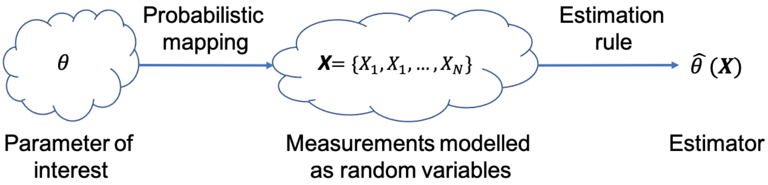
Wiener filters

Raj Thilak Rajan

Estimation Philosophy



- Let $X = \{X_1, X_2, \ldots, X_N\}$ be a set of random samples drawn from probability distributions $f_{X_n}(x_n; \boldsymbol{\theta}) \; \forall \; 1 \leq n \leq N$, where $\boldsymbol{\theta}$ is the parameter of interest
- We aim to
  (a) recover the unknown $\boldsymbol{\theta}$ from the measurements X, and
  (b) provide a performance measure of the estimated $\boldsymbol{\theta}$
- Bayesian philosophy : θ is a random variable and the statistics of θ is known.

Bayesian mean square error (Bmse)

- θ is viewed as a random variable
- We would like to minimize the MSE

$$\text{Bmse}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$$

where both $\mathbf{x}$ and θ are random, and the statistics of $\hat{\theta}$ depend on the statistics of both $\mathbf{x}$ and θ.

- Note the difference between these two MSEs:

$$\mathrm{mse}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \int (\hat{\theta} - \theta)^2 p(\mathbf{x}; \theta) d\mathbf{x}$$

$$\text{Bmse}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \int \int (\hat{\theta} - \theta)^2 p(\mathbf{x}, \theta) d\mathbf{x} d\theta$$

- Note that *mse* depends on θ, but *Bmse* does not, only on its statistics.

Minimum mean square estimation (MMSE)

- We know from Bayes' theorem $p(\mathbf{x}, \theta) = p(\theta|\mathbf{x})p(\mathbf{x})$, and hence

$$Bmse(\hat{\theta}) = \int \int (\hat{\theta} - \theta)^2 p(\mathbf{x}, \theta) d\mathbf{x} d\theta = \int \left[\int (\hat{\theta} - \theta)^2 p(\theta|\mathbf{x}) d\theta \right] p(\mathbf{x}) d\mathbf{x},$$

and since $p(\mathbf{x}) \geq 0$ for all $\mathbf{x}$, we minimize the inner integral for each $\mathbf{x}$ i.e.,

$$\text{Solve:} \qquad \underset{\theta}{min} \int (\hat{\theta} - \theta)^2 p(\theta|\mathbf{x}) d\theta$$

- Solution: Setting the derivative with respect to $\hat{\theta}$ to zero we obtain:

$$\begin{aligned} \frac{\partial}{\partial \hat{\theta}} \int (\hat{\theta} - \theta)^2 p(\theta|\mathbf{x}) d\theta &= 2 \int (\hat{\theta} - \theta) p(\theta|\mathbf{x}) d\theta \\ &= 2\hat{\theta} - 2 \int \theta p(\theta|\mathbf{x}) d\theta = 0 \end{aligned}$$

or

$$\hat{\theta} = \int \theta p(\theta|\mathbf{x}) d\theta = \mathbb{E}(\theta|\mathbf{x})$$

Maximum a posteriori (MAP)

- The MAP estimator corresponds to

$$\hat{\theta} = \arg\max_{\theta} p(\theta|\mathbf{x}).$$

- Using Bayes' rule, this is identical to

$$\hat{\theta} = \arg\max_{\theta} p(\mathbf{x}|\theta)p(\theta) = \arg\max_{\theta} log\left(p(\mathbf{x}|\theta)\right) + log\left(p(\theta)\right).$$

- MAP properties:
  - If $N \to \infty$, the pdf $p(\mathbf{x}|\theta)$ becomes dominant over $p(\theta)$ and the MAP becomes thus identical to the Bayesian MLE.
  - If the $\mathbf{x}$ and $\boldsymbol{\theta}$ are jointly Gaussian, then the MAP estimator is identical to the MMSE estimator.

Linear MMSE estimator

- Problem: Constrain the estimator to be linear i.e.,

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] + a_N$$

and choose the weighting coefficients $a_n \forall\ 0 \leq n \leq N$ to minimize

$$\text{Bmse}(\hat{\theta}) = \mathbb{E}\left[(\theta - \hat{\theta})^2\right]$$

- Solution: The LMMSE estimator and the corresponding Bayesian MSE is

$$\begin{aligned} \hat{\theta} &= \mathbb{E}(\theta) + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \mathbb{E}(\mathbf{x})) \\ \mathbf{M}(\hat{\theta}) &= C_{\theta\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} \end{aligned}$$

or for a vector parameter

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \mathbb{E}(\boldsymbol{\theta}) + \mathbf{C}_{\boldsymbol{\theta}x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \mathbb{E}(\mathbf{x})) \\ \mathbf{M}(\hat{\boldsymbol{\theta}}) &= \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} \end{aligned} \tag{1}$$

LMMSE estimator: Properties

- Bayesian Gauss-Markov model:

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

with $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_w)$ and $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_\theta, \mathbf{C}_\theta)$, the LMMSE estimator is

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \boldsymbol{\mu}_\theta + \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H}\mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_\theta) \\ &= \boldsymbol{\mu}_\theta + (\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H} + \mathbf{C}_\theta^{-1})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_\theta) \end{aligned}$$

and for $\epsilon = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$, the performance of the estimator is

$$\mathbf{C}_\epsilon = \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) = (\mathbf{C}_\theta^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1}$$

- LMMSE estimators are
  - identical in form to the MMSE estimator for jointly Gaussian $\mathbf{x}$ and θ
  - commutative and additive for affine transformations

Wiener filter: Problem formulation (1)

- We aim to estimate $\boldsymbol{\theta}$, from the measurements/data

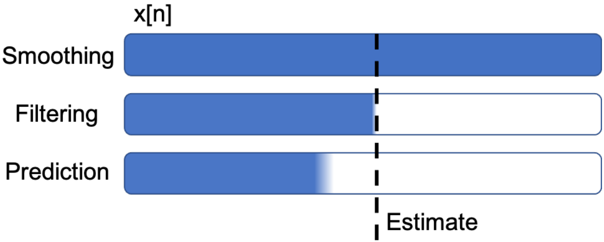
$$\mathbf{x} = [x[0], x[1], \ldots, x[N-1]]^T$$

which is WSS and zero mean, with a Toeplitz covariance structure

$$\mathbf{C}_{xx} = \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \ldots & r_{xx}[N-1] \\ r_{xx}[1] & r_{xx}[0] & \ldots & r_{xx}[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}[N-1] & r_{xx}[N-2] & \ldots & r_{xx}[0] \end{bmatrix} = \mathbf{R}_{xx}$$

- Smoothing:
  - Given $x[n] = s[n] + w[n],\ n = 0, 1, \ldots, N-1$, estimate $\theta = s[n]$
- Filtering:
  - Given $x[m] = s[m] + w[m],\ m = 0, 1, \ldots, n$, estimate $\theta = s[n]$
- Prediction:
  - Given $\{x[0], x[1], \ldots, x[N-1]\}$, estimate $\theta = s[N-1+l]$ for $l \geq 1$

Wiener filter: Problem formulation (2)



- Smoothing:
  - Given $x[n] = s[n] + w[n],\ n = 0, 1, \ldots, N-1$, estimate $\theta = s[n]$
- Filtering:
  - Given $x[m] = s[m] + w[m],\ m = 0, 1, \ldots, n$, estimate $\theta = s[n]$
- Prediction:
  - Given $\{x[0], x[1], \ldots, x[N-1]\}$, estimate $\theta = s[N-1+l]$ for $l \geq 1$

Wiener filter: Problem formulation (3)

- Problem: Design filters for Smoothing, Filtering and Prediction
- Assumptions:
  1. $\mathbb{E}(\mathbf{x}) = \mathbb{E}(\boldsymbol{\theta}) = \mathbf{0}$
  2. The signal and noise processes are uncorrelated i.e.,

$$r_{xx}[k] = r_{ss}[k] + r_{ww}[k], \quad \text{or} \quad \mathbf{C}_{xx} = \mathbf{R}_{xx} = \mathbf{R}_{ss} + \mathbf{R}_{ww}$$

- For $\mathbb{E}(\mathbf{x}) = \mathbb{E}(\boldsymbol{\theta}) = \mathbf{0}$, the vector LMMSE estimator and the Bayesian MSE matrix for θ are

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \mathbb{E}(\boldsymbol{\theta}) + \mathbf{C}_{x\theta}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \mathbb{E}(\mathbf{x})) = \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{x} \\ \mathbf{M}_{\hat{\theta}} &= \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} \end{aligned} \tag{2}$$

Wiener filter: Smoothing

- Problem
  - Given $x[n] = s[n] + w[n],\ n = 0, 1, \ldots, N-1$, estimate $\theta = s[n]$
  - Recollect, $\hat{\boldsymbol{\theta}} = \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{x}$, $\mathbf{M}_{\hat{\theta}} = \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}$
  - Define $\mathbf{x} = [x[0], x[1], \ldots, x[N-1]]^T$ and $\mathbf{s} = [s[0], s[1], \ldots, s[N-1]]^T$
- The covariance matrices are

$$\begin{aligned} \mathbf{C}_{xx} &= \mathbf{R}_{xx} = \mathbf{R}_{ss} + \mathbf{R}_{ww} \qquad \text{(Note: } N \times N \text{ matrices)} \\ \mathbf{C}_{\theta x} &= \mathbb{E}(\mathbf{s}\mathbf{x}^T) = \mathbb{E}(\mathbf{s}(\mathbf{s}+\mathbf{w})^T) = \mathbf{R}_{ss} \end{aligned}$$

- The Wiener estimator and the corresponding BMSE are

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \hat{\mathbf{s}} = \mathbf{R}_{ss}(\mathbf{R}_{ss} + \mathbf{R}_{ww})^{-1}\mathbf{x} = \mathbf{W}\mathbf{x} \\ \mathbf{M}_{\hat{\theta}} &= \mathbf{R}_{ss} - \mathbf{R}_{ss}(\mathbf{R}_{ss} + \mathbf{R}_{ww})^{-1}\mathbf{R}_{ss} = (\mathbf{I} - \mathbf{W})\mathbf{R}_{ss} \end{aligned}$$

where $\mathbf{W}$ is the Wiener smoothing matrix.

Wiener filter: Filtering

- Problem
  - Given $x[m] = s[m] + w[m],\ m = 0, 1, \ldots, n$, estimate $\theta = s[n]$
  - Recollect, $\hat{\boldsymbol{\theta}} = \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{x}$, $\mathbf{M}_{\hat{\theta}} = \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}$
  - Define $\mathbf{x}' = [x[0], x[1], \ldots, x[n]]^T$, $\mathbf{s}' = [s[0], s[1], \ldots, s[n]]^T$ and $\mathbf{r}_{ss} = [r_{ss}[0], r_{ss}[1], \ldots, r_{ss}[n]]^T$
- The covariance matrices are

$$\begin{aligned} \mathbf{C}_{xx} &= \mathbf{R}_{xx} = \mathbf{R}_{ss} + \mathbf{R}_{ww} \qquad \text{(Note: } n+1 \times n+1 \text{ matrices)} \\ \mathbf{C}_{\theta x} &= \mathbb{E}(s[n]\mathbf{x}'^T) = \mathbb{E}(s[n]\mathbf{s}'^T) = [r_{ss}[n], r_{ss}[n-1], \ldots, r_{ss}[0]] = \mathbf{r}'^T_{ss} \end{aligned}$$

- The Wiener estimator and the corresponding Bayesian MSE are

$$\begin{aligned} \hat{\theta} &= \hat{s}[n] = \mathbf{r}'^T_{ss}(\mathbf{R}_{ss} + \mathbf{R}_{ww})^{-1}\mathbf{x} = \mathbf{a}^T\mathbf{x} \\ \mathbf{M}_{\hat{\theta}} &= r_{ss}[0] - \mathbf{r}'^T_{ss}(\mathbf{R}_{ss} + \mathbf{R}_{ww})^{-1}\mathbf{r}'_{ss} \end{aligned}$$

- Relationship to Wiener-Hopf filtering equations

$$(\mathbf{R}_{ss} + \mathbf{R}_{ww})\mathbf{a} = \mathbf{r}'_{ss} \quad \leftrightarrow \quad (\mathbf{R}_{ss} + \mathbf{R}_{ww})\mathbf{h} = \mathbf{r}_{ss} \quad \leftrightarrow \quad \mathbf{R}_{xx}\mathbf{h} = \mathbf{r}_{ss}$$

where $\mathbf{h} = \{h^{(n)}[k] \triangleq a_{n-k}\}$ for $k = 0, 1, \ldots, n$.

Wiener filter: Prediction

- Problem
  - Given $\{x[0], x[1], \ldots, x[N-1]\}$, estimate $\theta = s[N-1+l]$ for $l \geq 1$
  - Recollect, $\hat{\boldsymbol{\theta}} = \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{x}$, $\mathbf{M}_{\hat{\theta}} = \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}$
  - Define $\mathbf{x} = [x[0], x[1], \ldots, x[n]]^T$ and $\mathbf{r}_{xx} = [r_{xx}[0], r_{xx}[1], \ldots, r_{xx}[N-1]]^T$
- The covariance matrices are

$$\begin{aligned} \mathbf{C}_{xx} &= \mathbf{R}_{xx} = \mathbf{R}_{ss} + \mathbf{R}_{ww} \qquad \text{(Note: } N \times N \text{ matrices)} \\ \mathbf{C}_{\theta x} &= \mathbb{E}(x[N-1+l]\mathbf{x}^T) = [r_{xx}[N-1+l], \ldots, r_{xx}[l]] = \mathbf{r'}_{xx}^T \end{aligned}$$

- The l-step linear predictor and the corresponding Bayesian MSE are

$$\begin{aligned} \hat{\theta} &= \hat{x}[N-1+l] = \mathbf{r'}_{xx}^T\mathbf{R}_{xx}^{-1}\mathbf{x} = \mathbf{a}^T\mathbf{x} \\ \mathbf{M}_{\hat{\theta}} &= r_{xx}[0] - \mathbf{r}'_{xx}{}^T\mathbf{R}_{xx}^{-1}\mathbf{r}'_{xx} \end{aligned}$$

- Relationship to Wiener-Hopf filtering equations

$$\mathbf{R}_{xx}\mathbf{a} = \mathbf{r}'_{xx} \quad \leftrightarrow \quad \mathbf{R}_{xx}\mathbf{h} = \mathbf{r}_{xx}$$

where $\mathbf{h}$ is the vector '$\mathbf{a}$' when flipped upside down.

Summary

Key points:

- Wiener filter is a special case (or an application) of the LMMSE estimator, leading to the Wiener-Hopf equations
- Smoothing : $\hat{\mathbf{s}} = \mathbf{R}_{ss}(\mathbf{R}_{ss} + \mathbf{R}_{ww})^{-1}\mathbf{x} = \mathbf{W}\mathbf{x}$
- Filtering : $\hat{s}[n] = \mathbf{r'}_{ss}^{T}(\mathbf{R}_{ss} + \mathbf{R}_{ww})^{-1}\mathbf{x}$
- Prediction : $\hat{x}[N-1+l] = \mathbf{r'}_{xx}^{T}\mathbf{R}_{xx}^{-1}\mathbf{x}$, for $l \geq 1$
- Reading: Kalman filter (not for examination)

Next session:

- Introduction to Neyman Pearson detector