舒尔补

2019年12月10日 20:34

给定任意矩阵块M，如下所示：

$$M=\begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

如果，矩阵块D是可逆的，则 $A-BD^{-1}C$ 称之为 D 关于M的舒尔补.
如果，矩阵块A是可逆的，则 $D-CA^{-1}B$ 称之为 A 关于M的舒尔补

舒尔补的由来：
将M矩阵变成上三角形或下三角形的过程中，都会遇到舒尔补：

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}=\begin{bmatrix} A & B \\ -CA^{-1}A+C & -CA^{-1}B+D \end{bmatrix}=\begin{bmatrix} A & B \\ 0 & \Delta_A \end{bmatrix}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}=\begin{bmatrix} A & 0 \\ C & -CA^{-1}B+D \end{bmatrix}$$

联合起来：M变成对角形：

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}=\begin{bmatrix} A & B \\ 0 & -CA^{-1}B+D \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}$$

$$=\begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix}$$

反过来，我们又能从对角形恢复成矩阵M：

$$\begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}=\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

由此可得：矩阵M的逆：

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1}=\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}$$

因为：$\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}=I.$

$- \; 0 \quad I \;] \; [\; 0 \quad I \;]$

⇒ 使得一个大问题变成很多个小问题

Toy Example:

假设多元变量 x 服从高斯分布，且由两部分组成：$x=\begin{bmatrix}a\\ b\end{bmatrix}$，变量之间构成的协方差矩阵为：

$$K=\begin{bmatrix}A & C^T\\ C & D\end{bmatrix}$$

其中 $A=cov(a,a)$ $D=cov(b,b)$ $C=cov(a,b)$，由此变量 x 的概率分布为：　利用舒尔补

$$P(a,b)=\underbrace{P(a)P(b|a)}_{\text{Bayes 法则可得}}\propto \exp\left(-\frac{1}{2}\begin{bmatrix}a\\ b\end{bmatrix}^T\begin{bmatrix}A & C^T\\ C & D\end{bmatrix}^{-1}\begin{bmatrix}a\\ b\end{bmatrix}\right) \rightarrow \text{由 Gaussian 形式可得}$$

$$\Rightarrow P(a,b)\propto \exp\left(-\frac{1}{2}\begin{bmatrix}a\\ b\end{bmatrix}^T\begin{bmatrix}I & -A^{-1}C^T\\ 0 & I\end{bmatrix}\begin{bmatrix}A^{-1} & 0\\ 0 & \Delta_A^{-1}\end{bmatrix}\begin{bmatrix}I & 0\\ -CA^{-1} & I\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix}\right)$$

$$\propto \exp\left(-\frac{1}{2}\begin{bmatrix}a & b-A^{-1}C^Tb\end{bmatrix}\begin{bmatrix}A^{-1} & 0\\ 0 & \Delta_A^{-1}\end{bmatrix}\begin{bmatrix}a\\ b-CA^{-1}a\end{bmatrix}\right)$$

↓ 分别提出 a 相关的和 b 相关的

$$\propto \exp\left(-\frac{1}{2}(a^TA^{-1}a)+(b-A^{-1}C^Tb)^T\Delta_A^{-1}(b-A^{-1}C^Tb)\right)$$

$$\propto \underbrace{\exp\left(-\frac{1}{2}a^TA^{-1}a\right)}_{P(a)}\underbrace{\exp\left(-\frac{1}{2}(b-A^{-1}C^Tb)^T\Delta_A^{-1}(b-A^{-1}C^Tb)\right)}_{P(b|a)}$$

⇒ 这意味着，我们能从多元高斯分布 $P(a,b)$ 中分别得到边际概率 $P(a)$ 和条件概率 $P(b|a)$

$P(a)$:

$$P(a)=\int_b P(a,b)$$

$$P(a)\propto \exp\left(-\frac{1}{2}a^TA^{-1}a\right)\sim N(0,A)$$

⇒ 边际概率的协方差就是从联合分布中取对应的矩阵块就行了.

$P(b|a)$:

$$P(b|a)\propto \exp\left(-\frac{1}{2}(b-A^{-1}C^Tb)^T\Delta_A^{-1}(b-A^{-1}C^Tb)\right)$$

$$\sim N(A^{-1}C^Tb,\ \Delta_A).$$

协方差为 a 对应的舒尔补，均值也变了.

为什么要讨论 $P(a)$, $P(b|a)$ 的信息矩阵 ?

因为基于优化的 SLAM 问题中，我们往往直接操作的是信息矩阵，而不是协方差矩阵.

保持着稀疏性.

假设我们已知信息矩阵：

$$J^T\Sigma^{-1}J\,\delta x=b$$

假设我们已知信息矩阵:

$$\begin{bmatrix} A & C^T \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$$

保证求出的:

$$J^T \Sigma^{-1} J \, \delta x = b$$

或: $\underline{J^T \sqrt{\Sigma^{-1}}} \; \underline{\sqrt{\Sigma^{-1}} \, \delta x} \; \delta x = b$

由公式易得:

$$\begin{bmatrix} A & C^T \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1} C^T \Delta_A^{-1} C A^{-1} & -A^{-1} C^T \Delta_A^{-1} \\ -\Delta_A^{-1} C A^{-1} & \Delta_A^{-1} \end{bmatrix} \triangleq \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix} \quad ☆$$

由条件概率 $P(b|a)$ 的协方差 Δ_A 以及公式☆，可得其信息矩阵:

$$\Delta_A^{-1} = \Lambda_{bb}$$

由边缘概率 $P(a)$ 的协方差为 A，以及公式☆，可得其信息矩阵:

$$A^{-1} = \Lambda_{aa} - \Lambda_{ab} \Lambda_{bb}^{-1} \Lambda_{ba}.$$

回顾 toy example 中去除变量 x_3 的操作:

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{w_1}{\sigma_1^2} & 0 \\ -\frac{w_1}{\sigma_1^2} & \frac{w_1^2}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{w_3^2}{\sigma_3^2} & -\frac{w_3}{\sigma_3^2} \\ 0 & -\frac{w_3}{\sigma_3^2} & \frac{1}{\sigma_3^2} \end{bmatrix} \qquad \begin{matrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{matrix}$$

从联合概率分布 $P(x_1, x_2, x_3)$ 中 marg 掉变量 x_3，即 $P(x_1, x_2)$ 对应的信息矩阵利用公式☆得:

$$\begin{aligned} K_2^{-1} &= \Lambda_{aa} - \Lambda_{ab} \Lambda_{bb}^{-1} \Lambda_{ba} \\ &= \Lambda_{aa} - \begin{bmatrix} 0 \\ -\frac{w_3}{\sigma_3^2} \end{bmatrix} \sigma_3^2 \begin{bmatrix} 0 & -\frac{w_3}{\sigma_3^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{w_1}{\sigma_1^2} \\ -\frac{w_1}{\sigma_1^2} & \frac{w_1^2}{\sigma_1^2} + \frac{1}{\sigma_2^2} \end{bmatrix} \end{aligned}$$

总结:

边缘概率对于协方差矩阵的操作是容易的，但不好操作信息矩阵。

条件概率对于信息矩阵容易操作，但不好操作协方差矩阵。

$$P(a, b) = \mathcal{N}\left(\begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} \right) = \mathcal{N}^{-1}\left(\begin{bmatrix} \eta_a \\ \eta_b \end{bmatrix}, \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix} \right)$$

	边际概率. $P(a) = \int P(a,b)\,d(b)$	条件概率: $P(a\|b) = P(a,b)/P(b)$
协方差矩阵:	$\mu = \mu_a$ $\Sigma = \Sigma_{aa}$	$\mu' = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(b - \mu_a)$ $\Sigma' = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$.
信息矩阵:	$y = y_a - \Lambda_{ab}\Lambda_{bb}^{-1}y_b$ $\Lambda = \Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}$.	$y' = y_a - \Lambda_{ab}b$ $\Lambda' = \Lambda_{aa}$