Aalto University

T61.3040

Wiener Filter(ing)

Amaury Lendasse and
Yoan Miche

Diagram of the content of first part of the course

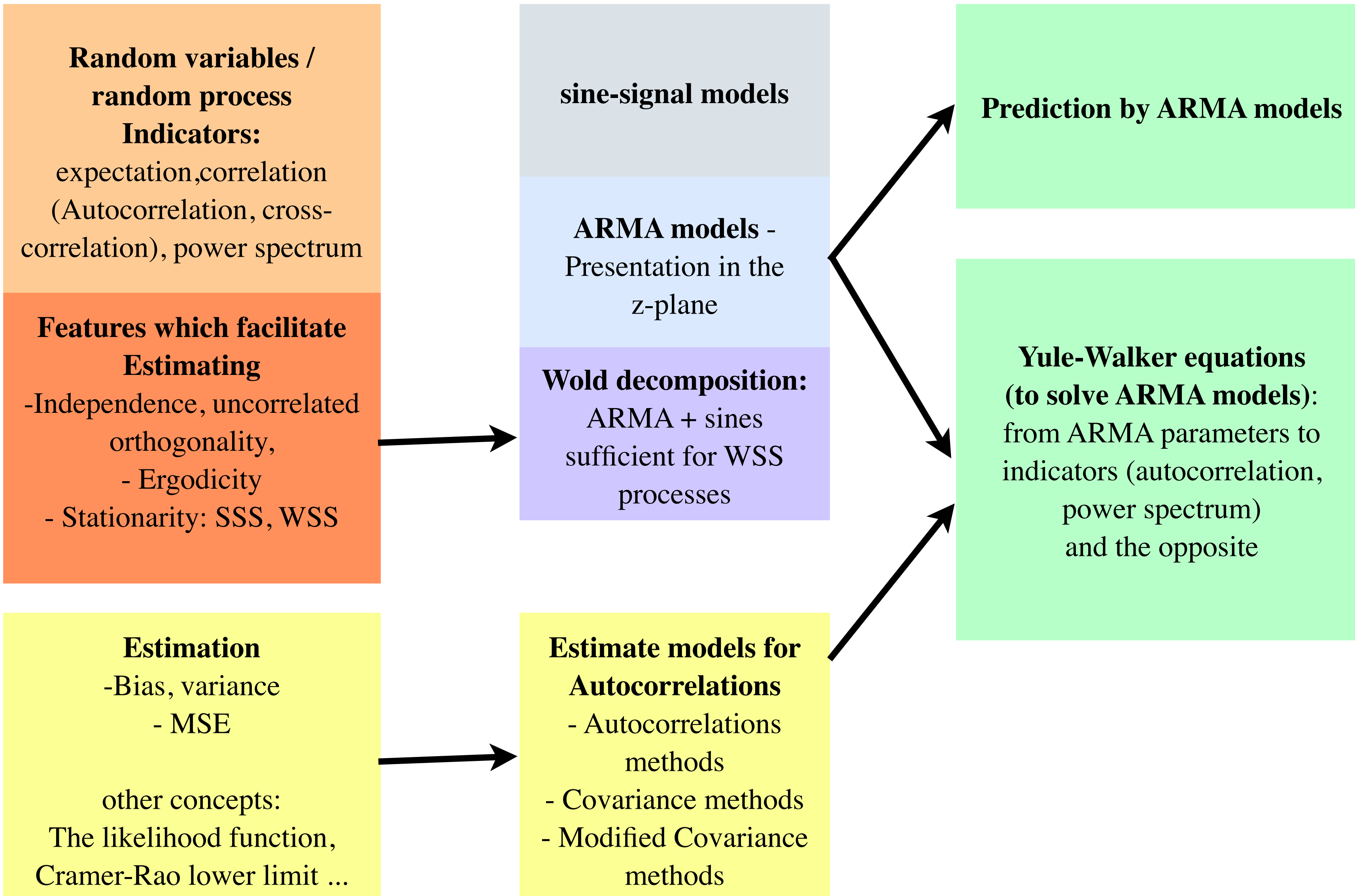
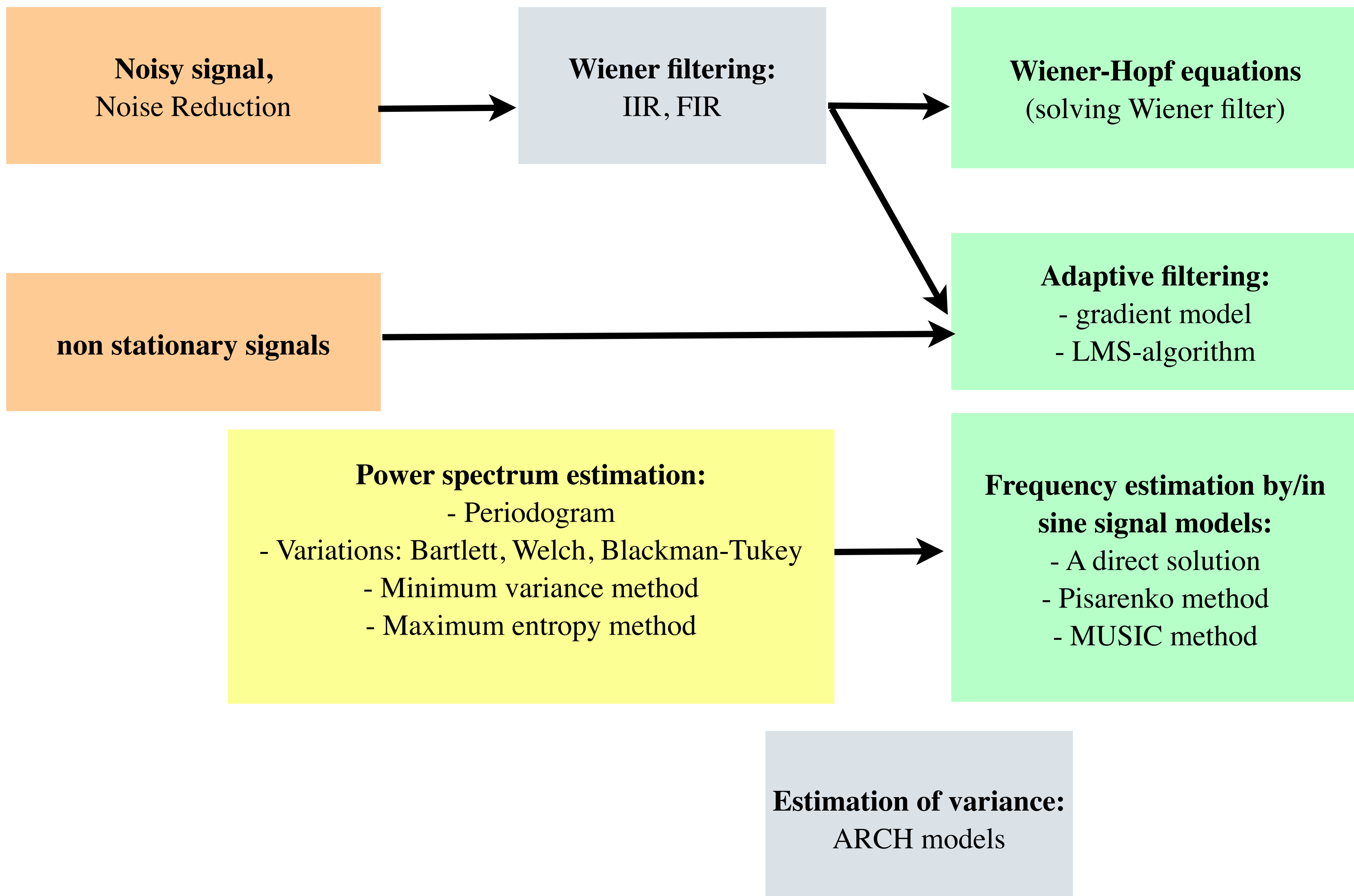


Diagram of content of the final part of the course



Today

Noisy signal,
Noise Reduction

→

Wiener filtering:
IIR, FIR

→

Wiener-Hopf equations
(solving Wiener filter)

Noisy signal

- Earlier, we predicted the process $x(n)$ from the previous values
- We can also predict the process $d(n)$ which is not observed, but it depends on the observed process $x(n)$
- in general, it can be written:

$$x(n) = d(n) + v(n)$$

- by assumptions of the processes and their inter-dependence, then we can estimate $d(n)$ using the observed process $x(n)$

Noisy signal

- Wiener filtering can be interpreted as noise reduction because the interesting process $d(n)$ is observed after addition of the noise $v(n)$
- In this context, the noise does not always mean white noise: the process $v(n)$ can in principle be any kind of noise
- It can be called noise because it contains the part of the detected process that we want to filter out

Noisy signal

Examples:

- Speech signal: the incoming signal $x(n)$ measured by the microphone may contain echoes and noise
- Heart beat measurement, in addition to the heart beats, there is also breathing and sounds from outside of the body
- Image and video signals may contain noise, which reduces the image quality

Wiener Filtering

- The prediction of the process $d(n)$ can be thought by filtering out the noise $v(n)$
- The familiar “pass filtering” from DSP can remove noise, but the best filter choice is not trivial
- Wiener filter is the optimal filter in the sense of MSE, which estimates $d(n)$ from observations

$$x(n) = d(n) + v(n)$$

Wiener Filtering

- Comparing prediction and Wiener filtering
- Linear prediction (AR (p)-process):

$$\hat{x}(n) = \sum_{k=1}^{p} -a(k)x(n-k),$$ we have to know $r_x(k)$

- Linear Wiener filtering:

$$\hat{d}(n) = \sum_{k=0}^{p-1} w(k)x(n-k),$$ we have to know $r_{dx}(k)$

Special cases:

Application	Measured	to estimate
prediction	$x(n) = d(n) + v(n)$	$d(n+k)$, $k>0$
smoothing	$x(n) = d(n) + v(n)$	$d(n-k)$, $k>0$
linear prediction	$x(n) = d(n-1) + v(n)$	$d(n)$
deconvolution	$x(n) = d(n) * g(\mathrm{n}) + v(n)$	$d(n)$

Optimal FIR filter

- First we solve the Wiener filter, which is a finite impulse response (FIR) filter:

$$W(z) = \sum_{n=0}^{p-1} w(n) z^{-n}$$

- Correlations are assumed to be known: $r_x(k)$, $r_d(k)$ and $r_{dx}(k)$
- $x(n)$ and $d(n)$ are WSS processes and $r_{dx}(k) < \infty$ and does not depend on n

Optimal FIR filter

- Wiener filter, where *p* is the order, it is therefore in the time domain

$$\hat{d}(n) = \sum_{l=0}^{p-1} w(l)x(n-l)$$

- Due to the orthogonality principle, we assume that $\hat{d}(n)$ linear combination of values:

$$x(n), x(n-1), \ldots, x(n-p+1)$$

Optimal FIR Filter Solution

- orthogonality principle gives orthogonality conditions

$$
\begin{aligned}
\mathrm{E}(e(n)x^*(n-k)) &= \mathrm{E}(d(n)x^*(n-k)) \\
&\quad - \sum_{l=0}^{p-1} w(l)\,\mathrm{E}(x(n-l)x^*(n-k)) \\
&= 0, \quad k = 0, 1, \ldots, p-1
\end{aligned}
$$

Optimal FIR Filter Solution

- by writing the expectation as correlation, you get the Wiener-Hopf equations:

$$\sum_{l=0}^{p-1} w(l) r_x(k-l) = r_{dx}(k), \quad k = 0, 1, \ldots, p-1$$

Optimal FIR Filter Solution

- Equations in matrix form are $R_x\, w = r_{dx}$

$$\begin{bmatrix} r_x(0) & r_x^*(1) & \dots & r_x^*(p-1) \\ r_x(1) & r_x(0) & \dots & r_x^*(p-2) \\ \vdots & \vdots & & \vdots \\ r_x(p-1) & r_x(p-2) & \dots & r_x(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ \vdots \\ w(p-1) \end{bmatrix} = \begin{bmatrix} r_{dx}(0) \\ r_{dx}(1) \\ \vdots \\ r_{dx}(p-1) \end{bmatrix}$$

- The solution is $R_x^{-1} r_{dx}$

Optimal FIR Filter Solution

- Previously we obtained parameters minimizing MSE
- Calculating the minimum of the error $\mathrm{E}(|e(n)|^2)$

$$
\begin{aligned}
\mathrm{E}(|e(n)|^2) &= \mathrm{E}(e(n)(d(n) - \hat{d}(n))^*) \\
&= \mathrm{E}(e(n)d^*(n)) \quad \text{orthogonality conditions} \\
&= \mathrm{E}(d(n)d^*(n)) - \sum_{l=0}^{p-1} w(l)\,\mathrm{E}(x(n-l)d^*(n))
\end{aligned}
$$

Optimal FIR Filter Solution

- Writing expectation as correlations and using $R_x^{-1} r_{dx}$ gives results

$$
\begin{aligned}
\mathrm{E}\left(|e(n)|^2 \right) &= r_d(0) - \sum_{l=0}^{p-1} w(l) r_{dx}^*(l) \\
&= r_d(0) - r_{dx}^H w \\
&= r_d(0) - r_{dx}^H R_x^{-1} r_{dx}
\end{aligned}
$$

- Result holds only when *w* is the solution of the optimal Wiener filter

Optimal FIR Filter Solution

- WH-equations provide the solution as a function of correlations $r_x(k)$ and $r_{dx}(k)$
- Autocorrelation $r_x(k)$ can be estimated using observations similarly to solving ARMA processes
- $r_{dx}(k)$ estimation is generally impossible, since $d(n)$ is not detected at any time n
- solution requires additional assumptions
- Let's considered the 4 example cases

Example 1:

- $x(n) = d(n) + v(n)$, where the zero mean $v(n)$ is not correlated with $d(n)$
- often a reasonable assumption, if the cause of the noise can not depend on $d(n)$
- e.g., the speech signal and the air conditioner noise, satisfy previous assumptions

Example 1:

- We calculate $r_x(k)$ and $r_{dx}(k)$ using previous assumptions,
- We obtain

$$\begin{aligned} r_{dx}(k) &= \mathrm{E}\big(d(n)x^*(n-k)\big) \\ &= \mathrm{E}\big(d(n)d^*(n-k)\big) + \mathrm{E}\big(d(n)v^*(n-k)\big) \\ &= \mathrm{E}\big(d(n)d^*(n-k)\big) \quad \text{d and v are not correlated} \\ &= r_d(k) \end{aligned}$$

Example 1:

- In addition, we get

$$
\begin{aligned}
r_x(k) &= \mathrm{E}\left(x(n+k)x^*(n)\right) \\
&= \mathrm{E}\left(d(n+k)d^*(n)\right) + \mathrm{E}\left(v(n+k)v^*(n)\right) \\
&\quad + \mathrm{E}\left(d(n+k)v^*(n)\right) + \mathrm{E}\left(v(n+k)d^*(n)\right) \\
&= r_d(k) + r_v(k) + r_{dv}(k) + r_{vd}(k) \\
&= r_d(k) + r_v(k)
\end{aligned}
$$

- Because $d(n)$ and $v(n)$ are uncorrelated

Example 1:

- We obtain the Wiener-Hopf equations

$$(R_d + R_v)w = r_d$$

where $r_d = \left[r_d(0), r_d(1), \ldots, r_d(p-1)\right]^T$

- If the noise is white noise then we get

$$(R_d + \sigma^2 I)w = r_d$$

Example 1:

- The matrices $R_d + R_v = R_x$ estimated from observations
- Also needed additional assumptions in order to obtain r_d
- How good is a filter if you need to know already r_d ?
  Answer: r_d tells only the statistical properties of $d(n)$.
- We want to estimate by filter the actual values $d(n)$!
- Example: when observing $x(n) = A\sin(\omega n + \phi) + v(n)$

  where $v(n)$ is a Gaussian white noise with variance σ^2

- demo: wfilt.R

Example 2:

- Linear prediction, $d(n) = x(n+1)$
- In WH-equations we have r_{dx} which is now

$$
\begin{aligned}
r_{dx}(k) &= \mathrm{E}\left(d(n)x^*(n-k)\right) \\
&= \mathrm{E}\left(x(n+1)x^*(n-k)\right) \\
&= r_x(k+1)
\end{aligned}
$$

Example 2:

- We get rid of $d(n)$ and get the Wiener-Hopf equations

$$\begin{bmatrix} r_x(0) & r_x^*(1) & \cdots & r_x^*(p-1) \\ r_x(1) & r_x(0) & \cdots & r_x^*(p-2) \\ \vdots & \vdots & & \vdots \\ r_x(p-1) & r_x(p-2) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ \vdots \\ w(p-1) \end{bmatrix} = \begin{bmatrix} r_x(1) \\ r_x(2) \\ \vdots \\ r_x(p) \end{bmatrix}$$

Example 2:

- The minimum error is obtained

$$\varepsilon = r_x(0) - \sum_{l=0}^{p-1} w(l) r_x^*(l+1)$$

- These are similar to the YW equations for the AR-model
- Note! signs of the coefficients have changed, because correspondence between the AR model and the Wiener filter

Example 3:

- Noisy signal prediction: $x(n) = y(n) + v(n)$

- The desired signal $d(n) = y(n+1)$,
- Assuming y and v non-correlated: $\hat{d}(n) = w(0)x(n) + \ldots$

$$r_x(k) = \mathrm{E}\left(x(n)x^*(n-k)\right) = r_y(k) + r_v(k) = r_d(k) + r_v(k)$$

$$\begin{aligned} r_{dx}(k) &= \mathrm{E}\left(d(n)x^*(n-k)\right) = \mathrm{E}\left(y(n+1)x^*(n-k)\right) \\ &= \mathrm{E}\left(y(n+1)y^*(n-k)\right) = r_y(k+1) = r_d(k+1) \end{aligned}$$

- The only difference between with the noise-free case is that r_v is added to r_d

Example 3:

- In matrix form, we get the Wiener-Hopf equations:

$$(R_d + R_v)w = r_{dx}$$

Where $r_{dx} = [r_d(1), r_d(2), \ldots, r_d(p)]^T$ as in the noiseless case

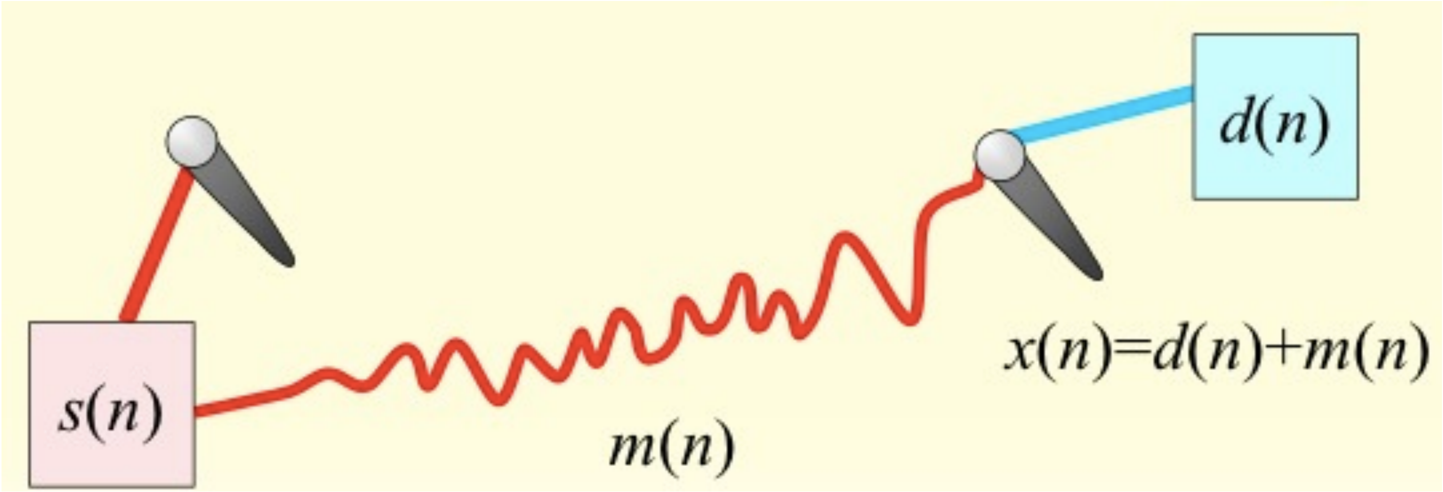
- demo: wpred.R

Example 4:

- Noise reduction using the observed noise
- Measuring by microphone for the speech signal $d(n)$
- Microphone are noisy signal $x(n) = d(n) + \mathrm{m}(n)$
- The objective is to estimate $d(n)$

Example 4:

- in addition we assume that the noise is caused by a noise source $s(n)$
- which can be observed directly
- Noise $m(n)$ entering the microphone is changed along the way
- So m ≠ s, but processes are not uncorrelated



Example 4:

- The problem can be solved by estimating the microphone noise $m(n)$
- If this is done well, $d(n)$ may be obtained by subtracting the estimated $m(n)$ from $x(n)$
- Estimation of microphone noise is possible because we directly observe process $s(n)$

Example 4:

- Now the process $s(n)$ has been observed and the desired process is $m(n)$
- General WH equations: $R_x\, w = r_{dx}$
- Substituting $s \rightarrow x$ and $m \rightarrow d$

  we obtain the WH equations: $R_s\, w = r_{ms}$

- R_s can be estimated directly from observations

Example 4:

- Calculated the cross-correlation is obtained developing:

$$
\begin{aligned}
r_{ms}(k) &= \mathrm{E}\left(m(n)s^*(n-k)\right) \\
&= \mathrm{E}\left(x(n)s^*(n-k)\right) - \mathrm{E}\left(d(n)s^*(n-k)\right) \\
&= r_{xs}(k)
\end{aligned}
$$

s and *d* uncorrelated

- So the $R_{ms} = R_{xs}$ can be estimated from the observations
- $m(n)$ is obtained using the Wiener filter $w = R_s^{-1} R_{xs}$

Example 4:

- Test in practice:

$$d(n) = \sin(n\omega_0 + \phi)$$
$$m(n) = a_1(1)m(n-1) + a_1(2)m(n-2) + g(n)$$
$$s(n) = a_2(1)s(n-1) + a_2(2)s(n-2) + g(n)$$

- AR (2)-processes, $m(n)$ and $s(n)$ are correlated, since $g(n)$ is the same for both process
- demo: noisec.R

Optimal IIR filter

- Causal FIR filter: $h(n) = 0$ when $n<0$ or $n>p-1$
- Generally LSI (linear shift-invariant) filter is a non-causal IIR filter
- Then the impulse response may differ from zero for all values of n
- We solve the non-causal IIR Wiener filter

Optimal IIR filter

- Error signal and the MSE are

$$e(n) = d(n) - \hat{d}(n) = d(n) - \sum_{l=-\infty}^{\infty} h(l)x(n-l)$$

$$\varepsilon = \mathrm{E}(|e(n)|^2)$$

- WH equations follow from orthogonality condition

Optimal IIR filter

- Orthogonality conditions:

$$\mathrm{E}(e(n)x^*(n-k)) = 0; \quad -\infty < k < \infty$$

- Difference to the the FIR case is the infinite number of equations
- Substituting the error signal to the orthogonality conditions, we get

$$\sum_{l=-\infty}^{\infty} h(l) r_x(k-l) = r_{dx}(k); \quad -\infty < k < \infty$$

Optimal IIR filter

- Orthogonality condition is a convolution

$$h(k) * r_x(k) = r_{dx}(k), \quad -\infty < k < \infty$$

- z-transformation of both sides will give

- $$H(z)P_x(z) = P_{dx}(z) \Rightarrow$$ $$H(z) = \frac{P_{dx}(z)}{P_x(z)}$$ (IIR Wiener - Filter)

Optimal IIR filter

- Example: $x(n) = d(n) + v(n)$, $E(v(n)) = 0$ and v, d are not correlated
- Then $r_x(k) = r_d(k) + r_v(k)$ so

$$P_x(z) = P_d(z) + P_v(z)$$

- Cross-correlation is $r_{dx}(k) = r_d(k)$ which gives

$$P_{dx}(z) = P_d(z)$$

Optimal IIR filter

- IIR Wiener filter is

$$H(z) = \frac{P_d(z)}{P_d(z) + P_v(z)}$$

- Interpretation of power changes in frequency omega:

$$— \quad P_d(\exp(j\omega)) << P_v(\exp(j\omega)): \quad H(\exp(j\omega)) \approx 0,$$

Omega is filtered off

$$— \quad P_d(\exp(j\omega)) >> P_v(\exp(j\omega)): \quad H(\exp(j\omega)) \approx 1,$$

Omega's power remains