Lecture 34: Quick review from previous lecture

Let A be a real, symmetric matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$.

- Let $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \ldots + x_n^2}$. The natural matrix norm of A is

  $$\|A\| = \max\{\|A\mathbf{u}\|_2 : \|\mathbf{u}\|_2 = 1\}.$$

  Then $\|A\| = \max_{1\le i\le n} |\lambda_i|$

- The **Frobenius norm** of a matrix A is defined by

  $$\|A\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2}$$

  Then $\|A\|_F = \sqrt{\lambda_1^2 + \ldots + \lambda_n^2}$.

- Suppose that $|\lambda_1| \ge \cdots \ge |\lambda_n|$. Then

  $$|\lambda_1| = \max\{\|A\mathbf{x}\|_2 : \ \|\mathbf{x}\|_2 = 1\}, \qquad |\lambda_n| = \min\{\|A\mathbf{x}\|_2 : \ \|\mathbf{x}\|_2 = 1\}.$$

  The **maximal value** of $\|A\mathbf{x}\|_2$ is achieved when $\mathbf{x} = \pm\mathbf{u}_1$.

  The **minimal value** of $\|A\mathbf{x}\|_2$ is achieved when $\mathbf{x} = \pm\mathbf{u}_n$.

Today we will discuss Singular Value Decomposition.

- Lecture will be recorded -

- Information for Final Exam and Course Grade has been posted on Canvas, see "Announcements".

§ Optimization principles for eigenvalues of symmetric matrices

Suppose that a symmetric matrix A has real eigenvalues

$$\lambda_1 \geq \cdots \geq \lambda_n$$

and has an orthonormal eigenvector basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Its spectral factorization is

$$A = QDQ^T. = [u_1 \cdots u_n]\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}[u_1 \cdots u_n]^T.$$

Consider the associated quadratic form:

Taking any x in $\mathbb{R}^n$, $x = c_1u_1 + \cdots + c_nu_n$ with $\|x\| = 1 = c_1^2 + \cdots + c_n^2$.

• $c_j = \langle x, u_j \rangle$.

① $$q(\mathbf{x}) = \langle A\mathbf{x}, \mathbf{x} \rangle.$$

$$= \langle A(c_1u_1 + \cdots + c_nu_n),\ c_1u_1 + \cdots + c_nu_n \rangle$$

$$= \langle c_1\lambda_1u_1 + \cdots + c_n\lambda_nu_n,\ \sim \rangle$$

NOTE

$\langle u_i, u_j \rangle = 0,\ i \neq j$

$= 1,\ i = j$

$$= c_1^2\lambda_1 + c_2^2\lambda_2 + \cdots + c_n^2\lambda_n.$$

$$\leq c_1^2\lambda_1 + c_2^2\lambda_1 + \cdots + c_n^2\lambda_1$$

$$= \lambda_1(c_1^2 + \cdots + c_n^2) = \lambda_1$$

$$q(u_1) = \langle Au_1, u_1 \rangle = \lambda_1\langle u_1, u_1 \rangle = \lambda_1$$

Thus, we have the result:

② $$q(x) = \langle Ax, x \rangle \geq c_1^2\lambda_n + \cdots + c_n^2\lambda_n$$

$$= \lambda_n(c_1^2 + \cdots + c_n^2) = \lambda_n$$

$$q(u_n) = \lambda_n.$$

Fact: Suppose that a symmetric matrix A has real eigenvalues

$$\lambda_1 \geq \cdots \geq \lambda_n.$$

Then

$$\lambda_1 = \max\{\langle A\mathbf{x}, \mathbf{x} \rangle : \|\mathbf{x}\|_2 = 1\}, \qquad \lambda_n = \min\{\langle A\mathbf{x}, \mathbf{x} \rangle : \|\mathbf{x}\|_2 = 1\}.$$

The **maximal value** is achieved when $\mathbf{x} = \pm\mathbf{u}_1$, the unit eigenvector associated with the largest eigenvalue λ_1.
The **minimal value** is achieved when $\mathbf{x} = \pm\mathbf{u}_n$, the unit eigenvector associated with the smallest eigenvalue λ_n.

Example. Consider the matrix

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

Find $\max\{\langle A\mathbf{x}, \mathbf{x}\rangle : \|\mathbf{x}\|_2 = 1\} = 4$ and $\min\{\langle A\mathbf{x}, \mathbf{x}\rangle : \|\mathbf{x}\|_2 = 1\} = 2$.

$0 = \det(A - \lambda I) = \lambda^2 - 6\lambda + 8, \quad \lambda = 4, 2.$

§ Similar matrices

Definition: Let $A,\ B$ be two square matrices. We say that A is **similar** to B if there exists an invertible (nonsingular) matrix S such that

$$A = SBS^{-1}.$$

Fact: If A and B are similar, then they have the same eigenvalues.

Suppose that $A\mathbf{v} = \lambda\mathbf{v}$ and λ is the eigenvalue. Then $S^{-1}\mathbf{v}$ is an eigenvector of B with eigenvalue λ.

Similarly, if $\mathbf{w}$ is an eigenvector of B with eigenvalue μ, then $S\mathbf{w}$ is an eigenvector of A with eigenvalue μ.

1. Characteristic polynomial of A

$$\begin{aligned} P_A(\lambda) &= \det(A - \lambda I) \\ &= \det(SBS^{-1} - \lambda I) \\ &= \det(SBS^{-1} - \lambda S S^{-1}) \\ &= \det(S(B - \lambda I)S^{-1}) \\ &= \det S \cdot \det(B - \lambda I) \det(S^{-1}) = \det(B - \lambda I) \\ &= P_B(\lambda). \end{aligned}$$

~~For example~~,

2. $Av = \lambda v$

$SBS^{-1}v = \lambda v. \Rightarrow B(S^{-1}v) = \lambda(S^{-1}v)$ eigenvector of B

Example. Let's consider $A = SBS^{-1} = \begin{pmatrix} 3 & 2 \\ 0 & 2 \end{pmatrix}$, where $S = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$.

It is clear that $\cancel{B}$ A is similar to $\cancel{A}$ B. The matrix B has eigenvalues 3 and 2, with eigenvectors $\mathbf{w}_1 = (1, 1)^T$ and $\mathbf{w}_2 = (2, 1)^T$ (check this!).

Then A also has eigenvalues 3 and 2, but has eigenvectors

$$S\mathbf{w}_1 = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad S\mathbf{w}_2 = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

Remark: In particular, if A is complete (implies A is diagonalizable), then A is similar to a diagonal matrix D:

$$A = VDV^{-1}.$$

8.7 Singular Values

While we've seen that eigenvectors and eigenvalues are powerful tools for understanding matrices and operators, they have limitations.

1. Only square matrices can have eigenvectors.
2. Not every matrix has a basis of eigenvectors (only complete/diagonalizable matrices do).
3. Even when an eigenbasis exists, unless the matrix is symmetric this basis will not be orthogonal.
4. Also, non-symmetric matrices may have complex eigenvalues/eigenvectors.

§ Singular value decomposition (SVD)

We study the factorization of a non-square matrix. The technique is widely used in data analysis.

The key observation is that for *any* real matrix $A = A_{m\times n}$ (not necessarily square), the matrices

$$AA^T, \quad A^TA$$

are both real, symmetric matrices (of sizes m-by-m and n-by-n, respectively).

Let's start by reviewing some facts.

Fact: Let $A \in M_{m\times n}$ ($m \times n$ real matrices). Then the following are true.

1. A^TA and AA^T are symmetric.
2. The kernel of A^TA = the kernel of A.
3. $rank(A) = rank(A^TA)$.

Consequently, the spectral theorem applies to these matrices AA^T, A^TA, even when A is NOT square. (Recall that A^TA is called the *Gram matrix* of A, while AA^T is the *Gram matrix* of A^T.)

$(A = A_{m\times n})$

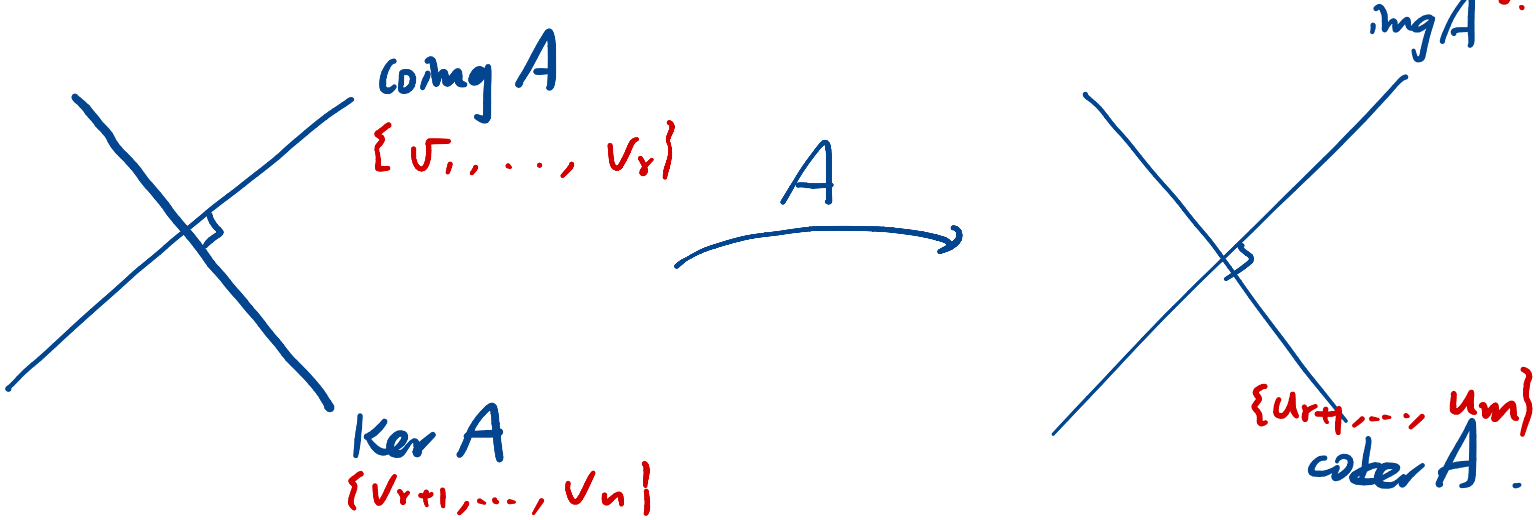
The SVD will be based on a factorization of the $n \times n$ matrix $A^T A$ with rank r:

Since A^TA is symmetric with rank $A^TA = r$, we can find eigenvalues $\underbrace{\lambda_1, \ldots, \lambda_r}_{\text{nonzero}}, \underbrace{0, \ldots, 0}_{r+1, \ldots, n}$ with orthonormal eigenvectors $v_1, \ldots, v_r, v_{r+1}, \ldots, v_n$, that is,

$$\begin{cases} (A^TA)v_j = \lambda_j v_j, & 1 \le j \le r \\ (A^TA)v_k = 0, & k = r+1, \ldots, n. \end{cases}$$

Thus, Fact in Lecture 32, coimg (A^TA) has orthonormal (ONB) basis $\{v_1, \ldots, v_r\}$ and Ker (A^TA) has ONB $\{v_{r+1}, \ldots, v_n\}$.

In addition, A^TA and A have the same Kernel and coimg.



Now, recall $\{v_1, \ldots, v_r\}$ ONB for coimg A, then $\{Av_1, \ldots, Av_r\}$ is a basis for img A

$$\langle Av_i, Av_j \rangle = \langle A^TAv_i, v_j \rangle$$
$$= \langle \lambda_i v_i, v_j \rangle$$

$$= \lambda_i \langle v_i, v_j \rangle$$

$$= \begin{cases} \lambda_i & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

It implies $\|Av_i\| = \sqrt{\lambda_i} \underset{\text{denote}}{=} \boxed{\sigma_i}$, $1 \le i \le r$.

Let $u_i := \dfrac{Av_i}{\sigma_i}$. Then $\{u_1, \dots, u_r\}$ ONB for img A.

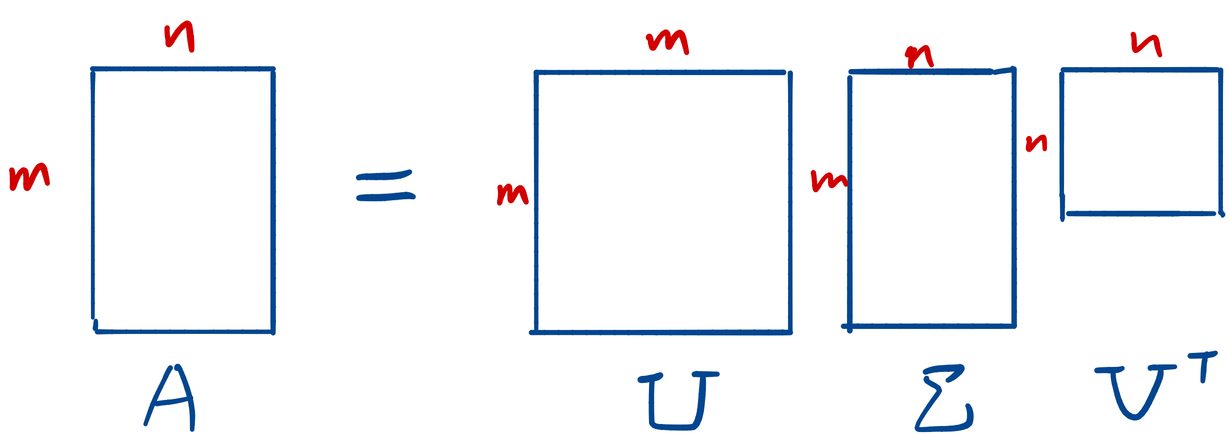
Finally, finding ONB $\{u_{r+1}, \dots, u_m\}$ for coker A.

$$A \underbrace{\left[v_1 \cdots v_r \; v_{r+1} \cdots v_n \right]}_{V} = \underbrace{\left[u_1 \cdots u_r \; u_{r+1} \cdots u_m \right]}_{U} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & 0 \cdots 0 \\ & & 0 & \end{bmatrix}}_{\Sigma}$$

$$\Rightarrow \quad A = U \Sigma V^T,$$

where U : $m \times m$ orthogonal matrix.

V : $n \times n$ — .



Definition: The square roots of the eigenvalues of $A^T A$ are called the **singular values** $\sigma_1, \sigma_2, \cdots, \sigma_n$ of an $m \times n$ matrix A.

Thus, we have shown that

Full SVD for a matrix:
Let A be an $m \times n$ matrix of rank r with the positive singular values

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r,$$

and let Σ be the $m \times n$ matrix defined by

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma_r & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & \cdots & 0 \end{bmatrix}_{m\times n}$$

Then there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

$$A = U\Sigma V^T. \quad \text{(Full SVD).}$$

Example. Find SVD for $A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix}$.

Solution:

To be continued.