Lesson 23

Read ch 8

Rank k approximation

matrix norms

Rank k approximation

$A = U \Sigma V^T \qquad A \in \mathbb{R}^{m \times n}$ of rank r

$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T + \sigma_{k+1} u_{k+1} v_{k+1} + \cdots + \sigma_r u_r v_r$

$A_k = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T$ rank k approximation

$k < r$

why rank k ?

If $\sigma_{k+1} \cdots \sigma_r$ are small intuitively

A_k is close to A

If A is $m \times n$ and each entry of A is a number between 0 & 255 (1 byte) it takes $m \cdot n$ bytes to store A and $k(m+n+1)$ to store A_k

Q how big can σ_1 be ?

$\sigma_1 = \sqrt{\lambda_1}$ where λ_1 is the biggest eigenvalue of $A^T A$, $n \times n$ matrix

$\lambda_1 \leq tr(A^T A)$

$A = (a_{ij})$ $A^T = (b_{ij})$ $A^T A = (c_{ij})$

$$Tr\, A^T A = \sum_{i=1}^{n} c_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ji} =$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} a_{ij}$$

so $\sigma_1 \leq \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2} \leq \sqrt{m \cdot n}\ 255$

If $m = 1280$ $n = 720$, common for HD

so $\sigma_1 \leq 960 \cdot 255 \leq 2^{18}$; need ≤ 3 bytes to store

Q: can you find a 3×3 matrix with entries between 0 and 255 and $\sigma_1 = 3 \cdot 255$?

How about $\begin{bmatrix} 255 & 255 & 255 \\ 255 & 255 & 255 \\ 255 & 255 & 255 \end{bmatrix}$?

Examples: 1 Term

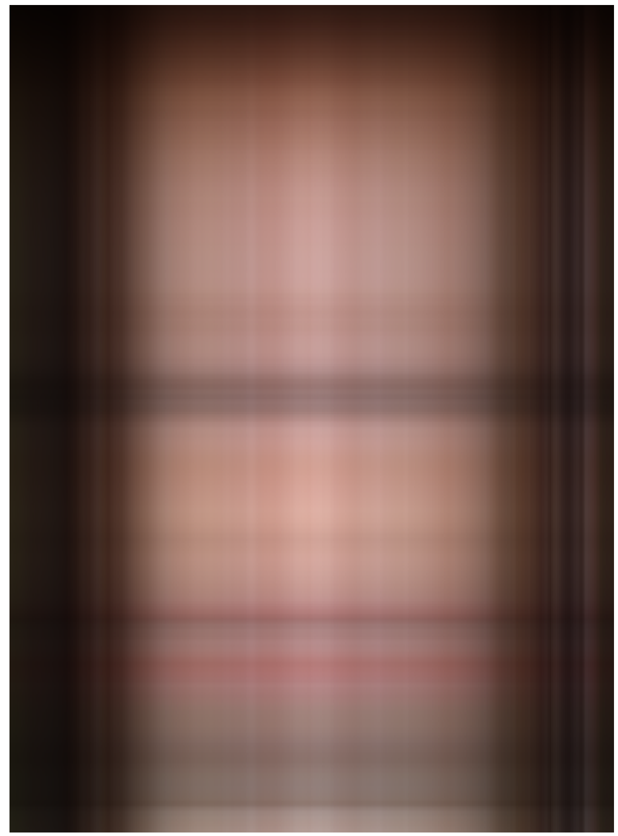
The following is a 500×500 image. The reduced SVD was applied equally to each color:

A

Original

$A_1 = \sigma_1 U_1 V_1^T$

Using 1 terms



Examples: 20 Terms

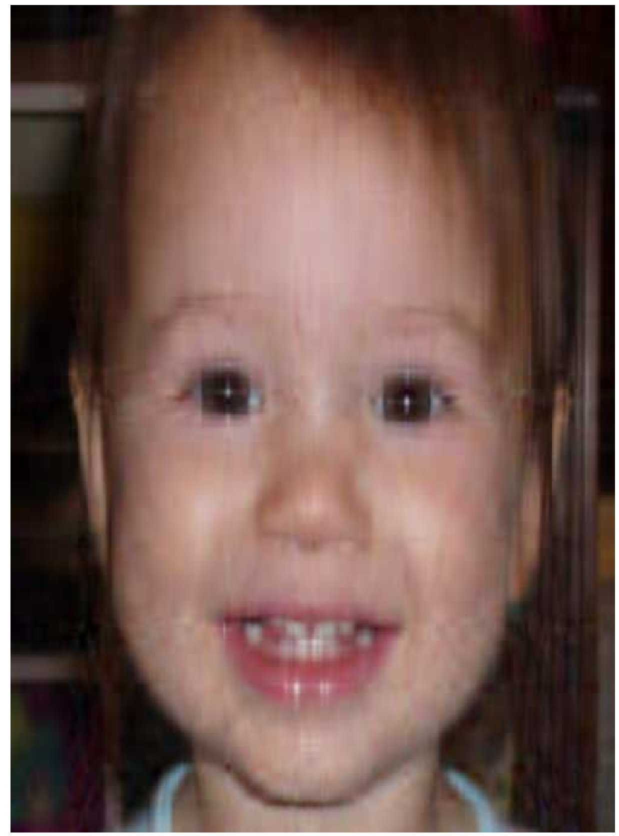
The following is a 500×500 image. The reduced SVD was applied equally to each color:

$A_{20} = \sigma_1 u_1 v_1^T + \cdots + \sigma_{20} u_{20} v_{20}^T$

A

Original

Using 20 terms



Examples: 50 Terms

The following is a 500×500 image. The reduced SVD was applied equally to each color:

Cost $3 \cdot 500 \times 500 = 750{,}000$ bytes

Cost $3 \cdot (2 \cdot 500 \cdot 50 + 50) = 150{,}150$ bytes

$A_{50} = \sigma_1 u_1 v_1^T + \cdots + \sigma_{50} u_{50} v_{50}^T$

Original 750 kB

Using 50 terms $\approx$ 150 kB

Th : $A_k = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_k u_k v_k^T$ has rank k

proof: $\text{Null}(A_k) = \text{span}(v_{k+1} \dots v_n)$
therefore, by rank nullity theorem,
$\text{rank}(A_k) = n - (n-k) = k$

(In hw 7 you showed $\text{rank}(A+B) \leq \text{rank}A + \text{rank}B$)

Proof: $A_k = \underbrace{\begin{bmatrix} u_1 \dots u_k \\ \end{bmatrix}}_{U_k} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix}}_{\Sigma_k} \underbrace{\begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \end{bmatrix}}_{V_k^T}$

$m \times n \qquad k \times k \qquad k \times n$

1) if $j > k$ $\quad A_k v_j = U_k \Sigma_k \begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \end{bmatrix} v_j = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

so $\text{span}(v_{k+1} \dots v_n) \leq \text{Null}(A_k)$

2) Suppose $A_K y = \vec{0}$ $\qquad y = \sum_{l=1}^{n} h_l v_l$

because $v_1 \ldots\ldots v_n$ is a basis for R^n

so $A_K y = A_K \sum_{l=1}^{n} h_l v_l = \sum_{l=1}^{n} h_l A_K v_l = 0$

which means

$$\sum_{l=1}^{n} h_l \, U_K \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_K \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_K^T \end{bmatrix} v_l =$$

$$= \sum_{l=1}^{K} h_l U_K \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_K \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_K^T \end{bmatrix} v_l = \sum_{l=1}^{K} h_l U_K \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_K \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

$$= \sum_{l=1}^{K} h_l U_K \begin{bmatrix} 0 \\ \vdots \\ \sigma_l \\ \vdots \\ 0 \end{bmatrix} = \sum_{l=1}^{K} h_l \sigma_l U_l = \vec{0}$$

but U_l are linearly independent

so $h_1 \sigma_1 = \cdots\cdots = h_K \sigma_K = 0$ $\qquad$ then

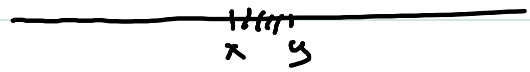
$h_1 = h_2 = \cdots = h_K = 0$

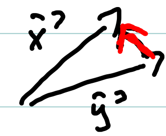
so $y \in \text{span}(v_{K+1} \cdots v_n)$

Two numbers x, y are close if $|x-y|$ is almost 0

Two vectors $\vec{x}$ and $\vec{y}$ are close when $\|\vec{x}-\vec{y}\|$ is close to 0.

When are 2 matrices close?





Def: $\boxed{\text{Norms}}$ are functions $\| \ \|$ $V \longrightarrow R$ satifying

1) $\|v\| \geq 0$ for any $v \in V$
2) $\|v\| = 0 \iff v = 0$
3) $\|kv\| = |k|\|v\|$ $k \in R$
4) $\|v + w\| \leq \|v\| + \|w\|$

$\|v - w\|$ is small intuitively means v and w are close to each other

v could be vectors in R^n, matrices, polynomials in $R[x]_{\leq d}$ ....

Ex $\quad | \ | \quad R \to R$ is a norm
absolute value

$\| \ \|_2 \quad R^n \to R$

$$\left\| \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$ is a norm

Now we want to define a function
$\| \ \|_2 \quad R^{m \times n} \longrightarrow R$ which is
a norm on the space of $m \times n$ matrices.

Spectral norm:

Def : $\quad \| \ \|_2 \quad R^{m \times n} \to R$

For $A \in R^{m \times n}$, $\quad \|A\|_2 = \max\limits_{\substack{v \neq \vec{0} \\ v \in R^n}} \dfrac{\|Av\|}{\|v\|}$

How do I know there is a max ?

Fact 1) $\quad \dfrac{\|Av\|}{\|v\|} = \dfrac{\|A(kv)\|}{\|kv\|}$ since

$$\frac{\|A(kv)\|}{\|kv\|} = \frac{\|k\, A v\|}{\|kv\|} = \frac{|k| \ \|Av\|}{|k| \ \|v\|}$$

Therefore we have

Th: $\max_{\vec{v} \neq 0} \frac{\|A v\|}{\|v\|} = \max_{\|v\|=1} \|A v\|$

closed and bounded
calculus tells me max exists

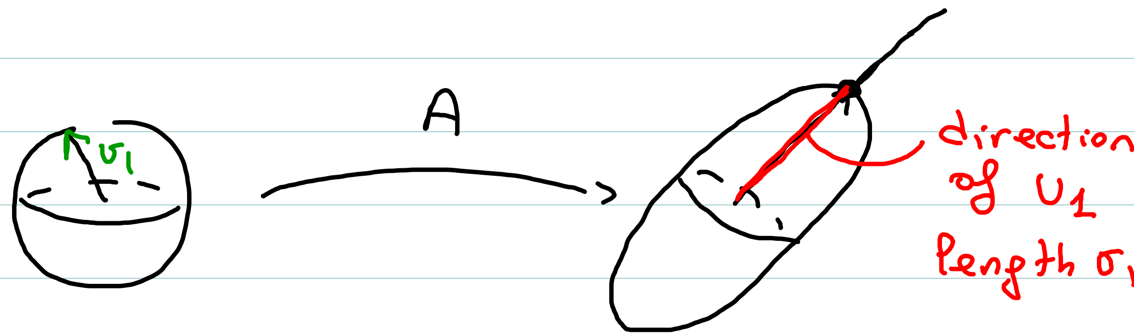
Th: $\|A\|_2$ is a norm on $V = \mathbb{R}^{m \times n}$

no proof.

$\|A\|_2$ is the maximum stretch that A applies to a unit vector

What is this maximum stretch?

Th : Given $A \in \mathbb{R}^{m \times n}$, $\|A\|_2 = \sigma_1$ first singular value



Proof :

$\|A\|_2 = \max\limits_{\|w\|=1} \|Aw\|$ Suppose $w \in \mathbb{R}^n$ $\|w\| = 1$

$\|Aw\| = \|U \Sigma V^T w\|$

① $\|V^T w\| = 1$ because V^T is orthogonal

② Let $V^T w = z \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$, $\Sigma z = \begin{bmatrix} \sigma_1 z_1 \\ \vdots \\ \sigma_r z_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

③ $\|U \Sigma z\| = \|\Sigma z\|$ (because U is orthogonal)

So $\|Aw\| = \left\| \begin{pmatrix} \sigma_1 z_1 \\ \vdots \\ \sigma_r z_r \\ \vdots \\ 0 \end{pmatrix} \right\| = \sqrt{\sigma_1^2 z_1^2 + \dots + \sigma_r^2 z_r^2} \leq$

$\sqrt{\sigma_1^2 z_1^2 + \sigma_1^2 z_2^2 + \dots + \sigma_1^2 z_r^2} \leq \sigma_1 \sqrt{z_1^2 + \dots + z_r^2} \leq \sigma_1$

This tells us $\|A\|_2 \leq \sigma_1$

If $w = v_1$ $\quad \|U \Sigma V^T v_1\| = \|U \Sigma \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\|$

$= \|U \begin{bmatrix} \sigma_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\| = \|\begin{bmatrix} \sigma_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\| = \sigma_1$

so $\|Aw\| \geq \sigma_1$ $\quad$ so $\|A\|_2 \geq \sigma_1$

Since $\|A\|_2 \leq \sigma_1$ and $\|A\|_2 \geq \sigma_1$ we must have $\|A\|_2 = \sigma_1$

Th $\quad \| A - \underbrace{\sigma_1 u_1 v_1^T - \sigma_2 u_2 v_2^T - \cdots \sigma_k u_k v_k^T}_{A_k} \|_2 = \sigma_{k+1}$

Proof : $A - \sigma_1 u_1 v_1^T - \cdots - \sigma_k u_k v_k^T = \sigma_{k+1} u_{k+1} v_{k+1}^T + \cdots + \sigma_r u_r v_r^T$

$$= \underset{m \times r-k}{\begin{bmatrix} u_{k+1} \cdots u_r \end{bmatrix}} \underset{r \times k \times r-k}{\begin{bmatrix} \sigma_{k+1} & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}} \underset{r-k \times n}{\begin{bmatrix} v_{k+1}^T \\ \vdots \\ v_r^T \end{bmatrix}}$$

reduced SVD of $A - A_k$

The norm of a matrix is its first singular value, which is σ_{k+1} for the matrix above.

$$\| A - A_k \|_2 = \sigma_{k+1}$$

This tells us : if σ_{k+1} is small A and A_k are close

Th: (Eckart-Young) $\quad \| A - A_k \|_2 = \min\limits_{B \text{ has rank } k} \| A - B \|_2$

This says A_k (rank k approximation to A) is the matrix of rank k that is closest to A.