# Lecture 10

**Last time**

- Matrix norms
- Singular Value Decomposition
- Spectral Factorization

**Today**

- Best low-rank approximation
- Perturbation theory for eigenvalues
- Distances between subspaces

## Low-rank approximation

In turn, SVD and spectral decompositions are useful to find low-rank approximations of matrices.

**Def:** The best low-rank $k$ approximation of a matrix $A$ is

$$A_{[k]} = \underset{\text{rank}(B)=k}{\operatorname{argmin}} \|A - B\|_{op}$$

We could take $F$ here but it wouldn't make a difference.

Low-rank approximations are routinely used in data science because

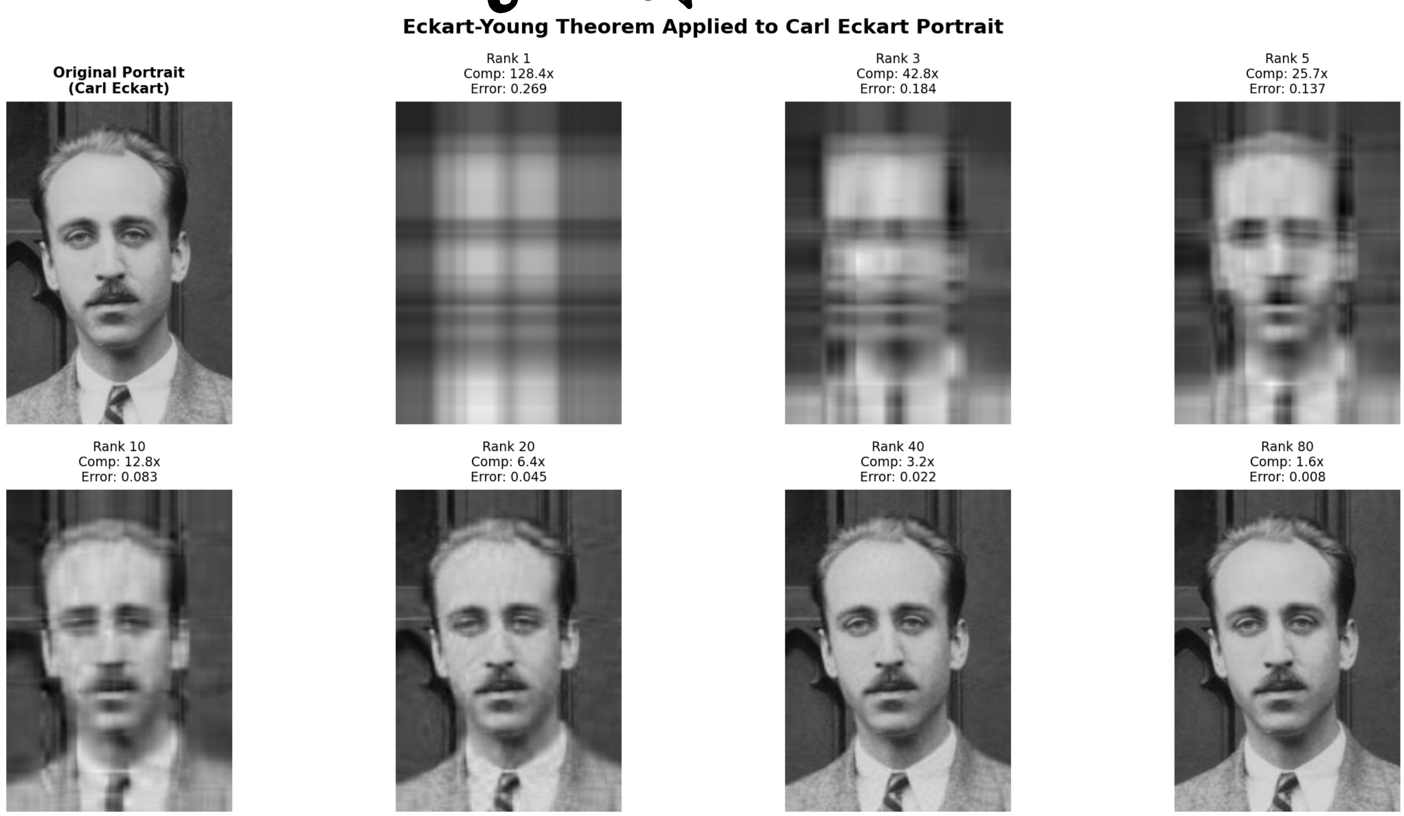
(1) We can store them with $dr$ scalars (better than $d^2$)

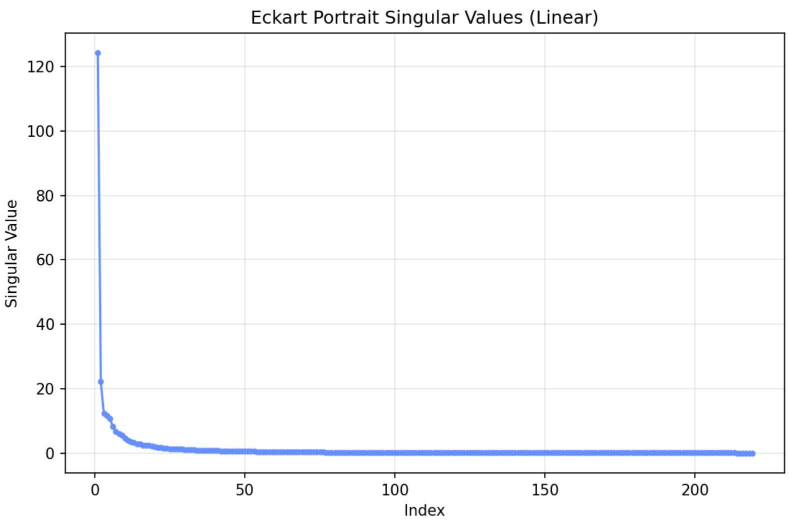
(2) Many computations can be speed up for low-rank matrices (e.g., matrix-vector products require $O(dr)$ floops.)

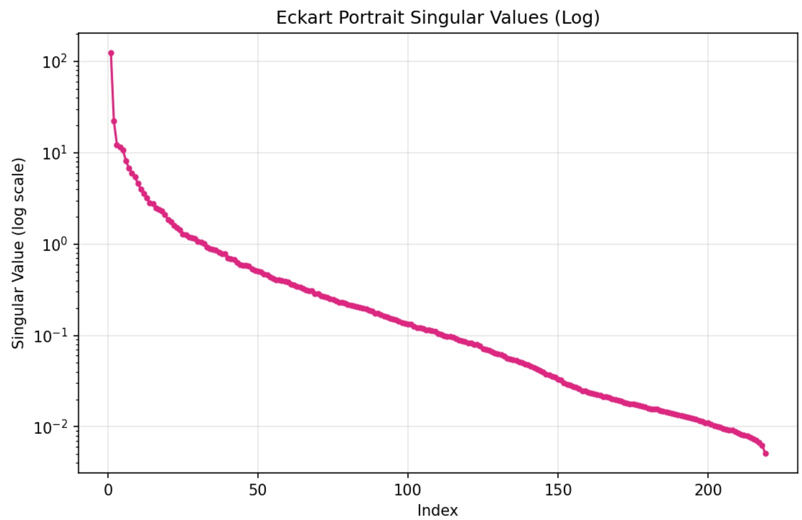
(3) Often low-rank approximations are accurate. (Udell & Townsend, '19).

Potential project

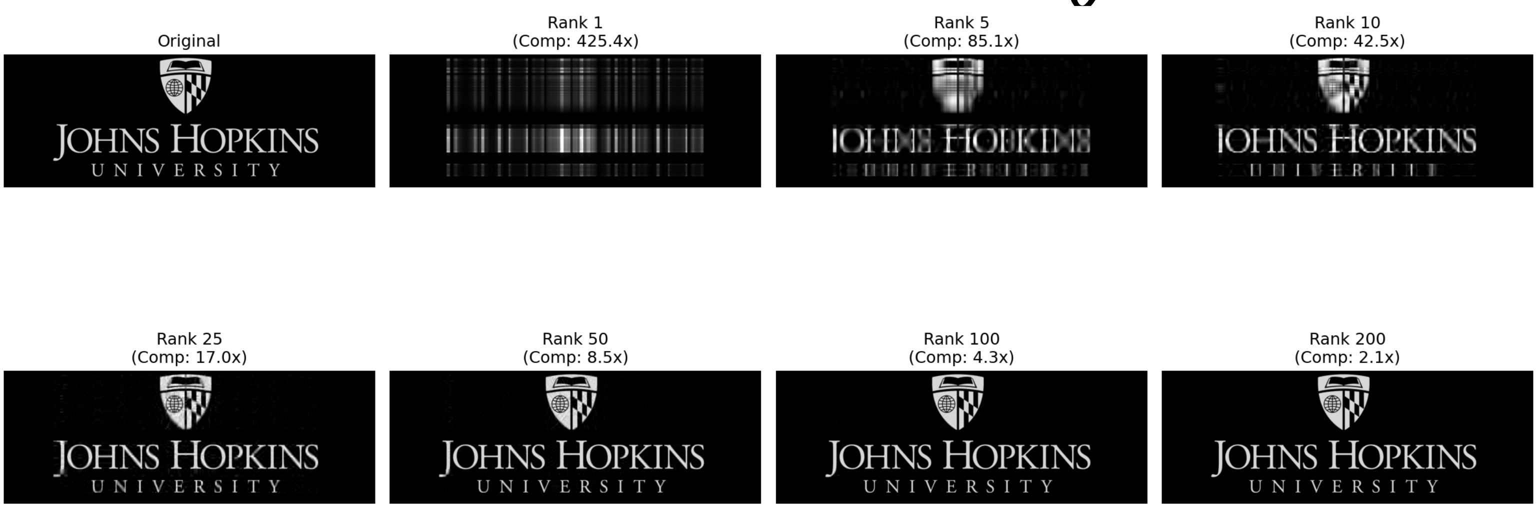
Let's see some practical examples where we take a grey-scale image and treat each pixel as an entry of a matrix:

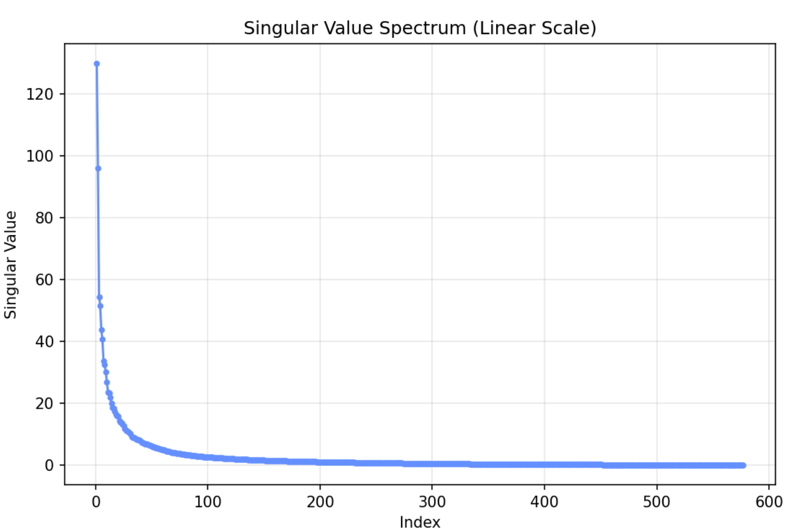


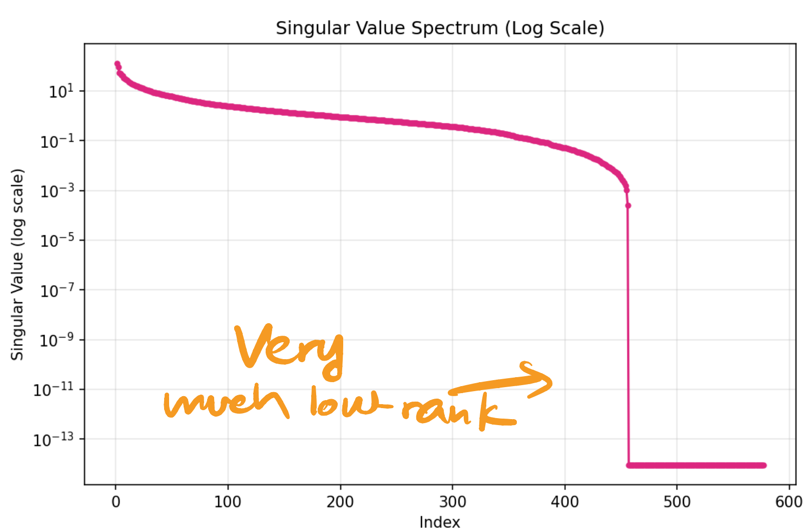



Now with the JHU logo:



A natural question now is how do we obtain ~~a~~ the best rank $k$ approximation?

**Theorem (Eckart-Young):** Let $A \in \mathbb{R}^{n \times m}$ and pick $k \le \operatorname{rank}(A)$. Take

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T .$$

Then, $A_k$ is rank $k$ and

$$\min_{\operatorname{rank}(B)=k} \|A - B\|_{op} = \|A - A_k\|_{op} = \sigma_{k+1} .$$

$\dashv$

**Proof:** We can diagonalize

$$U^T A_k V = \operatorname{diag}(\sigma_1, \dots, \sigma_k, 0, \dots, 0)$$

and so $A_k$ has rank $k$.
Moreover, thank to Fact ⊙ we have

$$\|A - A_k\|_{op} = \|\operatorname{diag}(0, \dots, 0, \sigma_{k+1}, \dots)\|_{op}$$
$$= \sigma_{k+1} .$$

Let $B$ be any rank $k$ matrix. Then, there exists an orthogonal basis

$x_1, \ldots, x_{m-k}$ of $\text{null}(B)$.

Moreover, a dimension counting argument yields

$$\text{span}\{x_1, \ldots, x_{m-k}\} \cap \text{span}\{v_1, \ldots, v_{k+1}\} \neq \emptyset.$$

Let $z$ be a unit norm in this intersection. Then

$$\begin{aligned}
\|A-B\|_{op}^2 &\geq \|(A-B)z\|_2^2 \\
&= \|Az\|_2^2 \\
&= \sum_{i=1}^{k+1} \sigma_i^2 (v_i^T z)^2 \\
&\geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} (v_i^T z)^2 \\
&= \sigma_{k+1}^2 \\
&= \|A - A_k\|_{op}^2.
\end{aligned}$$

$\square$

Exercise: Given a symmetric matrix $A$ and its spectral factorization $U\Lambda U^T$, how can you construct $A_{[k]}$? $\dagger$

It turns out that the same result follows if we substitute $\|\cdot\|_{op}$ by $\|\cdot\|_F$ in the definition of best rank $k$ approximation.

Theorem: Let $A \in \mathbb{R}^{n \times m}$ and pick $k \leq \text{rank}(A)$. Take

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T.$$

Then,

$$\min_{\text{rank}(B)=k} \|A - B\|_F = \|A - A_k\|_{op} = \sigma_{k+1}.$$

$\dagger$

# Fundamentals of perturbation theory

As we saw with our image examples, often we don't have a low-rank matrix exactly but rather an approximate one. Thus, we might expect to have

$$M = \underbrace{A}_{\text{Truly low-rank}} + \overbrace{E}^{\text{Noisy}}.$$

This yields the question of

How far are singular/eigen values and vectors of $M$ from those of $A$?

Today we will cover fundamental results that answer this question. We will not prove the results concerning the "values" as they would require a lot of background, instead we inclu de relevant pointers.

## Perturbation of eigenvalues and singular values

The next two results are proven in Tao (2012) Chapter 1 or in Bhatia (1997) Chapter III.2.

Lemma (Weyl's ineq. for eigenvalues)
Let $A, E \in S^n$. Then, for all $i \in [n]$ we have

$$|\lambda_i(A) - \lambda_i(A+E)| \leq \|E\|_{op}.$$

$\dashv$

Lemma (Weyl's ineq. for singular values)
Let $A, E \in \mathbb{R}^{n \times m}$. Then, for all $i \in [\min\{n,m\}]$ we have that

$$|\sigma_i(A) - \sigma_i(A+E)| \leq \|E\|_{op}.$$

$\dashv$

Thus, eigenvalues and singular values are Lipschitz (and hence stable to small perturbations).

## Interlude: distances and angles between subspaces.

Next we want to measure distances between eigenvectors. Notice that comparing vectors directly is not necessary what we want since we can have entire subspaces associated with a single eigenvalue. Thus we want to measure distances between supspaces generated by eigenvectors.

We will use the following fact.

**Fact ☆ (Stewart & Sun, 1990, Thm. 3.9):** Suppose that $|||\cdot|||$ is a norm that is invariant under rotations. Then, for any $A, B$ we have

$$|||A|||\,\sigma_{min}(B) \leq |||AB||| \leq |||A|||\cdot\|B\|_{op},$$

and

$$\sigma_{min}(A)\,|||B||| \leq |||AB||| \leq \|A\|_{op}\,|||B|||.$$

+

Consider two supspaces

$\mathcal{U} = \text{span}\{U\}$ and $\mathcal{U}^* = \text{span}\{U^*\}$.

$(u_1 \cdots u_r)$ $\qquad$ $(u_1^* \cdots u_r^*)$

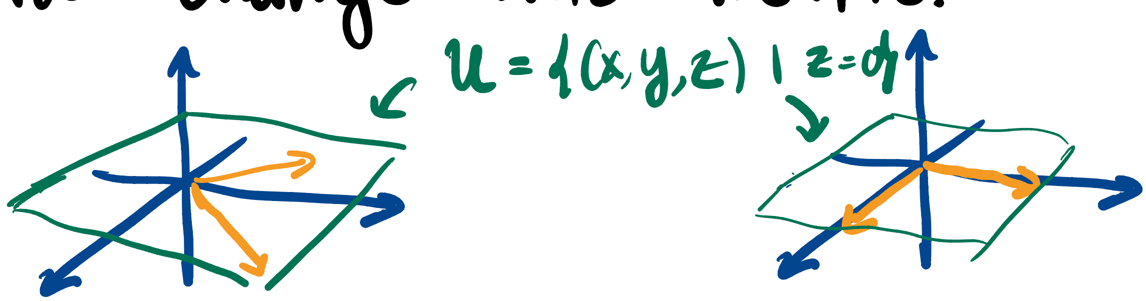
We write $U_\perp$ and $U_\perp^*$ for the $n \times (n-r)$ matrices s.t. $[U, U_\perp] \in O(n)$ and $[U^*, U_\perp^*] \in O(n)$. Thus, $U_\perp$ and $U_\perp^*$ are bases for $\mathcal{U}^\perp$ and $(\mathcal{U}^*)^\perp$, resp.

A very naive idea to measure the distance between $\mathcal{U}$ and $\mathcal{U}^*$ would be to use

some norm

$$|||\, U - U^* \,|||.$$

This is a bad idea because I can find another basis for $\mathcal{U}$ and that would change this metric.



Insight: We need metrics that are invariant under rotations $UR$ with $R \in O(r)$.

Here are some choices

1) Distance with optimal rotation

$$\text{dist}_{|||\cdot|||}(U, U^*) = \min_{R \in O(r)} |||UR - U^*|||.$$

e.g., Frobenius, operator

2) Distance between projections

$$|||UU^T - U^*(U^*)^T|||$$

This matrix projects onto $\mathcal{U}$.

3) Pricipal angles

Let $\sigma_1 \geq \dots \geq \sigma_r \geq 0$ be the singular values of $U^TU^*$. Since

$$\|U^TU^*\|_{op} \leq \|U\|_{op}\|U^*\|_{op} \leq 1.$$

Fact ★

then, $\sigma_i \in [0, 1]$. Define $\theta_i = \arccos \sigma_i$,

$$\Theta = \begin{bmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_r \end{bmatrix}, \quad \sin \Theta = \begin{bmatrix} \sin\theta_1 & & \\ & \ddots & \\ & & \sin\theta_r \end{bmatrix}.$$

and we can measure

$$|||\sin\theta|||.$$

**Exercise:** Show that all of these are the same regardless of the bases we chose.

$\dashv$