某一时刻，VIO 的状态集合为相机的位姿 $\left\{\mathbf{x}_{\mathbf{p}_j}, 1 \leq j \leq m\right\}$，以及特征点的位姿 $\left\{\mathbf{x}_{\mathbf{m}_i}, 1 \leq i \leq m\right\}$ 令

$$\mathbf{x}=\left[\begin{array}{c}\mathbf{x}_{\mathbf{p}} \\ \mathbf{x}_{\mathbf{m}}\end{array}\right]=\left[\begin{array}{c}\mathbf{x}_{\mathbf{p}_1} \\ \vdots \\ \mathbf{x}_{\mathbf{p}_m} \\ \mathbf{x}_{\mathbf{m}_1} \\ \vdots \\ \mathbf{x}_{\mathbf{m}_n}\end{array}\right] \tag{1}$$

并且有：

$$\mathbf{x}_{\mathbf{p}_j}=f_j\left(\mathbf{x}_{\mathbf{p}_{j-1}}, \mathbf{u}_j\right)+\mathbf{w}_j \tag{2}$$

上式描述了相邻时刻相机的位姿的约束关系，这个约束关系可以从预积分中得到，亦可以从相机的运动模型中得到。$\mathbf{w}_j \sim \mathcal{N}\left(0, \mathbf{Q}_j\right)$ 是高斯白噪声。当 $\mathbf{p}_{j-1}$ 作为先验时，$\mathbf{x}_{\mathbf{p}_j}$ 服从正态分布，$p\left(\mathbf{x}_{\mathbf{p}}\right)=\mathcal{N}\left(\mu_{\mathbf{p}}, \mathbf{Q}\right)$, 其中

$$\mu_{\mathbf{p}}=f(\mathbf{x})=\left[\begin{array}{c}\mathbf{x}_{\mathbf{p}_1} \\ f_1\left(\mathbf{x}_{\mathbf{p}_1}, \mathbf{u}_2\right) \\ \vdots \\ f_m\left(\mathbf{x}_{\mathbf{p}_{m-1}}, \mathbf{u}_m\right)\end{array}\right], \mathbf{Q}=\left[\begin{array}{ccc}\mathbf{Q}_1 & & \\ & \ddots & \\ & & \mathbf{Q}_m\end{array}\right] \tag{3}$$

另外摄像机，加速度计等传感器会提供一些量测信息：

$$\mathbf{z}_{ij}=h_{ij}\left(\mathbf{x}_{\mathbf{m}_i}, \mathbf{x}_{\mathbf{p}_j}\right)+\mathbf{v}_{ij} \tag{4}$$

其中 $\mathbf{v}_{ij} \sim \mathcal{N}\left(0, \mathbf{R}_{ij}\right)$。在基于非线性最小二乘的框架中，记：

$$g(\mathbf{x})=\left[\begin{array}{c}\mathbf{z}-h(\mathbf{x}) \\ \mathbf{x}_{\mathbf{p}}-f\left(\mathbf{x}_{\mathbf{p}}\right)\end{array}\right] \tag{5}$$

优化的目标是使下式最小, 注意 $\mathbf{W}$ 为对称矩阵，即 $\mathbf{W}=\mathbf{W}^T$：

$$\ell(\mathbf{x})=\frac{1}{2}\left(g(\mathbf{x})^T \mathbf{W} g(\mathbf{x})\right), \mathbf{W}=\left[\begin{array}{cc}\mathbf{R}^{-1} & 0 \\ 0 & \mathbf{Q}^{-1}\end{array}\right] \tag{6}$$

在高斯牛顿算法的框架下，在每一次迭代时，目标是求解 h 使得下式最小：

$$\ell(\mathbf{x}+\mathbf{h})=\frac{1}{2}(g(\mathbf{x})+\mathbf{J}_{\mathbf{g}} \mathbf{h})^T \mathbf{W}(g(\mathbf{x})+\mathbf{J}_{\mathbf{g}} \mathbf{h}) \tag{7}$$

其中 $\mathbf{J_g h} = \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}$，

$$\frac{\partial \ell(\mathbf{x}+\mathbf{h})}{\partial \mathbf{x}} = \frac{1}{2} g(\mathbf{x})^T \mathbf{W} g(\mathbf{x}) + g(\mathbf{x})^T \mathbf{W J_g h} + \frac{1}{2} \mathbf{h}^T \mathbf{J_g}^T \mathbf{W J_g h} \tag{8}$$

$$\frac{\partial^2 \ell(\mathbf{x}+\mathbf{h})}{\partial \mathbf{x^2}} = \mathbf{J_g}^T W g(\mathbf{x}) + \mathbf{J_g}^T \mathbf{W J_g h} \tag{9}$$

令上式等于 $\mathbf{0}$，并令 $\mathbf{H} = \mathbf{J_g}^T \mathbf{W J_g}$，得到正规方程 (Normal Equation):

$$\mathbf{Hh} = -\mathbf{J_g}^\mathbf{T} \mathbf{W g}(\mathbf{x}) \tag{10}$$

将 $\mathbf{H}$ 展开：

$$\mathbf{H} = \begin{bmatrix} \mathbf{J_f} \\ \mathbf{J_h} \end{bmatrix}^T \begin{bmatrix} \mathbf{R}^{-1} & 0 \\ 0 & \mathbf{Q}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{J_f} \\ \mathbf{J_h} \end{bmatrix} \tag{11}$$

$$= \mathbf{J_f R^{-1} J_f} + \mathbf{J_g Q^{-1} J_g} \tag{12}$$

$$= \begin{bmatrix} \mathbf{\Lambda_p} & \mathbf{\Lambda_{pm}} \\ \mathbf{\Lambda_{pm}^T} & \mathbf{\Lambda_m} \end{bmatrix} \tag{13}$$

因此正规方程可写成以下的形式：

$$\begin{bmatrix} \mathbf{\Lambda_p} & \mathbf{\Lambda_{pm}} \\ \mathbf{\Lambda_{pm}^T} & \mathbf{\Lambda_m} \end{bmatrix} \begin{bmatrix} \delta \mathbf{x_p} \\ \delta \mathbf{x_m} \end{bmatrix} = \begin{bmatrix} \mathbf{b_p} \\ \mathbf{b_m} \end{bmatrix} \tag{14}$$

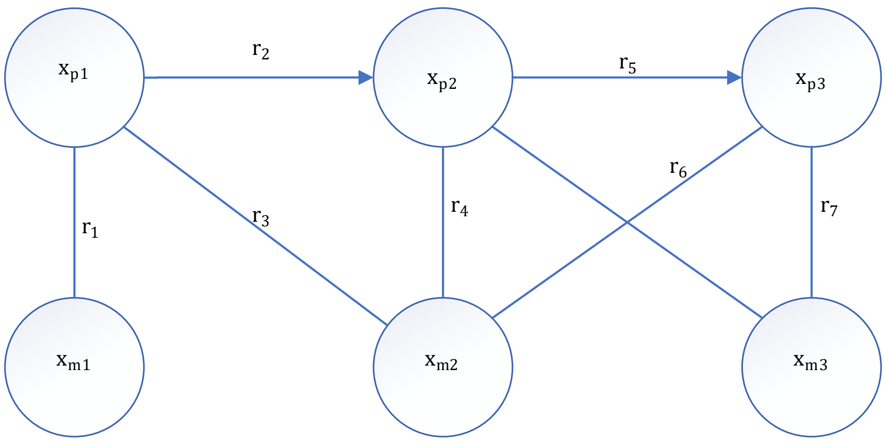

图 1　一个 VIO 的简单例子，图中的节点表示待优化的变量，边表示节点之间的误差项，第一时刻的相机 $\mathbf{x}_{p1}$ 能观测到特征点 $\mathbf{x}_{m1}$，在第二时刻的相机 $\mathbf{x}_{p2}$ 能观测到特征点 $\mathbf{x}_{m2}$ 和特征点 $\mathbf{x}_{m3}$，在第三时刻的相机 $\mathbf{x}_{p3}$ 能观测到特征点 $\mathbf{x}_{m2}$ 和特征点 $\mathbf{x}_{m3}$，这些观测对应 $\mathbf{r}_1 \cdots \mathbf{r}_7$ 共七组误差。注意

在图1中给出了一个 VIO 的简单例子，第一时刻的相机 $\mathbf{x}_{p1}$ 能观测到特征点 $\mathbf{x}_{m1}$，在第二时刻的相机 $\mathbf{x}_{p2}$ 能观测到特征点 $\mathbf{x}_{m2}$ 和特征点 $\mathbf{x}_{m3}$，在第三时刻的相机 $\mathbf{x}_{p3}$ 能

观测到特征点 $\mathbf{x}_{m2}$ 和特征点 $\mathbf{x}_{m3}$, 状态变量依次为 $\mathbf{x}_{p1}, \mathbf{x}_{p2}, \mathbf{x}_{p3}, \mathbf{x}_{m1}, \mathbf{x}_{m2}, \mathbf{x}_{m3}$，其中共有七条边，对应误差为 $\mathbf{r}_1 \cdots \mathbf{r}_7$，$\mathbf{r}_i, 1 \le i \le 7$ 对应的信息矩阵为 $\mathbf{\Lambda}_i, 1 \le i \le 7$, 其中，$\mathbf{\Lambda}_i$ 为：

$$\mathbf{\Lambda}_1 = \begin{bmatrix} {\mathbf{J}^{r1}_{\mathbf{x}_{p1}}}^\top \mathbf{W}_{\mathbf{r_1}} \mathbf{J}^{r1}_{\mathbf{x}_{p1}} & 0 & 0 & {\mathbf{J}^{r1}_{\mathbf{x}_{p1}}}^\top \mathbf{W}_{\mathbf{r_1}} \mathbf{J}^{r1}_{\mathbf{x}_{m1}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ {\mathbf{J}^{r1}_{\mathbf{x}_{m1}}}^\top \mathbf{W}_{\mathbf{r_1}} \mathbf{J}^{r1}_{\mathbf{x}_{p1}} & 0 & 0 & {\mathbf{J}^{r1}_{\mathbf{x}_{m1}}}^\top \mathbf{W}_{\mathbf{r_1}} \mathbf{J}^{r1}_{\mathbf{x}_{m1}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{15}$$

$$\cdots \tag{16}$$

$$\mathbf{\Lambda}_7 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & {\mathbf{J}^{r7}_{\mathbf{x}_{p3}}}^\top \mathbf{W}_{\mathbf{r_7}} \mathbf{J}^{r7}_{\mathbf{x}_{p3}} & 0 & 0 & {\mathbf{J}^{r7}_{\mathbf{x}_{p3}}}^\top \mathbf{W}_{\mathbf{r_7}} \mathbf{J}^{r7}_{\mathbf{x}_{m3}} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & {\mathbf{J}^{r7}_{\mathbf{x}_{m3}}}^\top \mathbf{W}_{\mathbf{r_7}} \mathbf{J}^{r7}_{\mathbf{x}_{p1}} & 0 & 0 & {\mathbf{J}^{r7}_{\mathbf{x}_{m3}}}^\top \mathbf{W}_{\mathbf{r_7}} \mathbf{J}^{r7}_{\mathbf{x}_{m3}} \end{bmatrix} \tag{17}$$

其中 $\mathbf{J}^{r_i}_{\mathbf{x}_j}$ 为 r_i 对 $\mathbf{x}_j$ 的雅可比，$\mathbf{W}_{\mathbf{r_i}}$ 为残差项对应的权重，该权重根据实际情况通常设置为对应残差的协方差的逆。总体的信息矩阵为：$\mathbf{\Lambda} = \sum_{i=1}^{7} \mathbf{\Lambda_i}$。在图**??** 中给出了这个信息矩阵累加的过程，在图中可以看到，每一个残差项提供的信息矩阵都是稀疏的，最终累加得到的稀疏矩阵 $\mathbf{\Lambda}$ 也是稀疏的。

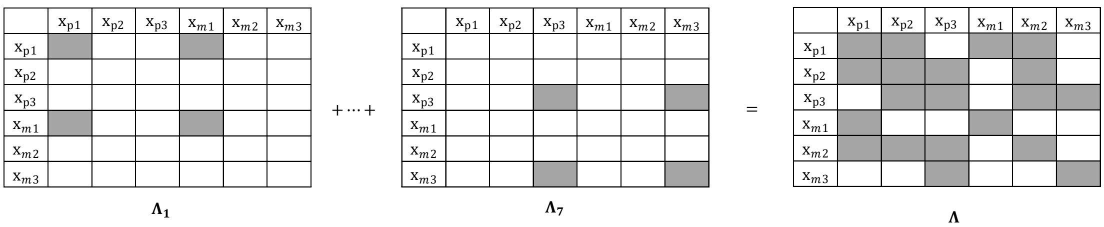

图 2　信息矩阵累加过程示意图，最后的信息矩阵 $\mathbf{\Lambda}$ 由每一个小的残差项的信息矩阵 $\mathbf{\Lambda}_i$ 累加而成，每一个小的残差项的信息矩阵 $\mathbf{\Lambda}_i$ 是稀疏的，最后的信息矩阵 $\mathbf{\Lambda}$ 也是稀疏的。$\mathbf{\Lambda}$ 的稀疏性主要有两点：(1). 它的右下角子矩阵只在对角线上有元素; (2). 它的右上角矩阵和左下角矩阵互为转置关系

对于一个基于图优化的视觉里程计系统来说，优化的大部分时间被消耗在了信息矩阵的维护和求逆上面。首先由于信息矩阵的右上角子矩阵和左下角子矩阵互为转置关系，因此可以使用舒尔补加速矩阵的求解。另外，由于假设特征点之间是独立的，对其中一个特征点的观测并不会影响对其它特征点的观测，因此信息矩阵的右下角子

矩阵的对角上的子矩阵都为对角矩阵。

根据信息矩阵的特点，将其分为四个子矩阵：

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_p & \mathbf{\Lambda}_{pm} \\ \mathbf{\Lambda}_{mp} & \mathbf{\Lambda}_m \end{bmatrix} \tag{18}$$

利用 $\mathbf{H}$ 的性质，可以用舒尔补加速正规方程的求解：

$$\begin{bmatrix} \mathbf{\Lambda_p} - \mathbf{\Lambda_{pm}} \left(\mathbf{\Lambda_m}\right)^{-1} \mathbf{\Lambda_{pm}^T} & 0 \\ \mathbf{\Lambda}_{\mathbf{pm}}^{T} & \mathbf{\Lambda_m} \end{bmatrix} \begin{bmatrix} \delta \mathbf{x_p} \\ \delta \mathbf{x_m} \end{bmatrix} = \begin{bmatrix} \mathbf{g_p} - \mathbf{\Lambda_{pm}} \left(\mathbf{\Lambda_m}\right)^{\mathbf{-1}} \mathbf{g_m} \\ \mathbf{g_m} \end{bmatrix} \tag{19}$$

可以得到：

$$\begin{aligned} \delta \mathbf{x_p} &= \left(\Lambda_p - \mathbf{\Lambda_{pm}} \left(\mathbf{\Lambda_m}\right)^{\mathbf{-1}} \mathbf{\Lambda_{pm}^T}\right)^{\mathbf{-1}} \left(\mathbf{g_p} - \mathbf{\Lambda_{pm}} \left(\mathbf{\Lambda_m}\right)^{\mathbf{-1}} \mathbf{g_m}\right) \\ \delta \mathbf{x_m} &= \quad \left(\mathbf{\Lambda_m}\right)^{\mathbf{-1}} \left(\mathbf{g_m} - \mathbf{\Lambda_{pm}^T} \delta \mathbf{x_p}\right) \end{aligned} \tag{20}$$

根据前文的描述，$\mathbf{\Lambda}_m$ 只在对角线附近有元素，因此 $\mathbf{\Lambda}_m^{-1}$ 的计算的时间复杂度可以从 $O(n^3)$ 降低到 $O(n)$，而在基于 SLAM 系统中，特征点通常非常多，从数千个到数万个不等，需要优化的位姿通常较少，因此 SLAM 的稀疏性可以确保整个问题是可以被优化的。