Math 3272: Linear Programming[1] Mikhail Lavrov

# Lecture 9: More about corner points

*September 13, 2022* *Kennesaw State University*

## 1 Matrix calculations for the dictionary

Let's review a fact from linear algebra: a matrix-vector product $A\mathbf{x}$ can be viewed as taking a linear combination of the columns of $A$ with weights from the vector $\mathbf{x}$. For example:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = x \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} + y \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} + z \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$$

This is just a different way of using the definition of matrix multiplication. We can check that in the equation above, for example, both sides give $1 \cdot x + 2 \cdot y + 3 \cdot z$ for the first component of the result.

To add a bit of a twist to this idea: once we've split the product $A\mathbf{x}$ up into columns like this, we can recombine some of the columns into smaller matrix-vector products. For example:

$$\begin{aligned} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} &= x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + x_5 \begin{bmatrix} 1 \\ 5 \end{bmatrix} \\ &= \left( x_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + x_5 \begin{bmatrix} 1 \\ 5 \end{bmatrix} \right) + \left( x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right) \\ &= \begin{bmatrix} 1 & 1 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x_2 \\ x_5 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_3 \\ x_4 \end{bmatrix}. \end{aligned}$$

To generalize the notation $x_i$ for the $i^{\text{th}}$ component of a vector, if $\mathcal{I}$ is a sequence of several indices, like $\mathcal{I} = (1, 3, 4)$, then we will write $\mathbf{x}_{\mathcal{I}}$ for the smaller vector with just the components numbered by $\mathcal{I}$ picked out. For example, if $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$, then $\mathbf{x}_{(1,3,4)} = (x_1, x_3, x_4)$.

In the case of a matrix $A$, we will write $A_i$ for the $i^{\text{th}}$ column of $A$. (Picking out a column is more useful to us than picking out a row, which is why we've made this decision for what the notation

[1]This document comes from the Math 3272 course webpage: https://facultyweb.kennesaw.edu/mlavrov/courses/3272-fall-2022.php

means.) Just as with vectors, if $\mathcal{I}$ is a sequence of several indices, then we'll write $A_{\mathcal{I}}$ for the matrix we get by picking out the columns numbered by $\mathcal{I}$ from $A$.

With this notation, the equation we wrote down a bit ago can be written more compactly as

$$A\mathbf{x} = A_{(2,5)}\mathbf{x}_{(2,5)} + A_{(1,3,4)}\mathbf{x}_{(1,3,4)}.$$

This is true for any 5-column matrix $A$ multiplied by any $\mathbf{x} \in \mathbb{R}^5$.

We are interested in doing this for one specific purpose: separating the basic variables from the nonbasic variables. For example, suppose that we have a system of equations

$$\begin{cases} x_1 + \phantom{2}x_2 + \phantom{3}x_3 + \phantom{4}x_4 + \phantom{5}x_5 = 4 \\ x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 10 \end{cases} \iff \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 4 \\ 10 \end{bmatrix}$$

which we write compactly as $A\mathbf{x} = \mathbf{b}$. We decide that we want our basic variables to be $x_2, x_5$ and we want to solve for them in terms of the nonbasic variables $x_1, x_3, x_4$. To do this, set $\mathcal{B} = (2,5)$ and $\mathcal{N} = (1,3,4)$, so that $\mathbf{x}_{\mathcal{B}} = (x_2, x_5)$ is the vector of basic variables and $\mathbf{x}_{\mathcal{N}} = (x_1, x_3, x_4)$ is the vector of nonbasic variables. Then we can turn the matrix equation $A\mathbf{x} = \mathbf{b}$ into the matrix equation

$$A_{\mathcal{B}}\mathbf{x}_{\mathcal{B}} + A_{\mathcal{N}}\mathbf{x}_{\mathcal{N}} = \mathbf{b}.$$

Now we can solve for $\mathbf{x}_{\mathcal{B}}$ in the same way that we'd solve a two-variable equation $2x + 3y = 4$ for $x$. We move $A_{\mathcal{N}}\mathbf{x}_{\mathcal{N}}$ to the other side, and then multiply by the inverse of $A_{\mathcal{B}}$:

$$\begin{aligned} A_{\mathcal{B}}\mathbf{x}_{\mathcal{B}} + A_{\mathcal{N}}\mathbf{x}_{\mathcal{N}} = \mathbf{b} &\implies A_{\mathcal{B}}\mathbf{x}_{\mathcal{B}} = \mathbf{b} - A_{\mathcal{N}}\mathbf{x}_{\mathcal{N}} \\ &\implies \mathbf{x}_{\mathcal{B}} = (A_{\mathcal{B}})^{-1}\mathbf{b} - (A_{\mathcal{B}})^{-1}A_{\mathcal{N}}\mathbf{x}_{\mathcal{N}}. \end{aligned}$$

This is just like the usual dictionary form of our answer, except in matrix form. In particular, we know that setting $\mathbf{x}_{\mathcal{N}} = \mathbf{0}$ (that is, setting all the nonbasic variables to 0) gives us the basic solution. In this case, doing that in the equation above turns it into $\mathbf{x}_{\mathcal{B}} = (A_{\mathcal{B}})^{-1}\mathbf{b}$. In other words:

**Fact 1.** *When $\mathcal{B}$ lists the basic variables and $\mathcal{N}$ lists the nonbasic variables, the basic solution to $A\mathbf{x} = \mathbf{b}$ is to set $\mathbf{x}_{\mathcal{B}} = (A_{\mathcal{B}})^{-1}\mathbf{b}$ and $\mathbf{x}_{\mathcal{N}} = \mathbf{0}$.*

We will make use of this fact again in the next lecture, when we talk about the revised simplex method—a way to do the simplex method with fewer unnecessary calculations. Today, we will use it to prove some claims we've previously explained by intuition.

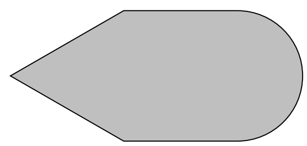
## 2 Definitions of corner points

What is a "corner point" of a subset of $\mathbb{R}^n$? There are three relevant definitions we'll discuss today.

The first definition is the definition of an **extreme point**. An extreme position in politics is the opposite of a moderate position: it's a position that does not compromise about anything. Similarly, an extreme point in a set $S \subseteq \mathbb{R}^n$ is a point that is not in the middle between any other points of $S$.

What does being between two points mean? Algebraically, given two points $\mathbf{p}$ and $\mathbf{q}$, the set of points that lie between them is the line segment of all points $t\mathbf{p} + (1-t)\mathbf{q}$ where $0 \le t \le 1$: the set of weighted averages of $\mathbf{p}$ and $\mathbf{q}$. We make our definition accordingly: a point $\mathbf{x} \in S$ is an extreme point of $S$ if it cannot be written as $t\mathbf{y} + (1-t)\mathbf{z}$ for $\mathbf{y}, \mathbf{z} \in S$ and $0 \le t \le 1$, *except* by taking $\mathbf{y} = \mathbf{x}$ and/or $\mathbf{z} = \mathbf{x}$.

Let's look at an example. Suppose set $S \subseteq \mathbb{R}^2$ is the region below. (Note that this is not the feasible region of any linear program: it has a curved boundary!)



What are the extreme points? Nothing in the middle will do: if we can go a little bit right and a little bit left from a point and stay in $S$, for example, then you're in the middle between "a little bit right" and "a little bit left", so you're not an extreme point. Also, a point that lies on a straight-line boundary is also not an extreme point: it is between two points obtained by going a little bit in one direction along the boundary, and a little bit in the other direction.

The three corner points of the triangle attached on the left of $S$ are all extreme points. Also, *every single point* on the curved boundary on the right of $S$ is an extreme point: from a point on the boundary of a circle, if you pick two opposite directions to go in, one of them will leave the circle.

The definition of an extreme point describes a geometric intution. We can also define a corner point in terms of what we *want* corner points to do. This gives us the definition of a vertex. When $S \subseteq \mathbb{R}^n$, a point $\mathbf{x} \in S$ is a **vertex** of $S$ if there is some nonzero vector $\mathbf{a} \in \mathbb{R}^n$ such that the dot product $\mathbf{a}^\mathsf{T}\mathbf{x} = a_1x_1 + \cdots + a_nx_n$ is *strictly bigger* (no ties allowed!) than $\mathbf{a}^\mathsf{T}\mathbf{y}$ for any $\mathbf{y} \in S$ with $\mathbf{y} \neq \mathbf{x}$.

In other words, the vertices are the points in $S$ that are the unique optimal solutions to a linear maximization problem over $S$.

Looking at the region drawn above, its vertices are almost the same as its extreme points. For the corner points between two straight boundaries, there are many vectors $\mathbf{a}$ we could choose to justify that the corner point is a vertex. For a point on the boundary of the circular arc, pick $\mathbf{a}$ to be the direction away from the center of that circle.

There's only one exception, which is very subtle: the points where the circular arc meets the straight boundary are extreme points, but they're not vertices. That's because if we optimize along the vector $\mathbf{a}$ which points from the center of the circle toward one of these points, then $\mathbf{a}$ points vertically, and all the points along that straight boundary will be tied with $\mathbf{x}$.

The last definition of a corner point only applies to the regions we care about: regions of the form

$S = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b} \text{ and } \mathbf{x} \geq \mathbf{0}\}$. We will assume that the system of equations $A\mathbf{x} = \mathbf{b}$ has no redundant or inconsistent equations: this assumption holds whenever we're using the simplex method, though sometimes we need a two-phase method to check it. Let $m$ be the number of equations (the number of rows in $A$).

In this setting, a **basic feasible solution** $\mathbf{x}$ is any $\mathbf{x} \in S$ such that we can split $(1, 2, \dots, n)$ into $m$ basic variables $\mathcal{B}$ and $n - m$ nonbasic variables $\mathcal{N}$ to have $\mathbf{x}_\mathcal{B} = (A_\mathcal{B})^{-1}\mathbf{b}$ and $\mathbf{x}_\mathcal{N} = \mathbf{0}$. (Note that from $\mathbf{x} \in S$, it follows that $\mathbf{x} \geq \mathbf{0}$.) The basic feasible solutions are exactly the solutions that the simplex method explores. What relationship do they have to the extreme points and the vertices?

# 3 Relationships between the definitions

We've seen by example that for *general* sets $S$, extreme points and vertices don't have to be the same. However, in the case $S = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b} \text{ and } \mathbf{x} \geq \mathbf{0}\}$, they will be the same, and they will always be the same as the basic feasible solutions! We will prove this in three steps.

## 3.1 From basic feasible solutions to vertices

**Theorem 1.** *Any basic feasible solution is a vertex of the feasible region* $S = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b} \text{ and } \mathbf{x} \geq \mathbf{0}\}$.

*Proof.* Suppose that $\mathbf{x}$ is a basic feasible solution: choose $\mathcal{B}$ and $\mathcal{N}$ such that $\mathbf{x}_\mathcal{B} = (A_\mathcal{B})^{-1}\mathbf{b}$ and $\mathbf{x}_\mathcal{N} = \mathbf{0}$. Then, define $\mathbf{a}$ by setting $\mathbf{a}_\mathcal{B} = 0$ and $\mathbf{a}_\mathcal{N} = -1$.

Then the dot product $\mathbf{a}^\mathsf{T}\mathbf{y} = a_1y_1 + a_2y_2 + \dots + a_ny_n$ simplifies to the sum

$$\mathbf{a}^\mathsf{T}\mathbf{y} = \sum_{i \in \mathcal{N}} (-1) \cdot y_i.$$

When $\mathbf{y} = \mathbf{x}$, this is equal to 0: each term of the sum is 0.

In fact, the only way for $\mathbf{a}^\mathsf{T}\mathbf{y} = 0$ to hold is to have $\mathbf{y}_\mathcal{N} = \mathbf{0}$: otherwise, there will be a negative term in the sum! And when $\mathbf{y}_\mathcal{N} = \mathbf{0}$, the equation $A\mathbf{y} = \mathbf{b}$ turns into $A_\mathcal{B}\mathbf{y}_\mathcal{B} = \mathbf{b}$, or $\mathbf{y}_\mathcal{B} = (A_\mathcal{B})^{-1}\mathbf{b}$. Therefore $\mathbf{a}^\mathsf{T}\mathbf{y} = 0$ only if $\mathbf{y} = \mathbf{x}$.

We've checked exactly the condition for $\mathbf{x}$ to be a vertex of the feasible region. □

## 3.2 From vertices to extreme points

We'll actually be able to show that for *any* set $S$, all vertices are also extreme points, even if $S$ is not the feasible region of a linear program (though if $S$ has curved boundaries, the reverse might not hold).

**Theorem 2.** *Any vertex of any set* $S \subseteq \mathbb{R}^n$ *is also an extreme point of* $S$.

*Proof.* Let $\mathbf{x} \in S$ be a vertex of $S$, and let $\mathbf{a}$ be the vector such that $\mathbf{a}^\mathsf{T}\mathbf{x} < \mathbf{a}^\mathsf{T}\mathbf{y}$ for all $\mathbf{y} \in S$ with $\mathbf{y} \neq \mathbf{x}$.

Suppose that $\mathbf{x}$ *is not* an extreme point: then there are $\mathbf{y}, \mathbf{z} \in S$ not equal to $\mathbf{x}$ and some $0 \le t \le 1$ such that $\mathbf{x} = t\mathbf{y} + (1-t)\mathbf{z}$. Multiplying by $\mathbf{a}^\mathsf{T}$ on both sides and distributing, we get

$$\mathbf{a}^\mathsf{T}\mathbf{x} = t(\mathbf{a}^\mathsf{T}\mathbf{y}) + (1-t)(\mathbf{a}^\mathsf{T}\mathbf{z}).$$

Because $\mathbf{y} \neq \mathbf{x}$ and $\mathbf{z} \neq \mathbf{x}$, we know that $\mathbf{a}^\mathsf{T}\mathbf{y} < \mathbf{a}^\mathsf{T}\mathbf{x}$ and $\mathbf{a}^\mathsf{T}\mathbf{z} < \mathbf{a}^\mathsf{T}\mathbf{x}$. Therefore

$$t(\mathbf{a}^\mathsf{T}\mathbf{y}) + (1-t)(\mathbf{a}^\mathsf{T}\mathbf{z}) < t(\mathbf{a}^\mathsf{T}\mathbf{x}) + (1-t)(\mathbf{a}^\mathsf{T}\mathbf{x}).$$

But the right-hand side of this inequality just simplifies to $\mathbf{a}^\mathsf{T}\mathbf{x}$, and we get the ridiculous inequality $\mathbf{a}^\mathsf{T}\mathbf{x} < \mathbf{a}^\mathsf{T}\mathbf{x}$. Therefore assuming $\mathbf{x}$ is not an extreme point has led us to a contradiction, and $\mathbf{x}$ must be an extreme point. □

### 3.3 From extreme points to basic feasible solutions

The last step, going from extreme points to basic feasible solutions, is trickier.

**Theorem 3.** *Any extreme point of the feasible region $S = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \ge \mathbf{0}\}$ is a basic feasible solution.*

*Proof.* Let $\mathbf{x}$ be any extreme point of the feasible region. Split up $(1, 2, \dots, n)$ into $\mathcal{P}$ and $\mathcal{Z}$ such that $\mathbf{x}_\mathcal{Z} = \mathbf{0}_\mathcal{Z}$ and $\mathbf{x}_\mathcal{P} > \mathbf{0}_\mathcal{P}$: the positive entries of $\mathbf{x}$ and the zero entries of $\mathbf{x}$.

What we'd *like* to be the case is that $\mathbf{x}_\mathcal{P}$ is $m$-dimensional (remember, $m$ is the number of rows in $A$) and that $A_\mathcal{P}$ is invertible. Then we can take $\mathcal{B} = \mathcal{P}$ and $\mathcal{N} = \mathcal{Z}$, and $\mathbf{x}$ will be forced to be the basic feasible solution with basic variables $\mathcal{B}$.

This can go wrong in a few ways. First of all, $\mathcal{P}$ might be too small. This is still fine; sometimes we have basic variables equal to 0. If the columns of $A_\mathcal{P}$ are at least linearly independent, then we can pick some more columns to add to $\mathcal{P}$ to make $\mathcal{B}$ in such a way that the columns of $A_\mathcal{B}$ are still linearly independent, making $A_\mathcal{B}$ invertible. Remove those same columns from $\mathcal{Z}$ to get $\mathcal{N}$. Now, once again, $\mathbf{x}$ will be the basic feasible solution with basic variables $\mathcal{B}$.

There are two more things that can go wrong:

- Maybe $\mathcal{P}$ is too small, but the columns of $A_\mathcal{P}$ are already *linearly dependent*. In that case, we can't add any columns to get an invertible matrix, and the procedure above won't work.
- Maybe $\mathcal{P}$ is not too small, but too big: it has more than $m$ entries. In that case, the columns of $A_\mathcal{P}$ are also linearly dependent: they are more than $m$ vectors in $\mathbb{R}^m$.

So if anything goes wrong, then it's because the columns of $A_\mathcal{P}$ are linearly dependent. In this case, we'll try to arrive at a contradiction by showing that $\mathbf{x}$ is not actually an extreme point.

If the columns of $A_\mathcal{P}$ are linearly dependent, then we can take a nontrivial linear combination of them to get $\mathbf{0}$. This linear combination can be written as

$$\sum_{i \in \mathcal{P}} y_i A_i = \mathbf{0}$$

where not all the $y_i$ are $\mathbf{0}$. Let's turn these numbers $y_i$ into an $n$-dimensional vector $\mathbf{y}$, by setting $y_j = 0$ for every $j \in \mathcal{Z}$. Then the linear combination above is just $A_\mathcal{P}\mathbf{y}_\mathcal{P}$.

Now pick a very very very very small value $r > 0$, and consider the points $\mathbf{x} + r\mathbf{y}$ and $\mathbf{x} - r\mathbf{y}$. We'll show that these are two points in $S$ and $\mathbf{x}$ is between them, concluding that $\mathbf{x}$ is not an extreme point.

- We know $A_{\mathcal{P}}\mathbf{y}_{\mathcal{P}} = \mathbf{0}$. We also know $A_{\mathcal{Z}}\mathbf{y}_{\mathcal{Z}} = \mathbf{0}$, because $\mathbf{y}_{\mathcal{Z}} = 0$. Therefore $A\mathbf{y} = \mathbf{0}$.

  It follows that $A(\mathbf{x} \pm r\mathbf{y}) = A\mathbf{x} \pm rA\mathbf{y} = \mathbf{b} \pm r\mathbf{0} = \mathbf{b}$. So the two points $\mathbf{x} + r\mathbf{y}$ and $\mathbf{x} - r\mathbf{y}$ satisfy our system of equations.

- For every position $i$ in $\mathcal{Z}$, both $x_i$ and $y_i = 0$, so $x_i + ry_i = 0$ as well.

  Meanwhile, for every position $i \in \mathcal{P}$, we know $x_i > 0$, so $x_i \pm ry_i \geq 0$ provided that $r$ is sufficiently small.

  Therefore $\mathbf{x} + r\mathbf{y} \geq \mathbf{0}$ and $\mathbf{x} - r\mathbf{y} \geq \mathbf{0}$, provided that $r$ is sufficiently small.

These are the two checks we need to know that $\mathbf{x} + r\mathbf{y} \in S$ and $\mathbf{x} - r\mathbf{y} \in S$. However, $\mathbf{x}$ can be written as $\frac{1}{2}(\mathbf{x} + r\mathbf{y}) + \frac{1}{2}(\mathbf{x} - r\mathbf{y})$, so $\mathbf{x}$ is not an extreme point in this case, contradicting the assumption we started with. $\square$