第三讲 线性最小二乘问题

问题介绍 • Householder/Givens 变换

目录

3-1 问题介绍

3.1 问题介绍

线性最小二乘问题

线性最小二乘问题 就是求解下面的最值问题:

$$\boxed{\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2} \tag{3.1}$$

其中 $A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$. 问题 (3.1) 的解称为**最小二乘解**.

线性最小二乘问题

线性最小二乘问题 就是求解下面的最值问题:

$$\boxed{\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2} \tag{3.1}$$

其中 $A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$. 问题 (3.1) 的解称为**最小二乘解**.

- 当 $m=n$ 且 A 非奇异时, 这就是一个线性方程组
- 当 $m>n$ 时, 约束个数大于未知量个数, 问题为 **超定**的;
- 当 $m<n$ 时, 未知量个数大于约束个数, 问题 (3.1) 为 **欠定** (或亚定) 的.

为了讨论方便, 除非特别说明, 本讲总是假定 A 是满秩的.

3-1-1 超定方程组

当 $m > n$ 时, 方程个数大于未知量个数, $Ax = b$ 的解可能不存在.

3-1-1 超定方程组

当 $m > n$ 时, 方程个数大于未知量个数, $Ax = b$ 的解可能不存在.

记目标函数: $$J(x) \triangleq \|Ax - b\|_2^2$$

易知: $J(x)$ 是关于 x 的二次函数, 而且是凸函数.

因此, x_* 是问题 (3.1) 的解当且仅当 x_* 是 $J(x)$ 的稳定点, 即满足

$$\frac{\partial J}{\partial x} = 0 \quad \Longleftrightarrow \quad A^{\intercal}Ax - A^{\intercal}b = 0$$

如果 A 不是满秩, 则 $A^{\intercal}A$ 半正定, 此时解不唯一.

3-1-2 欠定方程组

当 $m < n$ 时, 方程个数小于未知量个数, 存在无穷多个解 (假定 A 满秩), 是不适定问题.

3-1-2 欠定方程组

当 $m < n$ 时, 方程个数小于未知量个数, 存在无穷多个解 (假定 A 满秩), 是不适定问题.

这时我们通常寻求最小范数解, 于是原问题就转化为下面的 **约束优化问题**

$$\min_{Ax=b} \frac{1}{2}\|x\|_2^2 \tag{3.2}$$

对应的 **Lagrange 函数**为

$$\mathcal{L}(x,\lambda) = \frac{1}{2}\|x\|_2^2 + \lambda^{\mathsf{T}}(Ax-b),$$

其中 $\lambda = [\lambda_1, \ldots, \lambda_m]^{\mathsf{T}}$ 是 **Lagrange 乘子**. 问题 (3.2) 的解就是 $\mathcal{L}(x,\lambda)$ 的**鞍点**, 即:

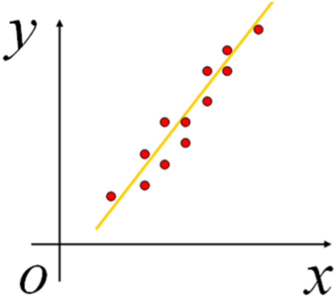
$$\frac{\partial \mathcal{L}}{\partial x} = x + A^{\mathsf{T}}\lambda = 0, \quad \frac{\partial \mathcal{L}}{\partial \lambda} = Ax - b = 0.$$

写成矩阵形式为 $\begin{bmatrix} I & A^{\mathsf{T}} \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix}$. 如果 A 满秩则存在唯一解.

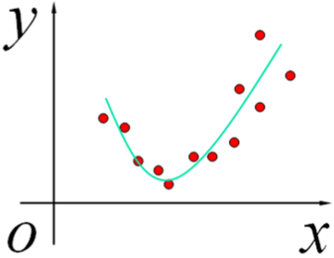
本课程主要讨论 超定问题

应用: 多项式数据拟合

已知 m 个数据样本 $\{(x_i, f_i)\}_{i=1}^m$, 寻找一个低次多项式来拟合这些数据.



$y = ax + b$



$y = ax^2 + bx + c$

3-2 Householder 变换与 Givens 变换

3.2 Householder 变换与 Givens 变换

矩阵变换

矩阵变换/线性变换

矩阵计算的一个基本思想就是把复杂的问题转化为等价的且易于求解的问题.

完成这个转化的基本工具就是 **矩阵变换**.

矩阵变换

矩阵变换/线性变换

矩阵计算的一个基本思想就是把复杂的问题转化为等价的且易于求解的问题.
完成这个转化的基本工具就是 **矩阵变换**.

- 比如高等代数中的三类初等变换.

- 除此之外, 在矩阵计算中常用的**重要矩阵变换**有:
  - Householder 变换
  - Givens 变换

以上两类矩阵变换都是正交变换, 可用于求解最小二乘问题、特征值、奇异值等问题

3-2-1 Householder 变换

定义 我们称矩阵

$$H = I - \frac{2}{v^*v}vv^* = I - \frac{2}{\|v\|_2^2}vv^*, \quad 0 \neq v \in \mathbb{C}^n, \tag{3.3}$$

为 **Householder 变换** (或 **Householder 矩阵**, 或 **Householder 反射**), 向量 v 称为 **Householder 向量**. 我们通常将矩阵 (3.3) 记为 $H(v)$.

✍ Householder 矩阵是单位矩阵的秩-1 修正.

✍ Householder 变换也可以定义为

$$H = I - 2vv^*, \quad v \in \mathbb{C}^n \text{ 且 } \|v\|_2 = 1.$$

✍ Householder 矩阵由 Householder 向量唯一确定.

Householder 变换的几何意义

从几何上看, 一个 Householder 变换是一个关于超平面 $\mathrm{span}\{v\}^{\perp}$ 的反射.

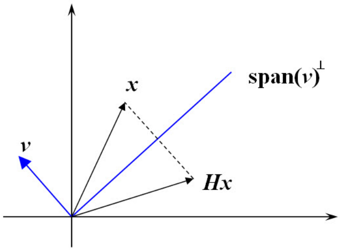
由于 $\mathbb{C}^2 = \mathrm{span}\{v\} \oplus \mathrm{span}\{v\}^{\perp}$, 因此对任意一个向量 $x \in \mathbb{C}^n$, 都可写成

$$x = \frac{v^* x}{v^* v} v + y \triangleq \alpha v + y,$$

其中 $\alpha v \in \mathrm{span}\{v\}$, $y \in \mathrm{span}\{v\}^{\perp}$. 于是

$$Hx = x - \frac{2}{v^* v} v v^* x = x - 2\alpha v = -\alpha v + y,$$

即 Hx 与 x 在 $\mathrm{span}\{v\}^{\perp}$ 方向有着相同的分量, 而在 v 方向的分量正好相差一个符号. 也就是说, Hx 是 x 关于超平面 $\mathrm{span}\{v\}^{\perp}$ 的镜面反射. 因此, Householder 变换也称为 Householder 反射.



Householder 矩阵的基本性质

定理 设 $H \in \mathbb{C}^{n\times n}$ 是一个 Householder 矩阵, 则

(1) $H^* = H$, 即 H Hermite 的;

(2) $H^*H = I$, 即 H 是酉矩阵;

(3) $H^2 = I$, 所以 $H^{-1} = H$;

(4) $\det(H) = -1$;

(5) H 有两个互异的特征值: $\lambda = 1$ 和 $\lambda = -1$, 其中 $\lambda = 1$ 的代数重数为 $n-1$.

(留作课外自习)

Householder 变换的重要应用: 化零

Householder 矩阵的一个非常重要的应用就是可以将一个向量除第一个元素以外的所有元素都化为零.

引理 设 $x, y \in \mathbb{C}^n$ 为任意两个互异的向量, 则存在一个 Householder 矩阵 $H(v)$ 使得 $y = H(v)x$ 的充要条件是 $\|x\|_2 = \|y\|_2$ 且 $y^*x \in \mathbb{R}$. (板书)

如果 x, y 都是实向量, 则条件 $y^*x \in \mathbb{R}$ 自然成立, 此时充要条件就是 $\|x\|_2 = \|y\|_2$.

Householder 变换的重要应用: 化零

Householder 矩阵的一个非常重要的应用就是可以将一个向量除第一个元素以外的所有元素都化为零.

引理 设 $x, y \in \mathbb{C}^n$ 为任意两个互异的向量, 则存在一个 Householder 矩阵 $H(v)$ 使得 $y = H(v)x$ 的充要条件是 $\|x\|_2 = \|y\|_2$ 且 $y^*x \in \mathbb{R}$. (板书)

如果 x, y 都是实向量, 则条件 $y^*x \in \mathbb{R}$ 自然成立, 此时充要条件就是 $\|x\|_2 = \|y\|_2$.

定理 设 $x = [x_1, x_2, \ldots, x_n]^\mathsf{T} \in \mathbb{R}^n$ 是一个非零向量, 则存在 Householder 矩阵 $H(v)$ 使得 $H(v)x = \alpha e_1$, 其中 $\alpha = \|x\|_2$ (或 $\alpha = -\|x\|_2$), $e_1 = [1, 0, \ldots, 0]^\mathsf{T} \in \mathbb{R}^n$.

在后面的讨论中, 我们将定理的向量 v 称为 x 对应的 Householder 向量.

Householder 向量的计算

设 $x = [x_1, x_2, \ldots, x_n]^\intercal \in \mathbb{R}^n$ 非零, 如何计算 Householder 向量 v, 使得 $H(v)x = \alpha e_1$?

Householder 向量的计算

设 $x = [x_1, x_2, \dots, x_n]^{\mathsf{T}} \in \mathbb{R}^n$ 非零, 如何计算 Householder 向量 v, 使得 $H(v)x = \alpha e_1$?

由前面的证明过程可知

$$v = x - \alpha e_1 = [x_1 - \alpha, x_2, \dots, x_n]^{\mathsf{T}}$$

Householder 向量的计算

设 $x = [x_1, x_2, \ldots, x_n]^\mathsf{T} \in \mathbb{R}^n$ 非零, 如何计算 Householder 向量 v, 使得 $H(v)x = \alpha e_1$?

由前面的证明过程可知

$$\boxed{v = x - \alpha e_1 = [x_1 - \alpha, x_2, \ldots, x_n]^\mathsf{T}}$$

实际计算时, 为了尽可能减少舍入误差, 避免两个相近的数相减, 我们通常取

$$\boxed{\alpha = -\operatorname{sign}(x_1) \cdot \|x\|_2} \tag{3.4}$$

事实上, 我们也可以取 $\alpha = \text{sign}(x_1)\|x\|_2$, 但需要改变计算方法:

$$\alpha = \text{sign}(x_1)\|x\|_2, \quad v_1 = x_1 - \alpha = \frac{x_1^2 - \|x\|_2^2}{x_1 + \alpha} = \frac{-(x_2^2 + x_3^2 + \cdots + x_n^2)}{x_1 + \alpha}. \tag{3.5}$$

事实上, 我们也可以取 $\alpha = \operatorname{sign}(x_1)\|x\|_2$, 但需要改变计算方法:

$$\alpha = \operatorname{sign}(x_1)\|x\|_2, \quad v_1 = x_1 - \alpha = \frac{x_1^2 - \|x\|_2^2}{x_1 + \alpha} = \frac{-(x_2^2 + x_3^2 + \cdots + x_n^2)}{x_1 + \alpha}. \tag{3.5}$$

在 v_1 的两种计算方法 (3.4) 和 (3.5) 中, α 的取值都与 x_1 的符号有关. 但在某些应用中, 我们需要确保 α 非负, 此时我们可以将这两种方法结合起来使用, 即:

$$v_1 = \begin{cases} x_1 - \alpha, & \text{if } \operatorname{sign}(x_1) < 0 \\ \dfrac{-(x_2^2 + x_3^2 + \cdots + x_n^2)}{x_1 + \alpha}, & \text{otherwise} \end{cases}$$

无论怎样计算 v, 我们都有 $H = I - \beta v v^*$, 其中

$$\beta = \frac{2}{v^* v} = \frac{2}{(x_1 - \alpha)^2 + x_2^2 + \cdots + x_n^2} = \frac{2}{2\alpha^2 - 2\alpha x_1} = -\frac{1}{\alpha v_1}.$$

算法 计算 Householder 向量 (总运算量大约为 $2n$)

% Given $x \in \mathbb{R}^n$, compute $v \in \mathbb{R}^n$ such that $Hx = \|x\|_2 e_1$, where $H = I - \beta vv^*$

function $[\beta, v]$ = **House**(x)
$n = \texttt{length}(x)$ (here $\texttt{length}(x)$ denotes the dimension of x)
$\sigma = x_2^2 + x_3^2 + \cdots + x_n^2, \quad v = x$
if $\sigma = 0$ **then**
    if $x_1 < 0$ **then**
        $v_1 = 2x_1, \beta = 2/v_1^2$
    else
        $v_1 = 0, \beta = 0$
    end if
else
    $\alpha = \sqrt{x_1^2 + \sigma}$ % $\alpha = \|x\|_2$
    if $x_1 < 0$ **then**
        $v_1 = x_1 - \alpha$
    else
        $v_1 = -\sigma/(x_1 + \alpha)$
    end if
    $\beta = 2/(v_1^2 + \sigma)$
end if

算法稳定性

可以证明, 上述算法具有很好的数值稳定性, 即

$$\|\tilde{H} - H\|_2 = \mathcal{O}(\varepsilon_u),$$

其中 $\tilde{H}$ 是由上述算法计算得到的近似矩阵, ε_u 是机器精度.

- ✍ 在实际计算时, 我们可以将向量 v 规范化, 使得 $v_1 = 1$. 这样, 我们就无需为 v 另外分配空间, 而是将 $v(2:n)$ 存放在 $x(2:n)$ 中.
- ✍ 为了避免可能产生的溢出, 我们也可以事先将 x 单位化, 即令 $x = x/\|x\|_2$

Householder 变换与向量和矩阵的乘积

对向量和矩阵做 Householder 变换时, 无需显式写出 Householder 矩阵.

设 $x \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$, $H = I - \beta vv^* \in \mathbb{R}^m$, 则

$$Hx = (I - \beta vv^*)x = x - \beta(v^*x)v,$$

$$HA = (I - \beta vv^*)A = A - \beta v(v^*A).$$

运算量大约分别为 4m 和 $4mn$, 远小于一般的矩阵-向量乘积和矩阵-矩阵乘积的运算量.

3-2-2 Givens 变换

为简单起见, 我们这里这讨论**实数域**中的 Givens 变换.

设 $\theta \in [0, 2\pi]$, 我们称矩阵

$$G(i,j,\theta) = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & c & & s & & \\ & & & \ddots & & & \\ & & -s & & c & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}, \qquad (i<j)$$

为 **Givens 变换** (或 **Givens 旋转**, 或 **Givens 矩阵**), 其中 $c = \cos(\theta)$, $s = \sin(\theta)$, 即将单位矩阵的 (i,i) 和 (j,j) 位置上的元素用 c 代替, 而 (i,j) 和 (j,i) 位置上的元素分别用 s 和 $-s$ 代替.

Givens 变换的基本性质

定理 $G(i,j,\theta)$ 是正交矩阵, 且 $\det(G(i,j,\theta)) = 1$.

Givens 变换的基本性质

定理 $G(i,j,\theta)$ 是正交矩阵, 且 $\det(G(i,j,\theta)) = 1$.

✍ Givens 变换可以看作是单位矩阵的一个秩-2 修正, 即

$$G(i,j,\theta) = I + [e_i, e_j] \begin{bmatrix} \cos(\theta) - 1 & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) - 1 \end{bmatrix} [e_i, e_j]^{\mathsf{T}}.$$

因此当一个矩阵左乘一个 Givens 矩阵时, 只会影响其第 i 行和第 j 行的元素.

而当一个矩阵右乘一个 Givens 矩阵时, 只会影响其第 i 和第 j 列的元素.

Givens 变换化零

例 设 $x=[x_1,x_2]^\mathsf{T}\in\mathbb{R}^2$, 则存在一个 Givens 变换 $G=\begin{bmatrix} c & s \\ -s & c\end{bmatrix}\in\mathbb{R}^{2\times 2}$ 使得 $Gx=[r,0]^\mathsf{T}$, 其中 c,s 和 r 的值如下:

$$c=\frac{x_1}{r},\quad s=\frac{x_2}{r},\quad r=\sqrt{x_1^2+x_2^2},$$

即 $G=\dfrac{1}{r}\begin{bmatrix} x_1 & x_2 \\ -x_2 & x_1\end{bmatrix}$

也就是说, 通过 Givens 变换, 我们可以将向量 $x\in\mathbb{R}^2$ 的第二个分量化为 0.

Givens 变换化零

例 设 $x = [x_1, x_2]^\mathsf{T} \in \mathbb{R}^2$, 则存在一个 Givens 变换 $G = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \in \mathbb{R}^{2\times 2}$ 使得 $Gx = [r, 0]^\mathsf{T}$, 其中 c, s 和 r 的值如下:

$$c = \frac{x_1}{r}, \quad s = \frac{x_2}{r}, \quad r = \sqrt{x_1^2 + x_2^2},$$

即 $G = \dfrac{1}{r} \begin{bmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{bmatrix}$

也就是说, 通过 Givens 变换, 我们可以将向量 $x \in \mathbb{R}^2$ 的第二个分量化为 0.

- 事实上, 我们都可以通过 Givens 变换将 $x \in \mathbb{R}^n$ 任意一个位置上的分量化为 0.
- 进一步, 可通过若干 Givens 变换, 将 x 中除第一个分量外的所有元素都化为 0.

算法 Givens 变换

% Given $x = [a, b]^\intercal \in \mathbb{R}^2$, compute c and s such that $Gx = [r, 0]^\intercal$ where $r = \|x\|_2$

function $[c, s]$ = **givens**(a, b)

if $b = 0$ **then**

  if $a \geq 0$ **then**

    $c = 1, \quad s = 0$

  else

    $c = -1, \quad s = 0$

  end if

else

  if $|b| > |a|$ **then** % 绝对值大的数作为分母

    $\tau = \dfrac{a}{b}, \quad s = \dfrac{\mathrm{sign}(b)}{\sqrt{1+\tau^2}}, \quad c = s\tau$

  else

    $\tau = \dfrac{b}{a}, \quad c = \dfrac{\mathrm{sign}(a)}{\sqrt{1+\tau^2}}, \quad s = c\tau$

  end if

end if

注记

正交变换与 Householder/Givens 变换

任何一个正交矩阵都可以写成若干个 Householder 矩阵或 Givens 矩阵的乘积 (见习题), 所以正交矩阵所对应的线性变换可以看作是反射变换和旋转变换的组合, 因此它不会改变向量的长度与 (不同向量之间的) 角度.

3-2-3 正交变换的舍入误差分析

引理 设 $P \in \mathbb{R}^{n\times n}$ 是一个精确的 Householder 或 Givens 变换, $\tilde{P}$ 是其浮点运算近似, 则

$$\mathrm{fl}(\tilde{P}A) = P(A+E), \quad \mathrm{fl}(A\tilde{P}) = (A+F)P,$$

其中 $\|E\|_2 = \mathcal{O}(\varepsilon_u)\cdot\|A\|_2$, $\|F\|_2 = \mathcal{O}(\varepsilon_u)\cdot\|A\|_2$. (这里 ε_u 表示机器精度)

这说明对一个矩阵做 Householder 变换或 Givens 变换是向后稳定的.

正交变换的稳定性

定理 考虑对矩阵 A 做一系列的正交变换 $P_1, P_2, \ldots, P_k$ 和 $Q_1, Q_2, \ldots, Q_k$, 则有

$$\mathrm{fl}(\tilde{P}_k \cdots \tilde{P}_1 A \tilde{Q}_1 \cdots \tilde{Q}_k) = P_k \cdots P_1 (A + E) Q_1 \cdots Q_k,$$

其中 $\|E\|_2 = \mathcal{O}(\varepsilon_u) \cdot (k\|A\|_2)$. 这说明整个计算过程是向后稳定的

正交变换的稳定性

定理 考虑对矩阵 A 做一系列的正交变换 $P_1, P_2, \ldots, P_k$ 和 $Q_1, Q_2, \ldots, Q_k$, 则有

$$\mathrm{fl}(\tilde{P}_k \cdots \tilde{P}_1 A \tilde{Q}_1 \cdots \tilde{Q}_k) = P_k \cdots P_1 (A+E) Q_1 \cdots Q_k,$$

其中 $\|E\|_2 = \mathcal{O}(\varepsilon_u) \cdot (k\|A\|_2)$. 这说明整个计算过程是向后稳定的

一般地, 设 X 是非奇异线性变换, $\tilde{X}$ 是其浮点运算近似, 当 X 作用到 A 上时, 我们有

$$\mathrm{fl}(\tilde{X}A) = XA + E = X(A + X^{-1}E) \triangleq X(A+F),$$

其中 $\|E\|_2 = \mathcal{O}(\varepsilon_u) \cdot \|XA\|_2 \leq \mathcal{O}(\varepsilon_u) \cdot \|X\|_2 \cdot \|A\|_2$, 故

$$\|F\|_2 = \|X^{-1}E\|_2 \leq \mathcal{O}(\varepsilon_u) \cdot \|X^{-1}\|_2 \cdot \|X\|_2 \cdot \|A\|_2 = \mathcal{O}(\varepsilon_u) \cdot \kappa_2(X) \cdot \|A\|_2,$$

因此, 舍入误差可能会被放大 $\kappa_2(X)$ 倍. 当 X 是正交变换时, $\kappa_2(X)$ 达到最小值 1, 这就是为什么在浮点运算中尽量使用正交变换的原因.

谢谢

THANK YOU