第十五讲:逼近与回归

凸优化问题的重要应用

杨 林

大 纲

1. 逼近

2. LASSO

大 纲

1. 逼近

2. LASSO

1 逼近

■ 逼近

在凸优化中，逼近是指一系列解决难以处理的优化问题（如：问题非凸，目标函数或约束函数过于复杂）的策略. 其核心思想是：用一个易于求解的凸优化问题，去近似一个原本难以求解的问题（或者是通过一个容易获得的次优解，代替原问题的最优解）

通常采取改变目标函数或约束集，实现模型简化的效果. 包括四种逼近策略：

- 范数逼近
- 罚函数逼近
- 正则化逼近
- 鲁棒逼近

1.1 范数逼近

$$\text{最小化} \qquad \|Ax - b\|$$

$(A \in \mathbb{R}^{m\times n}, m \geq n)$ Ax可以表示为$Ax = x_1 a_1 + \cdots + x_n a_n$，用$A$的列向量尽可能的逼近或者拟合向量$b$

对最优解 $x = \underset{x}{\operatorname{argmin}} \|Ax - b\|$ 的解释：

1. **几何**：Ax^* 是 $\mathcal{R}(A)$ 中离 b 最近的点
2. **估计**：线性测量模型

$$y = Ax + v$$

是测量值，x 是未知量，v 是测量误差；给定 $y = b$， x 的最佳猜测是 x^*

3. **最优设计**：x 是设计变量（输入），x^*是能最佳逼近期望结果 b 的设计

1.1 范数逼近

最小二乘逼近 ($\|\cdot\|$): 解满足正规方程

$$A^T A x = A^T b$$

$(x^* = (A^T A)^{-1} A^T b$,若 $A = n)$

切比雪夫逼近($\| \quad \|_\infty$): 可以作为线性规划问题求解

$$\begin{array}{ll} \text{最小化} & t \\ \text{约束条件} & -t\mathbf{1} \preceq Ax - b \preceq t\mathbf{1} \end{array}$$

绝对残差和的逼近 ($\| \quad \|_1$): 可以作为线性规划问题求解

$$\begin{array}{ll} \text{最小化} & \mathbf{1}^T y \\ \text{约束条件} & -y \preceq Ax - b \preceq y \end{array}$$

1.2 罚函数逼近

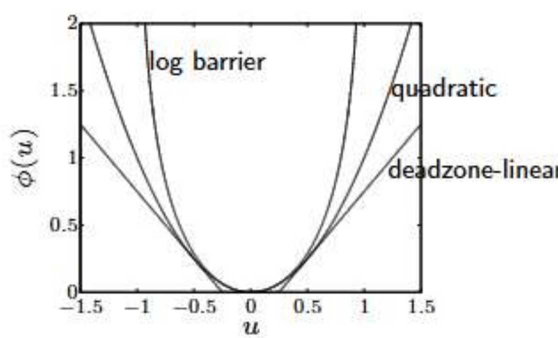
$$\begin{array}{ll} \text{最小化} & \phi(r_1) + \cdots + \phi(r_m) \\ \text{约束条件} & r = Ax - b \end{array}$$

$A \in \mathbb{R}^{m \times n}$, $m \geq n$, $\phi: \mathbb{R} \to \mathbb{R}$是一个凸惩罚函数，残差可代表对回归模型的估计误差

□ 可以使用不同的罚函数形式：

1. 二次函数：$\phi(u) = u^2$
2. 死区-线性宽度为 a: $\phi(u) = \max\{0, |u| - a\}$
3. 对数障碍限制为 a:

$$\phi(u) = \begin{cases} -a^2 \log(1 - (u/a)^2) & |u| < a \\ \infty & \text{其他情况} \end{cases}$$



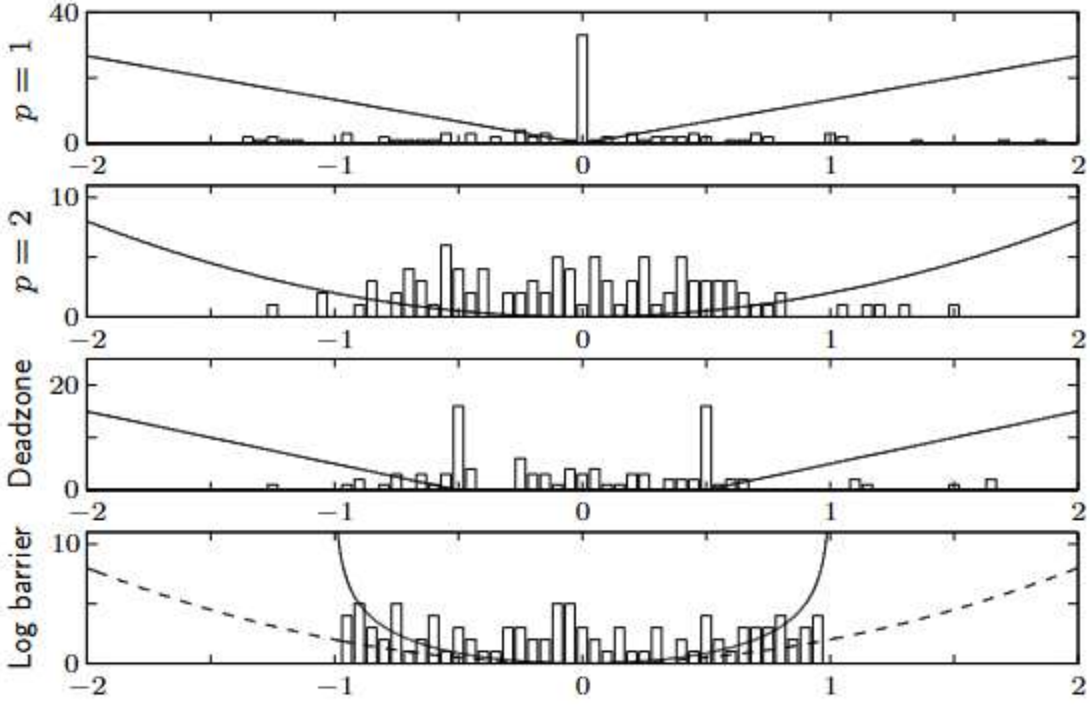
1.2 罚函数逼近

- **例** 1: ($m = 100, n = 30$): 惩罚的残差直方图

$$\phi(u) = |u|, \phi(u) = u^2,$$

$$\phi(u) = \max\{0, |u| - a\}, \phi(u) = -\log(1 - u^2)$$

罚函数的形状对残差的分布有重大影响

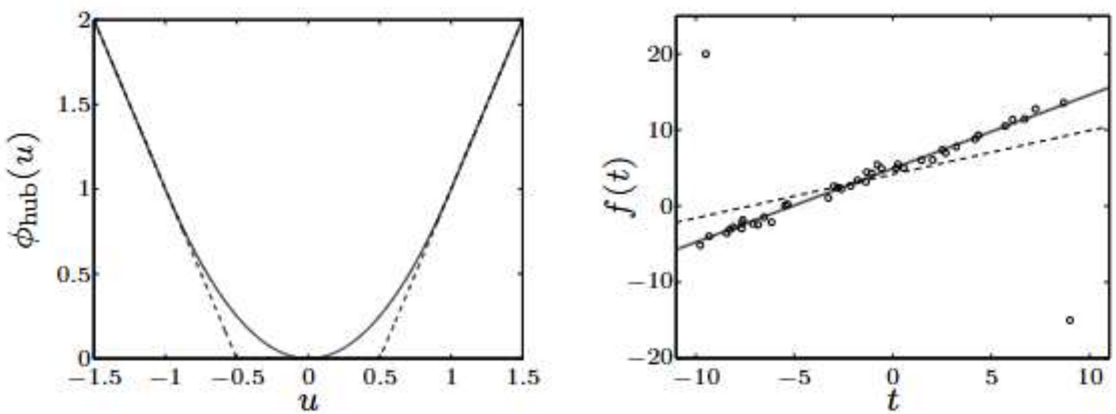


1.2 罚函数逼近

□ **Huber 罚函数**(带参数 M)，也叫做鲁棒最小二乘：

$$\phi_{\text{hub}}(u) = \begin{cases} u^2 & |u| \le M \\ M(2|u| - M) & |u| > M \end{cases}$$

线性增长对于较大的 u 使得逼近对离群值不敏感



左侧:$M = 1$ 时的 Huber 惩罚

右侧:通过二次(虚线)和 Huber (实线)惩罚拟合 42 个点 t_i, y_i (圆圈)的仿射函数 $f(t) = \alpha + \beta t$

1.3 正则化逼近

$$最小化(\mathrm{w.r.t.}\,\mathbb{R}_+^2) \quad (\|Ax-b\|, \|x\|)$$

$(A \in \mathbb{R}^{m\times n}, m \le n)$

- 解释:找到良好的逼近解 $Ax \approx b$,且 x 的值较小

1. **估计:**线性测量模型 $y = Ax + v$, 已知 $\| x \|$ 较小
2. **最优设计:**x 较小更经济或更高效,或线性模型 $y = Ax$ 仅在 x 较小的情况下有效
3. **鲁棒逼近:**x 较小的良好逼近解 $Ax \approx b$ 对 A 带来的误差不敏感,而 x 较大的良好逼近解则敏感

1.3 正则化逼近

■ **标量化逼近因子和正则化项**

最小化 $\|Ax-b\|+\gamma\|x\|$

对于不同 $\gamma > 0$，求解上述问题可以给出最优权衡曲线

Tikhonov 正则化

最小化 $\|Ax-b\|_2^2+\delta\|x\|_2^2$

可以作为一个最小二乘问题来求解

最小化 $\left\|\begin{bmatrix} A \\ \sqrt{\delta} I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix}\right\|_2^2$

可以得到解析解为 $x^* = (A^TA+\delta I)^{-1}A^Tb$

1.3 正则化逼近

□ 例 2:最优输入设计

线性时不变系统, 其脉冲响应为 h:

$$y(t)=\sum_{\tau=0}^{t} h(\tau)u(t-\tau)\,, t=0,1,\cdots,N$$

输入设计问题:具有 3 个目标的多元目标问题

跟踪误差与期望输出 y_{des} : $J_{\mathrm{track}}=\sum_{t=0}^{N}\left(y(t)-y_{\mathrm{des}}(t)\right)^2$

输入幅度:$J_{\mathrm{mag}}=\sum_{t=0}^{N} u(t)^2$

输入变化:$J_{\mathrm{der}}=\sum_{t=0}^{N-1}\left(u(t+1)-u(t)\right)^2$

使用一个微小且缓慢变化的输入信号来跟踪期望输出

1.3 正则化逼近

□ 例 2:最优输入设计

正则化最小二乘法公式

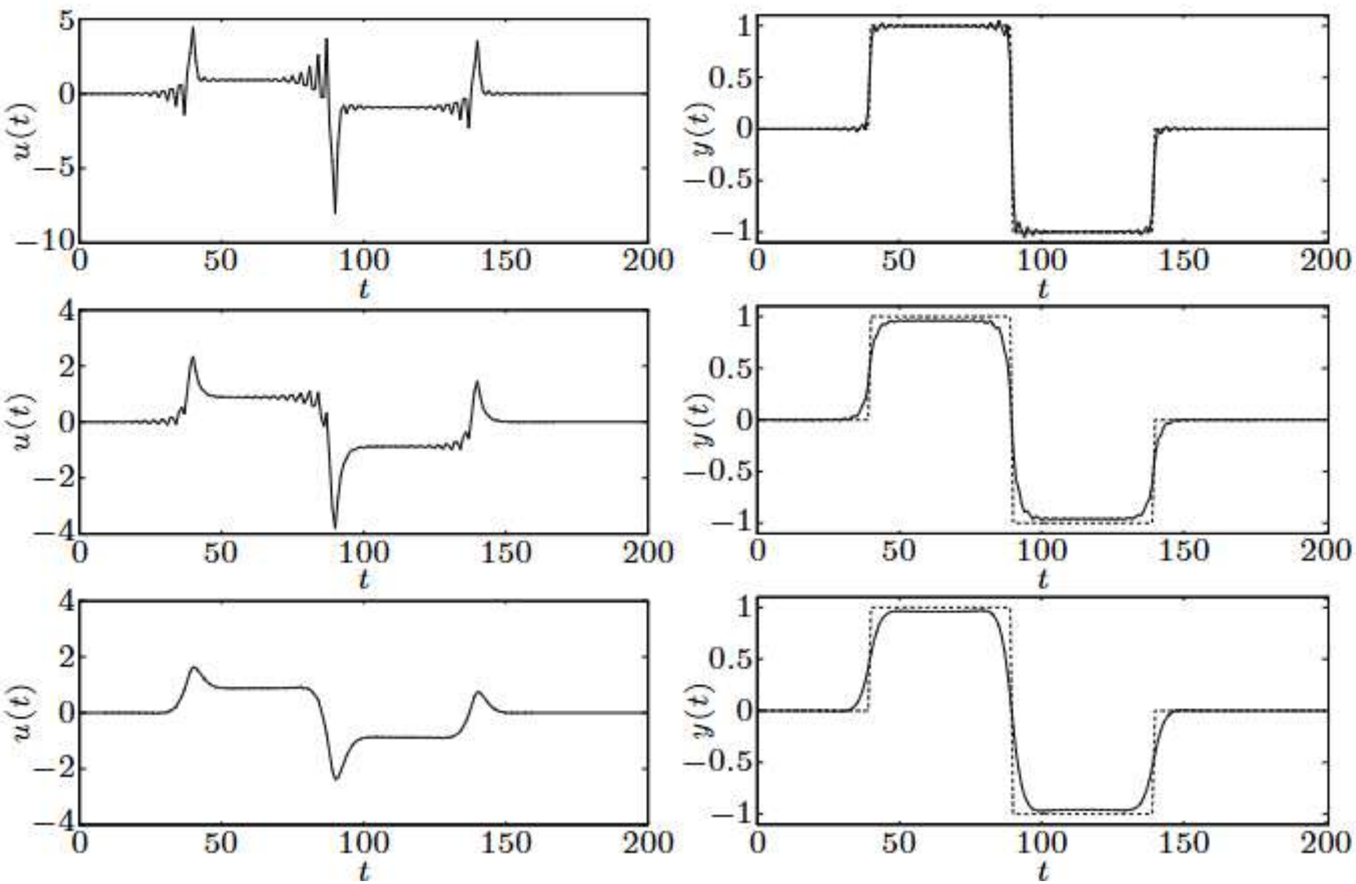
$$\text{最小化} \qquad J_{\text{track}} + \delta J_{\text{der}} + \eta J_{\text{mag}}$$

对于固定的 $\delta, \eta,$ 上述问题也是一个最小二乘问题

1.3 正则化逼近

☐ 例 2:最优输入设计

(顶部) $\delta = 0$, 小 η; (中间) $\delta = 0$, 大 η; (底部) 大 δ, η



1.3 正则化逼近

■ **例 3：信号重建**

$$\text{最小化}\ (\text{w.r.t.}\ \mathbb{R}_+^2)\quad (\|\hat{x}-x_{\text{cor}}\|,\phi(\hat{x}))$$

$x\in\mathbb{R}^n$ 是未知信号

$x_{\text{cor}}=x+v$ 是(已知)被损坏的 x 版本，带有加性噪声

变量 $\hat{x}$ (重建信号)是 x 的估计

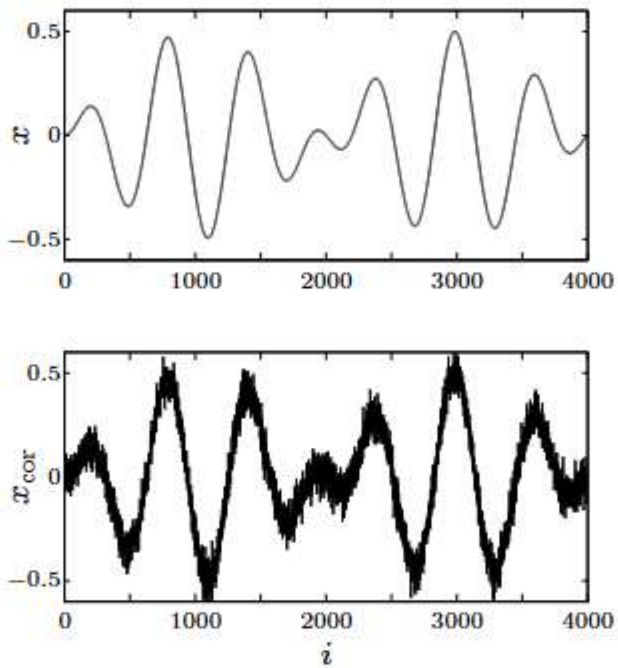
$\phi:\mathbb{R}^n\rightarrow\mathbb{R}$ 是正则化函数或平滑目标

示例：二次平滑，总变化平滑：

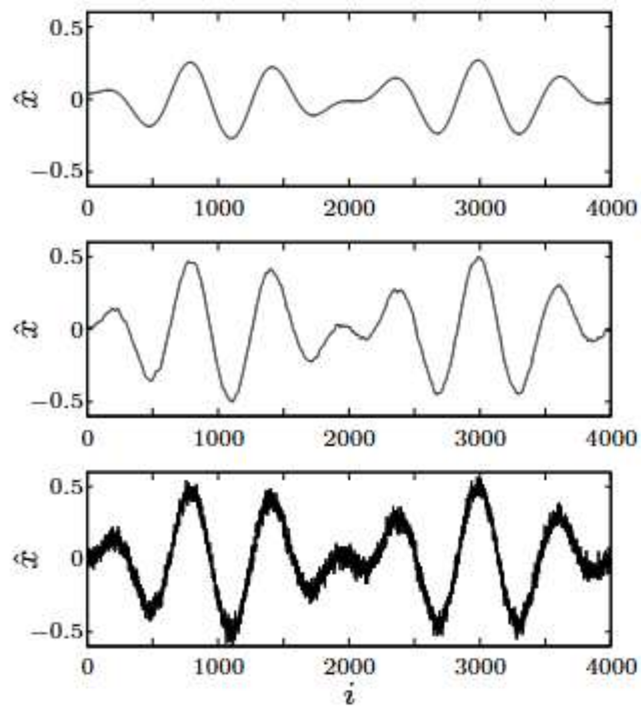
$$\phi_{\text{quad}}(\hat{x})=\sum_{i=1}^{n}(\hat{x}_{i+1}-\hat{x}_i)^2,\phi_{\text{tv}}(\hat{x})\ =\sum_{i=1}^{n}|\hat{x}_{i+1}-\hat{x}_i|$$

1.3 正则化逼近

□ 二次平滑示例



original signal x and noisy signal x_{cor}



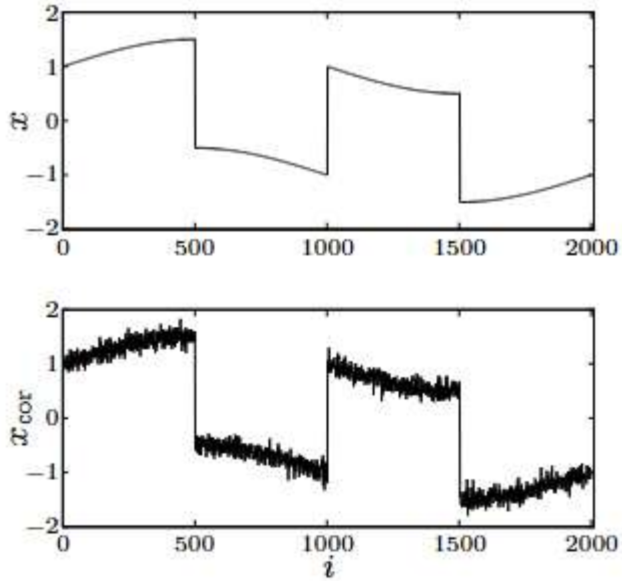
three solutions on trade-off curve $\|\hat{x} - x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{quad}}(\hat{x})$

中间达到最优折衷效果：在保持原信号形状前提下滤除毛刺噪声

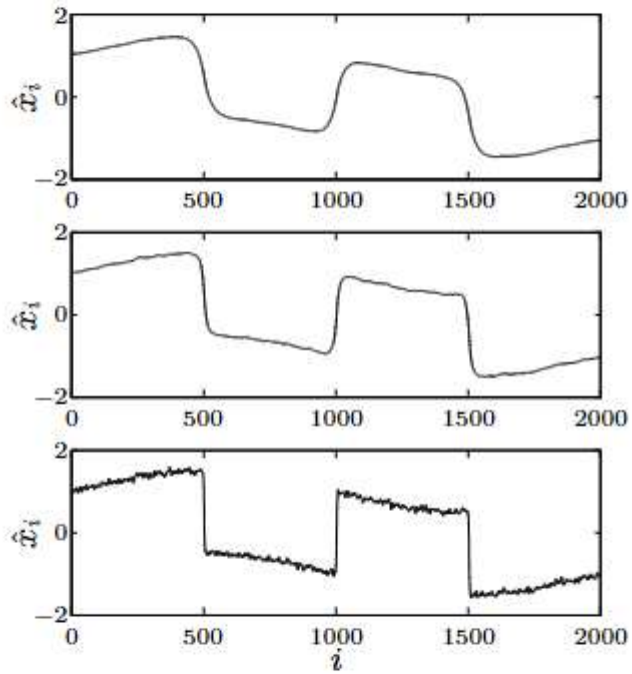
1.3 正则化逼近

□ 总变差重建示例

二次平滑可以消除信号中的噪声和急剧变化



original signal x and noisy signal x_{cor}

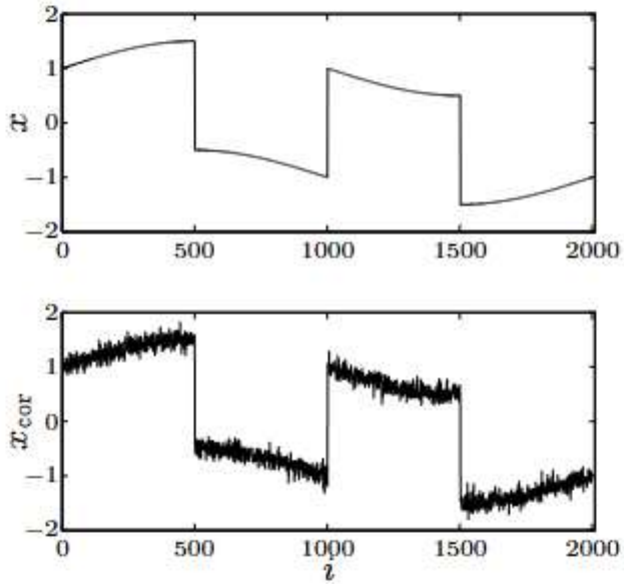


three solutions on trade-off curve $\|\hat{x}-x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{quad}}(\hat{x})$

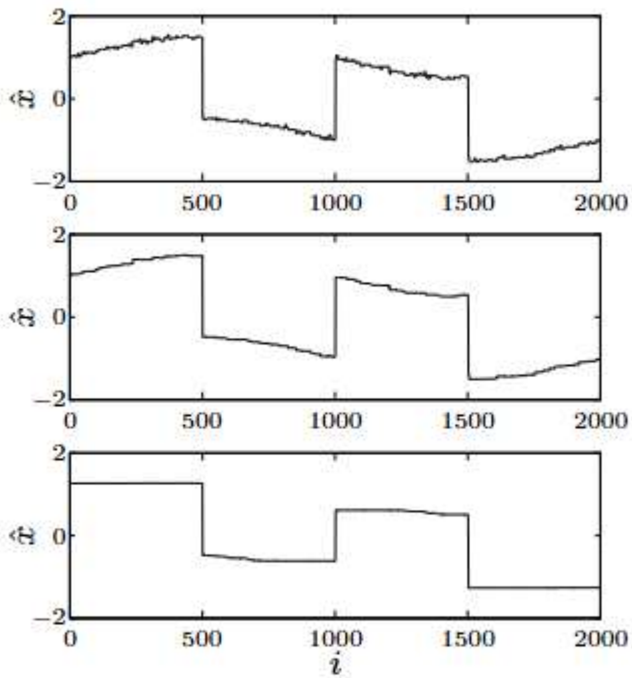
效果均无法达到最优，二次平滑也会消除信号中的急剧变化

1.3 正则化逼近

- 总变分平滑保留了信号中的锐利的过渡



original signal x and noisy signal x_{cor}



three solutions on trade-off curve $\|\hat{x} - x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{tv}}(\hat{x})$

1.4 鲁棒逼近

最小化 $\|Ax-b\|$，其中 A 不确定，有两种方法：

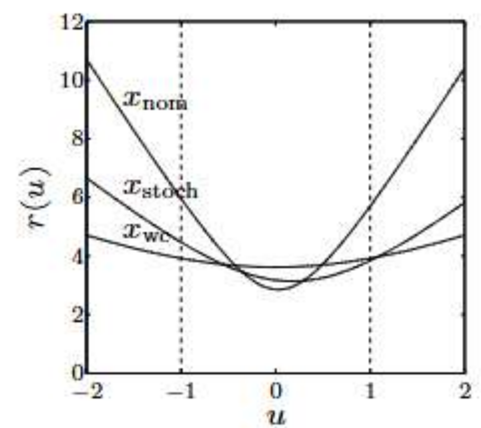
1. 随机：假设 A 是随机的，最小化 $\mathbf{E}\|Ax-b\|$
2. 最坏情况（min-max）：设置 $\mathcal{A}$ 为 A 的可能值，最小化 $\sup_{A\in\mathcal{A}}\|Ax-b\|$，仅在特殊情况（特定的范数 $\|\cdot\|$，分布，集合 $\mathcal{A}$）下可解

- **例** 4：$A(u)=A_0+uA_1$

a. x_{norm} 最小化 $\|A_0x-b\|$

b. x_{stoch} 最小化 $\mathbf{E}\|A(u)x-b\|_2^2$，其中 u 在 $[-1,1]$ 上均匀分布

c. 最小化 $\sup_{-1\le u\le 1}\|A(u)x-b\|_2^2$ 图示出了 $r(u)=\|A(u)x-b\|_2$



1.4 鲁棒逼近

随机鲁棒 LS, 其中 $A = \bar{A} + U, U$ 随机, $\mathbf{E}U = 0, \mathbf{E}U^T U = P$

$$\text{最小化} \quad \mathbf{E}\|(\bar{A} + U)_x - b\|_2^2$$

1. 目标函数的显式表达式:

$$\mathbf{E}\|Ax - b\|_2^2 = \mathbf{E}\|\bar{A}x - b + Ux\|_2^2$$
$$= \|\bar{A}x - b\|_2^2 + \mathbf{E}x^T U^T U x = \|\bar{A}x - b\|_2^2 + x^T P x$$

2. 因此, 鲁棒最小二乘问题等价于如下最小二乘问题

$$\text{最小化} \quad \|\bar{A}x - b\|_2^2 + \left\|P^{1/2}x\right\|_2^2$$

3. 当 $P = \delta I$ 时, 得到 Tikhonov 正则化问题

$$\text{最小化} \quad \|\bar{A}x - b\|_2^2 + \delta\|x\|_2^2$$

1.4 鲁棒逼近

- **最坏情况鲁棒 LS 满足** $\mathcal{A}=\{\bar{A}+u_1A_1+\cdots+u_pA_p \mid \|u\|_2\le 1\}$

$$\text{最小化} \sup_{A\in\mathcal{A}}\|\bar{A}x-b\|_2^2=\sup_{\|u\|_2\le 1}\|P(x)u+q(x)\|_2^2$$

其中$P(x)=[A_1x,A_2x,\cdots,A_px],q(x)=\bar{A}x-b$

1. 以下问题之间存在强对偶性

$$\begin{array}{ll}\text{最大化} & \|P(x)u+q(x)\|_2^2 \\ \text{约束条件} & \|u\|_2^2\le 1\end{array}$$

$$\begin{array}{ll}\text{最小化} & t+\lambda \\ \text{约束条件} & \begin{bmatrix} I & P & q \\ P^T & \lambda I & 0 \\ q(x)^T & 0 & t \end{bmatrix}\succeq 0\end{array}$$

1.4 鲁棒逼近

2. 因此, 鲁棒 LS 问题是等价于 SDP

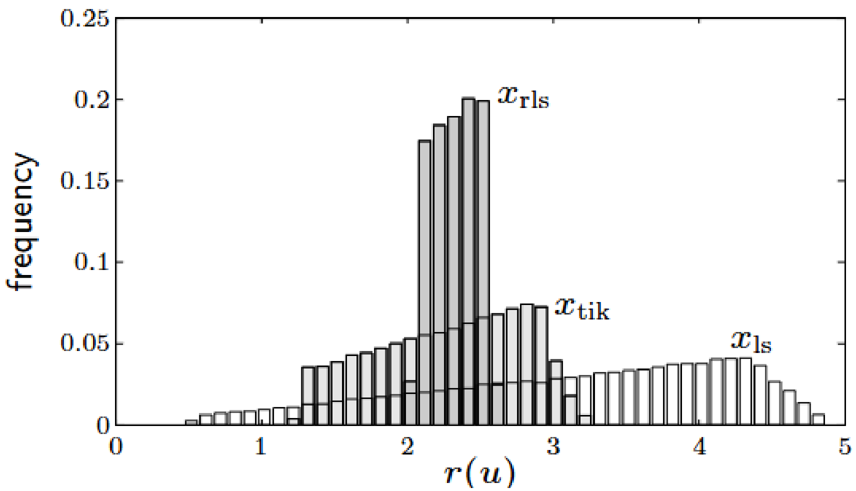
$$
\begin{array}{ll}
\text{最小化} & t+\lambda \\
\text{约束条件} & \begin{bmatrix} I & P(x) & q(x) \\ P(x)^T & \lambda I & 0 \\ q(x)^T & 0 & t \end{bmatrix} \succeq 0
\end{array}
$$

1.4 鲁棒逼近

□ 例 5:

$$r(u) = \|(A_0 + u_1A_1 + u_2A_2)x - b\|_2$$

其中 u 在单位圆上均匀分布,对于 x 的三个值



大 纲

1. 逼近

2. LASSO

2 LASSO

■ 什么是正则化？

在凸优化中，正则化是一个修改目标函数的特定技术，其目的是保证问题的良性态或诱导解具有某种期望的性质

通常是在目标函数后添加一个惩罚项

$$minimize \quad f_0(x) + \lambda R(x)$$

其中，$R(x)$ 是正则化项，是一个凸函数，用来惩罚我们不希望看到的解的性质；$\lambda > 0$ 是正则化参数，用于控制我们对 $R(x)$ 的重视程度

2 LASSO

■ 解决病态问题

$$\text{最小化} \qquad \|Ax-b\|_2^2$$

当 A 的列近似线性相关时,解 $x*$ 可能对数据 b 中的噪声极其敏感,且数值不稳定,使用1.3.1中介绍过的 **Tikhonov 正则化**

$$\text{最小化} \qquad \|Ax-b\|_2^2+\lambda\|x\|_2^2$$

这个新问题被称为**岭回归**. 附加项 $\lambda\|x\|_2^2$ 确保了目标函数是严格凸的,从而有唯一解. 这个解对噪声的敏感度更低,数值计算也更稳定

2 LASSO

■ 诱导稀疏性

$$\text{最小化} \qquad \|Ax - b\|_2^2$$

我们可以通过精心选择正则化函数 $R(x)$ 来鼓励解 x^* 拥有我们想要的某种性质, 如下列正则化函数

$$\text{最小化} \qquad \|Ax + b\|_2^2 + \lambda\|x\|_1$$

这个新问题被称为 **Lasso 回归**. l_1-范数正则化项 $\|x\|_1$ 具有一个很好的性质: 它倾向于产生稀疏解,即最优解 x^* 中许多分量恰好为零;当 λ 逐渐增大时,某些特征的系数会直接变为零,从而自动选择重要特征

谢 谢 ！