Chapter 5

Path-Following Algorithms

In Chapters 1 and 2, we described the central path $\mathcal{C}$, a path of points $(x_\tau, \lambda_\tau, s_\tau)$ that leads to the set Ω of primal-dual solutions. Points in Ω satisfy the KKT conditions (2.4), whereas points in $\mathcal{C}$ are defined by conditions that differ from the KKT conditions only by the presence of a positive scalar parameter $\tau > 0$, namely,

$$\begin{aligned} A^T\lambda + s &= c, && (5.1a)\\ Ax &= b, && (5.1b)\\ (x, s) &\geq 0, && (5.1c)\\ x_i s_i &= \tau, \qquad i = 1, 2, \ldots, n. && (5.1d) \end{aligned}$$

We showed in Chapter 2 that this system has a unique solution $(x_\tau, \lambda_\tau, s_\tau)$ for each $\tau > 0$ whenever the problem is feasible (although the KKT conditions, for which $\tau = 0$ in (5.1d), may have multiple solutions).

Path-following methods follow $\mathcal{C}$ in the direction of decreasing τ to the solution set Ω. They do not necessarily stay exactly *on* $\mathcal{C}$ or even particularly close to it. Rather, they stay within a loose but well-defined neighborhood of $\mathcal{C}$ while steadily reducing the duality measure μ to zero. Each search direction is a Newton step toward a point on $\mathcal{C}$, a point for which the duality measure τ is equal to or smaller than the current duality measure μ. The target value $\tau = \sigma\mu$ is used, where $\sigma \in [0, 1]$ is the centering parameter introduced in Chapter 1.

The algorithms of this chapter generate strictly feasible iterates (x^k, λ^k, s^k) that satisfy the first three KKT conditions (5.1a), (5.1b), and (5.1c). They deviate from the central path $\mathcal{C}$ only because the pairwise products $x_i s_i$ are generally not identical, so the condition (5.1d) is not satisfied exactly. This

deviation is measured by comparing the pairwise products with their average value $\mu = x^T s/n = (\sum x_i s_i)/n$, using, for example, a scaled norm defined by

$$\frac{1}{\mu}\|XSe - \mu e\| = \frac{1}{\mu}\left\| \begin{bmatrix} x_1 s_1 \\ \vdots \\ x_n s_n \end{bmatrix} - \left(\frac{x^T s}{n}\right) e \right\|. \tag{5.2}$$

In the literature, both the 2-norm and the ∞-norm have been used in this definition. For both norms, we can ensure that x and s are *strictly* positive by requiring that $(1/\mu)\|XSe - \mu e\| < 1$. (If component i of x or s is zero, we have $\|XSe - \mu e\| \geq |x_i s_i - \mu| = \mu$.)

By using the 2-norm in (5.2) and restricting the deviation to be less than a constant $\theta \in [0, 1)$, we obtain the neighborhood $\mathcal{N}_2(\theta)$ defined in (1.15):

$$\mathcal{N}_2(\theta) = \{(x, \lambda, s) \in \mathcal{F}^o \mid \|XSe - \mu e\|_2 \leq \theta\mu\}. \tag{5.3}$$

By using the ∞-norm in (5.2), we obtain the neighborhood $\mathcal{N}_\infty(\theta)$. We can motivate another neighborhood $\mathcal{N}_{-\infty}(\gamma)$ by noting that our chief concern is to keep the products $x_i s_i$ from becoming too much *smaller* than their average value μ and therefore to prevent x and s from approaching the boundary of the region $(x, s) \geq 0$ prematurely. We do not mind if some of these products are somewhat *larger* than μ, so the neighborhood $\mathcal{N}_{-\infty}(\gamma)$ uses a one-sided bound on $x_i s_i$ in place of the two-sided bound in (5.2), that is,

$$\mathcal{N}_{-\infty}(\gamma) = \{(x, \lambda, s) \in \mathcal{F}^o \mid x_i s_i \geq \gamma\mu \quad \text{for all } i = 1, 2, \ldots, n\}, \tag{5.4}$$

where $\gamma \in (0, 1)$.

Path-following methods follow Framework PD of Chapter 1. They select one of the neighborhood types $\mathcal{N}_2$, $\mathcal{N}_\infty$, or $\mathcal{N}_{-\infty}$ and choose the centering parameter σ and the step length parameter α to ensure that every iterate (x^k, λ^k, s^k) stays within the chosen neighborhood.

Methods based on the neighborhood $\mathcal{N}_2$ have $O(\sqrt{n}\log 1/\epsilon)$ complexity, matching the complexity estimate for the potential-reduction method of Chapter 4 (see Corollary 4.7). We describe two such methods in this chapter: the short-step path-following algorithm (Algorithm SPF) and the predictor-corrector algorithm (Algorithm PC). Algorithm SPF, the simplest of all interior-point methods, chooses a constant value $\sigma_k \equiv \sigma$ for the centering parameter and fixes the step length at $\alpha_k \equiv 1$ for all iterations k. This method was introduced by Kojima, Mizuno, and Yoshise [66] and Monteiro and Adler [94]; our analysis follows the latter paper. Algorithm PC alternates between two types of steps: *predictor* steps, which improve the value

of μ but which also tend to worsen the centrality measure (5.2), and *corrector* steps, which have no effect on the duality measure μ but improve centrality. Various aspects of this algorithm were foreshadowed by a number of authors, including Monteiro and Adler [94] and Sonnevend, Stoer, and Zhao [122, 123], but it was first stated and analyzed in the simple form used here by Mizuno, Todd, and Ye [92]. The algorithm is sometimes called "Mizuno–Todd–Ye predictor-corrector" to distinguish it from the quite different Mehrotra predictor-corrector algorithm of Chapter 10.

A disadvantage of the $\mathcal{N}_2(\theta)$ neighborhood is its restrictive nature. From the definition (5.3), we have for $(x, \lambda, s) \in \mathcal{N}_2(\theta)$ that

$$\sum_{i=1}^{n} \left(\frac{x_i s_i}{\mu} - 1 \right)^2 \leq \theta^2 < 1$$

so that the sum of squares of *all* relative deviations of $x_i s_i$ from their average value μ cannot exceed 1. Even if θ is close to its upper bound of 1, the neighborhood $\mathcal{N}_2(\theta)$ contains only a small fraction of the points in the strictly feasible set $\mathcal{F}^o$, so algorithms based on this neighborhood do not have much room in which to maneuver and the amount of progress they can achieve at each iteration is limited. The neighborhood $\mathcal{N}_{-\infty}(\gamma)$, on the other hand, is much more expansive: When γ is small, it takes up almost the entire strictly feasible set $\mathcal{F}^o$. We discuss a long-step path-following algorithm based on this neighborhood—Algorithm LPF—that makes more aggressive (that is, smaller) choices of centering parameter σ than does Algorithm SPF. Instead of taking unit steps, however, Algorithm LPF performs a line search along the direction obtained from (1.13), choosing α_k to be as large as possible subject to the restriction of remaining within $\mathcal{N}_{-\infty}(\gamma)$. Algorithm LPF is closely related to the very first polynomial primal-dual algorithm proposed by Kojima, Mizuno, and Yoshise [67]. It is more closely related to practical algorithms than are Algorithms SPF and PC, but its complexity bound is worse: $O(n \log 1/\epsilon)$ vs. $O(\sqrt{n} \log 1/\epsilon)$.

Although the kth step taken by a path-following algorithm aims for the point on the central path $\mathcal{C}$ whose duality measure is $\sigma_k \mu_k$, it rarely hits this target. The reason is that there is a discrepancy between the nonlinear equations (5.1) and the linear approximation on which the Newton-like step equations (1.13) are based. This discrepancy is quantified by the pairwise products $\Delta x_i \Delta s_i$. Much of the analysis in this chapter is concerned with finding bounds on these products and showing that the step $(\Delta x^k, \Delta \lambda^k, \Delta s^k)$ makes significant progress toward its target without actually scoring a direct hit.

In this chapter, we focus mainly on the convergence of the sequence $\{\mu_k\}$ of duality measures to zero. We close the chapter by looking at a different, but related, issue: convergence of the primal-dual iteration sequences $\{(x^k, \lambda^k, s^k)\}$. We prove that the x^k and s^k components are bounded and that strictly complementary solutions to (2.1), (2.2) can be recovered from the limit points of these sequences.

All three algorithms described in this chapter are remarkable for their simplicity. Despite their strong theoretical properties, they are easy to state and to analyze, once we are past the hurdle of a few tricky technical results. They also provide the foundation for more powerful algorithms, including algorithms that allow infeasible starting points and rapid local convergence. Unfortunately, as we see in subsequent chapters, our claim of simple analysis does not necessarily hold when these capabilities are added.

The Short-Step Path-Following Algorithm

We start with the short-step path-following algorithm, Algorithm SPF. This method starts at a point $(x^0, \lambda^0, s^0) \in \mathcal{N}_2(\theta)$ and uses uniform values $\alpha_k = 1$ and $\sigma_k = \sigma$, where θ and σ satisfy a certain relationship, described below. All iterates (x^k, λ^k, s^k) stay inside $\mathcal{N}_2(\theta)$, and the duality measure μ_k converges linearly to zero at the constant rate $1 - \sigma$.

The method is defined by filling in the Framework PD from Chapter 1. We assign specific values to θ and σ, justifying them in the analysis that follows.

Algorithm SPF
Given $\theta = 0.4$, $\sigma = 1 - 0.4/\sqrt{n}$, and $(x^0, \lambda^0, s^0) \in \mathcal{N}_2(\theta)$;
for $k = 0, 1, 2, \ldots$
    set $\sigma_k = \sigma$ and solve (1.13) to obtain $(\Delta x^k, \Delta \lambda^k, \Delta s^k)$;
    set $(x^{k+1}, \lambda^{k+1}, s^{k+1}) = (x^k, \lambda^k, s^k) + (\Delta x^k, \Delta \lambda^k, \Delta s^k)$;
end (for).

This algorithm is illustrated in Figure 5.1, which plots the first few iterates of the algorithm projected into an unusual space. The horizontal and vertical axes represent the pairwise products $x_1 s_1$ and $x_2 s_2$ for this two-dimensional problem, so the central path is the line emanating from $(0, 0)$ at an angle of $\pi/4$ radians. In this (nonlinear) space, the search directions transform to curves rather than straight lines. The solution is at the origin, and the challenge facing the algorithm is to reach this point while maintaining the feasibility conditions $Ax = b$, $A^T\lambda + s = c$ at all iterates.

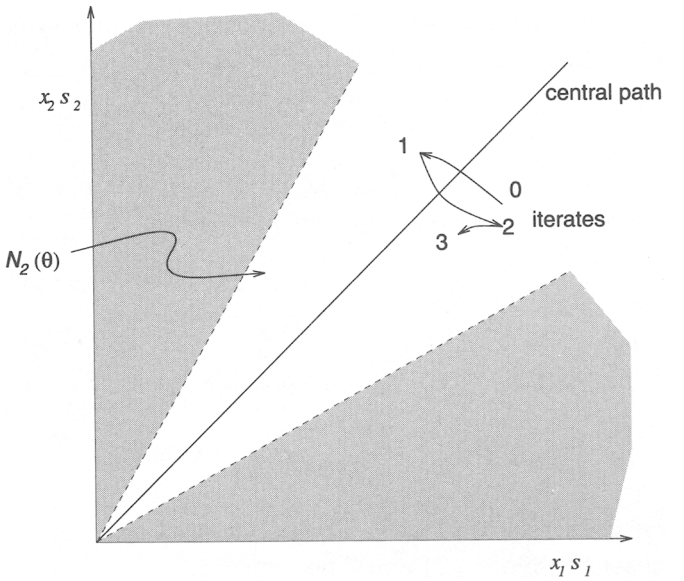

Figure 5.1. Iterates of Algorithm SPF, plotted in (xs) space.

The representations of the central path $\mathcal{C}$ and the neighborhoods $\mathcal{N}_2(\theta)$ and $\mathcal{N}_{-\infty}(\gamma)$ in Figures 5.1, 5.2, and 5.3 make it easy to see how step lengths are affected by neighborhood boundaries.

Most of the effort in the analysis of Algorithm SPF goes into showing that all its iterates stay in the neighborhood $\mathcal{N}_2(\theta)$. We prove this claim below in Lemmas 5.4 and 5.5 and Theorem 5.6. For now, let us take this claim on trust and prove the global convergence and polynomial complexity results. Linear convergence follows from Lemma 5.1, which we state in general terms because it applies to *all* algorithms that obtain their search directions from the system (1.12). The following notation is useful:

$$(x(\alpha), \lambda(\alpha), s(\alpha)) = (x, \lambda, s) + \alpha(\Delta x, \Delta\lambda, \Delta s), \tag{5.5a}$$

$$\mu(\alpha) = x(\alpha)^T s(\alpha)/n. \tag{5.5b}$$

Lemma 5.1 *Let the step* $(\Delta x, \Delta\lambda, \Delta s)$ *be defined by* (1.12). *Then*

$$\Delta x^T \Delta s = 0 \tag{5.6}$$

and

$$\mu(\alpha) = (1 - \alpha(1 - \sigma))\mu. \tag{5.7}$$

Proof. The first result is left as an exercise. For the second result, we use the third row of (1.12), namely,

$$S\Delta x + X\Delta s = -XSe + \sigma\mu e. \tag{5.8}$$

By summing the n components of this equation, we obtain $s^T\Delta x + x^T\Delta s = -(1-\sigma)x^Ts$. From this formula and (5.6), we obtain

$$x(\alpha)^Ts(\alpha) = x^Ts + \alpha(s^T\Delta x + x^T\Delta s) + \alpha^2\Delta x^T\Delta s = x^Ts(1-\alpha(1-\sigma)),$$

giving the result. □

For Algorithm SPF, we have by our specific choices of σ_k and α_k that

$$\mu_{k+1} = \sigma\mu_k = \left(1 - \frac{0.4}{\sqrt{n}}\right)\mu_k, \qquad k = 0, 1, \ldots, \tag{5.9}$$

so global linear convergence follows. The polynomial complexity result is an immediate consequence of (5.9) and Theorem 3.2.

Theorem 5.2 *Given $\epsilon > 0$, suppose that the starting point $(x^0, \lambda^0, s^0) \in \mathcal{N}_2(0.4)$ in Algorithm SPF has*

$$\mu_0 \leq 1/\epsilon^\kappa$$

for some positive constant κ. Then there is an index K with $K = O(\sqrt{n}\log 1/\epsilon)$ such that

$$\mu_k \leq \epsilon \qquad \text{for all } k \geq K.$$

Proof. Because of (5.9), we can set $\delta = 0.4$ and $\omega = 0.5$ in Theorem 3.2, and the result follows immediately. □

In the next section, we return to the task of showing that the iterates (x^k, λ^k, s^k) stay inside $\mathcal{N}_2(\theta)$.

Technical Results

The first result is purely technical; its proof can be found at the end of the chapter.

Lemma 5.3 *Let u and v be any two vectors in $\mathbb{R}^n$ with $u^Tv \geq 0$. Then*

$$\|UVe\| \leq 2^{-3/2}\|u+v\|^2,$$

where

$$U = \text{diag}(u_1, u_2, \ldots, u_n), \qquad V = \text{diag}(v_1, v_2, \ldots, v_n).$$

We showed in Lemma 5.1 that the inner product $\Delta x^T \Delta s = \sum \Delta x_i \Delta s_i$ is zero, but there is no reason for the individual pairwise products to be zero. In the next result, Lemma 5.4, we find a bound on the vector of these pairwise products.

Lemma 5.4 *If* $(x, \lambda, s) \in \mathcal{N}_2(\theta)$, *then*

$$\|\Delta X \Delta S e\| \leq \frac{\theta^2 + n(1-\sigma)^2}{2^{3/2}(1-\theta)}\mu.$$

Proof. Recall the definition of the diagonal matrix D as $X^{1/2}S^{-1/2}$ from (1.24). If we multiply (5.8) by $(XS)^{-1/2}$, we obtain

$$D^{-1}\Delta x + D\Delta s = (XS)^{-1/2}(-XSe + \sigma\mu e). \tag{5.10}$$

Now, apply Lemma 5.3 with $u = D^{-1}\Delta x$ and $v = D\Delta s$ to obtain

$$\begin{aligned}
\|\Delta X \Delta S e\| &= \|(D^{-1}\Delta X)(D\Delta S)e\| \\
&\leq 2^{-3/2}\|D^{-1}\Delta x + D\Delta s\|^2 && \text{from Lemma 5.3} \\
&= 2^{-3/2}\|(XS)^{-1/2}(-XSe + \sigma\mu e)\|^2 && \text{from (5.10)} \\
&= 2^{-3/2}\sum_{i=1}^{n}\frac{(-x_i s_i + \sigma\mu)^2}{x_i s_i} \\
&\leq 2^{-3/2}\frac{\|XSe - \sigma\mu e\|^2}{\min_i x_i s_i}. && (5.11)
\end{aligned}$$

Since $(x, \lambda, s) \in \mathcal{N}_2(\theta)$, we have

$$\min_i x_i s_i \geq (1-\theta)\mu. \tag{5.12}$$

For a bound on the numerator in (5.11), note first that

$$e^T(XSe - \mu e) = x^T s - \mu e^T e = 0,$$

and therefore

$$\begin{aligned}
&\|XSe - \sigma\mu e\|^2 \\
&= \|(XSe - \mu e) + (1-\sigma)\mu e\|^2 \\
&= \|XSe - \mu e\|^2 + 2(1-\sigma)\mu e^T(XSe - \mu e) + (1-\sigma)^2\mu^2 e^T e \\
&\leq \theta^2\mu^2 + (1-\sigma)^2\mu^2 n. && (5.13)
\end{aligned}$$

We obtain the result by substituting (5.12) and (5.13) into (5.11). □

Lemma 5.10 is a result similar to Lemma 5.4 that is proven during our analysis of Algorithm LPF. Lemma 5.4 gives a tighter bound, however, because the current point (x, λ, s) lies in the more restrictive neighborhood $\mathcal{N}_2(\theta)$. The difference in the bounds obtained in Lemmas 5.4 and 5.10 accounts for the difference in polynomial complexity estimate for Algorithms SPF and LPF.

We know from Lemma 5.1 that μ decreases linearly as we move along the direction $(\Delta x, \Delta\lambda, \Delta s)$. But how far does the point $(x(\alpha), \lambda(\alpha), s(\alpha))$ stray from the central path in the 2-norm measure? The answer is provided by the following result, which is a simple consequence of Lemma 5.4.

Lemma 5.5 *If* $(x, \lambda, s) \in \mathcal{N}_2(\theta)$, *we have*

$$\begin{aligned} &\|X(\alpha)S(\alpha)e - \mu(\alpha)e\| \\ &\quad \leq |1-\alpha|\,\|XSe - \mu e\| + \alpha^2\|\Delta X \Delta S e\| \qquad (5.14a) \\ &\quad \leq |1-\alpha|\theta\mu + \alpha^2\left[\frac{\theta^2 + n(1-\sigma)^2}{2^{3/2}(1-\theta)}\right]\mu. \qquad (5.14b) \end{aligned}$$

Proof. We use the third row of (1.12) to resolve the components of $X(\alpha)S(\alpha)e - \mu(\alpha)e$. From this equation and Lemma 5.1, we have

$$\begin{aligned} &x_i(\alpha)s_i(\alpha) - \mu(\alpha) \\ &\quad = x_i s_i + \alpha(s_i\Delta x_i + x_i\Delta s_i) + \alpha^2\Delta x_i\Delta s_i - (1-\alpha(1-\sigma))\mu \\ &\quad = x_i s_i(1-\alpha) + \alpha\sigma\mu + \alpha^2\Delta x_i\Delta s_i - (1-\alpha+\alpha\sigma)\mu \\ &\quad = x_i s_i(1-\alpha) + \alpha^2\Delta x_i\Delta s_i - (1-\alpha)\mu. \end{aligned}$$

Reassembling these components into a vector, we obtain

$$\begin{aligned} &\|X(\alpha)S(\alpha)e - \mu(\alpha)e\| \\ &\quad = \left\|\left[x_i s_i(1-\alpha) - (1-\alpha)\mu + \alpha^2\Delta x_i\Delta s_i\right]_{i=1}^{n}\right\| \\ &\quad \leq |1-\alpha|\,\|XSe - \mu e\| + \alpha^2\|\Delta X\Delta S e\| \\ &\quad \leq |1-\alpha|\,\theta\mu + \alpha^2\left[\frac{\theta^2 + n(1-\sigma)^2}{2^{3/2}(1-\theta)}\right]\mu. \quad \square \end{aligned}$$

Theorem 5.6 defines a relationship between θ and σ and shows that even a full step ($\alpha = 1$) along the search direction will not take the new iterate outside the neighborhood $\mathcal{N}_2(\theta)$.

Theorem 5.6 *Let the parameters $\theta \in (0,1)$ and $\sigma \in (0,1)$ be chosen to satisfy*

$$\frac{\theta^2 + n(1-\sigma)^2}{2^{3/2}(1-\theta)} \leq \sigma\theta. \tag{5.15}$$

Then if $(x, \lambda, s) \in \mathcal{N}_2(\theta)$, we have

$$(x(\alpha), \lambda(\alpha), s(\alpha)) \in \mathcal{N}_2(\theta)$$

for all $\alpha \in [0,1]$.

Proof. Substituting (5.15) into (5.14b), we have for $\alpha \in [0,1]$ that

$$\begin{aligned} \|X(\alpha)S(\alpha)e - \mu(\alpha)e\| &\leq (1-\alpha)\theta\mu + \alpha^2\sigma\theta\mu \\ &\leq (1-\alpha+\sigma\alpha)\theta\mu \quad \text{since } \alpha \in [0,1] \\ &= \theta\mu(\alpha) \qquad\qquad \text{from (5.7).} \end{aligned} \tag{5.16}$$

Hence the point $(x(\alpha), \lambda(\alpha), s(\alpha))$ satisfies the proximity condition for $\mathcal{N}_2(\theta)$.

We still have to check that $(x(\alpha), \lambda(\alpha), s(\alpha)) \in \mathcal{F}^o$. It is easy to verify that

$$Ax(\alpha) = b, \qquad A^T\lambda(\alpha) + s(\alpha) = c.$$

To check the positivity condition $(x(\alpha), s(\alpha)) > 0$, note first that $(x, s) = (x(0), s(0)) > 0$. It follows from (5.16) that

$$x_i(\alpha)s_i(\alpha) \geq (1-\theta)\mu(\alpha) = (1-\theta)(1-\alpha(1-\sigma))\mu > 0, \tag{5.17}$$

where the strict inequality is a consequence of $\theta \in (0,1)$, $\alpha \in (0,1]$, and $\sigma \in (0,1)$. Hence, we cannot have $x_i(\alpha) = 0$ or $s_i(\alpha) = 0$ for any index i when $\alpha \in [0,1]$. Therefore $(x(\alpha), s(\alpha)) > 0$ for all $\alpha \in [0,1]$, and so $(x(\alpha), \lambda(\alpha), s(\alpha)) \in \mathcal{F}^o$, as claimed. □

At this point, the proof of validity of Algorithm SPF is nearly complete. It remains only to check that our specific choice of parameters θ and σ, namely,

$$\theta = 0.4, \qquad \sigma = 1 - 0.4/\sqrt{n},$$

satisfies the condition (5.15) for all $n \geq 1$. We leave this as an exercise.

The Predictor-Corrector Method

In Algorithm SPF, we chose $\sigma_k \equiv \sigma$ to lie strictly between 0 and 1. This choice achieves the twin goals of improving centrality and reducing the duality measure μ into a single step. The predictor-corrector method,

Algorithm PC, takes two different kinds of steps to achieve each of these two goals. Successive iterations of Algorithm PC alternate between the two types of steps, which are

- *predictor* steps ($\sigma_k = 0$) to reduce μ,
- *corrector* steps ($\sigma_k = 1$) to improve centrality.

The other important ingredient in Algorithm PC is a *pair* of $\mathcal{N}_2$ neighborhoods, nested one inside the other. Even-index iterates ((x^k, λ^k, s^k), with k even) are confined to the inner neighborhood, whereas odd-index iterates are allowed to stray into the outer neighborhood, but not beyond.

The term *predictor-corrector* arose because of the analogy with predictor-corrector algorithms in ordinary differential equations (ODEs). These algorithms follow the solution trajectory of an initial-value ODE problem by alternating between predictor steps (which move along a tangent to the trajectory) and corrector steps (which move back toward the trajectory from the predicted point).

We examine the first two iterations of Algorithm PC, which suffice to illustrate the whole algorithm. Starting from a point (x^0, λ^0, s^0) in the inner neighborhood, we calculate a predictor step by setting $\sigma_0 = 0$. We move along this direction until we reach the boundary of the outer neighborhood. We stop at this point and define the new iterate (x^1, λ^1, s^1). A corrector step is now calculated by setting $\sigma_1 = 1$. A unit step along this direction ($\alpha = 1$) leads to a new iterate (x^2, λ^2, s^2) that is back inside the inner neighborhood. The two-step cycle then repeats, generating a sequence of iterates with the even-index iterates inside the inner neighborhood and the odd-index iterates on the boundary of the outer region. See Figure 5.2 for a depiction of this process.

Predictor steps reduce the value of μ by a factor of $(1 - \alpha)$, where α is the step length. Corrector steps leave μ unchanged, but by moving back into the inner neighborhood, they give the algorithm more room to maneuver on the next (predictor) iteration.

We obtain a formal specification of Algorithm PC by again filling in Framework PD of Chapter 1. For simplicity, we define the inner neighborhood to be $\mathcal{N}_2(0.25)$ and the outer neighborhood to be $\mathcal{N}_2(0.5)$. Other choices are possible, provided that the two neighborhoods are related to each other in a way that we define below.

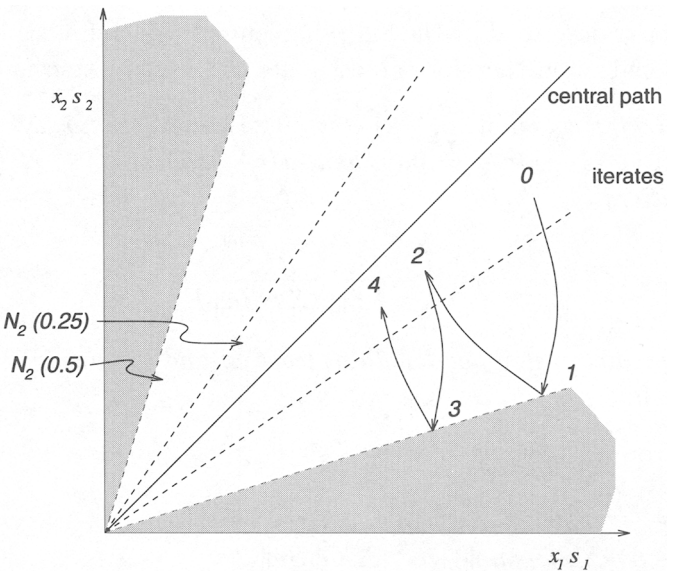

Figure 5.2. Iterates of Algorithm PC, plotted in (xs) space.

Algorithm PC
Given $(x^0, \lambda^0, s^0) \in \mathcal{N}_2(0.25)$;
for $k = 0, 1, 2, \ldots$
    if k is even
        (* predictor step *)
        solve (1.13) with $\sigma_k = 0$ to obtain $(\Delta x^k, \Delta\lambda^k, \Delta s^k)$;
        choose α_k as the largest value of α in $[0, 1]$ such that

$$(x^k(\alpha), \lambda^k(\alpha), s^k(\alpha)) \in \mathcal{N}_2(0.5); \tag{5.18}$$

        set $(x^{k+1}, \lambda^{k+1}, s^{k+1}) = (x^k(\alpha_k), \lambda^k(\alpha_k), s^k(\alpha_k))$;
    else
        (* corrector step *)
        solve (1.13) with $\sigma_k = 1$ to obtain $(\Delta x^k, \Delta\lambda^k, \Delta s^k)$;
        set $(x^{k+1}, \lambda^{k+1}, s^{k+1}) = (x^k, \lambda^k, s^k) + (\Delta x^k, \Delta\lambda^k, \Delta s^k)$;
    end (if)
end (for).

Our analysis of Algorithm PC is brief because most of the work has already been done in the analysis of Algorithm SPF. The behavior of each

predictor step is described by the following lemma, which finds a lower bound on its step length and therefore an estimate of the reduction in μ.

Lemma 5.7 *Suppose that* $(x, \lambda, s) \in \mathcal{N}_2(0.25)$, *and let* $(\Delta x, \Delta\lambda, \Delta s)$ *be calculated from* (1.12) *with* $\sigma = 0$. *Then* $(x(\alpha), \lambda(\alpha), s(\alpha)) \in \mathcal{N}_2(0.5)$ *for all* $\alpha \in [0, \bar{\alpha}]$, *where*

$$\bar{\alpha} = \min\left(\tfrac{1}{2}, \left(\frac{\mu}{8\|\Delta X \Delta S e\|}\right)^{1/2}\right). \tag{5.19}$$

Hence, the predictor step has length at least $\bar{\alpha}$, *and the new value of* μ *is at most* $(1 - \bar{\alpha})\mu$.

Proof. From (5.14a), we have

$$\begin{aligned}
&\|X(\alpha)S(\alpha)e - \mu(\alpha)e\| \\
&\quad \le (1-\alpha)\|XSe - \mu e\| + \alpha^2\|\Delta X \Delta S e\| \\
&\quad \le (1-\alpha)\|XSe - \mu e\| + \frac{\mu}{8\|\Delta X \Delta S e\|}\|\Delta X \Delta S e\| && \text{from (5.19)}, \\
&\quad \le \tfrac{1}{4}(1-\alpha)\mu + \frac{1}{8(1-\alpha)}(1-\alpha)\mu && \text{since } (x,\lambda,s) \in \mathcal{N}_2(0.25), \\
&\quad \le \tfrac{1}{4}(1-\alpha)\mu + \tfrac{1}{4}(1-\alpha)\mu && \text{since } \alpha \le \tfrac{1}{2}, \\
&\quad \le \tfrac{1}{2}\mu(\alpha) && \text{by (5.7), with } \sigma = 0.
\end{aligned}$$

Hence, the point $(x(\alpha), \lambda(\alpha), s(\alpha))$ satisfies the proximity condition for $\mathcal{N}_2(0.5)$. The remaining strict feasibility conditions can be verified as in the proof of Theorem 5.6. □

Lemma 5.4 can be used to find a lower bound on $\bar{\alpha}$. Setting $\theta = 0.25$ and $\sigma = 0$ in this result, we obtain

$$\frac{\mu}{8\|\Delta X \Delta S e\|} \ge \frac{2^{3/2}(1-0.25)}{8((0.25)^2 + n)} = \frac{3\sqrt{2}}{1+16n} \ge \frac{0.16}{n},$$

since $n \ge 1$. Hence, from (5.19) we have

$$\bar{\alpha} \ge \min\left(\tfrac{1}{2}, \left(\frac{0.16}{n}\right)^{1/2}\right) = \frac{0.4}{\sqrt{n}}.$$

Since predictor steps are taken at even-index iterates, this bound implies that

$$\mu_{k+1} \le \left(1 - \frac{0.4}{\sqrt{n}}\right)\mu_k, \qquad k = 0, 2, 4, \ldots. \tag{5.20}$$

Corrector steps are described by the following lemma, which shows that they return any point in $\mathcal{N}_2(0.5)$ to the inner neighborhood $\mathcal{N}_2(0.25)$ without changing the value of μ.

Lemma 5.8 *Suppose that* $(x, \lambda, s) \in \mathcal{N}_2(0.5)$, *and let* $(\Delta x, \Delta\lambda, \Delta s)$ *be calculated from* (1.12) *with* $\sigma = 1$. *Then we have*

$$(x(1), \lambda(1), s(1)) \in \mathcal{N}_2(0.25), \qquad \mu(1) = \mu.$$

Proof. By substituting $\sigma = 1$ into (5.7), we have immediately that $\mu(1) = \mu$. (In fact, $\mu(\alpha) = \mu$ for all $\alpha \in [0, 1]$.)

By substituting $\theta = 0.5$, $\alpha = 1$, and $\sigma = 1$ into (5.14b), we find that

$$\|X(1)S(1)e - \mu(1)e\| \leq \mu/4 = \mu(1)/4.$$

Hence, $(x(1), \lambda(1), s(1))$ satisfies the proximity conditions for $\mathcal{N}_2(0.25)$. The proof is completed by verifying that $(x(1), \lambda(1), s(1))$ is also strictly feasible, which follows as in Theorem 5.6. □

As we see from this lemma, corrector iterations leave the value of the duality measure μ unchanged. However, because the predictor iterations achieve a substantial reduction in μ (5.20), we can prove the same kind of polynomial complexity result as for the short-step algorithm.

Theorem 5.9 *Given* $\epsilon > 0$, *suppose that the starting point* $(x^0, \lambda^0, s^0) \in \mathcal{N}_2(0.25)$ *in Algorithm PC has*

$$\mu_0 \leq 1/\epsilon^\kappa$$

for some positive constant κ. *Then there is an index* K *with* $K = O(\sqrt{n} \log 1/\epsilon)$ *such that*

$$\mu_k \leq \epsilon \qquad \textit{for all } k \geq K.$$

Proof. Combining (5.20) with Lemma 5.8, we have

$$\mu_{k+2} = \mu_{k+1} \leq \left(1 - \frac{0.4}{\sqrt{n}}\right) \mu_k, \qquad k = 0, 2, 4, \ldots.$$

Hence, the reduction requirement (3.10) of Theorem 3.2 is *almost* satisfied when we set $\delta = 0.4$ and $\omega = 0.5$, except that the reduction in μ occurs over a span of *two* iterations instead of just one. The proof of Theorem 3.2 can be modified easily to handle this slightly different condition (as we

showed in the exercises for Chapter 3) without affecting the conclusion of the theorem. □

The predictor-corrector algorithm is a definite improvement over the short-step algorithm because of the adaptivity that is built into the choice of predictor step length. In Algorithm SPF, the values of σ and α are fixed at conservative values so that they confine the iterates (x^k, λ^k, s^k) to the neighborhood $\mathcal{N}_2(\theta)$ under all circumstances. By contrast, the predictor step lengths of Algorithm PC are longer when the predictor direction is a good search direction, that is, when it produces a large reduction in μ without moving away from the central path too rapidly. During the final stages of the algorithm, the predictor directions become better and better, and Algorithm PC is eventually able to use step lengths close to 1. In fact, convergence of the duality measures μ_k to zero is superlinear, as we see in Chapter 7.

Despite its adaptivity, Algorithm PC is still restricted by the cramped nature of the $\mathcal{N}_2$ neighborhoods, particularly during early iterations when far from the solution. We now describe a long-step path-following algorithm that combines flexibility in the choice of step length with the use of a more liberal neighborhood $\mathcal{N}_{-\infty}(\gamma)$.

A Long-Step Path-Following Algorithm

Algorithm LPF generates a sequence of iterates in the neighborhood $\mathcal{N}_{-\infty}(\gamma)$, which, for small values of γ (say, $\gamma = 10^{-3}$), occupies most of the set $\mathcal{F}^o$ of strictly feasible points. At each iterate of Algorithm LPF, we choose the centering parameter σ_k to lie between the two fixed limits $\sigma_{\min}$ and $\sigma_{\max}$, where $0 < \sigma_{\min} < \sigma_{\max} < 1$. The search direction is, as usual, obtained by solving (1.13), and we choose the step length α_k to be as large as possible, subject to staying inside $\mathcal{N}_{-\infty}(\gamma)$.

A formal statement of the algorithm follows.

Algorithm LPF

Given γ, $\sigma_{\min}$, $\sigma_{\max}$ with $\gamma \in (0,1)$, $0 < \sigma_{\min} < \sigma_{\max} < 1$,
and $(x^0, \lambda^0, s^0) \in \mathcal{N}_{-\infty}(\gamma)$;

for $k = 0, 1, 2, \ldots$

choose $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$;

solve (1.13) to obtain $(\Delta x^k, \Delta \lambda^k, \Delta s^k)$;

choose α_k as the largest value of α in $[0,1]$ such that

$$(x^k(\alpha), \lambda^k(\alpha), s^k(\alpha)) \in \mathcal{N}_{-\infty}(\gamma); \tag{5.21}$$

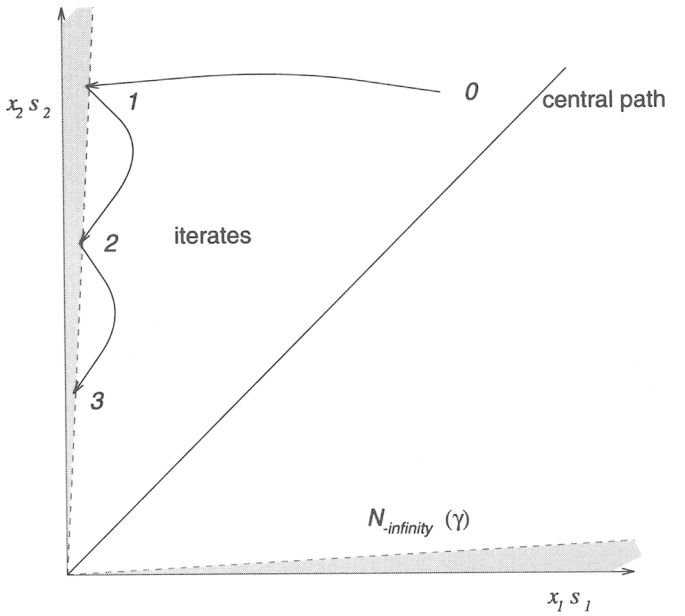

Figure 5.3. Iterates of Algorithm LPF, plotted in (xs) space.

set $(x^{k+1}, \lambda^{k+1}, s^{k+1}) = (x^k(\alpha_k), \lambda^k(\alpha_k), s^k(\alpha_k))$;
end (for).

Typical behavior of the algorithm is illustrated in Figure 5.3. As this figure shows (and the analysis confirms), the lower bound $\sigma_{\min}$ on the centering parameter ensures that each search direction starts out by moving off the boundary of $\mathcal{N}_{-\infty}(\gamma)$ and into the interior of this neighborhood. That is, small steps along the search direction improve the centrality. Larger values of α take us outside the neighborhood again, since the error of approximating the nonlinear system (5.1) by the linear step equations (1.12) becomes more pronounced as α increases. Still, we are guaranteed that a certain minimum step can be taken before we reach the boundary of $\mathcal{N}_{-\infty}(\gamma)$. Lemma 5.10 and Theorem 5.11 find a lower bound on α_k and a corresponding estimate of the reduction in μ at each iteration. Theorem 5.12 states the usual polynomial complexity result.

As an aside, we note that the representation of the neighborhoods $\mathcal{N}_2$ and $\mathcal{N}_{-\infty}$ in Figures 5.1, 5.2, and 5.3 is identical—both neighborhoods are demarcated by straight lines emanating from the origin. We see in the exercises that these two kinds of neighborhoods are closely related when

$n = 2$ but that this relationship breaks down for larger values of n. The two neighborhoods $\mathcal{N}_2$ and $\mathcal{N}_{-\infty}$ would have different shapes if we extended Figures 5.1, 5.2, or 5.3 into a third dimension.

Lemma 5.10 *If* $(x, \lambda, s) \in \mathcal{N}_{-\infty}(\gamma)$, *then*

$$\|\Delta X \Delta S e\| \leq 2^{-3/2}(1 + 1/\gamma)n\mu.$$

Proof. As in the proof of Lemma 5.4, we have

$$\|\Delta X \Delta S e\| \leq 2^{-3/2}\|(XS)^{-1/2}(-XSe + \sigma\mu e)\|^2.$$

Expanding the squared Euclidean norm and using such relationships as $x^T s = n\mu$ and $e^T e = n$, we obtain

$$\begin{aligned}
\|\Delta X \Delta S e\| &\leq 2^{-3/2}\left\|-(XS)^{1/2}e + \sigma\mu(XS)^{-1/2}e\right\|^2 \\
&\leq 2^{-3/2}\left[x^T s - 2\sigma\mu e^T e + \sigma^2\mu^2 \sum_{i=1}^{n} \frac{1}{x_i s_i}\right] \\
&\leq 2^{-3/2}\left[x^T s - 2\sigma\mu e^T e + \sigma^2\mu^2 \frac{n}{\gamma\mu}\right] \qquad \text{since } x_i s_i \geq \gamma\mu \\
&\leq 2^{-3/2}\left[1 - 2\sigma + \frac{\sigma^2}{\gamma}\right] n\mu \\
&\leq 2^{-3/2}(1 + 1/\gamma)n\mu,
\end{aligned}$$

as claimed. □

Theorem 5.11 *Given the parameters* γ, $\sigma_{\min}$, *and* $\sigma_{\max}$ *in Algorithm LPF, there is a constant* δ *independent of* n *such that*

$$\mu_{k+1} \leq \left(1 - \frac{\delta}{n}\right)\mu_k$$

for all $k \geq 0$.

Proof. We start by proving that

$$(x^k(\alpha), \lambda^k(\alpha), s^k(\alpha)) \in \mathcal{N}_{-\infty}(\gamma) \text{ for all } \alpha \in \left[0, 2^{3/2}\gamma\frac{1-\gamma}{1+\gamma}\frac{\sigma_k}{n}\right], \quad (5.22)$$

from which we deduce that α_k is bounded below as follows:

$$\alpha_k \geq 2^{3/2}\frac{\sigma_k}{n}\gamma\frac{1-\gamma}{1+\gamma}. \quad (5.23)$$

For any $i = 1, 2, \ldots, n$, we have from Lemma 5.10 that

$$|\Delta x_i^k \Delta s_i^k| \leq \|\Delta X^k \Delta S^k e\|_2 \leq 2^{-3/2}(1 + 1/\gamma) n \mu_k. \tag{5.24}$$

Using (5.8), we have from $x_i^k s_i^k \geq \gamma \mu_k$ and (5.24) that

$$\begin{aligned} x_i^k(\alpha) s_i^k(\alpha) &= (x_i^k + \alpha \Delta x_i^k)(s_i^k + \alpha \Delta s_i^k) \\ &= x_i^k s_i^k + \alpha(x_i^k \Delta s_i^k + s_i^k \Delta x_i^k) + \alpha^2 \Delta x_i^k \Delta s_i^k \\ &\geq x_i^k s_i^k (1 - \alpha) + \alpha \sigma_k \mu_k - \alpha^2 |\Delta x_i^k \Delta s_i^k| \\ &\geq \gamma(1 - \alpha)\mu_k + \alpha \sigma_k \mu_k - \alpha^2 2^{-3/2}(1 + 1/\gamma) n \mu_k. \end{aligned}$$

Meanwhile, we have from (5.7) that

$$\mu_k(\alpha) = (1 - \alpha(1 - \sigma_k))\mu_k.$$

From these last two formulas, we can see that the proximity condition

$$x_i^k(\alpha) s_i^k(\alpha) \geq \gamma \mu_k(\alpha)$$

is satisfied, provided that

$$\gamma(1 - \alpha)\mu_k + \alpha \sigma_k \mu_k - \alpha^2 2^{-3/2}(1 + 1/\gamma) n \mu_k \geq \gamma(1 - \alpha + \alpha \sigma_k)\mu_k.$$

Rearranging this expression, we obtain

$$\alpha \sigma_k \mu_k (1 - \gamma) \geq \alpha^2 2^{-3/2} n \mu_k (1 + 1/\gamma),$$

which is true if

$$\alpha \leq \frac{2^{3/2}}{n} \sigma_k \gamma \frac{1 - \gamma}{1 + \gamma}.$$

We have proved that $(x^k(\alpha), \lambda^k(\alpha), s^k(\alpha))$ satisfies the proximity condition for $\mathcal{N}_{-\infty}(\gamma)$ when α lies in the range stated in (5.22). We can also show, as in the proof of Theorem 5.6, that $(x^k(\alpha), \lambda^k(\alpha), s^k(\alpha)) \in \mathcal{F}^o$ for all α in the given range. Hence, we have proved (5.22) and therefore (5.23).

We complete the proof of the theorem by estimating the reduction in μ on the kth step. From (5.7) and (5.23), we have

$$\begin{aligned} \mu_{k+1} &= (1 - \alpha_k(1 - \sigma_k))\mu_k \\ &\leq \left(1 - \frac{2^{3/2}}{n} \gamma \frac{1 - \gamma}{1 + \gamma} \sigma_k (1 - \sigma_k)\right) \mu_k. \end{aligned} \tag{5.25}$$

Now, the function $\sigma(1-\sigma)$ is a concave quadratic function of σ, so on any given interval it attains its minimum value at one of the endpoints. Hence, we have

$$\sigma_k(1-\sigma_k) \geq \min\{\sigma_{\min}(1-\sigma_{\min}), \sigma_{\max}(1-\sigma_{\max})\} \quad \text{for all } \sigma_k \in [\sigma_{\min}, \sigma_{\max}].$$

The proof is completed by substituting this estimate into (5.25) and setting

$$\delta = 2^{3/2}\gamma\frac{1-\gamma}{1+\gamma}\min\{\sigma_{\min}(1-\sigma_{\min}), \sigma_{\max}(1-\sigma_{\max})\}. \quad \square$$

The complexity result is an immediate consequence of Theorems 5.11 and 3.2.

Theorem 5.12 *Given $\epsilon > 0$ and $\gamma \in (0,1)$, suppose that the starting point $(x^0, \lambda^0, s^0) \in \mathcal{N}_{-\infty}(\gamma)$ in Algorithm LPF has*

$$\mu_0 \leq 1/\epsilon^{\kappa}$$

for some positive constant κ. Then there is an index K with $K = O(n \log 1/\epsilon)$ such that

$$\mu_k \leq \epsilon \qquad \text{for all } k \geq K.$$

Limit Points of the Iteration Sequence

The convergence results of this chapter have focused thus far on convergence of the sequence $\{\mu_k\}$ to zero, without saying anything about the sequence of iterates $\{(x^k, \lambda^k, s^k)\}$. The behavior of the iterate sequence is a little more complicated than one might expect. The main issue is to show that the sequence $\{(x^k, s^k)\}$ has a limit point, because we can construct a primal-dual solution from any such limit point by the following argument: If $\mathcal{K}$ is the subsequence for which $\lim_{k\in\mathcal{K}}(x^k, s^k) = (x^*, s^*)$, we have for all $k \in \mathcal{K}$ that

$$Ax^k = b, \qquad c - s^k \in \text{Range}(A^T), \qquad (x^k, s^k) > 0.$$

Taking limits and using the facts that $\text{Range}(A^T)$ is closed and $\mu_k \downarrow 0$, we find that (x^*, s^*) satisfies

$$Ax^* = b, \qquad c - s^* \in \text{Range}(A^T), \qquad (x^*, s^*) \geq 0, \qquad (x^*)^T s^* = 0.$$

Hence, $c - s^* = A\lambda^*$ for some λ^*. Comparing these conditions with the KKT conditions (1.4), we conclude that $(x^*, \lambda^*, s^*) \in \Omega$ as claimed.

In this section, we look at the limiting behavior of the sequence $\{(x^k, s^k)\}$ generated by each algorithm in this chapter. We show that the sequence is bounded and therefore has at least one limit point. Further, all limit points correspond to strictly complementary solutions, that is, solutions (x^*, λ^*, s^*) for which

$$x_i^* > 0 \ \ (i \in \mathcal{B}), \qquad s_i^* > 0 \ \ (i \in \mathcal{N}), \tag{5.26}$$

where $\mathcal{B} \cup \mathcal{N}$ is the partition of $\{1, 2, \ldots, n\}$ defined in (2.11).

Lemma 5.13 *Let $\mu_0 > 0$ and $\gamma \in (0, 1)$. Then for all points (x, λ, s) with*

$$(x, \lambda, s) \in \mathcal{N}_{-\infty}(\gamma) \subset \mathcal{F}^o, \qquad \mu \le \mu_0 \tag{5.27}$$

(where $\mu = x^T s/n$), there are constants C_0 and C_3 such that

$$\|(x, s)\| \le C_0, \tag{5.28}$$

$$0 < x_i \le \mu/C_3 \ \ (i \in \mathcal{N}), \qquad 0 < s_i \le \mu/C_3 \ \ (i \in \mathcal{B}), \tag{5.29}$$

$$s_i \ge C_3\gamma \ \ (i \in \mathcal{N}), \qquad x_i \ge C_3\gamma \ \ (i \in \mathcal{B}). \tag{5.30}$$

Proof. The first result (5.28) follows immediately from Lemma 2.5 if we set $K = n\mu_0$.

For (5.29) and (5.30), let (x^*, λ^*, s^*) be any primal-dual solution. Since this solution and the point (x, λ, s) are both feasible, we have

$$Ax = Ax^* = b, \qquad A^T\lambda + s = A^T\lambda^* + s^* = c,$$

and therefore

$$(x - x^*)^T(s - s^*) = -(x - x^*)^T A^T(\lambda - \lambda^*) = 0.$$

Since (2.11) implies that $x_i^* = 0$ for $i \in \mathcal{N}$ and $s_i^* = 0$ for $i \in \mathcal{B}$, we can rearrange this expression to obtain

$$n\mu = x^T s^* + s^T x^* = \sum_{i \in \mathcal{N}} x_i s_i^* + \sum_{i \in \mathcal{B}} s_i x_i^*.$$

Since each term in the summations is nonnegative, each term is bounded by $n\mu$. Hence, for any $i \in \mathcal{N}$ with $s_i^* > 0$, we have

$$0 < x_i s_i^* \le n\mu \ \ \Rightarrow \ \ 0 < x_i \le \frac{n}{s_i^*}\mu. \tag{5.31}$$

Since the expression (5.31) holds for *any* solution (x^*, λ^*, s^*) with $s_i^* > 0$, we choose the one that yields the tightest bound on x_i, that is,

$$0 < x_i \leq \frac{n}{\sup_{(\lambda^*,s^*)\in\Omega_D} s_i^*}\,\mu.$$

Taking the maximum of this bound over the indices $i \in \mathcal{N}$, we obtain

$$\max_{i\in\mathcal{N}} x_i \leq \frac{n}{\min_{i\in\mathcal{N}} \sup_{(\lambda^*,s^*)\in\Omega_D} s_i^*}\,\mu.$$

Similarly,

$$0 < \max_{i\in\mathcal{B}} s_i \leq \frac{n}{\min_{i\in\mathcal{B}} \sup_{x^*\in\Omega_P} x_i^*}\,\mu.$$

Combining the two estimates, we obtain

$$\begin{aligned}\max\left(\max_{i\in\mathcal{N}} x_i, \max_{i\in\mathcal{B}} s_i\right) &\leq n\left[\min\left(\min_{i\in\mathcal{B}} \sup_{x^*\in\Omega_P} x_i^*, \min_{i\in\mathcal{N}} \sup_{(\lambda^*,s^*)\in\Omega_D} s_i^*\right)\right]^{-1}\mu \\ &= \frac{n}{\epsilon(A,b,c)}\,\mu,\end{aligned}$$

where $\epsilon(A,b,c)$ was defined in (3.5). The result (5.29) follows immediately when we set

$$C_3 = \frac{\epsilon(A,b,c)}{n}. \tag{5.32}$$

Existence of a strictly complementary solution (Theorem 2.4) guarantees that $\epsilon(A,b,c) > 0$, so C_3 is positive.

Finally, since $(x,\lambda,s) \in \mathcal{N}_{-\infty}(\gamma)$, we have $x_i s_i \geq \gamma\mu$ for all $i = 1,2,\ldots,n$. Hence, we have from (5.29) that

$$s_i \geq \frac{\gamma\mu}{x_i} \geq \frac{\gamma\mu}{\mu/C_3} = C_3\gamma \qquad \text{for all } i \in \mathcal{N}.$$

In the same way, we can show that $x_i \geq C_3\gamma$ for $i \in \mathcal{B}$, proving (5.30). □

Theorem 5.14 *Let $\{(x^k, \lambda^k, s^k)\}$ be a sequence of iterates generated by Algorithm SPF, PC, or LPF, and suppose that $\mu_k \downarrow 0$ as $k \to \infty$. Then the sequence $\{(x^k, s^k)\}$ is bounded and therefore has at least one limit point. Each limit point corresponds to a strictly complementary primal-dual solution.*

Proof. All three algorithms confine their iteration sequences to a neighborhood $\mathcal{N}_{-\infty}(\gamma)$ for some $\gamma > 0$. In Algorithm SPF, we have

$$(x^k, \lambda^k, s^k) \in \mathcal{N}_2(0.4) \subset \mathcal{N}_{-\infty}(0.4).$$

In Algorithm PC, all iterates belong to $\mathcal{N}_2(0.5)$, which is a subset of $\mathcal{N}_{-\infty}(0.5)$, whereas in Algorithm LPF, the value of γ is chosen explicitly. Also, the sequences $\{\mu_k\}$ are nonincreasing for each method; in particular, $\mu_k \leq \mu_0$ for all $k \geq 0$. Hence each iterate (x^k, λ^k, s^k) satisfies the assumptions of Lemma 5.13.

Boundedness of $\{(x^k, s^k)\}$ follows from (5.28). If (x^*, s^*) is a limit point, we can find $\lambda^* \in \mathbb{R}^m$ such that $(x^*, \lambda^*, s^*) \in \Omega$ (see the discussion above). Because of (5.30), we must have

$$s_i^* \geq C_3\gamma > 0 \ \ (i \in \mathcal{N}), \qquad x_i^* \geq C_3\gamma > 0 \ \ (i \in \mathcal{B}),$$

so the solution is strictly complementary. □

When the problem has a unique primal-dual solution (x^*, λ^*, s^*), it follows immediately from Theorem 5.14 that the iteration sequences for all three algorithms converge to this point.

Proof of Lemma 5.3

We return to the technical lemma stated earlier in the chapter, which claimed that for any vector pair u, v with $u^T v \geq 0$, we have

$$\|UVe\| \leq 2^{-3/2}\|u + v\|^2.$$

First, note that for any two scalars α and β with $\alpha\beta \geq 0$, we have from the algebraic-geometric mean inequality that

$$\sqrt{|\alpha\beta|} \leq \frac{1}{2}|\alpha + \beta|. \tag{5.33}$$

Since $u^T v \geq 0$, we have

$$0 \leq u^T v = \sum_{u_i v_i \geq 0} u_i v_i + \sum_{u_i v_i < 0} u_i v_i = \sum_{i \in \mathcal{P}} |u_i v_i| - \sum_{i \in \mathcal{M}} |u_i v_i|, \tag{5.34}$$

where we partitioned the index set $\{1, 2, \ldots, n\}$ as

$$\mathcal{P} = \{i \,|\, u_i v_i \geq 0\}, \qquad \mathcal{M} = \{i \,|\, u_i v_i < 0\}.$$

Now,

$$\begin{aligned}
\|UVe\| &= \left(\|[u_iv_i]_{i\in\mathcal{P}}\|^2 + \|[u_iv_i]_{i\in\mathcal{M}}\|^2\right)^{1/2} \\
&\le \left(\|[u_iv_i]_{i\in\mathcal{P}}\|_1^2 + \|[u_iv_i]_{i\in\mathcal{M}}\|_1^2\right)^{1/2} && \text{since } \|\cdot\|_2 \le \|\cdot\|_1 \\
&\le \left(2\,\|[u_iv_i]_{i\in\mathcal{P}}\|_1^2\right)^{1/2} && \text{from (5.34)} \\
&\le \sqrt{2}\left\|\left[\frac{1}{4}(u_i+v_i)^2\right]_{i\in\mathcal{P}}\right\|_1 && \text{from (5.33)} \\
&= 2^{-3/2}\sum_{i\in\mathcal{P}}(u_i+v_i)^2 \\
&\le 2^{-3/2}\sum_{i=1}^{n}(u_i+v_i)^2 \\
&= 2^{-3/2}\|u+v\|^2,
\end{aligned}$$

completing the proof.

Notes and References

Xu [152] has described a way to avoid the tight confines of the neighborhood $\mathcal{N}_2(\theta)$, $\theta < 1$ without sacrificing $O(\sqrt{n}L)$ complexity. He uses the neighborhood $\mathcal{N}_2(\theta) \cap \mathcal{N}_{-\infty}(\gamma)$, where $\gamma > 0$ is small but θ may be *much larger* than 1. Steps in this neighborhood can be almost as long as in the neighborhood $\mathcal{N}_{-\infty}$ alone, and the strategy has been successfully implemented by Xu, Hung, and Ye [153].

Lemma 5.13 and Theorem 5.14 are due to Güler and Ye [50]. The set of limit points of $\{(x^k, s^k)\}$ actually forms a continuum (see Tapia, Zhang, and Ye [126, Theorem 4.1]). Convergence of $\{(x^k, s^k)\}$ is discussed further in Chapters 6 and 7.

Exercises

1. Check that the choices of θ and σ used in Algorithm SPF satisfy the relationship (5.15).

2. Prove (5.6). Does this result still hold for the infeasible step (1.20)?

3. For a given $\theta \in (0,1)$, the problem of finding a value of σ that satisfies (5.15) while maximizing the decrease in μ for a unit step can be posed as a simple constrained optimization problem. Write down this problem. Does this problem have a solution for all $\theta \in (0,1)$? Explain.

4. Theorem 5.6 shows that the unit step for Algorithm SPF keeps the new iterate inside the neighborhood $\mathcal{N}_2(\theta)$. Express the problem of finding the maximal α such that $\|X(\alpha)S(\alpha)e - \mu(\alpha)e\|_2 \leq \theta\mu(\alpha)$ as a quartic polynomial in α.

5. In Algorithms SPF and LPF, is there any benefit to taking a step of length *longer* than 1, if such a step remains inside the required neighborhood?

6. In Algorithm PC, we chose the inner and outer neighborhoods to be $\mathcal{N}_2(0.25)$ and $\mathcal{N}_2(0.5)$, respectively. In this exercise, you are asked to consider more general neighborhoods $\mathcal{N}_2(\theta_{\text{in}})$ and $\mathcal{N}_2(\theta_{\text{out}})$ and look for conditions on the scalars θ_{in} and θ_{out} for which the analysis of the algorithm continues to hold.

   (i) Redefine the lower bound $\bar{\alpha}$ on the predictor step length obtained in Lemma 5.7 for arbitrary values of θ_{in} and θ_{out}. That is, find a value η such that

$$\bar{\alpha} = \min\left(\tfrac{1}{2}, \left(\frac{\eta\mu}{\|\Delta X \Delta S e\|}\right)^{1/2}\right).$$

   (ii) Modify the analysis of Lemma 5.8 to find the relationship between θ_{in} and θ_{out} that guarantees that the corrector step returns to the inner neighborhood.

   (iii) What is the largest value of θ_{out} for which the θ_{in} of part (ii) satisfies $\theta_{\text{in}} \in (0, \theta_{\text{out}})$?

7. Prove that when $n = 2$, the neighborhoods $\mathcal{N}_2(\theta)$ and $\mathcal{N}_{-\infty}(1 - \theta/\sqrt{2})$ are identical. Does a similar relationship hold when $n = 3$?