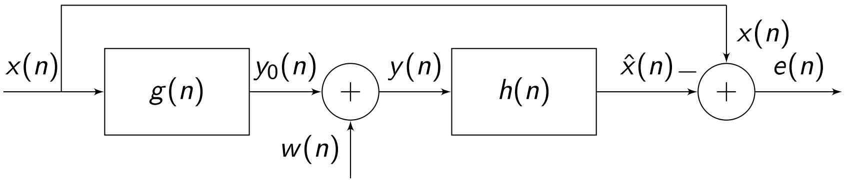
Applied Signal Processing
Lecture 10

Tomas McKelvey

Department of Electrical Engineering
Chalmers University of Technology

Agenda

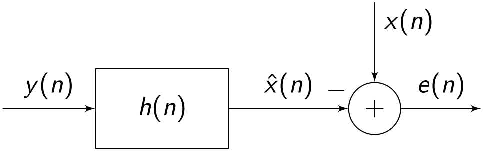
- Optimal Filtering
  - Wiener Filter
  - FIR Wiener Filter
- Deriving filter from data, solving the data based Least-Squares problem
- Recursive Least-Squares (RLS)



Assume $w(n)$ and $x(n)$ uncorrelated.

$$S_{yy}(\omega) = |G(\omega)|^2 S_{xx}(\omega) + S_{ww}(\omega)$$
$$S_{xy}(\omega) = G(\omega) S_{xx}(\omega)$$

How to select $h(n)$?



A different signal setup is

$$y(n) = s(n) + w(n) \tag{1}$$

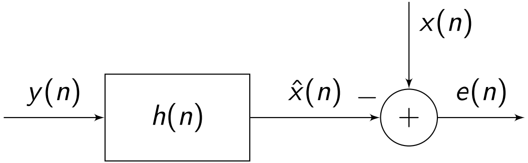
where $s(n)$ is a signal of importance and $w(n)$ is a disturbance.

Filtering If we want to make the best causal estimate of the signal $s(n)$ from $y(n)$ we set the desired signal $x(n) = s(n)$.

Prediction If we want to use the signal $y(n)$ to produce estimates of future values of $s(n)$ we set $x(n) = s(n + k)$ for some positive prediction horizon k.

Smoothing If the objective is to obtain a good estimate of past values of $s(n)$ based on observed values $y(n)$ we set $x(n) = s(n - k)$ for some positive smoothing lag k.

Optimal Filtering



We want to find filters h which make e small.

We assume $y(n)$ and $x(n)$ are zero mean, stochastic processes.

Making $e(n)$ small $\Rightarrow$ minimize $\mathbf{E}\, e^2(n)$

The variance of $e(n)$

We have $e(n) = x(n) - \hat{x}(n)$ and hence

$$e(n) = x(n) - \sum_{k=-\infty}^{\infty} h(k)y(n-k)$$

The variance is given by

$$\begin{aligned}
\mathbf{E}\, e^2(n) &= \mathbf{E}\left\{ \left(x(n) - \sum_{k=-\infty}^{\infty} h(k)y(n-k) \right)^2 \right\} \\
&= \mathbf{E}\left\{ \left(x(n) - \sum_{k=-\infty}^{\infty} h(k)y(n-k) \right) \left(x(n) - \sum_{m=-\infty}^{\infty} h(m)y(n-m) \right) \right\} \\
&= \mathbf{E}\left\{ x^2(n) - 2\sum_{m=-\infty}^{\infty} h(m)x(n)y(n-m) \right. \\
&\quad \left. + \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h(k)h(m)y(n-k)y(n-m) \right\}
\end{aligned}$$

The Variance

$$
\begin{aligned}
\mathbf{E}\, e^2(n) &= \mathbf{E}\left\{ x^2(n) - 2 \sum_{m=-\infty}^{\infty} h(m) x(n) y(n-m) \right. \\
&\quad \left. + \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h(k) h(m) y(n-k) y(n-m) \right\} \\
&= \mathbf{E}\, x^2(n) - 2 \sum_{m=-\infty}^{\infty} h(m)\, \mathbf{E}\left\{ x(n) y(n-m) \right\} \\
&\quad + \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h(k) h(m)\, \mathbf{E}\left\{ y(n-k) y(n-m) \right\} \\
&= \phi_{xx}(0) - 2 \sum_{m=-\infty}^{\infty} h(m) \phi_{yx}(m) + \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h(k) h(m) \phi_{yy}(m-k)
\end{aligned}
$$

Minimize the variance

The variance $e^2(n)$ is a non-negative quadratic function of $h(n)$.

The minimum is attained when the gradient w.r.t. $h(j)$ fo$j$ is equal to zero.

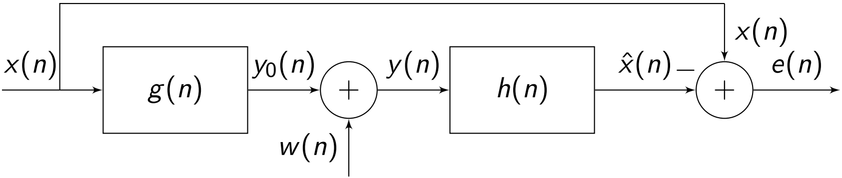
$$\frac{d}{d\,h(j)}\left(\phi_{xx}(0) - 2\sum_{m=-\infty}^{\infty} h(m)\phi_{yx}(m) + \sum_{k=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} h(k)h(m)\phi_{yy}(m-k)\right)$$

$$= -2\phi_{yx}(j) + 2\sum_{k=-\infty}^{\infty} h(k)\phi_{yy}(j-k) = 0$$

The optimal filter satisfy the equation

$$\phi_{yx}(j) = \sum_{k=-\infty}^{\infty} h(k)\phi_{yy}(j-k) \quad \Leftrightarrow \quad S_{yx}(\omega) = H(\omega)S_{yy}(\omega)$$

$$H(\omega) = \frac{S_{yx}(\omega)}{S_{yy}(\omega)} = \frac{S^*_{xy}(\omega)}{S_{yy}(\omega)}$$



We have

$$S_{yy}(\omega) = |G(\omega)|^2 S_{xx}(\omega) + S_{ww}(\omega)$$
$$S_{xy}(\omega) = G(\omega) S_{xx}(\omega)$$

Optimal filter

$$H(\omega) = \frac{S_{yx}(\omega)}{S_{yy}(\omega)} = \frac{S_{xy}^*(\omega)}{S_{yy}(\omega)}$$
$$= \frac{G^*(\omega) S_{xx}(\omega)}{|G(\omega)|^2 S_{xx}(\omega) + S_{ww}(\omega)} = \frac{1}{G(\omega)} \times \frac{1}{1 + \frac{S_{ww}(\omega)}{|G(\omega)|^2 S_{xx}(\omega)}}$$

FIR Wiener filter

If we restrict h to be a causal FIR filter of length M. The optimal filter satisfies, for $j = 0, 1, \ldots, M-1$

$$\phi_{yx}(j) = \sum_{k=0}^{M-1} h(k)\phi_{yy}(j-k)$$

$$\underbrace{\begin{bmatrix} \phi_{yx}(0) \\ \phi_{yx}(1) \\ \vdots \\ \phi_{yx}(M-1) \end{bmatrix}}_{\Phi_{yx}} = \underbrace{\begin{bmatrix} \phi_{yy}(0) & \phi_{yy}(1) & \cdots & \phi_{yy}(M-1) \\ \phi_{yy}(1) & \phi_{yy}(0) & \cdots & \phi_{yy}(M-2) \\ \vdots & & \ddots & \\ \phi_{yy}(M-1) & \phi_{yy}(M-2) & \cdots & \phi_{yy}(0) \end{bmatrix}}_{\Phi_{yy}} \underbrace{\begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(M-1) \end{bmatrix}}_{\mathbf{h}}$$

which yields

$$\mathbf{h} = \Phi_{yy}^{-1}\Phi_{yx}$$

and the minimum variance is

$$\mathbf{E}\, e^2(n) = \phi_{xx}(0) - \mathbf{h}^T\Phi_{yx} = \phi_{xx}(0) - \Phi_{yx}^T\Phi_{yy}^{-1}\Phi_{yx}$$

Determine the filter from data

- *The estimation problem:* How to derive the (stationary) optimal filter based on measured data.
- *The adaption problem:* How to adaptively update the filter when the signal properties slowly changes over time (quasi-stationary).

Introduce the data based criterion

$$L_N(\mathbf{h}) \triangleq \frac{1}{N+1}\sum_{n=0}^{N} e^2(n) = \frac{1}{N+1}\sum_{n=0}^{N}\Big(x(n) - \mathbf{h}^T\mathbf{y}(n)\Big)^2$$

where

$$\mathbf{y}(n) \triangleq \begin{bmatrix} y(n) & y(n-1) & \cdots & y(n-M+1) \end{bmatrix}^T$$

$$L_N(\mathbf{h}) = \frac{1}{N+1}\sum_{n=0}^{N}\Big(x^2(n) - 2\mathbf{h}^T\mathbf{y}(n)x(n) + \mathbf{h}^T\mathbf{y}(n)\mathbf{y}(n)^T\mathbf{h}\Big)$$

Optimal sample based filter

Introducing the notation

$$\mathbf{R}_{yy}(N) \triangleq \sum_{n=0}^{N} \mathbf{y}(n)\mathbf{y}^T(n), \quad \mathbf{R}_{yx}(N) \triangleq \sum_{n=0}^{N} \mathbf{y}(n)x(n)$$

The filters which minimizing L_N should satisfy

$$\mathbf{R}_{yy}(N)\mathbf{h} = \mathbf{R}_{yx}(N)$$

and if $\mathbf{R}_{yy}(N)$ has full rank the solution is unique and can be written as

$$\hat{\mathbf{h}}(N) = \mathbf{R}_{yy}^{-1}(N)\mathbf{R}_{yx}(N).$$

If the signals $y(n)$ and $x(n)$ are ergodic processes we have that

$$\lim_{N\to\infty} \frac{1}{N+1}\mathbf{R}_{yy}(N) = \Phi_{yy}, \quad \lim_{N\to\infty} \frac{1}{N+1}\mathbf{R}_{yx}(N) = \Phi_{yx} \tag{2}$$

and, if for some N_0, $\mathbf{R}_{yy}(N)$ is non-singular for all $N > N_0$ we obtain

$$\lim_{N\to\infty} \hat{\mathbf{h}}(N) = \Phi_{yy}^{-1}\Phi_{yx} = \mathbf{h}_{\text{opt}} \tag{3}$$

Comments

- Sample based filter converges to the optimal
- Expensive if we want to estimate a filter for each sample
    - Complexity of order M^3 for each sample (Matrix inversion)

Can we do better?

Recursive Least-Squares (RLS)

Can we recursively do the filtering and filter update efficiently?

Assume we already have $\mathbf{R}_{yy}^{-1}(N-1)$ and $\hat{\mathbf{h}}(N-1)$ then

$$\hat{x}(N) = \hat{\mathbf{h}}^T(N-1)\mathbf{y}(N), \quad \text{with error} \quad e(N) = x(N) - \hat{x}(N).$$

It can be shown that the inverse has the update (Sherman-Morrison-Woodbury formula)

$$\mathbf{R}_{yy}^{-1}(N) = \mathbf{R}_{yy}^{-1}(N-1) - \frac{\mathbf{R}_{yy}^{-1}(N-1)\mathbf{y}(N)\mathbf{y}^T(N)\mathbf{R}_{yy}^{-1}(N-1)}{1 + \mathbf{y}^T(N)\mathbf{R}_{yy}^{-1}(N-1)\mathbf{y}(N)}$$

and finally the filter update is given by

$$\hat{\mathbf{h}}(N) = \hat{\mathbf{h}}(N-1) + \mathbf{R}_{yy}^{-1}(N)\mathbf{y}(N)e(N)$$

- The complexity is of order M^2
- If initialized in the true LS-solution at $N = N_0$ (assuming $\mathbf{R}_{yy}(N_0)$ non-singular) then RLS provides the optimal solution for all $N > N_0$

We can introduce adaptivity be "forgetting" old data in the LS-criterion:

$$L_N(\mathbf{h}) = \sum_{n=0}^{N} e^2(n)\alpha^{N-n}$$

where $0 < \alpha \leq 1$. The RLS algorithm needs the following change:

$$\mathbf{R}_{yy}^{-1}(N) = \frac{1}{\alpha}\left[\mathbf{R}_{yy}^{-1}(N-1) - \frac{\mathbf{R}_{yy}^{-1}(N-1)\mathbf{y}(N)\mathbf{y}^T(N)\mathbf{R}_{yy}^{-1}(N-1)}{\alpha + \mathbf{y}^T(N)\mathbf{R}_{yy}^{-1}(N-1)\mathbf{y}(N)}\right]$$