Chapter 5

Characterization of instrumental response

5.1 Characteristic parameters

Characterization of the performance and response of a physical measurement system can be achieved by considering a multidimensional parameter space, in which each coordinate axis relates to a specific instrument parameter that has a major influence on the quality of the measurement. For astrophysical observations these parameters comprise, among others, bandwidth, field of view, precision and resolution, limiting sensitivity.

In the following paragraphs a brief description of these parameters is given. It should be kept in mind, however, that a certain "bandwidth" in the definitions of these parameters is inevitable and due account must be taken of their exact meaning in a particular observational context.

5.1.1 Bandwidth (symbol: $\lambda\lambda$, $\nu\nu$, or $\epsilon\epsilon$)

The *bandwidth*, or rather *spectral bandwidth*, is defined as the wavelength (or frequency, or energy) interval over which the instrument has adequate detection efficiency, i.e. over which it is observationally employed. The long and short wavelength cut-offs are derived from the wavelength dependent detection efficiency, which is governed by the physical interaction process (see chapter 4). The precise criteria for choosing the cut-off values are mostly rather arbitrary, e.g. cut-off is defined at 50 % or 10 % of the value of maximum efficiency in the bandpass.

5.1.2 Field of view (FOV, symbol: Ω_{FOV})

The *field of view* is defined as the solid angle subtended on the sky by the selected telescope configuration. If the effective area of the telescope is a continuously declining function of the off-axis (optical axis) angle, the choice of the effective FOV is somewhat arbitrary (like in the case of spectral bandwidth), e.g. decline to half the maximum sensitive area (FOV at Full Width at Half Maximum (FWHM)) or to zero sensitive area (FOV at Full Width at Zero Response (FWZR)). The effective FOV is also dependent on the angular resolution which is required for an observation: due to the potential decline of the image quality with off-axis angle not the whole field of view may qualify for a particular observation.

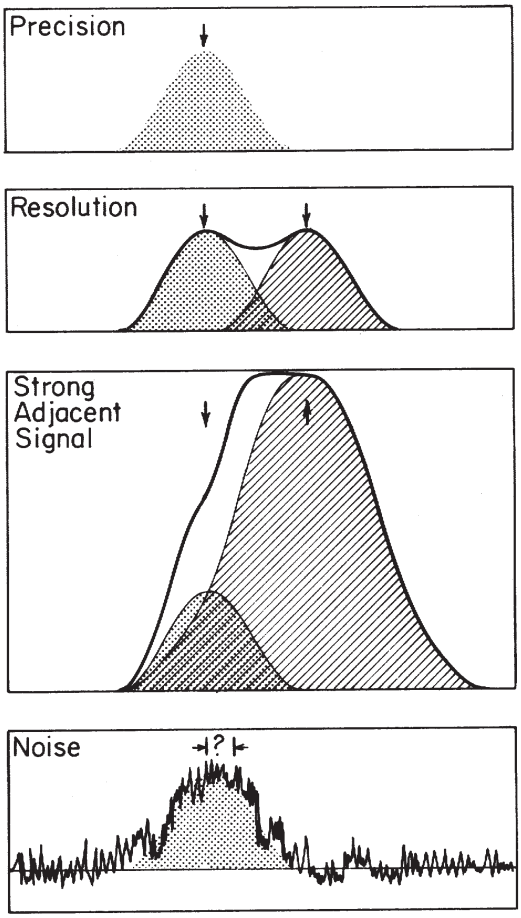


Figure 5.1: *Precision and resolution. The top panel shows that precision is the accuracy with which the centroid of a point (line) spread function can be determined. The panel below shows that the resolution is the interval which two signals of equal strength should be apart to be recognised separately. It is harder to resolve a weak signal adjacent to a strong one (panel below). The bottom panel shows the deterioration of precision in the presence of noise. Figure taken from Harwitt (1984).*

5.1.3 Precision and resolution

First of all, it is important to make a proper distinction between precision and resolution.

Precision

Precision represents the accuracy with which the exact value of a certain quantity can be established. In astrometry precision reflects the ability to accurately measure the position of a star, in spectroscopy precision reflects the ability to accurately pin down the exact wavelength (or frequency, or energy) of a spectral line. The precision of which an instrument is capable may substantially exceed its resolution.

For example: a point source of radiation. The image is blurred by the finite angular

resolution of the telescope, however if this blur is more or less symmetical in shape, the centre of the image can be determined with substantially higher precision than the spot size (see figure 5.1). The ultimate positional accuracy will be governed by the brightness of the spot with respect to the sky background, by instrumental noise sources and by a potential off-set between the measured position and the true position as a result of systematic errors. An example of such a systematic error is the misalignment between the telescope axis and the attitude reference system in a space-based astronomical observatory. The optical axis of the star sensors provide the attitude information and consequently the pointing position on the sky and may be slightly misaligned with the main telescope axis due to thermal gradients in the spacecraft. Of course, this can be calibrated on celestial sources visible in both the main telescope and the optical reference sensors, but presumably not always. Consequently, for high absolute position accuracies, this "bias" error may be non-negligible compared to the statistical spread in the centroid determination of the point source image. The final accuracy to which a position can be derived is sometimes referred to as *position resolution*, not to be confused with angular resolution (see later).

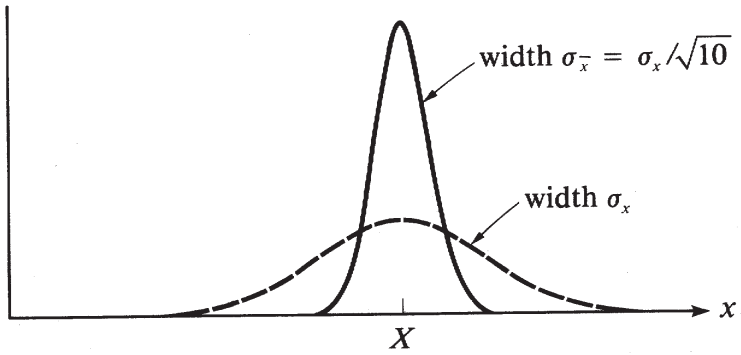
Suppose that the measurements of the stochastic variable X are normally distributed about the true value μ with width parameter σ_X. N independent measurements of X are done and the average is calculated:

$$\bar{X} = \frac{X_1 + \cdots + X_N}{N} \tag{5.1}$$

The standard deviation in the mean is:

$$\begin{aligned} \sigma_{\bar{X}} &= \sqrt{\left(\frac{\partial \bar{X}}{\partial X_1}\,\sigma_{X_1}\right)^2 + \cdots + \left(\frac{\partial \bar{X}}{\partial X_N}\,\sigma_{X_N}\right)^2} \\ &= \sqrt{\left(\frac{1}{N}\,\sigma_X\right)^2 + \cdots + \left(\frac{1}{N}\,\sigma_X\right)^2} \\ &= \frac{\sigma_X}{\sqrt{N}} \end{aligned} \tag{5.2}$$

So the width of the Gaussian curve will reduce by the square root of the number of measurements, which is shown in the following figure for the case of $N = 10$ (taken from Taylor (1982)):



Another example is the timing precision (symbol: $\Delta t_{abs} = t_{abs} - t_{observatory}$), which equals the difference between the measured time and true time. The absolute timing accuracy is particularly important for synchronisation of periodic phenomena over time series with repeated interupts (satellite data) or over different observations. For example in pulsar observations, the period of a pulsar (and other parameters like period derivative or orbital period and orbital period derivative in case of a binary) can be determined to a much higher precision, when two observations at different times are synchronised: i.e. the exact number of pulse periods between the two observations should be known. From the time of arrival of the pulse at the first observation and the predicted period (and period derivative etc.) a prediction is made of the time of arrival of the pulses in the second observation. From the deviation of the measured time of arrival a better value of the period (and period derivative etc.) can be determined. (The absolute time of an observation is usually determined by comparing the observatory maser clock with a GPS system (Global Positioning System) which is locked to Univeral Time (UT) determined by the world's best atomic clocks.)

Resolution

Resolution (or *resolving power*) represents the capability of measuring the separation between two closely spaced features, e.g. two spectral lines or two point like radiation sources. The following resolutions are commonly used in astrophysical data:

- *Angular resolution* (symbol: $\Delta\theta$)
  This is defined as the minimum angular separation needed between two equally bright point sources to ensure that they can be individually just separated ("resolved").

  If an angle of $\Delta\theta$ radians can just be resolved, the angular resolving power is

$$R_\theta = \frac{1}{\Delta\theta} \tag{5.3}$$

- *Spectral resolution* (symbol: $\Delta\lambda, \Delta\nu$ or $\Delta\epsilon$)
  This is defined as the separation in wavelength (or frequency, or energy) between two spectral lines of equal intensity, that is just large enough to resolve the lines individually.

  The spectral resolving power is the ratio of the wavelength λ at which the observation is carried out, to the wavelength difference $\Delta\lambda$ which can just be resolved:

$$R_S = \frac{\lambda}{\Delta\lambda} = \frac{\nu}{\Delta\nu} = \frac{\epsilon}{\Delta\epsilon} \tag{5.4}$$

- *Charge and mass resolution* (symbol: ΔZ and ΔA)
  These characteristic parameters are relevant for the measurement of the composition of a particle beam (e.g. cosmic-rays). It is defined as the fractional charge and mass which can just be separated by a particle telescope.

  The charge and mass resolving power is the ratio of the charge Z, mass A, which is being measured, to the charge difference ΔZ, mass difference ΔA, which can be just resolved:

$$R_Z = \frac{Z}{\Delta Z} \tag{5.5}$$

$$R_A = \frac{A}{\Delta A} \tag{5.6}$$

- *Temporal resolution* (symbol: Δt)
  This parameter can be defined as representing the minimum time interval between two consecutive samples of a realisation of a stochastic process for which these samples can be regarded as uncorrelated (independent) at the output of the measuring device.

It is important to realise that the characteristic resolution parameters defined above, apply to closely spaced features of comparable strength. If, for instance, a faint star lies very close to a brighter companion, or a faint spectral line lies near a far brighter spectral feature, the two may not be separable with a given resolution, even though a pair of equally bright stars or spectral lines would be easily resolved at the same separation.

5.1.4 Limiting sensitivity [symbol: $(F(\lambda, \nu, \epsilon)_{min}$]

The limiting sensitivity achievable with a specific observation facility is defined as the minimum value of the monochromatic flux density arising from a celestial point source, that can be detected significantly over a certain measuring period T in a predefined wavelength (frequency, energy) interval.

This parameter is a function of the effective telescope area A_{eff}, integration time T and angular resolution $\Delta\theta$, moreover it may strongly depend on sky position (background light) and instrumental noise (potentially variable). If the nature of the radiation source is truely diffuse in origin, like a cosmic-ray beam incident on a particle telescope, the collecting power of the instrument is normally related to the so-called *geometry factor* or *grasp*. This parameter is defined as the telescope effective area integrated over the FOV of the telescope:

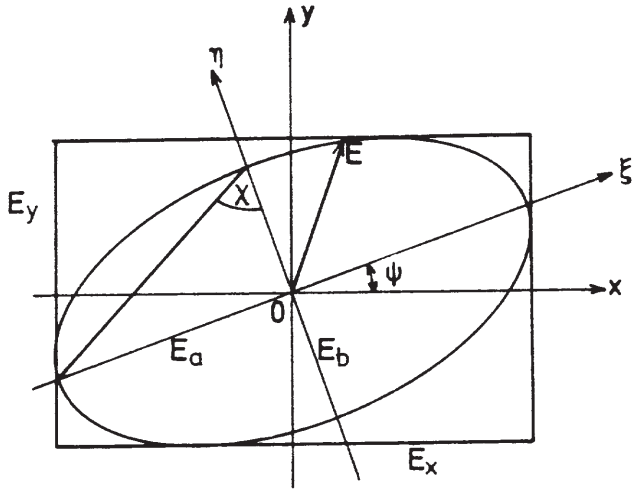
$$GF = \int_{\Omega_{FOV}} A_{eff}\, d\Omega \tag{5.7}$$

Derivation of expressions for the limiting sensitivity will be discussed in some detail in chapter 6.

5.1.5 Polarisation sensitivity (symbol: Π_{min})

The information carried by the polarisation of electromagnetic radiation is important, since it characterizes the physical conditions of the emission region. The polarisation is defined by using four so-called *Stokes parameters*. The detection system may only be sensitive to one polarisation component (such as a radio dipole antenna) and the telescope (optics, waveguides) may potentially alter it. The polarisation sensitivity is therefore defined as the accuracy to which the Stokes parameters can be measured.

Consider a monochromatic wave. This wave is always fully elliptically polarised and can be seen as a superposition of two orthogonal linearly polarised waves. In the following figure (taken from Rohlfs (1986)) the electric vector is drawn. ξ and η are the major and the minor axes of the ellipse, x and y are the directions of the dipole receivers.



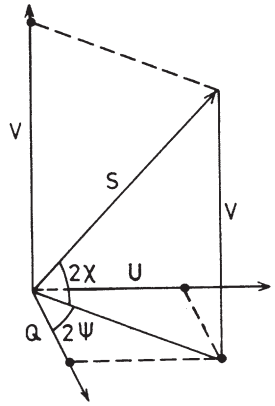
E_x and E_y are the observables. They are oscillating with the same frequency ν but with a phase difference δ. If $\delta > 0$ we call the wave right-handed polarised.
But in general the major axis ξ and the minor axes η of the ellipse do not coincide with the x- and y-dipoles of the receiver. The relation between the two coordinate systems is given by a simple linear rotation

$$\begin{aligned} E_\xi = E_a \cos(\tau + \delta) &= E_x \cos\psi + E_y \sin\psi \\ E_\eta = E_b \sin(\tau + \delta) &= -E_x \sin\psi + E_y \cos\psi \qquad (0 \le \psi \le \pi) \end{aligned} \tag{5.8}$$

in which ψ is the angle over which the ξ-axis is rotated with respect to the x-axis. Another angle χ can now be defined:

$$\tan\chi = \frac{E_a}{E_b} \qquad (0 \le \chi \le \frac{1}{2}\pi) \tag{5.9}$$

In 1892 Poincaré showed that there exists a one-to-one relation between polarisation states of a wave and points on a sphere with radius $E_a^2 + E_b^2$. The angles 2ψ and 2χ can be seen as the longitude and the lattitude of the sphere. The three coordinates of each point are the Stokes parameters Q, U and V, see the following figure (taken from Rohlfs (1986)).



Points on the equator represents linear polarised waves ($V = 0$), the north pole represents a right-circular polarised wave and the south pole a left-circular. The Stokes parameters are defined as

$$I = E_a^2 + E_b^2 \tag{5.10}$$
$$Q = I \cos 2\psi \cos 2\chi \tag{5.11}$$
$$U = I \sin 2\psi \cos 2\chi \tag{5.12}$$
$$V = I \sin 2\chi \tag{5.13}$$

By definition a monochromatic wave is fully polarised:

$$I^2 = Q^2 + U^2 + V^2 \tag{5.14}$$

Now the Stokes parameters can be expressed in the observable parameters E_1, E_2 and δ:

$$I = E_1^2 + E_2^2 \tag{5.15}$$
$$Q = E_1^2 - E_2^2 \tag{5.16}$$
$$U = 2\,E_1\,E_2 \cos\delta \tag{5.17}$$
$$V = 2\,E_1\,E_2 \sin\delta \tag{5.18}$$

5.2 Convolutions

5.2.1 General

In the preceeding paragraph several characteristic parameters have been defined which relate to the resolution of the astrophysical instrumentation in use. Although this gives a handle on capability and potential performance, for quantitative handling of the data full account needs

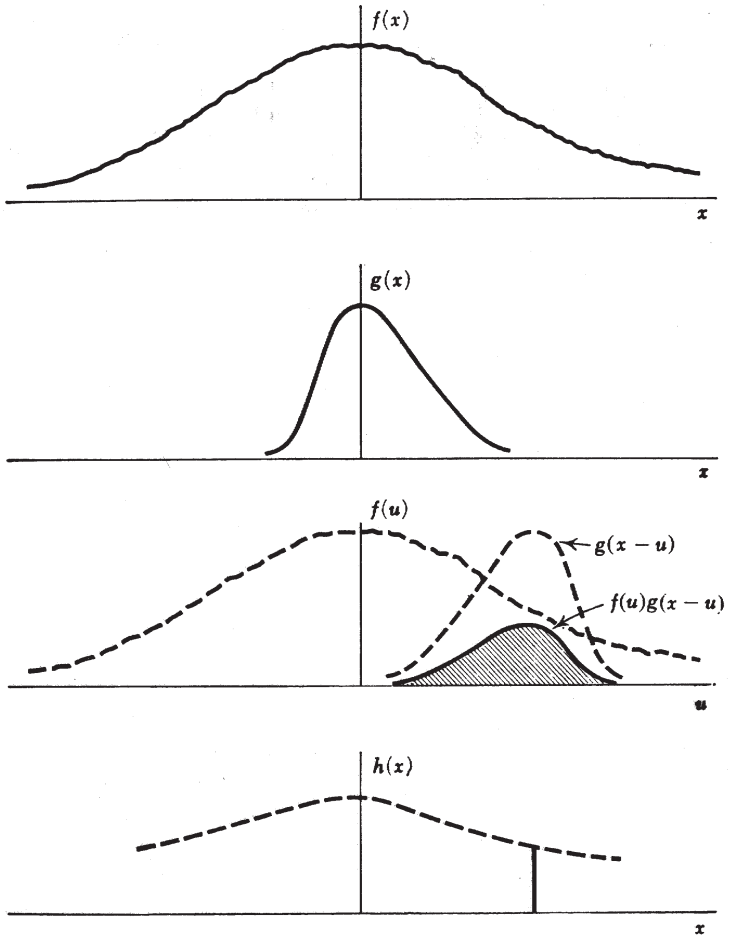


Figure 5.2: *The convolution h at point x (lower panel) of two functions f and g is equal to the integral of $f(u)$ and $g(x-u)$. Note that $g(x-u)$ is mirrored with respect to $g(x)$.*

to be taken of the influence of the instrument response on the incoming information carriers. This process can be described with the aid of convolutions: the measurement result can be described by a convolution of the measurand under study and the response function of the measurement device.

The convolution $h(x)$ of two functions $f(x)$ and $g(x)$ is:

$$h(x) = f(x) * g(x) = \int_{-\infty}^{+\infty} f(u) \cdot g(x-u)\, du \tag{5.19}$$

In words: a convolution describes the impact of a measuring device, described by $g(x)$, when it takes a weighted mean of some physical quantity $f(x)$ over a narrow range of the variable x, see figure 5.2. The simplest form of $g(x)$ is the window function $\frac{1}{\Delta x_0}\Pi\left(\frac{x}{\Delta x_0}\right)$, for which

the convolution integral reduces to the expression for the so-called *running average*:

$$\frac{1}{\Delta x_0}\int_{-\infty}^{+\infty} f(u)\Pi\left(\frac{x-u}{\Delta x_0}\right)du \;=\; \frac{1}{\Delta x_0}\int_{x-\frac{1}{2}x_0}^{x+\frac{1}{2}x_0} f(u)du \tag{5.20}$$

Convolutions are commutative, associative and distributive, i.e.:

$$f(x)*g(x) \;=\; g(x)*f(x) \tag{5.21}$$
$$f(x)*(g(x)*h(x)) \;=\; (f(x)*g(x))*h(x) \tag{5.22}$$
$$f(x)*(g(x)+h(x)) \;=\; f(x)*g(x)+f(x)*h(x) \tag{5.23}$$

5.2.2 Fourier transforms

Convolutions can be easily and elegantly handled by employing the Fourier transform technique. The Fourier transform of a function $g(x)$ exists if it satisfies the following conditions:

1. The integral of $|g(x)|$ exists, i.e. $\int_{-\infty}^{\infty}|g(x)|\,dx$ is convergent.
2. Discontinuities in $g(x)$ are finite.

The Fourier transform $\mathcal{F}(g(x))$ of $g(x)$ is called G and is a function of the Fourier domain variable s:

$$\mathcal{F}(g(x)) = G(s) = \int_{-\infty}^{+\infty} g(x)\; e^{-2\pi isx}\; dx \tag{5.24}$$

The inverse transform is given by

$$g(x) = \int_{-\infty}^{+\infty} G(s)\; e^{2\pi isx}\; ds \tag{5.25}$$

For a function $h(t)$ depending on time, its Fourier transform will depend on the frequency variable f:

$$H(f) = \int_{-\infty}^{+\infty} h(t)\; e^{-2\pi ift}\; dt \tag{5.26}$$

Sometimes the frequency variable f is replaced by the angular frequency $\omega = 2\pi f$, yielding

$$H(\omega) \;=\; \int_{-\infty}^{+\infty} h(t)\; e^{-i\omega t}\; dt \tag{5.27}$$

$$h(t) \;=\; \frac{1}{2\pi}\int_{-\infty}^{+\infty} H(\omega)\; e^{i\omega t}\; d\omega \tag{5.28}$$

which lacks the symmetry of the expressions in f.

A Fourier transform pair is often indicated in a symbolic way by a double arrow: $g(x) \Leftrightarrow G(s)$ or $h(t) \Leftrightarrow H(f)$.

A particular useful theorem from Fourier theory states that the Fourier transform of the convolution of two functions equals the product of the Fourier transforms of the individual functions

$$\mathcal{F}(f * g) = \mathcal{F}(f) \cdot \mathcal{F}(g) \qquad \text{or, in shorthand :} \quad f * g \Leftrightarrow F(s) \cdot G(s) \tag{5.29}$$

This is called the *convolution theorem*. Other useful and frequently applied theorems are

$$f(ax) \quad \Leftrightarrow \quad \frac{1}{|a|} F\left(\frac{s}{a}\right) \tag{5.30}$$

$$f(x) + g(x) \quad \Leftrightarrow \quad F(s) + G(s) \tag{5.31}$$

$$f(x-a) \quad \Leftrightarrow \quad e^{-2\pi i a s} F(s) \quad \textit{Shift theorem} \tag{5.32}$$

$$f(x)\ \cos \omega x \quad \Leftrightarrow \quad \frac{1}{2} F\left(s - \frac{\omega}{2\pi}\right) + \frac{1}{2} F\left(s + \frac{\omega}{2\pi}\right) \tag{5.33}$$

$$f(x) * f^*(-x) \quad \Leftrightarrow \quad |F(s)|^2 \tag{5.34}$$

$$f'(x) \quad \Leftrightarrow \quad 2\pi i s\ F(s) \tag{5.35}$$

$$\frac{d}{dx}(f(x) * g(x)) \quad = \quad f'(x) * g(x) \ = \ f(x) * g'(x) \tag{5.36}$$

$$\int_{-\infty}^{+\infty} |f(x)|^2\ dx \quad = \quad \int_{-\infty}^{+\infty} |F(s)|^2\ ds \quad \textit{Parseval's theorem} \tag{5.37}$$

$$\int_{-\infty}^{+\infty} f(x) g^*(x)\ dx \quad = \quad \int_{-\infty}^{+\infty} F(s) G^*(s)\ ds \tag{5.38}$$

$$f \text{ and } g \text{ real} \quad \int_{-\infty}^{+\infty} f(x) g(-x)\ dx \quad = \quad \int_{-\infty}^{+\infty} F(s) G(s)\ ds \tag{5.39}$$

5.3 Instrument response and data sampling

5.3.1 Response fuctions

In observational astrophysics it is often necessary to have a full and quantitative knowlegde of the angular and spectral reponse functions in order to perform a proper *deconvolution* of the measurement data to extract optimally the characteristics of the information source.

Consider the example of a spectrometer which possesses a response function $R(\lambda, \lambda')$. In many cases, the *smearing* effect (or *blurring*) of the spectrometer is solely a function of the wavelength difference $\lambda' - \lambda$, i.e. $R(\lambda, \lambda') = R(\lambda' - \lambda)$. If the information source emits a spectrum $S(\lambda)$, the source function, the measured distribution $M(\lambda)$ at the output of the spectrometer is given by the convolution:

$$M(\lambda) = \int_{-\infty}^{+\infty} S(\lambda') R(\lambda - \lambda')\ d\lambda' = \int_{0}^{\infty} S(\lambda') R(\lambda - \lambda')\ d\lambda' \tag{5.40}$$

since $S(\lambda) = 0$ for $\lambda \leq 0$. Applying the convolution theorem, with $S(\lambda) \Leftrightarrow S(s)$, $R(\lambda) \Leftrightarrow R(s)$ and $M(\lambda) \Leftrightarrow M(s)$

$$S(s) = \frac{M(s)}{R(s)} \tag{5.41}$$

The spectrum of the source can be reconstructed by deconvolution of the measured spectrum. This is accomplished through the inverse Fourier transform:

$$S(\lambda) = \mathcal{F}^{-1}\left[\frac{M(s)}{R(s)}\right] \tag{5.42}$$

To illustrate the notions just described, consider a spectrometer with a Gaussian response function:

$$R(\lambda) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{\lambda^2}{2\sigma^2}} \tag{5.43}$$

Suppose an infinitely narrow spectral line at wavelength λ_0 is measured: $S(\lambda) = S_0 \cdot \delta(\lambda - \lambda_0)$. The convolution of $S(\lambda)$ and $R(\lambda)$ gives the measured result, which is (of course) a Gaussian distributed profile around λ_0:

$$M(\lambda) = \frac{S_0}{\sqrt{2\pi}\sigma}\, e^{-\frac{(\lambda-\lambda_0)^2}{2\sigma^2}} \tag{5.44}$$

Suppose an observer has measured $M(\lambda)$ and $R(\lambda)$ is known. He can now reconstruct $S(\lambda)$. First the Fourier transform of $M(\lambda)$ is taken:

$$\begin{aligned}
M(s) &= \int_{-\infty}^{+\infty} M(\lambda)\, e^{-2\pi i s\lambda}\, d\lambda \\
&= \int_{-\infty}^{+\infty} \frac{S_0}{\sqrt{2\pi}\sigma}\, e^{-\frac{(\lambda-\lambda_0)^2}{2\sigma^2}}\, e^{-2\pi i s\lambda}\, d\lambda \\
&= \frac{S_0}{\sqrt{2\pi}\sigma}\, e^{-2\pi i s\lambda_0 - 2\pi^2 s^2\sigma^2} \int_{-\infty}^{+\infty} e^{-\frac{(\lambda-\lambda_0)^2}{2\sigma^2} - 2\pi i s(\lambda-\lambda_0) + 2\pi^2\sigma^2 s^2}\, d\lambda \\
&= \frac{S_0}{\sqrt{2\pi}\sigma}\, e^{-2\pi i s\lambda_0 - 2\pi^2 s^2\sigma^2} \int_{-\infty}^{+\infty} e^{-\pi\lambda'^2}\sqrt{2\pi}\sigma\, d\lambda' \\
&= S_0\, e^{-2\pi i s\lambda_0 - 2\pi^2 s^2\sigma^2}
\end{aligned} \tag{5.45}$$

where $\lambda' = \frac{\lambda-\lambda_0}{\sqrt{2\pi}\sigma} + i\sqrt{2\pi}\sigma s$. Analogously

$$R(s) = e^{-2\pi^2 s^2\sigma^2} \tag{5.46}$$

(Note that the Fourier transform of a Gaussian function is again Gaussian.) Applying equation 5.42 gives

$$S(\lambda) = \mathcal{F}^{-1}[S_0\, e^{-2\pi i s\lambda_0}] = S_0 \cdot \delta(\lambda - \lambda_0) \tag{5.47}$$

i.e. the observer has recovered the exact form of the spectral source function.

Consider now a spectrometer with a Gaussian response function with width σ. If σ is large, i.e. the line spread function is wide, the Fourier domain only contains low frequencies as can be seen from equation 5.46 and figure 5.3. By decreasing σ, the range of frequencies in the Fourier domain increases and consequently the range of *spectral frequencies* over which the measurement provides information. The $\ln R(s)$ is a parabola, which tends to a δ-function for the limit $\sigma \to \infty$ (no spectral resolution) and becomes flat for $\sigma \downarrow 0$ (perfect transmission at all frequencies).
Obtaining $S(\lambda)$ by this method is, in practise, complicated by two factors: *data sampling* and *noise*. This will be adressed in the following paragraphs.

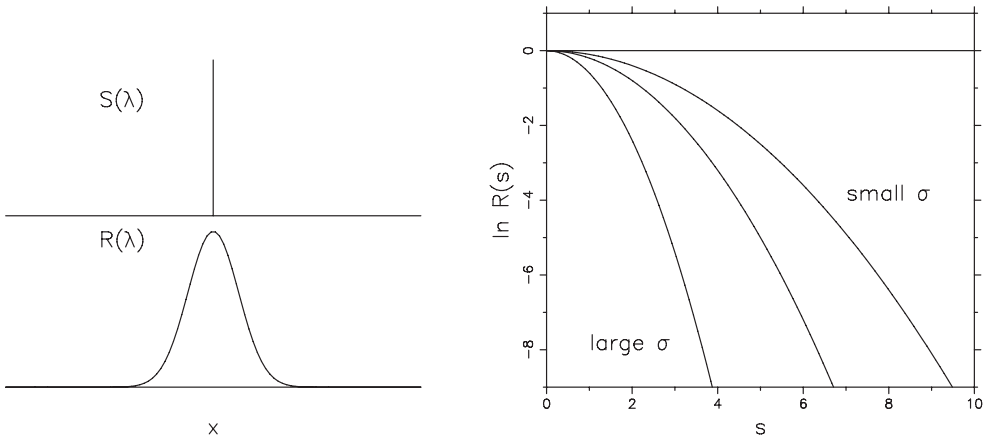


Figure 5.3: *The line $S(\lambda)$ is widened by the response function $R(\lambda)$. For a broad (narrow) response the Fourier transform of the measured function $M(s)$ contains information over a narrow (broad) range of frequencies.*

5.3.2 Discrete measurement intervals, the Nyquist criterion

First of all the measurement distribution $M(\lambda)$ is not a continuous function as assumed in equation 5.40, but is always sampled in discrete intervals or bins and is also not available over the whole interval from $-\infty$ to ∞.
The problem of discrete intervals can be overcome by using the discrete version of the Fourier transform.
Suppose there are N consecutive sampled values $g(x_k)$ with $x_k = k\tau$ $(k = 0, 1, 2, ..., N-1)$, the sampling interval is τ. With N *input numbers*, evidently no more than N independent *output numbers* can be produced. Estimates of the Fourier transform $G(s)$ can now be sought at discrete frequency values $s_n = \frac{n}{N\tau}$ $(n = -N/2, ..., N/2)$. The extreme values of n correspond in this case to the lower and upper limits of the Nyquist critical frequency (see later in this paragraph). Notice that in this case n takes $N+1$ rather than N values, this is because the two extreme values of n are not independent (i.e. they are equal), this reduces the count of independent numbers to N.

$$G(s_n) = \int_{-\infty}^{+\infty} g(x)\, e^{-2\pi i s_n x} dx \approx \sum_{k=0}^{N-1} g(x_k)\, e^{-2\pi i s_n x_k}\, \tau = G_D(s_n) \tag{5.48}$$

Substitution of x_k and s_n yields:

$$G_D(s_n) = \tau \sum_{k=0}^{N-1} g(x_k)\, e^{-2\pi i k n/N} \tag{5.49}$$

where $G_D(s_n)$ is the n^{th} value of the discrete transform of $g(x)$.
The inverse Fourier tranform, which recovers the set of sampled $g(x_k)$ values exactly from $G_D(s_n)$ is given by:

$$g(x_k) = \frac{1}{N} \sum_{n=0}^{N-1} G_D(s_n)\, e^{2\pi i k n/N} \tag{5.50}$$

N.B. Outside the measurement range the function values are set to zero.
Using discrete Fourier transforms instead of continuous transforms does not lead to loss of

information, provided the interval between sampling points or the bin width τ satisfies the *Nyquist criterion* for optimum sampling. This can be qualitativily understood in the following way. Any physical measurement system has a finite frequency response, therefore the measured distribution $M(x)$, constituting some function of a parameter x, is contained in bandwidth, i.e. the Fourier transform $M(s) \Leftrightarrow M(x)$ is a bandlimited function, characterized by a maximum cut-off frequency s_{max}, also called the critical s_{cr} or Nyquist frequency. In the case of 'Gaussian' response for instance: the frequencies will never be distributed purely Gaussian since no physical system transmits the tail frequencies up to ∞.
Nyquist (and Shannon) established a theorem for optimum sampling of bandlimited observations. This theorem states that no information is lost if sampling occurs at intervals (or in bins) $\tau = \frac{1}{2s_{cr}}$. The formal derivation of this theorem will not be given here, it is treated in the follow-on course OAF2. Thus, the use of the discrete Fourier transform causes no loss of information, provided that the sampling frequency $\frac{1}{\tau}$ is twice the highest frequency in the continuous input function (i.e. the source function convolved with the response function). The maximum frequency s_{max} that can be determined for a given sampling interval equals therefore $\frac{1}{2\tau}$. If the input signal is sampled too slowly, i.e. if the signal contains frequencies higher than $\frac{1}{2\tau}$, then these cannot be determined after the sampling process and the finer details will be lost. More seriously however, the higher frequencies which are not resolved will beat with the measured frequencies and produce spurious components in the frequency domain below the Nyquist frequency. This effect is known as *aliasing* and may give rise to major problems and uncertainties in the determination of the source function.

Example: *Consider again the above spectrometer with the Gaussian response function. What is the proper sampling interval for the measured function $M(\lambda)$?*
As stated before: no physical system transmits the tail frequencies up to ∞. Suppose that the frequencies of the Gaussian shaped Fourier transform $R(s) = e^{-2\pi^2\sigma^2 s^2}$ are covered out to three standard deviations in the s domain, i.e. the exponent $2\pi^2\sigma^2 s_{cr}^2$ equals $\frac{9}{2}$ so that $s_{cr} \approx \frac{1}{2\sigma}$. As is clear from the right panel in the figure in section 5.3.1, the power in the Fourier domain above this value of s_{cr} can be neglected. The Nyquist frequency $s_{cr} = \frac{1}{2\sigma}$ can be considered as appropriate and since $s_{cr} = \frac{1}{2\tau}$ the proper sampling interval τ (or bin width) for $M(\lambda)$ in case of a Gaussian line spread function with width σ is:

$$\tau = \sigma \tag{5.51}$$

If sampling (binning) is performed at a higher frequency the information in $M(\lambda)$ is oversampled, at lower frequencies undersampled. It is important to note that although oversampling ensures the transfer of all information, it also leads to statistical dependence between adjacent points (bins). This implies that standard statistical tests, like the χ^2-test or other likelihood tests, do not apply anymore since they assume statistically independent samples.

5.3.3 Noise

Secondly the presence of noise (both intrinsic to the signal and background noise) and disturbances will produce ambiguities in the values derived for $S(\lambda)$. Handling the errors and their

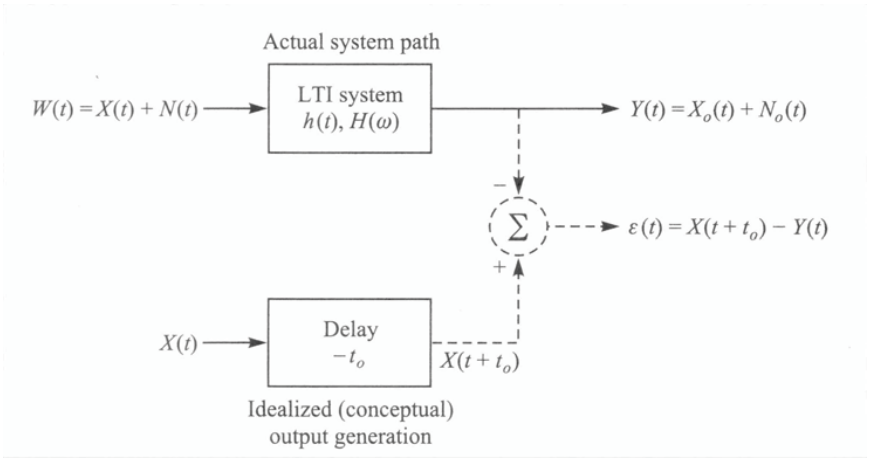


Figure 5.4: *Operation that defines the Wiener filter problem*

propagation is a major subject in deconvolution processing to arrive at reliable confidence intervals for the physical parameters estimated for the radiation source. Sophisticated filter algorithms have been developed to optimally reduce the influence of noise. Random noise can be most efficiently reduced by employing so-called optimal or *Wiener* filters. If applied to a stochastic process, the Wiener filter is the optimum result if the filter is designed such that its output is a good estimate of either the past, the present or the future value of the input signal. The basic problem to be addressed is depicted in figure 5.4. The input signal $W(t)$ is a stochastic process comprising the sum of a signal component $X(t)$ and a noise component $N(t)$:

$$W(t) \;=\; X(t) \;+\; N(t) \tag{5.52}$$

The system is assumed to be linear with a tranfer function $H(f)$, being the Fourier transform of the impulse response function $h(t)$. The output of the system is denoted $Y(t)$.
In general , $H(f)$ is selected such that $Y(t)$ is the best possible estimate of the input signal $X(t)$ at the time $t + t_0$, that is the best estimate of $X(t + t_0)$. If $t_0 > 0$, $Y(t)$ is an estimate of a *future* value of $X(t)$ corresponding to a *prediction filter*. If $t_0 < 0$, $Y(t)$ is an estimate of the *past* value of $X(t)$, corresponding to a *smoothing filter*. If $t = 0$, $Y(t)$ is an estimate of the current value of $X(t)$.
If $Y(t)$ differs from the desired true value of $X(t + t_0)$, the error is:

$$\epsilon(t) \;=\; X(t + t_0) - Y(t) \tag{5.53}$$

This error is illustrated conceptually in figure 5.4 by dashed lines. The optimum filter will be chosen so as to minimize the mean-squared value of $\epsilon(t)$:

$$\text{minimize} \quad \mathbf{E}\left\{\epsilon^2(t)\right\} \;=\; \mathbf{E}\left\{[X(t + t_0) - Y(t)]^2\right\} \tag{5.54}$$

We shall not give the derivation of the expression for the optimum filter transfer function, but only give the result for the case of an uncorrelated input signal $X(t)$ and noise $N(t)$:

$$H_{opt}(f) \;=\; \frac{S_X(f)}{S_X(f) \;+\; S_N(f)}\, e^{2\pi i f t_0} \tag{5.55}$$

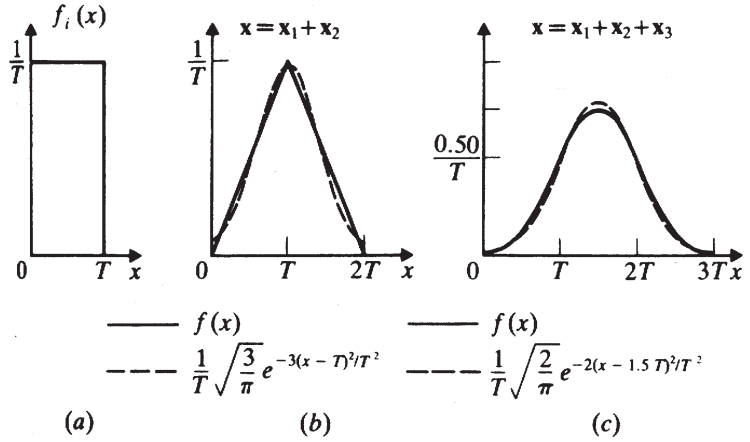


Figure 5.5: *Example of central limit theorem: window function and two consecutive selfconvolutions. The last one already resembles closely a Gaussian function.*

in which $S_X(f) = |X(f)|^2$ and $S_N(f) = |N(f)|^2$ represent the so-called *power spectral densities* of the input signal $X(t)$ and the noise $N(t)$ respectively.
Note: Let us check the value of $H_{opt}(f)$ for a noise free input, i.e. $S_N(f) = 0$. Equation (5.55) then reduces to:

$$H_{opt}(f) \;=\; e^{2\pi i f t_0} \tag{5.56}$$

This expression corresponds to an ideal delay line with delay $-t_0$ (recall the *shift theorem*!). If $t_0 > 0$, corresponding to prediction, we require an unrealizable negative delay line. If $t_0 < 0$, corresponding to a smoothing filter, the required delay is positive and realizable. $H_{opt}(f) = 1$ follows for $t_0 = 0$, this result is intuitively obvious.

5.4 The Central Limit Theorem

If a large number of functions are convolved together, the resultant function becomes increasingly smooth. Consider now n independent random variables X_i, each distributed according to a stationary probability density function $p_{X_i}(x)$. The sum X of these variables, i.e. $X = \sum_n X_i$ has a mean $\mu = \sum_n \mu_i$ and a variance $\sigma^2 = \sum_n \sigma_i^2$. Moreover, the probability density function of the sum X follows from the convolution of the individual probability densities:

$$p_X(x) = p_{X_1}(x) * p_{X_2}(x) * \;\cdots\; * p_{X_n}(x) \tag{5.57}$$

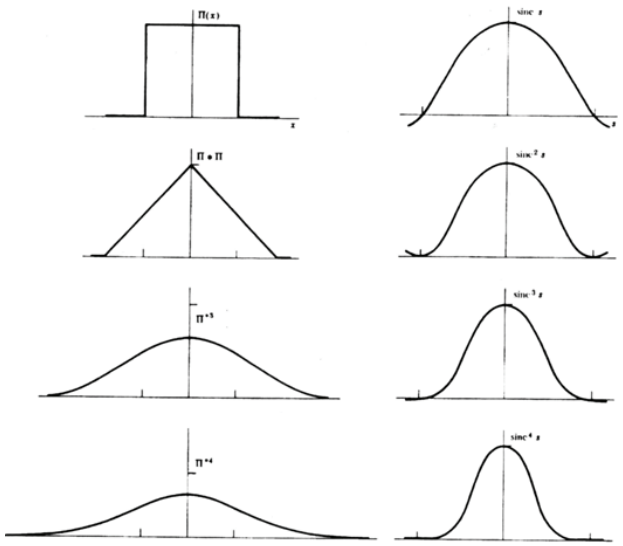
The central limit theorem (CLT) in its general form states that under applicable conditions the convolution of n functions (not necessarily the same) is a Gaussian function whose mean value is μ and whose variance is σ^2 plus a remainder which vanishes as $n \to \infty$:

$$\lim_{n\to\infty} p_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2} \tag{5.58}$$

The central limit theorem is basically a property of convolutions: the convolution of a large number of positive functions is approximately Gaussian. An example is given in figure 5.5, which shows that two self-convolutions of the window function $\frac{1}{T}\Pi\left(\frac{x-\frac{1}{2}T}{T}\right)$ already produces a

function which closely approximates a Gaussian density distribution. A starting condition for the CLT to hold in the case of repeated self-convolution is the requirement that the probability density function $p(x)$ or its Fourier transform can be approximated by a parabolic function in a small region around its maximum value.

Most physical measurement systems comprise a chain of elements, each contributing noise and changing response through it particular response function. The central limit theorem shows that in many cases these successive convolutions lead to an integral response, which can be approximated by a Gaussian function.



This figure shows the window function and three successive selfconvolutions and the associated Fourier transforms. The window function $\Pi(x)$ *forms a Fourier pair with the function* $\mathrm{sinc}(x)$.
Successive self-convolutions of $\Pi(x)$ *yield successive self-multiplications of* $\mathrm{sinc}(s)$ *(application of the convolution theorem, i.e.* $\Pi^{*n} \Leftrightarrow \mathrm{sinc}^n(s)$*). In a small range* Δs *around* $s = 0$ *(maximum)* $\mathrm{sinc}(s)$ *can be approximated by the parabolic curve* $(1 - as^2)$*, the* n^{th} *power of* $\mathrm{sinc}(s)$ *is then approximated by* $(1 - as^2)^n$*. Taylor expansion of this function yields*

$$\mathrm{sinc}^n(s) \approx (1 - as^2)^n = e^{-nas^2} + \mathsf{rest\ term} \tag{5.59}$$

Keeping s *fixed and letting* $n \to \infty$ *causes the Gaussian (exponential) term to get continually narrower, its width goes as* $n^{-1/2}$*, and the rest term tends to zero. The Gaussian term in the* s*-domain has a Gaussian Fourier transform, i.e.* $\sqrt{\frac{\pi}{na}} \cdot e^{-\frac{\pi^2 x^2}{na}}$*, which continually becomes wider, a property that variances add under convolution. The interesting phenomenon that emerges from this illustration is the tendency towards Gaussian form under successive self-convolution and, also, under successive self-multiplication.*