Statistical Digital Signal Processing

Chapter 4: Linear Estimation of Signals

Bisrat Derebssa, SECE, AAiT, AAU

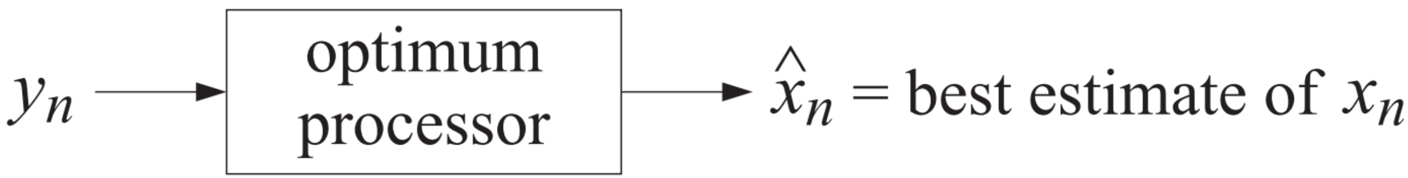
4.1. Introduction

- The problem of estimating one signal from another is one of the most important in signal processing.
- In many applications, the desired signal is not available or observable directly.
  - The observable signal is a degraded or distorted version of the original signal.
- The signal estimation problem is to recover, in the best way possible, the desired signal from its degraded replica.

- Typical Examples:
  - Desired signal may be corrupted by strong additive noise.
  - Signal distorted by magnitude and phase distortion due to channel properties.
  - Image blurring due to motion between camera and object.
  - Infer one signal from observation of another.

General Problem

- Estimate a random signal x_n on the basis of available observations of a related signal y_n.



- What does best estimate mean?
  - The maximum a posteriori (MAP) criterion.
  - The maximum likelihood (ML) criterion.
  - The mean square (MS) criterion.
  - The linear mean-square (LMS) criterion.

- Assume that the desired signal x_n is to be estimated over a finite time interval $n_a \leq n \leq n_b$.

$$\mathbf{x} = \begin{bmatrix} x_{n_a} \\ x_{n_a+1} \\ \vdots \\ x_{n_b} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_{n_a} \\ y_{n_a+1} \\ \vdots \\ y_{n_b} \end{bmatrix}$$

- Estimation is therefore determining the functional dependence

$$\hat{x}_n = \hat{x}_n(\mathbf{y})$$

The maximum a posteriori (MAP) criterion

- Maximizes the a posteriori conditional density of $\mathbf{x}_n$ given that $\mathbf{y}$ already occurred.

$$p(x_n|\mathbf{y}) = \text{maximum}$$

- Therefore, the estimate is the most probable choice resulting from the given observation.
- In general, results in non-linear equations.

Maximum Likelihood (ML)

- Selects estimate that maximizes the conditional density of **y** given $\mathbf{x}_n$, that is

$$p(\mathbf{y}|x_n) = \text{maximum}$$

- Selects estimates as though the collected observations **y** are the most likely to occur.
- In general, results in non-linear equations.

Mean Square (MS)

- Minimizes the mean-square estimation error

$$\mathcal{E} = E[e_n^2] = \min, \quad \text{where} \quad e_n = x_n - \hat{x}_n$$

- This solution is the conditional mean

$$\hat{x}_n = E[x_n|\mathbf{y}]$$

- That is, the estimate is the expected value given the observation **y**.
- In general, results in non-linear equations.

Linear Mean Square (LMS)

- Minimizes the mean-square estimation error assuming the estimate is **linear function of the observations**,

$$\mathcal{E} = E[e_n^2] = E[(x_n - \hat{x}_n)^2]$$

$$\hat{x}_n = \sum_{i=n_a}^{n_b} h(n,i) y_i$$

- Results in linear equations.
  - Easier to solve

Example:

- A discrete-amplitude, constant in time signal x can take on three values -1, 0, or 1 each with probability of 1/3. This signal is placed on a known carrier waveform c_n and transmitted over a noisy channel. The received samples are of the form

$$y_n = c_n x + v_n\,, \quad n = 1, 2, \ldots, M$$

  where v_n are white Gaussian noise IID(0,σ2).
- Obtain and compare the four alternative estimates.

- Obtain the conditional probabilities
  - p(y|x)

$$p(\mathbf{y}|x) = p(\mathbf{v}) = \prod_{n=1}^{M} p(v_n) = (2\pi\sigma_v^2)^{-M/2} \exp\left[-\frac{1}{2\sigma_v^2}\sum_{n=1}^{M} v_n^2\right]$$

If x is given the only randomness in *y* is *v*

$$= (2\pi\sigma_v^2)^{-M/2} \exp\left[-\frac{1}{2\sigma_v^2}\mathbf{v}^2\right] = (2\pi\sigma_v^2)^{-M/2} \exp\left[-\frac{1}{2\sigma_v^2}(\mathbf{y}-\mathbf{c}x)^2\right]$$

  - p(x|y)

Using Bayes' rule we find $p(x|\mathbf{y}) = p(\mathbf{y}|x)\,p(x)/p(\mathbf{y})$. Since

$$p(x) = \frac{1}{3}\left[\delta(x-1)+\delta(x)+\delta(x+1)\right]$$

we find

$$p(x|\mathbf{y}) = \frac{1}{A}\left[p(\mathbf{y}|1)\,\delta(x-1)+p(\mathbf{y}|0)\,\delta(x)+p(\mathbf{y}|-1)\,\delta(x+1)\right]$$

where the constant A is

Law of total probability

$$A = 3p(\mathbf{y}) = 3\int p(\mathbf{y}|x)\,p(x)\,dx = p(\mathbf{y}|1)+p(\mathbf{y}|0)+p(\mathbf{y}|-1)$$

- MAP approximation
  - Since p(x|y) is three impulses, it is the maximum of these

$$p(\mathbf{y}|x) = \text{maximum of}\{p(\mathbf{y}|1),\, p(\mathbf{y}|0),\, p(\mathbf{y}|-1)\}$$

Using the gaussian nature of $p(\mathbf{y}|x)$, we find equivalently

$$(\mathbf{y} - \mathbf{c}\,x)^2 = \text{minumum of}\{(\mathbf{y} - \mathbf{c})^2,\, \mathbf{y}^2,\, (\mathbf{y} + \mathbf{c})^2\}$$

Subtracting $\mathbf{y}^2$ from both sides, dividing by $\mathbf{c}^T\mathbf{c}$, and denoting

$$\bar{y} = \frac{\mathbf{c}^T\mathbf{y}}{\mathbf{c}^T\mathbf{c}}$$

we find the equivalent equation

$$x^2 - 2x\bar{y} = \min\{1 - 2\bar{y},\, 0,\, 1 + 2\bar{y}\}$$

and in particular, applying these for $+1, 0, -1$, we find

$$\hat{x}_{\text{MAP}} = \begin{cases} 1, & \text{if } \bar{y} > \frac{1}{2} \\ 0, & \text{if } -\frac{1}{2} < \bar{y} < \frac{1}{2} \\ -1, & \text{if } \bar{y} < -\frac{1}{2} \end{cases}$$

- ML approximation

$$\frac{\partial}{\partial x} p(\mathbf{y}|x) = 0 \quad \text{or} \quad \frac{\partial}{\partial x} \ln p(\mathbf{y}|x) = 0 \quad \text{or} \quad \frac{\partial}{\partial x} (\mathbf{y} - \mathbf{c}\,x)^2 = 0$$

which gives

$$\hat{x}_{\text{ML}} = \frac{\mathbf{c}^T \mathbf{y}}{\mathbf{c}^T \mathbf{c}} = \bar{y}$$

- MS approximation

$$E[x|\mathbf{y}] = \int x p(x|\mathbf{y}) dx = \int x \frac{1}{A} [p(\mathbf{y}|1)\delta(x-1) + p(\mathbf{y}|0)\delta(x) + p(\mathbf{y}|-1)\delta(x+1)] dx$$

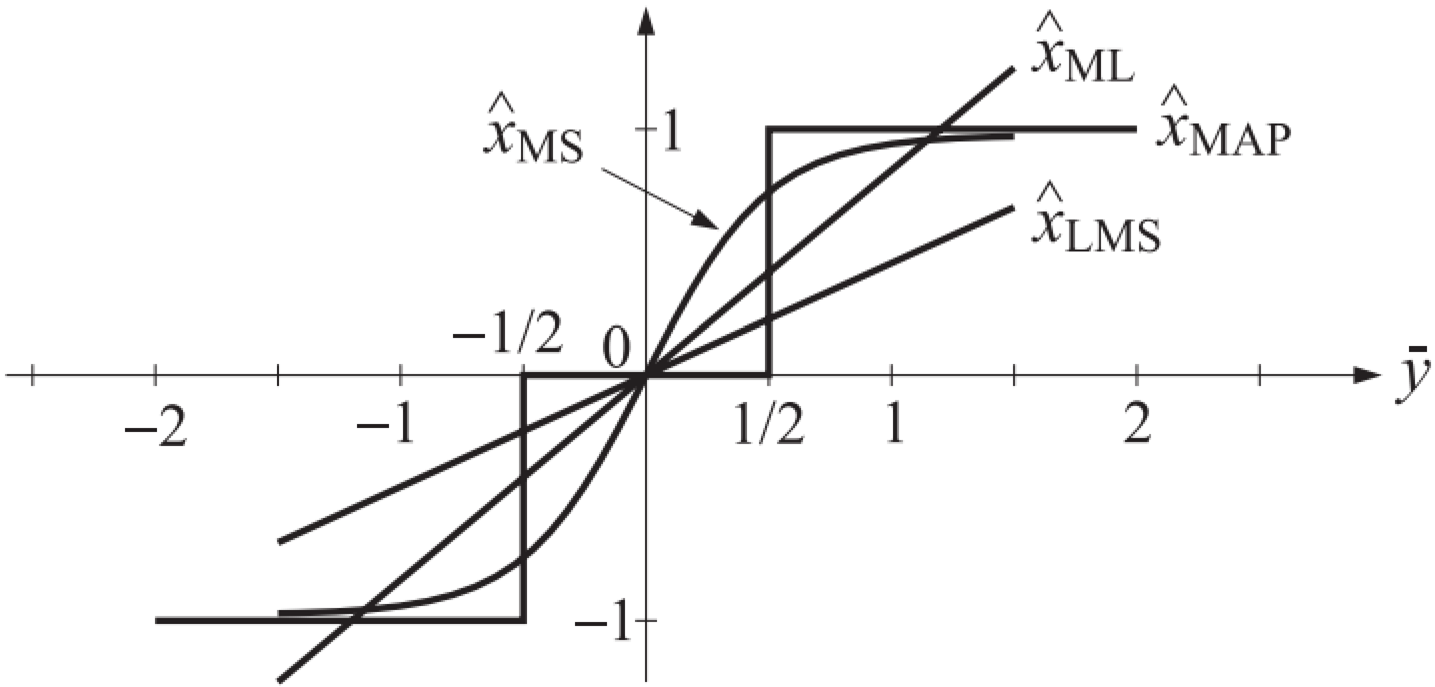
$$= \frac{1}{A}[p(\mathbf{y}|1) - p(\mathbf{y}|-1)], \quad \text{or,}$$

$$\hat{x}_{\text{MS}} = \frac{p(\mathbf{y}|1) - p(\mathbf{y}|-1)}{p(\mathbf{y}|1) + p(\mathbf{y}|0) + p(\mathbf{y}|-1)}$$

$$\hat{x}_{\text{MS}} = \frac{2\sinh(2a\bar{y})}{e^{a} + 2\cosh(2a\bar{y})}, \quad \text{where} \quad a = \frac{\mathbf{c}^T\mathbf{c}}{2\sigma_v^2}$$

- LMS approximation (to be derived later)

$$\hat{x}_{\text{LMS}} = \frac{\mathbf{c}^T\mathbf{y}}{\frac{\sigma_v^2}{\sigma_x^2} + \mathbf{c}^T\mathbf{c}} = \frac{\mathbf{c}^T\mathbf{c}}{\frac{\sigma_v^2}{\sigma_x^2} + \mathbf{c}^T\mathbf{c}} \bar{y}$$



$\hat{x}_{ML}$
$\hat{x}_{MS}$
1
$\hat{x}_{MAP}$
$\hat{x}_{LMS}$
−1/2
0
$\bar{y}$
−2
−1
1/2
1
2
−1

4.2. Linear Mean Square Error Estimation

- Design an estimator that provides an estimate of desired y(n) using linear combination of observed data x(n)

$$\hat{y}(n) \triangleq \sum_{k=1}^{M} c_k^*(n) x_k(n)$$

- Goal is to obtain the coefficients $c_k(n)$ such that the MSE is minimized.

$$P(\mathbf{c}) = E\{|e|^2\} \qquad e \triangleq y - \hat{y}$$

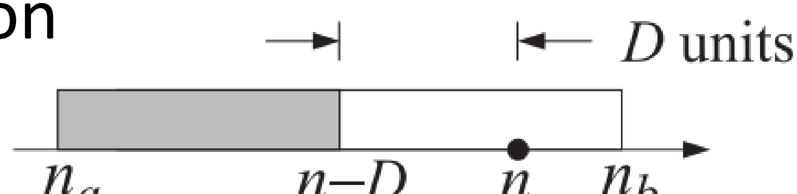
- In general, the coefficients are different at each time instant.
  - Assuming the index n is obvious, it is usually dropped in the equation.

$$\hat{y} \triangleq \sum_{k=1}^{M} c_k^* x_k = \mathbf{c}^H \mathbf{x}$$

- Depending on the location of n with respect to the data segment
  - Smoothing

  - Filtering
    - Observations in the present and past are taken

  - Prediction



- Conventions
  - All random variables are assumed to have zero mean.
    - If not, the mean is subtracted first.
  - The number M of data components used in the estimation is called the order of estimator.
  - The coefficient values that minimize the mean square error are called the linear MMSE (LMMSE) estimator, c_0.
  - The estimate found using c_0 is called the LMMSE estimate

- By using the linearity property of the expectation operator

$$P(\mathbf{c}) = E\{|e|^2\} = E\{(y - \mathbf{c}^H\mathbf{x})(y^* - \mathbf{x}^H\mathbf{c})\}$$

$$= E\{|y|^2\} - \mathbf{c}^H E\{\mathbf{x}y^*\} - E\{y\mathbf{x}^H\}\mathbf{c} + \mathbf{c}^H E\{\mathbf{x}\mathbf{x}^H\}\mathbf{c}$$

$$P(\mathbf{c}) = P_y - \mathbf{c}^H\mathbf{d} - \mathbf{d}^H\mathbf{c} + \mathbf{c}^H\mathbf{R}\mathbf{c}$$

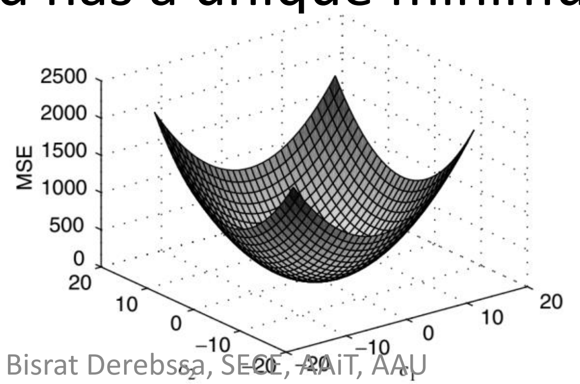
- Where

$$P_y \triangleq E\{|y|^2\}$$ Power of desired response

$$\mathbf{d} \triangleq E\{\mathbf{x}y^*\}$$ Cross-correlation between observed data vector x and desired response y.

$$\mathbf{R} \triangleq E\{\mathbf{x}\mathbf{x}^H\}$$ Autocorrelation matrix of the observed data vector x

- Note the MSE function P(c) : $P(\mathbf{c}) = P_y - \mathbf{c}^H\mathbf{d} - \mathbf{d}^H\mathbf{c} + \mathbf{c}^H\mathbf{R}\mathbf{c}$
  - depends only on the second-order moments of the desired response and the data,
  - is a quadratic function of the estimator coefficients and represents an (M + 1) dimensional surface with M degrees of freedom.
  - If the autocorrelation matrix, R, is positive definite, then P(c) is bowl-shaped and has a unique minimum.



- Putting the previous equation in “perfect square”.

$$P(\mathbf{c}) = P_y - \mathbf{d}^H \mathbf{R}^{-1} \mathbf{d} + (\mathbf{Rc} - \mathbf{d})^H \mathbf{R}^{-1} (\mathbf{Rc} - \mathbf{d})$$

- Note that only third term is dependent on c.
- In addition, since R is positive definite, its inverse is also positive define.
  - The third term is always non-negative definite.
  - The lowest error is when the third term is zero.

$$\mathbf{Rc} - \mathbf{d} = \mathbf{0}.$$

$$\mathbf{Rc}_o = \mathbf{d}$$

- These are called the normal equations:

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1M} \\ r_{21} & r_{22} & \cdots & r_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ r_{M1} & r_{M2} & \cdots & r_{MM} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_M \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_M \end{bmatrix}$$

$$r_{ij} \triangleq E\{x_i x_j^*\} = r_{ji}^*$$

$$d_i \triangleq E\{x_i y^*\}$$

- The minimum error is then

$$P_o = P_y - \mathbf{d}^H \mathbf{R}^{-1} \mathbf{d} = P_y - \mathbf{d}^H \mathbf{c}_o$$

- Note that:
  - If x and y are uncorrelated, the error is the worst $P_0 = P_y$.
  - As long as d is not zero, the MSE decreases the error.
- The normalized MSE is used for comparison

$$\mathcal{E} \triangleq \frac{P_o}{P_y} = 1 - \frac{P_{\hat{y}_o}}{P_y} \qquad 0 \le \mathcal{E} \le 1$$

- Any deviation from the optimum vector increases the error.

$$\mathbf{c} = \mathbf{c}_o + \tilde{\mathbf{c}},$$

$$P(\mathbf{c}_o + \tilde{\mathbf{c}}) = P(\mathbf{c}_o) + \tilde{\mathbf{c}}^H \mathbf{R} \tilde{\mathbf{c}}$$

- Note that the excess MSE depends on the input correlation matrix only.
  - Any deviation from the optimum can be detected by monitoring the MSE.

4.3. Principal component analysis of LMSE

- If R is expressed in terms of its eigenvalues and eigenvectors,

$$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H = \sum_{i=1}^{M} \lambda_i \mathbf{q}_i \mathbf{q}_i^H$$

$$\mathbf{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_M\}$$

$$\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_M]$$

$$\mathbf{Q}^H\mathbf{Q} = \mathbf{I}$$

- Transforming the optimum parameter vector by the eigenvectors

$$\mathbf{c}_o' \triangleq \mathbf{Q}^H \mathbf{c}_o \qquad \text{or} \qquad \mathbf{c}_o \triangleq \mathbf{Q}\mathbf{c}_o'$$

- Substituting these into the normal equations

$$\mathbf{\Lambda}\mathbf{c}_o' = \mathbf{d}' \qquad \mathbf{d}' \triangleq \mathbf{Q}^H \mathbf{d}$$

- Since Λ is diagonal,

$$\lambda_i c'_{o,i} = d'_i \qquad 1 \le i \le M$$

- The MMSE becomes

$$\begin{aligned} P_o &= P_y - \mathbf{d}^H \mathbf{c}_o \\ &= P_y - (\mathbf{Q}\mathbf{d}')^H \mathbf{Q}\mathbf{c}'_o = P_y - \mathbf{d}'^H \mathbf{c}'_o \\ &= P_y - \sum_{i=1}^{M} d_i'^* c'_{o,i} = P_y - \sum_{i=1}^{M} \frac{|d'_i|^2}{\lambda_i} \end{aligned}$$

- Note that since all the eigen values are positive, increasing the estimator order will always result in lower estimation error.

Assignment 4.1

- Show that the estimation error is orthogonal to the data used for estimation.

4.4. LDL^H Decomposition of the Normal Equations

- By taking the Hermatian and positive definiteness of the autocorrelation matrix, it can be written as

$$\mathbf{R} = \mathbf{L}\mathbf{D}\mathbf{L}^H$$

$$\mathbf{L} \triangleq \begin{bmatrix} 1 & 0 & \cdots & 0 \\ l_{10} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{M-1,0} & l_{M-1,1} & \cdots & 1 \end{bmatrix}$$

$$\mathbf{D} = \text{diag}\{\xi_1, \xi_2, \ldots, \xi_M\}$$

- When the decomposition is known

$$\mathbf{R}\mathbf{c}_o = \mathbf{L}\mathbf{D}(\mathbf{L}^H\mathbf{c}_o) = \mathbf{d}$$

$$\mathbf{L}^H\mathbf{c}_o = \mathbf{k}$$

$$\mathbf{L}\mathbf{D}\mathbf{k} \triangleq \mathbf{d}$$

Solution is trivial

Example: when M=4

- LDL^H decomposition

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & r_{23} & r_{24} \\ r_{31} & r_{32} & r_{33} & r_{34} \\ r_{41} & r_{42} & r_{43} & r_{44} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ l_{10} & 1 & 0 & 0 \\ l_{20} & l_{21} & 1 & 0 \\ l_{30} & l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} \xi_1 & 0 & 0 & 0 \\ 0 & \xi_2 & 0 & 0 \\ 0 & 0 & \xi_3 & 0 \\ 0 & 0 & 0 & \xi_4 \end{bmatrix} \begin{bmatrix} 1 & l_{10}^* & l_{20}^* & l_{30}^* \\ 0 & 1 & l_{21}^* & l_{31}^* \\ 0 & 0 & 1 & l_{32}^* \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$r_{11} = \xi_1 \quad \Rightarrow \quad \xi_1 = r_{11}$$

$$r_{21} = \xi_1 l_{10} \quad \Rightarrow \quad l_{10} = \frac{r_{21}}{\xi_1}$$

$$r_{22} = \xi_1 |l_{10}|^2 + \xi_2 \quad \Rightarrow \quad \xi_2 = r_{22} - \xi_1 |l_{10}|^2$$

$$r_{31} = \xi_1 l_{20} \quad \Rightarrow \quad l_{20} = \frac{r_{31}}{\xi_1}$$

$$r_{32} = \xi_1 l_{20} l_{10}^* + \xi_2 l_{21} \quad \Rightarrow \quad l_{21} = \frac{r_{32} - \xi_1 l_{20} l_{10}^*}{\xi_2}$$

$$r_{33} = \xi_1 |l_{20}|^2 + \xi_2 |l_{21}|^2 + \xi_3 \quad \Rightarrow \quad \xi_3 = r_{33} - \xi_1 |l_{20}|^2 - \xi_2 |l_{21}|^2$$

$$r_{41} = \xi_1 l_{30} \quad \Rightarrow \quad l_{30} = \frac{r_{41}}{\xi_1}$$

$$r_{42} = \xi_1 l_{30} l_{10}^* + \xi_2 l_{31} \quad \Rightarrow \quad l_{31} = \frac{r_{42} - \xi_1 l_{30} l_{10}^*}{\xi_2}$$

$$r_{43} = \xi_1 l_{30} l_{20}^* + \xi_2 l_{31} l_{21}^* + \xi_3 l_{32} \quad \Rightarrow \quad l_{32} = \frac{r_{43} - \xi_1 l_{30} l_{20}^* - \xi_2 l_{31} l_{21}^*}{\xi_3}$$

$$r_{44} = \xi_1 |l_{30}|^2 + \xi_2 |l_{31}|^2 + \xi_3 |l_{32}|^2 + \xi_4 \quad \Rightarrow \quad \xi_4 = r_{44} - \xi_1 |l_{30}|^2 - \xi_2 |l_{31}|^2 - \xi_3 |l_{32}|^2$$

- Solving for k

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ l_{10} & 1 & 0 & 0 \\ l_{20} & l_{21} & 1 & 0 \\ l_{30} & l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} \xi_1 k_1 \\ \xi_2 k_2 \\ \xi_3 k_3 \\ \xi_4 k_4 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix}$$

$$\xi_1 k_1 = d_1 \Rightarrow k_1 = \frac{d_1}{\xi_1}$$

$$l_{10}\xi_1 k_1 + \xi_2 k_2 = d_2 \Rightarrow k_2 = \frac{d_2 - l_{10}\xi_1 k_1}{\xi_2}$$

$$l_{20}\xi_1 k_1 + l_{21}\xi_2 k_2 + \xi_3 k_3 = d_3 \Rightarrow k_3 = \frac{d_3 - l_{20}\xi_1 k_1 + l_{21}\xi_2 k_2}{\xi_3}$$

$$l_{30}\xi_1 k_1 + l_{31}\xi_2 k_2 + l_{32}\xi_3 k_3 + \xi_4 k_4 = d_4 \Rightarrow k_4 = \frac{d_4 - l_{30}\xi_1 k_1 + l_{31}\xi_2 k_2 + l_{32}\xi_3 k_3}{\xi_4}$$

Obtained in a forward way, Previous values of k do not depend on next values

- Solving for c

$$\begin{bmatrix} 1 & l_{10}^* & l_{20}^* & l_{30}^* \\ 0 & 1 & l_{21}^* & l_{31}^* \\ 0 & 0 & 1 & l_{32}^* \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1^{(4)} \\ c_2^{(4)} \\ c_3^{(4)} \\ c_4^{(4)} \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} \Rightarrow \begin{aligned} c_4^{(4)} &= k_4 \\ c_3^{(4)} &= k_3 - l_{32}^* c_4 \\ c_2^{(4)} &= k_2 - l_{21}^* c_3 - l_{31}^* c_4 \\ c_1^{(4)} &= k_1 - l_{10}^* c_2 - l_{20}^* c_3 - l_{30}^* c_4 \end{aligned}$$

Obtained in a backward way

Solution of normal equations using triangular decomposition.

For $i = 1, 2, \ldots, M$ and for $j = 0, 1, \ldots, i-1$,

$$l_{ij} = \frac{1}{\xi_i}\left(r_{i+1,j+1} - \sum_{m=0}^{j-1} \xi_{m+1} l_{im} l_{jm}^*\right) \qquad \text{(not executed when } i = M\text{)}$$

$$\xi_i = r_{ii} - \sum_{m=1}^{i-1} \xi_m |l_{i-1,m-1}|^2$$

For $i = 1, 2, \ldots, M$,

$$k_i = \frac{d_i}{\xi_i} - \sum_{m=0}^{i-2} l_{i-1,m} k_{m+1}$$

Obtained in a forward way, Previous values of k do not depend on next values

For $i = M, M-1, \ldots, 1$,

$$c_i = k_i - \sum_{m=i+1}^{M} l_{m-1,i-1}^* c_m$$

Obtained in a backward way

- By using the LDL^H decomposition, the MMSE can be obtained without using the optimum estimator coefficients

$$P_o = P_y - \mathbf{c}_o^H \mathbf{R} \mathbf{c}_o = P_y - \mathbf{k}^H \mathbf{L}^{-1} \mathbf{R} (\mathbf{L}^{-1})^H \mathbf{k} = P_y - \mathbf{k}^H \mathbf{D} \mathbf{k}$$

$$P_o = P_y - \sum_{i=1}^{M} \xi_i |k_i|^2$$

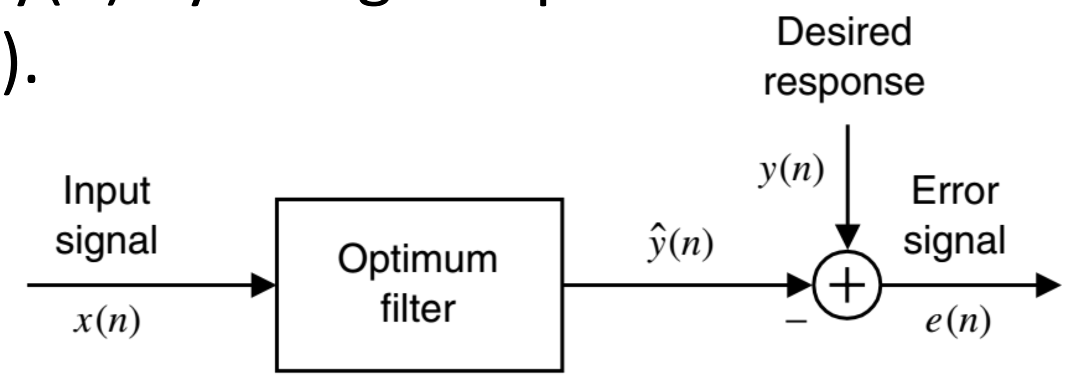
- Since $\xi_i > 0$, increasing the order of the filter can only reduce the MMSE.
  - Therefore better estimate.

A computationally effective linear MMSE estimation

1. $\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\}, \mathbf{d} = E\{\mathbf{x}y^*\}$	Normal equations $\mathbf{R}\mathbf{c}_o = \mathbf{d}$
2. $\mathbf{R} = \mathbf{L}\mathbf{D}\mathbf{L}^H$	Triangular decomposition
3. $\mathbf{L}\mathbf{D}\mathbf{k} = \mathbf{d}$	Forward substitution $\rightarrow \mathbf{k}$
4. $\mathbf{L}^H\mathbf{c}_o = \mathbf{k}$	Backward substitution $\rightarrow \mathbf{c}_o$
5. $P_o = P_y - \mathbf{k}^H\mathbf{D}\mathbf{k}$	MMSE computation
6. $e = y - \mathbf{c}_o^H\mathbf{x}$	Computation of residuals

4.5. Optimum FIR Filters

- The optimum filter forms an estimate of the desired response y(n) by using samples from a related input signal x(n).



$$\hat{y}(n) = \sum_{k=0}^{M-1} h(n,k)x(n-k)$$

$$\triangleq \sum_{k=1}^{M} c_k^*(n)x(n-k+1) \triangleq \mathbf{c}^H(n)\mathbf{x}(n)$$

$$\mathbf{c}(n) \triangleq [c_1(n)\ c_2(n)\ \cdots\ c_M(n)]^T$$

$$\mathbf{x}(n) \triangleq [x(n)\ x(n-1)\ \cdots\ x(n-M+1)]^T$$

- Note also that, the above problem is identical to the problem of linear MSE estimation.
- Therefore,

$$\mathbf{R}(n)\mathbf{c}_o(n) = \mathbf{d}(n)$$

$$\mathbf{R}(n) \triangleq E\{\mathbf{x}(n)\mathbf{x}^H(n)\}$$ Hermitian, but not Toeplitz

$$\mathbf{d}(n) \triangleq E\{\mathbf{x}(n)y^*(n)\}$$

$$P_o(n) = P_y(n) - \mathbf{d}^H(n)\mathbf{c}_o(n)$$

Stationary Wiener Filter

- If the input and desired response are jointly wide-sense stationary, the correlation matrix and cross-correlation vector do not depend on the index n,

$$\mathbf{R}\mathbf{c}_o = \mathbf{d}$$

Wiener-Hopf equation

$$P_o = P_y - \mathbf{d}^H \mathbf{c}_o$$

$$\mathbf{R} \triangleq \begin{bmatrix} r_x(0) & r_x(1) & \cdots & r_x(M-1) \\ r_x^*(1) & r_x(0) & \cdots & r_x(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_x^*(M-1) & r_x^*(M-2) & \cdots & r_x(0) \end{bmatrix}$$

Hermitian and Toeplitz

$$\mathbf{d} \triangleq [d_1 \; d_2 \; \cdots \; d_M]^T \triangleq [r_{yx}^*(0) \; r_{yx}^*(1) \; \cdots \; r_{yx}^*(M-1)]^T$$

Comparison with conventional frequency selective filters

Conventional frequency-selective filters	Optimum Filters
Designed to shape the spectrum of the input signal within a specific frequency band in which it operates.	Designed using the second-order moments of the processed signals and have the same effect on all classes of signals with the same second-order moments.
Effective only if the components of interest in the input signal have their energy concentrated within non-overlapping bands.	Effective even if the signals of interest have overlapping spectra
To design the filters, we need to know the limits of these bands, not the values of the sequences to be filtered.	These filters are optimized to the statistics of the data and thus provide superior performance when judged by the statistical criterion.

Optimum Infinite Impulse Response Filters

- The normal equations for optimum IIR filters are the same for FIR filters;
  - Only the limits in the convolution summation and the range of values for which the normal equations hold are different.
- Both are determined by the limits of summation in the filter convolution equation.

$$\hat{y}(n) = \sum_{k} h_o(k) x(n-k)$$

$$\sum_{k} h_o(k) r_x(m-k) = r_{yx}(m)$$

$$P_o = r_y(0) - \sum_{k} h_o(k) r_{yx}^*(k)$$

Only analytical solutions are possible due to infinite summation

Non-causal IIR Filters

- By using the Z-transform, for non-causal IIR filters, the convolution becomes

$$H_{\mathrm{nc}}(z)R_x(z) = R_{yx}(z)$$

$$H_{\mathrm{nc}}(z) = \frac{R_{yx}(z)}{R_x(z)}$$

Causal IIR Filters

- Since the convolution is defined for positive index values for causal filters, Z-transform cannot be used

$$\hat{y}(n) = \sum_{k=0}^{\infty} h_c(k)x(n-k)$$

- However, if the input is white, the solution is trivial.

$$r_x(l) = \sigma_x^2 \delta(l)$$

$$h_c(m) * \delta(m) = \frac{r_{yx}(m)}{\sigma_x^2} \qquad 0 \le m < \infty$$

- Since it is causal

$$H_c(z) = \frac{1}{\sigma_x^2}[R_{yx}(z)]_+$$

$$[R_{yx}(z)]_+ \triangleq \sum_{l=0}^{\infty} r_{yx}(l)z^{-l}$$

Construction of the Wiener Filter by Prewhitening

- If we are estimating the signal from white-noise sequence with delta-function autocorrelation, the solution would be much simpler.

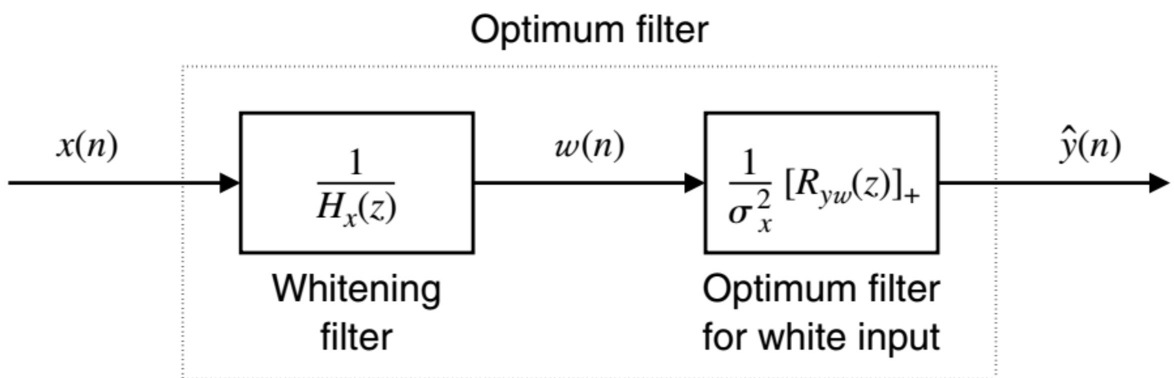
$$r_x(l) = \sigma_x^2 \delta(l)$$

- Then

$$h_c(m) * \delta(m) = \frac{r_{yx}(m)}{\sigma_x^2} \qquad 0 \le m < \infty$$

$$h_c(m) = \begin{cases} \dfrac{1}{\sigma_x^2} r_{yx}(m) & 0 \le m < \infty \\ 0 & m < 0 \end{cases}$$

- By using the fact that any random process can be modeled as the output of an LTI system with white-noise excitation,



- Where, the whitening filter has to satisfy

$$R_x(z) = \sigma_x^2 H_x(z) H_x^*\left(\frac{1}{z^*}\right)$$

Spectral factorization

- Then y(n) is estimated from the white noise with

$$H_c^{'}(z) = \frac{1}{\sigma_x^2}[R_{yw}(z)]_+$$

  – This is the one-sided z-transform of $r_{yw}(l)$.

- To obtain $R_{yw}(Z)$

$$E\{y(n)x^*(n-l)\} = \sum_{k=0}^{\infty} h_x^*(k) E\{y(n)w^*(n-l-k)\}$$

$$r_{yx}(l) = \sum_{k=0}^{\infty} h_x^*(k) r_{yw}(l+k)$$

$$R_{yw}(z) = \frac{R_{yx}(z)}{H_x^*(1/z^*)}$$

$$H_c^{'}(z) = \frac{1}{\sigma_x^2}\left[\frac{R_{yx}(z)}{H_x^*(1/z^*)}\right]_+$$

TABLE 6.5

Design of FIR and IIR optimum filters for stationary processes.

Filter type	Solution	Required quantities
FIR	$e(n) = y(n) - \mathbf{c}_o^H \mathbf{x}(n)$ $\mathbf{c}_o = \mathbf{R}^{-1}\mathbf{d}$ $P_o = r_y(0) - \mathbf{d}^H \mathbf{c}_o$	$\mathbf{R} = [r_x(m-k)]$, $\mathbf{d} = [r_{yx}(m)]$ $0 \le k, m \le M-1$, $M = \text{finite}$
Noncausal IIR	$H_{\text{nc}}(z) = \dfrac{R_{yx}(z)}{R_x(z)}$	$R_x(z) = \mathcal{Z}\{r_x(l)\}$ $R_{yx}(z) = \mathcal{Z}\{r_{xy}^*(l)\}$
	$P_{\text{nc}} = r_y(0) - \sum_{k=-\infty}^{\infty} h_{\text{nc}}(k) r_{yx}^*(k)$	
Causal IIR	$H_{\text{c}}(z) = \dfrac{1}{\sigma_x^2 H_x(z)} \left[\dfrac{R_{yx}(z)}{H_x^*(1/z^*)} \right]_+$	$R_x(z) = \sigma_x^2 H_x(z) H_x^*(1/z^*)$
	$P_{\text{c}} = r_y(0) - \sum_{k=0}^{\infty} h_{\text{nc}}(k) r_{yx}^*(k)$	$R_{yx}(z) = \mathcal{Z}\{r_{xy}(l)\}$

4.6. Filtering Additive Noise

- Assume a signal has been corrupted by additive noise and it is required to estimate the undegraded signal

$$x(n) = y(n) + v(n)$$

  - Where, the signal and noise are uncorrelated and zero mean.
  - They have known autocorrelation sequences $r_y(l)$ and $r_v(l)$.

- To design the optimum filter, the autocorrelation matrix $r_x(l)$ and the cross-correlation $r_{yx}(l)$ are required.

$$r_x(l) = E\{x(n)x^*(n-l)\} = r_y(l) + r_v(l)$$

$$r_{yx}(l) = E\{y(n)x^*(n-l)\} = r_y(l)$$

- Taking the z-transform of the above equations

$$R_x(z) = R_y(z) + R_v(z)$$

$$R_{yx}(z) = R_y(z)$$

- The noncausal optimum filter is then given by

$$H_{\mathrm{nc}}(z) = \frac{R_{yx}(z)}{R_x(z)} = \frac{R_y(z)}{R_y(z) + R_v(z)}$$

Example:

- Extract a random signal with known autocorrelation sequence which is corrupted by additive white noise.

$$r_y(l) = \alpha^{|l|} \qquad -1 < \alpha < 1$$
$$r_v(l) = \sigma_v^2 \delta(l)$$

The signal and noise are uncorrelated.

$\alpha = 4/5$ and

variance of noise$=1$

- Solution:
  - Since they are uncorrelated

$$r_x(l) = \alpha^{|l|} + \sigma_v^2 \delta(l)$$

$$r_{yx}(l) = \alpha^{|l|}$$

  - The complex PSD are

$$R_y(z) = \frac{(\frac{3}{5})^2}{(1 - \frac{4}{5}z^{-1})(1 - \frac{4}{5}z)} \qquad \frac{4}{5} < |z| < \frac{5}{4}$$

$$R_v(z) = \sigma_v^2 = 1$$

$$R_x(z) = \frac{8}{5}\frac{(1 - \frac{1}{2}z^{-1})(1 - \frac{1}{2}z)}{(1 - \frac{4}{5}z^{-1})(1 - \frac{4}{5}z)}$$

- Non-causal filter is

$$H_{\mathrm{nc}}(z) = \frac{R_{yx}(z)}{R_x(z)} = \frac{9}{40}\frac{1}{(1-\frac{1}{2}z^{-1})(1-\frac{1}{2}z)} \qquad \frac{1}{2} < |z| < 2$$

$$h_{\mathrm{nc}}(n) = \tfrac{3}{10}(\tfrac{1}{2})^{|n|} \qquad -\infty < n < \infty$$

- The MMSE is then

$$P_{\mathrm{nc}} = 1 - \tfrac{3}{10}\sum_{k=-\infty}^{\infty}(\tfrac{1}{2})^{|k|}(\tfrac{4}{5})^{|k|} = \tfrac{3}{10}$$

- Causal filter

$$\sigma_x^2 = \frac{8}{5}$$

$$R_x(z) = \sigma_x^2 H_x(z) H_x(z^{-1}) \qquad H_x(z) = \frac{1 - \frac{1}{2}z^{-1}}{1 - \frac{4}{5}z^{-1}}$$

$$R_{yw}(z) = \frac{R_{yx}(z)}{H_x(z^{-1})} = \frac{0.36}{(1 - \frac{4}{5}z^{-1})(1 - \frac{1}{2}z)} = \frac{0.6}{1 - \frac{4}{5}z^{-1}} + \frac{0.3z}{1 - \frac{1}{2}z}$$

  - Taking the causal part

$$\left[\frac{R_{yx}(z)}{H_x(z^{-1})}\right]_+ = \frac{\frac{3}{5}}{1 - \frac{4}{5}z^{-1}}$$

$$H_c(z) = \frac{1}{\sigma_x^2 H_x(z)} \left[\frac{R_{yx}(z)}{H_x^*(1/z^*)}\right]_+$$

$$H_c(z) = \frac{5}{8}\left(\frac{1 - \frac{4}{5}z^{-1}}{1 - \frac{1}{2}z^{-1}} \frac{\frac{3}{5}}{1 - \frac{4}{5}z^{-1}}\right) = \frac{3}{8}\left(\frac{1}{1 - \frac{1}{2}z^{-1}}\right) \qquad |z| < \frac{1}{2}$$

$$h_c(n) = \frac{3}{8}(\tfrac{1}{2})^n u(n)$$

$$P_c = r_y(0) - \sum_{k=0}^{\infty} h_c(k) r_{yx}(k) = 1 - \frac{3}{8}\sum_{k=0}^{\infty} (\tfrac{1}{2})^k (\tfrac{4}{5})^k = \frac{3}{8}$$

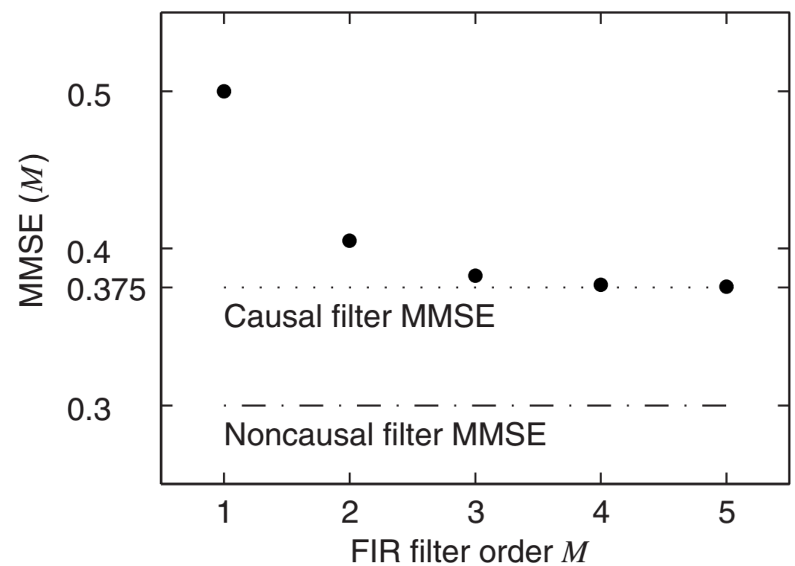
- Mth-order FIR filter

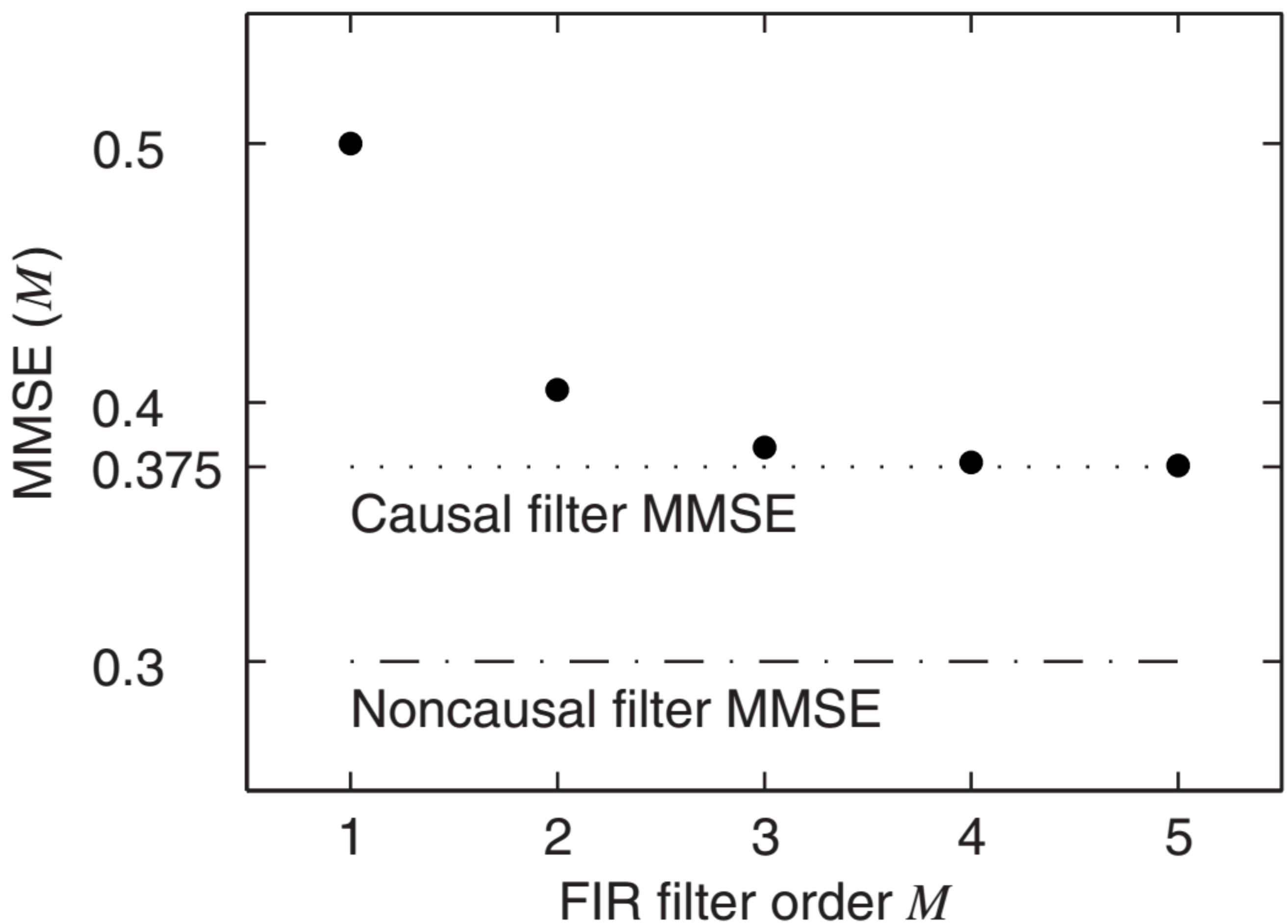
$$\mathbf{Rh} = \mathbf{d}$$

$$\mathbf{R} = \text{Toeplitz}(1 + \sigma_v^2, \alpha, \ldots, \alpha^{M-1})$$

$$\mathbf{d} = [1\, \alpha \ \cdots \ \alpha^{M-1}]^T$$

$$P_o = r_y(0) - \sum_{k=0}^{M-1} h_o(k) r_{yx}(k)$$




0.5
0.4
0.375
0.3
MMSE (M)
Causal filter MMSE
Noncausal filter MMSE
1
2
3
4
5
FIR filter order M

Assignment 4.2

- Suppose it is desired to estimate a signal x_n on a basis of noisy observation

$$y_n = x_n + v_n$$

  - The noise is white noise with 1 variance and uncorrelated with $x_{n.}$
  - Suppose the signal x_n is a first order Markov process with variance of w 0.82.

$$x_{n+1} = 0.6x_n + w_n$$

- Design the optimum linear filter.

Assignment 4.3

- A random process x(n) is said to be exactly predictable if

$$P_e = E\{|e^{\mathrm{f}}(n)|^2\} = 0$$

- Show that a random process is exactly predictable if its PSD consists of impulses.

$$R_x(e^{j\omega}) = \sum_k A_k \delta(\omega - \omega_k)$$