Columbia MA Math Camp

Linear Algebra

Vinayak Iyer [a]

August 21, 2019

[a] Material adapted from notes by David Thompson and Xingye Wu

Motivation

- Linear systems show up all the time in economics
  - Systems because we deal with more than one quantity at a time (multiple agents, multiple goods/prices, multiple choice variables, etc.)
  - Linearity sometimes comes naturally (e.g. budget constraints), and sometimes we impose it by necessity (fully nonlinear system too hard to analyze) i.e. we "linearize" the equations.
- Linear algebra provides tools for working with these kinds of systems: can we solve them? If so, how? Many different techniques
- My two cents: get comfortable with this section. It's important to be comfortable working with vectors and matrices "as a single object" - it will save you notation and brain space (and computing time if you're into that kind of stuff)

Table of Contents

Vectors

The basic unit in linear algebra is a **vector**. A vector $\mathbf{v}$ is an element of $\mathbb{R}^n$: $\mathbf{v} = (v_1, v_2, ..., v_n)$, where each $v_i \in \mathbb{R}$. In these notes I will denote vectors with boldface, lowercase type.

Two basic operations on vectors are **addition** and **scalar multiplication**:

- Addition: for two vectors of the same length, $\mathbf{v}$ and $\mathbf{w}$

$$\mathbf{v} + \mathbf{w} = (v_1 + w_1, ..., v_n + w_m)$$

- Scalar multiplication: given a vector $\mathbf{v}$ and a scalar $\alpha \in \mathbb{R}$

$$\alpha\mathbf{v} = (\alpha v_1, ..., \alpha v_n)$$

Inner Product

There's another common operation between vectors, known as the **inner product** (or dot product). For two vectors, $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have:

$$\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n} v_i w_i$$

You may also see the inner product written as $\langle \mathbf{v}, \mathbf{w} \rangle$.

While it's not immediately clear that the dot product is a useful notion, the following hints at its importance:

- $\|\mathbf{v}\|^2 = \sum_{i=1}^{n} v_i^2 = \mathbf{v} \cdot \mathbf{v}$, where $\| \cdot \|$ represents the **norm**, or length, of a vector.
- $d(\mathbf{v}, \mathbf{w})^2 = \sum_{i=1}^{n}(v_i - w_i)^2 = (\mathbf{v} - \mathbf{w}) \cdot (\mathbf{v} - \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|^2$

Theorem 1.1
(Cauchy-Schwarz) *For any vectors* $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$*,* $|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\|\|\mathbf{w}\|$*.*

Proof.
We'll show this in $\mathbb{R}^2$. The law of cosines tells us:

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\|\mathbf{v}\|\|\mathbf{w}\| \cos\theta$$

Note $\|\mathbf{v} - \mathbf{w}\|^2 = (\mathbf{v} - \mathbf{w}) \cdot (\mathbf{v} - \mathbf{w}) = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\mathbf{v} \cdot \mathbf{w}$. Simplify:

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\| \cos\theta$$

The result follows since $\cos\theta \leq 1$ $\square$

Cauchy-Schwarz (cont.)

In $\mathbb{R}^n$, we use Cauchy-Schwarz to *define* the angle between two vectors.

$$\cos\theta = \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|}$$

We say two vectors are **orthogonal** to each other if $\mathbf{v}\cdot\mathbf{w} = 0$.

Inner product (cont.)

Let's note a few things about the inner product:

- The inner product is **symmetric** : $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$
- The inner product is **linear** :

$$\begin{aligned} (\alpha \mathbf{v}) \cdot \mathbf{w} &= \alpha(\mathbf{v} \cdot \mathbf{w}) \\ (\mathbf{v} + \mathbf{z}) \cdot \mathbf{w} &= \mathbf{v} \cdot \mathbf{w} + \mathbf{z} \cdot \mathbf{w} \end{aligned}$$

- The inner product is **positive definite**: $\mathbf{v} \cdot \mathbf{v} \geq 0$, with equality iff $\mathbf{v} = \mathbf{0}$

Matrices

A matrix is just a rectangular array of numbers. An $m \times n$ matrix has m rows and n columns:

$$\mathbf{A} = (a_{ij})_{m \times n} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

A vector $\mathbf{v}$ is a $n \times 1$ matrix (a column vector) or a $1 \times n$ matrix (a row vector).

Addition and scalar multiplication are defined just as with vectors:

$$\mathbf{A} + \mathbf{B} = (a_{ij} + b_{ij})_{m \times n}, \quad \alpha\mathbf{A} = (\alpha a_{ij})_{m \times n}$$

Addition and scalar multiplication

Matrix addition and scalar multiplication are well-behaved:

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \mathbf{B} + \mathbf{A} \text{ (commutative)} \\ \mathbf{A} + (\mathbf{B} + \mathbf{C}) &= (\mathbf{A} + \mathbf{B}) + \mathbf{C} \text{ (associative)} \\ \mathbf{A} + \mathbf{0} &= \mathbf{A} \text{ (zero element)} \\ \mathbf{A} + (-1)\mathbf{A} &= \mathbf{0} \text{ (additive inverse)} \\ (\alpha + \beta)(\mathbf{A} + \mathbf{B}) &= \alpha\mathbf{A} + \beta\mathbf{A} + \alpha\mathbf{B} + \beta\mathbf{B} \text{ (distributive)} \end{aligned}$$

Matrix Multiplication

Matrix multiplication is hugely useful, but a little strange at first glance. We do not simply multiply element-by-element.

Let $\mathbf{A}$ be an $m \times n$ matrix and $\mathbf{B}$ a $n \times p$ matrix. Their product, $\mathbf{C} = \mathbf{AB}$ is the $m \times p$ matrix whose ij element is the inner product of the i-th row of $\mathbf{A}$ with the j-th column of $\mathbf{B}$:

$$c_{ij} = \sum_{r=1}^{n} a_{ir} b_{rj}$$

- Matrices must be **conformable**: No. cols of $\mathbf{A}$ = no. rows of $\mathbf{B}$
- Matrix multiplication lets us write inner products: $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w}$

Matrix Multiplication: Perspectives

- A collection of dot products
- Linear combinations of columns/rows
  - Let $\mathbf{A}_i$ denote the i-th column of $\mathbf{A}$
  - If $\mathbf{A}$ is $m \times n$ and x is an $n \times 1$ vector, then:

$$\mathbf{Ax} = \mathbf{A}_1 x_1 + ... + \mathbf{A}_n x_n$$

  - If $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times p$:

$$\mathbf{AB} = \begin{pmatrix} \mathbf{AB}_1 & \mathbf{AB}_2 & \dots & \mathbf{AB}_p \end{pmatrix}$$

- A linear function: $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $f(\mathbf{x}) = \mathbf{Ax}$ where $\mathbf{A}$ is an $m \times n$ matrix.

Matrix Multiplication: Properties

Matrix multiplication is generally well-behaved, with the important exception that it is not commutative.

- $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ (associative)
- $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ (left distributive)
- $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$ (right distributive)
- $\mathbf{AB} \neq \mathbf{BA}$ generally
- $\mathbf{AB} = \mathbf{0}$ does not imply $\mathbf{A}$ or $\mathbf{B}$ is $\mathbf{0}$

Matrix Multiplication: Properties (cont.)

Matrices have an identity element:

$$(\mathbf{I}_n)_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- For any $m \times n$ matrix $\mathbf{A}$, $\mathbf{A}\mathbf{I}_n = \mathbf{I}_m\mathbf{A} = A$.
- For a square matrix $\mathbf{A}$, if $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$, we call $\mathbf{B}$ the **inverse** of $\mathbf{A}$, and write $\mathbf{B} = \mathbf{A}^{-1}$.

Matrix Multiplication: Why?

Why do we have such a strange definition for matrix multiplication? It's useful for representing **linear systems**. Consider:

$$\begin{aligned} 2x_1 + x_2 &= 3 \\ x_1 + 2x_2 &= 3 \end{aligned}$$

We can write this as

$$\underbrace{\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} 3 \\ 3 \end{pmatrix}}_{\mathbf{b}}$$

Let's step back and think for a bit :

- Our goal is to find a tuple $(x_1, x_2) \in \mathbb{R}^2$ that satisfies both equations simultaneously.
- How do we know that a solution exists? Does it always? Can there be many?
- Is there a general method to solve linear systems, or must it be "by inspection" all the time?

Note: if we knew $\mathbf{A}^{-1}$ we could find $\mathbf{x}$ by calculating $\mathbf{A}^{-1}\mathbf{b}$. We'll come back to the question of how to (and when we can) find inverses of a square matrix

Matrices: Two Last Operations

The **transpose** of a $m \times n$ matrix $\mathbf{A}$, written $\mathbf{A}'$ or $\mathbf{A}^T$, is the $n \times m$ matrix with $a'_{ij} = a_{ji}$. A square matrix is **symmetric** if $\mathbf{A} = \mathbf{A}'$.

- $(\mathbf{A}')' = \mathbf{A}$
- $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$
- $(\alpha\mathbf{A})' = \alpha\mathbf{A}'$
- $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

The **trace** of a $n \times n$ matrix $\mathbf{A}$ is the sum of its diagonal elements:

$$tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$$

A little more about the trace

Being able to manipulate traces effectively can make some calculations dramatically simpler. Here are a few useful properties to keep in mind :

- For a scalar α, $tr(\alpha) = \alpha$
- So long as **A** and **B** are conformable, the trace commutes:

$$tr(\mathbf{AB}) = tr(\mathbf{BA})$$

- The above implies that the trace is invariant under **cyclic permutations**:

$$tr(\mathbf{ABC}) = tr(\mathbf{CAB}) = tr(\mathbf{BCA})$$

Table of Contents

Approach to solving linear systems

Consider how you would solve the system

$$\begin{aligned} 2x_1 + x_2 &= 3 \\ x_1 + 2x_2 &= 3 \end{aligned}$$

One solution might be:

- Add the second equation to the first: $3x_1 + 3x_2 = 6$
- Divide by 3: $x_1 + x_2 = 2$
- Subtract this equation from the second: $x_2 = 1$
- Insert $x_2 = 1$ into the first equation: $x_1 = 1$

So the solution is $(x_1, x_2) = (1, 1)$

Elementary row operations

The types of steps we just performed are called the **elementary row operations** for matrices.

- Switching two rows of a matrix
- Multiplying one row by a non-zero scalar
- Adding a multiple of one row to another row

We could replicate the steps above in matrix notation:

$$\begin{aligned}
&\left(\begin{array}{cc|c} 2 & 1 & 3 \\ 1 & 2 & 3 \end{array}\right) \to \left(\begin{array}{cc|c} 3 & 3 & 6 \\ 1 & 2 & 3 \end{array}\right) \\
\to \quad &\left(\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & 2 & 3 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & 1 \end{array}\right) \\
\to \quad &\left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right)
\end{aligned}$$

The corresponding action for columns are called **elementary column operations**

Matrix representation for elementary operations

Switching

- Let $\mathbf{T}_{ij}$ to be the identity matrix with rows i, j switched; $\mathbf{T}_{ij}\mathbf{A}$ is the matrix with rows i, j of $\mathbf{A}$ switched
- T_{ij} is its own inverse

Scalar multiplication

- Let $\mathbf{D}_i(\alpha)$ be the identity matrix with α on the i-th diagonal; $\mathbf{D}_i(\alpha)\mathbf{A}$ is the matrix with the i-th row multiplied by α
- $\mathbf{D}_i\left(\frac{1}{\alpha}\right)$ is the inverse of $\mathbf{D}_i(\alpha)$

Row addition

- Let $\mathbf{L}_{i,j}(m)$ be the identity matrix with m in the (i, j) position; $\mathbf{L}_{i,j}(m)\mathbf{A}$ is the matrix with m times row j added to row i
- $\mathbf{L}_{ij}(-m)$ is the inverse of $\mathbf{L}_{ij}(m)$

To get column operations, multiply on the right instead of on the left

Using row operations to solve linear systems

- Let $\mathbf{R}$ be some row operation.
- Since $\mathbf{R}$ is invertible, a vector $\mathbf{x}$ solves the system $\mathbf{Ax} = \mathbf{b}$ iff it solves $\mathbf{RAx} = \mathbf{Rb}$
- To solve the system, we simply apply row operations on both sides until the solution is "easy" to read off
- What's "easy"? One common setup is **row echelon form**:
  - All non-zero rows are above all zero rows
  - The leading coefficient (first non-zero entry) of each row is strictly to the right of the leading coefficient of the prior row
- Another common setup is **reduced row echelon form**, which adds the following requirements:
  - All leading coefficients are 1
  - The leading coefficients are the only nonzero entries in their column

- Finding a matrix inverse is the same as finding vectors $\mathbf{x}_i$ such that $\mathbf{A}\mathbf{x}_i = \mathbf{e}_i$, the i-th canonical basis vector.
- So just solve all n equations at once using the augmented matrix $\left(\begin{array}{c|c} \mathbf{A} & \mathbf{I} \end{array} \right)$.

Example:

$$\left(\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array} \right) \to \left(\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 2 & 4 & 0 & 2 \end{array} \right) \to \left(\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 0 & 3 & -1 & 2 \end{array} \right)$$

$$\to \left(\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 0 & 1 & -\frac{1}{3} & \frac{2}{3} \end{array} \right) \to \left(\begin{array}{cc|cc} 2 & 0 & \frac{4}{3} & -\frac{2}{3} \\ 0 & 1 & -\frac{1}{3} & \frac{2}{3} \end{array} \right)$$

$$\to \left(\begin{array}{cc|cc} 1 & 0 & \frac{2}{3} & -\frac{1}{3} \\ 0 & 1 & -\frac{1}{3} & \frac{2}{3} \end{array} \right)$$

A note about column operations

- In general, column operations do not preserve the solutions of systems of equations. If $\mathbf{Ax} = \mathbf{b}$, can we say anything about $\mathbf{ACx}$?
- Interestingly, we *can* use column operations to find inverses. This is due to the fact that left inverses are equal to right inverses, so if $\mathbf{R}_n ... \mathbf{R}_1 \mathbf{A} = \mathbf{I}$, then $\mathbf{A}\mathbf{R}_n ... \mathbf{R}_1 = \mathbf{I}$
- **Warning:** do not mix and match column and row operations to find an inverse.

Table of Contents

Motivation

- Our ultimate goal is to understand the behavior of linear systems of equations
- To facilitate this, it's useful to develop a few concepts from linear spaces
- These phrases appear often enough that it's worth knowing what they are, even if you don't use them every day

Let $W \subseteq \mathbb{R}^n$. We say that W is a **vector subspace** or **linear subspace** of $\mathbb{R}^n$ if:

- W contains $\mathbf{0}$
- W is closed under addition: $u, v \in W \Rightarrow u + v \in W$
- W is closed under scalar multiplication: $u \in W, \alpha \in \mathbb{R} \Rightarrow \alpha u \in W$

Linear Independence

Let $\mathbf{x}_1, ..., \mathbf{x}_k$ be k vectors in $\mathbb{R}^n$.

- A **linear combination** of $\mathbf{x}_1, ..., \mathbf{x}_k$ is a vector $\lambda_1 \mathbf{x}_1 + ... + \lambda_k \mathbf{x}_k$.
- The vectors $\mathbf{x}_1, ..., \mathbf{x}_k$ are **linearly dependent** if there exist numbers $c_1, ..., c_k$, not all equal to 0, such that
$$c_1 \mathbf{x}_1 + ... + c_k \mathbf{x}_k = \mathbf{0}$$
- If this equation only holds when $c_1 = ... = c_k = 0$ we say the vectors are **linearly independent**.

Linear Independence (cont.)

Proposition 3.1
Let $\mathbf{x}_1, ..., \mathbf{x}_k$ be linearly independent vectors and suppose there are 2 different representations of the same vector $\mathbf{y}$ i.e.

$$\lambda_1\mathbf{x}_1 + \ldots + \lambda_k\mathbf{x}_k = \mathbf{y} = \mu_1\mathbf{x}_1 + \ldots + \mu_k\mathbf{x}_k$$

Then the representation is unique i.e. $\lambda_i = \mu_i$ for all $i = 1, \ldots, k$.

Proof : Move all terms to one side and so $\lambda_i - \mu_i = 0\ \forall i$ **Note :** This is a nice result

because any vector that is a linear combination of the $\mathbf{x}$'s can be written so in a unique way. Will use this property soon.

Corollary: If the columns of $\mathbf{A}$ are linearly independent, the system $\mathbf{Ax} = \mathbf{b}$ has at most one solution.

Why? Note that you can think of the vector $\mathbf{b}$ as a linear combination of the columns of $\mathbf{A}$

Span

Let $\mathbf{x}_1, ..., \mathbf{x}_k$ be k vectors of $\mathbb{R}^n$. The **span** of $\mathbf{x}_1, .., \mathbf{x}_k$ is the collection of all linear combinations of $\mathbf{x}_1, ..., \mathbf{x}_k$:

$$\text{Span}(\mathbf{x}_1, ..., \mathbf{x}_k) = \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{x}_i | \{\lambda_i\}_{i=1}^{k} \in \mathbb{R}^k \right\}$$

Claim: the span of a collection of vectors is a vector subspace. (Why?)

Definition 3.1
Suppose W is a subspace of $\mathbb{R}^n$, and that $\mathbf{x_1}, ..., \mathbf{x}_k$ has the following two properties:

- $Span(\mathbf{x}_1, \ldots, \mathbf{x}_k) = W$
- $\mathbf{x}_1, ..., \mathbf{x}_k$ *are linearly independent*

Then $\mathbf{x}_1, ..., \mathbf{x}_k$ is called a **basis** *for W.*

Notes :

- By our earlier result, every element of W can be uniquely written as a linear combination of elements of $\mathbf{x}_1, ..., \mathbf{x}_k$
- If $\mathbf{w} = \lambda_1 \mathbf{x}_1 + \ldots + \lambda_k \mathbf{x}_k$, we call $\lambda_1, \ldots, \lambda_k$ the **coordinates** of w

In $\mathbb{R}^n$, we typically use the **canonical basis vectors**: $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, \ldots, 0)$ and so on

Proposition 3.2
Let $\mathbf{x}_1, ..., \mathbf{x}_j$ be a basis for W. Then any collection of more than j vectors of W is linearly dependent.

Proof :

- Let $\mathbf{w}_1, ..., \mathbf{w}_k$ be a collection of vectors of W with $k > j$.
- By definition of a basis, $\mathbf{x}_1, ..., \mathbf{x}_j, \mathbf{w}_1$ are linearly dependent:

$$\lambda_1 \mathbf{x}_1 + \ldots + \lambda_j \mathbf{x}_j = w_1$$

  with λ_i not all 0.
- WLOG, assume $\lambda_1 \neq 0$
- **Claim:** $\mathbf{w}_1, \mathbf{x}_2, ..., \mathbf{x}_j$ is a basis for W
- Repeat this process j times, and we find $\mathbf{w}_1, \ldots, \mathbf{w}_j$ is a basis for W
- Therefore $\mathbf{w}_1, ..., \mathbf{w}_j, \mathbf{w}_{j+1}, ..., \mathbf{w}_k$ is linearly dependent

Dimension (cont.)

The result above has two nice corollaries. Let W be a subspace of $\mathbb{R}^n$:

- All bases of W have the same number of elements. This is called the **dimension** of W. For example in $\mathbb{R}^2$, the basis has 2 elements – For example, $e_1 = (1, 0)$ and $e_2 = (0, 1)$
- If W has dimension j, any collection of j linearly independent vectors of W forms a basis for W (**proof:** if it didn't, we could find a set of $j + 1$ linearly independent vectors)
- Note $\{\mathbf{0}\}$ is subspace of $\mathbb{R}^n$. We say it has dimension 0.

Rank

Let $\mathbf{x}_1, \ldots, \mathbf{x}_k$ be a family of vectors of $\mathbb{R}^n$

- The **rank** of $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is the dimension of $\mathrm{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$
- Equivalently, the rank is the largest group of linearly independent vectors of $\mathbf{x}_1, \ldots, \mathbf{x}_k$.

Given an $m \times n$ matrix $\mathbf{A}$, its rank, $r(\mathbf{A})$ is the rank of the columns of $\mathbf{A}$, which are elements of $\mathbb{R}^m$.

- The span of the columns of $\mathbf{A}$ is also called the image of $\mathbf{A}$ or the **column space** of $\mathbf{A}$. In other words it is the set of vectors that can be expressed as linear combinations of the columns of $\mathbf{A}$
- Note $r(\mathbf{A}) \leq \min(m, n)$

Kernel

Definition 3.2
Let $\mathbf{A}$ *be an* $m \times n$ *matrix. Define the* **kernel** *of* $\mathbf{A}$ *as*

$$ker(A) \equiv \{\mathbf{x} \in \mathbb{R}^n | \mathbf{A}\mathbf{x} = \mathbf{0}\}$$

Claim: The kernel of $\mathbf{A}$ is a subpsace of $\mathbb{R}^n$ (problem set)

Theorem 3.1
Let $\mathbf{A}$ be an $m \times n$ matrix with rank k. Then the kernel of $\mathbf{A}$ is a subspace of $\mathbb{R}^n$ with dimension $n - k$.

- Essentially implies that Rank of a Matrix + Nullity = Number of Columns of the Matrix

Consider a $m \times n$ matrix $\mathbf{A}$ as a collection of n columns vectors. We need one key result:

Proposition 3.3
The rank of $\mathbf{A}$ *is unaffected by elementary row and column operations.*

Proof.
It should be clear that column operations do not affect the dimension of column space of $\mathbf{A}$. For row operations, note $\mathbf{RAx} = \mathbf{0} \Leftrightarrow \mathbf{Ax} = \mathbf{0}$, so row operations do not affect the kernel of A, so by the Rank-Nullity Theorem, the rank is preserved. $\square$

An implication of this theorem is that the rank of a matrix is equal to the rank of its transpose.

Calculating the rank (cont.)

There are two nice implications of this result:

- The rank of a matrix is the number of nonzero rows when in reduced row echelon form
- The rank of a matrix is equal to the rank of its transpose.
- **Idea:** row operations on $\mathbf{A}$ are column operations on $\mathbf{A}^T$ and vice-versa. Put $\mathbf{A}^T$ in reduced column echelon form

Results for square systems

Let $\mathbf{A}$ be an $n \times n$ matrix. The following are equivalent:

(a) $\mathbf{A}$ is invertible

(b) $\mathbf{A}$ is rank n (i.e. the columns of $\mathbf{A}$ are linearly independent)

(c) The kernel of $\mathbf{A}$ is trivial: $ker(\mathbf{A}) = \{0\}$

We'll show (1) $\Leftrightarrow$ (2). The fact that (2) $\Leftrightarrow$ (3) is immediate.

- $\implies$: Assume $\mathbf{A}$ is invertible. Then $\mathbf{Ax} = \mathbf{0}$ only has the trivial solution, so the columns of $\mathbf{A}$ are linearly independent, so $\mathbf{A}$ is rank n.

- $\impliedby$: Now assume $\mathbf{A}$ is rank n. The columns of $\mathbf{A}$ form a basis for $\mathbb{R}^n$, so there exist $\mathbf{b}_i$ such that $\mathbf{Ab}_i = \mathbf{e}_i$. Let $\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_n \end{pmatrix}$. Then

$$\mathbf{AB} = \mathbf{I}$$

  Finally, we need to show $\mathbf{BA} = I$. You'll do this on your problem set.

Non-square, homogeneous systems

Let $\mathbf{A}$ be an $m \times n$ matrix and consider the equation $\mathbf{Ax} = \mathbf{0}$.

- From Rank-Nullity Theorem, $dim(ker(\mathbf{A})) = n - k$

Now let's suppose $\mathbf{A}$ is full rank:

- If $m < n$, $rank(\mathbf{A}) = m$, so $dim(ker(\mathbf{A})) = n - m$. **Idea:** more unknowns than equations, so we get many solutions. $n - m$ free variables
- If $m \geq n$, $rank(\mathbf{A}) = n$, so $dim(ker(\mathbf{A})) = 0$.

Nonhomogeneous systems: $m > n$

Consider the system $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A}$ is $m \times n$ with $m > n$ and rank $r \leq n$

- **Overconstrained system**: more equations than unknowns
- Span of the columns of $\mathbf{A}$ is r-dimensional subspace of $\mathbb{R}^m$ - much "smaller" than $\mathbb{R}^m$. For most vectors $\mathbf{b}$, a solution will not exist
- $\mathbf{A} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then $\mathbf{Ax} = \begin{pmatrix} x \\ x \end{pmatrix}$
- For $\mathbf{b} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, there is no x that can satisfy both equations
- This is similar to regression contexts: many observations and only a few parameters to match the data with. Focus on solutions that minimize $\|\mathbf{b} - \mathbf{Ax}\|$.

Nonhomogeneous systems: $m < n$

Consider the system $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A}$ is $m \times n$ with $m < n$

- Underconstrained system: more unknowns than equations
- If $\mathbf{A}$ is full rank, columns of $\mathbf{A}$ are a basis for $\mathbb{R}^m$, so a solution x^* exists
- However, for any $\mathbf{z} \in ker(A)$, $\mathbf{A}(\mathbf{x}^* + \mathbf{z}) = \mathbf{b}$, so $x^* + z$ is also a solution
- Set of solutions is essentially $n - m$ dimensional

This situation can also arise in regression settings, when the number of regressors exceeds the number of data points. Trick is to restrict the set of $\mathbf{x}$'s you consider.

Table of Contents

Motivation

Calculating matrix inverses is an important part of solving systems of equations. How do we know when an inverse exists? The **determinant** helps us answer this question.

Consider the 2×2 case. Let

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

- This matrix is not invertible iff its columns are linearly dependent
- This happens iff $a = \lambda b$ and $c = \lambda d$ for some $\lambda \neq 0$
- This happens iff $\lambda ad = \lambda bc$, or if $ad - bc = 0$

To check whether a 2×2 matrix is invertible, we simply calculate $ad - bc$ and check whether it is 0. Therefore we define:

$$det(\mathbf{A}) \equiv |\mathbf{A}| = ad - bc$$

We won't prove this result, but there is a nice recursive formula for calculating determinants

Definition 4.1
Let $\mathbf{A}$ *be an* $n \times n$ *matrix, and let* $\mathbf{A}_{ij}$ *denote the matrix formed by deleting the* i*-th row and* j*-th column of* $\mathbf{A}$*. The* **determinant** *of* $\mathbf{A}$, $det(\mathbf{A})$ *or* $|\mathbf{A}|$ *is the real number defined recursively as:*

- *If* $n = 1$ *(that is, if* $\mathbf{A} = a_{11}$*),* $|\mathbf{A}| = a_{11}$
- *If* $n \geq 2$, $|\mathbf{A}| = (-1)^{1+1} a_{11} |\mathbf{A}_{11}| + ... + (-1)^{1+n} a_{1n} |\mathbf{A}_{1n}|$

For a 3×3 matrix:

$$\begin{aligned} \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= a(ei - hf) - b(di - fg) + c(dh - eg) \end{aligned}$$

Determinant: Properties

- If two rows (columns) of $\mathbf{A}$ are interchanged, $|A|$ changes sign
- If a row (column) of $\mathbf{A}$ is multiplied by c, $|A|$ is multiplied by c
- If a multiple of one row (column) is added to another row (column), $|A|$ is unchanged
- If two rows (columns) of $\mathbf{A}$ are proportional, $|A| = 0$
- $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$
- $|\mathbf{A}'| = |\mathbf{A}|$
- $\mathbf{A}^{-1}$ exists iff $|A| \neq 0$
- There's actually an explicit formula for $\mathbf{A}^{-1}$ (FMEA Section 1.1); the only one worth memorizing is the 2×2 case

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\underbrace{ad - bc}_{|\mathbf{A}|}} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Proposition 4.1
Consider the system of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ *where* $\mathbf{A}$ *is a* $n \times n$ *matrix. If* $\mathbf{A}$ *is invertible, then*

$$x_j = \frac{|\mathbf{A}_j|}{|\mathbf{A}|}$$

where $\mathbf{A}_j$ *is the matrix with* $\mathbf{b}$ *in place of the* j*-th column of* $\mathbf{A}$*.*

Proof.
Define

$$X_1 = \begin{pmatrix} x_1 & 0 & \dots & 0 \\ x_2 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ x_n & 0 & \dots & 1 \end{pmatrix}$$

We see $x_1 = det(X_1)$. Note also that $\mathbf{A}X_1 = \mathbf{A}_1$. Taking determinants on both sides gives $det(\mathbf{A})det(X_1) = det(\mathbf{A}_1)$. □