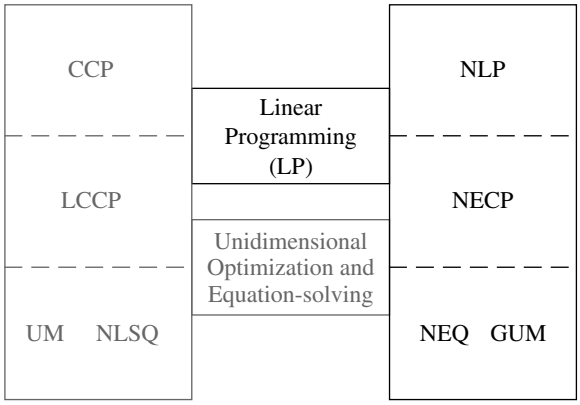
Part IV

Linear Programming Post-Karmarkar



10

LP from the Euler–Newton Perspective

We consider again the primal (P) and dual (D) linear programs (8.1) of Chapter 8. The assumptions stated after expressions (8.1) ensure that optimal solutions of the primal, say $\mathbf{x}^*$, and dual, say $(\pi^*, \mathbf{v}^*)$, simultaneously exist and satisfy primal feasibility, dual feasibility, and complementary slackness; i.e., they satisfy the well-known Karush–Kuhn–Tucker (KKT) optimality conditions:

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b}, \\ \mathbf{A}^T\pi + \mathbf{v} &= \mathbf{c}, \\ \mathbf{X}\mathbf{V}\mathbf{e} &= \mathbf{0}, \end{aligned} \tag{10.1}$$

and $\mathbf{x} \in R^n_+$, $\mathbf{v} \in R^n_+$. In the third set of equations,

$$\mathbf{X} = \text{diag}[x_1, \ldots, x_n], \qquad \mathbf{V} = \text{diag}[v_1, \ldots, v_n],$$

and $\mathbf{e}$ is an n-vector of 1's. The system (10.1) represents a set of $(m + 2n)$ linear and nonlinear equations in $(m+2n)$ variables $\mathbf{w} = (\mathbf{x}, \pi, \mathbf{v})$. Feasible solutions of the linear programs must also satisfy the nonnegativity bounds on $\mathbf{x}$ and $\mathbf{v}$.

The optimal solutions of (P) and (D) are unique under assumptions of nondegeneracy. Such assumptions are not needed here, but they can be made, if desired, for convenience of discussion.

10.1 The Parameterized Homotopy System

We build on the fundamental EN approach of Section 3.2 and the global homotopy (3.10), now applied to the system (10.1).

10.1.1 Interior Paths

Given initial interior vectors $\mathbf{x}^{(0)} > \mathbf{0}$, $\pi^{(0)}$, and $\mathbf{v}^{(0)} > \mathbf{0}$, which are permitted to be infeasible; i.e., they need not satisfy the equality constraints of (P) or (D), define the global homotopy parameterization of the equations (10.1) as follows:

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} + \mu\mathbf{q}_{x^{(0)}}, \\ \mathbf{A}^T\pi + \mathbf{v} &= \mathbf{c} + \mu\mathbf{q}_{v^{(0)}}, \\ \mathbf{X}\mathbf{V}\mathbf{e} &= \mu\mathbf{X}_0\mathbf{V}_0\mathbf{e}, \end{aligned} \tag{10.2}$$

where $\mathbf{x}, \mathbf{v} \in R^n_+$ and $\mu \in [0, 1]$ is a parameter. In the foregoing equations (10.2), $\mathbf{q}_{x^{(0)}} = \mathbf{A}\mathbf{x}^{(0)} - \mathbf{b}$, $\mathbf{q}_{v^{(0)}} = \mathbf{A}^T\pi^{(0)} + \mathbf{v}^{(0)} - \mathbf{c}$, $\mathbf{X}_0 = \operatorname{diag}\left[x_1^{(0)}, \ldots, x_n^{(0)}\right]$, and $\mathbf{V}_0 = \operatorname{diag}\left[v_1^{(0)}, \ldots, v_n^{(0)}\right]$. When $\mu = 1$, a solution of (10.2) is obviously the given initial point $\mathbf{w}^{(0)} = \left(\mathbf{x}^{(0)}, \pi^{(0)}, \mathbf{v}^{(0)}\right)$, and when $\mu = 0$, a solution of (10.2) with $\mathbf{x}, \mathbf{v} \in R^n_+$ is an optimal solution of (P) and (D) in (8.1), say $\mathbf{w}^* = (\mathbf{x}^*, \pi^*, \mathbf{v}^*)$. For each $\mu > 0$, a *unique* solution of (10.2) exists that lies in the nonnegative orthant with respect to the $\mathbf{x}$ and $\mathbf{v}$ variables. A convenient proof is given in the next chapter—see the discussion immediately following expression (11.4)—and an alternative proof is outlined in the paragraphs following expression (10.4) below.

The set D of Section 3.2.1 is now given by vectors $\mathbf{w} = (\mathbf{x}, \pi, \mathbf{v})$ with $\mathbf{x} \in R^n_+$, $\pi \in R^m$, and $\mathbf{v} \in R^n_+$. Note that D is closed but unbounded; i.e., it is not compact. When $\mathbf{x}$ and $\mathbf{v}$ have strictly positive components, then $\mathbf{w}$ is in D^0, the interior of D. Otherwise, it lies on the boundary ∂D. In addition, we define the following quantities for the homotopy system (10.2):

1. For each $\mu > 0$, let us denote the unique solution of (10.2) by $\mathbf{w}(\mu) = (\mathbf{x}(\mu), \pi(\mu), \mathbf{v}(\mu))$. Note, in particular, that $\mathbf{w}(1) = \mathbf{w}^{(0)}$ and $\mathbf{w}(0) = \mathbf{w}^*$.
2. The scalar quantity given by

$$\frac{\max_i \left(x_i^{(0)} v_i^{(0)}\right)}{\min_i \left(x_i^{(0)} v_i^{(0)}\right)} \tag{10.3}$$

will be called the *condition number* associated with the homotopy system (10.2).

3. Let us denote the homotopy system of equations (10.2) more compactly by $\mathbf{H}(\mathbf{w}, \mu) = \mathbf{0}$. Its Jacobian matrix, namely, the matrix of partial derivatives of $\mathbf{H}(\mathbf{w}, \mu)$ with respect to $\mathbf{w}$, is

$$\mathbf{H}'_w = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^T & \mathbf{I} \\ \mathbf{V} & \mathbf{0} & \mathbf{X} \end{bmatrix}. \tag{10.4}$$

   When $\mathbf{A}$ is of full rank and the matrices $\mathbf{X}$ and $\mathbf{V}$ have positive diagonal elements, then $\mathbf{H}'_w$ is *nonsingular*.

Now, intuitively, it is plausible that the points $\mathbf{w}(\mu) = (\mathbf{x}(\mu), \pi(\mu), \mathbf{v}(\mu))$, $1 \geq \mu \geq 0$, form a *unique differentiable path* of solutions of the homotopy system (10.2), leading from the given point $\mathbf{w}^{(0)} = \left(\mathbf{x}^{(0)}, \pi^{(0)}, \mathbf{v}^{(0)}\right)$ to a solution of the original system. Moreover, the assumption that $\mathbf{w}^{(0)}$ is in the positive orthant with respect to the variables $\mathbf{x}$ and $\mathbf{v}$ implies that $\mu\mathbf{X}_0\mathbf{V}_0\mathbf{e} > \mathbf{0}, \forall\mu > 0$, and then the form of the third set of equations in (10.2) suggests that the *entire* path $\mathbf{w}(\mu)$, $\mu \in (0, 1]$, remains in the positive orthant with respect to $\mathbf{x}$ and $\mathbf{v}$. (The exception is the terminal point corresponding to $\mu = 0$, which can lie on the boundary.) To be more specific, suppose that a component of the variables $\mathbf{x}$ associated with a point on the path, say the ith, crosses its axis and thus takes on the value zero. Then the ith component of the variables $\mathbf{v}$ must assume an infinite value. *But this cannot happen if the path is shown to be bounded.* A similar argument applies with the roles of $\mathbf{x}$ and $\mathbf{v}$ reversed. The third set of equations of the square system (10.2) defining the path thus serves as a "proxy" for the nonnegativity bounds on $\mathbf{x}$ and $\mathbf{v}$ in the optimality conditions (10.1).

Path existence can be formalized following a line of verification that is standard in the theory of homotopy-based methods, which is based on the implicit function theorem as outlined in Section 3.2, and the subsidiary proof of boundedness can be obtained by reduction to a theorem due to Dikin [1974]. See Nazareth [1991, Section 2.2] for an example of this type of argument when the starting vectors are feasible; i.e., $\mathbf{q}_{x^{(0)}} = \mathbf{0}$ and $\mathbf{q}_{v^{(0)}} = \mathbf{0}$. The extension to the more general case can be achieved along similar lines. Such proofs are facilitated by log-barrier transformations of the original linear programs and their relationship to (10.2), which are discussed in the next chapter, Section 11.1. But detailed demonstrations of path existence and smoothness will not be pursued here.

In addition to definitions itemized above, we will use the following definition concerning paths:

4. The scalar $\mathbf{x}(\mu)^T\mathbf{v}(\mu)$ will be called the *"duality gap"* associated with the point $(\mathbf{w}(\mu), \mu)$ on the homotopy path. It follows from (10.2) that

$$\mathbf{x}(\mu)^T\mathbf{v}(\mu) = (\mathbf{c} + \mu\mathbf{q}_{v^{(0)}})^T\mathbf{x}(\mu) - (\mathbf{b} + \mu\mathbf{q}_{x^{(0)}})^T\pi(\mu).$$

If the components of $\mathbf{w}(\mu)$ are feasible for (P) and (D); i.e., if $\mathbf{q}_{x^{(0)}} = \mathbf{0}$ and $\mathbf{q}_{v^{(0)}} = \mathbf{0}$, then $\mathbf{x}(\mu)^T\mathbf{v}(\mu)$ is a true duality gap for the linear programs (P) and (D). Otherwise, it represents a duality gap for a linear program with suitably modified objective and right-hand side vectors. Also, by summing the third set of equations in (10.2), it follows that

$$\mu = \frac{\mathbf{x}(\mu)^T\mathbf{v}(\mu)}{\left(\mathbf{x}^{(0)}\right)^T\mathbf{v}^{(0)}}. \tag{10.5}$$

This corresponds to the choice $\mathbf{z} = \begin{bmatrix}\mathbf{0}\\ \mathbf{e}\end{bmatrix} \in R^{m+2n}$ in expression (3.12), where $\mathbf{0} \in R^{m+n}$ and $\mathbf{e} \in R^n$ is a vector of 1's.

When the LP matrix $\mathbf{A}$ is of full rank and $\mathbf{w}(\mu)$, $\mu \neq 0$, lies within the positive orthant R^n_{++} with respect to $\mathbf{x}$ and $\mathbf{v}$ then, as noted under item 3 above, the Jacobian matrix (10.4) is *nonsingular*. This, in turn, implies that *the homotopy path has no turning points*, which greatly facilitates path-following strategies to be discussed shortly, in Section 10.2. Each point $(\mathbf{w}(\mu), \mu)$ on the homotopy path is defined by a unique μ, which, from (10.5), is the ratio of the duality gap for that point and the duality gap for the initial point. Thus, $\mu \downarrow 0$ is equivalent to saying that the duality gap associated with $(\mathbf{w}(\mu), \mu)$ tends to zero.

With each point $(\mathbf{w}(\mu), \mu)$, $\mu > 0$, on the path, define a condition number

$$\max_i(x_i(\mu)v_i(\mu))/\min_i(x_i(\mu)v_i(\mu)).$$

The third set of equations in (10.2) implies that the foregoing condition number, for any value of $\mu > 0$, is equal to the condition number (10.3) defined from the path's initial point $(\mathbf{w}(1), 1)$; i.e., the condition number associated with the homotopy system (10.2) can be obtained from any other interior point on the path.

Note that μ need not be confined to the interval $[0, 1]$. As μ is increased beyond 1, the homotopy system (10.2) is deformed away from optimal solutions of the original linear programs. Continuation of μ below zero yields solutions to (10.2) that lie outside the nonnegative $(\mathbf{x}, \mathbf{v})$ orthant. The latter noninterior case presents some interesting issues that will be pursued only very briefly in the present monograph, in Section 10.1.2.

Categorization of Interior Paths

We will refine the path terminology henceforth as follows. The path associated with the system (10.2) will be called an *interior homotopy path*. Note that it lies in a space of dimension $(m + 2n + 1)$ corresponding to the variables $(\mathbf{w}, \mu)$. Its projection into the space of variables $\mathbf{w}$ of dimension $(m + 2n)$ is called a *lifted* interior path. When the context is clear, we will

often say "interior path" or "general path" or simply "path" when referring to an interior homotopy path or to its lifted counterpart.

We emphasize that points $\mathbf{w}(\mu)$ on a (lifted) interior path are *not* required to be feasible; i.e., the components of $\mathbf{w}^{(0)}$ are not required to satisfy the equality constraints of (P) and (D). For this case, it is easy to ensure that all components of the vector $\mathbf{X}_0\mathbf{V}_0\mathbf{e}$ in (10.2) are equal, which implies that the homotopy system (10.2) and associated path have the optimal condition number of unity. There are two simple ways to achieve this property, namely:

- Given any $\mathbf{x}^{(0)} > \mathbf{0}$, let $v_i^{(0)} = 1/x_i^{(0)}, \forall i$ (or vice versa), and let $\pi^{(0)} = \mathbf{0}$. We will call the associated lifted path an *optimally conditioned* (infeasible) *interior path*.
- A special case is to choose $\mathbf{x}^{(0)} = \mathbf{v}^{(0)} = \mu^{(0)}\mathbf{e}$, $\mu^{(0)} > 0$, and $\pi^{(0)} = \mathbf{0}$. The corresponding lifted path of solutions of (10.2) will then be termed a *central ray-initiated* (optimally conditioned, infeasible) *interior path*.

When the initial point $\mathbf{w}^{(0)}$ is *feasible*, the path associated with (10.2) is a *feasible interior homotopy path*, and the corresponding lifted path, which consists of feasible points, is a *feasible interior path*.

Does there exist an optimally conditioned feasible interior path? The answer is yes, and this *unique* path is called the *central path* of the linear program. A very convenient way to establish its properties of existence and uniqueness is via the log-barrier transformations of the next chapter; see Section 11.1.1 and the discussion after expression (11.4). The central path is a fundamental object of linear programming. It was introduced in an alternative setting in Section 8.4 and will be discussed further in Section 10.2. We will let the context determine whether the term "central path" applies to the homotopy path in a space of dimension $(m + 2n + 1)$ or to its lifted counterpart in the primal–dual space (or even its restriction to the space of primal or dual variables).

We illustrate the paths defined above using a simple example, similar to the one given at the beginning of Chapter 7 and depicted in Figure 7.1. (The axes are reordered here, for convenience). The corresponding dual linear program has variables $\pi \in R$ and $\mathbf{v} \in R^3$. The feasible region for the π variable is the half-line $(-\infty, \pi^*)$, where π^* is the optimal value, and the feasible region for the $\mathbf{v}$ variables is also a half-line in R^3_+. The feasible regions for both primal and dual are depicted in Figure 10.1, with the π and $\mathbf{v}$ variables shown separately. The $\mathbf{w}$-space has 7 dimensions. A (general) homotopy path is projected separately into the $\mathbf{x}$, π, and $\mathbf{v}$ spaces, and the resulting lifted paths (G) are depicted. Lifted paths are also shown for several other cases: the central path (C), a feasible interior path (F), a central ray-initiated interior path (Ia), and an optimally conditioned interior path (Ib).

For some additional detail on interior path categorization, see Nazareth [1998].

10.1.2 Noninterior Paths

Now consider an *arbitrary* initial point $\mathbf{w}^{(0)}$; i.e., its components $\mathbf{x}^{(0)}$ and $\left(\pi^{(0)}, \mathbf{v}^{(0)}\right)$ are not required to have positive elements or satisfy the equality constraints of (P) or (D). Let us choose positive constants ω_i and γ_i, $i = 1, \ldots, n$, and define associated quantities

$$x_i^{+(0)} = x_i^{(0)} + \omega_i, \qquad v_i^{+(0)} = v_i^{(0)} + \gamma_i, \qquad i = 1, \ldots, n, \tag{10.6}$$

so that $\mathbf{x}^{+(0)} > \mathbf{0}$, $\mathbf{v}^{+(0)} > \mathbf{0}$, and the condition number of the diagonal matrix

$$\text{diag}\left[x_1^{+(0)} v_1^{+(0)}, \ldots, x_n^{+(0)} v_n^{+(0)}\right]$$

is bounded by any desired constant, say C. Let ω and γ be n-vectors with components ω_i and γ_i.

We can now define a homotopy system associated with the starting point $\mathbf{w}^{(0)}$ as follows:

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} + \mu \mathbf{q}_{x^{(0)}}, \\ \mathbf{A}^T\pi + \mathbf{v} &= \mathbf{c} + \mu \mathbf{q}_{v^{(0)}}, \\ \mathbf{X}^+\mathbf{V}^+\mathbf{e} &= \mu \mathbf{X}_0^+ \mathbf{V}_0^+ e, \\ \mathbf{x}^+ &= \mathbf{x} + \mu\omega, \\ \mathbf{v}^+ &= \mathbf{v} + \mu\gamma, \end{aligned} \tag{10.7}$$

where $\mathbf{x}^+, \mathbf{v}^+ \in R_+^n$ and $\mu \in [0, 1]$. The components of the n-vectors $\mathbf{x}^+$ and $\mathbf{v}^+$ are newly introduced variables x_i^+ and v_i^+, respectively, and the quantities $\mathbf{X}^+$, $\mathbf{V}^+$, $\mathbf{X}_0^+$, and $\mathbf{V}_0^+$ are diagonal matrices with corresponding diagonal elements x_i^+, v_i^+, $x_i^{+(0)}$, and $v_i^{+(0)}$, $i = 1, \ldots, n$. As in (10.2), the vectors $\mathbf{q}_{x^{(0)}}$ and $\mathbf{q}_{v^{(0)}}$ are defined to be $\mathbf{q}_{x^{(0)}} = \mathbf{A}\mathbf{x}^{(0)} - \mathbf{b}$ and $\mathbf{q}_{v^{(0)}} = \mathbf{A}^T\pi^{(0)} + \mathbf{v}^{(0)} - \mathbf{c}$.

When $\mu = 1$, the homotopy system (10.7) has a solution $\left(\mathbf{x}^{(0)}, \pi^{(0)}, \mathbf{v}^{(0)}\right)$ along with the vectors $\mathbf{x}^{+(0)}, \mathbf{v}^{+(0)}$ defined by (10.6). When $\mu = 0$, the system (10.7) is equivalent to the original KKT equations. For $0 \leq \mu \leq 1$, it defines a *noninterior homotopy path*. Using the illustrative example of the previous subsection, the associated noninterior lifted path is also depicted in Figure 10.1; see the path identified by N.

Observe that the foregoing homotopy system (10.7) can be converted to a system identical in structure to (10.2) by eliminating the original variables $\mathbf{x}$ and $\mathbf{v}$ to obtain

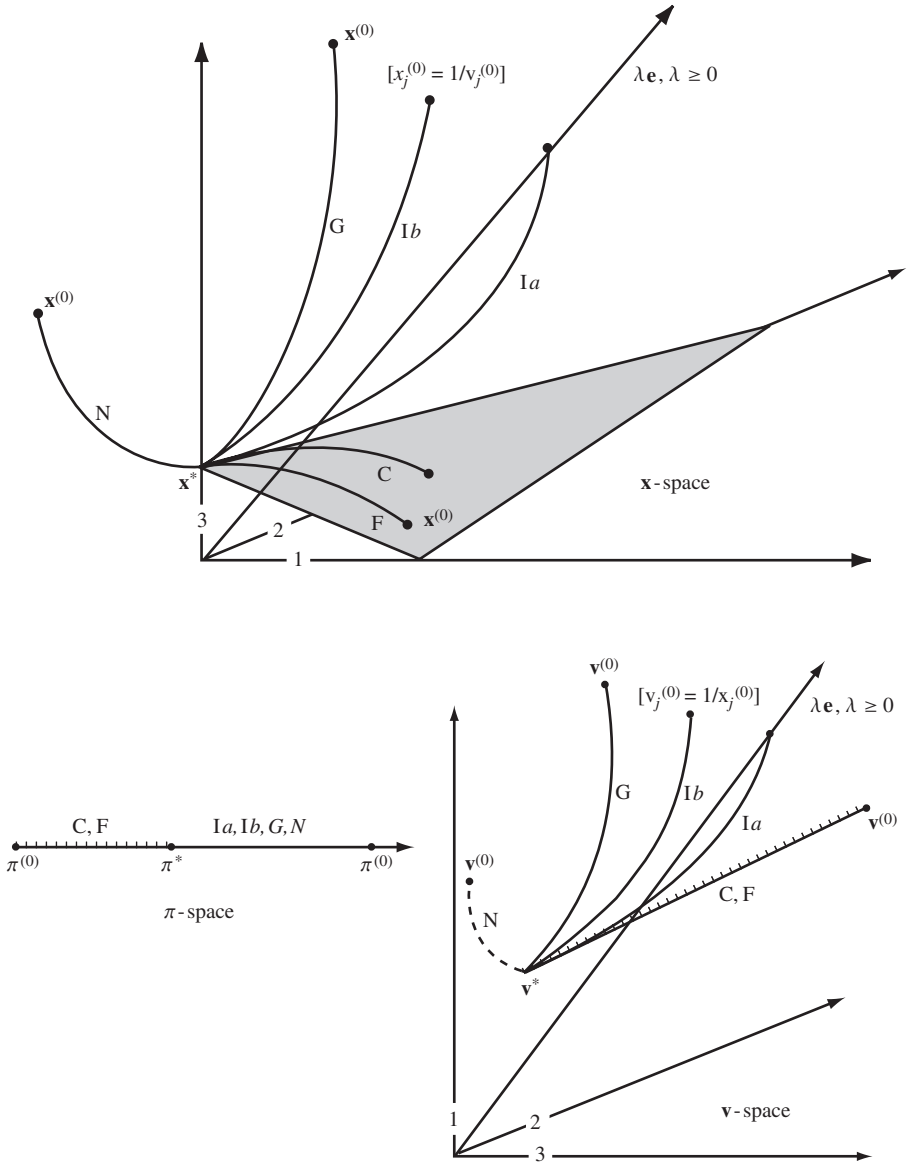


FIGURE 10.1 Illustration of paths.

$$\begin{aligned} \mathbf{A}\mathbf{x}^{+} &= \mathbf{b} + \mu \mathbf{q}^{+}_{x^{(0)}}, \\ \mathbf{A}^{T}\pi + \mathbf{v}^{+} &= \mathbf{c} + \mu \mathbf{q}^{+}_{v^{(0)}}, \\ \mathbf{X}^{+}\mathbf{V}^{+}\mathbf{e} &= \mu \mathbf{X}^{+}_{0}\mathbf{V}^{+}_{0} e, \end{aligned} \tag{10.8}$$

and $\mathbf{x}^+, \mathbf{v}^+ \in R^n_+$. The vectors in the right-hand sides of the first two equations are given by $\mathbf{q}^+_{x^{(0)}} = \mathbf{q}_{x^{(0)}} + \mathbf{A}\omega$ and $\mathbf{q}^+_{v^{(0)}} = \mathbf{q}_{v^{(0)}} + \gamma$. This transformation is useful for purposes of mathematical analysis, because all previous results of Section 10.1.1 along with subsequent discussion on path-following in Section 10.2 can be carried over to the noninterior case. For conceptual and implementational purposes, however, the form (10.7) may be preferable.

We do not consider the noninterior case in any detail in this monograph. For some related references, see, for example, Chen and Harker [1993] and Burke and Xu [2000].

10.2 Path-Following Building Blocks

We now turn to the formulation of algorithms for the system (10.2) premised on the homotopy method of Sections 3.2.2–3.2.4.

From a conceptual standpoint, homotopy, or path-following,[1] techniques lie at the foundation of post-Karmarkar interior-point methods for linear programming. They have found *fundamental new expression* within the linear programming context, and these LP adaptations and developments, in turn, have considerably reilluminated usage within the traditional general *nonlinear* equations setting. The situation is analogous to that for classical log-barrier techniques of nonlinear programming, which also play a fundamental role and have found fresh expression within the LP setting, as we shall see in the next chapter.

The homotopy system (10.2) is a mixture of $(n + m)$ linear equations and n nonlinear equations that have special structure. These particular $(2n + m)$ equations, in turn, give the path-following approach some very distinctive characteristics in the LP setting. The most prominent of these characteristics is the existence and special role of the *central path*: the unique optimally conditioned lifted path within the feasible polytope of a linear program and its dual. One must be careful, however, not to overly diminish the status of other homotopy paths discussed earlier (and their lifted counterparts), because they can play a role very similar to that of the central path within a path-following algorithm, as will be seen in subsequent discussion in this chapter. Put another way, the central path is a fundamental object of linear programming, but only a *first among equals*.

As discussed in Section 10.1.1, LP homotopy interior paths have no turning points when $\mathbf{A}$ is of full rank, and they are relatively simple to characterize. Convenient expressions can be found for their tangents and curvature, which will be given in Sections 10.2.2–10.2.3. When devising path-following algorithms, one can draw on the extensive knowledge of homotopy tech-

[1] We will use these two names interchangeably in this monograph.

niques surveyed in Chapter 3, in particular, techniques of the predictor–corrector variety. The latter are capable of significant enhancement in the LP setting, because of the linearity/nonlinear structure described above. These techniques are the "fundamentally different algorithms" of the second quotation in Section 3.2.4 and the subsequent considerations of that section.

We now discuss the set of *simple building blocks* from which such path-following algorithms can be composed.

10.2.1 Notation

For direction vectors below, we use the symbols $\mathbf{p}$ for predictor, $\mathbf{s}$ for second-order corrector, $\mathbf{n}$ for Newton corrector, and $\mathbf{c}$ for centering component.

Points or iterates of an algorithm that lie *off* the global homotopy path that is being followed are distinguished by an overbar, for example, $\overline{\mathbf{w}}$ or $\overline{\mathbf{w}}^{(k)}$. Points that lie *on* the path always have a superscript or a parameter μ associated with them, and the overbar is omitted, for example, $\mathbf{w}^{(0)}$ or $\mathbf{w}(\mu)$.

An arbitrary point of the primal–dual space of variables is denoted by $\mathbf{w} = (\mathbf{x}, \pi, \mathbf{v})$. The interior of the primal and dual feasible polytopes is denoted by $\mathcal{F}^0$.

We use $\mathbf{q}$ for residuals associated with the original KKT system (10.1) and $\mathbf{r}$ for residuals associated with the homotopy system (10.2). These have three main components, corresponding to the three sets of equations in (10.1) and (10.2), and the first two components are identified by the subscripts x and v, respectively (with an attached overbar when defined at points off the path).

10.2.2 Euler Predictor

Differentiating the homotopy system (10.2), namely $\mathbf{H}(\mathbf{w}(\mu), \mu) = \mathbf{0}$, with respect to μ yields the homotopy differential equation (HDE) analogous to (3.11). Linearizing the HDE at the initial point $\mathbf{w}^{(0)}$ yields the equations that define the tangent vector to the lifted path along the direction of increasing μ. The Euler predictor direction is the negative of this vector. Its components $\mathbf{p}_{w^{(0)}} = (\mathbf{p}_{x^{(0)}}, \mathbf{p}_{\pi^{(0)}}, \mathbf{p}_{v^{(0)}})$ are along the (negative) tangent vector to the lifted path at the point $\mathbf{w}^{(0)} = \left(\mathbf{x}^{(0)}, \pi^{(0)}, \mathbf{v}^{(0)}\right)$ and are defined by the following set of linear equations:

$$\begin{aligned} \mathbf{A}\mathbf{p}_{x^{(0)}} &= -\mathbf{q}_{x^{(0)}}, \\ \mathbf{A}^T\mathbf{p}_{\pi^{(0)}} + \mathbf{p}_{v^{(0)}} &= -\mathbf{q}_{v^{(0)}}, \\ \mathbf{X}_0\mathbf{p}_{v^{(0)}} + \mathbf{V}_0\mathbf{p}_{x^{(0)}} &= -\mathbf{X}_0\mathbf{V}_0\mathbf{e}, \end{aligned} \tag{10.9}$$

where in the right-hand side of the last equation, the components of the vectors are multiplied pairwise to form a vector. The vectors and diagonal

matrices above were defined in Section 10.1.1. Note that the matrix in the equations defining the Euler predictor is the Jacobian matrix (10.4) at the point $\mathbf{w} = \mathbf{w}^{(0)}$.

Consider any other point *on the path* leading from $\left(\mathbf{w}^{(0)}, 1\right)$ to $(\mathbf{w}^*, 0)$, say the point $(\mathbf{w}(\mu), \mu)$, $0 < \mu < 1$. Recalling the discussion on restart strategies of Section 3.2.4, suppose one *restarted* the homotopy at this point. By uniqueness of the original path, this must correspond simply to a reparameterization: rescaling the initial value of the parameter to 1. The restarted lifted path must overlap the original path to $\mathbf{w}^*$. Thus the (negative) tangent vector $(\mathbf{p}_x, \mathbf{p}_\pi, \mathbf{p}_v)$ at any point $(\mathbf{x}(\mu), \pi(\mu), \mathbf{v}(\mu))$ of the lifted path corresponding to the original system (10.2) is given by (10.9) with a simple transcription of symbols. For later reference, let us restate these equations:

$$\begin{aligned} \mathbf{A}\mathbf{p}_x &= -\mathbf{q}_x, \\ \mathbf{A}^T\mathbf{p}_\pi + \mathbf{p}_v &= -\mathbf{q}_v, \\ \mathbf{X}\mathbf{p}_v + \mathbf{V}\mathbf{p}_x &= -\mathbf{X}\mathbf{V}\mathbf{e}, \end{aligned} \tag{10.10}$$

where the diagonal matrices $\mathbf{X}$ and $\mathbf{V}$ have diagonal elements given by the components of $\mathbf{x}(\mu)$ and $\mathbf{v}(\mu)$, and the right-hand-side vectors are defined by $\mathbf{q}_x = \mathbf{A}\mathbf{x}(\mu) - \mathbf{b}$ and $\mathbf{q}_v = \mathbf{A}^T\pi(\mu) + \mathbf{v}(\mu) - \mathbf{c}$.

The following properties of the Euler predictor are easy to verify, and details are left to the reader:

- The Euler predictor direction at the point $\mathbf{w}(\mu)$ is parallel to the search direction defined by Newton's method applied to the KKT system of optimality equations (10.1) at the point $\mathbf{w}(\mu)$. In particular, this equivalence of directions holds at the initiating point $\mathbf{w}^{(0)}$. Note that the only restriction placed on $\mathbf{w}^{(0)}$ is that its components satisfy $\mathbf{x}^{(0)} > \mathbf{0}$ and $\mathbf{v}^{(0)} > \mathbf{0}$.
- Suppose that the path under consideration is the *central* path. Then the Euler predictor at any point on this path lies along the (negative) tangent vector to the path, and its primal and dual components are parallel to the primal affine-scaling and the dual affine-scaling directions, respectively, as defined in Sections 8.1 and 8.2. These directions coincide with the corresponding primal and dual components of the primal–dual affine-scaling directions given in Section 8.3.
- In the case of a *feasible* interior path, the Euler predictor direction $\mathbf{p}_w$ at a point $\mathbf{w}(\mu)$ on the path is the primal–dual affine-scaling direction as developed in expressions (8.18) and (8.19) with the appropriate transcription of symbols; i.e.,

$$\mathbf{X} \leftrightarrow \mathbf{D}_X; \qquad \mathbf{V} \leftrightarrow \mathbf{D}_V. \tag{10.11}$$

But note that the primal and dual components of $\mathbf{p}_w$ are now *distinct* from the primal affine-scaling and dual affine-scaling directions defined at the feasible interior primal and dual components of $\mathbf{w}(\mu)$.

- In the case of a *general* interior path; i.e., a path that need not lie within the feasible polytope of (P) and (D), the Euler predictor direction at $\mathbf{w}(\mu)$ is defined in terms of a primal–dual affine-scaling diagonal *matrix*, but it does *not* correspond to the direction that is defined in expressions (8.18) and (8.19), again with the transcription of symbols (10.11). It is standard terminology to continue to use the name primal–dual affine-scaling direction for the Euler predictor corresponding to the general (infeasible) interior case.

10.2.3 Second-Order Corrector

The vector $\mathbf{s}_w = (\mathbf{s}_x, \mathbf{s}_\pi, \mathbf{s}_v)$ defining the second derivative to the lifted path at the point $(\mathbf{x}(\mu), \pi(\mu), \mathbf{v}(\mu))$ can be obtained along lines analogous to the derivation of the first derivative above. It is given by solving the following set of linear equations:

$$\begin{aligned} \mathbf{A}\mathbf{s}_x &= \mathbf{0}, \\ \mathbf{A}^T\mathbf{s}_\pi + \mathbf{s}_v &= \mathbf{0}, \\ \mathbf{X}\mathbf{s}_v + \mathbf{V}\mathbf{s}_x &= -2\mathbf{p}_x\mathbf{p}_v, \end{aligned} \tag{10.12}$$

where in the right-hand side of the last equation, the components of the Euler predictor vector are multiplied pairwise to form a new vector. The diagonal matrices $\mathbf{X}$ and $\mathbf{V}$ are defined as in the Euler predictor equations (10.10).

10.2.4 Newton Corrector

Consider a point, denoted by $\overline{\mathbf{w}} = (\overline{\mathbf{x}}, \overline{\pi}, \overline{\mathbf{v}})$, that lies *off* the (general) homotopy path and an associated value μ. The Newton corrector $(\mathbf{n}_{\overline{x}}, \mathbf{n}_{\overline{\pi}}, \mathbf{n}_{\overline{v}})$ at this point is obtained, analogously to its definition in expression (3.2), by linearizing the system (10.2). It is given by the following system of linear equations:

$$\begin{aligned} \mathbf{A}\mathbf{n}_{\overline{x}} &= -\mathbf{r}_{\overline{x}}, \\ \mathbf{A}^T\mathbf{n}_{\overline{\pi}} + \mathbf{n}_{\overline{v}} &= -\mathbf{r}_{\overline{v}}, \\ \overline{\mathbf{V}}\mathbf{n}_{\overline{x}} + \overline{\mathbf{X}}\mathbf{n}_{\overline{v}} &= -\overline{\mathbf{X}\mathbf{V}}\mathbf{e} + \mu\mathbf{X}_0\mathbf{V}_0\mathbf{e}, \end{aligned} \tag{10.13}$$

where $\mathbf{e} \in R^n$ denotes the vector of all ones, and $\overline{\mathbf{X}}$ and $\overline{\mathbf{V}}$ denote diagonal matrices with positive diagonal elements defined by the components of $\overline{\mathbf{x}}$ and $\overline{\mathbf{v}}$, respectively. The right-hand sides of the first two equations are residual vectors, namely,

$$\mathbf{r}_{\overline{x}} = \mathbf{A}\overline{\mathbf{x}} - \mathbf{b} - \mu\mathbf{q}_{x^{(0)}} = \mathbf{q}_{\overline{x}} - \mu\mathbf{q}_{x^{(0)}},$$

where $\mathbf{q}_{\overline{x}} = \mathbf{A}\overline{\mathbf{x}} - \mathbf{b}$ and

$$\mathbf{r}_{\overline{v}} = \mathbf{A}^T\overline{\pi} + \overline{\mathbf{v}} - \mathbf{c} - \mu\mathbf{q}_{v^{(0)}} = \mathbf{q}_{\overline{v}} - \mu\mathbf{q}_{v^{(0)}}$$

with $\mathbf{q}_{\overline{v}} = \mathbf{A}^T\overline{\pi} + \overline{\mathbf{v}} - \mathbf{c}$.

Denote the Newton corrector direction more compactly by the vector $\mathbf{n}_{\overline{w}} = (\mathbf{n}_{\overline{x}}, \mathbf{n}_{\overline{\pi}}, \mathbf{n}_{\overline{v}})$. If the point $\overline{\mathbf{w}}$ is infeasible, but a *feasible* homotopy path is being followed, then the residuals in the right-hand side of the Newton corrector equations (10.13) correspond to the residuals for $\overline{\mathbf{w}}$ with respect to the original KKT equations (10.1). And in this case, if a *full*, i.e., a unit, step is taken along the Newton corrector direction, then the resulting point $\overline{\mathbf{w}} + \mathbf{n}_{\overline{w}}$ becomes *feasible*.

10.2.5 Restarts

Restarting and reparameterization of a homotopy have been described in Section 3.2.4, and they can be applied to the homotopy system (10.2). In particular, restarting allows two different HDEs to be used simultanously within a bipath-following strategy. This holds the key to formulating effective algorithms in the linear programming setting, as we shall see in the next section.

Consider a restart of the homotopy defined by (3.14), applied to the homotopy system (10.2) at the point $\overline{\mathbf{w}} = (\overline{\mathbf{x}}, \overline{\pi}, \overline{\mathbf{v}})$. For convenience of discussion, let us identify the original homotopy system and associated HDE by $\mathcal{A}$ and the restarted homotopy system and its HDE by $\mathcal{B}$. The Euler predictor (EP) for homotopy $\mathcal{B}$ at the point $\overline{\mathbf{w}}$ is given by (10.9) with the components of the vector $\mathbf{w}^{(0)} = \left(\mathbf{x}^{(0)}, \pi^{(0)}, \mathbf{v}^{(0)}\right)$ replaced by the components of $\overline{\mathbf{w}} = (\overline{\mathbf{x}}, \overline{\pi}, \overline{\mathbf{v}})$, and the diagonal matrices $\mathbf{X}_0$ and $\mathbf{V}_0$ replaced by $\overline{\mathbf{X}}$ and $\overline{\mathbf{V}}$, respectively. The Newton corrector (NC) for homotopy $\mathcal{A}$ is given by (10.13).

It is useful to define the centering component (CC) $\mathbf{c}_{\overline{w}} = (\mathbf{c}_{\overline{x}}, \mathbf{c}_{\overline{\pi}}, \mathbf{c}_{\overline{v}})$ for the original homotopy $\mathcal{A}$ at the point $\overline{\mathbf{w}} = (\overline{\mathbf{x}}, \overline{\pi}, \overline{\mathbf{v}})$ as the solution of the following system of linear equations:

$$\begin{aligned} \mathbf{A}\mathbf{c}_{\overline{x}} &= \mathbf{q}_{x^{(0)}}, \\ \mathbf{A}^T\mathbf{c}_{\overline{\pi}} + \mathbf{c}_{\overline{v}} &= \mathbf{q}_{v^{(0)}}, \\ \overline{\mathbf{V}}\mathbf{c}_{\overline{x}} + \overline{\mathbf{X}}\mathbf{c}_{\overline{v}} &= \mathbf{X}_0\mathbf{V}_0\mathbf{e}. \end{aligned} \tag{10.14}$$

Note that this system differs from the system defining the Newton corrector only in its right-hand-side vector. Also, the Euler predictor for homotopy $\mathcal{B}$ at $\overline{\mathbf{w}}$ has the same left-hand-side matrix as the Newton corrector and the centering component for homotopy $\mathcal{A}$. It then follows immediately that

$$[\text{NC for } \mathcal{A} \text{ at } \overline{\mathbf{w}}] = [\text{EP for } \mathcal{B} \text{ at } \overline{\mathbf{w}}] + \mu[\text{CC for } \mathcal{A} \text{ at } \overline{\mathbf{w}}]. \tag{10.15}$$

The relationship (10.15) will turn out to be very useful in the formulation of algorithms later. An important special case of (10.15) arises when the homotopy paths are *feasible*. In this case all residuals in the equations defining the EP, NC, and CC are zero. If, additionally, the homotopy path $\mathcal{A}$ is the central path, then the diagonal matrix $\mathbf{X}_0\mathbf{V}_0$ is replaced by the identity matrix.

It is also useful to note when a bipath strategy is employed that the option exists of using a starting point for an algorithm that does *not lie on the homotopy path* $\mathcal{A}$ being followed. This corresponds to performing a restart and defining a homotopy system $\mathcal{B}$ at the very outset. For example, homotopy $\mathcal{A}$ could correspond to the central path, and an algorithm could be (re-) started at a point that lies within some prespecified neighborhood of the central path.

10.3 Numerical Illustration

A pure Euler predictor strategy treats the HDE (3.11) as an initial value problem along lines discussed in Chapter 3, where $\mathbf{H}$, in the present context, is defined by (10.2) and the Euler predictor by (10.9). At a point that lies off the path, the Euler predictor direction would be derived from the same HDE and would be tangential to a path arising from use of *a different initial condition*. Small steps would be taken within a "tubular" neighborhood of the original homotopy path in order to follow the latter accurately.

A second basic strategy is to take an Euler predictor step along the direction defined by (10.9) and then immediately to restart the homotopy at the new iterate. Now the path and associated HDE are changed at each iteration, and the "terminal value" aspect of path following, as discussed in Section 3.2.4, is addressed. We have noted that this iteration is equivalent to Newton's method applied to the original KKT conditions of the linear program, and the convergence of the pure Newton iteration is not guaranteed, as is well known.

The algorithm implemented in the numerical illustration of this section extends the foregoing terminal-value strategy by inserting a Newton corrector phase between the Euler predictor phase and the restart as follows:

Algorithm EP/NC/R: Given an initial point $\mathbf{w}^{(0)} = \left(\mathbf{x}^{(0)}, \pi^{(0)}, \mathbf{v}^{(0)}\right)$ that need not be feasible for (P) and (D) but must satisfy $\mathbf{x}^{(0)} > \mathbf{0}$ and $\mathbf{v}^{(0)} > \mathbf{0}$:

1. *Predictor Phase*: Compute the Euler predictor $\mathbf{p}_{w^{(0)}} = (\mathbf{p}_{x^{(0)}}, \mathbf{p}_{\pi^{(0)}}, \mathbf{p}_{v^{(0)}})$ from (10.9). Take a step t_P, $0 < t_P \leq 1$, along it to a new point, $\overline{\mathbf{w}} = (\overline{\mathbf{x}}, \overline{\pi}, \overline{\mathbf{v}})$; i.e., $\overline{\mathbf{w}} = \mathbf{w}^{(0)} + t_P\mathbf{p}_{w^{(0)}}$. The step length t_P is obtained by computing the largest steps t_x and t_v in the interval

$[0,1]$ such that $\overline{\mathbf{x}} \geq \mathbf{0}$ and $\overline{\mathbf{v}} \geq \mathbf{0}$ and defining $t_P = \tau * \min[t_x, t_v]$ with $\tau \in (0,1)$. Correspondingly, compute the parameter $\mu = (1 - t_P)$.

2. *Corrector Phase*: The Euler predictor direction is obtained from a linearization of the system (10.2). The first two equations of the latter system are already linear, and thus the two equations will be satisfied *exactly* by the point $\overline{\mathbf{w}}$ with the homotopy parameter at the value of μ computed above. In other words, the relations $\mathbf{A}\mathbf{x}^{(0)} = \mathbf{b} + \mathbf{q}_{x^{(0)}}$ and $\mathbf{A}\mathbf{p}_{x^{(0)}} = -\mathbf{q}_{x^{(0)}}$ imply that

$$\mathbf{A}\overline{\mathbf{x}} = \mathbf{b} + (1 - t_P)\mathbf{q}_{x^{(0)}} = \mathbf{b} + \mu\mathbf{q}_{x^{(0)}}. \tag{10.16}$$

Similarly,

$$\mathbf{A}^T\overline{\pi} + \overline{\mathbf{v}} = \mathbf{c} + (1 - t_P)\mathbf{q}_{v^{(0)}} = \mathbf{c} + \mu\mathbf{q}_{v^{(0)}}. \tag{10.17}$$

The equations (10.13) defining the Newton corrector direction $\mathbf{n}_{\overline{w}}$ must therefore have zero right-hand-side residual vectors, and a Newton corrector step along the direction computed from (10.13) will produce an iterate that continues to satisy the first two sets of equations. Take a common step t_C no greater than unity that does not violate a bound constraint along the direction, in an analogous way to the predictor phase above. Update the current iterate and denote it again by $\overline{\mathbf{w}}$. Then repeat, if necessary. Since the first two sets of linear equations in (10.2) are already satisfied, the sum of squares of the residuals for the third set of equations of the homotopy system, namely, the function $F_\mu(\overline{\mathbf{w}}) = \frac{1}{2} r_\mu(\overline{\mathbf{w}})^T r_\mu(\overline{\mathbf{w}})$ with $r_\mu(\overline{\mathbf{w}}) = (\overline{\mathbf{X}}\overline{\mathbf{V}} - \mu\mathbf{X}_0\mathbf{V}_0)\mathbf{e}$, is a natural *merit function* for the Newton iteration. One or more Newton corrector steps are taken in sequence, and the corrector phase is terminated when the merit function is sufficiently small.

3. *Restart Phase*: Perform a restart by setting $\mathbf{w}^{(0)}$ to the current iterate $\overline{\mathbf{w}}$ at the end of the corrector phase. This is the most aggressive choice of periodic restart strategy, because the path and its associated HDE are changed after every predictor–corrector cycle.

In our implementation of the foregoing Algorithm EP/NC/R, the linear program being solved is assumed to be sparse and specified in standard MPS format—see, for example, Nazareth [1987]—but only nonnegative lower bounds on variables are allowed (instead of the more general lower and upper bounds of standard MPS format). The linear program is input and represented as a column list/row index packed data structure using subroutines described in Nazareth [1986f], [1987]. All operations involving the linear programming matrix $\mathbf{A}$ in (P) or (D) in (8.1) employ this packed data structure. The linear equations defining the Euler predictor and the Newton corrector directions are solved by a simple linear CG algorithm without preconditioning—see Section 5.1—and the procedure is

TABLE 10.1. SHARE2B: 97 rows, 79 columns (excluding slacks), 730 nonzero elements, $z^* = -415.732$.

It.	$\mathbf{c}^T\mathbf{x}^{(0)}$	$\mathbf{b}^T\pi^{(0)}$	$\|\mathbf{q}_{x^{(0)}}\|/\|\mathbf{b}\|$	$\|\mathbf{q}_{v^{(0)}}\|/\|\mathbf{c}\|$	$\max\left[x_i^{(0)}v_i^{(0)}\right]$	n_C
2	−1762.69	−22735.4	6698.03	110.590	3370.96	3
3	−897.465	−21380.3	2257.86	37.2790	1104.40	3
4	−590.548	−15098.9	739.723	12.2134	416.605	3
5	−439.873	−8861.92	279.040	4.60717	138.440	3
6	−373.861	−3784.46	92.7267	1.53099	36.0400	3
7	−358.176	−1131.10	24.1394	0.398560	6.08006	4
8	−377.788	−536.579	4.07240	0.067328	1.13203	4
9	−394.126	−464.520	0.758229	0.012519	0.491282	3
10	−405.324	−436.222	0.329061	0.005433	0.212608	3
11	−412.181	−422.425	0.142403	0.002351	0.071165	3
12	−414.612	−417.649	0.047666	0.000787	0.021150	3
13	−415.324	−416.351	0.014166	0.000234	0.007168	3
14	−415.568	−415.961	0.004801	0.000079	0.002747	2
15	−415.679	−415.803	0.001840	0.000030	0.000860	2
16	−415.731	−415.743	0.000575	0.000010	0.000134	2
17	−415.733	−415.733	0.002324	0.000002	0.000009	2
18	−415.731	−415.732	0.000272	0.000000	0.000000	1
19	−415.732	−415.732	0.000015	0.000000	0.000000	1
20	−415.732	−415.732	0.000008	0.000000	0.000000	1
21	−415.732	−415.732	0.000010	0.000000	0.000000	–

terminated when the norm of the residual of the linear system, relative to the norm of its initial right-hand side, is no greater than 10^{-6}. A standard choice for the parameter τ in the predictor phase is a number in the range $[0.95, 0.99]$, and in these experiments, $\tau = 0.95$ is used. The Newton iteration in the corrector phase is terminated when the norm of the merit function, relative to $\max_{i=1,\ldots,n} \overline{x}_i\overline{v}_i$, falls below 10^{-4}. For further details on the implementation, in particular, the choice of starting point and the step-length computation, see Nazareth [1996b].

Results for a realistic problem from the Netlib collection (Gay [1985]) are shown in Table 10.1. Problem statistics for this problem, called SHARE2B, are stated in the caption of the table. Its optimal value is z^*. At the end of each predictor, corrector(s), and restart cycle of the above Algorithm EP/NC/R, the table shows the attained objective values[2] for the primal

[2]The small loss of monotonicity in row 17 of the fourth column of the table is the result of insufficient accuracy in the computation of the predictor direction. Also, in this example, the maximum and minimum values of the quantities $x_i^{(0)}v_i^{(0)}$ are almost identical throughout, and the central-ray-initiated path is followed closely. In other examples reported in Nazareth [1996b], these quantities differ significantly.

and dual linear programs, the relative infeasibility for primal and dual, the largest element in the "duality gap," and the number of Newton corrector steps taken (n_C). These results are representative of performance on several other test problems from the Netlib collection, for which similar tables are given in Nazareth [1996b]. In addition, many refinements of the implementation are possible, for example, use of different steps in primal and dual space, extrapolation, inexact computation of search vectors, and so on. Details can be found in Nazareth [1996b].

It is evident that convergence can be ensured in a *conceptual* version of Algorithm EP/NC/R, where $\tau > 0$ can be chosen to be as small a fraction as is desired for theoretical purposes in the predictor phase, and the homotopy system (10.2) is solved *exactly* in the corrector phase. (The restart then corresponds to a reparameterization of the original homotopy.) In an *implementable* version of the algorithm, where larger values of τ are used and relatively high accuracy is requested in the residual of the corrector phase, the homotopy path will *not* change substantially between consecutive cycles. Although not formalized, a proof of convergence can be derived. This is the version of Algorithm EP/NC/R that is implemented in the foregoing numerical illustration.

From a practical standpoint, it is common to use values of τ very close to unity, for example, 0.99, and to enforce the condition $n_C = 1$. In this case the path can change drastically after a restart. Convergence requires the introduction of a bipath-following strategy so that the ability to correct to a guiding homotopy path from the original starting point is retained. We consider strategies of this type in the next subsection.

10.4 Path-Following Algorithms

10.4.1 Notation

We now use the more general notation of Chapter 3 for defining iterates of an algorithm; i.e., iteration numbers will be specified explicitly.

Points generated by a conceptual algorithm that lie *on* the homotopy path being followed will be denoted by $\mathbf{w}^{(k)} = \big(\mathbf{x}^{(k)}, \pi^{(k)}, \mathbf{v}^{(k)}\big)$. A homotopy parameter value will often be specified explicitly, for example, $\mathbf{x}\big(\mu^{(k)}\big)$, highlighting the fact that the point is a *function* of the homotopy parameter μ.

Points that lie *off* the path will be denoted by $\overline{\mathbf{w}}^{(k)} = \big(\overline{\mathbf{x}}^{(k)}, \overline{\pi}^{(k)}, \overline{\mathbf{v}}^{(k)}\big)$. If a target homotopy parameter value, say $\overline{\mu}^{(k)}$, is explicitly *associated* with the point, then, as in Chapter 3, it could be denoted by $\overline{\mathbf{x}}\big[\overline{\mu}^{(k)}\big]$. Note that the point is *not* a function of the parameter. However, this extension of the notation will rarely be needed here.

Within the sets of linear equations defining the Euler predictor, Newton corrector, and other quantities in Section 10.2, iteration numbers will

also be attached; for example, $\mathbf{X}$ will be replaced by $\mathbf{X}_k$, and $\mathbf{V}$ by $\mathbf{V}_k$. *Henceforth, in this chapter, the symbols* $\mathbf{X} = \text{diag}[x_1, \ldots, x_n]$ *and* $\mathbf{V} = \text{diag}[v_1, \ldots, v_n]$ *will be used to denote diagonal matrices at arbitrary points* $\mathbf{x} = (x_1, \ldots, x_n)$ *and* $\mathbf{v} = (v_1, \ldots, v_n)$*, respectively.*

10.4.2 Strategies

Figure 10.2 gives a sample of the many different path-following strategies that can be composed from the building blocks described in Section 10.2. *Lifted* paths are depicted. The strategies all employ the same basic template, and the continuous lines correspond to the paths used in a particular strategy. (The dotted lines in the depiction of that strategy can be ignored.) The initiating point for a path is denoted by $\left(\mathbf{x}^{(0)}, \pi^{(0)}, \mathbf{v}^{(0)}\right)$.

Paths can be feasible or infeasible, and a variety of such paths for a simple linear programming example, similar to the one of Figure 7.1, were illustrated earlier in Figure 10.1. Now Figure 10.2 gives a more generic depiction for six different path-following strategies that are discussed in the remainder of this section, and again in Section 10.4.4.

Strategy 1 and *Strategy 2* are embedding, unipath-following strategies that use a direct vertical-predictor decrease in the parameter, followed by one or more Newton corrector steps initiated from the current iterate $\overline{\mathbf{w}}^{(k)}$. *Strategy 1* is guided by the central path, and *Strategy 2* is guided by the path from the initial point $\mathbf{w}^{(0)}$.

Strategy 3 corresponds to the standard form of predictor–corrector unipath-following, where the Euler predictor is the (negative) tangent to a path through the current iterate that is determined from the *original* homotopy differential equation associated with the starting point $\mathbf{w}^{(0)}$. The latter point also determines the path that guides one or more Newton corrector steps, initiated at the point obtained from the predictor step. In *Strategy 4*, the path emanating from the current iterate defines both the Euler predictor and guides the Newton corrector steps. This is the unipath strategy used in the numerical illustration of Section 10.3.

Strategy 5 and *Strategy 6* are bipath-following strategies. In *Strategy 5*, the Euler predictor is the (negative) tangent to a path emanating from the current iterate $\overline{\mathbf{w}}^{(k)}$, and the central path guides the Newton corrector steps. In *Strategy 6*, the Euler predictor is the same as the one used in *Strategy 5*, but the Newton corrector steps are guided by the homotopy path from the initial point.

Numerous variants arise depending on whether or not the initial point is feasible[3] for the primal and dual problems, whether the initial point is additionally chosen to define an optimally conditioned or a central-ray-initiated path, and whether the Newton corrector is restricted to a *single*

[3]The primal variables and dual slacks must, of course, have positive components.

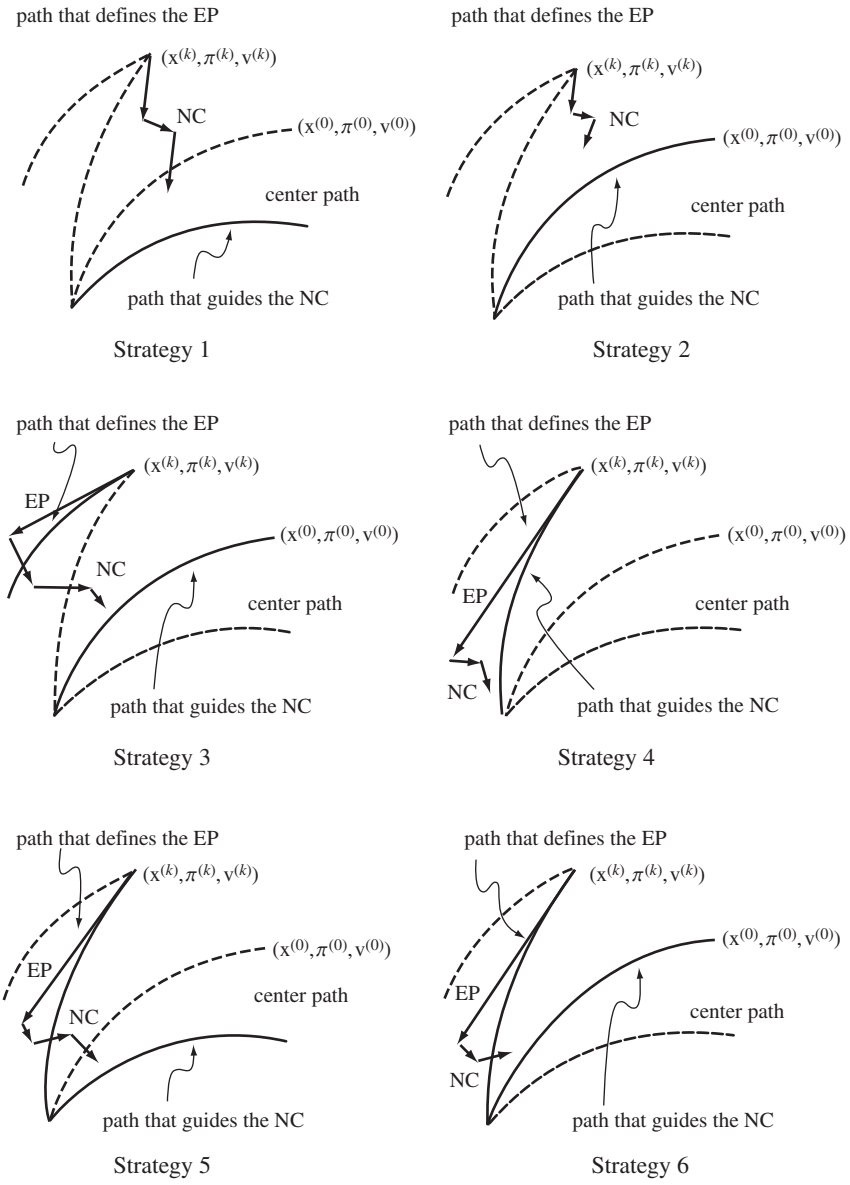


FIGURE 10.2 Strategies.

step. In all cases, an important issue in defining an algorithm is the choice of step lengths along predictor and corrector directions.

The foregoing discussion by no means exhausts the rich variety of path-following strategies that have been proposed and analyzed in the literature, but *it illustrates how they can be easily apprehended* within the

path-following framework considered in this section. It is interesting to note that the most basic form of homotopy path-following, corresponding to *Strategy 3* of Figure 10.2 and patterned on standard homotopy path-following theory, is only one of many used nowadays in the linear programming setting.

The homotopy or path-following approach has been promulgated from the outset in the sequence of articles of Nazareth [1986d], [1991], [1995a], [1996b] and Sonnevend et al. [1986], [1990], [1991]. The connection between Karmarkar's interior-point approach and homotopy (path-following) techniques was first noted in Nazareth [1986d]. This connection and an important relationship between interior-point techniques and log-barrier transformations of a linear program (discussed in the next chapter) served to motivate the *landmark contribution* of Megiddo [1989] on path-following in a primal–dual setting. This article, in turn, set the stage for a variety of primal–dual path-following algorithms by many different researchers.

Although not pursued in this monograph, strategies for *noninterior* path-following as described in Section 10.1.2 can be formulated along lines similar to the strategies described above, and they deserve further investigation.

10.4.3 Measures and Targets

The formulation of an algorithm based on the foregoing strategies requires measures of the "worth" of an iterate relative to a path being followed. This is the topic of the present subsection.

Merit Functions

In the Newton corrector phase, the Newton iteration defined by (3.2) is applied to the homotopy system (10.2) for given $\mu = \mu^{(k)} > 0$, starting from the point $\overline{\mathbf{w}}^{(k)}$. The Newton iteration has an associated merit function like (3.3) for which its search direction is a direction of descent. Thus the Newton corrector direction (10.13) is a direction of descent for the sum of squares of residuals merit function, say $F_\mu(\mathbf{w}) = \frac{1}{2}\mathbf{H}(\mathbf{w}, \mu^{(k)})^T\mathbf{H}(\mathbf{w}, \mu^{(k)})$, where $\mathbf{H}$ is defined prior to expression (10.4). The Jacobian matrix of the system is nonsingular at any interior point, and the gradient vector mapping ∇F_μ can vanish only at a minimizing point of F_μ. Thus F_μ is a natural merit function for globalizing Newton's method in a corrector phase, because difficulties illustrated by the examples of Chapter 3 cannot arise.

Let us focus on the feasible case so that the first two sets of equations of (10.2) are satisfied exactly. Then the merit function involves only residuals corresponding to the third set of equations, and thus

$$F_\mu(\mathbf{w}) = \tfrac{1}{2} r_\mu(\mathbf{w})^T r_\mu(\mathbf{w}), \tag{10.18}$$

where

$$r_\mu(\mathbf{w}) = (\mathbf{XV} - \mu^{(k)}\mathbf{X}_0\mathbf{V}_0)\mathbf{e} \tag{10.19}$$

and $\mathbf{X} = \text{diag}[x_1, \ldots, x_n]$, $\mathbf{V} = \text{diag}[v_1, \ldots, v_n]$, and $\mathbf{X}_0$ and $\mathbf{V}_0$ are defined immediately after (10.2). The Newton corrector (10.13) will have zero residuals in the right-hand side corresponding to the first two equations, and it is a direction of descent for the merit function (10.18).

Targets on the Homotopy Path

We have also seen in expression (3.12) and its application in the present setting, namely, (10.5) and discussion immediately afterwards, that the size of the duality gap for any point *on* the path relative to the size of the starting duality gap gives the associated value of μ. This provides a natural way to *associate a target point* on a feasible homotopy path with any *feasible* iterate $\overline{\mathbf{w}}^{(k)}$ that lies off the path, namely,

$$\overline{\mu}^{(k)} = \frac{\big(\overline{\mathbf{x}}^{(k)}\big)^T\overline{\mathbf{v}}^{(k)}}{\big(\mathbf{x}^{(0)}\big)^T\mathbf{v}^{(0)}}. \tag{10.20}$$

Also,

$$\big(\overline{\mathbf{x}}^{(k)}\big)^T\overline{\mathbf{v}}^{(k)} = \mathbf{c}^T\overline{\mathbf{x}}^{(k)} - \mathbf{b}^T\overline{\pi}^{(k)}.$$

Consider a horizontal Newton corrector iteration that is initiated at the point $\big(\overline{\mathbf{w}}^{(k)}, \overline{\mu}^{(k)}\big)$ in the space of dimension $(m+2n+1)$ and restricted to a horizontal hyperplane defined by $\eta^T\mathbf{w} = \overline{\mu}^{(k)}$, where $\eta \in R^{m+2n+1}$ is a unit vector with 1 in its last position and zeros elsewhere. This corresponds to the Newton iteration applied to the system (10.2) in the space of dimension $(m + 2n)$, with $\mu = \overline{\mu}^{(k)}$ and iterates confined to lie in a hyperplane, say $\mathcal{H}\big(\overline{\mathbf{w}}^{(k)}\big)$, that passes through $\overline{\mathbf{w}}^{(k)}$ and has a normal vector $(\mathbf{c}, -\mathbf{b}, \mathbf{0})$.

A more ambitious target would be given by replacing $\overline{\mu}^{(k)}$ by $\sigma\overline{\mu}^{(k)}$, where $\sigma \in (0, 1]$. This corresponds to *shifting* the above hyperplanes in their respective spaces.

Neighborhoods

Let us confine attention to points $\mathbf{w}$ in the foregoing hyperplane $\mathcal{H}\big(\overline{\mathbf{w}}^{(k)}\big)$, and consider a neighborhood of the feasible homotopy path defined by

$$N_2(\beta; \overline{\mu}^{(k)}) = \big\{\mathbf{w} \in \mathcal{F}^0 : \|\big(\mathbf{XV} - \overline{\mu}^{(k)}\mathbf{X}_0\mathbf{V}_0\big)\mathbf{e}\|_2 \le \beta\,\overline{\mu}^{(k)} \quad \text{and} \quad \mathbf{w} \in \mathcal{H}\big(\overline{\mathbf{w}}^{(k)}\big)\big\}, \tag{10.21}$$

where $\beta \in (0, 1)$, $\overline{\mu}^{(k)}$ is defined by (10.20), $\mathbf{X} = \text{diag}[x_1, \ldots, x_n]$, $\mathbf{V} = \text{diag}[v_1, \ldots, v_n]$, $\mathbf{X}_0$ and $\mathbf{V}_0$ are defined after (10.2), and $\mathcal{F}^0$ denotes the interior of the primal–dual feasible polytope. This neighborhood is closely related to the merit function (10.18). It measures distance as the norm of the residuals relative to the target duality gap $\overline{\mu}^{(k)}$, where the latter quantity would be the same if computed at any other point in the hyperplane.

If the point $\overline{\mathbf{w}}^{(k)}$ and its associated hyperplane $\mathcal{H}\big(\overline{\mathbf{w}}^{(k)}\big)$ are varied, one obtains a *conical* neighborhood of the (lifted) feasible homotopy path as follows:

$$N_2(\beta) = \left\{ (\mathbf{x}, \pi, \mathbf{v}) \in \mathcal{F}^0 : \|\mathbf{XVe} - \mu \mathbf{X}_0 \mathbf{V}_0 \mathbf{e}\|_2 \leq \beta\mu, \right.$$
$$\left. \text{where } \mu = \frac{\mathbf{x}^T \mathbf{v}}{\big(\mathbf{x}^{(0)}\big)^T \mathbf{v}^{(0)}} \right\}. \tag{10.22}$$

For the central path, this simplifies to

$$N_2(\beta) = \left\{ (\mathbf{x}, , \pi, \mathbf{v}) \in \mathcal{F}^0 : \|\mathbf{XVe} - \mu \mathbf{e}\|_2 \leq \beta\mu \text{ where } \mu = \frac{\mathbf{x}^T \mathbf{v}}{n} \right\}. \tag{10.23}$$

Various other neighborhoods have been employed in the literature, for example, the neighborhood obtained by using the ∞-norm in place of the Euclidean norm in (10.22). Neighborhoods of lifted homotopy paths, in particular, neighborhoods of the central path, play an important role in the complexity analysis of associated interior-point algorithms.

Extensions

The foregoing considerations can be appropriately enlarged to the case of *infeasible* iterates. For example, consider the Newton corrector phase in the illustrative Algorithm EP/NC/R of Section 10.3. For reasons discussed immediately after expressions (10.16)–(10.17), the same merit function as in the feasible case, namely (10.18), can be used. Also, after iteration superscripts are explicitly attached, it follows directly from these two expressions that

$$\big(\overline{\mathbf{x}}^{(k)}\big)^T \overline{\mathbf{v}}^{(k)} = \big(\mathbf{c} + \mu^{(k)} \mathbf{q}_{v^{(0)}}\big)^T \overline{\mathbf{x}}^{(k)} - \big(\mathbf{b} + \mu^{(k)} \mathbf{q}_{x^{(0)}}\big)^T \overline{\pi}^{(k)}. \tag{10.24}$$

If the associated target value $\overline{\mu}^{(k)} = \big(\overline{\mathbf{x}}^{(k)}\big)^T \overline{\mathbf{v}}^{(k)} / \big(\mathbf{x}^{(0)}\big)^T \mathbf{v}^{(0)}$ were used in a horizontal Newton corrector iteration started from $\overline{\mathbf{w}}^{(k)}$, then the iterates would lie in a hyperplane in primal–dual space that passes through $\overline{\mathbf{w}}^{(k)}$ and has the following normal vector:

$$\Big(\big(\mathbf{c} + \mu^{(k)} \mathbf{q}_{v^{(0)}}\big), -\big(\mathbf{b} + \mu^{(k)} \mathbf{q}_{x^{(0)}}\big), \mathbf{0} \Big).$$

Again a neighborhood analogous to (10.21) can be defined.

When the starting iterate for the Newton corrector is infeasible for the first two equations of the homotopy, then merit functions and neighborhoods must be suitably generalized. (This would occur in the foregoing example when a vertical predictor is substituted for the Euler predictor.)

These generalizations will not be pursued here, but further detail can be found in some of the relevant references cited in the next subsection.

10.4.4 Complexity Results

Theoretical algorithms based on the strategies summarized in Figure 10.2 and the measures and targets discussed in Section 10.4.3 have been extensively analyzed in the literature. We now give a short review[4] of some known complexity results.

Central path-following by Newton's method is based on *Strategy 1*. An embedding algorithm of this type, which is initiated from a suitably chosen feasible point, maintains feasible iterates, and uses a *single* Newton corrector step at each iteration, is shown to have polynomial complexity in Kojima, Mizuno, and Yoshise [1989]. Central path-following by Newton's method from an *infeasible* initial point—*Strategy 1* with $\mathbf{w}^{(0)} = \big(\mathbf{x}^{(0)}, \pi^{(0)}, \mathbf{v}^{(0)}\big)$ and subsequent iterates permitted to be infeasible—is proposed in Lustig [1991], thereby extending the foregoing Kojima–Mizuno–Yoshise [1989] algorithm. A general *globally convergent* algorithm of this type is given in Kojima, Megiddo, and Mizuno [1993], and a particular version is shown to have polynomial complexity in Mizuno [1994]. A related result can be found in Zhang [1994].

An embedding algorithm based on *Strategy 2*, which is initiated from a feasible, interior point $\mathbf{w}^{(0)}$, maintains feasible iterates, and uses a single Newton corrector step at each iteration, is shown to have polynomial complexity in Nazareth [1991] using complexity analysis proof techniques developed by Renegar and Shub [1992]. The homotopy parameter is reduced at each vertical-predictor step according to the following rule:

$$\mu^{(k+1)} = \left(1 - \frac{1}{40\sqrt{n}\sqrt{\|\mathbf{W}_0\|_2\|(\mathbf{W}_0)^{-1}\|_2}}\right)\mu^{(k)},$$

where $\mathbf{W}_0 \equiv \mathbf{X}_0\mathbf{V}_0$ and $\|\cdot\|_2$ denotes the largest element on the diagonal of $\mathbf{W}_0$; i.e., its spectral norm. Note that the reduction in the parameter depends on the *condition number* (10.3) associated with the homotopy system (10.2). All subsequent iterates generated by the algorithm remain feasible, and the duality gap at each iteration is bounded as follows:

$$\big(\overline{\mathbf{x}}^{(k)}\big)^T\overline{\mathbf{v}}^{(k)} = \mathbf{c}^T\overline{\mathbf{x}}\big(\mu^{(k)}\big) - \mathbf{b}^T\overline{\pi}\big(\mu^{(k)}\big) \le 2\mu^{(k)}n\|\mathbf{W}_0\|_2.$$

The associated target value on the path is (10.20) and is thus bounded by

$$\overline{\mu}^{(k)} \le \frac{2\mu^{(k)}n\|\mathbf{W}_0\|_2}{\big(\mathbf{x}^{(0)}\big)^T\mathbf{v}^{(0)}}.$$

[4]This is by no means a comprehesive survey of the available literature, and many important contributions are not mentioned.

The duality gap and the target value can be reduced to any fixed prespecified accuracy in $O(\sqrt{n})$ iterations. Related results are given in Tseng [1992]. An algorithm based on *Strategy 2* that follows a path lying *outside* the feasible polytope is given by Kojima, Megiddo, and Noma [1991].

Strategy 3 is the traditional EN path-following strategy. Central path-following algorithms along these lines are developed and analyzed in Sonnevend, Stoer, and Zhao [1990], [1991] for the case of feasible starting points and iterates, with one or more corrector steps being used to return to a neighborhood of the central path. These authors also mention the more general "weighted path" case, again for a feasible starting point, but do not pursue its details.

Strategy 4 is implemented in the numerical illustration of Section 10.3 and corresponds to general homotopy path-following with restarts. No convergence or complexity results are available for this algorithm, but the computational results given in Section 10.3 and, in more detail, in Nazareth [1996b] provide some useful information about its performance on realistic test problems.

A polynomial-time Euler predictor/Newton corrector algorithm based on *Strategy 5* is given by Mizuno, Todd, and Ye [1993]. This algorithm is initiated and remains within the feasible region, and it uses a *single* corrector step at each iteration. An adaptive choice is made for the predictor step length and the homotopy parameter so as to remain within the central-path neighborhood defined by (10.23). As noted in Section 10.2.2, the Euler predictor for the feasible case defines the primal–dual affine-scaling direction, and its primal and dual components are directions of descent/ascent for the primal/dual objective vectors, respectively. Thus the duality gap decreases monotonically with a step along the Euler predictor. The target values at consecutive iterations decrease monotonically as follows:

$$\overline{\mu}^{(k+1)} \leq \left(1 - 8^{-0.25} n^{-0.5}\right) \overline{\mu}^{(k)}.$$

Finally, consider *Strategy 6* of Figure 10.2. Kojima, Megiddo, and Mizuno [1993] distinguish their *Strategy 1* path-following approach from the *Strategy 2* path-following approach of Kojima, Megiddo, and Noma [1991], both of which are mentioned in the discussion above, as follows:[5]

> The trajectory traced by their algorithm runs *outside* the feasible region, but the central trajectory traced by our algorithm runs through the interior of the feasible region.

In the same vein, *Strategy 6* should be distinguished from the central-path-following *Strategy 5*. Tseng [1995] gives a very accessible proof of polynomial complexity for an Euler predictor/Newton corrector approach based on *Strategy 6*, which is started on (or suitably near) the central ray and follows

[5]Italics ours.

a central-ray-initiated interior path as defined in Section 10.1.1. After an Euler predictor step, only a *single* Newton corrector step is performed at each iteration. The homotopy parameter and the infeasibility at successive iterations are explicitly reduced as follows:

$$\mu^{(k+1)} = \big(1 - \theta^{(k)}\big)\big(1 - \gamma^{(k)}\big)\mu^{(k)}$$

and

$$\begin{bmatrix} \mathbf{q}_{x^{(k+1)}} \\ \mathbf{q}_{v^{(k+1)}} \end{bmatrix} = \mu^{(k+1)} \begin{bmatrix} \mathbf{q}_{x^{(0)}} \\ \mathbf{q}_{v^{(0)}} \end{bmatrix},$$

where $\theta^k \in [0, 1]$ is a step length along the predictor direction that keeps the iterate within a neighborhood of the central-ray-initiated path ($\theta^k = 0$ is permitted), and $\gamma^k \in (0, 1)$ is a quantity associated with the corrector, for which a lower bound that depends on both n and the starting point ensures polynomial complexity. The other quantities define the current and initial infeasibilities analogously to their definition in earlier sections.

In summary and referring to Figure 10.1 for purposes of illustration, we see that complexity results are known for the central path C, the feasible homotopy path F, and the central-ray-initiated and optimally conditioned paths Ia and Ib. When the initiating point $\big(\mathbf{x}^{(0)}, \pi^{(0)}, \mathbf{v}^{(0)}\big)$, where $\mathbf{x}^{(0)} > \mathbf{0}$ and $\mathbf{v}^{(0)} > \mathbf{0}$, defines a general interior homotopy path as in path G in Figure 10.1, it is possible to extend the proof of Tseng [1995] to obtain a polynomial complexity bound that depends also on the condition number of $\mathbf{W}_0$. Results for noninterior paths—Section 10.1.2 and path N of Figure 10.1—can also be derived from the results mentioned in the present section through the use of the transformation (10.8). In this case the complexity result will depend on the condition number of the system, the "weights" ω_i and γ_i, and the distance of the starting point from feasibility. A detailed complexity analysis of algorithms based on paths illustrated by G and N of Figure 10.1 is an open topic that deserves further investigation.

10.5 Mehrotra's Implementation

Mehrotra [1991], [1992] achieved an *implementational breakthrough* via a clever synthesis and extension of bipath-following techniques discussed in Sections 10.2 and 10.4. In particular, his approach incorporated the second-order corrector of Section 10.2.3 into the search direction, utilized efficient heuristics for defining targets on the central path, used different steps in primal and dual space, and reorganized the computational linear algebra, namely, the solution of the large-scale systems of linear equations that define the search direction at each iteration, in order to enhance overall efficiency.

We introduce these ideas in stages, beginning with use of the second-order corrector of Section 10.2.3. Let us resume the discussion in the introductory paragraph of Section 10.3. A refinement of the basic Euler predictor with restarts strategy mentioned there is to supplement the Euler predictor (10.9), which linearizes and thus approximates the path to first order, by employing the second-order corrector (10.12) to now define a *quadratic approximation* to the path being followed. The algorithm below, which is due to Zhang and Zhang [1995], forms this quadratic approximation to *a homotopy path that emenates from the current iterate* $\left(\overline{\mathbf{x}}^{(k)}, \overline{\pi}^{(k)}, \overline{\mathbf{v}}^{(k)}\right)$, and whose associated homotopy system is defined in an analogous manner to (10.2). In particular, the linear systems defining the algorithm's Euler predictor (10.25) and second-order corrector (10.26) are obtained from the systems of linear equations (10.9) and (10.12), respectively, with $\left(\mathbf{x}^{(0)}, \pi^{(0)}, \mathbf{v}^{(0)}\right)$ replaced by $\left(\overline{\mathbf{x}}^{(k)}, \overline{\pi}^{(k)}, \overline{\mathbf{v}}^{(k)}\right)$:

Algorithm EP/SOC/R: Given a starting point $(\overline{\mathbf{x}}^{(0)} > \mathbf{0}, \overline{\pi}^{(0)}, \overline{\mathbf{v}}^{(0)} > \mathbf{0})$, for $k = 0, 1, \ldots,$ do the following:

1. Solve the following set of linear equations for the Euler predictor direction $\mathbf{p}_{\overline{w}^{(k)}} = (\mathbf{p}_{\overline{x}^{(k)}}, \mathbf{p}_{\overline{\pi}^{(k)}}, \mathbf{p}_{\overline{v}^{(k)}})$:

$$\begin{aligned} \mathbf{A}\mathbf{p}_{\overline{x}^{(k)}} &= -\mathbf{q}_{\overline{x}^{(k)}}, \\ \mathbf{A}^T\mathbf{p}_{\overline{\pi}^{(k)}} + \mathbf{p}_{\overline{v}^{(k)}} &= -\mathbf{q}_{\overline{v}^{(k)}}, \\ \overline{\mathbf{X}}_k\mathbf{p}_{\overline{v}^{(k)}} + \overline{\mathbf{V}}_k\mathbf{p}_{\overline{x}^{(k)}} &= -\overline{\mathbf{X}}_k\overline{\mathbf{V}}_k\mathbf{e}, \end{aligned} \tag{10.25}$$

   where $\overline{\mathbf{X}}_k$ and $\overline{\mathbf{V}}_k$ are diagonal matrices defined by the components of $\overline{\mathbf{x}}^{(k)}$ and $\overline{\mathbf{v}}^{(k)}$, respectively, $\mathbf{q}_{\overline{x}^{(k)}} = \mathbf{A}\overline{\mathbf{x}}^{(k)} - \mathbf{b}$, and $\mathbf{q}_{\overline{v}^{(k)}} = \mathbf{A}^T\overline{\pi}^{(k)} + \overline{\mathbf{v}}^{(k)} - \mathbf{c}$.

2. Solve the following system of linear equations for the second-order corrector direction $\mathbf{s}_{\overline{w}^{(k)}} = (\mathbf{s}_{\overline{x}^{(k)}}, \mathbf{s}_{\overline{\pi}^{(k)}}, \mathbf{s}_{\overline{v}^{(k)}})$:

$$\begin{aligned} \mathbf{A}\mathbf{s}_{\overline{x}^{(k)}} &= \mathbf{0}, \\ \mathbf{A}^T\mathbf{s}_{\overline{\pi}^{(k)}} + \mathbf{s}_{\overline{v}^{(k)}} &= \mathbf{0}, \\ \overline{\mathbf{X}}_k\mathbf{s}_{\overline{v}^{(k)}} + \overline{\mathbf{V}}_k\mathbf{s}_{\overline{x}^{(k)}} &= -2\mathbf{p}_{\overline{x}^{(k)}}\mathbf{p}_{\overline{v}^{(k)}}, \end{aligned} \tag{10.26}$$

   where in the last set of equations of (10.26), the components of $\mathbf{p}_{\overline{x}^{(k)}}$ and $\mathbf{p}_{\overline{v}^{(k)}}$, which are obtained from (10.25), are multiplied pairwise.

3. Choose a step length $t^{(k)} > 0$ such that

$$\overline{\mathbf{w}}^{(k+1)} = \overline{\mathbf{w}}^{(k)} + t^{(k)}\mathbf{p}_{\overline{w}^{(k)}} + \tfrac{1}{2}\left(t^{(k)}\right)^2\mathbf{s}_{\overline{w}^{(k)}}$$

   remains positive in its $\overline{\mathbf{x}}^{(k+1)}$ and $\overline{\mathbf{v}}^{(k+1)}$ components.

   In a simpler version of Step 3, a direction-finding procedure (DfP) can be substituted for the quadratic approximation, by making a particular choice for $t^{(k)}$. Typically, choose $t^{(k)} = 1$, yielding a search

direction as the sum of the Euler predictor and second-order corrector directions. Then take a step along this search direction, subject to remaining in the positive orthant with respect to the $\mathbf{x}$ and $\mathbf{v}$ variables.

Note that the homotopy path used to compute the predictor and second-order corrector is restarted, in effect, at each iteration. Like Algorithm EP/NC/R of Section 10.3, Algorithm EP/SOC/R is not known to be convergent. We now enhance the latter by utilizing a bipath-following strategy. In addition to the homotopy path used in the algorithm that emenates from the current iterate, an original guiding path is retained. *In the remainder of the discussion in this section, let us take the latter to be the central path.*

Referring back to the building blocks of Section 10.2 from which algorithms can be composed, let us consider the Newton corrector obtained from (10.13) at the point $\overline{\mathbf{w}}^{(k)}$ with associated parameter value $\mu^{(k)}$. Because the central path is being followed, $\mathbf{q}_{x^{(0)}} = \mathbf{0}$ and $\mathbf{q}_{v^{(0)}} = \mathbf{0}$, and thus the residuals, simplify in these equations. In the new notation, where iteration superscripts are attached, the equations defining the Newton corrector are as follows:

$$\begin{aligned} \mathbf{A}\mathbf{n}_{\overline{x}^{(k)}} &= -\mathbf{q}_{\overline{x}^{(k)}}, \\ \mathbf{A}^T\mathbf{n}_{\overline{\pi}^{(k)}} + \mathbf{n}_{\overline{v}^{(k)}} &= -\mathbf{q}_{\overline{v}^{(k)}}, \\ \overline{\mathbf{V}}_k\mathbf{n}_{\overline{x}^{(k)}} + \overline{\mathbf{X}}_k\mathbf{n}_{\overline{v}^{(k)}} &= -\overline{\mathbf{X}}_k\overline{\mathbf{V}}_k\mathbf{e} + \mu^{(k)}\mathbf{e}. \end{aligned} \tag{10.27}$$

From the relationship (10.15), the Newton corrector at $\overline{\mathbf{w}}^{(k)}$ is the sum of the Euler predictor computed for the homotopy restarted at $\overline{\mathbf{w}}^{(k)}$, namely, (10.25) and the centering component derived from (10.14) and weighted by $\mu^{(k)}$. This centering component, again in the present notation (with iteration numbers attached) and simplified for the central path, is as follows:

$$\begin{aligned} \mathbf{A}\mathbf{c}_{\overline{x}^{(k)}} &= \mathbf{0}, \\ \mathbf{A}^T\mathbf{c}_{\overline{\pi}^{(k)}} + \mathbf{c}_{\overline{v}^{(k)}} &= \mathbf{0}, \\ \overline{\mathbf{V}}_k\mathbf{c}_{\overline{x}^{(k)}} + \overline{\mathbf{X}}_k\mathbf{c}_{\overline{v}^{(k)}} &= \mathbf{e}. \end{aligned} \tag{10.28}$$

Motivated by Algorithm EP/SOC/R and the direction-finding procedure (DfP) that can be used within Step 3, the Newton corrector direction (10.27), in particular, the part that corresponds to the Euler predictor, is augmented by adding to it the second-order corrector obtained from (10.26). This yields the DfP employed within Mehrotra-type path-following approaches. Its search direction, stated mathematically, is the sum of the Euler predictor direction for the interior path restarted at $\overline{\mathbf{w}}^{(k)}$, the corresponding second-order corrector direction for this path, and the centering component direction for the central guiding path, where the last direction is weighted by $\mu^{(k)}$. Let us first consider the choice of $\mu^{(k)}$, and then the effective computation of the search direction.

Mehrotra [1992] proposed an adaptive heuristic technique for choosing the weight $\mu^{(k)}$ as follows:

$$\mu^{(k)} = \sigma^{(k)}\overline{\mu}^{(k)}, \tag{10.29}$$

where $\overline{\mu}^{(k)} = \left(\overline{\mathbf{x}}^{(k)}\right)^T \overline{\mathbf{v}}^{(k)}/n$ is the target value (10.20) on the central path, for which the denominator $\left(\mathbf{x}^{(0)}\right)^T \mathbf{v}^{(0)}$ equals n. The parameter $\sigma^{(k)} \in (0, 1]$ is chosen adaptively at each iteration, using a heuristic based on the amount of progress that could be made along the pure Euler predictor (10.25). Recall from the itemized discussion at the end of Section 10.2.2 that this direction is equivalent to the primal–dual affine-scaling direction defined at the current iterate $\overline{\mathbf{w}}^{(k)}$. The heuristic proposed by Mehrotra [1992] is as follows. Compute steps along the components of the primal–dual affine-scaling direction in the usual way, for example, analogously to the computation of these quantities described in Chapters 7 and 8 for the feasible case. (This allows for different step lengths in the primal and dual spaces and thus permits greater potential progress.) Suppose the primal and dual points obtained using these step lengths are the components of the vector $\hat{\mathbf{w}}^{(k)} = \left(\hat{\mathbf{x}}^{(k)}, \hat{\pi}^{(k)}, \hat{\mathbf{v}}^{(k)}\right)$. A new target value $\hat{\mu}^{(k)} = \left(\hat{\mathbf{x}}^{(k)}\right)^T \hat{\mathbf{v}}^{(k)}/n$ is derived from them. Then let $\sigma^{(k)} = \left(\hat{\mu}^{(k)}/\overline{\mu}^{(k)}\right)^r$, where, typically, $r = 2$ or 3. This heuristic choice has been shown to work well in practice. The Newton corrector using (10.29) targets a point on a shifted hyperplane as described in Section 10.4.3. Also, it is useful to define

$$\sigma^{(k)}\overline{\mu}^{(k)} = \left(\overline{\mathbf{x}}^{(k)}\right)^T \overline{\mathbf{v}}^{(k)}/\rho, \tag{10.30}$$

and $\sigma^k \in (0, 1]$ implies $\rho \geq n$. The above equation (10.30) will he helpful later in seeing the connection with Karmarkar-like potential-function developments of Chapter 12.

Finally, the computational linear algebra is organized efficiently in the Mehrotra implementation by exploiting the fact that the Euler predictor, the second-order corrector, and the centering component are obtained from systems of linear equations that differ only in their right-hand sides. Thus the matrix factorization used to compute the Euler predictor can be preserved and used to solve a second system for $\overline{\mathbf{c}}_{\overline{w}^{(k)}} = (\overline{\mathbf{c}}_{\overline{x}^{(k)}}, \overline{\mathbf{c}}_{\overline{\pi}^{(k)}}, \overline{\mathbf{c}}_{\overline{v}^{(k)}})$ as follows:

$$\begin{aligned} \mathbf{A}\overline{\mathbf{c}}_{\overline{x}^{(k)}} &= \mathbf{0}, \\ \mathbf{A}^T\overline{\mathbf{c}}_{\overline{\pi}^{(k)}} + \overline{\mathbf{c}}_{\overline{v}^{(k)}} &= \mathbf{0}, \\ \overline{\mathbf{X}}_k\overline{\mathbf{c}}_{\overline{v}^{(k)}} + \overline{\mathbf{V}}_k\overline{\mathbf{c}}_{\overline{x}^{(k)}} &= -2\mathbf{p}_{\overline{x}^{(k)}}\mathbf{p}_{\overline{v}^{(k)}} + \sigma^{(k)}\overline{\mu}^{(k)}\mathbf{e}; \end{aligned} \tag{10.31}$$

i.e., instead of computing two separate directions—the second-order corrector and the weighted centering component—solve a single linear system, with a suitably chosen right-hand side as given in (10.31). Only one additional back-substitution operation is required to obtain the solution of this system, because the matrix factorization is already available. The

solution $\overline{\mathbf{c}}_{\overline{w}^{(k)}}$ is the required sum of the second-order corrector $\mathbf{s}_{\overline{w}^{(k)}}$ given by (10.26) and the weighted centering component $\left(\sigma^{(k)}\overline{\mu}^{(k)}\right)\mathbf{c}_{\overline{w}^{(k)}}$ derived from (10.28), which we call the *centered second-order corrector*.

The search direction employed in the Mehrotra-type DfP is the sum of the Euler predictor $\mathbf{p}_{\overline{w}^{(k)}}$ given by (10.25) and the centered second-order corrector $\overline{\mathbf{c}}_{\overline{w}^{(k)}}$ given by (10.31). A step is taken along this direction, subject to remaining in the positive orthant with respect to the $\mathbf{x}$ and $\mathbf{v}$ variables in the usual way. Again, greater progress can be obtained by allowing different step lengths along directions in primal and dual space. Note also that the techniques for *inexact* computation of direction vectors, along lines discussed in Chapter 7, can be adapted to the present setting. Details of this refinement are left to the reader.

Mehrotra [1992] used a potential function and a "fall-back" search direction in order to guarantee convergence of his algorithm. A theoretical version of the Mehrotra predictor–corrector approach was shown to be of polynomial-time complexity by Zhang and Zhang [1995]. In their approach, the algorithm was restored to the infeasible-interior-point, central-path-following tradition of Kojima, Megiddo, and Mizuno [1993], Mizuno [1994], and Zhang [1994] mentioned in Section 10.4.4. But instead of using the more standard Euler predictor and Newton corrector strategy, an Euler predictor and centered second-order corrector strategy was substituted.

The techniques described in this section (and variants on them) have been incorporated into a number of practical programs. See, in particular, Lustig, Marsden, and Shanno [1992], [1994a], Vanderbei [1995], Gondzio [1995], and Zhang [1995].

10.6 Summary

In the early days of the "Karmarkar Revolution," the term "interior" signified the interior of the feasible polytope. Other methods were termed "exterior." The terminology has since evolved, and "interior," nowadays, means the interior of the positive orthant with respect to the $\mathbf{x}$ and/or $\mathbf{v}$ variables; i.e., the terminology has evolved from *polytope-interior* to *orthant-interior*. Methods that do not maintain feasibility with respect to the equality constraints of the primal and dual linear programs are called *infeasible-interior*. If the $\mathbf{x} \geq \mathbf{0}$ and/or $\mathbf{v} \geq \mathbf{0}$ bounds are violated, they are called *noninterior*.[6]

From a conceptual standpoint, we have seen in this chapter that the homotopy approach lies at the foundation of interior-point methods for linear programming. It has found *fundamental new expression* within this specialized context, and these adaptations and developments, in turn, have

[6]And thus, if equality constraints were also violated, they would be called *infeasible-noninterior* methods—(un)naturally!

considerably reilluminated usage within the traditional general nonlinear equations setting.

Modern interior-point techniques are based on predictor–corrector, bipath-following strategies, where the guiding path is usually the central path. Many interesting refinements remain to be explored. For example:

- Alternative parameterizations, in particular, the use of two parameters simultaneously, can provide a conceptual foundation for taking different steps in the primal and dual spaces. For some early proposals along these lines, see Nazareth [1996b].
- Alternative guiding paths can be used in place of the central path in many algorithms, for example, in Mehrotra's algorithm of Section 10.5.
- Noninterior extensions described in Section 10.1.2 are an interesting avenue to pursue.
- Unification and integration of self-dual simplex homotopy techniques, which follow piecewise-linear paths, and primal–dual interior-point homotopy techniques, which follow smooth paths, could lead to useful enhancements of current, large-scale mathematical programming systems. For some further discussion, see Nazareth [1996b].

10.7 Notes

Sections 10.1–10.5: The material in this chapter is based on Nazareth [1996b], [1998].

http://www.springer.com/978-0-387-95572-8

Differentiable Optimization and Equation Solving
A Treatise on Algorithmic Science and the Karmarkar Revolution
Nazareth, J.
2003, XVII, 256 p. 2 illus., Hardcover
ISBN: 978-0-387-95572-8