Statistical Signal Processing
3. Discrete Wiener Filtering

Dr. Chau-Wai Wong

Electrical & Computer Engineering
North Carolina State University

Readings: Haykin 4th Ed. Chapter 2, Hayes Chapter 7

Contact: *chauwai.wong@ncsu.edu*. Updated: October 26, 2020.

Acknowledgment: ECE792-41 slides were adapted from ENEE630 slides developed by Profs. K.J. Ray Liu and Min Wu at the University of Maryland.

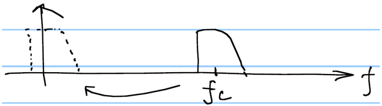
Preliminaries

- Why prefer FIR filters over IIR?

  $\Rightarrow$ FIR is inherently stable.

- Why consider complex signals?
  - Baseband representation is complex valued for narrow-band messages modulated at a carrier frequency.
  - Corresponding filters are also in complex form.

$u[n] = u_I[n] + ju_Q[n]$



- $u_I[n]$: in-phase component •

$u_Q[n]$: quadrature component

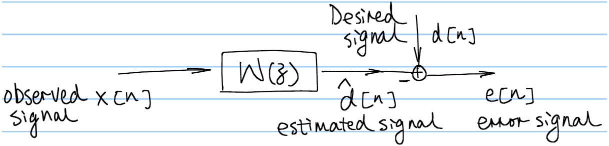
the two parts can be amplitude modulated by $\cos 2\pi f_c t$ and $\sin 2\pi f_c t$.

Preliminaries

- In many communication and signal processing applications, messages are modulated onto a carrier wave. The bandwidth of message is usually much smaller than the carrier frequency $\Rightarrow$ i.e., the signal modulated is "narrow-band".
- It is convenient to analyze in the baseband form to remove the effect of the carrier wave by translating signal down in frequency yet fully preserve the information in the message.
- The baseband signal so obtained is complex in general. $u[n] = u_I[n] + ju_Q[n]$
- Accordingly, the filters developed for the applications are also in complex form to preserve the mathematical formulations and elegant structures of the complex signal in the applications.

(1) General Problem

(Ref: Hayes §7.1)



Want to process $x[n]$ to minimize the difference between the estimate and the desired signal in some sense:

A major class of estimation (for simplicity & analytic tractability) is to use linear combinations of $x[n]$ (i.e. via linear filter).

When $x[n]$ and $d[n]$ are from two w.s.s. random processes, we often choose to minimize the mean-square error as the performance index.

$$\min_{\underline{w}} J \triangleq \mathbb{E}\left[|e[n]|^2\right] = \mathbb{E}\left[\left|d[n] - \hat{d}[n]\right|^2\right]$$

(2) Categories of Problems under the General Setup

1. Filtering
2. Smoothing
3. Prediction
4. Deconvolution

Wiener Problems: Filtering & Smoothing

- Filtering
    - The classic problem considered by Wiener
    - $x[n]$ is a noisy version of $d[n]$: $x[n] = d[n] + v[n]$
    - The goal is to estimate the true $d[n]$ using a causal filter (i.e., from the current and post values of $x[n]$)
    - The causal requirement allows for filtering on the fly
- Smoothing
    - Similar to the filtering problem, except the filter is allowed to be non-causal (i.e., all the $x[n]$ data is available)

Wiener Problems: Prediction & Deconvolution

- Prediction
  - The causal filtering problem with $d[n] = x[n+1]$, i.e., the Wiener filter becomes a linear predictor to predict $x[n+1]$ in terms of the linear combination of the previous value $x[n], x[n-1], , \dots$
- Deconvolution
  - To estimate $d[n]$ from its filtered (and noisy) version $x[n] = d[n] * g[n] + v[n]$
  - If $g[n]$ is also unknown $\Rightarrow$ blind deconvolution. We may iteratively solve for both unknowns

FIR Wiener Filter for w.s.s. processes

Design an FIR Wiener filter for jointly w.s.s. processes $\{x[n]\}$ and $\{d[n]\}$:

$W(z) = \sum_{k=0}^{M-1} a_k z^{-k}$ (where a_k can be complex valued)

$\hat{d}[n] = \sum_{k=0}^{M-1} a_k x[n-k] = \underline{a}^T \underline{x}[n]$ (in vector form)

$\Rightarrow e[n] = d[n] - \hat{d}[n] = d[n] - \underbrace{\sum_{k=0}^{M-1} a_k x[n-k]}_{\hat{d}[n]=\underline{a}^T \underline{x}[n]}$

By summation-of-scalar:

$$J = E\left[|e[n]|^2\right] = E\left[e[n]\, e^*[n]\right]$$

$$= E\left[|d[n]|^2\right] - E\left[d[n] \sum_{k=0}^{M-1} a_k^* x^*[n-k]\right] - E\left[d^*[n] \sum_{k=0}^{M-1} a_k x[n-k]\right] + E\left[\sum_{k=0}^{M-1} \sum_{l=0}^{M-1} a_k a_l^* x[n-k] x^*[n-l]\right]$$

$$= E\left[|d[n]|^2\right] - \sum_{k=0}^{M-1} a_k^* E\left[d[n]\, x^*[n-k]\right] - \sum_{k=0}^{M-1} a_k E\left[d^*[n]\, x[n-k]\right] + \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} a_k a_l^* \underbrace{E\left[x[n-k] x^*[n-l]\right]}_{r_x(l-k)}$$

FIR Wiener Filter: J in matrix-vector form

$$J = \mathbb{E}\left[(d[n] - \underline{a}^T \underline{x}[n])(d^*[n] - \underline{x}^H[n]\underline{a}^*)\right]$$

$$= \mathbb{E}\left[|d[n]|^2\right] - \underline{a}^H \underline{p}^* - \underline{p}^T \underline{a} + \underline{a}^H \mathbf{R}^T \underline{a}$$

where

$$\underline{x}[n] = \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-M+1 \end{bmatrix}, \quad \underline{p} = \begin{bmatrix} \mathbb{E}\left[x[n]d^*[n]\right] \\ \vdots \\ \mathbb{E}\left[x[n-M+1]d^*[n]\right] \end{bmatrix}, \quad \underline{a} = \begin{bmatrix} a_0 \\ \vdots \\ a_{M-1} \end{bmatrix}.$$

- $\mathbb{E}\left[|d[n]|^2\right]$: σ^2 for zero-mean random process
- $\underline{a}^H \mathbf{R}^T \underline{a}$: represent $\mathbb{E}\left[\underline{a}^T \underline{x}[n] \underline{x}^H[n] \underline{a}^*\right] = \underline{a}^T \mathbf{R} \underline{a}^*$

Perfect Square

1. If $\mathbf{R}$ is positive definite, $\mathbf{R}^{-1}$ exists and is positive definite.
2. $(\mathbf{R}\underline{a}^* - \underline{p})^H \mathbf{R}^{-1} (\mathbf{R}\underline{a}^* - \underline{p}) = (\underline{a}^T \mathbf{R}^H - \underline{p}^H)(\underline{a}^* - \mathbf{R}^{-1}\underline{p})$
   $= \underline{a}^T \mathbf{R}^H \underline{a}^* - \underline{p}^H \underline{a}^* - \underline{a}^T \underbrace{\mathbf{R}^H \mathbf{R}^{-1}}_{=\mathbb{I}} \underline{p} + \underline{p}^H \mathbf{R}^{-1} \underline{p}$

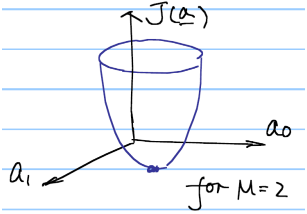
Thus we can write $J(\underline{a})$ in the form of perfect square:

$$J(\underline{a}) = \underbrace{\mathbb{E}\left[|d[n]|^2\right] - \underline{p}^H \mathbf{R}^{-1} \underline{p}}_{\text{Not a function of } \underline{a}\text{; Represent } J_{\min}.} + \underbrace{(\mathbf{R}\underline{a}^* - \underline{p})^H \mathbf{R}^{-1} (\mathbf{R}\underline{a}^* - \underline{p})}_{>0 \text{ except being zero if } \mathbf{R}\underline{a}^* - \underline{p} = 0}$$

Perfect Square

$J(\underline{a})$ represents the error performance surface:

convex and has unique minimum at $\mathbf{R}\underline{a}^* = \underline{p}$



Thus the necessary and sufficient condition for determining the optimal linear estimator (linear filter) that minimizes MSE is

$$\mathbf{R}\underline{a}^* - \underline{p} = 0 \Rightarrow \mathbf{R}\underline{a}^* = \underline{p}$$

This equation is known as the **Normal Equation**.
A FIR filter with such coefficients is called a **FIR Wiener filter**.

Perfect Square

$\mathbf{R}\underline{a}^* = \underline{p} \quad \therefore \underline{a}^*_{\text{opt}} = \mathbf{R}^{-1}\underline{p}$ if $\mathbf{R}$ is not singular
(which often holds due to noise)

When $\{x[n]\}$ and $\{d[n]\}$ are jointly w.s.s.
(i.e., crosscorrelation depends only on time difference)

$$\underset{R^T}{\begin{bmatrix} r_x(0) & r_x^*(1) & & \\ r_x(1) & r_x(0) & \ddots & \\ \vdots & & \ddots & \\ r_x(M-1) & & & \ddots \end{bmatrix}} \underset{\underline{a}}{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{M-1} \end{bmatrix}} = \underset{\underline{p}^*}{\begin{bmatrix} r_{dx}(0) \\ \vdots \\ r_{dx}(M-1) \end{bmatrix}}$$

This is also known as the Wiener-Hopf equation (the discrete-time counterpart of the continuous Wiener-Hopf integral equations)

Principle of Orthogonality

Note: to minimize a real-valued func. $f(z, z^*)$ that's analytic (differentiable everywhere) in z and z^*, set the derivative of f w.r.t. either z or z^* to zero.

- Necessary condition for minimum $J(\underline{a})$: (nece.&suff. for convex J)

$$\frac{\partial}{\partial a_k^*} J = 0 \text{ for } k = 0, 1, \ldots, M-1.$$

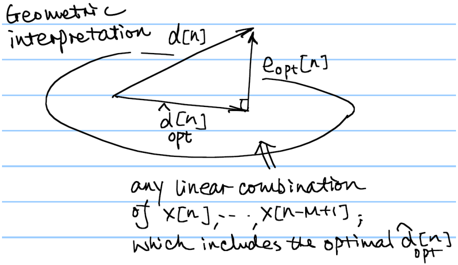
$$\Rightarrow \frac{\partial}{\partial a_k^*} \mathbb{E}\left[e[n]e^*[n]\right] = \mathbb{E}\left[e[n] \frac{\partial}{\partial a_k^*} \left(d^*[n] - \sum_{j=0}^{M-1} a_j^* x^*[n-j]\right)\right]$$
$$= \mathbb{E}\left[e[n] \cdot (-x^*[n-k])\right] = 0$$

Principal of Orthogonality

$$\mathbb{E}\left[e_{\text{opt}}[n] x^*[n-k]\right] = 0 \text{ for } k = 0, \ldots, M-1.$$

The optimal error signal $e_{\text{opt}}[n] = d[n] - \sum_{j=0}^{M-1} a_j^{\text{opt}} x[n-j]$ and each of the M samples of $x[n]$ that participated in the filtering are <u>statistically uncorrelated</u> (i.e., orthogonal in a statistical sense)

Principle of Orthogonality: Geometric View



Analogy:
r.v. $\Rightarrow$ vector;
E(XY) $\Rightarrow$ inner product of vectors

$\Rightarrow$ The optimal $\hat{d}[n]$ is the projection of $d[n]$ onto the subspace spanned by $\{x[n], \ldots, x[n-M+1]\}$ in a statistical sense.

The vector form: $\mathbb{E}\left[\underline{x}[n]e^*_{\text{opt}}[n]\right] = \underline{0}$.

This is true for any linear combination of $\underline{x}[n]$ and for FIR & IIR:

$$\mathbb{E}\left[\hat{d}_{\text{opt}}[n]e_{\text{opt}}[n]\right] = 0$$

Minimum Mean Square Error

Recall the perfect square form of J:

$$J(\underline{a}) = \underbrace{\mathbb{E}\left[|d[n]|^2\right] - \underline{p}^H \mathbf{R}^{-1} \underline{p}} + \underbrace{(\mathbf{R}\underline{a}^* - \underline{p})^H \mathbf{R}^{-1} (\mathbf{R}\underline{a}^* - \underline{p})}$$

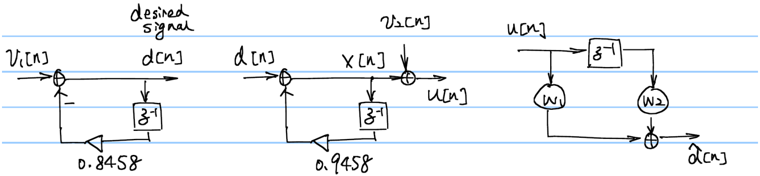
$$\therefore J_{\min} = \sigma_d^2 - \underline{a}_o^H \underline{p}^* = \sigma_d^2 - \underline{p}^H \mathbf{R}^{-1} \underline{p}$$

Also recall $d[n] = \hat{d}_{\text{opt}}[n] + e_{\text{opt}}[n]$. Since $\hat{d}_{\text{opt}}[n]$ and $e_{\text{opt}}[n]$ are uncorrelated by the principle of orthogonality, the variance is

$$\sigma_d^2 = \text{Var}(\hat{d}_{\text{opt}}[n]) + J_{\min}$$

$$\begin{aligned} \therefore \text{Var}(\hat{d}_{\text{opt}}[n]) &= \underline{p}^H \mathbf{R}^{-1} \underline{p} \\ &= \underline{a}_0^H \underline{p}^* = \underline{p}^H \underline{a}_o^* = \underline{p}^T \underline{a}_o \quad \text{real and scalar} \end{aligned}$$

Example and Exercise



$\{v_1[n]\} \sim$ white noise zero-mean σ_1^2 (real-valued)

distortion plus white noise $v_2[n]$ (σ_2^2, real valued)

Want to design a 2-tap Wiener filter to get the best estimate of $d[n]$.

We have $\sigma_1^2 = 0.27$, $\sigma_2^2 = 0.1$, $v_2 \perp v_1$, $v_2 \perp x$ (use "$\perp$" to represent 2 uncorrelated processes)

- What kind of process is $\{x[n]\}$?
- What is the correlation matrix of the channel output?
- What is the cross-correlation vector?
- $w_1 = ?$ $\quad w_2 = ?$ $\quad J_{\min} = ?$

Another Perspective (in terms of the gradient)

Theorem: If $f(\underline{z}, \underline{z}^*)$ is a **real-valued** function of complex vectors $\underline{z}$ and $\underline{z}^*$, then the vector pointing in the direction of the maximum rate of the change of f is $\nabla_{\underline{z}^*} f(\underline{z}, \underline{z}^*)$, which is a vector of the derivative of $f()$ w.r.t. each entry in the vector $\underline{z}^*$.

Corollary: Stationary points of $f(\underline{z}, \underline{z}^*)$ are the solutions to $\nabla_{\underline{z}^*} f(\underline{z}, \underline{z}^*) = 0$.

Complex gradient of a complex function:

	$\underline{a}^H \underline{z}$	$\underline{z}^H \underline{a}$	$\underline{z}^H A \underline{z}$
$\nabla_{\underline{z}}$	$\underline{a}^*$	0	$A^T \underline{z}^* = (A\underline{z})^*$
$\nabla_{\underline{z}^*}$	0	$\underline{a}$	$A\underline{z}$

Using the above table, we have $\nabla_{\underline{a}^*} J = -\underline{p}^* + \mathbf{R}^T \underline{a}$.

For optimal solution: $\nabla_{\underline{a}^*} J = \frac{\partial}{\partial \underline{a}^*} J = 0$

$\Rightarrow \mathbf{R}^T \underline{a} = \underline{p}^*$, or $\mathbf{R}\underline{a}^* = \underline{p}$, the Normal Equation. $\therefore \underline{a}^*_{\text{opt}} = \mathbf{R}^{-1} \underline{p}$

[Review on matrix & optimization: Hayes 2.3; Haykin (4th) Appendices A,B,C]

Review: differentiating complex functions and vectors

<1> Differentiable at z_0

$\lim_{\Delta z \to 0} \frac{f(z_0+\Delta z)-f(z_0)}{\Delta z}$ exist $\Rightarrow$ Need to converge in all directions for $\Delta z \to 0$

Recall: $f(z)$ is analytic (i.e. differentialable everywhere) on region D if $f(z) = u(x,y) + i\,v(x,y)$ is continuous and satisfy Cauchy-Riemann condition $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ & $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$.

<2> e.g. $f_1(z) = z z^* = |z|^2 = (x^2+y^2) + i \times 0$

$f_2(z) = z^* = x - iy$

$\Rightarrow$ DOES NOT satisfy Cauchy-Riemann.

Review: differentiating complex functions and vectors

unlike the real value optimiz. case, $\frac{df(x)}{dx}=0$.

Note: $f(z) = |z|^2$ has unique minimum at $z=0$, but not differentiable from complex analysis (any func. that depends on z^* is not differentiable)

We can either minimize $f(x,y)$ wrt x & y where $z = x+iy$, or treat z and z^* as indep. variables and minimize $f(z, z^*)$ w.r.t. both z and z^*: i.e. $\frac{\partial f}{\partial z} = 0$ and $\frac{\partial f}{\partial z^*} = 0$

Minimizing a real-valued func. of z and z^* (and the func. is analytic w.r.t. both z and z^*) is somewhat easier: the optimal points may be found by setting the derivative of $f(z, z^*)$ w.r.t. either z or z^* equal to zero and solve for z.

e.g. $f(z, z^*) = |z|^2 = z \cdot z^*$. sufficient to have $\frac{\partial f}{\partial z^*} = z = 0$.

Differentiating complex functions: More details

$z = x + iy$ $\qquad f(z) = u(x,y) + i\,v(x,y)$

Note: $x = \frac{1}{2}(z + z^*)$

$y = \frac{1}{2i}(z - z^*)$

$$\Rightarrow \begin{cases} \frac{\partial f}{\partial x} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} \\ \frac{\partial f}{\partial y} = \frac{\partial u}{\partial y} + i\frac{\partial v}{\partial y} \end{cases}$$

$$\begin{cases} \frac{\partial f}{\partial z} \overset{\text{def}}{=} \frac{1}{2}\left[\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}(-i)\right] \\ \frac{\partial f}{\partial z^*} \overset{\text{def}}{=} \frac{1}{2}\left[\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\cdot i\right] \end{cases}$$

i.e. $\frac{\partial f}{\partial z} = \frac{\partial f}{\partial x}\cdot\frac{\partial x}{\partial z} + \frac{\partial f}{\partial y}\cdot\frac{\partial y}{\partial z}$

For real-valued $f(z)$: i.e. $f(z) = u(x,y)$,

we have: ① $\frac{\partial f}{\partial z} = \left(\frac{\partial f}{\partial z^*}\right)^*$; ② Gradient $\nabla u = \begin{bmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial u}{\partial y} \end{bmatrix} \Rightarrow$ written as complex# $\frac{\partial u}{\partial x} + i\frac{\partial u}{\partial y} = \frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}$

E.g. ① if $f(z) = z = x + iy$

$\frac{\partial f}{\partial z^*} \overset{\text{def}}{=} \frac{1}{2}\left[\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right] = \frac{1}{2}[1 + i\cdot i] = 0$; $\frac{\partial f}{\partial z} \overset{\text{def}}{=} \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right) = \frac{1}{2}(1 - i\cdot i) = 1$

E.g. ② $f(z) = |z|^2$.

Let $A \overset{\text{def}}{=} \frac{f(z+\Delta z) - f(z)}{\Delta z} = \frac{(z+\Delta z)(z^* + (\Delta z)^*) - z\cdot z^*}{\Delta z} = z^* + (\Delta z)^* + z\frac{(\Delta z)^*}{\Delta z}$

for $\Delta z = \Delta x + 0\cdot i$: $\lim_{\Delta x \to 0} \frac{(\Delta z)^*}{\Delta z} = 1 \Rightarrow A \to z^* + z$

for $\Delta z = 0 + \Delta y\cdot i$: $\lim_{\Delta y \to 0} \frac{(\Delta z)^*}{\Delta z} = -1 \Rightarrow A \to z^* - z$

$\Rightarrow$ A converges to different results for different directions as $\Delta z \to 0$ except for $z = 0$

$\therefore$ the limit doesn't exist, except for $z = 0$ (and thus not differentiable)

Detailed Derivations

Example: solution

① What is $\{x[n]\}$?

$$d[n] = -0.8458\, d[n-1] + v_1[n] \Rightarrow H_1(z) = \frac{1}{1+0.8458 z^{-1}}$$

$$x[n] = 0.9458\, x[n-1] + d[n] \Rightarrow H_2(z) = \frac{1}{1-0.9458 z^{-1}}$$

$$H(z) = H_1(z) H_2(z) = \frac{1}{1+0.8458 z^{-1}} \cdot \frac{1}{1-0.9458 z^{-1}}$$

$$= \frac{1}{1-0.1 z^{-1} - 0.8 z^{-2}} = \frac{1}{1+a_1 z^{-1} + a_2 z^{-2}}$$

i.e. $\{x[n]\}$ is an AR(2) process of

$$x[n] - 0.1\, x[n-1] - 0.8\, x[n-2] = v_1[n]$$

Example: solution

② The channel output is $u[n] = x[n] + v_2[n]$

$$\underset{2\times 2}{R_u} = E\left[\underline{u}[n]\underline{u}^H[n]\right] = \begin{bmatrix} r_u(0) & r_u(1) \\ r_u(1)^* & r_u(0) \end{bmatrix} = R_x + R_{v_2}$$

$$= \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} + \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} = \begin{bmatrix} 1.1 & 0.5 \\ 0.5 & 1.1 \end{bmatrix}$$

where $r_x(\cdot)$ can be obtained from the AR parameter equation as seen from the example at the end of §2.1:

$$r_x(0) = \left(\frac{1+a_2}{1-a_2}\right)\frac{\sigma_1^2}{(1+a_2)^2 - a_1^2} = 1, \qquad r_x(1) = \frac{-a_1}{1+a_2} r_x(0) = 0.5$$

Example: solution

③ Obtain the cross correlation vector $\underline{P} = E\left[d^*[n]\begin{pmatrix} u[n] \\ u[n-1] \end{pmatrix}\right]$

$$E[d[n]u[n]] = E[(x[n] - 0.9458x[n-1])(x[n] + v_2[n])]$$
$$= r_x(0) - 0.9458\, r_x(-1) = 1 - 0.9458 \times 0.5$$
$$= 1 - 0.4729 = 0.5271$$

Similarly,
$$E[d[n]u[n-1]] = r_x(1) - 0.9458\, r_x(0) = -0.4458$$

$$\therefore \underline{P} = \begin{bmatrix} 0.5271 \\ -0.4458 \end{bmatrix}$$

④ optimal weights are
$$\underline{W}_o = \underline{R}^{-1}\underline{P} = \begin{bmatrix} 0.8360 \\ -0.7853 \end{bmatrix}$$

$$J(w_1, w_2) = 0.9486 - 1.0544w_1 + 0.8916w_2 + w_1w_2 + 1.1(w_1^2 + w_2^2)$$
$$\Rightarrow J_{min} = 0.1579$$