其他一些最优估计方法介绍

2023年1月12日 14:44

卡尔曼滤波与组合导航原理(6-7讲)

卡尔曼滤波与组合导航原理

西北工业大学自动化学院 严恭敏

2020-04

1

18 其他一些最优估计方法介绍

1. 最小方差估计和线性最小方差估计
2. 极大似然估计
3. 极大验后估计
4. 加权最小二乘估计
5. 贝叶斯估计
6. 维纳滤波
7. 递推贝叶斯估计

2

18 其他最优估计方法介绍

估计方法的主要发展历程：

估计理论是概率论与数理统计的一个分支，它是根据受扰动的观测数据来提取系统某些参数或状态的一种数学方法。

- 1795年高斯（K. Gauss）提出了最小二乘法，成功应用于谷神星轨道估计；
- 1912年费歇尔（R. A. Fisher）提出了极大似然估计法，从概率密度角度考虑估计问题；
- 1940年维纳（N. Wiener）提出了在频域中设计统计最优滤波器的方法，称为维纳滤波，但它只能处理平稳随机过程问题且滤波器设计复杂，应用受到很大限制；

静态的概念.

- 1960年卡尔曼（R. E. Kalman）提出了一种最优递推滤波方法，称为Kalman滤波，它既适用于平稳随机过程，也适用于非平稳过程，一经提出便得到了广泛应用。

3

18.1 随机向量的概率密度函数

记连续型随机向量

$$\boldsymbol{X}=\begin{bmatrix}X_1\\X_2\\\vdots\\X_n\end{bmatrix}\qquad \boldsymbol{Z}=\begin{bmatrix}Z_1\\Z_2\\\vdots\\Z_m\end{bmatrix}\qquad \boldsymbol{Y}=\begin{bmatrix}\boldsymbol{X}\\\boldsymbol{Z}\end{bmatrix}=\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_{n+m}\end{bmatrix}$$

Y 的联合概率密度函数记为

$$p(\boldsymbol{y})=p(y_1,y_2,\cdots,y_{n+m})=p(\boldsymbol{x},\boldsymbol{z})=p(x_1,x_2,\cdots,x_n,z_1,z_2,\cdots,z_m)$$

数学期望和方差阵

$$\boldsymbol{m}_Y=\mathrm{E}[\boldsymbol{Y}]=\int_{-\infty}^{\infty}\boldsymbol{y}p(\boldsymbol{y})\,\mathrm{d}\,\boldsymbol{y}=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\begin{bmatrix}y_1\\\vdots\\y_{n+m}\end{bmatrix}p(y_1,\cdots,y_{n+m})\,\mathrm{d}\,y_1\cdots\mathrm{d}\,y_{n+m}$$

$$\boldsymbol{C}_Y=\mathrm{Var}[\boldsymbol{Y}]=\mathrm{Cov}(\boldsymbol{Y},\boldsymbol{Y})=\mathrm{E}\left[(\boldsymbol{Y}-\mathrm{E}[\boldsymbol{Y}])(\boldsymbol{Y}-\mathrm{E}[\boldsymbol{Y}])^{\mathrm{T}}\right]=\int_{-\infty}^{\infty}(\boldsymbol{y}-\mathrm{E}[\boldsymbol{Y}])(\boldsymbol{y}-\mathrm{E}[\boldsymbol{Y}])^{\mathrm{T}}p(\boldsymbol{y})\,\mathrm{d}\,\boldsymbol{y}$$

恒等式：$\boldsymbol{C}_{XZ}=\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})=\mathrm{E}\left[\boldsymbol{X}\boldsymbol{Z}^{\mathrm{T}}\right]-\boldsymbol{m}_X\boldsymbol{m}_Z^{\mathrm{T}}$

4

18.1 随机向量的概率密度函数

Y_i 的边缘概率密度函数

$$p_{Y_i}(y_i)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(y_1,\cdots y_{i-1},y_i,y_{i+1},\cdots,y_{n+m})\,\mathrm{d}\,y_1\cdots\mathrm{d}\,y_{i-1}\,\mathrm{d}\,y_{i+1}\cdots\mathrm{d}\,y_{n+m}$$

Z 的边缘概率密度函数

$$p_Z(\boldsymbol{z})=\int_{-\infty}^{\infty}p(\boldsymbol{x},\boldsymbol{z})\,\mathrm{d}\,\boldsymbol{x}=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x_1,\cdots,x_n,z_1\cdots,z_m)\,\mathrm{d}\,x_1\cdots\mathrm{d}\,x_n$$

在 ***Z=z*** 条件下 ***X*** 的条件概率密度函数

$$p(\boldsymbol{x}\mid\boldsymbol{z})=\frac{p(\boldsymbol{x},\boldsymbol{z})}{p_Z(\boldsymbol{z})}$$

条件数学期望和条件方差阵

$$\boldsymbol{m}_{X|Z}=\mathrm{E}[\boldsymbol{X}\mid\boldsymbol{z}]=\int_{-\infty}^{\infty}\boldsymbol{x}p(\boldsymbol{x}\mid\boldsymbol{z})\,\mathrm{d}\,\boldsymbol{x}$$

$$\boldsymbol{C}_{X|Z}=\mathrm{Cov}(\boldsymbol{X}\mid\boldsymbol{z},\boldsymbol{X}\mid\boldsymbol{z})=\mathrm{E}\left[\left(\boldsymbol{x}-\mathrm{E}[\boldsymbol{X}\mid\boldsymbol{z}]\right)\left(\boldsymbol{x}-\mathrm{E}[\boldsymbol{X}\mid\boldsymbol{z}]\right)^{\mathrm{T}}\right]$$

$$=\int_{-\infty}^{\infty}\left(\boldsymbol{x}-\mathrm{E}[\boldsymbol{X}\mid\boldsymbol{z}]\right)\left(\boldsymbol{x}-\mathrm{E}[\boldsymbol{X}\mid\boldsymbol{z}]\right)^{\mathrm{T}}p(\boldsymbol{x}\mid\boldsymbol{z})\,\mathrm{d}\,\boldsymbol{x}$$

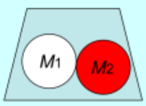
5

18.1 随机向量的概率密度函数

贝叶斯公式

$$\begin{cases}p(\boldsymbol{x}\mid\boldsymbol{z})=\dfrac{p(\boldsymbol{x},\boldsymbol{z})}{p_Z(\boldsymbol{z})}\qquad p(\boldsymbol{z}\mid\boldsymbol{x})=\dfrac{p(\boldsymbol{x},\boldsymbol{z})}{p_X(\boldsymbol{x})}\\ p(\boldsymbol{x},\boldsymbol{z})=p(\boldsymbol{x}\mid\boldsymbol{z})p_Z(\boldsymbol{z})=p(\boldsymbol{z}\mid\boldsymbol{x})p_X(\boldsymbol{x})\end{cases}$$

直观例子：罐子里有 M_1 个白球和 M_2 个红球。
(1)第一次取到红球的概率为 $M_2/(M_1+M_2)$；
(2)第二次又取到红球的概率为 $(M_2-1)/(M_1+M_2-1)$，有



$$\frac{M_2-1}{M_1+M_2-1}=\left(\frac{M_2}{M_1+M_2}\cdot\frac{M_2-1}{M_1+M_2-1}\right)\Big/\frac{M_2}{M_1+M_2}$$

即条件概率公式 $P(2^{nd}\ \mathrm{red}\mid 1^{st}\ \mathrm{red})=\dfrac{P(1^{st}\ \mathrm{red},2^{nd}\ \mathrm{red})}{P(1^{st}\ \mathrm{red})}$

6

18.2 最优估计的基本概念

观测模型方程

$$\boldsymbol{Z} = \boldsymbol{h}(\boldsymbol{X}, \boldsymbol{V})$$

m维观测向量　n维系统状态向量　干扰噪声向量　m维向量函数

估计

$$\hat{\boldsymbol{X}} = \hat{\boldsymbol{X}}(\boldsymbol{Z})$$ 估计是以系统的输出作为输入得到状态的观测

估计误差

$$\tilde{\boldsymbol{X}} = \boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})$$

估计的均方误差（Mean Square Error，MSE）阵

$$\mathrm{MSE}\left[\hat{\boldsymbol{X}}\right] = \mathrm{E}\left[\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\right] = \mathrm{E}\left[\left[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})\right]\left[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})\right]^{\mathrm{T}}\right]$$

所谓最优估计，是指在某一估计准则（指标函数）条件下，按照统计意义使估计达到最优。准则不同，估计方法就不同，估计结果往往也不相同。

7

→ 基本理论相同，比较 最优估计 → 对状态量大小的估计

机器学习 → 主要用于分类/预测

?

关心已有状态的不同　关心未来的变化

最大区别：估计中有量测和状态方程；

机器学习仅需一组数据

18.2 最优估计的基本概念

1）平方指标函数（损失函数，cost function）→ 机器学习中的损失函数?

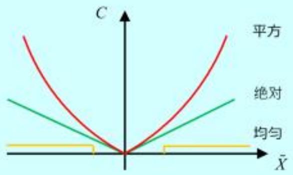
$$C(\hat{\boldsymbol{X}}) = (\boldsymbol{X} - \hat{\boldsymbol{X}})^{\mathrm{T}}(\boldsymbol{X} - \hat{\boldsymbol{X}})$$

2）绝对指标函数

$$C(\hat{\boldsymbol{X}}) = \left\|\boldsymbol{X} - \hat{\boldsymbol{X}}\right\|$$

3）均匀指标函数

$$C(\hat{\boldsymbol{X}}) = \begin{cases} 0, & \left\|\boldsymbol{X} - \hat{\boldsymbol{X}}\right\| < \varepsilon/2 \\ 1/\varepsilon, & \left\|\boldsymbol{X} - \hat{\boldsymbol{X}}\right\| \ge \varepsilon/2 \end{cases}$$



8

18.3 最小方差估计和线性最小方差估计

（1）最小方差估计（Minimum Variance，MV）

$$J(\hat{\boldsymbol{X}}) = \mathrm{E}\left[\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{X}}\right]\Big|_{\hat{\boldsymbol{X}}=\hat{\boldsymbol{X}}_{\mathrm{MV}}} = \min$$

$$J(\hat{\boldsymbol{X}}) = \mathrm{E}\left[\left[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})\right]^{\mathrm{T}}\left[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})\right]\right]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\boldsymbol{x} - \hat{\boldsymbol{X}}(\boldsymbol{z})\right]^{\mathrm{T}}\left[\boldsymbol{x} - \hat{\boldsymbol{X}}(\boldsymbol{z})\right]p(\boldsymbol{x},\boldsymbol{z})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$ 定义

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\boldsymbol{x} - \hat{\boldsymbol{X}}(\boldsymbol{z})\right]^{\mathrm{T}}\left[\boldsymbol{x} - \hat{\boldsymbol{X}}(\boldsymbol{z})\right]p(\boldsymbol{x}\mid\boldsymbol{z})p_Z(\boldsymbol{z})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}$$ 分为联合和边缘.

$$= \int_{-\infty}^{\infty}p_Z(\boldsymbol{z})\left\{\int_{-\infty}^{\infty}\left[\boldsymbol{x} - \hat{\boldsymbol{X}}(\boldsymbol{z})\right]^{\mathrm{T}}\left[\boldsymbol{x} - \hat{\boldsymbol{X}}(\boldsymbol{z})\right]p(\boldsymbol{x}\mid\boldsymbol{z})\,\mathrm{d}\boldsymbol{x}\right\}\mathrm{d}\boldsymbol{z}$$

$$= \int_{-\infty}^{\infty}p_Z(\boldsymbol{z})\left\{\int_{-\infty}^{\infty}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}p(\boldsymbol{x}\mid\boldsymbol{z})\,\mathrm{d}\boldsymbol{x} - 2\hat{\boldsymbol{X}}^{\mathrm{T}}(\boldsymbol{z})\int_{-\infty}^{\infty}\boldsymbol{x}p(\boldsymbol{x}\mid\boldsymbol{z})\,\mathrm{d}\boldsymbol{x} + \hat{\boldsymbol{X}}^{\mathrm{T}}(\boldsymbol{z})\hat{\boldsymbol{X}}(\boldsymbol{z})\int_{-\infty}^{\infty}p(\boldsymbol{x}\mid\boldsymbol{z})\,\mathrm{d}\boldsymbol{x}\right\}\mathrm{d}\boldsymbol{z}$$

$$= \int_{-\infty}^{\infty}p_Z(\boldsymbol{z})\left\{\mathrm{E}_X\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\mid\boldsymbol{z}\right] - 2\hat{\boldsymbol{X}}^{\mathrm{T}}(\boldsymbol{z})\mathrm{E}_X\left[\boldsymbol{X}\mid\boldsymbol{z}\right] + \hat{\boldsymbol{X}}^{\mathrm{T}}(\boldsymbol{z})\hat{\boldsymbol{X}}(\boldsymbol{z})\right\}\mathrm{d}\boldsymbol{z}$$

$$= \int_{-\infty}^{\infty}p_Z(\boldsymbol{z})\Big\{\mathrm{E}_X\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\mid\boldsymbol{z}\right] - \mathrm{E}_X^{\mathrm{T}}\left[\boldsymbol{X}\mid\boldsymbol{z}\right]\mathrm{E}_X\left[\boldsymbol{X}\mid\boldsymbol{z}\right] + \mathrm{E}_X^{\mathrm{T}}\left[\boldsymbol{X}\mid\boldsymbol{z}\right]\mathrm{E}_X\left[\boldsymbol{X}\mid\boldsymbol{z}\right] - 2\hat{\boldsymbol{X}}^{\mathrm{T}}(\boldsymbol{z})\mathrm{E}_X\left[\boldsymbol{X}\mid\boldsymbol{z}\right] + \hat{\boldsymbol{X}}^{\mathrm{T}}(\boldsymbol{z})\hat{\boldsymbol{X}}(\boldsymbol{z})\Big\}\mathrm{d}\boldsymbol{z}$$

$$= \int_{-\infty}^{\infty}\underbrace{p_Z(\boldsymbol{z})}_{>0}\left\{\mathrm{E}_X\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\mid\boldsymbol{z}\right] - \mathrm{E}_X^{\mathrm{T}}\left[\boldsymbol{X}\mid\boldsymbol{z}\right]\mathrm{E}_X\left[\boldsymbol{X}\mid\boldsymbol{z}\right]\right\}\mathrm{d}\boldsymbol{z} + \int_{-\infty}^{\infty}\underbrace{p_Z(\boldsymbol{z})}_{>0}\left[\mathrm{E}_X\left[\boldsymbol{X}\mid\boldsymbol{z}\right] - \hat{\boldsymbol{X}}(\boldsymbol{z})\right]^{\mathrm{T}}\left[\mathrm{E}_X\left[\boldsymbol{X}\mid\boldsymbol{z}\right] - \hat{\boldsymbol{X}}(\boldsymbol{z})\right]\mathrm{d}\boldsymbol{z}$$

9

18.3 最小方差估计和线性最小方差估计

$$J(\hat{X})=\int_{-\infty}^{\infty} p_Z(z)\left\{\underbrace{\mathrm{E}_X\left[X^{\mathrm{T}}X\mid z\right]-\mathrm{E}_X^{\mathrm{T}}[X\mid z]\mathrm{E}_X[X\mid z]}_{\text{非负}}\right\}\mathrm{d}z+$$

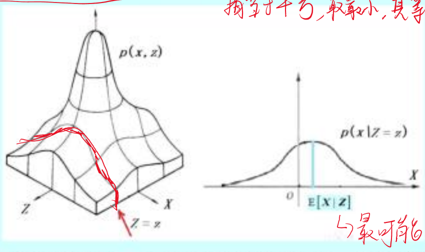
$$\int_{-\infty}^{\infty} p_Z(z)\underbrace{\left[\mathrm{E}_X[X\mid z]-\hat{X}(z)\right]^{\mathrm{T}}\left[\mathrm{E}_X[X\mid z]-\hat{X}(z)\right]}_{\text{与 }\hat{X}\text{ 无关}}\mathrm{d}z=\min$$

相当于平方，取最小，其等于0

$\Rightarrow$ $\mathrm{E}_X[X\mid z]-\hat{X}(z)=0$

$\Rightarrow$ $\hat{X}(z)=\mathrm{E}_X[X\mid z]$

$\Rightarrow$ $\hat{X}_{\mathrm{MV}}(Z)=\mathrm{E}[X\mid Z]$



看图解理解

与最可能取值的地方

最小方差估计等于某一观测实现**Z=z**条件下的条件均值，所以有时也称最小方差估计为(验后)条件期望估计。

18.3 最小方差估计和线性最小方差估计

最小方差估计的无偏性

$$\mathrm{E}_Z\left[\hat{X}_{\mathrm{MV}}(Z)\right]=\mathrm{E}_Z\left[\mathrm{E}_X[X\mid Z]\right]=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} xp(x\mid z)\mathrm{d}x\right]p(z)\mathrm{d}z$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xp(x\mid z)p(z)\mathrm{d}x\mathrm{d}z=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xp(x,z)\mathrm{d}z\mathrm{d}x$$

$$=\int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} p(x,z)\mathrm{d}z\mathrm{d}x=\int_{-\infty}^{\infty} xp_X(x)\mathrm{d}x=\mathrm{E}[X]$$

均方误差阵

$$\mathrm{E}\left[\tilde{X}_{\mathrm{MV}}\tilde{X}_{\mathrm{MV}}^{\mathrm{T}}\right]=\mathrm{E}\left[\left[X-\hat{X}_{\mathrm{MV}}(z)\right]\left[X-\hat{X}_{\mathrm{MV}}(z)\right]^{\mathrm{T}}\right]$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[x-\hat{X}_{\mathrm{MV}}(z)\right]\left[x-\hat{X}_{\mathrm{MV}}(z)\right]^{\mathrm{T}} p(x,z)\mathrm{d}x\mathrm{d}z$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x-\mathrm{E}[X\mid z]\right)\left(x-\mathrm{E}[X\mid z]\right)^{\mathrm{T}} p(x\mid z)\mathrm{d}x p_Z(z)\mathrm{d}z$$

$$=\int_{-\infty}^{\infty} C_{X\mid Z}\,p_Z(z)\mathrm{d}z$$

18.3 最小方差估计和线性最小方差估计

X

正态分布的最小方差估计

$$\hat{X}(z)=\mathrm{E}_X[X\mid z]=\int_{-\infty}^{\infty} xp(x\mid z)\mathrm{d}x=\int_{-\infty}^{\infty} x\frac{p(x,z)}{p_Z(z)}\mathrm{d}x$$

最小方差估计需要知道联合密度函数或条件概率密度函数，这在很多情况下是得不到的，或者是难以计算的。但是，如果被估计量和观测量都服从正态分布，则问题的求解会得到极大简化。

假设随机向量的正态分布

$$Y\sim N\left(m_Y,C_Y\right)\quad\text{即}\quad\begin{bmatrix}X\\Z\end{bmatrix}\sim N\left(\begin{bmatrix}m_X\\m_Z\end{bmatrix},\begin{bmatrix}C_X & C_{XZ}\\C_{ZX} & C_Z\end{bmatrix}\right)$$

$$p(y)=p(x,z)=\frac{1}{(2\pi)^{(n+m)/2}\left|C_Y\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(y-m_Y\right)^{\mathrm{T}}C_Y^{-1}\left(y-m_Y\right)\right\}$$

$$p(\boldsymbol{y})=p(\boldsymbol{x},\boldsymbol{z})=\frac{1}{(2\pi)^{(n+m)/2}\left|\boldsymbol{C}_Y\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{m}_Y)^{\mathrm{T}}\boldsymbol{C}_Y^{-1}(\boldsymbol{y}-\boldsymbol{m}_Y)\right\}$$

18.3 最小方差估计和线性最小方差估计

预备公式

$$\boldsymbol{C}_Y^{-1}=\begin{bmatrix}\boldsymbol{C}_X & \boldsymbol{C}_{XZ}\\ \boldsymbol{C}_{ZX} & \boldsymbol{C}_Z\end{bmatrix}^{-1}=\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ -\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{ZX} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}(\boldsymbol{C}_X-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{ZX})^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{C}_Z^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & -\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}\\ \boldsymbol{0} & \boldsymbol{I}\end{bmatrix}$$ →矩阵求逆

$$\left|\boldsymbol{C}_Y\right|=\begin{vmatrix}\boldsymbol{C}_X & \boldsymbol{C}_{XZ}\\ \boldsymbol{C}_{ZX} & \boldsymbol{C}_Z\end{vmatrix}=\left|\boldsymbol{C}_X-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{ZX}\right|\cdot\left|\boldsymbol{C}_Z\right|$$ →行列式计算

记号

$$\begin{cases}\boldsymbol{m}_{X|Z}=\boldsymbol{m}_X+\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}(\boldsymbol{z}-\boldsymbol{m}_Z)\\ \boldsymbol{C}_{X|Z}=\boldsymbol{C}_X-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{ZX}\end{cases}$$

则有

$$p(\boldsymbol{x},\boldsymbol{z})=\frac{1}{(2\pi)^{(n+m)/2}\left(\left|\boldsymbol{C}_{X|Z}\right|\cdot\left|\boldsymbol{C}_Z\right|\right)^{1/2}}\exp\left\{-\frac{1}{2}\left(\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{z}\end{bmatrix}-\begin{bmatrix}\boldsymbol{m}_X\\ \boldsymbol{m}_Z\end{bmatrix}\right)^{\mathrm{T}}\times\right.$$

$$\left.\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ -\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{ZX} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{C}_{X|Z}^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{C}_Z^{-1}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{I} & -\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}\\ \boldsymbol{0} & \boldsymbol{I}\end{bmatrix}\left(\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{z}\end{bmatrix}-\begin{bmatrix}\boldsymbol{m}_X\\ \boldsymbol{m}_Z\end{bmatrix}\right)\right\}$$

$$=\frac{1}{(2\pi)^{n/2+m/2}\left|\boldsymbol{C}_{X|Z}\right|^{1/2}\cdot\left|\boldsymbol{C}_Z\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})^{\mathrm{T}}\boldsymbol{C}_{X|Z}^{-1}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{m}_Z)^{\mathrm{T}}\boldsymbol{C}_Z^{-1}(\boldsymbol{z}-\boldsymbol{m}_Z)\right\}$$

$$=\frac{1}{(2\pi)^{n/2}\left|\boldsymbol{C}_{X|Z}\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})^{\mathrm{T}}\boldsymbol{C}_{X|Z}^{-1}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})\right\}\times$$

$$\frac{1}{(2\pi)^{m/2}\left|\boldsymbol{C}_Z\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{m}_Z)^{\mathrm{T}}\boldsymbol{C}_Z^{-1}(\boldsymbol{z}-\boldsymbol{m}_Z)\right\}$$

18.3 最小方差估计和线性最小方差估计

$$p(\boldsymbol{x},\boldsymbol{z})=\frac{1}{(2\pi)^{n/2}\left|\boldsymbol{C}_{X|Z}\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})^{\mathrm{T}}\boldsymbol{C}_{X|Z}^{-1}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})\right\}\times\frac{1}{(2\pi)^{m/2}\left|\boldsymbol{C}_Z\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{m}_Z)^{\mathrm{T}}\boldsymbol{C}_Z^{-1}(\boldsymbol{z}-\boldsymbol{m}_Z)\right\}$$

$$\Longrightarrow p(\boldsymbol{x},\boldsymbol{z})=p(\boldsymbol{x}\,|\,\boldsymbol{z})p_Z(\boldsymbol{z})$$ →贝叶斯公式成立

$$p(\boldsymbol{x}\,|\,\boldsymbol{z})=\frac{1}{(2\pi)^{n/2}\left|\boldsymbol{C}_{X|Z}\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})^{\mathrm{T}}\boldsymbol{C}_{X|Z}^{-1}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})\right\}$$

正态分布的最小方差估计及其均方误差阵：

$$\begin{cases}\hat{\boldsymbol{X}}_{\mathrm{MV}}=\mathrm{E}\left[\boldsymbol{X}\,|\,\boldsymbol{z}\right]=\int_{-\infty}^{\infty}\boldsymbol{x}p(\boldsymbol{x}\,|\,\boldsymbol{z})\mathrm{d}\boldsymbol{z}=\boldsymbol{m}_{X|Z}=\boldsymbol{m}_X+\underline{\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}(\boldsymbol{z}-\boldsymbol{m}_Z)}\\ \mathrm{E}\left[\tilde{\boldsymbol{X}}_{\mathrm{MV}}\tilde{\boldsymbol{X}}_{\mathrm{MV}}^{\mathrm{T}}\right]=\int_{-\infty}^{\infty}\boldsymbol{C}_{X|Z}p_Z(\boldsymbol{z})\mathrm{d}\boldsymbol{z}=\int_{-\infty}^{\infty}(\boldsymbol{C}_X-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{ZX})p_Z(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\\ \qquad=(\boldsymbol{C}_X-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{ZX})\int_{-\infty}^{\infty}p_Z(\boldsymbol{z})\mathrm{d}\boldsymbol{z}=\boldsymbol{C}_X-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{ZX}\end{cases}$$

当状态与观测之间服从联合正态分布时，只需使用一、二阶矩参数就能很容易地求得最小方差估计的表达式！

18.3 最小方差估计和线性最小方差估计

（2）线性最小方差估计（Linear Minimum Variance，LMV）

不论理论观测模型是线性的还是非线性的，它都采用观测量的线性组合建模来对状态进行估计

$$\hat{\boldsymbol{X}}=\boldsymbol{A}\boldsymbol{Z}+\boldsymbol{b}$$

$\boldsymbol{b}$ → n维待定常值向量

$\boldsymbol{A}$ → $n\mathrm{x}m$维待定常值矩阵

性能指标函数

$$J(\hat{\boldsymbol{X}})=\mathrm{E}\left[\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{X}}\right]=\mathrm{tr}\left(\mathrm{E}\left[\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\right]\right)$$

$$=\mathrm{tr}\left(\mathrm{E}\left[(\boldsymbol{X}-\boldsymbol{A}\boldsymbol{Z}-\boldsymbol{b})(\boldsymbol{X}-\boldsymbol{A}\boldsymbol{Z}-\boldsymbol{b})^{\mathrm{T}}\right]\right)$$

$$=\mathrm{tr}\left(\mathrm{E}\left[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\boldsymbol{A}\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{X}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}-\boldsymbol{A}\boldsymbol{Z}\boldsymbol{X}^{\mathrm{T}}-\boldsymbol{X}\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{b}\boldsymbol{X}^{\mathrm{T}}+\boldsymbol{b}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\boldsymbol{Z}\boldsymbol{b}^{\mathrm{T}}\right]\right)$$

$$=\mathrm{tr}\left((\boldsymbol{C}_X+\boldsymbol{m}_X\boldsymbol{m}_X^{\mathrm{T}})+\boldsymbol{A}(\boldsymbol{C}_Z+\boldsymbol{m}_Z\boldsymbol{m}_Z^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}-(\boldsymbol{C}_{XZ}+\boldsymbol{m}_X\boldsymbol{m}_Z^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}-\boldsymbol{A}(\boldsymbol{C}_{ZX}+\boldsymbol{m}_Z\boldsymbol{m}_X^{\mathrm{T}})\right.$$

$$\left.-\boldsymbol{m}_X\boldsymbol{b}^{\mathrm{T}}-\boldsymbol{b}\boldsymbol{m}_X^{\mathrm{T}}+\boldsymbol{b}\boldsymbol{m}_Z^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\boldsymbol{m}_Z\boldsymbol{b}^{\mathrm{T}}\right)$$

$$=\mathrm{tr}\left((\boldsymbol{A}-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1})\boldsymbol{C}_Z(\boldsymbol{A}-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1})^{\mathrm{T}}+(\boldsymbol{C}_X-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{ZX})+(\boldsymbol{m}_X-\boldsymbol{A}\boldsymbol{m}_Z-\boldsymbol{b})(\boldsymbol{m}_X-\boldsymbol{A}\boldsymbol{m}_Z-\boldsymbol{b})^{\mathrm{T}}\right)$$

$$=\mathrm{tr}\left((\boldsymbol{A}-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1})\boldsymbol{C}_Z(\boldsymbol{A}-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1})^{\mathrm{T}}\right)+\mathrm{tr}\left(\boldsymbol{C}_X-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{ZX}\right)$$

$$+\mathrm{tr}\left((\boldsymbol{m}_X-\boldsymbol{A}\boldsymbol{m}_Z-\boldsymbol{b})(\boldsymbol{m}_X-\boldsymbol{A}\boldsymbol{m}_Z-\boldsymbol{b})^{\mathrm{T}}\right)$$

18.3 最小方差估计和线性最小方差估计

非负

$$J(\hat{X}) = \mathrm{tr}\left((A - C_{XZ}C_Z^{-1})C_Z(A - C_{XZ}C_Z^{-1})^{\mathrm{T}}\right) + \mathrm{tr}\left(C_X - C_{XZ}C_Z^{-1}C_{ZX}\right)$$

→ 与 A, b 无关

$$+\,\mathrm{tr}\left((m_X - Am_Z - b)(m_X - Am_Z - b)^{\mathrm{T}}\right) = \min$$

$$\Longrightarrow \begin{cases} A - C_{XZ}C_Z^{-1} = 0 \\ m_X - Am_Z - b = 0 \end{cases} \Longrightarrow \begin{cases} A = C_{XZ}C_Z^{-1} \\ b = m_X - C_{XZ}C_Z^{-1}m_Z \end{cases}$$

$$\Longrightarrow \hat{X}_{\mathrm{LMV}} = AZ + b = m_X + C_{XZ}C_Z^{-1}(Z - m_Z)$$

→与正态分布下一致

线性最小方差估计只需已知一、二阶矩，不需已知联合密度函数。

→最小方差估计没法用，根据密度分布函数找不到
↓
正态分布下线性最小方差估计就能用了

线性最小方差估计的无偏性

$$\begin{aligned} \mathrm{E}_Z\left[\hat{X}_{\mathrm{LMV}}\right] &= \mathrm{E}_Z\left[m_X + C_{XZ}C_Z^{-1}(Z - m_Z)\right] \\ &= m_X + C_{XZ}C_Z^{-1}(\mathrm{E}_Z[Z] - m_Z) \\ &= m_X + C_{XZ}C_Z^{-1}(m_Z - m_Z) = m_X = \mathrm{E}[X] \end{aligned}$$

16

18.3 最小方差估计和线性最小方差估计

线性最小方差估计的均方差阵

$$\begin{aligned} \mathrm{E}\left[\tilde{X}_{\mathrm{LMV}}\tilde{X}_{\mathrm{LMV}}^{\mathrm{T}}\right] &= \mathrm{E}\left[\left\{X - \left[m_X + C_{XZ}C_Z^{-1}(Z - m_Z)\right]\right\}\left\{X - \left[m_X + C_{XZ}C_Z^{-1}(Z - m_Z)\right]\right\}^{\mathrm{T}}\right] \\ &= \mathrm{E}\left[\left[(X - m_X) - C_{XZ}C_Z^{-1}(Z - m_Z)\right]\left[(X - m_X) - C_{XZ}C_Z^{-1}(Z - m_Z)\right]^{\mathrm{T}}\right] \\ &= \mathrm{E}\left[(X - m_X)(X - m_X)^{\mathrm{T}}\right] - \mathrm{E}\left[(X - m_X)\left[C_{XZ}C_Z^{-1}(Z - m_Z)\right]^{\mathrm{T}}\right] \\ &\quad - \mathrm{E}\left[C_{XZ}C_Z^{-1}(Z - m_Z)(X - m_X)^{\mathrm{T}}\right] + \mathrm{E}\left[C_{XZ}C_Z^{-1}(Z - m_Z)\left[C_{XZ}C_Z^{-1}(Z - m_Z)\right]^{\mathrm{T}}\right] \\ &= C_X - C_{XZ}C_Z^{-1}C_{ZX} - C_{XZ}C_Z^{-1}C_{ZX} + C_{XZ}C_Z^{-1}C_ZC_Z^{-1}C_{ZX} = C_X - C_{XZ}C_Z^{-1}C_{ZX} \end{aligned}$$

对于正态分布而言，最小方差估计等价于线性最小方差估计，究其原因在于两正态分布变量之间的关系总可以用线性关系来描述。但是，一般在任意概率分布情况下，线性最小方差估计的精度往往不如最小方差估计。

→正态分布 $\xrightarrow{\text{线性拟合}}$ 正态分布

$\xrightarrow{\text{非线性拟合}}$ 不一定.

17

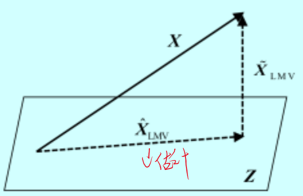
↳总结：要用肯定假设用从正态分布，然后是最小方差估计。

18.3 最小方差估计和线性最小方差估计

协相关阵

$$\begin{aligned} \mathrm{Cov}\left(\tilde{X}_{\mathrm{LMV}}, Z\right) &= \mathrm{E}\left[\left\{X - \left[m_X + C_{XZ}C_Z^{-1}(Z - m_Z)\right]\right\}(Z - m_Z)^{\mathrm{T}}\right] \\ &= \mathrm{E}\left[(X - m_X)(Z - m_Z)^{\mathrm{T}}\right] - C_{XZ}C_Z^{-1}\mathrm{E}\left[(Z - m_Z)(Z - m_Z)^{\mathrm{T}}\right] \\ &= C_{XZ} - C_{XZ}C_Z^{-1}C_Z = 0 \end{aligned}$$

不相关从几何角度看，称为正交，估计值是被估计量在观测空间上的正交投影。



正交性原理

估计肯定是已知空间内的线性组合
量测

正交投影示意图

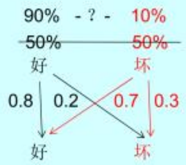
18

18.4 极大似然估计(Maximum Likelihood, ML)

→概率论

为了估计 $\boldsymbol{X}$，对 $\boldsymbol{Z}$ 进行观测，如果观测值为 $\boldsymbol{z}$ ，则出现该值的概率密度为 $L(\boldsymbol{x})=p(\boldsymbol{z}\mid\boldsymbol{x})$ 。

将使 $L(\boldsymbol{x})=p(\boldsymbol{z}\mid\boldsymbol{x})$ 取得最大值时的 $\boldsymbol{x}$ 作为 $\boldsymbol{X}$ 的最优估计 $\hat{\boldsymbol{X}}$ ，这时 $\hat{\boldsymbol{X}}$ 是准确值的可能性最大，这便是极大似然估计准则(导致此观测结果的最可能原因是什么，认为各原因是等概率发生的/等先验概率)。



似然函数 $L(\boldsymbol{x})|_{x=\hat{X}_{\mathrm{ML}}}=\max$

⇨ 似然方程 $\dfrac{\partial L(\boldsymbol{x})}{\partial \boldsymbol{x}}\Big|_{x=\hat{X}_{\mathrm{ML}}}=\boldsymbol{0}$

⇨ 对数似然方程 $\dfrac{\partial \ln L(\boldsymbol{x})}{\partial \boldsymbol{x}}\Big|_{x=\hat{X}_{\mathrm{ML}}}=\boldsymbol{0}$

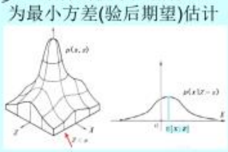
18.5 极大验后估计(Maximum A Posteriori, MAP)

极大验后估计准则：给定某一观测值 $\boldsymbol{Z}=\boldsymbol{z}$ ，使条件密度函数 $p(\boldsymbol{x}\mid\boldsymbol{z})$ 达到极大的那个 $\boldsymbol{x}$ 值，就是最可能的估计值，即

验后函数 $p(\boldsymbol{x}\mid\boldsymbol{z})|_{x=\hat{X}_{\mathrm{MAP}}}=\max$ ←→ 条件均值最大指标为最小方差(验后期望)估计

⇨ 验后方程 $\dfrac{\partial p(\boldsymbol{x}\mid\boldsymbol{z})}{\partial \boldsymbol{x}}\Big|_{x=\hat{X}_{\mathrm{MAP}}}=\boldsymbol{0}$

⇨ 对数验后方程 $\dfrac{\partial \ln p(\boldsymbol{x}\mid\boldsymbol{z})}{\partial \boldsymbol{x}}\Big|_{x=\hat{X}_{\mathrm{MAP}}}=\boldsymbol{0}$



在贝叶斯公式 $p(\boldsymbol{x}\mid\boldsymbol{z})=\dfrac{p(\boldsymbol{z}\mid\boldsymbol{x})p_X(\boldsymbol{x})}{p_Z(\boldsymbol{z})}$ 中：

右端分子中边缘密度函数 $p_X(\boldsymbol{x})$ 已知，它表示在未作观测之前就已经知道了状态 $\boldsymbol{X}$ 的概率密度函数，所以又可称 $p_X(\boldsymbol{x})$ 为***验前概率密度函数***；

左端条件概率密度函数 $p(\boldsymbol{x}\mid\boldsymbol{z})$ 意为作出观测 $\boldsymbol{Z}=\boldsymbol{z}$ 之后的状态 $\boldsymbol{X}$ 的概率密度函数，因而 $p(\boldsymbol{x}\mid\boldsymbol{z})$ 通常被称为***验后概率密度函数***。

18.5 极大验后估计

→对z没有量测，也不知道x的分布

在缺乏状态任何先验知识的情况下，极大验后估计等价于极大似然估计。

验证如下：

$$p(\boldsymbol{x}\mid\boldsymbol{z})=\frac{p(\boldsymbol{z}\mid\boldsymbol{x})p_X(\boldsymbol{x})}{p_Z(\boldsymbol{z})} \Longrightarrow \frac{\partial \ln p(\boldsymbol{x}\mid\boldsymbol{z})}{\partial \boldsymbol{x}}=\frac{\partial \ln p(\boldsymbol{z}\mid\boldsymbol{x})}{\partial \boldsymbol{x}}+\frac{\partial \ln p_X(\boldsymbol{x})}{\partial \boldsymbol{x}}$$

$$p_X(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}|\boldsymbol{C}_X|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_X)^{\mathrm{T}}\boldsymbol{C}_X^{-1}(\boldsymbol{x}-\boldsymbol{m}_X)\right\}$$

→假设正态分布

$$\frac{\partial \ln p_X(\boldsymbol{x})}{\partial \boldsymbol{x}}=\frac{\partial}{\partial \boldsymbol{x}}\left\{-\ln\left[(2\pi)^{n/2}|\boldsymbol{C}_X|^{1/2}\right]-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_X)^{\mathrm{T}}\boldsymbol{C}_X^{-1}(\boldsymbol{x}-\boldsymbol{m}_X)\right\}=-\boldsymbol{C}_X^{-1}(\boldsymbol{x}-\boldsymbol{m}_X)\to\boldsymbol{0}$$

$$\frac{\partial \ln p(\boldsymbol{x}\mid\boldsymbol{z})}{\partial \boldsymbol{x}}=\frac{\partial \ln p(\boldsymbol{z}\mid\boldsymbol{x})}{\partial \boldsymbol{x}}$$

18.6 加权最小二乘估计(Weighted Least Square, WLS)

观测$\boldsymbol{Z}$与状态$\boldsymbol{X}$之间具有线性函数关系

$$\boldsymbol{Z}=\boldsymbol{HX}+\boldsymbol{V} \qquad \mathrm{E}\left[\boldsymbol{V}\right]=\boldsymbol{0} \quad \mathrm{E}\left[\boldsymbol{VV}^{\mathrm{T}}\right]=\boldsymbol{C}_V$$ 量测的预测

指标函数是使**观测**与**由状态估计确定的观测估计**之间的加权误差平方和达到最小

m维正定对称加权阵

$$J(\hat{\boldsymbol{X}})=(\boldsymbol{Z}-\boldsymbol{H}\hat{\boldsymbol{X}})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{Z}-\boldsymbol{H}\hat{\boldsymbol{X}})\Big|_{\hat{\boldsymbol{X}}=\hat{\boldsymbol{X}}_{\mathrm{WLS}}}=\min$$

求导

$$\frac{\partial J(\hat{\boldsymbol{X}})}{\partial\hat{\boldsymbol{X}}}=\frac{\partial}{\partial\hat{\boldsymbol{X}}}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{WZ}-\hat{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WZ}-\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{WH}\hat{\boldsymbol{X}}+\hat{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH}\hat{\boldsymbol{X}})$$

$$=\boldsymbol{0}-\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WZ}-\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WZ}+2\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH}\hat{\boldsymbol{X}}$$

$$=2(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH}\hat{\boldsymbol{X}}-\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WZ})=\boldsymbol{0}$$

得估计 $\hat{\boldsymbol{X}}_{\mathrm{WLS}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WZ}$ 　　若 $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH}$ 可逆

无偏性 $\mathrm{E}\left[\hat{\boldsymbol{X}}_{\mathrm{WLS}}\right]=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}\mathrm{E}\left[\boldsymbol{Z}\right]=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}\mathrm{E}\left[\boldsymbol{HX}+\boldsymbol{V}\right]$

$$=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}\left(\boldsymbol{H}\mathrm{E}\left[\boldsymbol{X}\right]+\mathrm{E}\left[\boldsymbol{V}\right]\right)=\mathrm{E}\left[\boldsymbol{X}\right]$$

22

18.6 加权最小二乘估计

$$\hat{\boldsymbol{X}}_{\mathrm{WLS}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WZ}$$

估计误差 $\tilde{\boldsymbol{X}}_{\mathrm{WLS}}=\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{WLS}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})\boldsymbol{X}-(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WZ}$

$$=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{HX}-\boldsymbol{Z})=-(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WV}$$

均方误差阵 $\mathrm{E}\left[\tilde{\boldsymbol{X}}_{\mathrm{WLS}}\tilde{\boldsymbol{X}}_{\mathrm{WLS}}^{\mathrm{T}}\right]=\mathrm{E}\left[(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WV}\left[(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WV}\right]^{\mathrm{T}}\right]$

$$=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}\mathrm{E}\left[\boldsymbol{VV}^{\mathrm{T}}\right]\left[(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}\right]^{\mathrm{T}}$$

$$=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WC}_V\boldsymbol{WH}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}$$ W取何值，其取最小

$$\geq(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_V^{-1}\boldsymbol{H})^{-1}$$

等号成立的条件是 $\boldsymbol{W}=\boldsymbol{C}_V^{-1}\sigma^2$（$\sigma^2$为任意正常数），这时加权最小二乘估计的均方误差阵最小（即精度最高），称为最优加权最小二乘估计（或称马尔可夫估计）。

普通最小二乘估计 $\hat{\boldsymbol{X}}_{\mathrm{LS}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{Z}$ 　　$\boldsymbol{W}=\boldsymbol{I}$

$$\mathrm{E}[\tilde{\boldsymbol{X}}_{\mathrm{LS}}\tilde{\boldsymbol{X}}_{\mathrm{LS}}^{\mathrm{T}}]=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_V\boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}$$

23

18.6 加权最小二乘估计

证明 $\mathrm{E}\left[\tilde{\boldsymbol{X}}_{\mathrm{WLS}}\tilde{\boldsymbol{X}}_{\mathrm{WLS}}^{\mathrm{T}}\right]=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WC}_V\boldsymbol{WH}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\geq(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_V^{-1}\boldsymbol{H})^{-1}$

记 $\begin{cases}\boldsymbol{C}_V=\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S}\\ \boldsymbol{A}=\boldsymbol{H}^{\mathrm{T}}\boldsymbol{S}^{-1}\\ \boldsymbol{B}=\boldsymbol{SWH}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\end{cases}$ 　　$\boldsymbol{AB}=\boldsymbol{I}$

由许瓦茨不等式

$$\left[\boldsymbol{B}-\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{AA}^{\mathrm{T}})^{-1}\boldsymbol{AB}\right]^{\mathrm{T}}\left[\boldsymbol{B}-\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{AA}^{\mathrm{T}})^{-1}\boldsymbol{AB}\right]$$

$$=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}-2\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{AA}^{\mathrm{T}})^{-1}\boldsymbol{AB}+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{AA}^{\mathrm{T}})^{-1}\boldsymbol{AA}^{\mathrm{T}}(\boldsymbol{AA}^{\mathrm{T}})^{-1}\boldsymbol{AB}$$

$$=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}-(\boldsymbol{AB})^{\mathrm{T}}(\boldsymbol{AA}^{\mathrm{T}})^{-1}\boldsymbol{AB}\geq\boldsymbol{0}$$

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\geq(\boldsymbol{AB})^{\mathrm{T}}(\boldsymbol{AA}^{\mathrm{T}})^{-1}\boldsymbol{AB}=(\boldsymbol{AA}^{\mathrm{T}})^{-1}$$

则有

$$\mathrm{E}\left[\tilde{\boldsymbol{X}}_{\mathrm{WLS}}\tilde{\boldsymbol{X}}_{\mathrm{WLS}}^{\mathrm{T}}\right]=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WS}^{\mathrm{T}}\boldsymbol{SWH}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$$

$$\geq(\boldsymbol{AA}^{\mathrm{T}})^{-1}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_V^{-1}\boldsymbol{H})^{-1}$$

24

18.6 加权最小二乘估计

【举例】（**线性正态**观测下的各种估计方法）

$$Z = HX + V \qquad X \sim N(m_X, C_X) \quad V \sim N(0, C_V) \qquad \text{Cov}(X,V)=0$$

先验

预先计算

$$m_Z = E[Z] = E[HX+V] = Hm_X$$

一）一阶矩

$$C_Z = \text{Cov}(Z,Z) = \text{Cov}(HX+V, HX+V)$$
$$= \text{Cov}(HX,HX) + \text{Cov}(HX,V) + \text{Cov}(V,HX) + \text{Cov}(V,V)$$
$$= HC_XH^{\mathrm{T}} + HC_{XV} + C_{VX}H^{\mathrm{T}} + C_V = HC_XH^{\mathrm{T}} + C_V$$

$$C_{XZ} = \text{Cov}(X,Z) = \text{Cov}(X, HX+V) = C_XH^{\mathrm{T}}$$

$$\Longrightarrow \begin{bmatrix} X \\ Z \end{bmatrix} \sim N\left(\begin{bmatrix} m_X \\ m_Z \end{bmatrix}, \begin{bmatrix} C_X & C_{XZ} \\ C_{ZX} & C_Z \end{bmatrix}\right) \quad \Leftrightarrow p(x,z) \ \checkmark$$

联合正态分布

25

18.6 加权最小二乘估计

$$Z = HX + V$$

1）线性最小方差估计

矩阵求逆原理

$$C_XH^{\mathrm{T}}(HC_XH^{\mathrm{T}}+C_V)^{-1} = (C_X^{-1} + H^{\mathrm{T}}C_V^{-1}H)^{-1}H^{\mathrm{T}}C_V^{-1}$$

$$\hat{X}_{\text{LMV}} = m_X + C_{XZ}C_Z^{-1}(Z - m_Z)$$
$$= m_X + C_XH^{\mathrm{T}}(HC_XH^{\mathrm{T}} + C_V)^{-1}(Z - Hm_X)$$
$$= (C_X^{-1} + H^{\mathrm{T}}C_V^{-1}H)^{-1}\left[(C_X^{-1} + H^{\mathrm{T}}C_V^{-1}H)m_X + H^{\mathrm{T}}C_V^{-1}(Z - Hm_X)\right]$$
$$= (C_X^{-1} + H^{\mathrm{T}}C_V^{-1}H)^{-1}(H^{\mathrm{T}}C_V^{-1}Z + C_X^{-1}m_X)$$

$$(C_X^{-1} + H^{\mathrm{T}}C_V^{-1}H)^{-1} = C_X - C_XH^{\mathrm{T}}(HC_XH^{\mathrm{T}} + C_V)^{-1}HC_X$$

$$E\left[\tilde{X}_{\text{LMV}}\tilde{X}_{\text{LMV}}^{\mathrm{T}}\right] = C_X - C_{XZ}C_Z^{-1}C_{ZX}$$
$$= C_X - C_XH^{\mathrm{T}}(HC_XH^{\mathrm{T}} + C_V)^{-1}HC_X$$
$$= (C_X^{-1} + H^{\mathrm{T}}C_V^{-1}H)^{-1}$$

相当于 P_k

P_{k-1}

26

18.6 加权最小二乘估计

$$Z = HX + V \qquad \begin{bmatrix} X \\ Z \end{bmatrix} \sim N\left(\begin{bmatrix} m_X \\ m_Z \end{bmatrix}, \begin{bmatrix} C_X & C_{XZ} \\ C_{ZX} & C_Z \end{bmatrix}\right)$$

2）极大似然估计

$$p(z\,|\,x) = \frac{1}{(2\pi)^{m/2}|C_{Z|X}|^{1/2}}\exp\left\{-\frac{1}{2}(z - m_{Z|X})^{\mathrm{T}}C_{Z|X}^{-1}(z - m_{Z|X})\right\}$$
$$= \frac{1}{(2\pi)^{m/2}|C_V|^{1/2}}\exp\left\{-\frac{1}{2}(z - Hx)^{\mathrm{T}}C_V^{-1}(z - Hx)\right\}$$

其中

$$\begin{cases} m_{Z|X} = m_Z + C_{ZX}C_X^{-1}(x - m_X) = Hm_X + HC_XC_X^{-1}(x - m_X) = Hx \\ C_{Z|X} = C_Z - C_{ZX}C_X^{-1}C_{XZ} = (HC_XH^{\mathrm{T}} + C_V) - HC_XC_X^{-1}C_XH^{\mathrm{T}} = C_V \end{cases}$$

对数似然方程

$$\frac{\partial \ln p(z\,|\,x)}{\partial x} = \frac{\partial}{\partial x}\left\{-\ln\left[(2\pi)^{m/2}|C_V|^{1/2}\right] - \frac{1}{2}(z - Hx)^{\mathrm{T}}C_V^{-1}(z - Hx)\right\}$$
$$= H^{\mathrm{T}}C_V^{-1}(z - Hx) = 0$$

求得 $\hat{X}_{\text{ML}} = (H^{\mathrm{T}}C_V^{-1}H)^{-1}H^{\mathrm{T}}C_V^{-1}Z$ 均方差阵同最小二乘估计 27

就是最小二乘估计

18.6 加权最小二乘估计

$$Z = HX + V \qquad \begin{bmatrix} X \\ Z \end{bmatrix} \sim N\left(\begin{bmatrix} m_X \\ m_Z \end{bmatrix}, \begin{bmatrix} C_X & C_{XZ} \\ C_{ZX} & C_Z \end{bmatrix}\right)$$

3）极大验后估计

$$p(x\,|\,z) = \frac{1}{(2\pi)^{n/2}|C_{X|Z}|^{1/2}}\exp\left\{-\frac{1}{2}(x - m_{X|Z})^{\mathrm{T}}C_{X|Z}^{-1}(x - m_{X|Z})\right\}$$

其中

$$\begin{cases} m_{X|Z} = m_X + C_{XZ}C_Z^{-1}(z - m_Z) = m_X + C_XH^{\mathrm{T}}(HC_XH^{\mathrm{T}} + C_V)^{-1}(z - m_Z) \\ C_{X|Z} = C_X - C_{XZ}C_Z^{-1}C_{ZX} = C_X - C_XH^{\mathrm{T}}(HC_XH^{\mathrm{T}} + C_V)^{-1}C_{ZX}HC_X \end{cases}$$

对数似然方程

18.6 加权最小二乘估计

$$Z = HX + V$$

$$\begin{bmatrix} X \\ Z \end{bmatrix} \sim N\left(\begin{bmatrix} m_X \\ m_Z \end{bmatrix}, \begin{bmatrix} C_X & C_{XZ} \\ C_{ZX} & C_Z \end{bmatrix}\right)$$

3）极大验后估计

$$p(x \mid z) = \frac{1}{(2\pi)^{n/2}\left|C_{X|Z}\right|^{1/2}} \exp\left\{-\frac{1}{2}(x - m_{X|Z})^{\mathrm{T}} C_{X|Z}^{-1}(x - m_{X|Z})\right\}$$

其中

$$\begin{cases} m_{X|Z} = m_X + C_{XZ}C_Z^{-1}(z - m_Z) = m_X + C_X H^{\mathrm{T}}(HC_X H^{\mathrm{T}} + C_V)^{-1}(z - m_Z) \\ C_{X|Z} = C_X - C_{XZ}C_Z^{-1}C_{ZX} = C_X - C_X H^{\mathrm{T}}(HC_X H^{\mathrm{T}} + C_V)^{-1} C_{ZX} H C_X \end{cases}$$

对数似然方程

$$\frac{\partial \ln p(x \mid z)}{\partial x} = \frac{\partial}{\partial x}\left\{-\ln\left[(2\pi)^{n/2}\left|C_{X|Z}\right|^{1/2}\right] - \frac{1}{2}(x - m_{X|Z})^{\mathrm{T}} C_{X|Z}^{-1}(x - m_{X|Z})\right\}$$

$$= -C_{X|Z}^{-1}(x - m_{X|Z}) = 0$$

求得 $\hat{X}_{\mathrm{MAP}} = \underbrace{m_{X|Z}}_{\text{条件均值}} = m_X + C_X H^{\mathrm{T}}(HC_X H^{\mathrm{T}} + C_V)^{-1}(Z - Hm_X)$ →就是最小方差估计

均方差阵同最小方差估计

28

18.6 加权最小二乘估计

4）加权最小二乘估计

一样

$$\hat{X}_{\mathrm{WLS}} = (H^{\mathrm{T}}C_V^{-1}H)H^{\mathrm{T}}C_V^{-1}Z$$

$$\mathrm{E}\left[\tilde{X}_{\mathrm{WLS}}\tilde{X}_{\mathrm{WLS}}^{\mathrm{T}}\right] = (H^{\mathrm{T}}C_V^{-1}H)^{-1}$$

总结对比：极大似然估计

$$\hat{X}_{\mathrm{ML}} = (H^{\mathrm{T}}C_V^{-1}H)^{-1}H^{\mathrm{T}}C_V^{-1}Z$$

$$\mathrm{RLS}: \begin{cases} P_k^{-1}\hat{X}_k = P_{k-1}^{-1}\hat{X}_{k-1} + H_k^{\mathrm{T}}R_k^{-1}Z_k \\ P_k^{-1} = P_{k-1}^{-1} + H_k^{\mathrm{T}}R_k^{-1}H_k \end{cases}$$

→线性系统的Kalman，就是最小方差/极大后验估计

一样

线性最小方差估计

$$\hat{X}_{\mathrm{LMV}} = (C_X^{-1} + H^{\mathrm{T}}C_V^{-1}H)^{-1}(H^{\mathrm{T}}C_V^{-1}Z + C_X^{-1}m_X)$$

$$\mathrm{E}\left[\tilde{X}_{\mathrm{LMV}}\tilde{X}_{\mathrm{LMV}}^{\mathrm{T}}\right] = (C_X^{-1} + H^{\mathrm{T}}C_V^{-1}H)^{-1}$$

极大验后估计

$$\hat{X}_{\mathrm{MAP}} = (C_X^{-1} + H^{\mathrm{T}}C_V^{-1}H)^{-1}(H^{\mathrm{T}}C_V^{-1}Z + C_X^{-1}m_X)$$

残差小，估计精度更高，因为有先验。

29

18.7 贝叶斯估计 →只是一个框架

定义损失函数（Lost function，或称代价函数）为

$$L(\tilde{X}) = L\left(X - \hat{X}(Z)\right)$$

损失函数是标量函数，须满足如下3个条件： →误差越大，损失越大

（1）当$\|\tilde{X}_1\| \geq \|\tilde{X}_2\|$时，有$L(\tilde{X}_1) \geq L(\tilde{X}_2) \geq 0$；

（2）当$\|\tilde{X}\| = 0$时，有 $L(\tilde{X}) = 0$；

（3）$L(\tilde{X}) = L(-\tilde{X})$ 。→对称

即非负性、对称性、估计误差为零时取最小 →机器学习

损失函数是观测样本的函数，当样本取值不同时，带来的损失一般也不相同。为了从整体上评价损失函数的性能，定义平均损失函数，即贝叶斯风险（Bayesian Risk），为

$$R(\hat{X}) = \mathrm{E}\left[L(\tilde{X})\right] = \mathrm{E}\left[L\left(X - \hat{X}(Z)\right)\right]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} L\left(x - \hat{X}(z)\right) p(x, z)\,\mathrm{d}x\,\mathrm{d}z$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} L\left(x - \hat{X}(z)\right) p(x \mid z)\,\mathrm{d}x\right] p_Z(z)\,\mathrm{d}z$$

30

18.7 贝叶斯估计

使贝叶斯风险达到最小的估计称为贝叶斯估计 >=0

$$R(\hat{X})\ \big|_{X=\hat{X}_B}=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}L\big(x-\hat{X}(z)\big)p(x\,|\,z)\,\mathrm{d}\,x\right]p_Z(z)\,\mathrm{d}\,z=\min$$

$$\Longleftrightarrow\quad R(\hat{X}\,|\,z)=\int_{-\infty}^{\infty}L\big(x-\hat{X}(z)\big)p(x\,|\,z)\,\mathrm{d}\,x\ \big|_{X=\hat{X}_B}=\min$$

在实际应用贝叶斯估计时，如何选取合适的损失函数 L 很重要，多数时候总可以选择平方损失函数，比如 → 和机器学习一致

（1）最小方差估计实际上就是贝叶斯估计的一个特例，即有

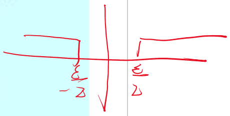
$$L(\tilde{X})=\tilde{X}^{\mathrm{T}}\tilde{X}$$

31

18.7 贝叶斯估计

（2）极大验后估计也是贝叶斯估计的一个特例，验证如下

定义均匀损失函数 $L(\tilde{X})=\begin{cases}0, & \|X-\hat{X}\|<\varepsilon/2\\ 1/\varepsilon, & \|X-\hat{X}\|\geq\varepsilon/2\end{cases}\qquad \varepsilon>0$



贝叶斯风险 $R(\hat{X})=\mathrm{E}\left[L\big(X-\hat{X}(Z)\big)\right]=\int_{-\infty}^{\infty}\left[\int_{\|X-\hat{X}\|\geq\varepsilon/2}\frac{1}{\varepsilon}p(x\,|\,z)\,\mathrm{d}\,x\right]p_Z(z)\,\mathrm{d}\,z$

$$=\int_{-\infty}^{\infty}\frac{1}{\varepsilon}\left[1-\int_{\|X-\hat{X}\|<\varepsilon/2}p(x\,|\,z)\,\mathrm{d}\,x\right]p_Z(z)\,\mathrm{d}\,z=\min$$

$$\int_{\|X-\hat{X}\|<\varepsilon/2}p(x\,|\,z)\,\mathrm{d}\,x=\max\xrightarrow{\varepsilon\to 0}p(x\,|\,z)=\max$$ 为极大验后估计

（3）最小二乘法能否当作贝叶斯估计的一个特例？

32

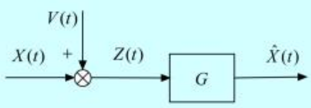
18.8 维纳滤波 → 了解，知道其概念即可（静态→动态）

前述讨论的几种估计问题可视为系统状态在某一观测时刻的“静态估计”，而维纳滤波研究的是从动态系统的长时间观测中求解状态变量最优估计的“动态估计”问题。 → 类似随机过程与概率论的区别

模型：一维连续时间系统的观测方程

$$Z(t)=X(t)+V(t)$$

观测信号　有用信号　噪声



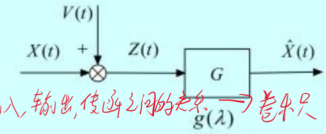
怎么设计G是维纳滤波的目的

根据观测去估计状态，结果记为 $\hat{X}(t)$，使得估计值尽量接近于真实状态，用统计的术语表示就是使估计的均方误差最小

$$J(\hat{X}(t))=\mathrm{E}\left[\tilde{X}^2(t)\right]=\mathrm{E}\left[\left[X(t)-\hat{X}(t)\right]^2\right]=\min$$

33

18.8 维纳滤波



状态的滤波估计

$$\hat{X}(t)=\int_0^{\infty} g(\lambda)Z(t-\lambda)\mathrm{d}\lambda$$ ←输入,输出,传函之间的关系 → 卷积

估计的无偏性

$$\mathrm{E}\left[\tilde{X}(t)\right]=\mathrm{E}\left[X(t)-\hat{X}(t)\right]=\mathrm{E}\left[X(t)\right]-\mathrm{E}\left[\int_0^{\infty} g(\lambda)Z(t-\lambda)\mathrm{d}\lambda\right]$$

$$=\mathrm{E}\left[X(t)\right]-\int_0^{\infty} g(\lambda)\mathrm{E}\left[Z(t-\lambda)\right]\mathrm{d}\lambda=0$$

估计误差与观测正交（借用正交性原理）

$$\mathrm{E}\left[\tilde{X}(t)Z(t-\tau)\right]=\mathrm{E}\left[\left[X(t)-\int_0^{\infty} g(\lambda)Z(t-\lambda)\mathrm{d}\lambda\right]Z(t-\tau)\right]$$

$$=\mathrm{E}\left[X(t)Z(t-\tau)\right]-\mathrm{E}\left[\int_0^{\infty} g(\lambda)Z(t-\lambda)\mathrm{d}\lambda Z(t-\tau)\right]$$

$$=\mathrm{E}\left[X(t)Z(t-\tau)\right]-\int_0^{\infty} g(\lambda)\mathrm{E}\left[Z(t-\lambda)Z(t-\tau)\right]\mathrm{d}\lambda$$

$$=\underbrace{R_{XZ}(\tau)}_{\text{互相关函数}}-\int_0^{\infty} g(\lambda)\underbrace{R_Z}_{\text{自相关函数}}(\tau-\lambda)\mathrm{d}\lambda=0$$

18.8 维纳滤波

都已知

$$\mathrm{E}\left[\tilde{X}(t)Z(t-\tau)\right]=R_{XZ}(\tau)-\int_0^{\infty} g(\lambda)R_Z(\tau-\lambda)\mathrm{d}\lambda=0$$

$$\Longrightarrow\quad R_{XZ}(\tau)=\int_0^{\infty} g(\lambda)R_Z(\tau-\lambda)\mathrm{d}\lambda$$

上式称为维纳—霍夫（Wiener—Hopf）积分方程。理论上通过求解该积分方程即可获得滤波器的脉冲响应函数 $g(t)$，完成维纳滤波器设计。但这一过程通常相当的繁琐和复杂。

维纳滤波是20世纪40年代在随机滤波理论上的一个重大突破，在方法论上具有比较深远的影响，但是在实际使用中却受到很大限制：

- 维纳—霍夫方程一般很难求解；
- 传递函数在工程上往往很难实现；
- 仅适用于处理单输出平稳随机过程，难以应用于复杂的高维随机系统。

18.9 估计量的性质 → 概统结论

（1）无偏性(unbiasness)

$$\mathrm{E}\left[\hat{X}\right]=\mathrm{E}\left[X\right]\qquad\Leftrightarrow\mathrm{E}\left[\tilde{X}\right]=\mathrm{E}\left[X-\hat{X}\right]=0$$

（2）有效性(effectiveness)

两种方法获得的无偏估计 $\hat{X}_{(1)}$ 和 $\hat{X}_{(2)}$，如果 $\mathrm{E}\left[\tilde{X}_{(1)}^2\right]<\mathrm{E}\left[\tilde{X}_{(2)}^2\right]$ 则认为前者比后者更有效。那么是否存在最小方差极限呢？

$$\mathrm{E}\left[\tilde{X}\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\hat{X}(z))p(x,z)\,\mathrm{d}x\,\mathrm{d}z=0$$

$$\frac{\partial\mathrm{E}\left[\tilde{X}\right]}{\partial x}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\partial}{\partial x}\left[(x-\hat{X}(z))p(x,z)\right]\mathrm{d}x\,\mathrm{d}z$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,z)\,\mathrm{d}x\,\mathrm{d}z+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\hat{X}(z))\frac{\partial}{\partial x}p(x,z)\,\mathrm{d}x\,\mathrm{d}z$$

$$=1+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\hat{X}(z))\frac{\partial}{\partial x}p(x,z)\,\mathrm{d}x\,\mathrm{d}z=0$$

18.9 估计量的性质

$$\frac{\partial \ln f(x)}{\partial x}=\frac{1}{f(x)}\frac{\partial f(x)}{\partial x} \Rightarrow \frac{\partial f(x)}{\partial x}=\frac{\partial \ln f(x)}{\partial x} f(x)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\hat{X}(z)-x)\frac{\partial}{\partial x}p(x,z)\,\mathrm{d}x\,\mathrm{d}z=1$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\hat{X}(z)-x)\frac{\partial \ln p(x,z)}{\partial x}p(x,z)\,\mathrm{d}x\,\mathrm{d}z=1$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\partial \ln p(x,z)}{\partial x}\sqrt{p(x,z)}\cdot\sqrt{p(x,z)}(\hat{X}(z)-x)\,\mathrm{d}x\,\mathrm{d}z=1$$

许瓦兹不等式 $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f^2(x,z)\,\mathrm{d}x\,\mathrm{d}z\cdot\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g^2(x,z)\,\mathrm{d}x\,\mathrm{d}z\ge\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,z)g(x,z)\,\mathrm{d}x\,\mathrm{d}z\right]^2$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\frac{\partial \ln p(x,z)}{\partial x}\right]^2 p(x,z)\,\mathrm{d}x\,\mathrm{d}z\cdot\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,z)(\hat{X}(z)-x)^2\,\mathrm{d}x\,\mathrm{d}z\ge 1$$

$$\mathrm{E}\left[(\hat{X}-X)^2\right]\ge \mathrm{E}\left\{\left[\frac{\partial \ln p(x,z)}{\partial x}\right]^2\right\}^{-1}=I^{-1}(x)$$

Rao-Cramer下界，$I(x)$称为Fisher信息量

状态向量情形 $\mathrm{E}\left[\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\right]\ge \mathrm{E}\left[\frac{\partial \ln p(\boldsymbol{x},\boldsymbol{z})}{\partial \boldsymbol{x}}\cdot\frac{\partial \ln p(\boldsymbol{x},\boldsymbol{z})}{\partial \boldsymbol{x}^{\mathrm{T}}}\right]^{-1}$

37

18.9 估计量的性质

（3）一致性(consistency)

$$\lim_{N\to\infty}P\left\{\left|\hat{X}-X\right|<\varepsilon\right\}=1$$

含义：当观测Z的次数N增多时，估计量取被估计量的可能性为100%。

38

18.9 估计量的性质

$$\mathrm{E}\left[\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\right]\ge \mathrm{E}\left[\frac{\partial \ln p(\boldsymbol{x},\boldsymbol{z})}{\partial \boldsymbol{x}}\cdot\frac{\partial \ln p(\boldsymbol{x},\boldsymbol{z})}{\partial \boldsymbol{x}^{\mathrm{T}}}\right]^{-1}$$

【举例】正态分布估计的Rao-Cramer下界。

$$p(\boldsymbol{x},\boldsymbol{z})=\frac{1}{(2\pi)^{n/2}\left|\boldsymbol{C}_{X|Z}\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})^{\mathrm{T}}\boldsymbol{C}_{X|Z}^{-1}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})\right\}\times$$
$$\frac{1}{(2\pi)^{m/2}\left|\boldsymbol{C}_Z\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{m}_Z)^{\mathrm{T}}\boldsymbol{C}_Z^{-1}(\boldsymbol{z}-\boldsymbol{m}_Z)\right\}$$

$$\frac{\partial \ln p(\boldsymbol{x},\boldsymbol{z})}{\partial \boldsymbol{x}}=\boldsymbol{0}+\frac{\partial}{\partial \boldsymbol{x}}\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})^{\mathrm{T}}\boldsymbol{C}_{X|Z}^{-1}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})\right\}+\boldsymbol{0}$$
$$=-\boldsymbol{C}_{X|Z}^{-1}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})$$

$$\begin{cases}\hat{\boldsymbol{X}}_{\mathrm{LMV}}=\boldsymbol{m}_{X|Z}=\boldsymbol{m}_X+\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}(\boldsymbol{Z}-\boldsymbol{m}_Z)\\ \mathrm{E}\left[\tilde{\boldsymbol{X}}_{\mathrm{LMV}}\tilde{\boldsymbol{X}}_{\mathrm{LMV}}^{\mathrm{T}}\right]=\boldsymbol{C}_{X|Z}=\boldsymbol{C}_X-\boldsymbol{C}_{XZ}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{ZX}\end{cases}$$

LMV达到了Rao-Cramer下界！

不可能还有估计方法比LMV好
所以Kalman已经是最优估计了、

$$\boldsymbol{I}^{-1}(x)=\mathrm{E}\left[\frac{\partial \ln p(\boldsymbol{x},\boldsymbol{z})}{\partial \boldsymbol{x}}\cdot\frac{\partial \ln p(\boldsymbol{x},\boldsymbol{z})}{\partial \boldsymbol{x}^{\mathrm{T}}}\right]^{-1}$$
$$=\mathrm{E}\left[\boldsymbol{C}_{X|Z}^{-1}(\boldsymbol{x}-\boldsymbol{m}_{X|Z})(\boldsymbol{x}-\boldsymbol{m}_{X|Z})^{\mathrm{T}}\boldsymbol{C}_{X|Z}^{-\mathrm{T}}\right]^{-1}$$
$$=\boldsymbol{C}_{X|Z}\mathrm{E}\left[(\boldsymbol{x}-\boldsymbol{m}_{X|Z})(\boldsymbol{x}-\boldsymbol{m}_{X|Z})^{\mathrm{T}}\right]^{-1}\boldsymbol{C}_{X|Z}=\boldsymbol{C}_{X|Z}$$

39

18.10 递推贝叶斯估计 —> 与时间有关系

状态空间模型与概率模型描述的等价性

非线性 ←
$$\begin{cases} \boldsymbol{X}_k = \boldsymbol{f}(\boldsymbol{X}_{k-1}, \boldsymbol{W}_{k-1}) \\ \boldsymbol{Z}_k = \boldsymbol{h}(\boldsymbol{X}_k, \boldsymbol{V}_k) \end{cases} \Longleftrightarrow \left\{ \begin{array}{l} p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1}) \\ p(\boldsymbol{z}_k \mid \boldsymbol{x}_k) \end{array} \right.$$
—> HMM 模型

状态转移具有马尔可夫性 $p\big(\boldsymbol{x}_k \mid (\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{k-1})\big) = p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1})$ —> Markov 链

贝叶斯估计目的 $\overline{\boldsymbol{z}}_k = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_k\} \Rightarrow p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_k)$ 分布函数

$\Rightarrow \hat{\boldsymbol{X}}_{k,\mathrm{MV}}(\overline{\boldsymbol{z}}_k) = \mathrm{E}\left[\boldsymbol{X}_k \mid \overline{\boldsymbol{z}}_k\right] = \int \boldsymbol{x}_k p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_k) \mathrm{d}\boldsymbol{x}_k$

$$p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_k) = \frac{p(\overline{\boldsymbol{z}}_k \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k)}{p(\overline{\boldsymbol{z}}_k)} = \frac{p\big((\boldsymbol{z}_k, \overline{\boldsymbol{z}}_{k-1}) \mid \boldsymbol{x}_k\big) p(\boldsymbol{x}_k)}{p(\boldsymbol{z}_k, \overline{\boldsymbol{z}}_{k-1})}$$

改成递推

$$\begin{aligned} &= p(\boldsymbol{z}_k \mid \boldsymbol{x}_k, \overline{\boldsymbol{z}}_{k-1} \mid \boldsymbol{x}_k) \\ &= p\big((\boldsymbol{z}_k \mid \boldsymbol{x}_k) \mid (\overline{\boldsymbol{z}}_{k-1} \mid \boldsymbol{x}_k)\big) p(\overline{\boldsymbol{z}}_{k-1} \mid \boldsymbol{x}_k) \\ &= p\big(\boldsymbol{z}_k \mid (\overline{\boldsymbol{z}}_{k-1}, \boldsymbol{x}_k)\big) p(\overline{\boldsymbol{z}}_{k-1} \mid \boldsymbol{x}_k) \\ &= p\big(\boldsymbol{z}_k \mid \boldsymbol{x}_k\big) p(\overline{\boldsymbol{z}}_{k-1} \mid \boldsymbol{x}_k) \end{aligned}$$

$$= \frac{p(\boldsymbol{z}_k \mid \boldsymbol{x}_k) p(\overline{\boldsymbol{z}}_{k-1} \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k)}{p(\boldsymbol{z}_k, \overline{\boldsymbol{z}}_{k-1})}$$

$$= \frac{p(\boldsymbol{z}_k \mid \boldsymbol{x}_k)\left[p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_{k-1}) p(\overline{\boldsymbol{z}}_{k-1}) / p(\boldsymbol{x}_k)\right] p(\boldsymbol{x}_k)}{p(\boldsymbol{z}_k \mid \overline{\boldsymbol{z}}_{k-1}) p(\overline{\boldsymbol{z}}_{k-1})} = \frac{p(\boldsymbol{z}_k \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_{k-1})}{p(\boldsymbol{z}_k \mid \overline{\boldsymbol{z}}_{k-1})}$$

40

18.10 递推贝叶斯估计

$$p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_k) = \frac{p(\boldsymbol{z}_k \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_{k-1})}{p(\boldsymbol{z}_k \mid \overline{\boldsymbol{z}}_{k-1})}$$

$$\begin{aligned} &= \int p\big((\boldsymbol{x}_k, \boldsymbol{x}_{k-1}) \mid \overline{\boldsymbol{z}}_{k-1}\big) \mathrm{d}\boldsymbol{x}_{k-1} \\ &= \int p\big(\boldsymbol{x}_k \mid (\boldsymbol{x}_{k-1}, \overline{\boldsymbol{z}}_{k-1})\big) p(\boldsymbol{x}_{k-1} \mid \overline{\boldsymbol{z}}_{k-1}) \mathrm{d}\boldsymbol{x}_{k-1} \\ &= \int p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1}) p(\boldsymbol{x}_{k-1} \mid \overline{\boldsymbol{z}}_{k-1}) \mathrm{d}\boldsymbol{x}_{k-1} \end{aligned}$$

$$\begin{aligned} &= \int p\big((\boldsymbol{z}_k, \boldsymbol{x}_k) \mid \overline{\boldsymbol{z}}_{k-1}\big) \mathrm{d}\boldsymbol{x}_k \\ &= \int p\big(\boldsymbol{z}_k \mid (\boldsymbol{x}_k, \overline{\boldsymbol{z}}_{k-1})\big) p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_{k-1}) \mathrm{d}\boldsymbol{x}_k \\ &= \int p(\boldsymbol{z}_k \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_{k-1}) \mathrm{d}\boldsymbol{x}_k \end{aligned}$$

$$= \frac{p(\boldsymbol{z}_k \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_{k-1})}{\int p(\boldsymbol{z}_k \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_{k-1}) \mathrm{d}\boldsymbol{x}_k}$$

递推贝叶斯估计的概率密度**预测**和**更新**公式

$$p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_{k-1}) = \int p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1}) p(\boldsymbol{x}_{k-1} \mid \overline{\boldsymbol{z}}_{k-1}) \mathrm{d}\boldsymbol{x}_{k-1} \qquad \leftarrow p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1})$$

$$p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_k) = \frac{p(\boldsymbol{z}_k \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_{k-1})}{\int p(\boldsymbol{z}_k \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k \mid \overline{\boldsymbol{z}}_{k-1}) \mathrm{d}\boldsymbol{x}_k} \qquad \leftarrow p(\boldsymbol{z}_k \mid \boldsymbol{x}_k)$$

- 卡尔曼滤波：线性正态模型情况下简化为一二阶矩传播；
- UKF滤波：采用少量粒子点近似一二阶矩传播；
- PF(粒子)滤波：采用大量粒子点近似概率密度传播。

41