Linear Programming and Simplex Method

We give here a brief introduction to linear programming and the simplex method. The problem is to find a solution of

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{subject} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0. \end{array}$$

This optimization problem arises in many settings, including in machine learning, statistics, economics, business logistics, optimal transport, physics, and engineering [1–3]. While the objective function and the constraints being linear may seem to suggest a simple solution, in practice these problems exhibit interesting behaviors and challenges for efficient numerical solution. As we shall also discuss, many interesting problems can also be transformed to be expressed as linear programs.

Primal and Dual Problems for Linear Programming

We first derive some necessary and sufficient conditions to achieve the minimizer of the linear program based on the *Karuch-Kahn-Tucker (KKT) conditions*. We consider the KKT conditions for the constrained problem

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{subject} & \mathbf{h}(\mathbf{x}) = 0 \\ & \mathbf{g}(\mathbf{x}) \leq 0. \end{array}$$

This has the Lagrangian

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) + \mathbf{s}^T \mathbf{g}(\mathbf{x}), \tag{1}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^n$ and $\mathbf{s} \in \mathbb{R}^m$ with $\mathbf{s} \geq 0$. The linear program (LP) has the Lagrangian

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s}) = \mathbf{c}^T\mathbf{x} - \boldsymbol{\lambda}^T (A\mathbf{x} - \mathbf{b}) - \mathbf{s}^T \mathbf{x}, \tag{2}$$

where we let $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$, $\mathbf{h}(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$, and $g(\mathbf{x}) = -\mathbf{x}$. This can also be expressed as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s}) = \mathbf{b}^T\boldsymbol{\lambda} - \mathbf{x}^T \left(A^T\boldsymbol{\lambda} + \mathbf{s} - \mathbf{c}\right). \tag{3}$$

We obtain the KKT conditions for LP by computing $\nabla_{\mathbf{x},\boldsymbol{\lambda}}\mathcal{L} = 0$ and $\nabla_{\mathbf{s}}\mathcal{L} \leq 0$. This yields

$$\nabla_{\mathbf{x}}\mathcal{L} = \mathbf{c} - A^T\boldsymbol{\lambda} - \mathbf{s} = 0, \quad \nabla_{\boldsymbol{\lambda}}\mathcal{L} = A\mathbf{x} - \mathbf{b} = 0, \quad \nabla_{\mathbf{s}}\mathcal{L} = \mathbf{x} \geq 0, \quad \mathbf{s} \geq 0, \quad \mathbf{s}^T\mathbf{x} = 0. \tag{4}$$

The KKT conditions for the LP are

$$A^T\boldsymbol{\lambda} + \mathbf{s} = \mathbf{c} \tag{5}$$
$$A\mathbf{x} - \mathbf{b} = 0 \tag{6}$$
$$\mathbf{x} \geq 0 \tag{7}$$
$$\mathbf{s} \geq 0 \tag{8}$$
$$s_i x_i = 0. \tag{9}$$

For a triple $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \mathbf{s}^*)$ that satisfies the KKT conditions, we have

$$\mathbf{c}^T\mathbf{x}^* = (A\boldsymbol{\lambda}^* + \mathbf{s}^*)^T \mathbf{x}^* = (A\mathbf{x}^*)^T \boldsymbol{\lambda}^* = \mathbf{b}^T\boldsymbol{\lambda}^* \quad \Rightarrow \quad \mathbf{b}^T\boldsymbol{\lambda}^* = \mathbf{c}^T\mathbf{x}^*. \tag{10}$$

If we instead consider any feasible point $\bar{\mathbf{x}}$, we have for the triple $(\bar{\mathbf{x}}, \boldsymbol{\lambda}^*, \mathbf{s}^*)$ that

$$\mathbf{c}^T\bar{\mathbf{x}} = (A\boldsymbol{\lambda}^* + \mathbf{s}^*)^T \bar{\mathbf{x}} = (A\bar{\mathbf{x}})^T \boldsymbol{\lambda}^* + \mathbf{s}^{*T}\bar{\mathbf{x}} = \mathbf{b}^T\boldsymbol{\lambda}^* + \mathbf{s}^{*T}\bar{\mathbf{x}} = \mathbf{c}^T\mathbf{x}^* + \mathbf{s}^{*T}\bar{\mathbf{x}}, \tag{11}$$

which yields

$$\mathbf{c}^T\bar{\mathbf{x}} = \mathbf{c}^T\mathbf{x}^* + \mathbf{s}^{*T}\bar{\mathbf{x}}. \tag{12}$$

Since $\mathbf{s}^{*,T}\bar{\mathbf{x}} \geq 0$, we have $\mathbf{c}^T\bar{\mathbf{x}} \geq \mathbf{b}^T\boldsymbol{\lambda}^* = \mathbf{c}^T\mathbf{x}^*$. We see that when $\mathbf{s}^{*T}\bar{\mathbf{x}} \neq 0$ the inequality is strict. This also shows that $\bar{\mathbf{x}}$ is a minimizer if and only if $\mathbf{s}^{*T}\bar{\mathbf{x}} = 0$. As a consequence, if we can find a triple $(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s})$ that satisfies the KKT conditions in equation 5- 9, we will have found a minimizer for the LP.

These considerations are also related to the dual optimization problem. For the LP, the dual problem can be obtained from

$$\begin{aligned} &\max_{\boldsymbol{\lambda},\mathbf{s}} \; q(\boldsymbol{\lambda}, \mathbf{s}) \\ &\text{subject } \mathbf{s} \geq 0, \end{aligned} \tag{13}$$

where $q(\boldsymbol{\lambda}, \mathbf{s}) = \inf_{\mathbf{x}\in\mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s})$. For LP, the dual objective function is

$$q(\boldsymbol{\lambda}, \mathbf{s}) = \inf_{\mathbf{x}\in\mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s}) = \begin{cases} \mathbf{b}^T\boldsymbol{\lambda}, & \text{when } A^T\boldsymbol{\lambda} + \boldsymbol{s} = \mathbf{c}. \\ -\infty, & \text{otherwise.} \end{cases} \tag{14}$$

This follows, since if we have a non-zero residual $\mathbf{v} = A^T\boldsymbol{\lambda} + \boldsymbol{s} - \mathbf{c} \neq 0$, we can take $\mathbf{x}(\beta) = \beta\mathbf{v}$. This gives $\mathcal{L}(\mathbf{x}(\beta), \boldsymbol{\lambda}, \mathbf{s}) = \mathbf{b}^T\boldsymbol{\lambda} - \beta\|\mathbf{v}\|^2 \to -\infty$ as $\beta \to \infty$. When $\mathbf{v} = 0$ we see the Lagrangian is simply $\mathbf{b}^T\boldsymbol{\lambda}$. Since $q(\boldsymbol{\lambda}, \mathbf{s}) = -\infty$ when $A^T\boldsymbol{\lambda} + \boldsymbol{s} \neq \mathbf{c}$, we can add this as a constraint above and obtain an optimization problem with the same solutions as the dual problem in equation 13. This yields the dual problem for the LP

$$\begin{array}{ll} \max_{\boldsymbol{\lambda},\mathbf{s}} & \mathbf{b}^T\boldsymbol{\lambda} \\ \text{subject} & A^T\boldsymbol{\lambda} + \boldsymbol{s} = \mathbf{c}, \;\; \mathbf{s} \geq 0. \end{array} \tag{15}$$

Since these solve the same KKT conditions, we have from our results above that for any feasible $\boldsymbol{\lambda}, \mathbf{s}$ to the dual problem and any feasible $\mathbf{x}$ from the primal problem that

$$\mathbf{c}^T\mathbf{x} \geq \mathbf{b}^T\boldsymbol{\lambda}. \tag{16}$$

We already showed that this becomes equal when the KKT conditions are satisfied. This is equivalent to having a triple of points $(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s})$ where $\mathbf{x}$ is feasible for the primal problem and $(\boldsymbol{\lambda}, \mathbf{s})$ is feasible for the dual problem. For such a triple, the $\mathbf{x}$ is a solution minimizing the primal problem and $(\boldsymbol{\lambda}, \mathbf{s})$ are a solution maximizing the dual problem. This provides useful criteria for characterizing the optimal solution and for developing optimization algorithms.

Geometric Intuition of Linear Programming and Constraints

For the LP using the formulation in equation 28, we have that the constraints $A\mathbf{x} = \mathbf{b}$ give a geometry that requires the solution to lie within the intersection of a collection of hyperplanes and within the positive orthant $\mathbf{x} \geq 0$. To help with some intuition, we illustrate some basic cases for different choices of m in three spatial dimensions $n = 3$ in Figure 1. In higher dimensions, while the geometry can become more rich it shares many similar properties.

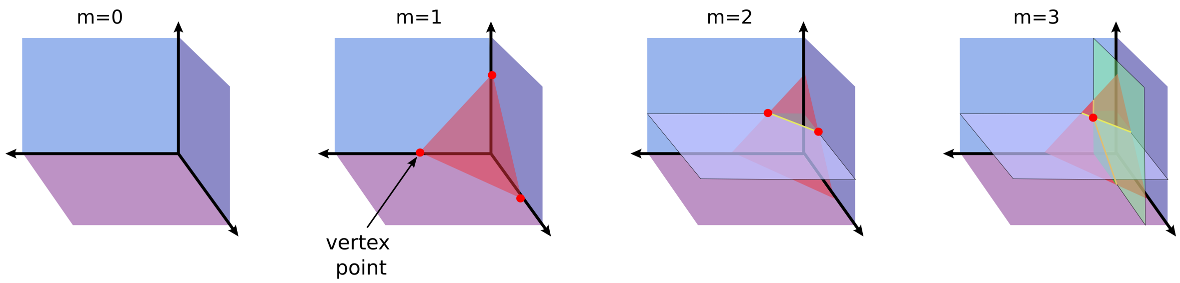

Figure 1: **Geometry of Linear Programming (LP).** We show the basic feasible points that arise from the constraints $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq 0$ *(red points)*.

The condition $\mathbf{x} \geq 0$ requires the solution remain within the positive orthant ($x_i \geq 0$ for all i). In the case of $m = 1$ the constraint $A\mathbf{x} = \mathbf{b}$ requires the solution also to be on a hyperplane of points orthogonal to $\mathbf{u}_1 = (a_{11}, \ldots, a_{1n})$. In the case each a_{1j} is non-zero, this will intersect with all three coordinate axes and form a simplex. Since the objective function $\mathbf{c}^T\mathbf{x}$ is linear it will always achieve its minimum value at one of the vertex points. An interesting property of the vertex points for $m = 1$ is they can only have at most one non-zero component $\mathbf{v} = (v_1, 0, 0)$, $\mathbf{v} = (0, v_2, 0)$, or $\mathbf{v} = (0, 0, v_3)$. We next consider the case with $m = 2$, which involves adding another constraint to $A\mathbf{x} = \mathbf{b}$ requiring the solution also to lie on the hyperplane of points orthogonal to $\mathbf{u}_2 = (a_{21}, \ldots, a_{2n})$. This requires the feasible points to lie on the line of intersection. An interesting property of the vertex points for $m = 2$ is they can only have at most two non-zero components $\mathbf{v} = (v_1, v_2, 0)$, $\mathbf{v} = (v_1, 0, v_3)$, or $\mathbf{v} = (0, v_2, v_3)$. Again, these vertex points arise from the intersection of the hyperplanes and the positive orthant conditions $x_i \geq 0$ for all i. The linear objective function $\mathbf{c}^T\mathbf{x}$ again will always achieve its minimum value at one of the vertex points. For $m = 3$, we introduce an addition constraint which gives an additional hyperplane. This results in the solution having to lie at the intersection of three hyperplanes, so in $n = 3$ if these are consistent this defines a single point. This gives the vertex points for $m = 3$ can have at most three non-zero components $\mathbf{v} = (v_1, v_2, v_3)$, which while not really a restriction it follows the same pattern as above. In particular, for m constraints the vertex points $\mathbf{v}$ have at most m non-zero components. This motivates the following definitions which are closely related to these properties of the vertices.

Def: A *basic feasible point* $\mathbf{x}$ is a point that is feasible and for which there exists a collection of indices $\mathcal{B}$ satisfying the following properties

(i) $|\mathcal{B}| = m$, contains exactly m indices.

(ii) if $i \notin \mathcal{B}$ then $x_i = 0$.

(iii) The $m \times m$ matrix $B = [A_i]_{i \in \mathcal{B}}$ is non-singular.

The A_i denotes the i^{th} column of the matrix A. We call $\mathcal{B}$ the *basis set*. We call the matrix constructed from this subset of columns of A as the *basis matrix*.

It can be proven that the basic feasible points are precisely the vertices of the constraint set of $A\mathbf{x} = b$ and $\mathbf{x} \geq 0$. This provides a useful characterization and way to generate vertex points. Since the objective is linear, the minimum value of the LP must occur at one of the basic feasible points provided the problem is bounded and an optimal solution exists. This suggests one strategy to find a minimizer $\mathbf{x}$ is to search through the basic feasible points. However, from the definitions we see a brute-force search could be expensive since there can be up to $\binom{n}{m}$ basic feasible points. We show how the objective function and feasibility conditions derived above can be used to try to more efficiently guide traversing of the basic feasible points to search for a minimum.

Simplex Method Motivations

Suppose we have a basic feasible point $\mathbf{x}$ for the LP with basis set $\mathcal{B}$. We define the set of indices not in the basis as $\mathcal{N} = \{i \mid i \notin \mathcal{B}\}$. WLOG we can arrange at a given stage our indices so $\mathcal{B} = \{1, 2, \ldots, m\}$ are the first m indices. It is convenient to express the point in terms of components as $\mathbf{x} = (\mathbf{x}_B, \mathbf{x}_N)$, where $\mathbf{x}_B = (x_1, \ldots, x_m)$ and $\mathbf{x}_N = (x_{m+1}, \ldots, x_n)$. We will sometimes abuse notation and also express this as $\mathbf{x} = \mathbf{x}_B + \mathbf{x}_N$.

A good initial step would be to see if we can construct a triple $(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s})$ that satisfies the KKT conditions, since then we would have the minimum. We can construct $\boldsymbol{\lambda}, \mathbf{s}$ from $\mathbf{x}$ using B. We first express $\mathbf{x}_B$ in terms of $\mathbf{b}$ by

$$A\mathbf{x} = \mathbf{b} \;\Rightarrow\; B\mathbf{x}_B + N\mathbf{x}_N = \mathbf{b} \;\Rightarrow\; B\mathbf{x}_B = \mathbf{b} \;\Rightarrow\; \mathbf{x}_B = B^{-1}\mathbf{b}. \tag{17}$$

The N is the $m \times (n - m)$ matrix consisting of the columns of A_i with indices $i \in \mathcal{N}$. We can relate any of these vertices to the $(\boldsymbol{\lambda}, \mathbf{s})$ by using that they need to satisfy the condition $A^T\boldsymbol{\lambda} + \mathbf{s} = \mathbf{c}$. This yields

$$B^T\boldsymbol{\lambda} + \mathbf{s}_B = \mathbf{c}_B, \quad N^T\boldsymbol{\lambda} + \mathbf{s}_N = \mathbf{c}_N. \tag{18}$$

The complimentary conditions would be

$$\mathbf{x}^T\mathbf{s} = \mathbf{x}_B^T\mathbf{s}_B + \mathbf{x}_N^T\mathbf{s}_N = 0. \tag{19}$$

For a non-degenerate basis $\mathcal{B}$ we would have $[\mathbf{x}_B]_i > 0$, which requires $\mathbf{s}_B = 0$. This gives

$$B^T\boldsymbol{\lambda} = \mathbf{c}_B, \quad N^T\boldsymbol{\lambda} + \mathbf{s}_N = \mathbf{c}_N. \tag{20}$$

We can solve these to obtain

$$\boldsymbol{\lambda} = B^{-T}\mathbf{c}_B, \quad \mathbf{s}_N = \mathbf{c}_N - N^T\boldsymbol{\lambda} = \mathbf{c}_N - \left(B^{-1}N\right)^T\mathbf{c}_B. \tag{21}$$

This gives $\boldsymbol{\lambda}$ and $\mathbf{s}$ for the basis $\mathcal{B}$. The only condition not explicitly enforced above is the non-negativity $\mathbf{s} \geq 0$. If we have that this holds for $\mathbf{s}$ then the triple $(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s})$ would satisfy the KKT conditions and we would have that $\mathbf{x}$ is the optimal solution. When this fails, then we must have that one of the components of $\mathbf{s}_N$ is negative. Also notice that the above construction for $\boldsymbol{\lambda}$ and $\mathbf{s}$ do not depend on $\mathbf{x}$ but only on the choice of basis $\mathcal{B}$.

Similar to the calculation in equation 11, we also have for this $\boldsymbol{\lambda}, \mathbf{s}$ that for any feasible $\mathbf{x}$ the objective function satisfies

$$\mathbf{c}^T\mathbf{x} = \mathbf{b}^T\boldsymbol{\lambda} + \mathbf{s}^T\mathbf{x} = \mathbf{b}^T\boldsymbol{\lambda} + \mathbf{s}_N^T\mathbf{x}_N. \tag{22}$$

Since $\mathbf{s}_N$ has at least one negative component corresponding to an index $q \notin \mathcal{B}$, this suggests if we are at the basic feasible point for $\mathcal{B}$ we can make the objective function even smaller by increasing the q^{th} component so that $x_q > 0$, so that $\mathbf{s}_N^T\mathbf{x}_N = s_q x_q < 0$. We mention again that we are assuming throughout we are in the non-degenerate case.

Since we need to keep $\mathbf{x}(x_q)$ a feasible point, this would also require adjusting the other components of $\mathbf{x}$ so that $A\mathbf{x}(x_q) = \mathbf{b}$ is maintained. We now show how it is possible to do this. We use that B is non-singular which implies that all of its columns $A_1, \ldots, A_m$ are linearly independent and form a basis for $\mathbb{R}^m$. This allows us to express the q^{th} column as

$$A_q \;=\; \sum_{i=1}^m v_i A_i, \Rightarrow \;\sum_{i=1}^m v_i A_i - A_q = 0. \tag{23}$$

WLOG, we arrange the indexing so that $q = m + 1$. This shows that if we let $\mathbf{v} = (v_1, v_2, \ldots, v_m, -1, 0, \ldots, 0)$ then $A\mathbf{v} = 0$. Since we have the constraints $\mathbf{x} \geq 0$, it is convenient to let $\mathbf{w} = -\mathbf{v} = (w_1, w_2, \ldots, w_m, 1, 0, \ldots, 0)$, where $w_i = -v_i$. We now consider $\mathbf{x}(\epsilon) = \mathbf{x}_0 + \epsilon\mathbf{w} = \mathbf{x}_0 - \epsilon\mathbf{v}$, which satisfies $A\mathbf{x}(\epsilon) = \mathbf{b}$ for any ϵ provided $\mathbf{x}_0$ is feasible. When $\mathbf{x}_0$ is taken to be the basic feasible point for $\mathcal{B}$ we have the q^{th} component is given by $[\mathbf{x}(\epsilon)]_q = x_q = \epsilon$, since $[\mathbf{x}_0]_q = 0$. For $\mathbf{x}(\epsilon)$ to be feasible for the LP, we also need to have $\mathbf{x}(\epsilon) \geq 0$. In the non-degenerate case, we have components

$$\begin{array}{ll} (\mathbf{x}(\epsilon))_i = x_i - \epsilon v_i, & i \in \mathcal{B} \\ (\mathbf{x}(\epsilon))_q = \epsilon, & i = q \\ (\mathbf{x}(\epsilon))_i = 0, & i \in \mathcal{N} \setminus \{q\}. \end{array} \tag{24}$$

To reduce the objective, we try to take $x_q = \epsilon$ as large as possible until one of the components becomes zero, $x_i - \epsilon v_i = 0$. Making x_q any larger would result in violating the constraint $\mathbf{x} \geq 0$. In particular, we take

$$x_q = \min\left\{\frac{x_i}{v_i} \mid v_i > 0\right\}, \;\; p = \arg\min_i\left\{\frac{x_i}{v_i} \mid v_i > 0\right\}. \tag{25}$$

If there are no $v_i > 0$ then we can make x_q as large as we like and we can conclude that the objective function is unbounded from equation 22. If the objective is bounded, the index p

denotes the component for which the zero occurs and we take $x_q = \epsilon = x_p/v_p$. If there is a tie, we take the index of smallest value. WLOG we take $p = m-1$ and $q = m+1$ for convenience, which always can be arranged by adjusting the index ordering.

Now the new point we end up at is of the form

$$\mathbf{x}(\epsilon) = (x_1(\epsilon), \ldots, x_{p-1}(\epsilon),\ 0,\ x_{p+1}(\epsilon),\ x_q(\epsilon),\ 0, \ldots, 0),$$

where $x_i(\epsilon) = x_i - \epsilon v_i \geq 0$, $x_q = \epsilon \geq 0$, and $x_p(\epsilon) = 0$. This point only has m non-zero components making it a good candidate for a basic feasible point. To construct the basis set we take $\mathcal{B}^+ = (\mathcal{B} \setminus \{p\}) \cup \{q\} = \{1, \ldots, p-1, p+1, q\}$. This is obtained by removing the index p and adding the index q.

The main property we have to verify is that $B^+ = [A_i]_{i\in\mathcal{B}^+}$ is non-singular. We need to show $\sum_{i\in\mathcal{B}^+} \beta_i A_i = 0$ implies $\beta_i = 0$. This follows from $v_p \neq 0$ and equation 23, which gives

$$0 \;=\; \sum_{i\in\mathcal{B}^+, i\neq q} \beta_i A_i + \beta_q A_q = \sum_{i\in\mathcal{B}, i\neq p} (\beta_i + v_i\beta_q)\, A_i + \beta_q v_p A_p \tag{26}$$

$$\Rightarrow \quad \beta_q v_p = 0,\ (\beta_i + v_i\beta_q) = 0, \Rightarrow \quad \beta_q = 0 \quad \Rightarrow \beta_i = 0. \tag{27}$$

We used the linear independence of $\mathcal{B}$. This shows that $\mathcal{B}^+$ is a basis set and that $\mathbf{x}(\epsilon)$ is a basic feasible point for $\epsilon = x_q = x_p/v_p$. From equation 22, we see in this case that the new basic feasible point has smaller value for the objective. We obtained this result under the assumption that the basis set was non-degenerate and A had full rank m. In practice, these assumptions can be relaxed with some additional technical considerations for handling potential cycling and other issues, which we do not go into detail here. The above was meant to help motivate the intuition for the steps of the simplex method in one of the most common cases.

Now we can again use equations 5- 9 to construct for the new feasible point $\mathbf{x}^+$ the $\boldsymbol{\lambda}^+, \mathbf{s}^+$ to see if the triple $(\mathbf{x}^+, \boldsymbol{\lambda}^+, \mathbf{s}^+)$ satisfy the KKT conditions. In particular, if $\mathbf{s}^+ \geq 0$ holds, then the $\mathbf{x}^+$ would be the optimal solution to the LP. The simplex method effectively uses the procedures we mentioned above successively until either finding a point that satisfies the KKT conditions or determines the problem is unbounded.

Simplex Method: Summary

Start with a basic feasible point $\mathbf{x}$ and see if is optimal by checking the KKT conditions. If KKT is satisfied, then we are done and $\mathbf{x}$ is optimal. Otherwise, proceed by performing the following modifications to the basis set $\mathcal{B} = \{1, 2, \ldots m\}$.

(i) Choose an index $q \in \mathcal{N}$ for which $s_q < 0$, which we call the *entering index*.

(ii) Construct $\mathbf{x}(\epsilon) = \mathbf{x} - \epsilon\mathbf{v}$, with $\mathbf{v} = (v_1, \ldots, v_m, -1, 0, \ldots, 0)$ by adding A_q to the column set so $\{A_1, \ldots, A_m, A_q\}$ is linearly dependent and $A_q = \sum_{i=1}^m v_i A_i$.

(iii) Increase $x_q = \epsilon$ away from zero until one of the components of $\mathbf{x}_B$ becomes zero, $x_q^+ = \min\left\{\frac{x_i}{v_i} \mid v_i > 0\right\}$, $p = \arg\min_i \left\{\frac{x_i}{v_i} \mid v_i > 0\right\}$. Let $\epsilon^+ = x_q^+ = x_p/v_p$. If no such $x_q > 0$ exists, then the objective function can be made arbitrarily negative and we halt by declaring the LP is unbounded.

(iv) Remove the index p (*leaving index*) from $\mathcal{B}$ and replace it with the *entering index* q to obtain the set of indices $\mathcal{B}^+$.

(v) The $\mathcal{B}^+$ gives the *basis set* for the new basic feasible point $\mathbf{x}^+ = \mathbf{x}(\epsilon^+)$. If it satisfies the KKT conditions ($s_i^+ \geq 0$ for all i), then we have found the minimum and are done. If not, then proceed to step (i) to move to a new basic feasible point.

The above procedure is sometimes referred to as *pivoting*. This moves from one vertex of the polytope to another vertex. In practice, degenerate bases can be encountered, in which case additional technical steps are needed to prevent cycling and other issues, which we do not discuss here. For more details, see [1–3]. In the non-degenerate case with A having full rank m this results in a strict decrease in the objective function each step. Since there are only a finite number of basic feasible points, the algorithm will eventually terminate by declaring unboundedness or finding an optimal solution.

Two-Phase Method for LPs

An important issue that arises in practice for many LP problems is to find an initial basic feasible point. For this purpose an LP with *artificial variables* $\mathbf{y}$ is introduced that has an easy to find initial basic feasible point. The design of this LP is such that it has solutions that are basic feasible points of the original LP problem. The *artificial LP problem* is

$$\begin{array}{ll} \min & y_1 + y_2 + \cdots + y_m \\ \text{subject} & [A, I_m][\mathbf{x}; \mathbf{y}] = \mathbf{b} \\ & [\mathbf{x}; \mathbf{y}] \geq 0. \end{array}$$

In our notation $[\mathbf{x}; \mathbf{y}]$ indicates row concatenation of the vectors to form a larger column vector. The $[A, I_m]$ denotes column concatenation to form an enlarged matrix. The I_m indices the $m \times m$ identity matrix. This problem will always have the feasible initial point $\mathbf{x} = [0; \mathbf{b}]$. If the original problem has a feasible point $\mathbf{x}$ then this LP problem has minimizer $[\mathbf{x}; 0]$ with $\mathbf{y} = 0$ and objective 0. If no such minimizer can be obtained, we can conclude the original problem does not have any feasible points.

We split solving an LP into two phases. In the first phase, we transform the LP into the artificial problem and apply the simplex method to search for the solution $[\mathbf{x}; 0]$. If no such solution exists, we declare the LP is infeasible. If such a solution exists, we use this solution as the initial basic feasible point $\mathbf{x}$ for the original LP problem. In the second phase, we proceed to solve the LP by using $\mathbf{x}$ as the initial basic feasible point in the simplex method. In this way, we can handle general LP problems in the standardized form even when there is not an obvious initial basic feasible point.

Tableau Notation

For book-keeping the steps of the simplex method various types of tableau notations are used. We discuss a few of the most common ones here along with how they can be interpreted. The starting point for encoding the simplex method is to construct a table of the standardized problem for $A, \mathbf{b}, \mathbf{c}$ of the form

$$\begin{array}{c|cccc} & \mathbf{a}_1 & \cdots & \mathbf{a}_n & \mathbf{b} \\ \hline & a_{11} & \cdots & a_{1n} & b_1 \\ & \vdots & & \vdots & \vdots \\ & a_{m1} & \cdots & a_{mn} & b_m \\ \mathbf{c}^T & c_1 & \cdots & c_n & 0 \end{array} \tag{28}$$

This corresponds to the composite matrix

$$\begin{bmatrix} A & \mathbf{b} \\ \mathbf{c}^T & 0 \end{bmatrix} = \begin{bmatrix} B & N & \mathbf{b} \\ \mathbf{c}_B^T & \mathbf{c}_N^T & 0 \end{bmatrix}. \tag{29}$$

In this representation we take the basis set to be $\mathcal{B} = \{1, \ldots, m\}$ without loss of generality. Now for the basis set, we want to convert all columns to be represented over the basis vectors $A_1, A_2, \ldots, A_m$. This is equivalent to computing $\mathbf{y}_i = B^{-1}A_i$ which gives components $[\mathbf{y}_i]_j = y_{ij}$. In the special case of indices with $i \in \mathcal{B}$, we obtain $\mathbf{y}_i = \mathbf{e}_i$ where $\mathbf{e}_i$ is the vector of all zeros except for a one in component i. This can also be expressed as $y_{ji} = \delta_{ij}$ for the Kronecker delta function. For $i \notin \mathcal{B}$, we obtain y_{ij} from $B^{-1}A_i$. We also represent $\mathbf{b}$ over the basis as $\mathbf{y}_0 = B^{-1}\mathbf{b}$.

This conversion can be computed through the action of B^{-1}, which is equivalent to performing row reductions of the entire matrix so the components of B are transformed it into the identity matrix I_m. This would be equivalent to multiplying on the left to obtain

$$\begin{bmatrix} B^{-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} B & N & \mathbf{b} \\ \mathbf{c}_B^T & \mathbf{c}_N^T & 0 \end{bmatrix} = \begin{bmatrix} I_m & B^{-1}N & B^{-1}\mathbf{b} \\ \mathbf{c}_B^T & \mathbf{c}_N^T & 0 \end{bmatrix} = \begin{bmatrix} I_m & [y_{ji}]_{i\notin\mathcal{B}} & [y_{j0}] \\ \mathbf{c}_B^T & \mathbf{c}_N^T & 0 \end{bmatrix}. \tag{30}$$

The $[y_{ji}]_{i\notin\mathcal{B}}$ give the matrix of components of the columns A_i represented over the basis. The $[y_{j0}]$ gives the vector of components for $\mathbf{b}$ represented over the basis. We next perform further row operations to reduce the $\mathbf{c}_B^T$ term to zero. This can be accomplished by

$$\begin{bmatrix} I_m & 0 \\ -\mathbf{c}_B^T & 1 \end{bmatrix} \begin{bmatrix} I_m & B^{-1}N & B^{-1}\mathbf{b} \\ \mathbf{c}_B^T & \mathbf{c}_N^T & 0 \end{bmatrix} = \begin{bmatrix} I_m & B^{-1}N & B^{-1}\mathbf{b} \\ 0^T & \mathbf{c}_N^T - \mathbf{c}_B^T B^{-1}N & -\mathbf{c}_B^T B^{-1}\mathbf{b} \end{bmatrix}. \tag{31}$$

This was done to obtain some useful terms in the last row that are helpful for determining the pivot operations. This provides the final form for the *canonical tableau*

$$\begin{bmatrix} I_m & B^{-1}N & B^{-1}\mathbf{b} \\ 0^T & \mathbf{s}_N^T & -\mathbf{c}_B^T B^{-1}\mathbf{b} \end{bmatrix}. \tag{32}$$

The $\mathbf{s}_N^T = \mathbf{c}_N^T - \mathbf{c}_B^T B^{-1}N$ are the reduced cost coefficients used to determine the entering index q. We also remark that in the lower right is the negative objective function value $\mathbf{c}^T\mathbf{x} = \mathbf{c}_B^T B^{-1}\mathbf{b}$ for the current basic feasible point $\mathbf{x} = (\mathbf{x}_B, \mathbf{x}_N)$. The upper right also gives the non-zero components of the basic feasible point $\mathbf{x}_B = B^{-1}\mathbf{b}$. The utility of the tableau is that it presents all the quantities explicitly that we need to decide each pivoting step of the simplex method.

We next discuss how the *canonical tableau* is transformed after each pivoting operation from the basis $\mathcal{B}$ to the basis $\mathcal{B}^+$. While in principle, we could recompute the tableau from scratch each iteration this would be inefficient. Instead, we can use how the representations transform when we change the basis by removing the column A_p and adding to the basis the column A_q. In particular,

$$A_q = \sum_{i=1}^{m} y_{iq} A_i = \sum_{i=1,i\neq p}^{m} y_{iq} A_i + y_{pq} A_p \Rightarrow \quad A_p = \frac{1}{y_{pq}} A_q - \sum_{i=1,i\neq p}^{m} \frac{y_{iq}}{y_{pq}} A_i. \tag{33}$$

For any vector $\mathbf{u}$ we have

$$\mathbf{u} = \sum_{i=1}^{m} u_i A_i = \sum_{i=1,i\neq p}^{m} u_i A_i + u_p A_p = \sum_{i=1,i\neq p}^{m} \left(u_i - \frac{y_{iq}}{y_{pq}} u_p \right) A_i + \frac{u_p}{y_{pq}} A_q. \tag{34}$$

In the case that $\mathbf{u} = A_j = \mathbf{y}_j$ or $\mathbf{u} = \mathbf{b} = \mathbf{y}_0$ we obtain the transformation rule

$$\begin{aligned} y_{ij}^{+} &= y_{ij} - \frac{y_{iq}}{y_{pq}} y_{pj}, \;\; \text{if } i \neq p \\ y_{pj}^{+} &= \frac{y_{pj}}{y_{pq}}, \;\; \text{if } i = p, \end{aligned} \tag{35}$$

where $j = 0, 1, \ldots, n$. This is equivalent to multiplying the row p by y_{iq}/y_{pq} and subtracting it from row i, when $i \neq p$. In the case of row $i = p$, this row is simply modified by multiplying it by the factor $1/y_{pq}$. To handle the last row of cost reduction coefficients, we can compute $z_j^+ = [\mathbf{c}_B^T B^{-1} N]_j = \mathbf{c}_B^T \mathbf{y}_j^+ = \sum_{i=1}^m c_i y_{ij}^+$ to obtain $s_j^+ = [\mathbf{s}_N^{T+}]_j = c_j - z_j^+$. We obtain the last component of the row giving the negative objective function by computing $-z_0^+ = \mathbf{c}_B^T \mathbf{y}_0^+$. This gives all the row operations to perform to update the *canonical tableau* for the basis $\mathcal{B}$ to the basis $\mathcal{B}^+$.

There are also more parsimonious representations often used in calculations. This is based on the key driver of each step of the simplex method being B^{-1}. The basis set of variables and components of this matrix are the main items to be tracked. In matrix form $[B^{-1}, \mathbf{y}_0]$. This gives the *revised tableau* notation for basis $\mathcal{B} = \{i_1, i_2, \ldots, i_m\}$ of the form

variable	B^{-1}			$\mathbf{y}_0$
x_{i_1}	$\beta_{i_1 i_1}$	$\cdots$	$\beta_{i_1 i_m}$	$y_{i_1,0}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_{i_m}	$\beta_{i_m i_1}$	$\cdots$	$\beta_{i_m i_m}$	$y_{i_m,0}$

(36)

This notation tracks $[B^{-1}]_{ij} = \beta_{ij}$ with β_{ij} the components of the inverse matrix. This also tracks the representation of $\mathbf{b}$ over this basis, $\mathbf{y}_0 = B^{-1}\mathbf{b} = \mathbf{x}_B = [y_{i_1,0}, \ldots, y_{i_m,0}]$. We then decide the next step of pivoting using the information that derives from B^{-1} and $\mathbf{y}_0$. We compute the updated cost reduction coefficients $\mathbf{s}_N^T = \mathbf{c}_N^T - \mathbf{c}_B^T B^{-1} N$ to decide on optimality or on the next pivot q and p to use.

To obtain the updated revised tableau for the new basis B^+, we can similarly update the matrix by using the row operations based on the transforms in equation 35. We first need to compute the entering index basis element A_q as $\mathbf{y}_q = B^{-1} A_q = [y_{i_1,q}, \ldots, y_{i_m,q}]$. We also need this representation in the new basis. For this purpose we extend temporally the tableau to $[B^{-1}, \mathbf{y}_0, \mathbf{y}_q]$. We then perform row operations on the entire extended tableau. In particular, we multiply row p by y_{iq}/y_{pq} and subtract it from row i when $i \neq p$. For $i = p$, this row is just multiplied by $1/y_{pq}$. In fact, this transforms row $i = p$ into the needed form for q, see equation 34. As a result, we change the label for row p to q to represent the variable $\mathbf{x}_q$. This gives the updated tableau $[B^{-1+}, \mathbf{y}_0^+]$ for basis $\mathcal{B}^+$.

The presented tableaus provide a few approaches that are used to perform book-keeping when solving linear programs in practice. Many of these steps also can be made more efficient by utilizing sparsity of the matrix A or other specialized data structures within implementations.

Conclusion

These notes are meant to serve as a brief introduction to linear programming and some related algorithms for finding their solutions. Linear programming is an active field with applications in machine learning, AI, statistics, economics, engineering, sciences, and other disciplines. Additional discussions and details on these algorithms also can be found in the references.

For comments or errors concerning these notes, please email: `atzberg@gmail.com`.

References

[1] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 1999.

[2] Edwin KP Chong, Wu-Sheng Lu, and Stanislaw H Zak. *An Introduction to Optimization: With Applications to Machine Learning*. John Wiley & Sons, 2023.

[3] David G Luenberger, Yinyu Ye, et al. *Linear and nonlinear programming*. Vol. 2. Springer, 1984.