Linear Algebra

CS771: Introduction to Machine Learning

Purushottam Kar

Announcements

Deadline for submitting regrading requests for midsem: **Sept 30, 2019**

Assignment 2 was released last week – deadline **Oct 26, 9:59PM IST**

- *Note the differences from the previous assignment*
- *Submission includes a PDF file (Gradescope) and a ZIP file (website)*
  - ZIP file should be password protected (8-10 alphanumeric password – no special chars)
- *Allowed to freely use code downloaded from the internet*
  - Extreme Classification Repository: manikvarma.org/downloads/XC/XMLRepository.html
  - Must give credit to original author of code though – no penalty for using other's code but heavy penalty for using code from the internet without giving due credit
- *ZIP file must be self contained*
  - Must contain a file called *predict.py* containing a method *getReco* (don't change names)
  - Must also contain model and any non *pip* libraries required to run prediction code
  - No GPU access available for prediction (no restrictions on training though)

Recap of Last Lecture

Using latent variable modelling and the AltOpt/EM algorithms to solve the mixed regression problem

Using Generative Learning for Classification

Saw the simple case of binary classification – extendible to multiclass too

Model feature vectors of each class as a single Gaussian

May use multiple Gaussians too and model each class as a GMM instead

Special cases correspond to linear classifiers/LwP classifiers

Both Gaussians identical (standard or else spherical): linear classifier

Uniform prior class probabilities and standard Gaussians: LwP classifier

General case: quadratic classifier

Generative methods to deal with missing data/data reconstruction

Linear Algebra

The study of vectors and certain types of operations on vectors

Eventual goal: build generative models that are smaller and faster

As was the case with calculus as well as probability theory

- *Do try to get intuition (geometric intuition in this case) for toy cases*
- *Use this to convince yourself that the operations/results do make sense*
- *Difficult to get geometric intuition in $d > 3$ – intuition may even be wrong!*
  - Curse of dimensionality – a collection of unintuitive things that happen in large dims
- *Thus, for high dimensional cases, trust the math*

Recall: several (equally correct) ways of describing vectors. Vectors are

- ***Physics***: *things that look like arrows and have a magnitude and a direction*
- ***Math***: *things that can be added together, or scaled by multiplying a scalar*
- ***CS/ML***: *a list of real numbers/integers that describe features/quantities*

Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

- *Scale individual vectors using the scalars, for example* $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$
- *Add/subtract the two vectors e.g.* $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$
- *Add/subtract the scaled vectors e.g.* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

- *Name comes from the word "line"*
- *If we fix* $\mathbf{a}$ *and vary* p*, vectors* $p \cdot \mathbf{a}$ *lie on a line*
- *Fix* $\mathbf{a}, \mathbf{b}, q$ *and vary* p*, then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*
- *Don't read too much into the name though*
- *If we vary both* p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*
- *In math in general, the word "linear" loses its connection to "line" quickly*

Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

- *Scale individual vectors using the scalars, for example* $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$
- *Add/subtract the two vectors e.g.* $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$
- *Add/subtract the scaled vectors e.g.* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

- *Name comes from the word "line"*
- *If we fix* $\mathbf{a}$ *and vary* p*, vectors* $p \cdot \mathbf{a}$ *lie on a line*
- *Fix* $\mathbf{a}, \mathbf{b}, q$ *and vary* p*, then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*
- *Don't read too much into the name though*
- *If we vary both* p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*
- *In math in general, the word "linear" loses its connection to "line" quickly*

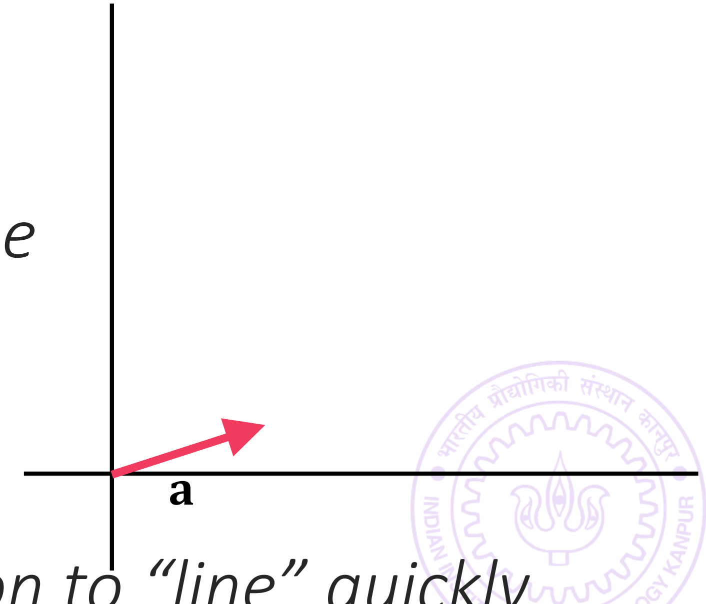
Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

- *Scale individual vectors using the scalars, for example* $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$
- *Add/subtract the two vectors e.g.* $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$
- *Add/subtract the scaled vectors e.g.* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

- *Name comes from the word "line"*
- *If we fix* $\mathbf{a}$ *and vary* p*, vectors* $p \cdot \mathbf{a}$ *lie on a line*
- *Fix* $\mathbf{a}, \mathbf{b}, q$ *and vary* p*, then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*
- *Don't read too much into the name though*
- *If we vary both* p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*
- *In math in general, the word "linear" loses its connection to "line" quickly*



Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

Scale individual vectors using the scalars, for example $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$

Add/subtract the two vectors e.g. $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$

Add/subtract the scaled vectors e.g. $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

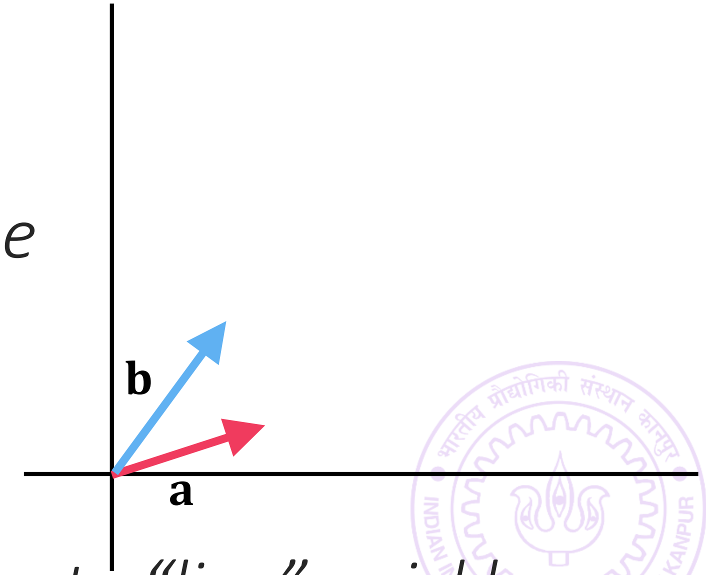
Name comes from the word "line"

If we fix $\mathbf{a}$ *and vary* p*, vectors* $p \cdot \mathbf{a}$ *lie on a line*

Fix $\mathbf{a}, \mathbf{b}, q$ *and vary* p*, then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*

Don't read too much into the name though

If we vary both p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*



In math in general, the word "linear" loses its connection to "line" quickly

Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

Scale individual vectors using the scalars, for example $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$

Add/subtract the two vectors e.g. $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$

Add/subtract the scaled vectors e.g. $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

Name comes from the word "line"

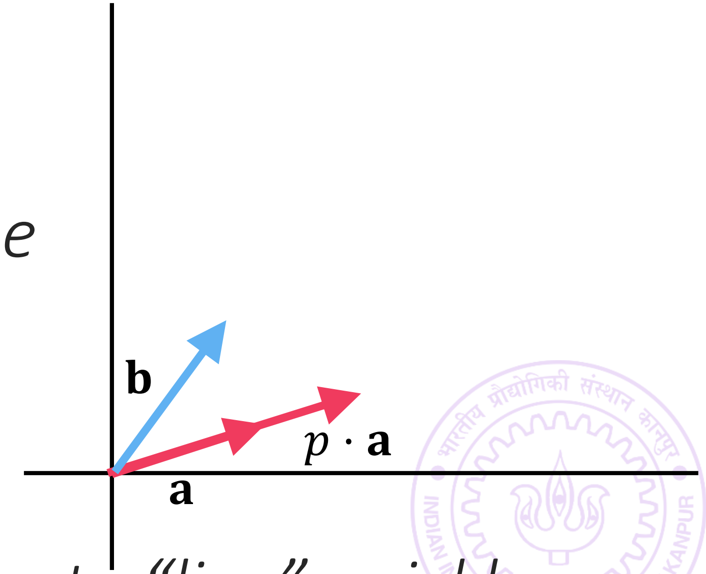
If we fix $\mathbf{a}$ *and vary* p, *vectors* $p \cdot \mathbf{a}$ *lie on a line*

Fix $\mathbf{a}, \mathbf{b}, q$ *and vary* p, *then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*

Don't read too much into the name though

If we vary both p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*

In math in general, the word "linear" loses its connection to "line" quickly



Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

Scale individual vectors using the scalars, for example $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$

Add/subtract the two vectors e.g. $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$

Add/subtract the scaled vectors e.g. $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

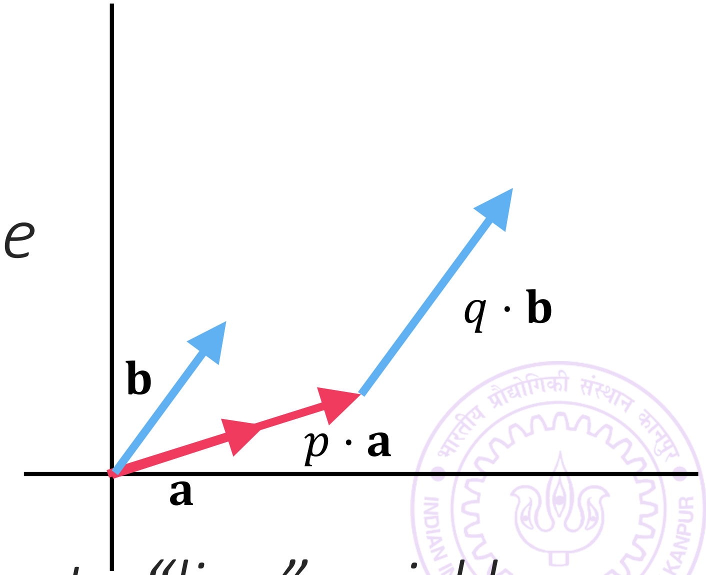
Name comes from the word "line"

If we fix $\mathbf{a}$ *and vary* p, *vectors* $p \cdot \mathbf{a}$ *lie on a line*

Fix $\mathbf{a}, \mathbf{b}, q$ *and vary* p, *then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*

Don't read too much into the name though

If we vary both p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*



In math in general, the word "linear" loses its connection to "line" quickly

Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

Scale individual vectors using the scalars, for example $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$

Add/subtract the two vectors e.g. $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$

Add/subtract the scaled vectors e.g. $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

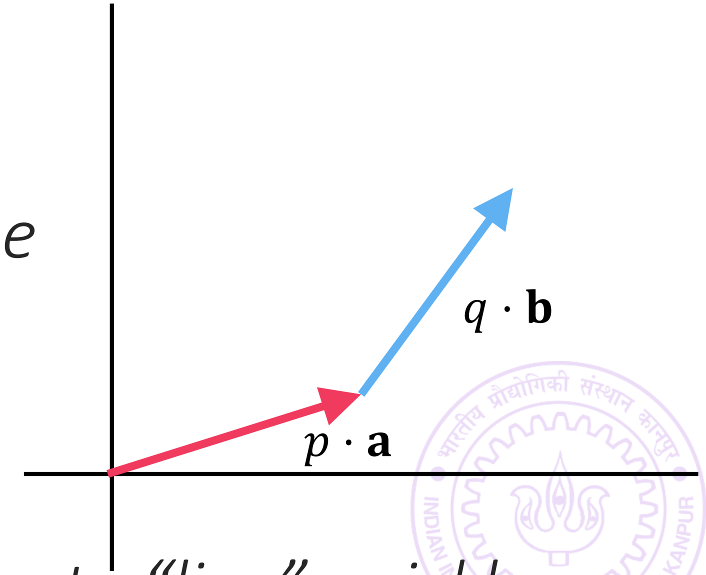
Name comes from the word "line"

If we fix $\mathbf{a}$ *and vary* p*, vectors* $p \cdot \mathbf{a}$ *lie on a line*

Fix $\mathbf{a}, \mathbf{b}, q$ *and vary* p*, then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*

Don't read too much into the name though

If we vary both p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*



In math in general, the word "linear" loses its connection to "line" quickly

Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

Scale individual vectors using the scalars, for example $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$

Add/subtract the two vectors e.g. $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$

Add/subtract the scaled vectors e.g. $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

Name comes from the word "line"

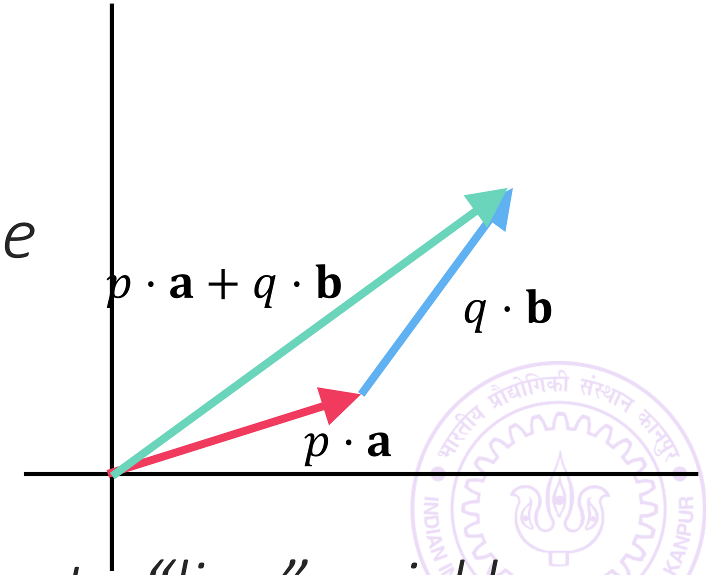
If we fix $\mathbf{a}$ *and vary* p*, vectors* $p \cdot \mathbf{a}$ *lie on a line*

Fix $\mathbf{a}, \mathbf{b}, q$ *and vary* p*, then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*

Don't read too much into the name though

If we vary both p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*

In math in general, the word "linear" loses its connection to "line" quickly



Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

Scale individual vectors using the scalars, for example $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$

Add/subtract the two vectors e.g. $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$

Add/subtract the scaled vectors e.g. $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

Name comes from the word "line"

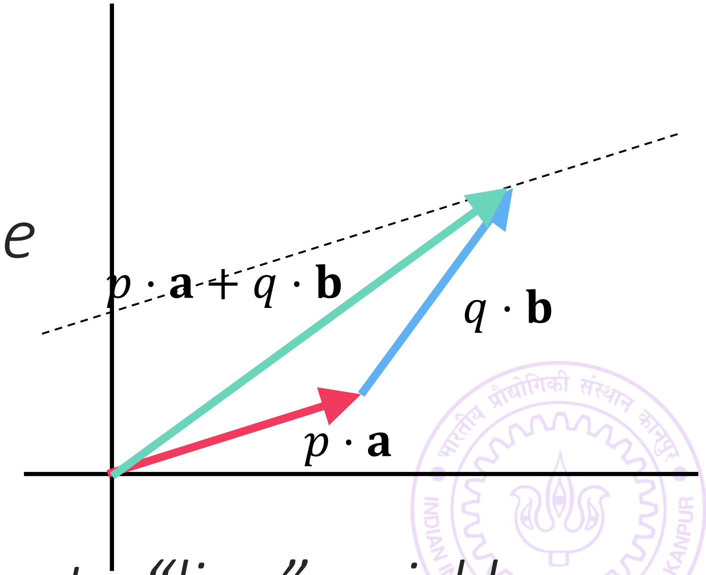
If we fix $\mathbf{a}$ *and vary* p*, vectors* $p \cdot \mathbf{a}$ *lie on a line*

Fix $\mathbf{a}, \mathbf{b}, q$ *and vary* p*, then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*

Don't read too much into the name though

If we vary both p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*

In math in general, the word "linear" loses its connection to "line" quickly



Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

Scale individual vectors using the scalars, for example $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$

Add/subtract the two vectors e.g. $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$

Add/subtract the scaled vectors e.g. $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

Name comes from the word "line"

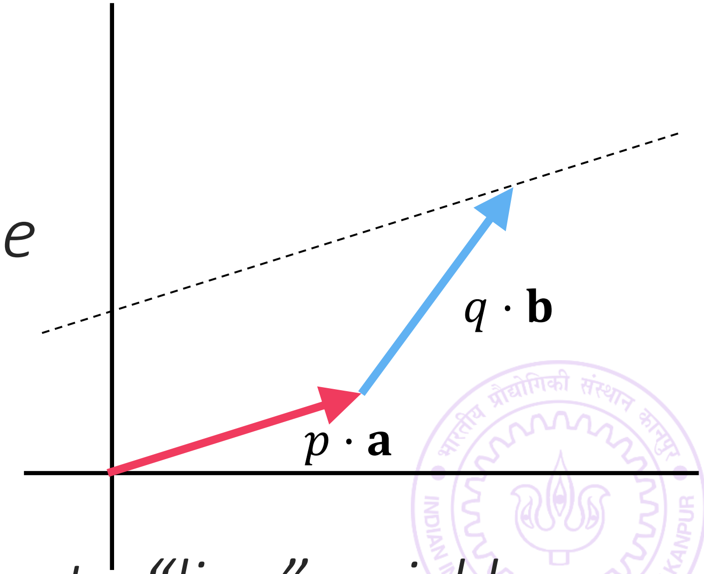
If we fix $\mathbf{a}$ *and vary* p*, vectors* $p \cdot \mathbf{a}$ *lie on a line*

Fix $\mathbf{a}, \mathbf{b}, q$ *and vary* p*, then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*

Don't read too much into the name though

If we vary both p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*

In math in general, the word "linear" loses its connection to "line" quickly



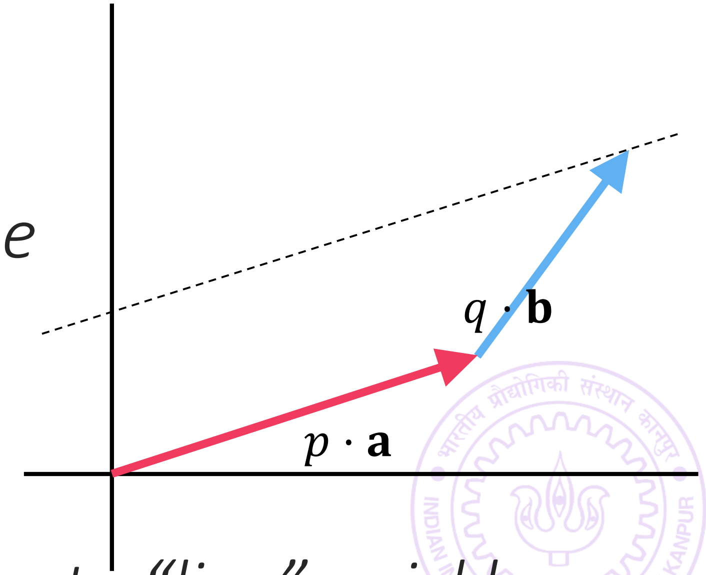
Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

- *Scale individual vectors using the scalars, for example* $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$
- *Add/subtract the two vectors e.g.* $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$
- *Add/subtract the scaled vectors e.g.* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

- *Name comes from the word "line"*
- *If we fix* $\mathbf{a}$ *and vary* p*, vectors* $p \cdot \mathbf{a}$ *lie on a line*
- *Fix* $\mathbf{a}, \mathbf{b}, q$ *and vary* p*, then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*
- *Don't read too much into the name though*
- *If we vary both* p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*
- *In math in general, the word "linear" loses its connection to "line" quickly*



Linear Combinations of Vectors

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$, we can

Scale individual vectors using the scalars, for example $p \cdot \mathbf{a}$ *or* $q \cdot \mathbf{b}$

Add/subtract the two vectors e.g. $\mathbf{a} + \mathbf{b}$ *or* $\mathbf{a} - \mathbf{b}$

Add/subtract the scaled vectors e.g. $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *or* $q \cdot \mathbf{a} - p \cdot \mathbf{b}$

These operations are called *linear* operations on vectors

Name comes from the word "line"

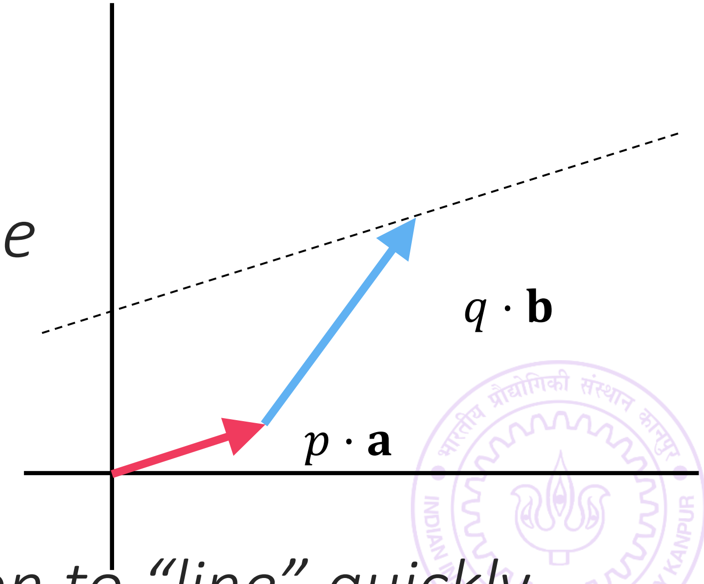
If we fix $\mathbf{a}$ *and vary* p*, vectors* $p \cdot \mathbf{a}$ *lie on a line*

Fix $\mathbf{a}, \mathbf{b}, q$ *and vary* p*, then* $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ *will line on a line*

Don't read too much into the name though

If we vary both p *and* q *then resulting vectors need no longer lie on a line (but they will lie on a plane)*

In math in general, the word "linear" loses its connection to "line" quickly



Special kinds of linear combinations

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$

Convex combinations: we insist $p, q \geq 0$ and $p + q = 1$.
Set of all convex combinations of two vectors is simply the line segment joining them

Affine combinations: we only insist that $p + q = 1$.
The set of all affine combinations of two vectors is the entire line (not segment) passing through them

Linear combinations: no such restrictions on p, q.
The set of all linear combinations of two vectors is called the *span* of those two vectors

Special case: when $\exists c \in \mathbb{R}$ s.t. $\mathbf{b} = c \cdot \mathbf{a}$ (possibly $c < 0$). Can you find out what happens in this case??

Special kinds of linear combinations

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$

Convex combinations: we insist $p, q \geq 0$ and $p + q = 1$.
Set of all convex combinations of two vectors is simply the line segment joining them

Affine combinations: we only insist that $p + q = 1$.
The set of all affine combinations of two vectors is the entire line (not segment) passing through them

Linear combinations: no such restrictions on p, q.
The set of all linear combinations of two vectors is called the *span* of those two vectors

Special case: when $\exists c \in \mathbb{R}$ s.t. $\mathbf{b} = c \cdot \mathbf{a}$ (possibly $c < 0$). Can you find out what happens in this case??

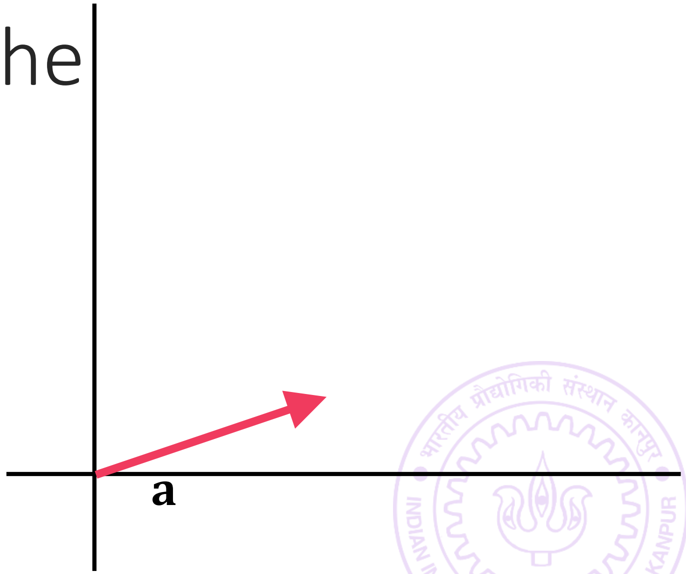
Special kinds of linear combinations

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$

Convex combinations: we insist $p, q \geq 0$ and $p + q = 1$. Set of all convex combinations of two vectors is simply the line segment joining them

Affine combinations: we only insist that $p + q = 1$. The set of all affine combinations of two vectors is the entire line (not segment) passing through them

Linear combinations: no such restrictions on p, q. The set of all linear combinations of two vectors is called the *span* of those two vectors



Special case: when $\exists c \in \mathbb{R}$ s.t. $\mathbf{b} = c \cdot \mathbf{a}$ (possibly $c < 0$). Can you find out what happens in this case??

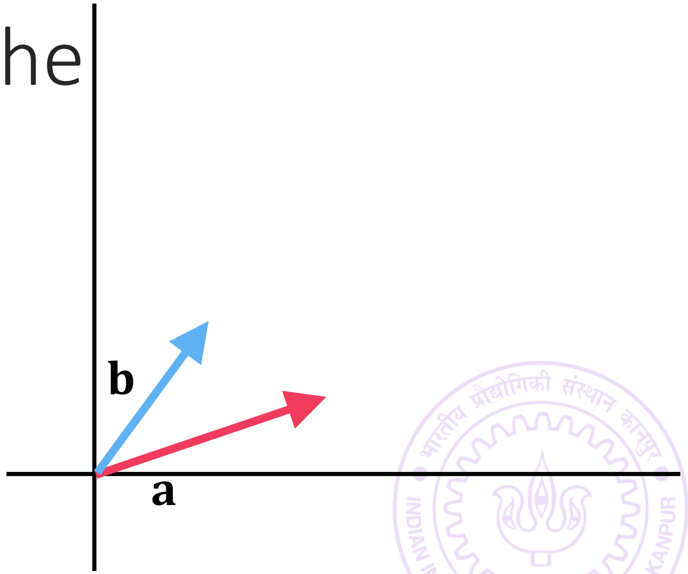
Special kinds of linear combinations

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$

Convex combinations: we insist $p, q \geq 0$ and $p + q = 1$. Set of all convex combinations of two vectors is simply the line segment joining them

Affine combinations: we only insist that $p + q = 1$. The set of all affine combinations of two vectors is the entire line (not segment) passing through them

Linear combinations: no such restrictions on p, q. The set of all linear combinations of two vectors is called the *span* of those two vectors



Special case: when $\exists c \in \mathbb{R}$ s.t. $\mathbf{b} = c \cdot \mathbf{a}$ (possibly $c < 0$). Can you find out what happens in this case??

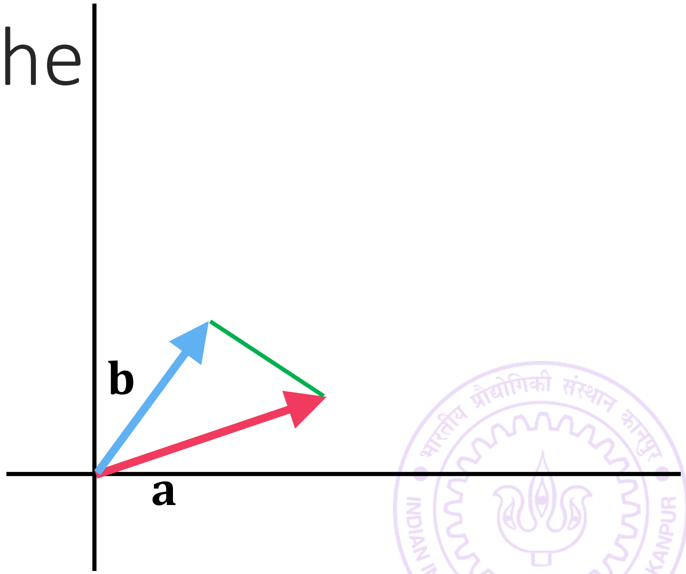
Special kinds of linear combinations

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$

Convex combinations: we insist $p, q \geq 0$ and $p + q = 1$. Set of all convex combinations of two vectors is simply the line segment joining them

Affine combinations: we only insist that $p + q = 1$. The set of all affine combinations of two vectors is the entire line (not segment) passing through them

Linear combinations: no such restrictions on p, q. The set of all linear combinations of two vectors is called the *span* of those two vectors



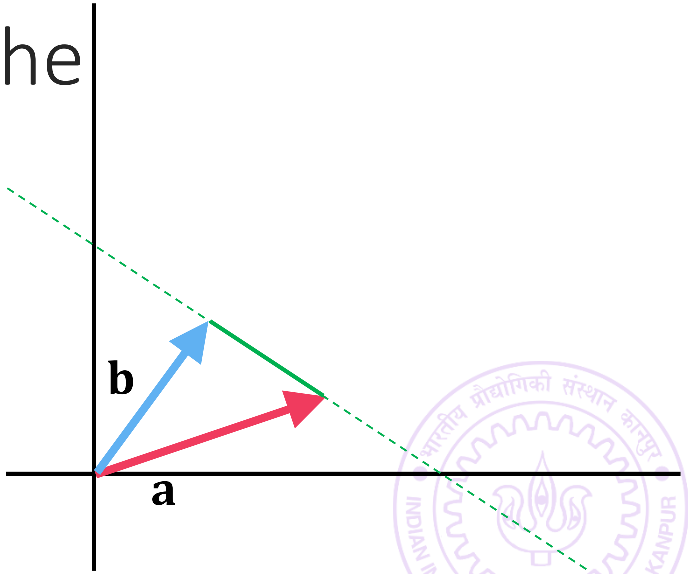
Special case: when $\exists c \in \mathbb{R}$ s.t. $\mathbf{b} = c \cdot \mathbf{a}$ (possibly $c < 0$). Can you find out what happens in this case??

Special kinds of linear combinations

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$

Convex combinations: we insist $p, q \geq 0$ and $p + q = 1$. Set of all convex combinations of two vectors is simply the line segment joining them

Affine combinations: we only insist that $p + q = 1$. The set of all affine combinations of two vectors is the entire line (not segment) passing through them



Linear combinations: no such restrictions on p, q. The set of all linear combinations of two vectors is called the *span* of those two vectors

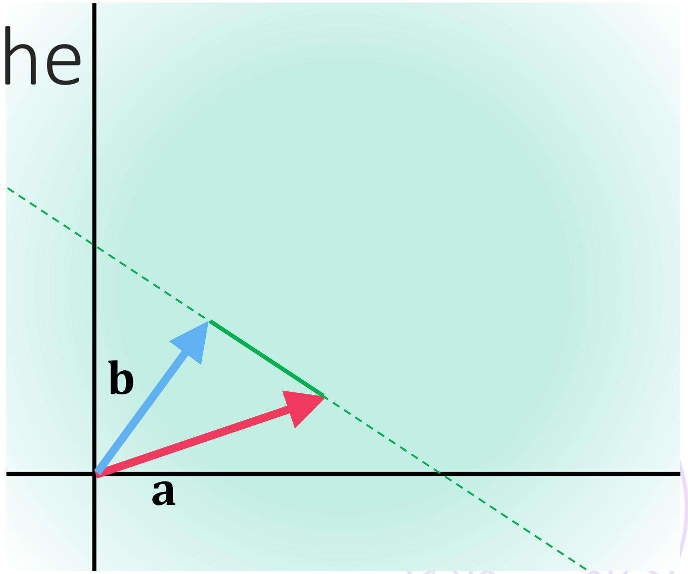
Special case: when $\exists c \in \mathbb{R}$ s.t. $\mathbf{b} = c \cdot \mathbf{a}$ (possibly $c < 0$). Can you find out what happens in this case??

Special kinds of linear combinations

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$

Convex combinations: we insist $p, q \geq 0$ and $p + q = 1$.
Set of all convex combinations of two vectors is simply the line segment joining them

Affine combinations: we only insist that $p + q = 1$.
The set of all affine combinations of two vectors is the entire line (not segment) passing through them



Linear combinations: no such restrictions on p, q.
The set of all linear combinations of two vectors is called the *span* of those two vectors

Special case: when $\exists c \in \mathbb{R}$ s.t. $\mathbf{b} = c \cdot \mathbf{a}$ (possibly $c < 0$). Can you find out what happens in this case??

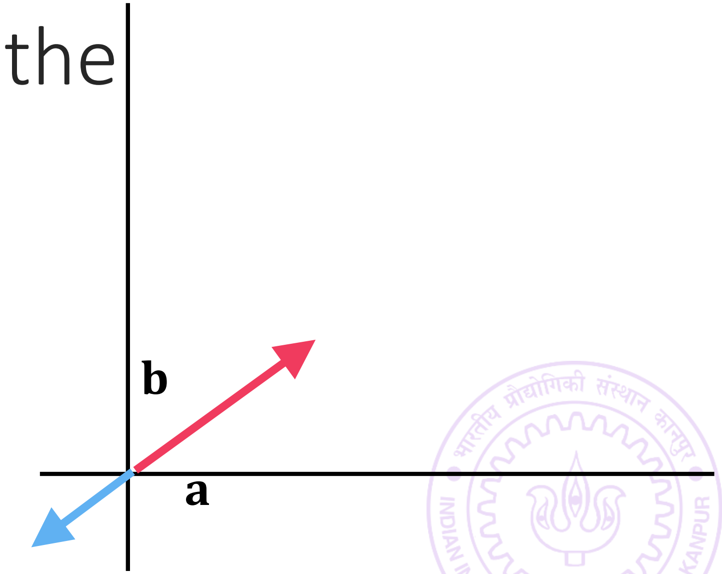
Special kinds of linear combinations

Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and scalars $p, q \in \mathbb{R}$

Convex combinations: we insist $p, q \geq 0$ and $p + q = 1$.
Set of all convex combinations of two vectors is simply the line segment joining them

Affine combinations: we only insist that $p + q = 1$.
The set of all affine combinations of two vectors is the entire line (not segment) passing through them

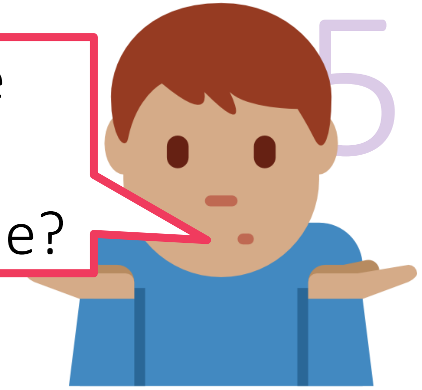
Linear combinations: no such restrictions on p, q.
The set of all linear combinations of two vectors is called the *span* of those two vectors



Special case: when $\exists c \in \mathbb{R}$ s.t. $\mathbf{b} = c \cdot \mathbf{a}$ (possibly $c < 0$). Can you find out what happens in this case??

Special kin

We can convince ourselves why the convex and affine combinations are respectively the line and the line segment but why should the span be the entire $\mathbb{R}^2$ plane?

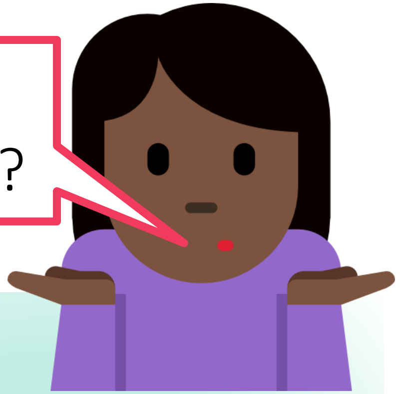


Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}$ and scalars $p, q \in \mathbb{R}$

Convex combinations: we insist $p, q \geq 0$ and $p + q = 1$.
Set of all convex comb ns of t
t n

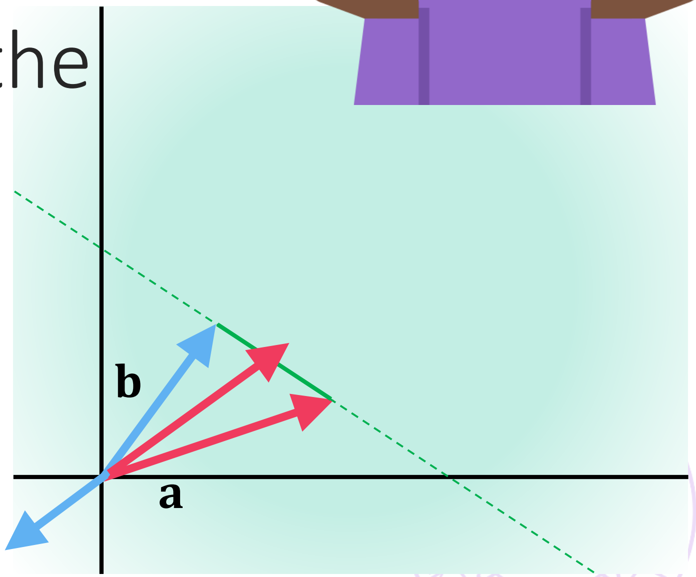
Let us study these in more detail to find out

What if we have more than two points. How are these defined then?



Affine combinations: we insist that $p + q = 1$.
The set of all affine combinations of two vectors is the entire line (not segment) passing through them

Linear combinations: no such restrictions on p, q.
The set of all linear combinations of two vectors is called the *span* of those two vectors



Special case: when $\exists c \in \mathbb{R}$ s.t. $\mathbf{b} = c \cdot \mathbf{a}$ (possibly $c < 0$). Can you find out what happens in this case??

Span of Vectors

Span of $\mathbf{a}, \mathbf{b}$: the set of all vectors that can be obtained by scaling the two vectors using (possibly negative) scalars p, q and adding them

Trick: First set $q = 0$ and see what happens when we vary only p

$p \cdot \mathbf{a}$ for all values of p gives us all vectors on the line along $\mathbf{a}$

Now let see add $\mathbf{b}$ and see the set of possible vectors of the form $p \cdot \mathbf{a} + \mathbf{b}$

This means that the set of all vectors of the form $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ where q is varied as well can be got simply by sweeping this line

We can convince ourselves that every point in $\mathbb{R}^2$ can be reached this way which is why span is $\mathbb{R}^2$

Span of Vectors

Span of $\mathbf{a}, \mathbf{b}$: the set of all vectors that can be obtained by scaling the two vectors using (possibly negative) scalars p, q and adding them

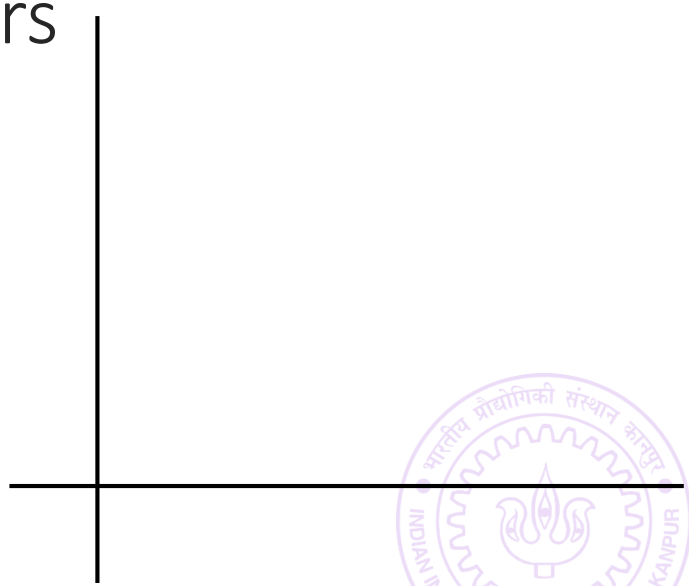
Trick: First set $q = 0$ and see what happens when we vary only p

$p \cdot \mathbf{a}$ for all values of p gives us all vectors on the line along $\mathbf{a}$

Now let see add $\mathbf{b}$ and see the set of possible vectors of the form $p \cdot \mathbf{a} + \mathbf{b}$

This means that the set of all vectors of the form $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ where q is varied as well can be got simply by sweeping this line

We can convince ourselves that every point in $\mathbb{R}^2$ can be reached this way which is why span is $\mathbb{R}^2$



Span of Vectors

Span of $\mathbf{a}, \mathbf{b}$: the set of all vectors that can be obtained by scaling the two vectors using (possibly negative) scalars p, q and adding them

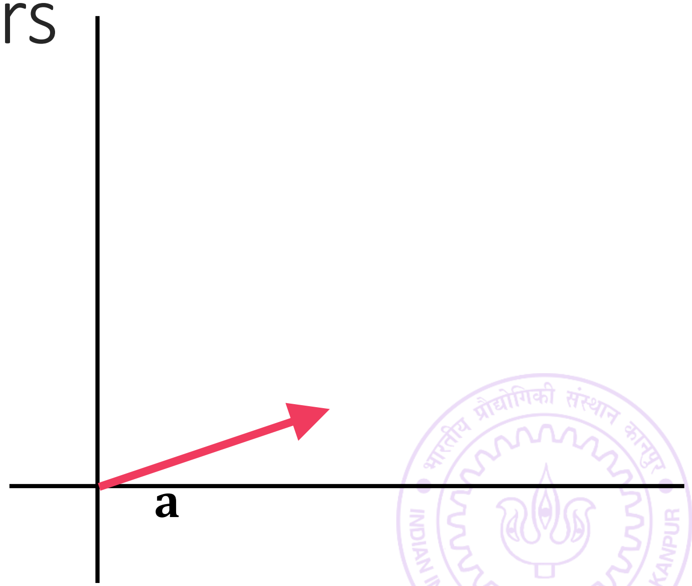
Trick: First set $q = 0$ and see what happens when we vary only p

$p \cdot \mathbf{a}$ for all values of p gives us all vectors on the line along $\mathbf{a}$

Now let see add $\mathbf{b}$ and see the set of possible vectors of the form $p \cdot \mathbf{a} + \mathbf{b}$

This means that the set of all vectors of the form $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ where q is varied as well can be got simply by sweeping this line

We can convince ourselves that every point in $\mathbb{R}^2$ can be reached this way which is why span is $\mathbb{R}^2$



Span of Vectors

Span of $\mathbf{a}, \mathbf{b}$: the set of all vectors that can be obtained by scaling the two vectors using (possibly negative) scalars p, q and adding them

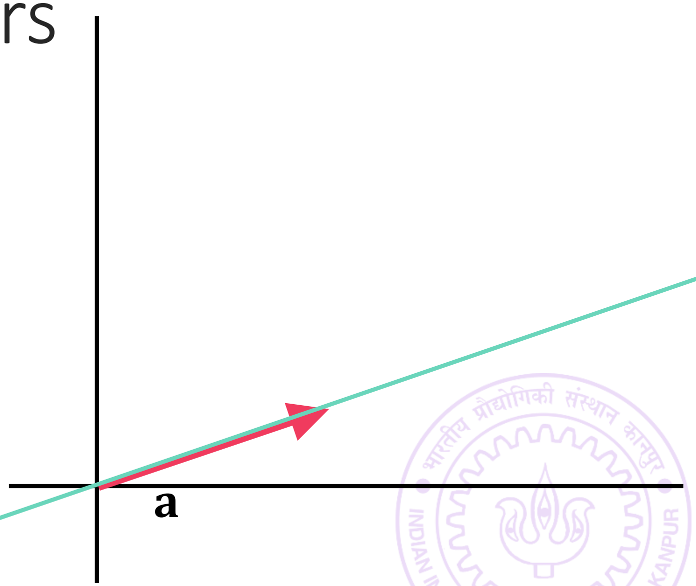
Trick: First set $q = 0$ and see what happens when we vary only p

$p \cdot \mathbf{a}$ for all values of p gives us all vectors on the line along $\mathbf{a}$

Now let see add $\mathbf{b}$ and see the set of possible vectors of the form $p \cdot \mathbf{a} + \mathbf{b}$

This means that the set of all vectors of the form $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ where q is varied as well can be got simply by sweeping this line

We can convince ourselves that every point in $\mathbb{R}^2$ can be reached this way which is why span is $\mathbb{R}^2$



Span of Vectors

Span of $\mathbf{a}, \mathbf{b}$: the set of all vectors that can be obtained by scaling the two vectors using (possibly negative) scalars p, q and adding them

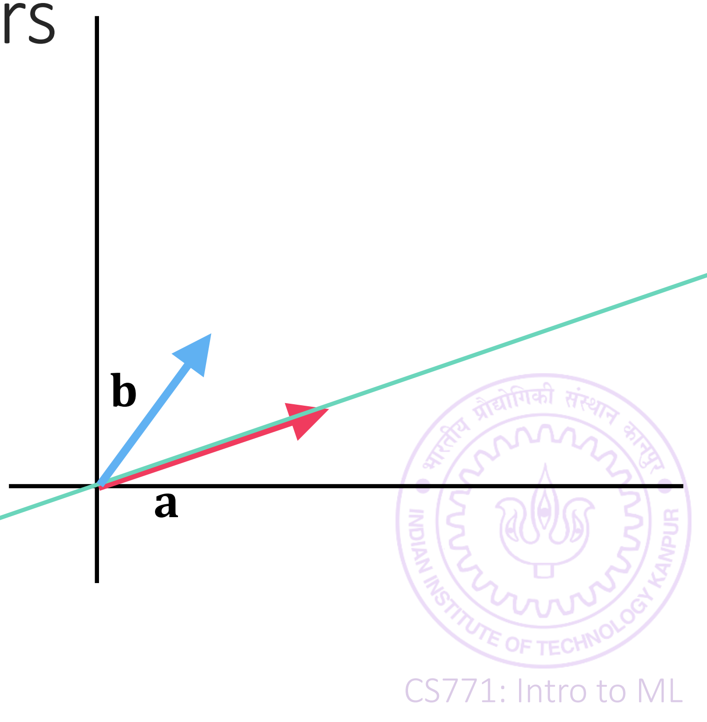
Trick: First set $q = 0$ and see what happens when we vary only p

$p \cdot \mathbf{a}$ for all values of p gives us all vectors on the line along $\mathbf{a}$

Now let see add $\mathbf{b}$ and see the set of possible vectors of the form $p \cdot \mathbf{a} + \mathbf{b}$

This means that the set of all vectors of the form $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ where q is varied as well can be got simply by sweeping this line

We can convince ourselves that every point in $\mathbb{R}^2$ can be reached this way which is why span is $\mathbb{R}^2$



Span of Vectors

Span of $\mathbf{a}, \mathbf{b}$: the set of all vectors that can be obtained by scaling the two vectors using (possibly negative) scalars p, q and adding them

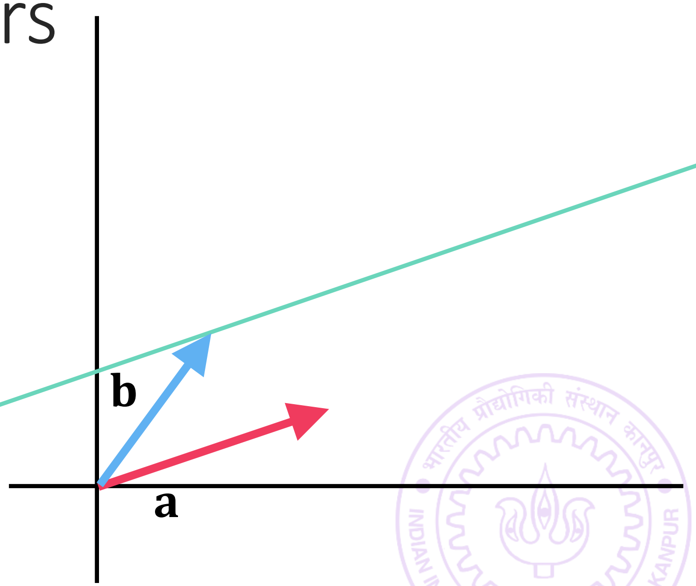
Trick: First set $q = 0$ and see what happens when we vary only p

$p \cdot \mathbf{a}$ for all values of p gives us all vectors on the line along $\mathbf{a}$

Now let see add $\mathbf{b}$ and see the set of possible vectors of the form $p \cdot \mathbf{a} + \mathbf{b}$

This means that the set of all vectors of the form $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ where q is varied as well can be got simply by sweeping this line

We can convince ourselves that every point in $\mathbb{R}^2$ can be reached this way which is why span is $\mathbb{R}^2$



Span of Vectors

Span of $\mathbf{a}, \mathbf{b}$: the set of all vectors that can be obtained by scaling the two vectors using (possibly negative) scalars p, q and adding them

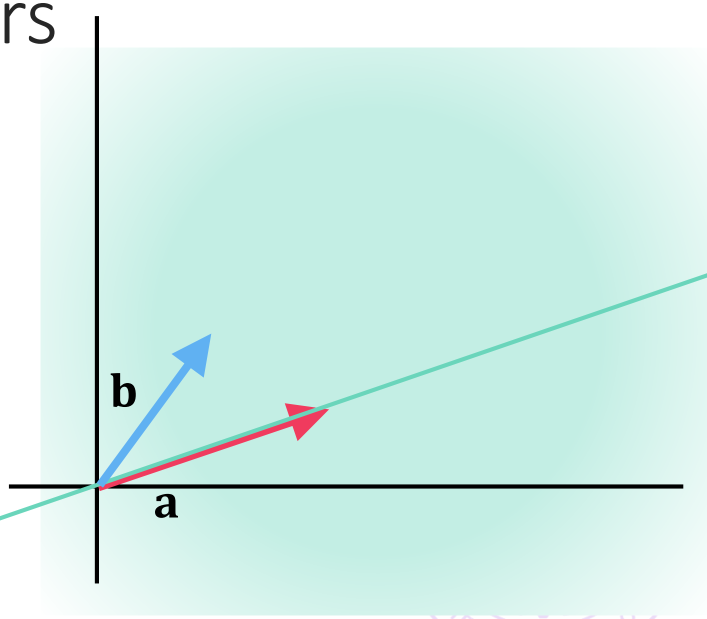
Trick: First set $q = 0$ and see what happens when we vary only p

$p \cdot \mathbf{a}$ for all values of p gives us all vectors on the line along $\mathbf{a}$

Now let see add $\mathbf{b}$ and see the set of possible vectors of the form $p \cdot \mathbf{a} + \mathbf{b}$

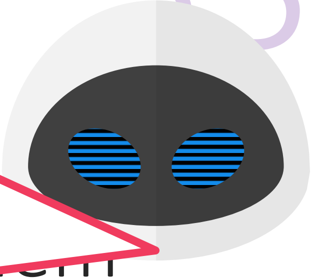
This means that the set of all vectors of the form $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ where q is varied as well can be got simply by sweeping this line

We can convince ourselves that every point in $\mathbb{R}^2$ can be reached this way which is why span is $\mathbb{R}^2$



Sp

The only time this procedure will not work is when $\mathbf{b} = c \cdot \mathbf{a}$ for some $c \in \mathbb{R}$. In that case, the vector $\mathbf{b}$ will lie on the line generated by $\mathbf{a}$ itself and adding multiples of $\mathbf{b}$ to that line will still give us that same line. To see this, note that for any p, q in this case we would have $p \cdot \mathbf{a} + q \cdot \mathbf{b} = (p + qc) \cdot \mathbf{a}$ which lies on the line along $\mathbf{a}$ itself! Thus, we are stuck to points on this line alone!!



Sp ... i
tw ... em

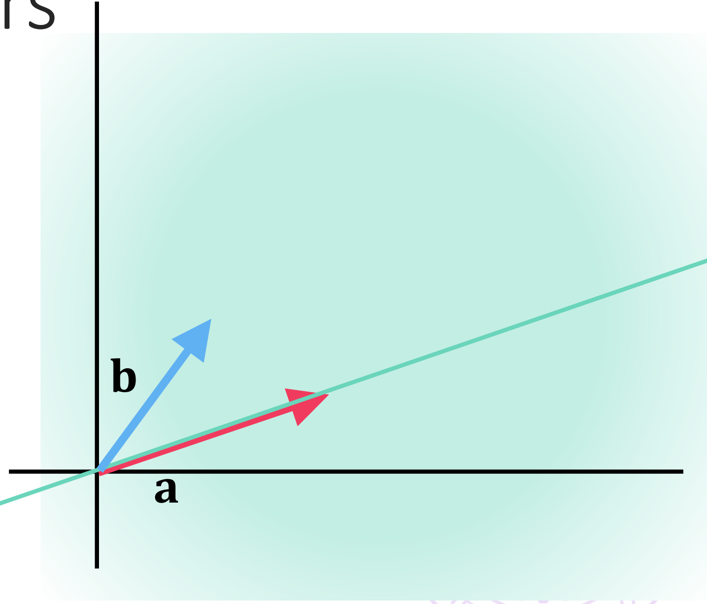
Trick: First set $q = 0$ and see what happens when we vary only p

$p \cdot \mathbf{a}$ for all values of p gives us all vectors on the line along $\mathbf{a}$

Now let see add $\mathbf{b}$ and see the set of possible vectors of the form $p \cdot \mathbf{a} + \mathbf{b}$

This means that the set of all vectors of the form $p \cdot \mathbf{a} + q \cdot \mathbf{b}$ where q is varied as well can be got simply by sweeping this line

We can convince ourselves that every point in $\mathbb{R}^2$ can be reached this way which is why span is $\mathbb{R}^2$



Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^n p_1 = 1$ and computing the point $\sum_{i=1}^n p_i \cdot \mathbf{x}^i$

The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

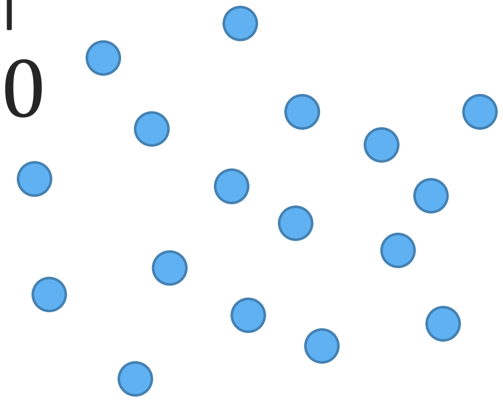
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \dots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \dots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^n p_1 = 1$ and computing the point $\sum_{i=1}^n p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

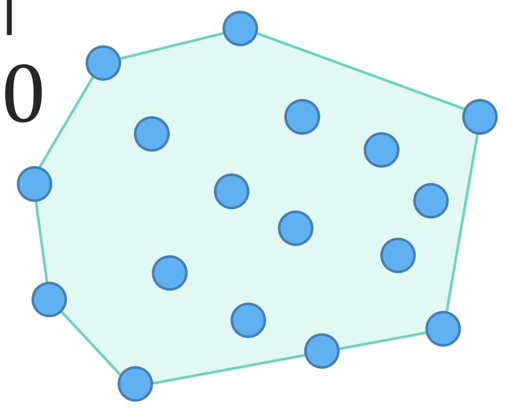
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^n p_1 = 1$ and computing the point $\sum_{i=1}^n p_i \cdot \mathbf{x}^i$



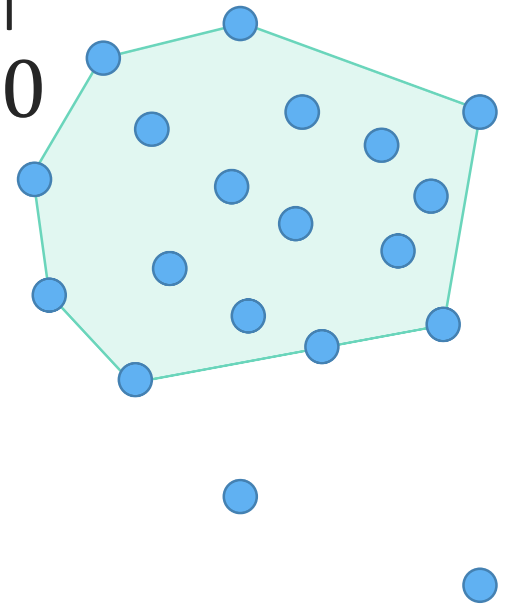
The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors



Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^n p_1 = 1$ and computing the point $\sum_{i=1}^n p_i \cdot \mathbf{x}^i$

The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

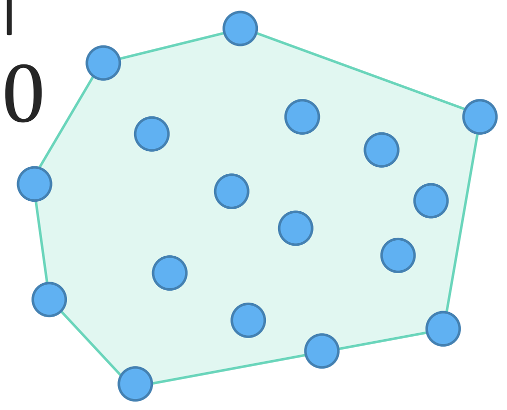
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

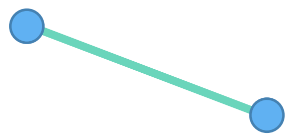
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^n p_1 = 1$ and computing the point $\sum_{i=1}^n p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points



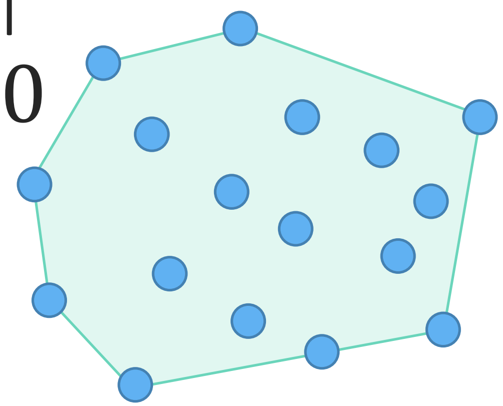
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

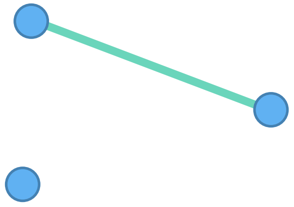
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points



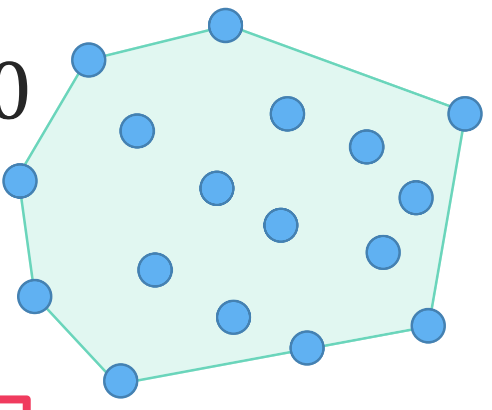
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

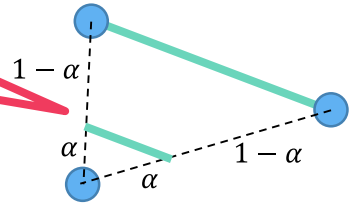
Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \dots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \dots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

Ratio of these lengths is $(1-\alpha):\alpha$



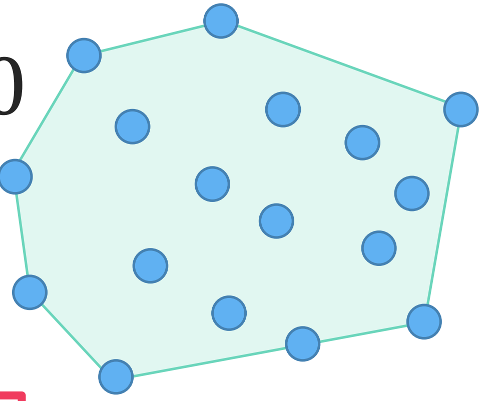
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

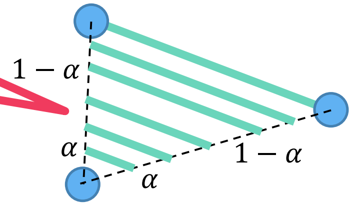
Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

Ratio of these lengths is $(1-\alpha):\alpha$



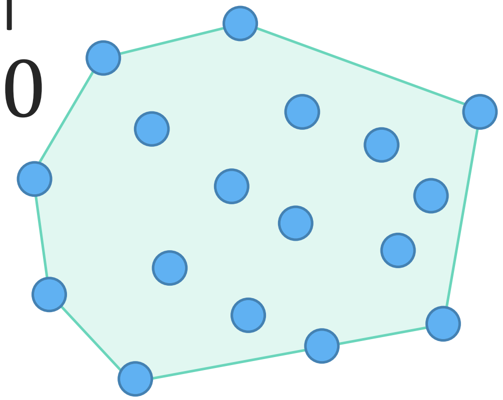
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

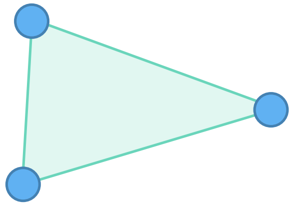
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^n p_1 = 1$ and computing the point $\sum_{i=1}^n p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points



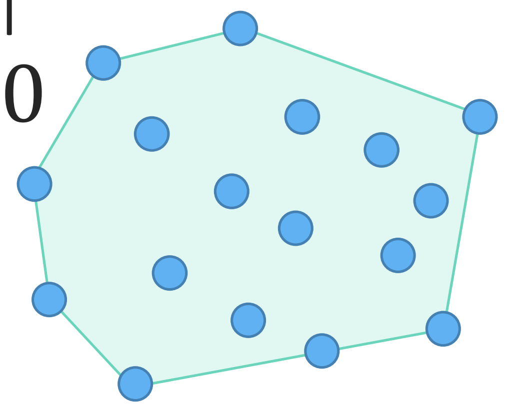
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

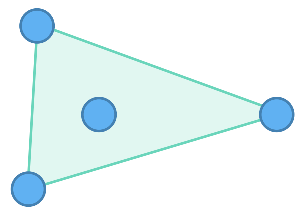
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points



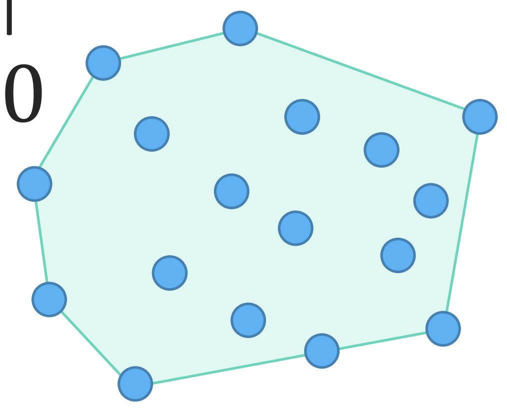
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

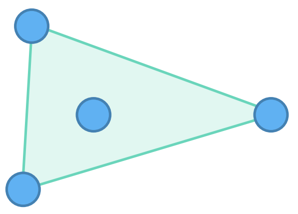
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points



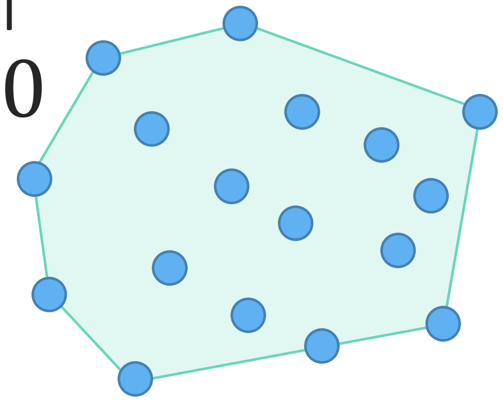
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

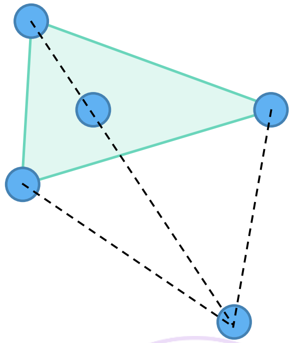
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \dots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \dots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points



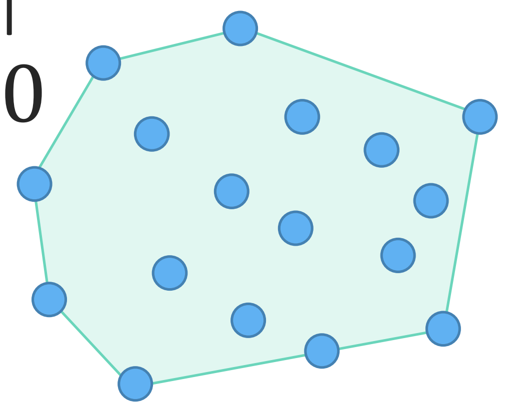
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

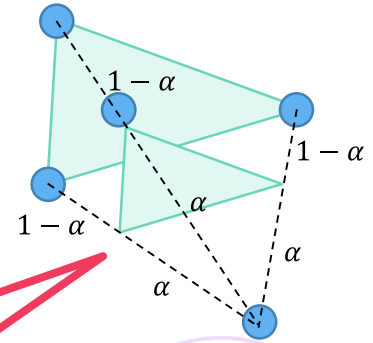
Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

Trick: just as before, start small – start with two points and look at their convex combinations (line segment)



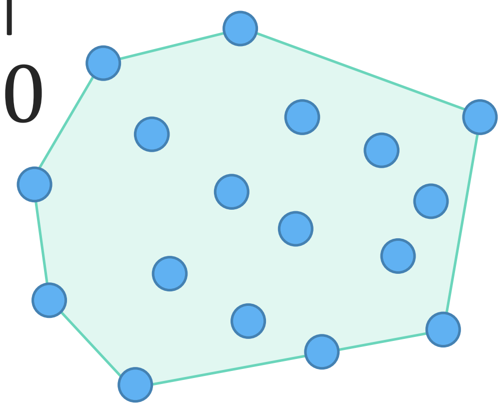
Then add a third point by taking some weig the previous two points i.e. $p_1 + p_2 = \alpha$ ar

Ratio of these lengths is $(1-\alpha){:}\,\alpha$

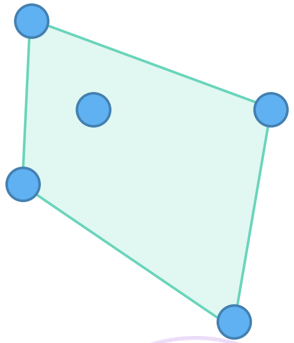
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

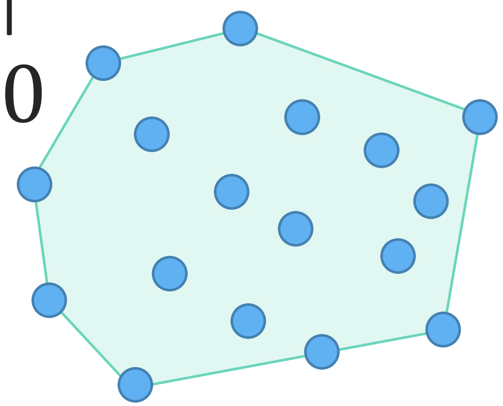


Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

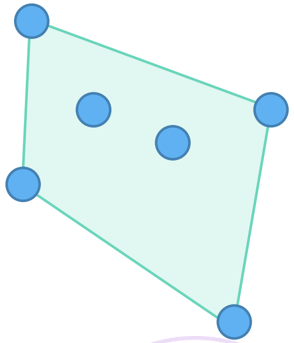
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors



Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^n p_1 = 1$ and computing the point $\sum_{i=1}^n p_i \cdot \mathbf{x}^i$

The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points



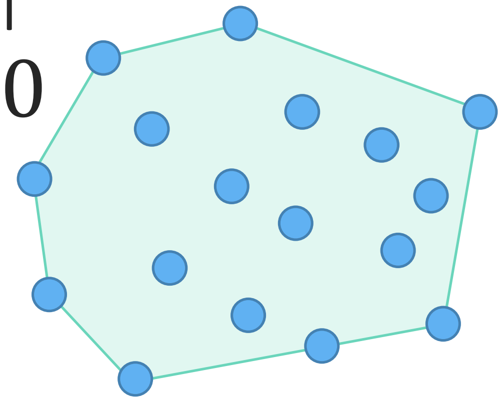
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

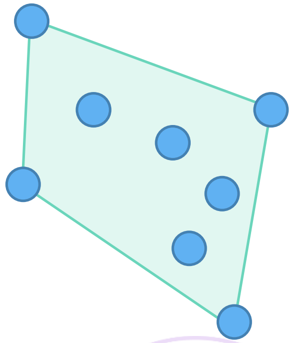
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

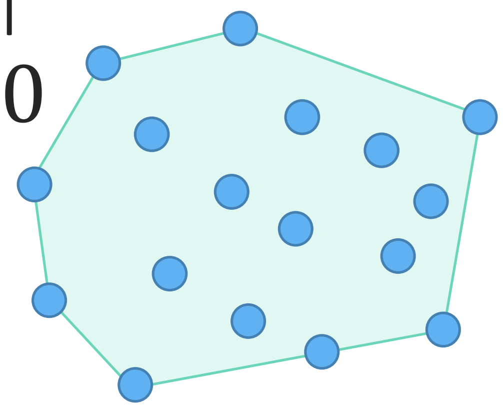


Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

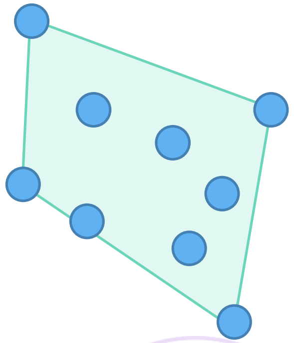
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors



Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$

The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points



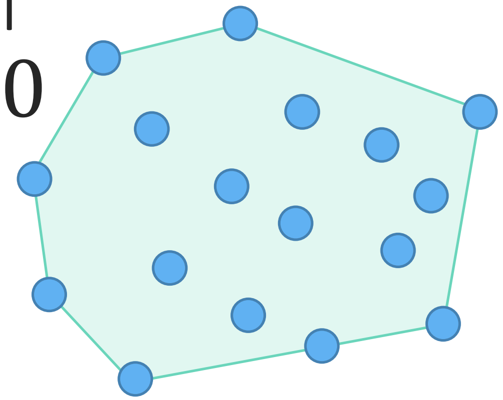
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

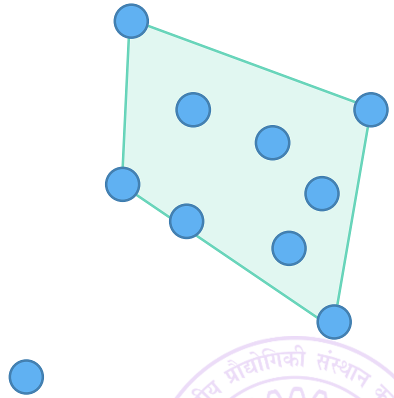
Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^n p_1 = 1$ and computing the point $\sum_{i=1}^n p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

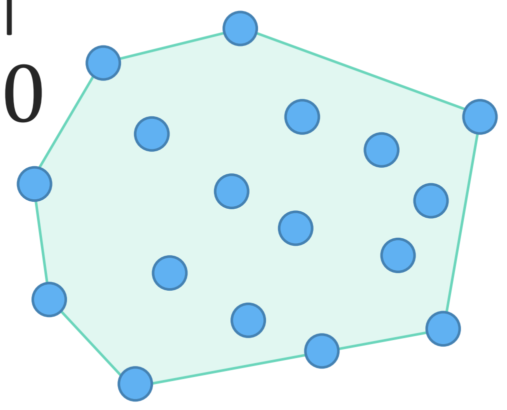


Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

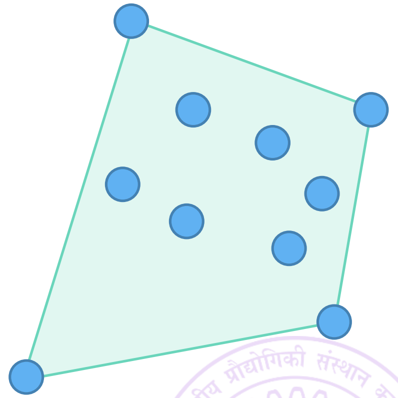
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^n p_1 = 1$ and computing the point $\sum_{i=1}^n p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points



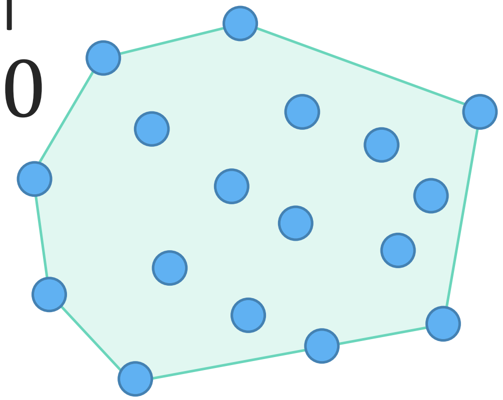
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

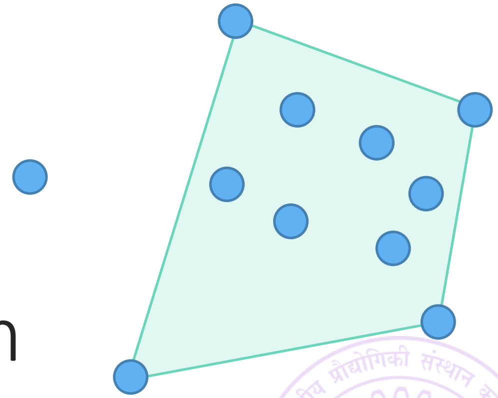
Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \dots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \dots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

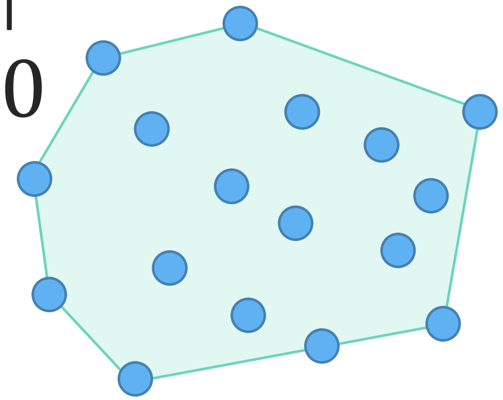


Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

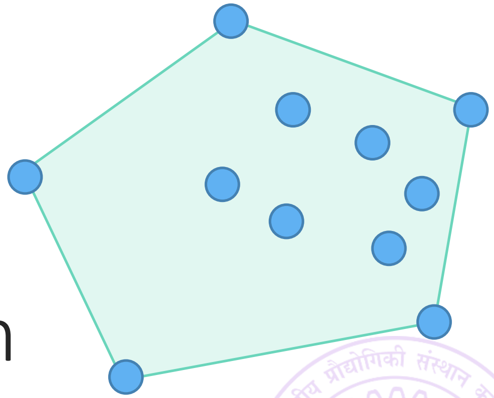
Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

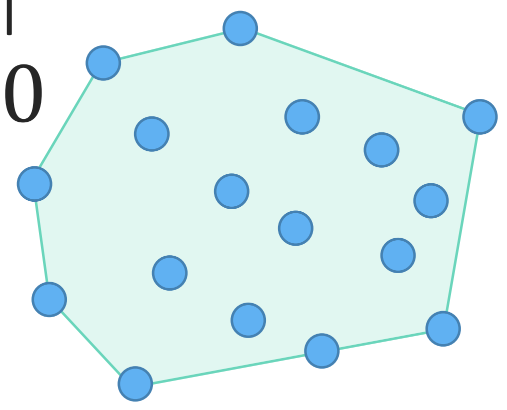


Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

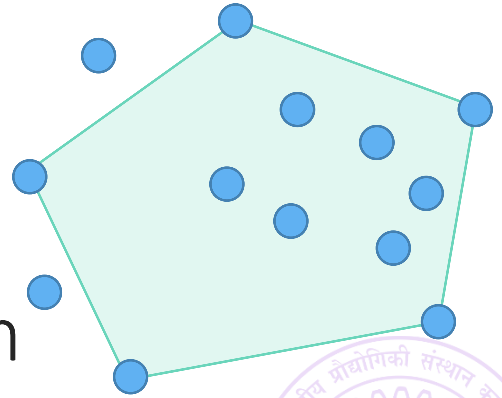
Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^n p_1 = 1$ and computing the point $\sum_{i=1}^n p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

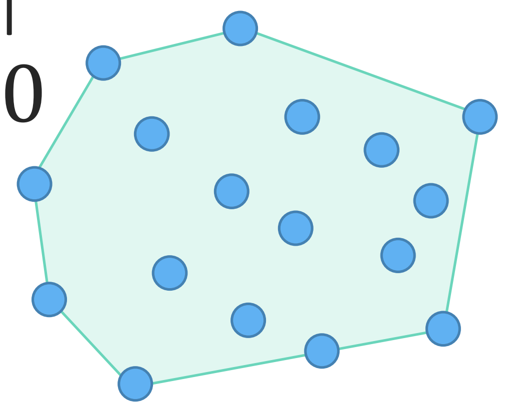
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)



Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

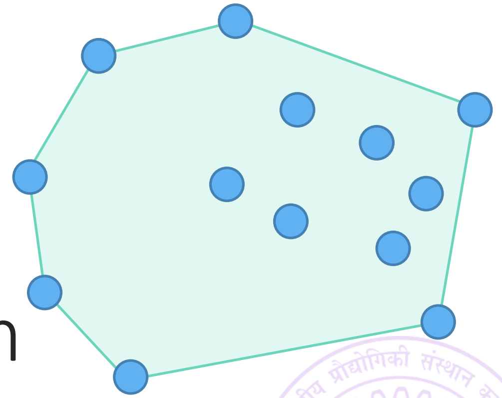
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinations of More than 2 Vectors



Given n points $\mathbf{x}^1, \dots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \dots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$

The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points



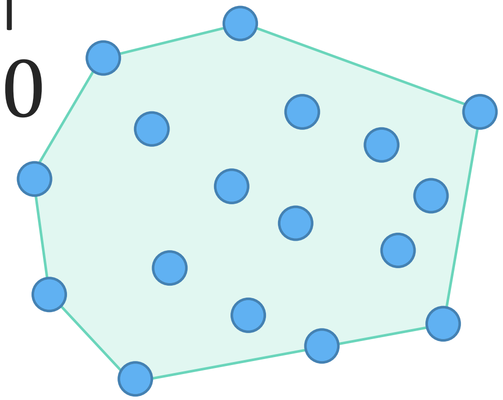
Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

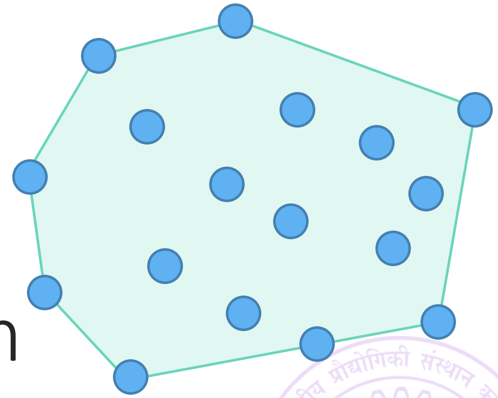
Combinations of More than 2 Vectors

Given n points $\mathbf{x}^1, \dots, \mathbf{x}^n \in \mathbb{R}^d$, a convex combination of these points is obtained by selecting $p_1, \dots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

Trick: just as before, start small – start with two points and look at their convex combinations (line segment)

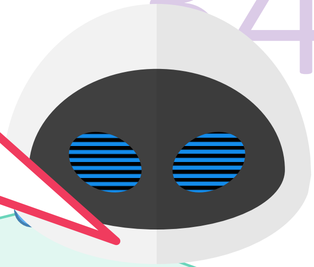


Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

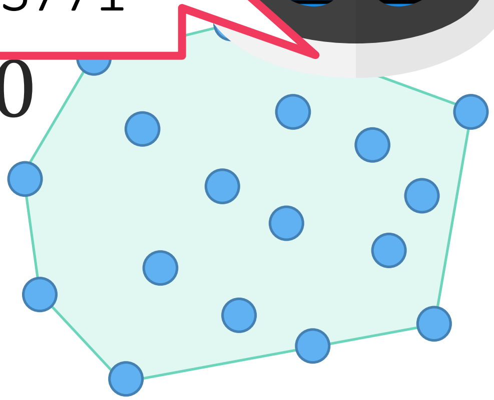
Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Combinat

It does not matter in which order do you consider the points. Eventually you will land up with the same convex hull. This claim requires a proof which is beyond the scope of CS771

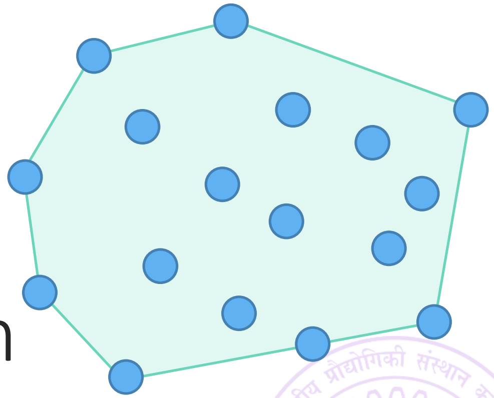


Given n points these points is obtained by selecting $p_1, \ldots, p_n$ s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_1 = 1$ and computing the point $\sum_{i=1}^{n} p_i \cdot \mathbf{x}^i$



The set of all convex combinations gives us the *convex hull* of these points, defined as the the smallest convex set that contains all these points

Trick: just as before, start small – start with two points and look at their convex combinations (line segment)



Then add a third point by taking some weight away from the previous two points i.e. $p_1 + p_2 = \alpha$ and $p_3 = 1 - \alpha$

Go over all $\alpha \in [0,1]$. Add more points similarly. If a new point already lies inside previous hull, nothing to do!

Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ *is covered so adding more vectors in 2 dims is not interesting anymore*

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$

Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

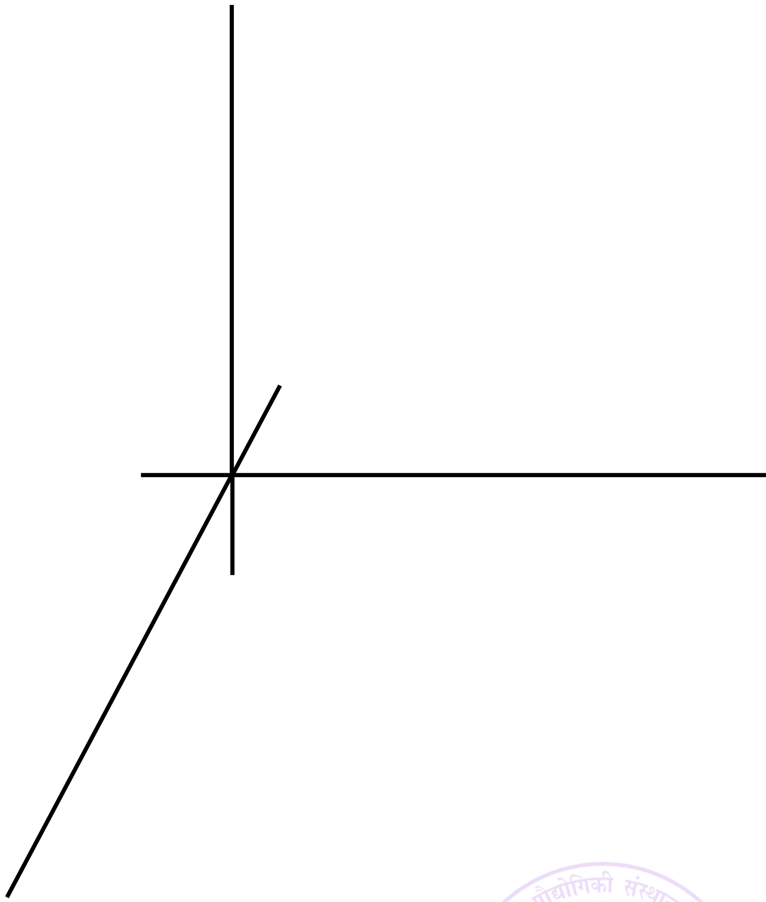
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ is covered so adding more vectors in 2 dims is not interesting anymore

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

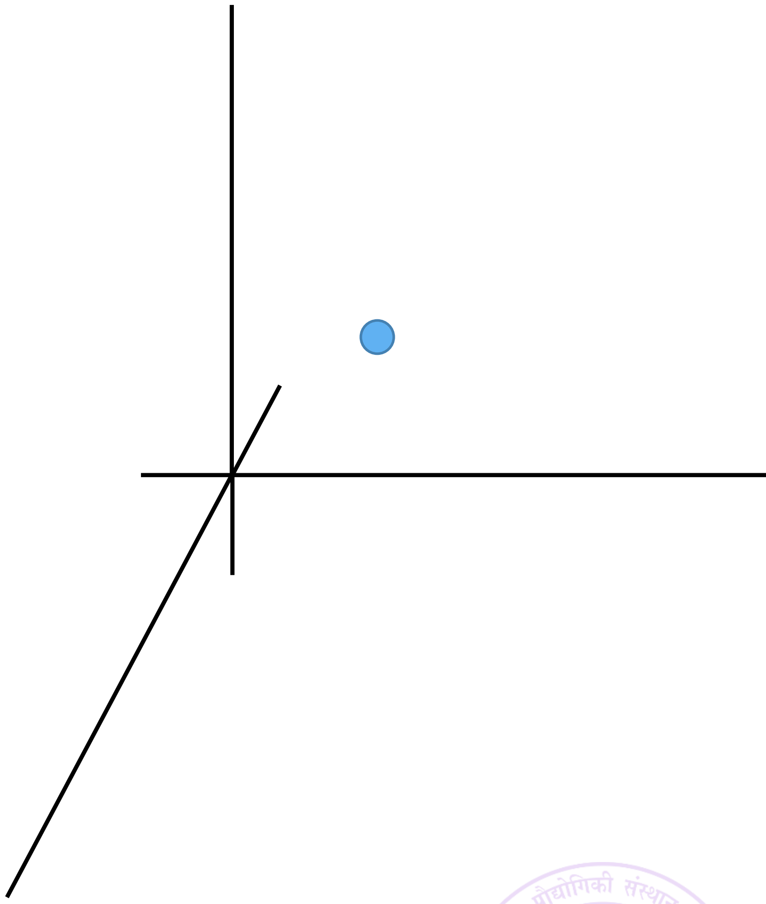
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ *is covered so adding more vectors in 2 dims is not interesting anymore*

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

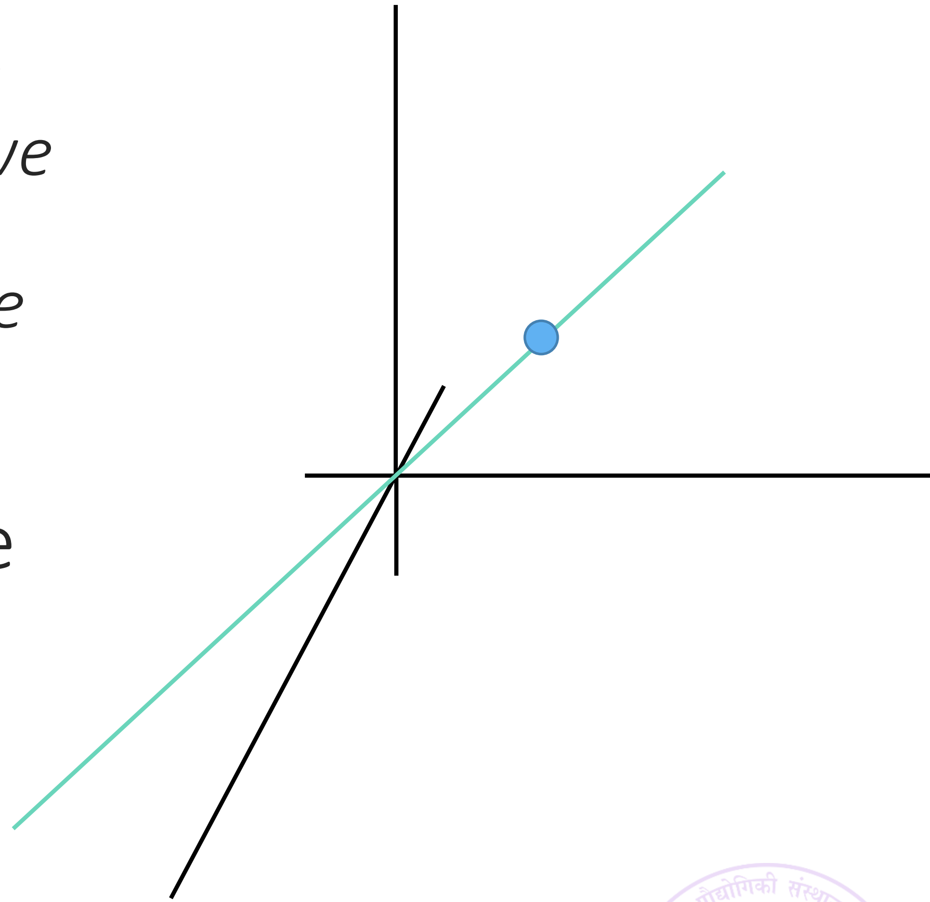
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ is covered so adding more vectors in 2 dims is not interesting anymore

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

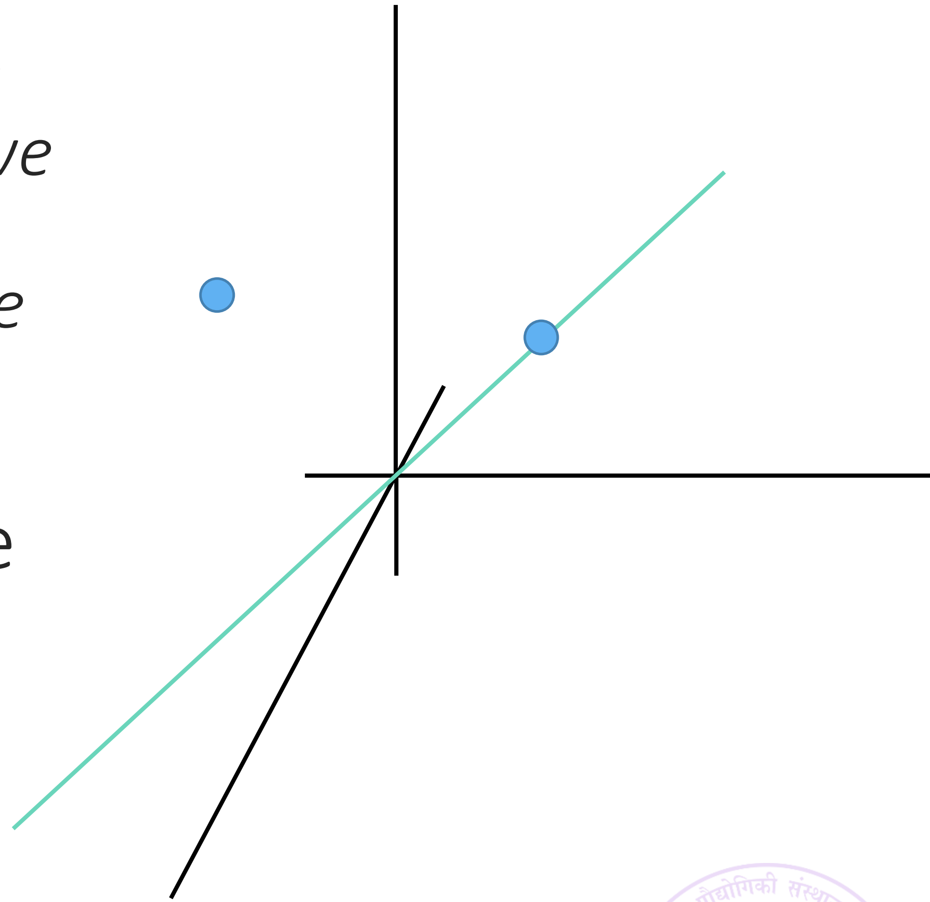
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ is covered so adding more vectors in 2 dims is not interesting anymore

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

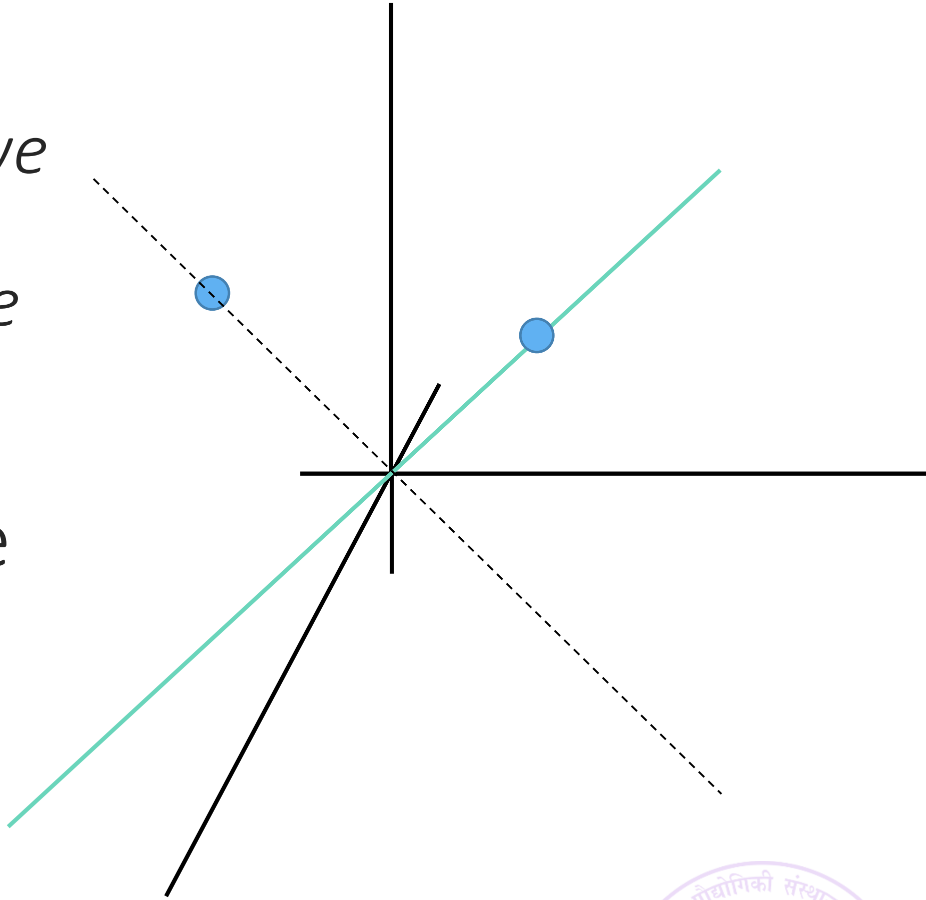
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ *is covered so adding more vectors in 2 dims is not interesting anymore*

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

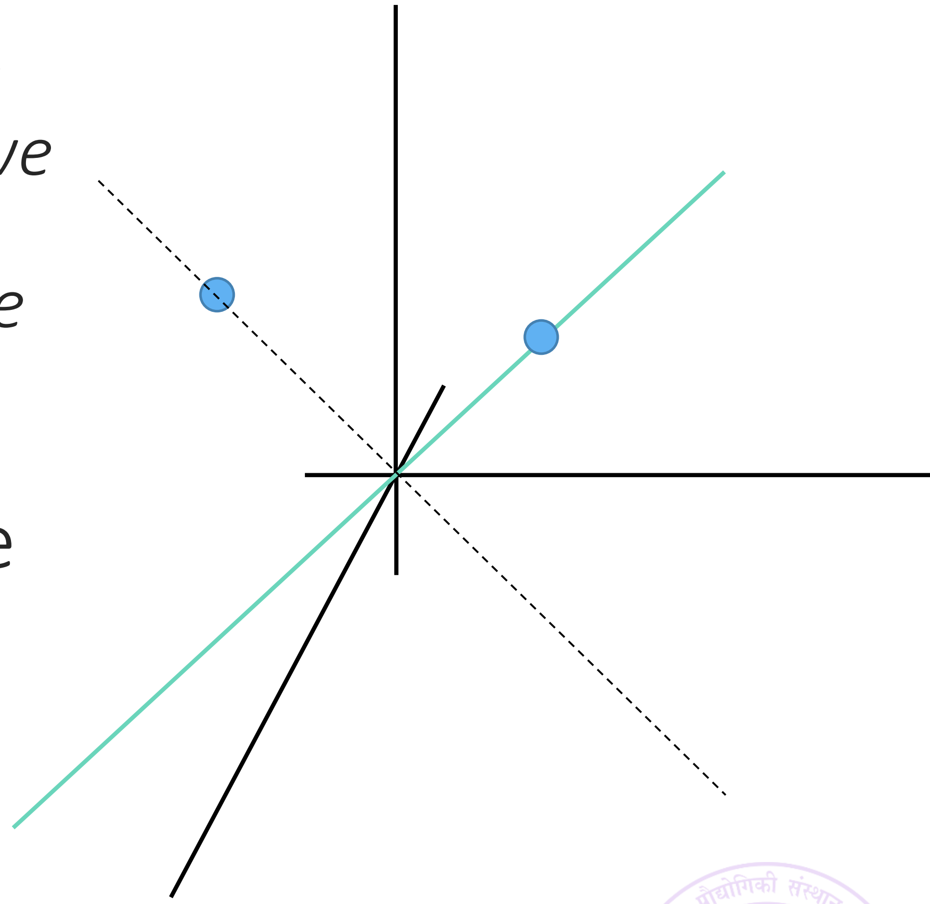
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ *is covered so adding more vectors in 2 dims is not interesting anymore*

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

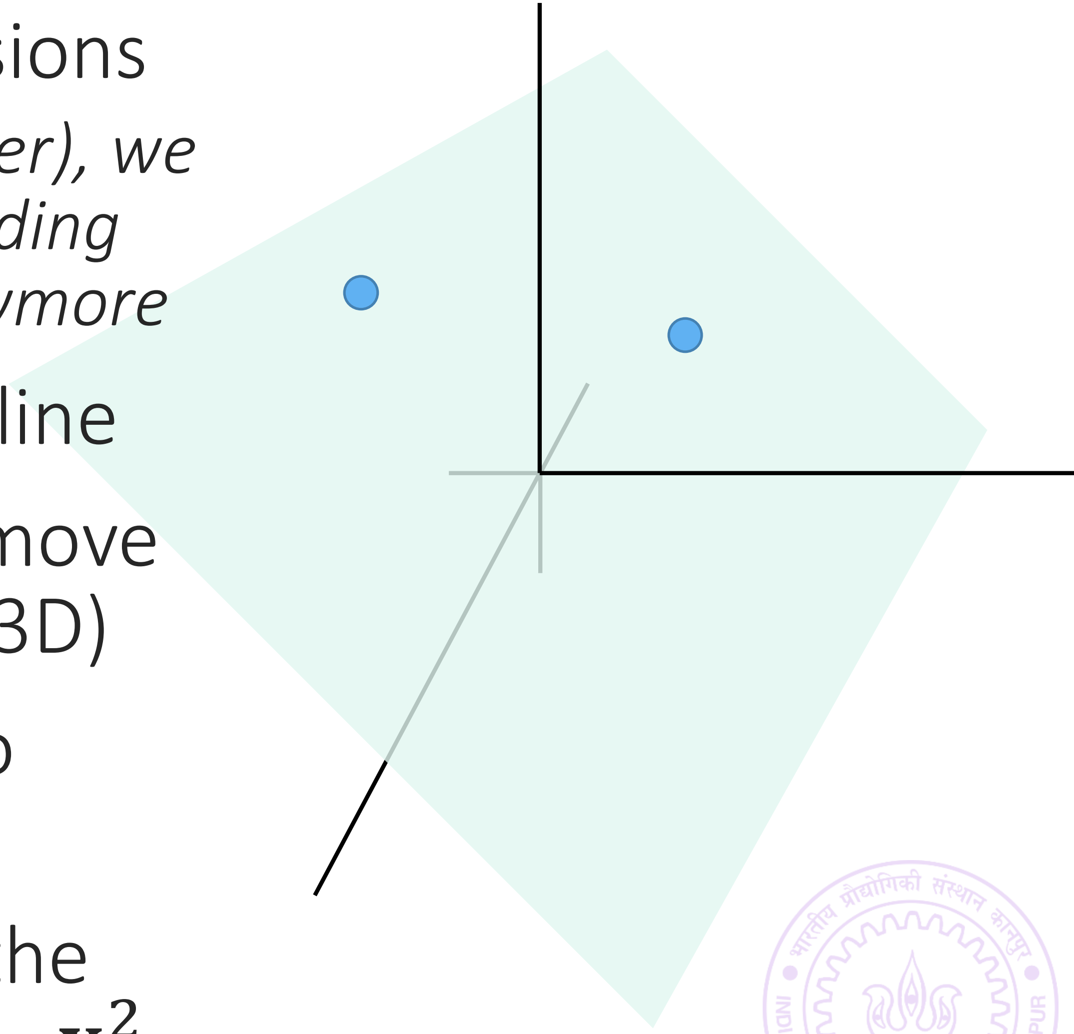
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ is covered so adding more vectors in 2 dims is not interesting anymore

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

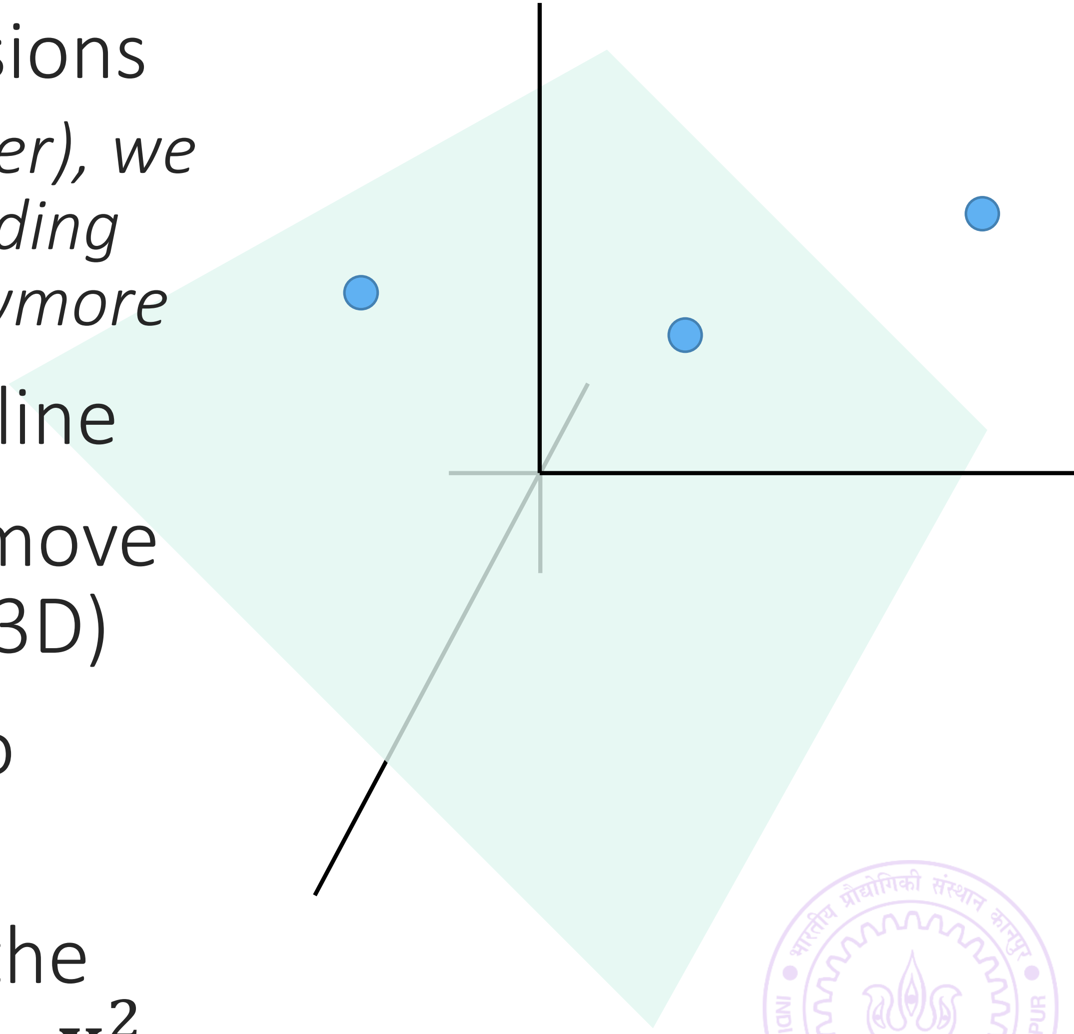
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ is covered so adding more vectors in 2 dims is not interesting anymore

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

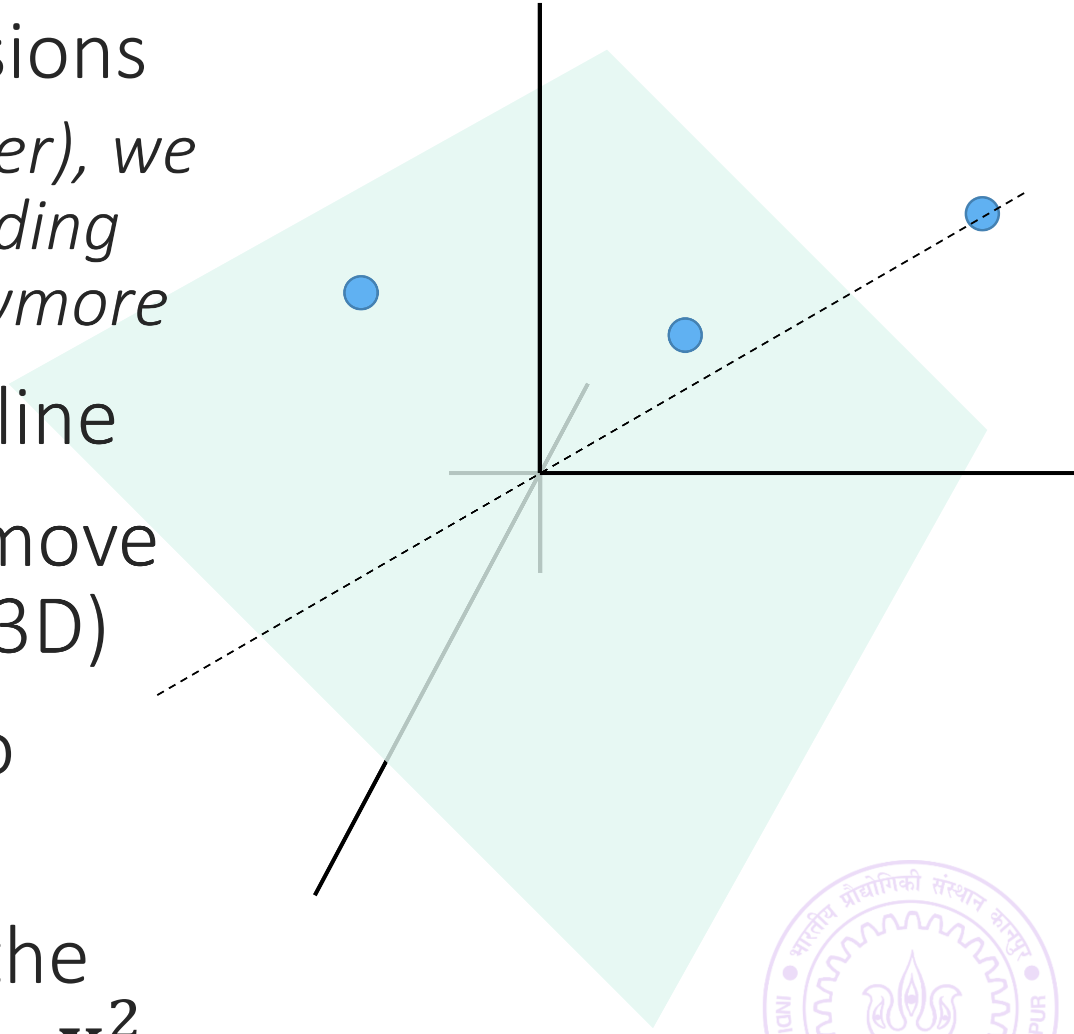
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ *is covered so adding more vectors in 2 dims is not interesting anymore*

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

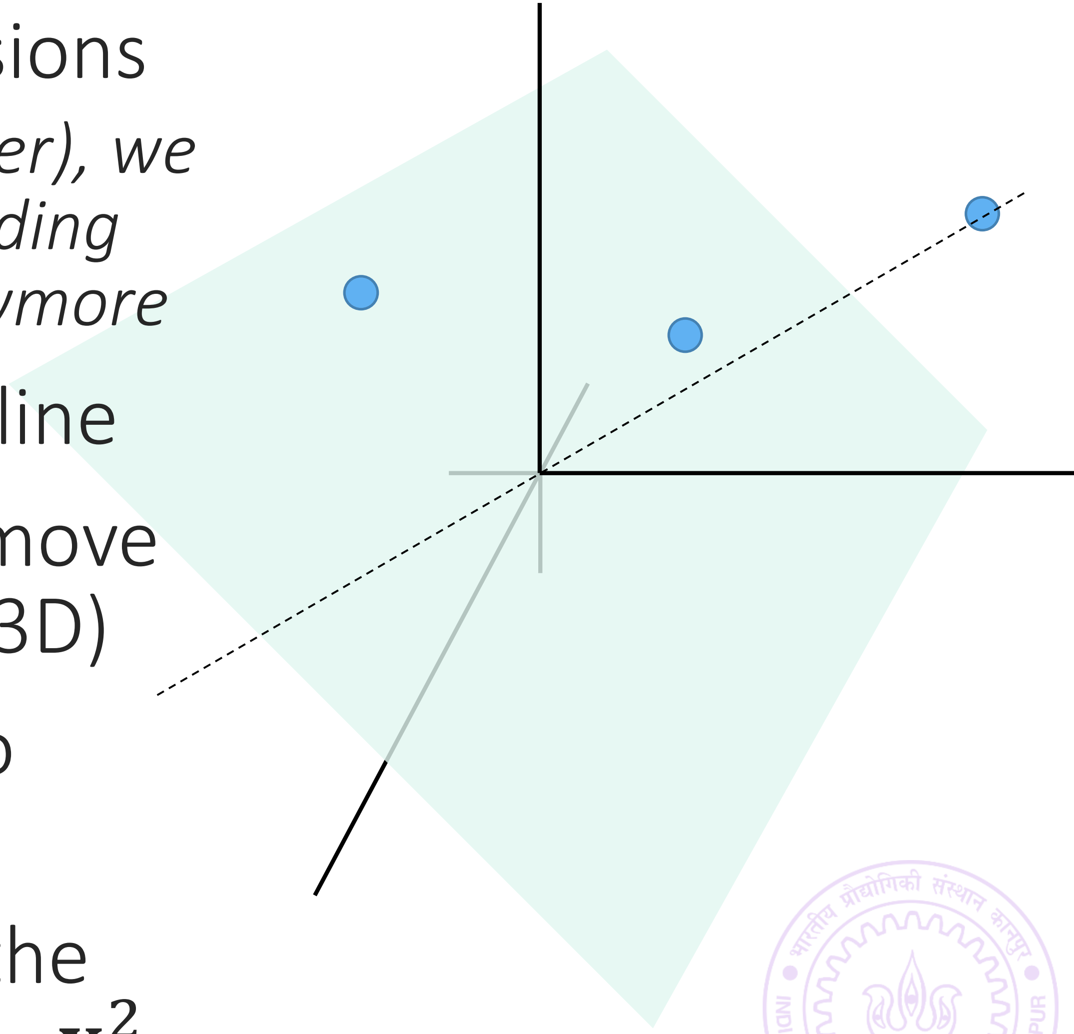
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ is covered so adding more vectors in 2 dims is not interesting anymore

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

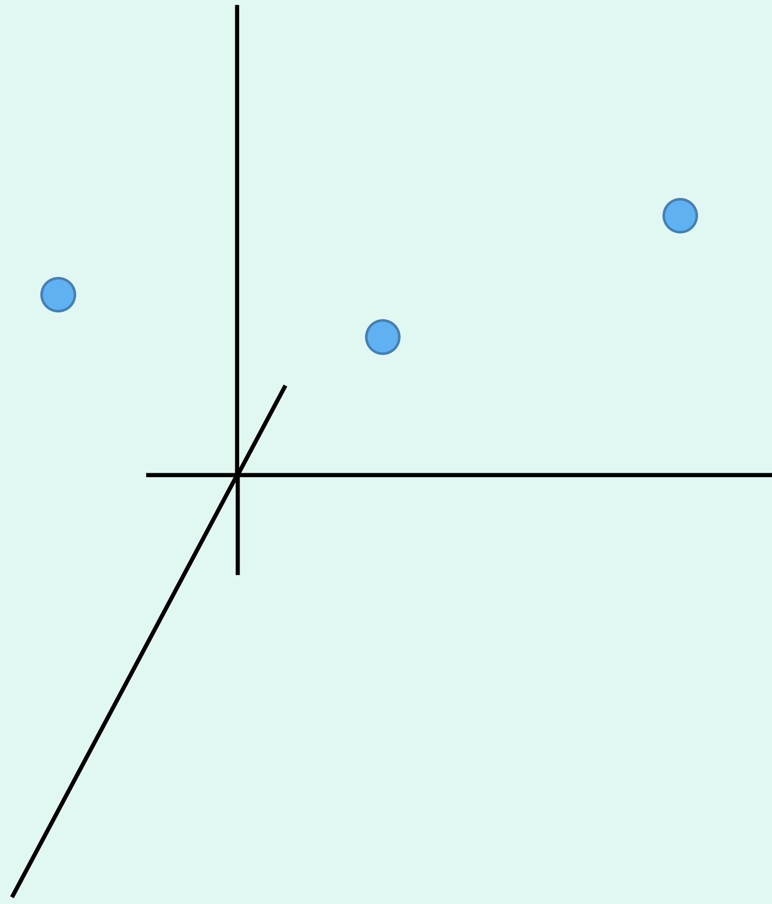
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ is covered so adding more vectors in 2 dims is not interesting anymore

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of More than Two Vectors

Let us consider three vectors in 3 dimensions

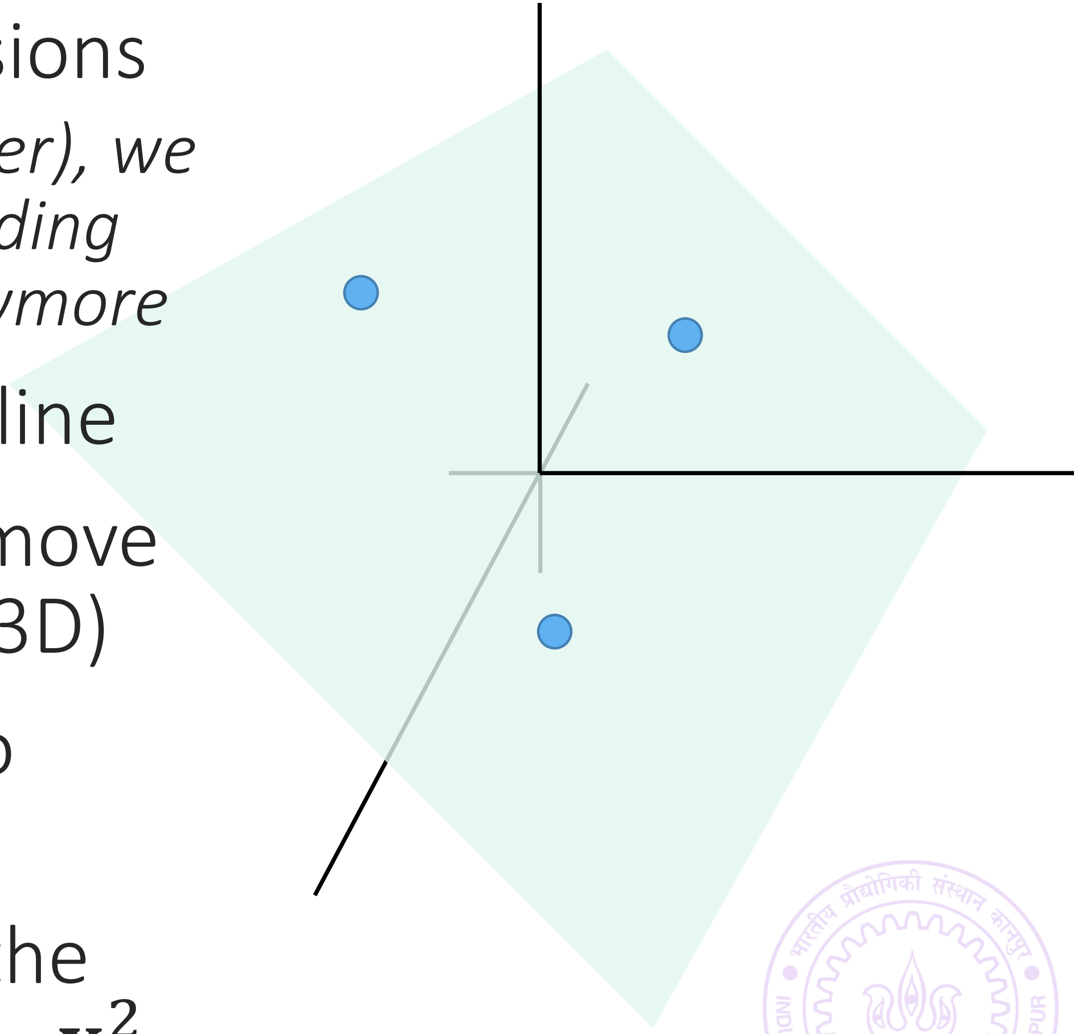
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ is covered so adding more vectors in 2 dims is not interesting anymore

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of [partially hidden]

Same reason as before. For any α, β, γ, we have
$$\alpha \cdot \mathbf{x}^1 + \beta \cdot \mathbf{x}^2 + \gamma \cdot \mathbf{x}^3 = (\alpha + p^1\gamma) \cdot \mathbf{x}^1 + (\beta + p^2\gamma) \cdot \mathbf{x}^2$$
i.e. we get no new points outside of the plane by adding $\mathbf{x}^3$

Let us consid

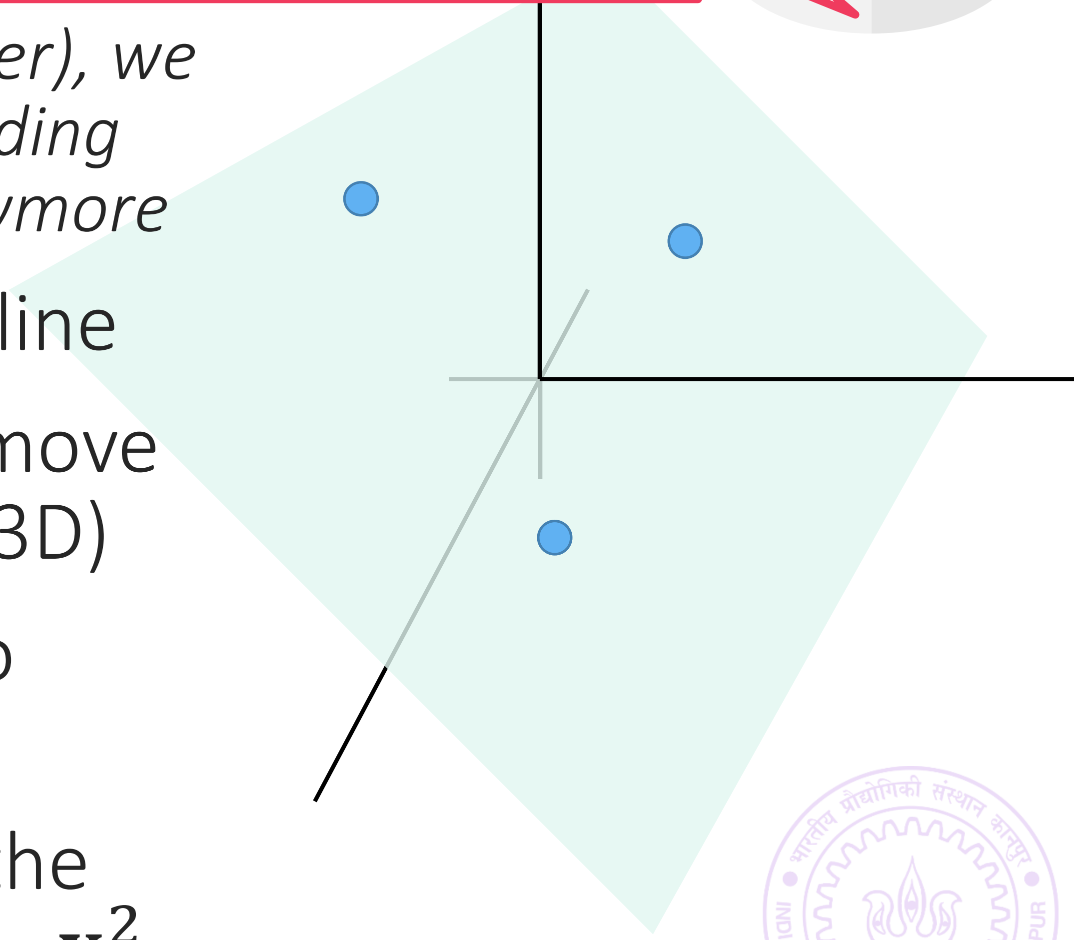
With two vectors (not multiples of each other), we already saw that entire $\mathbb{R}^2$ is covered so adding more vectors in 2 dims is not interesting anymore

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$

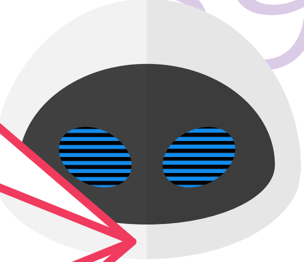


Span of

Same reason as before. For any α, β, γ, we have

$$\alpha \cdot \mathbf{x}^1 + \beta \cdot \mathbf{x}^2 + \gamma \cdot \mathbf{x}^3 = (\alpha + p^1\gamma) \cdot \mathbf{x}^1 + (\beta + p^2\gamma) \cdot \mathbf{x}^2$$

i.e. we get no new points outside of the plane by adding $\mathbf{x}^3$



Let us consid

With two ve
already saw
more vecto

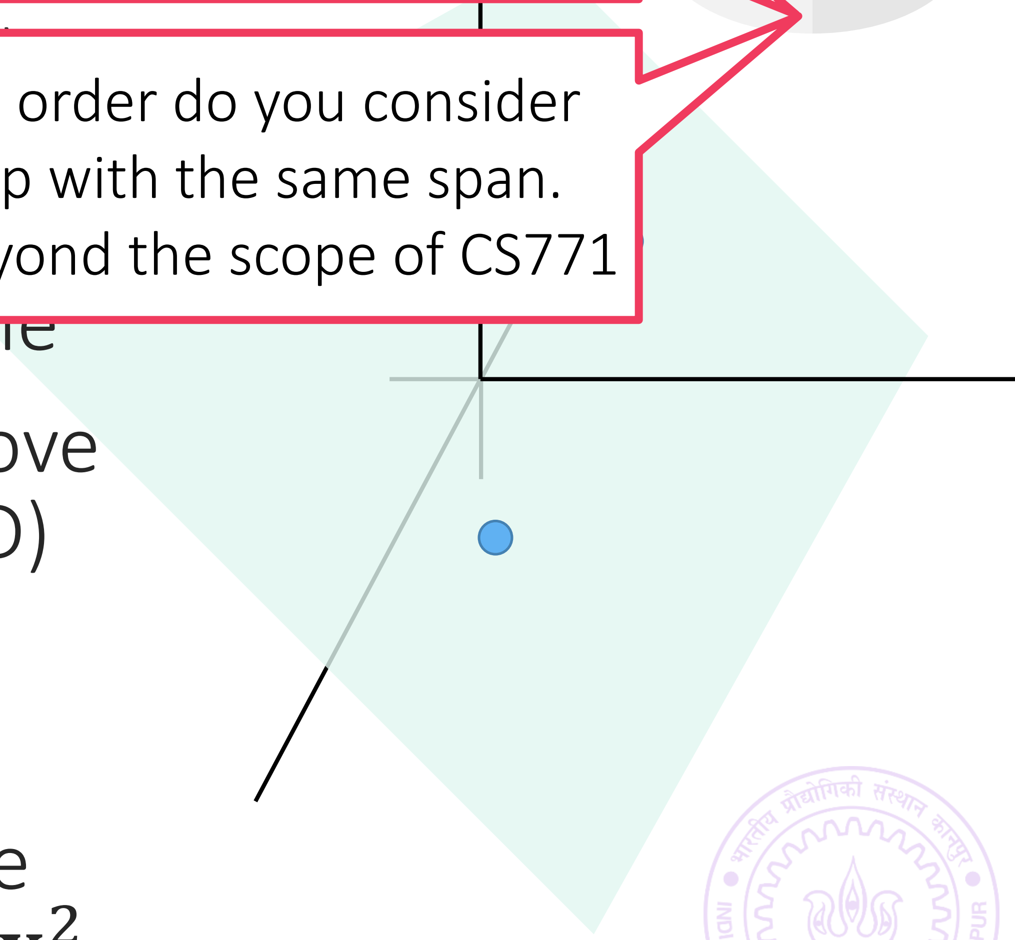
Yet again, it does not matter in which order do you consider the points. Eventually you will land up with the same span. This claim requires a proof which is beyond the scope of CS771

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a new vector $\mathbf{x}^2$: causes the line to move about and become a plane (flat sheet in 3D)

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$



Span of I

Same reason as before. For any α, β, γ, we have
$$\alpha \cdot \mathbf{x}^1 + \beta \cdot \mathbf{x}^2 + \gamma \cdot \mathbf{x}^3 = (\alpha + p^1\gamma) \cdot \mathbf{x}^1 + (\beta + p^2\gamma) \cdot \mathbf{x}^2$$
i.e. we get no new points outside of the plane by adding $\mathbf{x}^3$

Let us consid

With two ve
already saw
more vecto

Yet again, it does not matter in which order do you consider the points. Eventually you will land up with the same span. This claim requires a proof which is beyond the scope of CS771

Trick: start with span of one vector $\mathbf{x}^1$: a line

Add a ne
about an

Affine combinations of more than two vectors can also produce planes/hyperplanes. We will not require affine combinations much in our discussions so we are not going into details of those here

Add a new vector $\mathbf{x}^3$: causes the sheet to move about and cover entire $\mathbb{R}^3$

If $\mathbf{x}^3$ had been expressible as a comb of the previous 2 vectors i.e. $\mathbf{x}^3 = p^1 \cdot \mathbf{x}^1 + p^2 \cdot \mathbf{x}^2$ for some p^1, p^2, then would not cover $\mathbb{R}^3$

Dependence and Independence

If we can express $\mathbf{x}^n$ as a linear combination of some other vectors $\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}$, then we say that $\mathbf{x}^n$ is *linearly dependent* on $\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}$

- *$\mathbf{x}^n$ in such cases is redundant: linear combinations of $\mathbf{x}^1, \ldots, \mathbf{x}^n$ dont give us any new vector that some linear combination of $\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}$ didnt already give*
- *In other words,* $\mathrm{span}(\mathbf{x}^1, \ldots, \mathbf{x}^n) = \mathrm{span}(\mathbf{x}^1, \ldots, \mathbf{x}^{n-1})$
- *Show that if $\mathbf{x}^n$ is lin-dep on $\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}$ then there must exist at least one $i \in [n-1]$ s.t. $\mathbf{x}^i$ is also lin-dep on $\{\mathbf{x}^j\}_{1 \leq j \leq n, j \neq i}$ – linear dependence is infectious* ☹

A set of vectors $\mathbf{x}^1, \ldots, \mathbf{x}^n$ is said to be *linearly independent* if no vector $\mathbf{x}^i$ can be written as a linear combination of the other vectors $\{\mathbf{x}^j\}_{j \neq i}$

Basis

Given a set of vectors $S = \{\mathbf{x}^i\}$, consider the set $\mathrm{span}(S)$

A set of vectors B is called a basis for the set S if vectors in the set B are linearly independent as well as $\mathrm{span}(B) = \mathrm{span}(S)$

This is true if and only if (often written as iff) B is lin-indep and $\mathrm{span}(B) \supseteq S$

Linear independence assures us that the basis cannot be shrunk any further

If vectors in B are orthogonal *i.e.* $\mathbf{b}^1 \perp \mathbf{b}^2$ *if* $\mathbf{b}^1, \mathbf{b}^2 \in B, \mathbf{b}^1 \neq \mathbf{b}^2$ *then B is called an* orthogonal basis

If vectors in B are orthonormal *i.e. they are orthogonal as well as* $\|\mathbf{b}\|_2 = 1$ *for all* $\mathbf{b} \in B$*, then B is called an* orthonormal basis

These definitions hold true even if the set S contains infinitely many vectors

Need to define $\mathrm{span}(S)$ more carefully in this case but relatively simple

Exists a simple algo to extract a basis out of a finite set of vectors

Gram-Schmidt Process

Gram-Schmidt Process

Given $S = \{\mathbf{x}^1, \ldots, \mathbf{x}^n\}$, this process yields an orthonormal basis of S

Initialize basis by removing an (arbitrary) element from S, say $B = \{\mathbf{x}^i\}$

Repeat until all the set S is not empty

- *Remove an arbitrary element from the set S, say $\mathbf{x}$*
- *Compute $\tilde{\mathbf{x}} = \mathbf{x} - \sum_{\mathbf{b} \in B} (\mathbf{b}^\top \mathbf{x}) \cdot \mathbf{b}$*
- *If $\tilde{\mathbf{x}} \neq \mathbf{0}$, add $\tilde{\mathbf{x}} / \|\tilde{\mathbf{x}}\|_2$ to the basis B else throw $\tilde{\mathbf{x}}$ away*

The exact basis we get depends on the order in which we process the vectors but we will get an orthonormal basis every time

- *We will also get a basis of the same size every time (proof omitted)*
- *The above algorithm is simple but can be numerically imprecise (recall overflow issues), numerically stable versions of GM also exist*

Linear Maps/Transformations

These maps vectors to other vectors, but in a way that preserves lines

The mapped vectors may have the same/smaller/larger dimensionality

If a bunch of vectors were lying on a line earlier (any line), the mapped vectors would also lie on a (possibly different line)

Linear transformations always map the origin to the origin itself. If a map preserves all lines but shifts the origin – called an affine transformation *instead*

Mathematically, function $f : \mathbb{R}^d \to \mathbb{R}^k$ is a linear map/transformation if

$f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$ *for* **any** *pair of two vectors* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$

Note: The vectors $f(\mathbf{x} + \mathbf{y}), f(\mathbf{x}), f(\mathbf{y})$ are all k-dimensional, not d-dimensional

The emphasis on the word ***any*** is important. It wont do if this holds only for some pairs

$f(c \cdot \mathbf{x}) = c \cdot f(\mathbf{x})$ *for* **any** $c \in \mathbb{R}$ *and* **any** $\mathbf{x} \in \mathbb{R}^d$

Note: taking $c = 0$ shows us that $f(\mathbf{0}) = \mathbf{0}$ where $\mathbf{0} \in \mathbb{R}^d$ and $\mathbf{0} \in \mathbb{R}^k$

Linear Maps/Transformations

These maps vectors to other vectors, but in a way that preserve

Consider the line joining vectors $\mathbf{a}, \mathbf{b}$: all points on this line can be described as an affine combination of $\mathbf{a}, \mathbf{b}$ i.e. $\mathbf{v}_\lambda = \lambda \cdot \mathbf{a} + (1 - \lambda) \cdot \mathbf{b}$ for some $\lambda \in \mathbb{R}$. Now, given a linear map $f: \mathbb{R}^d \to \mathbb{R}^k$ we have

$$f(\mathbf{v}_\lambda) = f(\lambda \cdot \mathbf{a}) + f\big((1 - \lambda) \cdot \mathbf{b}\big) = \lambda \cdot f(\mathbf{a}) + (1 - \lambda) \cdot f(\mathbf{b})$$

Thus, $f(\mathbf{v}_\lambda)$ also lies on the line joining $f(\mathbf{a})$ and $f(\mathbf{b})$. Conversely, consider a point $\mathbf{x} \in \mathbb{R}^k$ on the line joining $f(\mathbf{a})$ and $f(\mathbf{b})$. This point must be of the form $t \cdot f(\mathbf{a}) + (1 - t) \cdot f(\mathbf{b})$ for some $t \in \mathbb{R}$. But this means that $\mathbf{x} = f(t \cdot \mathbf{a} + (1 - t) \cdot \mathbf{b})$ which means that $\mathbf{x}$ was the result of f mapping some point on the line joining $\mathbf{a}, \mathbf{b}$

ectors

nap

n instead

ation if

Note: The vectors $f(\mathbf{x} + \mathbf{y}), f(\mathbf{x}), f(\mathbf{y})$ are all k-dimensional, not d-dimensional

The emphasis on the word ***any*** is important. It wont do if this holds only for some pairs

$f(c \cdot \mathbf{x}) = c \cdot f(\mathbf{x})$ *for* **any** $c \in \mathbb{R}$ *and* **any** $\mathbf{x} \in \mathbb{R}^d$

Note: taking $c = 0$ shows us that $f(\mathbf{0}) = \mathbf{0}$ where $\mathbf{0} \in \mathbb{R}^d$ and $\mathbf{0} \in \mathbb{R}^k$

Encoding Linear Transformations

Although it may seem very tedious to write down such a function that maps vectors to vectors in a “linear” way, it is actually very easy

Let $\mathbf{e}_i = (0,0,\ldots,0,1,0,\ldots,0,0) \in \mathbb{R}^d$ be the vector in $\mathbb{R}^d$ which has $\mathbf{1}$ in the i-th coordinate and $\mathbf{0}$ everywhere else

Often called the $\boldsymbol{i}$-th canonical/standard/natural basic vector

Suppose someone tells me $f(\mathbf{e}_1), f(\mathbf{e}_2), \ldots, f(\mathbf{e}_d)$ (all these are $\in \mathbb{R}^k$)

For any new $\mathbf{v} = (v_1, \ldots, v_d) \in \mathbb{R}^d$, I can use linearity to find $f(\mathbf{v})$

$$f(\mathbf{v}) = f\left(\sum_{i=1}^{d} v_i \cdot \mathbf{e}_i\right) = \sum_{i=1}^{d} f(v_i \cdot \mathbf{e}_i) = \sum_{i=1}^{d} v_i \cdot f(\mathbf{e}_i)$$

Thus, fixing where the canonical vectors get mapped completely fixes where ***any*** vector $\mathbf{v} \in \mathbb{R}^d$ must get mapped

Matrices **ARE** Linear Transformations!

If we arrange all these d vectors (all of which are k-dimensional) as column vectors side by side, we get a $k \times d$ matrix $A \in \mathbb{R}^{k \times d}$

$$A = \begin{bmatrix} \begin{bmatrix} \\ f(\mathbf{e}_1) \\ \\ \end{bmatrix} \begin{bmatrix} \\ f(\mathbf{e}_2) \\ \\ \end{bmatrix} \dots \begin{bmatrix} \\ f(\mathbf{e}_d) \\ \\ \end{bmatrix} \end{bmatrix}$$

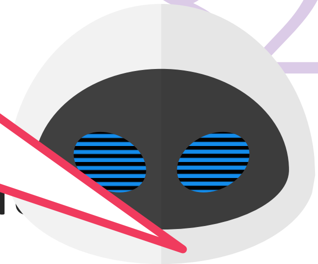
Note: another way of writing $\sum_{i=1}^{d} v_i \cdot f(\mathbf{e}_i)$ is to simply write $A\mathbf{v}$

Thus, all linear transformations from $\mathbb{R}^d \to \mathbb{R}^k$ can be expressed as a $k \times d$ matrix A such that $A\mathbf{v}$ gives us the transformed vectors!

It is easy to verify that for any matrix $B \in \mathbb{R}^{k \times d}$, the transformation defined as $g: \mathbf{v} \mapsto B\mathbf{v}$ is always linear

Ma

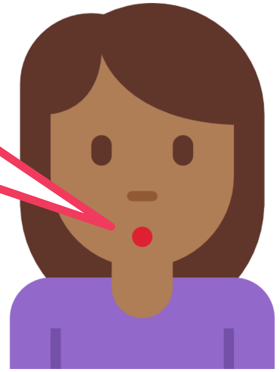
Be careful – this works only for linear transformations. If the transformation you have in mind is non-linear, then merely specifying what happens to $\mathbf{e}_1, \dots \mathbf{e}_d$ on that transformation will not tell us complete details of that map



If w

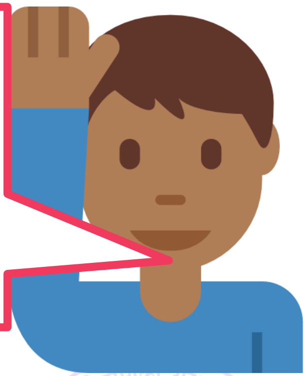
column vec

This means every linear transformation uniquely corresponds to a matrix and every matrix uniquely defines a linear transformation!



$$A = \left[f(\mathbf{e}_1) \middle| f(\mathbf{e}_2) \middle| \dots \middle| f(\mathbf{e}_d) \right]$$

We use feature matrices in ML, say $X \in \mathbb{R}^{n \times d}$. Even these matrices correspond to maps – they map models to scores i.e. $X: \mathbf{w} \mapsto \hat{\mathbf{y}} = X\mathbf{w}$. Thus, features define a linear map too! The goal of ML is to find a suitable input for this map (the model $\mathbf{w}$) such that the output is as desired by us (e.g. we may want $\hat{\mathbf{y}} \approx \mathbf{y}$ or $\mathbf{sign}(\hat{\mathbf{y}}) \approx \mathbf{y}$)



gives us the transformed vectors!

It is easy to verify that for any matrix $B \in \mathbb{R}^{k \times d}$, the transformation defined as $g: \mathbf{v} \mapsto B\mathbf{v}$ is always linear

Special Linear Transformations

Special Linear Transformations

The first kind of linear transformations that we study are *scaling transformations* that simply scale the various dimensions by different amounts

Special Linear Transformations

The first kind of linear transformations that we study are *scaling transformations* that simply scale the various dimensions by different amounts

- *Scaling by a negative quantity is permitted – amounts to flipping that dimension*

Special Linear Transformations

The first kind of linear transformations that we study are *scaling transformations* that simply scale the various dimensions by different amounts

- *Scaling by a negative quantity is permitted – amounts to flipping that dimension*
- *For scaling transformations, output vector is of the same dim as input vector*

Special Linear Transformations

The first kind of linear transformations that we study are *scaling transformations* that simply scale the various dimensions by different amounts

- *Scaling by a negative quantity is permitted – amounts to flipping that dimension*
- *For scaling transformations, output vector is of the same dim as input vector*
- *Sometimes, we may treat scaling by zero as equivalent to dropping a dimension*

Special Linear Transformations

The first kind of linear transformations that we study are *scaling transformations* that simply scale the various dimensions by different amounts

- *Scaling by a negative quantity is permitted – amounts to flipping that dimension*
- *For scaling transformations, output vector is of the same dim as input vector*
- *Sometimes, we may treat scaling by zero as equivalent to dropping a dimension*
- *Scaling transformations correspond to diagonal matrices*

Special Linear Transformations

The first kind of linear transformations that we study are *scaling transformations* that simply scale the various dimensions by different amounts

- *Scaling by a negative quantity is permitted – amounts to flipping that dimension*
- *For scaling transformations, output vector is of the same dim as input vector*
- *Sometimes, we may treat scaling by zero as equivalent to dropping a dimension*
- *Scaling transformations correspond to diagonal matrices*

Easy to see that scaling maps are linear

Special Linear Transformations

The first kind of linear transformations that we study are *scaling transformations* that simply scale the various dimensions by different amounts

- *Scaling by a negative quantity is permitted – amounts to flipping that dimension*
- *For scaling transformations, output vector is of the same dim as input vector*
- *Sometimes, we may treat scaling by zero as equivalent to dropping a dimension*
- *Scaling transformations correspond to diagonal matrices*

Easy to see that scaling maps are linear

e.g. $A = \begin{bmatrix} -0.5 & 0 \\ 0 & 1.5 \end{bmatrix}$

Scale x axis by factor of 0.5 and flip it

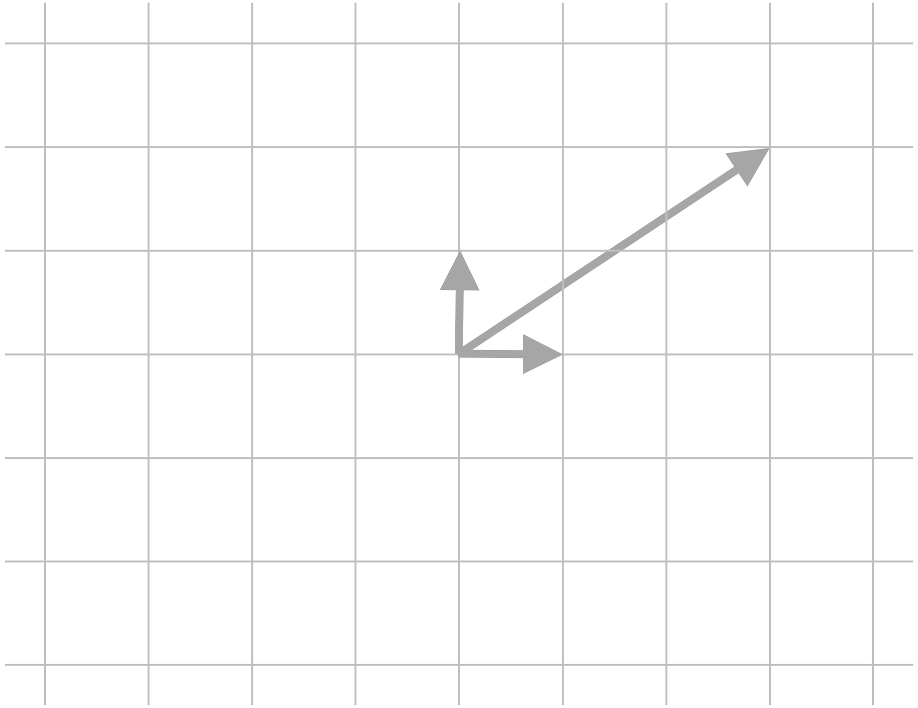
Scale y axis by a factor of 1.5

Special Linear Transformations

The first kind of linear transformations that we study are *scaling transformations* that simply scale the various dimensions by different amounts

- *Scaling by a negative quantity is permitted – amounts to flipping that dimension*
- *For scaling transformations, output vector is of the same dim as input vector*
- *Sometimes, we may treat scaling by zero as equivalent to dropping a dimension*
- *Scaling transformations correspond to diagonal matrices*

Easy to see that scaling maps are linear



e.g. $A = \begin{bmatrix} -0.5 & 0 \\ 0 & 1.5 \end{bmatrix}$

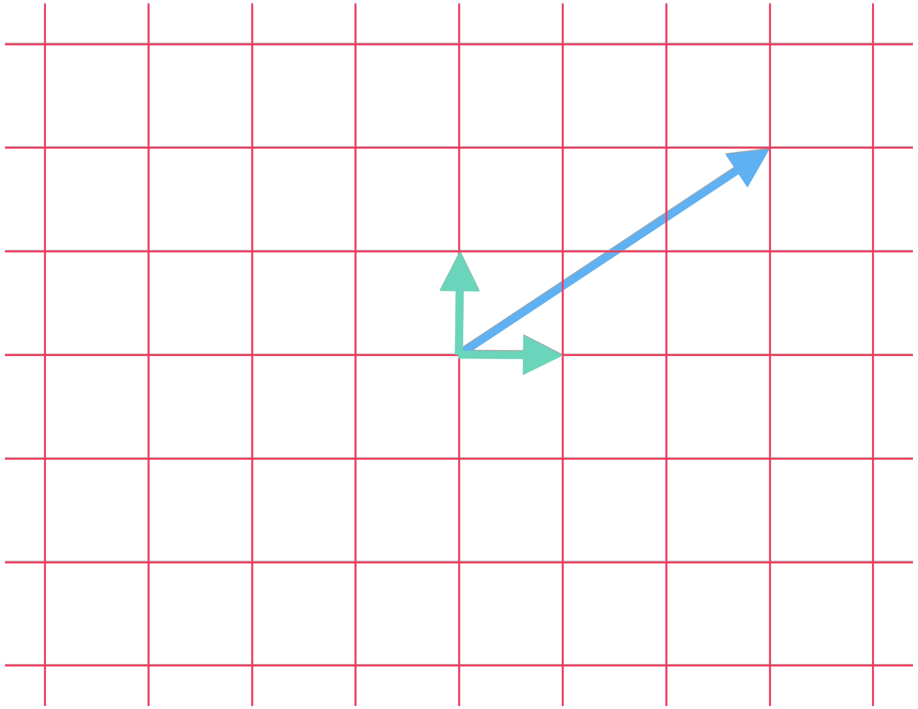
Scale x axis by factor of 0.5 and flip it
Scale y axis by a factor of 1.5

Special Linear Transformations

The first kind of linear transformations that we study are *scaling transformations* that simply scale the various dimensions by different amounts

- *Scaling by a negative quantity is permitted – amounts to flipping that dimension*
- *For scaling transformations, output vector is of the same dim as input vector*
- *Sometimes, we may treat scaling by zero as equivalent to dropping a dimension*
- *Scaling transformations correspond to diagonal matrices*

Easy to see that scaling maps are linear



e.g. $A = \begin{bmatrix} -0.5 & 0 \\ 0 & 1.5 \end{bmatrix}$

Scale x axis by factor of 0.5 and flip it

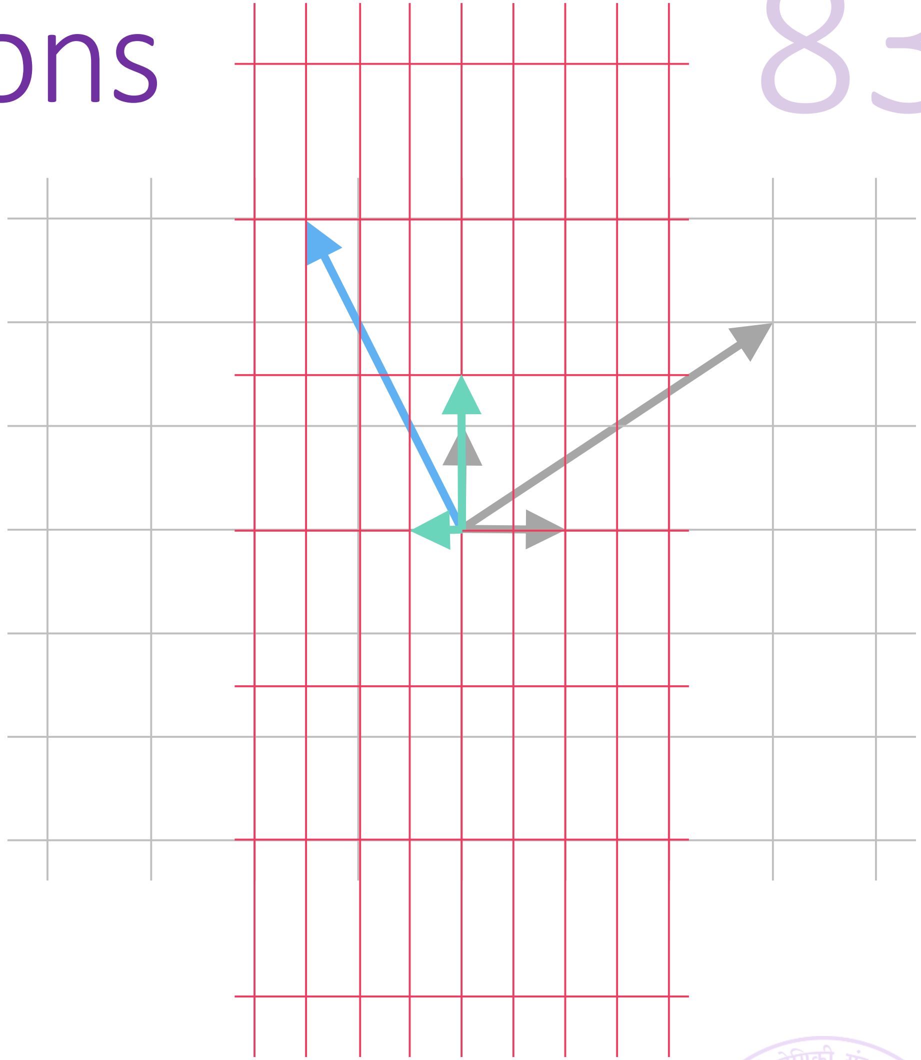
Scale y axis by a factor of 1.5

Special Linear Transformations

The first kind of linear transformations that we study are *scaling transformations* that simply scale the various dimensions by different amounts

- *Scaling by a negative quantity is permitted – amounts to flipping that dimension*
- *For scaling transformations, output vector is of the same dim as input vector*
- *Sometimes, we may treat scaling by zero as equivalent to dropping a dimension*
- *Scaling transformations correspond to diagonal matrices*

Easy to see that scaling maps are linear



e.g. $A = \begin{bmatrix} -0.5 & 0 \\ 0 & 1.5 \end{bmatrix}$

Scale x axis by factor of 0.5 and flip it

Scale y axis by a factor of 1.5

Special Linear Transformations

Rotation transformations are linear maps that simply rotate the whole space

- *Also preserve dot products b/w any two vecs*
  - **Corollary**: also preserve Euclid. distances b/w any two points since $\|\mathbf{x}-\mathbf{y}\|_2 = \sqrt{\langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y}\rangle}$
  - Such maps are called *isometric transformations*
  - *iso = same, metricus = measurement*
- *Output vec always of same dim as input vec*
- *Rotation transformations always correspond to matrices whose columns are orthonormal*
  - Since $\mathbf{e}_i \perp \mathbf{e}_j$ if $i \neq j$ and rot. maps preserve dot products, we must have $f(\mathbf{e}_i) \perp f(\mathbf{e}_j)$ too!
  - Since $\|\mathbf{e}_i\|_2 = 1$ hence $\|f(\mathbf{e}_i)\|_2 = 1$ too!

Special Linear Transformations

Rotation transformations are linear maps that simply rotate the whole space

- *Also preserve dot products b/w any two vecs*
  - **Corollary**: also preserve Euclid. distances b/w any two points since $\|\mathbf{x}-\mathbf{y}\|_2 = \sqrt{\langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y} \rangle}$
  - Such maps are called *isometric transformations* *iso = same, metricus = measurement*
- *Output vec always of same dim as input vec*
- *Rotation transformations always correspond to matrices whose columns are orthonormal*
  - Since $\mathbf{e}_i \perp \mathbf{e}_j$ if $i \neq j$ and rot. maps preserve dot products, we must have $f(\mathbf{e}_i) \perp f(\mathbf{e}_j)$ too!
  - Since $\|\mathbf{e}_i\|_2 = 1$ hence $\|f(\mathbf{e}_i)\|_2 = 1$ too!

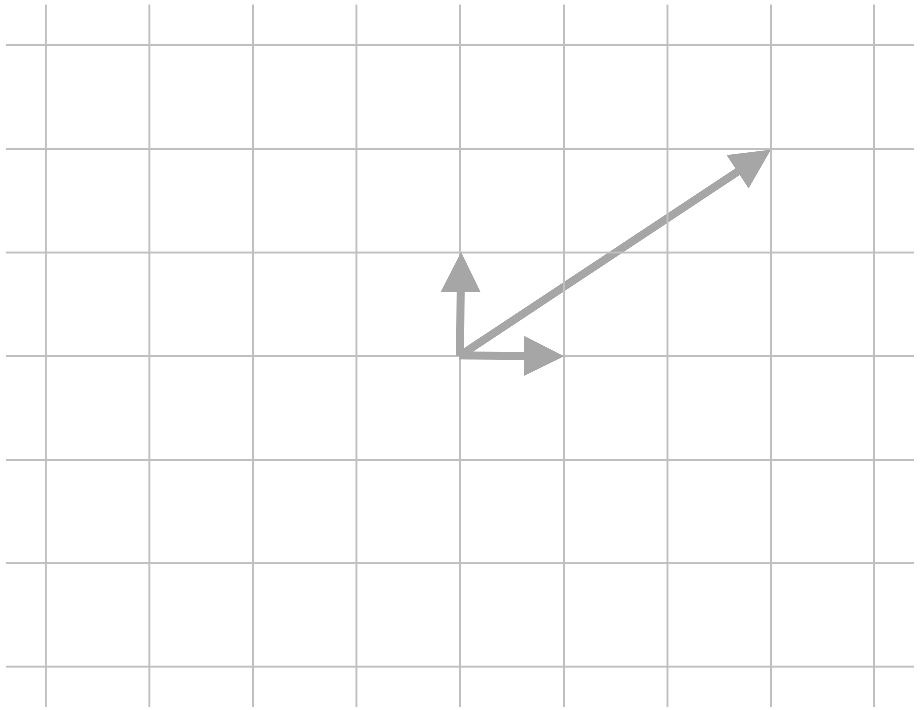
e.g. $A = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Rotate counter-clockwise by 45 degrees

Special Linear Transformations

Rotation transformations are linear maps that simply rotate the whole space

- *Also preserve dot products b/w any two vecs*
  - **Corollary**: also preserve Euclid. distances b/w any two points since $\|\mathbf{x}-\mathbf{y}\|_2 = \sqrt{\langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y}\rangle}$
  - Such maps are called *isometric transformations*
  - *iso = same, metricus = measurement*
- *Output vec always of same dim as input vec*
- *Rotation transformations always correspond to matrices whose columns are orthonormal*
  - Since $\mathbf{e}_i \perp \mathbf{e}_j$ if $i \neq j$ and rot. maps preserve dot products, we must have $f(\mathbf{e}_i) \perp f(\mathbf{e}_j)$ too!
  - Since $\|\mathbf{e}_i\|_2 = 1$ hence $\|f(\mathbf{e}_i)\|_2 = 1$ too!



e.g. $A = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Rotate counter-clockwise by 45 degrees

Special Linear Transformations

Rotation transformations are linear maps that simply rotate the whole space

Also preserve dot products b/w any two vecs

Corollary: also preserve Euclid. distances b/w any two points since $\|\mathbf{x}-\mathbf{y}\|_2 = \sqrt{\langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y}\rangle}$

Such maps are called *isometric transformations*

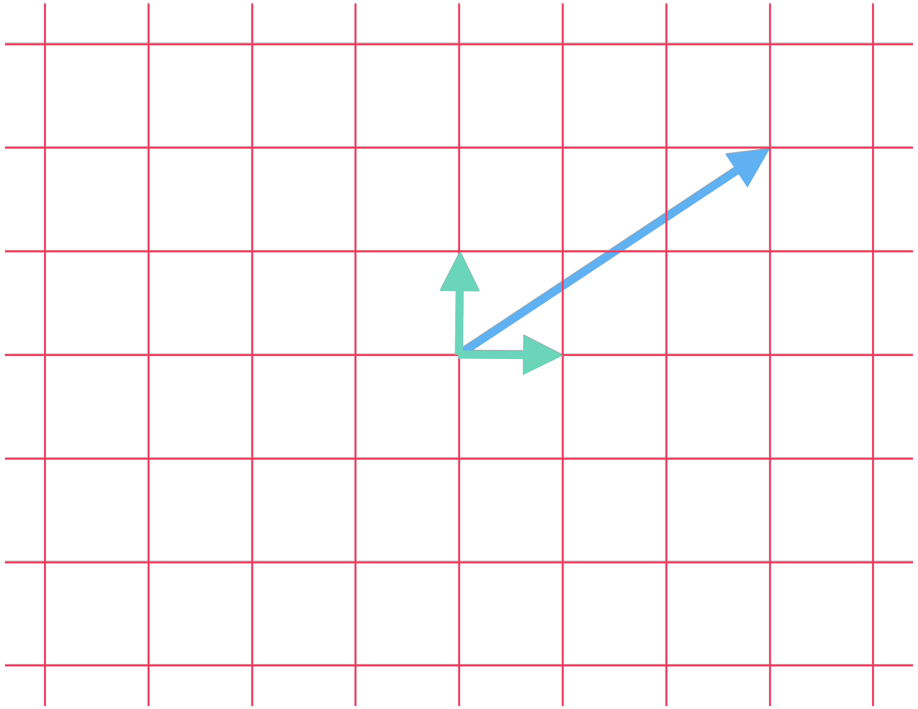
iso = same, metricus = measurement

Output vec always of same dim as input vec

Rotation transformations always correspond to matrices whose columns are orthonormal

Since $\mathbf{e}_i \perp \mathbf{e}_j$ if $i \neq j$ and rot. maps preserve dot products, we must have $f(\mathbf{e}_i) \perp f(\mathbf{e}_j)$ too!

Since $\|\mathbf{e}_i\|_2 = 1$ hence $\|f(\mathbf{e}_i)\|_2 = 1$ too!



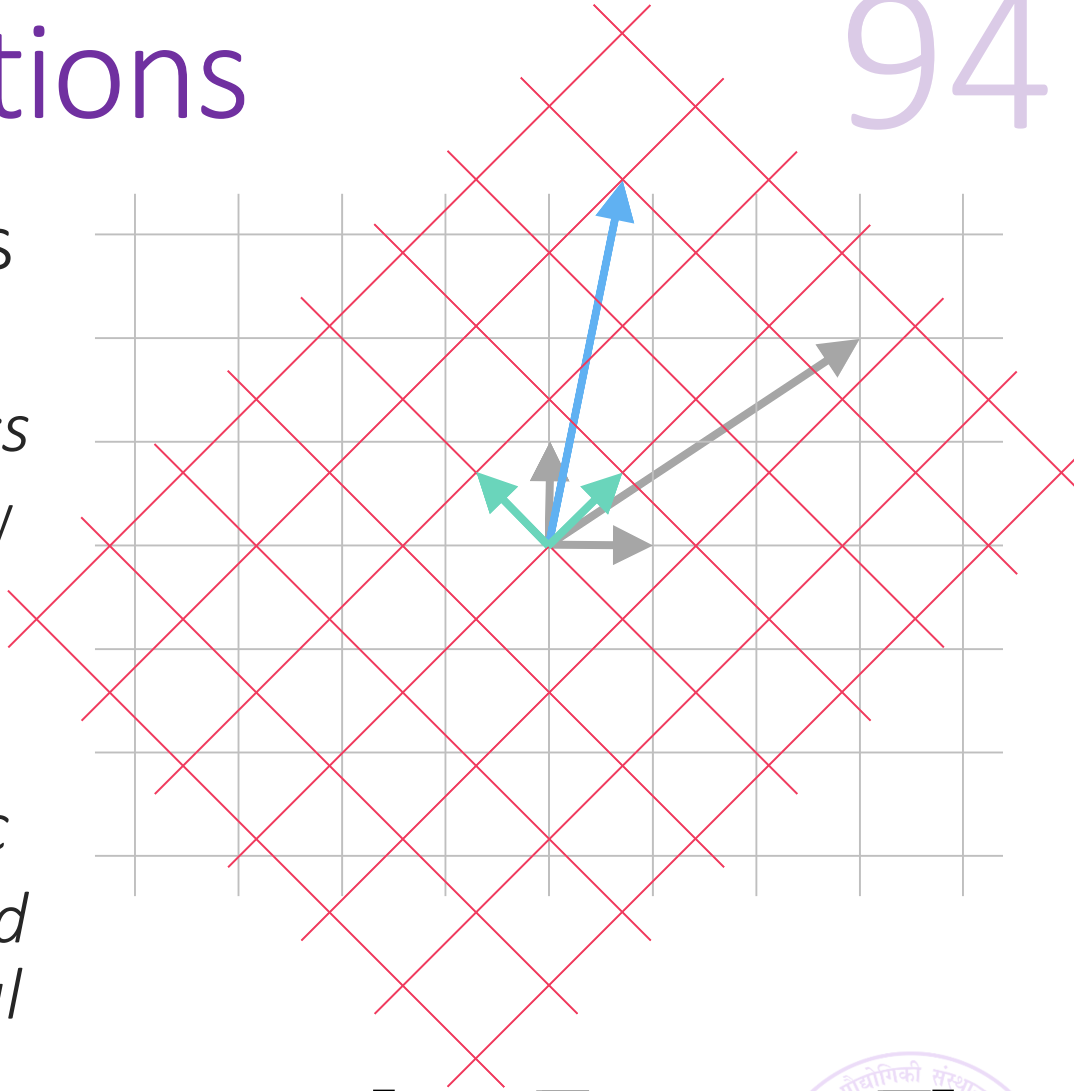
e.g. $A = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Rotate counter-clockwise by 45 degrees

Special Linear Transformations

Rotation transformations are linear maps that simply rotate the whole space

- *Also preserve dot products b/w any two vecs*
  - **Corollary**: also preserve Euclid. distances b/w any two points since $\|\mathbf{x}-\mathbf{y}\|_2 = \sqrt{\langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y} \rangle}$
  - Such maps are called *isometric transformations*
  - *iso = same, metricus = measurement*
- *Output vec always of same dim as input vec*
- *Rotation transformations always correspond to matrices whose columns are orthonormal*
  - Since $\mathbf{e}_i \perp \mathbf{e}_j$ if $i \neq j$ and rot. maps preserve dot products, we must have $f(\mathbf{e}_i) \perp f(\mathbf{e}_j)$ too!
  - Since $\|\mathbf{e}_i\|_2 = 1$ hence $\|f(\mathbf{e}_i)\|_2 = 1$ too!

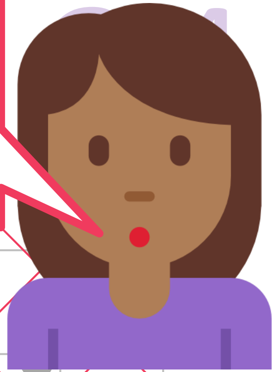


e.g. $A = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Rotate counter-clockwise by 45 degrees

Special Lin

Caution: not all orthonormal matrices correspond to rotation maps though. Orthonormal matrices can also give maps that flip the sign of an axis (after rotating it) or even exchange axes



Rotation transformations are linear maps that simply rotate the whole space

Also preserve dot products b/w any two vecs

Corollary: also preserve Euclid. distances b/w any two points since $\|\mathbf{x}-\mathbf{y}\|_2 = \sqrt{\langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y}\rangle}$

Such maps are called *isometric transformations*

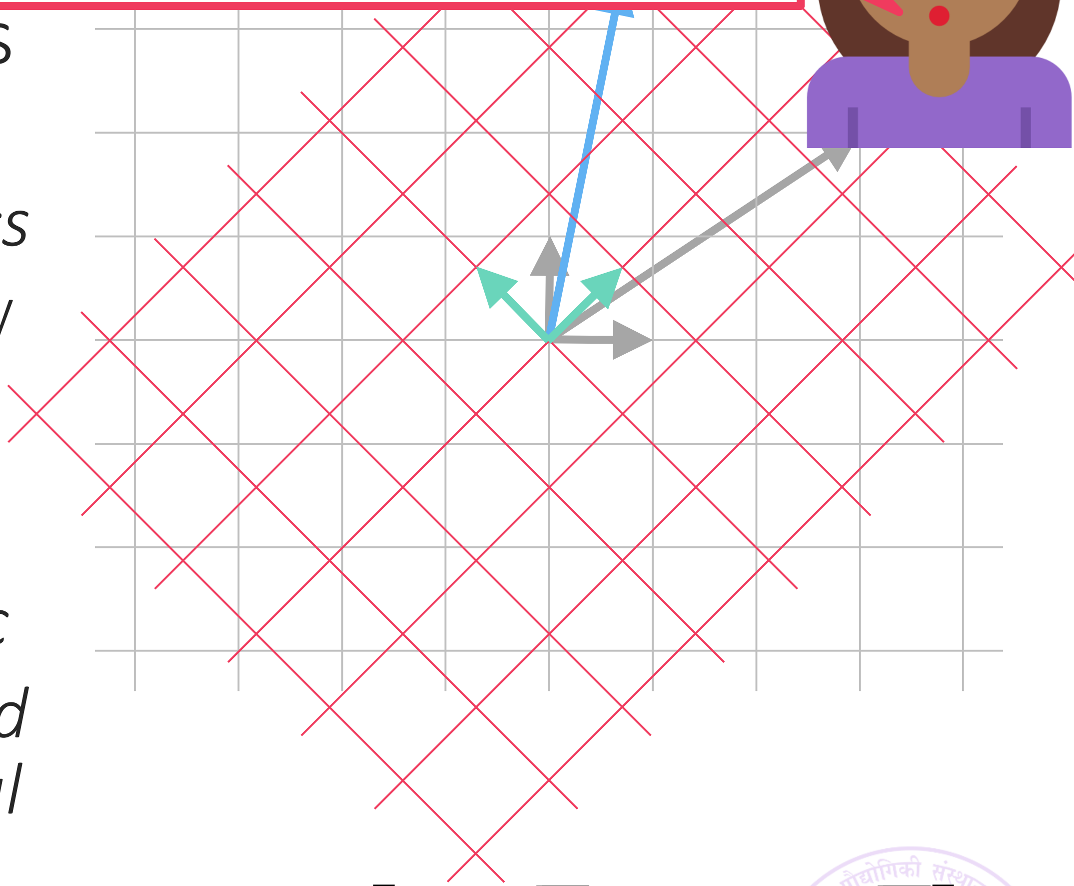
iso = same, metricus = measurement

Output vec always of same dim as input vec

Rotation transformations always correspond to matrices whose columns are orthonormal

Since $\mathbf{e}_i \perp \mathbf{e}_j$ if $i \neq j$ and rot. maps preserve dot products, we must have $f(\mathbf{e}_i) \perp f(\mathbf{e}_j)$ too!

Since $\|\mathbf{e}_i\|_2 = 1$ hence $\|f(\mathbf{e}_i)\|_2 = 1$ too!



e.g. $A = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Rotate counter-clockwise by 45 degrees

Special Linear Transformations

Beware of linear transformations where the vectors $f(\mathbf{e}_1), \dots, f(\mathbf{e}_d)$ are not linearly independent

In such case, the entire space is mapped to a lower-dimensional hyperplane/line

Sure, this low dimensional (hyper)plane/line could be sitting inside a higher dimensional space but it would itself be low dimensional

This could be the case even if k is a large number, even if $k \gg d$

Thus, notice that the possible values of $A\mathbf{v}$ are exactly $\mathrm{span}\big(f(\mathbf{e}_1), \dots, f(\mathbf{e}_d)\big)$

Special Linear Transformations

Beware of linear transformations where the vectors $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_d)$ are not linearly independent

In such case, the entire space is mapped to a lower-dimensional hyperplane/line

Sure, this low dimensional (hyper)plane/line could be sitting inside a higher dimensional space but it would itself be low dimensional

This could be the case even if k is a large number, even if $k \gg d$

Thus, notice that the possible values of $A\mathbf{v}$ are exactly $\mathrm{span}\big(f(\mathbf{e}_1), \ldots, f(\mathbf{e}_d)\big)$

e.g. $A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}$

Special Linear Transformations

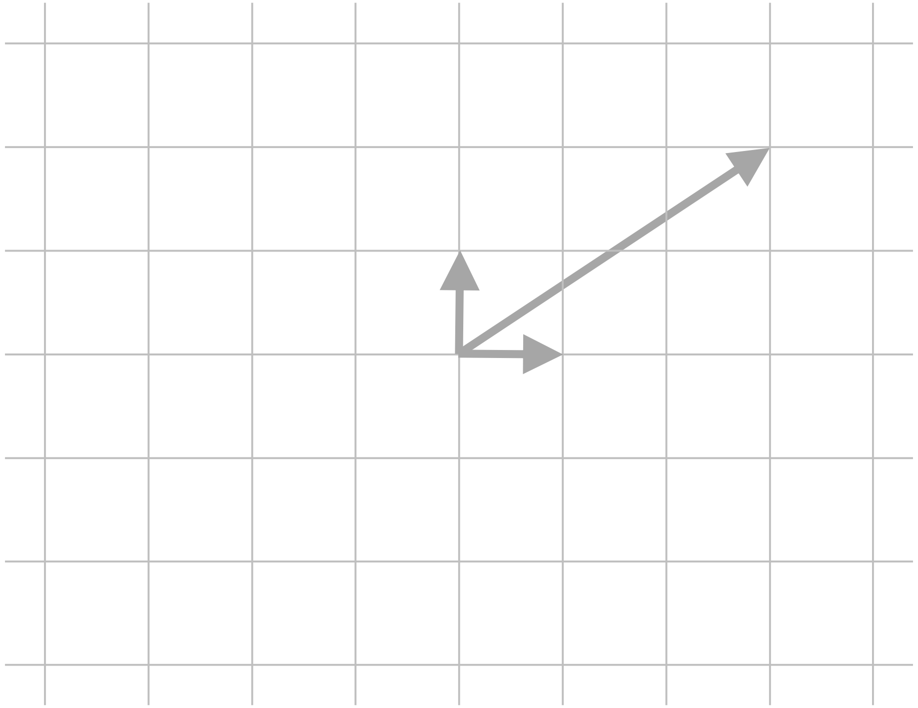
Beware of linear transformations where the vectors $f(\mathbf{e}_1), \dots, f(\mathbf{e}_d)$ are not linearly independent

In such case, the entire space is mapped to a lower-dimensional hyperplane/line

Sure, this low dimensional (hyper)plane/line could be sitting inside a higher dimensional space but it would itself be low dimensional

This could be the case even if k is a large number, even if $k \gg d$

Thus, notice that the possible values of $A\mathbf{v}$ are exactly $\mathrm{span}\big(f(\mathbf{e}_1), \dots, f(\mathbf{e}_d)\big)$



e.g. $A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}$

Special Linear Transformations

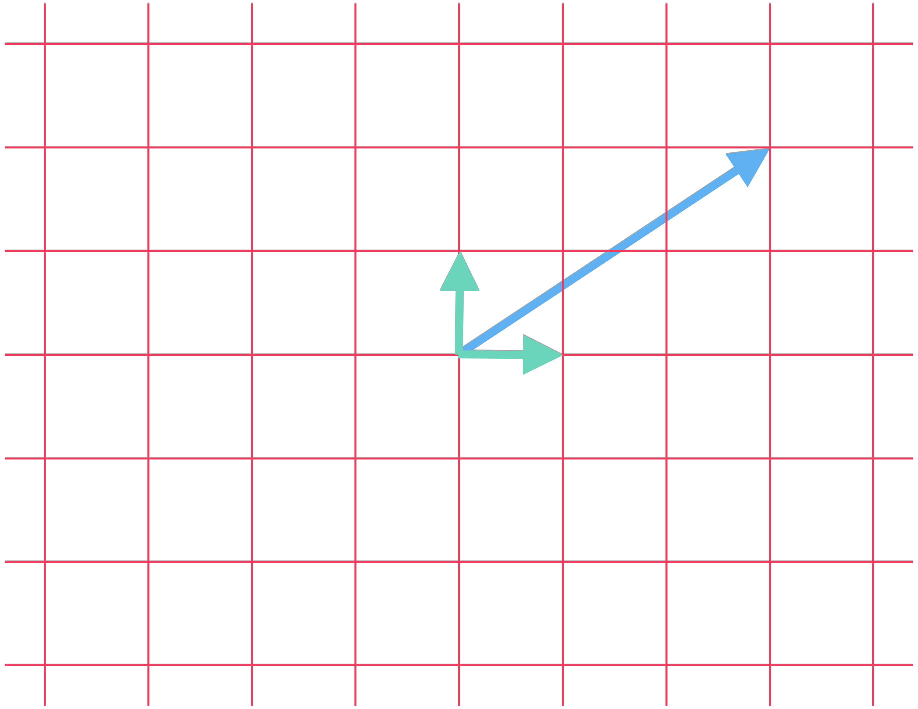
Beware of linear transformations where the vectors $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_d)$ are not linearly independent

In such case, the entire space is mapped to a lower-dimensional hyperplane/line

> *Sure, this low dimensional (hyper)plane/line could be sitting inside a higher dimensional space but it would itself be low dimensional*

This could be the case even if k is a large number, even if $k \gg d$

Thus, notice that the possible values of $A\mathbf{v}$ are exactly $\mathrm{span}\big(f(\mathbf{e}_1), \ldots, f(\mathbf{e}_d)\big)$



e.g. $A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}$

Special Linear Transformations

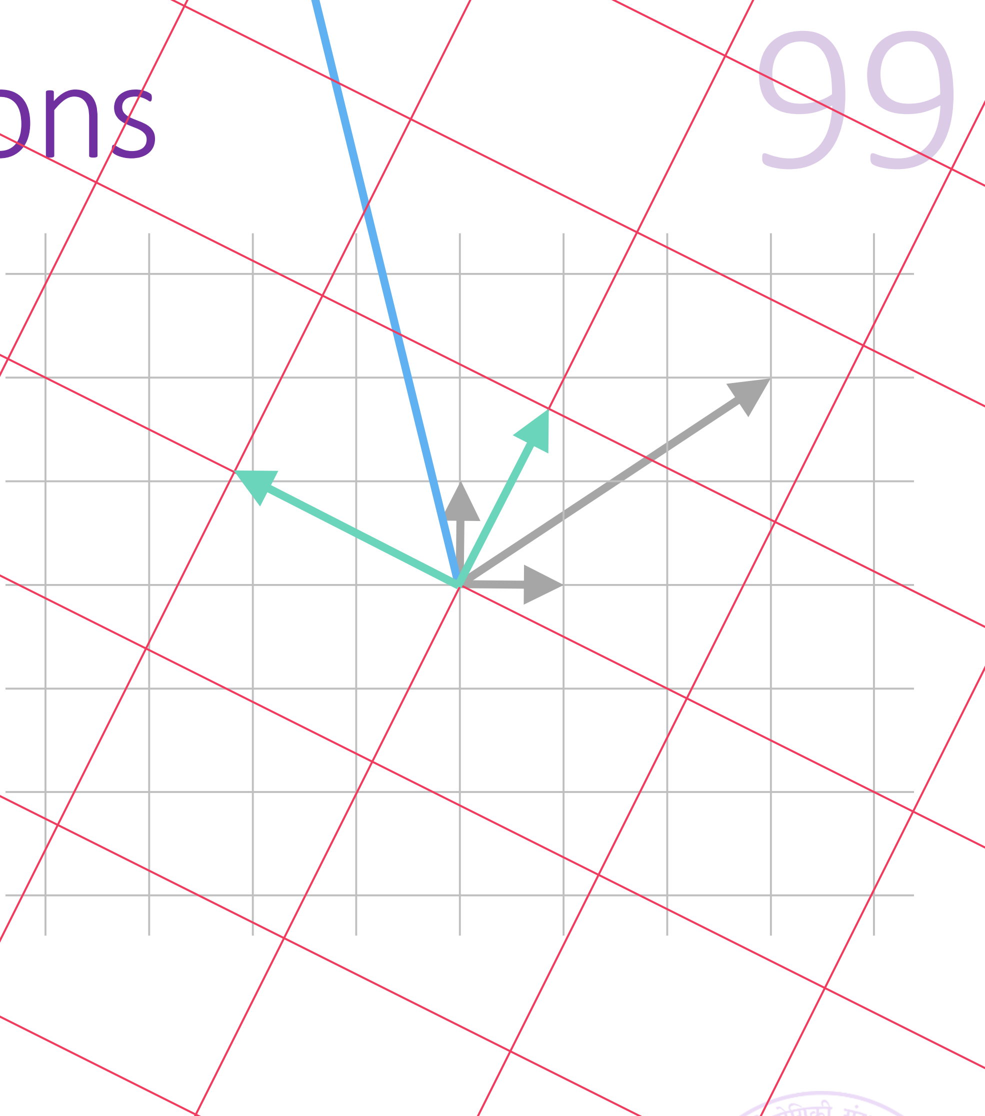
Beware of linear transformations where the vectors $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_d)$ are not linearly independent

In such case, the entire space is mapped to a lower-dimensional hyperplane/line

Sure, this low dimensional (hyper)plane/line could be sitting inside a higher dimensional space but it would itself be low dimensional

This could be the case even if k is a large number, even if $k \gg d$

Thus, notice that the possible values of $A\mathbf{v}$ are exactly $\mathbf{span}\big(f(\mathbf{e}_1), \ldots, f(\mathbf{e}_d)\big)$



e.g. $A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}$

Special Linear Transformations

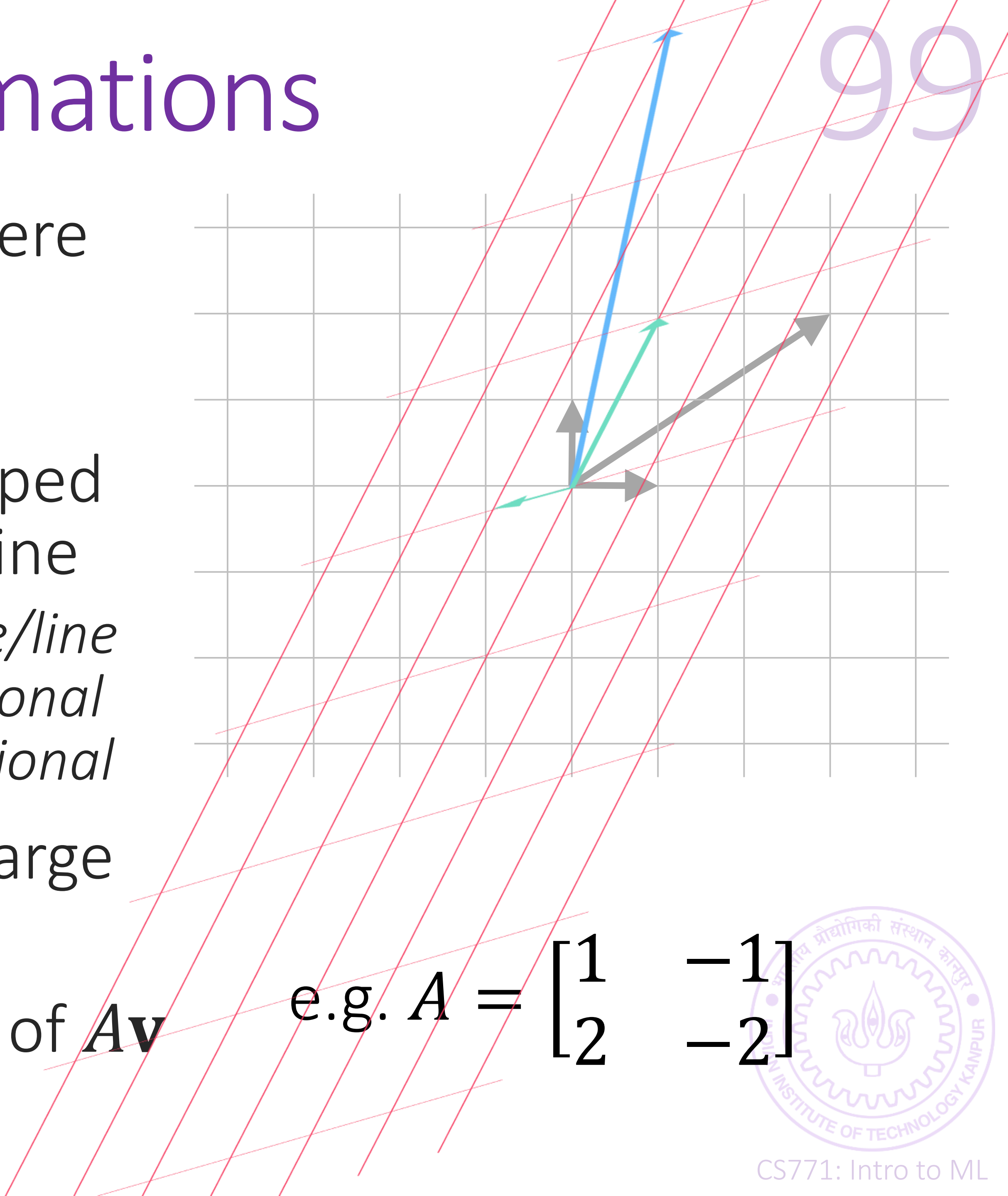
Beware of linear transformations where the vectors $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_d)$ are not linearly independent

In such case, the entire space is mapped to a lower-dimensional hyperplane/line

Sure, this low dimensional (hyper)plane/line could be sitting inside a higher dimensional space but it would itself be low dimensional

This could be the case even if k is a large number, even if $k \gg d$

Thus, notice that the possible values of $A\mathbf{v}$ are exactly $\mathrm{span}\left(f(\mathbf{e}_1), \ldots, f(\mathbf{e}_d)\right)$



e.g. $A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}$

Special Linear Transformations

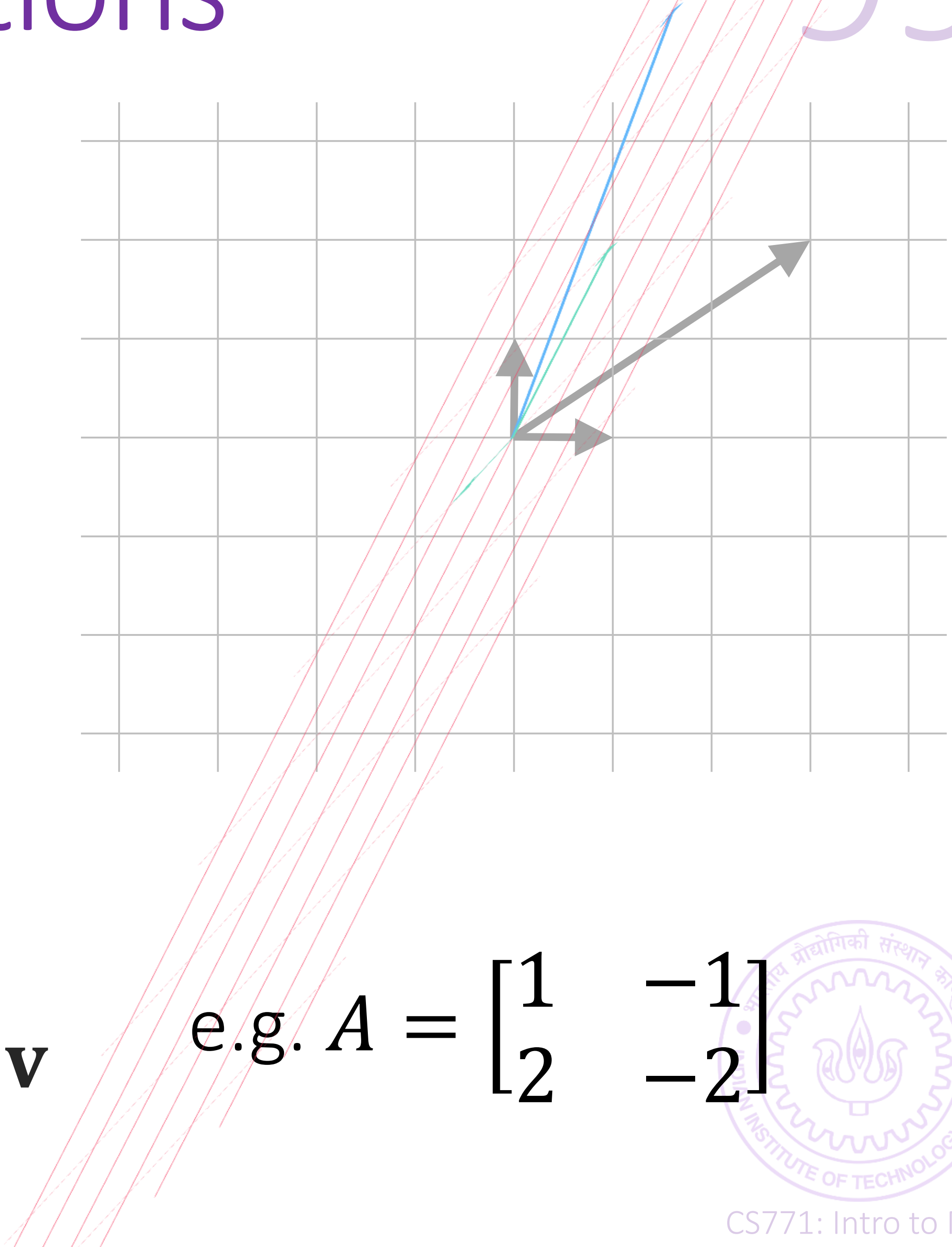
Beware of linear transformations where the vectors $f(\mathbf{e}_1), \dots, f(\mathbf{e}_d)$ are not linearly independent

In such case, the entire space is mapped to a lower-dimensional hyperplane/line

Sure, this low dimensional (hyper)plane/line could be sitting inside a higher dimensional space but it would itself be low dimensional

This could be the case even if k is a large number, even if $k \gg d$

Thus, notice that the possible values of $A\mathbf{v}$ are exactly $\mathbf{span}\big(f(\mathbf{e}_1), \dots, f(\mathbf{e}_d)\big)$



e.g. $A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}$

Special Linear Transformations

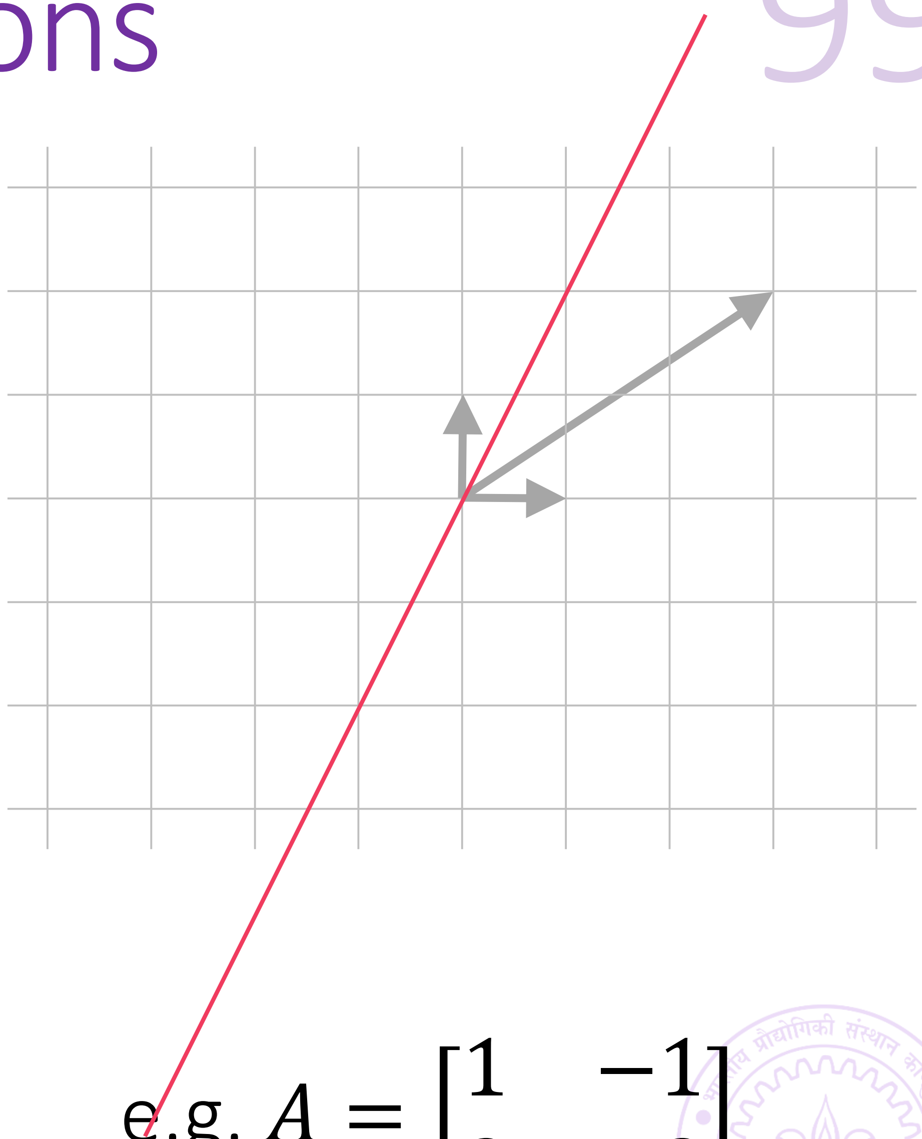
Beware of linear transformations where the vectors $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_d)$ are not linearly independent

In such case, the entire space is mapped to a lower-dimensional hyperplane/line

Sure, this low dimensional (hyper)plane/line could be sitting inside a higher dimensional space but it would itself be low dimensional

This could be the case even if k is a large number, even if $k \gg d$

Thus, notice that the possible values of $A\mathbf{v}$ are exactly $\mathrm{span}\left(f(\mathbf{e}_1), \ldots, f(\mathbf{e}_d)\right)$



e.g. $A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}$

Applying Multiple Linear Transformations

We can apply multiple linear transformations to a single vector

Let $f: \mathbb{R}^d \to \mathbb{R}^k$ and $g: \mathbb{R}^k \to \mathbb{R}^l$ be two linear transformations

f is represented by a matrix $A \in \mathbb{R}^{k \times d}$ and g by a matrix $B \in \mathbb{R}^{l \times k}$

The transformation $h: \mathbf{x} \mapsto g\big(f(\mathbf{x})\big)$ is then simply $B(A\mathbf{x}) = (BA)\mathbf{x}$ i.e. h is represented by the matrix $C = BA$

The claim $B(A\mathbf{x}) = (BA)\mathbf{x}$ requires a proof but it is a relatively simple proof

***Hint**: first show that h must be linear as well and then use linearity*

This is why can't multiply A and C together if $d \neq l$ since $A \in \mathbb{R}^{k \times d}$, $C \in \mathbb{R}^{l \times d}$ (the linear maps do not make sense!)

Also shows why matmul is associative i.e $(PQ)R = P(QR) = PQR$

Can also be used to show (by constructing an example) why matmul is not commutative i.e. if $L, M \in \mathbb{R}^{d \times d}$ why it may be the case that $LM \neq ML$

The Universal Linear Transformations

It turns out that scaling and rotation transformations are all we need to understand in order to understand **any** linear transformation

A superbly powerful result in linear algebra assures us that every linear transformation can be expressed as a composition of two rotation transformations and one scaling transformation

This result is known as the singular value decomposition (SVD) theorem and it underlies a very useful ML technique known as principal component analysis (PCA)