Matrix Decompositions

Mathematical Models For Data Science

Pratyush Dayal

Indian Institute of Technology Gandhinagar

Why and what?

Analogy

- Factoring of numbers, such as the factoring of 21 into prime numbers $7 \cdot 3$.
- Also called Matrix Factorization.
- Different representation using factors of interpretable matrices. For example:
  - Cholesky decomposition: Square-root-like operation for symmetric, positive definite matrices
  - Singular value decomposition: Extends this factorization to non-square matrices
- Represent very large numerical data in convenient form for analysis.

Determinant I

Definition

- The determinant of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is denoted as $\det(\boldsymbol{A})$ or $|\boldsymbol{A}|$

$$\det(\boldsymbol{A}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

- Determinant of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a function that maps $\boldsymbol{A}$ onto a real number.

Existence of determinant

Testing for Matrix Invertibility

For the smallest cases, we determine if a square matrix $\boldsymbol{A}$ is invertible.

- For a 1×1 matrix $A = [a]$, where $a \neq 0$, the inverse is $A^{-1} = \frac{1}{a}$. Thus, A is invertible if $a \neq 0$.
- For 2×2 matrices, the inverse A^{-1} is given by:

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

  Matrix A is invertible if $a_{11}a_{22} - a_{12}a_{21} \neq 0$, which is the determinant of A, denoted as $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Relationship with inverse

Theorem

For any square matrix $A \in \mathbb{R}^{n\times n}$, A is invertible if and only if $\det(A) \neq 0$.

Expressions for determinants (small matrices)

- For $n = 2$, $\det(\boldsymbol{A}) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$
- For $n = 3$(Sarrus' rule), $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{aligned} =& a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} \\ &- a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} \end{aligned}$

Determinant for special matrices

Triangular matrices

- A square matrix $\boldsymbol{T}$ is an upper-triangular matrix if $T_{ij} = 0 \forall i > j$ (zero below the diagonal).
- A lower-triangular matrix has zeros above its diagonal.
- For a triangular matrix $\boldsymbol{T} \in \mathbb{R}^{n \times n}$, the determinant is the product of its diagonal elements:

$$\det(\boldsymbol{T}) = \prod_{i=1}^{n} T_{ii}.$$

Physical significance

Example

- Consider the three linearly independent vectors $\mathbf{r}, \mathbf{g}, \mathbf{b} \in \mathbb{R}^3$ given as

$$\mathbf{r} = \begin{bmatrix} 2 \\ 0 \\ -8 \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} 6 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 4 \\ -1 \end{bmatrix}$$

- Writing these vectors as the columns of a matrix, $\boldsymbol{A} = [\mathbf{r}, \mathbf{g}, \mathbf{b}] = \begin{bmatrix} 2 & 6 & 1 \\ 0 & 1 & 4 \\ -8 & 0 & -1 \end{bmatrix}$
- The volume is computed as $V = |\det(\boldsymbol{A})| = 186$

Determinant and Trace

- Computing the determinant of an $n \times n$ matrix requires a general algorithm for $n > 3$. Theorem 4.2 reduces this to computing the determinant of $(n-1) \times (n-1)$ matrices. By recursively applying the Laplace expansion, we compute determinants of $n \times n$ matrices by ultimately computing determinants of 2×2 matrices.

Determinant and Trace

Theorem 4.2 (Laplace Expansion)

1. Expansion along column j:

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(A_{k,j}). \tag{4.1}$$

Here, $\det(A_{k,j})$ is called a minor and $(-1)^{k+j}\det(A_{k,j})$ is a cofactor.

2. Expansion along row j:

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(A_{j,k}). \tag{4.2}$$

Here, $A_{k,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ is the submatrix of A obtained by deleting row k and column j.

Determinant and Trace

Example (Laplace Expansion)
Compute the determinant of the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

using Laplace expansion along the first row. Applying (4.2), we have:

$$\det(A) = (-1)^{1+1} \cdot 1 \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+2} \cdot 2 \begin{vmatrix} 3 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+3} \cdot 3 \begin{vmatrix} 3 & 1 \\ 0 & 0 \end{vmatrix}.$$

Compute the determinants of the 2×2 matrices:

$$\det(A) = 1(1-0) - 2(3-0) + 3(0-0) = -5.$$

For completeness, compare this result to computing the determinant using Sarrus' rule:

Determinant and Trace

Properties of Determinant

Properties of Determinants:

1. **Multiplicativity:** $\det(AB) = \det(A)\det(B)$.
2. **Invariance under Transposition:** $\det(A) = \det(A^\top)$.
3. **Invertibility:** If A is invertible, then $\det(A^{-1}) = \frac{1}{\det(A)}$.
4. **Similar Matrices:** Similar matrices have the same determinant. Therefore, for a linear mapping $F : V \to V$, all transformation matrices $[F]$ have the same determinant, making the determinant invariant to the choice of basis.
5. **Column/Row Operations:** Adding a multiple of one column/row to another does not change the determinant. Multiplying a column/row by λ scales the determinant by λ. In particular, swapping two rows/columns changes the sign of the determinant: $\det(A) = -\det(A')$ where A' is A with two rows/columns swapped.

Determinant and Trace

Definition 4.1. The trace of a square matrix $A \in \mathbb{R}^{n \times n}$ is defined as

$$\mathrm{tr}(A) := \sum_{i=1}^{n} a_{ii}, \quad \text{i.e., the trace is the sum of the diagonal elements of } A.$$

The trace satisfies the following properties:

- $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$ for $A, B \in \mathbb{R}^{n \times n}$
- $\mathrm{tr}(\alpha A) = \alpha \mathrm{tr}(A)$ for $\alpha \in \mathbb{R}$ and $A \in \mathbb{R}^{n \times n}$
- $\mathrm{tr}(I_n) = n$, where I_n is the $n \times n$ identity matrix
- $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for $A \in \mathbb{R}^{n \times k}, B \in \mathbb{R}^{k \times n}$

Determinant and Trace

Characteristic Polynomial

For $\lambda \in \mathbb{R}$ and a square matrix $A \in \mathbb{R}^{n \times n}$,

$$p_A(\lambda) := \det(A - \lambda I)$$

is the characteristic polynomial of A, expressed as

$$p_A(\lambda) = c_0 + c_1\lambda + c_2\lambda^2 + \ldots + c_{n-1}\lambda^{n-1} + (-1)^n\lambda^n$$

where $c_0, \ldots, c_{n-1}$ are coefficients in $\mathbb{R}$. Specifically,

$$c_0 = \det(A)$$

$$c_{n-1} = (-1)^{n-1}\mathrm{tr}(A)$$

The characteristic polynomial allows computation of eigenvalues and eigenvectors

Eigenvalues and Eigenvectors

Definition 4.2. Let $A \in \mathbb{R}^{n\times n}$ be a square matrix. Then $\lambda \in \mathbb{R}$ is an eigenvalue of A and $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$ is the corresponding eigenvector of A if

$$A\mathbf{x} = \lambda\mathbf{x}. \tag{4.3}$$

The following statements are equivalent for a square matrix $A \in \mathbb{R}^{n\times n}$:

- λ is an eigenvalue of A.
- There exists a vector $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$ such that $A\mathbf{x} = \lambda\mathbf{x}$.
- $\mathbf{x}$ is a non-zero vector in $\mathbb{R}^n$ satisfying $(A - \lambda I)\mathbf{x} = 0$.
- $A - \lambda I$ is not invertible, i.e., $\det(A - \lambda I) = 0$.
- λ is a root of the characteristic polynomial $p_A(\lambda) = \det(A - \lambda I)$.

Eigenvalues and Eigenvectors

- **Collinearity and Codirection**: Two vectors that point in the same direction are called codirected. Two vectors are collinear if they codirected point in the same or the opposite direction.
- Let $\mathbf{x}$ be an eigenvector of A associated with eigenvalue λ. Then for any $c \in \mathbb{R} \setminus \{0\}$, it holds that $\mathbf{cx}$ is also an eigenvector of A associated with eigenvalue λ. This can be shown as follows:

$$A(\mathbf{cx}) = A(c\mathbf{x}) = cA\mathbf{x} = c\lambda\mathbf{x} = \lambda(c\mathbf{x}). \tag{4.4}$$

Thus, all vectors that are collinear to $\mathbf{x}$ (i.e., vectors of the form $c\mathbf{x}$, where $c \in \mathbb{R} \setminus \{0\}$) are also eigenvectors of A associated with eigenvalue λ.

Eigenvalues and Eigenvectors

Theorem 4.3

λ is an eigenvalue of A if and only if λ is a root of the characteristic polynomial $p_A(\lambda)$ of A.

Definition 4.3 (Eigenspace and Eigenspectrum). For $A \in \mathbb{R}^{n\times n}$, the set of all eigenvectors of A associated with an eigenvalue λ spans a subspace E_λ of $\mathbb{R}^n$, called the eigenspace of A with respect to λ. The eigenspace E_λ is denoted as

$$E_\lambda = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \lambda\mathbf{x}\}. \tag{4.5}$$

The set of all eigenvalues of A is called the eigenspectrum, or simply the spectrum, of A.

Eigenvalues and Eigenvectors

Example: The case of Identity Matrix

The identity matrix $I \in \mathbb{R}^{n\times n}$ has characteristic polynomial $p_I(\lambda) = \det(I - \lambda I) = (1-\lambda)^n = 0$, which has only one eigenvalue $\lambda = 1$ that occurs n times. Moreover, $I\mathbf{x} = \lambda\mathbf{x} = 1\mathbf{x}$ holds for all vectors $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$.
Because of this, the sole eigenspace E_1 of the identity matrix spans n dimensions, and all n standard basis vectors of $\mathbb{R}^n$ are eigenvectors of I.

Useful properties regarding eigenvalues and eigenvectors:

- A matrix A and its transpose $A^\top$ possess the same eigenvalues, but not necessarily the same eigenvectors.
- The eigenspace E_λ is the null space of $A - \lambda I$ since

$$A\mathbf{x} = \lambda\mathbf{x} \quad \Rightarrow \quad A\mathbf{x} - \lambda\mathbf{x} = 0 \quad \Rightarrow \quad (A - \lambda I)\mathbf{x} = 0 \Rightarrow \mathbf{x} \in \ker(A - \lambda I).$$

Eigenvalues and Eigenvectors

Example (Computing Eigenvalues, Eigenvectors, and Eigenspaces). Let us find the eigenvalues and eigenvectors of the 2×2 matrix

$$A = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}.$$

Step 1: Characteristic Polynomial. The characteristic polynomial $p_A(\lambda)$ is computed as

$$p_A(\lambda) = \det(A - \lambda I) = \det \begin{pmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{pmatrix} = (4-\lambda)(3-\lambda) - 2 \cdot 1.$$

Step 2: Eigenvalues. Solving $p_A(\lambda) = 0$, we obtain the eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 5$.

Step 3: Eigenvectors and Eigenspaces.

- For $\lambda = 5$:

$$A - 5I = \begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix}, \quad \text{Eigenspace } E_5 = \text{span}\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}.$$

Eigenvalues and Eigenvectors

- For $\lambda = 2$:

$$A - 2I = \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix}, \quad \text{Eigenspace } E_2 = \text{span}\left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}.$$

The two eigenspaces E5 and E2 in above example are one-dimensional as they are each spanned by a single vector.

Eigenvalues and Eigenvectors

- **Definition 4.4.** The geometric multiplicity of an eigenvalue λ_i of a matrix A is the number of linearly independent eigenvectors associated with λ_i, indicating the dimensionality of the eigenspace.

- Every eigenvalue has at least one associated eigenvector, ensuring its geometric multiplicity is at least one. The geometric multiplicity cannot exceed the eigenvalue's algebraic multiplicity but may be lower.

Example.

The matrix $A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ has two repeated eigenvalues $\lambda_1 = \lambda_2 = 2$ with an algebraic multiplicity of 2. The eigenvalue 2 has, however, only one distinct unit eigenvector $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, thus a geometric multiplicity of 1.

Eigenvalues and Eigenvectors

Theorem 4.4

The eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of a matrix $A \in \mathbb{R}^{n \times n}$ with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ are linearly independent. This implies that the eigenvectors of A with distinct eigenvalues form a basis of $\mathbb{R}^n$.

- **Definition 4.5.** A square matrix $A \in \mathbb{R}^{n \times n}$ is defective if it possesses fewer than n linearly independent eigenvectors.
- A non-defective matrix $A \in \mathbb{R}^{n \times n}$ does not necessarily require n distinct eigenvalues, but it does require that the eigenvectors form a basis of $\mathbb{R}^n$.
- Specifically, a defective matrix has at least one eigenvalue λ_i with an algebraic multiplicity $m > 1$ and a geometric multiplicity of less than m.

Eigenvalues and Eigenvectors

Theorem 4.5

Given a matrix $A \in \mathbb{R}^{m \times n}$, we can always obtain a symmetric, positive semidefinite matrix $S \in \mathbb{R}^{n \times n}$ by defining

$$S := A^\top A. \tag{4.6}$$

Remark. If $\mathrm{rk}(A) = n$, then $S := A^\top A$ is symmetric and positive definite.

- Understanding why Theorem 4.5 holds is insightful: Symmetry requires $S = S^\top$, and using (4.36) we get $S = A^\top A = (A^\top A)^\top = S^\top$. Moreover, positive semidefiniteness requires $x^\top S x \geq 0$, and with (4.6) this translates to $x^\top S x = x^\top A^\top A x = (Ax)^\top (Ax) \geq 0$, as the dot product computes a sum of squares, which are non-negative.

Eigenvalues and Eigenvectors

Theorem 4.6

The determinant of a matrix $A \in \mathbb{R}^{n \times n}$ is the product of its eigenvalues, i.e.,

$$\det(A) = \prod_{i=1}^{n} \lambda_i, \tag{4.7}$$

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

Theorem 4.7

The trace of a matrix $A \in \mathbb{R}^{n \times n}$ is the sum of its eigenvalues, i.e.,

$$\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i, \tag{4.8}$$