Reprinted from

N. Megiddo
Editor

Progress in Mathematical Programming

Springer-Verlag
New York Berlin Heidelberg
London Paris Tokyo

CHAPTER 8

Pathways to the Optimal Set in Linear Programming

Nimrod Megiddo

Abstract. This chapter presents continuous paths leading to the set of optimal solutions of a linear programming problem. These paths are derived from the weighted logarithmic barrier function. The defining equations are bilinear and have some nice primal-dual symmetry properties. Extensions to the general linear complementarity problem are indicated.

§1. Introduction

Algorithms for mathematical programming can often be interpreted as path-following procedures. This interpretation applies to the simplex method [4], Scarf's fixed-point algorithm [19], Lemke's algorithm [13] for the linear complementarity problem, homotopy methods for piecewise linear equations [5], and most of the methods for nonlinear optimization. This is the theme of the book by Garcia and Zangwill [8]. More recent algorithms for linear programming by Murty [17] and Mangasarian [15] are also based on natural paths that lead to optimal solutions. Iterative algorithms for nonlinear optimization usually assign to any point x in a certain set $S \subset R^n$ (usually convex) a "next point" $x' = f(x) \in S$. Given a starting point x^0, the iterative scheme generates a sequence of points $\{x^k\}$, where $x^{k+1} = f(x^k)$, that converges to a solution.

It is often instructive to consider "infinitesimal" versions of iterative algorithms in the following sense. Given the iterative scheme $x' = f(x)$, consider the differential equation

$$\dot{x} = f(x) - x.$$

This chapter was published in the Proceedings of the 7th Mathematical Programming Symposium of Japan, Nagoya, Japan, pp. 1–35, and is referenced by several other papers in this volume. Since those Proceedings are not highly available, it is reproduced here in its original form.

When this equation has a unique solution through x^0 then it determines a path $x = x(t)$ such that the tangent to the path at any x is equal to the straight line determined by x and x'. If the algorithm generates the point x' close to x then the path may be a good approximation to the sequence generated by the algorithm. This is true at least during later stages of the execution if the sequence converges to a solution point. If the algorithm makes large steps during the early stages then the path may be a bad approximation. Trajectories corresponding to discrete algorithms for nonlinear optimization were analyzed in [6, 8, 10]. The analogy to differential equations is well known.

Several people have recently worked on solution paths in linear programming. Nazareth [18] interprets Karmarkar's algorithm [12] as a homotopy method with restarts. Results about the infinitesimal version of Karmarkar's algorithm and related algorithms were recently obtained in [2] and [16]. Smale [20] showed that the path generated by the self-dual simplex algorithm [4] can be approximated by the Newton's method path for solving a certain system of nonlinear equations.

In this chapter we study solution paths related to barrier functions for linear programming. We believe the study of paths is essential for the design and analysis of algorithms for optimization. In Section 2 we describe the paths for linear programming problems in standard form. In Section 3 we develop essentially the same theory within a more symmetric framework. In Section 4 we analyze some properties of tangents to the trajectories, whereas Section 5 brings some observations on higher-order derivatives. In Section 6 we consider the behavior of trajectories near corners. In Section 7 we discuss generalizations to the linear complementarity problem.

§2. On the Logarithmic Barrier Function

In this section we consider the linear programming problem in the standard form

$$\text{(P)} \qquad \begin{aligned} &\text{Maximize } c^T x \\ &\text{subject to } Ax = b, \\ &\qquad\qquad\quad x \geq 0 \end{aligned}$$

where $A \in R^{m \times n}$, $b \in R^m$, and $c, x \in R^n$. We believe most of the readers are used to considering the linear programming problem in this form. However, an analogous analysis can be carried out with respect to other forms of the problem. We shall later discuss the problem in more detail, using another more symmetric variant. The presentation will therefore entail a fair amount of redundancy, which, we hope, will be of help to the reader.

The logarithmic barrier function technique, usually used in nonlinear constrained optimization, can of course be applied to the linear programming problem. This method recently came up in [9], where Karmarkar's algorithm

[12] was analyzed from the barrier function viewpoint, but the idea of using this function in the context of linear programming is usually attributed to Frisch [7]. The technique gives rise to the following problem:

$$(\mathrm{P}_\mu) \qquad \begin{aligned} &\text{Maximize } c^T x + \mu \sum_j \ln x_j \\ &\text{subject to } Ax = b, \\ &\qquad\qquad x > 0 \end{aligned}$$

where $\mu > 0$ is typically small. The barrier function approach is valid only if there exists an $x > 0$ such that $Ax = b$. However, it is easy to reformulate the problem, using one artificial variable, so that the feasible domain is of full dimension. We use e to denote a vector of 1's of any dimension as required by the context. Also, M denotes a real number always chosen to be sufficiently large or an "infinite" element adjoined to the ordered field of the reals. The following construction is well known. Given a problem in the form (P), consider the following problem:

$$(\mathrm{P}^*) \qquad \begin{aligned} &\text{Maximize } c^T x - M\xi \\ &\text{subject to } Ax + (b - Ae)\xi = b, \\ &\qquad\qquad x, \xi \geq 0. \end{aligned}$$

Obviously, x is an optimal solution for (P) if and only if $(x, 0)$ is an optimal solution for (P*). It follows that the vector $(x, \xi) = e$ satisfies the set of equations. Thus, without loss of generality, we may assume the problem is given in the form (P) and also $Ae = b$.

For any d-vector x, let D_x denote the diagonal matrix of order $d \times d$ whose diagonal entries are the components of x. A vector $x > 0$ is an optimal solution for (P_μ) if and only if there exists a vector $y \in R^m$ such that

$$(\mathrm{O}) \qquad \begin{aligned} \mu D_x^{-1} e \quad - A^T y &= -c, \\ Ax \qquad\qquad &= b. \end{aligned}$$

Obviously, the problem (P_μ) may be unbounded. Let us assume, for a moment, that the feasible domain $\{x\colon Ax = b, x \geq 0\}$ is bounded. At least in this case both (P) and (P_μ) have optimal solutions (for every μ). Under the boundedness assumption, (P_μ) has a *unique* optimal solution for every $\mu > 0$ since its objective function is strictly concave. Thus, under the boundedness assumption, the system (O) has a unique solution for x for every $\mu > 0$.

The left-hand side of the system (O) represents a nonlinear mapping $F_\mu(x, y)$ of R^{n+m} into itself. The Jacobian matrix of this mapping at (x, y) is obviously the following:

$$J = J_\mu(x, y) = \begin{pmatrix} -\mu D_x^{-2} & -A^T \\ A & 0 \end{pmatrix}.$$

Suppose A is of full rank m $(m \leq n)$. In this case, the value of y is uniquely determined by the value of x. Also, it is well known that in this case the matrix AA^T is positive definite and hence nonsingular. The linear system of equations

$$J_\mu(x, y)\begin{bmatrix}\xi \\ \eta\end{bmatrix} = \begin{bmatrix}\mu \\ 0\end{bmatrix}$$

can be interpreted as a "least squares" problem or a projection problem. A solution can be expressed in terms of the matrix $(AD_x^2A^T)^{-1}$, which is well defined since A is of full rank. It is interesting to observe the following:

Proposition 8.1. *The problem* (P_μ) *is either unbounded for every* $\mu > 0$ *or has a unique optimal solution for every* $\mu > 0$.

Proof: Consider the interval I of values t for which the set

$$L(t) = \{x \geq 0 : c^Tx = t, Ax = b\}$$

has a nonempty interior. Obviously, I is an open interval. If for any $t \in I$ the function $\phi(x) = \sum_j \ln x_j$ is unbounded on $L(t)$ then, of course, (P_μ) is unbounded for all positive values of μ. Without loss of generality, assume $\phi(x)$ is bounded over every $L(t)$ $(t \in I)$. Strict concavity of $\phi(x)$ implies that for each $t \in I$ there is a unique maximizer $x = x(t)$ of ϕ over $L(t)$. Let $g(t)$ denote the maximum value of $\phi(x)$ over $L(t)$. Consider first the case where $\infty \in I$; that is, the function c^Tx is unbounded. Here there is a *ray*, contained in the interior of the feasible region, along which c^Tx tends to infinity. Since the domain is polyhedral, the ray is bounded away from the boundary. Thus, on the ray the function $\phi(x)$ is bounded from below, and hence (P_μ) is unbounded for every $\mu > 0$. In the remaining case, notice that strict concavity of $g(t)$ implies that $t + \mu g(t)$ is bounded for every $\mu > 0$ if t is bounded. Thus, in the latter case (P_μ) has a unique optimal solution for every $\mu > 0$.

It follows from Proposition 8.1 that if the system (O) has a solution for any positive value of μ then it determines a unique and continuous path $x = x(\mu)$, where μ varies over the positive reals. When A is of full rank also a continuous path $y = y(\mu)$ is determined. We are interested in the limits of $x(\mu)$ and $y(\mu)$ as μ tends to zero. Suppose (for a moment) that the limits of $x(\mu)$ and $y(\mu)$ (as μ tends to 0) exist, and denote them by $\bar{x}$ and $\bar{y}$, respectively. It follows that $A\bar{x} = b$, $\bar{x} \geq 0$, and $A^T\bar{y} \geq c$. Moreover, for each j such that $\bar{x}_j > 0$, $A_j^T\bar{y} = c_j$. It follows that $\bar{x}$ and $\bar{y}$ are optimal solutions for (P) and its dual, respectively. To relate these paths to an algorithm for the linear programming problem, we have to address at least two issues. First, we have to know a solution for, say, $\mu = 1$. Second, the limit of $x(\mu)$ (as μ tends to zero) should exist.

It is easy to modify the objective function so that an initial solution becomes available. Note that instead of (P_μ) we can work with a problem of the form

$$(\mathrm{P}_\mu(w)) \qquad \begin{aligned} &\text{Maximize } c^Tx + \mu \sum_j w_j \ln x_j \\ &\text{subject to } Ax = b, \\ &\qquad\qquad x > 0, \end{aligned}$$

where $w \in R^n_+$ is any vector with positive components. Proposition 8.1 extends to this case. Suppose x^0 and y^0 are interior feasible solutions for the primal and the dual problems, respectively. We will show later that any problem can be reformulated so that such solutions are readily available. We can choose w so that the vectors x^0 and y^0 satisfy the optimality conditions with respect to $(\mathrm{P}_\mu(w))$ at $\mu = 1$:

$$
(\mathrm{P}(w)) \qquad \begin{aligned} \mu D_x^{-1} w \quad - A^T y &= -c \\ Ax \qquad\qquad &= b. \end{aligned}
$$

Specifically, $w = D_{x^0}(A^T y^0 - c)$. Thus, given any pair of interior feasible solutions for the primal and the dual problems, we can easily calculate a suitable weight vector w, which in turn determines paths $x = x(\mu)$ and $y = y(\mu)$ as explained above. We discuss the role of the weights in more detail in Section 3.

In view of the preceding discussion, let us assume that for every $\mu > 0$ the system (O) has a unique solution $(x(\mu), y(\mu))$. It is easy to show that $c^T x(\mu)$ tends to the optimal value of (P). This follows if we multiply the first row of (O) by $x(\mu)$, the second by $y(\mu)$, and then add them up. We get $b^T y(\mu) - c^T x(\mu) = n\mu$. The optimal value lies between $b^T y(\mu)$ and $c^T x(\mu)$ and this implies our claim that $c^T x(\mu)$ tends to the optimal value as μ tends to 0. We are interested in conditions under which the *point* $x(\mu)$ tends to an optimal solution of (P).

Let $V(\mu) = c^T x(\mu)$ (where $x(\mu)$ is the optimal solution of (P_μ)), and let $V(0)$ denote the optimal value of (P). We have just argued that $V(\mu)$ tends to $V(0)$ as μ tends to 0. Obviously, $x(\mu)$ is also the optimal solution of the following problem:

$$
(\bar{\mathrm{P}}_\mu) \qquad \begin{aligned} &\text{Maximize } c^T x + \mu \sum_j \ln x_j \\ &\text{subject to } Ax = b, \\ &\qquad\qquad c^T x = V(\mu), \\ &\qquad\qquad x > 0. \end{aligned}
$$

The latter is of course equivalent to

$$
\begin{aligned} &\text{Maximize } \sum_j \ln x_j \\ &\text{subject to } Ax = b, \\ &\qquad\qquad c^T x = V(\mu), \\ &\qquad\qquad x > 0. \end{aligned}
$$

Our assumption of existence of the path $x(\mu)$ is equivalent to existence of an optimal solution for the problem $(\bar{\mathrm{P}}_\mu)$ for any $\mu > 0$. Using the notation of Proposition 8.1, the function $\phi(x)$ is bounded on every $L(t)$ where $t = V(\mu)$ for some $\mu > 0$. We assert that this implies that the set $L(t)$ itself is bounded. The proof is as follows. If $L(t)$ is unbounded then there is a ray, bounded away from the boundary, along which at least one of the variables tends to infinity while the others are bounded away from zero. Along such a ray the function $\phi(x)$ tends to infinity. It follows that the set $L(V(0))$ is bounded. The maximum value $g(t)$ is a concave function of t. This concavity implies that $g(t)$ is bounded from above as t tends to $V(0)$. Let N denote the set of all indices j such that $x_j = 0$ in *every* optimal solution. Thus, the optimal face is the intersection of the feasible domain with subspace $\{x: x_j = 0, j \in N\}$. Let

$$\phi_N(x) = \sum_{j \in N} \ln x_j$$

and

$$\phi_B(x) = \phi(x) - \phi_N(x).$$

Let $\xi_j(\mu)$ denote the jth component of the vector $x(\mu)$, that is, the optimal solution at μ. Since $\phi_N(x)$ is constant on the set $\{x: x_j = \xi_j(\mu), j \in N\}$, it follows that the point $x(\mu)$ is actually the optimal solution of the problem

$$\begin{aligned} &\text{Maximize } \phi_B(x) \\ &\text{subject to } Ax = b, \\ &\qquad\qquad c^T x = V(\mu), \\ &\qquad\qquad x_j = \xi_j(\mu) \qquad (j \in N), \\ &\qquad\qquad x > 0. \end{aligned}$$

Since the optimal set is bounded, it follows that the problem corresponding to $\mu = 0$, that is,

$$\begin{aligned} &\text{Maximize } \phi_B(x) \\ &\text{subject to } Ax = b, \\ &\qquad\qquad c^T x = V(0), \\ &\qquad\qquad x_j = 0 \qquad (j \in N), \\ &\qquad\qquad x_j > 0 \qquad (j \notin N), \end{aligned}$$

has a unique optimal solution which we denote by $x(0)$. We claim that $x(0)$ is equal to the limit of $x(\mu)$ as μ tends to zero. This solution is also characterized by the following system:

$$\begin{aligned} \frac{1}{x_j} - A_j^T y = -\lambda c_j \qquad (j \notin N), \\ Ax = b, \\ x_j = 0 \qquad (j \in N), \\ c^T x = V(0), \end{aligned}$$

where λ is a multiplier corresponding to the equation $c^T x = V(0)$ and A_j^T is the jth row of A^T. Any limit of a convergent sequence of points $x(\mu_k)$ (where μ_k tends to 0 as k tends to infinity) satisfies the latter system of equations and hence equals $x(0)$. Thus, $x(\mu)$ tends to $x(0)$ as μ tends to zero. We can thus state the following proposition:

Proposition 8.2. *If for some $\mu > 0$ the system* (O) *has a solution $x > 0$ then for every $\mu > 0$ there is a solution $x(\mu)$ so that the path $x(\mu)$ is continuous and the limit of $x(\mu)$ as μ tends to zero exists and constitutes an optimal solution to the linear programming problem* (P).

The implication of Proposition 8.2 is that we can solve the linear programming problem by a "homotopy" approach. Starting from $\mu = 1$, where we readily have an optimal solution to problems of the form $(\mathrm{P}_\mu(w))$, we follow the path of optimal solutions for such problems while μ varies from 1 to 0. The limit as μ tends to zero is an optimal solution to the linear programming problem, namely the point $x(0)$. In the next section we will continue the study of the paths introduced above. However, henceforth we will consider a more symmetric form of the problem.

§3. Duality

We find it more instructive to consider the linear programming problem in the *symmetric* form (in the sense of the duality transformation):

$$
\text{(P)} \qquad \begin{aligned} &\text{Maximize } c^T x \\ &\text{subject to } Ax \le b, \\ &\qquad\qquad x \ge 0 \end{aligned}
$$

where $A \in R^{m \times n}$, $b \in R^m$, and $c, x \in R^n$. The system $\{Ax \le b\}$ can obviously be replaced by $\{Ax + u = b, u \ge 0\}$ where $u \in R^m$. In this section we complement the results of Section 2 and provide additional insights.

The following nonlinear concave optimization problem (where μ is a fixed positive number) can be considered an approximation to (P):

$$
(\mathrm{P}_\mu) \qquad \begin{aligned} &\text{Maximize } c^T x + \mu\left(\sum_j \ln x_j + \sum_i \ln u_i\right) \\ &\text{subject to } Ax + u = b, \\ &\qquad\qquad x, u > 0. \end{aligned}
$$

Notice that the gradient of the function $\phi(x) = \sum_j \ln x_j$ is equal to $D_x^{-1}e$ and also to $D_x^{-2}x$. A pair of vectors $x \in R_+^n$ and $u \in R_+^m$, such that $Ax + u = b$, constitutes an optimal solution for (P_μ) if and only if there exists a vector $y \in R^m$ such that

$$\begin{aligned} \mu D_x^{-2} x \quad - A^T y &= -c, \\ \mu D_u^{-1} e \quad\quad - y &= 0. \end{aligned}$$

It follows that such a vector y must satisfy $u = \mu D_y^{-1} e$ and hence x is optimal in (P_μ) if and only if there is $y \in R_+^m$ such that

$$(\mathrm{O}) \qquad \begin{aligned} \mu D_x^{-2} x \quad - A^T y \quad &= -c, \\ Ax \quad + \mu D_y^{-2} y &= \ b. \end{aligned}$$

The system (O) has some nice symmetry properties. Consider the dual of (P), namely

$$(\mathrm{D}) \qquad \begin{aligned} &\text{Minimize } b^T y \\ &\text{subject to } A^T y \geq c, \\ &\qquad\qquad y \geq 0. \end{aligned}$$

An approximate nonlinear convex optimization problem is as follows.

$$(\mathrm{D}_\mu) \qquad \begin{aligned} &\text{Minimize } b^T y - \mu\left(\sum_i \ln y_i + \sum_j \ln v_j\right) \\ &\text{subject to } A^T y - v = c, \\ &\qquad\qquad y, v > 0. \end{aligned}$$

It is easy to check that y is optimal in (D_μ) if and only if there exists an $x \in R_+^n$ such that (O) holds. Note that the nonlinear objective functions of (P_μ) and (D_μ) (for $\mu > 0$) are strictly concave and strictly convex, respectively. Thus, each of the problems (P_μ) and (D_μ) has at most one optimal solution. The relationship between the problems (P_μ) and (D_μ) is summarized in the following duality theorem:

Theorem 8.1.

(i) *If the problem (P_μ) is unbounded then the problem (D_μ) is infeasible, and if the problem (D_μ) is unbounded then the problem (P_μ) is infeasible.*
(ii) *The problem (P_μ) has a optimal solution if and only if the problem (D_μ) has an optimal solution.*
(iii) *If x and y are optimal solutions for (P_μ) and (D_μ), respectively, then the gap between the values of the objective functions of (P_μ) and (D_μ) is equal to $(m + n)\mu(1 + \ln \mu)$, whereas the gap between $c^T x$ and $b^T y$ equals $(m + n)\mu$.*

Proof. Suppose x and y are feasible solutions to the problems (P_μ) and (D_μ), respectively. Let u and v be as above. It follows that

$$\begin{aligned} y^T A x + u^T y &= b^T y, \\ y^T A x - v^T x &= c^T x. \end{aligned}$$

Thus, for $\mu \le 1$,

$$\begin{aligned} b^T y - c^T x &= u^T y + v^T x \ge \mu(u^T y + v^T x) \\ &> \mu\left(\sum_i \ln u_i + \sum_i \ln y_i + \sum_j \ln v_j + \sum_j \ln x_j\right). \end{aligned}$$

Thus,

$$b^T y - \mu\left(\sum_i \ln y_i + \sum_j \ln v_j\right) > c^T x + \mu\left(\sum_j \ln x_j + \sum_i \ln u_i\right).$$

The latter is analogous to the weak duality in linear programming. It implies that if (P_μ) is unbounded then (D_μ) is infeasible and if (D_μ) is unbounded then (P_μ) is infeasible.

We know from the preceding discussion of the system (O) that (P_μ) has an optimal solution if and only if (D_μ) has one. The optimal solutions are unique. If x and y are the optimal solutions for (P_μ) and (D_μ), respectively, then the system (O) implies

$$v_j x_j = u_i y_i = \mu.$$

Thus,

$$b^T y - c^T x = (m + n)\mu$$

and

$$\begin{aligned} &\left[b^T y - \mu\left(\sum_i \ln y_i + \sum_j \ln v_j\right)\right] - \left[c^T x + \mu\left(\sum_j \ln x_j + \sum_i \ln u_i\right)\right] \\ &\quad = (m + n)\mu(1 - \ln \mu). \end{aligned}$$

Interestingly, the gap between the optimal values depends only on μ and the dimensions m and n and not on the data. It follows from Theorem 8.1 that the optimal solutions $x = x(\mu)$ and $y = y(\mu)$ are such that $c^T x(\mu)$ and $b^T y(\mu)$ tend to the optimal value of (P) (which of course equals the optimal value of (D)). Moreover, the "duality gap" tends to zero *linearly* with the parameter μ. It can then be shown, as in the preceding section, that the points themselves tend to optimal solutions of (P) and (D), respectively.

For the symmetric primal-dual barrier approach to work, we need both (P) and (D) to have full-dimensional feasible domains. We note that every linear programming problem can be reformulated so that both the primal and the dual have full-dimensional feasible domains. Given a problem in the form (P), consider the following problem, where M is sufficiently large:

$$(\mathrm{P}^*) \qquad \begin{aligned} &\text{Maximize } c^T x - M\xi \\ &\text{subject to } Ax + (b - Ae - e)\xi \le b, \\ &\qquad (c - A^T e + e)^T x \le M, \\ &\qquad x, \xi \ge 0. \end{aligned}$$

It is easy to verify that if M is sufficiently large then x is an optimal solution for (P) if and only if $(x, 0)$ is an optimal solution for (P*). The point $e \in R^{n+1}$ lies in the interior of the feasible domain of (P*). Also, the point $e \in R^{m+1}$ lies in the interior of the feasible domain of the dual of (P*):

$$(\mathrm{D}^*) \qquad \begin{aligned} &\text{Minimize } b^T y + M\eta \\ &\text{subject to } A^T y + (c - A^T e + e)\eta \geq c, \\ &\qquad (b - Ae - e)^T y \geq -M, \\ &\qquad y, \eta \geq 0. \end{aligned}$$

Tricks of "Big M" are fairly standard in linear programming. Alternatively, to avoid numerical problems with large values of M, we can use here the equivalent of what is called "Phase I" in the linear programming literature.

For simplicity of notation, we write (x, y) for the column vector obtained by concatenating two column vectors x and y. We find it interesting to consider the mapping ψ: $R^{n+m} \to R^{n+m}$, defined by

$$\psi(x, y) = (\mu D_x^{-2} x - A^T y, Ax + \mu D_y^{-2} y).$$

This mapping underlies the system (O), which can be written as $\psi(x, y) = (-c, b)$. The Jacobian matrix of ψ at (x, y) is equal to

$$H = \begin{pmatrix} -\mu D_x^{-2} & -A^T \\ A & -\mu D_y^{-2} \end{pmatrix}.$$

Assuming x and y are positive, the matrix H is negative definite since for any $w \in R^n$ and $z \in R^m$,

$$(z, w)^T H(z, w) = -\mu(z^T D_x^{-2} z + w^T D_y^{-2} w).$$

In particular, H is nonsingular. It is also interesting to consider a related symmetric matrix

$$\tilde{H} = \begin{pmatrix} -\mu D_x^{-2} & -A^T \\ -A & +\mu D_y^{-2} \end{pmatrix}.$$

Obviously, $\tilde{H}$ is the Hessian matrix of the function

$$L_\mu(x, y) = c^T x + \mu \sum_j \ln x_j - y^T A x - \mu \sum_i \ln y_i + y^T b,$$

which is well defined for $x, y > 0$. Note that L is strictly concave in x for every y and strictly convex in y for every x. The pair $(x(\mu), y(\mu))$ (that is, the point where the gradient of $L(x, y)$ vanishes) constitutes the unique saddle point of $L(x, y)$, in the sense that x is a maximum and y is a minimum.

The sum of logarithms added to the linear objective function $c^T x$ plays the role of a "barrier" [6]. Suppose an algorithm for unconstrained optimization starts in the interior of the feasible domain and iterates by searching a line through the current point. The barrier "forces" the iterates to remain in the interior of the feasible domain. Another classical trick of nonlinear programming is to use a "penalty" function (where a penalty is incurred if a point outside the feasible domain is produced). Let us consider general algorithms that iterate on primal and dual interior points. Let μ denote a parameter that determines primal and dual interior feasible solutions $x(\mu) > 0$ and $y(\mu) > 0$, respectively. Let

$$u(\mu) = b - Ax(\mu) > 0$$

and

$$v(\mu) = A^T y(\mu) - c > 0.$$

If $x(\mu)$ and $y(\mu)$ tend (as μ tends to 0) to optimal solutions of the primal and dual problems, respectively, then necessarily the products $x_j(\mu)v_j(\mu)$ and $y_i(\mu)u_i(\mu)$ tend to 0 with μ. In other words, there exist functions $\mu_i(\mu)$ and $v_j(\mu)$ that tend to zero with μ so that

$$b_i - A_i x = \mu_i(\mu)\frac{1}{y_i}$$

and

$$A_j^T y - c_j = v_j(\mu)\frac{1}{x_j}.$$

The logarithmic barrier function method with uniform weights is *characterized* by the equations

$$\mu_i(\mu) = v_j(\mu) = \mu.$$

With general (not necessarily uniform) weights the functions μ_i and v_j remain linear in μ.

In pursuit of "natural" barrier or penalty functions, let us consider a problem in the following general form:

$$(\mathrm{P}_{f,\mu}) \qquad \begin{aligned} &\text{Maximize } c^T x + \mu \sum_j f(x_j) + \mu \sum_i f(u_i) \\ &\text{subject to } Ax + u = b \end{aligned}$$

where $f(\xi)$ is strictly concave. Let $g(\xi) = f'(\xi)$ and for any d-vector a let

$$G_a = \mathrm{Diag}(g(a_1), \ldots, g(a_d)).$$

A pair (x, u) is optimal for $(\mathrm{P}_{f,\mu})$ if and only if there exists a vector $y \in R^m$ such that

$$\begin{aligned} \mu G_x e \quad - A^T y &= -c, \\ \mu G_u e \quad\;\; - y &= \;\;0. \end{aligned}$$

We would like to have optimality conditions that are "primal-dual symmetric," that is, similar to the system (O) above. More precisely, we are interested in functions $f(\xi)$ where the optimal solution for the approximate dual problem provides Lagrange multipliers supporting the optimal solution of the approximate primal, and vice versa. Such functions would give rise to duality theorems similar to Theorem 8.1. When u is eliminated by the substitution $u_i = g^{-1}(y_i/\mu)$, we obtain a set of equations that we would like to have the same form as $\mu G_x e - A^T y = -c$. In other words, we need the function $g(\xi)$ to satisfy

$$\mu g(\xi) = g^{-1}\left(\frac{\xi}{\mu}\right)$$

for every ξ and $\mu > 0$. The latter requirement is very restrictive. It implies $g(\xi) = g^{-1}(\xi)$ so that $\mu g(\xi) = g(\xi/\mu)$. It follows that $g(\xi) = g(1)/\xi$. We reach the surprising conclusion that the only barrier or penalty functions that are primal-dual symmetric are of the form $f(\xi) = \kappa \ln(|\xi|)$, where κ is some constant. Such functions are appropriate only as barrier functions, that is, for interior point procedures, and not as penalty functions (for exterior point procedures).

We have already argued that for any pair (x^0, y^0) of interior feasible solutions (for (P) and (D), respectively), there exist weights that determine a pair of weighted barrier paths from x^0 and y^0 to the optimal sets. The characterization of these paths is simple. For simplicity of notation, let the indices of columns and rows vary over disjoint sets so that we can use w_i to denote a weight associated with a row and w_j to denote one associated with a column. Given the interior points x^0 and y^0, let

$$w_j = [A_j^T y^0 - c_j]x_j^0$$

and

$$w_i = [b_i - A_i x^0]y_i^0.$$

Then the function

$$c^T x + \mu\left(\sum_j w_j \ln x_j + \sum_i w_i \ln u_i\right)$$

has a maximum over the interior of the primal feasible region. Also, the function

$$b^T y - \mu\left(\sum_i w_i \ln y_i + \sum_j w_j \ln v_j\right)$$

has a minimum over the interior of the dual feasible region. If W is the total of the weights then the gap between the values of the linear functions is equal to $W\mu$. The gap between the values of the nonlinear functions is equal to $W\mu(1 - \ln \mu)$. The paths are characterized by the property that along each of them the products of complementary variables $x_j v_j$ and $y_i u_i$ are proportional to μ. In other words, the ratios across these products are kept constant. More explicitly,

$$x_j v_j = \mu w_j$$

and

$$y_i u_i = \mu w_i$$

along the paths. In the following section we will study the differential descriptions of these paths.

We have argued that a solution path determined by a pair of interior feasible solutions (for the primal and the dual problems) is the locus of interior feasible points with the same ratios across products of complementary variables. This interpretation suggests a natural generalization. Consider the following set of equations:

$$\text{(X)} \qquad \begin{aligned} x_j(A_j^T y - c_j) &= \mu w_j, \\ y_i(b_i - A_i x) &= \mu w_i. \end{aligned}$$

The original problem (P) requires that $Ax < b$ and $x \geq 0$. However, we can consider 2^{n+m} different problems, corresponding to the 2^{n+m} different ways of choosing the restrictions on the signs of the variables x_j and $u_i = b_i - A_i x$. The dual problem to each of these is obtained by suitable changes of sign of the complementary dual variables. For all such pairs of primal and dual problems, the products of complementary variables have to be nonnegative. In other words, if all the products $x_j v_j$ and $y_i u_i$ are nonnegative, then x and y are (respectively) primal and dual feasible solutions for at least one of these pairs of problems. In any case, x and y are feasible solutions of some pair of problems (not necessarily dual) that can be obtained from the original ones by changing the directions of some inequalities. The system (X) defines solution paths for all the feasible combinations of primal and dual problems.

A convenient description of the paths discussed above is obtained as follows. First, consider the problem

$$(\mathrm{P}_\infty) \qquad \begin{aligned} &\text{Maximize } \sum_j \ln x_j + \sum_i \ln u_i \\ &\text{subject to } Ax + u = b, \\ &\qquad\qquad x, u > 0. \end{aligned}$$

which is, in a sense, the limit of (P_μ) as μ tends to infinity. If (P_∞) has an optimal solution x^∞ then $x(\mu)$ tends to x^∞ as μ tends to infinity. Second, consider the problem of minimizing

$$c^T x - \mu\left(\sum_j \ln x_j + \sum_i \ln u_i\right)$$

subject to the same constraints. It is easy to see that as μ tends to infinity the path of the latter also approaches x^∞. It seems nice to apply at this point a change of parameter so that the paths of the two optimization problems can be described in a unified way. Consider the substitution $\mu = \tan\theta$. Equivalently, consider maximizing the following nonlinear objective function:

$$(\cos\theta)c^T x - (\sin\theta)\left(\sum_j \ln x_j + \sum_i \ln u_i\right).$$

For $0 < \theta < \pi/2$ we get the part of the path corresponding to the minimization problem, whereas the interval $\pi/2 < \theta < \pi$ corresponds to the maximization problem. The value $\theta = \pi/2$ corresponds to maximization of the sum of logarithms. If the intersections of level sets of $c^T x$ with the feasible polyhedron are bounded and the linear problem has a minimum then the path is well defined for $0 \le \theta < \pi/2$. If the feasible polyhedron is unbounded then the path is not defined at $\theta = \pi/2$. In fact, it diverges to infinity as θ tends to $\pi/2$. The defining equations of the path have the form

$$\begin{aligned} (\sin\theta)D_x^{-1} \qquad & - A^T y &&= -(\cos\theta)c, \\ Ax \qquad & + (\sin\theta)D_y^{-1} &&= \ (\cos\theta)b. \end{aligned}$$

Again, if the domain is bounded, this system defines a continuous path that leads from a minimum of $c^T x$ to a maximum of $c^T x$ through the maximum of the sum of logarithms.

It is interesting to consider the system discussed above in the neighborhood of $\theta = 0$. We know that the limits $x(0)$ and $y(0)$ exist (if the paths exist). We first prove.

Proposition 8.3. *Let $\bar{x}$, $\bar{y}$, $\bar{u}$, and $\bar{v}$ denote the optimal values of variables in* (P_μ) *and* (D_μ) *at the end of the paths (that is, when μ tends to zero, assuming the problem has an optimal solution). Then, for each pair of complementary variables, $(\bar{x}_i, \bar{u}_i)$ and $(\bar{y}_j, \bar{v}_j)$, one member of the pair is positive while the other equals zero.*

PROOF. Obviously, at least one of the members in each pair equals zero. It is well known that at degenerate vertices some pairs may have both members equal zero. However, degeneracy means that either the primal or the dual problem has an optimal face of dimension greater than zero. We claim that the solution paths converge to points in the relative interior of the optimal faces of the primal and dual problems. Consider, for example, the primal problem. The limit point $\bar{x}$ is where the sum $\sum_j \ln x_j + \sum_i \ln u_i$ (taken over all the variables that are not identically zero on the optimal face) is maximized relative to the optimal face. Obviously, each variable that is not identical to zero on the optimal face does not vanish at $\bar{x}$. This implies our proposition.

Assuming the limits $x(0)$ and $y(0)$ exist, consider the variables that vanish at this point. They also vanish at every other point of the optimal set. Let I denote the set of indices i such that $A_i x = b_i$ at every primal optimal solution x. Also, let J denote the set of indices j such that $x_j = 0$ at every primal optimal solution x. It follows that for every dual optimal solution y, $y_i = 0$ for $i \notin I$ and $A_j^T y = c_j$ for every $j \notin J$. Consider the following problem:

$$
(\tilde{\mathrm{P}}) \qquad
\begin{aligned}
&\text{Minimize } c^T x \\
&\text{subject to } A_i x \ge b_i && (i \in I), \\
&\qquad\qquad\quad A_i x \le b_i && (i \notin I), \\
&\qquad\qquad\quad x_j \le 0 && (j \in J), \\
&\qquad\qquad\quad x_j \ge 0 && (j \notin J).
\end{aligned}
$$

This problem is approximated by

$$
(\tilde{\mathrm{P}}_\mu) \qquad
\begin{aligned}
&\text{Minimize } c^T x - \mu\left(\sum_{j\in J} \ln(-x_j) + \sum_{j\notin J} \ln x_j + \sum_{i\in I} \ln(-u_i) + \sum_{i\notin I} \ln u_i\right) \\
&\text{subject to } Ax + u = b, \\
&\qquad\qquad\quad x_j < 0\ (j \in J),\ x_j > 0\ (j \notin J),\ u_i < 0\ (i \in I),\ u_i > 0\ (i \notin I).
\end{aligned}
$$

It follows that the optimality conditions for $(\tilde{\mathrm{P}}_\mu)$ are the same as those for (P_μ) in the sense that the solution paths (assuming they exist on both sides) can be joined continuously at the optimal face common to problems (P) and $(\tilde{\mathrm{P}})$. Recall that the function $c^T x$ increases monotonically as μ tends to zero. It follows that, as long as the path can be continued, it can be extended through the hyperplane arrangement so that in every cell it travels (monotonically in terms of $c^T x$) from a minimum of the cell to a maximum of the cell, which is also a minimum of an adjacent cell, then to a maximum of this adjacent cell, and so on. The substitution $\mu = \tan\theta$ yields a continuous representation of a combined path that travels through a sequence of bounded cells. Each sequence of bounded cells can be extended on both sides with unbounded cells where the path tends to infinity. Except in pathological cases, the paths do not visit cells in which the function has neither a maximum nor a minimum. A pathological case is, for example, a polyhedral cylinder on which the linear function is unbounded (both from above and from below).

§4. On Tangents to the Paths

Let (x^0, y^0) be a pair of interior feasible solutions (for problems (P) and (D), respectively). Let u^0 and v^0 denote the corresponding slack vectors of the primal and dual problems, respectively. We use the products $w_j = x_j^0 v_j^0$ and $w_i = y_i^0 u_i^0$ to define a pair of paths as explained earlier. Let us now examine the *tangent* to this path at the starting point.

The path is determined by the following equations:

$$
\begin{aligned}
\mu w_j \frac{1}{x_j} \quad - A_j^T y &= -c_j, \\
A_i x \quad + \mu w_i \frac{1}{y_i} &= b_i.
\end{aligned}
$$

Differentiation with respect to μ yields the following equations:

$$\begin{aligned} -\mu w_j \frac{\dot{x}_j}{x_j^2} \quad -A_j^T \dot{y} &= -w_j \frac{1}{x_j}, \\ A_i \dot{x} \quad -\mu w_i \frac{\dot{y}_i}{y_i^2} &= -w_i \frac{1}{y_i}. \end{aligned}$$

Consider a point (x', y') with $x' = x^0 - \delta \dot{x}$ and $y' = y^0 - \delta \dot{y}$, where $\dot{x}$ and $\dot{y}$ constitute the solution of the latter system of equations at $x = x^0$, $y = y^0$, and $\mu = 1$ and δ is any positive number. Obviously, (x', y') lies on the tangent to the curve at x^0 and y^0. It is easy to verify that

$$b^T y' - c^T x' = (1 - \delta)(b^T y - c^T x).$$

Let us denote the slack vectors corresponding to the pair (x', y') by u' and v', and let w_i' and w_j' denote the corresponding products of complementary variables. It follows that

$$\begin{aligned} w_j' &= x_j' v_j' \\ &= (x_j^0 - \delta \dot{x}_j)(v_j^0 - \delta A_j^T \dot{y}) \\ &= w_j \left[1 - \delta + \delta^2 \frac{\dot{x}_j}{x_j^0} \left(1 - \frac{\dot{x}_j}{x_j^0} \right) \right]. \end{aligned}$$

Similarly,

$$w_i' = y_i' u_i' = w_i \left[1 - \delta + \delta^2 \frac{\dot{y}_i}{y_i^0} \left(1 - \frac{\dot{y}_i}{y_i^0} \right) \right].$$

For the points x' and y' to remain feasible in their respective problems, it is necessary and sufficient that the following quantities be less than or equal to 1:

$$\delta \frac{\dot{x}_j}{x_j^0}, \qquad \delta \left(1 - \frac{\dot{x}_j}{x_j^0} \right), \qquad \delta \frac{\dot{y}_i}{y_i^0}, \qquad \delta \left(1 - \frac{\dot{y}_i}{y_i^0} \right).$$

It is interesting to examine properties of the tangents. To establish some connections to other interior point methods, we return for a moment to the problem in standard form as in Section 2. Thus, we now work with the problem in the form

$$\text{(P)} \qquad \begin{aligned} &\text{Maximize } c^T x \\ &\text{subject to } Ax = b, \\ &\qquad x \ge 0. \end{aligned}$$

Given a pair (x^0, y^0) where

$$Ax^0 = b, \qquad x^0 > 0 \quad \text{and} \quad A^T y^0 > c,$$

the pair of paths $x = x(\mu)$, $y = y(\mu)$ through (x^0, y^0) (that is, with $x(1) = x^0$ and $y(1) = y^0$) is determined by the following system of equations:

$$(A_j^T y - c_j)x_j = \mu(A_j^T y^0 - c_j)x_j^0 \qquad (j = 1, \ldots, n),$$
$$Ax = b.$$

By differentiation, at $\mu = 1$ we have the following system (where (x^0, y^0) was replaced by (x, y) for simplicity):

$$(A_j^T y - c_j)\dot{x}_j + x_j A_j^T \dot{y} = (A_j^T y - c_j)x_j \qquad (j = 1, \ldots, n),$$
$$A\dot{x} = 0.$$

Denote, as usual,

$$v_j = A_j^T y - c_j \qquad (j = 1, \ldots, n).$$

Thus,

$$v_j \dot{x}_j + x_j A_j^T \dot{y} = v_j x_j \qquad (j = 1, \ldots, n),$$
$$A\dot{x} = 0.$$

In matrix notation,

$$\begin{pmatrix} D_v & D_x A^T \\ A & 0 \end{pmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} D_v D_x e \\ 0 \end{bmatrix}.$$

Equivalently,

$$\begin{pmatrix} D_v D_x^{-1} & A^T \\ A & 0 \end{pmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} v \\ 0 \end{bmatrix}.$$

Let us substitute ξ_j for $\dot{x}_j$,

$$\xi_j = \sqrt{\frac{v_j}{x_j}}\,\dot{x}_j,$$

and write $D_x^{1/2} = \mathrm{Diag}(x_1^{1/2}, \ldots, x_n^{1/2})$. Thus,

$$\begin{pmatrix} D_v^{1/2} D_x^{-1/2} & A^T \\ A D_v^{-1/2} D_x^{1/2} & 0 \end{pmatrix} \begin{bmatrix} \xi \\ \dot{y} \end{bmatrix} = \begin{bmatrix} v \\ 0 \end{bmatrix}.$$

Finally, this is equivalent to

$$\begin{pmatrix} I & D_v^{-1/2} D_x^{1/2} A^T \\ A D_v^{-1/2} D_x^{1/2} & 0 \end{pmatrix} \begin{bmatrix} \xi \\ \dot{y} \end{bmatrix} = \begin{bmatrix} D_v^{1/2} D_x^{1/2} e \\ 0 \end{bmatrix}.$$

It turns out that the vector $\xi = (\xi_1, \ldots, \xi_n)^T$ is the orthogonal projection of the vector $D_v^{1/2} D_x^{1/2} e$ on the null space of the matrix $A D_v^{-1/2} D_x^{1/2}$. Thus, the interpretation of the direction in terms of x is as follows. Given a pair of primal and dual interior feasible solutions (x, y), the problem (P) is equivalent to the following problem, where z is the optimization variable and x and y are fixed:

$$\text{(P}'\text{)}\qquad \begin{aligned} &\text{Minimize } (A^T y - c)^T z \\ &\text{subject to } Az = 0, \\ &\qquad\quad x - z \geq 0. \end{aligned}$$

The gradient of the objective function is the vector

$$v = A^T y - c.$$

However, the algorithm takes a gradient step only after the following linear transformation has been applied:

$$T(z) = T_{x,y}(z) = D_v^{-1/2} D_x^{-1/2} z.$$

This transformation takes the current point x to the vector of geometric means of the values of complementary variables:

$$x' = T(x) = D_v^{1/2} D_x^{1/2} e,$$

that is,

$$x_j' = \sqrt{v_j x_j} \qquad (j = 1, \ldots, n).$$

The variable z is transformed into

$$\xi = T(z).$$

We thus have an equivalent problem

$$\begin{aligned} &\text{Minimize } (D_v^{1/2} D_x^{1/2} e)^T \xi \\ &\text{subject to } A D_v^{-1/2} D_x^{1/2} \xi = 0, \\ &\qquad\qquad x' - \xi \geq 0. \end{aligned}$$

Here the gradient is the same vector x' of geometric means. The projection of the gradient on the subspace of the feasible directions is as explained above.

§5. Differential Properties of the Solution Paths

In this section we consider higher-order derivatives associated with the curves $x = x(\mu)$ and $y = y(\mu)$ of primal and dual interior feasible solutions discussed in the preceding sections. For convenience, we introduce notation that is usually used in the context of the linear complementarity problem. We denote by $z = z(\mu)$ the $(n + m)$-vector obtained by concatenating $x(\mu)$ and $y(\mu)$, and we also use $s = s(\mu)$ to denote the $(n + m)$-vector obtained by concatenating the slack vectors $v(\mu)$ and $u(\mu)$. Let M denote the matrix

$$M = \begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix}$$

and let q denote the $(n+m)$-vector obtained by concatenating the vectors c and $-b$. Let $\dot{z}_i$ denote the derivative of z (as a function of μ), and $\dot{z} = (\dot{z}_1, \ldots, \dot{z}_{n+m})$. We also extend the arithmetic operations to vectors (applying them component by component) so, for example,

$$\frac{\dot{z}}{z^2} = \left(\frac{\dot{z}_1}{z_1^2}, \ldots, \frac{\dot{z}_{n+m}}{z_{n+m}^2}\right).$$

With the new notation the combined system of primal and dual constraints is the following:

$$s + Mz = -q,$$
$$z, s \geq 0.$$

A pair of optimal solutions is characterized by the complementary slackness conditions

$$s_i z_i = 0.$$

A solution path through a point (z^0, s^0) is determined by the equation

$$\mu\left(\frac{z^0 s^0}{z}\right) + Mz = -q,$$

which we wish to solve for z as μ approaches 0. Let

$$F(z; \mu) = \mu\left(\frac{z^0 s^0}{z}\right) + Mz + q$$

so we wish to solve $F(z; \mu) = 0$.

We can evaluate the derivatives $d^k z/d\mu^k$ by differentiating F. Let $w_i = z_i^0 s_i^0$ and let w denote the vector consisting of the w_i's. First,

$$\frac{dF}{d\mu} = -\mu w \frac{\dot{z}}{z^2} + M\dot{z} + w\frac{1}{z}.$$

Whenever a is a vector, let

$$\Delta(a) = D_a$$

denote as above a diagonal matrix whose diagonal entries are the components of a. It follows that the value of $\dot{z}$ at μ can be obtained by solving the following system of linear equations (where $\dot{z}$ is the unknown, assuming z is known):

$$\left[-\mu\Delta\left(\frac{w}{z^2}\right) + M\right]\dot{z} = -\frac{w}{z}.$$

Second,

$$\frac{d^2 F}{d\mu^2} = 2\mu w \frac{\dot{z}^2}{z^3} - \mu w \frac{\ddot{z}}{z^2} + M\ddot{z} - w\frac{\dot{z}}{z^2} - w\frac{\dot{z}}{z^2}.$$

Thus, the value of $\ddot{z}$ is obtained by solving the following system of linear equations (assuming μ, z, and $\dot{z}$ are known):

$$\left[-\mu\Delta\left(\frac{w}{z^2}\right) + M\right]\ddot{z} = 2w\frac{\dot{z}}{z^2} - 2\mu w\frac{\dot{z}^2}{z^3}.$$

Notice that the coefficient matrix

$$Q = -\mu\Delta\left(\frac{w}{z^2}\right) + M$$

is the same in the equations defining $\dot{z}$ and $\ddot{z}$. It can be shown that for every k, the value of the kth derivative $z^{(k)}$ can be obtained by solving a linear system, where the coefficient matrix is yet the same matrix Q, and the right-hand-side vector is a polynomial in terms of $1/z$ and the derivatives $\dot{z}, \ddot{z}, \ldots, z^{(k-1)}$.

The fact that all the derivatives of z can be evaluated as solutions of linear systems, with the same coefficient matrix, is due to the particular structure of the function F, namely

$$F(z;\mu) = \mu\alpha(z) + \beta(z),$$

where α and β are any C^∞ maps of R^{n+m} into itself. It follows that

$$\frac{dF}{d\mu} = \mu\frac{D\alpha}{Dz}\dot{z} + \frac{D\beta}{Dz}\dot{z} + \alpha(z),$$

so $\dot{z}$ is obtained from the following system:

$$\left[\mu\frac{D\alpha}{Dz} + \frac{D\beta}{Dz}\right]\dot{z} = -\alpha(z).$$

The second derivative has the form

$$\frac{d^2F}{d\mu^2} = \left[\mu\frac{D\alpha}{Dz} + \frac{D\beta}{Dz}\right]\ddot{z} + \mu\alpha_1(z,\dot{z}) + \beta_1(z,\dot{z}),$$

and it can be proved by induction on k that the kth derivative has the form

$$\frac{d^kF}{d\mu^k} = \left[\mu\frac{D\alpha}{Dz} + \frac{D\beta}{Dz}\right]z^{(k)} + \mu\alpha_{k-1}(z,\dot{z},\ldots,z^{(k-1)}) + \beta_{k-1}(z,\dot{z},\ldots,z^{(k-1)}).$$

§6. Behavior Near Vertices

It is convenient to consider in this section the linear programming problem in standard form, that is,

$$\begin{aligned} &\text{Maximize } c^T x \\ &\text{subject to } Ax = b, \\ &\qquad\qquad\quad x \geq 0, \end{aligned}$$

where $A \in R^{m \times n}$ $(m \leq n)$, $x, c \in R^n$, and $b \in R^m$. 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. We denote by $D = D(x)$ 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 x_B and x_N.

We assume that both the primal and dual problems have feasible regions of full dimension. The path is defined whenever a pair of interior feasible solutions for the primal and dual problems is given. Thus, let $x^0 \in R^n$ be such that $Ax^0 = b$ and $x^0 > 0$ and let $y^0 \in R^m$ be such that $A^T y^0 \geq c$. The path starting at (x^0, y^0) is given by the equations

$$\begin{aligned} x_j(A_j^T y - c_j) &= \mu x_j^0(A_j^T y^0 - c_j) \qquad (j = 1, \dots, n), \\ Ax &= b. \end{aligned}$$

It is obvious that for any point on this path, if we "restart" the path according to this definition then nothing changes since the products of complementary variables remain in the same proportions. Let

$$w_j = x_j^0(A_j^T y^0 - c_j).$$

Note that

$$Bx_B + Nx_N = b$$

so

$$x_B = B^{-1}(b - Nx_N).$$

Also, along the path

$$B^T y - c_B = \mu \left(\frac{w_1}{x_1}, \dots, \frac{w_m}{x_m} \right)^T.$$

It is convenient to denote the vector in the right-hand side of the latter by (w_B/x_B). We now have

$$y = B^{-T}\left[\mu \left(\frac{w_B}{x_B} \right) + c_B \right].$$

On the other hand, for every j,

$$x_j = \mu \frac{w_j}{A_j^T y - c_j}.$$

Thus,

$$x_j = \mu \frac{w_j}{A_j^T B^{-T}[\mu(w_B/x_B) + c_B] - c_j} = \mu \frac{w_j}{-\tilde{c}_j + A_j^T B^{-T}\mu(w_B/x_B)}.$$

Suppose $B^{-1}b > 0$ is the unique primal optimal solution and $B^{-1}c_B$ is the unique dual optimal solution, so the paths of the x_j's and the y_i's converge to these points, respectively. Asymptotically, as μ tends to zero, the "nonbasic" variables, that is, $x_j, j = m + 1, \ldots, n$, are

$$x_j \sim \mu \frac{w_j}{A_j^T B^{-T} c_B - c_j} \qquad (j = m + 1, \ldots, n).$$

The denominator in the right-hand is sometimes called the reduced cost with respect to the basis B, that is,

$$\tilde{c}_j = c_j - A_j^T B^{-T} c_B.$$

So,

$$x_j \sim -\mu \frac{w_j}{\tilde{c}_j} \qquad (j = m + 1, \ldots, n).$$

Note that if y^0 is close to the dual optimal solution $B^{-T}c_B$ then we have

$$x_j \sim \mu x_j^0 \qquad (j = m + 1, \ldots, n).$$

In other words, if we start close enough to an optimal solution, the path takes us approximately in a straight line to the optimal solution. This is different from the linear rescaling algorithm where all paths tend to a single direction of approach to the optimal solution [16].

§7. Extensions to the Linear Complementarity Problem

The trajectories described in the preceding sections lend themselves naturally to the general linear complementarity problem (LCP). The problem is as follows. Given a matrix $M \in R^{N \times N}$ and a vector $q \in R^N$, find a $z \in R^N$ such that

$$\text{(LCP)} \qquad \begin{aligned} Mz + q &\geq 0, \\ z &\geq 0, \\ z^T(Mz + q) &= 0. \end{aligned}$$

Note that if z is a solution to (LCP) then for every i, $i = 1, \ldots, N$, the complementarity condition holds:

$$z_i(M_i z + q_i) = 0.$$

It is well known that the (LCP) provides a unifying framework for a large number of problems, including of course the linear programming problem. The generic algorithm for the (LCP) was developed by Lemke [14], generalizing the self-dual simplex method of Dantzig [4]. The book by Garcia and Zangwill

[8] describes the method and the general underlying homotopy principle. The interested reader may refer to this book for more bibliographical notes. The paths described here can also be interpreted as homotopies. We are interested in the behavior of solution paths in some special cases of the (LCP). There has been considerable research on classes of matrices M for which Lemke's algorithm solves the problem. It would be interesting to investigate corresponding classes with respect to the paths described here.

The general idea is a simple generalization of the case of linear programming. Given an *interior* point z^0, that is,

$$z_i^0,\ M_i z^0 + q_i > 0 \qquad (i = 1, \ldots, N),$$

consider the following set of equations:

$$(\mathrm{LCP}(\mu)) \qquad z_i(M_i z + q_i) = \mu z_i^0 (M_i z^0 + q_i) \qquad (i = 1, \ldots, N),$$

where μ is a parameter. Starting at $\mu = 1$, we attempt to drive μ to zero while satisfying LCP(μ). If we succeed then we have solved the problem. However, in general we may generate a path that does not reach the level $\mu = 0$. It may, for example, diverge to infinity as μ approaches a certain positive limit. Moreover, unlike the case of linear programming, the value of μ does not always vary monotonically along a single path described by LCP(μ). Before addressing these issues, let us first consider the basic requirement of existence and uniqueness of a path through a given interior point. Consider the mapping

$$F(z; \mu) = \Delta(z)\Delta(Mz + q) - \mu\Delta(z^0)\Delta(Mz^0 + q),$$

where $\Delta(x) = \mathrm{Diag}(x)$. By classical theory, if the Jacobian matrix of $F(z; \mu)$ is nonsingular at $(z^0; 1)$ then a unique path exists through this point. Let $w = Mz + q$ as usual in the literature on (LCP) and notice that this w is not related to the weights introduced in the context of the linear programming problem. Let $D_z = \mathrm{Diag}(z_1, \ldots, z_N)$ and $D_w = \mathrm{Diag}(w_1, \ldots, w_N)$. It is easy to check that the derivative of $F(z)$ with respect to z is

$$J(z) = D_z M + D_w.$$

Obviously, $J(z)$ is nonsingular if and only if the matrix

$$\tilde{J}(z) = M + D_z^{-1} D_w$$

is nonsingular. In the linear programming problem the matrix $\tilde{J}(z)$ is positive definite (since M is skew-symmetric) and hence nonsingular for every interior point z. Obviously, whenever M is positive semidefinite the matrix $\tilde{J}(z)$ is positive definite at every interior z.

The (LCP) is intimately related to the quadratic programming problem. Consider first the following optimization problem:

$$\text{Minimize } \tfrac{1}{2} z^T M z - q^T z$$

$$\text{subject to } z \geq 0.$$

An approximate problem is

$$\text{Minimize } \tfrac{1}{2}z^T M z - q^T z - \mu \sum_i \ln z_i,$$

where μ is fixed. The necessary conditions for optimality of the approximate problem are

$$-\mu \frac{1}{z_i} + Mz = -q.$$

In other words,

$$z_i(M_i z + q_i) = \mu \qquad (i = 1, \ldots, N).$$

Obviously, we can also incorporate weights ω_i (as we did for the linear programming problem) so that the (LCP) path could start from any interior point and be interpreted as a weighted logarithmic barrier path. Specifically, if z^0 is an interior point then we can define

$$\omega_i = z_i^0(M_i z^0 + q_i)$$

and consider a path of optimal solutions (parametrized by μ) for the problem

$$\text{Minimize } \tfrac{1}{2}z^T M z - q^T z - \mu \sum_i \omega_i \ln z_i.$$

The path is of course described by the following system:

$$-\mu \frac{\omega_i}{z_i} + M_i z = -q_i \qquad i = 1, \ldots, N.$$

It is interesting to write the defining differential equations:

$$\left(\frac{\omega_i}{z_i^2} + M_i\right)\dot{z} = \frac{\omega_i}{z_i},$$

from which it follows that $\dot{z} \neq 0$ along the path. We now consider some special cases:

(1) The matrix M is positive semidefinite.

Here the objective function is convex and the function including the barrier

$$F_{\mu,\omega}(z) = \tfrac{1}{2}z^T M z - q^T z - \mu \sum_i \omega_i \ln z_i$$

is *strictly* convex. Thus, in this case there is at most one optimal solution to the approximate optimization problem, and it is characterized by the equations

$$z_i^T(M_i z + q_i) = \mu\omega_i \qquad (i = 1, \ldots, N).$$

Let $\omega = (\omega_1, \ldots, \omega_N)^T$. If there is an optimal solution for one value of μ then, because of nonsingularity of the Jacobian, the path extends. Moreover, uniqueness implies that μ varies monotonically. Let $z(\mu)$ denote the optimal solution as a function of μ and let

$$V(\mu) = \tfrac{1}{2}(z(\mu))^T(Mz(\mu) - q^T z(\mu)).$$

Obviously, $z(\mu)$ maximizes the sum $\sum_i \omega_i \ln z_i$ over the set

$$S(\mu) = \{z : \tfrac{1}{2}z^T M z - q^T z = V(\mu)\}.$$

(2) The quadratic programming problem.

Consider the following problem

$$\text{(QP)} \qquad \begin{aligned} &\text{Minimize } \tfrac{1}{2}x^T Q x + c^T x \\ &\text{subject to } Ax \geq b, \\ &\qquad\qquad x \geq 0. \end{aligned}$$

The approximate function, using the weighted logarithmic barrier, is

$$F(x) = \tfrac{1}{2}x^T Q x + c^T x - \sum_j \omega_j \ln x_j - \sum_i \omega_i \ln(A_i x - b_i).$$

It is easy to derive optimality conditions using dual variables y_i:

$$-\mu\frac{\omega_j}{x_j} + Q_j x - A^T y = -c_j,$$

$$A_i x - \mu\frac{\omega_i}{y_i} = b_i.$$

As in the usual linear complementarity theory, we obtain a representation of the approximate quadratic programming problem as an approximate complementarity problem with the matrix

$$\begin{pmatrix} Q & -A^T \\ A & 0 \end{pmatrix}$$

and the defining equations are

$$x_j(Q_j x + c_j - A^T y) = \mu\omega_j,$$

$$y_i(A_i x - b_i) = \mu\omega_i.$$

Obviously, if Q is positive semidefinite then the matrix M is positive semidefinite and hence $\tilde{J}(z)$ is positive definite at any interior z. This implies that the paths converge to optimal solutions.

(3) Equilibrium in bimatrix games.

The formulation of the problem of finding an equilibrium point in a bimatrix game as a linear complementarity problem is well known (see [3]). Let

$$\Sigma_n = \{x \in R^n : e^T x = 1, x_j \geq 0\}.$$

Given two matrices A, $B \in R^{m \times n}$, an equilibrium point is a pair of vectors

$x^* \in \Sigma_n$ and $y^* \in \Sigma_m$ such that for every $x \in \Sigma_n$ and $y \in \Sigma_m$,

$$y^T A x^* \geq (y^*)^T A x^*,$$
$$(y^*)^T B x \geq (y^*)^T B x^*.$$

The matrices A and B are assumed without loss of generality to have *positive* entries. The equilibrium conditions are equivalent to

$$A x^* \geq [(y^*)^T A x^*] e,$$
$$B^T y^* \geq [(y^*)^T B x^*] e.$$

By changing variables, one can set the equilibrium problem as follows. If x and y solve the following linear complementarity problem:

$$Ax \geq e, \qquad B^T y \geq 0, \qquad x \geq 0, \qquad y \geq 0,$$
$$x_j(B_j^T y - 1) = y_i(A_i x - 1) = 0,$$

then the normalized vectors

$$x^* = \frac{1}{e^T x} x, \qquad y^* = \frac{1}{e^T y} y$$

constitute an equilibrium point. Thus, the linear complementarity problem arising from bimatrix games has the underlying matrix

$$M = \begin{pmatrix} O & B^T \\ A & O \end{pmatrix}$$

and $q = -e$.

The fundamental equations are the following:

$$x_j(B_j^T y - 1) = w_j \mu,$$
$$y_i(A_i^T x - 1) = w_i \mu,$$

where the w_i's and w_j's are positive and can be chosen to suit the starting point. Since the matrices A and B are positive, it is easy to start the paths. We can choose any $x^0 \in R^n$ and $y^0 \in R^m$ with sufficiently large components that

$$Ax^0 > e, \qquad B^T y^0 > 0, \qquad x^0 > 0, \qquad y^0 > 0$$

and then define

$$w_j = x_j^0(B_j^T y^0 - 1),$$
$$w_i = y_i^0(A_i^T x^0 - 1).$$

Consider the mapping F: $R^{n+m+1} \to R^{n+m}$ defined by

$$F_j(x, y, \mu) = x_j(B_j^T y - 1) - w_j \mu,$$
$$F_i(x, y, \mu) = y_i(A_i^T x - 1) - w_i \mu.$$

The partial derivative of F with respect to (x, y) at a point (x, y, μ) where

$F(x, y, \mu) = 0$ is the following:

$$\begin{pmatrix} \mu D_x^{-1} D_{w_j} & D_x B^T \\ D_y A & \mu D_y^{-1} D_{w_i} \end{pmatrix}.$$

We have not yet studied this matrix to draw conclusions about convergence of paths to equilibrium points.

References

[1] E. R. Barnes, A variation on Karmarkar's algorithm for solving linear programming problems, Research Report No. RC 11136, IBM T. J. Watson Research Center, Yorktown Heights, N.Y. (May 1985).

[2] D. A. Bayer and J. C. Lagarias, The nonlinear geometry of linear programming I: Affine and projective rescaling trajectories, AT&T Preprint.

[3] R. W. Cottle and G. B. Dantzig, Complementary pivot theory of mathematical programming, in *Mathematics of Decision Sciences*, G. B. Dantzig and A. F. Veinott, Jr. (eds.), American Mathematical Society, Providence, R.I., 1968, pp. 115–136.

[4] G. B. Dantzig, *Linear Programming and Extensions*, Princeton University Press, Princeton, N.J., 1963.

[5] B. C. Eaves and H. Scarf, The solution of systems of piecewise linear equations, *Math. Operations Res.* **1** (1976), 1–27.

[6] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968.

[7] K. R. Frisch, The logarithmic potential method of convex programming, unpublished manuscript, University Institute of Economics, Oslo, Norway (1955).

[8] C. B. Garcia and W. I. Zangwill, *Pathways to Solutions, Fixed Points and Equilibria*, Prentice-Hall, Englewood Cliffs, N.J., 1981.

[9] P. E. Gill, W. Murray, M. A. Saunders, J. A. Tomlin, and M. H. Wright, 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, Calif. (July 1985).

[10] H. M. Greenberg, and J. E. Kalan, Methods of feasible paths in nonlinear programming, Technical Report CP 72004, Computer Science/Operations Research Center, Southern Methodist University (February 1972).

[11] P. Huard, Resolution of mathematical programming with nonlinear constraints by the method of centers, in *Nonlinear Programming*, J. Abadie (ed.), North-Holland, Amsterdam, 1967, pp. 207–219.

[12] N. Karmarkar, A new polynomial-time algorithm for linear programming, *Combinatorica* **2** (1984), 373–395.

[13] C. E. Lemke, Bimatrix equilibrium points and mathematical programming, *Management Sci.* **11** (1965), 681–689.

[14] C. E. Lemke, On complementary pivot theory, in *Mathematics of Decision Sciences*, G. B. Dantzig and A. F. Veinott, Jr. (eds.), American Mathematical Society, Providence, R.I., 1968, pp. 95–114.

[15] O. Mangasarian, Normal solutions of linear programs, *Math. Programming Study* **22** (1984), 206–216.

[16] N. Megiddo and M. Shub, Boundary behavior of interior point algorithms for linear programming, *Math. Operations Res.* **13** (1988), to appear.

[17] K. G. Murty, A new interior variant of the gradient projection method for linear

programming, Technical Paper 85-18, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor (May 1985).

[18] J. L. Nazareth, Homotopies in linear programming, *Algorithmica* **1** (1986), 529–536.

[19] H. E. Scarf, The approximation of fixed points of continuous mappings, *SIAM. J. Appl. Math.* **15** (1967), 1328–1343.

[20] S. Smale, Talk at Mathematical Sciences Research Institute (MSRI), Berkeley, California (January 1986).

[21] R. J. Vanderbei, M. J. Meketon, and B. A. Freedman, A modification of Karmarkar's linear programming algorithm, *Algorithmica* **1** (1986), 395–409.