

计算方法

Computational Methods

Lecture 4: 最小二乘、正交与 QR 分解

Least Squares, Orthogonality, and QR Decomposition

刘景铎

南京大学 · 计算机学院

课程信息

计算方法 Fall 2026

授课教师 刘景铎 liu@nju.edu.cn

上课 周四 5-6 节 · 1-18 周 · 仙 I-319

答疑 周四下午 4 点

助教 周宇恒 · 侯哲

作业 nm_nju_2026@163.com

课程主页 TCS Wiki · Fall 2026

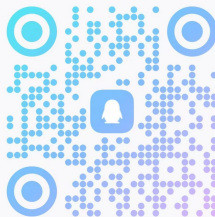
QQ 群 1054896504 加入时注明姓名、专业、
学号

教材: Timothy Sauer, 《数值分析》(原书第 2 版)。 最新公告以
课程主页为准。



计算方法 Fall 2026

群号: 1054896504



扫一扫二维码, 加入群聊



回顾：上一讲已经把问题化成“找最近点”

上一讲最后停在这里。设

$$A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m, \quad V := \text{col}(A).$$

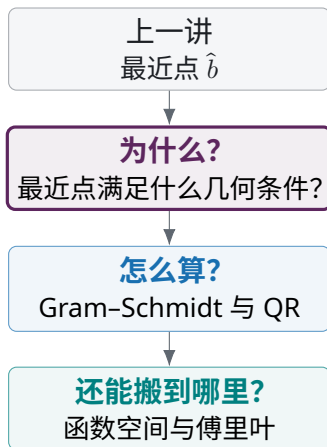
当 $b \notin V$ 时，不再要求 $Ax = b$ ，而是求

$$\min_x \|Ax - b\|_2^2.$$

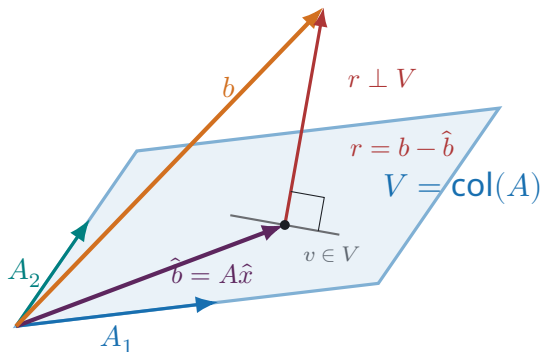
若最优解记为 $\hat{x}$ ，那么

$$\hat{b} := A\hat{x} \in V$$

就是 V 中离 b 最近的点。上一讲已经找到“问题”；
这一讲要找到“结构”。



最小二乘的几何直觉：投影到二维子空间



b 不一定在 V 中。

在 V 中离 b 最近的点记为

$$\hat{b} = A\hat{x}.$$

从 b 指向 V 的最短误差向量有什么特别？

它必须垂直于整个子空间。

先定义“垂直”：欧氏内积

对 $u, v \in \mathbb{R}^m$ ，定义标准欧氏内积

$$\langle u, v \rangle := u^\top v = \sum_{i=1}^m u_i v_i.$$

它与上一讲的 2-范数相容：

$$\|u\|_2 = \sqrt{\langle u, u \rangle}.$$

今天真正需要的新结构是“方向之间的关系”：

$$u \perp v \iff \langle u, v \rangle = 0.$$

对一个子空间 V ，记它的正交补为

$$V^\perp := \{u : \langle u, v \rangle = 0 \ \forall v \in V\}.$$

内积还能刻画夹角；本讲只需要其中最关键的特例： $90^\circ \iff$ **内积为 0**。

最近点判据：残差与子空间正交

令 $V = \text{col}(A)$ 、 $\hat{b} = A\hat{x}$ 、 $r = b - \hat{b}$ 。则

$$\boxed{\hat{b} \text{ 是 } V \text{ 中离 } b \text{ 最近的点}} \iff \boxed{r \perp V.}$$

必要性：若不垂直，就还能更近。

若存在 $v \in V$ 使 $\langle r, v \rangle \neq 0$ ，沿 $\hat{b} + tv \in V$ 移动：

$$\|r - tv\|_2^2 = \|r\|_2^2 - 2t \langle r, v \rangle + t^2 \|v\|_2^2.$$

取 $t_* = \langle r, v \rangle / \|v\|_2^2$ ，便有

$$\|r - t_* v\|_2^2 = \|r\|_2^2 - \frac{\langle r, v \rangle^2}{\|v\|_2^2} < \|r\|_2^2,$$

与“最近”矛盾。

充分性：垂直就已经最短。

若 $r \perp V$ ，任取 $y \in V$ 。因为

$$\hat{b} - y \in V, \quad r \perp (\hat{b} - y),$$

勾股定理给出

$$\|b - y\|_2^2 = \|r\|_2^2 + \|\hat{b} - y\|_2^2 \geq \|r\|_2^2.$$

因此 $\hat{b}$ 确实最近。

最近点 = 正交投影。

从正交条件得到法线方程

因为 $A_1, \dots, A_n$ 张成 $V = \text{col}(A)$,

$$r \perp V \iff A_j^\top r = 0 \quad \forall j.$$

代入 $r = b - A\hat{x}$:

$$A_j^\top (b - A\hat{x}) = 0 \quad \forall j.$$

把这些条件叠在一起:

$$A^\top (b - A\hat{x}) = 0.$$

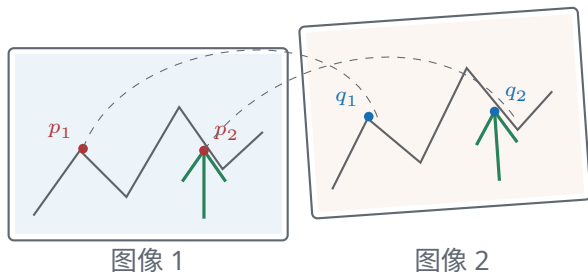
于是

$$\boxed{A^\top A\hat{x} = A^\top b.}$$

法线方程 (normal equations)

名字里的 *normal* 就是“法向 / 垂直”: 它是**残差垂直于列空间**的矩阵写法。

投影点 $\hat{b}$ 总是唯一; 若 A 的列线性相关, 同一个 $\hat{b}$ 可能对应多个 $\hat{x}$ 。列满秩时 $\hat{x}$ 才唯一。



图像 1

图像 2

在两张图片中选若干**对应点**：

$$p_i \longleftrightarrow q_i.$$

希望找到**同一个**二维仿射变换

$$q_i \approx Mp_i + t, \quad M \in \mathbb{R}^{2 \times 2}, t \in \mathbb{R}^2,$$

对所有参考点都尽量成立。

真实匹配带噪声，通常不可能让
每个等式都精确成立。

真正的问题是

$$\exists (M, t) \quad \text{s.t.} \quad \forall i, q_i \approx Mp_i + t.$$

而不是

$$\forall i, \exists (M_i, t_i) \quad \text{s.t.} \quad q_i = M_i p_i + t_i.$$

第二种写法对每个点都可以单独“量身定做”，几乎没有约束力。把 M, t 的 6 个未知参数堆成 $\theta \in \mathbb{R}^6$ 。若有 N 对对应点，就得到 $2N$ 个标量约束：

$$B\theta \approx y \quad \implies \quad \min_{\theta} \|B\theta - y\|_2^2.$$

通常 $N > 3$ ，因此这是一个**超定但带噪声**的线性系统。建模完成以后，后面仍统一把

B, θ, y 记回 A, x, b 。

同一个最优条件，还有第二种看法：局部线性化

第一讲中，导数描述一维函数的局部线性变化：

$$f(x + th) = f(x) + tf'(x)h + o(t).$$

对可微的标量函数 $F : \mathbb{R}^n \rightarrow \mathbb{R}$ ，在刚才的欧氏内积下，梯度 $\nabla F(x)$ 由

$$\boxed{F(x + th) = F(x) + t \langle \nabla F(x), h \rangle + o(t)} \quad (t \rightarrow 0)$$

刻画： $\langle \nabla F(x), h \rangle$ 就是沿方向 h 的**一阶变化率**。若 $\nabla F(x) \neq 0$ ，取 $h = -\nabla F(x)$ ，对充分小的 $t > 0$ 有

$$F(x + th) < F(x).$$

因此无约束局部极小点必须满足

$$\boxed{\nabla F(x) = 0.}$$

梯度不是新的“公式表”：它只是多维局部线性化的一种表示。

最小二乘的微积分解释：梯度为什么是 $2A^\top(Ax - b)$?

记

$$F(x) = \|Ax - b\|_2^2.$$

因为 $s \mapsto s^2$ 在 $s \geq 0$ 上单调，最小化 $\|Ax - b\|_2$ 与最小化 $F(x)$ 有完全相同的解。
沿固定方向 h 走一个标量步长 t :

$$\begin{aligned} F(x + th) &= \|Ax - b + tAh\|_2^2 \\ &= F(x) + 2t \langle Ax - b, Ah \rangle + t^2 \|Ah\|_2^2 \\ &= F(x) + 2t \langle A^\top(Ax - b), h \rangle + t^2 \|Ah\|_2^2. \end{aligned}$$

一次项的系数必须等于 $\langle \nabla F(x), h \rangle$ ，因此

$$\nabla F(x) = 2A^\top(Ax - b).$$

于是极小点满足

$$A^\top(A\hat{x} - b) = 0.$$

第三种看法：变分恒等式直接给出全局最优性

设 $F(x) = \|Ax - b\|_2^2$ 。对任意方向 h ,

$$F(x + h) = F(x) + 2h^\top A^\top (Ax - b) + \|Ah\|_2^2.$$

若 $A^\top (Ax - b) = 0$, 那么

$$F(x + h) = F(x) + \|Ah\|_2^2 \geq F(x) \quad \forall h.$$

所以 x 不只是局部极小点, 而是**全局最小点**。反过来, 令

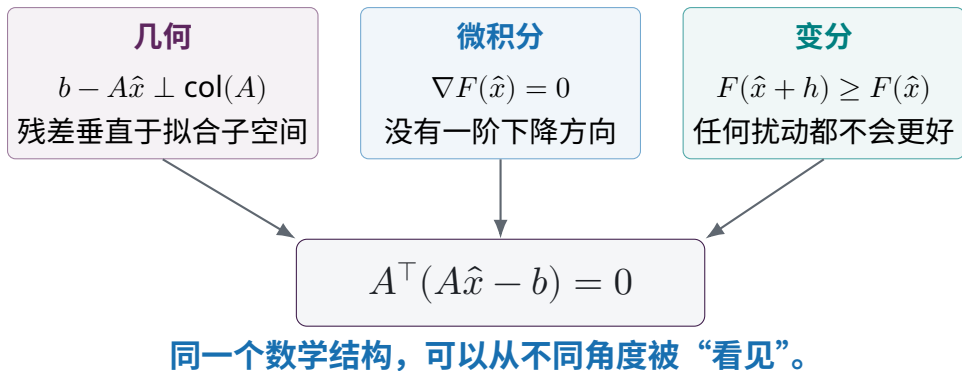
$$g := A^\top (Ax - b) \neq 0, \quad h = -\varepsilon g.$$

则

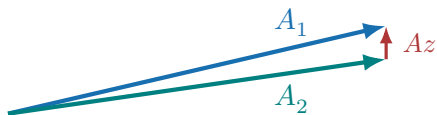
$$F(x + h) - F(x) = -2\varepsilon \|g\|_2^2 + \varepsilon^2 \|Ag\|_2^2 < 0$$

对充分小的 $\varepsilon > 0$ 成立。**对这个二次问题, 法线方程已经足以刻画全局最优解。**

一个条件，三种语言



为什么接近线性相关会麻烦？



先看只有两列的情形 $A = [A_1 \ A_2]$ 。若 $A_1 \approx A_2$ ，取

$$z := \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

那么

$$Az = A_1 - A_2 \approx 0.$$

对任意 $t \neq 0$ ，令系数变化 $\Delta x = tz$ ，则

$$\frac{\|A\Delta x\|_2}{\|\Delta x\|_2} = \frac{\|Az\|_2}{\|z\|_2} \approx 0.$$

沿 z 方向，系数可以变化很大而拟合向量几乎不动；
 A 几乎把 z 这个方向“压扁”了。

法线方程给出公式，但数值算法还没有结束

若 A 列满秩，则 $A^\top A$ 可逆，因此形式上

$$\hat{x} = (A^\top A)^{-1} A^\top b.$$

原问题本身可能敏感

$$Az \approx 0.$$

沿 z 方向大幅改变系数， Ax 却几乎不动。

这是问题本身的几何。

形成 $A^\top A$ 还会进一步压缩小方向

$$A = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix} \implies A^\top A = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon^2 \end{pmatrix}.$$

ε 方向变成 ε^2 。

QR 不会修复原问题的敏感性；它只是避免先形成 $A^\top A$ ，也不需要显式求逆。

如果列向量标准正交，最小二乘立刻解耦

设 $Q = [q_1 \cdots q_n] \in \mathbb{R}^{m \times n}$ 的列向量满足

$$\langle q_i, q_j \rangle = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad \Longleftrightarrow \quad Q^\top Q = I_n.$$

最小二乘的投影条件是

$$Q^\top (b - Q\hat{x}) = 0.$$

利用 $Q^\top Q = I_n$ ，立刻得到

$$Q^\top b - \hat{x} = 0, \quad \boxed{\hat{x} = Q^\top b}.$$

也就是逐个读取

$$\boxed{\hat{x}_j = q_j^\top b = \langle q_j, b \rangle}.$$

正交把原本耦合在一起的系数，变成了彼此独立的投影系数。

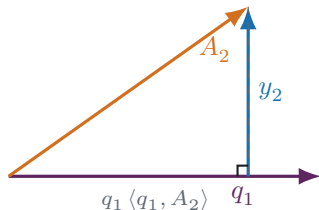
Gram-Schmidt: 把一组方向变成标准正交方向

给定线性无关向量 $A_1, A_2, \dots$ 。第一步:

$$y_1 = A_1, \quad q_1 = \frac{y_1}{\|y_1\|_2}.$$

第二步先从 A_2 中减去沿 q_1 的投影:

$$y_2 = A_2 - q_1 \langle q_1, A_2 \rangle, \quad q_2 = \frac{y_2}{\|y_2\|_2}.$$



Gram-Schmidt: 一般一步

已有标准正交向量 $q_1, \dots, q_{j-1}$ 时,

$$y_j = A_j - \sum_{i=1}^{j-1} q_i \langle q_i, A_j \rangle, \quad q_j = \frac{y_j}{\|y_j\|_2}.$$

于是

$$A_j = \sum_{i=1}^j r_{ij} q_i,$$

其中

$$r_{ij} = \langle q_i, A_j \rangle \quad (i < j), \quad r_{jj} = \|y_j\|_2.$$

第 j 列只会用到 $q_1, \dots, q_j$; 因此 $i > j$ 的系数必为 0。

这正是后面 QR 分解中的 R 为什么是上三角矩阵。

Gram-Schmidt 的实际实现：CGS 与 MGS

两种写法在**精确算术**下等价；真正的区别只在内层循环如何计算投影系数。

Classical Gram-Schmidt (CGS)

```
for  $j = 1, \dots, n$ :  $y \leftarrow A_j$   
    for  $i = 1, \dots, j - 1$ :  
         $r_{ij} \leftarrow q_i^\top A_j$   
         $y \leftarrow y - r_{ij}q_i$   
     $r_{jj} \leftarrow \|y\|_2, \quad q_j \leftarrow y/r_{jj}$ 
```

Modified Gram-Schmidt (MGS)

```
for  $j = 1, \dots, n$ :  $y \leftarrow A_j$   
    for  $i = 1, \dots, j - 1$ :  
         $r_{ij} \leftarrow q_i^\top y$   
         $y \leftarrow y - r_{ij}q_i$   
     $r_{jj} \leftarrow \|y\|_2, \quad q_j \leftarrow y/r_{jj}$ 
```

唯一差别：CGS 总是对原列 A_j 做投影；MGS 对**当前剩余向量** y 继续做投影。
精确算术下结果等价；有限精度下，MGS 通常更好地保持正交性。

QR 的几何意义：同一个子空间，换成标准正交基

若 $A \in \mathbb{R}^{m \times n}$ 列满秩，Gram-Schmidt 产生

$$\boxed{A = QR}, \quad Q \in \mathbb{R}^{m \times n}, Q^\top Q = I_n, \quad R \in \mathbb{R}^{n \times n} \text{ 可逆上三角}.$$

令

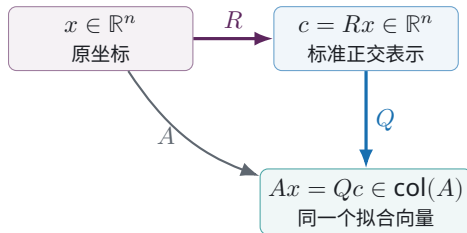
$$\boxed{c := Rx}.$$

因为 R 可逆， $x \leftrightarrow c$ 只是**系数空间中的换坐标**；而

$$Ax = Qc.$$

因此

$$\boxed{\text{col}(A) = \text{col}(Q)}.$$



$A = QR$: R 负责换坐标， Q 用标准正交基表示同一个拟合子空间。

用 QR 解最小二乘：正交投影 + 一个三角系统

由 $A = QR$ ，令 $c = Rx$ ，则

$$\min_x \|Ax - b\|_2^2 = \min_c \|Qc - b\|_2^2.$$

上一页已经把问题变成列向量标准正交的情形，因此

$$\hat{c} = Q^\top b.$$

最后把坐标换回去：

$$R\hat{x} = Q^\top b.$$

给定 QR 分解后，求解阶段只有两步：

1. 计算 $\hat{c} = Q^\top b$;
2. 解上三角系统 $R\hat{x} = \hat{c}$ 。

解 $R\hat{x} = \hat{c} \neq$ 先计算 R^{-1} 。

解线性方程组与矩阵求逆是两回事；后面的线性方程组讲座会专门讨论。

投影矩阵：把 b 正交分解成两个部分

若 Q 的列向量标准正交，定义

$$P := QQ^\top.$$

由上一页的投影系数 $Q^\top b$,

$$\hat{b} = Q(Q^\top b) = Pb, \quad r = b - \hat{b} = (I - P)b.$$

于是每个 b 都有正交分解

$$b = \underbrace{Pb}_{\in \text{col}(A)} + \underbrace{(I - P)b}_{\in \text{col}(A)^\perp}.$$

并且

$$\langle Pb, (I - P)b \rangle = 0.$$

坐标会变；“子空间分量 + 正交残差”只由子空间本身决定。

伪逆：真正能“逆”回来什么？

若 A 列满秩，则每个 b 都有唯一的最小二乘解。把这个线性映射记为

$$\boxed{A^+b := \hat{x}}, \quad \text{其中 } R\hat{x} = Q^\top b \quad (A = QR).$$

这就是 Moore–Penrose 伪逆在列满秩情形下的作用。

从系数出发：完全逆回来

$$x \xrightarrow{A} Ax \xrightarrow{A^+} x, \quad \boxed{A^+A = I_n}.$$

从任意数据出发：只回到可达部分

$$b \xrightarrow{A^+} \hat{x} \xrightarrow{A} Pb, \quad \boxed{AA^+ = P}.$$

作为矩阵， $A^+ = R^{-1}Q^\top = (A^\top A)^{-1}A^\top$ ；这是恒等式，不是要求显式求逆的算法。

另一侧（补充）：方程不足、精确解很多时，同一个伪逆会从所有解中选出**最小范数解**。

正交变换：保持几何，也保持误差尺度

前面的长方矩阵 $Q \in \mathbb{R}^{m \times n}$ 满足

$$Q^{\top} Q = I_n \implies \|Qc\|_2 = \|c\|_2.$$

它把 $\mathbb{R}^n$ 保持长度地映到列空间。

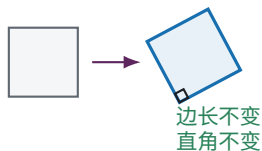
若 $Q \in \mathbb{R}^{n \times n}$ 还是方阵，则

$$Q^{-1} = Q^{\top}, \quad \langle Qx, Qy \rangle = \langle x, y \rangle.$$

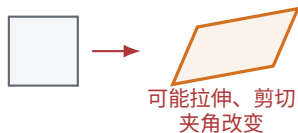
特别地，

$$\|Q \delta x\|_2 = \|\delta x\|_2.$$

正交变换



一般线性变换



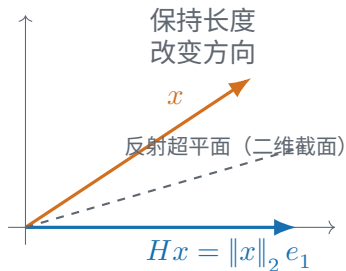
正交变换不拉伸、不剪切；2-范数误差也不会被单个正交变换放大。

实际 QR: Householder 用一次反射消掉一整列的尾部

Gram-Schmidt 最适合解释几何；实际稠密 QR 常用 Householder 反射。给定非零向量 x ，选择正交反射 H ，使

$$Hx = \pm \|x\|_2 e_1 = \begin{pmatrix} * \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

也就是：**一次反射同时消掉当前列的一整段尾部**。对剩余子矩阵递归重复，就得到上三角矩阵 R 。



$$A \xrightarrow{H_1} \begin{bmatrix} * & * \\ 0 & * \\ 0 & * \end{bmatrix} \xrightarrow{H_2, \dots} R$$

CGS 在有限精度下可能明显失去正交性；MGS 会改善这一点，而稠密直接 QR 通常更偏好 Householder。具体公式见补充材料。

从有限维到函数：正交投影的几何来自内积

前半讲的证明没有本质使用“矩阵”；它只需要**向量空间 + 内积**。

抽象内积提供几何

实向量空间上的内积满足

$$\begin{aligned}\langle x, y \rangle &= \langle y, x \rangle, \\ \langle ax + by, z \rangle &= a \langle x, z \rangle + b \langle y, z \rangle, \\ \langle x, x \rangle &> 0 \quad (x \neq 0).\end{aligned}$$

相应地定义长度

$$\|x\| := \sqrt{\langle x, x \rangle}.$$

于是**正交、勾股定理与投影**都具有与 $\mathbb{R}^m$ 中相同的形式。

函数也能有同样的几何吗？

第三讲已经知道函数也是向量，并见过

$$\|f\|_{\infty}, \quad \|f\|_2.$$

但**并非每种范数都来自内积**：

- ▶ L^2 范数来自积分内积；
- ▶ L^{∞} 范数一般没有这套内积几何。

所以最佳 L^2 逼近仍然可以做正交投影。

函数的内积：平方误差就是距离

在连续函数空间 $C[a, b]$ 上，选取合适的正权重 w （使积分存在），定义

$$\langle f, g \rangle_w := \int_a^b f(x)g(x)w(x) dx.$$

它正是 $x^\top y = \sum_i x_i y_i$ 的连续版本：**把求和换成积分**。当 $w \equiv 1$ 时，就是普通的 L^2 内积。更重要的是，

$$\|f - g\|_{2,w}^2 = \langle f - g, f - g \rangle_w = \int_a^b |f(x) - g(x)|^2 w(x) dx.$$

因此最小化平方误差，就是在这个内积几何中找最近点。

在选定的内积下，若

$$\langle \phi_i, \phi_j \rangle_w = 0 \quad (i \neq j),$$

就称 $\{\phi_j\}$ 为**正交函数族**；若再有 $\|\phi_j\|_{2,w} = 1$ ，则称为标准正交函数族。于是“**正交** → **投影** → **最佳平方逼近**”的有限维故事可以原封不动搬过来。

Chebyshev 正交性：权重其实是一种坐标变换

第三讲已经知道 $T_0, \dots, T_n$ 是 $\mathcal{P}_n$ 的一组基，并有 $T_k(\cos \theta) = \cos(k\theta)$ 。现在再看它们为什么在合适的内积下正交。对权重 $w(x) = 1/\sqrt{1-x^2}$ ，令 $x = \cos \theta$ 。

坐标换元

$$dx = -\sin \theta d\theta, \quad \sqrt{1-x^2} = \sin \theta,$$

所以

$$\frac{dx}{\sqrt{1-x^2}} = -d\theta.$$

Chebyshev $\Rightarrow$ 余弦

$$\langle T_i, T_j \rangle_w = \int_0^\pi \cos(i\theta) \cos(j\theta) d\theta.$$

因此

$$\langle T_i, T_j \rangle_w = 0 \quad (i \neq j).$$

$x: -1 \rightarrow 1$ 对应 $\theta: \pi \rightarrow 0$ ，积分上下限同时反转。

也就是说，切比雪夫多项式在这个带权内积下彼此正交。

权重 $1/\sqrt{1-x^2}$ 恰好把 x 坐标变成均匀的角度坐标 θ 。

上一讲的 L^∞ 极小极大与这里的带权 L^2 投影仍是不同的优化问题。

正交系统中的最佳平方逼近

在某个给定内积 $\langle \cdot, \cdot \rangle$ 下，希望在

$$V = \text{span}\{\phi_1, \dots, \phi_n\}$$

中近似 f ，写

$$g = \sum_{k=1}^n c_k \phi_k.$$

若 ϕ_k 两两正交，投影条件 $f - g \perp V$ 给出

$$c_j = \frac{\langle f, \phi_j \rangle}{\langle \phi_j, \phi_j \rangle}.$$

令 $\|u\|^2 := \langle u, u \rangle$ 。对任意 $h \in V$ ，因为 $(f - g) \perp (g - h)$ ，

$$\|f - h\|^2 = \|f - g\|^2 + \|g - h\|^2 \geq \|f - g\|^2.$$

和有限维完全一样：投影条件既给出系数，也证明它真的是最佳逼近。

同一个投影原则：有限维 QR 与函数投影

	有限维向量	函数空间
目标	b	f
基向量	q_j	ϕ_j
正交条件	$q_i^\top q_j = 0 \ (i \neq j)$	$\langle \phi_i, \phi_j \rangle = 0 \ (i \neq j)$
投影系数	$q_j^\top b$	$\frac{\langle f, \phi_j \rangle}{\langle \phi_j, \phi_j \rangle}$
残差	$b - QQ^\top b$	$f - \sum_j c_j \phi_j$

QR 中的 q_j 已经标准化，所以 $\langle q_j, q_j \rangle = 1$ ；右侧函数族这里只假设正交，因此保留分母。

无论“向量”是什么，最小二乘的结构都是正交投影。

傅里叶正交族：不同频率方向互相正交

现在改回 $[-\pi, \pi]$ 上的未加权内积

$$\langle f, g \rangle := \int_{-\pi}^{\pi} f(x)g(x) dx,$$

并把正弦也加入余弦族：

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \dots$$

对 $k, \ell \geq 1$ ，这些方向满足

$$\begin{aligned} \langle 1, 1 \rangle &= 2\pi, & \langle 1, \cos kx \rangle &= \langle 1, \sin kx \rangle = 0, \\ \langle \cos kx, \sin \ell x \rangle &= 0, & \langle \cos kx, \cos \ell x \rangle &= \langle \sin kx, \sin \ell x \rangle = \begin{cases} \pi, & k = \ell, \\ 0, & k \neq \ell. \end{cases} \end{aligned}$$

这组不同函数彼此正交；每个非零频率方向的长度平方都是 π 。

常数函数与非零频率的内积为 0；正弦-余弦交叉项用奇偶性；
同类不同频率用积化和差。详细验证见补充材料。

傅里叶系数 = 投影系数

用有限三角多项式近似

$$p_n(x) = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx).$$

非零频率：上一页给出 $\|\cos kx\|_2^2 = \|\sin kx\|_2^2 = \pi$ ，
所以

$$a_k = \frac{\langle f, \cos kx \rangle}{\pi}, \quad b_k = \frac{\langle f, \sin kx \rangle}{\pi} \quad (k \geq 1).$$

常数项为什么写成 $a_0/2$?

$$\langle 1, 1 \rangle = 2\pi, \quad c_0 = \frac{\langle f, 1 \rangle}{2\pi}.$$

传统傅里叶记号令 $c_0 = a_0/2$ ，于是

$$a_0 = \frac{\langle f, 1 \rangle}{\pi}.$$

因为 $\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x) dx$ ，这就是熟悉的积分公式。

也可从法线方程得到同样结果：正交性使各个系数彼此解耦。

从连续逼近到三角插值：同一空间，不同条件

对 2π 周期函数， $[-\pi, \pi]$ 与 $[0, 2\pi]$ 只是同一周期的平移；为配合等距采样，以下采用后者。

取奇数 $N = 2m + 1$ ，固定同一个 N 维频率空间

$$V_N = \text{span}\{1, \cos x, \sin x, \dots, \cos mx, \sin mx\}.$$

连续平方逼近：知道整个函数 f

$$\min_{p \in V_N} \int_0^{2\pi} |f(x) - p(x)|^2 dx.$$

目标是让整个周期上的平方误差尽量小。

不要求经过指定样本点。

三角插值：只知道 N 个等距样本

$$x_j = \frac{2\pi j}{N}, \quad j = 0, \dots, N-1.$$

记 $f_j := f(x_j)$ 。寻找 $p \in V_N$ ，满足

$$p(x_j) = f_j \quad \forall j.$$

必须精确通过每一个样本点。

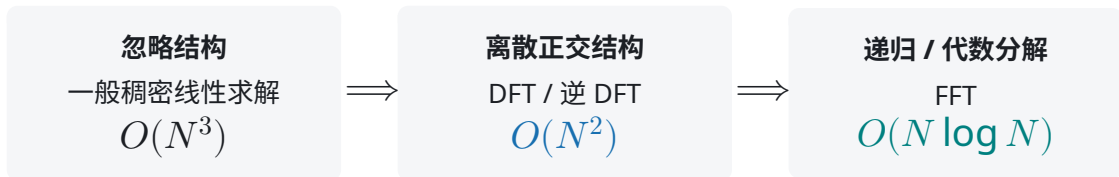
同一个表示空间 V_N ：一个做全局误差最小化，一个做精确插值。

N 个约束与 N 个未知系数只说明得到方阵；为什么它可逆？下一讲用 DFT 看见。

下一讲：从“解线性系统”到 DFT，再到 FFT

上一页的三角插值给出一个 $N \times N$ 的**结构化线性系统**：

等距样本 $f_0, \dots, f_{N-1} \longleftrightarrow N$ 个频率系数.



下一讲把正弦、余弦合并成**复指数频率**，回答两个问题：

1. 为什么变换可逆，而且逆也很简单？

来自等距采样与频率之间的**离散正交性**。

2. 为什么还能更快？

来自更深的**递归 / 代数分解结构**。

DFT 先把“求解”变成结构化变换；FFT 再把这个变换做快。

本讲主线：三个动作贯穿整讲

1. 找最近点 → 投影

$$\hat{b} \in V, \quad b - \hat{b} \perp V.$$

法线方程只是这个几何条件在坐标中的写法。

2. 换成正交表示

$$A = QR, \quad R\hat{x} = Q^T b.$$

标准正交表示让系数解耦；Householder 用保几何的反射实现 QR。

3. 把同一结构搬到函数

$$c_j = \frac{\langle f, \phi_j \rangle}{\langle \phi_j, \phi_j \rangle}.$$

切比雪夫多项式、傅里叶系数都来自同一个正交投影原则。

最小二乘的核心不是法线方程，而是正交投影。

更广的主线：选对表示以后，几何结构会直接变成计算结构；下一讲继续走向 DFT / FFT。

Appendix

欠定系统 · 坐标偏导 · Householder · 内积范数 · 三角恒等式

方程不足时：解很多，选哪一个？

若 $A \in \mathbb{R}^{m \times n}$ 且 $m < n$, $Ax = b$ 可能有无穷多解。一个自然选择是：

$$\min_x \|x\|_2 \quad \text{s.t.} \quad Ax = b.$$

若 A 行满秩，则这一侧的 Moore-Penrose 伪逆为

$$A^+ = A^\top (AA^\top)^{-1}, \quad x_* = A^+ b.$$

- ▶ 作业：验证 $Ax_* = b$;
- ▶ 加分题：证明 x_* 在所有解中长度最短。

当方程有很多解时，**最小范数**提供一种自然的规范化选择。

为什么 $A^\top(AA^\top)^{-1}b$ 是最短解?

设 A 行满秩, 并记

$$x_* = A^\top(AA^\top)^{-1}b.$$

首先 $Ax_* = b$ 。任意另一个解可写成

$$x = x_* + z, \quad Az = 0.$$

又因为 $x_* = A^\top y$, 其中 $y = (AA^\top)^{-1}b$, 所以

$$\langle x_*, z \rangle = y^\top Az = 0.$$

于是

$$\|x\|_2^2 = \|x_*\|_2^2 + \|z\|_2^2 \geq \|x_*\|_2^2.$$

因此 x_* 是唯一的最小范数解。

展开

$$F(x) = \|Ax - b\|_2^2 = x^\top A^\top Ax - 2b^\top Ax + b^\top b.$$

逐坐标求偏导得到

$$\nabla F(x) = 2A^\top Ax - 2A^\top b.$$

令梯度为 0:

$$A^\top Ax = A^\top b.$$

这和方向导数/局部线性化的推导完全一致；后者更能解释**为什么**梯度长成这样。

给定单位向量 v ，定义

$$H = I - 2vv^{\top}.$$

则

$$H^{\top} = H, \quad H^{\top}H = I,$$

所以 H 是正交矩阵；它把沿 v 的分量翻转，而保持 $v^{\perp}$ 中的分量不变。若希望把非零 x 反射到坐标轴，可令

$$v \propto x \mp \|x\|_2 e_1,$$

归一化后就有

$$Hx = \pm \|x\|_2 e_1.$$

因此对矩阵第一列做一次 Householder 反射，就能同时消掉首元素以下的分量；再递归处理右下角子矩阵即可构造 QR。

为什么 ∞ -范数不可能来自内积?

若某个范数来自内积，展开平方立即得到

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

这称为**平行四边形恒等式**。但在 $\mathbb{R}^2$ 的 ∞ -范数中，取

$$x = (1, 1), \quad y = (1, -1).$$

则

$$\|x + y\|_{\infty}^2 + \|x - y\|_{\infty}^2 = 2^2 + 2^2 = 8,$$

而

$$2\|x\|_{\infty}^2 + 2\|y\|_{\infty}^2 = 2 + 2 = 4.$$

因此 $\|\cdot\|_{\infty}$ 不可能由任何内积诱导。 **L^{∞} 范数并非由内积诱导，因此没有 L^2 那套由内积产生的正交投影几何。**

积化和差：

$$\sin y \cos x = \frac{1}{2}(\sin(y-x) + \sin(y+x)),$$

$$\cos y \cos x = \frac{1}{2}(\cos(y-x) + \cos(y+x)),$$

$$\sin y \sin x = \frac{1}{2}(\cos(y-x) - \cos(y+x)).$$

在 $[-\pi, \pi]$ 上，对整数 $k, \ell \geq 1$ ：

- ▶ 常数函数与 $\sin kx, \cos kx$ 的内积为 0；
- ▶ $\sin kx$ 与 $\cos \ell x$ 的内积对任意 k, ℓ 都为 0；
- ▶ 当 $k \neq \ell$ 时，同类不同频率的内积也为 0。

技术恒等式服务于一个几何事实：不同频率是正交方向。