vins的边缘化(marginalization)是VIO中比较难理解的部分,我整理了下之前看vins代码的这部分笔记,希望能给小伙伴们提供一点参考.

参考资料

[1].ceres-solver Covariance Estimation:http://www.ceres-solver.org/nnls_covariance.html

[2].SLAM中的marginalization 和 Schur complement:https://blog.csdn.net/heyijia0327/article/details/52822104

[3].Conditional and marginal distributions of a multivariate Gaussian:https://gbhqed.wordpress.com/2010/02/21/conditional-and-marginal-distributions-of-a-multivariate-gaussian/

[4].OKVIS 笔记：边缘化实现:https://fzheng.me/2018/03/23/okvis-marginalization/

[5].OKVIS 笔记：边缘化原理和策略:https://fzheng.me/2018/03/22/okvis-marginalization-base/

[6].论文:sliding window filter with application to planetary landing

残差的马氏范数与归一化

e的马氏范数和2-范数的转换:

$$||e||_{\Sigma}^{2}=e^{T}\Sigma^{-1}e=(\Sigma^{-1/2}e)^{T}(\Sigma^{-1/2}e)=||\Sigma^{-1/2}e||_{2}^{2}$$

由于ceres优化的最小二乘是$J=L^{T}L$的形式[1]，所以代码里通过$e'=\Sigma^{-1/2}e$ 对原始测量残差进行了处理。$\Sigma^{-1/2}$对应的即代码中的sqrt_info变量。

从概率的角度理解,马氏范数通过标准差将不同的观测的残差项的概率特性统一到同一个尺度下(残差项归一化)：

$$e\sim N(0,\Sigma)\Rightarrow\Sigma^{-1/2}e\sim N(0,1)$$

即每个观测残差除以标准差，在这个过程中消掉了观测残差的单位（长度，角度），使得不同观测来源的残差的2范数可以直接相加。

marginazation factor

舒尔补与边缘概率

根据最优化理论，采用高斯牛顿、LM等非线性优化方法对残差进行优化时，每次迭代等价于求解增量方程$H\delta x=b$,其中,令z为测量值，$h(x)$为观测模型,Σ为测量值的协方差矩阵，则变量间的关系为：

$$H=J^{T}J,b=-J^{T}e\,,\ J=\frac{\partial e'}{\partial x},e'=\sum_{k=0}^{n}(\Sigma^{-1/2}(z-h(x)))$$

令$\delta x=[\delta x_{1},\delta x_{2}]^{T}$,其中$\delta x_{1}$是需要margin的变量，$\delta x_{2}$是和$\delta x_{1}$有约束关系的变量。则有：

$$\begin{bmatrix} \Lambda_a, \Lambda_b \\ \Lambda_b^T, \Lambda_c \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

对于方程进行矩阵变化，两边都乘以舒尔项：

$$\begin{bmatrix} I, & 0 \\ -\Lambda_b^T \Lambda_a^{-1}, I \end{bmatrix} \begin{bmatrix} \Lambda_a, \Lambda_b \\ \Lambda_b^T, \Lambda_c \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix} = \begin{bmatrix} I, & 0 \\ -\Lambda_b^T \Lambda_a^{-1}, I \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$
$$\Downarrow$$
$$\begin{bmatrix} & \Lambda_a, & \Lambda_b \\ 0, & \Lambda_c - \Lambda_b^T \Lambda_a^{-1} \Lambda_b \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 - \Lambda_b^T \Lambda_a^{-1} b_1 \end{bmatrix}$$

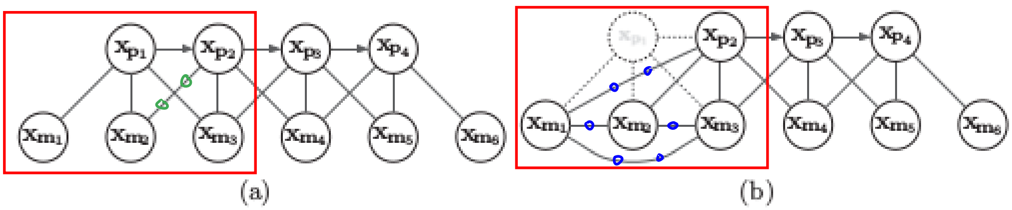
由此可以在不求解δx_1的情况下求解δx_2,并得到新的$H^* \delta x_2^* = b^*$形式：

$$(\Lambda_c - \Lambda_b^T \Lambda_a^{-1} \Lambda_b) \delta x_2 = b_2 - \Lambda_b^T \Lambda_a^{-1} b_1$$
$$\Downarrow$$
$$H^* \delta x_2^* = b^*$$

观察方程形式会发现，b^*是$[b_1, b_2] \propto [e_1, e_2]$的函数，即通过舒尔补过程，将要消去的变量$\delta x_1$构建的约束关系（残差）叠加到和其有观测关系的变量$\delta x_2$上，也就是方程$H^* \delta x^* = b^*$保留了先验信息（prior）。

H矩阵的稀疏结构只有在变量间存在约束关系的地方才有非零值，而通过舒尔补得到的新的H矩阵值，在之前0元素的地方出现了非零值，即出现了新的约束关系，如下图所示，消去x_{p1}的结果就是将x_{p1}对 x_{m1}的约束关系叠加到x_{p2}和其他有约束关系的变量上，比较明显的就是出现两条新的边:e_{p2m1}, e_{m1m3}(H^*矩阵中这两处的数值变成了非零)，其他和x_{p1}关联的变量间的约束关系数值上也有变化。

因此也会发现边缘化的代价是H矩阵变得越来越稠密。这也是一个需要处理的问题，具体在这里不展开了（其实我还没仔细研究过）。



从概率的角度解释，通过舒尔补得到的δx_2^*是边缘概率$\int_{\delta x_1} p(\delta x_2, \delta x_1)$的最大似然估计。可以参考[3].

$$\delta x_2^* = \arg\min_{\delta x_2} \log \int_{x_1} p(\delta x_2, \delta x_1)$$

更新先验残差

假设在变量χ_0进行了一次舒尔补的结果即论文中的prior为:

$$H_0^* \delta \chi_{02}^* = b_0^* \qquad \delta \chi_{02}^* = \arg\min_{\chi_{02}} \log \int_{\chi_{01}} p(\delta \chi_{02}, \delta \chi_{01})$$

基于χ_{02}^*处构建的先验信息，在之后的某个时刻对变量χ_k迭代优化的过程可以写做：

$$
\begin{gathered}
H\delta\chi_k = b \\
\Downarrow \\
\begin{bmatrix} H_{prior} & & \\ & H_B & \\ & & H_C \end{bmatrix}
\begin{bmatrix} \delta x_{prior} \\ \delta x_B \\ \delta x_C \end{bmatrix}
=
\begin{bmatrix} b_{prior} \\ b_B \\ b_C \end{bmatrix}
\end{gathered}
$$

$\delta x_{prior} = \chi_{k2} \boxminus \chi_{02}^{*}$,其中$\chi_{k2}$为$\chi_{02}^{*}$中的变量在$\chi_k$处的值.$\boxminus$是广义上的减法

为了构建χ_k状态的优化，在构建H时，需要得到χ_k处的H_{prior}和$b_{prior} \propto e_{prior}$。首先根据$\delta x_{prior}$的定义，其线性化点是$\chi_{02}^{*}$，所以

$$
H_{prior} = H_0^{*} = J_0^{*T} J_0^{*}
$$

根据H^{*}分解求得到对应的雅克比矩阵J^{*}:

H^{*}是对称矩阵，所以是hermitian矩阵，代码中使用Eigen::SelfAdjointEigenSolver完成对herrmitian的SVD分解。得到:

$$
\begin{gathered}
H_0^{*} = J_0^{*T} J_0^{*} = VSV^T \\
\Downarrow \\
J_0^{*} = S^{1/2} V^T \quad J_0^{*+} = VS^{-1/2}
\end{gathered}
$$

接着根据b_0和J^{*}更新求解χ_k处的先验项b_{prior}，通过一阶线性化求解b_{prior}:

$$
\begin{aligned}
b_{prior} &= b_0^{*} + \frac{\partial b}{\partial \chi}\Big|_{\chi_{02}^{*}} \delta x_{prior} \\
&= b_0^{*} - J_0^{*T} \frac{\partial e}{\partial \chi}\Big|_{\chi_{02}^{*}} \delta x_{prior} \\
&= b_0^{*} - J_0^{T*} J_0^{*} \delta x_{prior} \\
&= b_0^{*} - H_0^{*} \delta x_{prior}
\end{aligned}
$$

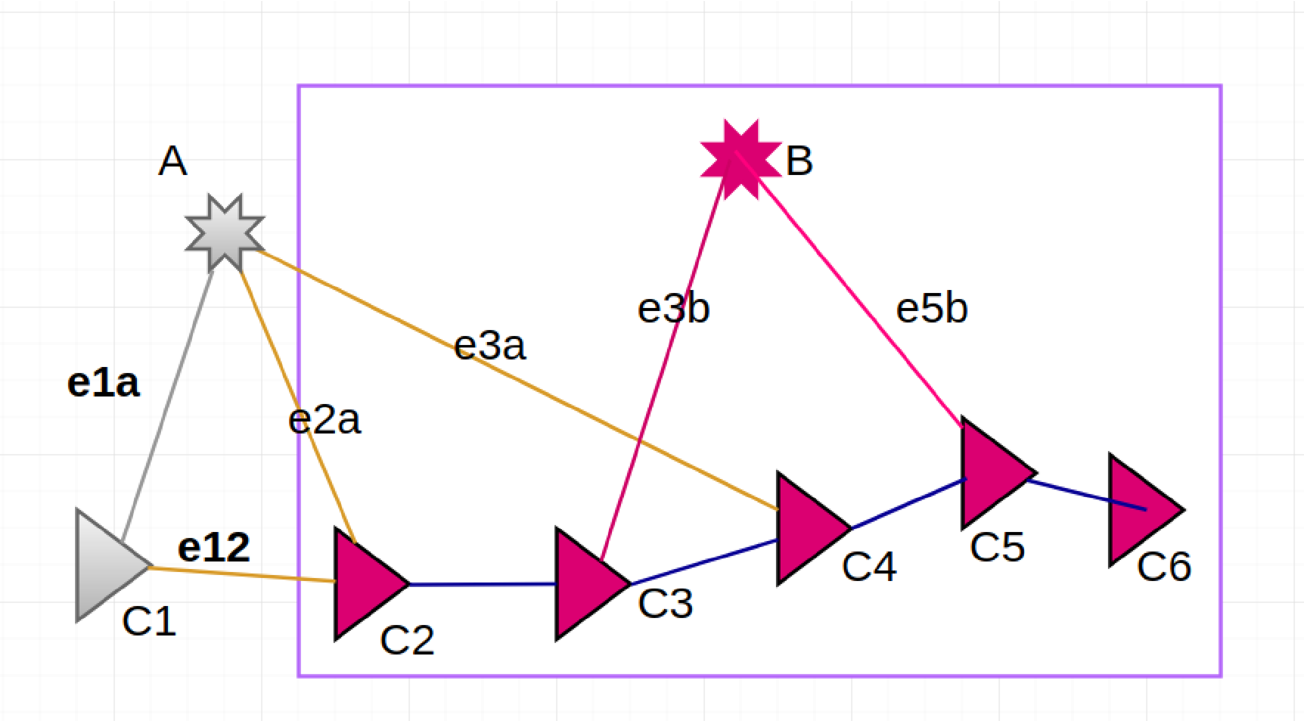
在ceres具体实现时，传入的是雅克比矩阵和残差项e_{prior}，所以需要求解出e_{prior}.利用b_{prior}的计算公式:

$$
\begin{gathered}
H_{prior} \delta x_{prior} = b_{prior} \\
\Downarrow \\
\begin{aligned}
J_0^{*T} J_0^{*} \delta x_{prior} &= b_0^{*} - H_0^{*} \delta x_{prior} \\
&= b_0^{*} - J_0^{T*} J_0^{*} \delta x_{prior} \\
&= -J_0^{*T} (-J_0^{T+} b_0^{*} + J_0^{*} \delta x_{prior})
\end{aligned} \\
\Downarrow \\
e_{prior} = -J_0^{*T+} b_0^{*} + J_0^{*} \delta x_{prior} \\
\Downarrow \\
e_0 = -J_0^{*T+} b_0^{*} \\
e_{prior} = e_0 + J_0^{*} \delta x_{prior}
\end{gathered}
$$

+ 表示逆运算

这里和代码上有点出入
因为H \delta x = -b , b = J^{T} e

公式推到这里其实先验要做的事情就比较清晰了,当一部分状态被消去(固定住,如下图的A, $C1$),而和消去状态有关联的状态(如下图的$C2$, $C3$, $C4$)变化时,通过一阶线性化更新两者之间的观测残差e_{prior}(下图的e_{12}, e_{2a}, e_{3a}).我的理解vins的优化并没有准从FEJ原则，对于先验部分使用J_0^{*}是因为后续优化中先验部分的线性化点一直是χ_{02}^{*}。



游振兴,youzx163@163.com