Optimization Techniques in Finance

3. Linear programming

Andrew Lesniewski

Baruch College
New York

Fall 2019

Outline

Example: graphical method

- *Linear programming* (LP) is concerned with optimization problems where the objective function and the constraint functions are linear. The constraints may be equalities or inequalities.
- *Example 1*. Consider the following problem with two *decision variables* (unknowns) and five (inequality) constraints

$$\min f(x) = x_1 + x_2, \quad \text{subject to} \begin{cases} x_1 - x_2 \geq -4, \\ 2x_1 + x_2 \leq 8, \\ x_i \geq 0, \text{ for } i = 1, 2. \end{cases}$$

- This problem is easy to solve using a *graphical method*.
- The feasible set of this problem is the quadrilateral $\mathcal{Q}$ with vertices at $(0, 0)$, $(0, 4)$, $(4/3, 16/3)$, and $(4, 0)$. Since the objective function is linear, its minimum can only be achieved at a vertex of $\mathcal{Q}$.
- Sliding the objective function through $\mathcal{Q}$ we see that its minimum is achieved at $x^* = (0, 0)$, where its value is $f(x^*) = 0$.

Example: the diet problem

- *Example 2*. There are n types of nutrients contained in m different types of food. How should a person's daily diet be structured so that he gets the required amount of each nutrient at the minimum cost?
- Let c_j, $j = 1, \ldots, m$ denote the unit price of food j, let b_i, $i = 1, \ldots, n$ denote the daily minimum required amount of nutrient i, let a_{ij} denote the content of nutrient i in food j, and let x_j denote the daily amount of food j.
- This leads to the following optimization problem:

$$\min f(x) = \sum_{j=1}^{m} c_j x_j, \quad \text{subject to} \begin{cases} \sum_{j=1}^{m} a_{ij} x_j \geq b_i, & \text{for } i = 1, \ldots, n, \\ x_j \geq 0, & \text{for } j = 1, \ldots, m. \end{cases}$$

Example: the optimal assignment problem

- *Example 3*. There are p people available to carry out q tasks. How should the tasks be assigned so that the total value is maximized?
- Let c_{ij} denote the daily value of person $i = 1, \ldots, p$ carrying out task $j = 1 \ldots, q$, and let x_{ij} denote the fraction of person's i workday to spend on task j.
- This leads to the following optimization problem:

$$\max f(x) = \sum_{i=1}^{p} \sum_{j=1}^{q} c_{ij} x_{ij}, \text{ subject to } \begin{cases} \sum_{j=1}^{q} x_{ij} \leq 1, & \text{for } i = 1, \ldots, p, \\ \sum_{i=1}^{p} x_{ij} \leq 1, & \text{for } j = 1, \ldots, q, \\ x_{ij} \geq 0, & \text{for } i = 1, \ldots, p, j = 1, \ldots, q. \end{cases}$$

- The first constraint guarantees that no one works more than a day, while the second condition means that no more than one person is assigned to a task.

Example: the transportation problem

- *Example 4*. There are p factories that supply a certain product, and there are q sellers to which this product is shipped. How should the market demand be met at a minimum cost?
- Assume that factory $i = 1, \ldots, p$ can supply s_i units of the product, while seller $j = 1, \ldots, q$ requires at least d_j units of the product. Let c_{ij} be the cost of transporting one unit of the product from factory i to seller j, and let x_{ij} denote the number of units shipped from factory i to seller j.
- This leads to the following optimization problem:

$$\min f(x) = \sum_{i=1}^{p} \sum_{j=1}^{q} c_{ij} x_{ij}, \text{ subject to } \begin{cases} \sum_{j=1}^{q} x_{ij} \leq s_i, & \text{for } i = 1, \ldots, p, \\ \sum_{i=1}^{p} x_{ij} \geq d_j, & \text{for } j = 1, \ldots, q, \\ x_{ij} \geq 0, & \text{for } i = 1, \ldots, p, j = 1, \ldots, q. \end{cases}$$

Example

- *Example 5*. Consider the following set of constraints in $\mathbb{R}^2$:

$$
\begin{aligned}
-x_1 + x_2 &\leq 1,\\
x_1, x_2 &\geq 0.
\end{aligned}
$$

  These inequalities define an *unbounded* subset of $\mathbb{R}^2$.

- LP problems with these constraints may show very different outcomes:
  (i) $\min f(x) = x_1 + x_2$ has a *unique solution* at $x = (0, 0)$.
  (ii) $\min f(x) = x_1$ has *infinitely many solutions* at $x = (0, t)$, $0 \leq t \leq 1$. These solutions are bounded.
  (iii) $\min f(x) = x_2$ has *infinitely many solutions* at $x = (t, 0)$, $t \geq 0$. These solutions are unbounded.
  (iv) For $\min f(x) = -x_1 - x_2$ *no feasible point is optimal* and the optimal cost is $-\infty$.

Examples

- Two dimensional LP problems with a moderate number of constraints can be solved using a graphical method.
- LP problems (especially if the number of decision variables and constraints are large) are notoriously difficult to solve.
- A breakthrough occurred in 1947, when George Dantzig developed the *simplex method*, which proved to be a very effective tool for solving linear programming problems.
- Another major breakthrough occurred in 1984, when Narendra Karmarkar developed an interior point method bearing his name. It is an efficient, polynomial runtime algorithm.

Standard form

- In general, we will be concerned with linear optimization problems of the form

$$\min f(x) = c^{\mathrm{T}} x, \quad \text{subject to } \begin{cases} a_i^{\mathrm{T}} x = b_i, & \text{for } i \in \mathcal{E}, \\ a_i^{\mathrm{T}} x_i \geq b_i, & \text{for } i \in \mathcal{I}, \end{cases} \tag{1}$$

  where $c \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$, and $b_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, are constant vectors.
- Note that there may be no constraints on some of the variables x_i, in which case their values are *unrestricted*.
- Recall that a *feasible solution* to the problem (1) (or *feasible vector*) is any x satisfying all the constraints. The feasible set is the the set of all feasible solutions.
- We will be assuming that the feasible set is nonempty.

Standard form

- Any LP problem can be written in the *standard form*, namely

$$\min c^{\mathrm{T}} x, \quad \text{subject to} \begin{cases} Ax = b, \\ x_i \geq 0, \text{ for } i = 1, \ldots, n. \end{cases} \tag{2}$$

  where $c \in \mathbb{R}^n$, $A \in \mathrm{Mat}_{mn}(\mathbb{R})$, and $b \in \mathbb{R}^m$.
- This form is convenient for describing solution algorithms for LP problems, and (unless otherwise stated) we will be assuming it in the following.
- Every problem of the form (1) can be transformed into the form (2).

Standard form

- This can be accomplished by means of two types of operations:
  (i) Elimination of inequality constraints: given an inequality of the form

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i,$$

we introduce a *slack variable* s_i, and the standard constraint:

$$\sum_{j=1}^{n} a_{ij} x_j + s_i = b_i,$$
$$s_i \ge 0.$$

  (ii) Elimination of free variables: if x_i is an unrestricted variable, we replace it by $x_i = x_i^+ - x_i^-$, with $x_i^+, x_i^- \ge 0$.

Example

- *Example 6*. Consider the following LP problem written in the non-standard form:

$$\min -x_1 - x_2, \quad \text{subject to} \begin{cases} 2x_1 + x_2 \leq 12, \\ x_1 + 2x_2 \leq 9, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

- Introducing slack variables x_3 and x_4, this problem can be written in the standard form:

$$\min -x_1 - x_2, \quad \text{subject to} \begin{cases} 2x_1 + x_2 + x_3 = 12, \\ x_1 + 2x_2 + x_4 = 9, \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0. \end{cases}$$

Example

- *Example 7.* Consider the following LP problem written in the non-standard form:

$$\min x_2, \quad \text{subject to} \begin{cases} x_1 + x_2 \geq 1, \\ x_1 - x_2 \leq 0. \end{cases}$$

  Notice that x_1 and x_2 are unrestricted.
- Writing $x_i = x_i^+ - x_i^-$, $i = 1, 2$, and introducing slack variables s_1 and s_2, this problem can be written in the standard form:

$$\min x_2^+ - x_2^-, \text{ subject to} \begin{cases} x_1^+ - x_1^- + x_2^+ - x_2^- - s_1 = 1, \\ x_1^+ - x_1^- - x_2^+ + x_2^- + s_2 = 0, \\ x_1^+ \geq 0, x_1^- \geq 0, x_2^+ \geq 0, x_2^- \geq 0, s_1 \geq 0, s_2 \geq 0. \end{cases}$$

Lower bound for the solution

- Let us go back to Example 6 written in the standard form, and consider the following few feasible solutions:

$$x = (0, 9/2, 15/2, 0), \quad f(x) = -9/2,$$
$$x = (6, 0, 0, 3), \quad f(x) = -6,$$
$$x = (5, 2, 0, 0), \quad f(x) = -7.$$

  Is the last solution optimal?
- One way to approach this question is to establish a lower bound on the objective function over the feasible set.
- For example, using the first constraint, we find that

$$-x_1 - x_2 \geq -2x_1 - x_2 - x_3 - 12.$$

Lower bound for the solution

- The second constraint helps a little more

$$
\begin{aligned}
-x_1 - x_2 &\geq -x_1 - 2x_2 - x_4 \\
&= -9.
\end{aligned}
$$

- An even tighter bound is obtained if we add both constraints multiplied by $-1/3$,

$$
\begin{aligned}
-x_1 - x_2 &\geq -x_1 - x_2 - \frac{1}{3}\,x_3 - \frac{1}{3}\,x_4 \\
&= -\frac{1}{3}\,(2x_1 + x_2 + x_3) - \frac{1}{3}\,(x_1 + 2x_2 + x_4) \\
&= -7.
\end{aligned}
$$

- The last lower bound means that $f(x) \geq -7$ for any feasible solution.
- Since we have already found a feasible solution saturating this bound, namely $x = (5, 2, 0, 0)$, it means that this x is an optimal solution to the problem.

Lower bound for the solution

- Let us formalize this process.
- Consider any linear combination of the two main constraints with coefficients y_1 and y_2:

$$y_1(2x_1+x_2+x_3)+y_2(x_1+2x_2+x_4) = (2y_1+y_2)x_1+(y_1+2y_2)x_2+y_1x_3+y_2x_4.$$

- Since $x_3, x_4 \geq 0$, this expression will provide a lower bound on the objective function $-x_1 - x_2$, if y_1 and y_2 satisfy the following conditions:

$$\begin{aligned} 2y_1 + y_2 &\leq -1, \\ y_1 + 2y_2 &\leq -1, \\ y_1, y_2 &\leq 0. \end{aligned}$$

- Indeed, in this case,

$$(2y_1 + y_2)x_1 + (y_1 + 2y_2)x_2 + y_1x_3 + y_2x_4 \leq -x_1 - x_2.$$

Lower bound for the solution

- In other words, we showed that

$$12y_1 + 9y_2 \leq -x_1 - x_2.$$

- This motivates the following idea.
- In order to obtain the largest possible lower bound, we should maximize the corresponding linear combination of the right hand sides of the constraints of the original problem, namely $12y_1 + 9y_2$, provided that y_1, y_2 satisfy the constraints introduced above.
- This leads us to the following *dual problem*:

$$\max 12y_1 + 9y_2, \quad \text{subject to} \begin{cases} 2y_1 + y_2 \leq -1, \\ y_1 + 2y_2 \leq -1, \\ y_1, y_2 \leq 0. \end{cases}$$

- This can be generalized to any LP problem in standard form.

Duality

- For a general problem (2) (called the *primal problem*), we consider the corresponding *dual problem*:

$$\max b^{\mathrm{T}} y, \quad \text{subject to } A^{\mathrm{T}} y \leq c. \tag{3}$$

- Introducing the slack variables s, we can state it in the standard form:

$$\max b^{\mathrm{T}} y, \quad \text{subject to } \begin{cases} A^{\mathrm{T}} y + s = c, \\ s_i \geq 0, \text{ for } i = 1, \ldots, n. \end{cases} \tag{4}$$

- As expected from the motivational example above, there is a relationship between the solutions of the primal and dual problems.
- The objective function of a feasible solution to the primal problem is bounded from below by the objective function of any feasible solution to the dual problem.

Duality

- *Weak duality theorem.* Let x be a feasible solution to the primal problem, and let y be a feasible solution to the dual problem. Then

$$c^{\mathrm{T}}x \geq b^{\mathrm{T}}y. \tag{5}$$

- For the proof, observe that, since $x \geq 0$, and $c - A^{\mathrm{T}}y \geq 0$ component-wise, the inner product of these vector must be nonnegative:

$$\begin{aligned} 0 &\leq (c - A^{\mathrm{T}}y)^{\mathrm{T}}x \\ &= c^{\mathrm{T}}x - y^{\mathrm{T}}Ax \\ &= c^{\mathrm{T}}x - y^{\mathrm{T}}b. \end{aligned}$$

Duality

- The quantity $c^{\mathrm{T}}x - y^{\mathrm{T}}b$ is called the *duality gap*.
- It follows immediately from the weak duality theorem that if
  (i) x is feasible for the primal problem,
  (ii) y is feasible for the dual problem,
  (iii) $c^{\mathrm{T}}x = y^{\mathrm{T}}b$,

  then x is an optimal solution to the primal problem, and y is an optimal solution to the dual problem.
- This condition gives thus a sufficient condition for optimality.

Duality

- It is also necessary. This is the content of the following theorem.
- *Strong duality theorem.* The primal problem has an optimal solution x if and only if the dual problem has an optimal solution y such that $c^{\mathrm{T}}x = y^{\mathrm{T}}b$.
- The following statement is a corollary of the strong duality theorem.
- It allows us to find an optimal solution to the primal problem given an optimal solution to the dual problem, and vice versa.
- *Complementary slackness.* Let x be an optimal solution to the primal problem, and let y be an optimal solution to the dual. Then the following two equations hold:

$$\begin{aligned} y^{\mathrm{T}}(Ax - b) &= 0, \\ (c - A^{\mathrm{T}}y)^{\mathrm{T}}x &= 0. \end{aligned} \tag{6}$$

The Lagrange multipliers perspective

- We can approach solving LP problems by means of the method of Lagrange multipliers.
- Only the first order conditions, the KKT conditions, will play a role: the Hessian of the Lagrange function is zero, as the objective function and the constraints are linear in x.
- Convexity of the problem is sufficient for the existence of a global minimum (we will discuss it later in these lectures).
- For an LP problem written in the standard form, the Lagrange function is

$$L(x, \lambda, s) = c^{\mathrm{T}} x + \lambda^{\mathrm{T}} (b - Ax) - s^{\mathrm{T}} x. \tag{7}$$

- Here, λ is the vector of Lagrange multipliers corresponding to the equality constraints $Ax = b$, and s is the vector of Lagrange multipliers corresponding to the inequality constraints $-x_i \leq 0$ (note the sign!).

The Lagrange multipliers perspective

- Applying the KKT necessary conditions, we find that

$$\begin{aligned} A^{\mathrm{T}}\lambda + s &= c, \\ Ax &= b, \\ x &\geq 0, \\ s &\geq 0, \\ x_i s_i &= 0,\ \text{for all } i = 1, \ldots, n. \end{aligned} \tag{8}$$

- Note that, as a consequence of the nonnegativity of the x_i's, the complementary slackness condition can equivalently be stated as a single condition $x^{\mathrm{T}}s = 0$.
- If (x^*, λ^*, s^*) is a solution to this system, then

$$\begin{aligned} c^{\mathrm{T}}x^* &= (A^{\mathrm{T}}\lambda^* + s^*)^{\mathrm{T}}x^* \\ &= \lambda^{*\mathrm{T}}Ax^* + s^{*\mathrm{T}}x^* \\ &= (Ax^*)^{\mathrm{T}}\lambda^* + x^{*\mathrm{T}}s^* \\ &= b^{\mathrm{T}}\lambda^*. \end{aligned} \tag{9}$$

The Lagrange multipliers perspective

- In other words, the Lagrange multipliers can be identified with the dual variables y in (4), and $b^{\mathrm{T}}\lambda$ is the objective function for the dual problem!
- Conversaly, we can apply the KKT conditions to the dual problem (3). The Lagrange function reads:

$$\bar{L}(y, x) = b^{\mathrm{T}} y + x^{\mathrm{T}}(c - A^{\mathrm{T}} y), \tag{10}$$

and the first order conditions are

$$\begin{aligned} Ax &= b, \\ A^{\mathrm{T}} y &\leq c, \\ x &\geq 0, \\ x^{\mathrm{T}}(c - A^{\mathrm{T}} y) &= 0. \end{aligned} \tag{11}$$

- The primal-dual relationship is symmetric: by taking the dual of the dual problem, we recover the original problem.

Interpretation of the dual problem

- The dual problem provides a useful intuitive interpretation of the primal problem.
- As an illustration, consider the nutrition problem in Example 2.
- The dual constraints read $\sum_{i=1}^{n} a_{ij}\lambda_i \leq c_j$, $j = 1, \ldots, m$, and so λ_i represents the unit price of nutrient i.
- Therefore, the dual objective function $\sum_{i=1}^{n} \lambda_i b_i$ represents the cost of the daily nutrients that the (imagined) salesman of nutrients is trying to maximize.
- The optimal values λ_i^* of the dual variables are called the *shadow prices* of the nutrients i.
- Even though the nutrients cannot be directly purchased, the shadow prices represent their actual market values.
- Another way of interpreting λ^* is as the sensitivities of the primal function.

Polyhedra

- We now turn to the simplex method, a numerical algorithm for solving LP problems.
- Our presentation of the simplex algorithm follows closely [1]. All the details left out from our discussion can be found in that book.
- Key to the formulation of the method is the *geometry of the feasible set* of an LP problem.
- Each such set forms a polyhedron.
- *Definition.* A *polyhedron* is a subset of $\mathbb{R}^n$ of the form $\mathcal{P} = \{x \in \mathbb{R}^n : Ax \geq b\}$, where $A \in \mathrm{Mat}_{mn}(\mathbb{R})$, and $b \in \mathbb{R}^m$.
- If the feasible set is presented in standard form $\mathcal{P} = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$, the polyhedron is said to be in a *standard form representation*.

Polyhedra

- A polyhedron may extend to infinity or be a bounded set. In the latter case, we refer to the polyhedron as *bounded*.
- A set $S \subset \mathbb{R}^n$ is *convex*, if for any $x, y \in S$ and $\lambda \in (0, 1)$, $\lambda x + (1 - \lambda)y \in S$.
- In other words, a convex set has the property that the line segment connecting any two of its points is contained in the set.
- Notice that a polyhedron $\mathcal{P}$ is a convex set.
- Namely, for $x, y \in \mathcal{P}$ and $\lambda \in (0, 1)$,

$$\begin{aligned} A(\lambda x + (1 - \lambda)y) &= \lambda Ax + (1 - \lambda)Ay \\ &\geq \lambda b + (1 - \lambda)b \\ &= b. \end{aligned}$$

- Polyhedra represented in standard form are, of course, convex as well.

Extreme points

- A vector $x \in \mathcal{P}$ is called an *extreme point* or *vertex*, if it *is not* a convex combination of two distinct points $y, z \in \mathcal{P}$. In other words, x is an extreme point, if

$$x = \lambda y + (1 - \lambda) z, \text{ with } \lambda \in (0, 1), \tag{12}$$

  implies $y = z = x$.
- In other words, an extreme point does not lie on a line between two other points of $\mathcal{P}$.
- Not every polyhedron has extreme points. For example, a half-space $\{x \in \mathbb{R}^n : a^{\mathrm{T}} x \geq b\}$ has no extreme points.
- Drawing a picture of a bounded polygon in the plane we verify that this geometric definition of a vertex conforms with the intuition.

Extreme points

- Let $a_1^{\mathrm{T}}, \ldots, a_m^{\mathrm{T}}$ denote the row vectors of the matrix A. In terms of these vectors, the feasible set can be characterized as $a_i^{\mathrm{T}} x \geq b_i$, $i = 1 \ldots, m$.
- We say that an inequality constraint $a_j^{\mathrm{T}} x \geq b_j$ or an equality constraint $a_j^{\mathrm{T}} x = b_j$ is *active* at a point x^*, if $a_j^{\mathrm{T}} x^* = b_j$.
- By $\mathcal{A}(x^*)$ we denote the set of (the indices of) all constraints active at x^*.
- If $x^* \in \mathbb{R}^n$, then the system

$$a_i^{\mathrm{T}} x = b_i, \quad \text{for all } i \in \mathcal{A}(x^*),$$

  has a unique solution, if and only if there exist n vectors in the set $\{a_i : i \in \mathcal{A}(x^*)\}$ which are linearly independent (why?).
- We will refer to the constraints as linearly independent, if the vectors a_i are linearly independent.

Basic solutions

- We will now define a basic feasible solution (BFS) of an LP problem.
- *Definition.* Consider a polyhedron $\mathcal{P}$, and let $x^* \in \mathbb{R}^n$ (not necessarily in $\mathcal{P}$!).
  - (a) The vector x^* is a *basic solution*, if
    - (i) all equality constraints are active,
    - (ii) n of the active (equality and inequality) constraints at x^* are linearly independent.
  - (b) If x^* is a basic solution that satisfies *all* of the constraints, it is called a *basic feasible solution*.

Extreme points

- *Theorem.* Suppose that the polyhedron $\mathcal{P} = \{x \in \mathbb{R}^n : a_i^{\mathrm{T}} x \geq b_i, i = 1, \ldots, m\}$ is nonempty. Then the following conditions are equivalent:
  (i) $\mathcal{P}$ has at least one extreme point.
  (ii) $\mathcal{P}$ does not contain a line (i.e. there is no direction $d \in \mathbb{R}^n$ such that $x + \lambda d \in \mathcal{P}$ for all $\lambda \in \mathbb{R}$.)
  (iii) There exist n linearly independent vectors among the vectors $a_1, \ldots, a_m$.
- The proof of this theorem is elementary but a bit too lengthy to discuss here, and it can be found in [1].

Extreme points

- *Corollary.* Every nonempty bounded polyhedron and every nonempty polyhedron in standard form has at least one extreme point.
- Indeed, in neither case the polyhedron contains a line.
- The following theorem justifies why we bother about extreme points.
- *Theorem.* Consider the LP problem $\min_{x \in \mathcal{P}} c^{\mathrm{T}} x$. Suppose that $\mathcal{P}$ has an extreme point and that there exists an optimal solution. Then, there exists an optimal solution that is an extreme point of $\mathcal{P}$.
- An immediate corollary to this theorem and the fact that every polyhedron in standard form has an extreme point is the following alternative:
  (i) either the optimization problem is unbounded (and the optimal value is $-\infty$),
  (ii) or there is an optimal solution.

Extreme points

- The proof of the theorem is fun: Let $\mathcal{Q}$ be the (nonempty) set of all optimal solutions, and let v be the optimal value of the objective function $c^{\mathrm{T}}x$.
- Then $\mathcal{Q} = \{x \in \mathbb{R}^n : \ Ax \geq b, c^{\mathrm{T}}x = v\}$, which is also a polyhedron. Since $\mathcal{Q} \subset \mathcal{P}$ and $\mathcal{P}$ does not contain a line, $\mathcal{Q}$ does not contain a line, and so $\mathcal{Q}$ has an extreme point x^*.
- We will show that x^* is also an extreme point of $\mathcal{P}$.
- Assume that $x^* = \lambda y + (1 - \lambda)z$, for $y, z \in \mathcal{P}$ and $\lambda \in (0, 1)$. Then

$$\begin{aligned} v &= c^{\mathrm{T}}x^* \\ &= \lambda c^{\mathrm{T}}y + (1 - \lambda)c^{\mathrm{T}}z. \end{aligned}$$

- Since $c^{\mathrm{T}}y, c^{\mathrm{T}}z \geq v$, this is possible only if $c^{\mathrm{T}}y = c^{\mathrm{T}}z = v$.
- Therefore, $y, z \in \mathcal{Q}$. But this contradicts the fact that, by assumption, x^* is an extreme point of $\mathcal{Q}$.
- This contradiction shows that x^* is an extreme point of $\mathcal{P}$. Also, since $x^* \in \mathcal{Q}$, it is optimal.

Degeneracy

- A basic solution $x \in \mathbb{R}^n$ is called *degenerate*, if more than n of the constraints are active at x.
- *Example 8*. Consider the polyhedron:

$$\begin{aligned} x_1 + x_2 + 2x_3 &\leq 8 \\ x_2 + 6x_3 &\leq 12 \\ x_1 &\leq 4 \\ x_2 &\leq 6. \\ x_1, x_2, x_3 &\geq 0. \end{aligned}$$

- The vector $x = (2, 6, 0)$ is a nondegenerate BFS, because there are exactly three active, linearly independent constraints: the first, the fourth, and $x_3 \geq 0$.
- The vector $x = (4, 0, 2)$ is degenerate, because there are four active constraints: the first three and $x_2 \geq 0$.

Degeneracy

- Generically, basic solutions are nondegenerate. If we generated the constraints coefficients purely randomly, we would end up, with probability 100%, with a fully nondegenerate problem.
- Degeneracies occur as a result of additional or coincidental dependencies that are common in real life situations. For this reason, they need to be addressed in any solution methodology.
- The simple method becomes a bit hairy if a degenerate BFS is encountered. In order to keep the discussion straightforward, from now on we assume that all BFSs are nondegenerate.
- We will make a few remarks regarding the degenerate case, following the presentation of the main outline of the simplex method.

Basic solutions for polyhedra in standard form

- From now on we assume that the polyhedron is represented in standard form $\mathcal{P} = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$.
- We will also assume that the matrix A is of full rank, i.e. exactly $m \leq n$ of its rows are linearly independent.
- This is no loss of generality, because if the rank of A is $k < m$, we can consider an equivalent problem with a rank k submatrix D of A with the redundant rows eliminated.
- *Example 9*. Consider the polyhedron defined by the constraints:

$$
\begin{aligned}
2x_1 + x_2 + x_3 &= 2, \\
x_1 + x_2 \qquad &= 1, \\
x_1 \qquad + x_3 &= 1, \\
x_1, x_2, x_3 &\geq 0.
\end{aligned}
$$

- The corresponding matrix A has rank 2. The first constraint is redundant (it is the sum of the second and third constraints), and can be eliminated without changing the problem.

Basic solutions for polyhedra in standard form

- At a basic solution of a polyhedron in standard form, the m equality constraints are always active.
- Then A has m columns that are linearly independent. Let $A_{r_1}, \ldots, A_{r_m}$, where each $A_{r_i} \in \mathbb{R}^n$, denote a set of *m linearly independent columns* of A.
- We must also have $x_j = 0$, for all $j \notin \{r_1, \ldots, r_m\}$.
- We let B denote the submatrix of A formed by these columns, i.e.

$$B = (A_{r_1} \quad \ldots \quad A_{r_m}).$$

  B is called a *basis matrix*.
- Since B is a square matrix of maximum rank, it is invertible.
- Clearly, a basic solution for polyhedra in standard form is degenerate if more than $n - m$ components of x are zero.

Basic solutions

- Permuting the columns of A we write it in the block form $\begin{pmatrix} B & N \end{pmatrix}$. Under the same permutation, a vector x can be written in the block form:

$$\begin{pmatrix} x_B \\ x_N \end{pmatrix}.$$

- Recall that in order for x to be a basic solution, we must have $x_N = 0$.
- The equation $Ax = b$ is equivalent to the block form equation

$$\begin{pmatrix} B & N \end{pmatrix} \begin{pmatrix} x_B \\ 0 \end{pmatrix} = b,$$

or

$$Bx_B = b.$$

Basic solutions

- Its solution reads
$$x_B = B^{-1}b. \tag{13}$$
- The variables x_B are called *basic variables*, while the variables x_N are called *nonbasic variables*.
- Recall that a basic solution is not guaranteed to be feasible, as it may violate the nonnegativity condition $x_B \geq 0$.

Basic solutions

- There is an key link the geometric concept of a vertex of a polyhedron and the analytic concept of a BFS given by the theorem below.
- *Theorem*. A vector x^* is a BFS, if and only if it is a vertex in $\mathcal{P}$.
- For the proof, we assume that x^* is *not* an extreme point of $\mathcal{P}$, i.e. it can be represented as $x^* = \lambda y + (1 - \lambda) z$, with $0 < \lambda < 1$, and distinct $y, z \in \mathcal{P}$.
- But then also $x^*_N = \lambda y_N + (1 - \lambda) z_N$. However, since $x^*_N = 0$, and $y, z \leq 0$ (since they are elements of $\mathcal{P}$), it follows that also $y_N = 0$ and $z_N = 0$.
- Since $Bx^*_B = b$, we also must have $By_B = b$ and $Bz_B = b$ (because $x^*_N = y_N = z_N = 0$).
- This implies that $x^*_B = y_B = z_B$ $(= B^{-1} b)$, and so $x^* = y = z$. This contradiction means that x^* is extreme.

Basic solutions

- In order to prove the converse statement, we suppose that x^* is not an BFS.
- Therefore, there do not exist n linearly independent active constraints, and thus the vectors a_i span a proper subspace of $\mathbb{R}^n$. We can thus find a direction d such that $a_i^{\mathrm{T}} d = 0$, for all $i \in \mathcal{A}(x^*)$.
- Consider now the vectors $y = x^* - \varepsilon d$ and $z = x^* + \varepsilon d$. Note that these vectors satisfy the active constraints, as $a_i^{\mathrm{T}}(x^* \pm \varepsilon d) = a_i^{\mathrm{T}} x^* \pm \varepsilon a_i^{\mathrm{T}} d = b$.
- Choosing ε sufficiently small, we can assure that they also satisfy the inactive constraint $a_i^{\mathrm{T}} x > b$.
- Consequently, $y, z \in \mathcal{P}$, and $x^* = \frac{1}{2}(y + z)$, which means that x^* is not an extreme point. This contradiction completes the proof.

Adjacent BFSs

- We will now proceed to describing an algorithm for moving from one BFS to another and decide when to stop the search.
- We start with the following definition.
- (i) Two BFSs are *adjacent*, if their basic matrices differ in one basic column only.
  (ii) Let $x \in \mathcal{P}$. A vector $d \in \mathbb{R}^n$ is a *feasible direction* at x, if there is a positive number θ such that $x + \theta d \in \mathcal{P}$.
  (iii) A vector $d \in \mathbb{R}^n$ is an *improving direction*, if $c^{\mathrm{T}} d < 0$.
- In other words, moving from x in an improving direction d lowers the value of the objective function $c^{\mathrm{T}} x$ by $c^{\mathrm{T}} d$.

Adjacent BFSs

- Note that if the new point $x + \theta d$ is feasible, then

$$Ad = 0. \tag{14}$$

  Indeed, with $\theta > 0$,

$$\begin{aligned} \theta A d &= A(x + \theta d) - Ax \\ &= b - b \\ &= 0. \end{aligned}$$

- The strategy is, starting from a BFS, to find an improving feasible direction towards an adjacent BFS.

Adjacent BFSs

- We move in the j-th *basic direction* $d = (d_B \;\; d_N)$ that has exactly one positive component corresponding to a *non-basic* variable.
- When moving in the basic direction, the nonbasic variable $x_j = 0$ becomes positive, while the other nonbasic variables remain zero. We say that x_j *enters the basis*.
- Specifically, we select a nonbasic variable x_j and set

$$\begin{aligned} d_j &= 1, \\ d_i &= 0, \text{ for every nonbasic index } i \neq j. \end{aligned}$$

Adjacent BFSs

- As a result, x changes to $x + d_B$, where

$$\begin{aligned} 0 &= Ad \\ &= \sum_{i=1}^{n} A_i d_i \\ &= \sum_{i=1}^{m} A_{r_i} d_{r_i} + A_j \\ &= B d_B + A_j, \end{aligned}$$

and so $d_B = -B^{-1} A_j$.

Adjacent BFSs

- We are now facing two cases.
- *Case 1*: x is a nondegenerate BFS. Then $x_B > 0$, which implies that $x_B + \theta d_B > 0$, and feasibility is assured by choosing θ sufficiently small.
- In particular, d is a feasible direction.
- *Case 2*: x is degenerate. In this case, d is not always a feasible direction. It is possible that a basic variable x_{r_i} is zero, while the corresponding component d_{r_i} is negative.
- In this case, if we follow the j-th basic direction, the nonnegativity constraint for d_{r_i} is violated, and we are led to nonfeasible solutions.

Reduced cost

- We will now study the effect of moving in the j-th basic direction on the objective function.
- Let x be a basic solution with basis matrix B, and let c_B be the vector of the costs of the basic variables.
- For each $i = 1, \ldots, n$ the *reduced cost* $\bar{c}_i$ of x_i is defined by

$$\bar{c}_i = c_i - c_B^{\mathrm{T}} B^{-1} A_i. \tag{15}$$

- The j-th basic direction is improving if and only if $\bar{c}_i < 0$.

Example

- *Example 10*. Consider the following problem. For $x \in \mathbb{R}^4$,

$$\min c^{\mathrm{T}} x \quad \text{subject to} \begin{cases} x_1 + x_2 + x_3 + x_4 = 2, \\ 2x_1 + 3x_3 + 4x_4 = 2, \\ x_i \geq 0,\ i = 1, \ldots, 4. \end{cases}$$

- The first two columns of the matrix A are

$$A_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

  Since they are linearly independent, we can choose x_1 and x_2 as our basic variables, and

$$B = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix}.$$

- We set $x_3 = x_4 = 0$ and solve the constraints to find $x_1 = 1$, and $x_2 = 1$. We have thus constructed a nondegenerate BFS.

Example

- A basic direction corresponding to the nonbaisc variable x_3 is obtained as follows:

$$\begin{aligned} d_B &= -B^{-1} A_3 \\ &= -\begin{pmatrix} 0 & 1/2 \\ 1 & -1/2 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} -3/2 \\ 1/2 \end{pmatrix}. \end{aligned}$$

- The cost of moving along this basic direction is $c^{\mathrm{T}} = -3c_1/2 + c_2/2 + c_3$.

Optimality condition

- The next result provides us with optimality conditions.
- *Theorem.* Let x be a BFS and let $\bar{c}$ be the corresponding vector of reduced costs.
  (i) If $\bar{c} \geq 0$, then x is optimal.
  (ii) If x is optimal and nondegenerate, then $\bar{c} \geq 0$.
- To prove (i) we let y be an arbitrary feasible solution, and $d = y - x$. Feasibility implies that $Ad = 0$, and so

$$Bd_B + \sum_j A_j d_j = 0,$$

  where the summation extends over the nonbasic indices j.
- Then

$$d_B = -\sum_j d_j B^{-1} A_j .$$

Optimality condition

- Therefore,

$$
\begin{aligned}
c^{\mathrm{T}} d &= c_B^{\mathrm{T}} d_B + \sum_j c_j d_j \\
&= \sum_j (c_j - c_B^{\mathrm{T}} B^{-1} A_j) d_j \\
&= \sum_j \overline{c}_j d_j .
\end{aligned}
$$

- For any nonbasic j we have $x_j = 0$ and $y_j \geq 0$ (since y is feasible). Therefore, $d_j \geq 0$, and $\overline{c}_j d_j \geq 0$.
- As a consequence, $c^{\mathrm{T}} y - c^{\mathrm{T}} x = c^{\mathrm{T}} d \geq$, and x is optimal.
- To prove (ii), assume that $\overline{c}_j < 0$. Since the reduced cost of a basic variable is zero, x_j must be nonbasic. Since x is nondegenerate, the j-th basic direction is a direction of cost decrease, and x cannot be optimal.

Step size

- Let d be a basic, feasible, improving direction from the current BFS x, and let B be the basis matrix for x.
- We wish to move by the amount of $\theta > 0$ in the direction d in order to find a BFS x' adjacent to x. This takes us to the point $x + \theta^* d$, where

$$\theta^* = \max\{\theta \geq 0 : \ x + \theta d \in \mathcal{P}\},$$

  and the resulting cost change is $\theta^* c^{\mathrm{T}} d$.
- Since $Ad = 0$, the equality constraints will never be violated, and $x + \theta d$ can become infeasible only if its components become negative.
- We want to find the largest possible θ such that $x_B + \theta d_B \geq 0$.
- If $d \geq 0$, then $x + \theta d \geq 0$, and $\theta^* = \infty$.
- If $d_i < 0$ for some i, then the condition becomes $x_i + \theta d_i \geq 0$, or $\theta \leq -x_i / d_i$. This condition has to hold for all i with $d_i < 0$.

Step size

- In other words, we are led to the following choice:

$$\theta^* = \min_{i:\, d_i < 0} \Big(-\frac{x_i}{d_i} \Big). \tag{16}$$

- If x_i is a nonbasic variable, then x_i is either entering and $d_i = 1$, or else $d_i = 0$. Therefore, it is sufficient to consider the basic variables only and thus

$$\theta^* = \min_{i:\, d_{r_i} < 0} \Big(-\frac{x_{r_i}}{d_{r_i}} \Big). \tag{17}$$

- Since we are assuming nondegeneracy, $x_{r_i} > 0$ and $\theta^* > 0$.
- Once θ^* has been chosen (and is finite), we move to the next feasible solution.

Step size

- Since $x_j = 0$ and $d_j = 1$, we have $y_j = \theta^* > 0$. Let l be the index saturating the minimum (17). Then $d_{r_l} < 0$ and

$$x_{r_l} + \theta^* d_{r_l} = 0.$$

- This means that the new basic variable has become 0, whereas the nonbasic variable x_j has become positive. This indicates that, in the next iteration, the index j should replace r_l.
- In other words, the new basis matrix B is obtained from B by replacing its column A_{r_l} with the column A_j.
- The columns A_{r_i}, $i \neq l$, and A_j are linearly independent and form a new basis matrix $\overline{B}$.
- The vector $y = x + \theta^* d$ is a BFS corresponding to $\overline{B}$.

An iteration of the simplex algorithm

- We can now summarize a typical iteration of the simplex algorithm as follows:
  (i) Start from a basis $A_{r_1}, \ldots, A_{r_m}$ of the columns of the matrix A and the associated BFS x.
  (ii) Compute the vector $\bar{c}$ of reduced costs corresponding to all nonbasic indices j. If $\bar{c} \geq 0$, then the current BFS x is optimal and the algorithm exits. Otherwise, choose a nonbasic index j for which $\bar{c}_j < 0$.
  (iii) Compute $u = B^{-1} A_j$. If no component of u is positive, then the optimal cost is $-\infty$, and the algorithm terminates.
  (iv) If some components of u are positive, compute the step size θ^* using (17).
  (v) Let l be the index saturating the minimum in (17). Form a new basis by replacing A_{r_l} with A_j. If y is a new BFS, the value of the new basic variables are $y_j = \theta^*$ and $y_{r_i} = y_{r_i} - \theta^* u_i$.
  (vi) Replace x_{r_i} with x_j in the list of basic variables.
- If $\mathcal{P}$ is nonempty and every BFS is *nondegenerate*, the simplex method terminates after finitely many steps. At termination, there are two possibilities:
  (i) We have an optimal basis B and an associated BFS that is optimal.
  (ii) We have found a d satisfying $Ad = 0$, $d \geq 0$, and $c^{\mathrm{T}} d < 0$, and the optimal value is $-\infty$.

Degenerate problems

- Degenerate problems present a number of technical issues such as
  - (i) What nonbasic variable should enter the BFS?
  - (ii) If more than one basic variable could leave (i.e. more than one basic variable attains the minimum that gives θ^*), which one should leave?
  - (iii) The algorithm may cycle forever at the some degenerate BFS.
- Various extensions and refinements to the basic method outlined in these notes have been developed to address the issues above.
- For example, *Bland's rule* is used to eliminate the risk of cycling:
  - (i) Among all nonbasic variables that can enter the new basis, select the one with the minimum index.
  - (ii) Among all basic variables that can exit the basis, select the one with the minimum index.
- The details are presented in Chapter 3 of [1].

Finding an initial BFS

- As usual, starting the iteration may sometimes not be easy, and finding an initial BFS may prove challenging.
- One strategy is to solve an auxiliary problem.
- For example, if we want a BFS with $x_i = 0$, we set the objective function to x_i and find the optimal solution to this problem.
- If the optimal value is 0 then we found a BFS with $x_i = 0$, otherwise there is no such feasible solution.

References

[1] Bertsimas, D., and Tsitsiklis, J. N.: *Linear Optimization*, Athena Scientific (1997).

[2] Nocedal, J., and Wright, S. J.: *Numerical Optimization*, Springer (2006).