Aalto University
School of Electrical
Engineering

Recap of Matrix Computations

Fatemeh Yaghoobi

Aalto University

2022

Contents

Basic Concepts and Notation

Linear algebra provides a way of compactly representing and operating on sets of linear equations.

- For example, consider the following system of equations:

$$
\begin{aligned}
4x_1 - 5x_2 &= -13 \\
-2x_1 + 3x_2 &= 9
\end{aligned}
$$

Basic Concepts and Notation

Linear algebra provides a way of compactly representing and operating on sets of linear equations.

- For example, consider the following system of equations:

$$\begin{aligned} 4x_1 - 5x_2 &= -13 \\ -2x_1 + 3x_2 &= 9 \end{aligned}$$

- Solving the system of 2 equations and 2 variables.

Basic Concepts and Notation

Linear algebra provides a way of compactly representing and operating on sets of linear equations.

- For example, consider the following system of equations:

$$\begin{aligned} 4x_1 - 5x_2 &= -13 \\ -2x_1 + 3x_2 &= 9 \end{aligned}$$

- Solving the system of 2 equations and 2 variables.
- Compact form: $\mathbf{A}\mathbf{x} = \mathbf{b}$

$$\mathbf{A} = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

Basic Concepts and Notation

Linear algebra provides a way of compactly representing and operating on sets of linear equations.

- For example, consider the following system of equations:

$$\begin{aligned} 4x_1 - 5x_2 &= -13 \\ -2x_1 + 3x_2 &= 9 \end{aligned}$$

- Solving the system of 2 equations and 2 variables.
- Compact form: $\mathbf{A}\mathbf{x} = \mathbf{b}$

$$\mathbf{A} = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

- What are the advantages of analyzing linear equations in this form?

Basic Notation

- **Scalars:** Quantities that are fully described by a magnitude (or numerical value) alone. Scalars are italic $(a, b, \ldots)$

Basic Notation

- **Scalars:** Quantities that are fully described by a magnitude (or numerical value) alone. Scalars are italic $(a, b, \dots)$
- $\mathbf{A} \in \mathbb{R}^{d_m \times d_n}$: a matrix with d_m rows and d_n columns, where the entries of $\mathbf{A}$ are real numbers. (Matrices are bold uppercase).

Basic Notation

- **Scalars:** Quantities that are fully described by a magnitude (or numerical value) alone. Scalars are italic $(a, b, \ldots)$
- $\mathbf{A} \in \mathbb{R}^{d_m \times d_n}$: a matrix with d_m rows and d_n columns, where the entries of $\mathbf{A}$ are real numbers. (Matrices are bold uppercase).
- $\mathbf{x} \in \mathbb{R}^{d_n}$: a vector with d_n entries. (Vectors are bold lowercase).

Basic Notation

- **Scalars:** Quantities that are fully described by a magnitude (or numerical value) alone. Scalars are italic $(a, b, \ldots)$

- $\mathbf{A} \in \mathbb{R}^{d_m \times d_n}$: a matrix with d_m rows and d_n columns, where the entries of $\mathbf{A}$ are real numbers. (Matrices are bold uppercase).
- $\mathbf{x} \in \mathbb{R}^{d_n}$: a vector with d_n entries. (Vectors are bold lowercase).

- **Column vector:** an d_n-dimensional vector is often thought of as a matrix with d_n rows and 1 column.

Basic Notation

- **Scalars:** Quantities that are fully described by a magnitude (or numerical value) alone. Scalars are italic $(a, b, \ldots)$
- $\mathbf{A} \in \mathbb{R}^{d_m \times d_n}$: a matrix with d_m rows and d_n columns, where the entries of $\mathbf{A}$ are real numbers. (Matrices are bold uppercase).
- $\mathbf{x} \in \mathbb{R}^{d_n}$: a vector with d_n entries. (Vectors are bold lowercase).
- **Column vector:** an d_n-dimensional vector is often thought of as a matrix with d_n rows and 1 column.
- **Row vector:** a matrix with 1 row and d_n columns — we typically write $\mathbf{x}^T$ (here $\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$, which we will define shortly).

Vectors

- A 3-dimensional vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ has 3 elements v_1, v_2, v_3 as in

$$\mathbf{v} = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$$

Vectors

- A 3-dimensional vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ has 3 elements v_1, v_2, v_3 as in

$$\mathbf{v} = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$$

- Number of elements is the dimension of the vector

Vectors

- A 3-dimensional vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ has 3 elements v_1, v_2, v_3 as in

  $$\mathbf{v} = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$$

- Number of elements is the dimension of the vector
- We use bold symbols to denote vectors, e.g., $\mathbf{v}, \mathbf{x}, \ldots$

Vectors

- A 3-dimensional vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ has 3 elements v_1, v_2, v_3 as in

  $$\mathbf{v} = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$$

- Number of elements is the dimension of the vector
- We use bold symbols to denote vectors, e.g., $\mathbf{v}, \mathbf{x}, \ldots$
- We add vectors $\mathbf{v} + \mathbf{w}$. We multiply them by numbers (scalars) like $c = 4$ and $d = 0$

  $$\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix}, \quad 4 \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 12 \\ 16 \\ 20 \end{bmatrix}, \quad 0 \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Vectors

- **Ones vector:** d_n-dimensional vector with all entries 1

Vectors

- **Ones vector:** d_n-dimensional vector with all entries 1
- **Unit vector:** a unit vector has one entry 1 and all others 0 denoted $\mathbf{e}_i$ where i is entry that is 1

Vectors

- **Ones vector:** d_n-dimensional vector with all entries 1
- **Unit vector:** a unit vector has one entry 1 and all others 0 denoted $\mathbf{e}_i$ where i is entry that is 1
- **Linear combinations:**
  for d_n-dimensional vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ and scalars $\beta_1, \ldots, \beta_m$,
  $$\beta_1 \mathbf{v}_1 + \ldots + \beta_m \mathbf{v}_m$$
  is a linear combination of the vectors
  $$1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + 4 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 11 \\ 10 \\ 13 \end{bmatrix}$$

Vectors

- **Ones vector:** d_n-dimensional vector with all entries 1
- **Unit vector:** a unit vector has one entry 1 and all others 0 denoted $\mathbf{e}_i$ where i is entry that is 1
- **Linear combinations:**
  for d_n-dimensional vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ and scalars $\beta_1, \ldots, \beta_m$,
  $$\beta_1 \mathbf{v}_1 + \ldots + \beta_m \mathbf{v}_m$$
  is a linear combination of the vectors
  $$1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + 4 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 11 \\ 10 \\ 13 \end{bmatrix}$$
- Sometimes a combination gives the zero vector. Then the vectors are **linearly dependent**
  $$-1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} - 1 \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$$

Vectors

- **Dot Products** $\mathbf{v} \cdot \mathbf{w}$ or $\mathbf{v}^\top \mathbf{w}$

$$\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = 3 \times 2 + 4 \times 0 + 5 \times 1 = 11 \; \boxed{\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}}$$

Vectors

- **Dot Products** $\mathbf{v} \cdot \mathbf{w}$ or $\mathbf{v}^\top \mathbf{w}$

$$\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = 3 \times 2 + 4 \times 0 + 5 \times 1 = 11 \; \boxed{\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}}$$

- $\mathbf{e_i}^\top \mathbf{v} = v_i$ (picks out the ith entry)

Vectors

- **Dot Products** $\mathbf{v} \cdot \mathbf{w}$ or $\mathbf{v}^\top \mathbf{w}$

$$\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = 3 \times 2 + 4 \times 0 + 5 \times 1 = 11 \quad \boxed{\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}}$$

- $\mathbf{e_i}^\top \mathbf{v} = v_i$ (picks out the ith entry)
- $\mathbf{1}^\top \mathbf{v} = v_1 + \ldots + v_n$

Vectors

- **Dot Products** $\mathbf{v} \cdot \mathbf{w}$ or $\mathbf{v}^\top \mathbf{w}$

$$\begin{bmatrix}3\\4\\5\end{bmatrix} \cdot \begin{bmatrix}2\\0\\1\end{bmatrix} = 3 \times 2 + 4 \times 0 + 5 \times 1 = 11 \quad \boxed{\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}}$$

- $\mathbf{e_i}^\top \mathbf{v} = v_i$ (picks out the ith entry)
- $\mathbf{1}^\top \mathbf{v} = v_1 + \ldots + v_n$
- **Length squared** of $\mathbf{v} = \begin{bmatrix}3\\4\end{bmatrix}$ is $\mathbf{v} \cdot \mathbf{v} = \mathbf{v}^\top \mathbf{v} = 3^2 + 4^2 = 9 + 16$

Vectors

- **Dot Products** $\mathbf{v} \cdot \mathbf{w}$ or $\mathbf{v}^\top \mathbf{w}$

$$\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = 3 \times 2 + 4 \times 0 + 5 \times 1 = 11 \; \boxed{\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}}$$

- $\mathbf{e_i}^\top \mathbf{v} = v_i$ (picks out the ith entry)
- $\mathbf{1}^\top \mathbf{v} = v_1 + \ldots + v_n$
- **Length squared** of $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ is $\mathbf{v} \cdot \mathbf{v} = \mathbf{v}^\top \mathbf{v} = 3^2 + 4^2 = 9 + 16$
- The dot product $\mathbf{v} \cdot \mathbf{w}$ reveals the **angle** θ between $\mathbf{v}$ and $\mathbf{w}$: $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos\theta$,

Vectors

- **Dot Products** $\mathbf{v} \cdot \mathbf{w}$ or $\mathbf{v}^\top \mathbf{w}$

$$\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = 3 \times 2 + 4 \times 0 + 5 \times 1 = 11 \quad \boxed{\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}}$$

- $\mathbf{e_i}^\top \mathbf{v} = v_i$ (picks out the ith entry)
- $\mathbf{1}^\top \mathbf{v} = v_1 + \ldots + v_n$
- **Length squared** of $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ is $\mathbf{v} \cdot \mathbf{v} = \mathbf{v}^\top \mathbf{v} = 3^2 + 4^2 = 9 + 16$
- The dot product $\mathbf{v} \cdot \mathbf{w}$ reveals the **angle** θ between $\mathbf{v}$ and $\mathbf{w}$: $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos\theta$,
- The angle between $\mathbf{v} = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$ is $\theta = 90^\circ$ because $\mathbf{v} \cdot \mathbf{w} = 0$

Vectors

- **Norm**: norm of a vector $\|\mathbf{x}\|$ is a measure of the “length” of the vector.
  For example, Euclidean or l_2 norm: $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$
  The l_1 norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$

Matrices

- **Matrices Multiplying Vectors:**
  There is a row way to multiply $\mathbf{A}\mathbf{x}$ and also a column way to compute the vector $\mathbf{A}\mathbf{x}$

Matrices

- **Matrices Multiplying Vectors:**
  There is a row way to multiply $\mathbf{Ax}$ and also a column way to compute the vector $\mathbf{Ax}$
- **Row way** $=$ **Dot product** of vector $\mathbf{x}$ with each row of $\mathbf{A}$

$$\mathbf{Ax} = \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 2v_1 + 5v_2 \\ 3v_1 + 7v_2 \end{bmatrix} \qquad \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}$$

Matrices

- **Matrices Multiplying Vectors:**
  There is a row way to multiply $\mathbf{Ax}$ and also a column way to compute the vector $\mathbf{Ax}$
- **Row way** $=$ **Dot product** of vector $\mathbf{x}$ with each row of $\mathbf{A}$

$$\mathbf{Ax} = \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 2v_1 + 5v_2 \\ 3v_1 + 7v_2 \end{bmatrix} \qquad \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}$$

- **Column way** $=$ $\mathbf{Ax}$ is a combination of the columns of $\mathbf{A}$

$$\mathbf{Ax} = \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1 \begin{bmatrix} \text{column} \\ 1 \end{bmatrix} + v_2 \begin{bmatrix} \text{column} \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}$$

Matrices

- The **identity matrix** has $\mathbf{I}\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Matrices

- The **identity matrix** has $\mathbf{I}\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

- **Dependent and Independent Columns**
  The columns of $\mathbf{A}$ are "dependent" if one column is a linear combination of the other columns
  Or $\mathbf{A}\mathbf{x} = \mathbf{0}$ for some vector $\mathbf{x}$ (other than $\mathbf{x} = \mathbf{0}$)

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 1 & 2 \end{bmatrix}$$ Reason: Column 2 of $\mathbf{A}_1$ = 2 (Column 1)

$$\mathbf{A}_2 = \begin{bmatrix} 1 & 4 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 0 \end{bmatrix}$$ Reason: $\mathbf{A}_2$ times $\mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ gives $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

Matrices

- The columns of $\mathbf{A}$ are "independent" if no column is a linear combination of the other columns
  Another way to say it : $\mathbf{Ax} = \mathbf{0}$ only when $\mathbf{x} = \mathbf{0}$

$$\mathbf{A}_3 = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 9 \end{bmatrix} \text{ and } \mathbf{A}_4 = \mathbf{I}$$

Matrices: matrix-matrix multiplication $\mathbf{AB}$

There are 4 ways to multiply matrices.

1. (Row i of $\mathbf{A}$) $\cdot$ (Column j of $\mathbf{B}$) produces one number: row i, column j of $\mathbf{AB}$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} 17 & . \\ . & . \end{bmatrix}$$

because $\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} = 17$ **Dot product**

Matrices: matrix-matrix multiplication $\mathbf{AB}$

There are 4 ways to multiply matrices.

1. (Row i of $\mathbf{A}$) $\cdot$ (Column j of $\mathbf{B}$) produces one number: row i, column j of $\mathbf{AB}$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} 17 & . \\ . & . \end{bmatrix}$$

because $\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} = 17$ **Dot product**

2. (Matrix $\mathbf{A}$) (Column j of $\mathbf{B}$) produces column j of $\mathbf{AB}$:
Combine columns of $\mathbf{A}$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} \mathbf{5} & 7 \\ \mathbf{6} & 8 \end{bmatrix} = \begin{bmatrix} \mathbf{17} & . \\ \mathbf{39} & . \end{bmatrix}$$

because $5 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 6 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 17 \\ 39 \end{bmatrix}$ **Linear combinations.**

Matrices: matrix-matrix Products $\mathbf{AB}$

3. (Row i of $\mathbf{A}$) (Matrix $\mathbf{B}$) produces row i of $\mathbf{AB}$: **Combine rows of $\mathbf{B}$**

$$\begin{bmatrix} \mathbf{1} & \mathbf{2} \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} \mathbf{17} & \mathbf{23} \\ . & . \end{bmatrix}$$

because $1 \begin{bmatrix} 5 & 7 \end{bmatrix} + 2 \begin{bmatrix} 6 & 8 \end{bmatrix} = \begin{bmatrix} 17 & 23 \end{bmatrix}$

Matrices: matrix-matrix Products $\mathbf{AB}$

3. (Row i of $\mathbf{A}$) (Matrix $\mathbf{B}$) produces row i of $\mathbf{AB}$: **Combine rows of $\mathbf{B}$**

$$\begin{bmatrix} \mathbf{1} & \mathbf{2} \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} \mathbf{17} & \mathbf{23} \\ . & . \end{bmatrix}$$

because $1\begin{bmatrix} 5 & 7 \end{bmatrix} + 2\begin{bmatrix} 6 & 8 \end{bmatrix} = \begin{bmatrix} 17 & 23 \end{bmatrix}$

4. (Column k of $\mathbf{A}$) (Row k of $\mathbf{B}$) produces a matrix : Add these matrices!

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 5 & 7 \end{bmatrix} = \begin{bmatrix} 5 & 7 \\ 15 & 21 \end{bmatrix} \text{ and } \begin{bmatrix} 2 \\ 4 \end{bmatrix} \begin{bmatrix} 6 & 8 \end{bmatrix} = \begin{bmatrix} 12 & 16 \\ 24 & 32 \end{bmatrix}$$

Now add: $\mathbf{AB} = \begin{bmatrix} 17 & 23 \\ 39 & 53 \end{bmatrix}$ **Outer product**

Operations and Properties

- **Transpose** of $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 4 \end{bmatrix}$ is $\mathbf{A}^\top = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 4 \end{bmatrix}$

$$\boxed{(\mathbf{A}^\top)_{ij} = \mathbf{A}_{ji}}$$

Operations and Properties

- **Transpose** of $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 4 \end{bmatrix}$ is $\mathbf{A}^\top = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 4 \end{bmatrix}$

$$\boxed{(\mathbf{A}^\top)_{ij} = \mathbf{A}_{ji}}$$

- Rules for sum and product

Transpose of $\mathbf{A} + \mathbf{B}$ is $\mathbf{A}^\top + \mathbf{B}^\top$

Transpose of $\mathbf{AB}$ is $\mathbf{B}^\top \mathbf{A}^\top$

Operations and Properties

- **Transpose** of $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 4 \end{bmatrix}$ is $\mathbf{A}^\top = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 4 \end{bmatrix}$

  $(\mathbf{A}^\top)_{ij} = \mathbf{A}_{ji}$

- Rules for sum and product

  Transpose of $\mathbf{A} + \mathbf{B}$ is $\mathbf{A}^\top + \mathbf{B}^\top$

  Transpose of $\mathbf{AB}$ is $\mathbf{B}^\top \mathbf{A}^\top$

- **A symmetric matrix has S**$^\top =$ **S** means $s_{ij} = s_{ji}$

Operations and Properties

- **Determinant:** The determinant of a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, is a function det : $\mathbb{R}^{n\times n} \to \mathbb{R}$, and is denoted as $|\mathbf{A}|$ or det $\mathbf{A}$
  The equations for determinants of matrices up to size 3×3:

$$
\begin{aligned}
|[a_{11}]| &= a_{11} \\
\left|\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right| &= a_{11}a_{22} - a_{12}a_{21} \\
\left|\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right| &= \begin{array}{l} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ \quad -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{array}
\end{aligned}
$$

Operations and Properties

- **Positive definite:** A symmetric matrix $\mathbf{A}$ is positive definite (PD) if for all non-zero vectors $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$. This is usually denoted as $\mathbf{A} > 0$.

Operations and Properties

- **Positive definite:** A symmetric matrix $\mathbf{A}$ is positive definite (PD) if for all non-zero vectors $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$. This is usually denoted as $\mathbf{A} > 0$.
- **Positive semidefinite:** A symmetric matrix $\mathbf{A}$ is positive semidefinite (PSD) if for all vectors $\mathbf{x} \in R^n$, $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$. This is usually denoted as $\mathbf{A} \geq 0$.

Operations and Properties

- **Positive definite:** A symmetric matrix $\mathbf{A}$ is positive definite (PD) if for all non-zero vectors $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$. This is usually denoted as $\mathbf{A} > 0$.
- **Positive semidefinite:** A symmetric matrix $\mathbf{A}$ is positive semidefinite (PSD) if for all vectors $\mathbf{x} \in R^n$, $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$. This is usually denoted as $\mathbf{A} \geq 0$.
- **Negative definite:** A symmetric matrix $\mathbf{A}$ is negative definite (ND) if for all non-zero vectors $\mathbf{x} \in R^n$, $\mathbf{x}^T\mathbf{A}\mathbf{x} < 0$. This is usually denoted as $\mathbf{A} < 0$.

Operations and Properties

- **Positive definite:** A symmetric matrix $\mathbf{A}$ is positive definite (PD) if for all non-zero vectors $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$. This is usually denoted as $\mathbf{A} > 0$.
- **Positive semidefinite:** A symmetric matrix $\mathbf{A}$ is positive semidefinite (PSD) if for all vectors $\mathbf{x} \in R^n$, $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$. This is usually denoted as $\mathbf{A} \geq 0$.
- **Negative definite:** A symmetric matrix $\mathbf{A}$ is negative definite (ND) if for all non-zero vectors $\mathbf{x} \in R^n$, $\mathbf{x}^T \mathbf{A} \mathbf{x} < 0$. This is usually denoted as $\mathbf{A} < 0$.
- **Negative semidefinite:** A symmetric matrix $\mathbf{A}$ is negative semidefinite (NSD) if for all vectors $\mathbf{x} \in R^n$, $\mathbf{x}^T \mathbf{A} \mathbf{x} \leq 0$. This is usually denoted as $\mathbf{A} \leq 0$.

Solving Linear Equations
$\mathbf{A}\mathbf{x} = \mathbf{b}$: $\mathbf{A}$ is n by n

- **Inverse Matrices $\mathbf{A}^{-1}$ and Solutions $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$**

The inverse of a square matrix $\mathbf{A}$ has $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}^{-1}$$

A has no inverse if $ad - bc = 0$

Solving Linear Equations
$\mathbf{Ax} = \mathbf{b}$: $\mathbf{A}$ is n by n

- **Inverse Matrices $\mathbf{A}^{-1}$ and Solutions $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$**

The inverse of a square matrix $\mathbf{A}$ has $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and $\mathbf{AA}^{-1} = \mathbf{I}$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}^{-1}$$

A has no inverse if $ad - bc = 0$

- Invertible ⇔ Rows are independent ⇔ Columns are independent.

Solving Linear Equations
$\mathbf{A}\mathbf{x} = \mathbf{b}$: $\mathbf{A}$ is n by n

- **Inverse Matrices $\mathbf{A}^{-1}$ and Solutions $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$**

The inverse of a square matrix $\mathbf{A}$ has $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$

  $$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}^{-1}$$

  A has no inverse if $ad - bc = 0$

- Invertible $\Leftrightarrow$ Rows are independent $\Leftrightarrow$ Columns are independent.
- Invertible $\Leftrightarrow$ The only solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$

Solving Linear Equations
$\mathbf{Ax} = \mathbf{b}$: $\mathbf{A}$ is n by n

- **Inverse Matrices $\mathbf{A}^{-1}$ and Solutions $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$**

The inverse of a square matrix $\mathbf{A}$ has $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}^{-1}$$

A has no inverse if $ad - bc = 0$

- Invertible $\Leftrightarrow$ Rows are independent $\Leftrightarrow$ Columns are independent.
- Invertible $\Leftrightarrow$ The only solution to $\mathbf{Ax} = \mathbf{b}$ is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$
- Computing $\mathbf{A}^{-1}$ is not efficient for $\mathbf{Ax} = \mathbf{b}$.   Use: elimination.

Solving Linear Equations
$\mathrm{Ax} = \mathrm{b}$: A is n by n

- **Elimination : Square Matrix A to Triangular U**

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 11 & 14 \\ 2 & 8 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 2 & 8 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 5 & 13 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 0 & 7 \end{bmatrix} = U$$

Solving Linear Equations
$\mathrm{Ax} = \mathrm{b}$: A is n by n

- **Elimination : Square Matrix A to Triangular U**

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 11 & 14 \\ 2 & 8 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 2 & 8 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 5 & 13 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 0 & 7 \end{bmatrix} = U$$

- One elimination step subtracts l_{ij} times row j from row i $(i > j)$

Solving Linear Equations
$\mathrm{Ax} = \mathrm{b}$: A is n by n

- **Elimination : Square Matrix A to Triangular U**

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 11 & 14 \\ 2 & 8 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 2 & 8 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 5 & 13 \end{bmatrix} \rightarrow$$
$$\begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 0 & 7 \end{bmatrix} = U$$

- One elimination step subtracts l_{ij} times row j from row i $(i > j)$
- Each step produces a zero below the diagonal of $\mathbf{U}: l_{21} = 2, l_{31} = l_{32} = 1$

Solving Linear Equations
$\mathrm{Ax} = \mathrm{b}$: A is n by n

- **Elimination : Square Matrix $\mathbf{A}$ to Triangular $\mathbf{U}$**

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 11 & 14 \\ 2 & 8 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 2 & 8 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 5 & 13 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 0 & 7 \end{bmatrix} = U$$

- One elimination step subtracts l_{ij} times row j from row i $(i > j)$
- Each step produces a zero below the diagonal of $\mathbf{U}: l_{21} = 2, l_{31} = l_{32} = 1$
- Elimination produced no zeros on the diagonal and created 3 zeros in $\mathbf{U}$.

Solving Linear Equations
$\mathbf{A}\mathbf{x} = \mathbf{b}$: $\mathbf{A}$ is n by n

- For $\mathbf{A}\mathbf{x} = \mathbf{b}$ add extra column $\mathbf{b}$, **Elimination and back substitution**

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 & 19 \\ 4 & 11 & 14 & 55 \\ 2 & 8 & 17 & 50 \end{bmatrix} \to \begin{bmatrix} \mathbf{U} & \mathbf{c} \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 & 19 \\ 0 & 5 & 6 & 17 \\ 0 & 0 & 7 & 14 \end{bmatrix}$$

Solving Linear Equations
$\mathbf{Ax} = \mathbf{b}$: $\mathbf{A}$ is n by n

- For $\mathbf{Ax} = \mathbf{b}$ add extra column $\mathbf{b}$, **Elimination and back substitution**

$$\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix} = \begin{bmatrix}2 & 3 & 4 & 19\\ 4 & 11 & 14 & 55\\ 2 & 8 & 17 & 50\end{bmatrix} \rightarrow \begin{bmatrix}\mathbf{U} & \mathbf{c}\end{bmatrix} = \begin{bmatrix}2 & 3 & 4 & 19\\ 0 & 5 & 6 & 17\\ 0 & 0 & 7 & 14\end{bmatrix}$$

- **Back substitution:** The last equation $7x_3 = 14$ gives $x_3 = 2$

Solving Linear Equations
$\mathbf{A}\mathbf{x} = \mathbf{b}$: $\mathbf{A}$ is n by n

- For $\mathbf{A}\mathbf{x} = \mathbf{b}$ add extra column $\mathbf{b}$, **Elimination and back substitution**

$$\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 & 19 \\ 4 & 11 & 14 & 55 \\ 2 & 8 & 17 & 50 \end{bmatrix} \rightarrow \begin{bmatrix}\mathbf{U} & \mathbf{c}\end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 & 19 \\ 0 & 5 & 6 & 17 \\ 0 & 0 & 7 & 14 \end{bmatrix}$$

- **Back substitution:** The last equation $7x_3 = 14$ gives $x_3 = 2$
- **Work upwards** The next equation $5x_2 + 6(2) = 17$ gives $x_2 = 1$

Solving Linear Equations
$\mathbf{A}\mathbf{x} = \mathbf{b}$: $\mathbf{A}$ is n by n

- For $\mathbf{A}\mathbf{x} = \mathbf{b}$ add extra column $\mathbf{b}$, **Elimination and back substitution**

$$\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix} = \begin{bmatrix}2 & 3 & 4 & 19 \\ 4 & 11 & 14 & 55 \\ 2 & 8 & 17 & 50\end{bmatrix} \to \begin{bmatrix}\mathbf{U} & \mathbf{c}\end{bmatrix} = \begin{bmatrix}2 & 3 & 4 & 19 \\ 0 & 5 & 6 & 17 \\ 0 & 0 & 7 & 14\end{bmatrix}$$

- **Back substitution:** The last equation $7x_3 = 14$ gives $x_3 = 2$
- **Work upwards** The next equation $5x_2 + 6(2) = 17$ gives $x_2 = 1$
- **Upwards again** The first equation $2x_1 + 3(1) + 4(2) = 19$ gives $x_1 = 4$

Solving Linear Equations
$\mathbf{A}\mathbf{x} = \mathbf{b}$: $\mathbf{A}$ is n by n

- For $\mathbf{A}\mathbf{x} = \mathbf{b}$ add extra column $\mathbf{b}$, **Elimination and back substitution**

$$\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix} = \begin{bmatrix}2 & 3 & 4 & 19\\ 4 & 11 & 14 & 55\\ 2 & 8 & 17 & 50\end{bmatrix} \to \begin{bmatrix}\mathbf{U} & \mathbf{c}\end{bmatrix} = \begin{bmatrix}2 & 3 & 4 & 19\\ 0 & 5 & 6 & 17\\ 0 & 0 & 7 & 14\end{bmatrix}$$

- **Back substitution:** The last equation $7x_3 = 14$ gives $x_3 = 2$
- **Work upwards** The next equation $5x_2 + 6(2) = 17$ gives $x_2 = 1$
- **Upwards again** The first equation $2x_1 + 3(1) + 4(2) = 19$ gives $x_1 = 4$
- **Conclusion** The only solution to this example is $\mathbf{x}^\top = (4, 1, 2)$

Cholesky decomposition

- **Matrix decompositions**: Describing a matrix by means of a different representation using factors of interpretable matrices.

Cholesky decomposition

- **Matrix decompositions**: Describing a matrix by means of a different representation using factors of interpretable matrices.
- An analogy for matrix decomposition is the factoring of numbers, such as the factoring of 21 into prime numbers 7×3.

Cholesky decomposition

- **Matrix decompositions**: Describing a matrix by means of a different representation using factors of interpretable matrices.
- An analogy for matrix decomposition is the factoring of numbers, such as the factoring of 21 into prime numbers 7×3.
- **Cholesky decomposition:** A square-root-like operation for symmetric, positive definite matrices $\mathbf{A}$.

Cholesky decomposition

- **Matrix decompositions**: Describing a matrix by means of a different representation using factors of interpretable matrices.
- An analogy for matrix decomposition is the factoring of numbers, such as the factoring of 21 into prime numbers 7×3.
- **Cholesky decomposition:** A square-root-like operation for symmetric, positive definite matrices $\mathbf{A}$.
- Matrix $\mathbf{A}$ can be factorized into a product $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$, where $\mathbf{L}$ is a lower-triangular matrix with positive diagonal elements:

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

Cholesky decomposition

- Cholsky factorization of $\mathbf{A} \in \mathbb{R}^{3\times 3}$

$$\begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \mathbf{L}\mathbf{L}^\top = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{31} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

Multiplying out the right-hand side yields:

$$\mathbf{A} = \begin{bmatrix} l_{11}^2 & l_{21}l_{11} & l_{31}l_{11} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{31}l_{21} + l_{32}l_{22} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

Cholesky decomposition

- Cholsky factorization of $\mathbf{A} \in \mathbb{R}^{3\times 3}$

$$\begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \mathbf{L}\mathbf{L}^\top = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{31} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

Multiplying out the right-hand side yields:

$$\mathbf{A} = \begin{bmatrix} l_{11}^2 & l_{21}l_{11} & l_{31}l_{11} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{31}l_{21} + l_{32}l_{22} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

- There is a simple pattern

$l_{11} = \sqrt{a_{11}}$, $l_{22} = \sqrt{a_{22} - l_{21}^2}$, $l_{33} = \sqrt{a_{33} - (l_{31}^2 + l_{32}^2)}$

$l_{21} = \frac{1}{l_{11}} a_{21}$, $l_{31} = \frac{1}{l_{11}} a_{31}$, $l_{32} = \frac{1}{l_{22}}(a_{32} - l_{31}l_{21})$

Cholesky decomposition

- What is the benefit of using matrix decomposition? The computations can be performed efficiently

  Example: Computing the determinant of a matrix

- **Differentiation of Univariate Functions**

$y = f(x), x, y \in \mathbb{R}$ | Function f of a scalar variable x

Definition : Difference Quotient

$$\frac{\delta y}{\delta x} := \frac{f(x + \delta x) - f(x)}{\delta x}$$

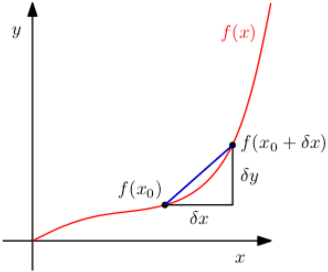

Figure: The average incline of a function f between x_0 and $x_0 + \delta x$

- **Derivative**

  For $h > 0$ the derivative of f at x is defined as the limit

$$\frac{\mathrm{d}f}{\mathrm{d}x} := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Matrix Calculus

- **Derivative**

  For $h > 0$ the derivative of f at x is defined as the limit

$$\frac{\mathrm{d}f}{\mathrm{d}x} := \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

- **Partial Differentiation and Gradients**

  $y = f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^{d_n}$ of $d_n n$ variables $x_1, x_2, \ldots x_{d_n}$

Function f depends on one or more variables $\mathbf{x}$

  Gradient:

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_{d_n}} \end{bmatrix} \in \mathbb{R}^n$$

  where e.g., $\frac{\partial f}{\partial x_1} = \lim_{h\to 0} \frac{f(x_1+h, x_2, \ldots, x_{d_n}) - f(x)}{h}$.

Matrix Calculus

- **Gradients of Vector-Valued Functions:** For a function $\mathbf{f}:\mathbb{R}^{d_n}\to\mathbb{R}^{d_m}$ and a vector $\mathbf{x}=\begin{bmatrix}x_1 & \dots & x_{d_n}\end{bmatrix}^\top$, the vector of value function is:

$$\mathbf{f}(\mathbf{x})=\begin{bmatrix}f_1(\mathbf{x})\\ \vdots\\ f_{d_m}(\mathbf{x})\end{bmatrix}\in\mathbb{R}^{d_m}$$

Jacobian: The matrix of all first-order partial derivatives of function $\mathbf{f}:\mathbb{R}^{d_n}\to\mathbb{R}^{d_m}$ is called Jacobian:

$$\mathbf{F}_{\mathbf{x}}(\mathbf{x})=\begin{bmatrix}\frac{\partial\mathbf{f}(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial\mathbf{f}(\mathbf{x})}{\partial x_{d_n}}\end{bmatrix}$$

$$=\begin{bmatrix}\frac{\partial f_1(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial f_1(\mathbf{x})}{\partial x_n}\\ \vdots & & \vdots\\ \frac{\partial f_{d_m}(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial f_m(\mathbf{x})}{\partial x_{d_n}}\end{bmatrix}\in\mathbb{R}^{d_m\times d_n}.$$

Linear and affine functions

- $f : \mathbb{R}^{d_n} \to \mathbb{R}$ means f is a function mapping d_n-dimensional vectors to numbers

Linear and affine functions

- $f : \mathbb{R}^{d_n} \to \mathbb{R}$ means f is a function mapping d_n-dimensional vectors to numbers
- f is linear if $f(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$

Linear and affine functions

- $f : \mathbb{R}^{d_n} \to \mathbb{R}$ means f is a function mapping d_n-dimensional vectors to numbers
- f is linear if $f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$
- when $\mathbf{a}$ is an d_n-dimensional vector, the function $f(\mathbf{x}) = \mathbf{a}^\top\mathbf{x} = a_1x_1 + a_2x_2 + \ldots + a_{d_n}x_{d_n}$ is the inner product function

Linear and affine functions

- $f : \mathbb{R}^{d_n} \to \mathbb{R}$ means f is a function mapping d_n-dimensional vectors to numbers
- f is linear if $f(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$
- when $\mathbf{a}$ is an d_n-dimensional vector, the function $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} = a_1 x_1 + a_2 x_2 + \ldots + a_{d_n} x_{d_n}$ is the inner product function
- a function that is linear plus a constant is called **affine**

Linear and affine functions

- $f : \mathbb{R}^{d_n} \to \mathbb{R}$ means f is a function mapping d_n-dimensional vectors to numbers
- f is linear if $f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$
- when $\mathbf{a}$ is an d_n-dimensional vector, the function $f(\mathbf{x}) = \mathbf{a}^\top\mathbf{x} = a_1x_1 + a_2x_2 + \ldots + a_{d_n}x_{d_n}$ is the inner product function
- a function that is linear plus a constant is called **affine**
- general form is $f(\mathbf{x}) = \mathbf{a}^\top\mathbf{x} + b$, with $\mathbf{a}$ an d_n-vector and b a scalar

Linear and affine functions

- $f : \mathbb{R}^{d_n} \to \mathbb{R}$ means f is a function mapping d_n-dimensional vectors to numbers
- f is linear if $f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$
- when $\mathbf{a}$ is an d_n-dimensional vector, the function $f(\mathbf{x}) = \mathbf{a}^\top\mathbf{x} = a_1x_1 + a_2x_2 + \ldots + a_{d_n}x_{d_n}$ is the inner product function
- a function that is linear plus a constant is called **affine**
- general form is $f(\mathbf{x}) = \mathbf{a}^\top\mathbf{x} + b$, with $\mathbf{a}$ an d_n-vector and b a scalar
- suppose $g : \mathbb{R}^{d_n} \to \mathbb{R}$ is a nonlinear function

Linear and affine functions

- $f : \mathbb{R}^{d_n} \to \mathbb{R}$ means f is a function mapping d_n-dimensional vectors to numbers
- f is linear if $f(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$
- when $\mathbf{a}$ is an d_n-dimensional vector, the function $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} = a_1 x_1 + a_2 x_2 + \ldots + a_{d_n} x_{d_n}$ is the inner product function
- a function that is linear plus a constant is called **affine**
- general form is $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$, with $\mathbf{a}$ an d_n-vector and b a scalar
- suppose $g : \mathbb{R}^{d_n} \to \mathbb{R}$ is a nonlinear function
- **First-order Taylor approximation** of g , near point $\hat{\mathbf{x}}$: $g(\mathbf{x}) \approx g(\hat{\mathbf{x}}) + \frac{\partial g(\hat{\mathbf{x}})}{\partial x_1}(x_1 - \hat{x}_1) + \ldots + \frac{\partial g(\hat{\mathbf{x}})}{\partial x_{d_n}}(x_{d_n} - \hat{x}_{d_n})$ **affine**

Linear and affine functions

- $f : \mathbb{R}^{d_n} \to \mathbb{R}$ means f is a function mapping d_n-dimensional vectors to numbers
- f is linear if $f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$
- when $\mathbf{a}$ is an d_n-dimensional vector, the function $f(\mathbf{x}) = \mathbf{a}^\top\mathbf{x} = a_1x_1 + a_2x_2 + \ldots + a_{d_n}x_{d_n}$ is the inner product function
- a function that is linear plus a constant is called **affine**
- general form is $f(\mathbf{x}) = \mathbf{a}^\top\mathbf{x} + b$, with $\mathbf{a}$ an d_n-vector and b a scalar
- suppose $g : \mathbb{R}^{d_n} \to \mathbb{R}$ is a nonlinear function
- **First-order Taylor approximation** of g , near point $\hat{\mathbf{x}}$:
  $g(\mathbf{x}) \approx g(\hat{\mathbf{x}}) + \frac{\partial g(\hat{\mathbf{x}})}{\partial x_1}(x_1 - \hat{x}_1) + \ldots + \frac{\partial g(\hat{\mathbf{x}})}{\partial x_{d_n}}(x_{d_n} - \hat{x}_{d_n})$ **affine**
- We can write using inner product as $g \approx g(\hat{\mathbf{x}}) + \nabla g(\hat{\mathbf{x}})^\top(\mathbf{x} - \hat{\mathbf{x}})$ where $\nabla g(\hat{\mathbf{x}}) = \begin{bmatrix} \frac{\partial g(\hat{\mathbf{x}})}{\partial x_1} & \ldots & \frac{\partial g(\hat{\mathbf{x}})}{\partial x_{d_n}} \end{bmatrix}$

References

[1] G. Strang, "Related resources," MIT OpenCourseWare. [Online]. Available: https://ocw.mit.edu/courses/18-06-linear-algebra-spring-2010/pages/related-resources/. [Accessed: 23-Aug-2022]

[2] S. Boy, and L. Vandenberghe, 2018. Introduction to applied linear algebra: vectors, matrices, and least squares. Cambridge university press.

[3] S. Boy, and L. Vandenberghe, Lecture slides for Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares.

[4] M.P. Deisenroth, A.A. Faisal, and C.S. Ong, 2020. Mathematics for machine learning. Cambridge University Press.