Chapter 2 :
Geophysical data processing

2.1 Introduction (p. 8)

Geophysical data :
measurment of some physical
quantity vs time and/or space

Ex : Gravity measurment
Magnetic
seismic
EM

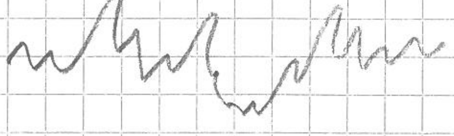
Geophysical data (GP-data) are the sum of
a) the responce from underlying geology
b) non-GP contrib
c) instrument noise

The job of the geophysicsist
* Isolate the responce from the
underlying geology
separate "signal from noise"

* Process / Interpret the signal
  How did the underlying geology look like

* GP-survey planning
  "Does one have sufficient signal-to-noise ratio"

2.2 Digitization of GP-data



Input signal: Cont. phys. signal

↓

Sensor — Hardware

↓

Digitization ⟹ AD-converter — Electronics (typically 24 bits)

↓

Computer

↓

Data.dat — File storage

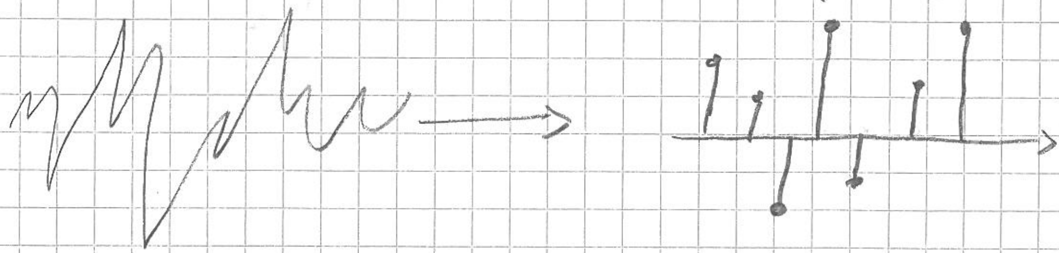
Continuous

Digital

$d(t)$

$d(t_i) = d_i \; , \; i \in N_0$

$t_i = t_0 + i \Delta t$



See Fig 2.2

Q : How accurate are discretely sampled data

A : It depends on the <u>dynamical range</u> and <u>sampling interval</u>.

① Dynamical range :

the ratio between the maximum to the minimum amplitude of a signal that can be rep. accurately.

$$\frac{A_{max}}{A_{min}}$$

Typically DR is expressed in decibel (dB)

Power : $P = |A|^2$

$$D_r = 10 \log_{10} \frac{P_2}{P_1}$$

$$= 20 \log_{10} \frac{|A_2|}{|A_1|}$$

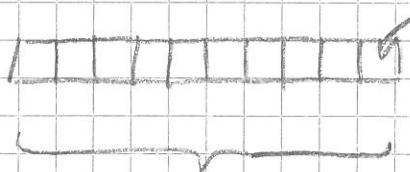
Ex : $A_1 = 1$; $A_2 = 1024$

$$D_r = 20 \log_{10} \frac{A_{max}}{A_{min}} = 60 \text{ dB}$$

— // —

In the computer numbers are represented in binary form :

a bit = 0,1



word length : # bits

Ex : $D_r = 84 \text{dB}$; what is the ampl. ratio

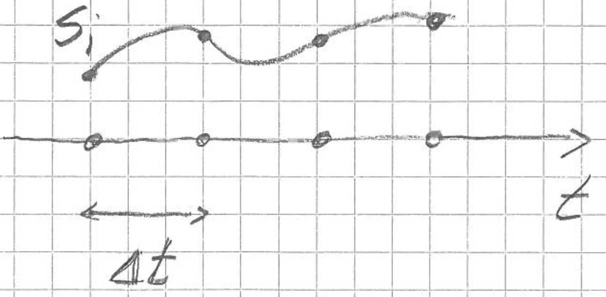
$$D_r = 20 \log_{10} \frac{A_{max}}{A_{min}}$$

$$\Rightarrow \frac{A_{max}}{A_{min}} = 10^{D_r/20} = 2^x$$

$$x = \frac{(D_r/20) \ln 10}{\ln 2} \simeq 14$$

15-bit words needed

② Sampling interval / Sampling freq.



Sampling interval (Δt):
time (space) difference between two consequtive samples

Sampling frequency:
\# of samples per unit time (or distance)

$$f_s = \frac{1}{\Delta t} \qquad \begin{matrix} (\text{or } \omega_s = 2\pi f_s) \\ (\quad k_s = 2\pi f_s) \end{matrix}$$

Q: How can we make sure that our discretely sampled signal d_i represents well the original continuous function

A: We can not in general !

This is only possible under certain requirements.

Nyquist - Shannon sampling theorem

Minimum of two samples per shortest period of the signal
$\Longrightarrow$ no loss of information

Nyquist frequency (critical freq.)

$$f_c = \frac{1}{2\Delta t} = \frac{f_s}{2} \qquad [f_N \text{ in the book}]$$

or

$$\omega_c = \frac{\pi}{\Delta t} \qquad (\omega = 2\pi f)$$

The highest freq component in an analog signal is its bandwidt

Let f_{max} be the maximum freq contained in an analog signal.
According to the sampling theorem, the analog signal is fairtfully represented by its discretly sampled series if

$$f_s \geq 2 f_{max}$$

or

$$2f_c \geq 2f_{max}$$

$$\Rightarrow \boxed{f_c \geq f_{max}} \qquad (\omega_c > \omega_{max})$$

Ex : $\Delta t = 2\ ms = 2\ 10^{-3} s$

$$f_s = \frac{1}{\Delta t} = 500\ Hz$$

$$f_c = \frac{1}{2\Delta t} = \frac{f_s}{2} = 250\ Hz$$

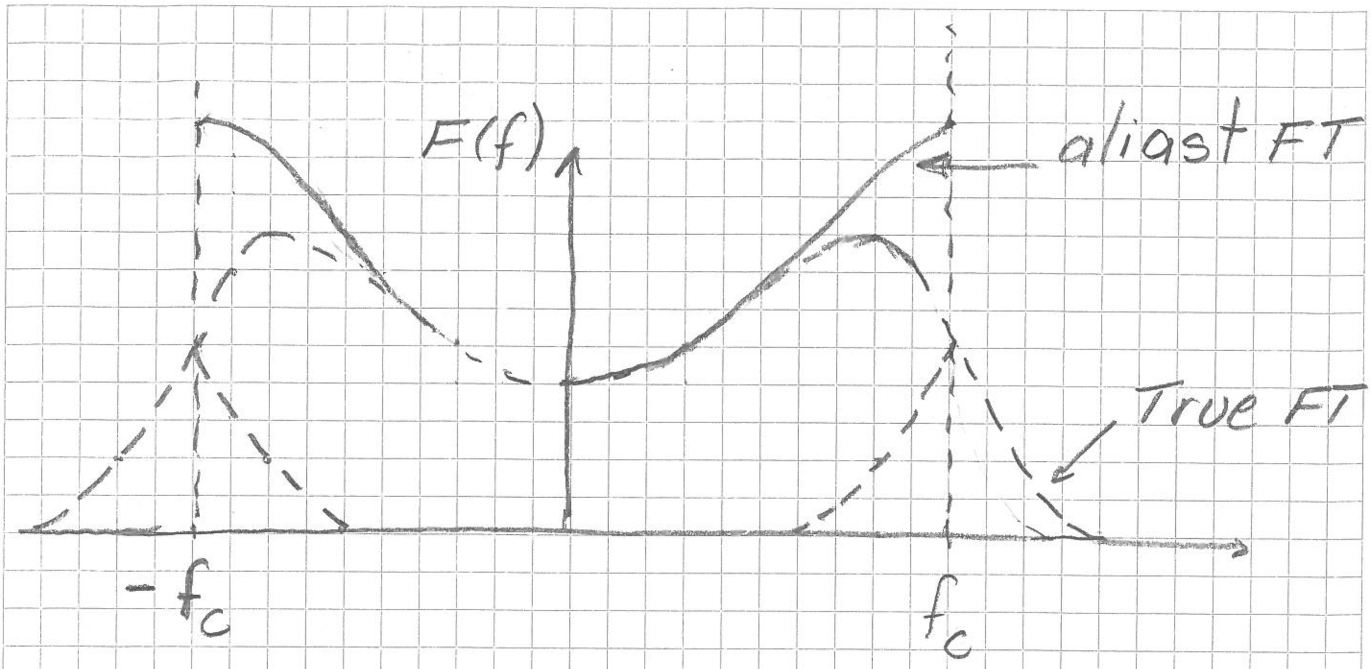
Alising : a phenomenon that happens if the analog signal has freq. components above the Nyquist freq. i.e., it happens if $f_c < f_{max}$.

See Fig 2.3

Alaising causes faulse freq. components to appear that were not in the original (analog) signal

DA convert.

$$\text{analog} \xrightarrow[f_s]{\Delta t} \text{digital} \longrightarrow \text{analog}$$



Aliasing is causing a distortion to occur for which $f_{max} > f_N$ are <u>folded</u> back into the Nyquist interval.

Q : What to do if your signal has $f_{max} > f_c$

A : Apply a low-pass frequency filter before digitizing so that the input signal has $f_{max} < f_c$

2.3 Spectral Analysis (p. 10)

time domain
space — " —

$\xleftrightarrow{FT}$

Freq. domain

real $\qquad$ complex

$$h(t) \xleftrightarrow{FT} H(f)$$

Def : Fourier transform

$$H(f) = \int_{-\infty}^{\infty} dt \; h(t) \, e^{i 2\pi f t}$$

$$h(t) = \int_{-\infty}^{\infty} df \; H(f) \, e^{-i 2\pi f t}$$

— " —

$\omega = 2\pi f \qquad d\omega = 2\pi \, df$

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

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

$$H(\omega) = \left[H(f) \right] \Big|_{f = \omega / 2\pi}$$

Typically $h(t)$ is real, so that it follows

$$H(-f) = [H(f)]^*$$

Show Fig 2.4

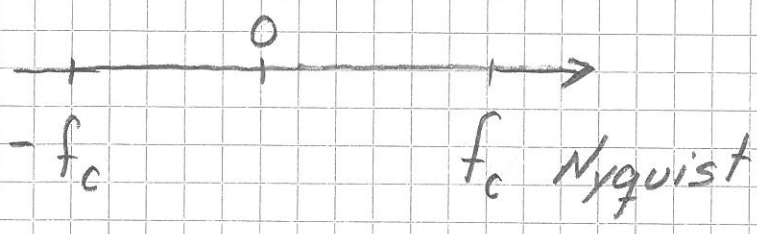
If the function $h(t)$ is periodic of period T

$$h_N(t) = \sum_{n=-N}^{N} C_n \, e^{-i\frac{2\pi n t}{T}}$$

$$C_n = \frac{1}{T} \int_{-T/2}^{T/2} dt \; h(t) \, e^{-i\frac{2\pi n t}{T}}$$

If one dels with a discretized signal of sampling interval Δt

$$h_k = h(t_k) \qquad t_k = k\,\Delta t \qquad k = 1, 2, \ldots, N$$



$$\underbrace{h(t_k)}_{h_k} = \int_{-\infty}^{\infty} df\ H(f)\, e^{-i2\pi f t_k}$$

$$\simeq \Delta f \sum_{n=1}^{N} H(f_n)\, e^{-i2\pi f_n t_k}$$

$$= \frac{1}{N\Delta t} \sum_{n=1}^{N} \Delta t\, H_n\, e^{-i2\pi f_n t_k}$$

$$= \frac{1}{N} \sum_{n=1}^{N} H_n\, e^{-i2\pi f_n t_k}$$

Note:

$$f_n t_k = \frac{n}{N\Delta t}\, k\Delta t = \underline{\frac{nk}{N}}$$

— // —

Since

$$e^{i\theta} = \cos\theta + i\sin\theta$$

$h_k = h(t_k)$ is decomposed into a sum of sine and cosines.

This is at the heart of Fourier anal.

Nyquist

$$\Delta f = \frac{2f_c}{N} = \frac{2}{N 2\Delta t} = \frac{1}{N\Delta t}$$

Therefore, we have discrete freq.

$$f_n = n\Delta f = \frac{n}{N\Delta t}, \quad n = -\frac{N}{2}, \ldots, \frac{N}{2}$$

$$H(f_n) = \int_{-\infty}^{\infty} dt \; h(t) \, e^{i2\pi f_n t}$$

$$\simeq \Delta t \sum_{k=1}^{N} h_k \, e^{i2\pi f_n t_k}$$

$$\equiv \Delta t \; H_n$$

The discrete FT maps N numbers h_k in time space into N numbers H_n in Fourier space.

The inverse transform:

$$h_k = \frac{1}{N} \sum_{n=1}^{N} H_n \, e^{-i2\pi f_n t_k}$$

which follows from $h(t_k)$

see next page

Fourier transforms converts

$$h(t) \underset{FT}{\longleftrightarrow} H(f)$$

Fourier spectrum

$$H(f) = A(f)\, e^{i\phi(f)}$$

$A(f)$: real amplitude

$\phi(f)$: real phase

Fourier pairs

$$h(t) \longleftrightarrow H(f)$$

Ex: Fig 2.8

Geophysical use: emgs

Fast Fourier Transform (FFT)

FFT algorithm : Cooley-Tukey (1974)

If

$$\vec{h} = [h_k] \qquad \text{(time-domain)}$$

$$\vec{H} = [H_n] \qquad \text{(fourier-domain)}$$

then

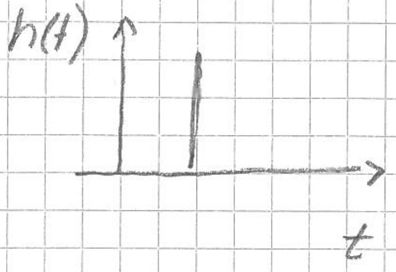
$$FFT(\vec{h}) = \vec{H}$$

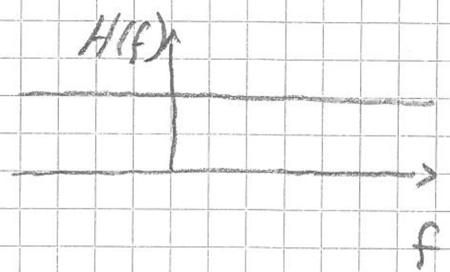
— " —

The FT can be generalized to higher dimensions

Note the following Fourier pairs

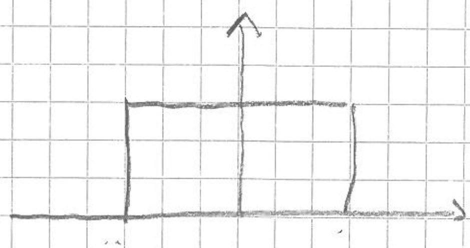
① $\delta(t) \longleftrightarrow 1$





② $1 \longleftrightarrow \delta(f)$

③



$\mathrm{sinc}(\pi t) \longleftrightarrow \mathrm{rect}(f)$

$= \dfrac{\sin(\pi t)}{\pi t}$

(rectangular func)
(puls-function)

④ $\mathrm{rect}(t) \longleftrightarrow \mathrm{sinc}(\pi f)$

$$\mathrm{rect}(t) = \begin{cases} 1 & |t| < \frac{1}{2} \\ \frac{1}{2} & |t| = \frac{1}{2} \\ 0 & |t| > \frac{1}{2} \end{cases}$$

2.4 Convolution (p. 13)

Def: Convolution of two (continuous) functions $g(t)$ and $h(t)$ is

$$g(t) * h(t) = \int_{-\infty}^{\infty} d\tau \, g(\tau) \, h(t-\tau)$$

— // —

You will also see the notation $(g*h)(t)$.

It can be shown (by a change of var.) that

$$g(t) * h(t) = h(t) * g(t).$$

— // —

Lets calculate the conv at $t_k = k\Delta t$, $k = 1,2,3,\ldots$

$$y(t_k) = g(t_k) * h(t_k)$$

$$\equiv \int_{-\infty}^{\infty} d\tau \, g(\tau) \, h(t_k - \tau)$$

$$\simeq \Delta t \sum_{i=1} g(\tau_i) \, h(t_k - \tau_i), \quad \tau_i = i\Delta t$$

$$= \Delta t \sum_{i=1} g_i \, h_{k-i}$$

In a discrete convolution one usually neglects the Δt-term and define for

$$\{g_i\}_{i=1}^{m} \quad \text{and} \quad \{h_j\}_{j=1}^{n}$$

$$\boxed{y_k = \sum_{i=1}^{m} g_i \, h_{k-i} \qquad k = 1, 2, \ldots, m+n-1}$$

Note: there are $m+n-1$ y_k values

Since $g * h = h * g$ it follows that

$$y_k = \sum_{i=1}^{n} h_i \, g_{k-i}$$

can alternatively used to calculate the convolution.

Ex : Calculate the convolution of the two vectors

$$\vec{g} = [2,0,1] \qquad m = 3$$

$$\vec{h} = [4,3,2,1] \qquad n = 4$$

See Fig. 2.12

$$\vec{y} = \text{conv}(\vec{g}, \vec{h}) \qquad [\text{matlab}]$$

$$= [8,6,8,5,2,1] \qquad 6\text{-elements}$$

— // —

The Convolution theorem

$$\boxed{\begin{aligned} \mathcal{F}(g(t) * h(t)) &= \mathcal{F}(g(t)) \cdot \mathcal{F}(h(t)) \\ &= G(f) \cdot H(f) \end{aligned}}$$

$\mathcal{F}$: The Fourier Transform

Octave / Matlab

conv$(\vec{g}, \vec{h})$

fftconv$(\vec{g}, \vec{h})$ - uses FFT

Q : What is the meaning of the convolution operator

A : It describes the effect on an input signal after passing throug a filter

Input signal $\longrightarrow$ Filter $\longrightarrow$ Output signal

Ex : What do you hear when the guy next-door is playing loud music?

You hear too much bass since the wall acts as a low-pass frequency filter.

Also see the animation on the wikipedia page dedicated to conv.

2.4.1 Impulse response / Transfer funct.

Input signal (time-domain)

Output signal (time-domain)

$\boxed{I(t): \text{ImpulseResponse}}$

$\Downarrow$ Fourier Trans.

$\boxed{T(f): \text{Transfer Function}}$

Both the Impulseresponce, $I(t)$, and the Transfer function, $T(f)$, characterize the filter and they contain the same amount of information.

Ex : Fig. 2.9

Note :

$I(t)$: time - domain
$T(f)$: frequency domain

$$\boxed{T(f) = \mathcal{F}(I(t))} \qquad I(t) = \mathcal{F}^{-1}(T(f))$$

Ex : We measure an input signal, $d(t)$, using a sensor of transfer function, $T(f)$, and obtain the output (measured) signal, $m(t)$.

$$d(t) \longrightarrow \boxed{\begin{array}{l} \text{Sensor} \\ T(f) \\ I(t) \end{array}} \longrightarrow m(t)$$

In the Fourier domain we have :

$$M(f) = T(f) \cdot D(f)$$

which in the time-domain becomes after using the convolution theorem :

$$m(t) = I(t) * d(t).$$

We observe that in general $m(t) \neq d(t)$.

2.4.2 Deconvolution (p. 16)

Q: From the measured signal $m(t)$, how can one obtain the input sig. $d(t)$?

A: Deconvolution (also called inv. filtering) solves this issue and works like this

$$m(t) = I(t) * d(t)$$

$$\downarrow FT$$

$$M(f) = T(f)\, D(f)$$

$$\Rightarrow D(f) = \frac{1}{T(f)} M(f) = T^{-1}(f)\, M(f)$$

$$d(t) = \mathcal{F}^{-1}\left(T^{-1}(f)\, M(f)\right)$$

Alternatively, one may ask for an inpulserespons $J(t)$ for deconvolution process that take $m(t)$ into $d(t)$, i.e

$$d(t) = J(t) * m(t)$$

or after using $m(t) = I(t) * d(t)$, it follows

$$d(t) = \underbrace{J(t) * I(t)}_{\delta(t)} * d(t)$$

Note: Applying two filters

$$d(t) \longrightarrow \underset{T_1(f)}{\boxed{\text{Filter 1}}} \longrightarrow \underset{T_2(f)}{\boxed{\text{Filter 2}}} \longrightarrow m_2(t)$$

$$M_2(f) = T_2(f) \cdot T_1(f) \cdot D(f)$$

Still $m_2(t) = \mathcal{F}^{-1}(M_2(f))$, but the inv. Fourier transform is more complicated to obtain.

The total transfer function of <u>combination</u> of filter (or sensor) 1 and 2 is

$$T(f) = T_2(f)\,T_1(f)$$

so that the corresponding total impuls response is

$$\underline{I}(t) = \mathcal{F}^{-1}(\,T_2(f) \cdot T_1(f))$$
$$= \underline{I}_2(t) * \underline{I}_1(t)$$

2.4.3 Application of convolutions in GP

- We have alredy seen the use conv. due to a non-trivial impulse response of a sensor. (instrument response)

- However, for instanse, the change in a seismic wave as it passes through the ground can be viewed as a filtering process.
  - with the ground we associate an impulse response
  - deconvolution is an essential part of seismic data processing (improve data quality)

Q : Why is deconvolution of GP data challenging

A : The impuls responce (or transfer function) of the Earth is typically not known.

2.5 Correlation (p. 16)

Definition:

Given two functions $g(t)$ and $h(t)$, then their correlation function is defined as

$$\phi_{gh}(t) = \int_{-\infty}^{\infty} d\tau \, g(\tau+t)\, h(\tau)$$

where t is known as the lag.

— // —

If $g=h$ we talk about the <u>auto-correlation function</u>.

In the discrete case it follows

$$\phi_{gh}(t_k) = \int_{-\infty}^{\infty} d\tau \, g(\tau+t_k)\, h(\tau)$$

$$\simeq \Delta t \sum_i g(\tau_i+t_k)\, h(\tau_i)$$

$$= \Delta t \sum_i g_{i+k}\, h_i$$

As for the convolution, one normally suppress Δt and under the assumption

$$\{g_i\}_{i=1}^{n} \qquad \{h_i\}_{i=1}^{n}$$

write

$$\phi_{gh}(\tau_k) = \sum_{i=1}^{n-k} g_{i+k} h_i , \quad -m \leq k \leq m$$

— // —

The correlation function $\phi_{gh}(\tau)$ quantify how <u>similar</u> the functions $g(t)$ and $h(t)$ are.

See Fig. 2.13

Where $\phi_{gh}(\tau) = 0$ (for given τ) we say that g and h are uncorrelated.

The auto-correlation function has the property

$$\phi_{xx}(\tau) = \phi_{xx}(-\tau)$$

The auto-correlation function ϕ_{xx} contains all the amplitude information of $x(t)$ but not any phase information.

Mathematically this is expressed by the so-called Wiener-Kchinchin theorem stating that

$$\phi_{xx}(t) \xleftrightarrow{FT} |\hat{\phi}_{xx}(f)|^2$$

where $\hat{\phi}(f) = \mathcal{F}(\phi_{xx}(t))$.

The quantity

$$P(f) = |G(f)|^2$$

is known as the power spectrum of the function $g(t)$.

The correlation theorem

$$\mathrm{corr}(g,h)_k \longleftrightarrow G_i(f)\, H_i^*(f)$$

For a real signal $x(t)$ it follows from the properties of the Fourier transform that

$$\hat{x}(f) = \mathcal{F}[x(t)]$$

$$\hat{x}^*(f) = \hat{x}(-f)$$

— " —

The power spectrum in more detail:

$$\hat{x}(f) = \mathcal{F}[x(t)]$$

$$P(f) \overset{\text{Def.}}{=\!=} |\hat{x}(f)|^2$$

$$= \hat{x}^*(f)\,\hat{x}(f)$$

$$= \mathcal{F}\left[\mathcal{F}^{-1}[\hat{x}^*(f)\,\hat{x}(f)]\right]$$

$$= \mathcal{F}[\phi_{xx}(t)]$$

2.5.1 Applications of correlations in GP

- Detection of weak signals embedded in noise
- Detecting periodicity (often via the power spectrum)

  see Fig 2.15
- Detection of multiple reflections in seismic reflection data.

2.6 Digital Filters (p. 17)

$$d(t) = s(t) + n(t)$$

data = signal + noise

signal : the part of the waveform data related to the geological structure under investigation

noise : all other components of the waveform

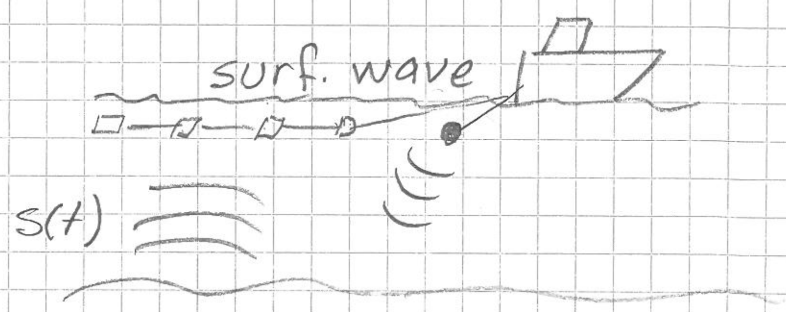
Two types of noise :

- random noise :
  statistically random signal typically unrelated to the GP survey

  Ex : colored instrument noise
  background vibration
  (due to wind, waves etc)

- coherent noise:
  components of the waveform generated by the GP experim. but not of particular interest for the geological interpretation

Ex.



Signal-to-noise ratio (SNR)

$$\frac{s(t)}{n(t)}$$

One want this signal to have a large amplitude.

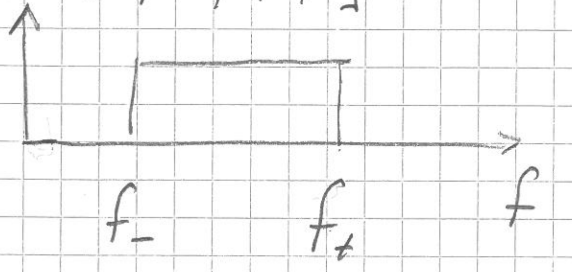
For low SNR one tries to increase the signal by processing to filter out eg coherent noise.

Two types of filters:

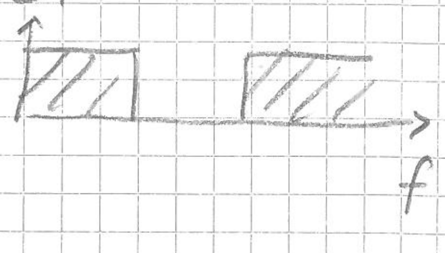
- Freq. Filters
- Inverse (deconvolution) filters

Frequency Filters

Lets through freq. components of the input signal only in a finite interval $[f_-, f_+]$



- low-pass (LP) filter : $f_- = 0$
- high-pass (HF) –"– : $f_+ = \infty$
- band-pass (BP) –"–
- band-reject



Inverse Filters:

Remove the effect of filters applied previously

(more in Ch 4)

2.7 Imaging and Modelling (Inversion) p 19.

Geophysical Waveform

↓

Processing

↓

Data with optimal SNR

↙ ↘

Imaging (data presented) | Modelling / Inversion

↓ ↓

Geological Interpretation