VINS 中的边缘化

Wang Changlong

2022 年 1 月 6 日

1 什么是边缘化

边缘化，总结为一句话就是从联合概率分布中求边际概率。即

$$P(X=x)=\int_{-\infty}^{\infty}P(X=x,Y=y)dy \tag{1}$$

具体到 vins 系统中，为了维护变量的维度可控以达到实时性，我们常常维护一个固定长度的滑窗。当窗口进行滑动时，某些旧的变量需要被舍弃，这个舍弃的过程就牵扯到了边缘化。

考虑一个长度为 4 的变量窗口，在 $t=2$ 的时刻，$X=[x_0,x_1,x_2]$。$t=3$ 时来一个新的变量 x_3，我们把 x_3 加入到变量窗口里，并对变量 $X_{old}=[x_0,x_1,x_2,x_3]$ 进行估计，这时我们得到了关于变量 $x_i, i=1,2,3,4$ 的联合概率分布。而后，我们舍弃掉变量 x_0，而保留关于变量 $X_{new}=[x_1,x_2,x_3]$ 的联合概率分布，这个舍弃就是通过边缘化来实现的。

$$\begin{aligned} P(X_{new}=x_{new}) &= \int_{-\infty}^{\infty}P(X_{old}=[x_{new},\tilde{x}_0])d\tilde{x}_0 \\ &= \int_{-\infty}^{\infty}P(X_{new}=x_{new},x_0=\tilde{x}_0)d\tilde{x}_0 \end{aligned} \tag{2}$$

2 协方差矩阵、信息矩阵和边缘化

假设变量 (x,y) 的概率密度函数符合均值 μ，协方差 Σ 的高斯分布，具有如下的矩阵形式：

$$p(x,y)=\mathcal{N}(\begin{bmatrix}\mu_x\\ \mu_y\end{bmatrix},\begin{bmatrix}\Sigma_{xx} & \Sigma_{xy}\\ \Sigma_{yx} & \Sigma_{yy}\end{bmatrix}) \tag{3}$$

那么边缘化就非常好操作了，对于变量 y，其对应的子块 μ_y 和 Σ_{yy} 就是边缘化掉变量 x 后，y 的均值和协方差。

但是在很多基于优化的 vins 中，我们对于变量的分布是以信息矩阵 Λ 和信息向量 η 的形式给出的，具有如下形式：

$$p(x,y) = \mathcal{N}(\Lambda^{-1}\begin{bmatrix}\eta_x \\ \eta_y\end{bmatrix}, \begin{bmatrix}\Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy}\end{bmatrix}^{-1}) \tag{4}$$

这时候我们想进行边缘化得到变量 y 的信息矩阵就没有那么直观了。舒尔补分解协方差矩阵并求逆，可以得到信息矩阵和协方差矩阵的关系：

$$\begin{aligned}
\begin{bmatrix}\Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy}\end{bmatrix} &= \begin{bmatrix}\Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy}\end{bmatrix}^{-1} \\
&= \begin{bmatrix}\mathbf{1} & \mathbf{0} \\ \text{-}\Sigma_{yy}^{-1}\Sigma_{yx} & \mathbf{1}\end{bmatrix}\begin{bmatrix}(\Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{yy}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{1} & \text{-}\Sigma_{xy}\Sigma_{yy}^{-1} \\ \mathbf{0} & \mathbf{1}\end{bmatrix} \\
&= \begin{bmatrix}(\Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1} & -(\Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} \\ -\Sigma_{yy}^{-1}\Sigma_{yx}(\Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1} & \Sigma_{yy}^{-1}+\Sigma_{yy}^{-1}\Sigma_{yx}(\Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\end{bmatrix}
\end{aligned} \tag{5}$$

由此可得，变量 y 的信息矩阵，即协方差 Σ_{yy} 的逆为：

$$\Sigma_{yy}^{-1} = \Lambda_{yy} - \Lambda_{yx}\Lambda_{xx}^{-1}\Lambda_{xy} \tag{6}$$

我们知道，对于上述高斯分布，我们可以采用**舒尔补**进行分解，得到边缘概率密度函数和条件概率密度函数。

$$\begin{aligned}
p(x,y) &= p(x|y)p(y) \\
p(x|y) &= \mathcal{N}(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}) \\
p(y) &= \mathcal{N}(\mu_y, \Sigma_{yy})
\end{aligned} \tag{7}$$

这时我们对照公式 Eq(7) 和公式 Eq(6)，会发现条件概率 $P(x|y)$ 的协方差就是 Λ_{xx}^{-1}！也就是说，我们如果直接从信息矩阵中取块，得到的是变量的条件概率分布的信息矩阵。

这引出了一个非常有趣的现象：边际概率对于协方差矩阵的操作是很容易的，对于信息矩阵不好操作，而条件概率恰好相反。

3 边缘化的平方根方法

上节中我们得到，直接对信息矩阵进行舒尔补分解是可以实现边缘化的，这一节我们学习边缘化的平方根（Square Root）方法。平方根方法的引入源自在

求解非线性最小二乘问题时，信息矩阵通常是形如 $H = J^T J$ 来进行构造的，如高斯牛顿方法和 LM 方法。如果我们对 J 进行 QR 分解，有

$$\mathbf{J} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix}^T \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}, \quad with \quad \mathbf{R} = \begin{bmatrix} R & T \\ 0 & P \end{bmatrix} \tag{8}$$

那么对于信息矩阵有

$$\mathbf{H} = \mathbf{R}^T \mathbf{R} = \begin{bmatrix} R & T \\ 0 & P \end{bmatrix}^T \begin{bmatrix} R & T \\ 0 & P \end{bmatrix} = \begin{bmatrix} R^T R & R^T T \\ T^T R & P^T P + T^T T \end{bmatrix}^T \tag{9}$$

对照公式 Eq(6) 和公式 Eq(9)，我们发现恰好有 $\Sigma_{yy}^{-1} = P^T P$。对于 QR 分解，我们知道，它得到的是一个正交矩阵和一个上三角矩阵的乘积。因此我们想要边缘化的子变量部分必须要放在变量向量的头部，才能直接使用上述方法进行边缘化。

实际上，我们进行边缘化的时候并不需要真的要求出正交矩阵 $\mathbf{Q}$，我们的目的是将求出矩阵 R，而 Givens Rotation 告诉我们，只需要不断地左乘 Givens 旋转矩阵，即可将消去矩阵 $\mathbf{J}$ 的下三角中的非零元素。