

计算方法

Computational Methods

Lecture 3: Chebyshev 多项式——如何控制最坏误差? Minimax Approximation, Model Spaces, and Least Squares

刘景钺

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

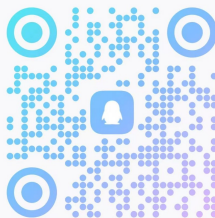
QQ 群 1054896504 加入时注明姓名、专业、
学号

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



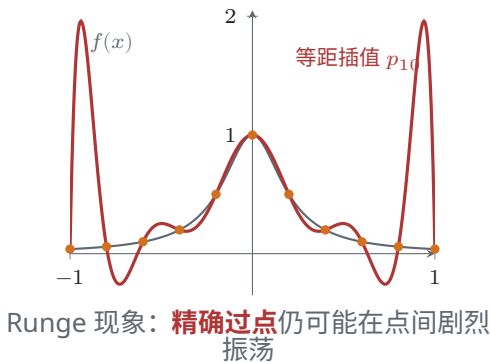
扫一扫二维码, 加入群聊



回顾：上一讲留下了什么问题？

给定 n 个节点 $x_1, \dots, x_n$ ：

- ▶ 存在唯一的次数 $< n$ 的插值多项式；
- ▶ Lagrange 公式给出一种表示；
- ▶ Shamir / Reed-Solomon：冗余让精确重建具备容错性；
- ▶ 但在实数上：精确通过所有点不代表逼近得好。



插值余项：节点如何进入误差？

设 P_f 是 f 在 $x_1, \dots, x_n$ 上的次数 $< n$ 插值多项式，记

$$\omega_n(t) = \prod_{j=1}^n (t - x_j).$$

则对任意 x ，存在位于 $x_1, \dots, x_n, x$ 所张成区间内的 ξ_x ，使

$$f(x) - P_f(x) = \frac{f^{(n)}(\xi_x)}{n!} \omega_n(x).$$

证明思路 (x 不是节点时)：

1. 令 $q(t) = P_f(t) + \lambda \omega_n(t)$ ，其中 $\lambda = \frac{f(x) - P_f(x)}{\omega_n(x)}$ 。于是 $q = f$ 在 $x_1, \dots, x_n, x$ 共 $n+1$ 个点成立。
2. 令 $\phi = f - q$ 。它有 $n+1$ 个零点；反复应用 Rolle 定理，存在 ξ_x 使 $\phi^{(n)}(\xi_x) = 0$ 。
3. 因为 $\deg P_f < n$ 且 ω_n 首一， $q^{(n)} = \lambda n!$ 。因此 $f^{(n)}(\xi_x) = \lambda n!$ ，整理即得余项公式。

节点如何影响误差？全部浓缩在 $\omega_n(x)$ 里。

今天的问题：先选节点，再面对最坏的光滑函数

设 P_f 是 f 在节点 $x_1, \dots, x_n$ 上的次数 $< n$ 插值多项式，记

$$\omega_n(x) = \prod_{j=1}^n (x - x_j), \quad \mathcal{F}_M := \{f \in C^n[-1, 1] : \|f^{(n)}\|_\infty \leq M\}.$$

对任意 $f \in \mathcal{F}_M$ ，插值余项给出

$$\|f - P_f\|_\infty \leq \frac{M}{n!} \|\omega_n\|_\infty.$$

而且这个上界**正好就是最坏情形**：

$$\sup_{f \in \mathcal{F}_M} \|f - P_f\|_\infty = \frac{M}{n!} \|\omega_n\|_\infty.$$

上界来自余项公式；下界取 $f(x) = Mx^n/n!$ 即可达到。

这不是在说“对每个固定 f 都最优”；

我们优化的是整个光滑函数类的最坏情形。

选点问题：先放松，再证明这个放松其实是紧的

$$\omega_n(x) := \prod_{j=1}^n (x - x_j).$$

因此，对上一页的光滑函数类，节点设计问题**恰好化为**

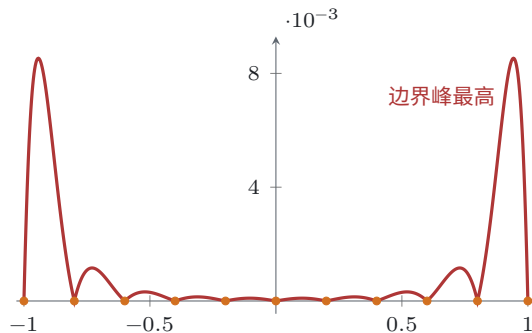
$$\min_{\substack{x_1, \dots, x_n \in [-1, 1] \\ x_i \neq x_j \ (i \neq j)}} \|\omega_n\|_\infty.$$

每个 ω_n 都是首一 n 次多项式，于是再把候选集合**放大**：

$$\min_{\substack{p: \deg p = n \\ [x^n]p = 1}} \|p\|_\infty \quad (\text{允许所有首一 } n \text{ 次多项式}).$$

放松后可行域更大，因此最优值**不会增大**：它给出原问题的下界。 若放松问题存在一个最优解，且它**恰好有 n 个位于 $[-1, 1]$ 的实根**，那么它同时也解出了原来的选点问题。

等距点的问题，真的出在边界附近



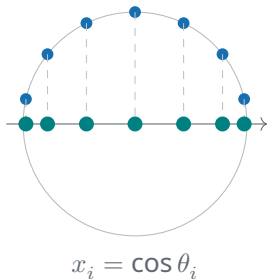
对 11 个等距节点，直接画出

$$|\omega_{11}(x)| = \prod_{j=1}^{11} |x - x_j|.$$

- ▶ 最大值不是平均分布的；
- ▶ **边界附近的峰最高；**
- ▶ 要压低最坏值，节点应当向边界**聚集**。

这提示我们：要控制最高峰，应该设法把各处的峰尽量“压平”。

Chebyshev 节点：在角度上等距，在 x 轴上不等距



蓝点在角度上等距；绿色投影在边界附近更密。

角度等距

$$\theta_i = \frac{(2i-1)\pi}{2n}, \quad i = 1, \dots, n.$$

投影到 x 轴

$$x_i = \cos \frac{(2i-1)\pi}{2n}.$$

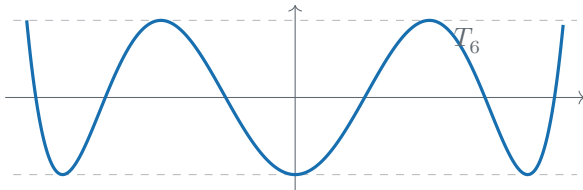
靠近 $0, \pi$ 的角度投影后会自然**挤到边界**，正好回应上一页“边界峰最高”的现象。

在这些节点上做 Lagrange 插值，得到的就是 **Chebyshev 插值多项式**。

Chebyshev 多项式：把圆上的均匀振荡搬到区间上

令 $x = \cos \theta$ ，定义

$$T_n(x) = \cos(n\theta) = \cos(n \arccos x), \quad x \in [-1, 1].$$



背景 在变量替换 $x = \cos \theta$ 下， $T_k(x) = \cos(k\theta)$ ：**Chebyshev 多项式对应余弦基函数。**

为什么 $\cos(n \arccos x)$ 是多项式？ 而且还能快速算？

先看代数：

倍角公式说明 $\cos(n\theta)$ 可以写成 $\cos \theta$ 的 n 次多项式。令 $x = \cos \theta$ ：

$$T_0 = 1, \quad T_1 = x, \quad T_2 = 2x^2 - 1.$$

更有用的是递推：

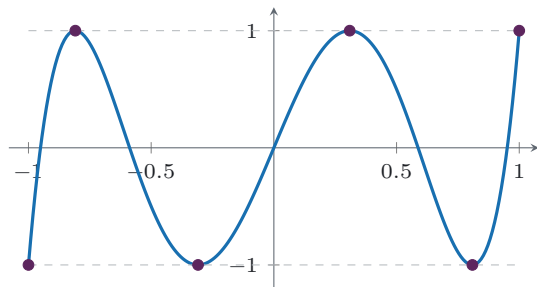
$$\begin{aligned} \cos((n+1)\theta) + \cos((n-1)\theta) \\ = 2 \cos(n\theta) \cos \theta. \end{aligned}$$

所以

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).$$

几何上来自余弦；代数上是多项式；计算上只需递推。

最关键的几何：把“最高峰”平均摊开



因为 $T_n(x) = \cos(n\theta)$:

- ▶ $|T_n(x)| \leq 1$;
- ▶ $T_n(\pm 1) = (\pm 1)^n$;
- ▶ 在 $n + 1$ 个点上交替达到 ± 1 ; 所有最高峰都**一样高**。

**这就是等振荡
(equioscillation)。**

两套余弦点：一个构造答案，一个证明最优

零点：构造答案 (construction)

$$x_i = \cos \frac{(2i-1)\pi}{2n}, \quad i = 1, \dots, n.$$

它们满足 $T_n(x_i) = 0$ ，因此正是**插值节点**。

极值点：最优性证书 (certificate)

$$y_k = \cos \frac{k\pi}{n}, \quad k = 0, \dots, n.$$

它们满足 $T_n(y_k) = (-1)^k$ ，给出 $n+1$ 个**交替见证点**。

零点告诉我们“怎么选”；极值点告诉我们“为什么不可能更好”。

背景 这种“给出构造 + 给出最优性证书”的结构，后面的线性规划对偶还会再出现。

归一化：把 Chebyshev 送进“首一多项式”的可行域

先把振幅归一化

由递推式可归纳

$$[x^n]T_n = 2^{n-1}.$$

所以定义

$$\widehat{T}_n(x) := 2^{1-n}T_n(x)$$

后, $\widehat{T}_n$ 是**首一 n 次多项式**, 且

$$\|\widehat{T}_n\|_{\infty} = 2^{1-n}.$$

零点仍然是合法节点

缩放不会改变零点, 因此

$$\widehat{T}_n(x_i) = 0, \quad x_i = \cos \frac{(2i-1)\pi}{2n}.$$

因为 $\widehat{T}_n$ 首一, 立刻有

$$\widehat{T}_n(x) = \prod_{i=1}^n (x - x_i) = \omega_n(x).$$

可行候选已经有了。 极值点仍交替达到 $\pm 2^{1-n}$; 下一步只需证明: **任何首一 n 次多项式都不可能更矮。**

Chebyshev 极小极大定理：下界 + 达到下界

任何首一候选都绕不过的下界

对任意首一 n 次多项式 p ,

$$\max_{x \in [-1, 1]} |p(x)| \geq 2^{1-n}.$$

Chebyshev 候选恰好达到下界

对 $\hat{T}_n = T_n/2^{n-1}$,

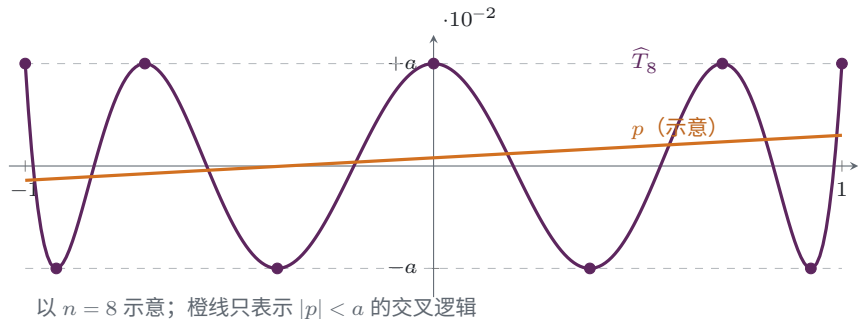
$$\max_{x \in [-1, 1]} |\hat{T}_n(x)| = 2^{1-n}.$$

因此 $\hat{T}_n$ 是最优解：普遍下界 + 匹配构造。

下一页：为什么 $n + 1$ 个交替极值点足以证明这个下界？

证明的核心: $n + 1$ 个交替极值点是一张“最优性证书”

设 $a = 2^{1-n}$, 反设存在更矮的首一候选 p , 即 $\|p\|_\infty < a$ 。



在每一对相邻极值点之间, $p - \hat{T}_n$ 的符号都被迫翻转, **两条曲线至少交叉一次。**

证明：一个更矮的候选会被迫产生很多零点

反设存在首一 n 次多项式 p_n ，使

$$|p_n(x)| < 2^{1-n} \quad \forall x \in [-1, 1].$$

令

$$r(x) = p_n(x) - \widehat{T}_n(x).$$

在 $n+1$ 个极值点 $y_k = \cos(k\pi/n)$ 上，

$$\operatorname{sgn} r(y_k) = -(-1)^k,$$

所以 $r(y_0), \dots, r(y_n)$ 正负交替。由连续性，每一对相邻极值点之间至少有一个零点：

r 至少有 n 个不同零点.

证明：但首一项抵消后，次数最多只有 $n - 1$

p_n 与 $\hat{T}_n$ 都是首一 n 次多项式，因此最高次项抵消：

$$\deg r \leq n - 1.$$

但上一页已经推出 r 至少有 n 个不同零点。

非零的 $n - 1$ 次多项式至多有 $n - 1$ 个根。 **矛盾！**

上一讲的老朋友又出现了：次数限制零点数。

因此 $\hat{T}_n$ 的最坏振幅已经不可能再降低。

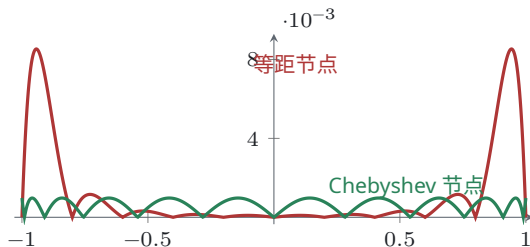
回到最初的问题：最优值 + 一个最优节点设计

把前面的两步合在一起：对任意 $M \geq 0$,

$$\min_{\substack{x_1, \dots, x_n \in [-1, 1] \\ x_i \neq x_j}} \sup_{f \in \mathcal{F}_M} \|f - P_f\|_{\infty} = \frac{M}{n!} 2^{1-n}$$

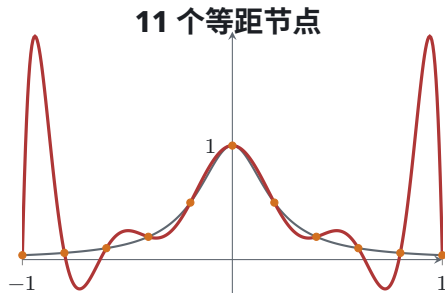
而一组最优节点是

$$x_i = \cos \frac{(2i-1)\pi}{2n}, \quad i = 1, \dots, n.$$

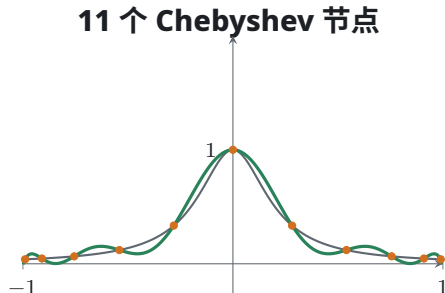


- ▶ 固定节点后，函数类的最坏误差**恰好由节点因子决定**；
- ▶ 再用 Chebyshev 极小极大定理解节点因子；
- ▶ 最后发现放松解本身就有合法实根：**放松是紧的**。

Runge 现象：同样是 11 个节点，只改变位置，误差差一个数量级



$$\max |f - p_{10}| \approx 1.92$$



$$\max |f - p_{10}| \approx 0.11$$

灰线: $f(x) = 1/(1 + 25x^2)$; 彩线: 次数 10 的插值多项式; 橙点: 实际插值节点。
这里只是一个具体例子; 前面的定理是函数类上的最坏情形保证, 不是对每个固定函数逐一比较。

从插值到逼近：不再只问“过不过点”，而要先做两次选择

插值 (Interpolation)

- ▶ 必须经过给定的所有点；
- ▶ 有些应用确实需要精确通过样本；
- ▶ 选点会强烈影响误差；
- ▶ 也可以改用**分段插值**。

函数逼近 (Approximation)

1. 先选允许答案落入的**模型空间** V ；
2. 再选怎样衡量误差的**范数** $\|\cdot\|$ ；
3. 然后求

$$\min_{g \in V} \|f - g\|.$$

逼近问题 = 模型空间 + 误差几何.

一致逼近：固定模型空间，再最小化最坏误差

次数 $< n$ 的多项式构成模型空间

$$\mathcal{P}_{n-1} = \{p : \deg p < n\}.$$

在区间 $[a, b]$ 上用无穷范数衡量误差：

$$\|g\|_{\infty} := \sup_{x \in [a, b]} |g(x)|.$$

于是最佳一致逼近误差就是

$$E_{n-1}(f) := \min_{p \in \mathcal{P}_{n-1}} \|f - p\|_{\infty}.$$

同样用 ∞ -范数看**最坏点**，但优化对象不同：
前半讲：**节点 / 插值规则** 这里：**多项式** $p \in \mathcal{P}_{n-1}$ 。

Weierstrass: 表达能力足够强, 但没有告诉我们怎样算

Weierstrass 逼近定理. 若 $f \in C[a, b]$, 则

$$E_{n-1}(f) \longrightarrow 0 \quad (n \rightarrow \infty).$$

等价地: 对任意 $\varepsilon > 0$, 总有某个多项式 p 使 $\|f - p\|_{\infty} < \varepsilon$ 。

它回答: 多项式**能不能**逼近?

它没有回答: 给定次数 / 给定选点规则,
哪个算法会逼近?

所以 **Runge 与 Weierstrass 不矛盾**: “存在很好的多项式” $\nRightarrow$ “等距插值得到的多项式会收敛”。

Lebesgue 常数：稳定性与近似质量由同一个因子控制

前半讲用 $\|f^{(n)}\|_\infty$ 控制绝对误差；这里换一个问题：不用导数，只比较“离最佳多项式有多远”。

$$I_n f(x) = \sum_{j=1}^n f(x_j) \ell_j(x),$$

$$\Lambda_n := \max_{x \in [-1, 1]} \sum_{j=1}^n |\ell_j(x)|.$$

背景 算子视角： $I_n : C[-1, 1] \rightarrow \mathcal{P}_{n-1}$ 是一个投影 ($I_n p = p$ 对所有 $p \in \mathcal{P}_{n-1}$)，且 $\|I_n\|_{\infty \rightarrow \infty} = \Lambda_n$ 。

数据扰动会被放大多少？

若节点值扰动 $y_j \mapsto y_j + \delta_j$ ，则

$$\|\Delta p\|_\infty \leq \Lambda_n \max_j |\delta_j|.$$

Λ_n 就是从节点值误差到函数误差的最坏放大倍数。

算法离最佳逼近有多远？

若 $E_{n-1}(f) = \min_{p \in \mathcal{P}_{n-1}} \|f - p\|_\infty$ ，
则

$$\|f - I_n f\|_\infty \leq (1 + \Lambda_n) E_{n-1}(f).$$

同一个放大因子也控制相对最佳逼近的损失。

Chebyshev 节点： $\Lambda_n = O(\log n)$ ；

等距节点： Λ_n 随 n **指数增长**。

更快的算法

- ▶ 更快地模拟随机游走；
- ▶ 迭代求解线性方程组：
Chebyshev 迭代 / 共轭梯度法；
- ▶ 最大特征值 / 特征方向：Lanczos 迭代；
- ▶ ...

复杂性理论

- ▶ 感知机表示奇偶函数的复杂性；
- ▶ 电路复杂度；
- ▶ 通信复杂度；
- ▶ ...

共同套路：先用低次多项式逼近一个标量函数，再把这个多项式变成矩阵上的算法。

接下来要补齐统一语言： 允许答案落在哪个空间？ 怎样衡量“误差”？

先回答第一个问题：允许的候选可以组成向量空间

典型的线性空间

$$\mathbb{R}^n, \quad \mathcal{P}_{n-1}, \quad \mathbb{R}[x],$$

$$C[a, b], \quad C^k[a, b].$$

这些对象看起来不同，但都允许

$$u + v, \quad cu$$

并且运算后仍留在同一个集合里。

“线性相关”也完全是同一句话

函数族 $\phi_1, \dots, \phi_k$ 线性相关，当且仅当

$$\exists (c_1, \dots, c_k) \neq 0 : \quad \sum_{i=1}^k c_i \phi_i(x) = 0, \quad \forall x.$$

否则线性无关。函数不是“像”向量；

这些运算一经固定，它们**就是**空间里的向量。

于是“允许答案落在哪个空间？”可以统一成一句话：**先选模型空间 V 。**

选定模型空间后，还要选表示：没有一组基对所有对象都“最好”

同一个空间

$$\mathcal{P}_8 = \text{span}\{1, x, \dots, x^8\} = \text{span}\{T_0, T_1, \dots, T_8\}$$

可以有完全不同的坐标。

对象一： T_8

$$T_8 = 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1.$$

单项式基：多个非零系数； Chebyshev 基：只有一个非零坐标。

对象二： x^8

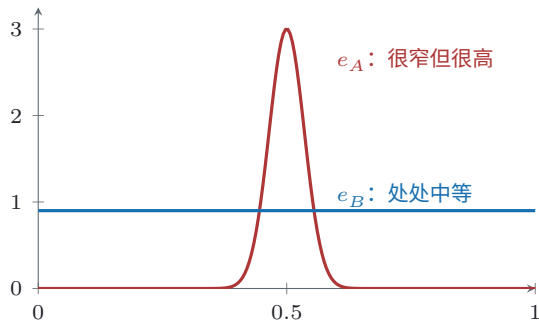
$$x^8 = \frac{T_8 + 8T_6 + 28T_4 + 56T_2 + 35T_0}{128}.$$

单项式基：只有一个非零坐标； Chebyshev 基：反而展开成多项。

基没有绝对的优劣：好的表示，要让当前对象或运算的结构变得简单、可见。

为什么 $T_0, \dots, T_8$ 确实是一组基？由 $\deg T_k = k$ 得到三角换基矩阵；证明见补充。

再回答第二个问题：换一把“尺子”，会反转哪个近似更好



最坏误差看：

$$\|e_A\|_{\infty} = 3 > 0.9 = \|e_B\|_{\infty}.$$

但按 L^2 误差衡量：

$$\|e_A\|_2 \approx 0.75 < 0.9 = \|e_B\|_2.$$

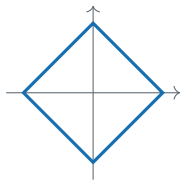
∞ -范数选 e_B ；

L^2 -范数反而选 e_A 。

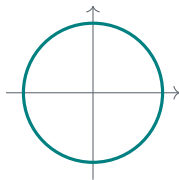
选择范数不是换一个记号，而是在改变优化问题本身。

同一个 $\mathbb{R}^2$ ，换范数就是换几何

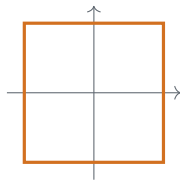
范数是满足**正定性**、**齐次性**、**三角不等式**的“长度”。对 $x = (x_1, x_2)^\top$ ：



$$\|x\|_1 = |x_1| + |x_2|$$



$$\|x\|_2 = \sqrt{x_1^2 + x_2^2}$$



$$\|x\|_\infty = \max_i |x_i|$$

对 $x \in \mathbb{R}^n$ ，更一般地， $1 \leq p < \infty$ ：

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

函数空间也可以选择不同的误差几何

对 $f : [a, b] \rightarrow \mathbb{R}$:

∞ -范数：最坏点

$$\|f\|_{\infty} = \sup_{x \in [a, b]} |f(x)|.$$

Chebyshev / 极小极大逼近关心的是这种几何。

L^2 -范数：整体误差

$$\|f\|_2 = \left(\int_a^b |f(x)|^2 dx \right)^{1/2}.$$

最小二乘最终最小化的是它的平方
 $\|f - g\|_2^2$ 。

同一个模型空间 V ，只改范数，就会得到**不同的最佳近似**。

为什么 L^2 -范数在线性逼近里特别好用？因为它来自内积

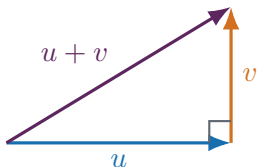
欧氏 2-范数满足

$$\|u + v\|_2^2 = \|u\|_2^2 + 2\langle u, v \rangle + \|v\|_2^2.$$

函数空间完全同理： $\langle f, g \rangle = \int_a^b f(x)g(x) dx$ ，且 $\|f\|_2^2 = \langle f, f \rangle$ 。

当 $u \perp v$ 时，中间的交叉项消失：

$$\|u + v\|_2^2 = \|u\|_2^2 + \|v\|_2^2.$$



对于线性子空间里的最佳逼近，这个结构尤其关键：

- ▶ 距离平方可以展开；
- ▶ “最短”可以转化成**正交**；
- ▶ 正交条件最终会变成线性代数。

下一讲会证明：对线性子空间 V ，最近点的残差恰好与 V 正交；完整证明见补充。

从线性方程到最小二乘：一个统一的优化视角

精确方程 $Ax = b$

$$\{x : Ax = b\} = \begin{cases} \emptyset, & b \notin \text{col}(A), \\ x_0 + \ker(A), & b \in \text{col}(A). \end{cases}$$

因此：

- ▶ $b \notin \text{col}(A)$ ：无解；
- ▶ $b \in \text{col}(A)$ 时：
 - $\ker(A) = \{0\}$ ：唯一解；
 - $\ker(A) \neq \{0\}$ ：无穷多解。

一旦有两个不同解，就已有非零核向量，因此自动有无穷多解。

统一成一个优化问题

无论是否存在精确解，都可以求

$$x^* \in \arg \min_x \|Ax - b\|_2^2.$$

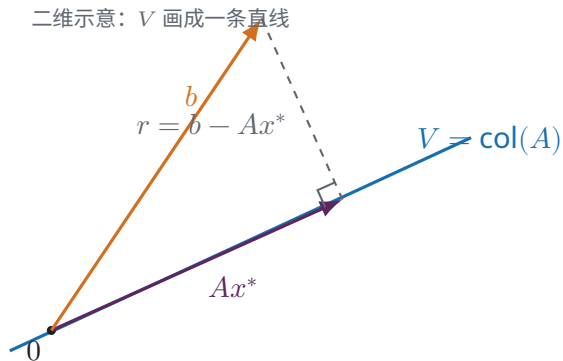
所有可能的 Ax 正好组成列空间

$$\text{col}(A) = \text{span}\{A_1, \dots, A_n\}.$$

- ▶ $b \in \text{col}(A)$ ：最小值为 0，最优解恰好就是精确解；
- ▶ $b \notin \text{col}(A)$ ：最小值 > 0 ，找列空间中离 b 最近的向量 Ax^* 。

注意两个层次：最近的向量 Ax^* 是唯一的；若 $\ker(A) \neq \{0\}$ ，实现它的参数 x^* 仍可能不唯一。

最小二乘：下一讲真正要解决的三个问题



1. “残差正交” 怎样变成关于 x 的**方程**？
2. 怎样用**正交基**直接算投影？
3. 为什么 QR 比直接形成 $A^T A$ 更**稳定**？

下一讲：最小二乘 → 正交投影 → **Gram-Schmidt** → **QR**.

本讲主线：先问“优化谁、怎样衡量误差”，再问“怎样算”

优化问题 = 可行对象 / 模型空间 + 误差几何

Chebyshev 节点设计

决策变量：节点 $x_1, \dots, x_n$



误差：函数类上的最坏 L^∞ 误差



等振荡 + 最优性证书



最优节点设计

最小二乘

决策变量： $Ax \in \text{col}(A)$



误差： $\|Ax - b\|_2^2$



内积 + 正交结构



最近点 / 正交投影

定义清楚“最好”以后，计算还要再选表示 / 基：

下一讲：正交基 $\rightarrow$ Gram-Schmidt $\rightarrow$ QR； 随后：傅里叶级数 / 三角插值 $\rightarrow$ FFT。

先明确决策变量与误差，再选择合适的表示与算法，把目标稳定、快速地算出来。

Appendix

区间缩放 · Lebesgue 常数 · 基与正交 · 范数补充

原讲义给出的三个方向：

- ▶ *The Polynomial Method in Quantum and Classical Computing;*
- ▶ *Faster Algorithms via Approximation Theory;*
- ▶ *Chebyshev Polynomials and Approximation Theory in Theoretical Computer Science and Algorithm Design.*

课外阅读，不要求掌握；核心目标只有一个：把“逼近”看成算法设计语言，而不是纯粹的函数分析问题。

补充 从 $[a, b]$ 到 $[-1, 1]$: 误差因子如何缩放?

作仿射变换

$$x = \frac{a+b}{2} + \frac{b-a}{2}t, \quad x_j = \frac{a+b}{2} + \frac{b-a}{2}t_j.$$

于是

$$\prod_{j=1}^n (x - x_j) = \left(\frac{b-a}{2}\right)^n \prod_{j=1}^n (t - t_j).$$

因此 $[-1, 1]$ 上的 Chebyshev 节点搬回 $[a, b]$ 后仍然最优; 最小节点因子为

$$\left[\left(\frac{b-a}{2}\right)^n 2^{1-n} \right].$$

补充 为什么插值误差至多比最佳逼近多一个 Lebesgue 常数?

插值算子 I_n 满足 $I_n p = p$ 对所有 $p \in \mathcal{P}_{n-1}$ 。因此任取 $p \in \mathcal{P}_{n-1}$ ：

$$\begin{aligned}\|f - I_n f\|_\infty &= \|(f - p) - I_n(f - p)\|_\infty \\ &\leq \|f - p\|_\infty + \|I_n(f - p)\|_\infty \\ &\leq (1 + \|I_n\|) \|f - p\|_\infty.\end{aligned}$$

取最优 p 得

$$\|f - I_n f\|_\infty \leq (1 + \Lambda_n) E_{n-1}(f), \quad \Lambda_n := \|I_n\|.$$

Chebyshev 节点满足 $\Lambda_n = O(\log n)$ 。

为什么 $T_0, \dots, T_{n-1}$ 构成 $\mathcal{P}_{n-1}$ 的一组基?

把 T_k 写成单项式基:

$$T_k(x) = 2^{k-1}x^k + \text{lower-degree terms} \quad (k \geq 1).$$

因此两组有序函数

$$(1, x, x^2, \dots, x^{n-1}) \quad \longleftrightarrow \quad (T_0, T_1, \dots, T_{n-1})$$

之间的换基矩阵是一个**对角元非零的三角矩阵**, 所以可逆。两组函数都线性无关, 并且生成同一个 n 维多项式空间。

最近点为什么等价于“残差正交”？

设 V 是线性子空间, $y^* \in V$, 残差 $r = b - y^*$ 。

若 y^* 是离 b 最近的点, 则对任意 $v \in V$, $t = 0$ 必须最小化

$$\begin{aligned}\phi(t) &= \|b - (y^* + tv)\|_2^2 \\ &= \|r - tv\|_2^2 = \|r\|_2^2 - 2t\langle r, v \rangle + t^2 \|v\|_2^2.\end{aligned}$$

因此

$$\phi'(0) = -2\langle r, v \rangle = 0 \quad \forall v \in V,$$

即 $r \perp V$ 。

反过来, 若 $r \perp V$, 则对任意 $v \in V$,

$$\|r - v\|_2^2 = \|r\|_2^2 + \|v\|_2^2 \geq \|r\|_2^2.$$

最近点 $\iff$ 残差 $\perp V$.

补充 p -范数之间没有谁永远“更正确”

对 $x \in \mathbb{R}^n$, $1 \leq p < \infty$:

$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{1/p}.$$

有限维中所有范数在拓扑意义下等价，但常数会依赖维数。例如

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n} \|x\|_\infty.$$

算法里真正重要的不是“范数是否等价”，
而是**误差界中的常数如何随维数增长**。