Week 1 (Hefferon I - I.III.1)

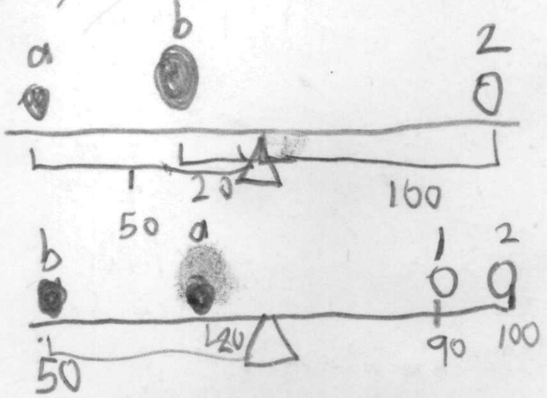
Linear algebra is the study of linear equations - equations where every term is a constant times a variable (or just a constant).

- Notice this is true of lines, $y = ax + b$

- It's also the study of matrices - representations of systems of linear equations.
- Linear algebra comes up in a variety of contexts -
  - Graphics. Matrices are used to rotate & transform shapes.
  - ML. Classifiers and decision-making rules can be implemented with matrices.
  - Simulation of linear systems. If we model systems with linear equations, we can predict their long-term behavior with linear algebra.
  - Pattern discovery. We can sometimes find new patterns in data with linear models.
  - Graph algorithms. Graphs can be represented as matrices so using linear algebra can give new insights.

Linear systems of equations - examples

(Physics & chem because CS is generally complicated.)



$50\underset{2}{a} + 20\underset{5}{b} = 200$

$20a + 50b = 290$

Chemical formula:

$wC_2H_6 + xO_2 \to yH_2O + zCO_2$

$w=2$
$x=7$
$y=6$
$z=4$

balance C: $2w = z$

balance H: $6w = 2y$

balance O: $2x = y + 2z$

Fewer equations than vars will mean no unique sol
(we can scale the quantities up or down)

Markov Chains (used in PageRank):

2 states for server up & down, 1% of going down every minute. If down, 20% chance of coming back up.

$Pr(\text{down}@t+1) = 0.01\,Pr(\text{up}@t) + 0.8\,Pr(\text{down}@t)$

$Pr(\text{up}@t+1) = 0.99\,Pr(\text{up}@t) + 0.2\,Pr(\text{down}@t)$

~ This is a linear dynamical system. Something is changing over time using linear equations.

A system of linear equations is a set of equations of the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = d$, where the a_i and d are all constants. A solution to the system is a tuple of n numbers that can be substituted for the x_i to make all the equations true. (Example: solution to 1st)

Solving With "Gauss's Method"

There are 3 moves we can perform on a system of equations that won't change the solution.

Swapping ~ Swap two equations. Looks useless now, but we'll use it as part of an algorithm.

Multiplying by a scalar ~ Replace an equation with the same equation multiplied by a constant on both sides.

Row combination ~ Replace an equation with the sum of that equation and a multiple of another.

A procedure for solving any linear system of equations:

For $i = 1$ to n:
  Use the operations above to eliminate variable x_i from all equations $i+1 \ldots n$

For $j = n$ to 1:
  Solve for x_j using scaling & substitution

Balancing example:

$$50a + 20b = 200$$
$$20a + 50b = 290$$

Goal: Eliminate "20a"

Multiply $\rho_1 * 2$
$\rho_2 * 5$
subtract ρ_1

$$100a + 40b = 400$$
$$100a + 250b = 1450$$
$$0 + 210b = 1050$$
$$b = 5$$

Goal: solve for b, then a

solve for b
sub into ρ_1

$$100a + 200 = 400$$
$$100a = 200$$
$$a = 2$$

Slightly bigger example

$x=7$
$y=3$
$z=5$

$$x + 2y + z = 18$$
$$x - 2y - z = -4$$
$$2x - y + z = 16$$

Goal: just top has x

Sub ρ_1 from ρ_2: $(\rho_2)\quad -4y - 2z = -22$

Sub $2\rho_1$ from ρ_3: $(\rho_3)\quad -5y - z = -20$

Goal: just 2nd row has y

$\rho_3 * 4$: $-20y - 4z = -80$

$\rho_2 * -5$: $20y + 10z = 110$

Add ρ_2 to ρ_3: $6z = 30$

Solve for z: $z = 5$

Backsub into ρ_2: $20y + 50 = 110$, $20y = 60$, $y = 3$

Backsub into ρ_1: $x + 6 + 5 = 18$, $x = 7$

We can write this in matrix form as shorthand.

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 18 \\ 1 & -2 & -1 & -4 \\ 2 & -1 & 1 & 16 \end{array}\right]$$

"augmented matrix"

$\rho_2 - \rho_1$, $\rho_3 - 2\rho_1$:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 18 \\ 0 & -4 & -2 & -22 \\ 0 & -5 & -1 & -20 \end{array}\right]$$

$\rho_3 = 4\rho_3 - 5\rho_2$:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 18 \\ 0 & -4 & -2 & -22 \\ 0 & 0 & 6 & 30 \end{array}\right]$$

"echelon form": each leading var of row to the right of the one above it (any 0 rows at bottom)

$\rho_3/6$, $\rho_2/-4$:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 18 \\ 0 & 1 & \frac{1}{2} & 5\frac{1}{2} \\ 0 & 0 & 1 & 5 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 18 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 5 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & 2 & 0 & 13 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 5 \end{array}\right]$$

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

Gauss-Jordan elimination is similar to the Gauss's Method already presented, ~~but~~ is more explicit about what to do after the step of eliminating variables down:

- Divide each row by its leading value, so it has a 1 there.
- Use normal row combination to eliminate the entries above the leading 1's.

("reduced echelon form") (refers to echelon form w/ leading 1's and zeros above & below them)

2supp

$$\begin{aligned} x + y - z &= 0 \\ 3x - 2y + 2z &= 5 \\ 5x - y - z &= 0 \end{aligned} \longrightarrow \left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 3 & -2 & 2 & 5 \\ 5 & -1 & -1 & 0 \end{array}\right]$$

subtract 3 * p1 and 5 * p1

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & -5 & 5 & 5 \\ 0 & -6 & 4 & 0 \end{array}\right] \xrightarrow{\text{divide by } -5 \text{ and } -6} \left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 1 & -\frac{2}{3} & 0 \end{array}\right]$$

$$\xrightarrow{r3 -= r2} \left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & \frac{1}{3} & 1 \end{array}\right] \xrightarrow{r3 *= 3} \left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

$$\xrightarrow{\substack{r2 += r3 \\ r1 += r3}} \left[\begin{array}{ccc|c} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right] \xrightarrow{r1 -= r2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Check: $1 + 2 - 3 = 0$ ✓

$3(1) - 2(2) + 2(3) = 3 - 4 + 6 = 5$ ✓

$5(1) - 1(2) - 3 = 0$ ✓

Number of solutions

What happens if we try to solve an underconstrained system – more vars than eqs?

(Chem eq)

$$
\begin{aligned}
2w - z &= 0 \\
6w - 2y &= 0 \\
2x - y - 2z &= 0
\end{aligned}
\qquad
\left[\begin{array}{cccc|c} 2 & 0 & 0 & -1 & 0 \\ 6 & 0 & -2 & 0 & 0 \\ 0 & 2 & -1 & -2 & 0 \end{array}\right]
$$

Subtract $3\rho_1$ from ρ_2:
$$
\left[\begin{array}{cccc|c} 2 & 0 & 0 & -1 & 0 \\ 0 & 0 & -2 & 3 & 0 \\ 0 & 2 & -1 & -2 & 0 \end{array}\right]
$$

Swap
$$
\left[\begin{array}{cccc|c} 2 & 0 & 0 & -1 & 0 \\ 0 & 2 & -1 & -2 & 0 \\ 0 & 0 & -2 & 3 & 0 \end{array}\right]
$$

Now the bottom row says $-2y + 3z = 0$ and that's as close as we'll get. We can still solve up in terms of an unknown variable y:

$$z = \tfrac{2}{3}y$$
$$x = \tfrac{7}{6}y$$
$$w = \tfrac{1}{3}y$$

We get ratios instead of a straight answer. The chemical equation, if we let $y = 6$ to make everything an integer, is

$$\underset{\uparrow \atop w}{2}\, C_2H_6 + \underset{\uparrow \atop x}{7}\, O_2 \longrightarrow \underset{\uparrow \atop y}{6}\, H_2O + \underset{\uparrow \atop z}{4}\, CO_2$$

But we could pick a bigger y to make a bigger batch.

In general one of 3 cases applies:

- Infinite solutions. True if an equation is redundant given the others, leaving at least one param free.
- One solution. If each equation is "linearly independent" of the others (more on that later)
- No solutions. If the equations are mutually inconsistent, like $x+y=2$ and also $x+y=3$.

A linear system is <u>homogeneous</u> if it's in the form $a_1x_1 + a_2x_2 + \dots + a_nx_n = 0$. Homogeneous systems help us learn about non-homogeneous ones.

Homogeneous systems must have at least one solution - the zero vector, $(0, 0, \dots, 0)$. If they have more, then they have an infinite number because scaling the nonzero solution produces other viable solutions.

If a square matrix representing a homogeneous system of equations has one solution, it's <u>nonsingular</u>. If it has infinite solutions, it's <u>singular</u>. ("Singular" here means something like "remarkable.")

$\boxed{p. 30}$

Singular: $\begin{aligned} 2x + y &= 0 \\ 4x + 2y &= 0 \end{aligned}$ Infinite solutions: $x = -y/2$ $\begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$ singular

Not singular: $\begin{aligned} 2x + y &= 0 \\ x + 2y &= 0 \end{aligned}$ One solution: "zero vector" $x = y = 0$ $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ nonsingular

~~If a system of equations (not homogeneous) has a unique solution, so will its homogeneous version. And if it has infinite solutions, so will its homogeneous version. But if it has no solutions the homogeneous could have one or many.~~

Example: $\begin{aligned} 2x + y &= 2 \\ 4x + 2y &= 4 \end{aligned}$ $x=1, y=0$ is a sol, and singular $\Rightarrow$ infinite sols. But if the bottom $= 3$, then no solutions.

In the general case for linear systems, the homogeneous matrix tells us whether there will be one or infinite solutions - <u>if</u> there is a solution at all.

mv

$2x + y = 2$
$4x + 2y = 3$ no sol - doesn't matter if singular (week 1) 4

If a particular solution exists, the set of solutions is given by $\{\vec{p} + \vec{h} \mid \vec{h}$ solves the homogeneous system$\}$

$\left[\begin{array}{cc|c} 2 & 1 & 2 \\ 4 & 2 & 4 \end{array}\right]$ $\vec{p} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ homogeneous sols $x = -y/2 \Rightarrow c\begin{bmatrix} 1 \\ -2 \end{bmatrix}$

Total sols $\begin{bmatrix} 1 \\ 0 \end{bmatrix} + c\begin{bmatrix} 1 \\ -2 \end{bmatrix}$

Try $c=3$ arbitrarily: $\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 \\ -6 \end{bmatrix} = \begin{bmatrix} 4 \\ -6 \end{bmatrix}$

$2(4) + 1(-6) = 2$ ✓
$4(4) + 2(-6) = 4$ ✓

Proof idea (one direction): Substituting $\vec{p} + \vec{h}$ into $a_1x_1 + a_2x_2 + \dots = d$ gives

$a_1(p_1 + h_1) + a_2(p_2 + h_2) + \dots + a_n(p_n + h_n)$

$= a_1p_1 + a_1h_1 + a_2p_2 + a_2h_2 + \dots$

$= \underbrace{(a_1p_1 + a_2p_2 + \dots)}_{\text{this is the solution that works}} + \underbrace{(a_1h_1 + a_1h_2 + \dots)}_{\text{this is zero}}$

$= d$ so the equation works, adding homogeneous sols has a net effect of zero

Matrices & vectors

So we can see matrices as systems of linear equations, and vectors as solutions to those equations. This leads us to one interpretation of matrix multiplication: we can see what we get when substituting a solution by multiplying a matrix (system of equations) by a vector (solution).

$$\begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ -6 \end{bmatrix} = \begin{bmatrix} 2(4) + 1(-6) \\ 4(4) + 2(-6) \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

To produce each element of the solution consult the corresponding row of the matrix and sum (matrix element * vector element) for each place.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}\begin{bmatrix} 10 \\ 11 \\ 12 \end{bmatrix} = \begin{bmatrix} 1*10 + 2*11 + 3*12 \\ 4*10 + 5*11 + 6*12 \\ 7*10 + 8*11 + 9*12 \end{bmatrix} = \begin{bmatrix} 10+22+36 \\ 40+55+72 \\ 70+88+108 \end{bmatrix} = \begin{bmatrix} 68 \\ 167 \\ 266 \end{bmatrix}$$

This gives us the results of
$$\begin{matrix} x + 2y + 3z \\ 4x + 5y + 6z \\ 7x + 8y + 9z \end{matrix} \quad \text{getting} \quad \begin{matrix} x = 10 \\ y = 11 \\ z = 12 \end{matrix}$$

The row length & vector length need to match — you can't get a meaningful answer by providing the wrong number of variable values.

~~There are other interpretations of matrix multiplication,~~

In a linear dynamical system, the equations are describing how a system evolves from one time to the next. If I write a game where for (p)opulation & (h)appiness

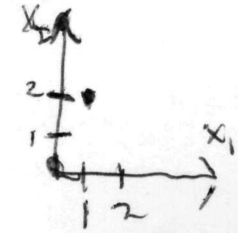
$$p' = p + \tfrac{1}{2}h$$
$$h' = \cancel{h - p/100}\ -p/100 + h$$

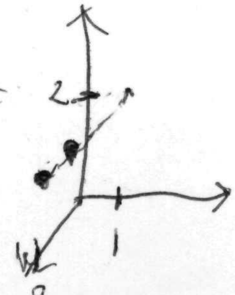
so population causes happiness to fall but happiness causes population to rise, this is equivalent to multiplying the $\begin{bmatrix} p \\ h \end{bmatrix}$ vector by $\begin{bmatrix} 1 & 1/2 \\ -\frac{1}{100} & 1 \end{bmatrix}$

to get the levels at the next time step. By the end of the module, we'll be able to predict whether this population will eventually collapse or grow forever, just by analyzing the matrix.

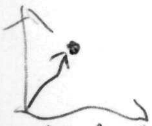
Geometric Interpretations

A solution to a system of equations isn't the only interpretation of a vector, we can also plot vectors in space.

$\begin{bmatrix} 1 \\ 2 \end{bmatrix} =$ 

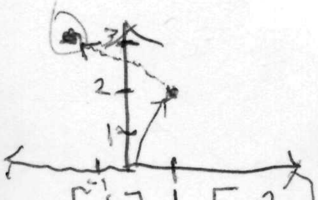
$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} =$ 

We can think of vectors as points or as arrows from the origin (zeros).



In the second case, <u>vector addition</u> tells us where we would end up if we moved to the ~~first~~ first location, then treated that as the starting point for the second vector. The dimensions should be the same. The result is a vector where each element is the sum of the other vectors' elements in that place:

$$\begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$



$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

Multiplication by a matrix can be seen as a transformation of a vector. Rotation, scaling, and other operations can be performed by matrix multiplication — a method often used in graphics. We'll come back to that.

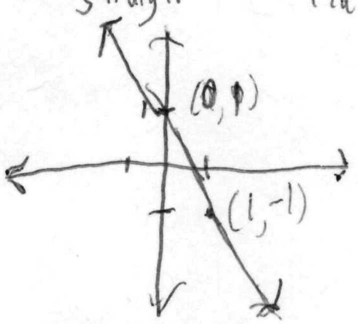
$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

rotate 90°

Equations as lines & planes, etc

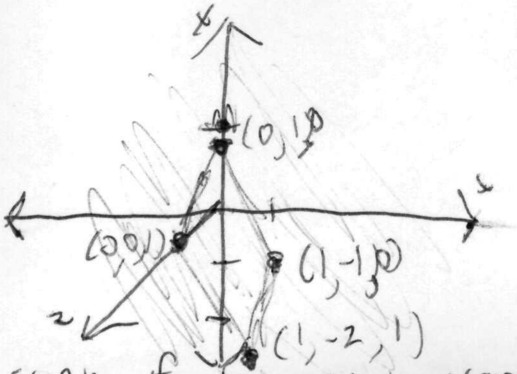
Linear equations may have an infinite number of solutions, but we can visualize those solutions as straight lines or flat planes in 2D and 3D, respectively.

$$2x + y = 1 \Rightarrow y = -2x + 1$$



This is some of the reason we use the term "linear" to describe these equations.

$$2x + y + z = 1$$



Curvature would come from some variable being squared or otherwise passed through a non-linear function. In a linear equation, an adjustment to one variable can be compensated for by a proportional change to another. The slope according to any particular variable is constant. (which we can show in multivar)

It's hard to visualize 4D, 5D etc equivalents to lines and planes, though they do come up in machine learning (where # dimensions = # of features for each datapoint). We know this much about them:

- They're still flat, not curved.
- From some perspective, we could still think of them as planes, just as looking at a plane edge-on makes it look like a line.

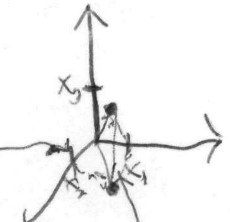
We call these surfaces "hyperplanes."

$w + 2x + y + z = 1 \Rightarrow$ $w=0$ looks like above, but $w=1$ shifts everything ....

Vector length & normalization

Length of a vector in 2D uses the familiar distance equation $d = \sqrt{x^2+y^2}$ where x and y are the coordinates ⊥ this is distance from the origin $\sqrt{(x-0)^2+(y-0)^2}$ and is just using what we know about triangles.

This generalizes in more dimensions (n) to $\sqrt{x_1^2+x_2^2+x_3^2+\dots+x_n^2}$ (the distance equation generalizes in the same way). Notice it's still just a square root.



$\sqrt{x_1^2+x_2^2}$

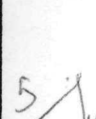
We could think of this as finding the length of a vector that takes 2 dimensions into account first then finding a vector that takes into account the third using that length, & repeating for higher dims. Getting the answer one triangle at a time.

$\sqrt{\left(\sqrt{x_1^2+x_2^2}\right)^2+x_3^2} = \sqrt{x_1^2+x_2^2+x_3^2}$

...

⇒ We write this as $|V|$ if V was the original vector.

Sometimes we may want to use vectors to just represent a direction in space, throwing out the magnitude. The convention is to rescale the vector to have a length of 1. This is called "normalization". Dividing each component by the magnitude will accomplish this.



$\begin{bmatrix}3\\4\end{bmatrix} \xrightarrow[\text{normalize}]{} \begin{bmatrix}\frac{3}{5}\\\frac{4}{5}\end{bmatrix}$ $\quad \underbrace{\sqrt{\left(\frac{3}{5}\right)^2+\left(\frac{4}{5}\right)^2}}_{\text{check}} = \sqrt{\frac{9}{25}+\left(\frac{16}{25}\right)} = \sqrt{\frac{25}{25}} = \sqrt{1} = 1$

$d=\sqrt{a^2+b^2}$ $\quad \begin{bmatrix}a\\b\end{bmatrix} \Rightarrow \begin{bmatrix}\frac{a}{d}\\\frac{b}{d}\end{bmatrix} \Rightarrow \sqrt{\left(\frac{a}{d}\right)^2+\left(\frac{b}{d}\right)^2} \Rightarrow \sqrt{\frac{a^2+b^2}{d^2}} = \sqrt{\frac{a^2+b^2}{a^2+b^2}} = 1$

Normalized vectors have their endpoints on a sphere (or hypersphere)

Angles between vectors & the dot product

The dot product of two vectors is a sum of the products of their components.

$$\vec{a}\cdot\vec{b} = a_1b_1 + a_2b_2 + \dots + a_nb_n$$

So

$$\begin{bmatrix}0\\1\end{bmatrix}\cdot\begin{bmatrix}2\\3\end{bmatrix} = 0\cdot 2 + 1\cdot 3 = 3.$$

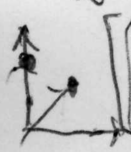
Using another result (Law of Cosines) that we won't prove, we could show that

$$\vec{a}\cdot\vec{b} = |\vec{a}||\vec{b}|\cos\theta$$

This gives us a way to calculate an angle between vectors - find the dot product, divide by both magnitudes, find the arccos (cos inverse).

$$\begin{bmatrix}0\\1\end{bmatrix}\cdot\begin{bmatrix}\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}\end{bmatrix} = 0 + \frac{\sqrt{2}}{2} = \underbrace{\frac{\sqrt{2}}{2}}_{\text{dot product}} = 1\cdot\sqrt{\left(\frac{\sqrt{2}}{2}\right)^2+\left(\frac{\sqrt{2}}{2}\right)^2}\cos\theta = \sqrt{\frac{2}{4}+\frac{2}{4}}\cos\theta$$

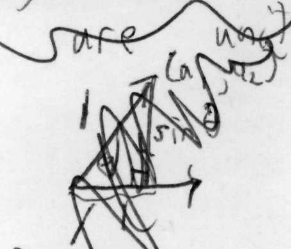
$$= \cos\theta$$



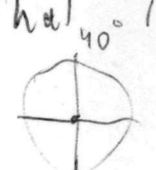
$\boxed{\text{(maybe scale both by 2)}}$

$$\cos\theta = \frac{\sqrt{2}}{2} \Rightarrow \theta = 45^\circ \text{ or } \frac{\pi}{4} \text{ radians}$$

Showing this works takes us a little deeper into trig than we want to go. ~~But notice that if a and b are unit vectors, this has a natural trig explanation,~~



Notice that if $\cos\theta = 0$ ~~Same if b is a~~ the angle is a right angle.



So if $\vec{a}\cdot\vec{b} = 0$ and $|\vec{a}| \neq 0$ and $|\vec{b}| \neq 0$, we know $\cos\theta = 0$ and the vectors are at right angles. We'll sometimes want this fact to find a vector that moves straight away from a surface  or to make good coordinate systems