = f(\bar{x})$.

PROOF. First, for any w such that $A^T w = c$,

$$f(x) \leqslant \|R(x) w\|^2,$$

and hence

$$\limsup_{x \to \bar{x}} f(x) \leqslant \|R(\bar{x})w\|^2.$$

Let $\bar{w}$ be an optimal solution for $(O(\bar{x}))$. It follows that

$$\limsup_{x \to \bar{x}} f(x) \leqslant \|R(\bar{x})\bar{w}\|^2 = f(\bar{x}).$$

Second,

$$f(x) = \|R^1(x)w^1(x)\|^2 + \|R^2(x)w^2(x)\|^2 \geqslant \|R^1(x)w^1(x)\|^2.$$

It follows that

$$\liminf_{x \to \bar{x}} f(x) \geqslant \liminf_{x \to \bar{x}} \|R^1(x)w^1(x)\|^2$$

and $\|R^1(x)w^1(x)\|$ is bounded in a neighborhood of $\bar{x}$, since $f(x)$ is. This implies that $w^1(x)$ is bounded in a neighborhood of $\bar{x}$. Now, let w^* be any accumulation point of $w^1(x)$ as x tends to $\bar{x}$ such that

$$\|R^1(\bar{x})w^*\| = \liminf_{x \to \bar{x}} \|R^1(x)w^1(x)\|.$$

Obviously,

$$\liminf_{x \to \bar{x}} \|R^1(x)w^1(x)\|^2 = \|R^1(\bar{x})w^*\|^2 \geqslant \|R(\bar{x})\bar{w}\|^2 = f(\bar{x}).$$

This finally implies our claim.

PROPOSITION D.3. *The vector $R(x)w(x)$ converges as x tends to $\bar{x}$.*

PROOF. Let L denote the set of all vectors $w^1 \in R^l$ for which there exists a vector $w^2 \in R^{m-l}$ such that $A_1^T w^1 + A_2^T w^2 = c$. Let us denote by $w^1(\bar{x})$ the unique solution of the following optimization problem (in terms of w^1):

$$(O_1(x)) \qquad \begin{array}{ll} \text{Minimize} & \|R^1(\bar{x})w^1\|^2 \\ \text{subject to} & w^1 \in L. \end{array}$$

Since $R^2(\bar{x}) = 0$,

$$f(\bar{x}) = \|R^1(\bar{x})w^1(\bar{x})\|^2.$$

On the other hand, for any x we have $w^1(x) \in L$, so

$$\|R^1(\bar{x})w^1(x)\|^2 \geqslant \|R^1(\bar{x})w^1(\bar{x})\|^2 = f(\bar{x}).$$

Since $R(x)$ tends to $R(\bar{x})$,

$$\liminf_{x \to \bar{x}} \|R^1(\bar{x})w^1(x)\|^2 \geqslant f(\bar{x}) \quad \text{and}$$

$$\liminf_{x \to \bar{x}} \|R^1(x)w^1(x)\|^2 \geqslant f(\bar{x}).$$

From Proposition D.2 it follows that

$$f(\bar{x}) \geq \limsup_{x \to \bar{x}} \|R^1(x) w^1(x)\|^2.$$

This implies that $\|R^2(x)w^2(x)\|^2$ tends to zero, so $R^2(x)w^2(x)$ converges to zero. Moreover, it now follows that $\|R^1(x)w^1(x)\|^2$ converges, and its limit is necessarily equal to $f(\bar{x})$. Any accumulation point of $w^1(x)$ (as x tends to $\bar{x}$) is an optimal solution to $(O_1(x))$. However, the latter has a unique optimal solution. It finally follows that $w^1(x)$ converges to $w^1(\bar{x})$. This completes the proof.

The behavior of the direction of $v(x)$ as x approaches a boundary point is summarized in the following theorem.

THEOREM D.4. *Suppose the function $g(\xi)$ is convex, twice continuously differentiable, and $g''(\xi)$ tends to infinity as ξ tends to zero. Under these conditions, the vector field $v = v(x)$ extends continuously to the boundary of the feasible domain. Moreover, the direction of $v(x)$ tends to a direction parallel to any face Φ as x approaches Φ.*

PROOF. In Proposition D.2 we showed that, as x tends to $\bar{x}$, the vector Rw tends to a vector of minimum norm relative to $\overline{R}$. Thus, $(D_x'')^{-1}w$ also tends to a finite limit, and Av tends to the same. Since A is of full rank, v converges to a limit. Moreover, for every i such that $A_i\bar{x} = b_i$, $(D_x'')_{ii}^{-1}w_i$ tends to zero so, necessarily, $A_i v$ tends to zero. Obviously, this means that the direction of v tends to a direction parallel to the face that contains the point $\bar{x}$ in its interior.

Theorem D.4 generalizes to any fixed value of μ. The direction v is given in general by the equation

$$\left(A^T D_x'' A\right) v = c + \mu A^T d_x',$$

where $d_x' = D_x' e$. *Let*

$$u = D_x'' A v - \mu d_x'.$$

We now have an equivalent system

$$(D_x'')^{-1}(u + \mu d_x') - Av = 0$$

$$A^T u = c.$$

With R denoting the same matrix as above, the equations that determine v and u are precisely the optimality conditions of the problem

$$\text{Minimize} \quad \tfrac{1}{2}\|Ru\|^2 + \mu\left[(D_x'')^{-1} d_x'\right]^T u$$

$$\text{subject to} \quad A^T u = c.$$

We first observe the following:

PROPOSITION D.5. *Suppose $g(\xi)$ is a real-valued function satisfying the following conditions*:

(i) *$g(\xi)$ is differentiable in an open interval* $(0, a)$,

(ii) *$g(\xi)$ tends to infinity as ξ tends to zero,*

(iii) *the derivative $g'(\xi)$ is monotone.*

Under these conditions, the ratio $g(\xi)/g'(\xi)$ tends to zero with ξ.

PROOF. Since g' is monotone, g' tends to $-\infty$ at 0. Thus, g is monotone decreasing, and hence invertible, in a neighborhood of 0. For y near 0, let y^* denote the smallest value such that $g(y^*) = \frac{1}{2}g(y)$. Obviously, $y^* > y$. Since g is convex and differentiable,

$$\tfrac{1}{2}g(y) = g(y) - g(y^*) < g'(y)(y - y^*).$$

Thus,

$$2(y - y^*) < \frac{g(y)}{g'(y)} < 0.$$

It suffices to show that $y - y^*$ tends to 0 with y. Now pick x_0 and define x_i $(0 \leqslant x_i \leqslant x_{i-1})$ by $g(x_i) = 2g(x_{i-1})$ for $i \geqslant 1$. Obviously, the sequence $\{x_i\}$ is monotone decreasing and converges to 0, so $x_{i-2} - x_i$ tends to 0. If y lies between x_i and x_{i-1} then y^* lies between x_{i-1} and x_{i-2}. It follows that

$$y^* - y \leqslant x_{i-2} - x_i,$$

which tends to 0.

The asymptotic behavior of the direction of v in the general case is summarized as follows:

THEOREM D.6. *Suppose the function $g(\xi)$ is convex, twice differentiable continuously, and $g''(\xi)$ tends monotonically to infinity as ξ tends to zero. Under these conditions, the vector field*

$$v = v(x) = \left(A^T D_x'' A\right)^{-1}\left(c + \mu A^T d_x'\right)$$

(where μ is fixed) extends continuously to the boundary of the feasible domain. Moreover, the direction of $v(x)$ tends to a direction parallel to any face Φ as x approaches Φ.

PROOF. By Proposition D.5 and our assumptions about the underlying function $g(\xi)$, the vector $d = (D_x'')^{-1}d_x'$ tends to a finite limit and, moreover, if $A_i x - b_i$ tends to zero then also d_i tends to zero. As in the proof of Proposition D.3, it follows that here the vector Ru approaches a finite limit and hence the vector $Av = R^2u + \mu d$ approaches a finite limit. Moreover, it also follows that the direction of v tends to be parallel to the face as before.

Appendix E. The behavior of the barrier method in inequality form near vertices. Let us now consider the behavior of the *general* Newton barrier algorithm (for problems in inequality form) in the neighborhood of a nondegenerate vertex. Let V denote any nondegenerate vertex of the feasible polyhedron and suppose, without loss of generality, that the first n constraints are tight at V. Let B denote the $(n \times n)$-submatrix of A consisting of the first n rows. Thus, B is nonsingular. Also, let N denote the submatrix consisting of the other $m - n$ rows of A. Let D_B', D_B'', D_N' and D_N'' denote the square submatrices of D_x' and D_x'' corresponding to the indices of B and N as suggested by the notation. Obviously,

$$A^T D_x'' A = B^T D_B'' B + N^T D_N'' N.$$

When the point x tends to the vertex V, the diagonal entries of D_B'' tend to infinity while those of D_N'' tend to some finite limits. It follows that

$$\left(A^T D_x'' A\right)^{-1} = \left(I + \left(B^T D_B'' B\right)^{-1} N^T D_N'' N\right)^{-1} \left(B^T D_B'' B\right)^{-1}$$

$$= \left(I + B^{-1}(D_B'')^{-1} B^{-T} N^T D_N'' N\right)^{-1} B^{-1}(D_B'')^{-1} B^{-T}$$

so $(A^T D_x'' A)^{-1}$ is asymptotically equal to $B^{-1}(D_B'')^{-1}B^{-T}$.

Consider the approximate field in the neighborhood of the vertex V. The underlying differential equation of the approximate field is the following:

$$\dot{x} = -B^{-1}(D_B'')^{-1} B^{-T}\left(c + \mu A^T D' \frac{1}{\sqrt{n-m}} e\right).$$

We gain more insight if we change variables as follows. Let $s = Bx - b_B$. The problem in terms of s is

$$\begin{aligned} &\text{Minimize} \quad c^T B^{-1} s \\ &\text{subject to} \quad AB^{-1} \geqslant b - AB^{-1} b_B. \end{aligned}$$

Let $\tilde{c} = B^{-T}c$. Also, note that $\dot{s} = B\dot{x}$. It follows that the differential equation in terms of s is the following:

$$\dot{s} = -(D_B'')^{-1}\tilde{c} - \mu (D_B'')^{-1} B^{-T} A^T D' e.$$

Note that

$$B^{-T}A^T = (I\ B^{-T}N^T)$$

so

$$\dot{s} = -(D_B'')^{-1}\left(\tilde{c} + \mu B^{-T} N^T D_N' e\right) - \mu (D_B'')^{-1} D_B' e.$$

Under our assumptions about the function g, the dominant term in the latter is $-\mu(D_B'')^{-1}D_B'e$. Interestingly, when $g(\xi) = -\ln(\xi)$ this has a very simple form: $-D(''_B)^{-1}D_B'e = s$. This shows that for any fixed $\mu > 0$, if x is sufficiently close to the vertex V then x is *repelled* from V.

Appendix F. Differentiability of the linear rescaling vector field. In this appendix we prove the differentiability of the linear rescaling vector field on the entire feasible region. Thus, we have here another proof of the continuity already proven in §2. We use the same notation as in §2.

We consider points x in $P = \{x: Ax = b, x \geqslant 0\}$, where the problem is to maximize $c^T x$. Suppose the point x tends to $\bar{x}$. Denote $N = \{1, \dots, n\}$. Let I_1 denote the set of indices i such that $\bar{x}_i > 0$ and let $I_2 = N \setminus I_1$. Let A_i, $i = 1, 2$, denote the submatrix of A consisting of the columns with indices in I_i. Let R_1 and R_2 denote the subspaces of R^n corresponding to the sets I_1, I_2. Also, for any n-vector x denote by x^i a subvector corresponding to I_i.

We use the following notation:

$E = \{y: Ay = 0\}$; this is a fixed subspace in R^n,

c—The objective function vector; this is a fixed vector in R^n,

$E_1 = E \cap R_1$; this is a fixed subspace in R^n, PROOF. Since g' is monotone, g' tends to $-\infty$ at 0. Thus, g is monotone decreasing, and hence invertible, in a neighborhood of 0. For y near 0, let y^* denote the smallest value such that $g(y^*) = \frac{1}{2}g(y)$. Obviously, $y^* > y$. Since g is convex and differentiable,

$$\tfrac{1}{2}g(y) = g(y) - g(y^*) < g'(y)(y - y^*).$$

Thus,

$$2(y - y^*) < \frac{g(y)}{g'(y)} < 0.$$

It suffices to show that $y - y^*$ tends to 0 with y. Now pick x_0 and define x_i $(0 \leqslant x_i \leqslant x_{i-1})$ by $g(x_i) = 2g(x_{i-1})$ for $i \geqslant 1$. Obviously, the sequence $\{x_i\}$ is monotone decreasing and converges to 0, so $x_{i-2} - x_i$ tends to 0. If y lies between x_i and x_{i-1} then y^* lies between x_{i-1} and x_{i-2}. It follows that

$$y^* - y \leqslant x_{i-2} - x_i,$$

which tends to 0.

The asymptotic behavior of the direction of v in the general case is summarized as follows:

THEOREM D.6. *Suppose the function $g(\xi)$ is convex, twice differentiable continuously, and $g''(\xi)$ tends monotonically to infinity as ξ tends to zero. Under these conditions, the vector field*

$$v = v(x) = \left(A^T D_x'' A\right)^{-1}\left(c + \mu A^T d_x'\right)$$

(where μ is fixed) extends continuously to the boundary of the feasible domain. Moreover, the direction of $v(x)$ tends to a direction parallel to any face Φ as x approaches Φ.

PROOF. By Proposition D.5 and our assumptions about the underlying function $g(\xi)$, the vector $d = (D_x'')^{-1}d_x'$ tends to a finite limit and, moreover, if $A_i x - b_i$ tends to zero then also d_i tends to zero. As in the proof of Proposition D.3, it follows that here the vector Ru approaches a finite limit and hence the vector $Av = R^2u + \mu d$ approaches a finite limit. Moreover, it also follows that the direction of v tends to be parallel to the face as before.

Appendix E. The behavior of the barrier method in inequality form near vertices. Let us now consider the behavior of the *general* Newton barrier algorithm (for problems in inequality form) in the neighborhood of a nondegenerate vertex. Let V denote any nondegenerate vertex of the feasible polyhedron and suppose, without loss of generality, that the first n constraints are tight at V. Let B denote the $(n \times n)$-submatrix of A consisting of the first n rows. Thus, B is nonsingular. Also, let N denote the submatrix consisting of the other $m - n$ rows of A. Let D_B', D_B'', D_N' and D_N'' denote the square submatrices of D_x' and D_x'' corresponding to the indices of B and N as suggested by the notation. Obviously,

$$A^T D_x'' A = B^T D_B'' B + N^T D_N'' N.$$

When the point x tends to the vertex V, the diagonal entries of D_B'' tend to infinity while those of D_N'' tend to some finite limits. It follows that

$$\left(A^T D_x'' A\right)^{-1} = \left(I + \left(B^T D_B'' B\right)^{-1} N^T D_N'' N\right)^{-1} \left(B^T D_B'' B\right)^{-1}$$

$$= \left(I + B^{-1}(D_B'')^{-1} B^{-T} N^T D_N'' N\right)^{-1} B^{-1}(D_B'')^{-1} B^{-T}$$

so $(A^T D_x'' A)^{-1}$ is asymptotically equal to $B^{-1}(D_B'')^{-1}B^{-T}$.

Consider the approximate field in the neighborhood of the vertex V. The underlying differential equation of the approximate field is the following:

$$\dot{x} = -B^{-1}(D_B'')^{-1} B^{-T}\left(c + \mu A^T D' \frac{1}{\sqrt{n-m}} e\right).$$

We gain more insight if we change variables as follows. Let $s = Bx - b_B$. The problem in terms of s is

$$\text{Minimize} \quad c^T B^{-1} s$$

$$\text{subject to} \quad AB^{-1} \geqslant b - AB^{-1} b_B.$$

Let $\tilde{c} = B^{-T}c$. Also, note that $\dot{s} = B\dot{x}$. It follows that the differential equation in terms of s is the following:

$$\dot{s} = -(D_B'')^{-1}\tilde{c} - \mu(D_B'')^{-1} B^{-T} A^T D' e.$$

Note that

$$B^{-T}A^T = (I \; B^{-T}N^T)$$

so

$$\dot{s} = -(D_B'')^{-1}\left(\tilde{c} + \mu B^{-T} N^T D_N' e\right) - \mu(D_B'')^{-1} D_B' e.$$

Under our assumptions about the function g, the dominant term in the latter is $-\mu(D_B'')^{-1}D_B'e$. Interestingly, when $g(\xi) = -\ln(\xi)$ this has a very simple form: $-D(''_B)^{-1}D_B'e = s$. This shows that for any fixed $\mu > 0$, if x is sufficiently close to the vertex V then x is *repelled* from V.

Appendix F. Differentiability of the linear rescaling vector field. In this appendix we prove the differentiability of the linear rescaling vector field on the entire feasible region. Thus, we have here another proof of the continuity already proven in §2. We use the same notation as in §2.

We consider points x in $P = \{x\colon Ax = b, x \geqslant 0\}$, where the problem is to maximize c^Tx. Suppose the point x tends to $\bar{x}$. Denote $N = \{1, \ldots, n\}$. Let I_1 denote the set of indices i such that $\bar{x}_i > 0$ and let $I_2 = N \setminus I_1$. Let A_i, $i = 1, 2$, denote the submatrix of A consisting of the columns with indices in I_i. Let R_1 and R_2 denote the subspaces of R^n corresponding to the sets I_1, I_2. Also, for any n-vector x denote by x^i a subvector corresponding to I_i.

We use the following notation:

$E = \{y\colon Ay = 0\}$; this is a fixed subspace in R^n,

c—The objective function vector; this is a fixed vector in R^n,

$E_1 = E \cap R_1$; this is a fixed subspace in R^n,

F—The orthogonal complement of E_1 in E; also fixed,

η_x—The orthogonal projection of $D_x c$ into $D_x^{-1}E$, eventually extended continuously to P;

F_x—The orthogonal complement of $D_x^{-1}E_1$ in $D_x^{-1}E$,

η_{F_x}—The orthogonal projection of $D_x v$ into F_x.

We wish to show that $D_x\eta_x$ is differentiable as a function of the point x even at boundary points $\bar{x}$ of P. We will show below that $D_x\eta_{F_x}$ is differentiable with zero derivative on the boundary, and that η_{F_x} tends to zero as x tends to $\bar{x}$.

PROPOSITION F.1. *For every $K > 0$, there exists a neighborhood N_K of $\bar{x}$ such that for all interior points $x \in N_K$ and $u \in D_x^{-1}F$, $\|u^2\| \geqslant K\|u^1\|$.*

PROOF. Since $E \cap R_1 = E_1$, we have $F \cap R_1 = 0$. Thus, the angle between any vector in F and any vector in R_1 is bounded away from zero. In other words, there exists a constant $C > 0$ such that, for $u = (u^1, u^2) \in F$, $C\|u^2\| \geqslant \|u^1\|$ (with equality holding at $u^1 = u^2 = 0$). It follows that for every $x > 0$,

$$\|D_x^{-1}u^2\| \geqslant C\frac{\min\{x_j\colon j \in I_1\}}{\max\{x_j\colon j \in I_2\}}\|D_x^{-1}u^1\| \quad \text{and}$$

$$\lim_{x\to\bar{x}} \frac{\|D_x^{-1}u^2\|}{\|D_x^{-1}u^1\|} = \infty.$$

(Note that $\min\{x_j\colon j \in I_1\}$ tends to a positive limit whereas $\max\{x_j\colon j \in I_2\}$ tends to zero.) This implies our claim.

PROPOSITION F.2. *Given $K > 0$, let N_K be a neighborhood of $\bar{x}$ satisfying the condition of Proposition F.1. Let $x \in N_K$ be an interior point, let $w = (w^1, w^2) \in D_x^{-1}F$, and let u be a unit vector in $D_x^{-1}E_1$. Under these conditions,*

$$u^Tw \leqslant \|w^1\| \leqslant \frac{1}{\sqrt{K^2+1}}\|w\|.$$

PROOF. Since $u^2 = 0$, we have

$$u^Tw = (u^1)^T w^1 \leqslant \|w^1\|.$$

By Proposition F.1,

$$\|w\|^2 = \|w^1\|^2 + \|w^2\|^2 \geqslant (K^2+1)\|w^1\|^2$$

and this completes the proof.

For any flat $M \subset R^n$ and any $w \in R^n$, let $\Pi(w; M)$ denote the orthogonal projection of w into M.

PROPOSITION F.3. *Given $K > 0$, let N_K be a neighborhood of $\bar{x}$ as in Proposition F.1. Let l denote the dimension of F_1. Under these conditions, for every $w \in D_x^{-1}F$,*

$$\|\Pi(w; D_x^{-1}E_1)\| \leqslant \sqrt{\frac{l}{K^2+1}}\,\|w\| \quad \textit{and}$$

$$\|\Pi(w; F_x)\| \geqslant \sqrt{1 - \frac{l}{K^2+1}}\,\|w\|.$$

PROOF. Let $\{u^1, \ldots, u^l\}$ be an orthonormal basis of $D_x^{-1}E_1$. Then, by Proposition F.2,

$$\|\Pi(w;\, D_x^{-1}E_1)\|^2 = \sum_{i=1}^{l} (w^T u^i)^2 \leqslant \frac{l}{K^2+1}\|w\|^2.$$

Recall that F_x is the orthogonal complement of $D_x^{-1}E_1$ in $D_x^{-1}E$, so

$$\|\Pi(w;\, D_x^{-1}E_1)\|^2 + \|\Pi(w;\, F_x)\|^2 = \|w\|^2$$

since $w \in D_x^{-1}F$. This implies the rest of the claim.

PROPOSITION F.4. *Under the conditions of Proposition F.3, for every $f = (f^1, f^2) \in F_x$,*

$$\|f^2\| \geqslant \frac{K}{1+\sqrt{l}}\|f^1\|.$$

PROOF. There is a point $e = e(f) \in D_x^{-1}E_1$ such that $f + e(f) \in D_x^{-1}F$. Thus by Proposition F.2

$$K\|f^1 + e(f)\| \leqslant \|f^2\|$$

and since f and $e(f)$ are orthogonal, Proposition F.3 gives

$$\|e(f)\| \leqslant \sqrt{\frac{l}{K^2+1}}\,\|f + e(f)\| \leqslant \sqrt{\frac{l}{K^2+1}}\,\sqrt{\frac{K^2+1}{K^2}}\,\|f^2\|.$$

PROPOSITION F.5. *The orthogonal projection $\Pi(D_x c;\, F_x)$ tends to zero as x tends to $\bar{x}$.*

PROOF. Let $\{f^1, \ldots, f^s\}$ be an orthonormal basis of F_x. We represent the subvectors of f^i corresponding to the sets I_1 and I_2 by $f^{(i,1)}$ and $f^{(i,2)}$, respectively. We have for every i

$$(D_x c)^T f^i = \left(D_{x_1} c^1\right)^T f^{(i,1)} + \left(D_{x_2} c^2\right)^T f^{(i,2)}.$$

By Proposition F.4, there exist constants K_x and C_x, both tending to zero as x tends to $\bar{x}$, such that

$$\|f^{(i,1)}\| \leqslant K_x\|f^{(i,2)}\| \quad \text{and} \quad \|f^{(i,1)}\| \leqslant C_x.$$

We now use the Cauchy-Schwartz inequality to estimate

$$\begin{aligned}\left(D_{x_1} c^1\right)^T f^{(i,1)} + \left(D_{x_2} c^2\right)^T f^{(i,2)} &\leqslant \|x^1\|_\infty\, \|c^1\|\, \|f^{(i,1)}\| + \|x^2\|_\infty\, \|c^2\|\, \|f^{(i,2)}\| \\ &\leqslant \|x^1\|_\infty\, \|c^1\| C_x + \|x^2\|_\infty\, \|c^2\|\, \|f^{(i,2)}\|.\end{aligned}$$

Thus, $(D_x c)^T f^i$ tends to 0 as x tends to $\bar{x}$ and

$$\eta_{F_x} = \sum_i \left((D_x v)^T f^i\right) f^i$$

tends to zero.

Let $\eta_{D_x^{-1}E_1}$ denote the orthogonal projection of $D_x c$ on $D_x^{-1}E_1$. Then $\eta_{D_x^{-1}E_1}$ is real analytic in a neighborhood of $\bar{x}$. Thus the vector $\eta_x = \eta_{F_x} + \eta_{D_x^{-1}E_1}$ extends continuously to the boundary point $\bar{x}$ with $\eta_{F_{\bar{x}}} = 0$. Note that at a boundary point $\bar{x}$, $\eta_{\bar{x}}$ is the orthogonal projection of $D_{x^1}c^1$ into $D_{x^1}E_1$. Now,

$$D_x\eta_x = D_x\eta_{F_x} + D_x\eta_{D_x^{-1}E_1} \quad \text{and}$$

$$D_x\eta_{F_x} = D_{x^1}\eta_{F_{x^1}} + D_{x^2}\eta_{F_{x^2}}$$

From the proof of Proposition F.1 and Proposition F.4,

$$\eta_{f_{x^2}} \geqslant \frac{C}{\sqrt{l}+1}\frac{\min\{x_j\colon j\in I_1\}}{\max\{x_j\colon j\in I_2\}}\eta_{F_{x^1}}$$

and $\eta_{F_{x^2}}$ tends to zero, so

$$D_x\eta_{F_x} \leqslant \frac{\max\{x_j\colon j\in I_2\}}{C\min\{x_j\colon j\in I_1\}}\|\eta_{F_{x^2}}\|\max\{x_j\colon j\in I_2\}(\sqrt{l}+1)$$

$$+\max\{x_j\colon j\in I_2\}\|\eta_{F_{x^2}}\|.$$

Since $\max\{x_j\colon j\in I_2\}/\|x-\bar{x}\|$ is bounded and $\|\eta_{F_{x^2}}\|$ tends to zero we have $\|D_x\eta_{F_x}\|/\|x-\bar{x}\|$ tending to zero as x tends to $\bar{x}$. We finally have

THEOREM F.6. *$D_x\eta_x$ is differentiable at $\bar{x}$ and $D_x\eta_{F_x}$ has zero derivative at $\bar{x}$.*

PROOF. The computation above applies to every face in which $\bar{x}$ lies, to prove that $D_x\eta_{F_x}$ has zero derivative at $\bar{x}$.

REMARK F.7. Having proven the differentiability, we can now give an explicit expression for the derivative. Let us denote $U = U(x) = D_x\eta_x \in R^n$ and let $U'(x) \in R^{n\times n}$ denote the derivative. If x is interior the vector U is determined by the following set of equations in U and W:

$$D_x^{-2}U(x) - A^TW(x) = c$$

$$AU(x) = 0.$$

Differentiation gives

$$D_x^{-2}U'(x) - A^TW'(x) = 2D_x^{-3}D_{U(x)}$$

$$AU'(x) = 0.$$

The interpretation of this system is that the jth column of $U'(x)$ is equal to D_x times the orthogonal projection of the jth column of the matrix $M = 2D_x^{-3}D_{U(x)}$ into the nullspace of the matrix AD_x. However, since M is diagonal, the jth column of U' turns out to be the orthogonal projection of $U_j(x)e^j/(x_j^2)$ into that nullspace. Interpretations can be developed for higher order derivatives using the same idea.

Acknowledgements. Michael Shub thanks Earl Barnes for introducing him to the linear rescaling algorithm and also acknowledges conversations with Lenore Blum.

References

[Bar] Barnes, E. R. (1986). A Variation on Karmarkar's Algorithm for Solving Linear Programming Problems. *Math. Programming* **36** 174–182.

[BayL] Bayer, D. A. and Lagarias, J. C. (1986). The Nonlinear Geometry of Linear Programming. I. Affine and Projective Rescaling Trajectories. AT & T preprint.

[Blu] Blum, L. (1985). Towards an Asymptotic Analysis of Karmarkar's Algorithm. Extended abstract.

[CaS] Cavalier, T. M. and Soyster, A. L. (1985). Some Computational Experience and a Modification of the Karmarkar Algorithm. Presented at the 12th Sympos. on Mathematical Programming, Cambridge, MA, August.

[GMSTW] Gill, P. E., Murray, W., Saunders, M. A., Tomlin, J. A. and Wright, M. H. (1985). On Projected Newton Barrier Methods for Linear Programming and an Equivalence to Karmarkar's Projective Method. Technical report SOL 85-11, Systems Optimization Laboratory, Department of Operations Research, Stanford University, Stanford, CA 94305, July.

[G] Gomory, R. E. (1955). Trajectories Tending to a Critical Point in 3-Space. *Ann. Math.* **61** 140–153.

[H] Hartman, P. (1964). *Ordinary Differential Equations*, J. Wiley and Sons, New York.

[Kar] Karmarkar, N. (1984). A New Polynomial-Time Algorithm for Linear Programming. in: *Proc. 16th Annual ACM Sympos. Theory of Computing* (1984), ACM, New York, 302–311; revised version: *Combinatorica* **4** 373–395.

[Me1] Megiddo, N. (1984). A Variation on Karmarkar's Algorithm. unpublished manuscript, December.

[Me2] _____ (1986). Pathways to the Optimal Set in Linear Programming. in: *Proc. 7th Math. Programming Sympos. Japan* (1986), 1–36; also available as Research Report RJ 5295, IBM Almaden Research Center, San Jose CA.

[P] Pyle, L. D. (1967). A Generalized Inverse ϵ-Algorithm for Constructing Intersection Projection Matrices with Application. *Numer. Math.* **10** 86–102.

[Re] Renegar, J. (1988). A Polynomial-Time Algorithm, Based on Newton's Method, for Linear Programming. *Math. Programming* **40** 59–94.

[Sh] Shub, M. (1986). *Global Stability of Dynamical Systems*. Springer, New York.

[VMF] Vanderbei, R. J., Meketon, M. J. and Freedman, B. A. (1986). A Modification of Karmarkar's Linear Programming Algorithm. *Algorithmica* **1** 395–408.

MEGIDDO: IBM ALMADEN RESEARCH CENTER, 650 HARRY ROAD, SAN JOSE, CALIFORNIA 95120-6099
TEL AVIV UNIVERSITY, TEL AVIV, ISRAEL

SHUB: IBM T. J. WATSON RESEARCH CENTER, BOX 218, YORKTOWN HEIGHTS, NEW YORK 10598