Lecture 10: Orthogonality

Thursday, September 24, 2015 9:30 AM

Admin: Homework 3 due, 4 out tonight.

Outline:

Lengths

Inner products & Angles

Orthogonality

More inner products

TODAY:

Angles between subspaces: Orthogonal spaces

Rank-Nullity Theorem

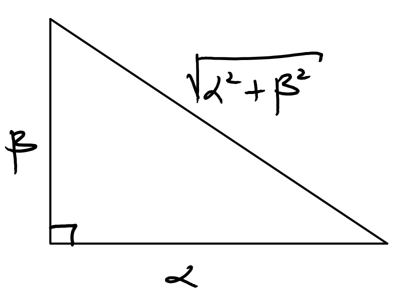
NEXT:

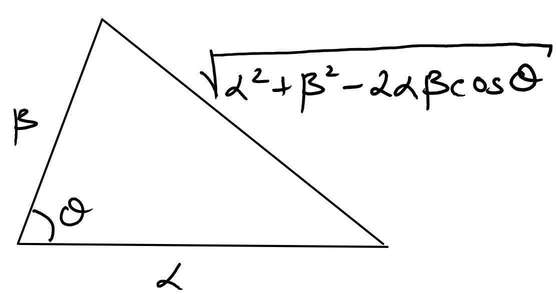
Mutually orthogonal sets of vectors & Bases

Projections

Gram-Schmidt orthogonalization

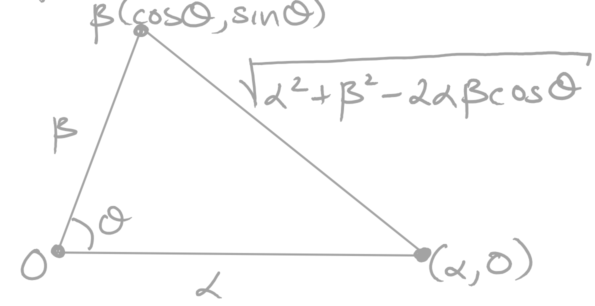
Recall: PYTHAGOREAN THEOREM & LAW OF COSINES





We'll use these to find expressions for lengths and angles in higher dimensions.

Proof: Write it out in coordinates:

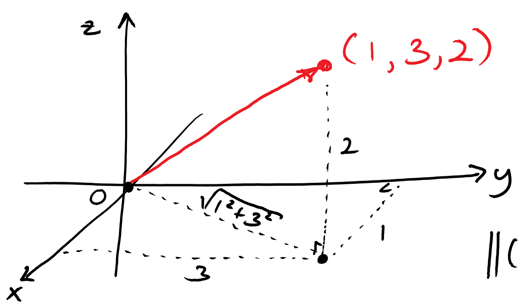


LENGTHS

Definition: For a vector $v \in \mathbb{R}^n$ or $v \in \mathbb{C}^n$, the **Euclidean length** of v is given by

$$\|v\| = \sqrt{|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2}$$

Example:



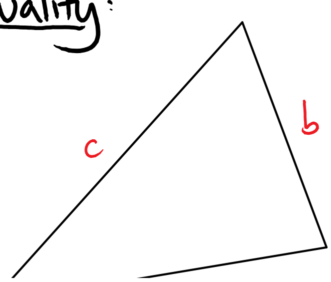
$$\|(1,3,2)\| = \sqrt{1^2+3^2+2^2} = \sqrt{14}$$

Observe:
- For $v \in \mathbb{R}^n$, $\|v\|^2 = v^T v = \begin{pmatrix} \text{——} v^T \text{——} \end{pmatrix} \begin{pmatrix} | \\ v \\ | \end{pmatrix}$
- Scaling: $\|\alpha \cdot \vec{v}\| = |\alpha| \cdot \|\vec{v}\|$

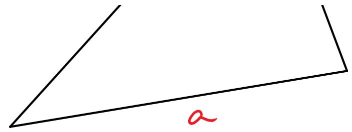
  $\Rightarrow \dfrac{v}{\|v\|}$ is a **unit vector**
- For $v \in \mathbb{C}^n$, $\|v\|^2 = \bar{v}^T v$,

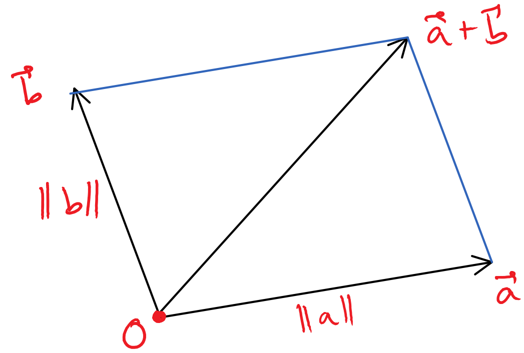
  e.g. $\|(1,i)\|^2 = 1^2 + |i|^2 = 2$

Triangle inequality:



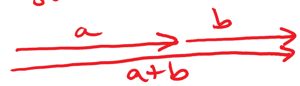
$a + b \geq c$
$a + c \geq b$
$b + c \geq a$





$$\boxed{\|\vec{a}+\vec{b}\| \leq \|\vec{a}\| + \|\vec{b}\|}$$

(with equality if and only if a and b point in the same direction along the same line



)

Inner Products & Angles

Definition: The Euclidean **inner product** between two vectors $\vec{u},\vec{v} \in \mathbb{R}^n$ is given by

$$\boxed{\vec{u}\cdot\vec{v} = u^T v = \sum_{i=1}^{n} u_i v_i} \in \mathbb{R}$$

For two complex vectors $\vec{u},\vec{v} \in \mathbb{C}^n$,

$$\vec{u}\cdot\vec{v} = \sum_i \bar{u}_i v_i \in \mathbb{C}.$$

(Other notation: dot product, scalar product, $\langle u|v\rangle$, (u,v).)

Observe:

* $\vec{u}\cdot\vec{v} = (\vec{v}\cdot\vec{u})^*$ ← complex conjugate
* $\vec{u}\cdot\vec{u} = \|\vec{u}\|^2$
* It is "bilinear":

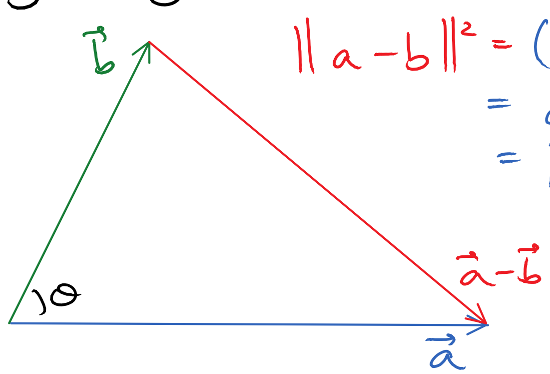
$\vec{u}\cdot(\vec{v}+\vec{w}) = \vec{u}\cdot\vec{v} + \vec{u}\cdot\vec{w}$ $\qquad$ $(\vec{u}+\vec{v})\cdot\vec{w} = \vec{u}\cdot\vec{w} + \vec{v}\cdot\vec{w}$

$\vec{u}\cdot(\alpha\vec{v}) = \alpha(\vec{u}\cdot\vec{v})$ $\qquad$ $(\alpha\vec{u})\cdot\vec{v} = \alpha^*(\vec{u}\cdot\vec{v})$

Definition: The **angle** between two nonzero vectors a and b (in $\mathbb{R}^n$ ~~or $\mathbb{C}^n$~~) is that $\Theta \in [0,\pi]$ satisfying

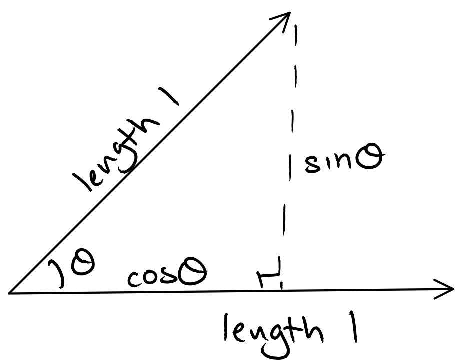
$$\boxed{\cos\Theta = \frac{a}{\|a\|}\cdot\frac{b}{\|b\|}}$$

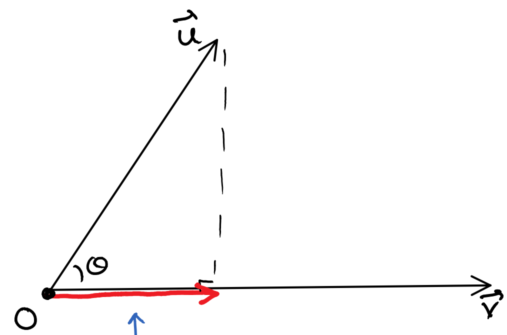
Observe: This definition agrees with what we know from geometry:



$$\|a-b\|^2 = (a-b)\cdot(a-b)$$
$$= a\cdot a + b\cdot b - 2a\cdot b$$
$$= \|a\|^2 + \|b\|^2 - 2\|a\|\cdot\|b\|\cdot\cos\theta$$

Observe: The magnitude $|\vec{u}\cdot\vec{v}|$ measures the "overlap" between $\vec{u}$ and $\vec{v}$:





projection of $\vec{u}$ in the direction of $\vec{v}$

$$= \frac{\vec{v}}{\|\vec{v}\|}\cdot\|\vec{u}\|\cos\theta$$
$$= \boxed{\frac{(\vec{u}\cdot\vec{v})}{\|\vec{v}\|^2}\,\vec{v}}$$

– the projection is 0 if $\vec{u}\cdot\vec{v}=0$, or $\theta=\frac{\pi}{2}$

Definition: $\vec{u}$ and $\vec{v}$ are **orthogonal** (denoted $\vec{u}\perp\vec{v}$) if

$$\vec{u}\cdot\vec{v}=0 \qquad (\Leftrightarrow 90^\circ \text{ angle, or } \tfrac{\pi}{2}\text{, between them})$$

(**Note**: Unlike the definition of angle, orthogonality makes sense even in vector spaces over finite fields — not just $\mathbb{R}$ or $\mathbb{C}$.)

Theorem (Cauchy-Schwarz inequality):

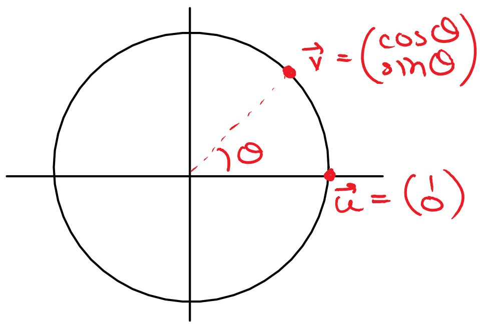
$$|\vec{u}\cdot\vec{v}| \le \|\vec{u}\|\cdot\|\vec{v}\|$$

Example:

$(\cos\theta)$

$\vec{u}\cdot\vec{v} = \|\vec{u}\|\cdot\|\vec{v}\|$ if they point

Example:



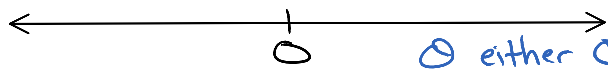
$\vec{u} \cdot \vec{v} = \|\vec{u}\| \cdot \|\vec{v}\|$ if they point in the same direction

$\vec{u} \cdot \vec{v} = -\|\vec{u}\| \cdot \|\vec{v}\|$ if they point in opposite directions

otherwise, $|\vec{u} \cdot \vec{v}| < \|\vec{u}\| \cdot \|\vec{v}\|$ strictly

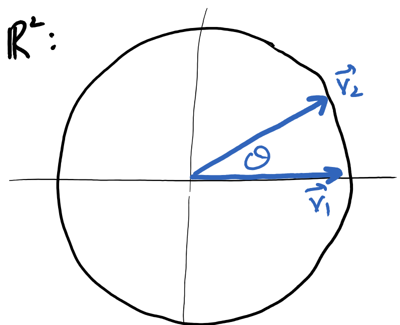
High-dimensional intuition:
What is the angle between two random unit vectors in $\mathbb{R}^n$?

$\mathbb{R}^1$:



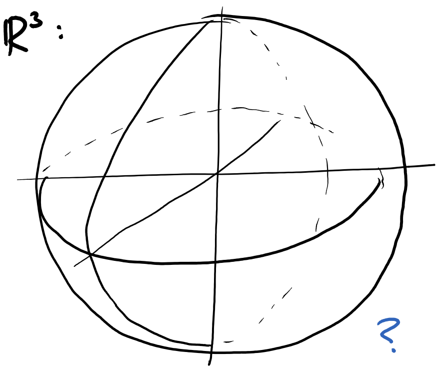
θ either 0 or π

$\mathbb{R}^2$:



all angles $\theta \in [0, \pi]$ are equally likely

$\mathbb{R}^3$:



How to generate random directions in $\mathbb{R}^n$?

"direction" = unit vector $\|\vec{v}\| = 1$

= point on the unit sphere

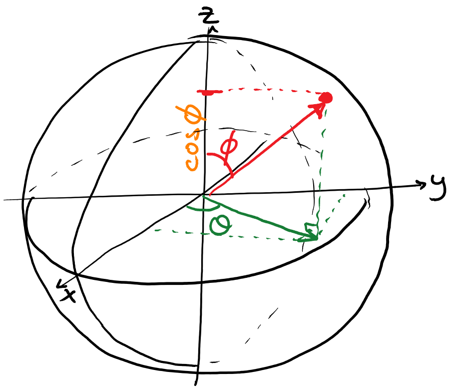
$\mathbb{R}^2$: $\theta = 2\pi \cdot \text{rand}(1)$

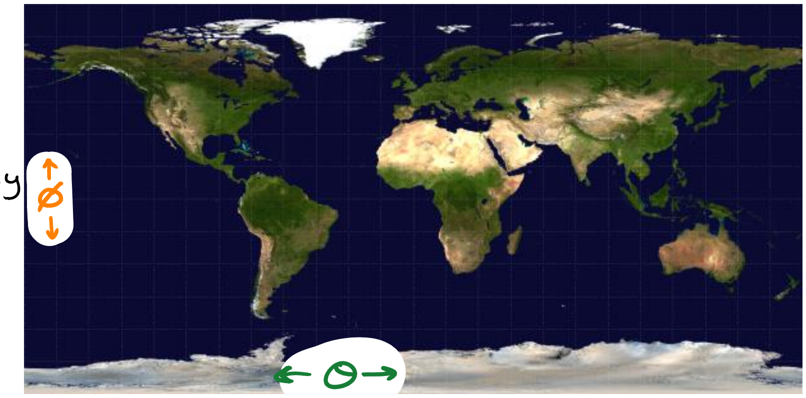
$\vec{v} = (\cos\theta, \sin\theta)$ ✓

```
>> help rand
 rand Uniformly distributed pseudorandom numbers.
    R = rand(N) returns an N-by-N matrix containing pseudorandom values drawn
    from the standard uniform distribution on the open interval(0,1).  rand(M,N)
    or rand([M,N]) returns an M-by-N matrix.  rand(M,N,P,...) or
    rand([M,N,P,...]) returns an M-by-N-by-P-by-... array.  rand returns a
    scalar.  rand(SIZE(A)) returns an array the same size as A.
```

$\mathbb{R}^3$: ① Use longitude θ and latitude ϕ?

$\vec{v} = (\cos\theta \cdot \sin\phi, \sin\theta \cdot \sin\phi, \cos\phi)$





No! This gives a non-uniform distribution (points near the poles occur too often)

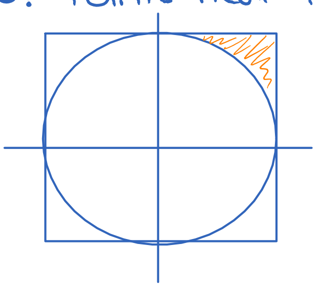
② Pick 3 coordinates at random, and renormalize?

```
>> v = 2 * rand(3, 1) - 1

v =

    0.8268
    0.2647
   -0.8049

>> v = v / norm(v)

v =

    0.6984
    0.2236
   -0.6799
```

← each coordinate is drawn uniformly from $[-1, 1]$.

← renormalize so $\|\vec{v}\| = 1$

No! Points near the corners occur too often



Better: Use the normal distribution

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}$$

$$\Rightarrow p(x)p(y) = \frac{1}{2\pi} e^{-(x^2+y^2)}$$

$= \|(x,y)\|^2$ independent of the angle $\tan^{-1}\frac{y}{x}$

```
>> help randn
 randn Normally distributed pseudorandom numbers.
    R = randn(N) returns an N-by-N matrix containing pseudorandom values drawn
    from the standard normal distribution.  randn(M,N) or randn([M,N]) returns
    an M-by-N matrix. randn(M,N,P,...) or randn([M,N,P,...]) returns an
    M-by-N-by-P-by-... array. randn returns a scalar.  randn(SIZE(A)) returns
    an array the same size as A.
```

```
>> n = 5; v = randn(n, 1); v = v / norm(v)

v =

   -0.0912
    0.0721
    0.7527
    0.5825
   -0.2839
```

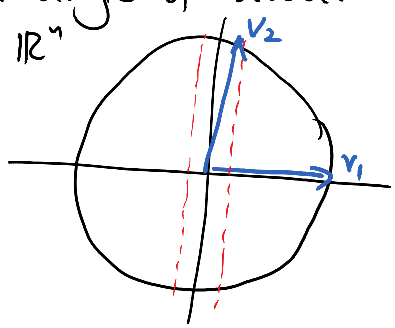
Intuition: What is the angle between two random vectors?

in $\mathbb{R}^n$?

```
trials = 100;
for pow = 1:15
  n = 2^pow;
  minangle = 180;
  for i = 1:trials
    x = randn(n,1);   } two random
    y = randn(n,1);   } n-dim vectors
    angle = 180/pi * acos(x'*y / sqrt((x'*x) * (y'*y)));
    minangle = min(angle, minangle);
  endfor
  disp([n, minangle]);
endfor
```

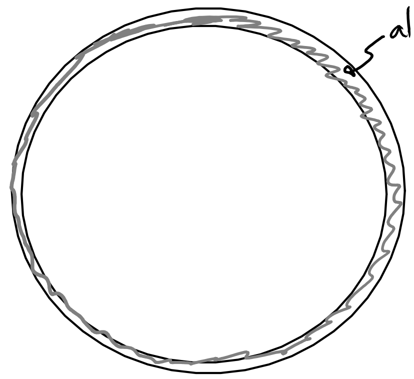
n	minangle
2.00000	0.87482
4.0000	16.8946
8.0000	36.7159
16.000	54.869
32.000	57.955
64.000	75.011
128.000	78.256
256.000	79.392
512.000	83.195
1024.000	84.732
2048.000	85.614
4096.000	87.519
8192.000	87.944
1.6384e+04	8.8889e+01
3.2768e+04	8.9228e+01
6.5536e+04	8.9405e+01
1.3107e+05	8.9468e+01
2.6214e+05	8.9661e+01
5.2429e+05	8.9794e+01
1.0486e+06	8.9824e+01

$\Rightarrow$ In high dimensions, random vectors are almost always at an angle of about 90 degrees.



Why?

① Almost all the volume of the unit ball is near the surface.



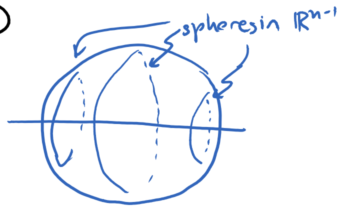
Why?

$$V(r,n) = v_n \cdot r^n$$

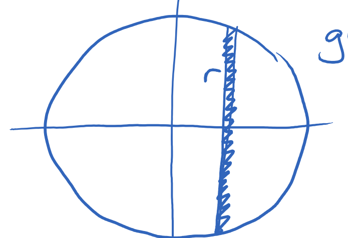
$\Rightarrow$ fraction of unit ball with length $\leq r$

$$= \frac{V(r,n)}{V(1,n)} = \frac{v_n \cdot r^n}{v_n} = r^n$$

(2)

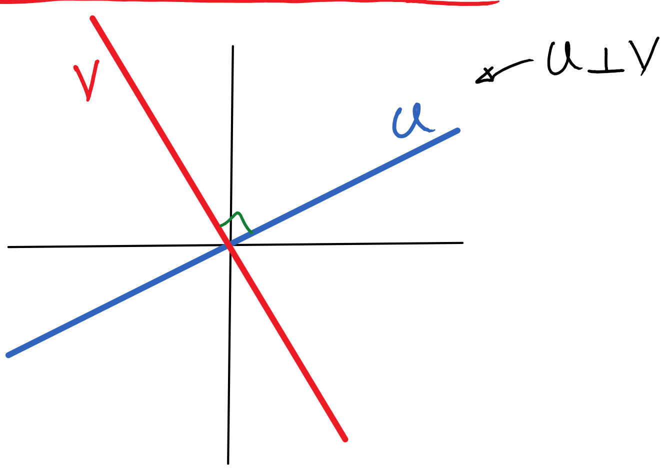


the total surface area here goes like r^{n-2}



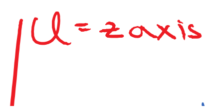
$\Rightarrow$ since r^{n-2} drops so quickly as r drops from 1 to 0, almost all the surface area is concentrated in the middle (along the "equator")

Orthogonal Subspaces

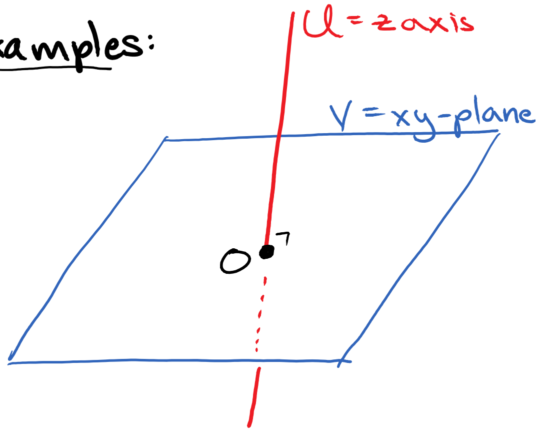


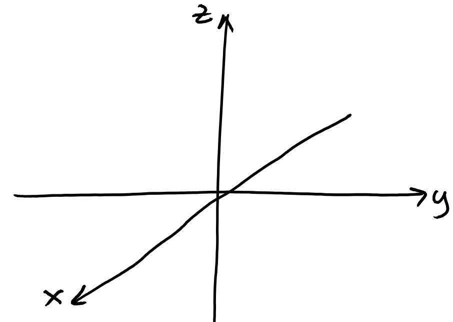
Definition: Two subspaces $\mathcal{U}$ and V are orthogonal (denoted $\mathcal{U} \perp V$, "$\mathcal{U}$ perp V") if every nonzero vector in $\mathcal{U}$ is orthogonal to every nonzero vector in V.

Examples:

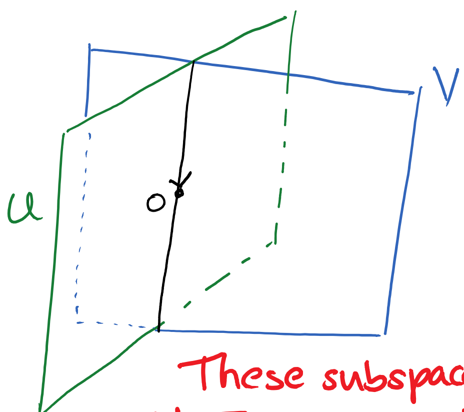


Examples:





coordinate axes are pairwise orthogonal



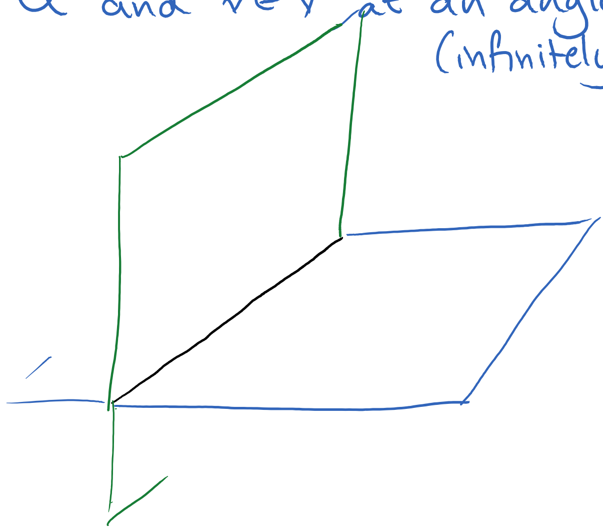
These subspaces are NOT orthogonal!!!

— because they intersect nontrivially

for $u \in \mathcal{U} \cap V$, u makes angle 0 with itself

In fact, for any angle θ, there are vectors $u \in \mathcal{U}$ and $v \in V$ at an angle θ to each other.

(infinitely many angles!)



Key example: Rank-nullity II.

Theorem: For any linear transformation A,

- $R(A) \perp N(A^T)$
  columnspace, left nullspace
- $R(A^T) \perp N(A)$
  rowspace, nullspace

Proof: In the standard bases,

$$A = \begin{pmatrix} \underline{\text{row 1 of } A} \\ \vdots \\ \underline{\text{row } m \text{ of } A} \end{pmatrix}$$

For $x \in \mathbb{R}^n$, $Ax = \begin{pmatrix} (\text{row 1}) \cdot x \\ \vdots \\ (\text{row } m) \cdot x \end{pmatrix} \in \mathbb{R}^m$.

Let $x \in N(A)$, $y \in R(A^T)$

$y = \sum_{i=1}^{m} \alpha_i r_i$

$Ax = \begin{pmatrix} r_1 \cdot x \\ \vdots \\ r_m \cdot x \end{pmatrix} = 0$

$\Rightarrow y \cdot x = \sum_i \alpha_i (r_i \cdot x) = 0$

$\Rightarrow N(A) \perp R(A^T)$. ✓

To prove $N(A^T) \perp R(A)$, repeat the argument with $A \leftrightarrow A^T$.

$\square$

Example:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 0 & 4 \\ 2 & -1 & 1 \end{pmatrix} \xrightarrow{-4,\ -2} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -8 & -8 \\ 0 & -5 & -5 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$R(A) = \text{Span}\{(1,4,2),(2,0,-1)\} \perp N(A^T) = \text{Span}((4,-5,8))$

$R(A^T) = \text{Span}\{(1,2,3),(0,1,1)\} \perp N(A) = \text{Span}((1,1,-1))$

$\text{rank}(A) = \dim R(A) = \dim R(A^T) = 2$

$\dim N(A) + \dim R(A^T) = 3$

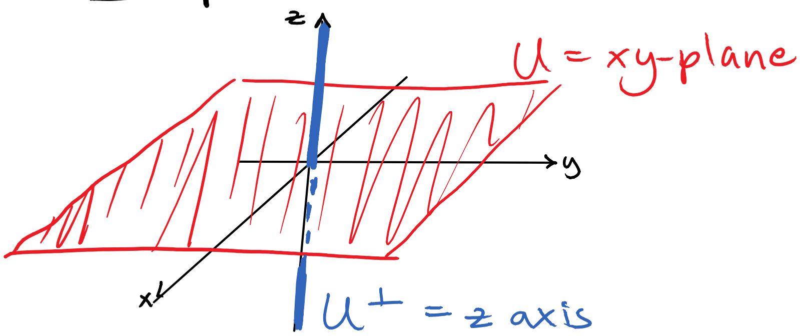
$\dim N(A^T) + \dim R(A) = 3$

Definition: For a subspace $\mathcal{U}$,

$\mathcal{U}^\perp = \{\text{all vectors orthogonal to } \mathcal{U}\}$

"orthogonal complement of $\mathcal{U}$", "$\mathcal{U}$ perp"

Example:



- $\mathcal{U} \cap \mathcal{U}^\perp = \{\vec{0}\}$
- $(\mathcal{U}^\perp)^\perp = \mathcal{U}$
- x and z axes are orthogonal, but **not** complements

How to find $\mathcal{U}^\perp$:

Start with a basis for $\mathcal{U}$.

$\Rightarrow \mathcal{U}^\perp = \{$all vectors $\perp$ to $\mathcal{U}\}$

$= \{$all vectors $\perp$ to all basis elements$\}$

(ie. $b_i \cdot v = 0$ for each basis elt. b_i)

Let $A = \begin{pmatrix} \text{———} \\ \text{———} \\ \text{rows a basis for } \mathcal{U} \\ \text{———} \end{pmatrix}$ $\Rightarrow$ $R(A^T) = \mathcal{U}$, $N(A) = \mathcal{U}^\perp$

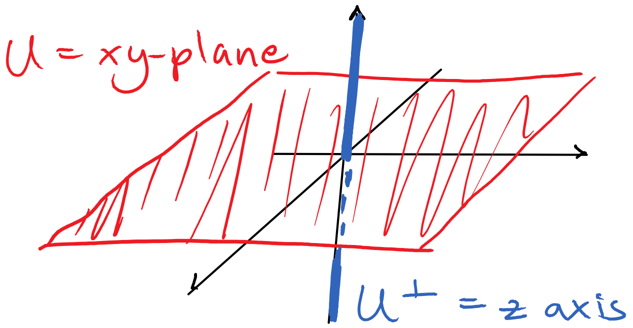
Corollary 1: In a space of dimension n,

$$\dim(\mathcal{U}) + \dim(\mathcal{U}^\perp) = n$$

Proof: Rank-nullity Thm. $\Rightarrow \dim R(A^T) + \dim N(A) = n$ $\square$

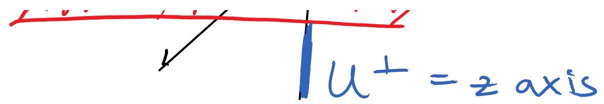
Corollary 1.5: Span$(\mathcal{U} \cup \mathcal{U}^\perp)$ = everything.

$\Rightarrow$ any vector v can be broken into $\mathcal{U}$ and $\mathcal{U}^\perp$ parts



eg.,

$$(\alpha, \beta, \gamma) = \underbrace{(\alpha, \beta, 0)}_{\mathcal{U}} + \underbrace{(0, 0, \gamma)}_{\mathcal{U}^\perp}$$



Proof: Combine a basis for $\mathcal{U}$ ($\dim \mathcal{U}$ elements) with a basis for $\mathcal{U}^\perp$ ($n - \dim \mathcal{U}$ elements). The combined set is still linearly independent. But n linearly independent vectors in a space of dimension n must be a basis, spanning everything.

Why are the bases together linearly indep.?
No nontrivial combinations of the basis vectors for $\mathcal{U}$ alone, or for $\mathcal{U}^\perp$ alone, can equal 0.
If a nontrivial combination of all the basis vectors

$$\begin{pmatrix}\text{comb. of basis}\\ \text{for } \mathcal{U}\end{pmatrix} + \begin{pmatrix}\text{comb. of}\\ \text{basis for } \mathcal{U}^\perp\end{pmatrix} = 0,$$

$$\Rightarrow \underbrace{\begin{pmatrix}\text{comb. of basis}\\ \text{for } \mathcal{U}\end{pmatrix}}_{\in \mathcal{U}} = \underbrace{-\begin{pmatrix}\text{comb. of}\\ \text{basis for } \mathcal{U}^\perp\end{pmatrix}}_{\in \mathcal{U}^\perp}$$

$\Rightarrow$ contradicts $\mathcal{U} \cap \mathcal{U}^\perp = \{\vec{0}\}$. □

Exercise: Prove that this decomposition is unique.

Rank-Nullity III:

Corollary 2: For any $m \times n$ matrix A,

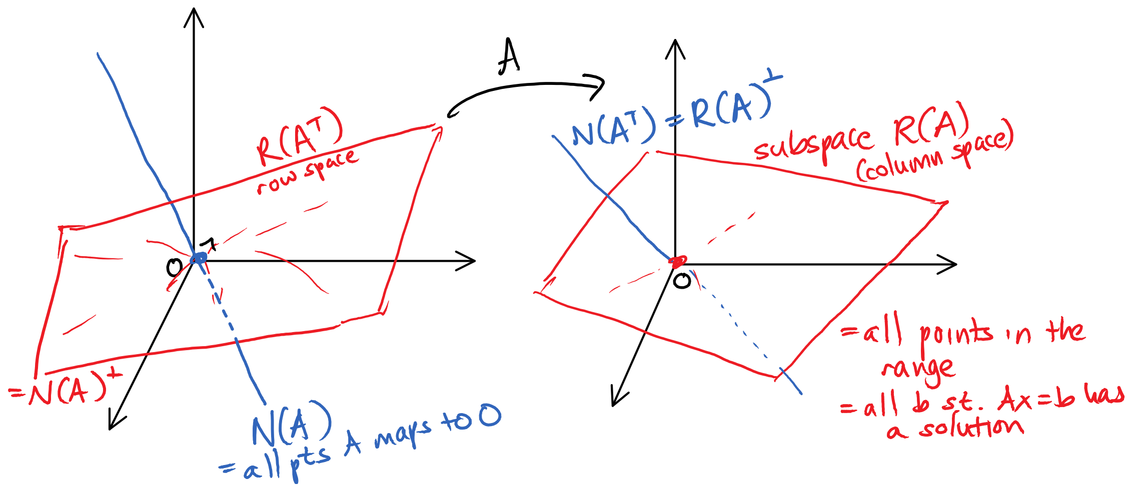
$$\boxed{\begin{aligned} N(A) &= R(A^T)^\perp \\ N(A^T) &= R(A)^\perp \end{aligned}}$$

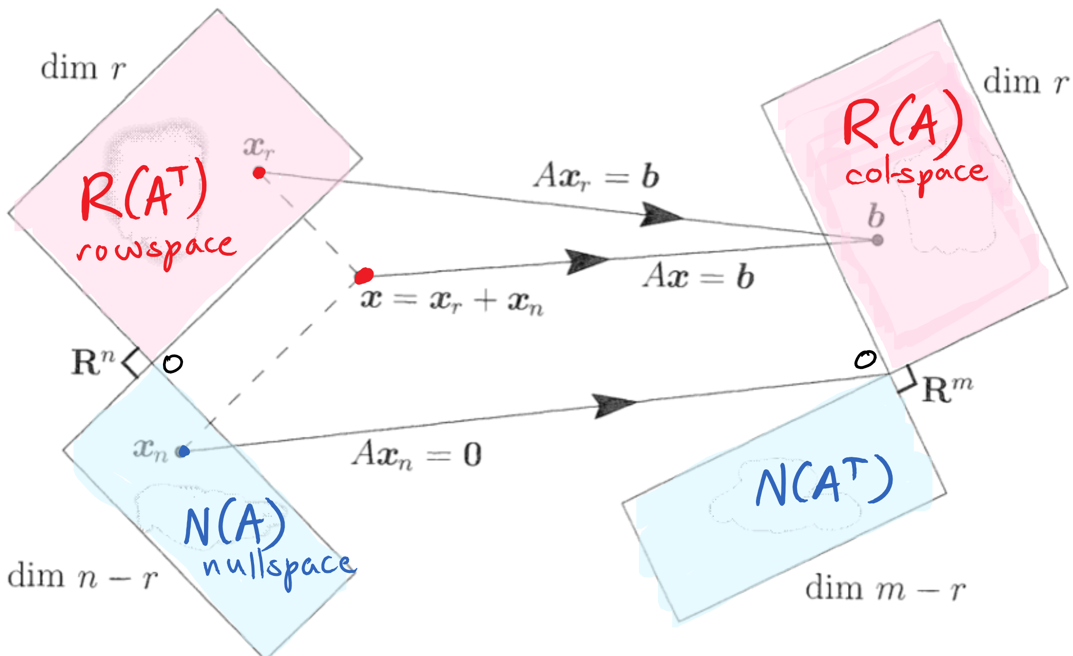
$\dim N(A) = n - \text{rank}(A)$
$\dim N(A^T) = m - \text{rank}(A)$.

Proof: Let $\mathcal{U} = R(A^T)$.

$$A = \begin{pmatrix} \text{———} \\ \text{———} \\ \text{rows span } \mathcal{U} \\ \text{———} \end{pmatrix} \Rightarrow N(A) = \mathcal{U}^\perp$$

□





$$N(A) = R(A^T)^\perp \qquad\qquad N(A^T) = R(A)^\perp$$

Todo: Understand better how $R(A^T)$ is mapped onto $R(A)$. [Singular-value decomposition]

Example:

$$A = \begin{pmatrix} -2 & 1 & & & & 1 \\ 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & 1 & \ddots & & \\ & & & & -2 & 1 \\ 1 & & & & 1 & -2 \end{pmatrix}$$

($n \times n$)

discretized 2nd derivative on the circle

Recall: $N(A) = \operatorname{Span}((1,1,\dots,1))$

$$A = A^T \Rightarrow N(A^T) = N(A)$$

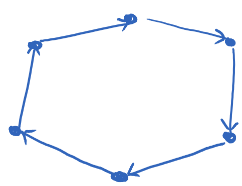
$$\Rightarrow R(A) = N(A^T)^\perp \qquad n-1 \text{ dimensional}$$
$$= \{\text{all } \vec{b} \in \mathbb{R}^n \mid \vec{b} \cdot (1,1,\dots,1) = 0\}$$
$$= \{\text{all } \vec{b} \in \mathbb{R}^n \mid b_1 + b_2 + \cdots + b_n = 0\}$$

Hence, $\boxed{\begin{array}{c} Ax = b \text{ has a solution} \\ \Updownarrow \\ b_1 + \cdots + b_n = 0 \end{array}}$

This is <u>much</u> easier than using Gaussian elimination to find a basis for $R(A)$. (Essentially because $N(A^T)$ is low-dimensional.)

This example is actually very important; similar matrices arise frequently in applications.

Note that $A = -E_G E_G^T$, where G is the cycle graph



and E_G is the incidence matrix (from last time).

We proved $N(E_G^T) = \text{Span}((1,1,\dots,1))$, which gives an alternate proof that $N(A) = \text{Span}((1,\dots,1))$.

Inner Products for More General Vector Spaces

I. Matrices

Example: For 2×2 real matrices

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

let their inner product be

$$\langle A, B \rangle = \text{Trace}(A^T B)$$
$$= \text{Trace}\begin{pmatrix} a_{11}b_{11} + a_{21}b_{21} & \text{something} \\ \text{something} & a_{12}b_{12} + a_{22}b_{22} \end{pmatrix}$$
$$= a_{11}b_{11} + a_{12}b_{12} + a_{21}b_{21} + a_{22}b_{22}$$

Symmetric matrices $= \text{Span}\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ \end{pmatrix} \right.$

$(M = M^T)$

Symmetric matrices $(M = M^T)$ $= \text{Span}\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}$

$\perp$

Anti-symmetric/skew-symmetric matrices $(M = -M^T)$ $= \text{Span}\left\{ \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right\}$

II. Vector spaces over finite fields

Let $V = \{0,1\}^n$ over $\mathbb{F}_2$ (addition/mult. mod 2)

$n=3$: $V = \left\{ \begin{matrix} (0,0,0), (0,0,1), (0,1,0), (0,1,1), \\ (1,0,0), (1,0,1), (1,1,0), (1,1,1) \end{matrix} \right\}$

- lengths don't mean so much

  $\|(1,1,0)\| = \sqrt{1^2+1^2+0^2} = \sqrt{2}$

  but $\frac{1}{\sqrt{2}}(1,1,0) \notin V$; you can't renormalize!

  (length^2 = Hamming weight)

- angles don't mean ~~much~~ anything

  there's no Euclidean geometry here

★ but orthogonality still makes sense!

$\vec{x} \perp \vec{y}$ if $\sum_{i=1}^{n} x_i y_i = 0 \mod 2$

<u>Note</u>: $(0,1,1) \perp (0,1,1)$!!

any even-weight vector is orthogonal to itself

★ the Rank-Nullity Theorem still holds!

<u>Definition</u>: A binary linear error correcting code is a subspace of $\{0,1\}^n$.

<u>Example</u>: 3-bit repetition code

$0 \mapsto 000 \qquad 1 \mapsto 111$

code $= \{(0,0,0), (1,1,1)\}$ 1 dimensional

$= R(A^T)$ for $A = (1 \ 1 \ 1)$

generator matrix

$N(A) = R(A^T)^\perp$

$= R\left(\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\right) = \{(0,0,0), (0,1,1),$

$N(A) = R(A')$

$$= R\left(\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}\right) = \left\{\begin{matrix}(0,0,0), (0,1,1), \\ (1,1,0), (1,0,1)\end{matrix}\right\}$$

$\uparrow$ parity check matrix

2 dimensional

$\rightsquigarrow$ a linear code can be specified by either the generating matrix or the parity-check matrix

III. Function spaces with more general inner products

$\left\{\begin{matrix}\text{all functions mapping} \\ \text{interval } [-1,1] \to \mathbb{R}\end{matrix}\right\}$ is a vector space

inner product?

$$\langle f, g\rangle \overset{?}{=} \cancel{\sum_{x\in[0,1]} f(x)\cdot g(x)}$$

$$= \int_{-1}^{1} dx\, f(x) g(x) \qquad \text{(for square-integrable functions)}$$

angles don't make sense
— but orthogonality does!!

Example: $1 \perp x$ since $\int_{-1}^{1} dx\, x = 0$
w.r.t. this inner product

$1 \perp (x^2 - \frac{1}{3})$ since $\int_{-1}^{1} dx\, (x^2 - \frac{1}{3}) = 0$

$x \perp (x^2 - \frac{1}{3})$ since $\int_{-1}^{1} dx (x^3 - \frac{x}{3}) = 0$

$1, x, x^2 - \frac{1}{3}$ all $\perp$ to $x^3 - \frac{1}{2}x$

$\Rightarrow \{1, x, x^2 - \frac{1}{3}, x^3 - \frac{1}{2}x\}$ is an orthogonal basis (w.r.t. this inner product) for polynomials of degree ≤ 3

Recall: Hermite polynomials $\{1, x, x^2 - 1, x^3 - 3x, x^4 - 6x^2 + 3\}$
These form an orthogonal basis w.r.t. a different inner product.