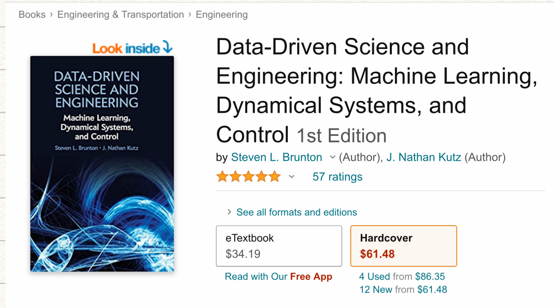
INTRODUCTION

- Welcome back
- Questions / missed points from last time

(1)



(2)

$$\begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix}$$

$$\begin{bmatrix} — & x_1^T & — \\ — & x_2^T & — \\ & \vdots & \\ — & x_n^T & — \end{bmatrix} = X^T = (U\Sigma V^T)^T = V\Sigma^T U^T$$

$$= U'\Sigma' V'^T$$

Opposite meanings

COLOUR KEY
- Helpful
- Important
- Definition
- Theorem
- Example

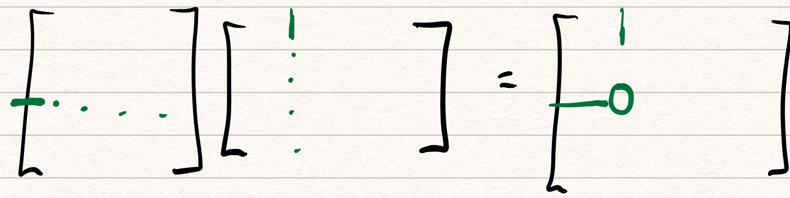
Session Breakdown

1: Theory	2: Practice	3: Application
- Preliminaries	- Eigenvector recap	- Image compression
- Decompositions	- Existence of SVD	- Linear regression
- The SVD	- Comparison of implementations	- Dimensionality reduction
- Low-rank approximations	- The economy SVD	- Wider uses
- Geometric interpretation	- Optimal truncation	- Limitations

RECAP

Rethinking multiplication

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$



$$\begin{bmatrix} — & a_1^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} \begin{bmatrix} | & & | \\ b_1 & \cdots & b_p \\ | & & | \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & \cdots & a_1^T b_p \\ \vdots & \ddots & \vdots \\ a_m^T b_1 & \cdots & a_m^T b_p \end{bmatrix}$$

Note for ADB $[D = \mathrm{diag}(d_1, \dots, d_n)]$, use weighted inner product $\langle u, v \rangle_d = \sum d_i u_i v_i$

$$\begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix} \begin{bmatrix} — & b_1^T & — \\ & \vdots & \\ — & b_n^T & — \end{bmatrix} = \underbrace{a_1 b_1^T}_{\text{rank } 1} + \dots + a_n b_n^T$$

For ADB, D as above, we have $\sum d_i a_i b_i^T$

The SVD

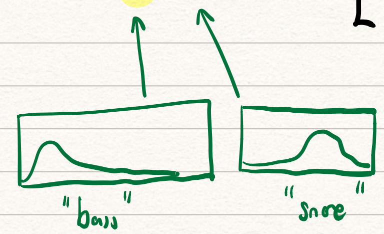
$$\overset{\text{time}\ \longrightarrow}{\begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix}} = \underset{(p \times n)}{X} = \underset{(p\times p)\,(p\times n)\,(n\times n)}{U \Sigma V^T} = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_p \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ \hline & & 0 & \end{bmatrix} \begin{bmatrix} — & v_1^T & — \\ — & v_2^T & — \\ & \vdots & \\ — & v_n^T & — \end{bmatrix}$$

"Time series" for u_1

$p \gg n$

$(0\ 1\ 2\ 1\ 2 \quad)$

Oversimplification warning



Coefficients for mixtures of $u_1, \dots, u_p$ to reconstruct original

- U, V orthogonal
- $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$
- Correction: unique up to sign if σ_i distinct

$$X = \underbrace{\sigma_1 u_1 v_1^T}_{\text{rank } 1} + \sigma_2 u_2 v_2^T + \sigma_3 u_3 v_3^T + \dots \Big|_r + \sigma_n u_n v_n^T \longrightarrow \underset{(p\times n)}{\tilde{X}_r} = \underset{(p\times r)}{\hat{U}_r} \underset{(r\times r)}{\tilde{\Sigma}_r} \underset{(r\times n)}{\tilde{V}_r^T}$$

Why only up to n?

$$U\Sigma = u_1 \Sigma_{1,:} + u_2 \Sigma_{2,:} + \dots + u_n \Sigma_{n,:} + u_{n+1} \underline{0}^T + \dots + u_p \underline{0}^T$$

Are $\tilde{U}_r$, $\hat{V}_r^T$ orthogonal?

No! But they are semi-orthogonal

$$\tilde{U}_r^T \tilde{U}_r = \begin{bmatrix} — & u_1 & — \\ & \vdots & \\ — & u_r & — \end{bmatrix} \begin{bmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{bmatrix} = I$$

$$\tilde{U}_r \hat{U}_r^T = u_1 u_1^T + \dots + u_r u_r^T + \underbrace{\quad\quad}_{\text{missing other terms.}}$$

At least $\tilde{\Sigma}_r$ is square.

Might as well use $\tilde{U}_n \hat{\Sigma}_n \hat{V}_n^T =: \hat{U}\hat{\Sigma}V^T$ (Economy SVD)

Not an approximation: $X = U\Sigma V^T = \hat{U}\hat{\Sigma}V^T$ exactly

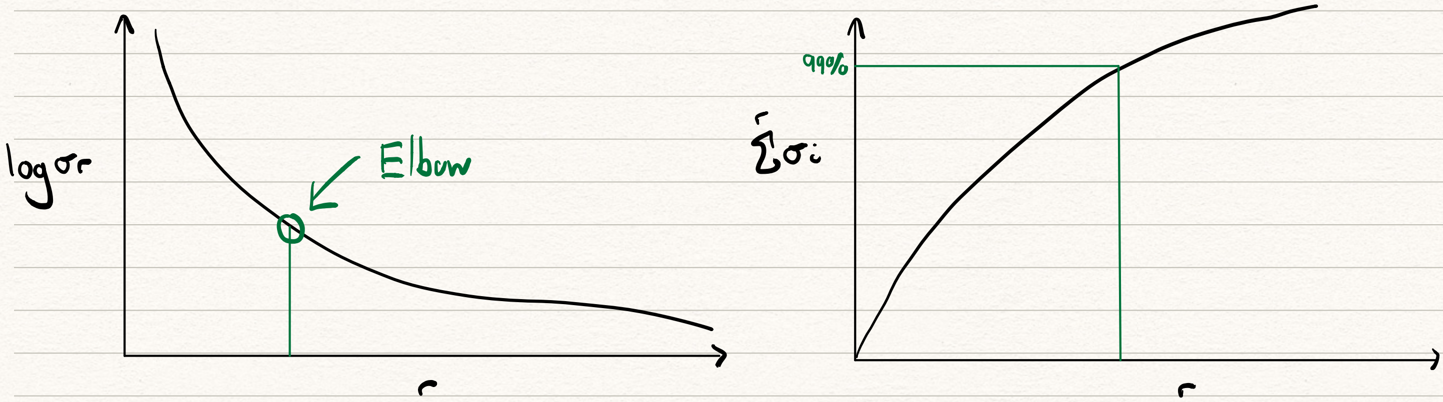
No longer orthogonal. $\rightarrow$

$$\underset{(\cancel{p\times p})\ (p\times n)}{\begin{bmatrix} | & & | & & \cancel{|} \\ u_1 & \cdots & u_n & \cdots & \cancel{u_p} \\ | & & | & & \cancel{|} \end{bmatrix}} \underset{(\cancel{p\times n})\ (n\times n)}{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & \cancel{0} & \end{bmatrix}} \underset{(n\times n)}{\begin{bmatrix} — & v_1^T & — \\ & & \\ — & v_n^T & — \end{bmatrix}}$$

(EK Theorem)

"Best" approximation of X using a rank r matrix $\tilde{X}_r$ is $\tilde{U}_r \tilde{\Sigma}_r \tilde{V}_r^T$

$\uparrow$ Fobenius norm (or spectral norm)



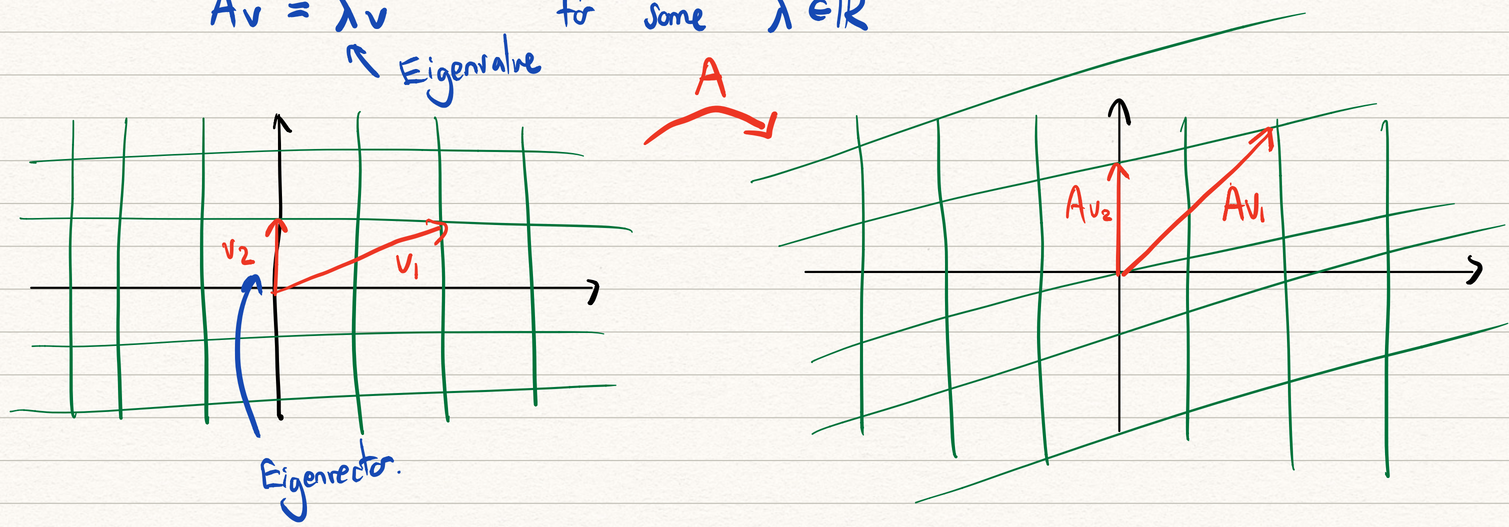
Eigenvalues/eigenvectors

Square matrix A

Eigenvector is a vector v $(v \neq \underline{0})$ s.t.

$Av = \lambda v$ for some $\lambda \in \mathbb{R}$

Eigenvalue



$$Av = \lambda v$$

$$Av - \lambda v = \underline{0}$$

$$Av - \lambda I v = \underline{0}$$

$$(A - \lambda I)v = \underline{0}$$

Not invertible

$$|A - \lambda I| = 0$$

$$\| \quad C_A(x) = 0$$

$v_1, \dots, v_n$

$\lambda_1, \dots, \lambda_n$

$$V = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$$

$$AV = \begin{bmatrix} | & | & & | \\ Av_1 & Av_2 & \cdots & Av_n \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} | & | & & | \\ \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} | & | & & | \\ v_1 & v_2 & & v_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$= V\Lambda$ ← diagonal Λ

COMPUTING THE SVD

Look at $X^T X = \begin{bmatrix} - & x_1^T & - \\ & \vdots & \\ - & x_n^T & - \end{bmatrix} \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} = \begin{bmatrix} x_1^T x_1 & \cdots & x_1^T x_n \\ \vdots & & \vdots \\ x_n^T x_1 & \cdots & x_n^T x_n \end{bmatrix}$ $(n \times n)$

(column-wise) correlation matrix of X $(p \times n)$

"similarity" between two vectors

Symmetric. PSD $\quad v^T X^T X v = (Xv)^T (Xv) = y^T y \geq 0.$

Diagonalisable + Eigenvalues are non-negative.

$$X^T X = (\hat{U}\hat{\Sigma}V^T)^T(\hat{U}\hat{\Sigma}V^T)$$
$$= V\hat{\Sigma}\cancel{\hat{U}^T\hat{U}}\hat{\Sigma}V^T \quad (I)$$
$$= V\hat{\Sigma}^2 V^T$$
$$(X^T X)V = V\hat{\Sigma}^2$$

Eigenvectors of $X^T X$; Eigenvalues

$$\hat{\Sigma}^2 = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}$$

$XX^T \Rightarrow \quad \underset{(p \times p)}{(XX^T)}\hat{U} = \hat{U}\hat{\Sigma}^2$

$X = \hat{U}\hat{\Sigma}V^T \Rightarrow \quad \hat{U} = XV\hat{\Sigma}^{-1}$

Ex. $X = \begin{bmatrix} 0 & 1 & 1 \\ 3 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix}$

$X^T X = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 5/2 & 3/2 \\ 0 & 3/2 & 5/2 \end{bmatrix}$

$\lambda_1 = 9$

$$0 \overset{\text{set}}{=} \begin{vmatrix} 5/2-\lambda & 3/2 \\ 3/2 & 5/2-\lambda \end{vmatrix} = (5/2-\lambda)^2 - \frac{9}{4}$$

$$= \frac{25}{4} - 5\lambda + \lambda^2 - \frac{9}{4} = 4 - 5\lambda + \lambda^2 = (\lambda-1)(\lambda-4)$$

$\lambda_2 = 4$

$\lambda_3 = 1$

$V_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

For $\lambda_2 = 4$

$$\begin{bmatrix} 5 & 0 & 0 \\ 0 & -3/2 & 3/2 \\ 0 & 3/2 & -3/2 \end{bmatrix} \Rightarrow v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \rightarrow \hat{v}_2 = \begin{pmatrix} 0 \\ \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}$$

$$v_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \rightarrow \hat{v}_3 = \begin{pmatrix} 0 \\ \sqrt{2}/2 \\ -\sqrt{2}/2 \end{pmatrix}$$

$$V^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2}/2 & \sqrt{2}/2 \\ 0 & \sqrt{2}/2 & -\sqrt{2}/2 \end{pmatrix} \qquad \hat{\Sigma} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\hat{U} = XV\hat{\Sigma} = \cdots = \begin{pmatrix} 0 & \sqrt{2}/2 & 0 \\ 1 & 0 & 0 \\ 0 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Problems:
- Inefficient

Better solution:
- QR factorisation.

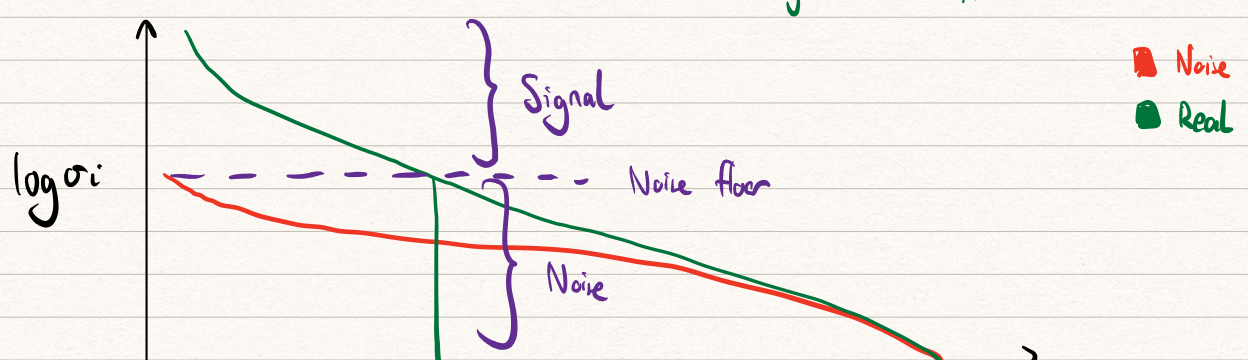
Problem:
What if X is massive.

OPTIMAL TRUNCATION

Gavish – Donoho (2014 IEEE).

$$X = X_{TRUE} + \gamma X_{NOISE}$$

$(X_{NOISE})_{ij} \overset{iid}{\sim} N(0,1)$

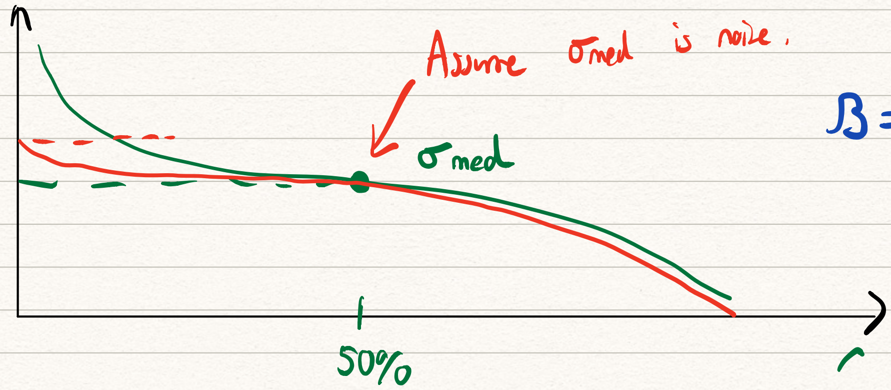


Simple case $\quad X$ square $\quad \gamma$ known.

Noise floor $:= \tau$

$$\tau = \frac{4}{\sqrt{3}} \gamma \sqrt{n}$$

General case $\quad X$ general $\quad \gamma$ unknown.



$\beta = \max\{n/p,\ p/n\}$

"aspect ratio"

$$\tau = \underset{\uparrow}{\omega(\beta)}\, \sigma_{med}$$

Correct factor for aspect ratio.