Columbia MA Math Camp

Linear Algebra

Vinayak Iyer [a]

July 22, 2020

[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, $\mathrm{v}, \mathrm{w} \in \mathbb{R}^n$, we have:

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

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

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

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

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

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

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

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

$$v \cdot w = \|v\|\|w\| \cos\theta$$

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.

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

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

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

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

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

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

Matrices

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

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

A vector 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:

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

Addition and scalar multiplication

Matrix addition and scalar multiplication are well-behaved:

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

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

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

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

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

Matrix Multiplication: Perspectives

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

$$Ax = A_1 x_1 + ... + A_n x_n$$

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

$$AB = \begin{pmatrix} AB_1 & AB_2 & \dots & AB_p \end{pmatrix}$$

- A linear function: $f(x) : \mathbb{R}^n \to \mathbb{R}^m$ with $f(x) = Ax$ where 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.

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

Matrix Multiplication: Properties (cont.)

Matrices have an identity element:

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

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

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

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

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 A^{-1} we could find x by calculating $\mathsf{A}^{-1}\mathsf{b}$. 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 A, written A' or A^T, is the $n \times m$ matrix with $a'_{ij} = a_{ji}$. A square matrix is **symmetric** if $\mathrm{A} = \mathrm{A}'$.

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

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

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

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:

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

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

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

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

Scalar multiplication

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

Row addition

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

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

Using row operations to solve linear systems

- Let R be some row operation.
- Since R is invertible, a vector x solves the system $Ax = b$ iff it solves $RAx = Rb$
- 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

Using row operations to find inverses

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

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 $Ax = b$, can we say anything about 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 $R_n ... R_1 A = I$, then $AR_n ... R_1 = I$
- **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

Subspaces

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

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

Proposition 3.1
Let $x_1, ..., x_k$ *be linearly independent vectors and suppose there are 2 different representations of the same vector* y *i.e.*

$$\lambda_1 x_1 + \ldots + \lambda_k x_k = y = \mu_1 x_1 + \ldots + \mu_k x_k$$

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 x's can be written so in a unique way. Will use this property soon.

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

Why? Note that you can think of the vector b as a linear combination of the columns of A

Span

Let $x_1, ..., x_k$ be k vectors of $\mathbb{R}^n$. The **span** of $x_1, .., x_k$ is the collection of all linear combinations of $x_1, ..., x_k$:

$$\text{Span}(x_1, ..., x_k) = \left\{ \sum_{i=1}^{k} \lambda_i x_i | \{\lambda_i\}_{i=1}^{k} \in \mathbb{R}^k \right\}$$

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

Basis

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

- *$Span(x_1, \ldots, x_k) = W$*
- *$x_1, ..., x_k$ are linearly independent*

Then $x_1, ..., x_k$ 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 $x_1, ..., x_k$
- If $w = \lambda_1 x_1 + \ldots + \lambda_k x_k$, 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 $x_1, ..., x_j$ *be a basis for* W*. Then any collection of more than* j *vectors of* W *is linearly dependent.*

Proof :

- Let $w_1, ..., w_k$ be a collection of vectors of W with $k > j$.
- By definition of a basis, $x_1, ..., x_j, w_1$ are linearly dependent:

$$\lambda_1 x_1 + \ldots + \lambda_j x_j = w_1$$

  with λ_i not all 0.
- WLOG, assume $\lambda_1 \neq 0$
- **Claim:** $w_1, x_2, ..., x_j$ is a basis for W
- Repeat this process j times, and we find $w_1, \ldots, w_j$ is a basis for W
- Therefore $w_1, ..., w_j, w_{j+1}, ..., w_k$ 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 $\{0\}$ is subspace of $\mathbb{R}^n$. We say it has dimension 0.

Rank

Let $x_1, \ldots, x_k$ be a family of vectors of $\mathbb{R}^n$

- The **rank** of $x_1, \ldots, x_k$ is the dimension of $\mathrm{Span}(x_1, \ldots, x_k)$
- Equivalently, the rank is the largest group of linearly independent vectors of $x_1, \ldots, x_k$.

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

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

Kernel

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

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

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

Theorem 3.1
Let A *be an* $m \times n$ *matrix with rank* k*. Then the kernel of* 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 A as a collection of n columns vectors. We need one key result:

Proposition 3.3
The rank of 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 A. For row operations, note $\mathrm{RAx} = 0 \Leftrightarrow \mathrm{Ax} = 0$, 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 A are column operations on A^T and vice-versa. Put A^T in reduced column echelon form

Results for square systems

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

(a) A is invertible
(b) A is rank n (i.e. the columns of A are linearly independent)
(c) The kernel of A is trivial: $ker(\mathrm{A}) = \{0\}$

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

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

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

$$\mathrm{AB} = \mathrm{I}$$

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

Non-square, homogeneous systems

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

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

Now let's suppose A is full rank:

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

Nonhomogeneous systems: $m > n$

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

- **Overconstrained system**: more equations than unknowns
- Span of the columns of A is r-dimensional subspace of $\mathbb{R}^m$ - much "smaller" than $\mathbb{R}^m$. For most vectors b, a solution will not exist
- $\mathrm{A} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then $\mathrm{Ax} = \begin{pmatrix} x \\ x \end{pmatrix}$
- For $\mathrm{b} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, 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 $\|\mathrm{b} - \mathrm{Ax}\|$.

Nonhomogeneous systems: $m < n$

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

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

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

- 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:

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

The Determinant

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

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

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

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 A are interchanged, $|A|$ changes sign
- If a row (column) of 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 A are proportional, $|A| = 0$
- $|\mathrm{AB}| = |\mathrm{A}||\mathrm{B}|$
- $|\mathrm{A}'| = |\mathrm{A}|$
- A^{-1} exists iff $|A| \neq 0$
- There's actually an explicit formula for A^{-1} (FMEA Section 1.1); the only one worth memorizing is the 2×2 case

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

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

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

where A_j *is the matrix with* b *in place of the* j*-th column of* 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 $\mathrm{A}X_1 = \mathrm{A}_1$. Taking determinants on both sides gives $det(\mathrm{A})det(X_1) = det(\mathrm{A}_1)$. □

Table of Contents

Motivation

Consider the following simplified system of equations from the New Keynesian model:

$$\begin{aligned} \pi_t &= \beta \pi_{t+1} + \kappa y_t \\ y_t &= y_{t+1} - \sigma(i - \pi_{t+1}) \end{aligned}$$

These types of systems are common in economic analysis: several interrelated variables reflecting the actions from distinct groups. Notice we can write this system as:

$$\begin{pmatrix} 1 & -\kappa \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \pi_t \\ y_t \end{pmatrix} = \begin{pmatrix} \beta & 0 \\ \sigma & 1 \end{pmatrix} \begin{pmatrix} \pi_{t+1} \\ y_{t+1} \end{pmatrix} + \begin{pmatrix} 0 \\ \sigma i \end{pmatrix}$$

Motivation (cont.)

Define $\mathsf{x}_t = \begin{pmatrix} \pi_t \\ y_t \end{pmatrix}$. This system is of the form:

$$\begin{aligned} \mathsf{x}_{t+1} &= \mathsf{A}\mathsf{x}_t + \mathsf{b} \\ &= \mathsf{A}(\mathsf{A}\mathsf{x}_{t-1} + \mathsf{b}) + \mathsf{b} = \mathsf{A}^2 \mathsf{x}_{t-1} + (\mathsf{I} + \mathsf{A})\mathsf{b} \\ &= \ldots \\ &= \mathsf{A}^{t+1}\mathsf{x}_0 + (\mathsf{I} + \mathsf{A} + \ldots + \mathsf{A}^t)\mathsf{b} \end{aligned}$$

Takeaway:

- The long-term behavior of this system depends on the power of a matrix.
- Given a matrix, can we easily tell how A^t will evolve? Turns out we can by studying the **eigenvalues** of A

Definition 5.1
A nonzero vector x *of a matrix* A *is a vector such that* $\mathsf{Ax} = \lambda\mathsf{x}$ *for some* $\lambda \in \mathbb{R}$ *is called an* **eigenvector** *of* A. *The value* λ *is called the* **eigenvalue**.

Example:

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

In this example, $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is the eigenvector with associated eigenvalue 3.

Finding Eigenvalues

- $Ax = \lambda x$ iff $(A - \lambda I)x = 0$.
- This implies $A - \lambda I$ has a nontrivial solution, which happens iff $det(A - \lambda I) = 0$.

Approach: calculate $det(A - \lambda I)$. This is known as the **characteristic polynomial** of A. The roots of this polynomial are the eigenvalues of A.

Example: $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

$$det(A - \lambda I) = \begin{vmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{vmatrix} = (2 - \lambda)^2 - 1$$

The roots of this equation are: $\lambda = 1$ and $\lambda = 3$.

Finding Eigenvectors

Once we know the eigenvalues of A, plug them into the equation $(A - \lambda I)x = 0$ and solve.

Let's find the eigenvector associated with $\lambda = 1$ in the previous example:

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Both equations implies $x_1 + x_2 = 0$, so for example $x = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ is an eigenvector of A with eigenvalue 1.

Proposition 5.1
If A *is an* $n \times n$ *matrix with eigenvalues* $\lambda_1, \ldots, \lambda_n$*, then*

- $|\mathrm{A}| = \lambda_1 \lambda_2 \ldots \lambda_n$
- $tr(\mathrm{A}) = \lambda_1 + \lambda_2 + \ldots + \lambda_n$

Proof.
(First result) Consider the characteristic polynomial $p(\lambda) = det(\mathrm{A} - \lambda \mathrm{I})$. According to the Fundamental Theorem of Algebra, we can factor

$$p(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \ldots (\lambda - \lambda_n)$$

where λ_i is an eigenvalue of A. Letting $\lambda = 0$, we see:

$$|\mathrm{A}| = p(0) = \lambda_1 \ldots \lambda_n = \lambda_1 \ldots \lambda_n$$

(Second result) Similar; look at coefficient on λ^{n-1} (use induction) $\square$

Properties of eigenvalues (cont.)

Proposition 5.2

Let $\lambda_1, ..., \lambda_m$ *be distinct eigenvalues of* A, *with associated eigenvectors* $\mathsf{v}_1, ..., \mathsf{v}_m$. *Then* $\mathsf{v}_1, ..., \mathsf{v}_m$ *are linearly independent*

Proof.
By way of contradiction, suppose $\mathsf{v}_1, ..., \mathsf{v}_m$ are linearly dependent.

- Let k be the smallest integer such that $\mathsf{v}_1, ..., \mathsf{v}_k$ are linearly dependent, and assume $\alpha_1 \mathsf{v}_1 + ... + \alpha_k \mathsf{v}_k = 0$.
- Applying A on both sides gives $\alpha_1 \lambda_1 \mathsf{v}_1 + ... + \alpha_k \lambda_k \mathsf{v}_k = 0$.
- Multiplying the first equation by λ_k and subtracting gives

$$\alpha_1(\lambda_1 - \lambda_k)\mathsf{x}_1 + ... + \alpha_{k-1}(\lambda_{k-1} - \lambda_k)\mathsf{x}_{k-1} = 0$$

- Since $\mathsf{v}_1, ..., \mathsf{v}_{k-1}$ are linearly independent and the eigenvalues are distinct, we must have $\alpha_1 = ... = \alpha_{k-1} = 0$.
- This implies $\alpha_k = 0$, so $\mathsf{v}_1, ..., \mathsf{v}_k$ are linearly independent ; which is a contradiction.

$\square$

Diagonalization

Remember our goal is to understand how A^t behaves.

For diagonal matrices, this is easy:

$$\mathrm{D}^t = \begin{pmatrix} d_1^t & 0 & \dots & 0 \\ 0 & d_2^t & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d_n^t \end{pmatrix}$$

Suppose we could write $\mathrm{A} = \mathrm{PDP}^{-1}$, where D is a diagonal matrix. Then

$$\mathrm{A}^2 = \mathrm{PD}\underbrace{\mathrm{P}^{-1}\mathrm{P}}_{\mathrm{I}}\mathrm{DP}^{-1} = \mathrm{PD}^2\mathrm{P}^{-1}$$

Likewise, $\mathrm{A}^t = \mathrm{PD}^t\mathrm{P}^{-1}$

Given a matrix A, when can we write $\mathsf{A} = \mathsf{PDP}^{-1}$?

- Can do this iff $\mathsf{AP} = \mathsf{PD}$ for some invertible matrix P, or:

$$\begin{pmatrix} \mathsf{A}\mathsf{p}_1 & \mathsf{A}\mathsf{p}_2 & \dots & \mathsf{A}\mathsf{p}_n \end{pmatrix} = \begin{pmatrix} d_1\mathsf{p}_1 & d_2\mathsf{p}_2 & \dots & d_n\mathsf{p}_n \end{pmatrix}$$

- That is, if $\mathsf{A}\mathsf{p}_i = d_i\mathsf{p}_i$ for each i. Equivalently, if p_i are the eigenvectors of A, and d_i the associated eigenvalues

Proposition 5.3
An $n \times n$ matrix A is diagonalizable if and only if it has a set of n linearly independent eigenvectors. In that case, $\mathsf{A} = \mathsf{PDP}^{-1}$, where P is a matrix of eigenvectors and D a diagonal matrix of corresponding eigenvalues.

In a sense, most matrices are diagonalizable:

Proposition 5.4
If a matrix has n distinct eigenvalues, it is diagonalizable.

Proof.
By Proposition 5.2, the eigenvalues are linearly independent. The result follows from the previous slide. □

- Distinct eigenvalues are sufficient but not necessary
- For matrices that aren't diagonalizable, there's a more general procedure: Jordan canonical form. We won't pursue this here.

Symmetric Matrices

A matrix P is called **orthogonal** if $\mathsf{PP}' = \mathsf{P}'\mathsf{P} = \mathsf{I}$

Proposition 5.5
If A *is symmetric:*

- *All eigenvalues of* A *are real*
- *Eigenvectors that correspond to distinct eigenvalues are orthogonal*
- A *is orthogonally diagonalizable: there exists an orthogonal matrix* P *such that* $\mathsf{A} = \mathsf{PDP}'$

Proof.
(Claim 2) Suppose $Ax_1 = \lambda_1 x_1$ and $Ax_2 = \lambda_2 x_2$ with $\lambda_1 \neq \lambda_2$.

- ${x_2}' A x_1 = \lambda_1 x_2' x_1$
- ${x_2}' A x_1 = x_1' A' x_2 = x_1' A x_2 = \lambda_2 x_1' x_2 = \lambda_2 x_2' x_1$

Therefore $\lambda_1 x_2' x_1 = \lambda_2 x_2' x_1$. Since $\lambda_1 \neq \lambda_2$, $x_2' x_1 = 0$

□

Table of Contents

Motivation

Quadratic forms are polynomials where every term is of degree two. For example:

$$x_1^2, \quad x_1^2 + 2x_1x_2, \quad x_1^2 + x_1x_3 + x_3^2$$

In economics, quadratic forms typically arise from **Taylor Series Expansion** (we will cover this next week)

- Tell us about the curvature of a function at a particular point
- Helpful in characterizing whether functions are convex/concave
- Helpful in determining whether critical points are max/min/other (2nd order tests)

Quadratic Forms: Definition

Definition 6.1
A **quadratic form** *is a function* $Q : \mathbb{R}^n \to \mathbb{R}$*:*

$$Q(x_1, ..., x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

Notes : Every quadratic form can be represented by a matrix. Let $\mathrm{A} = (a_{ij})$;

$$Q(x_1, ..., x_n) = \mathrm{x}' A \mathrm{x}$$

Moreover, every quadratic form can be represented by a **symmetric** matrix

Certain quadratic forms have attractive properties that will be useful when we discuss convexity for multivariable functions:

Definition 6.2
Let Q be a quadratic form

- *A quadratic form is* **positive definite** *if $Q(x) > 0$ for all $x \neq 0$*
- *A quadratic form is* **positive semidefinite** *if $Q(x) \geq 0$ for all x*
- *A quadratic form is* **negative definite** *if $Q(x) < 0$ for all $x \neq 0$*
- *A quadratic form is* **negative semidefinite** *if $Q(x) \leq 0$ for all x*
- *A quadratic form is* **indefinite** *if it is neither positive semidefinite nor negative semidefinite*

Definiteness in $\mathbb{R}^2$

Let $Q(x_1, x_2)$ be a quadratic form represented by $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$.

$$Q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2$$

- Q is positive definite iff $a > 0$ and $ac - b^2 > 0$
- Q is negative definite iff $a < 0$ and $ac - b^2 > 0$

Proof.
($\Rightarrow$) Suppose Q is positive definite. Then $Q(1, 0) = a > 0$. Similarly, $Q(-b, a) = -ab^2 + ca^2 > 0$, so $ac - b^2 > 0$.

($\Leftarrow$). Suppose $a > 0$ and $ac - b^2 > 0$. Then for $x \neq 0$:

$$\begin{aligned} Q(x_1, x_2) &= a\left(x_1^2 + \frac{2b}{a}x_1x_2 + \frac{c}{a}x_2^2\right) \\ &= a\left(\left(x_1 + \frac{b}{a}x_2\right)^2 + \frac{ac - b^2}{a^2}x_2^2\right) > 0 \end{aligned}$$

$\square$

Semidefiniteness in $\mathbb{R}^2$

- Q is positive semidefinite iff $a \geq 0, c \geq 0$ and $ac - b^2 \geq 0$
- Q is negative semidefinite iff $a \leq 0, c \leq 0$ and $ac - b^2 \geq 0$

Proof is similar, but note you need to check c as well!

Definiteness in $\mathbb{R}^n$

There is a generalization of the results in $\mathbb{R}^2$, but first we need a little vocabulary:

- A **principal minor** of order k a $n \times n$ matrix A is the determinant of a matrix consisting of k rows of A and the same k columns of A
- A **leading principal minor** of order k a $n \times n$ matrix A is the determinant of the matrix consisting of the first k rows and columns of A

Definiteness in $\mathbb{R}^n$

Let D_k be the **leading principal minor** of order k and Δ_k an arbitrary principal minor of order k.

- Q is **positive definite** $\Leftrightarrow D_k > 0$ for $k = 1, ..., n$
- Q is **negative definite** $\Leftrightarrow (-1)^k D_k > 0$ for $k = 1, ..., n$
- Q is **positive semidefinite** $\Leftrightarrow \Delta_k \geq 0$ for $k = 1, ..., n$ and all Δ_k
- Q is **negative semidefinite** $\Leftrightarrow (-1)^k \Delta_k \geq 0$ for $k = 1, ..., n$ and all Δ_k

A few notes:

- Generalizes the result in $\mathbb{R}^2$
- Checking semi-definiteness is more demanding - can't just check the principal minors

Example

Consider the quadratic form represented by

$$A = \begin{pmatrix} -2 & 6 & 0 \\ 6 & -18 & 0 \\ 0 & 0 & -4 \end{pmatrix}$$

Is this negative definite? The leading principal minors are:

- Order 1: $(-1)^1(-2) = 2 > 0$
- Order 2: $(-1)^2(36 - 36) = 0 \not> 0$
- Order 3: $(-1)^3(-2 * 72 + 6 * 24) = 0 \not> 0$

A is not negative definite, but could still be negative semidefinite: we need to check the remaining principal minors:

- Order 1: $(-1)^1(-18) \geq 0, (-1)^1(-4) \geq 0$
- Order 2: $(-1)^2(8 - 0) \geq 0, (-1)^2(72 - 0) \geq 0$

All the principal minors are the correct sign, so A is negative semidefinite

An eigenvalue characterization of definiteness

Let Q be represented by the symmetric matrix A with eigenvalues λ_i

- Q is **positive definite** $\Leftrightarrow \lambda_i > 0$ for all i
- Q is **negative definite** $\Leftrightarrow \lambda_i < 0$ for all i
- Q is **positive semidefinite** $\Leftrightarrow \lambda_i \geq 0$ for all i
- Q is **negative semidefinite** $\Leftrightarrow \lambda_i \leq 0$ for all i
- Q is **indefinite** $\Leftrightarrow$ A has positive and negative eigenvalues

Proof.
Since A is symmetric, it is orthogonally diagonalizable, so $\mathrm{x}'\mathrm{A}\mathrm{x} = \mathrm{x}'\mathrm{P}\mathrm{D}\mathrm{P}'\mathrm{x}$. Define $\mathrm{y} = \mathrm{P}'\mathrm{x}$. Then

$$\mathrm{x}'\mathrm{A}\mathrm{x} = \mathrm{y}'\mathrm{D}\mathrm{y} = \sum_{i=1}^{n} \lambda_i y_i^2$$

If each $\lambda_i > 0$, then $\mathrm{x}'\mathrm{A}\mathrm{x} > 0$, so Q is positive definite. If some $\lambda_i \leq 0$, we can find an $\mathrm{x} \neq 0$ such that $\mathrm{x}'\mathrm{A}\mathrm{x} \leq 0$, so Q is not positive definite. $\square$