Linear Algebra II

CS771: Introduction to Machine Learning

Purushottam Kar

Recap of Last Lecture

Linear combination of vectors

Special cases: affine combinations, convex combinations

Span of a set of vectors S: all vectors obtainable as linear combinations of S

Incremental trick to finding convex hull/span of a set of vectors

Notions of a set of vectors being linearly dependent/independent

Other definitions e.g. only null combination yields null vector are equivalent

Basis of a set S: a linearly independent set B s.t. $\text{span}(B) \supseteq S$

Gram-Schmidt process to identify an orthonormal basis for a finite set S

Linear Maps/Transformations

Every linear map from $\mathbb{R}^d \to \mathbb{R}^k$ corresponds uniquely to a $k \times d$ matrix

Every $m \times n$ matrix corresponds uniquely to a linear map from $\mathbb{R}^n \to \mathbb{R}^m$

Special kinds of linear maps: scaling maps, rotation maps

More on Compositional Linear Maps

Matrix operations assume very natural interpretations as linear maps

Let $f, g: \mathbb{R}^d \to \mathbb{R}^d$ be linear maps corresp. to matrices $A, B \in \mathbb{R}^{d \times d}$

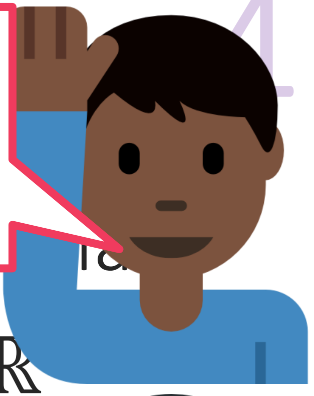
- ***Matrix addition***: *$A + B$ corresponds to the map* $p(\mathbf{x}) \triangleq f(\mathbf{x}) + g(\mathbf{x}), \forall \mathbf{x}$
- ***Matrix multiplication***: *AB corresponds to the map* $q(\mathbf{x}) \triangleq f\big(g(\mathbf{x})\big), \forall \mathbf{x}$
- ***Identity matrix***: *corresponds to the map* $i(\mathbf{x}) = \mathbf{x}, \forall \mathbf{x}$
- ***Matrix scaling***: *$c \cdot A$ corresponds to the map* $r(\mathbf{x}) \triangleq c \cdot f(\mathbf{x}), \forall \mathbf{x}$
- ***Inverse***: *A^{-1} corresp. to $s: \mathbb{R}^d \to \mathbb{R}^d$ s.t.* $f\big(s(\mathbf{x})\big) = s\big(f(\mathbf{x})\big) = i(\mathbf{x}) = \mathbf{x}, \forall \mathbf{x}$
- *You may verify that all the above compositional maps are indeed linear maps!*

Let $h: \mathbb{R}^d \to \mathbb{R}^k$ be a linear map with matrix $C \in \mathbb{R}^{k \times d}$

- ***Matrix transpose***: *$C^\top$ is a map $t: \mathbb{R}^k \to \mathbb{R}^d$*
- *Need not have any special relation with C in general, it is just another map*
- *A symmetric matrix is one whose transpose gives the same map as itself*

More

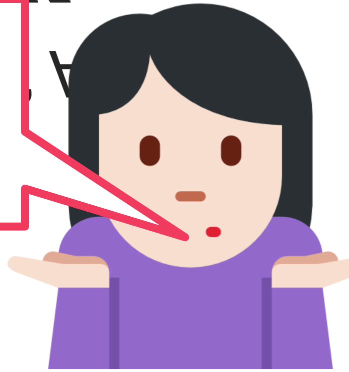
Note that this makes the result $(AB)^{-1} = B^{-1}A^{-1}$ almost immediately intuitive. If we applied a map g then a map f, then to invert/undo this composite map, we must first undo f then undo g ☺



Matrix op

Let $f, g: \mathbb{R}^d \to \mathbb{R}^d$ be linear maps corresp. to matrices $A, B \in \mathbb{R}$

Can you see why some of the other commonly known results also make sense this way? For example why $(A+B)^{-1} \neq A^{-1} + B^{-1}$ or why $c \cdot (A+B) = c \cdot A + c \cdot B$ or why $(c+d) \cdot A = c \cdot A + d \cdot A$ or why $A + B = B + A$



Identity matrix: corresponds to the map $i(\mathbf{x}) = \mathbf{x}, \forall \mathbf{x}$

Matrix scaling: $c \cdot A$ *corresponds to the map* $r(\mathbf{x}) \triangleq c \cdot f(\mathbf{x}), \forall \mathbf{x}$

Inverse: A^{-1} *corresp. to* $s: \mathbb{R}^d \to \mathbb{R}^d$ *s.t.* $f(s(\mathbf{x})) = s(f(\mathbf{x})) = i(\mathbf{x}) = \mathbf{x}, \forall \mathbf{x}$

You may verify that all the above compositional maps are indeed linear maps!

Let $h: \mathbb{R}^d \to \mathbb{R}^k$ be a linear map with matrix $C \in \mathbb{R}^{k \times d}$

Matrix transpose: $C^\top$ *is a map* $t: \mathbb{R}^k \to \mathbb{R}^d$

Need not have any special relation with C *in general, it is just another map*

A symmetric matrix is one whose transpose gives the same map as itself

More on Compositional Linear Maps

A vector (column by default) $\mathbf{v} \in \mathbb{R}^n = \mathbb{R}^{n\times 1}$ is a map from $\mathbb{R} \to \mathbb{R}^n$

It maps a scalar $c \in \mathbb{R}$ to a vector $c \cdot \mathbf{v} \in \mathbb{R}^n$

A row vector $\mathbf{u}^\top \in \mathbb{R}^{1\times n}$ is a map from $\mathbb{R}^n \to \mathbb{R}$

It maps a vector $\mathbf{v} \in \mathbb{R}^n$ to a scalar $\mathbf{u}^\top\mathbf{v} \in \mathbb{R}$

Thus, inner (aka dot) products can be seen as the application of a linear map!

Outer product: given a col. vector $\mathbf{v} \in \mathbb{R}^m$ and row vector $\mathbf{u}^\top \in \mathbb{R}^{1\times n}$, can think of $\mathbf{v}\mathbf{u}^\top$ as a composition map from $\mathbb{R}^n \to \mathbb{R} \to \mathbb{R}^m$ i.e. from $\mathbb{R}^n \to \mathbb{R}^m$. Indeed, for any $\mathbf{x} \in \mathbb{R}^n$ we have $\mathbf{v}\mathbf{u}^\top\mathbf{x} = (\mathbf{u}^\top\mathbf{x}) \cdot \mathbf{v} \in \mathbb{R}^m$

Not surprising then that $U = \mathbf{v}\mathbf{u}^\top \in \mathbb{R}^{m\times n}$ is a matrix

Example for matmul not symmetric: $UU^\top = \mathbf{v}\mathbf{u}^\top\mathbf{u}\mathbf{v}^\top = (\mathbf{u}^\top\mathbf{u}) \cdot \mathbf{v}\mathbf{v}^\top$ whereas $U^\top U = \mathbf{u}\mathbf{v}^\top\mathbf{v}\mathbf{u}^\top = (\mathbf{v}^\top\mathbf{v}) \cdot \mathbf{u}\mathbf{u}^\top$. $UU^\top$ and $U^\top U$ not even same dims if $m \neq n$

Can you show above identities using the linear map interpretation of matrices?

Column Space

Recall: every linear map $f: \mathbb{R}^d \to \mathbb{R}^k$ is represented by a matrix, say A of col. vectors telling us how f maps canonical vectors i.e. $\mathbf{e}_i, i \in [d]$

$$A = \left[\begin{bmatrix} \\ f(\mathbf{e}_1) \\ \\ \end{bmatrix} \begin{bmatrix} \\ f(\mathbf{e}_2) \\ \\ \end{bmatrix} \cdots \begin{bmatrix} \\ f(\mathbf{e}_d) \\ \\ \end{bmatrix} \right]$$

Note: $\sum_{i=1}^{d} v_i \cdot f(\mathbf{e}_i)$ is the same as $A\mathbf{v}$ where $\mathbf{v} = (v_1, \ldots, v_d) \in \mathbb{R}^d$

This means as $\mathbf{v}$ takes all possible values $\in \mathbb{R}^d$, $A\mathbf{v}$ will take all possible values in $\mathbf{span}\big(f(\mathbf{e}_1), \ldots, f(\mathbf{e}_d)\big)$ – called the *column space* of A

The column space of A tells us all possible outputs that map can give!

Rank

Given a set of vectors S, the **number** of vectors in its basis is called the *rank* of the set S – often written as $\mathbf{rank}(S)$ or even $\mathbf{dim}(S)$

Warning: a set of vectors can have more than one basis. However, all of them will have the same size. Gram-Schmidt will also give a basis of that same size

Special case 1: $\dim(S) = 0$. *This can happen only if the set contains only one element (or many copies of it) and that element is the zero vector* $\mathbf{0}$

Special case 2: $\dim(S) = 1$. This happens when vectors in S all lie along a line that passes through origin, even if the vectors are d-dimensional for $d > 1$

Special case 3: $\dim(S) = 2$. This happens when vectors in S all lie on a 2D plane passing through origin, even if the vectors are d-dimensional for $d > 2$

Not so special case: $\dim(S) = k$. This happens when vectors in S are all d-dim vectors but all lie on a k-dimensional subspace of $\mathbb{R}^d$

For example, vectors in $\mathbb{R}^5$ lying on a 3-dimensional subspace

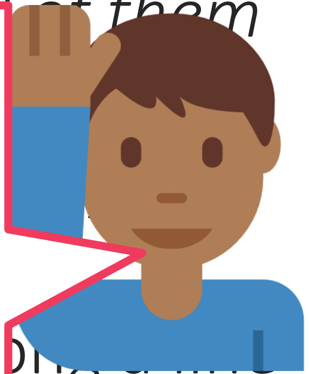
Rank

Can you see why we must always have $\dim(S) = \dim(\text{span}(S))$?
Can you see why we must always have $\dim(S) \leq |S|$?

Given a set of vectors S, the number of vectors in its basis is cal… *rank* of the set S – often written as $\text{rank}(S)$ or even $\dim(S)$

…ll of them

Be careful about lines/planes that do not pass through origin. The rank of a line that does not pass through the origin is 2 not 1. Verify this yourself by thinking of the line $x + y = 1$. It contains the points $(1,0)$ as well as $(0,1)$ which means the span of points on this line is the entire space $\mathbb{R}^2$ i.e. its rank is 2. Similarly, a plane in 3D that does not pass through the origin has rank 3 not 2



…ong a line that passes through origin, even if the vectors are d-dimensional for $d > 1$

Special case 3: $\dim(S) = 2$. This happens when vectors in S all lie on a 2D plane passing through origin, even if the vectors are d-dimensional for $d > 2$

Not so special case: $\dim(S) = k$. This happens when vectors in S are all d-dim vectors but all lie on a k-dimensional subspace of $\mathbb{R}^d$

For example, vectors in $\mathbb{R}^5$ lying on a 3-dimensional subspace

Column Rank

For a matrix A, rank of its set of columns is called *column rank* of A

For a matrix A, rank of its set of rows (can be defined as the column rank of $A^\top$ whose cols are rows of A) is called *row rank* of A

Claim without proof: the row rank and the column rank of any matrix are always the same so we simply talk about the *rank* of A

A matrix $A \in \mathbb{R}^{m \times n}$ for which rank is equal to either the # of rows or the # of columns i.e. $\text{rank}(A) = \min\{m, n\}$ is called a *full-rank* matrix

Remember, for every set S, *we always have* $\dim(S)$ *(aka* $\text{rank}(S)$*)* $\leq |S|$

A matrix that is not full rank is often called *rank-deficient*

“Spaces”

Term used a lot in linalg – vector *space*, column *space, subspace*

A *space* is simply a set of vectors that is “self-contained” in two senses

If we take any vector in that space and scale it by any real number, the resulting vector is already there in that space

If we take any two vectors in that space and add them together, the resulting vector is already there in that space

A fancy way of saying the above is to say that a space is “closed” with respect to vector addition and scalar multiplication

Note that zero vector must be inside every space S

Proof: Take any $\mathbf{v} \in S$ i.e. $-\mathbf{v} \in S$ *too i.e.* $\mathbf{v} + (-\mathbf{v}) \in S$ too i.e. $\mathbf{0} \in S$

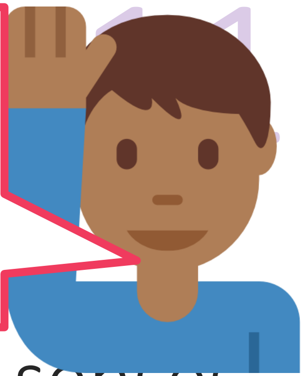
Can you see why every line passing through the origin is a space?

Can you see why a line segment/line not passing through origin is not?

“Spaces

Term used

Vector spaces like $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^d$ etc are all spaces. Spaces can contain other spaces too (the contained spaces are called *subspaces*). For example, the set $\left\{\mathbf{v} \in \mathbb{R}^5 : \sum_{j=1}^{5} \mathbf{v}_j = 0\right\}$ is a subspace of $\mathbb{R}^5$ (verify)



A *space* is simply a set of vectors that is “self-contained” in two senses

If we take any vector in that space and scale it by any real number, the resulting vector is already there in that space

If we take any two vectors in that space and add them together, the resulting vector is already there in that space

A fancy way of saying the above is to say that a space is “closed” with respect to vector addition and scalar multiplication

Note that zero vector must be inside every space S

Proof: Take any $\mathbf{v} \in S$ i.e. $-\mathbf{v} \in S$ *too i.e.* $\mathbf{v} + (-\mathbf{v}) \in S$ too i.e. $\mathbf{0} \in S$

Can you see why every line passing through the origin is a space?

Can you see why a line segment/line not passing through origin is not?

Terms Related to Spaces

Affine Spaces: a “displaced” (sub)space. Suppose $S \subseteq \mathbb{R}^d$ is a space. Then for any $\mathbf{v} \in \mathbb{R}^d$, the set $T_{\mathbf{v}} \triangleq \{\mathbf{x} + \mathbf{v} : \mathbf{x} \in S\}$ is an affine space.

Note: spaces are affine spaces (take $\mathbf{v} \in S$ e.g. $\mathbf{v} = \mathbf{0}$)

Hyperplane: in $\mathbb{R}^d$ is a subspace of rank $d - 1$

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} = 0\}$ for some $\mathbf{w} \in \mathbb{R}^d$

Affine Hyperplane: a “displaced” hyperplane

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} + b = 0\}$ for some $\mathbf{w} \in \mathbb{R}^d$ and some $b \in \mathbb{R}$. Note that $\mathbf{w}^\top \mathbf{x} = -b$ iff $\mathbf{x} = \mathbf{v} - \frac{b}{\|\mathbf{w}\|_2^2} \cdot \mathbf{w}$ for some $\mathbf{v}$ such that $\mathbf{w}^\top \mathbf{v} = 0$

In ML, often we use the term hyperplane when we should really be using the term affine hyperplane ☺

Terms Related to Spaces

Affine Spaces: a "displaced" (sub)space. Suppose $S \subseteq \mathbb{R}^d$ is a space. Then for any $\mathbf{v} \in \mathbb{R}^d$, the set $T_{\mathbf{v}} \triangleq \{\mathbf{x} + \mathbf{v} : \mathbf{x} \in S\}$ is an affine space.

Note: spaces are affine spaces (take $\mathbf{v} \in S$ e.g. $\mathbf{v} = \mathbf{0}$)

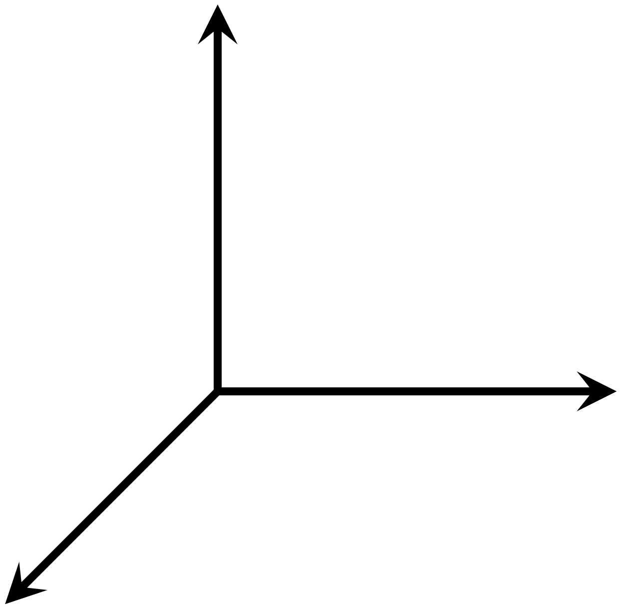
Hyperplane: in $\mathbb{R}^d$ is a subspace of rank $d - 1$

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} = 0\}$ for some $\mathbf{w} \in \mathbb{R}^d$

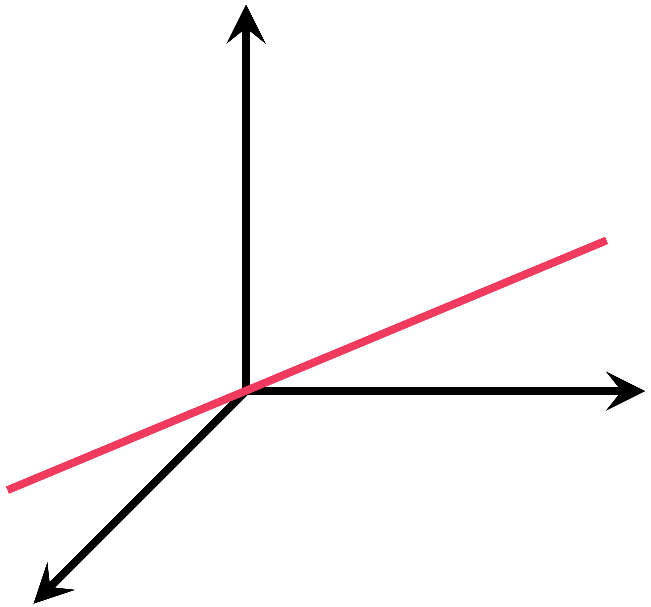
Affine Hyperplane: a "displaced" hyperplane

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} + b = 0\}$ for some $\mathbf{w} \in \mathbb{R}^d$ and some $b \in \mathbb{R}$. Note that $\mathbf{w}^\top \mathbf{x} = -b$ iff $\mathbf{x} = \mathbf{v} - \frac{b}{\|\mathbf{w}\|_2^2} \cdot \mathbf{w}$ for some $\mathbf{v}$ such that $\mathbf{w}^\top \mathbf{v} = 0$

In ML, often we use the term hyperplane when we should really be using the term affine hyperplane ☺



Terms Related to Spaces



Affine Spaces: a "displaced" (sub)space. Suppose $S \subseteq \mathbb{R}^d$ is a space. Then for any $\mathbf{v} \in \mathbb{R}^d$, the set $T_{\mathbf{v}} \triangleq \{\mathbf{x} + \mathbf{v} : \mathbf{x} \in S\}$ is an affine space.

Note: spaces are affine spaces (take $\mathbf{v} \in S$ e.g. $\mathbf{v} = \mathbf{0}$)

Hyperplane: in $\mathbb{R}^d$ is a subspace of rank $d - 1$

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} = 0\}$ for some $\mathbf{w} \in \mathbb{R}^d$

Affine Hyperplane: a "displaced" hyperplane

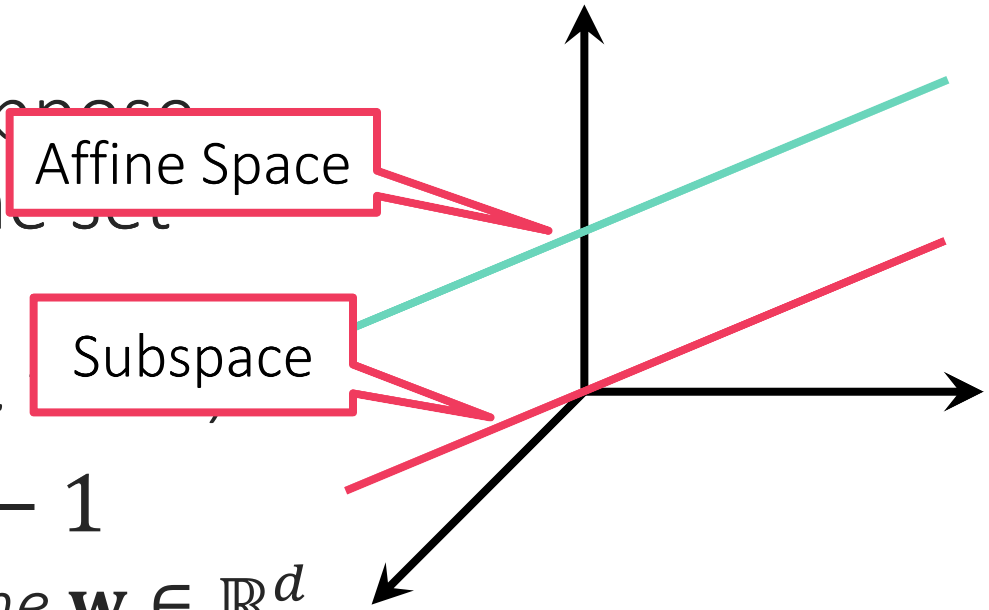
All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} + b = 0\}$ for some $\mathbf{w} \in \mathbb{R}^d$ and some $b \in \mathbb{R}$. Note that $\mathbf{w}^\top \mathbf{x} = -b$ iff $\mathbf{x} = \mathbf{v} - \frac{b}{\|\mathbf{w}\|_2^2} \cdot \mathbf{w}$ for some $\mathbf{v}$ such that $\mathbf{w}^\top \mathbf{v} = 0$

In ML, often we use the term hyperplane when we should really be using the term affine hyperplane ☺

Terms Related to Spaces

Affine Spaces: a "displaced" (sub)space. Sup
$S \subseteq \mathbb{R}^d$ is a space. Then for any $\mathbf{v} \in \mathbb{R}^d$, the
$T_{\mathbf{v}} \triangleq \{\mathbf{x} + \mathbf{v} : \mathbf{x} \in S\}$ is an affine space.

Note: spaces are affine spaces (take $\mathbf{v} \in S$ e.g.



Hyperplane: in $\mathbb{R}^d$ is a subspace of rank $d - 1$

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} = 0\}$ for some $\mathbf{w} \in \mathbb{R}^d$

Affine Hyperplane: a "displaced" hyperplane

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} + b = 0\}$ for some $\mathbf{w} \in \mathbb{R}^d$ and some $b \in \mathbb{R}$. Note that $\mathbf{w}^\top \mathbf{x} = -b$ iff $\mathbf{x} = \mathbf{v} - \frac{b}{\|\mathbf{w}\|_2^2} \cdot \mathbf{w}$ for some $\mathbf{v}$ such that $\mathbf{w}^\top \mathbf{v} = 0$

In ML, often we use the term hyperplane when we should really be using the term affine hyperplane ☺

Terms Related to Spaces

Affine Spaces: a "displaced" (sub)space. Suppose $S \subseteq \mathbb{R}^d$ is a space. Then for any $\mathbf{v} \in \mathbb{R}^d$, the set $T_{\mathbf{v}} \triangleq \{\mathbf{x} + \mathbf{v} : \mathbf{x} \in S\}$ is an affine space.

Note: *spaces are affine spaces (take* $\mathbf{v} \in S$ *e.g.*)

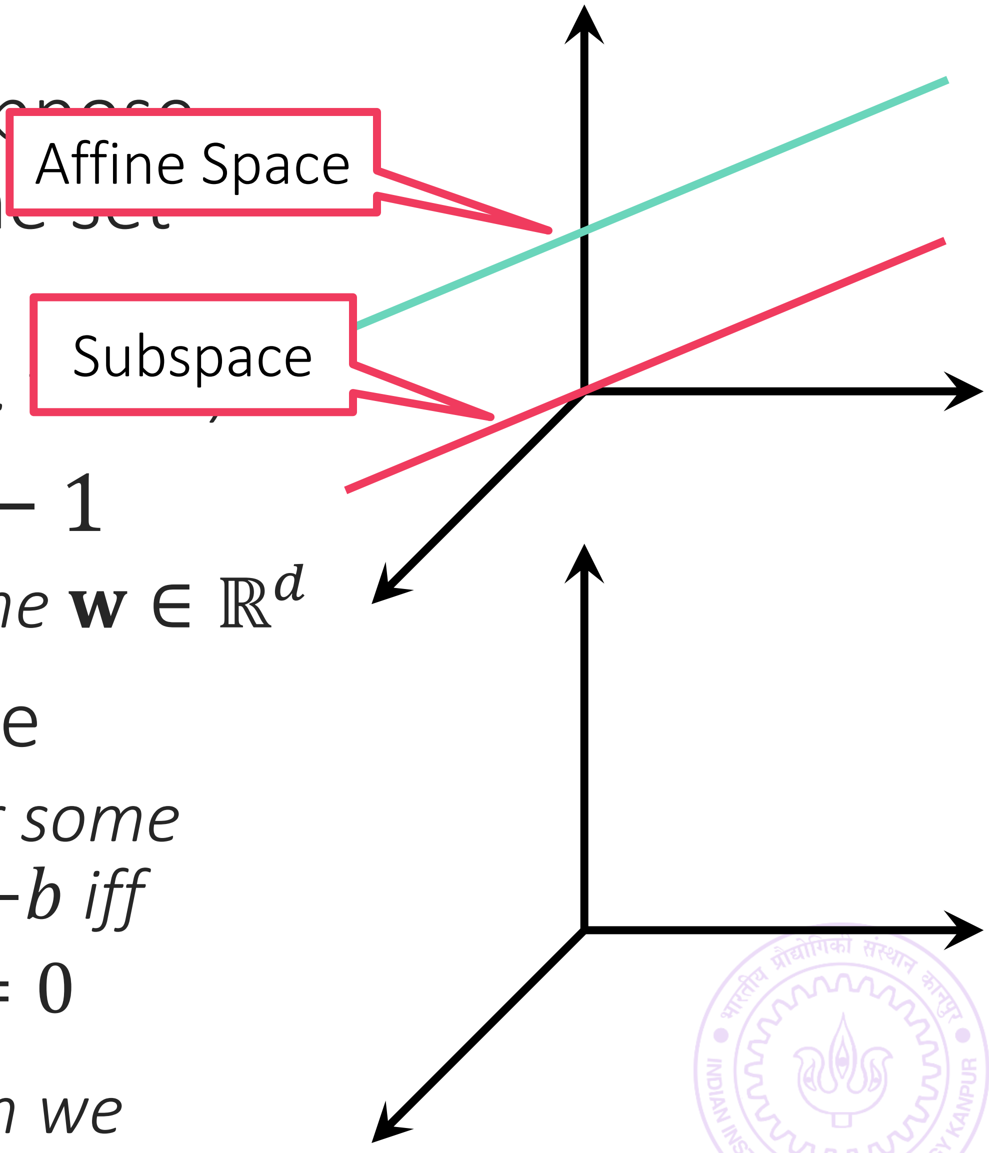
Hyperplane: in $\mathbb{R}^d$ is a subspace of rank $d - 1$

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} = 0\}$ *for some* $\mathbf{w} \in \mathbb{R}^d$

Affine Hyperplane: a "displaced" hyperplane

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} + b = 0\}$ *for some* $\mathbf{w} \in \mathbb{R}^d$ *and some* $b \in \mathbb{R}$. *Note that* $\mathbf{w}^\top \mathbf{x} = -b$ *iff* $\mathbf{x} = \mathbf{v} - \frac{b}{\|\mathbf{w}\|_2^2} \cdot \mathbf{w}$ *for some* $\mathbf{v}$ *such that* $\mathbf{w}^\top \mathbf{v} = 0$

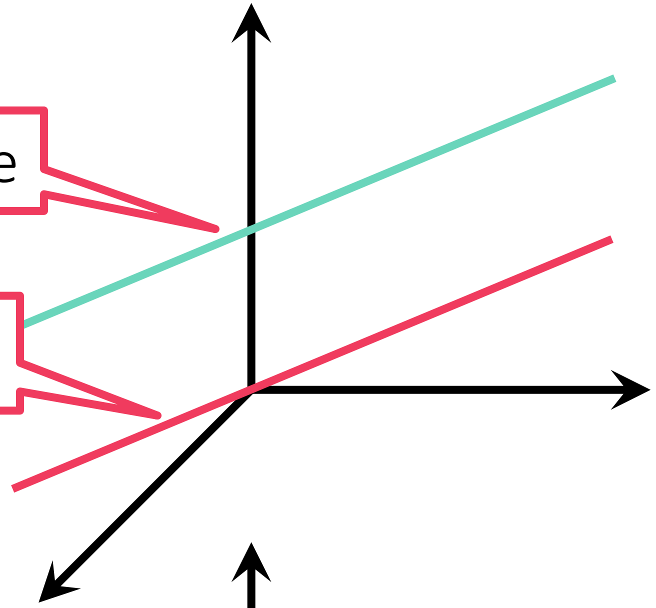
In ML, often we use the term hyperplane when we should really be using the term affine hyperplane ☺



Terms Related to Spaces

Affine Spaces: a "displaced" (sub)space. Suppose $S \subseteq \mathbb{R}^d$ is a space. Then for any $\mathbf{v} \in \mathbb{R}^d$, the set $T_\mathbf{v} \triangleq \{\mathbf{x} + \mathbf{v} : \mathbf{x} \in S\}$ is an affine space.

Note: *spaces are affine spaces (take $\mathbf{v} \in S$ e.g.*

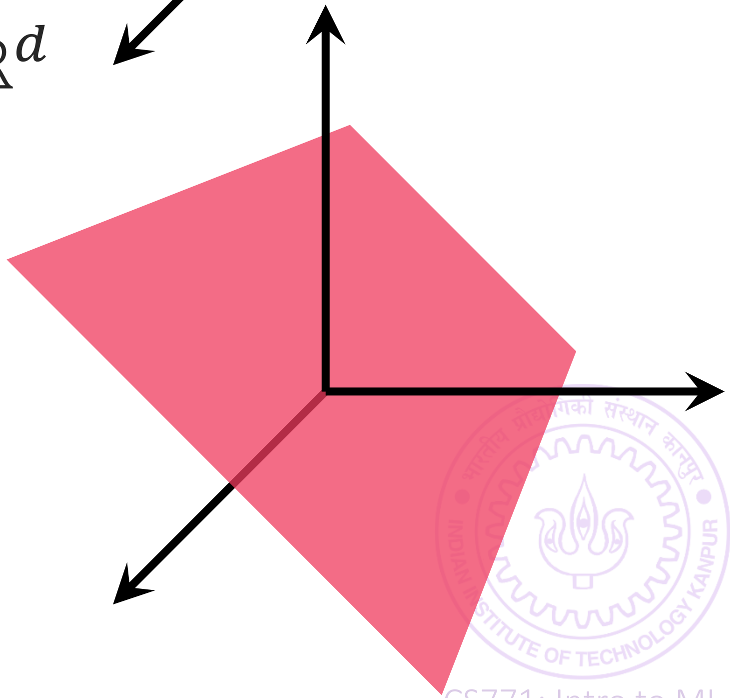


Hyperplane: in $\mathbb{R}^d$ is a subspace of rank $d - 1$

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} = 0\}$ *for some* $\mathbf{w} \in \mathbb{R}^d$

Affine Hyperplane: a "displaced" hyperplane

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} + b = 0\}$ *for some* $\mathbf{w} \in \mathbb{R}^d$ *and some* $b \in \mathbb{R}$. *Note that* $\mathbf{w}^\top \mathbf{x} = -b$ *iff* $\mathbf{x} = \mathbf{v} - \frac{b}{\|\mathbf{w}\|_2^2} \cdot \mathbf{w}$ *for some* $\mathbf{v}$ *such that* $\mathbf{w}^\top \mathbf{v} = 0$



In ML, often we use the term hyperplane when we should really be using the term affine hyperplane ☺

Terms Related to Spaces

Affine Spaces: a "displaced" (sub)space. Suppose $S \subseteq \mathbb{R}^d$ is a space. Then for any $\mathbf{v} \in \mathbb{R}^d$, the set $T_{\mathbf{v}} \triangleq \{\mathbf{x} + \mathbf{v} : \mathbf{x} \in S\}$ is an affine space.

Note: spaces are affine spaces (take $\mathbf{v} \in S$ e.g.

Hyperplane: in $\mathbb{R}^d$ is a subspace of rank

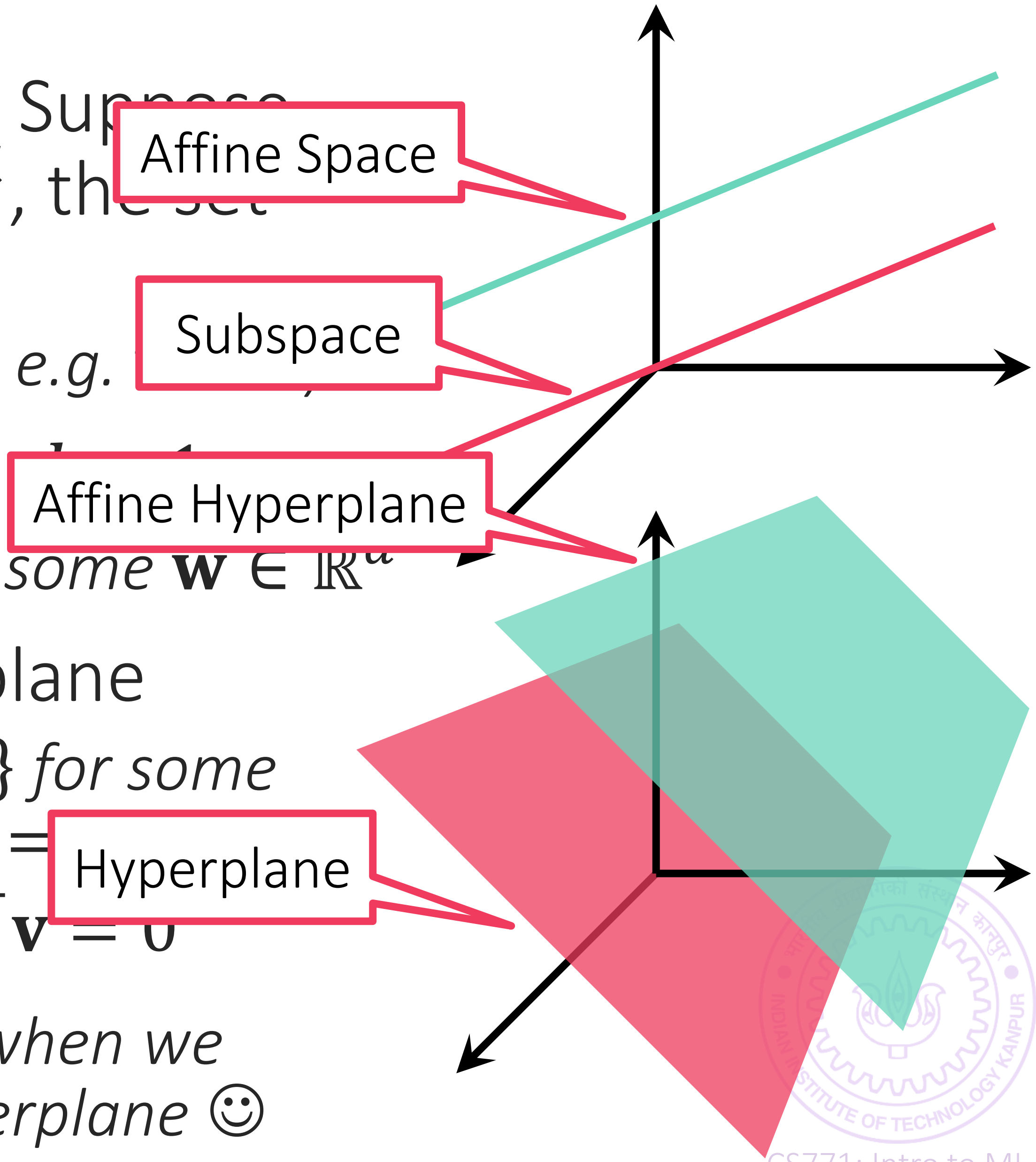
All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} = 0\}$ for some $\mathbf{w} \in \mathbb{R}^d$

Affine Hyperplane: a "displaced" hyperplane

All hyperplanes look like $\{\mathbf{x} : \mathbf{w}^\top \mathbf{x} + b = 0\}$ for some $\mathbf{w} \in \mathbb{R}^d$ and some $b \in \mathbb{R}$. Note that $\mathbf{w}^\top \mathbf{x} =$

$\mathbf{x} = \mathbf{v} - \frac{b}{\|\mathbf{w}\|_2^2} \cdot \mathbf{w}$ for some $\mathbf{v}$ such that $\mathbf{w}^\top \mathbf{v} = 0$

In ML, often we use the term hyperplane when we should really be using the term affine hyperplane ☺



Null Space

For any linear map $\mathbb{R}^d \to \mathbb{R}^k$ with matrix $A \in \mathbb{R}^{k \times d}$, its *null space* (aka *kernel*) is the set of inputs that get mapped to zero vector

$$\ker(A) = \{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} = \mathbf{0}\}$$

Can you show that $\ker(A)$ *is always a subspace of* $\mathbb{R}^d$*?*

The rank/dimension of $\ker(A)$ is called the *nullity* of A

If $\dim(\ker(A)) = 0$ *then this means that* A *only maps* $\mathbf{0} \in \mathbb{R}^d$ *to* $\mathbf{0} \in \mathbb{R}^k$

If $\dim(\ker(A)) = 1$ *then this means that* A *maps an entire line passing through the origin (i.e. a subspace of dimension/rank 1) to* $\mathbf{0}$

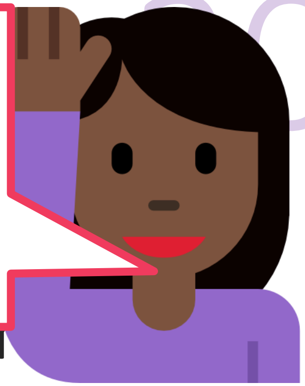
If $\dim(\ker(A)) = 2$ *then this means that* A *maps an entire 2D plane passing through the origin (i.e. a subspace of dimension/rank 2) to* $\mathbf{0}$ *... etc etc*

Rank-Nullity Theorem: $\dim(\ker(A)) + \operatorname{rank}(A) = d$

This implies that a matrix with nullity zero will always be full rank!

Null Spac

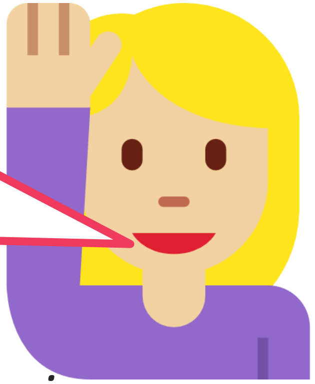
Elements in the null space are impotent in the sense that they cannot change the output of the linear map. Note that we have $A\mathbf{u} = A(\mathbf{u} + \mathbf{v})$ for any $\mathbf{u} \in \mathbb{R}^d$ and any $\mathbf{v} \in \ker(A)$.



For any linear

(aka *kernel*) is the set of inputs that get mapped to zero vector

Note that if $\dim(\ker(A)) > 0$, this means that the kernel has an infinite number of elements which means that for any $\mathbf{u} \in \mathbb{R}^d$, there exist infinitely many other vectors $\tilde{\mathbf{u}} \in \mathbb{R}^d$ such that $A\mathbf{u} = A\tilde{\mathbf{u}}$. Each such $\tilde{\mathbf{u}}$ is found by adding a kernel element to $\mathbf{u}$ (to see this, note $A\mathbf{u} = A\tilde{\mathbf{u}} \Rightarrow A(\mathbf{u} - \tilde{\mathbf{u}}) = \mathbf{0} \Rightarrow \mathbf{u} - \tilde{\mathbf{u}} \in \ker(A)$)



If $\dim(\ker(A)) = 0$ *then this means that* A *only maps* $\mathbf{0} \in \mathbb{R}^d$ *to* $\mathbf{0}$

If $\dim(\ker(A)) = 1$ *then this means that* A *maps an entire line passing through the origin (i.e. a subspace of dimension/rank 1) to* $\mathbf{0}$

If $\dim(\ker(A)) = 2$ *then this means that* A *maps an entire 2D plane passing through the origin (i.e. a subspace of dimension/rank 2) to* $\mathbf{0}$ *… etc etc*

Rank-Nullity Theorem: $\dim(\ker(A)) + \operatorname{rank}(A) = d$

This implies that a matrix with nullity zero will always be full rank!

Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

- *If we take an orthonormal matrix and multiply one or more columns by* -1*, it amounts to flipping that/those axes*

Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

If we take an orthonormal matrix and multiply one or more columns by -1*, it amounts to flipping that/those axes*

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

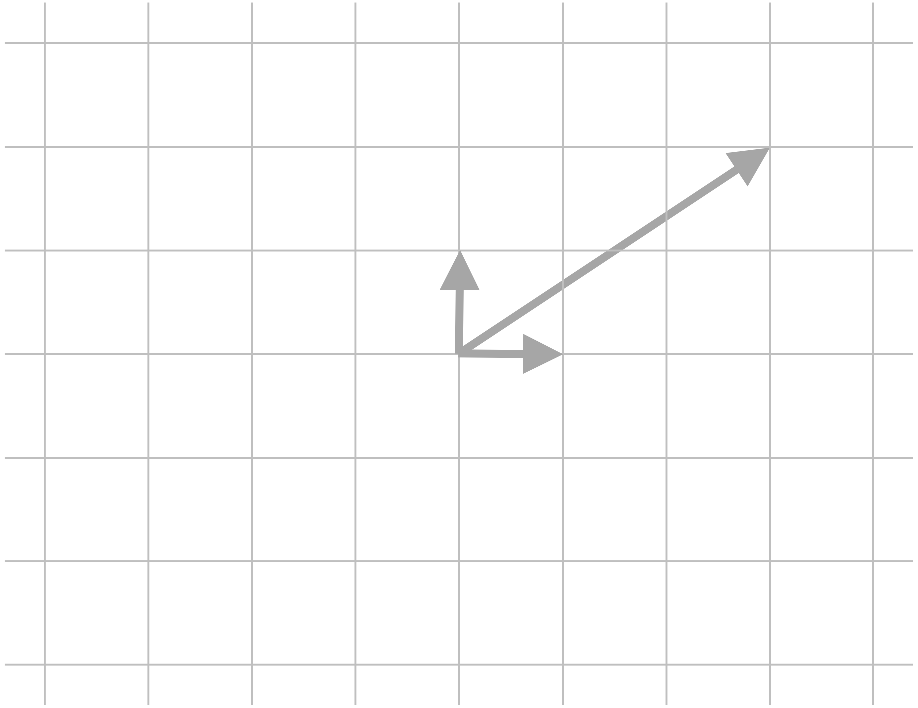
Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

If we take an orthonormal matrix and multiply one or more columns by -1*, it amounts to flipping that/those axes*



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

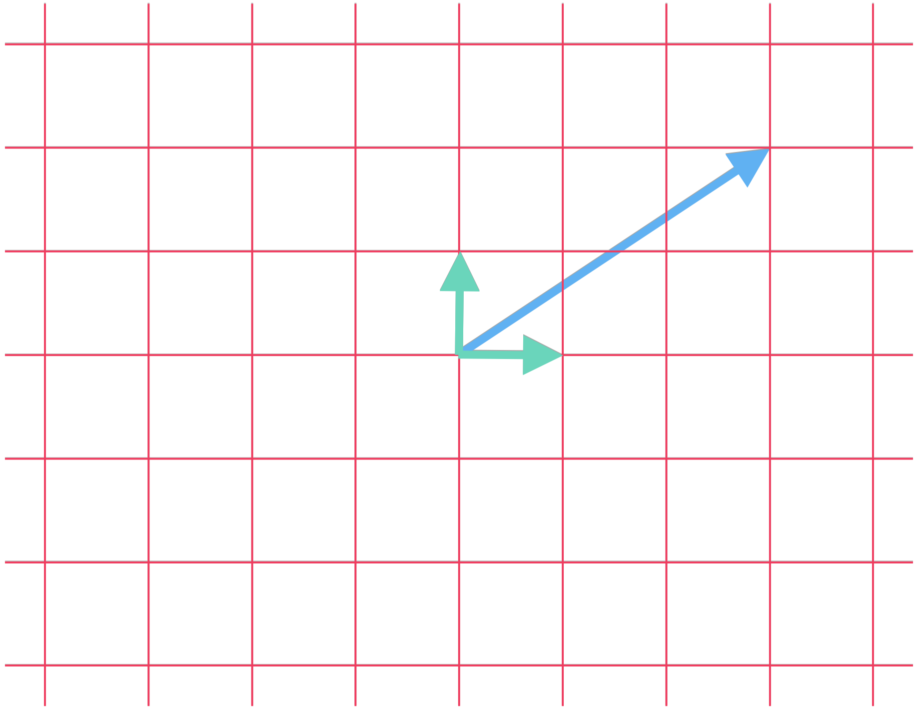
Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

If we take an orthonormal matrix and multiply one or more columns by -1*, it amounts to flipping that/those axes*



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

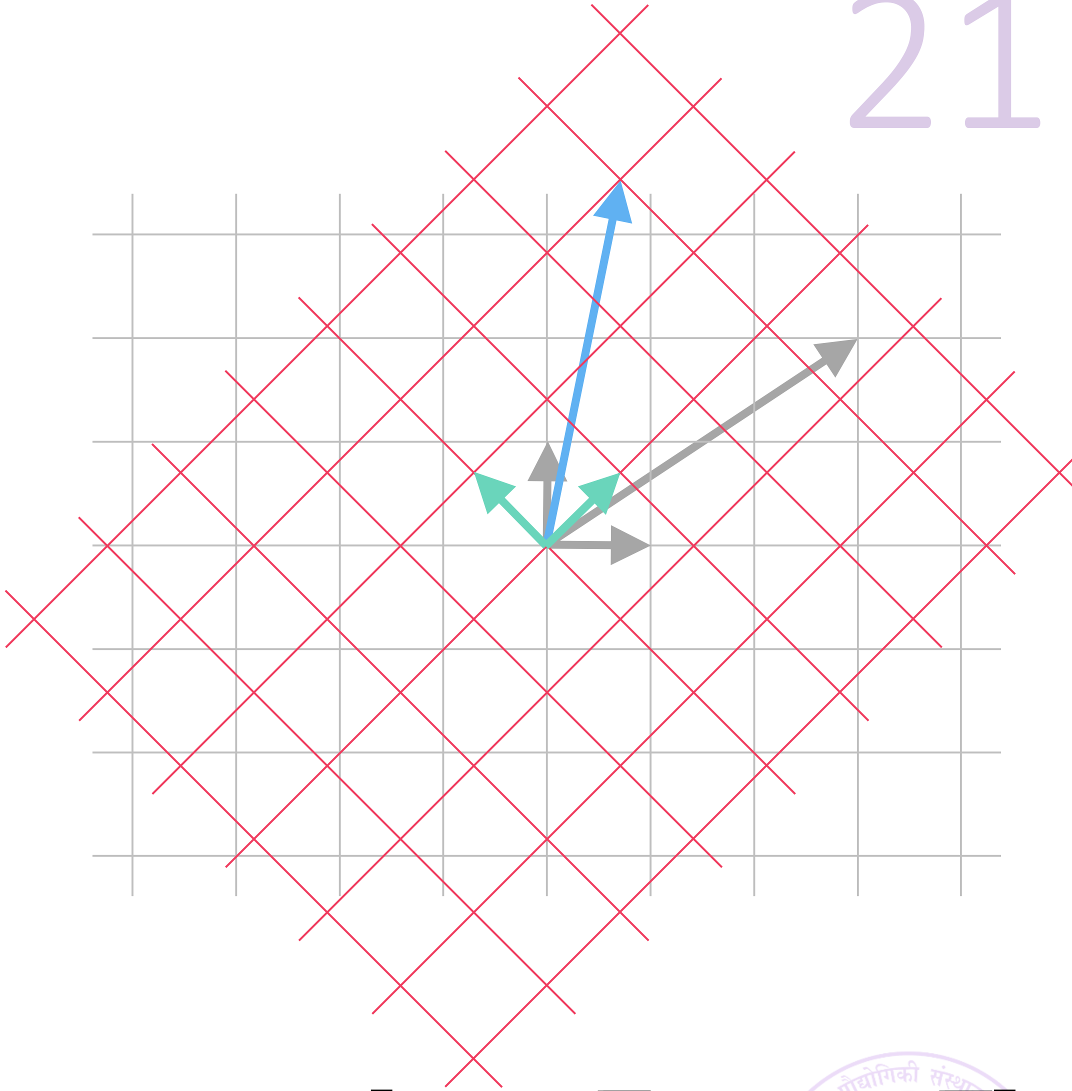
Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

If we take an orthonormal matrix and multiply one or more columns by -1*, it amounts to flipping that/those axes*



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

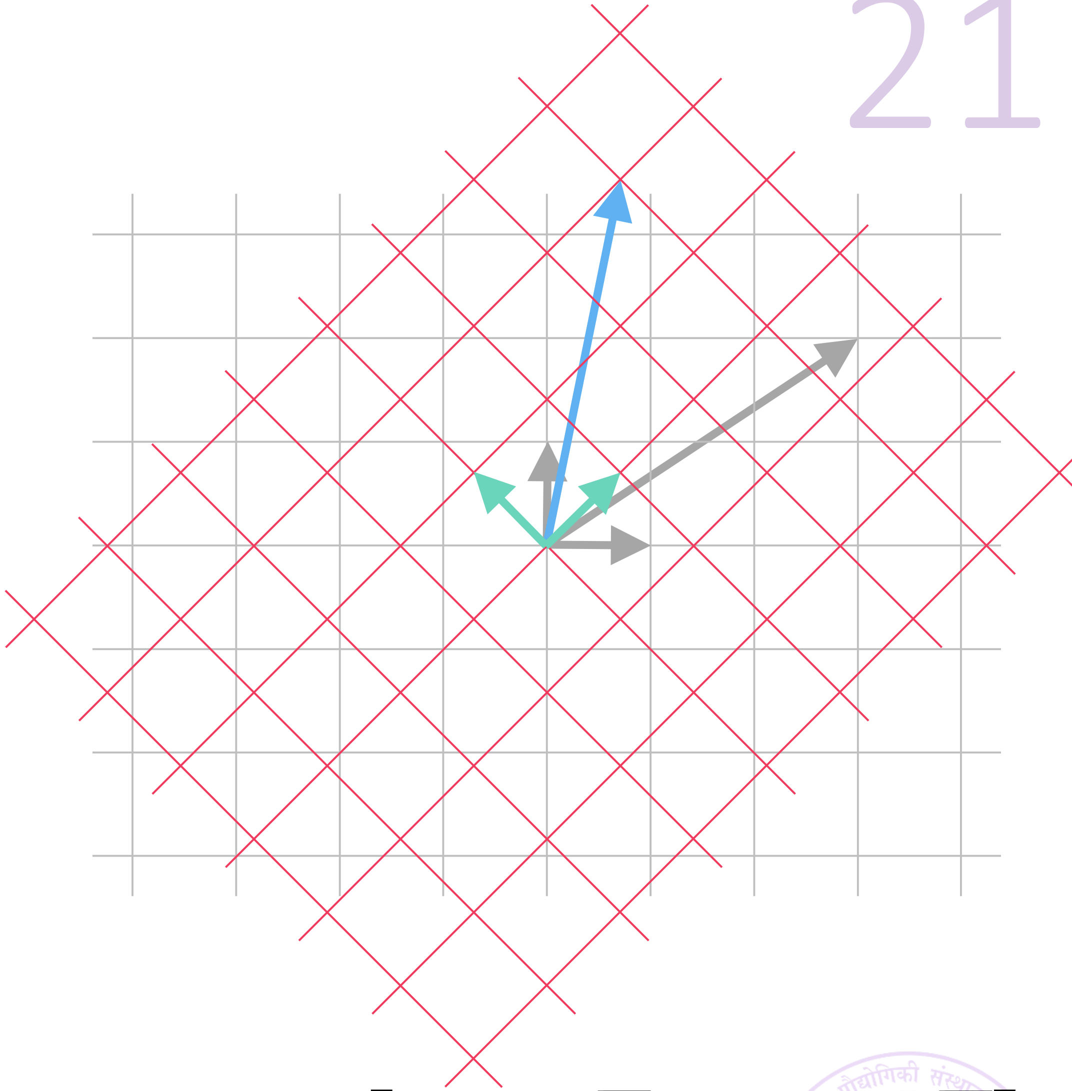
Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

If we take an orthonormal matrix and multiply one or more columns by -1*, it amounts to flipping that/those axes*



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

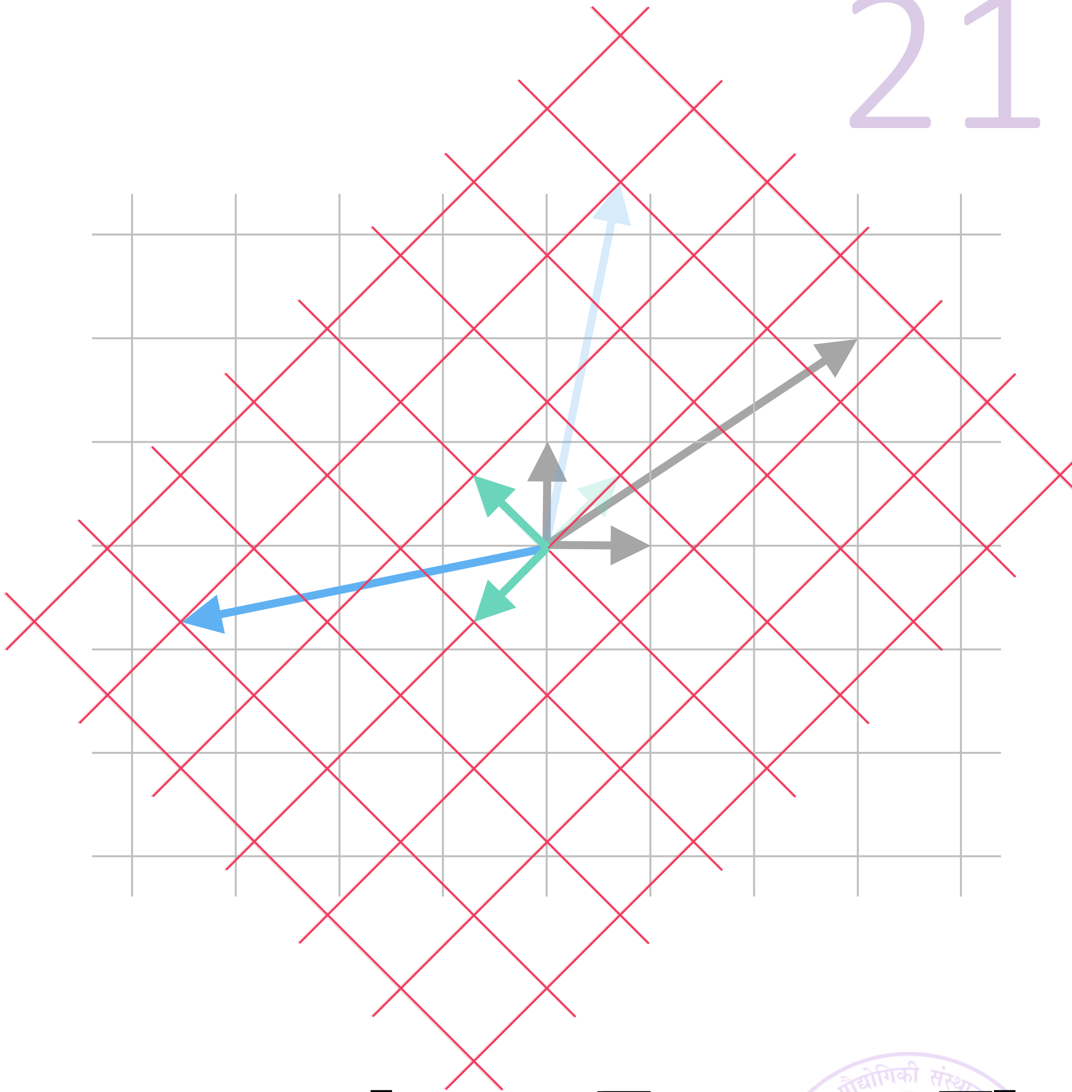
Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

If we take an orthonormal matrix and multiply one or more columns by -1*, it amounts to flipping that/those axes*



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

Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

- *If we take an orthonormal matrix and multiply one or more columns by* -1*, it amounts to flipping that/those axes*
- *If we take an orthonormal matrix and shift its columns around, it amounts to exchanging/reordering those axes*
- *In general, orthonormal matrices give us a new (rotated/flipped/exchanged) basis*

Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

- *If we take an orthonormal matrix and multiply one or more columns by -1, it amounts to flipping that/those axes*
- *If we take an orthonormal matrix and shift its columns around, it amounts to exchanging/reordering those axes*
- *In general, orthonormal matrices give us a new (rotated/flipped/exchanged) basis*

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

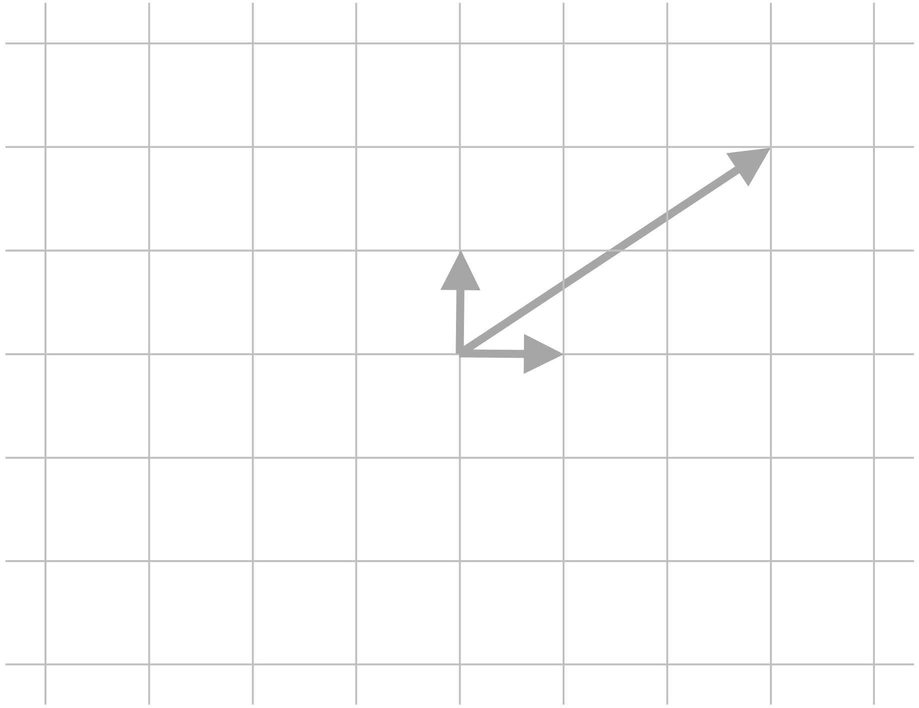
Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

- *If we take an orthonormal matrix and multiply one or more columns by -1, it amounts to flipping that/those axes*
- *If we take an orthonormal matrix and shift its columns around, it amounts to exchanging/reordering those axes*
- *In general, orthonormal matrices give us a new (rotated/flipped/exchanged) basis*



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

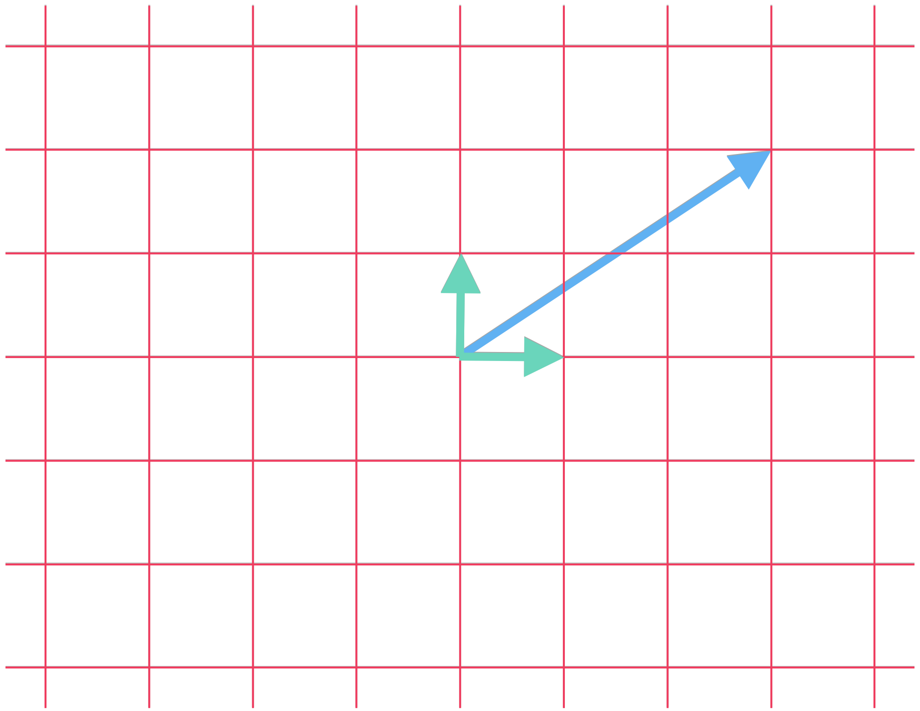
Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

- *If we take an orthonormal matrix and multiply one or more columns by -1, it amounts to flipping that/those axes*
- *If we take an orthonormal matrix and shift its columns around, it amounts to exchanging/reordering those axes*
- *In general, orthonormal matrices give us a new (rotated/flipped/exchanged) basis*



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

Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

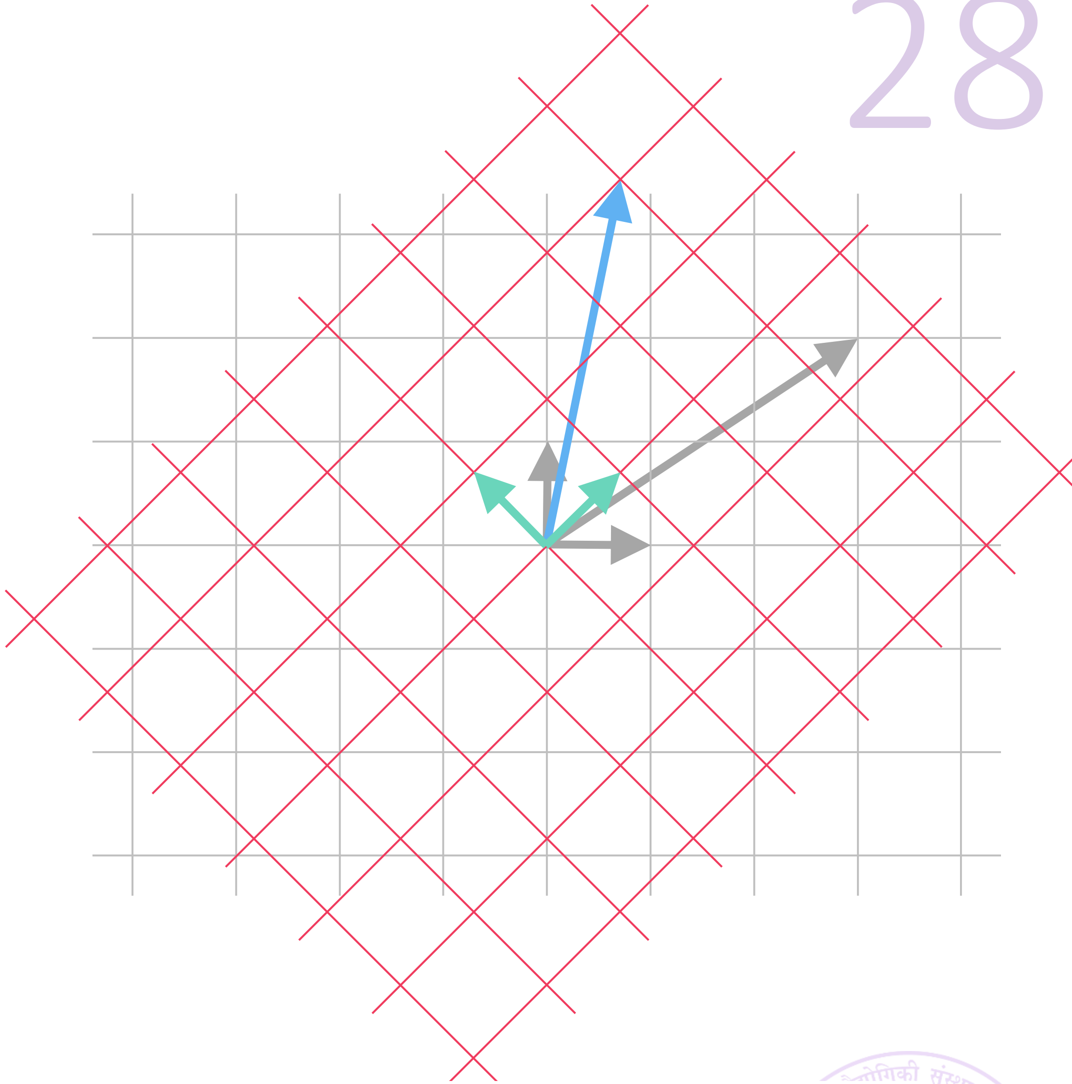
Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

If we take an orthonormal matrix and multiply one or more columns by -1*, it amounts to flipping that/those axes*

If we take an orthonormal matrix and shift its columns around, it amounts to exchanging/reordering those axes

In general, orthonormal matrices give us a new (rotated/flipped/exchanged) basis



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

Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

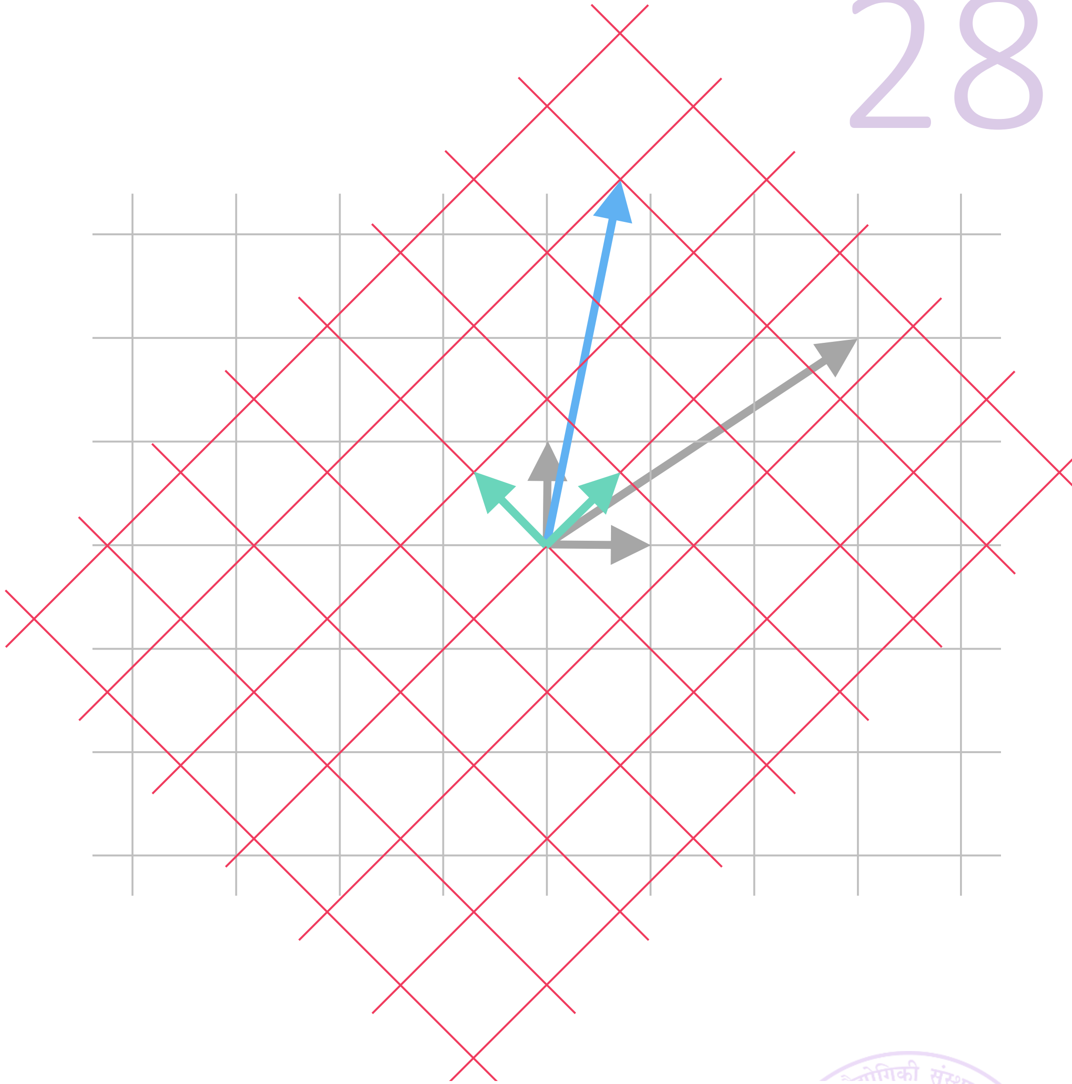
Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

If we take an orthonormal matrix and multiply one or more columns by -1*, it amounts to flipping that/those axes*

If we take an orthonormal matrix and shift its columns around, it amounts to exchanging/reordering those axes

In general, orthonormal matrices give us a new (rotated/flipped/exchanged) basis



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

Orthonormal Matrices

Matrices whose columns are unit L2 norm and perp. to each other

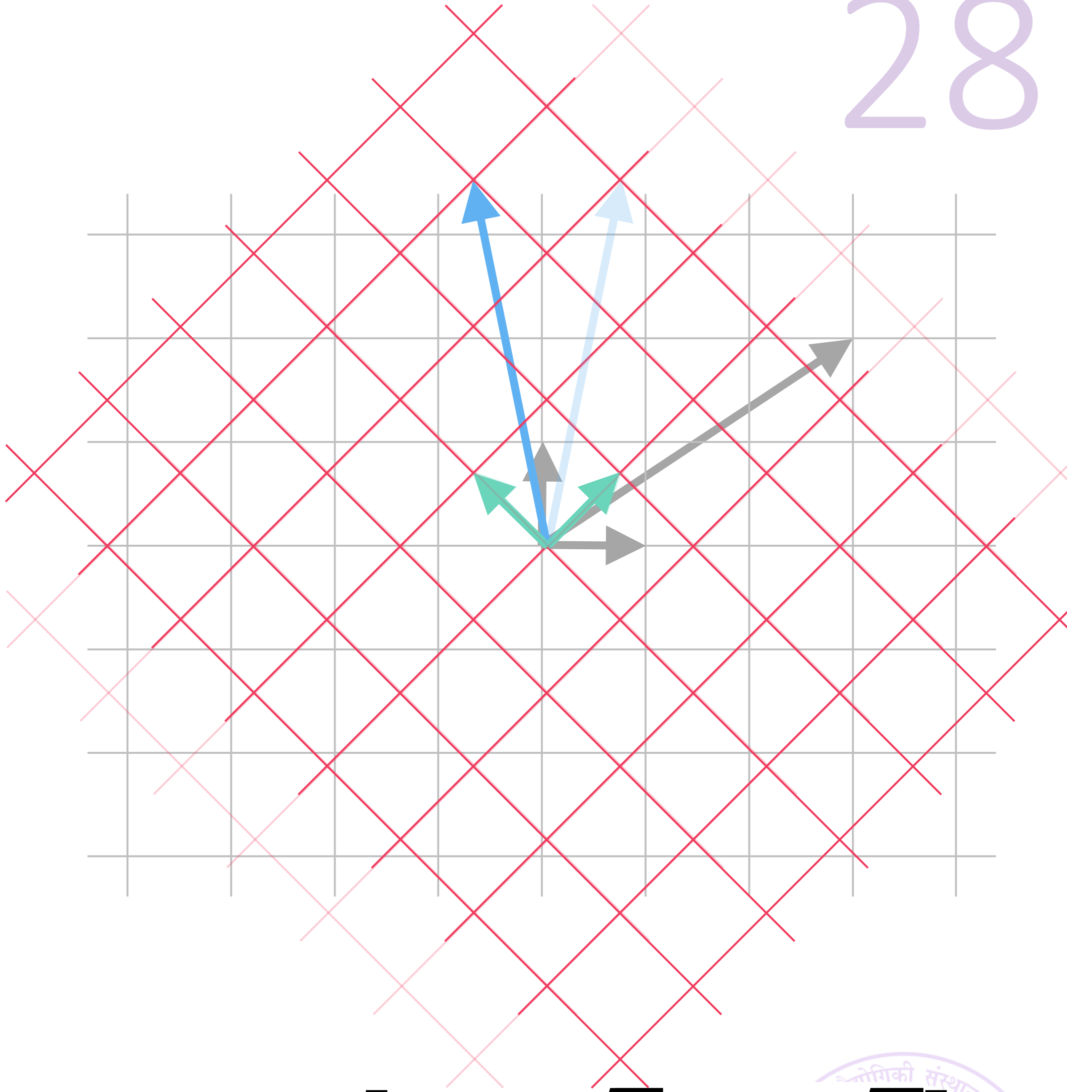
Last saw them w.r.t rotation maps

Orthonormal matrices can rotate for sure but can do much more

If we take an orthonormal matrix and multiply one or more columns by -1, it amounts to flipping that/those axes

If we take an orthonormal matrix and shift its columns around, it amounts to exchanging/reordering those axes

In general, orthonormal matrices give us a new (rotated/flipped/exchanged) basis



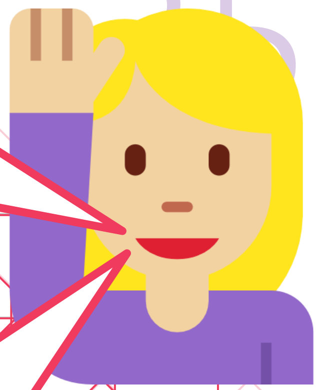
e.g. $A = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Orthonorm

Matrices whose

It is easy to see that if A is orthonormal i.e. its columns are unit L2 norm as well as pairwise orthogonal, then $A^\top A = I$ i.e. A has an inverse and that inverse is simply $A^\top$

It is also true that if $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_d]$ is orthonormal, then $AA^\top = I$ as well. To see this, notice that for any $\mathbf{a}_i$, we have $AA^\top \mathbf{a}_i = A\mathbf{e}_i = \mathbf{a}_i$. This means $AA^\top$ acts as an identity map to all the d orthonormal columns of A. However, any vector $\mathbf{v} \in \mathbb{R}^d$ can be written as a linear combination of the columns of A (since they are pairwise independent and there are d of them so their span is $\mathbb{R}^d$). Thus, $AA^\top$ must act as an identity map to all vectors $\mathbf{v} \in \mathbb{R}^d$ and so $AA^\top = I$



multiply one or more columns by -1, it amounts to flipping that/those axes

If we take an orthonormal matrix and shift its columns around, it amounts to exchanging/reordering those axes

In general, orthonormal matrices give us a new (rotated/flipped/exchanged) basis

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

Singular Value Decomposition

Let $A \in \mathbb{R}^{d\times d}$ be any (possibly non-symmetric) square matrix. Then we can **always** write A as a product of three matrices

$$A = U\Sigma V^\top$$

where $U, V \in \mathbb{R}^{d\times d}$ are orthonormal matrices and Σ is a scaling (diagonal) matrix with all entries non-negative

Thus, every linear map is simply a rotation (+ possibly axes flips, swaps), followed by axes scaling followed by another rotation (+ axes flips/swaps)

Singular Value Decomposition

Let $A \in \mathbb{R}^{d \times d}$ be any (possibly non-symmetric) square matrix. Then we can **always** write A as a product of three matrices

$$A = U\Sigma V^\top$$

where $U, V \in \mathbb{R}^{d \times d}$ are orthonormal matrices and Σ is a scaling (diagonal) matrix with all entries non-negative

Thus, every linear map is simply a rotation (+ possibly axes flips, swaps), followed by axes scaling followed by another rotation (+ axes flips/swaps)

$$A = \frac{3}{8\sqrt{2}} \cdot \begin{bmatrix} 2\sqrt{3} - 1 & 2\sqrt{3} + 1 \\ -2 - \sqrt{3} & -2 + \sqrt{3} \end{bmatrix}$$

Singular Value Decomposition

Let $A \in \mathbb{R}^{d \times d}$ be any (possibly non-symmetric) square matrix. Then we can **always** write A as a product of three matrices

$$A = U\Sigma V^\top$$

where $U, V \in \mathbb{R}^{d \times d}$ are orthonormal matrices and Σ is a scaling (diagonal) matrix with all entries non-negative

Thus, every linear map is simply a rotation (+ possibly axes flips, swaps), followed by axes scaling followed by another rotation (+ axes flips/swaps)

$$A = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Singular Value Decomposition

Let $A \in \mathbb{R}^{d \times d}$ be any (possibly non-symmetric) square matrix. Then we can **always** write A as a product of three matrices

$$A = U\Sigma V^\top$$

where $U, V \in \mathbb{R}^{d \times d}$ are orthonormal matrices and Σ is a scaling (diagonal) matrix with all entries non-negative

Thus, every linear map is simply a rotation (+ possibly axes flips, swaps), followed by axes scaling followed by another rotation (+ axes flips/swaps)

clockwise rotation 120   scaling   counter clockwise rotation 45

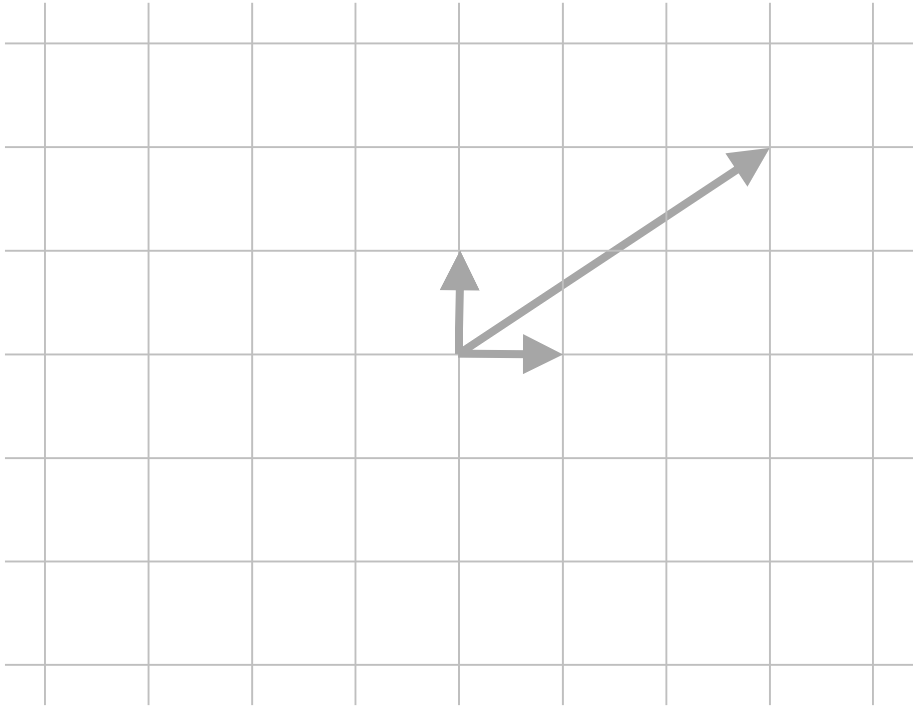
$$A = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Singular Value Decomposition

Let $A \in \mathbb{R}^{d \times d}$ be any (possibly non-symmetric) square matrix. Then we can **always** write A as a product of three matrices

$$A = U\Sigma V^\top$$

where $U, V \in \mathbb{R}^{d \times d}$ are orthonormal matrices and Σ is a scaling (diagonal) matrix with all entries non-negative



Thus, every linear map is simply a rotation (+ possibly axes flips, swaps), followed by axes scaling followed by another rotation (+ axes flips/swaps)

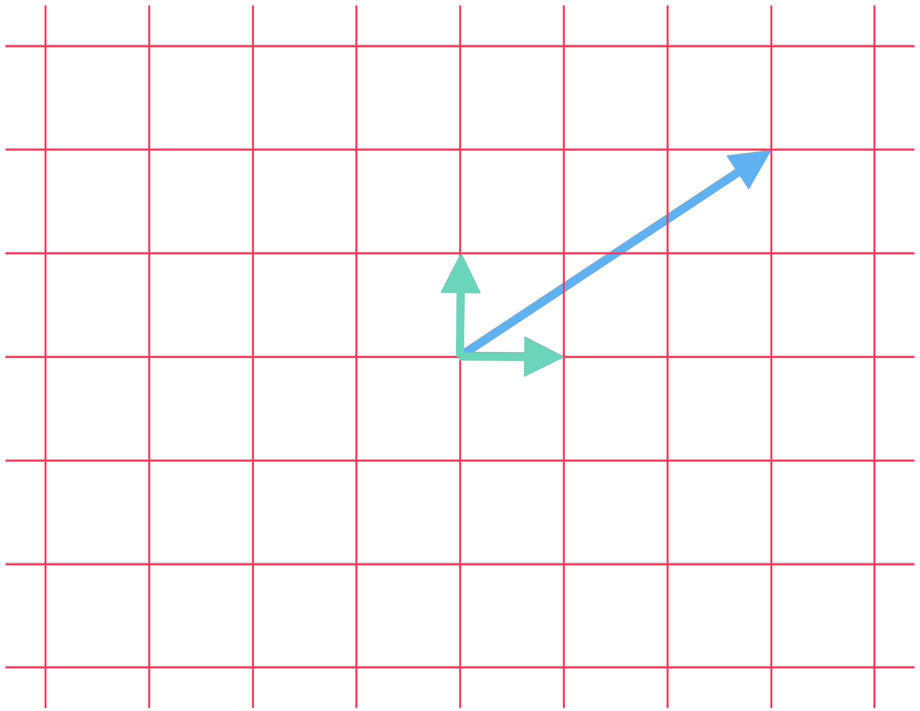
$$A = \underbrace{\begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}}_{\text{clockwise rotation 120}} \cdot \underbrace{\begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}}_{\text{scaling}} \cdot \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{\text{counter clockwise rotation 45}}$$

Singular Value Decomposition

Let $A \in \mathbb{R}^{d \times d}$ be any (possibly non-symmetric) square matrix. Then we can **always** write A as a product of three matrices

$$A = U\Sigma V^\top$$

where $U, V \in \mathbb{R}^{d \times d}$ are orthonormal matrices and Σ is a scaling (diagonal) matrix with all entries non-negative



Thus, every linear map is simply a rotation (+ possibly axes flips, swaps), followed by axes scaling followed by another rotation (+ axes flips/swaps)

clockwise rotation 120 | scaling | counter clockwise rotation 45

$$A = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

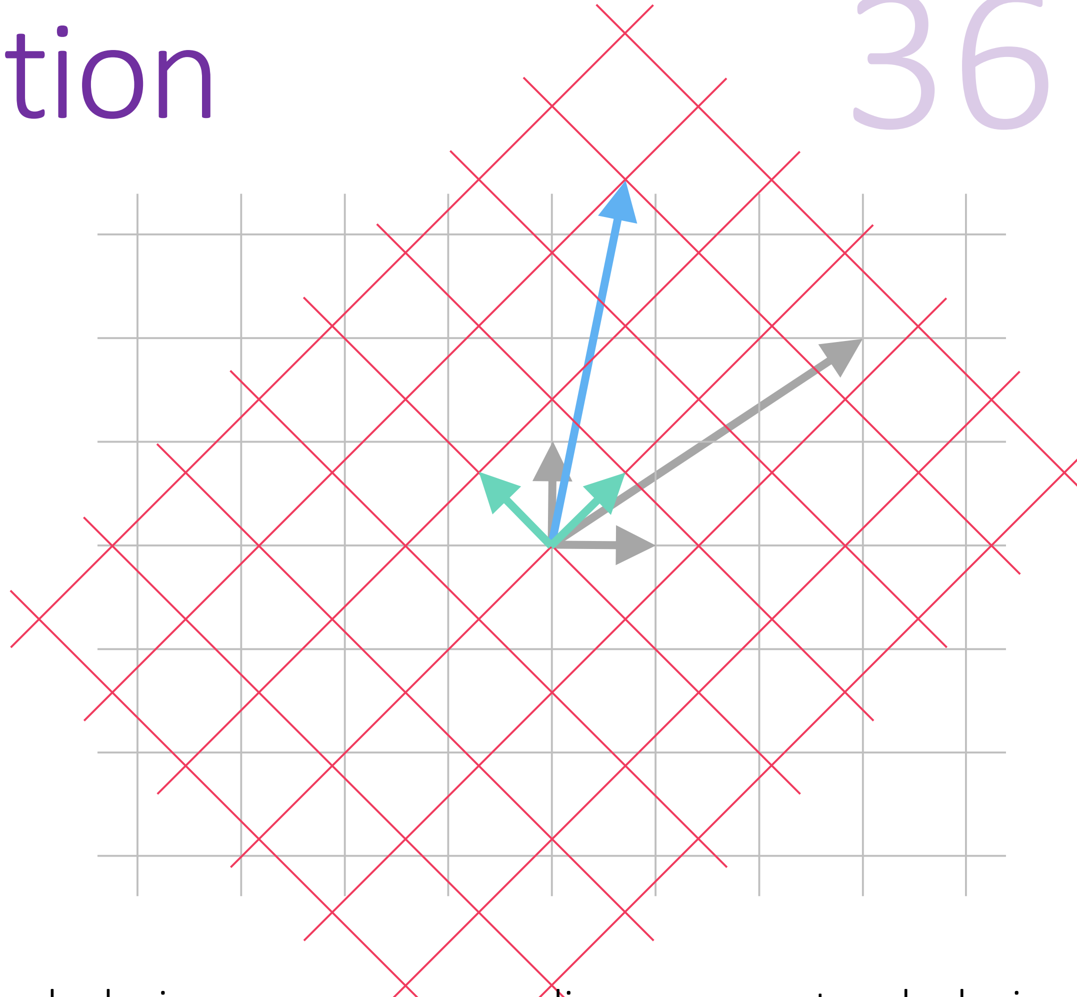
Singular Value Decomposition

Let $A \in \mathbb{R}^{d\times d}$ be any (possibly non-symmetric) square matrix. Then we can **always** write A as a product of three matrices

$$A = U\Sigma V^\top$$

where $U, V \in \mathbb{R}^{d\times d}$ are orthonormal matrices and Σ is a scaling (diagonal) matrix with all entries non-negative

Thus, every linear map is simply a rotation (+ possibly axes flips, swaps), followed by axes scaling followed by another rotation (+ axes flips/swaps)



clockwise rotation 120 | scaling | counter clockwise rotation 45

$$A = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

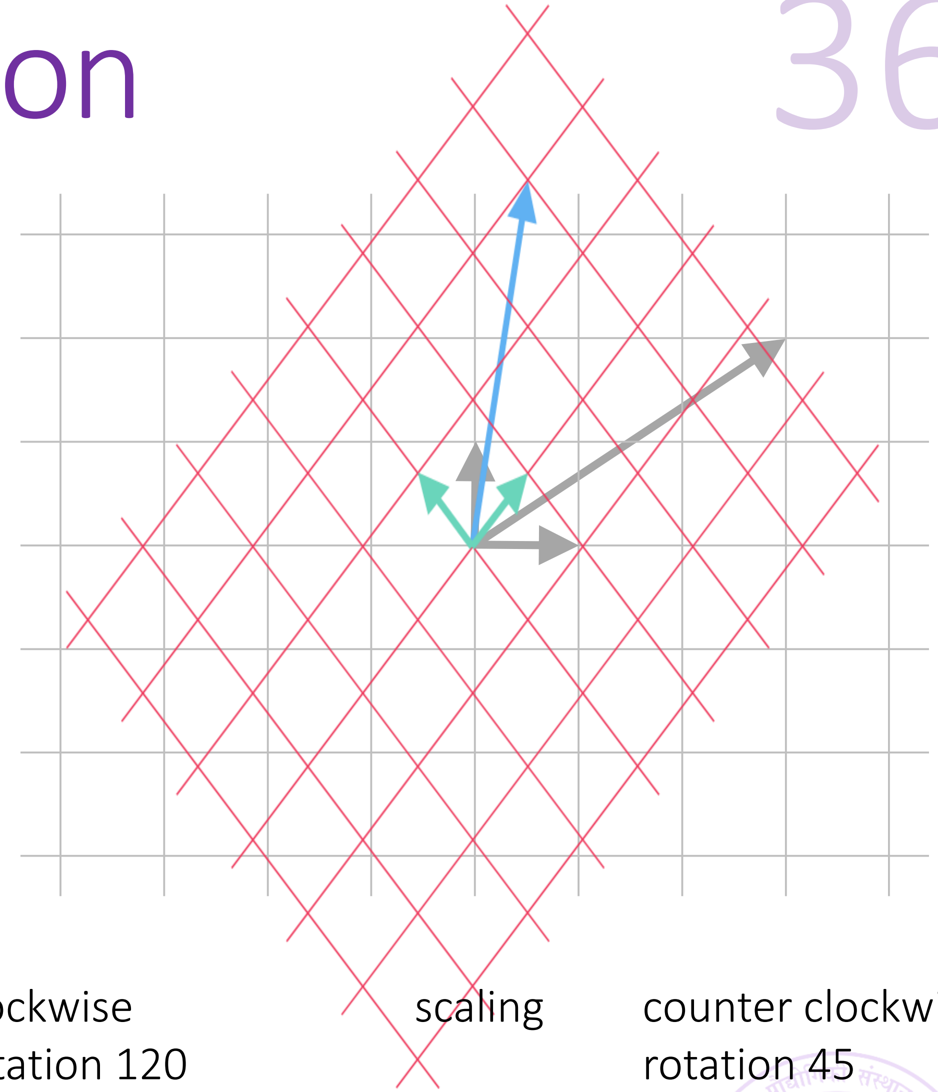
Singular Value Decomposition

Let $A \in \mathbb{R}^{d \times d}$ be any (possibly non-symmetric) square matrix. Then we can **always** write A as a product of three matrices

$$A = U\Sigma V^\top$$

where $U, V \in \mathbb{R}^{d \times d}$ are orthonormal matrices and Σ is a scaling (diagonal) matrix with all entries non-negative

Thus, every linear map is simply a rotation (+ possibly axes flips, swaps), followed by axes scaling followed by another rotation (+ axes flips/swaps)



$$A = \underbrace{\begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}}_{\text{clockwise rotation 120}} \cdot \underbrace{\begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}}_{\text{scaling}} \cdot \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{\text{counter clockwise rotation 45}}$$

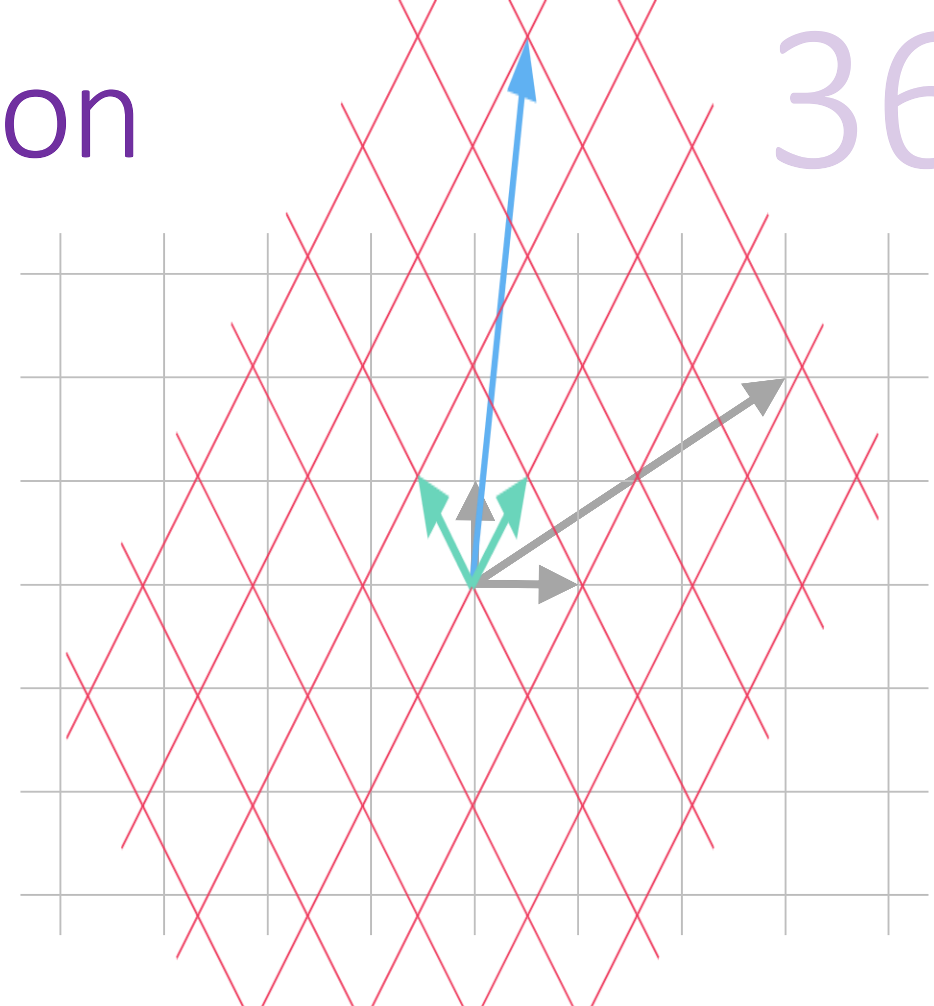
Singular Value Decomposition

Let $A \in \mathbb{R}^{d\times d}$ be any (possibly non-symmetric) square matrix. Then we can **always** write A as a product of three matrices

$$A = U\Sigma V^\top$$

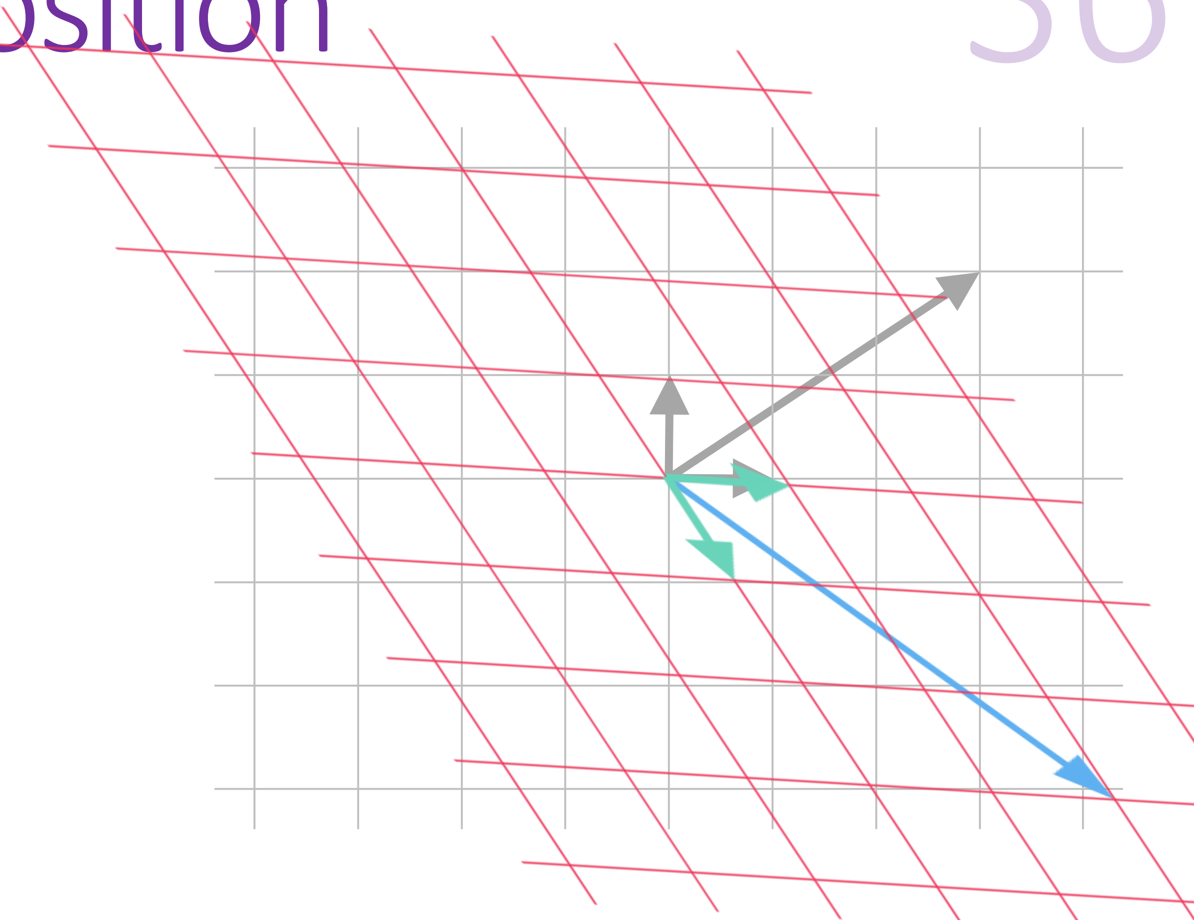
where $U, V \in \mathbb{R}^{d\times d}$ are orthonormal matrices and Σ is a scaling (diagonal) matrix with all entries non-negative

Thus, every linear map is simply a rotation (+ possibly axes flips, swaps), followed by axes scaling followed by another rotation (+ axes flips/swaps)



$$A = \underbrace{\begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}}_{\text{clockwise rotation 120}} \cdot \underbrace{\begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}}_{\text{scaling}} \cdot \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{\text{counter clockwise rotation 45}}$$

Singular Value Decomposition



Let $A \in \mathbb{R}^{d\times d}$ be any (possibly non-symmetric) square matrix. Then we can **always** write A as a product of three matrices

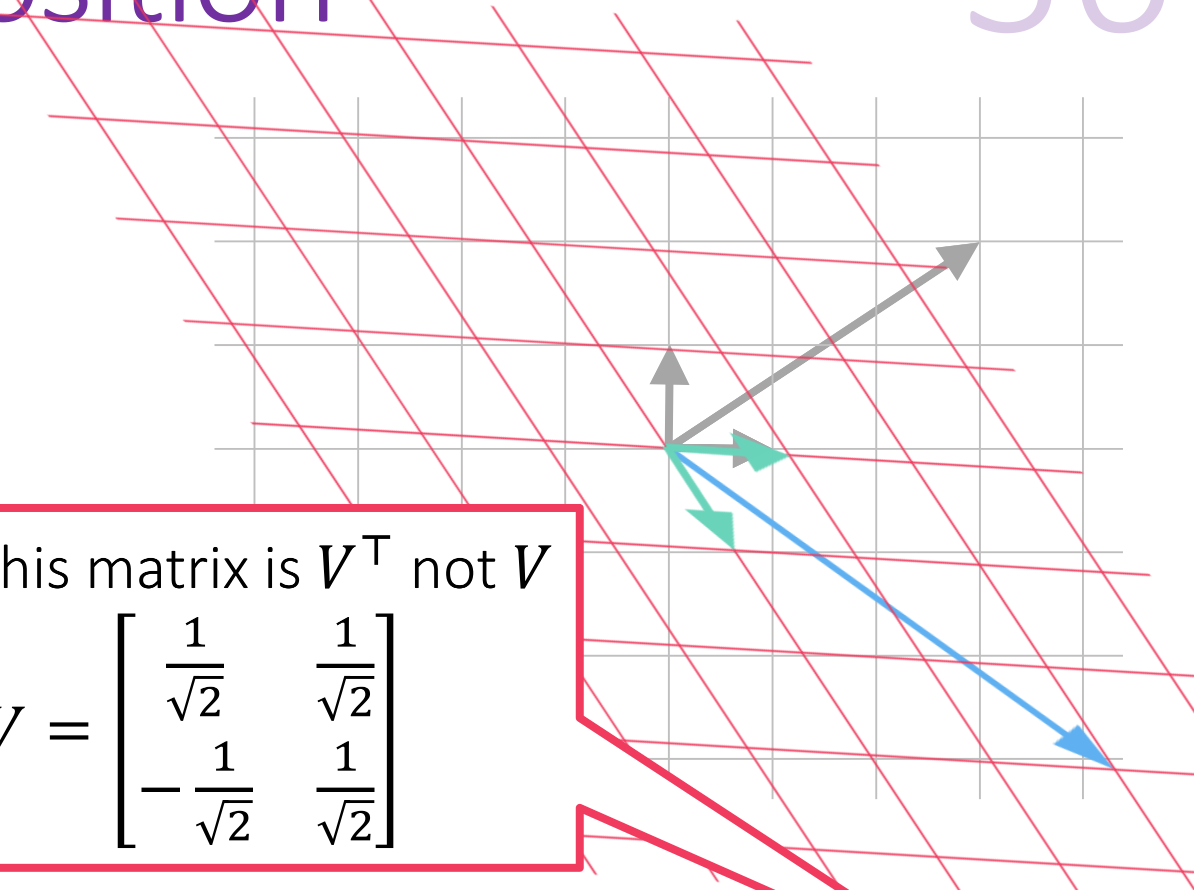
$$A = U\Sigma V^\top$$

where $U, V \in \mathbb{R}^{d\times d}$ are orthonormal matrices and Σ is a scaling (diagonal) matrix with all entries non-negative

Thus, every linear map is simply a rotation (+ possibly axes flips, swaps), followed by axes scaling followed by another rotation (+ axes flips/swaps)

$$A = \underbrace{\begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}}_{\text{clockwise rotation 120}} \cdot \underbrace{\begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}}_{\text{scaling}} \cdot \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{\text{counter clockwise rotation 45}}$$

Singular Value Decomposition



Let $A \in \mathbb{R}^{d\times d}$ be any (possibly non-symmetric) square matrix. Then we can **always** write A as a product of three matrices

$$A = U\Sigma V^\top$$

where $U, V \in \mathbb{R}^{d\times d}$ are orth
matrices and Σ is a scaling (d
matrix with all entries non-ne

Caution: This matrix is $V^\top$ not V

$$\text{i.e. } V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Thus, every linear map is simply a rotation (+ possibly axes flips, swaps), followed by axes scaling followed by another rotation (+ axes flips/swaps)

clockwise rotation 120

scaling

unter clockwise rotati n 45

$$A = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Singular Value Decomposition

$$A = U\Sigma V^\top$$

Diagonal elements in Σ are called the *singular values* of A (always $\geq \mathbf{0}$)

Column vectors of U are called the *left singular vectors* of A

Column vectors of V (row vecs of $V^\top$) called *right singular vectors* of A

If $U = [\mathbf{u}^1, \ldots, \mathbf{u}^d], \mathrm{V} = [\mathbf{v}^1, \ldots, \mathbf{v}^d], \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_d)$, then

$$A = \sum_{j=1}^{d} \sigma_j \, \mathbf{u}^j \left(\mathbf{v}^j\right)^\top$$

Singular values of a matrix are always unique, singular vectors are not

Given one set of left+right singular vectors, can obtain other sets of left+right singular vectors using "certain" orthonormal maps – details a bit tedious

Tons of things to study about singular decompositions – too little time

Singular Value Decomposition

$$A = U\Sigma V^\top$$

Diagonal elements in Σ are called the *singular values* of A (always $\geq \mathbf{0}$)

Column vectors of U are called the *left singular vectors* of A

Note that each matrix $\mathbf{u}^j(\mathbf{v}^j)^\top$ has unit (row and column) rank

Co … ecs of $V^\top$) called *right singular vectors* of A

If U … $\mathbf{v}^d], \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_d)$, then

$$A = \sum_{j=1}^{d} \sigma_j \, \mathbf{u}^j (\mathbf{v}^j)^\top$$

Singular values of a matrix are always unique, singular vectors are not

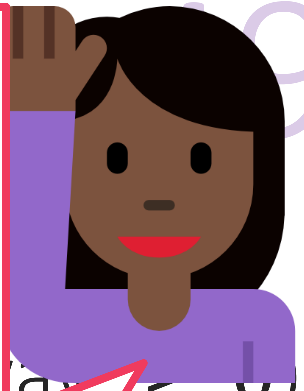
Given one set of left+right singular vectors, can obtain other sets of left+right singular vectors using "certain" orthonormal maps – details a bit tedious

Tons of things to study about singular decompositions – too little time

For example, consider the following four different but equivalent SVDs for A

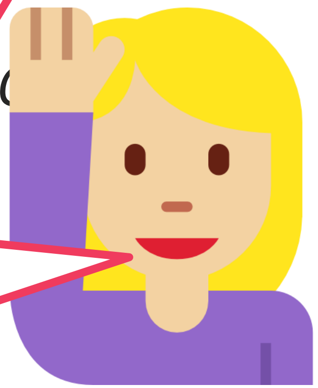
$$A = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{3}}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{3}{4} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{3}{4} \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$



If $U = [\mathbf{u}^1, \ldots, \mathbf{u}$ ... then

In order to minimize this ambiguity, people commonly write SVD such that $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_d)$ and $\sigma_1 \geq \sigma_2 \geq \cdots$



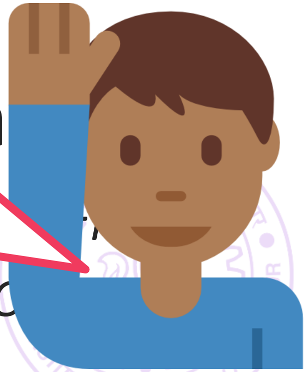
Singular values of a ma ... a

Given one set of left+rig ...
singular vectors using " ... dic

If $A = [\mathbf{a}^1, \ldots, \mathbf{a}^n] = U\Sigma V^\top \in \mathbb{R}^{d\times d}$ then we have

$$\mathbf{a}^j = \sum_{i=1}^d \sigma_i \mathbf{v}_j^i \cdot \mathbf{u}^i$$

Can you find an expression for rows of A too?



Tons of things to study ... ttle time

Singular Value Decomposition

SVD is defined even for matrices that are not square

Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 1: $n > m$ i.e. output dim < input dim i.e. A reduces dim of vectors

In this case Σ throws out last $n - m$ dimensions after rotation by $V^\top$

Singular Value Decomposition

SVD is defined even for matrices that are not square

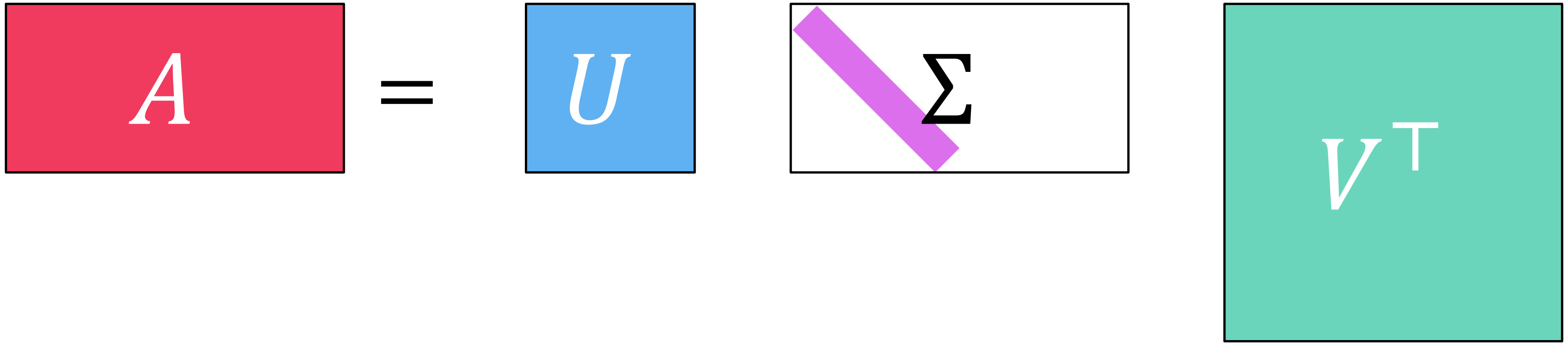
Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 1: $n > m$ i.e. output dim < input dim i.e. A reduces dim of vectors

In this case Σ throws out last $n - m$ dimensions after rotation by $V^\top$



Singular Value Decomposition

SVD is defined even for matrices that are not square

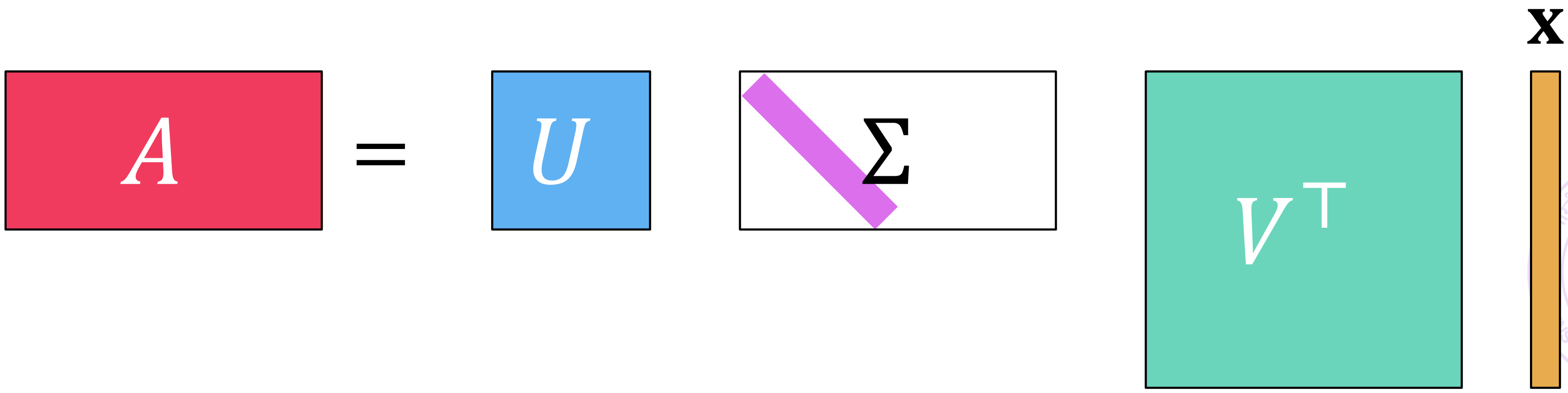
Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 1: $n > m$ i.e. output dim < input dim i.e. A reduces dim of vectors

In this case Σ throws out last $n - m$ dimensions after rotation by $V^\top$



Singular Value Decomposition

SVD is defined even for matrices that are not square

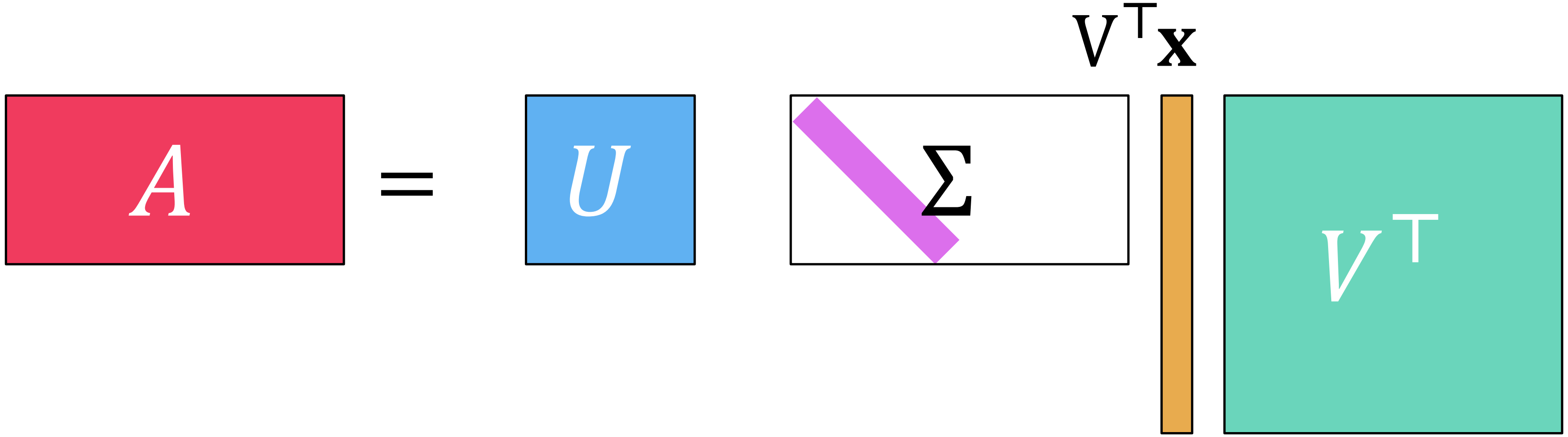
Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 1: $n > m$ i.e. output dim < input dim i.e. A reduces dim of vectors

In this case Σ throws out last $n - m$ dimensions after rotation by $V^\top$



Singular Value Decomposition

SVD is defined even for matrices that are not square

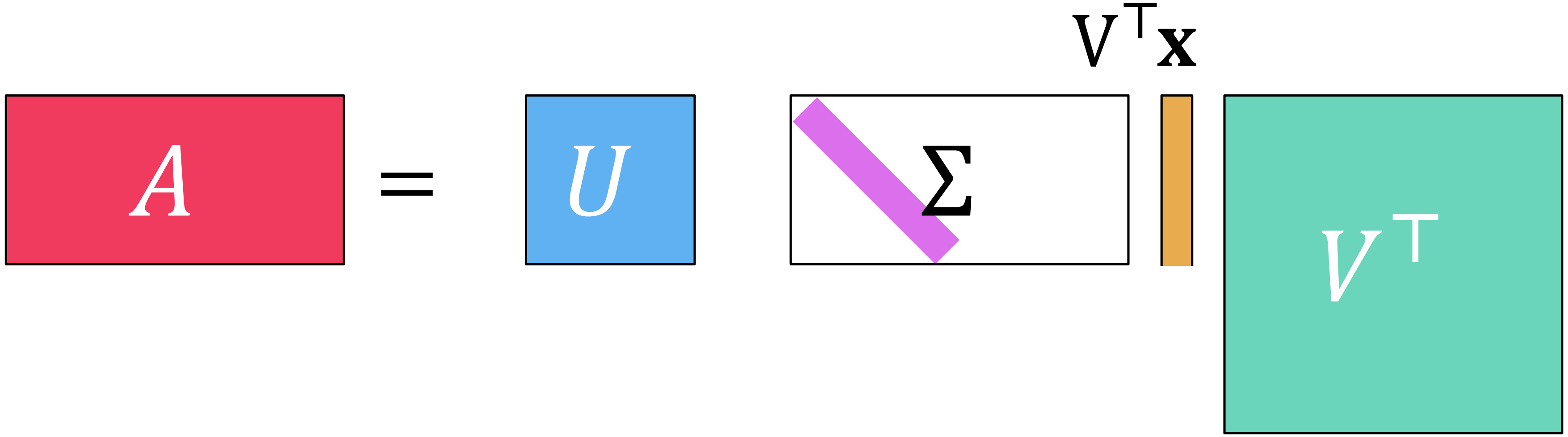
Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 1: $n > m$ i.e. output dim < input dim i.e. A reduces dim of vectors

In this case Σ throws out last $n - m$ dimensions after rotation by $V^\top$



Singular Value Decomposition

SVD is defined even for matrices that are not square

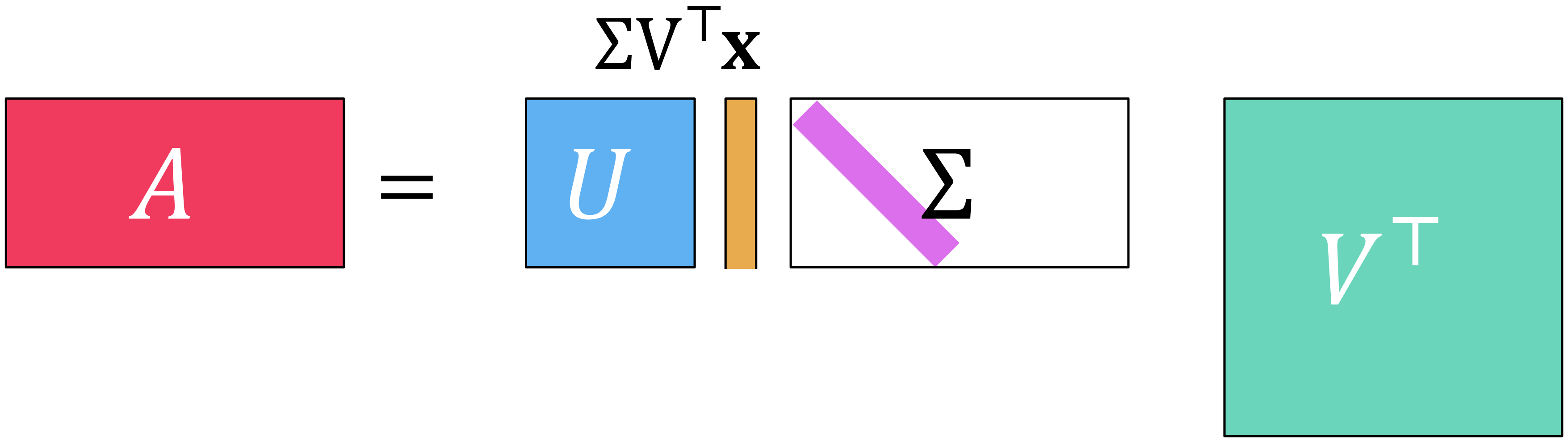
Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 1: $n > m$ i.e. output dim < input dim i.e. A reduces dim of vectors

In this case Σ throws out last $n - m$ dimensions after rotation by $V^\top$



Singular Value Decomposition

SVD is defined even for matrices that are not square

Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

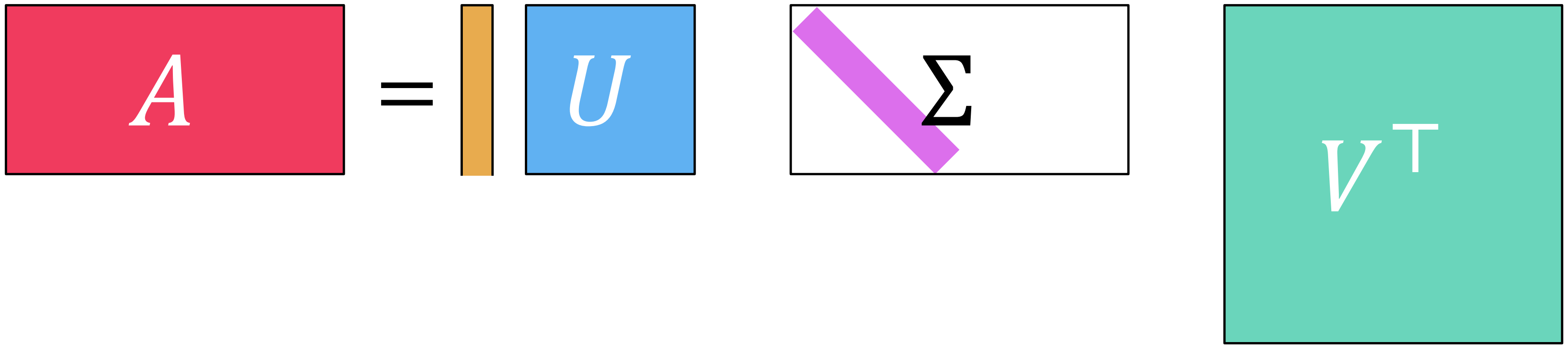
$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 1: $n > m$ i.e. output dim < input dim i.e. A reduces dim of vectors

In this case Σ throws out last $n - m$ dimensions after rotation by $V^\top$

$A\mathbf{x}$



Singular Value Decomposition

SVD is defined even for matrices that are not square

Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 2: $n < m$ i.e. output dim > input dim. A increases dim of vectors

In this case Σ adds $n - m$ dummy dimensions in addition to scaling

Singular Value Decomposition

SVD is defined even for matrices that are not square

Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

- $U \in \mathbb{R}^{m \times m}$ *and* $V \in \mathbb{R}^{n \times n}$ *are orthonormal matrices of different sizes*
- $\Sigma \in \mathbb{R}^{m \times n}$ *is a rectangular diagonal matrix with* $\Sigma_{ii} \geq 0$ *but* $\Sigma_{ij} = 0$ *for* $i \neq j$

Case 2: $n < m$ i.e. output dim > input dim. A increases dim of vectors

In this case Σ adds $n - m$ dummy dimensions in addition to scaling

Singular Value Decomposition

SVD is defined even for matrices that are not square

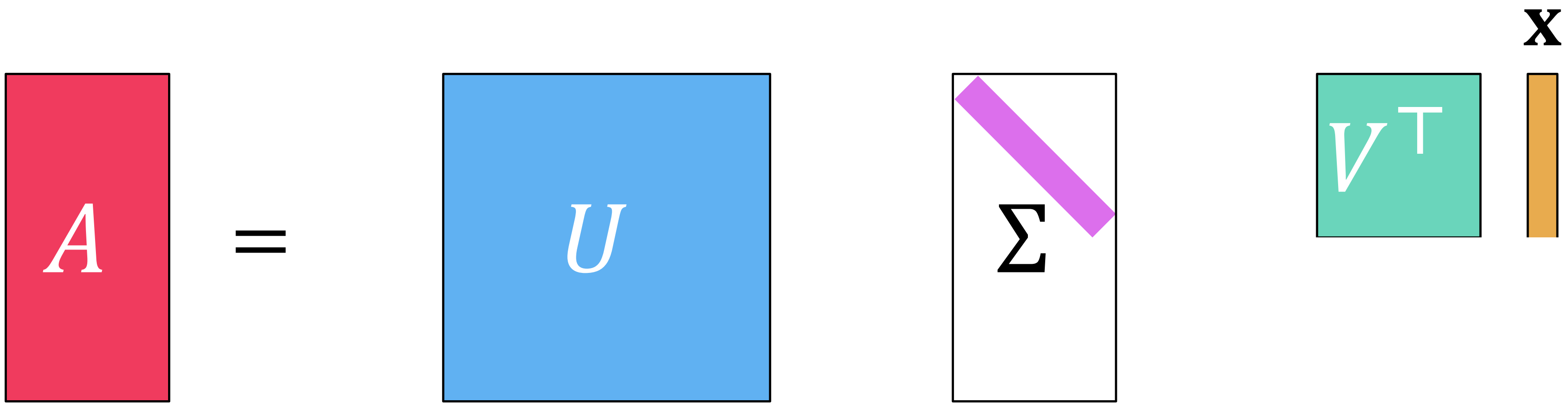
Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 2: $n < m$ i.e. output dim > input dim. A increases dim of vectors

In this case Σ adds $n - m$ dummy dimensions in addition to scaling



Singular Value Decomposition

SVD is defined even for matrices that are not square

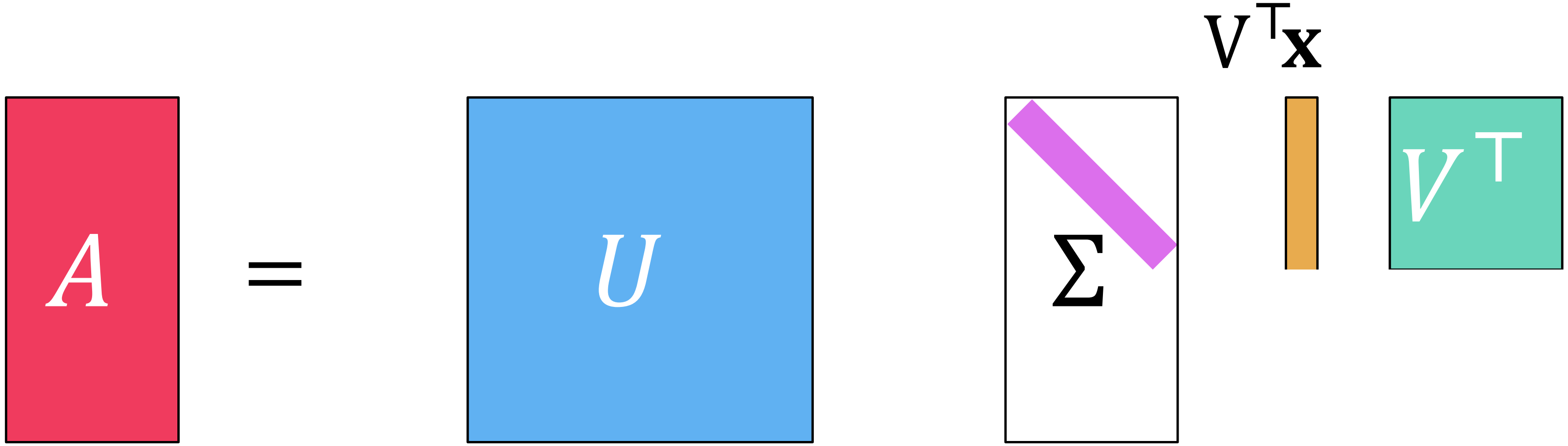
Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 2: $n < m$ i.e. output dim > input dim. A increases dim of vectors

In this case Σ adds $n - m$ dummy dimensions in addition to scaling



Singular Value Decomposition

SVD is defined even for matrices that are not square

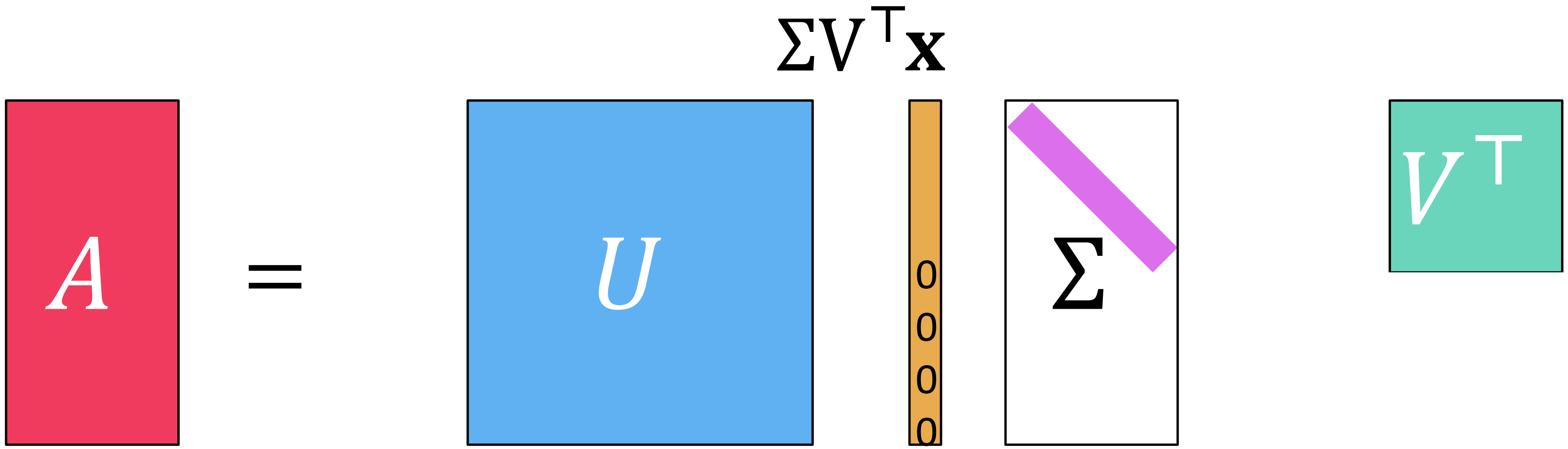
Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 2: $n < m$ i.e. output dim > input dim. A increases dim of vectors

In this case Σ adds $n - m$ dummy dimensions in addition to scaling



Singular Value Decomposition

SVD is defined even for matrices that are not square

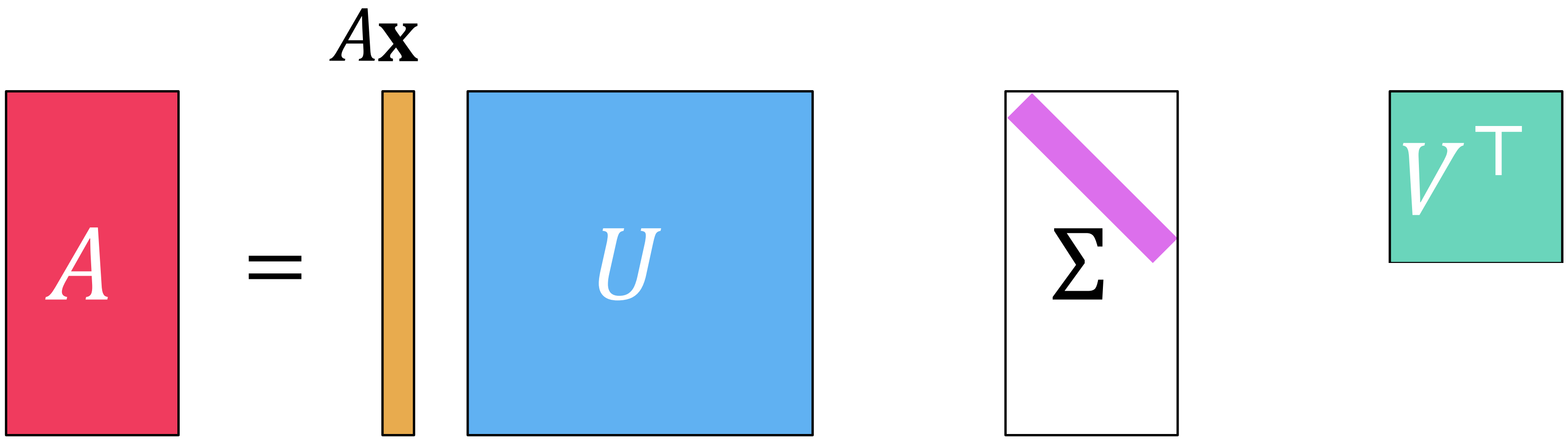
Suppose $A \in \mathbb{R}^{m \times n}$ then we can always write $A = U\Sigma V^\top$ where

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthonormal matrices of different sizes

$\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with $\Sigma_{ii} \geq 0$ but $\Sigma_{ij} = 0$ for $i \neq j$

Case 2: $n < m$ i.e. output dim > input dim. A increases dim of vectors

In this case Σ adds $n - m$ dummy dimensions in addition to scaling



Rank

Getting the SVD of a matrix $A = U\Sigma V^\top$ immediately tells us a lot

Rank: We always have $\mathbf{rank}(A) =$ number of non-zero entries in Σ

To see why, notice that if some diagonal entries of Σ are zero, we can remove those rows and the corresponding columns of U

If $A \in \mathbb{R}^{m\times n}$ and only r entries of Σ are non-zero, then we can equivalently write A as

$$A = \tilde{U}\tilde{\Sigma}V^\top$$

where $\tilde{U} \in \mathbb{R}^{m\times r}$ and $\tilde{\Sigma} \in \mathbb{R}^{r\times n}$. This is often called the thin SVD *of A*

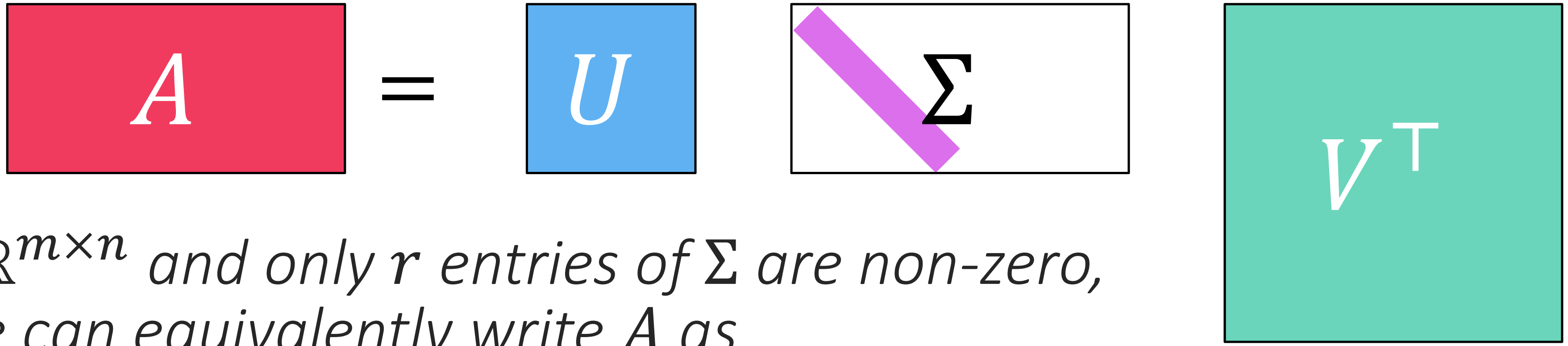
Finally, notice that $A\mathbf{v} = \tilde{U}\mathbf{z}$ where $\mathbf{z} = \tilde{\Sigma}V^\top\mathbf{v} \in \mathbb{R}^r$ and thus, the column space of A is nothing but the span of columns of $\tilde{U}$ and clearly $\mathbf{dim}(\tilde{U}) = r$

Rank

Getting the SVD of a matrix $A = U\Sigma V^\top$ immediately tells us a lot

Rank: We always have $\mathbf{rank}(A) =$ number of non-zero entries in Σ

To see why, notice that if some diagonal entries of Σ are zero, we can remove those rows and the corresponding columns of U



If $A \in \mathbb{R}^{m\times n}$ and only r entries of Σ are non-zero, then we can equivalently write A as

$$A = \tilde{U}\tilde{\Sigma}V^\top$$

where $\tilde{U} \in \mathbb{R}^{m\times r}$ and $\tilde{\Sigma} \in \mathbb{R}^{r\times n}$. This is often called the thin SVD *of A*

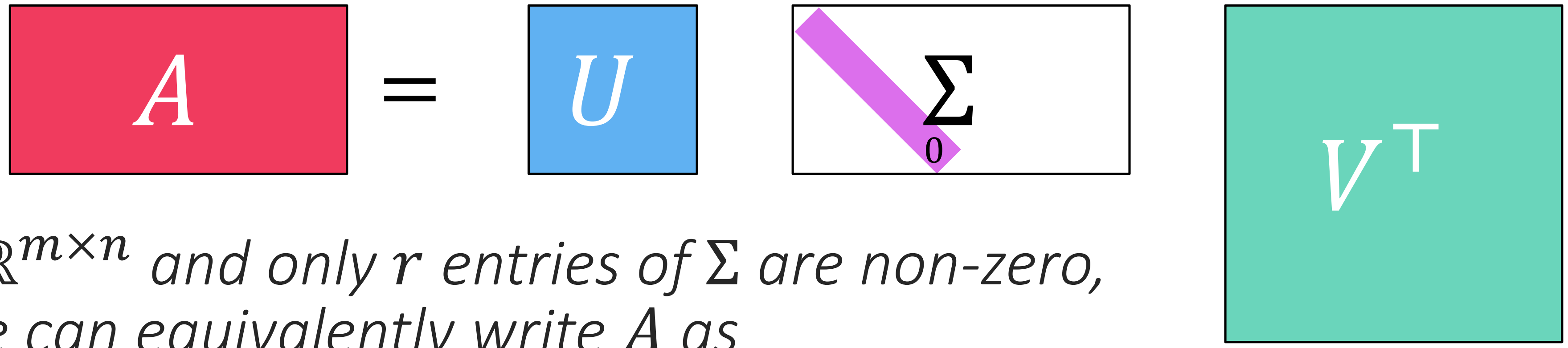
Finally, notice that $A\mathbf{v} = \tilde{U}\mathbf{z}$ where $\mathbf{z} = \tilde{\Sigma}V^\top\mathbf{v} \in \mathbb{R}^r$ and thus, the column space of A is nothing but the span of columns of $\tilde{U}$ and clearly $\dim(\tilde{U}) = r$

Rank

Getting the SVD of a matrix $A = U\Sigma V^\top$ immediately tells us a lot

Rank: We always have $\mathbf{rank}(A) =$ number of non-zero entries in Σ

To see why, notice that if some diagonal entries of Σ are zero, we can remove those rows and the corresponding columns of U



If $A \in \mathbb{R}^{m \times n}$ and only r entries of Σ are non-zero, then we can equivalently write A as

$$A = \tilde{U}\tilde{\Sigma}V^\top$$

where $\tilde{U} \in \mathbb{R}^{m \times r}$ and $\tilde{\Sigma} \in \mathbb{R}^{r \times n}$. This is often called the thin SVD *of A*

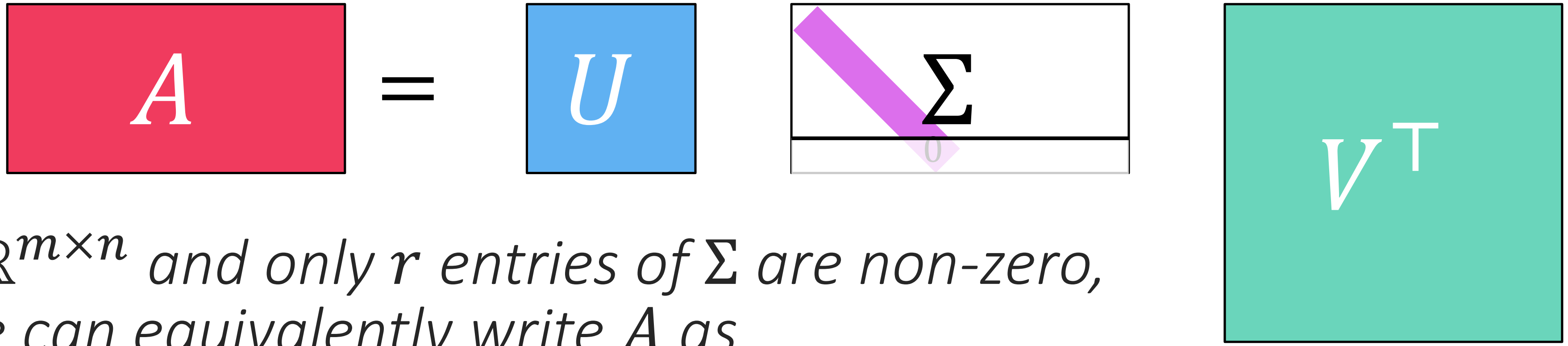
Finally, notice that $A\mathbf{v} = \tilde{U}\mathbf{z}$ where $\mathbf{z} = \tilde{\Sigma}V^\top\mathbf{v} \in \mathbb{R}^r$ and thus, the column space of A is nothing but the span of columns of $\tilde{U}$ and clearly $\dim(\tilde{U}) = r$

Rank

Getting the SVD of a matrix $A = U\Sigma V^\top$ immediately tells us a lot

Rank: We always have $\mathbf{rank}(A) =$ number of non-zero entries in Σ

To see why, notice that if some diagonal entries of Σ are zero, we can remove those rows and the corresponding columns of U



If $A \in \mathbb{R}^{m\times n}$ and only r entries of Σ are non-zero, then we can equivalently write A as

$$A = \tilde{U}\tilde{\Sigma}V^\top$$

where $\tilde{U} \in \mathbb{R}^{m\times r}$ and $\tilde{\Sigma} \in \mathbb{R}^{r\times n}$. This is often called the thin SVD *of A*

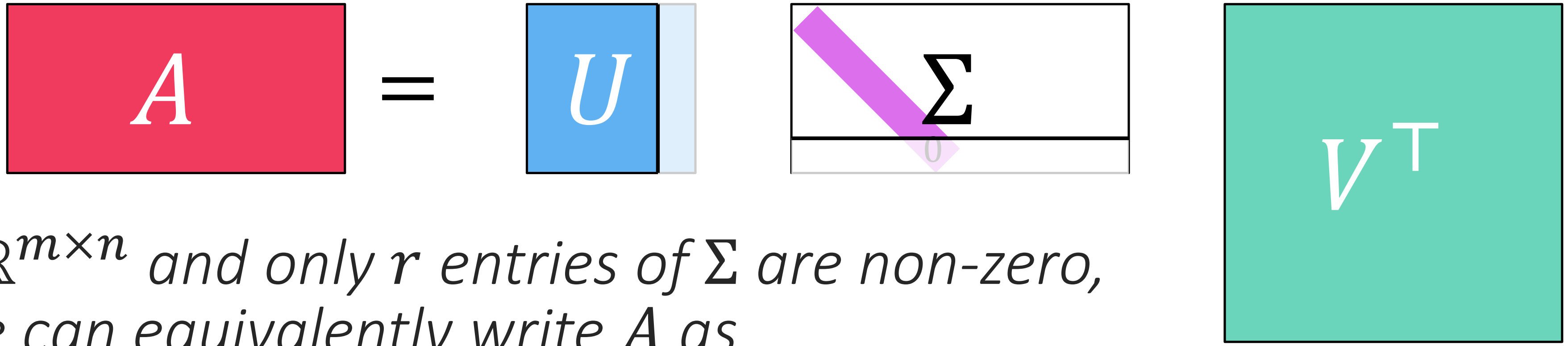
Finally, notice that $A\mathbf{v} = \tilde{U}\mathbf{z}$ where $\mathbf{z} = \tilde{\Sigma}V^\top\mathbf{v} \in \mathbb{R}^r$ and thus, the column space of A is nothing but the span of columns of $\tilde{U}$ and clearly $\dim(\tilde{U}) = r$

Rank

Getting the SVD of a matrix $A = U\Sigma V^\top$ immediately tells us a lot

Rank: We always have $\mathbf{rank}(A)$ = number of non-zero entries in Σ

To see why, notice that if some diagonal entries of Σ are zero, we can remove those rows and the corresponding columns of U



If $A \in \mathbb{R}^{m\times n}$ and only r entries of Σ are non-zero, then we can equivalently write A as

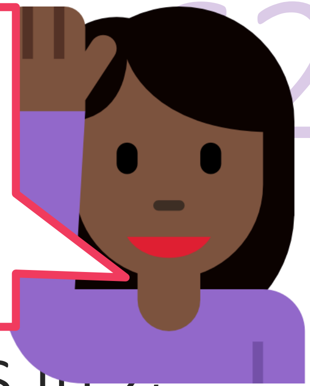
$$A = \tilde{U}\tilde{\Sigma}V^\top$$

where $\tilde{U} \in \mathbb{R}^{m\times r}$ and $\tilde{\Sigma} \in \mathbb{R}^{r\times n}$. This is often called the thin SVD *of A*

Finally, notice that $A\mathbf{v} = \tilde{U}\mathbf{z}$ where $\mathbf{z} = \tilde{\Sigma}V^\top\mathbf{v} \in \mathbb{R}^r$ and thus, the column space of A is nothing but the span of columns of $\tilde{U}$ and clearly $\dim(\tilde{U}) = r$

Rank

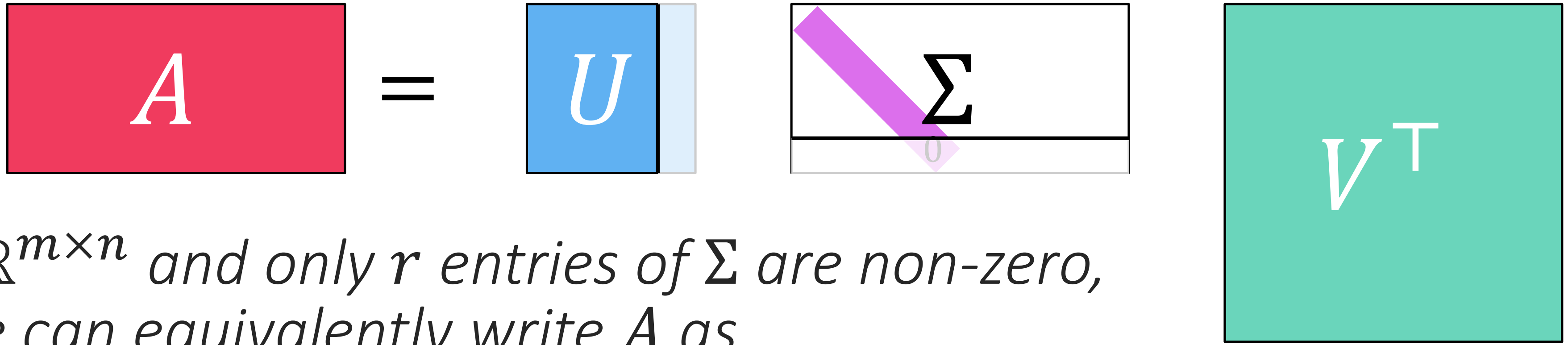
Not that this automatically shows that row rank and column rank are the same. $A^\top = V\Sigma^\top U^\top$ and the column rank of this matrix is clearly the same as that of the matrix A since the number of non-zeros in $\Sigma^\top$ does not change at all.



Getting the S

Rank: We always have $\mathbf{rank}(A)$ = number of non-zero entries in Σ

To see why, notice that if some diagonal entries of Σ are zero, we can remove those rows and the corresponding columns of U



If $A \in \mathbb{R}^{m\times n}$ and only r entries of Σ are non-zero, then we can equivalently write A as

$$A = \tilde{U}\tilde{\Sigma}V^\top$$

where $\tilde{U} \in \mathbb{R}^{m\times r}$ and $\tilde{\Sigma} \in \mathbb{R}^{r\times n}$. This is often called the thin SVD *of A*

Finally, notice that $A\mathbf{v} = \tilde{U}\mathbf{z}$ where $\mathbf{z} = \tilde{\Sigma}V^\top\mathbf{v} \in \mathbb{R}^r$ and thus, the column space of A is nothing but the span of columns of $\tilde{U}$ and clearly $\dim(\tilde{U}) = r$

Trace and Determinant

Trace and determinant are defined only for square matrices

Trace: For a square matrix $A \in \mathbb{R}^{d \times d}$, $\operatorname{tr}(A) = \sum_{j=1}^{d} A_{jj}$

Can use SVD to get another definition of trace

Trace has a funny property: $\operatorname{tr}(PQ) = \operatorname{tr}(QP)$ *for any* $P, Q \in \mathbb{R}^{d \times d}$

Trace also satisfies linearity properties if $P, Q \in \mathbb{R}^{d \times d}, c \in \mathbb{R}$, *then we have* $\operatorname{tr}(P + Q) = \operatorname{tr}(P) + \operatorname{tr}(Q)$ *and* $\operatorname{tr}(c \cdot P) = c \cdot \operatorname{tr}(P)$. *This gives us*

$$\operatorname{tr}(A) = \operatorname{tr}\left(\sum_{j=1}^{d} \sigma_i \, \mathbf{u}^i \left(\mathbf{v}^i\right)^\top\right) = \sum_{j=1}^{d} \sigma_i \cdot \operatorname{tr}\left(\mathbf{u}^i \left(\mathbf{v}^i\right)^\top\right)$$

$$= \sum_{j=1}^{d} \sigma_i \cdot \operatorname{tr}\left(\langle \mathbf{u}^i, \mathbf{v}^i \rangle\right) = \sum_{j=1}^{d} \sigma_i \cdot \langle \mathbf{u}^i, \mathbf{v}^i \rangle$$

Determinant: For a square matrix A, $|\det(A)| = \prod_j \sigma_j$

Sign of the $\det(A)$ depends on how many axes did U, V flip/swap

Inverses

A square matrix A is invertible iff all its singular values are non-zero

A singular value being zero means that A squishes vectors along some direction to the origin. This means that multiple input vectors map to the same output vector. This means that we cannot undo the linear map of A

Note that this means A is invertible iff $\det(A) \neq 0$

If $A = U\Sigma V^\top$ is invertible, then we always have $A^{-1} = V\Sigma^{-1}U^\top$

Nice because inverse of a diagonal matrix is simply element wise inverse

If $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_d)$, then $\Sigma^{-1} = \text{diag}\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_d}\right)$

Recall that we do insist $\sigma_i \neq 0$ i.e. $\sigma_i > 0$ and so $1/\sigma_i$ is well defined

Indeed $A^{-1}A = V\Sigma^{-1}U^\top A = V\Sigma^{-1}U^\top U\Sigma V^\top = V\Sigma^{-1}\Sigma V^\top = VV^\top = I$

Verify that $AA^{-1} = I$ as well

Pseudo Inverse

Even if a matrix $A = U\Sigma V^\top \in \mathbb{R}^{m\times n}$ is not invertible (even if A is not square), we can still define the Moore-Penrose inverse of A

$$A^+ = V\Sigma^+ U^\top \in \mathbb{R}^{n\times m}$$

if $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_\mathrm{k}) \in \mathbb{R}^{m\times n}$, where $k = \min\{m, n\}$ we define $\Sigma^+ = \mathrm{diag}(\tilde{\sigma}_1, \ldots, \tilde{\sigma}_\mathrm{k}) \in \mathbb{R}^{n\times m}$ where $\tilde{\sigma}_i = 1/\sigma_i$ if $\sigma_i > 0$ else $\tilde{\sigma}_i = 0$

The pseudo inverse satisfies a few nice properties

If A is indeed invertible then $A^+ = A^{-1}$

If $A = [\mathbf{a}^1, \ldots, \mathbf{a}^n]$ with $\mathbf{a}^j \in \mathbb{R}^m$, then $AA^+\mathbf{a}^j = \mathbf{a}^j$ i.e. AA^+ acts as identity for columns of A. This means that $AA^+A\mathbf{z} = A\mathbf{z}$ for all $\mathbf{z} \in \mathbb{R}^n$. However this means that if $\mathbf{u} = A\mathbf{z}$, then $A^+\mathbf{u}$ is a vector $\mathbf{v}$ such that $A\mathbf{v} = \mathbf{u} = A\mathbf{z}$. Thus, A^+ does invert the map of A by sending a valid output $\mathbf{u}$ of the map A to an input $\mathbf{v}$ which does indeed generate that very output with that map.

Can be shown that v is the vector with smallest L2 norm s.t. $A\mathbf{v} = \mathbf{u}$

Pseudo Inverse

Even if a matrix $A = U\Sigma V^\top \in$ ...

square), we can still define the Moore-Penrose inverse of A

$$A^+ = \ldots$$

if $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_k) \in \mathbb{R}^{m \ldots}$

$\Sigma^+ = \text{diag}(\tilde{\sigma}_1, \ldots, \tilde{\sigma}_k) \in \mathbb{R}^{n \times \ldots}$ where $\tilde{\sigma}_i = 1/\sigma_i$ if $\sigma_i > 0$ else 0

The pseudo inverse satisfies a few nice properties

If A is indeed invertible ...

If $A = [\mathbf{a}^1, \ldots, \mathbf{a}^n]$ with ... *is i*...

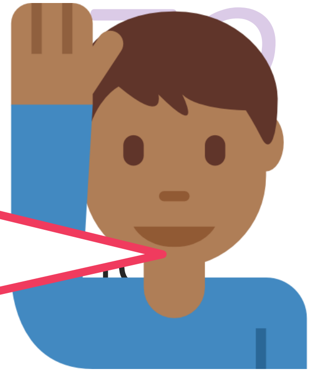
for columns of A. This means that ... *for all* $\mathbf{z} \in \mathbb{R}^{\ldots}$*. Howev*...

me... *Thus,*

A^+ ... *A to an*

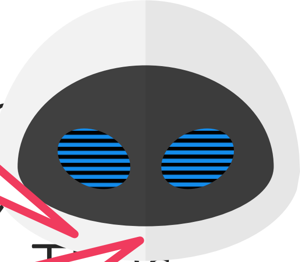
inp...

Ca...



Indeed, the pseudo inverse always gives us the "least squares" solution. If $\mathbf{u} = A\mathbf{z}$, then $A^+\mathbf{u} = \arg\min \|\mathbf{x}\|_2^2$ s.t. $A\mathbf{x} = \mathbf{u}$

Least squares?? Does this have anything to do with the least squares we did in regression?



It sure does. If $X^\top X$ is invertible, then we can show that $X^+ = (X^\top X)^{-1} X^\top$

Thus, least squares solution simply gives us $X^+\mathbf{y}$

In fact, even if there exists no $\mathbf{z} \in \mathbb{R}^m$, s.t. $\mathbf{u} = A\mathbf{z}$, even then, $A^+\mathbf{u}$ returns the solution which has the smallest error i.e.

$\|AA^+\mathbf{u} - \mathbf{u}\|_2^2 = \arg\min \|A\mathbf{x} - \mathbf{u}\|_2^2$ and among all vectors with this smallest error, $A^+\mathbf{u}$ is the one with smallest L2 norm