Optimization

3. Linear programming

Andrew Lesniewski

Baruch College
New York

Fall 2018

Outline

Motivation: short term financing problem

- As a motivation we consider the following example, taken from [1].
- A company needs to finance its short term term cash commitments. Specifically, for the next six months its cash requirements are given by the following table:

Month	Jan	Feb	Mar	Apr	May	Jun
Cashflow	-15	-100	200	-200	50	300

  The cashflows are given in thousands of USD.
- The company has the following sources of funding:
  (i) a line of credit for up to \$100k a rate of 1% per month,
  (ii) for the first thre months, it can issue 90-day commercial paper at the total rate of 2%,
  (iii) it can deposit excess funds at the rate of 0.3% per month.
- The company's objective is to decide how to use each source of funding to fulfill the liability requirements, and maximize its cash position at the end of June.

Motivation: short term financing problem

- In order to formulate the problem in mathematical terms, we will introduce the following *decision variables*:
  (i) $x_i \in [0, 100]$ is the amount drawn from the line of credit in month i,
  (ii) $y_i \geq 0$ is the face value of commercial paper issued in month i,
  (iii) $z_i \geq 0$ is the excess cash in month i,
  (iv) $f \in \mathbb{R}$ is the cash position in June.
- The variables x_i, y_i, z_i satisfy additional relations that we will formulate month by month.
- *Jan.* The company can borrow x_1 from the line of credit, and issue y_1 face value of commercial paper to meet the liability of $150k, i.e.

$$x_1 + y_1 - z_1 = 150.$$

- *Feb.* The cash liability is $100k. Additionally, principal plus interest $1.01x_1$ on the credit line is due, and $1.003z_1$ is received. The cash balance equation reads:

$$x_2 + y_2 - 1.01x_1 + 1.003z_1 - z_2 = 100.$$

Motivation: short term financing problem

- *Mar.* Similarly, the Mar equation reads:

$$x_3 + y_3 - 1.01x_2 + 1.003z_2 - z_3 = -200.$$

- *Apr.* The company can no longer issue commercial paper, but it has to pay the 2% interest on the amount y_1 of commercial paper issued in Jan. This yields:

$$x_4 - 1.02y_1 - 1.01x_3 + 1.003z_3 - z_4 = 200.$$

- *May.* Similarly, the May equation reads:

$$x_5 - 1.02y_2 - 1.01x_4 + 1.003z_4 - z_5 = -50.$$

- *Jun.* Credit line no longer available, and the cash position at the end of June is the excess cash:

$$-1.02y_3 - 1.01x_5 + 1.003z_5 - f = -300.$$

Motivation: short term financing problem

- The full model can be thus formulated as the following optimization problem:

$$
\max f \quad \text{subject to} \begin{cases} x_1 + y_1 - z_1 = 150, \\ x_2 + y_2 - 1.01x_1 + 1.003z_1 - z_2 = 100, \\ x_3 + y_3 - 1.01x_2 + 1.003z_2 - z_3 = -200, \\ x_4 - 1.02y_1 - 1.01x_3 + 1.003z_3 - z_4 = 200, \\ x_5 - 1.02y_2 - 1.01x_4 + 1.003z_4 - z_5 = -50, \\ -1.02y_3 - 1.01x_5 + 1.003z_5 - f = -300, \\ x_i \leq 100,\ i = 1, \ldots, 5, \\ x_i \geq 0,\ i = 1, \ldots, 5, \\ y_i \geq 0,\ i = 1, 2, 3, \\ z_i \geq 0,\ i = 1, \ldots, 5. \end{cases} \tag{1}
$$

Motivation: short term financing problem

- The problem formulated above is an example of a *linear programming* problem: the objective function as well as the constraint functions are first order polynomials in the decision variables.
- Linear programming problems (especially if the number of decision variables and constraints are large) are notoriously difficult to solve. Two dimensional problems with a moderate number of constraints can be solved using a graphical method.
- 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

- *Linear programming* (LP) is concerned with problems in which both the objective function $f(x)$ and the constraint functions $c_i(x)$, $i = 1, \ldots, m$, are linear (or, more precisely, affine), i.e.

$$\min f(x) = c^{\mathrm{T}} x, \quad \text{subject to } \begin{cases} a_i^{\mathrm{T}} x = b_i, & \text{for } i \in \mathcal{E}, \\ a_i^{\mathrm{T}} x_i \geq b_i, & \text{for } i \in \mathcal{I}, \end{cases} \tag{2}$$

  where $c \in \mathbb{R}^n$, and $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, 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 (2) (or *feasible vector*) is any x satisfying all the constraints. The feasible set is the the set of all feasible solutions.

Standard form

- Any linear programming can be written in the *standard form*, namely

$$\min c^{\mathrm{T}} x, \quad \text{subject to } \begin{cases} Ax = b, \\ x_i \geq 0, \text{ for } i = 1, \ldots, n. \end{cases} \tag{3}$$

  where $c \in \mathbb{R}^n$, $A \in \mathrm{Mat}_n(\mathbb{R})$,and $b \in \mathbb{R}^n$.

- This form is convenient for describing solution algorithms for LP problems, and we will be assuming it in the following.

Standard form

- Every problem of the form (2) can be transformed into the form (3). This can be accomplished by means of two types of operations:
  (i) Elimination of inequality constraints: given an inequality of the form

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

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

$$\begin{aligned} \sum_{j=1}^{n} a_{ij} x_j + s_i &= b_i, \\ s_i &\geq 0. \end{aligned}$$

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

Example

- *Example 1.* 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 2.* 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:

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

Lower bound

- Let us go back to Example 1 written in the standard form, and consider the following few feasible solutions:

$$\begin{aligned} x &= (0, 9/2, 15/2, 0), \quad f(x) = -9/2, \\ x &= (6, 0, 0, 3), \quad f(x) = -6, \\ x &= (5, 2, 0, 0), \quad f(x) = -7. \end{aligned}$$

  Is the last solution the optimal solution?
- 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

$$\begin{aligned} -x_1 - x_2 \geq & -2x_1 - x_2 - x_3 \\ & - 12. \end{aligned}$$

Lower bound for the solution

- The second constraint doeas a little better

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

- 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.
- Let us formalize this procedure.

Lower bound for the solution

- We can consider a linear combination of the two constraints with coefficients y_1 and y_2:

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

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

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

- We obtain the largest possible lower bound, we should maximize the corresponding linear combination of the right hand sides of the constraints, namely $12y_1 + 9y_2$.
- This leads us to the following *dual problem*:

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

Duality

- For a general problem (3) (called the *primal problem*), we consider the corresponding *dual problem*:

$$\max b^{\mathrm{T}} y, \quad \text{subject to } A^{\mathrm{T}} y \leq c. \tag{4}$$

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

$$\max b^{\mathrm{T}} y, \quad \text{subject to } \begin{cases} A^{\mathrm{T}} y + s = c, \\ s_i \geq 0, \text{ for } i = 1, \ldots, n. \end{cases} \tag{5}$$

- There is a relationship between 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:
- *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

$$c^{\mathrm{T}} x \geq b^{\mathrm{T}} y. \tag{6}$$

Duality

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

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

- 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 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:

$$\begin{aligned} y^{\mathrm{T}}(Ax - b) &= 0, \\ (c - A^{\mathrm{T}}y)^{\mathrm{T}}x &= 0. \end{aligned} \tag{7}$$

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

$$L(x, \lambda, s) = c^{\mathrm{T}} x + \lambda^{\mathrm{T}} (b - Ax) - s^{\mathrm{T}} x. \tag{8}$$

  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 $-x_i \leq 0$.

The Lagrange multipliers perspective

- Applying the KKT necessary conditions, we find that

$$
\begin{aligned}
A^{\mathrm{T}}\lambda + s &= c, \\
Ax &= b, \\
x &\geq 0, \\
s &\geq 0, \\
x_i s_i &= 0, \text{ for all } i = 1, \ldots, n.
\end{aligned}
\tag{9}
$$

  Note that the complementary slackness condition can equivalently be stated as $x^{\mathrm{T}} s = 0$ as a consequence of the nonnegativity of the x_i's.

- If (x^*, λ^*, s^*) is a solution to this system, then

$$
\begin{aligned}
c^{\mathrm{T}} x^* &= (A^{\mathrm{T}}\lambda^* + s^*) x^* \\
&= (Ax^*)^{\mathrm{T}} \lambda^* \\
&= b^{\mathrm{T}} \lambda^*.
\end{aligned}
\tag{10}
$$

The Lagrange multipliers perspective

- In other words, the Lagrange multipliers can be identified with the dual variables y in (5), and $b^{\mathrm{T}}\lambda$ is the objective function for the dual problem!
- Conversaly, we can apply the KKT conditions to the dual problem (4). The Lagrange function reads:

$$\bar{L}(y, x) = b^{\mathrm{T}} y + x^{\mathrm{T}}(c - A^{\mathrm{T}} y), \tag{11}$$

and the first order conditions are

$$\begin{aligned} Ax &= b, \\ A^{\mathrm{T}} y &\leq c, \\ x &\geq 0, \\ x^{\mathrm{T}}(c - A^{\mathrm{T}} y) &= 0. \end{aligned} \tag{12}$$

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

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 $\{x \in \mathbb{R}^n : Ax \geq b\}$, where $A \in \mathrm{Mat}_{mn}(\mathbb{R})$, and $b \in \mathbb{R}^m$.
- In particular, if the feasible set is presented in standard form $\{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$, the polyhedron is said to be in a *standard form representation*. In the following, we will denote polyhedra by the symbol $\mathcal{P}$.

Polyhedra

- 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 $a_i^{\mathrm{T}} x \geq b_i$, $i = 1 \ldots, m$.
- As usual, an inequality constraint $a_j^{\mathrm{T}} x \geq b_j$ is active at x^* if $a_j^{\mathrm{T}} x^* = b_j$. By $\mathcal{A}(x^*)$ we denote the set of all constraints active at x^*.
- The following observation will play a role in the following: Let $x^* \in \mathbb{R}^n$ and let $\mathcal{A}(x^*)$ denote the set of active constraints at x^*. Then the system

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

  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.
- Sometimes, for convenience, we will refer to the constraints as linearly independent, if the vectors a_i are linearly independent.

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*, 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

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

implies $y = z = x$.

- 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.
- *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 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.

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*.

Degeneracy

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

- Assume that the polyhedron is represented in standard form.
- From now on we will assume that the matrix A is of full rank, i.e. exactly m 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.* Consider the polyhedron defined by the constraints:

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

  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 et B denote the submatrix of A formed by these columns, $B = (A_{r_1} \quad \ldots \quad A_{r_m})$. 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 have to 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

$$Bx_B = b.$$

- 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*.

Basic solutions

- Recall that a basic solution is no guaranteed to be feasible, as it may violate the nonnegativity condition $x_B \geq 0$.
- There is an important link between BFSs and extreme points of the polyhedron $\mathcal{P}$: every BFS corresponds to one and only one extreme point of $\mathcal{P}$.
- The proof of this theorem is a bit long (see [1]), we will verify only the fun part of it, namely that a BFS is an extreme point of $\mathcal{P}$.
- Namely, 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.

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 $c^{\mathrm{T}} d$.
- Note that if d is feasible,

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

Indeed, for some $\theta > 0$,

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

Adjacent BFSs

- The strategy is, starting from a BFS, to find an improving feasible direction towards an adjacent BFS.
- We move in the j-th *basic direction* $d = (d_B \;\; d_N)$ which is defined by

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

- 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:
  - (i) 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.
  - (ii) 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.
- We will now study the effect of moving in the j-th basic direction on the objective function.

Reduced cost

- 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

- 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 fist two colums 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 $\bar{c} > 0$, then x is optimal.
  (ii) If x is optimal and nondegenerate, then $\bar{c} > 0$.
- (i) Let y be an arbitrary feasible solution, and $d = y - x$. Then d is a linear combination of $-B^{-1}A_j$, the j-th basic directions:

$$d = -\sum_j d_j B^{-1} A_j,$$

  where the summation extends over the nonbasic indices j. Note that $b_j \geq 0$.
- It follows that

$$\begin{aligned} c^{\mathrm{T}} d &= c_B^{\mathrm{T}} d_B + \sum_j c_j d_j \\ &= \sum_j \bar{c} j d_j \\ &\geq 0. \end{aligned}$$

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

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

- We want to find the largest possible θ such that $x_B + \theta d_B \geq 0$. If $d \geq 0$, this yields $\theta^* = \infty$. If one of the components $d_i < 0$, then the condition becomes $x_i + \theta d_i \geq 0$, or $\theta \leq -x_i/d_i$.
- In other words, we are led to the following choice. Let $u = -d_B = B^{-1}A_j$. Then

$$\theta^* = \min_{r_i:\, u_{r_i} > 0} \frac{x_{r_i}}{u_{r_i}} \,. \tag{16}$$

- 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 (16). 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 frm B by replacing its column A_{r_l} with the column A_j.

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 $u < 0$, then the optimal cost is $-\infty$, and the algorithm terminates.
  (iv) If some components of u are positive, compute the step size θ^* using (16).
  (v) Let l be the index saturating the minimum in (16). 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

- Bertsimas, D., and Tsitsiklis, J. N.: *Linear Optimization*, Athena Scientific (1997).
- Cornuejos, G., and Tutuncu, R.: *Optimization Methods in Finance*, Cambridge University Press (2007).
- Nocedal, J., and Wright, S. J.: *Numerical Optimization*, Springer (2006).