CHAPTER 17

The Central Path

In this chapter, we begin our study of an alternative to the simplex method for solving linear programming problems. The algorithm we are going to introduce is called a *path-following method*. It belongs to a class of methods called *interior-point methods*. The path-following method seems to be the simplest and most natural of all the methods in this class, so in this book we focus primarily on it. Before we can introduce this method, we must define the path that appears in the name of the method. This path is called the *central path* and is the subject of this chapter. Before discussing the central path, we must lay some groundwork by analyzing a nonlinear problem, called the *barrier problem*, associated with the linear programming problem that we wish to solve.

Warning: Nonstandard Notation Ahead

Starting with this chapter, given a lower-case letter denoting a vector quantity, we shall use the upper-case form of the same letter to denote the diagonal matrix whose diagonal entries are those of the corresponding vector. For example,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \Longrightarrow \quad X = \begin{bmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_n \end{bmatrix}.$$

This notation is nonstandard in mathematics at large, but has achieved a certain amount of acceptance in the interior-point-methods community.

1. The Barrier Problem

In this chapter, we consider the linear programming problem expressed, as usual, with inequality constraints and nonnegative variables:

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & Ax \leq b \\ & \quad x \geq 0. \end{array}$$

The corresponding dual problem is

$$\begin{array}{ll} \text{minimize} & b^T y \\ \text{subject to} & A^T y \geq c \\ & \quad\;\; y \geq 0. \end{array}$$

R.J. Vanderbei, *Linear Programming*, International Series in Operations Research & Management Science 196, DOI 10.1007/978-1-4614-7630-6_17,

As usual, we add slack variables to convert both problems to equality form:

$$
\begin{array}{ll}
\text{maximize} & c^T x \\
\text{subject to} & Ax + w = b \\
& x, w \geq 0
\end{array}
\tag{17.1}
$$

and

$$
\begin{array}{ll}
\text{minimize} & b^T y \\
\text{subject to} & A^T y - z = c \\
& y, z \geq 0.
\end{array}
$$

Given a constrained maximization problem where some of the constraints are inequalities (such as our primal linear programming problem), one can consider replacing any inequality constraint with an extra term in the objective function. For example, in (17.1) we could remove the constraint that a specific variable, say, x_j, is nonnegative by adding to the objective function a term that is negative infinity when x_j is negative and is zero otherwise. This reformulation doesn't seem to be particularly helpful, since this new objective function has an abrupt discontinuity that, for example, prevents us from using calculus to study it. However, suppose we replace this discontinuous function with another function that is negative infinity when x_j is negative but is finite for x_j positive and approaches negative infinity as x_j approaches zero. In some sense this smooths out the discontinuity and perhaps improves our ability to apply calculus to its study. The simplest such function is the logarithm. Hence, for each variable, we introduce a new term in the objective function that is just a constant times the logarithm of the variable:

$$
\begin{array}{ll}
\text{maximize} & c^T x + \mu \sum_j \log x_j + \mu \sum_i \log w_i \\
\text{subject to} & Ax + w = b.
\end{array}
\tag{17.2}
$$

This problem, while not equivalent to our original problem, seems not too different either. In fact, as the parameter μ, which we assume to be positive, gets small, it appears that (17.2) becomes a better and better stand-in for (17.1). Problem (17.2) is called the *barrier problem* associated with (17.1). Note that it is not really one problem, but rather a whole family of problems indexed by the parameter μ. Each of these problems is a nonlinear programming problem because the objective function is nonlinear. This nonlinear objective function is called a *barrier function* or, more specifically, a *logarithmic barrier function*.

It is instructive to have in mind a geometric picture of the barrier function. Recall that, for problems expressed in standard form, the set of feasible solutions is a polyhedron with each face being characterized by the property that one of the variables is zero. Hence, the barrier function is minus infinity on each face of the polyhedron. Furthermore, it is finite in the interior of the polyhedron, and it approaches minus infinity as the boundary is approached. Figure 17.1 shows some level sets for the barrier function for a specific problem and a few different choices of μ. Notice that, for each μ, the maximum is attained at an interior point, and as μ gets closer to zero this interior point moves closer to the optimal solution of the original linear programming problem (which is at the top vertex). Viewed as a function of μ, the set of optimal solutions to the barrier problems forms a path through

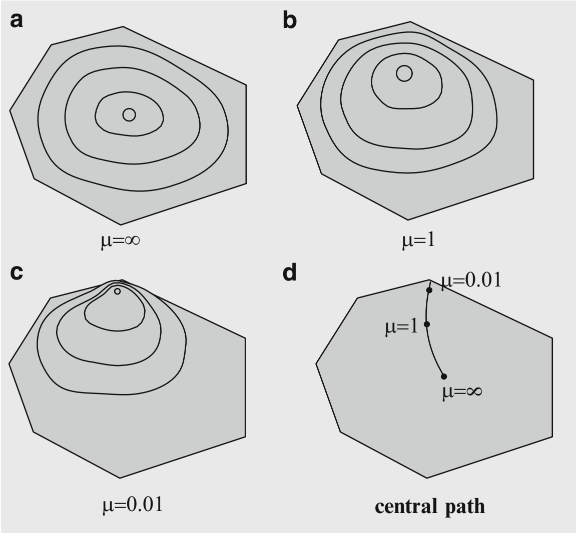

FIGURE 17.1. Parts (**a**) through (**c**) show level sets of the barrier function for three values of μ. For each value of μ, four level sets are shown. The maximum value of the barrier function is attained inside the innermost level set. The drawing in part (**d**) shows the central path.

the interior of the polyhedron of feasible solutions. This path is called the *central path*. Our aim is to study this central path. To this end, we need to develop some machinery, referred to as *Lagrange multipliers*.

2. Lagrange Multipliers

We wish to discuss briefly the general problem of maximizing a function subject to one or more equality constraints. Here, the functions are permitted to be nonlinear, but are assumed to be smooth, say, twice differentiable.

For the moment, suppose that there is a single constraint equation so that the problem can be formally stated as

$$\begin{array}{ll} \text{maximize} & f(x) \\ \text{subject to} & g(x) = 0. \end{array}$$

In this case, the geometry behind the problem is compelling (see Figure 17.2). The gradient of f, denoted ∇f, is a vector that points in the direction of most rapid increase of f. For unconstrained optimization, we would simply set this vector equal to zero to determine the so-called *critical points* of f, and the maximum, if it exists, would have to be included in this set. However, given the constraint, $g(x) = 0$, it is no longer correct to look at points where the gradient vanishes. Instead, the gradient must be orthogonal to the set of feasible solutions $\{x : g(x) = 0\}$. Of course, at each point x in the feasible set, $\nabla g(x)$, is a vector that is orthogonal to the feasible

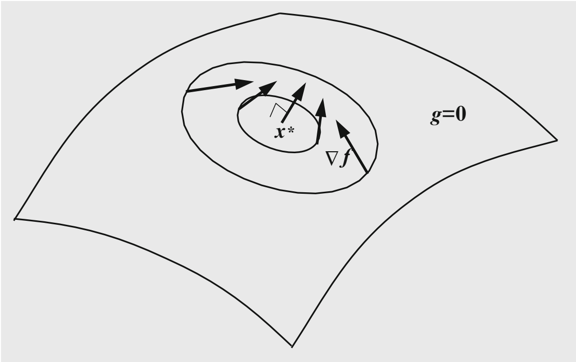

FIGURE 17.2. The concentric rings illustrate a few level sets of f. Clearly, at the optimal solution, x^*, the gradient must be perpendicular to the feasible set.

set at this point x. Hence, our new requirement for a point x^* to be a critical point is that it is feasible and that $\nabla f(x^*)$ be proportional to $\nabla g(x^*)$. Writing this out as a system of equations, we have

$$\begin{aligned} g(x^*) &= 0 \\ \nabla f(x^*) &= y \nabla g(x^*). \end{aligned}$$

Here, y is the proportionality constant. Note that it can be any real number, either positive, negative, or zero. This proportionality constant is called a *Lagrange multiplier*.

Now consider several constraints:

$$\begin{array}{ll} \text{maximize} & f(x) \\ \text{subject to} & g_1(x) = 0 \\ & g_2(x) = 0 \\ & \quad \vdots \\ & g_m(x) = 0. \end{array}$$

In this case, the feasible region is the intersection of m hypersurfaces (see Figure 17.3). The space orthogonal to the feasible set at a point x is no longer a one-dimensional set determined by a single gradient, but is instead a higher-dimensional space (typically m), given by the span of the gradients. Hence, we require that $\nabla f(x^*)$ lie in this span. This yields the following set of equations for a critical point:

$$\begin{aligned} g(x^*) &= 0 \\ \nabla f(x^*) &= \sum_{i=1}^{m} y_i \nabla g(x^*). \end{aligned} \tag{17.3}$$

The derivation of these equations has been entirely geometric, but there is also a simple algebraic formalism that yields the same equations. The idea is to introduce the so-called *Lagrangian* function

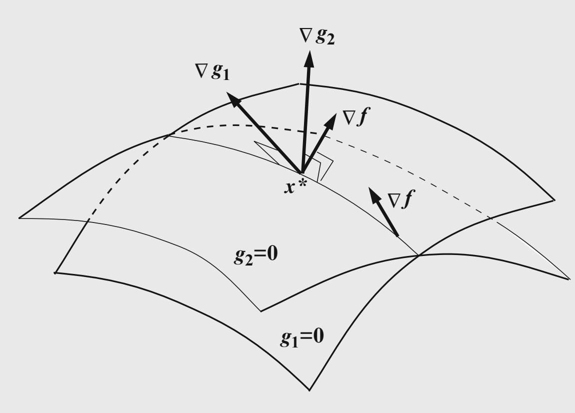

FIGURE 17.3. The feasible set is the curve formed by the intersection of $g_1 = 0$ and $g_2 = 0$. The point x^* is optimal, since the gradient of f at that point is perpendicular to the feasible set.

$$L(x, y) = f(x) - \sum_i y_i g_i(x)$$

and to look for its critical points over both x and y. Since this is now an unconstrained optimization problem, the critical points are determined by simply setting all the first derivatives to zero:

$$\begin{aligned} \frac{\partial L}{\partial x_j} &= \frac{\partial f}{\partial x_j} - \sum_i y_i \frac{\partial g_i}{\partial x_j} &= 0, \qquad & j = 1, 2, \ldots, n, \\ \frac{\partial L}{\partial y_i} &= -g_i &= 0, \qquad & i = 1, 2, \ldots, m. \end{aligned}$$

Writing these equations in vector notation, we see that they are exactly the same as those derived using the geometric approach. These equations are usually referred to as the *first-order optimality conditions*.

Determining whether a solution to the first-order optimality conditions is indeed a global maximum as desired can be difficult. However, if the constraints are all linear, the first step (which is often sufficient) is to look at the matrix of second derivatives:

$$Hf(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right].$$

This matrix is called the *Hessian* of f at x. We have

THEOREM 17.1. *If the constraints are linear, a critical point x^* is a local maximum if*

$$\xi^T Hf(x^*)\xi < 0 \tag{17.4}$$

for each $\xi \neq 0$ *satisfying*

$$\xi^T \nabla g_i(x^*) = 0, \qquad i = 1, 2, \ldots, m. \tag{17.5}$$

PROOF. We start with the two-term Taylor series expansion of f about x^*:

$$f(x^* + \xi) = f(x^*) + \nabla f(x^*)^T \xi + \frac{1}{2}\xi^T H f(x^*)\xi + o(\|\xi\|^2).$$

The vector ξ represents a displacement from the current point x^*. The only displacements that are relevant are those that lie in the feasible set. Hence, let ξ be a direction vector satisfying (17.5). From (17.3) and (17.5), we see that $\nabla f(x^*)^T \xi = 0$, and so

$$f(x^* + \xi) = f(x^*) + \frac{1}{2}\xi^T H f(x^*)\xi + o(\|\xi\|^2).$$

Employing (17.4) finishes the proof. □

It is worth remarking that if (17.4) is satisfied not just at x^* but at all x, then x^* is a unique global maximum.

In the next section, we shall use Lagrange multipliers to study the central path defined by the barrier problem.

3. Lagrange Multipliers Applied to the Barrier Problem

In this section, we shall use the machinery of Lagrange multipliers to study the solution to the barrier problem. In particular, we will show that (subject to some mild assumptions) for each value of the barrier parameter μ, there is a unique solution to the barrier problem. We will also show that as μ tends to zero, the solution to the barrier problem tends to the solution to the original linear programming problem. In the course of our study, we will stumble naturally upon the central path for the dual problem. Taken together, the equations defining the primal and the dual central paths play an important role, and so we will introduce the notion of a primal–dual central path.

We begin by recalling the barrier problem:

$$\begin{array}{ll} \text{maximize} & c^T x + \mu \sum_j \log x_j + \mu \sum_i \log w_i \\ \text{subject to} & Ax + w = b. \end{array}$$

This is an equality-constrained optimization problem, and so it is a problem to which we can apply the Lagrange multiplier tools developed in the previous section. The Lagrangian for this problem is

$$L(x, w, y) = c^T x + \mu \sum_j \log x_j + \mu \sum_i \log w_i + y^T (b - Ax - w).$$

Taking derivatives with respect to each variable and setting them to zero, we get the first-order optimality conditions:

$$\begin{aligned}
\frac{\partial L}{\partial x_j} &= c_j + \mu \frac{1}{x_j} - \sum_i y_i a_{ij} &&= 0, \quad j = 1, 2, \ldots, n, \\
\frac{\partial L}{\partial w_i} &= \mu \frac{1}{w_i} - y_i &&= 0, \quad i = 1, 2, \ldots, m. \\
\frac{\partial L}{\partial y_i} &= b_i - \sum_j a_{ij} x_j - w_i &&= 0, \quad i = 1, 2, \ldots, m.
\end{aligned}$$

Writing these equations in matrix form, we get

$$\begin{aligned}
A^T y - \mu X^{-1} e &= c \\
y &= \mu W^{-1} e \\
Ax + w &= b.
\end{aligned}$$

Here, as warned at the beginning of the chapter, X denotes the diagonal matrix whose diagonal entries are the components of x, and similarly for W. Also, recall that we use e to denote the vector of all ones.

Introducing an extra vector defined as $z = \mu X^{-1} e$, we can rewrite the first-order optimality conditions like this:

$$\begin{aligned}
Ax + w &= b. \\
A^T y - z &= c \\
z &= \mu X^{-1} e \\
y &= \mu W^{-1} e.
\end{aligned}$$

Finally, if we multiply the third equation through by X and the fourth equation by W, we arrive at a primal–dual symmetric form for writing these equations:

$$\begin{aligned}
Ax + w &= b \\
A^T y - z &= c \\
XZe &= \mu e \\
YWe &= \mu e.
\end{aligned} \tag{17.6}$$

Note that the first equation is the equality constraint that appears in the primal problem, while the second equation is the equality constraint for the dual problem. Furthermore, writing the third and fourth equations out componentwise,

$$\begin{aligned}
x_j z_j &= \mu \qquad j = 1, 2, \ldots, n \\
y_i w_i &= \mu \qquad i = 1, 2, \ldots, m,
\end{aligned}$$

we see that they are closely related to our old friend: complementarity. In fact, if we set μ to zero, then they are exactly the usual complementarity conditions that must be satisfied at optimality. For this reason, we call these last two equations the *μ-complementarity* conditions.

The first-order optimality conditions, as written in (17.6), give us $2n + 2m$ equations in $2n + 2m$ unknowns. If these equations were linear, they could be solved using Gaussian elimination, and the entire subject of linear programming would be no more difficult than solving systems of linear equations. But alas, they are nonlinear—but just barely. The only nonlinear expressions in these equations are simple multiplications such as $x_j z_j$. This is about the closest to being linear that one could imagine. Yet, it is this nonlinearity that makes the subject of linear programming nontrivial.

We must ask both whether a solution to (17.6) exists and if so is it unique. We address these questions in reverse order.

4. Second-Order Information

To show that the solution, if it is exists, must be unique, we use second-order information on the barrier function:

$$f(x, w) = c^T x + \mu \sum_j \log x_j + \mu \sum_i \log w_i. \tag{17.7}$$

The first derivatives are

$$\begin{aligned} \frac{\partial f}{\partial x_j} &= c_j + \frac{\mu}{x_j}, & j &= 1, 2, \ldots, n, \\ \frac{\partial f}{\partial w_i} &= \frac{\mu}{w_i}, & i &= 1, 2, \ldots, m, \end{aligned}$$

and the pure second derivatives are

$$\begin{aligned} \frac{\partial^2 f}{\partial x_j^2} &= -\frac{\mu}{x_j^2}, & j &= 1, 2, \ldots, n, \\ \frac{\partial^2 f}{\partial w_i^2} &= -\frac{\mu}{w_i^2}, & i &= 1, 2, \ldots, m. \end{aligned}$$

All the mixed second derivatives vanish. Therefore, the Hessian is a diagonal matrix with strictly negative entries. Hence, by Theorem 17.1, there can be at most one critical point and, if it exists, it is a global maximum.

5. Existence

So, does a solution to the barrier problem always exist? It might not. Consider, for example, the following trivial optimization problem on the nonnegative half-line:

$$\begin{array}{ll} \text{maximize} & 0 \\ \text{subject to} & x \geq 0. \end{array}$$

For this problem, the barrier function is

$$f(x) = \mu \log x,$$

which doesn't have a maximum (or, less precisely, the maximum is infinity which is attained at $x = \infty$). However, such examples are rare. For example, consider modifying the objective function in this example to make $x = 0$ the unique optimal solution:

$$\begin{array}{ll} \text{maximize} & -x \\ \text{subject to} & x \geq 0. \end{array}$$

In this case, the barrier function is

$$f(x) = -x + \mu \log x,$$

which is a function whose maximum is attained at $x = \mu$.

In general, we have the following result:

THEOREM 17.2. *There exists a solution to the barrier problem if and only if both the primal and the dual feasible regions have nonempty interior.*

PROOF. The "only if" part is trivial and less important to us. Therefore, we only prove the "if" part. To this end, suppose that both the primal and the dual feasible regions have nonempty interior. This means that there exists a primal feasible point $(\bar{x}, \bar{w})$ with $\bar{x} > 0$ and $\bar{w} > 0$ and there exists a dual feasible point $(\bar{y}, \bar{z})$ with $\bar{y} > 0$ and $\bar{z} > 0$.[1] Now, given any primal feasible point (x, w), consider the expression $\bar{z}^T x + \bar{y}^T w$. Replacing the primal and dual slack variables with their definitions, we can rewrite this expression as follows:

$$\begin{aligned} \bar{z}^T x + \bar{y}^T w &= \left(A^T \bar{y} - c\right)^T x + \bar{y}^T (b - Ax) \\ &= b^T \bar{y} - c^T x. \end{aligned}$$

Solving this equation for the primal objective function $c^T x$, we get that

$$c^T x = -\bar{z}^T x - \bar{y}^T w + b^T \bar{y}.$$

Therefore, the barrier function f defined in equation (17.7) can be written as follows:

$$\begin{aligned} f(x, w) &= c^T x + \mu \sum_j \log x_j + \mu \sum_i \log w_i \\ &= \quad \sum_j (-\bar{z}_j x_j + \mu \log x_j) \\ &\quad + \sum_i (-\bar{y}_i w_i + \mu \log w_i) \\ &\quad + b^T \bar{y}. \end{aligned}$$

Note that the last term is just a constant. Also, each summand in the two sums is a function of just one variable. These functions all have the following general form:

$$h(\xi) = -a\xi + \mu \log \xi, \qquad 0 < \xi < \infty,$$

where $a > 0$. Such functions have a unique maximum (at μ/a) and tend to $-\infty$ as ξ tends to ∞. From these observations, it is easy to see that, for every constant c, the set

$$\left\{(x, w) \in \mathbb{R}^{n+m} : f(x, w) \geq c\right\}$$

is bounded.

Put

$$\bar{f} = f(\bar{x}, \bar{w})$$

[1] Recall that we write $\xi > 0$ to mean that $\xi_j > 0$ for all j.

and let

$$\bar{P} = \{(x, w) : Ax + w = b, x \geq 0, w \geq 0, f(x, w) \geq \bar{f}\}.$$

Clearly, $\bar{P}$ is nonempty, since it contains $(\bar{x}, \bar{w})$. From the discussion above, we see that $\bar{P}$ is a bounded set.

This set is also closed. To see this, note that it is the intersection of three sets,

$$\{(x, w) : Ax + w = b\} \cap \{(x, w) : x \geq 0, w \geq 0\} \cap \{(x, w) : f(x, w) \geq \bar{f}\}.$$

The first two of these sets are obviously closed. The third set is closed because it is the inverse image of a closed set, $[\bar{f}, \infty]$, under a continuous mapping f. Finally, the intersection of three closed sets is closed.

In Euclidean spaces, a closed bounded set is called compact. A well-known theorem from real analysis about compact sets is that a continuous function on a nonempty compact set attains its maximum. This means that there exists a point in the compact set at which the function hits its maximum. Applying this theorem to f on $\bar{P}$, we see that f does indeed attain its maximum on $\bar{P}$, and this implies it attains its maximum on all of $\{(x, w) : x > 0, w > 0\}$, since $\bar{P}$ was by definition that part of this domain on which f takes large values (bigger than $\bar{f}$, anyway). This completes the proof. □

We summarize our main result in the following corollary:

COROLLARY 17.3. *If a primal feasible set (or, for that matter, its dual) has a nonempty interior and is bounded, then for each* $\mu > 0$ *there exists a unique solution*

$$(x_\mu, w_\mu, y_\mu, z_\mu)$$

to (17.6).

PROOF. Follows immediately from the previous theorem and Exercise 10.7. □

The path $\{(x_\mu, w_\mu, y_\mu, z_\mu) : \mu > 0\}$ is called the *primal–dual central path*. It plays a fundamental role in interior-point methods for linear programming. In the next chapter, we define the simplest interior-point method. It is an iterative procedure that at each iteration attempts to move toward a point on the central path that is closer to optimality than the current point.

Exercises

17.1 Compute and graph the central trajectory for the following problem:

$$\begin{array}{lrcr} \text{maximize} & -x_1 + x_2 & & \\ \text{subject to} & x_2 & \leq & 1 \\ & -x_1 \quad & \leq & -1 \\ & x_1,\ x_2 & \geq & 0\,. \end{array}$$

Hint: The primal and dual problems are the same—exploit this symmetry.

17.2 Let θ be a fixed parameter, $0 \leq \theta \leq \frac{\pi}{2}$, and consider the following problem:

$$\begin{array}{ll} \text{maximize} & (\cos\theta)x_1 + (\sin\theta)x_2 \\ \text{subject to} & x_1 \leq 1 \\ & x_2 \leq 1 \\ & x_1, x_2 \geq 0. \end{array}$$

Compute an explicit formula for the central path $(x_\mu, w_\mu, y_\mu, z_\mu)$, and evaluate $\lim_{\mu\to\infty} x_\mu$ and $\lim_{\mu\to 0} x_\mu$.

17.3 Suppose that $\{x : Ax \leq b, x \geq 0\}$ is bounded. Let $r \in \mathbb{R}^n$ and $s \in \mathbb{R}^m$ be vectors with positive elements. By studying an appropriate barrier function, show that there exists a unique solution to the following nonlinear system:

$$\begin{aligned} Ax + w &= b \\ A^T y - z &= c \\ XZe &= r \\ YWe &= s \\ x, y, z, w &> 0. \end{aligned}$$

17.4 Consider the linear programming problem in equality form:

$$\begin{array}{lll} & \text{maximize} & \sum_j c_j x_j \\ (17.8) & \text{subject to} & \displaystyle\sum_j a_j x_j = b \\ & & x_j \geq 0, \qquad j = 1, 2, \ldots, n, \end{array}$$

where each a_j is a vector in $\mathbb{R}^m$, as is b. Consider the change of variables,

$$x_j = \xi_j^2,$$

and the associated maximization problem:

$$\begin{array}{lll} (17.9) & \text{maximize} & \sum_j c_j \xi_j^2 \\ & \text{subject to} & \sum_j a_j \xi_j^2 = b \end{array}$$

(note that the nonnegativity constraints are no longer needed). Let V denote the set of basic feasible solutions to (17.8), and let W denote the set of points $(\xi_1^2, \xi_2^2, \ldots, \xi_n^2)$ in $\mathbb{R}^n$ for which $(\xi_1, \xi_2, \ldots, \xi_n)$ is a solution to the first-order optimality conditions for (17.9). Show that $V \subset W$. What does this say about the possibility of using (17.9) as a vehicle to solve (17.8)?

Notes

Research into interior-point methods has its roots in the work of Fiacco and McCormick (1968). Interest in these methods exploded after the appearance of the seminal paper Karmarkar (1984). Karmarkar's paper uses clever ideas from *projective geometry*. It doesn't mention anything about central paths, which have become fundamental to the theory of interior-point methods. The discovery that Karmarkar's algorithm has connections with the primal–dual central path introduced in this chapter can be traced to Megiddo (1989). The notion of central points can be traced to pre-Karmarkar times with the work of Huard (1967). D.A. Bayer and J.C. Lagarias, in a pair of papers (Bayer and Lagarias 1989a,b), give an in-depth study of the central path.

Deriving optimality conditions and giving conditions under which they are necessary and sufficient to guarantee optimality is one of the main goals of *nonlinear programming*. Standard texts on this subject include the books by Luenberger (1984), Bertsekas (1995), and Nash and Sofer (1996).