7.9. For the dual affine scaling algorithm, explain the meaning of "primal estimate" as defined in (7.59).

7.10. For the primal-dual algorithm, try to decompose $\mathbf{d}_w^k$ and $\mathbf{d}_s^k$ as we did for $\mathbf{d}_x^k$ in (7.113). Then analyze different components.

7.11. We take $\mathbf{x}^0 = \mathbf{e}$, $\mathbf{w}^0 = \mathbf{0}$, and $\mathbf{s}^0 = \mathbf{e}$.

(a) Show that (7.112a) becomes $\pi > 0$.

(b) Show that (7.112b) becomes $\lambda > n - \mathbf{c}^T\mathbf{e}$.

(c) What about (7.112c) and (7.112d)?

7.12. Derive the power-series expansions for $\mathbf{x}(\beta)$, $\mathbf{w}(\beta)$, $\mathbf{s}(\beta)$, $\mathbf{t}(\beta)$, $\mathbf{u}(\beta)$, and $\mathbf{v}(\beta)$ in the primal-dual algorithm.

7.13. Develop computer codes for the primal affine scaling, dual affine scaling, and primal-dual algorithms and test those problems in Exercise 3.16.

8

Insights into the Interior-Point Methods

In Chapter 7 we have studied three polynomial-time interior-point algorithms, namely the primal affine scaling with logarithmic barrier function, the dual affine scaling with logarithmic barrier function, and the primal-dual algorithms. Actually they are strongly connected and can be treated by an integrated approach. In this chapter we first show that the moving directions of these three algorithms are merely the *Newton directions* along three different *algebraic paths* that lead to the solution of the Karush-Kuhn-Tucker conditions of a given linear programming problem under suitable assumptions. Moreover, the dual information embedded in the primal algorithm and the primal information embedded in the dual algorithm can be recovered in the primal-dual algorithm but with different scaling matrices. Based on these findings, we then introduce a general theory of constructing new interior-point methods.

8.1 MOVING ALONG DIFFERENT ALGEBRAIC PATHS

Let us consider a linear programming problem (Program P) in its standard form:

$$\text{Minimize} \quad \mathbf{c}^T\mathbf{x} \tag{8.1a}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0} \tag{8.1b}$$

where $\mathbf{A}$ is an $m \times n$ matrix. Its dual problem (Program D) is in the following form:

$$\text{Maximize} \quad \mathbf{b}^T\mathbf{w} \tag{8.2a}$$

$$\text{subject to} \quad \mathbf{A}^T\mathbf{w} + \mathbf{s} = \mathbf{c}, \quad \mathbf{s} \geq \mathbf{0} \tag{8.2b}$$

For any positive scalar μ, we can incorporate a logarithmic barrier function either into the primal program P and consider a corresponding problem (Program P_μ):

$$\text{Minimize} \quad \mathbf{c}^T\mathbf{x} - \mu \sum_{j=1}^{n} \log_e x_j \tag{8.3a}$$

$$\text{subject to} \quad \mathbf{Ax} = \mathbf{b}, \qquad \mathbf{x} > \mathbf{0} \tag{8.3b}$$

or into the dual program D and consider a corresponding problem (Program D_μ):

$$\text{Maximize} \quad \mathbf{b}^T\mathbf{w} + \mu \sum_{j=1}^{n} \log_e s_j \tag{8.4a}$$

$$\text{subject to} \quad \mathbf{A}^T\mathbf{w} + \mathbf{s} = \mathbf{c}, \qquad \mathbf{s} > \mathbf{0} \tag{8.4b}$$

In Chapter 7, we have seen that the Karush-Kuhn-Tucker conditions of programs P_μ and D_μ lead to the same system:

$$\mathbf{A}^T\mathbf{w} + \mathbf{s} - \mathbf{c} = \mathbf{0} \tag{8.5a}$$

$$\mathbf{Ax} - \mathbf{b} = \mathbf{0} \tag{8.5b}$$

$$\mathbf{XSe} - \mu\mathbf{e} = \mathbf{0} \tag{8.5c}$$

$$\mathbf{x} > \mathbf{0}, \mathbf{s} > \mathbf{0} \tag{8.5d}$$

where $\mathbf{X}$ and $\mathbf{S}$ are diagonal matrices using the components of vectors $\mathbf{x}$ and $\mathbf{s}$ as diagonal elements, respectively.

To assure the existence of a unique optimal solution to program P_μ and program D_μ, or equivalently the existence of a unique solution to the system (8.5), we assume that

(A1) There exists a primal interior feasible solution, i.e.,

$$\mathbf{S} \equiv \left\{\mathbf{x} \in \mathbf{R}^n \mid \mathbf{Ax} = \mathbf{b}, \mathbf{x} > \mathbf{0}\right\} \neq \phi$$

(A2) There exists a dual interior feasible solution, i.e.,

$$\mathbf{T} \equiv \left\{(\mathbf{w}; \mathbf{s}) \in \mathbf{R}^m \times \mathbf{R}^n \mid \mathbf{A}^T\mathbf{w} + \mathbf{s} = \mathbf{c}, \mathbf{s} > \mathbf{0}\right\} \neq \phi$$

(A3) The constraint matrix $\mathbf{A}$ has full row rank.

Notice that, under the above assumptions, as μ approaches 0, the unique solution to the system of equations (8.5) solves the given linear program P and its dual problem D. However, for $\mu > 0$, we can actually approach the solution of $\mathbf{XSe} - \mu\mathbf{e} = \mathbf{0}$ from different but equivalent *algebraic paths*. To be more specific, for $x_j > 0$ and $s_j > 0$ $(j = 1, \ldots, n)$, consider the following functions:

$$f\left(x_j, s_j\right) = \mu - x_j s_j \tag{8.6a}$$

$$g\left(x_j, s_j\right) = \frac{\mu}{x_j} - s_j \tag{8.6b}$$

$$h\left(x_j, s_j\right) = \frac{\mu}{s_j} - x_j \tag{8.6c}$$

Although they are different in format, the above three functions are all algebraically equivalent to the condition (8.5c), since

$$\begin{aligned}
&\left\{(\mathbf{x}, \mathbf{s}) \in R^{2n} \mid f(x_j, s_j) = 0, x_j > 0, s_j > 0, \text{ for } j = 1, \ldots, n\right\} \\
&= \left\{(\mathbf{x}, \mathbf{s}) \in R^{2n} \mid g(x_j, s_j) = 0, x_j > 0, s_j > 0, \text{ for } j = 1, \ldots, n\right\} \\
&= \left\{(\mathbf{x}, \mathbf{s}) \in R^{2n} \mid h(x_j, s_j) = 0, x_j > 0, s_j > 0, \text{ for } j = 1, \ldots, n\right\} \\
&= \left\{(\mathbf{x}, \mathbf{s}) \in R^{2n} \mid \mathbf{XSe} - \mu\mathbf{e} = \mathbf{0}, \mathbf{x} > \mathbf{0}, \mathbf{s} > \mathbf{0}\right\}
\end{aligned}$$

In this way solving system (8.5) is equivalent to solving one of the following three systems:

$$\mathbf{A}^T\mathbf{w} + \mathbf{s} - \mathbf{c} = \mathbf{0} \tag{8.7a}$$

$$\mathbf{Ax} - \mathbf{b} = \mathbf{0} \tag{8.7b}$$

$$f(x_j, s_j) = 0, \qquad \text{for } j = 1, \ldots, n \tag{8.7c}$$

$$\mathbf{x} > \mathbf{0}, \mathbf{s} > \mathbf{0} \tag{8.7d}$$

$$\mathbf{A}^T\mathbf{w} + \mathbf{s} - \mathbf{c} = \mathbf{0} \tag{8.8a}$$

$$\mathbf{Ax} - \mathbf{b} = \mathbf{0} \tag{8.8b}$$

$$g(x_j, s_j) = 0, \qquad \text{for } j = 1, \ldots, n \tag{8.8c}$$

$$\mathbf{x} > \mathbf{0}, \mathbf{s} > \mathbf{0} \tag{8.8d}$$

$$\mathbf{A}^T\mathbf{w} + \mathbf{s} - \mathbf{c} = \mathbf{0} \tag{8.9a}$$

$$\mathbf{Ax} - \mathbf{b} = \mathbf{0} \tag{8.9b}$$

$$h(x_j, s_j) = 0, \qquad \text{for } j = 1, \ldots, n \tag{8.9c}$$

$$\mathbf{x} > \mathbf{0}, \mathbf{s} > \mathbf{0} \tag{8.9d}$$

To solve any one of the above three systems, let us assume that $\left(\mathbf{x}^k; \mathbf{w}^k; \mathbf{s}^k\right) \in R^n \times R^m \times R^n$ such that $\mathbf{A}^T\mathbf{w}^k + \mathbf{s}^k = \mathbf{c}$, $\mathbf{Ax}^k = \mathbf{b}$, $\mathbf{x}^k > \mathbf{0}$, and $\mathbf{s}^k > \mathbf{0}$. We shall apply the famous Newton method to solve these systems at $(\mathbf{x}^k; \mathbf{w}^k; \mathbf{s}^k)$. Note that only functions f, g, and h are nonlinear in these three systems. Therefore, when the Newton method is applied, we need only linearize them for obtaining a moving direction.

8.1.1 Primal Affine Scaling with Logarithmic Barrier Function

Let us focus on system (8.8) first. Taking one Newton step with a linearization of the function $g(x_j, s_j) = 0$, we have

$$0 - g(x_j^k, s_j^k) = \left[\nabla g(x_j^k, s_j^k)\right]^T \begin{pmatrix} x_j - x_j^k \\ s_j - s_j^k \end{pmatrix}$$

Substituting (8.6b) for the function g and multiplying it out, we see that

$$s_j^k - \frac{\mu}{x_j^k} = \left(-\frac{\mu}{(x_j^k)^2}, -1\right)\begin{pmatrix} x_j - x_j^k \\ s_j - s_j^k \end{pmatrix}$$

Consequently, we have

$$s_j = \frac{2\mu}{x_j^k} - \left(\frac{\mu}{(x_j^k)^2}\right) x_j$$

Since the above equation holds for $j = 1, \ldots, n$, by taking matrix $\mathbf{X}_k = \text{diag}\,(\mathbf{x}^k)$, we have

$$\mathbf{s} = 2\mu \mathbf{X}_k^{-2}\mathbf{x}^k - \mu \mathbf{X}_k^{-2}\mathbf{x} \tag{8.10}$$

Moving along Newton direction, the linear equations (8.8a) and (8.8b) are preserved. By (8.8a), $\mathbf{s} = \mathbf{c} - \mathbf{A}^T\mathbf{w}$ and (8.10) becomes

$$\mathbf{x} = \frac{1}{\mu}\mathbf{X}_k^2(\mathbf{A}^T\mathbf{w} + 2\mu\mathbf{X}_k^{-2}\mathbf{x}^k - \mathbf{c})$$

Multiplying both sides by matrix A and applying (8.8b), we see that

$$\mathbf{b} = \mathbf{A}\mathbf{x} = \frac{1}{\mu}\mathbf{A}\mathbf{X}_k^2(\mathbf{A}^T\mathbf{w} + 2\mu\mathbf{X}_k^{-2}\mathbf{x}^k - \mathbf{c})$$

Consequently,

$$\mathbf{w} = (\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{X}_k^2\mathbf{c} - \mu\mathbf{b})$$

Plugging $\mathbf{w}$ into the formula of $\mathbf{x}$, we have

$$\mathbf{x} = \frac{1}{\mu}\mathbf{X}_k^2[\mathbf{A}^T(\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{X}_k^2\mathbf{c} - \mu\mathbf{b}) + 2\mu\mathbf{X}_k^{-2}\mathbf{x}^k - \mathbf{c}]$$

Notice that $\mathbf{A}\mathbf{X}_k\mathbf{e} = \mathbf{A}\mathbf{x}_k = \mathbf{b}$, we see the Newton direction is given by

$$\begin{aligned}\Delta\mathbf{x}_k &= \mathbf{x} - \mathbf{x}^k \\ &= -\frac{1}{\mu}\mathbf{X}_k[\mathbf{I} - \mathbf{X}_k\mathbf{A}^T(\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{X}_k](\mathbf{X}_k\mathbf{c} - \mu\mathbf{e})\end{aligned}$$

Since the above direction is exactly the same as the direction given by formula (7.49a) at $\mathbf{x} = \mathbf{x}^k$, we can conclude that the primal affine scaling algorithm with logarithmic barrier function actually takes the Newton direction along the algebraic path of $g(\mathbf{x}, \mathbf{s}) = \mathbf{0}$.

8.1.2 Dual Affine Scaling with Logarithmic Barrier Function

This time, let us focus on the system (8.9) to show that the dual affine scaling algorithm with logarithmic barrier function actually takes the Newton direction along the algebraic path of $h(\mathbf{x}, \mathbf{s}) = \mathbf{0}$. Note that one Newton step with the linearization of $h(x_j, s_j) = 0$ results in

$$0 - h(x_j^k, s_j^k) = [\nabla h(x_j^k, s_j^k)]^T \begin{pmatrix} x_j - x_j^k \\ s_j - s_j^k \end{pmatrix}$$

Using formula (8.6c) for function h, we have

$$x_j^k - \frac{\mu}{s_j^k} = \left(-1, -\frac{\mu}{(s_j^k)^2}\right)\begin{pmatrix} x_j - x_j^k \\ s_j - s_j^k \end{pmatrix}$$

and

$$x_j = \frac{2\mu}{s_j^k} - \left(\frac{\mu}{(s_j^k)^2}\right) s_j$$

Note the above equation holds for $j = 1, \ldots, n$. By taking matrix $\mathbf{S}_k = \text{diag}\,(\mathbf{s}^k)$, we have

$$\mathbf{x} = 2\mu\mathbf{S}_k^{-1}\mathbf{e} - \mu\mathbf{S}_k^{-2}\mathbf{s} \tag{8.11}$$

Again, moving along the Newton direction preserves the linear equations (8.9a) and (8.9b). By (8.9b), we have

$$\begin{aligned}\mathbf{b} = \mathbf{A}\mathbf{x} &= 2\mu\mathbf{A}\mathbf{S}_k^{-1}\mathbf{e} - \mu\mathbf{A}\mathbf{S}_k^{-2}\mathbf{s} \\ &= 2\mu\mathbf{A}\mathbf{S}_k^{-1}\mathbf{e} - \mu\mathbf{A}\mathbf{S}_k^{-2}(\mathbf{c} - \mathbf{A}^T\mathbf{w})\end{aligned}$$

However (8.9a) says that $\mathbf{c} = \mathbf{A}^T\mathbf{w}^k + \mathbf{s}^k$, hence

$$\mathbf{b} = 2\mu\mathbf{A}\mathbf{S}_k^{-1}\mathbf{e} - \mu\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T\mathbf{w}^k - \mu\mathbf{A}\mathbf{S}_k^{-2}\mathbf{s}^k + \mu\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T\mathbf{w}$$

Therefore, we finally obtain the Newton direction

$$\begin{aligned}\Delta\mathbf{w}_k &= \mathbf{w} - \mathbf{w}^k \\ &= \frac{1}{\mu}(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T)^{-1}\mathbf{b} - (\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{S}_k^{-1}\mathbf{e}\end{aligned}$$

Comparing this direction to (7.73), we see that the dual affine scaling algorithm with logarithmic barrier function takes the Newton direction along the algebraic path of $h(\mathbf{x}, \mathbf{s}) = \mathbf{0}$.

8.1.3 The Primal-Dual Algorithm

Finally, we work on the system (8.7) to derive the moving directions of the primal-dual algorithm. Simply by taking one Newton step with a linearization of the function $f(x_j, s_j) = 0$, we have

$$0 - f(x_j^k, s_j^k) = [\nabla f(x_j^k, s_j^k)]^T \begin{pmatrix} x_j - x_j^k \\ s_j - s_j^k \end{pmatrix}$$

Using formula (8.6a) for function f, we have

$$x_j^k s_j^k - \mu = -\left(s_j^k, x_j^k\right)\begin{pmatrix} x_j - x_j^k \\ s_j - s_j^k \end{pmatrix}$$

Note the above equation holds for $j = 1, \ldots, n$, hence

$$\mathbf{X}_k \Delta\mathbf{s}_k + \mathbf{S}_k \Delta\mathbf{x}_k = -\mathbf{X}_k\mathbf{S}_k\mathbf{e} + \mu\mathbf{e} \tag{8.12a}$$

Moreover, moving along the Newton direction assures that

$$\mathbf{A}\Delta\mathbf{x}_k = \mathbf{0} \tag{8.12b}$$

and

$$\mathbf{A}^T\Delta\mathbf{w}_k + \Delta\mathbf{s}_k = \mathbf{0} \tag{8.12c}$$

Note that (8.12a), (8.12b), and (8.12c) form a system of linear equations with unknown variables $\Delta\mathbf{x}_k$, $\Delta\mathbf{w}_k$, and $\Delta\mathbf{s}_k$. Using (8.12b) and (8.12c) to eliminate $\Delta\mathbf{x}_k$ and $\Delta\mathbf{s}_k$ in (8.12a), we obtain

$$\begin{aligned}\Delta\mathbf{w}_k &= (\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{S}_k^{-1}(\mathbf{X}_k\mathbf{S}_k\mathbf{e} - \mu\mathbf{e}) \\ &= -(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{S}_k^{-1}\mathbf{v}^k(\mu).\end{aligned} \tag{8.13a}$$

where $\mathbf{v}^k(\mu) = \mu\mathbf{e} - \mathbf{X}_k\mathbf{S}_k\mathbf{e}$.

Plugging $\Delta\mathbf{w}_k$ into (8.12c), we have

$$\begin{aligned}\Delta\mathbf{s}_k &= -\mathbf{A}^T\Delta\mathbf{w}_k \\ &= \mathbf{A}^T(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{S}_k^{-1}\mathbf{v}^k(\mu).\end{aligned} \tag{8.13b}$$

After $\Delta\mathbf{s}_k$ is known, $\Delta\mathbf{x}_k$ immediately follows from (8.12a) as

$$\begin{aligned}\Delta\mathbf{x}_k &= -[\mathbf{S}_k^{-1} - \mathbf{S}_k^{-1}\mathbf{X}_k\mathbf{A}^T(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{S}_k^{-1}](\mathbf{X}_k\mathbf{S}_k\mathbf{e} - \mu\mathbf{e}) \\ &= [\mathbf{S}_k^{-1} - \mathbf{S}_k^{-1}\mathbf{X}_k\mathbf{A}^T(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{S}_k^{-1}]\mathbf{v}^k(\mu) \\ &= \mathbf{S}_k^{-1}[\mathbf{v}^k(\mu) - \mathbf{X}_k\Delta\mathbf{s}_k]\end{aligned} \tag{8.13c}$$

Comparing (8.13) to formula (7.90), we clearly see that the primal-dual algorithm takes the Newton direction along the algebraic path of $f(\mathbf{x}, \mathbf{s}) = \mathbf{0}$.

Now, combining the results we obtained in the previous three subsections results in the following theorem:

Theorem 8.1. The moving directions of the primal affine scaling algorithm with logarithmic barrier function, the dual affine scaling algorithm with logarithmic barrier function, and the primal-dual algorithm are the Newton directions along three different and yet equivalent algebraic paths that lead to the solution of the Karush-Kuhn-Tucker conditions (8.5).

8.2 MISSING INFORMATION

In Chapter 7, the primal approach and dual approach were treated separately as if they were independent problems. However, Theorem 8.1 indicates that the moving directions of both the primal affine scaling and dual affine scaling with logarithmic barrier function are closely related to that of the primal-dual method. Hence we shall further exploit the dual information in the primal approach and the primal information in the dual approach.

8.2.1 Dual Information in the Primal Approach

We first study the dual information in the primal affine scaling algorithm. From (8.10), we have

$$\begin{aligned}\mathbf{s} &= 2\mu\mathbf{X}_k^{-2}\mathbf{x}^k - \mu\mathbf{X}_k^{-2}\mathbf{x} \\ &= 2\mu\mathbf{X}_k^{-2}\mathbf{x}^k - \mu\mathbf{X}_k^{-2}(\mathbf{x}^k + \Delta\mathbf{x}_k) \\ &= \mu\mathbf{X}_k^{-2}\mathbf{x}^k - \mu\mathbf{X}_k^{-2}\mathbf{X}_k[\mathbf{I} - \mathbf{X}_k\mathbf{A}^T(\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{X}_k]\left(-\frac{1}{\mu}\mathbf{X}_k\mathbf{c} + \mathbf{e}\right) \\ &= \mu\mathbf{X}_k^{-2}\mathbf{x}^k - \mu\mathbf{X}_k^{-2}\mathbf{X}_k\left(-\frac{1}{\mu}\mathbf{X}_k\mathbf{c} + \mathbf{e}\right) + \mu\mathbf{A}^T(\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{X}_k\left(-\frac{1}{\mu}\mathbf{X}_k\mathbf{c} + \mathbf{e}\right) \\ &= \mathbf{c} - \mathbf{A}^T(\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{X}_k(\mathbf{X}_k\mathbf{c} - \mu\mathbf{e})\end{aligned}$$

Since we are moving along the Newton direction, both the primal and dual feasibility conditions are preserved. Hence we can define that

$$\mathbf{w} = (\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{X}_k(\mathbf{X}_k\mathbf{c} - \mu\mathbf{e})$$

In this way, we find the dual information

$$\begin{aligned}\Delta\mathbf{w}_k = \mathbf{w} - \mathbf{w}^k &= \left(\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k(\mathbf{X}_k\mathbf{c} - \mu\mathbf{e}) - \mathbf{w}^k \\ &= \left(\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k^2(\mathbf{c} - \mathbf{A}^T\mathbf{w}^k - \mu\mathbf{X}_k^{-1}\mathbf{e}) \\ &= \left(\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k(\mathbf{X}_k\mathbf{S}_k\mathbf{e} - \mu\mathbf{e}) \\ &= -\left(\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k\mathbf{v}^k(\mu)\end{aligned} \tag{8.14}$$

Comparing (8.14) with (8.13a), we see that the dual moving direction embedded in the primal affine scaling algorithm with logarithmic barrier function has exactly the same form as that of the primal-dual algorithm but with different scaling matrix, which, of course, depends on the primal information $\mathbf{X}_k$ only.

8.2.2 Primal Information in the Dual Approach

Similar to what we did in the last subsection, we can derive the embedded primal information of the dual affine scaling. Starting from Equation (8.11), we have

$$
\begin{aligned}
\mathbf{x} &= 2\mu \mathbf{S}_k^{-1}\mathbf{e} - \mu \mathbf{S}_k^{-2}\mathbf{s} \\
&= 2\mu \mathbf{S}_k^{-1}\mathbf{e} - \mu \mathbf{S}_k^{-2}\left[\mathbf{s}^k - \frac{1}{\mu}\mathbf{A}^T\left(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T\right)^{-1}\left(\mathbf{b} - \mu \mathbf{A}\mathbf{S}_k^{-1}\mathbf{e}\right)\right] \\
&= \mu \mathbf{S}_k^{-1}\left[\mathbf{e} + \mathbf{S}_k^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T\right)^{-1}\left(\frac{1}{\mu}\mathbf{A}\mathbf{X}_k\mathbf{e} - \mathbf{A}\mathbf{S}_k^{-1}\mathbf{e}\right)\right] \\
&= \mu \mathbf{S}_k^{-1}\left[\mathbf{e} + \mathbf{S}_k^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{S}_k^{-1}\left(\frac{1}{\mu}\mathbf{S}_k\mathbf{X}_k\mathbf{e} - \mathbf{e}\right)\right] \\
&= \mu \mathbf{S}_k^{-1}\left[\mathbf{I} - \mathbf{S}_k^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{S}_k^{-1}\right]\left(\frac{-1}{\mu}\mathbf{S}_k\mathbf{X}_k\mathbf{e} + \mathbf{e}\right) + \mathbf{x}^k
\end{aligned}
$$

Hence we know

$$
\begin{aligned}
\Delta \mathbf{x}_k &= -\left[\mathbf{S}_k^{-1} - \mathbf{S}_k^{-2}\mathbf{A}^T\left(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{S}_k^{-1}\right](\mathbf{X}_k\mathbf{S}_k\mathbf{e} - \mu\mathbf{e}) \\
&= \left[\mathbf{S}_k^{-1} - \mathbf{S}_k^{-2}\mathbf{A}^T\left(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{S}_k^{-1}\right]\mathbf{v}^k(\mu) \qquad (8.15)
\end{aligned}
$$

Comparing (8.15) to (8.13c), we see that the primal moving direction embedded in the dual affine scaling algorithm with logarithmic barrier function has exactly the same form as that of the primal-dual algorithm but with a different scaling matrix.

The results we found in the above two subsections can be summarized in the following theorem:

Theorem 8.2. The form of either the dual moving direction embedded in the primal affine scaling or the primal moving direction embedded in the dual affine scaling can be found in the primal-dual algorithm but with different scaling matrices.

8.3 EXTENSIONS OF ALGEBRAIC PATHS

The concept of "moving along the Newton direction on different algebraic paths" not only provides us a unified view to examine the primal affine scaling with logarithmic barrier function, dual affine scaling with logarithmic barrier function, and primal-dual algorithms but also serves as a platform to study new interior-point algorithms. At least in theory there are infinitely many algebraic paths that could lead to the solution of the given Karush-Kuhn-Tucker conditions, and each path may generate a new moving direction associated with a potential interior-point algorithm. If a suitable step-length can be decided at each iteration for a convergence proof, then a new interior-point algorithm is introduced for further studies. An example of moving along a new path is given below.

Consider the function

$$\tau(x_j, s_j) = \log_e \frac{x_j s_j}{\mu}$$

defined on $x_j > 0, s_j > 0, j = 1, 2, \ldots, n$, and $\mu > 0$. In this way, solving system (8.5) is equivalent to solving the following system:

$$
\begin{aligned}
\mathbf{A}^T\mathbf{w} + \mathbf{s} - \mathbf{c} &= \mathbf{0} && (8.16a) \\
\mathbf{A}\mathbf{x} - \mathbf{b} &= \mathbf{0} && (8.16b) \\
\tau(x_j, s_j) &= 0, \quad \text{for } j = 1, \ldots, n && (8.16c) \\
\mathbf{x} > \mathbf{0}, \mathbf{s} &> \mathbf{0} && (8.16d)
\end{aligned}
$$

We consider the moving direction at a given point $(\mathbf{x}^k; \mathbf{w}^k; \mathbf{s}^k)$ such that $\mathbf{A}\mathbf{x}^k = \mathbf{b}$, $\mathbf{A}^T\mathbf{w}^k + \mathbf{s}^k = \mathbf{c}$, $\mathbf{x}^k > \mathbf{0}$, and $\mathbf{s}^k > \mathbf{0}$. One Newton step at this point with a linearization of the function $\tau(x_j, s_j) = 0$ yields that

$$-\log_e \frac{x_j^k s_j^k}{\mu} = \left(\frac{1}{x_j^k}, \frac{1}{s_j^k}\right)\begin{pmatrix} x_j - x_j^k \\ s_j - s_j^k \end{pmatrix}$$

Since the above expression holds for $j = 1, 2, \ldots, n$, its vector form becomes

$$\mathbf{X}_k^{-1}\Delta\mathbf{x}_k + \mathbf{S}_k^{-1}\Delta\mathbf{s}_k = -\boldsymbol{\theta}(\mu) \qquad (8.17)$$

where
$$\boldsymbol{\theta}(\mu) = \left(\log_e \frac{x_1^k s_1^k}{\mu}, \log_e \frac{x_2^k s_2^k}{\mu}, \ldots, \log_e \frac{x_n^k s_n^k}{\mu}\right)$$

Moreover, moving along the Newton direction preserves the linear equations, hence we have $\mathbf{A}\Delta\mathbf{x}_k = \mathbf{0}$ and $\mathbf{A}^T\Delta\mathbf{w}_k + \Delta\mathbf{s}_k = \mathbf{0}$. These two equations together with (8.17) form a system of linear equations in terms of $\Delta\mathbf{x}_k$, $\Delta\mathbf{w}_k$, and $\Delta\mathbf{s}_k$. The solution of this system becomes

$$
\begin{aligned}
\Delta\mathbf{x}_k &= -\left[\mathbf{X}_k - \mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k\right]\boldsymbol{\theta}(\mu) && (8.18a) \\
\Delta\mathbf{w}_k &= \left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k\boldsymbol{\theta}(\mu) && (8.18b) \\
\Delta\mathbf{s}_k &= -\mathbf{A}^T\left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k\boldsymbol{\theta}(\mu) && (8.18c)
\end{aligned}
$$

Comparing (8.18) with (8.13), we see that the moving directions along this new path are different from previous results. Which algebraic path leads to computational superiority remains an unanswered theoretical question.

8.4 GEOMETRIC INTERPRETATION OF THE MOVING DIRECTIONS

Different geometric viewpoints have been proposed to interpret the moving directions of each individual affine scaling algorithm. Our objective in this section is to provide a geometric view which at least interprets the moving directions of the primal affine with logarithmic barrier, the dual affine with logarithmic barrier, and the primal-dual algorithms in a unified way. Later on, we show that, for each of the above three algorithms, an associated minimization problem can be defined such that the solution of the associated problem becomes the moving direction of the corresponding affine scaling algorithm.

In order to achieve the goal, the concept of *null space of a matrix* needs to be strengthened through the following two lemmas:

Lemma 8.1. Assume that $m \le n$ and $\mathbf{A}$ is an $(m \times n)$-dimensional matrix with full row rank. If $\mathbf{U}$ is an $[n \times (n-m)]$-dimensional matrix of full rank such that $\mathbf{AU} = \mathbf{0}$, then

$$\left[\mathbf{I} - \mathbf{A}^T\left(\mathbf{AA}^T\right)^{-1}\mathbf{A}\right]\mathbf{x} = \mathbf{U}\left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{x}$$

for each $\mathbf{x} \in R^n$.

Proof. For each $\mathbf{x} \in R^n$, since matrix $\mathbf{A}$ has full row rank, $\mathbf{x}$ can be decomposed as

$$\mathbf{x} = \mathbf{A}^T\mathbf{u}^1 + \mathbf{U}\mathbf{u}^2$$

where $\mathbf{u}^1 \in R^m$ and $\mathbf{u}^2 \in R^{n-m}$. Hence $\mathbf{Ax} = \mathbf{AA}^T\mathbf{u}^1 + \mathbf{AUu}^2 = \mathbf{AA}^T\mathbf{u}^1$ and, consequently, $\mathbf{u}^1 = (\mathbf{AA}^T)^{-1}\mathbf{Ax}$. Similarly, we see $\mathbf{U}^T\mathbf{x} = \mathbf{U}^T\mathbf{A}^T\mathbf{u}^1 + \mathbf{U}^T\mathbf{Uu}^2 = \mathbf{U}^T\mathbf{Uu}^2$ and, consequently, $\mathbf{u}^2 = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{x}$. In other words, we have

$$\mathbf{x} = \mathbf{A}^T(\mathbf{AA}^T)^{-1}\mathbf{Ax} + \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{x} \qquad \text{for } \mathbf{x} \in R^n$$

and

$$\left[\mathbf{I} - \mathbf{A}^T(\mathbf{AA}^T)^{-1}\mathbf{A}\right]\mathbf{x} = \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{x} \qquad \text{for } \mathbf{x} \in R^n$$

Notice that, if we define an *operator* $\mathbf{P} = \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T = [\mathbf{I} - \mathbf{A}^T(\mathbf{AA}^T)^{-1}\mathbf{A}]$, then $\mathbf{P}^2 = \mathbf{P}$ and $\mathbf{AP} = \mathbf{0}$. Also note that, since matrix $\mathbf{A}$ is assumed to be of full row rank, the null space of $\mathbf{A}$ is an $(n-m)$-dimensional subspace of R^n. This subspace is, of course, isomorphic to R^{n-m}, and matrix $\mathbf{U}$ in Lemma 8.1 actually serves as an isomorphism between the null space of $\mathbf{A}$ and the Euclidean space R^{n-m}. Furthermore, we can prove the following result:

Lemma 8.2. Let $\mathbf{A}$ and $\mathbf{U}$ be defined as in Lemma 8.1 and $\mathbf{Q}$ be an $(n \times n)$-dimensional matrix which is symmetric and positive definite. Then, we have

$$\mathbf{Q}\left[\mathbf{I} - \mathbf{QA}^T\left(\mathbf{AQ}^2\mathbf{A}^T\right)^{-1}\mathbf{AQ}\right]\mathbf{Qx} = \mathbf{U}\left(\mathbf{U}^T\mathbf{Q}^{-2}\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{x}$$

Proof. Since $\mathbf{Q}$ is positive definite, $\mathbf{Q}^{-1}$ exists. If we define $\hat{\mathbf{A}} = \mathbf{AQ}$ and $\hat{\mathbf{U}} = \mathbf{Q}^{-1}\mathbf{U}$, then $\hat{\mathbf{A}}$ is an $m \times n$ matrix with full row rank and $\hat{\mathbf{U}}$ is an $n \times (n-m)$ matrix of full rank. Moreover, $\hat{\mathbf{A}}\hat{\mathbf{Q}} = \mathbf{AU} = \mathbf{0}$. The result follows from Lemma 8.1.

With these two lemmas, we can start developing a unified geometric interpretation of the moving directions in different affine scaling algorithms.

8.4.1 Primal Affine Scaling with Logarithmic Barrier Function

For the primal affine scaling algorithm with logarithmic barrier function, consider Program P_μ of (8.3). For a positive μ, we define

$$\mathbf{p}(\mathbf{x}) = \mathbf{c}^T\mathbf{x} - \mu\sum_{j=1}^{n}\log_e x_j$$

Then $\mathbf{p}(\mathbf{x})$ is a convex and continuously differentiable function defined over the constraint set (8.3b). In particular, for a given interior feasible solution $\mathbf{x}^k$, we have a first-order approximation

$$\mathbf{p}(\mathbf{x}) \approx \mathbf{p}\left(\mathbf{x}^k\right) + [\nabla\mathbf{p}(\mathbf{x}^k)]^T(\mathbf{x} - \mathbf{x}^k)$$

where $\nabla\mathbf{p}(\mathbf{x}^k) = \mathbf{c} - \mu\mathbf{X}_k^{-1}\mathbf{e}$. Finding the steepest descent direction at $\mathbf{x}^k$ is equivalent to minimizing $[\nabla\mathbf{p}(\mathbf{x}^k)]^T(\mathbf{x} - \mathbf{x}^k)$. Thus we consider a subproblem P_s of P_μ:

$$\begin{aligned} \text{Minimize} \quad & \left[\nabla\mathbf{p}\left(\mathbf{x}^k\right)\right]^T\left(\mathbf{x} - \mathbf{x}^k\right) \\ \text{subject to} \quad & \mathbf{A}\left(\mathbf{x} - \mathbf{x}^k\right) = \mathbf{0} \\ & ||\mathbf{Q}^{-1}\left(\mathbf{x} - \mathbf{x}^k\right)||^2 \le \beta^2 \end{aligned}$$

where $\mathbf{Q}^{-1}$ is the inverse matrix of an $(n \times n)$-dimensional symmetric positive definite matrix $\mathbf{Q}$ and $\beta < 1$ is a well-chosen scalar such that the surface of the ellipsoid $\{\mathbf{x} \in R^n \mid ||\mathbf{Q}^{-1}(\mathbf{x}-\mathbf{x}^k)||^2 = \beta^2\}$ becomes inscribed in the feasible domain of program P_μ. In this case, the principal axes of the ellipsoid are the eigenvectors of $\mathbf{Q}^{-1}$. In particular, if we choose $\mathbf{Q} = \mathbf{X}_k$, then $\mathbf{Q}^{-1} = \mathbf{X}_k^{-1}$ and problem P_s can be treated in a null-space version. To be more specific, by noting that $(\mathbf{x} - \mathbf{x}^k)$ is in the null space of matrix $\mathbf{A}$, we can find a vector $\mathbf{h} \in R^{n-m}$ and use the isomorphism $\mathbf{U}$ (between R^{n-m} and the null space of $\mathbf{A}$) to replace $(\mathbf{x}-\mathbf{x}^k)$ by $\mathbf{Uh}$ in problem P_s. Consequently, problem P_s becomes

$$\begin{aligned} \text{Minimize} \quad & [\nabla\mathbf{p}\left(\mathbf{x}^k\right)]^T\mathbf{Uh} \\ \text{subject to} \quad & ||\mathbf{X}_k^{-1}\mathbf{Uh}||^2 \le \beta^2 \\ & \mathbf{h} \in R^{n-m} \end{aligned}$$

Note that the above problem is solvable by considering its Lagrangian:

$$L_1(\mathbf{h}, \lambda) = [\nabla\mathbf{p}(\mathbf{x}^k)]^T\mathbf{Uh} + \lambda\left(\left(||\mathbf{X}_k^{-1}\mathbf{Uh}||\right)^2 - \beta^2\right)$$

where $\lambda \ge 0$ is a Lagrangian multiplier. Taking the partial derivative of L_1 with respect to $\mathbf{h}$ and setting it to be zero at optimum $\mathbf{h}^k$, we have

$$\mathbf{U}^T\nabla\mathbf{p}\left(\mathbf{x}^k\right) + 2\lambda\left(\mathbf{U}^T\mathbf{X}_k^{-2}\mathbf{U}\right)\mathbf{h}^k = \mathbf{0}$$

Because matrix $\mathbf{U}$ has full rank and $\mathbf{X}_k$ is a diagonal matrix, $\mathbf{U}^T\mathbf{X}_k^{-2}\mathbf{U}$ is a nonsingular square matrix. Consequently,

$$\mathbf{h}^k = \frac{-1}{2\lambda}\left(\mathbf{U}^T\mathbf{X}_k^{-2}\mathbf{U}\right)^{-1}\mathbf{U}^T\nabla\mathbf{p}\left(\mathbf{x}^k\right)$$

Remember that $\mathbf{U}$ is an isomorphism between R^{n-m} and the null space of matrix $\mathbf{A}$ in R^n. We transform $\mathbf{h}^k$ back to the null space of matrix $\mathbf{A}$ by

$$\Delta\mathbf{x}_k = \mathbf{U}\mathbf{h}^k = -\frac{1}{2\lambda}\mathbf{U}\left(\mathbf{U}^T\mathbf{X}_k^{-2}\mathbf{U}\right)^{-1}\mathbf{U}^T\nabla\mathbf{p}\left(\mathbf{x}^k\right) \tag{8.19}$$

Noting that $\nabla\mathbf{p}(\mathbf{x}^k) = \mathbf{c} - \mu\mathbf{X}_k^{-1}\mathbf{e}$, we apply Lemma 8.2 to (8.19) with $\mathbf{Q} = \mathbf{X}_k$. In this way, we see that

$$\begin{aligned}\Delta\mathbf{x}_k &= -\frac{1}{2\lambda}\mathbf{X}_k[\mathbf{I} - \mathbf{X}_k\mathbf{A}^T\left(\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k]\mathbf{X}_k\left(\mathbf{c} - \mu\mathbf{X}_k^{-1}\mathbf{e}\right)\\ &= -\frac{1}{2\lambda}\mathbf{X}_k[\mathbf{I} - \mathbf{X}_k\mathbf{A}^T\left(\mathbf{A}\mathbf{X}_k^2\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k](\mathbf{X}_k\mathbf{c} - \mu\mathbf{e})\end{aligned} \tag{8.20}$$

Comparing (8.20) with (7.49a) and noting that $1/2\lambda$ is a positive scalar which does not alter the direction of a vector, we can conclude that the moving direction of the primal affine scaling with logarithmic barrier function algorithm is provided by the solution of the subprogram P_s. This also provides a geometric interpretation of the abovementioned moving direction.

8.4.2 Dual Affine Scaling with Logarithmic Barrier Function

With the same idea, we now consider the dual case. This time, we define

$$\mathbf{q}(\mathbf{w}, \mathbf{s}) = \mathbf{b}^T\mathbf{w} + \mu\sum_{j=1}^{n}\log_e s_j$$

and assume that $(\mathbf{w}^k, \mathbf{s}^k)$ is a solution to program D_μ of (8.4). In this way, $[\nabla\mathbf{q}(\mathbf{w}^k, \mathbf{s}^k)]^T = (\mathbf{b}^T, \mu\mathbf{e}^T\mathbf{S}_k^{-1})$. Since w-variables are unrestricted, we only have to construct an ellipsoid in the s-space and consider the following subproblem D_s of program D_μ:

$$\begin{aligned}\text{Maximize}\quad & (\mathbf{b}^T, \mu\mathbf{e}^T\mathbf{S}_k^{-1})\begin{pmatrix}\mathbf{w} - \mathbf{w}^k\\ \mathbf{s} - \mathbf{s}^k\end{pmatrix}\\ \text{subject to}\quad & [\mathbf{A}^T \mid \mathbf{I}_n]\begin{pmatrix}\mathbf{w} - \mathbf{w}^k\\ \mathbf{s} - \mathbf{s}^k\end{pmatrix} = \mathbf{0}\\ & \left\|\mathbf{Q}^{-1}\left(\mathbf{s} - \mathbf{s}^k\right)\right\|^2 \le \beta^2\end{aligned}$$

where $\mathbf{Q}^{-1}$ is the inverse matrix of an $(n \times n)$-dimensional symmetric positive definite matrix $\mathbf{Q}$ and $\beta < 1$ is a well-chosen scalar such that the nonnegativity constraints of program D_μ are replaced by the inscribing ellipsoid $\{\mathbf{s} \in R^n \mid \|\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{s}^k)\|^2 \le \beta^2\}$. In particular, $\mathbf{Q}^{-1}$ can be chosen as $\mathbf{S}_k^{-1}$ for the consideration of a null space version of program D_s.

To be more specific, we let

$$\hat{\mathbf{U}} = \begin{bmatrix}\mathbf{I}_m\\ -\mathbf{A}^T\end{bmatrix} \quad\text{and}\quad \hat{\mathbf{A}} = [\mathbf{A}^T \mid \mathbf{I}_n]$$

Then $\hat{\mathbf{A}}\hat{\mathbf{U}} = \mathbf{0}$ and $\hat{\mathbf{U}}^T$ can be considered as an isomorphism between R^m and the null space of $\hat{\mathbf{A}}$, i.e.,

$$\begin{pmatrix}\mathbf{w} - \mathbf{w}^k\\ \mathbf{s} - \mathbf{s}^k\end{pmatrix} = \hat{\mathbf{U}}\mathbf{v}, \quad \text{for some} \quad \mathbf{v} \in R^m$$

In other words, we have $\Delta\mathbf{w}_k = \mathbf{v}$, $\Delta\mathbf{s}_k = -\mathbf{A}^T\mathbf{v}$. Moreover, the subproblem D_s becomes

$$\begin{aligned}\text{Maximize}\quad & (\mathbf{b}^T, \mu\mathbf{e}^T\mathbf{S}_k^{-1})\hat{\mathbf{U}}\mathbf{v}\\ \text{subject to}\quad & \|-\mathbf{S}_k^{-1}\mathbf{A}^T\mathbf{v}\|^2 \le \beta^2\\ & \mathbf{v} \in R^m\end{aligned}$$

To solve this problem, we consider its Lagrangian

$$L_2(\mathbf{v}, \lambda) = \mathbf{b}^T\mathbf{v} - \mu\mathbf{e}^T\mathbf{S}_k^{-1}\mathbf{A}^T\mathbf{v} - \lambda\left(\left\|-\mathbf{S}_k^{-1}\mathbf{A}^T\mathbf{v}\right\|^2 - \beta^2\right) \tag{8.21}$$

By setting its partial derivative with respect to $\mathbf{v}$ to be zero at the optimal solution $\mathbf{v}^k$ and applying Lemma 8.2, we eventually have

$$\Delta\mathbf{w}_k = \mathbf{v}^k = \frac{1}{2\lambda}\left[\left(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T\right)^{-1}\mathbf{b} - \mu\left(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{S}_k^{-1}\mathbf{e}\right] \tag{8.22a}$$

and

$$\Delta\mathbf{s}_k = -\mathbf{A}^T\mathbf{v}^k = -\mathbf{A}^T\Delta\mathbf{w}_k \tag{8.22b}$$

Note that $1/2\lambda$ is only a positive scalar. By comparing (8.22) to (7.73), we conclude that the moving direction of the dual affine scaling with logarithmic barrier function algorithm is provided by the solution of the subprogram D_s. This is consistent with the geometric interpretation we derived for the primal case.

8.4.3 The Primal-Dual Algorithm

In order to interpret the moving directions of the primal-dual interior-point algorithm in the same context, we need to construct a primal-dual optimization problem $(\mathrm{PD})_\mu$ such that its subproblem $(\mathrm{PD})_s$ produces the desired directions. Note that the barrier function method requires the parameter μ to decrease to zero. Therefore, without loss of generality, we may assume that $\mu < 1$ in this subsection. In this case, if $\mathbf{x}$ is a feasible solution to problem P_μ and $(\mathbf{w}, \mathbf{s})$ is a feasible solution to problem D_μ, then

$$\mathbf{c}^T\mathbf{x} - \mu\sum_{j=1}^{n}\log_e x_j - \left(\mathbf{b}^T\mathbf{w} + \mu\sum_{j=1}^{n}\log_e s_j\right) = (\mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{w})$$

$$- \mu\left(\sum_{j=1}^{n}\log_e x_j + \sum_{j=1}^{n}\log_e s_j\right)$$

$$= \mathbf{x}^T\mathbf{s} - \mu\sum_{j=1}^{n}\log_e(x_j s_j)$$

$$\geq \mu\sum_{j=1}^{n}[x_j s_j - \log_e(x_j s_j)] \geq 0$$

The desired primal-dual optimization problem can be defined as a problem which minimizes the gap between problems P_μ and D_μ subject to the primal and dual interior feasibility conditions, i.e., problem $(\mathrm{PD})_\mu$ has the following form:

$$\text{Minimize} \quad \mathbf{c}^T\mathbf{x} - \mu\sum_{j=1}^{n}\log_e x_j - \left(\mathbf{b}^T\mathbf{w} + \mu\sum_{j=1}^{n}\log_e s_j\right) \tag{8.23a}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x} > \mathbf{0} \tag{8.23b}$$

$$\mathbf{A}^T\mathbf{w} + \mathbf{s} = \mathbf{c}, \quad \mathbf{s} > \mathbf{0} \tag{8.23c}$$

If we define

$$\overline{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^T & \mathbf{I}_n \end{bmatrix}$$

and use $\mathbf{p}(\mathbf{x})$ and $\mathbf{q}(\mathbf{w}, \mathbf{s})$ representing the objective function of P_μ and D_μ respectively, then problem $(\mathrm{PD})_\mu$ is simplified as follows:

$$\text{Minimize} \quad \mathbf{r}(\mathbf{x}; \mathbf{w}; \mathbf{s}) \equiv \mathbf{p}(\mathbf{x}) - \mathbf{q}(\mathbf{w}, \mathbf{s})$$

$$\text{subject to} \quad \overline{\mathbf{A}}\begin{pmatrix} \mathbf{x} \\ \mathbf{w} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{c} \end{pmatrix}$$

$$\mathbf{x} > \mathbf{0}, \quad \mathbf{s} > \mathbf{0}$$

Suppose that $(\mathbf{x}^k; \mathbf{w}^k; \mathbf{s}^k)$ is a feasible solution of $(\mathrm{PD})_\mu$. The steepest descent direction suggests us to consider the following subproblem $(\mathrm{PD})_s$:

$$\text{Minimize} \quad [\nabla\mathbf{r}(\mathbf{x}^k; \mathbf{w}^k; \mathbf{s}^k)]^T\begin{pmatrix} \mathbf{x} - \mathbf{x}^k \\ \mathbf{w} - \mathbf{w}^k \\ \mathbf{s} - \mathbf{s}^k \end{pmatrix}$$

$$\text{subject to} \quad \overline{\mathbf{A}}\begin{pmatrix} \mathbf{x} - \mathbf{x}^k \\ \mathbf{w} - \mathbf{w}^k \\ \mathbf{s} - \mathbf{s}^k \end{pmatrix} = \mathbf{0}$$

$$||\mathbf{Q}_1^{-1}(\mathbf{x} - \mathbf{x}^k)|| \leq \beta_1^2$$

$$||\mathbf{Q}_2^{-1}(\mathbf{s} - \mathbf{s}^k)|| \leq \beta_2^2$$

where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are symmetric positive definite matrices and $\beta_1, \beta_2 < 1$ are well-chosen scalars such that the corresponding ellipsoids are inscribed to the feasible domain of $(\mathrm{PD})_\mu$. In particular, we can choose $\mathbf{Q}_1 = \mathbf{X}_k^{1/2}\mathbf{S}_k^{-1/2}$ for the primal variables and $\mathbf{Q}_2 = \mathbf{X}_k^{-1/2}\mathbf{S}_k^{1/2}$ for the dual slacks and consider a null-space version of problem $(\mathrm{PD})_s$.

To do so, let us start with a matrix $\mathbf{U}$ satisfying $\mathbf{A}\mathbf{U} = \mathbf{0}$. By defining

$$\overline{\mathbf{U}} = \begin{pmatrix} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \\ \mathbf{0} & -\mathbf{A}^T \end{pmatrix}$$

we see $\overline{\mathbf{A}\mathbf{U}} = \mathbf{0}$ and $\overline{\mathbf{U}}$ serves as an isomorphism between R^n and the null space of $\overline{\mathbf{A}}$. More explicitly, we have

$$\begin{pmatrix} \mathbf{x} - \mathbf{x}^k \\ \mathbf{w} - \mathbf{w}^k \\ \mathbf{s} - \mathbf{s}^k \end{pmatrix} = \overline{\mathbf{U}}\begin{pmatrix} \mathbf{u}^1 \\ \mathbf{u}^2 \end{pmatrix}, \quad \text{where } \mathbf{u}^1 \in R^{n-m}, \quad \mathbf{u}^2 \in R^m$$

Therefore, $\Delta\mathbf{x}_k = \mathbf{U}\mathbf{u}^1$, $\Delta\mathbf{w}_k = \mathbf{u}^2$, and $\Delta\mathbf{s}_k = -\mathbf{A}^T\mathbf{u}^2$. Consequently, problem $(\mathrm{PD})_s$ becomes a null-space version:

$$\text{Minimize} \quad [\nabla\mathbf{r}(\mathbf{x}^k; \mathbf{w}^k; \mathbf{s}^k)]^T\begin{pmatrix} \mathbf{U}\mathbf{u}^1 \\ \mathbf{u}^2 \\ -\mathbf{A}^T\mathbf{u}^2 \end{pmatrix}$$

$$\text{subject to} \quad \left\|\mathbf{X}_k^{-1/2}\mathbf{S}_k^{1/2}\mathbf{U}\mathbf{u}^1\right\| \leq \beta_1^2$$

$$\left\|-\mathbf{X}_k^{1/2}\mathbf{S}_k^{-1/2}\mathbf{A}^T\mathbf{u}^2\right\| \leq \beta_2^2$$

$$\mathbf{u}^1 \in R^{n-m}, \quad \mathbf{u}^2 \in R^m$$

The Lagrangian of the above problem is given by

$$L_3(\mathbf{u}^1, \mathbf{u}^2, \lambda_1, \lambda_2) = \left(\frac{\partial \mathbf{r}(\mathbf{x}^k, \mathbf{w}^k, \mathbf{s}^k)}{\partial \mathbf{x}}\right)^T \mathbf{U}\mathbf{u}^1 + \left(\frac{\partial \mathbf{r}(\mathbf{x}^k, \mathbf{w}^k, \mathbf{s}^k)}{\partial \mathbf{w}}\right)^T \mathbf{u}^2$$
$$- \left(\frac{\partial \mathbf{r}(\mathbf{x}^k, \mathbf{w}^k, \mathbf{s}^k)}{\partial \mathbf{s}}\right)^T \mathbf{A}^T\mathbf{u}^2 + \lambda_1 \left(\left\|\mathbf{X}_k^{-1/2}\mathbf{S}_k^{1/2}\mathbf{U}\mathbf{u}^1\right\|^2 - \beta_1^2\right)$$
$$- \lambda_2 \left(\left\|-\mathbf{X}_k^{1/2}\mathbf{S}_k^{-1/2}\mathbf{A}^T\mathbf{u}^2\right\|^2 - \beta_2^2\right)$$

where λ_1 and λ_2 are nonnegative Lagrange multipliers.

Recall that $\mathbf{r}(\mathbf{x}; \mathbf{w}; \mathbf{s}) = \mathbf{p}(\mathbf{x}) - \mathbf{q}(\mathbf{w}, \mathbf{s})$. We now solve the subproblem $(\mathrm{PD})_s$ by setting that

$$\frac{\partial L_3}{\partial \mathbf{u}^1} = 0 \quad \text{and} \quad \frac{\partial L_3}{\partial \mathbf{u}^2} = 0$$

In this way, we have

$$\mathbf{u}^1 = \frac{-1}{2\lambda_1}\left(\mathbf{U}^T\mathbf{X}_k^{-1}\mathbf{S}_k\mathbf{U}\right)^{-1}\mathbf{U}^T\nabla\mathbf{p}(\mathbf{x}^k) \tag{8.24a}$$

and

$$\begin{aligned}\mathbf{u}^2 = &\frac{-1}{2\lambda_2}\left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\left(-\frac{\partial \mathbf{q}(\mathbf{w}^k, \mathbf{s}^k)}{\partial \mathbf{w}}\right) \\ &+ \frac{1}{2\lambda_2}\left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\left(-\frac{\partial \mathbf{q}(\mathbf{w}^k, \mathbf{s}^k)}{\partial \mathbf{s}}\right)\end{aligned} \tag{8.24b}$$

Transforming back to the original space, we see that

$$\Delta\mathbf{x}_k = \mathbf{U}\mathbf{u}^1 = -\frac{1}{2\lambda_1}\mathbf{U}\left(\mathbf{U}^T\mathbf{X}_k^{-1}\mathbf{S}_k\mathbf{U}\right)^{-1}\mathbf{U}^T\left(\mathbf{c} - \mu\mathbf{X}_k^{-1}\mathbf{e}\right)$$

Applying Lemma 8.2 results in

$$\begin{aligned}\Delta\mathbf{x}_k &= -\frac{1}{2\lambda_1}\mathbf{X}_k^{1/2}\mathbf{S}_k^{-1/2}\left[I - \mathbf{X}_k^{1/2}\mathbf{S}_k^{-1/2}\mathbf{A}^T\left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k^{1/2}\mathbf{S}_k^{-1/2}\right] \\ &\quad \mathbf{X}_k^{1/2}\mathbf{S}_k^{-1/2}\left(\mathbf{c} - \mu\mathbf{X}_k^{-1}\mathbf{e}\right) \\ &= -\frac{1}{2\lambda_1}\left(\mathbf{X}_k\mathbf{S}_k^{-1} - \mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\right)\left(\mathbf{c} - \mathbf{A}^T\mathbf{w}^k - \mu\mathbf{X}_k^{-1}\mathbf{e}\right) \\ &= -\frac{1}{2\lambda_1}\left(\mathbf{S}_k^{-1} - \mathbf{S}_k^{-1}\mathbf{X}_k\mathbf{A}^T\left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{S}_k^{-1}\right)\left(\mathbf{X}_k\mathbf{S}_k\mathbf{e} - \mu\mathbf{e}\right)\end{aligned} \tag{8.25a}$$

Similarly, we have

$$\begin{aligned}\Delta\mathbf{w}_k = \mathbf{u}^2 &= \frac{1}{2\lambda_2}\left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\left(\mathbf{b} - \mu\mathbf{A}\mathbf{S}_k^{-1}\mathbf{e}\right) \\ &= \frac{1}{2\lambda_2}\left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\left(\mathbf{A}\mathbf{X}_k\mathbf{e} - \mu\mathbf{A}\mathbf{S}_k^{-1}\mathbf{e}\right) \\ &= \frac{1}{2\lambda_2}\left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{S}_k^{-1}\left(\mathbf{X}_k\mathbf{S}_k\mathbf{e} - \mu\mathbf{e}\right)\end{aligned} \tag{8.25b}$$

Moreover,

$$\Delta\mathbf{s} = -\mathbf{A}^T\mathbf{u}^2 = -\frac{1}{2\lambda_2}\mathbf{A}^T\left(\mathbf{A}\mathbf{X}_k\mathbf{S}_k^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{S}_k^{-1}\left(\mathbf{X}_k\mathbf{S}_k\mathbf{e} - \mu\mathbf{e}\right) \tag{8.25c}$$

Noting that both λ_1 and λ_2 are nonnegative, and, comparing (8.25) to (7.90), we can confirm that the moving directions of the primal-dual affine scaling algorithm are given by the solution of the subproblem $(\mathrm{PD})_s$.

8.5 GENERAL THEORY

The geometric interpretation of the moving directions of the affine scaling algorithms suggests that we study two crucial factors. First, we need a symmetric positive definite scaling matrix to open an appropriate ellipsoid in the null space of the constraint matrix for consideration. Second, we need an appropriate objective function such that its first-order approximation is optimized. In the previous section, we have incorporated logarithmic barrier functions into the original objective and applied diagonal matrices for scaling. Here we want to further extend this approach to study more general results.

8.5.1 General Primal Affine Scaling

In this subsection, we focus on the primal program P defined by (8.1). Instead of choosing the logarithmic barrier function, for $\mu > 0$, let us use a general concave *barrier function* $\phi(\mathbf{x})$ which is well defined and differentiable on the relative interior of the primal feasible domain and consider the following problem $(\mathrm{P}\phi)_\mu$:

$$\text{Minimize} \quad \mathbf{c}^T\mathbf{x} - \mu\phi(\mathbf{x}) \tag{8.26a}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b} \tag{8.26b}$$

$$\mathbf{x} > \mathbf{0} \tag{8.26c}$$

Under the interior-point assumption (A1) on problem (P), let $\mathbf{x}^k$ be a feasible solution to problem $(\mathrm{P}\phi)_\mu$ and $\nabla\phi$ be the gradient of ϕ. We also let $\mathbf{Q}^{-1}$ be an arbitrary symmetric positive definite matrix, $\beta < 1$ be a positive scalar such that the ellipsoid $\{\mathbf{x} \in R^n \mid \|\mathbf{Q}^{-1}(\mathbf{x} - \mathbf{x}^k)\|^2 \le \beta^2\}$ becomes inscribed in the feasible domain of problem $(\mathrm{P}\phi)_\mu$. Our focus is to find a good moving direction vector $\Delta\mathbf{x}_k = \mathbf{x} - \mathbf{x}^k$ from the ellipsoid such that $\mathbf{x}$ is still feasible, i.e., $\mathbf{A}\Delta\mathbf{x}_k = \mathbf{0}$, and the objective value $\mathbf{c}^T\mathbf{x} - \mu\phi(\mathbf{x})$ is minimized.

Taking the first-order approximation of the objective function at the current interior solution $\mathbf{x}^k$, we have

$$\mathbf{c}^T\mathbf{x} - \mu\phi(\mathbf{x}) \approx \mathbf{c}^T\mathbf{x}^k - \mu\phi\left(\mathbf{x}^k\right) + \left[\mathbf{c} - \mu\nabla\phi\left(\mathbf{x}^k\right)\right]^T \Delta\mathbf{x}_k$$

Therefore, we focus on the following subproblem $(\mathrm{P}\phi)_s$:

$$\text{Minimize} \quad \left[\mathbf{c} - \mu\nabla\phi\left(\mathbf{x}^k\right)\right]^T \Delta\mathbf{x}_k \tag{8.27a}$$

$$\text{subject to} \quad \mathbf{A}\Delta\mathbf{x}_k = \mathbf{0} \tag{8.27b}$$

$$\|\mathbf{Q}^{-1}\Delta\mathbf{x}_k\|^2 \le \beta^2 \tag{8.27c}$$

In order to solve (8.27), we make use of the isomorphism $\mathbf{U}$ between the R^{n-m} and the null space of matrix $\mathbf{A}$ such that $\Delta\mathbf{x}_k$ is replaced by $\mathbf{h} \in R^{n-m}$ to eliminate the constraint $\mathbf{A}\Delta\mathbf{x}_k = \mathbf{0}$ in a null-space version problem. In this way, we have an equivalent problem:

$$\text{Minimize} \quad \left[\mathbf{c} - \mu\nabla\phi\left(\mathbf{x}^k\right)\right]^T \mathbf{U}\mathbf{h} \tag{8.28a}$$

$$\text{subject to} \quad \left\|\mathbf{Q}^{-1}\mathbf{U}\mathbf{h}\right\|^2 \le \beta^2 \tag{8.28b}$$

$$\mathbf{h} \in R^{n-m} \tag{8.28c}$$

The Lagrangian of problem (8.28) is given by

$$L(\mathbf{h}, \lambda) = \left(\mathbf{c} - \mu\nabla\phi\left(\mathbf{x}^k\right)\right)^T \mathbf{U}\mathbf{h} + \lambda\left(\left\|\mathbf{Q}^{-1}\mathbf{U}\mathbf{h}\right\|^2 - \beta^2\right)$$

where $\lambda \ge 0$ is the Lagrange multiplier associated with the inequality constraint. Setting $\partial L/\partial\mathbf{h} = 0$ and solving for $\mathbf{h}$ results in a solution

$$\mathbf{h}^k = -\frac{1}{2\lambda}\left(\mathbf{U}^T\mathbf{Q}^{-2}\mathbf{U}\right)^{-1}\mathbf{U}^T\left(\mathbf{c} - \mu\nabla\phi\left(\mathbf{x}^k\right)\right)$$

Consequently, from Lemma 8.2, a moving direction

$$\Delta\mathbf{x}_k = -\mathbf{Q}[\mathbf{I} - \mathbf{Q}\mathbf{A}^T\left(\mathbf{A}\mathbf{Q}^2\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{Q}]\mathbf{Q}\left(\mathbf{c} - \mu\nabla\phi\left(\mathbf{x}^k\right)\right) \tag{8.29}$$

is generated for the general primal affine scaling algorithm. Also note that, when $\phi(\mathbf{x})$ is strictly concave and twice differentiable, the Hessian matrix $\mathbf{H}$ of $-\phi(\mathbf{x})$ becomes symmetric positive definite. Actually, $\mathbf{H}$ is the Hessian of the objective function $\mathbf{c}^T\mathbf{x} - \mu\phi(\mathbf{x})$ of program $(\mathrm{P}\phi)_\mu$. If we choose $\mathbf{H}^{1/2}$ to be the scaling matrix $\mathbf{Q}^{-1}$ (or equivalently, $\mathbf{H} = \mathbf{Q}^{-2}$), then

$$\Delta\mathbf{x}_k = -\mathbf{H}^{-1/2}\left[\mathbf{I} - \mathbf{H}^{-1/2}\mathbf{A}^T\left(\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{H}^{-1/2}\right]\mathbf{H}^{-1/2}\left(\mathbf{c} - \mu\nabla\phi\left(\mathbf{x}^k\right)\right) \tag{8.30}$$

is the *projected Newton direction* with respect to the general barrier function.

Note that the classic inverse function can be used as a barrier function, i.e.,

$$\phi(\mathbf{x}) = -\frac{1}{r}\sum_{j=1}^{n}\frac{1}{x_j^r} \qquad \text{for } r > 0$$

In this case,

$$\nabla\phi(\mathbf{x}) = \mathbf{X}^{-r-1}\mathbf{e} \tag{8.31a}$$

and

$$\mathbf{H} = -(r+1)\mathbf{X}^{-r-2} \tag{8.31b}$$

The Karush-Kuhn-Tucker conditions becomes

$$\mathbf{A}^T\mathbf{w} + \mathbf{s} = \mathbf{c}, \qquad \mathbf{s} > \mathbf{0} \tag{8.32a}$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \qquad \mathbf{x} > \mathbf{0} \tag{8.32b}$$

$$\mathbf{X}^{r+1}\mathbf{S}\mathbf{e} = \mu\mathbf{e} \tag{8.32c}$$

Comparing (8.32) with (8.5), as $r \to 0$, we see the two systems are closely related. Plugging (8.31a) and (8.31b) into formula (8.30), den Hertog, C. Roos and T. Terlaky designed their inverse barrier method by moving along the projected Newton direction

$$\Delta\mathbf{x}_k = -\mathbf{H}^{-1/2}\left[\mathbf{I} - \mathbf{H}^{-1/2}\mathbf{A}^T\left(\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{H}^{-1/2}\right]\mathbf{H}^{-1/2}\left(\mathbf{c} - \mu\mathbf{X}^{-r-1}\mathbf{e}\right)$$

with a proper step-length such that the algorithm terminates after at most

$$O\left(\sqrt{n}\left(\frac{n}{\epsilon}\right)^{r/2}\log_e\left(\frac{n}{\epsilon}\right)\right)$$

iterations to reach an ϵ-optimal solution, under the assumptions of having an interior feasible solution and bounded primal feasible domain. As $r \to 0$, the inverse barrier function algorithm approaches the logarithmic barrier function algorithm with the same complexity bound. Moreover, R. Sheu and S-C. Fang used the general direction (8.29) to construct a generic path-following algorithm for linear programming and imposed some sufficient conditions on a general barrier function to achieve polynomial-time performance.

8.5.2 General Dual Affine Scaling

In this subsection, we shift our focus to the dual program D defined by (8.2). As in the general primal affine scaling, we replace the logarithmic barrier function, for $\mu > 0$, by a general concave barrier function $\psi(\mathbf{x})$ which is well defined and differentiable on the relative interior of the dual feasible domain. Now consider the following problem $(\mathrm{D}\psi)_\mu$:

$$\text{Maximize} \quad \mathbf{b}^T\mathbf{w} + \mu\psi(\mathbf{s}) \tag{8.33a}$$

$$\text{subject to} \quad \mathbf{A}^T\mathbf{w} + \mathbf{s} = \mathbf{c} \tag{8.33b}$$

$$\mathbf{s} > \mathbf{0} \tag{8.33c}$$

Under the interior-point assumption (A2) on problem (D), let $(\mathbf{w}^k; \mathbf{s}^k)$ be a feasible solution to problem $(\mathrm{D}\psi)_\mu$ and $\nabla\psi$ be the gradient of ψ. Again we let $\mathbf{Q}^{-1}$ be an arbitrary symmetric positive definite matrix, $\beta < 1$ be a positive scalar such that the ellipsoid $\{\mathbf{s} \in R^n \mid \|\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{s}^k)\|^2 \le \beta^2\}$ becomes inscribed in the feasible domain of problem $(\mathrm{D}\psi)_\mu$. In order to find good moving direction vectors $\Delta\mathbf{w}_k = \mathbf{w} - \mathbf{w}^k$ and $\Delta\mathbf{s}_k = \mathbf{s} - \mathbf{s}^k$, we focus on the following subproblem $(\mathrm{D}\psi)_s$:

$$\text{Maximize} \quad \left[\mathbf{b}^T \mid \left(\mu\nabla\psi\left(\mathbf{s}^k\right)\right)^T\right]\begin{pmatrix}\Delta\mathbf{w}_k \\ \Delta\mathbf{s}_k\end{pmatrix} \tag{8.34a}$$

$$\text{subject to} \quad \mathbf{A}^T\Delta\mathbf{w}_k + \Delta\mathbf{s}_k = \mathbf{0} \tag{8.34b}$$

$$\left\|\mathbf{Q}^{-1}\Delta\mathbf{s}_k\right\|^2 \leq \beta^2 \tag{8.34c}$$

Remember the isomorphism

$$\hat{\mathbf{U}} = \begin{bmatrix} \mathbf{I}_m \\ -\mathbf{A}^T \end{bmatrix}$$

between the null space of matrix $\hat{\mathbf{A}}^T = [\mathbf{A}^T \mid \mathbf{I}_n]$ and R^m such that $\Delta\mathbf{w}_k = \mathbf{v}$ and $\Delta\mathbf{s}_k = -\mathbf{A}^T\mathbf{v}$, for $\mathbf{v} \in R^m$. A null-space version of problem $(\mathrm{D}\psi)_s$ becomes

$$\text{Maximize} \quad \left[\mathbf{b}^T \mid (\mu\nabla\psi(\mathbf{s}^k))^T\right]\hat{\mathbf{U}}\mathbf{v} \tag{8.35a}$$

$$\text{subject to} \quad \|-\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{v}\|^2 \leq \beta^2 \tag{8.35b}$$

$$\mathbf{v} \in R^m \tag{8.35c}$$

The Lagrangian of problem (8.35) is given by

$$L(\mathbf{v}, \lambda) = \mathbf{b}^T\mathbf{v} - \left[\mu\nabla\psi(\mathbf{s}^k)\right]^T\mathbf{A}^T\mathbf{v} - \lambda\left(\left\|-\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{v}\right\|^2 - \beta^2\right)$$

where $\lambda \geq 0$ is the Lagrange multiplier associated with the inequality constraint (8.35b). Setting $\partial L/\partial\mathbf{v} = 0$ and solving for $\mathbf{v}$, we have

$$\mathbf{v}^k = \frac{1}{2\lambda}\left[\left(\mathbf{A}\mathbf{Q}^{-2}\mathbf{A}^T\right)^{-1}\mathbf{b} - \left(\mathbf{A}\mathbf{Q}^{-2}\mathbf{A}^T\right)^{-1}\mathbf{A}\left(\mu\nabla\psi(\mathbf{s}^k)\right)\right]$$

Consequently, we have

$$\Delta\mathbf{w}_k = \mathbf{v}^k = \frac{1}{2\lambda}\left[\left(\mathbf{A}\mathbf{Q}^{-2}\mathbf{A}^T\right)^{-1}\mathbf{b} - \left(\mathbf{A}\mathbf{Q}^{-2}\mathbf{A}^T\right)^{-1}\mathbf{A}\left(\mu\nabla\psi(\mathbf{s}^k)\right)\right] \tag{8.36a}$$

and

$$\Delta\mathbf{s}_k = -\mathbf{A}^T\Delta\mathbf{w}_k \tag{8.36b}$$

for the general dual affine scaling algorithm. Also note that, when $\psi(\mathbf{x})$ is strictly concave and twice differentiable, the Hessian matrix $\mathbf{H}$ of $\psi(\mathbf{x})$ becomes symmetric positive definite. If we choose $\mathbf{H}^{1/2}$ to be the scaling matrix $\mathbf{Q}^{-1}$ (or equivalently, $\mathbf{H} = \mathbf{Q}^{-2}$), then the corresponding formulas for $\Delta\mathbf{w}_k$ and $\Delta\mathbf{s}_k$ can be derived. When the classic inverse function is taken to be the barrier function for the dual approach, i.e.,

$$\psi(\mathbf{s}) = \frac{1}{r}\sum_{j=1}^{n}\frac{1}{s_j^r} \quad \text{for } r > 0$$

a corresponding dual algorithm can be further developed.

8.6 CONCLUDING REMARKS

In this chapter we have provided an algebraic view as well as a geometric interpretation to gain more insights into the primal affine scaling, dual affine scaling, and primal-dual algorithms. From the algebraic point of view, at least in theory, we may have infinitely

many algebraic paths that lead to the solutions of the Karush-Kuhn-Tucker conditions. Moving along the Newton direction of each such path with appropriate step-lengths may result in a new algorithm for further analysis.

The geometric interpretation relies on the special structure of a corresponding subproblem. Basically, it takes an appropriate scaling matrix and a scalar to open an inscribed ellipsoid in the feasible domain such that the inequality constraints can be replaced. Then we consider the projected (negative) gradient of the objective function in the null space of the constraint matrix as a potential moving direction. The shape of the inscribed ellipsoid is certainly determined by the scaling matrix, and the projected gradient is dependent on the barrier function applied.

Based on the geometric view, a general scheme which generates the moving directions of the generalized primal affine scaling and dual affine scaling has been included. As to the generalization of the primal-dual algorithm, the difficulty lies in finding a pair of primal barrier function $\phi(\mathbf{x})$ and dual barrier function $\psi(\mathbf{s})$ such that both programs $(\mathrm{P}\phi)_\mu$ and $(\mathrm{D}\psi)_\mu$ have a common system of Karush-Kuhn-Tucker conditions. If this can be done, the generalization follows immediately. But so far, except by using the logarithmic barrier function for both the primal and dual, no other successful case has been reported.

REFERENCES FOR FURTHER READING

8.1 Gonzaga, C., "Search directions for interior linear programming methods," *Algorithmica* 6, 153–181 (1991).

8.2 den Hertog, D., Roos, C., and Terlaky, T., "Inverse barrier methods for linear programming," Report of the Faculty of Technical Mathematics and Informatics, No. 90-27, Delft University of Technology (1990).

8.3 Sheu, R. L., and Fang, S. C., "Insights into the interior-point methods," OR Report No. 252, North Carolina State University, Raleigh, NC (1990), *Zeischrift für Operations Research* 36, 200–230 (1992).

8.4 Sheu, R. L., and Fang, S. C., "On the generalized path-following methods for linear programming," OR Report No. 261, North Carolina State University, Raleigh, NC (1992).

8.5 Ye, Y., "An extension of Karmarkar's algorithm and the trust region method for quadratic programming," in *Progress in Mathematical Programming: Interior-Point and Related Methods*, ed. N. Megiddo, Springer Verlag, New York, 49–64 (1989).

8.6 Zimmermann, U., "Search directions for a class of projective methods," *Zeischrift für Operations Research* 34, 353–379 (1990).

EXERCISES

8.1. Show that (8.18) is indeed a solution to the system (8.17) together with $\mathbf{A}\Delta\mathbf{x}_k = \mathbf{0}$ and $\mathbf{A}^T\Delta\mathbf{w}_k + \Delta\mathbf{s}_k = \mathbf{0}$.

8.2. If we define $\mathbf{P} = \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T = [\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}]$, show that $\mathbf{P}^2 = \mathbf{P}$ and $\mathbf{AP} = \mathbf{0}$.

8.3. Show that $\mathbf{v}^k$ in (8.22a) is indeed an optimal solution to the null-space version of program D_s.

8.4. From (8.19), we know that

$$\Delta \mathbf{x}_k = \frac{1}{2\lambda}\mathbf{U}\left(\mathbf{U}^T\mathbf{X}_k^{-2}\mathbf{U}\right)^{-1}\mathbf{U}^T\left(-\nabla \mathbf{p}\left(\mathbf{x}^k\right)\right)$$
$$= \frac{1}{2\lambda}\mathbf{U}\left(\mathbf{U}^T\mathbf{X}_k^{-2}\mathbf{U}\right)^{-1}\mathbf{U}^T\left(-\mathbf{c}+\mu\mathbf{X}_k^{-1}\mathbf{e}\right)$$

Show that $\mathbf{U}(\mathbf{U}^T\mathbf{X}_k^{-2}\mathbf{U})^{-1}\mathbf{U}^T$ is a projection mapping. Hence the moving direction in the primal affine scaling with logarithmic barrier function can be viewed as the negative gradient of $\mathbf{p}(\mathbf{x})$ projected into the null space of matrix $\mathbf{A}$.

8.5. From (8.22), first try to show that

$$\begin{pmatrix}\Delta \mathbf{w}_k\\ \Delta \mathbf{s}_k\end{pmatrix} = \frac{1}{2\lambda}\hat{\mathbf{U}}\left(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T\right)^{-1}\hat{\mathbf{U}}^T\left(\nabla \mathbf{q}\left(\mathbf{w}^k,\mathbf{s}^k\right)\right)$$
$$= \frac{1}{2\lambda}\hat{\mathbf{U}}(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T)^{-1}\hat{\mathbf{U}}^T\begin{pmatrix}\mathbf{b}\\ \mu\mathbf{S}_k^{-1}\mathbf{e}\end{pmatrix}$$

Then show that $\hat{\mathbf{U}}(\mathbf{A}\mathbf{S}_k^{-2}\mathbf{A}^T)^{-1}\hat{\mathbf{U}}^T$ is not a projection mapping. Hence the moving direction in the dual affine scaling with logarithmic barrier function cannot be viewed as the negative gradient of $\mathbf{q}(\mathbf{w};\mathbf{s})$ projected into the null space. The reason is mostly due to the unrestricted variables $\mathbf{w}$. This phenomenon will not happen for the symmetric dual problem, which requires both $\mathbf{w}$ and $\mathbf{s}$ to be nonnegative.

8.6. Derive (8.24a) and (8.24b) from

$$\frac{\partial L_3}{\partial \mathbf{u}^1} = 0 \quad \text{and} \quad \frac{\partial L_3}{\partial \mathbf{u}^2} = 0$$

8.7. In order to derive a geometric interpretation of the moving directions (8.18) associated with the algebraic path $r(x_j, s_j) = \log_e\left(x_j s_j/\mu\right) = 0$, we define

$$t(\mathbf{x};\mathbf{w};\mathbf{s}) = -\mathbf{x}^T\mathbf{s} + \sum_{j=1}^{n} x_j s_j \log_e\left(\frac{x_j s_j}{\mu}\right)$$

Now consider the following subproblem:

$$\left\{\text{minimize } \left[\nabla t\left(\mathbf{x}^k;\mathbf{w}^k;\mathbf{s}^k\right)\right]^T\begin{pmatrix}\mathbf{x}-\mathbf{x}^k\\ \mathbf{w}-\mathbf{w}^k\\ \mathbf{s}-\mathbf{s}^k\end{pmatrix} \;\middle|\; \mathbf{A}\Delta\mathbf{x}_k = \mathbf{0},\ \mathbf{A}^T\Delta\mathbf{w}_k + \Delta\mathbf{s}_k = \mathbf{0}\right\}$$

By choosing $\mathbf{X}_k^{-1/2}\mathbf{S}_k^{1/2}$ and $\mathbf{X}_k^{1/2}\mathbf{S}_k^{-1/2}$ as the scaling matrix for $\mathbf{x}$ and $\mathbf{s}$, respectively, show that the solution of this subproblem provides the moving directions (8.18).

8.8. Replace the objective function of (8.3a) by

$$\frac{\mathbf{c}^T\mathbf{x}}{\mu} + \frac{1}{r}\sum_{j=1}^{n}\frac{1}{x_j^r}$$

and verify that its Karush-Kuhn-Tucker conditions are given by (8.32).

8.9. Taking the inverse barrier function

$$\psi(\mathbf{s}) = \frac{1}{r}\sum_{j=1}^{n}\frac{1}{s_j^r} \qquad \text{for } r > 0$$

and using $\mathbf{H}^{1/2}$ as the scaling matrix $\mathbf{Q}^{-1}$, derive corresponding dual moving directions $\Delta\mathbf{w}_k$ and $\Delta\mathbf{s}_k$.

8.10. Consider using the entropy function

$$\phi(\mathbf{x}) = -\sum_{j=1}^{n} x_j \log_e x_j$$

as the barrier function.

(a) By plugging the entropic barrier function into the objective function of (8.26a), derive the corresponding Karush-Kuhn-Tucker conditions.

(b) Find the gradient and Hessian of the entropic barrier function at a given solution $\mathbf{x}^k$.

(c) Take $\mathbf{Q}^{-1} = \mathbf{H}^{1/2}$ as the scaling matrix to derive the corresponding formula of the primal moving direction $\Delta\mathbf{x}_k$.

8.11. Take

$$\psi(\mathbf{s}) = -\sum_{j=1}^{n} s_j \log_e s_j$$

as the barrier function.

(a) Derive the K-K-T conditions for problem (8.33).

(b) Compare your results with 8.10(a) to see if they represent the same system.