Linear Programming and the Simplex Method

Linear programs form a subclass of optimization problems. Before now, it may be that your one exposure to optimization has been in calculus:

- **Prototype optimization from MATH 271**: Determine the global extreme values of the function $f(x, y) = x^3 = 2y$ on the domain $0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1$.
- **Prototype optimization from MATH 171**: Find the dimensions of a cylindrical can with volume 900 cm^3 that uses the least amount of metal (i.e., that has minimum surface area).

Optimization methods for various sorts of problems and contexts present an active area of mathematical research, and there are many books on the subject.

The problems at hand, called **linear programs**, involve optimizing (maximizing or minimizing) a linear functional subject to linear equations and/or linear inequalities. A typical problem could take the form

$$\text{minimize } \ \mathbf{c}^{\mathrm{T}}\mathbf{x} \quad \text{subject to} \quad \mathbf{x} \geqslant \mathbf{0} \ \text{ and } \ \mathbf{A}\mathbf{x} \geqslant \mathbf{b}, \tag{1}$$

where $\mathbf{A}$ is a given m-by-n matrix, $\mathbf{b}$ is a given vector in $\mathbb{R}^m$, and $\mathbf{c}$ is a given vector in $\mathbb{R}^n$. Vectors $\mathbf{x} \in \mathbb{R}^n$ which satisfy the constraints—in this case, the requirements $\mathbf{A}\mathbf{x} \geqslant \mathbf{b}$ and $\mathbf{x} \geqslant \mathbf{0}$—are said to be **feasible**.

The constraints

For a vector $\mathbf{x} \in \mathbb{R}^n$, the notation $\mathbf{x} \geqslant \mathbf{0}$ simply means that each component x_j of $\mathbf{x}$ must satisfy $x_j \geqslant 0$, a natural requirement when the variables $x_1, \ldots, x_n$ represents amounts of physical quantities present. The equation $x_j = 0$ in $\mathbb{R}^n$ is an $(n-1)$-dimensional **hyperplane**, fixing the value of one component while leaving all the other variables in $\mathbf{x} = (x_1, \ldots, x_n)$ *free*. The inequality $x_j \geqslant 0$ has the hyperplane $x_j = 0$ as boundary, dividing $\mathbb{R}^n$ into two parts and selecting out the **half-space** $x_j \geqslant 0$. A feasible $\mathbf{x}$ must live in that half-space. In fact, the **feasible set** consists only of $\mathbf{x}$ which simultaneously live in many half-spaces dictated by $\mathbf{A}\mathbf{x} \geqslant \mathbf{b}$ and $\mathbf{x} \geqslant 0$:

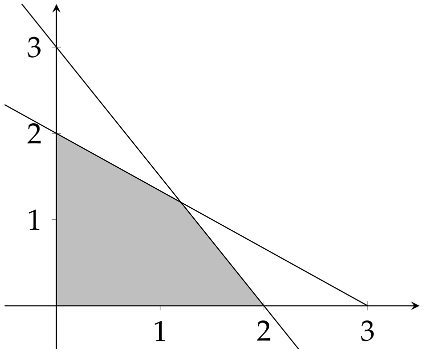
$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &\geqslant b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &\geqslant b_2 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &\geqslant b_m \\
x_1 &\geqslant 0 \\
&\vdots \\
x_n &\geqslant 0
\end{aligned}$$

In the next few examples, we consider several sorts of feasible sets that our constraints could present.

Example 1:

Let $\mathbf{A} = \begin{bmatrix} -2 & -3 \\ -3 & -2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -6 \\ -6 \end{bmatrix}$. For $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, the constraints $\mathbf{Ax} \geqslant \mathbf{b}$, $\mathbf{x} \geqslant 0$ translate into

$$\begin{aligned} -2x - 3y &\geqslant -6 \\ -3x - 2y &\geqslant -6 \\ x &\geqslant 0 \\ y &\geqslant 0 \end{aligned}$$



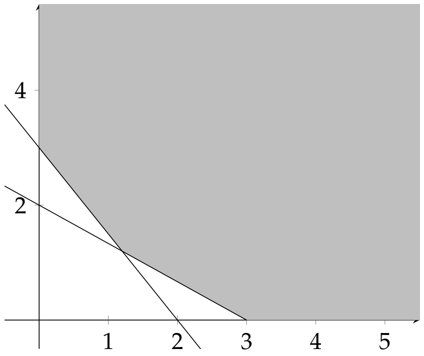
with the feasible set shaded at right. This region is closed and bounded, and the Extreme Value Theorem from Calculus says an continuous function on such a set attains global extrema.

■

Example 2:

Let $\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$. For $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, the constraints $\mathbf{Ax} \geqslant \mathbf{b}$, $\mathbf{x} \geqslant 0$ translate into

$$\begin{aligned} 2x + 3y &\geqslant 6 \\ 3x + 2y &\geqslant 6 \\ x &\geqslant 0 \\ y &\geqslant 0 \end{aligned}$$



with the feasible set shaded at right. This region is unbounded.

■

The simplex method, first look

This method for solving a problem of the form (3) was introduced by George Dantzig in 1946. It is a fast alternative to an exhaustive search of all corners of the feasible set which, in most settings, are quite numerous. The idea is as follows.

- Find, initially, one corner of the feasible set.
- If any of the n edges of the (polytope) feasible set which join at this corner represent directions which decrease cost, find the edge that leads to the greatest decrease, and follow it to a new corner.
- Repeat the previous step iteratively until reaching a corner where no edge can lead to lower cost.

Corners occur where n of the dividing hyperplanes that serve as boundaries of half-spaces come together. If we set n of the components of $\tilde{\mathbf{x}}$ equal to 0 (that is, take them as *free variables* and set them to zero), the remaining m components are pivot, or **basic variables**, and the solution of $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{b}$ that arises from taking the free variables to be zero is called a **basic solution**. If all basic variables in a basic solution are positive, then $\tilde{\mathbf{x}}$ is a **(nondegenerate) basic feasible solution**, corresponding to a corner of the feasible set where n hyperplanes (faces of the feasible set) come together.

We next consider the process of **pivoting**, exchanging one of the free variables for one of the basic ones. Upon generating relative cost coefficients for each of the free variables, we look to see if any are negative. If all are positive, then we have already arrived at the **optimal feasible solution**, and stop. If, however, one of these relative cost coefficients is negative, we choose the free variable with the largest (in magnitude) negative relative cost coefficient as one to take the place of a current basic variable.

By trading out just one of the free variables (i.e., keeping $n-1$ of them the same, while adopting one of the former basic variables as the n^{th} free one), we ensure that the new point we find by solving is joined to the previous one by an edge. We follow a system for choosing which variables are traded:

- The free variable with the most negative coefficient in $\mathbf{c}$ becomes basic.
- We look at positive ratios of values of the basic variables to the coefficients of the incoming variable, and choose the leaving variable to be the one with the smallest positive ratio.

Example 3:

We seek a solution of

$$\min_{\mathbf{x}\in\mathbb{R}^4} \begin{bmatrix} 1 & 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \qquad \text{subject to} \qquad \begin{bmatrix} 1 & 0 & 2 & -2 \\ 0 & 1 & 1 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 2 \\ 6 \end{bmatrix} \quad \text{and} \quad \mathbf{x} \geqslant \mathbf{0}.$$

We set $z = x_1 + x_2 + 2x_3 + x_4$, giving us the equations

$$\begin{aligned} z - x_1 - x_2 - 2x_3 - x_4 &= 0 \\ x_1 + 2x_3 - 2x_4 &= 2 \\ x_2 + x_3 + 4x_4 &= 6 \end{aligned} .$$

Arranging in a *tableau* we have

z	x_1	x_2	x_3	x_4	RHS
1	-1	-1	-2	-1	0
0	1	0	2	-2	2
0	0	1	1	4	6

or, in matrix RREF form, $\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 8 \\ 0 & 1 & 0 & 2 & -2 & 2 \\ 0 & 0 & 1 & 1 & 4 & 6 \end{bmatrix}$

The submatrix corresponding to the matrix equation $\begin{bmatrix} 1 & 0 & 2 & -2 & 2 \\ 0 & 1 & 1 & 4 & 6 \end{bmatrix}$ has x_1 and x_2 in the role of basic variables, x_3 and x_4 as free variables. We set the free variables to zero, and read off $x_1 = 2$, $x_2 = 6$. The vector $\mathbf{x} = (2, 6, 0, 0)$ is feasible, and the value of the cost function is 8.

From this feasible $\mathbf{x}$, a corner of the feasible set, we progress to an adjacent corner via pivoting. We must first decide which free variable becomes basic. The top row of our RREF matrix can be interpreted as saying

$$z + x_3 + x_4 = 8.$$

The relationship between z, our variable to be optimized, and the current free variables is uncluttered by the presence of basic variables in the equation, a byproduct of reduced row echelon form. Both x_3 and x_4 have coefficients 1, so a rise in either one has the same effect of forcing a fall in z. (If one had a larger coefficient, its effects on z would be stronger.) So, we can opt to pivot on either one, and I choose here to migrate x_3 from free to basic variable.

To decide on which basic variable becomes free, we look at ratios

$$\frac{\text{current size of basic variable}}{\text{coefficient of } x_3 \text{ in determining the basic variable}}.$$

We only consider positive ratios, and the smallest one is the winner:

For basic variable x_1, the ratio (RHS)/(x_3 coeff) $= \frac{2}{2} = 1$.
For basic variable x_2, the ratio (RHS)/(x_3 coeff) $= \frac{6}{1} = 6$.

So, as x_3 moves from free to basic, x_1 moves from basic to free.

Our new tableau and RREF:

z	x_3	x_2	x_1	x_4	RHS
1	-2	-1	-1	-1	0
0	2	0	1	-2	2
0	1	1	0	4	6

leading to RREF $\begin{bmatrix} 1 & 0 & 0 & -1/2 & 2 & 7 \\ 0 & 1 & 0 & 1/2 & -1 & 1 \\ 0 & 0 & 1 & -1/2 & 5 & 5 \end{bmatrix}$

The relevant submatrix is $\begin{bmatrix} 1 & 0 & 1/2 & -1 & 1 \\ 0 & 1 & -1/2 & 5 & 5 \end{bmatrix}$ with the first two columns now representing x_3 and x_2, our basic variables, in that order. Taking free variables x_1 and x_4 to be zero, we have solution $\mathbf{x} = (0, 5, 1, 0)$, a feasible vector with associated cost $z = 0 + 5 + 2(1) + 0 = 7$, a decrease from the cost of 8 at the previous corner $(2, 6, 0, 0)$.

Repeating our analysis on the latest RREF matrix, we see x_4 should more from free to basic status (has the only positive coefficient among free variables in the first row), and that x_2 is the basic variable that will move to free. (Two RHS/coefficient ratios are viewed, with the one corresponding to basic variable x_3 being $1/(-1) = -1$, and the other, corresponding to x_2 being the only positive one at $5/5 = 1$.) Trading x_4 and x_2 columns, our newest tableau is

z	x_3	x_4	x_1	x_2	RHS
1	-2	-1	-1	-1	0
0	2	-2	1	0	2
0	1	4	0	1	6

with RREF

$$\begin{bmatrix} 1 & 0 & 0 & -0.3 & -0.4 & 5 \\ 0 & 1 & 0 & 0.4 & 0.2 & 2 \\ 0 & 0 & 1 & -0.1 & 0.2 & 1 \end{bmatrix}$$

This yields feasible vector $\mathbf{x} = (0, 0, 2, 1)$, with corresponding cost $z = 2(2) + 1 = 5$. This is the optimal $\mathbf{x}$, as there are no positive coefficients for free variables x_1, x_2 in the relationship (as expressed in the top line of the most recent RREF) $z - (0.3)x_1 - (0.4)x_2 = 5$.

■

Standard form

In the problem

$$\min \quad \mathbf{c}^{\mathrm{T}}\mathbf{x} \qquad \text{subject to} \qquad \mathbf{A}\mathbf{x} \geqslant \mathbf{b}, \;\; \mathbf{x} \geqslant \mathbf{0}, \tag{2}$$

we assume $\mathbf{A}$ is m-by-n, and that the rank of $\mathbf{A}$ is m (so the reduced row echelon form of $\mathbf{A}$ does not have a row of zeros at the bottom). We define m **slack variables** $w_1, \ldots, w_m$ as

$$\mathbf{w} \;=\; \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \;=\; \mathbf{A}\mathbf{x} - \mathbf{b},$$

so that the i^{th} row of $\mathbf{A}\mathbf{x} \geqslant \mathbf{b}$ which read

$$a_{i1}x_1 + \ldots + a_{in}x_n \;\geqslant\; b_i \qquad \text{is translated to} \qquad \begin{cases} a_{i1}x_1 + \ldots + a_{in}x_n - w_i \;=\; b_i \\ w_i \;\geqslant\; 0 \end{cases}$$

Standard form is the reformulated optimization problem

$$\text{minimize} \quad \tilde{\mathbf{c}}^{\mathrm{T}}\tilde{\mathbf{x}} \qquad \text{subject to} \qquad \tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{b}, \quad \tilde{\mathbf{x}} \geqslant \mathbf{0}, \tag{3}$$

involving the new m-by-$(m+n)$ matrix

$$\tilde{\mathbf{A}} \;=\; \begin{bmatrix}\mathbf{A} & (-\mathbf{I})\end{bmatrix}, \qquad \tilde{\mathbf{x}} \;=\; \begin{bmatrix}\mathbf{x}\\ \mathbf{w}\end{bmatrix} \in \mathbb{R}^{m+n}, \qquad \text{and} \qquad \tilde{\mathbf{c}} \;=\; \begin{bmatrix}\mathbf{c}\\ \mathbf{0}\end{bmatrix} \in \mathbb{R}^{m+n}.$$

Several modifications:

- If the goal is to *maximize*, not minimize, the objective function $\mathbf{c}^{\mathrm{T}}\mathbf{x}$ then this deviation from standard form is erased by seeking, instead, to *minimize* $(-\mathbf{c})^{\mathrm{T}}\mathbf{x}$.

- Our modified $\tilde{\mathbf{A}}$, $\tilde{\mathbf{x}}$, $\tilde{\mathbf{c}}$ have been described under the assumption that each of the constraints involving entries from $\mathbf{A}$ are inequalities of the greater-than-or-equal type:

  $$a_{i1}x_1 + \ldots + a_{in}x_n \;\geqslant\; b_i.$$

  If, instead, it is of the less-than-or-equal variety, we replace

  $$a_{i1}x_1 + \ldots + a_{in}x_n \;\leqslant\; b_i \qquad \text{with} \qquad \begin{cases} a_{i1}x_1 + \ldots + a_{in}x_n + w_i \;=\; b_i \\ w_i \;\geqslant\; 0 \end{cases}$$

  And if there is an equality rather than inequality constraint, as in

  $$a_{i1}x_1 + \ldots + a_{in}x_n \;=\; b_i,$$

  then there is no need to introduce the slack variable w_i at all.

The simplex method, our 2nd look