Sequential LMMSE

1. Sequential LMMSE

다음과 같은 데이터가 주어졌다고 가정하자.

$$x[n] = A + w[n], \quad n = 0, 1, \ldots, N-1 \text{ where } A \sim N(\mu_A, \sigma_A^2),\ w[n] \sim N(0, \sigma^2)$$

해당 데이터에 대해서 MMSE는 이미 구한 바 있다.

$$\text{MMSE :} \qquad \hat{A} = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \bar{x} + \frac{\frac{\sigma^2}{N}}{\sigma_A^2 + \frac{\sigma^2}{N}} \mu_A$$

$$Bmse(\hat{A}) = \frac{\sigma^2}{N} \left(\frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \right)$$

또한, LMMSE도 동일한 값을 가진다.

여기서 $\mu_A = 0$이라고 가정하면,

$$\hat{A}[N-1] = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \bar{x}$$

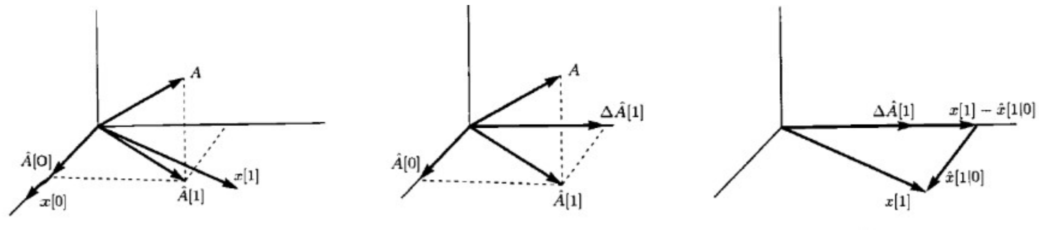
$$Bmse(\hat{A}[N-1]) = \frac{\sigma^2 \sigma_A^2}{N\sigma_A^2 + \sigma^2}$$

이에 따라 $\hat{A}[N]$을 구해보면,

$$\begin{aligned}
\hat{A}[N] &= \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N+1}} \frac{1}{N+1} \sum_{n=0}^{N} x[n] \\
&= \frac{N\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \frac{1}{N} \left(\sum_{n=0}^{N-1} x[n] + x[N] \right) \\
&= \frac{N\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \frac{\sigma_A^2 + \frac{\sigma^2}{N}}{\sigma_A^2} \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\
&= \frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\
&= \hat{A}[N-1] + \left(\frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} - 1 \right) \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\
&= \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} (x[N] - \hat{A}[N-1]) \\
&= \hat{A}[N-1] + K[N](x[N] - \hat{A}[N-1])
\end{aligned}$$

$$\begin{aligned}
Bmse(\hat{A}[N]) &= \frac{\sigma_A^2 \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \\
&= \frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \frac{\sigma_A^2 \sigma^2}{N\sigma_A^2 + \sigma^2} \\
&= (1 - K[N]) Bmse(\hat{A}[N-1])
\end{aligned}$$

2. Geometrical Interpretation



(a) (b) (c)

$\hat{A}[1]$은 A의 $x[0], x[1]$이 span하는 subspace 위로의 projection이다.

$\hat{A}[0]$은 A의 $x[0]$ 위로의 projection이다.

따라서 $\hat{A}[1] = \hat{A}[0] + \Delta\hat{A}[1]$이며 $\Delta\hat{A}[1]$는 $\hat{A}[0]$와 orthogonal하다.

$\hat{x}[1|0]$이 $x[0]$이 주어졌을 때 $x[1]$의 LMMSE라고 하면,

error값 $\tilde{x}[1] = x[1] - \hat{x}[1|0]$은 $x[0]$에 orthogonal하다.

따라서 $\Delta\hat{A}[1]$은 A의 $\tilde{x}[1] = x[1] - \hat{x}[1|0]$로의 projection이다.

그러므로 다음과 같은 수식으로 표현 가능하다.

$$\Delta\hat{A}[1] = (A, \frac{\tilde{x}[1]}{\|\tilde{x}[1]\|})\frac{\tilde{x}[1]}{\|\tilde{x}[1]\|} = \frac{E(A, \tilde{x}[1])}{E(\tilde{x}^2[1])}\tilde{x}[1]$$

$$K[1] = \frac{E(A, \tilde{x}[1])}{E(\tilde{x}^2[1])}$$

$$\hat{A}[1] = \hat{A}[0] + K[1](x[1] - \hat{x}[1|0])$$

일반적으로는 다음과 같이 표현 가능하다.

$$\hat{A}[N-1] = \sum_{n=0}^{N-1} K[n](x[n] - \hat{x}[n|n-1])$$

$$\hat{A}[N] = \hat{A}[N-1] + K[N](x[N] - \hat{x}[N|N-1])$$

$$K[n] = \frac{E[A(x[n] - \hat{x}[n|n-1])]}{E[(x[n] - \hat{x}[n|n-1])^2]}$$

이러한 Sequential LMMSE는 다음과 같은 4가지 명제를 만족한다.

$$\hat{\theta}_i[n] = \hat{\theta}_i[n-1] + K_i[n](x[n] - \hat{x}[n|n-1])$$

$$\hat{\theta}[n] = \hat{\theta}[n-1] + K[n](x[n] - \hat{x}[n|n-1])$$

LMMSE가 다음과 같을 때,

1) $A\theta$의 LMMSE는 $A\hat{\theta}$이다.

2) $\theta_1 + \theta_2$의 LMMSE는 $\hat{\theta}_1 + \hat{\theta}_2$이다.

3) $\hat{\theta}_i[n-1]$과 $x[n] - \hat{x}[n|n-1]$은 서로 uncorrelated하므로 orthogonal하다. 따라서 $E[\hat{\theta}_i[n-1](x[n] - \hat{x}[n|n-1])] = 0$을 만족한다.

4) Bayesian linear model의 가정에 따라 θ와 $w[n]$은 uncorrelated하고, $\hat{\theta}[n-1]$과 $w[n]$도 uncorrelated하므로

$E[(\theta - \hat{\theta}[n-1])w[n]] = 0$이다.

우선 다음과 같은 Bayesian linear model의 데이터를 가정하자.

$$x[n] = h^T[n]\theta + w[n]$$

위의 1), 2)번 명제에 따라 다음이 성립한다.

$$\hat{x}[n|n-1] = h^T[n]\hat{\theta}[n-1] + \hat{w}[n|n-1]$$

또한 위의 4)번 명제에 따라 다음이 성립한다.

$$\hat{x}[n|n-1] = h^T[n]\hat{\theta}[n-1]$$

여기서 $K_i[n] = \dfrac{E[\theta_i(x[n]-\hat{x}[n|n-1])]}{E[(x[n]-\hat{x}[n|n-1])^2]}$ 인데 분자 분모를 각각 계산해보면

$$\begin{aligned}
E[(x[n]-\hat{x}[n|n-1])^2] &= E[(h^T[n]\theta + w[n] - h^T[n]\hat{\theta}[n-1])^2] \\
&= E[(h^T[n](\theta - \hat{\theta}[n-1]) + w[n])^2] \\
&= h^T[n]E[(\theta-\hat{\theta}[n-1])(\theta-\hat{\theta}[n-1])^T]h[n] + E(w^2[n]) \\
&= h^T[n]M[n-1]h[n] + \sigma_n^2 \\
E[\theta_i(x[n]-\hat{x}[n|n-1])] &= E[(\theta_i - \hat{\theta}_i[n-1])(x[n]-\hat{x}[n|n-1])] \\
&= E[(\theta_i - \hat{\theta}_i[n-1])(h^T[n]\theta + w[n] - h^T[n]\hat{\theta}[n-1])] \\
&= E[(\theta_i - \hat{\theta}_i[n-1])(h^T[n](\theta - \hat{\theta}[n-1]))] \\
&= E[(\theta_i - \hat{\theta}_i[n-1])(\theta - \hat{\theta}[n-1])]h[n] \\
&= M[n-1]h[n]
\end{aligned}$$

따라서 구하고자 하는 $K[n]$은 다음과 같다.

$$K[n] = \frac{M[n-1]h[n]}{h^T[n]M[n-1]h[n] + \sigma_n^2}$$

또한 $M[n]$도 다음과 같이 sequential하게 구할 수 있다.

$$\begin{aligned}
M[n] &= E[(\theta - \hat{\theta}[n])(\theta - \hat{\theta}[n])^T] \\
&= E[(\theta - \hat{\theta}[n-1] - K[n](x[n]-\hat{x}[n|n-1]))(\theta - \hat{\theta}[n-1] - K[n](x[n]-\hat{x}[n|n-1]))^T] \\
&= M[n-1] - E[(\theta - \hat{\theta}[n-1])(x[n]-\hat{x}[n|n-1])K^T[n]] \\
&\quad - K[n]E[(x[n]-\hat{x}[n|n-1])(\theta - \hat{\theta}[n-1])^T] + K[n]E[(x[n]-\hat{x}[n|n-1])^2]K^T[n]
\end{aligned}$$

여기서

$$K[n]E[(x[n]-\hat{x}[n|n-1])^2] = E[\theta(x[n]-\hat{x}[n|n-1])] = M[n-1]h[n]$$

$$E[(\theta - \hat{\theta}[n-1])(x[n]-\hat{x}[n|n-1])] = E[\theta(x[n]-\hat{x}[n|n-1])] = M[n-1]h[n]$$

이기 때문에 결과적으로 다음과 같이 전개된다.

$$\begin{aligned}
M[n] &= M[n-1] - M[n-1]h[n]K^T[n] - K[n]h^T[n]M[n-1] + M[n-1]h[n]K^T[n] \\
&= (1 - K[n]h^T[n])M[n-1]
\end{aligned}$$

또한 mean이 0이 아닌 LMMSE에 대해서도 다음의 과정에 따라 동일하다.

$$\begin{aligned}
\hat{\theta}[n] - E(\theta) &= \hat{\theta}[n-1] - E(\theta) + K[n](x[n] - E(x[n]) - h^T[n](\hat{\theta}[n-1] - E(\theta))) \\
\hat{\theta}[n] &= \hat{\theta}[n-1] + K[n](x[n] - h^T[n]\hat{\theta}[n-1] - E(x[n]) + h^T[n]E(\theta)) \\
&= \hat{\theta}[n-1] + K[n](x[n] - h^T[n]\hat{\theta}[n-1])
\end{aligned}$$

전체적인 과정을 살펴보면 다음과 같다.

Estimator Update:

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \mathbf{K}[n](x[n] - \mathbf{h}^T[n]\hat{\boldsymbol{\theta}}[n-1])$$

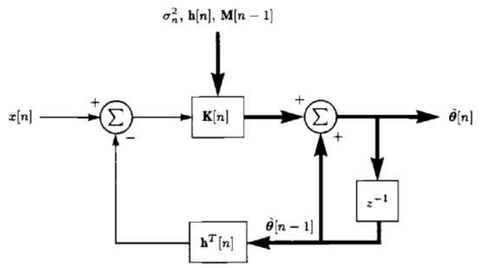
where

$$\mathbf{K}[n] = \frac{\mathbf{M}[n-1]\mathbf{h}[n]}{\sigma_n^2 + \mathbf{h}^T[n]\mathbf{M}[n-1]\mathbf{h}[n]}.$$

Minimum MSE Matrix Update:

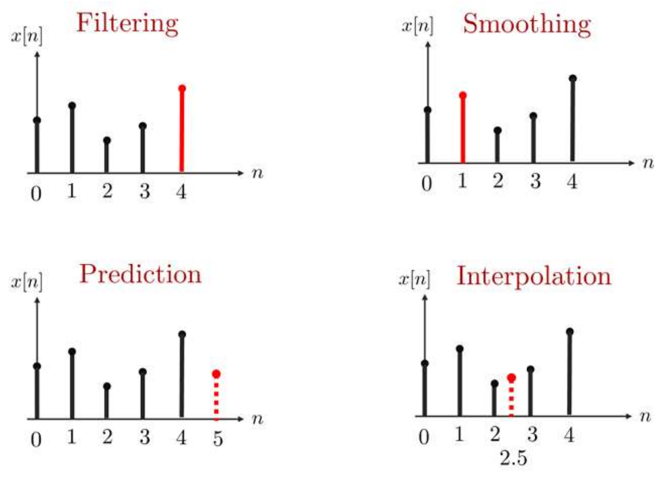
$$\mathbf{M}[n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{h}^T[n])\mathbf{M}[n-1].$$

$$\mathbf{M}[n] = E\left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n])(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n])^T\right]$$



(cf. Sequential Least Squares)

3. Estimation Problems
Filtering : 이전의 데이터들로 추정
Smoothing : 전체 데이터들로 추정
Prediction : 이전 데이터들로 미래 데이터 추정
Interpolation : 데이터들로 가운데 데이터 추정



1) Smoothing
$x[n] = s[n] + w[n]$에서 $x = [x[0], x[1], ..., x[N-1]]^T$로부터 $s = [s[0], s[1], ..., s[N-1]]^T$추정하는 문제이다.

여기서 $r_{xx}[k] = E(x[n]x[n+k]) = r_{ss}[k] + r_{ww}[k]$일 때,
$C_{xx} = R_{xx} = R_{ss} + R_{ww}$이고, $C_{\theta x} = E(sx^T) = E(s(s+w)^T) = R_{ss}$이므로

$$\begin{aligned} \hat{s} &= E(s) + C_{sx}C_{xx}^{-1}(x - E(x)) \\ &= C_{sx}C_{xx}^{-1}x = R_{ss}(R_{ss} + R_{ww})^{-1}x \end{aligned}$$

가 된다.

2) Filtering

$x[n] = s[n] + w[n]$에서 $x = [x[0], x[1], ..., x[n]]^T$로부터 $\theta = s[n]$을 추정하는 문제이다.
앞서 했던 것과 비슷하게 전개해보면

$$C_{xx} = R_{ss} + R_{ww}$$

$$\begin{aligned} C_{\theta x} &= E(s[n][x[0]\, x[1] \ldots x[n]]) \\ &= E(s[n][s[0]\, s[1] \ldots s[n]]) \\ &= [r_{ss}[n]\, r_{ss}[n-1] \ldots r[0]] \end{aligned}$$

$$\begin{aligned} \hat{s}[n] &= r'^{T}_{ss}(R_{ss} + R_{ww})^{-1}x \\ &= a^T x \\ a &= (R_{ss} + R_{ww})r'_{ss} \\ &= [a_0\, a_1 \ldots a_n] \end{aligned}$$

여기서 $h^{(n)}[k] = a_{n-k}$라고 한다면

$$\hat{s}[n] = \sum_{k=0}^{n} a_k x[k] = \sum_{k=0}^{n} h^{(n)}[n-k]x[k] = \sum_{k=0}^{n} h^{(n)}[k]x[n-k]$$

가 되며 다음을 만족하게 된다.

$$(R_{ss} + R_{ww})a = r'_{ss}$$
$$(R_{ss} + R_{ww})h = r_{ss}$$

$$\begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \ldots & r_{xx}[n] \\ r_{xx}[1] & r_{xx}[0] & \ldots & r_{xx}[n-1] \\ \ldots & \ldots & \ldots & \ldots \\ r_{xx}[n] & r_{xx}[n-1] & \ldots & r_{xx}[0] \end{bmatrix} \begin{bmatrix} h^{(n)}[0] \\ h^{(n)}[1] \\ \ldots \\ h^{(n)}[n] \end{bmatrix} = \begin{bmatrix} r_{ss}[0] \\ r_{ss}[1] \\ \ldots \\ r_{ss}[n] \end{bmatrix}$$

위 행렬식을 위너-호프 방정식이라고 하며 위너-호프 방정식은 정상 랜덤 프로세스에 대해 최적의 필터 계수 $h^{(n)}[k]$를 결정하는 방정식이다.