Columbia MA Math Camp

Linear Algebra

Vinayak Iyer [a]

August 20, 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: $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$ 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 $\mathbf{A}$ and $\mathbf{B}$ 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

Using row operations to find inverses

- 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:

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

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.