Lecture Notes on Wiener Filtering

Man-Wai MAK

Dept. of Electronic and Information Engineering,
The Hong Kong Polytechnic University
enmwmak@polyu.edu.hk

Abstract

This lecture note provides the detailed formulations and derivations of Wiener filtering and its application to speech enhancement. Its relationship with linear prediction will also be discussed. For simplicity, we will restrict to the case where the Wiener filter is an FIR filter. For IIR Wiener filters, refer to [1].

Please cite this document as: M.W. Mak, "Lecture Notes on Wiener Filtering", *Technical Report and Lecture Note Series, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University*, Feb 2016.

Keywords: Wiener filters, Speech enhancement

1. Wiener Filtering

Denote $y(n)$ as an observable signal and $x(n)$ as an unobservable target signal that we would like to recovered from $y(n)$. We use a linear time invariant FIR filter $H(z)$ with impulse response $h(n)$ to estimate $s(n)$ as follows:

$$\hat{x}(n) = h(n) * y(n) \approx x(n), \tag{1}$$

where $\hat{x}(n)$ is an estimate of $x(n)$. Eq. 1 can also be written as

$$\hat{x}(n) = \sum_{k=0}^{L-1} h(k)y(n-k) \tag{2}$$

where L is the length of the FIR filter $H(z)$. The error signal ϵ is given by

$$\epsilon = \mathbb{E}\left\{\left(\sum_{k=0}^{L-1} h(k)y(n-k) - x(n)\right)^2\right\} = \mathbb{E}\{[e(n)]^2\}, \tag{3}$$

where $\mathbb{E}$ is the expectation operator and

$$e(n) = \sum_{k=0}^{L-1} h(k)y(n-k) - x(n) \tag{4}$$

is the estimation error. Setting $\frac{\partial \epsilon}{\partial h(m)} = 0$, where $m = 0, \dots, L-1$, we have

$$\mathbb{E}\left\{\left(\sum_{k=0}^{L-1} h(k)y(n-k) - x(n)\right) y(n-m)\right\} = \mathbb{E}\{e(n)y(n-m)\} = R_{ey}(m) = 0, \tag{5}$$

where R_{ey} is the cross-correlation between $e(n)$ and $y(n)$. Note that Eq. 5 agrees with the orthogonality principle, which states that when the filter $\{h(m)\}_{m=0}^{L-1}$ is optimal, the error $e(n)$ is orthogonal to the data $y(n)$ that are used to compute the estimate $\hat{x}(n)$. Eq. 5 suggests that when the filter $H(z)$ is optimal, we have

$$\begin{aligned} R_{\hat{x}y}(m) &= \mathbb{E}\{\hat{x}(n)y(n-m)\} \\ &= \mathbb{E}\{x(n)y(n-m)\} \\ &= R_{xy}(m). \end{aligned} \tag{6}$$

To find $\{h(m)\}_{m=0}^{L-1}$, we express the cross-correlation between the estimated signal and the observed signal:

$$\begin{aligned} R_{\hat{x}y}(m) &= \mathbb{E}\left\{\hat{x}(n)y(n-m)\right\} \\ &= \mathbb{E}\left\{\left[\sum_{k=-\infty}^{\infty} h(k)y(n-k)\right] y(n-m)\right\} \\ &= \mathbb{E}\left\{\left[\sum_{k=-\infty}^{\infty} h(n-k)y(n)\right] y(n-m)\right\} \\ &= \sum_{k=-\infty}^{\infty} h(n-k)\mathbb{E}\left\{y(n)y(n-m)\right\} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=-\infty}^{\infty} h(n-k) R_{yy}(m) \\
&= \sum_{k=-\infty}^{\infty} h(m-k) R_{yy}(m) \\
&= h(m) * R_{yy}(m).
\end{aligned} \tag{7}
$$

Substituting Eq. 7 into Eq. 6, we have

$$
\begin{aligned}
h(m) * R_{yy}(m) &= R_{xy}(m) \\
\Longrightarrow \sum_{k=0}^{L-1} h(k) R_{yy}(m-k) &= R_{xy}(m).
\end{aligned} \tag{8}
$$

This gives L equations for $m = 0, \ldots, L-1$:

$$
\begin{aligned}
\begin{bmatrix} R_{yy}(0) & R_{yy}(-1) & \cdots & R_{yy}(1-L) \\ R_{yy}(1) & R_{yy}(0) & \cdots & R_{yy}(2-L) \\ \vdots & \vdots & \ddots & \vdots \\ R_{yy}(L-1) & R_{yy}(L-2) & \cdots & R_{yy}(0) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(L-1) \end{bmatrix} &= \begin{bmatrix} R_{xy}(0) \\ R_{xy}(1) \\ \vdots \\ R_{xy}(L-1) \end{bmatrix} \\
\Longrightarrow \quad \mathbf{R}\mathbf{h} &= \mathbf{r} \\
\Longrightarrow \quad \mathbf{h} &= \mathbf{R}^{-1}\mathbf{r}.
\end{aligned} \tag{9}
$$

The frequency domain of Eq. 8 is

$$
H(\omega) S_{yy}(\omega) = S_{xy}(\omega), \tag{10}
$$

where $H(\omega)$, $S_{yy}(\omega)$ and $S_{xy}(\omega)$ are the Fourier transform of $h(m)$, $R_{yy}(m)$, and $R_{xy}(m)$, respectively. Therefore, we have

$$
H(\omega) = \frac{S_{xy}(\omega)}{S_{yy}(\omega)}. \tag{11}
$$

2. Wiener Filter for Speech Enhancement

When applying the Wiener filter for speech enhancement, we assume the following additive noise model:

$$
y(n) = x(n) + b(n) \tag{12}
$$

where $y(n)$ is the observed noisy speech, $x(n)$ is the unobservable clean speech, and $b(n)$ is the stationary background noise which is assumed to be uncorrelated with $x(n)$. We use an optimal Wiener filter $H(z)$ to estimate the clean speech:

$$\hat{x}(n) = h(n) * y(n). \tag{13}$$

Based on the development of Section 1, when $H(z)$ is optimal we have the relationship in Eq. 11. To determine $S_{xy}(\omega)$ and $S_{yy}(\omega)$ in Eq. 11, we compute the following autocorrelation and cross-correlation:

$$\begin{aligned} R_{yy}(m) &= \mathbb{E}\{y(n)y(n-m)\} \\ &= \mathbb{E}\{[x(n)+b(n)][x(n-m)+b(n-m)]\} \\ &= \mathbb{E}\{x(n)x(n-m)\} + \mathbb{E}\{b(n)b(n-m)\} + 0 \\ &= R_{xx}(m) + R_{bb}(m) \end{aligned} \tag{14}$$

$$\begin{aligned} R_{xy}(m) &= \mathbb{E}\{[x(n)+b(n)]x(n-m)\} \\ &= \mathbb{E}\{x(n)x(n-m)\} + \mathbb{E}\{b(n)x(n-m)\} \\ &= R_{xx}(m) + 0 \\ &= R_{xx}(m) \end{aligned} \tag{15}$$

Taking Fourier transform of Eq. 14 and Eq. 15, we have

$$S_{yy}(\omega) = S_{xx}(\omega) + S_{bb}(\omega) \quad \text{and} \quad S_{xy}(\omega) = S_{xx}(\omega). \tag{16}$$

Substituting Eq. 16 into Eq. 11, we have

$$H(\omega) = \frac{S_{xx}(\omega)}{S_{xx}(\omega) + S_{bb}(\omega)}.$$

Note that $S_{xx}(\omega)$ and $S_{bb}(\omega)$ are the power spectrum of $x(n)$ and $b(n)$, respectively. However, we do not have $x(n)$. Therefore, we approximate $S_{xx}(\omega)$ by $S_{\hat{x}\hat{x}}(\omega)$, which results in

$$H(\omega) = \frac{S_{\hat{x}\hat{x}}(\omega)}{S_{\hat{x}\hat{x}}(\omega) + S_{bb}(\omega)}. \tag{17}$$

For the first frame, neither $H(z)$ nor $\hat{x}(n)$ is known. It is a kind of chicken-and-egg problem, because according to Eq. 17, $H(\omega)$ depends on $S_{\hat{x}\hat{x}}$, which in turn depends on $H(\omega)$ [see Eq. 1]. A possible solution is to set $H(z) = 1$ so that $S_{\hat{x}\hat{x}}(\omega) = S_{yy}(\omega)$. This estimate is then substituted into Eq. 17

to obtain $H(\omega)$. Then, this $H(\omega)$ is used to reestimate the speech, i.e., $S_{\hat{x}\hat{x}}(\omega) = H(\omega)S_{yy}(\omega)$, and the cycle is repeated a few time. This will give a good initial estimate of $H(\omega)$ for the first frame and also a good starting point for estimating the $S_{\hat{x}\hat{x}}(\omega)$ of the next frame.

3. Relation with Linear Prediction

In linear prediction, the current sample $s(n)$ is predicted from the past P samples:

$$s(n) = \sum_{k=1}^{P} a_k s(n-k) + e_s(n), \tag{18}$$

where $a_k, k = 1, \ldots, P$, are the linear prediction coefficients and $e_s(n)$ is the prediction error. Note that Eq. 18 is very similar to Eq. 4, which can be rewritten as

$$x(n) = \sum_{k=0}^{L-1} h(k)y(n-k) + e_x(n). \tag{19}$$

Therefore, minimizing $\mathbb{E}\{e_s(n)^2\}$ with respect to $\{a_k\}_{k=1}^{P}$ will lead to a set of equations very similar to the minimization of $\mathbb{E}\{e_x(n)^2\}$, i.e., Eq. 8. More specifically, when $\{a_k\}_{k=1}^{P}$ are optimal, we have

$$\sum_{k=1}^{P} a_k R_{ss}(m-k) = R_{ss}(m), \qquad m = 1, \ldots, P \tag{20}$$

where $R_{ss}(m)$ is the autocorrelation of $s(n)$ at time lag m. Note that because $R_{ss}(m)$ is the autocorrelation of $s(n)$, we have $R_{ss}(m-k) = R_{ss}(k-m)$. This property leads to the matrix equation:

$$\begin{bmatrix} R_{ss}(0) & R_{ss}(1) & \cdots & R_{ss}(P-1) \\ R_{ss}(1) & R_{ss}(0) & \cdots & R_{ss}(P-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_{ss}(P-1) & R_{ss}(P-2) & \cdots & R_{ss}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} R_{ss}(1) \\ R_{ss}(2) \\ \vdots \\ R_{ss}(P) \end{bmatrix}$$
$$\Longrightarrow \quad \mathbf{R}_s \mathbf{a} = \mathbf{r}_s$$
$$\Longrightarrow \quad \mathbf{a} = \mathbf{R}_s^{-1} \mathbf{r}_s. \tag{21}$$

References

[1] Alan Oppenheim and George Verghese, *Introduction to Communication, Control, and Signal Processing*, Massachusetts Institute of Technology: MIT OpenCourseWare), http://ocw.mit.edu, Spring 2010.