THE GEOMETRY OF SIMPLEX METHOD FOR LINEAR PROGRAMS

David F. Gleich

February 10, 2023

The material here is from Chapter 13 in Nocedal and Wright, but some of the geometry comes from Griva, Sofer, and Nash.

Recall the standard form for a linear program:

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \boldsymbol{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0. \end{aligned}$$

The feasible solution of an LP define a polytope:

$$\mathcal{F} = \{\mathbf{x} : \mathbf{x} \geq 0, \boldsymbol{A}\mathbf{x} = \mathbf{b}.\}$$

What is a polytope?

> Wikipedia:Polytype
> In elementary geometry, a polytope is a geometric object with flat sides, which exists in any general number of dimensions. A polygon is a polytope in two dimensions, a polyhedron in three dimensions, and so on in higher dimensions (such as a polychoron in four dimensions). Some theories further generalize the idea to include such things as unbounded polytopes (apeirotopes and tessellations), and abstract polytopes.

Here, we will only consider *bounded* polytopes.
Consider the linear program:

$$\begin{aligned} \underset{x_1,x_2}{\text{minimize}} \quad & -x_1 - 2x_2 \\ \text{subject to} \quad & x_1 \leq 3 \\ & -x_1 + 2x_2 \leq 7 \\ & -2x_1 + x_2 \leq 2 \\ & x_1, x_2 \geq 0. \end{aligned}$$

Show Julia figure for $\mathbf{A}\mathbf{x} \leq \mathbf{b}$, $\mathbf{x} \geq 0$, *see the* `Lecture-14-LP-Polytope.jl` *file.*

Discuss or contemplate What are the slack variables that are introduced when we convert this problem into standard form?

A vertex of the polytope is any point $\mathbf{f} \in \mathcal{F}$ such that

$$\mathbf{f} \neq \alpha\mathbf{y} + (1-\alpha)\mathbf{z} \qquad \text{for any } \mathbf{y}, \mathbf{z} \in \mathcal{F}.$$

Vertices are important because of the following result:

THEOREM 1 (Fundamental Theorem of Linear Programming)
If an LP has a solution, then a solution must occur at a vertex of the feasible polytope.

Discuss Find a partner, and discuss this theorem. I claim that the reason this theorem is true is "obvious" from these pictures. Think about the gradient and objective.

Proof (This is a sketch, as it doesn't handle the case of unbounded directions) There are really two things in the proof. First, we need to show that any feasible point can be expressed as a linear combination of vertices. Second, we need to show that any solution can be expressed as the sum of a single vertex.

Step 1 Any feasible point $\mathbf{f} \in \mathcal{F}$ can be written as:

$$\mathbf{f} = \sum_i \alpha_i \mathbf{v}_i$$

where $\mathbf{v}_i$ is a vertex of the feasible polytope and $\sum_i \alpha_i = 1$. (Assume that $\boldsymbol{A}$ is full-rank, if not, we can replace it by a full-rank matrix.) The formal argument here is somewhat tedious. Let $\mathbf{x}$ be a feasible point. If it's already a vertex, we are done. If it isn't a vertex, then we can find $\mathbf{y}, \mathbf{z}$ with the same non-zero structure as $\mathbf{x}$ such that $\mathbf{x} = \alpha\mathbf{y} + (1-\alpha)\mathbf{z}$. Let $\mathbf{p} = \mathbf{y} - \mathbf{z}$. Note that $\boldsymbol{A}\mathbf{p} = \boldsymbol{A}\mathbf{y} - \boldsymbol{A}\mathbf{z} = \mathbf{b} - \mathbf{b} = 0$. Note also that $\mathbf{p}$ has the same non-zero structure as $\mathbf{x}$. We can then use this $\mathbf{p}$ to find two new points $\mathbf{f}_1 = \mathbf{x} + \gamma_1\mathbf{p}, \gamma_1 > 0$ and $\mathbf{f}_2 = \mathbf{x} + \gamma_2\mathbf{p}, \gamma_2 < 0$ that have one more zero component in them than in $\mathbf{x}$. Hence $\mathbf{x} = \alpha_1\mathbf{y}_1 + \alpha_2\mathbf{y}_2$. Now we inductively repeat this argument on $\mathbf{y}_1$ and $\mathbf{y}_2$, and give them an additional non-zero component or find that they are a vertex. There are only a finite number of non-zeros, so the argument must end with all points at vertices. Thus, we have a convex combination of vertices.

We are basically moving along the direction $\mathbf{p}$ until we hit the constraint $\mathbf{x} = 0$ in a single component.

See the figure in my notes that I drew on the board

Step 2 Again, we assume that the feasible set is bounded (which can be removed, but it makes the argument more complicated, see the books.) Let $\mathbf{x}$ be a solution. Then $\mathbf{x} = \sum_i \alpha_i \mathbf{v}_i$ in terms of vertices by step 1. Now, pick out the vertex j which minimizes the objective:

$$\mathbf{c}^T\mathbf{v}_j = \min_i\{\mathbf{c}^T\mathbf{v}_i\}.$$

Then $\mathbf{v}_j$ must also be a solution because it is feasible and

$$\mathbf{c}^T\mathbf{y} = \sum_i \alpha_i \mathbf{c}^T\mathbf{v}_i \geq \mathbf{c}^T\mathbf{v}_j$$

but $\mathbf{y}$ was a solution, so we must have equality. ∎

This theorem is the heart of the simplex method. What the simplex method does is move between vertices of the polytope while improving the objective function. Because one of those vertices must be the solution, the method will find it eventually.

There are two things we still need, however. The first is a way of dealing with vertices of the polytope. The next theorem provides that. Second, we need a way of determining if a vertex point is optimal. That's covered in the next section.

The final step in the methods is to show the following characterization of vertices of the feasible polytope:

THEOREM 2 *A point* $\mathbf{v}$ *is a vertex of the feasible polytope if and only if it is a* basic feasible point, *that is a point* $\mathbf{x}$ *such that* $\mathbf{x}$ *is feasible and there is a subset* $\mathcal{B}$ *of the indices* $\{1, \ldots, n\}$ *such that* $\mathcal{B}$ *has m indices for the m equations in* $\boldsymbol{A}$, $i \notin \mathcal{B}$ *implies that* $\mathbf{x}_i = 0$ *and* $\boldsymbol{B} = [\boldsymbol{A}_i] i \in \mathcal{B}$ *is non-singular (where* $\boldsymbol{A}_i$ *is the ith column of* $\boldsymbol{A}$*).*

This theorem is characterizing vertices by a subset of columns of the matrix $\boldsymbol{A}$. This will be our computational handle on vertices in the simplex method.

Proof If $\mathbf{x}$ is a basic feasible point that is not a vertex, then we can find feasible $\mathbf{y}$ and $\mathbf{z}$ such that $\mathbf{x} = \alpha\mathbf{y} + (1-\alpha)\mathbf{z}$. Further note that $\mathbf{y}$ and $\mathbf{z}$ must have the same non-zero structure as $\mathbf{x}$ because $\mathbf{y} \geq 0$, $\mathbf{z} \geq 0$, and $\alpha, 1-\alpha > 0$. This means we can write $\mathbf{x}_{\mathcal{B}}$ to be the vector of non-zeros in $\mathbf{x}$, and $\mathbf{y}_{\mathcal{B}}$ and $\mathbf{z}_{\mathcal{B}}$ likewise. Hence $\boldsymbol{B}\mathbf{x}_{\mathcal{B}} = \mathbf{b}$ but also $\boldsymbol{B}\mathbf{y}_{\mathcal{B}} = \boldsymbol{B}\mathbf{z}_{\mathcal{B}} = \mathbf{b}$ and thus they are all the same via the non-singularity of $\boldsymbol{B}$.

The other direction is just a formalization of the argument we gave before. If we have a vertex that is not a basic feasible point, then we can find a direction $\mathbf{p}$ that stays feasible and preserves the non-zero structure of $\mathbf{v}$ because the columns corresponding to the non-zeros in $\mathbf{v}$ are linearly dependent. Thus, we can write $\mathbf{v}$ as a linear combo of these feasible points and we have our contradiction.

See Equations 13.14 and 13.15 in Nocedal and Wright for more detail. ∎

Again, let me reiterate the importance of this result. We can now use the fact that sets of m indicies of the matrix $\boldsymbol{A}$ give rise to vertices of the feasible polytope.

Overview of Simplex Method The simplex method, then, moves from vertex to vertex on the feasible polytope by adding and subtracting indicies from the subset $\mathcal{B}$ to move between vertices. At every stage, it moves to another vertex with a non-increasing objective value.

```
"Presolve" (Remove redundant rows from A)
"Phase 1" Find a feasible vertex / basic point
"Simplex" Find the dual variables / Lagrange multipliers for that point
While KKT conditions are not true
  Move to another point
  Find the dual variables / Lagrange multipliers for new point
```

The first step of finding a feasible vertex or basic point is called a Phase 1 solution. We'll cover that soon.

Columns to basic feasible points and Lagrange multipliers What we now need to address is how to find the dual variables or Lagrange multipliers at a basic feasible point (vertex). *If* we can get these (and we can) then we can determine when to stop the simplex method.

See Section 13.3 for more about this next piece.

Let $\mathbf{x}$ be a basic feasible point. Let $\boldsymbol{B}$ be the subset of columns at any basic feasible point. Thne $\boldsymbol{B}\mathbf{x}_B = \mathbf{b}$ and we can permute $\mathbf{x}$ into $\begin{bmatrix} \mathbf{x}_B^T & \mathbf{x}_N^T \end{bmatrix}^T$, where $\mathbf{x}_B$ are non-zero elements ("the basic elements") of $\mathbf{x}$ and $\mathbf{x}_N = 0$ be the zero elements ("the non-basic elements"). Partition $\mathbf{s}$ (multipliers) and $\mathbf{c}$ (objective) conformally into $\mathbf{s}_B, \mathbf{s}_N$ and $\mathbf{c}_B, \mathbf{c}_N$.

We can write:

$$\boldsymbol{A}\mathbf{x} = \boldsymbol{B}\mathbf{x}_B + \boldsymbol{N}\mathbf{x}_N = \mathbf{b}$$

where $\mathbf{x}_B = \boldsymbol{B}^{-1}\mathbf{b}$ and $\mathbf{x}_B \geq 0$. We set $\mathbf{s}_B = 0$ because all the elements of $\mathbf{x}_B$ handle that portion of the constraint. We now use the remaining KKT conditions to find $\boldsymbol{\lambda}$ and $\mathbf{s}_N$. Note that

$$\boldsymbol{B}^T\boldsymbol{\lambda} = \mathbf{c}_B \text{ and } \boldsymbol{N}^T\boldsymbol{\lambda} + \mathbf{s}_N = \mathbf{c}_N.$$

The first equation defines $\boldsymbol{\lambda}$ because $\boldsymbol{B}^T$ is non-singular. The second equation can then be solved for $\mathbf{s}_N$. Consequently, we can find the dual variables for a solution. *However*, $\mathbf{s}$ need not be non-negative.