Experimental Methods
Lecture 26

November 19th, 2020

Noise estimation

Introduction

- Estimation of noise properties is key for many reasons. Among them:
  - All techniques to extract signal form noise in some optimal or sub-optimal way, require some knowledge of the noise properties of the apparatus.
  - In some experiments, the objective of the measurement might not be a deterministic signal but directly a stochastic process of which we need to estimate the statistical properties
- Estimation of the mean value of a stochastic process bears no difference to signal extraction and has then been already discussed
- Next in importance is then the estimate of second order moments, i.e., of autocorrelation (and cross-correlation). The most common, robust and accurate practice to achieve an estimate of autocorrelation, for a stationary process, is that of estimating the PSD. This is the subject of this section.
- Estimation of autocorrelation for non stationary processes is a sophisticated subject beyond the scope of this course.
- The estimation of higher order moments is only needed for processes that cannot be reasonably treated as Gaussian. Again a subject beyond the scope of the course

A statistical interlude

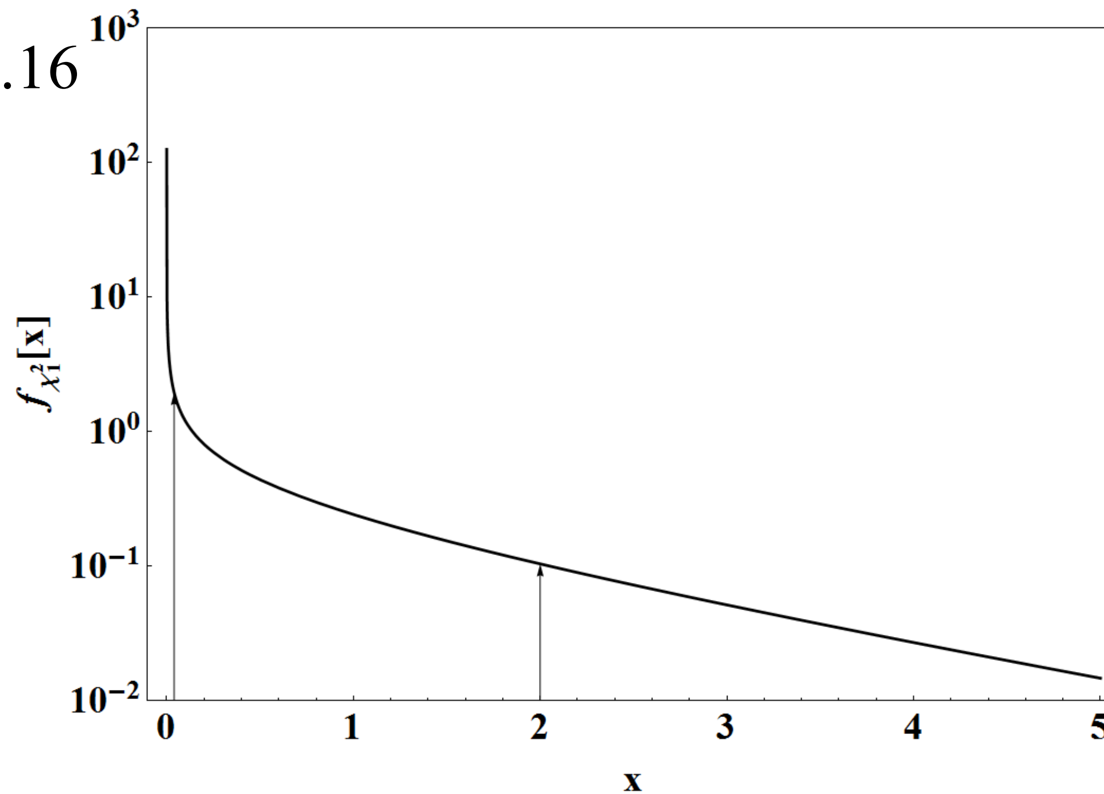
- Noise estimation reduce one way or the other, to the estimate of the fluctuations of some quantity around its mean value.
- In order to estimate the mean amplitude of fluctuations, fluctuating quantities are squared in order to take their sign out. Indeed averaging, for instance, independent fluctuations *with their sign* converges to zero with the square root of the number of sample that have been averaged.
- The most elementary estimation of noise is the estimate of the variance of a zero-mean Gaussian random variable v. Let's discuss the accuracy of using as an estimator the square v^2 of one observed sample of the random variable itself.
- It is useful to remind a few properties of the square of Gaussian variables:
- The square of the sum N random Gaussian variables all with zero-mean and standard deviation σ=1 $\quad \chi_N^2 = \sum_{k=1}^{N} v_k^2$
- is distributed as a chi-square variable with N degrees of freedom. The probability density function (PDF) of such a distribution is:

$$f_{\chi_N^2}(x) = \Theta(x) 2^{-N/2} e^{-x/2} x^{-1+\frac{N}{2}} \Big/ \mathrm{Gamma}(N/2)$$

Properties of chi-square distributions

- Thus PDF: $f_{\chi^2_N}(x) = \Theta(x)\, 2^{-N/2} e^{-x/2} x^{-1+\frac{N}{2}} \Big/ \mathrm{Gamma}(N/2)$
- Mean and standard deviation $\langle \chi^2_N \rangle = N \quad \sigma_{\chi^2_N} = \sqrt{2N}$
- The special case for N=1 $\quad f_{\chi^2_1}(x) = \Theta(x)\, e^{-x/2} \Big/ \sqrt{2\pi x}$
- The tails corresponding to the same probability as that for a Gaussian variable to fall outside the $\pm 1\ \sigma$ interval:

$$\int_0^{0.04} f_{\chi^2_1}(x)\,dx \approx \int_2^{\infty} f_{\chi^2_1}(x)\,dx \approx 0.16$$



Properties of chi-square distributions

- Suppose we have a zero mean random variable v with arbitrary standard deviation $\sigma_{v.}$ The random variable v/σ has unit variance.
- From the calculation in the previous page with 68% probability

$$.04 \le \left(v/\sigma_v\right)^2 \le 2$$

- That is

$$0.5v^2 \le \sigma_v^2 \le 25v^2$$

- If we are interested in the standard deviation

$$0.7\left|v\right| \le \sigma_v \le 5\left|v\right|$$

- Thus estimating the variance from one sample of the square of the random variable itself is very inaccurate!
- Thinks improve if you have many independent samples of the same random variable. See next pages

Increasing the number of degrees of freedom

- Take a "reduced" chi-square $\chi^2_{N,r} = \chi^2_N / N = \sum_{k=1}^{N} v_k^2 / N$
- It is straightforward to calculate its PDF

$$f_{\chi^2_{N,r}}(x) = \Theta(x) N\, 2^{-N/2} e^{-Nx/2} (Nx)^{-1+\frac{N}{2}} \Big/ \mathrm{Gamma}(N/2)$$

- Corresponding to mean value and standard deviation given by

$$\langle \chi^2_{N,r} \rangle = \langle \chi^2_N \rangle / N = N/N = 1 \quad \sigma_{\chi^2_{N,r}} = \left(\sigma_{\chi^2_N} / N\right) = \sqrt{2N}/N = \sqrt{2/N}$$

- By increasing N, the PDF tends to be Gaussian (see next page) so that with 68% probability $\chi^2_{N,r} = 1 \pm \sqrt{2/N}$
- If all the v_k have standard deviation equal to σ_v then $\tilde{v} = \dfrac{1}{\sigma_v^2} \dfrac{\sum_{k=1}^{N} v_k^2}{N}$ is distributed like a reduced chi square. Then with 68% probability

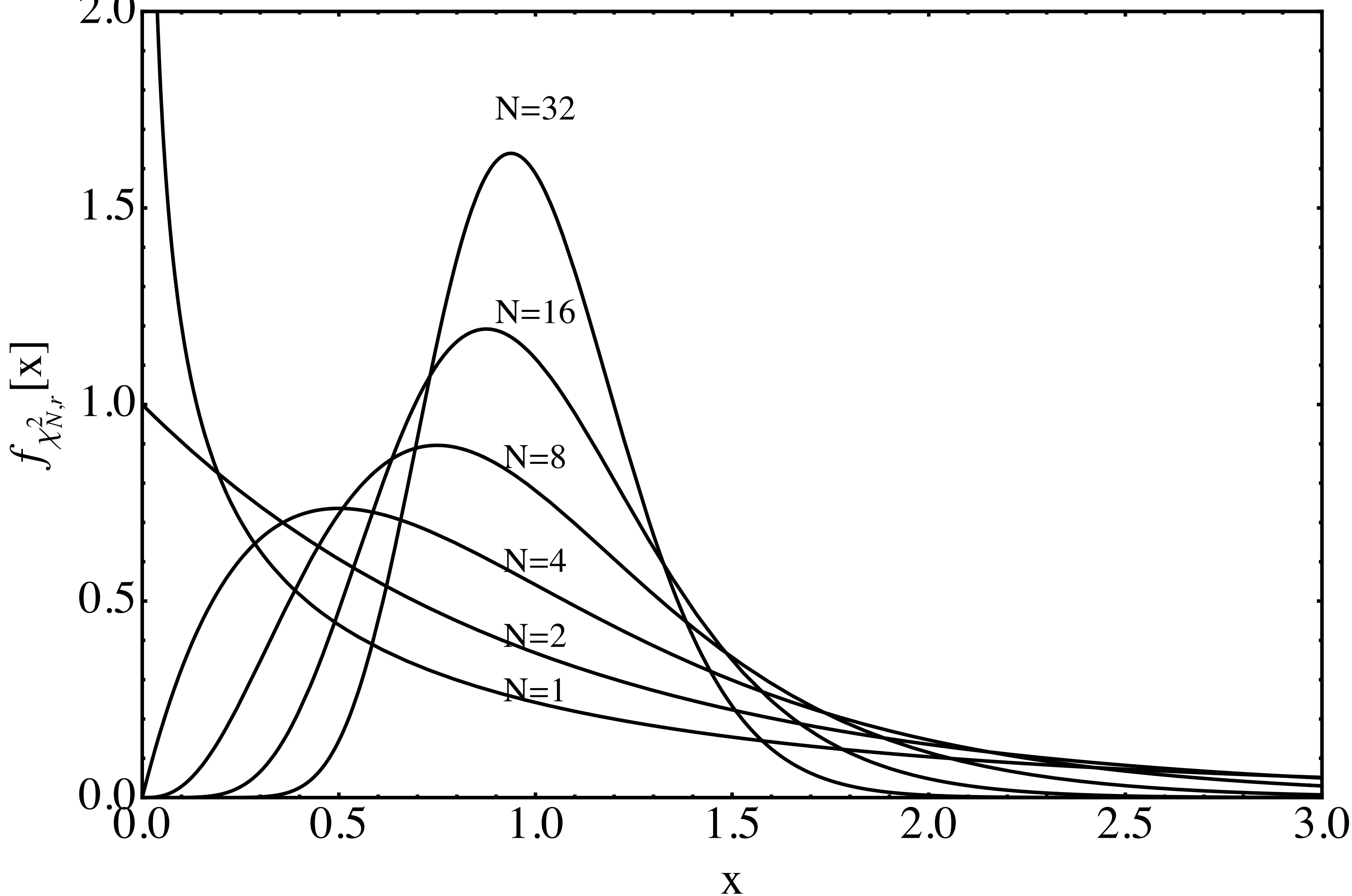
$$\sigma_v^2 \approx \left(\sum_{k=1}^{N} v_k^2 / N\right) \Big/ \left(1 \pm \sqrt{2/N}\right) \approx \left(\sum_{k=1}^{N} v_k^2 / N\right)\left(1 \mp \sqrt{2/N}\right)$$

- A case of interest is that of N=2 in this case

$$\int_0^{0.18} f_{\chi^2_{2,r}}(x)\,dx \approx \int_{1.8}^{\infty} f_{\chi^2_{2,r}}(x)\,dx \approx 0.16$$

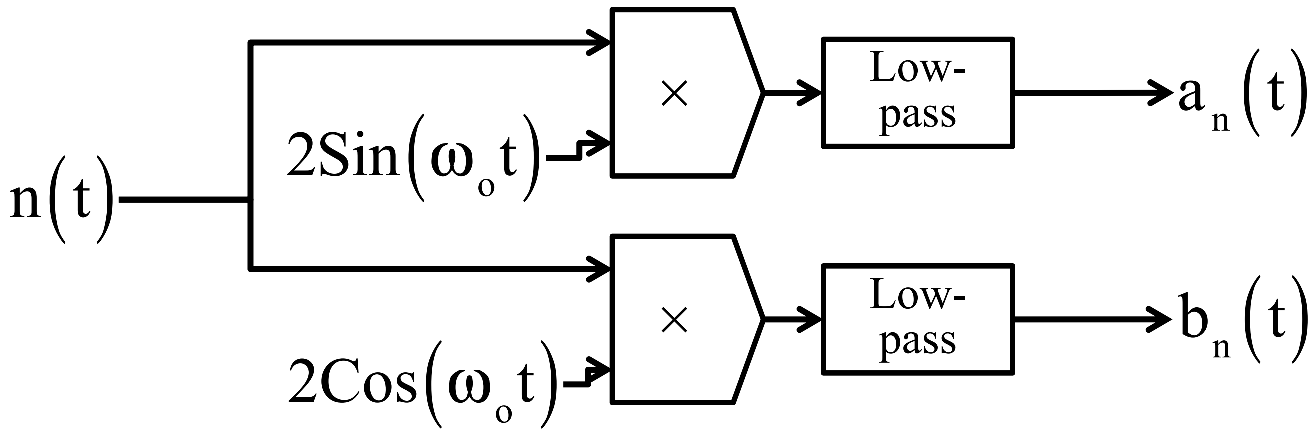
- And the 68% interval is $0.7\sqrt{\left(v_1^2 + v_2^2\right)/2} \le \sigma_v \le 2.4\sqrt{\left(v_1^2 + v_2^2\right)/2}$

Reduced chi-square PDF as a function of the number N of degrees of freedom



PSD estimation

- Now let's turn to PSD estimation. The first method that can be used to estimate the PSD of a stochastic process n(t) makes use of a PhSD



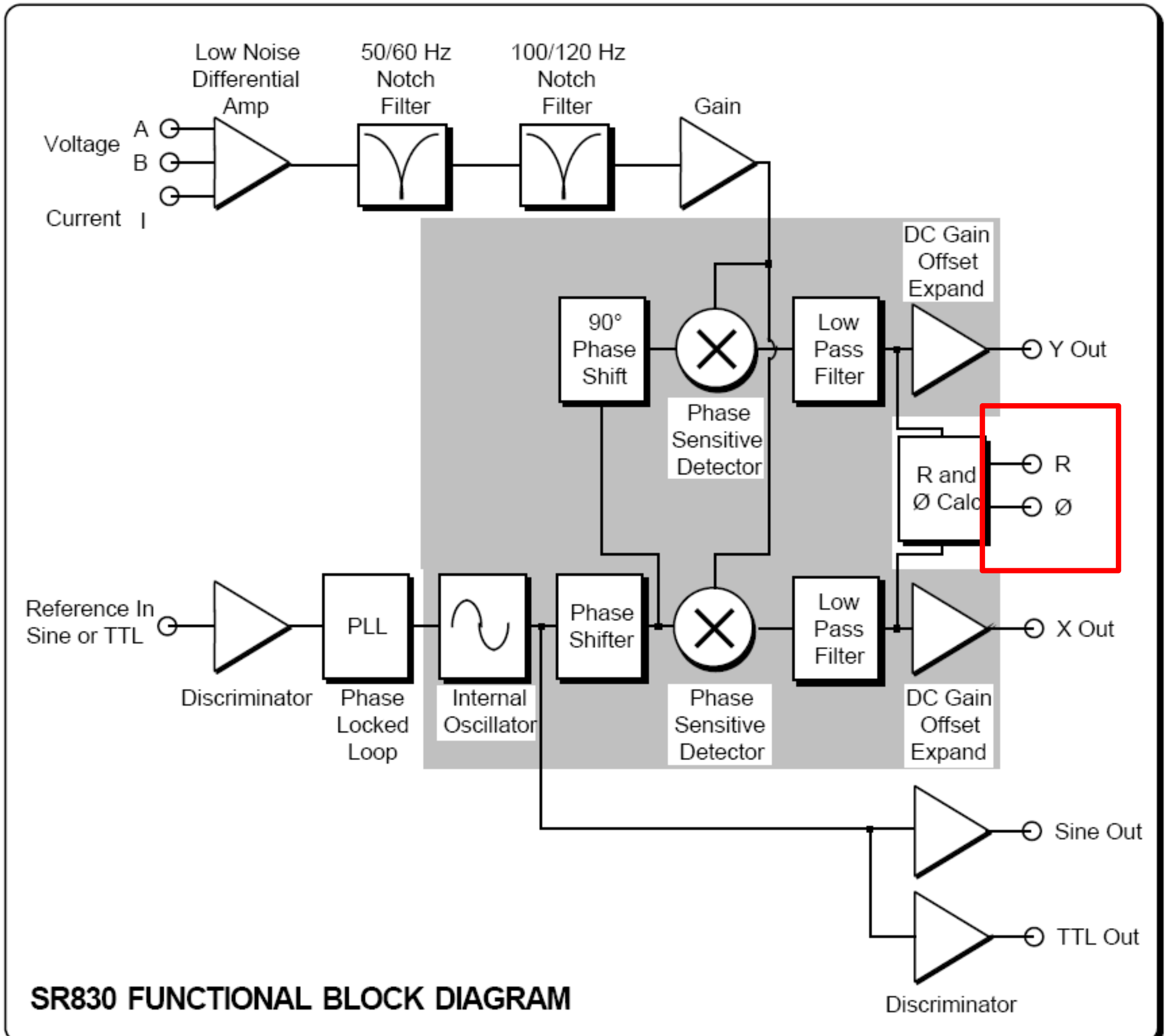
- Indeed, if n(t) is a stationary stochastic process, the two output phases are two zero-mean, uncorrelated stochastic processes both with PSD

$$S_{a_n a_n}(\omega) = S_{b_n b_n}(\omega) \simeq 2\left|H(\omega)\right|^2 S_{nn}(\omega_o)$$

- It follows that $\left\langle a_n^2 \right\rangle = \left\langle b_n^2 \right\rangle \simeq 2S_{nn}(\omega_o)\left\{(1/2\pi)\int_{-\infty}^{\infty}\left|H(\omega)\right|^2 d\omega\right\}$
- According to chi-square statistics, from a sample for both $a_n(t)$ and $b_n(t)$, the following quantity is a reduced chi-square with 2 degrees of freedom

$$a_n^2(t) + b_n^2(t) \Big/ 4S_{nn}(\omega_o)\left\{(1/2\pi)\int_{-\infty}^{\infty}\left|H(\omega)\right|^2 d\omega\right\}$$

The amplitude channel in the lock-in amplifier



PSD estimation

- As the following quantity is a reduced chi-square with 2 degrees of freedom

$$a_n^2(t)+b_n^2(t)\Big/4S_{nn}(\omega_o)\left\{(1/2\pi)\int_{-\infty}^{\infty}\left|H(\omega)\right|^2 d\omega\right\}$$

- Then

$$2S(\omega_o)=\frac{\dfrac{a_n^2(t)+b_n^2(t)}{2}}{\dfrac{1}{2\pi}\int_{-\infty}^{\infty}|H(\omega)|^2 d\,\omega}(1+4.7/-0.45)$$

which is an estimate of the one-sided spectral density
that can be underestimated by a factor 5 or overestimated by a factor 2 with 68% probability.

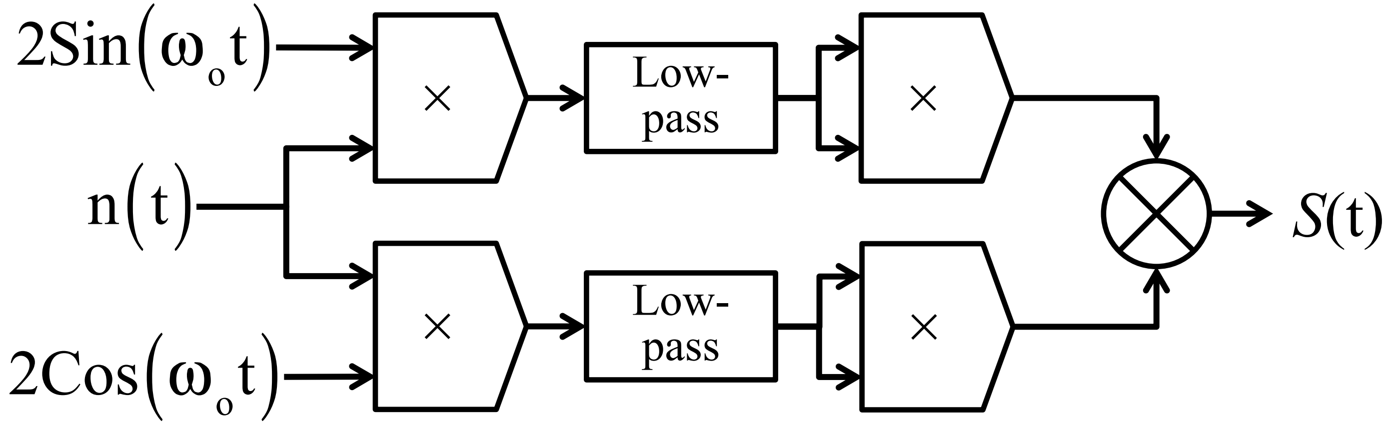
- Notice that in reality $\left\langle a_n^2\right\rangle=(1/2\pi)\int_{-\infty}^{\infty}\left|H(\omega)\right|^2\left\{S_{nn}(\omega-\omega_o)+S_{nn}(\omega+\omega_o)\right\}d\omega$

- So that the procedure gives an estimate of the PSD *averaged over the bandwidth of the filter*

- The estimate requires the knowledge of $\frac{1}{2\pi}\int_{-\infty}^{\infty}|H(\omega)|^2 d\,\omega$
- However the relative error is independent of the filter shape.

The phase sensitive detector as an estimator

- In conclusion one sample of the output S(t) of the following scheme:



  if properly calibrated, is an estimate of the one sided PSD of n(t) *at the carrier frequency ω_o* with a frequency resolution equal to the bandwidth of the output filter and a relative accuracy of less than one order of magnitude.

  To facilitate calculations, we will assume from now on that the filter has been calibrated such that $\frac{1}{2\pi}\int_{-\infty}^{\infty}|H(\omega)|^2 d\,\omega = 1$.

- We are going to discuss in the following the effect of averaging $S(t)$ over some period of time.

Correlation properties of a_n and b_n

- Remember that $S_{a_n,a_n}(\omega) = S_{b_n,b_n}(\omega) = 2|\mathrm{H}(\omega)|^2 \mathrm{S}_{\mathrm{n,n}}(\omega_o)$
- In order to calculate the autocorrelation we have to make assumptions about H(ω). We will take the filter as a simple low pass, properly normalized to fulfill the condition $\frac{1}{2\pi}\int_{-\infty}^{\infty}|H(\omega)|^2 d\,\omega = 1.$
- Then: $H(\omega) = \frac{(2\,\tau)^{\frac{1}{2}}}{1+i\omega\tau}$ and $S_{a_n,a_n}(\omega) = S_{b_n,b_n}(\omega) = \frac{4\tau \mathrm{S}_{\mathrm{n,n}}(\omega_o)}{1+\omega^2\tau^2}$
- It follows that the autocorrelation of a_n is
- $R_{a_n a_n}(\Delta t) = R_{b_n b_n}(\Delta t) = 2\mathrm{S}_{\mathrm{n,n}}(\omega_o)\, e^{-\frac{|\Delta t|}{\tau}}$

Correlation properties of a_n and b_n

- $R_{a_n a_n}(\Delta t) = R_{b_n b_n}(\Delta t) = 2\mathrm{S}_{\mathrm{n,n}}(\omega_o)\, e^{-\frac{|\Delta t|}{\tau}}$
- Thus data points spaced by Δt of order a few times τ will be approximately independent.
- Actually a more quantitative criterion for quasi-independence might be that the conditional variance $\langle a_n^2(t+\Delta t)|a(t)\rangle = 2\mathrm{S}_{\mathrm{n,n}}(\omega_o)\,\left(1 - e^{-\frac{2|\Delta t|}{\tau}}\right)$ is within some small factor from the unconditional one $2\mathrm{S}_{\mathrm{n,n}}(\omega_o)$.
- For instance, with $\Delta t = \tau$, the conditional variance is 13% smaller than the unconditional one, and the standard deviation within 7% of the unconditional one. For *noise* estimation these are pretty decent approximations!

Averaging

- Let's take now N samples of both a_n and b_n spaced by $\Delta t = \tau$
- In a measurement of duration T there will be approximately $N = \frac{T}{\tau}$ samples, the average $\frac{1}{4NS(\omega_o)}\sum_{k=1}^{N}(a_{n,k}^2+b_{n,k}^2)$ is a reduced chi-square with 2N degrees of freedom.
- Then, for large N

$$\frac{1}{4NS(\omega_o)}\sum_{k=1}^{N}(a_{n,k}^2+b_{n,k}^2) \equiv \frac{\overline{S}}{2S(\omega_o)} = 1 \pm \frac{1}{\sqrt{N}} = 1 \pm \frac{1}{\sqrt{T/\tau}}$$

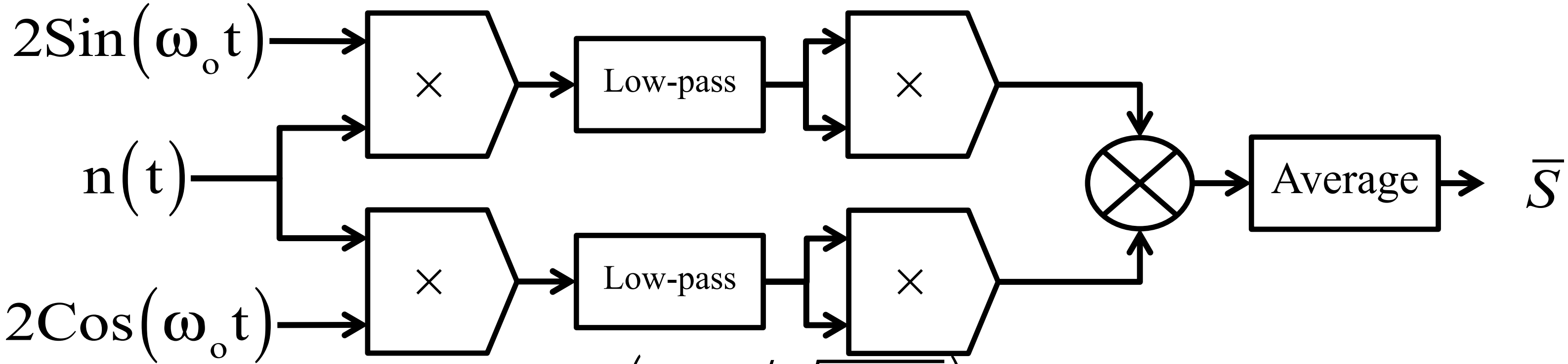
- Now consider that if you write the filter frequency response as $\mathrm{H}(\omega) = \frac{\sqrt{(2\Delta\omega)}}{1+i\omega/\Delta\omega}$, you realize that $\frac{1}{\tau} \equiv \Delta\omega$ is the bandwidth of your low-pass filter (and then also you spectral resolution). Then

$$2\,S(\omega_o) = \overline{S}(1 \pm \frac{1}{\sqrt{T\Delta\omega}})$$

- In conclusion this method gives an estimate of the one sided PSD at the carrier frequency ω_o with a relative error $1/\sqrt{T\Delta\omega}$

The basic scheme for a spectrum analyzer

Thus in summary the following is the ideal scheme for a spectrum analyzer:



with $$2S_{nn}(\omega_o) \approx \bar{S}\left(1 \pm 1/\sqrt{\Delta\omega T}\right)$$

The relative error on this estimate is then (Dicke's radiometric law)

$$\Delta\bar{S}/\bar{S} \approx 1/\sqrt{\text{Spectral resolution} \times \text{Duration of measurement}}$$

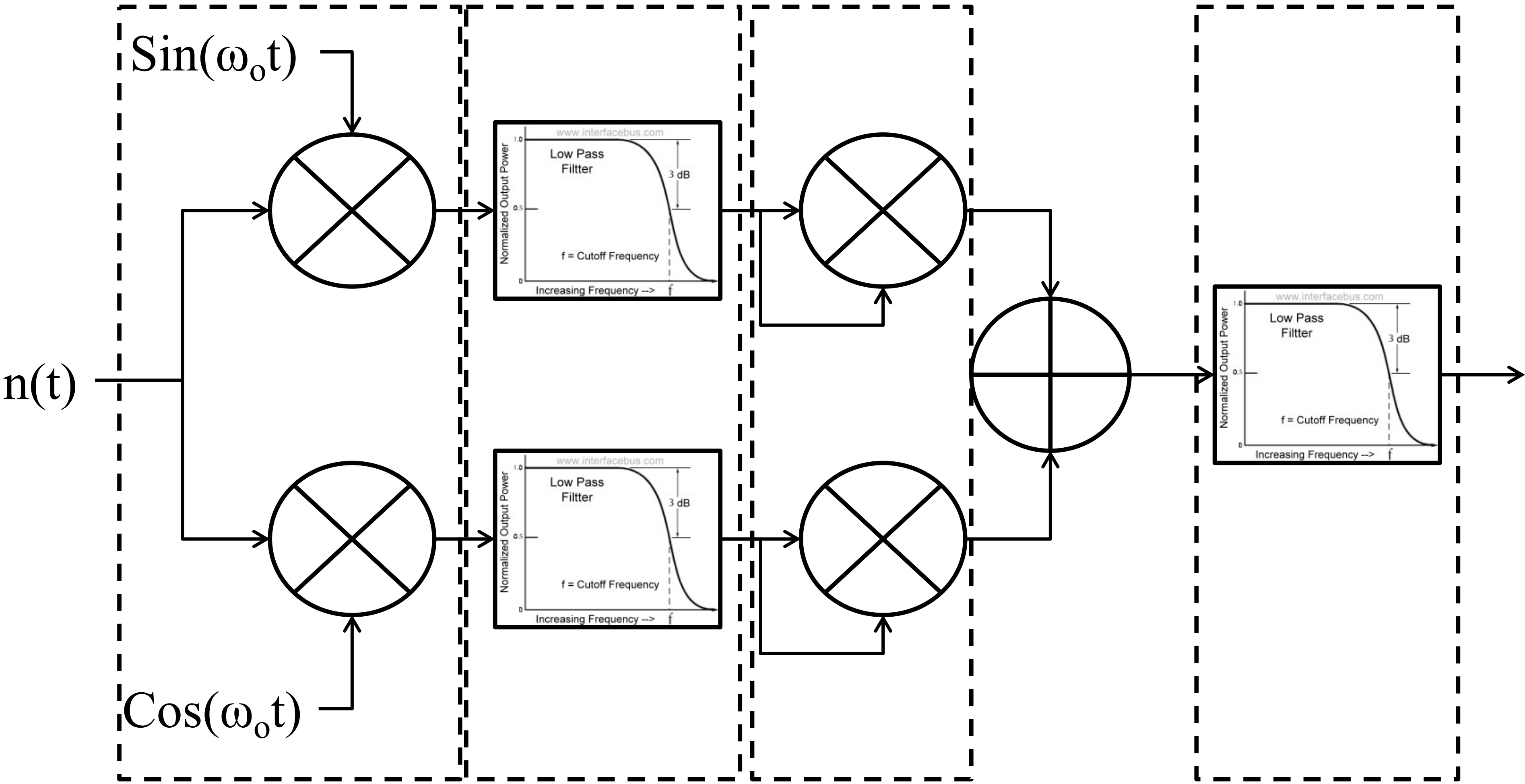
Notice that

The formula contains some approximation. Often one uses the more cautious formula (f is frequency) $2S_{nn}(\omega_o) \approx \bar{S}\left(1 \pm 1/\sqrt{\Delta f T}\right)$

The low pass filter of the PhSD (averaging before squaring) sets the spectral resolution

The integration time (averaging after squaring) sets the precision

A general scheme for PSD estimation



sets frequency of estimation

sets frequency resolution

squares noise

sets precision

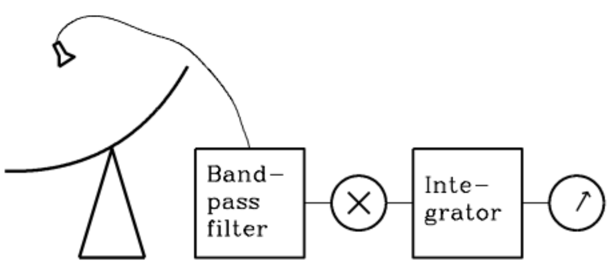
$$\frac{\Delta S}{S} \simeq \frac{1}{\sqrt{\Delta f\ T}}$$

An interesting application: radiometric measurements

- Radio telescopes measure PSD of celestial micro-wave sources
- Basic scheme is the same as for our spectral analyzer
- Relative error is predicted to be $\Delta\bar{S}/\bar{S} \approx 1/\sqrt{\Delta f T}$
- However the microwave signal is the sum of the one due to the source plus the environment back-ground, plus the electronics etc.

$$V_{\mu w}^{tot}(t) = V_{\mu w}^{source}(t) + V_{\mu w}^{background}(t)$$

- The contributions can be separated by pointing to empty sky



PARKES RADIO TELESCOPE (CSIRO / ATNF)

The simplest radiometer filters the broadband noise coming from the telescope, multiplies the filtered voltage by itself (square-law detection), smoothes the detected voltage, and measures the smoothed voltage. The function of the detector is to convert the noise voltage, which has zero mean, to noise power, which is proportional to the square of voltage.

An interesting application: radiometric measurements

- Thus, indicating with “on” the PSD measured while pointing at the source and “off” that collected from empty sky, we have

$$PSD_{\mu w}^{on} = PSD_{\mu w}^{source} + PSD_{\mu w}^{background}; \quad PSD_{\mu w}^{off} = PSD_{\mu w}^{background}$$

- And the contribution due to the source is $PSD_{\mu w}^{source} = PSD_{\mu w}^{on} - PSD_{\mu w}^{off}$
- The error on this estimate can be obtained from the errors over the two separate measurements $\sigma_{PSD_{\mu w}^{source}} = \sqrt{\sigma^2_{PSD_{\mu w}^{on}} + \sigma^2_{PSD_{\mu w}^{off}}}$
- and these are given by

$$\sigma^2_{PSD_{\mu w}^{on}} = \left(PSD_{\mu w}^{source} + PSD_{\mu w}^{background}\right)^2 \Big/ \Delta fT; \quad \sigma^2_{PSD_{\mu w}^{off}} = \left(PSD_{\mu w}^{background}\right)^2 \Big/ \Delta fT$$

- then

$$\sigma_{PSD_{\mu w}^{source}} = \frac{PSD_{\mu w}^{source}}{\sqrt{\Delta fT}} \sqrt{1 + 2\left(\frac{PSD_{\mu w}^{background}}{PSD_{\mu w}^{source}}\right) + 2\left(\frac{PSD_{\mu w}^{background}}{PSD_{\mu w}^{source}}\right)^2}$$

- or

$$\frac{\sigma_{PSD_{\mu w}^{source}}}{PSD_{\mu w}^{source}} = \frac{1}{\sqrt{\Delta fT}} \sqrt{1 + 2\left(\frac{PSD_{\mu w}^{background}}{PSD_{\mu w}^{source}}\right) + 2\left(\frac{PSD_{\mu w}^{background}}{PSD_{\mu w}^{source}}\right)^2}$$

An interesting application: radiometric measurements

There are two interesting limiting case for the relative error

$$\frac{\sigma_{PSD_{\mu w}^{source}}}{PSD_{\mu w}^{source}} = \frac{1}{\sqrt{\Delta f T}}\sqrt{1+2\left(\frac{PSD_{\mu w}^{background}}{PSD_{\mu w}^{source}}\right)+2\left(\frac{PSD_{\mu w}^{background}}{PSD_{\mu w}^{source}}\right)^2}$$

Depending on the value of the signal to noise ratio

$$SNR = PSD_{\mu w}^{source} \Big/ PSD_{\mu w}^{background}$$

They are

$$SNR >> 1 \rightarrow \frac{\sigma_{PSD_{\mu w}^{source}}}{PSD_{\mu w}^{source}} = \frac{1}{\sqrt{\Delta f T}} \quad SNR << 1 \rightarrow \frac{\sigma_{PSD_{\mu w}^{source}}}{PSD_{\mu w}^{source}} = \frac{\sqrt{2}}{SNR\sqrt{\Delta f T}}$$

Thus at high signal to noise ratio detectability implies

$$\sqrt{\Delta f T} > 1$$

While for low signal to noise ratio it implies

$$\sqrt{\Delta f T} > \frac{\sqrt{2}}{SNR}$$

Which is a bigger number