最优滤波理论及其应用

段玉波 编著
刘铁男 主审

哈尔滨工程大学出版社

最优滤波理论及其应用

段玉波　编著
刘铁男　主审

哈尔滨工程大学出版社

(黑)新登字第9号

内容简介

本书是作者吸收国内外学者在最优滤波理论方面最新研究成果的基础上，结合自己多年科研和应用的实践而写成的。书中阐述了最优滤波、自适应滤波和最优反卷积滤波的理论和应用技术。

本书的使用对象是电气专业和工业自动化专业的教师、研究生、高年级大学生、科研和工程技术人员。

最优滤波理论及其应用

段玉波　编著

刘铁男　主审

责任编辑　陈晓军

*

哈尔滨工程大学出版社出版发行

哈尔滨市海鸿电脑制版中心制版

哈尔滨师范大学印刷厂印刷

*

开本 850×1168 1/32　印张 9.0625 字数 220 千字

1994年10月　第1版　1994年10月　第1次印刷

印数：1—500册

ISBN 7-81007-480-6

TP·23　　定价：12.00元

前　言

当今的时代是知识爆炸的时代。在现代科学技术飞速发展的今天，一方面，现代科学技术为科研和生产实践奠定了科学的基础；另一方面，科研和生产实践的不断进步也为新理论和技术的持续发展提供了源动力。本书的特点是力争做到理论与实践相结合，并且有关应用部分，侧重最优滤波理论在油田生产中的应用。

本书共分如下五章：

第一章介绍了估计理论的基本知识，包括常用的误差函数、估值好坏的标准、Cramer－Rao 不等式及极大后验估计、极大似然估计和最小方差估计的概念。

第二章介绍有关的数学模型：随机差分方程模型、状态空间模型和两种模型之间的关系。这些模型将用于最优滤波和自适应滤波问题。

第三章阐述了最优滤波器和平滑器理论。

第四章论述了具有重要实用价值的自适应滤波器理论，其中包括了作者的一些研究成果。

第五章讨论了滤波理论在油田地震勘探数据处理方面应用的新成果。

由于作者的水平有限，不足之处在所难免，希望读者批评指正。

编著者

1994 年 5 月

目　录

第一章 估计理论的基础

系统状态的最优滤波(或状态估计)问题属于估计理论的范畴。估计理论的另一分支——系统辨识理论,这里不赘述了,读者请参阅文献〔1〕~〔61〕。本章将介绍估计理论的一些基本概念,以及估计量的某些重要性质和经典估计方法。

§1.1 误差函数

从问题的一般提法开始,我们把待估计的量记成 θ,把 θ 的估计量(或估值)记成 $\hat{\theta}$。θ 一般的是一个向量,设其维数为 r。z 表示观测数据,一般的说,它也是一个向量,设其维数为 q。$\mathbf{Z}$ 表示观测数据的集合,它是我们对 θ 进行估计的依据。θ 的估值 $\hat{\theta}$ 一般的说是 $\mathbf{Z}$ 的函数。故可记成:$\theta=\hat{\theta}(\mathbf{Z})$。

为了具有足够的广泛性,我们把 θ 看成是随机变量。由于观测数据的随机性,故估值 $\hat{\theta}=\hat{\theta}(\mathbf{Z})$ 也是随机变量。$\hat{\theta}$ 当然与 θ 有相同的维数,估计的误差是 $\tilde{\theta}=\theta-\hat{\theta}$,它也是一个与 θ 有相同维数的随机变量。

为判断 $\hat{\theta}$ 是否可以作为 θ 的一个估计量,应该有一个判别准则。作为这种准则,自然可以用由于估计的误差而引起的某种损失来表示。这种损失的数值应该由误差 $\tilde{\theta}$ 来确定,所以我们可以引入一个误差函数:

定义 1.1 定义在 r 维欧氏空间 R^r 上的一个函数 $L(\tilde{\theta})$,如果

满足条件

1. 对一切 $\tilde{\theta}\in R^r$，皆有 $L(\tilde{\theta})\geqslant 0$；

2. 如果 $\tilde{\theta}=0$，则 $L(\tilde{\theta})=0$；

3. 对于任何的 $\tilde{\theta}\in R^r$ 皆有 $L(\tilde{\theta})=L(-\tilde{\theta})$；

4. 如果 $\|\tilde{\theta}_1\|\leqslant\|\tilde{\theta}_2\|$，$\|\cdot\|$ 是欧式范数，则 $L(\tilde{\theta}_1)\leqslant L(\tilde{\theta}_2)$。

就说，它是一个误差函数。

注意到 $\tilde{\theta}$ 是随机变量，所以 $L(\tilde{\theta})$ 也是一个随机变量。从统计的观点来看，作为估计的准则，不应该直接取 $L(\tilde{\theta})$，而应该取它的统计均值，即数学期望。故我们有：

定义 1.2 设 θ 是待估计向量，$\hat{\theta}$ 是它的估值，$\tilde{\theta}$ 是估计的误差，$L(\tilde{\theta})$ 是一个误差函数，则称 $B(\tilde{\theta})=E\{L(\tilde{\theta})\}$ 为估计 θ 的贝叶斯(Bayes)风险。

$L(\tilde{\theta})$ 常见的形式是：

1. 平方误差函数

$$L(\tilde{\theta}) = \|\tilde{\theta}\|_S^2 = \tilde{\theta}' S\tilde{\theta}$$

其中 S 是一个正定的对称矩阵

2. 绝对误差函数

$$L(\tilde{\theta}) = \|\tilde{\theta}\|$$

3. 一致误差函数

$$L(\tilde{\theta})\begin{cases}0 & \text{当 } \|\tilde{\theta}\| < \dfrac{C}{2}\\[2ex] 1 & \text{当 } \|\tilde{\theta}\| \geqslant \dfrac{C}{2}\end{cases}$$

其中 C 是一个正的常数。

§1.2 估值好坏的标准

我们当然希望所得到的关于 θ 的估值 $\hat{\theta}$ 尽可能的好。因此，必

须首先给出“好”的标准是什么

1. 无偏性

θ 的估值 $\hat{\theta}=\hat{\theta}(Z)$ 说是无偏的，当 θ 不是随机变量时，如果

$$E\{\hat{\theta}\}=\theta$$

如果被估计的量 θ 是一个随机变量，$\hat{\theta}$ 是它的一个估计，我们说这个估值是无偏的，若

$$E\{\hat{\theta}\}=E\{\theta\} \tag{1.2}$$

2. 一致性

θ 的估计量 $\hat{\theta}_n$ 说是一致的，如果对于每个 $\varepsilon>0$，皆有

$$\lim_{n\to\infty}P\{\|\hat{\theta}_n-\theta\|\geqslant\varepsilon\}=0 \tag{1.3}$$

或者等价的

$$\lim_{n\to\infty}P\{\|\hat{\theta}_n-\theta\|<\varepsilon\}=1 \tag{1.3}$$

估计量 $\hat{\theta}_n$ 说是均方一致的，如果

$$\lim_{n\to\infty}E\{(\hat{\theta}_n-\theta)(\hat{\theta}-\theta)^{\tau}\}=0$$

估计量 $\hat{\theta}_n$ 说是强一致的，如果

$$P\{\|\hat{\theta}_n-\theta\|\geqslant\varepsilon\}=0$$

或等价的

$$P\{\|\hat{\theta}_n-\theta\|<\varepsilon\}=1$$

3. 有效性

设 $\hat{\theta}$ 及 $\hat{\theta}'$ 都是 θ 的无偏估值，如果

$$\mathrm{Var}\{\hat{\theta}\}\leqslant\mathrm{Var}\{\hat{\theta}'\}$$

就说 $\hat{\theta}$ 较 θ' 有效；如果当样本容量 n 固定时，$\mathrm{Var}\{\hat{\theta}\}$ 达到了最小值，就说 $\hat{\theta}$ 是 θ 的有效估计。

例 设 X 是一个随机变量，$x_1,x_2,\cdots,x_n$ 是 X 的一个子样。

$\bar{x}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$是$E\{x\}$的无偏估计,对于$n$个非负数$C_i$,使 $\sum_{i=l}^{n} C_i = 1$,$x_n = \sum_{k=1}^{n} C_k x_k$也是$E\{x\}$的一个无偏估计。但我们有

$$\left(\sum_{k=l}^{n} C_k\right)^2 \leqslant n \sum_{k=l}^{n} C_k^2$$

所以

$$\begin{aligned}
\mathrm{Var}\{\bar{x}_n\} &= \frac{1}{n}\mathrm{Var}\{x\} \\
&= \frac{\mathrm{Var}\{x\}}{n}\left(\sum_{k=l}^{n} C_k\right)^2 \leqslant \mathrm{Var}\{x\}\left(\sum_{k=l}^{n} C_k^2\right) \\
&= \mathrm{Var}\{\dot{x}_n\}
\end{aligned}$$

故$\bar{x}_n$较一切$\sum_{k=l}^{n} C_k x_k$为有效。

§1.3 Cramer—Rao 不等式

可以说明,无论是那一种估计,其误差方差皆是下方有界的。这种界限关系通常被称为克拉美--罗(Cramer--Rao)不等式。以下,我们就无偏估计的情形给出这个不等式。

我们对参量$\boldsymbol{x}$进行估计,观测模型是

$$\boldsymbol{z}_k = h_k(\boldsymbol{x}, \boldsymbol{v}_k) \qquad k = 1, 2, \cdots$$

$\mathbf{Z}_N = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_n\}$表示观测数据集。集中$\boldsymbol{v}_k$是具有已知分布的噪声。从而条件密度$P(\mathbf{Z}_N|\boldsymbol{x})$是能够被确定的。我们有:

定理(Cramer－Rao 不等式) 如果$\hat{\boldsymbol{x}}$是$\boldsymbol{x}$是基于观测数据$\mathbf{Z}_N$的任一个无偏估计量,假定$\frac{\partial}{\partial x}[P(\mathbf{Z}_N|\boldsymbol{x})]$和$\frac{\partial}{\partial x}\left[\frac{\partial}{\partial x}P(\mathbf{Z}_N|\boldsymbol{x})\right]^T$皆存在,作为$\mathbf{Z}_N$的函数关于$\boldsymbol{x}$一致绝对可积,则$\hat{\boldsymbol{x}}$在$\boldsymbol{x}$给定下的条件误差是下方有界的,其下界是费歇(Fisher)信息矩阵J的逆(当然要假定这个逆存在),即

$$E\{(\boldsymbol{x}-\hat{\boldsymbol{x}})(\boldsymbol{x}-\hat{\boldsymbol{x}})^{\tau}|x\}\geqslant J^{-1} \tag{1.4}$$

其中

$$J=E\left\{\left[\frac{\partial}{\partial x}\ln P(\mathbf{Z}_N|\boldsymbol{x})\right]\left[\frac{\partial}{\partial x}\ln P(\mathbf{Z}_N|\boldsymbol{x})\right]^{\tau}|\boldsymbol{x}\right\} \tag{1.5}$$

或者

$$J=-E\left\{\frac{\partial}{\partial x}[\frac{\partial}{\partial x}\ln P(\mathbf{Z}_N|\boldsymbol{x})^{\tau}|\boldsymbol{x}\right\} \tag{1.6}$$

在(1.4)式中等号成立当且仅当

$$\frac{\partial}{\partial x}\ln P(\mathbf{Z}_N|\boldsymbol{x})=K(\boldsymbol{x})(\boldsymbol{x}-\boldsymbol{x}) \tag{1.7}$$

证明:参见文献[13]。

§1.4 极大后验估计

设 θ 是待估向量,$\mathbf{Z}$ 是观测数据集合。在对 θ 进行具体观测之前,人们对 θ 根据过去的经验积累了一些知识。用数学形式把这种知识整理出来,就可以提出一个模式:虽然 θ 的具体值未知,但它服从一概率分布。其概率密度 $P(\theta)$ 称为“先验密度”。当得到观测 $\mathbf{Z}$ 后,可对 θ 的先验知识进行调整,给出在条件 $\mathbf{Z}$ 下的 θ 的概率密度 $P(\theta|\mathbf{Z})$ 称为后验密度。而把在 θ 的条件下观测 $\mathbf{Z}$ 的概率密度 $P(\mathbf{Z}|\theta)$ 称为似然密度。

考虑一致误差函数

$$L(\tilde{\theta})=\begin{cases}0 & \|\tilde{\theta}\|<\dfrac{C}{2}\\[2ex] 1 & \|\tilde{\theta}\|\geqslant\dfrac{C}{2}\end{cases} \tag{1.8}$$

使得相应的贝叶斯(Bayes)风险取最小的估值 $\hat{\theta}$ 称为一致误差估记,记作 $\hat{\theta}_{unf}$。因为

$$E\{L(\tilde{\theta})\}=E\{E\{L(\tilde{\theta})|\mathbf{Z}\}\}$$

$$
\begin{aligned}
&= \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} L(\boldsymbol{x}) P(\boldsymbol{x}|z) \mathrm{d}\boldsymbol{x} \right\} P(\mathbf{Z}) \mathrm{d}\mathbf{Z} \\
&= \int_{-\infty}^{+\infty} \left\{ 1 - \int_{\| \boldsymbol{x} - \theta_{unf} \| < \frac{C}{2}} P(\boldsymbol{x}|\mathbf{Z}) \mathrm{d}\boldsymbol{x} \right\} P(\mathbf{Z}) \mathrm{d}\mathbf{Z}
\end{aligned} \tag{1.9}
$$

由此可见，$P(\boldsymbol{x}|\mathbf{Z})$愈大，$E\{L(\tilde{\theta})\}$愈小。为使(Bayes)风险 $E\{L(\tilde{\theta})\}$最小，只须 $P(\boldsymbol{x}|\mathbf{Z})$最大。为明确起见，把 $P(\boldsymbol{x}|\mathbf{Z})$仍记成 $P(\theta|\mathbf{Z})$。所以选取 θ 估计 $\hat{\theta}$ 使 $P(\hat{\theta}|\mathbf{Z}) = \max P(\theta|\mathbf{Z})$是有明显的意义的。这样的估值 $\hat{\theta}$ 称为 θ 的极大后验估值，记作 θ_{MAP}。显然 θ_{MAP} 必满足

$$
\frac{\partial P(\theta|\mathbf{Z})}{\partial\theta} = 0 \tag{1.10}
$$

$\hat{\theta}$ 是极大后验估值的充分条件是

$$
\frac{\partial^2 P(\theta|\mathbf{Z})}{\partial\theta^2} < 0 \tag{1.11}
$$

注意到

$$
P(\theta|\mathbf{Z}) = \frac{P(\theta|\mathbf{Z})P(\theta)}{P(\mathbf{Z})}
$$

所以

$$
\begin{aligned}
\ln P(\theta|\mathbf{Z}) &= \ln P(\mathbf{Z}|\theta) + \ln P(\theta) - \ln P(\mathbf{Z}) \\
&= \ln P(\mathbf{Z},\theta) - \ln P(\mathbf{Z})
\end{aligned}
$$

但 $P(\mathbf{Z})$中不含 θ，故 $P(\theta|\mathbf{Z})$最大相当于 $P(\mathbf{Z},\theta)$最大。从而极大后验(MAP)估值能由下式求出。

$$
\frac{\partial \ln P(\mathbf{Z}|\theta)}{\partial\theta} + \frac{\partial \ln P(\theta)}{\partial\theta} = 0 \tag{1.12}
$$

或者

$$
\frac{\partial P(\mathbf{Z},\theta)}{\partial\theta} = 0 \tag{1.13}
$$

§1.5 极大似然估计

应用极大后验估计方法，必须事先知道密度函数 $P(\mathbf{Z}|\theta)$和 $P(\theta)$或者等价的要求知道联合分布密度 $P(\mathbf{Z},\theta)$。如果仅仅知道 $P(\mathbf{Z}|\theta)$，上述的估计方法就不能应用了。此时，由于 $P(\theta)$未知，从而 θ 的统计特性可看成是未加任何限制，即它的统计特性具有某种任意性。比如说我们可以考虑 θ 具有均匀分布的情形，此时必有

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

和

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

回忆估值 θ_{AMP}应满足方程

$$\frac{\partial \ln P(\mathbf{Z}|\theta)}{\partial\theta} + \frac{\partial \ln P(\theta)}{\partial\theta} = 0$$

于是，我们有

$$\frac{\partial \ln P(\mathbf{Z}|\theta)}{\partial\theta} = 0 \tag{1.14}$$

方程(1.14)也作为寻求 θ 的估值的一种手段。

我们说 θ 的估值 $\hat{\theta}$ 是它的极大似然估值，并记作 $\hat{\theta}_{ML}$，如果它满足

$$\ln P(\mathbf{Z}|\hat{\theta}_{ML}) = \max \ln P(\mathbf{Z}|\theta) \tag{1.15}$$

或者等价的它满足

$$P(\mathbf{Z}|\hat{\theta}_{ML}) = \max P(\mathbf{Z}|\theta) \tag{1.16}$$

此时，称 $P(\mathbf{Z}|\theta)$或 $\ln P(\mathbf{Z}|\theta)$为似然函数。

因为 $P(\mathbf{Z}|\theta)\mathrm{d}z$ 可理解为参数值是 θ 时，观测值 z 落在 $\mathbf{Z}$ 的一个小邻域中的概率。确定 θ 的估值的信息来自于 $\mathbf{Z}$。于是估值 $\hat{\theta}$ 愈近于真值，$P(\mathbf{Z}|\theta)$之值就应该愈大。所以，极大似然估值 $\hat{\theta}_{ML}$，取作

使 $P(\mathbf{Z}|\theta)$ 达到最大的 θ 值，是有道理的。

§1.6 最小方差估计

我们假设 θ 是系统的待估计向量，$\hat{\theta}=\hat{\theta}(\mathbf{Z})$ 为依据观测 $\mathbf{Z}$ 对 θ 的某种估计，它是和 θ 同维的向量函数，其自变量为向量 $\mathbf{Z}$。记这个估计的误差为

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

要衡量一个估计量的优劣，应该研究这个估计误差的整个统计规律。直观的想法是希望 $\hat{\theta}(\mathbf{Z})$ 与 θ 越接近越好。按统计规律而言，不管对 θ 重复多少次量测，用同样的估计方法，由每次量测所得到的估计的误差，大部分应当密集在零的附近，这才能认为这一估计方法是好的。我们知道，估计误差 $\tilde{\theta}$ 二阶原点矩 $E(\tilde{\theta}\tilde{\theta})^{\tau}$（称为均方误差阵）正是表示误差分布在零附近密集的程度。所以，如果存在某种估计，能使其均方误差阵比其它任何估计的均方误差阵都小，我们就可以说，这个估计是最优的。于是在这个意义上的所谓最优估计问题，就是录求一个适当的函数 $\hat{\theta}(\mathbf{Z})$，使 $E(\tilde{\theta}\tilde{\theta})$ 达到最小。

记 θ 的概率密度为 $P_1(\theta)$，$(\mathbf{Z})$ 的概率密度为 $P_2(\mathbf{Z})$，θ 和 $\mathbf{Z}$ 的联合密度为 $P(\theta,\mathbf{Z})$，在已知 θ 时 $\mathbf{Z}$ 的条件密度为 $P(\mathbf{Z}|\theta)$，在已知 $\mathbf{Z}$ 时 θ 的条件密度为 $P(\theta|\mathbf{Z})$。由贝叶斯公式有

$$P(\theta,\mathbf{Z})=P(\theta|\mathbf{Z})P_2(\mathbf{Z})=P(\mathbf{Z}|\theta)P_1(\theta) \qquad (1.17)$$

对某一估计量 $\hat{\theta}(\mathbf{Z})$，其均方误差阵 $E(\tilde{\theta}\tilde{\theta}^{\tau})$ 可以表示为

$$\begin{aligned}E(\tilde{\theta}\tilde{\theta})^{\tau}&=E[\theta-\hat{\theta}(\mathbf{Z})][\theta-\hat{\theta}(\mathbf{Z})]^{\tau}\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\theta-\hat{\theta}(\mathbf{Z})][\theta-\hat{\theta}(\mathbf{Z})]P(\theta,\mathbf{Z})\mathrm{d}\theta\mathrm{d}z\\&=\int_{-\infty}^{\infty}[\int_{-\infty}^{\infty}[\theta-\hat{\theta}(\mathbf{Z})][\theta\\&\quad-\hat{\theta}(\mathbf{Z})]^{\tau}P(\theta|Z)]\mathrm{d}\theta]P_2(z)\mathrm{d}z\end{aligned} \qquad (1.18)$$

(1.18)式对任意 $\hat{\theta}(\mathbf{Z})$ 都成立，我们的目的是选一个估计量 $\hat{\theta}_{MV}(\mathbf{Z})$，使(1.18)达最小。由(1.18)式最后一个式易知，只需对 $\hat{\theta}(Z)$ 求下式的最小值即可。

$$\int_{-\infty}^{\infty}[\theta-\hat{\theta}(\mathbf{Z})][\theta-\hat{\theta}(\mathbf{Z})]^{\tau}P(\theta|\mathbf{Z})\mathrm{d}\theta \tag{1.19}$$

下面来证明，使(1.19)式达到最小的 $\hat{\theta}(\mathbf{Z})$，就是在给定 $\mathbf{Z}$ 时，θ 的条件均值，即

$$\hat{\theta}_{MV}(\mathbf{Z})=E(\theta|\mathbf{Z}) \tag{1.20}$$

事实上，设 $\hat{\theta}(\mathbf{Z})$ 是 $\mathbf{Z}$ 的任一向量函数，则从(1.20)式出发，可以推出关系式

$$\begin{aligned}
&\int_{-\infty}^{\infty}[\theta-\hat{\theta}(\mathbf{Z})][\theta-\hat{\theta}(\mathbf{Z})]^{\tau}P(\theta|\mathbf{Z})\mathrm{d}\theta\\
=&\int_{-\infty}^{\infty}[\theta-E(\theta|\mathbf{Z})+E(\theta|\mathbf{Z})-\hat{\theta}(\mathbf{Z})][\theta-E(\theta|\mathbf{Z})\\
&+E(\theta|\mathbf{Z})-\hat{\theta}(\mathbf{Z})]^{\tau}P(\theta|\mathbf{Z})\mathrm{d}\theta\\
=&\int_{-\infty}^{\infty}[\theta-E(\theta|\mathbf{Z})][\theta-E(\theta|\mathbf{Z})]^{\tau}P(\theta|\mathbf{Z})\mathrm{d}\theta\\
&+[E(\theta|\mathbf{Z})-\hat{\theta}(\mathbf{Z})][E(\theta|\mathbf{Z})-\hat{\theta}(\mathbf{Z})]^{\tau}\int_{-\infty}^{\infty}P(\theta|\mathbf{Z})\mathrm{d}\theta\\
&+\int_{-\infty}^{\infty}[\theta-E(\theta|\mathbf{Z})]P(\theta|\mathbf{Z})\mathrm{d}\theta\cdot[E(\theta|\mathbf{Z})-\hat{\theta}(\mathbf{Z})]^{\tau}\\
&+[E(\theta|\mathbf{Z})-\hat{\theta}(\mathbf{Z})]\int_{-\infty}^{\infty}[\theta-E(\theta|\mathbf{Z})]^{\tau}P(\theta|\mathbf{Z})\mathrm{d}\theta\\
=&\int_{-\infty}^{\infty}[\theta-E(\theta|\mathbf{Z})][\theta-E(\theta|\mathbf{Z})]^{\tau}P(\theta|\mathbf{Z})\mathrm{d}\theta\\
&+[E(\theta|\mathbf{Z})-\hat{\theta}(\mathbf{Z})][E(\theta|\mathbf{Z})-\hat{\theta}(\mathbf{Z})]^{\tau}
\end{aligned} \tag{1.21}$$

(1.21)式右边第二项是一个非负定矩阵，所以由(1.21)立即得到不等式

$$\int_{-\infty}^{\infty}[\theta-\hat{\theta}(\mathbf{Z})][\theta-\hat{\theta}(\mathbf{Z})]^{\tau}P(\theta|\mathbf{Z})\mathrm{d}\theta$$

$$\geqslant \int_{-\infty}^{\infty} [\theta - E(\theta|\mathbf{Z})][\theta - E(\theta|\mathbf{Z})]^{\mathrm{T}} P(\theta|\mathbf{Z})\mathrm{d}\theta \qquad (1.22)$$

而且(1.22)式等号成立的充分必要条件是 $\hat{\theta}(\mathbf{Z})=E(\theta\ \mathbf{Z})$，这就证明(1.20)式。

由(1.20)可得

$$\begin{aligned} E[\hat{\theta}_{MV}(\mathbf{Z})] &= \int_{-\infty}^{\infty} E(\theta|\mathbf{Z}) p_2(\mathbf{Z})\mathrm{d}\mathbf{Z} \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta P(\theta|\mathbf{Z})\mathrm{d}\theta\right] P_2(\mathbf{Z})\mathrm{d}z \\ &= \int_{-\infty}^{\infty} \theta \left[\int_{-\infty}^{\infty} P(\theta,\mathbf{Z})\mathrm{d}\mathbf{Z}\right]\mathrm{d}\theta \\ &= \int_{-\infty}^{\infty} \theta P_1(\theta)\mathrm{d}\theta = E(\theta) \qquad (1.23) \end{aligned}$$

即估值 $\hat{\theta}_{MV}(\mathbf{Z})$ 是无偏的。这时最小均方误差阵 $E[\theta-\hat{\theta}_{MV}(\mathbf{Z})][\theta-\hat{\theta}_{MV}(\mathbf{Z})]^{\mathrm{T}}$ 就是估计误差 $\theta-\hat{\theta}_{MV}(\mathbf{Z})$ 的方差阵 $\mathrm{Var}(\theta-\hat{\theta}_{MV}(\mathbf{Z}))$。称 $\hat{\theta}_{MV}(\mathbf{Z})$ 为 θ 的(基于观测 $\mathbf{Z}$)最小方差估计，$\mathrm{Var}(\theta-\hat{\theta}_{MV}(\mathbf{Z})$ 称为估计误差的最小方差阵。由于有(1.20)式成立，有时又称 $\hat{\theta}_{MV}(\mathbf{Z})$ 为 θ 的条件均值估计。

如果 θ 是观测 $\mathbf{Z}$ 的线性函数，则 θ 的最小方差估计称为线性最小方差估计，关于线性最小方差估计的定义和性质将在第三章讨论。