Chapter 4. Matrix Decomposition

FPT Education
FPT UNIVERSITY

Chapter 4. Matrix Decomposition

FPT Education
FPT UNIVERSITY

Denote by $\mathbb{R}^{m\times n}$ the set of all matrices of size $m \times n$.

Theorem (Laplace expansion)

Consider a matrix $A \in \mathbb{R}^{n\times n}$. *For all* $i = 1, \ldots, n$:

1. *Expansion along row* i:

$$det(A) = |A| = \sum_{k=1}^{n} a_{ik}(-1)^{i+k} det(A_{ik}) = \sum_{k=1}^{n} a_{ik} c_{ik}(A).$$

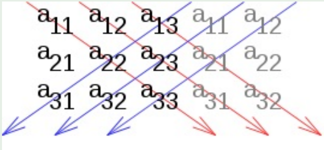
2. *Expansion along column* i:

$$det(A) = |A| = \sum_{k=1}^{n} a_{ki}(-1)^{k+i} det(A_{ki}) = \sum_{k=1}^{n} a_{ki} c_{ki}(A),$$

where $A_{ik} \in \mathbb{R}^{(n-1)\times(n-1)}$ *is the submatrix of* A *that we obtain when deleting row* i *and column* k, *and* $c_{ik}(A) = (-1)^{i+k} det(A_{ik})$ *is the* (i, k)*-cofactor of* A.

Example

1. For $A = [a] \in \mathbb{R}^{1\times 1}$, then $det(A) = a$.
2. For $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{R}^{2\times 2}$, $det(A) = ad - bc$.
3. For $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \in \mathbb{R}^{3\times 3}$, we can compute the determinant of A using Sarrus' rule:



$$det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}.$$

- $det(AB) = det(A)det(B)$.
- $det(A^T) = det(A)$.
- If A is regular (invertible) then $det(A^{-1}) = 1/det(A)$.
- If A and B are similar (there exists a regular matrix C such that $B = C^{-1}AC$) then $det(A) = det(B)$.
- Adding a multiple of a column/row to another one does not change $det(A)$.
- Swapping two rows/columns changes the sign of $det(A)$.
- $det(kA) = k^n det(A)$.

Theorem

Let A be in $\mathbb{R}^{n\times n}$. The followings are equivalent:

1. $det(A) \neq 0$.
2. *A has full rank, i.e.* $rk(A) = n$.
3. *A is regular.*
4. *The homogeneous system* $AX = 0$ *for* $X \in \mathbb{R}^n$ *has only trivial solution.*
5. *The linear system* $AX = b$, *for* $X, b \in \mathbb{R}^n$ *has unique solution.*

Definition

The trace of a matrix $A \in \mathbb{R}^{n \times n}$ is defined as:

$$tr(A) = \sum_{i=1}^{n} a_{ii},$$

the sum of the diagonal elements of A.

Example

$$tr\left(\begin{bmatrix} 2 & 3 & 1 \\ 0 & -3 & 4 \\ 1 & -2 & 4 \end{bmatrix}\right) = 2 - 3 + 4 = 3.$$

- $tr(A+B) = tr(A) + tr(B)$, for $A, B \in \mathbb{R}^{n \times n}$
- $tr(kA) = ktr(A)$, for $k \in \mathbb{R}, A \in \mathbb{R}^{n \times n}$
- $tr(I_n) = n$
- $tr(AB) = tr(BA)$, for $A \in \mathbb{R}^{n \times m}, B \in \mathbb{R}^{m \times n}$.

Definition

The characteristic polynomial of $A \in \mathbb{R}^{n \times n}$

$$p_A(\lambda) := det(A - \lambda I_n) = c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1} + (-1)^n\lambda^n,$$

where $c_0, c_1, \ldots, c_n \in \mathbb{R}$.

We have

$$\begin{aligned} c_0 &= det(A) \\ c_{n-1} &= (-1)^{n-1} tr(A). \end{aligned}$$

Definition

Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. Then $\lambda \in \mathbb{R}$ is an eigenvalue of A and nonzero vector $x \in \mathbb{R}^n$ is the corresponding eigenvector of A if

$$Ax = \lambda x$$

Theorem

The following statements are equivalent:

- *λ is an eigenvalue of A.*
- *There exists an $x \in \mathbb{R}^n - \{0\}$ with $Ax = \lambda x$.*
- *The homogeneous system $(A - \lambda I_n)x = 0$ has non-trivial solution.*
- *$rk(A - \lambda I_n) < n$.*
- *$det(A - \lambda I_n) = 0$.*
- *λ is a root of the characteristic polynomial $p_A(\lambda)$ of A.*

Definition

Let a square matrix A have an eigenvalue λ_i.

- The algebraic multiplicity of λ_i is the number of times the root appears in the characteristic polynomial.
- The eigenspace of A w.r.t λ_i, denoted by E_{λ_i} is the subspace spanned by all eigenvectors of A corresponding to λ_i.
- The geometric multiplicity of λ_i is the dimension of E_{λ_i}.
- The spectrum of A is the set of all eigenvalues of A.

Note. The algebraic multiplicity of an eigenvalue is not smaller than its geometric multiplicity. For example, $A = \begin{bmatrix} -1 & -1 \\ 4 & 3 \end{bmatrix}$ has only eigenvalue $\lambda = 1$ with algebraic multiplicity 2. And the eigenspace $E_1 = span\{\begin{bmatrix} 1 & -2 \end{bmatrix}^T$, hence geometric multiplicity 1.

Example

Find the eigenvalues and eigenvectors of the 2×2 matrix $A = \begin{bmatrix} 2 & 5 \\ 6 & 3 \end{bmatrix}$

- The characteristic polynomial of A

$$p_A(\lambda) = \begin{pmatrix} 2-\lambda & 5 \\ 6 & 3-\lambda \end{pmatrix} = -24 - 5\lambda + \lambda^2 = 0 \Leftrightarrow \lambda = -3 \text{ or } \lambda = 8.$$

  Thus, A has two eigenvalues -3 and 8.

- For $\lambda = -3$: Solve the system

$$\begin{bmatrix} 5 & 5 \\ 6 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = t \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

  Hence the eigenspace corresponding to -3: $E_{-3} = span\{\begin{bmatrix} 1 & -1 \end{bmatrix}^T\}$.

- Similarly for $\lambda = 8$, we have $E_8 = span\{\begin{bmatrix} 5 & 6 \end{bmatrix}^T\}$.

Theorem (Spectral Theorem)

If $A \in \mathbb{R}^{n\times n}$ is symmetric, there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A, and each eigenvalue is real.

Example

Matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ have two eigenvalues $\lambda_1 = 1$ with algebraic multiplicity 2 and $\lambda_2 = 4$ with algebraic multiplicity 1. Moreover

$$E_1 = span\{x_1 = \begin{bmatrix} -1 & 1 & 0 \end{bmatrix}^T, x_2 = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}^T\},$$
$$E_4 = span\{x_3 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T\}$$

Apply Gram-Schmidt algorithm:

$$E_1 = span\{\begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}^T, \begin{bmatrix} -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \end{bmatrix}^T\}$$
$$E_4 = span\{\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}^T\}.$$

Theorem

For $A \in \mathbb{R}^{n \times n}$,

$$det(A) = \prod_{i=1}^{n} \lambda_i$$

$$tr(A) = \sum_{i=1}^{n} \lambda_i$$

where $\lambda_i \in \mathbb{C}$ are (possible repeated) eigenvalues of A.

Example

The matrix $A = \begin{bmatrix} 1 & -3 \\ 1 & -2 \end{bmatrix}$ has two eigenvalues $\lambda_1 = (-1 - \sqrt{3}i)/2$ and $\lambda_2 = (-1 + \sqrt{3}i)/2$. We can see that

$$det(A) = 1 = \lambda_1 \lambda_2$$

$$tr(A) = -1 = \lambda_1 + \lambda_2.$$

FPT Education FPT UNIVERSITY

Definition

A matrix $A \in \mathbb{R}^{n \times n}$ is diagonalizable if it is similar to a diagonal matrix, i.e., if there exists an invertible matrix $P \in \mathbb{R}^{n \times n}$ such that $D = P^{-1}AP$.

Suppose $P = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix}$, $p_i \in \mathbb{R}^n$ and $D = diag(\lambda_1, \lambda_2, \dots, \lambda_n)$:

$$AP = A\begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} = \begin{bmatrix} Ap_1 & Ap_2 & \cdots & Ap_n \end{bmatrix}$$

$$PD = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \dots & \ddots & \dots \\ 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 p_1 & \lambda_2 p_2 & \cdots & \lambda_n p_n \end{bmatrix}$$

Hence, $AP = PD$ implies $Ap_i = \lambda_i p_i$.

Theorem

The followings are equivalent:

1) $A = PDP^{-1}$ *where* $P \in \mathbb{R}^{n \times n}$ *and* D *is a diagonal matrix whose diagonal entries are the eigenvalues of* A.
2) *The eigenvectors of* A *form a basis of* $\mathbb{R}^n$.
3) *The algebraic multiplicity of any eigenvalue of* A *is equal to its geometric multiplicity.*

Theorem

Every symmetric matrix $S \in \mathbb{R}^{n \times n}$ *can always be diagonalized.*

Example

Compute the eigendecomposition of matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

- The characteristic polynomial of A: $det(A - \lambda I_3) = (\lambda - 1)^2(\lambda - 4)$. Hence A has two eigenvalues: $\lambda_1 = 1$ (multiplicity 2) and $\lambda_2 = 4$ (multiplicity 1).
- All eigenvalues corresponding to λ_1 are in form: $\begin{bmatrix} -s-t & t & s \end{bmatrix}^T$. Choose $p_1 = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}^T, p_2 = \begin{bmatrix} -1 & 1 & 0 \end{bmatrix}^T$.
- All eigenvectors corresponding to λ_2 are in form: $\begin{bmatrix} s & s & s \end{bmatrix}^T$. Choose $p_3 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$.

Hence $A = \begin{bmatrix} -1 & -1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} -1 & -1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}^{-1}$.

Note.

1. If $D = diag(\lambda_1, \dots, \lambda_n)$ then $D^k = diag(\lambda_1^k, \dots, \lambda_n^k)$.
2. If $A = PDP^{-1}$ then $A^k = PD^kP^{-1}$.
3. Assume that the eigendecomposition $A = PDP^{-1}$ exists. Then

$$\begin{aligned} det(A) &= det(PDP^{-1}) = det(P)det(D)det(P^{-1}) \\ &= det(D) \\ &= \prod_{i=1}^{n} \lambda_i. \end{aligned}$$

Definition

Let $A \in \mathbb{R}^{m \times n}$ be a rectangular matrix of rank $r \in \{0, \cdots, \min(m, n)\}$. The singular value decomposition (SVD) of A is a decomposition of the form

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

- $U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \in \mathbb{R}^{m \times m}$ is orthonormal matrix. $u_i \in \mathbb{R}^m$ are called the left-singular vectors.
- $V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \in \mathbb{R}^{n \times n}$ is an orthonormal matrix. $v_j \in \mathbb{R}^n$ are called the right-singular vectors.
- Σ is an $m \times n$ matrix with $\Sigma_{ii} = \sigma_i \geq 0$ and $\Sigma_{ij} = 0$ for $i \neq j$. $\sigma_1 \geq \sigma_2 \geq \sigma_r > 0$ are called the singular values.

Construction of the right-singular vectors:

$A^T A \in \mathbb{R}^{n \times n}$ is a symmetric, positive semidefinite matrix.

$$A^T A = PDP^T = P \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} P^T,$$

where P is an orthogonal matrix, $\lambda_i \geq 0$ are eigenvalues of $A^T A$.
Otherwise,

$$\begin{aligned} A^T A &= \left(U \Sigma V^T\right)^T (U \Sigma V^T) = V \Sigma^T U^T U \Sigma V^T = V \Sigma^T \Sigma V^T \\ &= V \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix} V^T \end{aligned}$$

Hence $V = P$ and $\sigma_i^2 = \lambda_i$.

Construction of the left-singular vectors:

$$SDS^T = AA^T = \left(U\Sigma V^T\right)(U\Sigma V^T)^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T$$
$$= U \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix} U^T.$$

Since AA^T and A^TA have the same nonzero eigenvalues the nonzero entries of the Σ matrices in the SVD for both cases have to be the same. We can see

$$(Av_i)^T(Av_j) = v_i^T(A^TA)v_j = v_i(\lambda v_j) = 0, \ \forall\ i \neq j$$

which implies that $\{Av_1, \cdots, Av_r\}$ is an orthogonal basis of an r-dimensional subspace of $\mathbb{R}^m$, for $m \geq r$. Set

$$u_i := Av_i/||Av_i|| = Av_i/\sqrt{\lambda_i} = Av_i/\sigma_i,$$

showing us that the eigenvalues of AA^T are such that $\sigma_i^2 = \lambda_i$.

Example

Find the SVD of $A = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$.

$$A^T A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

$$\Rightarrow V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}, \Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}.$$

$$u_1 = \frac{1}{\sqrt{3}} A v_1 = \frac{1}{\sqrt{3}} A \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 2/\sqrt{6} & 1/\sqrt{6} & -1\sqrt{6} \end{bmatrix}^T$$

$$u_2 = \frac{1}{1} A v_2 = A \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^T \Rightarrow U = \begin{bmatrix} 2/\sqrt{6} & 0 \\ 1/\sqrt{6} & 1/\sqrt{2} \\ -1\sqrt{6} & 1/\sqrt{2} \end{bmatrix}.$$

4.6 Matrix Approximation

Through this section, consider a matrix $A \in \mathbb{R}^{m \times n}$ of rank r, we have the SVD of A:

$$A = U\Sigma V^T.$$

Define rank-1 matrix A_i as:

$$A_i := u_i v_i^T.$$

We obtain that:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T = \sum_{i=1}^{r} \sigma_i A_i.$$

Definition

For $k \leq r$, the rank-k approximation of A is defined as

$$\hat{A}(k) = \sum_{i=1}^{k} \sigma_i u_i v_i^T = \sum_{i=1}^{k} \sigma_i A_i, \text{ with } \operatorname{rk}\left(\hat{A}(k)\right) = k.$$

Error of Approximation:

Definition

The spectral norm of a matrix A is defined

$$\|A\|_2 := \max_{x \in \mathbb{R}^n - \{0\}} \frac{\|Ax\|_2}{\|x\|_2} = \max_{x \in \mathbb{R}^n - \{0\}} \frac{\sqrt{(Ax)^T (Ax)}}{\sqrt{x^T x}}.$$

Theorem

$\|A\|_2 = \sigma_1$, *the largest singular value of A.*

Theorem (Eckart-Young 1936)

$$\hat{A}(k) = \operatorname{argmin}_{rk(B)=k, B \in \mathbb{R}^{m \times n}} \|A - B\|_2 ,$$
$$\left\| A - \hat{A}(k) \right\|_2 = \sigma_{k+1}.$$

Proof of Eckart-Young theorem:

1) By above theorem, $\left\|A - \hat{A}(k)\right\|_2 = \sigma_{k+1}$.

2) Suppose that there is a matrix B with $rk(B) = k$ with

$$\|A - B\|_2 < \left\|A - \hat{A}(k)\right\|_2.$$

Set $Z = null(B)$, then $\dim(Z) = n - k$. Hence for $0 \neq x \in Z$, the Cauchy-Schwartz inequality implies that

$$\|Ax\|_2 = \|(A - B)x\|_2 \leq \|A - B\|_2 \|x\|_2 < \sigma_{k+1} \|x\|_2$$

Set $Z' = span\{v_1, \dots, v_{k+1}\}$. We have Z' is a subspace of $\mathbb{R}^n$ and $\dim Z' = k + 1$. Moreover, for $x \in Z'$:

$$\|Ax\|_2 \geq \sigma_{k+1} \|x\|_2.$$

Hence $Z \cap Z' = 0$ and $\dim(Z) + \dim(Z') = n + 1$ which is a contradiction (see the rank-null theorem).

Example

Find the rank-1 approximation of $A = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$.

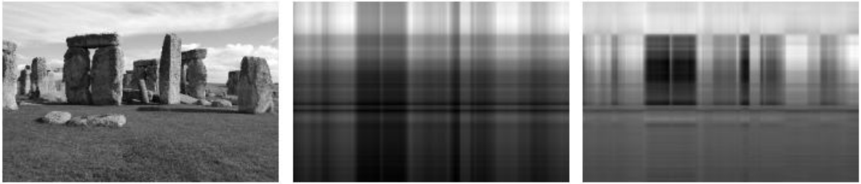
Answer. We have the SVD of A:

$$A = U\Sigma V^T = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

Hence the rank-1 approximation of A is

$$\sigma_1 u_1 v_1^T = \sqrt{6} \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \end{bmatrix} = \begin{bmatrix} 1 & \frac{-2}{5} & \frac{1}{5} \\ -2 & \frac{4}{5} & \frac{2}{5} \end{bmatrix}.$$

Image reconstruction with SVD



(a) Original image $\boldsymbol{A}$. (b) Rank-1 approximation $\widehat{\boldsymbol{A}}(1)$. (c) Rank-2 approximation $\widehat{\boldsymbol{A}}(2)$.

(d) Rank-3 approximation $\widehat{\boldsymbol{A}}(3)$. (e) Rank-4 approximation $\widehat{\boldsymbol{A}}(4)$. (f) Rank-5 approximation $\widehat{\boldsymbol{A}}(5)$.

We have learned

- determinant and trace of a square matrix, and their properties;
- how to find eigenvalues and eigenvectors of a square matrix;
- the diagonalization of a diagonalizable matrix;
- the singular value decomposition;
- k-rank approximation of a matrix.

Exercises for practice: 4.1- 4.12 (page 137, 138).