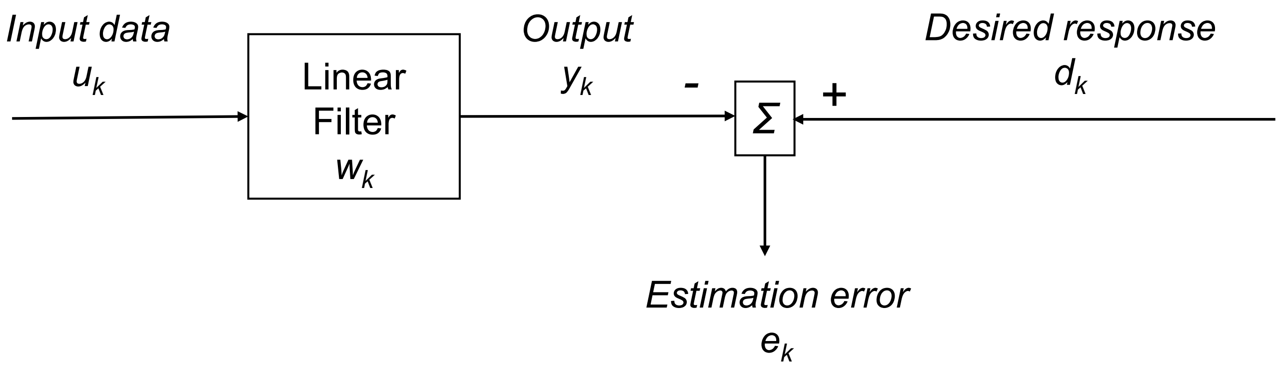
Adaptive Filtering

Luca Reggiani
luca.reggiani@polimi.it

Linear Optimum Filters

A linear discrete filter can be designed for obtaining an output as close as possible to a desired Response.



For a linear discrete filter, the typical cost function to be optimized is the mean value of the square of the estimation error $e_k = d_k - y_k$.

The filter coefficients are derived w.r.t. the optimization of this cost function.

Principle of orthogonality

Therefore, given the linear filtering operation

$$y_n = \sum_k w_k^* \cdot u_{n-k} \qquad \text{(Eq. 1)}$$

and the MSE (mean squared error) cost function

$$J = E[|d_n - y_n|^2] = E[|e_n|^2]$$

$$\nabla J = \frac{\partial J}{\partial \Re(w_k)} + j\frac{\partial J}{\partial \Im(w_k)}$$

we are imposing, in order to find the minimum, that the gradient is zero (in the complex sense, i.e. w.r.t. real and imaginary parts of the optimal filter coefficients) :

$$\nabla_k J = 0 \qquad \forall k$$

Principle of orthogonality

As

$$\nabla_k J = E\left[\frac{\partial e_n e_n^*}{\partial \Re(w_k)} + j\frac{\partial e_n e_n^*}{\partial \Im(w_k)}\right] = E\left[\frac{\partial e_n}{\partial \Re(w_k)} e_n^* + \frac{\partial e_n^*}{\partial \Re(w_k)} e_n + j\frac{\partial e_n}{\partial \Im(w_k)} e_n^* + j\frac{\partial e_n^*}{\partial \Im(w_k)} e_n\right]$$

and, from (Eq. 1),

$$\frac{\partial e_n}{\partial \Re(w_k)} = -u_{n-k} \qquad \frac{\partial e_n^*}{\partial \Re(w_k)} = -u_{n-k}^*$$

$$\frac{\partial e_n}{\partial \Im(w_k)} = ju_{n-k} \qquad \frac{\partial e_n^*}{\partial \Im(w_k)} = -ju_{n-k}^*$$

we obtain

$$\boxed{E\left[u_{n-k} \cdot e_{opt,n}^*\right] = 0 \qquad \forall k} \qquad \text{(Eq. 2)}$$

that means that the error of the optimal filter turns out to be orthogonal to each input sample.

Minimum Mean Squared Error

The cost function, evaluated for the optimal filter,

$$J_{min} = E\left[\left|e_{opt,n}\right|^2\right]$$

$$e_{opt,n} = d_n - y_{opt,n} = d_n - \hat{d}_n$$

thanks to the orthogonality, becomes

$$J_{min} = E[|d_n|^2] - E\left[\left|\hat{d}_n\right|^2\right] = \sigma_d^2 - \sigma_{\hat{d}}^2$$

In fact, it is easy to realize from (Eq. 1) and (Eq. 2), that the orthogonality principle holds also between filter outputs and the corresponding error:

$$E\left[y_{opt,n} \cdot e^*_{opt,n}\right] = 0$$

Wiener - Hopf equations

The W-H equations allow the computation of the filter coefficients.

For the FIR case, when the filter length *M* is finite, we have

$$E\left[u_{n-k}\left(d_n^* - \sum_{i=0}^{M-1} w_{opt,i} u_{n-i}^*\right)\right] = 0$$

and

$$\boxed{\sum_{i=0}^{M-1} w_{opt,i} r_{i-k} = p_{-k} \qquad k = 0, 1, \ldots, M-1}$$

where we have used the autocorrelation of the input data and the cross-correlation between desired and input data:

$$r_k = E[u_n u_{n-k}^*]$$

$$p_k = E[u_n d_{n-k}^*]$$

MMSE filters

- In the **matrix** form, the vector of optimal weights

$$\boxed{\begin{gathered}\boldsymbol{R}\cdot\boldsymbol{w}_{opt}=\boldsymbol{p}\\ \boldsymbol{w}_{opt}=\boldsymbol{R}^{-1}\cdot\boldsymbol{p}\end{gathered}} \qquad \text{(Eq. 3)}$$

with

$$\boldsymbol{R}=\begin{bmatrix} r_0 & \cdots & r_{M-1} \\ r_1^* & \cdots & r_{M-2} \\ r_{M-1}^* & \cdots & r_0 \end{bmatrix}$$

$$\boldsymbol{p}=[p_0 \quad p_{-1} \quad p_{-(M-1)}]^T$$

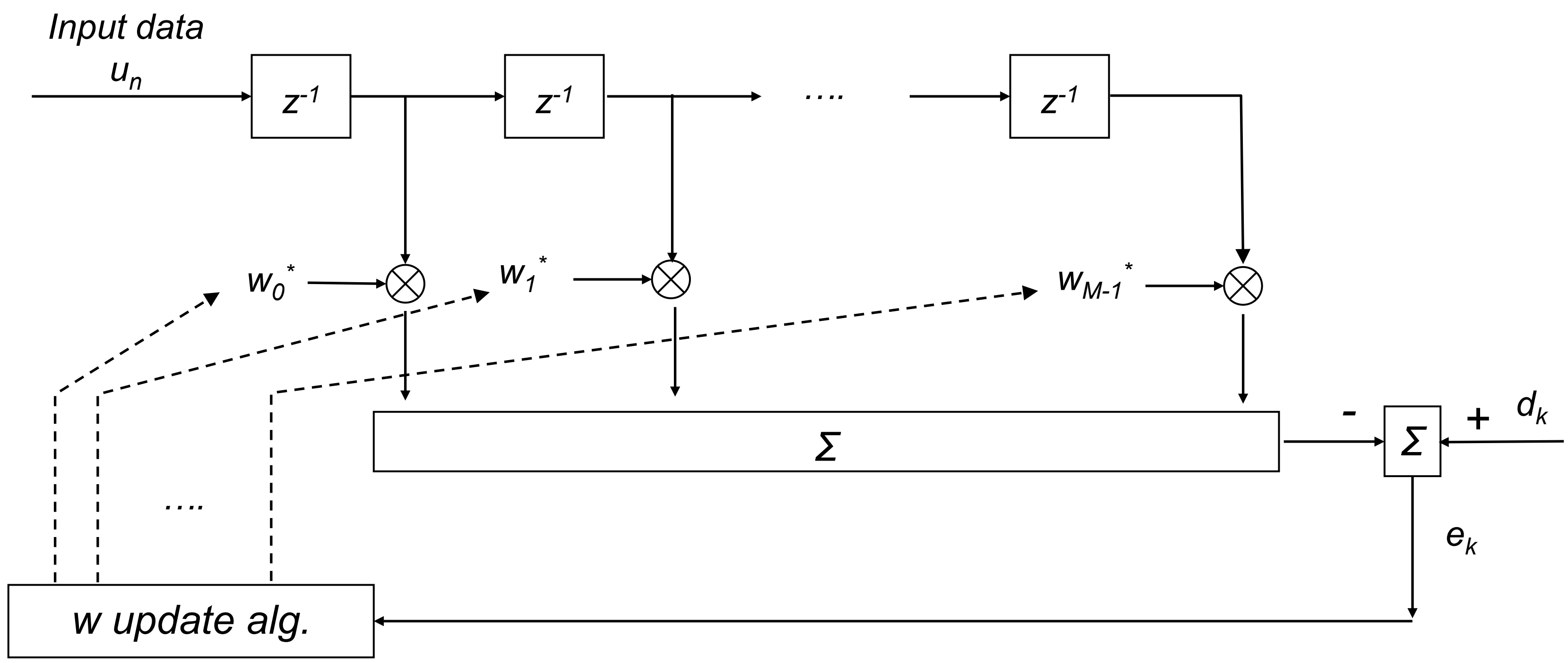
- **The minimum MSE**

It is not difficult to see that

$$J_{min}=\sigma_d^2-\sigma_{\hat{d}}^2=\sigma_d^2-\boldsymbol{w}_{opt}^H\cdot\boldsymbol{R}\cdot\boldsymbol{w}_{opt}=\sigma_d^2-\boldsymbol{w}_{opt}^H\cdot\boldsymbol{p}=\sigma_d^2-\boldsymbol{p}^H\cdot\boldsymbol{w}_{opt}$$

Adaptive architecture

The idea is to achieve the optimal solution by means of successive steps, iteratively, instead of directly with (Eq. 3). This allows the tracking in case of time-variant processes and, in addition, it can eliminate the estimation of the auto and cross correlations, which is typically difficult.



Method of the steepest gradient

Step size parameter

$$\boldsymbol{w}[\boldsymbol{n}+\mathbf{1}] = \boldsymbol{w}[\boldsymbol{n}] - \frac{1}{2}\mu \cdot \boldsymbol{\nabla J}[\boldsymbol{n}]$$

with

$$\boldsymbol{\nabla J}[\boldsymbol{n}] = -2\boldsymbol{p} + 2\boldsymbol{R} \cdot \boldsymbol{w}[\boldsymbol{n}]$$

It is a method for the minimum search of the cost function, which follows the opposite of the current gradient.

It can be seen that the sufficient and necessary stability condition for the convergence is

$$0 < \mu < \frac{2}{\lambda_{max}}$$

where λ_{max} is the maximum eigenvalue of ***R.***

Least Mean Square algorithm

Here we simplify the computation of the correlation vector and matrix, replacing them by an <u>instantaneous estimate</u>

$$\widehat{\boldsymbol{R}}[\boldsymbol{n}] = \boldsymbol{u}[\boldsymbol{n}] \cdot \boldsymbol{u}^H[\boldsymbol{n}]$$

$$\widehat{\boldsymbol{p}}[\boldsymbol{n}] = \boldsymbol{u}[\boldsymbol{n}] \cdot \boldsymbol{d}^*[\boldsymbol{n}]$$

$$\widehat{\nabla}\boldsymbol{J}[\boldsymbol{n}] = -2\boldsymbol{u}[\boldsymbol{n}] \cdot \boldsymbol{d}^*[\boldsymbol{n}] + 2\boldsymbol{u}[\boldsymbol{n}] \cdot \boldsymbol{u}^H[\boldsymbol{n}] \cdot \boldsymbol{w}[\boldsymbol{n}] = -2\boldsymbol{u}[\boldsymbol{n}] \cdot (\boldsymbol{d}^*[\boldsymbol{n}] - \boldsymbol{u}^H[\boldsymbol{n}] \cdot \boldsymbol{w}[\boldsymbol{n}])$$

Therefore

$$\boxed{\begin{aligned} \widehat{\boldsymbol{w}}[\boldsymbol{n}+\boldsymbol{1}] &= \widehat{\boldsymbol{w}}[\boldsymbol{n}] + \mu \cdot \boldsymbol{u}[\boldsymbol{n}] \cdot \boldsymbol{e}^*[\boldsymbol{n}] \\ &= \widehat{\boldsymbol{w}}[\boldsymbol{n}] + \mu \cdot \boldsymbol{u}[\boldsymbol{n}] \cdot (\boldsymbol{d}^*[\boldsymbol{n}] - \boldsymbol{u}^H[\boldsymbol{n}]\widehat{\boldsymbol{w}}[\boldsymbol{n}]) \end{aligned}}$$

Because of this simplification, the LMS suffers from gradient noise and, when it converges, it shows an excess mean square error w.r.t. Wiener, optimal solution.

References

[1] S. Haykin, ”Adaptive Filter Theory”, Prentice Hall.