Estimation of Autocorrelation Matrix

Lets consider $x(n) = A\cos(w_0 n + f)$, where f is random variable uniformly distributed between $0, 2p$.

$$g(\ell) = E\left[x_n^* x_{n+\ell}\right] = A^2 \frac{1}{2p}\int_0^{2p} \cos(w_0(n+\ell)+f)\cos(w_0 n + f)\,df$$

$$= \frac{A^2}{2}\frac{1}{2p}\int_o^{2p}\left[\cos w_0\ell + \underbrace{\cos(2w_0 n + w_0 l + 2j)}_{0}\right]df$$

$$= \frac{A^2}{2}\cos w_0\ell \,\frac{1}{2p} f\,\Big|_0^{2p} = \frac{A^2}{2}\cos w_0\ell$$

It is an even function too.

Now lets see what happens if we have a limited data points, N

$$\hat{g}(\ell) = \frac{1}{N}\sum_{n=0}^{N-\ell-1} x^*(n)\,x(n+\ell) \qquad for\ \ell \ge o$$

$$= \frac{A^2}{N}\sum_{n=0}^{N-\ell-1}\cos(w_0 n + f)\cos(w_0(n+\ell)+f)$$

$$= \frac{A^2}{2N}\left\{\underbrace{\sum_{n=0}^{N-\ell-1}\cos w_0\ell}_{1st\ part} + \underbrace{\sum_{n=0}^{N-\ell-1}\cos(2w_0 n + w_0\ell + 2f)}_{2nd\ part}\right\}$$

Lets simplify the second part first:

$$2^{\text{nd}}\text{ part}\ \begin{cases} = \operatorname{Re}\left\{\sum_{n=o}^{N-\ell-1} e^{j2w_0 n}\cdot e^{j(w_0\ell+2f)}\right\} \\ = \operatorname{Re}\left\{e^{j(w_0\ell+2f)}\,\dfrac{1-e^{j2w_0(N-\ell)}}{1-e^{+j2w_0}}\right\} \\ = \operatorname{Re}\left\{e^{jw_0\ell}\cdot e^{j2f}\cdot\dfrac{e^{jw_0(N-\ell)}\left(e^{-jw_0(N-\ell)} - e^{jw_0(N-\ell)}\right)}{e^{jw_0}\left(e^{-jw_0} - e^{jw_0}\right)}\right\} \\ = \operatorname{Re}\left\{e^{jw_0\ell}\cdot e^{j2f}\cdot e^{jw_0(N-\ell-1)}\,\dfrac{-2j\sin(w_0(N-\ell))}{-2j\sin w_0}\right\} \\ = \cos(w_0(N-1)+2f)\,\dfrac{\sin(w_0(N-\ell))}{\sin w_0} \end{cases}$$

and the 1st part:

$$\sum_{n=o}^{N-\ell-1} \cos w_0 \ell = (\cos w_0 \ell)(N-\ell)$$

$$\Rightarrow \hat{g}(\ell) = \frac{A^2}{2N}(N-\ell)\cos(w_0 \ell) + \frac{A^2}{2N}\cos(w_0(N-1)+2f)\frac{\sin(w_0(N-\ell))}{\sin w_0}$$

Now $\lim_{N\to\infty} \hat{g}(\ell) = \frac{A^2}{2}\cos w_0 \ell = g(\ell)$ if $w_0 \neq 0, p, 2p$

But f is a random variable and its presence in $\hat{g}(\ell)$ makes $\hat{g}(\ell)$ to be a random variable too. Let's see what kind of estimator it is.

$E[\hat{g}(\ell)] = \frac{N-\ell}{N} g(\ell) \Rightarrow$ it is a biased estimator. But, $E\left\{\underset{N\to\infty}{[\hat{g}(\ell)]}\right\} \to g(\ell)$. Therefore, it is asymptotically an unbiased estimator

How about its variance?

$$s^2_{\hat{g}(\ell)} = E\left\{\left|\hat{g}(\ell) - m_{\hat{g}(\ell)}\right|^2\right\} = \frac{A^2}{2\sqrt{2}}\left|\frac{\sin(N-\ell)w_0}{N\sin w_0}\right|$$

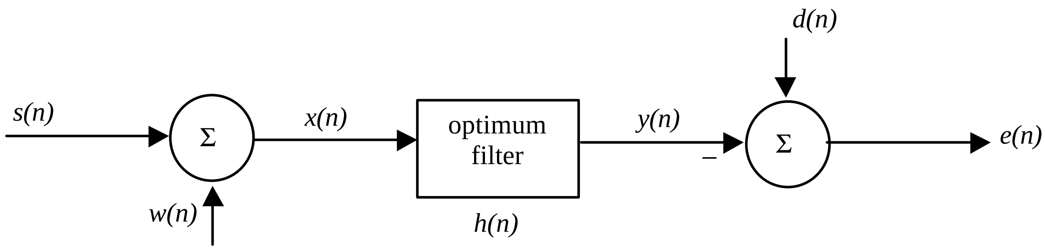
If we use the following definition for autocorrelation estimation, then it would become an unbiased estimator. In Matlab, you can also select this option.

$$\hat{\hat{g}}(\ell) = \frac{1}{N-\ell}\sum_{n=0}^{N-\ell-1} x_n^* x_{n+\ell}$$

This way, $E\left[\hat{\hat{g}}(\ell)\right] = g(\ell) \to$ unbiased estimator. But,

$s^2_{\hat{\hat{g}}(\ell)} = \frac{A^2}{2\sqrt{2}}\left|\frac{\sin(N-\ell)w_0}{(N-\ell)\sin w_0}\right| > s^2_{\hat{g}(\ell)}$ which means we get an unbiased estimator but at the cost of increasing the variance which is not desirable.

"Wiener Filters"



w(n) is white Gaussian noise and *d(n)* is the desired signal. We normally like to design the filter such that is suppresses the undesired interference component. There are 3 cases:

1) *d(n) = s(n)* then the linear estimation is referred to as filtering.
2) *d(n) = s (n + D)* and *D > o*, then the linear estimation is referred to as prediction. Note that this prediction is different from the predictions that we discussed so far.
3) *d(n) = s(n – D)*, then it is referred to as smoothing. The basic assumption is that *s(n)*, *w(n)* and *d(n)* are all WSS process with zero mean. The wiener filter is based on designing an optimum FIR/IIR filter in the minimum Mean-Square (MMSE) Error Sense.

FIR Wiener Filter

Since *h(n)* is finite, then *y(n)* (the output) depends on a finite data record *x(n)*, *x(n – 1)….*, *x(n – M + 1)*, where *M* is the order of the filter.

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

Cost function to be minimized is defined as

$$e_M = E\left\{|e(n)|^2\right\} = E\left\{|d(n) - y(n)|^2\right\}$$

Choosing this as the cost function has mathematical advantages such as having a unique minimum.

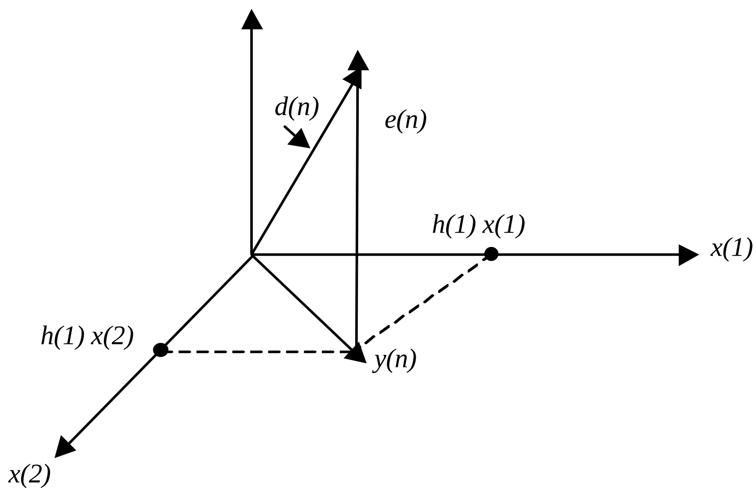
Orthogonality Principle in Linear MMSE Estimation

Output of the filter: $y(n) = \sum_{k=0}^{M-1} h(k) x(n-k)$

Error: $e(n) = d(n) - y(n)$

MSE: $e_M = E\left[|e(n)|^2\right]$ = the length of the vector *e(n)*

For a filter of degree *M = 2*, we have:



The principle of orthogonality states the length of this error vector is minimum, when *e(n)* is perpendicular to the data subspace (i.e. every *x(k)* point $o \le k \le M - 1$) or in other words $E\{e^*(n)\ x(n-k)\} = 0$. Where does this conclusion come from?

Let $h_k = a_k + jb_k$ for every filter coefficient. Then form the gradient vector:

$\nabla_k = \dfrac{\partial}{\partial a_k} + j\dfrac{\partial}{\partial b_k}$, $k = 0,1,...M-1$. Then

$$\nabla_k e_M = \nabla_k E\{|e(n)\cdot e^*(n)|\}$$
$$= E\left\{\frac{\partial e(n)}{a_k}e^*(n) + \frac{\partial e^*(n)}{a_k}e(n) + j\frac{\partial e(n)}{b_k}e^*(n) + j\frac{\partial e^*(n)}{\partial b_k}e(n)\right\}$$

but $e(n) = d(n) - y(n) = d(n) - \sum_{k=0}^{M-1}(a_k + jb_k)x(n-k)$. Therefore, we will have:

$$\frac{\partial e(n)}{\partial a_k} = -x(n-k) \qquad \frac{\partial e^*(n)}{\partial a_k} = -x^*(n-k)$$
$$\frac{\partial e(n)}{\partial b_k} = -jx(n-k) \qquad \frac{\partial e^*(n)}{\partial b_k} = +jx^*(n-k)$$

Substituting these into $\nabla_k e_M$ equation above, we get:

$$\nabla_k e_M = E\{-x(n-k)e^*(n) - x^*(n-k)e(n) + x(n-k)e^*(n) - x^*(n-k)e(n)\}$$
$$= E\{-2x^*(n-k)e(n)\} = -2E\{x^*(n-k)e(n)\}$$

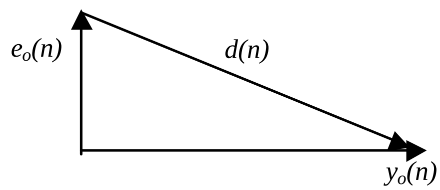
Therefore, minimizing $\nabla_K e_M = 0$, means that $E\{x(n-k)e_o^*(n)\} = 0$, where e_o is the estimation error that results when the filter operates at its optimum condition. This is called the principle of orthogonality.

Corollary to the Principle of Orthogonality

How about $E\{y(n)e^*(n)\}$?

$$E\{y(n)e^*(n)\} = E\left\{\sum_{k=o}^{M-1} h(k)x(n-k)e^*(n)\right\} = \sum_{k=o}^{M-1} h(k)E\{x(n-k)e^*(n)\}$$

Now when the filter is optimized, then $E\{x(n-k)e_o^*(n)\} = 0$ and given that we call the output at the optimal condition as $y_o(n)$, then we conclude: $E\{y_o(e)e_o^*(n)\} = 0$. It means that the optimum output is also orthogonal to the error.



Now we are ready to derive the filter coefficients for the optimum condition.

Using the principle of orthogonality, we have:

$$E\left\{x(n-\ell)\left[\underbrace{d^*(n) - \sum_{k=o}^{M-1} h_{ok}^* x^*(n-k)}_{e_o^*(n)}\right]\right\} = 0$$

$$\Rightarrow \sum_{k=o}^{M-1} h_{ok}^* E\left\{\underbrace{x(n-\ell)x^*(n-k)}_{g_{xx}(k-\ell)}\right\} = E\left\{\underbrace{x(n-\ell)d^*(n)}_{g_{xd}(-\ell)=g_{dx}^*(\ell)}\right\}$$

Getting the conjugate of the above equation:

Wiener-Hopf Equations:

$$\boxed{\sum_{k=0}^{M-1} h_{ok} g_{xx}(\ell-k) = g_{dx}(\ell), \qquad \ell = 0,1,\ldots, M-1}$$

g_{xx} is the teoplitz matrix of autocorrelation of *x(n)* and g_{dx} is the cross-correlation of *d(n)* and *x(n)* for *0 ≤ n ≤ M-1 (note this limits!)*

In matrix form: $\underline{\Gamma}_{xx_M} . \underline{h}_M = \underline{g}_{dx} \Rightarrow h_{opt} = \underline{\Gamma}_{xx_M}^{-1} \cdot \underline{g}_{dx}$

$$MMSE = Mine_M = s_d^2 - \sum_{k=0}^{n-1} h_{opt}(k) g_{dx}^*(k) = s_d^2 - \underline{g}_{dx}^{*T} \cdot \underline{\Gamma}_M^{-1} \cdot \underline{g}_{dx} \text{ and } s_d^2 = E\{|d(n)|\}^2.$$

Special Cases

If *d(n) = s(n)*, usually in practice *s(n)* and *w(n)* are uncorrelated. Therefore, $g_{xx}(k) = g_{ss}(k) + g_{ww}k$ and also $g_{dx} = g_{ss}(k) + \underbrace{g_{sw}(k)}_{0} = g_{ss}(k)$. Therefore, Wiener-Hopf equations become:

$$\sum_{k=0}^{M-1} h_o(k)[g_{ss}(\ell - k) + g_{ww}(\ell - k)] = g_{ss}(\ell) \quad \ell = 0,..., M-1$$

further $g_{ww}(\ell - k) = \begin{cases} s_w^2 & l = k \\ o & else \end{cases}$

If *d(n) = s(n + D), D > o*, then $g_{dx}(\ell) = g_{ss}(\ell + D)$. In both situations, the correlation matrix is teoplitz and Levinson-Durbin algorithm can be used to solve for optimum filter coefficients.

***Example*:**

We have a process *x(n) = s(n) + w(n)*, where we know $s_w^2 = 1$ and it is a white Gaussian noise. We also know that *s(n)* is an AR process described by the difference equation *s(n) = 0.6 s(n – 1) + v(n)*, where *v(n)* is also a white noise Gaussian noise with $s_v^2 = 0.64$. Design a Wiener Filter with degree of *2 = M* to estimate *s(n)*. Also determine MMSE at Stage 2.

Solution

Wiener-Hopf Equations: $\underline{\Gamma}_{xx} \cdot \underline{h}_{ok} = g_{dx}$. Assuming that *s(n)* and *w(n)* are uncorrelated, then we have $g_{xx} = g_{ss} + g_{ww}$ and also $g_{dx} = g_{ss}$ (note that *d(n)=s(n)*). Therefore, we need $g_{ss}(0)$ and $g_{ss}(1)$. In order to find g_{ss}, then we can use Yule-Walker equations:

$$\begin{bmatrix} g_{ss}(0) & g_{ss}(1) \\ g_{ss}(1) & g_{ss}(0) \end{bmatrix} \begin{bmatrix} 1 \\ -0.6 \end{bmatrix} = \begin{bmatrix} 0.64 \\ 0 \end{bmatrix}$$

$\Rightarrow g_{ss}(0) = 1$ and $g_{ss}(1) = 0.6$

$$\rightarrow g_{xx} = \underbrace{\begin{bmatrix} 1 \\ 0.6 \end{bmatrix}}_{g_{ss}} + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{g_{ww}} = \begin{bmatrix} 2 \\ 0.6 \end{bmatrix} \Rightarrow \Gamma_{xx} = \begin{bmatrix} 2 & 0.6 \\ 0.6 & 2 \end{bmatrix}$$

Now solving Wiener-Hopf Equations:

$$\begin{bmatrix} 2 & 0.6 \\ 0.6 & 2 \end{bmatrix} \begin{bmatrix} h_o(1) \\ h_o(2) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.6 \end{bmatrix} \Rightarrow h_o = \begin{bmatrix} 0.451 \\ 0.165 \end{bmatrix}$$

$$MMSE = s_d^2 - \sum_{k=0}^{1} h(k) g_{dx}(k) = 1 - \begin{bmatrix} 1 & 0.6 \end{bmatrix} \begin{bmatrix} 2 & 0.6 \\ 0.6 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0.6 \end{bmatrix} = 0.45$$