信号估值理论

赵丰

2018 年 4 月 13 日

1. Bayes 估值

   代价函数 $C(\hat{\theta},\theta)=C(\tilde{\theta})$, 其中 $\tilde{\theta}=\hat{\theta}-\theta$。平均风险

   $$\bar{R}=\int_{\mathcal{Z}}\int_{\Theta}C(\tilde{\theta})d\theta dz$$

   记 $\bar{R}(z)=\int_{\Theta}C(\tilde{\theta})p(\theta|z)d\theta$, 称为"给定 z 时的条件风险"。则 $\bar{R}=\int_{\mathcal{Z}}\bar{R}(z)p(z)dz$。Bayes 估值等价于求解 $\underset{\hat{\theta}}{\operatorname{argmin}}\,\bar{R}(z)$

2. 常见的代价函数

   (1) 平方误差代价：$C(\tilde{\theta})=\tilde{\theta}^2,\hat{\theta}_{\mathrm{ms}}=\int_{\Theta}\theta p(\theta|z)d\theta=\mathbb{E}[\theta|z],\hat{\theta}_{\mathrm{ms}}$ 被称为后验均值。

   (2) 绝对误差代价: $C(\tilde{\theta})=|\tilde{\theta}|,\hat{\theta}_{\mathrm{abs}}$ 满足 $\int_{-\infty}^{\hat{\theta}_{\mathrm{abs}}}p(\theta|z)d\theta=\int_{\hat{\theta}_{\mathrm{abs}}}^{+\infty}p(\theta|z)d\theta$, $\hat{\theta}_{\mathrm{abs}}$ 被称为后验中值。

   (3) 均匀误差代价：

   $$C(\tilde{\theta})=\begin{cases}0, & |\tilde{\theta}|\le\frac{\triangle}{2}\\ 1, & |\tilde{\theta}|>\frac{\triangle}{2}\end{cases}$$

   此时的条件风险为：

   $$\bar{R}(z)=1-\int_{\hat{\theta}-\frac{\triangle}{2}}^{\hat{\theta}+\frac{\triangle}{2}}p(\theta|z)d\theta$$

   当 $\triangle$ 很小时，使得条件风险 $\bar{R}(z)$ 最小有：$\hat{\theta}_{\mathrm{unf}}\approx\hat{\theta}_{\mathrm{map}}$，相当于后验概率密度的众数。

   (4) 对称下凸代价

   若代价函数满足

(a) $C(\tilde{\theta}) = C(-\tilde{\theta})$

(b) $C(b\theta_1 + (1-b)\theta_2) \le bC(\theta_1) + (1-b)C(\theta_2), \forall b \in (0,1)$

并且后验密度函数满足

(c) $p(\theta|z) = p(2\hat{\theta}_{\text{ms}} - \theta|z)$

则 $\hat{\theta}_{\text{ms}} = \underset{\hat{\theta}}{\text{argmin}}\, \bar{R}(z)$

证明.

$$\begin{aligned}
\bar{R}(z) &= \int_{\mathbb{R}} C(\theta - \hat{\theta}) p(\theta|z) d\theta &(1)\\
&= \int_{\mathbb{R}} C(\theta - \hat{\theta}) p(2\hat{\theta}_{\text{ms}} - \theta|z) d\theta\\
&= \int_{\mathbb{R}} C(2\hat{\theta}_{\text{ms}} - \theta - \hat{\theta}) p(\theta|z) d\theta\\
&= \int_{\mathbb{R}} C(-2\hat{\theta}_{\text{ms}} + \theta + \hat{\theta}) p(\theta|z) d\theta &(2)
\end{aligned}$$

对 (1) 和 (2) 作平均得:

$$\begin{aligned}
\bar{R}(z) &= \int_{\mathbb{R}} \frac{1}{2}\left(C(\theta - \hat{\theta}) + C(-2\hat{\theta}_{\text{ms}} + \theta + \hat{\theta})\right) p(\theta|z) d\theta\\
&\ge \int_{\mathbb{R}} C(\theta - \hat{\theta}_{\text{ms}}) p(\theta|z) d\theta
\end{aligned}$$

上式右端和 $\hat{\theta}$ 无关，当 $\theta - \hat{\theta} = -2\hat{\theta}_{\text{ms}} + \theta + \hat{\theta}$ 时不等式成立，即 $\hat{\theta} = \hat{\theta}_{\text{ms}}$ □

(5) 对称非下凸代价

3. 常见估值方式

   (a) 极大极小估值：先验未知，寻找先验概率 $p(\theta)$ 的最不利分布，再确定 Bayes 估值。

   (b) 最大后验估值：代价函数未知，求解最大后验方程，

$$\left[\frac{\partial \ln p(z|\theta)}{\partial \theta} + \frac{\partial p(\theta)}{\partial \theta}\right]\Bigg|_{\theta = \hat{\theta}_{\text{map}}} = 0 \tag{3}$$

   (c) 最大似然估值：先验和代价函数均未知 $\hat{\theta}_{\text{ML}} = \max_{\theta} p(z|\theta)$

4. 估计量的性质

(a) 无偏估计量

对于随机参量 θ, 无偏性要求 $\mathbb{E}[\hat{\theta}]=\mathbb{E}[\theta]$ 即 $\int_{\Theta}\int_{\mathcal{Z}}\hat{\theta}p(z,\theta)dzd\theta=\int_{\mathcal{Z}}\hat{\theta}p(z)dz$

(b) 一致估计量

基于 N 个观测值组成的观测矢量 $\boldsymbol{z}_N$ 作估值：$\hat{\theta}(\boldsymbol{z}_N)$ 是参数 θ 的一致估计量，当其满足

$$\lim_{N\to\infty}\Pr(|\hat{\theta}(\boldsymbol{z}_N)-\theta|<\epsilon)=1,\forall\epsilon>0$$

(c) 充分统计量:$\hat{\theta}(z)$ 是参数 θ 的充分统计量,当 $p(z|\theta)=p(\hat{\theta}(z)|\theta)h(z)$

(d) 有效 (efficient) 统计量：具有最小方差的无偏估计量。

5. C-R 不等式

	非随机	随机
	$\mathbb{E}[(\hat{\theta}-\theta)^2\|\theta]\geq 1/\mathbb{E}\left[\left(\frac{\partial\ln p(z\|\theta)}{\partial\theta}\right)^2\|\theta\right]$	$\mathbb{E}[(\hat{\theta}-\theta)^2]\geq 1/\mathbb{E}\left[\left(\frac{\partial\ln p(z,\theta)}{\partial\theta}\right)^2\right]$
等价形式	$\mathbb{E}\left[\left(\frac{\partial\ln p(z\|\theta)}{\partial\theta}\right)^2\|\theta\right]=-\mathbb{E}\left[\frac{\partial^2\ln p(z\|\theta)}{\partial\theta^2}\right]$	$\mathbb{E}\left[\left(\frac{\partial\ln p(z,\theta)}{\partial\theta}\right)^2\right]=-\mathbb{E}\left[\frac{\partial^2\ln p(z,\theta)}{\partial\theta^2}\right]$
存在条件	$\frac{\partial p(z\|\theta)}{\partial\theta}=(\hat{\theta}-\theta)\cdot K(\theta)$	$\frac{\partial p(z,\theta)}{\partial\theta}=(\hat{\theta}-\theta)\cdot K$
有效估值存在	有效估值就是最大似然估值	有效估值就是 $\hat{\theta}_{\mathrm{ms}}$ 或 $\hat{\theta}_{\mathrm{map}}$

表 1: 随机参量和非随机参量的对比

在随机参量的情况下，有 $\frac{\partial p(z,\theta)}{\partial\theta}=\frac{\partial p(\theta|z)}{\partial\theta}$。若有效估值存在，则后验概率密度满足方程 $\frac{\partial p(z,\theta)}{\partial\theta}=(\hat{\theta}-\theta)\cdot K$,从而解出 $p(\theta|z)=D\exp\left(-\frac{K}{2}(\theta-\hat{\theta})^2\right)$,即后验分布是高斯分布。反之，若后验分布是高斯分布，也满足有效估值存在性方程。所以对于随机参量，有效估值存在当且仅当后验分布是高斯分布。

6. 白高斯信道中单参量信号的估值

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

(1) θ 是未知的非随机参量时，求 $\hat{\theta}_{\mathrm{ML}}$

采用 K-L 展开求极限的方法,

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

$$\mathbb{E}[z_k]=s_k(\theta)$$
$$\mathrm{Var}[z_k]=\mathrm{Var}[n_k]=\frac{N_0}{2}$$
$$\mathrm{Cov}(z_k,z_l)=\frac{N_0}{2}\delta_{kl}$$

对观测矢量 $\boldsymbol{z}_N=(z_1,z_2,\ldots,z_N)$ 而言，联合概率密度函数为：

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

因此 $\dfrac{\partial\ln p(\boldsymbol{z}_N|\theta)}{\partial\theta}=\dfrac{2}{N_0}\displaystyle\sum_{k=1}^{N}(z_k-s_k(\theta))\dfrac{\partial s_k(\theta)}{\partial\theta}$ 取极限得：

$$\frac{\partial p(z(t)|\theta)}{\partial\theta}=\lim_{N\to\infty}\frac{\partial\ln p(\boldsymbol{z}|\theta)}{\partial\theta}=\frac{2}{N_0}\sum_{k=1}^{\infty}(z_k-s_k(\theta))\frac{\partial s_k(\theta)}{\partial\theta}$$

又根据 z_k 和 $s_k(\theta)$ 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：

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

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

$$\frac{\partial p(z(t)|\theta)}{\partial\theta}=\frac{2}{N_0}\int_0^T(z(t)-s(t,\theta))\frac{\partial s(t,\theta)}{\partial\theta}dt \tag{4}$$

由此可得最大似然方程：

$$\left.\int_0^T(z(t)-s(t,\theta))\frac{\partial s(t,\theta)}{\partial\theta}dt\right|_{\theta=\hat{\theta}_{\mathrm{ML}}}=0 \tag{5}$$

对上式求导可得 C-R 界为 $\frac{N_0/2}{\int_0^T\left(\frac{\partial s(t,\theta)}{\partial\theta}\right)^2}$

(2) 信号幅度的估值

此时，信号 $s(t,A)=As(t)$，其中 A 是待估值的非随机参量。由(4)式等于 0 得到：

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

不妨令 $\int_0^T s^2(t)dt=1$ （信号能量归一化），易验证 $\hat{A}_{\mathrm{ML}}$ 具有性质：

(a) 无偏性

(b) 方差为 $\frac{N_0}{2}$, 等于 C-R 界，因此是有效估值。

(3) 信号相位的估值

此时，信号 $s(t,\theta)=A\sin(\omega_0 t+\theta), 0\le t\le T$，其中 θ 是待估值的非随机参量，而 A 和 ω_0 是已知常数，假定 ωT 是 π 的整数倍。由(4)式等于 0 得到：

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

当信噪比足够大时，$n(t)/A$ 是小量，将上式中 arctan 在 $\tan(\theta)$ 处 Taylor 展开，得到

$$\hat{\theta}_{\mathrm{ML}}\approx\theta+\frac{2}{AT}\int_0^T\cos(\omega t+\theta)n(t)dt$$

因此 $\hat{\theta}_{\mathrm{ML}}$ 近似为均值为 θ（无偏），方差为 $\frac{N_0}{A^2T}$（有效）的高斯分布。

(4) 信号频率的估值

此时，信号 $s(t,\omega)=A\sin(\omega t+\theta), 0\le t\le T$，其中 ω 是待估值的非随机参量，而 A 是已知常数，θ 是杂散参量。可以求出信号频率的最大似然估值为

$$\hat{\omega}_{\mathrm{ML}}=\underset{\omega}{\operatorname{argmin}}\left(\int_0^T z(t)\cos\omega t dt\right)^2+\left(\int_0^T z(t)\sin\omega t dt\right)^2$$

(5) θ 为已知先验概率的随机参量时，求 $\hat{\theta}_{\mathrm{map}}$

首先，我们通过最大后验方程(3)式来求其估值：第一项已经由前面(4)求出，最大后验方程可重写为

$$\left[\frac{2}{N_0}\int_0^T(z(t)-s(t,\theta))\frac{\partial s(t,\theta)}{\partial\theta}dt+\frac{\partial p(\theta)}{\partial\theta}\right]\Bigg|_{\theta=\hat{\theta}_{\mathrm{map}}}=0 \qquad (6)$$

在（2）中，若 $A\sim N(0,\sigma_A^2)$，根据(6) 可得信号的幅度估值为

$$\hat{A}_{\mathrm{map}}=\frac{\int_0^T z(t)s(t)dt}{\int_0^T s^2(t)dt+\frac{N_0}{2\sigma_A^2}}$$

$\hat{A}_{\mathrm{ML}}$ 具有性质：

(a) 无偏性：$\mathbb{E}[A]=0, \mathbb{E}[\hat{A}_{\text{map}}]=0$

(b) 利用随机参数的 C-R 界的公式,$\mathbb{E}[(\hat{A}_{\text{map}}-A)^2]$ 为 $1/[\frac{2}{N_0}\int_0^T s^2(t)dt+\frac{1}{\sigma_A^2}]$ 等于 C-R 界，因此是有效估值。

7. 色高斯信道非随机参量的估值

观测方程 $z(t)=s(t,\theta)+n(t), 0\le t\le T$, 其中 $n(t)$ 为色高斯平稳噪声，其相关函数记为 $R_n(t-\tau)$。沿用信号检测理论中的 K-L 展开法，$z(t)=\sum_{k=1}^{\infty}z_k g_k(t)$, 系数 $\{z_k\}$ 互不相关，且 $z_k\sim N(s_k,\lambda_k)$。似然函数为：

$$p(z(t))=\prod_{k=1}^{\infty}\frac{1}{\sqrt{2\pi\lambda_k}}\exp\left(-\frac{(z_k-s_k)^2}{2\lambda_k}\right)$$

取其对数形式，有：

$$\ln p(z(t))=C-\sum_{k=1}^{\infty}\frac{z_k^2}{2\lambda_k}+\sum_{k=1}^{\infty}\frac{z_k s_k}{\lambda_k}-\sum_{k=1}^{\infty}\frac{s_k^2}{2\lambda_k} \tag{7}$$

其中 C 为与信号参量不相关的部分项。而在信号估计理论一节已经推导出

$$\sum_{k=1}^{\infty}\frac{z_k s_k}{\lambda_k}=\int_0^T z(t)h(t)dt \tag{8a}$$

$$\sum_{k=1}^{\infty}\frac{s_k^2}{2\lambda_k}=\frac{1}{2}\int_0^T s(t,\theta)h(t)dt \tag{8b}$$

其中 $h(t)=\displaystyle\sum_{k=1}^{\infty}\frac{s_k g_k(t)}{\lambda_k}$，为将 $h(t)$ 表示成积分的形式，引入反核函数 $R_n^{-1}(t-\tau)$，相对于 $R_n(t-\tau)$，我们有

$$R_n(t-\tau)=\sum_{k=1}^{\infty}\lambda_k g_k(t)g_k(\tau) \tag{9a}$$

$$R_n^{-1}(t-\tau)=\sum_{k=1}^{\infty}\frac{1}{\lambda_k}g_k(t)g_k(\tau) \tag{9b}$$

满足 $\int_0^T R_n^{-1}(t-\tau)R_n(\tau-\mu)d\tau=\delta(t-u)$ 针对(9b), 两边乘以 $s(\tau,\theta)$ 对 τ 积分得：$h(t)=\int_0^T s(\tau,\theta)R_n^{-1}(t-\tau)d\tau$ 利用反核函数的表达式：

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{z_k^2}{2\lambda_k}&=\frac{1}{2}\int_0^T\int_0^T z(t)z(\tau)\sum_{k=1}^{\infty}\frac{g_k(t)g_k(\tau)}{\lambda_k}dtd\tau\\&=\frac{1}{2}\int_0^T\int_0^T z(t)z(\tau)R_n^{-1}(t,\tau)dtd\tau\end{aligned}$$

因此(7)化为:

$$
\begin{aligned}
\ln p(z(t)) &= C - \frac{1}{2}\int_0^T \int_0^T [z(t)z(\tau) - 2z(t)s(\tau,\theta) + s(t,\theta)s(\tau,\theta)]R_n^{-1}(t-\tau)dtd\tau \\
&= C - \frac{1}{2}\int_0^T \int_0^T [z(t) - s(t,\theta)]R_n^{-1}(t-\tau)[z(\tau) - s(\tau,\theta)]dtd\tau \\
\frac{\partial \ln p(z(t)|\theta)}{\partial \theta} &= \int_0^T \int_0^T [z(t) - s(t,\theta)]R_n^{-1}(t-\tau)\frac{\partial s(\tau,\theta)}{\partial \theta}dtd\tau \\
&= \int_0^T (z(t) - s(t,\theta))\frac{\partial h(t,\theta)}{\partial \theta}dt
\end{aligned}
$$

利用极大似然法可以写出似然方程为

$$
\int_0^T (z(t) - s(t,\theta))\frac{\partial h(t,\theta)}{\partial \theta}dt\Big|_{\theta=\hat{\theta}_{\mathrm{ML}}} = 0 \tag{10}
$$

对比(5)式和(10)式可以发现,色高斯信道的 ML 估计中求偏导用 $h(t,\theta)$ 代替了 $s(t,\theta)$。当 $\lambda_k = \frac{N_0}{2}$ 时, $R_n(t-\tau) = \frac{N_0}{2}\delta(t-\tau), R_n^{-1}(t-\tau) = \frac{2}{N_0}\delta(t-\tau), h(t,\theta) = \frac{2}{N_0}s(t,\theta)$

另外不难求出估值的 C-R 界为 $1/\int_0^T \frac{\partial s(t,\theta)}{\partial \theta}\frac{\partial h(t,\theta)}{\partial \theta}dt$

例:信号振幅估计 $s(t,A) = As(t)$,由(10)式解得:

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

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

8. 多参量估值:可由单参数估值形式推广

9. 线性最小均方估值

   放宽对信号已知知识的要求,假设

   (a) 对信号的统计知识只知道信号参量 θ 和观测值 z 的一、二阶矩:$\mathbb{E}[\boldsymbol{\theta}], \mathbb{E}[\boldsymbol{z}], \mathrm{Var}[\boldsymbol{\theta}], \mathrm{Var}[\boldsymbol{z}], \mathrm{Cov}[\boldsymbol{\theta}, \boldsymbol{z}]$

   (b) 估值是观测值的线性组合 $\hat{\boldsymbol{\theta}} = \boldsymbol{A}\boldsymbol{z} + \boldsymbol{b}$

   MSE 估值准则为:$\min_{\boldsymbol{A},\boldsymbol{b}} \mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]$, 其中 $\mathbb{E}$ 是关于 $\boldsymbol{\theta}, \boldsymbol{z}$ 联合分布的期望。

首先说明最小均方估值是无偏的，即 $\mathbb{E}[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}]=\mathbf{0}$，反设 $\mathbb{E}[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}]=\boldsymbol{b}\neq\mathbf{0}$ 构造 $\hat{\boldsymbol{\theta}}'=\hat{\boldsymbol{\theta}}-\boldsymbol{b}$, 有：

$$
\begin{aligned}
\mathbb{E}[(\hat{\boldsymbol{\theta}}'-\boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}}'-\boldsymbol{\theta})] &= \mathbb{E}[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})]-\boldsymbol{b}^T\boldsymbol{b} \\
&< \mathbb{E}[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})]
\end{aligned}
$$

因此 $b=\mathbb{E}[\boldsymbol{\theta}]-\boldsymbol{A}\mathbb{E}[\boldsymbol{z}]$ 其次说明最小均方估值的误差与观测 $\boldsymbol{z}$ 正交，反设 $\mathbb{E}[(\hat{\boldsymbol{\theta}}_{\text{LMS}}-\boldsymbol{\theta})\boldsymbol{z}^T]=\boldsymbol{\Lambda}_{\boldsymbol{ez}}\neq\mathbf{0}_{M\times N}$，其中 $\boldsymbol{\theta}$ 是 M 维向量，$\boldsymbol{z}$ 是 N 维向量。构造 $\hat{\boldsymbol{\theta}}'=\hat{\boldsymbol{\theta}}-\boldsymbol{\Lambda}_{\boldsymbol{ez}}\text{Var}^{-1}[\boldsymbol{z}](\boldsymbol{z}-\mathbb{E}[\boldsymbol{z}])$,（不妨设 $M=1$，否则分别考虑 $\boldsymbol{\theta}$ 的各分量）有：

$$
\begin{aligned}
\mathbb{E}[(\hat{\boldsymbol{\theta}}'-\boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}}'-\boldsymbol{\theta})] &= \mathbb{E}[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})]-\boldsymbol{\Lambda}_{\boldsymbol{ez}}\text{Var}^{-1}[y]\boldsymbol{\Lambda}_{\boldsymbol{ez}} \\
&< \mathbb{E}[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})]
\end{aligned}
$$

因此 $\hat{\boldsymbol{\theta}}_{\text{LMS}}-\boldsymbol{\theta}$ 与 $\boldsymbol{z}-\mathbb{E}[\boldsymbol{z}]$ 正交，已有 $\hat{\boldsymbol{\theta}}_{\text{LMS}}=\boldsymbol{A}(\boldsymbol{z}-\mathbb{E}[\boldsymbol{z}])+\mathbb{E}[\boldsymbol{\theta}]$ 。由 $\mathbb{E}[(\hat{\boldsymbol{\theta}}_{\text{LMS}}-\boldsymbol{\theta})(\boldsymbol{z}-\mathbb{E}[\boldsymbol{z}])^T]=\mathbf{0}$ 得 $\boldsymbol{A}=\text{Cov}[\boldsymbol{\theta},\boldsymbol{z}]\text{Var}^{-1}[\boldsymbol{z}]$ 因此

$$
\hat{\boldsymbol{\theta}}_{\text{LMS}}=\text{Cov}[\boldsymbol{\theta},\boldsymbol{z}]\text{Var}^{-1}[\boldsymbol{z}](\boldsymbol{z}-\mathbb{E}[\boldsymbol{z}])+\mathbb{E}[\boldsymbol{\theta}] \tag{11}
$$

线性最小均方估值的误差矩阵为：

$$
\begin{aligned}
\text{Var}[\boldsymbol{\theta}_{\text{LMS}}-\boldsymbol{\theta}] &= \mathbb{E}[(\boldsymbol{\theta}_{\text{LMS}}-\boldsymbol{\theta})(\boldsymbol{\theta}_{\text{LMS}}-\boldsymbol{\theta})^T] \\
&= \text{Var}[\boldsymbol{\theta}]-\text{Cov}[\boldsymbol{\theta},\boldsymbol{z}]\text{Var}^{-1}[\boldsymbol{z}]\text{Cov}[\boldsymbol{z},\boldsymbol{\theta}]
\end{aligned}
$$

10. 最小二乘估值：仅有观测，无先验

假定观测模型为：

$$
\boldsymbol{z}=\boldsymbol{C}\boldsymbol{\theta}+\boldsymbol{n} \tag{12}
$$

其中 $\boldsymbol{n}$ 是观测噪声，$\boldsymbol{\theta}$ 是 M 维向量，$\boldsymbol{z}$ 是 N 维向量，$N>M$。最小二乘估值准则为：

$$
\min_{\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}_{\text{LS}}}(\boldsymbol{z}-\boldsymbol{C}\boldsymbol{\theta})^T(\boldsymbol{z}-\boldsymbol{C}\boldsymbol{\theta}) \tag{13}
$$

解得

$$
\hat{\boldsymbol{\theta}}_{\text{LS}}=(\boldsymbol{C}^T\boldsymbol{C})^{-1}\boldsymbol{C}^T\boldsymbol{z} \tag{14a}
$$

$$
=\boldsymbol{\theta}+(\boldsymbol{C}^T\boldsymbol{C})^{-1}\boldsymbol{C}^T\boldsymbol{n} \tag{14b}
$$

因为 $\mathbb{E}[\boldsymbol{n}]=\mathbf{0}$，所以最小二乘估值对给定的观测模型是无偏估计。最小二乘估值的方差矩阵为

$$
\text{Var}[\boldsymbol{\theta}_{\text{LS}}-\boldsymbol{\theta}]=(\boldsymbol{C}^T\boldsymbol{C})^{-1}\boldsymbol{C}^T\text{Var}[\boldsymbol{n}]\boldsymbol{C}(\boldsymbol{C}^T\boldsymbol{C})^{-1}
$$

11. 加权最小二乘估值 (LSW) 对观测模型 $\boldsymbol{z} = \boldsymbol{C\theta} + \boldsymbol{n}$ 引入新的目标函数 $(\boldsymbol{z} - \boldsymbol{C\theta})^T \boldsymbol{W} (\boldsymbol{z} - \boldsymbol{C\theta})$，其中 $\boldsymbol{W}$ 是对称正定的加权矩阵，加权最小二乘估值准则可描述为：

$$\min_{\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\mathrm{LSW}}} (\boldsymbol{z} - \boldsymbol{C\theta})^T \boldsymbol{W} (\boldsymbol{z} - \boldsymbol{C\theta})$$

从而得到加权最小二乘估值为：$\boldsymbol{\theta}_{\mathrm{LSW}} = (\boldsymbol{C}^T \boldsymbol{W} \boldsymbol{C})^{-1} \boldsymbol{C}^T \boldsymbol{W} \boldsymbol{z}$ 估值的方差矩阵为

$$\mathrm{Var}[\boldsymbol{\theta}_{\mathrm{LSW}} - \boldsymbol{\theta}] = (\boldsymbol{C}^T \boldsymbol{W} \boldsymbol{C})^{-1} \boldsymbol{C}^T \boldsymbol{W} \mathrm{Var}[\boldsymbol{n}] \boldsymbol{W} \boldsymbol{C} (\boldsymbol{C}^T \boldsymbol{W} \boldsymbol{C})^{-1}$$

下面求方差矩阵的下限：

设 $\mathrm{Var}[\boldsymbol{n}] = \boldsymbol{H}^T \boldsymbol{H}, \boldsymbol{B} = \boldsymbol{H} \boldsymbol{W} \boldsymbol{C} (\boldsymbol{C}^T \boldsymbol{W} \boldsymbol{C})^{-1}$，则 $\mathrm{Var}[\boldsymbol{\theta}_{\mathrm{LSW}} - \boldsymbol{\theta}] = \boldsymbol{B}^T \boldsymbol{B}$，取 $\boldsymbol{A} = \boldsymbol{C}^T H^{-1}$，则 $\boldsymbol{AB} = \boldsymbol{I}$

使用矩阵不等式

$$\boldsymbol{B}^T \boldsymbol{B} \succcurlyeq (\boldsymbol{AB})^T (\boldsymbol{A}\boldsymbol{A}^T)^{-1} (\boldsymbol{AB}) \tag{15}$$

证明. 设 $\boldsymbol{x}$ 是与 $\boldsymbol{B}$ 的列数相同的列向量。只需证明 $\boldsymbol{x}^T \boldsymbol{B}^T \boldsymbol{B} \boldsymbol{x} \geq \boldsymbol{x}^T (\boldsymbol{AB})^T (\boldsymbol{A}\boldsymbol{A}^T)^{-1} (\boldsymbol{AB}) \boldsymbol{x}$, 则只需证明 $\boldsymbol{I} - \boldsymbol{A}^T (\boldsymbol{A}\boldsymbol{A}^T)^{-1} \boldsymbol{A}$ 半正定。设 $\boldsymbol{y}$ 是 $\boldsymbol{A}^T (\boldsymbol{A}\boldsymbol{A}^T)^{-1} \boldsymbol{A}$ 的特征向量，特征值为 σ，则 $\boldsymbol{A}^T (\boldsymbol{A}\boldsymbol{A}^T)^{-1} \boldsymbol{A} \boldsymbol{y} = \sigma \boldsymbol{y}$，等式两端左乘以 $\boldsymbol{A} \Rightarrow \boldsymbol{Ay} = \sigma \boldsymbol{Ay}$，若 $\boldsymbol{Ay} \neq \boldsymbol{0} \Rightarrow \sigma = 1$, 否则 $\sigma = 0$，因此 $\boldsymbol{A}^T (\boldsymbol{A}\boldsymbol{A}^T)^{-1} \boldsymbol{A}$ 的特征值非 0 则 1，半正定的结论成立。 □

对于给定的下限问题，由不等式(15)直接得到下限为 $(\boldsymbol{A}\boldsymbol{A}^T)^{-1}$，即

$$\mathrm{Var}[\boldsymbol{\theta}_{\mathrm{LSW}} - \boldsymbol{\theta}] \succcurlyeq (\boldsymbol{C}^T \mathrm{Var}[\boldsymbol{n}] \boldsymbol{C})^{-1}$$

直接验证可得 $\boldsymbol{W} = \mathrm{Var}^{-1}[\boldsymbol{n}]$ 时取等号，此时称 $\boldsymbol{W}$ 为最佳加权矩阵。