Statistical Signal Processing

Lecture 8

chapter 2: Spectrum Estimation

- AR modeling motivations: LP of an AR(N) process, asymptotics
- AR modeling interpretations, techniques, model order selection

chapter 3: Optimal Filtering

Wiener filtering

- non-causal Wiener filtering
- signal in noise case
- equalization
- causal Wiener filtering

Noncausal Wiener Filtering

- parameter estimation: estimation of a finite number of random/deterministic parameters given (stochastically) related measurements
- spectrum estimation: estimation of an infinite (nonparametric)/finite (parametric) number of deterministic parameters
- here: estimation of an unknown *random process* $\{x_k,\ k \in \mathcal{Z}\}$ given a (cor-)related random process $\{y_k,\ k \in \mathcal{Z}\}$
- can be reduced to the estimation of a random parameter x_k:

$$\{y_n,\ n \in \mathcal{Z}\} \quad \rightarrow \quad \widehat{x}_k$$

- linear estimator = filter:

$$y_n \longrightarrow \boxed{h_{k,n}} \longrightarrow \widehat{x}_k = \sum_{n=-\infty}^{\infty} h_{k,n} y_n$$

Noncausal Wiener Filtering (2)

- determine the filter coefficients $h_{k,n}$ from the following LMMSE estimation problem

$$\min_{h_{k,n}} E(x_k - \widehat{x}_k)^2 \quad = \quad \min_{h_{k,n}} E(x_k - \sum_{n=-\infty}^{\infty} h_{k,n} y_n)^2$$

- the optimal $h_{k,n}$ satisfy the LMMSE orthogonality conditions

$$E\,(x_k - \widehat{x}_k) y_m = 0\,, \quad \forall m \in \mathcal{Z} \;\Rightarrow\; E\,\widehat{x}_k y_m = \sum_{n=-\infty}^{\infty} h_{k,n}\, E\, y_n y_m = E\, x_k y_m, \;\forall m \in \mathcal{Z}$$

- assume $\{x_k,\ k \in \mathcal{Z}\}$ and $\{y_k,\ k \in \mathcal{Z}\}$ to be jointly (wide sense) stationary (and zero mean): orthogonality conditions $\rightarrow$ normal equations

$$\sum_{n=-\infty}^{\infty} h_{k,n}\, r_{yy}(n-m) \;=\; r_{xy}(k-m)\;, \quad \forall m \in \mathcal{Z}$$

- substitute $k-m \;\rightarrow\; m$ and $n-k \;\rightarrow\; -n$

$$\sum_{n=-\infty}^{\infty} h_{k,n}\, r_{yy}(n-k+m) \;=\; r_{xy}(m) \;\;, \quad \sum_{n=-\infty}^{\infty} h_{k,k-n}\, r_{yy}(m-n) \;=\; r_{xy}(m)\;, \quad \forall m \in \mathcal{Z}$$

- solution for the $h_{k,k-n}$ is the same for any $k \;\Rightarrow\; h_{k,k-n} = h_{0,-n} = h_n$: stationarity $\Rightarrow$ the optimal linear filter is time-invariant

Noncausal Wiener Filtering (3)

- the optimal linear time-invariant filter satisfies

$$\sum_{n=-\infty}^{\infty} h_n\, r_{yy}(m-n) \;=\; r_{xy}(m)\;, \quad \forall m \in \mathcal{Z}$$

  This is an infinite set of equations in an infinite number of unknowns h_n.

- convolution $\Rightarrow$ take z-transform to obtain a simple product. Let

$$\mathrm{S}_{xy}(z) \;=\; \sum_{m=-\infty}^{\infty} r_{xy}(m) z^{-m} \quad , \quad \mathrm{H}(z) \;=\; \sum_{m=-\infty}^{\infty} h_m z^{-m}$$

  then

$$\mathrm{H}(z)\mathrm{S}_{yy}(z) = \mathrm{S}_{xy}(z) \quad \Rightarrow \quad \mathrm{H}(z) \;=\; \frac{\mathrm{S}_{xy}(z)}{\mathrm{S}_{yy}(z)}$$

Frequency Domain Interpretation

- the Fourier transform of $\widehat{x}_k = \Sigma_{m=-\infty}^{\infty} h_m y_{k-m}$ is
  (the Fourier transform is the z-transform evaluated at $z = e^{j2\pi f}$)

$$\widehat{X}(f) = H(f)Y(f) \ , \quad H(f) = \frac{S_{xy}(f)}{S_{yy}(f)}$$

  where $Y(f) = \mathrm{Y}(e^{j2\pi f}) = \sum_{m=-\infty}^{\infty} y_m e^{-j2\pi f m}$ etc.

- this resembles the scalar LMMSE problem:

$$\widehat{x} = h\,y \ , \ h = \frac{R_{xy}}{R_{yy}} \ , \quad \widehat{X}(f) = H(f)\,Y(f)$$

- can show:

$$H(f) = \frac{R_{X(f)Y(f)}}{R_{Y(f)Y(f)}} = \frac{S_{xy}(f)}{S_{yy}(f)}$$

- Wiener filtering of one random process from another is like a scalar LMMSE estimation of the Fourier transforms of those processes at every frequency.

MMSE Expressions

- The orthogonality property of the LMMSE estimator implies

$$E\,(x_k - \widehat{x}_k)\widehat{x}_k = \sum_{n=-\infty}^{\infty} h_n \underbrace{E\,(x_k - \widehat{x}_k)y_{k-n}}_{=\,0} = 0 \ \Rightarrow \ E\,x_k\widehat{x}_k = E\,\widehat{x}_k^2$$

  This allows us to demonstrate the following Pythagorean property

$$\begin{aligned}\text{MMSE} &= E\,\widetilde{x}_k^2 = E\,(x_k - \widehat{x}_k)^2 = E\,(x_k - \widehat{x}_k)x_k - \underbrace{E\,(x_k - \widehat{x}_k)\widehat{x}_k}_{=\,0} \\ &= E\,x_k^2 - E\,x_k\widehat{x}_k = E\,x_k^2 - \underbrace{E\,\widehat{x}_k^2}_{\geq\,0} \leq E\,x_k^2\end{aligned}$$

  $Ex_k^2 = r_{xx}(0)$ is the MSE if we have no observations. In that case, $\widehat{x}_k = Ex_k = 0$ is our best estimator, leading indeed to Ex_k^2 as MSE. $E\widehat{x}_k^2 = r_{\widehat{x}\widehat{x}}(0) \geq 0$ is the reduction in MSE by estimating x_k using the Wiener filter on the data $\{y_n,\ n \in \mathcal{Z}\}$.

- analyzing the MMSE in the frequency domain:

$$\text{MMSE} = r_{xx}(0) - r_{\widehat{x}\widehat{x}}(0) = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{xx}(f)df - \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{\widehat{x}\widehat{x}}(f)df$$

- Since $\widehat{x}_k = h_k * y_k$, we have for the power spectral density functions

$$S_{\widehat{x}\widehat{x}}(f) = |H(f)|^2 S_{yy}(f) = |S_{xy}(f)|^2 / S_{yy}(f)$$

MMSE Expressions in the Frequency Domain

- Let us introduce the *cross-power spectral density coefficient*

$$\rho_{xy}(f) \;=\; \frac{S_{xy}(f)}{\sqrt{S_{xx}(f)S_{yy}(f)}}$$

which is defined as zero whenever $S_{xx}(f) = 0$ or $S_{yy}(f) = 0$.

- $\Rightarrow$ $\text{MMSE} \;=\; r_{\tilde{x}\tilde{x}}(0) \;=\; \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{\tilde{x}\tilde{x}}(f)df \;=\; \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{xx}(f)\left[1-|\rho_{xy}(f)|^2\right]df$ (*)

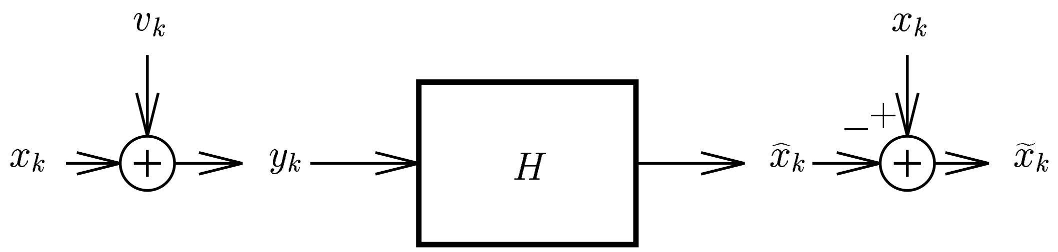
In particular we find that

$$1-|\rho_{xy}(f)|^2 \;=\; \frac{S_{\tilde{x}\tilde{x}}(f)}{S_{xx}(f)} \geq 0 \quad\Rightarrow\quad |\rho_{xy}(f)| \leq 1$$

$\rho_{xy}(f)$ is normalized and can be interpreted as the correlation coefficient between $X(f)$ and $Y(f)$.

- Since now $1-|\rho_{xy}(f)|^2 \in [0,1]$, (*) shows how the power spectral density of x_k gets attenuated as a function of frequency to obtain the power spectral density of the estimation error $\tilde{x}_k$. At frequencies where $X(f)$ and $Y(f)$ are strongly correlated, $S_{\tilde{x}\tilde{x}}(f)$ will be significantly reduced w.r.t. $S_{xx}(f)$ whereas this will not be the case at frequencies where $X(f)$ and $Y(f)$ are hardly correlated.

Signal in Noise



- case of $y_k = x_k + v_k$

- The quantities that determine the Wiener filter are

$$\begin{aligned}
r_{xy}(n) &= r_{xx}(n) + \underbrace{r_{xv}(n)}_{=0} = r_{xx}(n) &&, \quad \mathrm{S}_{xy}(z) = \mathrm{S}_{xx}(z)\\
r_{yy}(n) &= r_{xx}(n) + \underbrace{r_{xv}(n)}_{=0} + \underbrace{r_{vx}(n)}_{=0} + r_{vv}(n) &&\\
&= r_{xx}(n) + r_{vv}(n) &&, \quad \mathrm{S}_{yy}(z) = \mathrm{S}_{xx}(z) + \mathrm{S}_{vv}(z)
\end{aligned}$$

- Wiener filter depends on SNR as a function of frequency

$$H(f) \;=\; \frac{S_{xx}(f)}{S_{xx}(f)+S_{vv}(f)} \;\in\; [0,1] \quad \text{weighting filter}$$

Signal in Noise (2)

- can show

$$\frac{1}{S_{\tilde{x}\tilde{x}}(f)} = \frac{1}{S_{xx}(f)} + \frac{1}{S_{vv}(f)} \geq \max\{\frac{1}{S_{xx}(f)}, \frac{1}{S_{vv}(f)}\}$$

- and for the MMSE

$$\text{MMSE} = E\,\tilde{x}_k^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{\tilde{x}\tilde{x}}(f)\,df = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{S_{xx}(f)\,S_{vv}(f)}{S_{xx}(f)+S_{vv}(f)} df$$

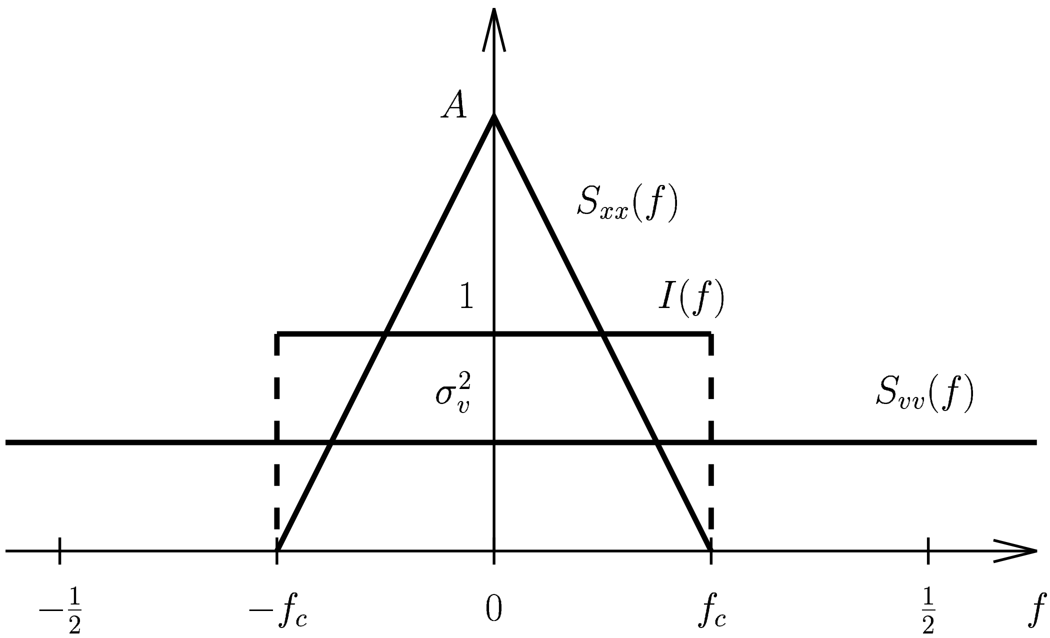
$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{xx}(f)\ \underbrace{\frac{S_{vv}(f)}{S_{xx}(f)+S_{vv}(f)}}_{0\leq\,.\,\leq 1} df = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{vv}(f)\ \underbrace{\frac{S_{xx}(f)}{S_{xx}(f)+S_{vv}(f)}}_{0\leq\,.\,\leq 1} df$$

$$\leq \min\left\{\int_{-\frac{1}{2}}^{\frac{1}{2}} S_{xx}(f)\,df\ ,\ \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{vv}(f)\,df\right\} = \min\{\underbrace{E\,x_k^2}_{\hat{x}_k=0},\ \underbrace{E\,v_k^2}_{\hat{x}_k=y_k}\}$$

and also

$$\text{MMSE} \leq \int_{-\frac{1}{2}}^{\frac{1}{2}} \min\{S_{xx}(f)\,,\,S_{vv}(f)\}\ df \leq \min\{E\,x_k^2,\ E\,v_k^2\}$$

Signal in Noise: Example



- example: Bandlimited Random Process in White Noise
- x_k with a triangular power spectral density function

$$S_{xx}(f) = \begin{cases} A(1-\frac{|f|}{f_c}) & ,\ |f| \leq f_c \\ 0 & ,\ f_c \leq |f| \leq \frac{1}{2} \end{cases}$$

Signal in Noise: Example (2)

- We receive $y_k = x_k + v_k$. The additive noise v_k is white with variance σ_v^2: $S_{vv}(f) = \sigma_v^2$.
- The task is to filter y_k so that the filter output $\widehat{x}_k$ approximates x_k well.
- Classical (non-statistical) filter design is based on the notion of distortion: pass x_k without distortion but for the rest cut out the noise as much as possible. Since x_k is bandlimited, we can choose an ideal low-pass filter $I(f)$ matched to the bandwidth f_c of x_k:

$$I(f) = \begin{cases} 1 \ , \ |f| \leq f_c \\ 0 \ , \ f_c < |f| \leq \frac{1}{2} \end{cases}$$

The output $\widehat{x}_k(I)$ of the filter $I(f)$ will be equal to x_k minus an error $\widetilde{x}_k(I)$ which is a low-pass filtered version of v_k. Hence the variance of the error is

$$E\,\widetilde{x}_k^2(I) = \int_{-f_c}^{f_c} S_{vv}(f)df = 2\sigma_v^2 f_c$$

Since $f_c \leq 0.5$, $E\,\widetilde{x}_k^2(I) \leq E\,v_k^2$. Hence, the filtering operation with $I(f)$ has reduced the noise level w.r.t. the measurement y_k while leaving the signal component x_k undistorted.

Signal in Noise: Example (3)

- Wiener approach: trades some signal distortion for a further reduction in overall error variance:

$$H(f) = \frac{S_{xx}(f)}{S_{xx}(f) + S_{vv}(f)} = \begin{cases} \dfrac{A}{\sigma_v^2} \dfrac{1 - \frac{|f|}{f_c}}{1 + \frac{A}{\sigma_v^2}(1 - \frac{|f|}{f_c})} & , \ |f| \leq f_c \\ 0 & , \ f_c < |f| \leq \frac{1}{2} \end{cases}$$

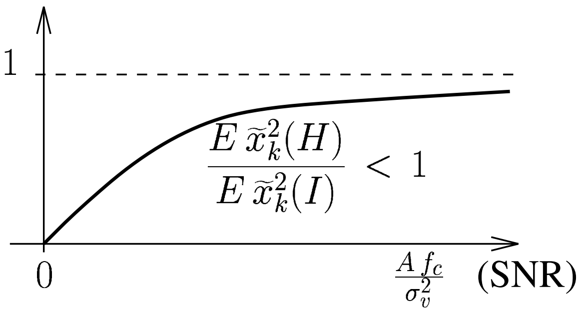
We can analyze the nature of the optimal filter at high or low signal-to-noise ratio (SNR):

$$\text{low SNR: } \frac{A}{\sigma_v^2} \to 0 \ : \ H(f) \to \frac{S_{xx}(f)}{S_{vv}(f)} = \begin{cases} \dfrac{A}{\sigma_v^2}(1 - \dfrac{|f|}{f_c}) & , \ |f| \leq f_c \\ 0 & , \ f_c < |f| \leq \frac{1}{2} \end{cases}$$

$$\text{high SNR: } \frac{A}{\sigma_v^2} \to \infty \ : \ H(f) \to I(f)$$

So for low SNR, the filter $H(f)$ becomes proportional to the ratio of the psdf's of the signal of interest and the noise. For high SNR, the filter $H(f)$ approaches the classical distortion-based design, passing perfectly all frequencies where the signal of interest is present.

Signal in Noise: Example (4)



- We find for the MMSE:

$$\text{MMSE} \;=\; E\,\tilde{x}_k^2(H) \;=\; [1-\frac{\sigma_v^2}{A}\ln(1+\frac{A}{\sigma_v^2})]\,E\,\tilde{x}_k^2(I)$$

We can again analyze the limiting behavior for high or low SNR:

$$\text{low SNR: } \frac{A}{\sigma_v^2}\to 0 \;:\; E\,\tilde{x}_k^2(H)\to E\,x_k^2 = A\,f_c$$
$$\text{high SNR: } \frac{A}{\sigma_v^2}\to \infty \;:\; E\,\tilde{x}_k^2(H)\to\; E\,\tilde{x}_k^2(I) = 2f_c\sigma_v^2$$

- high SNR: optimal filter // classical distortion criterion based design: the variance of the error is the variance of the noise in the signal band.
- low SNR: optimal filter works much better than the classical one. Indeed, the variance of the error becomes equal to the signal variance, even though the noise level is much higher! It is true though that at low SNR, the performance of even the optimal filter is not very good in this example.

Channel Equalization

- digital communications: $y_k \;=\; \mathrm{C}(q)\,x_k + v_k$.
  x_k: symbols, y_k: received signal, v_k: additive noise
  $\mathrm{C}(z) = \Sigma\, c_k z^{-k}$: channel (cascade of transmission (pulse shaping) filter, actual channel and receiver filter)
  sampling at the symbol rate

- assume: x_k and v_k independent white stationary sequences with zero mean and variances σ_x^2, σ_v^2. We also assume here that all signals involved are real and scalar.

- The x_k have a discrete distribution and take on values in a finite alphabet. The problem of deciding on the basis of the y_k which discrete values x_k have been sent is called the *detection* problem.

- If the channel impulse response c_k has only one non-zero sample, then the optimal detection can be done instantaneously since the noise samples v_k are independent. This means that if w.l.o.g. we consider c_0 to be the non-zero sample, x_k can be detected from the sample

$$y_k = c_0\,x_k + v_k\ .$$

Channel Equalization

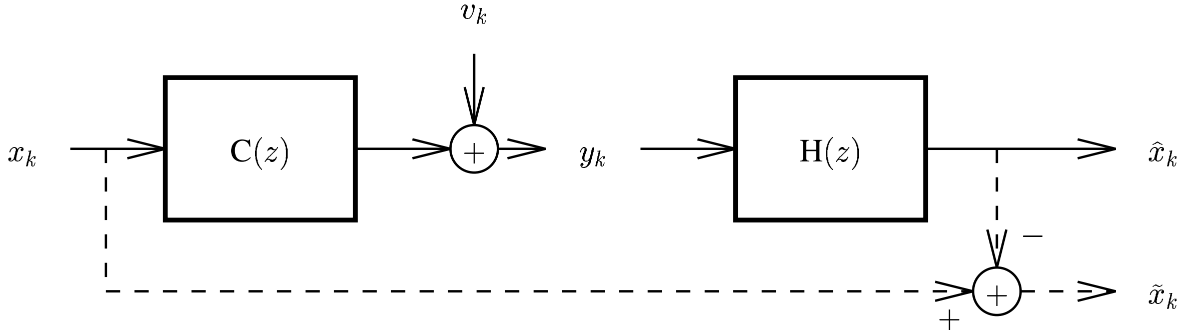
- If c_k contains more than one non-zero sample, then the symbols appear superimposed in the received signal y_k: *intersymbol interference* (ISI). The problem of detecting the symbols in the presence of ISI is called *equalization*. Optimal (maximum-likelihood) equalization can be done for FIR channels using the so-called Viterbi algorithm (which corresponds to dynamic programming).

- performance of Viterbi: bound by the Matched Filter Bound (MFB). The MFB expresses the maximum SNR achievable for the detection of a certain symbol x_k assuming that all other symbols have been correctly detected. This means that all other symbols are known so that their contribution can be subtracted from the received signal y_k. If we assume w.l.o.g. that the symbol we want to detect is x_0, then the resulting received signal with the contributions of the $x_k, k \neq 0$, removed is
$$y_k \;=\; c_k\, x_0 + v_k \quad \longrightarrow \quad Y = C\, x_0 + V \;\; .$$
If the white noise v_k is Gaussian, then it turns out that the optimal detection of x_0 can be done on the basis of the unconstrained ML estimate
$$\widehat{x}_0^{ML} = \frac{\Sigma_k\, c_k\, y_k}{\Sigma_k\, c_k^2} \;=\; (C^TC)^{-1}C^TY \;\; . \quad \text{MFB} \;=\frac{\sigma_x^2}{\sigma_{\tilde{x}}^2} = \frac{\sigma_x^2}{\sigma_v^2/||C||^2} = \frac{\sigma_x^2}{\sigma_v^2}\,||C||^2$$

Channel Equalization

- other interpretation: Consider filtering the y_k with a filter $\mathrm{H}(z)$ to obtain the signal $\widehat{x}_k$ and consider in particular the output at time 0, $\widehat{x}_0$. The part of $\widehat{x}_0$ due to x_0 is called the signal part while the part due to the v_k is called the noise part.

- the filter $\mathrm{H}(z)$ that maximizes the SNR in $\widehat{x}_0$ is the *matched filter* $\mathrm{H}(z) = \mathrm{C}^\dagger(z) = \mathrm{C}(1/z)$ $(= \mathrm{C}^H(1/z^*)$ in general), matched to the channel $\mathrm{C}(z)$. The SNR at the output of the matched filter is the MFB
$$\text{MFB} \;= \frac{\sigma_x^2}{\sigma_v^2}\frac{1}{2\pi j}\oint \frac{dz}{z}\mathrm{C}^\dagger(z)\mathrm{C}(z) = \frac{\sigma_x^2}{\sigma_v^2}\int_{-\frac{1}{2}}^{\frac{1}{2}} df\; C^*(f)C(f) = \frac{\sigma_x^2}{\sigma_v^2}\sum_k |c_k|^2 = \frac{\sigma_x^2}{\sigma_v^2}||C||^2$$
The MFB is proportional to the energy in the channel response. (The term MFB is also often used to denote the corresponding probability of error in the detection of $\widehat{x}_0^{ML}$).



Simpler Equalizers

- MFB optimization problem: $\widehat{x}_0 = x_0 \frac{1}{2\pi j} \oint \frac{dz}{z} \mathrm{H}(z)\mathrm{C}(z) \; + \mathrm{H}(q)\, v_k|_{k=0}$

  $$\mathrm{SNR} = \frac{\sigma_x^2}{\sigma_v^2} \frac{\left|\frac{1}{2\pi j} \oint \frac{dz}{z} \mathrm{H}(z)\mathrm{C}(z)\right|^2}{\frac{1}{2\pi j} \oint \frac{dz}{z} \mathrm{H}(z)\mathrm{H}^\dagger(z)} = \frac{\sigma_x^2}{\sigma_v^2} \frac{\left|\langle \mathrm{H}^\dagger(.), \mathrm{C}(.)\rangle\right|^2}{\left\|\mathrm{H}^\dagger(.)\right\|^2} \leq \frac{\sigma_x^2}{\sigma_v^2} \left\|\mathrm{C}(.)\right\|^2$$

  Cauchy-Schwartz inequality, equality (max) reached for $\mathrm{H}^\dagger(z) = \mathrm{C}(z)$.

- The Viterbi equalizer can be fairly complex however. A class of simple sub-optimal equalizers can be obtained as the cascade of a simple linear estimator followed by an instantaneous detector (decision element).

- The class of so-called *linear equalizers* (LEs) performs linear estimation on the basis of the signal y_k alone.

- The more sophisticated class of *decision-feedback equalizers* (DFEs) performs linear estimation on the basis of all y_k plus also the previously detected x_k (hence, feedback of previous decisions).

Zero-Forcing Linear Equalizers

- Linear equalizers are simply linear filters $\mathrm{H}(z)$ filtering the signal y_k and their output $\widehat{x}_k = \mathrm{H}(q) y_k$ gets processed by a decision element.

- classical point of view of filtering a signal in noise: zero distortion for the signal part. The signal part in $\widehat{x}_k$ is $\mathrm{H}(q)\mathrm{C}(q)x_k$. To obtain zero distortion (signal part of $\widehat{x}_k$ equal to x_k) would mean that

  $$\mathrm{H}(z)\mathrm{C}(z) = 1 \quad \Rightarrow \quad \mathrm{H}_{ZF}(z) = \frac{1}{\mathrm{C}(z)} = \frac{1}{\mathrm{C}_{min}(z)} \frac{1}{\mathrm{C}_{max}(z)}$$

  Since this equalizer forces the resulting ISI to zero, this solution is called the *zero-forcing* (ZF) equalizer.

- $\mathrm{C}_{min}(z)$ and $\mathrm{C}_{max}(z)$ are the minimum-phase and maximum-phase factors of $\mathrm{C}(z)$ (in general not minimum-phase nor maximum-phase) The inverses of $\mathrm{C}_{min}(z)$ and $\mathrm{C}_{max}(z)$ are causal and anti-causal resp. This means that in general $\mathrm{H}_{ZF}(z)$ will have an impulse response that extends from $-\infty$ to $+\infty$ (even if $\mathrm{C}(z)$ is FIR).

Zero-Forcing Linear Equalizers (2)

- In practice, $\mathrm{H}(z)$ will normally be approximated with an FIR filter. This FIR approximation will have to be non-causal in order to get a good approximation. Such finite non-causality can be dealt with by introducing a corresponding delay. The FIR approximation will have to be longer as the zeros of $\mathrm{C}(z)$ approach the unit circle. The ZF equalizer does not exist if zeros of $\mathrm{C}(z)$ on the unit circle.

- error signal $\widetilde{x}_k = x_k - \widehat{x}_k = \mathrm{H}(q) v_k = \frac{1}{\mathrm{C}(q)} v_k$ only contains noise, due to ZF

$$MSE_{ZF-LE} = E\,\widetilde{x}_k^2 = \frac{\sigma_v^2}{2\pi j}\oint \frac{dz}{z}\frac{1}{\mathrm{C}^\dagger(z)\mathrm{C}(z)}$$

The MSE can get very big as some zeros of $\mathrm{C}(z)$ approach the unit circle. This phenomenon is called *noise enhancement*. The perfect cancellation of the ISI is obtained at the cost of enhancing the noise. The SNR of the linear equalizer is defined as

$$SNR_{ZF-LE} = \frac{\sigma_x^2}{MSE_{ZF-LE}} \quad \text{or more precisely} \quad SINR = \frac{E\,(E\,\widehat{x}_k|_{x_k})^2}{E\,(\widehat{x}_k - E\,\widehat{x}_k|_{x_k})^2}$$

Using the Cauchy-Schwarz inequality, one can show that

$$SNR_{ZF-LE} \leq MFB\ .$$

MMSE Linear Equalizers

- The optimal filtering point of view simply takes the MSE as optimality criterion. Hence we get the Wiener filter

$$\mathrm{H}_{MMSE-LE}(z) = \frac{\mathrm{S}_{xy}(z)}{\mathrm{S}_{yy}(z)} = \frac{\sigma_x^2\mathrm{C}^\dagger(z)}{\sigma_x^2\mathrm{C}^\dagger(z)\mathrm{C}(z)+\sigma_v^2} \quad \left(\xrightarrow{\sigma_v^2\to 0} \quad \mathrm{H}_{ZF-LE}(z) = \frac{1}{\mathrm{C}(z)}\right)$$

Remark: the factor $\mathrm{C}^\dagger(z)$ in the numerator represents the matched filter.

$$MSE_{MMSE-LE} = E\,(x_k - \widehat{x}_k)x_k = \sigma_x^2 - \frac{1}{2\pi j}\oint\frac{dz}{z}\mathrm{S}_{\widehat{x}x}(z) = \sigma_x^2 - \frac{1}{2\pi j}\oint\frac{dz}{z}\mathrm{H}(z)\mathrm{S}_{yx}(z)$$

$$= \sigma_x^2\left(1 - \frac{\sigma_x^2}{2\pi j}\oint\frac{dz}{z}\mathrm{C}^\dagger(z)\mathrm{S}_{yy}^{-1}(z)\mathrm{C}(z)\right) = \frac{\sigma_x^2\sigma_v^2}{2\pi j}\oint\frac{dz}{z}\mathrm{S}_{yy}^{-1}(z) = \frac{\sigma_v^2}{2\pi j}\oint\frac{dz}{z}\frac{1}{\mathrm{C}^\dagger(z)\mathrm{C}(z)+\frac{\sigma_v^2}{\sigma_x^2}}\ .$$

From the previous expression, it is clear that

$$MSE_{MMSE-LE} \leq \min\left\{\sigma_x^2, MSE_{ZF-LE}\right\} \Rightarrow SNR_{MMSE-LE} \geq \max\left\{1, SNR_{ZF-LE}\right\}.$$

This means that the MMSE-LE always does at least as well as the ZF-LE, at least in terms of MSE. Using the Cauchy-Schwarz inequality, one can show that

$$SNR_{MMSE-LE} \leq MFB + 1\ .$$

Remark that the MMSE equalizer converges to the ZF equalizer as $\sigma_v^2 \to 0$

Unbiased MMSE Linear Equalizers

- remark: $SNR_{MMSE-LE}$ can be larger than the MFB. This is due to the fact that the MMSE LE gives a *biased* estimate of x_k: the coefficient of x_k appearing in $\widehat{x}_k$ is not equal to 1.
- Although the unconstrained MMSE LE gives the lowest MSE, the bias in $\widehat{x}_k$ will increase the *probability of error* in the decision process. The decision element expects indeed to see x_k plus some random deviations at its input, whereas a bias (in the form of $\alpha\, x_k$) is not a random perturbation w.r.t. x_k. Random perturbations are perturbations due to the $x_i, i \neq k$, and the v_i. The unbiased MMSE (UMMSE) LE minimizes the MSE subject to the constraint that the estimator be unbiased, i.e. $E\left[\widehat{x}_k | x_k\right] = x_k$, or $\frac{1}{2\pi j}\oint \frac{dz}{z}\mathrm{H}(z)\mathrm{C}(z) = \Sigma_k\, h_k c_{-k} = 1$.
- The MMSE equalizer takes the Bayesian viewpoint in which all the x_k and the v_k are considered random. The UMMSE equalizer takes the more deterministic viewpoint in which the symbol to be estimated x_k is considered deterministic, but the other $x_i, i \neq k$, and the v_i are still random (this the viewpoint of the BLUE estimator, considered here for an infinite number of measurements y_k).

Unbiased MMSE Linear Equalizers (2)

- The UMMSE equalizer design problem is hence

$$\min_{h_i:\ \sum_i h_i c_{-i} = 1} E\left(x_k - \sum_i h_i y_{k-i}\right)^2 .$$

We can turn this constrained optimization problem into an unconstrained optimization problem by introducing a Lagrange multiplier λ:

$$\min_{h_i,\lambda} f(\mathrm{H}(.),\lambda) \;=\; \min_{h_i,\lambda}\left\{E\left(x_k - \sum_i h_i y_{k-i}\right)^2 + \lambda\left(\sum_i h_i c_{-i} - 1\right)\right\} .$$

By setting derivatives equal to zero, we find

$$\begin{aligned}\frac{\partial f}{\partial h_i} &= 2\,E\left(x_k - \sum_n h_n y_{k-n}\right)(-y_{k-i}) + \lambda c_{-i} = 0\\ \frac{\partial f}{\partial \lambda} &= \sum_i h_i c_{-i} - 1 = 0\end{aligned}$$

The first equation leads to

$$\begin{aligned}&-r_{xy}(i) + \sum_n h_n r_{yy}(i-n) + \frac{\lambda}{2}c_{-i} = 0\\ &\Rightarrow\ -\mathrm{S}_{xy}(z) + \mathrm{H}(z)\mathrm{S}_{yy}(z) + \frac{\lambda}{2}\mathrm{C}^{\dagger}(z) = 0 \ \Rightarrow\ \mathrm{H}(z) = (\sigma_x^2 - \frac{\lambda}{2})\mathrm{C}^{\dagger}(z)\mathrm{S}_{yy}^{-1}(z) .\end{aligned}$$

Unbiased MMSE Linear Equalizers (3)

- From the constraint, we find

$$\frac{1}{2\pi j}\oint\frac{dz}{z}\mathrm{H}(z)\mathrm{C}(z) = 1 = (\sigma_x^2-\frac{\lambda}{2})\frac{1}{2\pi j}\oint\frac{dz}{z}\mathrm{C}^\dagger(z)\mathrm{S}_{yy}^{-1}(z)\mathrm{C}(z)\ .$$

Hence

$$\mathrm{H}_{UMMSE}(z) = \left(\frac{1}{2\pi j}\oint\frac{dz}{z}\mathrm{C}^\dagger(z)\mathrm{S}_{yy}^{-1}(z)\mathrm{C}(z)\right)^{-1}\mathrm{C}^\dagger(z)\mathrm{S}_{yy}^{-1}(z)$$

Hence, the UMMSE equalizer is simply proportional to the MMSE equalizer, with the proportionality factor adjusted for unbiasedness.

- We find for the MSE

$$MSE_{UMMSE-LE} = E\,(x_k-\widehat{x}_k)^2 = \tfrac{1}{2\pi j}\oint\frac{dz}{z}(\mathrm{S}_{xx}(z)-\mathrm{S}_{x\widehat{x}}(z)-\mathrm{S}_{\widehat{x}x}(z)+\mathrm{S}_{\widehat{x}\widehat{x}}(z))$$

$$= \tfrac{1}{2\pi j}\oint\frac{dz}{z}(\mathrm{S}_{xx}(z)-\mathrm{S}_{xx}(z)-\mathrm{S}_{xx}(z)+\mathrm{S}_{\widehat{x}\widehat{x}}(z)) = \underbrace{\left(\frac{1}{2\pi j}\oint\frac{dz}{z}\mathrm{C}^\dagger(z)\mathrm{S}_{yy}^{-1}(z)\mathrm{C}(z)\right)^{-1}}_{=\ \sigma^2_{\widehat{x}UMMSE}} - \sigma_x^2$$

from which we can find the SNR

$$SNR_{UMMSE-LE} = \frac{\sigma_x^2}{MSE_{UMMSE-LE}} = \frac{\frac{\sigma_x^2}{2\pi j}\oint\frac{dz}{z}\mathrm{C}^\dagger(z)\mathrm{S}_{yy}^{-1}(z)\mathrm{C}(z)}{1-\frac{\sigma_x^2}{2\pi j}\oint\frac{dz}{z}\mathrm{C}^\dagger(z)\mathrm{S}_{yy}^{-1}(z)\mathrm{C}(z)} = \frac{1}{\frac{\sigma_v^2}{2\pi j}\oint\frac{dz}{z}\mathrm{S}_{yy}^{-1}(z)} - 1$$

Unbiased MMSE Linear Equalizers (4)

- we find that

$$SNR_{UMMSE}\ =\ SNR_{MMSE}-1 \quad !$$

Even though the UMMSE LE has a lower SNR than the MMSE LE, it can be shown that its probability of error is lower (using the Gaussian assumption on the interfering symbols - Central Limit Theorem). Since the UMMSE has the highest SNR of all unbiased LEs, it has the lowest probability of error.

- We also have

$$SNR_{ZF-LE}\ \le\ SNR_{UMMSE-LE}\ \le\ MFB\ =\ \frac{1}{\sigma_v^2\,2\pi j}\oint\frac{dz}{z}\mathrm{S}_{yy}(z)-1$$

where the second inequality is due to the Cauchy-Schwarz inequality

$$\left(\frac{1}{2\pi j}\oint\frac{dz}{z}\mathrm{S}_{yy}^{-1}(z)\right)^{-1} \le \frac{1}{2\pi j}\oint\frac{dz}{z}\mathrm{S}_{yy}(z)$$

with equality iff $S_{yy}(f)$ and hence $|C(f)|$ is constant as a function of f ($c_k = c_i\,\delta_{ki}$ for one certain i). In that case also $SNR_{ZF-LE} = MFB$.

Causal Wiener Filtering

- temporal processing: often need to constrain the filter to be causal (real time implementation)
- using the data $\{y_n,\, n \leq k\}$, estimate $x_{k+\lambda}$:

  $\lambda = 0$: *filtering*

  $\lambda > 0$: *prediction*

  $\lambda < 0$: *smoothing*
- Instead of y_k, consider its whitened version $f_{\infty,k}$ which is obtained by filtering y_k with $\mathrm{A}_\infty(z) = \mathrm{S}^+_{yy}(\infty)/\mathrm{S}^+_{yy}(z)$, a causal and causally invertible filter.

  $f_{\infty,k}$ = *innovations* of y_k = white noise with variance $\sigma^2_{f,\infty}$
- $\mathrm{S}_{yy}(z) = \mathrm{S}^+_{yy}(z)\,\mathrm{S}^+_{yy}(z^{-1})$: *spectral factorization*. Subject to certain conditions, a psdf can be factored into its causal minimum-phase factor $\mathrm{S}^+_{yy}(z)$ and its anti-causal maximum-phase counterpart $\mathrm{S}^+_{yy}(z^{-1})$.
- $\mathrm{S}^+_{yy}(\infty) = \sigma_{f,\infty}$ since $\mathrm{S}_{yy}(z) = \frac{\sigma^2_{f,\infty}}{\mathrm{A}_\infty(z)\mathrm{A}_\infty(z^{-1})}$

Causal Wiener Filtering (2)

- LMMSE estimation:

$$\min_{h^f_{k,n}} E(x_{k+\lambda} - \widehat{x}_{k+\lambda})^2\ ,\quad \widehat{x}_{k+\lambda|k} = \sum_{m=-\infty}^{k} h_{k,m} y_m = \sum_{m=-\infty}^{k} h^f_{k,m} f_{\infty,m}$$

- orthogonality conditions $\Rightarrow$ ∞ normal equations: decoupled!

$$E(x_{k+\lambda} - \widehat{x}_{k+\lambda}) f_{\infty,n} = 0 = r_{x\,f_\infty}(\lambda + k - n) - h^f_{k,n}\sigma^2_{f,\infty}$$

  Hence

$$h^f_{k,n} = \frac{r_{x\,f_\infty}(\lambda + k - n)}{\sigma^2_{f,\infty}} = h^f_{0,n-k} = h^f_{k-n},\ n \leq k$$

- $h^f_k = \frac{r_{x\,f_\infty}(\lambda + k)}{\sigma^2_{f,\infty}},\ k \geq 0 \quad \rightarrow \quad \mathrm{H}^f(z) = \frac{1}{\sigma^2_{f,\infty}} \left\{ \mathrm{S}_{x\,f_\infty}(z)\, z^\lambda \right\}_+$

  $\{.\}_+$: "take the causal part of"
- $\mathrm{H}(z) \;=\; \mathrm{H}^f(z)\,\mathrm{A}_\infty(z) = \frac{1}{\sigma^2_{f,\infty}} \left\{ \mathrm{S}_{x\,f_\infty}(z)\, z^\lambda \right\}_+ \frac{\sigma_{f,\infty}}{\mathrm{S}^+_{yy}(z)} = \frac{1}{\mathrm{S}^+_{yy}(z)} \left\{ \frac{\mathrm{S}_{xy}(z)\, z^\lambda}{\mathrm{S}^+_{yy}(z^{-1})} \right\}_+$
- if drop causality constraint: $\mathrm{H}(z) = \frac{\mathrm{S}_{xy}(z)\, z^\lambda}{\mathrm{S}_{yy}(z)}$