ME 536

Weeks 4: Some more math
2 remember

Matrices

Collection of vectors, where a vector generally represents a data point.

Vectors can be in the rows or columns of the **matrix**.

<u>**Convention for the rest of the course:**</u>

Columns *of a data matrix represent* ***data points*** *in the rest of the class unless otherwise mentioned*

Let's refresh our memories on ***Matrices*** *and* ***Vectors*** *in action*

Matrix multiplied by a Vector

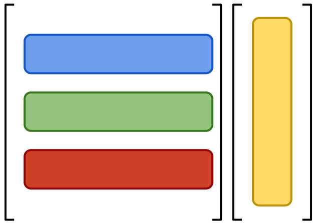
$$\mathbf{M}\mathbf{x} = \mathbf{a}$$

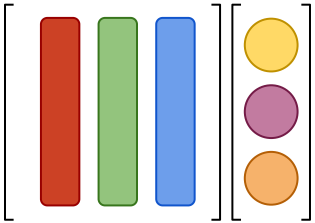
Vector multiplies the matrix on the right.

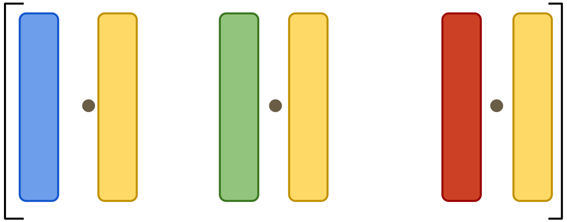
Recall: $\mathbf{M}$ is the matrix and $\mathbf{x}$, $\mathbf{a}$ are vectors.

Matrix multiplied by a Vector: 2 Main Interpretations

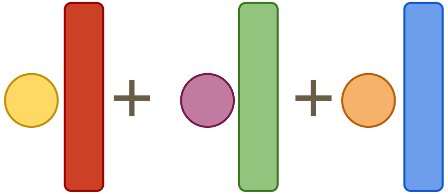
$$\mathbf{Mx} = \mathbf{a}$$





$\mathbf{a}^T =$ 

Dot (or inner) product version

$\mathbf{a} =$ 

Linear Combination version

Generally we are used to: square $\mathbf{M}$ with no rank deficiency, i.e. $\mathbf{M}$ is full rank

Row and Column Space of a Matrix

$$\begin{bmatrix} 4 & 2 & -1 & -8 & 12 & 1 \\ 2 & 6 & -3 & -4 & 6 & 3 \\ 5 & 4 & -2 & -10 & 15 & 2 \end{bmatrix}$$

Given $\mathbf{M}$

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

Column space of $\mathbf{M}$: $\mathbf{C(M)}$

Spanned by columns of $\mathbf{M}$

Row space of $\mathbf{M}$: $\mathbf{R(M)}$

Spanned by rows of $\mathbf{M}$

Row and Column Space of a Matrix

Column space of $\mathbf{M}$: $\mathbf{C(M)} \longrightarrow$ Spanned by columns of $\mathbf{M}$

Row space of $\mathbf{M}$: $\mathbf{R(M)} \longrightarrow$ Spanned by rows of $\mathbf{M}$

$$\mathbf{C(M)} \stackrel{?}{=} \mathbf{R(M)}$$

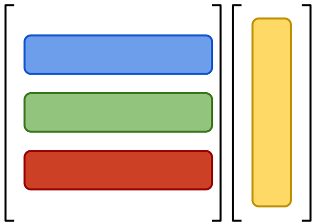
1- Fix for general case if it is not <u>always</u> `True`

2- Special case when this is correct?

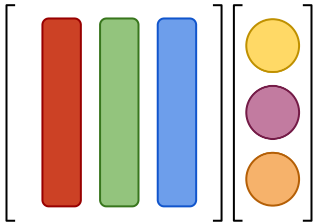
Matrix multiplied by a Vector: 2 Main Interpretations

$$\mathbf{Mx} = \mathbf{0}$$

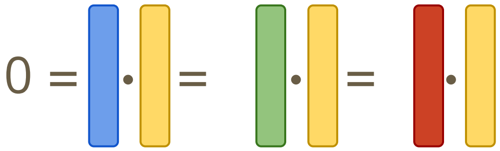
What if the trivial is not the only alternative: $\mathbf{x} \neq \mathbf{0}$

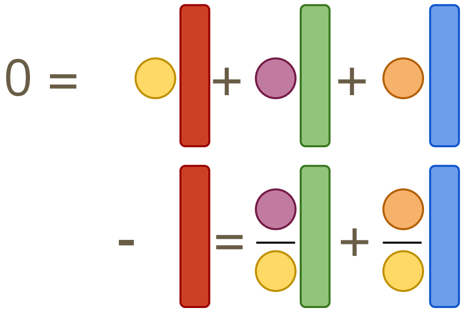


Dot (or inner) product version



Linear Combination version

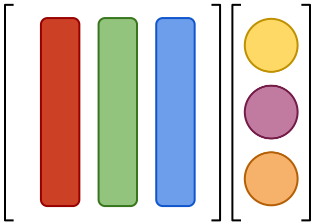


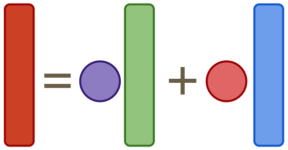


$\mathbf{Mx} = \mathbf{0}$: Different Cases

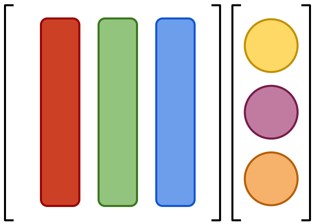
Let:

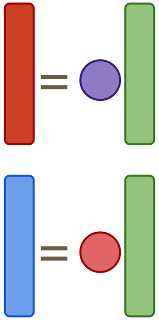
rank $(\mathbf{M}_{3x3}) = 2$





rank $(\mathbf{M}_{3x3}) = 1$





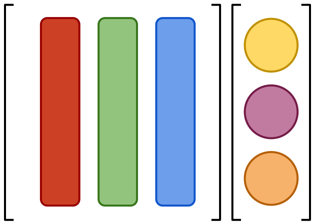
What does $C(\mathbf{M})$ represent in each case?

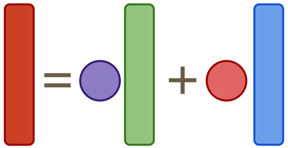
What about the basis of $\mathbf{M}$?

$\mathbf{Mx} = \mathbf{0}$: NULL Space - N($\mathbf{M}$)

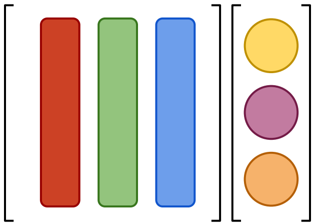
Let:

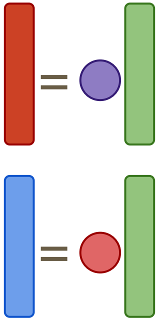
rank ($\mathbf{M}_{3x3}$) = $\mathbf{2}$





rank ($\mathbf{M}_{3x3}$) = $\mathbf{1}$

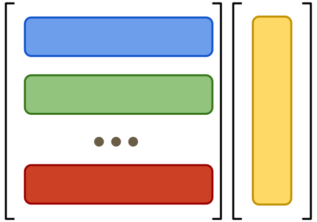


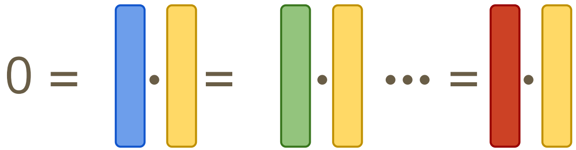


Where does $\mathbf{x}$ come from? Where do all $\mathbf{x}$ live?

Dimension of the N($\mathbf{M}$)? Number of free Parameters?

WRAP UP for $N(\mathbf{M}_{mxn})$





Let: $\mathbf{M}_{mxn}$

$N(\mathbf{M}) \perp R(\mathbf{M})$

similarly

$N(\mathbf{M}) \perp C(\mathbf{M}^T)$

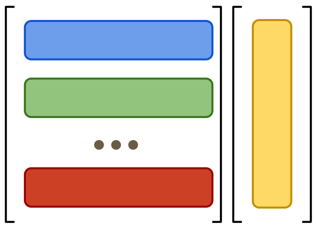
where

$\dim(R(\mathbf{M})) = \dim(C(\mathbf{M}^T)) = \operatorname{rank}(\mathbf{M}) = r$

$\dim(N(\mathbf{M})) =$ **?**

WRAP UP for N($\mathbf{M}_{\mathbf{mxn}}$)

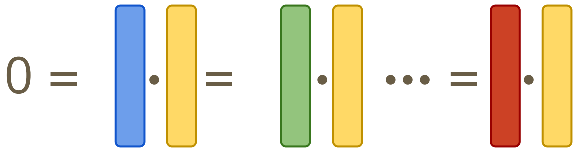
Similarly for $\mathbf{M}^T_{\mathbf{nxm}}$



N($\mathbf{M}^T$) $\perp$ R($\mathbf{M}^T$)

hence

N($\mathbf{M}^T$) $\perp$ C($\mathbf{M}$)



where

dim(R($\mathbf{M}^T$)) = dim(C($\mathbf{M}$)) = rank($\mathbf{M}$) = r

dim(N($\mathbf{M}^T$)) = **?**

Time Capsule: note to the Future

Recall that choosing a basis **from the data** matrix $\mathbf{M}$ is:

- Practical
- Yet subject to *ill-conditioned* cases!
  - For a square matrix, you can check for very small $|\cdot|$
  - But what if the Basis vectors do not form a square matrix?

Time Capsule: note to the Future

Given a full rank square matrix $\mathbf{A}_{nxn}$

We can consider this as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$, i.e. $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^n$

> For $\mathbf{Ax} = \mathbf{a}$, we have $\mathbf{x}, \mathbf{a} \in \mathbb{R}^n$

What if $\mathbf{A}_{nxm}$ is not square?

Then we have a mapping from $\mathbb{R}^m$ to $\mathbb{R}^n$, i.e. $\mathbf{A} : \mathbb{R}^m \rightarrow \mathbb{R}^n$

> For $\mathbf{Ax} = \mathbf{a}$, we have $\mathbf{x} \in \mathbb{R}^m$, and $\mathbf{a} \in \mathbb{R}^n$

Consider cases:

1. $\mathbf{A}_{3x2}$
2. $\mathbf{A}_{200x10}$
3. $\mathbf{A}_{2x3}$

Can you move back and forth?

Try to think about it, drawing might help

A famous Space: $\mathbb{R}^n$

- All of us know a very well-known vector space: $\mathbb{R}^n$

For $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, we know that the length of vector x

$$\|x\| = \left(\sum_{k=1}^{n} |x_k|^2 \right)^{1/2}$$

- For general vector spaces, we need a concept that corresponds to length in $\mathbb{R}^n$
- We use "norm" instead of "length"

Good news:

Our data will almost exclusively live in $\mathbb{R}^n$

So no weird brain freezing spaces

A famous Norm: length

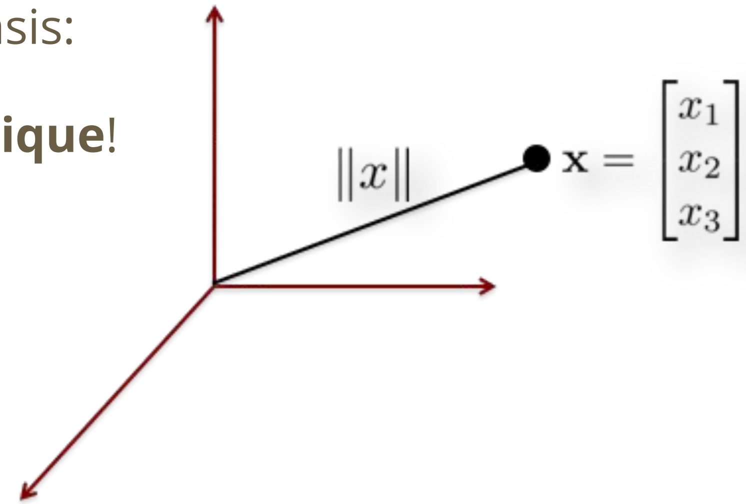
Euclidean Vector Norm

For a vector $x \in \mathbb{R}^n$, the euclidean norm is defined by:

$$\|x\| = (x_1^2 + x_2^2 + \ldots + x_n^2)^{1/2}$$

Just like the standard basis:

- *Famous*, but **not unique**!



In general: Norm

A norm for a real or complex vector space $\mathcal{V}$ is a function

$$||.|| : \mathcal{V} \to \mathbb{R}$$

with three properties:

1. $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0 \iff x = 0$
2. $\|\alpha x\| = |\alpha| \|x\|$ for all scalars α and for all $x \in \mathcal{V}$
3. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathcal{V}$

p-Norms

p-Norms

For $p \geq 1$, the p-norm of a vector $x \in \mathbb{R}^n$ is defined by:

$$\|x\|_p = (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{1/p}$$

l_1 norm $\rightarrow$	• $\|\mathbf{x}\|_1$ - a.k.a Taxicab or Manhattan norm
l_2 norm $\rightarrow$	• $\|\mathbf{x}\|_2$- Euclidean norm
…	• …
l_∞ norm $\rightarrow$	• $\|\mathbf{x}\|_\infty$ - max($\|x_1\|$, … , $\|x_n\|$)

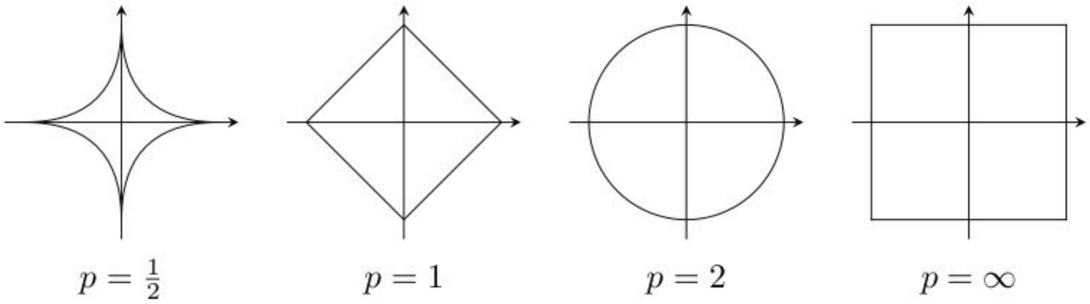
Even though value of *p* determines the norm, *in general* they are referred to as l_p norm. (with a *lowercase L*)

NOTE: shape of l_p norms:

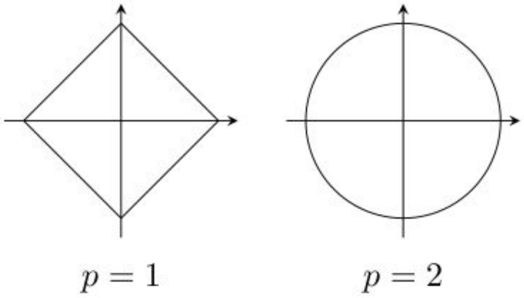
What are the set of **points with unit norm** look like:

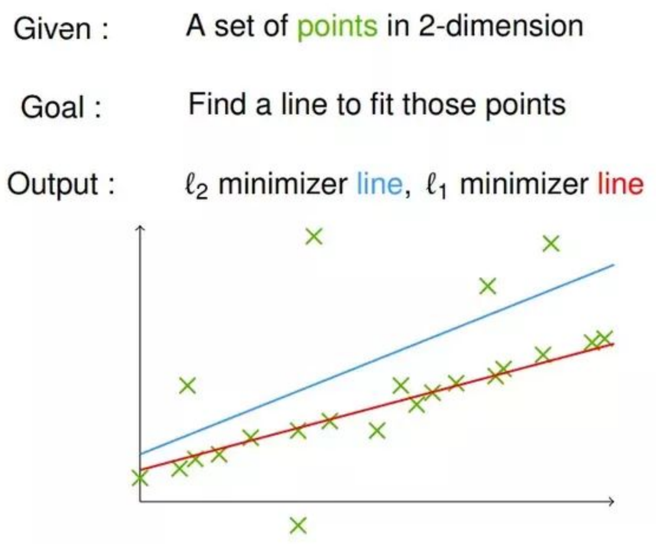
$$l_p = 1$$

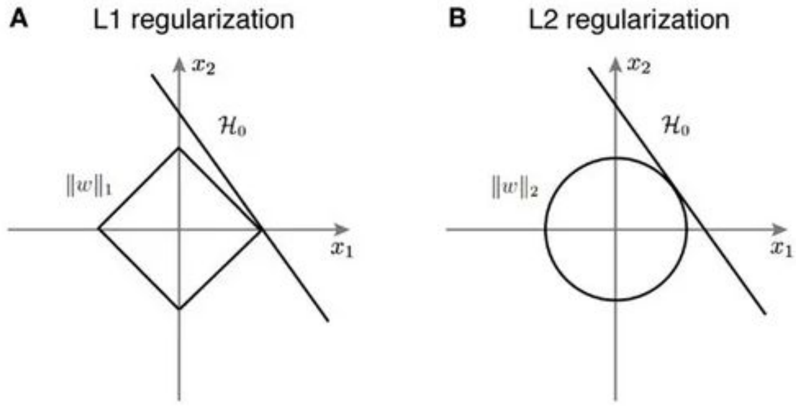
for different values of p = { ½, 1, 2, ∞} in 2D.



NOTE: l_1 over l_2 ?







Images are from: https://www.quora.com/When-would-you-chose-L1-norm-over-L2-norm

NOTE: l_0 norm

Despite the fact that l_0 norm is **not a norm by definition** but it is very often *used in practice*.

l_0 norm is defined by the **number of non-zero elements** in the vector.

→ i.e. minimizing l_0 results in sparse solution

NOTE: yet another norm

Note that if $\mathbf{D}$ is a symmetric matrix with positive diagonal terms, then for the vector $\mathbf{v}$:

$$\mathbf{v}^{\mathrm{T}}\mathbf{D}\mathbf{v}$$

defines a norm in the form of a weighted sum of squares.

Example: a 2x2 case

$$\mathbf{D} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\mathbf{v}^T\mathbf{D}\mathbf{v} = av_1^2 + bv_2^2$$

Good news:

Every finite dimensional real or complex topological vector space has a norm

So no hunt for finding norms, well mostly...

A famous Inner Product in: $\mathbb{R}^n$ a.k.a *DOT Product*

$\mathbb{R}^n$: Collection of all finite sequences

$$x = (x_1, x_2, \ldots, x_n), \;\; x_k \in \mathbb{R} \text{ for } k = 1, 2, \ldots, n$$

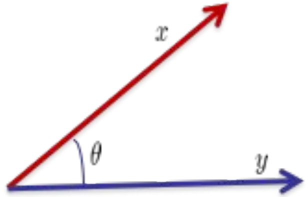
We can define norm for $\|x\| = \left(\sum_{k=1}^{n} x_k^2\right)^{1/2}$

Recall: inner product of $x, y \in \mathbb{R}^n$, $\langle x, y\rangle = \sum_{k=1}^{n} x_k y_k$

Clearly: $\langle x, x\rangle = \|x\|^2$

Inner product is a very important tool for analysis in $\mathbb{R}^n$.

It is a measure of angle between vectors



$\langle x, y\rangle = \|x\|\|y\| \cos\theta$

In general: Inner Product

Definition: An *inner product* on a real (or complex) vector space $\mathcal{V}$ is a function that maps each ordered pairs of vectors v and w to a real (or complex) scalar $\langle v, w \rangle$ such that the following properties hold:

1. $\langle v, v \rangle$ is real with $\langle v, v \rangle \geq 0$ for all $v \in V$ and $\langle v, v \rangle = 0 \iff v = 0$
2. $\langle v, \alpha w \rangle = \overline{\alpha} \langle v, w \rangle$ for all scalars α and for all $v, w \in \mathcal{V}$
3. $\langle v, w + u \rangle = \langle v, w \rangle + \langle v, u \rangle$ for all $v, u, w \in \mathcal{V}$
4. $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for all $v, w \in \mathcal{V}$ (note that $\langle v, w \rangle = \langle w, v \rangle$ for real spaces)

Brain teaser:

Try to come up with *your custom:*

- norm for vectors in $\mathbb{R}^n$

with three properties:

1. $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0 \iff x = 0$
2. $\|\alpha x\| = |\alpha| \|x\|$ for all scalars α and for all $x \in \mathcal{V}$
3. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathcal{V}$

- inner product for vector pairs in $\mathbb{R}^n$

1. $\langle v, v \rangle$ is real with $\langle v, v \rangle \geq 0$ for all $v \in V$ and $\langle v, v \rangle = 0 \iff v = 0$
2. $\langle v, \alpha w \rangle = \overline{\alpha} \langle v, w \rangle$ for all scalars α and for all $v, w \in \mathcal{V}$
3. $\langle v, w + u \rangle = \langle v, w \rangle + \langle v, u \rangle$ for all $v, u, w \in \mathcal{V}$
4. $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for all $v, w \in \mathcal{V}$ (note that $\langle v, w \rangle = \langle w, v \rangle$ for real spaces)

Inner Product Spaces

Simple Recipe: *A vector space* with an *Inner Product*

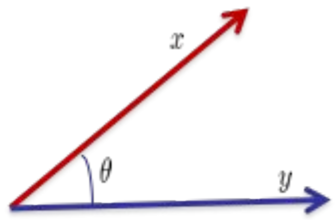
Examples: $\mathbb{R}^n$ with $\langle x, y \rangle = \sum_{k=1}^{n} x_k y_k$

$\mathbb{C}^n$ with $\langle x, y \rangle = \sum_{k=1}^{n} x_k \overline{y_k}$

Cauchy-Schwarz Inequality

Inner product for $\mathbb{R}^2$

$$\langle x, y \rangle = \|x\| \|y\| \cos(\theta)$$



$$\langle x, y \rangle = \|x\| \|y\| \cos \theta$$

$$|\langle x, y \rangle| \leq \|x\| \|y\| = \langle x, x \rangle^{1/2} \langle y, y \rangle^{1/2}$$

For a general normed vector space $\mathcal{V}$

Theorem. $|\langle v, w \rangle| \leq \langle v, v \rangle^{1/2} \langle w, w \rangle^{1/2}, \quad \forall v, w \in \mathcal{V}$

Good to know

- In any inner product vector space, regardless of the inner product we can always define a norm

Lemma. *If $\mathcal{V}$ has the inner product $\langle .,. \rangle$, then*

$$\|v\| = |\langle v, v \rangle|^{1/2} \text{ is a norm}$$

- But opposite is not true. We may not always define an inner product from a given norm

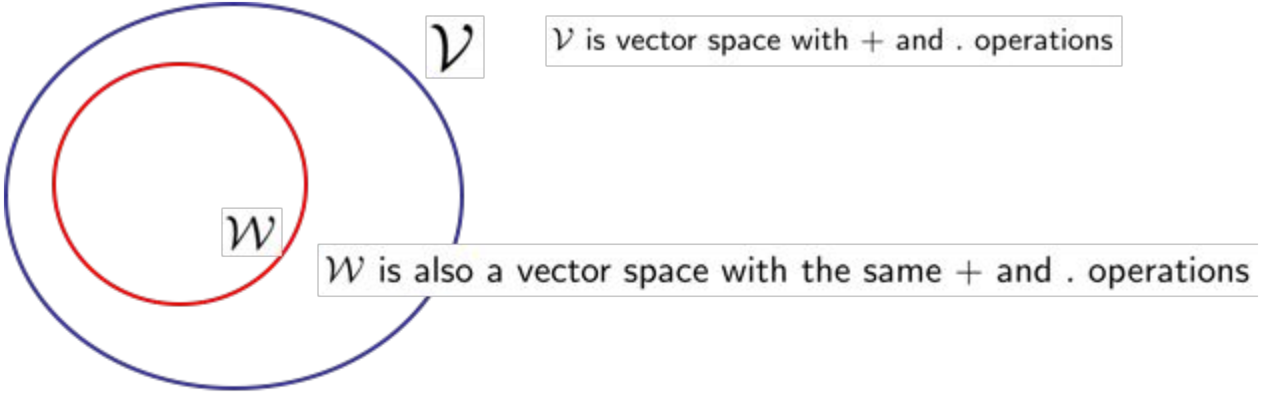
But...

- We may define an inner product from a given norm if the parallelogram law holds for the norm

$$\|v+w\|^2 + \|v-w\|^2 = 2\left(\|v\|^2 + \|w\|^2\right)$$

$$\text{for all } v, w \in \mathcal{V}$$

Subspaces

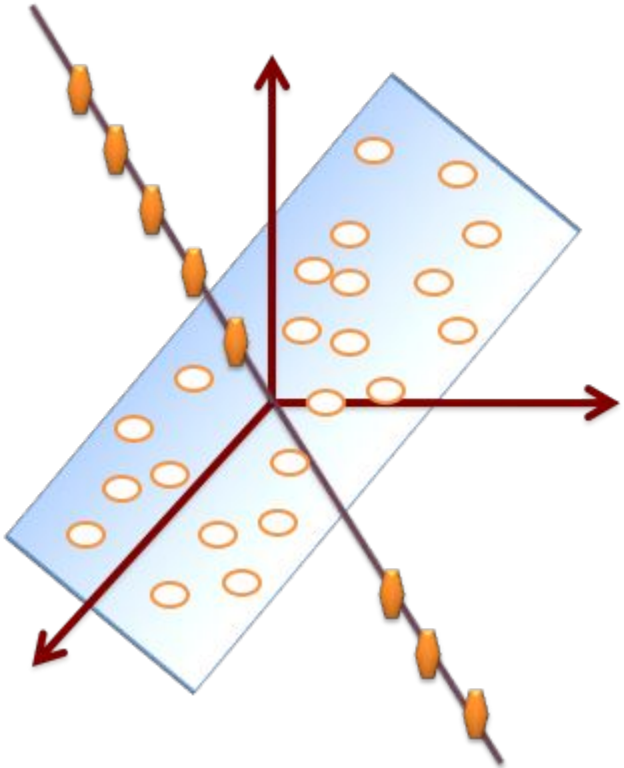


Lemma. *A subset* $\mathcal{W} \subset \mathcal{V}$ *is a subspace if*
$\alpha, \beta \in \mathcal{F}$ *and* $v, w \in \mathcal{W} \Rightarrow \alpha v + \beta w \in \mathcal{W}$

<u>Practical consequence:</u>

Any *line*, *plane*, etc are **subspaces** if they **include the origin**

Subspaces: Examples



A line through origin in $\mathbb{R}^3$ is a 1-dimensional subspace of $\mathbb{R}^3$

A plane through origin in $\mathbb{R}^3$ is a 2-dimensional subspace of $\mathbb{R}^3$

Sum of Subspaces

If $\mathcal{X}$ and $\mathcal{Y}$ are subspaces of a vector space $\mathcal{V}$, then the *sum* of $\mathcal{X}$ and $\mathcal{Y}$ is defined as all possible sums of vectors from $\mathcal{X}$ and $\mathcal{Y}$:

$$\mathcal{X} + \mathcal{Y} = \{x + y : x \in \mathcal{X} \text{ and } y \in \mathcal{Y}\}$$

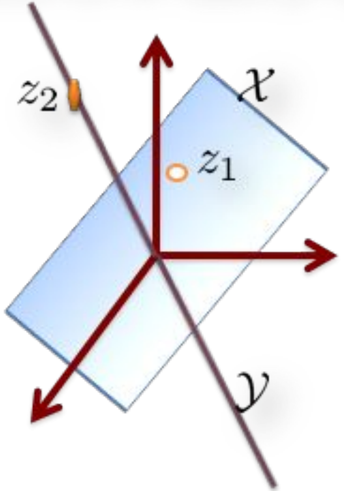
The sum of $\mathcal{X}$ and $\mathcal{Y}$ is a subspace of $\mathcal{V}$

Union of Subspaces

If $\mathcal{X}$ and $\mathcal{Y}$ are subspaces of a vector space $\mathcal{V}$, then the *union* of $\mathcal{X}$ and $\mathcal{Y}$ is defined:

$$\mathcal{X} \cup \mathcal{Y} = \{z : z \in \mathcal{X} \text{ or } z \in \mathcal{Y}\}$$

The union of $\mathcal{X}$ and $\mathcal{Y}$ may not be a subspace of $\mathcal{V}$



$$z_1 \in \mathcal{X} \cup \mathcal{Y}, z_2 \in X \cup \mathcal{Y}$$

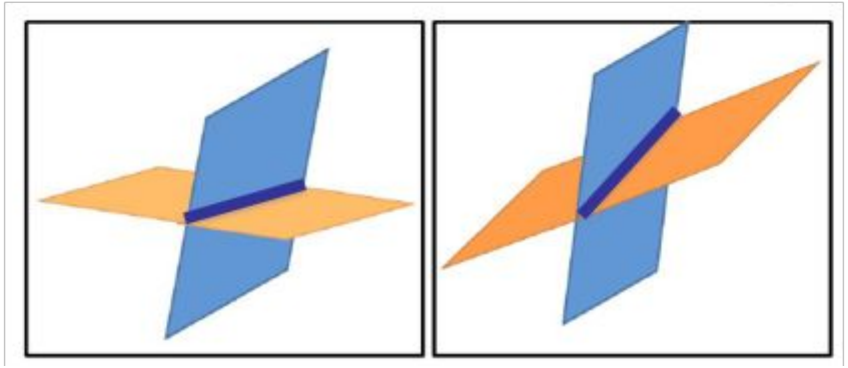
$$\text{but } z_1 + z_2 \notin \mathcal{X} \cup \mathcal{Y}$$

Intersection of Subspaces

If $\mathcal{X}$ and $\mathcal{Y}$ are subspaces of a vector space $\mathcal{V}$, then the *intersection* of $\mathcal{X}$ and $\mathcal{Y}$ is defined:

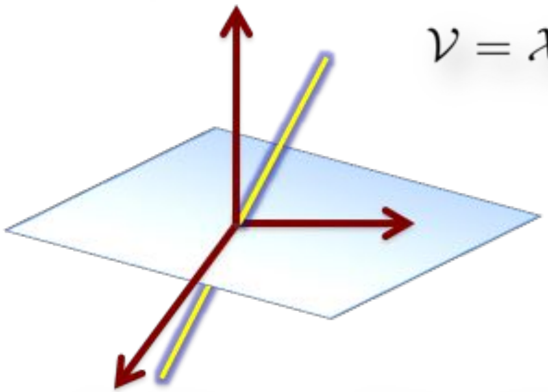
$$\mathcal{X} \cap \mathcal{Y} = \{z : z \in \mathcal{X} \ \text{ and } \ z \in \mathcal{Y}\}$$

The intersection of $\mathcal{X}$ and $\mathcal{Y}$ is a subspace of $\mathcal{V}$



Complementary Subspaces

Subspaces $\mathcal{X}$ and $\mathcal{Y}$ a space $\mathcal{V}$ are *complementary* if



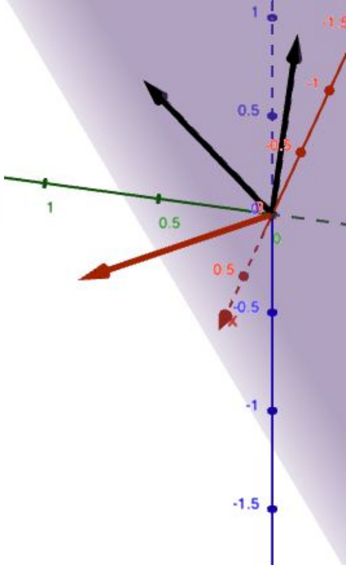
$$\mathcal{V} = \mathcal{X} + \mathcal{Y} \text{ and } \mathcal{X} \cap \mathcal{Y} = \emptyset$$

In this case, $\mathcal{V}$ is said to be the *direct sum* of $\mathcal{X}$ and $\mathcal{Y}$

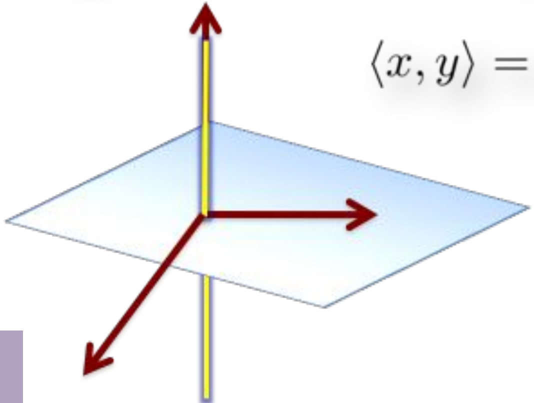
$$\mathcal{V} = \mathcal{X} \bigoplus \mathcal{Y}$$

For each $v \in \mathcal{V}$, there are *unique* vectors $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $v = x + y$

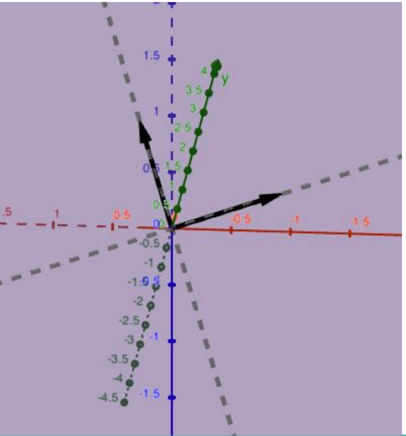
Orthogonal Subspaces



If $\mathcal{X}$ and $\mathcal{Y}$ are subspaces of an *inner-product* vector space $\mathcal{V}$, then $\mathcal{X}$ and $\mathcal{Y}$ are orthogonal $\mathcal{X} \perp \mathcal{Y}$ if



$$\langle x, y \rangle = 0 \text{ for all } x \in \mathcal{X} \text{ and } y \in \mathcal{Y}$$



If $\mathcal{X}$ is a subspace of a finite dimensional inner-product space $\mathcal{V}$, then $\mathcal{X}^{\perp}$ is its orthogonal complement if

$$\mathcal{V} = \mathcal{X} \bigoplus \mathcal{X}^{\perp}$$

Independent vs Disjoint

Linear subspaces $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_k$ of $\mathbb{R}^n$ are independent

if and only if

$$\dim(S_1 + S_2 + \ldots + S_k) = \dim(S_1) + \dim(S_2) + \ldots + \dim(S_k)$$

Linear subspaces $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_k$ of $\mathbb{R}^n$ are disjoint

if they intersect only at the origin

Independence is stronger than disjointedness

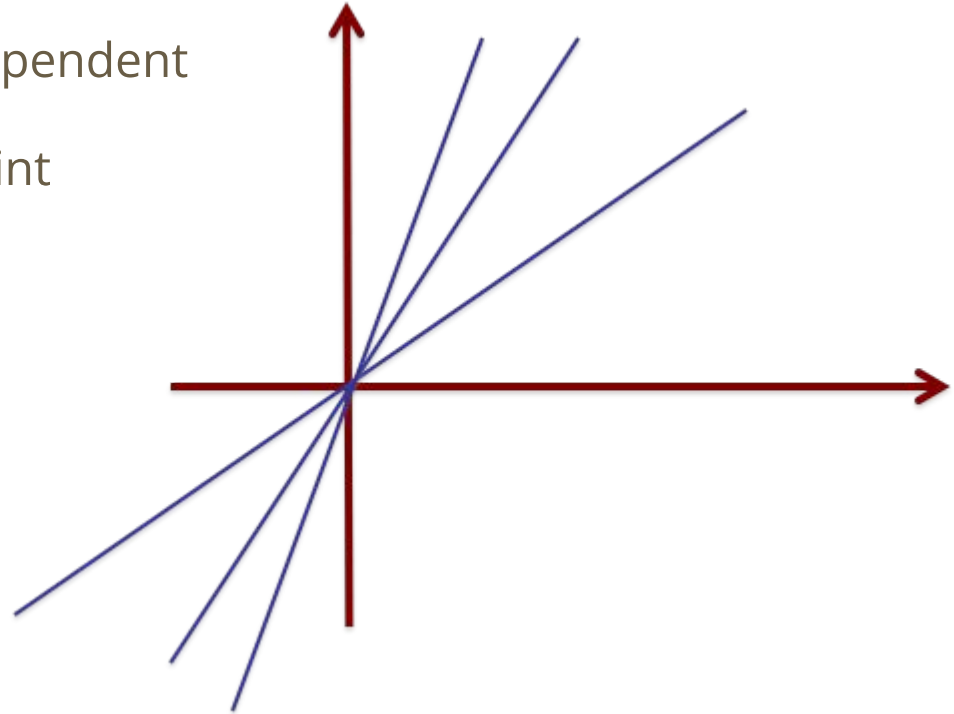
Three lines in $\mathbb{R}^2$ intersecting at 0 are disjoint

But they are not independent

Independent vs Disjoint

In R^2 any 2 lines are independent

3 or more lines are disjoint

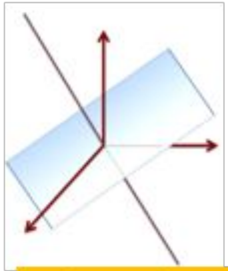


Minimal Angle

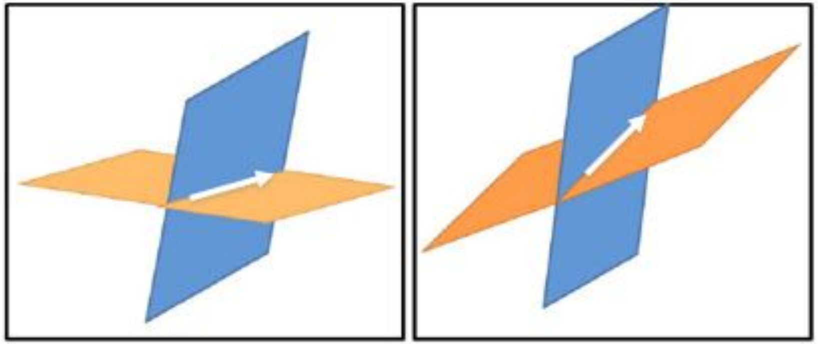
Minimum Angle

Let $\mathcal{F}$ and $\mathcal{G}$ be subspaces of $\mathbb{R}^D$. The minimal angle between $\mathcal{F}$ and $\mathcal{G}$ is defined as

$$\theta_{\min} = \arccos \left[\max_{\substack{f \in \mathcal{F} \\ g \in \mathcal{G} \\ ||f||_2 = ||g||_2 = 1}} f^T g \right]$$



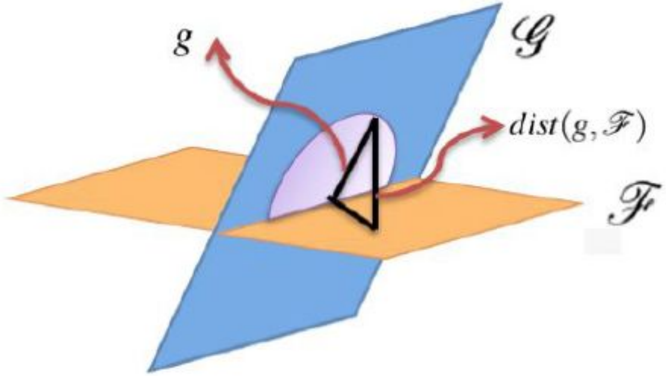
It is a good measure for complementary subspaces



It is not a good measure for non-complementary subspaces

Gap between Subspaces

$$d(\mathscr{F},\mathscr{G}) = \max_{\substack{g\in\mathscr{G}\\ ||g||_2=1}} dist(g,\mathscr{F}) = \max_{\substack{g\in\mathscr{G}\\ ||g||_2=1}} ||(I-P_{\mathscr{F}})g||_2$$



$$gap(\mathscr{F},\mathscr{G}) = \min(d(\mathscr{F},\mathscr{G}), d(\mathscr{G},\mathscr{F}))$$

Maximal Angle

Maximum Angle

The maximal angle between $\mathcal{F}$ and $\mathcal{G}$ is defined as

$$\theta_{\text{max}} = \arcsin(gap(\mathcal{F}, \mathcal{G})),$$

where $0 \leq \theta_{\text{max}} \leq \pi/2$.

It is useful for subspaces of equal dimension

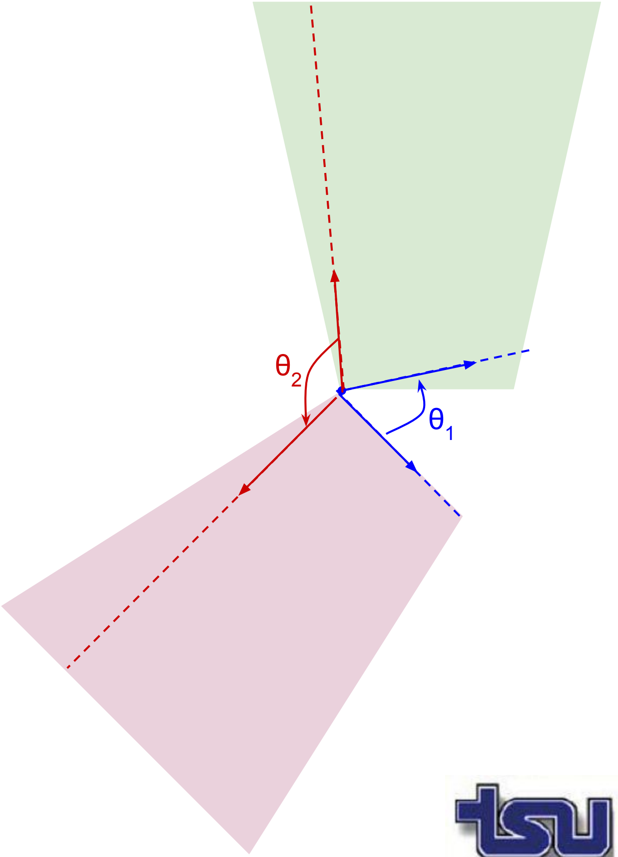
Principal Angles: An iterative process

Principle Angles

Let $\mathcal{F}$ and $\mathcal{G}$ be subspaces of $\mathbb{R}^D$. Let $k = \min(dim\mathcal{F}, dim\mathcal{G})$. Then, the principle angles $\theta_1, \theta_2, \ldots, \theta_k$ are the numbers $0 \leq \theta_i \leq \pi/2$ and they are defined as

$$\cos\theta_i = \max_{\substack{f\in\mathcal{F}_i \\ g\in\mathcal{G}_i \\ ||f||_2=||g||_2=1}} f^t g = f_i^t g_i \quad i = 1, \ldots, k$$

where $\mathcal{F}_1 = \mathcal{F}$ and $\mathcal{G}_1 = \mathcal{G}$, $||f_i||_2 = 1$, $||g_i||_2 = 1$, $\mathcal{F}_i = f_{i-1}^{\perp} \bigcap \mathcal{F}_{i-1}$, and $\mathcal{G}_i = g_{i-1}^{\perp} \bigcap \mathcal{G}_{i-1}$. Note that $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$.



Principal Angles

Principle Angles

Let $\mathcal{F}$ and $\mathcal{G}$ be subspaces of $\mathbb{R}^D$. Let $k = \min(dim\mathcal{F}, dim\mathcal{G})$. Then, the principle angles $\theta_1, \theta_2, \ldots, \theta_k$ are the numbers $0 \le \theta_i \le \pi/2$ and they are defined as

$$\cos\theta_i = \max_{\substack{f\in\mathcal{F}_i \\ g\in\mathcal{G}_i \\ f||_2=||g||_2=1}} f^t g = f_i^t g_i \quad i = 1, \ldots, k$$

where $\mathcal{F}_1 = \mathcal{F}$ and $\mathcal{G}_1 = \mathcal{G}$, $||f_i||_2 = 1$, $||g_i||_2 = 1$, $\mathcal{F}_i = f_{i-1}^{\perp} \bigcap \mathcal{F}_{i-1}$, and $\mathcal{G}_i = g_{i-1}^{\perp} \bigcap \mathcal{G}_{i-1}$. Note that $\theta_1 \le \theta_2 \le \cdots \le \theta_k$.

Time Capsule: note to the Future

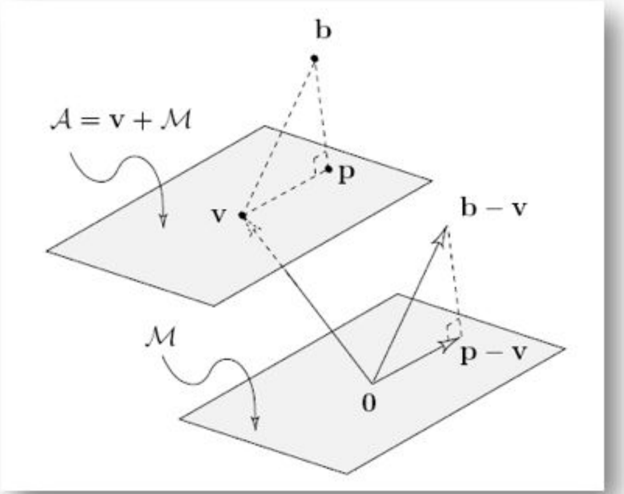
Given two subspaces $\mathscr{F}$ and $\mathscr{G}$ let $\mathbf{F}, \mathbf{G}$ form orthogonal bases for subspaces $\mathscr{F}$ and $\mathscr{G}$ respectively.

$$cos\theta_i = S(\mathbf{F}^T\mathbf{G})$$

where $S(\cdot)$ correspond to the *singular values* of $\cdot$

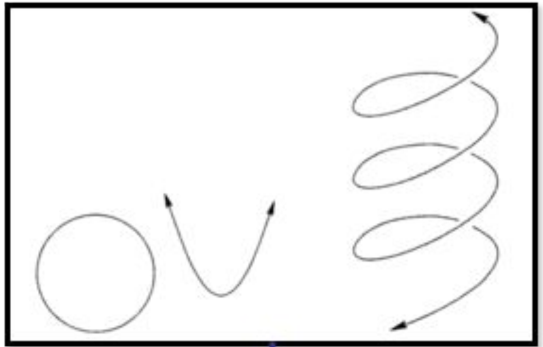
Affine Space

Affine Projections. If $\mathbf{v} \neq \mathbf{0}$ is a vector in a space $\mathcal{V}$, and if $\mathcal{M}$ is a subspace of $\mathcal{V}$, then the set of points $\mathcal{A} = \mathbf{v} + \mathcal{M}$ is called an ***affine space*** in $\mathcal{V}$.

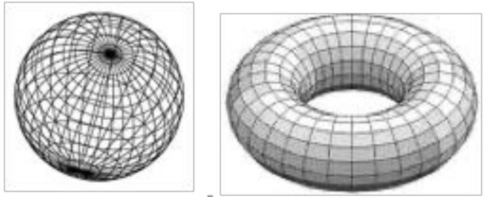


Manifolds

- A manifold is a mathematical space that (on a sufficiently small scale) resembles to the Euclidean space of a specific dimension



1-dimensional



2-dimensional

Manifolds

- Manifolds are like curves and surfaces, except that they might be of higher dimension
- Every manifold has a dimension
  - The number of independent parameters to specify a point
- *n* dimensional manifold is an object modeled *locally* on
  - It takes exactly *n* numbers to specify a point

Recall ME210: Parametric definition of curves and surfaces

to be continued...

With matrix decomposition methods