3 Optimal and Adaptive Filtering

3.1: Wiener-Hopf filter

Optimal and Adaptive Filtering

3.1

1. **Wiener-Hopf filter**
    - Minimum Mean Square Error Estimation
    - The Wiener-Hopf solution
2. **Applications of the Wiener-Hopf filter**
    - Interference cancelation in biological signals
    - Linear prediction for signal coding
3. **Adaptive filtering**
    - Steepest descend
    - Least Mean Square approach
4. **Applications of adaptive filtering**
    - …

Wiener-Hopf filter

3.1

1. **Introduction**
   - Problem modelling and filter configuration
2. **Minimum Mean Square Error (MMSE) prediction**
   - Principle of orthogonality
   - Some results of the MMSE prediction
3. **The Wiener-Hopf filter**
   - The Wiener-Hopf solution
   - The error performance surface
   - The Wiener-Hopf filter using a finite number of samples
4. **Conclusions**

Usefulness of the Wiener-Hopf filter

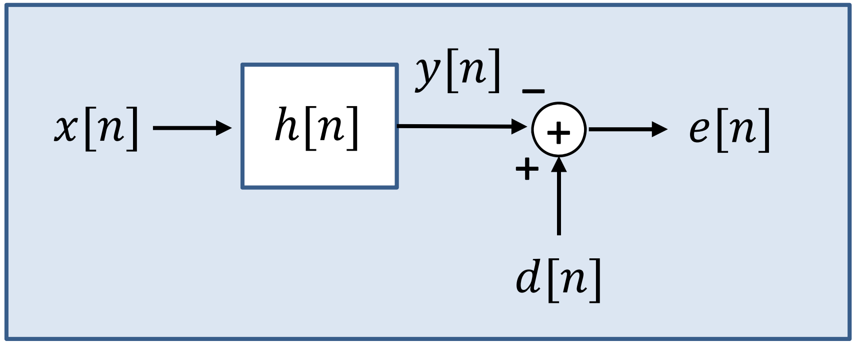
3.1

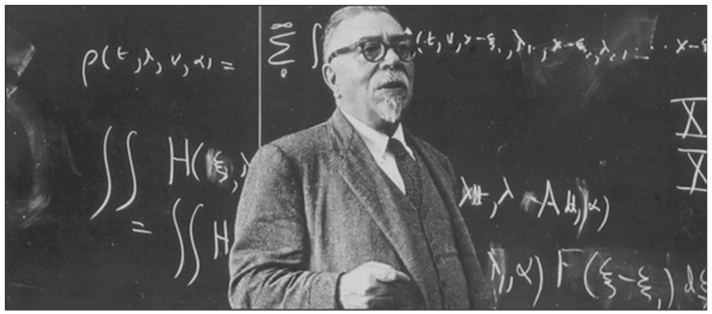
Several **estimation problems** can be modeled relying on a similar formulation:

Given a set of data from an observed noisy process (observations $x[n]$) and a desired target process that we want to estimate (reference $d[n]$), produce an estimated of the target process (estimation $y[n]$) by linear time-invariant (LTI) filtering ($T[n] = h[n]$) of the observed samples.

We assume **known stationary signal and noise spectra** (**correlation**) as well as **additive noise**.

Note: We will assume **FIR filters** and, in the second part of the Unit, **non-stationary** scenarios





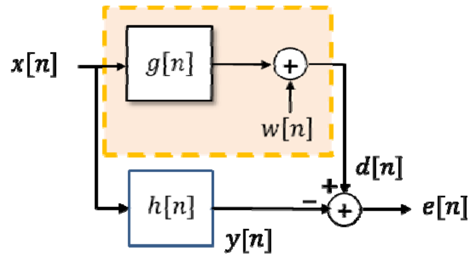
Norbert Wiener: Research Laboratory of Electronics MIT

Filter configuration

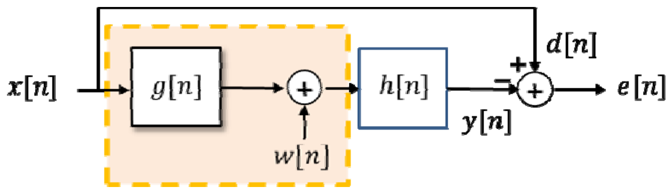
3.1

This formulation can be applied to a large family of problems that are commonly sorted into four wide classes:

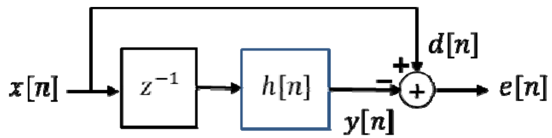
- **System identification**:
  - Noisy reference
  - Noise-free observations



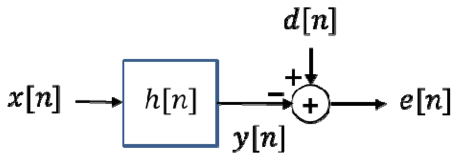
- **System inversion**:
  - Noisy observations
  - Noise-free reference



- **Signal prediction**:
  - Observations and reference are samples of the same noisy process



- **Signal cancellation**:
  - Noisy observations with interferences
  - Noisy interferences as reference(s)



System identification

3.1

System identification: We want to identify a given system, that can be real or some abstraction of a complex nature.

We **model** this system as an LTI system plus an additive noise source ($w[n]$).

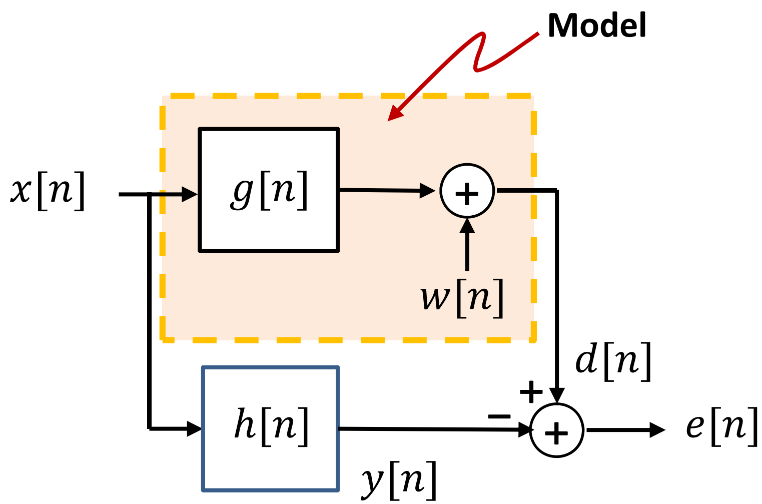
Design and Use: We excite the system with a known signal ($x[n]$) and obtain the filter that models the system.

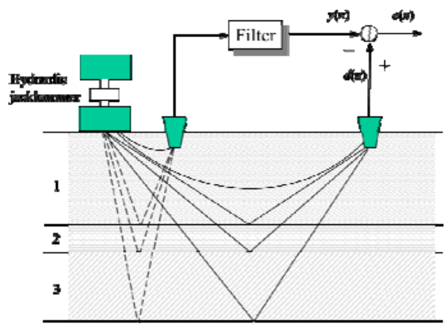
The application assumes:

- Noisy reference
- Noise-free observations

Example of application:

- Geological prospections





The features of the various geological layers can be estimated through the modeling of a wave distortion

System inversion

System inversion: We want to estimate a system an apply its inverse to the signal.

We **model** this system as an LTI system plus an additive noise source ($w[n]$).

Design: we excite the system with a known signal ($x[n]$) and obtain the filter that models the system.

Use: The filter is concatenated to the system to recover the estimated signal

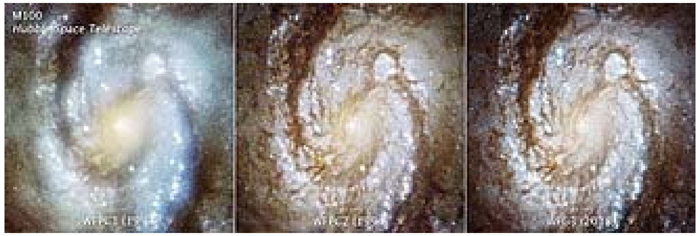
The application assumes:

- Noisy observations
- Noise-free reference

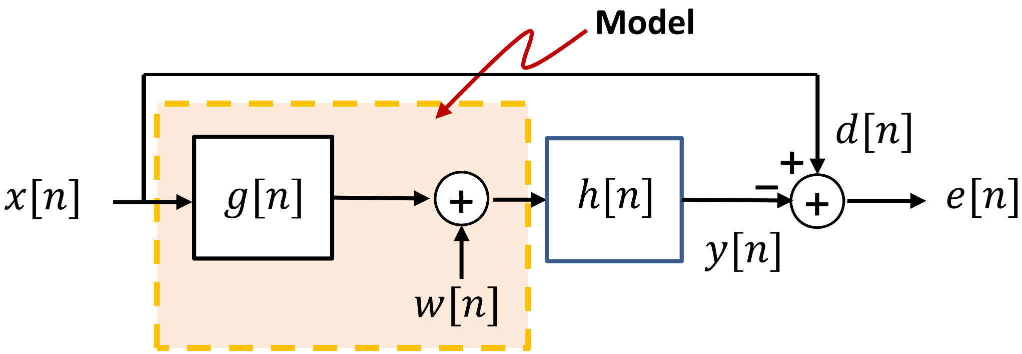
Example of application:

- Optical deconvolution

S. Bikkannavar and D. Redding, "Software for Optical Systems Spells the End of Blur" IEEE Spectrum, Feb. 2010



Optical evolution of Hubble's primary camera system: Spiral galaxy M100 as seen with WFPC1 in 1993 before corrective optics (left), with WFPC2 in 1994 after correction (center), and with WFC3 in 2018 (right)



Signal prediction

Signal prediction: We estimate the value of a random signal at a given time instance ($x[n_0]$) , based on other time instance values (e.g.: $x[n_0 - 1], x[n_0 - 2], \cdots$) .

Design: We compare the current signal value ($x[n_0]$) with its estimation ($y[n_0]$)

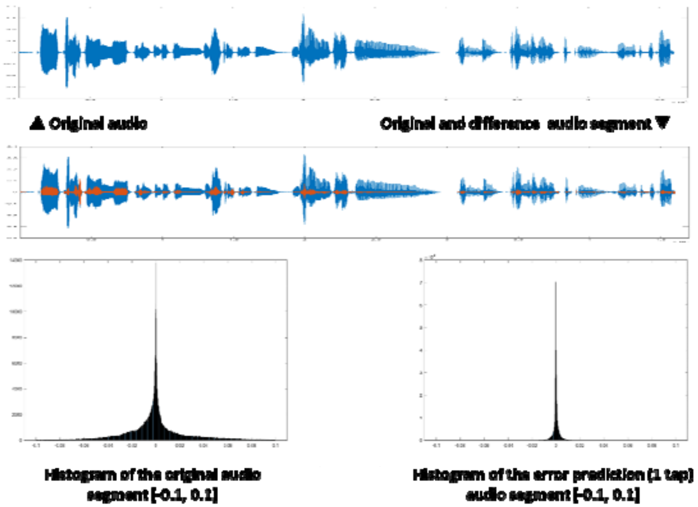
Use: The current signal value ($x[n_0]$) may not be available and we produce an estimation. If $x[n_0]$ is available, we produce the prediction error ($e[n_0]$)

The application assumes:

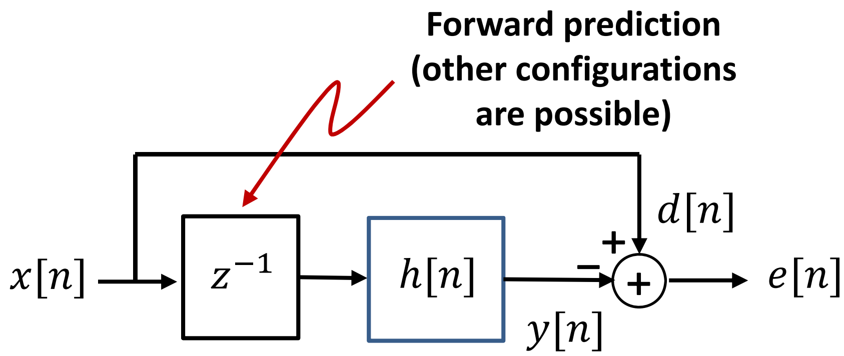
- Observations and reference belong to the same noisy process

Example of application:

- Speech coding and synthesis



The prediction error has a lower dynamic range and its quantization decreases the quantization noise power



Signal cancelation

3.1

Signal cancelation: We estimate the value of a primary signal which contains an interference. This interference has been isolated through other sensors in additional signals.

Design: We compare the primary signal ($d[n]$) with the interference ($x[n]$)

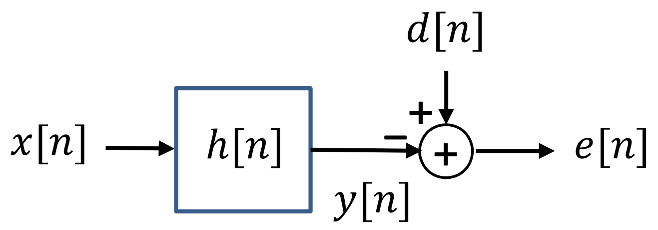
Use: We obtain the clean signal as the estimation error ($e[n]$)

The application assumes:

- Noisy interferences as observations
- Noisy signal and interferences as reference

Example of application:

- Interference cancelation



The sound of the engine interferes with the pilot communications

http://www.wildlandfirefighter.com/2019/11/21/meet-cal-fires-first-female-helicopter-pilot/

Wiener-Hopf filter

1. **Introduction**
   - Problem modelling and filter configuration
2. **Minimum Mean Square Error (MMSE) prediction**
   - Principle of orthogonality
   - Some results of the MMSE prediction
3. **The Wiener-Hopf filter**
   - The Wiener-Hopf solution
   - The error performance surface
   - The Wiener-Hopf filter using a finite number of samples
4. **Conclusions**

Minimum MSE prediction

Given the generic formulation, we restrict the analysis to the **FIR filter** case:

- It is the **optimal solution** if $x[n]$ and $d[n]$ are Gaussian jointly distributed processes.
- The filter is assumed to have $\boldsymbol{N}$ **coefficients**

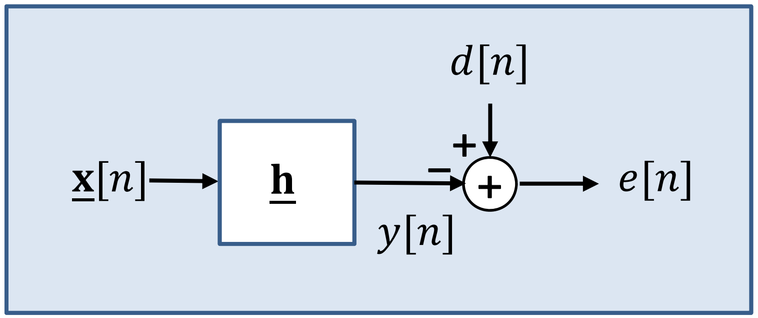
$$h[n] * x[n] = \underline{\mathbf{h}}^T \underline{\mathbf{x}}[n]$$

$$\underline{\mathbf{x}}[n] = \begin{bmatrix} x[n] \\ x[n-1] \\ \cdots \\ x[n-N+1] \end{bmatrix}$$

The **MSE** is used as optimization criterion:

- It is mathematically tractable.
- It leads to useful solutions for real applications
- It can be used as benchmark for other solutions

$$\min_{\underline{\mathbf{h}}} E\{e[n]^2\}$$



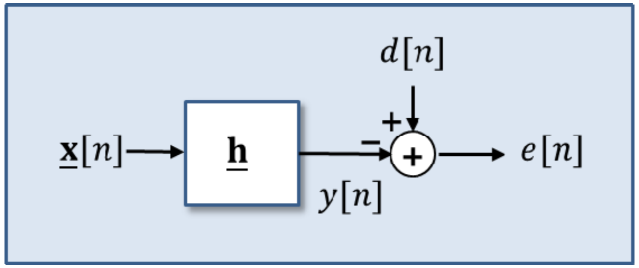
$$e[n] = d[n] - y[n] = d[n] - \underline{\mathbf{h}}^T \underline{\mathbf{x}}[n]$$

$$\boxed{\min_{\underline{\mathbf{h}}} E\{e[n]^2\} = \min_{\underline{\mathbf{h}}} E\left\{\left(d[n] - \underline{\mathbf{h}}^T \underline{\mathbf{x}}[n]\right)^2\right\}}$$

Principle of orthogonality

3.1

The **minimization** of the Mean Square Error implies:



$$\min_{\underline{\mathbf{h}}} E\{e[n]^2\} \quad \Rightarrow \quad \nabla_{\underline{\mathbf{h}}} E\{e[n]^2\} = \underline{0}$$

Note: We assume that the observations and the reference ($x[n]$ and $d[n]$) have zero mean.

$$\nabla_{\underline{\mathbf{h}}} E\left\{\left(d[n] - \underline{\mathbf{h}}^T \underline{\mathbf{x}}[n]\right)^2\right\} = \underline{0}$$

$$E\left\{\nabla_{\underline{\mathbf{h}}} \left(d[n] - \underline{\mathbf{h}}^T \underline{\mathbf{x}}[n]\right)^2\right\} = E\{2\left(d[n] - \underline{\mathbf{h}}^T \underline{\mathbf{x}}[n]\right)\underline{\mathbf{x}}[n](-1)\} = \underline{0}$$

$$\nabla_{\underline{\mathbf{h}}} E\{e[n]^2\} = \underline{0} \quad \Rightarrow \quad \boxed{E\{e[n]\, \underline{\mathbf{x}}[n]\} = \underline{0}}$$

The **error** is said to be **orthogonal to the observations**

Wiener-Hopf filter

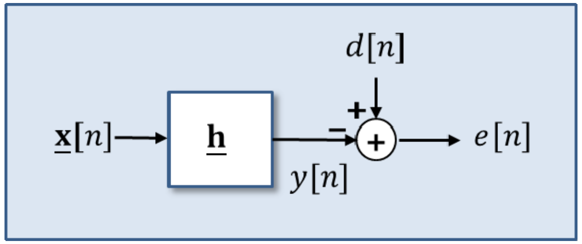
3.1

1. **Introduction**
   - Problem modelling and filter configuration
2. **Minimum Mean Square Error (MMSE) prediction**
   - Principle of orthogonality
   - Some results of the MMSE prediction
3. **The Wiener-Hopf filter**
   - The Wiener-Hopf solution
   - The error performance surface
   - The Wiener-Hopf filter using a finite number of samples
4. **Conclusions**

Some results of the MMSE prediction

3.1

In order to analyze some results of the MMSE prediction, let us define a **specific signal scenario**:



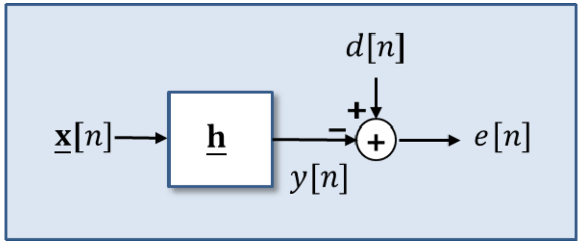
- The **observation process** can be split into two parts: $x[n] = a[n] + b[n]$
- The **reference process** can be split into two parts: $d[n] = a'[n] + c[n]$
- These parts have the following **correlation properties**:
  - $r_{ab}[l] = E\{a[n+l]b[n]\} = 0$
  - $r_{a'c}[l] = E\{a'[n+l]c[n]\} = 0$
  - $r_{aa'}[l] = E\{a[n+l]a'[n]\} \neq 0$ $\Leftarrow$ **The only two parts that are correlated**
  - $r_{ac}[l] = E\{a[n+l]c[n]\} = 0$
  - $r_{a'b}[l] = E\{a'[n+l]b[n]\} = 0$
  - $r_{bc}[l] = E\{b[n+l]c[n]\} = 0$

❑ **How does the optimal filter behave in this scenario?**

Some results of the MMSE prediction

3.1

When using the filter that minimizes the MSE ($\underline{\mathbf{h}}_{opt}$), the following property stands:



Note: Analyze a generic case and the specific previous one:
$x[n] = a[n] + b[n]$ and $d[n] = a'[n] + c[n]$

- At a time instance, the **estimation** and the **error signals** are not correlated:

$$E\{y[n]e[n]\} = 0$$

a) $E\{e[n]\cdot y[n]\} = [\, y[n] = \underline{h}_{opt}^T \, \underline{x}[n] \,] =$

$= E\{e[n]\cdot \underline{h}_{opt}^T \cdot \underline{x}[n]\} = [\, \underline{h}^T \text{ DET.} \,]$

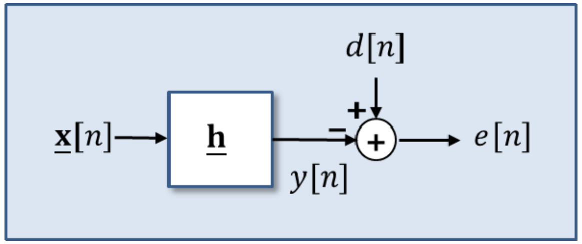
$= \underline{h}^T E\{e[n]\cdot \underline{x}[n]\} = [\text{NORMAL EQ.}] = \emptyset$

Some results of the MMSE prediction

3.1

When using the filter that minimizes the MSE ($\underline{\mathbf{h}}_{opt}$), the following property stands:

Note: Analyze a generic case and the specific previous one: $x[n] = a[n] + b[n]$ and $d[n] = a'[n] + c[n]$



- The **variance of the reference** signal is greater than or equal to the **variance of the error** signal:

$$E\{(d[n])^2\} \geq E\{(e[n])^2\}$$

$$b)\quad E\{(d[n])^2\} = [d[n] = y[n] + e[n]] = E\{(y[n]+e[n])^2\} =$$

$$= E\{y^2[n]\} + 2E\{y[n]\cdot e[n]\} + E\{e^2[n]\} = [\underline{h} = \underline{h}_{opt}] =$$

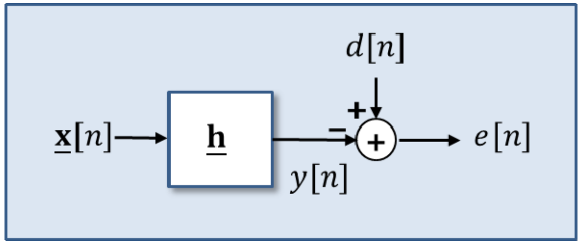
$$= [E\{y[n]e[n]\} = 0] = E\{y^2[n]\} + E\{e^2[n]\} \geq E\{e^2[n]\}$$

Some results of the MMSE prediction

3.1

When using the filter that minimizes the MSE ($\underline{\mathbf{h}}_{opt}$), the following property stands:

Note: Analyze a generic case and the specific previous one: $x[n] = a[n] + b[n]$ and $d[n] = a'[n] + c[n]$



- If the observation and the reference signals are not correlated, the **variance of the estimation** is zero:

$$E\{\underline{\mathbf{x}}[n]d[n]\} = \underline{0} \quad \Rightarrow \quad E\{(y[n])^2\} = 0$$

a) $E\{y^2[n]\} = E\{y[n]\,(d[n] - e[n])\} =$

$= E\{y[n]d[n]\} - E\{y[n]e[n]\} = [E\{y[n]e[n]\} = 0] =$

$= E\{y[n]d[n]\} = [y[n] = \underline{h}^T_{opt}\,\underline{x}[n]] = E\{\underline{h}^T_{opt}\,\underline{x}[n]\,d[n]\} =$

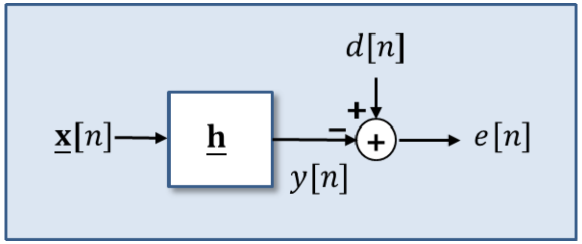
$= [\underline{h}^T \text{ DET}] = \underline{h}^T_{opt}\,E\{\underline{x}[n]\,d[n]\} = [E\{\underline{x}[n]\,d[n]\} = \underline{0}] = 0.$

Some results of the MMSE prediction

3.1

When using the filter that minimizes the MSE ($\underline{\mathbf{h}}_{opt}$), the following property stands:

Note: Analyze a generic case and the specific previous one: $x[n] = a[n] + b[n]$ and $d[n] = a'[n] + c[n]$



- The **minimum variance** (power) of the error signal is:

$$\varepsilon = r_d[0] - \underline{\mathbf{h}}_{opt}^T \, \underline{\mathbf{r}}_{xd}$$

d) $E\{(e[n])^2\}\big|_{min} = E\{e[n] \cdot (d[n] - y[n])\} =$

$= E\{e[n]\, d[n]\} - E\{e[n]\, y[n]\} = \left[E\{e[n]\, y[n]\} = \emptyset \right] =$

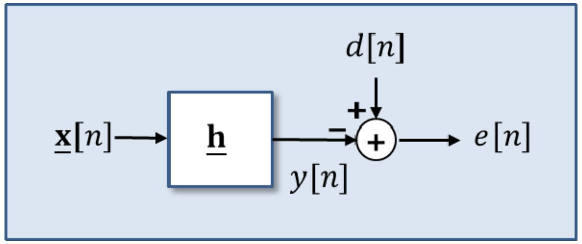
$= E\{e[n]\, d[n]\} = \left[e[n] = d[n] - \underline{h}_{opt}^T \underline{x}[n] \right] =$

$= E\{(d[n] - \underline{h}_{opt}^T \underline{x}[n])\, d[n]\} = E\{d^2[n]\} - E\{\underline{h}_{opt}^T \underline{x}[n]\, d[n]\} =$

Some results of the MMSE prediction

3.1

The conditions on the processes of the previous **signal scenario** can be **redefined** (relaxed) taken into account the **actual samples** involved in the filtering problem:



- The **observation process** can be split into two parts: $x[n] = a[n] + b[n]$
- The **reference process** can be split into two parts: $d[n] = a'[n] + c[n]$
- These parts have the following **correlation properties**:

- $r_{ab}[l] = E\{a[n+l]b[n]\} = 0$
- $r_{a'c}[l] = E\{a'[n+l]c[n]\} = 0$
- $r_{aa'}[l] = E\{a[n+l]a'[n]\} \neq 0$
- $r_{ac}[l] = E\{a[n+l]c[n]\} = 0$
- $r_{a'b}[l] = E\{a'[n+l]b[n]\} = 0$
- $r_{bc}[l] = E\{b[n+l]c[n]\} = 0$

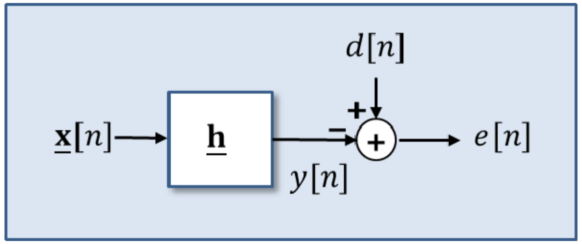
- $E\{\underline{\mathbf{a}}[n]\,\underline{\mathbf{b}}^T[n]\} = \underline{\underline{\mathbf{0}}}$
- $E\{a'[n]c[n]\} = 0$
- $E\{\underline{\mathbf{a}}[n]\,a'[n]\} \neq \underline{\mathbf{0}}$
- $E\{\underline{\mathbf{a}}[n]\,c[n]\} = \underline{\mathbf{0}}$
- $E\{\underline{\mathbf{b}}[n]\,a'[n]\} = \underline{\mathbf{0}}$
- $E\{\underline{\mathbf{b}}[n]\,c[n]\} = \underline{\mathbf{0}}$

Wiener-Hopf filter

1. **Introduction**
   - Problem modelling and filter configuration
2. **Minimum Mean Square Error (MMSE) prediction**
   - Principle of orthogonality
   - Some results of the MMSE prediction
3. **The Wiener-Hopf filter**
   - The Wiener-Hopf solution
   - The error performance surface
   - The Wiener-Hopf filter using a finite number of samples
4. **Conclusions**

The Wiener-Hopf solution

So far, we have analyze some properties of the optimal filter, but not yet obtain it:



$$e[n] = d[n] - \underline{\mathbf{h}}^T \underline{\mathbf{x}}[n]$$

$$E\{\, \underline{\mathbf{x}}[n] e[n] \} = \underline{\mathbf{0}}$$

$$\Rightarrow \quad E\{\underline{\mathbf{x}}[n]\left(d[n] - \underline{\mathbf{h}}^T \underline{\mathbf{x}}[n]\right)\} = \underline{\mathbf{0}}$$

$$E\{\underline{\mathbf{x}}[n]\left(d[n] - \underline{\mathbf{h}}^T \underline{\mathbf{x}}[n]\right)\} = E\{\underline{\mathbf{x}}[n] d[n]\} - E\{\underline{\mathbf{x}}[n] \underline{\mathbf{h}}^T \underline{\mathbf{x}}[n]\} = \underline{\mathbf{0}}$$

$$E\{\underline{\mathbf{x}}[n] d[n]\} - E\{\underline{\mathbf{x}}[n]\, \underline{\mathbf{x}}^T[n]\, \underline{\mathbf{h}}\} = \underline{r}_{xd}[0] - E\{\underline{\mathbf{x}}[n]\, \underline{\mathbf{x}}^T[n]\}\underline{\mathbf{h}} = \underline{\mathbf{0}}$$

$$\underline{\mathbf{r}}_{xd}[0] - \underline{\underline{\mathbf{R}}}_x[0]\underline{\mathbf{h}} = \underline{\mathbf{0}}$$

$$\boxed{\underline{\mathbf{h}}_{opt} = \underline{\underline{\mathbf{R}}}_x^{-1} \underline{\mathbf{r}}_{xd}}$$

◄ Commonly, we drop the evaluation in 0 in the notation

The Wiener-Hopf solution

3.1

The **optimal filter** in the sense of the MSE criterion is:

$$\boxed{\underline{\mathbf{h}}_{opt} = \underline{\underline{\mathbf{R}}}_x^{-1}\underline{\mathbf{r}}_{xd}}$$

$$\underline{\mathbf{r}}_{xd} = E\{\underline{\mathbf{x}}[n]d[n]\} = \begin{bmatrix} E\{x[n]\,d[n]\} \\ E\{x[n-1]\,d[n]\} \\ \cdots \\ E\{x[n-N+1]\,d[n]\} \end{bmatrix} = \begin{bmatrix} r_{xd}[0] \\ r_{xd}[-1] \\ \cdots \\ r_{xd}[-N+1] \end{bmatrix}$$

◀ **Cross-correlation vector**

$$\underline{\underline{\mathbf{R}}}_x = E\{\underline{\mathbf{x}}[n]\,\underline{\mathbf{x}}^T[n]\} = \begin{bmatrix} r_x[0] & r_x[1] & & r_x[N-1] \\ r_x[-1] & r_x[0] & \cdots & r_x[N-2] \\ \cdots & \cdots & & \cdots \\ r_x[-N+1] & r_x[-N+2] & & r_x[0] \end{bmatrix}$$

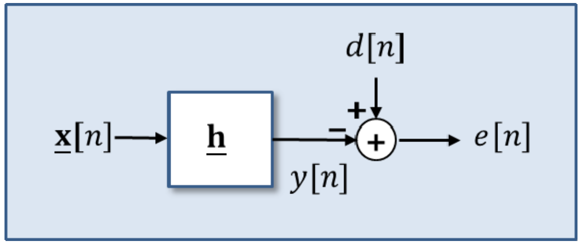
◀ **Correlation matrix**

The **optimal filter** depends on the second order statistics of the processes:

- We will further analyze the **properties of the correlation matrix**
- We will study how to proceed **when such statistics are not available**

The error performance surface

The Wiener-Hopf filter is optimal in the sense that it **minimizes the MSE of the prediction**; that is, the variance (power) of $e[n]$.



- For any filter, the MSE can be expressed as:

$$E\{(e[n])^2\} = \varepsilon + \left(\underline{\mathbf{h}}_{opt} - \underline{\mathbf{h}}\right)^T \underline{\underline{\mathbf{R}}}_x \left(\underline{\mathbf{h}}_{opt} - \underline{\mathbf{h}}\right)$$

The error performance surface

3.1

$$E\{(e[n])^2\} = \varepsilon + \left(\underline{\mathbf{h}}_{opt} - \underline{\mathbf{h}}\right)^T \underline{\underline{\mathbf{R}}}_x \left(\underline{\mathbf{h}}_{opt} - \underline{\mathbf{h}}\right)$$

$$= r_d[0] - 2\,\underline{h}^T \underline{r}_{xd} + \underline{h}^T \underline{\underline{R}}_x \underline{h} =$$

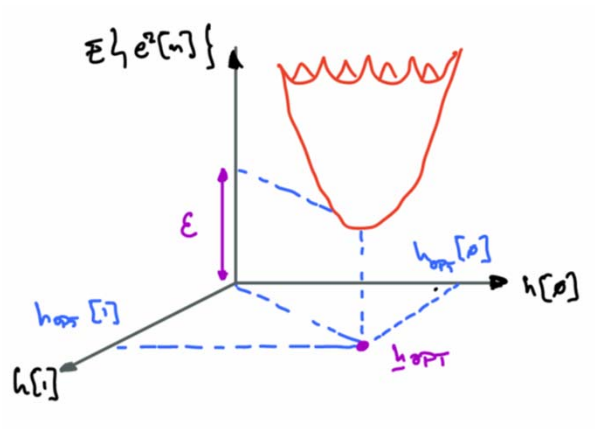
The error performance surface

3.1

The Wiener-Hopf filter is optimal in the sense that it **minimizes the MSE of the prediction**; that is, the variance (power) of $e[n]$.

- For any filter, the MSE can be expressed as:

$$\boxed{E\{(e[n])^2\} = \ \varepsilon + \left(\underline{\mathbf{h}}_{opt} - \underline{\mathbf{h}}\ \right)^T \underline{\underline{\mathbf{R}}}_x \left(\underline{\mathbf{h}}_{opt} - \underline{\mathbf{h}}\ \right)}$$



The **error performance surface**:

- Is a **quadratic function of the filter coefficients** and represents an M-dimensional surface:
  - Mathematically treatable
- As $\underline{\underline{\mathbf{R}}}_x$ is positive definite, the quadratic function is **convex** (bowl-shaped):
  - A unique extreme that is a **minimum**
- Provides a simple way to **assess the quality** of any filter implementation:
  - Very useful when working with **quantized filter coefficients**

The error performance surface

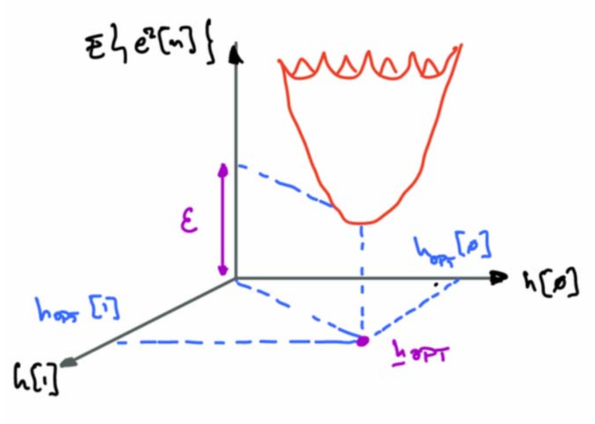
3.1

For any filter, the MSE can be expressed as:

$$E\{(e[n])^2\} = \varepsilon + \left(\underline{\mathbf{h}}_{opt} - \underline{\mathbf{h}}\right)^T \underline{\underline{\mathbf{R}}}_x \left(\underline{\mathbf{h}}_{opt} - \underline{\mathbf{h}}\right)$$

$$\underline{\mathbf{h}}_{opt} = \underline{\underline{\mathbf{R}}}_x^{-1} \underline{\mathbf{r}}_{xd}$$

$$\varepsilon = r_d[0] - \underline{\mathbf{h}}_{opt}^T \underline{\mathbf{r}}_{xd}$$



The **error performance surface**:

- The reference signal $d[n]$ only impacts on the **optimal solution**:
  - Position and value of the **minimum**.
- As $\underline{\underline{\mathbf{R}}}_x$ is positive definite, any deviation from the optimum **increases the MSE**. The increase depends on $\underline{\underline{\mathbf{R}}}_x$ only; that is, on $x[n]$only:
  - Very useful in the design of **adaptive filters**

Example: Interference cancelation

3.1

Signal cancelation:

We estimate the value of a primary signal which contains an interference. This interference has been isolated through other sensors in additional signals

MICROPHONE :

REFERENCE SENSOR (S) :

Example: Interference cancelation

3.1

→ MIC : $m[n] =$

→ SENSOR : $s[n] =$

Example: Interference cancelation

3.1

HOW TO SOLVE THE PROBLEM?

Example: Interference cancelation

3.1

→ FINAL RESULT : $\underline{\underline{R}}_x \cdot \underline{h}_{opt} = \underline{r}_{nd}$

$\Rightarrow \underline{\underline{R}}_x =$

Example: Interference cancelation

3.1

→ FINAL RESULT : $\underline{\underline{R}}_x \cdot \underline{h}_{OPT} = \underline{r}_{xd}$

⇒ $\underline{r}_{xd}$ =

Example: Interference cancelation

Signal cancelation:

We estimate the value of a primary signal which contains an interference. This interference has been isolated through other sensors in additional signals

Microphone signal: $m[n] = v[n] + e_m[n] + \cancel{w_m[n]} \Rightarrow d[n]$

Sensor signal: $s[n] = e_s[n] + w_s[n] \Rightarrow x[n]$

The analysis of this problem leads to the following solution:

$$\boxed{\left(\underline{\underline{\mathbf{R}}}_{e_s} + \underline{\underline{\mathbf{R}}}_{w_s}\right)\underline{\mathbf{h}}_{opt} = \underline{\mathbf{r}}_{e_s e_m}}$$

- The double function of the filter is evident in this solution:
  - It adapts the correlated part of $s[n]$ (with $m[n]$) while cancelling the uncorrelated one.
- What would be the impact of including the noise in the mic ($w_m[n]$)?

Wiener-Hopf filter

3.1

1. **Introduction**
   - Problem modelling and filter configuration
2. **Minimum Mean Square Error (MMSE) prediction**
   - Principle of orthogonality
   - Some results of the MMSE prediction
3. **The Wiener-Hopf filter**
   - The Wiener-Hopf solution
   - The error performance surface
   - The Wiener-Hopf filter using a finite number of samples
4. **Conclusions**

W-H filter using a finite number of samples

3.1

The **optimal filter** depends on the second order statistics of the processes. However, in a typical case, we only have a (small) **finite number of available samples** from both the observable and reference signals.

$$\underline{\underline{\mathbf{R}}}_x \, \underline{\mathbf{h}} = \underline{\mathbf{r}}_{xd}$$

$$\boxed{\underline{\mathbf{h}}_{opt} = \underline{\underline{\mathbf{R}}}_x^{-1} \underline{\mathbf{r}}_{xd}}$$

In that case, we can minimize **an estimation of the Mean Square Error** over the available set of samples. Let us assume that we have M samples of the reference signal and, given that the filter has N coefficients, $M + N - 1$ samples of the observation signal. We can define:

$$\boxed{\mathrm{MSE} \equiv \frac{1}{M} \sum_{m=0}^{M-1} (e[m])^2 = \frac{1}{M} \sum_{m=0}^{M-1} \left(d[m] - \underline{\mathbf{h}}^T \underline{\mathbf{x}}[m]\right)^2}$$

W-H filter using a finite number of samples

3.1

Let us assume that we have M samples of the reference signal and, given that the filter has N coefficients, $M+N-1$ samples of the observation signal:

$$\boxed{\mathrm{MSE} \equiv \frac{1}{M}\sum_{m=0}^{M-1}(e[m])^2 = \frac{1}{M}\sum_{m=0}^{M-1}\left(d[m] - \underline{\mathbf{h}}^T\underline{\mathbf{x}}[m]\right)^2}$$

Let us write this expression as combination of vectors. If we arrange the M sample equations ($e[n] = d[n] - \underline{\mathbf{h}}^T\underline{\mathbf{x}}[n]$) in a vector:

$$\underline{\mathbf{h}} = \begin{bmatrix} h[0] \\ h[1] \\ \cdots \\ h[N-1] \end{bmatrix} \quad \underline{\mathbf{x}}[n] = \begin{bmatrix} x[n] \\ x[n-1] \\ \cdots \\ x[n-N+1] \end{bmatrix}$$

$$\underline{\mathbf{e}}^T = \left[e[0], e[1], \cdots, e[M-1]\right]$$

$$\underline{\mathbf{d}}^T = \left[d[0], d[1], \cdots, d[M-1]\right]$$

$$\underline{\underline{\mathbf{X}}} = \begin{bmatrix} x[0] & x[1] & & x[M-1] \\ x[-1] & x[0] & & x[M-2] \\ \cdots & \cdots & \cdots & \cdots \\ x[-N+1] & x[-N+2] & & x[M-N] \end{bmatrix}$$

$$\boxed{\underline{\mathbf{e}}^T = \underline{\mathbf{d}}^T - \underline{\mathbf{h}}^T\underline{\underline{\mathbf{X}}}} \Rightarrow$$

W-H filter using a finite number of samples

3.1

The optimal filter should minimize the MSE:

$$\boxed{\underline{\mathbf{e}}^T = \underline{\mathbf{d}}^T - \underline{\mathbf{h}}^T \underline{\underline{\mathbf{X}}}}$$

$$\text{MSE} = \frac{1}{M}\sum_{m=0}^{M-1}(e[m])^2 = \frac{1}{M}\underline{\mathbf{e}}^T\underline{\mathbf{e}} = \frac{1}{M}\left(\underline{\mathbf{d}}^T - \underline{\mathbf{h}}^T\underline{\underline{\mathbf{X}}}\right)\left(\underline{\mathbf{d}}^T - \underline{\mathbf{h}}^T\underline{\underline{\mathbf{X}}}\right)^T \Rightarrow \quad \nabla_{\underline{\mathbf{h}}}\text{MSE} = \underline{\mathbf{0}}$$

$$\nabla_{\underline{\mathbf{h}}}\text{MSE} = \left[\left(\underline{\mathbf{d}}^T - \underline{\mathbf{h}}^T\underline{\underline{\mathbf{X}}}\right)^T = \left(\underline{\mathbf{d}} - \underline{\underline{\mathbf{X}}}^T\underline{\mathbf{h}}\right)\right] = \nabla_{\underline{\mathbf{h}}}\frac{1}{M}\left(\underline{\mathbf{d}}^T - \underline{\mathbf{h}}^T\underline{\underline{\mathbf{X}}}\right)\left(\underline{\mathbf{d}} - \underline{\underline{\mathbf{X}}}^T\underline{\mathbf{h}}\right) = \underline{\mathbf{0}}$$

$$\nabla_{\underline{\mathbf{h}}}\text{MSE} = \nabla_{\underline{\mathbf{h}}}\frac{1}{M}\left(\underline{\mathbf{d}}^T\underline{\mathbf{d}} - \underline{\mathbf{d}}^T\underline{\underline{\mathbf{X}}}^T\underline{\mathbf{h}} - \underline{\mathbf{h}}^T\underline{\underline{\mathbf{X}}}\,\underline{\mathbf{d}} + \underline{\mathbf{h}}^T\underline{\underline{\mathbf{X}}}\,\underline{\underline{\mathbf{X}}}^T\underline{\mathbf{h}}\right) = \underline{\mathbf{0}}$$

$$\nabla_{\underline{\mathbf{h}}}\text{MSE} = \frac{1}{M}\left(-2\underline{\underline{\mathbf{X}}}\,\underline{\mathbf{d}} + \mathbf{2}\underline{\underline{\mathbf{X}}}\,\underline{\underline{\mathbf{X}}}^T\underline{\mathbf{h}}\right) = \underline{\mathbf{0}}$$

$$\boxed{\underline{\mathbf{h}}_{opt} = \left(\underline{\underline{\mathbf{X}}}\,\underline{\underline{\mathbf{X}}}^T\right)^{-1}\underline{\underline{\mathbf{X}}}\,\underline{\mathbf{d}}}$$

W-H filter using a finite number of samples

3.1

By comparison with the optimal expression when having infinite samples:

$$\boxed{\underline{\mathbf{h}}_{opt} = (\underline{\underline{\mathbf{X}}}\,\underline{\underline{\mathbf{X}}}^T)^{-1}\underline{\underline{\mathbf{X}}}\,\underline{\mathbf{d}}}$$

Optimal filter (MMSE) using a finite number of samples

$$\boxed{\underline{\mathbf{h}}_{opt} = \underline{\underline{\mathbf{R}}}_x^{-1}\underline{\mathbf{r}}_{xd}}$$

Optimal filter (MMSE) using the exact second order statistics

We can see that we are implicitly estimating the cross-correlation vector and correlation matrix, based on the available samples:

$$\hat{\underline{\mathbf{r}}}_{xd}(\underline{\mathbf{x}}, \underline{\mathbf{d}}) = \frac{1}{M}\underline{\underline{\mathbf{X}}}\,\underline{\mathbf{d}} = \frac{1}{M}\sum_{m=1}^{M}\underline{\mathbf{x}}[m]d[m]$$

$$\hat{\underline{\underline{\mathbf{R}}}}_x(\underline{\mathbf{x}}) = \frac{1}{M}\underline{\underline{\mathbf{X}}}\,\underline{\underline{\mathbf{X}}}^T = \frac{1}{M}\sum_{m=1}^{M}\underline{\mathbf{x}}[m]\,\underline{\mathbf{x}}^T[m]$$

- ❑ **How have these estimates been built up?**

W-H filter using a finite number of samples

3.1

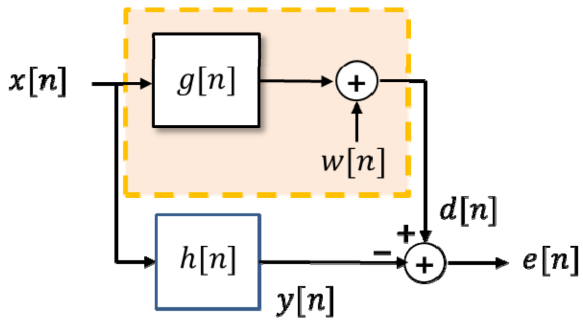
We can interpret the optimal filter (MMSE) using a finite number of samples, as an estimate of the Wiener-Hopf filter using exact second order statistics:

$$\underline{\mathbf{h}}_{opt} = \underline{\underline{\mathbf{R}}}_x^{-1}\underline{\mathbf{r}}_{xd} \quad \Rightarrow \quad \hat{\underline{\mathbf{h}}}_{opt} = (\underline{\underline{\mathbf{X}}}\,\underline{\underline{\mathbf{X}}}^T)^{-1}\underline{\underline{\mathbf{X}}}\,\underline{\mathbf{d}}$$

In order to assess this estimator, let us fix a (simple) **system identification scenario**. The application assumes:

- Noise-free observations (known signal: $\underline{\underline{\mathbf{X}}}$)
- Noisy reference: $\underline{\mathbf{d}}^T = \underline{\mathbf{g}}^T\underline{\underline{\mathbf{X}}} + \underline{\mathbf{w}}^T$

Note: The additive noise is modeled as white and Gaussian



How do we assess the quality of this estimator?

Performance of the estimator

3.1

- Analyze the performance of the optimal filter (MMSE) using a finite number of samples, as an estimator of the Wiener-Hopf filter using second order statistics

  Note: $w[n]$ is a **Gaussian, stationary, white noise**.

$$\underline{d}^T = \underline{g}^T \underline{\underline{X}} + \underline{w}^T \Rightarrow \underline{w}^T = \underline{d}^T - \underline{g}^T \underline{\underline{X}}$$

Wiener-Hopf filter