5.1 子空间与生成 (Subspace and Spanning)

5.1.1 $\mathbb{R}^n$ 的子空间

Definition 5.1 Subspace of $\mathbb{R}^n$

A set[1] U of vectors in $\mathbb{R}^n$ is called a **subspace** *of $\mathbb{R}^n$ if it satisfies the following properties:*

S1. The zero vector $\mathbf{0} \in U$.

S2. If $\mathbf{x} \in U$ and $\mathbf{y} \in U$, then $\mathbf{x}+\mathbf{y} \in U$.

S3. If $\mathbf{x} \in U$, then $a\mathbf{x} \in U$ for every real number a.

S2. 对加法封闭

S3. 对数乘封闭

{0}: 零空间 (zero space)

除 {0} 和 $\mathbb{R}^n$ 的 $\mathbb{R}^n$ 子空间: 真子空间 (proper subspace)

三维空间内过原点的直线和平面都是 $\mathbb{R}^3$ 的子空间

子空间也被用来描述矩阵的重要性质.

A的零空间(null space) $= \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = 0\}$

记作 null A

A的像空间(image space) $= \{A\vec{x} \mid \vec{x} \in \mathbb{R}^n\}$

记作 im A

$$E_\lambda(A) = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \lambda\vec{x}\}$$

其中A为 $n\times n$ 方阵, λ 为一个数

注: 并不是所有 $\mathbb{R}^n$ 的子集都是子空间.

如: $U_1 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| x \geq 0 \right\}$

不满足 S_3

$U_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| x^2 = y^2 \right\}$

不满足 S_2

5.1.2 生成集合（Spanning Sets）

Definition 5.2 Linear Combinations and Span in $\mathbb{R}^n$

The set of all such linear combinations is called the **span** *of the* $\mathbf{x}_i$ *and is denoted*

$$\operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\} = \{t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k \mid t_i \text{ in } \mathbb{R}\}$$

If $V = \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$, *we say that* V *is* **spanned** *by the vectors* $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$, *and that the vectors* $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ **span** *the space* V.

Theorem 5.1.1: Span Theorem

Let $U = \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ *in* $\mathbb{R}^n$. *Then:*

1. U *is a subspace of* $\mathbb{R}^n$ *containing each* $\mathbf{x}_i$.
2. *If* W *is a subspace of* $\mathbb{R}^n$ *and each* $\mathbf{x}_i \in W$, *then* $U \subseteq W$.

$n \times n$ 单位矩阵的第 j 列被称为第 j 个坐标向量（jth coordinate vector），记作 $\vec{e}_j$．集合 $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$ 记作 $\mathbb{R}^n$ 的标准基

5.2 独立性和维度（Independence and Dimension）

我们现在探究仅有一种向量的线性组合的表示方法的集合

5.2.1 线性无关（Linear Independence）

Definition 5.3 Linear Independence in $\mathbb{R}^n$

With this in mind, we call a set $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ of vectors ***linearly independent*** (or simply ***independent***) if it satisfies the following condition:

$$\text{If } t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k = \mathbf{0} \text{ then } t_1 = t_2 = \cdots = t_k = 0$$

零空间！

由 $s_1\vec{x}_1 + s_2\vec{x}_2 + \cdots + s_t\vec{x}_t = r_1\vec{x}_1 + \cdots + r_t\vec{x}_t$

$\Rightarrow (s_1 - r_1)\vec{x}_1 + (s_2 - r_2)\vec{x}_2 + \cdots + (s_t - r_t)\vec{x}_t = \vec{0}$

推出.

记 $t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k = \vec{0}$ 为线性组合成为零（vanish）

推出 $t_1 = t_2 = \cdots = t_k = 0$ 为平凡（trivial）

A set of vectors is linear independent if and only if the only linear combination that vanishes is trivial.

Independence Test

To verify that a set $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ of vectors in $\mathbb{R}^n$ is independent, proceed as follows:

1. *Set a linear combination equal to zero: $t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k = \mathbf{0}$.*
2. *Show that $t_i = 0$ for each i (that is, the linear combination is trivial).*

Of course, if some nontrivial linear combination vanishes, the vectors are not independent.

例、证明$\{\vec{v}, \vec{w}\}$是$\mathbb{R}^3$内相关(dependent)的，当且仅当$\vec{v}$. $\vec{w}$共线.

证：若$\vec{v}$, $\vec{w}$共线. 记$\vec{v} = a\vec{w}$,

有 $\vec{v} - \vec{v} = \vec{v} - a\vec{w} = \vec{0}$

$\Rightarrow$ Vanish问题有nontrivial solution

$\Rightarrow$ 相关.

反过来，记$p\vec{v} + q\vec{w} = \vec{0}$. p. q均不为0. 则有$\vec{v} = -\dfrac{q}{p}\vec{w}$

$\Leftrightarrow$ $\vec{v}$、$\vec{w}$ 共线.

得证.

Theorem 5.2.2

If A is an $m \times n$ matrix, let $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ denote the columns of A.

1. *$\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ is independent in $\mathbb{R}^m$ if and only if $A\mathbf{x} = \mathbf{0}$, $\mathbf{x}$ in $\mathbb{R}^n$, implies $\mathbf{x} = \mathbf{0}$.*
2. *$\mathbb{R}^m = \operatorname{span}\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ if and only if $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$ for every vector $\mathbf{b}$ in $\mathbb{R}^m$.*

$\Rightarrow$ 很难不联想到矩阵的可逆性质.

Theorem 5.2.3

The following are equivalent for an $n \times n$ matrix A:

1. *A is invertible.*
2. *The columns of A are linearly independent.*
3. *The columns of A span $\mathbb{R}^n$.*
4. *The rows of A are linearly independent.*
5. *The rows of A span the set of all $1 \times n$ rows.*

5.2.2 维度 (Dimension)

Theorem 5.2.4: Fundamental Theorem

Let U be a subspace of $\mathbb{R}^n$. If U is spanned by m vectors, and if U contains k linearly independent vectors, then $k \leq m$.

Definition 5.4 Basis of $\mathbb{R}^n$

*If U is a subspace of $\mathbb{R}^n$, a set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ of vectors in U is called a **basis** of U if it satisfies the following two conditions:*

1. $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ *is linearly independent.*
2. $U = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$.

The most remarkable result about bases[7] is:

Theorem 5.2.5: Invariance Theorem

If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$ are bases of a subspace U of $\mathbb{R}^n$, then $m = k$.

Proof. We have $k \leq m$ by the fundamental theorem because $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ spans U, and $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$ is independent. Similarly, by interchanging $\mathbf{x}$'s and $\mathbf{y}$'s we get $m \leq k$. Hence $m = k$. $\square$

The invariance theorem guarantees that there is no ambiguity in the following definition:

Definition 5.5 Dimension of a Subspace of $\mathbb{R}^n$

*If U is a subspace of $\mathbb{R}^n$ and $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ is any basis of U, the number, m, of vectors in the basis is called the **dimension** of U, denoted*

$$\dim U = m$$

Theorem 5.2.6

Let $U \neq \{\mathbf{0}\}$ be a subspace of $\mathbb{R}^n$. Then:

1. *U has a basis and $\dim U \leq n$.*
2. *Any independent set in U can be enlarged (by adding vectors from the standard basis) to a basis of U.*
3. *Any spanning set for U can be cut down (by deleting vectors) to a basis of U.*

Theorem 5.2.7

Let U be a subspace of $\mathbb{R}^n$ where $\dim U = m$ and let $B = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ be a set of m vectors in U. Then B is independent if and only if B spans U.

Theorem 5.2.8

Let $U \subseteq W$ be subspaces of $\mathbb{R}^n$. Then:

1. $\dim U \leq \dim W$.
2. *If $\dim U = \dim W$, then $U = W$.*

5.3 正交 (Orthogonality)

拓展点乘到n维向量，赋予n维空间欧氏几何

5.3.1 点乘、长度和距离.

$$\vec{x} = (x_1, x_2, x_3, \cdots, x_n)$$

$$\vec{y} = (y_1, y_2, \cdots, y_n)$$

$$\vec{x} \cdot \vec{y} = (\vec{x}^T)\vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

Definition 5.6 Length in $\mathbb{R}^n$

As in $\mathbb{R}^3$, the ***length*** $\|\mathbf{x}\|$ *of the vector is defined by*

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

Where $\sqrt{(\ \ )}$ indicates the positive square root.

Theorem 5.3.1

Let $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ denote vectors in $\mathbb{R}^n$. Then:

1. $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.
2. $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$.
3. $(a\mathbf{x}) \cdot \mathbf{y} = a(\mathbf{x} \cdot \mathbf{y}) = \mathbf{x} \cdot (a\mathbf{y})$ for all scalars a.

4. $\|\mathbf{x}\|^2 = \mathbf{x} \cdot \mathbf{x}$.

5. $\|\mathbf{x}\| \geq 0$, and $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.

6. $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$ *for all scalars* a.

由任意向量的夹角 θ 之 $|\cos\theta| \leq 1$，我们有

Theorem 5.3.2: Cauchy Inequality[9]

If $\mathbf{x}$ and $\mathbf{y}$ *are vectors in* $\mathbb{R}^n$, *then*

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$$

Moreover $|\mathbf{x} \cdot \mathbf{y}| = \|\mathbf{x}\|\|\mathbf{y}\|$ *if and only if one of* $\mathbf{x}$ and $\mathbf{y}$ *is a multiple of the other.*

↳ 柯西不等式

对于 $\mathbb{R}^n$，我们有

$$(x_1y_1 + x_2y_2 + \cdots + x_ny_n)^2 \leq (x_1^2 + x_2^2 + \cdots + x_n^2) \cdot (y_1^2 + y_2^2 + \cdots + y_n^2)$$

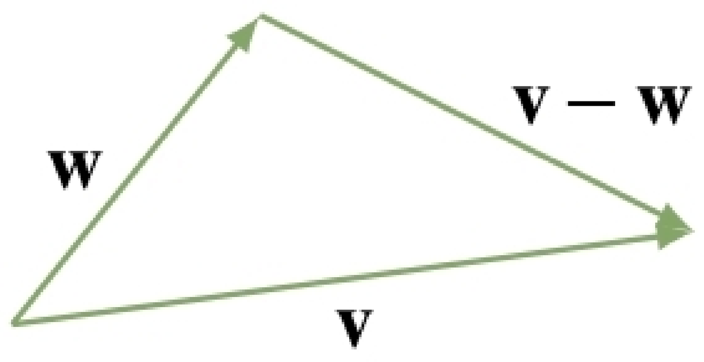
Corollary 5.3.1: Triangle Inequality

If $\mathbf{x}$ and $\mathbf{y}$ *are vectors in* $\mathbb{R}^n$, *then* $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Definition 5.7 Distance in $\mathbb{R}^n$

*If $\mathbf{x}$ and $\mathbf{y}$ are two vectors in $\mathbb{R}^n$, we define the **distance** $d(\mathbf{x},\ \mathbf{y})$ between $\mathbf{x}$ and $\mathbf{y}$ by*

$$d(\mathbf{x},\ \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$$



The motivation again comes from $\mathbb{R}^3$ as is clear in the diagram. This distance function has all the intuitive properties of distance in $\mathbb{R}^3$, including another version of the triangle inequality.

Theorem 5.3.3

If $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are three vectors in $\mathbb{R}^n$ we have:

1. $d(\mathbf{x},\ \mathbf{y}) \geq 0$ *for all* $\mathbf{x}$ *and* $\mathbf{y}$.
2. $d(\mathbf{x},\ \mathbf{y}) = 0$ *if and only if* $\mathbf{x} = \mathbf{y}$.
3. $d(\mathbf{x},\ \mathbf{y}) = d(\mathbf{y},\ \mathbf{x})$ *for all* $\mathbf{x}$ *and* $\mathbf{y}$.
4. $d(\mathbf{x},\ \mathbf{z}) \leq d(\mathbf{x},\ \mathbf{y}) + d(\mathbf{y},\ \mathbf{z})$*for all* $\mathbf{x}$, $\mathbf{y}$, *and* $\mathbf{z}$.   *Triangle inequality.*

$$|\vec{z} - \vec{x}| = |(\vec{z} - \vec{y}) + (\vec{y} - \vec{x})|$$
$$\leq |\vec{y} - \vec{x}| + |\vec{z} - \vec{y}|$$

5.3.2 正交集与扩展定理

Definition 5.8 Orthogonal and Orthonormal Sets

*We say that two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ are **orthogonal** if $\mathbf{x} \cdot \mathbf{y} = 0$, extending the terminology in $\mathbb{R}^3$ (See Theorem 4.2.3). More generally, a set $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ of vectors in $\mathbb{R}^n$ is called an **orthogonal set** if*

$$\mathbf{x}_i \cdot \mathbf{x}_j = 0 \text{ for all } i \neq j \quad \text{and} \quad \mathbf{x}_i \neq \mathbf{0} \text{ for all } i^{10}$$

*Note that $\{\mathbf{x}\}$ is an orthogonal set if $\mathbf{x} \neq \mathbf{0}$. A set $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ of vectors in $\mathbb{R}^n$ is called **orthonormal** if it is orthogonal and, in addition, each $\mathbf{x}_i$ is a unit vector:*

$$\|\mathbf{x}_i\| = 1 \text{ for each } i.$$

- $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$ 是 $\mathbb{R}^n$ 的标准正交集

- $\{a_1\vec{e}_1, a_2\vec{e}_2, \cdots, a_n\vec{e}_n\}$ $(a_i \neq 0)$ 是 $\mathbb{R}^n$ 的正交集

Definition 5.9 Normalizing an Orthogonal Set

Hence if $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ *is an orthogonal set, then* $\{\frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1, \frac{1}{\|\mathbf{x}_2\|}\mathbf{x}_2, \cdots, \frac{1}{\|\mathbf{x}_k\|}\mathbf{x}_k\}$ *is an orthonormal set, and we say that it is the result of* **normalizing** *the orthogonal set* $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k\}$.

正交集的标准化

Theorem 5.3.4: Pythagoras' Theorem

If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ *is an orthogonal set in* $\mathbb{R}^n$, *then*

$$\|\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_k\|^2 = \|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 + \cdots + \|\mathbf{x}_k\|^2.$$

证：$(\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_k)^2$

$$= |\vec{x}_1|^2 + |\vec{x}_2|^2 + \cdots + |\vec{x}_k|^2 + \sum_{i \neq j} \vec{x}_i \vec{x}_j$$

$$= |\vec{x}_1|^2 + |\vec{x}_2|^2 + \cdots + |\vec{x}_k|^2$$

Theorem 5.3.5

Every orthogonal set in $\mathbb{R}^n$ is linearly independent.

Proof. Let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ be an orthogonal set in $\mathbb{R}^n$ and suppose a linear combination vanishes, say: $t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k = \mathbf{0}$. Then

$$\begin{aligned} 0 = \mathbf{x}_1 \cdot \mathbf{0} &= \mathbf{x}_1 \cdot (t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k) \\ &= t_1(\mathbf{x}_1 \cdot \mathbf{x}_1) + t_2(\mathbf{x}_1 \cdot \mathbf{x}_2) + \cdots + t_k(\mathbf{x}_1 \cdot \mathbf{x}_k) \\ &= t_1\|\mathbf{x}_1\|^2 + t_2(0) + \cdots + t_k(0) \\ &= t_1\|\mathbf{x}_1\|^2 \end{aligned}$$

Since $\|\mathbf{x}_1\|^2 \neq 0$, this implies that $t_1 = 0$. Similarly $t_i = 0$ for each i. □

Theorem 5.3.5 suggests considering orthogonal bases for $\mathbb{R}^n$, that is orthogonal sets that span $\mathbb{R}^n$. These turn out to be the best bases in the sense that, when expanding a vector as a linear combination of the basis vectors, there are explicit formulas for the coefficients.

Theorem 5.3.6: Expansion Theorem

Let $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ be an orthogonal basis of a subspace U of $\mathbb{R}^n$. If $\mathbf{x}$ is any vector in U, we have

$$\mathbf{x} = \left(\frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\right)\mathbf{f}_1 + \left(\frac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\right)\mathbf{f}_1 + \cdots + \left(\frac{\mathbf{x}\cdot\mathbf{f}_m}{\|\mathbf{f}_m\|^2}\right)\mathbf{f}_m$$

Proof. Since $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ spans U, we have $\mathbf{x} = t_1\mathbf{f}_1 + t_2\mathbf{f}_2 + \cdots + t_m\mathbf{f}_m$ where the t_i are scalars. To find t_1 we take the dot product of both sides with $\mathbf{f}_1$:

$$\begin{aligned} \mathbf{x} \cdot \mathbf{f}_1 &= (t_1\mathbf{f}_1 + t_2\mathbf{f}_2 + \cdots + t_m\mathbf{f}_m) \cdot \mathbf{f}_1 \\ &= t_1(\mathbf{f}_1 \cdot \mathbf{f}_1) + t_2(\mathbf{f}_2 \cdot \mathbf{f}_1) + \cdots + t_m(\mathbf{f}_m \cdot \mathbf{f}_1) \\ &= t_1\|\mathbf{f}_1\|^2 + t_2(0) + \cdots + t_m(0) \\ &= t_1\|\mathbf{f}_1\|^2 \end{aligned}$$

$$\Rightarrow \quad t_1 = \frac{\vec{x} \cdot \vec{f}_1}{|\vec{f}_1|^2}$$

类似地， $t_i = \dfrac{\vec{x} \cdot \vec{f}_i}{|\vec{f}_i|^2}$ 对所有 i 成立

对 $\vec{x}$ 用正交集的向量的线性组合来表达

被称为傅立叶展开（Fourier expansion）

所有的 t_i 被称为傅立叶系数

(Fourier coefficient)

⇒ 可以看作从 $\mathbb{R}^2$ 中 u 在 d 上的投影计算拓展而来.

$$\vec{u}_1 = \text{proj}_{\vec{d}}\,\vec{u} = \frac{\vec{d}}{|\vec{d}|} \cdot \frac{\vec{d}\cdot\vec{u}}{|\vec{d}|}$$

$$= \frac{\vec{u}\cdot\vec{d}}{|\vec{d}|^2} \cdot \vec{d}$$

所有方向上的投影累加得到 $\vec{x}$

5.4 矩阵的秩 (Rank of a Matrix)

这里我们用维度 (dimension) 的概念解释秩 (rank)

Definition 5.10 Column and Row Space of a Matrix

The **column space**, col A, of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A.
The **row space**, row A, of A is the subspace of $\mathbb{R}^n$ spanned by the rows of A.

行空间: 矩阵 A 的行张成

列空间: A 的列张成

Lemma 5.4.1

Let A and B denote $m \times n$ matrices.

1. *If $A \to B$ by elementary row operations, then $\operatorname{row} A = \operatorname{row} B$.*
2. *If $A \to B$ by elementary column operations, then $\operatorname{col} A = \operatorname{col} B$.*

对矩阵进行初等行/列变换不会改变矩阵的行/列空间

⇒ 由子空间的 S_2. S_3 可以轻松地证明

Lemma 5.4.2

If R is a row-echelon matrix, then

1. *The nonzero rows of R are a basis of $\operatorname{row} R$.*
2. *The columns of R containing leading ones are a basis of $\operatorname{col} R$.*

第1条证明了:

$$\dim(\operatorname{row} A) = \operatorname{rank} A$$

对所有 $m\times n$ 矩阵 A 成立，与最终化为的 R 无关.

(无论R怎么化. 非零行数是不变的. 而 A 经一系列初等变换得到 R. 秩是一样的)

Theorem 5.4.1: Rank Theorem

Let A denote any $m \times n$ matrix of rank r. Then

$$\dim(\operatorname{col} A) = \dim(\operatorname{row} A) = r$$

Moreover, if A is carried to a row-echelon matrix R by row operations, then

1. *The r nonzero rows of R are a basis of row A.*
2. *If the leading 1s lie in columns $j_1, j_2, \dots, j_r$ of R, then columns $j_1, j_2, \dots, j_r$ of A are a basis of col A.*

所有的A都成立！

Example 5.4.2

Compute the rank of $A = \begin{bmatrix} 1 & 2 & 2 & -1 \\ 3 & 6 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{bmatrix}$ and find bases for row A and col A.

Solution. The reduction of A to row-echelon form is as follows:

$$\begin{bmatrix} 1 & 2 & 2 & -1 \\ 3 & 6 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 2 & -1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & -1 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 2 & -1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Hence rank $A = 2$, and $\{\begin{bmatrix} 1 & 2 & 2 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & -3 \end{bmatrix}\}$ is a basis of row A by Lemma 5.4.2. Since the leading 1s are in columns 1 and 3 of the row-echelon matrix, Theorem 5.4.1 shows that columns 1 and 3 of A are a basis $\left\{ \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix} \right\}$ of col A.

Corollary 5.4.1

If A is any matrix, then rank A = rank (A^T).

If A is an $m \times n$ matrix, we have col $A \subseteq \mathbb{R}^m$ and row $A \subseteq \mathbb{R}^n$. Hence Theorem 5.2.8 shows that $\dim(\operatorname{col} A) \le \dim(\mathbb{R}^m) = m$ and $\dim(\operatorname{row} A) \le \dim(\mathbb{R}^n) = n$. Thus Theorem 5.4.1 gives:

Corollary 5.4.2

If A is an $m \times n$ matrix, then rank $A \le m$ and rank $A \le n$.

Corollary 5.4.3

rank A = rank (UA) = rank (AV) whenever U and V are invertible.

Lemma 5.4.3

Let A, U, and V be matrices of sizes $m \times n$, $p \times m$, and $n \times q$ respectively.

1. $\operatorname{col}(AV) \subseteq \operatorname{col} A$, *with equality if* $VV' = I_n$ *for some* V'.
2. $\operatorname{row}(UA) \subseteq \operatorname{row} A$, *with equality if* $U'U = I_m$ *for some* U'.

如果U、V是可逆的，则A的秩不变，否则会降秩.

我们已经知道了 $\operatorname{rank} A = \operatorname{rank} UA$

$$\Rightarrow \operatorname{rank}(AV) = \operatorname{rank}(AV)^T = \operatorname{rank} V^T A^T$$
$$= \operatorname{rank} A^T$$
$$= \operatorname{rank} A$$

Corollary 5.4.4

If A is $m \times n$ and B is $n \times m$, then $\operatorname{rank} AB \leq \operatorname{rank} A$ *and* $\operatorname{rank} AB \leq \operatorname{rank} B$.

Proof. By Lemma 5.4.3, $\operatorname{col}(AB) \subseteq \operatorname{col} A$ and $\operatorname{row}(BA) \subseteq \operatorname{row} A$, so Theorem 5.4.1 applies. □

In Section 5.1 we discussed two other subspaces associated with an $m \times n$ matrix A: the null space $\operatorname{null}(A)$ and the image space $\operatorname{im}(A)$

$$\operatorname{null}(A) = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\} \text{ and } \operatorname{im}(A) = \{A\mathbf{x} \mid \mathbf{x} \text{ in } \mathbb{R}^n\}$$

Using rank, there are simple ways to find bases of these spaces. If A has rank r, we have $\operatorname{im}(A) = \operatorname{col}(A)$ by Example 5.1.8, so $\dim[\operatorname{im}(A)] = \dim[\operatorname{col}(A)] = r$. Hence Theorem 5.4.1 provides a method of finding a basis of $\operatorname{im}(A)$. This is recorded as part (2) of the following theorem.

Theorem 5.4.2

Let A denote an $m \times n$ matrix of rank r. Then

1. *The $n - r$ basic solutions to the system $A\mathbf{x} = \mathbf{0}$ provided by the gaussian algorithm are a basis of* $\operatorname{null}(A)$, *so* $\dim[\operatorname{null}(A)] = n - r$.
2. *Theorem 5.4.1 provides a basis of* $\operatorname{im}(A) = \operatorname{col}(A)$, *and* $\dim[\operatorname{im}(A)] = r$.

↳ 对于零空间，有秩-零度定理

$$\operatorname{rank} A + \operatorname{nullity} A = \text{column numbers of } A$$

例、如果 $A = \begin{bmatrix} 1 & -2 & 1 & 1 \\ -1 & 2 & 0 & 1 \\ 2 & -4 & 1 & 0 \end{bmatrix}$，求 A 的零空间、像空间.

解：$\operatorname{null} A = \{ \vec{x} \in \mathbb{R}^4 \mid A\vec{x} = \vec{0} \}$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \quad A\vec{x} = \vec{0}$$

$$\Leftrightarrow \begin{cases} x_1 - 2x_2 + x_3 + x_4 = 0 \\ -x_1 + 2x_2 + x_4 = 0 \\ 2x_1 - 4x_2 + x_3 = 0 \end{cases}$$

$$\Rightarrow \vec{x} = s\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix}$$

$$\operatorname{null} A = \operatorname{span}\left(\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix} \right)$$

$\operatorname{im} A = \{ A\vec{x} \mid \vec{x} \in \mathbb{R}^4 \}$

因为 leading 1s 在第1、3列

$$\Rightarrow \operatorname{im} A = \operatorname{span}\left(\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right)$$

Theorem 5.4.3

The following are equivalent for an $m \times n$ matrix A:

1. rank $A = n$.
2. *The rows of A span $\mathbb{R}^n$.*
3. *The columns of A are linearly independent in $\mathbb{R}^m$.*
4. *The $n \times n$ matrix $A^T A$ is invertible.*
5. *$CA = I_n$ for some $n \times m$ matrix C.*
6. *If $A\mathbf{x} = \mathbf{0}$, $\mathbf{x}$ in $\mathbb{R}^n$, then $\mathbf{x} = \mathbf{0}$.*

Theorem 5.4.4

The following are equivalent for an $m \times n$ matrix A:

1. rank $A = m$.
2. *The columns of A span $\mathbb{R}^m$.*
3. *The rows of A are linearly independent in $\mathbb{R}^n$.*
4. *The $m \times m$ matrix AA^T is invertible.*
5. *$AC = I_m$ for some $n \times m$ matrix C.*
6. *The system $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b}$ in $\mathbb{R}^m$.*

5.5 相似与对角化

(Similarity and Diagonalization)

5.5.1 相似矩阵

Definition 5.11 Similar Matrices

If A and B are $n \times n$ matrices, we say that A and B are **similar***, and write $A \sim B$, if $B = P^{-1}AP$ for some invertible matrix P.*

满足自反性、传递性、对称性.

① $A \sim A$

② $A \sim B \Rightarrow B \sim A$

③ $A \sim B$, $B \sim C \Rightarrow A \sim C$

Definition 5.12 Trace of a Matrix

The **trace** *$\operatorname{tr} A$ of an $n \times n$ matrix A is defined to be the sum of the main diagonal elements of A.*

In other words:

$$\text{If } A = [a_{ij}], \text{ then } \operatorname{tr} A = a_{11} + a_{22} + \cdots + a_{nn}.$$

It is evident that $\operatorname{tr}(A+B) = \operatorname{tr} A + \operatorname{tr} B$ and that $\operatorname{tr}(cA) = c \operatorname{tr} A$ holds for all $n \times n$ matrices A and B and all scalars c. The following fact is more surprising.

Lemma 5.5.1

Let A and B be $n \times n$ matrices. Then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

trace：对角线元素和

Theorem 5.5.1

If A and B are similar $n \times n$ matrices, then A and B have the same determinant, rank, trace, characteristic polynomial, and eigenvalues.

Proof. Let $B = P^{-1}AP$ for some invertible matrix P. Then we have

$$\det B = \det\left(P^{-1}\right) \det A \det P = \det A \text{ because } \det\left(P^{-1}\right) = 1/\det P$$

Similarly, $\operatorname{rank} B = \operatorname{rank}\left(P^{-1}AP\right) = \operatorname{rank} A$ by Corollary 5.4.3. Next Lemma 5.5.1 gives

$$\operatorname{tr}\left(P^{-1}AP\right) = \operatorname{tr}\left[P^{-1}(AP)\right] = \operatorname{tr}\left[(AP)P^{-1}\right] = \operatorname{tr} A$$

As to the characteristic polynomial,

$$\begin{aligned} c_B(x) = \det(xI - B) &= \det\left\{x(P^{-1}IP) - P^{-1}AP\right\} \\ &= \det\left\{P^{-1}(xI - A)P\right\} \\ &= \det(xI - A) \\ &= c_A(x) \end{aligned}$$

Finally, this shows that A and B have the same eigenvalues because the eigenvalues of a matrix are the roots of its characteristic polynomial. □

等秩：乘 P、P^{-1} 可以看作初等行变换、列变换，不会改变秩.

5.5.2 对角化复习

Theorem 5.5.3

Let A be an $n \times n$ matrix.

1. *A is diagonalizable if and only if $\mathbb{R}^n$ has a basis $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ consisting of eigenvectors of A.*
2. *When this is the case, the matrix $P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$ is invertible and $P^{-1}AP = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ where, for each i, λ_i is the eigenvalue of A corresponding to $\mathbf{x}_i$.*

The next result is a basic tool for determining when a matrix is diagonalizable. It reveals an important connection between eigenvalues and linear independence: Eigenvectors corresponding to distinct eigenvalues are necessarily linearly independent.

Theorem 5.5.4

Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ be eigenvectors corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$ of an $n \times n$ matrix A. Then $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ is a linearly independent set.

特征向量 ⇒ 两两线性无关

Theorem 5.5.5

If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalizable.

Proof. Choose one eigenvector for each of the n distinct eigenvalues. Then these eigenvectors are independent by Theorem 5.5.4, and so are a basis of $\mathbb{R}^n$ by Theorem 5.2.7. Now use Theorem 5.5.3. □

Example 5.5.4

Show that $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 3 \\ -1 & 1 & 0 \end{bmatrix}$ is diagonalizable.

Solution. A routine computation shows that $c_A(x) = (x-1)(x-3)(x+1)$ and so has distinct eigenvalues 1, 3, and -1. Hence Theorem 5.5.5 applies.

对于不可完全对角化的矩阵：

Lemma 5.5.2

Let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ be a linearly independent set of eigenvectors of an $n \times n$ matrix A, extend it to a basis $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_n\}$ of $\mathbb{R}^n$, and let

$$P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$$

be the (invertible) $n \times n$ matrix with the $\mathbf{x}_i$ as its columns. If $\lambda_1, \lambda_2, \dots, \lambda_k$ are the (not necessarily distinct) eigenvalues of A corresponding to $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ respectively, then $P^{-1}AP$ has block form

$$P^{-1}AP = \begin{bmatrix} \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_k) & B \\ 0 & A_1 \end{bmatrix}$$

where B has size $k \times (n-k)$ and A_1 has size $(n-k) \times (n-k)$.

Definition 5.13 Eigenspace of a Matrix

If λ is an eigenvalue of an $n \times n$ matrix A, define the **eigenspace** *of A corresponding to λ by*

$$E_\lambda(A) = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid A\mathbf{x} = \lambda\mathbf{x}\}$$

$\Leftrightarrow$ null $(A - \lambda I)$

Theorem 5.5.6

The following are equivalent for a square matrix A for which $c_A(x)$ factors completely.

1. *A is diagonalizable.*
2. *$\dim[E_\lambda(A)]$ equals the multiplicity of λ for every eigenvalue λ of the matrix A.*

总结： A 可对角化 (n×n)

$\Leftrightarrow$ n 个线性无关的 eigenvectors

$\Leftrightarrow$ n 个 eigenvalues

$\Leftrightarrow$ $\forall \lambda_i$ of A, $\dim E_{\lambda_i}(A)$ = multiplicity of λ_i

5.5.3 复特征值 (Complex Eigenvalues)

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$ characteristic polynomial:

$$x^2 + 1 = 0 .$$

$$\Rightarrow \lambda = \pm i$$

$$\Rightarrow P = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$$

$$D = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$$

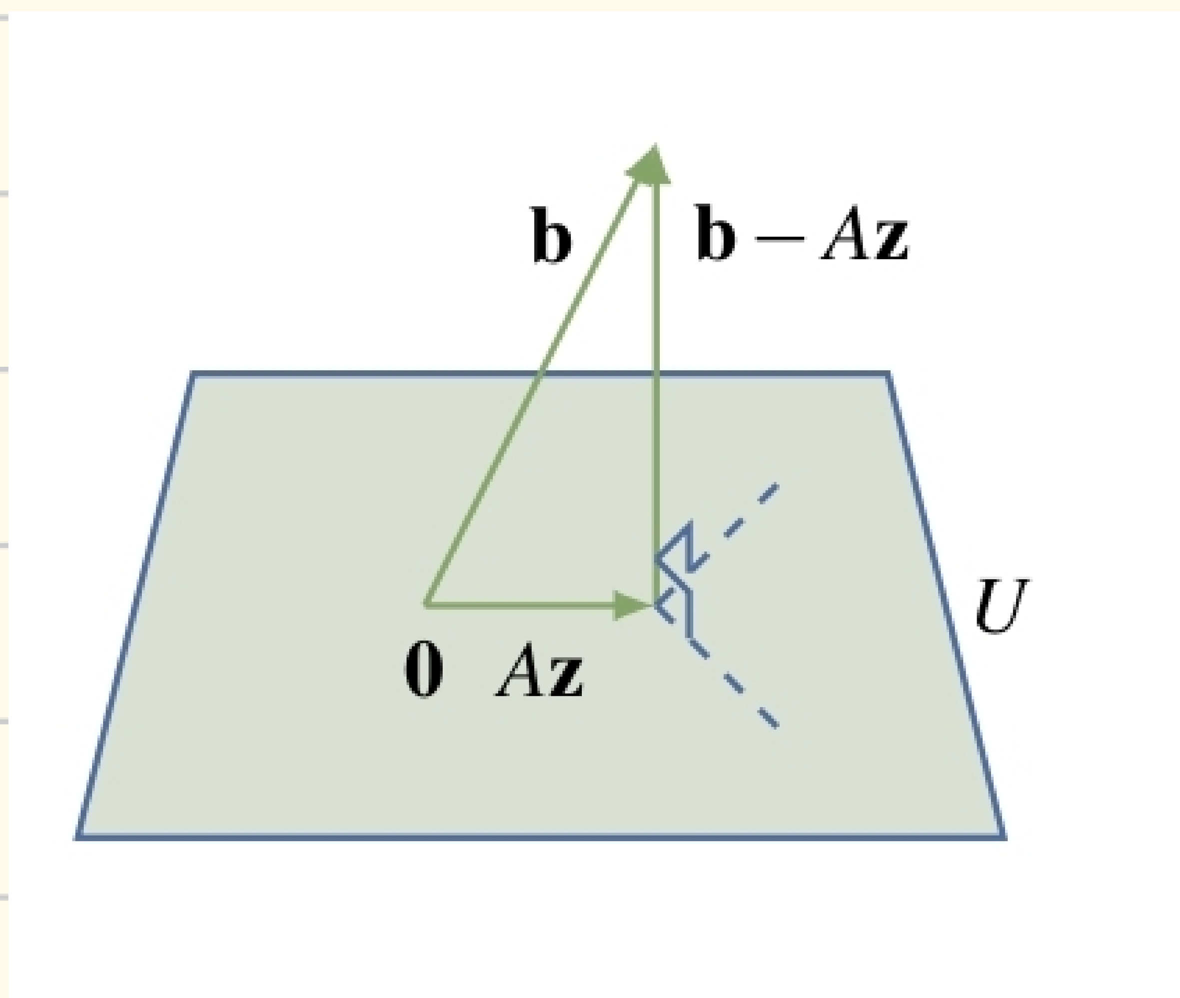
5.5.4 对称矩阵 (Symmetric Matrices)

一句化概括就是：对称矩阵的 eigenvalue 全为实数.

5.6 最佳逼近与最小二乘

(Best Approximation and Least Square)

对于无解的系统，得到逼近的解.



$A\vec{x} = \vec{b}$ is an inconsistent system.

$\Leftrightarrow$ $\vec{b}$ is not in col A

$\Leftrightarrow$ $\vec{b}$ is not in Im A

$= \{ A\vec{x} \mid \vec{x} \in \mathbb{R}^n \}$

$\exists\ \vec{z} \in \mathbb{R}^n,\ |A\vec{z} - \vec{b}|$ min

$\vec{z}$ 即是最佳逼近

求解方法：

$|\vec{b} - A\vec{z}|$ min $\Leftrightarrow$ $\vec{b} - A\vec{z} \perp$ col A

$\Leftrightarrow$ $\vec{b} - A\vec{z} \perp \vec{a_i}$, $\vec{a_i} \in$ col A

$$\Rightarrow \quad A^T \cdot (\vec{b} - A\vec{z}) = \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vdots \\ \vec{a_n} \end{bmatrix} \cdot (\vec{b} - A\vec{z}) = \vec{0}$$

$$\Rightarrow \quad A^T\vec{b} = A^TA\vec{z}$$

可解 $\vec{z}$.

高斯消元法 $> \vec{z} = (A^TA)^{-1}A^T\vec{b}$

Definition 5.14 Normal Equations

This is a system of linear equations called the ***normal equations*** *for* $\mathbf{z}$*.*

Note that this system can have more than one solution (see Exercise 5.6.5). However, the $n \times n$ matrix A^TA is invertible if (and only if) the columns of A are linearly independent (Theorem 5.4.3); so, in this case, $\mathbf{z}$ is uniquely determined and is given explicitly by $\mathbf{z} = (A^TA)^{-1}A^T\mathbf{b}$. However, the most efficient way to find $\mathbf{z}$ is to apply gaussian elimination to the normal equations.

This discussion is summarized in the following theorem.

Theorem 5.6.1: Best Approximation Theorem

Let A be an $m \times n$ matrix, let $\mathbf{b}$ be any column in $\mathbb{R}^m$, and consider the system

$$A\mathbf{x} = \mathbf{b}$$

of m equations in n variables.

1. *Any solution $\mathbf{z}$ to the normal equations*

   $$(A^TA)\mathbf{z} = A^T\mathbf{b}$$

   is a best approximation to a solution to $A\mathbf{x} = \mathbf{b}$ in the sense that $\|\mathbf{b} - A\mathbf{z}\|$ is the minimum value of $\|\mathbf{b} - A\mathbf{x}\|$ as $\mathbf{x}$ ranges over all columns in $\mathbb{R}^n$.

2. *If the columns of A are linearly independent, then A^TA is invertible and $\mathbf{z}$ is given uniquely by $\mathbf{z} = (A^TA)^{-1}A^T\mathbf{b}$.*

(2) 对于线性独立的列向量，$\vec{z}$ 可由 $(A^TA)^{-1}A^T\vec{b}$ 唯一得到

注意到，特殊地，若果 A 是 $n\times n$ 可逆方阵，

有 $\vec{z} = (A^TA)^{-1}A^T\vec{b}$

$$= A^{-1}(A^T)^{-1}A^T\vec{b}$$

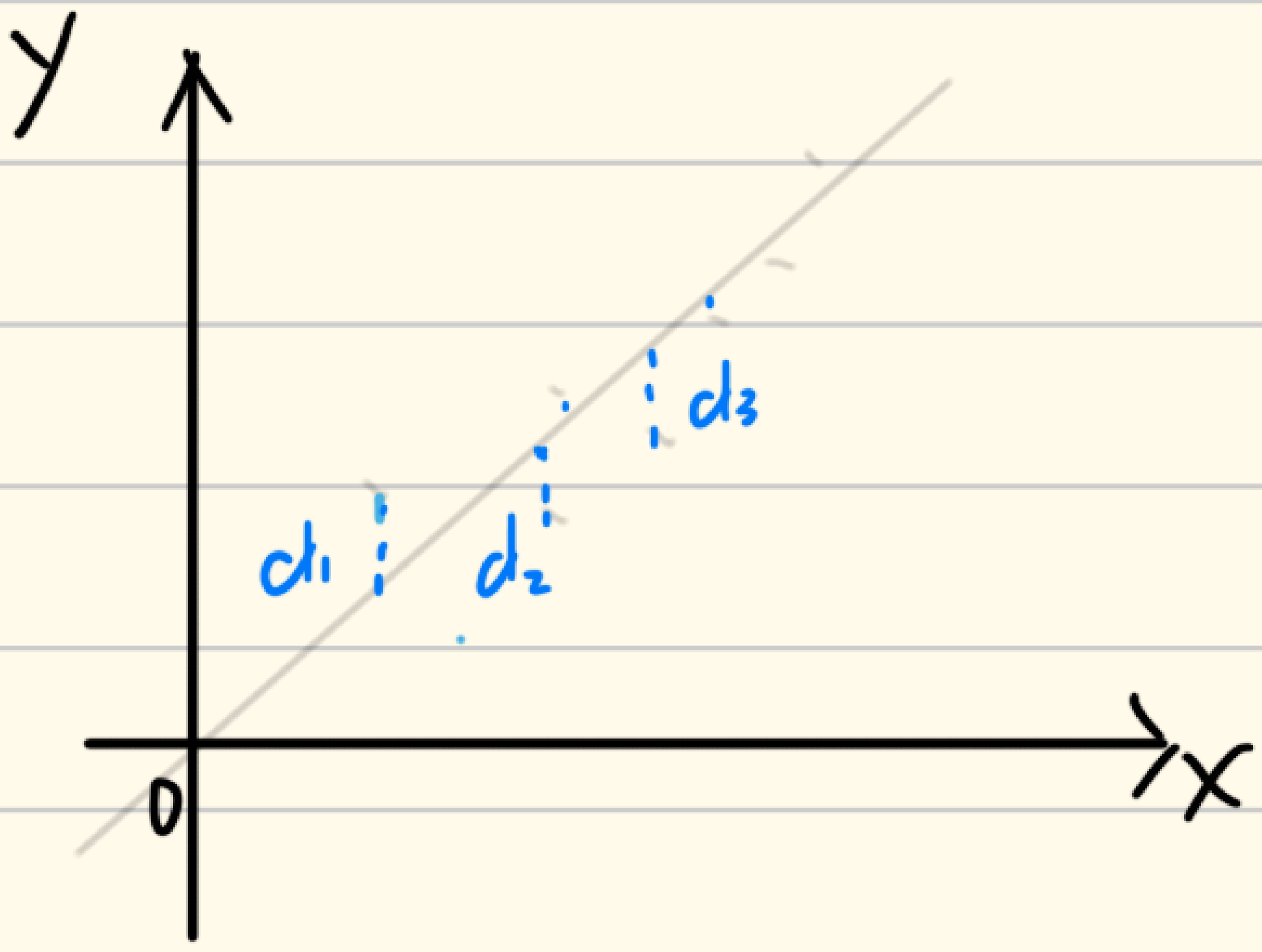
$$= A^{-1}\vec{b}$$

这即是 $|A\vec{x}-\vec{b}|=0$ 的解.

如果列线性无关，$(A^TA^{-1})A^T$ 起到了为一个非方阵矩阵构建逆的作用、称之为广义逆（generalized inverse）

行亦如此

5.6.1 最小二乘逼近（Least Square Approximation）



用直线拟合散点、探究变元 x、y 之间的关系.

记 $d_i = y_i - f(x_i)$

用 $S^2 = \sum_i d_i^2$ 来度量拟合效果.

s^2 最小的直线称为最小二乘逼近线.

l: $y = a_0 + a_1 x$, n 组数据 (x_i, y_i)

有 $\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}$

用最小二乘逼近,

$(A^T A)\vec{z} = A^T \vec{b}$ 得 $\vec{z}$, 令 s^2 最小

Theorem 5.6.2

Suppose that n data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ are given, where at least two of $x_1, x_2, \ldots, x_n$ are distinct. Put

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

Then the least squares approximating line for these data points has equation

$$y = z_0 + z_1 x$$

where $\mathbf{z} = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix}$ is found by gaussian elimination from the normal equations

$$(M^T M)\mathbf{z} = M^T \mathbf{y}$$

The condition that at least two of $x_1, x_2, \ldots, x_n$ are distinct ensures that $M^T M$ is an invertible matrix, so $\mathbf{z}$ is unique:

$$\mathbf{z} = (M^T M)^{-1} M^T \mathbf{y}$$

5.6.2 最小二乘逼近多项式

Definition 5.15 Least Squares Approximation

*A polynomial $f(x)$ satisfying this condition is called a **least squares approximating polynomial** of degree m for the given data pairs.*

If we write

$$M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} r_0 \\ r_1 \\ \vdots \\ r_m \end{bmatrix}$$

we see that $f(\mathbf{x}) = M\mathbf{r}$. Hence we want to find $\mathbf{r}$ such that $\|\mathbf{y} - M\mathbf{r}\|^2$ is as small as possible; that is, we want a best approximation $\mathbf{z}$ to the system $M\mathbf{r} = \mathbf{y}$. Theorem 5.6.1 gives the first part of Theorem 5.6.3.

Theorem 5.6.3

Let n data pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be given, and write

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \quad \mathbf{z} = \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_m \end{bmatrix}$$

1. *If $\mathbf{z}$ is any solution to the normal equations*

   $$(M^T M)\mathbf{z} = M^T\mathbf{y}$$

   then the polynomial

   $$z_0 + z_1 x + z_2 x^2 + \cdots + z_m x^m$$

   is a least squares approximating polynomial of degree m for the given data pairs.

2. *If at least $m+1$ of the numbers $x_1, x_2, \ldots, x_n$ are distinct (so $n \geq m+1$), the matrix $M^T M$ is invertible and $\mathbf{z}$ is uniquely determined by*

   $$\mathbf{z} = (M^T M)^{-1} M^T \mathbf{y}$$

对(2)的证明：

令 $\vec{r} = \begin{bmatrix} r_0 \\ r_1 \\ \cdots \\ r_m \end{bmatrix}$

$$M^T\vec{r} = r_0 [1 \ x_1 \ x_1^2 \cdots x_1^m]^T$$
$$+ r_1 [1 \ x_2 \ x_2^2 \cdots x_2^m]^T$$
$$\cdots$$
$$+ r_{n-1} [1 \ x_n \ x_n^2 \cdots x_n^m]^T$$

对于 $r_{i-1} [1 \ x_i \ x_i^2 \ \cdots \ x_i^m]^T = \vec{0}$

有 $r_0 + r_1 x + r_2 x^2 + \cdots + r_m x_i^m = 0$

因为至少有 $m+1$ 个根．而 x_i 最高 m 次

故 $r_0 = r_1 = \cdots = r_m = 0$

M^T 的列向量线性无关

$\Rightarrow$ M^T 为可逆

$\Rightarrow$ M^TM 可逆

(2) 翻译：n组数据中 m+1 个点横坐标不一样，则最小二乘最佳逼近向量唯一

Example 5.6.4

Find the least squares approximating quadratic $y = z_0 + z_1 x + z_2 x^2$ for the following data points.

$$(-3,\ 3),\ (-1,\ 1),\ (0,\ 1),\ (1,\ 2),\ (3,\ 4)$$

$$\vec{z} = \begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix}$$

$$M = \begin{bmatrix} 1 & -3 & 9 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 4 \\ 1 & 3 & 9 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \\ 4 \end{bmatrix}$$

$$M\vec{z} = \vec{y}$$

$$\hat{z} = (M^T M)^{-1} M^T \vec{y}$$

$$= \begin{bmatrix} 1.15 \\ 0.20 \\ 0.26 \end{bmatrix}$$

5.6.3 其它方程

Theorem 5.6.4

Let n data pairs (x_1, y_1), (x_2, y_2), $\dots$, (x_n, y_n) be given, and suppose that $m+1$ functions $f_0(x)$, $f_1(x)$, $\dots$, $f_m(x)$ are specified. Write

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad M = \begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_m(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_m(x_2) \\ \vdots & \vdots & & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_m(x_n) \end{bmatrix} \quad \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}$$

1. *If $\mathbf{z}$ is any solution to the normal equations*

$$(M^T M)\mathbf{z} = M^T \mathbf{y}$$

then the function

$$z_0 f_0(x) + z_1 f_1(x) + \cdots + z_m f_m(x)$$

is the best approximation for these data among all functions of the form $r_0 f_0(x) + r_1 f_1(x) + \cdots + r_m f_m(x)$ where the r_i are in $\mathbb{R}$.

2. *If $M^T M$ is invertible (that is, if* rank $(M) = m+1$*), then $\mathbf{z}$ is uniquely determined; in fact, $\mathbf{z} = (M^T M)^{-1}(M^T \mathbf{y})$.*

Example 5.6.5

Given the data pairs $(-1, 0)$, $(0, 1)$, and $(1, 4)$, find the least squares approximating function of the form $r_0 x + r_1 2^x$.

$$M = \begin{bmatrix} -1 & \frac{1}{2} \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \qquad \vec{r} = \begin{bmatrix} r_0 \\ r_1 \end{bmatrix}$$

$$\vec{y} = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}$$

$\Rightarrow \hat{r} = (M^T M)^{-1} M^T \vec{y}$.

$\Rightarrow \hat{r} = \cdots$

对 M 是每个参数 r_i 的系数矩阵，由 $f(x)$ 决定值

5.7 相关性和方差的应用

一组样本的均值（sample mean）

$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$

方差：

Variance $s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$

均值和方差可以用向量点乘方便地表示、

令 $\vec{1} = [1 \quad 1 \quad 1 \cdots \quad 1]$

$$\bar{x} = \frac{1}{n}(\vec{1} \cdot \vec{x})$$

记 $\vec{x}_c = [\, x_1-\bar{x} \quad x_2-\bar{x} \quad \cdots \quad x_n-\bar{x} \,]$

$$s^2 = \frac{1}{n-1} \left| \vec{x}_c \cdot \vec{1} \right|^2$$

描述两组数据的关系 x, y

相关系数 (correlation coefficient)

$$r = r(\vec{x}_c, \vec{y}_c) = \frac{\vec{x}_c \cdot \vec{y}_c}{|\vec{x}_c| \cdot |\vec{y}_c|}$$

可以将 r 看作 $\vec{x}_c$、$\vec{y}_c$ 的夹角余弦值

Theorem 5.7.1

Let $\mathbf{x} = [\, x_1 \quad x_2 \quad \cdots \quad x_n \,]$ *and* $\mathbf{y} = [\, y_1 \quad y_2 \quad \cdots \quad y_n \,]$ *be (nonzero) paired samples, and let* $r = r(\mathbf{x}, \mathbf{y})$ *denote the correlation coefficient. Then:*

1. $-1 \le r \le 1$.
2. $r = 1$ *if and only if there exist* a *and* $b > 0$ *such that* $y_i = a + bx_i$ *for each* i.
3. $r = -1$ *if and only if there exist* a *and* $b < 0$ *such that* $y_i = a + bx_i$ *for each* i.

$$\vec{x}_c \cdot \vec{y}_c = \vec{x} \cdot \vec{y} - n\,\bar{x} \cdot \bar{y}$$

证： $\vec{x}_c \cdot \vec{y}_c = (\vec{x} - \bar{x} \cdot \vec{1})(\vec{y} - \bar{y} \cdot \vec{1})$

$$= \vec{x} \cdot \vec{y} - \vec{x} \cdot \bar{y} \cdot \vec{1} - \bar{x}\,\vec{y}\,\vec{1} + xy \cdot n$$

$$= \vec{x} \cdot \vec{y} - nxy - nxy + nxy$$

$$= \vec{x} \cdot \vec{y} - nxy$$

Variance Formula

If x is a sample vector, then $s_x^2 = \frac{1}{n-1}\left(\|\cancel{\mathbf{x}_c}\|^2 - n\bar{x}^2\right)$.

$|X^2|$

We also get a convenient formula for the correlation coefficient, $r = r(\mathbf{x},\ \mathbf{y}) = \frac{\mathbf{x}_c \cdot \mathbf{y}_c}{\|\mathbf{x}_c\|\ \|\mathbf{y}_c\|}$. Moreover, (5.7) and the fact that $s_x^2 = \frac{1}{n-1}\|\mathbf{x}_c\|^2$ give:

Correlation Formula

If $\mathbf{x}$ and $\mathbf{y}$ are sample vectors, then

$$r = r(\mathbf{x},\ \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y} - n\bar{x}\,\bar{y}}{(n-1)s_x s_y}$$

Finally, we give a method that simplifies the computations of variances and correlations.

Data Scaling

Let $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}$ be sample vectors. Given constants a, b, c, and d, consider new samples $\mathbf{z} = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix}$ where $z_i = a + bx_i$, for each i and $w_i = c + dy_i$ for each i. Then:

a. $\bar{z} = a + b\bar{x}$

b. $s_z^2 = b^2 s_x^2$, so $s_z = |b| s_x$

c. *If b and d have the same sign, then $r(\mathbf{x},\ \mathbf{y}) = r(\mathbf{z},\ \mathbf{w})$.*