第三讲　线性最小二乘问题

最小二乘问题

最小二乘问题有着广泛的应用背景，如数据拟合，最优控制，信号与图像处理，压缩感知，机器学习，数据科学等, 是计算数学的一个重要研究分支, 也是一个非常活跃的研究领域.

最小二乘问题

线性最小二乘问题, 总体最小二乘问题, 约束最小二乘问题, ...

本讲主要介绍求解**线性最小二乘问题**的三种直接法.

1 问题介绍

考虑**线性最小二乘问题**

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

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, 解为 $x=A^{-1}b$;
- 当 $m<n$ 时, 未知量个数大于约束个数, **欠定** (或亚定) 方程组
- 当 $m>n$ 时, 约束个数大于未知量个数, **超定**方程组

- 为了讨论方便, 本讲总是假定 A 是满秩的.
- 本讲我们主要讨论超定线性最小二乘问题的求解.

2 初等变换矩阵

把复杂的问题转化为等价的较简单的且易于求解问题

矩阵变换

- 完成这个转化的基本工具就是矩阵变换
- 除了三类初等变换外, 矩阵计算中常用的矩阵变换还有:

Householder 变换 和 **Givens 变换**

2.1 基本变换矩阵

我们称矩阵

$$E(u, v, \tau) = I - \tau uv^*$$

为**基本变换矩阵**，其中 $u, v \in \mathbb{C}^n$ 是非零向量，τ 是一个非零复数。

- ✍ 基本变换矩阵是单位矩阵的一个秩 1 扰动/修正。
- ✍ 三类初等矩阵都是基本变换矩阵。

定理 设 $E(u,v,\tau)$ 是一个基本变换矩阵, 我们有

(1) $\det(E(u,v,\tau)) = 1-\tau v^*u$;

(2) 若 $1-\tau v^*u \neq 0$, 则 $E(u,v,\tau)$ 非奇异, 且

$$(E(u,v,\tau))^{-1} = E(u,v,\gamma), \quad \text{其中 } \gamma = \frac{\tau}{\tau v^*u - 1}.$$

(板书)

2.2 Householder 变换

定义 我们称矩阵

$$H = I - \frac{2}{v^*v} vv^* = I - \frac{2}{\|v\|_2^2} vv^*, \quad 0 \neq v \in \mathbb{C}^n, \tag{3.2}$$

为 Householder **矩阵** (或 Householder **变换**, 或 Householder **反射**), 向量 v 称为 Householder **向量**. 我们通常将矩阵 (3.2) 记为 $H(v)$.

几何意义

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

对任意一个向量 $x \in \mathbb{C}^n$, 可将其写为

$$x = \alpha v + y, \qquad \left(\text{with } \alpha = \frac{v^* x}{v^* v}\right)$$

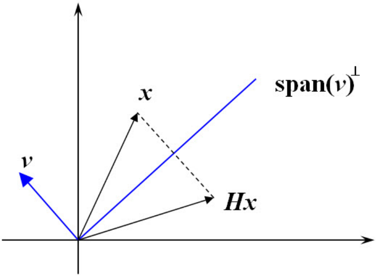
其中 $\alpha v \in \mathrm{span}\{v\}$, $y \in \mathrm{span}\{v\}^{\perp}$. 则

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

即 Hx 与 x 在 $\mathrm{span}\{v\}^{\perp}$ 方向有相同分量, 而在 v 方向的分量相差一个符号.

也就是说, Hx 是 x 关于超平面 $\mathrm{span}\{v\}^{\perp}$ 的镜面反射.

因此, Householder 矩阵有时也称为 **Householder 反射** 或 **反射矩阵** .



Householder 变换的几何意义

Householder 矩阵的几个基本性质

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 $n - 1$.

Householder 矩阵的重要应用

Householder 变换可以将一个向量除第一个元素以外的所有元素都化为零

先给出一个一般性结论

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 $x^*y \in \mathbb{R}$.

(取 $v = x - y$ 即可) (板书)

✍ 若 x, y 都是实向量，则 $x^*y \in \mathbb{R}$ 自然成立，此时充要条件为 $\boxed{\|x\|_2 = \|y\|_2}$

定理 设 $x = [x_1, x_2, \ldots, x_n]^{\mathsf{T}} \in \mathbb{R}^n$ 非零, 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，我们把定理中的向量 v 称为 x 对应的 **Householder 向量**

Householder 向量的计算

设 $x = [x_1, x_2, \ldots, x_n]^\mathsf{T} \in \mathbb{R}^n$, 由引理证明过程可知, 对应的 Householder 向量为

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

在实际计算中, 为尽可能地减少舍入误差, 应避免两个相近的数湘减, 因此取

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

✍ 事实上, 我们也可以取 $\alpha = \operatorname{sign}(x_1)\|x\|_2$, 但此时为了减少舍入误差, 我们需要通过下面的公式来计算 v 的第一个分量 v_1

$$v_1 = x_1 - \alpha = \frac{x_1^2 - \|x\|_2^2}{x_1 + \alpha} = \frac{-(x_2^2 + x_3^2 + \cdots + x_n^2)}{x_1 + \alpha}.$$

✍ 在前面两种计算方法中，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:

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

无论怎样选取 α, 我们都有 $H = I - \beta vv^*$ 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 2.1 计算 Householder 向量

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

✍ 总运算量大约为 $2n$ (乘法 + 加法), 且具有很好的数值稳定性。

✍ 在实际计算时, 我们可以将向量 v 单位化, 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, 因为变换后的向量 x 除第一个分量外, 其它都为零.

✍ 为了避免可能产生的溢出, 可事先将 x 单位化, 即令 $x = x/\|x\|_2$.

思考：这里要求 $v_1 \neq 0$, 那么什么情况下 $v_1 = 0$?

Householder 变换与矩阵的乘积

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

$$\boxed{HA = (I - \beta vv^*)A = A - \beta v(v^*A)}$$

因此, 在做 Householder 变换时, **不需要生成 Householder 矩阵, 只需要 Householder 向量即可**.

✍ 上面矩阵相乘的总运算量大约为 $4mn$.

2.3 Givens 变换

$$G(i,j,\theta)=\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & c & & s & & \\ & & & \ddots & & & \\ & & -s & & c & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \in \mathbb{R}^{n\times n} \qquad (\theta\in[0,2\pi],\ i\le j)$$

Givens 变换 (或 **Givens 旋转**, **Givens 矩阵**), 其中 $c=\cos(\theta)$, $s=\sin(\theta)$.

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

✍ 左乘 Givens 矩阵: 只会影响第 i 行和第 j 行的元素.

✍ 右乘 Givens 矩阵: 只会影响第 i 列和第 j 列的元素.

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 $Gx = \begin{bmatrix} r \\ 0 \end{bmatrix}$, 其中 c, s 和 r 的值如下:

- 若 $x_1 = x_2 = 0$, 则 $c = 1$, $s = 0$, $r = 0$;
- 若 $x_1 = 0$ 但 $x_2 \neq 0$, 则 $c = 0$, $s = x_2/|x_2|$, $r = |x_2|$;
- 若 $x_1 \neq 0$ 但 $x_2 = 0$, 则 $c = \mathrm{sign}(x_1)$, $s = 0$, $r = |x_1|$;
- 若 $x_1 \neq 0$ 且 $x_2 \neq 0$, 则 $c = x_1/r$, $s = x_2/r$, $r = \sqrt{x_1^2 + x_2^2}$.

✍ 通过 Givens 变换, 可以将 $x \in \mathbb{R}^n$ 的任意一个分量化为 0.

✍ 通过若干个 Givens 变换, 可以将 x 中除某个分量外的所有分量都化为 0.

算法 2.2 Givens 变换

% Given $x = [a, b]^{\mathsf{T}}$, compute c, s such that $Gx = [r, 0]^{\mathsf{T}}$ 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**   % 考虑计算稳定性
  $\tau = a/b, \quad s = \mathrm{sign}(b)/\sqrt{1+\tau^2}, \quad c = s\tau$
 else
  $\tau = b/a, \quad c = \mathrm{sign}(a)/\sqrt{1+\tau^2}, \quad s = c\tau$
 end if
end if

2.4 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 $\|E\|_2 = \mathcal{O}(\varepsilon_u)\cdot\|A\|_2$, $\|F\|_2 = \mathcal{O}(\varepsilon_u)\cdot\|A\|_2$.

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

* 本小结内容只需了解。

考虑对矩阵 A 做一系列的正交变换, 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.

矩阵乘积浮点计算误差分析

设 X 是一个非奇异的线性变换（矩阵），$\tilde{X}$ 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 $\|E\|_2 = O(\varepsilon_u) \cdot \|XA\|_2 \leq O(\varepsilon_u) \cdot \|X\|_2 \cdot \|A\|_2$, 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, 舍入误差将被放大 $\kappa_2(X)$ 倍.

若 X 是正交变换, 则 $\kappa_2(X) = 1$, 这就是为什么在浮点运算中尽量使用正交变换的原因.

3 QR 分解

3.1 QR 分解的存在唯一性

定理 (QR 分解) 设 $A \in \mathbb{C}^{m\times n}$ $(m \geq n)$. 则存在一个单位列正交矩阵 $Q \in \mathbb{C}^{m\times n}$ 和一个上三角矩阵 $R \in \mathbb{C}^{n\times n}$, 使得

$$A = QR$$

证明. (构造法) 设 $A = [a_1, a_2, \ldots, a_n] \in \mathbb{C}^{m\times n}$.

若 A 列满秩, 即 $\mathrm{rank}(A) = n$

则 QR 分解就是对 A 的列向量组进行 Gram-Schmidt 正交化过程。

算法 3.1 Gram-Schmidt Process

$r_{11} = \|a_1\|_2$

$q_1 = a_1 / r_{11}$

for $j = 2$ to n **do**

  $q_j = a_j$

  for $i = 1$ to $j - 1$ **do**

    $r_{ij} = q_i^* a_j$ % q_i^* 表示共轭转置

    $q_j = q_j - r_{ij} q_i$

  end for

  $r_{jj} = \|q_j\|_2$

  $q_j = q_j / r_{jj}$

end for

由 Gram-Schmidt 正交化过程可知

$$a_1 = r_{11} q_1, \quad a_j = r_{1j} q_1 + r_{2j} q_2 + \cdots + r_{jj} q_j = [q_1, q_2, \ldots, q_j] \begin{bmatrix} r_{1j} \\ r_{rj} \\ \vdots \\ r_{jj} \end{bmatrix}$$

记 $Q = [q_1, q_2, \ldots, q_n]$, $R = [r_{ij}]_{n \times n}$, 其中

$$r_{ij} = \begin{cases} q_i^* a_j, & \text{for } i \leq j \\ 0, & \text{for } i > j \end{cases}$$

则

$$[a_1, a_2, \ldots, a_n] = [q_1, q_2, \ldots, q_n] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}, \text{ 即 } A = QR$$

如果 A 不是列满秩

我们可以通过下面的方式做类似的正交化过程:

- 如果 $a_1 = 0$, 则令 $q_1 = 0$; 否则令 $q_1 = a_1/\|a_1\|_2$;
- 对于 $j = 2, 3, \ldots, n$, 计算 $\tilde{q}_j = a_j - \sum_{i=1}^{j-1}(q_i^* a_j)q_i$.

  如果 $\tilde{q}_j = 0$, 则表明 a_j 可以由 $a_1, a_2, \ldots, a_{j-1}$ 线性表出, 令 $q_j = 0$.

  否则令 $q_j = \tilde{q}_j/\|\tilde{q}_j\|_2$.

于是我们有

$$A = QR,$$

其中 $Q = [q_1, q_2, \ldots, q_n]$ 列正交且有为零的列向量。

✍ 注意: 如果 Q 的某一列 $q_k = 0$, 则 R 中对应的第 k 行就全部为 0.

设 $\text{rank}(A)=l<n$, 则 $\text{rank}(Q)=l$, 即 Q 有 l 个非零列, 设为 $q_{i_1},q_{i_2},\ldots,q_{i_l}$, 它们构成 $\mathbb{C}^m$ 中的一个单位正交向量组.

将其扩展成 $\mathbb{C}^m$ 中的一组标准正交基, 即

$$q_{i_1},q_{i_2},\ldots,q_{i_l},\quad \tilde{q}_1,\ldots,\tilde{q}_{m-l}.$$

然后我们用 $\tilde{q}_1$ 替换 Q 中的第一个零列, 用 $\tilde{q}_2$ 替换 Q 中的第二个零列.

依此类推, 将所有零列都替换, 最后得到的矩阵记为 $\tilde{Q}\in\mathbb{C}^{m\times n}$ (单位列正交).

由于 $\tilde{Q}$ 中新添加的列向量正好与 R 中的零行相对应, 所以

$$\boxed{\tilde{Q}R=QR=A} \Longrightarrow \boxed{A \text{ 的 QR 分解}}$$

$\square$

满秩矩阵 QR 分解的存在唯一性

定理 若 A 列满秩, 并要求 R 的对角线元素都为正, 则 A 的 QR 分解存在且唯一. (板书)

实矩阵的 QR 分解

若 A 是实矩阵, 则所有运算都是实运算, 因此 Q 和 R 都是实矩阵.

QR 分解的另一种形式

✍ 有时也将 QR 分解定义为: 存在酉矩阵 $Q \in \mathbb{C}^{m\times m}$ 使得

$$A = QR,$$

其中 $R = \begin{bmatrix} R_{11} \\ 0 \end{bmatrix} \in \mathbb{C}^{m\times n}$ 是上三角矩阵.

需注意：这里的 Q 是方阵，而 R 是长方阵。

秩亏矩阵的 QR 分解

若 A 不满秩, 则存在置换矩阵 P, 使得 AP 的前 l 列线性无关, 其中 $l=\mathrm{rank}(A)$.

对 AP 进行 QR 分解, 可得:

推论 设 $A\in\mathbb{C}^{m\times n}$ $(m\geq n)$, 则存在一个置换矩阵 P, 使得

$$AP = Q\begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix}_{n\times n},$$

其中 $Q\in\mathbb{C}^{m\times n}$ 单位列正交, $R_{11}\in\mathbb{C}^{l\times l}$ 是非奇异上三角矩阵.

上述结论可简化为

$$\boxed{AP = Q\begin{bmatrix} R_{11} & R_{12} \end{bmatrix}},$$

其中 $Q\in\mathbb{C}^{m\times l}$ 单位列正交, $R_{11}\in\mathbb{C}^{l\times l}$ 非奇异上三角.

满秩分解

推论 设 $A \in \mathbb{C}^{m\times n}$ 且 $\mathrm{rank}(A) = l \leq \min\{m, n\}$, 则存在满秩矩阵 $F \in \mathbb{C}^{m\times l}$ 和 $G \in \mathbb{C}^{l\times n}$, 使得

$$A = FG.$$

✍ 如果 A 是非奇异方阵, 则 QR 分解可用来求解方程组 $Ax = b$.

练习：基于 G-S 正交化的 QR 分解的运算量大约是多少?

3.2 基于 MGS 的 QR 分解

为了保证算法的数值稳定性，实际计算中不会直接采用 G-S 过程, 而是用 **修正的 G-S 过程** (modified Gram-Schmidt process, MGS)

算法 3.2 基于 MGS 的 QR 分解

% Given $A \in \mathbb{R}^{m\times n}$, compute $Q = [q_1, \dots, q_n]$ and R such that $A = QR$

Set $R = [r_{ik}] = 0_{n\times n}$ (the $n \times n$ zero matrix)

if $a_1 = 0$ **then**

  $q_1 = 0$

else

  $r_{11} = \|a_1\|_2$

  $q_1 = a_1/\|a_1\|_2$

end if

for $k = 2$ to n **do**

  $q_k = a_k$

  for $i = 1$ to $k-1$ **do**

    $r_{ik} = q_i^{\mathsf{T}} q_k$ % 注意与 G-S 的区别

    $q_k = q_k - r_{ik} q_i$

  end for

  if $q_k \neq 0$ **then**

    $r_{kk} = \|q_k\|_2$

    $q_k = q_k / r_{kk}$

  end if

end for

思考：基于 MGS 的 QR 分解，R 的对角线元素是非负的吗?

3.3 基于 Householder 变换的 QR 分解

Householder 变换：可以将任意非零变量除第一个元素外其它都化为零。

下面以 $m=n$ 为例，介绍基于 Householder 变换的 QR 分解

假定 $m=n$, 即 $A \in \mathbb{R}^{n\times n}$, 令 $H_1 \in \mathbb{R}^{n\times n}$ 为 Householder 变换, 满足

$$H_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} = \begin{bmatrix} r_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Longrightarrow H_1 A = \left[\begin{array}{c|ccc} r_1 & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ \hline 0 & & & \\ \vdots & & \tilde{A}_2 & \\ 0 & & & \end{array}\right], \quad \tilde{A}_2 \in \mathbb{R}^{(n-1)\times(n-1)}$$

同样, 构造 Householder 变换 $\tilde{H}_2 \in \mathbb{R}^{(n-1)\times(n-1)}$, 使得

$$\tilde{H}_2\tilde{A}_2 = \left[\begin{array}{c|ccc} r_2 & \tilde{a}_{23} & \cdots & \tilde{a}_{2n} \\ \hline 0 & & & \\ \vdots & & \tilde{A}_3 & \\ 0 & & & \end{array}\right], \quad \text{其中 } \tilde{A}_3 \in \mathbb{R}^{(n-2)\times(n-2)}$$

令 $H_2 = \begin{bmatrix} 1 & 0 \\ 0 & \tilde{H}_2 \end{bmatrix} \in \mathbb{R}^{n\times n}$, 则

$$H_2H_1A = \left[\begin{array}{cc|ccc} r_1 & \tilde{a}_{12} & \tilde{a}_{13} & \cdots & \tilde{a}_{1n} \\ 0 & r_2 & \tilde{a}_{23} & \cdots & \tilde{a}_{2n} \\ \hline 0 & 0 & & & \\ \vdots & \vdots & & \tilde{A}_3 & \\ 0 & 0 & & & \end{array}\right].$$

不断重复上述过程. 这样, 我们就得到一系列的矩阵

$$H_k = \begin{bmatrix} I_{k-1} & 0 \\ 0 & \tilde{H}_k \end{bmatrix}, \quad k = 2, 3, \ldots, n-1$$

使得

$$H_{n-1} \cdots H_2 H_1 A = \begin{bmatrix} r_1 & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ 0 & r_2 & \cdots & \tilde{a}_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & r_n \end{bmatrix} \triangleq R.$$

令 $Q = (H_{n-1} \cdots H_2 H_1)^{-1} = H_1 H_2 \cdots H_{n-1}$, 则

$$A = (H_{n-1} \cdots H_2 H_1)^{-1} R = QR$$

✍ 如果 $m > n$, 我们仍然可以通过上面的过程进行 QR 分解

✍ 如果不需要生成 Q, 则运算量大约为 $2mn^2 - 2/3n^3$

矩阵 Q 的计算

方法一 可在 QR 分解过程中直接计算，即

$$Q = I_n, \quad Q = QH_k, \quad k = 1, 2, \ldots, n-1.$$

方法二 **向后累积法** （需保留所有 Householder 向量）

$$Q = I_n, \quad Q = H_kQ, \quad k = n-1, n-2, \ldots, 1.$$

方法二的优点：一开始 Q 会比较稀疏, 随着迭代的进行, Q 才会慢慢变满.

运算量大约为 $4(m^2n - mn^2 + \frac{1}{3}n^3)$.

> 基于 MGS 和 Householder 变换的 QR 分解，所得的 Q 和 R 有何不同?

算法 3.3 基于 Householder 变换的 QR 分解

% Given $A \in \mathbb{R}^{m\times n}$, compute $Q \in \mathbb{R}^{m\times m}$, $R \in \mathbb{R}^{m\times n}$ such that $A = QR$

% The upper triangular part of R is stored in the upper triangular part of A

Set $Q = I_{m\times m}$

for $k = 1$ to n **do**

  $x = A(k:m, k)$

  $[\beta, v_k]$ = **house**(x)

  $v_k = v_k/\|v_k\|_2$

  $A(k:m, k:n) = (I_{m-k+1} - 2v_kv_k^\intercal)A(k:m, k:n)$

  $Q(1:k-1, k:m) = Q(1:k-1, k:m)(I_{m-k+1} - 2v_kv_k^\intercal)$

  $Q(k:m, k:m) = Q(k:m, k:m)(I_{m-k+1} - 2v_kv_k^\intercal)$

end for

✍ 注：这里只是算法的一个简单描述, 并没有考虑运算量问题.

3.4 列主元 QR 分解

当 A 不是满秩时, 我们可以进行**列主元 QR 分解**.

定理 (列主元 QR 分解) 设 $A \in \mathbb{C}^{m\times n}$ $(m \geq n)$, 且 $\mathrm{rank}(A) = l < n$. 则存在置换矩阵 P, 正交矩阵 $Q \in \mathbb{C}^{m\times m}$, 使得

$$AP = Q \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix}_{m\times n},$$

其中 $R_{11} \in \mathbb{C}^{l\times l}$ 是非奇异上三角矩阵, 且对角线元素满足

$$r_{11} \geq r_{22} \geq \cdots \geq r_{ll} > 0.$$

✍ 实现过程与 QR 分解类似, 只是在每一步需要选列主元, 同时做一个列交换.

列主元 QR 分解的实现

从第 $k-1$ 步到第 k 步 假设经过 $k-1$ 步后, 我们得到下面的分解

$$AP^{(k-1)} = Q^{(k-1)} \begin{bmatrix} R_{11}^{(k-1)} & R_{12}^{(k-1)} \\ 0 & R_{22}^{(k-1)} \end{bmatrix} \triangleq Q^{(k-1)} R^{(k-1)},$$

其中 $P^{(k-1)}$ 是置换矩阵, $Q^{(k-1)}$ 是正交矩阵, $R_{11}^{(k-1)} \in \mathbb{R}^{(k-1)\times(k-1)}$ 是非奇异上三角矩阵. 于是有

$$\left(Q^{(k-1)}\right)^{\intercal} AP^{(k-1)} = R^{(k-1)}.$$

下面考虑第 k 步

(1) 计算 $R_{22}^{(k-1)}$ 所有列的范数, 如果都为 0, 则 $R_{22}^{(k-1)}=0$, 算法结束.

(2) 当 $k\leq l$ 时, $R_{22}^{(k-1)}\neq 0$, 记范数最大的列为第 i_k 列.

若 $i_k\neq 1$, 则**交换 $R^{(k-1)}$ 的第 k 列与第 i_k+k-1 列**, 相应置换矩阵记为 P_k.

(3) 由于列交换不影响 $R^{(k-1)}$ 的前 $k-1$ 列, 因此列交换后的矩阵可记为

$$R^{(k-1)}P_k \triangleq \begin{bmatrix} R_{11}^{(k-1)} & \tilde{R}_{12}^{(k-1)} \\ 0 & \tilde{R}_{22}^{(k-1)} \end{bmatrix}.$$

构造 $\tilde{R}_{22}^{(k-1)}$ 的第 1 列所对应的 Householder 变换 $\tilde{H}_k$, 并令

$$H_k=\begin{bmatrix} I_{k-1} & 0 \\ 0 & \tilde{H}_k \end{bmatrix}, \quad P^{(k)}=P^{(k-1)}P_k.$$

于是

$$H_k\left(Q^{(k-1)}\right)^{\mathsf T}AP^{(k)}=H_kR^{(k-1)}P_k=\begin{bmatrix} R_{11}^{(k-1)} & \tilde{R}_{12}^{(k-1)} \\ 0 & \tilde{H}_k\tilde{R}_{22}^{(k-1)} \end{bmatrix}\triangleq R^{(k)},$$

其中 $\tilde{H}_k\tilde{R}_{22}^{(k-1)}$ 的第一列除第一个元素外, 其余都是零.

记 $Q^{(k)}\triangleq Q^{(k-1)}H_k^{\mathsf T}$, 则

$$AP^{(k)}=Q^{(k)}R^{(k)}\triangleq\begin{bmatrix} R_{11}^{(k)} & R_{12}^{(k)} \\ 0 & R_{22}^{(k)} \end{bmatrix},$$

其中 $R_{11}^{(k)}\in\mathbb{R}^{k\times k}$ 为非奇异上三角矩阵.

依此类推, 直到第 l 步, 我们就可以得到 A 的列主元 QR 分解.

✍ 由列主元的选取方法可知, R_{11} 的对角线元素非负且按降序排列.

3.5 基于 Givens 变换的 QR 分解

QR 分解也可以通过 Givens 变换来实现

设 $A \in \mathbb{R}^{n\times n}$, 构造 Givens 变换 G_{21}, 作用在 A 的最前面的两行上, 使得其第一列变为

$$G_{21}\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{bmatrix} = \begin{bmatrix} \tilde{a}_{11} \\ 0 \\ a_{31} \\ \vdots \\ a_{n1} \end{bmatrix} \qquad \Longrightarrow \text{将 } a_{21} \text{ 化为 } 0$$

构造 Givens 变换 G_{31}, 作用在 $G_{21}A$ 的第 1 行和第 3 行上, 将 a_{31} 化为零.

由于 G_{31} 只影响第 1 行和第 3 行, 所以第二行的零元素维持不变.

以此类推, 我们可以构造一系列的 Givens 变换 $G_{41}, G_{51}, \ldots, G_{n1}$, 使得

$$G_{n1} \cdots G_{21} A = \begin{bmatrix} * & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & * & \cdots & * \end{bmatrix}.$$

类似地，我们可以构造 Givens 变换 $G_{32}, G_{42}, \ldots, G_{n2}$, 将第二列的第 3 至第 n 个元素全化为零, 同时保持第一列不变.

以此类推，对其他列做类似的处理，最后将 A 转化成一个上三角矩阵 $R = G_{n,n-1} \cdots G_{21} A$，即

$$A = (G_{n,n-1} \cdots G_{21})^{\intercal} R \triangleq QR$$

✍ 与 Householder 变换一样, 在进行 Givens 变换时, 我们不需要显式地写出 Givens 矩阵.

✍ 对于稠密矩阵而言, 基于 Givens 变换的 QR 分解的运算量比 Householder 变换要多很多, 因此 该方法主要用于当矩阵的非零下三角元素相对较少时的情形, 比如上 Hessenberg 矩阵的 QR 分解.

✍ 如果 $A \in \mathbb{R}^{m\times n}$, $m > n$, 仍可以通过 Givens 变换进行 QR 分解.

算法 3.4 基于 Givens 变换的 QR 分解

% Given $A \in \mathbb{R}^{m\times n}$, compute $Q \in \mathbb{R}^{m\times m}$ and $R \in \mathbb{R}^{m\times n}$ such that $A = QR$

% The upper triangular part of R is stored in the upper triangular part of A

Set $Q = I_{m\times m}$

for $k = 1$ to n **do**

  for $i = k+1$ to m **do**

    $[c, s]$=**givens**(a_{kk}, a_{ik})

    $$\begin{bmatrix} A(k, k:n) \\ A(i, k:n) \end{bmatrix} = G \begin{bmatrix} A(k, k:n) \\ A(i, k:n) \end{bmatrix} \text{ where } G = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}$$

    $[Q(1:m,k), Q(1:m,i)] = [Q(1:m,k), Q(1:m,i)]G^{\mathsf{T}}$

  end for

end for

3.6 QR 分解的稳定性 *

基于 Householder 变换和 Givens 变换的 QR 分解都具有很好的数值稳定性, 基于 MGS 的 QR 分解也是向后稳定的.

✍ 如果需要计算 Q, 则 MGS 的运算量相对较少, 因此当 A 的列向量具有很好的线性无关性时, 可以使用 MGS 来计算 QR 分解. **但需要注意的是, MGS 得到的 Q 不是方阵 (除非 A 是方阵).**

* 本小结内容只需了解。

Björck [Björck 1967] 证明了通过 MGS 计算的矩阵 Q 满足

$$Q^\intercal Q = I + E_{MGS} \quad \text{其中} \quad \|E_{MGS}\|_2 \approx \varepsilon_u \kappa_2(A).$$

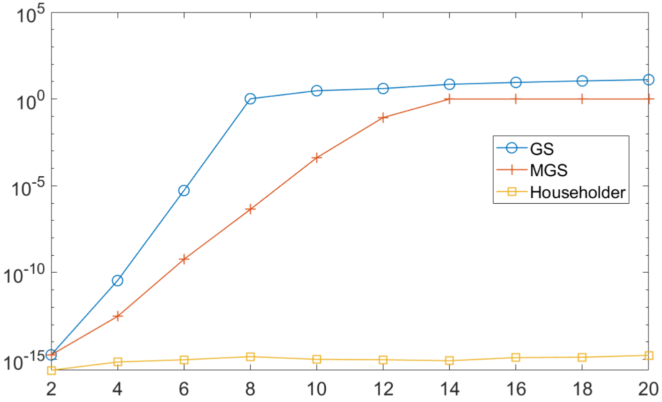
而 Householder 变换计算的矩阵 Q 满足

$$Q^\intercal Q = I + E_H \quad \text{其中} \quad \|E_H\|_2 \approx \varepsilon_u.$$

也就是说，Householder 变换得到的 Q 具有更好的正交性.

如果正交性至关重要, 建议使用 Householder 变换, 特别是当 A 的列向量接近线性相关时.

例 比较基于 GS, MGS 和 Householder 变换 QR 分解算法的稳定性, 即观察 $\|\tilde{Q}^\intercal\tilde{Q} - I\|_2$ 的值 (以 Hilbert 矩阵为例) QR_3methods.m



以 Hilbert 矩阵为例