3. Optimal and Adaptive Filtering

3.1. Wiener-Hopf filter

3.1.1. Introduction

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

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

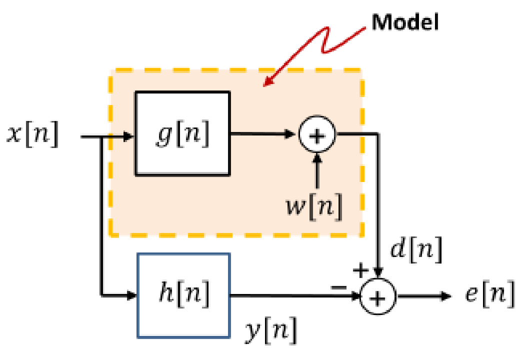
We assume **known stationary signal and noise spectra** (**correlation**), as well as **additive noise**. We will first assume Finite Impulse Response (FIR) filters, and afterwars non-stationary scenarios.

Filter configuration

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

- **System identification**

We want to **identify a given system**, that can be real or some abstraction. We model this system as an LTI system plus an additive noise source $w[n]$.

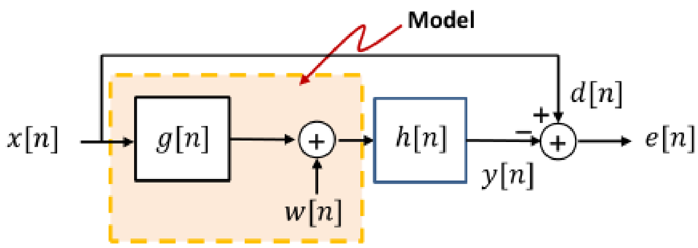


Design and use: we excite the system with a known signal $x[n]$ and obtain the filter that models the system.

The application assumes a noisy reference and noise-free observations.

- **System inversion**

We want to **estimate a system and apply its inverse to the signal**. We model this system as an LTI system plus an additive noise source $w[n]$.

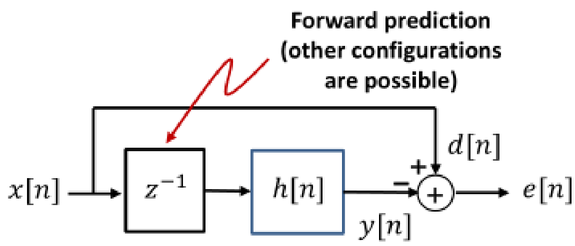


Design: we excite the system with a known signal $x[n]$ and obtain the filter that models the system.
Use: the filter is concatenated to the system to recover the estimated signal.

The application assumes noisy observations and a noise-free reference.

- **Signal prediction**

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



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 that observations and reference belong to the same noisy process.

- **Signal cancellation**

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 that the noisy interferences are the observations, while the noisy signal together with the interferences are the reference.

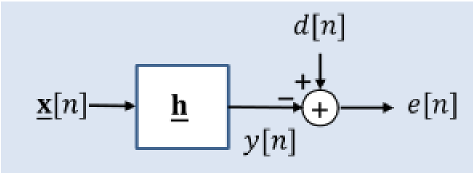
3.1.2. Minimum Mean Square Error (MMSE) prediction

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. Then, the filter is assumed to have a finite number of coefficients N. We use the **MSE** as an optimization criterion because it is mathematically tractable, leads to useful solutions and can be used as a benchmark for other solutions.

We will use the following notation:

$$h[n] * x[n] = \underline{h}^T \underline{x}[n], \quad \underline{x}[n] = \begin{bmatrix} x[n] \\ x[n-1] \\ \cdots \\ x[n-N+1] \end{bmatrix}.$$

So, say we have this generic signal case:



Then, the error is

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

and we want to minimize it:

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

In order to do this, we will first prove what is called the **principle of orthogonality**: if the MSE is minimum, then it holds that

$$\mathbb{E}[e[n]\underline{x}[n]] = \underline{0}.$$

If the MSE is minimum with respect to the filter, we know that the gradient is $\underline{0}$. Then, if we develop this mathematically,

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

So, we know that **the error is orthogonal to the observations**.

Now we want to develop some useful results of the MMSE prediction in a specific signal scenario:

- The observation process $x[n]$ can be split into two parts, $x[n] = a[n] + b[n]$.
- The reference process $d[n]$ can be split into two parts, $d[n] = a'[n] + c[n]$.

These parts of each process have the following correlation properties:

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

When using the filter that minimizes the MSE, $\underline{h}_{opt}$, the following properties hold:

1. At any point in time, the signal estimation and the error signal are not correlated:

$$\mathbb{E}[e[n]y[n]] = \left[y[n] = \underline{h}_{opt}^T \underline{x}[n]\right] = \mathbb{E}\left[e[n]\underline{h}_{opt}^T \underline{x}[n]\right] = \underline{h}_{opt}^T \mathbb{E}\left[e[n]\underline{x}[n]\right] = \underline{0}.$$

2. The variance of the reference signal is greater or equal than the variance of the error signal:

$$\begin{aligned} \mathbb{E}\left[d^2[n]\right] = \left[d[n] = y[n] + e[n]\right] = \mathbb{E}\left[(y[n] + e[n])^2\right] = \\ \mathbb{E}\left[d^2[n]\right] + 2\mathbb{E}\left[y[n]e[n]\right] + \mathbb{E}\left[e^2[n]\right] = \left[\text{we are using the optimal filter}\right] = \\ \mathbb{E}\left[d^2[n]\right] + \mathbb{E}\left[e^2[n]\right] \geq \mathbb{E}\left[e^2[n]\right]. \end{aligned}$$

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

$$\begin{aligned}\mathbb{E}\left[y^2[n]\right] = \mathbb{E}\left[y[n](d[n]-e[n])\right] = \mathbb{E}\left[y[n]d[n]\right] - \mathbb{E}\left[y[n]e[n]\right] = \\ \mathbb{E}\left[y[n]d[n]\right] = \mathbb{E}\left[\underline{h}_{opt}^T\underline{x}[n]d[n]\right] = \underline{h}_{opt}^T\mathbb{E}\left[\underline{x}[n]d[n]\right] = \\ \left[\mathbb{E}\left[\underline{x}[n]d[n]\right] = \underline{0}\right] = 0.\end{aligned}$$

4. The minimum variance of the error signal is $\varepsilon = r_d[0] - \underline{h}_{opt}^T\underline{r}_{xd}$:

$$\begin{aligned}\varepsilon := \mathbb{E}\left[e^2[n]\right]\Big|_{\min} = \mathbb{E}\left[e[n](d[n]-y[n])\right] = \mathbb{E}\left[e[n]d[n]\right] - \mathbb{E}\left[e[n]y[n]\right] = \\ \mathbb{E}\left[e[n]d[n]\right] = \mathbb{E}\left[\left(d[n]-\underline{h}_{opt}^T\underline{x}[n]\right)d[n]\right] = \mathbb{E}\left[d^2[n]\right] - \mathbb{E}\left[\underline{h}_{opt}^T\underline{x}[n]d[n]\right] = \\ r_d[0] - \underline{h}_{opt}^T\mathbb{E}\left[\underline{x}[n]d[n]\right] = \\ r_d[0] - \underline{h}_{opt}^T\underline{r}_{xd}.\end{aligned}$$

We will analyze the previous properties for the signal setting we stated before:

(sic.)

3.1.3. The Wiener-Hopf filter

The Wiener-Hopf solution

So far, we have analyzed some properties of the optimal filter, but we are yet to obtain it:

$$\begin{aligned}\left.\begin{array}{c} e[n] = d[n] - \underline{h}^T\underline{x}[n] \\ \mathbb{E}\left[\underline{x}[n]e[n]\right] = \underline{0}\end{array}\right| \implies \mathbb{E}\left[\underline{x}[n]\left(d[n]-\underline{h}^T\underline{x}[n]\right)\right] = \underline{0}. \\ \mathbb{E}\left[\underline{x}[n]\left(d[n]-\underline{h}^T\underline{x}[n]\right)\right] = \mathbb{E}\left[\underline{x}[n]d[n]\right] - \mathbb{E}\left[\underline{x}[n]\underline{h}^T\underline{x}[n]\right] = \underline{0} \iff \\ \mathbb{E}\left[\underline{x}[n]d[n]\right] - \mathbb{E}\left[\underline{x}[n]\underline{x}^T[n]\underline{h}\right] = \underline{r}_{xd}[0] - \mathbb{E}\left[\underline{x}[n]\underline{x}^T[n]\right]\underline{h} = \underline{0} \iff \\ \underline{r}_{xd}[0] - \underline{\underline{R}}_x[0]\underline{h} = \underline{0} \iff \boxed{\underline{h}_{opt} = \underline{\underline{R}}_x^{-1}\underline{r}_{xd}.}\end{aligned}$$

So, this is the optimal filter in the sense of MSE minimization, with the matrix and the vector involved being

$$\underline{r}_{xd} = \mathbb{E}\left[\underline{x}[n]d[n]\right] = \begin{bmatrix}\mathbb{E}\left[x[n]d[n]\right] \\ \mathbb{E}\left[x[n-1]d[n]\right] \\ \cdots \\ \mathbb{E}\left[x[n-N+1]d[n]\right]\end{bmatrix} = \begin{bmatrix} r_{xd}[0] \\ r_{xd}[-1] \\ \cdots \\ r_{xd}[-N+1]\end{bmatrix} : \text{the Cross-correlation vector}$$

$$\underline{\underline{R}}_x = \mathbb{E}\left[\underline{x}[n]\underline{x}^T[n]\right] = \begin{bmatrix} r_x[0] & r_x[1] & \cdots & r_x[N-1] \\ r_x[-1] & r_x[0] & \cdots & r_x[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_x[-N+1] & r_x[-N+2] & \cdots & r_x[0]\end{bmatrix} : \text{the Correlation matrix}$$

The **optimal filter**, hence, depends on the second order statistics of the processes. We will analyze the properties of this correlation matrix, and also we will study what to do when such statistics are not available.

The error performance surface

The Wiener-Hopf filter is optimal at minimizing the MSE of the prediction; that is, the variance (power) of the error signal $e[n]$ (assuming it is zero-mean). We now want to study how does the MSE behave for an arbitrary filter. In order to do this, let us first develop a useful result: for any filter, the MSE can be expressed as

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

So, let's see this result:

$$\begin{aligned}
\mathbb{E}\left[e^2[n]\right] = \mathbb{E}\left[\left(d[n] - \underline{h}^T \underline{x}[n]\right)\left(d[n] - \underline{h}^T \underline{x}[n]\right)\right] = \\
\mathbb{E}\left[d^2[n]\right] - 2\mathbb{E}\left[\underline{h}^T \underline{x}[n] d[n]\right] + \mathbb{E}\left[\underline{h}^T \underline{x}[n] \underline{h}^T \underline{x}[n]\right] = \\
r_d[0] - 2\mathbb{E}\left[\underline{h}^T \underline{x}[n] d[n]\right] + \mathbb{E}\left[\underline{h}^T \underline{x}[n] \underline{x}^T[n] \underline{h}\right] = \\
r_d[0] - 2\underline{h}^T \mathbb{E}\left[\underline{x}[n] d[n]\right] + \underline{h}^T \mathbb{E}\left[\underline{x}[n] \underline{x}^T[n]\right] \underline{h} = \\
r_d[0] - 2\underline{h}^T \underline{r}_{xd} + \underline{h}^T \underline{\underline{R}}_x \underline{h} = \left[\varepsilon = r_d[0] - \underline{h}_{opt}^T \underline{r}_{xd}\right] = \\
r_d[0] - \underline{h}_{opt}^T \underline{r}_{xd} + \underline{h}_{opt}^T \underline{r}_{xd} - 2\underline{h}^T \underline{r}_{xd} + \underline{h}^T \underline{\underline{R}}_x \underline{h} = \\
\varepsilon + \underline{h}_{opt}^T \underline{\underline{R}}_x \underline{h}_{opt} - 2\underline{h}^T \underline{\underline{R}}_x \underline{h}_{opt} + \underline{h}^T \underline{\underline{R}}_x \underline{h} = \\
\varepsilon + \underline{h}_{opt}^T \underline{\underline{R}}_x \underline{h}_{opt} - \underline{h}^T \underline{\underline{R}}_x \underline{h}_{opt} - \underline{h}_{opt}^T \underline{\underline{R}}_x \underline{h} + \underline{h}^T \underline{\underline{R}}_x \underline{h} \implies \\
\boxed{\mathbb{E}\left[e^2[n]\right] = \varepsilon + \left(\underline{h}_{opt} - \underline{h}\right)^T \underline{\underline{R}}_x \left(\underline{h}_{opt} - \underline{h}\right).}
\end{aligned}$$

So, as we can see, the MSE of **any filter** is a quadratic function of the filter coefficients and always lies in an $N-$dimensional surface. As $\underline{\underline{R}}_x$ is positive definite, the quadratic function is **convex** and hence it has a unique extreme that is a **minimum**, which is exactly what we wanted. The reference signal $d[n]$ only impacts on the position and value of the optimal solution, and not on the shape of the surface. We can also note that as $\underline{\underline{R}}_x$ is positive definite, **any deviation from the optimum filter increases the MSE**. This increase depends only on $\underline{\underline{R}}_x$, and so, only on $x[n]$. This fact will be very useful in the design of adaptive filters.

Wiener-Hopf filter using a finite number of samples