Course Notes: Deep Learning for Visual Computing

Peter Wonka

August 28, 2022

Contents

1 Linear Algebra Review

1.1 Vectors

- A vector can specify a point or a direction in n-dimensional space
- Notation: $\mathbf{x} \in \mathbb{R}^n$
- A vector comes from a vector space $< \mathbb{R}^n, +, \mathbb{R} >$
- We generally assume column vectors
- Example:

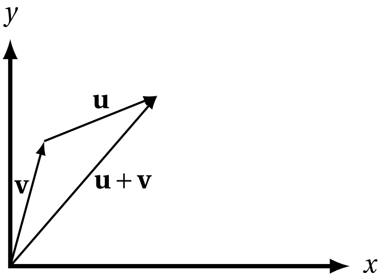
$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3, \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$$

- Length (2-norm): $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$
- Unit vector (normalized vector): $\|\mathbf{x}\|_2 = 1$
- Zero vector: all components are zeros, e.g. $(0,0,0)^T$
- Vector addition:

- $\mathbf{u}+\mathbf{v}$, e.g.:

$$\mathbf{u}+\mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{pmatrix}$$

- Graphical interpretation:

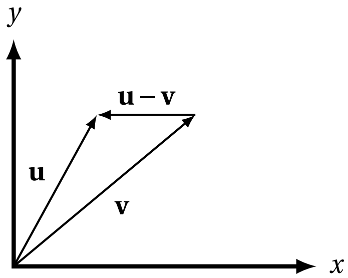


- Vector subtraction:

- $\mathbf{u}-\mathbf{v}$, e.g.:

$$\mathbf{u}-\mathbf{v}=\begin{pmatrix}u_1\\u_2\end{pmatrix}-\begin{pmatrix}v_1\\v_2\end{pmatrix}=\begin{pmatrix}u_1-v_1\\u_2-v_2\end{pmatrix}$$

- Graphical interpretation:

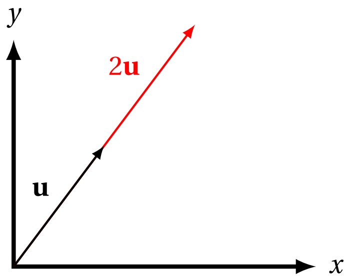


- Multiplication with a scalar:

- Computation:

$$a\mathbf{u} = a\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} au_1 \\ au_2 \\ au_3 \end{pmatrix}$$

- Graphical interpretation:



- Hadamard product:

- Terms: Hadamard product, Schur product, entrywise product
- Entry-wise multiplication
- Notation: $\mathbf{w} = \mathbf{u} \odot \mathbf{v}$
- Computation:

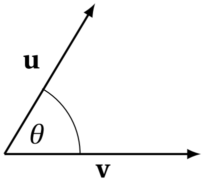
$$\mathbf{w} = \mathbf{u} \odot \mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \odot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} u_1 v_1 \\ u_2 v_2 \\ u_3 v_3 \end{pmatrix}$$

1.1.1 Inner Product

- Terms: dot-product, inner product, scalar product
- Measures how much two vectors are aligned
- Notation: $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v} = \mathbf{u} \cdot \mathbf{v}$
- Example:

$$< \mathbf{u}, \mathbf{v} >= \left(u_1, u_2, u_3\right) \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = u_1 v_1 + u_2 v_2 + u_3 v_3$$

- Graphical interpretation:



- The angle θ between two vectors $\mathbf{u}$ and $\mathbf{v}$ is:

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2}$$

- Two vectors are **perpendicular** or **orthogonal** iff:

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0 \quad \Rightarrow \quad \cos\theta = 0, \quad \theta = \frac{\pi}{2}$$

- Relationship to Euclidean norm: $\|\mathbf{x}\|_2^2 = \langle \mathbf{x}, \mathbf{x} \rangle$
- Commutative law: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
- Distributive law: $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$

1.1.2 Vector Norms

- A **norm** [wiki] is a function that assigns a strictly positive length or size to each vector in a vector space, except for the zero vector which is assigned a length of zero. A norm must also satisfy certain properties pertaining to **scalability** and **additivity**. The norm is generally denoted as $\|\cdot\|$.
- Specifically, a vector norm is a mapping $\mathbb{R}^n \to \mathbb{R}$, with the properties:
  - **non-negativity**: $\|\mathbf{x}\| \geq 0, \forall \mathbf{x}$; $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$
  - **scalability**: $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|, \forall \alpha \in \mathbb{R}$
  - **additivity**:$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. (also called **sub-additivity** or **triangle inequality**)
- Example definitions of a norm on a vector $\mathbf{x} \in \mathbb{R}^n$
  - 1-norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$
  - 2-norm: $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$
  - max-norm: $\|\mathbf{x}\|_\infty = \max_{1\leq i\leq n} |x_i|$

- All of the above are special cases of the L_p-norm (or p-norm):

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

- Two norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are equivalent if there exist constants C_1 and C_2 such that

$$C_1 \|\mathbf{x}\|_\alpha \leq \|\mathbf{x}\|_\beta \leq C_2 \|\mathbf{x}\|_\alpha, \forall \mathbf{x}$$

  - $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent with $\mathbf{x} \in \mathbb{R}^n$:

$$\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n} \|\mathbf{x}\|_2$$

1.1.3 Distance between vectors

- How to measure *distance* between vectors?
  - Obvious answer: the distance between two vectors $\mathbf{x}$ and $\mathbf{y}$ is $\|\mathbf{x}-\mathbf{y}\|$, where $\|\cdot\|$ is some vector norm.
  - Alternative: use the *angle* between two vectors $\mathbf{x}$ and $\mathbf{y}$ to measure the distance between them.
  - How to calculate the angle between two vectors?
  - The cosine of the angle between two vectors $\mathbf{x}$ and $\mathbf{y}$ can be used to measure the similarity between two vectors: if $\mathbf{x}$ and $\mathbf{y}$ are close, the angle between them is small and $\cos\theta \approx 1$; if $\mathbf{x}$ and $\mathbf{y}$ are orthogonal, i.e., $\langle \mathbf{x}, \mathbf{y} \rangle = 0, \cos\theta = 0, \theta = \frac{\pi}{2}$
- Why not just use the Euclidean distance?
  - Example: term-document matrix
  - Each entry tells how many times a term appears in the document:

	Doc 1	Doc 2	Doc 3
Term 1	10	1	0
Term 2	10	1	0
Term 3	0	1	0

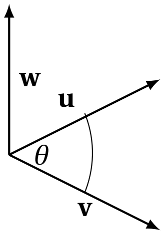
- Using the Euclidean distance Documents 1 and 2 look dissimilar, and Documents 2 and 3 look similar. This is just due to the length of the documents.
- Using the cosine of the angle between document vectors Documents 1 and 2 are similar to each other and dissimilar to Document 3.

1.1.4 Cross Product

- Notation: $\mathbf{w} = \mathbf{u} \times \mathbf{v}$
- The cross product is a vector
- $\mathbf{w}$ is orthogonal to $\mathbf{u}$ and $\mathbf{v}$
- Terms: cross product or vector product
- The direction of u follows the right hand rule:
  $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} \times \mathbf{v}$ form a right handed coordinate system
- Magnitude of $\mathbf{w}$ proportional to the sine of the angle between $\mathbf{u}$ and $\mathbf{v}$:

$$\|\mathbf{w}\|_2 = \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \sin\theta$$

- graphical interpretation:



- Computation:

$$\mathbf{w} = \mathbf{u} \times \mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{pmatrix}$$

- Cross product laws:
  - Distributive law:

$$\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$$

- Commutative law:

$$\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$$

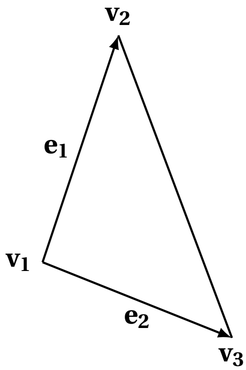
  Note: minus sign, cross product is not symmetric.

- Associative law:

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} \neq \mathbf{u} \times (\mathbf{v} \times \mathbf{w})$$

  Note: not associative

- Cross product is related to triangle area:
  - $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$: vertices of the triangle
  - $\mathbf{e}_1$, $\mathbf{e}_2$: two edges of the triangle
  - S: area of the triangle



$$\mathbf{e}_1 = \mathbf{v}_2 - \mathbf{v}_1$$

$$\mathbf{e}_2 = \mathbf{v}_3 - \mathbf{v}_1$$

$$S = \frac{1}{2}\|\mathbf{e_1} \times \mathbf{e_2}\|_2$$

1.2 Matrices

- Matrices are rectangular arrays that store numbers. An m-by-n matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ can be written as:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

1.3 Special Matrices

- The zero matrix is a matrix where all entries are 0
- The identity matrix $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$ has all entries 0, except for the diagonal entries are 1

$$\boldsymbol{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

1.4 Transpose Matrix

- Notation: $\boldsymbol{A}^T$
- The transpose matrix flips the entries accross the diagonal
- $\boldsymbol{A} \in \mathbb{R}^{m \times n} \rightarrow \boldsymbol{A}^T \in \mathbb{R}^{n \times m}$
- $\boldsymbol{A}$ has elements $a_{ij} \rightarrow \boldsymbol{A}^T$ has elements a_{ji}
- Example:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad \boldsymbol{A}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

1.5 Matrix-vector multiplication

- A matrix vector multiplication $\boldsymbol{A}\mathbf{x}$ is computed as follows:

$$\boldsymbol{A}\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix} = \mathbf{y}$$

- Each element of the output vector y is an inner product of a matrix row with the vector $\mathbf{x}$.

- Alternative interpretation: The output vector is a linear combination of column vectors of $\boldsymbol{A}$. The weight of column vector i is given by x_i:

$$\boldsymbol{A}\mathbf{x} = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n)\mathbf{x} = \sum_{i=1}^n \mathbf{a}_i x_i$$

1.6 Matrix-matrix multiplication

- Let $\boldsymbol{A} = \left(a_{ij}\right) \in \mathbb{R}^{m\times p}$ and $\boldsymbol{B} = \left(b_{ij}\right) \in \mathbb{R}^{p\times n}$. Then $\boldsymbol{C} = \boldsymbol{AB} = \left(c_{ij}\right) \in \mathbb{R}^{m\times n}$ is defined as follows:

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}, \forall i = 1, 2, ..., m, j = 1, ..., n$$

- Each column vector in $\boldsymbol{B}$ is multiplied $\boldsymbol{A}$

1.7 Linear Independence

- Given a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\} \in \mathbb{R}^m$ with $m \geq n$, consider the set of linear combinations $\mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$ for arbitrary coefficients α_i's.
- The vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ are linearly independent if:

$$\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = 0, \text{ if and only if } \alpha_i = 0, \forall i = 1, 2, ..., n$$

- A set of m linearly independent vectors of $\mathbb{R}^m$ is called a **basis** in $\mathbb{R}^m$: any vector in $\mathbb{R}^m$ can be expressed as a linear combination of the basis vectors.

1.8 Matrix Rank

- The **rank**(wiki) of a matrix is the maximum number of linearly independent column vectors.
- A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ with rank n is called **nonsingular** (also **invertible** or **nondegenerate**).
- A nonsingular matrix $\boldsymbol{A}$ has an inverse $\boldsymbol{A}^{-1}$ satisfying:

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}_n$$

- A square matrix is **not** invertible is called **singular** or **degenerate**. A square matrix is singular if and only if its determinant is 0.
- What is the rank of an outer-product matrix $\mathbf{x}\mathbf{y}^T \in \mathbb{R}^{m \times n}$ with $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$?

1.9 Range and Null Space

- $\mathbb{V}$ is a **subspace** of $\mathbb{R}^m$, if and only if $\alpha\mathbf{v}_1 + \beta\mathbf{v}_2 \in \mathbb{V}$, $\forall \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{V}$ and $\forall \alpha, \beta \in \mathbb{R}$.
  - Let $\mathbb{W} = \{(x, y) \in \mathbb{R}^2 | x \geq 0\}$, i.e., $\mathbb{W}$ is the set of all vectors $(x, y) \in \mathbb{R}^2$ in which $x \geq 0$. Is this a subspace of $\mathbb{R}^2$?
  - Let $\mathbb{W} = \left\{(0, x_2, x_3) \in \mathbb{R}^3 | \forall x_2, x_3\right\}$. Is this a subspace of $\mathbb{R}^3$?
  - Let $\mathbb{W} = \left\{(1, x_2, x_3) \in \mathbb{R}^3 | \forall x_2, x_3\right\}$. Is this a subspace of $\mathbb{R}^3$?
- The **range space**(wiki) (also called **image space** (wiki)) of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is defined by:

$$\textbf{range}(\boldsymbol{A}) = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = \boldsymbol{A}\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^n\}$$

- The **null space**(wiki) (also called **kernel**) of a matrix $\boldsymbol{A}$ is defined by:

$$\textbf{null}(A) = \{\mathbf{x} \in \mathbb{R}^n : \boldsymbol{A}\mathbf{x} = \mathbf{0}\}$$

- If follows from the definition that $\textbf{rank}(\boldsymbol{A}) = \textbf{dim}(\textbf{range}(\boldsymbol{A}))$.

  - The dimension of a space $\mathbb{V}$ denoted as $\mathbf{dim}(\mathbb{V})$, denotes the maximum number of linearly independent vectors in $\mathbb{V}$

- Show that $\forall \boldsymbol{A} \in \mathbb{R}^{m \times n}$ (this equation is also known as **Rank-nullity theorem** (wiki)):

$$\mathbf{dim}(\mathbf{null}(\boldsymbol{A})) + \mathbf{rank}(\boldsymbol{A}) = n$$

1.10 Eigenvalues and Eigenvectors

- **Definition**: Let $\boldsymbol{A}$ be an $n \times n$ matrix. A vector $\mathbf{v} \neq \mathbf{0}$ that satisfies:

$$\boldsymbol{A}\mathbf{v} = \lambda \mathbf{v}$$

  for some scalar λ is called an **eigenvector** of $\boldsymbol{A}$ and λ is the **eigenvalue** corresponding to the eigenvector $\mathbf{v}$.

- Example:

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix}, \quad \mathbf{v}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}$$

  - Check that:

$$\boldsymbol{A}\mathbf{v}_1 = -\mathbf{v}_1, \boldsymbol{A}\mathbf{v}_2 = -2\mathbf{v}_2$$

  - Let $\lambda_1 = -1, \lambda_2 = -2$, therefore: $\mathbf{v}_i$ is an eigenvector of $\boldsymbol{A}$ and λ_i is the corresponding eigenvalue.

- Suppose that λ is an eigenvalue of the matrix $\boldsymbol{A}$ with the corresponding eigenvector $\mathbf{x}$. Then if k is a positive integer, λ^k is an eigenvalue of the matrix $\boldsymbol{A}^k$ with corresponding eigenvector $\mathbf{x}$.
- Suppose $\boldsymbol{A} = (a_{ij})$ is an $n \times n$ triangular matrix:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & a_{nn} \end{pmatrix}$$

  - Check or convince yourself that the main diagonal entries $\{a_{ii}\}_{i=1}^{n}$ are the eigenvalues of the matrix $\boldsymbol{A}$.
- Suppose that $\boldsymbol{A}$ is a square matrix and further suppose that there exists an invertible matrix $\boldsymbol{P}$ (of the same size as $\boldsymbol{A}$) such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix. In such a case, we call $\boldsymbol{A}$ **diagonalizable**(wiki) and say that $\boldsymbol{P}$ diagonalizes $\boldsymbol{A}$.
  - Sufficient condition: if $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ has n distinct eigenvalues, $\boldsymbol{A}$ is diagonalizable.

- Let $\boldsymbol{P}$ be an invertible matrix. Show that $\boldsymbol{A}$ and $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ contain the same set of eigenvalues.
    - Check: $\mathbf{det}(\boldsymbol{I}_n - \boldsymbol{A}) = \mathbf{det}(\boldsymbol{I}_n - \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P})$

1.11 Matrix Norm

- A **matrix norm**(wiki) is a natural extension of the notion of a vector norm to matrices. Specifically, a matrix norm is a mapping $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ that satisfies the following properties:
  - $\|\boldsymbol{A}\| \geq 0$, $\|\boldsymbol{A}\| = 0$ if and only if $\boldsymbol{A} = \mathbf{0}$
  - $\|\alpha \boldsymbol{A}\| = |\alpha| \|\boldsymbol{A}\|$
  - $\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$
- Matrix norms induced by vector norm
  - Let $\|\cdot\|$ be a vector norm and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. The corresponding **matrix norm** is:

$$\|\boldsymbol{A}\| = \sup_{\mathbf{x}\in\mathbb{R}^n:\,\mathbf{x}\neq\mathbf{0}} \frac{\|\boldsymbol{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \sup_{\mathbf{x}\in\mathbb{R}^n:\,\|\mathbf{x}\|=1} \|\boldsymbol{A}\mathbf{x}\|$$

  - Show that:

- This is indeed a matrix norm (i.e., the non-negativity, scalability and triangle inequality hold).
- $\|\boldsymbol{A}\mathbf{x}\| \leq \|\boldsymbol{A}\|\|\mathbf{x}\|$
- $\|\boldsymbol{A}\boldsymbol{B}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|$

- Particularly, the induced matrix norm from the p-norm for vectors:

$$\|\boldsymbol{A}\|_p = \sup_{\mathbf{x}\in\mathbb{R}^n:\,\mathbf{x}\neq\mathbf{0}} \frac{\|\boldsymbol{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$

- In the special cases of $p = 1, 2, \infty$, the induced matrix norms can be computed by:
  - $\|\boldsymbol{A}\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^m |a_{ij}|$: maximum over columns.
  - $\|\boldsymbol{A}\|_2 = \sqrt{\lambda_{\max}(\boldsymbol{A}^T\boldsymbol{A})} = \sigma_{\max}(\boldsymbol{A})$: square root of the largest eigenvalue of $\boldsymbol{A}^T\boldsymbol{A}$ or the largest singular value of $\boldsymbol{A}$.
  - $\|\boldsymbol{A}\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^n |a_{ij}|$: maximum over rows.

- **Frobenius norm**: does not correspond to any vector norm:

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

  - Define **trace**$(\boldsymbol{B}) = \sum_{i=1}^{n} b_{ii}$ for any matrix $\boldsymbol{B} = (b_{ij}) \in \mathbb{R}^{n \times n}$.
  - Show that $\|\boldsymbol{A}\|_F^2 = \textbf{trace}(\boldsymbol{A}^T \boldsymbol{A})$

1.12 Condition Number

- Let $\|\cdot\|$ be some matrix norm, the **condition number**(wiki) of a matrix w.r.t this norm is defined as

$$\kappa(\boldsymbol{A}) = \|\boldsymbol{A}\| \|\boldsymbol{A}^{-1}\|$$

- Particularly, if $\|\cdot\|$ is the norm induced from the vector's L_2 norm, (i.e. $\|\cdot\|_2$), then

$$\kappa(\boldsymbol{A}) = \frac{\sigma_{\mathsf{max}}(\boldsymbol{A})}{\sigma_{\mathsf{min}}(\boldsymbol{A})}$$

  where $\sigma_{\mathsf{max}}(\boldsymbol{A})$ and $\sigma_{\mathsf{min}}(\boldsymbol{A})$ are maximal and minimal singular value of $\boldsymbol{A}$ respectively.

- If the condition number of a matrix is very large, then the matrix is called **ill-conditioned**. Practically, such a matrix is almost singular, and the computation of its inverse, or solution of a linear system of equations is prone to large numerical errors. A matrix that is not invertible has condition number equal to infinity.

- Example: consider the following matrix $\boldsymbol{A}$ with $a \neq 1$. $\boldsymbol{A}$ is nonsingular, but this is not enough, since the norm of $\boldsymbol{A}^{-1}$ tends to infinity as $a \to 1$

$$\boldsymbol{A} = \begin{pmatrix} a & 1 \\ 1 & 1 \end{pmatrix}, \quad \boldsymbol{A}^{-1} = \frac{1}{a-1} \begin{pmatrix} 1 & -1 \\ -1 & a \end{pmatrix}$$

1.13 Orthogonality

- Two vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal, if $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = 0$
- Given a set of **orthogonal** vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\} \in \mathbb{R}^m$ with $m \geq n$, i.e., $\mathbf{v}_i^T \mathbf{v}_j = 0$, $\forall i \neq j$, then they are linearly independent. Why?
- Let the set of orthogonal vectors $\mathbf{v}_i, i = 1, ..., n$ in $\mathbb{R}^m$ be **normalized**, i.e., $\|\mathbf{v}_i\| = 1$. Then they are **orthonormal**, and constitute an orthonormal basis in $\mathbb{R}^m$.
- An **orthogonal matrix**(wiki) is a square matrix whose columns and rows are orthonormal (orthogonal unit vectors). Prove the following properties of an orthogonal matrix:
  - An orthogonal matrix $\boldsymbol{Q} \in \mathbb{R}^{m \times m}$ has rank m.
  - $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$, that is $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{Q}^T \boldsymbol{Q} = \boldsymbol{I}_m$.
  - The Euclidean length of a vector $\mathbf{x} \in \mathbb{R}^m$ is invariant under an orthogonal transformation on $\boldsymbol{Q}$, that is:

$$\|\boldsymbol{Q}\mathbf{x}\|_2^2 = \|\mathbf{x}\|_2^2$$

– The product of two orthogonal matrices $\boldsymbol{Q}$ and $\boldsymbol{P}$ is orthogonal.

1.14 Eigenvalues and eigenvectors of a symmetric matrix

- The eigenvectors of a symmetric matrix are mutually orthogonal and its eigenvalues are real.
  - $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a symmetric matrix if $\boldsymbol{A} = \boldsymbol{A}^T$.
- A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ can be written in the form $\boldsymbol{A} = \boldsymbol{U} \Sigma \boldsymbol{U}^T$, where the columns of $\boldsymbol{U}$ ($\boldsymbol{U}$ is an orthogonal matrix) are the eigenvectors of $\boldsymbol{A}$ and Σ is a diagonal matrix, the diagonal elements of Σ which are the corresponding eigenvalues of $\boldsymbol{A}$. This is called the eigendecomposition of $\boldsymbol{A}$.
- Example of symmetric matrices: graphs:
  - The adjacency matrix of an undirected graph is a symmetric matrix.
- **Courant-Fischer Min-max Theorem**(wiki):
  - If $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is symmetric, then

$$\lambda_k(\boldsymbol{A}) = \max_{\dim(\mathbb{V})=k} \min_{\mathbf{0} \neq \mathbf{y} \in \mathbb{V}} \frac{\mathbf{y}^T \boldsymbol{A} \mathbf{y}}{\mathbf{y}^T \mathbf{y}}$$

- Assume both $\boldsymbol{A}$ and $\boldsymbol{E}$ are $n \times n$ symmetric matrices. Show that
  - $\lambda_k(\boldsymbol{A}) + \lambda_n(\boldsymbol{E}) \leq \lambda_k(\boldsymbol{A} + \boldsymbol{E}) \leq \lambda_k(\boldsymbol{A}) + \lambda_1(\boldsymbol{E})$
  - $|\lambda_k(\boldsymbol{A} + \boldsymbol{E}) - \lambda_k(\boldsymbol{A})| \leq \|\boldsymbol{E}\|_2$

1.15 Positive Semi-definite Matrix and Positive Definite Matrix

- A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is **positive semi-definite** (denoted as $\boldsymbol{A} \succeq 0$), if and only if $\mathbf{x}^T \boldsymbol{A}\mathbf{x} \geq 0$, for any $\mathbf{x} \in \mathbb{R}^n$.
  - All eigenvalues of $\boldsymbol{A}$ are non-negative.
  - $\boldsymbol{X}^T \boldsymbol{A} \boldsymbol{X}$ for any $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ is positive semi-definite.
- A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is **positive definite** (denoted as $\boldsymbol{A} \succ 0$), if and only if $\mathbf{x}^T \boldsymbol{A}\mathbf{x} > 0$, for any $\mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$.
  - All eigenvalues of $\boldsymbol{A}$ are positive.
  - All principal submatrices of $\boldsymbol{A}$ are positive definite.
  - All diagonal entries of $\boldsymbol{A}$ are positive.