

滑动窗口是滤波和优化方法的折衷

第 4 讲 基于滑动窗口算法的 VIO 系统：可观性和一致性

贺一家，高翔，崔华坤

2019 年 10 月 31 日

- ① 从高斯分布到信息矩阵
SLAM 和高斯分布
从两个简单例子窥探协方差和信息矩阵
- ② 舒尔补应用：边际概率，条件概率
舒尔补
舒尔补与条件概率和边际概率
总结
- ③ 滑动窗口算法
图优化基础
基于边际概率的滑动窗口算法
- ④ 滑动窗口中的 FEJ 算法
滑动窗口算法的系统回顾及深入剖析
系统可观性
滑动窗口算法中存在的问题以及 FEJ 算法
- ⑤ 作业

Section 1

从高斯分布到信息矩阵

SLAM 问题概率建模

考虑某个状态 ξ ，以及一次与该变量相关的观测 \mathbf{r}_i 。由于噪声的存在，观测服从概率分布 $p(\mathbf{r}_i|\xi)$ 。

多次观测时，各个测量值相互独立，则多个测量 $\mathbf{r} = (\mathbf{r}_1, \dots, \mathbf{r}_n)^\top$ 构建的似然概率为：

$$p(\mathbf{r}|\xi) = \prod_i p(\mathbf{r}_i|\xi) \quad (1)$$

如果知道机器人状态的先验信息 $p(\xi)$ ，如 GPS，车轮码盘信息等，则根据 Bayes 法则，有后验概率：

$$p(\xi|\mathbf{r}) = \frac{p(\mathbf{r}|\xi)p(\xi)}{p(\mathbf{r})} \quad (2)$$

通过最大后验估计，获得系统状态的最优估计：

$$\xi_{\text{MAP}} = \arg \max_{\xi} p(\xi|\mathbf{r}) \quad (3)$$

SLAM 问题求解

后验公式 (2) 中分母跟状态量无关，舍弃。最大后验变成了：

$$\xi_{\text{MAP}} = \arg \max_{\xi} \prod_i p(\mathbf{r}_i | \xi) p(\xi) \quad (4)$$

即

$$\xi_{\text{MAP}} = \arg \min_{\xi} \left[- \sum_i \log p(\mathbf{r}_i | \xi) - \log p(\xi) \right] \quad (5)$$

如果假设观测值服从多元高斯分布：

$$p(\mathbf{r}_i | \xi) = \mathcal{N}(\mu_i, \Sigma_i), p(\xi) = \mathcal{N}(\mu_{\xi}, \Sigma_{\xi}) \quad (6)$$

则有：

$$\xi_{\text{MAP}} = \underset{\xi}{\operatorname{argmin}} \sum_i \|\mathbf{r}_i - \mu_i\|_{\Sigma_i}^2 + \|\xi - \mu_{\xi}\|_{\Sigma_{\xi}}^2 \quad (7)$$

这个最小二乘的求解可以使用上节课的解法：

$$\mathbf{J}^{\top} \Sigma^{-1} \mathbf{J} \delta \xi = -\mathbf{J}^{\top} \Sigma^{-1} \mathbf{r}$$

可以人为设定，如果特别相信一个观测值，可以把 Σ_i 设大一点。也会有一些算法去动态调节 Σ_i 。可以去看动态估计和自适应估计的相关论文。

高斯分布和协方差矩阵

信息矩阵可以看作是每一项的权重

多元高斯分布

零均值的多元高斯分布有如下概率形式：

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left(-\frac{1}{2}\mathbf{x}^\top \Sigma^{-1}\mathbf{x}\right)$$

其中 Σ 是协方差矩阵，协方差矩阵的逆记作 $\Lambda = \Sigma^{-1}$ 。

比如变量 \mathbf{X} 为三维的变量时，协方差矩阵为：

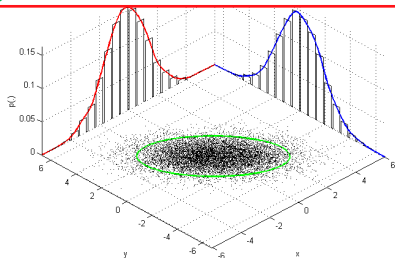
$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix} \quad (9)$$

其中 $\Sigma_{ij} = E(x_i x_j)$ 为对应元素求期望。

协方差矩阵的几何意义：

- 1、对角线上反映的是x和y的分布集中程度。
- 2、非对角线方向反映的是x和y的相关性。

通过贝叶斯公式变成一个最小二乘问题，最小二乘问题的解法就是：对每一个残差（观测预测-观测）求它的雅可比，拼成一个大的H矩阵，再去不断迭代。



当维数变大时，求解计算量巨大。如何优雅的从多元高斯分布中丢弃变量？

$$p(\mathbf{x}, \mathbf{y}) \rightarrow p(\mathbf{x}) \text{ or } p(\mathbf{y})$$

toy example 1

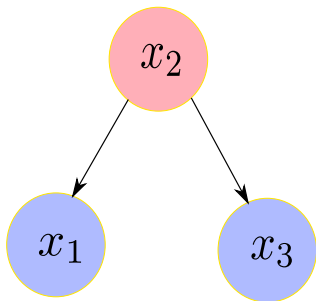
样例

设 x_2 为室外的温度, x_1, x_3 分别为房间 1 和房间 3 的室内温度:

$$\begin{aligned}x_2 &= v_2 \\x_1 &= w_1 x_2 + v_1 \\x_3 &= w_3 x_2 + v_3\end{aligned}\tag{10}$$

其中, v_i 相互独立, 且各自服从零均值, 协方差为 σ_i^2 的高斯分布。

这两个例子来源于文档¹



¹David Mackay. "The humble Gaussian distribution". In: (2006).

样例对应的状态量协方差矩阵

从上述关系，根据协方差公式的计算方式，我们可以写出 x 的协方差矩阵，先从对角元素开始计算：

$$\begin{aligned}
 \Sigma_{11} &= E(x_1 x_1) = E((w_1 v_2 + v_1)(w_1 v_2 + v_1)) \\
 &= w_1^2 E(v_2^2) + 2w_1 E(v_1 v_2) + E(v_1^2) \\
 &= w_1^2 \sigma_2^2 + \sigma_1^2
 \end{aligned} \tag{11}$$

同理有 $\Sigma_{22} = \sigma_2^2$, $\Sigma_{33} = w_3^2 \sigma_2^2 + \sigma_3^2$. 对于协方差矩阵的非对角元素：

$$\begin{aligned}
 \Sigma_{12} &= E(x_1 x_2) = E((w_1 v_2 + v_1)v_2) = w_1 \sigma_2^2 \\
 \Sigma_{13} &= E((w_1 v_2 + v_1)(w_3 v_2 + v_3)) = w_1 w_3 \sigma_2^2
 \end{aligned} \tag{12}$$

以此类推，可以得到整个协方差矩阵：

$$\Sigma = \begin{bmatrix} w_1^2 \sigma_2^2 + \sigma_1^2 & w_1 \sigma_2^2 & w_1 w_3 \sigma_2^2 \\ w_1 \sigma_2^2 & \sigma_2^2 & w_3 \sigma_2^2 \\ w_1 w_3 \sigma_2^2 & w_3 \sigma_2^2 & w_3^2 \sigma_2^2 + \sigma_3^2 \end{bmatrix} \tag{13}$$

对应的协方差矩阵的逆呢？

通过计算联合高斯分布从而得到协方差矩阵的逆：

$$\begin{aligned}
 p(x_1, x_2, x_3) &= p(x_2)p(x_1|x_2)p(x_3|x_2) \\
 &= \frac{1}{Z_2} \exp\left(-\frac{x_2^2}{2\sigma_2^2}\right) \frac{1}{Z_1} \exp\left(-\frac{(x_1 - w_1 x_2)^2}{2\sigma_1^2}\right) \frac{1}{Z_3} \exp\left(-\frac{(x_3 - w_3 x_2)^2}{2\sigma_3^2}\right)
 \end{aligned} \tag{14}$$

利用指数性质求出联合概率分布：

$$\begin{aligned}
 p(x_1, x_2, x_3) &= \frac{1}{Z} \exp\left(-\frac{x_2^2}{2\sigma_2^2} - \frac{(x_1 - w_1 x_2)^2}{2\sigma_1^2} - \frac{(x_3 - w_3 x_2)^2}{2\sigma_3^2}\right) \\
 &= \frac{1}{Z} \exp\left(-x_2^2 \left[\frac{1}{2\sigma_2^2} + \frac{w_1^2}{2\sigma_1^2} - \frac{w_3^2}{2\sigma_3^2}\right] - x_1^2 \frac{1}{2\sigma_1^2} + 2x_1 x_2 \frac{w_1}{2\sigma_1^2} - x_3^2 \frac{1}{2\sigma_3^2} + 2x_3 x_2 \frac{w_3}{2\sigma_3^2}\right) \\
 &= \frac{1}{Z} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{w_1}{\sigma_2^2} & 0 \\ -\frac{w_1}{\sigma_1^2} & \frac{w_1^2}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{w_3^2}{\sigma_3^2} & -\frac{w_3}{\sigma_3^2} \\ 0 & -\frac{w_3}{\sigma_3^2} & \frac{1}{\sigma_3^2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right)
 \end{aligned} \tag{15}$$

样例 1 对应的信息矩阵

由此得到协方差矩阵的逆，即信息矩阵：

$$\Lambda = \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{w_1}{\sigma_1^2} & 0 \\ -\frac{w_1}{\sigma_1^2} & \frac{w_1^2}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{w_3^2}{\sigma_3^2} & -\frac{w_3}{\sigma_3^2} \\ 0 & -\frac{w_3}{\sigma_3^2} & \frac{1}{\sigma_3^2} \end{bmatrix} \quad (16)$$

注意：信息矩阵中有两个元素为 0，它有什么具体含义呢？协方差逆矩阵中如果坐标为 (i, j) 的元素为 0，表示元素 i 和 j 关于其他变量条件独立，上面的例子中意味着变量 x_1 和 x_3 关于 x_2 条件独立。

协方差元素 vs 信息矩阵元素

假设室内温度和室外温度正相关 ($w_i > 0$),

- 协方差中非对角元素 $\Sigma_{ij} > 0$ 表示两变量是正相关。
- 信息矩阵中非对角元素为负数，甚至为 0。 $\Lambda_{12} < 0$ 表示在变量 x_3 发生的条件下，元素 x_1 和 x_2 正相关。

toy example 2

样例

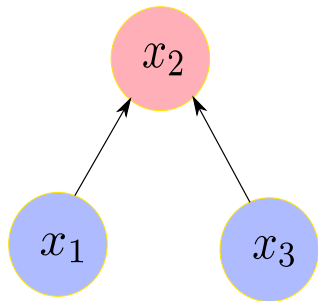
比如特征三角化，两个相机 pose 得到特征三维坐标：

$$x_2 = w_1 x_1 + w_3 x_3 + v_2 \quad (17)$$

同理，根据协方差矩阵的定义，可以得到协方差矩阵：

$$\Sigma = \begin{bmatrix} \sigma_1^2 & w_1 \sigma_1^2 & 0 \\ w_1 \sigma_1^2 & \sigma_2^2 + w_1^2 \sigma_1^2 + w_3^2 \sigma_3^2 & w_3 \sigma_3^2 \\ 0 & w_3 \sigma_3^2 & \sigma_3^2 \end{bmatrix}$$

协方差矩阵中非对角元素为 0 表示变量之间没有相关性。这是否意味着信息矩阵中也会为 0 呢？



样例 2 协方差矩阵的逆

按照例子 1 中的方式，求取协方差矩阵的逆：

$$\begin{aligned}
 p(x_1, x_2, x_3) &= p(x_1)p(x_3)p(x_2|x_1, x_3) \\
 &= \frac{1}{Z_1} \exp\left(-\frac{x_1^2}{2\sigma_1^2}\right) \frac{1}{Z_3} \exp\left(-\frac{x_3^2}{2\sigma_3^2}\right) \frac{1}{Z_2} \exp\left(-\frac{(x_2 - w_1x_1 - w_3x_3)^2}{2\sigma_2^2}\right)
 \end{aligned} \tag{18}$$

将变量整成向量形式：

$$\begin{aligned}
 p(x_1, x_2, x_3) &= \frac{1}{Z} \exp\left(-\frac{x_1^2}{2\sigma_1^2} - \frac{x_3^2}{2\sigma_3^2} - \frac{(x_2 - w_1x_1 - w_3x_3)^2}{2\sigma_2^2}\right) \\
 &= \frac{1}{Z} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} + \frac{w_1^2}{\sigma_2^2} & -\frac{w_1}{\sigma_2^2} & \frac{w_1w_3}{\sigma_2^2} \\ -\frac{w_1}{\sigma_2^2} & \frac{1}{\sigma_2^2} & -\frac{w_3}{\sigma_2^2} \\ \frac{w_1w_3}{\sigma_2^2} & -\frac{w_3}{\sigma_2^2} & \frac{1}{\sigma_3^2} + \frac{w_3^2}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right)
 \end{aligned} \tag{19}$$

样例 2 中的有趣现象

从上面推导出的信息矩阵来看,

- 虽然 x_1 和 x_3 不相关, 但是不说明他们的信息矩阵对应元素 Λ_{13} 为 0。
- 恰恰信息矩阵中 $\Lambda_{13} > 0$, 表示的是在变量 x_2 发生的条件下, 变量 x_1, x_3 成负相关。
- 对应上面的例子即 x_2 为常数, 如果 x_1 大, 则 x_3 小。

样例的意义: 非常有助于我们了解 SLAM 的信息矩阵 H 的组成。

疑问: 如果我们移除变量, 信息矩阵或协方差矩阵如何变化呢?

高翔: 这些东西在概率图模型和概率图理论里面有详细的介绍

toy example 1 中去除变量 x_3

协方差如何变化？

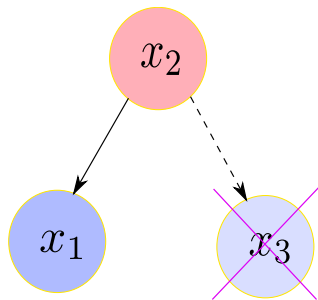
利用协方差的计算公式可知， x_1, x_2 计算协方差时跟 x_3 ，并无关系，所以

$$\Sigma = \begin{bmatrix} w_1^2 \sigma_2^2 + \sigma_1^2 & w_1 \sigma_2^2 & w_1 w_3 \sigma_2^2 \\ w_1 \sigma_2^2 & \sigma_2^2 & w_3 \sigma_2^2 \\ w_1 w_3 \sigma_2^2 & w_3 \sigma_2^2 & w_3^2 \sigma_2^2 + \sigma_3^2 \end{bmatrix} \quad (20)$$

就能得到去除 x_3 后的协方差矩阵：

$$\Sigma_2 = \begin{bmatrix} w_1^2 \sigma_2^2 + \sigma_1^2 & w_1 \sigma_2^2 \\ w_1 \sigma_2^2 & \sigma_2^2 \end{bmatrix} \quad (21)$$

这两个例子来源于文档²



²David Mackay. "The humble Gaussian distribution". In: (2006).

toy example 1 中去除变量 x_3

信息矩阵如何变化？

同样，我们只需要把信息矩阵公式 (16) 中 x_3 对应的部分 (蓝色) 去掉就可以：

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{w_1}{\sigma_1^2} & \theta \\ -\frac{w_1}{\sigma_1^2} & \frac{w_1^2}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{w_3^2}{\sigma_3^2} & -\frac{w_3}{\sigma_3^2} \\ \theta & -\frac{w_3}{\sigma_3^2} & \frac{1}{\sigma_3^2} \end{bmatrix} \quad (22)$$

从而得到：

$$\Sigma_2^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{w_1}{\sigma_1^2} \\ -\frac{w_1}{\sigma_1^2} & \frac{w_1^2}{\sigma_1^2} + \frac{1}{\sigma_2^2} \end{bmatrix} \quad (23)$$

是不是非常简单？但是问题在于：实际操作过程中并不会会有这种颜色标记。

这时，需要引入 marginalization (边缘化) 和 Schur's complement (舒尔补) 来解决这个问题。

本节课的主题：

当最小二乘的状态越来越多，维度越来越大的时候，如何把旧的变量丢掉，合理加入新的变量。

程序以信息矩阵形式出现时，要移除哪些项，怎么操作

Section 2

舒尔补应用：边际概率, 条件概率

舒尔补的概念

舒尔补定义

给定任意的矩阵块 M ，如下所示：

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (24)$$

- 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。

更多定义参见 [wiki](#)³.

³Wiki. Schur Complement. https://en.wikipedia.org/wiki/Schur_complement

舒尔补的来由

将 M 矩阵变成上三角或者下三角形过程中，都会遇到舒尔补：

$$\begin{aligned}
 \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{CA}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} &= \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \Delta_{\mathbf{A}} \end{bmatrix} \\
 \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} &= \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \Delta_{\mathbf{A}} \end{bmatrix}
 \end{aligned} \tag{25}$$

其中： $\Delta_{\mathbf{A}} = \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}$ 。联合起来，将 M 变形成对角形：

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{CA}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \Delta_{\mathbf{A}} \end{bmatrix} \tag{26}$$

反过来，我们又能从对角形恢复成矩阵 M ：

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{CA}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \Delta_{\mathbf{A}} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \tag{27}$$

使用舒尔补分解的好处

快速求解矩阵 M 的逆

矩阵 M 可写成：

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \quad (28)$$

由此可得矩阵 M 的逆：

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \quad (29)$$

因为：

$$\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} = I \quad (30)$$

这个快速分解有什么好处呢？

舒尔补应用于多元高斯分布

toy example

假设多元变量 \mathbf{x} 服从零均值高斯分布，且由两部分组成： $\mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}$ ，
变量之间构成的协方差矩阵为：

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

其中 $A = \text{cov}(a, a)$, $D = \text{cov}(b, b)$, $C = \text{cov}(a, b)$. 由此变量 \mathbf{x} 的概率分布为：

$$P(a, b) = P(a)P(b|a) \propto \exp \left(-\frac{1}{2} \begin{bmatrix} a \\ b \end{bmatrix}^\top \begin{bmatrix} A & C^\top \\ C & D \end{bmatrix}^{-1} \begin{bmatrix} a \\ b \end{bmatrix} \right) \quad (31)$$

利用舒尔补一节公式(29)，对高斯分布进行分解，得⁴

$$\begin{aligned}
 & P(a, b) \\
 & \propto \exp \left(-\frac{1}{2} \begin{bmatrix} a \\ b \end{bmatrix}^\top \begin{bmatrix} A & C^\top \\ C & D \end{bmatrix}^{-1} \begin{bmatrix} a \\ b \end{bmatrix} \right) \\
 & \propto \exp \left(-\frac{1}{2} \begin{bmatrix} a \\ b \end{bmatrix}^\top \begin{bmatrix} I & -A^{-1}C^\top \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \right) \\
 & \propto \exp \left(-\frac{1}{2} \begin{bmatrix} a^\top & (b - CA^{-1}a)^\top \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix} \begin{bmatrix} a \\ b - CA^{-1}a \end{bmatrix} \right) \\
 & \propto \exp \left(-\frac{1}{2} (a^\top A^{-1}a) + (b - CA^{-1}a)^\top \Delta_A^{-1} (b - CA^{-1}a) \right) \\
 & \propto \underbrace{\exp \left(-\frac{1}{2} a^\top A^{-1}a \right)}_{p(a)} \underbrace{\exp \left(-\frac{1}{2} (b - CA^{-1}a)^\top \Delta_A^{-1} (b - CA^{-1}a) \right)}_{p(b|a)}
 \end{aligned} \tag{32}$$

这意味着我们能从多元高斯分布 $P(a, b)$ 中分解得到**边际概率** $p(a)$ 和**条件概率** $p(b|a)$ 。

⁴Huang. Conditional and marginal distributions of a multivariate Gaussian. <https://gbhqed.wordpress.com/2010/02/21/conditional-and-marginal-distributions-of-a-multivariate-gaussian/>

关于 $P(a)$, $P(b|a)$ 的协方差矩阵

$P(a)$ 的启示

对于协方差：
边缘分布直接取
条件分布舒尔补

$$\begin{aligned}
 P(a) &= \int_b P(a, b) \\
 P(a) &\propto \exp\left(-\frac{1}{2}a^\top A^{-1}a\right) \sim \mathcal{N}(0, A)
 \end{aligned} \tag{33}$$

启示：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。

$P(b|a)$ 的启示

$$P(b|a) \propto \exp\left(-\frac{1}{2}\left(b - CA^{-1}a\right)^\top \Delta_A^{-1}\left(b - CA^{-1}a\right)\right) \tag{34}$$

启示： $P(b|a) \sim \mathcal{N}(CA^{-1}a, \Delta_A)$ 。协方差变为 a 对应的舒尔补，均值也变了。

疑问三连：信息矩阵也是这样的吗？如果不是， $P(a)$ 的信息矩阵如何算？ $P(b|a)$ 的呢？

关于 $P(a)$, $P(b|a)$ 的信息矩阵

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。所以，有必要知道边际概率，条件概率的信息矩阵是何形式。

也就是说，当我们已知联合分布的信息矩阵形式时，我们怎么知道边缘概率和条件概率的信息矩阵形式。
求得边缘分布的过程叫做边缘化，如：

$$p(a, b, c, d) \quad p(a, b, c)$$

$P(a)$, $P(b|a)$ 的信息矩阵

假设我们已知信息矩阵：

$$\begin{bmatrix} A & C^T \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix} \quad (35)$$

另外，由公式 (29) 可知，协方差矩阵各块和信息矩阵之间有：

$$\begin{bmatrix} A & C^T \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}C^T\Delta_A^{-1}CA^{-1} & -A^{-1}C^T\Delta_A^{-1} \\ -\Delta_A^{-1}CA^{-1} & \Delta_A^{-1} \end{bmatrix} \triangleq \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix} \quad (36)$$

小结：以上讨论的问题有两个：

- 1、如何从联合分布的协方差矩阵得到边缘分布的协方差矩阵和条件分布的协方差矩阵
- 2、如何从联合分布的信息矩阵得到边缘分布的信息矩阵和条件分布的信息矩阵

$P(a)$, $P(b|a)$ 的信息矩阵

由条件概率 $P(b|a)$ 的协方差为 Δ_A 以及公式(36)，易得其信息矩阵为：

$$\Delta_A^{-1} = \Lambda_{bb} \quad (37)$$

由边际概率 $P(a)$ 的协方差为 A 以及公式(36)，易得其信息矩阵为：

$$A^{-1} = \Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba} \quad (38)$$

回顾样例

回顾 toy example 1 中去除变量 x_3 的操作

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

从联合分布 $P(x_1, x_2, x_3)$ 中 marg 掉变量 x_3 ，即 $P(x_1, x_2)$ 对应的信息矩阵可以用公式(38)得到。


$$\begin{aligned} \mathbf{K}_2^{-1} &= \Lambda_{aa} - \Lambda_{ab} \Lambda_{bb}^{-1} \Lambda_{ba} \\ &= \Lambda_{aa} - \begin{bmatrix} 0 \\ -\frac{w_3}{\sigma_3^2} \end{bmatrix} \sigma_3^2 \begin{bmatrix} 0 & -\frac{w_3}{\sigma_3^2} \end{bmatrix} \\ &= \Lambda_{aa} - \begin{bmatrix} 0 & 0 \\ 0 & \frac{w_3^2}{\sigma_3^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{w_1}{\sigma_1^2} \\ -\frac{w_1}{\sigma_1^2} & \frac{w_1^2}{\sigma_1^2} + \frac{1}{\sigma_2^2} \end{bmatrix} \end{aligned} \quad (39)$$

总结

边际概率对于协方差矩阵的操作是很容易的，但不好操作信息矩阵。条件概率恰好相反，对于信息矩阵容易操作，不好操作协方差矩阵。对于非零均值的高斯分布，表格总结如下⁵，可课后自己证明：

$$P(\mathbf{a}, \mathbf{b}) = \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix} \right) = \mathcal{N}^{-1} \left(\begin{bmatrix} \boldsymbol{\eta}_a \\ \boldsymbol{\eta}_b \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{bmatrix} \right) \quad (40)$$

	边际概率 $p(\mathbf{a}) = \int p(\mathbf{a}, \mathbf{b}) d\mathbf{b}$	条件概率 $p(\mathbf{a} \mathbf{b}) = p(\mathbf{a}, \mathbf{b})/p(\mathbf{b})$
协方差矩阵	$\boldsymbol{\mu} = \boldsymbol{\mu}_a$ $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{aa}$	$\boldsymbol{\mu}' = \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{b} - \boldsymbol{\mu}_b)$ $\boldsymbol{\Sigma}' = \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}$
信息矩阵	$\boldsymbol{\eta} = \boldsymbol{\eta}_a - \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\eta}_b$ $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_{aa} - \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_a - \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\eta}_b$ $\boldsymbol{\Lambda}' = \boldsymbol{\Lambda}_{aa}$

⁵Matthew R Walter, Ryan M Eustice, and John J Leonard. "Exactly sparse extended information filters for feature-based SLAM". In: *The International Journal of Robotics Research* 26.4 (2007), pp. 335+359. 

小结：SLAM里面通常用到的是：
求边缘概率的信息矩阵

Section 3

滑动窗口算法

最小二乘的图表示 顶点是优化变量，边是残差

toy example 3

有如下最小二乘系统，对应的图模型如有图所示：

$$\xi = \operatorname{argmin}_{\xi} \frac{1}{2} \sum_i \|\mathbf{r}_i\|_{\Sigma_i}^2 \quad (41)$$

其中

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \dots \\ \xi_6 \end{bmatrix}, \mathbf{r} = \begin{bmatrix} \mathbf{r}_{12} \\ \mathbf{r}_{13} \\ \mathbf{r}_{14} \\ \mathbf{r}_{15} \\ \mathbf{r}_{56} \end{bmatrix} \quad (42)$$



- 圆圈：表示顶点，需要优化估计的变量.
- 边：表示顶点之间构建的残差。有一元边（只连一个顶点），二元边，多元边...

最小二乘问题信息矩阵的构成

上述最小二乘问题，对应的高斯牛顿求解为：

$$\underbrace{\mathbf{J}^\top \Sigma^{-1} \mathbf{J}}_{\mathbf{H} \text{ or } \Lambda} \delta \xi = \underbrace{-\mathbf{J}^\top \Sigma^{-1} \mathbf{r}}_{\mathbf{b}} \quad (43)$$

注意，这里的 $\Lambda = \Sigma_{new}^{-1} \neq \Sigma^{-1}$ 反应的是方程解的方差，而 Σ^{-1} 是残差的方差，两者含义不同。 另外，公式中的雅克比矩阵为：

$$\mathbf{J} = \frac{\partial \mathbf{r}}{\partial \xi} = \begin{bmatrix} \frac{\partial \mathbf{r}_{12}}{\partial \xi} \\ \frac{\partial \mathbf{r}_{13}}{\partial \xi} \\ \frac{\partial \mathbf{r}_{14}}{\partial \xi} \\ \frac{\partial \mathbf{r}_{15}}{\partial \xi} \\ \frac{\partial \mathbf{r}_{56}}{\partial \xi} \end{bmatrix} = \begin{bmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \\ \mathbf{J}_3 \\ \mathbf{J}_4 \\ \mathbf{J}_5 \end{bmatrix}, \quad \mathbf{J}^\top = \begin{bmatrix} \mathbf{J}_1^\top & \mathbf{J}_2^\top & \mathbf{J}_3^\top & \mathbf{J}_4^\top & \mathbf{J}_5^\top \end{bmatrix} \quad (44)$$

再次白话三要素：

残差是什么？残差是观测云测减观测

方差是什么？方差是残差的方差

雅克比是什么？雅克比是残差相对状态变量的雅克比！

矩阵乘法公式(43)可以写成连加：

$$\sum_{i=1}^5 \mathbf{J}_i^\top \Sigma_i^{-1} \mathbf{J}_i \delta \xi = - \sum_{i=1}^5 \mathbf{J}_i^\top \Sigma_i^{-1} \mathbf{r}_i \quad (45)$$

信息矩阵的稀疏性

雅克比矩阵，信息矩阵的稀疏性

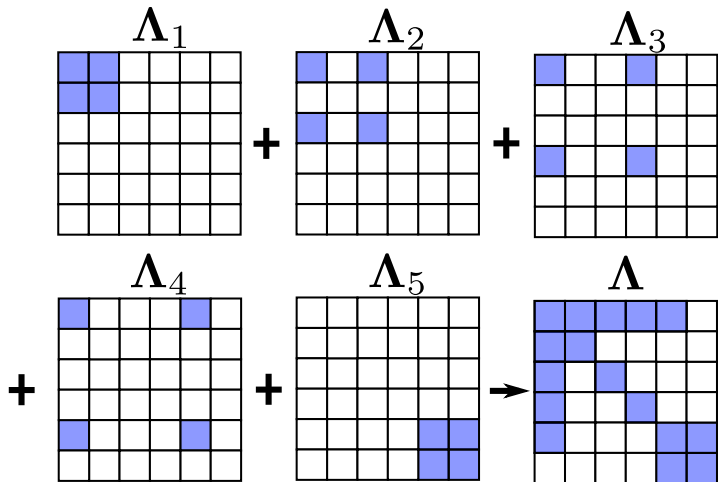
由于每个残差只和某几个状态量有关，因此，雅克比矩阵求导时，无关项的雅克比为 0。比如

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

同理，可以计算 $\mathbf{\Lambda}_1, \mathbf{\Lambda}_3, \mathbf{\Lambda}_4, \mathbf{\Lambda}_5$ ，并且也是稀疏的。

信息矩阵组装过程的可视化

将五个残差的信息矩阵加起来，得到样例最终的信息矩阵 Λ ，可视化如下：
注意这里的信息矩阵是 $J^T \Sigma^{-1} J$ 叠加起来的



基于边际概率的滑动窗口算法

为什么 SLAM 需要滑动窗口算法？

- 随着 VSLAM 系统不断往新环境探索，就会有新的相机姿态以及看到新的环境特征，最小二乘残差就会越来越多，信息矩阵越来越大，计算量将不断增加。
- 为了保持优化变量的个数在一定范围内，需要使用滑动窗口算法动态增加或移除优化变量。

滑动窗口算法大致流程

- ① 增加新的变量进入最小二乘系统优化
- ② 如果变量数目达到了一定的维度，则移除老的变量。
- ③ SLAM 系统不断循环前面两步

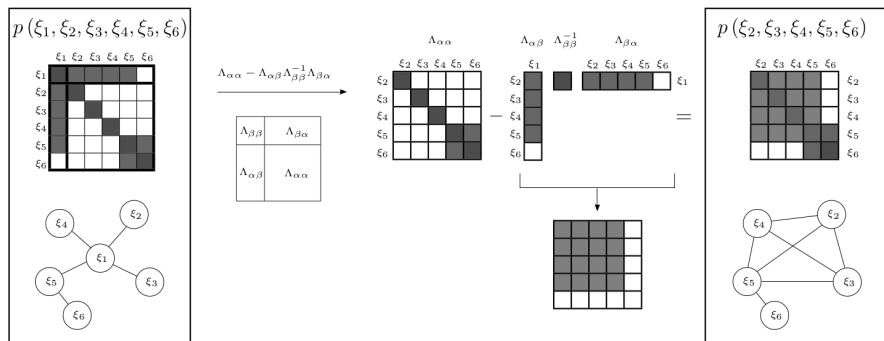
疑问：怎么移除老的变量？直接丢弃这些变量吗？

滑动窗口就是解决如何移除老变量以及加入新变量的问

利用边际概率移除老的变量

直接丢弃变量和对应的测量值，会损失信息。正确的做法是使用边际概率，将丢弃变量所携带的信息传递给剩余变量。

下图为 toy example 3 中使用边际概率移除变量 ξ_1 ，信息矩阵的变化过程⁶：

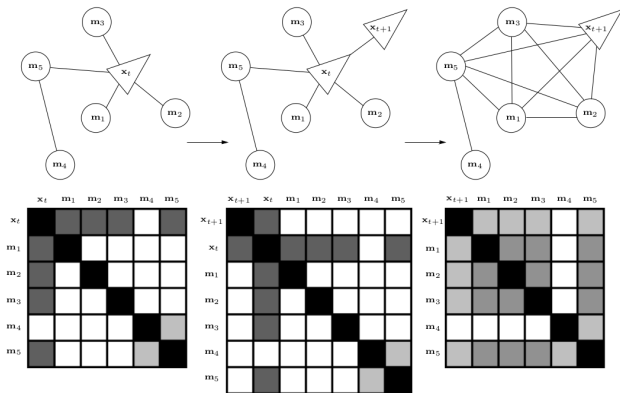


思考：如果是直接丢弃，信息矩阵如何变化？用边际概率来操作又会带来什么问题？

⁶Matthew R Walter, Ryan M Eustice, and John J Leonard. "Exactly sparse extended information filters for feature-based SLAM". In: *The International Journal of Robotics Research* 26.4 (2007), pp. 335-359. [↩](#) [↪](#) [↶](#) [↷](#) [↺](#) [↻](#)

toy example 4

如下图优化系统中，随着 x_{t+1} 的进入，变量 x_t 被移除⁷。



marginalization 会使得信息矩阵变稠密！原先条件独立的变量，可能变得相关。

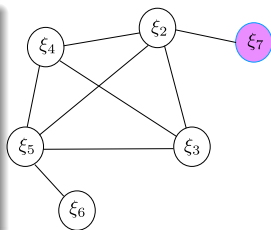
⁷Matthew R Walter, Ryan M Eustice, and John J Leonard. "Exactly sparse extended information filters for feature-based SLAM". In: *The International Journal of Robotics Research* 26.4 (2007), pp. 335-359.

Section 4

滑动窗口中的 FEJ 算法

toy example 3 再移除变量 ξ_1 后加入新变量

- 新的变量 ξ_7 跟老的变量 ξ_2 之间存在观测信息，能构建残差 r_{27} 。图模型如右图所示。
- 新残差加上之前 marg 留下的信息，构建新的最小二乘系统，对应的信息矩阵的变化如下图所示：



$$\Lambda \rightarrow \Lambda' + \Lambda_6 = \Lambda_{new}$$

The diagram illustrates the update of the information matrix Λ when a new variable ξ_7 is added. The matrices are represented as grids with columns labeled $\xi_2, \xi_3, \xi_4, \xi_5, \xi_6, \xi_7$.

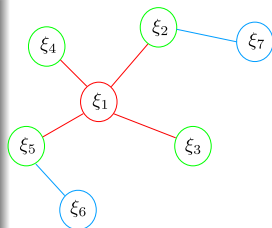
- Λ (6x6): Blue blocks at (2,2)-(5,5) and (6,6).
- Λ' (7x7): Blue blocks at (2,2)-(5,5) and (6,6).
- Λ_6 (7x7): Red blocks at (2,2), (2,7), (7,2), and (7,7).
- Λ_{new} (7x7): Blue blocks at (2,2)-(5,5) and (6,6), with red blocks at (2,7), (7,2), and (7,7).

注意： ξ_2 自身的信息矩阵由两部分组成，这会使得系统存在潜在风险。

系统回顾 toy example 3

toy example 3

- ① 如右图所示，在 $t \in [0, k]s$ 时刻，系统中状态量为 $\xi_i, i \in [1, 6]$ 。第 k' 时刻，加入新的观测和状态量 ξ_7 。
- ② 在第 k 时刻，最小二乘优化完以后，marg 掉变量 ξ_1 。被 marg 的状态量记为 x_m ，剩余的变量 $\xi_i, i \in [2, 5]$ 记为 x_r 。
- ③ marg 发生以后， x_m 所有的变量以及对应的测量将被丢弃。同时，这部分信息通过 marg 操作传递给了保留变量 x_r 。marg 变量的信息跟 ξ_6 不相关。
- ④ 第 k' 时刻，加入新的状态量 ξ_7 (记作 x_n) 以及对应的观测，开始新一轮最小二乘优化。



- 红色为被 marg 变量以及测量约束。
- 绿色为跟 marg 变量有关的保留变量。
- 蓝色为和 marg 变量无关联的变量。

marg 发生后，留下的到底是什么信息？

marg 前，变量 \mathbf{x}_m 以及对应测量 \mathcal{S}_m 构建的最小二乘信息矩阵为：

$$\begin{aligned} \mathbf{b}_m(k) &= \begin{bmatrix} \mathbf{b}_{mm}(k) \\ \mathbf{b}_{mr}(k) \end{bmatrix} = - \sum_{(i,j) \in \mathcal{S}_m} \mathbf{J}_{ij}^\top(k) \boldsymbol{\Sigma}_{ij}^{-1} \mathbf{r}_{ij} \\ \boldsymbol{\Lambda}_m(k) &= \begin{bmatrix} \boldsymbol{\Lambda}_{mm}(k) & \boldsymbol{\Lambda}_{mr}(k) \\ \boldsymbol{\Lambda}_{rm}(k) & \boldsymbol{\Lambda}_{rr}(k) \end{bmatrix} = \sum_{(i,j) \in \mathcal{S}_m} \mathbf{J}_{ij}^\top(k) \boldsymbol{\Sigma}_{ij}^{-1} \mathbf{J}_{ij}(k) \end{aligned} \quad (47)$$

marg 后，变量 \mathbf{x}_m 的测量信息传递给了变量 \mathbf{x}_r ：

$$\begin{aligned} \mathbf{b}_p(k) &= \mathbf{b}_{mr}(k) - \boldsymbol{\Lambda}_{rm}(k) \boldsymbol{\Lambda}_{mm}^{-1}(k) \mathbf{b}_{mm}(k) \\ \boldsymbol{\Lambda}_p(k) &= \boldsymbol{\Lambda}_{rr}(k) - \boldsymbol{\Lambda}_{rm}(k) \boldsymbol{\Lambda}_{mm}^{-1}(k) \boldsymbol{\Lambda}_{mr}(k) \end{aligned} \quad (48)$$

下标 p 表示 prior. 即这些信息将构建一个关于 \mathbf{x}_r 的 **先验信息**.

先验

我们可以从 $\mathbf{b}_p(k), \boldsymbol{\Lambda}_p(k)$ 中反解出一个残差 $\mathbf{r}_p(k)$ 和对应的雅克比矩阵 $\mathbf{J}_p(k)$. 需要注意的是，随着变量 $\mathbf{x}_r(k)$ 的后续不断优化变化，残差 $\mathbf{r}_p(k)$ 或者 $\mathbf{b}_p(k)$ 也将跟着变化，但雅克比 $\mathbf{J}_p(k)$ 则固定不变了。

新测量信息和旧测量信息构建新的系统

在 k' 时刻, 新残差 \mathbf{r}_{27} 和先验信息 $\mathbf{b}_p(k), \Lambda_p(k)$ 以及残差 \mathbf{r}_{56} 构建新的最小二乘问题, $\mathbf{J}_p(k)$ 和 $\mathbf{r}_p(k)$ 为先验部分对应的雅克比和残差:

$$\begin{aligned}\mathbf{b}(k') &= \Pi^\top \mathbf{J}_p(k) \mathbf{r}_p(k') - \sum_{(i,j) \in S_a(k')} \mathbf{J}_{ij}^\top(k') \Sigma_{ij}^{-1} \mathbf{r}_{ij}(k') \\ \Lambda(k') &= \Pi^\top \Lambda_p(k) \Pi + \sum_{(i,j) \in S_a(k')} \mathbf{J}_{ij}^\top(k') \Sigma_{ij}^{-1} \mathbf{J}_{ij}(k')\end{aligned}\quad (49)$$

其中 $\Pi = \begin{bmatrix} \mathbf{I}_{\dim \mathbf{x}_r} & \mathbf{0} \end{bmatrix}$ 用来将矩阵的维度进行扩张。 S_a 用来表示除被 marg 掉的测量以外的其他测量, 如 $\mathbf{r}_{56}, \mathbf{r}_{27}$ 。

出现的问题

- 由于被 marg 的变量以及对应的测量已被丢弃, 先验信息 $\Lambda_p(k)$ 中关于 \mathbf{x}_r 的雅克比在后续求解中没法更新。
- 但 \mathbf{x}_r 中的部分变量还跟其他残差有联系, 如 ξ_2, ξ_5 。这些残差如 \mathbf{r}_{27} 对 ξ_2 的雅克比会随着 ξ_2 的迭代更新而不断在最新的线性化点处计算。

信息矩阵的零空间变化

滑动窗口算法导致的问题

滑动窗口算法优化的时候，信息矩阵如公式(49)变成了两部分，且这两部分计算雅克比时的线性化点不同。这可能会导致信息矩阵的零空间发生变化，从而在求解时引入错误信息。

比如：求解单目 SLAM 进行 Bundle Adjustment 优化时，问题对应的信息矩阵 Λ 不满秩，对应的零空间为 N ，用高斯牛顿求解时有

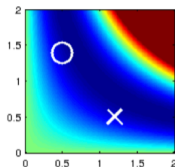
$$\begin{aligned}\Lambda \delta x &= b \\ \Lambda \delta x + \Lambda N &= b\end{aligned}\tag{50}$$

咦，增量 δx 在零空间维度下变化，并不会改变我们的残差。这意味着系统可以有多个满足最小化损失函数的解 x 。

toy example

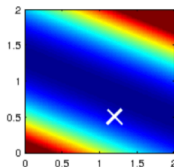
Windowed, real-time optimization: Consistency.

(for now, let's assume we have initializations, and know which points to use and where they are visible.)



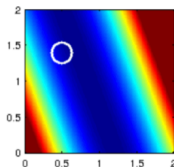
$$E = E_1 + E_2$$

$$= (xy - 1)^2 + (xy - 1)^2$$



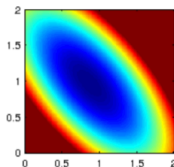
$$E'_1 = E_1$$

Lin. around (0.5, 1.4)



$$E'_2 = E_2$$

Lin. around (1.2, 0.5)



$$E' = E'_1 + E'_2$$

nullspace disappears!

never combine linearizations around different linearization points,
especially in the presence of non-linear nullspaces!

It will render unobservable dimensions observable, and corrupt the system.

多个解的问题，变成了一个确定解。不可观的变量，变成了可观的。

可观性的一种定义

对于测量系统 $\mathbf{z} = h(\boldsymbol{\theta}) + \varepsilon$, 其中 $\mathbf{z} \in \mathbb{R}^n$ 为测量值, $\boldsymbol{\theta} \in \mathbb{R}^d$ 为系统状态量, ε 为测量噪声向量。 $h(\cdot)$ 是个非线性函数, 将状态量映射成测量。对于理想数据, 如果以下条件成立, 则系统状态量 $\boldsymbol{\theta}$ 可观:

$$\forall \boldsymbol{\theta}, \forall \boldsymbol{\theta}' \in \mathbb{R}^d, \{\boldsymbol{\theta} \neq \boldsymbol{\theta}'\} \Rightarrow \{h(\boldsymbol{\theta}) \neq h(\boldsymbol{\theta}')\} \quad (51)$$

更多定义见论文⁸

对于 SLAM 系统而言 (如单目 VO), 当我们改变状态量时, 测量不变意味着损失函数不会改变, 更意味着求解最小二乘时对应的信息矩阵 Λ 存在着零空间。

⁸Claude Jauffret. "Observability and Fisher information matrix in nonlinear regression". In: *IEEE Transactions on Aerospace and Electronic Systems* 43.2 (2007), pp. 756–759.

单目 SLAM 系统 7 自由度不可观：6 自由度姿态 + 尺度。



单目 + IMU 系统是 4 自由度不可观：yaw 角 + 3 自由度位置不可观。
roll 和 pitch 由于重力的存在而可观，尺度因子由于加速度计的存在而可观。

白话：不同残差计算雅克比时线性化点不一致

信息矩阵多加了信息

零空间发生变化（退化）

多解问题变成单解

不可观问题变成可观

滑动窗口中的问题

滑动窗口算法中，对于同一个变量，不同残差对其计算雅克比矩阵时线性化点可能不一致，导致信息矩阵可以分成两部分，相当于在信息矩阵中多加了一些信息，使得其零空间出现了变化。

解决办法：First Estimated Jacobian

FEJ 算法：不同残差对同一个状态求雅克比时，线性化点必须一致。这样就能避免零空间退化而使得不可观变量变得可观。

论文推荐⁹

比如：toy example 3 中计算 r_{27} 对 ξ_2 的雅克比时， ξ_2 的线性化点必须和 r_{12} 对其求导时一致。

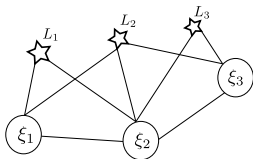
⁹Tue-Cuong Dong-Si and Anastasios I Mourikis. "Consistency analysis for sliding-window visual odometry". In: 2012 IEEE International Conference on Robotics and Automation. IEEE, 2012, pp. 5202–5209. 

Section 5

作业

作业

- ① 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 $r(\xi_i, L_k)$ 。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 Λ .
 - 2 请绘制相机 ξ_1 被 marg 以后的信息矩阵 Λ' .
- 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。
- 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.