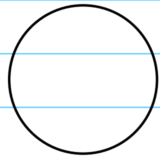
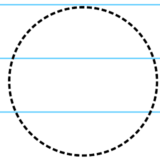
Notation for sets and mathematical logic

set notation

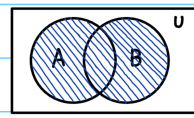


a **closed set** includes sample points on the barieer



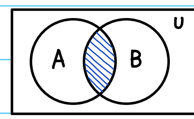
an **open set** does not

union:



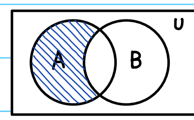
$A \cup B$

intersection:



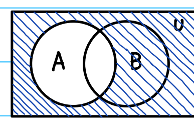
$A \cap B$

"difference":



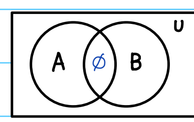
$A - B$ or $A \setminus B$ (A "remove" B)

"complement":



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

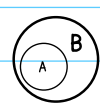
disjoint



$A \cap B = \emptyset$

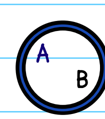
where $\emptyset$ denotes an "empty" set

subsets:



$A \subset B$

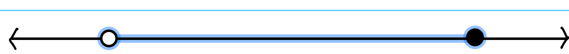
is a "proper" subset



$A \subseteq B$

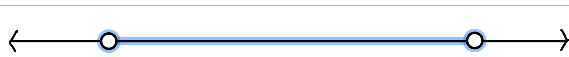
Interval notation

$[a, b]$ or $[a, b]$ or $a \leqslant x \leqslant b$



$(a, b]$ or $]a, b]$ or $a < x \leqslant b$

$[a, b)$ or $[a, b[$ or $a \leqslant x < b$



(a, b) or $]a, b[$ 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
$x \mid y$	x "divides" y i.e. $\exists\, m \in \mathbb{N}$ s.t. $y = mx$
$\{\ldots\}$	set
$\{\ldots \mid \ldots\}$ or $\ni$	... s.t. "such that" ...
!	only one; unique
$\equiv$ or IET or $\longleftrightarrow$	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 $B = \{x,y,z\}$

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

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

more Logic/Logical notation

QUALIFIERS

$\forall$	for every
$\exists$	there exists
$\neg$ or $\sim$ $\overline{A}$	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)

Common Logical Equivalences

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) $\sim (P \wedge Q) \equiv (\sim P) \vee (\sim Q)$.

.

Indicies, collections, & sets of sets

$I = A_i$ = the "index set"

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

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

two examples of an indexed collection of sets

Properties of Cartesian products

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

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

(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 × B. That means if n(A) = m and n(B) = n, then n(A × B) = mn.

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

(iv) A × A × A = {(a, b, c) : a, b, c ∈ A}. Here (a, b, c) is called an ordered triplet.

(v) The Cartesian product of sets is not commutative, i.e. A × B ≠ B × A

(vi) The Cartesian product of sets is not associative, i.e. A × (B × C) ≠ (A × B) × C

(vii) If A is a set, then A × ∅ = ∅ and ∅ × A = ∅.

(viii) If A and B are two sets, A × B = B × A if and only if A = B, or A = ∅, or B = ∅.

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

- A × (B ∩ C) = (A × B) ∩ (A × C)
- A × (B ∪ C) = (A × B) ∪ (A × C)
- (A ∩ B) × C = (A × C) ∩ (B × C)
- (A ∪ B) × C = (A × C) ∪ (B × 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 its up to the analyst to decide which method is most appropriate.

3) Extremely satisfying notation; can represent multidimensional data very intuitively with vectors, scalars, & matricies

4) A "must know" for statistics/econometrics, computer sci/programming, and quantitative finance

5) Unit length = 1 (normalization/direct comparisons)

6) MATRIX multiplication is <u>not</u> communative! But note that vectors & scalars have different properties than matricies.

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) span(A) and rk(A) are interrelated concepts that describe different aspects of the structure and <u>capabilities</u> 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 magnitude.

12) To understand orthogonality/orthonormality watch Khan Academy Intro to Orthonormal bases

The Basics: Matrix Multiplication, Addition, and Subtraction

$$\overset{n \rightarrow}{m \downarrow}\begin{bmatrix} a_{i,j} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \\ a_{m,1} & a_{m,2} & a_{m,3} & & a_{m,n} \end{bmatrix}_{m \times n} \qquad \text{Scalar} \rightarrow \alpha \quad \vec{X} = \begin{bmatrix} x_1 \\ x_2 \\ \\ x_n \end{bmatrix}_n$$

a matrix has m rows, n cols.

"single column" vector

- matrix multiplication vs. the "dot product" :

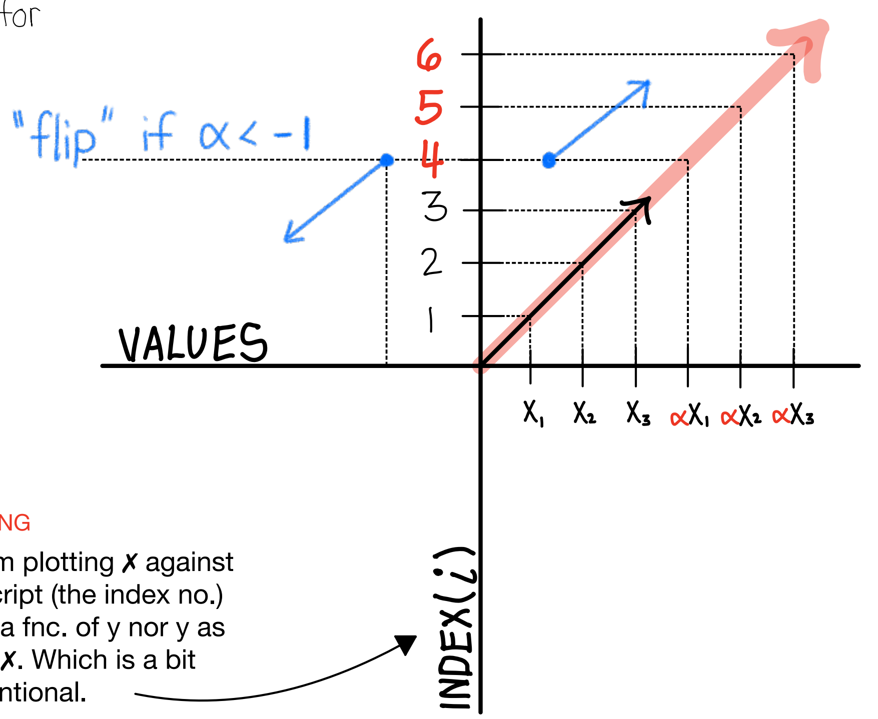
~ note that #cols (in the first matrix) must equal #rows (in the second matrix) or else two matricies cannot be multiplied together.

① A SCALAR AND A VECTOR :

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

$$\underset{3\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}} \alpha = \alpha \underset{3\times 1}{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}} = \underset{3\times 1}{\begin{bmatrix} (\alpha)(x_1) \\ (\alpha)(x_2) \\ (\alpha)(x_3) \end{bmatrix}}$$

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):

$$\vec{y}^{\,T}\vec{x} = \vec{x}^{\,T}\vec{y} = \underbrace{\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}}_{1\times 3} \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}}_{3\times 1} = \sum_{i=1}^{n=3} x_i y_i = (x_1)(y_1)+(x_2)(y_2)+(x_3)(y_3)$$

Communative property holds for dot products, too

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

③ A MATRIX AND A VECTOR:

non-communative; AB ≠ BA

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

✓

④ TWO SQUARE MATRICES:

non-communative; AB ≠ BA

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

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 $\vec{x}$:

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

Can we divide matrices?

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

- numbers: let $x = \frac{a}{b}$, then $x^{-1} = \frac{b}{a}$ $\Rightarrow$ $x\,x^{-1} = 1$
- numbers: $\frac{a}{b} = a * b^{-1} \equiv \dfrac{\frac{a}{1}}{\frac{b}{1}} = \left(\frac{a}{1}\right)\left(\frac{1}{b}\right)$
- 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**.

$$\begin{array}{cc}
 & \text{3 columns} \\
\text{2 rows} & \begin{array}{ccc} \downarrow & \downarrow & \downarrow \end{array} \\
\begin{array}{c} \rightarrow \\ \rightarrow \end{array} & \begin{bmatrix} -2 & 5 & 6 \\ 5 & 2 & 7 \end{bmatrix}
\end{array}$$

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.

$$\begin{aligned}
\mathbf{A} + \mathbf{B} &= \begin{bmatrix} 4 & 8 \\ 3 & 7 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 5 & 2 \end{bmatrix} \\
&= \begin{bmatrix} 4+1 & 8+0 \\ 3+5 & 7+2 \end{bmatrix} \\
&= \begin{bmatrix} 5 & 8 \\ 8 & 9 \end{bmatrix}
\end{aligned}$$

Basic/General Knowledge

- Main **Diagonal** of a Matrix: $\text{diag}(A) = \underset{3\times 3}{\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}} = \begin{bmatrix} a_{1,1} \\ a_{2,2} \\ a_{3,3} \end{bmatrix}$

- Trace: trace(A) = sum of all main diagonal entries
- To "diagonalize" a square matrix: $A = PDP^{-1} \iff D = P^{-1}AP$
- cardinality: the number of elements in a set
- dimensions of a matrix: 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 <u>not</u> "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 $\dim(A) = \dim(B)$ 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 <u>subset</u> of real numbers that form linear combinations among the matrix's cols (if the matrix is <u>not</u> full-rank). Logically, if 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.

$$\text{span}(V) \overset{\text{Generally}}{=\!=\!=} \text{all } v_i \in V \text{ that can be written as } \underset{\substack{\uparrow \\ \text{linear} \\ \text{scalars}}}{c_1} v_1 + c_2 v_2 + \cdots + c_n v_n$$

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

Square Matrices

- symmetric matrix: $A = A^T$
- a square "matrix of ones": self explainatory...
- 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:

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

- 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, $A_{3x3} I_3 = A_{3x3}$

- 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.)

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

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

the submatrix corresponding to a_{11} is: $\begin{bmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{bmatrix}$

the submatrix corresponding to a_{12} is: $\begin{bmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{bmatrix}$

the submatrix corresponding to a_{13} is: $\begin{bmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{bmatrix}$

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

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

  - <u>1-norm</u> $\|\vec{X}\|_1 = \left(\sum_{i=1}^{n} |x_i| 1\right)^{\frac{1}{1}}$ = absolute value of largest col. sum

  - <u>P-norm (general case)</u> $\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$

  - <u>∞-norm</u> $\|\vec{x}\|_{\inf} = \left(\sum_{i=1}^{n} |x_i|^{\inf}\right)^{\frac{1}{\inf}}$ = absolute value of largest row sum

- Determinant of a matrix*: det(A) = |A|:

  - 1×1 matrix: Let $A = [a]$, Then $|A| = a$

  - 2×2 matrix: Let $A = \underset{2\times2}{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}$, Then $|A| =$

  - 3×3 matrix: Let $A = \underset{3\times3}{\begin{array}{c} + \quad - \quad + \\ \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} \end{array}}$, $|A| = +a_{1,1} \underset{2\times2}{\begin{vmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{vmatrix}} - a_{1,2} \underset{2\times2}{\begin{vmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{vmatrix}} + a_{1,3} \underset{2\times2}{\begin{vmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{vmatrix}}$

* 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): $\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{trace\,(A^* A)}$,

  A^* is "the conjugate transpose of A" (see appendix for full detail)

- $A\vec{x} = \vec{b}$, A is called the "transition" or "coefficient" matrix
- a single col. vector $\vec{v}$ times a <u>basis</u> 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 $\vec{v}$"
  - $$B\vec{v} = \underbrace{\begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 \\ 0 & 0 & 0 & \lambda_4 & 0 \\ 0 & 0 & 0 & 0 & \lambda_5 \end{bmatrix}}_{B_{n\times n}} \underbrace{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix}}_{\vec{v}_n} = \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 + \lambda_4 v_4 + \lambda_5 v_5$$
  - B is very important; its struture determines everything
    (1) eigen-basis via finding the eigenspaces for A
    (2) orthogonal basis via Gram-Schmidt
    (3) orthonormal basis via Modified Gram-Schmidt
  - desireable 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:

i is the row index

j is the col. index

$$A\vec{x} = \vec{b}, \; a_{ij} \in A \equiv \begin{array}{c} \\ m \\ \text{rows} \end{array} \overbrace{\left[\begin{array}{ccccc|c} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & a_{1,5} & b_1 \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & a_{2,5} & b_2 \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & a_{3,5} & b_3 \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} & a_{4,5} & b_4 \\ a_{5,1} & a_{5,2} & a_{5,} & a_{5,4} & a_{5,5} & b_5 \end{array}\right]}^{A_{m\times n} \quad \vec{b}}_{n \text{ cols.}} = \overbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}}^{\vec{x}} \equiv \begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + a_{15}x_5 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + a_{25}x_5 = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + a_{35}x_5 = b_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + a_{45}x_5 = b_4 \\ a_{51}x_1 + a_{52}x_2 + a_{53}x_3 + a_{54}x_4 + a_{55}x_5 = b_5 \end{array}$$

KNOWN OBSERVATIONS

UNKNOWN (SOLVE FOR)

KNOWN

System of Linear Equations

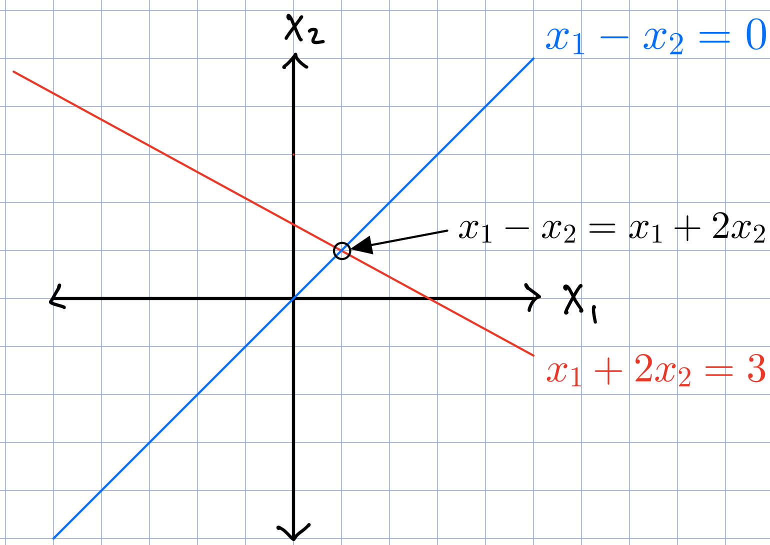
def: 2 or more linear equations

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

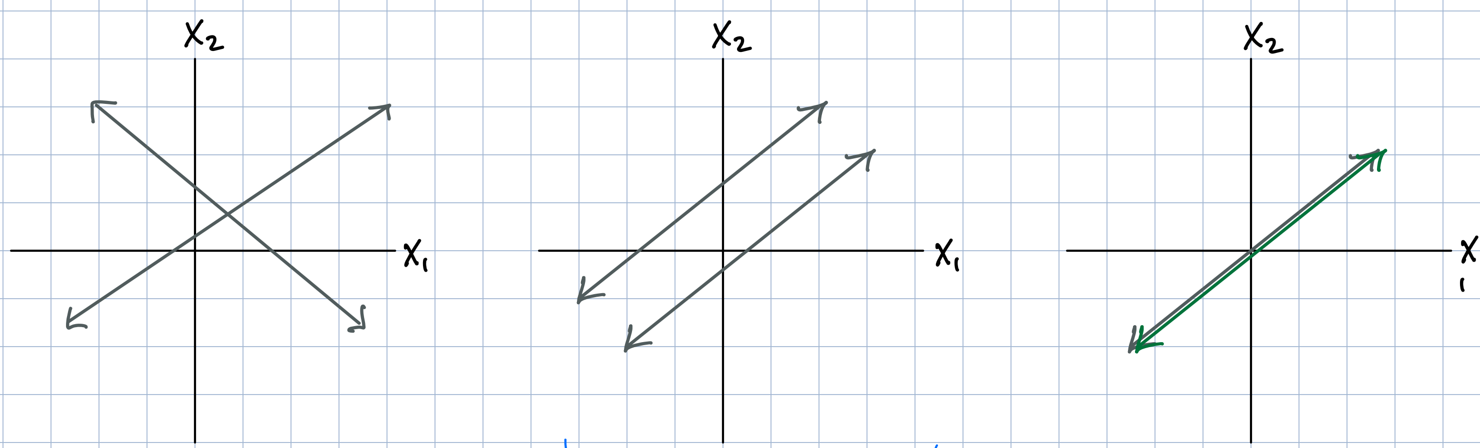
Ex:

$$x_1 + 2x_2 = 3 \Rightarrow x_1 = 3 - 2x_2$$
$$x_1 - x_2 = 0 \Rightarrow x_1 = x_2$$

The sol. set is visualized by the intersection:



one (unique) sol. | no sol. | many sol.



INCONSISTENT

CONSISTENT

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. $R_1 \longleftrightarrow R_2$
(2) Scaling e.g. $3R_1 \longrightarrow R_1$
(3) Replacement e.g. $3R_1 + R_2 \longrightarrow R_2$

* 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 invirtible $\longrightarrow$ proceed w/ Gaussian Elimination, LU Factorization, Cholesky Factorization, or QR Factorization (more on these later...)
b) if A is not invirtible, then no unique sol. exists $\Rightarrow$ 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 ref(A) or rref(A)

def: a **pivot column** is the corresponding column

* cannot see the pivots of matrix A unless converted to ref(A) <u>or</u> rref(A) *

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

0's 0's 0's

<u>Steps to covert A into rref(A) — ALGORITHM</u>

* the whole point of converting A to rref(A) is to reveal the pivots *

Forward Phase

- **step 1** Begin w/ the left-most, non-zero col
- **step 2** Make the first entry a non-zero "pivot"; interchange rows if needed
- **step 3** Use elementary row operations to set the pivot = 1 and to put all 0's below the pivot

Backward Phase

- **step 4** Repeat steps 1-3 for the submatrix
- **step 5** Beginning w/ the right-most pivot, work up & left to make all 0's above and below it

<u>Parametric Description / Parametric Form</u>

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

ex) $\text{rref}(A) = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & \end{matrix} \\ \left[\begin{matrix} 1 & -1 & 0 & 2 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & -1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}\right] \end{array}$

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 $a_{4,7} \neq 0$

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

consistent estimators?

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

Thm: a linear sys. is consistent iff the right-most col of augmented matrix is <u>not</u> a piviot 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

$$\text{rref}(A) = \left[\begin{array}{ccc|c} 1 & 3 & 0 & -5 \\ 0 & 0 & 1 & 3 \end{array}\right] \Rightarrow \begin{array}{l} x_1 + 3x_2 = -5 \Rightarrow x_1 = -5 - 3x_2 \\ x_3 = 3 \end{array}$$

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 ≡ refers to a matrix's column(s)

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

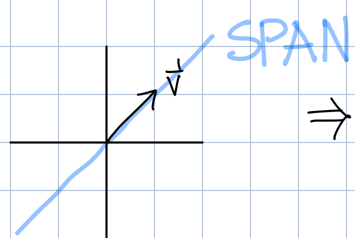
- use vectors to form systems of linear equations (i.e., a "linear combination")
- use matrices to represent systems of linear equations concisely
- solve systems of linear equations using Gaussian Ellim. or factorization methods

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

where $A = \left[\vec{v}_1 \middle| \vec{v}_2 \middle| \vec{v}_3 \middle| \vec{b}\right]$

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

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



$\Rightarrow$ $\text{span}\{\vec{v}\} = \{c_1\vec{v}\} \neq \mathbb{R}^2$

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
$c_1 = c_1 = \dots = c_k = 0$

def: a set of vectors $\{v_1, v_2, \dots, v_3\}$ are **linearly independent** if the vector equation $c_1v_1 + c_2v_2 + \dots + c_kv_k$ has only the trival 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 $\left[\begin{array}{cc|c} 2 & -4 & 0 \\ -3 & 6 & 0 \end{array}\right] \xrightarrow{\text{REF}} \left[\begin{array}{cc|c} 6 & -12 & 0 \\ 0 & 0 & 0 \end{array}\right]$ * c_2 is a free variable $\Rightarrow$ inf. many sol. $\Rightarrow$
$\Rightarrow$ 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 <u>not</u> 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} ITEATION
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} ITEATION
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} ITEATION
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} ITEATION
0	0	0	0	a_{55}	a_{56}	b_5	c_5 is determined by c_6 and row 5; c_4, c_3, c_2, c_1 are free	2^{nd} ITEATION
0	0	0	0	0	a_{66}	b_6	c_6 is determined by row 6 equality; c_2, c_3, c_4, c_5, c_6 are free	1^{st} ITEATION

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 <u>SINCE EVERY ROW IN A HAS A NONZERO PIVOT</u>, WOULD RESULT IN ONLY THE TRIVIAL SOL. $\Longrightarrow$ 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 $a_1x_1+a_2x_2+\ldots+a_nx_n=0$ 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:

$$\left[\begin{array}{ccccccc:c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & b_1 \\ 0 & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & b_2 \\ 0 & 0 & 0 & a_{34} & a_{35} & a_{36} & a_{37} & b_3 \\ 0 & 0 & 0 & 0 & a_{45} & a_{46} & a_{47} & b_4 \\ 0 & 0 & 0 & 0 & 0 & a_{56} & a_{57} & b_5 \\ 0 & 0 & 0 & 0 & 0 & 0 & a_{77} & b_6 \end{array}\right]$$

Row 1: c_1 now depends on c_2 & c_2 i.e., $c_1=f(c_2,c_3)$

Row 2: c_2 now depends on c_3 i.e., $c_2=f(c_3)\equiv c_3=f(c_2)$

$\Rightarrow$ inf. many sol ☹

Row 3: ✓ EREYTHING OK SO FAR... we have unique solutions for c_7, c_6, c_5, c_4

Rows 4–6: ✓ ✓ ✓

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 indedendence when dealing w/ a homogenous sys?

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

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

A:

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

(2) we solve $A\vec{c}=\vec{b}$ for $\vec{c}$, which are our regression coeficients 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

$$\vec{C} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix}, \quad \vec{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \quad A = \begin{array}{cccc} \vec{x}_1 & \vec{x}_2 & & \vec{x}_n \end{array} \begin{bmatrix} x_{11} & x_2 & \cdots & x_{mn} \\ x_{11} & x_2 & \cdots & x_{mn} \\ \vdots & \vdots & \ddots & \\ x_{1n} & x_{2n} & & x_{mn} \end{bmatrix} \longrightarrow \text{solve } A\vec{\beta} = \vec{C}$$

same as: $\hat{y} \overset{\text{zero-intercept MLR}}{=\!=} \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n + \varepsilon$

Best fit is qualified by the sum of squared errors:

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

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 you decision variable. To save space on the page I've decided to do it this way instead!

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

IMPLICIT (not explicit) CONSTRAINT

s.t. $\hat{y}_i = \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_n x_{in} \quad \forall \; x_{\underset{\text{row index}}{i}\overset{\text{col. index}}{j}} \in A$

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

which occurs where

LOCAL MINIMUM CONDITIONS

$\frac{\partial \hat{y}}{\partial \varepsilon} = 0$ and $\frac{\partial^2 \hat{y}}{\partial \varepsilon^2} > 0$

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

Eigenvalues and Eigenvectors:

def: **Eigenvalues** are simply (and most generally) a linear transformation – let "transformation" be denoted by " $\longrightarrow$ " or $T(\bullet)$ – of a vector

$T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is controlled by the "eigenvalue" λ: eigenvalues quantify the scaling factor applied by the transformation

$T(\vec{v}) = \lambda \vec{v}, \ \{v_1, v_2, \ldots, v_n\} \in \mathbb{R}^n$ vectors $\vec{v}$ which satisfy this equality are called **Eigenvectors**

exs: FINDING EIGENVALUES FINDING EIGENVECTORS

- $T(\vec{x})$ represented as a matrix-value product: $T(\vec{x}) = A_{n\times n}\vec{x}_n$, A is called the "transformation matrix"

$$T(\vec{x}) = A\vec{x}, \ A = \left[T\begin{pmatrix}1\\0\end{pmatrix} \ T\begin{pmatrix}0\\1\end{pmatrix} \right] \Rightarrow T(\vec{x}) = A\vec{x} = \begin{bmatrix}.4 & 0\\0 & 2.5\end{bmatrix}\begin{bmatrix}5\\2\end{bmatrix} = \begin{bmatrix}(.4)(5) + 0\\0 + (2.5)(2)\end{bmatrix} = \begin{bmatrix}2\\5\end{bmatrix} \checkmark$$

$I_{2\times2}$

- $A\vec{x} = \lambda\vec{x}$ iff $\boxed{\det(\lambda I_n - A) = 0} \equiv \lambda$ is an eigenvalue of A iff $\boxed{\det(\lambda I_n - A) = 0}$

exact same condition

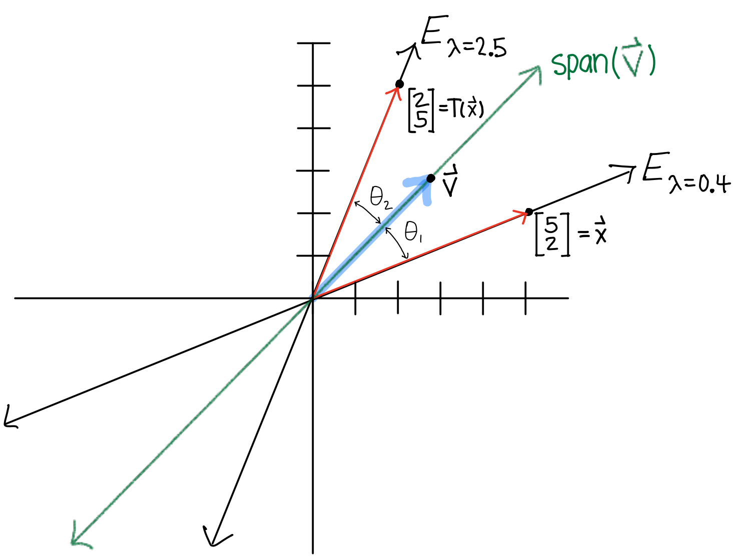
- $A = \begin{bmatrix}.4 & 0\\0 & 2.5\end{bmatrix}$ so we want to show that $\det\left(\lambda\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} - \begin{bmatrix}.4 & 0\\0 & 2.5\end{bmatrix}\right) = 0$ The "characteristic polynomial"

- Solving for λ: $\det\left(\lambda\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} - \begin{bmatrix}.4 & 0\\0 & 2.5\end{bmatrix}\right) = \det\left(\begin{bmatrix}\lambda - .4 & 0\\0 & \lambda - 2.5\end{bmatrix}\right) = \overbrace{(\lambda - .4)(\lambda - 2.5) - 0}^{\lambda^2 - 5.5\lambda + 6} \Rightarrow \begin{matrix}\lambda = 2.5\\ \lambda = .4\end{matrix}$ are the two eigenvalues of A

- solving for the **eigenspaces** (i.e., set of all $\vec{x}$'s that satisfy $A\vec{v} = \lambda\vec{x}$)

eigenvalue for $\lambda = 2.5$: $A\vec{x} = \lambda\vec{x} \Rightarrow \underset{A\vec{x}}{\begin{bmatrix}2\\5\end{bmatrix}} = \underset{\lambda}{2.5}\underset{\vec{x}}{\begin{bmatrix}5\\2\end{bmatrix}} = \begin{bmatrix}12.5\\5\end{bmatrix} \Rightarrow E_{\lambda=2.5} = \text{span}\left(\begin{bmatrix}12.5\\5\end{bmatrix}\right) = \text{span}(T(\vec{x}))$

eigenvalue for $\lambda = .4$: $A\vec{x} = \lambda\vec{x} \Rightarrow \underset{A\vec{x}}{\begin{bmatrix}2\\5\end{bmatrix}} = \underset{\lambda}{.4}\underset{\vec{x}}{\begin{bmatrix}5\\2\end{bmatrix}} = \begin{bmatrix}2\\.8\end{bmatrix} \Rightarrow E_{\lambda=0.4} = \text{span}\left(\begin{bmatrix}2\\.8\end{bmatrix}\right) = \text{span}(\vec{x})$



- $T(\vec{x})$ "refects" $\vec{x}$ about span($\vec{V}$); $\theta_1 = \theta_2$
- eigen "vectors" = $E_{\lambda=2.5}$ and $E_{\lambda=0.4}$
- eigen "values" = $\lambda = 2.5$ and $\lambda = 0.4$
- eigen "space" = $\text{span}\left\{\begin{bmatrix}12.5\\5\end{bmatrix}, \begin{bmatrix}2\\.8\end{bmatrix}\right\}$
- eigen "basis" = $\left\{\begin{bmatrix}12.5\\5\end{bmatrix}, \begin{bmatrix}2\\.8\end{bmatrix}\right\}$