3: Algebra

Math camp 2019

George Mason University

August 22, 2019

- Me: Arthur Dolgopolov,
- Don't hesitate to email me questions! Seriously.
  adolgopo@gmu.edu
- Lecture slides:
  https://arthurdolgopolov.net/teaching/mathcamp
- Wolfram Mathematica (free for students):
  https://cos.gmu.edu/mathematica/

- 0. Review
- 1. Linear algebra, matrices, eigenvalues and eigenvectors.
- + A bit of time for questions / More practice
- Extra References for linear algebra and calculus:
  - Sargent, Stachurski lecture in quantecon, https://lectures.quantecon.org/jl/linear_algebra.html
  - Essence of linear algebra and calculus on YouTube by 3Blue1Brown https://www.youtube.com/playlist?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab
  - Mathematical methods for economic theory Martin J. Osborne (Chapter 1.2) https://mjo.osborne.economics.utoronto.ca/index.php/tutorial/index/1/toc

Review...

1. Donuts example,
2. http://tutorial.math.lamar.edu/Classes/CalcI/ComputingIndefiniteIntegrals.aspx
3. Derivative of absolute value.

$$
\begin{aligned}
&= \lim_{\Delta x \to 0} \frac{|x+\Delta x| - |x|}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{(|x+\Delta x| - |x|)(|x+\Delta x| + |x|)}{\Delta x(|x+\Delta x| + |x|)} \\
&= \lim_{\Delta x \to 0} \frac{|x+\Delta x|^2 - |x|^2}{\Delta x(|x+\Delta x| + |x|)} \\
&= \lim_{\Delta x \to 0} \frac{(x+\Delta x)^2 - x^2}{\Delta x(|x+\Delta x| + |x|)} \\
&= \lim_{\Delta x \to 0} \frac{x^2 + 2x\Delta x + (\Delta x)^2 - x^2}{\Delta x(|x+\Delta x| + |x|)} \\
&= \lim_{\Delta x \to 0} \frac{2x + \Delta x}{|x+\Delta x| + |x|}
\end{aligned}
$$

Linear algebra

Algebra	Calculus	Geometry	Foundations	Number theory
Elementary	Analysis	Discrete	Philosophy of mathematics	
Linear	Differential equations	Algebraic	Mathematical logic	
Multilinear	Numerical analysis	Analytic	Set theory	
Abstract	Optimization	Differential	Category theory	
Combinatorics	Functional analysis	Finite		
Group theory	Dynamical systems	Topology		
Representation theory		Trigonometry		
Lie theory				

- Algebra - defining and manipulating things (symbols).
  - Linear algebra - defining and manipulating nice simple things.
- Fundamental object - **system of linear equations**:

$$\begin{cases} 3x + 5y + z = 2 \\ 7x - 2y + 4z = 0 \\ -6x + 3y + 2z = 5 \end{cases}$$

What is "nice and simple"? Formally linearity means:

Definition

A map (function, transformation) is linear if it preserves the operations of **addition** and **scalar multiplication**

1. Additivity

$$f(u+v) = f(u) + f(v)$$

2. Scalar multiplication

$$f(cu) = cf(u)$$

$$<=>$$

Intuitively, grid lines remain parallel and evenly spaced.
As long as we have this, we can use the results of linear algebra (whatever our actual problem and space may be).

More generally,

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + ...a_{1k}x_k \\ y_2 = a_{21}x_1 + a_{22}x_2 + ...a_{2k}x_k \\ ... \\ y_n = a_{n1}x_1 + a_{n2}x_2 + ...a_{nk}x_k \end{cases} \qquad \begin{bmatrix} a_{11} & a_{12} & ... & a_{1k} \\ a_{21} & a_{22} & ... & a_{2k} \\ ... & & & \\ a_{n1} & a_{n2} & ... & a_{nk} \end{bmatrix} \times X = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}$$

We would like to know:

- Does a solution exist?
- Unique or many?
- How difficult is it to find the solution?
- Can we well approximate a solution instead?

Definition

Vector of length n is a 1-dimensional array of numbers.

$x = (x_1 ... x_n)$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$x = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$

Euclidian vectors are usually represented by $\vec{x}$, but this notation is not used in econ. The set of all n vectors is $\mathbb{R}^n$.

Definition

Matrix is a 2-dimensional array of numbers.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \; ((a_{11}, a_{21}), (a_{12}, a_{22}))$$

Vectors have 3 competing intuitions: length+direction (Physics), array (CS), an element of vector space (Math).
Unlike sets, the order is important, hence $()$, not $\{\}$.

Definition

Vector of length n is a 1-dimensional array of numbers.

This is a natural definition for any n. Another way of looking at vectors is Euclidian (e.g. high school physics):

Definition

Euclidian Vector is a geometric object that has magnitude (or length) and direction.

Addition, scalar multiplication:

$$x + \gamma y = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} + \begin{bmatrix} \gamma y_1 \\ \gamma y_2 \\ \dots \\ \gamma y_n \end{bmatrix} = \begin{bmatrix} x_1 + \gamma y_1 \\ x_2 + \gamma y_2 \\ \dots \\ x_n + \gamma y_n \end{bmatrix}$$

Transpose:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

Dot (scalar) product:
$\langle x, y \rangle = x \cdot y = \sum_i^n x_i y_i = xy^T = ||x|| \cdot ||y|| \cdot cos\theta$
x and y are orthogonal if it is zero. $x \perp y$

The norm:
$||x|| = \sqrt{(x^T x)} = (\sum_{i=1}^n x_i^2)^{1/2}$

Matrix multiplication

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \dots \\ b_{n1} & b_{n2} & \dots & b_{np} \end{bmatrix}$$

$c_{ij} = a_{i1} b_{1j} + \dots + a_{im} b_{mj} = \sum_{k=1}^{m} a_{ik} b_{kj}$

A is $n \times m$, B is $m \times p$, C is $n \times p$.

Intuition: this is composition ($\circ$).

A third definition of vector/matrix:

Definition

Matrix is a (linear) function.

1. Systems of linear equations

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + ...a_{1k}x_k \\ y_2 = a_{21}x_1 + a_{22}x_2 + ...a_{2k}x_k \\ ... \\ y_n = a_{n1}x_1 + a_{n2}x_2 + ...a_{nk}x_k \end{cases}$$

is just $Ax = y$.

2. Matrix is a (linear) function $f : V \to W$, $V = \mathbb{R}^k$, $W = \mathbb{R}^n$. If the two spaces are the same, it is called a **linear operator**.

Any $n \times k$ matrix maps a vector x in $\mathbb{R}^k$ into vector $y = Ax \in \mathbb{R}^n$
In other words, if any matrix is a function, linear algebra solves problems of the form $y = f(x)$ when f is linear.

What is "nice and simple"? Formally linearity means:

Definition

A map (function, transformation) is linear if it preserves the operations of **addition** and **scalar multiplication**

1. Additivity

$$f(u+v) = f(u) + f(v)$$

2. Scalar multiplication

$$f(cu) = cf(u)$$

$$<=>$$

Intuitively, grid lines remain parallel and evenly spaced.
As long as we have this, we can use the results of linear algebra (whatever our actual problem and space may be).

A set of vectors $A = \{a_1, ...a_k\}$, $a_i \in \mathbb{R}^n$
$y \in \mathbb{R}^n$ is a linear combination of A if $y = \sum \beta_i a_i$ for some scalars $\beta_1 ... \beta_k$
Set of all linear combinations is called a **span**.

What if I want to span the whole $\mathbb{R}^3$? What set of vectors do I need?

Canonical basis (vectors are linearly independent and every vector in the vector space is a linear combination of this set):

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Definition

A **vector (linear) space** is a collection of vectors. In general a **space** is just a set with some structure.

A set of vectors is **linearly dependent** if some strict subset has the same span.

Definition

set of vectors A is linearly independent if:

1. It is not linearly dependent (see above)
2. No vector in A can be formed as linear combination of others
3. if $\sum \beta_i a_i = 0$ for some $\beta_1 ... \beta_k$ then $\beta_i = 0$ for all i.

Each element in the span has a unique representation (no other combination of β will produce the same vectors).

Proof.

Suppose $y = \gamma_1 a_1 + ... \gamma_k a_k$
Then $(\beta_1 - \gamma_1)a_1 + ... + (\beta_k - \gamma_k)a_k = 0$, and then $\gamma = \beta$, because these vectors are linearly independent. □

Linear transformation

Keeps origin in place, keeps grid lines parallel and evenly spaced.

In a two-dimensional space we can describe a linear transformation by a 2x2 matrix. Why?

Let's return to our original problem. A system of linear equations can have no solution, a unique solution or infinitely many solutions.

Ax is a linear combination of A. So we want to check if y is in the span of A.

Theorem

For any square linear system these are equivalent:

1. Columns of A are linearly independent.

2. A is full rank.

3. **Determinant** *of A is not zero (the matrix is nonsingular).*

4. For any $y \in \mathbb{R}^n$, $y = Ax$ has a unique solution.

A most general rule:

Theorem (Rouché–Capelli theorem)

A system of linear equations with n variables has a solution iff $rank(A) = rank(A|b)$. If there are solutions, they form a subspace of R of dimension $n - rank(A)$.

In particular: if $n = rank(A)$, the solution is unique,

otherwise there are infinitely many solutions.

In other terms: dimension of column space (rank), the span of columns, tells us if the solution exists and dimension of null space or a kernel describes how many solutions we have.

Examples:

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

$$A = \begin{pmatrix} 1 & 1 & 2 & |3 \\ 1 & 1 & 1 & |1 \\ 2 & 2 & 2 & |5 \end{pmatrix}$$

The rank of a matrix A is the number of rows and columns in the largest square matrix obtained by deleting rows and columns of A that has a determinant different from 0.

The rank of a matrix A is the maximum number of linearly independent vectors in a matrix.

Interpretation follows the same logic as for functions in general:

Below A is $m \times n$ matrix and $f(x) = Ax$,
f is injective (or "one-to-one") if and only if A has rank n (in this case, we say that A has full column rank).
f is surjective (or "onto") if and only if A has rank m (in this case, we say that A has full row rank).

Suppose that we have $n = m$ at first (a square matrix). A is an operator that maps from some space to the same space.

Injective: $f(x) = f(y) \implies x = y$

Surjective: $\forall y \in Y, \exists x \in X, f(x) = y$

f is injective (or "one-to-one") if and only if A has rank n (in this case, we say that A has full column rank).
$\sim$ The output has as many dimensions as the number of variables.

f is surjective (or "onto") if and only if A has rank m (in this case, we say that A has full row rank).
$\sim$ The output has as many dimensions as the number of equations.

The columns of the matrix are the images of the unit vectors. If they are not linear independent, the null space can't be trivial.
In other words take a homogeneous system of equations $Ax = 0$. It always has a zero solution, but it can have more solutions. These solutions form a null-space of the matrix. If there are more than one, then by definition it is not injective.

https://www.youtube.com/watch?v=uQhTuRlWMxw

A common approach to finding the rank of a matrix is to reduce it to a simpler form, generally row echelon form, by elementary row operations.
This approach can also be used to solve the system altogether.
Gaussian elimination: swapping, scalar multiplication and addition.

System of equations	Row operations	Augmented matrix
$\begin{aligned} 2x + y - z &= 8 \\ -3x - y + 2z &= -11 \\ -2x + y + 2z &= -3 \end{aligned}$		$\left[\begin{array}{ccc\|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right]$
$\begin{aligned} 2x + y - z &= 8 \\ \tfrac{1}{2}y + \tfrac{1}{2}z &= 1 \\ 2y + z &= 5 \end{aligned}$	$L_2 + \tfrac{3}{2}L_1 \to L_2$ $L_3 + L_1 \to L_3$	$\left[\begin{array}{ccc\|c} 2 & 1 & -1 & 8 \\ 0 & \tfrac{1}{2} & \tfrac{1}{2} & 1 \\ 0 & 2 & 1 & 5 \end{array}\right]$
$\begin{aligned} 2x + y - z &= 8 \\ \tfrac{1}{2}y + \tfrac{1}{2}z &= 1 \\ -z &= 1 \end{aligned}$	$L_3 + -4L_2 \to L_3$	$\left[\begin{array}{ccc\|c} 2 & 1 & -1 & 8 \\ 0 & \tfrac{1}{2} & \tfrac{1}{2} & 1 \\ 0 & 0 & -1 & 1 \end{array}\right]$
The matrix is now in echelon form (also called triangular form)		
$\begin{aligned} 2x + y &= 7 \\ \tfrac{1}{2}y &= \tfrac{3}{2} \\ -z &= 1 \end{aligned}$	$L_2 + \tfrac{1}{2}L_3 \to L_2$ $L_1 - L_3 \to L_1$	$\left[\begin{array}{ccc\|c} 2 & 1 & 0 & 7 \\ 0 & \tfrac{1}{2} & 0 & \tfrac{3}{2} \\ 0 & 0 & -1 & 1 \end{array}\right]$
$\begin{aligned} 2x + y &= 7 \\ y &= 3 \\ z &= -1 \end{aligned}$	$2L_2 \to L_2$ $-L_3 \to L_3$	$\left[\begin{array}{ccc\|c} 2 & 1 & 0 & 7 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right]$
$\begin{aligned} x &= 2 \\ y &= 3 \\ z &= -1 \end{aligned}$	$L_1 - L_2 \to L_1$ $\tfrac{1}{2}L_1 \to L_1$	$\left[\begin{array}{ccc\|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right]$

What is this solution?
$x = A^{-1}y$
A^{-1} is the **inverse** matrix: $AA^{-1} = A^{-1}A = I$
If no such matrix exists, A is **singular**.

I is the identity matrix $AI = IA = A$. $I = \begin{bmatrix} 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & & & & \\ 0 & 0 & 0 & ... & 1 \end{bmatrix}$

Theorem

A square $(n \times n)$ matrix has at most one inverse.

Proof

Let A be a $n \times n$ matrix, and suppose that B and C are both inverses of A. Then by the definition of an inverse, $BA = AB = I$ and $CA = AC = I$, where I is the $n \times n$ identity matrix. Thus $C = CI = C(AB) = (CA)B = IB = B$, so that C and B are the same.

Nonsingular matrix

A matrix that has an inverse is called nonsingular.

$$(A^T)^T = A$$

$$(A + B)^T = A^T + B^T$$

$$(AB)^T = B^T A^T$$

$$(rA)^T = rA^T$$

$$A(rB) = (rA)B = r(AB)$$

$$(r + s)A = rA + sA$$

$$A(B + C) = AB + AC$$

$$(A + B)C = AC + BC$$

$$A(BC) = (AB)C$$

Simplify $(AB)(B^{-1}A^{-1})$

Let $B = A^{-1}$, then $b_{ij} = (-1)^{i+j}|A_{ji}|/|A|$

Theorem (Cramer's rule:)

Let A be an $n \times n$ matrix, let b be a $n \times 1$ column vector, and consider the system of equations $Ax = b$, where x is an $n \times 1$ column vector.

If A is nonsingular then the (unique) value of x that satisfies the system is given by $x_i = |A^*(b, i)|/|A|$ for $i = 1, ..., n$, where $A^*(b, i)$ is the matrix obtained from A by replacing the ith column with b.

Practice:
https://mjo.osborne.economics.utoronto.ca/index.php/tutorial/index/1/eqs/t
Solve exercises (they have solutions).

More practice:
National income model (Y, C, T are endogenous)

$$Y = C + I + G$$

$$C = a + b(Y - T)$$

$$T = d + tY$$

Solve for equilibrium.

We will use this again for systems of differential equations (i.e. Macro I).

$$\begin{pmatrix} 1 & -1 & 0 \\ -b & 1 & b \\ -t & 0 & 1 \end{pmatrix}^{-1}$$

Augment with identity matrix:

$$\left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ -b & 1 & b & 0 & 1 & 0 \\ -t & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

$L2 = L2 + b \times (L1)$

$$\left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1-b & b & b & 1 & 0 \\ -t & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

$L3 = L3 + t \times (L1)$

$$\left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1-b & b & b & 1 & 0 \\ 0 & -t & 1 & t & 0 & 1 \end{array}\right)$$

$$L3 = L3 + \frac{t}{1-b} \times (L2)$$

$$\begin{pmatrix} 1 & -1 & 0 & | & 1 & 0 & 0 \\ 0 & 1-b & b & | & b & 1 & 0 \\ 0 & -t & \frac{br}{1-b}+1 & | & \frac{t}{1-b} & \frac{t}{1-b} & 1 \end{pmatrix}$$

$$L3 = L3/(1 + \frac{br}{1-b})$$

$$\begin{pmatrix} 1 & -1 & 0 & | & 1 & 0 & 0 \\ 0 & 1-b & b & | & b & 1 & 0 \\ 0 & 0 & 1 & | & \frac{t}{b(t-1)+1} & \frac{t}{b(t-1)+1} & \frac{1}{\frac{bt}{1-b}+1} \end{pmatrix}$$

$$L2 = L2 - b \times L3$$

$$\begin{pmatrix} 1 & -1 & 0 & | & 1 & 0 & 0 \\ 0 & 1-b & 0 & | & b - \frac{bt}{b(t-1)+1} & 1 - \frac{bt}{b(t-1)+1} & \frac{b(b-1)}{b(t-1)+1} \\ 0 & 0 & 1 & | & \frac{t}{b(t-1)+1} & \frac{t}{b(t-1)+1} & \frac{1}{\frac{bt}{1-b}+1} \end{pmatrix}$$

$$L2 = L2/(1-b)$$

$$\left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & \frac{b-bt}{b(t-1)+1} & \frac{1}{b(t-1)+1} & -\frac{b}{b(t-1)+1} \\ 0 & 0 & 1 & \frac{t}{b(t-1)+1} & \frac{t}{b(t-1)+1} & \frac{1}{\frac{bt}{1-b}+1} \end{array}\right)$$

$$L1 = L2 + L1$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{b(t-1)+1} & \frac{1}{b(t-1)+1} & -\frac{b}{b(t-1)+1} \\ 0 & 1 & 0 & \frac{b-bt}{b(t-1)+1} & \frac{1}{b(t-1)+1} & -\frac{b}{b(t-1)+1} \\ 0 & 0 & 1 & \frac{t}{b(t-1)+1} & \frac{t}{b(t-1)+1} & \frac{1}{\frac{bt}{1-b}+1} \end{array}\right)$$

$$\frac{1}{1-b+bt} \times \begin{pmatrix} 1 & 1 & -b \\ b-bt & 1 & -b \\ t & t & 1-b \end{pmatrix}$$

A is $n \times n$ matrix, λ is a scalar, $v \in \mathbb{R}^n$.
Remember that A is a function.
When it is applied to v, v is simply scaled by λ. Can you write this formally?

We will need eigenvalues and eigenvectors for 1) convexity and 2) differential equations.

A is $n \times n$ matrix, λ is a scalar, $v \in \mathbb{R}^n$.
Remember that A is a function.
When it is applied to v, v is simply scaled by λ. Can you write this formally?

Definition

$$A\mathbf{v} = \lambda\mathbf{v}$$

We will need eigenvalues and eigenvectors for 1) convexity and 2) differential equations.

Determinant is the factor by which linear transformation scales areas (it can take negative values).
Additional nice fact is that the determinant is the product of eigenvalues.

$$det(A) = \sum_{\sigma \in S_n} (sgn(\sigma) \prod_{i=1}^{n} a_{i,\sigma_i})$$

S_n is the set of permutations of $\{1, 2 ... n\}$, e.g. for $\{1, 2, 3\} : (1, 2, 3), (2, 3, 1), (3, 2, 1), (1, 3, 2), (2, 1, 3), (3, 1, 2)$
$sgn(\sigma)$ is the signature ($+1$ if even, -1 if odd).

Recursive formula:

$$det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} |A_{1j}|$$

A_{1j} is a matrix obtained by deleting row 1 and column j.

$$A\mathbf{v} = \lambda\mathbf{v}$$

$$(A - \lambda I)\mathbf{v} = 0$$

$\mathbf{v}$ non-zero, therefore $A - \lambda I$ must not be invertible.

$$det(A - \lambda I) = 0$$

- **characteristic polynomial**.

Example. Find eigenvalues and eigenvectors of:

$$A = \begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix}$$

Projections!

Let's find a projection of b onto a.
Another way to think of it is that we are minimizing the error $e = b - p$ of approximating b with p when we have to stick to subspace formed by a.
$p = xa$ for some scalar x. And it is perpendicular to $e = b - xa$.

$$a^T(b - xa) = 0$$

$$xa^T a = a^T b$$

$$x = \frac{a^T b}{a^T a}$$

$$p = ax = a\frac{a^T b}{a^T a}$$

Or as a linear transformation $P = \frac{aa^T}{a^T a}$, $p = Pb$.

Projections in higher dimensions

Let's find a projection of b onto the closest point p in a plane given by A.

Same approach:

$p = Ax$ for some x. And $A^T(b - Ax) = 0$.

$$A^T Ax = A^T b$$

$a^T a$ was a scalar. $A^T A$ is a matrix, so

$$x = (A^T A)^{-1} A^T b$$

$$p = Ax = A(A^T A)^{-1} A^T b$$

$$P = A(A^T A)^{-1} A^T$$

Why do economists care?

a) Computers are really good at multiplying matrices.
b) Small Application: Revealed preference.
$a \succeq b$, $b \succeq c \implies a \succeq c$.
Write down all your preferences in a big matrix (leave zeros where you are not sure).

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Take n-th power of the matrix. Why does this work? This is called a transitive closure.
What do numbers in the matrix stand for?

c) Econometrics example

Let $n > k$ that is the system is overdetermined - more equations than variables. There is a chance that this system has no solution. Can we find the best approximate solution?

The unique minimizer of $||y - Xb||$ over $b \in \mathbb{R}^K$ is

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Proof:

$$X\hat{\beta} = X(X^T X)^{-1} X^T y = Py$$

Py is orthogonal projection onto $span(X)$, so

$$||y - Py|| \leq ||y - z||$$

for any $z \in span(X)$.

Since $Xb \in span(X)$

$||y - X\hat{\beta}|| \leq ||y - Xb||$ for any $b \in \mathbb{R}^K$. $\square$

Try proving this by taking derivative (Problem set). More about this in https://lectures.quantecon.org/py/orth_proj.html

.. and your Econometrics class.

1. Additivity

$$f(u+v) = f(u) + f(v)$$

$$\frac{d}{dx}(x^3 + x^2) = \frac{d}{dx}(x^3) + \frac{d}{dx}(x^2)$$

2. Scalar multiplication

$$f(cu) = cf(u)$$

$$\frac{d}{dx}(4x^3) = 4\frac{d}{dx}(x^3)$$

This means we can find a matrix that defines this linear transformation?

Yes! Here it is (at least the beginning of it):

$$A = \begin{bmatrix} 0 & 1 & 0 & 0... \\ 0 & 0 & 2 & 0... \\ 0 & 0 & 0 & 3... \\ 0 & 0 & 0 & 0... \end{bmatrix}$$

Basis:

$$b_0(x) = 1$$

$$b_1(x) = x$$

$$b_2(x) = x^2$$

...

This is nice, but is it useful in any way?

Notice that A^n is **"Nilpotent"**, that is $A^n = O$, a zero matrix. Why does this make sense?

Differentiating reduces the power of a polynomial by one.

Cross product

$$a \times b = ||a|| \cdot ||b|| sin(\theta) n$$

Returns a vector perpendicular to both a and b.

Length equals the signed area of the parallelogram between vectors.

Cross product can be viewed as a linear transformation to the number line.