§5 最佳近似和最小二乘

Def 5.1 (最佳估计) 给出了定义了距离的线性函数空间 Y 和它的子空间 $X\subset Y$, 称函数 $\hat{\varphi}\in X$ 是(距离定义为 $\|\cdot\|$ 下) $f\in Y$ 在 X 中的最佳估计, 当且仅当 $\forall\varphi\in X$ $\|f-\hat{\varphi}\|\le\|f-\varphi\|$

Ex 5.2 [Thm 2.4.6 切比雪夫定理] 给出了在模长定义为 $\|g\|_\infty=\max\limits_{x\in[-1,1]}|g(x)|$ 下, $f(x)=-x^n\in\mathbb{P}_n(\mathbb{R})$ 在 $\mathbb{P}_{n-1}(\mathbb{R})$ 中的最佳估计是 $\hat{\varphi}(x)=\frac{T_n(x)}{2^{n-1}}-x^n$, T_n 是 n 次切比雪夫多项式.

Def 5.3 (基本线性估计问题) 给定函数 $f\in Y$ 和 n 个 X 中元素 $u_1,u_2,\dots,u_n\in X\subset Y$ 寻找形如 $\varphi=\sum_{i=1}^{n}a_iu_i$ 的最佳估计. $a_i\in\mathbb{R}$.

Ex 5.4 对于 $f(x)=e^x\in C^\infty[-1,1]$, 寻找形如 $\varphi=\sum_{i=1}^{n}a_iu_i\in X=\text{span}\{1,x,x^2,\dots\}$ 的最佳估计, 其中 n 为正整数, $\|\cdot\|$ 定义为 ∞-范数 $\|\cdot\|_\infty$, 1-范数或 2-范数, $\|g\|_1=\int_{-1}^{1}|g(x)|dx$, $\|g\|_2=(\int_{-1}^{1}|g(x)|^2dx)^{\frac12}$

注: 三种范数对应不同的误差, $\|\cdot\|_\infty$ 对应最小-最大误差, $\|\cdot\|_1$ 对应 $g(x)$ 与 x 轴围成的面积 (平均误差), $\|\cdot\|_2$ 对应 $g(x)$ 与 0 的 (函数空间下的) 欧氏距离.

Ex 5.5 对简单闭曲线 $r:[0,1)\to\mathbb{R}^2$ 和曲线上的 n 个点 $x_i\in r$, 考虑样条估计 $p:[0,1)\to\mathbb{R}^2$, 结点为 x_i, 总长度单位化, 记 $\text{Int}(r)$ 为闭曲线 r 围成的区域, 记 $S_1=\text{Int}(r)$, $S_2=\text{Int}(p)$ 的面积差为 $\|S_1\oplus S_2\|=\int_{S_1\oplus S_2}d\vec{x}$, 其中 $S_1\oplus S_2=S_1\cup S_2\setminus(S_1\cap S_2)$.

注: 称 $\|S_1\oplus S_2\|$ 最小的样条估计为 r 在 1-范数下的最佳估计.

Thm 5.6 设 X 是内积空间 $(Y,\|\cdot\|)$ 的有限维子空间, 则有 $\forall y\in Y$, 定义 $B_y=\{x\in X:\|x\|\le 2\|y\|\}$, 显然 $0\in B_y$, 而 y 到 B_y 的距离为 $\text{dist}(y,B_y)=\inf\limits_{x\in B_y}\|y-x\|\le\|y-0\|=\|y\|$, $\Rightarrow\exists\hat{\varphi}\in X$, s.t. $\forall\varphi\in X$, $\|\hat{\varphi}-y\|\le\|\varphi-y\|$.

由定义 $\forall z\in X$, $z\notin B_y$, $\|z\|>2\|y\|$, $\|z-y\|\ge\|z\|-\|y\|>\|y\|$

$\therefore$ 若 y 的最佳估计存在, 一定在 B_y 中, $\because B_y\subset X$ $\therefore B_y$ 是有限维的有界闭集, 因此是紧集

由连续函数在紧集上的极值定理, $d(x)=\|x-y\|$ 在 B_y 上取得极小值, 极小值点为 $\hat{\varphi}\in X$.

Thm 5.7 集合 $C[a,b]$ ($[a,b]$ 上全体连续函数) 是 $\mathbb{C}$ 上的内积空间, 内积运算定义为 $\langle u,v\rangle=\int_a^b\rho(t)u(t)\overline{v(t)}dt$ 其中 $\overline{v(t)}$ 是 $v(t)$ 的复共轭函数, 权重函数 $\rho(x)\in C[a,b]$ 满足 $\rho(x)>0$, 且 $\|u\|_2=(\int_a^b\rho(t)|u(t)|^2dt)^{\frac12}$

Def 5.8 称 $C[a,b]$ 上函数的最小均方估计是模长定义同 Thm 5.7 的最佳估计.

§5.1 正交系统

Def 5.9 称内积空间 X 的子集 S 是正交的, 如果 $\forall u,v\in S$, $\langle u,v\rangle=\begin{cases}0, & u\ne v\\1, & u=v\end{cases}$.

Ex 5.10 $\mathbb{R}^n$ 上的标准基是正交的.

Ex 5.11 Def 2.4.1 中 $\{T_n(x)=\cos(n\arccos x)\}$ 是正交集, 其中 $[a,b]=[-1,1]$, $\rho=\frac{1}{\sqrt{1-x^2}}$.

Thm 5.12 任何有限正交集的元素线性无关.

Def 5.13 Schmidt 正交化可将有限/无限线性无关组 $(u_1,u_2,\dots)$ 转化为 $(u_1^*,u_2^*,\dots)$

$v_{n+1}=u_{n+1}-\sum_{k=1}^{n}\langle u_{n+1},u_k^*\rangle u_k^*$, $u_{n+1}^*=\frac{v_{n+1}}{\|v_{n+1}\|}$

Thm 5.14 对有限/无限的线性无关组 $(u_1, u_2, \dots)$，斯密特正交化过程给出了系数

$$\begin{matrix} a_{11} \\ a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} \\ \vdots \end{matrix} \quad s.t.\ a_{kk}=\frac{1}{\|v_k\|}>0,\ \text{且} \begin{cases} u_1^*=a_{11}u_1 \\ u_2^*=a_{21}u_1+a_{22}u_2 \\ \dots \end{cases} \text{是正交的.}$$

PF: 由Def 5.13归纳可知 u_k^* 是 $u_k, \dots u_1$ 的线性组合，下用归纳法证明正交性，若 $\forall m<n$，$k\le m$，$\langle u_m^*, u_k^* \rangle = \begin{cases} 1, k=m \\ 0, k<m \end{cases}$

因为 $u_n^*=\frac{1}{\|v_n\|}(u_n-\sum_{j=1}^{n-1}\langle u_n, u_j^*\rangle u_j^*)$，所以 $\forall k<n$，

$\langle u_n^*, u_k^* \rangle = \langle \frac{1}{\|v_n\|}(u_n-\sum_{j=1}^{n-1}\langle u_n, u_j^*\rangle u_j^*), u_k^* \rangle = \frac{1}{\|v_n\|}(\langle u_n, u_k^*\rangle-\sum_{j=1}^{n-1}\langle u_n, u_j^*\rangle\langle u_j^*, u_k^*\rangle)$

$= \frac{1}{\|v_n\|}(\langle u_n, u_k^*\rangle-\langle u_n, u_k^*\rangle\langle u_k^*, u_k^*\rangle)=0$，$\langle u_n^*, u_n^*\rangle=\|u_n^*\|=1$

Cor 5.16 对有限/无限的线性无关组，可找到系数

$$\begin{matrix} b_{11} \\ b_{21} & b_{22} \\ \dots \end{matrix} \ \text{和正交集}\ (u_1^*, u_2^*, \dots)\ s.t.\ b_{ij}>0\ \text{且} \begin{cases} u_1=b_{11}u_1^* \\ u_2=b_{21}u_1^*+b_{22}u_2^* \\ \dots \end{cases}$$

PF: $u^*=Au$，A为下三角矩阵，且对角元为正 $\Rightarrow A$ 可逆 $\Rightarrow u=A^{-1}u^*$

Cor 5.17 在Thm 5.14中，$\forall i=1,2,\dots,n-1$，$\langle u_n^*, u_i\rangle=0$

Def 5.18 在Thm 5.7的内积定义下，对线性无关组 $(1, x, x^2, \dots)$ 进行Schmidt正交化过程，

分别在 $a, b, \rho(x)$ 以如下定义，可得到如下标准多项式：

a	b	$\rho(x)$	正交多项式	
-1	1	$1/\sqrt{1-x^2}$	第1类切比雪夫多项式	
-1	1	$\sqrt{1-x^2}$	第2类切比雪夫多项式	
-1	1	1	勒让德多项式	
-1	1	$(1-x)^\alpha(1+\beta)^\beta$	亚可比多项式	$(\alpha,\beta>-1)$
0	$+\infty$	$x^\alpha e^{-x}$	拉盖尔多项式	$(\alpha>-1)$
$-\infty$	$+\infty$	e^{-x^2}	厄米特多项式	

Ex 5.19 勒让德多项式的前三项.

$u_1=1, v_1=1$，$\|v_1\|^2=\int_{-1}^{1}dx=2$　$u_1^*=\frac{1}{\sqrt{2}}$

$u_2=x$，$v_2=x-\langle x, \frac{1}{\sqrt{2}}\rangle\frac{1}{\sqrt{2}}=x$　$\|v_2\|^2=\int_{-1}^{1}x\,dx=\frac{2}{3}$　$u_2^*=\sqrt{\frac{3}{2}}x$

$u_3=x^2$，$v_3=x^2-\langle x^2, \sqrt{\frac{3}{2}}x\rangle\sqrt{\frac{3}{2}}x-\langle x^2, \frac{1}{\sqrt{2}}\rangle\frac{1}{\sqrt{2}}=x^2-\frac{1}{3}$　$\|v_3\|^2=\int_{-1}^{1}(x^2-\frac{1}{3})dx=\frac{8}{45}$　$\therefore u_3^*=\sqrt{\frac{45}{8}}(x^2-\frac{1}{3})$

§5.12 傅立叶展开

Def 5.20 设 (u_1^*, u_2^*) 是有限/无限的正交集，称 w 的正交展开或傅立叶展开为 $\sum_{n=1}^{m}\langle w, u_n^*\rangle u_n^*$ $(m\le+\infty)$

其中系数 $\langle w, u_n^*\rangle$ 称为傅里叶系数，$\langle w, u_n^*\rangle u_n^*$ 称为 w 在 u_n^* 上的投影，称为傅里叶展开的误差为 $\sum_{n=1}^{m}\langle w, u_n^*\rangle u_n^*-w$

Ex 5.21 在欧氏内积下，$\mathbb{R}^3$ 的正交集为 $u_1^*=(1,0,0)^T$，$u_2^*=(0,1,0)^T$，$u_3^*=(0,0,1)^T$

对向量 $w=(a,b,c)^T$，其傅里叶系数为 a,b,c，傅立叶展开为 $w=au_1^*+bu_2^*+cu_3^*$，误差为 0

Ex 5.22 对 $L^2_{\rho=1}[-\pi,\pi]$ 上的正交集 $\frac{1}{\sqrt{2\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos x}{\sqrt{\pi}}, \dots, \frac{\sin nx}{\sqrt{\pi}}, \frac{\cos nx}{\sqrt{\pi}}$，$f(x)\sim\frac{a_0}{2}+\sum_{k=1}^{+\infty}(a_k\cos kx+b_k\sin kx)$

其中 $a_k=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos kx\,dx$，$b_k=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin kx\,dx$

Thm 5.23 设$u_1,\dots u_n$线性无关，$\{u_i^*\}$是$\{u_i\}$经斯密特正交化得到的正交组

若$w=\sum_{i=1}^n a_i u_i$，则$w=\sum_{i=1}^n \langle w, u_i^*\rangle u_i^*$

PF: 由$w=\sum_{i=1}^n a_i u_i$及Cor 5.16得$\exists c_1,\dots,c_n$ s.t. $w=\sum_{i=1}^n c_i u_i^*$，由$\{u_i^*\}$线性无关，$c_i=\langle w, u_i^*\rangle$

Thm 5.24 (Frourier级数的最小性) 设$u_1^*,\dots u_n^*$是正交系统，则$\forall w$，$\|w-\sum_{i=1}^n\langle w,u_i^*\rangle u_i^*\|\le\|w-\sum_{i=1}^n a_i u_i^*\|$ $\forall a_i$

PF: $\|w-\sum_{i=1}^n a_i u_i^*\|^2=\langle w-\sum_i a_i u_i^*, w-\sum_i a_i u_i^*\rangle=\langle w,w\rangle-\langle w,\sum a_i u_i^*\rangle-\langle\sum a_i u_i^*, w\rangle+\sum_i\sum_j a_i\bar{a}_j\langle u_i^*,u_j^*\rangle$

$=\langle w,w\rangle-\sum_i\bar{a}_i\langle w,u_i^*\rangle-\sum_i a_i\langle u_i^*,w\rangle+\sum_i|a_i|^2=\|w\|^2-\sum_i|\langle w,u_i^*\rangle|^2+\sum_i|a_i-\langle w,u_i^*\rangle|^2$.

Cor 5.25 设$(u_1,u_2,\dots u_n)$是线性无关组，则w的最佳线性估计是$\hat{\varphi}=\sum_k\langle w,u_k^*\rangle u_k^*$，

其中$\{u_i^*\}$是$\{u_i\}$经过斯密特正交化得到的正交化得到正交组，误差为$\|w-\hat{\varphi}\|^2=\min_{a_k}\|w-\sum_{k=1}^n a_k u_k\|^2$

$=\|w\|^2-\sum_k|\langle w,u_k^*\rangle|^2$

Cor 5.26 (贝塞尔不等式) 若$u_1^*,u_2^*,\dots,u_n^*$正交，则$\forall w$，$\sum_{i=1}^n|\langle w,u_i^*\rangle|^2\le\|w\|^2$

Cor 5.27 Def 5.13中斯密特正交化的过程满足$\forall n\in\mathbb{N}^+$，$\|v_{n+1}\|^2=\|u_{n+1}\|^2-\sum_{i=1}^n|\langle u_{n+1},u_i^*\rangle|^2$

Ex 5.28 考虑5.4在$a=-1$，$b=1$，$\rho=1$的内积定义下求均方误差最小的线性估计，即$\min_{a_i}\int_{-1}^1(e^x-\sum_{i=0}^n a_i x^i)^2dx$

对$n=1,2$，由Ex 5.19中求得的勒让德多项式 $u_1^*=\frac{1}{\sqrt{2}}$，$u_2^*=\sqrt{\frac{3}{2}}x$，$u_3^*=\sqrt{\frac{5}{8}}(3x^2-1)$

e^x的傅里叶系数为 $b_1=\int_{-1}^1\frac{1}{\sqrt{2}}e^x dx=\frac{1}{\sqrt{2}}(e-\frac{1}{e})$，$b_2=\int_{-1}^1\sqrt{\frac{3}{2}}xe^x dx=\sqrt{6}e^{-1}$，$b_3=\int_{-1}^1\sqrt{\frac{5}{8}}(3x^2-1)e^x dx=\sqrt{\frac{5}{2}}(e-\frac{7}{e})$

$$\hat{\varphi}_n=\begin{cases}\frac{1}{2e}(e^2-1)+\frac{3}{e}x, & n=1\\ \frac{1}{2e}(e^2-1)+\frac{3}{e}x+\frac{5}{4e}(e^2-7)(3x^2-1), & n=2\end{cases}$$

§5.3 正则方程组

Thm 5.29 设$u_1,u_2,\dots,u_n\in X$线性无关，u_i^*是u_i经斯密特正交化后的正交基，则

$\forall w$，$\forall j=1,\dots,n$ $(w-\sum_{i=1}^n\langle w,u_i^*\rangle u_i^*)\perp u_j^*$

PF: 若$w\in X$，则$w-\sum_{i=1}^n\langle w,u_i^*\rangle u_i^*=0$.

若$w\notin X$，令$w=u_{n+1}$并对u_{n+1}作斯密特正交化，则$u_{n+1}^*=\frac{w-\sum_{i=1}^n\langle w,u_i^*\rangle u_i^*}{\|v_{n+1}\|}\perp u_i$，$i=1,\dots,n$ 由Cor 5.17

Cor 5.30 设$u_1,\dots,u_n\in X$线性无关，若$\hat{\varphi}=\sum_{k=1}^n a_k u_k$是$w$的最佳线性估计，则$\forall j=1,\dots,n$，$(w-\hat{\varphi})\perp u_j$

PF: $\hat{\varphi}=\sum_{k=1}^n a_k u_k=\sum_{k=1}^n\langle w,u_k^*\rangle u_k^*$，由Thm 5.29 $(w-\hat{\varphi})\perp u_j^*$，$j=1,\dots,n$，由Cor 5.16 $(w-\hat{\varphi})\perp u_j$，$j=1,\dots,n$

Def 5.31 设$u_1,\dots,u_n$是内积空间的一列元素，则记$n\times n$矩阵$G=G(u_1,\dots,u_n)=(\langle u_i,u_j\rangle)_{n\times n}$为$u_1,\dots,u_n$的Gram矩阵，记$g=g(u_1,\dots,u_n)=\det G$为Gram行列式.

Lem 5.32 设$w_i=\sum_{j=1}^n a_{ij}u_j$，$i=1,2,\dots,n$，$A=(a_{ij})_{n\times n}$，$A^H=(\overline{a_{ji}})_{n\times n}$ 则 $\begin{cases}G(w_1,\dots,w_n)=AG(u_1,\dots,u_n)A^H\\ g(w_1,\dots,w_n)=|\det A|^2 g(u_1,\dots,u_n)\end{cases}$

PF: $G(w_1,\dots,w_n)=\begin{bmatrix}a_{11}&\cdots&a_{1n}\\ \vdots&&\\ a_{n1}&\cdots&a_{nn}\end{bmatrix}\begin{bmatrix}\langle u_1,w_1\rangle&\cdots&\langle u_1,w_n\rangle\\ \vdots&&\vdots\\ \langle u_n,w_1\rangle&\cdots&\langle u_n,w_n\rangle\end{bmatrix}=\begin{bmatrix}a_{11}&\cdots&a_{1n}\\ \vdots&&\vdots\\ a_{n1}&\cdots&a_{nn}\end{bmatrix}\begin{bmatrix}\langle u_1,u_1\rangle&\cdots&\langle u_1,u_n\rangle\\ \vdots&&\vdots\\ \langle u_n,u_1\rangle&\cdots&\langle u_n,u_n\rangle\end{bmatrix}\begin{bmatrix}\overline{a_{11}}&\cdots&\overline{a_{n1}}\\ \vdots&&\\ \overline{a_{1n}}&\cdots&\overline{a_{nn}}\end{bmatrix}=AG(u_1,\dots,u_n)A^H$

$\det[G(w_1,\dots,w_n)]=\det A\cdot\det G(u_1,\dots,u_n)\cdot\det A^H=\det A\,\overline{\det A}\,g(u_1,\dots,u_n)=|\det A|^2 g(u_1,\dots,u_n)$

Thm 5.33 对非0元素 $u_1, u_2, \cdots u_n \in X$，$0 \le g(u_1, \cdots, u_n) \le \prod_{k=1}^{n} \|u_k\|^2$

左边等号成立当且仅当 $u_1, \cdots, u_n$ 线性相关，右边等号成立当且仅当 $u_1, \cdots, u_n$ 正交。

PF：若 $u_1, \cdots, u_n$ 线性相关，则 $\exists c_1, \cdots, c_n$，$\sum_{i=1}^{n} c_i u_i = 0$，$[c_1, \cdots, c_n]^T \ne 0$，不妨设 $c_j \ne 0$。

令 $w_k = \begin{cases} \sum_{i=1}^{n} c_i u_i = 0, & k = j \\ u_k, & k \ne j \end{cases}$，则 $g(w_1, \cdots, w_n) = 0$，而 $w = Cu$，$C = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ c_1 & c_2 & \cdots & c_n & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}$，$\det C = c_j \ne 0$

因此 $g(u_1, \cdots, u_n) = \dfrac{g(w_1, \cdots, w_n)}{|\det C|^2} = 0$。

若 $u_1, \cdots, u_n$ 线性无关，由 Thm 5.14 可知 $\exists$ 正交组 $\{v_k^*\}$，s.t. $v_k^* = \sum_{i=1}^{k} a_{ki} u_i$

由定义 $g(v_1^*, \cdots, v_n^*) = 1$ 又 $\because \det(A) = \prod_{k=1}^{n} a_{kk}$ $\therefore g(u_1, \cdots, u_n) = \prod_{k=1}^{n} \dfrac{1}{a_{kk}^2} > 0$ $\therefore$ 左边不等号即取等条件得证。

若 $u_1, \cdots u_n$ 正交，$g(u_1^*, \cdots, u_n^*) = \det \operatorname{diag}(\|u_i\|^2)_n = \prod_{k=1}^{n} \|u_k\|^2$

反之，若 $g(u_1, \cdots, u_n) = \prod_{k=1}^{n} \|u_k\|^2$，则 $u_1, \cdots, u_n$ 一定线性无关。

对 $(u_1, \cdots, u_n)$ 应用 schmidt 正交化，由 Thm 5.14 得 $\cancel{\|u_k\|^2 =}$ $\dfrac{1}{a_{kk}} = \|v_k\|$，分别对前 $1, 2, \cdots, n$ 个 u_k 应用 Thm 5.14

得 $\|u_k\|^2 = \|v_k\|^2$，由 Cor 5.27，$\forall k = 1, 2, \cdots, n$ $\sum_{j=1}^{k-1} |\langle u_k, u_j^* \rangle|^2 = 0$ $\therefore$ u_k 正交

且由 Cor 5.27 得 $g(u_1, \cdots, u_n) = \prod_{k=1}^{n} \dfrac{1}{a_{kk}^2} = \prod_{k=1}^{n} \|v_k\|^2 = \prod_{k=1}^{n} \left(\|u_k\|^2 - \sum_{j=1}^{k-1} |\langle u_k, u_j^* \rangle|^2 \right) < \prod_{k=1}^{n} \|u_k\|^2$，右边得证。

Thm 5.34 设 $\hat{\varphi} = \sum_{i=1}^{n} a_i u_i$ 是 w 在 $\operatorname{span}\{u_1, \cdots, u_n\}$ 中的最佳线性估计，则系数 $a = (a_1, a_2, \cdots, a_n)^T$ 由正则方程组

$G(u_1, \cdots, u_n)^T a = c$，$c = [\langle w, u_1 \rangle, \cdots, \langle w, u_n \rangle]^T$ 唯一确定。

PF：由 Cor 5.30 $\langle \hat{\varphi}, u_j \rangle = \langle w, u_j \rangle = \sum_{k=1}^{n} a_k \langle u_k, u_j \rangle$，$\forall j = 1, \cdots, n$，联立 n 个方程。

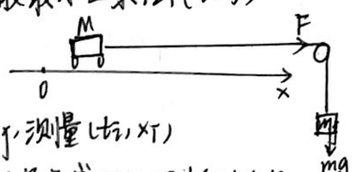
Ex 5.35 求 e^x 在 $\operatorname{sp}\{1, x, x^2\}$ 中的最佳线性估计。

$$G(1, x, x^2) = \begin{bmatrix} \langle 1, 1 \rangle, & \langle 1, x \rangle, & \langle 1, x^2 \rangle \\ \langle x, 1 \rangle, & \langle x, x \rangle, & \langle x, x^2 \rangle \\ \langle x^2, 1 \rangle, & \langle x^2, x \rangle, & \langle x^2, x^2 \rangle \end{bmatrix} = \begin{bmatrix} 2 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{3}{2} & 0 & \frac{2}{5} \end{bmatrix} \quad c = \begin{bmatrix} \langle e^x, 1 \rangle \\ \langle e^x, x \rangle \\ \langle e^x, x^2 \rangle \end{bmatrix} = \begin{bmatrix} e - \frac{1}{e} \\ \frac{2}{e} \\ e - \frac{5}{e} \end{bmatrix} \quad \text{解得} \begin{cases} a_0 = \dfrac{3(11 - e^2)}{4e} \\ a_1 = \dfrac{3}{e} \\ a_2 = \dfrac{15(e^2 - 7)}{4e} \end{cases}$$

Ex 5.36 对于 $f \in C[0,1]$ 在 $\operatorname{sp}\{1, x, x^2, \cdots\}$ 求最小二乘估计，则 $\langle x^j, x^k \rangle = \int_0^1 x^j x^k dx = \dfrac{1}{j+k+1}$ 为希尔伯特矩阵

§5.4 离散最小二乘法（DLS）

Ex 5.36



$m_j g = (m_j + M) a = (m_j + M) \dfrac{d^2 x}{dt^2}$

(1) 定 M, m_j，测量 (t_i, x_i)

(2) 拟合二次多项式 $p(t) = c_0 + c_1 t + c_2 t^2$

使 $\sum_i (x_i - p(t_i))^2$ 最小

(3) 取 $a_j = 2c_2$ 作为加速度的测量值，$F_j = m_j(g - a_j)$

(4) 改变 m_j，重复 (1)~(3)，得到若干 (a_j, F_j)

(5) 拟合出一次多项式 $f(x) = c_0 + c_1 x$，使 $\sum_j (F_j - f(x_j))$ 最小，

最终得到 M 近似值 c_1

* (2) 与 (5) 为"最小二乘问题"

§5.4.1 高斯函数和 Dirac delta 函数.

Def 5.38 Gauss函数形如 $f(x)=a\exp\left\{-\frac{(x-b)^2}{2c^2}\right\}$

Lem 5.39 Gauss函数的积分：$\int_{-\infty}^{+\infty} a\exp\left\{-\frac{(x-b)^2}{2c^2}\right\}dx = ac\sqrt{2\pi}$

Def 5.40 正态分布(高斯分布)：$f_{\mu,\sigma}=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$

Def 5.41 中心在 $\bar{x}$ 的 Dirac delta 函数：$\delta(x-\bar{x})=\lim\limits_{\varepsilon\to 0}\phi_\varepsilon(x-\bar{x})$，其中 $\phi_\varepsilon(x-\bar{x})=f_{\bar{x},\varepsilon}$

Lem 5.42 Dirac delta 函数满足 $\delta(x-\bar{x})=\begin{cases}+\infty, & x=\bar{x}\\ 0, & x\neq\bar{x}\end{cases}$ $\int_{-\infty}^{+\infty}\delta(x-\bar{x})dx=1$

Lem 5.43 (δ的筛选性) 设 $f:\mathbb{R}\to\mathbb{R}$ 连续，则 $\int_{-\infty}^{+\infty}\delta(x-\bar{x})f(x)dx=f(\bar{x})$

PF: $\forall\varepsilon>0$，$\int_{-\infty}^{+\infty}\delta(x-\bar{x})f(x)dx=\int_{\bar{x}-\varepsilon}^{\bar{x}+\varepsilon}\delta(x-\bar{x})f(x)dx=f(\xi_\varepsilon)\int_{\bar{x}-\varepsilon}^{\bar{x}+\varepsilon}\delta(x-\bar{x})dx=f(\xi_\varepsilon)$ $\xi_\varepsilon\in[\bar{x}-\varepsilon,\bar{x}+\varepsilon]$

令 $\varepsilon\to 0$ 则 $\int_{-\infty}^{+\infty}\delta(x-\bar{x})f(x)dx=f(\bar{x})$

Def 5.44 Heaviside 函数/阶梯函数 $H(x)=\begin{cases}0, & x<0\\ 1, & x>0\end{cases}$

~~Def~~ Lem 5.45 $H(x)=\int_{-\infty}^{x}\delta(t)dt$

§5.4.2 离散测度

Def 5.46 定义函数 $\lambda:\mathbb{R}\to\mathbb{R}$，$\lambda(t)=\begin{cases}0, & t\in(-\infty,a)\\ \int_a^t\rho(\tau)d\tau, & t\in[a,b]\\ \int_a^b\rho(\tau)d\tau, & t\in(b,+\infty)\end{cases}$；$d\lambda(t)=\begin{cases}\rho(t)d\lambda, & t\in[a,b]\\ 0, & \text{otherwise}\end{cases}$

Def 5.47 称点集 $\{t_1,t_2,\dots,t_N\}$ 的离散测度或 Dirac 测度为 $d\lambda(t)=\sum_{i=1}^{N}\rho_i\delta(t-t_i)dt$

相应地，$\lambda(t)=\sum_{i=1}^{N}\rho_iH(t-t_i)$

Lem 5.48 设 $u:\mathbb{R}\to\mathbb{R}$，定义 $\lambda(t)=\sum_{i=1}^{N}\rho_iH(t-t_i)$，则 $\int_{\mathbb{R}}u(t)d\lambda=\sum_{i=1}^{N}\rho_iu(t_i)$

PF: $\int_{\mathbb{R}}u(t)d\lambda=\int_{\mathbb{R}}\sum_{i=1}^{N}\rho_i\delta(t-t_i)u(t)dt=\sum_{i=1}^{N}\rho_iu(t_i)$

§5.4.3 DLS 和正则方程组.

Ex 5.49 考虑如下数据.

x	1	2	3	4	5	6	7	8	9	10	11	12
y	256	201	159	61	77	40	17	25	103	156	222	345

拟合二次多项式 $\sum_{j=0}^{2}a_jx^j$ 使 $\sum_{i=1}^{12}(y_i-\sum_{j=0}^{2}a_jx_i^j)^2$ 最小

Sol: 定义内积 $\langle f,g\rangle=\sum_{i=1}^{12}f(x_i)g(x_i)$，$G(1,x,x^2)=(\langle x^{i-1},x^{j-1}\rangle)_{3\times3}$，$c=\begin{bmatrix}\langle y,1\rangle\\ \langle y,x\rangle\\ \langle y,x^2\rangle\end{bmatrix}$，$a=G^{-1}c$

§5.4.3 DLS 的 QR 分解法

Def 5.50 称 $A\in\mathbb{R}^{n\times n}$ 正交 $\Leftrightarrow A^TA=I$

Def 5.51 称 $A\in\mathbb{R}^{n\times n}$ 上三角 $\Leftrightarrow\forall i>j$，$a_{ij}=0$，下三角 $\Leftrightarrow\forall i<j$，$a_{ij}=0$

Thm 5.52 (QR分解) $\forall A\in\mathbb{R}^{m\times n}$，$\exists$ 正交阵 Q 和上三角阵 R s.t. $A=QR$

PF: 记 $A=[\xi_1,\dots,\xi_n]\in\mathbb{R}^{m\times n}$，按如下步骤构造秩为 $r(A)$ 的矩阵 $A_r=[u_1,\dots,u_r]$

(S1) 令 $u_1=\xi_{k_1}$，k_1 满足 $\xi_{k_1}\neq 0$，$\forall l<k_1$ $\xi_l=0$.

(S2) $\forall j=2,\dots,r$ 令 $u_j=\xi_{k_j}$，k_j 满足 $K_j=\{\xi_{k_1},\dots,\xi_{k_j}\}$ 线性无关；$\forall l=k_{j-1}+1,\dots,k_j-1$，$\xi_l$ 可由 K_{j-1} 线性表示

$\Rightarrow$ 选出线性无关的 r 列.

由 Cor 5.16，斯密特正交化过程给出了唯一的正交阵 $A_r^*=[u_1^*,\dots,u_r^*]\in\mathbb{R}^{r\times r}$ 和上三角阵 $(b_{ij})_{r\times r}$

s.t. $A_r=A_r^*\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1r}\\ & b_{22} & \cdots & b_{2r}\\ & & \ddots & \vdots\\ & & & b_{rr}\end{bmatrix}$，$r\le m$. 接下来将其余列插入，对于全0列 $\xi_1,\dots,\xi_{k_1-1}$ $[\xi_1,\dots,\xi_{k_1-1},u_1,\dots,u_r]=A_r^*\left[\begin{array}{c|ccc} 0 & b_{11} & \cdots & b_{1r}\\ & & \ddots & \vdots\\ & & & b_{rr}\end{array}\right]$

对于 $l=k_{j-1}+1,\dots,k_j-1$ 列 $[u_1,\dots,u_{j-1},\xi_l]=[u_1^*,\dots,u_{j-1}^*]\left[\begin{array}{cccc} b_{11} & \cdots & b_{1,j-1} & c_{l1}\\ & \ddots & \vdots & \vdots\\ & & b_{j-1,j-1} & c_{l,j-1}\\ \hline & & 0 & \end{array}\right]$，由此可构造 $A=A_r^*R$.

Lem 5.53 正交阵是保2-测度的，即 $\forall x\in\mathbb{R}^n$，$\forall$ 正交阵 $Q\in\mathbb{R}^{n\times n}$，$\|Qx\|_2=\|x\|_2$

PF: $\|Qx\|_2^2=x^TQ^TQx=x^Tx=\|x\|_2^2$

Thm 5.54 考虑超定线性方程组 $Ax=b$，$A\in\mathbb{R}^{m\times n}$，$m\ge n$，则离散最小二乘问题 $\min\limits_{x\in\mathbb{R}^n}\|Ax-b\|_2^2$ 的解 x^* 为 $R_1x^*=c$ 的解，其中 $R_1\in\mathbb{R}^{n\times n}$，$c\in\mathbb{R}^n$，满足 $Q^TA=R=\begin{bmatrix}R_1\\0\end{bmatrix}$，$Q^Tb=\begin{bmatrix}c\\r\end{bmatrix}$

PF: $\forall x\in\mathbb{R}^n$，$\|Ax-b\|_2^2=\|Q^TAx-Q^Tb\|_2^2=\|R_1x-c\|_2^2+\|r\|_2^2$

§5.5 解决 ill-posed 线性系统

Def 5.55 称一个问题是 well-posed 的：(i) 有一个解 (ii) 解唯一 (iii) 条件数小. 否则，该问题为 ill-posed.

Lem 5.56 (线性系统的可解性) $A\in\mathbb{C}^{m\times n}$ 的线性系统 $Ax=b$ 有解，当且仅当 b 在 A 的范围内，或等价地说，b 与 A 的伴随矩阵的零空间垂直，即 $\forall z\in\{y\in\mathbb{C}^m: A^*y=0\}$，$\langle b,z\rangle=0$

在这种情况下，一个特解为 $x^{\oplus}=\sum\limits_{j=1}^{r}\frac{1}{\sigma_j}\langle b,v_j\rangle u_j$，其中 r 为 A 的秩，$\langle\sigma_j,u_j,v_j\rangle$ 为其奇异值和奇异向量.

PF: 若 $\exists x\in\mathbb{C}^n$，$Ax=b$，则 $\langle b,z\rangle=\langle Ax,z\rangle=\langle x,A^*z\rangle=0$

若 $\forall z\in\{y\in\mathbb{C}^m; A^*y=0\}$，$\langle b,z\rangle=0$，则 $b=\sum\limits_{j=1}^{r}\langle b,v_j\rangle v_j=\sum\limits_{j=1}^{r}\frac{1}{\sigma_j}\langle b,v_j\rangle Au_j=Ax^{\oplus}$

§5.5.1 无解或多解

Thm 5.57 假设 Lem 5.56 中 $Ax=b$ 无解，则 $x^{\oplus}$ 为其最小二乘问题的解，$x^{\oplus}=\arg\min\limits_{x\in\mathbb{C}^n}\|Ax-b\|_2^2$

PF: b 的傅里叶展开为 $b=\sum\limits_{j=1}^{m}\langle b,v_j\rangle v_j$，$\because b\notin\operatorname{range}A$，$Ax^{\oplus}-b\in\operatorname{span}(v_{r+1},\dots,v_m)$

由 A 的秩为 r，$\forall x\in\mathbb{C}^n$，$Ax-Ax^{\oplus}\in\operatorname{span}(v_1,\dots,v_r)$，因此 $\langle Ax^{\oplus}-b, Ax-Ax^{\oplus}\rangle=0$

由毕达哥拉斯定理，$\|Ax-b\|_2^2=\|Ax-Ax^{\oplus}\|_2^2+\|Ax^{\oplus}-b\|_2^2$

Thm 5.58 若 $Ax=b$ 有多解，则通解为 $y+x^{\oplus}$，$y\in\operatorname{null}A$，且 $x^{\oplus}$ 为具有最小2范数的解.

PF: 由 Lem 5.56 $x^{\oplus}\in\operatorname{span}(u_1,\dots,u_r)$，$\operatorname{null}A=\operatorname{span}(u_{r+1},\dots,u_n)$，因此 $\forall y\in\operatorname{null}A$，$\langle x^{\oplus},y\rangle=0$

由毕达哥拉斯定理 $\|x^{\oplus}+y\|_2=\|x^{\oplus}\|_2+\|y\|_2$

§5.5.2 The Moore-Penrose inverse

Def 5.59 矩阵 $A\in\mathbb{F}^{m\times n}$ ($\mathbb{F}=\mathbb{R},\mathbb{C}$) 的 $\begin{cases}\text{pseudo}\\ \text{Moore-Penrose}\\ \text{generalized}\end{cases}$ 逆为 $A^+\in\mathbb{F}^{n\times m}$. $A^+y=\sum_{j=1}^{r}\frac{1}{\sigma_j}\langle y, v_j\rangle u_j$

Thm 5.60 A^+ 满足以下性质：

(PDI-1) $AA^+A=A$ (PDI-2) $A^+AA^+=A^+$ (PDI-3) $(AA^+)^*=AA^+$, $(A^+A)^*=A^+A$, 即为 Hermitian 阵

Lem 5.61 若 A 的列向量线向无关，则 A^*A 可逆且 $A^+=(A^*A)^{-1}A^*$, $A^+A=I$

若 A 的行向量线向无关，则 AA^* 可逆，且 $A^+=A^*(AA^*)^{-1}$, $AA^+=I$.

§5.5.3 Spectral cutoff for ill-conditioning

Ex 5.62 求解 $Ax=b$ 的条件数受小奇异值支配，对 $\varepsilon\in\mathbb{C}$, 将 b 扰动为 $\hat{b}=b+\varepsilon v_j$, 由 Lem 5.56, 解将变为 $\hat{x}=x+\frac{\varepsilon}{\sigma_j}u_j$, 基于绝对误差的条件数为 $\frac{\|\hat{x}-x\|_2}{\|\hat{b}-b\|_2}=\frac{1}{\sigma_j}$, 因此若其奇异值很小，则问题为病态.

Def 5.63 谱截止 (spectral cutoff) 是通过忽略解 x 中奇异值较小的项来使病态问题更稳定的方法.

Def 5.64 给定 $Ax=b$ 的扰动 $\hat{b}$, 差异原理 (discrepancy principle) 是一种得出近似解的策略 使得 $\|A\hat{x}-\hat{b}\|_2=C\|b-\hat{b}\|_2$, 其中 $C\geq 1$ 为一个合适的常数.

Ex 5.65 对于谱截止，$x_p=\begin{cases}0, & p=0\\ \sum_{j=1}^{p}\sigma_j\langle\hat{b},v_j\rangle u_j & p=1,\dots,r\end{cases}$, 其中 p 由 $\|\hat{b}-b\|_2$ 决定，常令 $C=1$.

Thm 5.66 $A\in\mathbb{C}^{m\times n}$, 线性方程组 $Ax=b$ 可解，$\|\hat{b}-b\|_2\leq\varepsilon\leq\|\hat{b}\|_2$, 则存在一个最小的 $p=p(\varepsilon)$ 使得 $\|Ax_p-\hat{b}\|_2\leq\varepsilon$

谱截止的差异原理是收敛的，即 $\lim_{\varepsilon\to 0}x_p=A^+b$.

Pf: 定义 $f:\{0,1,\dots,r\}\to\mathbb{R}$, $f(p)=\|Ax_p-\hat{b}\|_2-\varepsilon^2=\sum_{j=p+1}^{m}|\langle\hat{b},v_j\rangle|^2-\varepsilon^2$, v_j 为 A 的右奇异向量.

$f(0)=\|\hat{b}\|-\varepsilon^2\geq 0$. $f(r)=-\varepsilon^2<0$ ($r=m$)

若 $r<m$, 由 Lem 5.56, $\forall j=r+1,\dots,m$, $\langle b,v_j\rangle=0$. $f(r)=\sum_{j=r+1}^{m}|\langle\hat{b}-b,v_j\rangle|^2-\varepsilon^2\leq\|\hat{b}-b\|_2^2-\varepsilon^2\leq 0$

f 单调递减，$f(0)\geq 0$, $f(r)\leq 0$, 则 $\exists p$, $\|Ax_p-\hat{b}\|_2\leq\varepsilon$

$\|Ax_p-b\|_2\leq\|Ax_p-\hat{b}\|_2+\|\hat{b}-b\|_2\leq 2\varepsilon$, $\because\forall v\in\text{span}(v_1,\dots,v_r)$ $A^+Av=v$ $\therefore\lim_{\varepsilon\to 0}x_p=A^+b$.

§5.5.4 Tikhonov 正则化

Def 5.67 (Tikhonov 正则化) 通过正则化线性方程组 $\alpha x+A^*Ax=A^*b$ 估计病态最小二乘问题 $Ax=b$ 的解. $\alpha>0$ 称为正则化系数.

Thm 5.68 设矩阵 $A\in\mathbb{C}^{m\times n}$ 的秩为 r, 则对任意 $b\in\mathbb{C}^m$, Tikhonov 正则化有唯一解.

$x_\alpha=\sum_{j=1}^{r}\frac{\sigma_j}{\alpha+\sigma_j^2}\langle b,v_j\rangle u_j$. 进一步的 Tikhonov 正则化满足 $\lim_{\alpha\to 0}(\alpha I+A^*A)^{-1}A^*b=A^+b$.

Pf: $B=\alpha I+A^*A$ 严格正定，故非奇异，且 $Bu_j=(\alpha I+A^*A)u_j=\alpha u_j+A^*Au_j=\alpha u_j+\sigma_jA^*v_j=\alpha u_j+\sigma_j^2u_j=(\alpha+\sigma_j^2)u_j$

$B^*=B\Rightarrow(\alpha+\sigma_j^2,u_j,u_j)$ 为 B 的奇异值，左奇异向量，右奇异向量.

将 Lem 5.56 应用于 $\alpha x+A^*Ax=A^*b$ 得 $x=\sum_{j=1}^{r}\frac{1}{\alpha+\sigma_j^2}\langle A^*b,u_j\rangle u_j=\sum_{j=1}^{r}\frac{\sigma_j}{\alpha+\sigma_j^2}\langle b,v_j\rangle u_j=x_\alpha$

Thm 5.69 $\forall b\in\mathbb{C}^m$, Tikhonov正则化的解 x_α 等价于 $\|Ax-b\|_2^2+\alpha\|x\|_2^2$ 的最小值点.

Pf: 令 $z_\alpha=\alpha x_\alpha+A^*Ax_\alpha-A^*b$, 由余弦定理,

$\|Ax-b\|_2^2+\alpha\|x\|_2^2=2\mathrm{Re}\langle x-x_\alpha,z_\alpha\rangle+\|Ax_\alpha-b\|_2^2+\alpha\|x_\alpha\|_2^2+\|Ax-Ax_\alpha\|_2^2+\alpha\|x-x_\alpha\|_2^2$

注: $2\mathrm{Re}\langle Ax-Ax_\alpha,Ax_\alpha-b\rangle+2\alpha\mathrm{Re}\langle x-x_\alpha,x_\alpha\rangle=2\mathrm{Re}\langle x-x_\alpha,A^*(Ax_\alpha-b)+\alpha x_\alpha\rangle=2\mathrm{Re}\langle x-x_\alpha,z_\alpha\rangle$

因此 x_α 满足 $z_\alpha=0$, 另一方面, 设 x_α 满足 $z_\alpha\neq 0$ 且是 $\|Ax-b\|_2^2+\alpha\|x\|_2^2$ 的最小值点.

则对 $x=x_\alpha-\varepsilon z_\alpha$, $\varepsilon\in\mathbb{R}$ $\|Ax-b\|_2^2+\alpha\|x\|_2^2=\|Ax_\alpha-b\|_2^2+\alpha\|x_\alpha\|_2^2-2\varepsilon a+\varepsilon^2 b$, $a=\|z_\alpha\|_2^2$, $b=\|Az_\alpha\|_2^2+\alpha\|z_\alpha\|_2^2$

取 $\varepsilon=\frac{a}{b}$, 则 $\|Ax-b\|_2^2+\alpha\|x\|_2^2<\|Ax_\alpha-b\|_2^2+\alpha\|x_\alpha\|_2$ 与 x_α 的最小性矛盾.

Ex 5.70 在实际应用中, 右端项通常不精确(在某个范围内), 例如 $\|\hat{b}-b\|_2<\varepsilon$

此时 Tikhonov 正则化对真解 $x=A^+b$ 的估计 x_α 满足 $Bx_\alpha=A^*\hat{b}$ ($B=\alpha I+A^*A$)

故绝对误差为 $x_\alpha-x=B^{-1}A^*(\hat{b}-b)+(B^{-1}A^*b-A^+b)$ (第1项来自 $\hat{b}-b$, 第2项来自正则化)

另一方面, 由 Lem 4.68 得 $\mathrm{cond}_2 B=\frac{\alpha+\sigma_1^2}{\alpha+\sigma_n^2}$, 为使 $\mathrm{cond}_2 B$ 小, α 应尽量大, 注意这个结论对 Ex 5.62 中

绝对误差的定义同样成立. α 应适当选择, 以保证准确性(α 不能太大)和稳定性(不能太小)

Thm 5.71 设线性方程组 $Ax=b$, $A\in\mathbb{C}^{m\times n}$ 可解, b 的测量值 $\hat{b}$ 满足 $\|\hat{b}-b\|_2\leq\varepsilon\leq\|\hat{b}\|_2$

则存在唯一 $\alpha=\alpha(\varepsilon)$ s.t. $\|Ax_\alpha-\hat{b}\|_2=\varepsilon$, 其中 x_α 是 $(\alpha I+A^*A)x=A^*\hat{b}$ 的唯一解.

进一步地, Tikhonov 正则化的差异原则是收敛的, 即 $\lim\limits_{\varepsilon\to 0}x_\alpha=A^+b$

Pf: 由 Thm 5.68 $f(\alpha)=\|Ax_\alpha-\hat{b}\|_2^2-\varepsilon^2=\sum\limits_{j=1}^{m}\left|\frac{\alpha^2}{(\alpha+\sigma_j^2)^2}\langle\hat{b},u_j\rangle\right|^2-\varepsilon^2$

其中 u_j 是 A 的左奇异向量, f 是连续单增函数且 $\lim\limits_{\alpha\to 0}f(\alpha)=-\varepsilon^2<0$. $\lim\limits_{\alpha\to+\infty}f(\alpha)=\|\hat{b}\|_2^2-\varepsilon^2>0$

故 f 有唯一根. $\exists!\,\alpha=\alpha(\varepsilon)$ s.t. $\|Ax_\alpha-\hat{b}\|_2=\varepsilon$

$\|\hat{b}\|_2-\varepsilon=\|\hat{b}\|_2-\|Ax_\alpha-\hat{b}\|_2\leq\|Ax_\alpha\|_2$

另一方面 $\because(\alpha I+A^*A)x_\alpha=A^*\hat{b}$ $\therefore\alpha\|Ax_\alpha\|_2=\|AA^*(\hat{b}-Ax_\alpha)\|_2\leq\|AA^*\|_2\varepsilon$

由 $\|\hat{b}\|_2\geq\|b\|-\varepsilon$, 结合上2个不等式.

$$\alpha\leq\frac{\|AA^*\|_2\varepsilon}{\|Ax_\alpha\|_2}\leq\frac{\|AA^*\|_2\varepsilon}{\|b\|_2-2\varepsilon}\quad 故\ \lim_{\varepsilon\to 0}\alpha=0\quad\lim_{\varepsilon\to 0}x_\alpha=\lim_{\alpha\to 0}x_\alpha=\lim_{\alpha\to 0}\sum_{j=1}^{r}\frac{\sigma_j}{\alpha+\sigma_j^2}\langle b,u_j\rangle v_j=\sum_{j=1}^{r}\frac{1}{\sigma_j}\langle b,u_j\rangle v_j=A^+b$$