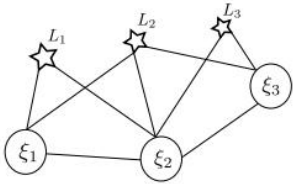
深蓝学院
shenlanxueyuan.com

第四章作业提示

主讲人 会打篮球的猫

第一题

❶ 某时刻，SLAM 系统中相机和路标点的观测关系如下图所示，其中 ξ 表示相机姿态，L 表示观测到的路标点。当路标点 L 表示在世界坐标系下时，第 k 个路标被第 i 时刻的相机观测到，重投影误差为 $\mathbf{r}(\xi_i, L_k)$。另外，相邻相机之间存在运动约束，如 IMU 或者轮速计等约束。



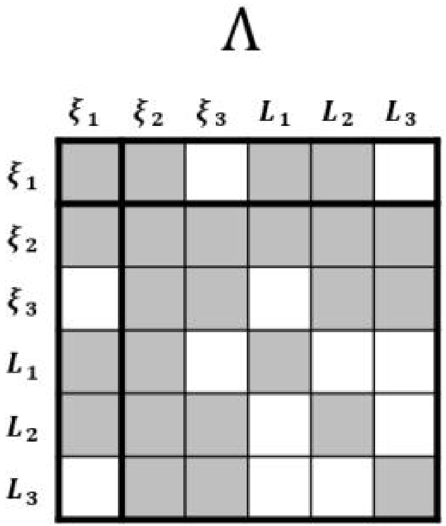
1 请绘制上述系统的信息矩阵 $\mathbf{\Lambda}$.
2 请绘制相机 ξ_1 被 marg 以后的信息矩阵 $\mathbf{\Lambda}'$.

$$\mathbf{J}_2 = \frac{\partial \mathbf{r}_{13}}{\partial \xi} = \begin{bmatrix} \frac{\partial \mathbf{r}_{13}}{\partial \xi_1} & \mathbf{0} & \frac{\partial \mathbf{r}_{13}}{\partial \xi_3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\mathbf{\Lambda}_2 = \mathbf{J}_2^\top \mathbf{\Sigma}_2^{-1} \mathbf{J} = \begin{bmatrix} (\frac{\partial \mathbf{r}_{13}}{\partial \xi_1})^\top \mathbf{\Sigma}_2^{-1} \frac{\partial \mathbf{r}_{13}}{\partial \xi_1} & \mathbf{0} & (\frac{\partial \mathbf{r}_{13}}{\partial \xi_1})^\top \mathbf{\Sigma}_2^{-1} \frac{\partial \mathbf{r}_{13}}{\partial \xi_3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ (\frac{\partial \mathbf{r}_{13}}{\partial \xi_3})^\top \mathbf{\Sigma}_2^{-1} \frac{\partial \mathbf{r}_{13}}{\partial \xi_1} & \mathbf{0} & (\frac{\partial \mathbf{r}_{13}}{\partial \xi_3})^\top \mathbf{\Sigma}_2^{-1} \frac{\partial \mathbf{r}_{13}}{\partial \xi_3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{46}$$

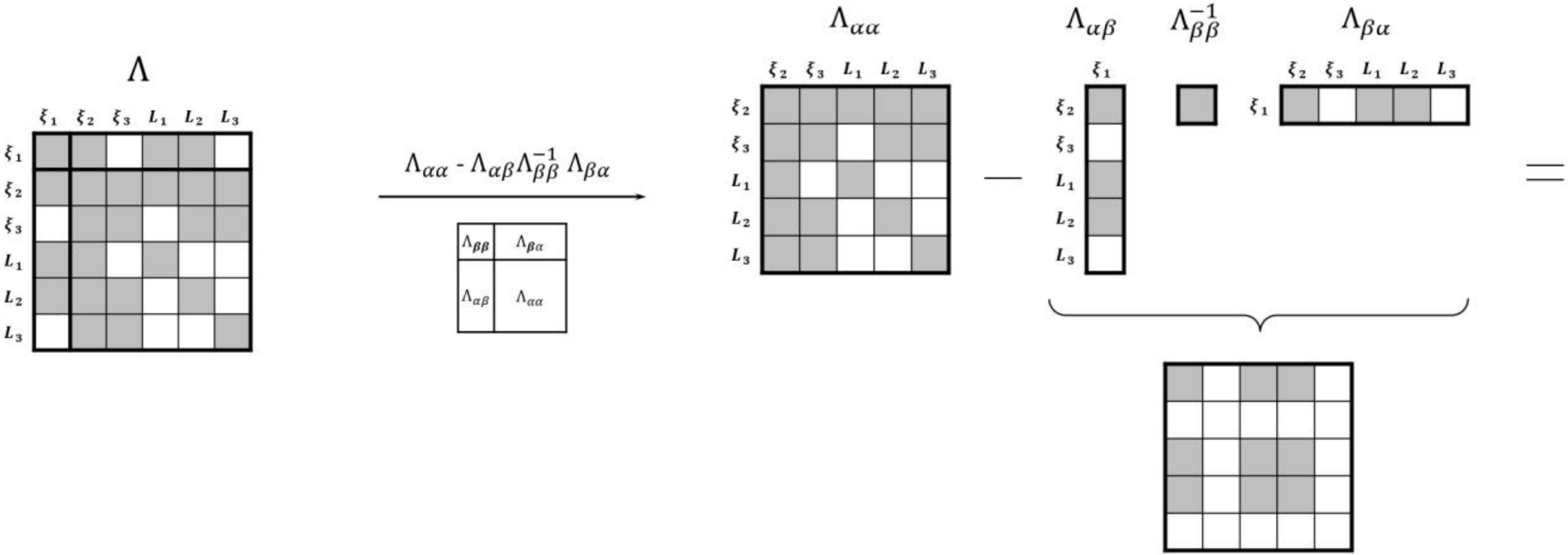
其实就是将系统所有约束对应的小信息矩阵累加的一个过程，只有该约束相连接的变量所对应位置才有值，其余为 0。

第一题



将ξ_1进行边缘化，根据原来的系统，与ξ_1相连的有ξ_1、L_1和L_2，边缘化操作后，这三者之间便会两两相连。

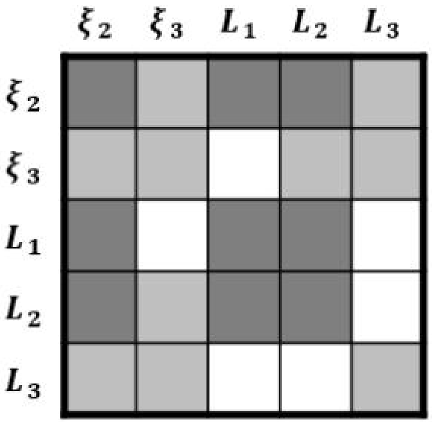
第一题



Λ
ξ_1 ξ_2 ξ_3 L_1 L_2 L_3
$\Lambda_{\alpha\alpha}$ - $\Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\beta\alpha}$
$\Lambda_{\beta\beta}$ $\Lambda_{\beta\alpha}$
$\Lambda_{\alpha\beta}$ $\Lambda_{\alpha\alpha}$
$\Lambda_{\alpha\alpha}$
$\Lambda_{\alpha\beta}$
$\Lambda_{\beta\beta}^{-1}$
$\Lambda_{\beta\alpha}$

第一题

Λ'



可以发现，和$\Lambda_{\alpha\alpha}$相比，增加了L_1和L_2之间的关联，信息矩阵更稠密了。

第二题

❷ 阅读《Relationship between the Hessian and Covariance Matrix for Gaussian Random Variables》. 证明信息矩阵和协方差的逆之间的关系。

- Hessian和协方差逆的关系

一个多元高斯随机变量的概率密度函数的负对数函数的Hessian矩阵，和其协方差矩阵的逆是相等的。

$$\mathbf{C}\delta\mathbf{x} = \delta\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_z)$$

$$p(\mathbf{C}\delta\mathbf{x}) = \alpha \exp\left(\delta\mathbf{x}^T\mathbf{C}^T\boldsymbol{\Sigma}_z^{-1}\mathbf{C}\delta\mathbf{x}\right) = \alpha \exp\left(\delta\mathbf{x}^T\boldsymbol{\Sigma}_x^{-1}\delta\mathbf{x}\right)$$

Consider a Gaussian random vector $\boldsymbol{\theta}$ with mean $\boldsymbol{\theta}^\star$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ so its joint probability density function (PDF) is given by:

$$p(\boldsymbol{\theta}) = (2\pi)^{-\frac{N_\theta}{2}}|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)\right] \tag{A.1}$$

The objective function can be defined as its negative logarithm:

$$J(\boldsymbol{\theta}) \equiv -\ln p(\boldsymbol{\theta}) = \frac{N_\theta}{2}\ln 2\pi + \frac{1}{2}\ln|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}^\star) \tag{A.2}$$

which is a quadratic function of the components in $\boldsymbol{\theta}$. By taking partial differentiations with respect to θ_l and $\theta_{l'}$, the (l, l') component of the Hessian matrix can be obtained:

$$\mathcal{H}^{(l,l')}(\boldsymbol{\theta}^\star) = \left.\frac{\partial^2 J(\boldsymbol{\theta})}{\partial\theta_l\partial\theta_{l'}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} = (\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1})^{(l,l')} \tag{A.3}$$

so the Hessian matrix is equal to the inverse of the covariance matrix:

$$\mathcal{H}(\boldsymbol{\theta}^\star) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \tag{A.4}$$

For Gaussian random variables, the second derivatives of the objective function are constant for all $\boldsymbol{\theta}$ because the objective function is a quadratic function of $\boldsymbol{\theta}$. Therefore, the Hessian matrix can be computed without obtaining the mean vector $\boldsymbol{\theta}^\star$.

第二题

深蓝学院 shenlanxueyuan.com

② 阅读《Relationship between the Hessian and Covariance Matrix for Gaussian Random Variables》. 证明信息矩阵和协方差的逆之间的关系。

- Hessian和信息矩阵的关系

We can then see it as an information. The covariance of score function above is the definition of Fisher Information. As we assume θ is a vector, the Fisher Information is in a matrix form, called Fisher Information Matrix:

$$\mathrm{F} = \mathop{\mathbb{E}}_{p(x|\theta)}\left[\nabla \log p(x|\theta)\, \nabla \log p(x|\theta)^{\mathrm{T}}\right].$$

However, usually our likelihood function is complicated and computing the expectation is intractable. We can approximate the expectation in $\mathbf{F}$ using empirical distribution $\hat{q}(x)$, which is given by our training data $X = \{x_1, x_2, \cdots, x_N\}$. In this form, $\mathbf{F}$ is called Empirical Fisher:

$$\mathrm{F} = \frac{1}{N}\sum_{i=1}^{N} \nabla \log p(x_i|\theta)\, \nabla \log p(x_i|\theta)^{\mathrm{T}}.$$

Claim: The negative expected Hessian of log likelihood is equal to the Fisher Information Matrix $\mathbf{F}$.

Proof. The Hessian of the log likelihood is given by the Jacobian of its gradient:

$$\begin{aligned}
\mathrm{H}_{\log p(x|\theta)} &= \mathrm{J}\left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right) \\
&= \frac{\mathrm{H}_{p(x|\theta)}\, p(x|\theta) - \nabla p(x|\theta)\, \nabla p(x|\theta)^{\mathrm{T}}}{p(x|\theta)\, p(x|\theta)} \\
&= \frac{\mathrm{H}_{p(x|\theta)}\, p(x|\theta)}{p(x|\theta)\, p(x|\theta)} - \frac{\nabla p(x|\theta)\, \nabla p(x|\theta)^{\mathrm{T}}}{p(x|\theta)\, p(x|\theta)} \\
&= \frac{\mathrm{H}_{p(x|\theta)}}{p(x|\theta)} - \left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)\left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)^{\mathrm{T}},
\end{aligned}$$

第二题

❷ 阅读《Relationship between the Hessian and Covariance Matrix for Gaussian Random Variables》. 证明信息矩阵和协方差的逆之间的关系。

- Hessian和信息矩阵的关系

where the second line is a result of applying quotient rule of derivative. Taking expectation wrt. our model, we have:

$$
\begin{aligned}
\underset{p(x|\theta)}{\mathbb{E}}\left[\mathrm{H}_{\log p(x|\theta)}\right] &= \underset{p(x|\theta)}{\mathbb{E}}\left[\frac{\mathrm{H}_{p(x|\theta)}}{p(x|\theta)} - \left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)\left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)^{\mathrm{T}}\right] \\
&= \underset{p(x|\theta)}{\mathbb{E}}\left[\frac{\mathrm{H}_{p(x|\theta)}}{p(x|\theta)}\right] - \underset{p(x|\theta)}{\mathbb{E}}\left[\left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)\left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)^{\mathrm{T}}\right] \\
&= \int \frac{\mathrm{H}_{p(x|\theta)}}{p(x|\theta)} p(x|\theta)\,\mathrm{d}x - \underset{p(x|\theta)}{\mathbb{E}}\left[\nabla \log p(x|\theta)\,\nabla \log p(x|\theta)^{\mathrm{T}}\right] \\
&= \mathrm{H}_{\int p(x|\theta)\,\mathrm{d}x} - \mathrm{F} \\
&= \mathrm{H}_1 - \mathrm{F} \\
&= -\mathrm{F}.
\end{aligned}
$$

Thus we have $\mathrm{F} = -\mathbb{E}_{p(x|\theta)}\left[\mathrm{H}_{\log p(x|\theta)}\right]$. □

Indeed knowing this result, we can see the role of F as a measure of curvature of the log likelihood function.

似然分布的概率密度函数的负对数函数的Hessian矩阵的期望等于信息矩阵。

第二题

❷ 阅读《Relationship between the Hessian and Covariance Matrix for Gaussian Random Variables》. 证明信息矩阵和协方差的逆之间的关系。

- Hessian和信息矩阵的关系

综上所述，负对数最大似然估计的 Hessian 矩阵就是信息矩阵，同时也等于协方差的逆。所以平时往往就直接将协方差矩阵的逆当做信息矩阵。

第三题

③ 请补充作业代码中单目 Bundle Adjustment 信息矩阵的计算，并输出正确的结果。正确的结果为：奇异值最后 7 维接近于 0，表明零空间的维度为 7.

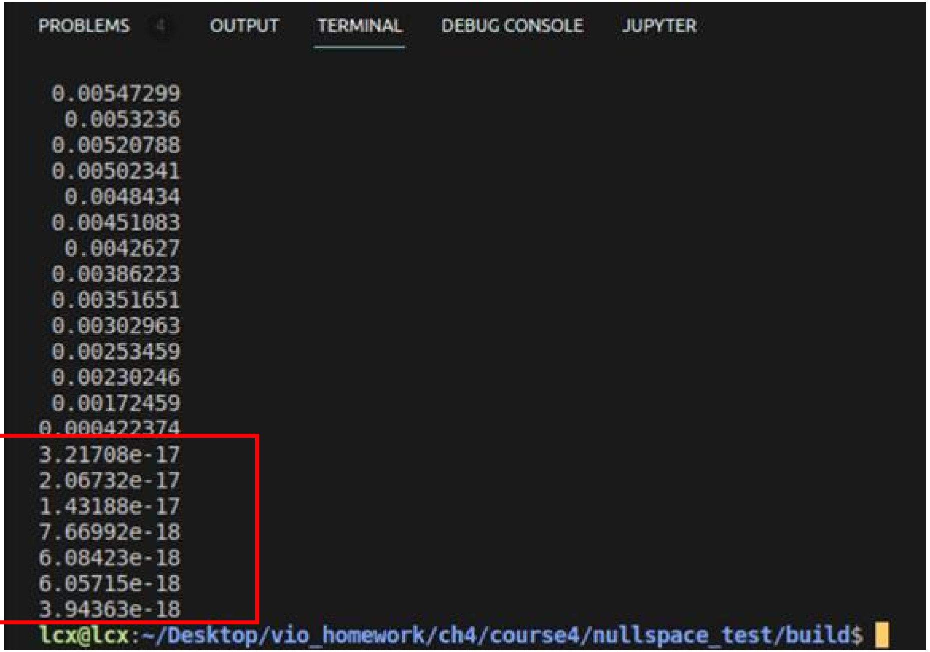
```
H.block(i*6,i*6,6,6) += jacobian_Ti.transpose() * jacobian_Ti;
/// 请补充完整作业信息矩阵块的计算
H.block(poseNums*6+j*3, i*6, 3, 6) += jacobian_Pj.transpose() * jacobian_Ti;
H.block(i*6, poseNums*6+j*3, 6, 3) += jacobian_Ti.transpose() * jacobian_Pj;
H.block(poseNums*6+j*3, poseNums*6+j*3, 3, 3) += jacobian_Pj.transpose() * jacobian_Pj;
```

$$\mathbf{J}_2 = \frac{\partial \mathbf{r}_{13}}{\partial \boldsymbol{\xi}} = \begin{bmatrix} \frac{\partial \mathbf{r}_{13}}{\partial \boldsymbol{\xi}_1} & \mathbf{0} & \frac{\partial \mathbf{r}_{13}}{\partial \boldsymbol{\xi}_3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\boldsymbol{\Lambda}_2 = \mathbf{J}_2^\top \boldsymbol{\Sigma}_2^{-1} \mathbf{J} = \begin{bmatrix} \left(\frac{\partial \mathbf{r}_{13}}{\partial \boldsymbol{\xi}_1}\right)^\top \boldsymbol{\Sigma}_2^{-1} \frac{\partial \mathbf{r}_{13}}{\partial \boldsymbol{\xi}_1} & \mathbf{0} & \left(\frac{\partial \mathbf{r}_{13}}{\partial \boldsymbol{\xi}_1}\right)^\top \boldsymbol{\Sigma}_2^{-1} \frac{\partial \mathbf{r}_{13}}{\partial \boldsymbol{\xi}_3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \left(\frac{\partial \mathbf{r}_{13}}{\partial \boldsymbol{\xi}_3}\right)^\top \boldsymbol{\Sigma}_2^{-1} \frac{\partial \mathbf{r}_{13}}{\partial \boldsymbol{\xi}_1} & \mathbf{0} & \left(\frac{\partial \mathbf{r}_{13}}{\partial \boldsymbol{\xi}_3}\right)^\top \boldsymbol{\Sigma}_2^{-1} \frac{\partial \mathbf{r}_{13}}{\partial \boldsymbol{\xi}_3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{46}$$

第三题

❸ 请补充作业代码中单目 Bundle Adjustment 信息矩阵的计算，并输出正确的结果。正确的结果为：奇异值最后 7 维接近于 0，表明零空间的维度为 7.



在单目视觉SLAM中，信息矩阵存在一个7维零空间，对应7自由度的不客观量。

深蓝学院
shenlanxueyuan.com

感谢各位聆听！

Thanks for Listening