# 4. Matrix Decompositions

## 4.1 Determinant and Trace

Determinant: A math. object in the analysis and solution of systems of linear equations

Only defined for square matrices

Det(A) or |A|

A function that maps A onto a real number

**Theorem 4.1.** *For any square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ it holds that $\boldsymbol{A}$ is invertible if and only if* $\det(\boldsymbol{A}) \neq 0$.

We have explicit (closed-form) expressions for determinants of small matrices in terms of the elements of the matrix. For $n = 1$,

$$\det(\boldsymbol{A}) = \det(a_{11}) = a_{11} . \tag{4.5}$$

For $n = 2$,

$$\det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} , \tag{4.6}$$

which we have observed in the preceding example.

For $n = 3$ (known as Sarrus' rule),

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} \\ &\quad - a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} . \end{aligned} \tag{4.7}$$

We call a square matrix $\boldsymbol{T}$ an *upper-triangular matrix* if $T_{ij} = 0$ for $i > j$, i.e., the matrix is zero below its diagonal. Analogously, we define a *lower-triangular matrix* as a matrix with zeros above its diagonal. For a triangular matrix $\boldsymbol{T} \in \mathbb{R}^{n\times n}$, the determinant is the product of the diagonal elements, i.e.,

$$\det(\boldsymbol{T}) = \prod_{i=1}^{n} T_{ii} . \tag{4.8}$$

For n〉3 cases, reduce the determinant of an nxn matrix to (n-1) x (n-1) matrices

By recursively applying the Laplace expansion, ultimately compute determinants of 2x2

For $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ the determinant exhibits the following properties:

- The determinant of a matrix product is the product of the corresponding determinants, $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$.
- Determinants are invariant to transposition, i.e., $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^\top)$.
- If $\boldsymbol{A}$ is regular (invertible), then $\det(\boldsymbol{A}^{-1}) = \frac{1}{\det(\boldsymbol{A})}$.
- Similar matrices (Definition 2.22) possess the same determinant. Therefore, for a linear mapping $\Phi : V \to V$ all transformation matrices $\boldsymbol{A}_\Phi$ of $\Phi$ have the same determinant. Thus, the determinant is invariant to the choice of basis of a linear mapping.
- Adding a multiple of a column/row to another one does not change $\det(\boldsymbol{A})$.
- Multiplication of a column/row with $\lambda \in \mathbb{R}$ scales $\det(\boldsymbol{A})$ by $\lambda$. In particular, $\det(\lambda\boldsymbol{A}) = \lambda^n \det(\boldsymbol{A})$.
- Swapping two rows/columns changes the sign of $\det(\boldsymbol{A})$.

**Theorem 4.3.** *A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ has $\det(\boldsymbol{A}) \neq 0$ if and only if $\operatorname{rk}(\boldsymbol{A}) = n$. In other words, $\boldsymbol{A}$ is invertible if and only if it is full rank.*

## 4.2 Eigenvalues and Eigenvectors

Every linear mapping has a unique transformation matrix given an ordered basis

Can interpret linear mapping and associated transformation matrices by performing an "eigen" analysis

Eigenvalues of a linear mapping tell us how a special set of vectors, the eigenvectors, is transformed by the linear mapping

Useful properties regarding eigenvalues and eigenvectors include the following:

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

$$\begin{aligned} \boldsymbol{Ax} = \lambda\boldsymbol{x} &\iff \boldsymbol{Ax} - \lambda\boldsymbol{x} = \boldsymbol{0} && (4.27a) \\ &\iff (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \iff \boldsymbol{x} \in \ker(\boldsymbol{A} - \lambda\boldsymbol{I}). && (4.27b) \end{aligned}$$

- Similar matrices (see Definition 2.22) possess the same eigenvalues. Therefore, a linear mapping $\Phi$ has eigenvalues that are independent of the choice of basis of its transformation matrix. This makes eigenvalues, together with the determinant and the trace, key characteristic parameters of a linear mapping as they are all invariant under basis change.
- Symmetric, positive definite matrices always have positive, real eigenvalues.

In other cases, may have multiple identical eigenvalues and the eigenspace may have more than one dimension

## 4.3 Cholesky Decomposition

Square-root operation that gives a decomposition of the number into identical components

Cholesky decomposition/Cholesky factorization: provides a square-root equivalent operation on symmetric, positive definite matrices

**Theorem 4.18** (Cholesky Decomposition). *A symmetric, positive definite matrix $\boldsymbol{A}$ can be factorized into a product $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$, where $\boldsymbol{L}$ is a lower-triangular matrix with positive diagonal elements:*

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & \cdots & l_{n1} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & l_{nn} \end{bmatrix}. \tag{4.44}$$

*$\boldsymbol{L}$ is called the Cholesky factor of $\boldsymbol{A}$, and $\boldsymbol{L}$ is unique.*

Cholesky factorization of this covariance matrix allows to generate samples from a Gaussian distribution and also to perform a linear transformation of random variable, compute determinants very efficiently

## 4.4 Eigendecomposition and Diagonalization

How to transform matrices into diagonal form

**Definition 4.19** (Diagonalizable). A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is *diagonalizable* if it is similar to a diagonal matrix, i.e., if there exists an invertible matrix $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ such that $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$.

In the following, we will see that diagonalizing a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is a way of expressing the same linear mapping but in another basis (see Section 2.6.1), which will turn out to be a basis that consists of the eigenvectors of $\boldsymbol{A}$.

Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, let $\lambda_1, \dots, \lambda_n$ be a set of scalars, and let $\boldsymbol{p}_1, \dots, \boldsymbol{p}_n$ be a set of vectors in $\mathbb{R}^n$. We define $\boldsymbol{P} := [\boldsymbol{p}_1, \dots, \boldsymbol{p}_n]$ and let $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ be a diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n$. Then we can show that

$$\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{D} \tag{4.50}$$

if and only if $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{p}_1, \dots, \boldsymbol{p}_n$ are corresponding eigenvectors of $\boldsymbol{A}$.

We can see that this statement holds because

$$\boldsymbol{A}\boldsymbol{P} = \boldsymbol{A}[\boldsymbol{p}_1, \dots, \boldsymbol{p}_n] = [\boldsymbol{A}\boldsymbol{p}_1, \dots, \boldsymbol{A}\boldsymbol{p}_n]\,, \tag{4.51}$$

$$\boldsymbol{P}\boldsymbol{D} = [\boldsymbol{p}_1, \dots, \boldsymbol{p}_n]\begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{bmatrix} = [\lambda_1\boldsymbol{p}_1, \dots, \lambda_n\boldsymbol{p}_n]\,. \tag{4.52}$$

Thus, (4.50) implies that

$$\boldsymbol{A}\boldsymbol{p}_1 = \lambda_1\boldsymbol{p}_1 \tag{4.53}$$

$$\vdots$$

$$\boldsymbol{A}\boldsymbol{p}_n = \lambda_n\boldsymbol{p}_n\,. \tag{4.54}$$

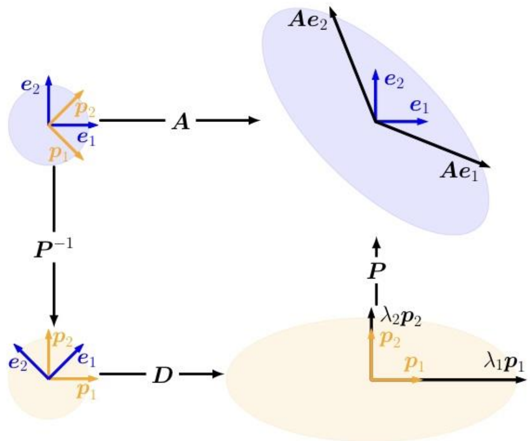
Therefore, the columns of $\boldsymbol{P}$ must be eigenvectors of $\boldsymbol{A}$.

Our definition of diagonalization requires that $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ is invertible, i.e., $\boldsymbol{P}$ has full rank (Theorem 4.3). This requires us to have $n$ linearly independent eigenvectors $\boldsymbol{p}_1, \dots, \boldsymbol{p}_n$, i.e., the $\boldsymbol{p}_i$ form a basis of $\mathbb{R}^n$.

**Theorem 4.20** (Eigendecomposition). *A square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ can be factored into*

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}\,, \tag{4.55}$$

*where $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{D}$ is a diagonal matrix whose diagonal entries are the eigenvalues of $\boldsymbol{A}$, if and only if the eigenvectors of $\boldsymbol{A}$ form a basis of $\mathbb{R}^n$.*




Theorem 4.20 implies that only non-defective matrices can be diagonalized and that the columns of $\boldsymbol{P}$ are the $n$ eigenvectors of $\boldsymbol{A}$. For symmetric matrices we can obtain even stronger outcomes for the eigenvalue decomposition.

**Theorem 4.21.** *A symmetric matrix $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ can always be diagonalized.*

Theorem 4.21 follows directly from the spectral theorem 4.15. Moreover, the spectral theorem states that we can find an ONB of eigenvectors of $\mathbb{R}^n$. This makes $\boldsymbol{P}$ an orthogonal matrix so that $\boldsymbol{D} = \boldsymbol{P}^\top \boldsymbol{A}\boldsymbol{P}$.

*Remark.* The Jordan normal form of a matrix offers a decomposition that works for defective matrices (Lang, 1987) but is beyond the scope of this book. $\diamondsuit$

- Diagonal matrices $\boldsymbol{D}$ can efficiently be raised to a power. Therefore, we can find a matrix power for a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ via the eigenvalue decomposition (if it exists) so that

$$\boldsymbol{A}^k = (\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1})^k = \boldsymbol{P}\boldsymbol{D}^k\boldsymbol{P}^{-1}\,. \tag{4.62}$$

  Computing $\boldsymbol{D}^k$ is efficient because we apply this operation individually to any diagonal element.
- Assume that the eigendecomposition $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ exists. Then,

$$\det(\boldsymbol{A}) = \det(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}) = \det(\boldsymbol{P})\det(\boldsymbol{D})\det(\boldsymbol{P}^{-1}) \tag{4.63a}$$