

MSCKF 公式推导及代码解析

崔华坤 2019.1.3

目录

一、简介.....	2
二、符号说明.....	3
三、状态向量.....	4
3.1 真实状态向量 true state	4
3.2 误差状态向量 error state	6
三、IMU 预测	6
3.1 离散形式下的状态向量预测	6
3.2 连续形式下误差状态的运动方程	6
3.3 离散状态下误差状态的运动方程	7
3.4 连续形式下协方差传播	7
四、相机位姿状态增广	9
4.1 误差状态向量的增广	9
4.3 协方差矩阵的增广	10
五、视觉测量模型	11
5.1 视觉测量残差.....	12
5.2 视觉 Jacobian 计算.....	12
5.3 视觉 Jacobian 处理.....	15
六、视觉更新.....	16
七、附录.....	16
7.1 ESKF 误差运动方程推导中的两处说明.....	16
7.2 增广的误差状态向量与原来误差状态的关系	17
7.3 噪声项与状态向量的相关性对结果的影响	19
八、参考文献.....	20

一、简介

MSCKF 是明尼苏达州大学 Mourikis 等人提出的一种基于 EKF 的 VIO 紧耦合的 SLAM 框架[1][2]。该框架的最大创新点在于未将路标点加入到状态向量中（因为加进去会导致状态向量一直增加，效率会越来越慢），而是等某个路标点不见或者太老时，先通过 GN 优化方法计算出该路标点的 3D 位置，然后将多个相机对这个路标点的观测作为一种约束，整合到 EKF 更新中，这样既不损失信息，又不增加状态向量，有点像边缘化 Marg 的思路。

MSCKF 代码[https://github.com/daniilidis-group/msckf_mono]算法流程如下图所示，主要分为预测、更新：

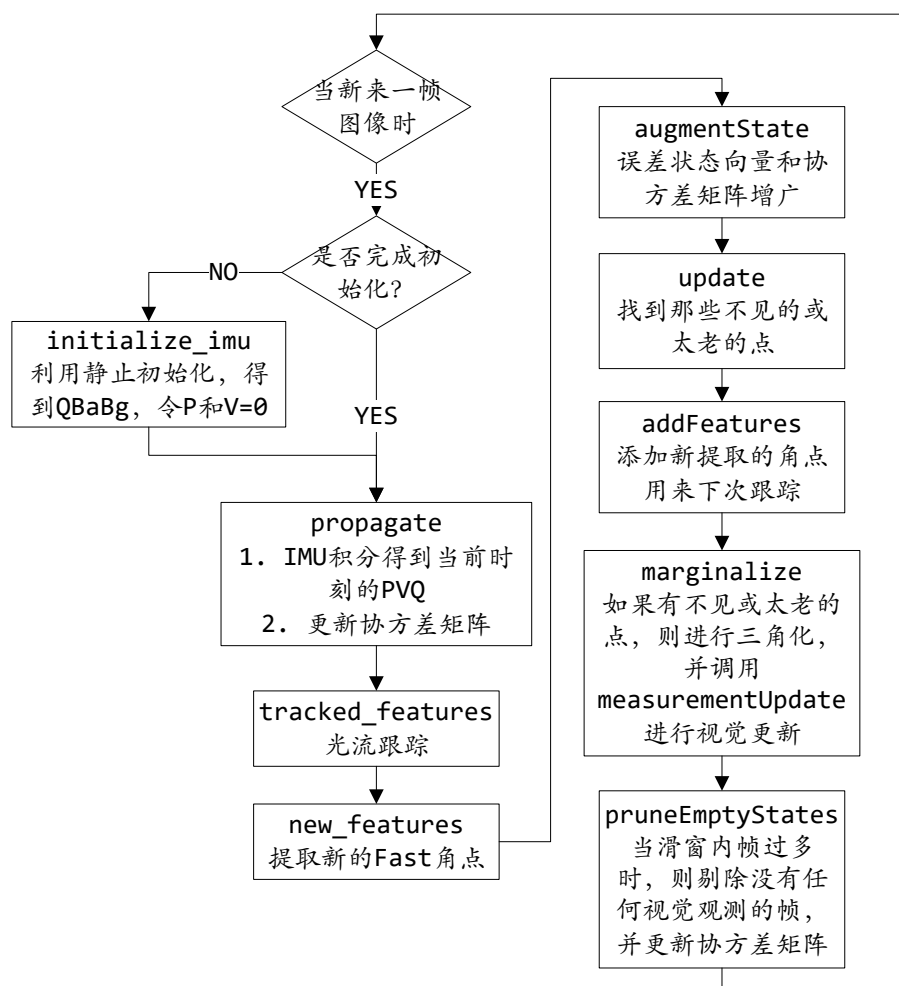


图 1 MSCKF 算法流程

下面对算法流程进行详细说明：

1. 初始化 initialize_imu: 当新来一帧图像时，先判断是否完成初始化。若未完

- 成，则进行静止初始化，即保持静止，利用平均的 `acc` 和 `gyr` 数据，计算相对重力的旋转角度、并计算 `acc` 和 `gyr` 的 `bias`，令初始位置和速度为 0；
2. 预测 `propagate`: 当完成初始化后，找出两帧图像之间的所有 IMU 数据进行预测：首先根据 IMU 积分出当前图像帧的 PVQ，另外更新整体的协方差矩阵；
 3. 视觉跟踪 `tracked_features`: 利用金字塔光流跟踪上一帧图像在当前帧上的特征点，跟踪时可以利用两帧之间 `gyr` 提供的帧间旋转，预测上一帧的特征点在当前帧上可能的初始位置，便于更准确地进行光流跟踪。当跟踪完成后，若当前帧跟踪的点数小于阈值，则重新提前些新的 `fast` 角点；之后将新点添加进来用于下一帧的跟踪 `addFeatures`；
 4. 视觉增广 `augmentState`: 将当前帧位姿的误差状态，添加到整体的误差状态向量中，并相应增广整体的协方差矩阵；
 5. 视觉更新 `update`: 分析跟踪的特征点，找到那些不被当前帧所观测的，或者跟踪帧数太老的点，并将滑窗中所有相机对这些点的观测信息去掉；
 6. 边缘化 `marginalize`: 对上面找到的那些不见点或老点进行三角化，在三角化之前先对这些老点进行检测，若某个点的共视帧的距离小于阈值则忽略该点（在 `checkMotion` 中），因为距离太近会导致三角化的误差增大；然后，再利用滑窗内所有观测到该点的共视帧进行优化求解，并调用 `measurementUpdate` 将这个点 `marg` 掉，并对共视帧添加约束，最后根据这些约束信息进行 EKF 更新，即计算卡尔曼增益、更新状态向量和协方差矩阵；
 7. 剔除老帧 `pruneEmptyStates`: 当滑窗内相机的图像帧大于阈值时，则找到那些已经没有任何观测信息的帧并把它剔除掉，并更新协方差矩阵。

二、符号说明

x : true state, 真实状态，可以理解为估计值，含有估计误差；

\hat{x} : nominal state, 名义状态，为理想的准确值，无误差，无噪声和扰动；

\tilde{x} : error state, 误差状态， $x = \tilde{x} + \hat{x}$ 。

三者关系可以形象地用下图表示，状态向量 x 是一个小车沿直线上的距离，那么，小车理想的准确位置就是 nominal \hat{x} ，但是这个理想的准确值，我们是无

法获得的，毕竟传感器都存在测量误差，导致我们只能得到一个实际的测量值或者估计值，就是 true state x 。这个测量值 x 跟理想值 \hat{x} 之间就差一个测量误差 \tilde{x} 。

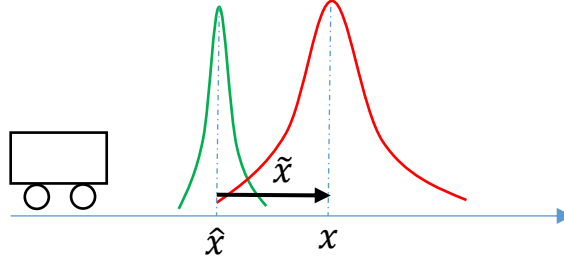


图 2 true 和 nominal 的关系

其中，四元数为：

$$q = \delta q \otimes \hat{q}, \quad \delta q \cong \begin{bmatrix} \frac{1}{2}\delta\theta \\ 1 \end{bmatrix} \quad (1)$$

true、nominal 和 error 的概念如下表所示，具体可参考论文[6]。

表 1 误差状态的卡尔曼滤波（ESKF）中的所有变量

Magnitude	True	Nominal	Error	Composition	Measured	Noise
Full state ⁽¹⁾	\mathbf{x}_t	\mathbf{x}	$\delta\mathbf{x}$	$\mathbf{x}_t = \mathbf{x} \oplus \delta\mathbf{x}$		
Position	\mathbf{p}_t	\mathbf{p}	$\delta\mathbf{p}$	$\mathbf{p}_t = \mathbf{p} + \delta\mathbf{p}$		
Velocity	\mathbf{v}_t	\mathbf{v}	$\delta\mathbf{v}$	$\mathbf{v}_t = \mathbf{v} + \delta\mathbf{v}$		
Quaternion ^(2,3)	\mathbf{q}_t	\mathbf{q}	$\delta\mathbf{q}$	$\mathbf{q}_t = \mathbf{q} \otimes \delta\mathbf{q}$		
Rotation matrix ^(2,3)	\mathbf{R}_t	\mathbf{R}	$\delta\mathbf{R}$	$\mathbf{R}_t = \mathbf{R} \delta\mathbf{R}$		
Angles vector ⁽⁴⁾			$\delta\boldsymbol{\theta}$	$\delta\mathbf{q} = e^{\delta\boldsymbol{\theta}/2}$ $\delta\mathbf{R} = e^{[\delta\boldsymbol{\theta}]_{\times}}$		
Accelerometer bias	\mathbf{a}_{bt}	\mathbf{a}_b	$\delta\mathbf{a}_b$	$\mathbf{a}_{bt} = \mathbf{a}_b + \delta\mathbf{a}_b$		\mathbf{a}_w
Gyrometer bias	$\boldsymbol{\omega}_{bt}$	$\boldsymbol{\omega}_b$	$\delta\boldsymbol{\omega}_b$	$\boldsymbol{\omega}_{bt} = \boldsymbol{\omega}_b + \delta\boldsymbol{\omega}_b$		$\boldsymbol{\omega}_w$
Gravity vector	\mathbf{g}_t	\mathbf{g}	$\delta\mathbf{g}$	$\mathbf{g}_t = \mathbf{g} + \delta\mathbf{g}$		
Acceleration	\mathbf{a}_t				\mathbf{a}_m	\mathbf{a}_n
Angular rate	$\boldsymbol{\omega}_t$				$\boldsymbol{\omega}_m$	$\boldsymbol{\omega}_n$

三、状态向量

3.1 真实状态向量 true state

MSCKF 的状态向量，包括两部分：当前时刻的 PVQB 即 X_{IMU_k} ，和 N 个相机位姿。以下图为例，状态向量包含 5 个相机的位姿 $T_{l \sim 5}$ 和当前时刻的位姿 X_{IMU_5} ，

其中, X_{IMU_5} 和 T_5 均表示当前时刻的状态, 只是局部坐标系不同: 一个是 IMU 系、一个是 Camera 系, 两者通过图中红色方块的外参进行约束。另外, 黄色小块为 IMU 预测因子, 蓝色小块为视觉更新因子。值得注意的是, 图中的路标点 $f_1 \sim f_3$ 并未放到优化变量中, 而是通过足够多的 KF 对路标点进行三角化, 并相信三角化的结果是准确的, 从而将其固定住并边缘化掉, 将边缘后的影响作为先验约束加到滑窗内的共视相机位姿上, 再进行视觉更新。下图中会用 f_1 和 f_2 这两个路标点进行视觉更新, f_1 是因为太老, 被滑窗所有帧都看见; f_2 是不被当前帧 T_5 所观测到了; f_3 因还能被当前帧观测到, 且不算太老, 因此留在窗口中继续跟踪。

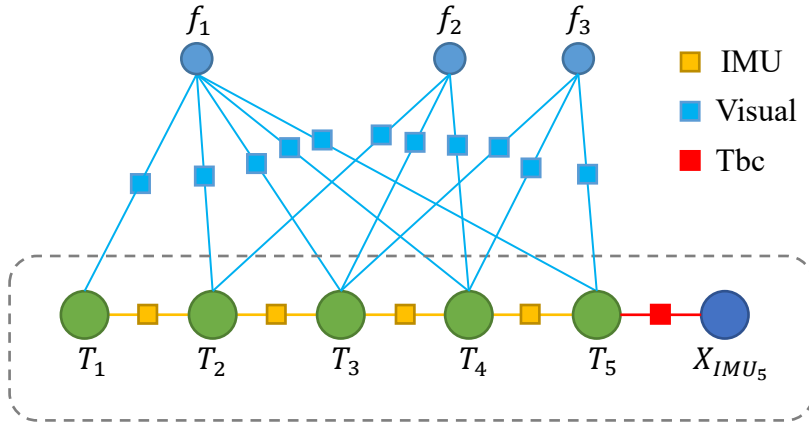


图 3 MSCKF 优化变量

优化变量的数学形式为:

$$X_k^{(16+7N) \times 1} = [X_{IMU_k} \quad {}^{c_1}_G \bar{q} \quad {}^G p_{c_1} \cdots {}^{c_N}_G \bar{q} \quad {}^G p_{c_N}]^T \quad (2)$$

其中, ${}^{c_1}_G \bar{q} = \bar{q}_{c_1 \leftarrow G}$ 为归一化四元数, ${}^G p_{c_1} = p_{G \leftarrow c_1}$, 注意旋转和平移的坐标系方向相反; 第 k 帧的优化变量 X_{IMU_k} 为:

$$X_{IMU_k}^{16 \times 1} = [{}^I_G \bar{q} \quad b_g \quad {}^G v_l \quad b_a \quad {}^G p_l]^T \quad (3)$$

$$T_i^{7 \times 1} = [{}^{c_i}_G \bar{q} \quad {}^G p_{c_i}]^T \quad (4)$$

严格来说, 优化变量矩阵中的各元素右上角也应加上转置符号“ T ”, 即优化变量为多行一列排布, 这里为了形式上简单, 不严谨地省去了转置符号, 仅在整体矩阵外面加上转置。

3.2 误差状态向量 error state

MSCKF 在预测和更新时的处理对象为误差状态向量，并非上面的实际状态向量，理由有四：

1. 朝向用四元数表示是过参的，但是角度误差的参数只有 3 个，与自由度是一致的；
2. 误差的均值为 0，更靠近初值，线性化点较准确；
3. 误差的值很小，可忽略二阶；
4. 误差的运动很慢。

那么，对应地，误差状态向量为：

$$\tilde{X}_k^{(15+6N) \times 1} = [\tilde{X}_{IMU_k} \ \delta\theta_{C_1} \ {}^G\tilde{p}_{C_1} \ \cdots \ \delta\theta_{C_N} \ {}^G\tilde{p}_{C_N}]^T \quad (5)$$

其中， $\delta\theta_{C_1} = \delta\theta_{C_1 \leftarrow G}$ ， ${}^G\tilde{p}_{C_1} = \tilde{p}_{G \leftarrow C_1}$ ；第 k 帧的误差状态向量 \tilde{X}_{IMU_k} 为：

$$\tilde{X}_{IMU_k}^{15 \times 1} = [\delta\theta_l \ \tilde{b}_g \ {}^G\tilde{v}_l \ \tilde{b}_a \ {}^G\tilde{p}_l]^T \quad (6)$$

其中， $\delta\theta_l = \delta\theta_{l \leftarrow G}$ ，表示从世界系到 IMU 局部系的旋转角度。

三、IMU 预测

3.1 离散形式下的状态向量预测

(补 PVQ 积分公式)

3.2 连续形式下误差状态的运动方程

$$\dot{\tilde{X}}_{IMU}^{15 \times 1} = F^{15 \times 15} \tilde{X}_{IMU}^{15 \times 1} + G^{15 \times 12} n_{IMU}^{12 \times 1} \quad (7)$$

其中， F 和 G 的详细推导可参考文献[6]或 VINS 的解析笔记，这里的区别在于考虑的地球旋转（但在代码中并未考虑）：

$$F = \begin{bmatrix} -[\hat{\omega} \times] & -I_3 & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ -C_q^T [\hat{a} \times] & 0_{3 \times 3} & -2[\omega_G \times] & -C_q^T & -[\omega_G \times]^2 \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & I_3 & 0_{3 \times 3} & 0_{3 \times 3} \end{bmatrix} \quad (8)$$

$$\mathbf{G} = \begin{bmatrix} -\mathbf{I}_3 & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{I}_3 & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & -\mathbf{C}_{\hat{q}}^T & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{I}_3 \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \end{bmatrix} \quad (9)$$

$$n_{IMU} = \begin{bmatrix} n_g \\ n_{wg} \\ n_a \\ n_{wa} \end{bmatrix} \quad (10)$$

其中, $C_{\hat{q}} = C(I_G \hat{q})$, ω_G 为地球自转角速度, 当前时刻的角速度和加速度

为: $\hat{w} = w_m - b_g$, $\hat{a} = a_m - b_a$ 。

3.3 离散状态下误差状态的运动方程

那么, 根据公式(7)的连续状态下的误差导数公式, 可推导出离散状态下相邻两时刻的误差关系为:

$$\tilde{X}_{IMU_{k+1}} = \Phi(t_k + \Delta T, t_k) \tilde{X}_{IMU_k} + (G \Delta T) n_{IMU} \quad (11)$$

其中, ΔT 为第 k 个 IMU 时刻到第 $k+1$ 个 IMU 时刻的间隔, Φ 为第 k 时刻的 IMU 误差项到第 $k+1$ 时刻的变换关系, 文献[1]将 Φ 写成指数形式, 文献[5]写成近似形式:

$$\Phi(t_k + \Delta T, t_k) = \exp(F \Delta T) \cong \mathbf{I}_{15 \times 15} + F \Delta T \quad (12)$$

但是, `msckf_mono` 代码 `propagate()`中与上面不同: 因借鉴 `Apply observability constraints - enforce nullspace of Phi`, 《*Observability-constrained Vision-aided Inertial Navigation*》 Hesch J, 深层原因待分析。

3.4 连续形式下协方差传播

假设第 k 时刻的整体的协方差矩阵如下所示:

$$P_{k|k}^{(15+6N) \times (15+6N)} = \begin{bmatrix} P_{II_{k|k}} & P_{IC_{k|k}} \\ P_{IC_{k|k}}^T & P_{CC_{k|k}} \end{bmatrix} \quad (13)$$

其中, $P_{II_{k|k}}^{15 \times 15}$ 为当前帧的 PVQB (即 \tilde{X}_{IMU_k}) 与自身的协方差矩阵;

$P_{IC_{k|k}}^{15 \times 6N}$ 为当前帧的 PVQB (即 \tilde{X}_{IMU_k}) 与滑窗内其他帧位姿的协方差矩阵;

$P_{CC_{k|k}}^{6N \times 6N}$ 为滑窗内所有相机的位姿与自身的协方差矩阵。下图给出了整体的误

差状态向量，以及整体的协方差矩阵：

$\delta\theta$	\tilde{b}_w	\tilde{v}	\tilde{b}_a	\tilde{p}	δT_1	δT_N
$P_{II_{k k}}$					$P_{IC_{k k}}$				
$P_{IC_{k k}}^T$					$P_{CC_{k k}}$				

图 4 第 k 时刻的整体误差状态向量和整体协方差矩阵

当下一个 IMU，即第 k+1 个 IMU 来到后，此时 k+1 时刻的整体协方差矩阵可写为：

$$P_{k+1|k}^{(15+6N) \times (15+6N)} = \begin{bmatrix} P_{II_{k+1|k}} & \Phi(t_k + \Delta T, t_k) P_{IC_{k|k}} \\ P_{IC_{k|k}}^T \Phi(t_k + \Delta T, t_k)^T & P_{CC_{k|k}} \end{bmatrix} \quad (14)$$

第 k+1 时刻的整体协方差矩阵可用下图表示：

$\delta\theta$	\tilde{b}_w	\tilde{v}	\tilde{b}_a	\tilde{p}	δT_1	δT_N
$P_{II_{k+1 k}}$					$\Phi P_{IC_{k k}}$				
$P_{IC_{k k}}^T \Phi^T$					$P_{CC_{k k}}$				

图 5 第 k+1 时刻的整体误差状态向量和整体协方差矩阵

那么，根据 Lyapunov equation（李亚普诺夫方程）的数值积分方法：

$$\dot{P}_{II} = F P_{II} + P_{II} F^T + G Q_{IMU} G^T \quad (15)$$

计算 $P_{II_{k+1}|k}$:

$$P_{II_{k+1}|k} = \Phi P_{II_k|k} \Phi^T + G Q_{IMU} G^T \Delta T \quad (16)$$

第 k+1 时刻的 IMU 跟 Camera 位姿的协方差计算公式为:

$$P_{IC_{k+1}|k} = \Phi(t_k + \Delta T, t_k) P_{IC_k|k} \quad (17)$$

上面公式可以理解为 (P 为条件概率):

$$P(I_{k+1}|C_k) = P(I_{k+1}|I_k)P(I_k|C_k) \quad (18)$$

MSCKF 在离散形式的误差运动方程采用五阶龙格-库塔数值积分方法。

四、相机位姿状态增广

4.1 误差状态向量的增广

当获得一张新的相机图像时, 假设为第 N+1 帧图像, 则需要将相机位姿的误差加入到当前优化状态向量中。首先, 我们给出第 N+1 帧图像时刻的位姿:

1. 无误差和噪声的相机位姿 nominal-state, 可写成如下形式:

$$\hat{T}_{C_{N+1} \leftarrow G}^{7 \times 1} = \begin{bmatrix} {}^{C_{N+1}}_G \hat{q} \\ {}^G \hat{p}_{C_{N+1}} \end{bmatrix} = \begin{bmatrix} {}^{C_{N+1}}_G \bar{q} \otimes {}^{I_{N+1}}_G \hat{q} \\ {}^G \hat{p}_{I_{N+1}} + {}^{I_{N+1}}_G \hat{q}^T {}^I p_C \end{bmatrix} \quad (19)$$

2. 根据 true-state = nominal-state + error-state, 可计算出带误差和噪声的相机位姿 true-state。首先, 根据公式(19)的中间一项 nominal, 加上 error 可得:

$$T_{C_{N+1} \leftarrow G}^{7 \times 1} = \delta T_{C_{N+1} \leftarrow G} \oplus \hat{T}_{C_{N+1} \leftarrow G} = \begin{bmatrix} \exp(\delta \theta_{C_{N+1} \leftarrow G}^\wedge) {}^{C_{N+1}}_G \hat{q} \\ {}^G \hat{p}_{C_{N+1}} + {}^G \tilde{p}_{C_{N+1}} \end{bmatrix} \quad (20a)$$

另外, 将公式(19)的最右边一项 nominal 中的变量, 加上 error 可的:

$$T_{C_{N+1} \leftarrow G}^{7 \times 1} = \begin{bmatrix} {}^{C_{N+1}}_G \bar{q} \otimes \exp(\delta \theta_{I_{N+1} \leftarrow G}^\wedge) {}^{I_{N+1}}_G \hat{q} \\ ({}^G \hat{p}_{I_{N+1}} + {}^G \tilde{p}_{I_{N+1}}) + [\exp(\delta \theta_{I_{N+1} \leftarrow G}^\wedge) {}^{I_{N+1}}_G \hat{q}]^T {}^I p_C \end{bmatrix} \quad (20b)$$

因为 MSCKF 中讨论的状态向量为状态向量的误差, 因此增广的误差状态向量为当前帧即第 N+1 帧的位姿误差。为了区别, 我们将位姿的误差用绿色表示, 原来的整体的误差状态向量 (如当前帧的 PVBQ 误差和滑窗内其他老帧的位姿误差) 用橙色表示。那么, 需要增广的位姿误差可写为:

$$\delta T_{C_{N+1} \leftarrow G} = \begin{bmatrix} \delta \theta_{C_{N+1} \leftarrow G} \\ {}^G \tilde{p}_{C_{N+1}} \end{bmatrix} \quad (21)$$

那么，根据上面公式（20）的 true-state 位姿状态公式，可继续推导出第 N+1 帧相机位姿的误差状态向量，跟原来的整体误差状态向量的关系，即将公式中的每一行单独写开为：

第一行为：

$$\begin{aligned} \exp(\delta \theta_{C_{N+1} \leftarrow G}^\wedge)^{C_{N+1} \hat{q}} &= {}^C \bar{q} \otimes \exp(\delta \theta_{I_{N+1} \leftarrow G}^\wedge)^{I_{N+1} \hat{q}} \\ \Leftrightarrow \delta \theta_{C_{N+1} \leftarrow G} &= {}^C \bar{q} \delta \theta_{I_{N+1} \leftarrow G} \end{aligned} \quad (22)$$

第二行为：

$$\begin{aligned} {}^G \hat{p}_{C_{N+1}} + {}^G \tilde{p}_{C_{N+1}} &= ({}^G \hat{p}_{I_{N+1}} + {}^G \tilde{p}_{I_{N+1}}) + [\exp(\delta \theta_{I_{N+1} \leftarrow G}^\wedge)^{I_{N+1} \hat{q}}]^T {}^I p_C \\ \Leftrightarrow {}^G \tilde{p}_{C_{N+1}} &= {}^G \tilde{p}_{I_{N+1}} + {}^{I_{N+1} \hat{q}}_G \hat{q}^T ({}^I p_C)^\wedge \delta \theta_{I_{N+1} \leftarrow G} \end{aligned} \quad (23)$$

上面两式的详细推导可参考附录 7.2。

4.3 协方差矩阵的增广

当来一帧新图像后，新的协方差矩阵可写成：

$$P_{k|k} \leftarrow \frac{\partial \delta T_{C_{N+1} \leftarrow G}}{\partial \tilde{X}} \cdot P_{k|k} \cdot \frac{\partial \delta T_{C_{N+1} \leftarrow G}^T}{\partial \tilde{X}}$$

即：

$$\begin{aligned} P_{k|k}^{(15+6(N+1)) \times (15+6(N+1))} &\leftarrow \begin{bmatrix} I_{6N+15} \\ J \end{bmatrix} P_{k|k}^{(15+6N) \times (15+6N)} \begin{bmatrix} I_{6N+15} \\ J \end{bmatrix}^T \\ &= \begin{bmatrix} P_{k|k} & PJ^T \\ JP & JPJ^T \end{bmatrix} \end{aligned} \quad (24)$$

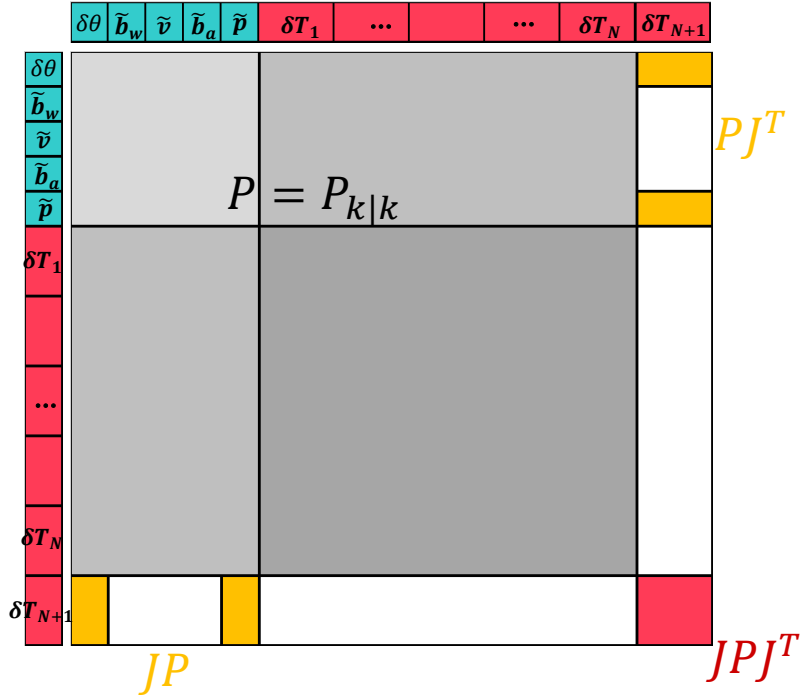


图 6 视觉增广后的协方差矩阵

其中，

$$\begin{aligned}
 J_{6 \times (15+6N)} &= \frac{\partial \delta T_{C_{N+1} \leftarrow G}}{\partial \tilde{X}} = \begin{bmatrix} \frac{\partial \delta \theta_{C_{N+1}}}{\partial \delta \theta_{I_{N+1}}} & 0_{3 \times 9} & \frac{\partial \delta \theta_{C_{N+1}}}{\partial {}^G \tilde{p}_{I_{N+1}}} & 0_{3 \times 6N} \\ \frac{\partial {}^G \tilde{p}_{C_{N+1}}}{\partial \delta \theta_{I_{N+1}}} & 0_{3 \times 9} & \frac{\partial {}^G \tilde{p}_{C_{N+1}}}{\partial {}^G \tilde{p}_{I_{N+1}}} & 0_{3 \times 6N} \end{bmatrix} \\
 &= \begin{bmatrix} C({}^C \bar{q}) & 0_{3 \times 9} & 0_{3 \times 3} & 0_{3 \times 6N} \\ {}^C \bar{q}^T ({}^I p_c)^\wedge & 0_{3 \times 9} & I_{3 \times 3} & 0_{3 \times 6N} \end{bmatrix} \quad (23)
 \end{aligned}$$

上式中与代码不同？

其中，详细的 Jacobian 推导可查看附录 7.2。

由上面分析可知，当新来一帧图像后，对整体协方差矩阵进行增广后，并不影响原有的整体协方差矩阵 $P_{k|k}$ ，只是在原协方差矩阵 $P_{k|k}$ 的最后新增 6 行 6 列而已。并且，增加的中只有与 $\delta \theta_{I_{N+1}}$ 和 ${}^G \tilde{p}_{I_{N+1}}$ 的两处及对角线处有值，其他处均为 0。

五、视觉测量模型

5.1 视觉测量残差

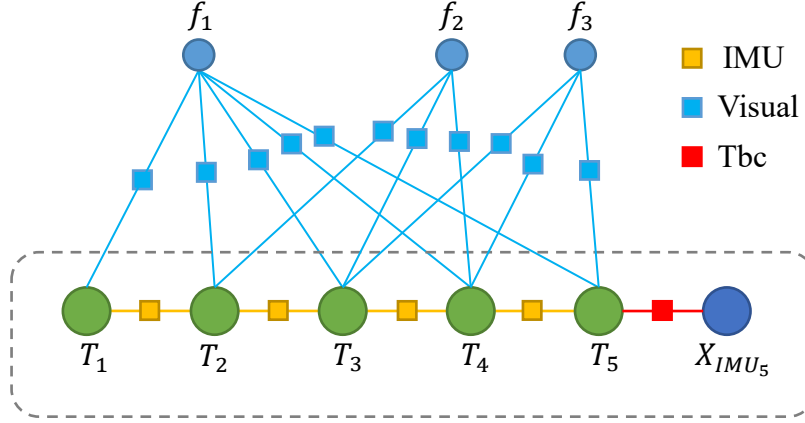


图 6 视觉重投影误差

对于第 j 个路标点，它在第 i 个相机中所观测到的重投影误差为：

$$r^{2 \times 1} = p_c - \hat{p}_c \quad (24)$$

其中， p_c 为该路标点被第 i 个相机实际观测到的，在归一化相机系下的坐标值，可根据 CCD 像素平面的坐标值推导可得：

$$p_c = \begin{bmatrix} \frac{x}{z} \\ \frac{y}{z} \end{bmatrix} + n = \begin{bmatrix} \frac{u - c_x}{f_x} \\ \frac{v - c_y}{f_y} \end{bmatrix} + n \quad (25)$$

其中， n 为在归一化坐标系下的视觉测量误差，可根据标定的像素误差（如 1.5 个像素）计算在归一化坐标系下的值来得到。

而 \hat{p}_c 为计算得到的在归一化相机系下的坐标值：

$$\hat{p}_c = \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} \quad (26)$$

其中，路标点在相机系下坐标值可通过下式获得：

$$\hat{p}'_c = \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} = c \left({}^{c_i} \hat{q} \right) \left({}^G \hat{p}_{f_j} - {}^G \hat{p}_{c_i} \right) \quad (27)$$

5.2 视觉 Jacobian 计算

下面计算视觉重投影误差相对于整体的误差状态向量和第 j 个路标点的 Jacobian，即 r 相对于当前时刻下的整体误差状态向量 \tilde{X} 和第 j 个路标点在世界系

下的坐标 $P_w = {}^G\hat{p}_{f_j}$ 来分别求 Jacobian。

根据常见的 Gauss-Newton 思路，可将误差在线性化点处进行展开：

$$e(x + \delta x) \approx e(x) + \frac{\partial e}{\partial x} \delta x。$$

对于我们这里讨论的误差状态向量，其线性化点在 0 处，即误差的状态分布为均值为 0 的高斯分布。那么，重投影误差可近似展开成如下形式：

$$r^{2 \times 1} \cong H_X^{2 \times (15+6N)} \tilde{X}^{(15+6N) \times 1} + H_{f_j}^{2 \times 3} {}^G\hat{p}_{f_j}^{3 \times 1} + n^{2 \times 1} \quad (28)$$

其中， n_i 的协方差矩阵为 $R = \sigma_{img}^2 I_2$ ；

根据链式法则，可计算出关于整体误差状态向量 \tilde{X} 的 Jacobian：

$$\begin{aligned} H_X &= \frac{\partial r}{\partial \tilde{X}} = \frac{\partial r}{\partial P'_c} \frac{\partial P'_c}{\partial \tilde{X}} \\ &= [0_{2 \times 15} \quad 0_{2 \times 6} \quad \cdots \quad J_{P'_c} [R_{C \leftarrow G} (t_{G \leftarrow C} - P_G)]^\wedge \quad -J_{P'_c} R_{C \leftarrow G} \quad \cdots] \end{aligned} \quad (29)$$

上面关于误差状态向量的 Jacobian 可以用下图形象表示，只有观测到该路标点的第 i 个相机位姿处（图中红色块）有值，其他处均为 0：

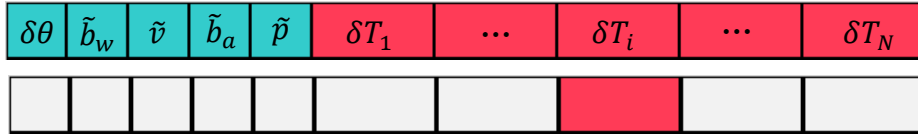


图 7 重投影误差关于整体误差状态向量的 Jacobian

另外，根据链式法则，同样可计算出关于路标点 ${}^G\hat{p}_{f_j}$ 的 Jacobian：

$$H_{f_j} = \frac{\partial r}{\partial {}^G\hat{p}_{f_j}} = \frac{\partial r}{\partial P'_c} \frac{\partial P'_c}{\partial {}^G\hat{p}_{f_j}} = J_{P'_c} \times R_{C \leftarrow G} \quad (30)$$

其中， $J_{P'_c}$ 为重投影误差相对于相机系下坐标 P'_c 的 Jacobian：

$$J_{P'_c} = \frac{\partial r}{\partial P'_c} = \frac{1}{z} \begin{bmatrix} 1 & 0 & -\frac{x}{z} \\ 0 & 1 & -\frac{y}{z} \end{bmatrix} \quad (31)$$

我们将对该路标点的所有 Camera 观测汇总可得（设该路标点被 M 个 Camera 观测到）：

$$r^{2M \times 1} \cong H_X^{2M \times (15+6N)} \tilde{X}^{(15+6N) \times 1} + H_f^{2M \times 3} {}^G\hat{p}_{f_j}^{3 \times 1} + n^{2M \times 1} \quad (32)$$

上式中与该路标点有关的所有视觉约束因子关于观测到的相机位姿的整体 Jacobian $H_X^{2M \times (15+6N)}$ 可表示为下图所示，每行表示一个相机观测到第 j 个路标点产生的视觉观测误差关于整体误差状态向量的 Jacobian，可见每行的矩阵中只有观测到路标点的相机位姿处（图中红色块）有值，其他处均为 0；最后的黄色列表示视觉观测误差关于路标点的世界坐标 ${}^G\hat{p}_{f_j}$ 的 Jacobian。

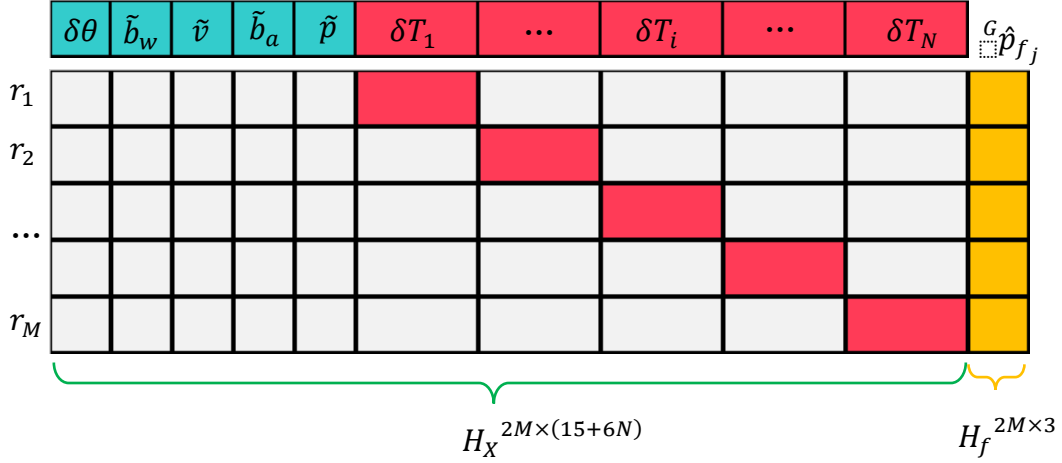


图 8 某个路标点的所有观测所产生的所有重投影误差关于整体误差状态向量和路标点的世界系坐标的 Jacobian

上面给出了当一个路标点看不见时的情况，那么当最新帧有多个路标点看不见时，对应的所有路标点的重投影误差对整体误差状态向量和相应路标点坐标值的 Jacobian 如下所示：

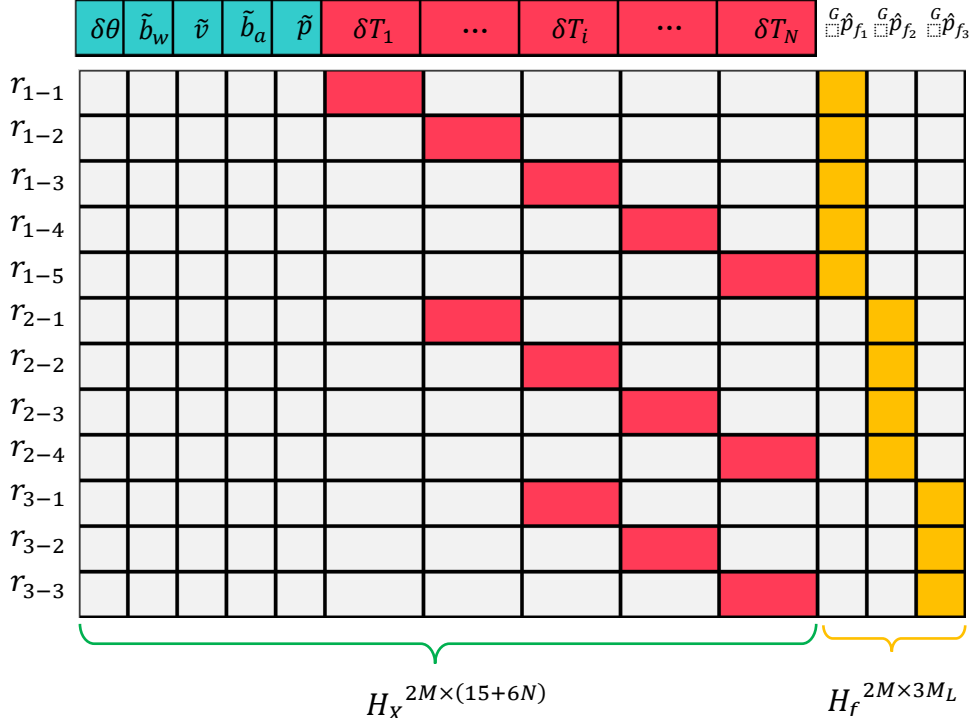


图 9 某帧视觉更新中所有参与边缘化的路标点及其对应的重投影误差

5.3 视觉 Jacobian 处理

对于 EKF，残差线性化需要满足如下形式，即残差仅与一组状态向量的误差项成线性化关系，且噪声项为与状态向量无关的零均值的高斯分布，即残差应能简化成这种形式： $r \cong H\delta x + n$ 。

而 MSCKF 的式(32)的残差与两个状态向量的误差项相关，不满足上式的线性化形式，因此不能直接应用 EKF 的测量更新流程。（附录 7.3 分别给出了两种不同的噪声项与状态向量误差相关性的方式对结果的影响）

为了克服这个问题，我们将原来的残差 $r^{2M \times 1}$ 投影到 $H_f^{2M \times 3M_L}$ 的左零空间（left nullspace）中（左零空间大小为 $2M \times (2M - 3M_L)$ ），并令投影后的残差为 $r_0^{(2M-3M_L) \times 1}$ 。令矩阵 A 为半正交矩阵（即 $A^T A = I$ ），A 的列组成了 H_f 的左零空间，可得：

$$\begin{aligned} r_0^{(2M-3M_L) \times 1} &= A^T r^{2M \times 1} \cong \underbrace{A^T H_x^{2M \times (15+6N)}}_{H_0} \tilde{X}^{(15+6N) \times 1} + \underbrace{A^T n^{2M \times 1}}_{n_0} \\ &= H_0^{(2M-3) \times (15+6N)} \tilde{X}^{(15+6N) \times 1} + n_0^{(2M-3) \times 1} \end{aligned} \quad (34)$$

其中， $A^{2M \times (2M-3)}$ 。值得说明的是，我们不需要计算 A ，而是根据吉文斯旋转

（Givens Rotation）公式可以高效地将向量 r 和矩阵 H_X 投影到 H_f 的左零空间中。

（具体怎么做？）

六、视觉更新

当某个路标点不见或者太老时，即因为遮挡或视野等问题不能被当前帧所观测到，或者该路标点被观测到的共视帧数量大于某个阈值时进行 EKF 更新过程。因为通常 H_0 的矩阵很大，为了降低计算量，我们对 H_0 进行 QR 分解：

$$H_0^{(2M-3) \times (15+6N)} = [Q_1 \quad Q_2] \begin{bmatrix} T_H \\ 0 \end{bmatrix} \quad (35)$$

其中， $Q_1^{(2M-3) \times (15+6N)}$ 、 $Q_2^{(2M-3) \times (15+6N)}$ 为正交矩阵， $T_H^{(15+6N) \times (15+6N)}$ 为上三角矩阵， $0^{(15+6N) \times (15+6N)}$ 。

那么，式(34)可化简为：

$$\begin{bmatrix} Q_1^T r_0 \\ Q_2^T r_0 \end{bmatrix} = \begin{bmatrix} T_H \\ 0 \end{bmatrix} \tilde{X} + \begin{bmatrix} Q_1^T n_0 \\ Q_2^T n_0 \end{bmatrix} \quad (36)$$

我们忽略仅为噪声项的 $Q_2^T r_0$ ，可得最后的观测方程：

$$\begin{aligned} Q_1^T r_0 &= T_H \tilde{X} + Q_1^T n_0 \\ \Rightarrow r_n^{(15+6N) \times 1} &= T_H \tilde{X} + n_n \end{aligned} \quad (37)$$

其中， n_n 的协方差矩阵为： $R_n = Q_1^T R_0 Q_1 = \sigma_{im}^2 I_r$ 。

最后，我们应用 EKF 的更新公式，可得卡尔曼增益为：

$$K = P T_H^T (T_H P T_H^T + R_n)^{-1} \quad (38)$$

则误差状态向量的增量可求解为：

$$\Delta X = K r_n \quad (39)$$

协方差矩阵为：

$$P_{k+1|k+1} = (I_\xi - K T_H) P_{k+1|k} (I_\xi - K T_H)^T + K R_n K^T \quad (40)$$

七、附录

7.1 ESKF 误差运动方程推导中的两处说明

1. 文献[6]中的(232)和(233)解释：

$$a_t = R_t(a_m - a_{bt} - a_n) + g_t \quad (232)$$

$$w_t = w_m - w_{bt} - w_n \quad (233)$$

上式给出了加速度和角速度的测量值与 true-state 的关系。这里，以(233)为例进行说明：

将(233)式整理成：

$$w_m = w_t + w_{bt} + w_n \quad (A1)$$

上式可结合 true-state 物理意义来理解：true-state 表示估计的值，含有估计误差。那么上式可表达为：实际测量得到的陀螺仪值=带误差的估计值+带误差的 bias 偏差+噪声项。

2. 文献[6]中的(236b)和(236c)：

$$\dot{v} = R(a_m - a_b) + g \quad (236b)$$

$$\dot{q} = \frac{1}{2}q \otimes (w_m - w_b) \quad (236c)$$

上式给出了 nominal-state 的速度和旋转的运动方程。仍是以陀螺仪为例，解释(236c)：首先，nominal-state 的物理意义是不含任何误差的理想值。我们从(A1)进行分析，将(A1)中的 true-state 写成 nominal-state+误差项的形式：

$$w_m = (w + \delta w) + (w_b + \delta w_b) + w_n \quad (A2)$$

若将上面状态认为是 nominal-state，则不考虑任何误差和噪声项，即上式中的 δw 、 δw_b 和 w_n ，则上式可写为：

$$w_m = (w + 0) + (w_b + 0) + 0 \quad (A3)$$

$$\Rightarrow w_m = w + w_b \quad (A4)$$

$$\Rightarrow w = w_m - w_b \quad (A5)$$

7.2 增广的误差状态向量与原来误差状态的关系

1. $\delta\theta_{C_{N+1} \leftarrow G}$

首先，将新来的帧的旋转四元数的 true-state，即从世界系到 Cam 系的旋转，转为从世界系到 IMU 系的旋转，再左乘 IMU 到 Cam 的外参：

$${}^C\hat{q} = {}^C\bar{q} \otimes {}^I\hat{q} \quad (A6)$$

将上式写成个人熟悉的形式：

$$R_{C \leftarrow G} = R_{C \leftarrow I} R_{I \leftarrow G} \quad (A7)$$

展开成 $\text{true} = \text{error} + \text{nominal}$ 的形式为:

$$\begin{aligned}
& \exp(\delta\theta_{C \leftarrow G}^\wedge) \hat{R}_{C \leftarrow G} = R_{C \leftarrow I} \exp(\delta\theta_{I \leftarrow G}^\wedge) \hat{R}_{I \leftarrow G} \\
& \Leftrightarrow (I + \delta\theta_{C \leftarrow G}^\wedge) \hat{R}_{C \leftarrow G} = R_{C \leftarrow I} (I + \delta\theta_{I \leftarrow G}^\wedge) \hat{R}_{I \leftarrow G} \\
& \Leftrightarrow \delta\theta_{C \leftarrow G}^\wedge \hat{R}_{C \leftarrow G} = R_{C \leftarrow I} \delta\theta_{I \leftarrow G}^\wedge \hat{R}_{I \leftarrow G} \\
& \Leftrightarrow \delta\theta_{C \leftarrow G}^\wedge = R_{C \leftarrow I} \delta\theta_{I \leftarrow G}^\wedge \hat{R}_{I \leftarrow G} \hat{R}_{C \leftarrow G}^T \\
& \Leftrightarrow \delta\theta_{C \leftarrow G}^\wedge = R_{C \leftarrow I} \delta\theta_{I \leftarrow G}^\wedge R_{C \leftarrow I}^T \\
& \Leftrightarrow \delta\theta_{C \leftarrow G}^\wedge = (R_{C \leftarrow I} \delta\theta_{I \leftarrow G}^\wedge)^\wedge \\
& \Leftrightarrow \delta\theta_{C \leftarrow G} = R_{C \leftarrow I} \delta\theta_{I \leftarrow G} \tag{A8}
\end{aligned}$$

那么, 最新帧的旋转误差相对于原有误差状态向量 (这里只跟 $\delta\theta_{I \leftarrow G}$ 有关)

的 Jacobian 为:

$$\frac{\partial \delta\theta_{C \leftarrow G}}{\partial \delta\theta_{I \leftarrow G}} = R_{C \leftarrow I} = \mathcal{C}(\mathcal{C}_I \bar{q}) \tag{A9}$$

2. ${}^G \tilde{p}_{C_N}$

首先我们写出最新帧的平移量与原有物理量的关系, 即将 Cam 系到 World 系的平移, 转成 IMU 系到 World 系的变换。我们给出 true-state 的形式如下:

$${}^G p_{C_{N+1}} = {}^G p_{I_{N+1}} + {}^{I_{N+1}}_G \bar{q}^T {}^I p_C \tag{A10}$$

根据 $\text{true} = \text{error} + \text{nominal}$, 可将上式展开成:

$$\begin{aligned}
& {}^G \hat{p}_{C_{N+1}} + {}^G \tilde{p}_{C_{N+1}} = ({}^G \hat{p}_{I_{N+1}} + {}^G \tilde{p}_{I_{N+1}}) + [\exp(\delta\theta_{I_{N+1} \leftarrow G}^\wedge) {}^{I_{N+1}}_G \hat{\bar{q}}]^T {}^I p_C \\
& \Leftrightarrow {}^G \hat{p}_{C_{N+1}} + {}^G \tilde{p}_{C_{N+1}} \approx ({}^G \hat{p}_{I_{N+1}} + {}^G \tilde{p}_{I_{N+1}}) + {}^{I_{N+1}}_G \hat{\bar{q}}^T (1 - \delta\theta_{I_{N+1} \leftarrow G}^\wedge) {}^I p_C \\
& \Leftrightarrow {}^G \hat{p}_{C_{N+1}} + {}^G \tilde{p}_{C_{N+1}} \approx ({}^G \hat{p}_{I_{N+1}} + {}^G \tilde{p}_{I_{N+1}}) + {}^{I_{N+1}}_G \hat{\bar{q}}^T {}^I p_C - {}^{I_{N+1}}_G \hat{\bar{q}}^T \delta\theta_{I_{N+1} \leftarrow G}^\wedge {}^I p_C \tag{A11}
\end{aligned}$$

为了快速地处理平移, 我们可以写成向量的形式, 举例说明:

${}^G \hat{r}_{G \rightarrow I} = p_{G \leftarrow I} = {}^G p_I$: 表示 IMU 系到世界系的平移, 也可用在世界系 G 中的从 G 系的坐标原点指向 IMU 局部系 I 的坐标原点的向量 \hat{r} 来表示; 换句话说, ${}^G \hat{r}_{G \rightarrow I}$ 表示 I 系原点在 G 系中的位置, 即表示从 I 系到 G 系的平移 $p_{G \leftarrow I}$, 而 MSCKF 中习惯表达成 ${}^G p_I$ 。

那么, (A11)式可以写成 (我们只重写了三处):

$$\begin{aligned} {}^G\vec{r}_{G \rightarrow C_{N+1}} + {}^G\tilde{p}_{C_{N+1}} \approx \\ {}^G\vec{r}_{G \rightarrow I_{N+1}} + {}^G\tilde{p}_{I_{N+1}} + {}^{I_{N+1}}_{\hat{q}}{}^T {}^I\vec{r}_{I \rightarrow C} - {}^{I_{N+1}}_{\hat{q}}{}^T \delta\theta_{I_{N+1} \leftarrow G} \wedge {}^I p_C \end{aligned} \quad (A12)$$

根据向量加法和旋转法则, 易知等号右边的绿色部分可化简为:

$${}^G\vec{r}_{G \rightarrow I_{N+1}} + {}^{I_{N+1}}_{\hat{q}}{}^T {}^I\vec{r}_{I \rightarrow C} = {}^G\vec{r}_{G \rightarrow I_{N+1}} + {}^G\vec{r}_{I \rightarrow C} = {}^G\vec{r}_{G \rightarrow C_{N+1}} = \text{右边} \quad (A13)$$

将上式代回(A12)可得:

$${}^G\tilde{p}_{C_{N+1}} = {}^G\tilde{p}_{I_{N+1}} + {}^{I_{N+1}}_{\hat{q}}{}^T ({}^I p_C)^\wedge \delta\theta_{I_{N+1} \leftarrow G} \quad (A14)$$

那么, 最新帧的平移误差量, 相对于原有的误差状态量的 Jacobian (这里只跟 ${}^G\tilde{p}_{I_{N+1}}$ 和 $\delta\theta_{I_{N+1} \leftarrow G}$ 有关):

$$\frac{\partial {}^G\tilde{p}_{C_{N+1}}}{\partial \delta\theta_{I_{N+1} \leftarrow G}} = {}^{I_{N+1}}_{\hat{q}}{}^T ({}^I p_C)^\wedge = C_{\hat{q}}^T ({}^I p_C)^\wedge \quad (A15)$$

$$\frac{\partial {}^G\tilde{p}_{C_{N+1}}}{\partial {}^G\tilde{p}_{I_{N+1}}} = I \quad (A16)$$

7.3 噪声项与状态向量的相关性对结果的影响

下图给出了直接将公式(32)用于 EKF 的结果, 对比公式(34)的结果可发现, 当噪声项与状态向量相关时, EKF 估计出来的结果将偏移得更严重^[5]。

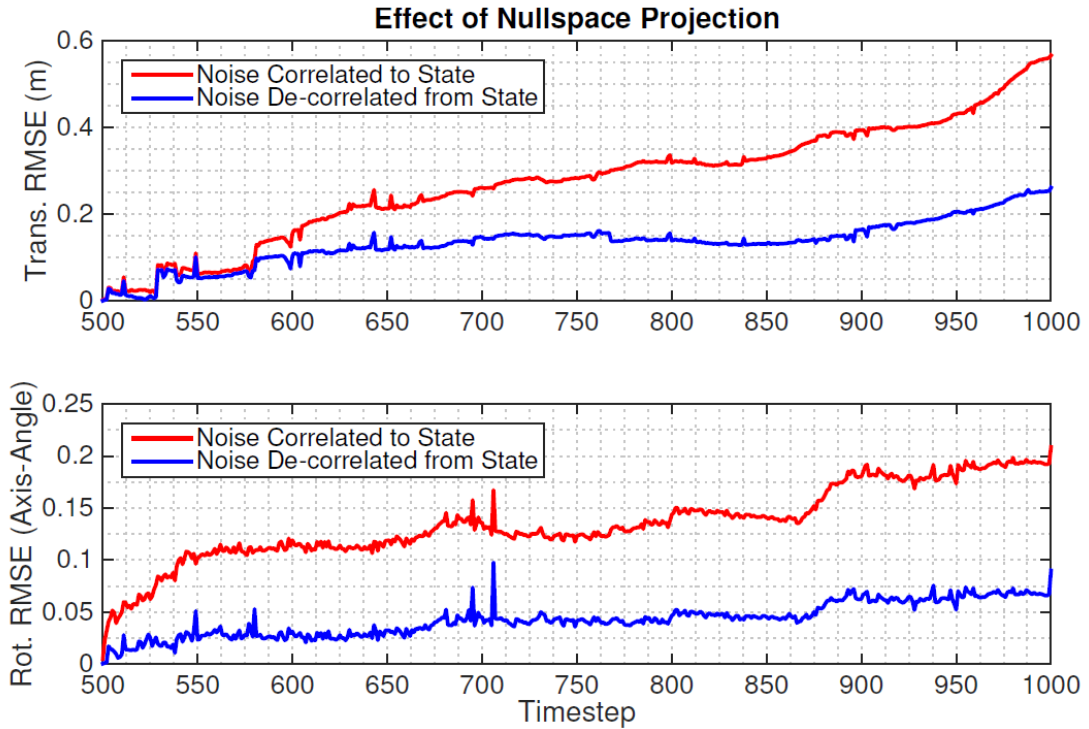


图 5 噪声项与状态向量的相关性对结果的影响

八、参考文献

- [1] A. I. Mourikis. A multi-state constraint kalman filter for vision-aided inertial navigation. Proc. ICRA, 2007.
- [2] A. I. Mourikis. A multi-state constraint kalman filter for vision-aided inertial navigation. 2006. Dept. of Computer Science and Engineering, University of Minnesota, Tech. Rep.
- [3] F. Gonzalez. Visual inertial odometry for mobile robotics. 2015.
- [4] N. Trawny. Indirect kalman filter for 3D attitude estimation. 2005.
- [5] L. Clement. The battle for filter supremacy: a comparative study of the multi-state constraint kalman filter and the sliding window filter. CVR. 2015.
- [6] J. Sola. Quaternion kinematics for the error-state Kalman filter. 2017.