Time Series Analysis:

3. Gaussian White Noise

The Wiener filter

P. F. Góra

https://zfs.fais.uj.edu.pl/pawel_gora

2022

White noise

The most popular form of a noise, or a random contamination of a signal, is called *the white noise*.

Formally, the white noise is a stochastic process $\eta(t)$ such that

- for any possible t, $\eta(t)$ is a Gaussian random variable with the standard distribution

$$N(0,1)(\eta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\eta^2} \tag{1}$$

- for any possible t, t'

$$\left\langle \eta(t)\eta(t') \right\rangle = \delta(t-t')\,, \tag{2}$$

  where $\langle \cdots \rangle$ stands for stochastic averaging and δ is the Dirac δ-function.

It follows immediately that

$$\langle \eta(t) \rangle = 0 \tag{3a}$$

$$\left\langle (\eta(t))^2 \right\rangle = 1 \tag{3b}$$

We will frequently use the abbreviation GWN for the Gaussian White Noise.

A discrete GWN is an infinite sequence $\{\eta_n\}_{-\infty}^{\infty}$ such that any η_n is a Gaussian random variable with the standard $N(0,1)$ distribution, satisfying

$$\langle \eta_n \rangle = 0 \tag{4a}$$

$$\langle \eta_n \eta_m \rangle = \delta_{nm} \tag{4b}$$

Equations (2) and the second of the equations (4) say that η's taken at different times are statistically independent.

Power spectrum of the white noise

. . . or why is this noise "white"?

The white noise is stationary and we can obtain its power spectrum from the Wiener-Khinchin theorem

$$N(f) = \left\langle |\mathcal{N}(f)|^2 \right\rangle \tag{5}$$

where $\mathcal{N}(f)$ is the Fourier transform of the GWN:

$$\mathcal{N}(t) = \int\limits_{-\infty}^{\infty} \eta(t) e^{2\pi i f t}\, dt = \lim_{T\to\infty} \int\limits_{-T/2}^{T/2} \eta(t) e^{2\pi i f t}\, dt\,. \tag{6}$$

In (5) we need to take the statistical average as the Fourier transform of a stochastic process is also a stochastic process, but in Fourier domain.

Setting aside the limit to simplify notation, we get

$$N(f) = \left\langle \left| \int\limits_{-T/2}^{T/2} \eta(t) e^{2\pi i f t}\, dt \right|^2 \right\rangle = \left\langle \int\limits_{-T/2}^{T/2} \int\limits_{-T/2}^{T/2} \eta(t)\eta(t') e^{2\pi i f t} e^{-2\pi i f t'}\, dt\, dt' \right\rangle$$

$$= \int\limits_{-T/2}^{T/2} \int\limits_{-T/2}^{T/2} \left\langle \eta(t)\eta(t') \right\rangle e^{2\pi i f t} e^{-2\pi i f t'}\, dt\, dt'$$

$$= \int\limits_{-T/2}^{T/2} \int\limits_{-T/2}^{T/2} \delta(t-t') e^{2\pi i f (t-t')}\, dt\, dt' = \int\limits_{-T/2}^{T/2} dt = T\,. \tag{7}$$

We can see that the power spectrum (7) is divergent as $T \to \infty$: the white noise contains infinite power. However, any *practically interesting* process must have a finite duration (the process is exactly zero outside a certain interval) and we do not need to evaluate infinite integrals. The mathematically correct white noise is only a model, an idealisation of more realistic processes.

Another consequence of (7) is that the power spectrum is flat: The power density does not depend on the frequency, all frequencies contribute equally to the power spectrum. Therefore we call this process *white*.

Reasons for using GWN

1. GWN has nice mathematical properties that are easy to handle ☺.

2. The other reason for using GWN comes by virtue of the **Central Limit Theorem**:

Theorem: Let $\{x_1, x_2, \ldots x_n\}$ be a random sample, or a sequence of independent and identically distributed (iid's) random variables, drawn from a distribution of expected value given by μ and finite variance given by σ^2. Then the distribution of the rescaled sample average

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \mu\right) \tag{8}$$

approaches the normal distribution $N(0, \sigma^2)$ as $n \to \infty$.

It is a consequence of the Central Limit Theorem that if a random distribution is an accumulation of *many* microscopic, very difficult to predict events, like collisions with air molecules, their collective effect is practically Gaussian. Therefore, many random events, like thermal fluctuations of local density in a gas, thermal currents, fluctuations of the magnetic field of the Earth and even measurement errors, are very well approximated by a GWN. A more thorough analysis shows that a series formed by fluctuations in any macroscopic body in thermal equilibrium are approximated by GWN. It was one of the discoveries of Marian Smoluchowski, a great Polish physicists.

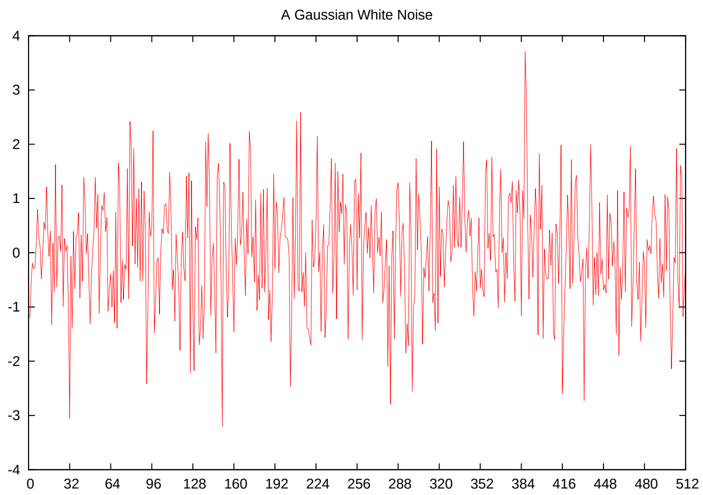
Marian Smoluchowski (1872-1917)

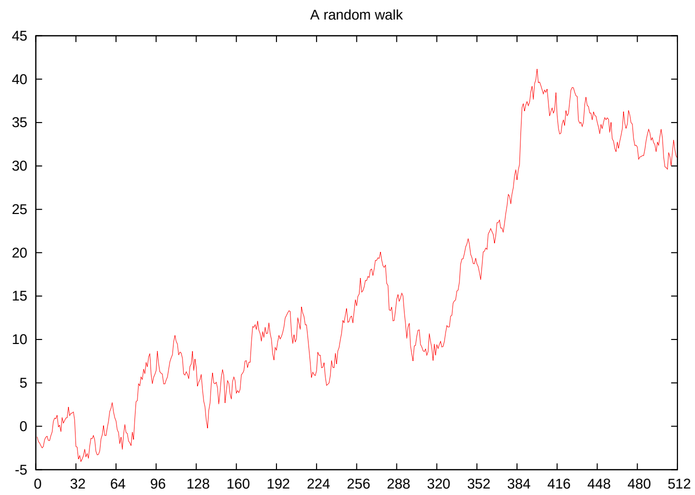
The random walk

GWN may be considered as a series of steps $\{\eta_n\}$ taken by a 1-d random walker. The cumulative effect of those steps

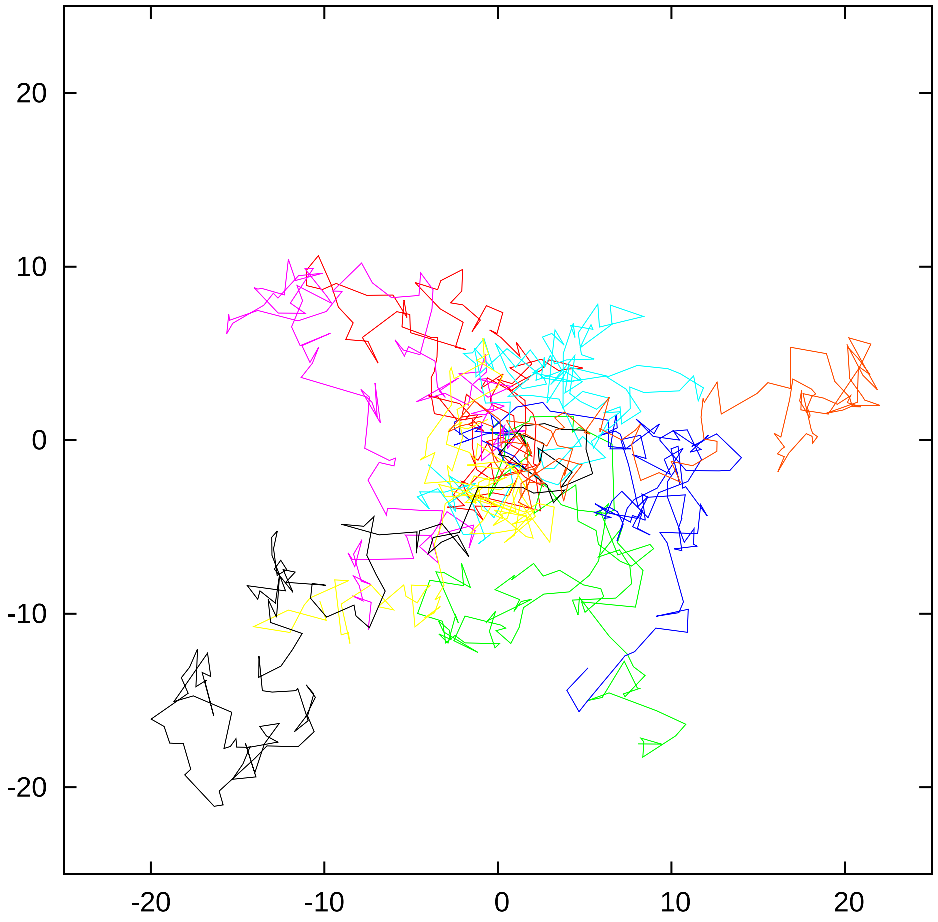
$$x_n = x_{n-1} + \eta_n \tag{9}$$

is called a random walk or a Brownian motion.





Several trajectories of a 2-d random walker



It is easy to see that the Brownian motion (9) satisfies

$$x_n = \sum_{i=1}^{n} \eta_i \,. \tag{10}$$

Therefore $\langle x_n \rangle = 0$ and

$$\begin{aligned}
\left\langle x_n^2 \right\rangle &= \left\langle \left(\sum_{i=1}^{n} \eta_i \right)^2 \right\rangle = \left\langle \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i \eta_j \right\rangle \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \left\langle \eta_i \eta_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij} = \sum_{i=1}^{n} 1 \\
&= n \,.
\end{aligned} \tag{11}$$

Probability: α-stable distributions

A probability distribution is α-stable if the sum of two independent random variables drawn from this distribution has the same distribution, possibly shifted and rescaled. The Gaussian (or normal) distribution is α-stable. There are infinitely many α-stable distributions but the Gaussian distribution is the only α-stable distribution that has a finite mean and a finite variance. Two other well known α-stable distributions are

- Cauchy distribution:

$$\rho(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2}$$

  (does not have a variance, has a mean as a principal value only)

- Lévy-Smirnov distribution ($x > 0$):

$$\rho(x) = \frac{1}{\sqrt{2\pi}} \, x^{-3/2} \, e^{-\frac{1}{2x}}$$

  (does not have neither a variance, nor a mean).

If we generalize the Central Limit Theorem and omit the part about a finite mean and a finite variance, it turns out that the distribution of the sample average approaches an α-stable distribution.

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

$$c(t) = s(t) + \eta(t) = \int\limits_{-\infty}^{\infty} r(t-\tau)u(\tau)\,d\tau + \eta(t)\,. \tag{12}$$

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.

$$\left\langle \int\limits_{-\infty}^{\infty} |u(t) - \tilde{u}(t)|^2 \, dt \right\rangle = \text{minimum}\,, \tag{13}$$

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

$$\left\langle \int\limits_{-\infty}^{\infty} \left| U(f) - \tilde{U}(f) \right|^2 \, df \right\rangle = \text{minimum}\,. \tag{14}$$

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

$$\tilde{U}(f) = \frac{C(f)\Phi(f)}{R(f)}\,. \tag{15}$$

*The errors are Gaussian!

We need to minimize with respect to Φ:

$$\left\langle \int\limits_{-\infty}^{\infty} \left| \frac{[S(f)+N(f)]\Phi(f)}{R(f)} - \frac{S(f)}{R(f)} \right|^2 df \right\rangle$$

$$= \int\limits_{-\infty}^{\infty} |R(f)|^{-2} \left\langle |(S(f)+N(f))\Phi(f) - S(f)|^2 \right\rangle df$$

$$= \int\limits_{-\infty}^{\infty} |R(f)|^{-2} \Big\langle |S(f)|^2|\Phi(f)|^2 + S(f)N^*(f)|\Phi(f)|^2$$
$$+ S(f)^*N(f)|\Phi(f)|^2 + |N(f)|^2|\Phi(f)|^2 - |S(f)|^2\Phi(f)$$
$$- N(f)S^*(f)\Phi(f) - |S(f)|^2\Phi^*(f) - N^*(f)S(f)\Phi^*(f) + |S(f)|^2 \Big\rangle df \qquad (16)$$

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

$$\int\limits_{-\infty}^{\infty} |R(f)|^{-2} \left(|S(f)|^2 \left(|\Phi(f)|^2 - \Phi(f) - \Phi^*(f) + 1 \right) \right.$$
$$\left. + \left\langle |N(f)|^2 \right\rangle |\Phi(f)|^2 \right) df =$$
$$\int\limits_{-\infty}^{\infty} |R(f)|^{-2} \left(|S(f)|^2 |1 - \Phi(f)|^2 + \left\langle |N(f)|^2 \right\rangle |\Phi(f)|^2 \right) df = \text{minimum}\,. \tag{17}$$

Assuming that Φ is real, (17) is minimized for:

$$\Phi(f) = \frac{|S(f)|^2}{|S(f)|^2 + \left\langle |N(f)|^2 \right\rangle} . \tag{18}$$

The power spectra are defined for positive frequancies only but we want to have a filter that acts on components with negative frequencies as well. We therefore generalize (18) to

$$\Phi(f) = \frac{|S(|f|)|^2}{|S(|f|)|^2 + \left\langle |N(|f|)|^2 \right\rangle} . \tag{19}$$

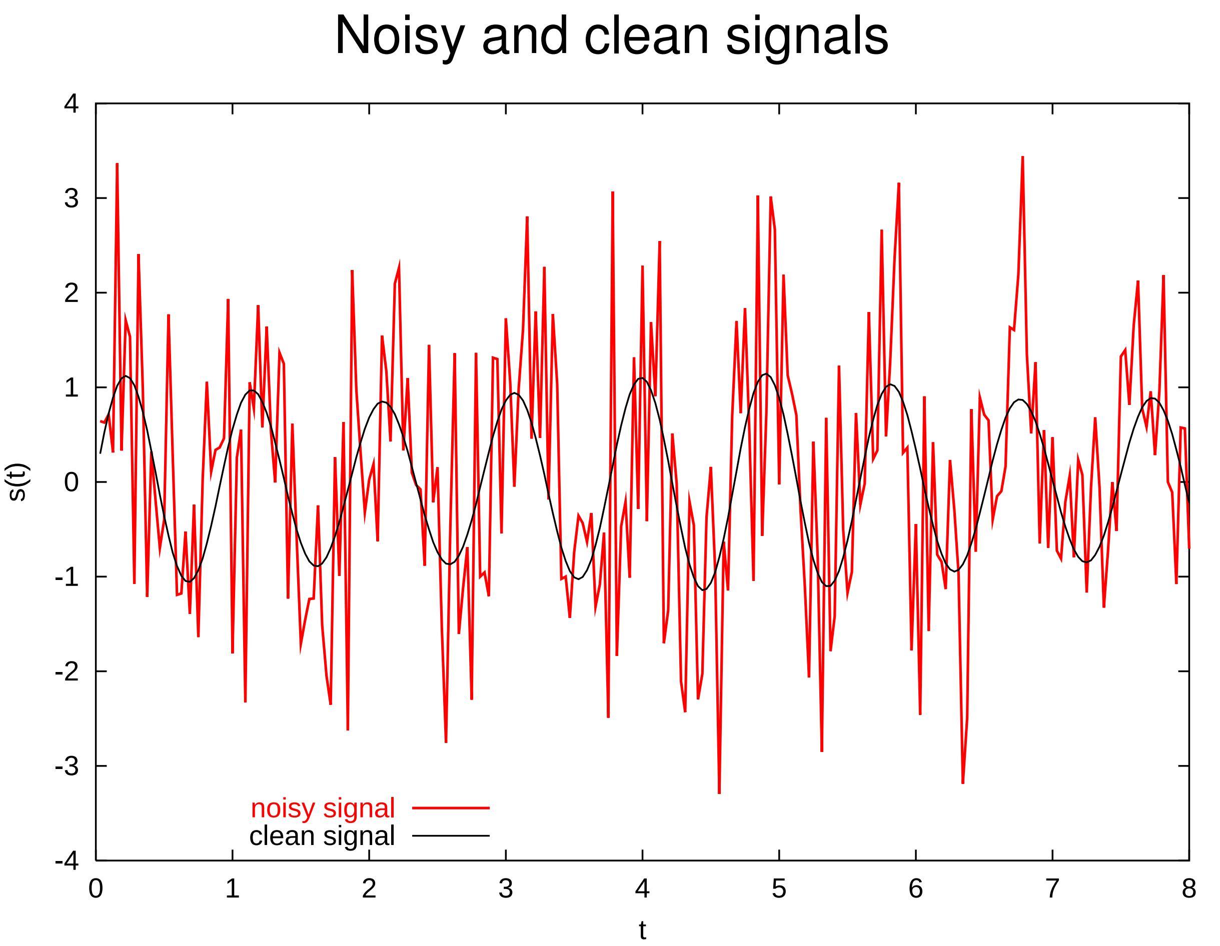
This is the *Wiener filter*, also known as (AKA) *the optimal filter*. We estimate the average noise level via the power spectrum. The filter (19) (almost) vanishes in a region where there is (almost) no signal, and is close to unity where there is (almost) no noise.

Caveat emptor!

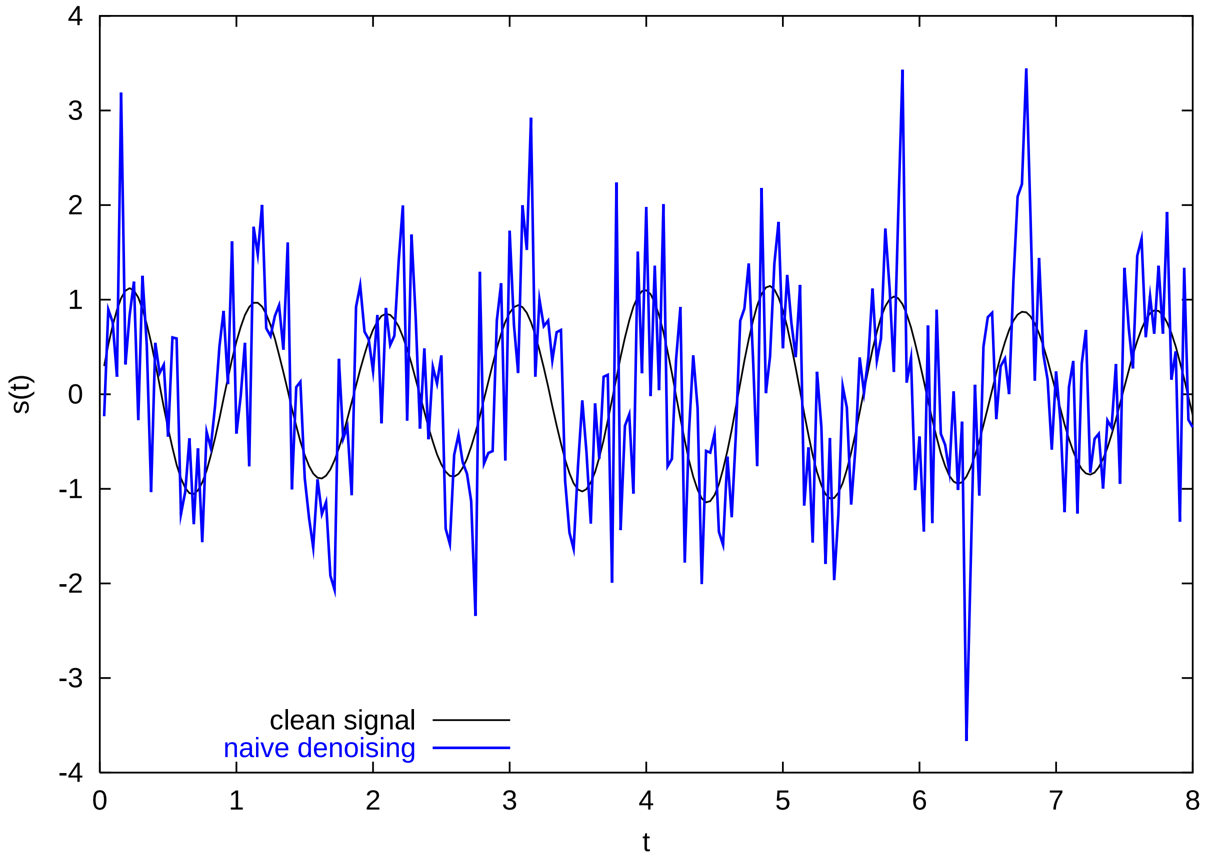
In practice, we try to *guess* both $|S(f)|^2$ and $|N(f)|^2$, using the same power spectrum. Let $P(f)$ be the power spectrum. For signals independent form the noise, $P(f) = |S(f)|^2 + \left\langle |N(f)|^2 \right\rangle$. Therefore, we need to use

$$\Phi(f) = \begin{cases} \dfrac{P(|f|) - |N(|f|)|^2}{P(|f|)} & \text{if } P(|f|) > |N(|f|)|^2\,, \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

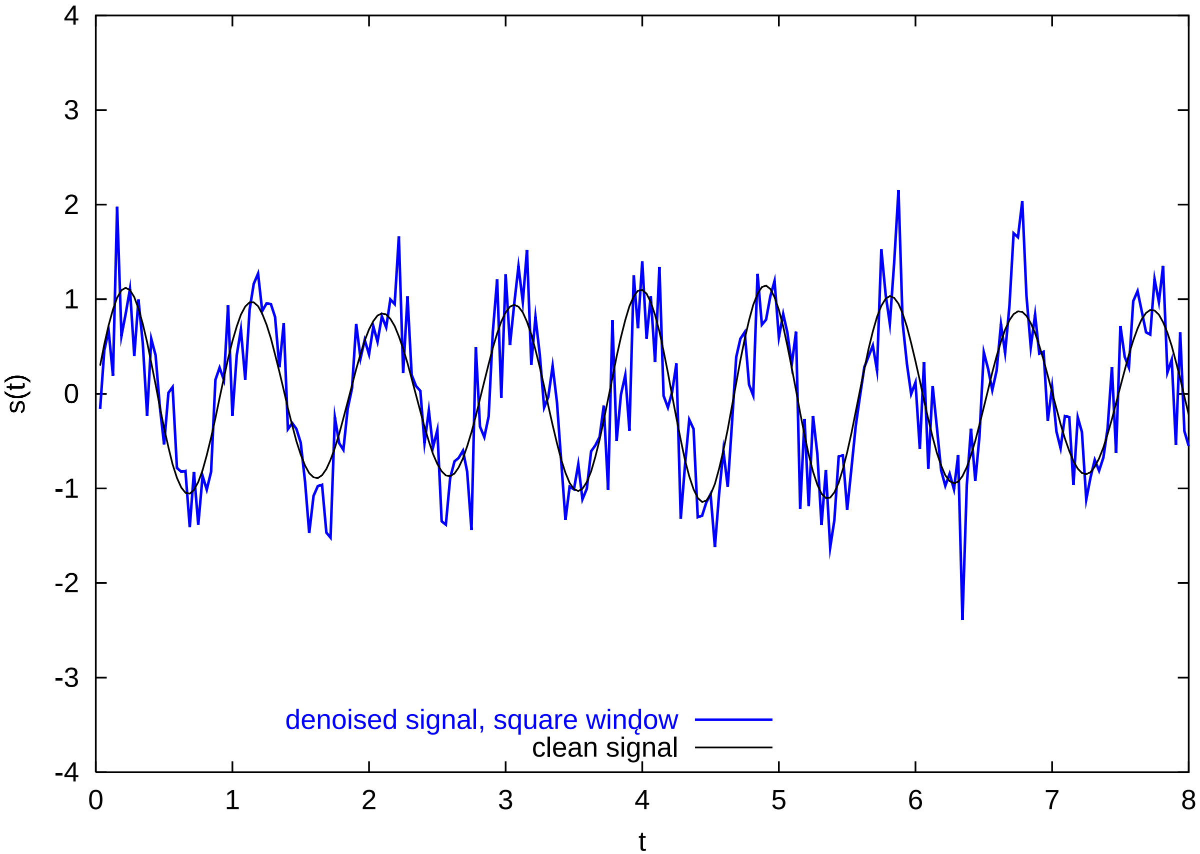
$P(f)$ stands for the power spectrum of the *full* signal $c(t)$. $|N(f)|^2$ is an approximate power spectrum of the noise, usually fitted to the high frequency part. It is most important that $\Phi(f) \geqslant 0$. $P(f)$ can be calculated with a window function.


Noisy and clean signals
4
3
2
1
0
-1
-2
-3
-4
s(t)
0
1
2
3
4
5
6
7
8
t
noisy signal
clean signal

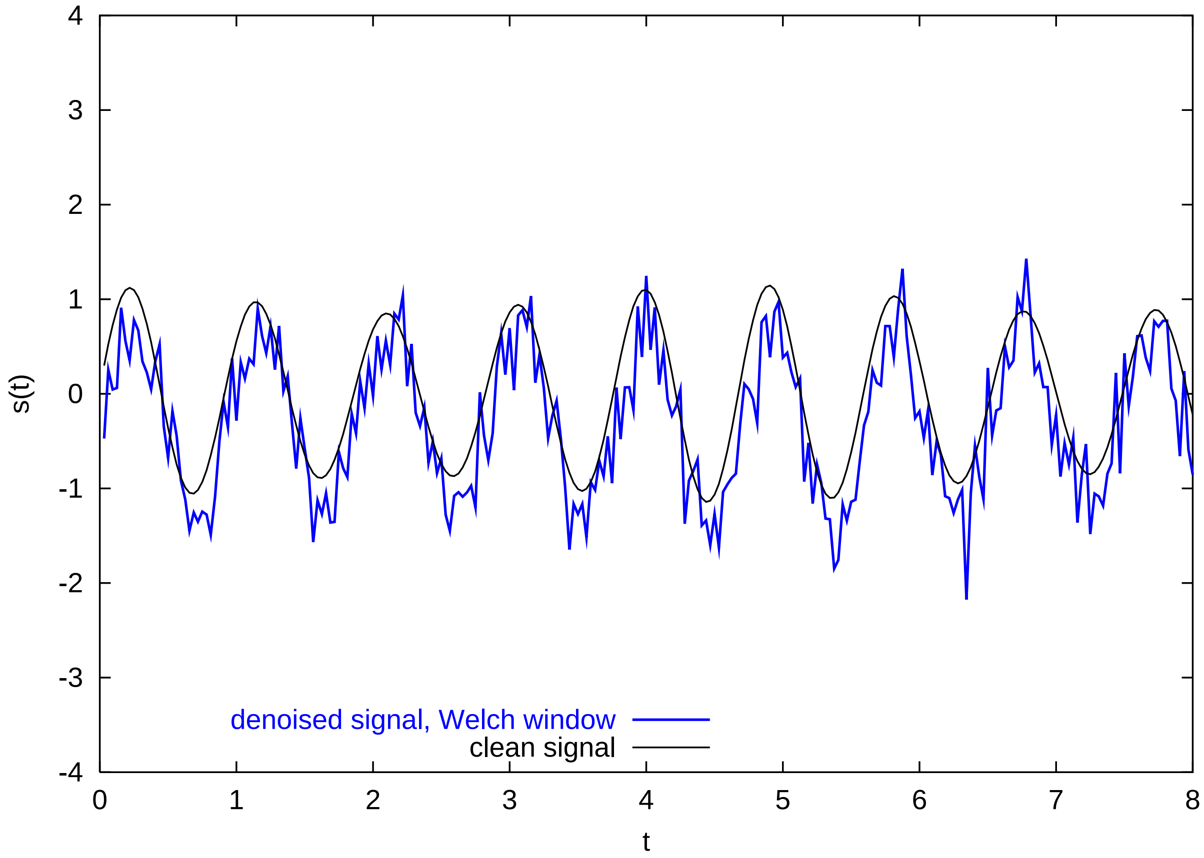
Naive denoising, $P(f) < \text{threshold} \Rightarrow C(f) = 0$, fails:



Wiener filter (square window)



Wiener filter (Welch window)

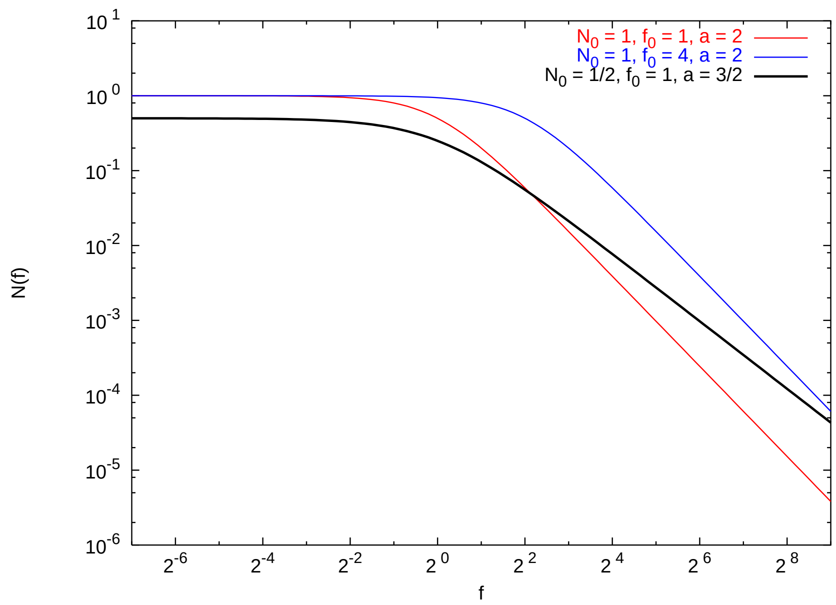


A common form of noise

Frequently, the power spectrum of the noise has the form

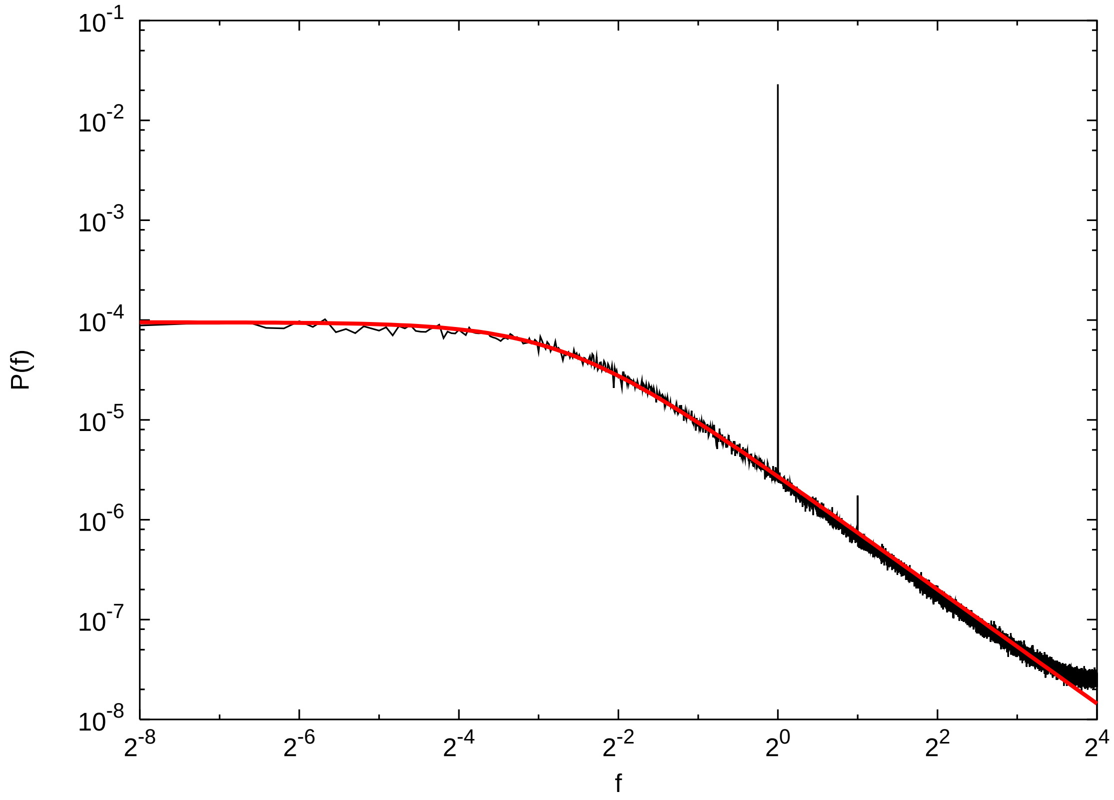
$$N(f) = \frac{N_0}{1 + \left|\frac{f}{f_0}\right|^a} . \tag{21}$$

We estimate N_0, f_0, a by fitting Eq. (21) to "experimental" data.



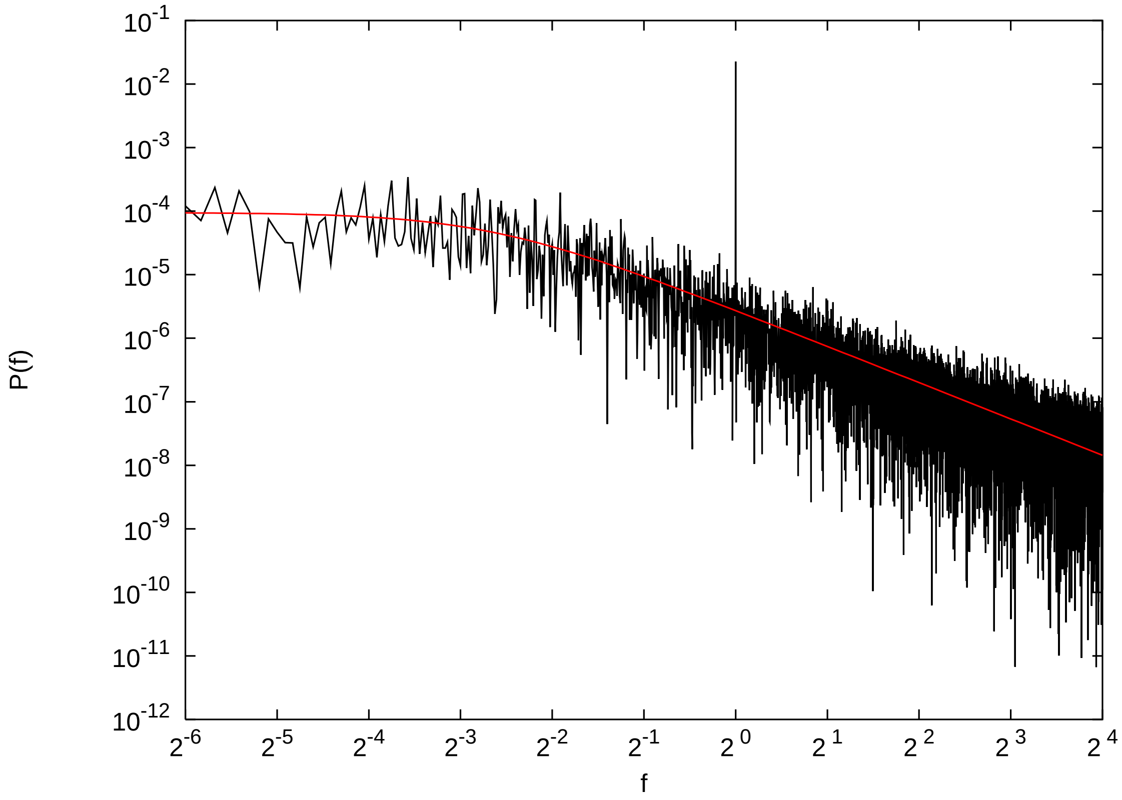
For $|f| \gg 1$, $N(f) \sim |f|^{-a}$. In the log-log plot, the "tail" is a straight line with a slope $-a$. $a > 1$!
If $1 < a < 2$, we have a *fractal noise*.

Example



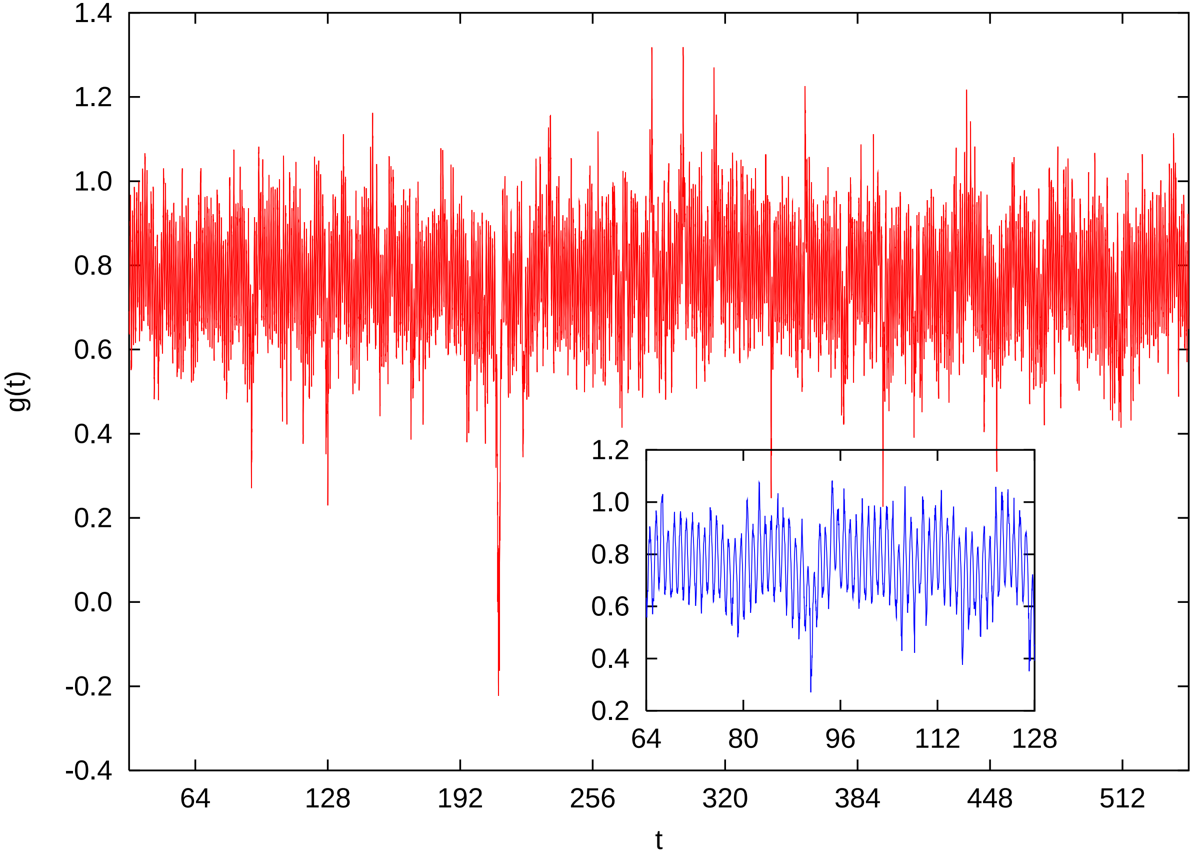
A fitted line (21) with $N_0 = 9.5 \cdot 10^{-4}$, $f_0 = 5/32$, $a = 1.9$.

In practice, the spectra are of *lower* quality. . .

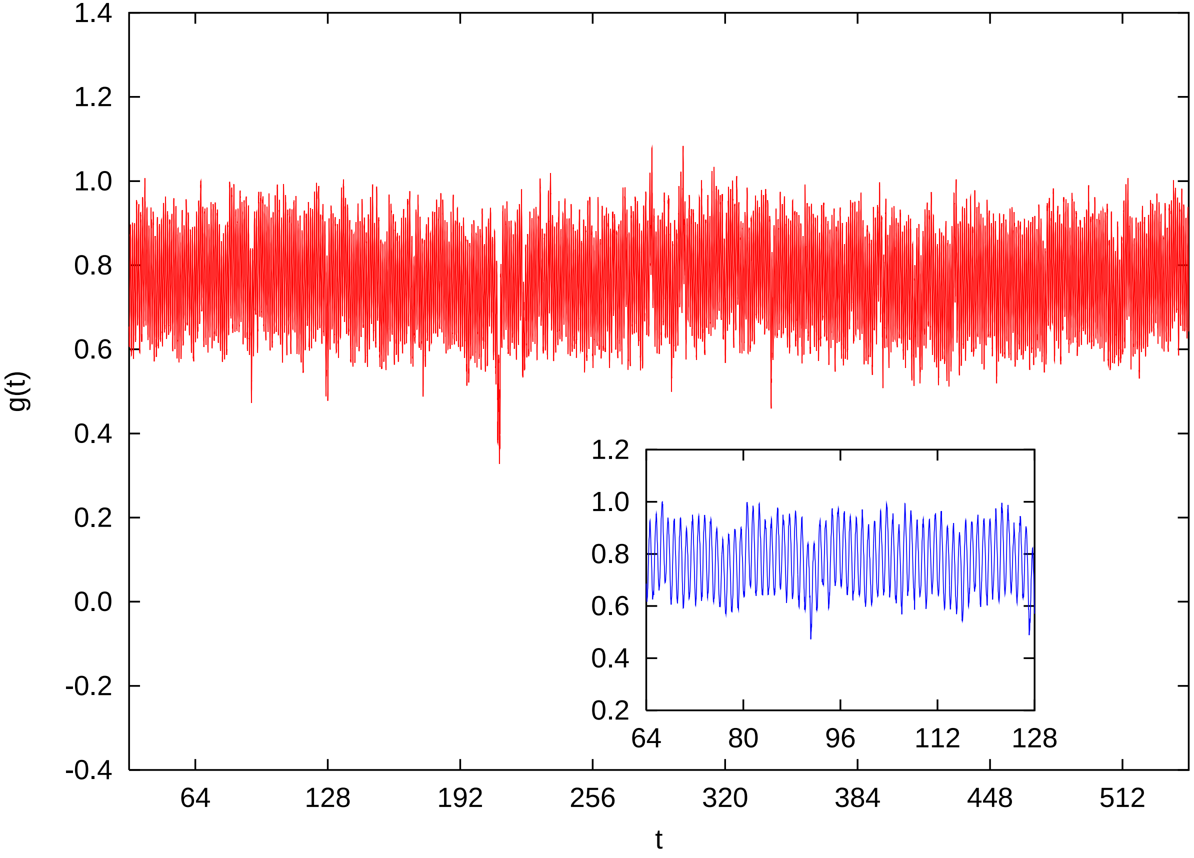


A fitted line (21) with $N_0 = 9.5 \cdot 10^{-4}$, $f_0 = 5/32$, $a = 1.9$.

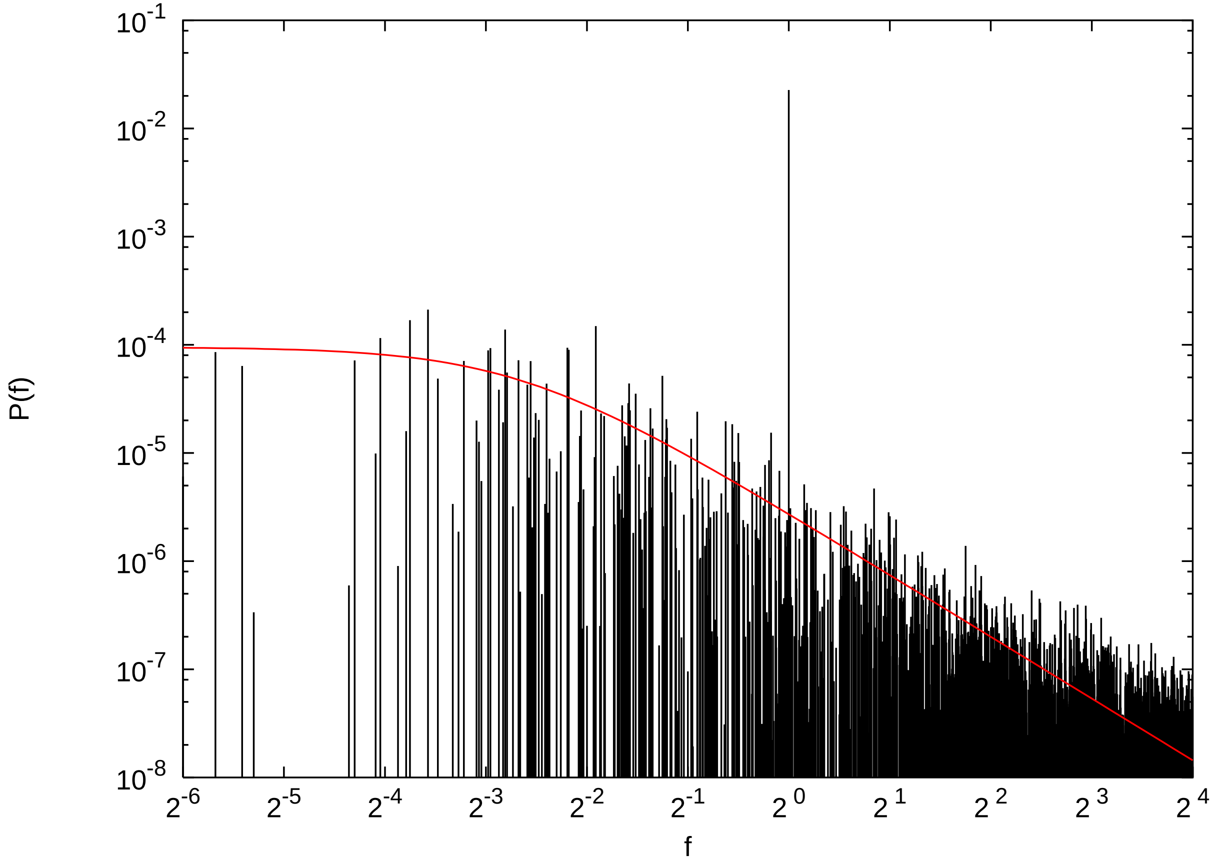
Before filtering



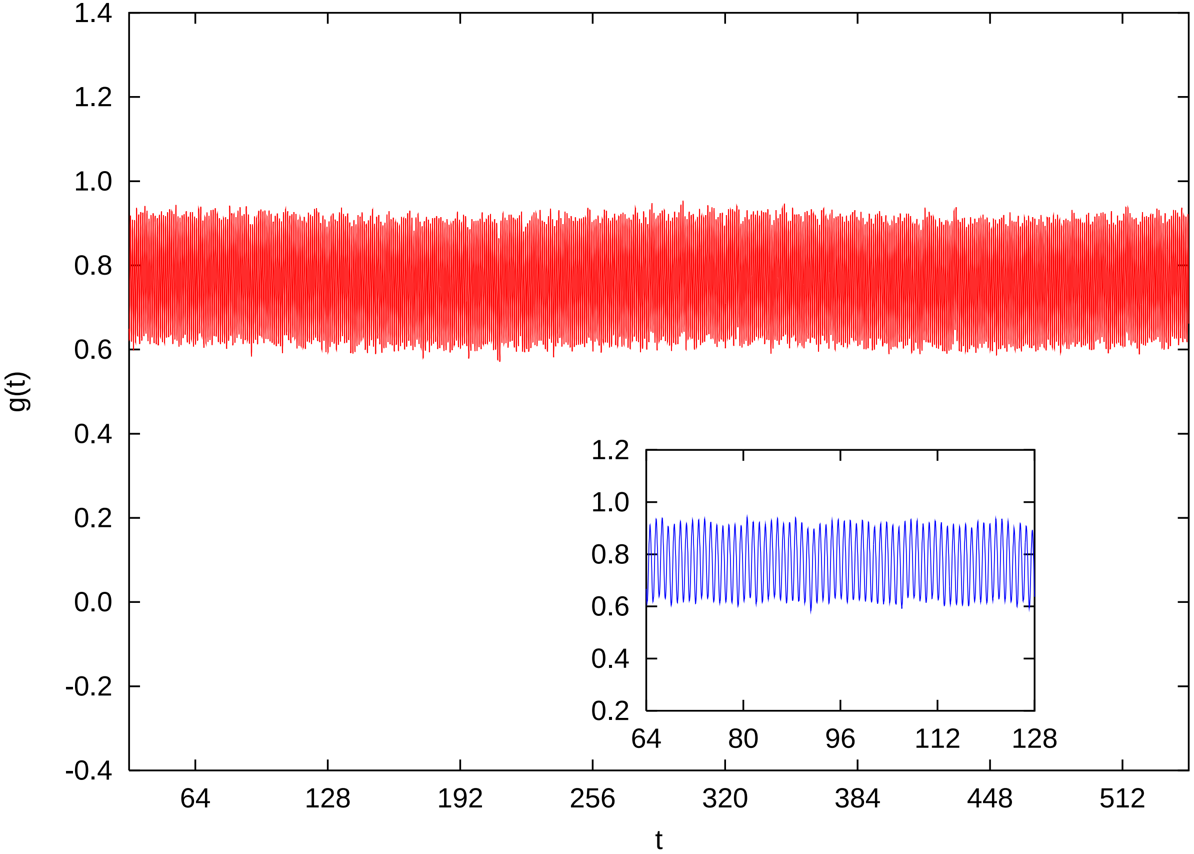
After filtering



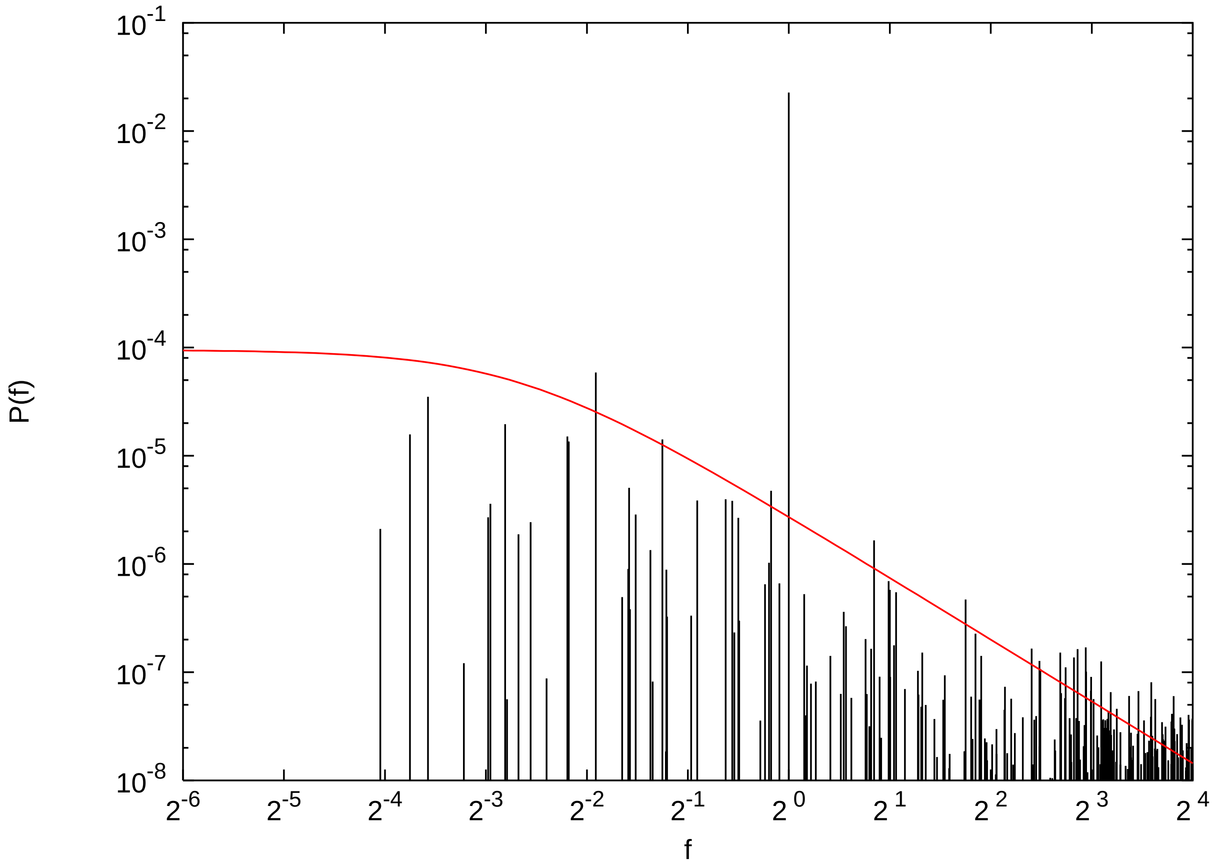
Spectrum after filtering



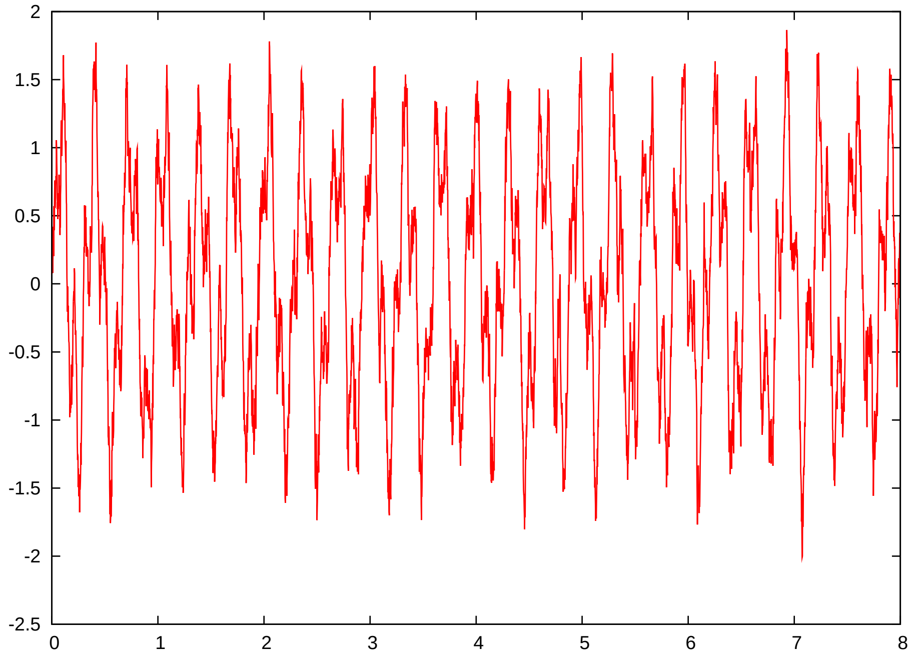
Filtering the filtered: A cascade Wiener filter



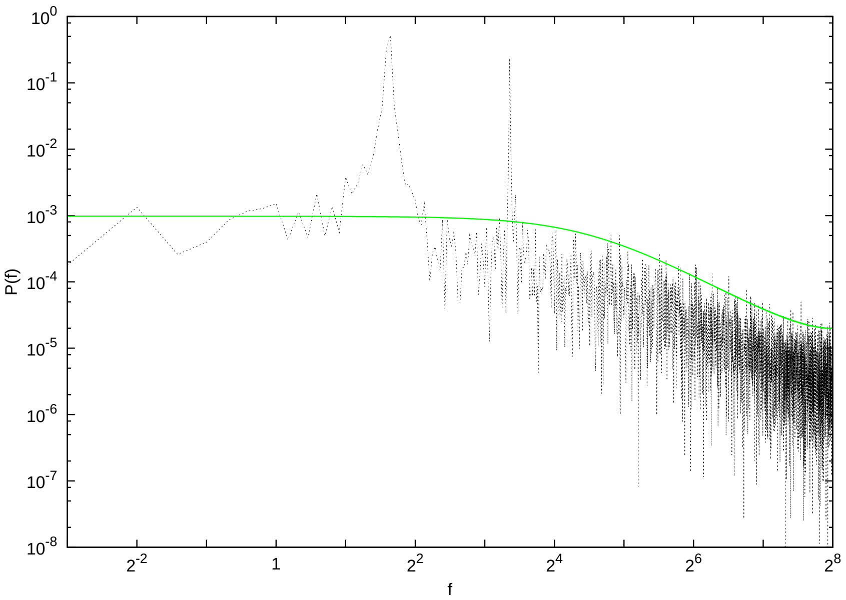
Spectrum after the cascade filter



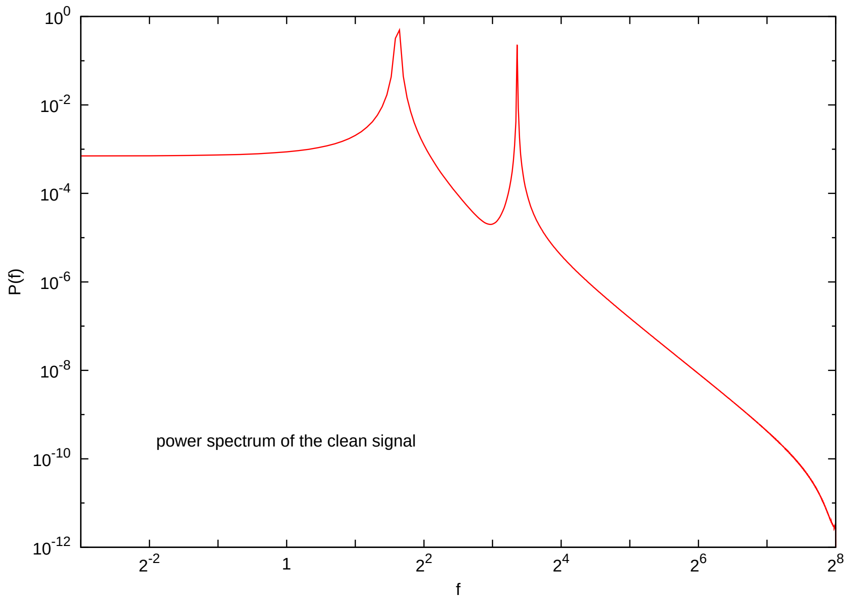
Yet another example



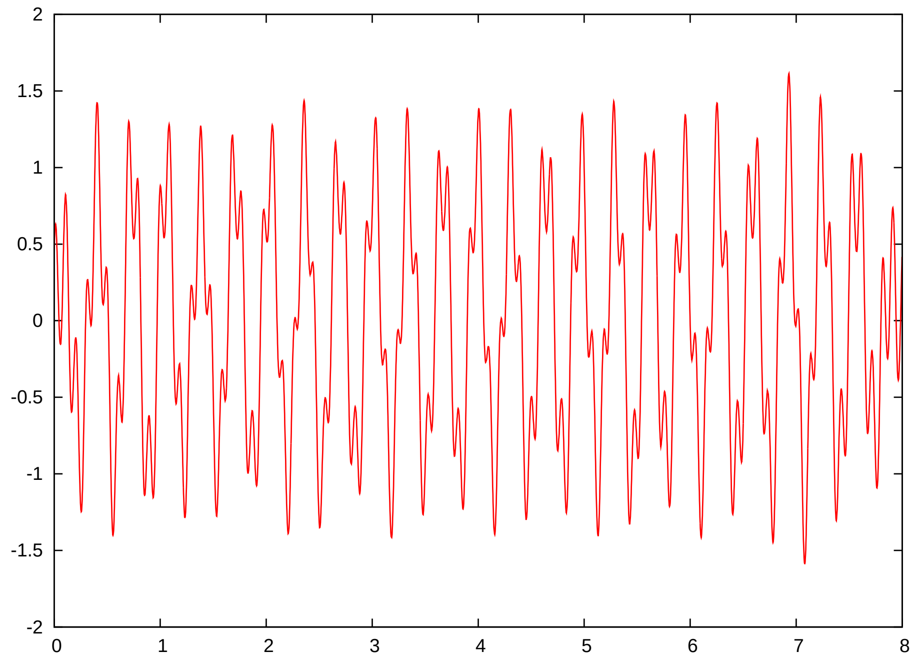
A quasiperiodic signal contaminated by a weak, Gaussian noise.
A quasiperiodic signal is a finite superposition of harmonic waves of incommensurate frequencies.



The power spectrum of the signal



The power spectrum of the *clean* signal



The filtered signal

Wiener filter and non-Gaussian noises

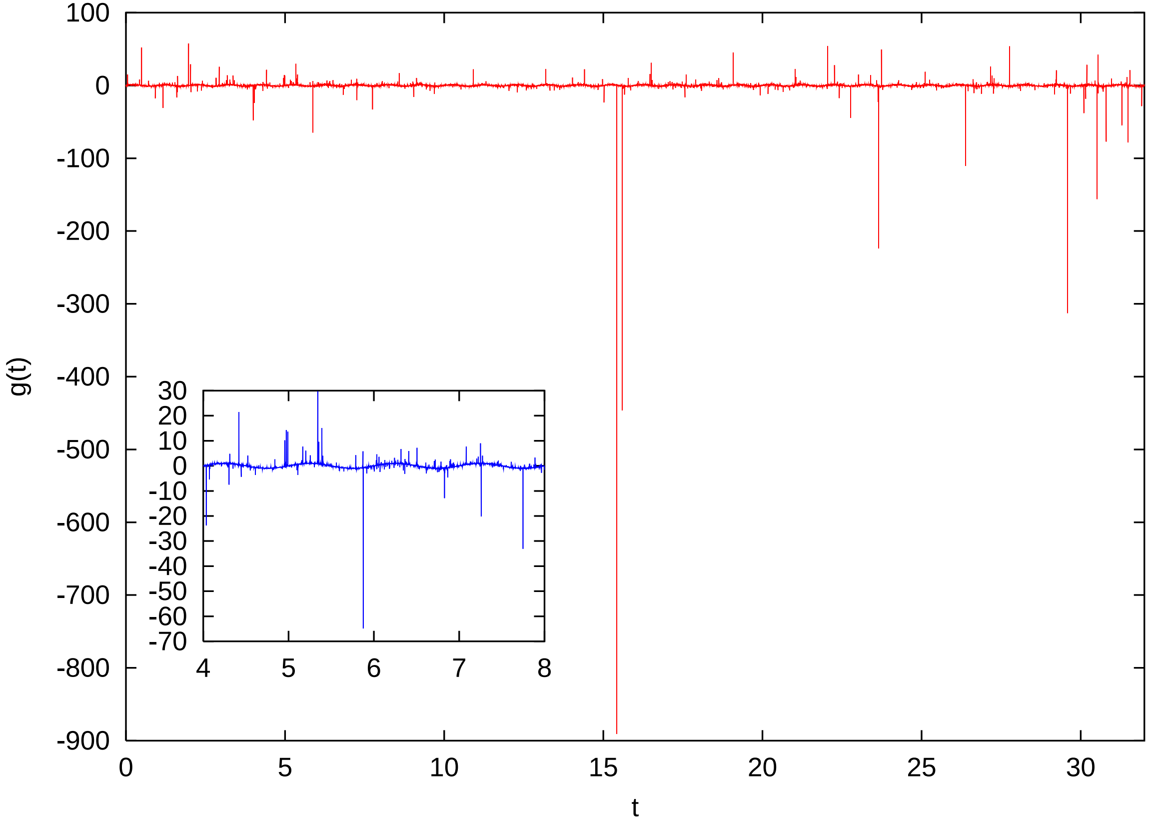
The Wiener filter is optimal under the assumption that the noise is Gaussian. How does the Wiener filter perform for non-Gaussian noises? In particular, how does the filter perform if the second or the first moments of the noise do not exist, or if the noise has "heavy tails"?

We will analyse a process generated with

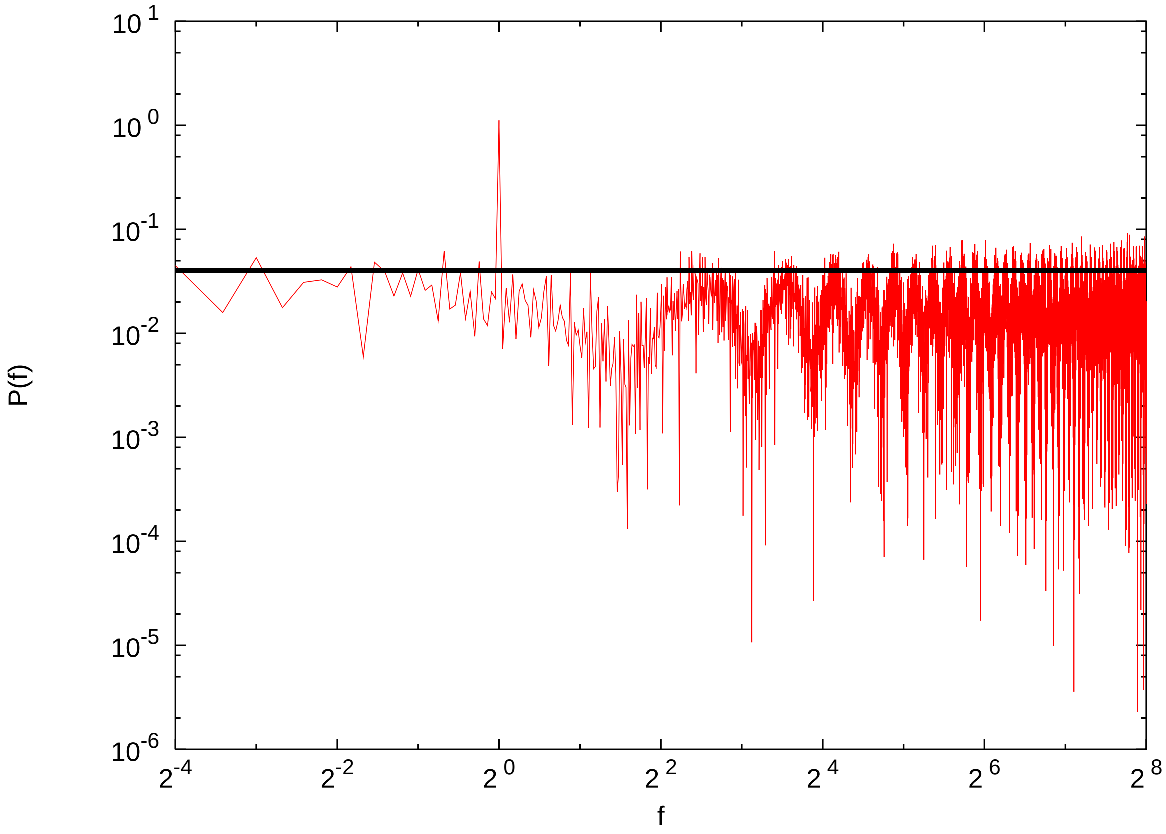
$$g(t) = \sin 2\pi t + \frac{1}{16}\zeta(t)\,, \tag{22}$$

with $\zeta(t)$ drawn from the Cauchy distribution and $\langle\zeta(t_1)\zeta(t_2)\rangle = 0$ for $t_1 \neq t_2$.

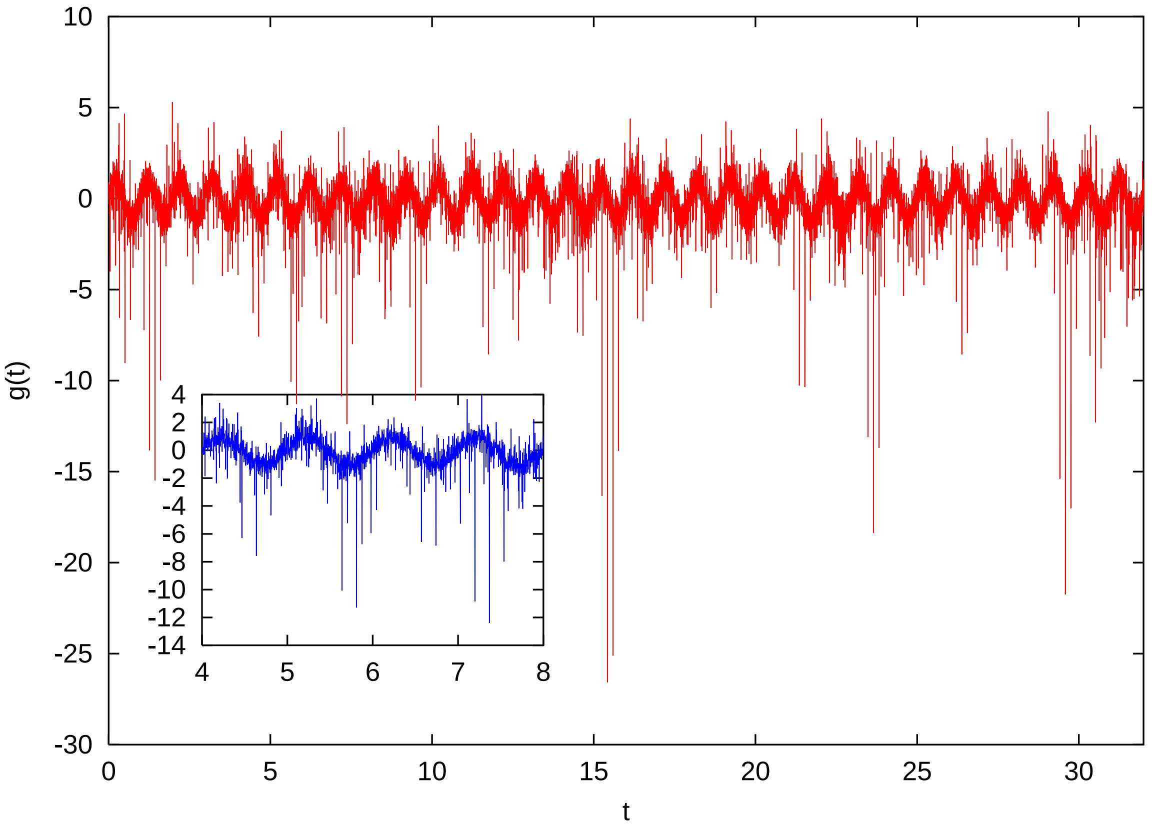
1st realization: The signal before filtering



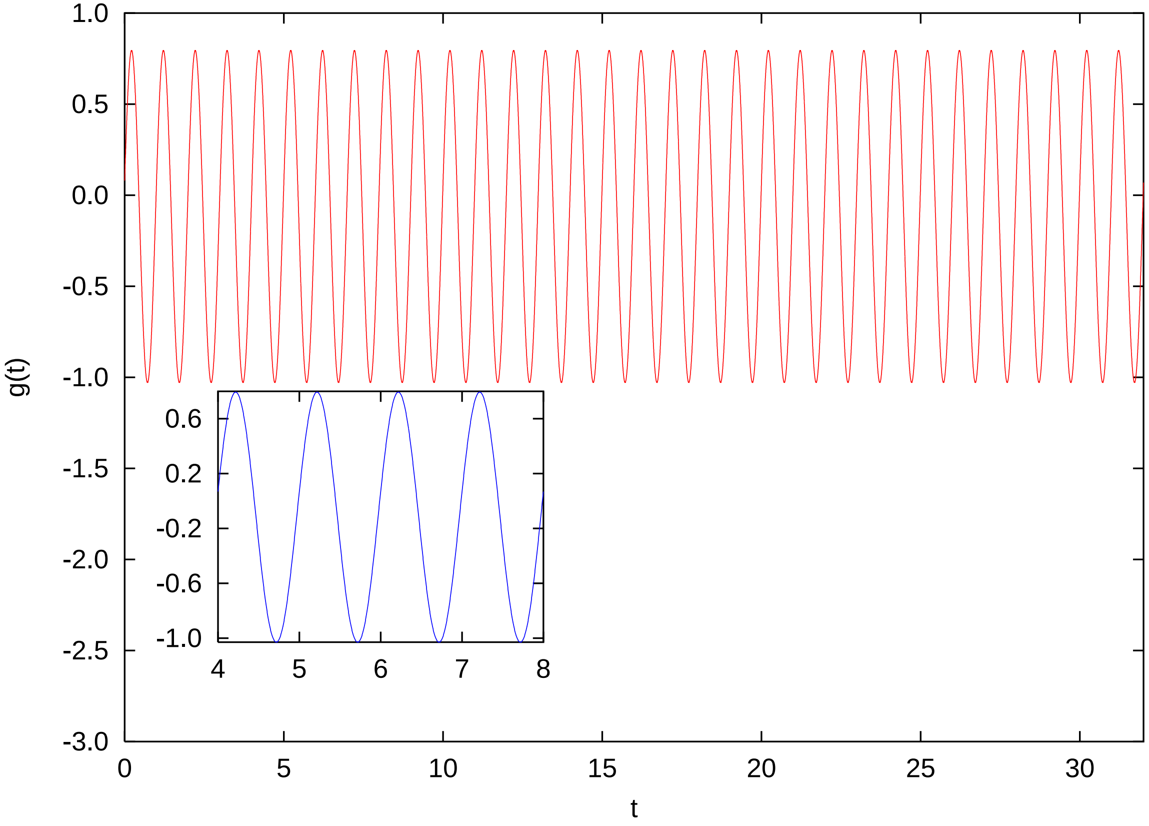
1st realization: The power spectrum of the unfiltered signal



1st realization: After filtering



1st realization: After filtering twice, in a cascade

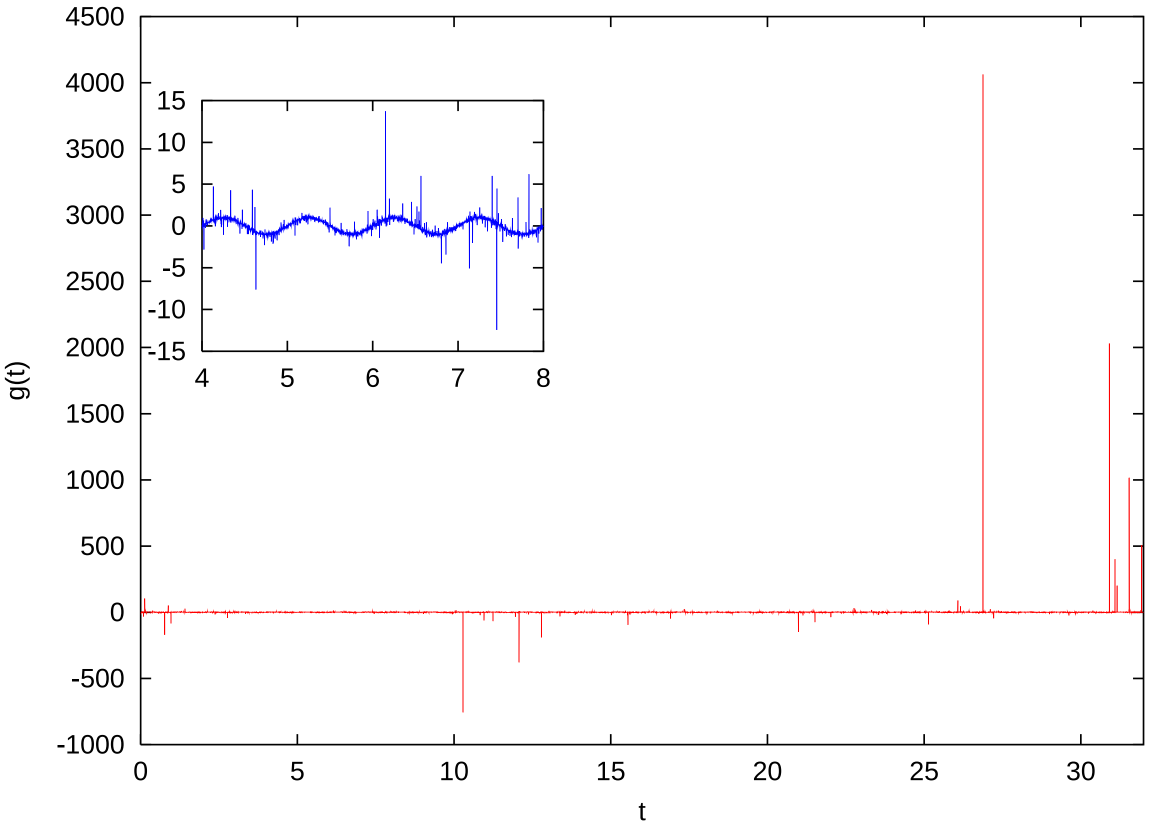


So far, so good.

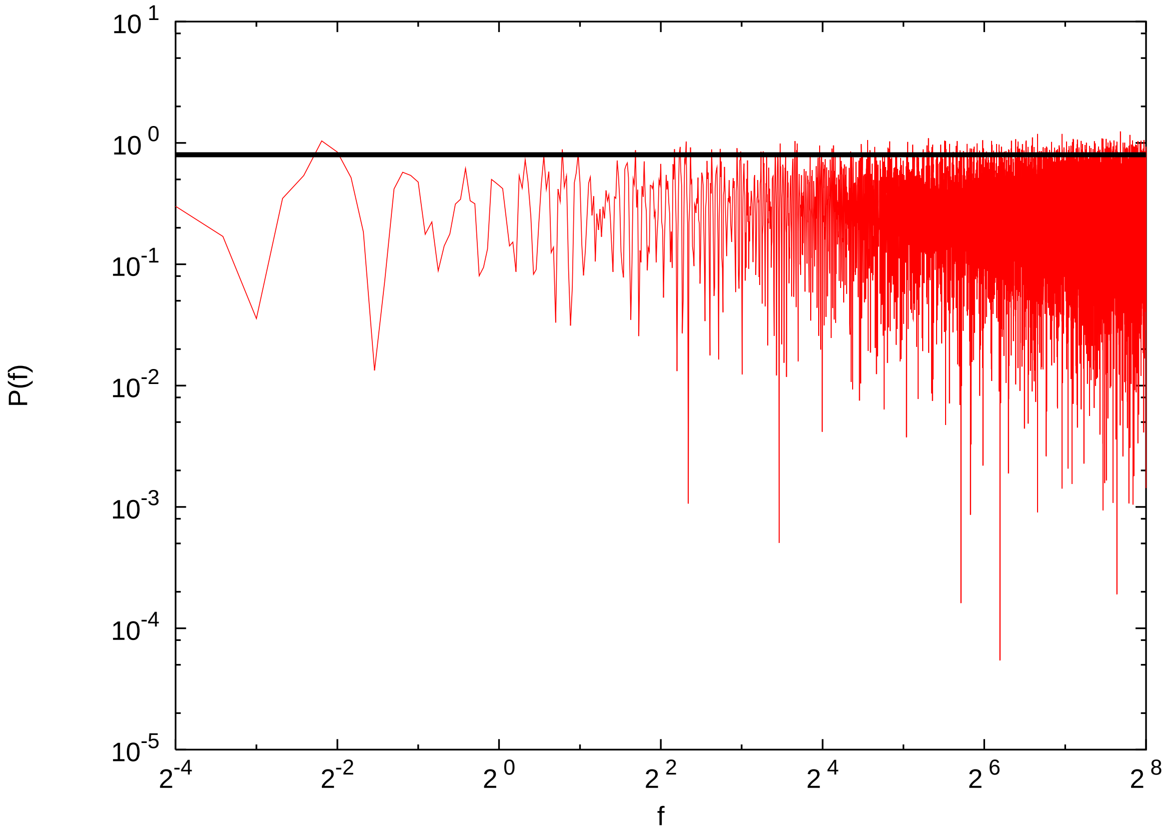
The signal has a unique, and *known*, characteristic frequency. Note that (i) the amplitude of the filtered signal has dropped (the filter is lossy), and (ii) the average of the denoised signal is shifted from zero,

Now, do the same trick with *a different realization of the same process* (22):

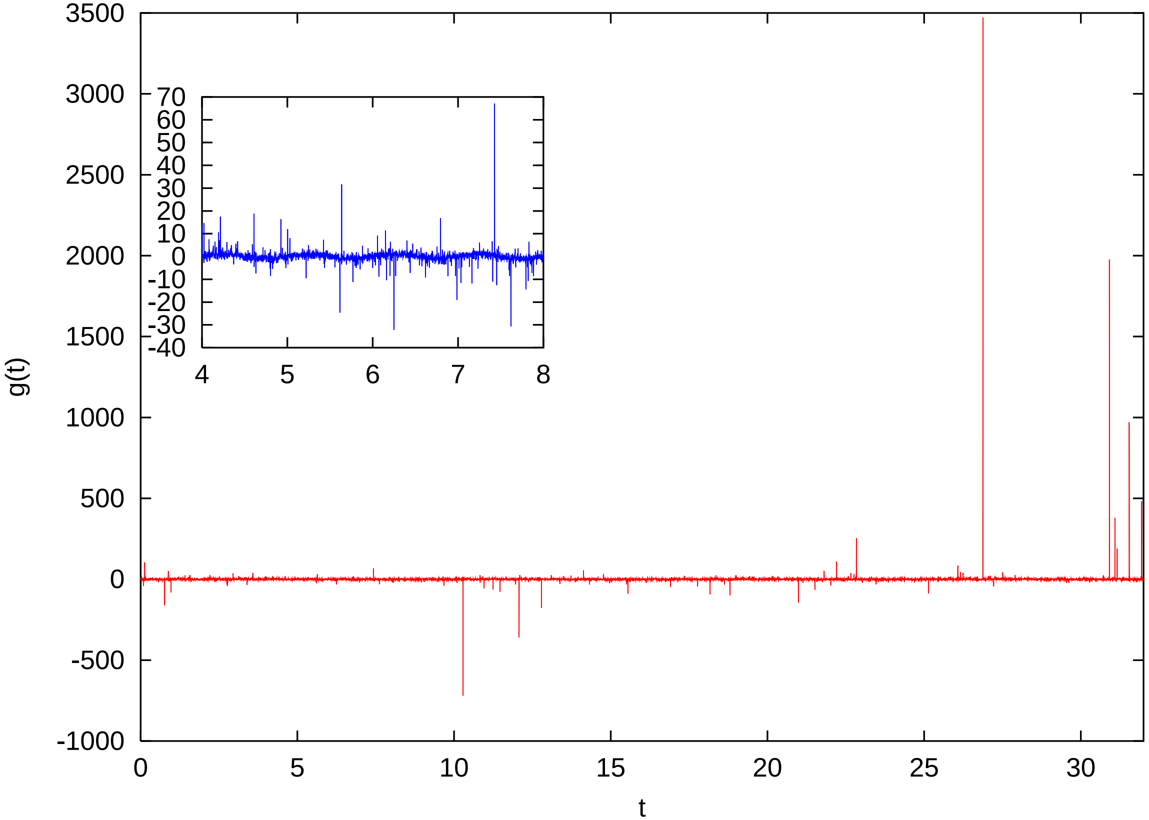
2nd realization: The signal before filtering



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).