# 最小均方误差估计理论

**Minimum Mean Square Error (MMSE) Estimation**

统计信号处理笔记

2025 年 5 月 24 日

## 目录



## 1 引言与动机

在统计学和信号处理中，我们经常面临从噪声观测中估计未知参数或信号的问题。最小均方误差（MMSE）估计是一种基础且重要的估计准则，它在理论上具有最优性，在实践中具有广泛的应用价值。

**注记 1.1: MMSE的重要性**

MMSE估计不仅提供了理论上的最优解，还为许多实际算法（如Wiener滤波、Kalman滤波等）提供了理论基础。

## 2 基本概念与定义

**定义 2.1: MMSE估计器**

设$\boldsymbol{X} \in \mathbb{R}^n$是未知随机向量，$\boldsymbol{Y} \in \mathbb{R}^m$是观测向量。MMSE估计器$\hat{\boldsymbol{X}}_{\text{MMSE}}$定义为最小化均方误差的估计器：

$$\hat{\boldsymbol{X}}_{\text{MMSE}} = \underset{\hat{\boldsymbol{X}}(\boldsymbol{Y})}{\operatorname{argmin}}\, \mathbb{E}\left[\|\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Y})\|^2\right] \tag{1}$$

其中期望运算是关于联合分布$p(\boldsymbol{X}, \boldsymbol{Y})$进行的，$\|\cdot\|$表示欧几里得范数。

**定义 2.2: 均方误差**

对于任意估计器$\hat{\boldsymbol{X}}(\boldsymbol{Y})$，其均方误差定义为：

$$\text{MSE}(\hat{\boldsymbol{X}}) = \mathbb{E}\left[\|\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Y})\|^2\right] = \mathbb{E}\left[\sum_{i=1}^{n}(X_i - \hat{X}_i(\boldsymbol{Y}))^2\right] \tag{2}$$

## 3 核心理论结果

### 3.1 MMSE估计器的最优性

**定理 3.1: MMSE估计器的解析形式**

MMSE估计器由条件期望给出：

$$\hat{\boldsymbol{X}}_{\text{MMSE}}(\boldsymbol{Y}) = \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}] \tag{3}$$

此时的最小均方误差为：

$$\text{MMSE} = \mathbb{E}\left[\|\boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}]\|^2\right] \tag{4}$$

*证明.* 对于任意估计器$\hat{\boldsymbol{X}}(\boldsymbol{Y})$，将估计误差分解为：

$$\boldsymbol{X} - \hat{\boldsymbol{X}} = \underbrace{(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}])}_{\boldsymbol{e}_{\text{irr}}} + \underbrace{(\mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}] - \hat{\boldsymbol{X}})}_{\boldsymbol{e}_{\text{red}}} \tag{5}$$

其中$\boldsymbol{e}_{\text{irr}}$为不可约误差，$\boldsymbol{e}_{\text{red}}$为可约误差。

计算均方误差：

$$\mathbb{E}[\|\boldsymbol{X} - \hat{\boldsymbol{X}}\|^2] = \mathbb{E}[\|\boldsymbol{e}_{\text{irr}} + \boldsymbol{e}_{\text{red}}\|^2] \tag{6}$$

$$= \mathbb{E}[\|\boldsymbol{e}_{\text{irr}}\|^2] + \mathbb{E}[\|\boldsymbol{e}_{\text{red}}\|^2] + 2\mathbb{E}[\boldsymbol{e}_{\text{irr}}^T \boldsymbol{e}_{\text{red}}] \tag{7}$$

关键是证明交叉项为零。利用条件期望的性质：

$$\begin{aligned}
\mathbb{E}[\boldsymbol{e}_{\text{irr}}^T \boldsymbol{e}_{\text{red}}] &= \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}])^T (\mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}] - \hat{\boldsymbol{X}})\right] && (8)\\
&= \mathbb{E}\left[\mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}])^T (\mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}] - \hat{\boldsymbol{X}}) \mid \boldsymbol{Y}]\right] && (9)\\
&= \mathbb{E}\left[(\mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}] - \hat{\boldsymbol{X}})^T \mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}]) \mid \boldsymbol{Y}]\right] && (10)\\
&= \mathbb{E}\left[(\mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}] - \hat{\boldsymbol{X}})^T \cdot \mathbf{0}\right] = 0 && (11)
\end{aligned}$$

因此：

$$\mathbb{E}[\|\boldsymbol{X} - \hat{\boldsymbol{X}}\|^2] = \underbrace{\mathbb{E}[\|\boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}]\|^2]}_{\text{不可约误差}} + \underbrace{\mathbb{E}[\|\mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}] - \hat{\boldsymbol{X}}\|^2]}_{\geq 0} \tag{12}$$

显然，当$\hat{\boldsymbol{X}} = \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}]$时，第二项为零，此时MSE达到最小值。 □

### 3.2 正交性原理

**定理 3.2: 正交性原理**

MMSE估计误差$\boldsymbol{e} = \boldsymbol{X} - \hat{\boldsymbol{X}}_{\text{MMSE}}$与观测$\boldsymbol{Y}$的任意函数正交：

$$\mathbb{E}[\boldsymbol{e}\, g(\boldsymbol{Y})^T] = \mathbf{0} \quad \forall \text{可测函数 } g(\cdot) \tag{13}$$

特别地，$\mathbb{E}[\boldsymbol{e} \mid \boldsymbol{Y}] = \mathbf{0}$。

证明. 由于$\boldsymbol{e} = \boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}]$，利用条件期望的性质：

$$\begin{aligned}
\mathbb{E}[\boldsymbol{e}\, g(\boldsymbol{Y})^T] &= \mathbb{E}\left[\mathbb{E}[\boldsymbol{e}\, g(\boldsymbol{Y})^T \mid \boldsymbol{Y}]\right] && (14)\\
&= \mathbb{E}\left[g(\boldsymbol{Y})^T \mathbb{E}[\boldsymbol{e} \mid \boldsymbol{Y}]\right] && (15)\\
&= \mathbb{E}\left[g(\boldsymbol{Y})^T \mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}]) \mid \boldsymbol{Y}]\right] && (16)\\
&= \mathbb{E}\left[g(\boldsymbol{Y})^T (\mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}] - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}])\right] && (17)\\
&= \mathbb{E}\left[g(\boldsymbol{Y})^T \cdot \mathbf{0}\right] = \mathbf{0} && (18)
\end{aligned}$$

□

**注记 3.1: 正交性的几何解释**

正交性原理表明，MMSE估计误差在由观测$\boldsymbol{Y}$的所有函数张成的空间中的投影为零，这是最优估计的几何特征。

## 4 高斯情形下的MMSE

## 4.1 联合高斯分布的MMSE

**定理 4.1: 高斯向量的MMSE估计**

若$\boldsymbol{X}$和$\boldsymbol{Y}$服从联合高斯分布：

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right) \tag{19}$$

则MMSE估计器为线性函数：

$$\hat{\boldsymbol{X}}_{\text{MMSE}} = \boldsymbol{\mu}_X + \Sigma_{XY}\Sigma_{YY}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_Y) \tag{20}$$

对应的最小均方误差为：

$$\text{MMSE} = \text{tr}(\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}) \tag{21}$$

证明. 对于联合高斯分布，条件分布$\boldsymbol{X} \mid \boldsymbol{Y}$也是高斯分布：

$$\boldsymbol{X} \mid \boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{X|Y}, \Sigma_{X|Y}) \tag{22}$$

其中：

$$\boldsymbol{\mu}_{X|Y} = \boldsymbol{\mu}_X + \Sigma_{XY}\Sigma_{YY}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_Y) \tag{23}$$

$$\Sigma_{X|Y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \tag{24}$$

由于$\mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}] = \boldsymbol{\mu}_{X|Y}$，所以MMSE估计器即为上述线性形式。

最小均方误差为：

$$\text{MMSE} = \mathbb{E}[\|\boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}]\|^2] = \text{tr}(\text{Var}(\boldsymbol{X} \mid \boldsymbol{Y})) = \text{tr}(\Sigma_{X|Y}) \tag{25}$$

□

## 4.2 线性MMSE (LMMSE)估计

**定义 4.1: 线性MMSE估计器**

当估计器限制为线性形式$\hat{\boldsymbol{X}} = A\boldsymbol{Y} + \boldsymbol{b}$时，线性MMSE（LMMSE）估计器定义为：

$$\hat{\boldsymbol{X}}_{\text{LMMSE}} = \underset{A,\boldsymbol{b}}{\text{argmin}}\, \mathbb{E}\left[\|\boldsymbol{X} - A\boldsymbol{Y} - \boldsymbol{b}\|^2\right] \tag{26}$$

**定理 4.2: LMMSE估计器的解析解**

LMMSE估计器为：

$$\hat{\boldsymbol{X}}_{\text{LMMSE}} = \boldsymbol{\mu}_X + \Sigma_{XY}\Sigma_{YY}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_Y) \tag{27}$$

$$A_{\text{opt}} = \Sigma_{XY}\Sigma_{YY}^{-1} \tag{28}$$

$$\boldsymbol{b}_{\text{opt}} = \boldsymbol{\mu}_X - A_{\text{opt}}\boldsymbol{\mu}_Y \tag{29}$$

证明. 将MSE展开：

$$\begin{aligned}&\mathbb{E}[\|\boldsymbol{X}-A\boldsymbol{Y}-\boldsymbol{b}\|^2] &(30)\\ &=\mathbb{E}[\|\boldsymbol{X}-\boldsymbol{\mu}_X-A(\boldsymbol{Y}-\boldsymbol{\mu}_Y)-(\boldsymbol{b}+A\boldsymbol{\mu}_Y-\boldsymbol{\mu}_X)\|^2] &(31)\end{aligned}$$

令$\tilde{\boldsymbol{X}}=\boldsymbol{X}-\boldsymbol{\mu}_X$，$\tilde{\boldsymbol{Y}}=\boldsymbol{Y}-\boldsymbol{\mu}_Y$，$\boldsymbol{c}=\boldsymbol{b}+A\boldsymbol{\mu}_Y-\boldsymbol{\mu}_X$，则：

$$\text{MSE}=\mathbb{E}[\|\tilde{\boldsymbol{X}}-A\tilde{\boldsymbol{Y}}-\boldsymbol{c}\|^2] \tag{32}$$

对$\boldsymbol{c}$求偏导并令其为零：

$$\frac{\partial}{\partial\boldsymbol{c}}\text{MSE}=-2\mathbb{E}[\tilde{\boldsymbol{X}}-A\tilde{\boldsymbol{Y}}]=0 \tag{33}$$

由于$\mathbb{E}[\tilde{\boldsymbol{X}}]=\mathbb{E}[\tilde{\boldsymbol{Y}}]=\boldsymbol{0}$，得$\boldsymbol{c}_{\text{opt}}=\boldsymbol{0}$。

对$A$求偏导并令其为零：

$$\frac{\partial}{\partial A}\text{MSE}=-2\mathbb{E}[\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Y}}^T]+2A\mathbb{E}[\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^T]=0 \tag{34}$$

因此：

$$A_{\text{opt}}=\mathbb{E}[\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Y}}^T](\mathbb{E}[\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^T])^{-1}=\Sigma_{XY}\Sigma_{YY}^{-1} \tag{35}$$

□

**定理 4.3: 高斯情形下MMSE与LMMSE的等价性**

当$\boldsymbol{X}$和$\boldsymbol{Y}$联合高斯分布时：

$$\hat{\boldsymbol{X}}_{\text{MMSE}}=\hat{\boldsymbol{X}}_{\text{LMMSE}} \tag{36}$$

## 5 实际应用示例

**例题 5.1: 一维标量估计**

考虑标量情形：$X\sim\mathcal{N}(0,\sigma_X^2)$，观测模型为$Y=X+N$，其中$N\sim\mathcal{N}(0,\sigma_N^2)$且$X\perp N$。求MMSE估计器及其性能。

**解：** 首先计算必要的统计量：

$$\begin{aligned}\mu_X&=0,\quad \mu_Y=0 &(37)\\ \Sigma_{XX}&=\sigma_X^2 &(38)\\ \Sigma_{YY}&=\sigma_X^2+\sigma_N^2 &(39)\\ \Sigma_{XY}&=\text{Cov}(X,Y)=\text{Cov}(X,X+N)=\sigma_X^2 &(40)\end{aligned}$$

MMSE估计器为：

$$\hat{X}_{\text{MMSE}}=\frac{\sigma_X^2}{\sigma_X^2+\sigma_N^2}Y=\frac{\text{SNR}}{1+\text{SNR}}Y \tag{41}$$

其中$\text{SNR}=\sigma_X^2/\sigma_N^2$为信噪比。

最小均方误差为：

$$\text{MMSE}=\sigma_X^2-\frac{(\sigma_X^2)^2}{\sigma_X^2+\sigma_N^2}=\frac{\sigma_X^2\sigma_N^2}{\sigma_X^2+\sigma_N^2} \tag{42}$$

**例题 5.2: 向量Wiener滤波**

考虑向量观测模型：

$$\boldsymbol{Y} = H\boldsymbol{X} + \boldsymbol{N} \tag{43}$$

其中$H \in \mathbb{R}^{m\times n}$为已知矩阵，$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \Sigma_X)$，$\boldsymbol{N} \sim \mathcal{N}(\boldsymbol{0}, \Sigma_N)$，且$\boldsymbol{X} \perp \boldsymbol{N}$。

**解：** 计算协方差矩阵：

$$\Sigma_{XY} = \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{Cov}(\boldsymbol{X}, H\boldsymbol{X} + \boldsymbol{N}) = \Sigma_X H^T \tag{44}$$

$$\Sigma_{YY} = \mathrm{Var}(\boldsymbol{Y}) = H\Sigma_X H^T + \Sigma_N \tag{45}$$

MMSE估计器为：

$$\hat{\boldsymbol{X}}_{\text{MMSE}} = \Sigma_X H^T (H\Sigma_X H^T + \Sigma_N)^{-1}\boldsymbol{Y} \tag{46}$$

这正是著名的Wiener滤波器公式。

# 6 重要性质与扩展

## 6.1 MMSE的不变性质

**定理 6.1: 仿射变换不变性**

若$\boldsymbol{Z} = A\boldsymbol{X} + \boldsymbol{b}$，则：

$$\hat{\boldsymbol{Z}}_{\text{MMSE}} = A\hat{\boldsymbol{X}}_{\text{MMSE}} + \boldsymbol{b} \tag{47}$$

## 6.2 贝叶斯观点

从贝叶斯角度看，MMSE估计器等价于在平方损失函数下的贝叶斯估计器：

$$\hat{\boldsymbol{X}}_{\text{Bayes}} = \underset{\hat{\boldsymbol{X}}}{\operatorname{argmin}}\, \mathbb{E}_{p(\boldsymbol{X}|\boldsymbol{Y})}[L(\boldsymbol{X}, \hat{\boldsymbol{X}})] \tag{48}$$

其中$L(\boldsymbol{X}, \hat{\boldsymbol{X}}) = \|\boldsymbol{X} - \hat{\boldsymbol{X}}\|^2$。

# 7 计算复杂度与实现

在实际应用中，MMSE估计器的计算复杂度主要来自：

1. 协方差矩阵的计算和存储：$O(n^2)$空间复杂度
2. 矩阵求逆运算：$O(m^3)$时间复杂度（其中$m$为观测维度）
3. 矩阵乘法运算：$O(nm^2)$时间复杂度

对于大规模问题，可采用以下优化策略：

- 利用矩阵的特殊结构（如对角、带状等）
- 使用迭代算法（如共轭梯度法）
- 采用低秩近似技术
- 利用递推算法（如Kalman滤波）

## 8 总结

MMSE估计理论为统计估计提供了坚实的理论基础。其主要优点包括：

- 理论最优性：在均方误差意义下达到最小值
- 解析解：在高斯情形下具有闭式解
- 几何直观：基于正交性原理的清晰几何解释
- 广泛应用：为许多实际算法提供理论支撑

然而，MMSE估计也有其局限性：

- 需要完整的统计信息（联合分布或二阶矩）
- 计算复杂度较高，特别是高维情形
- 对模型假设敏感，鲁棒性有限

在实际应用中，需要根据具体问题的特点，在估计性能、计算复杂度和鲁棒性之间进行权衡。