第五章 信号估计（）

内容提要

- ✓ 5.1 概述;
- ✓ 5.2 参数估计—最大似然估计（ML）;
- ✓ 5.3 参数估计—最大后验概率估计（MAP）;
- ✓ 5.4 贝叶斯估计
- ✓ 5.5 估计量的性质
- ✓ 5.6 最小二乘估计
- ✓ 5.7 线性最小均方误差估计
- ✓ 5.8 线性递推估计
- ✓ 5.9 正交原理

- ✓ 5.10 波形的线性最小均方误差估计
- ✓ 5.11 维纳滤波器
- ✓ 5.12 卡尔曼滤波

5.1 概言

一、什么是估计：

当信号肯定存在的情况，

测量值 $x(t) \xrightarrow{\text{某一准则}}$ 确定信号的数量特性。

二、估计模型：

$$x(t) = s(t) + n(t)$$

源 → 观察空间 → 某一准则下优化计算 → 估计量输出

三、发展

1795 年，高斯（18 岁）提出最小二乘估计，用于预测行星的轨道。

费希尔（fisher）发展了最大似然预测理论。

………………………

四、分类：

参量估计：

- 最大似然估计（ML）
- 最大后验概率估计（MAP）
- 贝叶斯估计（最小平均风险估计）
- 最小二乘估计（LS）
- 线性最小均方误差估计
- 线性递推估计

波形估计：

- 线性最小均方误差估计
- 维纳滤波器
- 卡尔曼滤波器

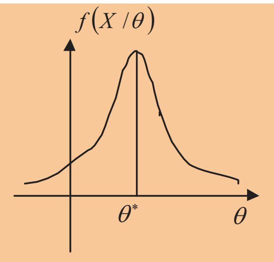
5.2 参数估计—最大似然估计（ML）

一、基本思想

$$x_i = \theta + n_i,$$

其中 θ 为源发出的信号，$i = 1,2\cdots N$

N 次观察矢量：$X = \{x_1, x_2 \cdots x_N\}$ 的似然函数 $f(X/\theta)$ 表示 θ 出现条件下，N 个观测值的联合密度函数，是 θ 的条件密度函数。



最大似然方程：

$$\frac{\partial f(X/\theta)}{\partial \theta} = 0$$
$$\Downarrow$$
$$\theta = \theta^* = \hat{\theta}_{ML}$$

表示在 $\theta = \theta^*$ 处似然函数最大，换一句话说 $\theta = \theta^*$ 的机会最高。

将 θ^* 作为 θ 最大似然准则下的估计值—最大似然估计.

二、准则的应用

$$x_i = \theta + n_i, \qquad i = 1,2\cdots N$$

单样本的似然函数： $f(x_i/\theta)$

样本矢量的似然函数： $f(X/\theta) = \prod_{i=1}^{N} f(x_i/\theta)$

两边取对数： $\ln f(X/\theta) = \sum_{i=1}^{N} \ln f(x_i/\theta)$

解最大似然方程：$\frac{\partial \ln f(X/\theta)}{\partial \theta} = \sum_{i=1}^{N}\left[f'(x_i/\theta) \bullet \frac{1}{f(x_i/\theta)} \right] = 0$

若有 p 个待估计的参数，则解联立方程组 $\frac{dl}{d\theta_k} = 0$， $k = 1,2\cdots p$.

例：对参数 θ 进行 N 测量，各次测量误差不相同，且均服从 $N(0,\sigma^2)$，求最大似然估计。

解：$x_i = \theta + n_i, \qquad i = 1,2\cdots N$

单样本的似然函数：

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

样本矢量的似然函数：

$$f(X/\theta)=\underbrace{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N}_{F}\exp\left[-\sum_{i=1}^{N}\frac{(x_i-\theta)^2}{2\sigma^2}\right]$$

$$\ln f(X/\theta)=\ln F-\frac{1}{2\theta^2}\sum_{i=1}^{N}(x_i-\theta)^2$$

$$\frac{\partial \ln f(X/\theta)}{\partial\theta}=0$$

$$\Rightarrow \theta=\theta^*=\hat{\theta}_{ML}=\frac{1}{N}\sum_{i=1}^{N}x_i$$

（观测样本均值）

最大似然估计 $\hat{\theta}_{ML}$ 的方差为：

$$E\left[(\theta-\hat{\theta}_{ML})^2\right]=E\left[(\theta-\frac{1}{N}\sum_{i=1}^{N}x_i)^2\right]$$

$$=E\left\{\left[\theta-\frac{1}{N}\sum_{i=1}^{N}(\theta+n_i)\right]^2\right\}=\frac{1}{N}\sigma_n^2$$

（满足均值中心极限定理）

例：考虑在均值为零，方差为 σ_n^2 的加性高斯白噪声 n 中接收信号 S，单次观测方程为 $x=s+n$，求信号的最大似然估计。

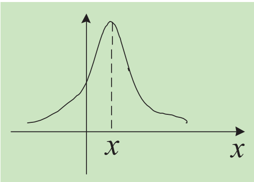
解：单样本的似然函数：

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

利用最大似然估计方程：

$$\frac{\partial \ln f(x/s)}{\partial s}=\frac{1}{\sigma_n^2}(x-s)\Big|_{s=\hat{s}_{ML}}=0$$

$\hat{s}_{ML}=x$ 单次观察时实测数据即为最大似然估计



三、准则的应用场合

■ 由于最大似然估计没有利用先验概率，其一般性能比利用先验信息的贝叶斯估计要差一些。

■当 θ 为未知非随机参量时

为随机参量，但不知其先验概率

为计算（获得）后验概率比计算（获得）似然函数要困难

最大似然估计不失为一种性能优良实用的估计

■ 对于绝大多数实用的最大似然估计，当观察数据足够多时，其性能是最优的。

5.3 参数估计—最大后验概率估计（MAP）

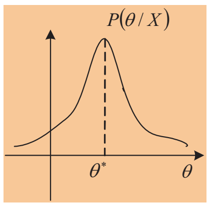
一、基本思想

$$x_i=\theta+n_i,\quad i=1,2\cdots N,$$

观测样本矢量 $X=\{x_1,x_2\cdots x_N\}$。

单样本后验概率为 $P(\theta/x_i)$。

样本矢量的后验概率为 $P(\theta/X)=\frac{P(\theta)f(X/\theta)}{f(X)}$ 是 θ 的函数。



最大后验估计方程：

$$\frac{\partial P(\theta/X)}{\partial\theta}=0$$

$$\Downarrow$$

$$\theta=\theta^*=\hat{\theta}_{MAP}$$

表示在 $\theta=\theta^*=\hat{\theta}_{MAP}$ 处后验概率最大，换一句话说 $\theta=\theta^*=\hat{\theta}_{MAP}$ 最有可能发生。

将 θ^* 作为 θ 的最大后验概率准则下的估计值—最大后验概率估计。

注意：

当 $P(\theta)$ 为均匀分布，即为常数时，$P(H_i/X)=\frac{P(H_i)f(X/H_i)}{f(X)}$

最大后验估计 $\xrightarrow{\text{先验概率为均匀分布时，等价于}}$ 最大似然估计。

例：对 θ 进行一次测量（单次观测），$x_1=\theta+n_1$，$n:N(0,\sigma_n^2)$，

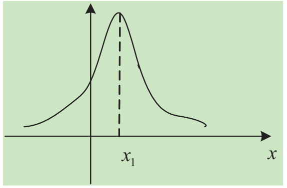
θ：均匀分布的随机变量，

$$f(\theta)=\begin{cases}\frac{1}{2\theta_M} & -\theta_M\le\theta\le\theta_M\\ 0 & else\end{cases}$$

求 $\hat{\theta}_{ML}$，$\hat{\theta}_{MAP}$。

解：（1）ML

x 的似然函数：$f(x/\theta)=\frac{1}{\sqrt{2\pi\delta_n^2}}\exp\left[-\frac{(x_1-\theta)^2}{2\delta_n^2}\right]$



$\hat{\theta}_{ML}=x_1$ 单次观察时实测数据即为最大似然估计

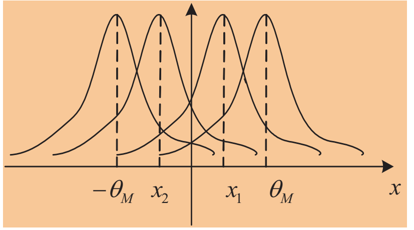
（2）MAP

$$P(\theta/X)=\frac{P(\theta)f(X/\theta)}{f(X)}$$

在 $x\in[-\theta_M,\theta_M]$内，$P(\theta)$为均匀分布，因而$P(\theta/x)$峰值和$f(x/\theta)$相同。即估计值皆为该峰值，也就是单次观察的实测数据。

在$x\in(\theta_M,\infty)$内，$P(\theta/x)$最大概率只能在θ_M处。

在$x\in(-\infty,-\theta_M)$内，$P(\theta/x)$最大概率只能在$-\theta_M$处。



$$\therefore\quad \hat{\theta}_{MAP}=\begin{cases} x_1 & -\theta_M\le x_1\le\theta_M \\ \theta_M & x_1>\theta_M \\ -\theta_M & x_1<-\theta_M \end{cases}$$

5.4 贝叶斯估计

一、估计误差：

$x=\theta+n$，x 为实测数据量，

$\hat{\theta}(X)$为基于观察数据 X 的 θ 的估计值。

$\tilde{\theta}=\theta-\hat{\theta}(X)$为 θ 的估计误差。

二、损失函数/代价函数

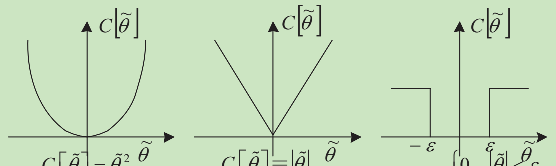
不同的估计精度将有不同的代价。

$C[\theta,\hat{\theta}]$，一个双变量的函数，

在许多情况下为 $C[\tilde{\theta}]$，仅与估计误差有关。

（1）一般地 $C[\tilde{\theta}]\ge 0$ （非负）（2）误差为零时代价最小

（3）常见的代价函数：



$C[\tilde{\theta}]=\tilde{\theta}^2$ 误差平方代价函数

$C[\tilde{\theta}]=|\tilde{\theta}|$ 误差绝对值代价函数

$C[\tilde{\theta}]=\begin{cases}0 & |\tilde{\theta}|<\varepsilon \\ 1 & |\tilde{\theta}|\ge\varepsilon\end{cases}$ 误差阱型代价函数 也叫均匀代价函数

三、贝叶斯估计：

作一般的讨论，待估计的参量为$\theta_1,\theta_2,\cdots\theta_M\to$矢量$\bar{\theta}$。

观测矢量X与待估计矢量$\bar{\theta}$的联合概率密度函数：

$$f(X,\bar{\theta})=\underbrace{f(X/\bar{\theta})}_{\text{似然函数}}\underbrace{p(\bar{\theta})}_{\text{先验概率}}=\underbrace{f(\bar{\theta}/X)}_{\text{后验概率}}f(X)$$

1、条件风险：

给定观察矢量X后，估计与$\bar{\theta}$ 相联系的平均代价。

$$R(\hat{\bar{\theta}}/X)=\int_{\{\bar{\theta}\}}C(\bar{\theta},\hat{\bar{\theta}})\underbrace{f(\bar{\theta}/X)}_{\text{后验概率}}d\bar{\theta}$$

2、风险：

平均代价总和，是条件风险对所有观测值X求平均。

$$\begin{aligned} R(\hat{\bar{\theta}}) &= E\left[R(\hat{\bar{\theta}}/X)\right] \\ &= \int_{\{X\}}R(\hat{\bar{\theta}}/X)f(X)dX \\ &= \int_{\{X\}}\int_{\{\bar{\theta}\}}C\left[\bar{\theta},\hat{\bar{\theta}}\right]f(\bar{\theta}/X)f(X)d\bar{\theta}dX \\ &= \int_{\{X\}}\int_{\{\bar{\theta}\}}C\left[\bar{\theta},\hat{\bar{\theta}}\right]\underbrace{f(X,\bar{\theta})}_{\text{联合概率密度}}d\bar{\theta}dX \end{aligned}$$

3、Bayes 估计：

使风险最小的估计为 Bayes 估计。

即 $\min R(\hat{\bar{\theta}})$ （风险）

给定 X，$f(X)$与θ无关。

等价于 $\min R(\hat{\bar{\theta}}/X)$ （条件风险）

$$\frac{\partial}{\partial\hat{\bar{\theta}}}R(\hat{\bar{\theta}}/X)=0$$

四、不同代价函数下的贝叶斯估计：

1、平方误差代价函数下的 Bayes 估计

（以单参量为例）

条件风险：$R(\hat{\theta}/X)=\int_{-\infty}^{\infty}(\theta-\hat{\theta})^2 f(\theta/X)d\theta$

↓ （恰好均方误差表达式：误差平方后的均值）

$\min(R(\hat{\theta}/X))$

↓

$$\frac{\partial R(\hat{\theta}/X)}{\partial\hat{\theta}}=2\hat{\theta}\underbrace{\int_{-\infty}^{\infty}f(\theta/X)d\theta}_{1}-2\int_{-\infty}^{\infty}\theta f(\theta/X)d\theta=0$$

$$\therefore \hat{\theta}_{MS}(X)=\int_{-\infty}^{\infty}\theta f(\theta/X)d\theta=E[\theta/X]$$

（条件均值，后验均值）

$$\because\quad \frac{\partial^2 R(\hat{\theta}/X)}{\partial^2\hat{\theta}}=2>0,\quad\therefore$$

$\hat{\theta}_{MS}(X)$ 确是是条件风险最小的估计量，即 **Bayes** 估计量。

结论：

a. 贝叶斯估计量 $\xrightarrow{\text{平方误差代价函数下}}$ 就是给定 X 时，θ 的条件均值，后验均值。

b. 风险 $\xrightarrow{\text{平方误差代价函数下}}$ 就是均方误差。

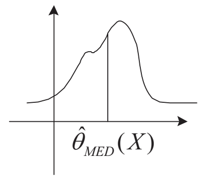
Bayes 估计使风险最小，也就是使均方误差最小。因此该估计又称最小均方误差估计

2．绝对误差代价函数下的 Bayes 估计

$$R(\hat{\theta}/X)=\int_{-\infty}^{\infty}\left|\theta-\hat{\theta}\right|f(\theta/X)d\theta$$
$$=\int_{-\infty}^{\hat{\theta}}(\hat{\theta}-\theta)f(\theta/X)d\theta+\int_{\hat{\theta}}^{\infty}(\theta-\hat{\theta})f(\theta/X)d\theta$$

$$\min R(\hat{\theta}/X),\quad \frac{\partial R(\hat{\theta}/X)}{\partial\hat{\theta}}=0$$

$$\int_{-\infty}^{\hat{\theta}}f(\theta/X)d\theta=\int_{\hat{\theta}}^{\infty}f(\theta/X)d\theta$$



（后验概率密度的中值 $\hat{\theta}_{MED}(X)$）

结论：

贝叶斯估计量 $\xrightarrow{\text{绝对误差代价函数下}}$ 就是后验中值

也称条件中值估计。

3．阱型误差代价函数下的 Bayes 估计

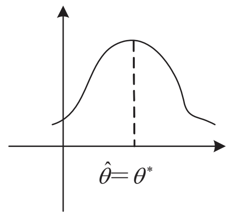
$$R(\hat{\theta}/X)=\int_{-\infty}^{\hat{\theta}-\varepsilon}f(\theta/X)d\theta+\int_{\hat{\theta}+\varepsilon}^{\infty}f(\theta/X)d\theta$$
$$=1-\int_{\hat{\theta}-\varepsilon}^{\hat{\theta}+\varepsilon}f(\theta/X)d\theta$$

$$\min R(\hat{\theta}/X)$$

当 ε 很小且不为0，

第二项最大 $\xrightarrow{\text{选择}\hat{\theta}\text{等于后验概率密度的众数}}$

后验概率密度最大点(众数)



结论：

贝叶斯估计量 $\xrightarrow{\text{阱型误差代价函数下}}$ 就是后验概率密度的众数，最大后验概率密度点，该估计就是最大后验概率估计。

$\hat{\theta}_{MAP}(X)$

五、不同代价函数下的贝叶斯估计量的不变性：

贝叶斯估计：为了计算方便，有时需要选择不同的代价函数，那么对不同的代价函数，估计量是否一样呢？

下面讨论什么类型的 $\begin{cases}\text{代价函数}\\ \text{后验概率密度函数}\end{cases}$，能使估计量具有不变性？

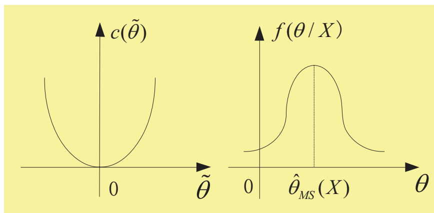
约束条件一：

a 代价函数为对称 $C(\tilde{\theta})=C(-\tilde{\theta})$

b 代价函数为下凸函数 $C[b\theta_1+(1-b)\theta_2]\le bC(\theta_1)+(1-b)C(\theta_2)$

$b\in[0,1]$，　　对任意 θ_1，θ_2

c 后验概率密度对其条件均值对称



证明见相关教材。

该条件中的下凸代价函数排除了均匀代价函数（阱型代价函数）。

为了包括非下凸代价函数，给出约束条件 2：

约束条件二：

a 代价函数为对称 $C(\tilde{\theta})=C(-\tilde{\theta})$

b 代价函数为非减函数 $C(\tilde{\theta}_1)\ge C(\tilde{\theta}_2),\ \left|\tilde{\theta}_1\right|\ge\left|\tilde{\theta}_2\right|$

c 后验概率密度是对称于条件均值的单峰函数

在这两种代价函数和后验概率密度函数条件下，最小均方误差估计、条件中值估计和最大后验概率估计等价，进一步称最小均方误差估计就是贝叶斯估计。

这就是最佳估计的不变性。

例：对参数 θ 进行估计，N 次观察互相独立。

$$x_i = \theta + n_i \qquad i = 1,2\cdots N$$

设 θ 为高斯分布，$N(0,\sigma_\theta^2)$，n 服从 $N(0,\sigma_n^2)$。

解：似然函数：

$$f(X/\theta) = \prod_{i=1}^{N} f(x_i/\theta) = \prod_{i=1}^{N}\left\{\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left[-\frac{(x_i-\theta)^2}{2\sigma_n^2}\right]\right\}$$

$$f(\theta) = \frac{1}{\sqrt{2\pi\delta_\theta^2}}\exp\left[-\frac{\theta^2}{2\sigma_\theta^2}\right], \quad f(\theta/X) = \frac{f(\theta)f(X/\theta)}{f(X)}$$

后验密度：

$$f(\theta/X) = \underbrace{\frac{1}{f(X)}\left(\frac{1}{\sqrt{2\pi}\delta_n}\right)^N\left(\frac{1}{\sqrt{2\pi}\sigma_\theta}\right)}_{K(X)}\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{N}\left(\frac{x_i-\theta}{\sigma_n}\right)^2 + \frac{\theta^2}{\sigma_\theta^2}\right]\right\}$$

$$= K(X)\exp\left\{\frac{-\left[\theta - \frac{\delta_m^2}{\delta_n^2}\sum_{i=1}^{N}x_i\right]^2}{2\sigma_m^2}\right\}$$

$$\left(\sigma_m^2 = \frac{\sigma_n^2\sigma_\theta^2}{N\sigma_\theta^2+\sigma_n^2}\right)$$

最大后验概率估计：$\hat{\theta}_{MAP} = \dfrac{\sigma_m^2}{\sigma_n^2}\sum_{i=1}^{N}x_i$

平方误差代价函数条件下的贝叶斯估计为条件均值，

$$\hat{\theta}_{ms} = \hat{\theta}_{MAP}$$

绝对误差代价函数条件下的贝叶斯估计为后验中值，

$$\hat{\theta}_{abs} = \hat{\theta}_{MAP}$$

阱型误差代价函数条件下的贝叶斯估计为后验最大点（众数），即为最大后验概率估计：

$$\hat{\theta}_{MAP} = \hat{\theta}_{MAP}$$

5.5 估计量的性质

待估计量为 θ，实测数据 N 个，$x_1, x_2\cdots x_N$

$\hat{\theta}_N = \hat{\theta}_N(X)$（估计器）

待估计量也是随机变量，它具有许多统计特性。

一、估计的无偏性

我们希望：

$E(\hat{\theta}_N) = \theta$，　对于非随机参量

$E(\hat{\theta}_N) = E(\theta)$，对于随机参量

无偏性：讨论估计量的均值问题

1、条件无偏估计：

非随机参量估计：

$$E\{\hat{\theta}_N\} = \int_X \hat{\theta}_N(X)f(X/\theta)dX = \theta$$

随机参量估计：

$$E\{\hat{\theta}_N\} = \int_\theta\int_X \hat{\theta}_N(X)f(X,\theta)dXd\theta = E(\theta)$$

2、渐进无偏估计：

非随机参量估计：

$$\lim_{N\to\infty} E\{\hat{\theta}_N\} = \theta$$

随机参量估计：

$$\lim_{N\to\infty} E\{\hat{\theta}_N\} = E(\theta)$$

3、有偏估计：

非随机参量估计：

$$E\{\hat{\theta}_N\} - \theta = b_N(\theta) \neq 0$$

随机参量估计：

$$E\{\hat{\theta}_N\} - E(\theta) = b_N(\theta) \neq 0$$

$b_N(\theta)$ 为估计偏差，若 $b_N(\theta)$ 不依赖于 θ，则可用一些简单的方法将它去掉。（如直流部分，去直流）。

例：（3.2 中例）$x_i = \theta + n_i$，$i = 1,2\cdots N$，n 服从 $N(0,\sigma^2)$，求 ML。

解：

$$\hat{\theta}_{ML} = \frac{1}{N}\sum_{i=1}^{N}x_i$$

$$E\{\hat{\theta}_{ML}\} = \frac{1}{N}\sum_{i=1}^{N}E[x_i] = \frac{1}{N}\times N\times\theta = \theta$$，为无偏估计。

二、估计一致性

我们希望随观察次数的增加，估计质量要进一步提高；

↓

即　估计值趋近于被估计量的真值（非随机参量）

　　估计值趋近于被估计量均值（随机参量）

↓

估计的均方误差不断减小。

1、　一致估计：

当观测样本数 N 增加时，估计量在其真值附近越来越集中，即方差越来越小。ε 为任意小的正数

$$\lim_{N\to\infty} P\left\{\left|\hat{\theta}_N - \theta\right| > \varepsilon\right\} = 0$$（非随机参量估计）

$\hat{\theta}_N$ 依概率收敛于 θ

$$\lim_{N\to\infty} P\left\{\left|\hat{\theta}_N - E(\theta)\right| > \varepsilon\right\} = 0$$（随机参量估计）

$\hat{\theta}_N$ 依概率收敛于 $E(\theta)$

2、无条件一致估计：

$$\lim_{N\to\infty} P\left\{ \left| E\left(\hat{\theta}_N\right)-\theta \right| > \varepsilon \right\} = 0$$

3、均方一致（均方收敛）估计：

$$\lim_{N\to\infty} E\left[\left(\hat{\theta}_N-\theta\right)^2\right] = 0$$

例：（3.2 中例）$x_i=\theta+n_i$， $i=1,2\cdots N$， n服从$N(0,\sigma^2)$

例：

$$\hat{\theta}_{ML}=\frac{1}{N}\sum_{i=1}^{N}x_i$$

可以证明，$\hat{\theta}_{ML}$服从均值为θ，方差为$\frac{\sigma^2}{N}$的高斯分布。$N\to\infty$时，$\hat{\theta}_{ML}$是θ的一致无偏估计。

统计学证明(均值中心极限定理)：无论什么样的总体分布，其样本均值都是总体均值的一致无偏估计。

三、有效估计与克拉美－罗不等式

无偏估计：估计量的均值＝参量的真值。

但不能说无偏估计就是好估计，如果方差很大，该估计误差仍可能很大。

估计量的方差愈小愈好，则估计量取其均值附近值的概率愈大，但其方差不会无穷小，有一个最好的界限。

1、估计的有效性

对于被估计量θ的任意无偏估计量$\hat{\theta}_1$和$\hat{\theta}_2$，若估计的均方误差满足

$$E\left[\left(\hat{\theta}_1-\theta\right)^2\right] < E\left[\left(\hat{\theta}_2-\theta\right)^2\right]$$

则称估计量$\hat{\theta}_1$比$\hat{\theta}_2$有效。

如果被估计量θ的无偏估计量$\hat{\theta}$的均方误差小于其他任意无偏估计量的均方误差，则称该估计为最小均方误差无偏估计量

2、克拉美－罗不等式：

$$E\left\{\left[\hat{\theta}_N-E\left(\hat{\theta}_N\right)\right]^2\right\} \ge \frac{\left[1+b_N'(\theta)\right]^2}{I_N(\theta)}$$

$$I_N(\theta)=E\left\{\frac{\partial}{\partial\theta}\ln f(X/\theta)\right\}^2 = -E\left\{\frac{\partial^2}{\partial\theta^2}\ln f(X/\theta)\right\}$$

参考刘有恒 P295

$$b_N'(\theta)=E\left\{\left[\hat{\theta}_N-E\left(\hat{\theta}_N\right)\right]\frac{\partial}{\partial\theta}\left[\ln f(X/\theta)\right]\right\}$$

$$b_N'(\theta)=E\left\{\left[\hat{\theta}_N-E\left(\hat{\theta}_N\right)\right]\frac{\partial}{\partial\theta}\left[\ln f(X/\theta)\right]\right\}$$

等式成立的条件：

$$\frac{\partial \ln f(X/\theta)}{\partial\theta}=K(\theta)\left[\hat{\theta}_N-\theta\right]$$

3、有效估计：上式取等号时，估计量具有最小均方误差的估计。

例：（3.2 中例）$x_i=\theta+n_i$， $i=1,2\cdots N$， n服从$N(0,\sigma^2)$

例：

$$\hat{\theta}_{ML}=\frac{1}{N}\sum_{i=1}^{N}x_i$$

似然函数： $f(X/\theta)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left[-\sum_{i=1}^{N}\frac{(x_i-\theta)^2}{2\sigma^2}\right]$

$$\frac{\partial \ln f(X/\theta)}{\partial\theta}=\sum_{i=1}^{N}\frac{(x_i-\theta)^2}{\sigma^2}=\frac{N}{\sigma^2}\left(\hat{\theta}_{ML}-\theta\right)$$

$\hat{\theta}_{ML}$是θ的有效估计。

结论：$\hat{\theta}_{ML}$是θ的无偏，一致的，有效估计。

5.6 最小二乘估计

一、引言

Bayes 估计：需知后验密度函数和损失函数

最大后验估计：需知后验密度函数（等效于的待估计量θ先验密度函数和似然函数）

最大似然估计：需知θ的似然函数

最小二乘估计：1795 年，18 岁的高斯在预测行星估计的研究中提出，省去了全部概率假设，无需知道更多的统计知识。

条件放宽

二、简单情况下的估计

$$x_i=\theta+n_i, \quad i=1,2\cdots N$$

设θ的估计量为$\hat{\theta}$

则第i次测量误差为$\tilde{\theta}_i=x_i-\hat{\theta}$，而非$\tilde{\theta}_i=\theta-\hat{\theta}$

N次测量后，误差平方和为

$$J(\tilde{\theta})=\sum_{i=1}^{N}\tilde{\theta}_i^2=\sum_{i=1}^{N}\left(x_i-\hat{\theta}\right)^2$$

最小二乘估计：使误差平方和最小的估计

$$\frac{\partial J(\tilde{\theta})}{\partial \hat{\theta}} = 0 = -2\sum_{i=1}^{N}(x_i - \hat{\theta})$$

$$\hat{\theta}_{LS} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

三、非直接测量时的估计

有时待估计量θ不能直接测量

需要中间装置如传感器才能测量，如测温（h_i：转换系数）

$$x_i = h_i\theta + n_i, \quad i = 1,2\cdots N$$

设θ 的估计量为$\hat{\theta}$

则第i 次测量误差为：$\tilde{\theta}_i = x_i - h_i\hat{\theta}$

N 次测量后，误差平方和为

$$J(\tilde{\theta}) = \sum_{i=1}^{N}\tilde{\theta}_i^2 = \sum_{i=1}^{N}(x_i - h_i\hat{\theta})^2$$

使误差平方和最小，$\dfrac{\partial J(\tilde{\theta})}{\partial \hat{\theta}} = 0$

且 $\dfrac{\partial^2 J(\tilde{\theta})}{\partial^2 \hat{\theta}} = 2\sum_{i=1}^{N} h_i^2 > 0$，$J(\tilde{\theta})$有最小值。

$$\hat{\theta}_{LS} = \frac{\sum_{i=1}^{N} h_i x_i}{\sum_{i=1}^{N} h_i^2}$$

四、非直接测量时估计的向量形式：

$$x_i = h_i\theta + n_i \qquad i = 1,2\cdots N$$

$$\begin{cases} x_1 = h_1\theta + n_1 \\ x_2 = h_2\theta + n_2 \\ \vdots \\ x_N = h_N\theta + n_N \end{cases}$$

$X = [x_1, x_2, \cdots x_N]^T$ 观察矢量

$H = [h_1, h_2, \cdots h_N]^T$ 转换系数矢量

$\underline{n} = [n_1, n_2, \cdots n_N]^T$ 噪声矢量

$$X = H\theta + \underline{n}$$

设估计量为$\hat{\theta}$，估计误差平方和为

$$J(\tilde{\theta}) = J(\hat{\theta}) = (X - H\hat{\theta})^T(X - H\hat{\theta})$$

使$J(\tilde{\theta})$最小，$\dfrac{\partial J(\tilde{\theta})}{\partial \hat{\theta}} = 0 = -2H^T(X - H\hat{\theta})$

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

例： $x_i = \theta + n_i$，$i = 1,2\cdots N$，θ 为待估计参量，n_i 服从$N(0,\sigma^2)$，求θ 的最小二乘估计。

解： 单样本的似然函数：

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

样本矢量的似然函数：

$$f(X/\theta) = \prod_{i=1}^{N} f(x_i/\theta)$$

两边取对数，

$$l = \ln f(X/\theta) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\theta)^2 + 常量$$

使l 最大的估计：最大似然估计，需使$\sum_{i=1}^{N}(x_i-\theta)^2$最小，类似于最小二乘估计中使$J(\hat{\theta}) = \sum_{i=1}^{N}(x_i-\theta)^2$最小。

结论：高斯噪声下，最大似然估计等价于最小二乘估计，

$$\hat{\theta}_{LS} = \hat{\theta}_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

四、最小二乘的性质

1、无偏性：

$$\hat{\theta}_{LS} = (H^T H)^{-1} H^T X \qquad X = H\theta + \underline{n}$$

$$\begin{aligned}\hat{\theta}_{LS} &= (H^T H)^{-1} H^T (H\theta + \underline{n}) \\ &= (H^T H)^{-1}(H^T H)\theta + (H^T H)^{-1} H^T \underline{n} \\ &= \theta + (H^T H)^{-1} H^T \underline{n}\end{aligned}$$

两边取期望，同时得到：$\hat{\theta}_{LS} - \theta = \tilde{\theta} = (H^T H)^{-1} H^T \underline{n}$

$$E[\hat{\theta}_{LS}] = E[\theta] + (H^T H)^{-1} H^T E[\underline{n}]$$

设$E[\underline{n}] = 0$，有 $E[\hat{\theta}_{LS}] = E[\theta]$

2、一致性

$$\tilde{\theta} = (H^T H)^{-1} H^T \underline{n}$$

$$\begin{aligned}E[\tilde{\theta}^2] &= E\left[(\hat{\theta}_{LS} - \theta)(\hat{\theta}_{LS} - \theta)^T\right] \\ &= (H^T H)^{-1} H^T E(\underline{n}\underline{n}^T) H (H^T H)^{-1}\end{aligned}$$

设噪声$\underline{n}$ 均值为0，方差为$\underline{R} = E[\underline{n}\underline{n}^T]$

$$E[\tilde{\theta}^2] = (H^T H)^{-1} H^T \underline{R} H (H^T H)^{-1}$$

设 $\underline{R}=E[\underline{n}\underline{n}^T]=\sigma^2 I$，$\underline{n}=(n_1,n_2,\cdots n_N)^T$，$n_i(i=1,2,\cdots N)$，
零均值，独立，同方差为 σ^2。

$$E[\tilde{\theta}^2]=\sigma^2(H^TH)^{-1}H^TH(H^TH)^{-1}=\sigma^2(H^TH)^{-1}$$

$$E[\tilde{\theta}^2]=\frac{\sigma^2}{N}\left(\frac{1}{N}H^TH\right)^{-1}$$

$$\lim_{N\to\infty}E[\tilde{\theta}^2]=0$$

3、高斯分布下，最小二乘等价于最大似然估计。

5.7 线性最小均方误差估计（）

一、.引言

最小二乘估计：

未规定 $\hat{\theta}_{LS}$ 与 x（观察）间的关系；

估计误差的平方和最小。

线性最小均方误差估计：

规定 $\hat{\theta}_L$ 与 x 间为线性关系；

估计误差的均方误差最小；

待估计量的一阶/二阶矩已知。

二、讨论单一估计量的估计

$$x_i=\theta+n_i,\quad i=1,2\cdots N$$

$X=(x_1,x_2,\cdots x_N)$，设估计量为 $\hat{\theta}_L$，且为观察数据的线性组合，

$$\hat{\theta}_L=\sum_{i=1}^{N}a_i x_i$$

估计误差：

$$\tilde{\theta}=\theta-\hat{\theta}_L=\theta-\sum_{i=1}^{N}a_i x_i$$

估计误差均方值（平均误差功率）：

$$E[\tilde{\theta}^2]=E\left[\left(\theta-\sum_{i=1}^{N}a_i x_i\right)^2\right]$$

线性最小均方误差估计：使 $E[\tilde{\theta}^2]$ 最小的估计。

$$\frac{\partial E[\tilde{\theta}^2]}{\partial a_i}=0\qquad i=1,2\cdots N$$

期望与微分可互换，

$$\frac{\partial E[\tilde{\theta}^2]}{\partial a_i}=E\left[\frac{\partial\tilde{\theta}^2}{\partial a_i}\right]=2E\left[\frac{\tilde{\theta}\partial\tilde{\theta}}{\partial a_i}\right]=2E\left[\frac{\tilde{\theta}\partial(\theta-\sum_{i=1}^{N}a_i x_i)}{\partial a_i}\right]=0$$

$$E[\tilde{\theta}x_i]=0\qquad i=1,2\cdots N$$

表示：在线性最小均方误差条件下，估计误差与观测值正交。

$$E\left[\left(\theta-\sum_{i=1}^{N}a_i x_i\right)x_j\right]=0\qquad j=1,2\cdots N$$

$$E[\theta x_j]-\sum_{i=1}^{N}a_i E[x_i x_j]=0\qquad j=1,2\cdots N$$

$$R_{ij}=E[x_i x_j],\qquad b_j=E[\theta x_j]$$

$$\sum_{i=1}^{N}a_i R_{ij}=b_j\qquad j=1,2\cdots N$$

$$\begin{cases}R_{11}a_1+R_{21}a_2+\cdots R_{N1}a_N=b_1\\ R_{12}a_1+R_{22}a_2+\cdots+R_{N2}a_N=b_2\\ \vdots\\ R_{1N}a_1+R_{2N}a_2+\cdots+R_{NN}a_N=b_N\end{cases}$$

$\underline{A}=(a_1,a_2,\cdots a_N)^T$ 线性均方误差最小的系数向量

$\underline{B}=(b_1,b_2,\cdots b_N)^T=E[\theta X]$ 互相关系数向量

$$\underline{R}=\begin{bmatrix}R_{11}&R_{21}&\cdots&R_{N1}\\ R_{12}&R_{22}&\cdots&R_{N2}\\ \vdots&\vdots&\ddots&\vdots\\ R_{1N}&R_{2N}&\cdots&R_{NN}\end{bmatrix}=E[XX^T]$$

自相关矩阵

$$\underline{R}\underline{A}=\underline{B}$$

$$\underline{A}=\underline{R}^{-1}\underline{B}$$

$$\hat{\theta}_L=\sum_{i=1}^{N}a_i x_i=\underline{A}^T X$$

（线性均方误差最小）θ 的估计量

例：$x_i=\theta+n_i$，$i=1,2\cdots N$，n_i 均值为0，方差为 σ_n^2，且采样之间噪声不相关，信号与噪声不想关，即有 $E[n_i]=0$，$E[n_i^2]=\sigma_n^2$，$E[n_i n_j]=\sigma_n^2\delta_{ij}$，$E[\theta n_i]=0$，令信号 θ 有 $E[\theta]=0$，$E[\theta^2]=C$，求 $\hat{\theta}_L$。

解：

$$E[\tilde{\theta}x_j]=0\qquad j=1,2\cdots N$$

$$\sum_{i=1}^{N}a_i R_{ij}=b_j\qquad j=1,2\cdots N$$

自相关矩阵

$$R_{ij}=E[x_i x_j]=E[(\theta+n_i)(\theta+n_j)]=C+\delta_n^2\delta_{ij}$$

互相关系数

$$b_j=E[\theta x_j]=E[\theta(\theta+n_j)]=C$$

$$\sum_{i=1}^{N}a_i(C+\sigma_n^2\delta_{ij})=C\qquad j=1,2\cdots N$$

$$C\sum_{i=1}^{N}a_i+\sigma_n^2\delta_{ij}\sum_{i=1}^{N}a_i=C\qquad j=1,2\cdots N$$

$$C\sum_{i=1}^{N}a_i+\sigma_n^2 a_j=C\qquad j=1,2\cdots N$$

将这上面的 $j=1,2\cdots N$ N 个方程相加，

$$NC\sum_{i=1}^{N}a_i+\sigma_n^2\sum_{j=1}^{N}a_j=NC$$

$$\sum_{i=1}^{N}a_i=\frac{NC}{NC+\sigma_n^2}$$

信噪比 $SNR=\dfrac{C}{\sigma_n^2}=\dfrac{1}{b}$，$b=\dfrac{\sigma_n^2}{C}$，

$$a_1=a_2=\cdots=a_N=\frac{C}{NC+\sigma_n^2}=\frac{1}{N+\frac{1}{SNR}}=\frac{1}{N+b}$$

$$\hat{\theta}_L=\sum_{i=1}^{N}a_ix_i=\frac{1}{N+b}\sum_{i=1}^{N}x_i$$

$$\tilde{\theta}=\theta-\hat{\theta}_L=\frac{1}{N+b}\left[b\theta-\sum_{i=1}^{N}n_i\right]$$

$$E\left[\tilde{\theta}^2\right]=\frac{1}{N+b}\sigma_n^2$$

5.8 线性递推估计

一、引言

$$x_i=\theta+n_i,\quad i=1,2\cdots N$$

当$N=K$时，$\hat{\theta}_K=\underline{A}_K^TX_K$　$\underline{A}_K^T=\underline{R}_K^{-1}\underline{B}_K$

$$\underline{R}_K=E[X_K^TX_K]$$

$$\underline{B}_K=E[\theta X_K]$$

（1）求$\underline{R}_K^{-1}$；（2）存储必要值

当$N=K+1$时，$\hat{\theta}_{K+1}=\underline{A}_{K+1}^TX$　$\underline{A}_{K+1}^T=\underline{R}_{K+1}^{-1}\underline{B}_{K+1}$

（1） 再求$\underline{R}_{K+1}^{-1}$；（2）存储

观察数据每增加一维，相应各量重算，费时，费存储空间。能否利用上一次估计量$\hat{\theta}_K$和这次实测值x_{K+1}来估计这次的估计量$\hat{\theta}_{K+1}$？

递推估计：

$$\hat{\theta}_{K+1}=a_{K+1}\hat{\theta}_K+b_{K+1}x_{K+1}$$

上一次估计量$\hat{\theta}_K$与这次实测数据x_{K+1}的线性组合。

二、线性递推估计

例：已知$x_i=\theta+n_i$，$i=1,2\cdots N$ $E[n_i]=0$，$E[n_in_k]=\sigma_n^2\delta_{ik}$，$E[\theta]=0$，$E[\theta n]=0$，$E[\theta^2]=C$，求$\theta$的线性递推估计。

解：

$N=1$　第一次测量值x_1

设估计量：$\hat{\theta}_1=b_1x_1$

误差：$\tilde{\theta}_1=\theta-\hat{\theta}_1$

均方误差：$e_1=E[\tilde{\theta}_1^2]=E\left[(\theta-\hat{\theta}_1)^2\right]$

准则：选择b_1，使$\min\left\{E\left[\tilde{\theta}_1^2\right]\right\}=\min\{e_1\}$

正交原理，估计误差与测量值正交。

$$E[\tilde{\theta}_1x_1]=0$$

$$x_1=\theta+n_1$$

$$\tilde{\theta}_1=\theta-\hat{\theta}_1=\theta-b_1x_1$$

$$E\left[\tilde{\theta}_1x_1\right]=E\left[(\theta-b_1x_1)(\theta+n_1)\right]=0$$

$$\Rightarrow E\left[\theta^2\right]=b_1\left\{E\left[\theta^2\right]+\sigma_n^2\right\}$$

$$C=b_1\left[C+\sigma_n^2\right]$$

$$b_1=\frac{C}{C+\sigma_n^2}=\frac{1}{1+b},$$

$b=\frac{\sigma_n^2}{C}$，信噪比倒数。

此时均方误差，

$$e_1=E\left[\tilde{\theta}_1^2\right]=E\left[(\theta-\hat{\theta}_1)^2\right]=E\left[(\theta-b_1x_1)^2\right]=E\left\{\left[\theta-b_1(\theta+n_1)\right]^2\right\}=\frac{\sigma_n^2}{1+b}$$

归一化误差，$P_1=\frac{e_1}{\sigma_n^2}=\frac{1}{1+b}=b_1$

$N=2$　第二次测量值x_2

设估计量：$\hat{\theta}_2=a_2\hat{\theta}_1+b_2x_2$

误差：$\tilde{\theta}_2=\theta-\hat{\theta}_2$

均方误差：$e_2=E[\tilde{\theta}_2^2]=E\left[(\theta-\hat{\theta}_2)^2\right]$

准则：选择a_2，b_2，使e_2最小。

$$\min\{e_2\}=\min\{E(\tilde{\theta}_2^2)\}$$

根据正交原理：

$$E[\tilde{\theta}_2x_2]=0$$

$$E[\tilde{\theta}_2\hat{\theta}_1]=E[\tilde{\theta}_2b_1x_1]=0\Rightarrow E[\tilde{\theta}_2x_1]=0$$

$E[\tilde{\theta}_2x_1]=0$　　$E[\tilde{\theta}_2x_2]=0$

$$E[(\theta-\hat{\theta}_2)x_1]=0$$

$$\hat{\theta}_2=a_2\hat{\theta}_1+b_2x_2$$

$$\hat{\theta}_1=\theta-\tilde{\theta}_1$$

$$x_1=\theta+n_1$$

$$x_2=\theta+n_2$$

$$E[\tilde{\theta}_1x_1]=0$$

$$a_2=1-b_2$$

$$\tilde{\theta}_2=\theta-\hat{\theta}_2=\theta-a_2\hat{\theta}_1-b_2x_2$$

$$x_2=\theta+n_2$$

$$\hat{\theta}_1=b_1x_1$$

$$x_1=\theta+n_1$$

$$b_2\sigma_n^2=C-b_2C-a_2E\left[\hat{\theta}_1x_2\right]$$

$$x_2=\theta+n_2$$

$$\hat{\theta}_1=b_1x_1=b_1(\theta+n_1)$$

$$b_2\sigma_n^2=C-b_2C-a_2E\left[\hat{\theta}_1\theta\right]$$

$$\because\ e_2=E[\tilde{\theta}_2]=E\left[(\theta-\hat{\theta}_2)^2\right]=C-b_2C-a_2E[\hat{\theta}_1\theta]$$

$$e_2=b_2\sigma_n^2$$

归一化误差量：$p_2=\frac{e_2}{\sigma_n^2}=b_2$

$$\hat{\theta}_1=b_1x_1=b_1(\theta+n_1)$$

$$b_2\sigma_n^2=C-b_2C-a_2b_1C$$

$$b=\frac{\sigma_n^2}{C}$$

$$a_2=1-b_2$$

$$(1-b_1)-b_2(1-b_1)=b_2b$$

$$1-b_1=\frac{b}{1+b}$$

$$\frac{1}{1+b}=p_1$$

$$b_2=p_2=\frac{p_1}{1+p_1}$$

$$a_2=1-b_2$$

$$\hat\theta_2=a_2\hat\theta_1+b_2x_2$$

同理： $N=K+1$

$$\hat\theta_{K+1}=a_{K+1}\hat\theta_K+b_{K+1}x_{K+1}$$

$$a_{K+1}=1-b_{K+1}$$

$$b_{K+1}=p_{K+1}=\frac{p_K}{1+p_K}$$

$$\hat\theta_{K+1}=\hat\theta_K+b_{K+1}\left(x_{K+1}-\hat\theta_K\right)$$

残差

5.9 正交原理

一、引言

上节：$E[\tilde\theta x_i]=0$，观察值与估计误差$\tilde\theta$正交。

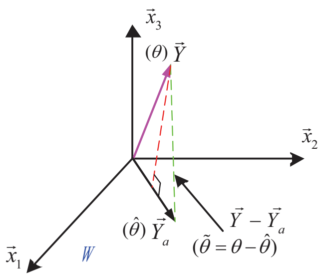
一般性正交原理：估计量$\hat\theta$与估计误差$\tilde\theta$正交，$E[\hat\theta\tilde\theta]=0$

二、几何解释

V：三个相互垂直的向量$\vec x_1$，$\vec x_2$，$\vec x_3$ 张成的三维空间。

W：V的子空间，由向量$\vec x_1$，$\vec x_2$张成的平面。

$\vec Y$：存在于V，但不存在于W。



在W中找一个向量$\vec Y_a$，最近似V中的$\vec Y$，希望$\vec Y$与$\vec Y_a$的向量差$\vec Y-\vec Y_a$的长度最小，即范数$\|\vec Y-\vec Y_a\|$取极小值。由几何学知识可知，投影原理（正交原理）$(\vec Y-\vec Y_a)\perp\vec Y_a$，即向量差$\perp W$平面上的投影。该结论可推广到高维空间。

差向量$\perp W$平面$\to$差向量$\perp W$平面上每个基向量（类似观察数据）。

用内积表示：

$$\left(\left(\vec Y-\vec Y_a\right),\vec x_j\right)=0$$

$$\vec Y_a=\sum_{i=1}^{N}a_i\vec x_i$$ （N维空间，a_i为标量）

N维空间 $$\left(\left(\vec Y-\sum_{i=1}^{N}a_i\vec x_i\right),\vec x_j\right)=0$$

内积性质：$(\alpha\vec X+\beta\vec Y,\vec Z)=\alpha(\vec X,\vec Z)+\beta(\vec Y,\vec Z)$

$$\left(\vec Y,\vec x_j\right)=\sum_{i=1}^{N}a_i\left(\vec x_i,\vec x_j\right)$$

若$\vec x$，$\vec Y$为随机量，则上面内积符号前加上均值算子。

定义1：N维向量空间， $X=(u_1,u_2,\cdots u_N)^T$

$Y=(v_1,v_2,\cdots v_N)^T$

则$(X,Y)=\sum_{i=1}^{N}u_iv_i^*=(Y^*)^TX$

定义 2：随机变量空间 $X=(u_1,u_2,\cdots u_N)^T$，0 均值，$Y=(v_1,v_2,\cdots v_N)^T$，0 均值

则$E(X,Y)=E\left(\sum_{i=1}^{N}u_iv_i^*\right)$

5.10 波形的线性最小均方误差估计

一、引言

以前讲参量估计（一个确定量，随机量，波形中的一个未知量）

本节讲：一个时间波形的估计，平稳随机过程。

本节讲：简单的波形估计，后面讲 Winer 和 Kalman。

二、问题

$x(t)=s(t)+n(t)$，$s(t),n(t)$是 0 均值的平稳随机过程。

由$x(t)$中特定数据$x[\xi]$（$\xi\in I$）来估计

$s(t)$ 滤波

$s(t+\tau)$ $\tau>0$ 预测

$s(t-\tau)$ $\tau>0$ 平滑（内插）

用$y(t)$统一表示。

Note： I为时间轴上某些离散点集合或一段时间区间。

三、线性最小均方误差准则

估计量：$\hat y(t)=L\{x(\xi):\xi\in I\}$，$x(\xi)$的线性组合。

估计误差：$\tilde y(t)=y(t)-\hat y(t)$

误差均方值：$E\left[\tilde y(t)^2\right]=E\left[(y(t)-\hat y(t))^2\right]$

准则： 选择一个合适的线性算子 L，$\min\left\{E\left[\tilde y(t)^2\right]\right\}$，

由正交原理：

$$E[\tilde y(t)x(\xi_i)]=0\qquad(\xi_i\in I)$$

$$E[(y(t)-\hat y(t))x(\xi_i)]=0\qquad(\xi_i\in I)$$

$$E[(y(t)-L[x(\xi),\xi\in I])x(\xi_i)]=0\qquad(\xi_i\in I)$$

此时，最小均方误差：

$$E[\tilde y^2(t)]=R_{LMS}=E[\tilde y(t)y(t)]$$
$$=E[(y(t)-L[x(\xi),\xi\in I])y(t)]$$

$$E[\tilde y^2(t)]=E\{\tilde y(t)[y(t)-\hat y(t)]\}$$

证明： $=E[\tilde y(t)y(t)]-E[\tilde y(t)\hat y(t)]$ 后面一项由

正交原理可知为零。

例1：已知随机过程$s(t)$，确定$s(t)$的预测值$s(t+\tau)$，$\tau>0$。

解：

由$s(t)$估计：$y(t)=s(t+\tau)$

设估计量：$\hat{y}(t)=as(t)=\hat{s}(t+\tau)$

估计误差：$\tilde{y}(t)=s(t+\tau)-as(t)$

根据正交原理：

$$E[\tilde{y}(t)s(t)]=0$$

$$E[(s(t+\tau)-as(t))s(t)]=0$$

$$E[s(t+\tau)s(t)]=aE[s(t)s(t)]$$

$$R(\tau)=aR(0)$$

$$\Rightarrow a=\frac{R(\tau)}{R(0)}<1$$

$$\therefore \hat{s}(t+\tau)=\frac{R(\tau)}{R(0)}s(t)$$

最小均方误差：

$$E\left[\tilde{y}^2(t)\right]=E\left[\tilde{y}(t)y(t)\right]$$
$$=E\left[\left(s(t+\tau)-as(t)\right)s(t+\tau)\right]$$
$$=R(0)-aR(\tau)=R(0)-\frac{R^2(\tau)}{R(0)}$$
$$\overset{令R(0)=1}{=} 1-R^2(\tau)$$

结论：

a．由相关函数特性，τ 增大，$R(\tau)$降低，$E[\tilde{y}^2(t)]$增大。

b．对于已知随机过程$s(t)$，估计性能取决于相关函数$R(\tau)$。

c．对于已知随机过程$s(t)$，如果$R(\tau)=\delta(t)$，前后数据无相关性，则无法作这样的预测，或者说就是利用随机信号的前后相关性来进行预测。

例2．$x(t)=s(t)+n(t)$，$s(t),n(t)$是互相正交的随机过程。由$x(t)$估计$s(t)$，即滤波。

解：$y(t)=s(t)$

设估计量：$\hat{y}(t)=\hat{s}(t)=ax(t)$

估计误差：$\tilde{y}(t)=y(t)-\hat{y}(t)=s(t)-ax(t)$

均方误差：$E[\tilde{y}^2(t)]=E[(s(t)-ax(t))^2]$

根据正交原理：$E[\tilde{y}(t)x(t)]=0$

$$E[(s(t)-ax(t))x(t)]=0$$

$$E[s(t)x(t)]=aE[x(t)x(t)]$$

$$R_{sx}(0)=aR_{xx}(0)$$

$$\Rightarrow a=\frac{R_{sx}(0)}{R_{xx}(0)}$$

$$\because R_{sx}(\tau)=R_{ss}(\tau)$$

$$R_{xx}(\tau)=R_{ss}(\tau)+R_{nn}(\tau)$$

$$a=\frac{R_{ss}(0)}{R_{ss}(0)+R_{nn}(0)}$$

$$\therefore \hat{s}(t)=\frac{R_{ss}(0)}{R_{ss}(0)+R_{nn}(0)}x(t)$$

最小均方误差：

$$E[\tilde{y}^2(t)]=E[\tilde{y}(t)y(t)]$$
$$=E[(s(t)-ax(t))s(t)]$$
$$=R_{ss}(0)-aR_{xs}(0)=R_{ss}(0)-aR_{ss}(0)$$
$$=\frac{R_{ss}(0)R_{nn}(0)}{R_{ss}(0)+R_{nn}(0)}$$

5.11 维纳滤波器

一、引言

前面讲简单波形估计时，$x(t)=s(t)+n(t)$，估计

$s(t)$ 滤波

$s(t+\tau)$ $\tau>0$ 预测

$s(t-\tau)$ $\tau>0$ 平滑

估计量：$\hat{y}(t)=\sum_{i=1}^{N}a_ix_i$

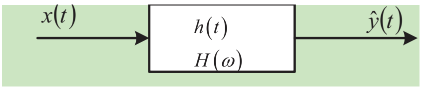
估计误差：$\tilde{y}(t)=y(t)-\hat{y}(t)$

$\min\left\{E\left[\tilde{y}(t)^2\right]\right\}$，求出$a_i$。

维纳滤波：

$$x(t)=s(t)+n(t)，估计\begin{cases}s(t)\\ s(t+\tau),\tau>0\\ s(t-\tau),\tau>0\end{cases}$$

a.基本思想



估计量： $\hat{y}(t)=x(t)\otimes h(t)$

估计误差: $\tilde{y}(t)=y(t)-\hat{y}(t)$

$$\min\left\{E\left[\tilde{y}(t)^2\right]\right\}=\min\left\{E\left[y(t)-x(t)\otimes h(t)\right]^2\right\}$$

求出 $h(t)$，此时为最佳滤波器。

Winner filter 的基本思想：寻求线性滤波器的最佳 $h(t)$，使其输出波形作为输入波形估计时均方误差达到最小。

b. Winner filter 是 Winner 于 40 年代为解决火炮射点预测问题而提出的，49 年美国解密，同时期，苏联学者哥尔莫哥霍夫也解决了该问题，所以该滤波器也叫维纳—霍夫滤波器。

c. Winner filter 特点：

必须用到无限过去的数据，不适合实时处理，费时，费存储空间。

不再涉及所有的概率假设，但要求一阶/二阶距。

二、维纳滤波器

$x(t)=s(t)+n(t)$，估计

$y(t)=s(t)$

$s(t+\alpha)$

$s(t-\alpha)$ （$\alpha>0$）

$x(t)$，$s(t)$互相统计相关且均值为 0，$x(t)$在时间间隔$[a,b]$内值由$x(\xi)$表示，设所求最佳 filter 为时不变系统。

估计量：

$$\hat{y}(t)=\int_a^b h(t-\xi)x(\xi)d\xi \overset{t-\xi=\lambda}{=} \int_{t-b}^{t-a} h(\lambda)x(t-\lambda)d\lambda$$

估计误差： $\tilde{y}(t)=y(t)-\hat{y}(t)$

均方误差： $E[\tilde{y}(t)^2]$

准则： $\min\left\{E\left[\tilde{y}(t)^2\right]\right\}$，利用正交原理

$$E[\tilde{y}(t)x(\xi)]=0$$

$$E\left\{\left[y(t)-\int_{t-b}^{t-a}h(\lambda)x(t-\lambda)d\lambda\right]x(\xi)\right\}=0$$

$E[y(t)x(\xi)]=R_{yx}(t-\xi)$ 互相关函数

$E[x(t-\lambda)x(\xi)]=R_{xx}(t-\lambda-\xi)$ 自相关函数

$$R_{yx}(t-\xi)=\int_{t-b}^{t-a}R_{xx}(t-\xi-\lambda)h(\lambda)d\lambda$$

$\lambda\in[t-b,t-a]$ $\xi\in[a,b]$

令 $\tau=t-\xi$，$\tau\in[t-b,t-a]$

$$R_{yx}(\tau)=\int_{t-b}^{t-a}R_{xx}(\tau-\lambda)h(\lambda)d\lambda,$$

此时有最小均方误差：

$$E\left[\tilde{y}^2(t)\right]=E\left[\tilde{y}(t)y(t)\right]$$

$$=E\left\{\left[y(t)-\int_{t-b}^{t-a}h(\lambda)x(t-\lambda)d\lambda\right]y(t)\right\}$$

$$=R_{yy}(0)-\int_{t-b}^{t-a}R_{yx}(-\lambda)h(\lambda)d\lambda$$

分两种情况求 $h(\lambda)$。

I 型非因果系统

$a=-\infty$，$b=\infty$

$\lambda\in[t-b,t-a]\Rightarrow\lambda\in(-\infty,\infty)$

$\tau\in[t-b,t-a]\Rightarrow\tau\in(-\infty,\infty)$

$$R_{yx}(\tau)=\int_{-\infty}^{\infty}R_{xx}(\tau-\lambda)h(\lambda)d\lambda$$

此时，$\lambda<0$时，$h(\lambda)\neq 0$，非因果系统。不仅用到过去数据，而且要用未来数据，不适合实时处理场合。

II 型因果系统

$a=-\infty$，$b=t$

$\lambda\in[t-b,t-a]\Rightarrow\lambda\in[0,\infty)$

$\tau\in[t-b,t-a]\Rightarrow\tau\in[0,\infty)$

$$R_{yx}(\tau)=\int_{0}^{\infty}R_{xx}(\tau-\lambda)h(\lambda)d\lambda$$ 维纳—霍夫方程

此时，$\lambda<0$时，$h(\lambda)=0$，因果系统。

分析非因果系统意义：

- 对记录数据的事后分析可用。
- 物理不可实现的 filter 的均方误差是物理可实现 filter 均方误差的下限，而且物理不可实现 filter 最小均方误差容易计算，这就值得研究。

三、非因果系统（I 型）

$$R_{yx}(\tau)=\int_{t-b}^{t-a}R_{xx}(\tau-\lambda)h(\lambda)d\lambda$$

$a=-\infty$，$b=\infty$

$\lambda\in[t-b,t-a]\Rightarrow\lambda\in(-\infty,\infty)$

$\tau\in[t-b,t-a]\Rightarrow\tau\in(-\infty,\infty)$

$$R_{yx}(\tau)=\int_{-\infty}^{\infty}R_{xx}(\tau-\lambda)h(\lambda)d\lambda \qquad \tau\in(-\infty,\infty)$$

两边取拉氏变换

$S_{yx}(s)=S_{xx}(s)H(s)$

$s=j\omega$，代入得功率谱密度

$S_{yx}(\omega)=S_{xx}(\omega)H(\omega)$

$S_{yx}(\omega)$为输入$x(t)$与待估计量$y(t)$的互谱密度；

$S_{xx}(\omega)$为输入$x(t)$的谱密度。

I 型最佳滤波器的传递函数

$$H(\omega)=\frac{S_{yx}(\omega)}{S_{xx}(\omega)}$$

当$x(t)=s(t)+n(t)$，$y(t)=s(t)$，

$s(t)$与$n(t)$相互独立时，

$R_{xx}(\tau)=R_{ss}(\tau)+R_{nn}(\tau)$

$S_{xx}(\omega)=S_{ss}(\omega)+S_{nn}(\omega)$

$R_{yx}(\tau)=R_{sx}(\tau)=R_{ss}(\tau)$

$S_{yx}(w)=S_{ss}(w)$

$$H(\omega)=\frac{S_{yx}(\omega)}{S_{xx}(\omega)}=\frac{S_{ss}(\omega)}{S_{ss}(\omega)+S_{nn}(\omega)}$$

注意：当$S_{nn}(\omega)$较小频段上，$H(\omega)$大，

当$S_{nn}(\omega)$较大频段上，$H(\omega)$小，**Winner filter** 就是这样抑制噪声复现信号的。

I 型最佳滤波器的最小均方误差为：

$$E\left[\tilde{y}^2(t)\right]=E\left[\tilde{y}(t)y(t)\right]$$

$$=E\left\{\left[y(t)-\int_{-\infty}^{\infty}h(\lambda)x(t-\lambda)d\lambda\right]y(t)\right\}$$

$$=R_{yy}(0)-\int_{-\infty}^{\infty}R_{yx}(-\lambda)h(\lambda)d\lambda$$

令$w(\tau)=R_{yy}(0)-\int_{-\infty}^{\infty}R_{yx}(\tau-\lambda)h(\lambda)d\lambda$

$E\left[\tilde{y}^2(t)\right]_{\min}=w(0)$　　两边取变换，且$s=j\omega$

$$W(\omega)=S_{yy}(\omega)-S_{yx}(\omega)H(\omega)$$

$$w(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}W(\omega)e^{j\omega\tau}d\omega$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[S_{yy}(\omega)-S_{yx}(\omega)H(\omega)\right]e^{j\omega\tau}d\omega$$

$$E\left[\tilde{y}^2(t)\right]_{\min}=w(0)$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[S_{yy}(\omega)-S_{yx}(\omega)H(\omega)\right]d\omega$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\frac{S_{yy}(\omega)S_{xx}(\omega)-S_{yx}^2(\omega)}{S_{xx}(\omega)}\right]d\omega$$

$$H(\omega)=\frac{S_{yx}(\omega)}{S_{xx}(\omega)}$$

当$x(t)=s(t)+n(t)$，$y(t)=s(t)$，

$s(t)$与$n(t)$相互独立时，

$$H(\omega)=\frac{S_{ss}(\omega)}{S_{ss}(\omega)+S_{nn}(\omega)}$$

$$E\left[\tilde{y}^2(t)\right]_{\min}=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\frac{S_{ss}(\omega)S_{nn}(\omega)}{S_{ss}(\omega)+S_{nn}(\omega)}\right]d\omega$$

例：设信号与噪声为零均值随机过程，且信号与噪声相互独立，它们的谱密度$S_{ss}(w)=\frac{1}{1+w^2}$，$S_{nn}(w)=1$，求$s(t)$最佳估计时滤波器。

解：当$\tau=0$时，

$$H(w)=\frac{S_{yx}(w)}{S_{xx}(w)}$$

$$=\frac{S_{ss}(w)}{S_{ss}(w)+S_{nn}(w)}$$

$$=\frac{\frac{1}{1+w^2}}{1+\frac{1}{1+w^2}}=\frac{1}{2+w^2}$$

$$h(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}H(w)e^{jwt}dw$$

$$=\frac{1}{2\sqrt{2}}\exp\left(\sqrt{2}|t|\right)$$

当$\tau\neq0$时，

$$H(w)=\frac{1}{2+w^2}e^{jw\tau}$$

$$h(t)=\frac{1}{2\sqrt{2}}\exp\left(\sqrt{2}\tau+|t|\right)$$

四．因果关系（II 型）

$$R_{yx}(\tau)=\int_{t-b}^{t-a}R_{xx}(\tau-\lambda)h(\lambda)d\lambda$$

$a=-\infty$，$b=t$

$$R_{yx}(\tau)=\int_{0}^{\infty}R_{xx}(\tau-\lambda)h(\lambda)d\lambda\quad \tau\in[0,\infty)$$

维纳－霍夫方程

（**Winer-Hoff** 方程）

求解上述方程有两种方法：

- 维纳－霍夫提出的频谱因式分解
- 伯德（**Bodo**）和仙农（**Shannon**）提出的预白化方法

$R_{yx}(\tau)=\int_0^\infty R_{xx}(\tau-\lambda)h(\lambda)d\lambda$　$\tau\in[0,\infty)$　维纳—霍夫方程

（注意：只对正时成立，这就导致求解的复杂性）

频谱因式分解：构造一个函数$\varphi(\tau)$，使$\varphi(\tau)$满足下列方程

$$\varphi(\tau)=R_{yx}(\tau)-\int_0^\infty R_{xx}(\tau-\lambda)h(\lambda)d\lambda$$

$\varphi(\tau)=0$，$\tau\ge 0$，相关的零极点全在右半平面。

负时间函数，又是非因果函数

$\varphi(\tau)$未知，$\tau<0$，

$h(\tau)=0$，$\tau<0$相关的零极点全在左半平面。

正时间因果函数

两边取拉氏变换

$$\varphi(p)=S_{yx}(p)-S_{xx}(p)H(p)$$

对$S_{xx}(p)$频谱因式分解：$S_{xx}(p)=S_{xx}^{+}(p)S_{xx}^{-}(p)$

$S_{xx}^{+}(p)$所有零极点在$p=\sigma+jw$的左半平面　镜像

$S_{xx}^{-}(p)$所有零极点在$p=\sigma+jw$的右半平面，　对称。

$$\varphi(p)=S_{yx}(p)-S_{xx}^{+}(p)S_{xx}^{-}(p)H(p)$$

$$H(p)S_{xx}^{+}(p)=\frac{-\varphi(p)}{S_{xx}^{-}(p)}+\frac{S_{yx}(p)}{S_{xx}^{-}(p)}$$

左　左　右/右　分解

左　右　左+右

（正时间函数）（负时间函数）正+负时间函数

$$\frac{S_{yx}(p)}{S_{xx}^{-}(p)}=\left[\frac{S_{yx}(p)}{S_{xx}^{-}(p)}\right]^{+}+\left[\frac{S_{yx}(p)}{S_{xx}^{-}(p)}\right]^{-}$$

左　右

只能相加，而非相乘，因为相乘不能分成两项。

$$H(p)S_{xx}^{+}(p)=\left[\frac{S_{yx}(p)}{S_{xx}^{-}(p)}\right]^{+}$$

左　左　左

II 型维纳滤波传递函数：

$$H(p)=\frac{1}{S_{xx}^{+}(p)}\left[\frac{S_{yx}(p)}{S_{xx}^{-}(p)}\right]^{+}$$

II 型维纳滤波的最小均方误差：

$$\begin{aligned}E[\tilde{y}^2(t)]&=E[\tilde{y}(t)y(t)]\\&=E\left\{\left[y(t)-\int_0^\infty h(\lambda)x(t-\lambda)d\lambda\right]y(t)\right\}\\&=R_{yy}(0)-\int_0^\infty R_{yx}(-\lambda)h(\lambda)d\lambda\end{aligned}$$

例：设$x(t)=s(t)+n(t)$，$R_{ss}(\tau)=Ae^{-a|\tau|}$，$R_{nn}(\tau)=\eta\delta(\tau)$，对信号$s(t)$进行估计的维纳滤波器。

解：

$$R_{ss}(\tau)\to S_{ss}(w)=\frac{2aA}{a^2+w^2}$$

简单考虑a=1，$A=1/2$，$p=jw$

$$=\frac{-1}{-1+p^2}$$

$$R_{nn}(\tau)\to S_{nn}(w)=\eta\overset{\eta=1}{=}1$$

I 型滤波器：$H(w)=\dfrac{S_{yx}(w)}{S_{xx}(w)}$

$$H_1(p)=\frac{S_{yx}(p)}{S_{xx}(p)}=\frac{S_{sx}(p)}{S_{xx}(p)}$$

$$\overset{s(t),n(t)\text{相对独立}}{=}\frac{S_{ss}(p)}{S_{xx}(p)}$$

$$=\frac{S_{ss}(p)}{S_{ss}(p)+S_{nn}(p)}=\frac{1}{p^2-2}$$（有一个极点$\sqrt{2}$在右半平面）

$p=jw$　为不可实现的 filter。

$$H_1(w)=\frac{1}{2+w^2}$$

$$h_1(t)=\frac{1}{2\sqrt{2}}e^{\sqrt{2}|t|}$$

I 型滤波器最小均方误差：

$$\begin{aligned}E[\tilde{y}^2(t)]_{\min}&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\frac{S_{ss}(w)S_{nn}(w)}{S_{ss}(w)+S_{nn}(w)}\right]dw\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{2+w^2}dw\approx 0.354\end{aligned}$$

II 型滤波器：$H(p)=\dfrac{1}{S_{xx}^{+}(p)}\left[\dfrac{S_{yx}(p)}{S_{xx}^{-1}(p)}\right]^{+}$

$$\begin{aligned}S_{xx}(p)&=S_{ss}(p)+S_{nn}(p)\\&=\frac{-1}{-1+p^2}+1=\frac{p^2-2}{p^2-1}=\frac{(p+\sqrt{2})(p-\sqrt{2})}{(p+1)(p-1)}\\&=S_{xx}^{+}(p)S_{xx}^{-}(p)\end{aligned}$$

$S_{xx}^{+}(p)=\dfrac{p+\sqrt{2}}{p+1}$　　$S_{xx}^{-}(p)=\dfrac{p-\sqrt{2}}{p-1}$

$$S_{yx}(p)=S_{sx}(p)=S_{ss}(p)=\frac{-1}{p^2-1}=\frac{-1}{(p+1)(p-1)}$$

$$\frac{S_{yx}(p)}{S_{xx}^{-}(p)}=\frac{-1}{(p+1)(p-\sqrt{2})}=\frac{\frac{1}{1+\sqrt{2}}}{p+1}+\frac{\frac{-1}{1+\sqrt{2}}}{p-\sqrt{2}}$$

左　　右

$$\left[\frac{S_{yx}(p)}{S_{xx}^{-}(p)}\right]^{+}=\frac{\frac{1}{1+\sqrt{2}}}{p+1}$$

$$H_2(p)=\frac{1}{S_{xx}^{+}(p)}\left[\frac{S_{yx}(p)}{S_{xx}^{-1}(p)}\right]^{+}=\frac{p+1}{p+\sqrt{2}}\times\frac{\frac{1}{1+\sqrt{2}}}{p+1}=\frac{\frac{1}{1+\sqrt{2}}}{p+\sqrt{2}}$$

$$h_2(t)=\frac{1}{1+\sqrt{2}}e^{-\sqrt{2}t}\qquad (t>0)$$

II 型滤波器最小均方误差：

$$E\left[\tilde{y}^2(t)\right]=R_{yy}(0)-\int_0^{\infty}R_{yx}(-\lambda)h(\lambda)d\lambda$$

$$R_{yy}(0)=R_{ss}(0)=A=\frac{1}{2}$$

$$R_{yx}(-\lambda)=R_{ss}(-\lambda)=\frac{1}{2}e^{-\lambda}$$

$$E\left[\tilde{y}^2(t)\right]=\frac{1}{2}-\int_0^{\infty}\frac{1}{2(1+\sqrt{2})}e^{-\lambda(1+\sqrt{2})}d\lambda\approx 0.415>0.354$$

进一步说明：物理不可实现 **filter** 的均方误差是物理可实现 **filter** 均方误差的下限。

第五章　信号估计

内容提要

- ✓ 5.1 概述;
- ✓ 5.2 参数估计—最大似然估计（**ML**）;
- ✓ 5.3 参数估计—最大后验概率估计（**MAP**）;
- ✓ 5.4 贝叶斯估计
- ✓ 5.5 估计量的性质
- ✓ 5.6 最小二乘估计
- ✓ 5.7 线性最小均方误差估计
- ✓ 5.8 线性递推估计
- ✓ 5.9 正交原理
- ✓ 5.10 波形的线性最小均方误差估计
- ✓ 5.11 维纳滤波器
- ✓ 5.12 卡尔曼滤波

5.12 卡尔曼滤波器

一、引言

40 年代

↓

维纳滤波问题：求滤波器的冲击响应，使估计的均方误差最小

↓

事实上，如何最好地加权过去的输入数据以决定现在的输出，权重就是冲击响应

↓

实时性差，不能满足对人造卫星、导弹等飞行体进行实时精密的跟踪、观测和控制

kalman 滤波：**60** 年代，**kalman** 最初提出是离散的线性滤波问题

↓

稍后，**kalman** 与数学家布西合伙解决了连续的线性滤波问题

↓

是建立在$\begin{cases}\text{信号模型}\\\text{观察模型}\end{cases}$上，线性无偏最小均方误差的递推估计

↓

突破了 **winer** 滤波的平稳过程的限制，没有 **winer** 滤波中无限数据的要求，

winer 单输入或多输入　　　→估计　一个量

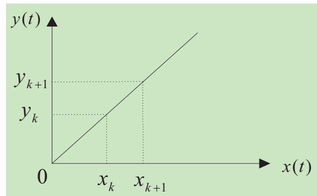
kalman多输入　　**filter**　　→可同时估计　多个量

二、信号方程与测量方程

① 设目标 **F** 在 **XOY** 平面上匀速运动

$t_k=kT$时，目标的位置为（x_k,y_k），速度为（$\dot{x}_k,\dot{y}_k$），

$t_{k+1}=(k+1)T$时，目标的位置为（x_{k+1},y_{k+1}），速度为（$\dot{x}_{k+1},\dot{y}_{k+1}$）。



$$
\begin{cases}
x_{k+1} = x_k + \dot{x}_k T \\
\dot{x}_{k+1} = \dot{x}_k \\
y_{k+1} = y_k + \dot{y}_k T \\
\dot{y}_{k+1} = \dot{y}_k
\end{cases}
$$

信号方程：
$$
\begin{pmatrix} x_{k+1} \\ \dot{x}_{k+1} \\ y_{k+1} \\ \dot{y}_{k+1} \end{pmatrix} = \begin{pmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_k \\ \dot{x}_k \\ y_k \\ \dot{y}_k \end{pmatrix}
$$

② 若目标非等速运动

$t_k = kT$ 时，目标的位置为（x_k, y_k），速度为（$\dot{x}_k, \dot{y}_k$），

$t_{k+1} = (k+1)T$ 时，目标的位置为（x_{k+1}, y_{k+1}），速度为（$\dot{x}_{k+1}, \dot{y}_{k+1}$）。

设从 t_k 到 t_{k+1} 内，速度有随机变化，即加速度分别随机分量 η_{xk}，η_{yk}。

$$
\begin{cases}
x_{k+1} = x_k + \dot{x}_k T + \eta_{xk} \dfrac{T^2}{2} \\
\dot{x}_{k+1} = \dot{x}_k + \eta_{xk} T \\
y_{k+1} = y_k + \dot{y}_k T + \eta_{yk} \dfrac{T^2}{2} \\
\dot{y}_{k+1} = \dot{y}_k + \eta_{yk} T
\end{cases}
$$

信号方程：
$$
\begin{pmatrix} x_{k+1} \\ \dot{x}_{k+1} \\ y_{k+1} \\ \dot{y}_{k+1} \end{pmatrix} = \begin{pmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_k \\ \dot{x}_k \\ y_k \\ \dot{y}_k \end{pmatrix} + \begin{pmatrix} \frac{T^2}{2} & 0 \\ T & 0 \\ 0 & \frac{T^2}{2} \\ 0 & T \end{pmatrix} \begin{pmatrix} \eta_{xk} \\ \eta_{yk} \end{pmatrix}
$$

信号方程反映该信号产生时的固有规律。

③ 今在原点处放一测量雷达，

$t_k = kT$ 时，测出的目标位置为（X_k, Y_k），目标实际位置为（x_k, y_k），测量中分别引入测量噪声 V_{xk}，V_{yk}，

测量方程：
$$
\begin{cases}
X_k = x_k + V_{xk} \\
Y_k = y_k + V_{yk}
\end{cases}
$$

$$
\begin{pmatrix} X_k \\ Y_k \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_k \\ \dot{x}_k \\ y_k \\ \dot{y}_k \end{pmatrix} + \begin{pmatrix} V_{xk} \\ V_{yk} \end{pmatrix}
$$

结论：**kalman filter** 就是基于信号方程和测量方程，利用实测数据 $\begin{pmatrix} X_k \\ Y_k \end{pmatrix}$ 以递推方式来估计信号 $\begin{pmatrix} x_k \\ \dot{x}_k \\ y_k \\ \dot{y}_k \end{pmatrix}$。

三、一维 kalman 滤波

① 信号方程（状态方程）：

$$S(k+1) = aS(k) + W(k)$$

（一个离散随机过程可由一个白噪声通过一阶回归系统产生,反映固有规律）

$S(k+1)$：$t = (k+1)T$ 时的采样信号

$S(k)$：$t = kT$ 时的采样信号

$W(k)$：$t = kT$ 时的白噪声

a：表示 $S(k+1)$ 与 $S(k)$ 间的相互系数，**a** 越大，相关性大，系统惯性大，信号变化慢

$$E[W(k)] = 0,\quad E[W(k)W(j)] = \sigma_w^2 \sigma_{kj}$$

$$E[S(k)] = 0$$

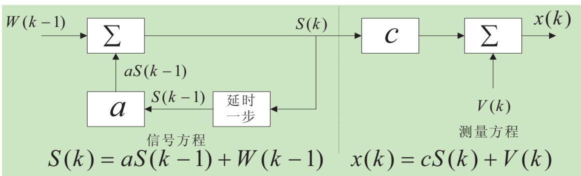
$$E[S^2(k)] = R_s(0) = \sigma_s^2 = \frac{\sigma_w^2}{1-a^2} = P_s(0)$$

$$E[S(k)S(k+j)] = R_s(j) = a^{|j|} R_s(0) = a^{|j|} P_s(0)$$

② 测量方程：

$$x(k) = cS(k) + V(k)$$

c：测量值与信号之间的转换系数



③ 滤波方程：

k 步滤波估计：$\hat{S}(k) = a(k)\hat{S}(k-1) + b(k)x(k)$

（上一步估计值和这次实测值的线性组合）

k 步滤波误差：$e(k) = S(k) - \hat{S}(k)$

k 步滤波估计验后均方误差：$P(k) = E\{e^2(k)\}$

$\min[P(k)]$ 正交原理

$$E[e(k)\hat{S}(k-1)] = 0 \qquad\qquad E[e(k)x(k)] = 0$$

$$E\{[S(k) - \hat{S}(k)]\hat{S}(k-1)\} = 0$$

$$E\{[S(k) - a(k)\hat{S}(k-1) - b(k)x(k)]\hat{S}(k-1)\} = 0$$

$$E\{[a(k)S(k-1) - a(k)\hat{S}(k-1) - a(k)S(k-1)]\hat{S}(k-1)\} = E\{[b(k)x(k) - S(k)]\hat{S}(k-1)\}$$

$\downarrow \quad x(k) = cS(k) + V(k)$

$$a(k)E\{[S(k-1) - \hat{S}(k-1) - S(k-1)]\hat{S}(k-1)\} = E\{[S(k)(cb(k)-1) + b(k)V(k)]\hat{S}(k-1)\}$$

$e(k-1) = S(k-1) - \hat{S}(k-1)$ $\qquad$ $E[v(k)\hat{S}(k-1)] = 0$

$E[e(k-1)\hat{S}(k-1)] = 0$，上一步正交原理

$$a(k)E[S(k-1)\hat{S}(k-1)] = [1 - cb(k)]E[S(k)\hat{S}(k-1)]$$

$S(k) = aS(k-1) + W(k-1)$

$E[W(k-1)\hat{S}(k-1)] = 0$

$$a(k)E[S(k-1)\hat{S}(k-1)] = a[1 - cb(k)]E[S(k-1)\hat{S}(k-1)]$$

$\downarrow$

两参数关系：$a(k) = a[1 - cb(k)]$

验后滤波估计均方误差：

$$P(k) = E\{e^2(k)\} = E\{e(k)S(k)\}$$

$x(k) = cS(k) + V(k) \longrightarrow S(k) = \frac{1}{c}[x(k) - V(k)]$

$E[e(k)x(k)] = 0$

$$P(k)=E\{e(k)\frac{1}{c}[x(k)-V(k)]\}=-\frac{1}{c}E[e(k)V(k)]$$

$$e(k)=S(k)-\hat{S}(k)=S(k)-a(k)\hat{S}(k-1)-b(k)x(k)$$

$$P(k)=-\frac{1}{c}E\{[S(k)-a(k)\hat{S}(k-1)-b(k)x(k)]V(k)\}$$

$E[S(k)V(k)]=0$， $E[\hat{S}(k-1)V(k)]=0$

$x(k)=cS(k)+V(k)$

$$P(k)=\frac{1}{c}b(k)E[V(k)V(k)]=\frac{1}{c}b(k)\sigma_v^2$$

滤波估计增益：$b(k)=\dfrac{cP(k)}{\sigma_v^2}$

下面求验后滤波估计均方误差 $P(k)$：

$$P(k)=E\{[S(k)-\hat{S}(k)]^2\}$$

$$=E\{[S(k)-a(k)\hat{S}(k-1)-b(k)x(k)]^2\}$$

$a(k)=a[1-cb(k)]$

$$=E\{[S(k)-a(1-cb(k))\hat{S}(k-1)-b(k)x(k)]^2\}$$

$x(k)=cS(k)+V(k)$

$$=E(\{[1-cb(k)]S(k)-a[1-cb(k)]\hat{S}(k-1)-b(k)V(k)\}^2)$$

$S(k)=aS(k-1)+W(k-1)$

$$=E\{(a[1-cb(k)][S(k)-\hat{S}(k-1)]+[1-cb(k)]W(k-1)-b(k)V(k))^2\}$$

$$e(k-1)=S(k-1)-\hat{S}(k-1)$$

$$=E\{(a[1-cb(k)]e(k-1)+[1-cb(k)]W(k-1)-b(k)V(k))^2\}$$

$P(k-1)=E[e^2(k-1)]$ 第 k-1 步验后滤波估计均方误差

$\sigma_w^2=E[W^2(k-1)]$，

$\sigma_v^2=E[V^2(k)]$， $E[e(k-1)W(k-1)]=0$

$E[e(k-1)V(k)]=0$ 但 $E[e(k)V(k)]\neq 0$

$$P(k)=a^2[1-cb(k)]^2P(k-1)+[1-cb(k)]^2\sigma_w^2+b^2(k)\sigma_v^2$$

$$=[1-cb(k)]^2[a^2P(k-1)+\sigma_w^2]+b^2(k)\sigma_v^2$$

令验前滤波估计均方误差 $P_1(k)=a^2P(k-1)+\sigma_w^2$

利用 $P(k)=\dfrac{b(k)}{c}\sigma_v^2 \longrightarrow cb(k)P(k)=b^2(k)\sigma_v^2$

$$P(k)=[1-cb(k)]^2P_1(k)+cb(k)P(k)$$

$$[1-cb(k)]P(k)=[1-cb(k)]^2P_1(k)$$

验后滤波估计均方误差

$$P(k)=[1-cb(k)]P_1(k)$$

验前滤波估计均方误差

$$P(k)=\frac{b(k)}{c}\sigma_v^2$$

$$\frac{b(k)}{c}\sigma_v^2=[1-cb(k)]P_1(k)$$

滤波器估计增益：$b(k)=\dfrac{cP_1(k)}{c^2P_1(k)+\sigma_v^2}$ 验前滤波估计均方误差

总结：一维 kalman 滤波估计方程：

① 滤波方程：$\hat{S}(k)=a(k)\hat{S}(k-1)+b(k)x(k)$

$\downarrow\ a(k)=a[1-cb(k)]$

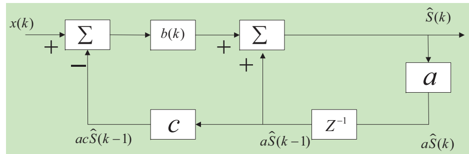
$$=a\hat{S}(k-1)+b(k)[x(k)-ac\hat{S}(k-1)]$$

$$=a[1-cb(k)]\hat{S}(k-1)+b(k)x(k)$$

② 验前滤波估计均方误差：$P_1(k)=a^2P(k-1)+\sigma_w^2$

③ 滤波估计增益：$b(k)=\dfrac{cP_1(k)}{c^2P_1(k)+\sigma_v^2}$ 验前滤波估计均方误差

④ 验后滤波估计均方误差：$P(k)=[1-cb(k)]P_1(k)$



$$\hat{S}(k)=a\hat{S}(k-1)+b(k)\underbrace{[x(k)-ac\hat{S}(k-1)]}_{\text{残差}}$$

例：设信号模型和测量模型为：

$S(k)=aS(k-1)+W(k-1)$， $x(k)=cS(k)+V(k)$， $R_s(j)=a^{|j|}\sigma_s^2$，

$\sigma_s^2=\dfrac{\sigma_w^2}{1-a^2}$， $P(0)=R_s(0)=\sigma_s^2=\dfrac{\sigma_w^2}{1-a^2}$， $\sigma_v^2=\sigma_w^2$， $a^2=\dfrac{1}{2}$，

$c=1$。求 $\hat{S}(k)$。

解：① $k=1$

a. $\hat{S}(1)=a[1-cb(1)]\hat{S}(k-1)+b(k)x(k)$

$\downarrow\ \hat{S}(0)=0$

$=b(1)x(1)$

b. 验前滤波估计均方误差：

$$P_1(1)=a^2P(0)+\sigma_w^2=a^2\frac{\sigma_w^2}{1-a^2}+\sigma_w^2=2\sigma_w^2$$

c. $$b(1)=\frac{cP_1(1)}{c^2P_1(1)+\sigma_v^2}=\frac{P_1(1)}{P_1(1)+\sigma_v^2}=\frac{\frac{\sigma_w^2}{1-a^2}}{\frac{\sigma_w^2}{1-a^2}+\sigma_v^2}=\frac{2}{3}\approx 0.67$$

$\hat{S}(1)=0.67x(1)$

d. 验后滤波估计均方误差：

$$P(1)=[1-cb(1)]P_1(1)=[1-\frac{2}{3}]P_1(1)=\frac{2}{3}\sigma_w^2$$

② $k=2$

a. $\hat{S}(2)=a[1-cb(2)]\hat{S}(1)+b(2)x(2)=a[1-b(2)]\hat{S}(1)+b(2)x(2)$

b. 验前滤波估计均方误差：

$$P_1(2)=a^2P(1)+\sigma_w^2=\frac{1}{2}P(1)+\sigma_w^2=\frac{1}{3}\sigma_w^2+\sigma_w^2=\frac{4}{3}\sigma_w^2$$

c. $$b(2)=\frac{cP_1(2)}{c^2P_1(2)+\sigma_v^2}=\frac{P_1(2)}{P_1(2)+\sigma_v^2}=\frac{4}{7}\approx 0.57$$

$$\hat{S}(2)=\sqrt{\frac{1}{2}}\cdot 0.43\hat{S}(1)+0.57x(2)$$

d. $$P(2)=[1-cb(2)]P_1(2)=[1-\frac{4}{7}]\frac{4}{3}\sigma_w^2=\frac{4}{7}\sigma_w^2\approx 0.57\sigma_w^2$$

③ $k=3$

a. $\hat{S}(3)=a[1-cb(3)]\hat{S}(2)+b(3)x(3)=a[1-b(3)]\hat{S}(2)+b(3)x(3)$

b. 验前滤波估计均方误差：

$$P_1(3)=a^2P(2)+\sigma_w^2=\frac{1}{2}P(2)+\sigma_w^2=\frac{1}{2}\cdot\frac{4}{7}\sigma_w^2+\sigma_w^2=\frac{9}{7}\sigma_w^2$$

c. $b(3)=\dfrac{cP_1(3)}{c^2P_1(3)+\sigma_v^2}=\dfrac{P_1(3)}{P_1(3)+\sigma_v^2}=\dfrac{9}{16}\approx 0.562$

$$\hat{S}(3)=\frac{7}{9}a\hat{S}(2)+\frac{7}{16}x(3)$$

d. $P(3)=[1-cb(3)]P_1(3)=[1-\frac{9}{16}]\frac{9}{7}\sigma_w^2=\frac{9}{16}\sigma_w^2\approx 0.562\sigma_w^2$

---

④ $k\uparrow$，$P(k)\downarrow$减小，$b(k)\to$一个稳定值，当$b(k)$稳定后会等于b，求b

$$b(k)=\frac{cP_1(k)}{c^2P_1(k)+\sigma_v^2}=\frac{P_1(k)}{P_1(k)+\sigma_v^2}$$

↓ $P_1(k)=a^2P(k-1)+\sigma_w^2$

↓ $P(k-1)=\dfrac{\sigma_v^2}{c}b(k-1)=\sigma_v^2 b(k-1)$

$$b(k)=\frac{a^2\sigma_v^2 b(k-1)+\sigma_w^2}{a^2\sigma_v^2 b(k-1)+\sigma_w^2+\sigma_v^2}$$

↓ $b(k-1)=b(k)=b$

$b=\dfrac{b+2}{b+4}\ \Rightarrow\ b=0.56$或$b=-0.356$，$b$负值无意义

$\therefore b$ 的稳定值为 **0.56**。

四、一维 kalman 预测器

① 信号方程：$S(k)=aS(k-1)+W(k-1)$

↓ $E[W(k-1)]=0$，$E[W(k)W(j)]=\sigma_w^2\delta_{kj}$

↓ $E[S^2(k)]=R_s(0)=\sigma_s^2=\dfrac{\sigma_w^2}{1-a^2}=P_s(0)$

↓ $E[S(k)S(k+j)]=R_s(j)=a^{|j|}P_s(0)$

② 测量方程：$x(k)=cS(k)+V(k)$

---

③ 预测方程：

k步预测估计：$\hat{S}(k+1/k)=\alpha(k)\hat{S}(k/k-1)+\beta(k)x(k)$

$\hat{S}(k/k-1)$：$k-1$步预估计，

$\beta(k)$：预测估计增益，$x(k)$：实测值

$\hat{S}(k+1/k)$：第k步预测出处$k+1$步的$S(k+1)$。

k步预测估计误差：$e(k+1/k)=S(k+1)-\hat{S}(k+1/k)$

k步预测估计均方误差：$P(k+1/k)=E\{e^2(k+1/k)\}$

↓ $\min P(k+1/k)$，正交原理

$E[e(k+1/k)\hat{S}(k/k-1)]=0$　　$E[e(k+1/k)x(k)]=0$

$E\{[S(k+1)-\hat{S}(k+1/k)]\hat{S}(k/k-1)\}=0$

$E\{[S(k+1)-\alpha(k)\hat{S}(k/k-1)-\beta(k)x(k)]\hat{S}(k/k-1)\}=0$

$\alpha(k)E\{[S(k)-\hat{S}(k/k-1)-S(k)]\hat{S}(k/k-1)\}=E\{[S(k+1)-\beta(k)x(k)]\hat{S}(k/k-1)\}$

↓ $S(k+1)=aS(k)+W(k)$

↓ $x(k)=cS(k)+V(k)$

↓ $e(k/k-1)=S(k)-\hat{S}(k/k-1)$

$\alpha(k)E\{[S(k)-e(k/k-1)]\hat{S}(k/k-1)\}$

$=E\{[(a-c\beta(k))S(k)+W(k)-\beta(k)x(k)]\hat{S}(k/k-1)\}$

↓ $E[e(k/k-1)\hat{S}(k/k-1)]=0$

↓ $E[W(k)\hat{S}(k/k-1)]=0$

↓ $E[V(k)\hat{S}(k/k-1)]=0$

$\alpha(k)E[S(k)\hat{S}(k/k-1)]=[a-c\beta(k)]E[S(k)\hat{S}(k/k-1)]$

↓

两参数关系：$\alpha(k)=a-c\beta(k)$

预测估计均方误差：

$P(k+1/k)=E\{e^2(k+1/k)\}=E\{e(k+1/k)S(k+1)\}$

↓ $S(k+1)=aS(k)+W(k)$

↓ $x(k)=cS(k)+V(k)\to S(k)=\dfrac{1}{c}[x(k)-V(k)]$

$=E\{e(k+1/k)[\dfrac{a}{c}x(k)-\dfrac{a}{c}V(k)+W(k)]\}$

↓ $E[e(k+1/k)x(k)]=0$

$=E\{e(k+1/k)[-\dfrac{a}{c}V(k)+W(k)]\}$

$=-\dfrac{a}{c}E[e(k+1/k)V(k)]+E[e(k+1/k)W(k)]$

↓ $e(k+1/k)=S(k+1)-\hat{S}(k+1/k)$

↓ $S(k+1)=aS(k)+W(k)$

↓ $e(k+1/k)=aS(k)+W(k)-\hat{S}(k+1/k)$

↓ $\hat{S}(k+1/k)=\alpha(k)\hat{S}(k/k-1)+\beta(k)x(k)$

↓ $=\alpha(k)\hat{S}(k/k-1)+c\beta(k)S(k)+\beta(k)V(k)$

↓ $e(k+1/k)=[a-c\beta(k)]S(k)-\alpha(k)\hat{S}(k/k-1)+W(k)-\beta(k)V(k)$

↓ $E[V(k)S(k)]=0$

↓ $E[V(k)\hat{S}(k/k-1)]=0$

↓ $E[V(k)W(k)]=0$

↓ $E[W(k)S(k)]=0$

↓ $E[W(k)\hat{S}(k/k-1)]=0$

预测估计均方误差：$P(k+1/k)=\dfrac{a}{c}\beta(k)\sigma_v^2+\sigma_w^2$

下面求$P(k+1/k)$

$P(k+1/k)=E\{[S(k+1)-\hat{S}(k+1/k)]^2\}$

↓ $\hat{S}(k+1/k)=\alpha(k)\hat{S}(k/k-1)+\beta(k)x(k)$

↓ $\alpha(k)=a-c\beta(k)$

↓ $x(k)=cS(k)+V(k)$

↓ $S(k+1)=aS(k)+W(k)$

$=E(\{[a-c\beta(k)][S(k)-\hat{S}(k/k-1)]+W(k)-\beta(k)V(k)\}^2)$

↓ $e(k/k-1)=S(k)-\hat{S}(k/k-1)$

$=E(\{[a-c\beta(k)]e(k/k-1)+W(k)-\beta(k)V(k)\}^2)$

↓ $P(k/k-1)=E[e^2(k/k-1)]$ 第$k-1$步

↓预测估计均方误差 $\sigma_w^2=E[W^2(k)]$

↓ $\sigma_v^2=E[V^2(k)]$

↓ $E[e(k/k-1)W(k)]=0$

↓ $E[e(k/k-1)V(k)]=0$

↓ $E[W(k)V(k)]=0$

$$P(k+1/k)=[a-c\beta(k)]^2P(k/k-1)+\sigma_w^2+\beta^2(k)\sigma_v^2$$

↓ $P(k+1/k)=\frac{a}{c}\beta(k)\sigma_v^2+\sigma_w^2$

预测估计增益：$\beta(k)=\frac{caP(k/k-1)}{c^2P(k/k-1)+\sigma_v^2}$

总结一维 kalman 一步预测估计方程组：

① 预测估计方程组：

$$\hat S(k+1/k)=\alpha(k)\hat S(k/k-1)+\beta(k)x(k)$$

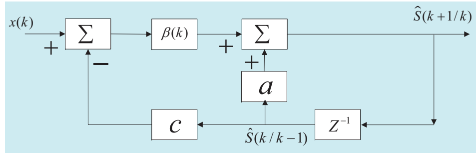
↓ $\alpha(k)=a-c\beta(k)$

$$=a\hat S(k/k-1)+\beta(k)[x(k)-c\hat S(k/k-1)]$$

$$=[a-c\beta(k)]\hat S(k/k-1)+\beta(k)x(k)$$

② 预测估计增益：$\beta(k)=\frac{caP(k/k-1)}{c^2P(k/k-1)+\sigma_v^2}$

③ 预测估计均方误差：$P(k+1/k)=\frac{a}{c}\beta(k)\sigma_v^2+\sigma_w^2$



一步预测

五、一维 kalman 同时滤波与一步预测

线性最小均方误差估计是无偏估计（王启才, P138）

↓

估计为条件均值（王启才, P96））

$\hat S(k)=E[S(k)/x(1),x(2),\cdots,x(k)]$ 条件均值

$\hat S(k+1/k)=E[S(k+1)/x(1),x(2),\cdots,x(k)]$ 条件均值

↓ $S(k+1)=aS(k)+W(k)$

$$=E[aS(k)/x(1),x(2),\cdots,x(k)]+E[W(k)/x(1),x(2),\cdots,x(k)]$$

可得到：$\hat S(k+1/k)=a\hat S(k)$

滤波估计方程：$\hat S(k)=a(k)\hat S(k-1)+b(k)x(k)$

↓ $a(k)=a[1-cb(k)]$

$$=a\hat S(k-1)+b(k)[x(k)-ac\hat S(k-1)]$$

↓ $\hat S(k/k-1)=a\hat S(k-1)$

$$=\hat S(k/k-1)+b(k)[x(k)-c\hat S(k/k-1)]$$

↓ $\hat S(k+1/k)=a\hat S(k)$，上式两边乘 a

$$\hat S(k+1/k)=a\hat S(k/k-1)+ab(k)[x(k)-c\hat S(k/k-1)]$$

预测估计方程：

$$\hat S(k+1/k)=\alpha(k)\hat S(k/k-1)+\beta(k)x(k)$$

↓ $\alpha(k)=a-c\beta(k)$

$$\hat S(k+1/k)=a\hat S(k/k-1)+\beta(k)[x(k)-c\hat S(k/k-1)]$$

由以上两式，$\beta(k)=ab(k)$（两种增益关系）

预测估计增益　　滤波估计增益

预测估计均方误差：

$$P(k+1/k)=\frac{a}{c}\beta(k)\sigma_v^2+\sigma_w^2$$

↓ $\beta(k)=ab(k)$

$$=\frac{a^2}{c}b(k)\sigma_v^2+\sigma_w^2$$

↓ $b(k)=\frac{cP(k)}{\sigma_v^2}$

$$P(k+1/k)=a^2P(k)+\sigma_w^2$$

预测估计均方误差　　验后滤波估计均方误差

总结一维 kalman 同时滤波与一步预测的方程组为：

预测估计方程：$\hat S(k/k-1)=a\hat S(k-1)$

验前滤波估计均方误差：$P_1(k)=a^2P(k-1)+\sigma_w^2$

（相当于预测误差协方差）$P_1(k/k-1)=a^2P(k-1)+\sigma_w^2$

滤波估计增益：$b(k)=\frac{cP_1(k)}{c^2P_1(k)+\sigma_v^2}$　$b(k)=\frac{cP_1(k/k-1)}{c^2P_1(k/k-1)+\sigma_v^2}$

滤波估计方程：$\hat S(k)=a(k)\hat S(k-1)+b(k)x(k)$

↓ $a(k)=a[1-cb(k)]$

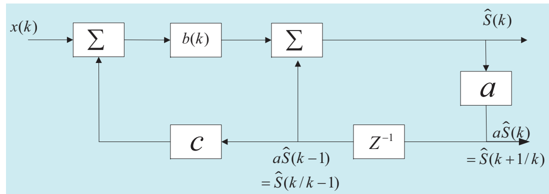
$$=a\hat S(k-1)+b(k)[x(k)-ac\hat S(k-1)]$$

↓ $\hat S(k/k-1)=a\hat S(k-1)$

$$\hat S(k)=\hat S(k/k-1)+b(k)[x(k)-c\hat S(k/k-1)]$$

验后滤波估计均方误差：$P(k)=[1-cb(k)]P_1(k)$

（相当于滤波误差协方差）$P(k)=[1-cb(k)]P_1(k/k-1)$



六、N 维 kalman 滤波器

kalman 滤波器的优点：多输入多输出，同时多个量估计。

1. 信号方程：k 步，N 个待估计量，$S_1(k), S_2(k), \cdots, S_N(k)$

$$
\begin{cases}
S_1(k) = a_{11}S_1(k-1) + a_{12}S_2(k-1) + \cdots + a_{1N}S_N(k-1) + W_1(k-1) \\
S_2(k) = a_{21}S_1(k-1) + a_{22}S_2(k-1) + \cdots + a_{2N}S_N(k-1) + W_2(k-1) \\
\vdots \\
S_N(k) = a_{N1}S_1(k-1) + a_{N2}S_2(k-1) + \cdots + a_{NN}S_N(k-1) + W_N(k-1)
\end{cases}
$$

$$
\begin{pmatrix} S_1(k) \\ S_2(k) \\ \vdots \\ S_N(k) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & & & \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix} \begin{pmatrix} S_1(k-1) \\ S_2(k-1) \\ \vdots \\ S_N(k-1) \end{pmatrix} + \begin{pmatrix} W_1(k-1) \\ W_2(k-1) \\ \vdots \\ W_N(k-1) \end{pmatrix}
$$

$$\underline{S}(k) = \underline{A}\underline{S}(k-1) + \underline{W}(k-1)$$

2. 测量方程：k 步，M 个测量值 $x_1(k), x_2(k), \cdots, x_M(k)$

$$
\begin{cases}
x_1(k) = c_{11}S_1(k) + c_{12}S_2(k) + \cdots + c_{1N}S_N(k) + V_1(k) \\
x_2(k) = c_{21}S_1(k) + c_{22}S_2(k) + \cdots + c_{2N}S_N(k) + V_2(k) \\
\vdots \\
x_M(k) = c_{M1}S_1(k) + c_{M2}S_2(k) + \cdots + c_{MN}S_N(k) + V_M(k)
\end{cases}
$$

$$
\begin{pmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_M(k) \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1N} \\ c_{21} & a_{22} & \cdots & c_{2N} \\ \vdots & & & \\ c_{M1} & c_{M2} & \cdots & c_{MN} \end{pmatrix} \begin{pmatrix} S_1(k) \\ S_2(k) \\ \vdots \\ S_N(k) \end{pmatrix} + \begin{pmatrix} V_1(k) \\ V_2(k) \\ \vdots \\ V_M(k) \end{pmatrix}
$$

$$\underline{X}(k) = \underline{C}\underline{S}(k) + \underline{V}(k)$$

七、 卡尔曼滤波的发散现象和克服发散的方法

从理论上讲，随着观测次数的增加，卡尔曼滤波的均方误差 P_k 只应该逐渐减小而最终趋于某个稳定值。但在实际应用中，有时会发生这样的现象：按公式计算的方差阵可能逐渐地趋于零，而实际滤波的均方误差会随着观测次数的增加而增大，这种现象称为卡尔曼滤波的发散现象。

产生发散的原因很多，其中信号模型不准确是重要原因之一。

卡尔曼滤波的发散现象产生的原因，除了信号模型不准的原因外，扰动噪声和观测噪声的统计特性取得不准，计算字长有限产生的量化误差等也是发散产生的原因。

克服卡尔曼滤波发散的方法归纳起来主要有如下几种。

①自适应滤波法　如果在滤波过程中，利用新的观测数据，对信号模型、噪声的统计特性等实时进行修正，以保持最优或次最优滤波，用这种自适应滤波方法可以抑制发散现象。

②限定滤波增益法　在卡尔曼滤波算法中，随着 k 的增大滤波增益将逐渐减小，即“新息”的修正作用越来越小。这样，由于模型等不准产生的误差得不到有效的抑制，容易产生滤波发散现象。这个方法就是强使滤波增益随着时间下降至某个数值后，不再随 k 的增加而减小，这是克服发散的方法之一。不过这样做的结果可能使滤波达不到最佳结果。

③渐消记忆法　卡尔曼滤波具有无限增长的记忆特性，它获得的滤波值 $\hat{x}_k$ 使用了 k 时刻以前的全部观测数据。但对动态模型来说，在进行滤波时，需加大新数据的作用，减小老数据的影响。这就是渐消记忆法滤波的基本思想。

④限定记忆法　这种方法与渐消记忆法的不同之处在于不是逐渐减小老数据的影响，而是在作 k 时刻滤波时，只利用离尼时刻最近的 N 个数据，把更前的数据丢掉，N 的数目的选择与信号模型的类型有关。

除上述克服滤波发散的方法外，还可以在状态方程模型中加大扰动白噪声 W_k，增加扰动方差，以避免均方误差阵减小太多，也是防止发散现象的一种方法。关于克服滤波发散现象的方法，我们就简单介绍这些，不再作具体分析和讨论。

八、 常增益滤波方法

滤波和预测的目的是估计当前和未来时刻目标的运动状态，包括位置、速度和加速度等。前面我们讨论了两种线性滤波器：维纳滤波器和卡尔曼滤波器，现在我们讨论另一种线性滤波器：常增益 $\alpha-\beta$ 与 $\alpha-\beta-\gamma$ 滤波器。

1　$\alpha-\beta$ 滤波

$\alpha-\beta$ 滤波器最早是为了改善边扫描边跟踪雷达的跟踪性能而提出的。这种体制的雷达，在天线波瓣一次扫过目标期间，大约能收到几个至几十个目标回波脉冲，但从一串回波脉冲到下一串回波脉冲则相隔一个天线扫描周期，可长达数秒。因而这种雷达的距离和角度跟踪系统处于采样跟踪状态，从而要求跟踪系统具有对目标状态进行预测的能力，以防止下一次取样之前雷达已失去对目标的跟踪。

$\alpha-\beta$ 滤波是以输入是均匀变化的信号为前提的。当得到 k 时刻的测量值 $x(k)$ 后，按下列方程对信号 $s(k)$ 进行估计

$$\hat{s}(k) = \hat{s}(k \mid k-1) + \alpha[x(k) - \hat{s}(k \mid k-1)] \quad (1)$$

$$\hat{\dot{s}}(k) = \hat{\dot{s}}(k \mid k-1) + \frac{\beta}{T}[x(k) - \hat{s}(k \mid k-1)] \quad (2)$$

一步预测为

$$\hat{s}(k \mid k-1) = \hat{s}(k-1) + T\,\hat{\dot{s}}(k-1) \quad (3)$$

$$\hat{\dot{s}}(k \mid k-1) = \hat{\dot{s}}(k-1) \quad (4)$$

式中 $\hat{s}(k)$ 是第 k 个周期的平滑坐标；

$\hat{s}(k/k-1)$ 是在第(k 一 1)个周期内计算所得的第 k 个周期的外推坐标；

$x(k)$ 是第 k 个周期录取的坐标；

$\hat{\dot{s}}(k)$ 是第 k 个周期的速度估计；

T 是天线扫描周期；

校正系数 α 和 β 分别为位置和速度平滑系数。它们可以是常数，也可以是随取样序列分段改变。我们引入以下记号：

滤波估计矢量　$\hat{\mathbf{s}}(k) = [\hat{s}(k)\,\hat{\dot{s}}(k)]^{\mathrm{T}}$

预测估计矢量

$$\hat{\mathbf{s}}(k|k-1)=[\hat{s}(k|k-1)\hat{\dot{s}}(k|k-1)]^{T}$$

状态转移矩阵 $A=\begin{bmatrix}1 & T\\ 0 & 1\end{bmatrix}$

观察转换矢量 $C=[10]$

滤波增益矩阵 $K=[\alpha\ \beta/T]^{T}$

则可写出 $\alpha-\beta$ 滤波的矢量矩阵表达式。式(1)和式(2)写成矩阵形式为

$$\hat{\mathbf{s}}(k\mid k-1)=A\hat{\mathbf{s}}(k-1) \quad (5)$$

同样，式(3)和式(4)写成矩阵形式为

$$\hat{\mathbf{s}}(k)=\mathbf{s}(k\mid k-1)+\mathbf{K}[x(k)-\mathbf{C}\hat{\mathbf{s}}(k\mid k-1)] \quad (6)$$

由上两式可知，$\alpha-\beta$ 滤波是一种递推滤波，其运算流程如图 1 所示。

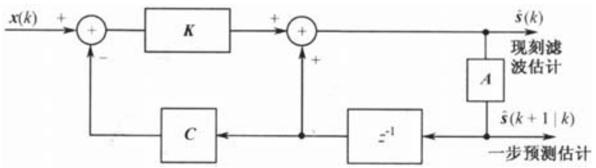
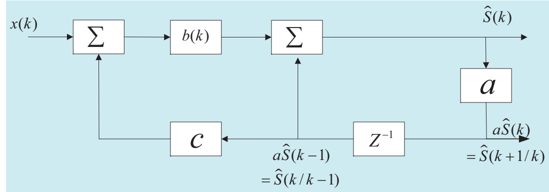


图 1　同时实现矢量信号滤波与预测的常增益 $\alpha-\beta$ 滤波运算流程



将式(5)和式(6)以及图 1 与卡尔曼滤波方程式和预测方程式以及相应图相比较，显然它们具有相同的结构形式，不同之处仅在于卡尔曼滤波增益 $b(k)$ 是时变的，而 $\alpha-\beta$ 滤波增益 K 是恒定的，非时变的。因此，可以认为 $\alpha-\beta$ 滤波是卡尔曼滤波的特例。

$\alpha-\beta$ 滤波器实质上是一个二阶线性定常系统，它有两个自由参数，即 α 和 β。因此可以用线性系统的理论讨论 $\alpha-\beta$ 滤波器的性质和选择 α 和 β 的数值。通过计算可以求得位置估值和速度估值的等效传递函数 $H_{s}(z)$ 和 $H_{\dot{s}}(z)$ 分别为

$$H_{s}(z)=\frac{\hat{s}(z)}{x(z)}=\frac{\alpha z\left(z+\frac{\beta-\alpha}{\alpha}\right)}{z^{2}-(2-\alpha-\beta)z+(1-\alpha)} \quad (7)$$

$$H_{\dot{s}}(z)=\frac{\hat{\dot{s}}(z)}{x(z)}=\frac{\frac{\beta}{T}z(z-1)}{z^{2}-(2-\alpha-\beta)z+(1-\alpha)} \quad (8)$$

由上两式可以看出，滤波器的特征方程为

$$z^{2}-(2-\alpha-\beta)z+(1-\alpha)=0 \quad (9)$$

根据稳定性判据可知，滤波器的稳定区域是由

$$2\alpha+\beta<4,\ 0<\alpha<2,\ 0<\beta<4$$

所规定的三角形，如图 2 所示。只要 α 和 β 的取值落在这个区域之内，滤波器就是稳定的。

α 和 β 值的选择除了满足稳定性条件以外，还要考虑对滤波器暂态和稳态性能的要求。就滤波器的暂态特性而言，可分为过阻尼、欠阻尼和临界阻尼三种情况。这三种情况的区域划分如图 3 所示。工程上常采用临界阻尼状态。

临界阻尼状态时，特征方程的根是两个正的重根 r，由此可建立起求解 $\alpha-\beta$ 参数值的关系式来，即

$$(z-r)^{2}=z^{2}-(2-\alpha-\beta)z-(1-\alpha)$$

比较上式两边的各个系数，可得

$$2r=2-\alpha-\beta$$

$$r^{2}=1-\alpha$$

因此临界阻尼状态时 α 和 β 的关系式为

$$\beta=2-\alpha-2\sqrt{1-\alpha} \quad (10)$$

α 是位于 0～1 的某一个数值，它的数值可以根据系统所要求的阻尼比和自然谐振频率来确定。若要求系统的带宽越大，α 就应越大，这时平滑越差；反之，α 越小，系统带宽就越小，平滑将越好。

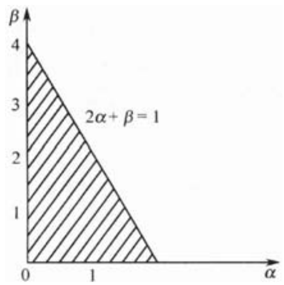


图 2　滤波器的稳定区域

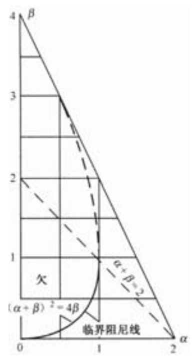


图 3　稳定区域的划分

2　$\alpha-\beta-\gamma$ 滤波

对于一个等加速输入过程可用 $\alpha-\beta-\gamma$ 法来实现递推滤波。

$\alpha-\beta-\gamma$ 滤波的滤波方程组为

$$\left.\begin{aligned}
\hat{s}(k)&=\hat{s}(k|k-1)+\alpha[x(k)-\hat{s}(k|k-1)]\\
\hat{\dot{s}}(k)&=\hat{\dot{s}}(k|k-1)+\frac{\beta}{T}[x(k)-\hat{s}(k|k-1)]\\
\hat{\ddot{s}}(k)&=\hat{\ddot{s}}(k|k-1)+\frac{2\gamma}{T^{2}}[x(k)-\hat{s}(k|k-1)]
\end{aligned}\right\} \quad (11)$$

预测方程组为

$$\left.\begin{aligned}
\hat{s}(k|k-1) &= \hat{s}(k-1)+\hat{\dot{s}}(k-1)T+\hat{\ddot{s}}(k-1)\frac{T^2}{2}\\
\hat{\dot{s}}(k|k-1) &= \hat{\dot{s}}(k-1)+\hat{\ddot{s}}(k-1)T\\
\hat{\ddot{s}}(k|k-1) &= \hat{\ddot{s}}(k-1)
\end{aligned}\right\} \quad (12)$$

其矩阵表达式为

$$\left.\begin{aligned}
\hat{\mathbf{s}}(k|k-1) &= \mathbf{A}\hat{\mathbf{s}}(k-1)\\
\hat{\mathbf{s}}(k) &= \hat{\mathbf{s}}(k|k-1)+\mathbf{K}[\mathbf{x}(k)-\mathbf{CA}\hat{\mathbf{s}}(k|k-1)]
\end{aligned}\right\} \quad (13)$$

式中

滤波估计矢量 $\hat{\mathbf{s}} = [\hat{s}\ \hat{\dot{s}}\ \hat{\ddot{s}}]_k^{\mathrm{T}}$

预测估计矢量 $\hat{\mathbf{s}}(k|k-1) = [\hat{s}\ \hat{\dot{s}}\ \hat{\ddot{s}}]_{k|k-1}^{\mathrm{T}}$

观察转换矢量 $\mathbf{C} = [1 0 0]$

滤波增益矩阵 $\mathbf{K} = \left[\alpha\ \frac{\beta}{T}\ \frac{2\gamma}{T^2}\right]^{\mathrm{T}}$

状态转移矩阵 $\mathbf{A} = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}$

系统的位置、速度和加速度估值对应输入测量值的等效传递函数为

$$\left.\begin{aligned}
H_1(z) &= \frac{[\alpha^3+(-2\alpha+\beta+\gamma)z^2+(\alpha-\beta+\gamma)]z}{z^3-(3-\alpha-\beta-\gamma)z^2+(3-2\alpha-\beta+\gamma)z-(1-\alpha)}\\
H_2(z) &= \frac{\left(\frac{\beta}{T}z\right)(z-1)\left[z+\frac{2\gamma-\beta}{\beta}\right]}{z^3-(3-\alpha-\beta-\gamma)z^2+(3-2\alpha-\beta+\gamma)z-(1-\alpha)}\\
H_3(z) &= \frac{\left(\frac{2\gamma}{T^2}z\right)(z-1)(z-1)}{z^3-(3-\alpha-\beta-\gamma)z^2+(3-2\alpha-\beta+\gamma)z-(1-\alpha)}
\end{aligned}\right\} \quad (14)$$

因此系统的特征方程为

$$z^3-(3-\alpha-\beta-\gamma)z^2+(3-2\alpha-\beta+\gamma)z-(1-\alpha)=0 \quad (15)$$

由稳定性判据可知，系统的稳定条件为

$$\alpha>0,\ \beta>0,\ \gamma>0,\ 2\alpha+\beta\leqslant 4,$$
$$2\alpha>\beta,\ \alpha(\beta+\gamma)>2\gamma \quad (16)$$

α、β 和 γ，的参数值选择与 $\alpha-\beta$ 滤波器的参数值选择一样，应考虑系统的稳定性以及暂态和稳态特性。

在临界阻尼状态选择时，系统特征方程的根是三重正实根 r，即

$$(z-r)^3 = z^3-(3-\alpha-\beta-\gamma)z^2+$$
$$(3-2\alpha-\beta-\gamma)z-(1-\alpha)$$

由此可得 α、β 和 γ 与根 r 的关系式

$$\alpha = 1-r^3 \quad (17)$$

$$\beta = 1.5(1-r^2)(1-r) \quad (18)$$

$$\gamma = 0.5(1-r)^3 \quad (19)$$

给定 α 值后，可求得 r 值，从而得到 β 和 γ 的数值。

值得指出的是，常增益滤波器具有卡尔曼滤波器的结构形式，因此可以根据卡尔曼滤波理论来选择滤波增益阵。这样就为设计常增益滤波器提供了更多的方法，同时也建立起常增益滤波器与卡尔曼滤波器之间的联系。

$\alpha-\beta$ 滤波适合于目标做匀速运动的情况，$\alpha-\beta-\gamma$ 滤波适合于目标做匀加速运动的情况。在稳态时它们与卡尔曼滤波是等效的，但是在暂态过程中，或者目标做随机机动变化时，它们的性能不如卡尔曼滤波。

由于 $\alpha-\beta$ 滤波和 $\alpha-\beta-\gamma$ 滤波是两种较简单且易于工程实现的常增益滤波方法，所以已被广泛应用于对数据处理精度要求不高或者数据处理量较大的跟踪滤波器的设计过程中。

九、小结

1. 卡尔曼和维纳滤波器都是在给定并应用随机信号噪声的前二阶矩的统计特性，以线性最小均方误差估计为准则来滤波。
2. 维纳滤波器要求信号和噪声的相关函数已知，适合平稳过程的波形估计；卡尔曼滤波器要消息模型和观察模型已知，可用于平稳或非平稳过程。
3. 卡尔曼滤波器是递推算法，只需当前观察数据和前一步处理数据，适合实时处理；维纳滤波器估计时需要全部的历史数据，但对于非实时处理，维纳滤波器仍是一种有效的估计方法。
4. 维纳滤波器是单输出滤波器，但可以多输入；卡尔曼滤波器可以多输入和多输出。