

---

清华大学 研究生课程讲义

# 统计信号处理

廖庆敏

清华大学

[liaoqm@tsinghua.edu.cn](mailto:liaoqm@tsinghua.edu.cn)

2024 年 9 月

# 目 录

导言 .....	3
第 1 章 信号检测理论.....	4
1.1 信号检测模型.....	4
1.2 常用判决准则.....	5
1.2.1 Bayes 准则（最小平均风险准则） .....	5
1.2.2 极大极小准则（安全平均风险准则）Minimax Criterion.....	8
1.2.3 Neyman-Pearson 准则.....	10
1.2.4 最小错误概率准则（理想观测者准则）Minimum Probability of Error Criterion.....	12
1.2.5 最大似然准则Maximum Likelihood Criterion.....	12
1.2.6 最大后验概率准则Maximum A Posteriori Probability Criterion.....	12
1.2.7 各类准则小结（自习） .....	13
1.3 在白高斯信道中的确知信号检测 .....	13
1.3.1 连续信号的似然比检验及其最佳接收机.....	13
1.3.2 二元数字通信系统 .....	17
1.3.3 雷达系统 .....	18
1.4 多元假设检验（MULTIPLE-HYPOTHESIS TESTING） .....	19
1.5 复合假设检验与随机参量信号检测 .....	21
1.5.1 复合假设检验（Composite Hypothesis Testing） .....	21
1.5.2 在白高斯信道中随机幅度信号检测.....	22
1.5.3 在白高斯信道中随机相位信号检测.....	23
1.5.4 在白高斯信道中随幅随相信号检测.....	26
1.6 序列检测（WALD 检测） .....	29
1.7 在色高斯信道中的信号检测 .....	32
1.7.1 Karhunen-Loève 正交展开法.....	32
1.7.2 白化滤波器 .....	33
1.7.3 接收机性能 .....	34
1.7.4 最佳信号波形 .....	35
1.8 小结.....	36
第 2 章 信号估值理论.....	38
2.1 信号参量估值模型 .....	38
2.2 常用估值准则.....	38
2.2.1 Bayes 估值.....	38
2.2.2 极大极小估值 .....	42
2.2.3 最大后验估值 .....	42
2.2.4 最大似然估值 .....	42
2.3 估计量的性质 .....	43
2.4 CRAMÉR-RAO 不等式 .....	44
2.4.1 非随机参量 .....	44
2.4.2 随机参量 .....	46
2.5 在高斯信道中单参量信号的估值 .....	48
2.5.1 白高斯信道 .....	48
2.5.2 色高斯信道 .....	52
2.6 多参量估值 .....	55
2.7 线性最小均方估值（LMS） .....	56
2.8 正交原理.....	58
2.9 最小二乘估值（LS）和加权最小二乘估值（LSW） .....	59
2.9.1 最小二乘估值（LS） .....	59
2.9.2 加权最小二乘估值（LSW） .....	60
2.10 小结.....	64
第 3 章 波形估计理论.....	66

3.1	波形的线性最小均方估计 .....	6 6
3.2	WIENER 滤波 .....	6 8
3.2.1	物理不可实现的 Wiener 滤波器 .....	6 8
3.2.2	物理可实现的 Wiener 滤波器 .....	6 9
3.3	离散时间 WIENER 滤波 .....	7 2
3.3.1	有限观测区间的离散 Wiener 滤波器 .....	7 2
3.3.2	物理可实现的离散 Wiener 滤波器 .....	7 3
3.3.3	物理不可实现的离散 Wiener 滤波器 .....	7 3
3.3.4	离散 Wiener 滤波矩阵 .....	7 4
3.4	KALMAN 滤波导言 .....	7 5
3.4.1	起因 .....	7 5
3.4.2	状态空间模型 .....	7 6
3.5	离散时间 KALMAN 滤波 .....	7 9
3.6	连续时间 KALMAN 滤波 .....	8 6
3.7	KALMAN 滤波的稳定与发散 .....	9 1
3.7.1	滤波的稳定性 .....	9 1
3.7.2	滤波的发散问题 .....	9 2
3.8	KALMAN 滤波的推广 (简介) .....	9 5
3.9	非线性滤波 .....	9 6
3.10	匹配滤波器 .....	1 0 1
3.11	小结 .....	1 0 2
第 4 章	非参量和鲁棒检测与估计 .....	1 0 3
4.1	非参量检测 (NONPARAMETRIC DETECTION) .....	1 0 3
4.1.1	简单非参量检测——符号检测 (Sign Test) .....	1 0 3
4.1.2	Wilcoxon (威尔柯克孙) 检测 (W 检测器) .....	1 0 5
4.1.3	渐近相对效率 (Asymptotic Relative Efficiency) .....	1 0 5
4.1.4	双输入系统 (Two-Input Systems) .....	1 0 8
4.2	局部最佳检测 (LOCALLY OPTIMAL DETECTION) .....	1 0 9
4.2.1	举例 .....	1 1 0
4.2.2	广义高斯噪声 (Generalized Gaussian Noise) 与局部最佳检测 .....	1 1 1
4.3	鲁棒检测和估值 (ROBUST DETECTION AND ESTIMATION) .....	1 1 2
4.3.1	鲁棒假设检验 (休伯 Huber 模型) .....	1 1 2
4.3.2	鲁棒估值 .....	1 1 4
4.4	小结 .....	1 1 6
第 5 章	总 结 .....	1 1 7

#### ◆参考书目:

1. T.A. Schonhoff & A.A. Giordano, 《Detection and Estimation: Theory and its Applications》, Pearson Education, Inc. 2007 (信号检测与估计——理论与应用, 关欣等译, 电子工业出版社, 2012 年)
2. M.D. Srinath, P.K. Rajasekaran & R. Viswanathan, 《Introduction to Statistical Signal Processing with Applications》, Prentice Hall, 1996.
3. Steven M. Kay, 《Fundamentals of statistical signal processing, Volume I: Estimation theory》(©1993) & 《Fundamentals of statistical signal processing, Volume II: Detection theory》(©1998), Pearson Education. (《统计信号处理基础——估计与检测理论 (卷 I、卷 II 合集)》, 罗鹏飞等译, 电子工业出版社, 2023 年)
4. James V. Candy, 《Bayesian Signal Processing: Classical, Modern, and Particle Filtering Methods》(2<sup>nd</sup> edition, ©2016), John Wiley & Sons, Inc. (宗华等译, 哈尔滨工业大学出版社, 2023 年).
5. Harry L. Van Trees, Kristine L. Bell, with Zhi Tian, 《Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory》(2nd Edition), John Wiley & Sons, Inc. 2013.
6. 周强、曲长文、张彬编著, 《信号检测与估计》(第 2 版), 电子工业出版社, 2021 年。

# 导 言

## ◆问题的提出：

在各种信息传输系统（如：通信、雷达、自动控制等系统）中，存在各种外部设备**干扰**（由电气设备、无线电台等产生的）和内部**噪声**（电子热噪声等），它们通常是**随机**的（不能用确定信号来描述，而只能用统计量或概率等来描述）。由于它们**混杂**在我们想传输的信息中，从而造成接收机难于从接收到的信号中**分离**出有用的**信息**来。

无线电定位：向目标发射电磁能，观测反射的电磁波。

问题一：目标是否存在？（检测问题）

问题二：如目标存在，距离为多少？速度为多少？（估计问题）

模拟通信：电视广播。无线广播：如短波接收机，高频干扰严重，低频相对较轻。

数字通信：发射“0”、“1”，接收机判决（二元检测）。

语音传输：发射语音谱参数，接收机根据参数综合出语音波形（估值）。

字符识别：中文 OCR，字样数千（手写各异），统计识别（多元检测）。

图象处理：噪声图象复原（Wiener、Kalman 滤波）（波形估计）。

## ◆研究内容：

如何从混入了干扰和噪声的信号中**最有效地提取出有用信息**。

## ◆研究方法：

概率论和数理统计方法。涉及的课程：概率论，随机过程，信号与系统，线性代数等。

## ◆应用领域：

通信、雷达、自动控制、遥测、天文、气象、地震、医学、社会等。

## ◆统计信号处理（统计检测、信号检测与估计）分类：

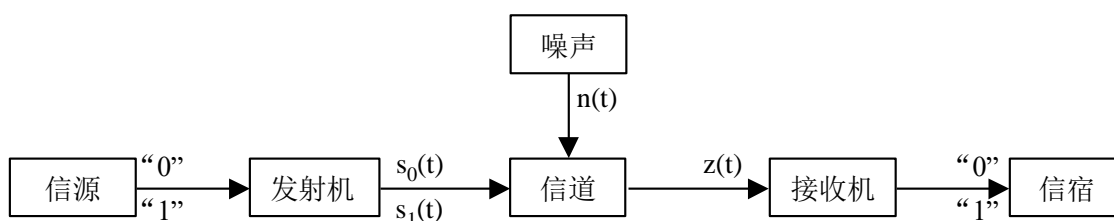
- |   |          |          |
|---|----------|----------|
| 1、确知信号是否存在： <b>简单假设检验</b> 。                   | } 狭义信号检测 | } 广义信号检测 |
| 2、含随机参量的信号是否存在： <b>复合假设检验</b> 。               |          |          |
| 3、对信号参量（观测时间 $T$ 内不变）的估计： <b>信号估值/参量估计</b> 。  |          |          |
| 4、将信号（观测时间 $T$ 内连续）过滤出来：信号波形复原/ <b>波形估计</b> 。 |          |          |

# 第1章 信号检测理论

## 1.1 信号检测模型

以二元通信系统为例。

◆物理过程：



① 加性噪声： $z(t)=s_i(t)+n(t)$ ,  $i=0,1$ 。其中， $s_i(t)$ 为确定性信号， $n(t)$ 为 Gauss 分布的噪声信号。

② 作一次观测： $z=s_i+n$ ,  $t=t_0$ 时刻抽样。

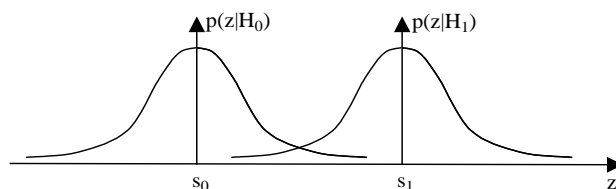
问题：如何根据  $z$  值来确定发射的是哪一个信号？

③ 假设： $H_0$ （发射  $s_0$ ）， $H_1$ （发射  $s_1$ ）。

$$H_0: z=s_0+n;$$

条件概率密度如下图：

$$H_1: z=s_1+n。$$



④ 假定先验概率：信源以概率  $P(H_0)$ 送出符号“0”；以概率  $P(H_1)$ 送出符号“1”。

⑤ 寻找  $\{z\}$ 值域（记为  $\mathcal{Z}$ ）的一个划分以给出适当的判决，即：判决  $\mathcal{D}_0$ （收到的是“0”）和判决  $\mathcal{D}_1$ （收到的是“1”）。

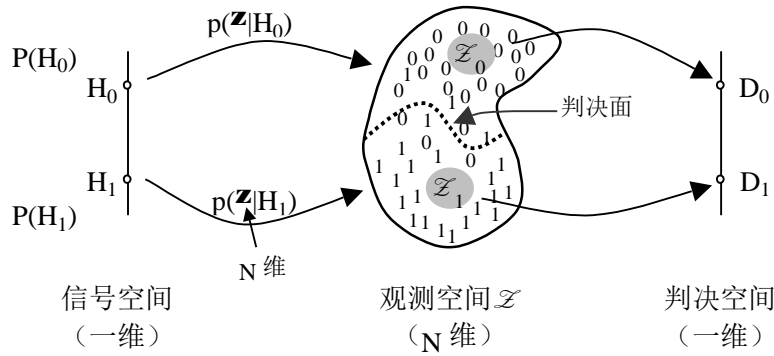
实际上，就是要寻找一个  $\mathcal{Z}$  域的划分： $\mathcal{Z}_0$  和  $\mathcal{Z}_1$ 。

然后，作出判决：
$$\begin{cases} \mathcal{D}_0, & \text{如果 } z \in \mathcal{Z}_0; \\ \mathcal{D}_1, & \text{如果 } z \in \mathcal{Z}_1. \end{cases}$$

问题：怎样对  $\mathcal{Z}$  进行适当的划分？

实际上， $\mathcal{Z}$  域划分的确定依赖于所使用的准则（和实际应用相结合的）。这就是我们要研究的。

◆检测模型：



**关键：**根据某个准则，确定  $\mathcal{Z}$  的划分（界限，判决面）。

## 1.2 常用判决准则

### 1.2.1 Bayes 准则（最小平均风险准则）

#### ◆二元假设：

每次判决有四种结果：

事件	概率	代价因子	通信应用（举例）	雷达应用（举例）
① $D_0, H_0$	$P(D_0, H_0)$	$C_{00}$	0	0
② $D_1, H_1$	$P(D_1, H_1)$	$C_{11}$	0	0
③ $D_1, H_0$	$P(D_1, H_0)$	$C_{10}$	1	$\neq$ (如：导弹检测)
④ $D_0, H_1$	$P(D_0, H_1)$	$C_{01}$	1	

#### ◆平均风险：

$$\begin{aligned}\bar{R} &= C_{00} P(D_0, H_0) + C_{11} P(D_1, H_1) + C_{10} P(D_1, H_0) + C_{01} P(D_0, H_1) \\ &= C_{00} P(H_0) P(D_0|H_0) + C_{11} P(H_1) P(D_1|H_1) + C_{10} P(H_0) P(D_1|H_0) + C_{01} P(H_1) P(D_0|H_1)\end{aligned}$$

这里的四个条件概率及其描述如下表所示：

条件概率表达式	名称	通信术语	雷达术语	符号表示
$P(D_0   H_0) = \int_{\mathcal{Z}_0} p(\mathbf{z}   H_0) d\mathbf{z}$	正确			
$P(D_1   H_1) = \int_{\mathcal{Z}_1} p(\mathbf{z}   H_1) d\mathbf{z}$	正确		检测	$P_D$
$P(D_1   H_0) = \int_{\mathcal{Z}_1} p(\mathbf{z}   H_0) d\mathbf{z}$	第一类错误	虚报	虚警	$P_F$
$P(D_0   H_1) = \int_{\mathcal{Z}_0} p(\mathbf{z}   H_1) d\mathbf{z}$	第二类错误	漏报	漏警	$P_M$

◆问题：寻找一个划分  $\mathcal{Z} = \mathcal{Z}_0 \cup \mathcal{Z}_1$  且  $\mathcal{Z}_0 \cap \mathcal{Z}_1 = \emptyset$  使  $\bar{R}$  最小。

推导：由  $\bar{R}$  的表达式，我们可得：

$$\begin{aligned}\bar{R}(\mathcal{Z}_0) &= [C_{11} P(H_1) + C_{10} P(H_0)] + \\ &\quad + \int_{\mathcal{Z}_0} [P(H_1) (C_{01} - C_{11}) p(\mathbf{z} | H_1) - P(H_0) (C_{10} - C_{00}) p(\mathbf{z} | H_0)] d\mathbf{z}\end{aligned}$$

由于第一项 $[C_{11}P(H_1)+C_{10}P(H_0)]$ 为固定值，故只考虑后一积分项的取值。

$$\text{令: } F(\mathcal{Z}_0) = \int_{\mathcal{Z}_0} G(\mathbf{z}) p(\mathbf{z}|H_0) d\mathbf{z}$$

$$\text{其中: } G(\mathbf{z}) = P(H_1)(C_{01} - C_{11})\lambda(\mathbf{z}) - P(H_0)(C_{10} - C_{00}),$$

$$\text{且定义 } \lambda(\mathbf{z}) \equiv \frac{p(\mathbf{z}|H_1)}{p(\mathbf{z}|H_0)}。$$

要使 $\bar{R}$ 取最小值，则要使 $\mathcal{Z}_0$ 的选取使 $F(\mathcal{Z}_0)$ 取负的最大值。而要想使 $F(\mathcal{Z}_0)$ 取负的最大值，又由于 $p(\mathbf{z}|H_0)>0$ ，应该这样来组成 $\mathcal{Z}_0$ ：

凡是使 $G(\mathbf{z})<0$ 的 $\mathbf{z}$ 应被划入 $\mathcal{Z}_0$ 。

当然， $\mathcal{Z}$ 中除去 $\mathcal{Z}_0$ 外的所有 $\mathbf{z}$ 应属于 $\mathcal{Z}_1$ ，即

凡是使 $G(\mathbf{z})\geq 0$ 的 $\mathbf{z}$ 应被划入 $\mathcal{Z}_1$ 。

$$\text{使 } \bar{R} \text{ 取最小值的 } \mathcal{Z} \text{ 的划分应遵循这样的规则: } G(\mathbf{z}) \begin{matrix} \mathbf{z} \in \mathcal{Z}_1 \\ > \\ < \\ \mathbf{z} \in \mathcal{Z}_0 \end{matrix} 0$$

又由 $G(\mathbf{z})$ 的表达式以及 $P(H_1)>0$ 、 $P(H_0)>0$ 和 $(C_{01} - C_{11})>0$ 、 $(C_{10} - C_{00})>0$ ，综上所述，我们可得：

$$\lambda(\mathbf{z}) \begin{matrix} \mathbf{z} \in \mathcal{Z}_1 \\ > \\ < \\ \mathbf{z} \in \mathcal{Z}_0 \end{matrix} \frac{P(H_0)(C_{10} - C_{00})}{P(H_1)(C_{01} - C_{11})}$$

◆最小平均风险准则（常称为 Bayes 准则）：

$$\lambda(\mathbf{z}) \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \lambda_B$$

其中， $\lambda(\mathbf{z}) = \frac{p(\mathbf{z}|H_1)}{p(\mathbf{z}|H_0)}$ ，称为“似然比”（Likelihood Ratio）；

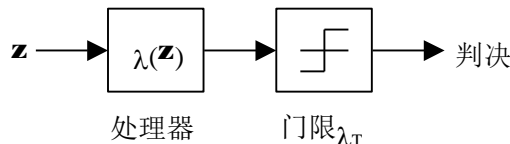
$$\lambda_B = \frac{P(H_0)(C_{10} - C_{00})}{P(H_1)(C_{01} - C_{11})}，\text{称为“Bayes 门限”。}$$

$$\text{对数形式: } \mathcal{L}(\mathbf{z}) \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \mathcal{L}_B$$

其中， $\mathcal{L}(\mathbf{z}) = \ln(\lambda(\mathbf{z}))$ ，称为“对数似然比”； $\mathcal{L}_B = \ln(\lambda_B)$ ，称为“对数 Bayes 门限”。

◆似然比检验（LRT: Likelihood Ratio Test）：

似然比接收机：



在实践中，似然比 $\lambda(\mathbf{z})$ 的计算不变，仅根据应用来改变似然比门限 $\lambda_T$ 。

因为 $\mathbf{z}$ 是一个随机变量，故 $\lambda(\mathbf{z})$ 也是一个随机变量。但前者是N维的，而后者是一维的。

如果我们采用 $\lambda(\mathbf{z})$ 作为检验统计量 $\gamma$ ，那么根据判决规则，我们有：

$$P_F = \int_{\lambda_B}^{+\infty} p(\gamma|H_0) d\gamma \quad \text{及} \quad P_M = \int_{-\infty}^{\lambda_B} p(\gamma|H_1) d\gamma$$

这里， $\gamma=\lambda(\mathbf{z})$ 的条件密度函数  $p(\gamma|H_i)$ 可由 $\lambda(\mathbf{z})$ 的表达式及  $\mathbf{z}$  的条件密度函数来得到。

◆例子：

二元通信系统， $C_{00}=C_{11}=0$ ， $C_{10}=C_{01}=1$ ，一次观测。

$P(H_0)=1/2$ （等先验概率）。

$$\begin{aligned} H_1: \quad z &= A + n \\ H_0: \quad z &= 0 + n \end{aligned} \quad \text{其中: } n \sim N(0, \sigma^2)$$

求：判决规则及判决划分。

解：由题意可得：  $p(z|H_0) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2} z^2)$

$$p(z|H_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2} (z-A)^2)$$

故，可算得对数似然比如下：

$$\mathcal{L}(z) = \ln(\lambda(z)) = \frac{1}{2\sigma^2} (z^2 - (z-A)^2) = \frac{A}{\sigma^2} (z - \frac{A}{2})$$

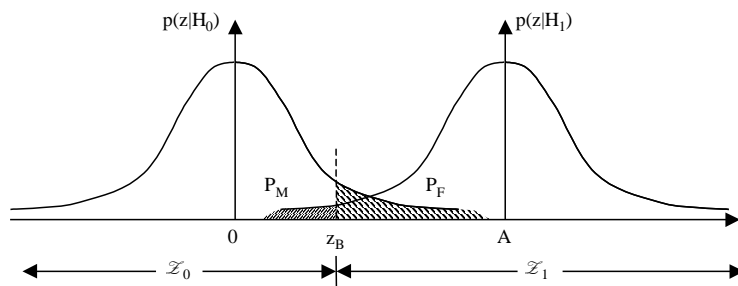
又，对数 Bayes 门限为：

$$\mathcal{L}_B = \ln(\lambda_B) = \ln\left(\frac{\frac{1}{2}(1-0)}{\frac{1}{2}(1-0)}\right) = 0$$

所以，根据 Bayes 准则，我们有：

$$\frac{A}{\sigma^2} (z - \frac{A}{2}) \underset{H_0}{\overset{H_1}{>}} 0$$

即：  $z \underset{H_0}{\overset{H_1}{>}} z_B$ ， 其中：  $z_B = \frac{A}{2}$ 。 或：  $\mathcal{Z}_0 = (-\infty, \frac{A}{2})$ ，  $\mathcal{Z}_1 = [\frac{A}{2}, +\infty)$ 。



#例毕

◆课外：

在上例中，令  $A=2$ ， $\sigma^2=1$ ，试将平均风险  $\bar{R}$  表述为先验概率  $P(H_0)$ 的函数形式（即  $\bar{R}(P(H_0))$ ），然后分别求当  $P(H_0)=1/3$ 、 $P(H_0)=1/2$  和  $P(H_0)=2/3$  时平均风险之值（记为  $\bar{R}_B(1/3)$ 、 $\bar{R}_B(1/2)$ 和  $\bar{R}_B(2/3)$ ）并比较它们的大小。

注：需利用  $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} u^2) du$  “正态概率积分函数”，重要性质： $\Phi(x) = 1 - \Phi(-x)$ 。

（答案提示：0.1450、0.1587、0.1450）



### 1.2.2 极大极小准则（安全平均风险准则）Minimax Criterion

由计算 Bayes 门限 $\lambda_B$ 的表达式可知：Bayes 准则的建立需事先知道先验概率。

◆问题：若事先并不知道准确的先验概率值，那么我们该怎样来建立一个适当的判决准则？

思路：当然，可先假定一先验概率值，再建立 Bayes 准则。但若此值与实际值不符，后果呢？

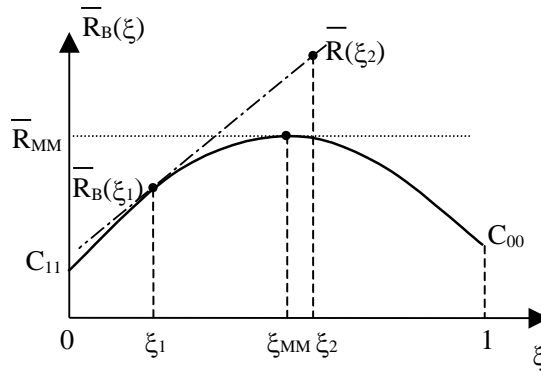
推导：令  $P(H_0)=\xi$ ，则  $P(H_1)=1-\xi$ 。那么，我们有：

$$\begin{aligned}\bar{R}_B(\xi) &= C_{00}\xi P(D_0|H_0) + C_{11}(1-\xi) P(D_1|H_1) + C_{10}\xi P(D_1|H_0) + C_{01}(1-\xi) P(D_0|H_1) \\ &= [C_{11}P(D_1|H_1) + C_{01}P(D_0|H_1)] + \\ &\quad + \xi[C_{00}P(D_0|H_0) + C_{10}P(D_1|H_0) - C_{11}P(D_1|H_1) - C_{01}P(D_0|H_1)] \\ &= \alpha(\mathbf{z}_B|\xi) + \xi \cdot \beta(\mathbf{z}_B|\xi)\end{aligned}$$

其中：  $\alpha(\mathbf{z}_B|\xi) = C_{11}P(D_1|H_1) + C_{01}P(D_0|H_1)$ ；

$\beta(\mathbf{z}_B|\xi) = C_{00}P(D_0|H_0) + C_{10}P(D_1|H_0) - C_{11}P(D_1|H_1) - C_{01}P(D_0|H_1)$ 。

我们可作出  $\bar{R}_B(\xi)-\xi$  的一条典型曲线如下图所示（实黑线）：



如果我们假定 $\xi=\xi_1$ 并据此来设计接收机（即确定 $\mathbf{z}_B$ ），那么这时的平均风险为：

$$\bar{R}_B(\xi_1) = \alpha(\mathbf{z}_B|\xi_1) + \xi_1 \cdot \beta(\mathbf{z}_B|\xi_1)$$

但实际上如果 $\xi \neq \xi_1$ 而是 $\xi=\xi_2$ ，我们仍使用同一接收机，那么此时的平均风险为：

$$\bar{R}(\xi_2) = \alpha(\mathbf{z}_B|\xi_1) + \xi_2 \cdot \beta(\mathbf{z}_B|\xi_1)$$

换句话说，此时的平均风险与先验概率成线性关系（如上图中的点划线所示）。因此我们可能冒很大的风险（如 $\bar{R}(\xi_2) \gg \bar{R}_B(\xi_2)$ ）。

那么，我们是否可采取措施来避免这种情况的产生呢？

是的。实际上我们可以选取使 $\beta(\mathbf{z}_B|\xi)=0$ 所对应的那个 $\xi$ 的值（记为 $\xi_{MM}$ ）来设计接收机。此时，不管先验概率如何变化，平均风险均为一恒定值 $\bar{R}_{MM}$ （如上图中的圆点线所示），即：

$$\bar{R}(\xi) \equiv \bar{R}_{MM} = \bar{R}_B(\xi_{MM}) = \alpha(\mathbf{z}_B|\xi_{MM})$$

参考上图，我们可知： $\bar{R}_{MM}$ 等于最小平均风险（Bayes 风险）的极大值。故，我们称由此而产生的判决准则为“极大极小准则”或“安全平均风险准则”。

◆极大极小化方程：

由 $\beta(\mathbf{z}_B|\xi_{MM})=0$ ，我们有：

$$C_{00}P(D_0|H_0) + C_{10}P(D_1|H_0) = C_{11}P(D_1|H_1) + C_{01}P(D_0|H_1)$$

$$\text{或} \quad C_{00}[1 - P_F(\xi_{MM})] + C_{10}P_F(\xi_{MM}) = C_{11}[1 - P_M(\xi_{MM})] + C_{01}P_M(\xi_{MM})$$

注意到它的左边实际上是与  $H_0$  有关的风险而其右边是与  $H_1$  有关的风险。

故，极大极小准则又称为“等风险准则”。

◆极大极小准则：

$$\lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{>}} \lambda_{MM}$$

$$\text{其中, } \lambda(\mathbf{z}) = \frac{p(\mathbf{z}|H_1)}{p(\mathbf{z}|H_0)}, \text{ 即前面定义的似然比; } \lambda_{MM} = \frac{\xi_{MM}(C_{10} - C_{00})}{(1 - \xi_{MM})(C_{01} - C_{11})}。$$

事实上，我们可由下述步骤求  $\xi_{MM}$  之值：首先，由判决准则给出含  $\lambda_{MM}$  的检验统计量判决不等式；再由极大极小化方程求出  $\lambda_{MM}$  之值；最后，由  $\lambda_{MM} = \lambda_B(\xi_{MM})$  求出  $\xi_{MM}$  之值。

◆例（除先验概率未知外其它条件同前例）：

二元通信系统， $C_{00}=C_{11}=0$ ， $C_{10}=C_{01}=1$ ，一次观测。

$$\begin{aligned} H_1: z &= A + n \\ H_0: z &= 0 + n \end{aligned} \quad \text{其中: } n \sim N(0, \sigma^2)$$

求：先验概率  $\xi_{MM}$  及门限  $z_{MM}$ 。

解：由前例中所得的对数似然比及极大极小准则有：

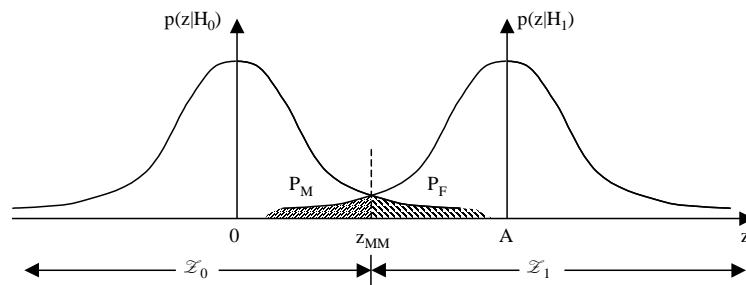
$$\mathcal{L}(z) = \frac{A}{\sigma^2} \left( z - \frac{A}{2} \right) \underset{H_0}{\overset{H_1}{>}} \mathcal{L}_{MM}$$

$$\text{即 } z \underset{H_0}{\overset{H_1}{>}} z_{MM} \equiv \frac{A}{2} + \frac{\sigma^2}{A} \mathcal{L}_{MM}$$

故，可将  $P_M$  和  $P_F$  表述如下：

$$P_M(\xi_{MM}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{z_{MM}} \exp\left(-\frac{1}{2\sigma^2}(z-A)^2\right) dz = \Phi\left(\frac{z_{MM}-A}{\sigma}\right)$$

$$P_F(\xi_{MM}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{z_{MM}}^{+\infty} \exp\left(-\frac{1}{2\sigma^2}z^2\right) dz = \Phi\left(\frac{-z_{MM}}{\sigma}\right)$$



将已知代价因子代入极大极小化方程得：

$$P_M(\xi_{MM}) = P_F(\xi_{MM}) \quad \text{即} \quad \Phi\left(\frac{z_{MM}-A}{\sigma}\right) = \Phi\left(\frac{-z_{MM}}{\sigma}\right)$$

$$\text{因此, } \frac{z_{MM}-A}{\sigma} = \frac{-z_{MM}}{\sigma}$$

故,  $z_{MM} = \frac{A}{2}$ 。

从而, 由  $z_{MM} = \frac{A}{2} + \frac{\sigma^2}{A} \mathcal{L}_{MM}$  可得:  $\mathcal{L}_{MM} = 0$

又由对数 Bayes 门限计算公式:  $\mathcal{L}_{MM} = \ln(\lambda_{MM}) = \ln\left(\frac{\xi_{MM}(1-0)}{(1-\xi_{MM})(1-0)}\right) = \ln\left(\frac{\xi_{MM}}{(1-\xi_{MM})}\right)$

因而,  $\ln\left(\frac{\xi_{MM}}{(1-\xi_{MM})}\right) = 0$

所以,  $\xi_{MM} = \frac{1}{2}$ 。

这个结果也可以解释为: 当先验概率未知时, 采用等先验概率可获得最小平均风险。

#例毕

### 1.2.3 Neyman-Pearson 准则

◆问题: 若我们不能确定先验概率和代价因子, 那么我们该怎样来建立一个适当的判决准则?

思路: 设计接收机的两个重要指标是它的虚警概率  $P_F$  和检测概率  $P_D$ 。通常, 我们希望  $P_F$  尽可能小而  $P_D$  尽可能大。然而, 这两者是相关的 (常常是  $P_F$  减小会导致  $P_D$  减小)。所以, 我们将前面的问题改为:

在固定  $P_F$  为一个较小值 (即  $P_F = \alpha$ ) 的情况下, 如何使  $P_D$  取得最大值?

推导: 这是一个求条件极值的问题, 我们可采用 Lagrange 乘子法来求解。

由于  $P_D = 1 - P_M$ , 所以使  $P_D$  最大等效于使  $P_M$  最小。

故, 我们希望在  $P_F = \alpha$  的情况下使  $P_M$  最小。为此, 构造一个如下的目标函数:

$$\begin{aligned} J &= P_M + \mu \cdot (P_F - \alpha) \quad \text{其中: } \mu \text{ (Lagrange 乘子)} \geq 0 \\ &= P(D_0|H_1) + \mu \cdot [1 - P(D_0|H_0) - \alpha] \\ &= \int_{\mathcal{Z}_0} p(\mathbf{z}|H_1) d\mathbf{z} + \mu \cdot (1 - \int_{\mathcal{Z}_0} p(\mathbf{z}|H_0) d\mathbf{z} - \alpha) \\ &= \mu \cdot (1 - \alpha) + \int_{\mathcal{Z}_0} [\lambda(\mathbf{z}) - \mu] p(\mathbf{z}|H_0) d\mathbf{z} \end{aligned}$$

与前面推导 Bayes 准则时类似, 要使  $J$  取最小值, 则:

凡是使  $[\lambda(\mathbf{z}) - \mu] < 0$  的  $\mathbf{z}$  应被划入  $\mathcal{Z}_0$ 。

换句话说, 如令  $\lambda_{NP} = \mu$ , 则我们有:  $\lambda(\mathbf{z}) \underset{\mathbf{z} \in \mathcal{Z}_0}{\overset{\mathbf{z} \in \mathcal{Z}_1}{>}} \lambda_{NP}$

◆Neyman-Pearson 准则:

$$\lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{>}} \lambda_{NP} \quad \text{或其对数形式:} \quad \mathcal{L}(\mathbf{z}) \underset{H_0}{\overset{H_1}{>}} \mathcal{L}_{NP}$$

其中,  $\lambda_{NP}$  可由这样来确定:  $P_F = \int_{\mathcal{Z}_1} p(\mathbf{z}|H_0) d\mathbf{z} \stackrel{\gamma = \lambda(\mathbf{z})}{=} \int_{\lambda_{NP}}^{+\infty} p(\gamma|H_0) d\gamma = \alpha$ 。

◆讨论 (思考 Neyman-Pearson 准则与 Bayes 准则之关系)

◆例:

同前例，并令  $A=1$ ,  $\sigma^2=1$ 。

求当  $P_F = \alpha = 0.1$  时  $\lambda_{NP}$  之值及对应的  $P_D$ 。

解：由前例并代入已知条件得： $\mathcal{L}(z) = \ln(\lambda(z)) = \frac{A}{\sigma^2} (z - \frac{A}{2}) = z - \frac{1}{2}$ 。

令  $\mathcal{L}(z)$  为检验统计量 (test statistic)，记为  $\gamma$ 。

由于  $z|H_0 \sim N(0,1)$ ，故： $\gamma|H_0 \sim N(-\frac{1}{2}, 1)$ 。

由 Neyman-Pearson 准则 (对数形式)，我们有：

$$P_F = \int_{\mathcal{L}_{NP}}^{+\infty} p(\gamma|H_0) d\gamma = 1 - \Phi(\mathcal{L}_{NP} + \frac{1}{2}) = \alpha = 0.1。$$

查表得： $\mathcal{L}_{NP} + \frac{1}{2} = 1.2816$

故： $\lambda_{NP} = \exp(\mathcal{L}_{NP}) = 2.1850$ 。

又  $z|H_1 \sim N(1,1)$ ，故  $\gamma|H_1 \sim N(\frac{1}{2}, 1)$ 。

因而， $P_D = \int_{\mathcal{L}_{NP}}^{+\infty} p(\gamma|H_1) d\gamma = 1 - \Phi(\mathcal{L}_{NP} - \frac{1}{2}) = 1 - \Phi(0.2816) = 0.3891$ 。

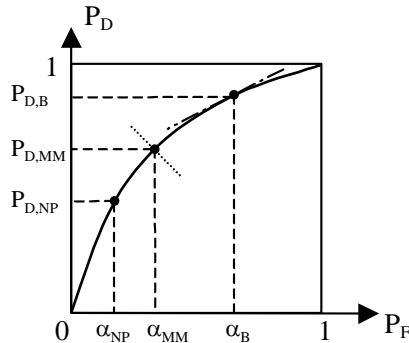
如此，可计算出其它几点  $(P_F, P_D)$  的坐标：(0.3, 0.6828)，(0.5, 0.8413)，(0.7, 0.9363)。

#解毕

#### ◆似然比检验工作特性：

Bayes 准则、极大极小准则、Neyman-Pearson 准则均可归结为似然比检验。

接收机工作特性 (ROC—Receiver Operating Characteristic)，它由  $P_F$ - $P_D$  曲线来描述 (见下图)。



接收机工作特性的几个重要性质：

① ROC 曲线上各点的斜率为似然比门限。

根据似然比检验准则  $\lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{>}} \lambda_T$ ，令  $\gamma = \lambda(\mathbf{z})$ ，我们有：

$$P_F = \int_{\lambda_T}^{+\infty} p(\gamma|H_0) d\gamma \quad \text{和} \quad P_D = \int_{\lambda_T}^{+\infty} p(\gamma|H_1) d\gamma$$

$$\text{则：} \quad \frac{dP_F}{d\lambda_T} = -p(\lambda_T|H_0) \quad \text{和} \quad \frac{dP_D}{d\lambda_T} = -p(\lambda_T|H_1)$$

$$\text{故：} \quad \frac{dP_D}{dP_F} = \frac{dP_D}{d\lambda_T} / \frac{dP_F}{d\lambda_T} = \frac{p(\lambda_T|H_1)}{p(\lambda_T|H_0)}$$

根据概率变换公式： $p(\gamma|H_i) |d\gamma| = p_z(\mathbf{z}|H_i) |d\mathbf{z}| \quad i=0,1$

所以,  $\frac{p(\gamma | H_1)}{p(\gamma | H_0)} = \frac{p_z(\mathbf{z} | H_1)}{p_z(\mathbf{z} | H_0)} = \lambda(\mathbf{z}) = \gamma$

当  $\gamma = \lambda_T$  时, 有:  $\frac{p(\lambda_T | H_1)}{p(\lambda_T | H_0)} = \lambda_T$

因此, ROC 曲线上各点的斜率  $\frac{dP_D}{dP_F}$  = 似然比门限  $\lambda_T$ 。

## ② 与 Bayes 准则的关系:

在 ROC 曲线上确定一个斜率为  $\lambda_T = \lambda_B$  的点, 其坐标值  $(P_F, P_{D,B})$  即为与此门限相对应的虚警概率和检测概率。

## ③ 与极大极小准则的关系:

由极大极小化方程  $C_{00}(1-P_F) + C_{10}P_F = C_{11}P_D + C_{01}(1-P_D)$  可得:

$$P_D = a - b \cdot P_F \quad \text{其中: } a = \frac{C_{01} - C_{00}}{C_{01} - C_{11}}, \quad b = \frac{C_{10} - C_{00}}{C_{01} - C_{11}}。$$

这是一条在  $P_F$ - $P_D$  平面上的截距为  $a$ 、斜率为  $b$  的直线 (如上图圆点线), 它与 ROC 曲线的交点  $(P_F, P_{D,MM})$  即为满足极大极小准则的解。

## ④ 与 Neyman-Pearson 准则的关系:

给定  $P_F = \alpha$ , ROC 曲线上对应的纵坐标  $P_{D,NP}$  即为满足 Neyman-Pearson 准则时的检测概率。

## ◆思考:

同前例, 并令  $A=0.5$ 、 $1.0$ 、 $2.0$ , 则 ROC 曲线的变化如何?

## 1.2.4 最小错误概率准则 (理想观测者准则) Minimum Probability of Error Criterion

在通信系统中, 我们常用 “最小错误概率准则” (或称为 “理想观测者准则”), 即: 使下面的平均错误概率  $P_e$  为最小:

$$P_e = P(D_1, H_0) + P(D_0, H_1) = P(H_0) P_F + P(H_1) P_M。$$

对比平均风险的表达式:

$$\bar{R} = C_{00} P(D_0, H_0) + C_{11} P(D_1, H_1) + C_{10} P(D_1, H_0) + C_{01} P(D_0, H_1)$$

可见: 最小错误概率准则是 Bayes 准则的特例。

故, 对比 Bayes 判决规则, 即可得最小错误概率判决规则为:

$$\lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{>}} \frac{P(H_0)}{P(H_1)}。$$

## 1.2.5 最大似然准则 Maximum Likelihood Criterion

判决规则:

$$\lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{>}} 1 \quad \text{或} \quad \frac{p(\mathbf{z} | H_1)}{p(\mathbf{z} | H_0)} \underset{H_0}{\overset{H_1}{>}} 1。$$

## 1.2.6 最大后验概率准则 Maximum A Posteriori Probability Criterion

在得到一个观测值  $\mathbf{z}$  之后, 我们也可以这样来作出一个合理的判断: 如果取假设  $H_1$  ( $H_0$ ) 的概率  $P(H_1 | \mathbf{z})$  ( $P(H_0 | \mathbf{z})$ ) 更高, 那么就确定假设  $H_1$  ( $H_0$ ) 成立。这就是 “最大后验概率准则” (MAP):

$$\frac{P(H_1 | \mathbf{z})}{P(H_0 | \mathbf{z})} \underset{H_0}{\overset{H_1}{>}} 1。$$

根据 Bayes 公式（后验概率公式）：

$$P(H_i | \mathbf{z}) \cdot p(\mathbf{z}) = p(\mathbf{z} | H_i) \cdot P(H_i), \quad i = 0, 1。$$

其中：  $p(\mathbf{z}) = p(\mathbf{z} | H_1) \cdot P(H_1) + p(\mathbf{z} | H_0) \cdot P(H_0)$ ，称为“全概率密度”。

故  $P(H_i | \mathbf{z}) = p(\mathbf{z} | H_i) \cdot P(H_i) / p(\mathbf{z})$ 。将之代入前面的判决准则，可得用似然比来表示的最大后验概率准则：

$$\lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{>}} \frac{P(H_0)}{P(H_1)}。$$

由此可见，最大后验概率准则有下面的特性：

- ① 与最小错误概率准则（理想观测者准则）等效；
- ② 与  $(C_{10}-C_{00})=(C_{01}-C_{11})$  时的 Bayes 准则等效；
- ③ 最大似然准则是最大后验概率准则在等先验概率（ $P(H_1)=P(H_0)$ ）时的特例。

### 1.2.7 各类准则小结（自习）

- ① Bayes 准则，基础；
- ② 各类准则的由来，推导思路，结论；
- ③ 各类准则间的关系；
- ④ 似然比检验，检验统计量；
- ⑤ 接收机工作特性（ROC）曲线；
- ⑥ 似然比门限及各类概率（ $P_F$ ,  $P_D$ ,  $P_M$  等）的计算。

## 1.3 在白高斯信道中的确知信号检测

### 1.3.1 连续信号的似然比检验及其最佳接收机

◆一次观测：

$$z(t) = s_i(t) + n(t), \quad i=0, 1$$

经一次取样后，可得：  $z = s_i + n$ ,  $n \sim N(0, \sigma_n^2)$ 。

$$p(z | H_i) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{1}{2\sigma_n^2} (z-s_i)^2\right)$$

$$\lambda(z) = \frac{p(z | H_1)}{p(z | H_0)} = \dots\dots$$

判决准则：  $\lambda(z) \underset{H_0}{\overset{H_1}{>}} \lambda_T$ 。根据不同的准则确定  $\lambda_T$ ，再用  $\lambda(z) = \lambda_T$  求出判决区域的划分。

◆多次观测：

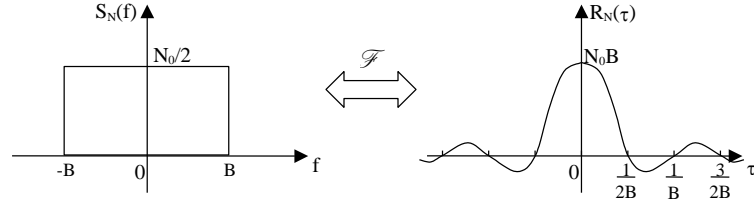
N 次观测可获得 N 个样值，记为：  $\mathbf{z} = (z_1, z_2, \dots, z_N)$ 。

现在，我们来讨论怎样计算  $\mathbf{z}$  的似然比：  $\lambda(\mathbf{z}) = \frac{p(\mathbf{z} | H_1)}{p(\mathbf{z} | H_0)}$ 。

若这  $N$  个样值独立，则：  $p(\mathbf{z} | H_i) = \prod_{k=1}^N p(z_k | H_i)$ 。

那么，什么条件下可获得独立的  $N$  个样值呢？

假设噪声为高斯带限白噪声，其频谱如下左图，相应的相关函数示于下右图。



设信号的观测区间为  $[0, T]$ ，则我们可得  $N = T / \frac{1}{2B} = 2BT$  个独立观测值（相关为 0 的高斯变量）。

又  $z_k = s_{ik} + n_k$ ， $i=0,1$ 。因为  $s_{ik}$  为确知信号，故  $z_{ik} \sim N(s_{ik}, \sigma_n^2)$ 。

从而，我们有这  $N$  个观测值的联合概率密度：

$$p(\mathbf{z} | H_i) = \prod_{k=1}^N p(z_k | H_i) = \left( \frac{1}{2\pi\sigma_n^2} \right)^{\frac{N}{2}} \prod_{k=1}^N \exp\left(-\frac{(z_k - s_{ik})^2}{2\sigma_n^2}\right)$$

因此， $N$  次观测时的似然比为：

$$\lambda(\mathbf{z}) = \frac{p(\mathbf{z} | H_1)}{p(\mathbf{z} | H_0)} = \exp\left(\sum_{k=1}^N \left( -\frac{(z_k - s_{1k})^2}{2\sigma_n^2} + \frac{(z_k - s_{0k})^2}{2\sigma_n^2} \right)\right) \quad (*)$$

下面，我们来讨论检验统计量的概念和  $N$  次取样的意义：

若在观测时间  $T$  内，信号为恒定值，即：  $s_{1k} = s_1$ ， $s_{0k} = s_0$ 。那么，该似然比可重写为：

$$\lambda(\mathbf{z}) = \exp\left(\frac{(s_1 - s_0)}{\sigma_n^2} \sum_{k=1}^N z_k - N \frac{(s_1^2 - s_0^2)}{2\sigma_n^2}\right)$$

利用对数形式，则有：

$$\mathcal{L}(\mathbf{z}) = \frac{(s_1 - s_0)}{\sigma_n^2} \sum_{k=1}^N z_k - N \frac{(s_1^2 - s_0^2)}{2\sigma_n^2}$$

根据对数似然比检验判决规则：  $\mathcal{L}(\mathbf{z}) \underset{H_0}{\overset{H_1}{>}} \mathcal{L}_T$ ，我们可得下述判决不等式：

$$\frac{1}{N} \sum_{k=1}^N z_k \underset{H_0}{\overset{H_1}{>}} \frac{\sigma_n^2/N}{(s_1 - s_0)} \mathcal{L}_T + \frac{(s_1 + s_0)}{2}$$

✓ 若令检验统计量  $y = \frac{1}{N} \sum_{k=1}^N z_k$ ，且令门限  $y_T = \frac{\sigma_n^2/N}{(s_1 - s_0)} \mathcal{L}_T + \frac{(s_1 + s_0)}{2}$

则上述判决规则可表示为：  $y \underset{H_0}{\overset{H_1}{>}} y_T$

思考：当  $N \uparrow$  时， $P_F$  和  $P_D$  的变化怎样？

#### ◆连续信号似然比的计算：

方法：利用多次观测的似然比计算公式，先对连续信号作  $N$  次抽样，再令  $N \rightarrow \infty$ ，从而得到连续情况下的似然比。

将上面  $N$  次观测时的似然比计算公式 (\*) 重新整理，可得：

$$\lambda(\mathbf{z}) = \exp\left(\sum_{k=1}^N \left( \frac{s_{1k} z_k}{\sigma_n^2} - \frac{s_{0k} z_k}{\sigma_n^2} - \frac{s_{1k}^2}{2\sigma_n^2} + \frac{s_{0k}^2}{2\sigma_n^2} \right)\right)$$

当  $N \rightarrow \infty$  时，抽样时间间隔  $\Delta t \rightarrow 0$ ，又  $\Delta t = \frac{1}{2B}$ ，故此时等效于  $B \rightarrow \infty$ （理想信道）。

又噪声功率  $\sigma_n^2 = N_0 B = \frac{N_0}{2\Delta t}$ 。

$$\text{从而 } \lim_{\substack{N \rightarrow \infty \\ \Delta t \rightarrow 0}} \sum_{k=1}^N \frac{s_{ik} z_k}{\sigma_n^2} = \lim_{\substack{N \rightarrow \infty \\ \Delta t \rightarrow 0}} \sum_{k=1}^N \frac{2}{N_0} s_{ik} z_k \Delta t = \frac{2}{N_0} \int_0^T s_i(t) z(t) dt, \quad i=0,1。$$

$$\text{同理：} \lim_{\substack{N \rightarrow \infty \\ \Delta t \rightarrow 0}} \sum_{k=1}^N \frac{s_{ik}^2}{2\sigma_n^2} = \lim_{\substack{N \rightarrow \infty \\ \Delta t \rightarrow 0}} \sum_{k=1}^N \frac{1}{N_0} s_{ik}^2 \Delta t = \frac{1}{N_0} \int_0^T s_i^2(t) dt, \quad i=0,1。$$

所以，我们导出连续信号的似然比为：

$$\lambda(z(t)) = \exp\left(\frac{2}{N_0} \left( \int_0^T s_1(t) z(t) dt - \int_0^T s_0(t) z(t) dt \right) - \frac{1}{N_0} \left( \int_0^T s_1^2(t) dt - \int_0^T s_0^2(t) dt \right)\right)$$

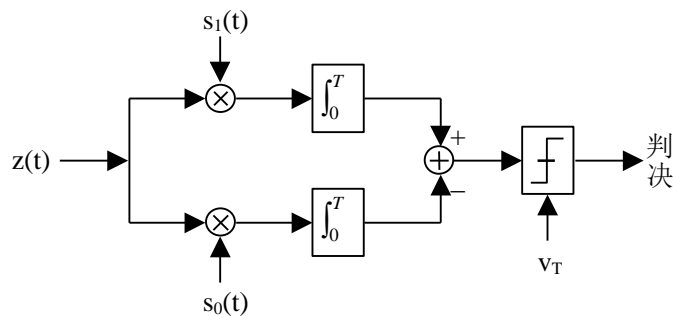
#### ◆最佳接收机：

定义：信号  $s_i(t)$  的能量为  $E_i = \int_0^T s_i^2(t) dt$ 。

由对数似然比检验公式  $\mathcal{L}(z(t)) \underset{H_0}{\overset{H_1}{>}} \mathcal{L}_T$  ( $\mathcal{L}_T$  由判决准则确定) 可得连续信号时的判决不等式：

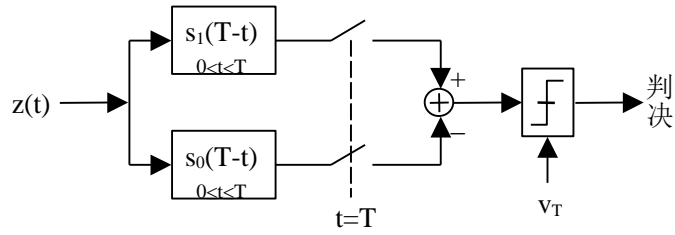
$$\int_0^T s_1(t) z(t) dt - \int_0^T s_0(t) z(t) dt \underset{H_0}{\overset{H_1}{>}} v_T, \quad \text{其中：} v_T = \frac{N_0}{2} \mathcal{L}_T + \frac{1}{2} (E_1 - E_0)$$

由此判决不等式，我们可设计出如下的最佳接收机：



此接收机又称为“相关接收机”。我们亦可以另一种形式来实现它，即所谓的“匹配滤波器”：





◆接收机的工作性能:

令检验统计量  $v = \int_0^T s_1(t)z(t)dt - \int_0^T s_0(t)z(t)dt$ 。

在通讯系统中，我们常使用最小平均错误概率准则。此时，它的判决规则为：

$$v \underset{H_0}{\overset{H_1}{>}} v_T, \quad \text{其中: } v_T = \frac{N_0}{2} \mathcal{L}_T + \frac{1}{2} (E_1 - E_0), \quad \text{而 } \lambda_T = \frac{P(H_0)}{P(H_1)}。$$

下面我们来求虚报概率  $P_F$  和漏报概率  $P_M$ 。

分析:  $v = \int_0^T [s_1(t) - s_0(t)] z(t) dt$ ,

而  $z(t) | H_i = s_i(t) + n(t)$ ,

其中:  $s_i(t)$  为确知信号,  $E(n(t)) = 0$ ,  $E(n(t_1)n(t_2)) = \frac{N_0}{2} \delta(t_1 - t_2)$ 。

由于高斯变量通过线性系统（积分系统）仍为高斯分布，故  $p(v|H_i)$  应为高斯密度函数。

定义: 信号平均能量  $E = \frac{1}{2} (E_1 + E_0)$ ;

信号互相关系数  $\rho = \frac{1}{E} \int_0^T s_1(t)s_0(t)dt$ 。注:  $|\rho| \leq 1$ 。

$H_0$ :  $E(v|H_0) = \rho E - E_0$ 。

$$\begin{aligned} E([v - E(v)]^2 | H_0) &= E\left(\left[\int_0^T (s_1(t) - s_0(t))(z(t) - s_0(t)) dt\right]^2\right) \\ &= E\left(\left[\int_0^T (s_1(t) - s_0(t)) n(t) dt\right]^2\right) \\ &= \int_0^T \int_0^T (s_1(t_1) - s_0(t_1))(s_1(t_2) - s_0(t_2)) E[n(t_1)n(t_2)] dt_1 dt_2 \\ &= \int_0^T \int_0^T (s_1(t_1) - s_0(t_1))(s_1(t_2) - s_0(t_2)) \frac{N_0}{2} \delta(t_1 - t_2) dt_1 dt_2 \\ &= \frac{N_0}{2} \int_0^T (s_1(t) - s_0(t))^2 dt \\ &= N_0 E(1 - \rho)。 \end{aligned}$$

$H_1$ :  $E(v|H_1) = E_1 - \rho E$ 。

$$E([v - E(v)]^2 | H_1) = N_0 E(1 - \rho)。$$

这样，我们可写出条件概率密度如下：

$$p(v|H_0) = \frac{1}{\sqrt{2\pi} \sqrt{N_0 E(1 - \rho)}} \exp\left(-\frac{[v - (\rho E - E_0)]^2}{2N_0 E(1 - \rho)}\right)$$

$$p(v|H_1) = \frac{1}{\sqrt{2\pi}\sqrt{N_0 E(1-\rho)}} \exp\left(-\frac{[v - (E_1 - \rho E)]^2}{2N_0 E(1-\rho)}\right)$$

所以，虚报概率和漏报概率可计算如下：

$$P_F = \int_{v_T}^{+\infty} p(v|H_0) dv$$

$$\text{作变量代换：令 } y = \frac{v - (\rho E - E_0)}{\sqrt{N_0 E(1-\rho)}} \text{ 及 } y_T = \frac{v_T - (\rho E - E_0)}{\sqrt{N_0 E(1-\rho)}} = \frac{\frac{N_0}{2} \mathcal{L}_T + E(1-\rho)}{\sqrt{N_0 E(1-\rho)}}$$

$$= \int_{y_T}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

$$= 1 - \Phi(y_T)$$

$$P_M = \int_{-\infty}^{v_T} p(v|H_1) dv$$

$$\text{作变量代换：令 } y = \frac{v - (E_1 - \rho E)}{\sqrt{N_0 E(1-\rho)}} \text{ 及 } y_T' = \frac{v_T - (E_1 - \rho E)}{\sqrt{N_0 E(1-\rho)}} = \frac{\frac{N_0}{2} \mathcal{L}_T - E(1-\rho)}{\sqrt{N_0 E(1-\rho)}}$$

$$= \int_{-\infty}^{y_T'} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

$$= \Phi(y_T')$$

### 1.3.2 二元数字通信系统

在通信系统中，常使用最小平均错误概率准则。

平均错误概率可表示为：  $P_e = P(H_0)P_F + P(H_1)P_M$

在等先验概率情况下，  $\mathcal{L}_T = 0 \Rightarrow y_T = -y_T' \Rightarrow \Phi(y_T') = 1 - \Phi(y_T)$

$$\Rightarrow P_e = P_F = P_M = 1 - \Phi\left(\sqrt{\frac{E(1-\rho)}{N_0}}\right) \quad (**)$$

◆讨论：

在设计通信系统时，常使用“最小平均错误概率准则”。

从上面的平均错误概率表达式 (\*\*):  $P_e = 1 - \Phi(x)$ ，其中：  $x = \sqrt{\frac{E(1-\rho)}{N_0}}$

我们可知： $P_e$ 的取值与三个因素有关，即  $E$ 、 $N_0$  和  $\rho$ 。

又由于  $x \uparrow \Rightarrow \Phi(x) \uparrow$ ，所以当信噪比  $E/N_0$  一定时，我们有： $\rho \downarrow \Rightarrow P_e \downarrow$ 。

而  $|\rho| \leq 1$ ，故：当  $\rho = -1$ （即  $s_0(t) = -s_1(t)$ ）时， $P_e$  取得最小值。

下面我们讨论几种实际通信系统：

① 相干移相键控 (CPSK)：

$$s_0(t) = A \sin \omega_0 t$$

$$s_1(t) = -A \sin \omega_0 t$$

其中： $0 \leq t \leq T$ 。

$$P_e = 1 - \Phi\left(\sqrt{\frac{2E}{N_0}}\right)。$$

判决规则可简化为：  $\int_0^T s_1(t)z(t)dt \underset{H_0}{\overset{H_1}{>}} 0$ ，此时的相关接收机可用单个相关器来实现。

② 相干移频键控 (CFSK)：

$$\begin{aligned} s_0(t) &= A \sin \omega_0 t \\ s_1(t) &= A \sin \omega_1 t \end{aligned} \quad \text{其中： } 0 \leq t \leq T。$$

若使两个频率  $\omega_0$  和  $\omega_1$  满足下述关系：

$$\begin{aligned} \omega_1 - \omega_0 &= n\pi/T \\ \omega_1 + \omega_0 &= m\pi/T \end{aligned} \quad \text{其中： } n \text{ 和 } m \text{ 是整数，}$$

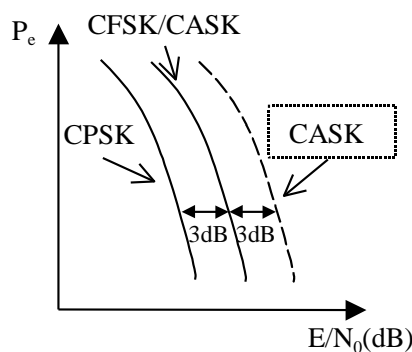
则：  $P_e = 1 - \Phi\left(\sqrt{\frac{E}{N_0}}\right)$ 。且相关接收机的判决门限  $v_T = 0$ 。

③ 相干移幅键控 (CASK) 或启闭键控 (OOK)：

$$\begin{aligned} s_0(t) &= 0 \\ s_1(t) &= A \sin \omega_1 t \end{aligned} \quad \text{其中： } 0 \leq t \leq T。$$

则：  $P_e = 1 - \Phi\left(\sqrt{\frac{E_1}{2N_0}}\right) = 1 - \Phi\left(\sqrt{\frac{E}{N_0}}\right)$ ，且相关接收机的判决门限  $v_T = (1/2)E_1$ 。

三种接收机的性能比较：



### 1.3.3 雷达系统

$$\text{雷达信号：} \begin{cases} s_0(t) = 0 \\ s_1(t) = A \sin \omega_0(t+\tau) \end{cases} \quad \text{其中： } 0 \leq t \leq T。$$

因而可知：  $\rho=0$ ，  $E_1 = \int_0^T s_1^2(t)dt = (1/2)A^2T = 2E$  以及 检验统计量  $v = \int_0^T s_1(t)z(t)dt$ 。

在雷达系统中，我们常常采用 Neyman-Pearson 准则：  $v \underset{H_0}{\overset{H_1}{>}} v_{NP}$ 。

$$\text{由上节的推导，我们有：} \begin{cases} v|H_0 \sim N(0, N_0 E_1/2) \\ v|H_1 \sim N(E_1, N_0 E_1/2) \end{cases}$$

根据 N-P 准则，虚警概率可计算如下：

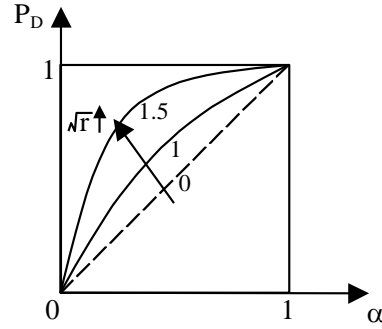
$$P_F = \int_{v_{NP}}^{+\infty} p(v|H_0)dv = 1 - \Phi(y_{NP}), \quad \text{其中： } y_{NP} = \frac{v_{NP}}{\sqrt{N_0 E_1/2}}。$$

将已知条件  $P_F = \alpha$  代入上式，我们即可确定  $y_{NP}$  及  $v_{NP}$  之值。

又，检测概率可计算如下：

$$P_D = \int_{V_{NP}}^{+\infty} p(v|H_1)dv = 1 - \Phi(x_{NP}), \text{ 其中: } x_{NP} = y_{NP} - \sqrt{\frac{2E_1}{N_0}}。$$

令信噪比  $r=2E_1/N_0$ ，则可作出  $P_D \sim \alpha$  曲线：



## 1.4 多元假设检验 (Multiple-hypothesis Testing)

“M 元择一”：需要将观测空间  $\mathcal{Z}$  划分为 M 个子空间。

M 元假设下的平均代价可计算如下：

$$\begin{aligned} \bar{R} &= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij} P(H_j) P(D_i|H_j) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij} P(H_j) \int_{\mathcal{Z}_i} p(\mathbf{z}|H_j) d\mathbf{z} \\ &= \sum_{i=0}^{M-1} C_{ii} P(H_i) \int_{\mathcal{Z}_i} p(\mathbf{z}|H_i) d\mathbf{z} + \sum_{i=0}^{M-1} \sum_{\substack{j=0 \\ j \neq i}}^{M-1} C_{ij} P(H_j) \int_{\mathcal{Z}_i} p(\mathbf{z}|H_j) d\mathbf{z} \\ &= \sum_{i=0}^{M-1} C_{ii} P(H_i) - \sum_{i=0}^{M-1} C_{ii} P(H_i) \sum_{\substack{j=0 \\ j \neq i}}^{M-1} \int_{\mathcal{Z}_j} p(\mathbf{z}|H_i) d\mathbf{z} + \sum_{i=0}^{M-1} \sum_{\substack{j=0 \\ j \neq i}}^{M-1} C_{ij} P(H_j) \int_{\mathcal{Z}_i} p(\mathbf{z}|H_j) d\mathbf{z} \\ &= \sum_{i=0}^{M-1} C_{ii} P(H_i) - \sum_{j=0}^{M-1} \sum_{\substack{i=0 \\ i \neq j}}^{M-1} C_{ii} P(H_i) \int_{\mathcal{Z}_j} p(\mathbf{z}|H_i) d\mathbf{z} + \sum_{i=0}^{M-1} \sum_{\substack{j=0 \\ j \neq i}}^{M-1} C_{ij} P(H_j) \int_{\mathcal{Z}_i} p(\mathbf{z}|H_j) d\mathbf{z} \\ &= \sum_{i=0}^{M-1} C_{ii} P(H_i) + \sum_{i=0}^{M-1} \int_{\mathcal{Z}_i} \left[ \sum_{\substack{j=0 \\ j \neq i}}^{M-1} (C_{ij} - C_{ji}) P(H_j) p(\mathbf{z}|H_j) \right] d\mathbf{z} \end{aligned}$$

显然，上式中第一项为固定代价，而第二项随观测空间  $\mathcal{Z}$  的划分而变化。

$$\text{令: } I_i(\mathbf{z}) = \sum_{\substack{j=0 \\ j \neq i}}^{M-1} (C_{ij} - C_{ji}) P(H_j) p(\mathbf{z}|H_j)$$

由于  $I_i(\mathbf{z}) > 0$ ，若要使平均代价  $\bar{R}$  最小，那么我们应按如下规则来划分观测空间：

首先，对每一个  $\mathbf{z}$  值，计算  $I_i(\mathbf{z})$ ,  $i=0, \dots, M-1$ ；

然后，将此  $\mathbf{z}$  值划入使  $I_i(\mathbf{z})$  值最小的子空间内。

也就是说，判决规则为：

假设  $H_i$  ( $i=0, \dots, M-1$ ) 成立，若  $I_i(\mathbf{z}) = \min\{I_k(\mathbf{z}) | k=0, \dots, M-1\}$ 。

在通信系统中，我们通常采用最小平均错误概率准则。

此时， $C_{ij} = 1 - \delta_{ij}$ ，其中  $\delta_{ij}$  为 Kronecker  $\delta$ -函数。

将此代价  $C_{ij}$  代入  $I_i(\mathbf{z})$  的表达式可得：

$$I_i(\mathbf{z}) = \sum_{\substack{j=0 \\ j \neq i}}^{M-1} P(H_j) p(\mathbf{z}|H_j) = \sum_{\substack{j=0 \\ j \neq i}}^{M-1} P(H_j|\mathbf{z}) p(\mathbf{z}) = [1 - P(H_i|\mathbf{z})] p(\mathbf{z})$$

故， $I_i(\mathbf{z})$ 取最小值等效于  $P(H_i|\mathbf{z})$ 取最大值（即等同于“最大后验概率准则”）。

在等先验概率情况下： $P(H_j)=P$ ，则： $I_i(\mathbf{z}) = \sum_{\substack{j=0 \\ j \neq i}}^{M-1} p(\mathbf{z}|H_j) P = p(\mathbf{z}) - p(\mathbf{z}|H_i) P$

故， $I_i(\mathbf{z})$ 取最小值等效于  $p(\mathbf{z}|H_i)$ 取最大值（即等同于“最大似然准则”）。

◆例：

三元假设问题。

假设  $H_i$ ：接收信号  $\sim N(m_i, \sigma^2)$ ，其中： $m_1=1$ ， $m_2=2$ ， $m_3=-1$ 。

等先验概率，最小平均错误概率准则，作  $N$  次独立观测。

求：判决规则表达式，检验统计量表达式及相应的判决域。

解：由题意：

$$\begin{aligned} p(\mathbf{z}|H_i) &= \left( \frac{1}{2\pi\sigma^2} \right)^{N/2} \exp\left( -\frac{\sum_{j=1}^N (z_j - m_i)^2}{2\sigma^2} \right), \quad i=1,2,3 \\ &= \left( \frac{1}{2\pi\sigma^2} \right)^{N/2} \exp\left( -\frac{\sum_{j=1}^N z_j^2 - 2\left(\sum_{j=1}^N z_j\right) \cdot m_i + N \cdot m_i^2}{2\sigma^2} \right), \quad i=1,2,3 \end{aligned}$$

在  $p(\mathbf{z}|H_i)$ 的表达式中，与假设  $H_i$ 有关的是指数上的最后两项。

由此可见，依据最大似然准则，若判决假设  $H_i$ 成立，则：

$$\max p(\mathbf{z}|H_i) \Rightarrow \max [2\left(\sum_{j=1}^N z_j\right)m_i - Nm_i^2], \quad \text{即} \quad \max \left[ \left(\frac{1}{N} \sum_{j=1}^N z_j\right)m_i - \frac{1}{2} m_i^2 \right].$$

令  $Q_i(\mathbf{z}) = \left(\frac{1}{N} \sum_{j=1}^N z_j\right)m_i - \frac{1}{2} m_i^2$ ，则判决规则为：

假设  $H_i$ 成立，若  $Q_i(\mathbf{z})$ 取得最大值。

又令检验统计量  $y = \frac{1}{N} \sum_{j=1}^N z_j$ ，则对于不同的假设，我们有：

$$Q_1(\mathbf{z}) = Q_1(y) = y - \frac{1}{2}, \quad Q_2(\mathbf{z}) = Q_2(y) = 2y - 2, \quad Q_3(\mathbf{z}) = Q_3(y) = -y - \frac{1}{2}.$$

故，判决域的划分应为：

$$\begin{aligned} \mathcal{Z}_1: \begin{cases} Q_1(y) \geq Q_2(y) \\ Q_1(y) \geq Q_3(y) \end{cases} &\Rightarrow \begin{cases} y \leq \frac{3}{2} \\ y \geq 0 \end{cases} \Rightarrow 0 \leq y \leq \frac{3}{2}; \\ \mathcal{Z}_2: \begin{cases} Q_2(y) > Q_1(y) \\ Q_2(y) > Q_3(y) \end{cases} &\Rightarrow \begin{cases} y > \frac{3}{2} \\ y > \frac{1}{2} \end{cases} \Rightarrow y > \frac{3}{2}; \\ \mathcal{Z}_3: \begin{cases} Q_3(y) > Q_1(y) \\ Q_3(y) > Q_2(y) \end{cases} &\Rightarrow \begin{cases} y < 0 \\ y < \frac{1}{2} \end{cases} \Rightarrow y < 0. \end{aligned}$$

## 1.5 复合假设检验与随机参量信号检测

### 1.5.1 复合假设检验 (Composite Hypothesis Testing)

考虑这样的情况：雷达系统中，接收到的目标反射信号，其幅度、相位和时延可能是随机的；通信系统中，接收到的多径信号，其幅度、相位和频率可能是随机的。这样，在描述信号统计特性的概率密度函数（似然函数）中就含有未知的参量——“复合假设”（与前面的简单假设不同）。

为简单明了起见，我们下面以二元假设为例来讨论，其结论可推广至多元假设情况下。

令：与假设有关的参量为 $\theta$ ，其取值空间为 $\Theta$ 。

根据未知参量的性质，我们分以下三种情况来讨论：

(1) 概率密度函数已知的随机参量：

- ① 给定 $\theta$ 时，求出条件似然函数；
- ② 将条件似然函数对参量平均，求出似然函数；
- ③ 按以前的简单假设检验来求解。

(2) 概率密度函数未知的随机参量：

- ① 找出未知参量的最不利密度函数： $p(\theta)$ ；
- ② 按上面(1)中的已知概率密度函数的随机参量来求解。

(3) 未知的常数：

采用 Neyman-Pearson 准则：假定 $\theta$ 值，寻求  $P_F=\alpha$  时使  $P_D$  最大的检验。

根据具体情况，我们可采用下面两种办法：

- ① 若  $P_D$  最大时与 $\theta$ 值无关：  
这时可实现最佳检验，称为“一致最大势（Uniformly Most Powerful - UMP）检验”。
- ② 若一致最大势检验不存在时：

广义似然比检验（generalized likelihood ratio）：

$$\lambda_g(\mathbf{z}) \underset{H_0}{\overset{H_1}{>}} \lambda_T \quad \text{其中：广义似然比 } \lambda_g(\mathbf{z}) = \frac{\max_{\theta=\theta_1} p(\mathbf{z} | H_1, \theta)}{\max_{\theta=\theta_0} p(\mathbf{z} | H_0, \theta)}。$$

◆例：

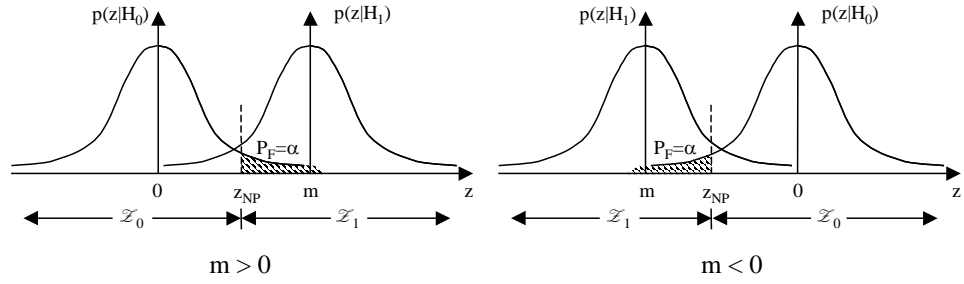
假设  $H_0$  下，观测值服从零均值、方差为 $\sigma^2$ 的高斯分布；

假设  $H_1$  下，观测值服从方差为 $\sigma^2$ 的高斯分布，但均值  $m$  为未知常数。

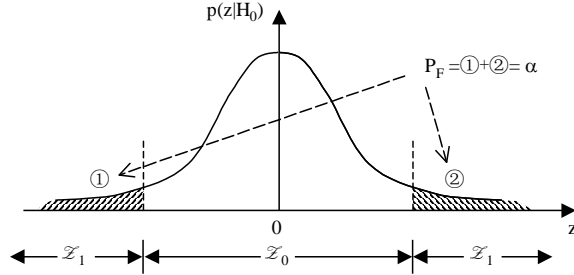
➤ 一次观测：

若已知  $m>0$ ，则 UMP 检验存在（参见下左图）；

若已知  $m<0$ ，则 UMP 检验亦存在（参见下右图）；



若  $m$  可正可负，则 UMP 检验不存在。此时，可采用双边检验（two-sided test）：



➤ 多次观测：

根据  $N$  次独立观测值，我们可得  $m$  的最大似然估值为： $\hat{m}_{ML} = \frac{1}{N} \sum_{j=1}^N z_j$ 。

故可采用广义似然比检验： $\frac{p(\mathbf{z}|H_1, \hat{m}_{ML})}{p(\mathbf{z}|H_0)} \underset{H_0}{\overset{H_1}{>}} \lambda_T$

经过运算，最后我们得到双边检验： $\left| \frac{1}{\sqrt{N}} \sum_{j=1}^N z_j \right| \underset{H_0}{\overset{H_1}{>}} \eta$ ，其中： $\eta = \sqrt{2\sigma^2 \ln \lambda_T}$ 。

### 1.5.2 在白高斯信道中随机幅度信号检测

假设  $H_0$ :  $s_0(t) = 0$ ;

假设  $H_1$ :  $s_1(t) = A$ ，其中： $A \sim N(0, \sigma_A^2)$ 。

高斯信道，加性噪声  $n(t)$ ， $N(0, \sigma_N^2)$ 。  $A$  与  $n(t)$  彼此相互独立。

一次观测时，似然函数为：

$$p(z|H_0) = \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{z^2}{2\sigma_N^2}\right)$$

$$p(z|H_1) = \int_{-\infty}^{+\infty} p(z|H_1, A) p(A) dA = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{(z-A)^2}{2\sigma_N^2}\right) \cdot \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left(-\frac{A^2}{2\sigma_A^2}\right) dA$$

从而，可得似然比：

$$\begin{aligned} \lambda(z) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left(-\frac{(\sigma_A^2 + \sigma_N^2)A^2 - 2\sigma_A^2 zA}{2\sigma_A^2 \sigma_N^2}\right) dA \\ &= \frac{\sigma_N}{\sqrt{\sigma_A^2 + \sigma_N^2}} \exp\left(\frac{\sigma_A^2}{2\sigma_N^2(\sigma_A^2 + \sigma_N^2)} z^2\right) \end{aligned}$$

根据 Bayes 准则： $\ln\lambda(z) \underset{H_0}{\overset{H_1}{>}} \ln\lambda_B$ ,

$$\text{经整理可得: } z^2 \underset{H_0}{\overset{H_1}{>}} \frac{2\sigma_N^2(\sigma_A^2 + \sigma_N^2)}{\sigma_A^2} \left[ \ln \lambda_B + \frac{1}{2} \ln \left( 1 + \frac{\sigma_A^2}{\sigma_N^2} \right) \right]$$

### 1.5.3 在白高斯信道中随机相位信号检测

在实际应用中，我们会遇到具有随机相位的信号检测问题。如：通信系统中的非相干 OOK 信号的检测，雷达系统中表明目标是否存在的反射信号的检测等。在这些情况中，我们可作如下的假设：

$H_0$ :  $z(t) = n(t)$ ,  $0 \leq t \leq T$ , 其中： $n(t)$ 是均值为零、谱密度为 $\frac{N_0}{2}$  的白高斯噪声；

$H_1$ :  $z(t) = \sqrt{\frac{2E_s}{T}} \sin(\omega_c t + \theta) + n(t)$ ,  $0 \leq t \leq T$ , 其中： $\theta$ 是随机变量， $E_s$ 是该信号的能量。

由 1.3.1 节中的公式：

$$\lambda(z(t)) = \exp \left( \frac{2}{N_0} \left( \int_0^T s_1(t) z(t) dt - \int_0^T s_0(t) z(t) dt \right) - \frac{1}{N_0} \left( \int_0^T s_1^2(t) dt - \int_0^T s_0^2(t) dt \right) \right)$$

故，对给定的 $\theta$ 值，此时的条件似然比为：

$$\begin{aligned} \lambda(z(t)|\theta) &= \exp \left( \frac{2}{N_0} \int_0^T z(t) \sqrt{\frac{2E_s}{T}} \sin(\omega_c t + \theta) dt - \frac{E_s}{N_0} \right) \\ &= \exp \left( \frac{2}{N_0} \sqrt{\frac{2E_s}{T}} \int_0^T z(t) (\sin \omega_c t \cos \theta + \cos \omega_c t \sin \theta) dt - \frac{E_s}{N_0} \right) \end{aligned}$$

令：同相分量  $y_c = \int_0^T z(t) \cos \omega_c t dt$ ，正交分量  $y_s = \int_0^T z(t) \sin \omega_c t dt$ ，

则有：

$$\lambda(z(t)|\theta) = \exp \left( \frac{2}{N_0} \sqrt{\frac{2E_s}{T}} (y_c \sin \theta + y_s \cos \theta) - \frac{E_s}{N_0} \right)$$

假定随机相位在 $[0, 2\pi]$ 区间上是均匀分布（最不利分布）的，即： $p(\theta) = \frac{1}{2\pi}$ ,  $0 \leq \theta \leq 2\pi$

那么，我们可得似然比为：

$$\begin{aligned} \lambda(z(t)) &= \int_0^{2\pi} \lambda(z(t)|\theta) p(\theta) d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} \exp \left( \frac{2}{N_0} \sqrt{\frac{2E_s}{T}} (y_c \sin \theta + y_s \cos \theta) - \frac{E_s}{N_0} \right) d\theta \\ &= \exp \left( -\frac{E_s}{N_0} \right) \cdot I_0 \left( \frac{2}{N_0} \sqrt{\frac{2E_s}{T}} \sqrt{y_c^2 + y_s^2} \right) \quad (***) \end{aligned}$$

其中： $I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} \exp(x \cos \phi) d\phi$ ，称为“零阶修正 Bessel 函数”。

令： $y = \sqrt{y_c^2 + y_s^2}$ ，根据似然比判决规则： $\lambda(z(t)) \underset{H_0}{\overset{H_1}{>}} \lambda_T$ ，可得：

$$I_0 \left( \frac{2}{N_0} \sqrt{\frac{2E_s}{T}} y \right) \underset{H_0}{\overset{H_1}{>}} \lambda_T \exp \left( \frac{E_s}{N_0} \right)。$$

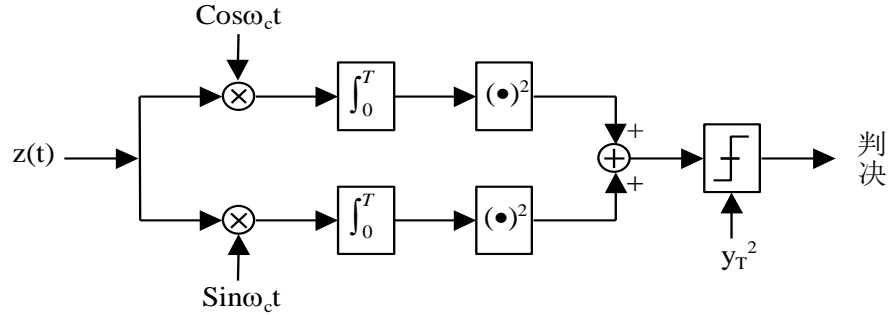


又由于  $I_0(x)$  是  $x$  的单调增函数（小的  $x$ ,  $\ln I_0(x) \approx x^2/4$ ; 大的  $x$ ,  $\ln I_0(x) \approx x$ ），故上述判决规则可改写为：

$$y \underset{H_0}{\overset{H_1}{>}} y_T, \quad \text{其中: } y_T \text{ 由方程 } I_0\left(\frac{2}{N_0} \sqrt{\frac{2E_s}{T}} y_T\right) = \lambda_T \exp\left(\frac{E_s}{N_0}\right) \text{ 确定。}$$

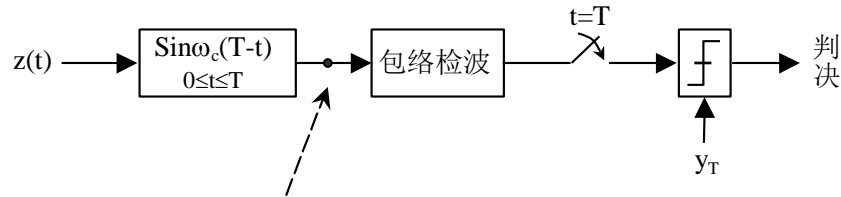
因此，我们有下述的最佳接收机实现形式：

➤ 正交接收机：



等效地，我们亦可采用匹配滤波器的形式来实现正交接收机。

➤ 非相干匹配滤波器：



$$\int_0^T z(\tau) \sin \omega_c(T-t+\tau) d\tau = \int_0^T z(\tau) \sin \omega_c \tau d\tau \cos \omega_c(T-t) + \int_0^T z(\tau) \cos \omega_c \tau d\tau \sin \omega_c(T-t)$$

接收机工作特性（ROC）分析：

➤ 首先，需要计算  $y$  的概率密度函数  $p(y|H_0)$  和  $p(y|H_1)$ 。

$$\text{由定义: } y = \sqrt{y_c^2 + y_s^2}, \quad y_c = \int_0^T z(t) \cos \omega_c t dt, \quad y_s = \int_0^T z(t) \sin \omega_c t dt.$$

当给定  $\theta$  值时， $z(t)$  为高斯过程，故  $y_c$  和  $y_s$  都是高斯变量，且两者互不相关，因为：

$$\begin{aligned} E[(y_c - E(y_c|\theta))(y_s - E(y_s|\theta))] &= E\left(\int_0^T \int_0^T n(t_1)n(t_2) \cos \omega_c t_1 \sin \omega_c t_2 dt_1 dt_2\right) \\ &= \int_0^T \int_0^T \frac{N_0}{2} \delta(t_1 - t_2) \cos \omega_c t_1 \sin \omega_c t_2 dt_1 dt_2 = \frac{N_0}{2} \int_0^T \cos \omega_c t \sin \omega_c t dt = 0. \end{aligned}$$

所以， $y_c$  和  $y_s$  相互独立。

在假设  $H_1$  下：

$$E(y_c|\theta) = E\left(\int_0^T [s(t)+n(t)] \cos \omega_c t dt\right) = \sqrt{\frac{2E_s}{T}} \int_0^T \sin(\omega_c t + \theta) \cos \omega_c t dt = \sqrt{\frac{E_s T}{2}} \sin \theta$$

$$E(y_s|\theta) = \sqrt{\frac{E_s T}{2}} \cos \theta$$

$$\text{Var}(y_c|\theta) = E[(y_c - E(y_c|\theta))^2] = E\left[\left(\int_0^T n(t) \cos \omega_c t dt\right)^2\right] = \frac{N_0}{2} \int_0^T \cos^2 \omega_c t dt = \frac{N_0 T}{4}$$

$$\text{Var}(y_s|\theta) = E([y_s - E(y_s|\theta)]^2) = E([\int_0^T n(t) \sin \omega_c t dt]^2) = \frac{N_0}{2} \int_0^T \sin^2 \omega_c t dt = \frac{N_0 T}{4}$$

令  $\sqrt{\frac{E_s T}{2}} = c$  且  $\frac{N_0 T}{4} = \sigma^2$ , 则  $y_c$  和  $y_s$  的联合分布为:

$$p(y_c, y_s | H_1, \theta) \equiv f(y_c, y_s) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} [(y_c - c \sin \theta)^2 + (y_s - c \cos \theta)^2]\right)$$

由随机变量代换 ( $y_c/y_s \leftrightarrow y/\theta_0$ ):  $\begin{cases} y_c = y \sin \theta_0 \\ y_s = y \cos \theta_0 \end{cases}$

其对应的 Jacobi 行列式为:  $J = -y$ , 从而可得  $y$  和  $\theta_0$  的联合条件概率密度:

$$p(y, \theta_0 | H_1, \theta) = |J| f(y \sin \theta_0, y \cos \theta_0) = y \cdot \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} [y^2 + c^2 - 2yc \cos(\theta - \theta_0)]\right)$$

通过联合分布求边际分布, 可得  $y$  的条件概率密度为:

$$p(y | H_1, \theta) = \int_0^{2\pi} p(y, \theta_0 | H_1, \theta) d\theta_0 = \frac{y}{\sigma^2} \exp\left(-\frac{1}{2\sigma^2} [y^2 + c^2]\right) \cdot I_0\left(\frac{yc}{\sigma^2}\right)$$

由于  $p(y | H_1, \theta)$  与参量  $\theta$  的取值无关, 故代入  $c$  的值, 我们有:

$$p(y | H_1) = \frac{y}{\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \left[y^2 + \frac{E_s T}{2}\right]\right) \cdot I_0\left(\frac{y}{\sigma^2} \sqrt{\frac{E_s T}{2}}\right), \quad y > 0, \quad \text{其中: } \sigma^2 = \frac{N_0 T}{4}.$$

此分布称为 “Rice 分布” 或 “广义 Rayleigh 分布”。

在假设  $H_0$  下:

此时等同于在假设  $H_1$  下信号能量  $E_s = 0$  的特殊情况, 即:

$$p(y | H_0) = p(y | H_1)_{|E_s=0} = \frac{y}{\sigma^2} \exp\left(-\frac{y^2}{2\sigma^2}\right), \quad y > 0$$

此概率密度称为 “Rayleigh 分布”。

➤ 然后, 求出虚警概率  $P_F$  和检测概率  $P_D$ 。

根据判决规则:  $y \underset{H_0}{\overset{H_1}{>}} y_T$ , 可得:

$$P_F = \int_{y_T}^{+\infty} p(y | H_0) dy = \int_{y_T}^{+\infty} \frac{y}{\sigma^2} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy = \exp\left(-\frac{1}{2} \left(\frac{y_T}{\sigma}\right)^2\right)$$

$$\begin{aligned} P_D &= \int_{y_T}^{+\infty} p(y | H_1) dy = \int_{y_T}^{+\infty} \frac{y}{\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \left[y^2 + \frac{E_s T}{2}\right]\right) \cdot I_0\left(\frac{y}{\sigma^2} \sqrt{\frac{E_s T}{2}}\right) dy \\ &= Q\left(\frac{1}{\sigma} \sqrt{\frac{E_s T}{2}}, \frac{y_T}{\sigma}\right) \end{aligned}$$

其中:  $Q(a, b) = \int_b^{+\infty} u \exp\left(-\frac{u^2 + a^2}{2}\right) I_0(au) du$  称为 “Marcum 函数”。

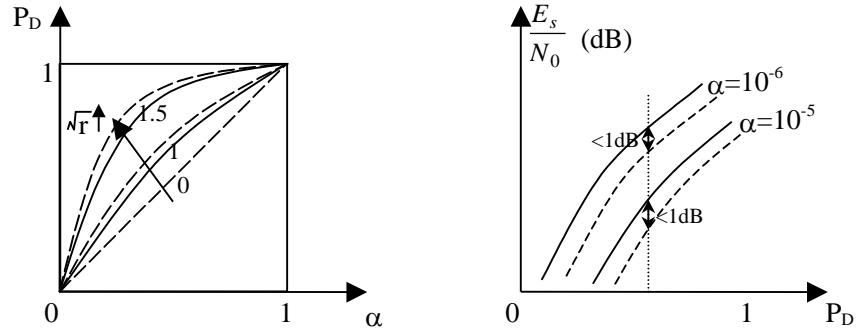
再部分代入  $\sigma$  的值, 可得:  $P_D = Q\left(\sqrt{\frac{2E_s}{N_0}}, \frac{y_T}{\sigma}\right)$ 。

➤ 最后, 分析接收机性能。

若采用 Neyman-Pearson 准则：

由  $P_F=\alpha$  可得：  $\frac{y_T}{\sigma} = \sqrt{-2\ln \alpha}$  。故，检测概率  $P_D = Q\left(\sqrt{\frac{2E_s}{N_0}}, \sqrt{-2\ln \alpha}\right)$ 。

令信噪比  $r = \frac{2E_s}{N_0}$ ，我们可作出接收机的 ROC 曲线（示于下左图中实线）以及当  $\alpha$  一定时检测概率与信噪比的关系曲线（示于下右图中实线）。为对比起见，我们同时将检测确知信号时对应的特性曲线用虚线示于图中。



若采用 最小平均错误概率准则：

由于：  $P_e = (1/2)(P_F + P_M) = \frac{1}{2} \left[ 1 + \exp\left(-\frac{1}{2}\left(\frac{y_T}{\sigma}\right)^2\right) - Q\left(\sqrt{\frac{2E_s}{N_0}}, \frac{y_T}{\sigma}\right) \right]$

为使  $P_e$  最小，我们令  $\frac{\partial P_e}{\partial y_T} = 0$ ，于是：  $\exp\left(-\frac{E_s}{N_0}\right) \cdot I_0\left(\frac{y_T}{\sigma} \sqrt{\frac{2E_s}{N_0}}\right) = 1$

此方程为“超越方程”。当信噪比很大时，可得近似解：  $\frac{y_T}{\sigma} = \sqrt{\frac{E_s}{2N_0}}$ 。

此时（当  $a$  和  $b$  很大时，  $Q(a,b) \approx \Phi(a-b)$ ）：  $P_e \approx \frac{1}{2} \exp\left(-\frac{E_s}{4N_0}\right)$ 。

#### 1.5.4 在白高斯信道中随幅随相信号检测

一种常用的信道模型是瑞利衰落信道。

假定：幅度和相位是相互独立的两个随机变量，且已知它们的先验概率密度。

$H_0: z(t) = n(t), 0 \leq t \leq T$ ,

$H_1: z(t) = A \sin(\omega_c t + \theta) + n(t), 0 \leq t \leq T$ ,

其中：—  $n(t)$  是均值为零、谱密度为  $\frac{N_0}{2}$  的白高斯噪声；

—  $\theta$  是随机变量，在  $[0, 2\pi]$  上均匀分布，在观测期间内保持恒值；

—  $A$  是随机变量，服从瑞利分布：  $p(A) = \frac{A}{\sigma_A^2} \exp\left(-\frac{A^2}{2\sigma_A^2}\right), A \geq 0$ 。

这样的信号叫作“瑞利慢衰落信号”。

此时，似然比检验的判决规则可写为：

$$\lambda(z(t)) = \frac{\int_{\{A\}} \int_{\{\theta\}} p(\mathbf{z} | H_1, A, \theta) p(A) p(\theta) d\theta dA}{p(\mathbf{z} | H_0)} = \int_{\{A\}} \lambda(z(t) | A) p(A) dA \underset{H_0}{\overset{H_1}{>}} \lambda_T$$

$$\text{其中: } \lambda(z(t) | A) = \int_{\{\theta\}} \frac{p(\mathbf{z} | H_1, A, \theta)}{p(\mathbf{z} | H_0)} p(\theta) d\theta$$

在给定 A 值时, 上述似然比  $\lambda(z(t) | A)$  的计算和随相信号的情形一样。因此, 由应用二中的 (\*\*\*) 式并代入  $E_s = (1/2)A^2T$  和 y 的表达式, 可得此时的似然比为:

$$\lambda(z(t) | A) = \exp\left(-\frac{A^2 T}{2N_0}\right) \cdot I_0\left(\frac{2}{N_0} Ay\right)$$

所以, 再代入 p(A) 的表达式, 我们有:

$$\begin{aligned} \lambda(z(t)) &= \int_0^{+\infty} \exp\left(-\frac{A^2 T}{2N_0}\right) I_0\left(\frac{2}{N_0} Ay\right) \cdot \frac{A}{\sigma_A^2} \exp\left(-\frac{A^2}{2\sigma_A^2}\right) dA \\ &= \int_0^{+\infty} \frac{A}{\sigma_A^2} \exp\left(-\frac{A^2}{2} \left(\frac{T}{N_0} + \frac{1}{\sigma_A^2}\right)\right) I_0\left(\frac{2y}{N_0} A\right) dA \end{aligned}$$

此积分的计算可用下述方法之一:

(一) 利用 Marcum 函数<sup>[张树京]</sup>:

$$\text{令 } u = A \sqrt{\frac{T}{N_0} + \frac{1}{\sigma_A^2}}, \text{ 则}$$

$$\lambda(z(t)) = \int_0^{+\infty} \frac{N_0}{\sigma_A^2 T + N_0} u \exp\left(-\frac{u^2}{2}\right) I_0\left(\frac{2y}{N_0 \sqrt{\frac{T}{N_0} + \frac{1}{\sigma_A^2}}} u\right) du$$

$$\text{再令 } a = \frac{2y}{N_0 \sqrt{\frac{T}{N_0} + \frac{1}{\sigma_A^2}}}, \text{ 则}$$

$$\begin{aligned} \lambda(z(t)) &= \frac{N_0}{\sigma_A^2 T + N_0} \exp\left(\frac{a^2}{2}\right) \int_0^{+\infty} u \exp\left(-\frac{u^2 + a^2}{2}\right) I_0(au) du \\ &= \frac{N_0}{\sigma_A^2 T + N_0} \exp\left(\frac{a^2}{2}\right) Q(a, 0) \end{aligned}$$

又  $Q(a, 0) = 1$  并代回 a 的表达式, 可得:

$$\lambda(z(t)) = \frac{N_0}{\sigma_A^2 T + N_0} \exp\left(\frac{2\sigma_A^2}{N_0(\sigma_A^2 T + N_0)} y^2\right)$$

(二) 利用下述公式<sup>[刘有恒]</sup>:

$$\int_0^{+\infty} x \exp(-ax^2) I_0(x) dx = \frac{1}{2a} \exp\left(\frac{1}{4a}\right)$$

$$\text{令 } x = \frac{2y}{N_0} A, \text{ 则 } \lambda(z(t)) = \int_0^{+\infty} \frac{N_0^2}{4y^2 \sigma_A^2} x \exp\left(-\frac{N_0(\sigma_A^2 T + N_0)}{8y^2 \sigma_A^2} x^2\right) I_0(x) dx$$

$$\text{又令 } a = \frac{N_0(\sigma_A^2 T + N_0)}{8y^2 \sigma_A^2}, \text{ 则}$$

$$\lambda(z(t)) = \frac{N_0^2}{4y^2\sigma_A^2} \int_0^{+\infty} x \exp(-ax^2) I_0(x) dx = \frac{N_0^2}{4y^2\sigma_A^2} \frac{1}{2a} \exp\left(\frac{1}{4a}\right)$$

$$\text{再代回 } a \text{ 的表达式, 可得: } \lambda(z(t)) = \frac{N_0}{\sigma_A^2 T + N_0} \exp\left(\frac{2\sigma_A^2}{N_0(\sigma_A^2 T + N_0)} y^2\right)$$

根据似然比检验的对数形式并整理后, 可得判决规则:

$$\begin{matrix} H_1 \\ y > y_T \\ H_0 \end{matrix} \quad \text{其中: } y_T = \sqrt{\frac{N_0(\sigma_A^2 T + N_0)}{2\sigma_A^2} \ln\left(\lambda_T \left(1 + \frac{\sigma_A^2 T}{N_0}\right)\right)}$$

故, 最佳接收机的实现形式与前面应用二中的接收机形式一样, 只是门限  $y_T$  不一样。

最后, 我们来分析最佳接收机的性能。

在假设  $H_0$  下, 由于  $y|H_0$  与信号幅度  $A$  无关, 故其概率密度函数不变。所以, 虚警概率的计算与前面随相信号情况时相同, 即:

$$P_F = \int_{y_T}^{+\infty} p(y|H_0) dy = \int_{y_T}^{+\infty} \frac{y}{\sigma^2} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy = \exp\left(-\frac{1}{2}\left(\frac{y_T}{\sigma}\right)^2\right)$$

在假设  $H_1$  下, 在给定  $A$  时, 检测概率的计算也与前面随相信号情况时相同, 即 (将  $E_s = \frac{A^2 T}{2}$  代入):

$$P_D(A) = \int_{y_T}^{+\infty} p(y|H_1, A) dy = \int_{y_T}^{+\infty} \frac{y}{\sigma^2} \exp\left[-\frac{1}{2\sigma^2} \left[y^2 + \left(\frac{AT}{2}\right)^2\right]\right] \cdot I_0\left(\frac{ATy}{2\sigma^2}\right) dy$$

现在  $A$  为随机变量, 故应将  $P_D(A)$  对  $A$  求平均:  $P_D = \int_{\{A\}} P_D(A) p(A) dA$

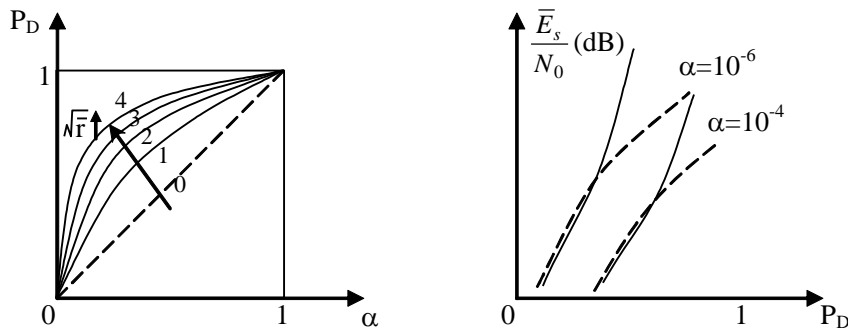
代入  $p(A)$  的表达式, 并经一些运算<sup>[张树京、刘有恒]</sup>, 最后可得:

$$P_D = \exp\left(-\frac{1}{2}\left(\frac{y_T}{\sigma}\right)^2 \left(1 + \frac{\bar{E}_s}{N_0}\right)^{-1}\right) \quad \text{其中: 平均信号能量 } \bar{E}_s = \sigma_A^2 T。$$

定义: 平均信噪比  $\bar{r} = \bar{E}_s / N_0$ 。

再考虑到虚警概率的表达式并令  $P_F = \alpha$ , 我们有:  $P_D = \alpha^{1/(1+\bar{r})}$

从而我们可得到接收机的 ROC 曲线如下图所示:



由图可见: 在较低的信噪比时, 随幅随相信号的检测性能 (实线) 与随相信号的 (虚线) 相比相差不多或更好; 而在高信噪比时, 随幅随相信号的检测性能与随相信号的相比就差得多了。其中的缘由可参见参考资料<sup>[刘有恒]</sup>。

## 1.6 序列检测 (Wald 检测)

现在, 我们来考虑非固定的样本总数情况。

定义: 观测样本矢量  $\mathbf{z}_j = (z_1, z_2, \dots, z_j)^T$ 。

序列检测是分阶段进行的。也就是说, 每获得一个新的样本后, 就按照下述的序列检测 (Wald 检测) 判决规则进行一次判决:

$$\lambda(\mathbf{z}_j) \begin{cases} > \lambda_1 \\ < \\ > \\ < \lambda_0 \end{cases} \begin{matrix} H_1 \\ \\ \\ H_0 \end{matrix} \quad \left. \begin{matrix} < \\ > \end{matrix} \right\} \text{继续观测}$$

其中,  $\lambda_1$  和  $\lambda_0$  分别由给定的虚警概率  $P_F$  和漏警概率  $P_M$  确定。

这就意味着: 如果我们已获得足够的信息, 则停止观测信号而迅速作出判决; 否则的话, 继续观测信号以逐步增加作出判决所必需的信息量。

下面, 我们来讨论几个关键量的计算。

➤ 根据  $P_F = \alpha$ ,  $P_M = \beta$  来计算  $\lambda_1$  和  $\lambda_0$  之值:

由定义:  $P_F = \int_{\mathcal{Z}_1} p(\mathbf{z}_j | H_0) d\mathbf{z}_j$  和  $P_D = \int_{\mathcal{Z}_1} p(\mathbf{z}_j | H_1) d\mathbf{z}_j$ 。

稍加推导, 可得:  $P_D = \int_{\mathcal{Z}_1} \frac{p(\mathbf{z}_j | H_1)}{p(\mathbf{z}_j | H_0)} p(\mathbf{z}_j | H_0) d\mathbf{z}_j = \int_{\mathcal{Z}_1} \lambda(\mathbf{z}_j) p(\mathbf{z}_j | H_0) d\mathbf{z}_j$

根据判决准则, 当假设  $H_1$  为真且判决  $H_1$  成立时, 应有:  $\lambda(\mathbf{z}_j) \geq \lambda_1$ 。代入上式可得:

$$P_D \geq \int_{\mathcal{Z}_1} \lambda_1 p(\mathbf{z}_j | H_0) d\mathbf{z}_j = \lambda_1 \int_{\mathcal{Z}_1} p(\mathbf{z}_j | H_0) d\mathbf{z}_j = \lambda_1 P_F$$

又  $P_D = 1 - \beta$  和  $P_F = \alpha$ , 故,  $1 - \beta \geq \lambda_1 \alpha$ , 即:  $\lambda_1 \leq \frac{1 - \beta}{\alpha}$ 。

类似地, 由  $P(D_0 | H_0) = 1 - \alpha$  可推得:  $\lambda_0 \geq \frac{\beta}{1 - \alpha}$ 。

若各取样值是相互独立的, 则有:  $\lambda(\mathbf{z}_j) = \frac{p(\mathbf{z}_j | H_1)}{p(\mathbf{z}_j | H_0)} = \prod_{k=1}^j \frac{p(z_k | H_1)}{p(z_k | H_0)} = \lambda(z_j) \lambda(\mathbf{z}_{j-1})$

两边取对数得:  $\mathcal{L}(\mathbf{z}_j) = \mathcal{L}(z_j) + \mathcal{L}(\mathbf{z}_{j-1})$

当上式中增量  $\mathcal{L}(z_j)$  很小时, 作出判决时  $\mathcal{L}(\mathbf{z}_j)$  应和门限相差很少。故, 在上面的两个不等式中可近似地取等号, 从而求得简单的设计公式:

$$\ln \lambda_1 \approx \ln\left(\frac{1 - \beta}{\alpha}\right) \text{ 和 } \ln \lambda_0 \approx \ln\left(\frac{\beta}{1 - \alpha}\right)。$$

➤ 每个假设下平均取样数的计算:

定义: 二元变量  $k$  为

$$k_m = \begin{cases} 1, & \text{直到第}(m-1)\text{步尚未作出判决} \\ 0, & \text{若在以前某一步已作出了判决} \end{cases}$$

若假定第  $N$  步结束了检验, 则:  $\mathcal{L}(\mathbf{z}_N) = \sum_{m=1}^N \mathcal{L}(z_m) = \sum_{m=1}^{\infty} k_m \mathcal{L}(z_m)$

因此, 给定假设  $H_i$  时  $\mathcal{L}(\mathbf{z}_N)$  的条件均值:

$$\begin{aligned}
E[\mathcal{L}(\mathbf{z}_N)|H_i] &= \sum_{n=1}^{\infty} E[\mathcal{L}(\mathbf{z}_N)|H_i, N=n] P(N=n|H_i) \\
&= \sum_{n=1}^{\infty} E\left[\sum_{m=1}^n k_m \mathcal{L}(z_m) | H_i, N=n\right] P(N=n|H_i)
\end{aligned}$$

由于  $k_m$  仅依赖于  $z_1, z_2, \dots, z_{m-1}$  并且与  $z_m$  无关，故亦与  $\mathcal{L}(z_m)$  无关。因而，

$$E[\mathcal{L}(\mathbf{z}_N)|H_i] = \sum_{n=1}^{\infty} P(N=n|H_i) \sum_{m=1}^n E[\mathcal{L}(z_m)|H_i] E[k_m|H_i, N=n]$$

又假定在各假设下观测值具有相同分布： $E[\mathcal{L}(z_m)|H_i] = E[\mathcal{L}(z)|H_i]$ ，则

$$\begin{aligned}
E[\mathcal{L}(\mathbf{z}_N)|H_i] &= E[\mathcal{L}(z)|H_i] \sum_{n=1}^{\infty} P(N=n|H_i) \sum_{m=1}^n E[k_m|H_i, N=n] \\
&= E[\mathcal{L}(z)|H_i] \sum_{n=1}^{\infty} P(N=n|H_i) E\left[\sum_{m=1}^n k_m | H_i, N=n\right] \\
&= E[\mathcal{L}(z)|H_i] \sum_{n=1}^{\infty} P(N=n|H_i) n = E[\mathcal{L}(z)|H_i] E(N|H_i)
\end{aligned}$$

从而，可得给定假设  $H_i$  时的平均取样数  $E(N|H_i) = \frac{E[\mathcal{L}(\mathbf{z}_N)|H_i]}{E[\mathcal{L}(z)|H_i]}$

由于上式中  $E[\mathcal{L}(z)|H_i]$  容易计算，故我们现在来考虑  $E[\mathcal{L}(\mathbf{z}_N)|H_i]$  的计算问题。

因为假定取得  $N$  个取样值时结束检验，所以根据判决规则，此时  $\mathcal{L}(\mathbf{z}_N)$  近似地只取两个值之一： $\ln\lambda_0$  或  $\ln\lambda_1$ 。

假设  $H_1$  为真时：

若  $\mathcal{L}(\mathbf{z}_N) \approx \ln\lambda_1$ ，则此时检测正确，相应的概率为  $P_D = 1 - \beta$ ；

若  $\mathcal{L}(\mathbf{z}_N) \approx \ln\lambda_0$ ，则此时检测错误，相应的概率为  $P_M = \beta$ 。

因而，我们有： $E[\mathcal{L}(\mathbf{z}_N)|H_1] = (1 - \beta) \ln\lambda_1 + \beta \ln\lambda_0$ 。

假设  $H_0$  为真时：

若  $\mathcal{L}(\mathbf{z}_N) \approx \ln\lambda_1$ ，则此时检测错误，相应的概率为  $P_F = \alpha$ ；

若  $\mathcal{L}(\mathbf{z}_N) \approx \ln\lambda_0$ ，则此时检测正确，相应的概率为  $(1 - \alpha)$ 。

因而，我们有： $E[\mathcal{L}(\mathbf{z}_N)|H_0] = \alpha \ln\lambda_1 + (1 - \alpha) \ln\lambda_0$ 。

所以，最终我们可得各种假设下的平均取样数为：

$$E(N|H_1) = \frac{(1 - \beta) \ln \lambda_1 + \beta \ln \lambda_0}{E[\mathcal{L}(z)|H_1]} \quad E(N|H_0) = \frac{\alpha \ln \lambda_1 + (1 - \alpha) \ln \lambda_0}{E[\mathcal{L}(z)|H_0]}$$

Wald 和 Wolfowitz 证明了：对于给定的  $P_F$  和  $P_M$  值，序列检测可使平均观测次数（取样数） $E(N|H_1)$  和  $E(N|H_0)$  为最小。一般来说，它比固定样本检测节省 1/2~3/4 的样本数。其存在的缺点是：检测的每一步都要重新调整检验统计量；需要储存较多的样本数等。

另外，还可以证明：序列检测以概率 1 来结束。但在低信噪比的情况下，可能需要的样本数太多而导致检测的时间太长。所以，我们最好将取样总数的上限固定。当到达其上限值时，采用固定样本的检测方法作出判决。这样的检验称为“截断的序列检测”。

◆例：

二元假设。  $H_1: z_j = 1 + n_j$ ;  $H_0: z_j = n_j$ 。  $n_j \sim N(0, 1)$ 。假定： $P_F = \alpha = 0.1$ ,  $P_M = \beta = 0.1$ 。

求：1) 序列检测时的判决规则；

- 2) 序列检测时每种假设下结束检验所需的平均样本数;  
3) 若固定样本数, 确定满足性能所需的样本数。

$$\text{解: } \lambda(z) = \frac{\frac{1}{2\pi} \exp\left(-\frac{(z-1)^2}{2}\right)}{\frac{1}{2\pi} \exp\left(-\frac{z^2}{2}\right)} = \exp\left(z - \frac{1}{2}\right)$$

$$\mathcal{L}(z) = z - \frac{1}{2}$$

$$\mathcal{L}(\mathbf{z}_j) = \sum_{k=1}^j z_k - \frac{j}{2}$$

$$1) \text{ 由公式可得: } \ln \lambda_1 = \ln \frac{1-\beta}{\alpha} = \ln 9 = 2.197$$

$$\ln \lambda_0 = \ln \frac{\beta}{1-\alpha} = \ln \frac{1}{9} = -2.197$$

$$\text{故 Wald 检验判决规则为: } \mathcal{L}(\mathbf{z}_j) \begin{matrix} >_{H_1} 2.197 \\ <_{H_0} -2.197 \end{matrix} \left. \vphantom{\begin{matrix} >_{H_1} 2.197 \\ <_{H_0} -2.197 \end{matrix}} \right\} \text{继续观测}$$

- 2) 平均样本总数为:

$$E(N|H_1) = \frac{(1-\beta) \ln \lambda_1 + \beta \ln \lambda_0}{E[\mathcal{L}(z) | H_1]} = \frac{(1-0.1) \ln 9 + 0.1(-\ln 9)}{E\left[\left(z - \frac{1}{2}\right) | H_1\right]} = \frac{(1-0.1) \ln 9 + 0.1(-\ln 9)}{1 - \frac{1}{2}} = 3.52$$

$$E(N|H_0) = \frac{\alpha \ln \lambda_1 + (1-\alpha) \ln \lambda_0}{E[\mathcal{L}(z) | H_0]} = \frac{0.1 \ln 9 + (1-0.1)(-\ln 9)}{E\left[\left(z - \frac{1}{2}\right) | H_0\right]} = \frac{0.1 \ln 9 + (1-0.1)(-\ln 9)}{0 - \frac{1}{2}} = 3.52$$

取整后, 可得:  $E(N|H_1) = E(N|H_0) = 4$ 。

- 3) 设固定样本总数为  $N$ , 并令检验统计量  $y = \frac{1}{\sqrt{N}} \sum_{k=1}^N z_k$ , 则

$$y|H_1 \sim N(\sqrt{N}, 1), \quad y|H_0 \sim N(0, 1)。$$

根据似然比检验判决规则:  $y \underset{H_0}{\overset{H_1}{>}} y_T$ , 我们有:

$$P_F = \int_{y_T}^{+\infty} p(y|H_0) dy = 1 - \Phi(y_T) \quad \text{且} \quad P_M = \int_{-\infty}^{y_T} p(y|H_1) dy = \Phi(y_T - \sqrt{N})。$$

由题意:  $P_F = P_M$ , 故有:  $y_T - \sqrt{N} = -y_T$ , 即:  $y_T = \sqrt{N}/2$ 。

又  $P_F = 1 - \Phi(y_T) = 0.1$ , 故查表可得:  $y_T = 1.2815$ 。

最终, 经运算并取整后可得:  $N = 7$ 。

可见, 此时的  $N$  值比序列检测的平均样本总数大  $3/4$ 。

#解毕



## 1.7 在色高斯信道中的信号检测

在前面 1.3 节中，我们曾讨论了在白高斯信道中的信号检测问题。在许多实际的信道中，由于存在着各种各样因素的影响（如：从噪声信号源到接收机之间通常存在各种滤波器），所以，接收端所获得的信号中不再只含有白噪声而是**色噪声**（具有某种形状功率谱的噪声），这时的信号样本值就不再是相互独立的了。那么，在这种情况下，我们如何解决信号检测的问题呢？

我们可以选择两种办法：**Karhunen-Loève 展开**和**白化滤波器**。前一种方法是将接收信号在归一化正交函数集上展开，这样便获得了该信号的另一等效表示法——**K-L 展开系数集**（因为通过这些系数我们亦可恢复接收信号），由于这些系数互不相关且服从高斯分布，故它们相互独立，从而可以写出它们的似然函数（和接收信号的幅度似然函数是等效的）。后一种方法是将接收信号先通过一个滤波器，使其输出的噪声信号成为白噪声，这样我们就可以利用前面白高斯信道中信号检测的结果了。

下面我们将更为具体地来讨论这两种方法。为了简明扼要，我们仍以二元假设为例，并假定：假设  $H_1$  为真时， $z(t)=s_i(t)+n(t)$ ，其中  $n(t)$  是均值为零、协方差函数为  $R_n(t_1, t_2)$  的色高斯噪声。

### 1.7.1 Karhunen-Loève 正交展开法

◆正交展开：

把  $z(t)$  在基函数  $\{g_k(t)\}$  上展开：
$$z(t) = \sum_{k=1}^{+\infty} z_k g_k(t) \quad \text{其中：} z_k = \int_0^T z(t) g_k(t) dt。$$

我们展开的目的是：选择基函数以使展开系数  $\{z_k\}$  相互独立。

由于  $z_k$  是高斯随机变量，所以我们只须使  $\{z_k\}$  互不相关即可，也就是说，其协方差函数应满足下述关系（以实函数为例）：

$$E[(z_k - \bar{z}_k)(z_l - \bar{z}_l)] = \lambda_k \delta_{kl} \quad \text{其中：} \bar{z}_k = E[z_k] = \int_0^T s_i(t) g_k(t) dt = s_{ik}$$

$$\text{又 } z_k - \bar{z}_k = z_k - s_{ik} = \int_0^T [z(t) - s_i(t)] g_k(t) dt = \int_0^T n(t) g_k(t) dt = n_k$$

$$\text{故：} E[(z_k - \bar{z}_k)(z_l - \bar{z}_l)] = E[n_k n_l] = E\left[\int_0^T n(t_1) g_k(t_1) dt_1 \int_0^T n(t_2) g_l(t_2) dt_2\right] = \int_0^T \int_0^T R_n(t_1, t_2) g_k(t_1) g_l(t_2) dt_1 dt_2$$

$$\text{若：} \int_0^T R_n(t_1, t_2) g_l(t_2) dt_2 = \lambda_l g_l(t_1) \quad (*)$$

且： $\{g_k(t)\}$  是归一化正交函数集  $\int_0^T g_k(t) g_l(t) dt = \delta_{kl}$ 。

$$\text{则，} E[(z_k - \bar{z}_k)(z_l - \bar{z}_l)] = \int_0^T g_k(t_1) \lambda_l g_l(t_1) dt_1 = \lambda_l \delta_{kl} = \lambda_k \delta_{kl}$$

因此，如果我们选择  $\{g_k(t)\}$  满足上述积分方程  $(*)$ ，那么  $z(t)$  的展开式系数  $\{z_k\}$  就互不相关。该展开式就称为“**Karhunen-Loève 展开式**”。

方程  $(*)$  称为“**齐次积分方程**”， $R_n(t_1, t_2)$  是该积分方程的核， $g_l(t)$  是本征函数， $\lambda_l$  是其对应的本征值。

◆似然比计算及判决规则：

我们以接收信号 K-L 展开式的前  $N$  个系数来建立“等效”观测向量： $\mathbf{z}_N = (z_1, z_2, \dots, z_N)^T$

由此，我们有似然比函数：

$$\lambda(\mathbf{z}_N) = \frac{\prod_{k=1}^N \frac{1}{\sqrt{2\pi\lambda_k}} \exp\left(-\frac{(z_k - s_{1k})^2}{2\lambda_k}\right)}{\prod_{k=1}^N \frac{1}{\sqrt{2\pi\lambda_k}} \exp\left(-\frac{(z_k - s_{0k})^2}{2\lambda_k}\right)} = \exp\left(\sum_{k=1}^N \frac{1}{2\lambda_k} [2z_k(s_{1k} - s_{0k}) - (s_{1k}^2 - s_{0k}^2)]\right)$$

取对数并整理后得： $\ln\lambda(\mathbf{z}_N) = \sum_{k=1}^N \frac{1}{\lambda_k} z_k(s_{1k}-s_{0k}) - \frac{1}{2} \sum_{k=1}^N \frac{1}{\lambda_k} (s_{1k}^2 - s_{0k}^2)$

令  $N \rightarrow \infty$ ，我们可得连续信号的等效似然比函数：

$$\ln\lambda(z(t)) = \lim_{N \rightarrow \infty} \ln\lambda(\mathbf{z}_N) = \sum_{k=1}^{\infty} \frac{1}{\lambda_k} z_k(s_{1k}-s_{0k}) - \frac{1}{2} \sum_{k=1}^{\infty} \frac{1}{\lambda_k} (s_{1k}^2 - s_{0k}^2)$$

将  $z_k$  的定义代入上式的第一项中、 $s_{ik}$  的定义取一个代入上式的第二项中可得：

$$\ln\lambda(z(t)) = \int_0^T z(t) \sum_{k=1}^{\infty} \frac{1}{\lambda_k} (s_{1k}-s_{0k}) g_k(t) dt - \frac{1}{2} \left[ \int_0^T s_1(t) \sum_{k=1}^{\infty} \frac{1}{\lambda_k} s_{1k} g_k(t) dt - \int_0^T s_0(t) \sum_{k=1}^{\infty} \frac{1}{\lambda_k} s_{0k} g_k(t) dt \right]$$

再令  $h_i(t) = \sum_{k=1}^{\infty} \frac{1}{\lambda_k} s_{ik} g_k(t)$ ，则上式可简化为：

$$\ln\lambda(z(t)) = \left[ \int_0^T z(t) h_1(t) dt - \int_0^T z(t) h_0(t) dt \right] - \frac{1}{2} \left[ \int_0^T s_1(t) h_1(t) dt - \int_0^T s_0(t) h_0(t) dt \right]$$

所以，根据似然比检验准则： $\ln\lambda(z(t)) \underset{H_0}{\overset{H_1}{>}} \ln\lambda_T$ ，我们可得判决规则如下：

$$\int_0^T z(t) h_1(t) dt - \int_0^T z(t) h_0(t) dt \underset{H_0}{\overset{H_1}{>}} v_T \quad \text{其中：} v_T = \ln\lambda_T + \frac{1}{2} \left[ \int_0^T s_1(t) h_1(t) dt - \int_0^T s_0(t) h_0(t) dt \right]$$

可以证明  $h_i(t)$  是下述积分方程的解： $\int_0^T R_n(t, \tau) h_i(\tau) d\tau = s_i(t)$

#### ◆最佳接收机：

- 1) 相关接收机
- 2) 匹配滤波器

### 1.7.2 白化滤波器

#### ◆思路：

可以设想将接收信号  $z(t)$  通过一个滤波器  $h_w(t, \tau)$ ： $z(t) \longrightarrow \boxed{h_w(t, \tau)} \longrightarrow z_w(t)$   
 $0 \leq \tau \leq T$

那么，我们可将输出信号  $z_w(t)$  写成两部分：

$$z_w(t) = \int_0^T h_w(t, \tau) z(\tau) d\tau \quad |H_i = \int_0^T h_w(t, \tau) (s_i(\tau) + n(\tau)) d\tau = \int_0^T h_w(t, \tau) s_i(\tau) d\tau + \int_0^T h_w(t, \tau) n(\tau) d\tau$$

$$\text{定义：} s_{wi}(t) = \int_0^T h_w(t, \tau) s_i(\tau) d\tau \quad n_w(t) = \int_0^T h_w(t, \tau) n(\tau) d\tau$$

$$\text{则有：} z_w(t) = s_{wi}(t) + n_w(t)$$

若我们能找到这样一个滤波器  $h_w(t, \tau)$ ，使其输出噪声信号  $n_w(t)$  为白噪声，即

$$R_{nw}(t, \tau) = E[n_w(t) n_w(\tau)] = \delta(t - \tau)$$

则，对滤波器输出信号的似然比检验的判决规则将与前面 § 1.3.1 节中连续信号在白高斯信道中的信号检测一样，即：

$$\int_0^T s_{w1}(t) z_w(t) dt - \int_0^T s_{w0}(t) z_w(t) dt \underset{H_0}{\overset{H_1}{>}} v_T, \quad \text{其中：} v_T = \ln\lambda_T + \frac{1}{2} (E_{w1} - E_{w0})$$

从而，可设计出与此相应的最佳接收机。

◆**关键：**设计满足要求的白化滤波器  $h_w(t, \tau)$ 。

◆**求解：**

$h_w(t, \tau)$ 可正交展开为如下形式: 
$$h_w(t, \tau) = \sum_{k=1}^{\infty} h_k g_k(t) g_k(\tau)$$

这样可得:

$$\begin{aligned} R_{nw}(t, \tau) &= E[n_w(t)n_w(\tau)] = E\left[\int_0^T \int_0^T h_w(t, s) h_w(\tau, u) n(s) n(u) ds du\right] \\ &= \int_0^T \int_0^T h_w(t, s) h_w(\tau, u) R_n(s, u) ds du = \int_0^T \int_0^T \left[ \sum_{k=1}^{\infty} h_k g_k(t) g_k(s) \right] \left[ \sum_{l=1}^{\infty} h_l g_l(\tau) g_l(u) \right] R_n(s, u) ds du \\ &= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \int_0^T h_k h_l g_k(t) g_k(s) g_l(\tau) \lambda_l g_l(s) ds = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} h_k h_l g_k(t) g_l(\tau) \lambda_l \delta_{kl} \\ &= \sum_{k=1}^{\infty} h_k^2 g_k(t) g_k(\tau) \lambda_k \end{aligned}$$

根据我们的要求:  $E[n_w(t)n_w(\tau)] = \delta(t-\tau)$ , 而 $\delta(t-\tau)$ 可表示为:  $\delta(t-\tau) = \sum_{k=1}^{\infty} g_k(t) g_k(\tau)$

因此, 我们有: 
$$\sum_{k=1}^{\infty} h_k^2 g_k(t) g_k(\tau) \lambda_k = \sum_{k=1}^{\infty} g_k(t) g_k(\tau)$$

故, 对比等式两边的系数, 可得: 
$$h_k^2 \lambda_k = 1 \quad \text{即} \quad h_k = \frac{1}{\sqrt{\lambda_k}}$$

所以, 最后我们得到: 
$$h_w(t, \tau) = \sum_{k=1}^{\infty} \frac{1}{\sqrt{\lambda_k}} g_k(t) g_k(\tau)$$

可以证明 (自习), 判决规则与上一节 § 1.7.1 中是一样的。

一般而言, 计算  $h_w(t, \tau)$  相当棘手 (特殊的积分方程才可解)。

#### ◆实际情况中:

一种重要的情形是: 平稳噪声, 有理谱, 半无限观测区间。此时,  $\Phi_n(s) = \Phi_n^+(s) \Phi_n^-(s)$

则, 时不变白化滤波器为:  $H_w(s) = 1 / \Phi_n^+(s)$

这样, 可以核实其输出的噪声频谱为 1:

$$\Phi_y(s) = \Phi_n(s) H_w(s) H_w(-s) = \Phi_n(s) / [\Phi_n^+(s) \Phi_n^+(-s)] = \Phi_n(s) / [\Phi_n^+(s) \Phi_n^-(s)] = 1$$

即意味着噪声被白化了。

### 1.7.3 接收机性能

在 § 1.3.2 二元数字通信系统中, 在白高斯信道和等先验概率情况下, 使用最小错误概率准则, 我们曾得:

$$P_e = P_F = P_M = 1 - \Phi\left(\sqrt{\frac{E(1-\rho)}{N_0}}\right)$$

由此, 在色高斯信道中, 我们可使用类推的方法来得到此时的平均错误概率。

首先, 在计算中, 色噪声时须用白化后的信号  $s_{w1}(t)$  来代替白噪声时的信号  $s_1(t)$ , 即

$$E(1-\rho) = \frac{1}{2} \int_0^T [s_{w1}(t) - s_{w0}(t)]^2 dt$$

其次, 考察  $N_0$  的值: 白噪声时,  $R_n(t, \tau) = (N_0/2) \delta(t-\tau)$ ;

而色噪声白化后,  $R_{nw}(t, \tau) = \delta(t-\tau)$ 。故, 色噪声等效于白噪声当  $N_0=2$  时的情形。

所以，在色高斯信道中平均错误概率为：

$$P_e = 1 - \Phi(d)$$

$$\text{其中： } d^2 = \frac{1}{4} \int_0^T [s_{w1}(t) - s_{w0}(t)]^2 dt = \frac{1}{4} \int_0^T \left[ \int_0^T h_w(t, \tau) (s_1(\tau) - s_0(\tau)) d\tau \right]^2 dt$$

可见，平均错误概率  $P_e$  的计算：

- 1) 需要解齐次积分方程以求得  $h_w(t, \tau)$  (这并非易事)；
- 2) 与信号波形有关 (而在白高斯信道中仅与信号能量有关)；
- 3)  $d \uparrow \Rightarrow P_e \downarrow$

#### 1.7.4 最佳信号波形

给定约束条件：平均信号能量  $E = \frac{1}{2} \int_0^T [s_1^2(t) + s_0^2(t)] dt = \text{常值}$

我们可利用 Lagrange 乘子法来求最佳信号波形，即使下面的目标函数取最大值：

$$Q = \int_0^T \left[ \int_0^T h_w(t, \tau) (s_1(\tau) - s_0(\tau)) d\tau \right]^2 dt - \mu \int_0^T [s_1^2(t) + s_0^2(t)] dt$$

在进行变分运算后，可得：

$$\int_0^T \int_0^T [s_1(\tau) - s_0(\tau)] h_w(t, \tau) h_w(t, u) dt d\tau = \mu s_1(u) \quad (*)$$

$$\text{且 } s_1(u) = -s_0(u) \quad u \in [0, T]$$

由此， $d^2$  可表达为：

$$\begin{aligned} d^2 &= \frac{1}{4} \int_0^T \left[ \int_0^T h_w(t, \tau) (s_1(\tau) - s_0(\tau)) d\tau \right] \left[ \int_0^T h_w(t, u) (s_1(u) - s_0(u)) du \right] dt \\ &= \frac{1}{4} \int_0^T (s_1(u) - s_0(u)) \mu s_1(u) du = \frac{\mu}{2} \int_0^T s_1^2(u) du = \frac{\mu}{2} \int_0^T s_0^2(u) du \end{aligned}$$

**推论：**给定信号能量时，要使  $P_e$  最小，则要使  $d^2$  最大。故， $\mu$  应选取它的最大值。

由上述 (\*) 式 (考虑到  $s_0(t) = -s_1(t)$ ) 可得：

$$\int_0^T \int_0^T s_1(\tau) h_w(t, \tau) h_w(t, u) dt d\tau = \frac{\mu}{2} s_1(u)$$

$$\text{对两边求积分：} \int_0^T R_n(u, w) du$$

$$\int_0^T s_1(\tau) \left[ \int_0^T \int_0^T R_n(u, w) h_w(t, \tau) h_w(t, u) du dt \right] d\tau = \int_0^T \frac{\mu}{2} s_1(u) R_n(u, w) du \quad (**)$$

又，由白化滤波器的求解过程 ( $E[n_w(t)n_w(\tau)] = \delta(t-\tau)$ ) 中知：

$$\int_0^T \int_0^T h_w(t, s) h_w(t, u) R_n(s, u) ds du = \delta(t-\tau)$$

$$\text{对两边求积分：} \int_0^T h_w(t, v) dt$$

$$\int_0^T h_w(t, v) \int_0^T \int_0^T h_w(t, s) h_w(t, u) R_n(s, u) ds du dt = \int_0^T \delta(t-\tau) h_w(t, v) dt = h_w(\tau, v)$$

$$\int_0^T h_w(\tau, u) \left[ \int_0^T \int_0^T h_w(t, s) h_w(t, v) R_n(s, u) ds dt \right] du = h_w(\tau, v)$$

$$\text{因而有：} \int_0^T \int_0^T h_w(t, s) h_w(t, v) R_n(s, u) ds dt = \delta(u-v)$$

经符号替换： $s \rightarrow u, v \rightarrow \tau, u \rightarrow w$

$$\int_0^T \int_0^T h_w(t,u) h_w(t,\tau) R_n(u,w) du d\tau = \delta(w-\tau)$$

将上式代入 (\*\*) 式中, 可得:  $\int_0^T s_1(u) R_n(u,w) du = \frac{2}{\mu} s_1(w)$  (\*\*\*)

可见, 最佳信号波形是以协方差  $R_n(u,w)$  为核的、与特征值  $\frac{2}{\mu}$  相对应的特征函数。

根据上面的推论 ( $\mu$  最大  $\Rightarrow P_e$  最小), 故应以下述方式选取最佳信号波形:

- 1) 以 (\*\*) 式的解中  $\frac{2}{\mu}$  为最小值时所对应的特征函数取为  $s_1(t)$ ;
- 2) 取  $s_0(t) = -s_1(t)$ 。

## 1.8 小结

信号检测模型

常用判决准则 (Bayes, Minimimum, Neyman-Pearson, MPE, ML, MAP)

在白高斯信道中的信号检测

多次观测时似然比的计算  $\Rightarrow$  连续信号时似然比的计算

通信系统 (相关接收机, 匹配滤波器)

雷达系统

多元假设检验

复合假设检验

随幅信号, 随相信号, 随幅随相信号的检测。(正交接收机, 非相干匹配滤波器)

序列检测

在色高斯信道中的信号检测

K-L 展开, 以及似然比计算和判决规则

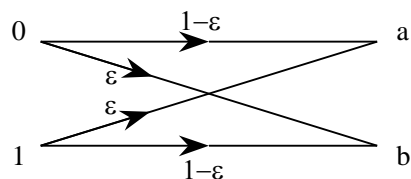
白化滤波器

接收机性能及最佳信号波形

## 第一章练习题

1. 试推导如下情况的似然比。在假设  $H_1$  下，观测值  $z$  是高斯变量，均值为  $m_1$ ，标准偏差为  $\sigma_1$ 。在假设  $H_0$  下， $z$  也是一个高斯变量，但均值为  $m_0$ ，标准偏差为  $\sigma_0$ 。并求判决域和错误概率。

2. 考察右图所示的二元对称信道。 $\varepsilon$  是交叉概率（假定  $\varepsilon < 1/2$ ），即信道输入为 0（或 1）时，输出为 b（或 a）的概率。若先验概率相等，试导出保证平均错误概率最小的判决准则，并求最小平均错误概率。



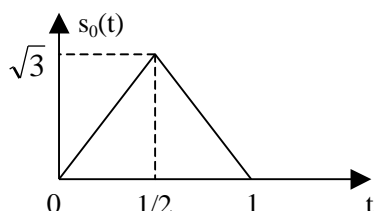
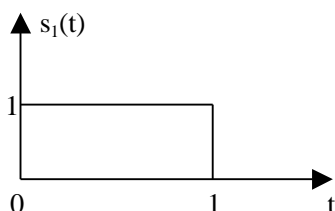
3. 请使用最小错误概率准则，设计一个在如下两种假设间作出选择的接收机（假定两种假设的先验概率相等）：

$$H_1: z(t) = s_1(t) + n(t)$$

$$H_0: z(t) = s_0(t) + n(t)$$

其中：信号  $s_1(t)$  和  $s_0(t)$  如下图所示；噪声  $n(t)$  是高斯随机变量，均值为零，谱密度为  $N_0/2$ 。

并画出平均错误概率与  $2E/N_0$  的函数关系（ $E$  为信号的平均能量）。



4. 已知  $K$  个独立的观测值

$$H_1: z_k = n_k.$$

其中： $n_k$  是均值为零、方差为 2 的高斯随机变量， $k = 1, 2, \dots, K$ 。

$$H_0: z_k = 1 + n_k.$$

- 设计似然比检验，并求  $P_F$  和  $P_M$ 。
- 画出  $K = 1$  时的接收机工作特性。
- 假定  $c_{00} = c_{11} = 0$ ， $c_{01} = 2$ ， $c_{10} = 1$ ， $P(H_0) = 0.7$ ，试求最小  $N$  值，使  $K = N$  时的风险不大于  $K = 1$  时风险的二分之一。

5. 对二元通信系统，其假设为：

$$H_1: z(t) = A \cos \omega_1 t + B \cos(\omega_2 t + \phi) + n(t)$$

$$H_0: z(t) = B \cos(\omega_2 t + \phi) + n(t)$$

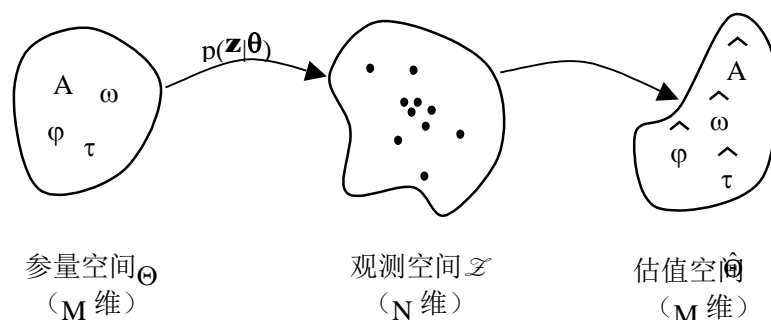
其中： $0 \leq t \leq T$ ，  
 $A$ 、 $B$ 、 $\omega_1$ 、 $\omega_2$ 、 $\phi$  均为已知常数。

假定： $\int_0^T \cos \omega_1 t \cos \omega_2 t dt = \int_0^T \cos \omega_1 t \sin \omega_2 t dt = 0$ ， $n(t)$  是谱密度为  $N_0/2$  的高斯白噪声。试画出其最佳接收机模型，并分析其误码率是否和  $A \cos \omega_1 t$  及  $B \cos(\omega_2 t + \phi)$  有关。计算误码率以证明你的分析结论。

## 第2章 信号估值理论

### 2.1 信号参量估值模型

估值：对信号的参量（一个或多个）进行**估计**，使所估计的信号参量尽量与真实参量一致。



信号的参量：雷达系统中，目标的距离、方位、速度等；通信系统中，脉幅调制 PAM、脉频调制 PFM、脉位调制 PPM。

特点：

- ① 参量**连续**变化（检测：有限数状态）；
- ② 参量的估值**逼近**其真值（检测：判决结果只与原假设相同或不同）；
- ③ 根据准则确定估值的**计算公式**（检测：划分判决空间界面）；
- ④ 估值质量的评价，即估值与真值的**误差**怎样（检测：接收机性能的评价，即  $P_F$ ,  $P_D$ ,  $P_e$  等）。

### 2.2 常用估值准则

#### 2.2.1 Bayes 估值

◆最小平均风险：

$$\text{检测： } \bar{R} = \sum_i \sum_j C_{ij} P(D_i, H_j),$$

由于假设-判决状态对  $(D_i, H_j)$  的数目有限，因而与之对应的是**有限个数**的代价因子  $\{C_{ij}\}$ 。

估值：由于信号参量及其估值具有取值的**连续性**，故上述代价因子应该用**代价函数**来代替。

记代价函数为  $C(\hat{\theta}, \theta)$ ，其中： $\theta$  为信号参量的真值， $\hat{\theta}$  为其估值。

通常，我们认为代价函数只与估值与真值间的误差有关，即令

$$C(\hat{\theta}, \theta) = C(\tilde{\theta}), \text{ 其中误差 } \tilde{\theta} = \hat{\theta} - \theta.$$

那么，我们可以将与此代价函数相对应的平均风险表示为：

$$\bar{R} = \int_{\Theta} \int_{\mathcal{Z}} C(\tilde{\theta}) p(\mathbf{z}, \theta) d\mathbf{z} d\theta$$

**Bayes 估值准则**是：取得使平均代价  $\bar{R}$  最小的估值。

为简化估值准则，我们可作下述的推导：

$$\bar{R} = \int_{\Theta} \int_{\mathcal{Z}} C(\tilde{\theta}) p(\mathbf{z}, \theta) d\mathbf{z} d\theta = \int_{\Theta} \int_{\mathcal{Z}} C(\tilde{\theta}) p(\theta | \mathbf{z}) p(\mathbf{z}) d\mathbf{z} d\theta = \int_{\mathcal{Z}} p(\mathbf{z}) \left[ \int_{\Theta} C(\tilde{\theta}) p(\theta | \mathbf{z}) d\theta \right] d\mathbf{z}$$

令：  $\bar{R}(\mathbf{z}) = \int_{\Theta} C(\tilde{\theta}) p(\theta|\mathbf{z}) d\theta$ ，称为“给定  $\mathbf{z}$  时的条件风险”。

$$\text{则： } \bar{R} = \int_{\mathcal{Z}} \bar{R}(\mathbf{z}) p(\mathbf{z}) d\mathbf{z}$$

由于  $p(\mathbf{z}) > 0$ ，所以 Bayes 估值准则等效于取得使条件风险  $\bar{R}(\mathbf{z})$  最小的估值。

为了进一步给出 Bayes 估值的计算公式，我们需要先确定代价函数的表达式。

现在，就让我们来讨论几种代价函数下的 Bayes 估值。为简明扼要，我们以下将采用一维信号参量  $\theta$  来代替信号参量矢量  $\boldsymbol{\theta}$ 。

#### ◆代价函数的选择：

我们常用的代价函数有：平方误差代价、绝对误差代价、均匀误差代价、对称下凸代价、对称非下凸代价等。

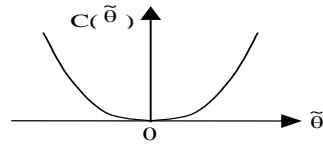
下面我们给出它们的具体表达式及相应的估值公式。

#### ◆平方误差代价：

$$C(\tilde{\theta}) = \tilde{\theta}^2 = (\hat{\theta} - \theta)^2$$

此时的条件风险为：

$$\bar{R}(\mathbf{z}) = \int_{\Theta} C(\tilde{\theta}) p(\theta|\mathbf{z}) d\theta = \int_{\Theta} (\hat{\theta} - \theta)^2 p(\theta|\mathbf{z}) d\theta。$$



$$\frac{\partial \bar{R}(\mathbf{z})}{\partial \hat{\theta}} = 2 \int_{\Theta} (\hat{\theta} - \theta) p(\theta|\mathbf{z}) d\theta = 2 \left[ \hat{\theta} \int_{\Theta} p(\theta|\mathbf{z}) d\theta - \int_{\Theta} \theta p(\theta|\mathbf{z}) d\theta \right] = 2 \left[ \hat{\theta} - \int_{\Theta} \theta p(\theta|\mathbf{z}) d\theta \right]$$

$$\text{令 } \left. \frac{\partial \bar{R}(\mathbf{z})}{\partial \hat{\theta}} \right|_{\hat{\theta} = \hat{\theta}_{\text{ms}}} = 0, \text{ 则我们可得： } \hat{\theta}_{\text{ms}} = \int_{\Theta} \theta p(\theta|\mathbf{z}) d\theta = E(\theta|\mathbf{z}), \text{ 称为“条件均值”。}$$

由于  $p(\theta|\mathbf{z})$  是后验概率密度，故  $\hat{\theta}_{\text{ms}}$  又被称为“后验均值”。

将  $\hat{\theta}_{\text{ms}}$  代入条件风险  $\bar{R}(\mathbf{z})$  的表达式，我们得到：

$$\bar{R}_{\text{ms}}(\mathbf{z}) = \int_{\Theta} [\theta - E(\theta|\mathbf{z})]^2 p(\theta|\mathbf{z}) d\theta = \sigma_{\theta}^2(\mathbf{z}), \text{ 称为“条件方差”或“后验方差”。}$$

可见，此时的条件风险  $\bar{R}_{\text{ms}}(\mathbf{z})$  就是信号参量的均方误差。

所以，该 Bayes 估值又被称为“最小均方估值”（或“最小方差估值”），且：

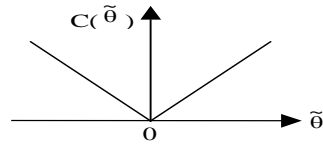
$$\hat{\theta}_{\text{ms}} = \text{后验概率密度 } p(\theta|\mathbf{z}) \text{ 的均值。}$$

#### ◆绝对误差代价：

$$C(\tilde{\theta}) = |\tilde{\theta}| = |\hat{\theta} - \theta|$$

此时的条件风险为：

$$\begin{aligned} \bar{R}(\mathbf{z}) &= \int_{\Theta} |\hat{\theta} - \theta| p(\theta|\mathbf{z}) d\theta \\ &= \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta|\mathbf{z}) d\theta + \int_{\hat{\theta}}^{+\infty} (\theta - \hat{\theta}) p(\theta|\mathbf{z}) d\theta。 \end{aligned}$$



$$\text{故有： } \frac{\partial \bar{R}(\mathbf{z})}{\partial \hat{\theta}} = \int_{-\infty}^{\hat{\theta}} p(\theta|\mathbf{z}) d\theta - \int_{\hat{\theta}}^{+\infty} p(\theta|\mathbf{z}) d\theta。$$

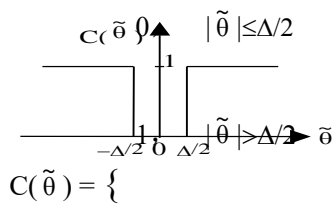
$$\text{令 } \left. \frac{\partial \bar{R}(\mathbf{z})}{\partial \hat{\theta}} \right|_{\hat{\theta} = \hat{\theta}_{\text{abs}}} = 0, \text{ 则我们可得： } \int_{-\infty}^{\hat{\theta}_{\text{abs}}} p(\theta|\mathbf{z}) d\theta = \int_{\hat{\theta}_{\text{abs}}}^{+\infty} p(\theta|\mathbf{z}) d\theta。$$

所以，我们可知：



$\hat{\theta}_{\text{abs}} =$  后验概率密度  $p(\theta|\mathbf{z})$  的中值 (中位数)。

◆均匀误差代价:



此时的条件风险为:

$$\bar{R}(\mathbf{z}) = \int_{-\infty}^{\hat{\theta} - \frac{\Delta}{2}} 1 \cdot p(\theta|\mathbf{z}) d\theta + \int_{\hat{\theta} + \frac{\Delta}{2}}^{+\infty} 1 \cdot p(\theta|\mathbf{z}) d\theta = 1 - \int_{\hat{\theta} - \frac{\Delta}{2}}^{\hat{\theta} + \frac{\Delta}{2}} p(\theta|\mathbf{z}) d\theta.$$

故, 当  $\Delta$  很小时, 要使条件风险  $\bar{R}(\mathbf{z})$  取最小值  $\Leftrightarrow$  使后验概率密度  $p(\theta|\mathbf{z})$  取最大值。

所以, 我们有:

$\hat{\theta}_{\text{unf}} \approx \hat{\theta}_{\text{map}} =$  后验概率密度  $p(\theta|\mathbf{z})$  的众数 ( $p(\theta|\mathbf{z})$  取最大值时的  $\theta$  值)。

◆对称下凸代价:

若代价函数满足下述二条件:

- 1) 对称性:  $C(\tilde{\theta}) = C(-\tilde{\theta})$ ;
- 2) 下凸性:  $C(b\theta_1 + (1-b)\theta_2) \leq bC(\theta_1) + (1-b)C(\theta_2)$ , 对  $\forall b \in (0, 1)$ ;

且后验密度函数满足下述条件:

- 3) 对条件均值  $\hat{\theta}_{\text{ms}}$  对称:  $p_{\theta}(\theta - \hat{\theta}_{\text{ms}}|\mathbf{z}) = p_{\theta}(\hat{\theta}_{\text{ms}} - \theta|\mathbf{z})$

令:  $\delta = \theta - \hat{\theta}_{\text{ms}}$ ,

则我们有:

$$\bar{R}(\mathbf{z}) = \int_{-\infty}^{+\infty} C(\tilde{\theta}) p_{\theta}(\theta|\mathbf{z}) d\theta = \int_{-\infty}^{+\infty} C(\theta - \hat{\theta}) p_{\theta}(\theta|\mathbf{z}) d\theta = \int_{-\infty}^{+\infty} C(\delta + \hat{\theta}_{\text{ms}} - \hat{\theta}) p(\delta|\mathbf{z}) d\delta \quad (*)$$

$$= \int_{-\infty}^{+\infty} C(-\delta - \hat{\theta}_{\text{ms}} + \hat{\theta}) p(\delta|\mathbf{z}) d\delta = \int_{-\infty}^{+\infty} C(\delta - \hat{\theta}_{\text{ms}} + \hat{\theta}) p(\delta|\mathbf{z}) d\delta \quad (**)$$

利用 (\*) 和 (\*\*) 两式, 我们可得:

$$\bar{R}(\mathbf{z}) = \int_{-\infty}^{+\infty} \left[ \frac{1}{2} C(\delta + \hat{\theta}_{\text{ms}} - \hat{\theta}) + \frac{1}{2} C(\delta - \hat{\theta}_{\text{ms}} + \hat{\theta}) \right] p(\delta|\mathbf{z}) d\delta \geq \int_{-\infty}^{+\infty} C(\delta) p(\delta|\mathbf{z}) d\delta$$

当  $\hat{\theta} = \hat{\theta}_{\text{ms}}$  时, 上式中等号成立。

可见: 当前述三个条件 1) - 3) 得到满足时, Bayes 估值就是最小均方估值  $\hat{\theta}_{\text{ms}}$  (后验均值)。

◆对称非下凸代价:

将上面的第二个条件 2) 改为非下凸性, 即:

$$\text{当 } \tilde{\theta} \geq 0 \text{ 时, } \frac{dC(\tilde{\theta})}{d\tilde{\theta}} \geq 0; \quad \text{当 } \tilde{\theta} < 0 \text{ 时, } \frac{dC(\tilde{\theta})}{d\tilde{\theta}} \leq 0.$$

并且要求:  $\lim_{\tilde{\theta} \rightarrow \infty} C(\tilde{\theta}) p(\tilde{\theta}|\mathbf{z}) = 0$ 。

则可证明: 此时的 Bayes 估值亦是最小均方估值  $\hat{\theta}_{\text{ms}}$  (后验均值)。

◆结论:

当后验概率密度  $p(\theta|\mathbf{z})$  对称于其均值 (如高斯分布) 时, 很大一类代价函数的 Bayes 估值就是最小

均方估值  $\hat{\theta}_{ms}$  (后验均值)。

◆例：观测样本为： $z_i = m + n_i$ ,  $i=1,2,\dots,N$  且相互独立。 $m \sim N(0, \sigma_m^2)$ ,  $n_i \sim N(0, \sigma_n^2)$ 。

求  $m$  的 Bayes 估计量  $\hat{m}$ 。

解：由于后验概率密度  $p(m|\mathbf{z}) = p(m)p(\mathbf{z}|m)/p(\mathbf{z})$

$$\text{又由题意： } p(m) = \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{m^2}{2\sigma_m^2}\right)$$

$$p(\mathbf{z}|m) = \prod_{i=1}^N p(z_i|m) = \left(\frac{1}{\sqrt{2\pi\sigma_n^2}}\right)^N \exp\left(-\frac{\sum_{i=1}^N (z_i - m)^2}{2\sigma_n^2}\right)$$

$p(\mathbf{z})$  与  $m$  无关。

故可得：

$$\begin{aligned} p(m|\mathbf{z}) &= A(\mathbf{z}) \exp\left[-\frac{1}{2}\left[\frac{m^2}{\sigma_m^2} + \frac{1}{\sigma_n^2} \sum_{i=1}^N (z_i - m)^2\right]\right], \text{ 其中 } A(\mathbf{z}) \text{ 表示不含 } m \text{ 的所有项。} \\ &= B(\mathbf{z}) \exp\left[-\frac{1}{2}\left[\left(\frac{1}{\sigma_m^2} + \frac{N}{\sigma_n^2}\right)m^2 - \frac{2}{\sigma_n^2}\left(\sum_{i=1}^N z_i\right)m\right]\right], \text{ 其中 } B(\mathbf{z}) \text{ 表示不含 } m \text{ 的所有项。} \end{aligned}$$

$$\text{令： } \sigma^2 = \left(\frac{1}{\sigma_m^2} + \frac{N}{\sigma_n^2}\right)^{-1} = \frac{\sigma_m^2 \sigma_n^2}{\sigma_n^2 + N\sigma_m^2}, \quad \bar{z} = \frac{1}{N} \sum_{i=1}^N z_i。$$

则有：

$$\begin{aligned} p(m|\mathbf{z}) &= B(\mathbf{z}) \exp\left[-\frac{1}{2}\left[\frac{1}{\sigma^2} m^2 - \frac{2N\bar{z}}{\sigma_n^2} m\right]\right] \\ &= C(\mathbf{z}) \exp\left[-\frac{1}{2\sigma^2}\left[m - \frac{N\sigma^2}{\sigma_n^2} \bar{z}\right]^2\right], \text{ 其中 } C(\mathbf{z}) \text{ 表示不含 } m \text{ 的所有项。} \end{aligned}$$

可见：后验概率密度  $p(m|\mathbf{z})$  为一高斯分布。

$$\text{所以， } \hat{m}_{ms} = \hat{m}_{abs} = \hat{m}_{map} = \frac{N\sigma^2}{\sigma_n^2} \bar{z} = \frac{\sigma_m^2}{\sigma_n^2 + N\sigma_m^2} \sum_{i=1}^N z_i。$$

此时的条件方差  $\bar{R}_{ms}(\mathbf{z}) = E[(m - \hat{m}_{ms})^2 | \mathbf{z}] = \sigma^2 = \frac{\sigma_m^2 \sigma_n^2}{\sigma_n^2 + N\sigma_m^2}$ ，由于它与观测值  $\mathbf{z}$  无关，故  $\bar{R}_{ms}(\mathbf{z}) = \bar{R}_{ms}$  且它就是估值方差  $E[(\hat{m}_{ms} - m)^2]$ 。

#解毕

下面我们来讨论一下本例中信噪比与估值及其方差的关系：

$$\text{令信噪比 } r = \frac{\sigma_m^2}{\sigma_n^2}, \text{ 则估值 } \hat{m}_{ms} = \frac{r}{1 + Nr} \sum_{i=1}^N z_i, \text{ 其方差 } \sigma^2 = \frac{r}{1 + Nr} \sigma_n^2。$$

可见：1) 当信噪比  $r$  很大时，估值  $\hat{m}_{ms} \rightarrow$  样本均值  $\bar{z} = \frac{1}{N} \sum_{i=1}^N z_i$ ，其方差  $\sigma^2 \rightarrow \frac{1}{N} \sigma_n^2$ 。换句话说，此时的估值基本上由观测值来确定（因为信噪比高，观测值具有较高的可信度）。

2) 当信噪比  $r$  很小时, 估值  $\hat{m}_{ms} \approx r \sum_{i=1}^N z_i \rightarrow 0 (= \bar{m})$ , 其方差  $\sigma^2 \rightarrow r \sigma_n^2 = \sigma_m^2$ 。换句话说, 此时的估值与观测值基本无关 (因为信噪比太低, 观测值无可信度可言) 而由信号参量的先验知识来确定 (此时先验知识更可信)。

### 2.2.2 极大极小估值

由上例可知: 求后验概率密度  $p(\mathbf{m}|\mathbf{z})$  需要知道信号参数的先验概率  $p(\mathbf{m})$ 。

如果先验概率未知, 则可采用极大极小估值方法:

寻求先验概率  $p(\theta)$  的最不利分布, 再确定 Bayes 估值。

◆例: 在上例中, 若令  $\sigma_m^2 \rightarrow \infty$ , 则信号参量  $\mathbf{m} \rightarrow$  均匀分布 (从  $-\infty$  到  $+\infty$ ), 即为最不利分布。

此时, 我们得到极大极小估值:  $\hat{m}_{MM} = \lim_{\sigma_m^2 \rightarrow \infty} \hat{m}_{ms} = \frac{1}{N} \sum_{i=1}^N z_i$

相应的平均风险为:  $\bar{R}_{MM} = \lim_{r \rightarrow \infty} \bar{R}_{ms} = \frac{1}{N} \sigma_n^2$ 。

可见: 极大极小估值的风险 ( $\frac{1}{N} \sigma_n^2$ ) 大于 (等于) Bayes 估值的风险 ( $\frac{r}{1+Nr} \sigma_n^2$ )。换句话说, 极大极小风险是 Bayes 风险的极大值, 说明和以前讨论信号检测时的极大极小准则意义是一样的。

### 2.2.3 最大后验估值

前面: 我们曾证明过均匀代价下的 Bayes 估值就是最大后验估值 (MAP, 后验众数)。

实际上, 如果规定不出代价函数  $C(\tilde{\theta})$  的表达式, 我们亦可采用最大后验估值:

$$\hat{\theta}_{\text{map}} = \max_{\theta=\hat{\theta}_{\text{map}}} p(\theta|\mathbf{z})$$

又  $p(\theta|\mathbf{z})$  取最大值的必要条件是:  $\left. \frac{\partial p(\theta|\mathbf{z})}{\partial \theta} \right|_{\theta=\hat{\theta}_{\text{map}}} = 0$  等效于  $\left. \frac{\partial \ln p(\theta|\mathbf{z})}{\partial \theta} \right|_{\theta=\hat{\theta}_{\text{map}}} = 0$ ,

而  $p(\theta|\mathbf{z}) = \frac{p(\theta) \cdot p(\mathbf{z}|\theta)}{p(\mathbf{z})}$  (以便将观测矢量  $\mathbf{z}$  和信号参量  $\theta$  的知识分开)

从而可知:  $p(\theta|\mathbf{z})$  取最大值的必要条件为

$$\left[ \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} + \frac{\partial \ln p(\theta)}{\partial \theta} \right] \bigg|_{\theta=\hat{\theta}_{\text{map}}} = 0, \text{ 称为 “最大后验方程”。}$$

观测知识    先验知识

### 2.2.4 最大似然估值

前面: 先验未知  $\Rightarrow$  极大极小估值; 代价未知  $\Rightarrow$  最大后验估值。

现在: 1) 先验、代价均未知  $\Rightarrow$  最大似然估值;

2) 若参量是非随机的, 无先验分布可言  $\Rightarrow$  只能采用最大似然估值。

最大似然估值:  $\hat{\theta}_{\text{ML}} = \max_{\theta=\hat{\theta}_{\text{ML}}} p(\mathbf{z}|\theta)$

似然函数取最大值的必要条件是:  $\left. \frac{\partial p(\mathbf{z}|\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_{\text{ML}}} = 0$  等效于  $\left. \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_{\text{ML}}} = 0$ ,

这就是 “似然方程”。

◆例：在前例中，求  $\hat{m}_{ML}$  = ?

解：由于似然函数  $p(\mathbf{z}|\mathbf{m}) = \left( \frac{1}{\sqrt{2\pi\sigma_n^2}} \right)^N \exp\left(-\frac{\sum_{i=1}^N (z_i - m)^2}{2\sigma_n^2}\right)$

则由  $\left. \frac{\partial \ln p(\mathbf{z}|\mathbf{m})}{\partial m} \right|_{m=\hat{m}_{ML}} = 0$  可得：  $\frac{1}{\sigma_n^2} \sum_{i=1}^N (z_i - \hat{m}_{ML}) = 0$  即  $\hat{m}_{ML} = \frac{1}{N} \sum_{i=1}^N z_i = \bar{z}$ 。

#解毕

可见：此例中， $\hat{m}_{ML} = \hat{m}_{MM}$ 。因为它们都没有利用信号参量的先验信息，所以，其估值方差比 Bayes 估值方差要大。当然，当信噪比  $r$  很大时， $\hat{m}_{ML} \approx \hat{m}_{ms}$ ，因为如前所述，此时的信息主要来自观测量。

## 2.3 估计量的性质

估计量是观测值  $\mathbf{z}$  的函数： $\hat{\theta}(\mathbf{z})$ 。

$\mathbf{z}$  是随机变量  $\Rightarrow \hat{\theta}$  也是随机变量。一组观测值给出一个估计值， $N$  组观测值给出  $N$  个估计值。所以，我们下面来讨论如何评价它们的质量。

### ① 无偏估计量：

对非随机参量，无偏估计量应满足： $E(\hat{\theta}) = \theta$ （真值），这里： $E(\hat{\theta}) = \int_{\mathcal{Z}} \hat{\theta} p(\mathbf{z}|\theta) d\mathbf{z}$ ；

对随机参量，无偏估计量应满足： $E(\hat{\theta}) = E(\theta)$ ，

这里： $E(\hat{\theta}) = \int_{\Theta} \int_{\mathcal{Z}} \hat{\theta} p(\mathbf{z}, \theta) d\mathbf{z} d\theta = \int_{\mathcal{Z}} \hat{\theta} p(\mathbf{z}) d\mathbf{z}$ 。

在 § 2.1.1 节的例题中： $E(m) = 0$ ， $\hat{m}_{ms} = \frac{N\sigma^2}{\sigma_n^2} \bar{z}$ ， $\hat{m}_{ML} = \bar{z}$ 。

而  $E(z_i) = E(m + n_i) = E(m) + E(n_i) = 0 \Rightarrow E(\bar{z}) = 0$

由于  $E(\hat{m}_{ms}) = E(\hat{m}_{ML}) = 0 = E(m)$ ，故  $\hat{m}_{ms}$  和  $\hat{m}_{ML}$  均为  $m$  的无偏估计。

### ② 一致估计量：

基于  $N$  个观测值组成的观测矢量  $\mathbf{z}_N$  作估值： $\hat{\theta}(\mathbf{z}_N)$ 。

希望  $N \uparrow$  时，（对非随机参量而言）估值趋近于真值： $\hat{\theta}(\mathbf{z}_N) \rightarrow \theta$ ，即

对任意小的  $\varepsilon > 0$ ，总有  $\lim_{N \rightarrow \infty} P(|\hat{\theta}(\mathbf{z}_N) - \theta| < \varepsilon) = 1$

称为“ $\hat{\theta}$  依概率收敛于  $\theta$ ”，此时的  $\hat{\theta}$  即称为“一致估计量”。

前例中：假定  $m$  为非随机参量。最大似然估值  $\hat{m}_{ML} = \bar{z}$ ，其方差为  $\frac{1}{N} \sigma_n^2$ 。

从而，当  $N \rightarrow \infty$  时，其方差  $\rightarrow 0$ 。故  $\hat{m}_{ML}$  是  $m$  的一致估计量。

### ③ 充分估计量：

如果： $p(\mathbf{z}|\theta) = p(\hat{\theta}(\mathbf{z})|\theta) \cdot h(\mathbf{z})$ ，其中  $h(\mathbf{z}) \geq 0$  且与  $\theta$  无关。

那么， $\hat{\theta}(\mathbf{z})$  是  $\theta$  的一个充分估计量。

前例中：假定  $m$  为非随机参量。最大似然估值  $\hat{m}_{ML} = \bar{z}$ 。

$$\begin{aligned}
p(\mathbf{z}|\mathbf{m}) &= \left( \frac{1}{\sqrt{2\pi\sigma_n^2}} \right)^N \exp\left(-\frac{\sum_{i=1}^N (z_i - m)^2}{2\sigma_n^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_n^2/N}} \exp\left(-\frac{(\hat{m}_{ML} - m)^2}{2\sigma_n^2/N}\right) \underbrace{\sqrt{2\pi\sigma_n^2/N} \left( \frac{1}{\sqrt{2\pi\sigma_n^2}} \right)^N \exp\left(-\frac{\sum_{i=1}^N z_i^2 - N\hat{m}_{ML}^2}{2\sigma_n^2}\right)}_{h(\mathbf{z})} \\
&\quad \underbrace{p(\hat{m}_{ML}|\mathbf{m})}_{\text{故 } \hat{m}_{ML} \text{ 是 } m \text{ 的一个充分估计量。}}
\end{aligned}$$

故  $\hat{m}_{ML}$  是  $m$  的一个充分估计量。

#### ④ 有效估计量:

具有最小方差的无偏估计量，称为“有效估计量”。

这里的最小方差即是下一节将要讨论的 Cramér-Rao 界限:

$$E((\hat{\theta} - \theta)^2) \geq \text{Cramér-Rao 界限}。$$

## 2.4 Cramér-Rao 不等式

按未知参量的性质分两种情况来讨论估计量的最小方差，即非随机参量和随机参量。

### 2.4.1 非随机参量

由无偏估计量性质，我们有： $E(\hat{\theta} - \theta) = \int_{\mathcal{Z}} (\hat{\theta} - \theta)p(\mathbf{z}|\theta)d\mathbf{z} = 0$

上式两边对  $\theta$  求导（假定求导和积分的次序可交换）:

$$\int_{\mathcal{Z}} (\hat{\theta} - \theta) \frac{\partial p(\mathbf{z}|\theta)}{\partial \theta} d\mathbf{z} + \int_{\mathcal{Z}} (-1)p(\mathbf{z}|\theta)d\mathbf{z} = 0$$

而  $\int_{\mathcal{Z}} p(\mathbf{z}|\theta)d\mathbf{z} = 1$ ，故可得： $\int_{\mathcal{Z}} (\hat{\theta} - \theta) \frac{\partial p(\mathbf{z}|\theta)}{\partial \theta} d\mathbf{z} = 1$ 。

$$\text{又 } \frac{\partial p(\mathbf{z}|\theta)}{\partial \theta} = \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \cdot p(\mathbf{z}|\theta) = \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \sqrt{p(\mathbf{z}|\theta)} \cdot \sqrt{p(\mathbf{z}|\theta)}$$

故，可重写前面的等式为： $\int_{\mathcal{Z}} \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \sqrt{p(\mathbf{z}|\theta)} (\hat{\theta} - \theta) \sqrt{p(\mathbf{z}|\theta)} d\mathbf{z} = 1$

又根据 Schwarz 不等式： $\int g^2(x)dx \cdot \int h^2(x)dx \geq [\int g(x)h(x)dx]^2$

并且当且仅当  $g(x) = K \cdot h(x)$  时，等号成立。

在前面的等式中，令： $\frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \sqrt{p(\mathbf{z}|\theta)} = g(x)$ ， $(\hat{\theta} - \theta) \sqrt{p(\mathbf{z}|\theta)} = h(x)$ ，

则，我们有： $\int_{\mathcal{Z}} \left[ \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \right]^2 p(\mathbf{z}|\theta) d\mathbf{z} \cdot \int_{\mathcal{Z}} (\hat{\theta} - \theta)^2 p(\mathbf{z}|\theta) d\mathbf{z} \geq 1$

即： $E\left[\left(\frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta}\right)^2 | \theta\right] \cdot E((\hat{\theta} - \theta)^2 | \theta) \geq 1$

所以，我们得到：估值方差  $\text{Var}(\hat{\theta} - \theta) = E((\hat{\theta} - \theta)^2 | \theta) \geq 1 / E\left[\left(\frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta}\right)^2 | \theta\right]$ 。

此不等式的右边（即  $1/E\left(\left[\frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta}\right]^2|\theta\right)$ ）就是无偏估值方差的下限，称为“Cramér-Rao 界限”。

当且仅当  $\frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \sqrt{p(\mathbf{z}|\theta)} = (\hat{\theta} - \theta) \sqrt{p(\mathbf{z}|\theta)} \cdot K(\theta)$  即

$$\frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} = (\hat{\theta} - \theta) \cdot K(\theta)$$

时，等号成立。此时，估值方差达到最小值，也就是说，这样的估值为有效估值。这也是有效估值存在的充要条件。

我们也可以推导出 Cramér-Rao 界限的另一种表达形式：

$$\text{由 } \int_{\mathcal{Z}} p(\mathbf{z}|\theta) d\mathbf{z} = 1 \text{ 对 } \theta \text{ 求导得： } \int_{\mathcal{Z}} \frac{\partial p(\mathbf{z}|\theta)}{\partial \theta} d\mathbf{z} = 0 \text{ 或 } \int_{\mathcal{Z}} \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \cdot p(\mathbf{z}|\theta) d\mathbf{z} = 0$$

$$\text{两边对 } \theta \text{ 再求导得： } \int_{\mathcal{Z}} \frac{\partial^2 \ln p(\mathbf{z}|\theta)}{\partial \theta^2} \cdot p(\mathbf{z}|\theta) d\mathbf{z} + \int_{\mathcal{Z}} \left[ \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \right]^2 \cdot p(\mathbf{z}|\theta) d\mathbf{z} = 0$$

$$\text{即 Cramér-Rao 界限 } E\left[\left[\frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta}\right]^2|\theta\right) = -E\left(\frac{\partial^2 \ln p(\mathbf{z}|\theta)}{\partial \theta^2}|\theta\right)$$

$$\text{则，估值方差 } \text{Var}(\hat{\theta} - \theta) = E((\hat{\theta} - \theta)^2|\theta) \geq -1/E\left(\frac{\partial^2 \ln p(\mathbf{z}|\theta)}{\partial \theta^2}|\theta\right)。$$

下面我们来看一看有效估值和最大似然（ML）估值之间的关系：

$$\text{根据似然方程（ML 估值的必要条件）： } \left. \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_{ML}} = 0$$

$$\text{若有效估值存在，则： } \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} = (\hat{\theta} - \theta) \cdot K(\theta)$$

故，此时由上两式可得： $\left. (\hat{\theta} - \theta) \cdot K(\theta) \right|_{\theta=\hat{\theta}_{ML}} = 0$  即  $\hat{\theta} = \hat{\theta}_{ML}$ 。这说明：所取得的有效估值就是最大似然估值。

若有效估值不存在，则我们不知道最大似然估值有多好。事实上，在这种情况下我们还可以取得比最大似然估值更好（方差更小）的估值，只是我们还未找出得到它们的一般规则。

在一般条件下，可以证明最大似然估值具有下述性质：

- ① 最大似然估值是一致估计量；
- ② 最大似然估值是渐近有效估计量；

设  $\hat{\theta}_{eff}$  是有效估值， $\hat{\theta}$  是我们通过某种方法得到的无偏估值，

$$\text{则定义估值效率为： } \eta = \frac{E\left([\hat{\theta}_{eff} - \theta]^2\right)}{E\left([\hat{\theta} - \theta]^2\right)} = \frac{\text{Var}(\hat{\theta}_{eff})}{\text{Var}(\hat{\theta})}。 \text{显然， } \eta \leq 1。$$

所谓“渐近有效估值”就是当观测样本总数  $N$  无限增加时估值效率趋于 1，即

$$\lim_{N \rightarrow \infty} \eta = 1。$$

$$\text{利用 Cramér-Rao 界限，可写为： } \lim_{N \rightarrow \infty} \frac{1}{E\left(\left[\frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta}\right]^2\right) \text{Var}(\hat{\theta})} = 1。$$

- ③ 最大似然估值是渐近高斯分布的。

在前面的例题 ( $z_i = m + n_i$ , 假定  $m$  为非随机参量,  $n_i \sim N(0, \sigma_n^2)$ ) 中, 我们曾得到:

$$\text{无偏最大似然估值 } \hat{m}_{ML} = \frac{1}{N} \sum_{i=1}^N z_i \text{ 和 } \ln p(\mathbf{z}|m) = K - \frac{1}{2\sigma_n^2} \sum_{i=1}^N (z_i - m)^2。$$

$$\text{从而, 可得: } \frac{\partial \ln p(\mathbf{z}|m)}{\partial m} = \frac{1}{\sigma_n^2} \sum_{i=1}^N (z_i - m) = \frac{N}{\sigma_n^2} (\hat{m}_{ML} - m)$$

可见, 有效估值存在的条件得到满足, 即等式  $\frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} = (\hat{\theta} - \theta) \cdot K(\theta)$  成立。此时,  $\hat{m}_{ML}$  就是有效估值, 其方差为最小方差, 即为

$$\text{Cramér-Rao 界限} = 1 / E\left(\frac{\partial^2 \ln p(\mathbf{z}|m)}{\partial m^2} | m\right) = \frac{\sigma_n^2}{N}。$$

另一方面, 由于  $z_i \sim N(m, \sigma_n^2)$ , 故  $\hat{m}_{ML} \sim N(m, \frac{\sigma_n^2}{N})$ , 因而可以核实最大似然估值  $\hat{m}_{ML}$  的方差  $\frac{\sigma_n^2}{N}$  确实达到了 Cramér-Rao 界限, 是有效估值。

## 2.4.2 随机参量

对随机参量  $\theta$ , 我们也可推导出它的 Cramér-Rao 界限的表达式。

$$\text{由 } E(\hat{\theta} - \theta) = \int_{\mathcal{Z}} (\hat{\theta} - \theta) p(\mathbf{z}|\theta) d\mathbf{z}$$

$$\text{两边乘以 } p(\theta) \text{ 得: } E(\hat{\theta} - \theta) p(\theta) = \int_{\mathcal{Z}} (\hat{\theta} - \theta) p(\mathbf{z}|\theta) p(\theta) d\mathbf{z}$$

$$\text{等式两边对 } \theta \text{ 求导: } \frac{\partial}{\partial \theta} [E(\hat{\theta} - \theta) p(\theta)] = -p(\theta) \int_{\mathcal{Z}} p(\mathbf{z}|\theta) d\mathbf{z} + \int_{\mathcal{Z}} (\hat{\theta} - \theta) \frac{\partial p(\mathbf{z}, \theta)}{\partial \theta} d\mathbf{z}$$

$$\text{即 } \frac{\partial}{\partial \theta} [E(\hat{\theta} - \theta) p(\theta)] = -p(\theta) + \int_{\mathcal{Z}} (\hat{\theta} - \theta) \frac{\partial p(\mathbf{z}, \theta)}{\partial \theta} d\mathbf{z}$$

$$\text{两边再求积分 } \int_{-\infty}^{+\infty} d\theta: [E(\hat{\theta} - \theta) p(\theta)]|_{-\infty}^{+\infty} = -1 + \int_{-\infty}^{+\infty} \int_{\mathcal{Z}} (\hat{\theta} - \theta) \frac{\partial p(\mathbf{z}, \theta)}{\partial \theta} d\mathbf{z} d\theta$$

$$\text{如果 } \lim_{\theta \rightarrow -\infty} [E(\hat{\theta} - \theta) \cdot p(\theta)] = 0 \text{ 且 } \lim_{\theta \rightarrow +\infty} [E(\hat{\theta} - \theta) \cdot p(\theta)] = 0,$$

$$\text{则, } \int_{-\infty}^{+\infty} \int_{\mathcal{Z}} (\hat{\theta} - \theta) \frac{\partial p(\mathbf{z}, \theta)}{\partial \theta} d\mathbf{z} d\theta = 1$$

此等式和非随机参量中的相类似, 故可采用同样的方法加以推导, 最后得到:

$$E((\hat{\theta} - \theta)^2) \geq 1/E\left(\left[\frac{\partial \ln p(\mathbf{z}, \theta)}{\partial \theta}\right]^2\right) = -1/E\left(\frac{\partial^2 \ln p(\mathbf{z}, \theta)}{\partial \theta^2}\right)$$

当且仅当

$$\frac{\partial \ln p(\mathbf{z}, \theta)}{\partial \theta} = (\hat{\theta} - \theta) \cdot K$$

时, 等号成立。此时, 我们得到估值的最小均方误差为  $1/E\left(\left[\frac{\partial \ln p(\mathbf{z}, \theta)}{\partial \theta}\right]^2\right)$  或  $-1/E\left(\frac{\partial^2 \ln p(\mathbf{z}, \theta)}{\partial \theta^2}\right)$ ,

这就是随机参量的 Cramér-Rao 界限, 达到此界限的估值称为有效估值。

讨论:

① Bayes 估值时, 在平方误差代价下, 我们取得的估值  $\hat{\theta}_{ms}$  使均方误差最小。

若有效估值存在, 则它就是  $\hat{\theta}_{ms}$ 。

② 最大后验估值时，估值  $\hat{\theta}_{\text{map}}$  满足最大后验方程：  $\left. \frac{\partial \ln p(\theta | \mathbf{z})}{\partial \theta} \right|_{\theta=\hat{\theta}_{\text{map}}} = 0$ 。

又  $p(\mathbf{z}, \theta) = p(\theta | \mathbf{z})p(\mathbf{z})$  或  $\ln p(\mathbf{z}, \theta) = \ln p(\theta | \mathbf{z}) + \ln p(\mathbf{z})$

由于  $p(\mathbf{z})$  与  $\theta$  无关，故有：  $\frac{\partial \ln p(\mathbf{z}, \theta)}{\partial \theta} = \frac{\partial \ln p(\theta | \mathbf{z})}{\partial \theta}$ 。

所以，若有效估值存在（即  $\frac{\partial \ln p(\mathbf{z}, \theta)}{\partial \theta} = (\hat{\theta} - \theta) \cdot K$ ），则有：

$$(\hat{\theta} - \theta) \cdot K \Big|_{\theta=\hat{\theta}_{\text{map}}} = 0 \quad \text{即} \quad \hat{\theta} = \hat{\theta}_{\text{map}}。$$

所以，此时的有效估值就是最大后验估值  $\hat{\theta}_{\text{map}}$ 。

综合①及②可见：当有效估值存在时，它就是  $\hat{\theta}_{\text{ms}}$  或  $\hat{\theta}_{\text{map}}$ ，但用最大后验方程求估值比用最小均方误差的方法要方便。

③ 后验密度分布与有效估值的关系：

若有效估值存在（即  $\frac{\partial \ln p(\mathbf{z}, \theta)}{\partial \theta} = (\hat{\theta} - \theta) \cdot K$ ），由于  $\frac{\partial \ln p(\mathbf{z}, \theta)}{\partial \theta} = \frac{\partial \ln p(\theta | \mathbf{z})}{\partial \theta}$ ，故：

$$\frac{\partial \ln p(\theta | \mathbf{z})}{\partial \theta} = (\hat{\theta} - \theta) \cdot K$$

解此微分方程得：  $\ln p(\theta | \mathbf{z}) = -\frac{K}{2}(\hat{\theta} - \theta)^2 + C$ ，其中  $C$  为一常数。

故，  $p(\theta | \mathbf{z}) = \exp(-\frac{K}{2}(\hat{\theta} - \theta)^2 + C) = D \cdot \exp(-\frac{K}{2}(\theta - \hat{\theta})^2)$ ，其中  $D$  为一常数。

可见，对参量  $\theta$  而言，后验概率密度  $p(\theta | \mathbf{z})$  为一高斯分布。

而由②可知：此时的有效估值就是最大后验估值  $\hat{\theta}_{\text{map}}$ 。

所以，我们得到如下的结论：

具有高斯型分布的后验概率密度的估值是有效估值，其估值方差达到 Cramér-Rao 界限。

在前面的例题（ $\mathbf{z}_i = \mathbf{m} + \mathbf{n}_i$ ， $\mathbf{m}$  为随机参量， $\mathbf{n}_i \sim N(0, \sigma_n^2)$ ）中，我们曾得到：

$$p(\mathbf{m} | \mathbf{z}) = C(\mathbf{z}) \exp \left( -\frac{1}{2\sigma^2} \left[ \mathbf{m} - \frac{N\sigma^2}{\sigma_n^2} \bar{\mathbf{z}} \right]^2 \right)， \quad \text{其中 } C(\mathbf{z}) \text{ 表示不含 } \mathbf{m} \text{ 的所有项。}$$

它满足高斯分布条件，故其估值为有效估值。

可核实如下：

$$\text{由 } \frac{\partial \ln p(\mathbf{z}, \mathbf{m})}{\partial \mathbf{m}} = \frac{\partial \ln p(\mathbf{m} | \mathbf{z})}{\partial \mathbf{m}} = -\frac{1}{\sigma^2} \left[ \mathbf{m} - \frac{N\sigma^2}{\sigma_n^2} \bar{\mathbf{z}} \right] \quad (\text{显然，有效估值存在。})$$

$$\text{可得 } \frac{\partial^2 \ln p(\mathbf{z}, \mathbf{m})}{\partial \mathbf{m}^2} = -\frac{1}{\sigma^2}， \quad \text{故 C-R 界限为 } \sigma^2。$$

$$\text{又前面计算得 } \text{Var}(\hat{\theta}_{\text{ms}}) = \text{Var}(\hat{\theta}_{\text{map}}) = \sigma^2，$$

可见， $\hat{\theta}_{\text{ms}}$  和  $\hat{\theta}_{\text{map}}$  均达到了 C-R 界限，都是有效估值。



## 2.5 在高斯信道中单参量信号的估值

假设观测信号为：  $z(t)=s(t,\theta)+n(t)$ ,  $0\leq t\leq T$ 。

其中：  $\theta$  是待估计的未知参量，  $n(t)$  是均值为零、服从高斯分布的噪声。

### 2.5.1 白高斯信道

$n(t)$ ：零均值、白高斯、谱密度  $N_0/2$ 。

◆  $\theta$  为未知非随机参量时，求其估值  $\hat{\theta}_{ML}$ 。

首先，我们需要计算似然函数  $p(z(t)|\theta)$ 。

方法一：利用 § 1.3.1 节求连续信号似然比的方法可得：

$$p(z(t)|\theta) = C \cdot \exp\left(-\frac{1}{N_0} \int_0^T [z(t) - s(t,\theta)]^2 dt\right)$$

故有对数似然函数的偏导数：

$$\frac{\partial \ln p(z(t)|\theta)}{\partial \theta} = \frac{2}{N_0} \int_0^T [z(t) - s(t,\theta)] \frac{\partial s(t,\theta)}{\partial \theta} dt$$

方法二：利用 K-L 展开及求极限的方式：

$$\text{将 } z(t) \text{ 在归一化正交基函数 } \{g_k(t)\} \text{ 上作 K-L 展开: } z(t) = \sum_{k=1}^{\infty} z_k g_k(t)$$

$$\text{其中, } z_k = \int_0^T z(t) g_k(t) dt = \int_0^T [s(t,\theta) + n(t)] g_k(t) dt = s_k(\theta) + n_k$$

由于  $n_k$  为高斯分布，故  $z_k$  亦为高斯分布，且有

$$E(z_k) = s_k(\theta), \quad \text{Var}(z_k) = \text{Var}(n_k) = N_0/2, \quad E[(z_k - \bar{z}_k)(z_l - \bar{z}_l)] = (N_0/2) \delta_{kl}.$$

对观测矢量  $\mathbf{z}_N = (z_1, z_2, \dots, z_N)^T$  而言，我们有：

$$p(\mathbf{z}_N|\theta) = \prod_{k=1}^N \frac{1}{\sqrt{2\pi N_0/2}} \exp\left(-\frac{(z_k - s_k(\theta))^2}{N_0}\right) = \left(\frac{1}{\sqrt{\pi N_0}}\right)^N \exp\left(-\frac{1}{N_0} \sum_{k=1}^N (z_k - s_k(\theta))^2\right)$$

$$\text{因而: } \frac{\partial \ln p(\mathbf{z}_N|\theta)}{\partial \theta} = \frac{2}{N_0} \sum_{k=1}^N [z_k - s_k(\theta)] \frac{\partial s_k(\theta)}{\partial \theta}$$

则，取极限后：

$$\frac{\partial \ln p(z(t)|\theta)}{\partial \theta} = \lim_{N \rightarrow \infty} \frac{\partial \ln p(\mathbf{z}_N|\theta)}{\partial \theta} = \frac{2}{N_0} \sum_{k=1}^{\infty} [z_k - s_k(\theta)] \frac{\partial s_k(\theta)}{\partial \theta}$$

又，根据  $z_k$  与  $s_k(\theta)$  的 K-L 展开式和公式  $\sum_{k=1}^{\infty} g_k(t) g_k(\tau) = \delta(t-\tau)$ ，我们有：

$$\sum_{k=1}^{\infty} z_k \frac{\partial s_k(\theta)}{\partial \theta} = \sum_{k=1}^{\infty} \int_0^T \int_0^T z(t) g_k(t) \frac{\partial s_k(\tau, \theta)}{\partial \theta} g_k(\tau) dt d\tau = \int_0^T z(t) \frac{\partial s(t, \theta)}{\partial \theta} dt$$

$$\text{同理: } \sum_{k=1}^{\infty} s_k(\theta) \frac{\partial s_k(\theta)}{\partial \theta} = \int_0^T s(t) \frac{\partial s(t, \theta)}{\partial \theta} dt$$

故，可得与方法一同样的结果：

$$\frac{\partial \ln p(z(t)|\theta)}{\partial \theta} = \frac{2}{N_0} \int_0^T [z(t) - s(t,\theta)] \frac{\partial s(t,\theta)}{\partial \theta} dt$$

然后，根据最大似然估值的必要条件——似然方程  $\left. \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_{ML}} = 0$ ，我们可通过解下面的方程来求得最大似然估值  $\hat{\theta}_{ML}$ ：

$$\int_0^T [z(t) - s(t, \theta)] \frac{\partial s(t, \theta)}{\partial \theta} dt \Big|_{\theta=\hat{\theta}_{ML}} = 0$$

这就是单参量信号估值时的似然方程。

现在，我们来考察一下此时的 C-R 界限的表达形式。

由对数似然函数的一阶导数及  $z(t)=s(t, \theta)+n(t)$ ，我们可得其二阶导数为

$$\frac{\partial^2 \ln p(z(t)|\theta)}{\partial \theta^2} = -\frac{2}{N_0} \int_0^T \left( \frac{\partial s(t, \theta)}{\partial \theta} \right)^2 dt + \frac{2}{N_0} \int_0^T n(t) \frac{\partial^2 s(t, \theta)}{\partial \theta^2} dt$$

$$\text{由于 } E[n(t)]=0, \text{ 故 } E\left[ \frac{\partial^2 \ln p(z(t)|\theta)}{\partial \theta^2} \right] = -\frac{2}{N_0} \int_0^T \left( \frac{\partial s(t, \theta)}{\partial \theta} \right)^2 dt$$

$$\text{故单参量信号估值时的 C-R 界限为 } -1/E\left[ \frac{\partial^2 \ln p(z(t)|\theta)}{\partial \theta^2} \right] = N_0 / \left[ 2 \int_0^T \left( \frac{\partial s(t, \theta)}{\partial \theta} \right)^2 dt \right]$$

最后，讨论在几个具体应用中的非随机单参量信号估值：

#### a) 信号幅度估值：

此时，信号  $s(t, \theta)$  可表示为

$$s(t, A) = A s(t) \quad 0 \leq t \leq T \quad \text{其中，} A \text{ 是待估值的非随机参量。}$$

由似然方程可得：

$$\hat{A}_{ML} = \frac{\int_0^T z(t) s(t) dt}{\int_0^T s^2(t) dt}$$

若令  $\int_0^T s^2(t) dt = 1$ （信号能量归一化），则  $\hat{A}_{ML} = \int_0^T z(t) s(t) dt$ （相关接收机）。

$\hat{A}_{ML}$  性质：

- 1) 无偏估值。
- 2) 有效估值。

#### b) 信号相位估值：

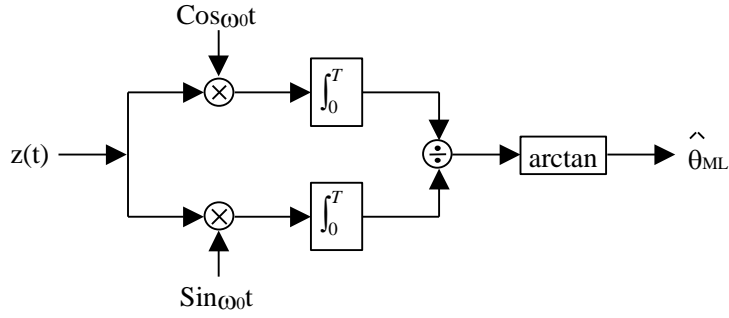
此时，信号  $s(t, \theta)$  可表示为  $s(t, \theta) = A \sin(\omega_0 t + \theta) \quad 0 \leq t \leq T$

其中， $\theta$  是待估值的非随机参量，而  $A$  和  $\omega_0$  是已知常数，假定  $\omega_0 T = k\pi$ 。

由似然方程可得：

$$\hat{\theta}_{ML} = \arctan \left( \frac{\int_0^T z(t) \cos \omega_0 t dt}{\int_0^T z(t) \sin \omega_0 t dt} \right)$$

从而，我们可以用如下的双通道相位测量装置来对信号的相位进行估值：



关于单通道相位测量装置、锁相环等内容，可参考[刘有恒]一书。

$\hat{\theta}_{ML}$  性质：

当信噪比足够大时（参见[张树京]一书），相位估值  $\hat{\theta}_{ML}$  的概率密度为一高斯分布，其均值为  $\theta$ （无偏估值）、方差为  $N_0/(2E_s)$ （有效估值）（C-R 界限 =  $N_0 / \left[ 2 \int_0^T \left( \frac{\partial s(t, \theta)}{\partial \theta} \right)^2 dt \right] = N_0 / \left( 2 \int_0^T [A \cos(\omega_0 t + \theta)]^2 dt \right) = N_0 / (A^2 T) = N_0 / (2E_s) \right)$ 。

c) 信号频率估值：

此时，信号  $s(t, \theta)$  可表示为

$$s(t, \omega) = A \sin(\omega t + \theta) \quad 0 \leq t \leq T$$

其中， $\omega$  是待估值的非随机参量，而  $A$  是已知常数、 $\theta$  是杂散参量。

根据 § 1.5 节中随相信号检测推导结果，在给定信号频率  $\omega$  时有如下似然函数：

$$\begin{aligned} \lambda(z(t)|\omega) &= \frac{p(z(t)|H_1)}{p(z(t)|H_0)} \\ &= \exp\left(-\frac{E_s}{N_0}\right) \cdot I_0\left[\frac{2}{N_0} \sqrt{\frac{2E_s}{T}} \sqrt{y_c^2 + y_s^2}\right] = \exp\left(-\frac{E_s}{N_0}\right) \cdot I_0\left[\frac{2A}{N_0} y\right] \end{aligned}$$

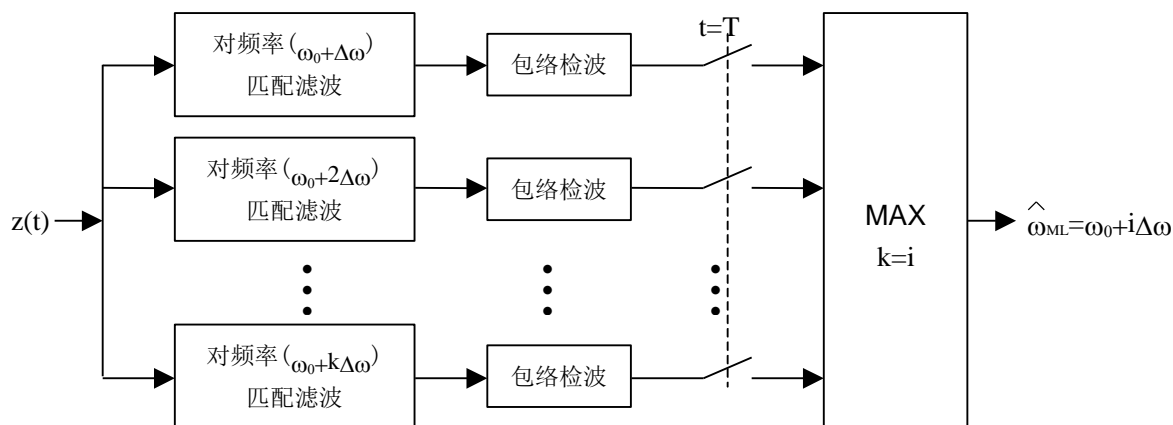
其中， $y = \sqrt{y_c^2 + y_s^2}$ ， $y_c = \int_0^T z(t) \cos \omega t dt$ ， $y_s = \int_0^T z(t) \sin \omega t dt$ ， $E_s = A^2 T / 2$ 。

由于假设  $H_0$  是有用信号不存在时，故  $p(z(t)|H_0)$  与信号频率  $\omega$  无关，令其为一常数  $C$ ，那么，我们可写出如下形式的似然函数

$$p(z(t)|\omega) = p(z(t)|H_1) = C \cdot \exp\left(-\frac{E_s}{N_0}\right) \cdot I_0\left[\frac{2A}{N_0} y\right]$$

因为  $I_0(x)$  是  $x$  的单调增函数，故信号频率的最大似然估值  $\hat{\omega}_{ML}$  就是对应于最大  $y$  值的频率。

由此，我们可构造如下的频率估值装置：



显然,  $\Delta\omega \downarrow \Rightarrow$  估值精度 $\uparrow$ , 但匹配滤波器数目 $\uparrow$ 。

当然,  $\Delta\omega$  应受限于 C-R 界限 (否则, 其估值精度无意义), 通常取  $\Delta\omega = 2\pi/T$  或  $\pi/T$ 。

近似地 (参见[张树京]一书), C-R 界限  $= \frac{3N_0}{2A^2T^3} = \frac{3N_0}{4E_sT^2}$ , 由此亦可看出:  $T \uparrow (\Delta\omega \downarrow) \Rightarrow$  C-R 界限 $\downarrow$  (估值精度 $\uparrow$ )。

◆  $\theta$  为已知先验概率的随机参量时, 求其估值  $\hat{\theta}_{\text{map}}$ 。

首先, 我们通过最大后验方程来求估值:

在最大后验方程  $\left[ \frac{\partial \ln p(\mathbf{z} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{\text{map}}} = 0$  中, 第一项已由前面求出, 故最大后验方程可重写

$$\text{为: } \left[ \frac{2}{N_0} \int_0^T [z(t) - s(t, \theta)] \frac{\partial s(t, \theta)}{\partial \theta} dt + \frac{\partial \ln p(\theta)}{\partial \theta} \right]_{\theta = \hat{\theta}_{\text{map}}} = 0$$

其次, 我们来计算估值的 C-R 界限:

由于  $E[z(t) - s(t, \theta)] = E[n(t)] = 0$ , 故  $\ln p(z(t), \theta)$  二阶偏导数的均值为

$$E \left[ \frac{\partial^2 \ln p(z(t), \theta)}{\partial \theta^2} \right] = -\frac{2}{N_0} E \int_0^T \left( \frac{\partial s(t, \theta)}{\partial \theta} \right)^2 dt + E \frac{\partial^2 \ln p(\theta)}{\partial \theta^2}$$

所以, 估值的 C-R 界限为

$$-1 / E \left[ \frac{\partial^2 \ln p(z(t), \theta)}{\partial \theta^2} \right] = 1 / \left[ \frac{2}{N_0} E \int_0^T \left( \frac{\partial s(t, \theta)}{\partial \theta} \right)^2 dt - E \frac{\partial^2 \ln p(\theta)}{\partial \theta^2} \right]$$

最后, 举一个例子来说明:

若  $s(t, A) = A s(t)$ , 其中  $A$  是待估的随机参量,

且  $A \sim N(0, \sigma_A^2)$ , 即  $p(A) = \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left(-\frac{A^2}{2\sigma_A^2}\right)$ ,

则, 最大后验方程可写为:  $\left[ \frac{2}{N_0} \int_0^T [z(t) - A s(t)] s(t) dt - \frac{A}{\sigma_A^2} \right]_{A = \hat{A}_{\text{map}}} = 0$

从而, 我们可得信号的幅度估值为  $\hat{A}_{\text{map}} = \frac{\int_0^T z(t) s(t) dt}{\int_0^T s^2(t) dt + \frac{N_0}{2\sigma_A^2}}$

令  $\int_0^T s^2(t)dt + \frac{N_0}{2\sigma_A^2} = 1/C$ , 则  $\hat{A}_{\text{map}} = C \int_0^T z(t)s(t)dt$  (相关接收机)。

$\hat{A}_{\text{map}}$  性质:

1) 无偏估值。

由于  $E[z(t)] = E[A s(t) + n(t)] = E(A) s(t) + E[n(t)] = 0$ , 故

$$E(\hat{A}_{\text{map}}) = C \int_0^T E[z(t)] A s(t) dt = 0 = E(A) \quad (\text{无偏估值}).$$

2) 有效估值。

$$\begin{aligned} \frac{\partial \ln p(z(t), A)}{\partial A} &= \frac{2}{N_0} \int_0^T [z(t) - A s(t)] s(t) dt - \frac{A}{\sigma_A^2} \\ &= \frac{2}{N_0} \left[ \int_0^T z(t)s(t)dt - \frac{A}{C} \right] \\ &= \frac{2}{N_0 C} [\hat{A}_{\text{map}} - A] \end{aligned}$$

由此可知:  $\hat{A}_{\text{map}}$  为有效估值。

其方差达到 C-R 界限, 即

$$1 / \left[ \frac{2}{N_0} E \int_0^T \left( \frac{\partial s(t, A)}{\partial A} \right)^2 dt - E \frac{\partial^2 \ln p(A)}{\partial A^2} \right] = 1 / \left[ \frac{2}{N_0} \int_0^T s^2(t)dt + \frac{1}{\sigma_A^2} \right]$$

## 2.5.2 色高斯信道

观测方程:  $z(t) = s(t, \theta) + n(t)$ ,  $0 \leq t \leq T$ ,

其中  $n(t)$  为色高斯平稳噪声, 其相关函数记为  $R_n(t-\tau)$ 。

现在, 我们希望求最大似然估值  $\hat{\theta}_{\text{ML}}$ 。

首先, 需要求出似然函数  $p(z(t)|\theta)$ 。

为简化符号, 我们暂时略去信号参数  $\theta$ , 即  $z(t) = s(t) + n(t)$ 。

现在, 我们利用信号的 K-L 展开式来求似然函数  $p(z(t))$ :

在 § 1.7.1 节 K-L 展开法中, 我们曾得:

$$z(t) = \sum_{k=1}^{\infty} z_k g_k(t), \quad \text{其中 } z_k = \int_0^T z(t) g_k(t) dt.$$

且若我们选择  $\{g_k(t)\}$  满足积分方程:  $\int_0^T R_n(t_1, t_2) g_l(t_2) dt_2 = \lambda_l g_l(t_1)$ ,

则  $z(t)$  的展开式系数  $\{z_k\}$  就互不相关, 即:  $E[(z_k - \bar{z}_k)(z_l - \bar{z}_l)] = \lambda_k \delta_{kl}$ 。

$$\text{由此, 我们可得似然函数: } p(z(t)) = \prod_{k=1}^{\infty} \frac{1}{\sqrt{2\pi\lambda_k}} \exp\left(-\frac{(z_k - s_k)^2}{2\lambda_k}\right)$$

$$\text{取其对数形式: } \ln p(z(t)) = C - \sum_{k=1}^{\infty} \frac{(z_k - s_k)^2}{2\lambda_k} = C - \sum_{k=1}^{\infty} \frac{z_k^2}{2\lambda_k} + \sum_{k=1}^{\infty} \frac{z_k s_k}{\lambda_k} - \sum_{k=1}^{\infty} \frac{s_k^2}{2\lambda_k}$$

其中  $C$  为与信号参量不相关的部分项。

下面，我们来讨论  $\ln p(z(t))$  的各项：

$$-\sum_{k=1}^{\infty} \frac{z_k^2}{2\lambda_k} = -\frac{1}{2} \int_0^T \int_0^T z(t) z(\tau) \sum_{k=1}^{\infty} \frac{g_k(t) g_k(\tau)}{\lambda_k} dt d\tau$$

$$\text{定义反核函数 } R_n^{-1}(t-\tau): \quad \int_0^T R_n^{-1}(t-\tau) R_n(\tau-u) d\tau = \delta(t-u) \quad 0 \leq t, u \leq T$$

$$\text{它可用本征函数和本征值展开为} \quad R_n^{-1}(t-\tau) = \sum_{k=1}^{\infty} \frac{1}{\lambda_k} g_k(t) g_k(\tau)$$

$$\text{故，我们可写出：} \quad -\sum_{k=1}^{\infty} \frac{z_k^2}{2\lambda_k} = -\frac{1}{2} \int_0^T \int_0^T z(t) z(\tau) R_n^{-1}(t, \tau) dt d\tau$$

又，在 § 1.7.1 节 K-L 展开法中已经导出：

$$\sum_{k=1}^{\infty} \frac{z_k s_k}{\lambda_k} = \int_0^T z(t) h(t) dt$$

$$-\sum_{k=1}^{\infty} \frac{s_k^2}{2\lambda_k} = -\frac{1}{2} \int_0^T s(t) h(t) dt$$

下面，我们将用反核函数  $R_n^{-1}(t-\tau)$  来表示  $h(t)$ ：

我们在 § 1.7.1 节 K-L 展开法中已知  $h(t)$  满足下述积分方程

$$\int_0^T R_n(\tau-u) h(u) du = s(\tau)$$

方程两边乘以  $R_n^{-1}(t-\tau)$  并作积分  $\int_0^T d\tau$ ：

$$\int_0^T \int_0^T h(u) \underline{R_n(\tau-u) R_n^{-1}(t-\tau)} du d\tau = \int_0^T s(\tau) R_n^{-1}(t-\tau) d\tau$$

利用反核函数定义  $\int_0^T R_n^{-1}(t-\tau) R_n(\tau-u) d\tau = \delta(t-u)$ ，可得：

$$h(t) = \int_0^T s(\tau) R_n^{-1}(t-\tau) d\tau$$

将此  $h(t)$  表达形式代入可得：

$$\sum_{k=1}^{\infty} \frac{z_k s_k}{\lambda_k} - \sum_{k=1}^{\infty} \frac{s_k^2}{2\lambda_k} = \int_0^T \int_0^T [z(t) - \frac{1}{2} s(t)] s(\tau) R_n^{-1}(t-\tau) dt d\tau$$

所以，我们可将对数似然函数表为

$$\begin{aligned} \ln p(z(t)) &= C - \sum_{k=1}^{\infty} \frac{z_k^2}{2\lambda_k} + \sum_{k=1}^{\infty} \frac{z_k s_k}{\lambda_k} - \sum_{k=1}^{\infty} \frac{s_k^2}{2\lambda_k} \\ &= C - \frac{1}{2} \int_0^T \int_0^T [z(t) z(\tau) - 2 z(t) s(\tau) + s(t) s(\tau)] R_n^{-1}(t-\tau) dt d\tau \\ &= C - \frac{1}{2} \int_0^T \int_0^T [z(t) z(\tau) - z(t) s(\tau) - \underline{z(t) s(\tau)} + s(t) s(\tau)] R_n^{-1}(t-\tau) dt d\tau \\ &= C - \frac{1}{2} \int_0^T \int_0^T [z(t) z(\tau) - z(t) s(\tau) - \underline{z(\tau) s(t)} + s(t) s(\tau)] R_n^{-1}(t-\tau) dt d\tau \\ &= C - \frac{1}{2} \int_0^T \int_0^T [z(t) - s(t)] R_n^{-1}(t-\tau) [z(\tau) - s(\tau)] dt d\tau \end{aligned}$$

考虑到信号参量，最终得到对数似然函数为

$$\ln p(z(t)|\theta) = C - \frac{1}{2} \int_{\theta}^T \int_{\theta}^T [z(t) - s(t, \theta)] R_n^{-1}(t-\tau) [z(\tau) - s(\tau, \theta)] dt d\tau$$

然后，我们根据所得到的对数似然函数列出似然方程以求估值  $\hat{\theta}_{ML}$ 。

对  $\ln p(z(t)|\theta)$  求偏导：

$$\begin{aligned} \frac{\partial \ln p(z(t)|\theta)}{\partial \theta} &= \frac{1}{2} \int_{\theta}^T \int_{\theta}^T \frac{\partial s(t, \theta)}{\partial \theta} R_n^{-1}(t-\tau) [z(\tau) - s(\tau, \theta)] dt d\tau + \\ &\quad + \frac{1}{2} \int_{\theta}^T \int_{\theta}^T [z(t) - s(t, \theta)] R_n^{-1}(t-\tau) \frac{\partial s(\tau, \theta)}{\partial \theta} dt d\tau \\ &= \int_{\theta}^T \int_{\theta}^T [z(t) - s(t, \theta)] R_n^{-1}(t-\tau) \frac{\partial s(\tau, \theta)}{\partial \theta} dt d\tau \\ &= \int_{\theta}^T [z(t) - s(t, \theta)] \frac{\partial h(t, \theta)}{\partial \theta} dt \end{aligned}$$

所以，似然方程可写为  $\int_{\theta}^T [z(t) - s(t, \theta)] \frac{\partial h(t, \theta)}{\partial \theta} dt \Big|_{\theta = \hat{\theta}_{ML}} = 0$

其中  $h(t, \theta) = \int_{\theta}^T s(\tau, \theta) R_n^{-1}(t-\tau) d\tau$  或  $s(t, \theta) = \int_{\theta}^T h(\tau, \theta) R_n(t-\tau) d\tau$ 。

与白高斯噪声情况相比较：

1) 偏导数中用  $h(t, \theta)$  代替了  $s(t, \theta)$ ；

2) 当  $R_n(t-\tau) = (N_0/2) \delta(t-\tau)$  时， $s(t, \theta) = (N_0/2) h(t, \theta)$ 。此时，两种情况下所得结果等效。

最后，我们来求估值的 C-R 界限：

由对数似然函数的一阶偏导数，可得其二阶偏导数为

$$\frac{\partial^2 \ln p(z(t)|\theta)}{\partial \theta^2} = \int_{\theta}^T \left[ -\frac{\partial s(t, \theta)}{\partial \theta} \right] \frac{\partial h(t, \theta)}{\partial \theta} dt + \int_{\theta}^T [z(t) - s(t, \theta)] \frac{\partial^2 h(t, \theta)}{\partial \theta^2} dt$$

求其平均，则右边第二项为零。故，我们得到

$$\text{C-R 界限} = -1 / E \left[ \frac{\partial^2 \ln p(z(t)|\theta)}{\partial \theta^2} \right] = 1 / \int_{\theta}^T \frac{\partial s(t, \theta)}{\partial \theta} \frac{\partial h(t, \theta)}{\partial \theta} dt$$

与白高斯噪声情况相比较：

1) 偏导数之一用  $h(t, \theta)$  代替了  $s(t, \theta)$ ；

2) 白高斯噪声时， $h(t, \theta) = (2/N_0) s(t, \theta)$ 。此时，两种情况下所得结果一样。

◆例：

信号振幅估计。此时， $s(t, \theta)$  可表示为  $s(t, A) = A s(t)$ 。求  $\hat{A}_{ML}$ 。

解：由似然方程：  $\int_{\theta}^T [z(t) - A s(t)] \frac{\partial h(t, A)}{\partial A} dt \Big|_{A = \hat{A}_{ML}} = 0$

$$\text{又：} \quad h(t, A) = \int_{\theta}^T A s(\tau) R_n^{-1}(t-\tau) d\tau = A \int_{\theta}^T s(\tau) R_n^{-1}(t-\tau) d\tau$$

令  $h(t) = \int_{\theta}^T s(\tau) R_n^{-1}(t-\tau) d\tau$ ，则  $h(t, A) = A h(t)$ 。将之代入似然方程，可得：

$$\int_{\theta}^T [z(t) - \hat{A}_{ML} s(t)] h(t) dt = 0$$

故，我们有：

$$\hat{A}_{ML} = \frac{\int_0^T z(t)h(t)dt}{\int_0^T s(t)h(t)dt}$$

此时，

$$\text{C-R 界限} = 1 / \int_0^T s(t) h(t) dt$$

与前面白噪声情况相比，用  $h(t)$  代替了  $s(t)$  而已。且若白噪声时， $h(t) = (2/N_0) s(t)$ ，二者等效。

#解毕

## 2.6 多参量估值

在前面几节中，我们讨论了单参量信号估值的问题。

在实际应用中，我们有时可能需要对多个信号参量同时进行估值，并且这些信号参量可能彼此相关，如：雷达系统中根据目标反射信号的频率和时延来同时估计目标的运动速度和距离。

设由信号的多个参量组成信号参量矢量为： $\theta = (\theta_1, \theta_2, \dots, \theta_M)^T$ 。

而其对应的参量估值矢量为： $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_M)^T$ 。

### ◆Bayes 估值：

在§2.2.1 节中，我们导出 Bayes 估值等效于使条件风险  $\bar{R}(\mathbf{z}) = \int_{\Theta} C(\tilde{\theta})p(\theta|\mathbf{z})d\theta$  最小的估值。当代价函数为平方误差代价函数时，我们利用单参量信号的估值公式即可得到多参量时的最小均方误差估值为：

$$\hat{\theta}_{ms} = \int_{\Theta} \theta p(\theta|\mathbf{z})d\theta = E(\theta|\mathbf{z}), \quad \text{即为矢量形式的条件均值。}$$

$M \times 1 \quad M \times 1 \quad M \times 1$

### ◆最大似然估值：

单参量时，我们有似然方程： $\left. \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_{ML}} = 0$

多参量时，对每个参量  $\theta_i$  我们应有： $\left. \frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta_i} \right|_{\theta_i=\hat{\theta}_{iML}} = 0$ ，写成矢量形式：

$$\left. \nabla_{\theta} \ln p(\mathbf{z}|\theta) \right|_{\theta=\hat{\theta}_{ML}} = \mathbf{0}, \quad \text{其中，偏微分算子 } \nabla_{\theta} = \left( \frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \dots, \frac{\partial}{\partial \theta_M} \right)^T.$$

$M \times 1 \quad M \times 1$

这就是矢量形式的、多参量估值的似然方程。

下面，我们讨论其估值方差（非严格推导）。

分析：对单参量估值情况，当且仅当  $\frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta} = (\hat{\theta} - \theta) \cdot K(\theta)$  时，估值方差达到 C-R 界限，

$$\text{即：} 1/E\left[\left(\frac{\partial \ln p(\mathbf{z}|\theta)}{\partial \theta}\right)^2 \middle| \theta\right] \text{ 或 } -1/E\left(\frac{\partial^2 \ln p(\mathbf{z}|\theta)}{\partial \theta^2} \middle| \theta\right).$$

对多参量估值情况，可推知：倒数将被矩阵求逆所代替，平方将被两矢量相乘所代替。

故，定义一个 Fisher 信息矩阵：

$$\mathbf{F} = E\left(\nabla_{\theta} \ln p(\mathbf{z}|\theta) \cdot \nabla_{\theta}^T \ln p(\mathbf{z}|\theta)\right) = -E\left(\nabla_{\theta} (\nabla_{\theta}^T \ln p(\mathbf{z}|\theta))\right)$$

$M \times M \quad M \times 1 \quad 1 \times M \quad M \times 1 \quad 1 \times M$



其逆矩阵记为 $\Psi$ ，即： $\Psi = \mathbf{F}^{-1}$ ，

而令逆矩阵 $\Psi$ 的对角线元素构成的矢量为： $\Sigma = (\psi_{11}, \psi_{22}, \dots, \psi_{MM})^T$

则，无偏估值的估值方差

$$\text{Var}(\hat{\boldsymbol{\theta}}_{\text{ML}}) \geq \Sigma$$

$M \times 1 \quad M \times 1$

当且仅当  $\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{z}|\boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{\text{ML}} - \boldsymbol{\theta})$  时，等号成立，此时  $\hat{\boldsymbol{\theta}}_{\text{ML}}$  称为“联合有效估值”。

$M \times 1 \quad M \times M \quad M \times 1$

#### ◆最大后验估值：

按照类似上面最大似然估值的推导方法，我们可得：

$$\nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{\theta}|\mathbf{z}) \big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{map}}} = \mathbf{0}, \quad \text{其中，偏微分算子 } \nabla_{\boldsymbol{\theta}} = \left( \frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \dots, \frac{\partial}{\partial \theta_M} \right)^T。$$

$M \times 1 \quad M \times 1$

这就是矢量形式的、多参量估值的最大后验方程。

此时的 Fisher 信息矩阵应改写为：

$$\mathbf{F} = E(\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{z}, \boldsymbol{\theta}) \bullet \nabla_{\boldsymbol{\theta}}^T \ln p(\mathbf{z}, \boldsymbol{\theta})) = -E(\nabla_{\boldsymbol{\theta}} (\nabla_{\boldsymbol{\theta}}^T \ln p(\mathbf{z}, \boldsymbol{\theta})))$$

$M \times M \quad M \times 1 \quad 1 \times M \quad M \times 1 \quad 1 \times M$

相应地，无偏估值的估值方差

$$\text{Var}(\hat{\boldsymbol{\theta}}_{\text{map}}) \geq \Sigma, \quad \text{其中，}\Sigma\text{的定义与最大似然估值时相同。}$$

$M \times 1 \quad M \times 1$

当且仅当  $\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{z}, \boldsymbol{\theta}) = \mathbf{C} \cdot (\hat{\boldsymbol{\theta}}_{\text{map}} - \boldsymbol{\theta})$  时，等号成立，此时  $\hat{\boldsymbol{\theta}}_{\text{map}}$  称为“联合有效估值”。

$M \times 1 \quad M \times M \quad M \times 1$

## 2.7 线性最小均方估值（LMS）

现在我们放宽对信号已知知识的要求，假设：

- ① 对信号的统计知识，我们只知信号参量和观测值的一、二阶矩，即

$$E(\boldsymbol{\theta}), E(\mathbf{z}), \text{Var}(\boldsymbol{\theta}), \text{Var}(\mathbf{z}), \text{Cov}(\boldsymbol{\theta}, \mathbf{z}).$$

$M \times 1 \quad N \times 1 \quad M \times M \quad N \times N \quad M \times N$

- ② 最佳估值是观测值的线性组合：

$$\hat{\boldsymbol{\theta}} = \mathbf{A} \mathbf{z} + \mathbf{B}$$

$M \times 1 \quad M \times N \quad N \times 1 \quad M \times 1 \quad \text{通常，} N \geq M。$

这样做（线性组合）的原因是：许多时候，最小均方估值  $\hat{\boldsymbol{\theta}}_{\text{MS}}$  很难求得（因需计算条件均值  $\int_{\boldsymbol{\theta}} \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{z}) d\boldsymbol{\theta}$ ），而通常它是非线性的，那么，线性估值虽然牺牲了最优性但换来了计算的简便性；但是，如果它是线性的，那么，此时  $\hat{\boldsymbol{\theta}}_{\text{LMS}} = \hat{\boldsymbol{\theta}}_{\text{MS}}$ 。

线性最小均方误差的估值准则为：

$$\min_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{LMS}}} E(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}}) \Leftrightarrow \min_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{LMS}}} E((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})) \Leftrightarrow \min_{\{\mathbf{A}, \mathbf{B}\}} E((\boldsymbol{\theta} - \mathbf{A}\mathbf{z} - \mathbf{B})^T (\boldsymbol{\theta} - \mathbf{A}\mathbf{z} - \mathbf{B}))$$

$1 \times M \quad M \times 1 \quad (\text{标量}) \quad 1 \times M \quad M \times 1 \quad 1 \times M \quad M \times 1 \quad (\text{误差平方和})$

下面我们对系数矩阵  $\mathbf{A}$  和系数矢量  $\mathbf{B}$  进行求解。

首先， $E(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})$  取得极值的必要条件为：

$$\frac{\partial E(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})}{\partial \mathbf{A}} = \mathbf{0} \quad \text{和} \quad \frac{\partial E(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})}{\partial \mathbf{B}} = \mathbf{0}$$

$M \times N \quad (\text{标量对矩阵求导}) \quad M \times 1 \quad (\text{标量对矢量求导})$

假定积分（求统计平均）和求导可交换，则：

$$E\left(\frac{\partial(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})}{\partial \mathbf{A}}\right) = \mathbf{0} \quad \text{和} \quad E\left(\frac{\partial(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})}{\partial \mathbf{B}}\right) = \mathbf{0}$$

将标量对矢量的求导公式  $\frac{\partial(\mathbf{X}^T \cdot \mathbf{Y})}{\partial \mathbf{B}} = \frac{\partial \mathbf{X}^T}{\partial \mathbf{B}} \cdot \mathbf{Y} + \frac{\partial \mathbf{Y}^T}{\partial \mathbf{B}} \cdot \mathbf{X}$  应用于  $E\left(\frac{\partial(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})}{\partial \mathbf{B}}\right)$  得：

$$E\left(\frac{\partial(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})}{\partial \mathbf{B}}\right) = E\left(2 \cdot \frac{\partial \tilde{\boldsymbol{\theta}}^T}{\partial \mathbf{B}} \cdot \tilde{\boldsymbol{\theta}}\right)$$

又  $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \mathbf{A}\mathbf{z} - \mathbf{B}$  且  $\frac{d\mathbf{B}^T}{d\mathbf{B}} = \mathbf{I}$  ( $M \times M$ )，故我们有：

$$E\left(\frac{\partial(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})}{\partial \mathbf{B}}\right) = 2E(-\mathbf{I} \cdot (\boldsymbol{\theta} - \mathbf{A}\mathbf{z} - \mathbf{B})) = -2[E(\boldsymbol{\theta}) - \mathbf{A}E(\mathbf{z}) - \mathbf{B}]$$

而  $E\left(\frac{\partial(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})}{\partial \mathbf{B}}\right) = \mathbf{0}$ ，所以，我们解得： $\mathbf{B} = E(\boldsymbol{\theta}) - \mathbf{A}E(\mathbf{z})$

现在考虑从方程  $E\left(\frac{\partial(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})}{\partial \mathbf{A}}\right) = \mathbf{0}$  ( $M \times N$ ) 求解  $\mathbf{A}$ 。

由于  $\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}} = \text{tr}(\tilde{\boldsymbol{\theta}} \cdot \tilde{\boldsymbol{\theta}}^T)$  (矩阵的迹  $\text{tr}$  等于矩阵对角线上的元素之和)，故：

$$\begin{aligned} E\left(\frac{\partial(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})}{\partial \mathbf{A}}\right) &= E\left(\frac{\partial}{\partial \mathbf{A}} \text{tr}(\tilde{\boldsymbol{\theta}} \cdot \tilde{\boldsymbol{\theta}}^T)\right) = E\left(\frac{\partial}{\partial \mathbf{A}} \text{tr}((\boldsymbol{\theta} - \mathbf{A}\mathbf{z} - \mathbf{B})(\boldsymbol{\theta} - \mathbf{A}\mathbf{z} - \mathbf{B})^T)\right) \\ &= E\left(\frac{\partial}{\partial \mathbf{A}} \text{tr}(\underline{\boldsymbol{\theta}}\underline{\boldsymbol{\theta}}^T - \underline{\boldsymbol{\theta}}\mathbf{z}^T\mathbf{A}^T - \underline{\mathbf{B}}\mathbf{B}^T - \mathbf{A}\mathbf{z}\boldsymbol{\theta}^T + \mathbf{A}\mathbf{z}\mathbf{z}^T\mathbf{A}^T + \mathbf{A}\mathbf{z}\mathbf{B}^T - \underline{\mathbf{B}}\boldsymbol{\theta}^T + \mathbf{B}\mathbf{z}^T\mathbf{A}^T + \underline{\mathbf{B}}\mathbf{B}^T)\right) \\ &= E\left(\frac{\partial}{\partial \mathbf{A}} \text{tr}(-\underline{\boldsymbol{\theta}}\mathbf{z}^T\mathbf{A}^T - \underline{\mathbf{A}}\mathbf{z}\boldsymbol{\theta}^T + \mathbf{A}\mathbf{z}\mathbf{z}^T\mathbf{A}^T + \underline{\mathbf{A}}\mathbf{z}\mathbf{B}^T + \mathbf{B}\mathbf{z}^T\mathbf{A}^T)\right) \\ &= E\left(\frac{\partial}{\partial \mathbf{A}} \text{tr}(-2\underline{\boldsymbol{\theta}}\mathbf{z}^T\mathbf{A}^T + \mathbf{A}\mathbf{z}\mathbf{z}^T\mathbf{A}^T + 2\mathbf{B}\mathbf{z}^T\mathbf{A}^T)\right) \\ &= E(-2\underline{\boldsymbol{\theta}}\mathbf{z}^T + 2\mathbf{A}\mathbf{z}\mathbf{z}^T + 2\mathbf{B}\mathbf{z}^T) \\ &= 2[\mathbf{A}E(\mathbf{z}\mathbf{z}^T) + \mathbf{B}E(\mathbf{z}^T) - E(\boldsymbol{\theta}\mathbf{z}^T)] \end{aligned}$$

由  $E\left(\frac{\partial(\tilde{\boldsymbol{\theta}}^T \cdot \tilde{\boldsymbol{\theta}})}{\partial \mathbf{A}}\right) = \mathbf{0}$  并将  $\mathbf{B} = E(\boldsymbol{\theta}) - \mathbf{A}E(\mathbf{z})$  代入上式可得：

$$\mathbf{A}[E(\mathbf{z}\mathbf{z}^T) - E(\mathbf{z})E(\mathbf{z}^T)] = E(\boldsymbol{\theta}\mathbf{z}^T) - E(\boldsymbol{\theta})E(\mathbf{z}^T)$$

或  $\mathbf{A}E([\mathbf{z} - E(\mathbf{z})][\mathbf{z} - E(\mathbf{z})]^T) = E([\boldsymbol{\theta} - E(\boldsymbol{\theta})][\mathbf{z} - E(\mathbf{z})]^T)$

最后，我们得到系数矩阵  $\mathbf{A}$  和  $\mathbf{B}$  解：

$$\mathbf{A} = \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \quad (M \times N = M \times N \times N \times N)$$

$$\mathbf{B} = E(\boldsymbol{\theta}) - \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot E(\mathbf{z}) \quad (M \times 1 = M \times 1 - M \times N \times N \times N \times N \times 1)$$

从而得到线性最小均方估值为： $\hat{\boldsymbol{\theta}}_{\text{LMS}} = E(\boldsymbol{\theta}) + \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot [\mathbf{z} - E(\mathbf{z})]$

下面我们来讨论估计量的性质：

- ① 估值的计算只需信号参量  $\boldsymbol{\theta}$  和观测值  $\mathbf{z}$  的一、二阶矩。
- ② 无偏估值，因  $E(\hat{\boldsymbol{\theta}}_{\text{LMS}}) = E(\boldsymbol{\theta}) + \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot [E(\mathbf{z}) - E(\mathbf{z})] = E(\boldsymbol{\theta})$ 。

③ 估值的误差矩阵为：

$$\begin{aligned}
 \text{Var}(\tilde{\boldsymbol{\theta}}) &= E(\tilde{\boldsymbol{\theta}} \cdot \tilde{\boldsymbol{\theta}}^T) \\
 &= E((\boldsymbol{\theta} - E(\boldsymbol{\theta}) - \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot [\mathbf{z} - E(\mathbf{z})]) (\boldsymbol{\theta} - E(\boldsymbol{\theta}) - \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot [\mathbf{z} - E(\mathbf{z})])^T) \\
 &= \text{Var}(\boldsymbol{\theta}) - \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot \text{Cov}(\mathbf{z}, \boldsymbol{\theta}) - \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot ([\text{Var}(\mathbf{z})]^{-1})^T \cdot [\text{Cov}(\boldsymbol{\theta}, \mathbf{z})]^T + \\
 &\quad + \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot \text{Var}(\mathbf{z}) \cdot ([\text{Var}(\mathbf{z})]^{-1})^T \cdot [\text{Cov}(\boldsymbol{\theta}, \mathbf{z})]^T \\
 &= \text{Var}(\boldsymbol{\theta}) - \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot \text{Cov}(\mathbf{z}, \boldsymbol{\theta})
 \end{aligned}$$

可见：根据已知的信号参量 $\boldsymbol{\theta}$ 和观测值 $\mathbf{z}$ 的一、二阶矩，我们可计算出信号参量的线性最小均方估值及其误差矩阵。

另外，我们可以导出线性最小均方估值误差 $\tilde{\boldsymbol{\theta}}$ 与观测值 $\mathbf{z}$ 无关的结论，因为：

$$\begin{aligned}
 E(\tilde{\boldsymbol{\theta}} \mathbf{z}^T) &= E((\boldsymbol{\theta} - E(\boldsymbol{\theta}) - \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot [\mathbf{z} - E(\mathbf{z})]) \mathbf{z}^T) \\
 &= E((\boldsymbol{\theta} - E(\boldsymbol{\theta}) - \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot [\mathbf{z} - E(\mathbf{z})]) [\mathbf{z} - E(\mathbf{z})]^T) \\
 &= \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) - \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot \text{Var}(\mathbf{z}) \\
 &= \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) - \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \\
 &= \mathbf{0}
 \end{aligned}$$

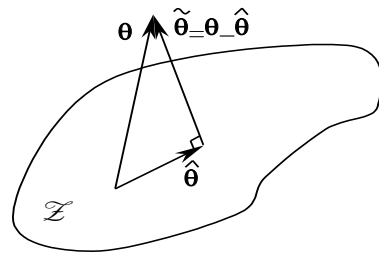
这就是我们下一节要讲述的正交原理。

## 2.8 正交原理

在上一节中，我们通过求极值的方法得到了线性最小均方估值。实际上，我们亦可以根据著名的正交原理来获得该线性最小均方估值。

几何解释如右图所示。

设观测空间 $\mathcal{Z}$ 是随机向量 $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$ 的有限集构成的线性空间，对 $\forall k$ 有 $E(\mathbf{z}_k \mathbf{z}_k^T) < \infty$ 。则 $\mathcal{Z}$ 是具有有限二阶矩的随机向量空间。定义：此空间中两个向量的内积为 $\langle \mathbf{x}, \mathbf{y} \rangle = E(\mathbf{x}^T \mathbf{y})$ ，相应的模（范数）为 $\|\mathbf{x}\| = [E(\mathbf{x}^T \mathbf{x})]^{1/2}$ 。若 $E(\mathbf{x}^T \mathbf{y}) = 0$ ，则称两个矢量 $\mathbf{x}$ 和 $\mathbf{y}$ 是正交的。



我们希望在观测空间 $\mathcal{Z}$ 中来估计信号的参量矢量 $\boldsymbol{\theta}$ ，令信号参量的估值 $\hat{\boldsymbol{\theta}}$ 为 $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$ 的线性组合。其估值误差矢量 $\tilde{\boldsymbol{\theta}}$ 为信号参量矢量 $\boldsymbol{\theta}$ 与估值矢量 $\hat{\boldsymbol{\theta}}$ 的差值，即 $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ 。根据线性空间的投影定理，使估值误差的模 $\|\tilde{\boldsymbol{\theta}}\|$ 最小（即 $E(\tilde{\boldsymbol{\theta}}^T \tilde{\boldsymbol{\theta}})$ 最小）的充要条件是 $\hat{\boldsymbol{\theta}}$ 为 $\boldsymbol{\theta}$ 在观测空间 $\mathcal{Z}$ 中的投影。故，估值误差矢量 $\tilde{\boldsymbol{\theta}}$ 与观测矢量 $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$ 正交，即：

$$\text{取得线性最小均方估值的充要条件是 } E(\tilde{\boldsymbol{\theta}} \cdot \mathbf{z}^T) = \mathbf{0}.$$

$M \times 1 \quad 1 \times N \quad M \times N$

这就是“正交原理”。

下面我们给予证明：

### ① 必要性

即要证明：若 $\hat{\boldsymbol{\theta}}_{\text{LMS}} = \mathbf{A}\mathbf{z} + \mathbf{B}$ ，其中 $\mathbf{A} = \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1}$ 、 $\mathbf{B} = E(\boldsymbol{\theta}) - \mathbf{A} \cdot E(\mathbf{z})$ ，则 $E(\tilde{\boldsymbol{\theta}} \cdot \mathbf{z}^T) = \mathbf{0}$ 。而这已经在上一节的最后加以推导证明了。

### ② 充分性

即要证明：若有线性无偏估值 $\hat{\boldsymbol{\theta}} = \mathbf{C}\mathbf{z} + \mathbf{D}$ 满足 $E(\tilde{\boldsymbol{\theta}} \cdot \mathbf{z}^T) = \mathbf{0}$ ，则一定有 $\mathbf{C} = \mathbf{A}$ 且 $\mathbf{D} = \mathbf{B}$ （即 $\hat{\boldsymbol{\theta}}$ 为线性最小均方估值）。

由估值无偏性： $E(\hat{\theta})=CE(\mathbf{z})+D=E(\theta)$ ，故  $D=E(\theta)-CE(\mathbf{z})$ 。

所以， $\tilde{\theta}=\theta-C\mathbf{z}-D=\theta-E(\theta)-C[\mathbf{z}-E(\mathbf{z})]$ 。

又由正交性  $E(\tilde{\theta} \cdot \mathbf{z}^T)=0$  和无偏性  $E(\tilde{\theta})=0$  有： $E(\tilde{\theta} \cdot \mathbf{z}^T)-E(\tilde{\theta})[E(\mathbf{z})]^T=0$ 。

而，等式的左端  $=E(\tilde{\theta} \cdot [\mathbf{z}-E(\mathbf{z})]^T)=E((\theta-E(\theta)-C[\mathbf{z}-E(\mathbf{z})]) \cdot [\mathbf{z}-E(\mathbf{z})]^T)$   
 $=\text{Cov}(\theta, \mathbf{z})-C\text{Var}(\mathbf{z})$

所以， $C=\text{Cov}(\theta, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1}$ 。

由此，我们得到： $C=A$  且  $D=B$ ，即充分性得证。

## 2.9 最小二乘估值 (LS) 和加权最小二乘估值 (LSW)

在前面几节中，Bayes 估值、最大后验估值和最大似然估值均要求知道概率密度函数，线性最小均方估值要求知道统计一、二阶矩。

现在，我们放宽对统计知识的要求，无须任何先验知识而对待估参量进行估值。当然，可以预计估值的质量不如以前，但在不少的实际应用中还是令人满意的。

### 2.9.1 最小二乘估值 (LS)

我们首先来看一个实例：

我们希望根据实际测得的目标距离  $d$  (米) 来估计目标的初始距离  $d_0$  (米) 和它的运动速度  $v$  (米/秒)。假设测量距离仪器的噪声 (如尺度不准等) 为  $n$  (米)，那么我们可以建立如下的观测模型：

$$d=d_0+vt+n$$

$$\text{作 } N \text{ 次观测，则有：} \begin{pmatrix} d_1 \\ \vdots \\ d_N \end{pmatrix} = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_N \end{pmatrix} \begin{pmatrix} d_0 \\ v \end{pmatrix} + \begin{pmatrix} n_1 \\ \vdots \\ n_N \end{pmatrix}$$

由此可看出，一般地，我们可假定观测模型为：

$$\begin{array}{ccccccc} \mathbf{z} & = & \mathbf{C} & \mathbf{\theta} & + & \mathbf{n} \\ \text{Nx1} & & \text{NxM} & \text{Mx1} & & \text{Nx1} \\ \text{观测矢量} & & \text{系数矩阵} & \text{待估参量} & & \text{观测噪声 (观测时产生的随机误差或外加干扰噪声)} \end{array}$$

令估值为  $\hat{\theta}$ ，则我们可构造目标 (性能) 函数为  $(\mathbf{z}-\mathbf{C}\hat{\theta})^T \cdot (\mathbf{z}-\mathbf{C}\hat{\theta})$ ，它是观测估值误差的平方和。那么，最小二乘估值准则可叙述为：

$$\min_{\hat{\theta}=\hat{\theta}_{LS}} (\mathbf{z}-\mathbf{C}\hat{\theta})^T \cdot (\mathbf{z}-\mathbf{C}\hat{\theta})$$

1xN          Nx1

$$\text{令 } \frac{\partial}{\partial \hat{\theta}} (\mathbf{z}-\mathbf{C}\hat{\theta})^T \cdot (\mathbf{z}-\mathbf{C}\hat{\theta})=0, \text{ 可得: } 2 \cdot \frac{\partial (\mathbf{z}-\mathbf{C}\hat{\theta})^T}{\partial \hat{\theta}} \cdot (\mathbf{z}-\mathbf{C}\hat{\theta})=0 \text{ 或 } \mathbf{C}^T (\mathbf{z}-\mathbf{C}\hat{\theta})=0$$

故，我们得到最小二乘估值为： $\hat{\theta}_{LS}=(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{z}$

下面我们来讨论最小二乘估值的性质：

- ①  $\hat{\theta}_{LS}$  是观测值  $\mathbf{z}$  的线性函数，但它与线性最小均方估值  $\hat{\theta}_{LMS}$  不同。
- ②  $\hat{\theta}_{LS}$  是无偏估值，若  $E(\mathbf{n})=0$ 。
- ③ 估值方差矩阵  $E(\tilde{\theta} \tilde{\theta}^T)$ ：

$$\begin{aligned} \tilde{\theta} &= \theta - \hat{\theta}_{LS} = \theta - (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{z} = \theta - (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T(\mathbf{C}\theta+\mathbf{n}) = -(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{n} \\ E(\tilde{\theta} \tilde{\theta}^T) &= (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T E(\mathbf{n}\mathbf{n}^T) \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1} = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T \mathbf{R}_n \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1} \end{aligned}$$

若  $E(\mathbf{n})=\mathbf{0}$ ，则噪声相关阵  $\mathbf{R}_n = \text{Var}(\mathbf{n})$ 。

此时， $E(\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^T) = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \text{Var}(\mathbf{n}) \mathbf{C} (\mathbf{C}^T \mathbf{C})^{-1}$ 。

## 2.9.2 加权最小二乘估值 (LSW)

在上一节中，我们构造目标函数  $(\mathbf{z}-\mathbf{C}\hat{\boldsymbol{\theta}})^T \cdot (\mathbf{z}-\mathbf{C}\hat{\boldsymbol{\theta}})$  时，实际上是将各次观测噪声同等对待，即各次观测噪声同等重要。

现在，若各次观测噪声  $\mathbf{n}_k$  ( $k=1,2,\dots,N$ ) 具有不同的功率，则“同等地”对待它们不尽合理。如果我们对大的观测误差加权小而对小的观测误差加权重，那么我们可以预期这样所得到的估值会更精确。

所以，对观测模型  $\mathbf{z}=\mathbf{C}\boldsymbol{\theta}+\mathbf{n}$  引入新的目标函数  $(\mathbf{z}-\mathbf{C}\hat{\boldsymbol{\theta}})^T \cdot \mathbf{W} \cdot (\mathbf{z}-\mathbf{C}\hat{\boldsymbol{\theta}})$ ，其中  $\mathbf{W}$  是对称正定的加权矩阵。那么，加权最小二乘估值准则可描述为：

$$\min_{\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}_{\text{LSW}}} (\mathbf{z}-\mathbf{C}\hat{\boldsymbol{\theta}})^T \cdot \mathbf{W} \cdot (\mathbf{z}-\mathbf{C}\hat{\boldsymbol{\theta}})$$

$\begin{matrix} 1 \times N & N \times N & N \times 1 \end{matrix}$

从而得到加权最小二乘估值为： $\hat{\boldsymbol{\theta}}_{\text{LSW}} = (\mathbf{C}^T \mathbf{W} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{W} \mathbf{z}$

下面我们来讨论最小二乘估值的性质：

- ①  $\hat{\boldsymbol{\theta}}_{\text{LSW}}$  是  $\mathbf{z}$  的线性函数，但与最小二乘估值  $\hat{\boldsymbol{\theta}}_{\text{LS}}$  和线性最小均方估值  $\hat{\boldsymbol{\theta}}_{\text{LMS}}$  不同。
- ②  $\hat{\boldsymbol{\theta}}_{\text{LSW}}$  是无偏估值，若  $E(\mathbf{n})=\mathbf{0}$ 。
- ③ 估值方差矩阵  $E(\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^T)$ ：

$$E(\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^T) = (\mathbf{C}^T \mathbf{W} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{W} \mathbf{R}_n \mathbf{W} \mathbf{C} (\mathbf{C}^T \mathbf{W} \mathbf{C})^{-1}$$

求估值方差矩阵的下限：

由于  $\mathbf{R}_n$  为正定的，故可令  $\mathbf{R}_n = \mathbf{H}^T \mathbf{H}$ 。

根据矩阵 Schwarz 不等式： $\mathbf{B}^T \mathbf{B} \geq (\mathbf{A} \mathbf{B})^T (\mathbf{A} \mathbf{A}^T)^{-1} (\mathbf{A} \mathbf{B})$

$\begin{matrix} L \times L & L \times N & N \times N & N \times L \end{matrix}$

当且仅当  $\mathbf{B} = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{B}$  时，上式中等号“=”成立。

令  $\mathbf{A} = \mathbf{C}^T \mathbf{H}^{-1}$ ， $\mathbf{B} = \mathbf{H} \mathbf{W} \mathbf{C} (\mathbf{C}^T \mathbf{W} \mathbf{C})^{-1}$ ，则有  $\mathbf{A} \mathbf{B} = \mathbf{I}$ 。从而可得：

$$\begin{aligned} \mathbf{B}^T \mathbf{B} &= (\mathbf{C}^T \mathbf{W} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{W}^T \mathbf{H}^T \mathbf{H} \mathbf{W} \mathbf{C} (\mathbf{C}^T \mathbf{W} \mathbf{C})^{-1} = (\mathbf{C}^T \mathbf{W} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{W} \mathbf{R}_n \mathbf{W} \mathbf{C} (\mathbf{C}^T \mathbf{W} \mathbf{C})^{-1} \\ &\geq (\mathbf{A} \mathbf{B})^T (\mathbf{A} \mathbf{A}^T)^{-1} (\mathbf{A} \mathbf{B}) = (\mathbf{A} \mathbf{A}^T)^{-1} = (\mathbf{C}^T \mathbf{H}^{-1} (\mathbf{H}^{-1})^T \mathbf{C})^{-1} = (\mathbf{C}^T (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{C})^{-1} = (\mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{C})^{-1} \end{aligned}$$

所以，我们最终得到： $E(\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^T) \geq (\mathbf{C}^T \mathbf{R}_n^{-1} \mathbf{C})^{-1}$

当且仅当  $\mathbf{W} = \mathbf{R}_n^{-1}$  时，上式中等号“=”成立（可代入前式验证）。此时的  $\mathbf{W}$  称为最佳加权矩阵，相应的估值方差矩阵为最小。

### ◆例 1：

设观测模型： $\mathbf{z} = \mathbf{C} \boldsymbol{\theta} + \mathbf{n}$ ，（ $\mathbf{C}$  称为调制矩阵）

$\begin{matrix} N \times 1 & N \times M & M \times 1 & N \times 1 & N \geq M \end{matrix}$

其中： $\boldsymbol{\theta}$  与  $\mathbf{n}$  相互独立，均服从高斯分布，且： $E(\boldsymbol{\theta})=\boldsymbol{\mu}_\theta$ ， $E(\mathbf{n})=\mathbf{0}$ ， $\text{Var}(\boldsymbol{\theta})=\mathbf{V}_\theta$ ， $\text{Var}(\mathbf{n})=\mathbf{V}_n=\mathbf{R}_n$ 。

试求对  $\boldsymbol{\theta}$  的：①最小均方误差估值  $\hat{\boldsymbol{\theta}}_{\text{ms}}$ 、②最大后验估值  $\hat{\boldsymbol{\theta}}_{\text{map}}$ 、③最大似然估值  $\hat{\boldsymbol{\theta}}_{\text{ML}}$ 、④线性最小均方估值  $\hat{\boldsymbol{\theta}}_{\text{LMS}}$ 。

解：

- ① 最小均方误差估值：

首先需要求出后验密度  $p(\boldsymbol{\theta}|\mathbf{z})$  的表达式。

因为  $p(\boldsymbol{\theta}|\mathbf{z}) = p(\mathbf{z}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) / p(\mathbf{z})$ ，而由已知条件，我们可得：

$$p(\mathbf{z}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2} |\mathbf{V}_n|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{z}-\mathbf{C}\boldsymbol{\theta})^T \mathbf{V}_n^{-1} (\mathbf{z}-\mathbf{C}\boldsymbol{\theta})\right) \quad \text{其中} |\mathbf{V}_n| \text{为 } \mathbf{V}_n \text{的行列式。}$$

$$p(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{M/2} |\mathbf{V}_\theta|^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{\theta}-\boldsymbol{\mu}_\theta)^T \mathbf{V}_\theta^{-1} (\boldsymbol{\theta}-\boldsymbol{\mu}_\theta)\right)$$

$$p(\mathbf{z}) = \frac{1}{(2\pi)^{N/2} |\mathbf{V}_z|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{z}-\boldsymbol{\mu}_z)^T \mathbf{V}_z^{-1} (\mathbf{z}-\boldsymbol{\mu}_z)\right) \quad (\mathbf{z} \text{ 服从 } \boldsymbol{\theta} \text{ 和 } \mathbf{n} \text{ 联合高斯分布})$$

$$\text{其中: } \boldsymbol{\mu}_z = E(\mathbf{z}) = \mathbf{C} \cdot \boldsymbol{\mu}_\theta$$

$$\begin{aligned} \mathbf{V}_z &= E((\mathbf{z}-\boldsymbol{\mu}_z)(\mathbf{z}-\boldsymbol{\mu}_z)^T) = E((\mathbf{z}-\mathbf{C} \cdot \boldsymbol{\mu}_\theta)(\mathbf{z}-\mathbf{C} \cdot \boldsymbol{\mu}_\theta)^T) \\ &= E((\mathbf{C}\boldsymbol{\theta} + \mathbf{n} - \mathbf{C} \cdot \boldsymbol{\mu}_\theta)(\mathbf{C}\boldsymbol{\theta} + \mathbf{n} - \mathbf{C} \cdot \boldsymbol{\mu}_\theta)^T) = E([\mathbf{C}(\boldsymbol{\theta}-\boldsymbol{\mu}_\theta) + \mathbf{n}][\mathbf{C}(\boldsymbol{\theta}-\boldsymbol{\mu}_\theta) + \mathbf{n}]^T) \\ &= E(\mathbf{C}(\boldsymbol{\theta}-\boldsymbol{\mu}_\theta)(\boldsymbol{\theta}-\boldsymbol{\mu}_\theta)^T \mathbf{C}^T) + E(\mathbf{n}\mathbf{n}^T) \\ &= \mathbf{C}\mathbf{V}_\theta \mathbf{C}^T + \mathbf{V}_n \end{aligned}$$

故，我们有：

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{z}) &= \frac{|\mathbf{V}_z|^{1/2}}{(2\pi)^{M/2} |\mathbf{V}_n|^{1/2} |\mathbf{V}_\theta|^{1/2}} \exp\left(-\frac{1}{2} [(\mathbf{z}-\mathbf{C}\boldsymbol{\theta})^T \mathbf{V}_n^{-1} (\mathbf{z}-\mathbf{C}\boldsymbol{\theta}) + (\boldsymbol{\theta}-\boldsymbol{\mu}_\theta)^T \mathbf{V}_\theta^{-1} (\boldsymbol{\theta}-\boldsymbol{\mu}_\theta) - (\mathbf{z}-\boldsymbol{\mu}_z)^T \mathbf{V}_z^{-1} (\mathbf{z}-\boldsymbol{\mu}_z)]\right) \\ &= \frac{1}{(2\pi)^{M/2} |\mathbf{V}_{\theta|z}|^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{\theta}-\boldsymbol{\mu}_{\theta|z})^T \mathbf{V}_{\theta|z}^{-1} (\boldsymbol{\theta}-\boldsymbol{\mu}_{\theta|z})\right) \end{aligned}$$

其中的均值 $\boldsymbol{\mu}_{\theta|z}$ 和方差 $\mathbf{V}_{\theta|z}$ 可由下述方法求得：

$$\text{首先, } \mathbf{V}_{\theta|z}^{-1} = \mathbf{V}_\theta^{-1} + \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C}$$

$$\text{而由矩阵引理: } (\mathbf{A} + \mathbf{B}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{I} + \mathbf{C}\mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{C}\mathbf{A}^{-1}$$

$$\begin{aligned} \text{故, } \mathbf{V}_{\theta|z} &= (\mathbf{V}_\theta^{-1} + \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1} = \mathbf{V}_\theta - \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_n^{-1} (\mathbf{I} + \mathbf{C}\mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_n^{-1})^{-1} \mathbf{C}\mathbf{V}_\theta \\ &= \mathbf{V}_\theta - \mathbf{V}_\theta \mathbf{C}^T (\mathbf{V}_n + \mathbf{C}\mathbf{V}_\theta \mathbf{C}^T)^{-1} \mathbf{C}\mathbf{V}_\theta \\ &= \mathbf{V}_\theta - \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_z^{-1} \mathbf{C}\mathbf{V}_\theta \end{aligned}$$

$$\text{其次, } \mathbf{V}_{\theta|z}^{-1} \boldsymbol{\mu}_{\theta|z} = \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{z} + \mathbf{V}_\theta^{-1} \boldsymbol{\mu}_\theta$$

$$\begin{aligned} \text{故, } \boldsymbol{\mu}_{\theta|z} &= \mathbf{V}_{\theta|z} (\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{z} + \mathbf{V}_\theta^{-1} \boldsymbol{\mu}_\theta) = (\mathbf{V}_\theta - \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_z^{-1} \mathbf{C}\mathbf{V}_\theta) (\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{z} + \mathbf{V}_\theta^{-1} \boldsymbol{\mu}_\theta) \\ &= \boldsymbol{\mu}_\theta + \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{z} - \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_z^{-1} \mathbf{C}\mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{z} - \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_z^{-1} \mathbf{C}\boldsymbol{\mu}_\theta \\ &= \boldsymbol{\mu}_\theta + \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{z} - \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_z^{-1} (\mathbf{V}_z - \mathbf{V}_n) \mathbf{V}_n^{-1} \mathbf{z} - \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_z^{-1} \mathbf{C}\boldsymbol{\mu}_\theta \\ &= \boldsymbol{\mu}_\theta + \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_z^{-1} \mathbf{z} - \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_z^{-1} \mathbf{C}\boldsymbol{\mu}_\theta \\ &= \boldsymbol{\mu}_\theta + \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_z^{-1} (\mathbf{z} - \boldsymbol{\mu}_z) \end{aligned}$$

所以， $\hat{\boldsymbol{\theta}}_{\text{ms}} = p(\boldsymbol{\theta}|\mathbf{z})$ 的均值 =  $\boldsymbol{\mu}_{\theta|z} = \boldsymbol{\mu}_\theta + \mathbf{V}_\theta \mathbf{C}^T \mathbf{V}_z^{-1} (\mathbf{z} - \boldsymbol{\mu}_z)$ 。可见：它与 $\boldsymbol{\theta}$ 的先验知识有关。

## ② 最大后验估值：

由于后验概率密度 $p(\boldsymbol{\theta}|\mathbf{z})$ 是高斯分布，所以， $\hat{\boldsymbol{\theta}}_{\text{map}} = \hat{\boldsymbol{\theta}}_{\text{ms}} = \boldsymbol{\mu}_{\theta|z}$ 。

也可由后验方程 $\nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{\theta}|\mathbf{z}) \big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{map}}} = \mathbf{0}$ 得： $-\frac{1}{2} [\mathbf{V}_{\theta|z}^{-1} + (\mathbf{V}_{\theta|z}^{-1})^T] (\hat{\boldsymbol{\theta}}_{\text{map}} - \boldsymbol{\mu}_{\theta|z}) = \mathbf{0}$

从而可得： $\hat{\boldsymbol{\theta}}_{\text{map}} = \hat{\boldsymbol{\theta}}_{\text{ms}} = \boldsymbol{\mu}_{\theta|z}$ 。

## ③ 最大似然估值：

由似然方程 $\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{z}|\boldsymbol{\theta}) \big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{ML}}} = \mathbf{0}$ 得： $\frac{1}{2} \mathbf{C}^T [\mathbf{V}_n^{-1} + (\mathbf{V}_n^{-1})^T] (\mathbf{z} - \mathbf{C} \hat{\boldsymbol{\theta}}_{\text{ML}}) = \mathbf{0}$

故可得： $\hat{\boldsymbol{\theta}}_{ML} = (\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{z}$ 。可见：它与 $\boldsymbol{\theta}$ 的先验知识无关。

$$\begin{matrix} M \times 1 & M \times M & M \times N & N \times N & N \times 1 \end{matrix}$$

其估值方差矩阵为：

$$E(\tilde{\boldsymbol{\theta}}_{ML} \tilde{\boldsymbol{\theta}}_{ML}^T) = E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{ML})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{ML})^T]$$

$$\text{而 } \hat{\boldsymbol{\theta}}_{ML} = (\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{z} = (\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{V}_n^{-1} (\mathbf{C}\boldsymbol{\theta} + \mathbf{n}) = \boldsymbol{\theta} + (\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{n}$$

$$\begin{aligned} \text{故：} E(\tilde{\boldsymbol{\theta}}_{ML} \tilde{\boldsymbol{\theta}}_{ML}^T) &= E[(\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{n}][(\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{n}]^T \\ &= (\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{V}_n \mathbf{V}_n^{-1} \mathbf{C} (\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1} \\ &= (\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1} \end{aligned}$$

④ 线性最小均方估值：

$$\hat{\boldsymbol{\theta}}_{LMS} = E(\boldsymbol{\theta}) + \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot [\mathbf{z} - E(\mathbf{z})]$$

$$\text{Var}(\tilde{\boldsymbol{\theta}}_{LMS}) = \text{Var}(\boldsymbol{\theta}) - \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) \cdot [\text{Var}(\mathbf{z})]^{-1} \cdot \text{Cov}(\mathbf{z}, \boldsymbol{\theta})$$

$$\text{在本例中：} E(\boldsymbol{\theta}) = \boldsymbol{\mu}_\theta, E(\mathbf{z}) = \boldsymbol{\mu}_z = \mathbf{C}\boldsymbol{\mu}_\theta, E(\mathbf{n}) = \mathbf{0}, \text{Var}(\mathbf{z}) = \mathbf{V}_z, \text{Var}(\boldsymbol{\theta}) = E(\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^T) = \mathbf{V}_\theta,$$

$$\begin{aligned} \text{Cov}(\boldsymbol{\theta}, \mathbf{z}) &= E[(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)(\mathbf{z} - \boldsymbol{\mu}_z)^T] = E[(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)[\mathbf{C}\boldsymbol{\theta} + \mathbf{n} - \mathbf{C}\boldsymbol{\mu}_\theta]^T] = E(\tilde{\boldsymbol{\theta}} [\mathbf{C}\tilde{\boldsymbol{\theta}} + \mathbf{n}]^T) \\ &= E(\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^T) \mathbf{C}^T + E(\tilde{\boldsymbol{\theta}} \mathbf{n}^T) = \mathbf{V}_\theta \mathbf{C}^T \end{aligned}$$

$$\text{Cov}(\mathbf{z}, \boldsymbol{\theta}) = \mathbf{C} \mathbf{V}_\theta$$

$$\text{所以，我们可得：} \hat{\boldsymbol{\theta}}_{LMS} = \boldsymbol{\mu}_\theta + \mathbf{V}_\theta \mathbf{C}^T \cdot \mathbf{V}_z^{-1} \cdot [\mathbf{z} - \boldsymbol{\mu}_z] = \hat{\boldsymbol{\theta}}_{ms}$$

$$\text{Var}(\tilde{\boldsymbol{\theta}}_{LMS}) = \mathbf{V}_\theta - \mathbf{V}_\theta \mathbf{C}^T \cdot \mathbf{V}_z^{-1} \cdot \mathbf{C} \mathbf{V}_\theta = \mathbf{V}_{\theta|z}$$

#解毕

讨论：

a) 最小均方估值 $\hat{\boldsymbol{\theta}}_{ms} (= \hat{\boldsymbol{\theta}}_{map} = \hat{\boldsymbol{\theta}}_{LMS})$ 与最大似然估值 $\hat{\boldsymbol{\theta}}_{ML}$ 的方差相比较：

$$\begin{aligned} \text{因 } \mathbf{V}_{\theta|z}^{-1} &= \mathbf{V}_\theta^{-1} + \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C}, \text{ 则 } \mathbf{V}_{\theta|z} = (\mathbf{V}_\theta^{-1} + \mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1}, \text{ 而 } \text{Var}(\tilde{\boldsymbol{\theta}}_{ML}) = (\mathbf{C}^T \mathbf{V}_n^{-1} \mathbf{C})^{-1}, \\ \text{故, } \mathbf{V}_{\theta|z} &\leq \text{Var}(\tilde{\boldsymbol{\theta}}_{ML}). \end{aligned}$$

当无 $\boldsymbol{\theta}$ 的先验知识时， $\boldsymbol{\theta}$ 的方差 $\mathbf{V}_\theta$ 为无穷大，其逆 $\mathbf{V}_\theta^{-1}$ 为零，上式中等号近似成立。

b) 最大似然估值 $\hat{\boldsymbol{\theta}}_{ML}$ 与加权最小二乘估值 $\hat{\boldsymbol{\theta}}_{LSW}$ 相比较：

当 $\mathbf{W}$ 为最佳加权矩阵，即 $\mathbf{W} = \mathbf{R}_n^{-1} = \mathbf{V}_n^{-1}$ 时，两者的方差完全一样， $\tilde{\boldsymbol{\theta}}_{LMS}$ 亦为最佳估值。

$$\text{实际上，似然函数 } p(\mathbf{z}|\boldsymbol{\theta}) \text{ 的指数为 } [-\frac{1}{2} (\mathbf{z} - \mathbf{C}\boldsymbol{\theta})^T \mathbf{V}_n^{-1} (\mathbf{z} - \mathbf{C}\boldsymbol{\theta})],$$

$$\text{而此时的加权最小二乘估值的目标函数为 } (\mathbf{z} - \mathbf{C}\boldsymbol{\theta})^T \mathbf{V}_n^{-1} (\mathbf{z} - \mathbf{C}\boldsymbol{\theta}),$$

所以，使目标函数 $(\mathbf{z} - \mathbf{C}\boldsymbol{\theta})^T \mathbf{V}_n^{-1} (\mathbf{z} - \mathbf{C}\boldsymbol{\theta})$ 最小 $\Leftrightarrow$ 使似然函数 $p(\mathbf{z}|\boldsymbol{\theta})$ 最大。

c) 最大似然估值 $\hat{\boldsymbol{\theta}}_{ML}$ 与最小二乘估值 $\hat{\boldsymbol{\theta}}_{LS}$ 相比较：

当 $\mathbf{n}$ 为白噪声时， $\mathbf{V}_n = \mathbf{R}_n = \sigma^2 \mathbf{I}$ 。

$$\text{此时，似然函数 } p(\mathbf{z}|\boldsymbol{\theta}) \text{ 的指数为 } [-\frac{1}{2\sigma^2} (\mathbf{z} - \mathbf{C}\boldsymbol{\theta})^T (\mathbf{z} - \mathbf{C}\boldsymbol{\theta})],$$

$$\text{而最小二乘估值的目标函数为 } (\mathbf{z} - \mathbf{C}\boldsymbol{\theta})^T (\mathbf{z} - \mathbf{C}\boldsymbol{\theta}),$$

故在这种情况下，使目标函数 $(\mathbf{z} - \mathbf{C}\boldsymbol{\theta})^T (\mathbf{z} - \mathbf{C}\boldsymbol{\theta})$ 最小 $\Leftrightarrow$ 使似然函数 $p(\mathbf{z}|\boldsymbol{\theta})$ 最大。

◆例 2：

对二维矢量 $\theta=(\theta_1, \theta_2)^T$ 进行两次观测:

$$\mathbf{z}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \theta + \mathbf{n}_1$$

$$\mathbf{z}_2 = 4 = [1, 2] \theta + \mathbf{n}_2$$

求 $\hat{\theta}_{LS}$ 。

解: 将二次观测统一为一个方程, 即 $\mathbf{z}=\mathbf{C}\theta+\mathbf{n}$ , 其中:  $\mathbf{z}=\begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix}$ ,  $\mathbf{C}=\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{pmatrix}$ ,  $\mathbf{n}=\begin{pmatrix} \mathbf{n}_1 \\ n_2 \end{pmatrix}$ 。

因为 $\hat{\theta}_{LS}=(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{z}$

而 $\mathbf{C}^T\mathbf{C}=\begin{pmatrix} 2 & 3 \\ 3 & 6 \end{pmatrix}$ , 则 $(\mathbf{C}^T\mathbf{C})^{-1}=\begin{pmatrix} 2 & -1 \\ -1 & 2/3 \end{pmatrix}$ ,

所以,  $\hat{\theta}_{LS}=\begin{pmatrix} 2 & -1 \\ -1 & 2/3 \end{pmatrix}\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix}=\begin{pmatrix} 1 \\ 4/3 \end{pmatrix}$ 。

#解毕

### ◆例 3:

用两个电表同时测量电网电压, 一个为 220V, 另一个为 210V。

已知: 表 1 的平均误差为 0, 标准偏差为 8; 表 2 的平均误差为 0, 标准偏差为 4。

试建立观测方程, 并求电网电压的最小二乘估值和加权最小二乘估值以及它们的估值方差, 比较并讨论之。

解: 据题意, 可有观测方程:  $\mathbf{z}=\mathbf{C}\theta+\mathbf{n}$ ,

其中:  $\mathbf{z}=\begin{pmatrix} 220 \\ 210 \end{pmatrix}$ ,  $\mathbf{C}=\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ,  $\theta=v$  (待估电网电压),  $\mathbf{n}=\begin{pmatrix} n_1 \\ n_2 \end{pmatrix}$ 。

且:  $E(n_1)=E(n_2)=0$ ,  $\mathbf{R}_n=\mathbf{V}_n=\begin{pmatrix} \sigma_{n_1}^2 & 0 \\ 0 & \sigma_{n_2}^2 \end{pmatrix}=\begin{pmatrix} 8^2 & 0 \\ 0 & 4^2 \end{pmatrix}=\begin{pmatrix} 64 & 0 \\ 0 & 16 \end{pmatrix}$ 。

须求 $\hat{v}_{LS}$ 、 $\sigma_{LS}^2$ 和 $\hat{v}_{LSW}$ 、 $\sigma_{LSW}^2$ 。

① 求最小二乘估值 $\hat{v}_{LS}$ 。

由最小二乘估值公式:  $\hat{v}_{LS}=(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{z}$ ,  $\sigma_{LS}^2=(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{R}_n\mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}$ 。

代入已知参数:  $\mathbf{C}^T\mathbf{C}=2$ , 则 $(\mathbf{C}^T\mathbf{C})^{-1}=1/2$ 。

所以,  $\hat{v}_{LS}=\frac{1}{2}(1,1)\begin{pmatrix} 220 \\ 210 \end{pmatrix}=215$ 。

$\sigma_{LS}^2=\frac{1}{2}(1,1)\begin{pmatrix} 64 & 0 \\ 0 & 16 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\frac{1}{2}=20$ 。

由此可见:  $\sigma_{n_2}^2 < \sigma_{LS}^2 < \sigma_{n_1}^2$ 。说明由于缺乏有关统计知识, 最小二乘估值并非最佳估值。

② 求加权最小二乘估值 $\hat{v}_{LSW}$ 。

由加权最小二乘估值公式:

$\hat{v}_{LSW}=(\mathbf{C}^T\mathbf{W}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{W}\mathbf{z}$ ,  $\sigma_{LSW}^2=(\mathbf{C}^T\mathbf{W}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{W}\mathbf{R}_n\mathbf{W}\mathbf{C}(\mathbf{C}^T\mathbf{W}\mathbf{C})^{-1}$ 。

1) 套公式求解:



当  $\mathbf{W}=\mathbf{R}_n^{-1}=\begin{pmatrix} 1/8^2 & 0 \\ 0 & 1/4^2 \end{pmatrix}$  时，取得最佳估值。

此时， $\mathbf{C}^T\mathbf{W}\mathbf{C}=(1,1)\begin{pmatrix} 1/8^2 & 0 \\ 0 & 1/4^2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}=\frac{1}{64}+\frac{1}{16}=\frac{5}{64}$ ，则  $(\mathbf{C}^T\mathbf{W}\mathbf{C})^{-1}=\frac{64}{5}$ 。

故， $\hat{v}_{LSW}=\frac{64}{5}(1,1)\begin{pmatrix} 1/8^2 & 0 \\ 0 & 1/4^2 \end{pmatrix}\begin{pmatrix} 220 \\ 210 \end{pmatrix}=212$ 。

且： $\sigma_{LSW}^2=(\mathbf{C}^T\mathbf{W}\mathbf{C})^{-1}=\frac{64}{5}=12.8$ 。

由此可见： $\sigma_{LSW}^2 < \sigma_{n2}^2 < \sigma_{LS}^2 < \sigma_{n1}^2$ 。说明由于具有有关统计知识，加权最小二乘估值是最佳估值，既优于最小二乘估值，又优于任一个单一估值。

2) 求极值方法：

设  $\mathbf{W}=\begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix}$ ，则  $\mathbf{C}^T\mathbf{W}\mathbf{C}=(1,1)\begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}=w_1+w_2$ 。

从而， $\sigma_{LSW}^2=(w_1+w_2)^{-1}(1,1)\begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix}\begin{pmatrix} 64 & 0 \\ 0 & 16 \end{pmatrix}\begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}(w_1+w_2)^{-1}$   

$$=\frac{64w_1^2+16w_2^2}{(w_1+w_2)^2}$$

为使  $\sigma_{LSW}^2$  最小，令偏导数为零，即  $\frac{\partial}{\partial w_1}\sigma_{LSW}^2=0$  及  $\frac{\partial}{\partial w_2}\sigma_{LSW}^2=0$ ，

解方程得到： $w_2=4w_1$ 。

所以，最小  $\sigma_{LSW}^2=\frac{64w_1^2+16(4w_1)^2}{(w_1+4w_1)^2}=12.8$ 。

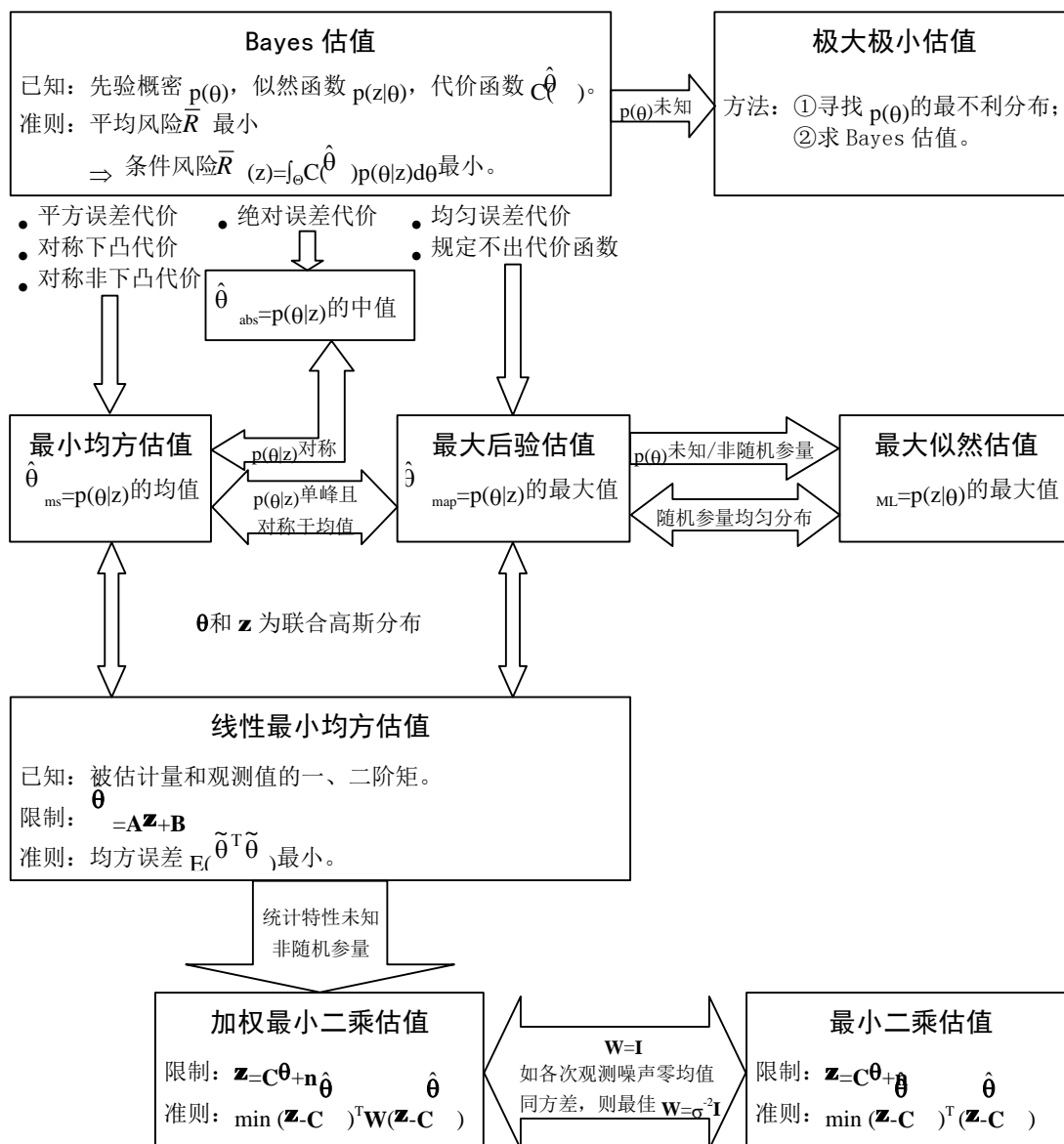
此时，我们有： $\hat{v}_{LSW}=(w_1+w_2)^{-1}(1,1)\begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix}\begin{pmatrix} 220 \\ 210 \end{pmatrix}=212$ 。

可见：所得结果与 1) 中套公式的结果一样。

#解毕

## 2.10 小结

◆ 各类估值准则之间的关系：



### ◆ 各类估值的计算公式及其性质：

- 常用估值准则：
  - Bayes 估值、极大极小估值、最大后验估值、最大似然估值
  - 多参量估值
  - 线性最小均方估值（正交原理）
  - 最小二乘估值、加权最小二乘估值
- 估计量的性质：
  - 无偏估值、一致估值、充分估值、有效估值
- Cramér-Rao 不等式：
  - 非随机参量、随机参量
- 高斯信道中单参量信号的估值：
  - 白高斯信道
  - 色高斯信道

## 第3章 波形估计理论

信号波形复原/波形估计：将信号（观测时间  $T$  内连续）过滤出来。

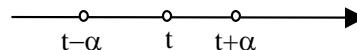
例如：模拟通信系统中，载送信息的调制信号作用在载波上产生已调信号，再传输。在接收端，对信号波形进行估计。

**估值描述**：给定两个有关联的随机过程  $z(t)$  和  $s(t)$ ，希望利用  $z(t)$  或它的某些值去估计（信号） $s(t)$  的各种参量  $g(t)$ 。

**估计基础**： $z(t)$  在一定时刻  $t=\xi$  的值， $\xi$  属于时间轴上集合  $I$ （离散点或区间）。

**问题**：寻找这些数据  $z(\xi)$  的适当变换，作出  $g(t)$  的最佳估计  $\hat{g}(t)$ 。

**问题类型**：



- ① **预测问题**： $g(t) = s(t+\alpha)$ ，其中  $\alpha > 0$ 。

滤波器实际输出    滤波器理想输出

- ② **滤波问题**： $g(t) = s(t)$ 。

- ③ **平滑（内插）问题**：对任意属于观测区间的  $t$ ，估计  $s(t)$ 。

即：待估计量为  $g(t)=s(t)$ ， $t \in$  观测区间。

固定滞后平滑：基于  $[t_0, t]$  的观测，估计  $s(t-\alpha)$ ，其中  $\alpha > 0$ 。

### 3.1 波形的线性最小均方估计

◆ **线性最小均方估计**：

寻找线性估值（变换）： $\hat{g}(t) = L[z(\xi)]$ ，其中  $\xi \in I$ ， $L[\bullet]$  为线性算子。

**目的**：使均方误差  $e^2 = E([g(t) - \hat{g}(t)]^2) = E([g(t) - L[z(\xi)]]^2)$  最小。

为使均方误差  $e^2$  最小，可应用正交原理，即：

若  $E([g(t) - L[z(\xi)]] \cdot z(\xi_i)) = 0$ ， $\xi_i \in I$ ，则均方误差最小。

此时， $e^2_{\min} = E([g(t) - L[z(\xi)]] \cdot [g(t) - L[z(\xi)]]) = E([g(t) - L[z(\xi)]] \cdot g(t))$ 。

证明：

假设对  $L[\bullet]$  算子正交原理成立。令  $L_1[\bullet]$  为另一线性算子，

$$\begin{aligned} E([g(t) - L_1[z(\xi)]]^2) &= E([g(t) - L[z(\xi)] + \underline{L[z(\xi)] - L_1[z(\xi)]}]^2) \\ &= E([g(t) - L[z(\xi)]] + L_2[z(\xi)]^2) \\ &= E([g(t) - L[z(\xi)]]^2) + 2E(\underline{[g(t) - L[z(\xi)]] \cdot L_2[z(\xi)]}) + E([L_2[z(\xi)]]^2) \\ &= E([g(t) - L[z(\xi)]]^2) + E([L_2[z(\xi)]]^2) \end{aligned}$$

故  $E([g(t) - L_1[z(\xi)]]^2) \geq E([g(t) - L[z(\xi)]]^2)$ ，

当且仅当  $E([L_2[z(\xi)]]^2) = E([L[z(\xi)] - L_1[z(\xi)]]^2) = 0$  时，上式中等号成立。

所以，我们以概率 1 有：当  $L_1[\bullet] = L[\bullet]$  时，估值的均方误差最小。

#证毕

◆ **例**：考虑一个内插问题，即：给定区间  $[0, T]$  两个端点的信号值（即  $s(0)$  和  $s(T)$ ），已知信号的自相关函数  $R_s(\tau)$ ，要求估计区间内任意时刻的信号值。

本问题中:  $g(t) = s(t)$ ,  $t \in [0, T]$  且  $z(t) = \{s(\xi): \xi=0, T\}$ 。

令线性估计量  $\hat{g}(t) = a \cdot s(0) + b \cdot s(T)$ ,  $t \in [0, T]$ 。

则由正交原理:  $E([s(t) - a \cdot s(0) - b \cdot s(T)] \cdot s(0)) = 0$  且  $E([s(t) - a \cdot s(0) - b \cdot s(T)] \cdot s(T)) = 0$

从而可得:  $a \cdot R_s(0) + b \cdot R_s(T) = R_s(t)$  且  $a \cdot R_s(T) + b \cdot R_s(0) = R_s(T-t)$

解方程得:  $a = \frac{R_s(0) \cdot R_s(t) - R_s(T) \cdot R_s(T-t)}{R_s^2(0) - R_s^2(T)}$ ,  $b = \frac{R_s(0) \cdot R_s(T-t) - R_s(t) \cdot R_s(T)}{R_s^2(0) - R_s^2(T)}$ 。

可见: 当  $t=0$  时,  $a=1$ ,  $b=0$ ; 当  $t=T$  时,  $a=0$ ,  $b=1$ 。

#例毕

#### ◆最佳线性滤波器:

用  $\{z(\xi): \xi \in [0, T]\}$  的线性组合来估计  $g(t)$ , 使均方误差最小。

将该线性组合用下式来表示:

$$\begin{aligned}\hat{g}(t) &= \lim_{\Delta\xi \rightarrow 0} \sum_{i=1}^N h(t, \xi_i) \cdot z(\xi_i) \cdot \Delta\xi, \text{ 其中 } N \cdot \Delta\xi = T \\ &= \int_0^T h(t, \xi) z(\xi) d\xi\end{aligned}$$

用滤波器来表示, 如右图。  $z(t) \longrightarrow \boxed{h(t, \xi)} \longrightarrow \hat{g}(t)$

下面讨论  $h(t, \xi)$  的求解。

方法之一: 用正交原理求解。

由  $E([g(t) - \hat{g}(t)] \cdot z(\tau)) = 0$ ,  $\tau \in [0, T]$  得:  $E([g(t) - \int_0^T h(t, \xi) z(\xi) d\xi] \cdot z(\tau)) = 0$

故, 我们得到 **Wiener-Hopf 方程**:  $R_{gz}(t, \tau) = \int_0^T h(t, \xi) R_z(\xi, \tau) d\xi$ ,  $\tau \in [0, T]$

解此方程即可得到  $h(t, \xi)$  的解。

此时, 均方误差  $e^2_{\min} = E([g(t) - \int_0^T h(t, \xi) z(\xi) d\xi] \cdot g(t)) = R_g(t, t) - \int_0^T h(t, \xi) R_{gz}(t, \xi) d\xi$ 。

方法之二: 用变分法求解。

令  $\hat{g}(t) = \int_0^T h_1(t, \xi) z(\xi) d\xi$ , 我们希望求最佳  $h_1(t, \xi)$ , 以使下面的均方误差为最小:

$$e^2 = E([g(t) - \int_0^T h_1(t, \xi) z(\xi) d\xi]^2)。$$

将均方误差重写为:  $e^2 = R_g(t, t) + \int_0^T \int_0^T h_1(t, \xi) h_1(t, \eta) R_z(\eta, \xi) d\eta d\xi - 2 \int_0^T h_1(t, \xi) R_{gz}(t, \xi) d\xi$

假定最佳滤波器为  $h(t, \xi)$ 。

我们可令  $h_1(t, \xi) = h(t, \xi) + \varepsilon f(t, \xi)$ , 其中  $\varepsilon$  为小的扰动因子,  $f(t, \xi)$  为扰动函数。

利用上式, 我们可将均方误差表达式记为  $e^2(\varepsilon)$ 。

采用变分法, 令  $\frac{de^2(\varepsilon)}{d\varepsilon} \big|_{\varepsilon=0} = 0$ , 可得:  $\int_0^T f(t, \xi) [R_{gz}(t, \xi) - \int_0^T h(t, \eta) R_z(\eta, \xi) d\eta] d\xi = 0$

故有:  $R_{gz}(t, \xi) = \int_0^T h(t, \eta) R_z(\eta, \xi) d\eta$ ,  $\xi \in [0, T]$

可见我们得到与方法一所得相同的结果。

## 3.2 Wiener 滤波

假定:  $z(t)$  平稳, 则  $R_z(t, \tau) = R_z(t - \tau)$ ;  $g(t)$  与  $z(t)$  联合平稳, 则  $R_{gz}(t, \tau) = R_{gz}(t - \tau)$ 。

那么:  $\hat{g}(t) = \int_0^T h(t - \xi) z(\xi) d\xi$

Wiener-Hopf 方程改写为:  $R_{gz}(t - \tau) = \int_0^T h(t - \xi) R_z(\xi - \tau) d\xi, \tau \in [0, T]$

估值均方误差:  $e_{\min}^2 = R_g(0) - \int_0^T h(t - \xi) R_{gz}(t - \xi) d\xi$ 。

### 3.2.1 物理不可实现的 Wiener 滤波器

在此, 我们将观测区间  $[0, T]$  拓展到  $(-\infty, +\infty)$ , 则:

- 1) 由于  $\hat{g}(t) = \int_{-\infty}^{+\infty} h(t - \xi) z(\xi) d\xi$ , 故估值不仅利用了估值时刻  $t$  以前的数据  $\{z(\xi): \xi \in (-\infty, t]\}$ , 而且也使用了估值时刻  $t$  以后的数据  $\{z(\xi): \xi \in (t, +\infty)\}$ ;
- 2) 作变量代换  $\eta = t - \xi$ , 我们可将估值表示为:  $\hat{g}(t) = \int_{-\infty}^{+\infty} h(\eta) z(t - \eta) d\eta$ 。可见: 滤波器的冲激响应  $h(\eta)$  存在于区间  $(-\infty, +\infty)$  上, 说明它是物理不可实现的。

在 Wiener-Hopf 积分方程中, 令  $t - \tau = u, t - \xi = v$ , 则可将该方程重写为:

$$R_{gz}(u) = \int_{-\infty}^{+\infty} h(v) R_z(u - v) dv, \quad -\infty < u < +\infty$$

等式右端为一卷积形式, 故作双边拉氏变换, 则在频域上可表示为:

$\Phi_{gz}(s) = H(s) \Phi_z(s)$  其中:  $\Phi_{gz}(s)$  称互谱函数,  $H(s)$  称传输函数,  $\Phi_z(s)$  称功率谱函数。

所以, 我们得到最佳线性滤波器的传输函数为:  $H(s) = \frac{\Phi_{gz}(s)}{\Phi_z(s)}$

此时, 估值方差为:  $e_{\min}^2 = R_g(0) - \int_{-\infty}^{+\infty} h(v) R_{gz}(v) dv$ 。

利用 Parseval 定理, 可得:  $e_{\min}^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi_g(\omega) d\omega - \frac{1}{2\pi} \int_{-\infty}^{+\infty} H(\omega) \Phi_{gz}^*(\omega) d\omega$

再代入  $H(\omega)$  的表达式:  $e_{\min}^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left( \Phi_g(\omega) - \frac{|\Phi_{gz}(\omega)|^2}{\Phi_z(\omega)} \right) d\omega$

◆ 加性噪声:

$z(t) = s(t) + n(t)$ ,  $s(t)$  和  $n(t)$  相互独立,  $n(t)$  的均值为零。

考虑滤波问题:  $g(t) = s(t)$ 。

从而可得:  $\Phi_{gz}(\omega) = \Phi_s(\omega)$ ,  $\Phi_z(\omega) = \Phi_s(\omega) + \Phi_n(\omega)$ 。

所以, 物理不可实现的 Wiener 滤波器的传输函数为:

$$H(\omega) = \frac{\Phi_s(\omega)}{\Phi_s(\omega) + \Phi_n(\omega)} = \frac{1}{1 + (\Phi_s(\omega)/\Phi_n(\omega))^{-1}}$$

此时,  $e_{\min}^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left( \Phi_s(\omega) - \frac{|\Phi_s(\omega)|^2}{\Phi_s(\omega) + \Phi_n(\omega)} \right) d\omega = \left( \text{若 } \Phi_s(\omega) \text{ 为实函数} \right) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\Phi_s(\omega) \Phi_n(\omega)}{\Phi_s(\omega) + \Phi_n(\omega)} d\omega$ 。

◆例：已知： $\Phi_s(s) = \frac{2}{1-s^2}$ ， $\Phi_n(s) = 1$ 。

求物理不可实现 Wiener 滤波器的冲激响应  $h(t)$  及其估值均方误差  $e_{\min}^2$ 。

解：由上述公式可得： $H(s) = \frac{1}{1 + \left(2/(1-s^2)\right)^{-1}} = \frac{2}{3-s^2}$ 。

作双边拉氏反变换得： $h(t) = \frac{1}{\sqrt{3}} e^{-\sqrt{3}|t|}$ 。它是物理不可实现的（在 $(-\infty, 0)$ 上有响应值）。

此时， $e_{\min}^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{2}{2 + (1+\omega^2)} d\omega = \frac{1}{\sqrt{3}} = 0.577$ 。

#解毕

### 3.2.2 物理可实现的 Wiener 滤波器

在此，我们将观测区间 $[0, T]$ 拓展到 $(-\infty, t)$ ，则：

1) 由于  $\hat{g}(t) = \int_{-\infty}^t h(t-\xi) \underline{z(\xi)} d\xi$ ，故估值仅利用了估值时刻  $t$  以前的数据，即  $\{z(\xi): \xi \in (-\infty, t]\}$ ；

2) 作变量代换  $\eta = t - \xi$ ，我们可将估值表示为： $\hat{g}(t) = \int_0^{+\infty} \underline{h(\eta)} z(t-\eta) d\eta$ 。可见：滤波器的冲激响应  $h(\eta)$  仅存在于区间 $[0, +\infty)$ 上，说明它是物理可实现的。

此时，Wiener-Hopf 积分方程重写为： $R_{gz}(t-\tau) = \int_{-\infty}^t h(t-\xi) R_z(\xi-\tau) d\xi$ ， $\tau \in (-\infty, t]$  或  $\tau \leq t$

令  $t-\tau=u$ ， $t-\xi=v$ ，则可将该方程重写为： $R_{gz}(u) = \int_0^{+\infty} h(v) R_z(u-v) dv$ ， $u \geq 0$

对此方程的求解，可采取下述的频谱因式分解法和预白化法两种方法之一。

#### ◆方法之一：频谱因式分解法

可将  $h(v)$  写为： $h(v) = \begin{cases} h(v), & v \geq 0 \\ 0, & v < 0 \end{cases} \xleftrightarrow{\mathcal{L}} H(s)$ ，其零极点在  $s$  平面的左半平面。

则 Wiener-Hopf 方程可写为： $\int_{-\infty}^{+\infty} h(v) R_z(u-v) dv - R_{gz}(u) = 0$ ， $u \geq 0$

为使方程对所有的  $u$  值（即 $-\infty < u < +\infty$ ）成立，须将其右端加以改造：

令  $q(u) = \begin{cases} 0, & u \geq 0 \\ q(u), & u < 0 \end{cases} \xleftrightarrow{\mathcal{L}} Q(s)$ ，其零极点在  $s$  平面的右半平面。

其中  $q(u)$  为未知函数。

这样有改写后的 Wiener-Hopf 方程： $\int_{-\infty}^{+\infty} h(v) R_z(u-v) dv - R_{gz}(u) = q(u)$ ， $-\infty < u < +\infty$

作双边拉氏变换得： $H(s) \Phi_z(s) - \Phi_{gz}(s) = Q(s)$

在实际应用中， $z(t)$  为实函数且为平稳随机过程，所以  $R_z(\tau)$  是实偶函数，则  $\Phi_z(s)$  为有理谱，它可分为零极点在左半平面和零极点在右半平面两部分乘积，即：

$$\Phi_z(s) = \Phi_z^+(s) \cdot \Phi_z^-(s),$$

其中： $\Phi_z^+(s)$  的零极点在左半平面， $\Phi_z^-(s) = \Phi_z^+(-s)$  的零极点在右半平面。

从而可得： $H(s) \cdot \Phi_z^+(s) = \frac{Q(s)}{\Phi_z^-(s)} + \frac{\Phi_{gz}(s)}{\Phi_z^-(s)}$

上式右端的第二项可再分解为两部分之和： $\frac{\Phi_{gz}(s)}{\Phi_z^-(s)} = \left[ \frac{\Phi_{gz}(s)}{\Phi_z^-(s)} \right]^{t^+} + \left[ \frac{\Phi_{gz}(s)}{\Phi_z^-(s)} \right]^{t^-}$ ，其中 $[\bullet]^{t^+}$ 的零极点在左半平面（在时域中对应于  $t > 0$  的那一部分冲激响应函数）， $[\bullet]^{t^-}$ 的零极点在右半平面（在时域中对应于  $t < 0$  的那一部分冲激响应函数）。

由于等式两边在  $s$  平面上左半平面中的零极点应相等，故得： $H(s) \cdot \Phi_z^+(s) = \left[ \frac{\Phi_{gz}(s)}{\Phi_z^-(s)} \right]^{t^+}$

$$\text{即：} \quad H(s) = \frac{1}{\Phi_z^+(s)} \left[ \frac{\Phi_{gz}(s)}{\Phi_z^-(s)} \right]^{t^+} \quad \text{或} \quad H(\omega) = \frac{1}{\Phi_z^+(\omega)} \left[ \frac{\Phi_{gz}(\omega)}{\Phi_z^-(\omega)} \right]^{t^+}$$

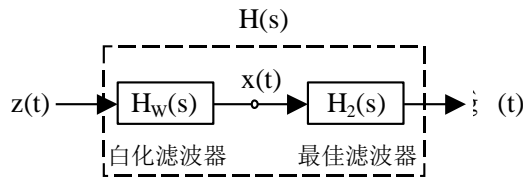
这就是物理可实现的 **Wiener 滤波器**。

#### ◆方法之二：预白化法

若  $z(t)$  为白信号，即  $R_z(\tau) = \delta(\tau)$ ，则由 Wiener-Hopf 方程得：

$$h(t) = \begin{cases} R_{gz}(t), & t \geq 0 \\ 0, & t < 0 \end{cases} \xleftrightarrow{\mathcal{L}} H(s) = [\Phi_{gz}(s)]^{t^+}$$

若  $z(t)$  为非白信号，则可先将  $z(t)$  通过白化滤波器  $H_W(s)$  变成白信号  $x(t)$ ，再对  $x(t)$  采用与上面类似的最佳滤波器  $H_2(s)$ ，从而得到最佳估计  $\hat{g}(t)$ 。那么，总的滤波器  $H(s) = H_W(s)H_2(s)$  就是我们要求的 **Wiener 滤波器**（见下图）。



与前面类似，我们可将  $\Phi_z(s)$  作有理分解： $\Phi_z(s) = \Phi_z^+(s) \Phi_z^-(s)$ ，且  $\Phi_z^-(s) = \Phi_z^+(-s)$ 。

则，白化滤波器应为： $H_W(s) = 1 / \Phi_z^+(s)$ 。

又，对应于  $x(t)$  的最佳滤波器应为（用  $x(t)$  代替  $z(t)$ ）： $H_2(s) = [\Phi_{gx}(s)]^{t^+}$ 。

现在，须求出  $\Phi_{gx}(s)$  的表达式。由于

$$R_{gx}(\tau) = E(g(t) x(t-\tau)) = E(g(t) \int_{-\infty}^{+\infty} h_W(\xi) z(t-\tau-\xi) d\xi) = \int_{-\infty}^{+\infty} h_W(\xi) R_{gz}(\tau+\xi) d\xi$$

$$\xrightarrow{\text{令 } u = -\xi} \int_{-\infty}^{+\infty} h_W(-u) R_{gz}(\tau-u) du$$

$$\text{等式两端作双边拉氏变换得：} \quad \Phi_{gx}(s) = H_W(-s) \Phi_{gz}(s) = \frac{\Phi_{gz}(s)}{\Phi_z^+(-s)} = \frac{\Phi_{gz}(s)}{\Phi_z^-(s)}。$$

$$\text{故，我们得到：} H_2(s) = \left[ \frac{\Phi_{gz}(s)}{\Phi_z^-(s)} \right]^{t^+}。$$

$$\text{因此，我们最终得到的 Wiener 滤波器为：} H(s) = H_W(s) H_2(s) = \frac{1}{\Phi_z^+(s)} \left[ \frac{\Phi_{gz}(s)}{\Phi_z^-(s)} \right]^{t^+}。$$

#### ◆Wiener 滤波器的估值方差：

对于物理可实现的 **Wiener 滤波器**，其均方误差为：

$$\begin{aligned} e_{\min}^2 &= R_g(0) - \int_{-\infty}^t h(t-\xi) R_{gz}(t-\xi) d\xi = [t-\xi=v] = R_g(0) - \int_0^{+\infty} h(v) R_{gz}(v) dv \\ &= R_g(0) - \int_{-\infty}^{+\infty} h(v) R_{gz}(v) dv \end{aligned}$$

$$\text{又 } H(s) = \frac{1}{\Phi_z^+(s)} [\Phi_{gx}(s)]^+ = \frac{1}{\Phi_z^+(s)} \int_0^{+\infty} R_{gx}(\tau) e^{-s\tau} d\tau$$

$$\text{故, } h(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} H(s) e^{st} ds = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{1}{\Phi_z^+(s)} \int_0^{+\infty} R_{gx}(\tau) e^{-s\tau} d\tau \cdot e^{st} ds。$$

将  $h(t)$  的表达式代入  $e_{\min}^2$  中的第二项  $\int_{-\infty}^{+\infty} h(v) R_{gz}(v) dv$  可得:

$$\begin{aligned} \int_{-\infty}^{+\infty} h(v) R_{gz}(v) dv &= \int_{-\infty}^{+\infty} R_{gz}(v) \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{1}{\Phi_z^+(s)} \int_0^{+\infty} R_{gx}(\tau) e^{-s\tau} d\tau \cdot e^{sv} ds dv \\ &= \int_0^{+\infty} R_{gx}(\tau) \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{1}{\Phi_z^+(s)} \int_{-\infty}^{+\infty} R_{gz}(v) e^{sv} dv e^{-s\tau} ds d\tau \\ &= \int_0^{+\infty} R_{gx}(\tau) \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{1}{\Phi_z^+(s)} \frac{\Phi_{gz}(-s) e^{-s\tau}}{ds} ds d\tau = \int_0^{+\infty} R_{gx}(\tau) \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{\Phi_{gx}(-s) e^{-s\tau}}{ds} ds d\tau \\ &= \int_0^{+\infty} R_{gx}(\tau) R_{gx}^*(\tau) d\tau = \int_0^{+\infty} R_{gx}^2(\tau) d\tau \end{aligned}$$

所以, 我们最后得到估值均方误差:  $e_{\min}^2 = R_g(0) - \int_0^{+\infty} R_{gx}^2(\tau) d\tau$

讨论:  $g(t)=s(t+\alpha)$  时,

$$R_g(0)=E(g(t)g(t))=R_s(0) \text{ 且 } R_{gx}(\tau)=E(g(t+\tau)x(t))=E(s(t+\tau+\alpha)x(t))=R_{sx}(\tau+\alpha)。$$

$$\text{故, } e_{\min}^2 = R_s(0) - \int_0^{+\infty} R_{sx}^2(\tau+\alpha) d\tau = R_s(0) - \int_{\alpha}^{+\infty} R_{sx}^2(t) dt。$$

可见: 当  $\alpha \rightarrow +\infty$  时,  $e_{\min}^2 \rightarrow R_s(0)$ , 此时的均方误差最大;

当  $\alpha \rightarrow -\infty$  时,  $e_{\min}^2$  达到下限 (此下限称为不可减误差)。此时, 我们利用了信号的全部信息, 其对应的滤波器是物理不可实现的, 或是非实时处理的;

当  $\alpha$  取其它值时,  $e_{\min}^2(\alpha > 0, \text{预测}) > e_{\min}^2(\alpha = 0, \text{滤波}) > e_{\min}^2(\alpha < 0, \text{平滑})。$

◆例: 已知:  $z(t) = s(t) + n(t)$ ,  $\Phi_s(s) = \frac{2}{1-s^2}$ ,  $\Phi_n(s) = 1$ 。

求物理可实现的 Wiener 滤波器  $h(t)$  及其估值方差  $e_{\min}^2$ 。

解: 由  $\Phi_z(s) = \Phi_s(s) + \Phi_n(s) = \frac{3-s^2}{1-s^2} = \frac{\sqrt{3}+s}{1+s} \frac{\sqrt{3}-s}{1-s}$  可知:

$$\Phi_z^+(s) = \frac{\sqrt{3}+s}{1+s} \quad \text{及} \quad \Phi_z^-(s) = \frac{\sqrt{3}-s}{1-s}。$$

对滤波问题,  $g(t) = s(t)$ , 故  $\Phi_{gz}(s) = \Phi_s(s) = \frac{2}{1-s^2}$ 。因此, 我们有:

$$\frac{\Phi_{gz}(s)}{\Phi_z^-(s)} = \frac{2}{1-s^2} \cdot \frac{\sqrt{3}-s}{1-s} = \frac{2}{(1+s)(\sqrt{3}-s)} = \frac{\sqrt{3}-1}{1+s} + \frac{\sqrt{3}-1}{\sqrt{3}-s}$$

$$\text{故, } \left[ \frac{\Phi_{gz}(s)}{\Phi_z^-(s)} \right]^+ = \frac{\sqrt{3}-1}{1+s}。$$



$$\text{所以, } H(s) = \frac{1}{\Phi_z^+(s)} \left[ \frac{\Phi_{gz}(s)}{\Phi_z^-(s)} \right]^{t^+} = \frac{1+s}{\sqrt{3}+s} \frac{\sqrt{3}-1}{1+s} = \frac{\sqrt{3}-1}{\sqrt{3}+s} \xleftrightarrow{\mathcal{L}} h(t) = (\sqrt{3}-1)e^{-\sqrt{3}t} u(t)。$$

由于  $e^2_{\min} = R_s(0) - \int_0^{+\infty} R_{gx}^2(\tau) d\tau$ , 故须先求  $R_s(0)$  及  $R_{gx}(\tau)$ 。

因  $\Phi_s(\omega) \xleftrightarrow{\mathcal{F}} R_s(\tau) = e^{-|\tau|}$ , 故  $R_s(0)=1$ 。

因  $\Phi_{gx}(\omega) = \frac{\Phi_{gz}(\omega)}{\Phi_z^-(\omega)} = \frac{\sqrt{3}-1}{1+j\omega} + \frac{\sqrt{3}-1}{\sqrt{3}-j\omega}$ , 而由于在  $e^2_{\min}$  的表达式中我们所求关于  $R_{gx}(\tau)$  的积分限为  $0 \rightarrow +\infty$ , 故我们只需考虑  $\Phi_{gx}(\omega)$  部分分解中极点在左半平面的那一部分 (对应于  $R_{gx}(\tau)$  中  $\tau \geq 0$  的部分), 即

$$[\Phi_{gx}(\omega)]^{\tau^+} = \frac{\sqrt{3}-1}{1+j\omega} \xleftrightarrow{\mathcal{F}} [R_{gx}(\tau)]^{\tau^+} = (\sqrt{3}-1)e^{-\tau} u(\tau)$$

最后, 我们得到:

$$e^2_{\min} = 1 - \int_0^{+\infty} (\sqrt{3}-1)^2 e^{-2\tau} d\tau = \sqrt{3}-1 = 0.732$$

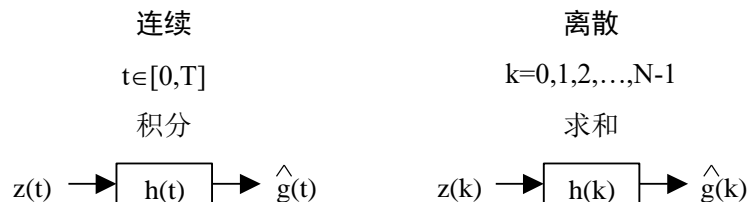
由此可见: 物理可实现  $h(t)$  的估值方差比物理不可实现  $h(t)$  的估值方差要大 (在前例中我们曾算得后者为 0.577)。

#解毕

### 3.3 离散时间 Wiener 滤波

#### 3.3.1 有限观测区间的离散 Wiener 滤波器

连续和离散情况比较:



设  $g(k)$  为理想输出, 而实际输出为  $\hat{g}(k) = \sum_{i=0}^{N-1} h(i) z(k-i)$ , 我们希望求得最佳  $\{h(i)\}$  以使均方误差  $E([g(k) - \hat{g}(k)]^2)$  最小。

由正交原理:  $E([g(k) - \hat{g}(k)] \cdot z(l)) = 0 \quad l = k, k-1, \dots, k-(N-1)$

展开并整理得:  $r_{gz}(k-l) = \sum_{i=0}^{N-1} h(i) r_z(k-i-l) \quad l = k, k-1, \dots, k-(N-1)$

替换变量得:  $r_{gz}(j) = \sum_{i=0}^{N-1} h(i) r_z(j-i) \quad j = k-l = 0, 1, \dots, N-1$

写成矩阵形式:  $\mathbf{R}_{gz} = \mathbf{R}_z \cdot \mathbf{H}$   
 $N \times 1 \quad N \times N \quad N \times 1$

其中  $\mathbf{H} = (h(0), h(1), \dots, h(N-1))^T$

$$\mathbf{R}_{gz} = (r_{gz}(0), r_{gz}(1), \dots, r_{gz}(N-1))^T$$

$$\mathbf{R}_z = \begin{pmatrix} r_z(0) & r_z(1) & \cdots & r_z(N-1) \\ r_z(1) & r_z(0) & \cdots & r_z(N-2) \\ \vdots & \vdots & \cdots & \vdots \\ r_z(N-1) & r_z(N-2) & \cdots & r_z(0) \end{pmatrix}$$

从而可得最佳滤波器:  $\mathbf{H} = \mathbf{R}_z^{-1} \mathbf{R}_{gz}$

其估值方差 (即最小均方误差)  $e_{\min}^2 = E([g(k) - \hat{g}(k)]^2) = E([g(k) - \hat{g}(k)] g(k))$

$$= r_g(0) - \sum_{i=0}^{N-1} h(i) r_{gz}(i) = r_g(0) - \mathbf{H}^T \cdot \mathbf{R}_{gz}$$

### 3.3.2 物理可实现的离散 Wiener 滤波器

若令上节中的  $N \rightarrow +\infty$ , 即可得到物理可实现的滤波器。

此时有:  $\hat{g}(k) = \sum_{i=0}^{+\infty} h(i) z(k-i)$

$$r_{gz}(j) = \sum_{i=0}^{+\infty} h(i) r_z(j-i) \quad j \geq 0 \quad \text{离散 Wiener-Hopf 方程}$$

其估值方差:  $e_{\min}^2 = r_g(0) - \sum_{i=0}^{+\infty} h(i) r_{gz}(i)$

与连续情况时的推导类似, 利用 Z 变换 ( $H(z) = \sum_{k=-\infty}^{+\infty} h(k) z^{-k}$ ) 及频谱分解法可得:

$$H(z) = \frac{1}{\Phi_z^+(z)} \left[ \frac{\Phi_{gz}(z)}{\Phi_z^-(z)} \right]^{k^+} \quad \text{离散 Wiener 滤波器}$$

其中:  $\Phi_z^+(z)$  的零极点均在单位圆内;  $\Phi_z^-(z)$  的零极点均在单位圆外;

$$\left[ \frac{\Phi_{gz}(z)}{\Phi_z^-(z)} \right]^{k^+} \text{ 是 } \left[ \frac{\Phi_{gz}(z)}{\Phi_z^-(z)} \right] = \left[ \frac{\Phi_{gz}(z)}{\Phi_z^-(z)} \right]^{k^+} + \left[ \frac{\Phi_{gz}(z)}{\Phi_z^-(z)} \right]^{k^-} \text{ 中的第一项 (其极点在单位圆内)}。$$

### 3.3.3 物理不可实现的离散 Wiener 滤波器

若令前一节中的  $[0, N] \rightarrow (-\infty, +\infty)$ , 即可得到物理不可实现的滤波器。

此时有:  $\hat{g}(k) = \sum_{i=-\infty}^{+\infty} h(i) z(k-i)$

$$r_{gz}(j) = \sum_{i=-\infty}^{+\infty} h(i) r_z(j-i) \quad -\infty < j < +\infty$$

$$\text{两边同求 Z 变换可得: } H(z) = \frac{\Phi_{gz}(z)}{\Phi_z(z)}$$

其均方误差:  $e_{\min}^2 = r_g(0) - \sum_{i=-\infty}^{+\infty} h(i) r_{gz}(i)$

◆例:

$z(k) = s(k) + n(k)$ ,  $g(k) = s(k)$  (滤波),  $r_{gn}(k) = 0$  (不相关)。

$n(k)$  为平稳白噪声,  $r_n(m) = V \delta(m) \xleftrightarrow{\mathcal{Z}} \Phi_n(z) = V$

$s(k)$  为一阶 AR 过程,  $r_s(m) = a^{|m|} (0 < a < 1) \xleftrightarrow{\mathcal{Z}} \Phi_s(z) = \frac{a - a^{-1}}{(z + z^{-1}) - (a + a^{-1})}$

分别求物理不可实现及物理可实现滤波器。

解：①物理不可实现滤波器：

$$\begin{aligned} H(z) &= \frac{\Phi_{gz}(z)}{\Phi_z(z)} = \frac{\Phi_s(z)}{\Phi_s(z) + \Phi_n(z)} = \frac{a - a^{-1}}{a - a^{-1} + V((z + z^{-1}) - (a + a^{-1}))} \\ &= \frac{(a - a^{-1})/V}{(z + z^{-1}) - (b + b^{-1})} \quad \text{其中 } b + b^{-1} = a + a^{-1} - (a - a^{-1})/V \quad 0 < b < 1 \end{aligned}$$

其对应的时域滤波器为：

$$h(k) = \frac{a^{-1} - a}{V(b^{-1} - b)} b^{|k|}$$

此时的均方误差：

$$e_{\min}^2 = 1 - \frac{a^{-1} - a}{V(b^{-1} - b)} \sum_{i=-\infty}^{+\infty} (ab)^{|i|} = 1 - \frac{a^{-1} - a}{V(b^{-1} - b)} \frac{1 + ab}{1 - ab}$$

②物理可实现滤波器：

$$\Phi_z(z) = \Phi_s(z) + \Phi_n(z) = \frac{z - b}{z - a} \bullet V \frac{z - b^{-1}}{z - a^{-1}}$$

$$\text{即 } \Phi_z^+(z) = \frac{z - b}{z - a}, \quad \Phi_z^-(z) = V \frac{z - b^{-1}}{z - a^{-1}}$$

$$\text{而 } \frac{\Phi_{gz}(z)}{\Phi_z^-(z)} = \frac{\Phi_s(z)}{\Phi_z^-(z)} = \frac{cz}{z - a} - \frac{cz}{z - b^{-1}} \quad \text{其中 } c = \frac{a^{-1} - a}{V(b^{-1} - a)}$$

$$\text{即 } \left[ \frac{\Phi_{gz}(z)}{\Phi_z^-(z)} \right]^{k+} = \frac{cz}{z - a}, \quad \left[ \frac{\Phi_{gz}(z)}{\Phi_z^-(z)} \right]^{k-} = -\frac{cz}{z - b^{-1}}$$

所以，我们有：

$$H(z) = \frac{1}{\Phi_z^+(z)} \left[ \frac{\Phi_{gz}(z)}{\Phi_z^-(z)} \right]^{k+} = \frac{cz}{z - b}$$

$$h(k) = c b^k u(k)$$

此时的均方误差：

$$e_{\min}^2 = 1 - c \sum_{i=0}^{+\infty} (ab)^{|i|} = 1 - \frac{a^{-1} - a}{V(b^{-1} - a)} \frac{1}{1 - ab}$$

若令  $a=0.8$ ,  $V=1$ , 则有：物理不可实现滤波器的均方误差为 0.3, 而物理可实现滤波器的均方误差为 0.375。可见：前者比后者小。

#解毕

### 3.3.4 离散 Wiener 滤波矩阵

前面所述的是根据一组观测值  $\{z(i), i=k, k-1, \dots, k-(N-1)\}$  来计算估值  $\hat{g}(k)$ , 现在我们希望根据  $N$  个观测值  $\mathbf{z}=(z_1, z_2, \dots, z_N)^T$  来计算  $N$  个信号参量  $\mathbf{G}=(g_1, g_2, \dots, g_N)^T$  的估值  $\hat{\mathbf{G}}=(\hat{g}_1, \hat{g}_2, \dots, \hat{g}_N)^T$ 。例如, 多个观测站对运动中的导弹进行测量, 在某一时刻得到多个观测量, 我们需要根据这些观测量值来估计导弹运动的一些参数。

我们采用的估值准则是线性最小均方估值, 即

令估值  $\hat{\mathbf{G}}$  为观测值  $\mathbf{z}$  的线性组合, 记为

$$\hat{\mathbf{G}} = \mathbf{H} \mathbf{z}, \quad \text{其中 } \mathbf{H} \text{ 为待定的滤波矩阵}$$

$\begin{matrix} \text{Nx1} & \text{NxN} & \text{Nx1} \end{matrix}$

寻找最佳的系数矩阵  $\mathbf{H}$ , 以使均方误差  $E(\tilde{\mathbf{G}}^T \tilde{\mathbf{G}}) = E[(\mathbf{G} - \hat{\mathbf{G}})^T (\mathbf{G} - \hat{\mathbf{G}})]$  为最小。

根据正交原理:

$$E(\tilde{\mathbf{G}} \mathbf{z}^T) = \mathbf{0} \quad \text{即} \quad E[(\mathbf{G} - \mathbf{H}\mathbf{z})\mathbf{z}^T] = \mathbf{0}$$

$\begin{matrix} \text{Nx1} & \text{1xN} & \text{NxN} \end{matrix}$

我们有:  $\mathbf{R}_{gz} = \mathbf{H} \mathbf{R}_z$

故, 可得最佳滤波矩阵 (即 Wiener 滤波矩阵):  $\mathbf{H} = \mathbf{R}_{gz} \mathbf{R}_z^{-1}$

此时, 滤波器的输出为:  $\hat{\mathbf{G}} = \mathbf{H} \mathbf{z} = \mathbf{R}_{gz} \mathbf{R}_z^{-1} \mathbf{z}$

其方差矩阵可计算如下:

$$\begin{aligned} E(\tilde{\mathbf{G}} \tilde{\mathbf{G}}^T) &= E[\tilde{\mathbf{G}} (\mathbf{G} - \hat{\mathbf{G}})^T] = E[\tilde{\mathbf{G}} (\mathbf{G} - \mathbf{R}_{gz} \mathbf{R}_z^{-1} \mathbf{z})^T] = E[\tilde{\mathbf{G}} \mathbf{G}^T] \\ &= E[(\mathbf{G} - \mathbf{R}_{gz} \mathbf{R}_z^{-1} \mathbf{z}) \mathbf{G}^T] = \mathbf{R}_g - \mathbf{R}_{gz} \mathbf{R}_z^{-1} \mathbf{R}_{zg} \end{aligned}$$

其估值的均方误差为:  $E(\tilde{\mathbf{G}}^T \tilde{\mathbf{G}}) = \text{Tr}[E(\tilde{\mathbf{G}} \tilde{\mathbf{G}}^T)] = \text{Tr}[\mathbf{R}_g - \mathbf{R}_{gz} \mathbf{R}_z^{-1} \mathbf{R}_{zg}]$

◆例: 已知  $\mathbf{z} = \mathbf{s} + \mathbf{n}$ ,  $\mathbf{G} = \mathbf{s}$ ,  $\mathbf{s}$  与  $\mathbf{n}$  不相关且其均值均为零。

求最佳滤波矩阵及其估值的均方误差。

解:  $\mathbf{R}_g = \mathbf{R}_s$ ,  $\mathbf{R}_{gz} = \mathbf{R}_s$ ,  $\mathbf{R}_z = \mathbf{R}_s + \mathbf{R}_n$ 。

因此, 我们得到:

$$\begin{aligned} \mathbf{H} &= \mathbf{R}_{gz} \mathbf{R}_z^{-1} = \mathbf{R}_s (\mathbf{R}_s + \mathbf{R}_n)^{-1} \\ \hat{\mathbf{G}} &= \mathbf{H} \mathbf{z} = \mathbf{R}_s (\mathbf{R}_s + \mathbf{R}_n)^{-1} \mathbf{z} \end{aligned}$$

其估值的均方误差为:

$$\begin{aligned} E(\tilde{\mathbf{G}}^T \tilde{\mathbf{G}}) &= \text{Tr}[\mathbf{R}_g - \mathbf{R}_{gz} \mathbf{R}_z^{-1} \mathbf{R}_{zg}] = \text{Tr}[\mathbf{R}_s - \mathbf{R}_s (\mathbf{R}_s + \mathbf{R}_n)^{-1} \mathbf{R}_s] \\ &= \text{Tr}[\mathbf{R}_s (\mathbf{R}_s + \mathbf{R}_n)^{-1} \mathbf{R}_n] \quad (\text{根据矩阵求逆公式}) \\ &= \text{Tr}[\mathbf{H} \mathbf{R}_n] \end{aligned}$$

#解毕

## 3.4 Kalman 滤波导言

### 3.4.1 起因

Wiener 滤波:

要求观测过程  $z(t)$  平稳, 以线性加权过去的观测数据来作估计  $\hat{g}(t) = \int_{-\infty}^t h(t-\xi) z(\xi) d\xi$ , 在使估值的均方误差为最小的准则下, 求最佳滤波器  $h(t)$ 。

缺点:

- 1) 若频域分解, 分析单个信号还行, 分析多个信号很难;
- 2) 若时域求解,

a) 存储量大: 由于须保存所有观测数据  $\{z(k)\}$ , 以计算估值  $\hat{g}(k) = \sum_{i=0}^{N-1} h(i) z(k-i)$ , 故当  $N \uparrow$  时, 计算机的存储量  $\uparrow$ 。

b) 运算量大：由于  $\mathbf{H} = \mathbf{R}_z^{-1} \mathbf{R}_{gz}$ ，每新获取一个数据就须重新计算自相关矩阵  $\mathbf{R}_z$  及其逆矩阵，故当  $N \uparrow$  时，计算机的运算量  $\uparrow\uparrow$ ，难以获得实时估值。

故，Wiener 滤波只适用于对有限观测数据进行“离线”（非实时）处理，而不适应于现代科技的要求，如：对卫星、导弹等目标进行实时的精密跟踪、控制等。

### Kalman 滤波：

六十年代初，Kalman 等人提出：不求传输函数，改用状态变量法，时域求解。

特点：

- 1) 建立状态空间模型；
- 2) 线性最小均方估计；
- 3) 采用递推算法，即：根据新的观测数据不断修正估计量；
- 4) 不要求平稳性，可用于时变系统。

优点：存储量小，具有实时性，序贯作出估计，多信号估计，同时算出估计误差。

### 3.4.2 状态空间模型

由经典物理学可知：一个动态系统的状态变化可由下述状态方程描述

$$\underset{\text{未来状态}}{\mathbf{X}_k} = \underset{\text{转移矩阵}}{\Phi_{k,k-1}} \underset{\text{当前状态}}{\mathbf{X}_{k-1}} + \underset{\text{外部干扰或由于对运动方程描述的不准确性而引入的误差}}{\mathbf{W}_{k-1}} \quad \text{其中 } k \text{ 表示离散信号 } k \text{ 时刻或连续信号 } t_k \text{ 时刻。}$$

对预测和滤波，我们需要用对系统状态的观测数据来对系统的状态进行估计，故需要建立一个观测方程（ $Z_k$  的维数  $N$  不一定与状态  $X_k$  的维数  $M$  相同）：

$$\underset{\text{观测值}}{Z_k} = \underset{\text{观测矩阵}}{H_k} \underset{\text{状态矩阵}}{X_k} + \underset{\text{观测噪声}}{V_k}$$

状态引起的观测值

目标：根据线性最小均方估计准则，对状态变量  $X_k$  作最佳估计  $\hat{X}_k$ 。

在求解最佳估计之前，先来看一看怎样建立一个随机过程的状态空间模型。

#### 1) 状态变量和状态方程：

状态变量：动态系统中一组数目最少的变量。

当已知系统的初始状态（ $t=t_0$ ）和系统输入（ $t \geq t_0$  时）时，根据状态变量可完全确定系统的状态（ $t \geq t_0$  时）。

状态方程：状态变量所满足的一阶微分方程。

例：RLC 串联电路（如右图）。

通过对外部输出电压  $y(t)$  的测量，估计内部电路的状态。

依据上面的定义，可选电压  $v(t)$  和电流  $i(t)$  作为状态变量。

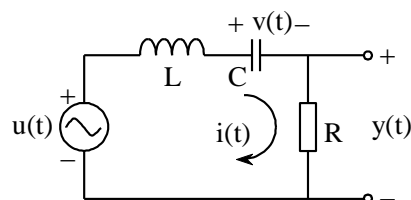
由此，我们可得到如下两个方程：

$$\text{电压方程} \quad L \frac{di}{dt} + R i(t) + v = u(t)$$

$$\text{电流方程} \quad C \frac{dv}{dt} = i(t)$$

令  $x_1=i(t)$ ,  $x_2=v(t)$ ,  $u=u(t)$ , 则上述方程可写为：

$$L \dot{x}_1 + R x_1 + x_2 = u$$



$$C\dot{x}_2 = x_1$$

用矩阵形式来表示：
$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} -\frac{R}{L} & -\frac{1}{L} \\ \frac{1}{C} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \frac{1}{L} \\ 0 \end{pmatrix} u$$
 连续状态方程

另外，根据输出电压  $y = Ri$ ，我们有：
$$y = (R, 0) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$
 输出方程

一般地，针对所观测的系统，我们可列出下面的两个方程：

$$\dot{X} = A X + B U \quad \text{状态方程}$$

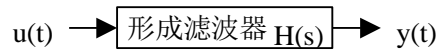
$$\begin{matrix} M \times 1 & M \times M & M \times 1 & M \times J & J \times 1 \end{matrix}$$

$$Y = C X \quad \text{输出方程}$$

$$\begin{matrix} L \times 1 & L \times M & M \times 1 \end{matrix}$$

## 2) 线性连续系统的状态模型：

**模型：**随机过程  $y(t)$  是由白噪声  $u(t)$  输入某线性系统而产生的，该系统的状态变量  $X$  是为得到  $y(t)$  而引用的中间变量。



对有理谱密度的随机过程可按下述步骤导出其状态模型：

a) 谱密度因式分解： $\Phi_y(s) = \Phi_y^+(s) \Phi_y^-(s)$ ;

b) 形成滤波器： $H(s) = \Phi_y^+(s)$ ;

c) 由  $H(s)$  推导状态方程。

**例：**已知  $\Phi_y(s) = \frac{16}{s^4 + 4s^2 + 16}$ ，试导出其状态模型及其输出方程。

**解：**按照上面的步骤，我们有：

a) 由谱密度因式分解  $\Phi_y(s) = \frac{4}{s^2 + 2s + 4} \frac{4}{s^2 - 2s + 4}$  可得：

$$\Phi_y^+(s) = \frac{4}{s^2 + 2s + 4} \quad (\text{其极点为 } s = -1 \pm \sqrt{3}i, \text{ 在左半平面})$$

$$\Phi_y^-(s) = \frac{4}{s^2 - 2s + 4} \quad (\text{其极点为 } s = 1 \pm \sqrt{3}i, \text{ 在右半平面})$$

b) 所构成的形成滤波器为：

$$H(s) = \Phi_y^+(s) = \frac{4}{s^2 + 2s + 4}$$

c) 由于  $Y(s) = H(s) U(s)$ ，将  $H(s)$  表达式代入得：

$$(s^2 + 2s + 4) Y(s) = 4 U(s)$$

作拉氏反变换，即得如下的微分方程：

$$\ddot{y} + 2\dot{y} + 4y = 4u$$

令状态变量： $x_1 = y$ ， $x_2 = \dot{x}_1$ ，则可将此方程表示为：

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -4 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 4 \end{pmatrix} u$$

此状态方程就与前面提到的其矩阵形式  $\dot{X} = AX + BU$  一样了。

其输出方程为：

$$y = (1, 0) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

亦与前面提到的其矩阵形式  $Y = CX$  一样。

#解毕

### 3) 线性连续系统状态方程的时域解：

已知： $t=t_0$  时系统的初始状态  $X(t_0)$ ， $A$ ， $B$ ， $U(t)$ 。

$$\begin{matrix} M \times 1 & M \times M & M \times J & J \times 1 \end{matrix}$$

需求解状态方程： $\dot{X} = AX + BU$ 。

定义：指数矩阵  $e^{At} = I + At + \frac{A^2 t^2}{2!} + \frac{A^3 t^3}{3!} + \dots + \frac{A^k t^k}{k!} + \dots$

则有： $\frac{d}{dt} e^{At} = A + A^2 t + \frac{A^3 t^2}{2!} + \dots + \frac{A^k t^{k-1}}{(k-1)!} + \dots = e^{At} A$

将状态方程两边乘以  $e^{-At}$  并移项得： $e^{-At} \dot{X} - e^{-At} AX = e^{-At} BU$

即  $\frac{d}{dt} (e^{-At} X) = e^{-At} BU$

对两边积分  $\int_{t_0}^t$  得： $e^{-At} X(t) - e^{-At} X(t_0) = \int_{t_0}^t e^{-A\tau} B U(\tau) d\tau$

故，我们得到状态方程的解为： $X(t) = e^{A(t-t_0)} X(t_0) + \int_{t_0}^t e^{A(t-\tau)} B U(\tau) d\tau$

它也被称为“状态转移方程”，因为它描述了系统的状态从  $t_0$  时刻到  $t$  时刻的转移过程。

令： $\Phi(t) = e^{At}$ ，

这里，它被称为“转移矩阵”，因为：若  $U(t)=0$ ，则  $X(t) = \Phi(t-t_0) X(t_0)$ 。

转移矩阵具有下面的重要特性：

a)  $\Phi(t) = \mathcal{L}^{-1}([sI - A]^{-1})$ 。

证明：由  $A \mathcal{L}(e^{At}) = A \int_0^{+\infty} e^{At} e^{-st} dt = \int_0^{+\infty} e^{-st} d e^{At} = e^{-st} e^{At} \Big|_0^{+\infty} - \int_0^{+\infty} e^{At} d e^{-st}$   
 $= -I + s \mathcal{L}(e^{At})$ ，

即  $(sI - A) \mathcal{L}(e^{At}) = I$  或  $\mathcal{L}(e^{At}) = (sI - A)^{-1}$

而  $\Phi(t) = e^{At}$ ，故有： $\Phi(t) = \mathcal{L}^{-1}([sI - A]^{-1})$ 。

b)  $\Phi(t_3-t_1) = \Phi(t_3-t_2) \Phi(t_2-t_1)$ 。

证明：由于  $X(t_3) = \Phi(t_3-t_2) X(t_2)$ ， $X(t_2) = \Phi(t_2-t_1) X(t_1)$ ，

则  $X(t_3) = \Phi(t_3-t_2) \Phi(t_2-t_1) X(t_1)$ ，而  $X(t_3) = \Phi(t_3-t_1) X(t_1)$ ，

所以， $\Phi(t_3-t_1) = \Phi(t_3-t_2) \Phi(t_2-t_1)$ 。

利用转移矩阵，状态转移方程可写为：

$$X(t) = \Phi(t-t_0) X(t_0) + \int_{t_0}^t \Phi(t-\tau) B U(\tau) d\tau \quad t \geq t_0$$

### 4) 线性系统的离散状态模型：

连续  $\rightarrow$  抽样  $\rightarrow$  离散。

连续随机过程的状态模型： $\dot{X} = AX + BU$  和  $Y = CX$

其时域解为： $X(t) = \Phi(t-t_0) X(t_0) + \int_{t_0}^t \Phi(t-\tau) B U(\tau) d\tau \quad t \geq t_0$

在离散时刻对连续过程进行抽样： $t = t_0, t_1, t_2, \dots, t_k, \dots$

使用简化记号： $X_k = X(t_k)$ ,  $\Phi_{k+1,k} = \Phi(t_{k+1}-t_k)$ ,  $W_k = \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}-\tau) B U(\tau) d\tau$ 。

那么，离散化后的状态方程，称之为“离散状态模型”，可写为：

$$X_{k+1} = \Phi_{k+1,k} X_k + W_k \quad k \geq 0 \quad \text{称为“离散状态方程”}$$

$$\begin{matrix} M \times 1 & M \times M & M \times 1 & M \times 1 \end{matrix}$$

$$Y_k = C_k X_k \quad \text{称为“输出方程”}$$

$$\begin{matrix} L \times 1 & L \times M & M \times 1 \end{matrix}$$

其中：— $\Phi_{k+1,k} = e^{A(t_{k+1}-t_k)} = e^{A\Delta t} = \mathcal{L}^{-1}([sI - A]^{-1})|_{t=\Delta t}$  均匀抽样；

— $W_k$  被假定为零均值白噪声（各分量互不相关），

其协方差矩阵为  $E(W_k W_k^T) = Q_k \delta_{ki}$ ，这里， $\delta_{ki}$  为 Kronecker- $\delta$  函数，而

$$Q_k = E(W_k W_k^T) = E\left[\left(\int_{t_k}^{t_{k+1}} \Phi(t_{k+1}-\xi) B U(\xi) d\xi\right)\left(\int_{t_k}^{t_{k+1}} \Phi(t_{k+1}-\eta) B U(\eta) d\eta\right)^T\right]$$

$$= \int_{t_k}^{t_{k+1}} \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}-\xi) B E(U(\xi) U^T(\eta)) B^T \Phi^T(t_{k+1}-\eta) d\xi d\eta$$

它的计算是比较繁杂的。

### 3.5 离散时间 Kalman 滤波

根据前面的讨论，我们有如下的离散系统模型：

$$\text{状态方程: } X_k = \Phi_{k,k-1} X_{k-1} + W_{k-1}, \quad k \geq 1$$

$$\begin{matrix} M \times 1 & M \times M & M \times 1 & M \times 1 \end{matrix}$$

$$\text{观测方程: } Z_k = H_k X_k + V_k, \quad k \geq 1$$

$$\begin{matrix} N \times 1 & N \times M & M \times 1 & N \times 1 \end{matrix}$$

其中：— $\Phi_{k,k-1}$ ，一步转移矩阵；

— $W_k$ ，动态噪声， $E(W_k)=0$ ，协方差矩阵  $\text{Cov}(W_k, W_i) = E(W_k W_i^T) = Q_k \delta_{ki}$ ；

— $V_k$ ，观测噪声， $E(V_k)=0$ ，协方差矩阵  $\text{Cov}(V_k, V_i) = E(V_k V_i^T) = R_k \delta_{ki}$ ；

$W_k$  和  $V_i$  互不相关，即  $\text{Cov}(W_k, V_i) = 0$

— $H_k$ ，观测矩阵。

已知：初始状态  $E(X_0) = \mu_0$ ,  $\text{Var}(X_0) = E((X_0 - \mu_0)(X_0 - \mu_0)^T) = P_0$ ,  $X_0$  与  $\{W_k\}$ 、 $\{V_k\}$  均不相关，即  $\text{Cov}(X_0, W_k) = 0$ ,  $\text{Cov}(X_0, V_k) = 0$ 。

须求： $X_k$  的估计  $\hat{X}_k$ ，使均方误差  $E(\tilde{X}_k^T \tilde{X}_k)$  为最小。

这里，定义  $\tilde{X}_k = \tilde{X}_{k/k}$ （取得  $z_k$  之后的估计误差） $= \hat{X}_{k/k} - X_k = \hat{X}_k - X_k$ ，

而定义  $\tilde{X}_{k/k-1} = \hat{X}_{k/k-1} - X_k$ （取得  $z_k$  之前的估计误差）。

解：为了简洁地说明问题，我们没有使用严格的数学证明方式，而是采用了直观推导方法。

假定在第  $(k-1)$  次观测之后，我们得到状态估计值  $\hat{X}_{k-1} = \hat{X}_{k-1/k-1}$ 。

一方面，我们可根据状态方程来预测第  $k$  次观测时的状态值：

因为  $E(W_{k-1})=0$ ，故我们可直观地将该状态值估计为： $\hat{X}_{k/k-1} = \Phi_{k,k-1} \hat{X}_{k-1}$ 。



另一方面，我们可根据观测方程来预测第  $k$  次的观测值 ( $Z_k$ ):

因为  $E(V_k)=0$ ，故我们可适当地推测第  $k$  次的观测值为  $H_k \hat{X}_{k/k-1}$ 。

这样，我们有理由认为：

在获得第  $k$  次实际观测数据  $Z_k$  之后来估计状态  $X_k$  的递推形式应是

$$\hat{X}_k = \hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k (Z_k - H_k \hat{X}_{k/k-1})$$

其中， $K_k$  为待定的增益矩阵，它的选取应使均方误差  $E(\tilde{X}_k^T \tilde{X}_k)$  最小。

定义：估计误差方差矩阵  $P_k = E(\tilde{X}_k \tilde{X}_k^T)$ 。

由于  $E(\tilde{X}_k^T \tilde{X}_k) = \text{Tr}(P_k)$ ，故我们可以证明：使  $P_k$  最小一定会使  $E(\tilde{X}_k^T \tilde{X}_k)$  最小（这可理解为误差向量长度最小）。因此，问题转化为：求增益矩阵  $K_k$ ，使  $P_k$  最小。

下面我们来求  $P_k$  的表达式。

$$\begin{aligned} \text{因 } \tilde{X}_k &= \hat{X}_k - X_k = \hat{X}_{k/k-1} + K_k (Z_k - H_k \hat{X}_{k/k-1}) - X_k = \hat{X}_{k/k-1} + K_k (H_k X_k + V_k - H_k \hat{X}_{k/k-1}) - X_k \\ &= (I - K_k H_k) (\hat{X}_{k/k-1} - X_k) + K_k V_k = (I - K_k H_k) \tilde{X}_{k/k-1} + K_k V_k \end{aligned}$$

$$\begin{aligned} \text{所以 } P_k &= E(\tilde{X}_k \tilde{X}_k^T) = E[(I - K_k H_k) \tilde{X}_{k/k-1} + K_k V_k] [\tilde{X}_{k/k-1}^T (I - K_k H_k)^T + V_k^T K_k^T] \\ &= (I - K_k H_k) E(\tilde{X}_{k/k-1} \tilde{X}_{k/k-1}^T) (I - K_k H_k)^T + K_k E(V_k V_k^T) K_k^T + \\ &\quad + (I - K_k H_k) E(\tilde{X}_{k/k-1} V_k^T) K_k^T + K_k E(V_k \tilde{X}_{k/k-1}^T) (I - K_k H_k)^T \end{aligned}$$

定义：一步预测估计误差方差矩阵  $P_{k/k-1} = E(\tilde{X}_{k/k-1} \tilde{X}_{k/k-1}^T)$ 。

又由于已知  $\{V_k\}$  与  $\{W_k\}$  互不相关且  $\{V_k\}$  与  $X_0$  不相关，故由状态方程可知： $V_k$  与  $X_k$  不相关，从而  $E(\tilde{X}_{k/k-1} V_k^T) = 0$ 、 $E(V_k \tilde{X}_{k/k-1}^T) = 0$ 。

$$\text{所以 } P_k = (I - K_k H_k) P_{k/k-1} (I - K_k H_k)^T + K_k R_k K_k^T。$$

下面，我们采用配平方法来求  $P_k$  的最小值：

将  $P_k$  表达式的右端展开，减、加项  $P_{k/k-1} H_k^T (H_k P_{k/k-1} H_k^T + R_k)^{-1} H_k P_{k/k-1}$ ，再把有关  $K_k$  的项归并在平方项里，即得：

$$\begin{aligned} P_k &= P_{k/k-1} - P_{k/k-1} H_k^T (H_k P_{k/k-1} H_k^T + R_k)^{-1} H_k P_{k/k-1} + \\ &\quad + [K_k - P_{k/k-1} H_k^T (H_k P_{k/k-1} H_k^T + R_k)^{-1}] (H_k P_{k/k-1} H_k^T + R_k) [K_k - P_{k/k-1} H_k^T (H_k P_{k/k-1} H_k^T + R_k)^{-1}]^T \end{aligned}$$

由于上面的  $P_k$  表达式中，前两项与  $K_k$  无关，故要使  $P_k$  最小，只要选择  $K_k$  满足下式即可：

$$K_k = P_{k/k-1} H_k^T (H_k P_{k/k-1} H_k^T + R_k)^{-1}$$

这样的增益矩阵  $K_k$  被称为“**Kalman 增益矩阵**”。

此时的估计误差方差矩阵为：

$$P_k = P_{k/k-1} - P_{k/k-1} H_k^T (H_k P_{k/k-1} H_k^T + R_k)^{-1} H_k P_{k/k-1} = P_{k/k-1} - K_k H_k P_{k/k-1} = (I - K_k H_k) P_{k/k-1}$$

这就是“**误差方差的迭代公式**”。

要完成以上的递推过程，还需要得到从  $P_{k-1}$  推出  $P_{k/k-1}$  的递推关系：

$$\text{由于 } \tilde{X}_{k/k-1} = \hat{X}_{k/k-1} - X_k = \Phi_{k,k-1} \hat{X}_{k-1} - X_k = \Phi_{k,k-1} \hat{X}_{k-1} - (\Phi_{k,k-1} X_{k-1} + W_{k-1}) = \Phi_{k,k-1} \tilde{X}_{k-1} - W_{k-1}$$

$$\text{所以 } P_{k/k-1} = E(\tilde{X}_{k/k-1} \tilde{X}_{k/k-1}^T) = E[(\Phi_{k,k-1} \tilde{X}_{k-1} - W_{k-1}) (\tilde{X}_{k-1}^T \Phi_{k,k-1}^T - W_{k-1}^T)]$$

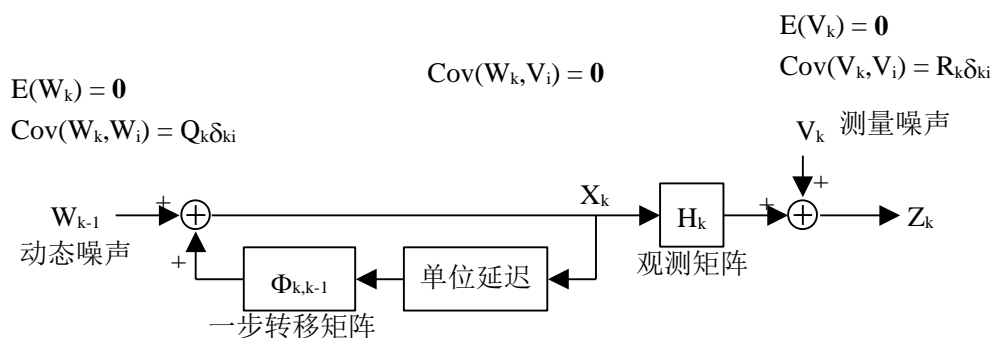
$$\text{而 } E(\tilde{X}_{k-1} W_{k-1}^T) = 0, E(W_{k-1} \tilde{X}_{k-1}^T) = 0 \quad (\text{依据与前面推导 } E(\tilde{X}_{k/k-1} V_k^T) = 0 \text{ 类似}),$$

$$\text{故, } P_{k/k-1} = \Phi_{k,k-1} P_{k-1} \Phi_{k,k-1}^T + Q_{k-1}。$$

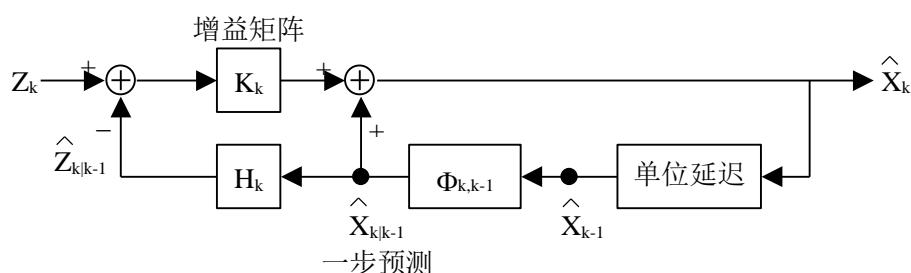
#解毕

## ◆小结：

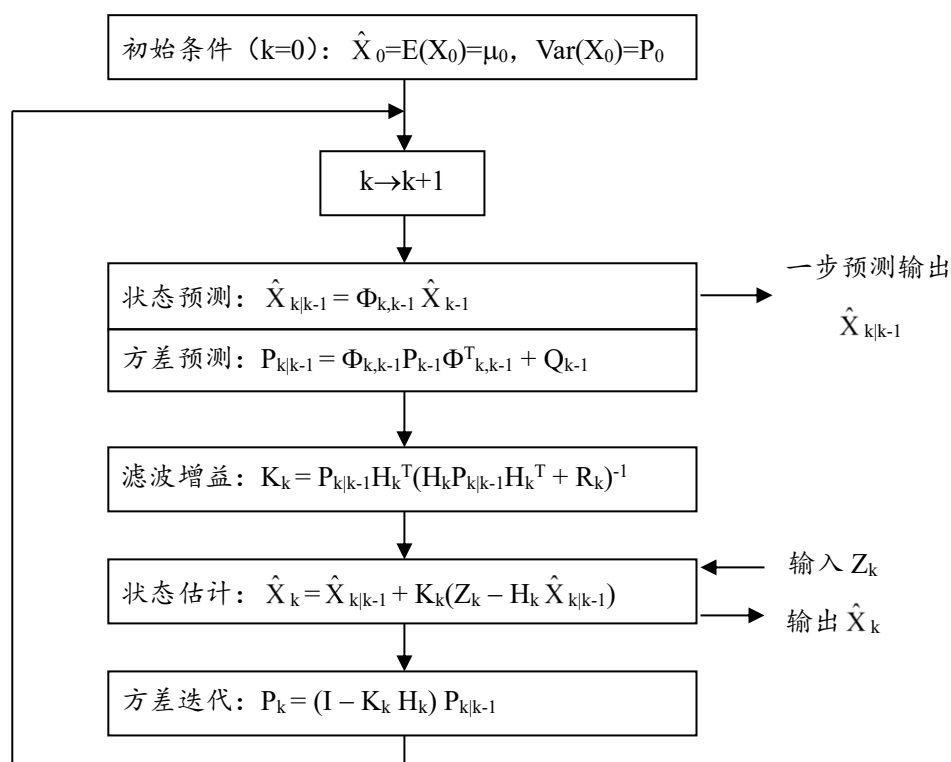
## 系统模型



## Kalman 滤波



## 递推流程



## ◆讨论:

①  $\hat{X}_k$  是  $X_k$  的线性最小均方估计，是无偏估计，即  $E(\hat{X}_k) = E(X_k)$ 。

若  $\{W_k\}$  和  $\{V_k\}$  均服从高斯分布，则  $\hat{X}_k$  是  $X_k$  的最小均方估计。

② 关于增益矩阵  $K_k$ :

a) 需计算  $(H_k P_{k|k-1} H_k^T + R_k)^{-1}$  矩阵，其维数=观测量的维数  $(N \times N) \neq$  观测值  $\{z_k\}$  的个数，即和 Wiener

滤波器中需计算的  $\mathbf{R}_z^{-1}$  (见 § 3.3.1 节中  $\mathbf{H} = \mathbf{R}_z^{-1} \mathbf{R}_{gz}$ ) 维数 (=观测值  $\{z_k\}$  的个数) 相比较小, 计算量不大。

b)  $\mathbf{K}_k$  的计算与  $\mathbf{Z}_k$  无关, 故可事先“离线”算出, 以减少实时的“在线”计算量。

c) 当初始方差  $\mathbf{P}_0$ 、动态噪声的  $\mathbf{Q}_{k-1}$ 、观测噪声的  $\mathbf{R}_k$  同乘以一个标量时,  $\mathbf{K}_k$  不变。即, 当三者同时增大时,  $\mathbf{K}_k$  不变, 但此时误差的方差增大。

d) 当  $\mathbf{R}_k \uparrow$  时,  $\mathbf{K}_k \downarrow$ 。

e) 当  $\mathbf{P}_0 \downarrow$  或  $\mathbf{Q}_{k-1} \downarrow$  时,  $\mathbf{P}_{k|k-1} \downarrow \Rightarrow \mathbf{P}_k \downarrow, \mathbf{K}_k \downarrow$ 。

综合 d)、e) 可知:  $\mathbf{K}_k \propto \mathbf{Q}_{k-1}, \mathbf{K}_k \propto \mathbf{R}_k^{-1}$ 。

#### ◆例 1: 地面雷达对空中目标的跟踪。

如果从雷达站到目标的斜距为  $x(t)$ , 其变化率  $\dot{x}(t)$  为常数。

从  $t=0$  时, 开始观测目标, 观测间隔为 1 秒, 测得  $x(t)$  的四个观测值为:

$$z(1) = 1.1\text{km}, z(2) = 2.0\text{km}, z(3) = 3.2\text{km}, z(4) = 3.8\text{km}。$$

已知:  $E(x(0))=0, E(\dot{x}(0))=0, E(x(0)\dot{x}(0))=0, E(x^2(0))=10 \text{ km}^2, E(\dot{x}^2(0))=10 (\text{km/s})^2,$

$$E(v(k))=0, R(k)=0.1 \text{ km}^2。$$

利用 Kalman 滤波方法求斜距  $x(t)$  的最佳滤波值与滤波误差方差。

解:

①建立系统的状态方程:

设目标位移为  $x_1(t)$ , 其速度为  $x_2(t)$ ,

则  $\dot{x}_1(t) = x_2(t), \dot{x}_2(t) = \ddot{x}_1(t) = 0$ 。

从而可得系统的状态方程为:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{即} \quad \dot{\mathbf{X}} = \mathbf{A} \mathbf{X}$$

②求解状态方程以得到转移矩阵:

对应于 § 3.4.2 中 3) 的模型  $\dot{\mathbf{X}} = \mathbf{A} \mathbf{X} + \mathbf{B} \mathbf{U}$ ,

本例中:  $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \mathbf{B} = \mathbf{0}$ 。

故得其时域解为:  $\mathbf{X}(t) = \Phi(t-t_0) \mathbf{X}(t_0)$

这里,  $\Phi(t) = e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \dots = \mathbf{I} + \mathbf{A}t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$

所以, 系统的转移矩阵为:  $\Phi(t-t_0) = \begin{pmatrix} 1 & t-t_0 \\ 0 & 1 \end{pmatrix}$

③给出离散系统模型:

状态方程:  $\mathbf{X}_k = \Phi_{k,k-1} \mathbf{X}_{k-1} + \mathbf{W}_{k-1}$

其中  $\mathbf{X}_k = \begin{pmatrix} x_1(k\Delta t) \\ x_2(k\Delta t) \end{pmatrix} = \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix}, \Phi_{k,k-1} = \Phi(k\Delta t, (k-1)\Delta t) = \Phi(\Delta t) = \Phi(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$

$$\mathbf{W}_{k-1} = \mathbf{0}$$

观测方程:  $\mathbf{Z}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{V}_k$  (直接测量值+观测噪声)

其中  $\mathbf{H}_k = \mathbf{H} = (1, 0)$

④确定系统的初始条件:

$$\hat{\mathbf{X}}_0 = E(\mathbf{X}(0)) = \begin{pmatrix} E(x_1(0)) \\ E(x_2(0)) \end{pmatrix} = \begin{pmatrix} E(\dot{x}(0)) \\ E(\ddot{x}(0)) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mathbf{0}$$

$$\mathbf{P}_0 = \text{Var}(\mathbf{X}_0) = \underset{\text{均值为零}}{E \left( \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} \begin{pmatrix} x_1(0) & x_2(0) \end{pmatrix} \right)} = \begin{pmatrix} E(x^2(0)) & E(x(0)\dot{x}(0)) \\ E(x(0)\dot{x}(0)) & E(\dot{x}^2(0)) \end{pmatrix} = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$$

⑤递推计算估值:

(注意: 已知  $\mathbf{Q}_k=0$ ,  $\mathbf{R}_k=0.1 \text{ km}^2$ )

$$\text{当 } k=1 \text{ 时, } \hat{\mathbf{X}}_{1|0} = \Phi_{1,0} \hat{\mathbf{X}}_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \mathbf{0} = \mathbf{0}$$

$$\mathbf{P}_{1|0} = \Phi_{1,0} \mathbf{P}_0 \Phi_{1,0}^T + \mathbf{Q}_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} + \mathbf{0} = \begin{pmatrix} 20 & 10 \\ 10 & 10 \end{pmatrix}$$

$$\begin{aligned} \mathbf{K}_1 &= \mathbf{P}_{1|0} \mathbf{H}^T (\mathbf{H} \mathbf{P}_{1|0} \mathbf{H}^T + \mathbf{R}_1)^{-1} \\ &= \begin{pmatrix} 20 & 10 \\ 10 & 10 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left( \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 20 & 10 \\ 10 & 10 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 0.1 \right)^{-1} = \frac{1}{20.1} \begin{pmatrix} 20 \\ 10 \end{pmatrix} \approx \begin{pmatrix} 0.9950 \\ 0.4975 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \hat{\mathbf{X}}_1 &= \hat{\mathbf{X}}_{1|0} + \mathbf{K}_1 (\mathbf{Z}_1 - \mathbf{H} \hat{\mathbf{X}}_{1|0}) \\ &= \mathbf{0} + \frac{1}{20.1} \begin{pmatrix} 20 \\ 10 \end{pmatrix} (1.1 - \begin{pmatrix} 1 & 0 \end{pmatrix} \mathbf{0}) = \frac{1}{20.1} \begin{pmatrix} 22 \\ 11 \end{pmatrix} \approx \begin{pmatrix} 1.0945 \\ 0.5473 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \mathbf{P}_1 &= (\mathbf{I} - \mathbf{K}_1 \mathbf{H}) \mathbf{P}_{1|0} \\ &= \left( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{1}{20.1} \begin{pmatrix} 20 \\ 10 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \right) \begin{pmatrix} 20 & 10 \\ 10 & 10 \end{pmatrix} = \frac{1}{20.1} \begin{pmatrix} 2 & 1 \\ 1 & 101 \end{pmatrix} \approx \begin{pmatrix} 0.09950 & 0.04975 \\ 0.04975 & 5.0249 \end{pmatrix} \end{aligned}$$

$$\text{当 } k=2 \text{ 时, } \hat{\mathbf{X}}_{2|1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \hat{\mathbf{X}}_1 = \frac{1}{20.1} \begin{pmatrix} 33 \\ 11 \end{pmatrix} \approx \begin{pmatrix} 1.6418 \\ 0.54726 \end{pmatrix}$$

$$\mathbf{P}_{2|1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \mathbf{P}_1 \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \frac{1}{20.1} \begin{pmatrix} 105 & 102 \\ 102 & 101 \end{pmatrix} \approx \begin{pmatrix} 5.2239 & 5.0746 \\ 5.0746 & 5.0249 \end{pmatrix}$$

$$\mathbf{K}_2 = \mathbf{P}_{2|1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left( \begin{pmatrix} 1 & 0 \end{pmatrix} \mathbf{P}_{2|1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 0.1 \right)^{-1} = \frac{1}{107.01} \begin{pmatrix} 105 \\ 102 \end{pmatrix} \approx \begin{pmatrix} 0.9812 \\ 0.9532 \end{pmatrix}$$

$$\hat{\mathbf{X}}_2 = \hat{\mathbf{X}}_{2|1} + \mathbf{K}_2 (\mathbf{Z}_2 - \begin{pmatrix} 1 & 0 \end{pmatrix} \hat{\mathbf{X}}_{2|1}) = \frac{1}{2150.901} \begin{pmatrix} 4287.33 \\ 1911.51 \end{pmatrix} \approx \begin{pmatrix} 1.9933 \\ 0.8887 \end{pmatrix}$$

$$\mathbf{P}_2 = \left( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \mathbf{K}_2 \begin{pmatrix} 1 & 0 \end{pmatrix} \right) \mathbf{P}_{2|1} = \frac{1}{2150.901} \begin{pmatrix} 211.05 & 205.02 \\ 205.02 & 404.01 \end{pmatrix} \approx \begin{pmatrix} 0.09812 & 0.09532 \\ 0.09532 & 0.1878 \end{pmatrix}$$

$$\text{当 } k=3 \text{ 时, } \hat{\mathbf{X}}_{3|2} = \frac{1}{2150.901} \begin{pmatrix} 6198.84 \\ 1911.51 \end{pmatrix} \approx \begin{pmatrix} 2.8820 \\ 0.8887 \end{pmatrix}$$

$$\mathbf{P}_{3|2} = \frac{1}{2150.901} \begin{pmatrix} 1025.1 & 609.03 \\ 609.03 & 404.01 \end{pmatrix} \approx \begin{pmatrix} 0.47659 & 0.28315 \\ 0.28315 & 0.18783 \end{pmatrix}$$

$$\mathbf{K}_3 = \frac{1}{1240.1901} \begin{pmatrix} 1025.1 \\ 609.03 \end{pmatrix} \approx \begin{pmatrix} 0.82657 \\ 0.49108 \end{pmatrix}$$

$$\hat{\mathbf{X}}_3 \approx \frac{1}{2150.901 \times 1240.1901} \begin{pmatrix} 8388952.7 \\ 2787238.6 \end{pmatrix} \approx \begin{pmatrix} 3.14484 \\ 1.04488 \end{pmatrix}$$

$$P_3 \approx \frac{1}{2150.901 \times 1240.1901} \begin{pmatrix} 220488.9 & 130996.3 \\ 130996.3 & 130131.7 \end{pmatrix} \approx \begin{pmatrix} 0.08266 & 0.04911 \\ 0.04911 & 0.04878 \end{pmatrix}$$

$$\text{当 } k=4 \text{ 时, } \hat{X}_{4|3} \approx \begin{pmatrix} 4.18972 \\ 1.04488 \end{pmatrix}$$

$$P_{4|3} \approx \begin{pmatrix} 0.22966 & 0.09789 \\ 0.09789 & 0.04878 \end{pmatrix}$$

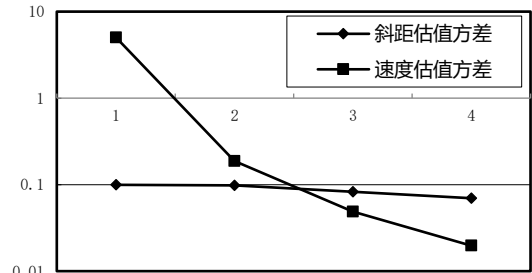
$$K_4 \approx \begin{pmatrix} 0.69666 \\ 0.29694 \end{pmatrix}$$

$$\hat{X}_4 \approx \begin{pmatrix} 3.9182 \\ 0.9292 \end{pmatrix}$$

$$P_4 \approx \begin{pmatrix} 0.06967 & 0.02969 \\ 0.02969 & 0.01971 \end{pmatrix}$$

根据以上数据，我们可将估值方差示于右图中。由图可见：随着  $k \uparrow$ ，估值精度  $\uparrow$ 。

#例毕



## ◆例 2：电话负载预测。

一电话公司在过去几年中其用户的实际电话负载量如右图所示，希望据此预测以后的电话负载量。

### ①建立连续状态方程：

根据以前的数据推测系统的状态模型是由“线性增长+正弦振荡+噪声”构成的，即

$$\text{线性} \quad \ddot{x} = n_1(t)$$

$$\text{振荡} \quad \ddot{y} + \omega^2 y = n_2(t)$$

其中  $n_1(t)$  和  $n_2(t)$  为相互统计独立的白噪声。

令状态矢量为  $X = (x_1, x_2, x_3, x_4)^T = (x, \dot{x}, y, \dot{y})^T$

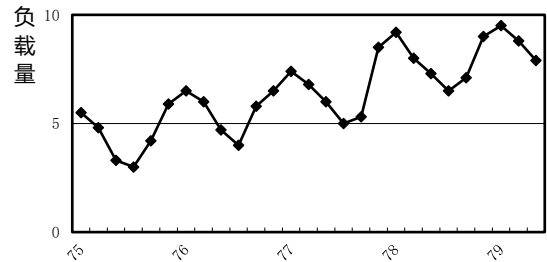
则相应的连续系统的状态方程可写为：

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\omega^2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} 0 \\ n_1(t) \\ 0 \\ n_2(t) \end{pmatrix} \quad \text{或 其矩阵形式} \quad \dot{X} = A X + U$$

### ②求转移矩阵：

$$\Phi_{k,k-1} = \mathcal{L}^{-1}[(sI - A)^{-1}] \Big|_{t=T} = \mathcal{L}^{-1} \left( \begin{pmatrix} s & -1 & 0 & 0 \\ 0 & s & 0 & 0 \\ 0 & 0 & s & -1 \\ 0 & 0 & \omega^2 & s \end{pmatrix}^{-1} \right) \Big|_{t=T}$$

$$= \mathcal{L}^{-1} \left( \begin{pmatrix} \frac{1}{s} & \frac{1}{s^2} & 0 & 0 \\ 0 & \frac{1}{s} & 0 & 0 \\ 0 & 0 & \frac{s}{s^2 + \omega^2} & \frac{1}{s^2 + \omega^2} \\ 0 & 0 & \frac{-\omega^2}{s^2 + \omega^2} & \frac{s}{s^2 + \omega^2} \end{pmatrix} \right) \Big|_{t=T} = \begin{pmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \omega T & \frac{1}{\omega^2} \sin \omega T \\ 0 & 0 & -\sin \omega T & \cos \omega T \end{pmatrix}$$



③给出离散系统模型：

$$\text{状态方程: } \mathbf{X}_k = \Phi_{k,k-1} \mathbf{X}_{k-1} + \mathbf{W}_{k-1}$$

$$\begin{matrix} 4 \times 1 & 4 \times 4 & 4 \times 1 & 4 \times 1 \end{matrix}$$

$$\text{观测方程: } z_k = \mathbf{H}_k \mathbf{X}_k + v_k$$

$$\begin{matrix} 1 \times 1 & 1 \times 4 & 4 \times 1 & 1 \times 1 \end{matrix}$$

$$\text{其中: } \mathbf{H}_k = \mathbf{H} = (1, 0, 1, 0)$$

$$\mathbf{Q}_k = E(\mathbf{W}_k \mathbf{W}_k^T) \quad (\text{需凭经验确定})$$

$$\mathbf{R}_k = E(v_k^2) \quad (\text{可推测为一小常量})$$

④确定系统的初始条件：

可利用过去的 16 个数据来启动滤波器，即初始条件为： $\hat{\mathbf{X}}_{16}$  及其误差协方差阵  $\mathbf{P}_{16}$ 。

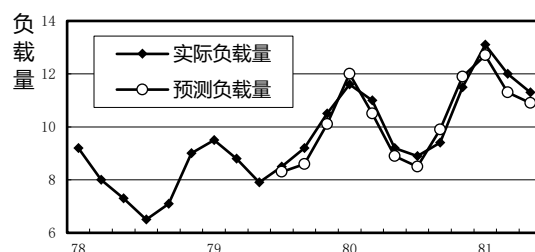
⑤递推计算（从  $k=17$  开始）：

$$\text{一步预测负载量: } \hat{\mathbf{X}}_{k|k-1} = \Phi_{k,k-1} \hat{\mathbf{X}}_{k-1}$$

一步预测方差：

$$\mathbf{P}_{k|k-1} = \Phi_{k,k-1} \mathbf{P}_{k-1} \Phi_{k,k-1}^T + \mathbf{Q}_{k-1}$$

实验结果证明：利用 Kalman 递推公式计算的预测值比较符合实际情况。



#例毕

◆Kalman 滤波的第二组公式：

在前面，我们导出的 Kalman 滤波公式称为“第一组公式”：

$$\mathbf{K}_k = \mathbf{P}_{k/k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k/k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (\text{a})$$

$$\mathbf{P}_{k/k-1} = \Phi_{k,k-1} \mathbf{P}_{k-1} \Phi_{k,k-1}^T + \mathbf{Q}_{k-1} \quad (\text{b})$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k/k-1} \quad (\text{c})$$

现在，我们要推导出 Kalman 滤波的第二组公式：

$$\mathbf{K}_k = \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \quad (\text{d})$$

$$\mathbf{P}_{k/k-1}^{-1} = (\Phi_{k,k-1} \mathbf{P}_{k-1} \Phi_{k,k-1}^T + \mathbf{Q}_{k-1})^{-1} \quad (\text{e})$$

$$\mathbf{P}_k^{-1} = \mathbf{P}_{k/k-1}^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \quad (\text{f})$$

证明：将矩阵求逆的公式  $(\mathbf{A} + \mathbf{BC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{I} + \mathbf{CA}^{-1} \mathbf{B})^{-1} \mathbf{CA}^{-1}$  应用于(a)中可得：

$$\begin{aligned} \mathbf{K}_k &= \mathbf{P}_k \mathbf{H}_k^T [\mathbf{R}_k^{-1} - \mathbf{R}_k^{-1} \mathbf{H}_k (\mathbf{I} + \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1} \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1}] = [\mathbf{I} - \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k (\mathbf{I} + \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1}] \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \\ &= [(\mathbf{I} + \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k) - \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k] (\mathbf{I} + \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1} \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \\ &= (\mathbf{I} + \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1} \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} = [\mathbf{P}_k^{-1} (\mathbf{I} + \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)]^{-1} \mathbf{H}_k^T \mathbf{R}_k^{-1} \\ &= [\mathbf{P}_k^{-1} (\mathbf{I} + \mathbf{P}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)]^{-1} \mathbf{H}_k^T \mathbf{R}_k^{-1} = (\mathbf{P}_{k/k-1}^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{R}_k^{-1} \end{aligned}$$

从而，我们又有：

$$\begin{aligned} \mathbf{I} - \mathbf{K}_k \mathbf{H}_k &= \mathbf{I} - (\mathbf{P}_{k/k-1}^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \\ &= (\mathbf{P}_{k/k-1}^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1} [(\mathbf{P}_{k/k-1}^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k) - \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k] \\ &= (\mathbf{P}_{k/k-1}^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1} \mathbf{P}_{k/k-1}^{-1} \end{aligned}$$

$$\text{所以，由(c)式可得: } \mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k/k-1} = (\mathbf{P}_{k/k-1}^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1}$$

因此，将上式代入前面  $\mathbf{K}_k$  的表达式即可证明(d)式；显然(b)式与(e)式等效；而对上式的两端矩阵求逆即可证明(f)式。

讨论:

- ① 由于第二组公式要求两个矩阵  $P_k$  及  $P_{k/k-1}$  的逆矩阵, 这将大大地增加计算时间, 因此, 在通常的递推计算中我们采用第一组公式;
- ② 在缺乏初始条件的先验知识时, 常先使用第二组公式, 而后使用第一组公式。因为此时初始的误差方差矩阵  $P_0=(1/\varepsilon)I$ , 其中  $\varepsilon$  为一小正数, 而由(b)式对  $P_{1/0}$  也只能取  $(1/\varepsilon)I$ , 因而由(a)和(c)两式无法直接计算  $K_k$  和  $P_k$ 。这是因为当  $P_0 \rightarrow \infty$  时, 第一组公式不便于计算。而使用第二组公式, 就没有这样的问题。
- ③ 在开展理论研究及研究观测误差对估值性能的影响时, 第二组公式有利于排除其它参数的影响。

## 3.6 连续时间 Kalman 滤波

连续时变系统模型:

$$\begin{array}{lcl} \dot{X}(t) = & A(t) X(t) + B(t) U(t) & \text{连续状态方程} \\ \text{Mx1} & \text{系统矩阵 MxM} \quad \text{状态变量 Mx1} \quad \text{系统噪声矩阵 MxJ} \quad \text{动态噪声变量 Jx1} & \end{array}$$

$$\begin{array}{lcl} Z(t) = & H(t) X(t) + V(t) & \text{观测方程} \\ \text{观测变量 Nx1} & \text{观测矩阵 NxM} \quad \text{Mx1} \quad \text{观测噪声变量 Nx1} & \end{array}$$

$$E(U(t))=0, E(U(t)U^T(\tau))=Q(t)\delta(t-\tau), E(V(t))=0, E(V(t)V^T(\tau))=R(t)\delta(t-\tau), E(U(t)V^T(\tau))=0, \\ JxJ \text{ (非负定)} \quad NxN \text{ (正定)}$$

已知: 初始状态  $X(t_0)$ ,  $E(X(t_0))=\mu_X(t_0)$  (一般=0),  $\text{Var}(X(t_0))=P(t_0)$ ,

且  $t \geq t_0$  时  $X(t_0)$  与  $U(t)$ 、 $V(t)$  均不相关, 即:  $\text{Cov}(X(t_0), U(t))=0$ ,  $\text{Cov}(X(t_0), V(t))=0$ 。

问题: 给定观测  $Z(t)$  ( $t \geq t_0$ ) 后, 要求状态  $X(t)$  的滤波估计  $\hat{X}(t)$ , 使均方误差方差矩阵  $P_{t/t}=P_t=E(\tilde{X}(t)\tilde{X}^T(t))$  为最小的线性估计, 这里  $\tilde{X}(t)=\hat{X}(t)-X(t)$ 。

推导: 采用方法: 离散系统的结果  $\Rightarrow$  采样间隔  $\Delta t \rightarrow 0 \Rightarrow$  连续系统的结果。

- ① 寻找对应的离散化方程:

首先, 使连续状态方程的解离散化。

前面我们已得连续状态方程的解为:

$$X(t) = \Phi(t, t_0) X(t_0) + \int_{t_0}^t \Phi(t, \tau) B(\tau) U(\tau) d\tau \quad t \geq t_0$$

这里, 我们稍加了一点推广以适应时变系统的情况: 用  $\Phi(t, t_0)$  代替了  $\Phi(t-t_0)$ , 而用  $B(\tau)$  代替了  $B$ 。

令  $t=k \cdot \Delta t$ ,  $t_0=(k-1)\Delta t$ , 则有:

$$X(k \cdot \Delta t) = \Phi(k \cdot \Delta t, (k-1)\Delta t) X((k-1)\Delta t) + \int_{(k-1)\Delta t}^{k \cdot \Delta t} \Phi(k \cdot \Delta t, \tau) B(\tau) U(\tau) d\tau$$

即  $X_k = \Phi_{k,k-1} X_{k-1} + W_{k-1}$

其中:  $W_{k-1} = \int_{(k-1)\Delta t}^{k \cdot \Delta t} \Phi(k \cdot \Delta t, \tau) B(\tau) U(\tau) d\tau$

$$\approx \Phi(k \cdot \Delta t, (k-1)\Delta t) B((k-1)\Delta t) \int_{(k-1)\Delta t}^{k \cdot \Delta t} U(\tau) d\tau$$

$E(W_k) = 0$  (因为  $E(U(t)) = 0$ )

$$\begin{aligned} E(W_k W_k^T) &\approx \Phi(t+\Delta t, t) B(t) \int_{k \cdot \Delta t}^{(k+1)\Delta t} \int_{i \cdot \Delta t}^{(i+1)\Delta t} E(U(\xi) U^T(\eta)) d\xi d\eta B^T(t) \Phi^T(t+\Delta t, t) \\ &= \Phi(t+\Delta t, t) B(t) \int_{k \cdot \Delta t}^{(k+1)\Delta t} \int_{i \cdot \Delta t}^{(i+1)\Delta t} Q(\xi) \delta(\xi-\eta) d\xi d\eta B^T(t) \Phi^T(t+\Delta t, t) \end{aligned}$$

$$\begin{aligned}
&= \Phi(t+\Delta t, t) B(t) \int_{k \cdot \Delta t}^{(k+1)\Delta t} Q(\xi) d\xi B^T(t) \Phi^T(t+\Delta t, t) \bullet \delta_{ki} \\
&= \Phi(t+\Delta t, t) B(t) Q(t) \Delta t B^T(t) \Phi^T(t+\Delta t, t) \bullet \delta_{ki} \\
&= Q_k \bullet \delta_{ki}
\end{aligned}$$

即  $Q_k = \Phi(t+\Delta t, t) B(t) Q(t) B^T(t) \Phi^T(t+\Delta t, t) \Delta t$

其次，使观测方程离散化。

由于我们不能采取直接抽样的方法来使观测方程  $Z(t) = H(t)X(t) + V(t)$  离散化成  $Z_k = H_k X_k + V_k$  (因为  $V(t)$  为零均值白噪声,  $R_v(\tau) = (N_0/2)\delta(\tau)$ , 则  $V_k$  的方差将为  $R_v(0) \rightarrow \infty$ ), 所以我们转而采取在时间间隔  $\Delta t$  上取平均, 即:

$$Z_k = H_k X_k + V_k$$

$$\text{其中: } Z_k = \frac{1}{\Delta t} \int_{k \cdot \Delta t}^{(k+1)\Delta t} z(\tau) d\tau \approx \frac{1}{\Delta t} z(k \cdot \Delta t) \cdot \Delta t = z(t), \quad H_k = H(t), \quad X_k = X(t),$$

$$V_k = \frac{1}{\Delta t} \int_{k \cdot \Delta t}^{(k+1)\Delta t} V(\tau) d\tau \quad \text{且} \quad E(V_k) = \mathbf{0}$$

$$\begin{aligned}
E(V_k V_i^T) &= \frac{1}{(\Delta t)^2} \int_{k \cdot \Delta t}^{(k+1)\Delta t} \int_{i \cdot \Delta t}^{(i+1)\Delta t} R(\xi) \delta(\xi - \eta) d\xi d\eta \\
&\approx \frac{1}{(\Delta t)^2} R(k \cdot \Delta t) \Delta t \delta_{ki} = \frac{R(t)}{\Delta t} \delta_{ki} = R_k \delta_{ki}
\end{aligned}$$

$$\text{即 } R_k = \frac{R(t)}{\Delta t}$$

② 写出相应的离散 Kalman 滤波公式 ( $t = k \cdot \Delta t$ ):

$$\hat{X}(t+\Delta t) = \hat{X}(t+\Delta t | t) + K(t+\Delta t) [Z(t+\Delta t) - H(t+\Delta t) \hat{X}(t+\Delta t | t)]$$

$$\text{其中: } \hat{X}(t+\Delta t | t) = \Phi(t+\Delta t, t) \hat{X}(t)$$

$$\hat{X}(t_0) = E(X(t_0))$$

$$K(t+\Delta t) = P(t+\Delta t) H^T(t+\Delta t) [R_{k+1}]^{-1} = P(t+\Delta t) H^T(t+\Delta t) R^{-1}(t+\Delta t) \Delta t \quad (\text{第二组})$$

$$P(t+\Delta t) = [I - K(t+\Delta t) H(t+\Delta t)] P(t+\Delta t | t) \quad (\text{第一组})$$

$$\text{其中: } P(t+\Delta t | t) = \Phi(t+\Delta t, t) P(t) \Phi^T(t+\Delta t, t) + Q_k$$

$$P(t_0) = \text{Var}(X(t_0))$$

③ 取极限, 导出连续 Kalman 滤波公式:

$$\text{因为: } \Phi(t+\Delta t, t) = \exp(A(t+\Delta t - t)) = \exp(A \Delta t) = I + A(t) \Delta t + \mathbf{O}((\Delta t)^2)$$

$$\text{所以: } \hat{X}(t) = \lim_{\Delta t \rightarrow 0} \frac{\hat{X}(t+\Delta t) - \hat{X}(t)}{\Delta t}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \{ [\Phi(t+\Delta t, t) - I] \hat{X}(t) + K(t+\Delta t) [Z(t+\Delta t) - H(t+\Delta t) \Phi(t+\Delta t, t) \hat{X}(t)] \}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \{ [A(t) \Delta t + \mathbf{O}((\Delta t)^2)] \hat{X}(t) + K(t+\Delta t) [Z(t+\Delta t) - H(t+\Delta t) \Phi(t+\Delta t, t) \hat{X}(t)] \}$$

$$= A(t) \hat{X}(t) + \lim_{\Delta t \rightarrow 0} \frac{K(t+\Delta t)}{\Delta t} \lim_{\Delta t \rightarrow 0} [Z(t+\Delta t) - H(t+\Delta t) [I + A(t) \Delta t + \mathbf{O}((\Delta t)^2)] \hat{X}(t)]$$

$$= A(t) \hat{X}(t) + \lim_{\Delta t \rightarrow 0} \frac{K(t+\Delta t)}{\Delta t} [Z(t) - H(t) \hat{X}(t)]$$



$$\text{令 } K(t) = \lim_{\Delta t \rightarrow 0} \frac{K(t+\Delta t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} [P(t+\Delta t) H^T(t+\Delta t) R^{-1}(t+\Delta t) \Delta t]$$

即：  $K(t) = P(t) H^T(t) R^{-1}(t)$  滤波增益方程

则得：  $\hat{X}(t) = A(t) \hat{X}(t) + K(t) [Z(t) - H(t) \hat{X}(t)]$  状态估计方程

下面还要求误差矩阵的表达式：

$$\begin{aligned} \text{因： } \Phi(t+\Delta t, t) P(t) \Phi^T(t+\Delta t, t) &= [I + A(t) \Delta t + O((\Delta t)^2)] P(t) [I + A^T(t) \Delta t + O((\Delta t)^2)]^T \\ &= P(t) + A(t) P(t) \Delta t + P(t) A^T(t) \Delta t + O((\Delta t)^2) \end{aligned}$$

所以，将  $P(t+\Delta t | t)$  的表达式代入  $P(t+\Delta t)$  的计算式中得：

$$\begin{aligned} P(t+\Delta t) &= [I - K(t+\Delta t) H(t+\Delta t)] [\Phi(t+\Delta t, t) P(t) \Phi^T(t+\Delta t, t) + Q_k] \\ &= [I - K(t+\Delta t) H(t+\Delta t)] [P(t) + A(t) P(t) \Delta t + P(t) A^T(t) \Delta t + O((\Delta t)^2) + Q_k] \\ &= P(t) + A(t) P(t) \Delta t + P(t) A^T(t) \Delta t - K(t+\Delta t) H(t+\Delta t) P(t) + Q_k + O((\Delta t)^2) \end{aligned}$$

$$\text{又： } \lim_{\Delta t \rightarrow 0} \frac{K(t+\Delta t)}{\Delta t} = K(t) = P(t) H^T(t) R^{-1}(t)$$

$$\begin{aligned} \text{且 } \lim_{\Delta t \rightarrow 0} \frac{Q_k}{\Delta t} &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} [\Phi(t+\Delta t, t) B(t) Q(t) B^T(t) \Phi^T(t+\Delta t, t) \Delta t] \\ &= \lim_{\Delta t \rightarrow 0} \Phi(t+\Delta t, t) \bullet B(t) Q(t) B^T(t) \bullet \lim_{\Delta t \rightarrow 0} \Phi^T(t+\Delta t, t) \\ &= \lim_{\Delta t \rightarrow 0} [I + O(\Delta t)] \bullet B(t) Q(t) B^T(t) \bullet \lim_{\Delta t \rightarrow 0} [I + O(\Delta t)] = B(t) Q(t) B^T(t) \end{aligned}$$

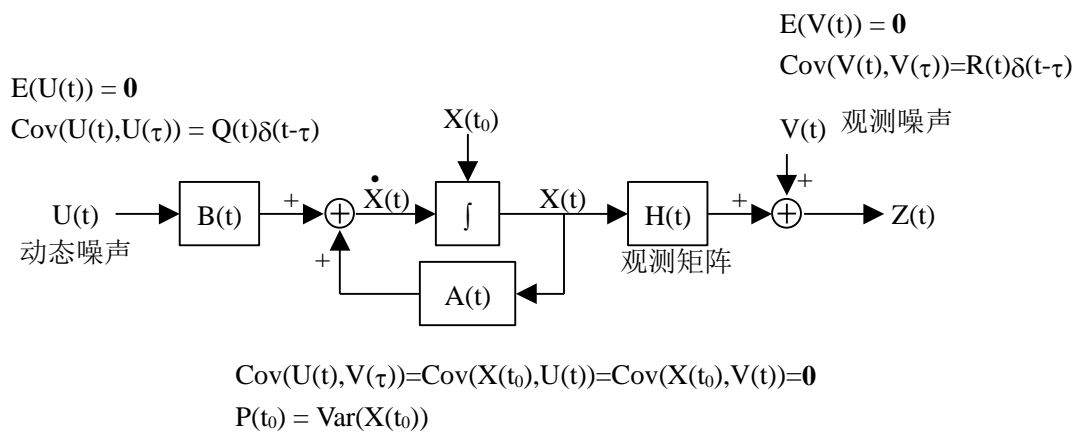
而  $\dot{P}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t+\Delta t) - P(t)}{\Delta t}$ ，故可得：

$$\dot{P}(t) = A(t) P(t) + P(t) A^T(t) - P(t) H^T(t) R^{-1}(t) H(t) P(t) + B(t) Q(t) B^T(t)$$

这就是误差方差方程，也称为“Ricatti 方程”。

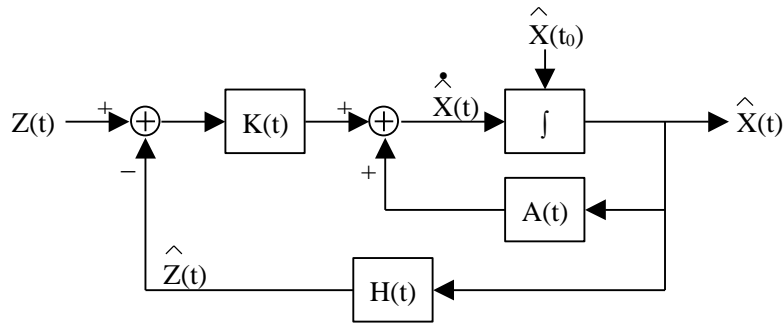
#### ◆小结：

连续系统模型：



(注：框图为乘法而非卷积关系)

连续 Kalman 滤波：



Kalman 滤波算法:

- ① 求解 Ricatti 方程:  $\dot{P}(t) = A(t)P(t) + P(t)A^T(t) - P(t)H^T(t)R^{-1}(t)H(t)P(t) + B(t)Q(t)B^T(t)$
- ② 计算滤波增益:  $K(t) = P(t)H^T(t)R^{-1}(t)$
- ③ 状态估计:  $\dot{\hat{X}}(t) = A(t)\hat{X}(t) + K(t)[Z(t) - H(t)\hat{X}(t)]$

#### ◆Ricatti 方程的求解:

我们可将该方程简记为:  $\dot{P} = AP + PA^T - PH^T R^{-1} HP + BQB^T$

初始条件为:  $P(0) = P_0$ 。

对于常参量情况, 此一阶非线性微分方程可转化为两个联立线性方程, 从而求得其解析解。

转化过程:

设  $P$  可写为:  $P = FG^{-1}$ ,  $G(0) = I$ 。

则  $PG = F$ , 两边对  $t$  求导得:  $\dot{P}G + P\dot{G} = \dot{F}$ 。

代入 Ricatti 方程得:  $(AP + PA^T - PH^T R^{-1} HP + BQB^T)G + P\dot{G} = \dot{F}$

利用  $PG=F$  并整理  $P$  项得:  $P(A^T G - H^T R^{-1} HF + \dot{G}) + (AF + BQB^T G - \dot{F}) = 0$

若令两括号中的项为零, 则上述方程成立。

此时, 我们可得如下的线性微分方程组:

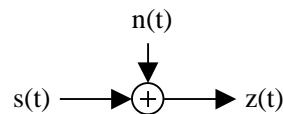
$$\begin{cases} \dot{F} = AF + BQB^T G \\ \dot{G} = H^T R^{-1} HF - A^T G \end{cases}$$

初始条件为:  $F(0) = P(0) = P_0$ ,  $G(0) = I$ 。

#### ◆例: 在加性噪声信道中传输信号问题。

已知: 信号  $s(t)$  的自相关函数为  $R_s(\tau) = e^{-|\tau|} \xleftrightarrow{\mathcal{L}} \frac{2}{1-s^2}$ ,

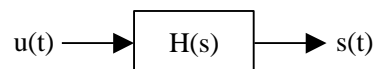
噪声  $n(t)$  的自相关函数为  $R_n(\tau) = \delta(\tau) \xleftrightarrow{\mathcal{L}} 1$ 。



试用 Kalman 滤波方法估计  $s(t)$ 。

解: 由  $s(t)$  的功率谱密度可得形成滤波器:

$$H(s) = \frac{\sqrt{2}}{s+1} \quad (\text{参见 § 3.4.2 节}).$$



令状态变量  $x(t) = s(t)$ , 噪声  $u(t)$  为单位白噪声,

则状态方程可写为:  $\dot{x}(t) = -x(t) + \sqrt{2} u(t)$

由题意, 观测方程为:  $z(t) = x(t) + v(t)$

这里,  $v(t) = n(t)$ 。

因此, 与连续系统模型相对应的参数为:

$$A = -1, \quad B = \sqrt{2}, \quad H = 1,$$

$$Q = 1 \quad (\text{因为 } E[u(t)u^T(\tau)] = \delta(t-\tau))$$

$$R = 1 \quad (\text{因为 } E[v(t)v^T(\tau)] = \delta(t-\tau))$$

根据转化方式求解 Riccati 方程，我们可写出如下的方程组：

$$\begin{cases} \dot{F} = -F + 2G \\ \dot{G} = F + G \end{cases}$$

$$\text{初始条件: } F(0) = P_0 = E[x^2(0)] = R_s(0) = 1, \quad G(0) = 1。$$

两边作拉氏变换：

$$\begin{cases} sF(s) - F(0) = -F(s) + 2G(s) \\ sG(s) - G(0) = F(s) + G(s) \end{cases}$$

解此方程组，得：

$$\begin{cases} F(s) = P_0 \frac{s}{s^2 - 3} + (2 - P_0) \frac{1}{s^2 - 3} \\ G(s) = \frac{s}{s^2 - 3} + (1 + P_0) \frac{1}{s^2 - 3} \end{cases}$$

$$\text{利用拉氏变换公式: } \text{ch}(\alpha t) = \frac{1}{2}(e^{\alpha t} + e^{-\alpha t}) \xleftrightarrow{\mathcal{L}} \frac{s}{s^2 - \alpha^2} \quad (\sigma \geq |\text{Re}(\alpha)|)$$

$$\text{sh}(\alpha t) = \frac{1}{2}(e^{\alpha t} - e^{-\alpha t}) \xleftrightarrow{\mathcal{L}} \frac{\alpha}{s^2 - \alpha^2} \quad (\sigma \geq |\text{Re}(\alpha)|)$$

对方程组的解作拉氏反变换得：

$$\begin{cases} F(t) = P_0 \text{ch}(\sqrt{3} t) + \frac{2 - P_0}{\sqrt{3}} \text{sh}(\sqrt{3} t) \\ G(t) = \text{ch}(\sqrt{3} t) + \frac{1 + P_0}{\sqrt{3}} \text{sh}(\sqrt{3} t) \end{cases}$$

由  $P = FG^{-1}$  得：

$$P(t) = \frac{F(t)}{G(t)} = \frac{\sqrt{3} \text{ch}(\sqrt{3}t) + \text{sh}(\sqrt{3}t)}{\sqrt{3} \text{ch}(\sqrt{3}t) + 2 \text{sh}(\sqrt{3}t)}$$

而最佳增益：

$$K(t) = P(t) H^T R^{-1} = P(t)$$

从而，我们可将状态估计方程表示为：

$$\dot{\hat{x}}(t) = -\hat{x}(t) + P(t) [z(t) - \hat{x}(t)]$$

即

$$\dot{\hat{s}}(t) = -\hat{s}(t) + P(t) [z(t) - \hat{s}(t)]$$

稳态解：

$$\text{方法一: } P = \lim_{t \rightarrow \infty} P(t) = \frac{\sqrt{3} + 1}{\sqrt{3} + 2} = \sqrt{3} - 1$$

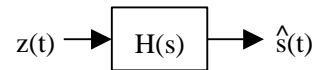
方法二：由于稳态时， $P(t) = P$ ， $\dot{P} = 0$ ，则将它们代入 Riccati 方程得：

$$0 = -P - P - P^2 + \sqrt{2}\sqrt{2} \quad \text{即} \quad P^2 + 2P - 2 = 0$$

解得：  $P = -1 \pm \sqrt{3}$ ，取其正值得：  $P = \sqrt{3} - 1$ 。

此时，状态估计方程为：

$$\hat{\dot{x}}(t) + \sqrt{3} \hat{x}(t) = (\sqrt{3} - 1) z(t)$$



也就是说，可将观测过程  $z(t)$  通过一个定常滤波器（其传递函数为  $H(s) = \frac{\sqrt{3}-1}{s+\sqrt{3}}$ ）来获得信号的估计  $\hat{s}(t)$ （即  $\hat{x}(t)$ ）（见右上图）。

#例毕

## 3.7 Kalman 滤波的稳定与发散

### 3.7.1 滤波的稳定性

如果我们对系统初始状态的均值和方差根本不知道或了解不准确，那么在进行 Kalman 滤波时我们所选取的初值就可能与系统本身的初值不同： $\hat{X}_0 \neq E(X_0)$ ， $P_0 \neq \text{Var}(X_0)$ 。在这种情况下，我们就需要研究初值的不准确性对以后计算所得的滤波值有何影响。比如：

——是否需要初值误差充分小，才能导致滤波估值达到其最优估值？

——是否无论怎样选取初值，都能保证随着  $t$  的增加使滤波值逐渐达到其最优估值？

滤波的稳定性是指：在时间充分长以后，初值的影响可以忽略。

设系统状态方程为： $X_k = \Phi_{k,k-1} X_{k-1} + W_{k-1}$

（这里讨论的是离散系统，但其结论对连续系统也是适用的。）

根据状态方程，若令初值取为  $X_0$ ，则可推得： $X_k = \Phi_{k,0} X_0 + \sum_{l=1}^k \Phi_{k,l} W_{l-1}$

若取另一初值为  $X'_0$ ，则相应地，我们有： $X'_k = \Phi_{k,0} X'_0 + \sum_{l=1}^k \Phi_{k,l} W_{l-1}$

故，系统  $k$  时刻状态之差为： $X'_k - X_k = \Phi_{k,0} (X'_0 - X_0)$

取范数（其定义为  $\|A\| = \sqrt{\text{Tr}(A^T A)}$ ）后得： $\|X'_k - X_k\| = \|\Phi_{k,0}\| \|X'_0 - X_0\|$

那么，对这样的系统，我们可知：

#### 1) 系统是稳定的：

如果（存在） $\exists c > 0$ ，对（所有） $\forall k \geq 0$ ，使得： $\|\Phi_{k,0}\| \leq c$ ，

那么，此时对任意  $\varepsilon > 0$ ，只要  $\|X'_0 - X_0\| < \delta = \varepsilon / c$ ，则可保证有： $\|X'_k - X_k\| < \varepsilon$ ，即是说：解误差在给定的范围内。

#### 2) 系统是渐近稳定的：

如果： $\lim_{k \rightarrow \infty} \|\Phi_{k,0}\| = 0$ ，

那么，对任何初值  $X'_0$  与  $X_0$ ，均有： $\lim_{k \rightarrow \infty} \|X'_k - X_k\| = 0$ ，即是说，在  $k$  充分大之后，初值的影响消失。

#### 3) 系统是一致渐近稳定的：

如果  $\exists c_1 > 0$ ， $\exists c_2 > 0$ ，对  $\forall k \geq l \geq 0$ ，使得： $\|\Phi_{k,l}\| \leq c_2 e^{-c_1(k-l)}$ ，

那么，系统除了具有上述 1) 和 2) 两条性质外，还具有另一条性质：有界输入必然导致有界输

出。即，如果  $\varepsilon > 0$ ，对  $\forall k \geq 0$ ，使得： $\|W_k\| \leq r$ ， $\|X_0\| \leq r$ ，则有： $\|X_k\| \leq \frac{c_2 r}{1 - e^{-c_1}}$ 。（此结论可由范数性质证明，即：

$$\|X_k\| \leq \|\Phi_{k,0}\| \|X_0\| + \sum_{l=1}^k \|\Phi_{k,l}\| \|W_{l-1}\| \leq c_2 e^{-c_1 k} r + c_2 r \sum_{l=1}^k e^{-c_1(k-l)} \leq \frac{c_2 r}{1 - e^{-c_1}}。$$

总之，对系统的稳定性而言，一致渐近稳定  $\Rightarrow$  渐近稳定  $\Rightarrow$  稳定。

下面，我们给出关于系统稳定性的定理：

**滤波稳定性定理：**

如果系统是一致完全能控制的和一致完全能观测的<sup>[王照林等编，《现代控制理论基础》]</sup>，那么 Kalman 滤波系统是一致渐近稳定的。

据此，我们可有如下的一些推论：

**推论 1：**

对这样的系统，当时间充分长之后，Kalman 滤波值将渐近地不依赖于滤波初值的选取，而且有界观测输入必导致有界滤波输出。

**推论 2：**

对这样的系统，当时间充分长之后，Kalman 滤波误差方差阵与增益矩阵也将渐近地不依赖于初始方差阵的选取。

**推论 3：**

对这样的定常系统，当时间充分长之后，Kalman 滤波误差方差阵  $P_k$  将趋向于一个唯一确定的正定矩阵  $P$ ，从而 Kalman 滤波增益矩阵也将趋向于一个唯一确定的增益矩阵  $K$ ，这时，滤波达到稳态，可免去递推计算过程。

离散定常系统： $\Phi_{k,k-1} = \Phi$ ， $H_k = H$ ， $Q_k = Q$ ， $R_k = R$ 。

当  $k \rightarrow \infty$  时， $P_k \rightarrow P$ ， $P_{k|k-1} \rightarrow P_{k+1|k} \rightarrow M$ ， $K_k \rightarrow K$ 。

为求得稳态解，我们将  $K_k = P_{k|k-1} H^T (H P_{k|k-1} H^T + R)^{-1}$  代入  $P_k = (I - K_k H) P_{k|k-1}$  中，再将所获得的  $P_k$  代入  $P_{k+1|k} = \Phi P_k \Phi^T + Q$  中，得 Ricatti 差分方程：

$$P_{k+1|k} = \Phi [P_{k|k-1} - P_{k|k-1} H^T (H P_{k|k-1} H^T + R)^{-1} H P_{k|k-1}] \Phi^T + Q$$

稳态时，此方程退化为如下的离散型 Ricatti 代数方程：

$$M = \Phi [M - M H^T (H M H^T + R)^{-1} H M] \Phi^T + Q$$

由此方程解出  $M$ （所存在的唯一稳态解），即可推得  $K$  及  $P$ 。

连续定常系统： $A$ 、 $B$ 、 $H$ 、 $Q$ 、 $R$  均为常数矩阵。

稳态时（参见上节例题），在 Ricatti 微分方程中，令  $\dot{P} = 0$ ，即可求出稳态解  $P$ 。

### 3.7.2 滤波的发散问题

上一节讨论了滤波初值的选取对滤波估值的影响，并且得出结论：如果系统是一致渐近稳定的，那么在时间充分长之后初值对滤波的影响可以忽略。现在，我们将讨论另一类问题：当系统的数学模型和实际的物理模型并不完全一致时，模型的差别会对滤波估值产生什么样的影响。

理论上，观测数据  $\uparrow \Rightarrow$  状态估计精确性  $\uparrow$ 。

实际上，在计算中误差方差可变得很小（滤波增益也相应地变得很小），而实际误差可能会很大，即，Kalman 滤波出现“发散”现象。

出现“发散”的原因可能有如下几个方面：

1) 对系统的物理问题了解不够精确，从而导致了所建立的数学模型与系统的实际物理状况不相符；

或者，虽然对系统的物理问题了解足够精确，但相应的数学模型过于复杂（如：非线性系统或系统的维数很高），须对模型进行简化（如：线性化或降维），从而带来系统描述的不精确性。

- 2) 对系统的动态噪声和（或）观测噪声的统计特性缺乏足够的了解，从而导致对它们的取值发生偏差。
- 3) 计算机固有的有限字长带来计算的舍入误差，从而使计算出的误差矩阵失去正定性，甚至失去对称性，导致计算值与理论值的偏差。

下面，我们举例说明这一问题。

◆例：

若实际系统的物理模型为： 
$$\begin{cases} x_k = x_{k-1} + \alpha, \text{ 其中 } \alpha \text{ 为常数} \\ z_k = x_k + v_k, \text{ 其中 } v_k \text{ 为观测噪声, } E(v_k)=0, \text{Var}(v_k)=\sigma^2 \end{cases}$$

而我们所建立的数学模型为： 
$$\begin{cases} x'_k = x'_{k-1} \\ z_k = x'_k + v_k \end{cases}$$

又设滤波的初始状态估计为：  $\hat{x}'_0 = 0, P'_0 = \infty$ ，即无系统初始状态的任何知识。

现在，我们来讨论系统模型的差异对 Kalman 滤波的影响。

首先，通过数学模型来计算估值及其均方误差：

由 Kalman 滤波公式（第二组）及已知条件可算出：

$$\Phi'_{k,k-1} = 1, H'_k = 1, R'_k = \sigma^2, Q'_k = 0,$$

$$P'_{-1} = P'_{1|0} = P'_0 + H'^T_1 R'^{-1}_1 H'_1$$

$$= (\Phi'_{1,0} P'_0 \Phi'^T_{1,0} + Q'_0) + H'^T_1 R'^{-1}_1 H'_1 = (1/\infty) + (1/\sigma^2) = \sigma^{-2}$$

$$P'_{k-1} = P'_{k|k-1} = P'_{k-1} + \sigma^{-2} \xrightarrow{\text{重复此公式}} (P'_{k-2} + \sigma^{-2}) + \sigma^{-2} = \dots = k \sigma^{-2}$$

$$P'_k = \sigma^2 / k$$

$$K'_k = P'_k H'^T_k R'^{-1}_k = (\sigma^2 / k) (1/\sigma^2) = 1/k$$

故可得状态估值：

$$\hat{x}'_k = \hat{x}'_{k|k-1} + K'_k (z_k - H'_k \hat{x}'_{k|k-1}) \xrightarrow{\hat{x}'_{k|k-1} = \Phi'_{k,k-1} \hat{x}'_{k-1} = \hat{x}'_{k-1}} \hat{x}'_{k-1} + (1/k)(z_k - H'_k \hat{x}'_{k-1})$$

$$= \frac{k-1}{k} \hat{x}'_{k-1} + \frac{1}{k} z_k \xrightarrow{\text{重复此公式}} \frac{k-1}{k} \left( \frac{k-2}{k-1} \hat{x}'_{k-2} + \frac{1}{k-1} z_{k-1} \right) + \frac{1}{k} z_k = \dots = \frac{1}{k} \sum_{i=1}^k z_i$$

因此，当  $k \rightarrow \infty$  时，状态估值  $\hat{x}'_k \rightarrow \bar{z}$ ，而其估值均方误差  $P'_k = \sigma^2/k \rightarrow 0$ 。即，从理论计算上讲，估值误差趋于零。

但是，实际观测值是对真实物理系统的观测，即：

$$z_i = x_i + v_i = x_0 + i\alpha + v_i$$

故，前面按数学模型所得的估值为：

$$\hat{x}'_k = \frac{1}{k} \sum_{i=1}^k z_i = \frac{1}{k} \sum_{i=1}^k (x_0 + i\alpha + v_i) = x_0 + \frac{k+1}{2} \alpha + \frac{1}{k} \sum_{i=1}^k v_i$$

这样，实际的估值误差为：

$$\tilde{x}_k = \hat{x}'_k - x_k = (x_0 + \frac{k+1}{2} \alpha + \frac{1}{k} \sum_{i=1}^k v_i) - (x_0 + k\alpha) = \frac{1}{k} \sum_{i=1}^k v_i - \frac{k-1}{2} \alpha$$

其均值为：

$$E(\tilde{x}_k) = -\frac{k-1}{2}\alpha \quad (\text{当 } k \rightarrow \infty \text{ 时, } E(\tilde{x}_k) \rightarrow \infty.)$$

相应的估值均方误差为:

$$P_k = E(\tilde{x}_k^2) = E\left[\left(\frac{1}{k} \sum_{i=1}^k v_i - \frac{k-1}{2}\alpha\right)^2\right] = \frac{1}{k}\sigma^2 + \frac{(k-1)^2}{4}\alpha^2 \quad (\text{当 } k \rightarrow \infty \text{ 时, } P_k \rightarrow \infty.)$$

可见: 当系统模型出现偏差时, 理论计算的估值均方误差 $\rightarrow 0$ , 而实际上其估值均方误差 $\rightarrow \infty$ , 即出现了所谓“发散”现象。

另外, 即使我们在数学模型中加入一个常数 $\beta$  ( $\beta \neq \alpha$ ) 以改善模型的精确性, 即

$$\mathbf{x}'_k = \mathbf{x}'_{k-1} + \beta$$

那么, 按新模型计算所得的估值、估值误差及估值均方误差分别为:

$$\hat{x}'_k = \frac{k-1}{2}\beta + \frac{1}{k} \sum_{i=1}^k z_i$$

$$\tilde{x}_k = \frac{1}{k} \sum_{i=1}^k v_i - \frac{k-1}{2}(\alpha - \beta)$$

$$P_k = \frac{1}{k}\sigma^2 + \frac{(k-1)^2}{4}(\alpha - \beta)^2 \quad (\text{当 } k \rightarrow \infty \text{ 时, } P_k \rightarrow \infty.)$$

可见: 当 $k \rightarrow \infty$ 时,  $P_k \rightarrow \infty$  (除非 $\beta = \alpha$ ), 即 Kalman 滤波的发散现象依然存在。

如果我们分析发散的原因, 那么我们看到:

$P'_k \downarrow \downarrow$  (过快)  $\Rightarrow K'_k (=P'_k H'^T R'^{-1}) \downarrow \downarrow \Rightarrow$  新观测值  $z_k$  对估值的修正作用 (新息的作用)  $\downarrow \downarrow \Rightarrow$  估值与观测值的分离  $\uparrow \uparrow \Rightarrow$  不精确模型在计算估值上所起的作用  $\uparrow \uparrow \Rightarrow$  估值误差的积累  $\uparrow \uparrow \Rightarrow$  滤波发散。

可采取的解决办法之一是:

在经过  $M$  次观测计算后, 限制滤波增益  $K_k$  的值不小于某个下限 (如:  $1/M$ ), 即令

$$K^*_k = \begin{cases} \text{计算值(本例中为 } 1/k), & \text{当 } k \leq M \text{ 时;} \\ 1/M, & \text{当 } k > M \text{ 时。} \end{cases}$$

可以证明<sup>[王宏禹, 《随机数字信号处理》, pp.235-239]</sup>, 我们这样所得的估值均方误差为:

$$P^*_k = (M-1)^2 \left[ 1 - \frac{1}{2} \left( 1 - \frac{1}{M} \right)^{k-M} \right]^2 \alpha^2 + \frac{\sigma^2}{2M-1} \left[ 1 + \left( 1 - \frac{1}{M} \right)^{2k-2M+1} \right]$$

从而, 当 $k \rightarrow \infty$ 时,  $P^*_k \rightarrow (M-1)^2 \alpha^2 + \frac{\sigma^2}{2M-1}$ , 即估值均方误差不再趋于无穷大了。

但是, 无论对原有的实际系统模型、还是对所选取的数学模型而言, 此时的估值都不是最佳的了。若实际系统中的 $\alpha=0$ , 则在 $k$ 充分大后, 修正滤波增益后的估值均方误差 $P^*_k$ 将大于未经修正的估值均方误差 $P_k$ ; 但若 $\alpha \neq 0$ , 则修正滤波增益后的估值避免了发散现象, 其所付出的代价是: 该估值为非最佳估值。

当然,  $M$  值的选取不是任意的, 为使估值均方误差 $P^*_k$ 小于观测噪声方差 $\sigma^2$  (否则还不如直接使用观测值 $z_k$ 为系统状态 $x_k$ 的估值), 由不等式 $(M-1)^2 \alpha^2 + \frac{\sigma^2}{2M-1} < \sigma^2$ 得:

$$1 < M < 1 + \frac{\sigma}{\alpha}$$

可见: 只有当 $\alpha \ll \sigma$ 时, 此改进方法才有价值。

所以，在实际应用中，如何选取克服滤波发散的方法需要作具体、仔细的分析。

#例毕

#### ◆克服发散现象的方法：

一般情况下，我们可采取下述的一些方法来克服滤波发散问题：

1) 限制滤波增益减少，使观测值始终能作用于估值：

a) 直接增加增益矩阵： $K_k \uparrow$  (限定下界法)；

b) 限制误差协方差阵： $P_k \uparrow \Rightarrow K_k \uparrow$ ；

c) 人为地增加动态噪声方差： $Q_k \uparrow \Rightarrow P_{k+1|k} \uparrow \Rightarrow K_{k+1} \uparrow$  (伪随机噪声法)。

2) 渐消记忆法：

滤波器逐渐“忘掉”老数据以增强新观测值的作用。例如：用指数因子 $\alpha^{k-l}$  ( $\alpha > 1, l < k$ ) 加权观测噪声方差阵  $R_l$ ，这意味着， $l \downarrow$  (离  $k$  时刻越“老”)  $\Rightarrow$  加权后的  $R_l \uparrow \Rightarrow$  对应的滤波增益  $K_l \downarrow \Rightarrow l$  时刻观测值对  $k$  时刻的估值作用  $\downarrow$ ，即逐渐削弱老数据对估值的作用，反之，就是加强了新观测数据对估值的修正作用。

3) 限定记忆法：

根据 Kalman 递推过程，Kalman 滤波利用了所有的观测数据。现在，如果我们只利用离当前  $k$  时刻最近的前  $N$  个观测值，那么这就相当于上述 2) 中的办法 (只是逐渐“忘掉”老数据改为“去掉”那些过旧的数据)。

4) 自适应滤波：

利用观测数据，对不精确的系统模型进行修正。(这是另一门课程的内容)

5) 采用平方根法克服计算机数值发散：

依据：若  $x \in [10^{-6}, 10^6]$ ，则  $\sqrt{x} \in [10^{-3}, 10^3]$ ，这样可降低计算的数值精度。在 Kalman 滤波中，我们使用矩阵的平方根公式来降低所要求的计算精度，即令：

$$P_k = G_k G_k^T, P_{k|k-1} = G_{k|k-1} G_{k|k-1}^T。$$

### 3.8 Kalman 滤波的推广 (简介)

在上面所讲述的 Kalman 滤波方法中，所涉及的系统为线性系统，系统的动态噪声和观测噪声均为白噪声，主要考虑的是滤波估值问题。

现在，我们将在更一般的情况下进行拓宽研究。

#### ◆在有色噪声情况下的滤波：

采用扩大状态向量法。

例如，设某系统模型如下：

$$\text{状态方程} \quad \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ \xi_1 \end{pmatrix}$$

$$\text{观测方程} \quad \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} c_1 & c_2 \\ c_3 & c_4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ \xi_2 \end{pmatrix}$$

其中， $w_1$  和  $v_1$  为白噪声，而  $\xi_1$  和  $\xi_2$  为有色噪声并分别由下述方程描述：

$$\dot{\xi}_1 = d \cdot \xi_1 + w_2$$

$$\dot{\xi}_2 = e \cdot \xi_2 + w_3$$

这里， $w_2$  和  $w_3$  均为白噪声。

那么，我们可将有色噪声  $\xi_1$  和  $\xi_2$  亦作为状态变量加入到系统原有的状态矢量  $(x_1, x_2)^T$  中而形成新的系统



状态矢量 $(x_1, x_2, \xi_1, \xi_2)^T$ 。相应地，新系统的数学模型可写为：

$$\text{新状态方程} \quad \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{\xi}_1 \\ \dot{\xi}_2 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & 0 & 0 \\ a_3 & a_4 & 1 & 0 \\ 0 & 0 & d & 0 \\ 0 & 0 & 0 & e \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ 0 \\ w_2 \\ w_3 \end{pmatrix}$$

$$\text{新观测方程} \quad \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} c_1 & c_2 & 0 & 0 \\ c_3 & c_4 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ 0 \end{pmatrix}$$

如此改造之后，新系统就只有白噪声了，从而原则上可以应用前面的 Kalman 滤波公式来求解（这里说“原则上”是因为与前面的滤波推导不同的是此时的观测噪声矢量中有零元素）。

但是，如果原来的系统观测方程中未含有白噪声而只有有色噪声（如：在上面的观测方程中  $v_1=0$ ），那么观测噪声在新的观测方程中为零，从而导致其协方差矩阵  $R$  变为奇异矩阵，使得 Kalman 滤波方法不再适用。故，此时须另寻其它解决方法（见参考书《现代控制理论基础》）。

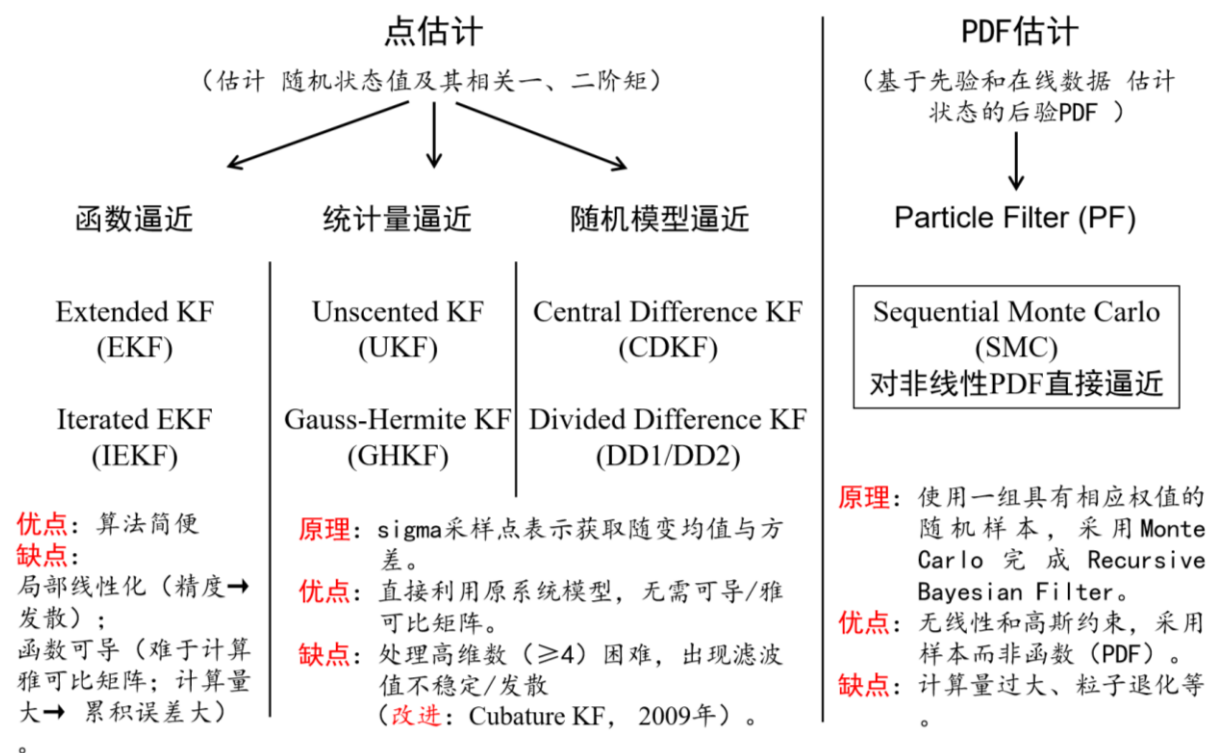
#### ◆预测问题和平滑问题：

预测：首先求得滤波估值  $\hat{X}(t)$  ( $t_0 \leq t \leq t_1$ )，然后求状态转移矩阵  $\Phi(t_2, t_1)$ ，最后得到预测估值  $\hat{X}(t_2|t_1) = \Phi(t_2, t_1) \hat{X}(t)$  ( $t_2 > t_1$ )。

平滑：非实时处理，需推导相应的算法。

### 3.9 非线性滤波

#### ◆概述



#### ◆非线性系统的线性化处理方法——EKF（点估计）：

设有非线性系统模型如下：

$$X_k = f_{k-1}(X_{k-1}) + \Gamma_{k-1}(X_{k-1}) W_{k-1}$$

$$Z_k = \mathbf{h}_k(\mathbf{X}_k) + \mathbf{V}_k$$

其中,  $\mathbf{W}_k \sim \mathcal{N}(0, \mathbf{Q}_k)$ ,  $\mathbf{V}_k \sim \mathcal{N}(0, \mathbf{R}_k)$ ,  $\mathbf{X}_0 \sim \mathcal{N}(\mu_0, \mathbf{P}_0)$ ,  $\{\mathbf{W}_k\}$ 、 $\{\mathbf{V}_k\}$ 、 $\mathbf{X}_0$  相互独立。

注：对于定常线性系统  $\mathbf{X}_k = \Phi_{k,k-1} \mathbf{X}_{k-1} + \Gamma_{k-1} \mathbf{W}_{k-1}$ , 其对应的 Kalman 滤波公式与前述系统模型  $\mathbf{X}_k = \Phi_{k,k-1} \mathbf{X}_{k-1} + \mathbf{W}_{k-1}$  的公式之间的差别仅在于一步预测误差方差的公式上, 即此时应改为:  $\mathbf{P}_{k|k-1} = \Phi_{k,k-1} \mathbf{P}_{k-1} \Phi_{k,k-1}^T + \Gamma_{k-1} \mathbf{Q}_{k-1} \Gamma_{k-1}^T$ 。

我们可采取下述的方法之一以使系统模型线性化, 从而达到利用已知的线性系统 Kalman 滤波方法来求得估值的目的。

**方法一：围绕标称状态的线性化**

定义：标称状态序列  $\{\mathbf{X}_k^0\}$

$\mathbf{X}_k^0 = \mathbf{f}_{k-1}(\mathbf{X}_{k-1}^0)$  (无噪声干扰的状态方程), 其初始值为  $\mathbf{X}_0^0 = \mu_0 = \mathbf{E}(\mathbf{X}_0)$ 。

那么, 利用 Taylor 展开式  $f(\mathbf{x}) = f(\mathbf{x}_0) + f'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + (1/2!)f''(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)^2 + \dots$ , 并忽略二次和二次以上的项 (假定围绕标称状态的偏离较小) 后, 得到:

$$\mathbf{f}_{k-1}(\mathbf{X}_{k-1}) \approx \mathbf{f}_{k-1}(\mathbf{X}_{k-1}^0) + \frac{\partial \mathbf{f}_{k-1}}{\partial \mathbf{X}_{k-1}^0} (\mathbf{X}_{k-1} - \mathbf{X}_{k-1}^0)$$

又定义：状态偏离  $\mathbf{X}_k^\Delta = \mathbf{X}_k - \mathbf{X}_k^0$ 。

那么, 将系统的状态方程  $\mathbf{X}_k = \mathbf{f}_{k-1}(\mathbf{X}_{k-1}) + \Gamma_{k-1}(\mathbf{X}_{k-1}) \mathbf{W}_{k-1}$  及标称状态序列的定义代入状态偏离的表达式, 可得:

$$\begin{aligned} \mathbf{X}_k^\Delta &= [\mathbf{f}_{k-1}(\mathbf{X}_{k-1}) + \Gamma_{k-1}(\mathbf{X}_{k-1}) \mathbf{W}_{k-1}] - \mathbf{f}_{k-1}(\mathbf{X}_{k-1}^0) \\ &\approx [\mathbf{f}_{k-1}(\mathbf{X}_{k-1}^0) + \frac{\partial \mathbf{f}_{k-1}}{\partial \mathbf{X}_{k-1}^0} (\mathbf{X}_{k-1} - \mathbf{X}_{k-1}^0) + \Gamma_{k-1}(\mathbf{X}_{k-1}) \mathbf{W}_{k-1}] - \mathbf{f}_{k-1}(\mathbf{X}_{k-1}^0) \\ &\approx \frac{\partial \mathbf{f}_{k-1}}{\partial \mathbf{X}_{k-1}^0} (\mathbf{X}_{k-1} - \mathbf{X}_{k-1}^0) + \Gamma_{k-1}(\mathbf{X}_{k-1}^0) \mathbf{W}_{k-1} \end{aligned}$$

这里, 作一近似替换, 认为  $\Gamma_{k-1}(\mathbf{X}_{k-1}) \approx \Gamma_{k-1}(\mathbf{X}_{k-1}^0)$ 。

由此产生了新的状态方程:

$$\mathbf{X}_k^\Delta \approx \frac{\partial \mathbf{f}_{k-1}}{\partial \mathbf{X}_{k-1}^0} \mathbf{X}_{k-1}^\Delta + \Gamma_{k-1}(\mathbf{X}_{k-1}^0) \mathbf{W}_{k-1}$$

$$\text{其中: } \frac{\partial \mathbf{f}_{k-1}}{\partial \mathbf{X}_{k-1}^0} \triangleq \frac{\partial \mathbf{f}_{k-1}(\mathbf{X}_{k-1})}{\partial \mathbf{X}_{k-1}} \bigg|_{\mathbf{X}_{k-1} = \mathbf{X}_{k-1}^0} \triangleq \begin{pmatrix} \frac{\partial f_{k-1}^{(1)}(\mathbf{X}_{k-1})}{\partial \mathbf{X}_{k-1}^{(1)}} & \dots & \frac{\partial f_{k-1}^{(1)}(\mathbf{X}_{k-1})}{\partial \mathbf{X}_{k-1}^{(M)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{k-1}^{(M)}(\mathbf{X}_{k-1})}{\partial \mathbf{X}_{k-1}^{(1)}} & \dots & \frac{\partial f_{k-1}^{(M)}(\mathbf{X}_{k-1})}{\partial \mathbf{X}_{k-1}^{(M)}} \end{pmatrix} \bigg|_{\mathbf{X}_{k-1} = \mathbf{X}_{k-1}^0}$$

(记号  $\mathbf{Y}^{(n)}$  表示矢量  $\mathbf{Y}$  中的第  $n$  个元素)

再定义：观测偏离  $\mathbf{Z}_k^\Delta = \mathbf{Z}_k - \mathbf{h}_k(\mathbf{X}_k^0)$ 。

类似于状态偏离的推导方法, 我们可得到新的观测方程:

$$\mathbf{Z}_k^\Delta \approx \frac{\partial \mathbf{h}_k}{\partial \mathbf{X}_k^0} \mathbf{X}_k^\Delta + \mathbf{V}_k$$

$$\text{其中: } \frac{\partial \mathbf{h}_k}{\partial \mathbf{X}_k^0} \bigg|_{\mathbf{X}_k = \mathbf{X}_k^0} \stackrel{\Delta}{=} \left( \begin{array}{ccc} \frac{\partial \mathbf{h}_k^{(1)}(\mathbf{X}_k)}{\partial \mathbf{X}_k^{(1)}} & \dots & \frac{\partial \mathbf{h}_k^{(1)}(\mathbf{X}_k)}{\partial \mathbf{X}_k^{(M)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{h}_k^{(N)}(\mathbf{X}_k)}{\partial \mathbf{X}_k^{(1)}} & \dots & \frac{\partial \mathbf{h}_k^{(N)}(\mathbf{X}_k)}{\partial \mathbf{X}_k^{(M)}} \end{array} \right) \bigg|_{\mathbf{X}_k = \mathbf{X}_k^0}$$

这样，我们就可以应用前面所讲述的 Kalman 滤波方法了。其初始条件为  $\hat{\mathbf{X}}_0^\Delta = \mathbf{E}(\mathbf{X}_0^\Delta) = \mathbf{0}$  及  $\mathbf{P}_0 = \text{Var}(\mathbf{X}_0^\Delta) = \text{Var}(\mathbf{X}_0)$ ，最后所得  $\mathbf{X}_k$  的最佳滤波估值为：

$$\hat{\mathbf{X}}_k = \hat{\mathbf{X}}_k^\Delta + \mathbf{X}_k^0$$

**缺点：**当真实状态  $\mathbf{X}_k$  偏离标称状态  $\mathbf{X}_k^0$  越来越大时，估值误差就会增加（因为在状态方程的近似过程中我们忽略了 Taylor 展开式中状态偏离的二次项及其上的高阶项）。

**改进：**采用下面将要叙述的方法。

**方法二：推广的 Kalman 滤波**

假设观测时刻  $k$  以前已得滤波估值  $\hat{\mathbf{X}}_{k-1}$ ，则可采用类似于方法一的近似方法，把状态方程围绕  $\hat{\mathbf{X}}_{k-1}$  展开：

$$\mathbf{X}_k \approx \mathbf{f}_{k-1}(\hat{\mathbf{X}}_{k-1}) + \frac{\partial \mathbf{f}_{k-1}}{\partial \hat{\mathbf{X}}_{k-1}} (\mathbf{X}_{k-1} - \hat{\mathbf{X}}_{k-1}) + \mathbf{\Gamma}_{k-1}(\hat{\mathbf{X}}_{k-1}) \mathbf{W}_{k-1}$$

再把观测方程围绕  $\hat{\mathbf{X}}_{k|k-1}$  展开：

$$\mathbf{Z}_k \approx \mathbf{h}_k(\hat{\mathbf{X}}_{k|k-1}) + \frac{\partial \mathbf{h}_k}{\partial \hat{\mathbf{X}}_{k|k-1}} (\mathbf{X}_k - \hat{\mathbf{X}}_{k|k-1}) + \mathbf{V}_k$$

整理后，可得新的系统模型如下：

$$\mathbf{X}_k \approx \frac{\partial \mathbf{f}_{k-1}}{\partial \hat{\mathbf{X}}_{k-1}} \mathbf{X}_{k-1} + \mathbf{U}_{k-1} + \mathbf{\Gamma}_{k-1}(\hat{\mathbf{X}}_{k-1}) \mathbf{W}_{k-1}$$

$$\mathbf{Z}_k \approx \frac{\partial \mathbf{h}_k}{\partial \hat{\mathbf{X}}_{k|k-1}} \mathbf{X}_k + \mathbf{Y}_k + \mathbf{V}_k$$

$$\text{其中: } \mathbf{U}_{k-1} = \mathbf{f}_{k-1}(\hat{\mathbf{X}}_{k-1}) - \frac{\partial \mathbf{f}_{k-1}}{\partial \hat{\mathbf{X}}_{k-1}} \hat{\mathbf{X}}_{k-1}, \quad \mathbf{Y}_k = \mathbf{h}_k(\hat{\mathbf{X}}_{k|k-1}) - \frac{\partial \mathbf{h}_k}{\partial \hat{\mathbf{X}}_{k|k-1}} \hat{\mathbf{X}}_{k|k-1}。$$

这样，我们就可以利用一般线性系统的 Kalman 滤波公式来得到最佳估值（关于一般线性系统的 Kalman 滤波公式请参见参考书《现代控制理论基础》）。

#### ◆非线性系统处理方法——粒子滤波 PF (Particle Filter)：

状态方程：  $\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}, \mathbf{v}_{k-1})$

观测方程：  $\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{n}_k)$

已知：前一时刻的后验概率密度  $p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})$ ，

目标：估计后验概率密度  $p(\mathbf{x}_k | \mathbf{y}_{1:k})$ 。

步骤：

1) 预测过程：

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k, \mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}$$

$$= \int p(x_k | x_{k-1}) p(x_{k-1} | y_{1:k-1}) dx_{k-1}$$

2) 更新过程:

$$p(x_k | y_{1:k}) = [p(y_k | x_k) p(x_k | y_{1:k-1})] / p(y_k | y_{1:k-1})$$

$$\text{其中, 归一化常数 } p(y_k | y_{1:k-1}) = \int p(y_k | x_k) p(x_k | y_{1:k-1}) dx_k$$

实际问题: 上述计算过程中需要用到积分, 这对于一般的**非线性、非高斯**系统, 很难得到后验概率密度的解析形式。

采样:

为了解决难以积分的问题, 可以用**蒙特卡洛采样**来代替计算后验概率密度:

假设从后验概率密度中采样到  $N$  个样本, 即  $N$  个粒子。

当  $N$  足够大时, 那么后验概率密度可近似计算为:

$$p(x_n | y_{1:k}) \approx \hat{p}(x_n | y_{1:k}) = (1/N) \sum_{i=1}^N \delta(x_n - x_n^{(i)})$$

以做图像目标跟踪为例, 我们希望知道当前状态的期望值, 即

$$\begin{aligned} E[f(x_n)] &\approx \int f(x_n) \hat{p}(x_n | y_{1:k}) dx_n \quad \text{其中: } f(x_n) \text{ 是粒子的状态函数} \\ &= (1/N) \sum_{i=1}^N \int f(x_n) \delta(x_n - x_n^{(i)}) dx_n \\ &= (1/N) \sum_{i=1}^N f(x_n^{(i)}) \quad (\text{采样粒子状态值的平均值}) \end{aligned}$$

但是, 后验概率密度是需要寻求的目标, 此时未知, 因此无法从后验概率密度函数直接采样。

#### ● 重要性采样 (Importance Sampling):

解决的办法, 就是从一个已知的密度函数 (记为  $q$ ——**重要性概率密度函数**) 进行采样。因而, 上述问题可演变为:

$$\begin{aligned} E[f(x_k)] &= \int f(x_k) \frac{p(x_k | y_{1:k})}{q(x_k | y_{1:k})} q(x_k | y_{1:k}) dx_k \\ &= \frac{E_{q(x_k | y_{1:k})}[W_k(x_k) f(x_k)]}{E_{q(x_k | y_{1:k})}[W_k(x_k)]} \\ &\quad \text{依据贝叶斯公式, 并定义: } W_k(x_k) = \frac{p(y_{1:k} | x_k) p(x_k)}{q(x_k | y_{1:k})} \propto \frac{p(x_k | y_{1:k})}{q(x_k | y_{1:k})} \\ &\approx \frac{\frac{1}{N} \sum_{i=1}^N W_k(x_k^{(i)}) f(x_k^{(i)})}{\frac{1}{N} \sum_{i=1}^N W_k(x_k^{(i)})} \quad \text{蒙特卡洛方法} \\ &= \sum_{i=1}^N \tilde{W}_k(x_k^{(i)}) f(x_k^{(i)}) \quad (\text{采样粒子状态值的直接平均值变成了加权平均}) \\ &\quad \text{其中: 定义归一化权重 } \tilde{W}_k(x_k^{(i)}) = \frac{W_k(x_k^{(i)})}{\sum_{i=1}^N W_k(x_k^{(i)})} \end{aligned}$$

如何获得权重  $W_k(x_k^{(i)})$  ?

考虑观测时序, 利用递归方式, 推导可得:

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(y_k | x_k^{(i)}) p(x_k^{(i)} | x_{k-1}^{(i)})}{q(x_k^{(i)} | x_{k-1}^{(i)}, y_k)}$$

如果选取重要性概率密度函数为：

$$q(x_k^{(i)} | x_{k-1}^{(i)}, y_k) = p(x_k^{(i)} | x_{k-1}^{(i)}) \quad (\text{可由状态转移方程得到})$$

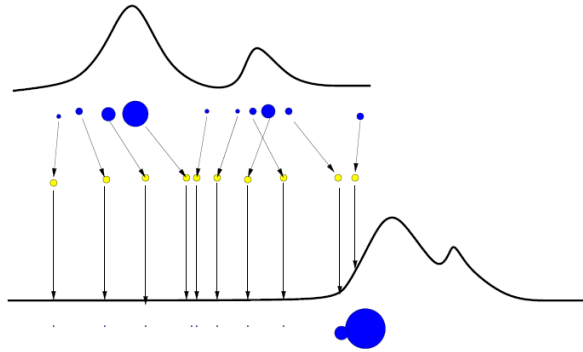
则权重计算简化为：

$$w_k^{(i)} \propto w_{k-1}^{(i)} p(y_k | x_k^{(i)}) \quad (\text{可由观测方程得到})$$

### ● 重采样 (Re-Sampling)：

上述重要性采样存在一个问题：

- 随着迭代，某些粒子的权重可能会越来越小，反之，个别粒子的权重会越来越大；
- 权重小的粒子几乎没有作用，称为“无效粒子”；
- 随着无效采样粒子数目的增加，使得大量的计算资源浪费在对估计后验滤波概率分布几乎不起作用的粒子上，使得估计性能下降。



重采样的思路：

- 舍弃“无效粒子”并保持粒子数目不变。因此，要生成“新粒子”来取代它们。
- “新粒子”是保留的“有效粒子”的复制品，权重越大，可由它复制生成的“新粒子”就越大。

从数学上来讲，就是把重要性采样改为**重采样**，即：将

$$p(x_k | y_{1:k}) = \sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$$

变成

$$\tilde{p}(x_k | y_{1:k}) = \sum_{j=1}^N \frac{1}{N} \delta(x_k - x_k^{(j)}) = \sum_{i=1}^N \frac{n_i}{N} \delta(x_k - x_k^{(i)})$$

其中  $n_i$  是第  $i$  个粒子被复制的次数。

注意到每次重采样后，权重都要满足：

$$w_{k-1}^{(i)} = \frac{1}{N}$$

因此，权重的迭代更新简化为：

$$w_k^{(i)} \propto w_{k-1}^{(i)} p(y_k | x_k^{(i)})$$

即

$$w_k^{(i)} \propto p(y_k | x_k^{(i)})$$

总结：

标准的粒子滤波算法流程如下：

1. 粒子集初始化 ( $k=0$ ):

对于  $i=1,2,\dots,N$ , 由先验  $p(x_0)$  生成采样粒子  $\{x_0^{(i)}\}_{i=1}^N$

2. 对于  $k=1,2,\dots$ , 循环执行以下步骤:

① 重要性采样: 对于  $i=1,2,\dots,N$ , 从重要性概率密度中生成采样粒子  $\{\tilde{x}_k^{(i)}\}_{i=1}^N$ , 计

算粒子权值  $\tilde{w}_k^{(i)}$ , 并进行归一化;

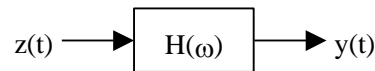
② 重采样: 对粒子集  $\{\tilde{x}_k^{(i)}, \tilde{w}_k^{(i)}\}$  进行重采样, 重采样后的粒子集为  $\{x_k^{(i)}, 1/N\}$ ;

③ 输出: 计算  $k$  时刻的状态估计值:  $\hat{x}_k = \sum_{i=1}^N \tilde{x}_k^{(i)} \tilde{w}_k^{(i)}$ 。

### 3.10 匹配滤波器

**匹配滤波器:** 瞬时输出信噪比最大的最佳线性滤波器。

在确定性信号  $s(t)$  通过加性白噪声  $n(t)$  (其功率谱密度为  $N_0/2$ ) 信道情况下, 滤波器的输入为:



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

相应地, 滤波器的输出  $y(t)$  为信号输出  $s_o(t)$  和噪声输出  $n_o(t)$  之和:

$$y(t) = s_o(t) + n_o(t)$$

这里, 输出信号可表示为  $s_o(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(\omega) H(\omega) e^{j\omega t} d\omega$ , 其瞬时功率 ( $t=t_0$  时) 为  $|s_o(t_0)|^2$ 。

而输出噪声的平均功率为  $E[n_o^2(t)] = \frac{N_0}{4\pi} \int_{-\infty}^{+\infty} |H(\omega)|^2 d\omega$ 。

从而可得  $t_0$  时刻的 (瞬时) 输出信噪比为:

$$SNR_0 = \frac{|s_o(t_0)|^2}{E[n_o^2(t)]} = \frac{\left| \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(\omega) H(\omega) e^{j\omega t_0} d\omega \right|^2}{\frac{N_0}{4\pi} \int_{-\infty}^{+\infty} |H(\omega)|^2 d\omega}$$

由 Schwartz 定理  $\left| \int f(x) f^*(x) dx \right| \leq \int |f(x)|^2 dx$  可知:

$$SNR_0 \leq \frac{2E_s}{N_0}, \quad \text{其中 } E_s = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |S(\omega)|^2 d\omega$$

所以, 当输出信噪比  $SNR_0$  为最大 (即上式取等号) 时, 我们可得到最佳滤波器 (即匹配滤波器) 的传输函数为:

$$H(\omega) = C S^*(\omega) e^{-j\omega t_0} \quad (\text{通常取 } C=1)$$

相应地, 在时域上, 滤波器的冲击响应为:

$$h(t) = s(t_0 - t)$$

---

## 3.11 小结

- 波形的线性最小均方估计  
Wiener-Hopf 方程
- Wiener 滤波器  
连续时间 Wiener 滤波器（物理不可实现与物理可实现）  
离散时间 Wiener 滤波器（有限观测区间、物理可实现与物理不可实现）  
离散 Wiener 滤波矩阵
- Kalman 滤波器  
状态空间模型  
离散时间 Kalman 滤波器、连续时间 Kalman 滤波器  
Kalman 滤波的稳定与发散、Kalman 滤波的推广
- 非线性滤波  
线性化方法  
粒子滤波
- 匹配滤波器

## 第4章 非参量和鲁棒检测与估计

### 4.1 非参量检测 (Nonparametric Detection)

参量检测：每个假设下，概率密度函数都有确定的形式。

非参量检测：未知函数形式  $\Rightarrow$  统计检测方法

- 应用：1) 难于获得输入数据的完全统计描述；  
2) 数据的统计特性是时变的；  
3) 最佳检测器结构过于复杂而不能实现。

非参量检测器的特点：对环境变化不敏感；易于实现。

代价：与最佳检测器相比稍差。

#### 4.1.1 简单非参量检测——符号检测 (Sign Test)

◆问题：考虑一组连续观测量  $X_1, X_2, \dots, X_n$ ，独立、同分布——累积分布函数  $F(\cdot)$ 。

希望确定：其分布的中值是零、或者是大于零。

$\Rightarrow$  非参量检测问题：

令  $p$  为随机样本  $X_i$  大于零的概率，

那么：若  $p=1/2$ ，则中值为零；若  $p>1/2$ ，则中值大于零。

故，可设计如下假设检验：

$$\begin{aligned} H_0: p=1/2 \\ H_1: p>1/2 \end{aligned} \quad X_i \sim \text{任意的连续累积分布函数 } F(\cdot)$$

对此问题，存在一个最佳检验：具有第一类错误概率（虚警）为  $\alpha$  的最有效检验。

证明：令  $f(\cdot)$  是  $F(\cdot)$  的密度函数。

定义：  $f^+(x_i) = f(x_i | X_i > 0)$

$f^-(x_i) = f(x_i | X_i \leq 0)$

$$f_0(x_i) = \frac{1}{2} f^+(x_i) + \frac{1}{2} f^-(x_i)$$

如此，可写出似然比：  $\lambda(\mathbf{X}) = \prod_{i=1}^n \frac{f(x_i | H_1)}{f(x_i | H_0)}$

$$\text{容易得到: } \frac{f(x_i | H_1)}{f(x_i | H_0)} = \begin{cases} \frac{pf^+(x_i)}{\frac{1}{2}f^+(x_i)} = 2p & \text{若 } x_i > 0 \\ \frac{(1-p)f^-(x_i)}{\frac{1}{2}f^-(x_i)} = 2(1-p) & \text{若 } x_i \leq 0 \end{cases}$$

$$\text{因此: } \lambda(\mathbf{X}) = 2^n p^{\sum_{i=1}^n u(x_i)} (1-p)^{n-\sum_{i=1}^n u(x_i)} \quad \text{这里 } u(x_i) = \begin{cases} 1, & x_i > 0 \\ 0, & x_i \leq 0 \end{cases}$$

故，对数似然比检验准则等效于：



$$T = \sum_{i=1}^n u(X_i) \begin{matrix} H_1 \\ > \\ H_0 \end{matrix} c \text{ 其中: } c \text{ 是使虚警为 } \alpha \text{ 的某个门限。}$$

这就是“符号检测 (Sign Test)”。

由于  $T$  取整数, 因此该规则等效于一个计数规则: 若  $T > c$ , 则判决  $H_1$  为真。

$$T \text{ 具有二项式分布: } \begin{cases} T|H_0 \sim (n, \frac{1}{2}) \\ T|H_1 \sim (n, p) \end{cases}$$

$$\text{所以, 虚警 } \alpha = \sum_{k=c+1}^n \binom{n}{k} \left(\frac{1}{2}\right)^n$$

$$\text{检测概率 } P_D = \sum_{k=c+1}^n \binom{n}{k} p^k (1-p)^{n-k}$$

一般而言, 在非参量检测中, 检测概率不能以可分析的形式确定。这里的符号检测是一个例外。

◆问题: 符号检测器也是一个 UMP 检验。但若观测的分布确知其函数形式, 则似然比检验将是最有效的吗? 比符号检测器更强?

答: 若  $X_i$  服从高斯分布  $N(\mu, \sigma^2)$ 。假设  $H_0: \mu=0$ ; 假设  $H_1: \mu>0$ 。

则似然比检验为一个线性检测器 (linear detector):

$$\sum_{i=1}^m X_i \begin{matrix} H_1 \\ > \\ H_0 \end{matrix} d \quad \text{其中: } d \text{ 由虚警 } \alpha \text{ 来确定。}$$

$$\text{从而可得: } P_D = P\left(\sum_{i=1}^m X_i > d \mid H_1\right) = \Phi\left(-\frac{d - m\mu}{\sqrt{m}\sigma}\right) \quad \text{其中: } d = \sqrt{m}\sigma\Phi^{-1}(1-\alpha)$$

所以, 对给定的  $\alpha$  而言, 线性检测器优于符号检测器 (对高斯分布有:  $p = \Phi(\sqrt{m}\mu/\sigma)$ )。

◆问题: 若希望确定其分布的中值是否为零, 怎么办?

答: 此时, 假设  $H_1: p \neq \frac{1}{2}$ 。

采用双边检验:  $T$  或太大、或太小, 而导致  $H_0$  不成立。

而  $p = \frac{1}{2}$  时,  $T$  的分布——二项式分布是对称的。

所以, 不符合  $H_0$  的区域 (两个对称的“尾”) 应为:

$$\begin{cases} T \geq c \\ T \leq c' \end{cases}, \quad \text{这里, } c' = n - c。$$

$$\text{根据虚警要求, } c \text{ 满足: } \sum_{k=c}^n \binom{n}{k} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2}$$

◆问题: 若希望确定其分布的中值为某个常量 (如:  $M_0$ ), 应该怎么处理?

答: 此问题较简单——令新的观测值为  $Z_i = X_i - M_0$ 。

#### 4.1.2 Wilcoxon (威尔柯克孙) 检测 (W 检测器)

对于确定分布中值是否为  $M_0$ , 符号检测对所有大于或小于中值的样本作同样的加权: 0 或 1。

若分布是对称的, 则  $H_0$  假设下约 50% 的样本  $\{Z_i = X_i - M_0\}$  将为正值。

考虑  $|Z_i|$  的秩排序, 即: 将  $Z_i$  的绝对值由小到大排序, 把秩 1 赋给其最小值, 而秩  $n$  给予其最大值。以

$r(|Z_i|)$  表示的秩, 则  $\sum_{i=1}^n r(|Z_i|) = \frac{n(n+1)}{2}$ 。

Wilcoxon 检测 (或称 “signed rank test”) 是基于下述统计量之一的:

$$T^+ = \sum_{i=1}^n u(Z_i) r(|Z_i|) \quad \text{或} \quad T^- = \sum_{i=1}^n [1 - u(Z_i)] r(|Z_i|)$$

显然, 我们有:  $T^+ + T^- = \frac{n(n+1)}{2}$ 。

等效地:  $T^+ = \sum_{i=1}^n i T_i$

其中  $T_i = \begin{cases} 1, & \text{若其样值 } Z \text{ (对应于 } \{|Z_i|, k=1,2,\dots,n\} \text{ 集合中第 } i \text{ 个最小值)} \text{ 是正值;} \\ 0, & \text{否则。} \end{cases}$

可证明:  $H_0$  假设下,  $T_1, T_2, \dots, T_n$  是相互独立的, 且  $P(T_i=0)=P(T_i=1)=1/2$ 。

所以,  $H_0$  假设下:  $E(T^+) = \sum_{i=1}^n i E(T_i) = \frac{n(n+1)}{4}$

$$\text{Var}(T^+) = \sum_{i=1}^n i^2 \text{Var}(T_i) = \frac{n(n+1)(2n+1)}{6} \cdot \frac{1}{4} = \frac{n(n+1)(2n+1)}{24}$$

对中值超过  $M_0$  的单边假设检验:

放弃  $H_0$ , 若  $T^+ - \frac{n(n+1)}{4} \geq k$ , 这里  $k$  被选为满足某个虚警  $\alpha$  的值。

若  $n$  很大, 则  $T^+$  近于高斯分布。  $E(T^+)$  和  $\text{Var}(T^+)$  可计算, 因此根据虚警  $\alpha$  可算出  $k$  值。

若  $n$  很小, 则可列举出  $T^+$  的所有可能取值及其概率, 再依据虚警  $\alpha$  算出  $k$  值。

若分布的对称性可以得到满足, 则对某些分布而言, Wilcoxon 检测比符号检测更有效。所谓“更有效”就是具有较高的渐近相对效率 (Asymptotic Relative Efficiency), 这是评价检测性能的一个重要参数。

#### 4.1.3 渐近相对效率 (Asymptotic Relative Efficiency)

若存在两个检测器  $T_1$  和  $T_2$ , 可达到相同的虚警  $\alpha$  和检测概率  $\beta$ , 但所需的样本数  $n_1$  和  $n_2$  不相同, 则  $T_1$  对  $T_2$  的相对效率定义为:

$$\text{RE}_{12} = \frac{n_2}{n_1}$$

若  $\text{RE}_{12} > 1$ , 则  $T_1$  比  $T_2$  更有效。

应用  $\text{RE}_{12}$  的难点是: 通常,  $n_1$  和  $n_2$  是  $\alpha$  和  $\beta$  的函数。可能对一些范围内取值的  $\alpha$  和  $\beta$ ,  $\text{RE}_{12} > 1$ ; 而对另一些范围内取值的  $\alpha$  和  $\beta$ ,  $\text{RE}_{12} < 1$ 。所以, 在整个  $\alpha$  和  $\beta$  取值范围内, 不能依据  $\text{RE}_{12}$  来判断一个检测器优于另一个检测器。

但是在许多情况下,  $n_1$  和  $n_2$  假定为大的值。在某些限制下, 当样本尺寸  $\rightarrow \infty$  时,  $\text{RE}_{12}$  趋于一个独立于  $\alpha$  和  $\beta$  的值。这样就导出了渐近相对效率 (Asymptotic Relative Efficiency: ARE) 的定义。

对于一个一致检验, 对任何给定的  $\alpha > 0$ , 随着样本数的增加 ( $\rightarrow \infty$ ), 检测概率  $\beta$  可  $\rightarrow 1$ 。任意给定的检

测概率 $\beta$  ( $<1$ ) 也可通过一个检验获得, 若备选假设  $H_1$  被允许随  $n \rightarrow \infty$  而逼近于零假设  $H_0$  的话。

考虑一个简单的零假设  $H_0: \theta = \theta_0$  相对于一个单边备选假设  $H_1: \theta \in \Theta_1 = \{\theta > \theta_0\}$  的检验问题。

在大量的信号检测问题中, 信号很弱。例如: 考虑一个序列  $\theta_n \in \Theta_1$ , 当  $n \uparrow$  时,  $\theta_n \rightarrow \theta_0$  (从上方逼近, 例:  $\theta_n = \theta_0 + \frac{k}{\sqrt{n}}$ ),  $H_1 \rightarrow H_0$ 。

形式上可定义:  $ARE_{12} = \lim_{\substack{n_1 \rightarrow \infty \\ n_2 \rightarrow \infty \\ H_1 \rightarrow H_0}} RE_{12}$

也可理解为: 对于极弱信号的检测, 在保证相同的性能 (相同的虚警 $\alpha$ 和检测概率 $\beta$ ) 前提下, 两个检测器所需要的极限样本数之比值。

◆ ARE 的另一种计算方法:

$$\text{检验 } T_n: \quad t_n(\mathbf{X}) \underset{H_0}{\overset{H_1}{>}} t_{n,\alpha}$$

其中:  $\mathbf{X}$ —— $n$  个样值的观测矢量;  $t_{n,\alpha}$ ——虚警为 $\alpha$ 的门限。

$$T_n \text{ 的效验 (efficacy): } \quad \varepsilon = \frac{\left[ \frac{dE_\theta(T_n)}{d\theta} \Big|_{\theta_0} \right]^2}{n \cdot \sigma_{\theta_0}^2(T_n)}$$

其中:  $E_\theta(T_n)$ —— $T_n$  的均值 ( $H_1$  假设下);

$\sigma_{\theta_0}^2(T_n)$ —— $\theta = \theta_0$  时  $T_n$  的方差。

可证明:  $ARE_{12} = \lim_{n \rightarrow \infty} \frac{\varepsilon_1}{\varepsilon_2}$  即:  $ARE_{12}$  为两检测器的效验之比。

◆ 一些特定检验的 ARE 计算:

• 符号检测:  $T_n = \sum_{i=1}^n u(X_i)$  渐近正态分布

$$\begin{cases} H_0: p=1/2 \\ H_1: p>1/2 \end{cases} \Leftrightarrow \begin{cases} H_0: X_i \sim F_0(x) \text{ 且 } F_0(0)=1/2 \\ H_1: X_i \sim F_0(x-\theta) \text{ 且 } \theta>0 \end{cases}$$

故  $\theta_0=0$ ,  $\theta_n \in \Theta_1 = \{\theta > 0\}$ 。

$$E_\theta(T_n) = E\left[\sum_{i=1}^n u(X_i)\right] = n(1 - F_0(-\theta))$$

$$\sigma_\theta^2(T_n) = nF_0(-\theta)(1 - F_0(-\theta))$$

$$\text{所以: } \varepsilon_{\text{sign}} = \frac{\left[ \frac{d[n(1-F_0(-\theta))]}{d\theta} \Big|_{\theta=0} \right]^2}{n[nF_0(-\theta)(1-F_0(-\theta)) \Big|_{\theta=0}]} = \frac{n^2 f_0^2(0)}{n^2 \cdot \frac{1}{2} \left(1 - \frac{1}{2}\right)} = 4 f_0^2(0)$$

$$\text{这里 } f_0(x) = \frac{dF_0(x)}{dx}$$

• 线性检测:  $T_n = \sum_{i=1}^n X_i$

$$\begin{cases} H_0: p=1/2 \\ H_1: p>1/2 \end{cases} \Rightarrow \begin{cases} H_0: \text{mean} = 0 \\ H_1: \text{mean} = \theta > 0 \end{cases} \quad \text{假定 } f_0(x) \text{ 对原点对称}$$

$$\text{所以: } \varepsilon_{\text{linear}} = \frac{\left[ \frac{dE\left[\sum_{i=1}^n X_i | \theta\right]}{d\theta} \right]_{\theta=0}^2}{n \cdot \text{Var}_{\theta}\left(\sum_{i=1}^n X_i\right)_{\theta=0}} = \frac{\left[ \frac{d(n\theta)}{d\theta} \right]_{\theta=0}^2}{n \cdot n\sigma_x^2} = \frac{1}{\sigma_x^2} \quad \text{这里 } \sigma_x^2 \text{ 为 } x_i \text{ 的方差。}$$

$$\text{所以: } \text{ARE}_{\text{sign,linear}} = \lim_{n \rightarrow \infty} \frac{\varepsilon_{\text{sign}}}{\varepsilon_{\text{linear}}} = 4f_0^2(0)\sigma_x^2$$

$$\text{若 } X_i \text{ 高斯分布, 则 } f_0(0) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \Rightarrow \text{ARE}_{s,1} = \frac{2}{\pi} \approx 0.64$$

$\Rightarrow$  符号检测差于线性检测 (UMP 检验)。

$$\text{若 } X_i \text{ 拉氏分布, } f_0(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \text{ 则 } f_0(0) = \frac{\lambda}{2}, \sigma_x^2 = \frac{2}{\lambda^2} \Rightarrow \text{ARE}_{s,1} = 2$$

$\Rightarrow$  符号检测优于线性检测。

#### •Student's t-test:

若  $\sigma_x^2$  未知, 则不能构成线性检测器。

所以, 对于检测具有未知方差的正态分布均值可利用本检测器:

$$S_n^* = \frac{\sqrt{n} \cdot \overline{x_n}}{S_n} \underset{H_0}{\overset{H_1}{>}} S_{n,\alpha}^* \quad (\text{门限, 大的值})$$

$$\text{其中: } \overline{x_n} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x_n})^2$$

性能: 渐近地,  $n \rightarrow \infty$ , t-test  $\rightarrow$  linear detector (具有同样的效验)。

#### •Wilcoxon 检验:

$$\text{可证明: } T^+ = \sum_{i=1}^n u(X_i)r(|X_i|) \Leftrightarrow T^+ = \sum_{i=1}^n \sum_{j=1}^i u(X_i + X_j)$$

证: 排列  $\{X_i\}$ :  $X_{i_1} < X_{i_2} < \dots < X_{i_{k-1}} < 0 < X_{i_k} < \dots < X_{i_n}$

$T^+$  为正样本秩之和, 而秩是按样本幅度排列的。

这等效于: 计算项数  $N_j$ ——在  $\{X_{i_j} + 0, X_{i_j} + X_{i_{j-1}}, \dots, X_{i_j} + X_{i_1}\}$  集合中的正项数

$$\text{并且计算 } T^+ = \sum_{j=i_k}^{i_n} N_j$$

$$\text{而 } N_j = 0, \text{ 当 } j < i_k \text{ 时 } \Rightarrow T^+ = \sum_{j=i_1}^{i_n} N_j$$

由于  $j$  的取值包含了所有的样本值，

$$\text{所以, } T^+ = \sum_{i=1}^n \sum_{j=1}^i u(X_i + X_j)$$

$$\text{故, } E(T^+) = \sum_{i=1}^n E[u(X_i)] + \sum_{i=2}^n \sum_{j=1}^{i-1} E[u(X_i + X_j)]$$

$$\text{而 } E[u(X_i)] = \int_0^{+\infty} f_0(x - \theta) dx = 1 - F_0(-\theta) = F_0(\theta)$$

$$E[u(X_i + X_j)] = P(X_i + X_j > 0) = 1 - \int_{-\infty}^{+\infty} f_0(\sigma - \theta) F_0(-\sigma - \theta) d\sigma$$

$$\text{注: } f_z(z = x + y) = \int_{-\infty}^{+\infty} f_x(z - \sigma) f_y(\sigma) d\sigma$$

因此，假定  $f_0(x)$  对原点对称（即：  $f_0(x) = f_0(-x)$ ），

$$\left. \frac{\partial E(T^+)}{\partial \theta} \right|_{\theta=0} = n f_0(0) + n(n-1)I \quad \text{其中 } I = \int_{-\infty}^{+\infty} f_0^2(\sigma) d\sigma$$

$$\text{又知: } \text{Var}(T^+) = \frac{n(n+1)(2n+1)}{24}$$

$$\text{故, 效验 } \mathcal{E}_{\text{Wilcoxon}} = \lim_{n \rightarrow \infty} \frac{[n f_0(0) + n(n-1)I]^2}{n \cdot \frac{n(n+1)(2n+1)}{24}} = 12I^2$$

$$\text{若分布为高斯型, 则 } I = \int_{-\infty}^{+\infty} \left( \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}} \right)^2 dx = \frac{1}{2\sqrt{\pi}\sigma_x} \Rightarrow \mathcal{E}_{\text{Wilcoxon}} = \frac{3}{\pi\sigma_x^2}$$

$$\text{若分布为拉氏型, 则 } I = \int_{-\infty}^{+\infty} \left( \frac{1}{\sqrt{2}\sigma_x} e^{-\frac{\sqrt{2}|x|}{\sigma_x}} \right)^2 dx = \frac{1}{2\sqrt{2}\sigma_x} \Rightarrow \mathcal{E}_{\text{Wilcoxon}} = \frac{3}{2\sigma_x^2}$$

●典型数据:

分布类型	$\text{ARE}_{\text{Wilcoxon}, t\text{-test}}$	$\text{ARE}_{\text{sign}, t\text{-test}}$	$\text{ARE}_{\text{sign}, \text{Wilcoxon}}$
均匀 (Uniform)	1	$\frac{1}{3}$	$\frac{1}{3}$
正态 (Normal)	$\frac{3}{\pi} = 0.955$	$\frac{2}{\pi} = 0.64$	$\frac{2}{3}$
拉氏 (Double exponential)	$\frac{3}{2}$	2	$\frac{4}{3}$
Logistic	$\frac{\pi^2}{9} = 1.097$	$\frac{\pi^2}{12} = 0.82$	$\frac{3}{4}$

注意到：最后一列数据可以从前两列的数据之比导出。

另外，实际上可证明： $\text{ARE}_{\text{Wilcoxon}, t\text{-test}}$  总是大于等于 0.864。所以，相对于最佳参量检测方法——线性检测器（和  $t\text{-test}$  检测器等效）而言，W 检测器是一个性能优良的鲁棒检测器。

#### 4.1.4 双输入系统 (Two-Input Systems)

令  $\mathbf{W} = ((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$

典型： $\mathbf{X} = (X_1, X_2, \dots, X_n)$  和  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$  是两个通道的输出。

应用：射电天文学、水下声检测（声纳）、地球物理等，一个公共信号在两个独立的噪声干扰信道中传输。

$$\begin{cases} H_0: \mathbf{X}, \mathbf{Y}, \text{ 独立序列, 零均值, 固定方差 } E(X_i^2)=\sigma_1^2, E(Y_i^2)=\sigma_2^2; \\ H_1: \text{ 随机信号 } \mathbf{S}=(S_1, S_2, \dots, S_n), \text{ 零均值, } \text{Var}(S_i)=\sigma_3^2, \text{ 存在于两个通道,} \\ \quad \mathbf{X}, \mathbf{Y} \text{ 的噪声分量分布如 } H_0, \text{ 且独立于 } \mathbf{S}. \end{cases}$$

考虑参量和非参量检测两种情况：

1) 所有过程是 Gaussian,  $\sigma_1=\sigma_2=\sigma$ :

$$\begin{cases} H_0: (X_i, Y_i) \text{ 联合高斯, } \rho=0, \sigma_x=\sigma_y=\sigma; \\ H_1: (X_i, Y_i) \text{ 联合高斯, } \rho=\frac{\sigma_3^2}{\sigma^2+\sigma_3^2}>0, \sigma_x^2=\sigma_y^2=\frac{\sigma^2}{1-\rho}. \end{cases}$$

$$\text{N-P 检测为: } \sum_{i=1}^n (x_i+y_i)^2 \underset{H_0}{\overset{H_1}{>}} t_1$$

2)  $X_i, Y_i$  和  $S_i$  的密度已知是对称和连续的：

$$F_1(x) = P(X_i \leq x | H_0)$$

$$F_2(y) = P(Y_i \leq y | H_0)$$

$$F_3(s) = P(S_i \leq s | H_0)$$

$$\begin{cases} H_0: F(x,y) = F_1(x) F_2(y); \\ H_1: F(x,y) = \int_{-\infty}^{+\infty} F_1(x-s) F_2(y-s) f_3(s) ds, \text{ 其中: } f_3(s) = \frac{dF_3(s)}{ds}. \end{cases}$$

对于符号检测器，N-P 检验为：

$$\sum_{i=1}^n u(x_i y_i) \underset{H_0}{\overset{H_1}{>}} t_2 \quad \text{其中 } u(\cdot) \text{ 是阶跃函数。}$$

这样的检验统计量是极性符合相关器（PCC: Polarity Coincidence Correlator）。

可证明： $\text{ARE}_{\text{PCC}, \text{参量 1}} \text{ 或 } \text{ARE}_{(2),1} = 0.202 \text{ (Gaussian)} = 3.5 \text{ (Laplace)}。$

## 4.2 局部最佳检测（Locally Optimal Detection）

在涉及到参量  $\theta$  的单边检验中，若：1) 观测分布属于指数类，或 2) 似然比为单调函数，则存在 UMP 检验。

若 1)、2) 均不满足，我们可以试图限制检验的种类，然后在该种检验中寻找最佳解（如同 UMP 无偏检验）。

在某些情况中，这样的检验可能也不存在，如：噪声中的弱信号检测（weak signals）。我们将在本节中描述一种最佳检验。

令：  $X_1, X_2, \dots, X_n$  是 i.i.d.（独立同分布）、具有绝对连续的 CDF—— $F(\cdot; \theta)$  的观测。

$$\begin{cases} H_0: \theta = \theta_0; \\ H_1: \theta > \theta_0, \text{ 这里 } \theta \text{ 接近于 } \theta_0. \end{cases}$$

设检验量  $T_n$ ，虚警  $\alpha_n$ ，误警  $\beta_n(\theta)$ ，这里的下标  $n$  表明对样本尺寸  $n$  的依赖性。

设另一检验量  $T_n^*$ ，具有相同的虚警  $\alpha_n$ ，而误警为  $\beta_n^*(\theta)$ ：

$$\left. \frac{\partial \beta_n^*(\theta)}{\partial \theta} \right|_{\theta=\theta_0} \leq \left. \frac{\partial \beta_n(\theta)}{\partial \theta} \right|_{\theta=\theta_0}$$

则  $T_n^*$  是局部最佳的。

等效地，利用 power function（检测概率，“效能函数”）： $\Phi_n(\theta) = 1 - \beta_n(\theta)$

则有：
$$\left. \frac{\partial \Phi_n^*(\theta)}{\partial \theta} \right|_{\theta=\theta_0} \geq \left. \frac{\partial \Phi_n(\theta)}{\partial \theta} \right|_{\theta=\theta_0}$$

$\Rightarrow$  若对  $T_n^*$  而言，与其它任一  $T_n$  相比较，其在  $\theta_0$  处的效能函数斜率更大的话，则在  $\theta = \theta_0 + \varepsilon$ ,  $\varepsilon > 0$ ,  $\varepsilon \rightarrow 0$  处， $T_n^*$  的效能（或检测概率）是在给定相同  $\alpha_n$  的检验中最大的。

令判断域为  $\mathbf{I}$ ，其补集为  $\mathbf{I}^*$ ，则：

$$\alpha_n = \int_{\mathbf{I}} \prod_{i=1}^n f(x_i; \theta_0) dx$$

$$\beta_n(\theta) = \int_{\mathbf{I}^*} \prod_{i=1}^n f(x_i; \theta) dx$$

$$\Phi_n(\theta) = \int_{\mathbf{I}} \prod_{i=1}^n f(x_i; \theta) dx$$

当  $f(\cdot; \theta)$  满足某些规定条件时，
$$\frac{\partial \Phi_n(\theta)}{\partial \theta} = \int_{\mathbf{I}} \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i; \theta) dx$$

$\Rightarrow$  因而，所要求解的问题等效于：

在  $\alpha_n$  一定的情况下，使 
$$\left. \frac{\partial \Phi_n(\theta)}{\partial \theta} \right|_{\theta=\theta_0}$$
 最大。

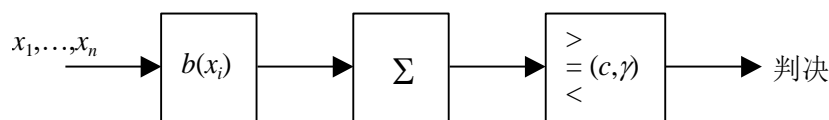
这是 N-P 问题的扩展，所以可得类似解，即：

局部最佳检验：

$$\left. \frac{\frac{\partial}{\partial \theta} \left[ \prod_{i=1}^n f(x_i; \theta) \right]}{\prod_{i=1}^n f(x_i; \theta)} \right|_{\theta=\theta_0} \begin{matrix} > c & \text{判决 } H_1 \\ = c & \text{以概率 } \gamma \text{ 判决 } H_1 \\ < c & \text{判决 } H_0 \end{matrix}$$

这里的  $c$  和  $\gamma$  ( $0 \leq \gamma \leq 1$ ) 被适当地选择以产生虚警  $\alpha_n$ 。

该检验统计量等效于：
$$\sum_{i=1}^n \left. \frac{\partial \ln f(x_i; \theta)}{\partial \theta} \right|_{\theta=\theta_0} \equiv \sum_{i=1}^n b(x_i)。$$



局部最佳检测器

性能：在  $\theta \rightarrow \theta_0$  时，局部最佳检验和最有效的 N-P 检验一样有效，即它们的 ARE 为 1。

#### 4.2.1 举例

##### ◆例 1：

若  $x_1, x_2, \dots, x_n$ ,  $x_i \sim \text{Cauchy}$ （柯西）分布（具有定位参量  $\theta$ ），

则 
$$f(x_1, x_2, \dots, x_n; \theta) = \frac{1}{\pi^n} \sum_{i=1}^n \frac{1}{1 + (x_i - \theta)^2}$$

$$\begin{cases} H_0: \theta=0 \\ H_1: \theta>0 \end{cases} \Rightarrow \text{可证明：具有虚警}\alpha\text{的 UMP 检验不存在。}$$

但可找出其局部最佳检验：

$$\frac{\partial \ln f(x_1, x_2, \dots, x_n; \theta)}{\partial \theta} = \sum_{i=1}^n \frac{2(x_i - \theta)}{1 + (x_i - \theta)^2}$$

$$\Rightarrow \sum_{i=1}^n \frac{2x_i}{1 + x_i^2} \underset{H_0}{\overset{H_1}{>}} k \quad \text{其中：} k \text{ 为满足}\alpha\text{的取值。 (连续情况下，无需}\gamma\text{，故}\gamma=0\text{)}$$

当  $n$  很大时，可依据中心极限定理证明  $\sum_{i=1}^n \frac{2x_i}{1 + x_i^2} \rightarrow \text{Gaussian 分布}$ ，故可找到满足  $\alpha$  的  $k$  值。

对强信号 ( $\theta$  很大) 而言，局部最佳解的可能是很差的。本例就是这样的情况：

$$\theta \rightarrow \infty \Rightarrow x_i \rightarrow \infty, \quad \frac{x_i}{1 + x_i^2} \rightarrow 0, \quad \text{而对于 } \alpha < 1/2, \text{ 应有 } k > 0.$$

所以， $P(\text{检验量} > k) \rightarrow 0$ ，即该检验的效能趋于零。

#### ◆例 2：

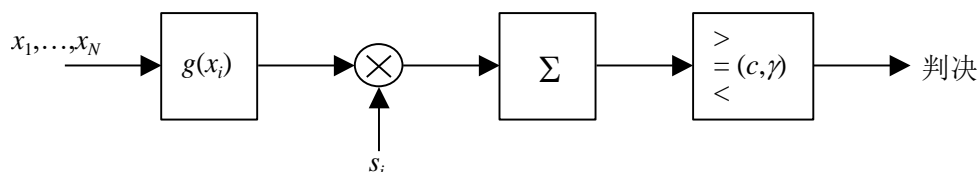
考虑在 i.i.d. (独立同分布) 噪声中 非常量 信号的局部最佳检测问题。

$$\begin{cases} H_0: X_i = N_i \\ H_1: X_i = N_i + \theta s_i \end{cases} \quad \text{这里：} N_i \sim f_n(\cdot), \quad \theta > 0, \quad \theta \rightarrow 0, \quad i=1, 2, \dots, N$$

其中： $s_i$  为非常量信号： $s_i \neq s_j$ 。

$$\text{则} \quad f(x_i | H_1) = f_n(x_i - \theta s_i)$$

$$\Rightarrow \quad \left. \frac{\partial}{\partial \theta} \ln[f(x_i | H_1)] \right|_{\theta=0} = -\frac{s_i}{f_n(x_i)} \frac{df_n(x_i)}{dx_i} = s_i \cdot g(x_i) \quad \text{其中：} g(x_i) = -\frac{1}{f_n(x_i)} \frac{df_n(x_i)}{dx_i}$$



#### 4.2.2 广义高斯噪声 (Generalized Gaussian Noise) 与局部最佳检测

$$\text{广义高斯类随机变量 } N: \quad f_N(x) = \frac{c \eta(\sigma_n, c)}{2 \Gamma(1/c)} e^{-[\eta(\sigma_n, c)|x|]^c} \quad -\infty < x < +\infty$$

$$\text{其中：} \eta(\sigma_n, c) = \sigma_n^{-1} \sqrt{\frac{\Gamma(3/c)}{\Gamma(1/c)}} \quad c > 0$$

两种尾部衰减程度不同的典型情况：

$$\begin{cases} c=1: \text{双指数函数;} \\ c=2: \text{高斯函数。} \end{cases}$$

若  $c < 1$ ，比双指数函数衰减还慢；

若  $c > 1$ ，则比高斯函数衰减还快。

对于噪声中常正量信号的检测：只有高斯密度函数，可存在 UMP 检验；但对不同的  $c$  值，可找到相应的局部最佳检验。

对于弱信号的双边检验 ( $H_0: \theta = \theta_0$ ;  $H_1: \theta \neq \theta_0, \theta \rightarrow \theta_0$ ) 问题：可找出一个局部最佳检验。



## 4.3 鲁棒检测和估值 (Robust Detection and Estimation)

参量检测：已知密度分布；

非参量检测：未知密度分布，仅假定原点对称、连续可导等性质；

鲁棒检测：密度分布属于某类  $F$ ，在标称密度（或称“名义密度”：nominal density） $f_0$  附近。

在鲁棒检测和估计中，我们试图使这类函数的最大风险极小化。这就是“Robust”的意义。

### 4.3.1 鲁棒假设检验 (休伯 Huber 模型)

令：  $(x_1, x_2, \dots, x_n)$  i.i.d. observations 具有 common density  $p$ 。

问题：检验  $p=p_1$  或  $p=p_0$ ？

这里，  $p_i = \{q \mid q = (1-\varepsilon_i)f_i + \varepsilon_i H_i, H_i \in \mathbf{H}\} \quad i=0,1$

其中：  $0 \leq \varepsilon_i < 1$ ；  $\mathbf{H}$ ：观测空间中的所有概率测量的类（class）。

通常：  $\varepsilon_i$  较小  $\Rightarrow p_i$  接近其标称密度  $f_i$ 。

令：  $R(q_i', \Phi)$ ——在  $q_i'$  密度（实际的真正密度）下检验  $\Phi(x)$  的风险，

则问题可描述为：

在  $\sup_{q_0'} R(q_0', \Phi) \leq \alpha$  条件下，找出  $\min_{\Phi} \sup_{q_1'} R(q_1', \Phi)$ 。

令  $q_0$ 、 $q_1$  是两个最不利的分布，即：

对任何检验  $\Phi(x)$  而言，  $R(q_i', \Phi) \leq R(q_i, \Phi) \quad i=0,1$

可证明：上面极大极小化问题可以通过两个最不利分布间的概率比检验（Probability Ratio Test）来解决。

所以，一旦找到了两个最不利的分布  $q_0$  和  $q_1$ ，则上述问题化为：

在  $R(q_0, \Phi) \leq \alpha$  条件下，找出  $\min_{\Phi} R(q_1, \Phi)$ 。

例如：若风险表述为错误概率，则问题就是标准的、在  $q_0$  和  $q_1$  间作判决的 Neyman-Pearson 检验。

寻找最不利分布可归结为下面的表达式：

$$q_0(x) = \begin{cases} (1-\varepsilon_0)f_0(x) & \frac{f_1(x)}{f_0(x)} < c_0 \\ \frac{1}{c_0}(1-\varepsilon_0)f_1(x) & \frac{f_1(x)}{f_0(x)} \geq c_0 \end{cases}$$
$$q_1(x) = \begin{cases} (1-\varepsilon_1)f_1(x) & \frac{f_1(x)}{f_0(x)} > c_1 \\ c_1(1-\varepsilon_1)f_0(x) & \frac{f_1(x)}{f_0(x)} \leq c_1 \end{cases}$$

其中：  $0 \leq c_1 < c_0 < \infty$ ， $c_0$  和  $c_1$  必须选为使  $q_0$  和  $q_1$  成为概率密度函数，即

$$(1-\varepsilon_0) \left\{ P\left(\frac{f_1(x)}{f_0(x)} < c_0 \mid f_0\right) + \frac{1}{c_0} P\left(\frac{f_1(x)}{f_0(x)} \geq c_0 \mid f_1\right) \right\} = 1$$

$$(1-\varepsilon_1) \left\{ P\left(\frac{f_1(x)}{f_0(x)} > c_1 \mid f_1\right) + c_1 P\left(\frac{f_1(x)}{f_0(x)} \leq c_1 \mid f_0\right) \right\} = 1$$

“最不利”含义：在各自标称函数  $f_i(x)$  占优势的区间， $q_i(x)$  相互模仿，导致错误概率最大。

$\{\varepsilon_i\}$  必须足够小以保证  $p_0$  和  $p_1$  不交叠（overlap，无公共元素），因为  $\varepsilon_1 \uparrow \Rightarrow c_1 \uparrow$ ， $\varepsilon_0 \uparrow \Rightarrow c_0 \downarrow$ 。在信号检

测的应用中，这导致了需要信号电平足够大的限制。

利用最不利分布，前面的问题就可以通过下面的概率比检验来解决：

$$\Phi^*(x) = \begin{cases} 1, & T_n(x) > k \\ \tau, & T_n(x) = k \\ 0, & T_n(x) < k \end{cases} \quad \text{其中: } T_n(x) \equiv \sum_{i=1}^n \ln \frac{q_1(x_i)}{q_0(x_i)}$$

这里：

$$\frac{q_1(x)}{q_0(x)} = \begin{cases} bc_1 & \frac{f_1(x)}{f_0(x)} \leq c_1 \\ b \frac{f_1(x)}{f_0(x)} & c_1 < \frac{f_1(x)}{f_0(x)} < c_0 \\ bc_0 & \frac{f_1(x)}{f_0(x)} \geq c_0 \end{cases} \quad \text{其中: } b = \frac{1-\varepsilon_1}{1-\varepsilon_0}$$

$\Phi^*(x)$ ——检验函数：给定  $x$ ，判决  $H_1$  的条件概率。

$T_n(x)$ ——检验统计量。

◆例：考虑 Huber 模型的特殊情况： $\varepsilon_0 = \varepsilon_1 = \varepsilon$ ,  $f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$  ——高斯,  $f_1(x) = f_0(x-\theta)$ ,  $\theta > 0$ 。

即：在标称高斯噪声中常正信号的检测问题。

$$b = \frac{1-\varepsilon_1}{1-\varepsilon_0} = 1$$

由  $c_0$ 、 $c_1$  需满足的条件得：

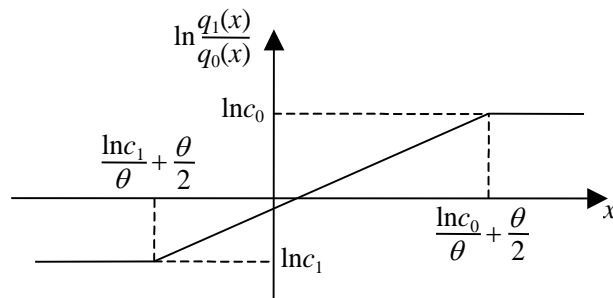
$$\Phi\left(\frac{\ln c_0}{\theta} + \frac{\theta}{2}\right) + \frac{1}{c_0} [1 - \Phi\left(\frac{\ln c_0}{\theta} - \frac{\theta}{2}\right)] = \frac{1}{1-\varepsilon}$$

$$[1 - \Phi\left(\frac{\ln c_1}{\theta} - \frac{\theta}{2}\right)] + c_1 \Phi\left(\frac{\ln c_1}{\theta} + \frac{\theta}{2}\right) = \frac{1}{1-\varepsilon}$$

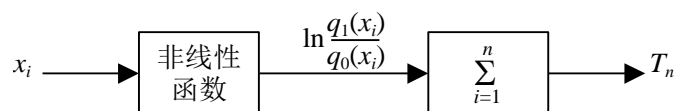
因为： $\frac{f_1(x)}{f_0(x)} \leq c \Leftrightarrow x \leq \frac{\ln c}{\theta} + \frac{\theta}{2}$ ，所以：

$$\frac{q_1(x)}{q_0(x)} = \begin{cases} c_1 & x \leq \frac{\ln c_1}{\theta} + \frac{\theta}{2} \\ e^{\theta x - \frac{\theta^2}{2}} & \frac{\ln c_1}{\theta} + \frac{\theta}{2} < x < \frac{\ln c_0}{\theta} + \frac{\theta}{2} \\ c_0 & x \geq \frac{\ln c_0}{\theta} + \frac{\theta}{2} \end{cases}$$

取其对数，可作出如下的非线性函数图：



对应的检测器结构如下图所示：



为了保证  $p_1$  和  $p_0$  不交叠（无公共元素），要求  $q_1 \neq q_0$

$\Rightarrow \theta > \theta_\varepsilon$ , 这里  $\theta_\varepsilon$  是  $2\Phi(\frac{\theta_\varepsilon}{2}) = \frac{1}{1-\varepsilon}$  的解：

$\varepsilon$	0.04	0.138
$\theta_\varepsilon$ （信号幅度）	0.11	0.40
SNR(dB)（噪声方差为 1）	-19	-8

可见：不适于低 SNR 的通信系统（如：扩频通信等）。

“鲁棒性”解释：

由于高斯拖尾衰减快，所以依据标称密度  $f_0(x)$ ，大的  $x$  很少出现。但在  $\varepsilon$ -污染模型中，污染的随意性导致了  $p_0$  的“重尾”，因而可能产生大的  $x$ 。因此，区分  $p_1$  和  $p_0$ ，大的样本值的影响就有限了。显然，对于检测高斯分布的均值，线性检测器是最佳的，但对于现在的  $p_0$ ，大的样本值却可能导致  $p_1$  的错误判决。所以，线性检测器和  $\varepsilon$ -污染模型相比就是极为敏感和不鲁棒的了。

此结果可推广到：在标称高斯噪声中非常量信号的检测。

观测模型： $X_i = \theta s_i + N_i \quad i=1,2,\dots,n$

其中： $s_i$  为确知信号； $N_i$  为 i.i.d. 的标称高斯随机变量。

$$\begin{cases} H_0: \theta=0 \\ H_1: \theta>0 \end{cases}$$

检测器与上面的类似，只是限幅电平取决于  $s_i$ 。

为达到鲁棒目的，需要一个强的信号电平  $\theta > \frac{\theta_\varepsilon}{(s_i)_{\min}}$ （ $\theta_\varepsilon$  的定义同前）。

#### 4.3.2 鲁棒估值

如同假设检验情况一样，观测样本的密度对于标称密度的偏离可能严重地影响一个密度参数估值器的质量。

例如，考虑对一个定位参数  $\theta$  的估计：

$X_1, X_2, \dots, X_n$ , i.i.d. 且  $f(x-\theta)$  为其概率密度。

估值器：样本平均  $\frac{1}{n} \sum_{i=1}^n x_i$  和 样本中值  $\text{med}\{x_1, x_2, \dots, x_n\}$ 。

若  $f(x)$  完全确定，则可获得  $\theta$  的最大似然估值（MLE）：

$$f(x) \text{ 是高斯时, } \hat{\theta}_{\text{ML}} = \frac{1}{n} \sum_{i=1}^n x_i ;$$

$$f(x) \text{ 是双指数时, } \hat{\theta}_{\text{ML}} = \text{med}\{x_1, x_2, \dots, x_n\} ;$$

但是，当甚至只有一、两个样本的取值与  $f(x)$  不相符（或很大、或很小）时，样本平均的估值质量将迅速变坏。

下面来说明（以数学推导形式）这个问题。

考虑： $\mathbf{F} = \{f | f = (1-\varepsilon)\Phi + \varepsilon h, h \in \mathbf{H}\}$

其中： $\Phi(x)$ ——标准正态分布； $\varepsilon \in (0,1)$ ； $\mathbf{H}$ ——所有对称有界密度函数类。

Huber 建议一个估值器类 (class) —— “M-estimators”:

$$\hat{\theta}(x) = \arg \left\{ \min_{\theta} \sum_{i=1}^n L(x_i - \theta) \right\} \quad \text{其中: } L(\cdot) \text{ 为确定估值器的函数。}$$

例: 样本平均对应于选择  $L(x)=x^2$  (因为: 求导  $\Rightarrow$  样本平均估值器)

MLE (最大似然估值器) 对应于选择  $L(x)=-\ln f(x)$ 。

假定:  $L$  是凸的 (convex)、原点对称、足够光滑的 (regular)。

则基于  $L$  的 M-检测器是一致的, 且有性质:  $\sqrt{n} [\hat{\theta}(x) - \theta]$  是具有零均值、方差  $V(l, f)$  的渐近高斯量。

$$\text{这里: } V(l, f) = \frac{\int l^2(x) f(x) dx}{\left[ \int l'(x) f(x) dx \right]^2} \quad \text{其中 } l = \frac{dL(x)}{dx}$$

应用 Schwarz 不等式可证明:

$$\text{对于一个固定的 } f_0 \text{ 有: } V(l, f_0) \geq V(l_0, f_0) = \frac{1}{I(f_0)}$$

$$\text{其中: } l_0(x) = -\frac{1}{f_0(x)} \frac{df_0(x)}{dx}$$

$$I(f) = \int \frac{[f'(x)]^2}{f(x)} dx \quad \text{—— “Fisher 定位信息量 (location information)”}$$

对于被污染的密度函数:

\*对应于  $l(x)=x$ , 基于标称模型  $f=\Phi$  的最佳估值器是样本平均

$$\text{此时, } V(l, f) = \int_{-\infty}^{+\infty} x^2 f(x) dx = (1-\varepsilon) + \varepsilon \int_{-\infty}^{+\infty} x^2 h(x) dx$$

由于  $\varepsilon > 0$ 、 $h$  是任意函数, 所以平均样本的渐近方差可能非常大, 故对  $\varepsilon$ -污染模型而言, 此估值器很不鲁棒。

\*对于样本中值估值器:

$$l(x) = \text{sgn}(x), \quad V(l, f) = \frac{1}{4f^2(0)} \leq \frac{1}{4(1-\varepsilon)^2 \Phi^2(0)} = \frac{\pi}{2(1-\varepsilon)^2}$$

显然, 它比样本平均更鲁棒。

但当  $f$  就是标称函数  $\Phi$  时, 其方差比样本平均的方差大  $\frac{\pi}{2} \approx 1.57 (\approx \frac{1}{0.64})$  倍。

我们希望找到一个比样本中值更好的、又具有鲁棒性质的估值器, Huber 建议用极大极小的形式 (minimax) 来设计定位估值器:

$$\min_l \max_{f \in F} V(l, f) \quad (*)$$

$$\text{在对作某些假定后, 可得: } l_R = -\frac{1}{f_R(x)} \frac{df_R(x)}{dx}$$

其中  $f_R(x)$  是由  $f_R = \arg \{ \max_{f \in F} I(f) \}$  所定义的最不利密度函数。

这样的一对  $(l_R, f_R)$  成为 (\*) 的鞍点 (saddle-point) 解。

---

对于标称正态分布：

$$l_R(x) = \begin{cases} x, & |x| < a \\ a \cdot \text{sgn}(x), & |x| \geq a \end{cases}$$

这里：  $\int_{-a}^a \Phi(x)dx + \frac{2\Phi(a)}{b} = \frac{1}{1-\varepsilon}$       其中：  $b = \frac{-\Phi'(a)}{\Phi(a)}$

因此， $l_R(x)$ 是一个软限幅器（soft limiter）。

在（\*）意义上的一个鲁棒估值器  $\hat{\theta}_R$  可得到为：

$$\sum_{i=1}^n l_R(x_i - \hat{\theta}_R) = 0$$

## 4.4 小结

- 非参量检测由来及应用
- 非参量检测方法（Sign Test, Wilcoxon Test, Two-Input Systems/Two-Channel Tests）  
    相关概念（Linear/Non-Linear detector, ARE-Asymptotic Relative Efficiency）
- 局部最佳检测（Locally Optimal Detection）  
    由来、应用及处理方法
- 鲁棒检测和估值（Robust Detection and Estimation）  
    由来、Huber Model（ $\varepsilon$ -contaminated mixtures）、

---

# 第5章 总 结

## 第 1 章 信号检测理论

- 信号检测模型及常用判决准则：

Bayes 准则（最小平均风险准则）；

极大极小准则（安全平均风险准则、Neyman-Pearson 准则、最小错误概率准则（理想观测者准则）、最大似然准则、最大后验概率准则）。

- 多元假设检验；
- 复合假设检验；
- 序列检测（WALD 检测）；
- 白高斯信道中的信号检测：

确知信号检测：连续信号的似然比检验及其最佳接收机（相关接收机/匹配滤波器）；应用：二元数字通信系统（误码率、CPSK、CFSK、CASK/OOK）和雷达系统。

随机参量信号检测：随幅、随相（正交接收机/非相干匹配滤波器）、随幅随相信号检测。

- 色高斯信道中的信号检测：

Karhunen-Loève 正交展开法/白化滤波器；

接收机性能及特点、最佳信号波形及特性。

## 第 2 章 信号估值理论

- 信号参量估值模型及常用估值准则：

Bayes 估值、各种不同代价下的估值；

极大极小估值、最大后验估值、最大似然估值。

- 估计量的性质；
- CRAMÉR-RAO 不等式：非随机参量、随机参量；
- 在高斯信道中单参量信号的估值：

白高斯信道：非随机参量估值及应用（幅度估值/相关接收机、相位估值/双通道相位测量和频率估值/匹配滤波器组）、随机参量估值及应用（幅度估值/相关接收机）；

色高斯信道：似然方程及应用（幅度估值/相关型接收机）。

- 多参量估值；
- 线性最小均方估值（LMS）及正交原理；
- 最小二乘估值（LS）和加权最小二乘估值（LSW）；
- 各类估值准则间关系总结。

## 第 3 章 波形估计理论

- 波形的线性最小均方估计；

- WIENER 滤波：

连续时间：物理不可实现 WF、物理可实现的 WF（频谱因式分解法、预白化法）；

离散时间：有限观测区间的、物理可实现的及物理不可实现的离散 WF；

- 
- 离散 Wiener 滤波矩阵。
  - KALMAN 滤波：  
状态空间模型；  
离散时间 KALMAN 滤波（估值递推直观表示、配平方法、应用示例、两组公式）；  
连续时间 KALMAN 滤波（离散化求解、Ricatti 方程及稳态解）；  
KALMAN 滤波的稳定性和发散问题的解决办法。
  - KALMAN 滤波的推广（简介）：色噪声（扩大状态向量法）、状态预测与平滑估计；
  - 非线性滤波：点估计（线性化：标称状态、预测状态/EKF）、PDF 估计（粒子滤波）；
  - 匹配滤波器

#### 第 4 章 非参量检测和 ROBUST 检测

- 非参量检测（NONPARAMETRIC DETECTION）：  
符号检测（Sign Test）；  
Wilcoxon（威尔柯克孙）检测（W 检测器）；  
渐近相对效率（Asymptotic Relative Efficiency）；  
双输入系统（Two-Input Systems）。
- 局部最佳检测（LOCALLY OPTIMAL DETECTION）、广义高斯噪声（Generalized Gaussian Noise）；
- 鲁棒检测和估值（ROBUST DETECTION AND ESTIMATION）：  
鲁棒假设检验（休伯 Huber 模型、非线性检测器）；  
鲁棒估值（软限幅器）。