LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO -
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Technical Note **LIGO-T11XXXXX–vX** 2011/07/14

Derivation of formula for optimum Wiener filter coefficients

Ishwita Saikia, Manuel Marchiò

California Institute of Technology
LIGO Project, MS 18-34
Pasadena, CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project, Room NW22-295
Cambridge, MA 02139
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

LIGO Hanford Observatory
Route 10, Mile Marker 2
Richland, WA 99352
Phone (509) 372-8106
Fax (509) 372-8137
E-mail: info@ligo.caltech.edu

LIGO Livingston Observatory
19100 LIGO Lane
Livingston, LA 70754
Phone (225) 686-3100
Fax (225) 686-7189
E-mail: info@ligo.caltech.edu

http://www.ligo.caltech.edu/

1 Introduction

Consider the following filter:

$$
\begin{aligned}
\text{INPUT (witness signal)} \quad & u(0), u(1), u(2)... \qquad (\textit{time series}) \\
\text{IMPULSE RESPONSE (coefficients)} \quad & w_o, w_1, w_2, ... \\
\text{OUTPUT (estimated signal)} \quad & y(n) = \sum_{k=0}^{\infty} w_k^* u(n-k) \qquad n = 0, 1, 2, ...
\end{aligned} \tag{1}
$$

we denote w_k as a complex coefficient, such that

$$w_k = a_k + ib_k; \qquad k = 0, 1, 2, ... \tag{2}$$

hence its complex conjugation is $w_k^* = a_k - ib_k$; $k = 0, 1, 2, ...$

The above filter will estimate the desired response $d(n)$ with an *estimation error* defined by the equation 3. We have assumed that $d(n)$ and $u(n)$ are single realizations of stationary stochastic processes, both with zero mean.

$$e(n) = d(n) - y(n) \tag{3}$$

$$\Rightarrow e(n) = d(n) - \sum_{k=0}^{\infty} w_k^* u(n-k)$$

$$\Rightarrow e(n) = d(n) - \sum_{k=0}^{\infty} (a_k - ib_k) u(n-k) \tag{4}$$

$$e^*(n) = d^*(n) - \sum_{k=0}^{\infty} (a_k + ib_k) u^*(n-k) \tag{5}$$

2 Principle of orthogonality

To optimize the filter the following cost function has to be minimized:

$$J = E\left[e(n)e^*(n)\right] = E\left[|e(n)|^2\right] \tag{6}$$

Lets define a gradient operator:

$$\nabla_k = \frac{\partial}{\partial a_k} + i \frac{\partial}{\partial b_k} \tag{7}$$

So, to minimize J we put

$$\nabla_k J = \frac{\partial J}{\partial a_k} + i \frac{\partial J}{\partial b_k} = 0; \qquad k = 0, 1, 2, ... \tag{8}$$

and using 6 ...

$$E\left[\frac{\partial e(n)}{\partial a_k}e^*(n) + \frac{\partial e^*(n)}{\partial a_k}e(n) + \frac{\partial e(n)}{\partial b_k}ie^*(n) + \frac{\partial e^*(n)}{\partial b_k}ie(n)\right] = 0 \tag{9}$$

Calculating the partial derivatives of equation 9 using equations 4 and 5 ...

$$\begin{aligned}
\frac{\partial e(n)}{\partial a_k} &= -u(n-k) \\
\frac{\partial e^*(n)}{\partial a_k} &= -u^*(n-k) \\
\frac{\partial e(n)}{\partial b_k} &= iu(n-k) \\
\frac{\partial e^*(n)}{\partial b_k} &= -iu^*(n-k)
\end{aligned}$$

Substituting the value of above partial derivatives in equation 9 ...

$$\begin{aligned}
& E[-u(n-k)e^*(n) - u^*(n-k)e(n) + iu(n-k)ie^*(n) - iu^*(n-k)ie(n)] = 0 \\
\Rightarrow \quad & E[-u(n-k)e^*(n) - u^*(n-k)e(n) - u(n-k)e^*(n) + u^*(n-k)e(n)] = 0 \\
\Rightarrow \quad & E[-u(n-k)e^*(n) - \cancel{u^*(n-k)e(n)} - u(n-k)e^*(n) + \cancel{u^*(n-k)e(n)}] = 0 \\
\Rightarrow \quad & -2E[u(n-k)e^*(n)] = 0 \\
& \therefore\ E[u(n-k)e^*(n)] = 0
\end{aligned} \tag{10}$$

Equation 10 is the mathematical form of the *Principle of orthogonality*.
This principle is very important for the study of linear optimum filters since it tells us whether the filter is optimum. It is a necessary and sufficient condition for J to obtain its minimum value. Equation 10 also tells us that the estimation error $e_o(n)$ is orthogonal to each sample of the witness signal.

2.1 Corollary

A corollary of the principle of orthogonality can be derived by finding the correlation between the filter output $y(n)$ and the estimation error $e(n)$ as shown below...

$$E\left[y(n)e^*(n)\right] = E\left[\sum_{k=0}^{\infty} w_k^* u(n-k)e^*(n)\right] \tag{11}$$

$$= \sum_{k=0}^{\infty} w_k^* E\left[u(n-k)e^*(n)\right] \tag{12}$$

using the principle of orthogonality 10 we can see that the expectation values inside the summation are all zero, so we get the result

$$E\left[y(n)e^*(n)\right] = 0 \tag{13}$$

Equation 13 tells us that in an optimum filter the filter output $y_o(n)$ is orthogonal to the estimation error $e_o(n)$.

3 Normalized mean-square error

When the filter is working in the optimum condition, the filtered output may be written as $y_o(n)$ and its corresponding estimation error as $e_o(n)$.

Hence,

$$e_o(n) = d(n) - y_o(n) \tag{14}$$

For $e_o(n)$,

$$J_{min} = E[|e(n)|^2] \tag{15}$$

where $J_{min}$is the minimum mean-square error.

Calculating mean-square values for both sides of equation 14 and using equation 15 ...

$$\begin{aligned}
J_{min} = E[|e_o(n)|^2] =& E[(d(n) - y_o(n))\,(d^*(n) - y_o^*(n))] \\
=& E\left[|d(n)|^2 + |y_o(n)|^2 - d(n)y_o^*(n) - d^*(n)y_o(n)\right] \\
=& E\left[|d(n)|^2 + |y_o(n)|^2 - (y_o(n) + e_o(n))\,y_o^*(n) - (y_o^*(n) + e_o^*(n))\,y_o(n)\right] \\
=& E\left[|d(n)|^2 + |y_o(n)|^2 - y_o(n)y_o^*(n) - e_o(n)y_o^*(n) - y_o^*(n)y_o(n) + e_o^*(n)y_o(n)\right] \\
=& E\left[|d(n)|^2 - \cancel{e(n)y_o^*(n)} - |y_o(n)|^2 + \cancel{e_o^*(n)y_o(n)}\right] && (16) \\
=& E\left[|d(n)|^2 - |y_o(n)|^2\right] \\
=& \left(E\left[|d(n)|^2\right] - E\left[|d(n)|\right]^2\right) - \left(E\left[|y_o(n)|^2\right] - E\left[|y_o(n)|\right]^2\right) && (17) \\
=& \sigma_d^2(n) - \sigma_{y_o}^2(n)
\end{aligned}$$

In step 16 the cancellation occurs due to the corollary of the *Principle of orthogonality* 10. In step 17 the zero means of the quantities $d(n)$ and $y_0(n)$ were added to obtain the variances.

Hence,

$$J_{min} = \sigma_d^2 - \sigma_{y_o}^2 \tag{18}$$

where σ_d^2 and $\sigma_{y_o}^2$ are the variances of $d(n)$ and $y_0(n)$.

normalizing equation 18 ...

$$\frac{J_{min}}{\sigma_d^2} = 1 - \frac{\sigma_{y_o}^2}{\sigma_d^2}$$

$$\Rightarrow \varepsilon = 1 - \frac{\sigma_{y_o}^2}{\sigma_d^2} \tag{19}$$

where $\varepsilon = \frac{J_{min}}{\sigma_d^2}$ is known as the *normalized mean-square error* and it is always positive since J_{min} and σ_d^2 are always positive. So from equation 19 we can conclude that $0 \leq \varepsilon \leq 1$.
The value of ε tells us how the filter is working. If $\varepsilon = 0$ then we say that the filter is working perfectly in an optimum condition and if $\varepsilon = 1$ then the filter is working completely wrong (worst case).

4 Wiener-Hopf equations

Using *Principle of orthogonality* 10 and equation 5...

$$E[u(n-k)e^*(n)] = 0; \qquad k = 0, 1, 2, ...$$

$$\Rightarrow \quad E[u(n-k)(d^*(n) - \sum_{j=0}^{\infty} w_{0j} u^*(n-j)] = 0; \qquad k = 0, 1, 2, ...$$

$$\Rightarrow \quad E[u(n-k)d^*(n)] - \sum_{j=0}^{\infty} w_{0j} E[u(n-k)u^*(n-i)] = 0; \qquad k = 0, 1, 2, ...$$

$$\Rightarrow \quad \sum_{j=0}^{\infty} w_{0j} E[u(n-k)u^*(n-i)] = E[u(n-k)d^*(n)]; \; k = 0, 1, 2, ... \tag{20}$$

The expectation values defined in 20 are special.
From LHS of equation 20 we get $E[u(n-k)u^*(n-i)]$ which is the autocorrelation function of the witness samples $u(n-k)$ for a lag of $i-k$. Thus,

$$r(i-k) = E[u(n-k)u^*(n-i)] \tag{21}$$

From RHS of equation 20 we get $E[u(n-k)d^*(n)]$ which is the cross-correlation between the witness samples $u(n-k)$ and the desired response $d(n)$ for a lag of $-k$. Thus,

$$p(-k) = E[u(n-k)d^*(n)] \tag{22}$$

Using 21 and 22 in equation 20 gives...

$$\sum_{j=0}^{\infty} w_{0j} r(i-k) = p(-k); \quad k = 0, 1, 2, ... \tag{23}$$

The system of equations 23 are known as the *Wiener-Hopf equations*. These equations define the optimum coefficients of the filter. Now, lets denote these equations in the matrix form.

Let $\mathbf{R}$ denote the *M-by-M correlation matrix* of the witness samples $u(n), u(n-1), ..., u(n-M+1)$, that is,

$$\mathbf{R} = E\left[\mathbf{u}(n)\mathbf{u}^H(n)\right]$$

where $\mathbf{u}(n)$ is the M-by-1 column vector whose elements are $u(n), u(n-1), ..., u(n-M+1)$, so we have

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

Let $\mathbf{p}$ denote the the *M-by1 cross-correlation vector* between the witness samples $u(n)$and the desired response $d(n)$:

$$\mathbf{p} = E\left[\mathbf{u}(n)d^*(n)\right]$$

so we have

$$\mathbf{p} = \left[p(0), p(-1), ..., p(1-M)\right]^T$$

Let $\mathbf{w}_o$denote the M-by-1 optimum filter weights, so

$$\mathbf{w}_o = \left[w_{o,0}, w_{o,1}, ..., w_{o,M-1}\right]^T$$

To solve the Wiener-Hopf equations 23 we can rewrite them in the compact matrix form

$$\mathbf{R}\mathbf{w}_o = \mathbf{p} \tag{24}$$

we assume the *correlation matrix* to be nonsingular and find the optimum filter weights by inverting equation 24, thus we get

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p} \tag{25}$$

Equation 25 computes the coefficients of an optimum filter.