Time Series Analysis:

4. Wiener filter

P. F. Góra

http://th-www.if.uj.edu.pl/zfs/gora/

2009
**Wiener filter (optimal filter)**

A system generates a *stationary* signal $u(t)$. We record this signal with a device that has a known response function $r(t)$. In addition, the signal is *contaminated* with Gaussian White Noise (GWN), *not correlated* with the signal. We actually register

\[
c(t) = s(t) + \eta(t) = \int_{-\infty}^{\infty} r(t - \tau) u(\tau) d\tau + \eta(t).
\]  

(1)

We record $c(t)$, we know $r(t)$, we make assumptions about $\eta(t)$. What can we say about $u(t)$?
We will construct an estimate $\tilde{u}(t)$ that is optimal in the least squares sense.

$$\langle \int_{-\infty}^{\infty} |u(t) - \tilde{u}(t)|^2 \, dt \rangle = \text{minimum} ,$$

(2)

$\langle \cdots \rangle$ stands for averaging over realizations of the noise. By Parseval identity we have for the Fourier transforms

$$\langle \int_{-\infty}^{\infty} \left| U(f) - \tilde{U}(f) \right|^2 \, df \rangle = \text{minimum} .$$

(3)

$S(f) = U(f)R(f)$. We seek an estimate in the Fourier domain:

$$\tilde{U}(f) = \frac{C(f)\Phi(f)}{R(f)} .$$

(4)

*The errors are Gaussian!
We need to minimize with respect to $\Phi$:

\[
\left\langle \int_{-\infty}^{\infty} \left| \frac{[S(f) + N(f)] \Phi(f)}{R(f)} - \frac{S(f)}{R(f)} \right|^2 df \rightangle
\]

\[
= \int_{-\infty}^{\infty} |R(f)|^{-2} \left\langle |(S(f) + N(f)) \Phi(f) - S(f)|^2 \right\rangle df
\]

\[
= \int_{-\infty}^{\infty} |R(f)|^{-2} \left\langle |S(f)|^2 |\Phi(f)|^2 + S(f)N^*(f)|\Phi(f)|^2 + S(f)*N(f)|\Phi(f)|^2 + |N(f)|^2 |\Phi(f)|^2 - |S(f)|^2 \Phi(f)
\right.
\]

\[
- N(f)S^*(f) \Phi(f) - |S(f)|^2 \Phi^*(f) - N^*(f)S(f) \Phi^*(f) + |S(f)|^2 \right\rangle df
\]

(5)
Products of the noise and the signal are marked in red. Their averages vanish by the assumption of their statistical independence. All that remains is

\[
\int_{-\infty}^{\infty} |R(f)|^{-2} \left( |S(f)|^2 \left( |\Phi(f)|^2 - \Phi(f) - \Phi^*(f) + 1 \right) \right. \\
+ \left. \left\langle |N(f)|^2 \right| \Phi(f) \right| \right)^2 \right) \, df = \\
\int_{-\infty}^{\infty} |R(f)|^{-2} \left( |S(f)|^2 \left| 1 - \Phi(f) \right|^2 + \left\langle |N(f)|^2 \right| \Phi(f) \right| \right)^2 \, df = \text{minimum}. 
\]

(6)
Assuming that $\Phi$ is real, (6) is minimized for:

$$\Phi(f) = \frac{|S(f)|^2}{|S(f)|^2 + \langle |N(f)|^2 \rangle}.$$ (7)

This is the Wiener filter, also known as (AKA) the optimal filter.

We estimate the average noise level via the power spectrum. The filter (7) (almost) vanishes in region where there is (almost) no signal, and is close to unity where there is (almost) no noise.
Caveat emptor!

In practice, we try to guess both $|S(f)|^2$ and $|N(f)|^2$, using the same power spectrum. Therefore, we need to use

$$
\Phi(f) = \begin{cases} 
\frac{P(f) - \hat{N}(f)}{P(f)} & \text{if } P(f) > \hat{N}(f), \\
0 & \text{otherwise}.
\end{cases}
$$

(8)

$P(f)$ stands for the power spectrum of the full signal $c(t)$. $\hat{N}(f)$ is an approximate power spectrum of the noise, usually fitted to the high frequency part. It is most important that $\Phi(f) \geq 0$. $P(f)$ can be calculated with a windowing function.
Noisy and clean signals

![Graph showing noisy and clean signals](image-url)
Wiener filter (square window)

Copyright © 2009 P. F. Góra
Wiener filter (Welch window)
Naive denoising, $P(f) < \text{threshold} \iff C(f) = 0$, fails:
A common form of noise

Frequently, the power spectrum of the noise has the form

\[ N(f) = \frac{N_0}{1 + |\frac{f}{f_0}|^a}. \]  \hspace{1cm} (9)

We estimate \( N_0, f_0, a \) by fitting Eq. (9) to “experimental” data.

For \( |f| \gg 1 \), \( N(f) \sim |f|^{-a} \). In the log-log plot, the “tail” is a straight line with a slope \(-a\). \( a > 1 \! \)!

If \( 1 < a < 2 \), we have a fractal noise.
Example

A fitted line (9) with $N_0 = 9.5 \cdot 10^{-4}$, $f_0 = 5/32$, $a = 1.9$. 
In practice, the spectra are of lower quality...

A fitted line (9) with $N_0 = 9.5 \cdot 10^{-4}$, $f_0 = 5/32$, $a = 1.9$. 
Before filtering
After filtering
Spectrum after filtering

![Spectrum graph]

Copyright © 2009 P. F. Góra
Filtering the filtered: A cascade Wiener filter
Spectrum after the cascade filter
**Wiener filter and non-Gaussian noises**

The Wiener filter is optimal under the assumption that the noise is Gaussian. How does the Wiener filter perform for non-Gaussian noises? In particular, how does the filter perform if the second or the first moments of the noise do not exist, or if the noise has “heavy tails”?

We will analyse a process generated with

\[ g(t) = \sin 2\pi t + \frac{1}{16} \zeta(t), \quad (10) \]

with \( \zeta(t) \) drawn from the Cauchy distribution and \( \langle \zeta(t_1)\zeta(t_2) \rangle = 0 \) for \( t_1 \neq t_2 \).
1st realization: The signal before filtering
1st realization: The power spectrum of the unfiltered signal
1st realization: After filtering
1st realization: After filtering twice, in a cascade
So far, so good.

The signal has a unique, and known, characteristic frequency. Note that (i) the amplitude of the filtered signal has dropped (the filter is lossy), and (ii) the average of the denoised signal is shifted from zero,

Now, do the same trick with a different realization of the same process (10):
2nd realization: The signal before filtering
2nd realization: Power spectrum of the unfiltered signal
2nd realization: After filtering

The signal has *deteriorated* as a result of filtering!
Conclusions?

- If the second moment of the noise contaminating the signal does not exist, in particular, if the noise is a Lévy process, *it may so happen* that the Wiener filter filters out the noise.

- It may also happen that the Wiener filter deteriorates the signal!

- Either of the above may happen for different realizations of *the same* process.

- There are no rules.

- If the noise is Gaussian *and the signal is stationary*, the filter cleans the noise almost surely (with probability 1 in the limit of an infinite signal).
Filtering in the Fourier domain

Filtering: Multiplying the transform by a *transform function*.

\[
g_n \xrightarrow{\text{DFT}} G_n \xrightarrow{\mathcal{H}(f_n)} G_n \xrightarrow{\text{inverse DFT}} \tilde{g}_n \tag{11}
\]

where \(\mathcal{H}(f_n)\) is the transfer function discretized over \(f_n\).

The most important drawback: You need to know the whole series, collect all data, *prior* to filtering.
An example is the characteristic function

\[ \mathcal{H}(f) = \begin{cases} 
1 & |f| \leq f_c \\
0 & |f| > f_c 
\end{cases} \]

Alternatively, it can be expressed as

\[ \mathcal{H}(f) = \frac{2 \exp \left( -\left( \frac{f}{f_c} \right)^4 \right)}{1 + \exp \left( -\left( \frac{f}{f_c} \right)^4 \right)} \]
The signal before and after filtering
A reason for the strange result of the “blue” filtering

When filtering in the Fourier domain, we usually do not smooth out the filters.