7 Eigenschaften linearer Messgeräte

7.1 Why use technical measuring instruments?

The primary and natural measuring instruments that a physicist can use are the senses. For example, colors, shapes, and patterns can be recognized with the bare eye, acoustic frequencies can be distinguished with the ears, and temperatures can be sensed with the skin. Similarly, we can use our body to change the experimental setting of such simple measurements. For example, when we face the sun and try to identify the color of an object, we may use our hand to produce a shadow covering our eyes in order to recognize the colors better. Putting one hand behind one ear, we may be able to distinguish particular acoustic sounds better from a noisy background.

However, the history of science has brought about numerous man-made technical measuring instruments and measurement setups that have a number of advantages for the user. For example, a thermometer measures temperature consistently on a well-defined temperature scale that can be calibrated. It is more 'objective' than our sense for temperature, which may depend on the temperature around us and on the current state of our body. A microscope can enhance the human visual sense by magnifying small objects. A voltmeter extends the human senses by measuring a quantity that is unaccessible for his senses otherwise. An analog-to-digital converter (ADC) connected to a computer may record long time traces of a particular quantity of interest. A measuring instrument can often take measurements faster and cheaper than a human. Technical measurement equipment can also facilitate reliable measurements by making sure by construction that certain experimental settings are repeadedly the same. For example, measuring a resistance with a mutimeter, we can safely assume that the applied voltage will be the same in each measurement of the current. Very accurate instruments will even make sure that the electronics inside the instrument is kept at a constant temperature at all times. This is why measuring instruments have become an indispensable part of modern physics experiments.

Looking at a measuring instrument in general, it will have one (or more) inputs and one (or more) outputs. For example, a digital voltmeter may have two connections at the input and a digital display as the output. A counter for cosmic rays may use the volume of the sensor crystal as its input and generate a current signal as the output. Linear measuring instruments will have an output signal that is linearly related to the input signal. In general, an ideal measuring instrument produces an output signal y that has a known functional relationship $y = f(x)$ to the input signal x.

In this sense, also an extended experimental setup can be seen as one single measuring instrument. In this case we have a compound multi-component instrument consisting of

many modular subunits, often also called a measuring system. In this case there may be a multitude of input signals $x_1, x_2, \ldots, x_n$ which we denote by a vector $\mathbf{x}$. Also the output of such a measurement system could consist of a number of values $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, and the relation between input and output is a function

$$\mathbf{y} = f(\mathbf{x}). \tag{36}$$

For example, the detector assemblies including all the read-out electronics used at CERN for modern particle experiments may be seen as gigantic measuring instruments or measuring systems.

In general, measuring instruments are not perfect. They have their limits regarding range, resolution, dynamic range, sensitivity, bandwidth, accuracy and precision. These characteristic quantities of a measuring instrument will be briefly described below.

Before an experimentalist can start to perform measurements, he or she has to get familiar with the characteristic performance of the technical measuring instruments of the experiment. This may at first sight seem unnecessary and cumbersome in situations like in a lab class or even in a research lab with experiments that were carefully designed and are well maintained. The previous student was also able to perform reasonable measurements. However, it is absolutely mandatory to familiarize oneself extensively with the measuring instruments in order to produce reliable measurement results, and to interpret them correctly. Not doing this would be similar to driving without a driving license. For example, would you be willing to trust a scientist, who reports results with a relative precision of 10^{-5}, but fails to see that his measurement equipment had accuracy problems of the order of 10^{-2}? Or would you, as a tax payer, fund a scientist, who claims that he needs expensive measurement equipment with an accuracy on the level of 10^{-8} if he uses it to measure a quantity with a precision of 10^{-2}?

7.2 Range, resolution, sensitivity

Each measuring instrument will be able to measure an input quantity x reliably only in a limited interval of values $[x_{\min}, x_{\max}]$. For example, a digital voltmeter may be made for measuring voltages in the interval between -10 and $+10\,\mathrm{V}$. A thermometer may be able to measure temperatures between -20 and $+50^\circ\,\mathrm{C}$. A ruler may allow you to measure lengths between 0 and 30 cm. This basic property of a measuring instrument is called the *measuring interval* [**?**]. In many cases the measuring interval is limited by saturation of a sensor. This occurs, for example, if an amplifier is overdriven.

Measuring interval

The difference $X = x_{\max} - x_{\min}$ between the upper limit and the lower limit of the measuring interval is called the *range* of the measuring instrument. For example, a volt-

Range of a measuring instrument

meter measuring values in the interval between $-10\,\mathrm{V}$ and $+10\,\mathrm{V}$ has a range of $20\,\mathrm{V}$. The ruler measuring lengths between 0 and 30 cm has a range of 30 cm, and the thermometer measuring temperatures between -20 and $+50°\mathrm{C}$ has a range of $70\,\mathrm{K}$.

Resolution

Within the measuring interval, each measuring instrument will be able to detect changes that are larger than some minimum amount Δx. This minimum amount is called the *resolution* of the measuring instrument. For example, if a thermometer is able to detect temperature changes of $0.1°\mathrm{C}$, but not $0.01°\mathrm{C}$, then its resolution is $0.1°\mathrm{C}$. If a digital voltmeter is able to detect changes in the input voltage that are bigger than $10\,\mathrm{mV}$, then this is its resolution. If we are confident that we can distinguish two objects with a length difference of $1/3\,\mathrm{mm}$ by using a ruler having ticks with a spacing of $1\,\mathrm{mm}$, then we are performing a length measurement with a resolution of $1/3\,\mathrm{mm}$. We will see in chapter 7.3 that the resolution of a measuring instrument may be influenced by its bandwidth. The resolution may also depend on the value of the quantity being measured. An example is a logarithmic amplifier, which provides a better resolution for small signals than for large signals.

The finite resolution of a measuring instrument can have different reasons. Often the components of the instruments are sources of noise at the output, which sets the resolution. In other cases the digital display of the instrument has a finite number of digits that limits the resolution. In digital instruments, the resolution may be limited by the number of bits of an analog-to-digital converter inside the instrument. In case of a particle counter the resolution (the smallest number of particles counted per second) could be limited by dark counts. In case of the ruler, it is our own visual ability that limits the resolution of the length measurement. A scanning force microscope, which is used to image surfaces on sub micrometer length scales, may have a lateral resolution of patterns on the surface of 1 nm, and a vertical resolution of 0.1 nm. Do you know the measurement resolutions of the equipment that you are using in your experiment? If not, look it up in the manuals of the instruments, or try to determine it experimentally.

Sensitivity

The term *sensitivity* of a measuring instrument is not used in a unique way. Some people use it as a synonym for resolution. However, the official use of the term [**?**] is the ratio of the change in output value of a measuring instrument or sensor and the corresponding change in the input value. This definition implies that the change in the input value is larger than the resolution of the measuring instrument. Like the resolution, also the sensitivity can depend on the input value of the measuring instrument.

Dynamic range

Consider a measuring instrument measuring a quantity x within the measurement range $X = x_{\max} - x_{\min}$. Let the resolution of the instrument be Δx. The *dynamic range* r of the

measuring instrument is then defined as

$$r = 10 \log \frac{X}{\Delta x} \, \mathrm{dB},$$

where dB is the abbreviation for decibel. For example, if a voltmeter can measure voltages between $100\,\mu\mathrm{V}$ and $10\,\mathrm{V}$, its dynamic range is

$$10 \log_{10} \frac{10\,\mathrm{V}}{100\,\mu\mathrm{V}} = 50\,\mathrm{dB}.$$

We will see in chapter 7.3 that the dynamic range of a measuring instrument (like the resolution) may depend on its bandwidth.

The dynamic range of a measuring instrument is related to the notion of the *signal-to-noise* ratio. The ratio $X/\Delta x$ is the best signal-to-noise ratio that can be achieved with a particular measuring instrument, if the input signal is noiseless.

Signal-to-noise ratio

7.3 Response functions and Bandwidth

The bandwidth of a measuring instrument is related to its response to time-dependent signals. We will see below that the bandwidth can be related to a time-scale called correlation time τ_c via Fourier transform. In a measurement setup, we often measure a signal as a function time, or of a parameter that is varied in time. This are situations, where the bandwidth and the correlation time are of crucial importance. If a signal $x(t)$ is applied to the input of the instrument, the output will show a corresponding time-dependent response $y(t)$, which depends on the correlation time or the bandwidth of the instrument.

Correlation time

Measurement data are always taken within a finite time span T that we call the *duration of the measurement*. For example, if we measure the dependence of the voltage drop across a resistor on the applied current, we will set the first current value at some time that we call $t = 0$, another current value some time later, and so on until we have reached the last value of the current that we want to set at time T. This time span of the measurement is an important characteristic parameter of the measurement. We say, it is an important *time scale* of the experiment.

Duration of a measurement

Another time scale of a measurement that arises frequently, if we vary a parameter of the measurement such as the current in the previous example, is the time interval Δt between setting two successive current values. Using a computer-controlled experiment, we may set currents I_n at times $n\Delta t$, where $n = 1, \ldots N$, and read corresponding voltage values V_n.

Sampling interval

The time scale Δt is then called the *sampling interval*, and N is the number of acquired data points. Of course, the measurement duration in such a case is $T = N\Delta t$.

It is the interplay of these measurement related time scales T, Δt and τ_c that we want to discuss below.

7.3.1 The response function for analog instruments

We confine our discussion here to the common case of linear time-invariant measuring systems. These systems have the property that an input signal $x(t) = a_1 x_1(t) + a_2 x_2(t)$ will produce an output signal $y(t) = a_1 y_1(t) + a_2 y_2(t)$, where $y_j(t)$ is the output signal that is produced by the input signal $x_j(t)$ (linearity). Time invariance implies that an input signal $x(t - T)$ produces an output signal $y(t - T)$, if $x(t)$ produces $y(t)$. The response to a general time-dependent signal is then given by

$$y(t) = \int_{-\infty}^{\infty} h(t - t')x(t')dt' = \int_{-\infty}^{t} h(t - t')x(t')dt'. \tag{37}$$

This relation is a convolution integral. The response function $h(t - t')$ is zero for $t' > t$ in order to ensure causality. We will see below that $h(t)$ is the so-called impulse response function. It may contain time-related parameters which are then called the characteristic time scales of the measuring instrument.

7.3.2 The response function for discrete sampling

If data is acquired point by point after discrete time intervals Δt, we have $x_n = x(n\Delta t)$ and $y_n = y(n\Delta t)$ with n being an integer number. The discrete relation between input and output corresponding to eq. (37) is then

$$y_n = \sum_m h_{n-m} x_m,$$

where causality requires that $h_{n-m} = 0$ for $m > n$. We may write this relation in matrix notation by defining the matrix $\mathcal{H}$ with elements $\mathcal{H}_{n,m} = h_{n-m}$, and by regarding the x_n and y_n as components of vectors $\mathbf{x}$ and $\mathbf{y}$, respectively. The relation, which is the discrete analogue to eq. (37) then reads

$$\mathbf{y} = \mathcal{H}\mathbf{x}. \tag{38}$$

As a result of causality, $\mathcal{H}$ is a matrix with nonzero elements only on or below the main diagonal. The matrix elements are the same on, and along each secondary diagonal below the main diagonal.

7.3.3 Impulse response function

We return to the continuous case. If $x(t) = \delta(t - t_0)$, i.e., a unit impulse at time t_0, the output signal will be $y(t) = h(t - t_0)$, where $h(t)$ is called the impulse response function. An example for the impulse response function of a measuring instrument that shows an ideal first order low-pass filter behavior is shown in Fig. 32. Its functional form is

$$h(t) = \frac{1}{\tau} e^{-t/\tau} \theta(t), \tag{39}$$

where

$$\theta(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } t \geq 0 \end{cases}$$

is Heaviside's step function. The time scale τ, called the *response time*, represents a characteristic time scale of the measuring instrument.

Response time

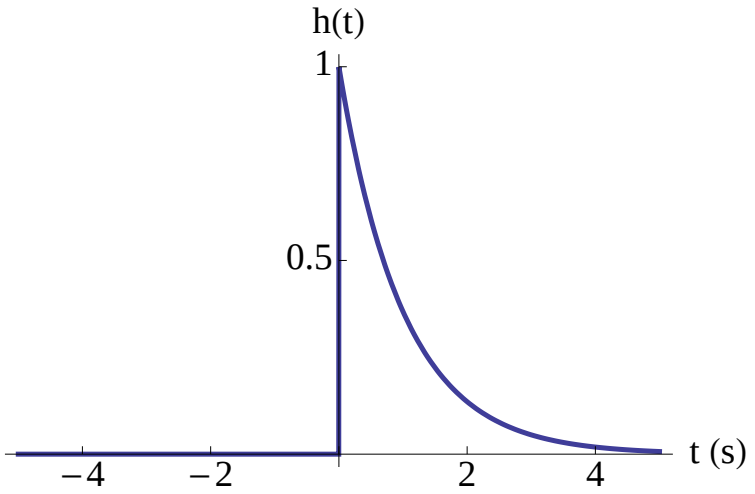

Abbildung 32: Impulse response of an ideal low-pass filter with a time constant of 1 second. The impulse is applied at $t_0 = 0$.

In the corresponding discrete case $x_n = \delta_{n,m_0}$, and the impulse response is given by h_{n-m_0}, which corresponds to the column with index m_0 of the matrix $\mathcal{H}$, i.e., to $\mathcal{H}_{n,m_0}$.

At this point we are in the position to appreciate the significance of the response time τ. Suppose our measuring instrument is a particle counter, where the absorption of each particle is a sharp impulse in time, which causes an impulse response of the instrument. If many particles are typically absorbed within time τ, then the individual absorption events cannot be resolved in time. If, however, incident particles are typically separated by far

more than τ, the counter can nicely resolve individual absorption events. The response time therefore sets the time-resolution of the counter. On the other hand, if we sample the output of the particle counter using a sampling interval Δt which is much larger than τ, we are likely missing out most events. In order to get the full information about the detected events, we need to sample at a rate $\Delta t \ll \tau$.

7.3.4 Step response function

If $x(t) = \theta(t)$ is a step-function of unit amplitude with the step occurring at time $t = 0$, then the output signal will be given by the step response function $H(t)$. It is obtained from the impulse response function by

$$H(t) = \int_0^t h(t - t')dt'.$$

A typical step response function is depicted in Fig. 33.

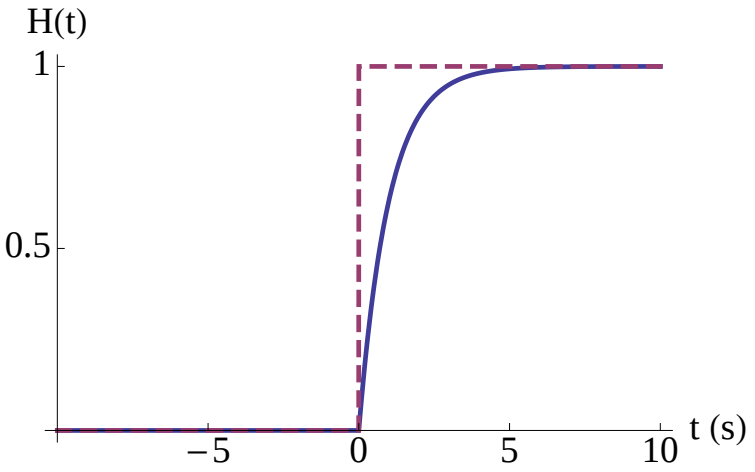

Abbildung 33: Step response of a measuring instrument with ideal low-pass behavior with a time constant of 1 second. The input step is shown as a dashed line.

The significance of the response time τ becomes evident, if we look at the response to a staircase at the input of the measuring instrument shown in Fig. 34. Steps follow each other after time intervals Δt. The response is plotted for different $\tau/\Delta t$ ratios. The staircase is recovered at the output only, if $\tau \ll \Delta t$. In the opposite case of $\tau \gg \Delta t$, the steps are completely washed out and the output is significantly delayed.

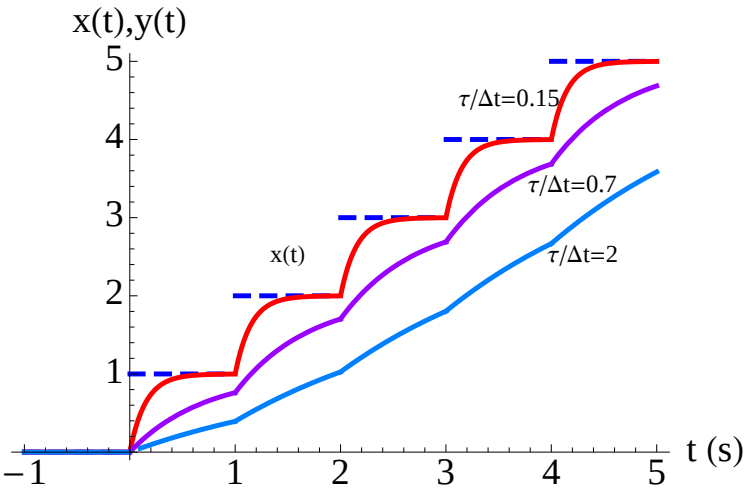

Abbildung 34: Response of a measuring instrument with ideal low-pass filter behavior with a time constant of 1 second to a signal staircase at the input. The input staircase is shown as a dashed line, the response is plotted for different ratios $\tau/\Delta t$.

7.3.5 Low-pass response to a ramp at the input

Suppose the input of our measuring instrument increases linearly at a rate r from the value x_0 to a value $x_1 = x_0 + rT$ within time span T. This ramp is described by

$$x(t) = x_0 + rt\theta(t) - r(t-T)\theta(t-T).$$

Suppose further that our measuring instrument has a low-pass impulse response function described by eq. (39). Using eq. (37) we find the output signal of the measuring instrument

$$y(t) = x_0 + \theta(t)\left[r(t-\tau) + r\tau e^{-\frac{t}{\tau}}\right] - \theta(t-T)\left[\tau r e^{-(t-T)/\tau} + r(t-\tau-T)\right].$$

This function is shown in Fig. 35 together with the input ramp $x(t)$. We can see two characteristic features of the low-pass response: first, the onset of the ramp appears with a time delay τ at the output such that the linear ramp at the output is shifted in time by τ. Second, the sharp kinks in the input at $t = 0$ and at $t = \tau$ are smoothened out by the exponential response function.

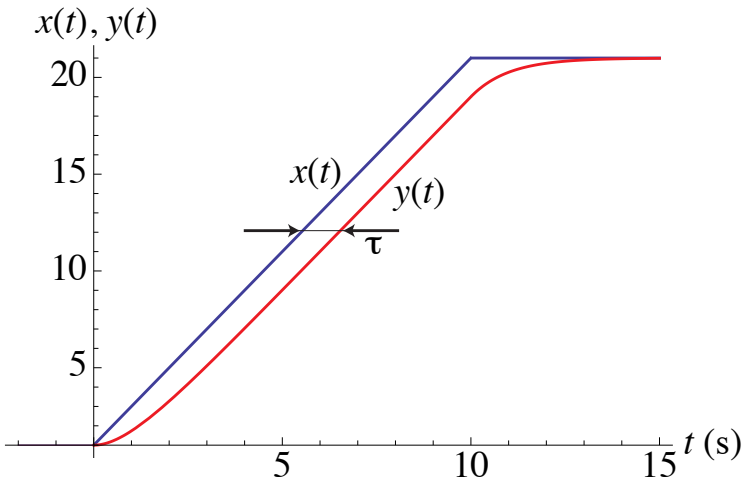

Abbildung 35: Blue curve: ideal ramp $x(t)$ at the input of the measuring instrument. Red curve: output of the measuring instrument, which is assumed to have an ideal low-pass response.

7.3.6 Frequency response function for analog instruments

If we perform a Fourier transform of eq. (37), we obtain (using the convolution theorem)

$$y(\omega) = h(\omega)x(\omega),$$

where $h(\omega)$ is the *frequency response function* given by

$$h(\omega) = \int_{-\infty}^{+\infty} dt h(t) e^{-i\omega t} \tag{40}$$

with $\omega = 2\pi f$ being the angular frequency.

low-pass response

For example, the frequency response of a measuring instrument with the low-pass characteristic of eq. (39) is given by

$$h(\omega) = \frac{1}{1 + i\omega\tau}.$$

We talk about a low-pass response, if

$$\log |h(\omega)|^2 = -\log(1 + (\omega\tau)^2).$$

This function is essentially constant for $\omega\tau \ll 1$, i.e., for frequencies below $f_{\mathrm{BW}} = \omega_{\mathrm{BW}}/2\pi = 1/2\pi\tau$. This frequency is called the bandwidth of the measuring instrument. At $f = f_{\mathrm{BW}}$

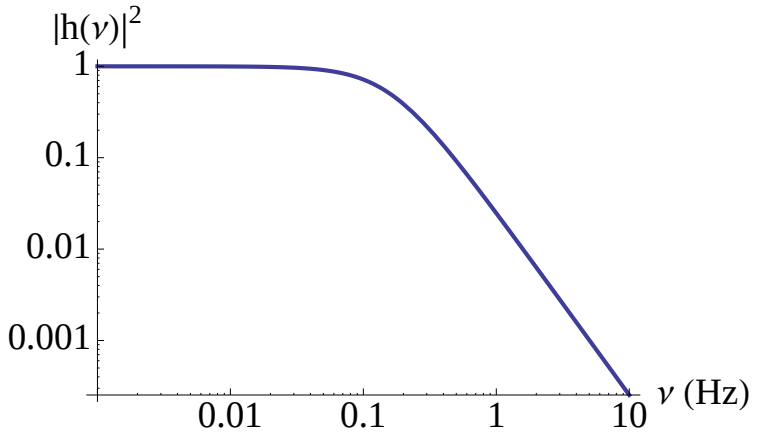

Abbildung 36: Frequency response function of a low-pass filter with a time constant of 1 second.

the response $|h(\omega)|^2$ has decayed to half its zero-frequency value, which corresponds to -3 dB. Engineers talk about the "3 dB point". For frequencies above f_{BW} it decays according to a power law in frequency with a slope of 6 dB/oct. Such a low-pass frequency response function is depicted in Fig. 36. 3 dB-point

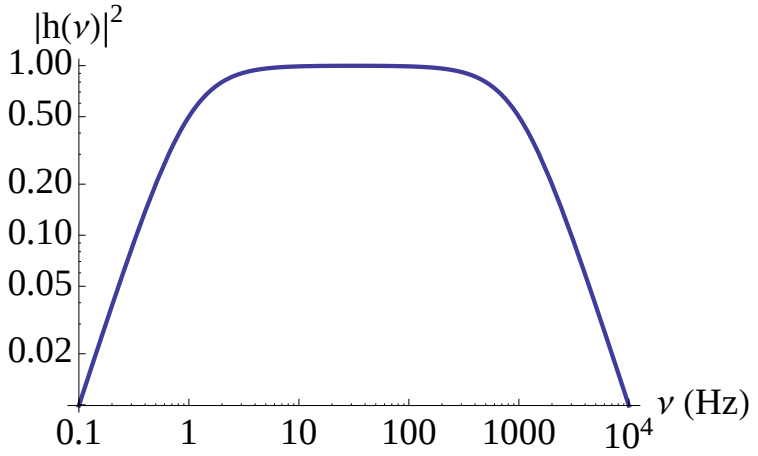

Abbildung 37: Bandpass response function with a passband between 1 Hz and 1 kHz.

We talk about a band-pass response, if $|h(\omega)|^2$ is essentially constant above f_{min} and below f_{max}. This interval is called the *passband*. An example is shown in Fig. 37.

7.3.7 Frequency response function for discrete sampling

If the response function is given in N discretely sampled points $h_n = h(n\Delta t)$ ($n = 1, \ldots, N$), then the analogue to the Fourier transform (40) is the discrete Fourier transform[6]

$$H_m = \sum_{n=1}^{N} h_n e^{-2\pi i(m-1)(n-1)/N},$$

with the inverse

$$h_n = \frac{1}{N} \sum_{m=1}^{N} H_m e^{2\pi i(m-1)(n-1)/N}.$$

Note that this is a linear transformation. The discrete Fourier transform can be seen as an approximation of the continuous Fourier transform in the frequency interval $[0, 1/2\Delta t]$, if you assign discrete frequencies

$$f_m = \frac{m-1}{N\Delta t}.$$

The continuous Fourier transform $h(2\pi f)$ is then approximately given at these discrete frequency points by

$$h(2\pi f_m) = H_m \Delta t.$$

Figure 38 shows, as an example, the discretely sampled response function of an instrument with low-pass filter characteristic. The corresponding discrete Fourier transform is shown in Fig. 39 on a double-logarithmic scale. Significant deviations arise only at the high frequency end of the spectrum. This could be improved by sampling with smaller Δt in the time domain.

7.3.8 Sampling theorem

In modern measurement systems, signals in the time-domain are often sampled, i.e., values are measured at discrete times $t_j = j\Delta t$ (j integer), where $f_\mathrm{s} = 1/\Delta t$ is called the sampling frequency. The purpose of sampling is usually to digitize analog signals for further digital processing. In fact, even when you read an analog measuring instrument and write down the number in your lab book, you are doing exactly the same thing. The Shannon-Nyquist

[6] Note that different definitions are used in different fields and in common software. Our definition agrees with the common definition in signal processing, which is also used in MATLAB.

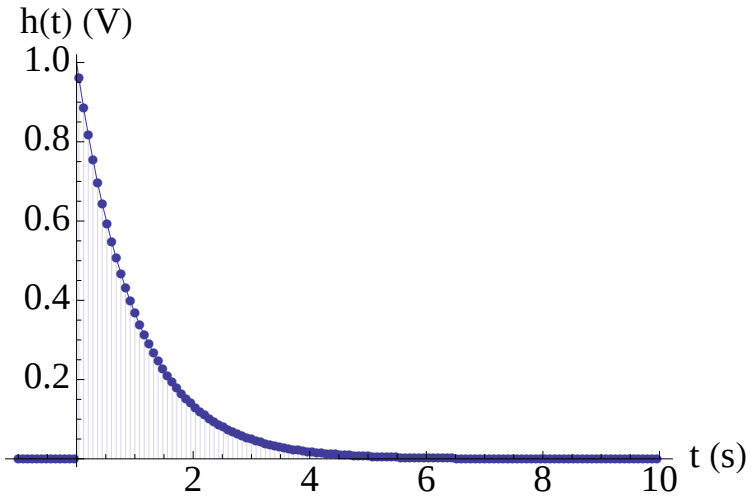

Abbildung 38: Impulse response function of an instrument with low-pass characteristic. The time constant is 1 s corresponding to a bandwidth $\Delta f = 0.16\,\mathrm{Hz}$. The sampling interval is $\Delta t = 80\,\mathrm{ms}$, and the total length of the time trace is $T = 11\,\mathrm{s}$. The solid line represents the continuous impulse response function.

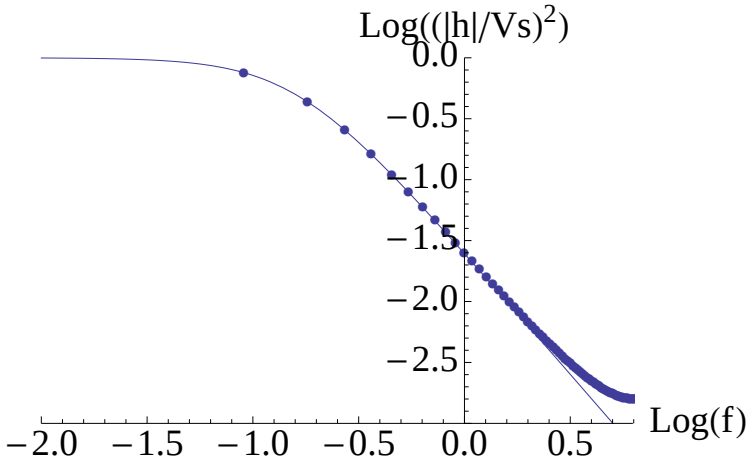

Abbildung 39: Discrete Fourier transform of the low-pass response of Fig. 38. The plotted points have coordinates $((n-1)/N\Delta t, h_n \Delta t)$. The solid line is the corresponding continuous Fourier transform.

sampling theorem [**?**] states that a bandwidth-limited signal $x(t)$ with bandwidth Δf is completely determined by sampling with

$$\Delta t = \frac{1}{f_s} = \frac{1}{2\Delta f}.$$

Figure 40 shows the example of a signal that is bandwidth limited to $\Delta f = 15.75\,\mathrm{Hz}$ and sampled with $f_s = 31.5\,\mathrm{Hz}$.

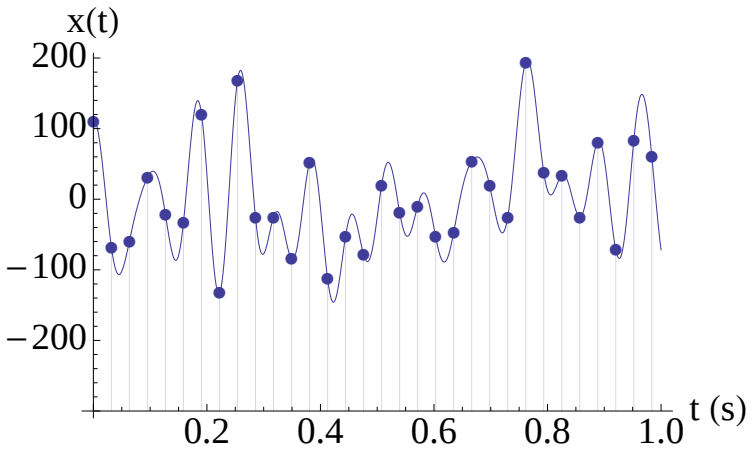

Abbildung 40: One-second time trace of a signal $s(t)$ bandwidth limited by $\Delta f = 15.75\,\mathrm{Hz}$, sampled at a frequency $f_s = 31.5\,\mathrm{Hz}$.

7.4 Precision and noise

The precision of a measuring instrument describes its ability to reproduce the same output value y for a constant (noiseless) input signal x under specified conditions [**?**]. Sources of noise internal to the instrument will produce an output signal $y(t)$ consisting of a noise contribution $n(t)$ fluctuating randomly in time superimposed on a constant value $f(x)$ (often called the signal"), such that

$$y(t) = f(x) + n(t). \tag{41}$$

Suppose we observe $n(t)$ for a certain time span T giving

$$n_T(t) = \begin{cases} n(t) & \text{for } |t| < T/2 \\ 0 & \text{for } |t| > T/2. \end{cases}$$

The fluctuating signal will then have the property that its time average

$$\overline{n(t)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} dt n(t) = \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} dt n_T(t) = 0.$$

Furthermore, we can say that

$$\overline{n(t)} = \overline{n(t+\tau)}, \tag{42}$$

which is one condition for describing the noise as a *stationary* random process. Such a signal may look like the curve depicted in Fig. 40. Note that eq. (41) represents a very common mathematical model for the appearance of noise in a measuring instrument. However, in some cases more sophisticated models may be required.

7.4.1 Mean-square fluctuation amplitude.

Reading the measuring instrument at different times will then give statistically fluctuating values. The mean square value of the fluctuations (provided that it exists) given by

$$\overline{n^2(t)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} n^2(t) dt = \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} n_T^2(t) dt$$

limits the precision of the measurement. This gives rise to *statistical errors* in the measurement. Statistical errors in measurements have nothing to do with mistakes. They are rather an inevitable property of each measuring instrument.

Statistical errors

We consider only cases here, where the noise $n(t)$ can be described as a stationary random process, which in addition to eq. (42) implies that

$$\overline{n^2(t)} = \overline{n^2(t+\tau)} \tag{43}$$

for arbitrary times τ.

Time-dependent noise signals $n(t)$ can be converted to a voltage and visualized with an oscilloscope. Figure 41 shows an example of such a noisy time-trace. The width of the trace indicated by the two dotted lines with the arrows is a direct measure of the fluctuation amplitude $\overline{n^2(t)}$.

Oscilloscope

7.4.2 Power spectral density of the noise

The function $n_T(t)$ can be Fourier-transformed to give $n_T(\omega)$. The quantity

$$S(\omega) = \ = \lim_{T\to\infty} \frac{|n_T(\omega)|^2}{T} =: \overline{|n(\omega)|^2}$$

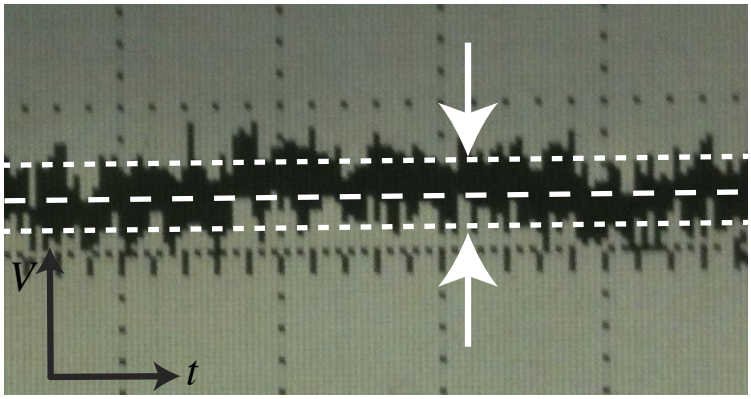

Abbildung 41: Image of the display of a modern digital oscilloscope showing a time-dependent noise signal $n(t)$. The horizontal axis of the display is the time axis with a scale of 250 μs per division. The vertical axis is the voltage axis with a scale of 2 mV per division.

is called the power spectral density of the noise of the measuring instrument. This function will usually have a high-frequency cut-off related to the noise-bandwidth of the measuring instrument. Typical noise spectral density curves are shown in Fig. 42. It can be shown that the mean squared fluctuation amplitude

$$\overline{n^2(t)} = \int S(\omega) d\omega, \tag{44}$$

i.e., it is the power spectral density integrated over the noise-bandwidth of the measuring instrument.

The power spectral density of the noise can be converted to a voltage and measured with a spectrum analyzer. This measuring instrument filters the signal $n(t)$ and samples a finite length time-trace. Using fast Fourier transform (FFT) algorithms, the power spectral density can be calculated and displayed. Figure 43 shows the display of such a spectrum analyzer. The dashed line indicates the value of a constant white-noise background. The displayed quantity is $\sqrt{S(\omega)}$.

Spectrum analyzer

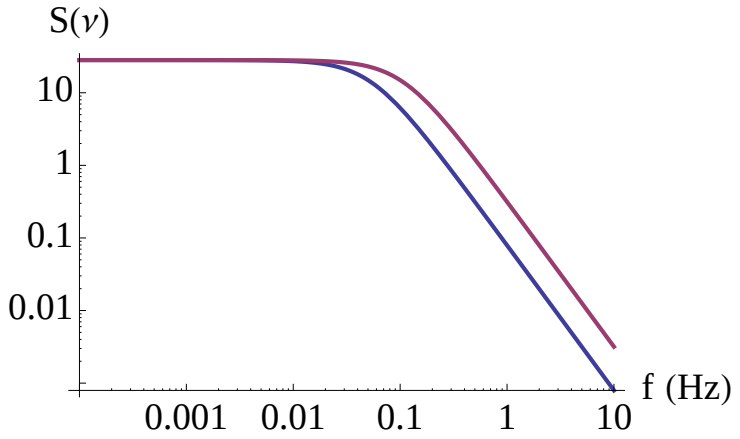

Abbildung 42: Typical noise spectral density functions with noise bandwidth of 0.05 Hz and 0.1 Hz.

7.4.3 Noise autocorrelation function

The Fourier transform of the power spectral density gives a time-domain function which is called the autocorrelation function of the noise

$$\begin{aligned} C(\tau) &= \frac{1}{2\pi}\int S(\omega)e^{i\omega\tau}d\omega \\ &= \lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty} n_T(t-\tau)n_T(t)dt =: \overline{n(t)n(t+\tau)} \end{aligned} \tag{45}$$

Usually this correlation function has a maximum at $\tau = 0$ and decays with increasing $|\tau|$ such that $\lim_{\tau\to\pm\infty} = 0$. At $\tau = 0$ the value is

$$C(\tau = 0) = \overline{n^2(t)}.$$

For the stationary noise that we consider here, the autocorrelation function has the property (c.f. eq. (43))

$$\overline{n(t)n(t+\tau)} = \overline{n(0)n(\tau)}.$$

We define the correlation coefficient

$$\rho(\tau) = C(\tau)/C(\tau = 0)$$

with the property $\rho(\tau = 0) = 1$ and $\lim_{\tau\to\infty}\rho(\tau) = 0$.

The *correlation time* τ_c of the noise is defined via

Correlation time

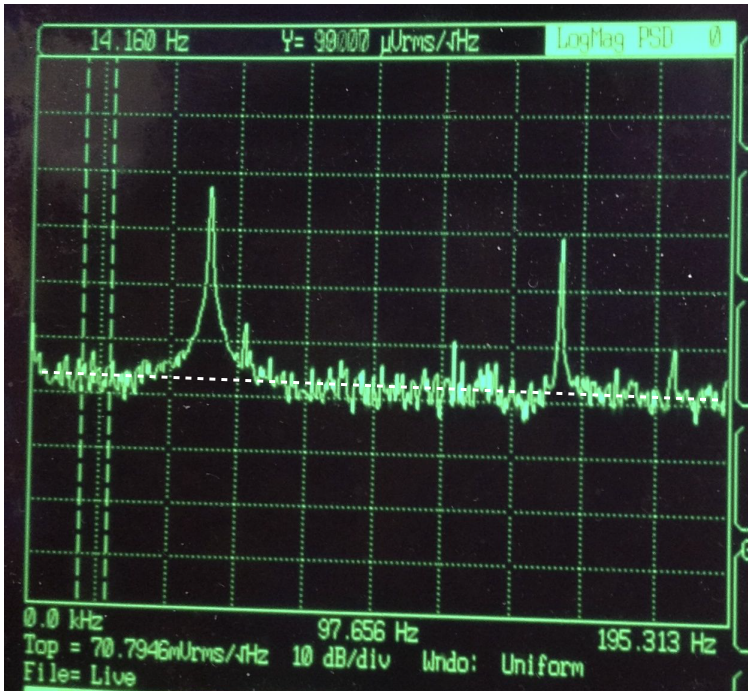

Abbildung 43: Measured power spectral density of the noise. The horizontal dashed line indicates the level of the white background noise, which is about $100\,\mu\mathrm{V}/\sqrt{\mathrm{Hz}}$. This value represents $\sqrt{S(\omega)}$. In addition, two large peaks are seen in the spectrum, which originate from 50 Hz and 150 Hz signals (mains frequency 50 Hz and its harmonic) originating from power transformers disturbing the measurement. The horizontal axis is the frequency axis ranging from 0 to 195.313 Hz.

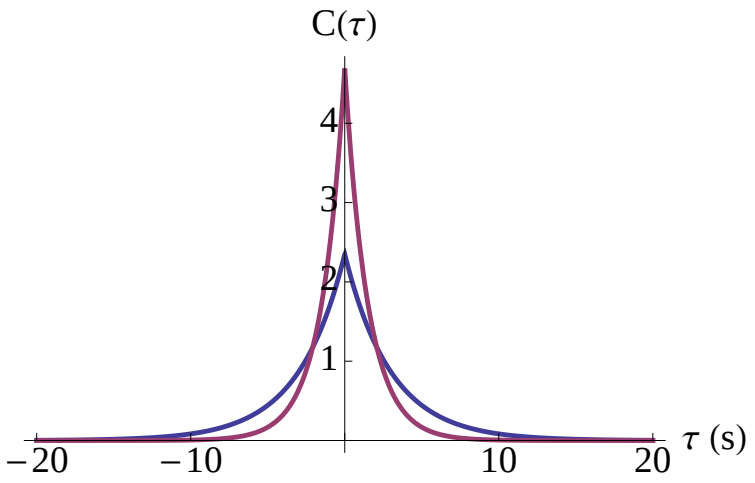

Abbildung 44: Typical correlation functions with correlation times of 2.1 s and 1.05 s.

$$C(\tau_c) = C(0)/2 = \overline{n^2(t)}/2 \quad \text{or equivalently as} \quad \rho(\tau_c) = 1/2.$$

The noise output $n(t)$ is strongly correlated with $n(t + \Delta t)$ if $|\Delta t| \ll \tau_c$ and essentially uncorrelated, if $|\Delta t| \gg \tau_c$. A large noise bandwidth will lead to short correlation times and vice versa. Typical correlation functions are shown in Fig. 44.

7.4.4 Probability distribution for uncorrelated noise

Suppose we sample the noisy output of an analog measuring instrument having a constant (in time) noiseless input signal applied. The output signal is then given by eq. (41), and it may look similar to the signal in Fig. 40. We choose the sampling interval $\Delta t \gg \tau_c$ such that subsequent readings are statistically independent. We will then get a distribution of output values y that can be represented in a histogram plot as shown in Fig. 45. It resembles a symmetric bell-shaped curve with maximum at $y = f(x)$ and width $\sqrt{\overline{n^2(t)}}$.

The red curve plotted into the histogram shows that this distribution can be well represented as a Gaussian probability density (also called normal distribution)

$$\mathrm{pdf}(y|x, \overline{n^2(t)}, f(x)) = \frac{1}{\sqrt{2\pi\overline{n^2(t)}}} \exp\left(-\frac{(y - f(x))^2}{2\overline{n^2(t)}}\right). \tag{46}$$

From the full width at half maximum of this distribution we can determine the mean square fluctuation amplitude $\overline{n^2(t)}$. We can interpret the above conditional probability

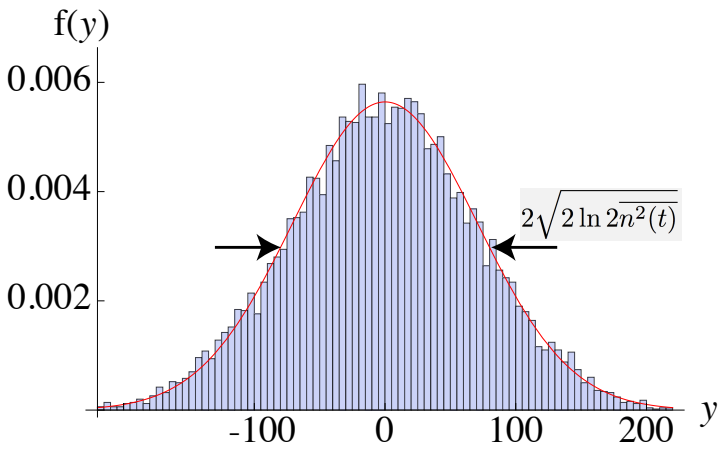

Abbildung 45: Histogram of 10'000 statistically independent data points sampled from a measuring instrument. The sampling interval Δt was chosen to be large compared to the correlation time τ_c in order to ensure statistical independence of the noise in the samples. The full width at half maximum of the distribution is related to the mean square fluctuation amplitude.

distribution in terms of the uncertainty about propositions introduced before. The value of the probability density function $\mathrm{pdf}(y|x, \overline{n^2(t)}, f(x))dy$ quantifies our uncertainty about (is the probability of) the proposition that we obtain a value within the interval $[y, y+dy[$ when sampling the output once, *given* that the input value x, the mean square noise of the measurement $\overline{n^2(t)}$ and the behaviour of the measuring instrument $f(x)$ *are exactly known*. In this sense, eq. (46) represents a predictive probability distribution. It predicts the measurement outcome, before it has been measured. We will talk more about the prediction of measurement results in chapter **??**.

Real data and approximate models for random processes

At this point it is important to notice the conceptional step that we have taken by approximating the histogram distribution, which represents real measured data, by a probability density function, which represents *an approximate model* of the measured distribution. The situation is very similar to other cases in physics that you know. When we describe the motion of an apple that falls down from a tree, we may have recorded data of the height $h(t)$ as a function of time. When we use Newton's laws of motion to describe the measured data with the equation $h(t) = h_0 - gt^2/2$, then this equation represents an approximate model of the real data at hand. While this equation is a model for the systematic change of $h(t)$, a probability (density) distribution is a model for the statistical properties of certain quantities in a measurement.

Since we will encounter the normal distribution frequently in the following, we introduce

the notation

$$N_{\mu,\sigma}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{47}$$

This would allow us to write $\mathrm{pdf}(y|x, \overline{n^2(t)}, f(x)) = N_{\mu,\sigma}(y)$, where $\mu = f(x)$ and $\sigma^2 = \overline{n^2(t)}$.

The normal distribution (47) allows us to make a prediction about a single datum x, if we know μ and σ. However, the histogram in Fig. 45 is the result of a large number of N data x_i ($i = 1 \ldots N$). What is the probability to obtain such a data set $\{x_i\}$ giving such a histogram? Using the product rule of probabilities and assuming statistical independence in the noise of our values, we can apply the product rule (**??**) for statistically independent propositions and write

$$\mathrm{pdf}(\{x_i\}|\mu,\sigma) = \prod_{i=1}^{N} N_{\mu,\sigma}(x_i) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}\right). \tag{48}$$

The expression in the numerator of the exponent can be rewritten as

$$\sum_{i=1}^{N}(x_i-\mu)^2 = N\overline{(x_i-\mu)^2} = N(\mu-\overline{x_i})^2 + N\overline{(x_i-\overline{x_i})^2}.$$

Here we have defined the sample mean

$$\overline{x_i} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

and the sample variance

$$\overline{(x_i-\overline{x_i})^2} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x_i})^2.$$

Maximum likelihood estimate of μ and σ

At this point we will make another important turn. The above expression gives us the probability that the dataset $\{x_i\}$ was drawn from a gaussian distribution *given* the mean μ and the width parameter σ. Of course, there is a whole class of distributions having this functional form for different parameters μ, σ. We may now ask the question, which distribution of this class would have produced the data with the largest probability? How would we have to choose μ and σ? The answer to this question is easily found: from this whole class of distribution functions, the one with

$$\mu = \overline{x_i} \quad \text{and} \quad \sigma^2 = \overline{(x_i-\overline{x_i})^2}$$

gives the data of our histogram with the highest probability. We have found these values by maximizing eq. (48) with respect to μ and σ. The result for μ and σ is called the maximum likelihood estimate of these quantities. However, we should warn the reader that a maximum likelihood estimate does not in general solve the problem of estimating μ and σ from the data consistently. What we would rather like to have for such an estimate is a statement like "what are the most plausible values of μ and σ given the data and assuming the validity of the gaussian model"(meaning that we aim at a probability distribution for μ and σ). The statement that we have now made is different in a subtle way: it states "which distribution of the class of all gaussian distributions would have predicted our data with the highest probability".

We see that a measuring instrument introduces uncertainty about a measurement quantity, because it suffers from internal noise quantified by $\overline{n^2(t)}$. Equation (44) shows that the internal noise can usually be suppressed by reducing the bandwidth. The price to pay for the reduction of the noise is a slower time-response of the system, and increased correlation time, and therefore an increased measurement time, if the noise on subsequently recorded data points is to remain uncorrelated.

The average uncertainty introduced by the measuring instrument can be quantified, if we inspect the probability density function in eq. (46) and use a uniform point density function $m(x) = 1/\Delta$ in eq. (**??**). The entropy can be worked out to be [**?**]

$$H = \frac{1}{2}\log_2(2\pi e\overline{n^2(t)}/\Delta^2), \tag{49}$$

where Δ is very small compared to σ. The average uncertainty is proportional to the logarithm of the width σ of the distribution function, or to the mean square noise of the measuring instrument.

Equations (44) and (49) allow doing a quantitative evaluation of information gain by bandwidth reduction. Assume for example, that the power spectral density of the instrument noise is white, but limited by a low-pass filter at the output to the bandwidth Δf. If we reduce this bandwidth to $\Delta f/4$, we reduce $\overline{n^2(t)}$ by a factor of four, and therefore gain one bit of information about the quantity to be measured. A bandwidth reduction by a factor of hundred to $\Delta f/100$ reduces $\overline{n^2(t)}$ by hundred and gives an information gain of $3.32\,\text{bit} = 1$ decimal digit. This means for example, that you will see one additional stable digit on the display of your digital multimeter.

7.5 Accuracy and Calibration

7.5.1 Accuracy

The accuracy of a measuring instrument describes in how far its real behaviour agrees with its ideal (theoretical) design behavior described by $y = f(x)$. The very common case of a linear amplifier, for example, may have the ideal behavior

$$y = Gx,$$

where G is called the gain of the amplifier, which is its only performance parameter. However, a real linear amplifier may have

$$y = y_{\text{offs}} + \tilde{G}(x, T, t)(x - x_{\text{offs}}).$$

Here, x_{offs} is an input signal offset, y_{offs} an output signal offset, and the gain G has been replaced by the effective gain $\tilde{G}(T, x, t)$, which depends on time (t), local temperature (T), and on the level of the input signal x itself (which makes the amplifier nonlinear). There is no doubt that a good amplifier will have very small offsets, very weak nonlinearities, very weak time-dependence and very weak temperature dependence. However, the accuracy will *always* be limited. The input and output offsets may even vary periodically in time (for example, picking up spurious signals at the mains frequency and its multiples is fairly common nowadays), or they may creep with time during instrument warm-up or when a person enters the room and the temperature rises by 0.3 degrees. Also degradation of the instrument accuracy (meaning a change of the function $\tilde{G}(x, T, t)$) after many operating hours may be a serious issue.

Figure 46 shows the characteristic of a linear amplifier with an offset. Its linear input–output characteristic is shifted with respect to the ideal one by a certain amount. Figure 47 shows the characteristic of an amplifier with a gain error. The slope of its input–output characteristic differs from that of the ideal amplifier. Nonlinearities of an amplifier are shown in Fig. 48.

Figure 49 shows another type of systematic deviation from ideal behaviour that may be encountered with a linear amplifier. Imagine a case where the input signal is continuously swept in time from one value to another and back. If the amplifier response has a finite bandwidth, the output signal will show hysteric behavior. Hysteretic behaviour can be identified and reduced by changing the bandwidth of the amplifier (if possible), or by reducing the sweep rate of the input signal.

In the general case, the ideal performance of a measuring instrument, given by

$$y = f(x)$$

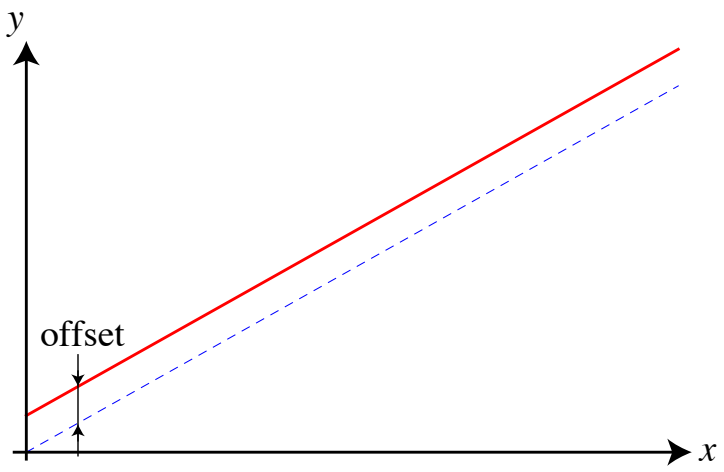

Abbildung 46: Characteristic of a measuring instrument with an offset. The dashed line is the ideal linear characteristic. The real characteristic is shifted by a constant amount from the ideal curve. Such behavior can be the source of systematic errors in a measurement.

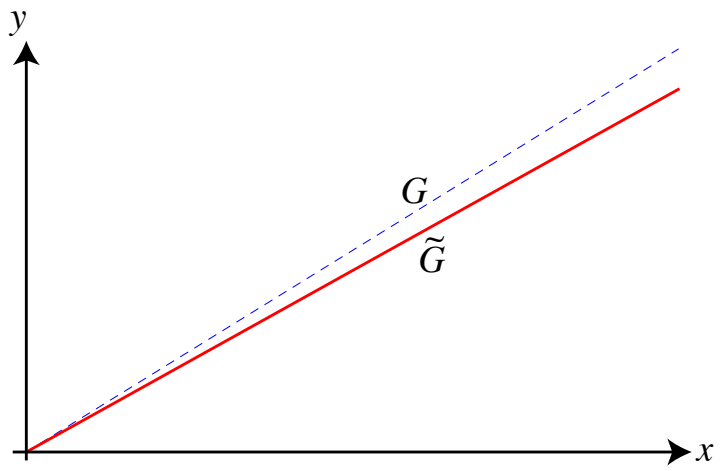

Abbildung 47: Characteristic of a measuring instrument with a gain error. The dashed line is the ideal linear characteristic. The real characteristic has a different slope than the ideal curve. This is another reason why measurements may suffer from systematic errors.

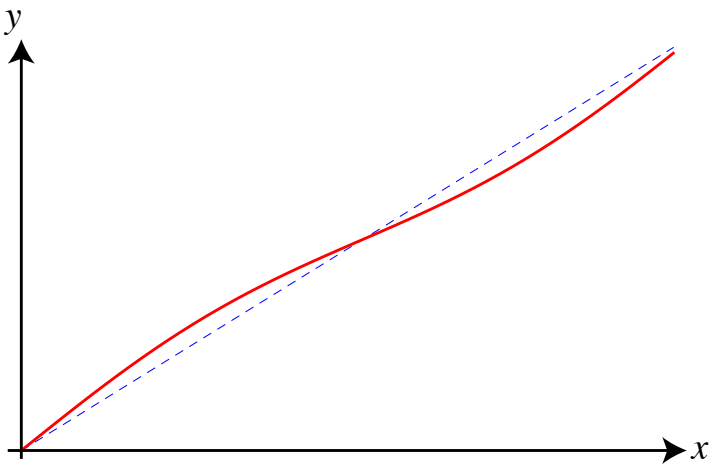

Abbildung 48: Characteristic of a measuring instrument with nonlinearities. The dashed line is the ideal linear characteristic. The real characteristic does not exactly follow the ideal curve, and therefore leads to systematic errors in a measurement.

may be spoiled in practice, and a deviating function

$$y = \tilde{f}(x)$$

may be in place. Different instruments of the same model from a particular manufacturer may differ in their functions $\tilde{f}(x)$, and the same measuring instrument may perform slightly different in different labs, because of differences in the environment (temperature, pressure, humidity, gravitational field, etc.).

7.5.2 Calibration

Systematic errors due to the limited accuracy of measuring instruments can be minimized by using modern, well made and well calibrated measurement equipment. If required, a very well calibrated measuring instrument can be used in the lab to calibrate less accurate equipment.

The process of calibration consists of comparing the outputs of two measuring instruments which are subject to the same input signal. One of the two measuring instruments is called the calibration standard and must have a measurement uncertainty much smaller

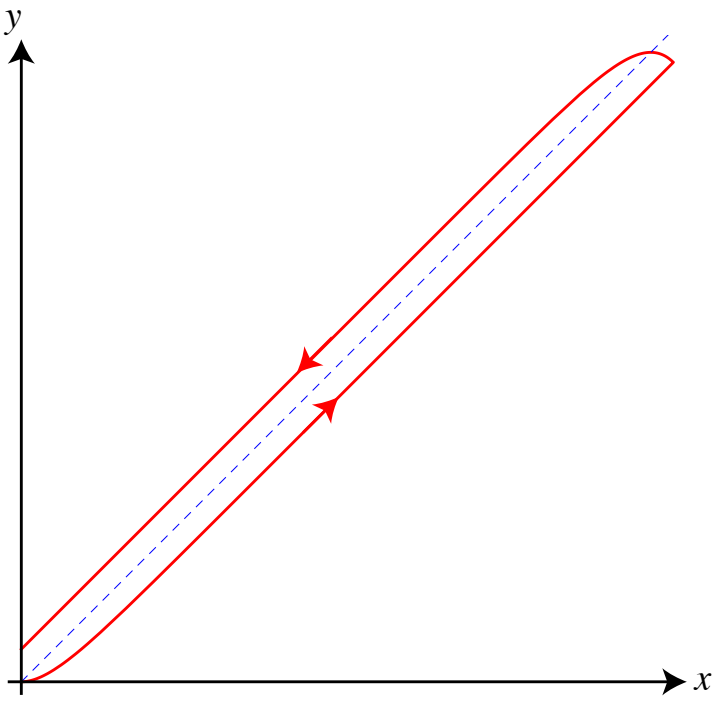

Abbildung 49: Characteristic of a measurement with hysteresis arising from the finite bandwidth of the linear amplifier. The dashed line is the ideal linear characteristic. The real output signals of up- and down-sweep differ.

than the other, which is being calibrated. The result of a calibration is a calibration function which translates the instrument reading into the value of the measured quantity. This function may be represented as a table, diagram, or curve.

Suppose for example, that we have an electromagnet consisting of two parallel coils producing a magnetic field along the axis of the coils. The current I through the coils determines the magnitude B of the magnetic field in the central point between the two coils on the coil axis. We may use a Hall-magnetometer as the calibration standard to relate the measured current through the coils with the magnetic field. The resulting $B(I)$ curve is the desired calibration curve.

The labs with the most accurate measuring instruments are the national institutes of standards. Although their measuring instruments also suffer from deficiencies in terms of accuracy, it is the sole purpose of these labs to make measuring instruments as accurate as possible, and to provide a calibration reference for manufacturers of instruments. For example they can compare masses with that of the kilogram standard with a relative uncertainty of 10^{-9}. Frequency (time) is measured even with a relative uncertainty of 10^{-16}. These are impressive numbers which are usually not reached in a typical research environment.

However, even the best calibration experiments will have limits in accuracy, such that some uncertainty remains. In order to incorporate this uncertainty into the mathematical description of the measurement equipment, we choose the following approach: suppose we have a measuring instrument that produces an output

$$y = \tilde{f}(x) = f(x) + y_{\text{offs}}.$$

The offset leads to a systematic change of the output reading y for a given input x. This is commonly called an undesired *systematic effect*. We have taken care of this effect by calibrating the instrument as good as we could, but there is a remaining calibration uncertainty in y_{offs} of σ_{offs}.

Formally we can say that

$$\begin{aligned}\text{pdf}(y|x, \overline{n^2(t)}, \sigma_{\text{offs}}, f(x)) = & \\ & \int dy_{\text{offs}} \text{pdf}(y, y_{\text{offs}}|x, \overline{n^2(t)}, \sigma_{\text{offs}}, f(x)) \\ & = \int dy_{\text{offs}} \text{pdf}(y_{\text{offs}}|x, \overline{n^2(t)}, \sigma_{\text{offs}}, f(x)) \times \\ & \qquad \text{pdf}(y|x, y_{\text{offs}}, \overline{n^2(t)}, \sigma_{\text{offs}}, f(x)).\end{aligned}$$

Here we have used in the first step the marginalization rule (**??**) for integrating out y_{offs}, and the product rule (**??**) in the second step for separating the probability density function for y_{offs}. This probability density represents the *insufficient knowledge* about y_{offs} that remained after calibration. It is in fact independent of x, $\overline{n^2(t)}$, and $f(x)$. We can therefore write

$$\text{pdf}(y|x, \overline{n^2(t)}, \sigma_{\text{offs}}, f(x)) = \int dy_{\text{offs}} \text{pdf}(y_{\text{offs}}|\sigma_{\text{offs}}) \text{pdf}(y|x, y_{\text{offs}}, \overline{n^2(t)}, f(x)). \quad (50)$$

The conditional probability density distribution $\text{pdf}(y|x, y_{\text{offs}}, \overline{n^2(t)}, f(x))$ is given in analogy to eq. (46) by

$$\text{pdf}(y|x, y_{\text{offs}}, \overline{n^2(t)}, f(x)) = N_{f(x)-y_{\text{offs}}, \overline{n^2(t)}}(y).$$

In order to see the effect of the integral in eq. (50), we may represent the insufficient knowledge about y_{offs} with a probability density distribution

$$\text{pdf}(y_{\text{offs}}|\sigma_{\text{offs}}) = N_{0,\sigma_{\text{offs}}}(y_{\text{offs}}).$$

Performing the integration we then find

$$\text{pdf}(y|x, \overline{n^2(t)}, \sigma_{\text{offs}}, f(x)) = N_{f(x),\sigma_{\text{tot}}}(y) \quad \text{where } \sigma_{\text{tot}}^2 = \overline{n^2(t)} + \sigma_{\text{offs}}^2. \quad (51)$$

This is a very important result, in particular comparing to eq. (46). We see here, that there are two kinds of uncertainties involved in the problem. On one hand we have the noise of the measuring instrument that leads to a statistical uncertainty about the value read at the output of the instrument. This uncertainty can be reduced experimentally by reducing the bandwidth of the measurement (see eq. (44)). On the other hand, we have an uncertainty about the offset remaining after the calibration of the measuring instrument. This uncertainty can in principle not be reduced by a reduction of the bandwidth, because it describes our insufficient knowledge about the apparatus. The total measurement uncertainty σ_{tot} can never become smaller than the uncertainty σ_{offs} about the systematic parameter y_{offs}. This is a first instance where we have encountered the effect of a *systematic error*. The quantity σ_{offs} describes the limited accuracy of the measuring instrument, whereas the parameter $\overline{n^2(t)}$ describes its limited precision.

7.5.3 Accuracy vs. Precision

In the above discussion we have introduced the notion of the accuracy, as the ability of a measuring instrument to follow some ideal design behavior, and the notion of precision, as

the ability of a measuring instrument to reproduce a certain output value given a noiseless constant input signal. As we have seen in the quantitative example given above, the overall performance of a measuring instrument in a particular application may be either governed by its accuracy or its precision. We illustrate this graphically in Fig. 50. In these histogram

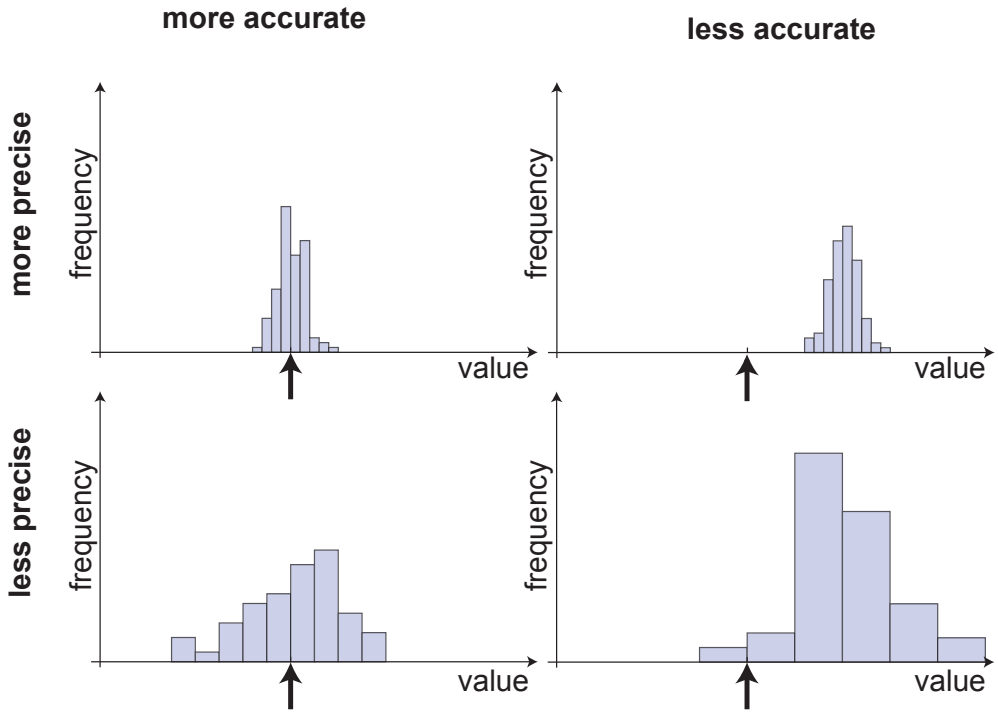

Abbildung 50: Graphical illustration of an accurate measurement with high precision (top left), a less accurate measurement with high precision (top right), an accurate measurement with low precision (bottom left), and a less accurate measurement with low precision.

plots, the true value is indicated by a black arrow. The larger the distance of the distribution's center of mass from the position of the arrow is, the larger is the deviation of the measurement outcome from the true value. Comparing Figs. 50(top left) and 50(bottom left) we notice that in both cases the measurements scatter randomly around the true value, which is characteristic for accurate measurements. However, the scatter in Fig. 50(top left) is much smaller than in Fig. 50(bottom left), meaning that the former is more precise than the latter. In contrast, Fig. 50(top right) and (bottom right) show measurements with lower accuracy. They both have a bias towards the right. Still the measurements in Fig. 50(top right) show higher precision than the measurements in Fig. 50(bottom right),

because they are concentrated sharper around the mean.