POLITECNICO DI MILANO

Wiener Filters

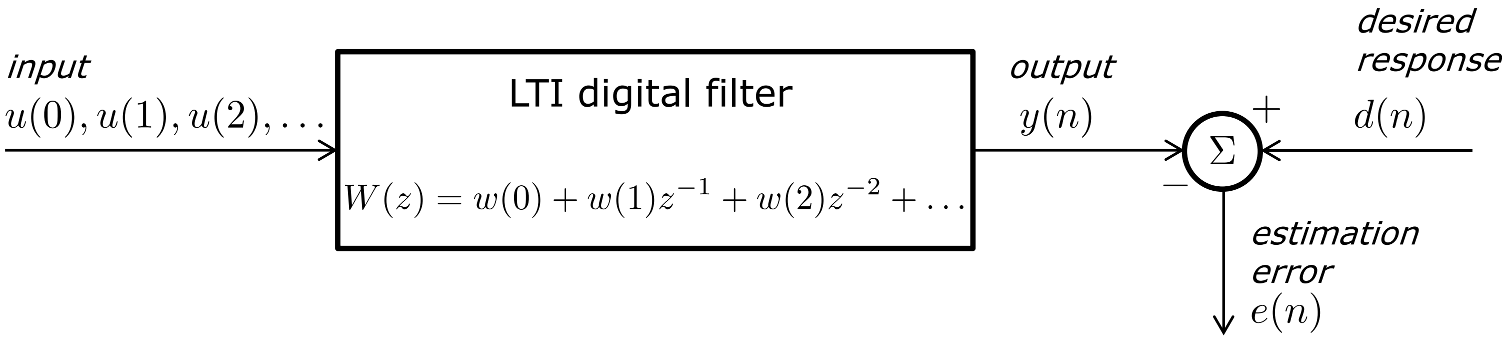
Antonio Canclini
Augusto Sarti

Introduction

- The purpose of Wiener filtering is to produce an estimate of a target signal by linear time-invariant (LTI) filtering of an observed noisy process

- In particular, the goal is to perform minimum mean-square-error (MMSE) estimation of a stationary random process of interest $d(n)$, given measurements of another related process $u(n)$

- We consider the following block schematic:



Data model and problem formulation

- All the signals are complex-valued stationary zero-mean random processes $u(n), d(n), y(n), e(n) \in \mathbb{C}$

- $W(z)$ is a complex-valued IIR (infinite impulse response) filter $w(0), w(1), w(2), \ldots \in \mathbb{C}$

- Filter output computed through linear convolution $$y(n) = \sum_{k=0}^{\infty} w_k^* u(n-k)$$

- We define a MSE cost function as follows

$$J = E\{e(n)e^*(n)\} = E\{|e(n)|^2\}\ , \quad \text{with} \quad e(n) = d(n) - y(n)$$

- Goal: find the filter coeffs. that minimize J

$$\hat{W}(z) = \arg \min_{W(z)} J$$

The orthogonality principle

- We are interested in computing the gradient of the cost function
- We express the filter coefficients as $w(k) = a(k) + jb(k)\,, \quad k = 0, 1, 2, \ldots$
- We define the (complex) gradient operator as $\nabla_k = \dfrac{\partial}{\partial a_k} + j\dfrac{\partial}{\partial b_k}$
- The gradient of the cost function is therefore given by

$$\nabla_k J = E\left\{ \frac{\partial e(n)}{\partial a(k)} e^*(n) + \frac{\partial e^*(n)}{\partial a(k)} e(n) + j\frac{\partial e(n)}{\partial b(k)} e^*(n) + j\frac{\partial e^*(n)}{\partial b(k)} e(n) \right\} \quad (1)$$

- We expect the cost function exhibits a minimum when its gradient is zero, thus we impose

$$\nabla_k J = 0 \quad \forall\, k = 0, 1, 2, \ldots \infty$$

The orthogonality principle

- To solve for $\nabla_k J = 0$ we compute all the partial derivatives included in the expression of the gradient, e.g.

$$\begin{aligned}\frac{\partial e(n)}{\partial a_k} &= \frac{\partial \left\{d(n) - \sum_{l=0}^{+\infty} w_l^* u(n-l)\right\}}{\partial a_k} \\ &= \frac{\partial \left\{d(n) - \sum_{l=0}^{+\infty} \left[a_l - jb_l\right] u(n-l)\right\}}{\partial a_k} \\ &= \frac{\partial d(n)}{\partial a_k} - \frac{\partial \sum_{l=0}^{+\infty} a_l u(n-l)}{\partial a_k} + j\frac{\partial \sum_{l=0}^{+\infty} b_l u(n-l)}{\partial a_k} \\ &= -u(n-k)\end{aligned}$$

- Similarly, we obtain:

$$\frac{\partial e^*(n)}{\partial a(k)} = -u^*(n-k) \qquad \frac{\partial e(n)}{\partial b(k)} = ju(n-k) \qquad \frac{\partial e^*(n)}{\partial b(k)} = -ju^*(n-k)$$

The orthogonality principle

- Replacing the partial derivatives in (1), after canceling common terms, we finally get the result

$$\nabla_k J = -2E\left\{u(n-k)e^*(n)\right\}$$

- We are now ready to specify the operating conditions required for minimizing the cost function:
  - Let $e_o(n)$ denote the special value of the estimation error that results when the filter operates in its optimum condition (i.e., the cost function has reached its minimum)
  - In this case the gradient must be zero, and therefore the **principle of orthogonality** holds:

$$\boxed{E\left\{u(n-k)e_o^*(n)\right\} = 0\ , \quad k = 0, 1, 2, \ldots}$$

the estimation error is orthogonal to the input samples that enter into the estimation of the desired response

Corollary of the orthogonality principle

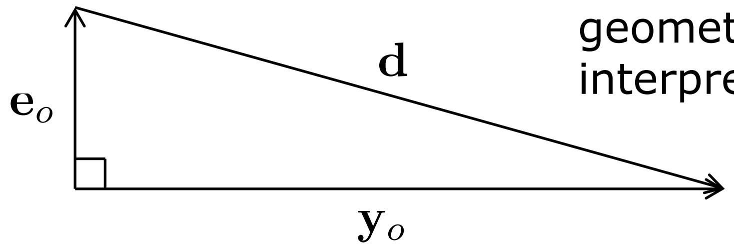
- We now analyze the correlation between the filter output and the estimation error:

$$E\left\{y(n)e^*(n)\right\} = E\left\{\sum_{k=0}^{\infty} w^*(k)u(n-k)e^*(n)\right\} = \sum_{k=0}^{\infty} w^*(k)E\left\{u(n-k)e^*(n)\right\}$$

- Denoting with $y_o(n)$ the output produced by the filter optimized in the MSE sense, using the orthogonality principle we get

$$\boxed{E\left\{y_o(n)e_o^*(n)\right\} = 0}$$

- This corollary states that, when the filter operates in its optimum conditions, the estimate of the desired response is orthogonal (uncorrelated) to the corresponding estimation error



geometric interpretation

Minimum mean-squared error

- Consider the following definitions:

$$\hat{d}(n) \triangleq y_o(n) \qquad \text{estimate of the desired response}$$

$$J_{\min} \triangleq E\left\{|e_o(n)|^2\right\} \qquad \text{minimum MSE}$$

$$\sigma_d^2 \triangleq E\left\{|d(n)|^2\right\} \qquad \text{variance of } d(n)$$

$$\sigma_{\hat{d}}^2 \triangleq E\left\{|\hat{d}(n)|^2\right\} \qquad \text{variance of } \hat{d}(n)$$

- It is easy to verify that

$$e_o(n) = d(n) - y_o(n) \quad \Longrightarrow \quad d(n) = \hat{d}(n) + e_o(n)$$

- Applying the corollary of the orthogonality principle, we obtain

$$\sigma_d^2 = \sigma_{\hat{d}}^2 + J_{\min} \quad \Longrightarrow \quad J_{\min} = \sigma_d^2 - \sigma_{\hat{d}}^2 \quad \Longrightarrow \quad \boxed{\varepsilon \triangleq \frac{J_{\min}}{\sigma_d^2} = 1 - \frac{\sigma_{\hat{d}}^2}{\sigma_d^2}}$$

normalized MSE

$0 \leq \varepsilon \leq 1$

0 → $\sigma_{\hat{d}}^2 = \sigma_d^2$ (complete agreement between $\hat{d}(n)$ and $d(n)$)

1 → $\sigma_{\hat{d}}^2 = 0$ (no agreement, worst possible situation)

- The principle of orthogonality specifies the necessary and sufficient condition for the optimum operation of the Wiener filter
- It can be rewritten as follows, denoting the optimum filter coefficients with $w_o(0), w_o(1), w_o(2), \dots$

$$E\left\{u(n-k)\left[d^*(n)-\sum_{i=0}^{\infty} w_o(i)u^*(n-i)\right]\right\}=0\ , \qquad k=0,1,2,\dots$$

- Expanding the equation and rearranging the terms we get

$$\sum_{i=0}^{\infty} w_o(i)\underbrace{E\left\{u(n-k)u^*(n-i)\right\}}_{r(i-k)}=\underbrace{E\left\{u(n-k)d^*(n)\right\}}_{p(-k)}\ , \qquad k=0,1,2,\dots$$

autocorrelation of the input

cross-correlation of the input and the desired response

Wiener-Hopf equations

$$\boxed{\sum_{i=0}^{\infty} w_o(i)r(i-k)=p(-k)\ , \qquad k=0,1,2,\dots}$$

linear system of equations, which define the optimum filter coefficients

Solution for FIR filters

- Wiener-Hopf equations define an infinite set of equations whose solution is the IIR Wiener filter

- An approximate solution can be obtained by truncating the summation, to compute a FIR filter with M taps

$$\sum_{i=0}^{M-1} w_o(i) r(i-k) = p(-k) \ , \qquad k = 0, 1, 2, \ldots, M-1$$

- Note that FIR filters are preferrable with respect to IIR ones, as they are always stable

- The choice of M depends on the specific applications, i.e. on the order of the linear system to be identified

Matrix formulation

- Consider the following definitions

input vector $$\mathbf{u}(n) = [u(n), u(n-1), \ldots, u(n-M+1)]^T$$

cross-correlation vector $$\mathbf{p} = E\{\mathbf{u}(n)d^*(n)\} = [p(0), p(-1), \ldots, p(1-M)]^T$$

vector of optimum coefficients $$\mathbf{w}_o = [w_o(0), w_o(1), \ldots, w_o(M-1)]^T$$

auto-correlation matrix $$\mathbf{R} = E\{\mathbf{u}(n)\mathbf{u}(n)^H\} =$$

$$= \begin{bmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r^*(1) & r(0) & \cdots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(M-1) & r^*(M-2) & \cdots & r(0) \end{bmatrix}$$

- The Wiener-Hopf equations can be rewritten in matrix form, and the solution is found by matrix inversion:

$$\mathbf{R}\mathbf{w}_o = \mathbf{p} \quad \Longrightarrow \quad \mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$$

Error performance surface

- The Wiener-Hopf equation can also be derived by examining the dependence of the cost function J on the tap weights of the filter
- We first rewrite the estimation error as

$$e(n) = d(n) - \sum_{k=0}^{M-1} w^*(k)u(n-k)$$

- Substituting in the expression of the cost function, we get

$$\begin{aligned} J &= E\{e(n)e^*(n)\} \\ &= E\{|d(n)|^2\} - \sum_{k=0}^{M-1} w^*(k)E\{u(n-k)d^*(n)\} - \sum_{k=0}^{M-1} w(k)E\{u^*(n-k)d(n)\} \\ &+ \sum_{k=0}^{M-1}\sum_{i=0}^{M-1} w^*(k)w(i)E\{u^*(n-k)u(n-i)\} \end{aligned}$$

$$\Rightarrow \boxed{J = \sigma_d^2 - \sum_{k=0}^{M-1} w^*(k)p(-k) - \sum_{k=0}^{M-1} w(k)p^*(-k) + \sum_{k=0}^{M-1}\sum_{i=0}^{M-1} w^*(k)w(i)r(i-k)}$$

Error performance surface

- J is a second-order function of the tap weights in the filter

- Thus, we may visualize the dependence of the cost function on the tap weights $w(0), w(1), \ldots, w(M-1)$ as a bowl-shaped *(M+1)*-dimensional surface with M degrees of freedom (**error performance surface**)

paraboloid

- Most important, this surface exhibits a unique minimum

convex => unique minimum

- The gradient of J can be easily computed from the previous equation

$$\begin{aligned}\nabla_k J &= \frac{\partial J}{\partial a(k)} + j\frac{\partial J}{\partial b(k)} \\ &= -2p(-k) + 2\sum_{i=0}^{M-1} w(i)r(i-k)\end{aligned}$$

Error performance surface

- At the minimum point of the error-performance surface, the cost function attains the minimum, and the gradient is identically zero

- Thus, in correspondence of the optimum filter weights

$$-2p(-k) + 2\sum_{i=0}^{M-1} w_o(i)r(i-k) = 0 \quad \Rightarrow \quad \boxed{\sum_{i=0}^{M-1} w_o(i)r(i-k) = p(-k)}$$

As expected, this system of equation coincides with the Wiener-Hopf equations found starting from the orthogonality principle

Minimum Mean-squared error

- We are now interested in computing the value of $J_{\min} = \sigma_d^2 - \sigma_{\hat{d}}^2$

- The only unknown is $\sigma_{\hat{d}}^2$, we need to compute the variance of

$$\hat{d}(n) = \sum_{k=0}^{M-1} w_o^*(k)u(n-k) = \mathbf{w}_o^H \mathbf{u} \Longrightarrow \begin{aligned} \sigma_{\hat{d}}^2 &= E\{\mathbf{w}_o^H \mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}_o\} \\ &= \mathbf{w}_o^H E\{\mathbf{u}(n)\mathbf{u}^H(n)\}\mathbf{w}_o \\ &= \mathbf{w}_o^H \mathbf{R}\mathbf{w}_o \end{aligned}$$

- Using the fact that $\mathbf{R}\mathbf{w}_o = \mathbf{p}$, we have $\sigma_{\hat{d}}^2 = \mathbf{w}_o^H \mathbf{p} = \mathbf{p}^H \mathbf{w}_o$

- Finally, using $\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$ we get the desired result

$$\boxed{J_{\min} = \sigma_d^2 - \mathbf{p}^H \mathbf{R}^{-1} \mathbf{p}}$$

Canonical form of the Error-Performance surface

- The error-performance surface can be rewritten in matrix form as

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{w}^H\mathbf{p} - \mathbf{p}^H\mathbf{w} + \mathbf{w}^H\mathbf{R}\mathbf{w}$$

- Rearranging the terms, we get

$$\begin{aligned} J(\mathbf{w}) &= \left[\sigma_d^2 - \mathbf{p}^H\mathbf{R}^{-1}\mathbf{p}\right] + (\mathbf{w} - \mathbf{R}^{-1}\mathbf{p})^H\mathbf{R}(\mathbf{w} - \mathbf{R}^{-1}\mathbf{p}) \\ &= J_{\min} + (\mathbf{w} - \mathbf{w}_o)^H\mathbf{R}(\mathbf{w} - \mathbf{w}_o) \end{aligned}$$

- The above equation shows explicitly the existance of an unique minimum of the cost function, occurring when $\mathbf{w} = \mathbf{w}_o$

- Netherveless, the expression of the cost function can be further simplified by operating a change of the basis on which it is defined

Canonical form of the Error-Performance surface

- We consider the eigenvalue decomposition of $\mathbf{R}$:

$$\mathbf{R} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^H \ , \qquad \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_M \end{bmatrix}$$

- Consider also the following transformed version of the tap-weights:

$$\mathbf{v} = \mathbf{Q}^H(\mathbf{w} - \mathbf{w}_o) = [v_1, v_2, \ldots, v_M]^T$$

- Using the above definitions, the canonical form is defined as

$$\boxed{\begin{aligned} J &= J_{\min} + \mathbf{v}^H\boldsymbol{\Lambda}\mathbf{v} \\ &= J_{\min} + \sum_{k=1}^{M} \lambda_k v_k v_k^* \\ &= J_{\min} + \sum_{k=1}^{M} \lambda_k |v_k|^2 \end{aligned}}$$

The components of the transformed coefficient vector $\mathbf{v}$ constitute the principal axes of the error-performance surface