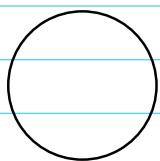
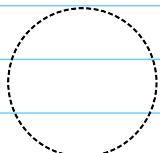


Notation for sets and mathematical logic

set
notation

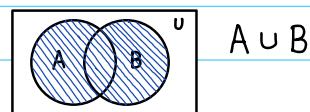


a **closed set** includes sample points on the barrier



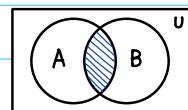
an **open set** does not

Union:



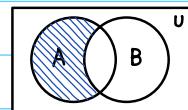
$$A \cup B$$

intersection:



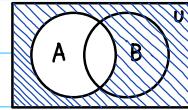
$$A \cap B$$

"difference":



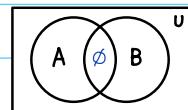
$$A - B \text{ or } A \setminus B \quad (A \text{ "remove" } B)$$

"complement":



$$A^c \text{ or } A' \text{ or } \bar{A}$$

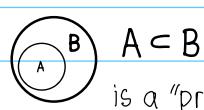
disjoint



$$A \cap B = \emptyset$$

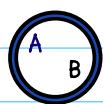
where \emptyset denotes an "empty" set

Subsets:



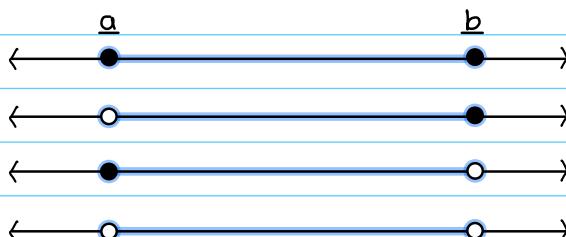
$$A \subset B$$

is a "proper" subset



$$A \subseteq B$$

Interval
notation



$$[a, b] \text{ or } [a, b] \text{ or } a \leq x \leq b$$

$$(a, b] \text{ or }]a, b] \text{ or } a < x \leq b$$

$$[a, b) \text{ or } [a, b[\text{ or } a \leq x < b$$

$$(a, b) \text{ or }]a, b[\text{ or } a < x < b$$

some
Logic/Logical
notation

|A| cardinality of A; no. elements in A; size of A

$a \in A$ elements a are in set A

$a \notin A$ elements a are **not** in set A

\forall for every

\exists there exists

$\Rightarrow \Leftarrow$ or c: contradiction

$x | y$ x "divides" y i.e. $\exists m \in \mathbb{N}$ s.t. $y = mx$

{...} set

{... | ...} or \exists ... s.t. "such that" ...

! only one; unique

\equiv or IET or \leftrightarrow equivalence; "is equivalent to"

\Leftrightarrow or iff biconditional equivalence; "is equivalent to if and only if"

$\therefore P$ and $\because P$ therefore P and because of P

$A \times B$ cartesian product:

given $A = \{1, 2, 3\}$ and $B = \{x, y, z\}$

$$A \times B = \{(1, x), (1, y), (1, z), (2, x), (2, y), (2, z), (3, x), (3, y), (3, z)\}$$

- `itertools.product` in Python \neq "zip" function
- creates a set of all possible "ordered pairs"
- **ordered pairs \neq cartesian coordinates**
- "order matters" and $|A \times B| = mn$, $m = |A|$ and $n = |B|$

| | |
|-------------------|---|
| QUALIFIERS | \forall for every |
| | \exists there exists |
| | \neg or \sim \bar{A} negation; "not" |
| | $a \wedge b$ a and b |
| | $a \vee b$ a or b (inclusive i.e., a or b or both) |
| | $a \Delta b$ $(a \vee b) \wedge \neg(a \wedge b)$ a or b (exclusive i.e., a or b but <u>not</u> both) |

Theorem

For statements P , Q and R ,

(1) Commutative Laws

- (a) $P \vee Q \equiv Q \vee P.$
- (b) $P \wedge Q \equiv Q \wedge P.$

(2) Associative Laws

- (a) $P \vee (Q \vee R) \equiv (P \vee Q) \vee R.$
- (b) $P \wedge (Q \wedge R) \equiv (P \wedge Q) \wedge R.$

(3) Distributive Laws

- (a) $P \vee (Q \wedge R) \equiv (P \vee Q) \wedge (P \vee R).$
- (b) $P \wedge (Q \vee R) \equiv (P \wedge Q) \vee (P \wedge R).$

(4) De Morgan's Laws

- (a) $\sim(P \vee Q) \equiv (\sim P) \wedge (\sim Q).$
- (b) $\sim(P \wedge Q) \equiv (\sim P) \vee (\sim Q).$

.

Indices,
collections,
& sets of
sets

$I = A_i$ = the "index set"

$$\bigcup_{i=1}^n = A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n$$

$$\bigcap_{i=1}^n = A_1 \cap A_2 \cap A_3 \cap \dots \cap A_n$$

} two examples of an indexed collection of sets

Properties of
Cartesian
products

(1) $A \times B \neq B \times A$

(2) If $A = \emptyset$ or $B = \emptyset$, then $A \times B = \emptyset$

(3) If A and B are finite sets, then $|A \times B| = |A| \times |B|$

Some of the important properties of Cartesian products of sets are given below.

(i) Two ordered pairs are equal, if and only if the corresponding first elements are equal and the second elements are also equal.

(ii) If there are m elements in A and n elements in B , then there will be mn elements in $A \times B$. That means if $n(A) = m$ and $n(B) = n$, then $n(A \times B) = mn$.

(iii) If A and B are non-empty sets and either A or B is an infinite set, then $A \times B$ is also an infinite set.

(iv) $A \times A \times A = \{(a, b, c) : a, b, c \in A\}$. Here (a, b, c) is called an ordered triplet.

(v) The Cartesian product of sets is not commutative, i.e. $A \times B \neq B \times A$

(vi) The Cartesian product of sets is not associative, i.e. $A \times (B \times C) \neq (A \times B) \times C$

(vii) If A is a set, then $A \times \emptyset = \emptyset$ and $\emptyset \times A = \emptyset$.

(viii) If A and B are two sets, $A \times B = B \times A$ if and only if $A = B$, or $A = \emptyset$, or $B = \emptyset$.

(ix) Let A, B and C be three non-empty sets, then,

- $A \times (B \cap C) = (A \times B) \cap (A \times C)$
- $A \times (B \cup C) = (A \times B) \cup (A \times C)$
- $(A \cap B) \times C = (A \times C) \cap (B \times C)$
- $(A \cup B) \times C = (A \times C) \cup (B \times C)$

LINEAR ALGEBRA

95% of applications
use LU Factorization

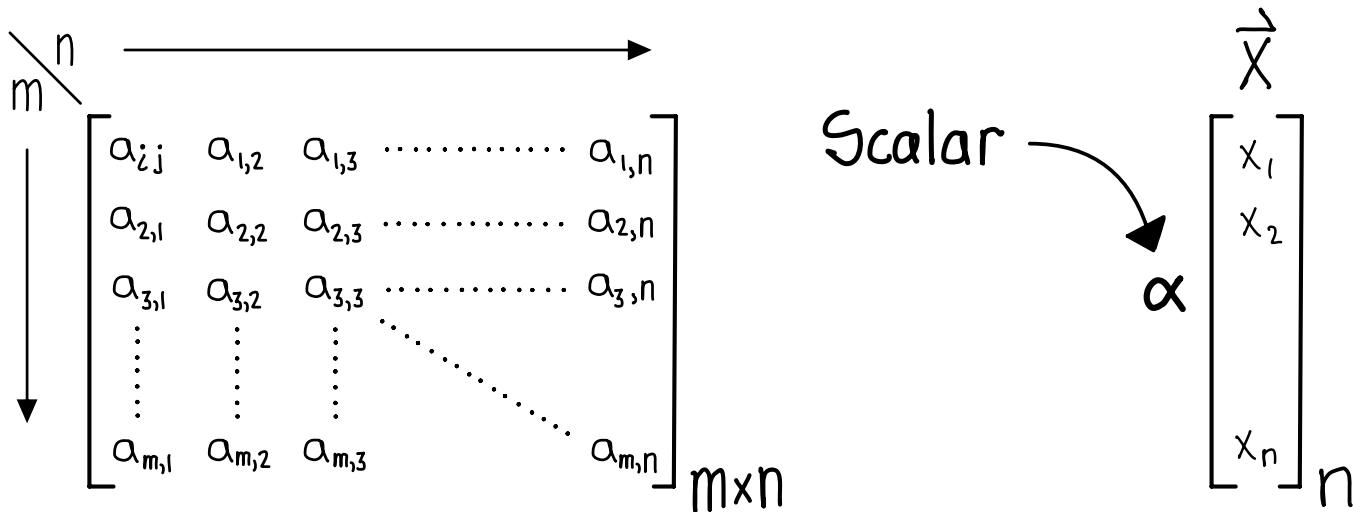
Highlight:

- 0) mostly about massaging data into a system of equations; the best methods "set the table" for a computationally efficient procedure
- 1) Linear Algebra is the study of linear transformations over data
- 2) Various methods for setting up and solving a system of (linear) equations:
 - Gaussian Elimination
 - LU & LDU Factorization
 - QR Factorization
 - Lagrange Multiplier

"linear" in parameters

all of these can be executed within software (MatLab), but it's up to the analyst to decide which method is most appropriate.
- 3) Extremely satisfying notation; can represent multidimensional data very intuitively with vectors, scalars, & matrices
- 4) A "must know" for statistics/econometrics, computer sci /programming, and quantitative finance
- 5) Unit length = 1 (normalization / direct comparisons)
- 6) MATRIX multiplication is not commutative! But note that vectors & scalars have different properties than matrices.
- 7) Scalars are element-wise operations used to scale up ($\alpha > 1$) or scale down ($0 \leq \alpha < 1$) a vector's magnitude as well as reverse its direction i.e. "flip" over the y-axis ($\alpha < 0$)
- 8) vectors are sequences; dot products are series
- 9) $\text{span}(A)$ and $\text{rk}(A)$ are interrelated concepts that describe different aspects of the structure and capabilities of matrix A
- 10) basic linear algebra assumes unique solutions; advanced linear algebra deals with "ill-conditioned" systems.
- 11) inner product (dot) is a vector operation that combines two vectors in the same vector space, resulting in a scalar that reflects aspects of their magnitude and angle. Note that $\sqrt{\langle a, a \rangle} = \|a\|_2 = a's \text{ magnitude}$.
- 12) To understand orthogonality/orthonormality watch Khan Academy Intro to Orthonormal bases

The Basics: Matrix Multiplication, Addition, and Subtraction



a matrix has m rows, n cols.

"single column" vector

- matrix multiplication vs. the "dot product" :

~ note that #cols (in the first matrix) must equal #rows (in the second matrix)
or else two matrices cannot be multiplied together.

① A SCALAR AND A VECTOR:

Commutative property holds for scalar/vector combo
and, although not shown here, it also holds for
a scalar/matrix combo.

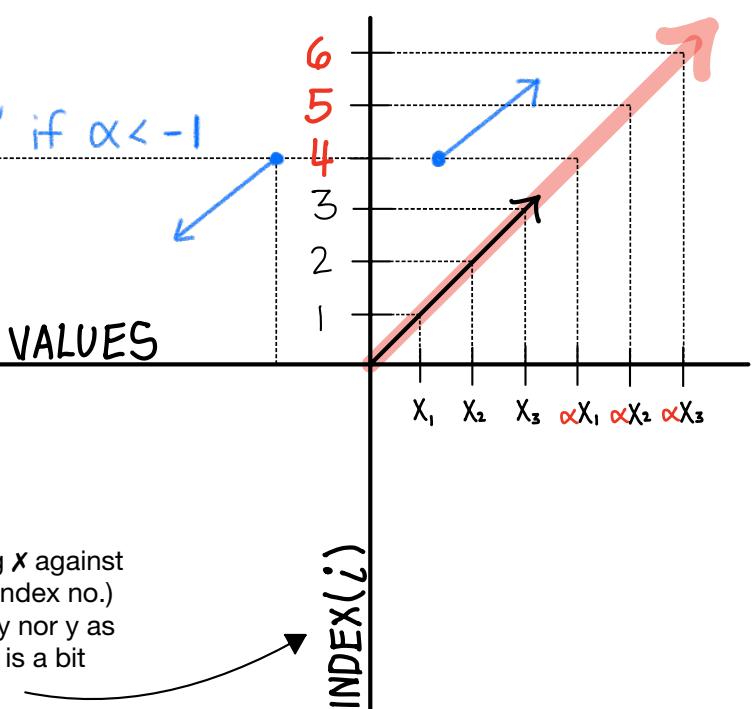
$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \alpha = \alpha \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} (\alpha)(X_1) \\ (\alpha)(X_2) \\ (\alpha)(X_3) \end{bmatrix}$$

$3 \times 1 \quad 3 \times 1 \quad 3 \times 1$

"flip" if $\alpha < -1$

VALUES

might look something like this ...



WARNING

here, I am plotting X against its subscript (the index no.)
not X as a fnc. of y nor y as a fnc. of X . Which is a bit
unconventional.

② TWO VECTORS (dot product): $\vec{y}^T \vec{x} = \vec{x}^T \vec{y} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}_{1 \times 3} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_{3 \times 1} = \sum_{i=1}^{n=3} x_i y_i$

↑

Communative property holds for dot products, too

$(x_1)(y_1) + (x_2)(y_2) + (x_3)(y_3)$

notice how we may use dot product to calculate a weighted average or expected value

- ③ A MATRIX AND A VECTOR:
non-communative; $AB \neq BA$

$$A\vec{x} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}_{2 \times 3} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{3 \times 1} = \begin{bmatrix} (a)(x_1) + (b)(x_2) + (c)(x_3) \\ (d)(x_1) + (e)(x_2) + (f)(x_3) \end{bmatrix}_{2 \times 1}$$

✓ 3x1

- ④ TWO SQUARE MATRICES:
non-communative; $AB \neq BA$

$$AB = \begin{bmatrix} a & b \\ c & d \end{bmatrix}_{2 \times 2} \begin{bmatrix} \omega & x \\ y & z \end{bmatrix}_{2 \times 2} = \begin{bmatrix} (a)(\omega) + (b)(y) & (a)(x) + (b)(z) \\ (c)(\omega) + (d)(y) & (c)(x) + (d)(z) \end{bmatrix}_{2 \times 2}$$

- ⑤ [special case] let \vec{x} be a sequence of real numbers $\{-1, .75, 2.5\}$ and let $\vec{y} = \{1, 1, 1\}$.
The the dot produce is the cum sum of all numbers in \vec{x} :

$$\vec{x} = \begin{bmatrix} -1 \\ .75 \\ 2.5 \end{bmatrix}, \vec{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Rightarrow \vec{x}^T \vec{y} = \begin{bmatrix} -1, .75, 2.5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = (-1)(1) + (.75)(1) + (2.5)(1) = (-1) + (.75) + (2.5)$$

Can we divide matrices?

Yes, but not how you think... just as numbers have reciprocals, matrices have inverses.

- numbers: let $x = \frac{a}{b}$, then $x^{-1} = \frac{b}{a} \Rightarrow x x^{-1} = 1$
- numbers: $\frac{a}{b} = a * b^{-1} \equiv \frac{a}{\frac{1}{b}} = (\frac{a}{1})(\frac{1}{b})$
- matrices: let A be a square ($n \times n$) matrix, then A is invertible iff $[A][A]^{-1} = [I]$; iff the equality holds, then $[A]^{-1}$ exists and also $[A]^{-1}[A] = [I]$ (the only time matrix multiplication is "communative")
- matrices: $[A][A]^{-1} = [I] \Rightarrow$ We may "divide" $[B]$ by $[A]$ as $[B][A]^{-1}$ iff A and B are both square matrices of the same dimensions.

- matrix addition & subtraction is an element-wise operation:

A **matrix** is a rectangular arrangement of numbers into rows and columns. Each number in a matrix is referred to as a **matrix element** or **entry**.

3 columns

2 rows ↓ ↓ ↓

$$\begin{array}{l} \rightarrow \\ \rightarrow \end{array} \begin{bmatrix} -2 & 5 & 6 \\ 5 & 2 & 7 \end{bmatrix}$$

The **dimensions** of a matrix give the number of rows and columns of the matrix *in that order*. Since matrix A has 2 rows and 3 columns, it is called a 2×3 matrix.

Adding matrices

Given $\mathbf{A} = \begin{bmatrix} 4 & 8 \\ 3 & 7 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 5 & 2 \end{bmatrix}$, let's find $\mathbf{A} + \mathbf{B}$.

We can find the sum simply by adding the corresponding entries in matrices \mathbf{A} and \mathbf{B} . This is shown below.

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 4 & 8 \\ 3 & 7 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 5 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 4+1 & 8+0 \\ 3+5 & 7+2 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 8 \\ 8 & 9 \end{bmatrix}$$

Basic/General Knowledge

- Main Diagonal of a Matrix: $\text{diag}(A) =$

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}_{3 \times 3} = \begin{bmatrix} a_{1,1} \\ a_{2,2} \\ a_{3,3} \end{bmatrix}$$

- Trace: $\text{trace}(A) = \text{sum of all main diagonal entries}$
- To "diagonalize" a square matrix: $A = PDP^{-1} \Leftrightarrow D = P^{-1}AP$
- cardinality: the number of elements in a set
- dimensions of a matrix: $\text{dim}(A)$: m rows, n cols
- eigenvalue/eigenvector: For a square matrix A , an eigenvalue (scalar) λ and its corresponding eigenvector \vec{v} (non-zero) satisfy $A\vec{v} = \lambda\vec{v}$
- spectrum: the spectrum of a matrix refers to its eigenvalues
- system of equations: a collection of at least two equations, which together imply solution(s) to unknowns.
- ill-conditioned system: Systems in which very small changes in the coef. matrix lead to large changes in the solution i.e., the system is not "numerically stable". Thus, ill-conditioned systems have ∞ -many numerical solutions, but perhaps only one true solution.
- perturb: means to make small adjustments to matrix entries i.e., to generate a matrix $C = A + B$ where $\text{dim}(A) = \text{dim}(B)$ and B contains very small values; an explicit test of ill-conditioning.
- inconsistent: systems with no solution(s) at all
- consistent: systems with at least one sol. exists; however may or may not be an ill-conditioned system
- span: The "span" of a matrix is the entire real no. space $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3$ etc. (if the matrix is full-rank) or a subset of real numbers that form linear combinations among the matrix's cols (if the matrix is not full-rank). Logically, if $\text{span}(A)$ is a subset of real numbers – as in the latter – matrix A suffers from multicollinearity. See appendix for more on span vs. rank and the implications of having linearly DEPENDENT column vectors within a matrix.

$\text{span}(V) \stackrel{\text{Generally}}{=} \text{all } v_i \in V \text{ that can be written as } c_1 v_1 + c_2 v_2 + \dots + c_n v_n$

$\text{span}(V) \stackrel{\text{Full Rank}}{=} \mathbb{R}^n, n = \# \text{ linearly INDEPENDENT cols} \stackrel{\text{full rank}}{=} \# \text{cols.}$

Square Matrices

- symmetric matrix:

$$A = A^T$$

- a square "matrix of ones":

self explanatory...

- a "diagonal matrix":

$$\begin{bmatrix} a_{1,1} & 0 & 0 & 0 \\ 0 & a_{2,2} & 0 & 0 \\ 0 & 0 & a_{3,3} & 0 \\ 0 & 0 & 0 & a_{4,4} \end{bmatrix}$$

4×4

- tri-diagonal matrix:

$$\begin{bmatrix} \dots & & & 0 \\ & \dots & & \\ & & \dots & \\ 0 & & & 0 \end{bmatrix}$$

- lower-triangular matrix:

$$\begin{bmatrix} a_{1,1} & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$$

3×3

- upper-triangular matrix:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ 0 & a_{2,2} & a_{2,3} \\ 0 & 0 & a_{3,3} \end{bmatrix}$$

3×3

, REF is one example

- Identity matrix:

$$I_3 =$$

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

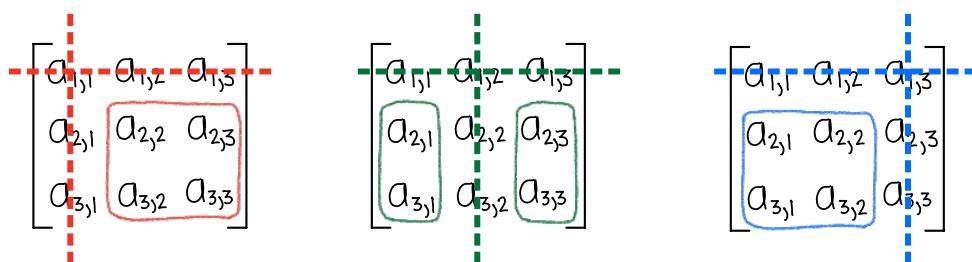
3×3

$$, A_{3 \times 3} I_3 = A_{3 \times 3}$$

- Submatrix: every matrix entry has exactly one submatrix; it is everything leftover in the parent matrix after excluding the entry's row & its column.

◦ e.g.)

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$$



the submatrix corresponding to $a_{1,1}$ is: $\begin{bmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{bmatrix}$

the submatrix corresponding to $a_{1,2}$ is: $\begin{bmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{bmatrix}$

the submatrix corresponding to $a_{1,3}$ is: $\begin{bmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{bmatrix}$

- Magnitude of a vector aka the norm*: $\|V\| = \sqrt{V_1^2 + V_2^2 + V_3^2 + \dots + V_n^2}$

Is conceptually similar to Euclidean distance, which is why we refer to magnitude as the Euclidean norm, or just "norm" for short; technically it's called the 2-norm. Other types of norms include:

- 1-norm $\|\vec{x}\|_1 = \left(\sum_{i=1}^n |x_i|^1 \right)^{\frac{1}{1}}$ = absolute value of largest col. sum
- p-norm (general case) $\|\vec{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$
- ∞ -norm $\|\vec{x}\|_{\infty} = \left(\sum_{i=1}^n |x_i|^{\infty} \right)^{\frac{1}{\infty}}$ = absolute value of largest row sum
- Determinant of a matrix*: $\det(A) = |A|$:
- $|1 \times 1$ matrix: Let $A = [a]$, Then $|A| = a$
- $|2 \times 2$ matrix: Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}_{2 \times 2}$, Then $|A| =$
- $|3 \times 3$ matrix: Let $A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}_{3 \times 3}$, $|A| = +a_{1,1} \begin{vmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{vmatrix}_{2 \times 2} - a_{1,2} \begin{vmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{vmatrix}_{2 \times 2} + a_{1,3} \begin{vmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{vmatrix}_{2 \times 2}$

* the determinant is to a matrix what magnitude is to a vector; can think of it as a matrix's "magnitude" or a measure of size

- Frobenius norm (an alternative to the determinant): $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\text{trace}(A^* A)}$, A^* is "the conjugate transpose of A " (see appendix for full detail)

- $\vec{A}\vec{x} = \vec{b}$, A is called the "transition" or "coefficient" matrix
- a single col. vector \vec{v} times a basis B is a linear combination and B is derived from A very intentionally

- Let $B = \{\vec{c}_1 | \vec{c}_2 | \vec{c}_3 | \vec{c}_4 | \vec{c}_5\}$ s.t. " B is a basis for \vec{V} "

$$\circ B\vec{v} = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 \\ 0 & 0 & 0 & \lambda_4 & 0 \\ 0 & 0 & 0 & 0 & \lambda_5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 + \lambda_4 v_4 + \lambda_5 v_5$$

$B_{n \times n}$ \vec{v}_n

- B is very important; its structure determines everything
 - (1) eigen-basis via finding the eigenspaces for A
 - (2) orthogonal basis via Gram-Schmidt
 - (3) orthonormal basis via Modified Gram-Schmidt

- desirable features of B
 - i. ensures numerical stability (that's the whole point)
 - ii. is a diagonal matrix (easy to work with), but doesn't have to be.

- In general:

\vec{i} is the row index
 \vec{j} is the col. index

$$A\vec{x} = \vec{b}, a_{i,j} \in A \quad \equiv \quad \underbrace{\begin{matrix} m \text{ rows} \\ \left(\begin{array}{ccccc|c} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & a_{1,5} & b_1 \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & a_{2,5} & b_2 \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & a_{3,5} & b_3 \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} & a_{4,5} & b_4 \\ a_{5,1} & a_{5,2} & a_{5,3} & a_{5,4} & a_{5,5} & b_5 \end{array} \right)}_{n \text{ cols.}} = \begin{matrix} \vec{x} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{matrix} \quad \begin{matrix} \vec{b} \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{matrix}$$

KNOWN OBSERVATIONS UNKNOWN (SOLVE FOR)

$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + a_{15}x_5 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + a_{25}x_5 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + a_{35}x_5 &= b_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + a_{45}x_5 &= b_4 \\ a_{51}x_1 + a_{52}x_2 + a_{53}x_3 + a_{54}x_4 + a_{55}x_5 &= b_5 \end{aligned}$

System of Linear Equations

def: 2 or more linear equations

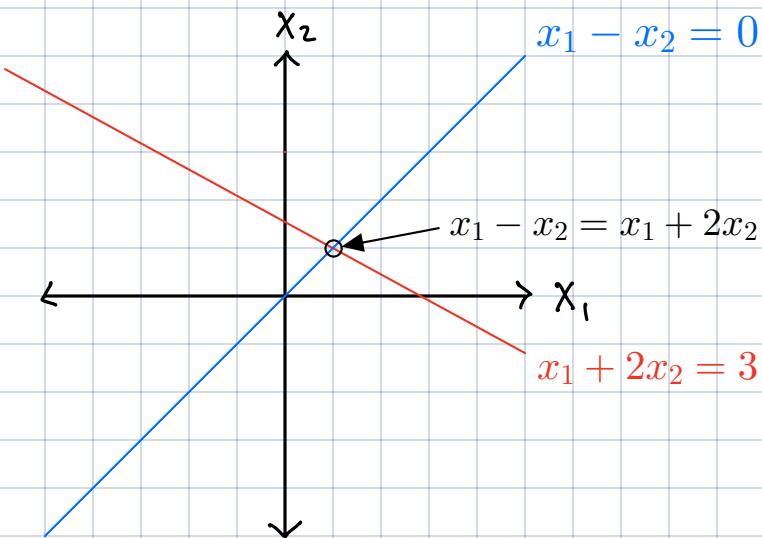
- Sol. is the set of points that satisfy all equations in the system

Ex:

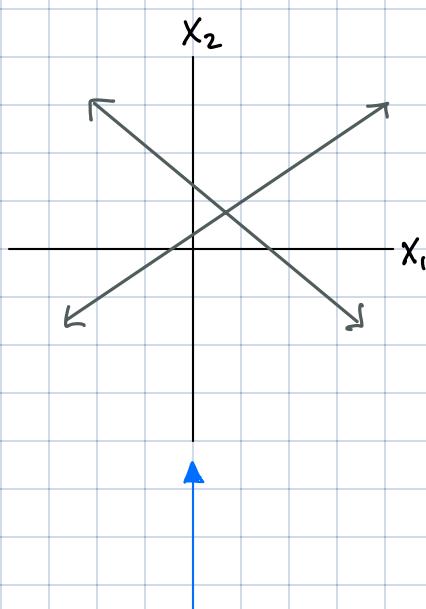
$$x_1 + 2x_2 = 3 \Rightarrow x_1 = 3 - 2x_2$$

$$x_1 - x_2 = 0 \Rightarrow x_1 = x_2$$

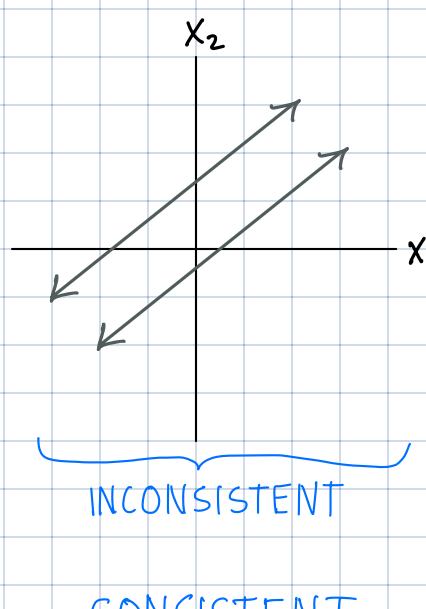
The sol. set is visualized by the intersection:



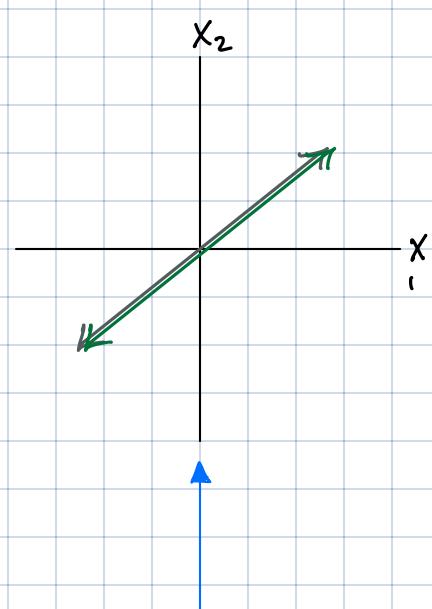
one (unique) sol.



no sol.



many sol.



def: a sys. of linear equations is consistent if at least one sol.

THE PURPOSE OF LINEAR ALGEBRA IS TO DEVELOP AND USE ALGORITHMS FOR SOLVING SYSTEMS OF LINEAR EQUATIONS.

Ex: Solve the following sys. of linear equations at a high level

$$4x_1 + 5x_2 + 3x_3 + 3x_4 = 1$$

$$x_1 + x_2 + x_3 + x_4 = 0$$

$$2x_1 + 3x_2 + x_3 + x_4 = 1$$

$$5x_1 + 7x_2 + 3x_3 + 3x_4 = 2$$

Step 1 The first step is to construct either a coefficient matrix or an augmented matrix

coefficient matrix:

$$\begin{bmatrix} 4 & 5 & 3 & 3 \\ 1 & 1 & 1 & 1 \\ 2 & 3 & 1 & 1 \\ 5 & 7 & 3 & 3 \end{bmatrix} = \{\vec{c}_1 | \vec{c}_2 | \vec{c}_3 | \vec{c}_4\}$$

augmented matrix:

$$\left[\begin{array}{cccc|c} 4 & 5 & 3 & 3 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 2 & 3 & 1 & 1 & 1 \\ 5 & 7 & 3 & 3 & 2 \end{array} \right]$$

- the augmented matrix nicely packages up the entire sys. of linear equations into a single matrix

Step 2 Use elementary row operations to convert augmented matrix into RREF

- (1) Interchange e.g. $R_1 \leftrightarrow R_2$
- (2) Scaling e.g. $3R_1 \rightarrow R_1$
- (3) Replacement e.g. $3R_1 + R_2 \rightarrow R_2$

* ALL ELEMENTARY ROW OPERATIONS ARE REVERSABLE *
* DOES NOT CHANGE THE SOLUTION SET *

def: Two matrices are row equivalent if a sequence of row operations transforms one into the other. For example, A is consistent w/ its RREF if its RREF exists.

Step 3 Q: Is the matrix consistent (i.e., at least one sol. exists)? If not, the sys. cannot be solved so stop here. If yes, proceed to step 4.

Step 4 Q: How many solutions exist? $[A]^{-1}$ exists iff $[A][A]^{-1} = [I]$

- a) Verify that A is invertible → proceed w/ Gaussian Elimination, LU Factorization, Cholesky Factorization, or QR Factorization (more on these later...)
- b) if A is not invertible, then no unique sol. exists ⇒ an infinite number of solutions exist *this does not mean that none are optimal*

def: a pivot position is the row/col index no. of the leading entry in $\text{ref}(A)$ or $\text{rref}(A)$

def: a pivot column is the corresponding column

*cannot see the pivots of matrix A unless converted to $\text{ref}(A)$ or $\text{rref}(A)$ *

ex) $\text{rref}(A) = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$

Steps to convert A into $\text{rref}(A)$ - ALGORITHM

the whole point of converting A to $\text{rref}(A)$ is to reveal the pivots

- | | |
|----------------|--|
| Forward Phase | <p>Step 1 Begin w/ the left-most, non-zero col</p> <p>Step 2 Make the first entry a non-zero "pivot"; interchange rows if needed</p> <p>Step 3 Use elementary row operations to set the pivot=1 and to put all 0's below the pivot</p> |
| Backward Phase | <p>Step 4 Repeat Steps 1-3 for the submatrix</p> <p>Step 5 Beginning w/ the right-most pivot, work up & left to make all 0's above and below it</p> |

Parametric Description / Parametric Form

Each col in A is either a basic variable aka pivot variable or a free variable aka free parameter.

ex) $\text{rref}(A) = \begin{bmatrix} 1 & -1 & 0 & 2 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & -1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Basic Vars: x_1, x_3, x_6
Free Vars: x_2, x_4, x_5
can take on any value

no sol. here if $a_{47} \neq 0$

From here, want to express basic vars. in terms of free vars...

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

consistent estimators?

Thm: a linear sys. is consistent iff the right-most col of augmented matrix is not a pivot col

💡 IS THIS IS WHAT ECONOMETRITIANS MEAN BY A "CONSISTENT ESTIMATOR"? (would make sense b/c "if one estimator is inconsistent, they all are" - Wooldridge) 💡

Requirements for a unique solution set

- ① $A\vec{x} = \vec{b}$ must be a consistent sys. of linear equations
- ② A must be square matrix and full-rank, and thus invertible
- ③ $\det(A)$ must be non-zero

Another Example of Parametric Form

$$\text{rref}(A) = \left[\begin{array}{ccc|c} 1 & 3 & 0 & -5 \\ 0 & 0 & 1 & 3 \end{array} \right] \Rightarrow X_1 + 3X_2 = -5 \Rightarrow X_1 = -5 - 3X_2$$

$$X_3 = 3$$

X_2 is a "free variable" to X_1 b/c X_1 is a function of X_2

We know that X_1 is the "basic variable" b/c **1st pivot** corresponds to X_1

Vector Equations

def: a vector is a matrix w/ exactly one column \equiv refers to a matrix's column(s)

ex) $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3$, $\mathbb{R}^n = \{\vec{v} \mid \vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, v_i \in \mathbb{R} \forall i \in \mathbb{N}\}$

- use vectors to form systems of linear equations (i.e., a "linear combination")
- use matrices to represent systems of linear equations concisely
- solve systems of linear equations using Gaussian Elim. or factorization methods

Thm: \vec{b} is a linear combination of \vec{v}_1, \vec{v}_2 , and \vec{v}_3 iff A is consistent

where $A = [\vec{v}_1 \vec{v}_2 \vec{v}_3 \vec{b}]$

def: the set of all possible linear combinations $\vec{b} \in \mathbb{R}^n$ is called **span**

ex) $\vec{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathbb{R}^2$

SPAN

$$\Rightarrow \text{span}\{\vec{v}\} = \{C_1 \vec{v}\} = \mathbb{R}^2$$

b/c C_1 is "free" and can therefore take on any value

Linear Independence

The "trivial solution" to a sys. of homogenous equations (i.e., equations whose terms are all of the same degree) is
 $c_1 = c_2 = \dots = c_k = 0$

def: a set of vectors $\{v_1, v_2, \dots, v_k\}$ are linearly independent if the vector equation $c_1v_1 + c_2v_2 + \dots + c_kv_k$ has only the trivial sol.

ex) $\left\{ \begin{bmatrix} 2 \\ -3 \end{bmatrix}, \begin{bmatrix} -4 \\ 6 \end{bmatrix} \right\}$ are these vectors linearly independent?

① $c_1 \begin{bmatrix} 2 \\ -3 \end{bmatrix} + c_2 \begin{bmatrix} -4 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ then yes, else no.

② $\left[\begin{array}{ccc|c} 2 & -4 & 0 \\ -3 & 6 & 0 \end{array} \right] \xrightarrow{\text{REF}} \left[\begin{array}{ccc|c} 6 & -12 & 0 \\ 0 & 0 & 0 \end{array} \right]$ *
 C₂ is a free variable \Rightarrow inf. many sol. \Rightarrow not only the trivial sol. \Rightarrow no

CONCLUSION: for every col. in A w/o a pivot, you get a free variable, which implies inf. many solutions.
 Therefore the two vectors are not linearly independent (since the trivial solution is not the only solution).

* This is why REF and RREF is so powerful. It enables systematic backward ellim. consider a homogenous sys. represented by the following augmented matrix:

| | | |
|---|---|---------------------------|
| $a_{11} a_{12} a_{13} a_{14} a_{15} a_{16} b_1$ | C ₁ is determined by C ₂ , C ₃ , C ₄ , C ₅ , C ₆ , and row1; no remaining parameters are free | 6 th ITERATION |
| $0 a_{22} a_{23} a_{24} a_{25} a_{26} b_2$ | C ₂ is determined by C ₃ , C ₄ , C ₅ , C ₆ , and row2; C ₁ is free | 5 th ITERATION |
| $0 0 a_{33} a_{34} a_{35} a_{36} b_3$ | C ₃ is determined by C ₄ , C ₅ , C ₆ , and row3; C ₁ and C ₂ are free | 4 th ITERATION |
| $0 0 0 a_{44} a_{45} a_{46} b_4$ | C ₄ is determined by C ₅ , C ₆ , and row4; C ₁ , C ₂ , and C ₃ are free | 3 rd ITERATION |
| $0 0 0 0 a_{55} a_{56} b_5$ | C ₅ is determined by C ₆ and row5; C ₁ , C ₂ , C ₃ , C ₄ , C ₅ , C ₆ are free | 2 nd ITERATION |
| $0 0 0 0 0 a_{66} b_6$ | C ₆ is determined by row6 equality; C ₁ , C ₂ , C ₃ , C ₄ , C ₅ , C ₆ are free | 1 st ITERATION |

IF THESE WERE ALL ZEROS, THE SYS. WOULD BE "HOMOGENOUS" (which in it of itself implies "consistency" i.e., there IS AT LEAST ONE SOL.) AND SINCE EVERY ROW IN A HAS A NONZERO PIVOT, WOULD RESULT IN ONLY THE TRIVIAL SOL. \Rightarrow INDEPENDENCE

THESE ZEROS SHOW THAT A IS IN REF

TO SUMMARIZE:

- "consistent" means "at least one sol."; occurs when every col. in REF(A) has a non-zero leading entry (i.e., A has non-zero "pivots")
 - "homogenous" means $\vec{b} = 0$ in $A\vec{x} = \vec{b}$; and all homogenous sys. are consistent (happens when $a_1x_1 + a_2x_2 + \dots + a_nx_n = 0$ for all rows)
 - If REF(A) is homogenous AND appears as a perfect "staircase", then $\vec{x} = 0$ (called the "trivial solution") is the only sol. and there are no "free" variables, signifying that the \vec{a} vectors are linearly independent of each other.
-

Q: what if A is not square?

A: the backwards ellim might fall apart at a certain point... consider:

| | | | | | | | | | |
|----------|----------|----------|----------|----------|----------|----------|-------|---|---------------------------------|
| a_{11} | a_{12} | a_{13} | a_{14} | a_{15} | a_{16} | a_{17} | b_1 | C_1 now depends on C_2 & C_3 i.e., $C_1 = f(C_2, C_3)$ | } \Rightarrow inf. many sol 😐 |
| 0 | a_{22} | a_{23} | a_{24} | a_{25} | a_{26} | a_{27} | b_2 | C_2 now depends on C_3 i.e., $C_2 = f(C_3) \equiv C_3 = f(C_2)$ | |
| 0 | 0 | 0 | a_{34} | a_{35} | a_{36} | a_{37} | b_3 | EVERYTHING OK SO FAR... we have unique solutions for C_7, C_6, C_5, C_4 | |
| 0 | 0 | 0 | 0 | a_{45} | a_{46} | a_{47} | b_4 | ✓ | |
| 0 | 0 | 0 | 0 | 0 | a_{56} | a_{57} | b_5 | ✓ | |
| 0 | 0 | 0 | 0 | 0 | 0 | a_{77} | b_6 | ✓ | |

The problem here is that C_2 and C_3 can be proportionally set to enforce the row 2 equality: $C_2 \propto C_3$ s.t. $C_1(0) + C_2(a_{22}) + C_3(a_{23}) + C_4(a_{24}) + C_5(a_{25}) + C_6(a_{26}) + C_7(a_{27}) = b_2$

Therefore, either C_2 or C_3 is "free" and there are inf. many sol. as a result.

Q: Why does "only the trivial sol." imply indepedence when dealing w/ a homogenous sys?

A: because otherwise at least one C_j is "free"

Q: how do $A\vec{z} = \vec{b}$ and multiple linear regression relate?

A: (1) the \vec{a} vectors are observed predictors of \vec{b}

(2) We solve $A\vec{z} = \vec{b}$ for \vec{z} , which are our regression coefficients aka our "estimators" *

(3) if A's rows are linearly independent, then so are A's columns (i.e., the \vec{a} vectors are then independent of each other) and the "no perfect collinearity" MLR assumption holds.

(4) $A_{m \times n} \equiv m$ # of observations, n # of independent variables

* it's a bit more complicated than this... see Wooldridge Appendix E along w/ this binder's appendix.

Regression Mechanics and the Least Squares Problem

$$\vec{C} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix}, \quad \vec{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \quad A = \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \vec{x}_n \\ x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{bmatrix} \longrightarrow \text{solve } A \vec{\beta} = \vec{C}$$

Same as: $\hat{y} \xrightarrow{\text{zero-intercept MLR}} \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n + \epsilon$

Best fit is qualified by the sum of squared errors:

$$\sqrt{\sum_{i=1}^n \epsilon_i^2}$$

WARNING

It's very unconventional to write out an implicit constraint; typically only explicit constraints are written out. Here, the Betas are your decision variables and the linear combination (representing a Multiple Linear Regression model) is the framework that feeds into our objective function. In Pyomo (Python) you would not actually state this as a constraint but rather write the objective function as a function of all other variables i.e., the error term/residual as a function of beta and x which are the underlying variables and declare the betas as your decision variable. To save space on the page I've decided to do it this way instead!

| IMPLICIT (not explicit) CONSTRAINT | |
|---|--|
| $\min \sqrt{\sum_{i=1}^n \epsilon_i^2}$ | s.t. $\hat{y}_i = \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_n x_{in} \quad \forall x_{ij} \in A$ <div style="text-align: right; margin-top: -20px;"> col. index \downarrow row index ↑ </div> |

where $\epsilon_i = \hat{y}_i - y_i \Rightarrow \epsilon_i^2 = (\hat{y}_i - y_i)^2$

which occurs where

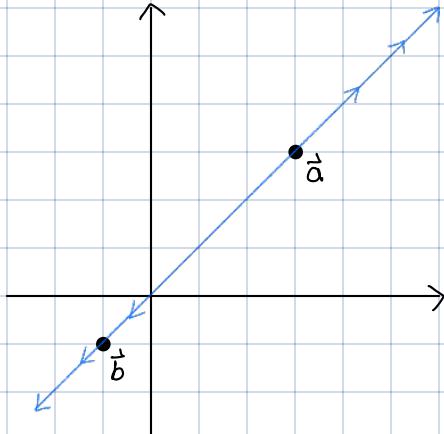
| LOCAL MINIMUM CONDITIONS | |
|---|--|
| $\frac{\partial \hat{y}}{\partial \epsilon} = 0$ and $\frac{\partial^2 \hat{y}}{\partial \epsilon^2} > 0$ | |

not entirely sure where an intercept term might come from...

Linear Independence (cont.)

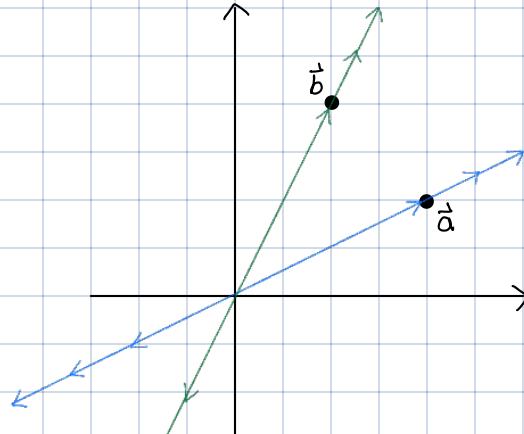
Thm: a set of vectors is linearly dependent iff at least one vector is a linear combination of the others i.e., \vec{a} and \vec{b} are dependent if $\text{span}\{\vec{a}\} = \text{span}\{\vec{b}\}$

DEPENDENT: $\text{span}\{\vec{a}\} = \text{span}\{\vec{b}\}$



$\text{span}\{\vec{a}, \vec{b}\} = \text{a LINE in } \mathbb{R}^2$

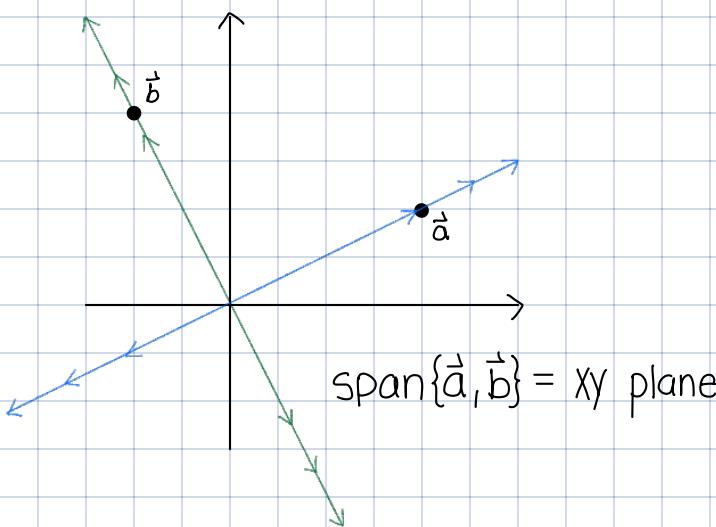
INDEPENDENT & CORRELATED: $\text{span}\{\vec{a}\} \neq \text{span}\{\vec{b}\}$



$\text{span}\{\vec{a}, \vec{b}\} = \text{xy plane}$

INDEPENDENT & UNCORRELATED:

$\text{span}\{\vec{a}\} \perp \text{span}\{\vec{b}\}$



$\text{span}\{\vec{a}, \vec{b}\} = \text{xy plane}$

$$\text{Cosine similarity} = \frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\| \|\vec{b}\|}$$

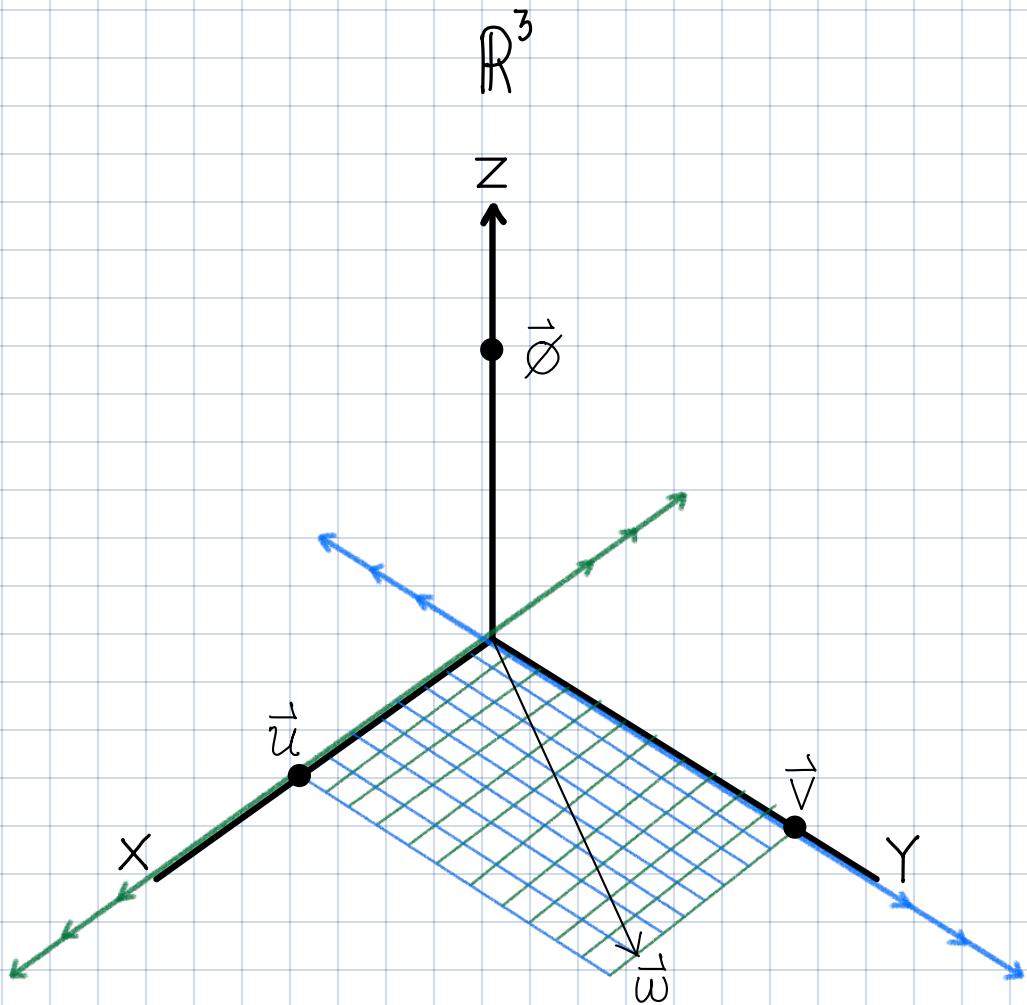
$[-1, 1]$

iff Pearson correlation (below)
centers \vec{a} and \vec{b} at their means

$$\rho_{\vec{a}, \vec{b}} = \frac{\text{cov}(a, b)}{\sigma_a \sigma_b} = \frac{\frac{1}{n-1} \sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (a_i - \bar{a})^2} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (b_i - \bar{b})^2}}$$

def: two vectors \vec{a} and \vec{b} are **orthogonal** when $\text{span}\{\vec{a}\}$ is perpendicular to $\text{span}\{\vec{b}\}$
(which also means that their dot product and cosine similarity equals zero!)

Two variables can be independent and still correlate (b/c correlation \neq causation), but they are "orthogonal" when independent AND uncorrelated.



The checkered space is all linear combinations of \vec{u} and \vec{v} and equals $\text{span}\{\vec{u}, \vec{v}\} \stackrel{\text{def}}{=} \{a\vec{u} + b\vec{v} \mid a, b \in \mathbb{R}\}$

$\{\vec{u}, \vec{v}, \vec{w}\}$ is linearly dependent b/c \vec{w} is redundant

$\{\vec{u}, \vec{v}\}$ and $\{\vec{u}, \vec{w}\}$ and $\{\vec{v}, \vec{w}\}$ are each linearly independent b/c their spans aren't on the same line

$\{\vec{u}, \vec{v}, \vec{\phi}\}$ is linearly independent and an "orthogonal basis" for \mathbb{R}^3 (since $\vec{u} \perp \vec{v} \perp \vec{\phi}$)

$\{\vec{u}, \vec{v}, \vec{\phi}, + \text{any more than 3 vectors in } \mathbb{R}^3\}$ is linearly dependent

Thm: if a set of vectors contains the zero vector in \mathbb{R}^n , the set is always automatically linearly dependent; also any set $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p\} \in \mathbb{R}^n$ is dependent if $p > n$.

 You

If vectors a and b are independent, $\text{span}\{a,b\} = xy$ plane. Is that the entire xy plane or a bounded xy plane?

 ChatGPT

If vectors a and b are independent and span the xy plane, it means that they form a basis for the xy plane. This implies that any vector in the xy plane can be expressed as a linear combination of a and b . The term "span" refers to the set of all possible linear combinations of a given set of vectors.

In this context, when we say that the span of a and b is the xy plane, we are referring to the entire xy plane, not just a bounded or limited region of it. The concept of independence ensures that a and b are not collinear, which means they don't lie on the same line, and thus can cover the whole plane through their combinations. This coverage is unbounded, meaning it extends infinitely in all directions within the plane.

 You

correct or incorrect: a and b do not have to be orthogonal for your explanation to hold.

 ChatGPT

Correct. Vectors a and b do not have to be orthogonal for the explanation to hold. The key requirement is that they are linearly independent, not orthogonal. Orthogonality is a stronger condition that implies independence (for non-zero vectors), but independence does not necessarily imply orthogonality. Two vectors are independent if one is not a scalar multiple of the other, meaning they don't point in the same or exactly opposite directions. This independence is sufficient to ensure that their span covers the entire plane they define, in this case, the xy plane.

 ChatGPT

Orthogonal vectors are vectors that meet at a right angle (90 degrees) to each other. In terms of their dot product, two vectors are orthogonal if and only if their dot product is zero. The dot product of two vectors $a = (a_1, a_2, \dots, a_n)$ and $b = (b_1, b_2, \dots, b_n)$ is calculated as $a \cdot b = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$. Thus, for orthogonal vectors a and b , $a \cdot b = 0$.

$$\text{ex)} \quad \vec{a} = \begin{bmatrix} 5 \\ 5 \\ 0 \end{bmatrix} \times \perp \vec{b} = \begin{bmatrix} -5 \\ 5 \\ 7 \end{bmatrix} \rightarrow \langle \vec{a} \mid \vec{b} \rangle = \begin{bmatrix} 5 & 5 & 0 \end{bmatrix} \begin{bmatrix} -5 \\ 5 \\ 7 \end{bmatrix} = (5)(-5) + (5)(5) + (0)(7) = 0$$

 You

Explain what about "spanning set" of vectors for \mathbb{R}^n is.

 ChatGPT

A "spanning set" of vectors for \mathbb{R}^n refers to a collection of vectors in \mathbb{R}^n from which any vector in \mathbb{R}^n can be expressed as a linear combination. A linear combination of a set of vectors $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ involves multiplying each vector by a scalar and then adding the results together. Formally, a vector \vec{v} in \mathbb{R}^n is a linear combination of a set of vectors if it can be written as:

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \dots + c_k\vec{v}_k$$

where c_1, c_2, \dots, c_k are scalars.

For a set of vectors to span \mathbb{R}^n , any vector in \mathbb{R}^n must be able to be represented in this way. This means that the spanning set "covers" the whole space in the sense that no point in \mathbb{R}^n is left out; every point can be reached by some combination of the vectors in the spanning set.

For example, in \mathbb{R}^2 , the set containing just the two vectors $\{(1, 0), (0, 1)\}$ is a spanning set because any vector (x, y) in \mathbb{R}^2 can be written as a linear combination of these two vectors:

$$(x, y) = x(1, 0) + y(0, 1)$$

Similarly, in \mathbb{R}^3 , the set of vectors $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ spans the space because any vector (x, y, z) in \mathbb{R}^3 can be expressed as:

$$(x, y, z) = x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1)$$

The concept of a spanning set is crucial in linear algebra because it helps in understanding the dimensions of a space, the basis of a vector space, and the solutions to systems of linear equations. A vector space is said to be finitely spanned if there exists a finite set of vectors that spans the space. The minimum number of vectors needed to span a vector space is called the dimension of the space. For \mathbb{R}^n , the dimension is n , and a set of n linearly independent vectors that spans the space is called a basis for the space.

INTRO TO LINEAR TRANSFORMATIONS AND THE LINEAR TRANSFORMATION MATRIX

" T is a map on vectors from \mathbb{R}^n to \mathbb{R}^m and a linear transformation on \vec{x} "

$$T: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

↑ ↑
"domain" "co-domain" $\stackrel{\text{def}}{=} \text{image}(T)$

$$T(\vec{x}): \vec{x} \xrightarrow{\text{"maps"} } A\vec{x}, \quad \vec{x} \subset \mathbb{R}^n, \quad A\vec{x} \subset \mathbb{R}^m$$

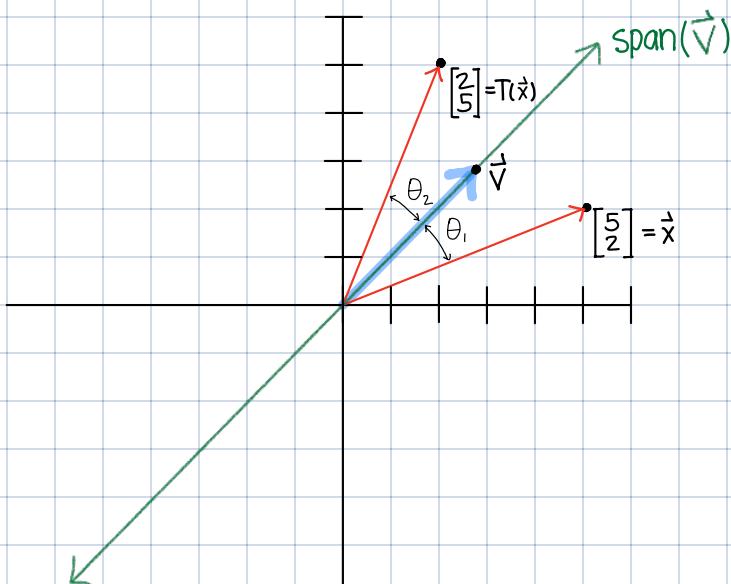
↓
"subspace"

common versions of $T(\vec{x})$ are:

- $T(\vec{x})$ as a scalar (analogous to a functional "stretch" in \mathbb{R}^2)
- $T(\vec{x})$ as a reflector (analogous to a functional "flip" in \mathbb{R}^2)
- $T(\vec{x})$ as a rotation (no functional analog in \mathbb{R}^2)
- $T(\vec{x})$ as a projection (no functional analog in \mathbb{R}^2)
- $T(\vec{x})$ as a shear
- $T(\vec{x})$ as a contraction or expansion

see coursera note packet on T7Shield for more detail

Sometimes, multiple of these options result in the exact same transformation, just via different execution; for example:



$T(\vec{x})$ is both a reflection of \vec{x} about $\text{span}\{\vec{v}\}$ and a rotation about the origin by $(\theta_1 + \theta_2)$

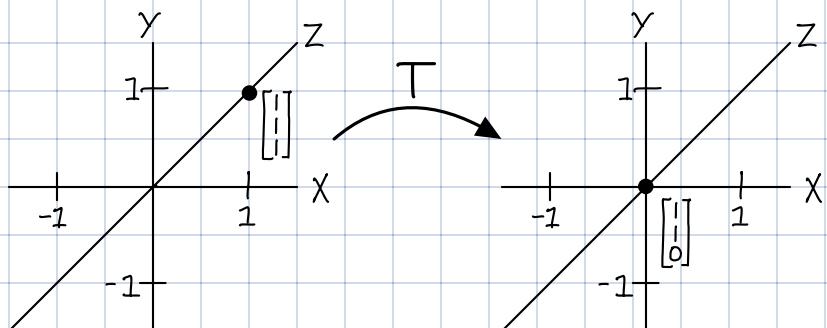
ex) $A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$ b/c let $\vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\rightarrow A\vec{x} = \underbrace{\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}}_{\vec{x} \mapsto A\vec{x}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \langle \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{x} \rangle \\ \langle \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \vec{x} \rangle \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} =$

def: the image of T is a linear combination of the cols. of A

$\text{image}(T) = \text{span}\{\text{cols. of } A\} = \text{col}(A)$ aka the "column space"
aka the "vector space"
aka the "output"

- if $\vec{b} \in \text{im}(A)$ means $\exists \vec{x}$ s.t. $A\vec{x} = \vec{b}$ holds $\Rightarrow A$ is "consistent"

ex) let $A\vec{x} = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array} \right] = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \mathbb{R}^3$:



- T is a "projection" down along the z axis ("sends z to the floor")

- $\text{im}(T) = \text{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} = \mathbb{R}^3$ is a linearly dependent set

$$\text{im}(T) = \text{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\} = \mathbb{R}^2 \text{ is a linearly independent set}$$

- $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $n = \# \text{ of cols in } A = 3$, $m = \# \text{ non-zero rows in } A = 2$ (dimensionality/reduction)

def: a transformation T is a linear transformation iff these two conditions hold

- $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$ i.e., T distributes over addition
- $T(C\vec{u}) = CT(\vec{u})$

def: a matrix A is positive definite if for any non-zero vector \vec{x} , the quadratic form $\vec{x}'A\vec{x} > 0$

def: a matrix A is positive definite if for any non-zero vector \vec{x} , the quadratic form $\vec{x}'A\vec{x} \geq 0$

def: a "non-zero" vector can contain 0's, but not exclusively; a non-zero vector is any vector that is not the zero vector

def: the trivial solution to $A\vec{x} = \vec{b}$ is the zero vector; a non-trivial solution to $A\vec{x} = \vec{b}$ is a non-zero vector

def: the image of a linear transformation or transformation matrix A wrt a non-zero vector $\vec{x} = \vec{b}$ in $A\vec{x} = \vec{b}$ i.e., the solution to $A\vec{x} = \vec{b}$

ChatGPT 4 >

 You

What is the "image" of a linear transformation T over a vector x

 ChatGPT

The "image" of a linear transformation T over a vector x , often denoted as $T(x)$, refers to the output vector obtained after applying the transformation T to the input vector x . In more formal terms, if you have a linear transformation $T: V \rightarrow W$, which maps vectors from space V to space W , then the image of $x \in V$ under T is the vector in W that is the result of applying T to x .

Mathematically, if x is a vector in V , then the image of x under T , denoted $T(x)$, is found by applying the transformation rules of T to x . This process often involves matrix multiplication when T is represented as a matrix and x as a column vector in the case of finite-dimensional vector spaces.

For example, if T is represented by a matrix A and x is a column vector, then the image of x under T , $T(x)$, is computed as Ax , which results in another vector. This vector Ax is the "image" of x under the transformation T .

The concept of the image is fundamental in understanding how linear transformations act on vector spaces, including their effects on the geometry of the space, such as stretching, compressing, rotating, or any combination thereof.

def: the standard vectors of \mathbb{R}^n are $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$, $\vec{e}_i = \begin{bmatrix} 0 \\ \vdots \\ i \\ \vdots \\ 0 \end{bmatrix}$ i^{th} coordinate

- ex) $\mathbb{R}^2 : \{e_1, e_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

- GOAL: find the standard matrix A s.t. $T(\vec{x}) = A\vec{x}$ in order to execute the desired linear transformation over \vec{x}

Thm: given a linear transformation $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$, find $A_{m \times n}$ s.t.

\exists a unique matrix A s.t. $T(\vec{x}) = A\vec{x}$.

- algorithm for finding A :

$A = [\vec{e}_1 | \vec{e}_2 | \cdots | \vec{e}_n]$ is called the "standard matrix" or "transition matrix"

def: Given $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is "onto" if $\forall \vec{b} \in \mathbb{R}^m, \exists \vec{x} \in \mathbb{R}^n$ s.t. $T(\vec{x}) = \vec{b}$

def: Given $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is "one-to-one" (1:1) if $\forall \vec{b} \in \underbrace{\text{im}(T)}_{\mathbb{R}^m}$ is the image of at most one $\vec{x} \in \mathbb{R}^n$

- if T is 1:1, then A is "full rank" (analogous to the horizontal line test)
- A being "full rank" is **necessary** for A to be invertible, it is not **sufficient** unless A is also square; if $n=m$ (specifically $n < m$), then A can still be "full rank" but it is not invertible b/c A is not square.

Thm: Given $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ T is 1:1 iff \exists pivot in every col A i.e., A 's cols are linearly independent iff $T(\vec{x}) = \vec{0} \iff A\vec{x} = 0$

Thm: Given $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$, T is "onto" iff \exists a pivot in every row of A iff $\text{im}(T) = \text{span}\{\text{cols. of } A\} = \mathbb{R}^m$ *

\uparrow to find A 's row pivots, must find $\text{ref}(A)$ or $\text{rref}(A)$

* these two theorems can be accurately stated in several ways, which can be confusing; a good rule-of-thumb is: $n < m \Rightarrow$ "onto" and $1:1 \Leftrightarrow Ax = 0$ has only the trivial solution