Lecture 17

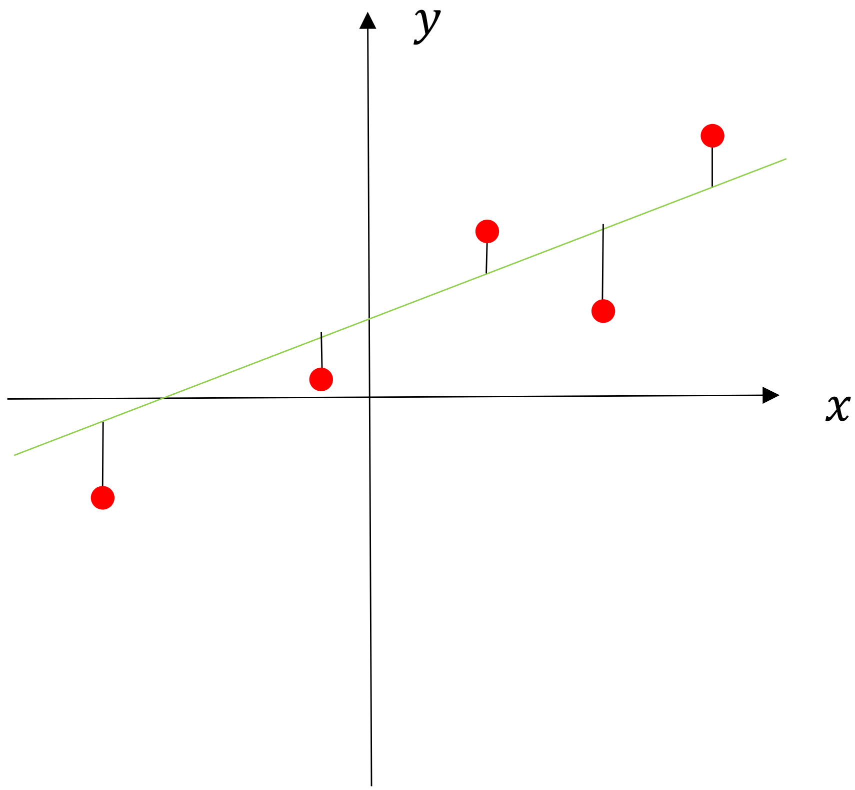
Least squares estimation

Singular value decomposition (SVD)

Recall from lecture 5.

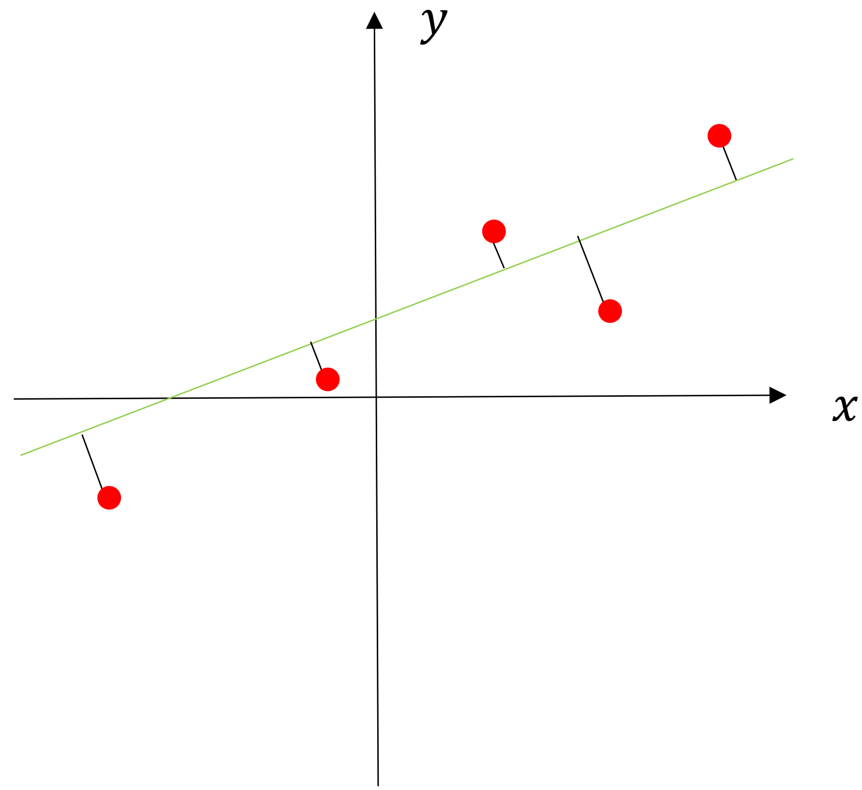
Version 1: linear regression

Error is distance to line in y direction only.



Version 2: "total least squares"

Error is distance perpendicular to line.



Least squares: version 2

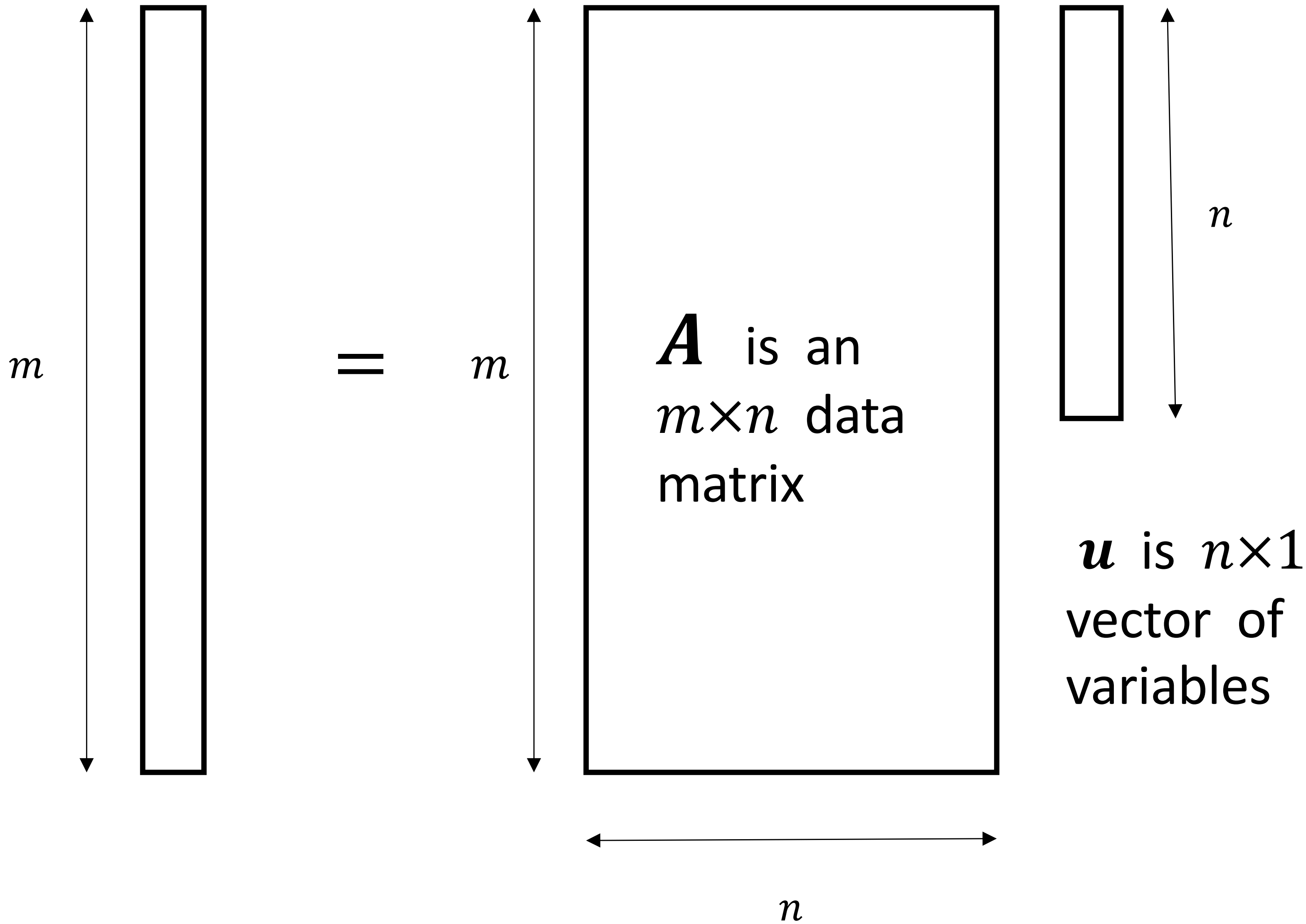
Find the $\boldsymbol{u}$ that minimizes L2 norm $\| \boldsymbol{A}\,\boldsymbol{u} \|^2$

subject to $\| \boldsymbol{u} \| = 1.$

where

- $\boldsymbol{A}$ is an $m \times n$ data matrix, $m \geq n$
- $\boldsymbol{u}$ is a $n \times 1$ vector of variables

Find unit vector $\boldsymbol{u}$ that minimizes the L2 norm of $\boldsymbol{Au}$.



Examples of Version 2 Problems (next two weeks)

- Camera calibration
- Image Stitching for panoramas
- Binocular Stereo

Least squares: version 2

Find the $\boldsymbol{u}$ that minimizes L2 norm $\| \boldsymbol{A}\, \boldsymbol{u} \|^2$

subject to $\| \boldsymbol{u} \| = 1$, where

- $\boldsymbol{A}$ is an $m \times n$ data matrix, $m \geq n$
- $\boldsymbol{u}$ is a $n \times 1$ vector of variables

<u>Solution (claimed back in lecture 5):</u>

Compute the eigenvectors of $\boldsymbol{n} \times \boldsymbol{n}$ matrix $\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A}$.

Take the unit eigenvector that has the smallest eigenvalue.

Why does the eigenvector of $\boldsymbol{A}^T\boldsymbol{A}$ with the minimum eigenvalue solve this problem ?

Main idea: first suppose $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}^T\boldsymbol{A}$:

$$\| \boldsymbol{A}\,\boldsymbol{u} \|^2 = \boldsymbol{u}^T\boldsymbol{A}^T\boldsymbol{A}\,\boldsymbol{u}$$

$$= \lambda\,\boldsymbol{u}^T\boldsymbol{u}$$

if $\boldsymbol{u}$ is an eigenvector (and λ is its eigenvalue)

$$= \lambda$$

when $\boldsymbol{u}$ has unit length

$$\geq 0$$

because L2 norm is non-negative.

So we want the eigenvector with smallest eigenvalue.

Why does the eigenvector of $A^T A$ with the minimum eigenvalue solve the problem ?

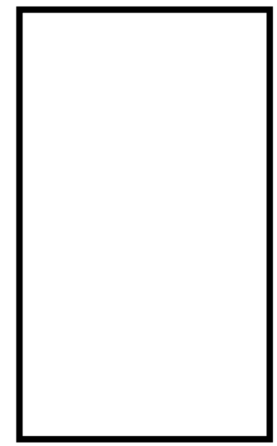
Linear algebra tells us that the $n \times n$ matrix $\boldsymbol{A}^T \boldsymbol{A}$ has

- n orthogonal eigenvectors
- non-negative eigenvalues

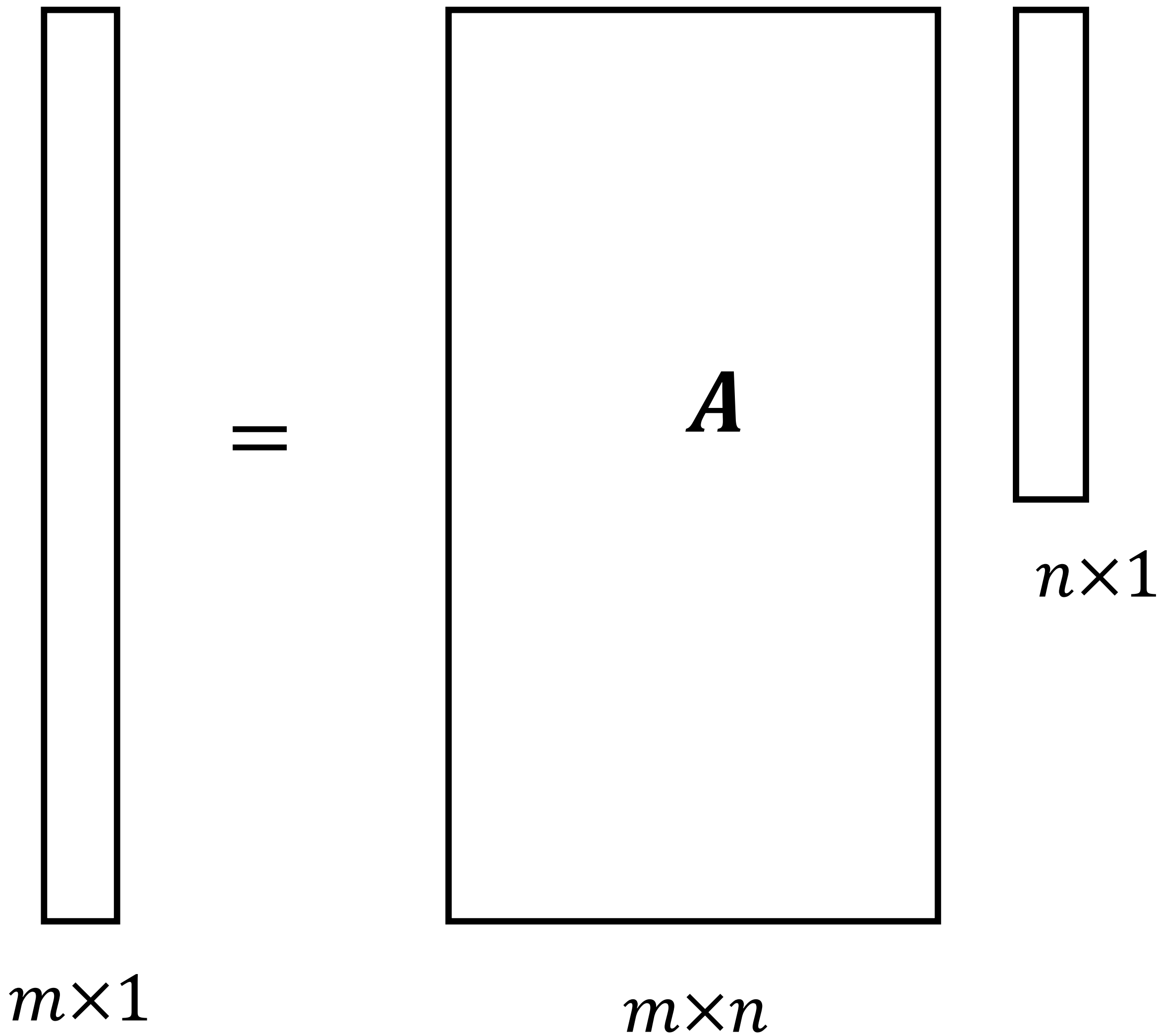
Therefore we can write any vector $\boldsymbol{u}$ as a sum of these eigenvectors. By inspection, $\boldsymbol{u}^T \boldsymbol{A}^T \boldsymbol{A} \, \boldsymbol{u}$ will be smallest when $\boldsymbol{u}$ is the eigenvector with smallest eigenvalue.

Singular Value Decomposition (SVD)

Let $\boldsymbol{A}$ be any $m \times n$ real data matrix.



In our examples, $m \geq n$.


$=$
A
$n\times 1$
$m\times 1$
$m\times n$

Since the eigenvectors of $\boldsymbol{A^T A}$ are orthogonal and the eigenvalues are non-negative, we can write:

$$\boldsymbol{A^T A\ V = V\ D}$$

where

- the columns of $\boldsymbol{V}$ are ortho*normal* eigenvectors of $\boldsymbol{A^T A}$
- $\boldsymbol{D}$ is a diagonal matrix of (non-negative) eigenvalues

Since the eigenvectors of $A^T A$ are orthogonal and the eigenvalues are non-negative, we can write:

$$\boldsymbol{A^T A\ V = V\ \Sigma^2}$$

where

- the columns of $\boldsymbol{V}$ are ortho*normal* eigenvectors of $\boldsymbol{A^T A}$
- $\boldsymbol{\Sigma}$ is a diagonal matrix, whose elements are called the *singular* values of $\boldsymbol{A}$

$$A^T A \; V \;\; = \;\; V \, \boldsymbol{\Sigma}^2$$

Multiplying on the left by V^T gives:

$$V^T A^T A \; V \;\; = \boldsymbol{\Sigma}^2$$

$$\boldsymbol{A}^T \boldsymbol{A} \ \boldsymbol{V} \ = \ \boldsymbol{V} \, \boldsymbol{\Sigma}^2$$

Multiplying on the left by $\boldsymbol{V}^T$ gives:

$$\boldsymbol{V}^T \boldsymbol{A}^T \boldsymbol{A} \ \boldsymbol{V} \ = \boldsymbol{\Sigma}^2$$

By inspection, the columns of $\boldsymbol{A} \ \boldsymbol{V}$ are orthogonal.
Therefore we can uniquely define a matrix $\boldsymbol{U}$ such that:

$$\boldsymbol{A} \ \boldsymbol{V} \ = \boldsymbol{U} \ \boldsymbol{\Sigma}$$

where the columns of $\boldsymbol{U}$ are parallel to columns of $\boldsymbol{A} \, \boldsymbol{V}$ (and orthonormal).

Q: What are the magnitudes of the columns of $\boldsymbol{A} \, \boldsymbol{V}$?

Singular Value Decomposition (SVD)

From the previous slide:

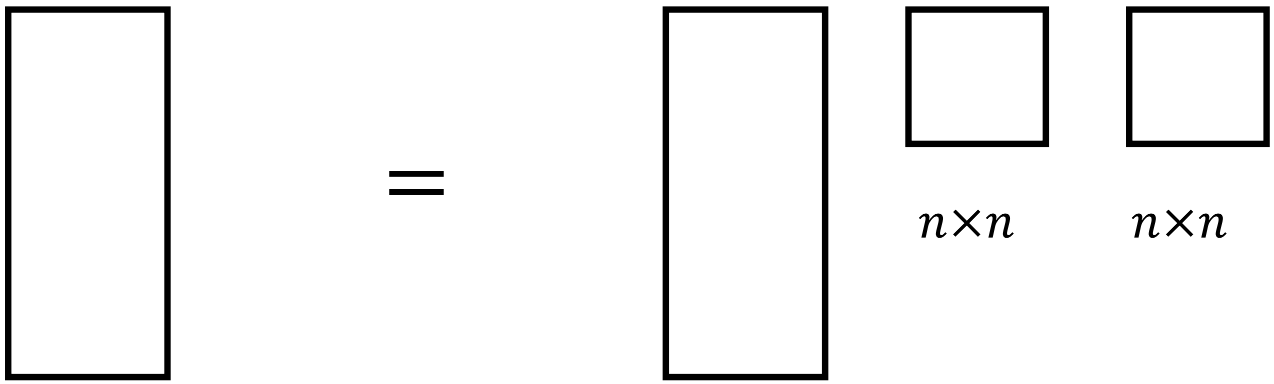
$$\boldsymbol{A}\ \boldsymbol{V} = \boldsymbol{U}\ \boldsymbol{\Sigma}$$

Since the columns of $\boldsymbol{V}$ are orthonormal, right multiplying by $\boldsymbol{V}^{\boldsymbol{T}}$ gives:

$$\boxed{\boldsymbol{A} = \boldsymbol{U}\ \boldsymbol{\Sigma}\ \boldsymbol{V}^{\boldsymbol{T}}}$$

This is known as the Singular Value Decomposition (SVD) of $\boldsymbol{A}$.

What does the matrix $\boldsymbol{A}$ do when you multiply it by an n vector: $\boldsymbol{Ax}$?



$n \times n$ $\quad$ $n \times n$

$m \times n$ $\quad$ $m \times n$

$$\boldsymbol{A} = \boldsymbol{U} \, {\color{red}\boldsymbol{\Sigma}} \, \boldsymbol{V}^{\boldsymbol{T}}$$

Use the coefficients as weights on $\boldsymbol{n}$ columns of $\boldsymbol{U}$ in $\boldsymbol{R}^{\boldsymbol{m}}$ and sum up.

Scale coefficients by multiplying by singular values.

Rotate to coordinates defined by orthonormal columns of $\boldsymbol{V}$ in $\boldsymbol{R}^{\boldsymbol{n}}$ (orthonormal rows of $\boldsymbol{V}^{\boldsymbol{T}}$)

Matlab

```
[U, S, V]  =  svd(A)
```

n singular values
returned as a vector
in decreasing order

Examples of Version 2 Problems (coming soon)

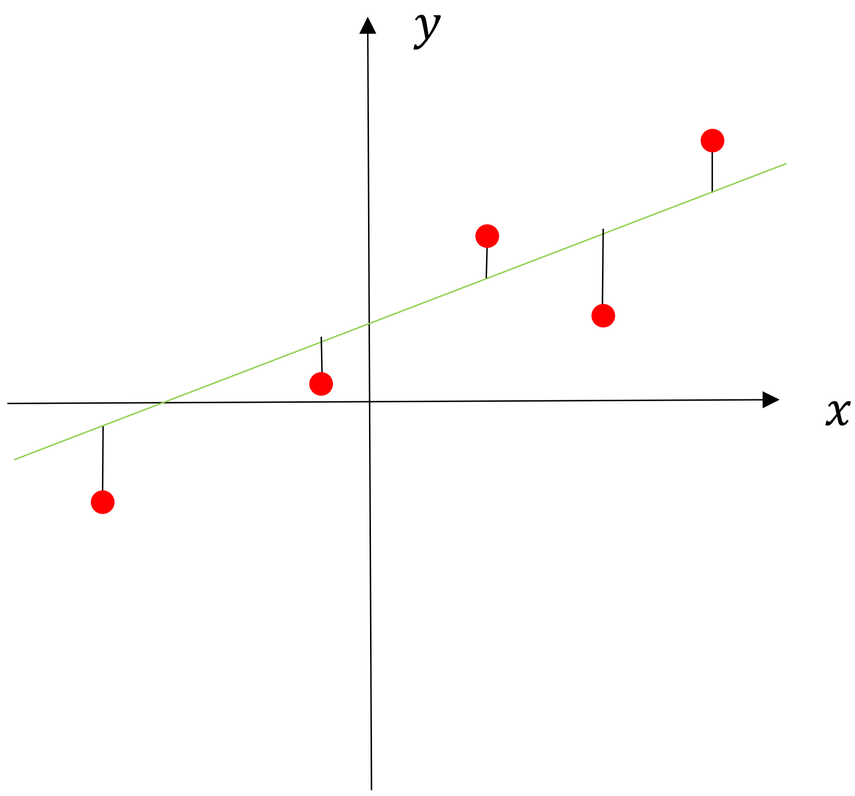
- Camera calibration
- Image Stitching for panoramas
- Binocular Stereo

For each of these problems, we will set up a data matrix `A` and solve the problem by taking the SVD. The solution will be the column of matrix `V` that corresponds to the smallest singular value.

Recall from lecture 5.

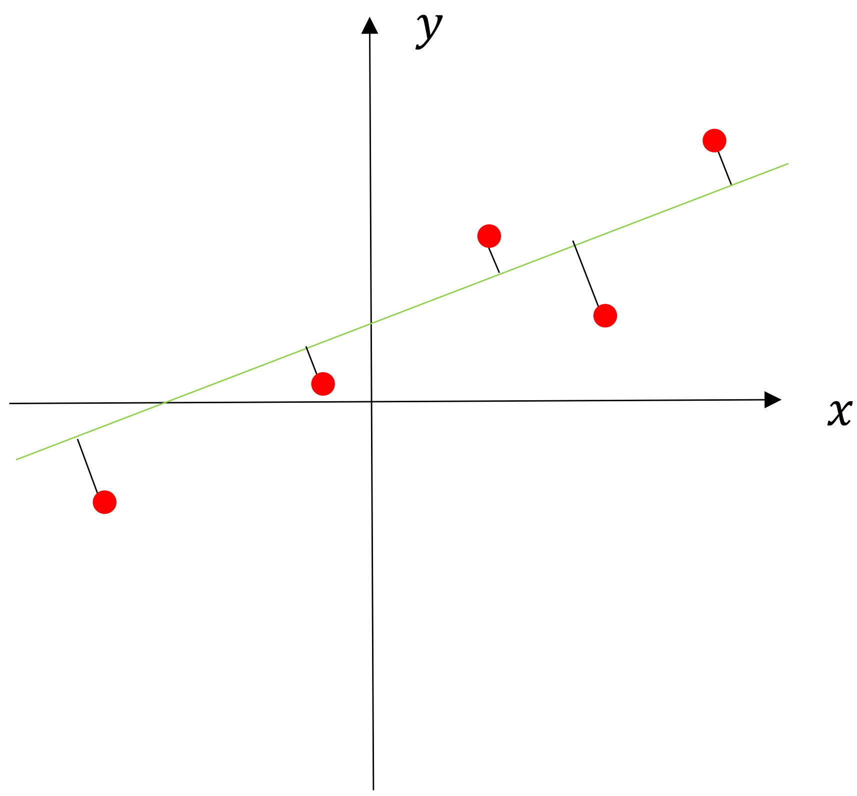
Version 1: linear regression

Error is distance to line in y direction only.



Version 2: "total least squares"

Error is distance perpendicular to line.

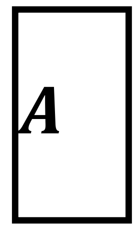


Given $\mathbf{A}$ and $\mathbf{b}$ defined below, find the $\mathbf{u}$ that minimizes:

$$\| \mathbf{A}\,\mathbf{u} - \mathbf{b} \|^2$$

where

$\mathbf{A}$ is an $m \times n$ matrix, where $m \geq n$



$\mathbf{u}$ is a $n \times 1$ vector of variables

$\mathbf{b}$ is a $m \times 1$ data vector (lecture 5) To find the $\mathbf{u}$ that minimizes:

$$\| \mathbf{A}\, \mathbf{u} - \mathbf{b} \|^2$$

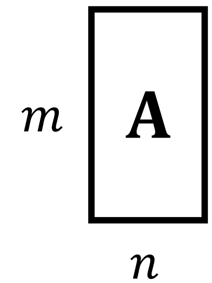
we solve for:

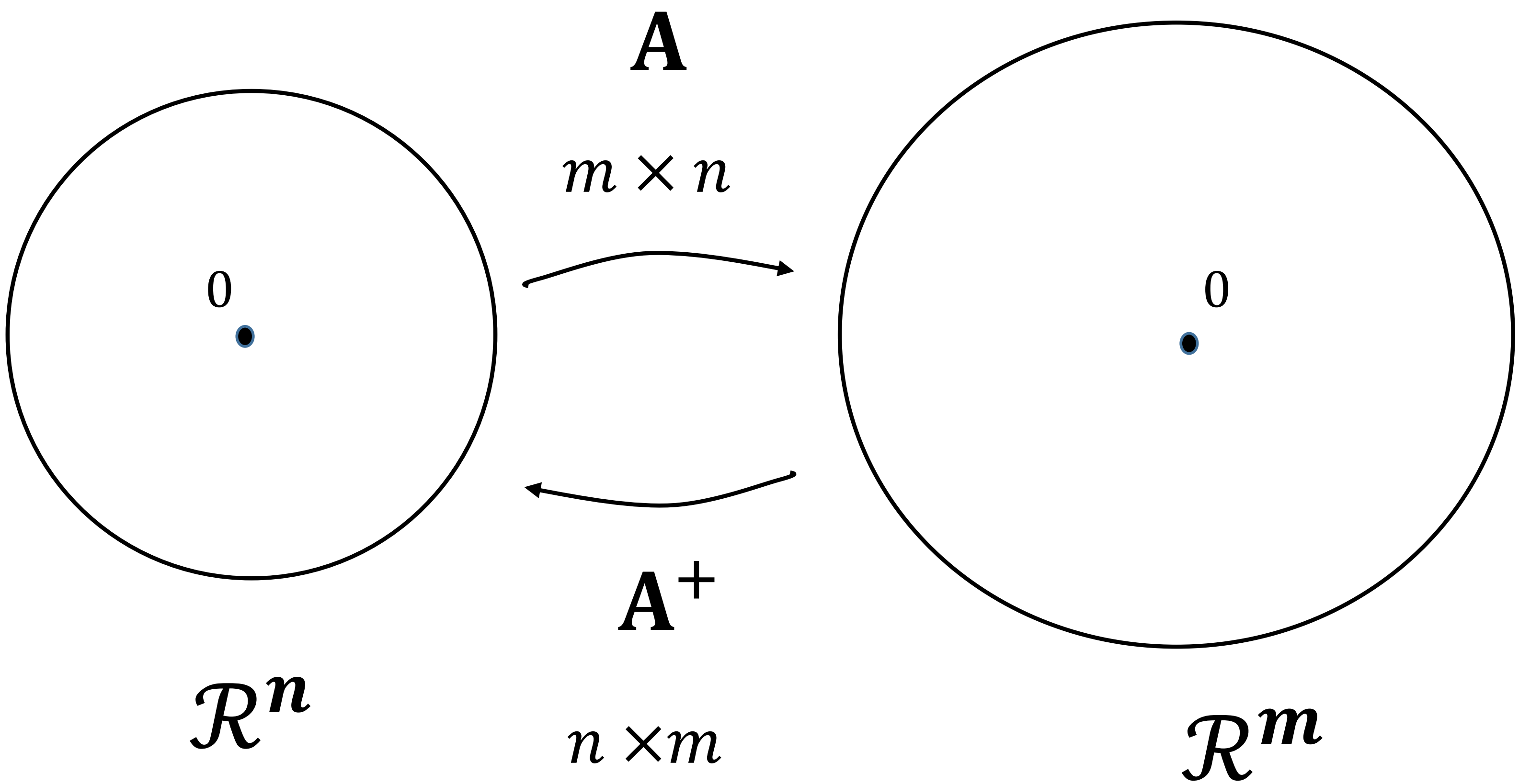
$$\mathbf{u} = \underbrace{(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T}\mathbf{b}$$

This matrix is called the *pseudoinverse* of **A**.
It is typically written $\boldsymbol{A}^+$.

Let's give a geometric interpretation of this matrix and then relate it to the SVD.

pseudoinverse $\mathbf{A}^{+} \equiv \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}$

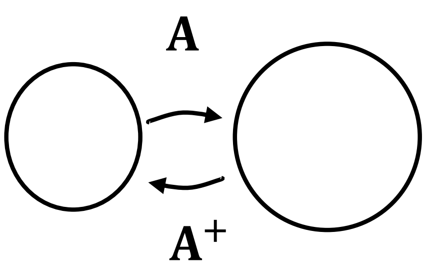




What does the pseudoinverse do?

$$\mathbf{A^{+} \equiv \left(A^{T}A\right)^{-1}A^{T}}$$

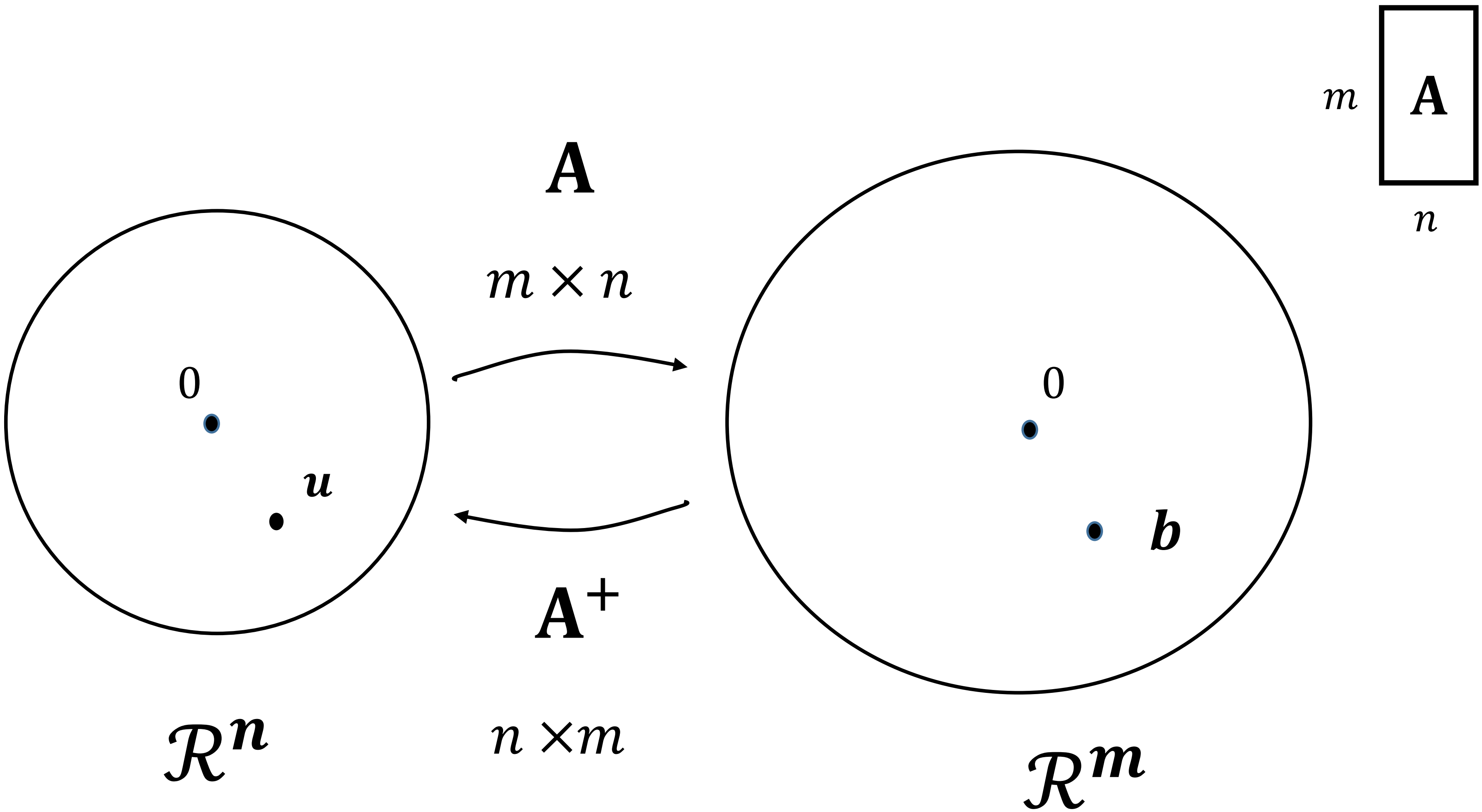
Thus,



$$\mathbf{A^{+}A = \left(A^{T}A\right)^{-1}A^{T}A \ = I}$$

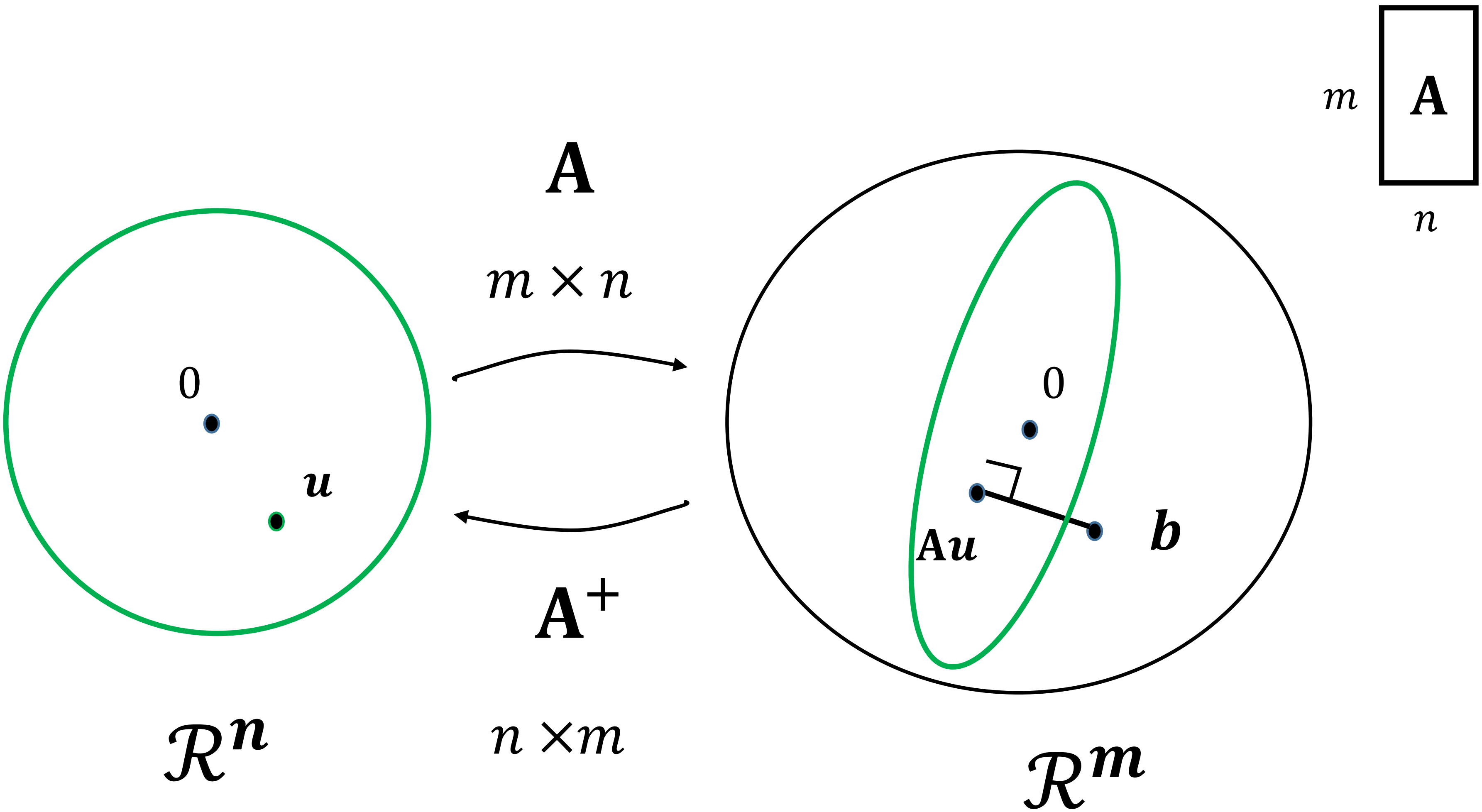
$$\mathbf{A\,A^{+} = \ A\left(A^{T}A\right)^{-1}A^{T} \neq I}$$

equality if $\boldsymbol{A}$ is invertible.



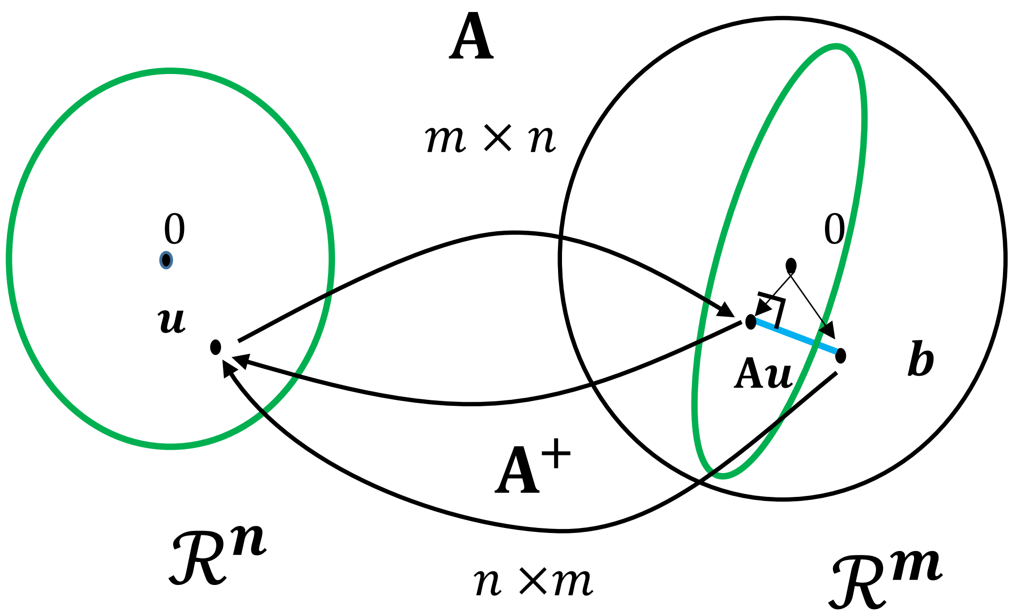
Given $\mathbf{A}$ and $\mathbf{b}$ in $\mathcal{R}^{\boldsymbol{m}}$, find the $\mathbf{u}$ in $\mathcal{R}^{\boldsymbol{n}}$ that minimizes $\| \mathbf{A}\,\mathbf{u} - \mathbf{b} \|^2$.

The solution is $\mathbf{u} = \mathbf{A}^{+}\mathbf{b}$.



Given $\mathbf{A}$ and $\mathbf{b}$ in $\mathcal{R}^{\boldsymbol{m}}$, find the $\mathbf{u}$ in $\mathcal{R}^{\boldsymbol{n}}$ that minimizes $\| \mathbf{A}\,\mathbf{u} - \mathbf{b} \|^2$.

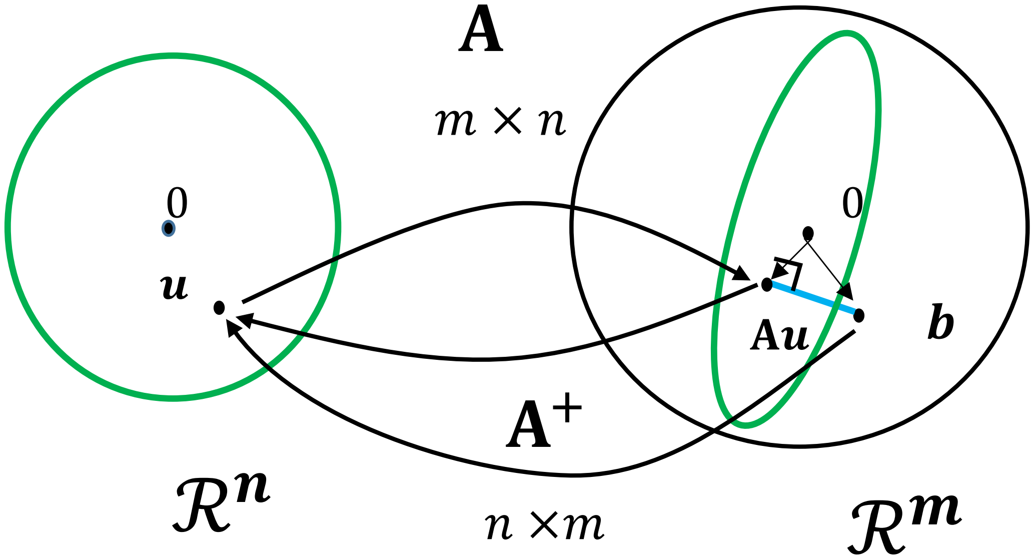
The solution is $\mathbf{u} = \mathbf{A}^+\mathbf{b}$. Intuitively (and as we'll argue next), $\mathbf{Au}$ is the orthogonal projection of $\boldsymbol{b}$ onto the column space of $\mathbf{A}$.

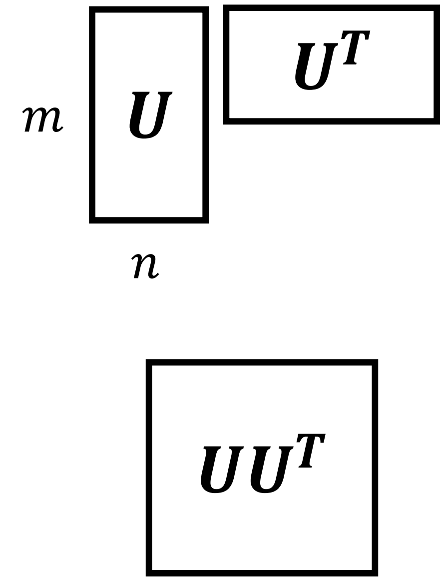


Substitute $\mathbf{A} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^{\mathrm{T}}$ into $\mathbf{A}^{+} \equiv \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}$

Exercise : Show this gives $\mathbf{A}^{+} = \mathbf{V}\,\boldsymbol{\Sigma}^{-1}\mathbf{U}^{\mathrm{T}}$.

Exercise : Show $\mathbf{A}\,\mathbf{A}^{+} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$.





$$\mathbf{V}\,\boldsymbol{\Sigma}^{-1}\mathbf{U}^{\mathbf{T}}\,.$$

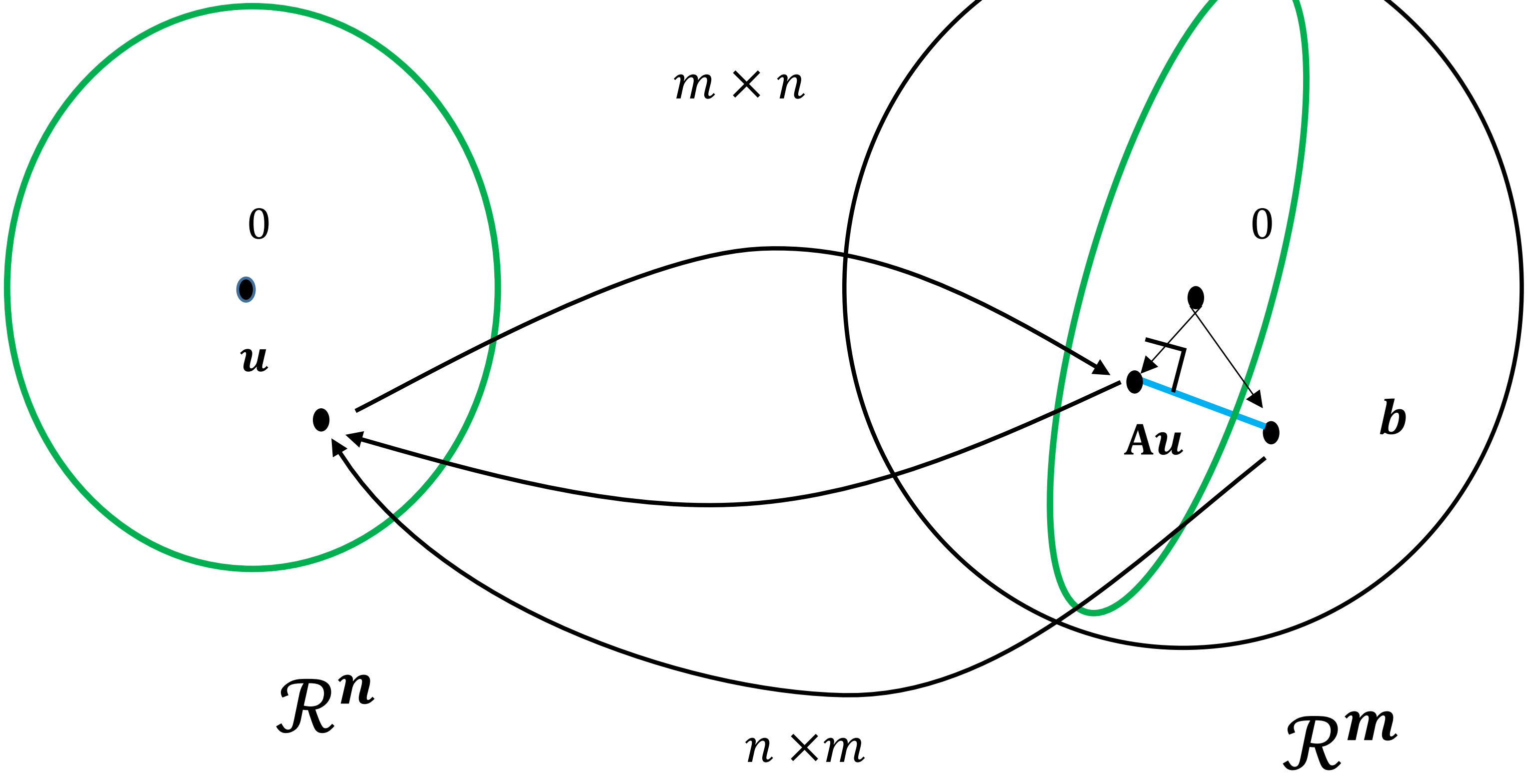
$$\boldsymbol{A}\,\boldsymbol{A}^{+} = \boldsymbol{U}\boldsymbol{U}^{\boldsymbol{T}}\,\boldsymbol{b}$$

$\boldsymbol{U}\boldsymbol{U}^{\boldsymbol{T}}$ is an $m \times m$ matrix which projects $\boldsymbol{b}$ to the column space of $\boldsymbol{A}$.

$$\mathbf{A} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^{\mathbf{T}}$$

V column vectors are here.

U column vectors are here.



$$\mathbf{A}^{+} = \boldsymbol{V}\,\boldsymbol{\Sigma}^{-\mathbf{1}}\mathbf{U}^{\boldsymbol{T}}$$

Next Two Weeks…

- Camera calibration
- Image Stitching for panoramas
- Binocular Stereo