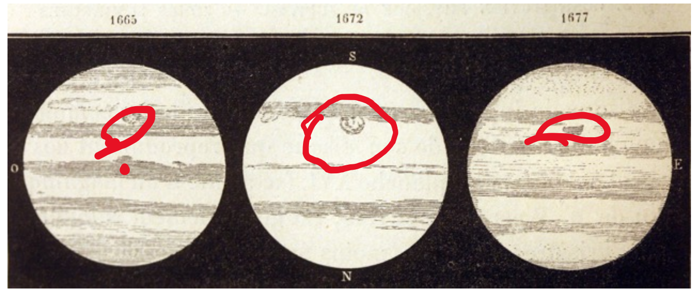
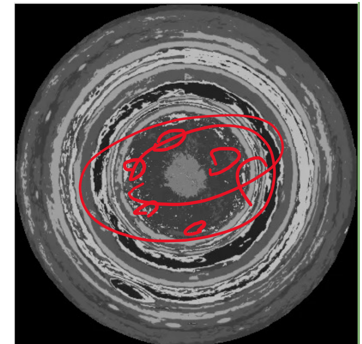
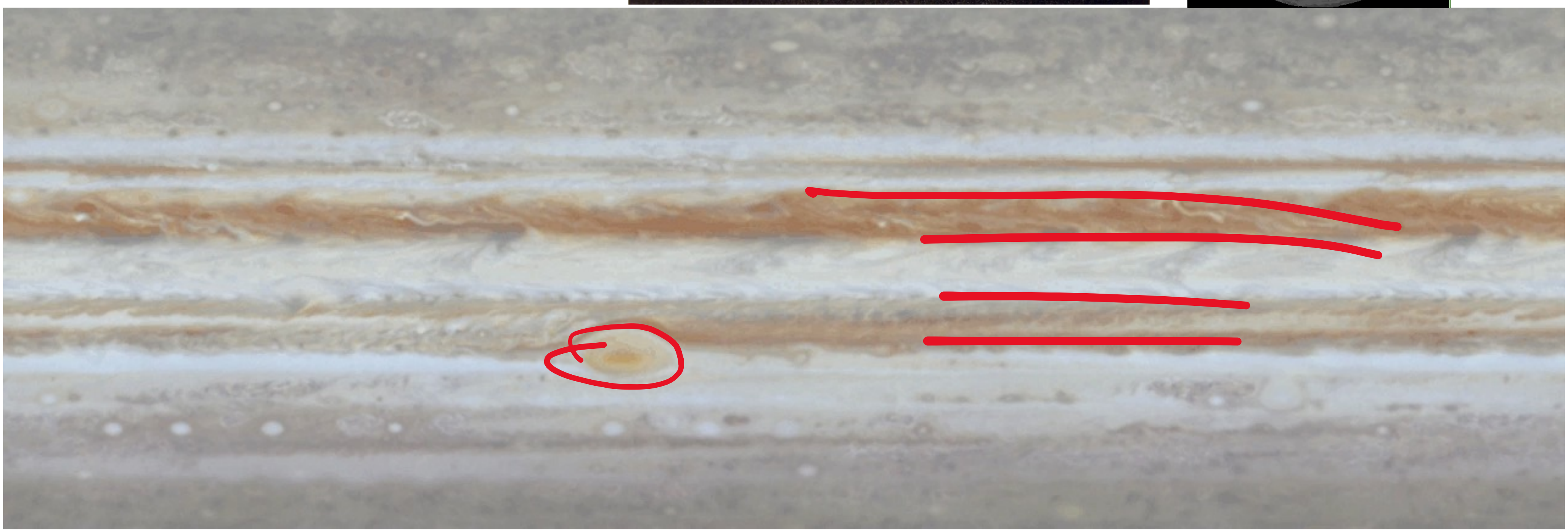
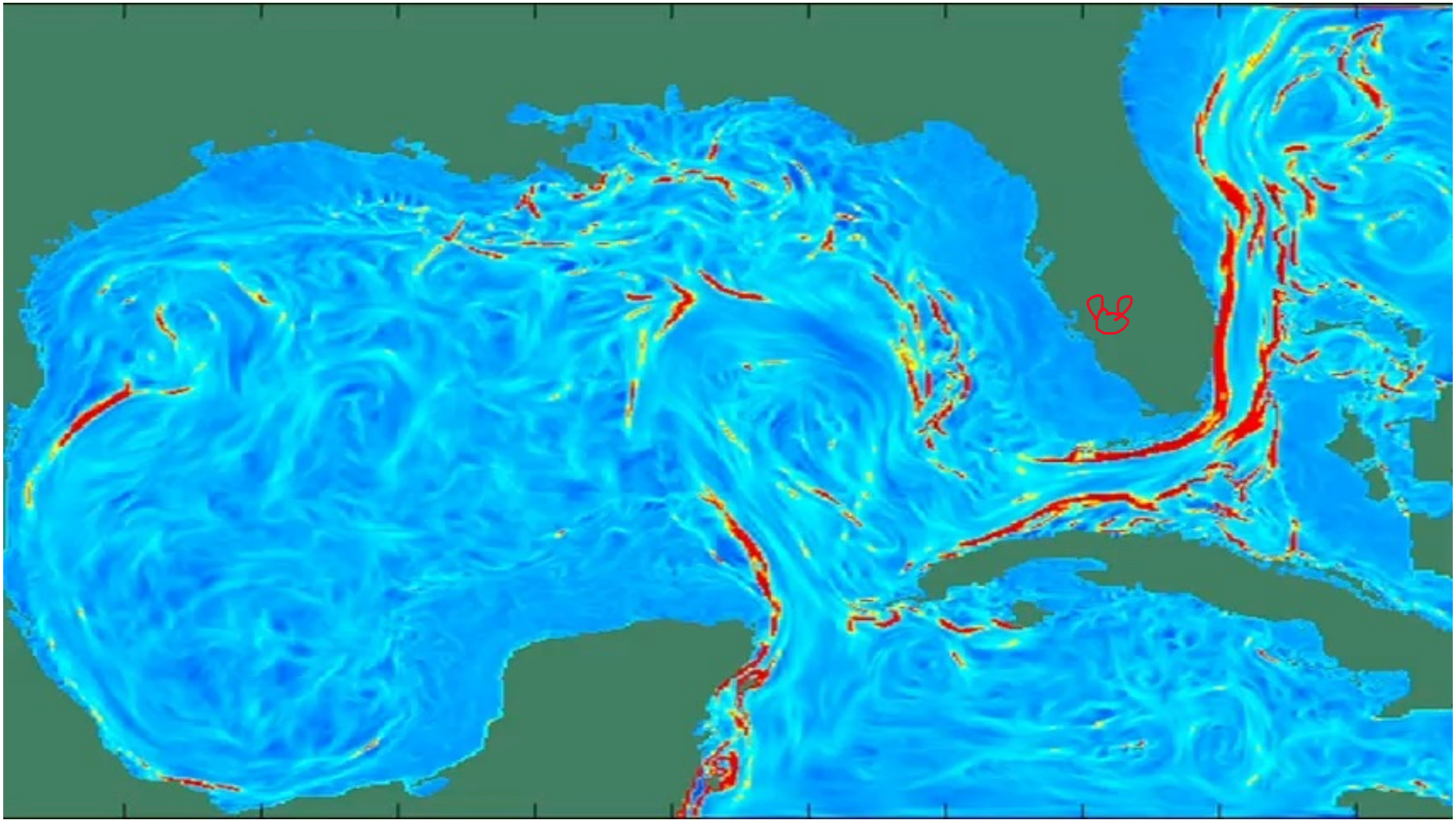
EE520 Data Driven Analysis of Complex Systems

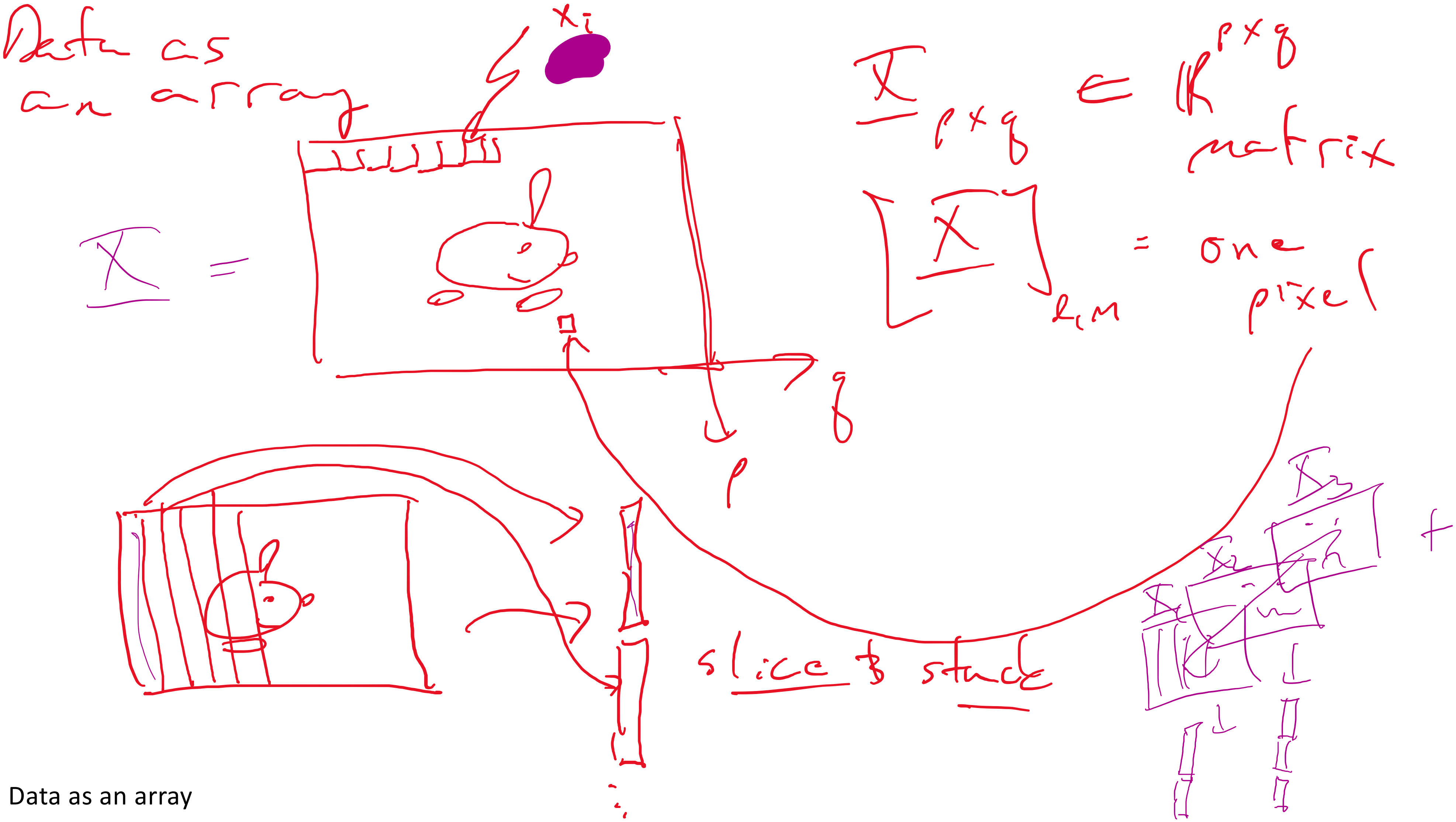
Erik Bollt











Data as an array

On Matrix Multiplication

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}, \text{ and each } a_{i,j} \in \mathbb{C}$$

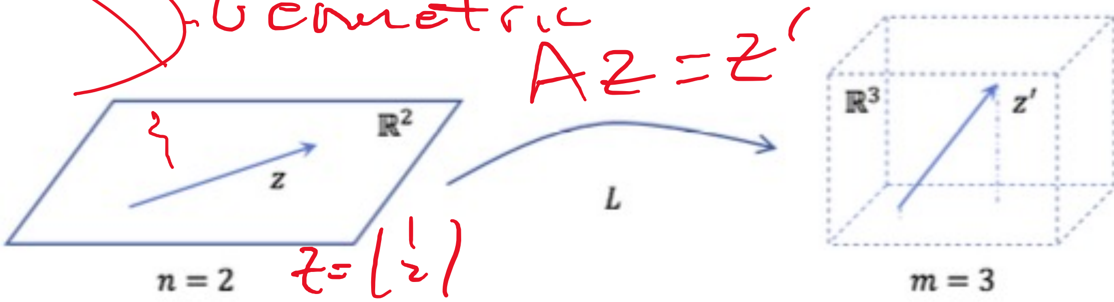
$$\begin{aligned} L(\mathbf{z}) : \mathbb{R}^n &\to \mathbb{R}^m \\ \mathbf{z} &\mapsto \mathbf{z}' = A\mathbf{z}, \end{aligned}$$

in terms of the usual matrix times vector multiplication,

$$[\mathbf{z}]'_i = \sum_{j=1}^{n} A_{i,j} [\mathbf{z}]_j, \text{ for each } i = 1, \cdots, m,$$

Geometric

$Az = z'$



$z = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

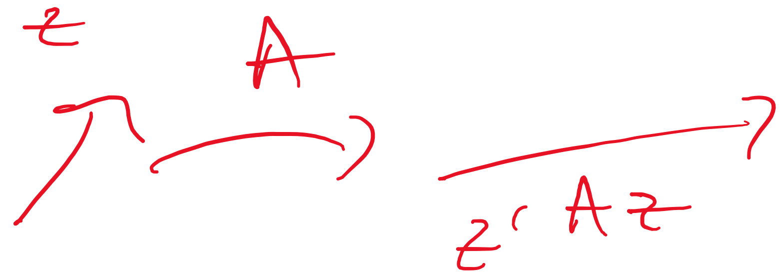
• a vector has length & direction.

But as linear algebra ; Matrices × vectors

$$A_{2\times 2} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

$$A\begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1\cdot 3 + 2\cdot 4 \\ 3\cdot 3 + 4\cdot 4 \end{pmatrix} = \begin{pmatrix} 11 \\ 25 \end{pmatrix}$$

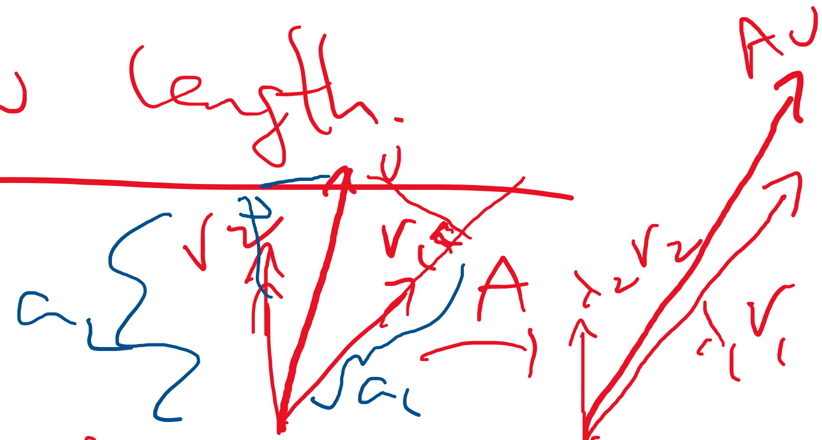
$z' = Az$



new direction, new length.

Eig for square.

① $\underset{2\times 2}{A}v = \lambda v$



Characterize matrices by knowing just these

② e.g. special directions

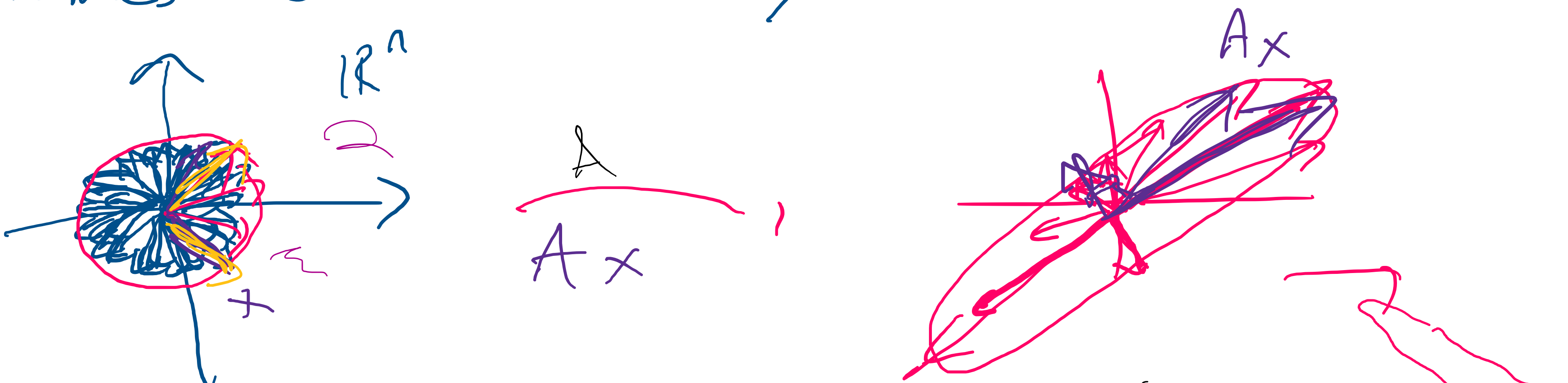
• $Au = A(a_1 v_1 + a_2 v_2) = a_1 A v_1 + a_2 A v_2$ $\quad \det(A - \lambda I) = 0$

$= a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2$ $\quad (A - \lambda I)v = 0$

Matrix time circle =

all vectors of length 1.

? Matrix x circle ?! But matrix times vector. ↗ = ↗



$S = \{ x \mid \|x\|_2 = 1,\ x \in E \cong \mathbb{R}^2 \}$; $A \cdot S \equiv \{ y : y = Ax,\ x \in S \}$

Theorem 2.1.1 — Singular Value Decomposition. Let A be an $m \times n$ matrix whose entries come from the field $\mathcal{K}$, which is either the field of real numbers or the field of complex numbers. Then the singular value decomposition of A exists, and it takes the form of a product of matrices:

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^*_{n\times n}, \tag{2.5}$$

where

- U is an $m \times m$ unitary matrix.
- Σ is a diagonal $m \times n$ matrix with non-negative real numbers on the diagonal.
- V is an $n \times n$ unitary matrix, and V^* is the conjugate transpose of V.

$$\Sigma := diag(\sigma_1, \sigma_2, \cdots, \sigma_p), p = min(m, n),$$
$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0.$$

$$\Sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots \end{pmatrix}$$

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

$$u_1^T u_1 = u_1 \cdot u_1 = 1$$

$$\begin{pmatrix} u_1^T u_1 & u_1^T u_2 & u_1^T u_3 & \cdots \\ u_2^T u_1 & u_2^T u_2 & \cdots & \end{pmatrix}$$

$$u_i^T u_j = \delta_{ij} = 1 \text{ or } 0$$

The singular values are the nonegative values: $\sigma_i \geq 0, i = 1, \cdots, n$,

The left singular vectors: u_i are the columns of $U = [u_1, u_2, ..., u_m]$.

The right singular vectors: v_i are the columns of $V = [v_1, v_2, ..., v_n]$.

$$\|u\|^2 \cos\theta = \vec{u_1} \cdot \vec{u_1} = I \qquad u_2^T u_1 = 0$$

$$\circ\ U \text{ unitary} \Leftrightarrow U^* U = U U^* = I$$

$$U^T U = \begin{pmatrix} u_1 & u_2 & \cdots & u_m \end{pmatrix}^T \begin{pmatrix} u_1 & u_2 & \cdots & u_m \end{pmatrix} = I$$

Definition 2.1.1 — Singular values and singular vectors. The singular values of A are the scalar values, σ_i, and the columns of U and V have columns that are the corresponding i^{th} left and right singular vectors, u_i and v_i:

The singular values are the nonegative values: $\sigma_i \geq 0, i = 1, \cdots, n$,

The left singular vectors: u_i are the columns of $U = [u_1, u_2, ..., u_m]$.

The right singular vectors: v_i are the columns of $V = [v_1, v_2, ..., v_n]$.

Since V is orthogonal, then right multiplying Eq. (2.5) by V,

$$AV = U\Sigma V^* V = U\Sigma, \tag{2.8}$$

Fri 08/21/20

$Ax = b$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

if

A is square

$AV = U\Sigma$

$$[A]\,[v_1 v_2 \cdots v_n] = [u_1 u_2 \cdots u_n]\, diag(\sigma_1, \sigma_2, \cdots, \sigma_n).$$

■ **Example 2.1** Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix}_{2\times 3}$. By SVD of the matrix A we have:

$$\begin{aligned} A &= U\Sigma V^T \\ &= \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \sqrt{70} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{14}} & \sqrt{\frac{2}{7}} & \frac{3}{\sqrt{14}} \\ \frac{-3}{\sqrt{10}} & 0 & \frac{1}{\sqrt{10}} \\ \frac{-1}{\sqrt{35}} & \sqrt{\frac{5}{7}} & \frac{-3}{\sqrt{35}} \end{pmatrix}. \end{aligned} \tag{2.28}$$

We see that the second singular value, $\sigma_2 = 2$, meaning that number of non-zero singular values $r < \min\{m, n\}$. Such matrix is called rank deficient matrix. If we take the economy version (with $r = 1$) of the SVD we will have:

$$\begin{aligned} u_1 \sigma_1 v_1^T &= \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} \end{pmatrix} \left(\sqrt{70}\right) \begin{pmatrix} \frac{1}{\sqrt{14}} & \sqrt{\frac{2}{7}} & \frac{3}{\sqrt{14}} \end{pmatrix} \\ &\approx \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix} \end{aligned} \tag{2.29}$$

■

$A^TA = V\Sigma^T\Sigma V^T$

$A^TAV = \Sigma^T\Sigma V$

$V = (v_1\ v_2 \cdots v_n)$

$$[A]\,[v_1 v_2 \cdots v_n] = [u_1 u_2 \cdots u_n]\, diag(\sigma_1, \sigma_2, \cdots, \sigma_n).$$

Full ☆

The Economy SVD, and Reduced Rank SVD

The general SVD, Eq. (2.5) may be written in terms of submatrices.

Definition 2.1.3 — The Economy SVD. For any matrix $A \in \mathbb{R}^{m\times n}$, the general SVD Eq. (2.5) can be written in terms of smaller matrices,

$$A_{m\times n} = \hat{U}_{m\times n}\hat{\Sigma}_{n\times n}V^*_{n\times n}, \tag{2.21}$$

and $U = [\hat{U}_{m\times n}|\hat{U}_{(n-m)\times n}]$, written in terms of an orthogonal "buffer" matrix

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = 0$

$= 0 \dots$

Definition 2.1.4 — Rank Deficient SVD. For a matrix $A \in \mathbb{R}^{m\times n}$ such that the SVD results in singular values

$$\sigma_r > \sigma_{r+1} = 0, \text{ for some } r < n. \tag{2.22}$$

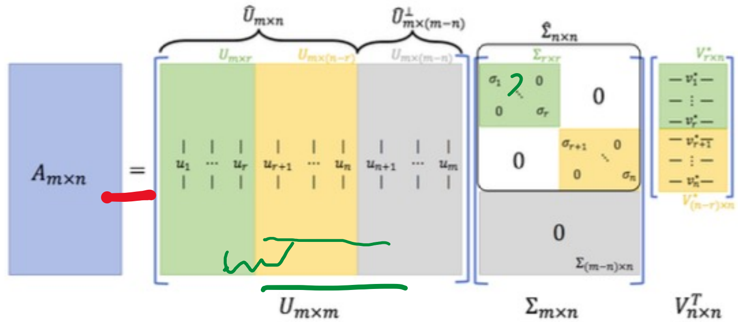
then the SVD can be written in terms of an economy form as smaller matrices,

$$A_{m\times n} = \hat{U}_{m\times r}\hat{\Sigma}_{n\times n}V^*_{n\times r}, \tag{2.23}$$

and related to the general SVD Eq. (2.5) by $U = [\hat{U}_{m\times r}|\hat{U}_{(n-r)\times n}]$, but $r < n$.

Rank ill conditioned

Full,
Economy,
Truncated
SVD



m>n

Figure 2.3: $m > n$ tall skinny

$\sigma_{r+1} = 0$

$\sigma_r = 10^{-6}$

$\sigma_{r+1} \sim 10^{-13}$

$< \varepsilon$

m<n

$A = [\quad]$

Recall that,

$$
\begin{aligned}
A_{m\times n} &= \hat{U}_{m\times n}\hat{\Sigma}_{n\times n}\hat{V}^T_{n\times n} \\
&= \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_n \\ | & | & & | \end{bmatrix}
\begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \end{bmatrix}
\begin{bmatrix} - & v_1^T & - \\ - & v_2^T & - \\ - & \vdots & - \\ - & v_n^T & - \end{bmatrix}
\end{aligned} \tag{2.24}
$$

but $V^TV = I$, orthogonality allows:

$$A_{m\times n}\hat{V}_{n\times n} = \hat{U}_{m\times n}\hat{\Sigma}_{n\times n} \tag{2.25}$$

so,

$$
A_{m\times n}\begin{bmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_n \\ | & | & & | \end{bmatrix}
\begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \end{bmatrix} \tag{2.26}
$$

but this just states n-matrix times vector statements:

$$
\begin{aligned}
Av_1 &= \sigma_1 u_1 \\
Av_2 &= \sigma_2 u_2 \\
&\vdots \\
Av_n &= \sigma_n u_n
\end{aligned} \tag{2.27}
$$

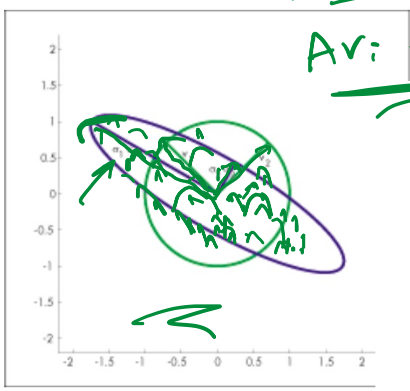
$A = \underline{U \Sigma V^T}$

$A V = U \Sigma$

$A v_i = \sigma_i u_i$

Geometry:.

1. V^* rotates to a standard configuration.
2. Σ stretches each orthogonal axis to the major covariance axis of the corresponding ellipsoid, and
3. U rotates results back to the configuration that associates with A.



$\sigma_1 \sim 1.5$

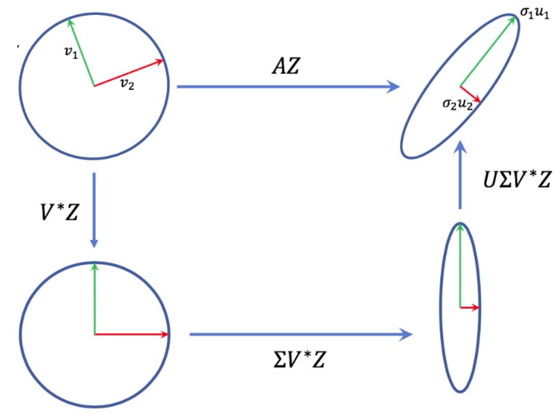
$\sigma_2 \sim 0.5$

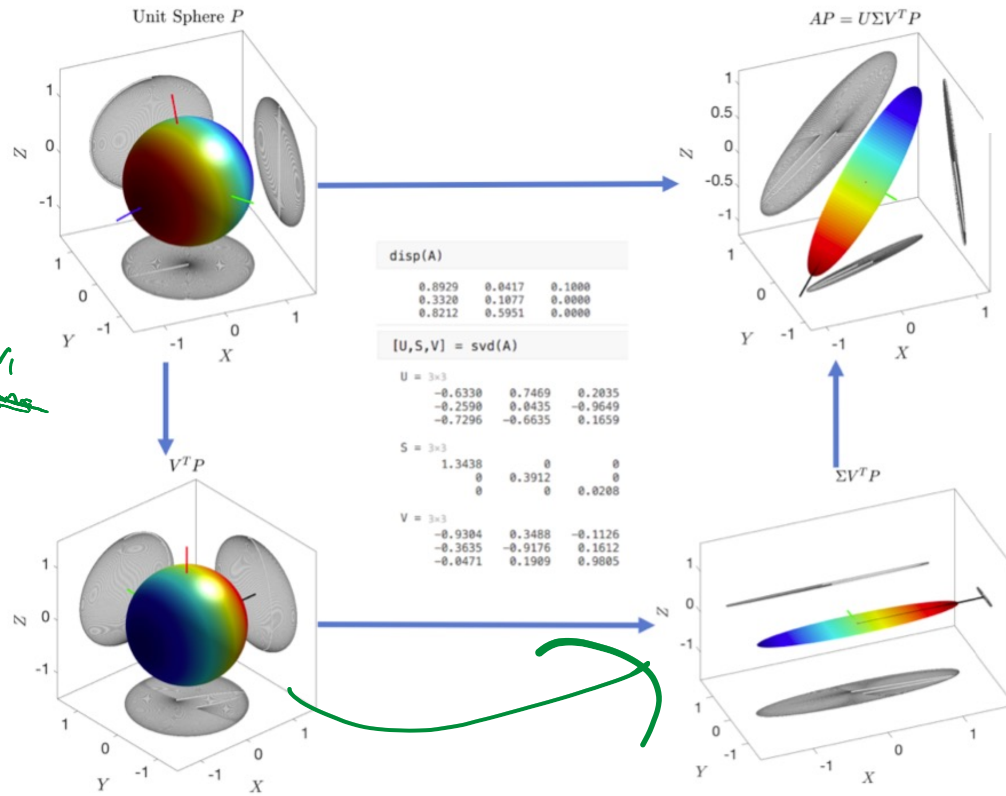
what if

$\sigma_1 = 2$

$\sigma_2 = 0.0003$

$\sigma_1 u_1$, v_1






```
disp(A)

    0.8929    0.0417    0.1000
    0.3320    0.1077    0.0000
    0.8212    0.5951    0.0000

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

U = 3x3
   -0.6330    0.7469    0.2035
   -0.2590    0.0435   -0.9649
   -0.7296   -0.6635    0.1659

S = 3x3
    1.3438         0         0
         0    0.3912         0
         0         0    0.0208

V = 3x3
   -0.9304    0.3488   -0.1126
   -0.3635   -0.9176    0.1612
   -0.0471    0.1909    0.9805
```

A

$A w = U \Sigma V^T w$

$v_i^T w$

$= v_i \cdot w$

1 2 3 4

$X = \left[\begin{array}{l} h \\ 4 \\ f \end{array} \quad \cdots \right]$

f

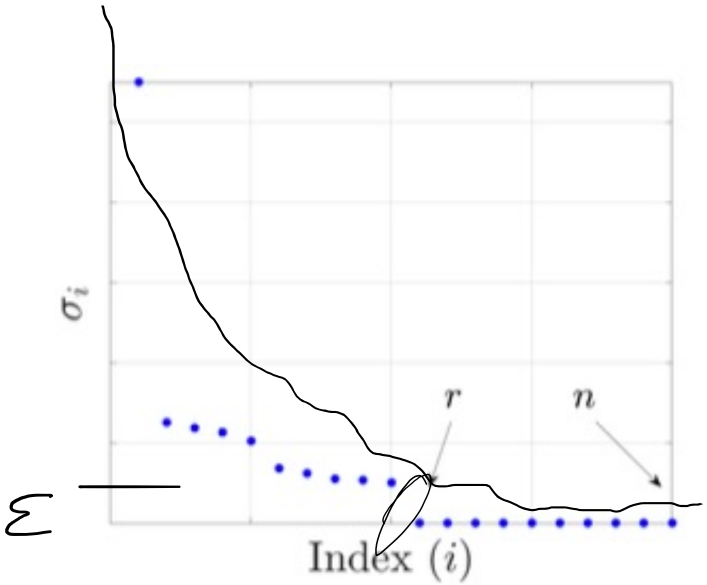
w

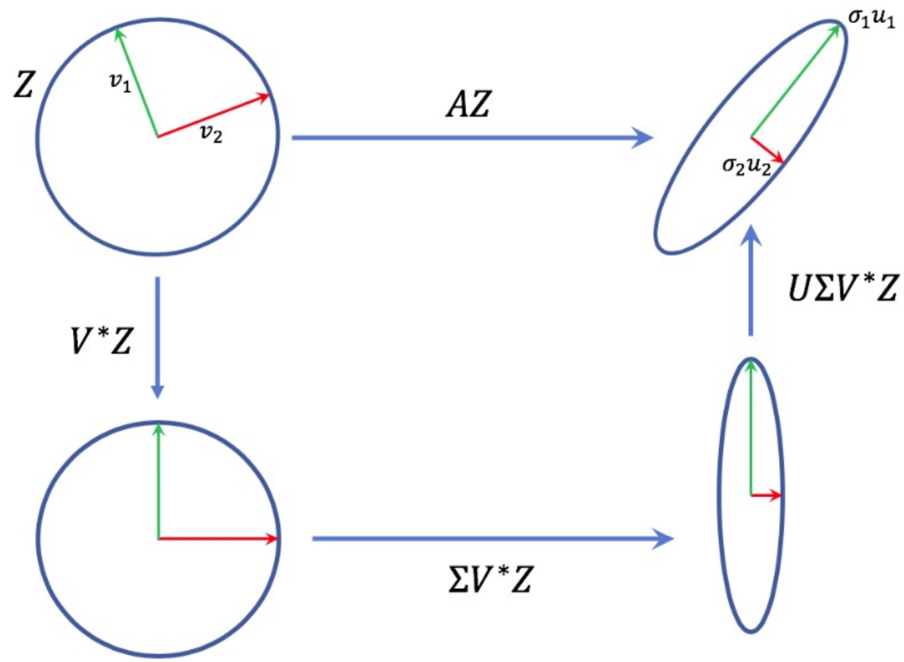
h

redundant

if $\sigma_{r+1} \sim 0$

And ROM





$\sigma_r > \varepsilon > \sigma_{r+1}$

~ 0

Bunny Compression

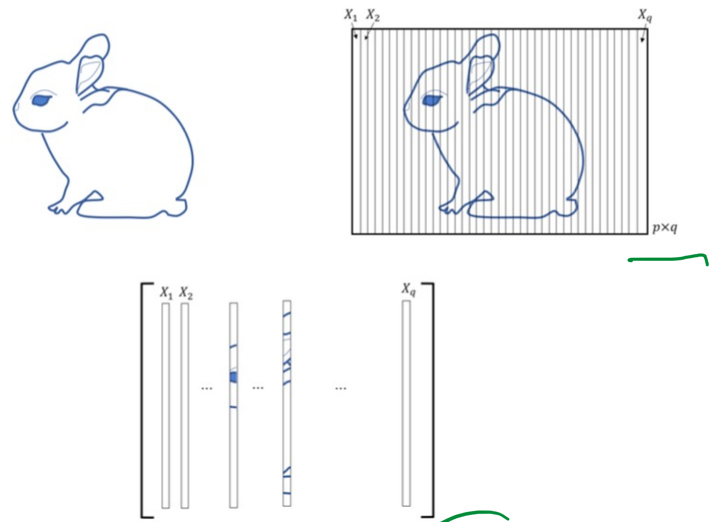


Figure 2.6: Caption

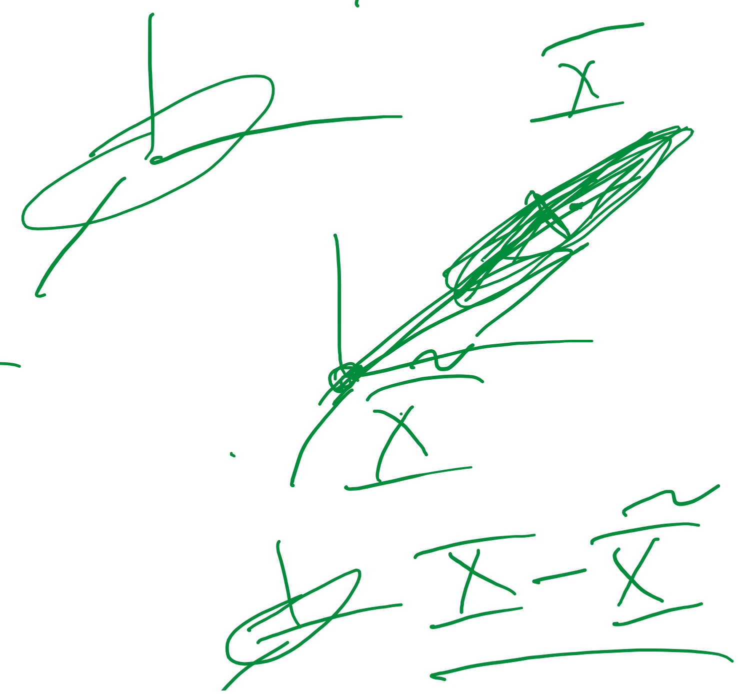
$\underline{X}$

$\underline{X}^T \underline{X}$

Covariance – notice the demean step

$$C_I = \frac{1}{n-1}\left(X - \tilde{X}\right)^T\left(X - \tilde{X}\right)$$

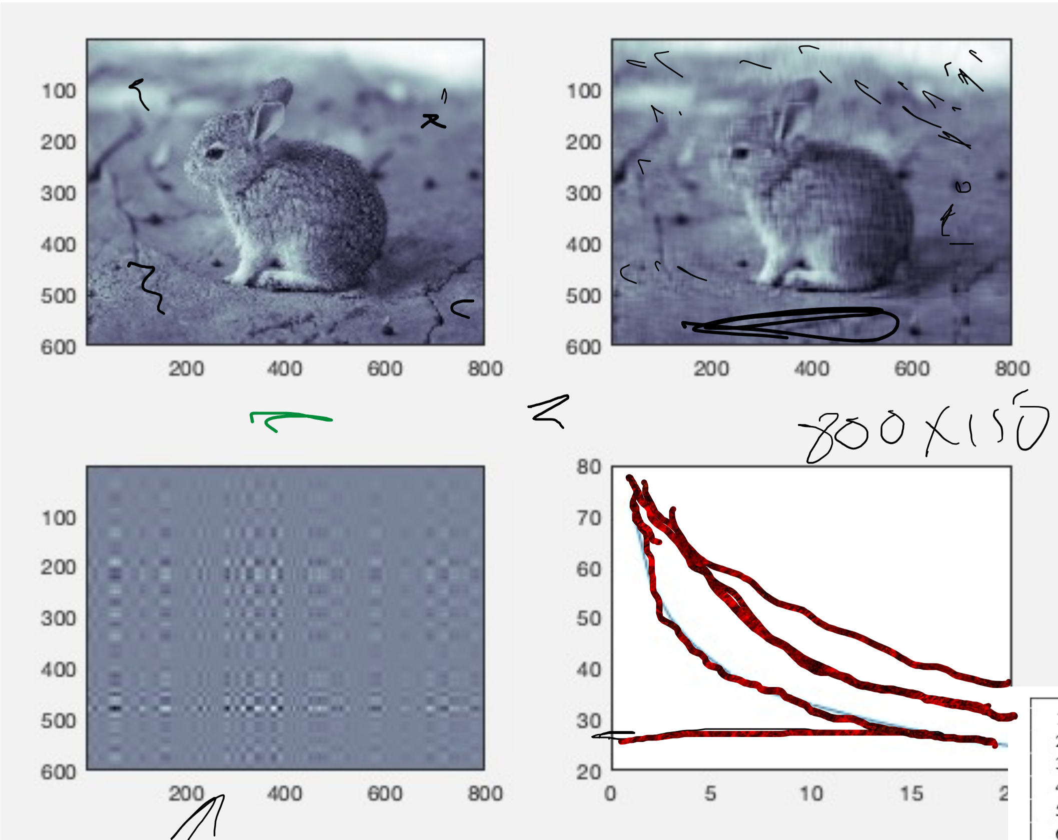
$A^T \quad A$



$\underline{X}$

$\underline{X}$

$\underline{X} - \tilde{X}$



$$u(x,t) = \sum a_i(t)\,\phi_i(x)$$

$$|a_i(t)| \xrightarrow{i \to \infty} 0$$

as fast as possible.

$$I = U \Sigma V^T$$

800 x 150

$$\approx \hat{U} \hat{\Sigma} \hat{V}^T$$

800 x 150

```
I = imread('Bunny.jpg');

figure
subplot(1,2,1)
imshow(I)
xticks({}); yticks({});
pbaspect([1 1 1])
title('RGB Image')

I = rgb2gray(I);  %Convert the 3D RGB color to 1D grayscale
I = im2double(I); %Convert integer value to double (scaled ...
    from 0 to 1)

subplot(1,2,2)
imshow(I)
xticks({}); yticks({});
pbaspect([1 1 1])
title('Grayscale Image')
```

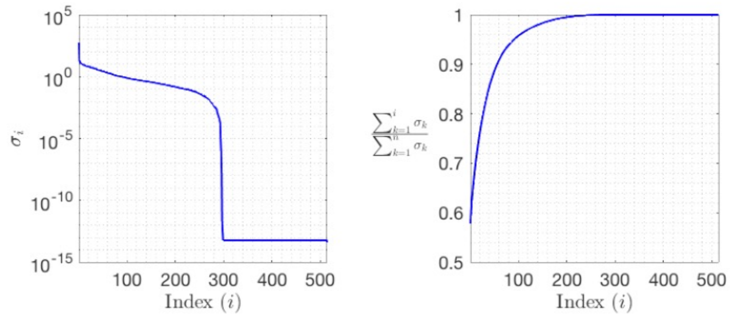
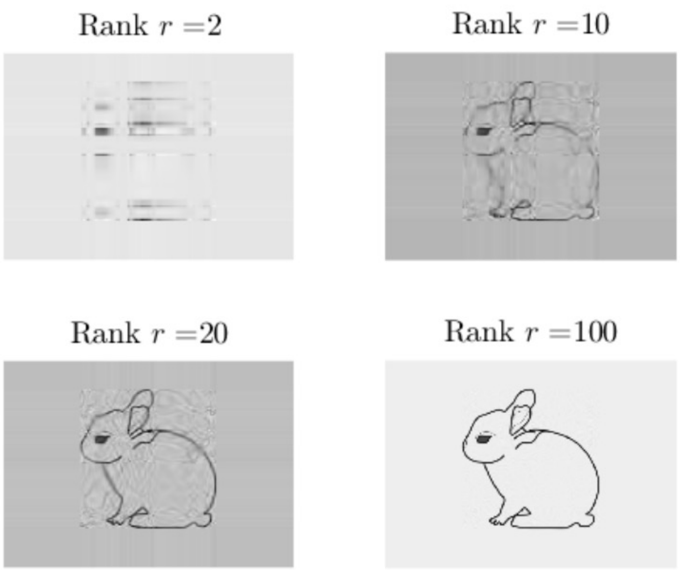
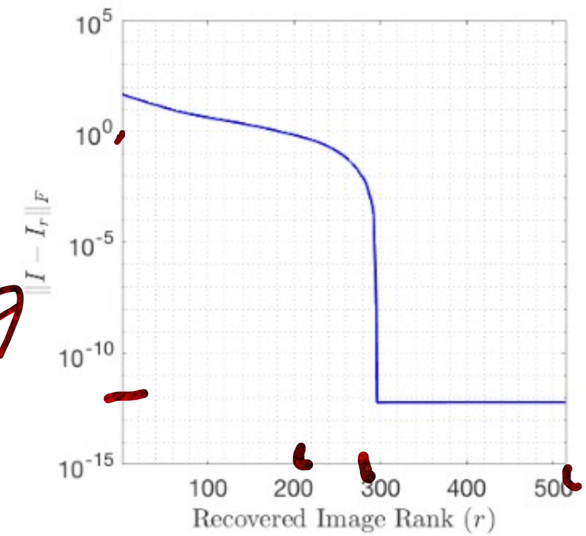


Figure 2.8: (Left) Singular Values. (Right) Energy





: Distance $\|I - I_r\|_F$, where I_r is the recovered image using the reduced

$$e_{rel} = \frac{\|I - I_i\|_F}{\|I\|_F}$$

History

Gene Golub's license plate, photographed by Professor P. M. Kroonenberg of Leiden University.Gene Howard Golub (February 29, 1932 – November 16, 2007), Fletcher Jones Professor of Computer Science at Stanford University. His work made fundamental contributions that have made the singular value decomposition practical as one of the most powerful and widely used tools in modern matrix computation.

Lot's of Machine ~~lea~~ learning
& Data Analysis
is solving an ill-posed
– optimize a cost function.

$$AA^T = U\Sigma \underbrace{V^T (V}\Sigma^T U^T)$$

$$= U\Sigma\Sigma^T U^T$$

$$(AA^T)\,\overline{\underline{U}} = U(\Sigma\Sigma^T) = (\Sigma\Sigma^T)\,\overline{\underline{U}}$$

$$\overline{\underline{U}} = \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_m \\ | & | & & | \end{pmatrix}$$

Definition 2.1.2 — Induced Norm. Suppose a vector norm $\|\cdot\|$ on $\mathcal{K}^m$ is given. Any matrix $A_{m\times n}$ induces a linear operator from $\mathcal{K}^n$ to $\mathcal{K}^m$ with respect to the standard basis, and one defines the corresponding induced norm or operator norm on the space $\mathcal{K}^{m\times n}$ of all $m \times n$ matrices as follows:

$$\|A\|_p = \sup_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p} \tag{2.14}$$

or, taking a vector x such that $\|x\|_p = 1$, then we have

$$\|A\|_p = \sup_{\|x\|_p=1} \|Ax\|_p \tag{2.15}$$

Some Special (Simple) Matrix Norms

The first 3 of these are induced norms, but the 4th is not.

- For $p = 1$:

$$\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}| \tag{2.16}$$

- For $p = \infty$:

$$\|A\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}| \tag{2.17}$$

- A special case is the spectral norm when $p = 2$, in which we have:

$$\|A\|_2 = \sqrt{\lambda_{max}(A^T A)} = \sigma_{max} \tag{2.18}$$

where σ_{max} is the maximum singular value of the matrix A.

- The Frobenius norm is given by:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2} \tag{2.19}$$

Theorem 2.1.2 For a matrix A, the product of the singular values of A, equals the absolute value of its determinant:

$$|det(A)| = \prod_{i=1}^{n} \sigma_i \tag{2.20}$$

$Ax = x'$

$p=1$: $\|(x_1, x_2)\|_1 = |x_1| + |x_2|$

$p=\infty$: $\|(x_1, x_2)\|_\infty = \max_i |x_i|$

$p=2$: $\|(x_1, x_2)\|_2 = \sqrt{x_1^2 + x_2^2}$

$x = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$; $\|x\|_1 = 1+3 = 4$

$\|x\|_\infty = 3$

$\|x\|_2 = \sqrt{1^2+3^2} = \sqrt{10}$

Definition 2.1.2 — Induced Norm. Suppose a vector norm $\|\cdot\|$ on $\mathcal{K}^m$ is given. Any matrix $A_{m\times n}$ induces a linear operator from $\mathcal{K}^n$ to $\mathcal{K}^m$ with respect to the standard basis, and one defines the corresponding induced norm or operator norm on the space $\mathcal{K}^{m\times n}$ of all $m \times n$ matrices as follows:

$$\|A\|_p = \sup_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p} \tag{2.14}$$

or, taking a vector x such that $\|x\|_p = 1$, then we have

$$\|A\|_p = \sup_{\|x\|_p=1} \|Ax\|_p \tag{2.15}$$

Some Special (Simple) Matrix Norms

The first 3 of these are induced norms, but the 4th is not.

- For $p = 1$:

$$\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}| \tag{2.16}$$

- For $p = \infty$:

$$\|A\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}| \tag{2.17}$$

- A special case is the spectral norm when $p = 2$, in which we have:

$$\|A\|_2 = \sqrt{\lambda_{max}(A^T A)} = \sigma_{max} \tag{2.18}$$

where σ_{max} is the maximum singular value of the matrix A.

- The Frobenius norm is given by:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2} \tag{2.19}$$

Theorem 2.1.2 For a matrix A, the product of the singular values of A, equals the absolute value of its determinant:

$$|det(A)| = \prod_{i=1}^{n} \sigma_i \tag{2.20}$$

Fun facts about matrix estimation (data estimation)

If A, $\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = 0$

- $\text{range}(A) = \text{span}(u_1, u_2, \dots, u_r)$
- $\text{null}(A) = \text{span}(v_{r+1}, v_{r+2}, \dots, v_n)$
- $\|A\|_2 = \sigma_1$; $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2}$
- $A = \sum_{i=1}^{r} \sigma_i u_i v_i^* = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \dots + \sigma_r u_r v_r^*$

outer

rank-1 outer products.

$\} \, 0 \xrightarrow{A}$

$\sigma_2 = 0$

$r = 1$

$A = U \Sigma V^T$

$\begin{pmatrix} u_1 & u_2 & \dots & u_m \end{pmatrix} \begin{pmatrix} \sigma_1 & \sigma_2 & \dots & \sigma_r \end{pmatrix} \begin{pmatrix} - v_1^T - \\ - v_2^T - \\ \vdots \end{pmatrix}$ $1 \times n$

$m \times 1$

$w_1^T w_2 = w_1 \cdot w_2$

$= \|w_1\| \|w_2\| \cos\theta$

$w_1 \to w_2$

$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$

Matrix Estimation / Data Estimation. $A_{m\times n}$

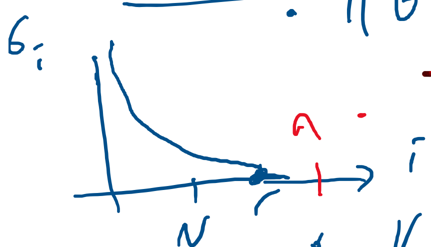
A

- let $0 \le N \le r$ and $A_N = \sum_{i=1}^{N} \sigma_i u_i v_i^*$

(so we may be skipping some of them ... $\sum_{i=N+1}^{r}$

Then. $\|A - A_N\|_2 = \sigma_{N+1}$ (first one skipped)

(what if it zero?)



$e_N = \|A - A_N\|_F = \sqrt{\sigma_{N+1}^2 + \sigma_{N+2}^2 + \dots + \sigma_r^2}$