

第七章 最优估计与最优滤波基础

- 一、最优估计理论概述
- 二、最优估计基本概念与准则
- 三、线性离散时间系统最优滤波方法

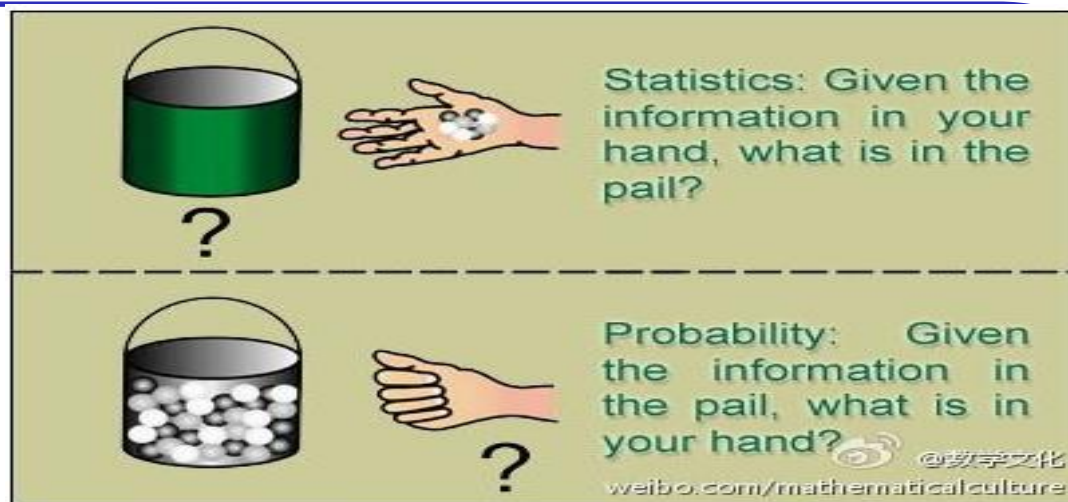
一、最优估计理论概述

- 对系统的输入和输出进行量测而得到的数据只能反映系统的外部特性。
- 系统的动态规律需要用内部状态变量来描述（通常无法直接测量，或有随机干扰存在）。
- 确定性系统的状态量：直接量测、构造状态观测器；
- 随机性系统的状态量：
 - ✓ 从被噪声污染的信号中恢复有用信号；
 - ✓ 利用最优估计技术，即研究如何利用测量数据按照某种判别规则获得对有用信号的最优估计。
- 由于随机干扰信号的描述建立在概率统计的基础之上，故随机干扰下系统状态量的确定也必须建立在概率统计的基础上。属于随机信号处理、统计推断范畴。

概率与统计:

From: 卡内基梅隆大学

Larry Wasserman教授



✓ The basic problem that we study in **probability** is:
Given a data generating process, what are the properties of the outcomes?

概率论是统计推断的基础

✓ The basic problem of **statistical inference** is the inverse of probability:

Given the outcomes, what can we say about the process that generated the data?

统计推断是概率的逆问题

包括：预测、分类、聚类、参数估计

什么叫估计？

估计是指从测量得出的与状态 $\mathbf{x}(t)$ 有关的数据 $\mathbf{z}(t) = \mathbf{h}[\mathbf{x}(t)] + \mathbf{v}(t)$ 中解算出 $\mathbf{x}(t)$ 的方法或技术。其中，随机向量 $\mathbf{v}(t)$ 为测量误差， $\hat{\mathbf{x}}(t)$ 称为对 $\mathbf{x}(t)$ 的估计， $\mathbf{z}(t)$ 称为对 $\mathbf{x}(t)$ 的量测。

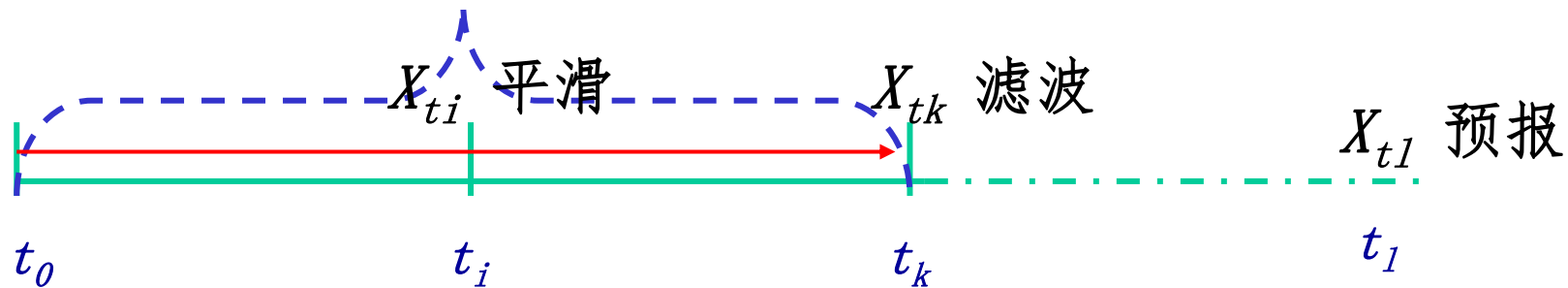
基于 $[t_0, t_k]$ 时间段内的量测数据对状态 $\mathbf{x}(t)$ 进行估计，则

当 $t = t_k$ 时，称 $\hat{\mathbf{x}}(t)$ 为对 $\mathbf{x}(t)$ 的**滤波**；

当 $t > t_k$ 时，称 $\hat{\mathbf{x}}(t)$ 为对 $\mathbf{x}(t)$ 的**预测**；

当 $t < t_k$ 时，称 $\hat{\mathbf{x}}(t)$ 为对 $\mathbf{x}(t)$ 的**平滑**；

最优估计和最优滤波是在某一指标函数达到极值时的估计和滤波结果。不同的优化指标导致不同的最优结果，在满足某些特定条件的情况下，不同的最优结果之间可以是等价的。



二、最优估计基本概念与估计准则

1. 最优估计的基本概念

设 $\mathbf{x} \in \mathbb{R}^n$ 是一个未知参数向量，量测 \mathbf{z} 是一个 m 维的随机向量， $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N\}$ 是 \mathbf{z} 的一组容量为 N 的样本集合，称统计量

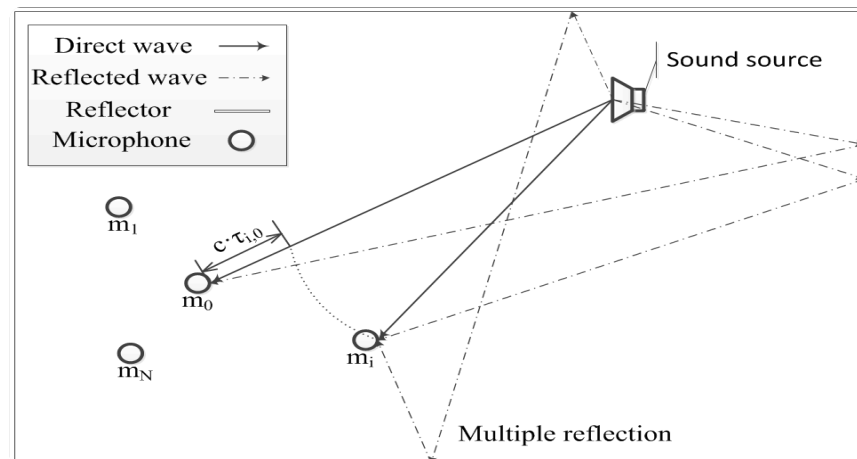
$$\hat{\mathbf{x}}^{(N)} = \varphi(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N) \quad (1)$$

为对 \mathbf{x} 的一个估计量，其中 $\varphi(\cdot)$ 称为估计准则。

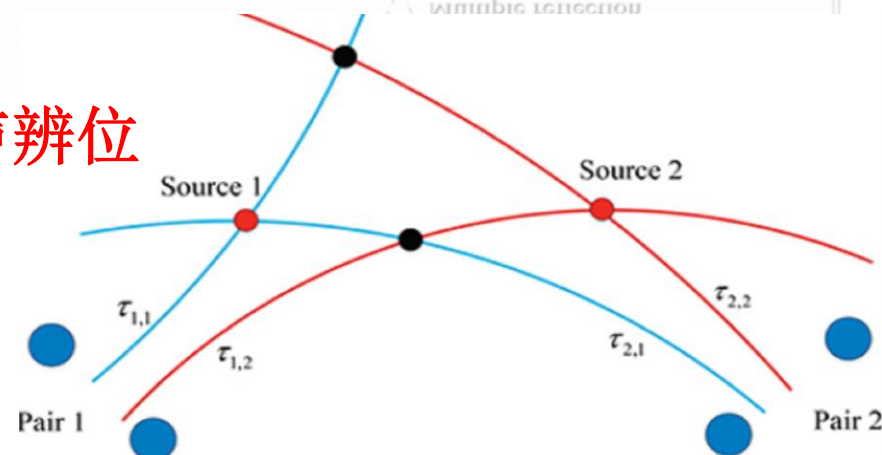
利用样本对参数进行估计本质上是随机的。当样本值给定时所得到的参数估计值一般与真值并不相同，因而需要用某些准则进行评价。

最优估计的典型应用领域：导航定位系统，电力系统，通信系统，经济系统等。

声源定位



听声辨位



任务：基于多个麦克风单元提供的复合信号，提取声音的能量、时差和方位特征，确定声源目标数目和位置信息。

无人船

感知，导航（避障、航迹规划），控制（航向、航速）



瞭望者 II



精海一号

搭载：GNSS接收机、测深仪、ADCP（声学多普勒流速剖面仪），电子罗盘，水质采样和在线监测仪等多种传感设备。

无人机



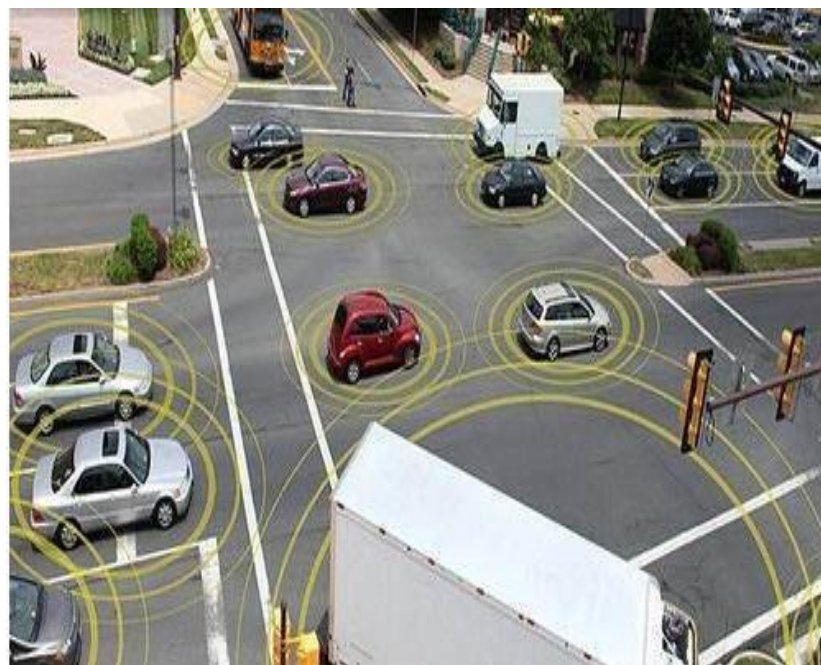
海上搜救



地面监视

搭载：惯性导航+GPS, 超声波测距传感器、激光测距传感器、空速计、气压计、视觉传感器

无人车



搭载：摄像头、激光雷达、毫米波雷达等，辅以 GPS 高精度地图
自主环境感知、路径规划、速度控制等 **（感知，导航，控制）**

定义1 对于式 (1) , 所得估计量如果满足

$$E[\hat{\mathbf{x}}^{(N)}] = E[\mathbf{x}]$$

则称 $\hat{\mathbf{x}}^{(N)}$ 是对参数 \mathbf{x} 的一个无偏估计; 如果满足

$$\lim_{N \rightarrow \infty} E(\hat{\mathbf{x}}^{(N)}) = E(\mathbf{x})$$

则称 $\hat{\mathbf{x}}^{(N)}$ 是对参数 \mathbf{x} 的一个渐近无偏估计。

例1 设 z 是任意随机变量, 期望 $E[z] = m$, 方差为 $\text{Var}[z] = \sigma^2$ 而 z 的一组容量为 N 的样本是 $\{z_1, z_2, \dots, z_N\}$, 假定它们之间相互独立且同分布 (IID: Independent Identical Distributed); 定义两个统计量分别为

$$\hat{m}_N = \frac{1}{N} \sum_{i=1}^N z_i, \quad \hat{\sigma}_N^2 = \frac{1}{N} \sum_{i=1}^N z_i^2 - \hat{m}_N^2$$

试证 \hat{m}_N 和 $\hat{\sigma}_N^2$ 分别为对 m 和 σ^2 无偏估计和渐进无偏估计。

证明：因为

$$E[\hat{m}_N] = \frac{1}{N} \sum_{i=1}^N E[z_i] = \frac{1}{N} (N \cdot m) = m$$

所以， \hat{m}_N 是 m 的一个无偏估计；而

$$\begin{aligned} E[\hat{\sigma}_N^2] &= \frac{1}{N} \sum_{i=1}^N E[z_i^2] - E[\hat{m}_N^2] = E[z^2] - E[\hat{m}_N^2] \\ &= E[z^2] - E\left\{\left[\frac{1}{N} \sum_{i=1}^N z_i\right]^2\right\} = E[z^2] - \frac{1}{N^2} E\left[\sum_{i=1}^N z_i^2 + \sum_{i=1, i \neq j}^N z_i z_j\right] \\ &= E[z^2] - \frac{1}{N} E[z^2] - \frac{N-1}{N} m^2 = \frac{N-1}{N} \sigma^2 \end{aligned}$$

所以， $\hat{\sigma}_N^2$ 是对 σ^2 的一个渐近无偏估计。

利用 $D(\xi) = E(\xi - E\xi)^2 = E(\xi^2) - (E\xi)^2$

例2 由 $x_i = s + v_i \quad (i = 1, 2, \dots, N)$

给定的观测样本，其中 s 为信号， v_i 为均值为0的随机噪声，信号 s 的一个可能估值是

$$\hat{s} = g(x_1, x_2, \dots, x_N) = \frac{1}{N}(x_1 + x_2 + \dots + x_N)$$

判断该估值是否是 s 的无偏估计？

解：

$$\begin{aligned} \because E[\hat{s}] &= \frac{1}{N} \{E[x_1] + E[x_2] + \dots + E[x_N]\} \\ &= \frac{1}{N} \{s \cdot N\} \quad (\because E[v_i] = 0 \quad i = 1, 2, \dots, N) \end{aligned}$$

$$\therefore E[\hat{s}] = s$$

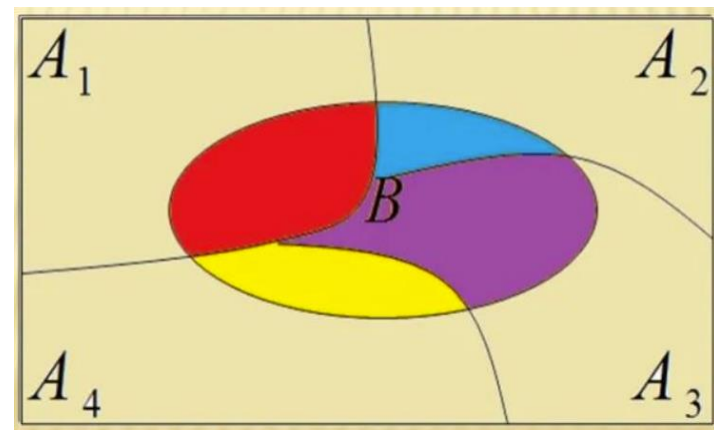
$\therefore \hat{s}$ 是 s 的一个无偏估计。

回顾

全概率公式： 将一个复杂事件(随机量)的概率计算问题，分解为若干个简单事件的概率计算问题，最后应用概率的可加性求出最终结果。

关键：

找到划分(完备事件组) $A_i (i = 1, 2, \dots, n)$ 在实际应用中，这个划分是导致 B 发生的各种可能原因， $P(B)$ 是事件 B 分别在这些原因下发生的条件概率的加权平均。



$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

主观Bayes公式：

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^n P(A_j)P(B|A_j)}$$

2、最大后验概率准则

设 \mathbf{x} 为随机向量， z 为对 \mathbf{x} 的量测， $f(\mathbf{x}|z)$ 称为 \mathbf{x} 的**后验概率密度函数**，它是**给定量测 z 条件下 \mathbf{x} 的条件概率密度函数**。最大后验概率准则又称为极大后验估计，是以待估参量 \mathbf{x} 的后验概率密度最大作为估计准则，即

$$\hat{\mathbf{x}}_{MAP} = \arg \max_{\hat{\mathbf{x}}} f(\hat{\mathbf{x}} | z)$$

其中， \arg 意为使得优化指标函数取得极值时的参数值。

以下方程称为**后验方程**：

$$\left. \frac{\partial \ln f(\hat{\mathbf{x}} | z)}{\partial \hat{\mathbf{x}}} \right|_{\hat{\mathbf{x}} = \hat{\mathbf{x}}_{MAP}} = 0$$

3、极大似然估计

极大似然估计是以观测值出现的概率最大作为估计准则，也就是似然函数 $\Lambda(\mathbf{x}) = f(\mathbf{z} | \mathbf{x})$ 最大，即

$$\hat{\mathbf{x}}_{ML} = \arg \max_{\hat{\mathbf{x}}} f(\mathbf{z} | \hat{\mathbf{x}})$$

与极大后验估计相比，极大似然估计有两个优点：

- (1) 确定似然函数比确定后验概率密度函数容易些。
- (2) 被估计量是随机的，也可以是非随机的，适用范围比极大后验估计要广。

但是，如果对待估计量有先验知识，则 $\hat{\mathbf{x}}_{ML}$ 精度不如 $\hat{\mathbf{x}}_{MAP}$ 。当不掌握待估计量先验信息时，二者具有相同的估计结果。

以下方程称为似然方程：

$$\left. \frac{\partial \ln \Lambda(\hat{\mathbf{x}})}{\partial \hat{\mathbf{x}}} \right|_{\hat{\mathbf{x}} = \hat{\mathbf{x}}_{ML}} = \left. \frac{\partial \ln f(\mathbf{z} | \hat{\mathbf{x}})}{\partial \hat{\mathbf{x}}} \right|_{\hat{\mathbf{x}} = \hat{\mathbf{x}}_{ML}} = 0$$

例3 令 x_1, x_2, \dots, x_N 是基于一个具有概率密度函数

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-(x - \mu)^2 / (2\sigma^2)\right]$$

的正态分布得到的随机样本。求均值 μ 和方差 σ^2

的最大似然估计。

解：似然函数是 μ 和 σ^2 二者的函数，所以有

$$\begin{aligned} L(x_1, \dots, x_N; \mu, \sigma^2) &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-(x_i - \mu)^2 / (2\sigma^2)\right] \\ &= (2\pi\sigma^2)^{-N/2} \exp\left[\frac{-1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2\right] \end{aligned}$$

由于对似然函数直接求偏导较复杂，因此可以先对L求对数，然后再求偏导，即令

$$L = \ln L(x_1, \dots, x_N, \mu, \sigma^2) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2$$

对待估参数分别求偏导，然后令偏导为零，得

$$\begin{cases} \frac{\partial L}{\partial \mu} = -\frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \mu) = 0 \\ \frac{\partial L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^N (x_i - \mu)^2 = 0 \end{cases}$$

整理得

$$\hat{\mu}_{ML} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}$$

无偏的；

$$\hat{\sigma}_{ML}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

$$= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^N (x_i)^2 - \bar{x}^2$$

$$E\{\hat{\sigma}_{ML}^2\} = \frac{N-1}{N} \hat{\sigma}^2$$

(例1已证) 有偏的；

但是这个偏差可以通过乘一常数 $\frac{N}{N-1}$ 得以消除。

4、加权最小二乘估计（掌握）

最小二乘法发表在1809年《天体运行论》中。但实际上早在1795年，高斯已经推导出，1801年应用这种理论思想推算了谷神星（位于火星和木星间）的轨道。

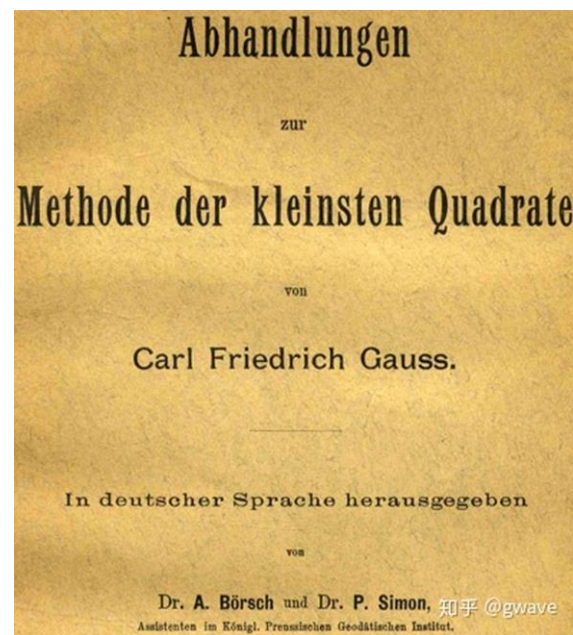
假定量测信息 \mathbf{z} 可以表示为参数 \mathbf{x} 的线性函数，即 $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$

其中 $\mathbf{H} \in \mathbb{R}^{m \times n}$ ， $\mathbf{v} \in \mathbb{R}^m$ 是一个零均值方差为 \mathbf{R} 的随机向量；设加权阵 $\mathbf{W} \in \mathbb{R}^{m \times m}$ 且有 $\mathbf{W} \geq 0$ 为对称非负定阵，则如下估计

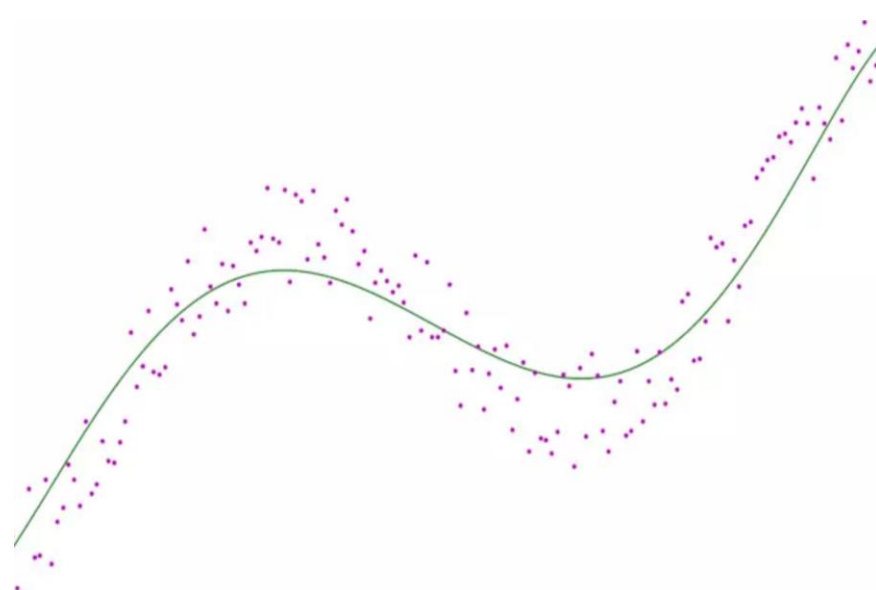
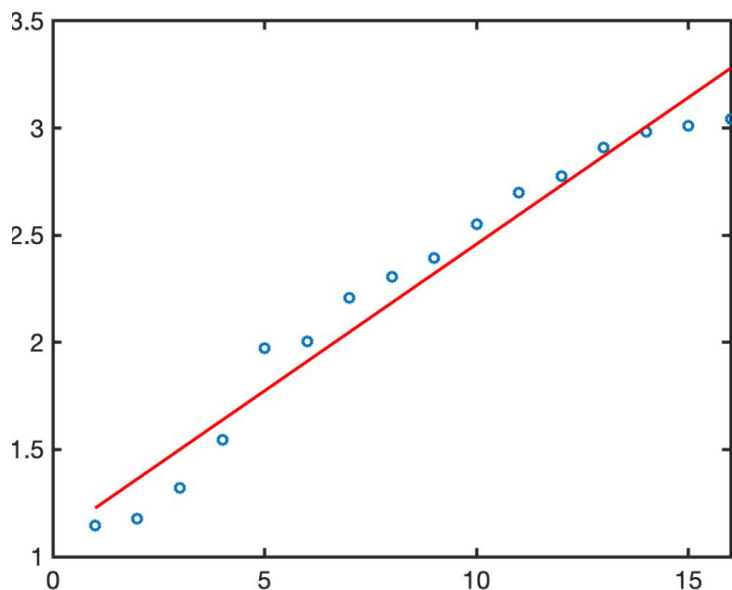
$$\hat{\mathbf{x}}_{WLS} = \arg \min_{\hat{\mathbf{x}}} (\mathbf{z} - \mathbf{H}\hat{\mathbf{x}})^T \mathbf{W} (\mathbf{z} - \mathbf{H}\hat{\mathbf{x}})$$

称为加权最小二乘（Weighted Least Squares, WLS）估计；如果 $\mathbf{W} = \mathbf{I}$ ，则称为最小二乘（Least Squares, LS）估计。

加权最小二乘法：残差的
加权平方和最小



曲线拟合问题中的应用



机器学习的目标为：

- 学习一个 $f(\mathbf{x})$ 多项式，可以拟合真实数据的变化趋势
- 准则：使每一个真实数值到 $f(\mathbf{x})$ 的拟合数值的距离之和最小

定理1 设 H^TWH 可逆, 则基于量测信息 \mathbf{z} 和加权矩阵 \mathbf{W} 对参数 \mathbf{x} 的WLS估计为

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

估计误差协方差矩阵为

$$E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] = (H^TWH)^{-1} H^T W R W H (H^TWH)^{-1}$$

其中, $\tilde{\mathbf{x}}$ 为估计误差 ($\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}_{WLS}$)。

证明：因为

$$\min_{\hat{x}} (z - H\hat{x})^T W (z - H\hat{x}) = \min_{\hat{x}} (z^T W z - 2z^T W H \hat{x} + \hat{x}^T H^T W H \hat{x})$$

而

$$\begin{aligned} & \frac{\partial}{\partial \hat{x}} (z^T W z - 2z^T W H \hat{x} + \hat{x}^T H^T W H \hat{x}) \\ &= \frac{\partial}{\partial \hat{x}} (z^T W z - 2\hat{x}^T H^T W z + \hat{x}^T H^T W H \hat{x}) \\ &= \underline{-2H^T W z + 2H^T W H \hat{x}} \\ &= 0 \end{aligned}$$

所以相应的最小二乘估计为：

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

利用

$$\frac{\partial \langle Ax, x \rangle}{\partial x} = Ax + A^T x$$

$$\frac{\partial \langle x, y \rangle}{\partial x} = y$$

关于估计误差协方差矩阵 $E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T]$:

估计误差可表示为:

$$\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}_{WLS}$$

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

所以 $E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] = E[(\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{v} (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{v}^T]$

$$= (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} E[\mathbf{v} \mathbf{v}^T] \mathbf{W} \mathbf{H} (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1}$$

$$= (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{R} \mathbf{W} \mathbf{H} (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1}$$

需要指出的是: 如果量测噪声 $\mathbf{v} \in \mathbb{R}^{N_m}$ 是零均值的随机向量, 则加权最小二乘估计也是无偏估计, 因为此时:

$$E[\tilde{\mathbf{x}}] = E[(\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{v}] = 0$$

$$E[\hat{\mathbf{x}}] = E[\mathbf{x}]$$

两个特例：

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

$$E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{R} \mathbf{W} \mathbf{H} (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1}$$

① 当 $\mathbf{W} = \mathbf{I}$ ，则加权最小二乘估计就是一般的最小二乘估计：

估计值： $\hat{\mathbf{x}}_{LS} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{z}$

掌握！！！！

估计误差的协方差阵为： $E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R} \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1}$

② 如 $\mathbf{W} = \mathbf{R}^{-1}$ ，则加权最小二乘估计又称为马尔科夫估计：

估计值： $\hat{\mathbf{x}}_M = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z}$

估计误差的协方差阵为： $E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1}$

马尔科夫估计的误差协方差阵最小，是加权最小二乘估计中的最优者。

(注意两个矩阵 \mathbf{A} , \mathbf{B} 比大小时, $\mathbf{A} > \mathbf{B}$ 指 $\mathbf{A} - \mathbf{B}$ 正定)

例 3 根据对二维矢量 θ 的两次观测

$$x_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \theta + n_1$$

$$x_2 = 4 = [1 \quad 2] \theta + n_2$$

求其最小二乘估计。

解：

令 $2 = \theta_1 + \theta_2 + n_{11}, \quad 1 = \theta_2 + n_{12}$

有 $4 = \theta_1 + 2\theta_2 + n_2, \quad x = H\theta + n$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}, \quad H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}, \quad n = \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$$

代入公式，得

$$\hat{\theta}_{LS} = (H^T H)^{-1} H^T x$$

$$= \left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}^T \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}^T \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix}$$

例4 用一台仪器对未知确定性标量 x 作 n 次测量，测量值分别为 z_1, \dots, z_n ，测量误差 $v_i (i=1, \dots, n)$ 为均值为0方差为 r 的随机标量，求其最小二乘估计，并估计协方差阵。

解：令 $\mathbf{Z} = [z_1, \dots, z_n]^T$ $\mathbf{H} = [1, \dots, 1]^T$ $\mathbf{V} = [v_1, \dots, v_n]^T$

有 $E[\mathbf{V}\mathbf{V}^T] = r\mathbf{I}$, $\mathbf{W} = \mathbf{I}$

估计结果为：

$$\hat{x} = \frac{1}{n} [z_1 + \dots + z_n]$$

相应的协方差阵

$$E[\tilde{x}\tilde{x}^T] = \frac{r}{n}$$

思考：说明什么？

例5 设估计确定性电压 x 有两种方案，[方案1]：用2个高精度的电压表各量测一次，测量噪声的均值为0、方差为2；[方案2]：用4个精度低的电压表各量测一次，测量噪声的均值为0、方差为3.5，若采用马尔科夫估计，试问哪一种方案好？

解：由题意，得测量方程： $\mathbf{z} = \mathbf{H}x + \mathbf{V}$ $\mathbf{W} = \mathbf{R}^{-1}$

① 若采用方案1，则： $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

从而有 $E[\tilde{x}^2] = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} = 1$

② 若采用方案2，则： $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} 3.5 & & & 0 \\ & 3.5 & & \\ & & 3.5 & \\ 0 & & & 3.5 \end{bmatrix}$

从而有 $E[\tilde{x}^2] = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} = 7/8$

方案2好。说明什么？

例6 用两台仪器对未知确定性标量 x 各直接测量一次，量测分别为 z_1 和 z_2 ，两台仪器测量误差均值为0，方差分别为 r 和 $4r$ 。求其最小二乘估计以及马尔科夫估计，并估计协方差阵。

解：由题意，得测量方程：

$$\mathbf{z} = \mathbf{H}x + \mathbf{V} \quad \text{其中} \quad \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} r & 0 \\ 0 & 4r \end{bmatrix}$$

若采用一般最小二乘估计， $\mathbf{W} = \mathbf{I}$ ，有

$$\hat{x} = \frac{1}{2}[z_1 + z_2],$$

$$E[\tilde{x}^2] = E[(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{v})(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{v})^T]$$

$$= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R} \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1}$$

$$= \frac{1}{2} \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{R} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{2} = \frac{5}{4} r > r$$

思考：说明什么？

上式说明，使用精度差一倍的两台仪器同时测量，最小二乘估计效果不如单独采用一台仪器。

如果采用马尔科夫估计，取 $W = R^{-1}$ ，有

$$\hat{x} = \frac{1}{5}[4z_1 + z_2]$$

$$E[\tilde{x}^2] = \frac{4}{5}r < r$$

可见，应用马尔科夫估计，可以获得较仅用一台精度高的仪器更好的效果。所以，增加不同的测量值，并根据其精度区别利用，能有效提高估计精度。

(了解)

其他**LS**相关。。。。

递推最小二乘

约束最小二乘

总体最小二乘

鲁棒最小二乘

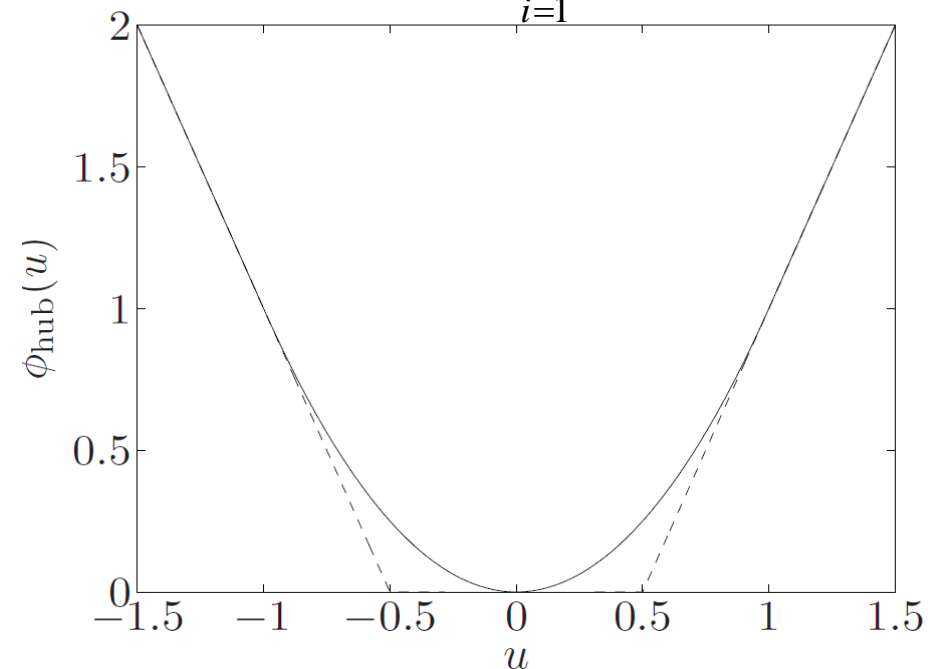
偏最小二乘

最小一乘

最小零乘

(了解-其它代价函数)

$$\min \sum_{i=1}^m \phi(r_i)$$

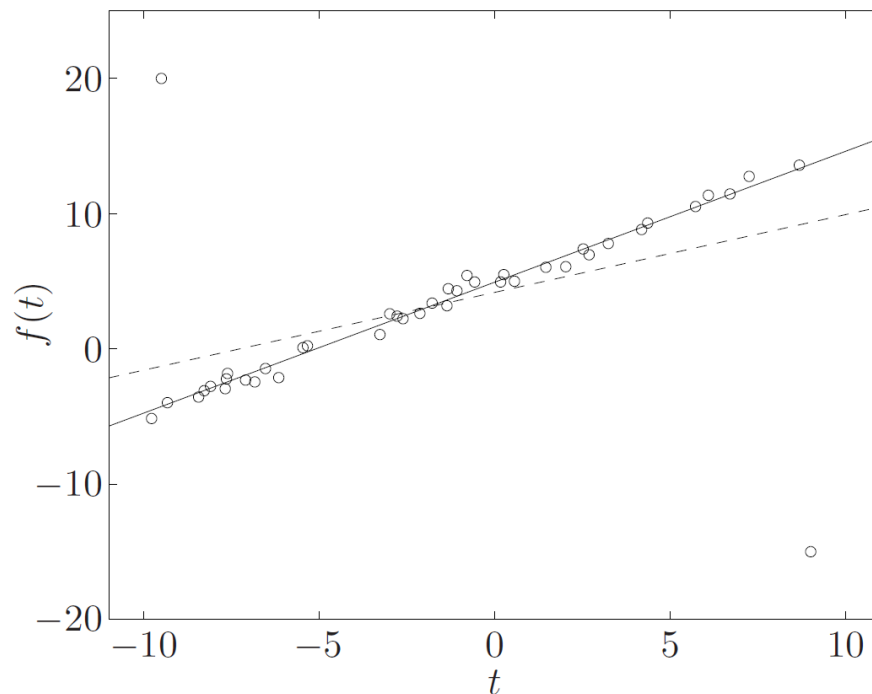


惩罚函数

鲁棒最小二乘 $\phi_{\text{hub}}(u) = \begin{cases} u^2 & |u| \leq M \\ M(2|u| - M) & |u| > M \end{cases}$

最小一乘 $\phi(r_i) = |r_i|$

s.t. $r = z - H\hat{x}$



拟合结果

5、最小均方误差估计 (Minimum Mean Square Error-MMSE)

设 \mathbf{x} 为随机向量， \mathbf{z} 为对 \mathbf{x} 的量测。显然，基于量测 \mathbf{z} 对 \mathbf{x} 的估计是 \mathbf{z} 的函数。由于随机误差的存在，需要按统计意义的最优标准求取。最小均方误差估计是使下述指标达到最小的估计

$$J = E \left[(\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) | \mathbf{z} \right]$$

$$= \int (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) p(\mathbf{x} | \mathbf{z}) d\mathbf{x}$$

$$= \int \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} p(\mathbf{x} | \mathbf{z}) d\mathbf{x}$$

$$\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$$

定理2: 随机参数 \mathbf{x} 的最小均方误差估计就是 \mathbf{x} 的后验期望。

优化的目标函数

$$E[(\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) | \mathbf{z}]$$

求一阶导，并令一阶导为0

$$\begin{aligned} & \frac{\partial}{\partial \hat{\mathbf{x}}} \left\{ E[(\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) | \mathbf{z}] \right\} \\ & \frac{\partial}{\partial \hat{\mathbf{x}}} \int_{-\infty}^{+\infty} (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) p(\mathbf{x} | \mathbf{z}) d\mathbf{x} \\ & = \int_{-\infty}^{+\infty} 2(\mathbf{x} - \hat{\mathbf{x}}) p(\mathbf{x} | \mathbf{z}) d\mathbf{x} = 0 \\ & \Rightarrow \hat{\mathbf{x}} \int_{-\infty}^{+\infty} p(\mathbf{x} | \mathbf{z}) d\mathbf{x} = \int_{-\infty}^{+\infty} \mathbf{x} p(\mathbf{x} | \mathbf{z}) d\mathbf{x} = E[\mathbf{x} | \mathbf{z}] \\ & \Rightarrow \hat{\mathbf{x}} = E[\mathbf{x} | \mathbf{z}] \quad \left(\text{同时 } \frac{\partial^2 J}{\partial^2 \hat{\mathbf{x}}} \Big|_{\hat{\mathbf{x}} = \hat{\mathbf{x}}_{\text{MMSE}}} = 2\mathbf{I} > 0 \right) \end{aligned}$$

利用

$$\frac{\partial \langle A\mathbf{x}, \mathbf{x} \rangle}{\partial \mathbf{x}} = A\mathbf{x} + A^T \mathbf{x}$$

$$\frac{\partial \langle \mathbf{x}, \mathbf{y} \rangle}{\partial \mathbf{x}} = \mathbf{y}$$

定理3 最小均方误差估计是 \mathbf{x} 的无偏估计，即证

$$E[\hat{\mathbf{x}}_{MMSE}] = E[\mathbf{x}]$$

证明：最小均方误差估计为 $\hat{\mathbf{x}}_{MMSE} = E[\mathbf{x} | \mathbf{z}]$ ，其中，求期望是对 \mathbf{x} 进行的，因此 $\hat{\mathbf{x}}_{MMSE}$ 是关于 \mathbf{z} 的函数。对 $\hat{\mathbf{x}}_{MMSE}$ 求期望实质上是对 \mathbf{z} 进行的，所以有：

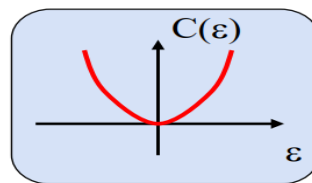
$$\begin{aligned} E[\hat{\mathbf{x}}_{MMSE}] &= E_z[E[\mathbf{x} | \mathbf{z}]] = \int_{-\infty}^{\infty} E[\mathbf{x} | \mathbf{z}] p_z(\mathbf{z}) d\mathbf{z} \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \mathbf{x} p(\mathbf{x} | \mathbf{z}) d\mathbf{x} \right] p_z(\mathbf{z}) d\mathbf{z} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{x} p(\mathbf{x} | \mathbf{z}) p_z(\mathbf{z}) d\mathbf{x} d\mathbf{z} \end{aligned}$$

应用贝叶斯公式，得

$$\begin{aligned} E[\hat{\mathbf{x}}_{MMSE}] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{x} p(\mathbf{x}, \mathbf{z}) d\mathbf{z} d\mathbf{x} = \int_{-\infty}^{\infty} \mathbf{x} \left[\int_{-\infty}^{\infty} p(\mathbf{x}, \mathbf{z}) d\mathbf{z} \right] d\mathbf{x} \\ &= \int_{-\infty}^{\infty} \mathbf{x} p_x(\mathbf{x}) d\mathbf{x} = E[\mathbf{x}] \end{aligned}$$

(了解)

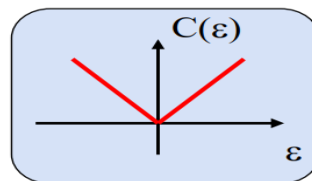
1. Quadratic: $C(\varepsilon) = \varepsilon^2$



$$\hat{\theta} = E\{\theta | \mathbf{x}\}$$

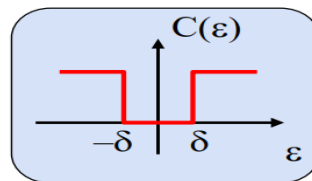
$$= \text{mean of } p(\theta | \mathbf{x})$$

2. Absolute: $C(\varepsilon) = |\varepsilon|$

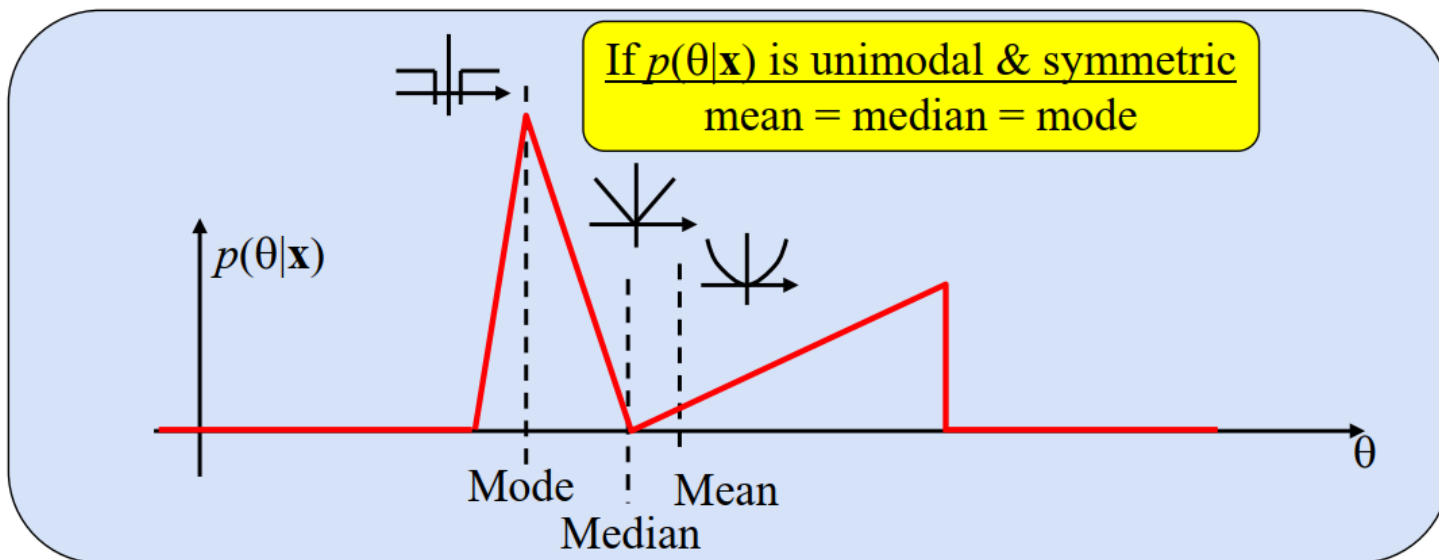


$$\hat{\theta} = \text{median of } p(\theta | \mathbf{x})$$

3. Hit-or-Miss: $C(\varepsilon) = \begin{cases} 0, & |\varepsilon| < \delta \\ 1, & |\varepsilon| \geq \delta \end{cases}$
 $\delta > 0$ and small



$$\hat{\theta} = \text{mode of } p(\theta | \mathbf{x})$$



定理4 若 n 维参数 \mathbf{x} 和 m 维样本 \mathbf{z} 均服从高斯分布, 令 $\mathbf{y} = [\mathbf{x} \ \mathbf{z}]^T$, 其均值和协方差阵分别为

$$\bar{\mathbf{y}} = E \begin{bmatrix} \mathbf{x} \\ \mathbf{z} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{z}} \end{bmatrix}, \quad \mathbf{R}_y = \text{Cov}(\tilde{\mathbf{y}}) = \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xz} \\ \mathbf{R}_{zx} & \mathbf{R}_{zz} \end{bmatrix}$$

其中 $\mathbf{R}_{xx} = \text{Cov}[\tilde{\mathbf{x}}] = E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T]$, $\mathbf{R}_{xz} = \text{Cov}[\tilde{\mathbf{x}}, \tilde{\mathbf{z}}] = E[\tilde{\mathbf{x}}\tilde{\mathbf{z}}^T]$

$$\mathbf{R}_{zz} = \text{Cov}[\tilde{\mathbf{z}}] = E[\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T]$$

并假定 \mathbf{R}_y 和 \mathbf{R}_{zz} 非奇异; 那么, 给定 \mathbf{z} 时 \mathbf{x} 也是 Gauss 的, 最小均方误差估计为

$$\hat{\mathbf{x}}_{MMSE} = E[\mathbf{x} | \mathbf{z}] = \bar{\mathbf{x}} + \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} (\mathbf{z} - \bar{\mathbf{z}})$$

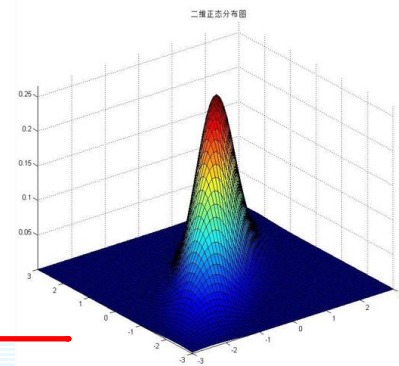
估计误差的协方差阵是

$$\mathbf{P} = \text{Cov}(\tilde{\mathbf{x}} | \mathbf{z}) = E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T | \mathbf{z}] = \mathbf{R}_{xx} - \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} \mathbf{R}_{zx}$$

多元正态分布

$$\mathcal{N}(\cdot; \mathbf{a}, \mathbf{B})$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{B}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{a})' \mathbf{B}^{-1} (\mathbf{x} - \mathbf{a}) \right\}$$



上述定理说明：当参数 \mathbf{x} 和样本 \mathbf{z} 是联合高斯分布的，最小方差估计不必通过对条件概率的积分求取，只需知道参数 \mathbf{x} 和样本 \mathbf{z} 的一、二阶矩。

6、线性最小均方误差估计（了解）

设 $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{B} \in \mathbb{R}^{n \times (Nm)}$, 若对参数 \mathbf{x} 的估计可表示为量测信息 \mathbf{z} 的线性函数

$$\hat{\mathbf{x}} = \mathbf{a} + \mathbf{B}\mathbf{z}$$

则称为线性估计；进而如果估计误差的均方值达到最小 ($\min E\{[\mathbf{x} - \hat{\mathbf{x}}]^T [\mathbf{x} - \hat{\mathbf{x}}]\}$) , 则称之为线性最小均方误差估计 (LMMSE) , 也称线性最小方差估计。我们不加证明给出如下定理，用以描述当参数 \mathbf{x} 和量测信息 \mathbf{z} 是任意分布时的线性最小均方误差估计。

定理5 设参数 \mathbf{z} 和量测信息 \mathbf{x} 是任意分布, \mathbf{z} 的协方差阵 \mathbf{R}_{zz} 非奇异, 则利用量测信息 \mathbf{z} 对参数 \mathbf{x} 的LMMSE估计唯一地表示为

$$\hat{\mathbf{x}}_{LMMSE} = E^*[\mathbf{x} | \mathbf{z}] = \bar{\mathbf{x}} + \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} (\mathbf{z} - \bar{\mathbf{z}})$$

此处 $E^*(\cdot | \cdot)$ 只是一个记号, 不表示条件期望; 而估计误差的协方差阵是

$$\mathbf{P} = \text{Cov}[\tilde{\mathbf{x}} | \mathbf{z}] = \mathbf{R}_{xx} - \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} \mathbf{R}_{zx}$$

例7： 设被估计量 \mathbf{x} 和观测量 \mathbf{z} 的联合分布如下表所示，求 \mathbf{x} 的最小均方误差估计。

$\mathbf{z} \backslash \mathbf{x}$	-3	-2	2	3
-1	1/4	1/4	0	0
1	0	0	1/4	1/4

解：

$$\hat{\mathbf{x}}_{MMSE} = E[\mathbf{x} | \mathbf{z}] = \sum_{i=1}^4 \mathbf{x}_i p(\mathbf{x}_i | \mathbf{z}) = \begin{cases} -\frac{5}{2} & \mathbf{z} = -1 \\ \frac{5}{2} & \mathbf{z} = 1 \end{cases}$$

其中，

$$p(\mathbf{x}_i | \mathbf{z} = -1) = \frac{p(\mathbf{x}_i, \mathbf{z} = -1)}{\sum_i p(\mathbf{x}_i, \mathbf{z} = -1)} = \begin{cases} \frac{1}{2}, i = 1, 2 \\ 0, i = 3, 4 \end{cases}$$

$$p(\mathbf{x}_i | \mathbf{z} = 1) = \frac{p(\mathbf{x}_i, \mathbf{z} = 1)}{\sum_i p(\mathbf{x}_i, \mathbf{z} = 1)} = \begin{cases} 0, i = 1, 2 \\ \frac{1}{2}, i = 3, 4 \end{cases}$$

例8： 设被估计量 \mathbf{x} 和观测量 \mathbf{z} 的联合分布如下表所示，分别求 \mathbf{x} 的最小均方误差（及线性最小均方误差）估计，并比较方差阵大小。

\mathbf{z}	$\mathbf{x}=-1$	$\mathbf{x}=0$	$\mathbf{x}=1$	$\mathbf{x}=2$
$\mathbf{z}=-1$	$1/10$	$2/10$	0	0
$\mathbf{z}=0$	0	$1/10$	$3/10$	0
$\mathbf{z}=1$	0	0	$1/10$	$2/10$

解： 若采用最小均方误差估计

$$\hat{\mathbf{x}}_{MMSE} = E[\mathbf{x} | \mathbf{z}] = \sum_{i=1}^4 \mathbf{x}_i p(\mathbf{x}_i | \mathbf{z}) = \begin{cases} -\frac{1}{3} & \mathbf{z} = -1 \\ \frac{3}{4} & \mathbf{z} = 0 \\ \frac{5}{3} & \mathbf{z} = 1 \end{cases}$$

\leftarrow	$x=-1\leftarrow$	$x=0\leftarrow$	$x=1\leftarrow$	$x=2\leftarrow$
$z=-1\leftarrow$	$1/10\leftarrow$	$2/10\leftarrow$	$0\leftarrow$	$0\leftarrow$
$z=0\leftarrow$	$0\leftarrow$	$1/10\leftarrow$	$3/10\leftarrow$	$0\leftarrow$
$z=1\leftarrow$	$0\leftarrow$	$0\leftarrow$	$1/10\leftarrow$	$2/10\leftarrow$

其中，条件概率

$$p(\mathbf{x} = -1 | \mathbf{z} = -1) = \frac{1/10}{3/10} = \frac{1}{3}, p(\mathbf{x} = 0 | \mathbf{z} = -1) = \frac{2/10}{3/10} = \frac{2}{3}$$

$$p(\mathbf{x} = 0 | \mathbf{z} = 0) = \frac{1/10}{4/10} = \frac{1}{4}, p(\mathbf{x} = 1 | \mathbf{z} = 0) = \frac{3/10}{4/10} = \frac{3}{4}$$

$$p(\mathbf{x} = 1 | \mathbf{z} = 1) = \frac{1/10}{3/10} = \frac{1}{3}, p(\mathbf{x} = 2 | \mathbf{z} = 1) = \frac{2/10}{3/10} = \frac{2}{3}$$

估计误差的方差是：

$$E_{x|z} (\mathbf{x} - \hat{\mathbf{x}}_{MMSE})^2$$

$$= \sum_{i=1}^6 \sum_{j=1}^3 (\mathbf{x}_i - E[\mathbf{x} | \mathbf{z}_j])^2 p(\mathbf{x}_i, \mathbf{z}_j)$$

$$\hat{\mathbf{x}}_{MMSE} = E[\mathbf{x} | \mathbf{z}] = \begin{cases} -\frac{1}{3} & \mathbf{z} = -1 \\ \frac{3}{4} & \mathbf{z} = 0 \\ \frac{5}{3} & \mathbf{z} = 1 \end{cases}$$

$$= \frac{1}{10} \left(\frac{2}{3} \right)^2 + \frac{2}{10} \left(\frac{1}{3} \right)^2 + \frac{1}{10} \left(\frac{3}{4} \right)^2 + \frac{3}{10} \left(\frac{1}{4} \right)^2 + \frac{1}{10} \left(\frac{2}{3} \right)^2 + \frac{2}{10} \left(\frac{1}{3} \right)^2 = \frac{5}{24}$$

\leftarrow	$x=-1\leftarrow$	$x=0\leftarrow$	$x=1\leftarrow$	$x=2\leftarrow$
$z=-1\leftarrow$	$1/10\leftarrow$	$2/10\leftarrow$	$0\leftarrow$	$0\leftarrow$
$z=0\leftarrow$	$0\leftarrow$	$1/10\leftarrow$	$3/10\leftarrow$	$0\leftarrow$
$z=1\leftarrow$	$0\leftarrow$	$0\leftarrow$	$1/10\leftarrow$	$2/10\leftarrow$

其中，条件概率

$$p(\mathbf{x} = -1 | \mathbf{z} = -1) = \frac{1/10}{3/10} = \frac{1}{3}, p(\mathbf{x} = 0 | \mathbf{z} = -1) = \frac{2/10}{3/10} = \frac{2}{3}$$

$$p(\mathbf{x} = 0 | \mathbf{z} = 0) = \frac{1/10}{4/10} = \frac{1}{4}, p(\mathbf{x} = 1 | \mathbf{z} = 0) = \frac{3/10}{4/10} = \frac{3}{4}$$

$$p(\mathbf{x} = 1 | \mathbf{z} = 1) = \frac{1/10}{3/10} = \frac{1}{3}, p(\mathbf{x} = 2 | \mathbf{z} = 1) = \frac{2/10}{3/10} = \frac{2}{3}$$

估计误差的方差是：

$$\begin{aligned}
 & E(\mathbf{x} - \hat{\mathbf{x}}_{MMSE})^2 \\
 &= \sum_{i=1}^6 \sum_{j=1}^3 \left(\mathbf{x}_i - E[\mathbf{x} | \mathbf{z}_j] \right)^2 p(\mathbf{x}_i, \mathbf{z}_j) \quad \bullet \quad \hat{\mathbf{x}}_{MMSE} = E[\mathbf{x} | \mathbf{z}] = \begin{cases} -\frac{1}{3} & \mathbf{z} = -1 \\ \frac{3}{4} & \mathbf{z} = 0 \\ \frac{5}{3} & \mathbf{z} = 1 \end{cases} \\
 &= \frac{1}{10} \left(\frac{2}{3} \right)^2 + \frac{2}{10} \left(\frac{1}{3} \right)^2 + \frac{1}{10} \left(\frac{3}{4} \right)^2 + \frac{3}{10} \left(\frac{1}{4} \right)^2 + \frac{1}{10} \left(\frac{2}{3} \right)^2 + \frac{2}{10} \left(\frac{1}{3} \right)^2 = \frac{5}{24}
 \end{aligned}$$

若采用线性最小均方误差估计, $\hat{\mathbf{x}}_{LMMSE} = E^*[\mathbf{x} | \mathbf{z}] = \bar{\mathbf{x}} + \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} (\mathbf{z} - \bar{\mathbf{z}})$

其中, 一阶矩 $\bar{\mathbf{x}} = \frac{1}{10}(-1) + \frac{3}{10} \times 1 + \frac{1}{10} \times 1 + \frac{2}{10} \times 2 = \frac{7}{10}$

$$\bar{\mathbf{z}} = \left(\frac{1}{10} + \frac{2}{10} \right) (-1) + \left(\frac{1}{10} + \frac{2}{10} \right) \times 1 = 0$$

\leftarrow	$x=-1\leftarrow$	$x=0\leftarrow$	$x=1\leftarrow$	$x=2\leftarrow$
$z=-1\leftarrow$	$1/10\leftarrow$	$2/10\leftarrow$	$0\leftarrow$	$0\leftarrow$
$z=0\leftarrow$	$0\leftarrow$	$1/10\leftarrow$	$3/10\leftarrow$	$0\leftarrow$
$z=1\leftarrow$	$0\leftarrow$	$0\leftarrow$	$1/10\leftarrow$	$2/10\leftarrow$

二阶矩

$$\mathbf{R}_{xx} = \frac{1}{10} \left(\frac{17}{10} \right)^2 + \frac{2}{10} \left(\frac{7}{10} \right)^2 + \frac{1}{10} \left(\frac{7}{10} \right)^2 + \frac{3}{10} \left(\frac{3}{10} \right)^2 + \frac{1}{10} \left(\frac{3}{10} \right)^2 + \frac{2}{10} \left(\frac{13}{10} \right)^2 = \frac{81}{100}$$

$$\mathbf{R}_{zz} = \left(\frac{1}{10} + \frac{2}{10} \right) (-1)^2 + \left(\frac{1}{10} + \frac{2}{10} \right) \times 1^2 = \frac{3}{5}$$

$$\mathbf{R}_{xz} = \frac{1}{10} \left(-\frac{17}{10} \right) (-1) + \frac{2}{10} \left(-\frac{7}{10} \right) (-1) + \frac{1}{10} \times \frac{3}{10} + \frac{2}{10} \times \frac{13}{10} = \frac{3}{5}$$

$$\begin{aligned}\hat{\mathbf{x}}_{LMMSE} &= E^*[\mathbf{x} | \mathbf{z}] = \bar{\mathbf{x}} + \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} (\mathbf{z} - \bar{\mathbf{z}}) \\ &= \frac{7}{10} + \mathbf{z} = \begin{cases} -\frac{3}{10} & \mathbf{z} = -1 \\ \frac{7}{10} & \mathbf{z} = 0 \\ \frac{17}{10} & \mathbf{z} = 1 \end{cases}\end{aligned}$$

估计误差的方差是：

$$\begin{aligned}\mathbf{P} = Var[\tilde{\mathbf{x}}] &= \mathbf{R}_{xx} - \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} \mathbf{R}_{zx} \\ &= \frac{81}{100} - \frac{3}{5} = \frac{21}{100} > \frac{5}{24}\end{aligned}$$

大于最小均方误差估计获得的估计误差方差！

三、线性离散时间系统最优滤波方法

- 1、卡尔曼滤波概述
- 2、卡尔曼滤波递推公式
- 3、估计滤波性能评价

1、卡尔曼滤波概述

R. E. Kalman 匈牙利裔美国数学家(1930—2016)



1960年提出了离散系统**卡尔曼滤波**。

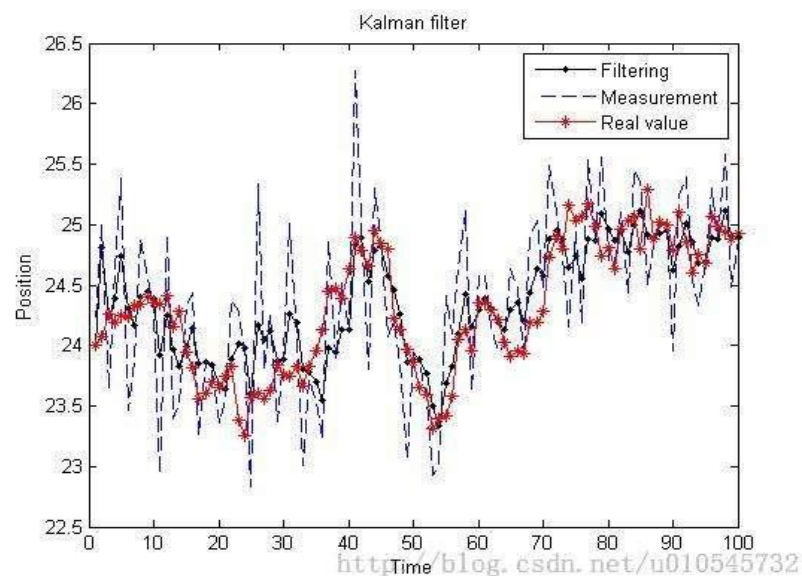
- 是一种时域滤波方法，采用状态空间方法描述系统；
- 算法采用递推机制，数据存储空间小；

• 卡尔曼滤波器源于他的博士论文和1960年发表的论文

《A New Approach to Linear Filtering and Prediction Problems》（线性滤波与预测问题的新方法）。

Online:

<http://www.cs.unc.edu/~welch/kalman/media/pdf/Kalman1960.pdf>



(1) Kalman 滤波理论的提出:

① 1959年, NASA开始研究载人太空飞船登月方案—阿波罗计划, 当时提出了两个主要问题:

- 飞船中途导航和制导, 即飞船的测轨定姿问题✓
- 液体燃料助推器大挠度条件下的自动驾驶问题。



Apollo

② 主要问题是解决对飞船运动状态的估计。当时，测量主要来自三个子系统：

➤ 飞船的惯性测量装置

➤ 天文观测仪

➤ 地面测轨系统(测轨数据经数据链传送至飞船)

那么导航制导问题的关键就是：三个子系统的信息融合与估计问题。当时，曾试图采用递推加权最小二乘和维纳滤波的方法，均因满足不了要求和过于复杂的计算而被迫放弃。

③ 1960年秋，Kalman访问了NASA，同年提出卡尔曼滤波算法。立即引起重视并投入研究。尔后，由MIT负责研制成功阿波罗计划中的导航系统，这也是Kalman理论早期最著名的应用之一。在以后的几十年里，该理论又被发展，应用领域也被广泛拓展。

卡尔曼滤波器：Optimal Recursive Data Processing Algorithm
(最优递推数据处理算法)

对于解决很多实际问题，它是最优、效率最高甚至是最有用的。

- 递推最小均方误差估计
- 最初提出的滤波基本理论只适用于线性系统
- 需建立较准确的动态模型和观测模型。

- 递推最小均方误差估计
- 最初提出的滤波基本理论只适用于线性系统
- 需建立较准确的动态模型和观测模型。

卡尔曼在NASA（美国国家航空航天局）埃姆斯研究中心访问时,发现他的方法对于解决阿波罗计划的轨道预测很有用,后来阿波罗飞船的导航系统使用了这种滤波器。



Apollo

2、卡尔曼滤波递推公式

状态空间描述

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{\Gamma}_k \mathbf{w}_k$$

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k$$

状态演化过程

测量过程

其中 $k \in \mathbb{N}$ 是时间指标, $\mathbf{x}_k \in \mathbb{R}^n$ 是 k 时刻的系统状态向量, \mathbf{F}_k 是系统状态转移矩阵, 而 \mathbf{w}_k 是过程噪声, $\mathbf{\Gamma}_k$ 是过程噪声加权矩阵; $\mathbf{z}_k \in \mathbb{R}^m$ 是 k 时刻对系统状态的量测向量, \mathbf{H}_k 是量测矩阵, 而 \mathbf{v}_k 是量测噪声。

定义: 直到 k 时刻所有的累积量测

$$\mathbf{Z}^k = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k\}$$

定理6 对于如上系统描述， 假定过程噪声 $w_k \sim \mathcal{N}(0, Q_k)$ 是一个独立过程 ($E\{w_k w_l^T\} = Q_k \delta_{kl}$) (白、高斯)， 量测噪声 $v_k \sim \mathcal{N}(0, R_k)$ 也是一个独立过程 ($E\{v_k v_l^T\} = R_k \delta_{kl}$) (白、高斯)； 它们之间相互独立， 而且二者还与初始状态估计误差也独立， 那么基于MMSE准则， 有如下基本Kalman滤波公式： (δ : **Kronecker delta function**)

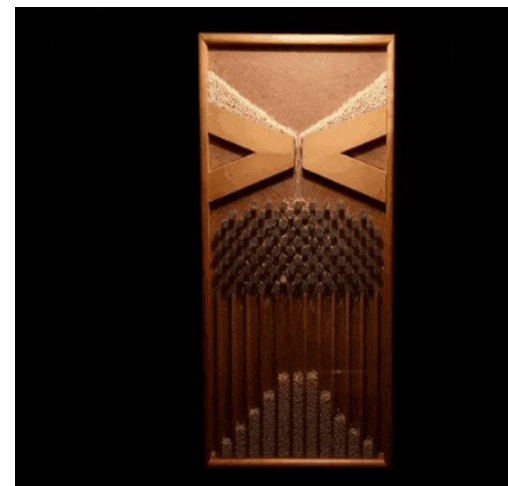
(i) 初始条件： $\hat{x}_{0|0} = \bar{x}_0$, $\tilde{x}_{0|0} = x_0 - \hat{x}_{0|0}$, $\text{cov}(\tilde{x}_{0|0}) = P_{0|0}$

(ii) 预报 (时间更新)： 一步提前预测值和预测误差的协方差阵分别是

$$\hat{x}_{k|k-1} = E(x_k | Z^{k-1}) = F_{k-1} \hat{x}_{k-1|k-1}, \quad (1)$$

$$P_{k|k-1} = \text{cov}(\tilde{x}_{k|k-1} | Z^{k-1}) = F_{k-1} P_{k-1|k-1} F_{k-1}^T + \Gamma_{k-1} Q_{k-1} \Gamma_{k-1}^T, \quad (2)$$

其中 $\tilde{x}_{k|k-1} = x_k - \hat{x}_{k|k-1}$ 是一步预测误差；



(iii) **量测更新**：获取新的量测 z_k 后，滤波更新值和相应的滤波误差的协方差阵分别是：

$$\hat{x}_{k|k} = E(x_k | Z^k) = \hat{x}_{k|k-1} + K_k (z_k - H_k \hat{x}_{k|k-1}), \quad (4)$$

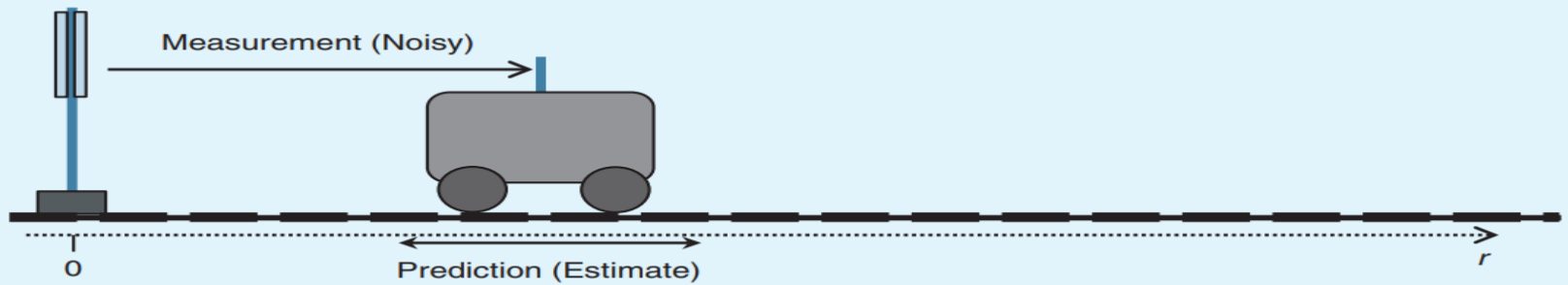
$$P_{k|k} = \text{COV}(\tilde{x}_{k|k} | Z^k) = P_{k|k-1} - P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} H_k P_{k|k-1}, \quad (5)$$

其中 $\tilde{x}_{k|k} = x_k - \hat{x}_{k|k}$ 是滤波误差。

k 时刻的**Kalman增益阵**为：

$$K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} \quad (3)$$

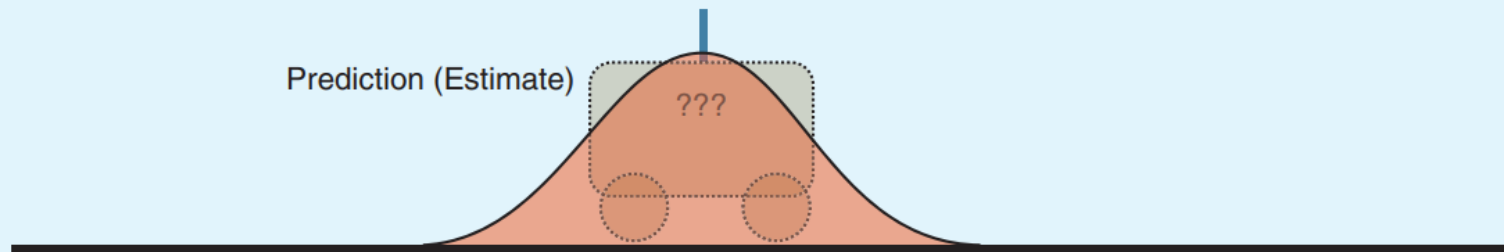
新息协方差矩阵



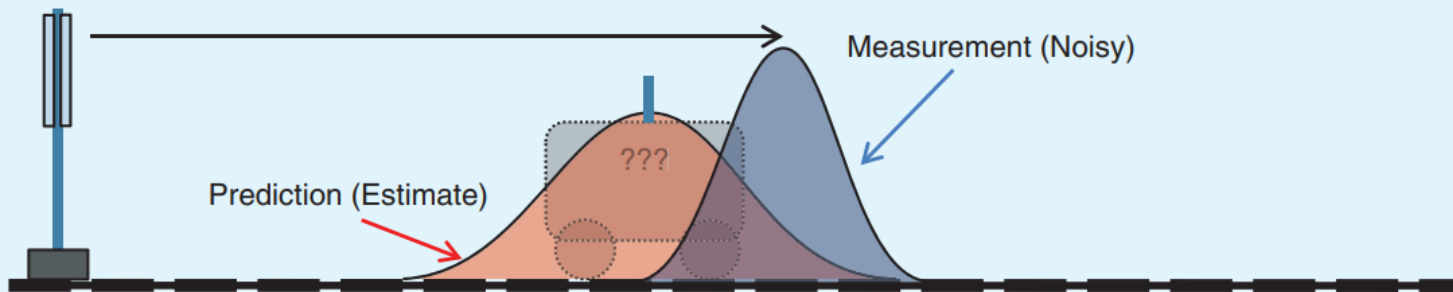
[FIG1] This figure shows the one-dimensional system under consideration.



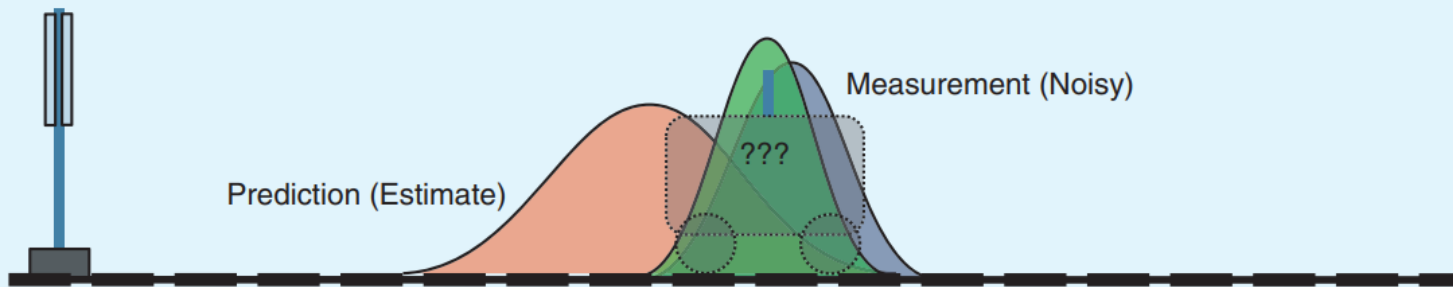
[FIG2] The initial knowledge of the system at time $t = 0$. The red Gaussian distribution represents the pdf providing the initial confidence in the estimate of the position of the train. The arrow pointing to the right represents the known initial velocity of the train.



[FIG3] Here, the prediction of the location of the train at time $t = 1$ and the level of uncertainty in that prediction is shown. The confidence in the knowledge of the position of the train has decreased, as we are not certain if the train has undergone any accelerations or decelerations in the intervening period from $t = 0$ to $t = 1$.



[FIG4] Shows the measurement of the location of the train at time $t = 1$ and the level of uncertainty in that noisy measurement, represented by the blue Gaussian pdf. The combined knowledge of this system is provided by multiplying these two pdfs together.



[FIG5] Shows the new pdf (green) generated by multiplying the pdfs associated with the prediction and measurement of the train's location at time $t = 1$. This new pdf provides the best estimate of the location of the train, by fusing the data from the prediction and the measurement.

思路: $\hat{\mathbf{x}}_{k-1|k-1}, P_{k-1|k-1} \xRightarrow{\text{时间更新}} \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{z}}_{k|k-1}; R(\mathbf{R}_{xx}, \mathbf{R}_{xz}, \mathbf{R}_{zz}) \xRightarrow{\text{卡尔曼增益计算}} K \xRightarrow{\text{量测更新}} \hat{\mathbf{x}}_{k|k}, P_{k|k}$

Remark: 在MMSE下

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} (\mathbf{z} - \bar{\mathbf{z}}), \mathbf{P} = \text{cov}(\tilde{\mathbf{x}}) = \mathbf{R}_{xx} - \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} \mathbf{R}_{zx}$$

$$\hat{\mathbf{x}}_{k|k} = E(\mathbf{x}_k | \mathbf{Z}^k) \quad \mathbf{z}_k \quad \mathbf{P}_{k|k} = \text{cov}(\tilde{\mathbf{x}}_{k|k} | \mathbf{Z}^k)$$

状态预报

$$\begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{z}} \end{bmatrix} \triangleq \begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \hat{\mathbf{z}}_{k|k-1} \end{bmatrix} = E \left[\begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{pmatrix} \middle| \mathbf{Z}^{k-1} \right] \Rightarrow \begin{bmatrix} \tilde{\mathbf{x}}_{k|k-1} \\ \tilde{\mathbf{z}}_{k|k-1} \end{bmatrix}$$

状态预报误差 \mathbf{R}_{xx}

量测预报误差 (新息) \mathbf{R}_{zz}

量测预报

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xz} \\ \mathbf{R}_{zx} & \mathbf{R}_{zz} \end{bmatrix} = \text{cov} \left[\begin{pmatrix} \tilde{\mathbf{x}}_{k|k-1} \\ \tilde{\mathbf{z}}_{k|k-1} \end{pmatrix} \middle| \mathbf{Z}^{k-1} \right] = E \left[\begin{pmatrix} \tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^T & \tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{z}}_{k|k-1}^T \\ \tilde{\mathbf{z}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^T & \tilde{\mathbf{z}}_{k|k-1} \tilde{\mathbf{z}}_{k|k-1}^T \end{pmatrix} \middle| \mathbf{Z}^{k-1} \right]$$

\mathbf{S}_{k+1} 新息协方差矩阵

证明：分两个步骤完成证明。

1、根据系统演化方程以及噪声假设，则有如下一步预测公式：

状态预报：
$$\hat{\mathbf{x}}_{k|k-1} = E(\mathbf{x}_k | \mathbf{Z}^{k-1}) = E(\mathbf{F}_{k-1} \mathbf{x}_{k-1} + \mathbf{\Gamma}_{k-1} \mathbf{w}_{k-1} | \mathbf{Z}^{k-1}) = \mathbf{F}_{k-1} \hat{\mathbf{x}}_{k-1|k-1}$$

状态预报协方差：
$$\begin{aligned} \mathbf{P}_{k|k-1} &= \text{cov}(\tilde{\mathbf{x}}_{k|k-1} | \mathbf{Z}^{k-1}) = E[\tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^T | \mathbf{Z}^{k-1}] \\ &= E\{[\mathbf{F}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}) + \mathbf{\Gamma}_{k-1} \mathbf{w}_{k-1}][\mathbf{F}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}) + \mathbf{\Gamma}_{k-1} \mathbf{w}_{k-1}]^T | \mathbf{Z}^{k-1}\} \\ &= \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{F}_{k-1}^T + \mathbf{\Gamma}_{k-1} \mathbf{Q}_{k-1} \mathbf{\Gamma}_{k-1}^T; \end{aligned}$$

量测预报：
$$\hat{\mathbf{z}}_{k|k-1} = E(\mathbf{z}_k | \mathbf{Z}^{k-1}) = E(\mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k | \mathbf{Z}^{k-1}) = \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}$$

量测预报误差（新息）：
$$\tilde{\mathbf{z}}_{k|k-1} = \mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}$$

量测预报协方差：
$$\begin{aligned} \text{cov}(\tilde{\mathbf{z}}_{k|k-1} | \mathbf{Z}^{k-1}) &= E[(\mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_k)(\mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_k)^T | \mathbf{Z}^{k-1}] = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k; \\ \text{cov}(\tilde{\mathbf{x}}_{k|k-1}, \tilde{\mathbf{z}}_{k|k-1} | \mathbf{Z}^{k-1}) &= E[\tilde{\mathbf{x}}_{k|k-1} (\mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_k)^T | \mathbf{Z}^{k-1}] = \mathbf{P}_{k|k-1} \mathbf{H}_k^T; \end{aligned}$$

2、根据MMSE估计准则，在获取新的量测后有

$$\hat{\mathbf{x}}_{k|k} = E(\mathbf{x}_k | \mathbf{Z}^k)$$

(考虑 $\hat{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} (\mathbf{z} - \bar{\mathbf{z}})$ ———— 定理3)

$$\hat{\mathbf{x}}_{k|k-1} \text{ --- } \bar{\mathbf{x}}$$

$$\hat{\mathbf{z}}_{k|k-1} \text{ ---- } \bar{\mathbf{z}}$$

$$\text{cov}(\tilde{\mathbf{x}}_{k|k-1}, \tilde{\mathbf{z}}_{k|k-1} | \mathbf{Z}^{k-1}) \text{ --- } > \mathbf{R}_{xz}$$

$$\text{cov}(\tilde{\mathbf{z}}_{k|k-1}, \tilde{\mathbf{z}}_{k|k-1} | \mathbf{Z}^{k-1}) \text{ --- } > \mathbf{R}_{zz}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})$$

$$= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}),$$

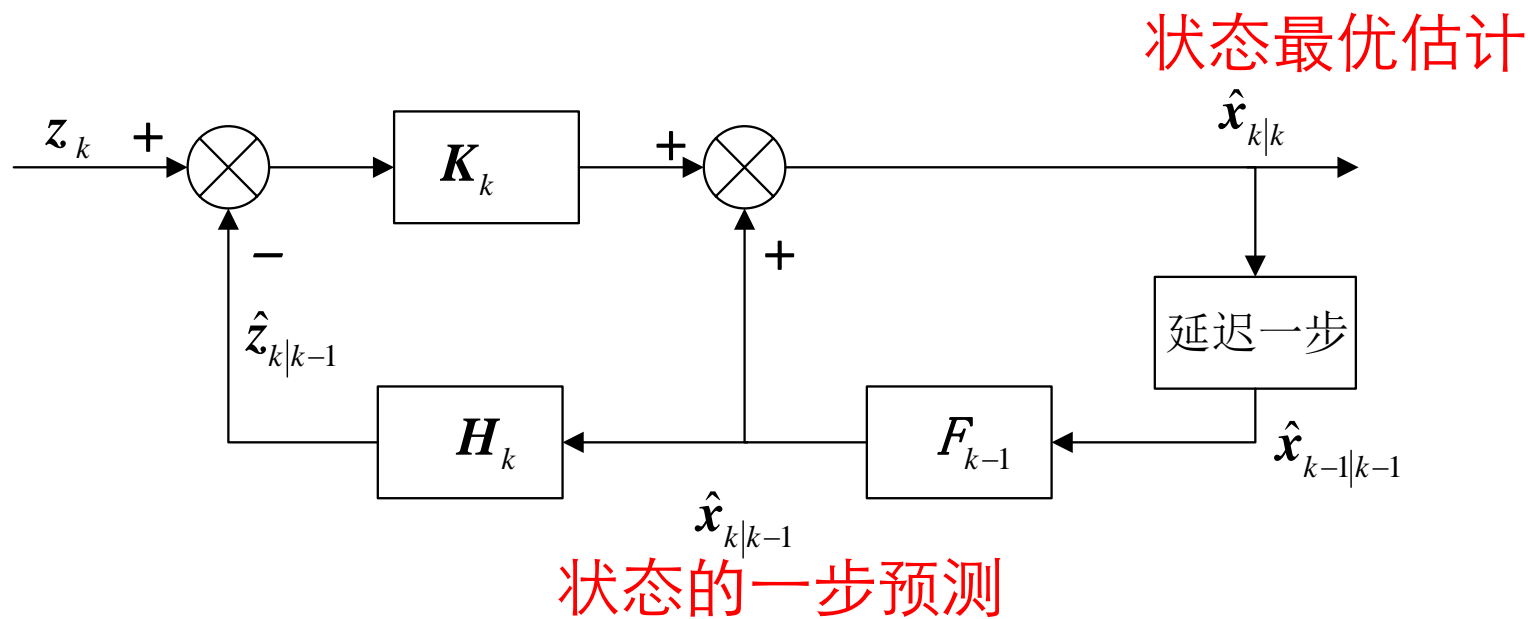
$$\text{从而, } \tilde{\mathbf{x}}_{k|k} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k} = \tilde{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \tilde{\mathbf{z}}_{k|k-1}$$

其中，Kalman 增益阵满足 $\mathbf{K}_k = \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$

同时有

$$\begin{aligned} P_{k|k} &= \text{cov}(\tilde{\mathbf{x}}_{k|k} | \mathbf{Z}^k) = E(\tilde{\mathbf{x}}_{k|k} \tilde{\mathbf{x}}_{k|k}^T | \mathbf{Z}^k) \\ &= E[(\tilde{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \tilde{\mathbf{z}}_{k|k-1})(\tilde{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \tilde{\mathbf{z}}_{k|k-1})^T | \mathbf{Z}^k] \\ &= P_{k|k-1} - \mathbf{K}_k \mathbf{H}_k P_{k|k-1} - P_{k|k-1} \mathbf{H}_k^T \mathbf{K}_k^T + \mathbf{K}_k (\mathbf{H}_k P_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k) \mathbf{K}_k^T \\ &= P_{k|k-1} - P_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k P_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \mathbf{H}_k P_{k|k-1}, \\ &(\text{or : } P_{k|k} = \mathbf{R}_{xx} - \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} \mathbf{R}_{zx}, \text{直接可得}) \end{aligned}$$

卡尔曼滤波示意图：



$$\begin{pmatrix} \text{状态估计值} \\ \text{(滤波更新值)} \end{pmatrix} = \begin{pmatrix} \text{状态的一步预测值} \\ \text{(状态预报值)} \end{pmatrix} + \begin{pmatrix} \text{Kalman} \\ \text{增益阵} \end{pmatrix} \begin{pmatrix} \text{量测的一步预测误差} \\ \text{(新息、量测预报误差)} \end{pmatrix}$$

状态空间模型示例（目标跟踪）

✓ 状态方程

1、近似匀速（CV）运动

连续形式： $\dot{\mathbf{x}}(t) = \mathbf{v}(t)$, $\ddot{\mathbf{x}}(t) = \mathbf{a}(t)$

即将加速度看作是随机扰动（状态噪声），一般假定 $\mathbf{a}(t) = \mathbf{w}_t$
服从零均值白高斯)

$$\text{令 } \mathbf{x}_t = \begin{bmatrix} \mathbf{x}(t) \\ \dot{\mathbf{x}}(t) \end{bmatrix}, \text{ 有 } \dot{\mathbf{x}}_t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x}_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathbf{w}_t$$

采用连续系统离散化的手段

高阶微分方程 $\longrightarrow \dot{x} = Ax + Bu$ 一阶微分方程组

↓ 求解状态方程

$$x(t) = \Phi(t - t_0)x(t_0) + \int_{t_0}^t \Phi(t - \tau)Bu(\tau)d\tau \quad \text{状态方程的解}$$

↓ 离散化

$$x[(k+1)T] = G(T)x(kT) + H(T)u(kT), \text{即 } x_{k+1} = G(T)x_k + H(T)u_k \quad \text{一阶差分方程组}$$

其中

$$G(T) = \Phi(T) = e^{AT},$$

$$H(T) = \left(\int_0^T \Phi(t)dt \right) B = \left(\int_0^T e^{At} dt \right) B$$

$$e^{AT} = \sum_{k=0}^{\infty} \frac{1}{k!} A^k T^k$$

状态空间模型示例

离散形式:

$$\text{令 } \mathbf{x}_k = \begin{bmatrix} \mathbf{x}_k \\ \dot{\mathbf{x}}_k \end{bmatrix}, \text{ 有 } \mathbf{x}_{k+1} = F_k \mathbf{x}_k + \Gamma_k w_k$$

$$\text{其中, } F_k = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \quad \Gamma_k = \begin{bmatrix} T^2/2 \\ T \end{bmatrix}$$

对于三维空间情形

$$\mathbf{x}_k = [x_k, \dot{x}_k, y_k, \dot{y}_k, z_k, \dot{z}_k]$$

$$F_k = \begin{bmatrix} 1 & T & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & T & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & T \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\Gamma_k = \begin{bmatrix} T^2/2 & 0 & 0 \\ T & 0 & 0 \\ 0 & T^2/2 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^2/2 \\ 0 & 0 & T \end{bmatrix}$$

状态空间模型示例

2、近似匀加速（CA）运动

（将加速度看作是随机扰动（状态噪声））

连续形式： $\dot{x}(t)=v(t)$, $\ddot{x}(t)=a(t)$, $\ddot{\ddot{x}}(t)=\dot{a}(t)=w_t$

若取 $x_k = [x_k, \dot{x}_k, \ddot{x}_k]$

$$x_{k+1} = F_k x_k + \Gamma_k w_k$$

$$F_k = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}$$

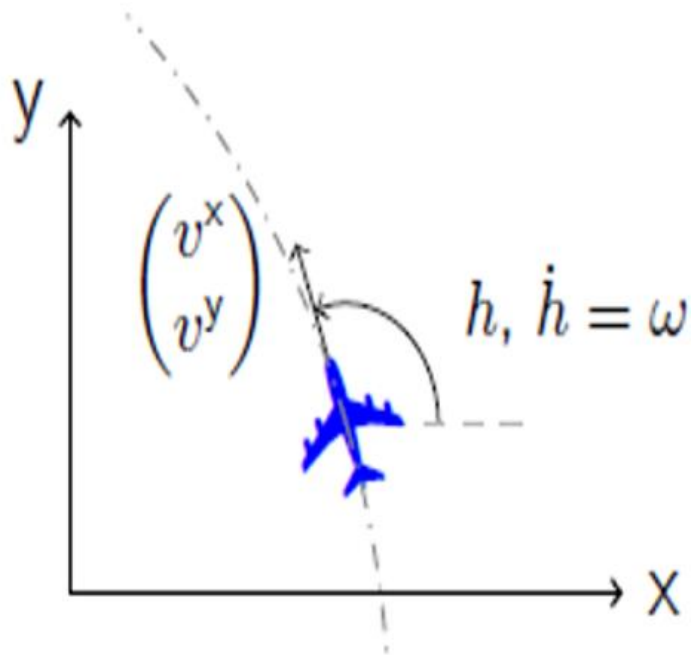
$$\Gamma_k = \begin{bmatrix} T^3/6 \\ T^2/2 \\ T \end{bmatrix}$$

带机动输入

$$x_{k+1} = F_k x_k + \Gamma_k w_k + G_k u_k$$

3、CT---转弯模型 (nearly) coordinated turn

v : 线速度; h : heading 朝向角



Continuous time description

$$\dot{x} = v \cos(h)$$

$$\dot{y} = v \sin(h)$$

from which the following differential equation is obtained

$$\ddot{x} = \frac{d}{dt}\dot{x} = -v\dot{h} \sin(h) = -\omega\dot{y}$$

$$\ddot{y} = \frac{d}{dt}\dot{y} = v\dot{h} \cos(h) = \omega\dot{x}$$

CT的连续时间系统形式:

定义

$$x_1 = x, x_2 = \dot{x}_1 = \dot{x}$$

$$y_1 = y, y_2 = \dot{y}_1 = \dot{y}$$

$$x = [x \quad \dot{x} \quad y \quad \dot{y}]^T$$

考虑在加速度分量上的随机扰动 $w = \begin{bmatrix} w_x \\ w_y \end{bmatrix}$, 有

$$\ddot{x} = -\omega \dot{y} + w_x, \quad \ddot{y} = \omega \dot{x} + w_y$$

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\omega \\ 0 & 0 & 0 & 1 \\ 0 & \omega & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix}$$

CT的离散时间系统形式:

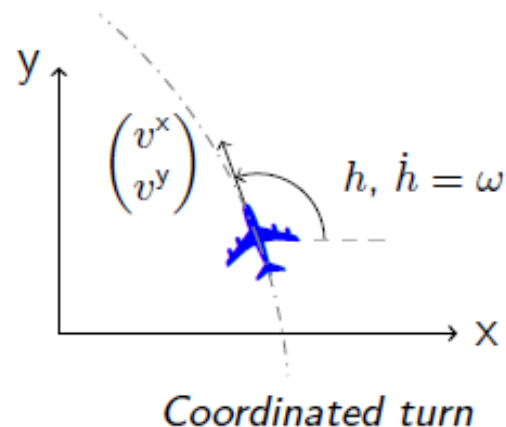
$$\underline{x_k} = [\underline{x_k} \quad \dot{\underline{x_k}} \quad \underline{y_k} \quad \dot{\underline{y_k}}]^T$$

$$\underline{x_{k+1}} = F_k \underline{x_k} + \Gamma_k \underline{w_k}$$

$$\underline{x_k} = [\underline{x_k} \quad \dot{\underline{x_k}} \quad \underline{y_k} \quad \dot{\underline{y_k}}]^T$$

$$F_k = \begin{bmatrix} 1 & \frac{\sin(\omega t)}{\omega} & 0 & -\frac{1 - \cos(\omega t)}{\omega} \\ 0 & \cos(\omega t) & 0 & -\sin(\omega t) \\ 0 & \frac{1 - \cos(\omega t)}{\omega} & 1 & \frac{\sin(\omega t)}{\omega} \\ 0 & \sin(\omega t) & 0 & \cos(\omega t) \end{bmatrix}$$

$$\Gamma_k = \begin{bmatrix} T^2 / 2 & 0 \\ T & 0 \\ 0 & T^2 / 2 \\ 0 & T \end{bmatrix}$$



状态空间模型示例

✓ 量测方程（以线性为例）

若已知状态向量为

$$\mathbf{x}_k = [x_k, \dot{x}_k, y_k, \dot{y}_k, z_k, \dot{z}_k]$$

假设量测仅包含位置信息。

则

$$z_k = H_k x_k + v_k$$

其中，

$$H_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

例9 用如下差分方程产生一个随机序列

$$x(n) = 1.74x(n-1) - 0.81x(n-2) + \underline{v(n)}$$

观测方程 $y(n) = x(n) + v_2(n)$

$v(n), v_2(n)$ 分别是方差0.04和9的白噪声。

$$\underline{x(n)} = \begin{bmatrix} x(n-1) \\ \underline{x(n)} \end{bmatrix}$$

等效状态演化方程为：
$$\mathbf{x}(n) = \begin{bmatrix} 0 & 1 \\ -0.81 & 1.74 \end{bmatrix} \mathbf{x}(n-1) + \begin{bmatrix} 0 \\ v(n) \end{bmatrix}$$

观测方程为：
$$y(n) = H\mathbf{x}(n) + v_2(n)$$

状态空间模型参数为

$$F = \begin{pmatrix} 0 & 1 \\ -0.81 & 1.71 \end{pmatrix} \quad Q_1(n) = \begin{pmatrix} 0 & 0 \\ 0 & 0.04 \end{pmatrix}$$

$$H = (0 \quad 1) \quad Q_2(n) = 9$$

实验的方法是：

给定双初值， $x(-1) = x(0) = 0$

由原始状态演化方程

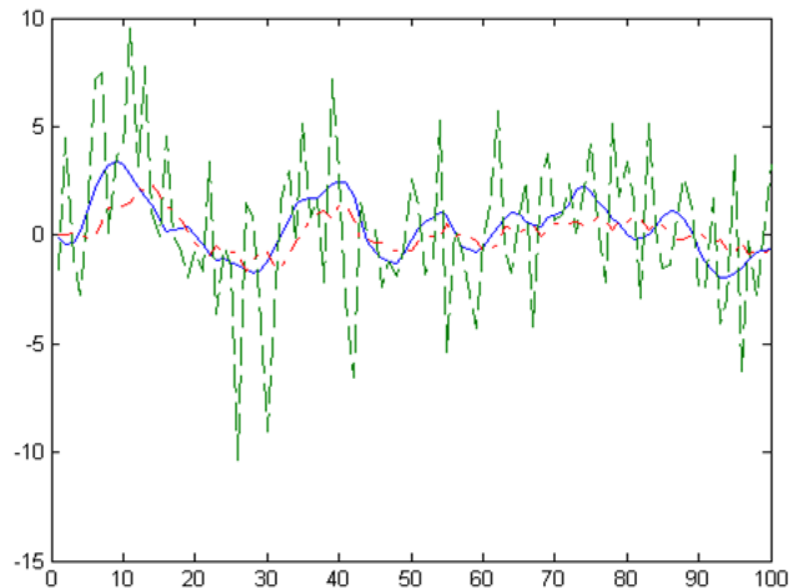
$$x(n) = 1.74x(n-1) - 0.81x(n-2) + v(n)$$

产生随机序列的一次实现。利用观测方程 $y(n) = x(n) + v_2(n)$ 产生观测数据。

在滤波过程中，采用新的状态空间模型：

$$\mathbf{x}(n) = \begin{bmatrix} 0 & 1 \\ -0.81 & 1.74 \end{bmatrix} \mathbf{x}(n-1) + \begin{bmatrix} 0 \\ v(n) \end{bmatrix}$$

$$y(n) = H\mathbf{x}(n) + v_2(n)$$



卡尔曼滤波的发散现象

1) 发散现象

当量测数目不断增加时，按滤波方程计算出的估计误差协方差阵趋于零或某一稳定值，但实际的估计误差却有可能越来越大，使滤波器逐渐失去估计作用，这种现象称为滤波器的发散。

2) 发散原因

滤波发散

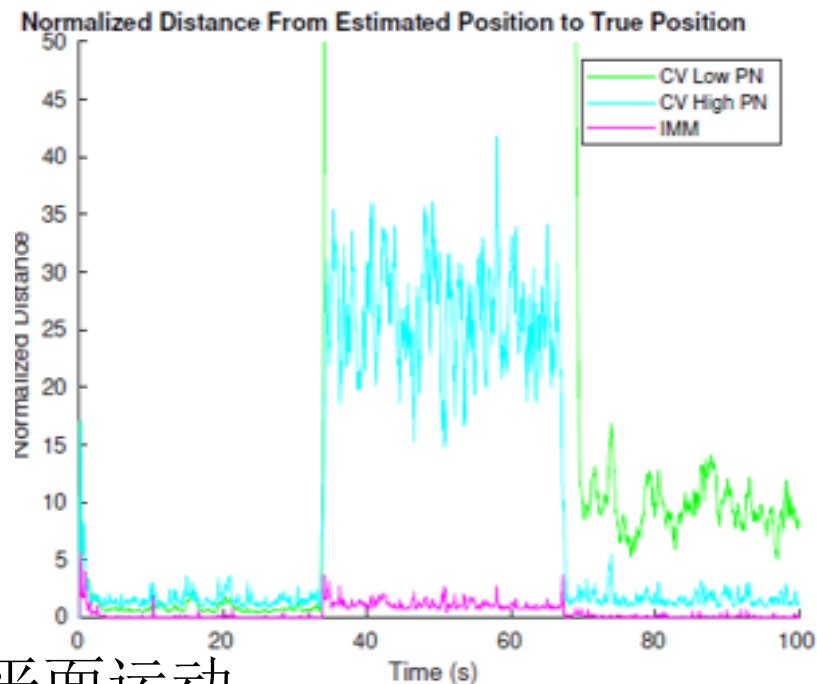
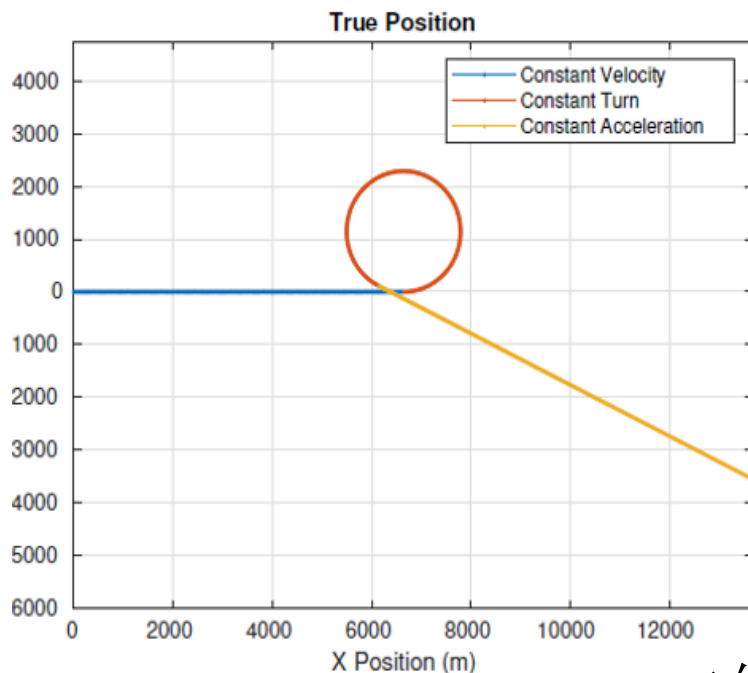
- 系统状态演化模型不够精确；
- 扰动噪声矢量和观测噪声矢量统计特性描述不准确；

计算发散

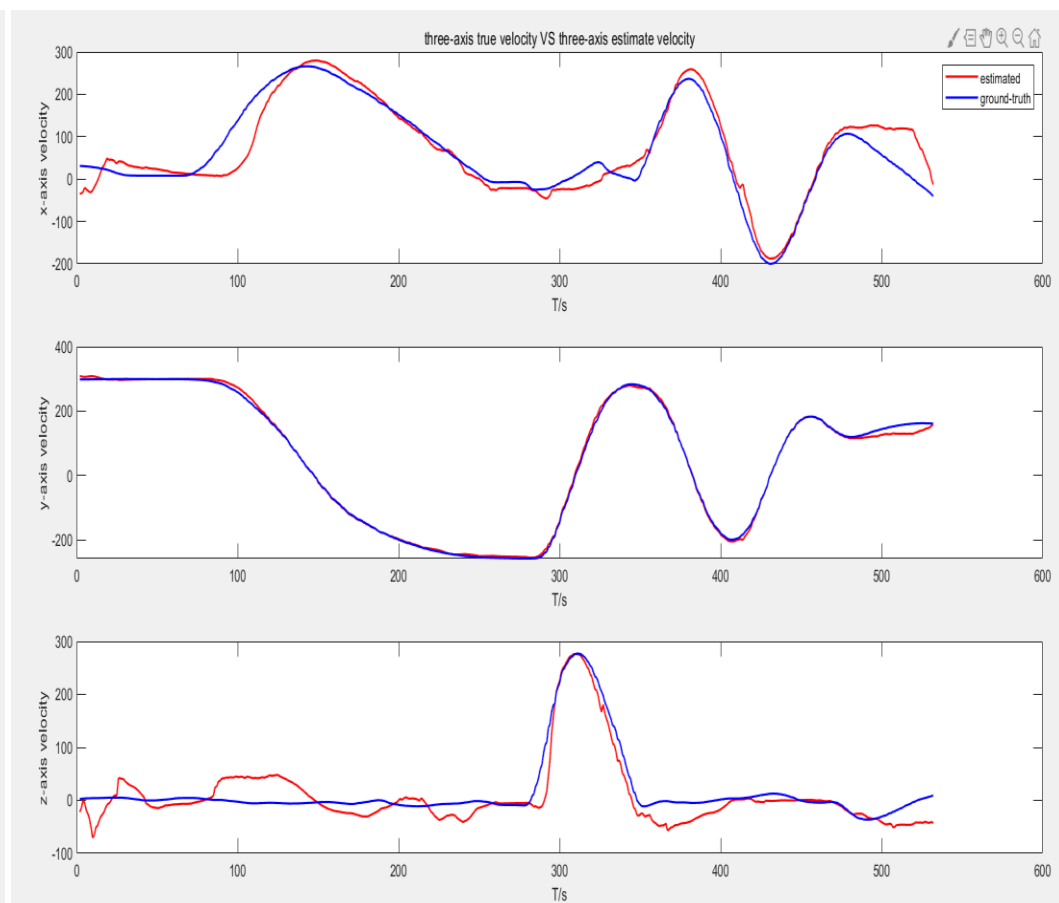
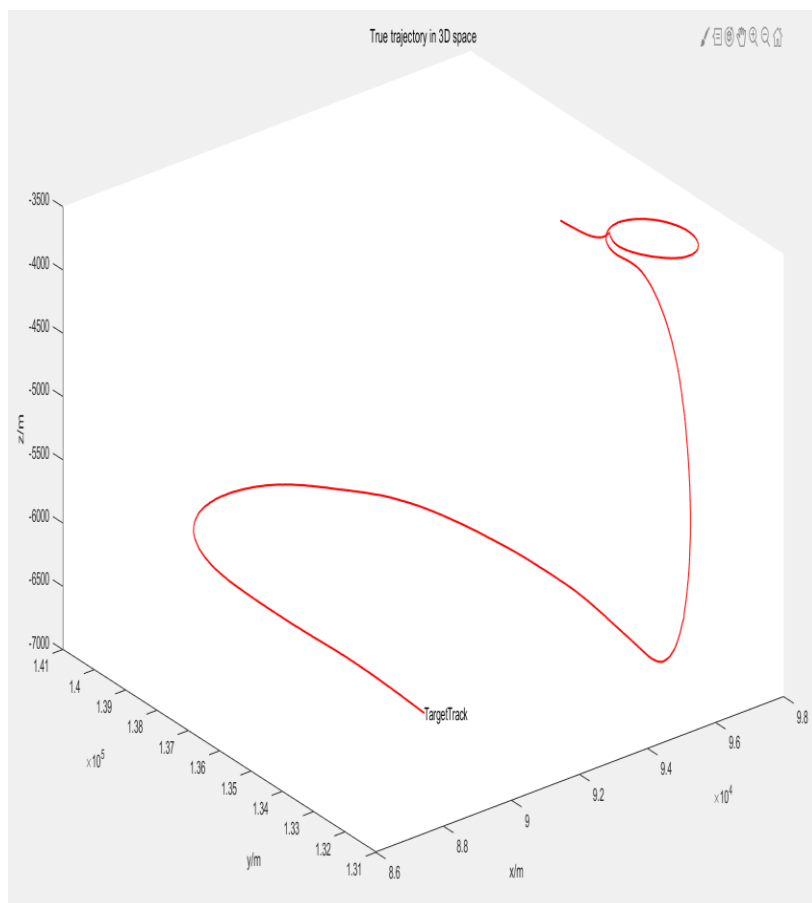
- 计算中的有限字长效应，舍入误差累积；
- 其他各种误差影响等。

3) 克服办法

- 建立尽可能精确的系统状态演化模型，如匀速直线运动模型，匀加速直线运动模型，机动目标当前统计模型等。
- 噪声模型（过程噪声和量测噪声）的统计特性描述合理。
- 自适应滤波法（例如IMM-交互式多模型算法）。



二维平面运动



三维空间运动

• 3、估计滤波性能评价

◆ 绝对误差度量

例如：

- 均方根误差 (RMSE-Root Mean Square Error)

$$RMSE(\hat{x}) = \left(\frac{1}{M} \sum_{i=1}^M \|\tilde{x}_i\|^2 \right)^{1/2} \quad \|\tilde{x}_i\| = \left(\tilde{x}_i' \tilde{x}_i \right)^{1/2}, \tilde{x}_i = \hat{x}_i - x$$

其中， M 为monte-carlo仿真的次数，其中 \tilde{x}_i 为第 i 次仿真得到的误差向量。

不同属性需要分开度量：

$$position \quad RMSE(\hat{x}) = \left(\frac{1}{M} \sum_{i=1}^M \left[\|\tilde{x}_i\|^2 + \|\tilde{y}_i\|^2 \right] \right)^{1/2}$$

$$velocity \quad RMSE(\hat{x}) = \left(\frac{1}{M} \sum_{i=1}^M \left[\|\dot{\tilde{x}}_i\|^2 + \|\dot{\tilde{y}}_i\|^2 \right] \right)^{1/2}$$

Monte-carlo 方法

冯.诺依曼（美籍匈牙利数学家、计算机科学家、物理学家）在洛斯阿拉莫斯国家实验室为核武器计划工作时，发明了MC方法。

也称为统计模拟方法，是通过从概率模型的随机抽样进行近似数学计算的方法。

核心：随机抽样。

包括直接抽样法，接受拒绝抽样法，重要性抽样法等。



John von Neumann
1903~1957)

- 平均欧式误差 (*Average Euclid Error-AEE*)
(Mean Absolute Error-MAE)

AEE定义为:

$$AEE(\hat{x}) = \frac{1}{M} \sum_{i=1}^M \|\tilde{x}_i\|$$

从几何上讲, AEE是欧式空间中真值与估值间的算数平均距离。

RMSE和AEE的大小都主要受大误差项控制, 但AEE受大误差影响相对较小。例如, 有99个约为1的项和一个约为400的项, AEE约为5, RMSE约为40。我们可以看出RSME过于关注大误差而几乎忽略小误差, 有些情形下这是不合理的甚至是不能接受的(比如当由于滤波发散而产生了几个大的误差)。

- 调和平均误差 (Harmonic Average Error, HAE)

某种意义上, RMSE和AEE是悲观的度量 (pessimistic measures)。因为, 如前所述, 大误差对它们具有统治性, 它们更关注不好的结果。但有些情况下我们需要关注好的结果, 这时候就要用到乐观的度量 (optimistic measure), 提出基于调和平均数 (harmonic average) 的HAE:

$$HAE(\hat{x}) = \left(\frac{1}{M} \sum_{i=1}^M \|\tilde{x}_i\|^{-1} \right)^{-1} = \frac{M}{1/\|\tilde{x}_1\| + \dots + 1/\|\tilde{x}_M\|}$$

倒数平均数

$$i.e. \frac{1}{HAE(\hat{x})} = \frac{1}{M} \left(\frac{1}{\|\tilde{x}_1\|} + \dots + \frac{1}{\|\tilde{x}_M\|} \right)$$

显然, HAE大小主要受小误差项控制。由于HAE主要注重好的行为, 所以它适用于关注好的行为的应用中。

- 几何平均误差 (Geometric Average Error, GAE)

由于RMSE, AEE和HAE都受极端误差项控制, 所以它们是不平衡的量度。但对于许多问题, 我们希望大误差 (不好的行为) 能被一个充分小的误差 (好的行为) 平衡, 反之亦然。所以引入基于几何平均数的GAE:

$$GAE(\hat{x}) = \left(\prod_{i=1}^M \|\tilde{x}_i\| \right)^{1/M}$$

◆ 相对误差度量

相对均方根误差、相对欧式误差、相对调和误差、相对几何平均误差。

例如:

$$RMSRE(\hat{x}) = \frac{\left(\frac{1}{M} \sum_{i=1}^M \|\tilde{x}_i\|^2 \right)^{1/2}}{\|x\|} = \frac{\left(\frac{1}{M} \sum_{i=1}^M \|\tilde{x}_i\| \right)^{1/2}}{\left(\frac{1}{M} \sum_{i=1}^M \|\hat{x}_i\|^2 \right)^{1/2}}$$

(RMSRE-Root Mean Square Relative Error)

练习题：卡尔曼滤波MATLAB 练习

设目标运动状态方程和测量方程分别为：

$$x_{k+1} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} T^2 / 2 \\ T \end{bmatrix} w_k, \quad z_k = \begin{bmatrix} 1 & 0 \end{bmatrix} x_k + v_k$$

其中： $w_k \sim N(0, 10^{-1})$, $v_k \sim N(0, 1)$, $x_k = \begin{bmatrix} x_k & \dot{x}_k \end{bmatrix}^T$

仿真时长100秒，采样周期 $T=1$ 秒。

初始状态自定，编程实现KF，并绘制RMSE。

本章小结

1. 最优估计

无偏估计、渐近无偏估计

估计方法：极大后验、极大似然、加权最小二乘、
最小均方误差、线性最小均方误差

2. 最优滤波

卡尔曼滤波的基本思想和递推方法