26:711:564 **OPTIMIZATION MODELS IN FINANCE**
Professor Andrzej Ruszczyński

Linear Programming

1 Standard Form

A linear programming problem is an optimization problem in which the objective function and all constraint functions are *affine* functions, that is, sums of linear functions and constants. Recall that every linear function $f(x)$ on $\mathbb{R}^n$ must have the form:

$$f(x) = \langle c, x \rangle = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n,$$

where $c \in \mathbb{R}^n$ is a vector of constant coefficients. The objective function and the constraint functions must be formed of expressions of such a form plus, perhaps, some constants.

Every linear programming problem can be transformed to an equivalent problem in a *standard form*:

$$\begin{aligned} \min\ & \langle c, x \rangle \\ \text{subject to } & Ax = b, \\ & x \geq 0. \end{aligned} \tag{1}$$

with a matrix A of dimension $m \times n$, $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$. The number of variables n may change in the transformation.

This can be achieved via the following transformations.

1. If the original problem is a maximization problem

   $$\max_{x \in X}\ c_0 + c_1 x_1 + c_2 x_2 + \cdots + c_n x_n,$$

   it is equivalent to the minimization problem having the sign of the function reversed, and with the constant skipped:

   $$\min_{x \in X}\ -c_1 x_1 - c_2 x_2 - \cdots - c_n x_n.$$

2. If the original problem contains an inequality

   $$a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{in} x_n \leq b_i,$$

   we can add one variable (called the *slack variable*) s_i, and equivalently express this inequality as a system of an equation and an inequality on the sign of s_i:

   $$a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{in} x_n + s_i = b_i \qquad \text{and} \qquad s_i \geq 0.$$

3. If the original problem contains an inequality

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \geq b_i,$$

   we can add one slack variable s_i and equivalently express this inequality as a system of an equation and an inequality on the sign of s_i:

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - s_i = b_i \qquad \text{and} \qquad s_i \geq 0.$$

4. If the original problem contains a variable x_j, which is not restricted to be nonnegative, we can represent x_j as a difference of two nonnegative variables:

$$x_j = y_j^1 - y_j^2, \qquad y_j^1 \geq 0, \qquad y_j^2 \geq 0.$$

As a result of these transformations, the number of decision variables increases, but the resulting problem is equivalent to the original problem in the sense that there is an obvious way to read the solution of the original problem from the solution of its equivalent problem in the standard form.

Modern modeling software does not require the users to specify linear programming problems in their standard forms; this is done automatically by the program. For theoretical purposes, the standard form is useful to develop a unified theory of optimality conditions and methods for solving linear programming problems. Due to the equivalence of every linear programming problem to its standard form, these theoretical results apply to all linear programming problems.

2 Direct Derivation of Optimality Conditions

Consider a linear programming problem in the standard form (1). The feasible set X of this problem is convex and an optimal solution, if it exists, is a basic feasible solution, an extreme point of X (see the notes on convexity). Our intention is to characterize the optimal solutions algebraically.

Assume for simplicity of presentation that the rank of A is equal to its number of rows m. Clearly, $m \leq n$. If $\hat{x}$ is a basic feasible solution, then the columns of A corresponding to the positive coordinates $\hat{x}_j$ of $\hat{x}$ are linearly independent. If their number is m, then they form a square nonsingular matrix B. If their number k is smaller than m, then we can add to them $m - k$ columns of A to create a square nonsingular matrix B.

After rearranging the columns of A and after the corresponding rearrangement of x, so that the selected m columns of A are first, we can write the system of equations in (1) as follows

$$\begin{bmatrix} B & N \end{bmatrix} \begin{bmatrix} x_B \\ x_N \end{bmatrix} = b.$$

This can be written explicitly as

$$Bx_B + Nx_N = b. \tag{2}$$

The variables x_B are called *basic variables*, while x_N are *nonbasic variables*.

System (2) is satisfied by the current point

$$\hat{x} = \begin{bmatrix} \hat{x}_B \\ \hat{x}_N \end{bmatrix}, \qquad \hat{x}_B = B^{-1}b, \qquad \hat{x}_N = 0.$$

Let us consider x_N in system (2) as independent variables, and x_B as dependent variables. As B is nonsingular, we can write

$$x_B = B^{-1}(b - Nx_N). \tag{3}$$

We also split c in the same way:

$$c = \begin{bmatrix} c_B \\ c_N \end{bmatrix}.$$

Let us calculate the difference of the objective function values at x and $\hat{x}$. We have

$$\begin{aligned} c^T x - c^T \hat{x} &= (c_B)^T x_B + (c_N)^T x_N - (c_B)^T \hat{x}_B + (c_N)^T \hat{x}_N \\ &= (c_B)^T B^{-1}(b - Nx_N) + (c_N)^T x_N - (c_B)^T B^{-1} b \\ &= \Big((c_N)^T - (c_B)^T B^{-1} N \Big) x_N. \end{aligned} \tag{4}$$

Denote

$$\pi = (B^T)^{-1} c_B, \qquad \bar{c} = c - A^T \pi = \begin{bmatrix} 0 \\ c_N - N^T \pi \end{bmatrix}$$

The vector π is called the vector of *multipliers*, while the vector $\bar{c}$ is the vector od *reduced costs*. We can rewrite (4) as follows:

$$c^T x - c^T \hat{x} = (\bar{c}_N)^T x_N.$$

Observe that $x_N \geq 0$. If

$$\bar{c}_N \geq 0, \tag{5}$$

then no other feasible solution of the problem can be better than $\hat{x}$, that is, $\hat{x}$ is optimal.

If $\bar{c}_j < 0$ for some nonbasic x_j, we can increase x_j, while keeping the other nonbasic variables at 0, and while adjusting the values of basic variables according to (3). In this way we can move to a better basic feasible solution. This is the idea of the simplex method.

3 Applying Farkas' Lemma

We can derive the optimality conditions from more general abstract considerations. For a point $\hat{x}$ we can define the set of *feasible directions*

$$D = \{d \in \mathbb{R}^n : d_j \geq 0 \text{ if } \hat{x}_j = 0,\ Ad = 0\}.$$

We can represent D as the set of solutions of the following system of linear inequalities

$$Ad \geq 0 \tag{6}$$

$$-Ad \geq 0 \tag{7}$$

$$d_j \geq 0,\ \ j \in J. \tag{8}$$

The point $\hat{x}$ is optimal if and only if every solution d of system (6)–(8) satisfies the inequality

$$c^T d \geq 0. \tag{9}$$

By Farkas' lemma, this is possible if and only if c in (9) is a positive combination of the rows of inequalities (6), (7), and (8). This means that there exist vectors $\mu \geq 0$, $\nu \geq 0$, and $\kappa \geq 0$, with $\kappa_j = 0$, $j \notin J$, such that

$$c = A^T \mu - A^T \nu + \kappa.$$

Denote $\lambda = \mu - \nu$. We conclude that

$$c - A^T \lambda \geq 0.$$

The vector λ is called the vector of *Lagrange multipliers*. The multipliers π calculated in the previous section become Lagrange multipliers at the optimal solution of the problem.