



机器人学中的状态估计

Timothy Barfoot 著 高翔, 谢晓佳等 译

Slides by Xiang Gao

2020年春



第4讲 偏差、匹配和外点

- 输入和测量的偏差
- 数据关联
- 外点



第4讲 偏差、匹配和外点

- 输入和测量的偏差
- 数据关联
- 外点



输入和测量的偏差

- 在上两节课我们讨论了LG系统与NLNG系统中的状态估计问题
- 实际问题与理论问题还存在各种可能的偏差：
 - 实际噪声并不是零均值高斯噪声
 - 实际当中可能存在错误观测
 - 观测数据还存在匹配问题（数据关联）
- 这些都可能状态估计与实际相差甚远
- 本节课主要讨论这几类问题的处理方式



输入和测量的偏差

- 首先考虑LG系统下，运动和观测存在偏差的情况。系统方程为：

$$x_k = Ax_{k-1} + B(u_k + \bar{u}) + w_k$$

$$y_k = Cx_k + \bar{y} + n_k$$

其中 \bar{u} \bar{y} 是输入与观测量的偏差， w 和 n 仍为白噪声

$$w_k \sim \mathcal{N}(0, Q), \quad n_k \sim \mathcal{N}(0, R)$$

- 在LG系统中，我们使用卡尔曼滤波器获得状态估计：

预测：

$$\check{P}_k = A_{k-1} \hat{P}_{k-1} A_{k-1}^T + Q_k$$

$$\check{x}_k = A_{k-1} \hat{x}_{k-1} + v_k$$

卡尔曼增益：

$$K_k = \check{P}_k C_k^T (C_k \check{P}_k C_k^T + R_k)^{-1}$$

更新：

$$\hat{P}_k = (1 - K_k C_k) \check{P}_k$$

$$\hat{x}_k = \check{x}_k + K_k \underbrace{(y_k - C_k \check{x}_k)}_{\text{更新量}}$$

定义误差： $\check{e}_k = \check{x}_k - x_k$

$$\hat{e}_k = \hat{x}_k - x_k$$

写出误差动态过程：

$$\check{e}_k = A \hat{e}_{k-1} - (B \bar{u} + w_k)$$

$$\hat{e}_k = (1 - K_k C) \check{e}_k + K_k (\bar{y} + n_k)$$



输入和测量的偏差

- 相比于LG系统，有偏情况下多出一项和偏差相关的项：

$$\check{e}_k = A\hat{e}_{k-1} - (B\bar{u} + w_k)$$

在白噪声时，卡尔曼滤波是无偏的，意味着：

$$\hat{e}_k = (1 - K_k C)\check{e}_k + K_k(\bar{y} + n_k)$$

$$E[\hat{e}_k] = 0, \quad E[\check{e}_k] = 0, \quad E[\hat{e}_k \hat{e}_k^T] = \hat{P}_k, \quad E[\check{e}_k \check{e}_k^T] = \check{P}_k$$

- 但是在有偏差存在时，误差会逐渐累积，不妨看k=1时刻：

$$E[\check{e}_1] = A \underbrace{E[\hat{e}_0]}_0 - \left(B\bar{u} + \underbrace{E[w_1]}_0 \right) = -B\bar{u}$$

$$E[\hat{e}_1] = (1 - K_1 C) \underbrace{E[\check{e}_1]}_{-B\bar{u}} + K_1 \left(\bar{y} + \underbrace{E[n_1]}_0 \right)$$

$$= -(1 - K_1 C)B\bar{u} + K_1 \bar{y}$$



输入和测量的偏差

• 方差部分:

$$\begin{aligned}
 E[\check{e}_1 \check{e}_1^T] &= E[(A\hat{e}_0 - (B\bar{u} + w_1))(A\hat{e}_0 - (B\bar{u} + w_1))^T] \\
 &= \underbrace{E[(A\hat{e}_0 - w_1)(A\hat{e}_0 - w_1)^T]}_{\check{P}_1} + \underbrace{(-B\bar{u}) E[(A\hat{e}_0 - w_1)^T]}_0 \\
 &\quad + \underbrace{E[(A\hat{e}_0 - w_1)](-B\bar{u})^T}_0 + (-B\bar{u})(-B\bar{u})^T \\
 &= \check{P}_1 + (-B\bar{u})(-B\bar{u})^T
 \end{aligned}$$

$$\check{P}_1 = E[\check{e}_1 \check{e}_1^T] - \underbrace{E[\check{e}_1]E[\check{e}_1]^T}_{\text{偏差的影响}}$$

$$\begin{aligned}
 E[\hat{e}_1 \hat{e}_1^T] &= E\left[((1 - K_1 C)\check{e}_1 + K_1(\bar{y} + n_1)) \right. \\
 &\quad \left. \times ((1 - K_1 C)\check{e}_1 + K_1(\bar{y} + n_1))^T \right] \\
 &= \underbrace{E[((1 - K_1 C)\check{e}_1 + K_1 n_1)((1 - K_1 C)\check{e}_1 + K_1 n_1)^T]}_{\hat{P}_1} \\
 &\quad + \underbrace{(K_1 \bar{y}) E[((1 - K_1 C)\check{e}_1 + K_1 n_1)^T]}_{(-(1 - K_1 C)B\bar{u})^T} \\
 &\quad + \underbrace{E[((1 - K_1 C)\check{e}_1 + K_1 n_1)] (K_1 \bar{y})^T}_{-(1 - K_1 C)B\bar{u}} + (K_1 \bar{y})(K_1 \bar{y})^T \\
 &= \hat{P}_1 + (-(1 - K_1 C)B\bar{u} + K_1 \bar{y}) \times (-(1 - K_1 C)B\bar{u} + K_1 \bar{y})^T
 \end{aligned}$$

$$\hat{P}_1 = E[\hat{e}_1 \hat{e}_1^T] - \underbrace{E[\hat{e}_1]E[\hat{e}_1]^T}_{\text{偏差的影响}}$$

方差部分是过于确信的 (overconfident)



输入和测量的偏差

- 如果已知偏差的具体值，那么显然可以直接进行修正：

预测：

$$\check{P}_k = A\hat{P}_{k-1}A^T + Q$$

$$\check{x}_k = A\hat{x}_{k-1} + Bu_k + \underbrace{B\bar{u}}_{\text{偏差}}$$

卡尔曼增益：

$$K_k = \check{P}_k C^T (C\check{P}_k C^T + R)^{-1}$$

更新：

$$\hat{P}_k = (1 - K_k C)\check{P}_k$$

$$\hat{x}_k = \check{x}_k + K_k (y_k - C\check{x}_k - \underbrace{\bar{y}}_{\text{偏差}})$$

- 当然这样问题也就不存在了
- 实际情况中，我们可以一边估计状态，一边估计其偏差
- 例子：GPS信号与高精地图中通常存在固定偏差，偏差值和地区相关
- 例子：IMU读取通常存在固定偏差



输入和测量的偏差

- 1. 考虑未知的输入偏差

- 我们增广系统状态，把输入偏差考虑进来： $x'_k = \begin{bmatrix} x_k \\ \bar{u}_k \end{bmatrix}$
- 并假设偏差随时间游走： $\bar{u}_k = \bar{u}_{k-1} + s_k$ $s_k \sim \mathcal{N}(0, W)$

- 那么增广的运动模型和观测模型变为：

$$x'_k = \underbrace{\begin{bmatrix} A & B \\ 0 & 1 \end{bmatrix}}_{A'} x'_{k-1} + \underbrace{\begin{bmatrix} B \\ 0 \end{bmatrix}}_{B'} u_k + \underbrace{\begin{bmatrix} w_k \\ s_k \end{bmatrix}}_{w'_k} \quad \text{其中: } w'_k \sim \mathcal{N}(0, Q'), \quad Q' = \begin{bmatrix} Q & 0 \\ 0 & W \end{bmatrix}$$

$$y_k = \underbrace{\begin{bmatrix} C & 0 \end{bmatrix}}_{C'} x'_k + n_k$$

- 那么可以继续使用卡尔曼滤波器的方式进行状态估计



输入和测量的偏差

- 来考察上述LG系统的能观性，能观性矩阵为：

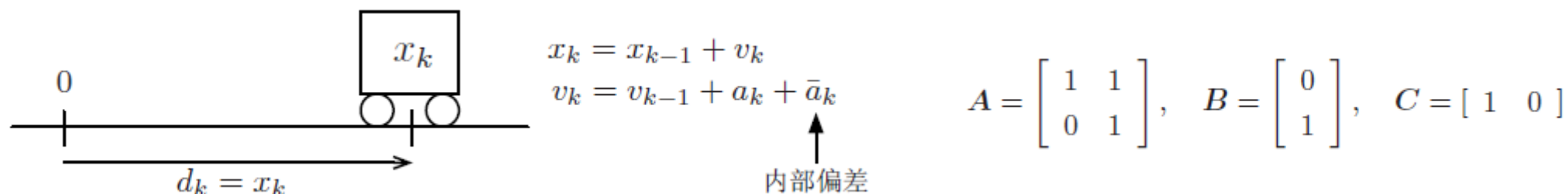
$$\mathcal{O}' = \begin{bmatrix} C' \\ C'A' \\ \vdots \\ C'A'^{(N+U-1)} \end{bmatrix}$$
- 若要状态估计问题存在唯一最优解，则要求：

$$\underbrace{Q' > 0, R > 0}_{\text{由定义得证}}, \quad \text{rank } \mathcal{O}' = N + U$$
- 但是在偏差存在的情况下，这一般是不成立的
- 下面举两个具体的例子



输入和测量的偏差

- 例1：水平运动的小车：状态变量为位移与速度，观测变量为位移，输入为加速度



- 对于原系统和增广系统，其能观性矩阵分别为：

$$\mathcal{O} = \begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \Rightarrow \text{rank } \mathcal{O} = 2 = N$$

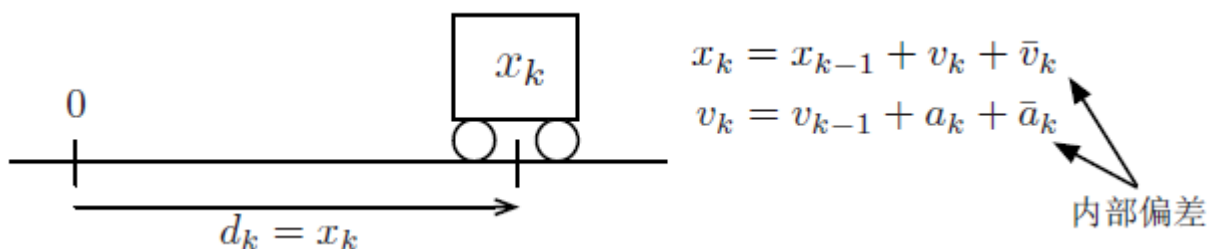
$$\mathcal{O}' = \begin{bmatrix} C' \\ C'A' \\ C'A'^2 \end{bmatrix} = \begin{bmatrix} C & 0 \\ CA & CB \\ CA^2 & CAB + CB \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \Rightarrow \text{rank } \mathcal{O}' = 3 = N + U$$

- 这种系统在有偏差时依然是能观的



输入和测量的偏差

- 例2：上述系统中，既能控制速度，也能控制加速度：



$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad C = [1 \quad 0]$$

- 原系统是能观的，但增广系统却是不能观的：

$$\mathcal{O}' = \begin{bmatrix} C' \\ C'A' \\ C'A'^2 \\ C'A'^3 \end{bmatrix} = \begin{bmatrix} C & 0 \\ CA & CB \\ CA^2 & C(A+1)B \\ CA^3 & C(A^2+A+1)B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 \\ 1 & 3 & 3 & 3 \end{bmatrix}$$

$$\Rightarrow \text{rank } \mathcal{O}' = 3 < 4 = N + U$$

也即无法根据输入和观测唯一地推测系统状态



输入和测量的偏差

- 2. 考虑未知的测量偏差
 - 同样假设偏差为随机游走：

$$x'_k = \begin{bmatrix} x_k \\ \bar{y}_k \end{bmatrix} \quad \bar{y}_k = \bar{y}_{k-1} + s_k \quad \text{其中: } s_k \sim \mathcal{N}(0, W)$$

- 增广后系统运动模型和观测模型变为：

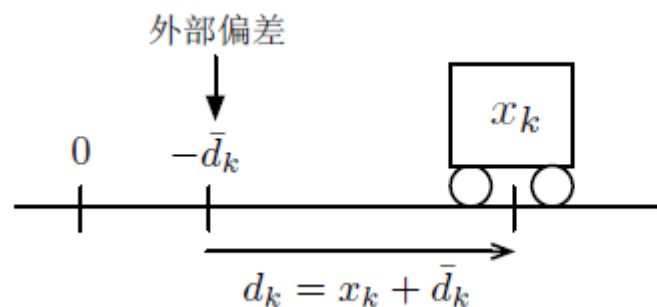
$$x'_k = \underbrace{\begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix}}_{A'} x'_{k-1} + \underbrace{\begin{bmatrix} B \\ 0 \end{bmatrix}}_{B'} u_k + \underbrace{\begin{bmatrix} w_k \\ s_k \end{bmatrix}}_{w'_k} \quad w'_k \sim \mathcal{N}(0, Q'), \quad Q' = \begin{bmatrix} Q & 0 \\ 0 & W \end{bmatrix}$$

$$y_k = \underbrace{[C \quad 1]}_{C'} x'_k + n_k$$



输入和测量的偏差

- 继续来看前面的例子，此时我们认为观测量存在偏差：
 - 车辆一直测量某个事先未知的路标，也就是我们所谓的SLAM问题



$$\begin{aligned} x_k &= x_{k-1} + v_k \\ v_k &= v_{k-1} + a_k \end{aligned}$$

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = [1 \quad 0]$$

- 原问题和增广问题的能观性矩阵为：

$$\mathcal{O} = \begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \Rightarrow \text{rank } \mathcal{O} = 2 = N$$

$$\begin{aligned} \mathcal{O}' &= \begin{bmatrix} C' \\ C'A' \\ C'A'^2 \end{bmatrix} = \begin{bmatrix} C & 1 \\ CA & 1 \\ CA^2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \\ &\Rightarrow \text{rank } \mathcal{O}' = 2 < 3 = N + U \end{aligned}$$

- 所以此问题也不是能观的。直观上，同时移动小车与路标时，测量值不变。



输入和测量的偏差

- 偏差问题的讨论：
 - 无论是运动还是测量方程，都可以把偏差引入到状态中，形成增广的状态估计问题；
 - （大）部分情况下，引入偏差后的问题是不完全能观的；
 - 对于不完全能观的情况，对应的批量估计：

$$(H^T W^{-1} H) \hat{x} = H^T W^{-1} z$$

左侧的矩阵存在零空间，使得它不能直接求逆（但可以求广义逆）

- 我们可以引入偏差量的初始估计，并在求解过程中保持零空间不变，来计算这种情况下的状态估计；



第4讲 偏差、匹配和外点

- 输入和测量的偏差
- 数据关联
- 外点



数据关联

- 数据关联问题实际当中非常普遍，有些非常容易，有些非常困难
 - 例：GPS卫星定位需要接收自身与卫星之间的距离；而卫星是人为编码的，所以很容易确定目前的距离信息是相对于哪一颗卫星；
 - 例：星敏感器测量自身观测到各星星的位置，与星表相比较，确定自身的姿态；然而，星表与观测数据之间的关联问题难以确定；
- 数据关联问题通常认为在状态估计外部解决，即估计状态时已经知道数据关联信息
 - 例如，VSLAM中通过特征点的描述信息来确定特征点与路标的对应关系；
- 也有一些估计器，一边计算状态，一边求解数据关联问题
 - 例如，直接法的VSLAM、激光SLAM中的ICP，等等
 - 在星敏感器的例子中，常见的做法是利用数据的内部特性（各星之间的距离和夹角）来确定数据关联
- 本课程不展开各种问题中数据关联问题的讨论；



第4讲 偏差、匹配和外点

- 输入和测量的偏差
- 数据关联
- 外点



外点

- 外点：测量信息当中的错误值
 - 例如：GPS信号的多路径效应，此时GPS接收器会以高置信度返回一个错误位置信息
- 从正常的噪声模型看来，外点是非常不可能产生的数据
- 一维高斯分布中，3倍标准差以内的数据占99.7%，通常认为大于3倍标准差的数据为外点；
- 实践当中也有一些广泛使用的处理外点方法，下面我们重点介绍两种：**RANSAC**和**M估计**

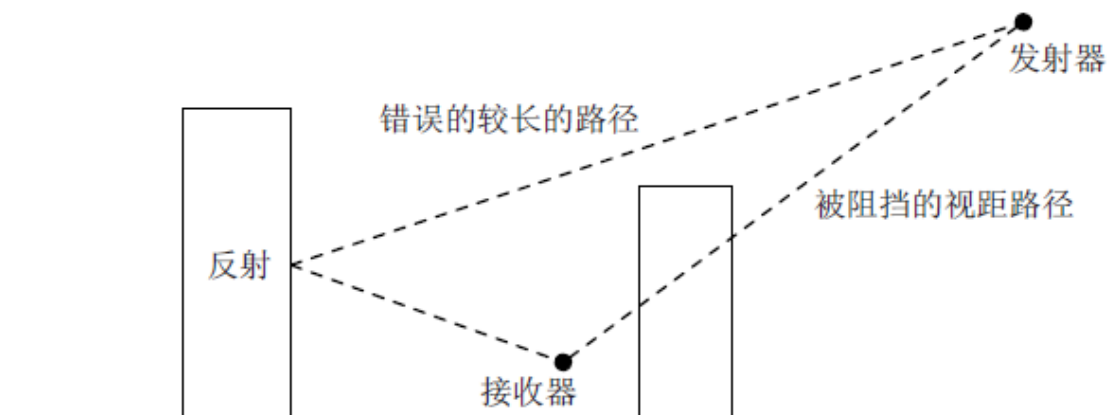
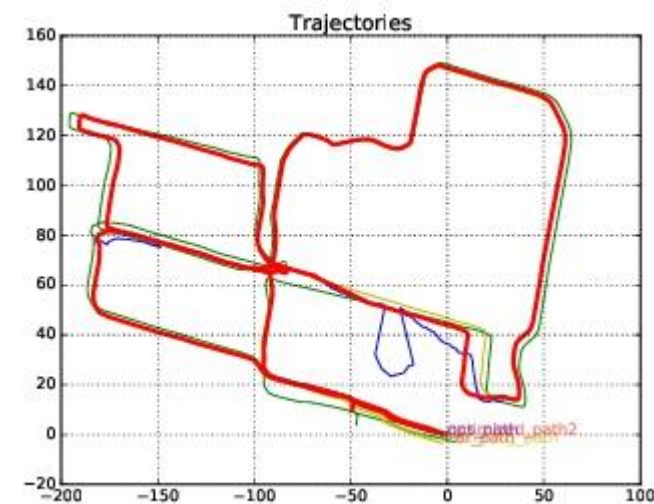


图 5-5 建筑物的复杂结构可能导致 GPS 系统产生错误的观测

实际案例：
红色（估计轨迹）
蓝色（GPS轨迹）



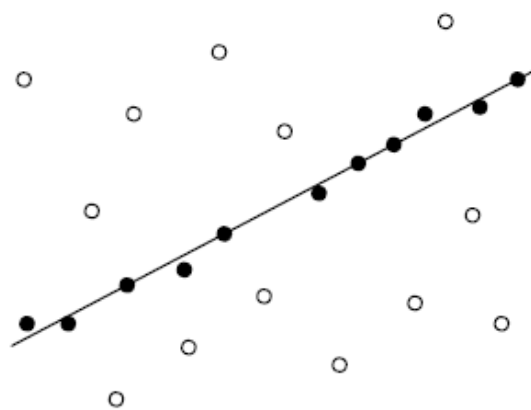


外点

- RANSAC (Random Sample Consensus) , 随机采样一致性
- 我们假设数据是由外点 (outlier) 和内点 (inlier) 混合的, 内点符合预设的模型, 而外点不符合



带有内点和外点的数据集合



RANSAC拟合的拥有最多内点的直线

RANSAC算法:

1. 随机选择 (最小) 子集, 假设它们为内点
2. 利用这些内点拟合出模型参数
3. 判断剩余的数据是否符合拟合模型, 若内点太少, 则放弃, 回到1;
4. 根据3判断的内点, 重新计算模型参数
5. 根据4计算的参数, 重新判断内外点, 直到收敛;



外点

- RANSAC的性能如何?
- 设我们最小需要 n 个内点来计算模型参数, 每个点为内点的概率为 w , 那么, 当计算了 k 次RANSAC后, 我们有概率 p 选到一个全部为内点的子集。显然有:

$$1 - p = (1 - w^n)^k$$

□ k 次 (至少有一个外点的概率)

- 那么 k 与 p 的关系为: $k = \frac{\ln(1 - p)}{\ln(1 - w^n)}$
- RANSAC的思想是非常简洁有效的, 我们设计算法时可以参照RANSAC的思路



外点

- M估计：修改二次函数的形状

- 前面提到的MAP目标函数的形式为： $J(x) = \frac{1}{2} \sum_{i=1}^N e_i(x)^T W_i^{-1} e_i(x)$

- 现在，用更一般的形式表达它：

$$J'(x) = \sum_{i=1}^N \alpha_i \rho(u_i(x)) \quad \text{其中 } \alpha_i > 0 \text{ 为权重值, } u_i(x) = \sqrt{e_i(x)^T W_i^{-1} e_i(x)}$$

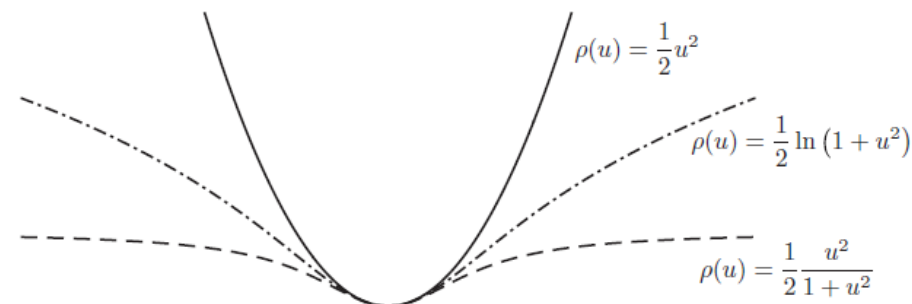
- 我们可以自由选择代价函数的形式，例如：
- $\rho(u) = \frac{1}{2} u^2$
 二次 (quadratic)

$\rho(u) = \frac{1}{2} \ln(1 + u^2)$
 柯西 (Cauchy)

$\rho(u) = \frac{1}{2} \frac{u^2}{1 + u^2}$
 Geman-McClure

- 此类函数称为鲁棒代价函数 (robust cost functions)，除上述的两种之外还有很多

- 相比二次函数，它们在误差较大时增长明显变慢





外点

- 鲁棒函数对自变量的导数:

$$\frac{\partial J'(x)}{\partial x} = \sum_{i=1}^N \alpha_i \frac{\partial \rho}{\partial u_i} \frac{\partial u_i}{\partial e_i} \frac{\partial e_i}{\partial x}$$

$$\text{其中: } \frac{\partial u_i}{\partial e_i} = \frac{1}{u_i(x)} e_i(x)^T W_i^{-1}$$

- 代入之, 得:

$$\frac{\partial J'(x)}{\partial x} = \sum_{i=1}^N e_i(x)^T Y_i(x)^{-1} \frac{\partial e_i(x)}{\partial x}$$

$$\text{其中: } Y_i(x)^{-1} = \frac{\alpha_i}{u_i(x)} \frac{\partial \rho}{\partial u_i} \bigg|_{u_i(x)} W_i^{-1}$$

- 而原问题的导数为:

$$J(x) = \frac{1}{2} \sum_{i=1}^N e_i(x)^T W_i^{-1} e_i(x)$$

$$\frac{\partial J(x)}{\partial x} = \sum_{i=1}^N e_i(x)^T W_i^{-1} \frac{\partial e_i(x)}{\partial x}$$

- 故鲁棒函数可看成对协方差进行了修改, 由固定的W变成了可变的Y



外点

- 在最小二乘的迭代过程中，可以使用当前的工作点计算Y:

$$Y_i(x_{\text{op}})^{-1} = \frac{\alpha_i}{u_i(x_{\text{op}})} \left. \frac{\partial \rho}{\partial u_i} \right|_{u_i(x_{\text{op}})} W_i^{-1}$$

- 这种做法也称为迭代重加权最小二乘 (Iteratively Reweighted Least Squares, IRLS)
- 此类方法实际上把目标函数变为了:

$$J''(x) = \frac{1}{2} \sum_{i=1}^N e_i(x)^T Y_i(x_{\text{op}})^{-1} e_i(x)$$

- 不过这个函数的极值点与原问题在同一点，所以不影响求解



外点

• 协方差估计

- 有时候我们无法十分确定协方差矩阵 W
- 那么可以选择同时估计状态与协方差：

$$J(x) = \frac{1}{2} \sum_{i=1}^N e_i(x)^T W_i^{-1} e_i(x)$$

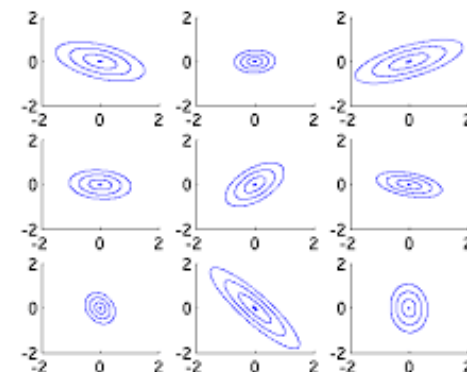
$$\{\hat{x}, \hat{M}\} = \arg \min_{\{x, M\}} J'(x, M) \quad \text{这种做法称为自适应估计 (Adaptive Estimation)}$$

- 假设协方差矩阵服从逆维希特分布 (inverse-Wishart distribution) : $M_i \sim W^{-1}(\Psi_i, \nu_i)$
- 该分布使用参数 Ψ_i 和 ν_i 描述概率分布：

$$p(M_i) = \frac{\det(\Psi_i)^{\frac{\nu_i}{2}}}{2^{\frac{\nu_i M_i}{2}} \Gamma_{M_i}(\frac{\nu_i}{2})} \det(M_i)^{-\frac{\nu_i + M_i + 1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\Psi_i M_i^{-1})\right)$$

- 其中 $\Gamma_{M_i}(\cdot)$ 为多变量Gamma函数

□ 不用害怕这里的分布形式，
因为MAP取对数之后和前面的系数项无关，可舍掉





外点

- 同时估计状态与协方差: $p(x, M|z) = p(x|z, M)p(M) = \prod_{i=1}^N p(x|z_i, M_i)p(M_i)$
- 目标函数变为:

$$J'(x, M) = \frac{1}{2} \sum_{i=1}^N (e_i(x)^T M_i^{-1} e_i(x) - \alpha_i \ln(\det(M_i^{-1})) + \text{tr}(\Psi_i M_i^{-1})) \quad \text{其中: } \alpha_i = \nu_i + M_i + 2$$

- 此时, 问题与x和M都相关, 但我们先固定x, 计算最优的M, 求上式关于M逆的导数, 得:

$$\frac{\partial J'(x, M)}{\partial M_i^{-1}} = \frac{1}{2} e_i(x) e_i(x)^T - \frac{1}{2} \alpha_i M_i + \frac{1}{2} \Psi_i \quad \text{令其为零, 得: } M_i(x) = \underbrace{\frac{1}{\alpha_i} \Psi_i}_{\text{常量}} + \underbrace{\frac{1}{\alpha_i} e_i(x) e_i(x)^T}_{\text{膨胀}}$$

- 该式表明, 最优方差是先验基础上增加一个和残差相关的项
 - 把它代入目标函数中, 目标函数变为:

$$J'(x) = \frac{1}{2} \sum_{i=1}^N \alpha_i \ln(1 + e_i(x)^T \Psi_i^{-1} e_i(x)) \quad \text{该式等价于IRLS中, 以}\Psi\text{作为先验的情况}$$



外点

- 于是，IRLS在这种特定的情况下等效于M估计
 - 所以IRLS是一种更普遍的方法，而不是状况不好时的补救方案
 - 实际当中，IRLS也比较实用，不引入额外的假设



小结

- 现实问题与理论问题是不一样的
 - 现实问题中输入与观测不一定只带有白噪声，可能含有偏差；
 - 现实问题存在数据关联问题，也存在异常值
 - 有时候噪声的协方差也是不确定的
- 本节课的结论
 - 可以把偏差引入到增广状态中，但系统不一定是能观的；
 - 我们通常都需要单独处理外点，使得系统变得鲁棒；
 - 外点的处理方式有RASNAC、M估计（鲁棒估计）、IRLS、自适应估计，等等



作业

- 课后习题1, 2, 3