

向量空间

Vector Spaces

Baobin Li

Email:libb@ucas.ac.cn

School of Computer and Control Engineering, UCAS

Spaces and Subspaces

- Many mathematical entities that were considered to be quite different from matrices were in fact quite similar.
- For example, objects such as points in the plane \mathbb{R}^2 and \mathbb{R}^3 , polynomials, continuous functions, and differentiable functions satisfy the same additive properties and scalar multiplication properties given for matrices.
- Rather than studying each topic separately, it is more efficient and productive to study many topics at one time by studying the common properties that they satisfy.
- This eventually led to the axiomatic definition of a vector space.
- A vector space involves four things: two sets \mathcal{V} and \mathcal{F} , and two algebraic operations called vector addition and scalar multiplication. 标量乘

Vector Space Definition

The set \mathcal{V} is called a *vector space over \mathcal{F}* when the vector addition and scalar multiplication operations satisfy the following properties.

- (A1) $x+y \in \mathcal{V}$ for all $x, y \in \mathcal{V}$. This is called the *closure property for vector addition.* iP1.12
- (A2) $(x+y)+z = x+(y+z)$ for every $x, y, z \in \mathcal{V}$ ✓
- (A3) $x+y = y+x$ for every $x, y \in \mathcal{V}$. ✓
- (A4) There is an element $\mathbf{0} \in \mathcal{V}$ such that $x+\mathbf{0}=x$ for every $x \in \mathcal{V}$.
- (A5) For each $x \in \mathcal{V}$, there is an element $(-x) \in \mathcal{V}$ such that $x+(-x)=\mathbf{0}$. ax
- (M1) $\alpha x \in \mathcal{V}$ for all $\alpha \in \mathcal{F}$ and $x \in \mathcal{V}$. This is the *closure property for scalar multiplication.* ✓
- (M2) $(\alpha\beta)x = \alpha(\beta x)$ for all $\alpha, \beta \in \mathcal{F}$ and every $x \in \mathcal{V}$. ✓
- (M3) $\alpha(x+y) = \alpha x + \alpha y$ for every $\alpha \in \mathcal{F}$ and all $x, y \in \mathcal{V}$ ✓
- (M4) $(\alpha+\beta)x = \alpha x + \beta x$ for all $\alpha, \beta \in \mathcal{F}$ and every $x \in \mathcal{V}$.
- (M5) $1x = x$ for every $x \in \mathcal{V}$.

- The formal definition of a vector space stipulates how these four things relate to each other. *规定了这四者之间的关系*
- \mathcal{V} is a nonempty set of objects called vectors. Although \mathcal{V} can be quite general, we will usually consider \mathcal{V} to be a set of n-tuples or a set of matrices.
- \mathcal{F} is a scalar field for us \mathcal{F} is either the field \mathfrak{R} of real numbers or the field \mathcal{C} of complex numbers.
- Vector addition (denoted by $x + y$) is an operation between elements of \mathcal{V} .
- Scalar multiplication (denoted by αx) is an operation between elements of \mathcal{F} and \mathcal{V} .

Example 1

The set $\mathfrak{R}^{m \times n}$ of $m \times n$ real matrices is a vector space over \mathfrak{R} .

The set $\mathcal{C}^{m \times n}$ of $m \times n$ real matrices is a vector space over \mathcal{C} .

Example 2

The real coordinate spaces

$$\mathfrak{R}^{1 \times n} = \{(x_1, x_2, \dots, x_n), x_i \in \mathfrak{R}\}$$

$$\mathfrak{R}^{n \times 1} = \{(x_1, x_2, \dots, x_n)^T, x_i \in \mathfrak{R}\}$$

Example 3

With function addition and scalar multiplication defined by

$$(f + g)(x) = f(x) + g(x) \text{ and } (\alpha f)(x) = \alpha f(x),$$

the following sets are vector spaces over \mathfrak{R} :

1. *The set of functions mapping the interval $[0, 1]$ into \mathfrak{R} .*
2. *The set of all real-valued continuous functions defined on $[0, 1]$.*
3. *The set of real-valued functions that are differentiable on $[0, 1]$.*
4. *The set of all polynomials with real coefficients.*

Subspaces

Let \mathcal{S} be a nonempty subset of a vector space \mathcal{V} over \mathcal{F} (symbolically, $\mathcal{S} \subseteq \mathcal{V}$). If \mathcal{S} is also a vector space over \mathcal{F} using the same addition and scalar multiplication operations, then \mathcal{S} is said to be a **subspace** of \mathcal{V} . It's not necessary to check all 10 of the defining conditions in order to determine if a subset is also a subspace—only the closure conditions **(A1)** and **(M1)** need to be considered. That is, a nonempty subset \mathcal{S} of a vector space \mathcal{V} is a subspace of \mathcal{V} if and only if

$$\text{(A1)} \quad \mathbf{x}, \mathbf{y} \in \mathcal{S} \implies \mathbf{x} + \mathbf{y} \in \mathcal{S}$$

and

$$\text{(M1)} \quad \mathbf{x} \in \mathcal{S} \implies \alpha \mathbf{x} \in \mathcal{S} \text{ for all } \alpha \in \mathcal{F}.$$

Example 4

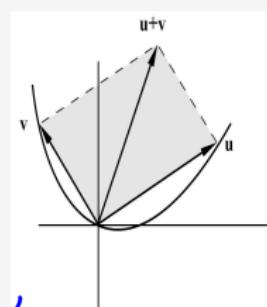
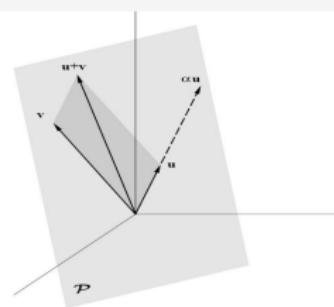
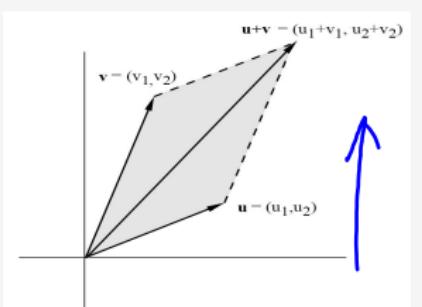
Given a vector space \mathcal{V} , the set $\mathcal{Z} = \{0\}$ containing only the zero vector is a subspace of \mathcal{V} . This subspace is called the trivial subspace.

Example 5

Straight lines through the origin in \mathbb{R}^2 and \mathbb{R}^3 are subspaces.

Example 6

In \mathbb{R}^3 , Planes through the origin are also subspaces.



Questions:

- ▶ What about straight lines not through the origin?
- ▶ What about curved lines through the origin?

- Visual interpretation: Subspaces are the flat surfaces passing through the origin.
- For a set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ from a vector space \mathcal{V} , the set of all possible linear combinations of the \mathbf{v}_i 's is denoted by

$$\text{span}(S) = \underbrace{\{\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_r \mathbf{v}_r \mid \alpha_i \in \mathbb{F}\}}_{\text{span}(S)}$$

- Notice that $\boxed{\text{span}(S)}$ is a subspace of \mathcal{V} .
- In fact, all subspaces of \mathbb{R}^n are of the type $\text{span}(S)$.
- If $\mathbf{u} \neq 0$ is a vector in \mathbb{R}^3 , then $\text{span}\{\mathbf{u}\}$ is the straight line passing through the origin and \mathbf{u} .
- If $S = \{\mathbf{u}, \mathbf{v}\}$, where \mathbf{u} and \mathbf{v} are nonzero vectors in \mathbb{R}^3 , $\text{span}(S)$ is the plane passing through the origin and the points \mathbf{u} and \mathbf{v} .

Spanning Sets

- For a set of vectors $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$, the subspace

$$\text{span}(\mathcal{S}) = \{\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_r \mathbf{v}_r\}$$

generated by forming all linear combinations of vectors from \mathcal{S} is called the space spanned by \mathcal{S} .

- If \mathcal{V} is a vector space such that $\mathcal{V} = \text{span}(\mathcal{S})$, we say \mathcal{S} is a *spanning set* for \mathcal{V} . In other words, \mathcal{S} spans \mathcal{V} whenever each vector in \mathcal{V} is a linear combination of vectors from \mathcal{S} .

Example 7

- The unit vectors $\{e_1, e_2, \dots, e_n\}$ form a spanning set for \mathbb{R}^n .
- The finite set $\{1, x, x^2, \dots, x^n\}$ spans the space of all polynomials such that $\deg p(x) \leq n$, and the infinite set $\{1, x, x^2, \dots\}$ spans the space of all polynomials.

Example 8

For a set of vectors $S = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ from a subspace $\mathcal{V} \subseteq \mathbb{R}^{m \times 1}$, let \mathbf{A} be the matrix containing the \mathbf{a}_i 's as its columns. S spans \mathcal{V} if and only if for each $\mathbf{b} \in \mathcal{V}$, there corresponds a column x such that $\mathbf{Ax} = \mathbf{b}$.

- This simple observation often is quite helpful. For example, to test whether or not $S = \{(1, 1, 1), (1, -1, -1), (3, 1, 1)\}$ spans \mathbb{R}^3 .
 - Just place these row as columns in a matrix \mathbf{A} .
 - Check "Is the system $\mathbf{Ax} = \mathbf{b}$ consistent for every $\mathbf{b} \in \mathbb{R}^3$?"

$$\begin{pmatrix} 1 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$



- As we know, $\mathbf{Ax} = \mathbf{b}$ is consistent if and only if $\text{rank}[\mathbf{A}|\mathbf{b}] = \text{rank}(\mathbf{A})$...
- In this case, $\text{rank}(\mathbf{A}) = 2$, but $\text{rank}[\mathbf{A}|\mathbf{b}] = 3$ for some \mathbf{b} (e.g., $b_1 = 0$, $b_2 = 1$, $b_3 = 0$), so S doesn't span \mathbb{R}^3 .

Sum of Subspaces

If \mathcal{X} and \mathcal{Y} are subspaces of a vector space \mathcal{V} , then the sum of \mathcal{X} and \mathcal{Y} is defined to be the set of all possible sums of vectors from \mathcal{X} with vectors from \mathcal{Y} . That is,

$$\mathcal{X} + \mathcal{Y} = \{x + y \mid x \in \mathcal{X} \text{ and } y \in \mathcal{Y}\}.$$

\Leftarrow $x \in \mathcal{X}, y \in \mathcal{Y}$

- The sum $\mathcal{X} + \mathcal{Y}$ is also a subspace of \mathcal{V} .
- If S_X, S_Y span \mathcal{X}, \mathcal{Y} , then $S_X \cup S_Y$ spans $\boxed{\mathcal{X} + \mathcal{Y}}$.

Example 9

If $\mathcal{X} \subseteq \mathbb{R}^2$ and $\mathcal{Y} \subseteq \mathbb{R}^2$ are subspaces defined by two different lines through the origin, then $\mathcal{X} + \mathcal{Y} = \mathbb{R}^2$. This follows from the parallelogram law.

Four Fundamental Subspaces

- Subspace are intimately related to linear functions as explain below.

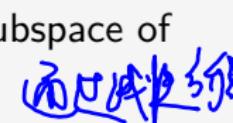
Subspaces and Linear Functions

Rf/1

For a linear function f mapping \mathbb{R}^n into \mathbb{R}^m , let $\mathcal{R}(f)$ denote the range of f . That is, $\mathcal{R}(f) = \{f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$ is the set of all “images” as \mathbf{x} varies freely over \mathbb{R}^n .

- The range of every linear function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a subspace of \mathbb{R}^m , and every subspace of \mathbb{R}^m is the range of some linear function.

For this reason, subspaces of \mathbb{R}^m are sometimes called linear spaces. 

- This result means that every matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ generates a subspace of \mathbb{R}^m by means of the range of the linear function $f(x) = \mathbf{A}x$. 

Proof. If $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear function, then the range of f is a subspace of \mathbb{R}^m because the closure properties **(A1)** and **(M1)** are satisfied. Establish **(A1)** by showing that $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{R}(f) \Rightarrow \mathbf{y}_1 + \mathbf{y}_2 \in \mathcal{R}(f)$. If $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{R}(f)$, then there must be vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ such that $\mathbf{y}_1 = f(\mathbf{x}_1)$ and $\mathbf{y}_2 = f(\mathbf{x}_2)$, so it follows from the linearity of f that

$$\mathbf{y}_1 + \mathbf{y}_2 = f(\mathbf{x}_1) + f(\mathbf{x}_2) = f(\mathbf{x}_1 + \mathbf{x}_2) \in \mathcal{R}(f).$$

Similarly, establish **(M1)** by showing that if $\mathbf{y} \in \mathcal{R}(f)$, then $\alpha\mathbf{y} \in \mathcal{R}(f)$ for all scalars α by using the definition of range along with the linearity of f to write

$$\mathbf{y} \in \mathcal{R}(f) \implies \mathbf{y} = f(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathbb{R}^n \implies \alpha\mathbf{y} = \alpha f(\mathbf{x}) = f(\alpha\mathbf{x}) \in \mathcal{R}(f).$$

Now prove that every subspace \mathcal{V} of \mathbb{R}^m is the range of some linear function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is a spanning set for \mathcal{V} so that

$$\mathcal{V} = \{\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n \mid \alpha_i \in \mathbb{R}\}.$$

Stack the \mathbf{v}_i 's as columns in a matrix $\mathbf{A}_{m \times n} = (\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_n)$, and put the α_i 's in an $n \times 1$ column $\mathbf{x} = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ to write

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n = (\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_n) \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \mathbf{Ax}.$$

The function $f(\mathbf{x}) = \mathbf{Ax}$ is linear and we have that

$$\mathcal{R}(f) = \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^{n \times 1}\} = \{\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n \mid \alpha_i \in \mathbb{R}\} = \mathcal{V}. \blacksquare$$

Range Spaces

- Likewise, the transpose of \mathbf{A} defines a subspace of \mathbb{R}^n by means of the range of the linear function $f(x) = \mathbf{A}^T x$.
- These two "range spaces" are two of the four fundamental subspace defined by a matrix.

范围空间

Range Spaces

The *range of a matrix* $\mathbf{A} \in \mathbb{R}^{m \times n}$ is defined to be the subspace $R(\mathbf{A})$ of \mathbb{R}^m that is generated by the range of $f(\mathbf{x}) = \mathbf{Ax}$. That is,

$$R(\mathbf{A}) = \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m.$$

Similarly, the range of \mathbf{A}^T is the subspace of \mathbb{R}^n defined by

$$R(\mathbf{A}^T) = \{\mathbf{A}^T \mathbf{y} \mid \mathbf{y} \in \mathbb{R}^m\} \subseteq \mathbb{R}^n$$

Because $R(\mathbf{A})$ is the set of all "images" of vectors $\mathbf{x} \in \mathbb{R}^m$ under transformation by \mathbf{A} , some people call $R(\mathbf{A})$ the image space of \mathbf{A} .

- As we know, that every matrixvector product $\mathbf{A}\mathbf{x}$ (i.e., every image) is a linear combination of the columns of \mathbf{A} provides a useful characterization of the range spaces.
- Therefore, $R(\mathbf{A})$ is nothing more than the space spanned by the columns of \mathbf{A} . $R(\mathbf{A})$ is often called the column space of \mathbf{A} .

Column and Row Spaces

For $\mathbf{A} \in \mathbb{R}^{m \times n}$, the following statements are true.

- $R(\mathbf{A}) =$ the space spanned by the columns of \mathbf{A} (column space).
- $R(\mathbf{A}^T) =$ the space spanned by the rows of \mathbf{A} (row space). ✓
- $\mathbf{b} \in R(\mathbf{A}) \iff \mathbf{b} = \mathbf{A}\mathbf{x}$ for some \mathbf{x} . ✓
- $\mathbf{a} \in R(\mathbf{A}^T) \iff \mathbf{a}^T = \mathbf{y}^T \mathbf{A}$ for some \mathbf{y}^T . ✓

Nullspace

By considering the linear functions $f(\mathbf{x}) = \mathbf{Ax}$ and $g(\mathbf{y}) = \mathbf{A}^T\mathbf{y}$, the other two fundamental subspaces defined by $\mathbf{A} \in \mathbb{R}^{m \times n}$ are obtained. They are $\mathcal{N}(f) = \{\mathbf{x}_{n \times 1} \mid \mathbf{Ax} = \mathbf{0}\} \subseteq \mathbb{R}^n$ and $\mathcal{N}(g) = \{\mathbf{y}_{m \times 1} \mid \mathbf{A}^T\mathbf{y} = \mathbf{0}\} \subseteq \mathbb{R}^m$.

Nullspace

- For an $m \times n$ matrix \mathbf{A} , the set $N(\mathbf{A}) = \{\mathbf{x}_{n \times 1} \mid \mathbf{Ax} = \mathbf{0}\} \subseteq \mathbb{R}^n$ is called the *nullspace* of \mathbf{A} . In other words, $N(\mathbf{A})$ is simply the set of all solutions to the homogeneous system $\mathbf{Ax} = \mathbf{0}$.
- The set $N(\mathbf{A}^T) = \{\mathbf{y}_{m \times 1} \mid \mathbf{A}^T\mathbf{y} = \mathbf{0}\} \subseteq \mathbb{R}^m$ is called the *left-hand nullspace* of \mathbf{A} because $N(\mathbf{A}^T)$ is the set of all solutions to the left-hand homogeneous system $\mathbf{y}^T \mathbf{A} = \mathbf{0}^T$.

$$(\mathbf{A}^T \mathbf{y})^T = \mathbf{0}^T$$

Summary



The four fundamental subspaces associated with $\mathbf{A}_{m \times n}$ are as follows.

- The range or column space: ✓ $R(\mathbf{A}) = \{\mathbf{Ax}\} \subseteq \mathbb{R}^m$.
- The row space or left-hand range: ✓ $R(\mathbf{A}^T) = \{\mathbf{A}^T \mathbf{y}\} \subseteq \mathbb{R}^n$.
- The nullspace: ✓ $N(\mathbf{A}) = \{\mathbf{x} \mid \mathbf{Ax} = \mathbf{0}\} \subseteq \mathbb{R}^n$.
- The left-hand nullspace: ✓ $N(\mathbf{A}^T) = \{\mathbf{y} \mid \mathbf{A}^T \mathbf{y} = \mathbf{0}\} \subseteq \mathbb{R}^m$.

Let \mathbf{P} be a nonsingular matrix such that $\mathbf{PA} = \mathbf{U}$, where \mathbf{U} is in row echelon form, and suppose $\underline{\text{rank } (\mathbf{A}) = r}$. ~~66~~ ~~1st 2nd~~

- Spanning set for $R(\mathbf{A})$ = the basic columns in \mathbf{A} .
- Spanning set for $R(\mathbf{A}^T)$ = the nonzero rows in \mathbf{U} . ~~2nd 0 0 0 0~~
- Spanning set for $N(\mathbf{A})$ = the \mathbf{h}_i 's in the general solution of $\underline{\mathbf{Ax} = 0}$.
- Spanning set for $N(\mathbf{A}^T)$ = the last $m - r$ rows of \mathbf{P} .

If \mathbf{A} and \mathbf{B} have the same shape, then ~~1st 2nd 3rd~~

- $\mathbf{A} \stackrel{\text{row}}{\sim} \mathbf{B} \iff N(\mathbf{A}) = N(\mathbf{B}) \iff R(\mathbf{A}^T) = R(\mathbf{B}^T)$
- $\mathbf{A} \stackrel{\text{col}}{\sim} \mathbf{B} \iff R(\mathbf{A}) = R(\mathbf{B}) \iff N(\mathbf{A}^T) = N(\mathbf{B}^T)$.

Linear Independence

- For a given set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, there may or may not exist dependency relationships in the sense that it may not be possible to express one vector as a linear combination of others.
- Consider two sets of vectors

$$\mathcal{A} = \left\{ \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 9 \\ -3 \\ 4 \end{pmatrix} \right\}, \quad \text{←}$$

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

- For the first set, $\mathbf{v}_3 = 3\mathbf{v}_1 + 2\mathbf{v}_2$, i.e., $3\mathbf{v}_1 + 2\mathbf{v}_2 - \mathbf{v}_3 = 0$.
- For the second set, there are no solutions for α_1, α_2 and α_3 in the homogeneous equation $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \alpha_3\mathbf{v}_3 = 0$, other than the trivial solution $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

Linear Independence

线性无关

A set of vectors $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is said to be a linearly independent set whenever the only solution for the scalars α_i in the homogeneous equation

$$\underbrace{\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n = \mathbf{0}}$$

is the trivial solution $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$.

若有非零解

- Whenever there is a nontrivial solution for α 's (i.e. at least one $\alpha_i \neq 0$), the set \mathcal{S} is said to be a Linearly dependent set.
- Linearly independent sets are those that contain no dependency relations.
- Linearly dependent sets are those in which at least one vector is a combination of the others.
- The empty set is always linearly independent.

- How to determine whether or not a set of vectors is linearly independent?
- Example

$$\Rightarrow \mathcal{S} = \left\{ \mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} \right\}.$$

- Solution: determine whether or not there exists a nontrivial solution for the following homogeneous equation:

$$\underbrace{\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \alpha_3 \mathbf{v}_3 = (0, 0, 0)^T,}_{\text{Ad.}}$$

- Equivalently, $\mathbf{A}\boldsymbol{\alpha} = \mathbf{0}$, where

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 5 \\ 2 & 0 & 6 \\ 1 & 2 & 7 \end{pmatrix}, \quad \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T, \quad \mathbf{0} = (0, 0, 0)^T$$

Linear Independence and Matrices

Let \mathbf{A} be an $m \times n$ matrix.

- Each of the following statements is equivalent to saying that the columns of \mathbf{A} form a linearly independent set.
 - ▷ $N(\mathbf{A}) = \{\mathbf{0}\}$.
 - ▷ $\text{rank}(\mathbf{A}) = n$.
- Each of the following statements is equivalent to saying that the rows of \mathbf{A} form a linearly independent set.
 - ▷ $N(\mathbf{A}^T) = \{\mathbf{0}\}$. ✓ A是-13行
 - ▷ $\text{rank}(\mathbf{A}) = m$. ✓
- When \mathbf{A} is a square matrix, each of the following statements is equivalent to saying that \mathbf{A} is nonsingular.
 - ▷ The columns of \mathbf{A} form a linearly independent set.
 - ▷ The rows of \mathbf{A} form a linearly independent set.

Special Types of Matrices

对角支配矩阵

■ Diagonally dominant matrices

- A matrix $A_{n \times n}$ is said to be diagonally dominant whenever

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}| \text{ for each } i = 1, 2, \dots, n.$$

- The magnitude of each diagonal entry exceeds the sum of the magnitudes of the off-diagonal entries in the corresponding row.
- Diagonally dominant matrices occur naturally in a wide variety of practical applications.
- In 1900, Minkowski discovered that all diagonally dominant matrices are nonsingular. ✓✓
- The strategy is to prove that if A is diagonally dominant, then $N(A) = \{0\}$. ✓

Vandermonde Matrices

- ▶ Matrices of the form

$$\mathbf{V}_{m \times n} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{pmatrix}$$

- ▶ This is named in honor of the French mathematician: **Alexandre Theophile Vandermonde (1735-1796)**.
- ▶ Columns constitute a linearly independent set whenever $n \leq m$.
- ▶ Problem: Given a set of m points $\mathcal{S} = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ in which the x_i 's are distinct, there is a unique polynomial.

$$l(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_{m-1} t^{m-1}$$

of degree $m - 1$ that passes through each point in \mathcal{S} .

- $l(t)$ must be given by

$$l(t) = \sum_{i=1}^m \left(y_i \frac{\prod_{j \neq i}^m (t - x_j)}{\prod_{j \neq i}^m (x_i - x_j)} \right).$$

拉格朗日插值法

- The polynomial $l(t)$ is known as the Lagrange interpolation polynomial of degree $m - 1$.
- If $\text{rank}(\mathbf{A}_{m \times n}) < n$, the columns of \mathbf{A} must be a dependent set.
- For such matrices, we often wish to extract a maximal linearly independent subset of columns.
- i.e., a linearly independent set containing as many columns from \mathbf{A} as possible.
- Although there can be several ways to make such a selection, the basic columns in \mathbf{A} always constitute one solution.

最大无关子集 (最大独立子集)

Maximal Independent Subsets

If $\text{rank}(\mathbf{A}_{m \times n}) = r$, then the following statements hold.

- Any maximal independent subset of columns from \mathbf{A} contains exactly r columns.
- Any maximal independent subset of rows from \mathbf{A} contains exactly r rows.
举反例
- In particular, the r basic columns in \mathbf{A} constitute one maximal independent subset of columns from \mathbf{A} .

Basic Facts of Independence

For a nonempty set of vectors $\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ in a space \mathcal{V} , the following statements are true.

- If \mathcal{S} contains a linearly dependent subset, then \mathcal{S} itself must be linearly dependent.
- If \mathcal{S} is linearly independent, then every subset of \mathcal{S} is also linearly independent.
- If \mathcal{S} is linearly independent and if $\mathbf{v} \in \mathcal{V}$, then the *extension set* $\mathcal{S}_{ext} = \mathcal{S} \cup \{\mathbf{v}\}$ is linearly independent if and only if $\mathbf{v} \notin \text{span}(\mathcal{S})$.
- If $\mathcal{S} \subseteq \mathbb{R}^m$ and if $n > m$, then \mathcal{S} must be linearly dependent.

■ Wronski matrix

- Let \mathcal{V} be the vector space of real-valued functions of a real variable, and let $\mathcal{S} = \{f_1(x), f_2(x), \dots, f_n(x)\}$ be a set of functions that are $n - 1$ times differentiable.
- The Wronski matrix is defined to be

$$\mathbf{W} = \begin{pmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f'_1(x) & f'_2(x) & \cdots & f'_n(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{pmatrix}$$



- If there is at least one point $x = x_0$ such that $\mathbf{w}(x_0)$ is nonsingular, \mathcal{S} must be a linearly independent set.
- For example, to verify that the set of polynomials $\mathcal{P} = 1, x, x^2, \dots, x^n$ is linearly independent, just observe that the associated Wronski matrix.

Basis and Dimension

- A linearly independent spanning set for a vector space \mathcal{V} is called a basis of \mathcal{V} .
 - It can be proven that every vector space \mathcal{V} possesses a basis.
 - Just as in the case of spanning sets, a space can possess many different bases.
- 单位向量
1. The unit vectors $S = \{e_1, e_2, \dots, e_n\}$ in \mathbb{R}^n are a basis for \mathbb{R}^n . This is called the standard basis for \mathbb{R}^n .
 2. The set $\{1, x, x^2, \dots, x^n\}$ is a basis for the vector space of polynomials having degree n or less.
 3. The infinite set $\{1, x, x^2, \dots\}$ is a basis for the vector space of all polynomials.
 4. If A is an $n \times n$ nonsingular matrix, then the set of rows in A as well as the set of columns from A constitute a basis for \mathbb{R}^n .
 5. For the trivial vector space $Z = \{0\}$, there is no nonempty linearly independent spanning set. Consequently, the empty set is considered to be a basis for Z .

- Spaces that possess a basis containing an infinite number of vectors are referred to as **infinite-dimensional spaces.** *无限维空间*
- Those that have a finite basis are called **finite-dimensional spaces.**

Characterizations of a Basis

Let \mathcal{V} be a subspace of \mathbb{R}^m , and let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\} \subseteq \mathcal{V}$. The following statements are equivalent.

- \mathcal{B} is a basis for \mathcal{V} .
- \mathcal{B} is a minimal spanning set for \mathcal{V} .
- \mathcal{B} is a maximal linearly independent subset of \mathcal{V} .

- Although a space \mathcal{V} can have many different bases, all bases for \mathcal{V} contain the same number of vectors.
- This number is quite important, which is called the dimension of \mathcal{V} .

Dimension

The **dimension** of a vector space \mathcal{V} is defined to be

$$\begin{aligned}\dim \mathcal{V} &= \text{number of vectors in any basis for } \mathcal{V} \\ &= \text{number of vectors in any minimal spanning set for } \mathcal{V} \\ &= \text{number of vectors in any maximal independent subset of } \mathcal{V}.\end{aligned}$$

■ Some examples:

1. $\dim \mathbb{R}^3 = 3$ because the three unit vectors $\{e_1, e_2, e_3\}$ constitute a basis for \mathbb{R}^3 .
2. If \mathcal{L} is a line through the origin in \mathbb{R}^3 , then $\dim \mathcal{L} = 1$ because a basis for \mathcal{L} consists of any nonzero vector lying along \mathcal{L} .
3. In \mathcal{P} is a plane through the origin in \mathbb{R}^3 , then $\dim \mathcal{P} = 2$ because a minimal spanning set for \mathcal{P} must contain two vectors from \mathcal{P} .
4. If $\mathcal{Z} = \{0\}$ is the trivial subspace, then $\dim \mathcal{Z} = 0$ because the basis for this space is the empty set.

■ In a loose sense the dimension of a space is a measure of the amount of "stuff" in the space

■ Dimension is in terms of degrees of freedom .

1. In the trivial space \mathcal{Z} , there are no degrees of freedom: you can move nowhere.
2. For a line, there is one degree of freedom: length.
3. In a plane there are two degrees of freedom: length and width.
4. In \mathbb{R}^3 there are three degrees of freedom: length, width and height.

- It is important not to confuse the dimension of a vector space \mathcal{V} with the number of components contained in the individual vector.
- For example, $\dim \mathcal{P} = 2$, but the individual vector in \mathcal{P} each have three components.
- There is the relationship between the dimension and the number of components contained in the individual vectors.

Subspace Dimension

For vector spaces \mathcal{M} and \mathcal{N} such that $\mathcal{M} \subseteq \mathcal{N}$, the following statements are true.

- $\dim \mathcal{M} \leq \dim \mathcal{N}$.
- If $\dim \mathcal{M} = \dim \mathcal{N}$, then $\mathcal{M} = \mathcal{N}$. ✓

Fundamental Subspaces—Dimension and Bases

For an $m \times n$ matrix of real numbers such that $\text{rank}(\mathbf{A}) = r$,

- $\dim \underline{R(\mathbf{A})} = r,$
- $\dim N(\mathbf{A}) = n - r,$
- $\dim R(\mathbf{A}^T) = r,$
- $\dim \underline{N(\mathbf{A}^T)} = m - r.$



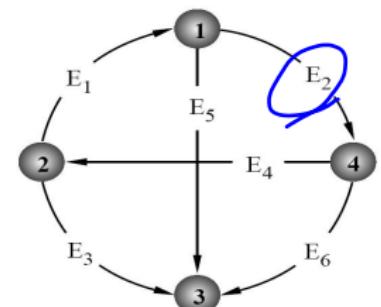
Let \mathbf{P} be a nonsingular matrix such that $\underline{\mathbf{PA}} = \mathbf{U}$ is in row echelon form, and let \mathcal{H} be the set of \mathbf{h}_i 's appearing in the general solution of $\mathbf{Ax} = \mathbf{0}$.

- The basic columns of \mathbf{A} form a basis for $\underline{R(\mathbf{A})}$.
- The nonzero rows of \mathbf{U} form a basis for $\underline{R(\mathbf{A}^T)}$.
- The set \mathcal{H} is a basis for $\underline{N(\mathbf{A})}$.
- The last $m - r$ rows of \mathbf{P} form a basis for $\underline{N(\mathbf{A}^T)}$.

For matrices with complex entries, the above statements remain valid provided that $\underline{\mathbf{A}^T}$ is replaced with $\underline{\mathbf{A}^*}$.

Rank and Connectivity

- A set of points (or nodes), $\{N_1, N_2, \dots, N_m\}$, together with a set of paths (or edges), $\{E_1, E_2, \dots, E_m\}$, between the nodes is called a **graph**.
- A **connected graph** is one in which there is a sequence of edges linking any pair of nodes.
- A **directed graph** is one in which each edge has been assigned a direction.
- There is a close relationship between the graph connectivity and matrix rank.
- The incidence matrix associated with a directed graph containing m nodes and n edges is defined to be the $m \times n$ matrix \mathbf{E} whose $e_{k,j}$ is
 - ▶ 1 if edge E_j is directed toward node N_k ;
 - ▶ -1 if edge E_j is directed away from node N_k ;
 - ▶ 0 if edge E_j neither begins nor ends at node N_k .



$$\mathbf{E} = \begin{pmatrix} E_1 & E_2 & E_3 & E_4 & E_5 & E_6 \\ N_1 & 1 & -1 & 0 & 0 & -1 & 0 \\ N_2 & -1 & 0 & -1 & 1 & 0 & 0 \\ N_3 & 0 & 0 & 1 & 0 & 1 & 1 \\ N_4 & 0 & 1 & 0 & -1 & 0 & -1 \end{pmatrix}.$$

- Each column in \mathbf{E} must contain exactly two nonzero entries: +1 and -1. Consequently, all column sums are zero. ✓.
- $\text{rank}(\mathbf{E}) = \text{rank}(\mathbf{E}^T) = m - \dim N(\mathbf{E}^T) \leq m - 1$.

Rank and Connectivity

Let \mathcal{G} be a graph containing m nodes. If \mathcal{G} is undirected, arbitrarily assign directions to the edges to make \mathcal{G} directed, and let \mathbf{E} be the corresponding incidence matrix.

- \mathcal{G} is connected if and only if $\text{rank}(\mathbf{E}) = m - 1$.

- If \mathcal{X} and \mathcal{Y} are subspaces of a vector space \mathcal{V} , then

$$\dim(\mathcal{X} + \mathcal{Y}) = \dim \mathcal{X} + \dim \mathcal{Y} - \dim(\mathcal{X} \cap \mathcal{Y}).$$

- $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$

Proof.

Observe that

$$\underline{R(\mathbf{A} + \mathbf{B}) \subseteq R(\mathbf{A}) + R(\mathbf{B})}. \checkmark$$

Recall that if $\mathcal{M} \subseteq \mathcal{N}$, then $\dim \mathcal{M} \leq \dim \mathcal{N}$, we have

待完成

$$\begin{aligned}\text{rank}(\mathbf{A} + \mathbf{B}) &= \dim R(\mathbf{A} + \mathbf{B}) \stackrel{\uparrow}{\leq} \dim(R(\mathbf{A}) + R(\mathbf{B})) \\ &= \dim R(\mathbf{A}) + \dim R(\mathbf{B}) - \dim(R(\mathbf{A}) \cap R(\mathbf{B})) \\ &\leq \dim R(\mathbf{A}) + \dim R(\mathbf{B}) = \underline{\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})}\end{aligned}$$



More About Rank

- Rank is invariant under multiplication by a nonsingular matrix.
- If P and Q are nonsingular matrices such that the product PAQ is defined, then
推导得证

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{PAQ}) = \text{rank}(\mathbf{PA}) = \text{rank}(\mathbf{AQ}). \Rightarrow$$

- However, multiplication by rectangular or singular matrices can alter the rank.

Rank of a Product

If \mathbf{A} is $m \times n$ and \mathbf{B} is $n \times p$, then



$$\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{B}) - \dim \underline{N(\mathbf{A}) \cap R(\mathbf{B})}.$$

- Although $\text{rank}(\mathbf{A})$ and $\text{rank}(\mathbf{B})$ are frequently or can be estimated, the term $\dim N(\mathbf{A}) \cap R(\mathbf{B})$ can be costly to obtain.
- Upper and lower bounds for $\text{rank}(\mathbf{AB})$ depends only on $\text{rank}(\mathbf{A})$ and $\text{rank}(\mathbf{B})$.

Bounds on the Rank of a Product

If \mathbf{A} is $m \times n$ and \mathbf{B} is $n \times p$, then

- $\text{rank}(\mathbf{AB}) \leq \min \{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\},$
- $\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - n \leq \text{rank}(\mathbf{AB}) \checkmark$

- The products $\mathbf{A}^T \mathbf{A}$ and $\mathbf{A} \mathbf{A}^T$ and their complex counterparts $\mathbf{A}^* \mathbf{A}$ and $\mathbf{A} \mathbf{A}^*$ deserve special attention because they naturally appear in a wide variety of applications.

Products $\mathbf{A}^T \mathbf{A}$ and $\mathbf{A} \mathbf{A}^T$

For $\mathbf{A} \in \mathbb{R}^{m \times n}$, the following statements are true.

- $\text{rank}(\mathbf{A}^T \mathbf{A}) = \text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A} \mathbf{A}^T)$. ✓ $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$
- $R(\mathbf{A}^T \mathbf{A}) = R(\mathbf{A}^T)$ and $R(\mathbf{A} \mathbf{A}^T) = R(\mathbf{A})$.
- $N(\mathbf{A}^T \mathbf{A}) = N(\mathbf{A})$ and $N(\mathbf{A} \mathbf{A}^T) = N(\mathbf{A}^T)$.

For $\mathbf{A} \in \mathbb{C}^{m \times n}$, the transpose operation $(\star)^T$ must be replaced by the conjugate transpose operation $(\star)^*$.

\rightarrow 表示转置

- Consider an $m \times n$ system of equations $\underline{\mathbf{A}x = b}$ that may or may not be consistent.
- Multiplying on the left-hand side by \mathbf{A}^T produces the $n \times n$ system $\mathbf{A}^T \mathbf{A}x = \mathbf{A}^T b$ called the associated system of normal equations.
- First, the normal equations are always consistent, regardless of whether or not the original system is consistent because $\mathbf{A}^T b \in R(\mathbf{A}^T) = R(\mathbf{A}^T \mathbf{A})$. ✓

- If $\underline{\mathbf{Ax} = b}$ happens to be consistent, then $\mathbf{Ax} = b$ and $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T b$ have the same solution set.
- If $\underline{\mathbf{Ax} = b}$ is consistent and has a unique solution, the same is true for $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T b$, and the unique solution common to both system is

$$\boxed{x = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T b.}$$

(Left side) *(Right side)*

- There is one outstanding question: what do the solutions of the normal equations $\underline{\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T b}$ represent when the original system $\underline{\mathbf{Ax} = b}$ is not consistent?
- Use of the product $\underline{\mathbf{A}^T \mathbf{A}}$ or the normal equations is not recommended for numerical computation.
- Any sensitivity to small perturbations that is present in the underlying matrix \mathbf{A} is magnified by forming the product $\underline{\mathbf{A}^T \mathbf{A}}$.
- Nevertheless, the normal equations are an important theoretical idea that leads to practical tools of fundamental importance such as the method least squares.

Normal Equations

- For an $m \times n$ system $\mathbf{Ax} = \mathbf{b}$, the associated system of *normal equations* is defined to be the $n \times n$ system $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$.
- $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$ is always consistent, even when $\mathbf{Ax} = \mathbf{b}$ is not consistent.
- When $\mathbf{Ax} = \mathbf{b}$ is consistent, its solution set agrees with that of $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$. the normal equations provide least squares solutions to $\mathbf{Ax} = \mathbf{b}$ when $\mathbf{Ax} = \mathbf{b}$ is inconsistent.
- $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$ has a unique solution if and only if $\text{rank}(\mathbf{A}) = n$, in which case the unique solution is $\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$.
- When $\mathbf{Ax} = \mathbf{b}$ is consistent and has a unique solution, then the same is true for $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$, and the unique solution to both systems is given by $\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$.

- One more way to think about rank:

最大非零子矩阵

Rank and the Largest Nonsingular Submatrix

The rank of a matrix $A_{m \times n}$ is precisely the order of a maximal square nonsingular submatrix of A . In other words, to say $\underline{\text{rank}}(A) = r$ means that there is at least one $r \times r$ nonsingular submatrix in A , and there are no nonsingular submatrices of larger order.

- It is impossible to increase the rank by means of matrix multiplication: $\underline{\text{rank}}(AE) \leq \underline{\text{rank}}(A)$.
- In a certain sense there is a dual statement for matrix addition that says that it is impossible to decrease the rank by means of a "small" matrix addition: $\underline{\text{rank}}(A + E) \geq \underline{\text{rank}}(A)$ whenever E has entries of small magnitude.

Summary of Rank



For $\mathbf{A} \in \mathbb{R}^{m \times n}$, each of the following statements is true.

- $\text{rank}(\mathbf{A}) =$ The number of nonzero rows in any row echelon form that is row equivalent to \mathbf{A} .
- $\text{rank}(\mathbf{A}) =$ The number of pivots obtained in reducing \mathbf{A} to a row echelon form with row operations.
- $\text{rank}(\mathbf{A}) =$ The number of basic columns in \mathbf{A} (as well as the number of basic columns in any matrix that is row equivalent to \mathbf{A}).
- $\text{rank}(\mathbf{A}) =$ The number of independent columns in \mathbf{A} —i.e., the size of a maximal independent set of columns from \mathbf{A} .
- $\text{rank}(\mathbf{A}) =$ The number of independent rows in \mathbf{A} —i.e., the size of a maximal independent set of rows from \mathbf{A} .
- $\text{rank}(\mathbf{A}) = \dim R(\mathbf{A})$.
- $\text{rank}(\mathbf{A}) = \dim R(\mathbf{A}^T)$.
- $\text{rank}(\mathbf{A}) = n - \dim N(\mathbf{A})$.
- $\text{rank}(\mathbf{A}) = m - \dim N(\mathbf{A}^T)$.
- $\text{rank}(\mathbf{A}) =$ The size of the largest nonsingular submatrix in \mathbf{A} .

For $\mathbf{A} \in \mathcal{C}^{m \times n}$, replace $(\star)^T$ with $(\star)^*$.

Classical Least Squares

~~Topics~~

- At discrete points t_i (often points in time), observations b_i of some phenomenon are made, and the results are records as a set of ordered pairs:

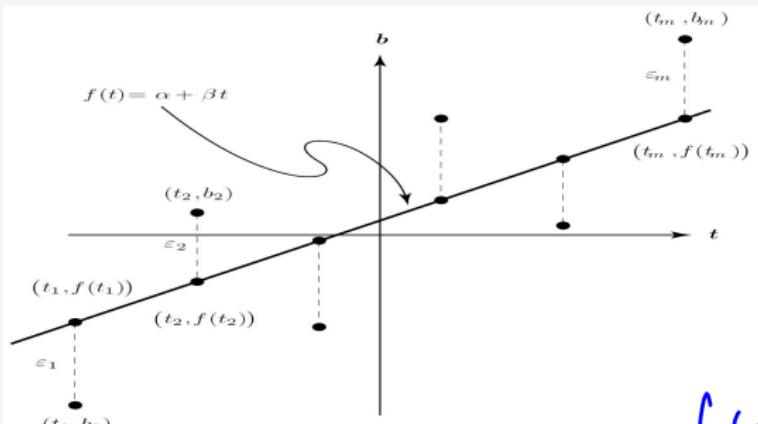
$$\mathcal{D} = \{(t_1, b_1), (t_2, b_2), \dots, (t_m, b_m)\}.$$

~~What's this~~

- The problem is to make estimations or predictions at points \hat{t} that are between or beyond the observation points t_i .
- A standard approach is to find the equation of a curve $y = f(t)$ that closely fits the points in \mathcal{D} .
- The phenomenon can be estimated at any nonobservation point \hat{t} with the value $\hat{y} = f(\hat{t})$.
- Let's begin by fitting a straight line to the points in \mathcal{D} .

~~最小二乘法~~

$$y = f(t)$$



$$f(t) = \alpha + \beta t -$$

- The strategy is to determine the coefficients α and β in the equation of the line $f(t) = \alpha + \beta t$ that best fits the points (t_i, b_i) in the sense that the sum of the squares of the vertical errors $\epsilon_1, \epsilon_2, \dots, \epsilon_m$ is minimal.
- The distance from (t_i, b_i) to a line $f(t) = \alpha + \beta t$ is

$$\epsilon_i = |f(t_i) - b_i| = |\alpha + \beta t_i - b_i|$$

- The objective is to find values for α and β such that

$$\sum_{i=1}^m \varepsilon_i^2 = \sum_{i=1}^m (\alpha + \beta t_i - b_i)^2 \text{ is minimal.}$$

- Minimization techniques tell us

$$\sum_{i=1}^m (\alpha + \beta t_i - b_i) = 0, \quad \sum_{i=1}^m (\alpha + \beta t_i - b_i) t_i = 0.$$



- It is easy to get the matrix form $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$, where setting

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{pmatrix}^T, \quad \mathbf{b} = (b_1, b_2, \dots, b_m)^T, \quad \mathbf{x} = (\alpha, \beta)^T.$$

- This is the system of normal equations associated with the $\mathbf{Ax} = \mathbf{b}$.

- The t_i 's are assumed to be distinct numbers, so $\text{rank}(\mathbf{A}) = 2$.
- The normal equations have a unique solution given by

$$\begin{aligned}x &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T b \\&= \frac{1}{m \sum t_i^2 - (\sum t_i)^2} \begin{pmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & m \end{pmatrix} \begin{pmatrix} \sum b_i \\ \sum t_i b_i \end{pmatrix} \\&= \frac{1}{m \sum t_i^2 - (\sum t_i)^2} \begin{pmatrix} \sum t_i^2 \sum b_i - \sum t_i \sum t_i b_i \\ m \sum t_i b_i - \sum t_i \sum b_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}\end{aligned}$$

- Finally, notice that the total sum of squares of the errors is given by

$$\sum_{i=1}^m \varepsilon_i^2 = (\mathbf{A}x - b)^T (\mathbf{A}x - b).$$

General Least Squares Problem

-般最小二乘法

For $\mathbf{A} \in \Re^{m \times n}$ and $\mathbf{b} \in \Re^m$, let $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\mathbf{x}) = \mathbf{Ax} - \mathbf{b}$. The general least squares problem is to find a vector \mathbf{x} that minimizes the quantity

$$\sum_{i=1}^m \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{Ax} - \mathbf{b})^T (\mathbf{Ax} - \mathbf{b}).$$

Any vector that provides a minimum value for this expression is called a *least squares solution*.

- The set of all least squares solutions is precisely the set of solutions to the system of normal equations $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$.
- There is a unique least squares solution if and only if $\text{rank}(\mathbf{A}) = n$, in which case it is given by $\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$.
- If $\mathbf{Ax} = \mathbf{b}$ is consistent, then the solution set for $\mathbf{Ax} = \mathbf{b}$ is the same as the set of least squares solutions.

- The classical least squares problem is part of a broader topic known as linear regression. 
- It is the study of situations where attempts are made to express y as a linear combination of other variables t_1, t_2, \dots, t_n .
- That means that $y = \alpha_0 + \alpha_1 t_1 + \alpha_2 t_2 + \dots + \alpha_n t_n + \varepsilon$.
- ε is a random function whose values average out to zero in some sense.
- In other words, a linear hypothesis is the supposition that the expected (or mean) value of y at each point where the phenomenon can be observed is given by a linear equation:

$$E(y) = \alpha_0 + \alpha_1 t_1 + \alpha_2 t_2 + \dots + \alpha_n t_n.$$
 

- Example: the problem of predicting the amount of weight that a pint of ice cream loses when it is stored at very low temperatures.

- There are many factors that may contribute to weight loss.
- It is reasonable to believe that storage time and temperature are the primary factors, so we will make a linear hypothesis of the form

$$\underline{y = \alpha_0 + \alpha_1 t_1 + \alpha_2 t_2 + \varepsilon}. \checkmark$$

- y = weight loss, t_1 = storage time, t_2 = storage temperature.
- ε is a random function to account for all other factors.
- The expected weight loss is $E(y) = \underline{\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2}$.
- Suppose that we conduct an experiment as shown below.

Time (weeks)	1	1	1	2	2	2	3	3	3
Temp ($^{\circ}F$)	-10	-5	0	-10	-5	0	-10	-5	0
Loss (grams)	.15	.18	.20	.17	.19	.22	.20	.23	.25

If

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & -10 \\ 1 & 1 & -5 \\ 1 & 1 & 0 \\ 1 & 2 & -10 \\ 1 & 2 & -5 \\ 1 & 2 & 0 \\ 1 & 3 & -10 \\ 1 & 3 & -5 \\ 1 & 3 & 0 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} .15 \\ .18 \\ .20 \\ .17 \\ .19 \\ .22 \\ .20 \\ .23 \\ .25 \end{pmatrix},$$

A x = b

- If $b_i = E(y_i)$, $\mathbf{Ax} = \mathbf{b}$ is a consistent system, so we could solve for the unknown parameters α_0 , α_1 and α_2 .
- However, it is impossible to observe the exact value of the mean weight loss for a given storage time and temperature.
- The system $\mathbf{Ax} = \mathbf{b}$ will be inconsistent-especially when the number of observations greatly exceeds the number of parameters.
- Since we can't find exact values for α_i 's, we hope for getting a set of good estimates for these parameters.
- The famous Gauss-Markov theorem states that the least squares estimates are the best way to estimate the α_i 's.

Returning to our ice cream example, it can be verified that $\mathbf{b} \notin R(\mathbf{A})$, so, as expected, the system $\mathbf{Ax} = \mathbf{b}$ is not consistent, and we cannot determine exact values for α_0, α_1 , and α_2 . The best we can do is to determine least squares estimates for the α_i 's by solving the associated normal equations $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$, which in this example are

$$\begin{pmatrix} 9 & 18 & -45 \\ 18 & 42 & -90 \\ -45 & -90 & 375 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1.79 \\ 3.73 \\ -8.2 \end{pmatrix}.$$

$$\cancel{\mathbf{A}^T \mathbf{A}} \left(\begin{matrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{matrix} \right) = \cancel{\mathbf{A}^T \mathbf{b}}$$

The solution is

$$\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} .174 \\ .025 \\ .005 \end{pmatrix},$$

$$\checkmark \quad \cancel{(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}}$$

and the estimating equation for mean weight loss becomes

$$\hat{y} = .174 + .025t_1 + .005t_2.$$

For example, the mean weight loss of a pint of ice cream that is stored for nine weeks at a temperature of $-35^\circ F$ is estimated to be

$$\hat{y} = .174 + .025(9) + .005(-35) = .224 \text{ grams.}$$

Least Squares Curve Fitting Problem

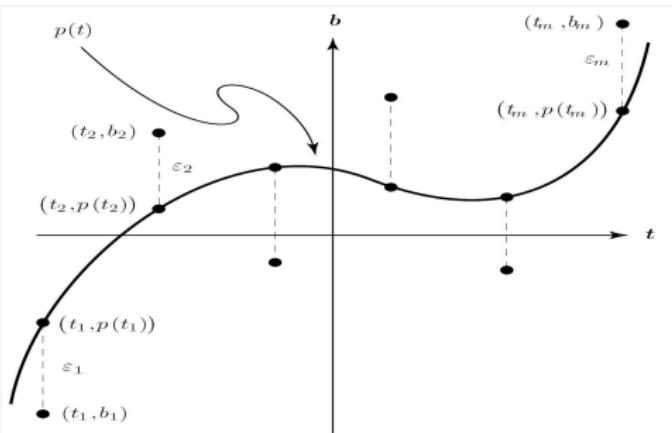
Least Squares Curve Fitting Problem: Find a polynomial

$$p(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_{n-1} t^{n-1}$$

with a specified degree that comes as close as possible in the sense of least squares to passing through a set of data points

$$\mathcal{D} = \{(t_1, b_1), (t_2, b_2), \dots, (t_m, b_m)\},$$

where the t_i 's are distinct numbers, and $n \leq m$. ✓.



- For the ε_i 's indicated in above figure, the objective is to minimize the sum of squares

$$\sum_{i=1}^m \varepsilon_i^2 = \sum_{i=1}^m (p(t_i) - b_i)^2 = (\mathbf{Ax} - \mathbf{b})^T (\mathbf{Ax} - \mathbf{b}),$$

where

$$\mathbf{A} = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_m & t_m^2 & \cdots & t_m^{n-1} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{n-1} \end{pmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

- The least squares polynomial of degree $n - 1$ is obtained from the least squares solution associated with the system $\mathbf{Ax} = \mathbf{b}$.
- Furthermore, this least squares polynomial is unique because \mathbf{A} is the Vandermonde matrix with $n \leq m$, so $\text{rank}(\mathbf{A}) = n$.
- $\mathbf{Ax} = \mathbf{b}$ has a unique least squares solution give by $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$.

Exercise

1. Which of the following are spanning sets for \mathbb{R}^3 ?

- (a) $\{(1, 1, 1)\}$
- (b) $\{(1, 0, 0), (0, 0, 1)\}$
- (c) $\{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$
- (d) $\{(1, 2, 1), (2, 0, -1), (4, 4, 1)\}$

2. Let \mathcal{X} and \mathcal{Y} be two subspaces of a vector space \mathcal{V} . Prove that the intersection $\mathcal{X} \cap \mathcal{Y}$ is also a subspace of \mathcal{V} .

3. If $A = \begin{pmatrix} -1 & 1 & 1 & -2 & 1 \\ -1 & 0 & 3 & -4 & 2 \\ -1 & 0 & 3 & -5 & 3 \\ -1 & 0 & 3 & -6 & 4 \\ -1 & 0 & 3 & -6 & 2 \end{pmatrix}$ and $b = (-2, -5, -6, -7, -7)^T$, is $b \in R(A)$?

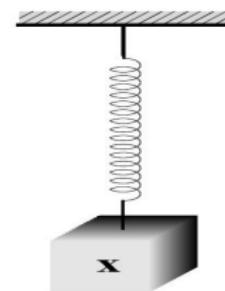
4. Which of the following sets of functions are linearly independent?
- (a) $\{\sin x, \cos x, x \sin x\}$.
 - (b) $\{e^x, xe^x, x^2 e^x\}$.
5. Determine the dimensions of each of the following vector spaces:
- (a) The space of polynomials having degree n or less.
 - (b) The space $\Re^{m \times n}$ of $m \times n$ matrices.
6. verify that $\text{rank}(\mathbf{A}^T \mathbf{A}) = \text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A} \mathbf{A}^T)$ for

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 1 & -4 \\ -1 & -3 & 1 & 0 \\ 2 & 6 & 2 & -8 \end{pmatrix}.$$

7. For $\mathbf{A} \in \Re^{m \times n}$, explain why $\mathbf{A}^T \mathbf{A} = 0$ implies $\mathbf{A} = 0$.

8. Hooke's law says that the displacement y of an ideal spring is proportional to the force x that is applied i.e., $y = kx$ for some constant k . Consider a spring in which k is unknown. Various masses are attached, and the resulting displacements shown in the following figure are observed. Using these observations, determine the least squares estimate for k .

x (lb)	y (in)
5	11.1
7	15.4
8	17.5
10	22.0
12	26.3



9. Using least squares techniques, fit the following data

x	-5	-4	-3	-2	-1	0	1	2	3	4	5
y	2	7	9	12	13	14	14	13	10	8	4

with a line $y = \alpha_0 + \alpha_1x$ and then fit the data with a quadratic $y = \alpha_0 + \alpha_1x + \alpha_2x^2$. Determine which of these two curves best fits the data by computing the sum of the squares of the errors in each case.