Recap: spectral norm of matrix

$$\|A\|_2 = \max_x \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1(A)$$

Proof: Use SVD

Recap: spectral norm of matrix

$$\|A\|_2 = \max_x \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1(A)$$

for any vector x s.t. $\|x\|_2 = 1$

Proof: Use SVD

$$\|Ux\|_2^2 = x^T U^T U x = x^T x = \|x\|_2^2$$

$$
\begin{aligned}
\|Ax\|_2 &= \|U\Sigma V^T x\|_2 \\
&= \|\Sigma V^T x\|_2 \quad \text{by unitary invariance} \\
&= \|\Sigma y\|_2 \quad \text{with } \|y\|_2 = 1 \\
&= \sqrt{\sum_{i=1}^n \sigma_i^2 y_i^2} \\
&\le \sqrt{\sum_{i=1}^n \sigma_1^2 y_i^2} = \sigma_1 \|y\|_2^2 = \sigma_1.
\end{aligned}
$$

$y = V^T x$

$y^T y = x^T x = 1$.

$\sigma_1 \ge \sigma_i \quad \forall i$

$\|Ax\| = \sigma_1$

now take $x = v_1$

$$Ax = U\Sigma V^T v_1 = U\Sigma \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = u_1 \sigma_1$$

Recap: spectral norm of matrix

$$\|A\|_2 = \max_x \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1(A)$$

Proof: Use SVD

$$\begin{aligned}
\|Ax\|_2 &= \|U\Sigma V^T x\|_2 \\
&= \|\Sigma V^T x\|_2 \quad \text{by unitary invariance} \\
&= \|\Sigma y\|_2 \quad \text{with } \|y\|_2 = 1 \\
&= \sqrt{\sum_{i=1}^n \sigma_i^2 y_i^2} \\
&\le \sqrt{\sum_{i=1}^n \sigma_1^2 y_i^2} = \sigma_1 \|y\|_2^2 = \sigma_1.
\end{aligned}$$

matrix ✓

unit. inv. norm

for any

$\|A\| = \|\underbrace{Q}_{\text{orth}} A \underbrace{W}_{\text{orth}}\|$

$Q = U^T$

$W = V$

$= \|\Sigma\|$

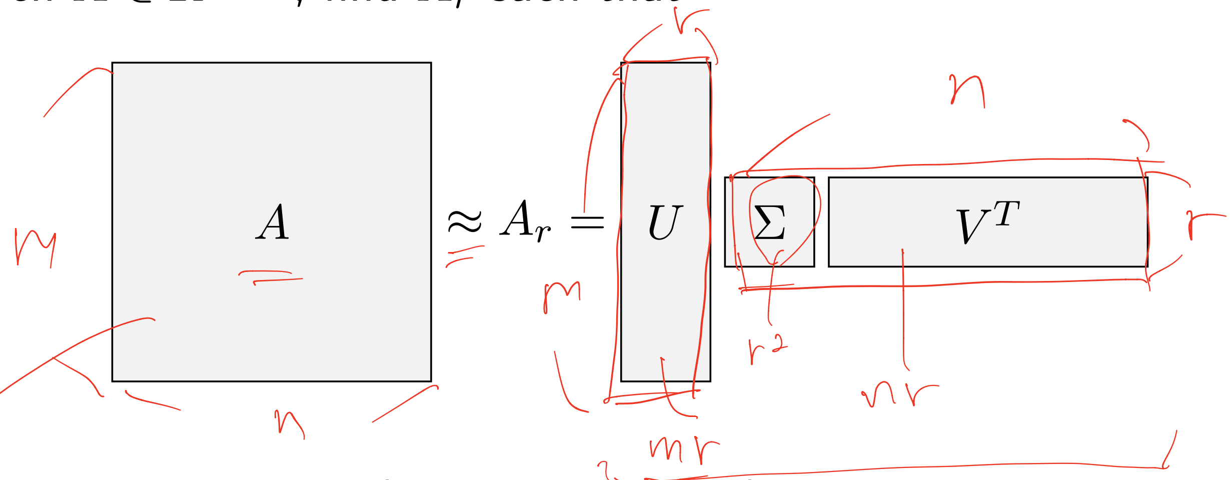
$= \|(\sigma_1, \ldots, \sigma_n)\|$

Frobenius norm: $\|A\|_F = \sqrt{\sum_i \sum_j |A_{ij}|^2} = \sqrt{\sum_{i=1}^n (\sigma_i(A))^2}$

(exercise)

Low-rank approximation of a matrix

Given $A \in \mathbb{R}^{m \times n}$, find A_r such that



- Storage savings (data compression)

mn vs. $(m+n)r$ to store.

10^6 if $r \ll \min(m,n)$ (r: 10-100)

10^{12} 10^7

Optimal low-rank approximation by SVD

Truncated SVD: $A_r = U_r\Sigma_r V_r^T$, $\Sigma_r = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$

$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\mathrm{rank}(B)=r} \|A - B\|_2$$

$C = \left[\ \right]\left[\ \right]_r$, $\mathrm{rank}(C) \le r$.

$= \left\| [u_{r+1} \cdots u_n] \begin{bmatrix} \sigma_{r+1} & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} v_{r+1}^T \\ \vdots \\ v_n^T \end{bmatrix} \right\|$

$B = A_r = U_r \Sigma_r V_r^T$

$A = [u_1 \cdots u_n] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$

$A_r = [u_1 \cdots u_r] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}$

- Good approximation if $\sigma_{r+1} \ll \sigma_1$:

relative error $\dfrac{\sigma_{r+1}}{\sigma_1}$

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

first r sing vecs   sing vals $r \ll n$.

Horn & Johnson,

- Optimality holds for any unitarily invariant norm
- Prominent application: PCA
- Many matrices have explicit or hidden low-rank structure (nonexaminable)

$\|A - A_r\|_F = \min_{\mathrm{rank}(B)=r} \|A - B\|_F$

low-rank!   rank-1!

SVD optimality proof in spectral norm

Truncated SVD: $A_r = U_r \Sigma_r V_r^T$, $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r)$

$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\text{rank}(B)=r} \|A - B\|_2$$

SVD optimality proof in spectral norm

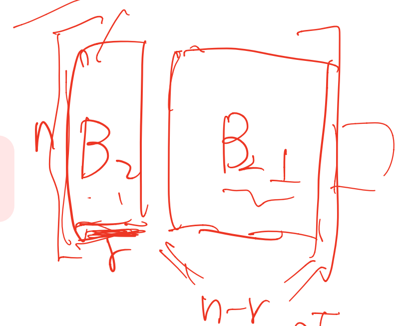
Truncated SVD: $A_r = U_r \Sigma_r V_r^T$, $\Sigma_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$

$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\mathrm{rank}(B)=r} \|A - B\|_2$$

- Since $\mathrm{rank}(B) \le r$, we can write $B = B_1 B_2^T$ where B_1, B_2 have r columns.

SVD optimality proof in spectral norm

$\mathbb{R}^n$: $\dim(\mathrm{span}(B_2)) \le r$



Truncated SVD: $A_r = U_r \Sigma_r V_r^T$, $\Sigma_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$

$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\mathrm{rank}(B)=r} \|A - B\|_2$$

$B_2^T B_{2\perp} = 0$

- Since $\mathrm{rank}(B) \le r$, we can write $B = B_1 B_2^T$ where B_1, B_2 have r columns.

normal

$BW = 0$ if $B_2^T W = 0$.

- There exists orthogonal $W \in \mathbb{C}^{n \times (n-r)}$ s.t. $BW = 0$. Then $\|A - B\|_2 \ge \|(A - B)W\|_2 = \|AW\|_2 = \|U\Sigma(V^* W)\|_2$.

SVD optimality proof in spectral norm

Truncated SVD: $A_r = U_r \Sigma_r V_r^T$, $\Sigma_r = \text{diag}(\sigma_1, \ldots, \sigma_r)$

$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\text{rank}(B)=r} \|A - B\|_2$$

if $m<n$, $\exists x \neq 0$ s.t. $\underbrace{\begin{bmatrix} A \end{bmatrix}}_{n} \begin{bmatrix} x \end{bmatrix} = 0$

- Since $\text{rank}(B) \le r$, we can write $B = B_1 B_2^T$ where B_1, B_2 have r columns.
- There exists orthogonal $W \in \mathbb{C}^{n \times (n-r)}$ s.t. $BW = 0$. Then $\|A - B\|_2 \ge \|(A - B)W\|_2 = \|AW\|_2 = \|U\Sigma(V^* W)\|_2$.
- Now since W is $(n-r)$-dimensional, there is is an interesection between $\boxed{W}$ and $\boxed{[v_1, \ldots, v_{r+1}]}$, the $(r+1)$-dimensional subspace spanned by the leading $r+1$ left singular vectors ($[W, v_1, \ldots, v_{r+1}]\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] = 0$ has a solution; then Wx_1 is such a vector).

$\exists \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \neq 0$ s.t.

$\begin{bmatrix} W & v_1, \ldots v_{r+1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$

$0 \neq \boxed{W x_1} = \boxed{[v_1 \cdots v_{r+1}] x_2}$

$\| \quad \| = \|x_1\|$

SVD optimality proof in spectral norm

Truncated SVD: $A_r = U_r \Sigma_r V_r^T$, $\Sigma_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$

$$\|A - A_r\|_2 = \sigma_{r+1} = \min_{\mathrm{rank}(B)=r} \|A - B\|_2$$

$A_r = U_r \Sigma_r V_r^T$ Then $\|A - A_r\| = \left\| [u_{r+1} \cdots u_n] \begin{bmatrix} \sigma_{r+1} & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} v_{r+1}^T \\ \vdots \\ v_n^T \end{bmatrix} \right\|_2$

$= \left\| \begin{bmatrix} \sigma_{r+1} & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \right\|_2$

$= \sigma_{r+1}$

- Since $\mathrm{rank}(B) \le r$, we can write $B = B_1 B_2^T$ where B_1, B_2 have r columns.
- There exists orthogonal $W \in \mathbb{C}^{n \times (n-r)}$ s.t. $BW = 0$. Then $\|A - B\|_2 \ge \|(A - B)W\|_2 = \|AW\|_2 = \|U\Sigma(V^*W)\|_2$.

  $\forall\, \mathrm{rank}(B) \le r$

- Now since W is $(n-r)$-dimensional, there is is an interesection between W and $[v_1, \ldots, v_{r+1}]$, the $(r+1)$-dimensional subspace spanned by the leading $r+1$ left singular vectors ($[W, v_1, \ldots, v_{r+1}]\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] = 0$ has a solution; then Wx_1 is such a vector).

  $\Rightarrow \|A - B\|_2 \ge \sigma_{r+1}$

  $\|(A-B)x_1\|_2 \ge \sigma_{r+1}$

  $\|x_1\|_2 = 1$

  $Wx_1 = -[v_1 \cdots v_{r+1}][x_2]$

- Then scale x_1 ($\ne 0$) to have unit norm, and $\|x_1\| = \|x_2\| = 1$

  $\|U\Sigma V^* W x_1\|_2 = \|U\Sigma_{r+1} y_1\|_2$, where $\|y_1\|_2 = 1$ and Σ_{r+1} is the leading $r+1$ part of Σ. Then $\|U\Sigma_{r+1} y_1\|_2 \ge \sigma_{r+1}$ can be verified directly.

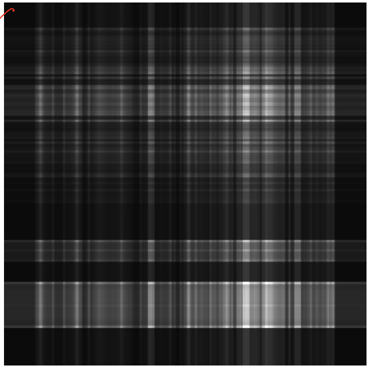
  $[u_1 \cdots u_{r+1}]$

Low-rank approximation: image compression

grayscale image=matrix

σ_i (mat) decay rapidly. e.g.
(image)
$\sigma_1 \geq \sigma_2 \gg \sigma_{100}$

original (mn)



rank 1 m+n

rank 5

rank 10

rank 20

rank 50