1 Linear Programming (Section 11.8)

In this section, we'll briefly go an overview of linear programming. There is both a linear convex objective and linear convex constraints. Let's define the following quantities:

$$c \in \mathbb{R}^n, \qquad x \in \mathbb{R}^n, \qquad A \in \mathbb{R}^{m\times n}, \qquad b \in \mathbb{R}^m.$$

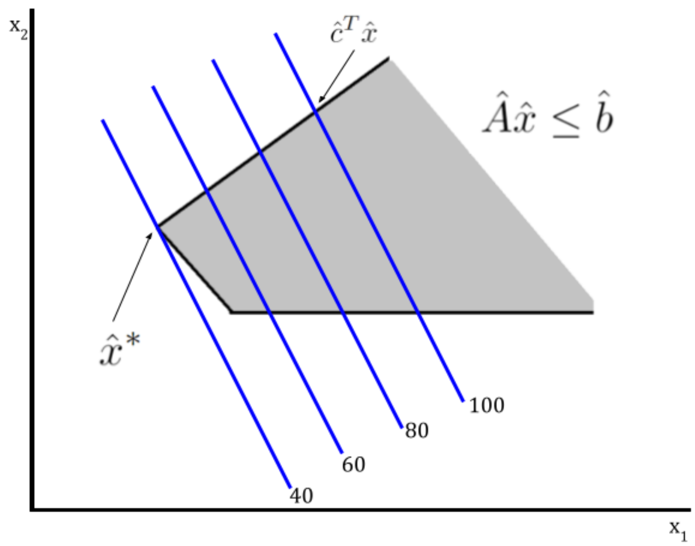
The linear program in **standard form** is defined by

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

In *practice*, however, problems are often formulated in non-standard form:

$$\underset{\hat{x}}{\text{minimize}}\ \hat{c}^T \hat{x} \quad \text{subject to} \quad \hat{A}\hat{x} \leq \hat{b}.$$

Suppose we defined a feasible region (defined by the grey region). The objective function is a linear function, and its contour lines are parallel lines that go through the feasible region. Each line is defined by $\hat{c}^T\hat{x}$ for different values of x_1 or x_2.



Suppose that the objective function decreases to the left and increases to the right. Then, the optimal point (minimizing point) $\hat{x}^*$ is the point that lies in the constraint region but is minimizing the objective function. In this sense, linear programs can either have

- unique solutions (like what we have in the visualization above),
- no solutions (there's two cases to this: infeasible, and unbounded objective), and
- infinitely many solutions.

1.1 Non-Standard to Standard

To bring our non-standard formulation to its standard form, we want to use "slack" variables $z \in \mathbb{R}^m$ so that we can essentially remove the inequality. That is,

$$\hat{A}\hat{x} \leq \hat{b} \implies \hat{A}\hat{x} + z = \hat{b}, \quad z \geq 0.$$

We want to decompose $\hat{x}$ into its positive and negative parts,

$$\hat{x} = x^+ - x^-,$$

such that $x^+ = \max(x, 0) \geq 0$, $x^- = \max(-x, 0) \geq 0$. Note that this notation isn't very precise; for the x^+ case, what this means is that, for each element $x \in \hat{x}$, we put $\max(x, 0)$ into the corresponding index in x^+.

(Example.) Suppose

$$\hat{x} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}.$$

We can decompose it as follows:

$$\hat{x} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{x^+} - \underbrace{\begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}}_{x^-}.$$

Let

$$c = \begin{bmatrix} \hat{c} \\ -\hat{c} \\ 0 \end{bmatrix}, \quad x = \begin{bmatrix} x^+ \\ x^- \\ z \end{bmatrix}$$

so that

$$c^T x = \hat{c}^T x^+ - \hat{c}^T x^- = \hat{c}^T \hat{x}.$$

We can define the matrix A and vector b as

$$A = \begin{bmatrix} \hat{A} & -\hat{A} & I_{m\times m} \end{bmatrix}, \quad \hat{b} = b.$$

Thus, we can reformulate the minimization problem mentioned above as

$$\begin{aligned} &\underset{x}{\text{minimize}}\ \hat{c}^T \hat{x} \quad \text{subject to} \quad \hat{A}\hat{x} \le \hat{b} \\ \iff &\underset{x}{\text{minimize}}\ c^T x \quad \text{subject to} \quad Ax = b, \quad x \ge 0 \end{aligned},$$

as desired. Keep in mind that this problem got a lot bigger: the quantities in standard form are larger than the quantities in non-standard form. For example, the vector x in the equivalent problem is about three times as large as $\hat{x}$, since it *stacks* the positive components, negative components, and the slack variable.

1.2 Solving the Problem: Optimality Conditions

We can use the optimality conditions to find a solution. We can use the **Lagrangian function**, which arguments the objective with *constraints* and "new" variables called **Lagrange multipliers**,

$$\lambda \in \mathbb{R}^m, \qquad s \in \mathbb{R}^n, \qquad s \ge 0.$$

For the problem in standard form, the Lagrangian function is defined by

$$L(x, \lambda) = c^T x - \lambda^T (Ax - b) - s^T x.$$

At a solution, $\nabla_x L(x^*, \lambda^*) = 0$ and $\nabla_\lambda L(x^*, \lambda^*) = 0$. This is valid as long as x^* is a feasible point. In particular,

$$\nabla_x L = 0 = c - A^T x^* - s^* \implies A^T x^* + s^* = c,$$
$$\nabla_\lambda L = 0 = -(Ax^* - b) \implies Ax^* = b.$$

Additionally, we impose a requirement on x^* and s^*:

$$x^* \ge 0, \qquad s^* \ge 0, \qquad s_i^* \cdot x_i^* = 0 \quad (1 \le i \le n)$$

In other words, if s is zero, then x is not zero, and vice versa. This means that

$$x^{*T} s^* = 0.$$

The optimality conditions are sufficient enough to define a global minimizer for the problem:

Suppose $\overline{x}$ is a feasible point, i.e., $A\overline{x} = b$ for $\overline{x} \geq 0$. Then,

$$\begin{aligned} c^T\overline{x} &= (A^T\lambda^* + s^*)^T\overline{x} \\ &= \lambda^* A\overline{x} + s^{*T}\overline{x} \\ &= \lambda^{*T} b + s^{*T}\overline{x} \\ &\geq \lambda^* b. \end{aligned}$$

At the solution,

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

Thus, for any feasible point, $c^T\overline{x} \geq c^T x^*$, and x^* is a global minimum.