线性贝叶斯估计问题

杨文　　　电子信息学院
教学实验大楼十楼1008室
E-mail: yw@eis.whu.edu.cn

贝叶斯估计

- 最小均方差估计（MMSE）总是后验概率密度的均值（后验均值）
- 最大后验概率估计（MAP）总是后验概率密度的峰值（后验峰值）
- 当后验概率密度函数为对称于其后验均值的单峰密度函数时，MMSE估计或MAP估计是众多类型代价函数的最佳估计量。

估计量的不变性

由于代价函数的选择往往带有某些主观的武断性，如能证明在一定条件下最佳估计量与所选定的代价函数无关是很有意义的。

性质1：若代价函数$C(\varepsilon)$是对称的凸U函数，且后验概率密度函数$p(\theta|x)$对称于其后验均值，即

1）$C(\varepsilon)=C(-\varepsilon)$（对称性）

2）$C(b\varepsilon_1+(1-b)\varepsilon_2)\le bC(\varepsilon_1)+(1-b)C(\varepsilon_2)$（凸性）$0\le b\le 1$

3）$p(\varphi|x)=p(-\varphi|x)$（对称性） $\varphi \overset{\text{def}}{=} \theta-\hat{\theta}_{MMSE}=\theta-E(\theta|x)$

在这种情况下，使上述这一类代价函数最小的最佳估计$\hat{\theta}$与$\hat{\theta}_{MAP}$或者$\hat{\theta}_{MMSE}$一致。若$C(\varepsilon)$是严格凸U函数，则最佳估计$\hat{\theta}$是惟一的，且等于$\hat{\theta}_{MAP}$或$\hat{\theta}_{MMSE}$。

估计量的不变性

性质2：若代价函数$C(\varepsilon)$是对称的非减函数，即

1）$C(\varepsilon)=C(-\varepsilon)$（对称性）

2）$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}C(\varepsilon)\begin{cases}\geq 0 & \varepsilon \geq 0 \\ \leq 0 & \varepsilon \leq 0\end{cases}$$

同时后验概率密度是对称于条件均值的单峰函数，即

3）$p(\varphi|x)=p(-\varphi|x)$（对称性） $\varphi \overset{\text{def}}{=} \theta-\hat{\theta}_{MMSE}=\theta-\hat{\theta}_{MAP}$

且满足条件4

4）$\lim\limits_{\varphi\to\infty} C(\varphi)\ p(\varphi|x)=0$

则使这一类代价函数最小的最佳估计$\hat{\theta}$与$\hat{\theta}_{MMSE}$或者$\hat{\theta}_{MAP}$一致。

估计量的不变性告诉我们：对相当广泛的一类代价函数，只要性质1和2的条件得以满足，则最小均方误差估计或最大后验概率估计总是使代价最小的最佳估计。

线性贝叶斯估计量的引出

- 最佳贝叶斯估计量是很难用闭合形式确定的，并且在实践中因其计算量太大而难以实现
  - MMSE估计量含有多重积分；
  - MAP估计量含有多维最大值求解问题
- 在不能做出高斯假定的时候，就必须利用另外的方法：**选择保留MMSE准则，但是限定估计量是线性的**，则估计量的显式表示可以很容易地根据PDF的前两阶矩来确定—实践中的维纳滤波器

线性贝叶斯估计器

由数据集 $\{x(0), x(1), \cdots, x(N-1)\}$ 估计标量参数 θ，θ 是随机变量的一个实现，如果将估计限制在一个线性估计器：

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n \cdot x(n) + a_N \tag{5}$$

选择系数集 $\{a_n, n = 0, \cdots N\}$，使 Bayesian MSE 最小，即：

$$\min_{\{a_n\}} \{mse(\hat{\theta})\} = E[(\theta - \hat{\theta})^2] = \iint (\theta - \hat{\theta})^2 p(\mathbf{x}, \theta) d\mathbf{x} d\theta$$

先求解 a_N

$$\frac{\partial}{\partial a_N} E\left[(\theta - \sum_{n=0}^{N-1} a_n x(n) - a_N)^2 \right] = -2E\left[\theta - \sum_{n=0}^{N-1} a_n \cdot x(n) \right] - a_N$$

得

$$a_N = E(\theta) - \sum_{n=0}^{N-1} a_n E(x(n))$$

线性贝叶斯估计器

将 a_N 表达式代入 $Bmse(\hat{\theta})$ 中

$$Bmse(\hat{\theta}) = E\left\{\left[\sum_{n=0}^{N-1} a_n (x(n) - E(x(n)) - (\theta - E(\theta))\right]^2\right\}$$

$$= E\left\{\left[\boldsymbol{a}^T (\boldsymbol{x} - E(\boldsymbol{x})) - (\theta - E(\theta))\right]^2\right\}$$

$$= \boldsymbol{a}^T C_{xx}\boldsymbol{a} - \boldsymbol{a}^T C_{x\theta} - C_{\theta x}\cdot \boldsymbol{a} + C_{\theta\theta}$$

为使其最小，令：

$$\frac{\partial Bmse(\hat{\theta})}{\partial \boldsymbol{a}} = 2C_{xx}\cdot \boldsymbol{a} - 2C_{x\theta}$$

得： $\boldsymbol{a} = C_{xx}^{-1}\cdot C_{x\theta}$

将 $\boldsymbol{a}$ 和 a_N 代入(5)：

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n \cdot x(n) + a_N$$

$$= \sum_{n=0}^{N-1} a_n x(n) + E(\theta) - \sum_{n=0}^{N-1} a_n E(x(n))$$

$$= \boldsymbol{a}^T (\boldsymbol{x} - E(\boldsymbol{x})) + E(\theta)$$

$$= E(\theta) + C_{x\theta}^T \cdot C_{xx}^{-1}(\boldsymbol{x} - E(\boldsymbol{x}))$$

将 $\boldsymbol{a}$ 代入 Bmse 表达式，得最小 Bmse 为：

$$Bmse(\hat{\theta}) = C_{\theta\theta} - C_{\theta x}\cdot C_{xx}^{-1}\cdot C_{x\theta}$$

线性贝叶斯估计器

若有 $E(\theta)=0$， $E(x(n))=0$，则上面各式简化为：

$$\begin{cases} \boldsymbol{a} = C_{xx}^{-1} C_{x\theta} \\ \theta = C_{x\theta}^{T} \cdot C_{xx}^{-1} \cdot \boldsymbol{x} \\ Bmse(\hat{\theta}) = C_{\theta\theta} - C_{x\theta}^{T} \cdot C_{xx}^{-1} \cdot C_{x\theta} \end{cases} \tag{6}$$

- 这组关系式可以直接联系到 Wiener 滤波问题。
- 线性 Bayesian 估计与高斯分布下的一般 Bayesian 估计是一致的，在高斯分布下，线性估计可达最优。

注：在以上推导和讨论中，

$$C_{x\theta} \overset{\Delta}{=} E\left[(\boldsymbol{x}-E(\boldsymbol{x}))(\theta-E(\theta))\right] \quad \text{（N×1 矩阵或列矢量）}$$

$$C_{\theta x} \overset{\Delta}{=} E\left[(\theta-E(\theta))\cdot(\boldsymbol{x}-E(\boldsymbol{x}))^{T}\right] \quad \text{（1×N 矩阵或行矢量）}$$

$$C_{\theta x} = C_{x\theta}^{T}$$

矢量LMMSE估计量

贝叶斯高斯－马尔可夫定理：

如果观测数据x可以使用贝叶斯线性模型表示－x=Hθ+w

其中x是一个N×1的数据矢量，H是一个已知的N×p矩阵,θ是一个p×1的的随机矢量，它的现实是要估计的，其均值和协方差分别为E(θ)和$C_{\theta\theta}$;w是一个N×1的噪声矢量，均值和协方差分别为零和C_w,且与θ是不相关的（另外，联合PDFp(w,θ)是任意的）。那么θ的$MMSE$估计量为

$$\hat{\theta} = E(\theta) + C_{\theta\theta} H^T (H C_{\theta\theta} H^T + C_w)^{-1} (x - HE(\theta))$$

$$= E(\theta) + \left(C_{\theta\theta}^{-1} + H^T C_w^{-1} H\right)^{-1} H^T C_w^{-1} \left(x - HE(\theta)\right)$$

估计量的性能是通过误差$\varepsilon = \theta - \hat{\theta}$来度量的，误差的均值为零，协方差矩阵为 $C_\varepsilon = E_{x,\theta}(\varepsilon\varepsilon^T) = C_{\theta\theta} - C_{\theta\theta} H^T (H C_{\theta\theta} H^T + C_w)^{-1} H C_{\theta\theta}$

$$= \left(C_{\theta\theta}^{-1} + H^T C_w^{-1} H\right)^{-1}$$

误差协方差矩阵也是最小的MSE矩阵$M_{\hat{\theta}}$,其对角线上的元素产生最小贝叶斯MSE,即$\left[M_{\hat{\theta}}\right]_{ii} = \left[C_\varepsilon\right]_{ii} = Bmse(\hat{\theta}_i)$

贝叶斯线性模型

贝叶斯一般线性模型为

$$\mathrm{x}=H\theta+w$$

其中x是一个N×1的数据矢量，H是一个已知的$N\times p$矩阵，θ是一个$p\times 1$的具有先验概率PDF$N(\mu_\theta, C_\theta)$的随机矢量，w是一个$N\times 1$的噪声矢量，具有PDF$N(0, C_w)$，且与θ无关。它和经典的一般线性模型的区别在于，将θ看作为一个具有高斯先验PDF的随机变量。

如果观测数据x满足上面的模型，那么后验PDF$p(\theta|\mathrm{x})$是高斯分布的，它的均值和协方差分别为

$$E(\theta|\mathrm{x})=\mu_\theta+C_\theta H^T(HC_\theta H^T+C_w)^{-1}(\mathrm{x}-H\mu_\theta)$$

$$C_{\theta|\mathrm{x}}=C_\theta-C_\theta H^T(HC_\theta H^T+C_w)^{-1}HC_\theta$$

为确保$HC_\theta H^T+C_w$的可逆性，那么H不必是满秩的。

贝叶斯线性模型

后验*PDF*的均值和协方差还可以表达为

$$E(\theta\,|\,\mathrm{x})=\mu_\theta+\left(C_\theta^{-1}+H^TC_w^{-1}H\right)^{-1}H^TC_w^{-1}\left(\mathrm{x}-H\mu_\theta\right)$$

且

$$C_{\theta|\mathrm{x}}=\left(C_\theta^{-1}+H^TC_w^{-1}H\right)^{-1},\text{即}C_{\theta|\mathrm{x}}^{-1}=C_\theta^{-1}+H^TC_w^{-1}H$$

对于无先验知识的情况，$C_\theta^{-1}\to 0$，因此

$$\hat{\theta}=E(\theta\,|\,\mathrm{x})\to\left(H^TC_w^{-1}H\right)^{-1}H^TC_w^{-1}\mathrm{x}$$

这是一般线性模型的*MVU*估计量

结论：在Bayes线性模型中没有先验知识的时候，MMSE估计量与经典线性模型的MVU估计量有着相同的形式。

贝叶斯线性模型下的MMSE估计量的性能

贝叶斯线性模型下MMSE估计量的性能：

如果观测数据x可以使用贝叶斯线性模型表示，那么*MMSE*估计量为

$$\hat{\theta} = \mu_\theta + C_\theta H^T (H C_\theta H^T + C_w)^{-1} (\mathrm{x} - H\mu_\theta)$$
$$= \mu_\theta + \left(C_\theta^{-1} + H^T C_w^{-1} H\right)^{-1} H^T C_w^{-1} \left(\mathrm{x} - H\mu_\theta\right)$$

估计量的性能是通过误差$\varepsilon = \theta - \hat{\theta}$来度量的，它的PDF是高斯的，均值为零，协方差矩阵为

$$C_\varepsilon = E_{\mathrm{x},\theta}(\varepsilon\varepsilon^T) = C_\theta - C_\theta H^T (H C_\theta H^T + C_w)^{-1} H C_\theta$$
$$= \left(C_\theta^{-1} + H^T C_w^{-1} H\right)^{-1}$$

误差协方差矩阵也是最小的*MSE*矩阵$M_{\hat{\theta}}$,其对角线上的元素产生最小贝叶斯*MSE*, 即$\left[\mathrm{M}_{\hat{\theta}}\right]_{ii} = \left[\mathrm{C}_\varepsilon\right]_{ii} = Bmse(\hat{\theta}_i)$

维纳滤波

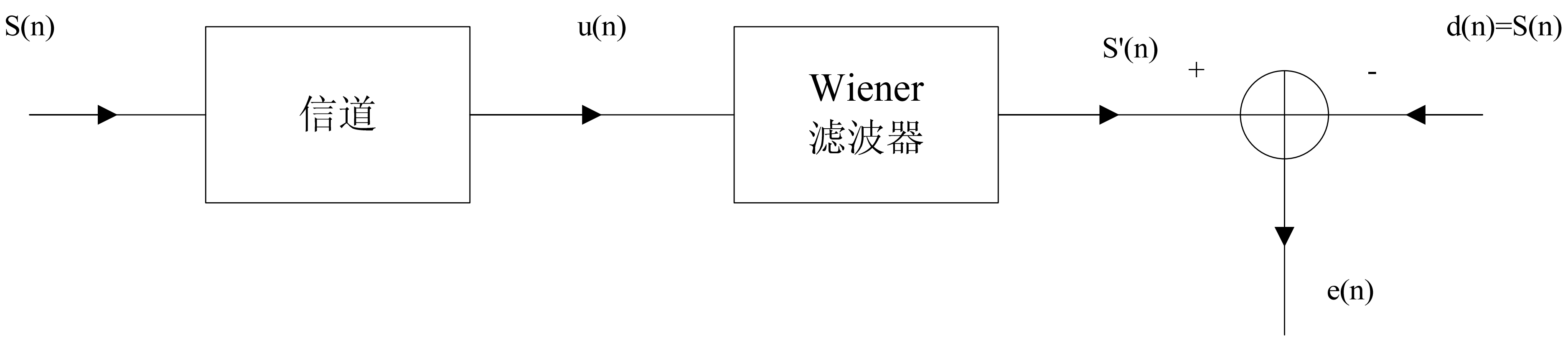
- Wiener滤波器是从统计意义上的最优滤波，它要求输入信号是宽平稳随机序列，我们主要集中在FIR结构的Wiener滤波器的讨论。
- 由信号当前值与它的各阶延迟，

$$\{u(n),u(n-1),\cdots,u(n-M+1)\}$$

估计一个期望信号$d(n)$，输入信号$u(n)$是宽平稳的，$u(n)$和$d(n)$是联合宽平稳的，要求这个估计的均方误差最小.。

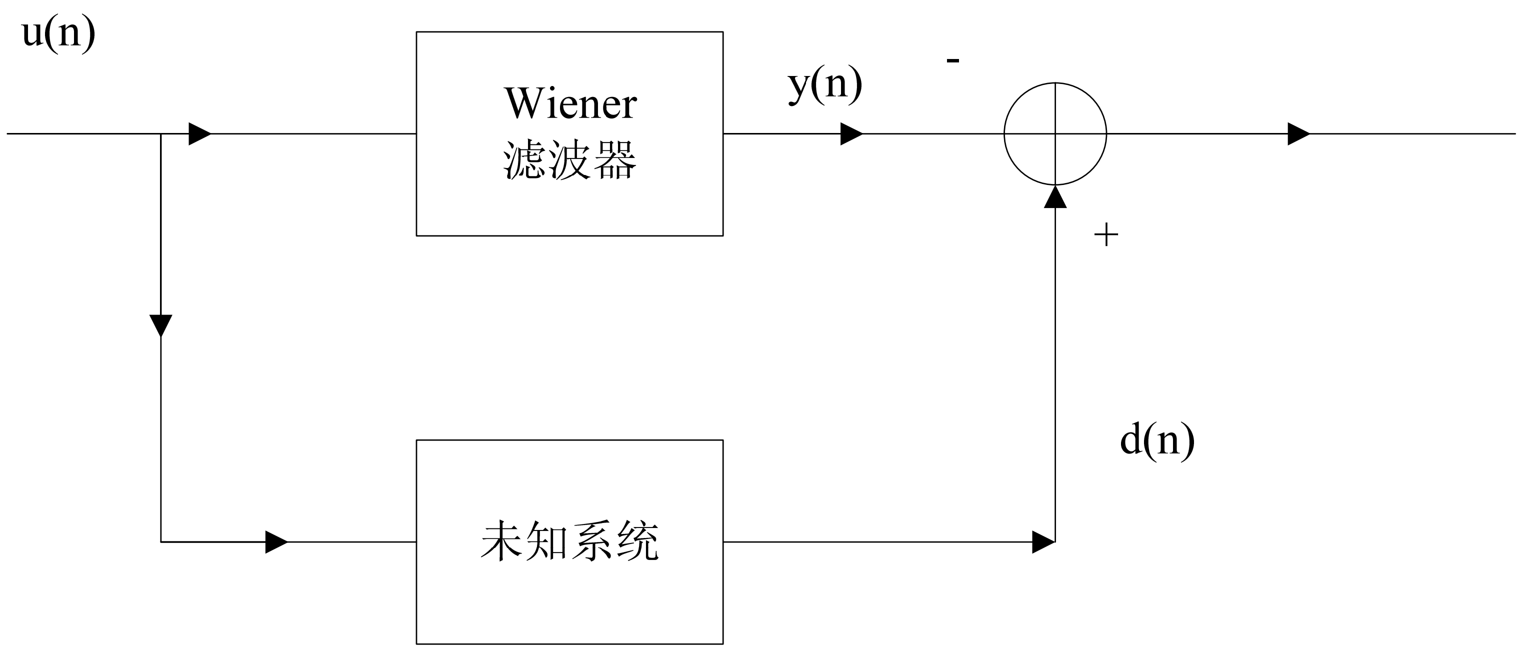
维纳滤波器的应用1

- 通信的信道均衡器



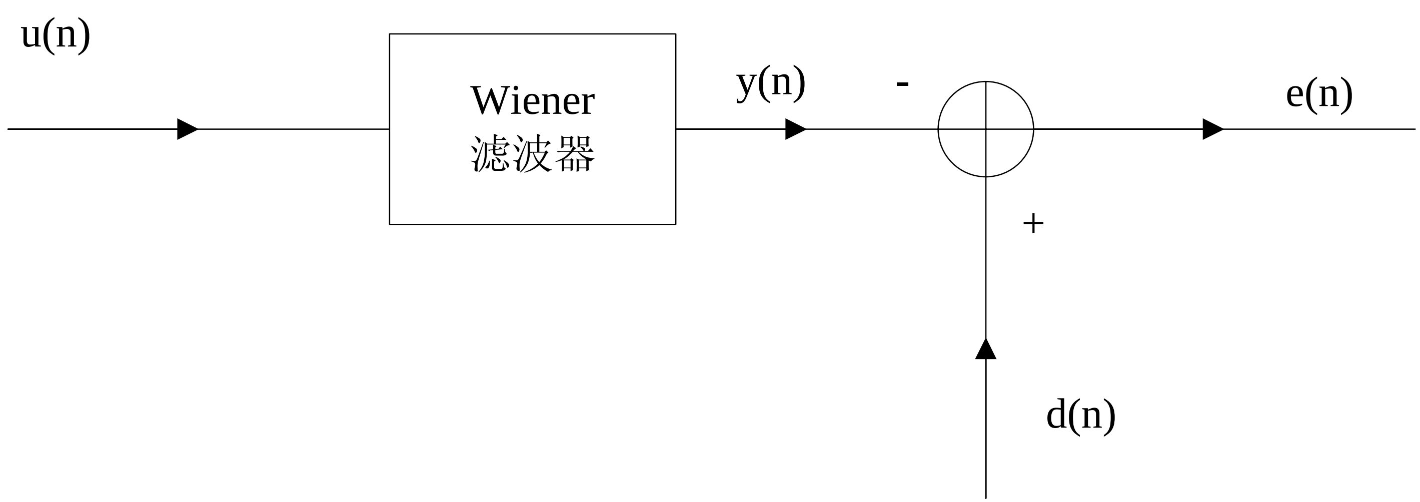
维纳滤波器的应用2

- 系统辨识



维纳滤波器的一般结构

- 一般结构

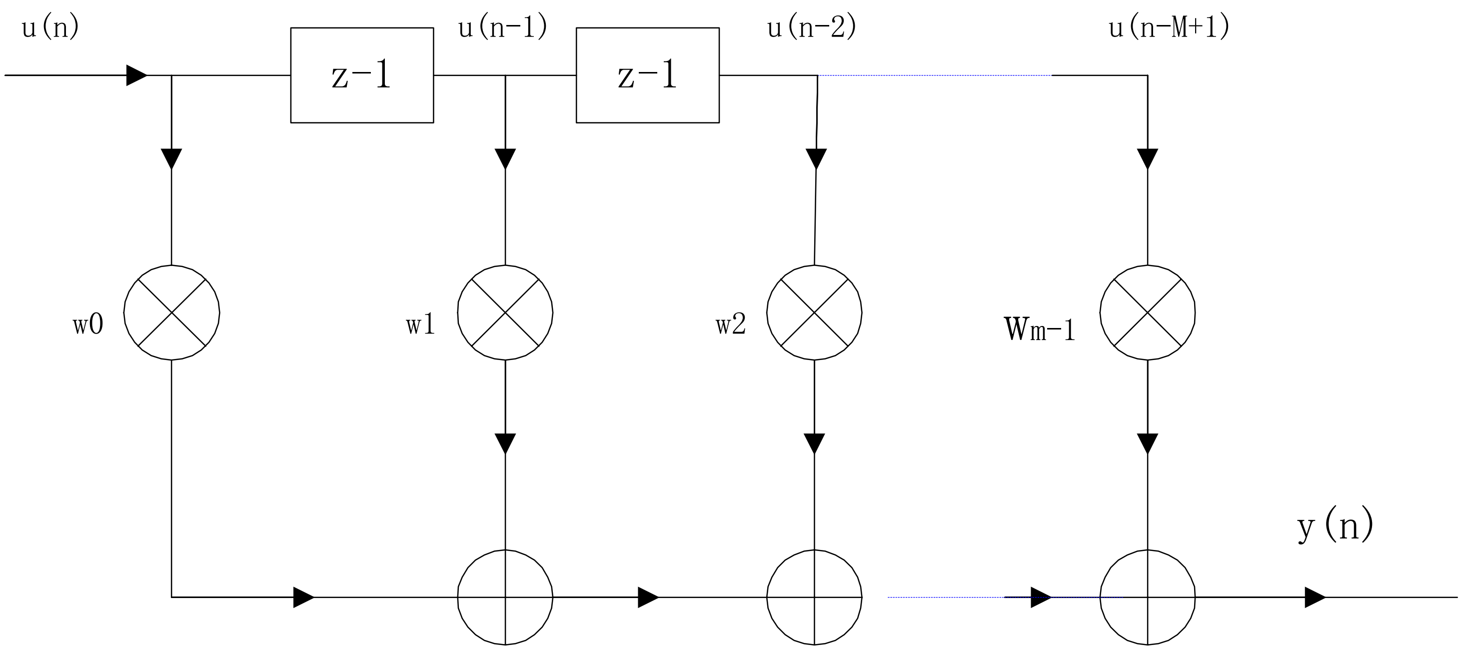


Wiener滤波器的目的是求最优滤波器系数 $\boldsymbol{w}_o$，使下式最小

$$J(n) = E[|e(n)|^2] = E\left[\left|d(n) - \hat{d}(n)\right|^2\right]$$

从估计的观点导出维纳滤波

- FIR结构(也称为横向)的Wiener滤波器的核心结构如下图所示



从估计的观点导出维纳滤波

为了与估计理论一致，假设信号，滤波器权值均为实数

由输入 $u(n)$ 和它的 1 至（M-1）阶延迟，估计期望信号 $d(n)$，确定权系数

$\{w_i, i=0,\cdots M-1\}$ 使估计误差均方值最小，均方误差定义为：

$$J = E[(d(n)-\hat{d}(n))^2]$$

这里估计 $\hat{d}(n)$ 写为：

$$\hat{d}(n) = \sum_{k=0}^{M-1} w_k \cdot u(n-k)$$

从估计的观点导出维纳滤波

除了现在是波形估计外，与线性 Bayesian 估计一一对应。

$\hat{\theta} = \sum_{k=0}^{N-1} a_k \cdot x(k)$	$\hat{d}(n) = \sum_{k=0}^{M-1} w_k \cdot u(n-k)$
$\boldsymbol{a} = [a_0, a_1, \cdots a_{N-1}]^T$	$\boldsymbol{w} = [w_0, w_1, \cdots w_{N-1}]^T$
$\boldsymbol{x} = [x(0), x(1), \cdots x(N-1)]^T$	$\boldsymbol{u}(n) = [u(n), u(n-1), \cdots u(n-M+1)]^T$
θ	$d(n)$
C_{xx}	R（零均值假设）
$C_{x\theta}$	$\boldsymbol{P} = E[\boldsymbol{u}(n) \cdot d(n)] = [p(0), p(-1), \cdots p(-M+1)]^T$

从估计的观点导出维纳滤波

这里($p(-k)=E[u(n-k)d(n)]$)，Wiener 滤波与线性 Bayesian 估计变量之间具有一一对应

关系，设最优滤波器系数为 $\boldsymbol{w}_0$，由线性 Bayesian 估计得到 Wiener 滤波器系数对应式：

$$C_{xx}\cdot \boldsymbol{a}=C_{x\theta} \quad \Rightarrow \quad R\cdot w_0=p$$

上式后一个方程称为 Wiener-Hopf 方程,

或

$$\boldsymbol{a}=C_{xx}^{-1}C_{x\theta} \quad \Rightarrow \quad w_0=R^{-1}\cdot p$$

$$\hat{\theta}=C_{x\theta}^{T}\cdot C_{xx}^{-1}\cdot \boldsymbol{x} \quad \Rightarrow \quad \hat{d}(n)=p^{T}\cdot R^{-1}\cdot \boldsymbol{u}(n)=w_0^{T}\cdot \boldsymbol{u}(n)$$

$$Bmse(\hat{\theta})=C_{\theta\theta}-C_{x\theta}^{T}\cdot C_{xx}^{-1}\cdot C_{x\theta} \quad \Rightarrow \quad J_{\min}=\sigma_d^2-p^{T}\cdot R^{-1}\cdot p$$

从估计的观点导出维纳滤波

结论：

1) Wiener 滤波器是线性 FIR 滤波器中的最优滤波器，但非线性滤波可能会达到更好结果。
2) 在联合高斯条件下，Wiener 滤波也是总体最优的（①从 Bayesian 估计意义上讲是这样，②要满足平稳条件）
3) 从线性贝叶斯估计推导过程知，在滤波器系数取非最优的 $\boldsymbol{w}$ 时，其误差性能表示：

$$J(\boldsymbol{w}) = \sigma_d^2 - \boldsymbol{w}^T \cdot \boldsymbol{p} - \boldsymbol{p}^T \cdot \boldsymbol{w} + \boldsymbol{w}^T \cdot R \cdot \boldsymbol{w}$$

它是 $\boldsymbol{w}$ 的二次曲面，只有一个最小点， $\boldsymbol{w} = \boldsymbol{w}_0$ 时， $J(w) = J_{\min}$

从正交原理和线性滤波观点分析维纳滤波器

Wiener 滤波器是一个最优线性滤波器，图 3 是一个一般表示框图，滤波器核是 IIR 或 FIR 的，为了导出后续常用的一些工具，我们导出最优滤波器的正交原理，并从正交原理出发重新导出一般的 Wiener 滤波器方程，先推导适应于 IIR 和 FIR 的一般结论，然后重点讨论 FIR。

讨论一般的复数形式。

- $u[0],\cdots u[n]\cdots$　输入过程。
- $w_0, w_1, w_2, \cdots$　滤波器系数，（权系数）
- 希望的响应　$d[n]$
- 输出误差：　$e[n]=d[n]-y[n]$

从正交原理和线性滤波观点分析维纳滤波器

- 正交性原理

对复数据情况，推导一般结论，实数据是特例。

$$y[n]=\sum_{R=0}^{\infty} w_k^* u(n-k)$$

$$e[n]=d[n]-y[n]$$

均方误差是：

$$J=E\{e[n]e^*[n]\}$$

$$=E\{|e[n]|^2\}$$

设权系数

$$w_k=a_k+jb_k$$

定义递度算子 $\nabla=[\nabla_0,\nabla_1,\cdots\nabla_k\cdots]^T$

其中 $\nabla_k=\frac{\partial}{\partial w_k}=\frac{\partial}{\partial a_k}+j\frac{\partial}{\partial b_k}$

符号 ∇J 是递度算子作用于 J，其中第 k 项为：

$$\nabla_k J=\frac{\partial J}{\partial a_k}+j\frac{\partial J}{\partial_{bk}}$$

要求 $w_0,w_1,\cdots$ 的值,使得 J 最小,即

$$\nabla J=\theta$$

或等价：

$$\nabla_k J=0 \qquad k=0,1,2\cdots$$

将 $J = E\{e[n]\cdot e^*[n]\}$ 代入递度算子

得到：

$$\nabla_k J = E\left[\frac{\partial e[n]}{\partial a_k} e^*[n] + \frac{\partial e^*[n]}{\partial a_k}\cdot e[n] + \frac{\partial e[n]}{\partial b_k}\cdot je^*[n] + \frac{\partial e^*[n]}{\partial b_k}\cdot je[n]\right]$$

由

$$e[n] = d[n] - \sum_{k=0}^{\infty} w_k^* \cdot u[n-k]$$

得到：

$$\frac{\partial e[n]}{\partial a_k} = -u[n-k]$$

$$\frac{\partial e[n]}{\partial b_k} = ju[n-k]$$

$$\frac{\partial e^*[n]}{\partial a_k} = -u^*[n-k]$$

$$\frac{\partial e^*[n]}{\partial b_k} = -ju^*[n-k]$$

代入 $\nabla_k J$ 表达式整理得：

$$\nabla_k J = -2E[u[n-k]\cdot e^*[n]]$$

当

$$\nabla_k J = 0 \qquad k = 0,1,\cdots$$

时，J 达到最小。

从正交原理和线性滤波观点分析维纳滤波器

设 J 达最小时，用 w_0, $e_0[n]$ 表示权系数和误差 e[n]，且 $J = J_{\min}$

则有：

$$E[u[n-k]e_0^*[n]] = 0\text{，}\quad k = 0, 1, \cdots$$

和

$$E[u^*[n-k] \cdot e_0[n]] = 0\text{，}\quad k = 0, 1, 2\cdots$$

以上两式为正交性原理，达到最优滤波时，误差和输入正交。

推论：

$$E[y_0[n] \cdot e_0^*[n]] = 0$$

从正交原理和线性滤波观点分析维纳滤波器

- 最小均方误差：

在达最优时，$y_0[n]$ 也写成：$\hat{d}[n|U_n]$，表示由 $u[n], u[n-1], \cdots\cdots$ 张成的空间对 d[n] 的估计（最优线性估计）。

$$e_0[n] = d[n] - y_0[n] = d[n] - \hat{d}[n|u_n]$$

也可以写成：

$$d[n] = e_0[n] + d[n|u_n]$$

由 $\hat{d}[n|u_n]$ 和 $e_0[n]$ 正交性得：

$$\sigma_d^2 = E\left[\left|e_o[n]\right|^2\right] + \sigma_{\hat{d}}^2$$

$$= J_{\min} + \sigma_{\hat{d}}^2$$

即：

$$J_{\min} = \sigma_d^2 - \sigma_{\hat{d}}^2$$

从正交原理和线性滤波观点分析维纳滤波器

- 维纳一霍夫方程（**Wiener-Hopf**）

由正交性原理

$$E[u[n-k]\cdot e_0^*[n]]=0 \qquad k=0,1,\cdots$$

即：

$$E\left[u[n-k]\cdot\left(d^*[n]-\sum_{i=0}^{\infty}w_{0i}\cdot u^*[n-i]\right)\right]=0 \qquad k=0,1,\cdots$$

得：

$$\sum_{i=0}^{\infty}w_{0i}E[u[n-k]\cdot u^*[n-i]]=E[u[n-k]\cdot d^*[n]]$$

由于：

$$r_{uu}[i-k]=E[u[n-k]\cdot u^*[n-i]]$$

$$p[-k]=E[u[n-k]\cdot d^*[n]]$$

得：

$$\sum_{i=0}^{\infty}w_{0i}\cdot r_{uu}[i-k]=p[-k] \quad k=0,1,\cdots$$

这就是 Wiener-Hopf 方程，解此方程，可得到最优权 $\{w_{0i}\}$。

对于 M 阶 FIR 滤波器，（横向滤波器）wiener-Hopf 方程变为：

$$\sum_{i=0}^{M-1} w_{0i} r[i-k] = p[-k], \qquad k = 0, 1, \cdots M-1$$

• 矩阵形式：

令

$$u[n] = [u[n], u[n-1], \cdots u[n-M+1]]^T$$

和

$$R = E[\boldsymbol{u}[n] \cdot \boldsymbol{u}^H[n]]$$

$$1 = \begin{pmatrix} r[0], & r[1], & \cdots & r[M-1] \\ r^*[1], & r[0], & \cdots & r[M-2] \\ \vdots & & & \\ r^*[M-1], & r^*[M-2], & \cdots & r[0] \end{pmatrix}$$

$$\boldsymbol{p} = E[\boldsymbol{u}[n] \cdot d^*[n]] = [p[n], p[-1], \cdots p[1-M]]^T$$

Winer-Hopf 方程写为：

$$R \cdot \boldsymbol{w}_0 = \boldsymbol{p}$$

这里 $\boldsymbol{w}_0 = [w_{00}, w_{01}, w_{02}, \cdots w_{0M-1}]^T$

解方程求得：

$$\boldsymbol{w}_0 = R^{-1}\boldsymbol{p}$$

- 再看最小均方误差

由

$$\hat{d}[n \mid U_n] = \sum_{k=0}^{M-1} w_k^* u[n-k]$$

$$= \boldsymbol{w}_0^H \cdot \boldsymbol{u}[n]$$

得

$$\sigma_{\hat{d}}^2 = E[\hat{d}[n \mid U_n] \cdot \hat{d}^*[n \mid U_n]]$$

$$= E[\boldsymbol{w}_0^H \cdot \boldsymbol{u}[n] \cdot \boldsymbol{u}^H[n] \cdot \boldsymbol{w}_0]$$

$$= \boldsymbol{w}_0^H \cdot E[\boldsymbol{u}[n] \cdot \boldsymbol{u}^H[n]] \cdot \boldsymbol{w}_0$$

$$= \boldsymbol{w}_0^H \cdot R_{uu} \cdot \boldsymbol{w}_0$$

$$= \boldsymbol{w}_0^H \cdot \boldsymbol{p}$$

$$= \boldsymbol{p}^H \cdot \boldsymbol{w}_0 = \boldsymbol{p}^H \cdot R^{-1}\boldsymbol{p}$$

则

$$J_{\min} = \sigma_d^2 - \sigma_{\hat{d}}^2$$

$$= \sigma_d^2 - p^H \cdot w_0$$

$$= \sigma_d^2 - p^H \cdot R^{-1} \cdot p$$

从正交原理和线性滤波观点分析维纳滤波器

- 分析：由正交原理导出Wiener滤波器和从线性贝叶斯估计得到wiener滤波器，分别得到一些有益的启示，从估计理论看,Wiener滤波器只是一个线性Beyesian估计，它是最优估计的一个线性逼近，只有在高斯情况下，它才是真正的最优滤波器，在其它分布情况下，非线性滤波器可以达到比线性最优滤波器更优的结果. 从一般线性最优滤波器的正交原理出发，我们容易忽视这些限制。

误差性能表面

由

$$e[n]=d[n]-\sum_{k=0}^{M-1}w_k^*u[n-k]$$

直接代入

$$J=E[e[n]\cdot e^*[n]]$$

整理得：

$$J=\sigma_d^2-\sum_{k=0}^{M-1}w_k^*p(-k)-\sum_{k=0}^{M-1}w_k\cdot p^*(-k)+\sum_{k=0}^{M-1}\cdot\sum_{i=0}^{M-1}w_k^*\cdot w_i r[i-k]$$

由上式，可以看出，J是W_k的二次曲面，是碗状曲面，碗口向上，J_{min}在碗底，其实，由上式直接对w_k求导，得到一组方程，正是wiener-Hopf方程。

误差性能曲面

上式也可以直接写成矩阵形式

$$J(w)=\sigma_d^2-\boldsymbol{w}^H p-\boldsymbol{p}^H\cdot\boldsymbol{w}+\boldsymbol{w}^H R\boldsymbol{w}$$

它可以整理成如下形式：

$$J(w)=\sigma_d^2-p^H\cdot R^{-1}\cdot p+(w-R^{-1}p)^H\cdot R\cdot(w-R^{-1}p)$$

上式在 $\boldsymbol{w}_0=R^{-1}p$ 时，达最小，

$$J_{\min}=\min_w J(w)=\sigma_d^2-p^H\cdot R^{-1}\cdot p$$

性能表面 $J(w)$ 也可以写成：

$$J(\boldsymbol{w})=J_{\min}+(\boldsymbol{w}-\boldsymbol{w}_0)^H\cdot R(\boldsymbol{w}-\boldsymbol{w}_0)$$

由于 $R=Q\Lambda Q^H$

故

$$J(w)=J_{\min}+(w-w_0)^H Q\Lambda Q^H\cdot(w-w_0)$$

误差性能表面

令 $\boldsymbol{v} = Q^H(\boldsymbol{w} - \boldsymbol{w}_0)$

$$J = J_{\min} + v^H \cdot \Lambda \cdot v$$

$$= J_{\min} + \sum_{k=1}^{M} \lambda_k \cdot v_k \cdot v_k^*$$

$$= J_{\min} + \sum_{k=1}^{N} \lambda_k \cdot |v_k|^2$$

$$J - J_{\min} = \sum_{k=1}^{M} \frac{|v_k|^2}{\frac{1}{\lambda_k}}$$

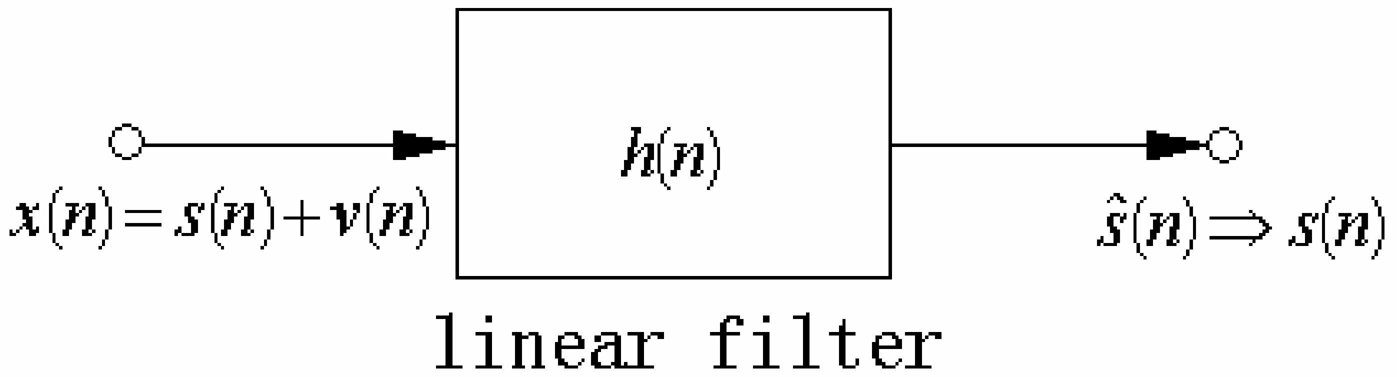
这是超椭圆，$\dfrac{1}{\lambda_k}$ 为其一个轴。

通过坐标变换，得到如上规范形式，对于一个给定 $J \neq J_{\min}$，有：

谢谢大家

Normal equations of Wiener filter

(1) Wiener Filtering Problem

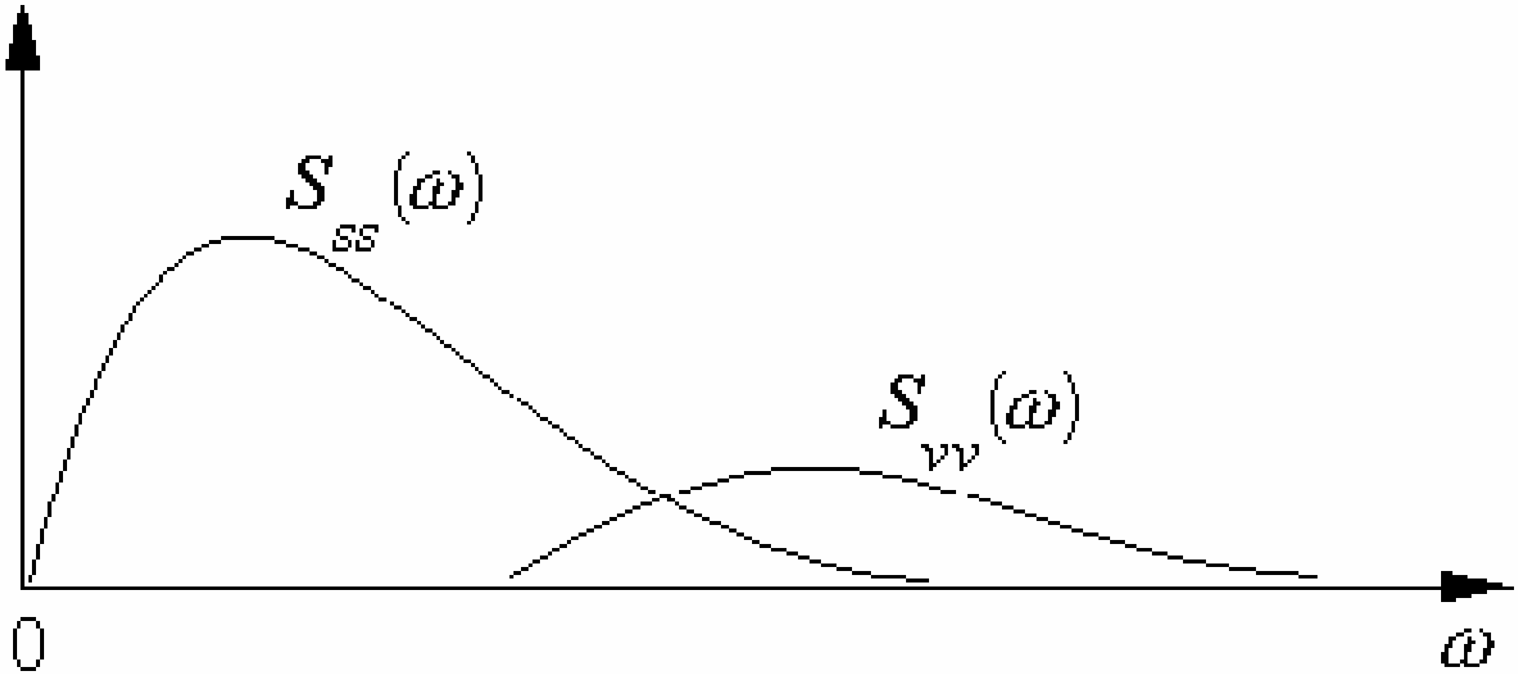


$x(n)$ Observed data or Measured data

$s(n)$ Signal

$v(n)$ Noise or Interference

Additive combination



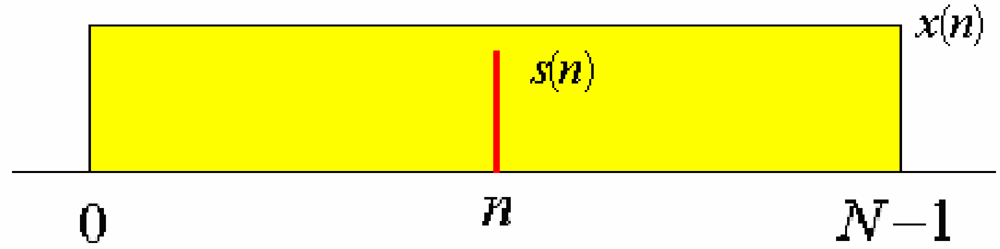
$$\hat{s}(n) = x(n) * h(n) = \sum_i h(i)x(n-i)$$ Linear estimation

$$\xi(n) = E\left[e^2(n)\right] = \min \Rightarrow h(n)$$ Optimum estimation (minimum mean-square error)

$$e(n) = s(n) - \hat{s}(n)$$ Estimation error

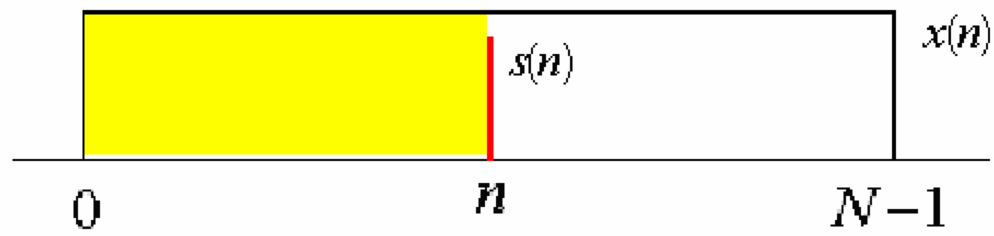
Smoothing

$$\hat{s}(n) = \sum_{i=0}^{N-1} h(n-i)x(i)$$



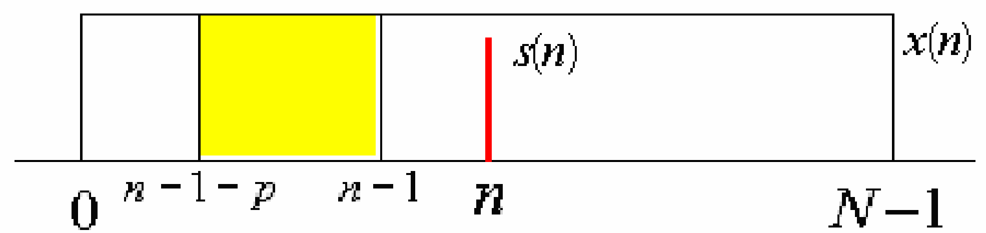
Filtering

$$\hat{s}(n) = \sum_{i=0}^{n} h(n-i)x(i)$$



Prediction

$$\hat{s}(n) = \sum_{i=n-1-p}^{n-1} h(n-i)x(i)$$



P-order forward one step linear prediction (LPC)

Wiener Filter

$$\hat{s}(n) = \sum_{i=0}^{n} h(i)x(n-i), \quad n = 0, \quad 1, \quad \cdots, \quad N-1$$

$$\begin{bmatrix} \hat{s}(0) \\ \hat{s}(1) \\ \hat{s}(2) \\ \vdots \\ \hat{s}(N-1) \end{bmatrix} = \begin{bmatrix} h(0) & 0 & 0 & \cdots & 0 \\ h(1) & h(0) & 0 & \cdots & 0 \\ h(2) & h(1) & h(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ h(N-1) & h(N-2) & h(N-3) & \cdots & h(0) \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(N-1) \end{bmatrix}$$

Causality: Low diagonal matrix

(2) Orthogonal equations

$$\frac{\partial \xi(n)}{\partial h(j)} = 2E\left[e(n)\frac{\partial e(n)}{\partial h(j)}\right] = -2E[e(n)x(n-j)] = 0, \quad j = 0, \quad 1, \quad \cdots$$

$$E[e(n)x(n-j)] = 0, \quad \forall j$$

(3) normal equations (Wiener-Hopf equations)

$$R_{sx}(m) = \sum_i h(i)R_{xx}(m-i), \quad \forall m$$

where $R_{xx}(m-i) = E[x(n-i)x(n-m)]$ autocorrelation sequence of $x(n)$

$R_{sx}(m) = E[s(n)x(n-m)]$ cross-correlation sequence of $s(n)$ and $x(n)$

The Normal Equations of Causal Wiener Filter

$$R_{sx}(m) = \sum_{i=0}^{\infty} h(i)R_{xx}(m-i), \quad m \geq 0$$