

# Review of Linear Algebra

*CSE 547 / STAT 548 at the University of Washington*

**Acknowledgment:** This note was originally compiled by Jessica Su for CS 224W at Stanford with substantial modifications by Yikun Zhang in Winter 2023 and Spring 2024 for CSE 547 / STAT 548 at UW. Parts of this note are adapted from Lecture 8 of Professor Yen-Chi Chen's <sup>1</sup> and Professor Michael Perlman's lecture notes (Perlman, 2020) for STAT 512 at UW. Other good references about linear algebra includes Horn and Johnson (2012); Axler (2015) and notes from CS 224W at Stanford:

- [http://snap.stanford.edu/class/cs224w-2014/recitation/linear\\_algebra/LA\\_Slides.pdf](http://snap.stanford.edu/class/cs224w-2014/recitation/linear_algebra/LA_Slides.pdf),
- [http://snap.stanford.edu/class/cs224w-2015/recitation/linear\\_algebra.pdf](http://snap.stanford.edu/class/cs224w-2015/recitation/linear_algebra.pdf).

**Note:** We only discuss the vectors and matrices with real entries in this note, though the stated results also hold for complex entries.

## 1 Vector Space, Span, and Linear Independence

**Vector space:** A *vector space* over the real numbers  $\mathbb{R}$  is a set of vectors that is closed under additions with an identity as the zero vector  $\mathbf{0}$  and additive inverses in the set. It is also closed under scalar multiplications of the vectors by elements in  $\mathbb{R}$ .

The most common vector space in Machine Learning is the Euclidean space  $\mathbb{R}^n$ , which consists of all ordered  $n$ -tuples of real numbers. A vector of  $\mathbb{R}^n$  can be denoted by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

or a row vector  $\mathbf{x}^T = [x_1, \dots, x_n]$ , where  $x_i, i = 1, \dots, n$  are called its *components* or *coordinates*.

### 1.1 Vector Operations

**Dot/Inner product:** The geometric properties of  $\mathbb{R}^n$  are derived from the *Euclidean dot product* defined as:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = x_1 y_1 + \dots + x_n y_n = \sum_{i=1}^n x_i y_i,$$

---

<sup>1</sup>See [http://faculty.washington.edu/yenchic/20A\\_stat512.html](http://faculty.washington.edu/yenchic/20A_stat512.html).

where  $\mathbf{x} = [x_1, \dots, x_n]^T$  and  $\mathbf{y} = [y_1, \dots, y_n]^T$  are in  $\mathbb{R}^n$ .

**Orthogonality:** Two vectors in  $\mathbb{R}^n$  are *orthogonal* if and only if their dot product is zero. In  $\mathbb{R}^2$ , we also call orthogonal vectors perpendicular.

**Norm:** The standard  $\ell_2$ -norm or length of a vector  $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{R}^n$  is given by

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \dots + x_n^2}.$$

Other possible norms in  $\mathbb{R}^n$  include

- $\ell_p$ -norm:  $\|\mathbf{x}\|_p = \left( \sum_{i=1}^n x_i^p \right)^{\frac{1}{p}}$ . It reduces to the above  $\ell_2$ -norm when  $p = 2$ .
- $\ell_\infty$ -norm:  $\|\mathbf{x}\|_\infty = \max_{i=1, \dots, n} |x_i|$ . Notice that  $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_p \leq n^{\frac{1}{p}} \|\mathbf{x}\|_\infty$ .

When the context is clear, we often write the norm of a vector  $\mathbf{x}$  as  $\|\mathbf{x}\|$ . The norms in  $\mathbb{R}^n$  can be used to measure distances between data points (or vectors) in  $\mathbb{R}^n$ .

**Triangle inequality:** For two vectors  $\mathbf{x}, \mathbf{y}$  and any norm  $\|\cdot\|$  in  $\mathbb{R}^n$ , the *triangle inequality* states that

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|,$$

and its reverse version goes as

$$\|\mathbf{x} - \mathbf{y}\| \geq \left| \|\mathbf{x}\| - \|\mathbf{y}\| \right|.$$

## 1.2 Subspaces and Span

**Subspace of  $\mathbb{R}^n$ :** A *subspace* of  $\mathbb{R}^n$  is a subset of  $\mathbb{R}^n$  that is, by itself, a vector space over  $\mathbb{R}$  using the same operations of vector addition and scalar multiplication in  $\mathbb{R}^n$ . In other words, a subset of  $\mathbb{R}^n$  is a subspace precisely when it is closed under these two operations.

**Linear combination:** A *linear combination* of the vectors  $\mathbf{v}_1, \dots, \mathbf{v}_k$  (in  $\mathbb{R}^n$ ) is any expression of the form  $a_1\mathbf{v}_1 + \dots + a_k\mathbf{v}_k$ , where  $k$  is a positive integer and  $a_1, \dots, a_k \in \mathbb{R}$ . Note that some of  $a_1, \dots, a_k$  may be zero.

**Span:** The *span* of a set  $\mathcal{S}$  of vectors consists of all possible linear combinations of finitely many vectors in  $\mathcal{S}$ , *i.e.*,

$$\text{span } \mathcal{S} = \{a_1\mathbf{v}_1 + \dots + a_k\mathbf{v}_k : \mathbf{v}_1, \dots, \mathbf{v}_k \in \mathcal{S}, a_1, \dots, a_k \in \mathbb{R}, \text{ and } k = 1, 2, \dots\}.$$

## 1.3 Linear Independence

The vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$  (in  $\mathbb{R}^n$ ) are *linearly dependent* if and only if there exist  $a_1, \dots, a_n \in \mathbb{R}$ , **not all zero**, such that  $a_1\mathbf{v}_1 + \dots + a_n\mathbf{v}_n = \mathbf{0}$ .

A finite set of vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$  (in  $\mathbb{R}^n$ ) is *linearly independent* if it is not linearly dependent. In other words, we cannot write any vector in  $\mathbf{v}_1, \dots, \mathbf{v}_n$  in terms of a linear combination of the other vectors.

## 2 Matrices

A  $m \times n$  matrix  $A \in \mathbb{R}^{m \times n}$  is an array of  $mn$  numbers as

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}.$$

It represents the *linear mapping* (or *linear transformation*) from  $\mathbb{R}^n$  to  $\mathbb{R}^m$  as

$$\mathbf{x} \mapsto A\mathbf{x} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n A_{1i}x_i \\ \sum_{i=1}^n A_{2i}x_i \\ \vdots \\ \sum_{i=1}^n A_{mi}x_i \end{bmatrix} \quad \text{for any } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n.$$

Here, the linearity means that  $A(a\mathbf{x} + b\mathbf{y}) = aA\mathbf{x} + bA\mathbf{y}$  for any  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  and  $a, b \in \mathbb{R}$ . In particular, when  $m = n$ ,  $A \in \mathbb{R}^{n \times n}$  is called a square matrix.

### 2.1 Matrix Operations

**Matrix addition:** If  $A, B$  are both  $m \times n$  matrices, then the matrix addition is defined as elementwise additions as:

$$[A + B]_{ij} = A_{ij} + B_{ij}.$$

**Example 1.** Here is an example of a matrix addition for two matrices in  $\mathbb{R}^{2 \times 2}$  as

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}.$$

**Matrix multiplication:** For two matrices  $A \in \mathbb{R}^{m \times n}$ ,  $B \in \mathbb{R}^{n \times p}$ , the product  $AB$  is a  $m \times p$  matrix, whose  $(i, j)$ -entry is

$$[AB]_{ij} = \sum_{k=1}^n A_{ik}B_{kj}$$

for all  $1 \leq i \leq m$  and  $1 \leq j \leq p$ .

**Example 2.** Here is an example of the matrix multiplication for two square matrices in  $\mathbb{R}^{2 \times 2}$  as

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 \times 5 + 2 \times 7 & 1 \times 6 + 2 \times 8 \\ 3 \times 5 + 4 \times 7 & 3 \times 6 + 4 \times 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}.$$

We can also multiply non-square matrices when their dimensions are matched (*i.e.*, the number of columns of the first matrix should be equal to the number of rows of the second

matrix) as

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 2 \cdot 4 & 1 \cdot 2 + 2 \cdot 5 & 1 \cdot 3 + 2 \cdot 6 \\ 3 \cdot 1 + 4 \cdot 4 & 3 \cdot 2 + 4 \cdot 5 & 3 \cdot 3 + 4 \cdot 6 \\ 5 \cdot 1 + 6 \cdot 4 & 5 \cdot 2 + 6 \cdot 5 & 5 \cdot 3 + 6 \cdot 6 \end{bmatrix} = \begin{bmatrix} 9 & 12 & 15 \\ 19 & 26 & 33 \\ 29 & 40 & 51 \end{bmatrix}.$$

### Properties of matrix multiplications:

- *Associativity*:  $(AB)C = A(BC)$ .
- *Distributivity*:  $A(B + C) = AB + AC$ .
- However, matrix multiplication is in general **not** commutative. That is,  $AB$  is not necessarily equal to  $BA$ .
- The matrix multiplication between a 1-by- $n$  matrix and an  $n$ -by-1 matrix is the same as taking the dot product of the corresponding vectors.

**Matrix transpose:** If  $A = [A_{ij}] \in \mathbb{R}^{m \times n}$ , then its *transpose*  $A^T$  is a  $n \times m$  matrix, whose  $(i, j)$ -entry is  $A_{ji}$ . That is,  $[A^T]_{ij} = A_{ji}$ .

**Example 3.** Here is an example of transposing a  $3 \times 2$  matrix, where we switch the matrix's rows with its columns as

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}.$$

### Properties of matrix transpose:

- $(A^T)^T = A$  for any matrix  $A \in \mathbb{R}^{m \times n}$ .
- $(A + B)^T = A^T + B^T$  with  $A, B \in \mathbb{R}^{m \times n}$ .
- $(AB)^T = B^T A^T$  with  $A \in \mathbb{R}^{m \times n}$  and  $B \in \mathbb{R}^{n \times p}$ .

*Proof.* Let  $AB = C$  and  $(AB)^T = D$ . Then,

$$\begin{aligned} (AB)_{ij}^T &= D_{ij} = C_{ji} \\ &= \sum_k A_{jk} B_{ki} \\ &= \sum_k (A^T)_{kj} (B^T)_{ik} \\ &= \sum_k (B^T)_{ik} (A^T)_{kj}. \end{aligned}$$

It shows that  $D = B^T A^T$  and the result follows. □

**Identity matrix:** The identity matrix  $\mathbf{I}_n$  is an  $n \times n$  (square) matrix given by

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

where it has all 1's on the diagonal and 0's everywhere else. It is sometimes abbreviated  $\mathbf{I}$  when the dimension of the matrix is clear. For any  $A \in \mathbb{R}^{m \times n}$ , it holds that  $A\mathbf{I}_n = \mathbf{I}_m A$ .

**Matrix inverse:** Given a square matrix  $A \in \mathbb{R}^{n \times n}$ , its *inverse*  $A^{-1}$  (if it exists) is the unique matrix satisfying

$$AA^{-1} = A^{-1}A = \mathbf{I}_n.$$

Notice that the inverse of a matrix may not always exist. Those matrices that have an inverse are called *invertible* or *nonsingular*.

**Properties of matrix inverse:** Whenever the matrices  $A, B \in \mathbb{R}^{n \times n}$  are invertible, we have the following properties.

- $(A^{-1})^{-1} = A$ .
- $(AB)^{-1} = B^{-1}A^{-1}$ .
- $(A^{-1})^T = (A^T)^{-1}$ . (It can be proved by noting that  $(A^{-1})^T(A^T) = (AA^{-1})^T = \mathbf{I}_n$ .)
- All the columns (or rows) of  $A$  are linearly independent, *i.e.*,  $\text{rank}(A) = n$ .
- $\det(A) \neq 0$ .

**Matrix rank:** The *rank* of a matrix  $A \in \mathbb{R}^{m \times n}$  is the dimension of the linear space spanned by its rows (or columns). One can verify that

- $\text{rank}(A) \leq \min\{m, n\}$  and  $\text{rank}(A) = \text{rank}(A^T)$ .
- $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$  for any  $A \in \mathbb{R}^{m \times n}$  and  $B \in \mathbb{R}^{n \times p}$ .

**Matrix trace:** For a square matrix  $A \in \mathbb{R}^{n \times n}$ , the *trace* of  $A$  is defined as

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

*i.e.*, it is the sum of all the diagonal entries of  $A$ . Specifically, the traces of matrices satisfy the following properties:

- $\text{tr}(aA + bB) = a\text{tr}(A) + b\text{tr}(B)$  for any  $A, B \in \mathbb{R}^{n \times n}$  and  $a, b \in \mathbb{R}$ .
- $\text{tr}(A) = \text{tr}(A^T)$  for any  $A \in \mathbb{R}^{n \times n}$ .
- $\text{tr}(AB) = \text{tr}(BA)$  for any  $A \in \mathbb{R}^{m \times n}$  and  $B \in \mathbb{R}^{n \times m}$ .

*Proof.* By direct calculations,

$$\begin{aligned}\operatorname{tr}(AB) &= \sum_{i=1}^m [AB]_{ii} = \sum_{i=1}^m \left( \sum_{k=1}^n A_{ik} B_{ki} \right) \\ &= \sum_{k=1}^n \left( \sum_{i=1}^m B_{ki} A_{ik} \right) = \sum_{k=1}^n [BA]_{kk} = \operatorname{tr}(BA).\end{aligned}$$

□

**Determinant:** For a square matrix  $A \in \mathbb{R}^{n \times n}$ , its *determinant*  $\det(A)$  or  $|A|$  is defined as

$$\det(A) = \sum_{\pi} \left( \operatorname{sign}(\pi) \prod_{i=1}^n A_{i\pi(i)} \right),$$

where the sum is over all  $n!$  permutations  $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$  and  $\operatorname{sign}(\pi) = 1$  or  $-1$  according to whether the minimum number of transpositions (*i.e.*, pairwise interchanges) necessary to achieve it starting from  $\{1, \dots, n\}$  is even or odd. One can also calculate  $\det(A)$  through the Laplace expansion by minor along row  $i$  or column  $j$  as

$$\det(A) = \sum_{k=1}^n (-1)^{i+k} A_{ik} \det(M_{ik}) = \sum_{k=1}^n (-1)^{k+j} A_{kj} \det(M_{kj}),$$

where  $M_{ik} \in \mathbb{R}^{(n-1) \times (n-1)}$  denotes the submatrix of  $A$  obtained by removing row  $i$  and column  $k$  of  $A$ . Geometrically, the determinant of  $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] \in \mathbb{R}^{n \times n}$  gives the signed volume of a  $n$ -dimensional parallelotope  $\mathcal{P} = \{c_1 \mathbf{a}_1 + \dots + c_n \mathbf{a}_n : c_1, \dots, c_n \in [0, 1]\}$ , *i.e.*,

$$\det A = \pm \operatorname{Volume}(\mathcal{P}),$$

where  $\mathbf{a}_1, \dots, \mathbf{a}_n$  are column vectors of  $A$ .

**Example 4.** We give explicit formulae for computing the determinants of square matrices with dimension less than 3 as:

$$\begin{aligned}\det[A_{11}] &= A_{11}, \\ \det \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} &= A_{11}A_{22} - A_{12}A_{21}, \\ \det \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} &= A_{11}A_{22}A_{33} + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32} \\ &\quad - A_{11}A_{23}A_{32} - A_{12}A_{21}A_{33} - A_{13}A_{22}A_{31}.\end{aligned}$$

**Properties of determinant:** For any  $A, B \in \mathbb{R}^{n \times n}$ ,

- $\det(AB) = \det(A) \cdot \det(B)$ .
- $\det(A^{-1}) = [\det(A)]^{-1}$  and  $\det(A^T) = \det(A)$ .

## 2.2 Special Types of Matrices

**Diagonal matrix:** A matrix  $D \in \mathbb{R}^{n \times n}$  is *diagonal* if  $D_{ij} = 0$  whenever  $i \neq j$ . We write a diagonal matrix  $D$  as

$$D = \text{diag}(d_1, d_2, \dots, d_n) = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}.$$

One can verify that

$$D^k = \begin{bmatrix} d_1^k & 0 & \cdots & 0 \\ 0 & d_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n^k \end{bmatrix}.$$

**Triangular matrix:** A matrix  $A \in \mathbb{R}^{n \times n}$  is *lower triangular* if  $A_{ij} = 0$  whenever  $i < j$ . That is, a lower triangular matrix has all its nonzero elements on or below the diagonal. Similarly, a matrix  $A$  is *upper triangular* if its transpose  $A^T$  is lower triangular. When  $A$  is a lower or upper triangular matrix,  $\det(A) = \prod_{i=1}^n A_{ii}$ .

**Orthogonal matrix:** A square matrix  $U \in \mathbb{R}^{n \times n}$  is orthogonal if  $UU^T = U^T U = \mathbf{I}_n$ . This implies that

- $U^{-1} = U^T$ , *i.e.*, the inverse of an orthogonal matrix is its transpose. Moreover,  $\det(U) = \pm 1$ .
- the rows (or columns) of  $U$  form an orthonormal basis for  $\mathbb{R}^n$ .
- $U$  preserves angles and lengths, *i.e.*, for any vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,

$$\langle U\mathbf{x}, U\mathbf{y} \rangle = (U\mathbf{x})^T (U\mathbf{y}) = \mathbf{x}^T U^T U \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle \quad \text{and} \quad \|U\mathbf{x}\|_2^2 = \|\mathbf{x}\|_2^2.$$

**Symmetric matrix:** A square matrix  $A \in \mathbb{R}^{n \times n}$  is symmetric if  $A = A^T$ , *i.e.*,  $A_{ij} = A_{ji}$  for all entries of  $A$ .

**Projection matrix:** A square matrix  $P \in \mathbb{R}^{n \times n}$  is a *projection matrix* if it is symmetric and idempotent:  $P^2 = P$ .

**Positive definite matrix:** A (real) symmetric matrix  $S \in \mathbb{R}^{n \times n}$  is *positive semi-definite* (PSD) if its quadratic form is nonnegative, *i.e.*,

$$\mathbf{x}^T S \mathbf{x} \geq 0$$

for all  $\mathbf{x} \in \mathbb{R}^n$ . Furthermore,  $S$  is *positive definite* (PD) if its quadratic form is strictly positive, *i.e.*,

$$\mathbf{x}^T S \mathbf{x} > 0$$

for all  $\mathbf{x} \in \mathbb{R}^n$  with  $\mathbf{x} \neq \mathbf{0}$ . Here are some useful properties of PSD or PD matrices.

- A diagonal matrix  $D = \text{diag}(d_1, \dots, d_n)$  is PSD if and only if  $d_i \geq 0$  for all  $i = 1, \dots, n$ . It is PD if and only if  $d_i > 0$  for all  $i = 1, \dots, n$ . In particular, the identity matrix  $\mathbf{I}_n$  is PD.
- If  $S \in \mathbb{R}^{n \times n}$  is PSD, then  $ASA^T$  is also PSD for any matrix  $A \in \mathbb{R}^{m \times n}$ .
- If  $S \in \mathbb{R}^{n \times n}$  is PD, then  $ASA^T$  is also PD for any matrix  $A \in \mathbb{R}^{m \times n}$  with full rank  $\text{rank}(A) = m \leq n$ .
- $AA^T$  is PSD for any matrix  $A \in \mathbb{R}^{m \times n}$ .  $AA^T$  is PD for any matrix  $A \in \mathbb{R}^{m \times n}$  with full rank  $\text{rank}(A) = m \leq n$ .
- $S$  is PD  $\implies S$  has full rank  $\implies S^{-1}$  exists  $\implies S^{-1} = (S^{-1})S(S^{-1})^T$  is PD.

## 2.3 Eigenvalues and Eigenvectors

Given a square matrix  $A \in \mathbb{R}^{n \times n}$ ,  $\lambda \in \mathbb{R}$  is an eigenvalue of  $A$  with the corresponding eigenvector  $\mathbf{x} \in \mathbb{R}^n$  and  $\mathbf{x} \neq \mathbf{0}$  if  $A\mathbf{x} = \lambda\mathbf{x}$ .

Here,  $\mathbf{0} \in \mathbb{R}^n$  stands for a vector whose entries are all zero. By convention, the zero vector cannot be an eigenvector of any matrix.

**Example 5.** If  $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ , then the vector  $\mathbf{x} = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$  is an eigenvector with eigenvalue 1, because

$$A\mathbf{x} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 3 \\ -3 \end{bmatrix} = 1 \times \begin{bmatrix} 3 \\ -3 \end{bmatrix}.$$

### 2.3.1 Solving for eigenvalues and eigenvectors

We exploit the fact that  $A\mathbf{x} = \lambda\mathbf{x}$  if and only if

$$(A - \lambda\mathbf{I}_n)\mathbf{x} = \mathbf{0}. \quad (1)$$

(Note that  $\lambda\mathbf{I}_n$  is the diagonal matrix where all the diagonal entries are  $\lambda$ , and all other entries are zero.)

The equation (1) has a nonzero solution  $\mathbf{x}$  if and only if  $\det(A - \lambda\mathbf{I}_n) = 0$ ; see Section 1.1 in [Horn and Johnson \(2012\)](#). Therefore, we can obtain the eigenvalues of a matrix  $A$  by solving the *characteristic equation*  $\det(A - \lambda\mathbf{I}_n) = 0$  for  $\lambda$ . Once we have done that, you can find the corresponding eigenvector for each eigenvalue  $\lambda$  by solving the system of equations  $(A - \lambda\mathbf{I}_n)\mathbf{x} = \mathbf{0}$  for  $\mathbf{x}$ .

**Example 6.** If  $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ , then

$$A - \lambda\mathbf{I}_n = \begin{bmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{bmatrix}$$

and

$$\det(A - \lambda\mathbf{I}_n) = (2 - \lambda)^2 - 1 = \lambda^2 - 4\lambda + 3.$$



Setting it to 0 yields that  $\lambda = 1$  and  $\lambda = 3$  are possible eigenvalues.

(i) To find the eigenvectors for  $\lambda = 1$ , we plug  $\lambda$  into the equation  $(A - \lambda \mathbf{I}_n)\mathbf{x} = 0$ . This gives us

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Any vector with  $x_2 = -x_1$  is a solution to this equation, and in particular,  $\begin{bmatrix} 3 \\ -3 \end{bmatrix}$  is one solution.

(ii) To find the eigenvectors for  $\lambda = 3$ , we again plug  $\lambda$  into the equation and obtain that

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Any vector where  $x_2 = x_1$  is a solution to this equation.

★ **Note:** The above method is never used to calculate eigenvalues and eigenvectors for large matrices in practice. We will introduce the power iterative method in the lecture (Lecture 6: Dimensionality Reduction) to find eigenpairs instead.

### 2.3.2 Properties of eigenvalues and eigenvectors

- If  $A \in \mathbb{R}^{n \times n}$  is symmetric, then all its eigenvalues are real.
- The eigenvalues of any (lower or upper) triangular matrix  $A \in \mathbb{R}^{n \times n}$  are its diagonal entries.
- The trace of a matrix  $A \in \mathbb{R}^{n \times n}$  is equal to the sum of its eigenvalues, *i.e.*,  $\text{tr}(A) = \sum_{i=1}^n \lambda_i$  with  $\lambda_1, \dots, \lambda_n$  being the eigenvalues of  $A$ .
- $\det(A) = \prod_{i=1}^n \lambda_i$ , where  $\lambda_1, \dots, \lambda_n$  is the eigenvalues of  $A \in \mathbb{R}^{n \times n}$ .
- A symmetric matrix is PSD (PD) if all its eigenvalues are nonnegative (positive).
- The eigenvalues of a projection matrix are either 1 or 0.

## 2.4 Matrix Norms

**Frobenius norm:** Given a matrix  $A \in \mathbb{R}^{m \times n}$ , its *Frobenius norm* is defined as

$$\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2} = \sqrt{\text{tr}(A^T A)}.$$

We can compute  $\|A\|_F$  as  $\|A\|_F = \sqrt{\sigma_1(A)^2 + \dots + \sigma_q(A)^2}$ , where  $\sigma_i(A), i = 1, \dots, q$  are singular values of  $A$  and  $q = \min\{m, n\}$ ; see [Section 3](#) for the definition of singular values. In particular, if  $A$  is a symmetric matrix in  $\mathbb{R}^{n \times n}$ , then  $\|A\|_F = \sqrt{\sum_{i=1}^n \lambda_i^2}$  with  $\lambda_1, \dots, \lambda_n$  being the eigenvalues of  $A$ .

**Maximum norm:** The maximum norm (or  $\ell_\infty$ -norm) for  $A \in \mathbb{R}^{m \times n}$  is defined as  $\|A\|_{\max} = \max_{i,j} |A_{ij}|$ . Strictly speaking,  $\|\cdot\|_{\max}$  is not a matrix norm because it does not satisfy the submultiplicativity  $\|AB\| \leq \|A\| \|B\|$ . However, it is a vector norm when we consider  $\mathbb{R}^{m \times n}$  as a  $mn$ -dimensional vector space; see Section 5.6 in [Horn and Johnson \(2012\)](#).

**Operator norm:** For any matrix  $A \in \mathbb{R}^{m \times n}$  and  $\ell_p$ -norm for vectors in  $\mathbb{R}^m$  and  $\mathbb{R}^n$ , then the corresponding operator norm  $\|A\|_p$  is defined as

$$\|A\|_p = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}.$$

For the special cases when  $p = 1, 2, \infty$ , these (induced) operator norms can be computed as

- $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |A_{ij}|$ , which is simply the maximum absolute column sum of the matrix.
- $\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |A_{ij}|$ , which is simply the maximum absolute row sum of the matrix.
- $\|A\|_2 = \sqrt{\lambda_{\max}(AA^T)} = \sigma_{\max}(A)$ , where  $\lambda_{\max}(AA^T)$  is the maximum eigenvalue of  $AA^T$  and  $\sigma_{\max}(A)$  is the maximum singular value of  $A$ .

There are several useful inequalities between these matrix norms. For any  $A \in \mathbb{R}^{m \times n}$ ,

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n} \|A\|_2, \quad \|A\|_{\max} \leq \|A\|_2 \leq \sqrt{mn} \|A\|_{\max}, \quad \text{and} \quad \|A\|_F \leq \sqrt{mn} \|A\|_{\max}.$$

### 3 Spectral Decomposition and Singular Value Decomposition (SVD)

**Theorem 1** (Spectral Decomposition of a Real Symmetric Matrix). *For a symmetric (square) matrix  $A \in \mathbb{R}^{n \times n}$ , there exists a real orthogonal matrix  $U \in \mathbb{R}^{n \times n}$  such that*

$$A = U \Lambda U^T = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^T,$$

where  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ ,  $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$ , and  $\mathbf{u}_1, \dots, \mathbf{u}_n$  are orthonormal eigenvectors of  $A$  associated with eigenvalues  $\lambda_1, \dots, \lambda_n$ .

The spectral decomposition also provides us with a convenient method for computing the power  $A^k = U \Lambda^k U^T$  and exponentiation  $\exp(A) = U \exp(\Lambda) U^T$  of a real symmetric matrix  $A \in \mathbb{R}^{n \times n}$ .

While the spectral decomposition ([Theorem 1](#)) only works for symmetric (square) matrices, it is also feasible to diagonalize a rectangular matrix  $A \in \mathbb{R}^{m \times n}$  through orthogonal matrices.

**Theorem 2** (Singular Value Decomposition (SVD)). *Let  $A \in \mathbb{R}^{m \times n}$  with  $q = \min\{m, n\}$ . There exist orthogonal matrices  $\tilde{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m] \in \mathbb{R}^{m \times m}$  and  $\tilde{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n] \in \mathbb{R}^{n \times n}$  as*

well as a (square) diagonal matrix  $\Sigma_q = \text{diag}(\sigma_1, \dots, \sigma_q) \in \mathbb{R}^{q \times q}$  such that

$$A = \tilde{U} \Sigma \tilde{V}^T = \sum_{i=1}^q \sigma_q \mathbf{u}_i \mathbf{v}_i^T = U \Sigma_q V^T,$$

where  $U = [\mathbf{u}_1, \dots, \mathbf{u}_q] \in \mathbb{R}^{m \times q}$ ,  $V = [\mathbf{v}_1, \dots, \mathbf{v}_q] \in \mathbb{R}^{n \times q}$ , and

$$\begin{aligned} \Sigma &= \Sigma_q \text{ if } m = n, \\ \Sigma &= [\Sigma_q \mathbf{0}] \in \mathbb{R}^{m \times n} \text{ if } n > m, \\ \Sigma &= \begin{bmatrix} \Sigma_q \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{m \times n} \text{ if } m > n. \end{aligned}$$

Here,  $\sigma_1, \dots, \sigma_q$  are called the **singular values** of  $A$ , which are eigenvalues of  $AA^T$  when  $m \leq n$  or  $A^T A$  when  $m > n$ .

Notice that the number of nonzero singular values of  $A$  determines the rank of  $A$ . During the lecture (Lecture 6: Dimensionality Reduction), we will leverage the singular value decomposition to reduce the dimension (or matrix rank) of a user-movie rating matrix.

## References

- S. Axler. *Linear algebra done right*. Springer, 3 edition, 2015.
- R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2 edition, 2012.
- M. Perlman. Probability and Mathematical Statistics I (STAT 512 Lecture Notes), 2020. URL <https://sites.stat.washington.edu/people/mdperlma/STAT%20512%20MDP%20Notes.pdf>.