统计计算

2016 年 3 月 8 日

目录

目录

定理

设 $\mathbf{A} \in \mathbb{C}^{n\times n}$，则 $\mathbf{A}$ 有唯一 LDR 分解的充要条件是 $\mathbf{A}$ 的顺序主子式 $\Delta_i \neq 0 (i = 1, \cdots, n-1)$，其中 $\mathbf{D} = \operatorname{diag}(d_1, \cdots, d_n)$，且

$$d_i = \Delta_i / \Delta_{i-1}, \quad i = 1, \cdots, n,$$

这里记 $\Delta_0 = 1$.

LDR 分解可以在 LR 分解的基础上进行.

1. 求出 $\mathbf{A}$ 的 LR 分解 $\mathbf{A} = \mathbf{LR}$.

   令 $\mathbf{D}$ 为 $\mathbf{R}$ 的主对角线上元素构成的对角阵,
   即 $\mathbf{D} = \mathrm{diag}(r_{11}, \cdots, r_{nn})$.
   此时有 $r_{ii} \neq 0 (i = 1, \cdots, n-1)$.

矩阵的 LDR 分解算法 (续)

2. 若 $\mathbf{A}$ 非奇异, 则 $r_{nn} \neq 0$, $\mathbf{D}$ 可逆.

   令 $\widetilde{\mathbf{R}} = \mathbf{D}^{-1}\mathbf{R}$, 则有 $\mathbf{A} = \mathbf{LR} = \mathbf{LD}\widetilde{\mathbf{R}}$. 由于 $\widetilde{\mathbf{R}}$ 实际上是将 $\mathbf{R}$ 的第 i 行除以 r_{ii} 后的结果, 因此, $\widetilde{\mathbf{R}}$ 是单位上三角矩阵.

3. 若 $\mathbf{A}$ 奇异, 则 $r_{nn}=0$, 此时有

$$\mathbf{D}=\mathrm{diag}(r_{11},\cdots,r_{n-1,n-1},0).$$

令 $\tilde{\mathbf{R}}$ 为将 $\mathbf{R}$ 的第 i 行除以 $r_{ii}(i=1,\cdots,n-1)$, 并将 r_{nn} 换成 1 所得的矩阵.

1. 方法 1:
   首先得到 LDR 分解, 再令 $\tilde{\mathbf{L}} = \mathbf{LD}$ 即得 $\mathbf{A}$ 的 Crout 分解 $\mathbf{A} = \mathbf{LDR} = \tilde{\mathbf{L}}\mathbf{R}$.

2. 方法 2:
   同 LR 分解算法的分析一样可得, 只是算法的计算顺序为: 第 1 列 $+$ 第 1 行 $\longrightarrow$ 第 2 列 $+$ 第 2 行 $\longrightarrow \cdots$

矩阵的 Crout 分解算法 (续)

Crout 分解算法

对 $i=1,\cdots,n$, 计算

$$\begin{cases} l_{ji} = a_{ji} - \sum_{k=1}^{i-1} l_{jk} r_{ki}, & j=i,\cdots,n, \\ r_{ij} = \dfrac{1}{l_{ii}}\left(a_{ij} - \sum_{k=1}^{i-1} l_{ik} r_{kj}\right), & j=i+1,\cdots,n. \end{cases}$$

矩阵的 Crout 分解算法 (续)

练习

要求自己写出节省存贮空间的算法.

问　题?

Cholesky 分解的定义

设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 为 Hermite 正定矩阵, 称

$$\mathbf{A} = \mathbf{T}\mathbf{T}^{\mathrm{H}}$$

为矩阵 $\mathbf{A}$ 的 Cholesky 分解, 其中 $\mathbf{T} \in \mathbb{C}^{n \times n}$ 为下三角矩阵.

Cholesky 分解也称为平方根分解 (注意也有其它形式的分解被称为平方根分解).

Cholesky 分解的存在唯一性

定理

若 $\mathbf{A} \in \mathbb{C}^{n\times n}$ 为 Hermite 正定矩阵, 则

1. $\mathbf{A}$ 存在 Cholesky 分解 $\mathbf{A} = \mathbf{T}\mathbf{T}^{\mathrm{H}}$.
2. 进一步, 若 $\mathbf{T}$ 的对角元素全取为正数, 则分解是唯一的.

证明　由 $\mathbf{A}$ 为 Hermite 正定矩阵知, 对任意 $i=1,\cdots,n$, $\mathbf{A}$ 的 i 阶顺序主子式 $\Delta_i>0$. 于是 $\mathbf{A}$ 有唯一的 LDR 分解 $\mathbf{A}=\mathbf{LDR}$, 其中, $\mathbf{D}=\mathrm{diag}(d_1,\cdots,d_n)$, 且 $d_i>0$ $(i=1,\cdots,n)$.

由 $\mathbf{A}$ 为 Hermite 矩阵知 $\mathbf{LDR}=\mathbf{R}^{\mathrm{H}}\mathbf{D}\mathbf{L}^{\mathrm{H}}$, 再由 LDR 分解的唯一性知 $\mathbf{L}=\mathbf{R}^{\mathrm{H}}$, 故 $\mathbf{A}=\mathbf{LDL}^{\mathrm{H}}$. 令

$$\mathbf{T}=\mathbf{L}\cdot\mathrm{diag}\left(\sqrt{d_1},\cdots,\sqrt{d_n}\right),$$

则 $\mathbf{T}$ 为下三角矩阵, $\mathbf{A}=\mathbf{T}\mathbf{T}^{\mathrm{H}}$ 为 $\mathbf{A}$ 的 Cholesky 分解. $\square$

方法 1:

设 $\mathbf{A}=(a_{ij})\in\mathbb{R}^{n\times n}$. 令

$$\mathbf{T}=\begin{bmatrix} t_{11} & & & \\ t_{21} & t_{22} & & \\ \vdots & \vdots & \ddots & \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{bmatrix}.$$

由 Cholesky 分解 $\mathbf{A}=\mathbf{T}\mathbf{T}^{\mathrm{T}}$ 两边对应元素相等知:
对任意 $i=1,\cdots,n$, 有

$$\begin{aligned} a_{ij} &= t_{i1}t_{j1}+t_{i2}t_{j2}+\cdots+t_{ij}t_{jj}, \quad j<i, \\ a_{ii} &= t_{i1}^2+t_{i2}^2+\cdots+t_{ii}^2. \end{aligned}$$

从而可得 Cholesky 分解的递推公式.

Cholesky 分解算法 1

对 $i = 1, \cdots, n$, 计算

$$
\begin{cases}
t_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} t_{ik}^2}, \\
t_{ji} = \dfrac{1}{t_{ii}} \left(a_{ij} - \sum_{k=1}^{i-1} t_{ik} t_{jk} \right), \quad j = i+1, \cdots, n.
\end{cases}
$$

练习

要求自己写出节省存贮空间的算法.

方法 2:

记 $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_n]$, 其中

$$\begin{aligned} \mathbf{t}_1 &= [t_{11}, \ldots, t_{n1}]^{\mathrm{T}}, \\ \mathbf{t}_i &= [\underbrace{0, \cdots, 0}_{i-1}, t_{ii}, \cdots, t_{ni}]^{\mathrm{T}}, \quad i = 2, \cdots, n, \end{aligned}$$

于是

$$\mathbf{A} = \mathbf{T}\mathbf{T}^{\mathrm{T}} = \sum_{i=1}^{n} \mathbf{t}_i \mathbf{t}_i^{\mathrm{T}}.$$

1. 记 $\mathbf{A}^{(0)} = \mathbf{A} = (a_{ij}^{(0)})$.

Cholesky 分解算法 (续)

2. 记 $\mathbf{A}^{(1)} = \mathbf{A}^{(0)} - \mathbf{t}_1\mathbf{t}_1^{\mathrm{T}} = (a_{ij}^{(1)})$, 则 $\mathbf{A}^{(1)} = \sum_{i=2}^{n} \mathbf{t}_i\mathbf{t}_i^{\mathrm{T}}$, 从而

$$
\begin{bmatrix}
a_{11}^{(0)} - t_{11}^2 & a_{12}^{(0)} - t_{11}t_{21} & \cdots & a_{1n}^{(0)} - t_{11}t_{n1} \\
a_{12}^{(0)} - t_{21}t_{11} & a_{22}^{(0)} - t_{21}^2 & \cdots & a_{2n}^{(0)} - t_{21}t_{n1} \\
\cdots & \cdots & \cdots & \cdots \\
a_{1n}^{(0)} - t_{n1}t_{11} & a_{2n}^{(0)} - t_{n2}t_{21} & \cdots & a_{nn}^{(0)} - t_{n1}^2
\end{bmatrix}
$$

$$
= \begin{bmatrix}
0 & 0 & \cdots & 0 \\
0 & * & \cdots & * \\
\vdots & \vdots & & \vdots \\
0 & * & \cdots & *
\end{bmatrix},
$$

于是可以计算 $\mathbf{t}_1$, 从而可知 $\mathbf{A}^{(1)}$.

3. 记 $\mathbf{A}^{(2)} = \mathbf{A}^{(1)} - \mathbf{t}_2\mathbf{t}_2^{\mathrm{T}} = (a_{ij}^{(2)})$, 用上面的方法类似可得 $\mathbf{t}_2$ 及 $\mathbf{A}^{(2)}$, 如此继续下去可得到 $\mathbf{T}$.

Cholesky 分解算法 (续)

Cholesky 分解算法 2

对 $k=1,\cdots,n$, 计算

$$\begin{cases} t_{jk} = \dfrac{a_{kj}^{(k-1)}}{\sqrt{a_{kk}^{(k-1)}}}, & j=k,k+1,\cdots,n, \\ a_{ij}^{(k)} = a_{ij}^{(k-1)} - \dfrac{a_{ik}^{(k-1)}a_{kj}^{(k-1)}}{a_{kk}^{(k-1)}}, & i,j=k+1,\cdots,n. \end{cases}$$

相应可以给出节省存贮空间的算法.

Cholesky 分解算法举例

教材第 232 页例 1.3.

作业

1. 习题
   P. 292: 5.4 — 5.7
2. 本节算法的详细推导与编码
3. 上机实习题
   P. 294: 1, 3, 4

矩阵三角分解的应用举例

教材第 233 — 234 页.

- 计算行列式

- 解线性方程组

  n 元线性方程组可写成矩阵形式

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

其中, $\mathbf{A} \in \mathbb{C}^{m\times n}$, $\mathbf{x} \in \mathbb{C}^{n}$, $\mathbf{b} \in \mathbb{C}^{m}$.

若 $\mathbf{A}$ 有 LR 分解 $\mathbf{A} = \mathbf{L}\mathbf{R}$, 则

$$\mathbf{A}\mathbf{x} = \mathbf{b} \Longleftrightarrow \mathbf{L}\mathbf{R}\mathbf{x} = \mathbf{b}$$

$$\Longleftrightarrow \begin{cases} \mathbf{L}\mathbf{y} = \mathbf{b} \\ \mathbf{R}\mathbf{x} = \mathbf{y}, \end{cases}$$

从而可以用回代法方便地求出 $\mathbf{x}$.

矩阵三角分解的应用举例 (续)

- 计算加权距离

  加权距离的意义.
  从加权范数的观点认识.
  权矩阵的 Cholesky 分解.

补充: 加权 LS 估计

对于线性模型 $(\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V})$, 如果我们知道 $\text{Var}\,\varepsilon = \sigma^2\mathbf{V}$, 其中 $\mathbf{V}$ 已知, 当然希望充分利用 ε 的统计信息以得到参数 β 的最优估计.

因为 $\mathbf{V} > \mathbf{O}$, 存在平分根 $\mathbf{V}^{\frac{1}{2}} > \mathbf{O}$. 作变换

$$\widetilde{\mathbf{y}} = \mathbf{V}^{-\frac{1}{2}}\mathbf{y}, \quad \widetilde{\mathbf{X}} = \mathbf{V}^{-\frac{1}{2}}\mathbf{X}, \quad \widetilde{\varepsilon} = \mathbf{V}^{-\frac{1}{2}}\varepsilon,$$

则模型 $(\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V})$ 转化为 $(\widetilde{\mathbf{y}}, \widetilde{\mathbf{X}}\beta, \sigma^2\mathbf{I})$.

补充: 加权 LS 估计 (续)

若 $\mathbf{X}$ 为满列秩的, 则 β 的加权 LS 估计唯一地表示为

$$\hat{\beta} = (\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{y}} = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{y},$$

且估计误差方差阵为

$$E(\beta - \hat{\beta})(\beta - \hat{\beta})^{\mathrm{T}} = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}.$$

按加权范数的观点, β 的加权 LS 估计是如下优化问题

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^r} \|\mathbf{y} - \mathbf{X}\beta\|_{\mathbf{V}^{-1}}^2$$

的解.

补充: 加权 LS 估计 (续)

对任意对称正定矩阵 $\mathbf{W} \in \mathbb{R}^{n\times n}$, 均可考虑如下形式的加权 LS 估计:

$$\hat{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{y}.$$

问题

权矩阵 $\mathbf{W}$ 取什么时, 估计 $\hat{\beta}$ 具有最优性 (在估计误差方差阵最小意义下)?

对任意对称正定矩阵 $\mathbf{W} \in \mathbb{R}^{n\times n}$, 均可考虑如下形式的加权 LS 估计:

$$\hat{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{y}.$$

问题

权矩阵 $\mathbf{W}$ 取什么时, 估计 $\hat{\beta}$ 具有最优性 (在估计误差方差阵最小意义下)?

补充: 加权 LS 估计 (续)

可以证明:

$\mathbf{W}=\mathbf{V}^{-1}$ 时, $\hat{\beta}$ 的估计误差方差阵

$$
\begin{aligned}
& E(\beta-\hat{\beta})(\beta-\hat{\beta})^{\mathrm{T}} \\
= & \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{V}\mathbf{W}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}
\end{aligned}
$$

达到最小.

此时, 加权 LS 估计也称为 Markov 估计.

事实上, 对正定矩阵 $\mathbf{V}$, 令

$$\mathbf{A} = \mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1/2}, \ \mathbf{B} = \mathbf{V}^{1/2}\mathbf{W}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1},$$

由矩阵型 Cauchy–Schwarz 不等式, 有

$$\begin{aligned}
&(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{V}\mathbf{W}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1} \\
&= \mathbf{B}^{\mathrm{T}}\mathbf{B} \\
&\geq (\mathbf{A}\mathbf{B})^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}(\mathbf{A}\mathbf{B}) \\
&= (\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1} \\
&= (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}.
\end{aligned}$$

问　题?

我们可以从另一角度来考察 LR 分解.

由 $\mathbf{A}$ 的 LR 分解知: 当 $\mathbf{A}$ 非奇异且满足分解条件时有

$$\mathbf{L}^{-1}\mathbf{A} = \mathbf{R}.$$

上式可以理解为对 $\mathbf{A}$ 作线性变换以后变成上三角阵 $\mathbf{R}$.

进一步的思考

问题

是否存在其它变换将 $\mathbf{A}$ 变成上三角阵?

问　题?