视觉SLAM十四讲|第10讲 后端1

咸鱼了好久。。。。起来接着学习。

上来就又被先验后验搞蒙了，重新理解一下：

> 先验概率：根据以往经验和分析得到的概率，它往往作为“由因求果”问题中的“因”出现。
> 后验概率：指在得到“结果”的信息后重新修正的概率，是“执果寻因”问题中的“因”。

最大的区别就在于结果/当前观测在后验概率中已知，在先验概率中未知。

1. 后端要做什么?

前端能够根据相邻的两幅图像判断出此时此刻的位姿，是暂时的；那么后端需要对前端测量以及计算的结果进行矫正，不仅用过去的信息，也用未来的信息更新自己，希望能够得到一个长时间的正确状态。

前面有讲过，在运动方程和观测方程中，如果把位姿和路标看成随机变量，就可以把问题变为已知观测数据和运动数据情况下，如何确定状态量的分布问题，是一个状态估计问题。假设噪声和状态量服从高斯分布，只需要估计它的均值和方差，即可确定状态量的分布。

完成后端优化可以使用滤波器，也可以使用非线性优化，本讲对两种方法都做了推导。

2. 滤波器推导

先做一些符号的定义和铺垫：

把位姿和路标写在一起，记为： $\boldsymbol{x}_k \triangleq \{\boldsymbol{x}_k, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\}$ ，用新符号写运动方程和观测方程，为： $\begin{cases} \boldsymbol{x}_k = f(\boldsymbol{x}_{k-1}, \boldsymbol{u}_k) + \boldsymbol{w}_k \\ \boldsymbol{z}_k = h(\boldsymbol{x}_k) + \boldsymbol{v}_k \end{cases} \quad k = 1, \ldots, N$ 。

需要求解的是后验概率问题，即在已知0时刻的状态、1:k时刻的观测下，k时刻的状态分布，写为 $P(\boldsymbol{x}_k | \boldsymbol{x}_0, \boldsymbol{u}_{1:k}, \boldsymbol{z}_{1:k})$ ，根据贝叶斯法则，将后验概率展开为似然和先验概率的乘积，写为 $P(\boldsymbol{x}_k | \boldsymbol{x}_0, \boldsymbol{u}_{1:k}, \boldsymbol{z}_{1:k}) \propto P(\boldsymbol{z}_k | \boldsymbol{x}_k) P(\boldsymbol{x}_k | \boldsymbol{x}_0, \boldsymbol{u}_{1:k}, \boldsymbol{z}_{1:k-1})$ 。

先验概率中k时刻的状态 $\boldsymbol{x}_k$ 受到前面0:k-1时刻状态的影响，对先验概率进行展开，可以写为 $P(\boldsymbol{x}_k | \boldsymbol{x}_0, \boldsymbol{u}_{1:k}, \boldsymbol{z}_{1:k-1}) = \int P(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}, \boldsymbol{x}_0, \boldsymbol{u}_{1:k}, \boldsymbol{z}_{1:k-1}) P(\boldsymbol{x}_{k-1} | \boldsymbol{x}_0, \boldsymbol{u}_{1:k}, \boldsymbol{z}_{1:k-1}) \mathrm{d}\boldsymbol{x}_{k-1}$ ，在进行滤波器模型推导时，假设马尔可夫性（k时刻状态只与k-1时刻状态有关，与之前状态无关）。可以根据马尔可夫性对先验概率的展开式进行化简，积分中第一个概率化简为

$P\left(\boldsymbol{x}_k|\boldsymbol{x}_{k-1},\boldsymbol{x}_0,\boldsymbol{u}_{1:k},\boldsymbol{z}_{1:k-1}\right)=P\left(\boldsymbol{x}_k|\boldsymbol{x}_{k-1},\boldsymbol{u}_k\right)$ ，第二个概率化简为 $P\left(\boldsymbol{x}_{k-1}|\boldsymbol{x}_0,\boldsymbol{u}_{1:k},\boldsymbol{z}_{1:k-1}\right)=P\left(\boldsymbol{x}_{k-1}|\boldsymbol{x}_0,\boldsymbol{u}_{1:k-1},\boldsymbol{z}_{1:k-1}\right)$ （运动方程中k时刻状态只受到k-1时刻的状态和k时刻的运动数据相关）。

2.1 卡尔曼滤波器（线性高斯系统）

在线性高斯系统中，用线性方程描述运动方程和观测方程，写为 $\begin{cases}\boldsymbol{x}_k=\boldsymbol{A}_k\boldsymbol{x}_{k-1}+\boldsymbol{u}_k+\boldsymbol{w}_k\\ \boldsymbol{z}_k=\boldsymbol{C}_k\boldsymbol{x}_k+\boldsymbol{v}_k\end{cases}\quad k=1,\ldots,N$ ，记噪声服从零均值高斯分布：$\boldsymbol{w}_k\sim N(\mathbf{0},\boldsymbol{R}).\quad \boldsymbol{v}_k\sim N(\mathbf{0},\boldsymbol{Q})$ ，令 $\hat{\boldsymbol{x}}_k$ 表示后验，$\overline{x}$ 表示先验。

因为各随机变量服从高斯分布，误差服从零均值高斯分布，根据高斯分布的特性（后面解释）和观测方程、运动方程，可以分别写出似然和先验的分布：

$$P\left(\boldsymbol{x}_k|\boldsymbol{x}_0,\boldsymbol{u}_{1:k},\boldsymbol{z}_{1:k-1}\right)=N\left(\boldsymbol{A}_k\hat{\boldsymbol{x}}_{k-1}+\boldsymbol{u}_k,\boldsymbol{A}_k\hat{\boldsymbol{P}}_{k-1}\boldsymbol{A}_k^T+\boldsymbol{R}\right)$$

$$P\left(\boldsymbol{z}_k|\boldsymbol{x}_k\right)=N\left(\boldsymbol{C}_k\boldsymbol{x}_k,\boldsymbol{Q}\right)$$

把先验均值和方差记为 $\overline{\boldsymbol{x}}_k=\boldsymbol{A}_k\hat{\boldsymbol{x}}_{k-1}+\boldsymbol{u}_k,\quad \overline{\boldsymbol{P}}_k=\boldsymbol{A}_k\hat{\boldsymbol{P}}_{k-1}\boldsymbol{A}_k^T+\boldsymbol{R}$

根据后验、似然、先验的关系，有：$N\left(\hat{\boldsymbol{x}}_k,\hat{\boldsymbol{P}}_k\right)=N\left(\boldsymbol{C}_k\boldsymbol{x}_k,\boldsymbol{Q}\right)\cdot N\left(\overline{\boldsymbol{x}}_k,\overline{\boldsymbol{P}}_k\right)$

高斯分布底数为e，比较等式两边的指数项（不考虑常数项）有，
$\left(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_k\right)^T\hat{\boldsymbol{P}}_k^{-1}\left(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_k\right)=\left(\boldsymbol{z}_k-\boldsymbol{C}_k\boldsymbol{x}_k\right)^T\boldsymbol{Q}^{-1}\left(\boldsymbol{z}_k-\boldsymbol{C}_k\boldsymbol{x}_k\right)+\left(\boldsymbol{x}_k-\overline{\boldsymbol{x}}_k\right)^T\overline{\boldsymbol{P}}_k^{-1}\left(\boldsymbol{x}_k-\overline{\boldsymbol{x}}_k\right)$

比较两边的系数，经过一系列推导可以得到后验均值和方差的表达式。

整个卡尔曼滤波的过程为：

1. 预测：

$$\bar{\boldsymbol{x}}_k = \boldsymbol{A}_k \hat{\boldsymbol{x}}_{k-1} + \boldsymbol{u}_k, \quad \bar{\boldsymbol{P}}_k = \boldsymbol{A}_k \hat{\boldsymbol{P}}_{k-1} \boldsymbol{A}_k^T + \boldsymbol{R}. \tag{10.24}$$

2. 更新：先计算 $\boldsymbol{K}$，它又称为卡尔曼增益：

$$\boldsymbol{K} = \bar{\boldsymbol{P}}_k \boldsymbol{C}_k^T \left(\boldsymbol{C}_k \bar{\boldsymbol{P}}_k \boldsymbol{C}_k^T + \boldsymbol{Q}\right)^{-1}. \tag{10.25}$$

然后计算后验概率的分布：

$$\begin{aligned} \hat{\boldsymbol{x}}_k &= \bar{\boldsymbol{x}}_k + \boldsymbol{K}\left(\boldsymbol{z}_k - \boldsymbol{C}_k \bar{\boldsymbol{x}}_k\right) \\ \hat{\boldsymbol{P}}_k &= \left(\boldsymbol{I} - \boldsymbol{K}\boldsymbol{C}_k\right) \bar{\boldsymbol{P}}_k. \end{aligned} \tag{10.26}$$

在第一步预测中，使用k-1时刻的后验分布估计k时刻的先验分布，这一步不确定性变大，第二步使用k时刻的先验分布估计k时刻的后验分布，对于结果进行修正，缩小不确定性。

2.2 拓展卡尔曼滤波器EKF（非线性系统）

将线性系统拓展到非线性系统，通常会在某个点处对方程进行展开，保留一阶项，按照线性系统进行推导。

对于k时刻的运动方程，展开有：
$\boldsymbol{x}_k \approx f\left(\hat{\boldsymbol{x}}_{k-1}, \boldsymbol{u}_k\right) + \left.\frac{\partial f}{\partial \boldsymbol{x}_{k-1}}\right|_{\hat{\boldsymbol{x}}_{k-1}} \left(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1}\right) + \boldsymbol{w}_k$ ，记偏导数为 $\boldsymbol{F} = \left.\frac{\partial f}{\partial \boldsymbol{x}_{k-1}}\right|_{\hat{\boldsymbol{x}}_{k-1}}$ 。

对于k时刻的观测方程，展开有： $z_k \approx h\left(\overline{\boldsymbol{x}}_k\right) + \left.\frac{\partial h}{\partial \boldsymbol{x}_k}\right|_{\overline{\boldsymbol{x}}_k} \left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k\right) + \boldsymbol{n}_k$ ，记偏导数为 $\boldsymbol{H} = \left.\frac{\partial h}{\partial \boldsymbol{x}_k}\right|_{\hat{\boldsymbol{x}}_k}$ 。

利用和线性系统类似的方法，推导得到预测步骤（没看明白怎么推的，如有大神请帮忙解释一下）： $\overline{\boldsymbol{x}}_k = f\left(\hat{\boldsymbol{x}}_{k-1}, \boldsymbol{u}_k\right), \quad \overline{\boldsymbol{P}}_k = \boldsymbol{F}\hat{\boldsymbol{P}}_k \boldsymbol{F}^T + \boldsymbol{R}_k$

卡尔曼增益： $\boldsymbol{K}_k = \overline{\boldsymbol{P}}_k \boldsymbol{H}^{\mathrm{T}} \left(\boldsymbol{H}\overline{\boldsymbol{P}}_k \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{Q}_k\right)^{-1}$

更新步骤： $\hat{\boldsymbol{x}}_k = \overline{\boldsymbol{x}}_k + \boldsymbol{K}_k\left(\boldsymbol{z}_k - h\left(\overline{\boldsymbol{x}}_k\right)\right), \hat{\boldsymbol{P}}_k = \left(\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}\right)\overline{\boldsymbol{P}}_k$

EKF的局限性：马尔可夫假设、一次线性化，这两个数学上的处理化简了问题，但是偏离了实际情况。记录均值、方差、路标等信息，存储量大。

3. 非线性优化（BA与图优化）

> Bundle Adjustment，是指从视觉重建中提炼出最优的3D模型和相机参数（内参数和外参数）。

3.1 优化目标

回顾整个投影过程，假设已知相机的外参数（R，t）和世界坐标系的p点坐标，求解p点的像素坐标过程。

世界坐标转化为相机坐标： $\boldsymbol{P}' = \boldsymbol{R}\boldsymbol{p} + \boldsymbol{t} = [X', Y', Z']^T$

相机坐标投影到归一化平面： $\boldsymbol{P}_c = [u_c, v_c, 1]^T = [X'/Z', Y'/Z', 1]^T$

去畸变（径向畸变）： $\begin{cases} u_c' = u_c\left(1 + k_1 r_c^2 + k_2 r_c^4\right) \\ v_c' = v_c\left(1 + k_1 r_c^2 + k_2 r_c^4\right) \end{cases}$

计算像素坐标： $\begin{cases} u_s = f_x u_c' + c_x \\ v_s = f_y v_c' + c_y \end{cases}$

整个过程中，用到了相机的位姿（R，t）和路标p的世界坐标，得到了路标的像素坐标，对应到观测方程z=h(x,y)中，观测z就是像素坐标 $[u_s, v_s]^T$ ，相机位姿x用李代数表示 ξ ，路标y用三维点p的坐标。那么通过降低观测到的像素坐标z和估计到的像素坐标 $h(\boldsymbol{\xi}, \boldsymbol{p})$ 之间的误差，就可以得到最优的x和y的解。用最小二乘表示整体的代价函数为：

$$\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\left\|e_{ij}\right\|^2 = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\left\|z_{ij} - h\left(\boldsymbol{\xi}_i, \boldsymbol{p}_j\right)\right\|^2$$

采用非线性优化更新增量的方式求解，令 $\boldsymbol{x} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_n]^T$ ， $f(x) = \boldsymbol{z}_{ij} - h\left(\boldsymbol{\xi}_i, \boldsymbol{p}_j\right)$ 。

引入增量后，目标函数变为：

$$\frac{1}{2}\left\|f(\boldsymbol{x} + \Delta\boldsymbol{x})\right\|^2 \approx \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\left\|\boldsymbol{e}_{ij} + \boldsymbol{F}_{ij}\Delta\boldsymbol{\xi}_i + \boldsymbol{E}_{ij}\Delta\boldsymbol{p}_j\right\|^2$$

$\boldsymbol{F}_{ij}$ 表示f(x)对于相机位姿的偏导数， $\boldsymbol{E}_{ij}$ 表示f(x)对于路标点位置的偏导数。

接下来使用第六章的非线性优化方法进行求解，在求解时不免要计算 $H\Delta x = g$ ，利用H矩阵的稀疏性能够加速求解。

3.2 稀疏性

在高斯牛顿法中 $\boldsymbol{H} = \boldsymbol{J}^T\boldsymbol{J}$ ，在列文伯格——马夸尔特方法中 $\boldsymbol{H} = \boldsymbol{J}^T\boldsymbol{J} + \lambda\boldsymbol{I}$ ，所以H的稀疏性主要是由J引起的。

对于代价函数中的 e_{ij} ，描述的是在 $\boldsymbol{\xi}_i$ 看到 $\boldsymbol{p}_j$ ，只与这两个量相关，所以

$$\boldsymbol{J}_{ij}(\boldsymbol{x}) = \left(\boldsymbol{0}_{2\times6}, \ldots \boldsymbol{0}_{2\times6}, \frac{\partial \boldsymbol{e}_{ij}}{\partial \boldsymbol{\xi}_i}, \boldsymbol{0}_{2\times6}, \ldots \boldsymbol{0}_{2\times3}, \ldots \boldsymbol{0}_{2\times3}, \frac{\partial \boldsymbol{e}_{ij}}{\partial \boldsymbol{p}_j}, \boldsymbol{0}_{2\times3}, \ldots \boldsymbol{0}_{2\times3}\right)$$

由于 $\boldsymbol{H} = \boldsymbol{J}^T\boldsymbol{J}$ ，J只在i和j处有非零块，所以矩阵H中只有(i,i),(i,j),(j,i),(j,j)处有非零块，其中i是相机部分，j是路标部分。

举例说明，假设有2个相机位姿 (C_1, C_2) 和6个路标 $(P_1, P_2, P_3, P_4, P_5, P_6)$ ，那么相机所对应的变量有 $\boldsymbol{\xi}_i, i = 1, 2$ ，路标对应变量有 $\boldsymbol{p}_j, j = 1, \ldots, 6$ 。如图10-4所示，连线表示相机可以看到路标。

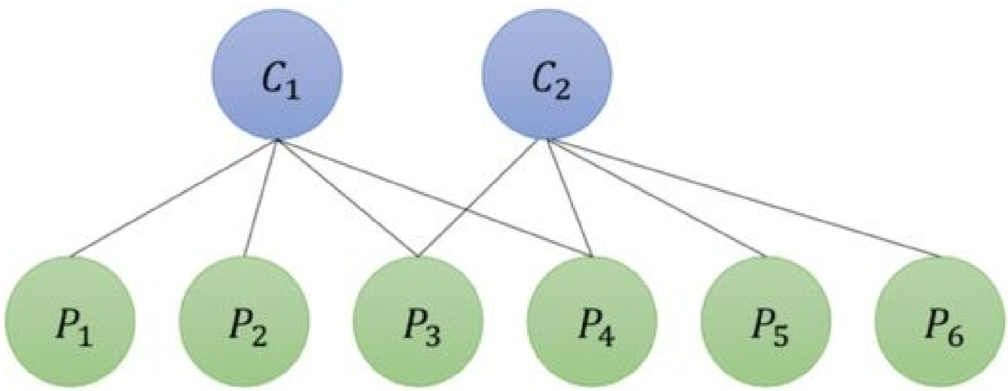

图 10-4　点和边组成的示意图。该图显示相机 C_1 观测到了路标点 P_1, P_2, P_3, P_4，相机 C_2 看到了 P_3 到 P_6。

图10-4的场景，用目标函数表示应该是:

$$\frac{1}{2}\left(\|e_{11}\|^2 + \|e_{12}\|^2 + \|e_{13}\|^2 + \|e_{14}\|^2 + \|e_{23}\|^2 + \|e_{24}\|^2 + \|e_{25}\|^2 + \|e_{26}\|^2\right)$$

其中 e_{11} 描述的是相机 C_1 观测到了 p_1 ，把所有变量按照 $\boldsymbol{x} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_2)^T$ 的顺序摆放，则雅各比矩阵为

$$J_{11} = \frac{\partial e_{11}}{\partial \boldsymbol{\xi}_1} = \left(\frac{\partial e_{11}}{\partial \xi_1}, \boldsymbol{0}_{2\times6}, \frac{\partial e_{11}}{\partial \boldsymbol{p}_1}, \boldsymbol{0}_{2\times3}, \boldsymbol{0}_{2\times3}, \boldsymbol{0}_{2\times3}, \boldsymbol{0}_{2\times3}, \boldsymbol{0}_{2\times3}\right)$$

，依次推类可以得到目标函数每一项的雅各比矩阵，叠放在一起，用方块表示非零块，则矩阵J和H的非零值如图10-6所示。

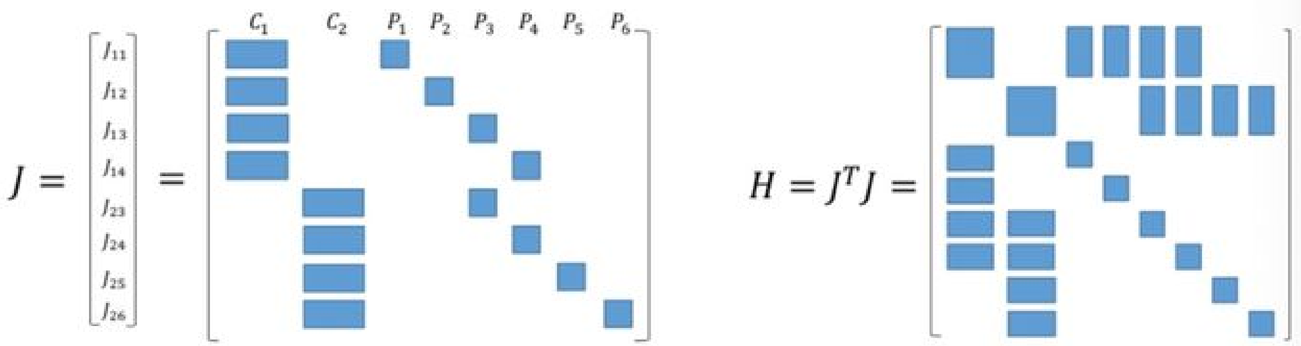

图 10-6　Jacobian 矩阵的稀疏性 (左) 和 $\boldsymbol{H}$ 矩阵的稀疏性 (右)，蓝色的方块表示矩阵在对应的矩阵块处有数值，其余没有颜色的部分表示矩阵在该处的数值始终为 0。知乎 @笨丫头

每个非零块都对应着相机和路标点的关系，拓展开来，可以把H矩阵划分成四个区域，如图10-9所示。

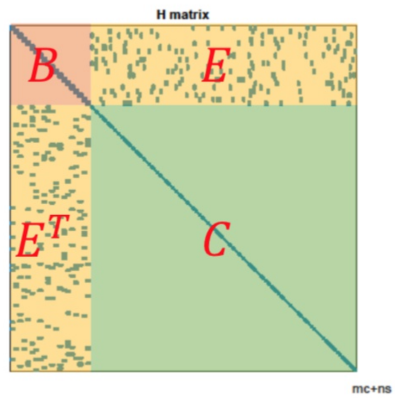

图 10-9　$\boldsymbol{H}$ 矩阵的区域划分。 知乎 @笨丫头

其中BC是对角块矩阵，B的对角块与相机位姿的维度相同，C的对角块与路标维度相同，E的数值和具体的观测相关。

3.3 边缘化（Marginalization）

把相机位姿变量放到一起，记为 $\boldsymbol{x}_c = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_m]^T \in \mathbb{R}^{6m}$ ，把空间点的变量放在一起，记为 $\boldsymbol{x}_p = [\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n]^T \in \mathbb{R}^{3n}$ 。

在非线性优化中，需要根据 $H\Delta x = g$ 求解增量，用分块矩阵表示H后变为：$\begin{bmatrix} \boldsymbol{B} & \boldsymbol{E} \\ \boldsymbol{E}^T & \boldsymbol{C} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{x}_c \\ \Delta \boldsymbol{x}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{w} \end{bmatrix}$ ，需要想办法计算两个增量。

先消去右上角的E：$\begin{bmatrix} I & -EC^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} B & E \\ E^T & C \end{bmatrix} \begin{bmatrix} \Delta x_c \\ \Delta x_p \end{bmatrix} = \begin{bmatrix} I & -EC^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix}$ ，

整理得到 [公式] ，

用第一行计算出 $\Delta \boldsymbol{x}_c$ 的值，再带入第二行，求解 $\Delta \boldsymbol{x}_p$ ，这个过程称为Marginalization。

3.4 鲁棒核函数

当输入的数据存在无匹配时，误差会很大，在二范式中产生的梯度也会很大，从而导致增量朝着错误的方向变化。选用一个具有光滑性质并且变化较小的函数可以使得系统更稳健，这种函数称为鲁棒核函数，比如Huber核： $H(e)=\begin{cases}\frac{1}{2}e^2 & \text{if } |e|\le\delta \\ \delta\left(|e|-\frac{1}{2}\delta\right) & \text{otherwise}\end{cases}$ 。

附录A.3

高斯函数性质：

对于随机变量 $\boldsymbol{x}\sim N\left(\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_{xx}\right)$ ，若有 $\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}+\boldsymbol{w}$ ，其中A，b为线性变量的稀疏矩阵核偏移量，w为噪声项，且 $w\sim N(\mathbf{0},\boldsymbol{R})$ ，有
$p(\boldsymbol{y})=N\left(\boldsymbol{A}\boldsymbol{\mu}_x+\boldsymbol{b},\boldsymbol{A}\boldsymbol{\Sigma}_{xx}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{R}\right)$ 。

参考文献

[1] 《视觉SLAM十四讲从理论到实践》 高翔，张涛