Chapter 1 Linear Programming

1.1 Transportation of Commodities

We consider a market consisting of a certain number of providers and demanders of a commodity and a network of routes between the providers and the demanders along which the commodity can be shipped from the providers to the demanders. In particular, we assume that the transportation network is given by a set $\mathcal{A}$ of arcs, where $(i,j) \in \mathcal{A}$ means that there exists a route connecting the provider i and the demander j.
We denote by c_{ij} the unit shipment cost on the arc (i,j), by s_i the available supply at the provider i , and by d_j the demand at the demander j. The variables are the quantities x_{ij} of the commodity that is shipped over the arc $(i,j) \in \mathcal{A}$ and the problem is to minimize the transportation costs

$$\text{minimize} \sum_{(i,j)\in\mathcal{A}} c_{ij}\, x_{ij} \quad \text{over all } x = (x_{ij}) \geq 0 \tag{1.1}$$

under the natural constraints:

- the supply s_i at i should exceed the sum of the demands at all j such that $(i,j) \in \mathcal{A}$, i.e.

$$\sum_{j:(i,j)\in\mathcal{A}} x_{ij} \;\leq\; s_i \quad \text{for all } i\ , \tag{1.2}$$

- the demand d_j at j must be satisfied in the sense that is less or equal the sum of the supplies at all i such that $(i,j) \in \mathcal{A}$, i.e.

$$\sum_{i:(i,j)\in\mathcal{A}} x_{ij} \;\geq\; d_j \quad \text{for all } j\ . \tag{1.3}$$

The problem (1.1)-(1.3) is a Linear Program (LP) whose solution by the simplex method and primal-dual interior-point methods will be considered in sections 1.2 and 1.3 below.

1.1.1 Dantzig's original transportation model

As an example we consider *G.B. Dantzig's* original transportation model: We assume two providers $i = 1$ and $i = 2$ of tin cans located at Seattle and San Diego and three demanders $j = 1, j = 2$, and $j = 3$ located at New York, Chicago, and Topeka, respectively:

Sets

i	**canning plants:**	**Seattle , San Diego**
j	**markets:**	**New York , Chicago , Topeka**

The following table contains the distances d_{ij} in thousands of miles between the providers and the demand centers

	New York	**Chicago**	**Topeka**
Seattle	**2.5**	**1.7**	**1.8**
San Diego	**2.5**	**1.8**	**1.4**

The freight F in dollars per case per thousand miles is $F = 90$, so that the transport cost c_{ij} in thousands of dollars per case is given by

$$c_{ij} \;=\; \frac{F\; d_{ij}}{1000}\;.$$

The supplies $s_i, 1 \leq i \leq 2$, and the demands $d_j, 1 \leq j \leq 3$, are given as follows

Supplies

$\mathbf{s_1}$	**Seattle**	**325**
$\mathbf{s_2}$	**San Diego**	**575**

Demands

$\mathbf{d_1}$	**New York**	**325**
$\mathbf{d_2}$	**Chicago**	**300**
$\mathbf{d_3}$	**Topeka**	**275**

1.1.2 The primal-dual and the dual problem

For the inequality constraints (1.2) and (1.3) we introduce Lagrange multipliers $p_i^s \geq 0$ and $p_j^d \geq 0$ which are also called dual variables or shadow prices. Then, the LP (1.1)-(1.3) can be restated as the saddle point problem

$$\min_{x \geq 0} \max_{p^s, p^d \geq 0} L(x, p^s, p^d) \tag{1.4}$$

where the Lagrangian L is given by

$$\begin{aligned} L(x, p^s, p^d) \;:=\; & \sum_{(i,j)\in\mathcal{A}} c_{ij}\; x_{ij} \;+ \\ & + \sum_{i:(i,j)\in\mathcal{A}} p_i^s \;\Big(\sum_{j:(i,j)\in\mathcal{A}} x_{ij} - s_i\Big) \;+ \\ & + \sum_{j:(i,j)\in\mathcal{A}} p_j^d \;\Big(d_j - \sum_{i:(i,j)\in\mathcal{A}} x_{ij}\Big)\;. \end{aligned} \tag{1.5}$$

The optimality conditions for (1.5) turn out to be

$$x_{ij} \geq 0 \ , \ L_{x_{ij}} \geq 0 \ , \ x_{ij} \cdot L_{x_{ij}} = 0 \quad \text{for all } (i,j) \in \mathcal{A} \ , \tag{1.6}$$

$$p_i^s \geq 0 \ , \ L_{p_i^s} \leq 0 \ , \ p_i^s \cdot L_{p_i^s} = 0 \quad \text{for all } i \ , \tag{1.7}$$

$$p_j^d \geq 0 \ , \ L_{p_j^d} \leq 0 \ , \ p_j^d \cdot L_{p_j^d} = 0 \quad \text{for all } j \ . \tag{1.8}$$

Computing the derivatives of the Lagrangian L, (1.6)-(1.8) represents the Linear Complementarity Problem (LCP)

$$\begin{aligned} & x_{ij} \ \geq \ 0 \ , \ p_i^s + c_{ij} - p_j^d \ \geq \ 0 \ , \\ & x_{ij} \cdot (p_i^s + c_{ij} - p_j^d) \ = \ 0 \quad \text{for all } (i,j) \in \mathcal{A} \ , \end{aligned} \tag{1.9}$$

$$\begin{aligned} & p_i^s \ \geq \ 0 \ , \ s_i - \sum_{j:(i,j)\in\mathcal{A}} x_{ij} \ \geq \ 0 \ , \\ & p_i^s \cdot (s_i - \sum_{j:(i,j)\in\mathcal{A}} x_{ij}) \ = \ 0 \quad \text{for all } i \ , \end{aligned} \tag{1.10}$$

$$\begin{aligned} & p_j^d \geq 0 \ , \ \sum_{i:(i,j)\in\mathcal{A}} x_{ij} - d_j \ \geq \ 0 \ , \\ & p_j^d \cdot (\sum_{i:(i,j)\in\mathcal{A}} x_{ij} - d_j) \ = \ 0 \quad \text{for all } j \ . \end{aligned} \tag{1.11}$$

Although, the complementarity conditions (1.9)-(1.11) can be derived rigorously, let us comment on them from an intuitive point of view:

(i) Complementarity condition (1.9):

The supply price p_i^s at i plus the transportation cost c_{ij} from i to j must exceed the market price p_j^d at d, i.e., $p_i^s + c_{ij} - p_j^d \geq 0$. Otherwise, in a competitive marketplace, another provider will replicate the supplier i increasing the supply of the commodity which drives down the market price. This chain would repeat until the inequality is satisfied. On the other hand, if the cost of delivery c_{ij} strictly exceeds the market price, i.e., $p_i^s + c_{ij} - p_j^d > 0$, then nothing is shipped from i to j because doing so would incur a loss and hence, $x_{ij} = 0$.

(ii) Complementarity condition (1.10):

In case $s_i > \sum_{j:(i,j)\in\mathcal{A}} x_{ij}$ there is an excessive supply at i. In a competitive marketplace, the provider is not willing to pay for more supply, because he is already over-supplied, and hence, $p_i^s = 0$. On the other hand, if $s_i = \sum_{j:(i,j)\in\mathcal{A}} x_{ij}$, the supplier might be willing to pay for additional supply of the commodity, whence $p_i^s \geq 0$.

(iii) Complementarity condition (1.11):
Assume $\sum_{i:(i,j)\in\mathcal{A}} x_{ij} > d_j$ which means that the supply exceeds the demand. Hence, the demander is not willing to pay for more goods, i.e., $p_j^d = 0$. Otherwise, if $\sum_{i:(i,j)\in\mathcal{A}} x_{ij} = d_j$, the demander might consider to order more commodities whence $p_j^d \geq 0$.

The LCP (1.9)-(1.11) represents the complementary slackness conditions of the LP (1.1)-(1.3) which are both the necessary and the sufficient optimality conditions for the LP. Moreover, the conditions (1.9)-(1.11) are also the necessary and sufficient optimality conditions of the problem

$$\max_{p^s,p^d\geq 0} \sum_j d_j\, p_j^d \;-\; \sum_i s_i\, p_i^s \;, \tag{1.12}$$

$$\text{subject to} \quad c_{ij} \;\geq\; p_j^d - p_i^s \quad \text{for all } (i,j)\in\mathcal{A} \;, \tag{1.13}$$

which is called the dual linear program
Since the LCP (1.9)-(1.11) contains both the primal and the dual variables, it is referred to as the primal-dual formulation of the LP.

1.1.3 Model generalization

The primal-dual formulation of the LP has the advantage that extensions of the underlying model can easily be accomplished. One of these extensions is to consider the demand as a function of the prices p, i.e., we replace d with some function $d(p)$. As long as this function is an affine one, the complementarity problem stays linear, as it is the case for $d_j(p) := d_j(1 - p_j^d)$. However, we note that in this case the demand behaves somewhat strangely, if $p_j^d > 1$.
A more realistic scenario occurs by introducing some price elasticity e_j and a reference price $\bar{p}_j^d$ at j and to define the demand according to

$$d_j(p_j^d) \;:=\; \Big(\frac{\bar{p}_j^d}{p_j^d}\Big)^{e_j} \;. \tag{1.14}$$

In this case, the demand depends nonlinearly on the price which means that we are faced with a nonlinear complementarity problem (NCP). Methods to solve nonlinear programming problems will be presented in Chapters 3 and 4.

1.2 The Simplex Method

A linear program is a problem of the following form

(1.15) $$\textit{minimize } c_1x_1 + c_2x_2 + \cdots + c_nx_n \equiv c^Tx$$

over all $x \in \mathbb{R}^n$ which satisfy finitely many linear equality and inequality constraints

(1.16) $$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i \quad , \quad 1 \leq i \leq m_1 \ ,$$

(1.17) $$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i \quad , \quad m_1 + 1 \leq i \leq m \ .$$

Here, c_k, a_{ik}, b_i are given real numbers. The function c^Tx is called the *objective function*, and any $x \in \mathbb{R}^n$ which satisfies (1.16),(1.17) is referred to as a *feasible point*. Introducing additional variables and equations, the linear program (1.15),(1.16),(1.17) can be transformed into a form where the constraints only consist of equations and elementary inequalities of the form $x_i \geq 0$. Moreover, it is useful to cast the objective functional c^Tx in the form $c^Tx \equiv -x_p$. For that purpose, in (1.16) any inequality

$$a_{i1}x_1 + \cdots + a_{in}x_n \leq b_i$$

is replaced with an equation and an elementary inequality by means of the *slack variable* x_{n+i}

$$a_{i1}x_1 + \cdots + a_{in}x_n + x_{n+i} = b_i, \quad x_{n+i} \geq 0 \ .$$

If the objective functional $c_1x_1 + \cdots + c_nx_n$ does not have the required form, we introduce an additional variable x_{n+m_1+1} and an additional equation according to

$$c_1x_1 + \cdots + c_nx_n + x_{n+m_1+1} = 0$$

which has to be added to the other constraints. Then, the minimization of c^Tx is equivalent to the maximization of x_p under the thus extended constraints.

Therefore, without restriction of generality we may assume that the linear program has the following *standard form*:

(1.18) $$LP(I,p): \quad \textit{maximize} \quad x_p$$

(1.19) $$x \in \mathbb{R}^n : \quad Ax = b,$$

(1.20) $$x_i \geq 0 \text{ for } i \in I.$$

Here, $I \subset N := \{1, 2, \ldots, n + m_1 + 1\}$ is an index set (possibly empty), p stands for a fixed index with $p \in N \backslash I$, $A = (a_1|a_2|\ldots|a_{n+m_1+1})$ is a real $(m+1) \times (n+m_1+1)$-matrix with columns a_i and $b \in \mathbb{R}^{m+1}$

is a given vector. The variables $x_i, i \in I$ resp. $x_i, i \notin I$ are called *constrained* resp. *free (unconstrained)* variables.

We denote by

$$P := \{x \in \mathbb{R}^n \mid Ax = b \;\&\; x_i \geq 0 \text{ for all } i \in I\}$$

the *feasible set* of $LP(I,p)$. A vector $\bar{x} \in P$ is said to be the *optimal solution* of $LP(I,p)$, if $\bar{x}_p = \max\{x_p \mid x \in P\}$.

To illustrate the ideas, we use the following example:

$$\begin{array}{ll} \textit{minimize} & -x_1 - 2x_2 \\ x: & -x_1 + x_2 \leq 2 \\ & x_1 + x_2 \leq 4 \\ & x_1 \geq 0,\; x_2 \geq 0. \end{array}$$

After introducing the slack variables x_3, x_4 and the additional variable x_5 for the objective functional, the problem can be stated in standard form $LP(I,p)$ with $p = 5$, $I = \{1,2,3,4\}$:

$$\begin{array}{lll} \textit{maximize} & x_5 & \\ x: & -x_1 + x_2 + x_3 & = 2 \\ & x_1 + x_2 + x_4 & = 4 \\ & -x_1 - 2x_2 + x_5 & = 0 \\ & x_i \geq 0 \text{ for } i \leq 4. & \end{array}$$

The problem can be graphically displayed in $\mathbb{R}^2$. The shaded set P in Figure 1 is a polygon.

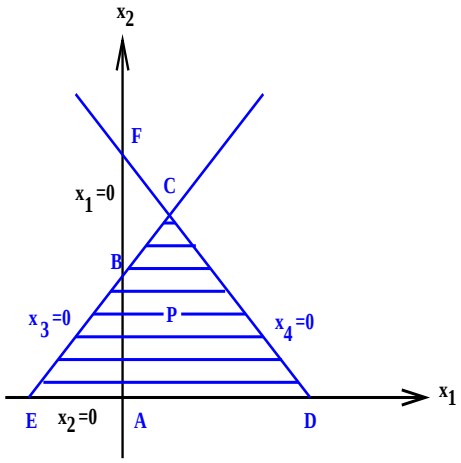

FIGURE 1. Feasible set

We first consider the linear system $Ax = b$ of $LP(I,p)$. For a vector $J = (j_1, \ldots, j_r)$, $j_i \in N$, of indices we denote by $A_J := (a_{j_1}|\ldots|a_{j_r})$ the

submatrix of A with columns a_{j_i}; x_J refers to the vector $(x_{j_1}, \ldots, x_{j_r})^T$. For notational simplicity, the set $\{j_i \mid i = 1, 2, \ldots, r\}$ of components of J will also be denoted by J, and we will use the notation $p \in J$, if there exists t such that $p = j_t$. We define

Definition 1.1: *An index vector $J = (j_1, \ldots, j_M)$ with $M := m+1$ different indices $j_i \in N$ is called a basis of $Ax = b$ resp. of $LP(I, p)$, if A_J is regular.*

Obviously, A has a basis if and only if the rows of A are linearly independent. Besides J, sometimes A_J is referred to as a basis as well; the variables x_i with $i \in J$ are called *basis variables*, the other variables x_k (indizes k) with $k \notin J$ are said to be *non-basis variables* (*non-basis indices*). In case an index vector K contains all non-basis indices, we will write $J \oplus K = N$.

In the previous example, $J_A := (3, 4, 5)$, $J_B := (4, 5, 2)$ are bases.

To a basis J, $J \oplus K = N$ we assign a uniquely determined solution $\bar{x} = \bar{x}(J)$ of $Ax = b$, called a *basis solution*, with the property $\bar{x}_K = 0$. Since

$$A\bar{x} = A_J\bar{x}_J + A_K\bar{x}_K = A_J\bar{x}_J = b \ ,$$

$\bar{x}$ is given by

$$\bar{x}_J := \bar{b}, \quad \bar{x}_K := 0 \text{ with } \bar{b} := A_J^{-1}b. \tag{1.21}$$

Moreover, for any basis J, a solution x of $Ax = b$ is uniquely determined by its non-basis part x_K and its basis part $\bar{x}$: This follows from the multiplication of $Ax = A_Jx_J + A_Kx_K = b$ by A_J^{-1} and (1.21):

$$\begin{aligned} x_J &= \bar{b} - A_J^{-1}A_Kx_K \\ &= \bar{x}_J - A_J^{-1}A_Kx_K. \end{aligned} \tag{1.22}$$

Choosing x_K arbitrarily and defining x_J and hence x by (1.22), we have that x solves $Ax = b$. (1.22) thus provides a specific parametrization of the solution set $\{x \mid Ax = b\}$ via the parameters x_K.

If the basis solution $\bar{x}$ associated with the basis J of $Ax = b$ is a feasible solution of $LP(I, p)$, $\bar{x} \in P$, i.e., due to $\bar{x}_K = 0$ there holds

$$\bar{x}_i \geq 0 \text{ for all } i \in I \cap J \ , \tag{1.23}$$

J is called a *feasible basis* of $LP(I, p)$ and $\bar{x}$ is said to be a *feasible basis solution.* Moreover, a feasible basis is called *non-degenerate*, if instead of (1.23) the sharper condition

$$\bar{x}_i > 0 \text{ for all } i \in I \cap J. \tag{1.24}$$

holds true. The linear program $LP(I, p)$ is said to be non-degenerate, if all feasible bases J of $LP(I, p)$ are non-degenerate.

Geometrically, the feasible basis solutions of the different bases of $LP(I, p)$ correspond to the vertices of the polyhedra P of feasible solutions, provided the set of vertices of P is non-empty. In the example (cf. Fig. 1), the vertex $A \in P$ corresponds to the feasible basis $J_A = (3, 4, 5)$, since A is determined by $x_1 = x_2 = 0$ and $\{1, 2\}$ is the complementary set of J_A with respect to $N = \{1, 2, 3, 4, 5\}$; B corresponds to $J_B = (4, 5, 2)$, C corresponds

to $J_C = (1, 2, 5)$ etc. The basis $J_E = (1, 4, 5)$ is non-feasible, since the associated basis solution E is not feasible ($E \notin P$).

The simplex method for the solution of linear programs due to G.B. Dantzig proceeds as follows: Starting from a feasible basis J_0 of $LP(I, p)$ with $p \in J_0$, the *simplex steps*

$$J_i \to J_{i+1}$$

recursively generate a sequence $\{J_i\}$ of additional feasible bases J_i of $LP(I, p)$ with $p \in J_i$ and the following property: The values $\bar{x}(J_i)_p$ of the objective functional associated with the feasible basis solutions $\bar{x}(J_i)$ w.r.t. J_i are non-decreasing

$$\bar{x}(J_i)_p \leq \bar{x}(J_{i+1})_p \text{ for } \; i \geq 0 \; .$$

If the bases J_i are non-degenerate and $LP(I, p)$ admits an optimal solution, the sequence $\{J_i\}$ terminates after finitely many steps with a basis J_L whose basis solution $\bar{x}(J_L)$ is an optimal solution of $LP(I, p)$, and we have

$$\bar{x}(J_i)_p < \bar{x}(J_{i+1})_p \text{ for } 0 \leq i \leq L-1 \; .$$

Furthermore, two subsequent bases $J = (j_1, \ldots, j_M)$ and $\tilde{J} = (\tilde{j}_1, \ldots, \tilde{j}_M)$ are *neighbors* in the following sense: J and $\tilde{J}$ possess exactly $M-1$ common components. J and $\tilde{J}$ are related by an exchange of indices: There exist exactly two indices $q, s \in N$ with $q \in J$, $s \notin J$ and $q \notin \tilde{J}$, $s \in \tilde{J}$, i.e., $\tilde{J} = (J \cup \{s\}) \backslash \{q\}$.

For non-degenerate problems, neighboring feasible bases correspond geometrically to neighboring vertices of P. In the example (cf. Fig. 1), we have that $J_A = (3, 4, 5)$ and $J_B = (4, 5, 2)$ are neighbors, A and B are neighboring vertices.

1.2.1 Phase II of the simplex method

The simplex method, to be more precise 'phase II' of the simplex method assumes the knowledge of a feasible basis J of $LP(I, p)$ with $p \in J$. Such a feasible basis can be obtained by the so-called 'phase I', provided $LP(I, p)$ has a feasible solution. We first describe a typical step of phase II which starts from a feasible basis J and generates a feasible neighboring basis $\tilde{J}$ of $LP(I, p)$:

Simplex step: *Assumption: $J = (j_1, \ldots, j_M)$ is a feasible basis of $LP(I, p)$ with $p = j_t \in J$, $J \oplus K = N$.*

Step 1: *Compute the vector*

$$\bar{b} := A_J^{-1} b$$

and thus the basis solution $\bar{x}$ associated with J satisfying $\bar{x}_J := \bar{b}$, $\bar{x}_K := 0$.

Step 2: *Compute the row vector*

$$\pi := e_t^T A_J^{-1},$$

where $e_t = (0, \ldots, 1, \ldots, 0)^T \in \mathbb{R}^M$ *is the* t*-th unit vector in* $\mathbb{R}^M$*. Using* π*, compute the numbers*

$$c_k := \pi a_k, \qquad k \in K.$$

Step 3: *Check whether*

$$\begin{aligned} c_k \geq 0 \quad & \textit{for all } k \in K \cap I \\ \textit{and } c_k = 0 \quad & \textit{for all } k \in K \setminus I \end{aligned} \tag{1.25}$$

If yes, stop: the basis solution $\bar{x}$ *is the optimal solution of* $LP(I, p)$*.*
If no, determine $s \in K$ *such that*

$$c_s = \min\{c_k < 0 \mid k \in K \cap I\} \quad \textit{or} \quad |c_s| = \max\{|c_k| \neq 0, \mid k \in K \setminus I\} \tag{1.26}$$

and set $\sigma := -\text{sign}(c_s)$*.*

Step 4: *Compute the vector*

$$\bar{a} := (\bar{\alpha}_1, \bar{\alpha}_2, \ldots, \bar{\alpha}_M)^T := A_J^{-1} a_s.$$

Step 5: *If*

$$\sigma\bar{\alpha}_i \leq 0 \quad \textit{for all } i \textit{ with } j_i \in I, \tag{1.27}$$

stop: $LP(I, p)$ *does not admit a finite optimum. Otherwise,*

Step 6: *determine an index* r *with* $j_r \in I$*,* $\sigma\bar{\alpha}_r > 0$ *and*

$$\frac{\bar{b}_r}{\sigma\bar{\alpha}_r} = \min\left\{ \frac{\bar{b}_i}{\sigma\bar{\alpha}_i} \;\middle|\; i : j_i \in I \;\&\; \sigma\bar{\alpha}_i > 0 \right\}.$$

Step 7: *Choose as* $\tilde{J}$ *any suitable index vector with*

$$\tilde{J} := (J \cup \{s\}) \setminus \{j_r\},$$

for instance

$$\tilde{J} := (j_1, \ldots, j_{r-1}, s, j_{r+1}, \ldots, j_M)$$

or

$$\tilde{J} := (j_1, \ldots, j_{r-1}, j_{r+1}, \ldots, j_M, s).$$

We will motivate these rules and assume that $J = (j_1, \ldots, j_M)$ is a feasible basis of $LP(I, p)$ with $p = j_t \in J$ and$J \oplus K = N$. In view of (1.21), step 1 provides the associated basis solution $\bar{x} = \bar{x}(J)$. Since all solution of $Ax = b$

can be represented in the form (1.22), due to $p = j_t$, the objective functional satisfies

$$x_p = e_t^T x_J = \bar{x}_p - e_t^T A_J^{-1} A_K x_K = \bar{x}_p - \pi A_K x_K = \bar{x}_p - c_K x_K, \tag{1.28}$$

if the row vector π and the components c_k of the row vector c_K, $c_K^T \in \mathbb{R}^{n-m}$ are chosen as in step 2. c_K is called the vector of *reduced costs:*

In view of (1.28), $c_k, k \in K$ stands for the amount by which the objective functional x_p decreases, if x_k is enlarged by one unit. Hence, if (1.25) is satisfied (cf. step 3), for each *feasible* solution x of $LP(I,p)$, (1.28) and $x_i \geq 0, i \in I$ imply

$$x_p = \bar{x}_p - \sum_{k \in K \cap I} c_k x_k \leq \bar{x}_p,$$

i.e., the basis solution $\bar{x}$ is the optimal solution of $LP(I,p)$. This motivates the test (1.25) and the assertion a) of step 3. If (1.25) does not old true, there exists an index $s \in K$ for which either

$$c_s < 0, \qquad s \in K \cap I, \tag{1.29}$$

or

$$|c_s| \neq 0, \qquad s \in K \setminus I\ . \tag{1.30}$$

Assume that s is such an index. We set $\sigma := -\text{sign}(c_s)$. Since due to (1.28) an increase in σx_s yields an increase in the objective functional x_p, we consider the following family of vectors $x(\theta) \in \mathbb{R}^{n+m_1+1}$, $\theta \in \mathbb{R}$,

$$x(\theta)_J := \bar{b} - \theta\sigma A_J^{-1} a_s = \bar{b} - \theta\sigma\bar{a}, \tag{1.31}$$

$$x(\theta)_s := \theta\sigma, \tag{1.32}$$

$$x(\theta)_k := 0 \quad \text{for } k \in K,\ k \neq s. \tag{1.33}$$

Here, $\bar{a} := A_J^{-1} a_s$ is chosen as in step 4.

In the example we have $I = \{1,2,3,4\}$, and $J_0 = J_A = (3,4,5)$ is a feasible basis, $K_0 = (1,2)$, $p = 5 \in J_0$, $t_0 = 3$. We obtain:

$$A_{J_0} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}, \qquad \bar{b} = \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix}$$

$\bar{x}(J_0) = (0,0,2,4,0)^T$ ($\hat{=}$ point A in Fig. 1) and $\pi A_{J_0} = e_{t_0}^T \Rightarrow \pi = (0,0,1)$. The reduced costs are $c_1 = \pi a_1 = -1$, $c_2 = \pi a_2 = -2$. Hence, J_0 is not optimal. Choosing in step 3 the index $s = 2$, we obtain

$$\bar{a} = A_{J_0}^{-1} a_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

The family of solutions $x(\theta)$ is given by

$$x(\theta) = (0, \theta, 2-\theta, 4-\theta, 2\theta)^T.$$

Geometrically, in Fig. 1 $x(\theta)$, $\theta \geq 0$ describes a ray pointing from vertex A ($\theta = 0$) towards the neighboring vertex B ($\theta = 2$) along an edge of the polyhedra P.

In view of (1.22), we have $Ax(\theta) = b$ for all $\theta \in \mathbb{R}$. In particular, observing (1.28), $\bar{x} = x(0)$ and the choice of σ

$$x(\theta)_p = \bar{x}_p - c_s x(\theta)_s = \bar{x}_p + \theta\,|c_s|\,, \tag{1.34}$$

such that the objective functional increases monotonically along the ray $x(\theta)$. Consequently, among the solutions $x(\theta)$ of $Ax = b$ we pick the best *feasible* solution, i.e., we are looking for the largest $\theta \geq 0$ with

$$x(\theta)_l \geq 0 \quad \text{for all } l \in I.$$

Taking (1.33) into account, this is equivalent to choosing the largest $\theta \geq 0$ with

$$x(\theta)_{j_i} \equiv \bar{b}_i - \theta\sigma\bar{\alpha}_i \geq 0 \quad \text{for all } i \text{ with } \quad j_i \in I\ , \tag{1.35}$$

since $x(\theta)_k \geq 0, k \in K \cap I$, $\theta \geq 0$ is automatically satisfied due to (1.33). Now, if $\sigma\bar{\alpha}_i \leq 0$ for all $j_i \in I$ (cf. step 5), (1.35) implies that $x(\theta), \theta \geq 0$ is a feasible solution of $LP(I,p)$ with $\sup\{x(\theta)_p \mid \theta \geq 0\} = +\infty$: Then, $LP(I,p)$ does not have a finite optimum. This justifies step 5. Otherwise, there is a largest $\theta =: \bar{\theta}$ such that (1.35) holds true:

$$\bar{\theta} = \frac{\bar{b}_r}{\sigma\bar{\alpha}_r} = \min\left\{ \frac{\bar{b}_i}{\sigma\bar{\alpha}_i} \;\middle|\; i : j_i \in I \;\&\; \sigma\bar{\alpha}_i > 0 \right\}.$$

This determines an index r with $j_r \in I$, $\sigma\bar{\alpha}_r > 0$ and

$$x(\bar{\theta})_{j_r} = \bar{b}_r - \bar{\theta}\sigma\bar{\alpha}_r = 0, \quad x(\bar{\theta}) \text{ is a feasible solution.} \tag{1.36}$$

In the example we have

$$\bar{\theta} = 2 = \frac{\bar{b}_1}{\bar{\alpha}_1} = \min\left\{ \frac{\bar{b}_1}{\bar{\alpha}_1},\ \frac{\bar{b}_2}{\bar{\alpha}_2} \right\},\ r = 1.$$

$x(\bar{\theta}) = (0, 2, 0, 2, 4)^T$ corresponds to the vertex B in Fig. 1.

Due to the feasibility of J we have $\bar{\theta} \geq 0$, and (1.34) implies

$$x(\bar{\theta})_p \geq \bar{x}_p\ .$$

If J is non-degenerate, there holds $\bar{\theta} > 0$, and hence

$$x(\bar{\theta})_p > \bar{x}_p.$$

In view of (1.22), (1.33), and (1.36) $x = x(\theta)$ is the uniquely determined solution of $Ax = b$ with the additional property

$$x_{j_r} = 0,\ x_k = 0 \quad \text{for } \ k \in K,\ k \neq s\ ,$$

i.e., $x_{\tilde{K}} = 0, \tilde{K} := (K \cup \{j_r\}) \setminus \{s\}$. The uniqueness of x implies the regularity of $A_{\tilde{J}}$, $\tilde{J} := (J \cup \{s\}) \setminus \{j_r\}$. Hence, $x(\theta) = \bar{x}(\tilde{J})$ is a basis solution associated with the neighboring feasible basis $\tilde{J}$, and there holds

$$\bar{x}(\tilde{J})_p > \bar{x}(J)_p, \qquad \text{if } J \text{ is non-degenerate,} \tag{1.37}$$

$$\bar{x}(\tilde{J})_p \geq \bar{x}(J)_p, \qquad \text{otherwise.} \tag{1.38}$$

In the example we obtain the new basis

$$J_1 = (2,4,5) = J_B, \; K_1 = (1,3) \; ,$$

which corresponds to the vertex B in Fig. 1. With regard to the objective functional x_5, we have that B is 'better' than $A : \bar{x}(J_B)_5 = 4 > \bar{x}(J_A)_5 = 0$.

According to the definition of r, we always have $j_r \in I$, whence

$$J \setminus I \subset \tilde{J} \setminus I,$$

i.e., in the step $J \to \tilde{J}$, there are only constrained variables x_{j_r}, $j_r \in I$ which are eliminated from the basis. As soon as a free variable x_s, $s \notin I$, has become a basis variable, it remains a basis variable in all subsequent simplex steps. In particular, $p \in \tilde{J}$ due to $p \in J$ and $p \notin I$. The new basis $\tilde{J}$ satisfies again the assumption of the simplex step such that the simplex step can be applied to $\tilde{J}$. Starting from a first feasible basis J_0 of $LP(I,p)$ with $p \in J_0$, we thus obtain a sequence

$$J_0 \to J_1 \to J_2 \to \cdots$$

of feasible bases J_i of $LP(I,p)$ with $p \in J_i$, for which in case of non-degeneracy of allJ_i there holds

$$\bar{x}(J_0)_p < \bar{x}(J_1)_p < \bar{x}(J_2)_p < \cdots \; .$$

In this case, J_i will not occur again. Since there are only finitely many index vectors J, the method must terminate after finitely many steps. Consequently, for the simplex method we have shown:

Theorem 1.1 *Assume that J_0 is a feasible basis of $LP(I,p)$ with $p \in J_0$. If $LP(I,p)$ is non-degenerate, starting from J_0, the simplex method generates a finite sequence of feasible bases J_i of $LP(I,p)$ with $p \in J_i$ and $\bar{x}(J_i)_p < \bar{x}(J_{i+1})_p$. The last basis solution either is an optimal solution of $LP(I,p)$ or $LP(I,p)$ does not admit a finite optimum.*

We proceed with the example: As a result of the first simplex step we have obtained the new feasible basis $J_1 = (2,4,5) = J_B$, $K_1 = (1,3)$, $t_1 = 3$, such that

$$A_{J_1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}, \bar{b} = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}, \quad \bar{x}(J_1) = (0,2,0,2,4)^T \; (\hat{=} \text{ vertex } B),$$

$$\pi A_{J_1} = e_{t_1}^T \Rightarrow \pi = (2,0,1).$$

The reduced costs are $c_1 = \pi a_1 = -3$, $c_3 = \pi a_3 = 2$. Hence,J_1 is not optimal:

$$s = 1,\ \bar{a} = A_{J_1}^{-1} a_1 \Rightarrow \bar{a} = \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix} \Rightarrow r = 2.$$

Therefore,

$$J_2 = (2,1,5) = J_C, \quad K_2 = (3,4), \quad t_2 = 3$$
$$A_{J_2} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ -2 & -1 & 1 \end{bmatrix}, \qquad \bar{b} = \begin{bmatrix} 3 \\ 1 \\ 7 \end{bmatrix},$$
$$\bar{x}(J_2) = (1,3,0,0,7) \ (\hat{=} \text{ vertex } C),$$
$$\pi A_{J_2} = e_{t_2}^T \quad \Rightarrow \quad \pi = (\frac{1}{2}, \frac{3}{2}, 1)$$

The reduced costs are $c_3 = \pi a_3 = \frac{1}{2} > 0$, $c_4 = \pi a_4 = \frac{3}{2} > 0$.

The optimality criterion is satisfied, and hence, $\bar{x}(J_2)$ is optimal, i.e., $\bar{x}_1 = 1$, $\bar{x}_2 = 3$, $\bar{x}_3 = 0$, $\bar{x}_4 = 0$, $\bar{x}_5 = 7$. The optimal value of the objective functional x_5 is given by $\bar{x}_5 = 7$.

The practical implementation of the simplex method, each step $J \to \tilde{J}$ (requires the solution of three linear systems with the matrix A_J

$$A_J \bar{b} = b \qquad \Rightarrow \bar{b} \qquad \text{(Step 1)}, \tag{1.39}$$
$$\pi A_J = e_t^T \qquad \Rightarrow \pi \qquad \text{(Step 2)}, \tag{1.40}$$
$$A_J \bar{a} = a_s \qquad \Rightarrow \bar{a} \qquad \text{(Step 4)}. \tag{1.41}$$

The computational work for the successive bases $J \to \tilde{J} \to \cdots$ can be significantly reduced by taking into account that the bases $J \to \tilde{J}$ are neighbors: the new basis matrix $A_{\tilde{J}}$ is obtained from A_J by replacing a column of A_J with another column of A. This can be used, e.g., in case of decompositions of the basis matrix A_J of type

$$FA_J = R, \qquad F \text{ regular}, \ R \text{ upper triangular matrix}.$$

Taking advantage of such a decomposition, the linear systems (1.41) can be easily solved:

$$R\bar{b} = Fb \Rightarrow \bar{b},$$
$$R^T z = e_t \Rightarrow z \Rightarrow \pi = z^T F,$$
$$R\bar{a} = Fa_s \Rightarrow \bar{a}.$$

Moreover, by means of a decomposition $FA_J = R$ of A_J, in each simplex step a similar decomposition $\tilde{F}A_{\tilde{J}} = \tilde{R}$ can be easily computed for the neighboring basis $\tilde{J} = (j_1, \ldots, j_{r-1}, j_{r+1}, \ldots, j_{M,s})$ (cf. Step 7): The matrix

$FA_{\tilde{J}}$ is an upper Hessenberg matrix of the form (shown here for $M = 4$, $r = 2$)

$$FA_{\tilde{J}} = \begin{bmatrix} x & x & x & x \\ & x & x & x \\ & x & x & x \\ & & x & x \end{bmatrix} =: R' \ .$$

The subdiagonal elements can be easily eliminated by elimination matrices $E_{r,r+1}$, $E_{r+1,r+2}$, $\ldots$, $E_{M-1,M}$ and hence, R' can be transformed into an upper triangular matrix $\tilde{R}$:

$$\tilde{F}A_{\tilde{J}} = \tilde{R}, \quad \tilde{F} := EF, \tilde{R} := ER', \quad E := E_{M-1,M}E_{M-2,M-1}\ldots E_{r,r+1}.$$

Therefore, it seems to be reasonable, to implement the simplex method in such a way that in each simplex step $J \to \tilde{J}$ a 4-tuple $\mathcal{M} = \{J; t; F, R\}$ with the property

$$j_t = p, \qquad FA_J = R,$$

is transformed into a similar 4-tuple $\tilde{\mathcal{M}} = \{\tilde{J}; \tilde{t}; \tilde{F}, \tilde{R}\}$. Besides a feasible basis J_0 with $p \in J_0$ of $LP(I, p)$, the initialization of this variant of the simplex method also requires a decomposition $F_0 A_{J_0} = R_0$ of A_{J_0}.

Instead of the decompositions $FA_J = R$, standard implementations of the simplex method use other quantities which enable an efficient solution of the linear systems (1.41). The so-called 'Inverse-Basis-Method' uses 5-tuples of the form

$$\hat{\mathcal{M}} = \{J; t; B, \bar{b}, \pi\}$$

with

$$j_t = p, \quad B := A_J^{-1}, \quad \bar{b} = A_J^{-1}b, \quad \pi := e_t^T A_J^{-1} \ .$$

Another variant implements 5-tuples

$$\bar{\mathcal{M}} = \{J; t; \bar{A}, \bar{b}, \pi\}$$

with

$$j_t = p, \quad \bar{A} := A_J^{-1}A_K, \quad \bar{b} := A_J^{-1}b, \quad \pi := e_t^T A_J^{-1}, \quad J \oplus K = N \ .$$

In the transition $J \to \tilde{J}$, computational work can be reduced in such a way that for neighboring bases J, $\tilde{J}$ the inverse $A_{\tilde{J}}^{-1}$ can be computed by multiplying A_J^{-1} with a suitable Frobenius matrix G: $A_{\tilde{J}}^{-1} = GA_J^{-1}$. Here, the computational work is even a little bit less as in case of the decomposition $FA_J = R$. However, a serious drawback is the numerical instability: if a basis matrix A_{J_i} is ill conditioned, inevitable large errors in $A_{J_i}^{-1}$, $A_{J_i}^{-1}A_{K_i}$ are amplified in $\hat{\mathcal{M}}_i$ and $\bar{\mathcal{M}}_i$ for all subsequent 5-tuples $\hat{\mathcal{M}}_j$, $\bar{\mathcal{M}}_j$, $j > i$.

The following practical example illustrates the gain in numerical stability, if instead of using the 'Inverse-Basis-Method' the triangular decomposition

will be employed. Consider a linear program with constraints of the form

$$\begin{aligned} Ax &= b, \qquad A = (A^{(1)}, A^{(2)}), \\ x &\geq 0. \end{aligned} \tag{1.42}$$

The matrix A is chosen as the 5×10-matrix given by the 5×5-submatrices $A^{(1)}$, $A^{(2)}$

$$\begin{aligned} &A^{(1)} = (a_{ik}^{(1)}), \quad a_{ik}^{(1)} := 1/(i+k), \quad i,k = 1,\ldots,5, \\ &A^{(2)} := I_5 = \text{5-row unit matrix .} \end{aligned}$$

Here, $A^{(1)}$ is badly conditioned, whereas $A^{(2)}$ is well conditioned. The right-hand side is chosen as the vector $b := A^{(1)} \cdot e$, $e := (1,1,1,1,1)^T$,

$$b_i := \sum_{k=1}^{5} \frac{1}{i+k},$$

so that both bases $J_1 := (1,2,3,4,5)$, $J_2 := (6,7,8,9,10)$ are feasible for (1.42) with the basis solutions

$$\begin{aligned} \bar{x}(J_1) &:= \bar{b}_1 \; 0 \; , \\ \bar{b}_1 &:= A_{J_1}^{-1} b \;=\; (A^{(1)})^{-1} b \;=\; e \; , \\ \bar{x}(J_2) &:= 0 \; \bar{b}_2 \; , \\ \bar{b}_2 &:= A_{J_2}^{-1} b \;=\; (A^{(2)})^{-1} b \;=\; b \; . \end{aligned} \tag{1.43}$$

As a start basis we choose $J_2 = (6,7,8,9,10)$ and transform it by the Inverse-Basis-Method resp. the triangular decomposition method into the new basis J_1 and then, using another sequence of exchange steps, return to the start basis J_2:

$$J_2 \to \cdots \to J_1 \to \cdots \to J_2.$$

For the associated basis solutions (1.43), this cycling process yields the following results (machine accuracy eps $\approx 10^{-11}$, inexact digits are underlined):

Basis	exact basis solution	Inverse-Basis-Method	Triangular decomposition
	$1.4500000000E+00$	$1.4500000000E+00$	$1.4500000000E+00$
	$1.0928571428E+00$	$1.0928571428E+00$	$1.0928571428E+00$
$J_2\bar{b}_2 =$	$8.8452380952E-01$	$8.8452380952E-01$	$8.8452380952E-01$
	$7.4563492063E-01$	$7.4563492063E-01$	$7.4563492063E-01$
	$6.4563492063E-01$	$6.4563492063E-01$	$6.4563492063E-01$
	1	$1.0000000\underline{182}E+00$	$1.0000000\underline{786}E+00$
	1	$9.9999\underline{84079}E-01$	$9.9999\underline{16035}E-01$
$J_1\bar{b}_1 =$	1	$1.0000004\underline{372}E+00$	$1.0000\underline{27212}E+00$
	1	$9.9999\underline{52118}E-01$	$9.9999\underline{56491}E-01$
	1	$1.0000001\underline{826}E+00$	$1.0000\underline{14837}E+00$
	$1.4500000000E+00$	$1.4500\underline{10511}E+00$	$1.4500000000E+00$
	$1.0928571428E+00$	$1.0928\underline{5}9\underline{972}E+00$	$1.0928571\underline{427}E+00$
$J_2\bar{b}_2 =$	$8.8452380952E-01$	$8.8452\underline{453057}E-01$	$8.8452380\underline{950}E-01$
	$7.4563492063E-01$	$7.4563\underline{554473}E-01$	$7.4563492\underline{060}E-01$
	$6.4563492063E-01$	$6.4563\underline{547103}E-01$	$6.4563492\underline{059}E-01$

We obtain the following result: Due to $A_{J_2} = I_5$, at the beginning both methods provide the exact solution. For the basis J_1, both methods yield the same inexact results which reflect the bad condition of A_{J_1}. This can not be avoided, unless the computations are carried out with higher accuracy. After the step with the badly conditioned basis matrix A_{J_1}, the situation changes drastically in favor of the triangular decomposition method. This method provides the basis solution associated with J_2 practically at machine accuracy, whereas the Inverse-Basis-Method reproduces the solution with the same inaccuracy as in case of the previous basis J_1. In case of the Inverse-Basis-Method, all subsequent bases inherit the bad condition of the basis matrix A_J.

1.2.2 Phase I of the simplex method

The initialization of phase II of the simplex method requires a feasible basis J_0 of $LP(I,p)$ with $p = j_{t_0} \in J_0$ resp. an associated 4-tuple $\mathcal{M}_0 = \{J_0; t_0; F_0, R_0\}$, where the regular matrix F_0 and the regular upper triangular matrix R_0 provide a decomposition $F_0 A_{J_0} = R_0$ of the basis matrix A_{J_0}.

In some special cases, it is easy to find a feasible basis $J_0(\mathcal{M}_0)$, e.g., when we are faced with a linear program of the following form:

$$\begin{aligned} \text{minimize} \quad & c_1 x_1 + \cdots + c_n x_n \\ x \in \mathbb{R}^n : \quad & a_{i1} x_1 + \cdots + a_{in} x_n \le b_i, \quad i = 1, 2, \ldots, m \\ & x_i \ge 0 \quad \text{for} \;\; i \in I_1 \subset \{1, 2, \ldots, n\}, \end{aligned}$$

where $b_i \geq 0$ for $i = 1, 2, \ldots, m$. Introducing slack variables, we obtain the equivalent problem

$$\begin{array}{rl} \text{maximize} & x_{n+m+1} \\ x \in \mathbb{R}^{n+m+1}: & a_{i1}x_1 + \cdots + a_{in}x_n + x_{n+i} = b_i,\ i = 1, 2, \ldots, m \\ & c_1x_1 + \cdots + c_nx_n + x_{n+m+1} = 0, \\ & x_i \geq 0 \text{ for } \ i \in I_1 \cup \{n+1, n+2, \ldots, n+m\}\ , \end{array}$$

representing the standard form $LP(I, p)$ of the previous section

$$A = \begin{bmatrix} a_{11} & \ldots & a_{1n} & 1 & & \\ \vdots & & \vdots & & \ddots & \\ a_{m1} & \ldots & a_{mn} & & & 1 \\ c_1 & \ldots & c_n & & & 1 \end{bmatrix}, b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \\ 0 \end{bmatrix},$$

$$p := n + m + 1, \quad I := I_1 \cup \{n+1, n+2, \ldots, n+m\}\ .$$

In view of $b_i \geq 0$, we have that $J_0 := (n+1, n+2, \ldots, n+m+1)$ is a feasible basis with $p = j_t \in J_0$, $t := m + 1$. A corresponding $\mathcal{M}_0 = (J_0; t_0; F_0, R_0)$ is given by $t_0 := m + 1$, $F_0 := R_0 := I_{m+1}$, where I_{m+1} denotes the unit matrix with $m + 1$ rows.

For more general linear programs (P), the so-called 'phase I' provides a feasible basis. This phase is based on techniques where phase II is applied to a modified linear program $(\tilde{P})$ which is such that a feasible start basis is known, i.e., $(\tilde{P})$ can be solved by means of phase II and each optimal basis of $(\tilde{P})$ yields a feasible start basis for (P). Here, we will only describe one of these techniques. We refer to the literature on linear programming with respect to other techniques.

We consider a linear program of the following form:

$$\begin{array}{rll} \text{minimize} & & c_1x_1 + \cdots + c_nx_n \\ x \in \mathbb{R}^n: & & a_{j1}x_1 + \cdots + a_{jn}x_n = b_j, \quad j = 1, 2, \ldots, m \\ & & x_i \geq 0 \text{ for } \ i \in I \subseteq \{1, 2, \ldots, n\}\ , \end{array} \tag{1.44}$$

and assume that $b_j \geq 0$ holds true for all j (multiply the j-th constraint by -1, if $b_j < 0$).

First, we extend the constraints by introducing artificial variables x_{n+1}, $\ldots$, x_{n+m},

$$\begin{array}{llllll} a_{11}x_1 & +\cdots+ & a_{1n}x_n & +x_{n+1} & & = b_1 \\ \vdots & & \vdots & & \ddots & \vdots \\ a_{m1}x_1 & +\cdots+ & a_{mn}x_n & & +x_{n+m} & = b_m \\ x_i \geq 0 \text{ for } & i \in I \cup \{n+1, \ldots, n+m\}\ . & & & & \end{array} \tag{1.45}$$

Obviously, the feasible solutions of (1.44) are uniquely assigned to those feasible solutions of (1.45) for which the artificial variables are zero:

$$x_{n+1} = x_{n+2} = \cdots = x_{n+m} = 0\ . \tag{1.46}$$

Now, we design a maximization problem with the constraints (1.45) whose optimal solutions satisfy (1.46), provided (1.44) admits feasible solutions. For this purpose, we consider $LP(\hat{I},\hat{p})$:

$$
\begin{array}{lllllllll}
\text{maximize} & x_{n+m+1} & & & & & & & \\
x: & a_{11}x_1 & +\cdots+ & a_{1n}x_n & +x_{n+1} & & & & = b_1 \\
 & \vdots & & \vdots & & \ddots & & & \vdots \\
 & a_{m1}x_1 & +\cdots+ & a_{mn}x_n & & & +x_{n+m} & & = b_m \\
 & & & & x_{n+1}+ & \cdots & +x_{n+m} & +x_{n+m+1} & = 0
\end{array}
$$

$x_i \geq 0$ for $i \in \hat{I} := I \cup \{n+1,\ldots,n+m\},\ \hat{p} := n+m+1$

A possible start basis for this problem is $\hat{J}_0 := (n+1,\ldots,n+m+1)$ which is feasible, since the associated basis solution $\bar{x}$ with

$$
\bar{x}_j = 0, \quad \bar{x}_{n+i} = b_i, \quad \bar{x}_{n+m+1} = -\sum_{i=1}^{m} b_i, \quad \text{for } 1 \leq j \leq n, 1 \leq i \leq m,
$$

is feasible due to $b_j \geq 0$.

A 4-tuple $\mathcal{M}_0 = \{\hat{J}_0; \hat{t}_0; \hat{F}_0, \hat{R}_0\}$ corresponding to $\hat{J}_0$ is given by

$$
\hat{t}_0 := m+1, \qquad \hat{F}_0 := \begin{bmatrix} 1 & & & 0 \\ \vdots & \ddots & & \\ 0 & \ldots & 1 & \\ -1 & \ldots & -1 & 1 \end{bmatrix}, \qquad \hat{R}_0 := \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}.
$$

Now, for the solution of $LP(\hat{I},\hat{p})$, phase II of the simplex method can be launched. Due to $x_{n+m+1} = -\sum_{i=1}^{m} x_{n+i} \leq 0$, $LP(\hat{I},\hat{p})$ has a finite maximum and hence, phase II provides an optimal basis $\bar{J}$ and the associated basis solution $\bar{x} = \bar{x}(\bar{J})$ which is the optimal solution of $LP(\hat{I},\hat{p})$.
We distinguish the three cases:

1: $\bar{x}_{n+m+1} < 0$, i.e., (1.46) does not hold true for $\bar{x}$,
2: $\bar{x}_{n+m+1} = 0$ and no artificial variable is a basis variable,
3: $\bar{x}_{n+m+1} = 0$ and there exists an artificial variable in $\bar{J}$.

In case 1, (1.44) is not solvable, since any feasible solution corresponds to a feasible solution of $LP(\hat{I},\hat{p})$ with $x_{n+m+1} = 0$. In case 2, the optimal basis $\bar{J}$ of $LP(\hat{I},\hat{p})$ readily gives a feasible start basis for phase II of the simplex method. Case 3 represents a degenerate problem, since the artificial variables in the basis $\bar{J}$ are zero. If necessary, by a re-numeration of the equations and the artificial variables we may achieve that the artificial variables in the basis $\bar{J}$ are the variables x_{n+1}, x_{n+2}, …, x_{n+k}. In $LP(\hat{I},\hat{p})$, we then eliminate the remaining artificial variables which are not in $\bar{J}$ and instead of x_{n+m+1} introduce a new variable $x_{n+k+1} := -x_{n+1} - \cdots - x_{n+k}$ and a new variable x_{n+k+2} for the objective functional. The optimal basis $\bar{J}$ of $LP(\hat{I},\hat{p})$ yields a feasible start basis $\bar{J} \cup \{x_{n+k+2}\}$ for the problem

equivalent to (1.44)

$$
\begin{array}{l}
\text{maximize } x_{n+k+2} \\
\begin{array}{lccccccccccc}
x: & a_{11}x_1 & +\cdots+ & a_{1n}x_n & +x_{n+1} & & & & & & & = b_1 \\
 & \vdots & & \vdots & & \ddots & & & & & & \vdots \\
 & a_{k1}x_1 & +\cdots+ & a_{kn}x_n & & & +x_{n+k} & & & & & = b_k \\
 & & & & x_{n+1}+ & \cdots & +x_{n+k} & +x_{n+k+1} & & & & = 0 \\
 & a_{k+1,1}x_1 & +\cdots+ & a_{k+1,n}x_n & & & & & & & & = b_{k+1} \\
 & \vdots & & \vdots & & & & & & & & \vdots \\
 & a_{m1}x_1 & +\cdots+ & a_{mn}x_n & & & & & & & & = b_m \\
 & c_1x_1 & +\cdots+ & c_nx_n & & & & & & & +x_{n+k+2} & = 0
\end{array} \\
x_i \geq 0 \text{ or } i \in I \cup \{n+1, \ldots, n+k+1\} \ .
\end{array}
$$

1.3 Primal-Dual Interior Point Methods

1.3.1 Primal-Dual Methods

1.3.1.1 Optimality Conditions for LP

We recall from Chapter 1.1 the definition of the standard form of an LP: Given vectors $b \in \mathbb{R}^m, c \in \mathbb{R}^n$, and a matrix $A \in \mathbb{R}^{m \times n}$, we are looking for a vector $x \in \mathbb{R}^n$ satisfying

$$\text{minimize } c^T x \quad \text{subject to} \quad Ax = b \ , \ x \geq 0 \ . \tag{1.47}$$

The sets

$$\mathcal{F}_P := \{x \in \mathbb{R}^n \mid Ax = b \ , \ x \geq 0\} \ , \tag{1.48}$$

$$\mathcal{F}_P^o := \{x \in \mathbb{R}^n \mid Ax = b \ , \ x > 0\} \tag{1.49}$$

are called the primal feasible set and the primal strictly feasible set, respectively.
The dual of the LP is given by: Find $\lambda \in \mathbb{R}^m, s \in \mathbb{R}^n$, such that

$$\text{maximize } b^T \lambda \quad \text{subject to} \quad A^T \lambda + s = c \ , \ s \geq 0 \ . \tag{1.50}$$

The sets

$$\mathcal{F}_D := \{(\lambda, s) \in \mathbb{R}^m \times \mathbb{R}^n \mid A^T \lambda + s = c \ , \ s \geq 0\} \ , \tag{1.51}$$

$$\mathcal{F}_D^o := \{(\lambda, s) \in \mathbb{R}^m \times \mathbb{R}^n \mid A^T \lambda + s = c \ , \ s > 0\} \tag{1.52}$$

are referred to as the dual feasible set and the dual strictly feasible set, respectively.

Theorem 1.2 (KKT conditions) A vector $(x^*, \lambda^*, s^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ is a solution of (1.47),(1.50) if and only if the following Karush-Kuhn-Tucker

(KKT) conditions are satisfied

$$A^T\lambda + s = c\ , \tag{1.53}$$

$$Ax = b\ , \tag{1.54}$$

$$(x,s) \geq 0\ , \tag{1.55}$$

$$x^T s = 0\ . \tag{1.56}$$

Proof: See, e.g., [2, 3].

The KKT conditions (1.53)-(1.56) represent an LCP which is called the primal-dual problem. The sets

$$\mathcal{F}_{PD} := \{(x,\lambda,s) \mid Ax = b\ ,\ A^T\lambda + s = c\ ,\ (x,s) \geq 0\}\ , \tag{1.57}$$

$$\mathcal{F}^o_{PD} := \{(x,\lambda,s) \mid Ax = b\ ,\ A^T\lambda + s = c\ ,\ (x,s) > 0\} \tag{1.58}$$

are said to be the primal-dual feasible set and the primal-dual strictly feasible set, respectively. Moreover, we denote by

$$\Omega_P := \{x^* \in \mathbb{R}^n \mid x^* \text{ solves (1.47)}\}\ , \tag{1.59}$$

$$\Omega_D := \{(\lambda^*, s^*) \in \mathbb{R}^m \times \mathbb{R}^n \mid (\lambda^*, s^*) \text{ solves (1.50)}\}\ , \tag{1.60}$$

$$\Omega := \Omega_P \times \Omega_D = \{(x^*, \lambda^*, s^*) \text{ solves } (1.53 - (1.56)) \mid \tag{1.61}$$

the primal, the dual, and the primal-dual solution set, respectively.

Theorem 1.3 (Characterization of solutions) There holds:

(i) If the primal and dual problems are feasible, i.e., $\mathcal{F}_{PD} \neq \emptyset$, the set Ω is nonempty.

(ii) If either the primal or the dual problem has an optimal solution, so does the other, and the values of the objective functionals are equal.

Proof of (i): Assertion **(i)** follows from the application of Farkas' lemma to the system

$$\begin{pmatrix} -A & 0 & 0 \\ 0 & I & 0 \\ c^T & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ s \\ \beta \end{pmatrix} + \begin{pmatrix} 0 \\ A^T \\ -b^T \end{pmatrix} \lambda = \begin{pmatrix} -b \\ c \\ 0 \end{pmatrix}\ ,\ (x,s,\beta) \geq 0\ ,$$

whose solutions coincide with Ω.

Proof of (ii): Assume that x^* is an optimal solution of the LP. Due to the necessity of the KKT conditions, there exist λ^*, s^* such that the 3-tuple (x^*, λ^*, s^*) satisfies (1.53)-(1.56). Since the KKT conditions are also sufficient, (x^*, λ^*, s^*) is a primal-dual solution and hence, (λ^*, s^*) is a dual solution. The same argument applies to an optimal solution of the dual problem. The proof that the optimal objective values are equal is left as an exercise.

Corollary (Bounds for the objective functionals) There holds:

(i) Assume that the LP is feasible. Then, the objective functional $c^T x$ is bounded from below on its feasible region if and only if the dual problem is feasible.

(ii) Assume that the dual problem is feasible. Then, the objective functional $b^T\lambda$ is bounded from above on its feasible region if and only if the primal problem is feasible.

Proof: The proof is left as an exercise.

The following result reveals a condition for the existence and boundedness of the primal and dual solution sets:

Theorem 1.4 (Existence of primal/dual solutions) Assume that the primal and dual problems are feasible, i.e., $\mathcal{F}_{PD} \neq \emptyset$. Then, there holds:

(i) If the dual problem has a strictly feasible point, the primal solution set Ω_P is nonempty and bounded.

(ii) If the primal problem has a strictly feasible point, the set

$$\{s^* \in \mathbb{R}^n \mid (\lambda^*, s^*) \in \Omega_D \text{ for some } \lambda^* \in \mathbb{R}^m\}$$

is nonempty and bounded.

Proof of (i): let $(\bar{\lambda}, \bar{s})$ be the strictly feasible dual point and assume that $\hat{x}$ is some primal feasible point. Then, we have

$$0 \;\leq\; \bar{s}^T\hat{x} \;=\; c^T\hat{x} \;-\; b^T\bar{\lambda}\;, \tag{1.62}$$

and the set

$$\mathcal{T} \;:=\; \{x \in \mathbb{R}^n \mid Ax = b\;,\; x \geq 0\;,\; c^Tx \leq c^T\hat{x}\}$$

is nonempty ($\hat{x} \in \mathcal{T}$) and closed. For any $x \in \mathcal{T}$, (1.62) implies

$$\sum_{i=1}^{n} \bar{s}_i x_i \;=\; \bar{s}^T x \;=\; c^Tx - b^T\bar{\lambda} \;\leq\; c^T\hat{x} - b^T\bar{\lambda} \;=\; \bar{s}^T\hat{x}\;.$$

Since $\bar{s}_i > 0, x_i \geq 0$, it follows that

$$x_i \;\leq\; \frac{1}{\bar{s}_i}\;\bar{s}^T\hat{x} \quad\Longrightarrow\quad \|x\|_\infty \;\leq\; \big(\max_{1\leq i\leq n}\frac{1}{\bar{s}_i}\big)\;\bar{s}^T\hat{x}\;.$$

Since x has been arbitrarily chosen from $\mathcal{T}$, we deduce that $\mathcal{T}$ is nonempty, bounded and closed. Consequently, c^Tx must attain its minimum on $\mathcal{T}$, i.e., there exists x^* such that $\mathcal{T}$

$$x^* \in \mathcal{T}\;,\; c^Tx^* \;\leq c^Tx \quad \text{for all } x \in \mathcal{T}\;.$$

Obviously, $x^* \in \Omega_P$. Therefore, Ω_P is nonempty and bounded as a subset of the bounded set $\mathcal{T}$.

Proof of (ii): The proof of **(ii)** is left as an exercise.

We note that there are LPs with $\mathcal{F}_{PD} \neq \emptyset$ but $\mathcal{F}^o_{PD} = \emptyset$, i.e., there is no strictly feasible point. An example is given by

$$\begin{aligned} &\min_{x\in\mathbb{R}^3} x_1 \\ &\text{subject to} \quad x_1 + x_3 \;=\; 0\;,\; x \geq 0\;. \end{aligned} \tag{1.63}$$

The associated dual problem is given by

$$\begin{aligned} (1.64) \qquad & \max_{\lambda\in\mathbb{R},s\in\mathbb{R}^3} 0 \\ & \text{subject to} \quad \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \lambda + \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} , \; s \geq 0 . \end{aligned}$$

Any feasible primal-dual $(x, \lambda, s) \in \mathcal{F}_{PD}$ is of the form $x_1 = x_3 = s_2 = 0$, whence $\mathcal{F}^o_{PD} = \emptyset$.
An important feature of the strictly feasible set $\mathcal{F}^o_{PD}$ is that $\mathcal{F}^o_{PD} \neq \emptyset$ implies that

$$(1.65) \qquad \{(x^*, s^*) \mid (x^*, \lambda^*, s^*) \in \Omega \text{ for some } \lambda^*\}$$

is a bounded set, which is an immediate consequence of the previous theorem.
An important property that will play a crucial role in the convergence analysis of primal-dual interior-point algorithms is strict complementarity:
Assume that (x^*, λ^*, s^*) is a solution of the LCP. Then, (1.56) implies

$$x^*_i = 0 \quad \text{and/or} \quad s^*_i = 0 \quad \text{for all } 1 \leq i \leq n .$$

We define the inactive sets

$$(1.66) \qquad \mathcal{I}_P := \{1 \leq i \leq n \mid x^*_i \neq 0 \text{ for some } x^* \in \Omega_P\} ,$$

$$(1.67) \qquad \mathcal{I}_D := \{1 \leq i \leq n \mid s^*_i \neq 0 \text{ for some } (\lambda^*, s^*) \in \Omega_D\} .$$

Theorem 1.5 (Goldman-Tucker theorem) There holds

$$(1.68) \qquad \mathcal{I}_P \cup \mathcal{I}_D = \{1, ..., n\} .$$

Hence, there exists a primal solution $x^* \in \Omega_P$ and a dual solution $(\lambda^*, s^*) \in \Omega_D$ such that $x^* + s^* > 0$.

Proof: We refer to [5].

Primal-dual solutions (x^*, λ^*, s^*) satisfying $x^* + s^* > 0$ are called strictly complementary solutions. The Goldman-Tucker theorem guarantees the existence of such a solution. The following example shows that an LP may have multiple solutions some of which are strictly complementary and others not: Consider the LP

$$\begin{aligned} (1.69) \qquad & \min_{x\in\mathbb{R}^3} x_1 \\ & \text{subject to} \quad x_1 + x_2 + x_3 = 1 , \; x \geq 0 . \end{aligned}$$

The associated dual problem is as follows

$$\begin{aligned} (1.70) \qquad & \max_{\lambda\in\mathbb{R},s\in\mathbb{R}^3} \lambda \\ & \text{subject to} \quad \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \lambda + s = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} , \quad s \geq 0 . \end{aligned}$$

Primal-dual solutions are given by

$$x^* = (0, t, 1-t)^T \ , \ \lambda^* = 0 \ , \ s^* = (1,0,0)^T \quad , \quad t \in [0,1] \ . \tag{1.71}$$

For $t \in (0,1)$ the solutions are strictly complementary, whereas for $t = 0$ and $t = 1$ they are not, because there is an index i such that both x_i^* and s_i^* are zero.

1.3.1.2 Central Path

The central path $\mathcal{C}$ is an arc of strictly feasible points parametrized by a scalar $\tau > 0$ such that each point $(x_\tau, \lambda_\tau, s_\tau) \in \mathcal{C}$ solves the system

$$A^T\lambda_\tau \ + \ s_\tau \ = \ c \ , \tag{1.72}$$

$$Ax_\tau \ = \ b \ , \tag{1.73}$$

$$x_{\tau,i}s_{\tau,i} \ = \ \tau \quad , \quad 1 \le i \le n \ , \tag{1.74}$$

$$(x,s) \ > \ 0 \ . \tag{1.75}$$

As we shall show, the existence of $\mathcal{C}$ is guaranteed, provided $\mathcal{F}_{PD}^o \neq \emptyset$. As a first result in this direction, we prove:

Lemma 1.1 Assume $\mathcal{F}_{PD}^o \neq \emptyset$ and let $K \ge 0$. Then, the set

$$\{(x,s) \mid (x,\lambda,s) \in \mathcal{F}_{PD} \text{ for some } \lambda \ , \ x^Ts \le K\}$$

is bounded.

Proof: Let $(\bar{x}, \bar{\lambda}, \bar{s}) \in \mathcal{F}_{PD}^o$ and $(x,\lambda,s) \in \mathcal{F}_{PD}, x^Ts \le K$ be arbitrarily given. Then, the two equations

$$\begin{aligned} A(\bar{x} - x) \ &= \ 0 \ , \\ A^T(\bar{\lambda} - \lambda) \ + \ (\bar{s} - s) \ &= \ 0 \end{aligned}$$

imply

$$(\bar{x} - x)^T(\bar{s} - s) \ = \ -\ (\bar{x} - x)^T A^T(\bar{\lambda} - \lambda) \ = \ 0 \ ,$$

whence

$$\bar{x}^Ts \ + \ \bar{s}^Tx \ \le \ K \ + \ \bar{x}^T\bar{s} \ .$$

Due to $(\bar{x}, \bar{s}) > 0$ we have

$$0 \ < \ \xi \ := \ \min_{1 \le i \le n} \ \min \ (\bar{x}_i, \bar{s}_i) \ .$$

Hence, from the previous inequality we deduce

$$\xi e^T(x+s) \ \le \ K \ + \ \bar{x}^T\bar{s} \ ,$$

where $e := (1, ..., 1)^T$, and further

$$0 \ \le x_i \ \le \ \frac{1}{\xi}(K + \bar{x}^T\bar{s}) \quad , \quad 0 \ \le s_i \ \le \ \frac{1}{\xi}(K + \bar{x}^T\bar{s}) \quad , \quad 1 \le i \le n \ .$$

Theorem 1.6 (Existence of the central path) Under the assumption $\mathcal{F}_{PD}^o \neq \emptyset$, for every $\tau > 0$ there is a solution $(x_\tau, \lambda_\tau, s_\tau)$ of (1.72)-(1.75). Moreover, the (x_τ, s_τ)-component of the solution is uniquely determined.

Proof: We prove that (x_τ, s_τ) is the unique minimizer of

$$\min_{(x,s)\in\mathcal{H}^o} f_\tau(x,s) \ , \tag{1.76}$$

where $f_\tau(x,s)$ is the logarithmic barrier function

$$f_\tau(x,s) \ := \ \frac{1}{\tau}x^T s \ - \ \sum_{j=1}^{n}\log(x_j s_j) \tag{1.77}$$

and $\mathcal{H}^o$ is the reduced strictly feasible set

$$\mathcal{H}^o \ := \ \{(x,s) \mid (x,\lambda,s) \in \mathcal{F}^o_{PD} \text{ for some } \lambda \in \mathbb{R}^m\} \ .$$

We first show that each level set of f_τ

$$\{(x,s) \in \mathcal{H}^o \mid f_\tau(x,s) \le \kappa\} \quad , \quad \kappa > 0 \ ,$$

is contained in a compact subset of $\mathcal{H}^o$. For that purpose, we rewrite f_τ according to

$$f_\tau(x,s) \ = \ \sum_{j=1}^{n} g(\frac{x_j s_j}{\tau}) + n - n\,\log\,\tau \ , \tag{1.78}$$

where g denotes the strictly convex, nonnegative function

$$g(t) \ := \ t - \log\,t - 1 \quad , \quad t \in R_+$$

satisfying

$$g(t) \ \to \ \infty \quad \text{for } t \to 0 \text{ and } t \to \infty \ . \tag{1.79}$$

In view of (1.78) we have

$$f_\tau(x,s) \le \kappa \quad \Longleftrightarrow \quad \sum_{j=1}^{n} g(\frac{x_j s_j}{\tau}) \le \bar{\kappa} := \kappa - n + n\,\log\,\tau \ ,$$

whence for each $1 \le i \le n$

$$g(\frac{x_i s_i}{\tau}) \ \le \ \bar{\kappa} - \sum_{j\neq i} g(\frac{x_j s_j}{\tau}) \ \le \ \bar{\kappa} \ .$$

Taking (1.79) into account, there exists $M > 0$ such that

$$\frac{1}{M} \ \le \ x_i s_i \ \le \ M \quad , \quad 1 \le i \le n \ , \tag{1.80}$$

and consequently,

$$x^T s \ \le \ nM \ . \tag{1.81}$$

Observing the previous lemma, from (1.81) we deduce the existence of $M_u > 0$ such that

$$x_i \in (0, M_u] \quad , \quad s_i \in (0, M_u] \quad , \quad 1 \le i \le n \ .$$

Using (1.80), it follows that for all $1 \leq i \leq n$

$$\begin{aligned} x_i &\geq \frac{1}{Ms_i} \geq \frac{1}{MM_u} , \\ s_i &\geq \frac{1}{Mx_i} \geq \frac{1}{MM_u} . \end{aligned}$$

Setting $M_\ell := 1/(MM_u)$, we conclude

$$x_i \in [M_\ell, M_u] \quad , \quad s_i \in [M_\ell, M_u] \quad , \quad 1 \leq i \leq n .$$

Since f_τ is bounded from below on $\mathcal{H}^o$ according to

$$f_\tau(x,s) \geq n\,(1 - \log\tau) ,$$

f_τ attains its minimum in $\mathcal{H}^o$. This minimum must be unique, since f_τ is strictly convex in $\mathcal{H}^o$ which follows from the fact that the first term in (1.77) is linear on $\mathcal{H}^o$, i.e.,

$$x^T s = c^T x - b^T\lambda = c^T x - \bar{x}^T A^T\lambda = c^T x - \bar{x}^T(c - s) = c^T x + \bar{x}^T s - \bar{x}^T c$$

for any $(x,s) \in \mathcal{H}^o$ and any $\bar{x}$ such that $A\bar{x} = b$, whereas the second summation term has a positive definite Hessian for all $(x,s) > 0$.
It remains to be shown that the unique minimizer (x_τ, s_τ) of (1.76) corresponds to the (x,s)-component of the solution of (1.72)-(1.75). We note that (x_τ, s_τ) solves the problem

$$\begin{aligned} &\min_{x,s} f_\tau(x,s) , \\ &\text{subject to} \quad Ax = b \ , \ A^T\lambda + s = c \ , \ (x,s) > 0 . \end{aligned} \tag{1.82}$$

Setting

$$X := \mathrm{diag}(x_1, ..., x_n) \quad , \quad S := \mathrm{diag}(s_1, ..., s_n) ,$$

the KKT conditions for (1.82) imply the existence of Lagrange multipliers ν and μ such that

$$\frac{\partial}{\partial x} f_\tau(x,s) = A^T\nu \implies \frac{s}{\tau} - X^{-1}e = A^T\nu , \tag{1.83}$$

$$\frac{\partial}{\partial \lambda} f_\tau(x,s) = A\mu \implies 0 = -\,A\mu , \tag{1.84}$$

$$\frac{\partial}{\partial s} f_\tau(x,s) = \mu \implies \frac{x}{\tau} - S^{-1}e = \mu . \tag{1.85}$$

Combining (1.84) and (1.85) yields

$$A(\frac{x}{\tau} - S^{-1}e) = 0 . \tag{1.86}$$

Taking the inner product of the left-hand side in (1.86) with ν and observing (1.83), we find

$$\begin{aligned} &(A(\frac{x}{\tau} - S^{-1}e))^T\nu = (\frac{x}{\tau} - S^{-1}e)^T A^T\nu = \\ &= (\frac{x}{\tau} - S^{-1}e)^T(\frac{s}{\tau} - X^{-1}e) = 0 . \end{aligned}$$

It follows that

$$\begin{aligned} 0 &= (\frac{1}{\tau}Xe - S^{-1}e)^T(X^{-1/2}S^{1/2})(X^{1/2}S^{-1/2})(\frac{1}{\tau}Se - X^{-1}e) = \\ &= \|\frac{1}{\tau}(XS)^{1/2}e - (XS)^{-1/2}e\|^2 , \end{aligned}$$

and hence,

$$\frac{1}{\tau}(XS)^{1/2}e - (XS)^{-1/2}e = 0 \quad \Longrightarrow \quad XSe = \tau e ,$$

which concludes the proof of the theorem.

A commonly used primal-dual interior-point method is to couple the inequality constraints $x \geq 0$ by a standard logarithmic barrier function parametrized by a barrier parameter $\tau > 0$ which leads to the family of parametrized minimization subproblems

$$\min_x \; c^T x \; - \; \tau \sum_{i=1}^{n} \log x_i \quad \text{subject to } Ax = b . \tag{1.87}$$

The domain of the logarithmic barrier function is the set of strictly feasible points for the LP, and the optimality conditions imply the existence of a Lagrange multiplier $\lambda \in \mathbb{R}^m$ such that

$$\begin{aligned} \tau X^{-1}e + A^T\lambda &= c , \\ Ax &= b , \\ x &> 0 . \end{aligned}$$

If we define $s \in \mathbb{R}^n$ by means of

$$s_i := \frac{\tau}{x_i} \quad , \quad 1 \leq i \leq n ,$$

we see that the minimizer x_τ of (1.87) is the x-component of the central path vector $(x_\tau, \lambda_\tau, s_\tau) \in \mathcal{C}$. Hence, we may refer to the path

$$\{x_\tau \in \mathbb{R}^n \mid x_\tau \text{ solves (1.87)} , \; \tau > 0\} \tag{1.88}$$

as the primal central path.

1.3.2 Path following algorithms

1.3.2.1 Preliminaries

The optimality conditions for the LP can be written as the nonlinear system

$$F(x,\lambda,s) \;=\; \begin{pmatrix} A^T\lambda + s - c \\ Ax - b \\ XSe \end{pmatrix} \;=\; 0 \quad , \quad (x,s) \geq 0 \; . \tag{1.89}$$

Using the same nonlinear function F, the central path vector $(x_\tau, \lambda_\tau, s_\tau) \in \mathcal{C}$ turns out to be the solutions of the nonlinear system

$$F(x_\tau,\lambda_\tau,s_\tau) \;=\; \begin{pmatrix} 0 \\ 0 \\ \tau e \end{pmatrix} \;=\; 0 \quad , \quad (x_\tau, s_\tau) > 0 \; . \tag{1.90}$$

Most primal-dual algorithms take Newton steps toward points on $\mathcal{C}$. To describe the search direction, a centering parameter $\sigma \in [0,1]$ and a duality measure μ according to

$$\mu \;:=\; \frac{1}{n} \sum_{i=1}^{n} x_i s_i \tag{1.91}$$

are introduced. Then, the Newton-like algorithm is as follows:

Step 1: Initialization
Choose $(x^0, \lambda^0, s^0) \in \mathcal{F}^o_{PD}$.

Step 2: Iteration loop
For $k \geq 0$ compute

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{pmatrix} \begin{pmatrix} \Delta x^k \\ \Delta\lambda^k \\ \delta s^k \end{pmatrix} \;=\; \begin{pmatrix} 0 \\ 0 \\ -X^k S^k e + \sigma_k \mu_k e \end{pmatrix} \; , \tag{1.92}$$

where $\sigma_k \in [0,1]$ and $\mu_k := (x^k)^T s^k / n$, and set

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \\ s^{k+1} \end{pmatrix} \;=\; \begin{pmatrix} x^k \\ \lambda^k \\ s^k \end{pmatrix} \;+\; \alpha_k \begin{pmatrix} \Delta x^k \\ \Delta\lambda^k \\ \Delta s^k \end{pmatrix} \; , \tag{1.93}$$

where α_k is such that $(x^{k+1}, s^{k+1}) > 0$.

Path following algorithms restrict the iterates to a neighborhood of $\mathcal{C}$ and follow $\mathcal{C}$ to a solution of the LP. The two most common neighborhoods of $\mathcal{C}$ are the the so-called 2-norm neighborhood

$$\mathcal{N}_2(\theta) := \{(x,\lambda,s) \in \mathcal{F}^o_{PD} \mid \|XSe - \mu e\|_2 \leq \theta\mu\} \; , \; \theta \in (0,1) \; , \tag{1.94}$$

where $\|\cdot\|_2$ stands for the Euclidean norm, and the one-sided ∞-norm neighborhood

$$\mathcal{N}_{-\infty}(\gamma) := \{(x,\lambda,s) \in \mathcal{F}^o_{PD} \mid x_i s_i \geq \gamma\mu, 1 \leq i \leq n\} \ , \ \gamma \in (0,1). \tag{1.95}$$

In the sequel, we will investigate three classes of methods:

- short-step path following methods,
- Mizuno-Todd-Ye predictor-corrector methods,
- long-step path following methods.

1.3.2.2 Short-step path following methods

This method starts at a point $(x^0,\lambda^0,s^0) \in \mathcal{N}_2(\theta)$ and uses uniform values

$$\alpha_k = 1 \quad , \quad \sigma_k = \sigma \quad , \quad k \geq 0 \ ,$$

where θ and σ satisfy some relationship (see the theorem below).

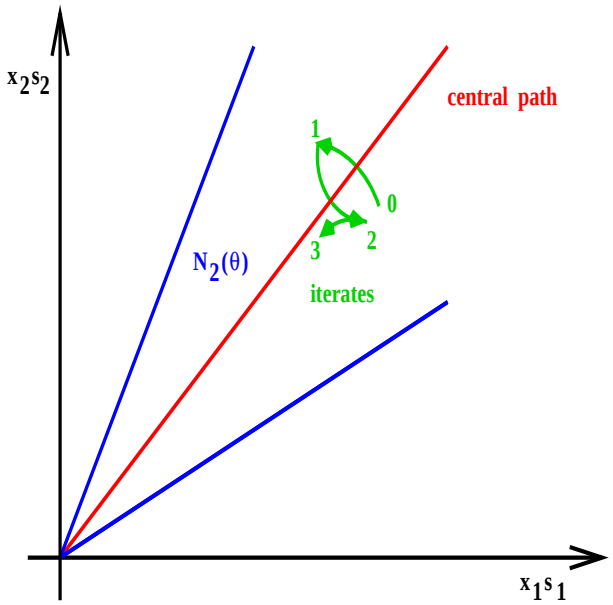

FIGURE 2. Short-step path following algorithm

Step 1: Initialization

Choose $(x^0,\lambda^0,s^0) \in \mathcal{F}^o_{PD}$ and set $\theta := 0.4$, $\sigma := 1 - 0.4/\sqrt{n}$.

Step 2: Iteration loop

For $k \geq 0$ set $\sigma_k = \sigma$ and compute

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{pmatrix} \begin{pmatrix} \Delta x^k \\ \Delta\lambda^k \\ \delta s^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -X^k S^k e + \sigma_k \mu_k e \end{pmatrix} \ , \tag{1.96}$$

where $\mu_k := (x^k)^T s^k/n$. Set

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \\ s^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \\ s^k \end{pmatrix} + \begin{pmatrix} \Delta x^k \\ \Delta \lambda^k \\ \Delta s^k \end{pmatrix} . \tag{1.97}$$

Figure 2 contains the first iterates of the algorithm. The horizontal and the vertical coordinate axes stand for the $x_1 s_1$ and the $x_2 s_2$ product, respectively. The central path is the line emanating from $(0, 0)$ at an angle of $\pi/4$. The search direction appear to be curves rather than straight lines. The solution is at $(0, 0)$ and the problem is to reach that point maintaining the feasibility conditions

$$Ax = b \quad , \quad A^T \lambda + s = c \, .$$

The choice of θ and σ is motivated by the following result:

Theorem 1.7 (Properties of the short-step path following algorithm) Let $\theta \in (0, 1)$ and $\sigma \in (0, 1)$ be given such that

$$\frac{\theta^2 + n(1-\sigma)^2}{2^{3/2}(1-\theta)} \leq \sigma \, \theta \, . \tag{1.98}$$

Then

$$(x, \lambda, s) \in \mathcal{N}_2(\theta) \implies (x(\alpha), \lambda(\alpha), s(\alpha)) \in \mathcal{N}_2(\theta) \, , \; \alpha \in [0, 1] \, . \tag{1.99}$$

Proof: For a proof we refer to [5].

1.3.2.3 Mizuno-Todd-Ye predictor-corrector methods

Predictor-corrector methods consist of predictor steps with $\sigma_k = 0$ to reduce the duality measure μ and corrector step with $\sigma_k = 1$ to improve centrality. They work with an inner neighborhood $\mathcal{N}_2(0.25)$ and an outer neighborhood $\mathcal{N}_2(0.5)$ such that even-index iterates are confined to the inner neighborhood, whereas odd-index iterates stay in the outer neighborhood.

Step 1: Initialization
Choose $(x^0, \lambda^0, s^0) \in \mathcal{N}_2(0.25)$.

Step 2: Iteration loop
For $k \geq 0$ do:
Predictor step: If k is even, set $\sigma_k = 0$ and solve

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{pmatrix} \begin{pmatrix} \Delta x^k \\ \Delta \lambda^k \\ \delta s^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -X^k S^k e \end{pmatrix} . \tag{1.100}$$

Choose α_k as the largest value of $\alpha \in [0, 1]$ such that

$$(x^k + \alpha \Delta x^k, \lambda^k + \alpha \Delta \lambda^k, s^k + \alpha \Delta s^k) \in \mathcal{N}_2(0.5) \, . \tag{1.101}$$

Set

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \\ s^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \\ s^k \end{pmatrix} + \alpha_k \begin{pmatrix} \Delta x^k \\ \Delta\lambda^k \\ \Delta s^k \end{pmatrix} . \tag{1.102}$$

Corrector step: If k is odd, set $\sigma_k = 1$ and solve

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{pmatrix} \begin{pmatrix} \Delta x^k \\ \Delta\lambda^k \\ \delta s^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -X^k S^k e + \mu_k e \end{pmatrix} , \tag{1.103}$$

where $\mu_k := (x^k)^T s^k/n$ and set

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \\ s^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \\ s^k \end{pmatrix} + \begin{pmatrix} \Delta x^k \\ \Delta\lambda^k \\ \Delta s^k \end{pmatrix} . \tag{1.104}$$

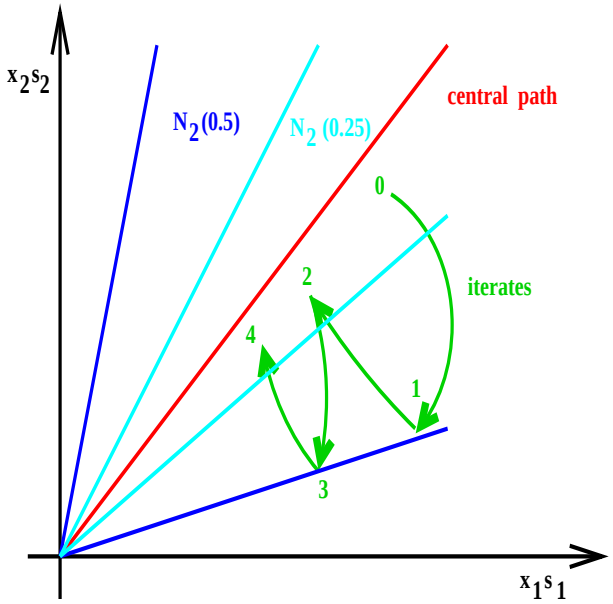

FIGURE 3. Path following predictor-corrector algorithm

Figure 3 displays the path following predictor-corrector algorithm. We start from (x^0, λ^0, s^0) in the inner neighborhood $\mathcal{N}_2(0.25)$ and use a predictor step with $\sigma_0 = 0$ to arrive at the new iterate (x^1, λ^1, s^1) at the boundary of the outer neighborhood $\mathcal{N}_2(0.5)$. The corrector step with $\sigma_1 = 1$ and unit step $\alpha = 1$ leads us back into the inner neighborhood $\mathcal{N}_2(0.25)$. This cycle then repeats.

The predictor steps achieve a substantial reduction of the duality measure:

Lemma 1.2 (Properties of the predictor step) Assume $(x, \lambda, s) \in \mathcal{N}_2(0.25)$ and that $(\Delta x, \Delta\lambda, \Delta s)$ has been computed as described in the predictor step (i.e., with $\sigma = 0$). Then, there holds

$$(x + \alpha\Delta x, \lambda + \alpha\Delta\lambda, s + \alpha\Delta s) \in \mathcal{N}_2(0.5) \quad \text{for all } \alpha \in [0, \overline{\alpha}] \ , \tag{1.105}$$

where

$$\overline{\alpha} \ := \ \min \ (\frac{1}{2}, (\frac{\mu}{8\|\Delta X \Delta S e\|})^{1/2}) \ . \tag{1.106}$$

For the new value μ^{new} of the duality measure we have

$$\mu^{new} \ \leq \ (1 - \overline{\alpha}) \ \mu \ . \tag{1.107}$$

Proof: We refer to [5].

Corrector steps return to the inner neighborhood without changing the duality measure:

Lemma 1.3 (Properties of the corrector step) Assume $(x, \lambda, s) \in \mathcal{N}_2(0.5)$ and that $(\Delta x, \Delta\lambda, \Delta s)$ has been computed by the corrector step with $\sigma = 1$ and $\alpha = 1$. Then, we have

$$(x + \Delta x, \lambda + \Delta\lambda, s + \Delta s) \in \mathcal{N}_2(0.25) \quad , \quad \mu^{new} = \mu \ . \tag{1.108}$$

Proof: We refer to [5].

1.3.2.4 Long-step path following methods

The long-step path following method generates iterates in the neighborhood $\mathcal{N}_{-\infty}(\gamma)$ which for small γ contains most of the strictly feasible points. At each step, the centering parameter σ stays between two fixed limits $0 < \sigma_{min} < \sigma_{max} < 1$ and the step length α_k of the search direction is chosen as large as possible, provided the new iterate stays in $\mathcal{N}_{-\infty}(\gamma)$.

Step 1: Initialization

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

Step 2: Iteration loop

For $k \geq 0$ choose $\sigma_k \in [\sigma_{min}, \sigma_{max}]$ and compute

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{pmatrix} \begin{pmatrix} \Delta x^k \\ \Delta\lambda^k \\ \delta s^k \end{pmatrix} \ = \ \begin{pmatrix} 0 \\ 0 \\ -X^k S^k e + \sigma_k \mu_k e \end{pmatrix} \ , \tag{1.109}$$

where $\mu_k := (x^k)^T s^k / n$. Choose α_k as the largest value such that

$$(x^k + \alpha\Delta x^k, \lambda^k + \alpha\Delta\lambda^k, s^k + \alpha\Delta s^k) \in \mathcal{N}_{-\infty}(\gamma) \tag{1.110}$$

and set

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \\ s^{k+1} \end{pmatrix} \ = \ \begin{pmatrix} x^k \\ \lambda^k \\ s^k \end{pmatrix} \ + \ \alpha_k \ \begin{pmatrix} \Delta x^k \\ \Delta\lambda^k \\ \Delta s^k \end{pmatrix} \ . \tag{1.111}$$

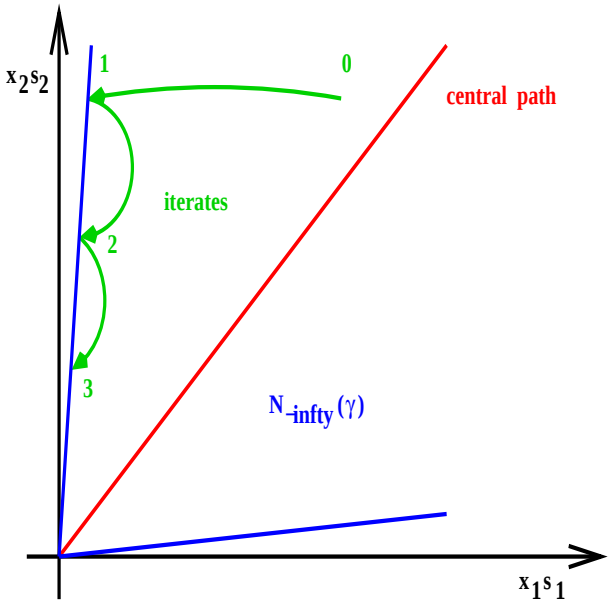

FIGURE 4. Long-step path following algorithm

The lower bound σ_{min} on the centering parameter guarantees that the search directions start out by moving off the boundary of $\mathcal{N}_{-\infty}(\gamma)$ and into its interior: Small steps would improve the centrality, whereas large steps lead outside the neighborhood. The step size selection α_k ensures that we stay at least at the boundary.

Lemma 1.4 (properties of the long-step path following algorithm) Fore given $\gamma \in (0,1)$ and $0 < \sigma_{min} < \sigma_{max} < 1$ there exists $\delta < n$, independent of n, such that

$$\mu_{k+1} \leq (1 - \frac{\delta}{n})\, \mu_k \quad , \quad k \geq 0 \; . \tag{1.112}$$

Proof: We refer to [5].

1.3.2.5 Convergence of the path following algorithms

As far as the convergence of the sequence of iterates of the three previously introduced path following primal-dual interior-point methods is concerned, we have the following result:

Theorem 1.8 (Convergence of iterates of path following methods) Assume that $\{(x^k, \lambda^k, s^k)\}_{k \in \mathbb{N}_0}$ is a sequence of iterates generated either by the short-step resp. long-step path following method or by the predictor-corrector path-following algorithm and suppose that the sequence $\{\mu_k\}_{k \in \mathbb{N}_0}$ of duality measures is going to zero as $k \to \infty$. Then, the sequence $\{(x^k, s^k)\}_{k \in \mathbb{N}_0}$ is bounded and thus contains a convergent subsequence. Each limit point is a strictly complementary primal-dual solution.

Proof: We refer to [5].

1.3.3 Mehrotra's predictor-corrector algorithm

In contrast to the algorithms treated in the previous subsection, Mehrotra's algorithm produces a sequence of infeasible iterates (x^k, λ^k, s^k) with $(x^k, s^k) > 0$. Each iteration step involves the following three components

- an 'affine-scaling' prediction step which is the Newton direction for the nonlinear function F as defined by (1.90),
- a centering term by means of an adaptively chosen centering parameter σ,
- a corrector step that compensates for some nonlinearity in the predictor step direction.

Predictor step: Given (x, λ, s) with $(x, s) > 0$, the affine scaling direction $(\Delta x^{aff}, \Delta\lambda^{aff}, \Delta s^{aff})$ is obtained by the solution of the system

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x^{aff} \\ \Delta\lambda^{aff} \\ \Delta s^{aff} \end{pmatrix} = \begin{pmatrix} -r_c \\ -r_b \\ -XSe \end{pmatrix} , \tag{1.113}$$

where r_b and r_c stand for the residuals

$$r_b := Ax - b \quad , \quad r_c := A^T\lambda + s - c \; . \tag{1.114}$$

The step lengths are chosen separately for the primal and dual components

$$\alpha^p_{aff} := \text{argmax } \{\alpha \in [0,1] \mid x + \alpha\Delta x^{aff} \geq 0\} \; , \tag{1.115}$$

$$\alpha^d_{aff} := \text{argmax } \{\alpha \in [0,1] \mid s + \alpha\Delta s^{aff} \geq 0\} \; . \tag{1.116}$$

Adaptive choice of the centering parameter: If we perform a full step to the boundary in the affine-scaling direction, the resulting value of the duality measure would be

$$\mu_{aff} = (x + \alpha^p_{aff}\Delta x^{aff})^T(s + \alpha^d_{aff}\Delta s^{aff})/n \; .$$

Mehrotra has suggested a heuristics for the choice of the centering parameter

$$\sigma = (\frac{\mu_{aff}}{\mu})^3 \; , \tag{1.117}$$

which can be motivated as follows:
If $\mu_{aff} \ll \mu$, the affine-scaling direction is such that it leads to a significant reduction of the duality measure. Consequently, the centering parameter σ should be chosen close to zero. On the other hand, if μ_{aff} is only a bit smaller than μ, the trajectory should lead closer to the central path $\mathcal{C}$ which can be realized by choosing the centering parameter σ closer to 1. In order to compute the centering step direction, we would have to solve

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x^c \\ \Delta\lambda^c \\ \Delta s^c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \sigma\mu e \end{pmatrix} \; . \tag{1.118}$$

Instead of doing so, we will combine the centering step with the corrector step.

Corrector step: The impact of a full step in the affine-scaling direction on the pairwise products $x_i s_i, 1 \leq i \leq n$, is as follows

$$(1.119)\quad \begin{aligned}(x_i + \Delta x_i^{aff})(s_i + \Delta s_i^{aff}) \;&=\\ &= \underbrace{x_i s_i + x_i \Delta s_i^{aff} + s_i \Delta x_i^{aff}}_{=\,0} + \Delta x_i^{aff} \Delta s_i^{aff} \;=\; \Delta x_i^{aff} \Delta s_i^{aff}\;,\end{aligned}$$

where we have used that the sum of the first three terms is zero due to (1.113). The corrector step is designed in such a way that the pairwise products $x_i s_i$ come closer to the target value of zero:

$$(1.120)\qquad \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x^{cor} \\ \Delta \lambda^{cor} \\ \Delta s^{cor} \end{pmatrix} \;=\; \begin{pmatrix} 0 \\ 0 \\ -\Delta X^{aff} \Delta S^{aff} e \end{pmatrix}\;,$$

where

$$\begin{aligned}\Delta X^{aff} \;&=\; \mathrm{diag}(\Delta x_1^{aff}, ..., \Delta x_n^{aff})\;,\\ \Delta S^{aff} \;&=\; \mathrm{diag}(\Delta s_1^{aff}, ..., \Delta s_n^{aff})\;.\end{aligned}$$

Now, it is an easy exercise to show that (1.119) and (1.120) imply

$$(1.121)\qquad \begin{aligned}&(x_i + \Delta x_i^{aff} + \Delta x_i^{cor})(s_i + \Delta s_i^{aff} + \Delta s_i^{cor}) \;=\\ &=\; \Delta x_i^{aff} \Delta s_i^{cor} + \Delta x_i^{cor} \Delta s_i^{aff} + \Delta x_i^{cor} \Delta s_i^{cor}\;.\end{aligned}$$

If for $\mu \to 0$ the coefficient matrix in (1.113) resp. (1.120) is approaching a nonsingular limit, we indeed have

$$\|(\Delta x^{aff}, \delta s^{aff})\| \;=\; O(\mu) \quad , \quad \|(\Delta x^{cor}, \delta s^{cor})\| \;=\; O(\mu^2)\;,$$

which implies

$$\begin{aligned}\Delta x_i^{aff} \Delta s_i^{aff} \;&=\; O(\mu^2)\;,\\ \Delta x_i^{aff} \Delta s_i^{aff} \Delta x_i^{aff} \Delta s_i^{cor} + \Delta x_i^{cor} \Delta s_i^{aff} + \Delta x_i^{cor} \Delta s_i^{cor} \;&=\; O(\mu^3)\;.\end{aligned}$$

However, if the limiting matrix is singular, it is not guaranteed that the corrector step is smaller in norm than the predictor step (in fact, often it is larger). Nevertheless, numerical evidence suggests that also in this case the corrector step improves the overall performance of the algorithm.

Combining the centering and the corrector step amounts to the solution of the linear system

$$(1.122)\qquad \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x^{cc} \\ \Delta \lambda^{cc} \\ \Delta s^{cc} \end{pmatrix} \;=\; \begin{pmatrix} 0 \\ 0 \\ \sigma \mu e - \Delta X^{aff} \Delta S^{aff} e \end{pmatrix}\;.$$

A commonly used variant of Mehrotra's predictor-corrector step is given as follows:

Step 1: Initialization

Choose (x^0, λ^0, s^0) with $(x^0, s^0) > 0$.

Step 2: Iteration loop

For $k \geq 0$ set

$$(x, \lambda, s) \;=\; (x^k, \lambda^k, s^k)$$

and solve (1.113) for $(\Delta x^{aff}, \Delta\lambda^{aff}, \Delta s^{aff})$. Compute

$$\alpha^p_{aff} \;:=\; \text{argmax}\,\{\alpha \in [0,1] \mid x^k + \alpha\Delta x^{aff} \geq 0\}\;, \tag{1.123}$$

$$\alpha^d_{aff} \;:=\; \text{argmax}\,\{\alpha \in [0,1] \mid s^k + \alpha\Delta s^{aff} \geq 0\}\;, \tag{1.124}$$

$$\mu_{aff} \;:=\; (x^k + \alpha^p_{aff}\Delta x^{aff})^T(s^k + \alpha^d_{aff}\Delta s^{aff})/n\;, \tag{1.125}$$

$$\sigma \;:=\; (\mu_{aff}/\mu)^3\;. \tag{1.126}$$

Solve (1.122) for $(\Delta x^{cc}, \Delta\lambda^{cc}, \Delta s^{cc})$ and compute the search direction and the step length to the boundary according to

$$\begin{aligned}(\Delta x^k, \Delta\lambda^k, \Delta s^k) \;&:=\; (\Delta x^{aff}, \Delta\lambda^{aff}, \Delta s^{aff}) \;+ \\ &\quad + (\Delta x^{cc}, \Delta\lambda^{cc}, \Delta s^{cc})\;,\end{aligned} \tag{1.127}$$

$$\alpha^p_{max} \;:=\; \text{argmax}\,\{\alpha \geq 0 \mid x^k + \alpha\Delta x^k \geq 0\}\;, \tag{1.128}$$

$$\alpha^d_{max} \;:=\; \text{argmax}\,\{\alpha \geq 0 \mid s^k + \alpha\Delta s^k \geq 0\}\;. \tag{1.129}$$

Set

$$\alpha^p_k \;:=\; \min(0.99 \cdot \alpha^p_{max}, 1) \quad , \quad \alpha^d_k \;:=\; \min(0.99 \cdot \alpha^d_{max}, 1)\;,$$

and compute

$$x^{k+1} \;:=\; x^k \;+\; \alpha^p_k \delta x^k\;, \tag{1.130}$$

$$(\lambda^{k+1}, s^{k+1}) \;:=\; (\lambda^k, s^k) \;+\; \alpha^d_k(\delta\lambda^k, \Delta s^k)\;. \tag{1.131}$$

References

[1] G.B. Dantzig; Linear Programming and Extensions. Princeton Univ. Press, Princeton, 1963

[2] R. Fletcher; Practical Methods of Optimization. Wiley, New York, 1987

[3] O.L. Mangarasian; Nonlinear Programming. McGraw-Hill, New York, 1969

[4] J. Stoer and R. Bulirsch; Introduction to Numerical Analysis. 3rd Edition. Springer, Berlin-Heidelberg-New York, 2002

[5] S.J. Wright; Primal-Dual Interior-Point Methods. SIAM, Philadelphia, 1997