Representation of Digital Signals

Why a special lecture?

from NMSOP- Bormann (2002)
Almost every analysis in Geophysics (and meanwhile also in Geology) uses digital data, successive spectral analysis and filtering.

Scherbaum, “Of poles and Zeroes” (1996)
Representation of Digital Signals

Richter et al. (2004)
I) Discrete Signals
   1) Discretization and Sampling Theorem
   2) Fourier Transform
   3) From Analog to Digital

II) Statistical Description of Signals
   1) Definition deterministic vs. stochastic (stationary)
   2) Some important statistical Moments
   3) The uncertainty principle

III) Spectral Estimation
    1) Non-Parametric Methods
    2) Parametric Methods
    3) Time-Frequency-Representation

IV) Digital Filter
    1) LTI Systems
    2) The concept of response and transfer functions
    3) The seismometer as filter
    4) Problems of FIR and IIR
I) Discrete Signals

1. Discretization - Sampling

A continuous Signal Function: \( x_a(t) \) taken at specific time steps \( T_s \) results in:

\[
x[n] = x_a(nT_s);
\]

\( T_s \) = sampling interval; \( f_s = \frac{1}{T_s} \) = sampling rate or sampling frequency

⚠️ Note! The amplitude values are still \( x_a(t) \in \mathbb{R} \)!!!
Different mathematical notation using the 1-Impulse:

$$\delta(t) = \begin{bmatrix} 1 \text{ if } t=0 \\ 0 \text{ sonst} \end{bmatrix};$$

and using a series of such “1-Impulses” describes the sampling:

$$\delta_{Ts}(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s)$$

This results in:

$$x[n] = x_a(t)\delta_{Ts}(t)$$
I) Discrete Signals

Discretization - Sampling

The Sampling Theorem

In order to describe a continuous signal or function complete and unique using amplitude values taken at discrete times $T_s$, the sampled signal \textbf{MUST NOT} HAVE energy above a certain frequency $f_s = \frac{1}{2T_s}$. This frequency is also called Nyquist-Frequency.

The corresponding continuous signal $x_a(t)$ could be reconstructed using a linear combination of the discrete function weighted by a function $\text{sinc}(t) = \frac{\sin(t)}{t}$:

$$x_a(t) = \sum_{n = -\infty}^{\infty} x[n] \text{sinc}(\pi f_s (t - n T_s))$$
2. Sampling - Fourier Transform

Definition:

\[
X_a(j\omega) = \int_{-\infty}^{\infty} x_a(t) e^{-j\omega t} \, dt \iff x_a(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_a(j\omega) e^{j\omega t} \, d\omega;
\]

Strictly only valid if:

the function is absolute integrable: \( \int |x_a(t)| \, dt \leq c < \infty \)

Latter point is not always the case in Geophysics!
## Important Properties of the Fourier Transform

<table>
<thead>
<tr>
<th>Time domain</th>
<th>Frequency domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>x(t) real</td>
<td>$X(-j\omega) = [X(j\omega)]^*$</td>
</tr>
<tr>
<td>x(t) imaginary</td>
<td>$X(-j\omega) = -[X(j\omega)]^*$</td>
</tr>
<tr>
<td>x(t) = x(-t) (x(t) even)</td>
<td>$X(-j\omega) = X(j\omega)$ (X(j\omega) even)</td>
</tr>
<tr>
<td>x(t) = - x(-t) (x(t) odd)</td>
<td>$X(-j\omega) = -X(j\omega)$ (X(j\omega) odd)</td>
</tr>
<tr>
<td>x(t) real and even</td>
<td>$X(j\omega)$ real and even</td>
</tr>
<tr>
<td>x(t) real and odd</td>
<td>$X(j\omega)$ imaginary and odd</td>
</tr>
<tr>
<td>x(t) imaginary and even</td>
<td>$X(j\omega)$ imaginary and even</td>
</tr>
<tr>
<td>x(t) imaginary and odd</td>
<td>$X(j\omega)$ real and odd</td>
</tr>
<tr>
<td>x(t) imaginary and odd</td>
<td>* complex conjugate</td>
</tr>
<tr>
<td>Time domain</td>
<td>Frequency domain</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>multiplication: $x_a(t)h_a(t)$</td>
<td>convolution: $\frac{1}{2\pi}X_a(j\omega) \bullet H_a(j\omega)$</td>
</tr>
<tr>
<td>convolution: $x_a(t) \bullet h_a(t)$</td>
<td>multiplication: $X_a(j\omega)H_a(j\omega)$</td>
</tr>
<tr>
<td>differentiation: $\frac{dx(t)}{dt}$</td>
<td>multiplication: $j\omega \cdot X(j\omega)$</td>
</tr>
<tr>
<td>integration: $\int x_a(t) \ dt$</td>
<td>multiplication: $\frac{1}{j\omega}X_a(j\omega)$</td>
</tr>
<tr>
<td>$x_a(t-a)$</td>
<td>$X_a(j\omega)e^{-j\omega a}$ for $a &gt; 0$</td>
</tr>
<tr>
<td></td>
<td>* complex conjugate</td>
</tr>
</tbody>
</table>
I) Discrete Signals  

**Sampling - Fourier Transform**

Parseval's Theorem:

\[
\int_{-\infty}^{\infty} |x_a(t)|^2 \, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X_a(j\omega)|^2 \, d\omega;
\]

Back to sampling process:

\[
FT \left( \sum_{n=-\infty}^{\infty} \delta(t-nT_s) \right) = \frac{2\pi}{T_s} \sum_{k=-\infty}^{\infty} \delta(\omega - k \frac{2\pi}{T_s}) = \Delta_T(j\omega)
\]

Using

\[
\omega_s = 2\pi f_s = \frac{2\pi}{T_s} \text{ this results in } \Delta_T(j\omega) = \frac{2\pi}{T_s} \sum_{k=-\infty}^{\infty} \delta(\omega - k\omega_s)
\]
I) Discrete Signals

Sampling - Fourier Transform

FT of a sampled signal can be represented by a convolution of $FT\{\Delta_T\}$ and $FT\{X_a(j\omega)\}$

$$X(j\omega) = \frac{1}{T_s} \sum_{k = -\infty}^{\infty} X_a(j\omega - k\omega_s)$$

What the heck does that mean????

⚠️ The sampled signal $x[n]$ will be periodic in frequency $\omega_s$ (sampling frequency). It follows that the continuous signal $x(t)$ can be reconstructed using only one period. Only valid if the sampling theorem is not violated and no energy above $\frac{\omega_s}{2}$ is present in the signal $x(t)$:

$$X_a(j\omega) = T_sX(j\omega)$$
The sampling theorem **MUST** be applied **BEFORE** the sampling process. Therefore an analog lowpass filter must be applied before sampling - regardless which sampling frequency is used. The corner frequency ($f_c$) of that filter should satisfy:

$$f_c = 0.4 \cdot f_s.$$
I) Discrete Signals

Sampling - Fourier Transform

a) FT of analog signal

b) FT of discrete signal (sampling theorem complied)

c) FT of discrete signal (sampling theorem violated)
Problem 1:
Assume we are sampling with 125 Hz without an analog lowpass. Estimate the alias frequencies of noise signals at 70 Hz, 120 Hz and 300 Hz!
Sampling Theorem complied
Sampling Theorem violated
I) Discrete Signals

Sampling - Fourier Transform

Consequence: error in reconstruction
3. A/D-Conversion

Decimal system:

\[ x_{(10)} = \sum_{i} d_i^{(10)} 10^i; \]

Example:

\[ 1024_{(10)} = 4 \cdot 10^0 + 2 \cdot 10^1 + 0 \cdot 10^2 + 1 \cdot 10^3 \]

<table>
<thead>
<tr>
<th>LSB</th>
<th>MSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Binary system:

\[ x_{(2)} = \sum_{i} d_i^{(2)} 2^i \]

Example:

\[ 512_{(10)} = 0 \cdot 2^0 + \ldots + 0 \cdot 2^8 + 1 \cdot 2^9 \]

represents “Little Endian”

<table>
<thead>
<tr>
<th>LSB</th>
<th>MSB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>000000001</td>
</tr>
</tbody>
</table>
A 16 bit A/D-converter could represent in principle $2^{16}$ output states in its maximum (values between $0 - (2^n - 1)$ are possible).

The LSB (least significant bit) or smallest step width of the A/D-converter (resolution) is defined by:

$$LSB = \frac{\text{Maximale Voltage}}{2^n} = Q.$$  

As the resolution is directly dependent on the number of bits, a n-bit A/D-converter has “n-bit” resolution. Unfortunately, there is no rule, which would specify a “critical” number of “must have” bits. It is simply like that: if we have more bits we will decrease the noise added to the signal.
I) Discrete Signals

A/D-Conversion

Figure 2.20: Conversion of an analog signal into a corresponding digital one involves quantizing both axes, sampling time and digitizing signal value. In the figure we see the original analog signal overlaid with the sampled time and digitized signal value grid. The resulting digital signal is depicted by the dots.

Figure 2.21: Conversion of an analog signal into the corresponding digital one with fewer digitizing levels. As in the previous figure the original analog signal has been overlaid with the sampled time and digitized signal value grid. However, here only 17 levels (about four bits) are used to represent the signal.

Figure 2.22: Conversion of an analog signal into the corresponding digital one with fewer digitizing levels. Once again the original analog signal has been overlaid with the sampled time and digitized signal value grid. Here only nine levels (a little more than three bits) are used to represent the signal.

Figure 2.23: Conversion of an analog signal into the corresponding digital one with fewer digitizing levels. Once again the original analog signal has been overlaid with the sampled time and digitized signal value grid. Here only five levels (about two bits) are used to represent the signal.
An equivalent important parameter of A/D conversion is the so called dynamic range:

\[ D = 20 \log_{10} \left( \frac{A_{\text{max}}}{A_{\text{min}}} \right) \]

and therefore

\[ D = 20 \log_{10} (2^n - 1) \approx n \log_{10} (2) = n \cdot 6 \]

**Note:** this definition intrinsically assumes proportionality to power (20*log) of the signal - **NOT** energy (10*log)!

16 bit A/D-converter: 90dB;
24 bit A/D-converter: 138 dB;

**Be aware of the sign!**
III) Statistic Description

1. Definition: deterministic vs. stochastic

![Diagram showing deterministic and stochastic signals](image-url)
III) Statistic Description

Definition: deterministic vs. stochastic

Deterministic signals are, at least in principle, absolute reproducible. If the boundary conditions are the same an event could be completely reproducible. Two earthquakes exactly at location and source mechanism will produce exactly identical seismograms (if the propagation properties did not change!). The propagation medium is modelled as a so called Linear-Time-Invariant Filter (LTI).

On the other hand seismic noise measurements used in microzonation analysis, volcanic tremor and microseisms defy themselves of this simple description. These signals could not be forecasted as well as exactly described.

To describe stochastic signals, we have to learn some theorems of probability

\[ Pr(x) \] and the probability density function: \[ Pr(x \leq a) = \int_{-\infty}^{a} f(x) \, dx. \]
III) Statistic Description

The complete description of a statistical parameter requires the estimation of probability of every event in the parameter space. However, this is not always possible and, fortunately, not always needed. Sometimes it is sufficient to describe the mean behavior of the variable.

2. Some Statistical Moments

a) Sample Mean

Example: Dice

$x$: random variable; the dice was thrown N-times; the number $k$ occurred $n_k$ times.

The **sample mean** is defined through:

$$
\langle x \rangle_N = \frac{1n_1 + 2n_2 + 3n_3 + 4n_4 + 5n_5 + 6n_6}{N}
$$
III) Statistic Description

Some Statistical Moments

Approximation for $N \gg 1$: $\frac{n_k}{N} \approx Pr(x=k)$ and therefore:

$$\langle x \rangle_N \approx \sum_{k=1}^{6} kPr(x=k)$$

The expectation value of the variable is defined by:

$$E\{x\} = \sum_k \alpha_k Pr(x=\alpha_k) \text{ bzw. } E\{x\} = \int_{-\infty}^{\infty} \alpha f_x(\alpha) \, d\alpha$$

**Problem 2:**
Can we expect an exact estimation of the expectation value using digital signals?
b) Variance

The variance describes the deviation of the actual value from the expectation value of a variable. It is defined by:

\[
\text{Var}\{x\} = \sigma^2 = E\{(x - E\{x\})^2\} = E\{x^2\} - E^2\{x\}
\]

c) Covariance and Correlation

Often we want to estimate the statistical relationship of two or more variables. A frequently asked questions is whether two variables are statistically independent or not:

Correlation:

\[
r_{xy} = E\{xy^*\}
\]

Covariance:

\[
c_{xy} = \text{Cov}\{x, y\} = E\{xy^*\} - m_x m_y^*
\]

* complex conjugate and \(m_x = E\{x\}, m_y = E\{y\}\)
Sometimes the correlation coefficient - the normalized covariance is used instead:

\[ \rho_{xy} = \frac{E\{xy^*\} - m_xm_y^*}{\sigma_x\sigma_y} \quad \text{with} \quad |\rho_{xy}| \leq 1; \]

**Definition:**
Two random variables are statistical independent if their joint probability density function is separable:

\[ f_{x,y}(\alpha, \beta) = f_x(\alpha)f_y(\beta) \]

A somewhat weaker but more frequently used formulation uses the separability of covariance or correlation:

If \( E\{xy^*\} = E\{x\}E\{y^*\} \) or \( r_{xy} = m_xm_y^* \) than the signals are called uncorrelated.
An additional property of uncorrelated signals is:

\[ Var\{x + y\} = Var\{x\} + Var\{y\} \]

If \( r_{xy} = 0 \), than the signals are called **orthogonal**.

Consequence: two signals with zero mean, which are uncorrelated are always orthogonal.

**d) Bias and Consistency**

Problem: only part of data are available (often the case in seismology/geophysics). Can we estimate the already defined statistical moments? **No!**

The deviation of the sample moments (e.g., mean and variance) is called **bias**.

Assume: the statistical moment \( \theta \) should be estimated from sample measurements of the random variable \( x_n \) with \( n = 1, \ldots, N \).
III) Statistic Description

Some Statistical Moments

Assume: the mean of $\hat{\theta}$ equals the true value $\theta$.

Than the bias of the estimate is defined by:

$$B\{\theta\} = \theta - E\{\hat{\theta}\}.$$ 

If $B\{\theta\} = 0$ the estimate is unbiased

The estimate is called asymptotic unbiased if:

$$\lim_{N \to \infty} E\{\hat{\theta}\} = \theta.$$ 

An estimate is consistent if it converge in some way towards its true value
Example: “Sample-Mean”

\[ m_x = \frac{1}{N} \sum_{n=1}^{N} x_n. \]

The expectation value is:

\[ E\{\hat{m}_x\} = \frac{1}{N} \sum_{n=1}^{N} E\{x_n\} = m_x. \]

Therefore the sample mean is a unbiased estimate.

The variance on the other hand:

\[ Var\{\hat{m}_x\} = \frac{1}{N^2} \sum_{n=1}^{N} Var\{x\} = \frac{\sigma_x^2}{N}. \]
III) Statistic Description

Some Statistical Moments

is a consistent estimate as its value converges towards zero if N approaches infinity.

e) Covariance- Autocorrelation and Cross correlation function

In order to describe a stochastic time series (also for deterministic signals), the definition of autocovariance, cross correlation and cross correlation function is often used. We will introduce these important analysis tools using 1D time series. The extension to n dimensions is easy, however:

**Autocovariance function:**
\[ c_{xx}(l) = E \{ (x(t) - m_x)(x(t + l) - m_x) \} ; \]

**Cross covariance function:**
\[ c_{xy}(l) = E \{ (x(t) - m_x) - (y(t + l) - m_y) \} ; \]

**Cross correlation function:**
\[ \rho_{xy}(l) = \frac{c_{xy}(l)}{c_{xy}(0)} \]

What does this mean? How can we use these function in our daily analysis?
To put it straight first: the covariance and correlation of deterministic and stochastic signals are formally different. However, this difference does not matter in our daily work!

Assume a periodic function: \( x(t) = A_0 \sin(\omega_0 t - \phi) \) with \( T_0 = \frac{2\pi}{\omega_0} \);

the autocovariance is then:

\[
    c_{xx}(l) = \frac{1}{T_0} \int_0^{T_0} A_0^2 \sin(\omega_0 t - \phi) \sin(\omega_0 (t + l) - \phi) \, dt = \frac{A_0^2}{2} \cos(\omega_0 l)
\]

It is easy to see that the resulting signal is again periodic with the same frequency. However, all phase information is lost!

This is different for a stochastic (random) signal. There will be only one maximum at time lag zero!

\[
|c_{xx}(l)| < c_{xx}(0) \text{ for } |l| > 0.
\]
The behavior of the autocovariance function for $l \to \infty$ gives also some information about the possibility to predict the corresponding signal!
III) Statistic Description

Some Statistical Moments
Definition of Stationary m-th Order Processes
A signal or process \( \{X(t_1 + l), \ldots, X(t_n + l)\} \) is called m-th order stationary if the statistical moments up to order m are existing and identical for all times \( t_1 \ldots t_n \) and all time lags \( l \).

Example: 2. order stationary process:

1) \( E\{x(t)\} = m_x = \text{const} \)
2) \( Var\{x(t)\} = \sigma^2 = \text{const} \)
3) \( c_{xx}(l) = E\{(x(t) - m_x)(x(t + l) - m_x)\} \) only dependent on time lag \( l \)
Properties of autocovariance and cross covariance function:

- the autocovariance and cross covariance function of real signals are real.
- the autocovariance and cross covariance function of stationary signals is only depending on time lag $l$. Different time segments will result in identical autocovariance.
- the autocovariance function contain information about the internal coherence (?) of different time segments within the signal $x(t)$.
- the autocovariance has its maximum at time lag 0. On the other hand, the cross covariance will have its maximum at arbitrary (dependent on the signals) time lag $l$. The cross covariance at time lag 0 corresponds to the mean of the product of both signals.
- the autocovariance function is even.
III) Statistic Description

Some Statistical Moments

The cross covariance and its normalized version the cross correlation functions are of great importance when analyzing the similarity of two signals or in estimating the time shift between two traces.

Example:
An earthquake signal \( x(t) \), recorded at two stations, is buried in noise \( n(t) \) at one station and, as the stations are at different locations, shifted by the travel time difference:
\[
x_1(t) = s(t); \quad x_2(t) = s(t - t_0) + n(t);
\]

The cross correlation function of those signal results in:
\[
\rho_{x_1x_2}(l) = \rho_{ss}(l + t_0) + \rho_{sn}(l). \quad \text{If: } |\rho_{ns}(l)| \ll \rho_{ss}(0) \quad \text{than: } \rho_{x_1x_2}(l) \approx \rho_{ss}(l + t_0)
\]
III) Statistic Description

Some Statistical Moments
Nice **but**, we can do a couple of things wrong! There are several pitfalls when applying the covariance functions. Therefore, we will hold a moment and look for it!

All given definitions assume infinite sequences or continuous signals. In reality measured signals are always finite. It is only possible to estimate the statistical properties (\(\rightarrow\) Bias)!

**Problem 3**: Give the formulas for mean, variance, covariance function and cross covariance function for finite, discrete time signals.
Pitfall 1: What would you expect cross correlating these two traces?
Cure and Pitfall 2: sliding window analysis
Cure and Pitfall 2: sliding window analysis
Pitfall 3: one-sided signals - Why that?
3. The uncertainty principle

We already know it from physics: it is not possible to resolve the impulse of a particle as well as its location in arbitrary precision at the same time. Same is true for digital signal processing:

It is not possible to increase time and frequency resolution at the same time. If we increase the frequency resolution, we MUST increase the corresponding time window (i.e., decrease the time resolution) and vice versa.

In its best time-frequency analysis will choose an “optimum” of Time/Frequency resolution.
III) Statistic Description

The uncertainty principle

2 spikes (0.6s separated)

(Bandpass: 1.0 - 15 Hz)

(Bandpass: 1.0 - 1.5 Hz)
The uncertainty principle in digital signal processing connects time- and frequency resolution in the following relation:

\[ \Delta f \Delta t \geq \frac{1}{4\pi} \quad \text{or} \quad \Delta \omega \Delta t \geq \frac{1}{2} \]

with \( \Delta \omega, \Delta f \) frequency resolution, and \( \Delta t \) time resolution.
Spectral Estimation

1. Non-Parametric Methods

We have already used the properties of Fourier Transform when discussing the effects of sampling a continuous signal. Mapping time domain signals into the corresponding frequency space is one of the most fundamental analysis tools in signal processing. To take a short cut, we will immediately start with discrete Fourier Transform based analysis.

First: What can we gain out of that FT-analysis? What does FT means mathematically?

Problem 4: What happens, when we estimate the cross covariance (NOT the function!) of a sinus and a cosine with the same frequency?
But wait: remembering definition II.2.c) a zero value means that sinus and cosine are orthogonal functions!

The FT is therefore nothing else but a convolution of the signal with a (infinitive) number of sinuses and cosines - also called basis functions. In other words, the signal is represented by a linear combination of sinuses and cosines. The functional space spanned by this “vectors” is called Hilbert space.
a) Discrete Fourier Transform (DFT)

Assume: aperiodic, discrete signal. The discrete Fourier Transform is defined as:

\[
X(j\omega) = \sum_{n=-\infty}^{\infty} x[nT_s]e^{-jnT_s\omega}
\]

Valid under the condition that the signal is square integrable:

\[
\sum_{n=-\infty}^{\infty} |x[nT_s]|^2 \leq c < \infty;
\]

The back transform is defined as:

\[
x[nT_s] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(j\omega)e^{jnT_s\omega} \ d\omega;
\]
NOTE:
1) sign convention in the argument of the exponential function;
2) FT of infinite length of a **discrete** signal will be **continuous**.
3) all properties of the FT are also valid in this case (Parseval and Convolution)

Be aware of fact 2)

⚠️ The FT (DFT) of a **discrete** and **finite** signal is defined as (IEEE):

\[
X[k\Delta f] = T_s \sum_{n=0}^{N-1} x[nT_s] e^{-j2\pi kn/N}
\]

and its back transform:
Spectral Estimation  

Non-Parametric Methods

The continuous frequency $\omega$ maps on $\frac{k}{T_s N}$ or $\frac{kf_s}{N}$. Here, $\frac{1}{NT_s} = \frac{1}{T}$ represents the window (length), we have transformed. It is immediately clear that the best frequency resolution $\Delta f$ is defined by the observation length of our signal:

$$\Delta f = \frac{1}{T}.$$ 

As shown before, a discrete signal will have a periodic spectrum. In this case the periodicity is in $\mathbf{N}$.
Amplitude and Phase of DFT:

Applying the DFT we map our signal \( x[nT_s] \) into the complex plane:
\[
X[k\Delta f] \in \mathbb{C}.
\]

An often used representation of the resulting complex series is in polar form:
\[
X[k\Delta f] = |X[k\Delta f]| e^{-j\Phi[k\Delta f]};
\]

The absolute value of the DFT is called amplitude-spectrum, while the quantity \( \Phi[k] \) stands for the phase-spectrum. The amplitude spectrum gives the amplitude present in the signal at a certain frequency. However, the phase spectrum of an aperiodic discrete time signal could be interpreted as:
\[ \Phi[k \Delta f] = 2\pi k \Delta f \tau; \]

with \( \tau \) time shift of the signal (regarding start of analysis window).
In other words, the time shift of the signal can be interpret as slope of the corresponding phase spectrum.

**This estimate is NOT depending on the sampling rate!!!**

**Note**: If the time shift of the signal is large, the phase exists as a multiple of $2\pi$ (Phase wrapping). This could be easily seen by:

\[
\Phi[k] = \text{atan} \left( \frac{\text{Im}[X[k\Delta f]]}{\text{Re}[X[k\Delta f]]} \right) . \quad \text{atan} \quad \text{is known to be periodic in} \quad 2\pi .
\]

Known Problem in: TDOA-estimation, InSAR, seismic Source decomposition, RADAR, GPS etc.
In order to avoid this $2\pi$ wrapping, so called **phase unwrapping** is applied. There are several techniques for doing so, but none of them are without drawbacks or doubts.
Windowing - Leakage

The finiteness of the signal bears an additional complexity. Selecting just a portion of a (physical) infinite signal is equivalent to a multiplication of the data with a box-car window:
The resulting signal can be written as:

\[ x_T[n] = x[n]w[n]. \]

The DFT of a box car window is simply:

\[
W(j\omega) = \sum_{n=-\infty}^{\infty} w[nT_s]e^{-j\omega nT_s} = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} e^{-j\omega nT_s} = 2 \frac{\sin\left(\frac{\omega T}{2}\right)}{\omega};
\]
Spectral Estimation

Non-Parametric Methods

Remember the rule that multiplication in time domain will result in convolution in frequency domain!
The convolution of the FT of a box car with a discrete periodic (cosine) function will result in:

\[
X_{Taper}[j2\pi f] = \int_{-\infty}^{\infty} S[jg] W[jf-jg] \, dg = \;
\]

\[
= \frac{T}{2} \left( \frac{\sin(\pi(f-f_0)T)}{\pi(f-f_0)T} + \frac{\sin(\pi(f+f_0)T)}{\pi(f+f_0)T} \right) \]

If time \( T \) is an even multiple \( n \) of the signals period \( \frac{1}{f_0} \) than:

\[
X_{Taper}[j2\pi f] = \frac{f \sin(2\pi fn(1/f_0))}{\pi(f^2-f_0^2)} ;
\]
Spectral Estimation

Non-Parametric Methods
We are of course not restricted to the boxcar window. We should choose a window, which has more “convenient” properties. Best would be a window function, which has a sharp main lobe and strongly suppressed side-lobes. Unfortunately, there is a trade off between sharpness and side-lobe suppression.
Spectral Estimation  Non-Parametric Methods

Boxcar

Barlett

Henning/Cosine

Hamming

Welch
Spectral Estimation

Non-Parametric Methods

Rechteck
Barlett/Dreieck
Henning/Cosinus
Hamming
Welch
The width of the main lobe has severe consequences. Let’s assume a signal, where two spectral lines at \( f \) and \( f + \Delta f \) exist. In order to detect both lines, the half width of the window must be smaller than the distance between the two spectral lines.

The half width of the window function is defined as width of the main lobe taken at 50% of its maximum.

For the box car that is:

\[
\frac{T \sin(\pi \Delta f_{Taper} T)}{\pi f T} = \frac{1}{2} T.
\]

and results in:

\[
T_{Boxcar} > \frac{1, 2}{\Delta f}; \quad T_{Barlett} > \frac{1, 8}{\Delta f}; \quad T_{Tukey} > \frac{2}{\Delta f}; \quad T_{Parzen} > \frac{2, 5}{\Delta f};
\]
Note: The boxcar window has the best frequency resolution! The actual frequency resolution will always be smaller than the formal $\Delta f = \frac{1}{T}$ (DFT).

Note: usually, today algorithms for computing the discrete FT use the so called FastFourierTransform (FFT), which is optimized (in some way) for computers. However, the data length (in points) must be given in power of 2 (e.g., Press et al., 1988, Numerical Recipes in Fortran (C; C++; Pascal)).
b) Cross spectral estimates, Convolution and Coherence

We have already learned how the convolution of two time traces is transformed in the frequency domain. The discrete version is defined as:

\[ g[mT_s] = \sum_{n=0}^{N-1} y[nT_s]x[(m-n)T_s] \]

for \( m = 0, 1, \ldots, N-1 \);

or using the properties of the DFT:

\[ g[mT_s] = \text{DFT}^{-1}\{G[k\Delta f]\} = \text{DFT}^{-1}\{Y[k\Delta f]X[k\Delta f]\}; \]

The phase properties can be estimated by:

\[ G[k\Delta f] = |Y[k\Delta f]|X[k\Delta f]| e^{j\Phi_y(k\Delta f)} e^{j\Phi_x(k\Delta f)} \]
Note a pitfall:
We already recognized that the FT of a finite, discrete signal will produce a periodic spectrum. This is also valid for a infinite, discrete signal, but the spectrum is only computed up to the Nyquist frequency as all frequencies above contain no new information. However, if we multiply two DFTs with length $M_1$ and $M_2$ where $M_1 < M_2$ and transform the resulting spectrum back into time using the length $M_1$ we will suffer from a so called “wrap around effect”. It appears that part of the convolution trace $> M_1$ is wrapped from the “beginning” into the convolution trace.
Spectral Estimation

Non-Parametric Methods

Diagram:

1. Test Signal

2. Time vs. Amplitude

3. Time vs. Amplitude (continued)
To be correct the resulting trace should have a spike at 11s (Why?). As the original trace is restricted to 10.24 s, the convolution results in a spike at position 0.76 s.

**Problem 5:** How many samples long should the zero padding be in the latter example in order to avoid the wrap around effect?

The **cross spectrum** of two discrete signals is defined as:

\[
C[k\Delta f] = Y[k\Delta f]X^*[k\Delta f]
\]

A not very precise but instructive interpretation of the cross spectrum is, that it estimates how much energy is present in **both** signals at a certain frequency.

A more important property will be obvious, if we rewrite the definition as:

\[
C[k\Delta f] = |Y[k\Delta f]| |X[k\Delta f]|e^{j\Phi_y(k\Delta f)} e^{-j\Phi_x(k\Delta f)}.
\]
Note: The cross phase spectrum is estimated by the difference of the individual phase spectra.

Remembering the relation $\Phi[k\Delta f] = 2\pi k\Delta f \tau$ it is easy to see that the phase of the cross spectrum contains the time difference between both signals.
Spectral Estimation Non-Parametric Methods
**Spectral Estimation**  

Using the properties of the DFT and the so called Wiener-Khintchine Theorem (we do not care about that at present) enables us to estimate the autocovariance and cross correlation function in a very effective way:

\[
c_{xy}[lT_s] = \frac{1}{2N+1} \sum_{n=\text{-}N+l}^{N-l} x[(n+l)T_s]y[nT_s]
\]

\[
= DFT^{-1}\{X[jk\Delta f]Y^*[jk\Delta f]\}
\]

\[
\rho_{xy}[lT_s] = \frac{1}{2N+1} \sum_{k=\text{-}N+l}^{N-l} x[(n+l)T_s]y[nT_s]
\]

\[
= DFT^{-1}\left\{ \frac{X[jk\Delta f]Y^*[jk\Delta f]}{C_{xx}^{1/2}C_{yy}^{1/2}} \right\}
\]
This leads us directly to a new quantity the so called **Coherency**:

\[
Coh_{XY}[jk\Delta f] = \frac{X[jk\Delta f]Y^*[jk\Delta f]}{C_{XX}^{1/2}C_{YY}^{1/2}} = \frac{Re[C_{XY}] + iIm[C_{XY}]}{C_{XX}^{1/2}C_{YY}^{1/2}};
\]

More commonly used is the so called **Coherence** - the absolute value of the coherency:

\[
CCoh_{XY}[jk\Delta f] = \frac{(Re[C_{XY}])^2 + (Im[C_{XY}])^2}{C_{XX}C_{YY}};
\]

The coherence is real - but what does it mean?
Let us compute the coherence of two signals:
non-averaged spectrum
Spectral Estimation

We observe, that the coherence is equal to one everywhere (actually that is not really surprising)!
Let’s start from scratch. The coherence should be one if two traces are identical (up to a certain phase shift) within a frequency band or conterminously one frequency and its phase are stable within both signals within a certain time window.

Consequently, we must either average our cross- and auto spectra within certain frequency bands (Daniell’s method) or we average the corresponding spectra of several time windows.
Using the Daniell approach (average within 2m+1 frequency steps) results in:

\[
\overline{C_{l}[jk\Delta f]} = \frac{1}{2m+1} \sum_{l=-m}^{m} C[j(k+l)\Delta f].
\]

The same average procedure must be also applied in case of the auto spectra:
Spectral Estimation

Non-Parametric Methods
The coherence is an important assessment tool for evaluating the similarity of two traces in frequency domain. This is especially important in array seismology when the coherent signal is buried in also coherent noise, acting in a different frequency range.
d) The Power Spectral Density (PSD)

The DFT (or more general FT) is only defined if the signal is square integrable. The overwhelming numbers of processes are violating this definition. However, these signals have infinite energy (density) but finite power (density). On the other hand, applying this statement to our data we neglect the finiteness of our data!

There is another reason for using power density, which is more important in daily life work: if we select a certain time window out of a random, stationary process and estimate its spectral content via a DFT, we suffer from a 100% variance in our estimate! I.e., applying the DFT for another time window results in a complete different spectrum:

Formal definition of power spectral density:

\[
P_X^t = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x^2(t) \ dt = \int_{-\infty}^{\infty} P(f) \ df;
\]
Spectral Estimation

Non-Parametric Methods

The PSD estimates the power/per frequency bin present in the signal. If we have a random, stationary and zero mean process we can adhere:

$$P_x^t = \sigma^2;$$
There are several ways for estimating the PSD, however, we will restrict ourself to the most simple one (regarding theory). Just for completeness we will introduce the

**Theorem of Wiener and Khintchine**

\[ c_{xx} = DFT^{-1}\{P(f)\} \text{ and vice versa } P(f) = DFT\{c_{xx}\}. \]

It follows that starting from the PSD, we can only reconstruct the autocovariance of the corresponding signal but **NOT** the signal itself.
PSD via Periodogram Estimate

In the most simple approach we use directly the DFT:

$$
\hat{P}_p(f) = \frac{T_s}{N} \left( \sum_{n=0}^{N-1} x_n e^{-i2\pi fnT_s} \right)^2 = \frac{T_s}{N} \left( \sum_{n=0}^{N-1} x_n e^{-i2\pi \frac{k_n}{N}} \right)^2 = \hat{P}[k\Delta f]
$$

However, as already noted, this estimate would have a variance of 100%. In order to reduce this variance, several modifications are on the market (Taper, averaging in frequency or time, etc.). We will focus on two of the various possibilities: The Daniell- and the Welch approach.

The Periodogram estimate intrinsically assumes a periodic repetition of the selected data outside the observed interval. Again we multiply the data to be analyzed with a box car window. Applying arbitrary window function we will deal with an expectation value of:

$$
E\{\hat{P}_p[k\Delta f]\} = P(f) \times \frac{|W[k\Delta f]|^2}{N};
$$
Spectral Estimation Non-Parametric Methods

and Bias:

\[ b\{\hat{P}_p[k\Delta f]\} = P(f) \times \frac{|W[k\Delta f]|^2}{N} - P(f) \]

Using a boxcar the window term will be (FT: \(\frac{\sin(x)}{x}\)):

\[ \lim_{N \to \infty} W[k\Delta f] \to 1. \]

The estimate of PSD using Periodogram is asymptotically bias free. But as the variance of a periodogram from a random process is:

\[ Var\{\hat{P}_p\} \geq (E\{\hat{P}_p\})^2 \]

The variance will be 100% if N approaches infinity. The Periodogram is not a consistent estimate.
There exists a trade-off between bias and variance reduction when analyzing finite data sequences using the periodogram approach.

Daniell-Estimate:
No modified taper is applied but the spectral estimates are averaged within frequency bins. This will reduce the variance significantly:

\[ \hat{P}_b(k\Delta f) = \frac{1}{2M+1} \sum_{m=-M}^{M} \hat{P}_p[m\Delta f] \]
Welch-Estimate
One of the best direct approaches. Here, segmentation in time, tapering, 50% overlapping data segments and averaging the resulting spectra are applied:

\[
P_{w}^{s}[k\Delta f] = \frac{T_{s}}{LU} \left| \sum_{l=0}^{L-1} x_{l + (s-1) \frac{L}{2}} e^{-j2\pi k\frac{l}{L}} \right|^2 \]

segments of length L and tapering

with \( U = \frac{1}{L} \sum_{l=0}^{L-1} u_{l}^2 \) the taper and finally averaging over S segments:

\[
\hat{P}_{w}[k\Delta f] = \frac{1}{S} \sum_{s} P_{w}^{s}[k\Delta f] 
\]

The variance reduction using a boxcar (not the best one! - leakage) would be:

\[
Var\{P_{w}[k\Delta f]\} = \frac{3}{2S} P(f) 
\]
Spectral Estimation

Non-Parametric Methods

Some words about confidence limits. They are used in order to quantify the quality of a spectral estimate.
Confidence intervals specify the upper and lower limit of the interval, in which regarding a certain probability, the true value of the spectral estimates will be found.

Estimation of confidence limits:

\[ \text{Prob}[x_1 \leq \hat{P}[k\Delta f] \leq x_2] = \beta \]

We must specify \( \beta = 1 - \alpha \) according the probability \( \frac{\alpha}{2} \) that the true \( P(f) \) is smaller than \( x_1 \) and \( 1 - \frac{\alpha}{2} \) that \( P(f) \) is smaller than \( x_2 \).

The bandwidth or spectral resolution using a periodogram and a box car taper results in: \( b_p = \frac{1}{NT_s} \) or segmenting in \( S \) segments of length \( L \): \( b_b = \frac{1}{LT_s} \).
Spectral Estimation  Non-Parametric Methods

The variable $\frac{\nu \hat{P}_p(f)}{P(f)}$ follows a $\chi^2$ distribution with $\nu$ degrees of freedom.

A single segment gives: $\nu = \frac{2E^2[\hat{P}_p(f)]}{Var[\hat{P}_p(f)]} \approx 2$;

Averaging S non-overlapping segments will result in $\nu = 2S$ degrees of freedom. Finally we can estimate the confidence interval of $P(f)$ with $1 - \alpha$ probability as:

$$\frac{\nu}{u^{\nu}} \hat{P}_p(f), \quad \frac{\nu}{u^{\nu}} \hat{P}_p(f)$$

$$u^{\nu} 2 |1 - \frac{\alpha}{2} | \frac{\nu}{u^{\nu}} 1 |\frac{\alpha}{2}$$

Which can be found in mathematical tables.

The Welch approach with its modified taper and overlapping segments results in slightly less degrees of freedom:

$$\frac{\nu}{2NTs} = b_w \text{ and } \nu = \frac{2E^2\{\hat{P}_w(f)\}}{Var\{\hat{P}_w(f)\}} = 1, 6S \text{ therefore } b_w = \frac{1, 16}{L}.$$
Finally some PSD values of common random processes:

white noise: $P(f) = P_0 \forall f$ with autocovariance function $c_{xx}[l] = P_0 \delta$;

red noise: $P(f) = \frac{\sigma^2}{\alpha} \sqrt{\frac{\pi}{\alpha^2}} e^{-\frac{\pi^2 f^2}{\alpha^2}}$ and $c_{xx}[l] = \sigma^2 e^{-\alpha^2 l}$;
2. Parametric Methods

We have already noticed that applying classical, non-parametric spectral estimators will result in a trade off between bias and variance reduction. A different class of spectral estimators is based on signal models (physical processes), which are generating the signal. As a by-product when estimating the parameters of these models we can also estimate the spectral content of the signal. We will focus only on one of the various existing methods to demonstrate the capabilities but also the drawbacks of that class of estimators.

AR - Estimate (similar to MEM)
An estimate of the PSD or the extrapolation of an autocovariance function will be more and more successful, if we have a deeper knowledge about the signal or the signal generating process. Using the theorem of Wold (1938) - all stationary signals can be separated in a deterministic (forecast) and a non-deterministic part - a whole class of signals “recursive processes” come to life. The most fundamental (stationary) process is called Autoregressive-Moving-Average (ARMA):


**Spectral Estimation**

**Parametric Methods**

\[
A \sum_{a=1}^{A} \alpha[a] y[i-k] + M \sum_{m=1}^{M} \beta[m] x[i-l] + x[i];
\]

(difference equation of ARMA-process), with the gaussian, random input signal \( x[i] \).

There are two special cases existing:

a) Moving Average (MA):

\[
y[i] = \sum_{m=1}^{M} \beta[m] x[i-m] + x[i];
\]

b) Autoregressive Model (AR):

\[
y[i] = - \sum_{a=1}^{A} \alpha[a] y[i-a] + x[i]
\]
Here, \( x[i] \) represents the non-deterministic part of the signal. The process is called \((A,M)\)-th order ARMA Process and AR-2 or MA-2 process, if \( A, M = 2 \).

Problem: Which process should be modelled?

A frequently used technique for estimating the signals parameter is based on the extrapolation of an AR autocovariance function. Let’s assume an A-th order AR-process. If we multiply the equation with \( y[i - a] \) it results in:

\[
c_{yy}[a] = \alpha_1 c_{yy}[a - 1] + \ldots + \alpha_A c_{yy}[a - A] + E\{y[i - a]x[j]\};
\]

(ACF)

We can re-write these equations for different lags a as “normal-equation” in matrix form:
If the ACF for $c_{yy}[0], \ldots, c_{yy}[A]$ are known, we can estimate the $A$ coefficients $\alpha[1], \ldots, \alpha[A]$.

After this parameter estimation, we can FT the equation and result in:

$$Y[j\omega] - Y[j\omega](\alpha[1]e^{-j\omega t} + \alpha[2]e^{-2j\omega t} + \ldots + \alpha[A]e^{-Aj\omega t}) = X[j\omega] \text{ or}$$

$$Y(j\omega) = \frac{1}{A[j\omega]} X[j\omega]$$

and therefore the PSD is given by:
Spectral Estimation

The nominator $\sigma_x^2$ represents the variance of the driving white noise (gaussian random process). This value must also be estimated while computing the coefficients.

\[
P_y[j\omega] = \frac{T_s \sigma_x^2}{\left| A \sum_{a=1}^{A} \alpha[a] e^{-2\pi faT_s} \right|^2};
\]
Let's have a look on two different AR-2 processes:
One of the problems when applying model based spectral estimators is hidden in the estimation of the process length (A & M). In the next example we will see, what happens if our process length is too small or too large.
There exist several different ways for fixing the length of the process in advance (data-driven). However, non is working really reliable. Of course this is different if we have some a priori knowledge of the physical process itself.

In order to estimate the filter (!) length, driven completely by the data, we use the definition of Burg (1967):

\[
P_{A+1} = \frac{1}{2(N-A)} \sum_{i=1}^{N-A} \left( \left( y[i+A] - \sum_{a=1}^{A} \alpha_A[a] y[i+A-a] \right)^2 + \left( y[i] - \sum_{a=1}^{A} \alpha_A[a] y[i+a] \right)^2 \right); \]

Here \( \alpha_A \) represents the coefficients estimated using a length \( A \).

If \( A \) is chosen correctly \( P_{A+1} \) represents the variance of a gaussian process.
Akaike’s **Final Prediction Error** (FPE):

\[ FPE(A) = \frac{N + A + 1}{N - A - 1} P_{A+1}; \]

That \( A \) will be used, at which the function \( FPE(A) \) has its (first!) minimum.

**Properties:**

- the **resolution** is, in principle, **infinite!**
- spectral estimates using a **too coarse** \( \Delta f \), (free to chose) will be **badly** resolved;
- the position of spectral peaks will be **shifted** to an unknown amount, if the **filter length** is wrong or **variable noise** conditions within the signal. When applying to **short data segments**, the result **depend** strongly on the **phase** of the signal.
- the most severe point is embedded in the estimate of the process length. If the **signal is noise and/or includes also harmonic components**, the computation of filter length \( A \) is **difficult** and **non-unique**.
3. Time-Frequency-Representation

Very often, not only the spectral content of a signal is the desired output but also its time-dependence. An adhoc method consists of a running segmentation of the data and successive FT of the corresponding segments. This is generally (with tapering and overlap) the procedure to estimate so called ShortTermFourierTransform (STFT):
Spectral Estimation

Time-Frequency-Representation

STFT.
Whatever, in this chapter all is about finding the optimum “time-frequency-atoms”. We will have a brief look onto the problem of optimum tapers as well as optimum overlap.

A general answer to the question about best taper was given by Gabor (1949). According to the uncertainty principle the best time-frequency-localization will be achieved using a gaussian function:

$$g_{\alpha}(t) = \frac{1}{2\sqrt{\pi\alpha}} e^{-\frac{t^2}{4\alpha}};$$

with $\alpha > 0 = \text{const}$. The **Gabor Transform** is given by:

$$G_{b}^{\alpha}\{x(t)\} = \int_{-\infty}^{\infty} (x(t)g_{\alpha}(t-b))e^{-j\omega t} \; dt;$$

The FT of the signal $x(t)$ is centered at $t = b$. The width of the taper is defined by a “damping” constant $\alpha$. This damping results in a time resolution of:

$$\Delta t = \sqrt{\alpha};$$
A different view on the use of the gaussian taper opens a complete new field in signal processing.

Rewriting the equation to:

\[ G_{\alpha_b, \omega}(t) = g_{\alpha}(t-b)e^{i\omega t} \]

the TF is defined as:

\[ G_\alpha \{ x(t) \} = \int_{-\infty}^{\infty} x(t)G_{\alpha_b, \omega}^*(t) \, dt = \langle x(t), G_{\alpha_b, \omega}(t) \rangle; \]

and therefore the GT consists in the inner product of two functions. The FT of this modified gaussian can be written as:

\[ \text{FT}\{ G_{\alpha_b, \omega_k}(t) \} = \sqrt{\frac{\alpha}{\pi}} e^{-\alpha(\omega - \omega_k)^2} \]

The gaussian is therefore also a gaussian in the frequency range!
Again some re-writing and choosing special parameters (octave bands!) results in an optimum localization in time and frequency!

Gabor Transform
Spectral Estimation

Time-Frequency-Representation

Gabor Transform - constantQ
Spectral Estimation  Time-Frequency-Representation

constant Q
As we have not averaged any spectrum, we can also display the phase properties.
This latter definition of the Gabor Transform and carefully chosen parameters (so called constant Q) end up in the new field of “Continuous Wavelet Transform” (Morlet, 1987).

**Continuous Wavelet Transform (CWT)**
The CWT is defined:

\[
W_{\psi}^{a,b}\{x(t)\} = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^\ast\left(\frac{t-b}{a}\right) \, dt;
\]

here \(a, b \in \mathbb{R}\) and \(a \neq 0\), with \(f \in L^2(\mathbb{R})\).

In other words, the Wavelet Transform is the inner product (convolution) of the signal with an analysis function. The analysis function, when orthogonal, forms again a Hilbert space!
The only difference to the constant Q filter introduced while focusing on the GT is its more general definition. Today, a complete library of different analyzing functions (wavelets) are existing with different properties and different advantages. Whatevover, this flexibility is at the same time also its drawback. In order to interpret the result of a CWT, the corresponding wavelet must be given. Different CWT with different wavelets can often not be compared directly.

It is also possible to define a coherence measure based on wavelet theory. The advantage is its scaling law, resulting in always constant averaging properties at different frequency bands.
Wavelet Coherence: