Fabio D'Andrea
LMD – 4e étage “dans les serres”
01 44 32 22 31

dandrea@lmd.ens.fr

http://www.lmd.ens.fr/dandrea/TEACH

Program

13/1	Elementary statistics – 1
20/1	Elementary statistics - 2
27/1	Exercises – Computer room
10/2	Fourier Analysis -1
17/2	Fourier Analysis -2, stochastic processes
24/2	Exercises – Computer room
2/3	Exercises – Computer room
9/3	Principal component analysis -1
16/3	Principal component analysis -2
23/3	Exercises – Computer room
13/4	Exercises – Computer room
27/4	Cluster analysis
4/5	Exercises – Computer room
25/5	Principal component analysis: Complements
6/6	Exam.

. Lesson 2.

Frequency domain methods.

Fourier transform. A reminder.

For any "well behaved" function x(t), the Fourier transform will be written:

$$\mathcal{F}(x(t)) = \hat{x}(s) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi st} dt$$

Joseph Fourier
1768 - 1830

Important theorems:

1) if $f * g = \int_{-\infty}^{\infty} f(t) g(\tau - t) dt$ then $\mathcal{F}(f * g) = \hat{f}(s) \cdot \hat{g}(s)$ $(Convolution)$

2) if $f \bullet g = \int_{-\infty}^{\infty} f(t) g(\tau + t) dt$ then $\mathcal{F}(f \bullet g) = \hat{f}(s)^* \cdot \hat{g}(s)$ $(Cross - correlation)$

3) $\int_{-\infty}^{\infty} x(t)^2 dt = \int_{-\infty}^{\infty} |\hat{x}(s)|^2 ds$ $(Parseval)$

In most practical applications, the function $x(t)$ is only known at discrete intervals of time Δt, and for a non-infinite time, $0 \le t \le T$.

Hence $x(t) = x(n\Delta t), \;\; n = 0,1,..N$ and $T = N\Delta t$

Δt is also called "sampling interval", and Δt^{-1} is called the "sampling rate".

There is also a special frequency that can be defined, the Nyquist frequency:

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

Fourier Series. A reminder.

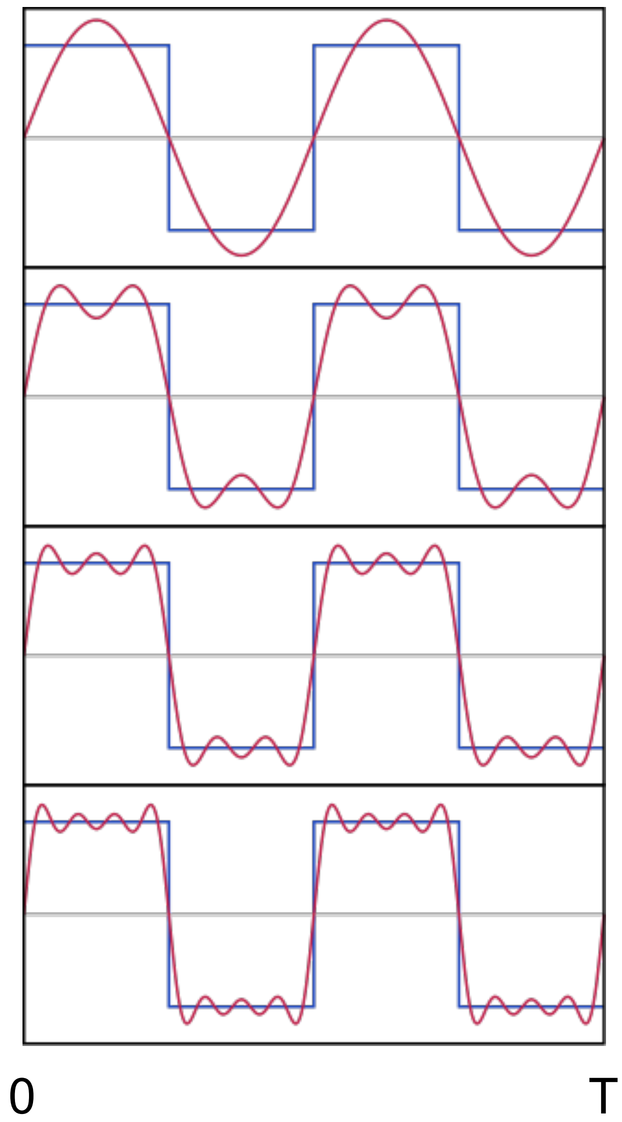
Any periodic function $x(t) = x(t+T)$ can be written as a Fourier Series

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{2\pi n t}{T}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{2\pi n t}{T}\right)$$

Where the Fourier coefficients are defined as:

$$a_n = \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \cos\left(\frac{2\pi n t}{T}\right) dt, \qquad for \quad n \geq 0$$

$$b_n = \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \sin\left(\frac{2\pi n t}{T}\right) dt, \qquad for \quad n \geq 1$$



Fourier Series. A reminder.

Any periodic function $x(t) = x(t+T)$ can be written as a Fourier Series

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{i\frac{2\pi nt}{T}}$$

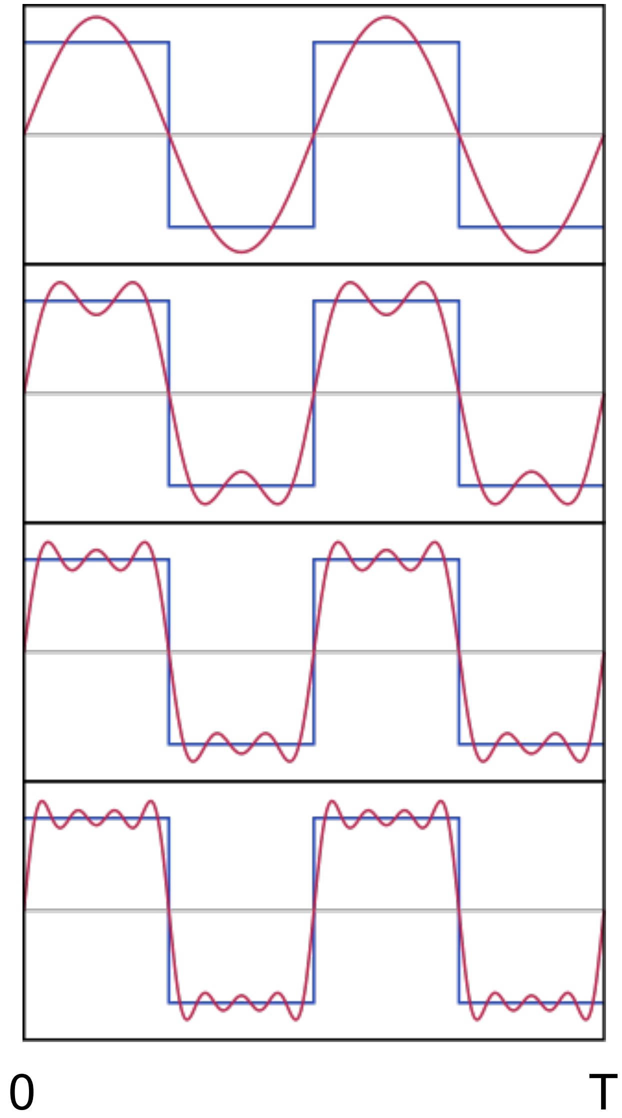
Where the Fourier coefficients are defined as:

$$c_n = \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-i\frac{2\pi nt}{T}} dt$$

This is a compact way of writing the original Fourier formula that made use of sines and cosines, Where the coefficients are related this way:

$$a_n = c_n + c_{-n} \qquad n = 0,1,2,3...$$

$$b_n = i(c_n - c_{-n}) \qquad n = 1,2,3.....$$



Interpretation as basis of a Hilbert space

Functions $e_n = \frac{1}{T} e^{-i\frac{2\pi nt}{T}}$ are an orthonormal basis of the space $\mathscr{L}^2([-T/2,\ T/2])$ of the square-Integrable fiunctions defined over $[-T/2,\ T/2]$. This space is called a Hilbert space with an inner product defined as:

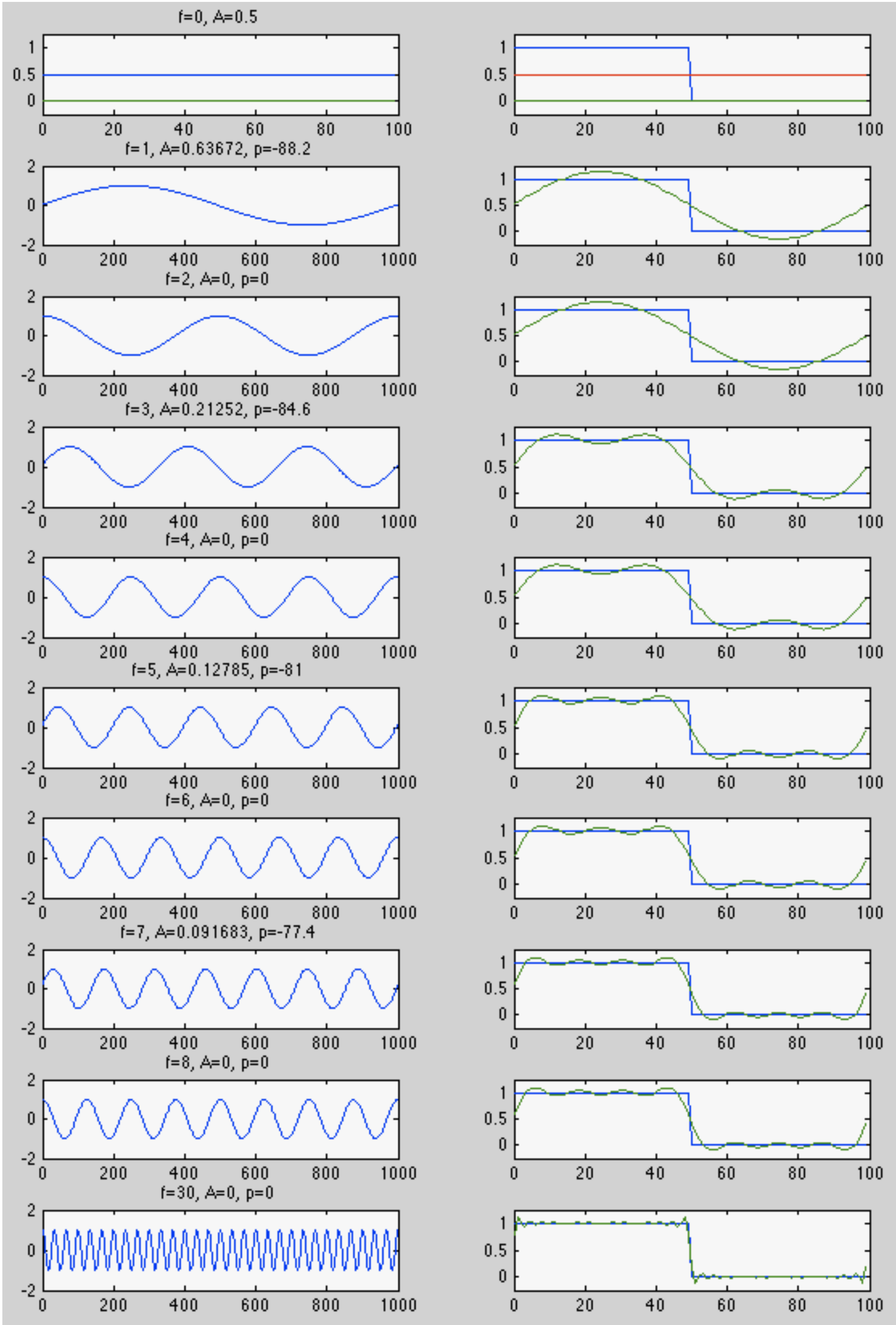
$$< f,g >= \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)g(t)^* dt$$

So that any function of the space can be expanded as:

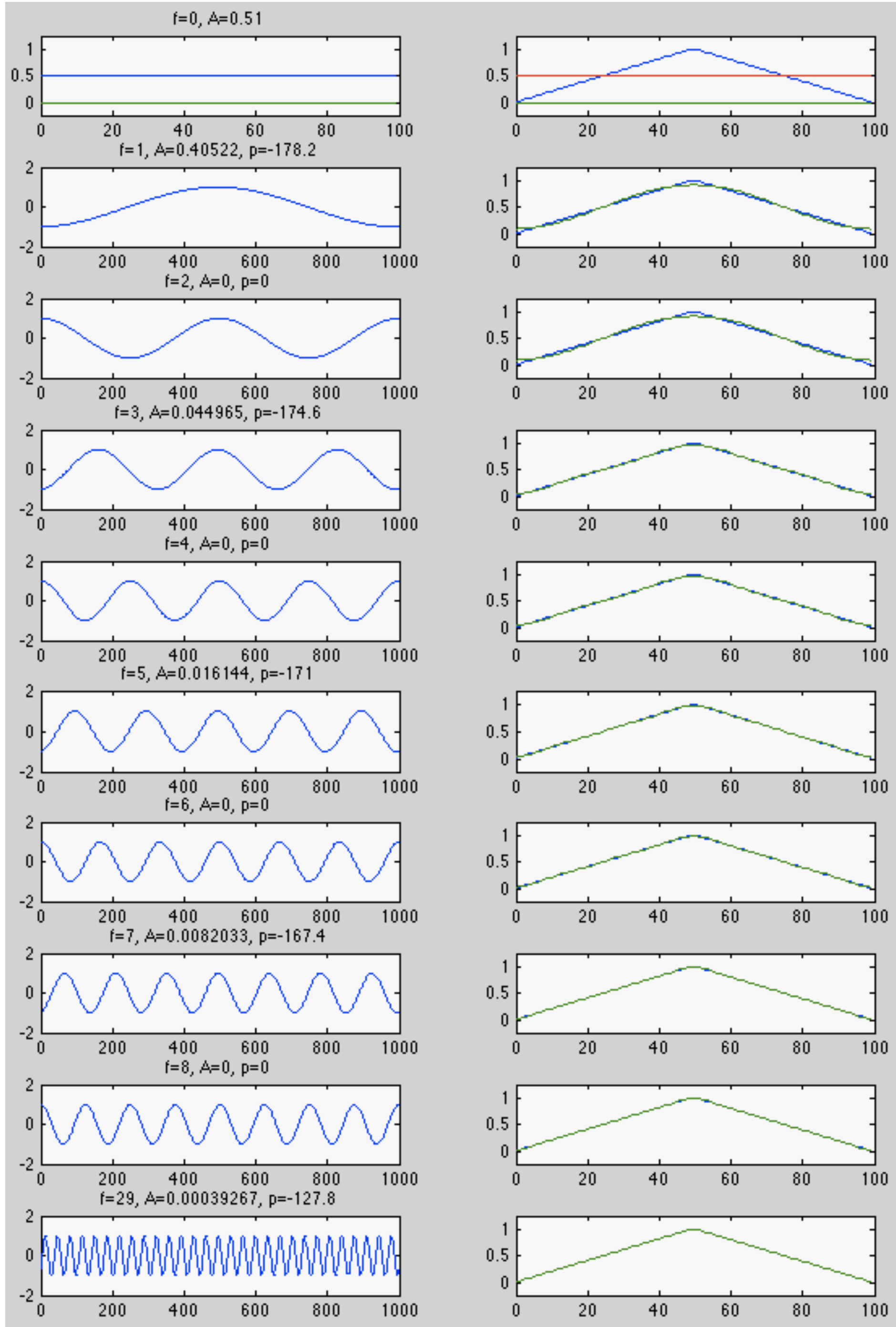
$$f = \sum_{n=-\infty}^{\infty} < f,e_n > e_n$$

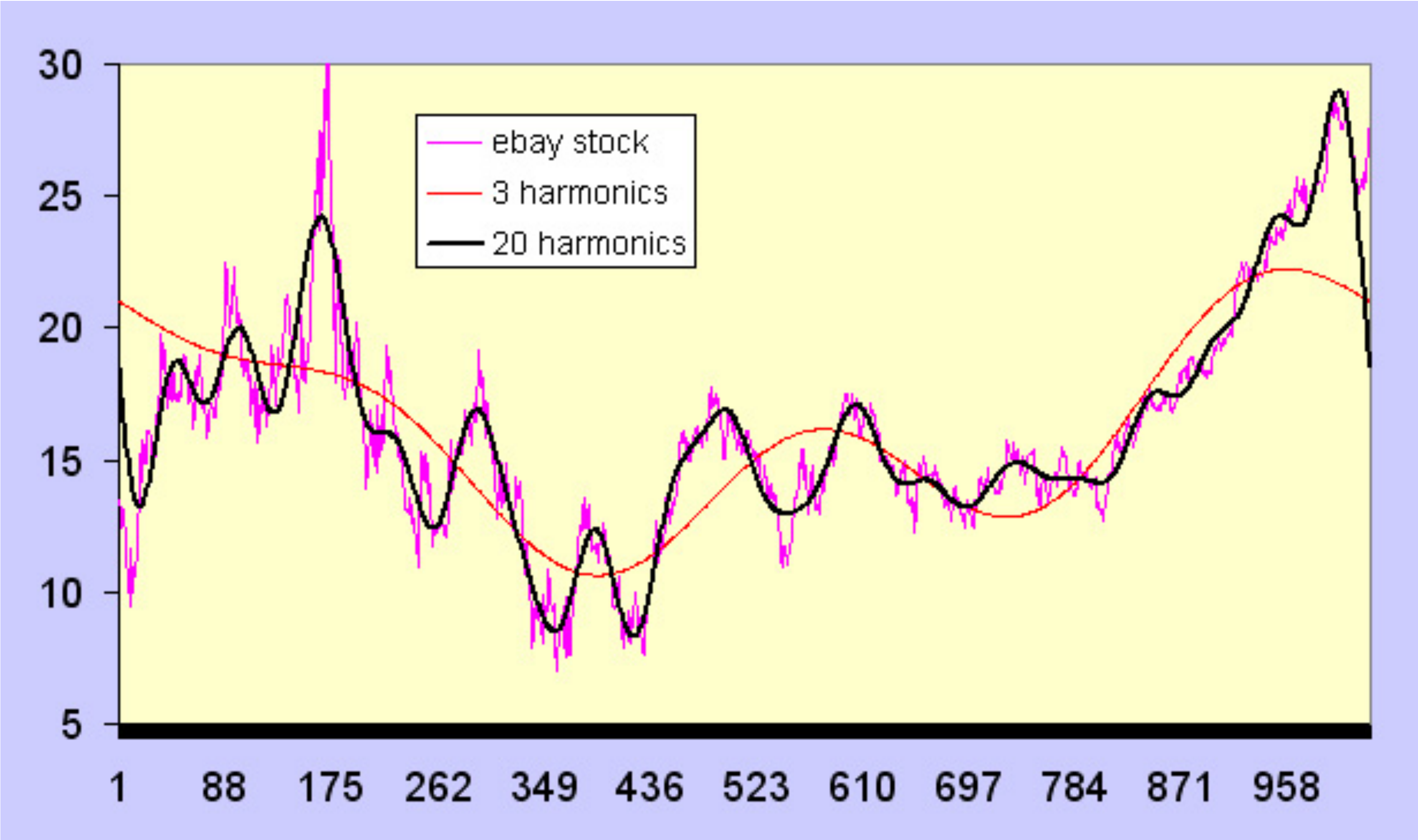
This also holds for the sines and cosines representation of the Fourier basis. In this case the basis would be formed by the functions 1, $\frac{2}{T}\cos\frac{2\pi nt}{T}$ and $\frac{2}{T}\sin\frac{2\pi nt}{T}$.

Step wave



Triangular wave





30
25
20
15
10
5
ebay stock
3 harmonics
20 harmonics
1
88
175
262
349
436
523
610
697
784
871
958

In most practical applications, the function $x(t)$ is only known at discrete intervals of time Δt, and for a non-infinite time, $0 \le t \le T$.

Hence $x(t) = x(n\Delta t),\ n = 0,1,..N$ and $T = N\Delta t$

Δt is also called "sampling interval", and Δt^{-1} is called the "sampling rate".

There is also a special frequency that can be defined, the Nyquist frequency:

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

Harry Nyquist
1889-1976

Claude Shannon
1916 – 2001

It can be shown that if a function is sampled with intervals Δt, and it is band-limited, $-1/2\Delta t \leq s \leq 1/2\Delta t$, then it can be reconstructed exactly. We can write:

$$x(t) = \Delta t \sum_{n=-\infty}^{\infty} x(n\Delta t) \frac{\sin(2\pi f\,(t - n\Delta t))}{\pi(t - n\Delta t))}$$

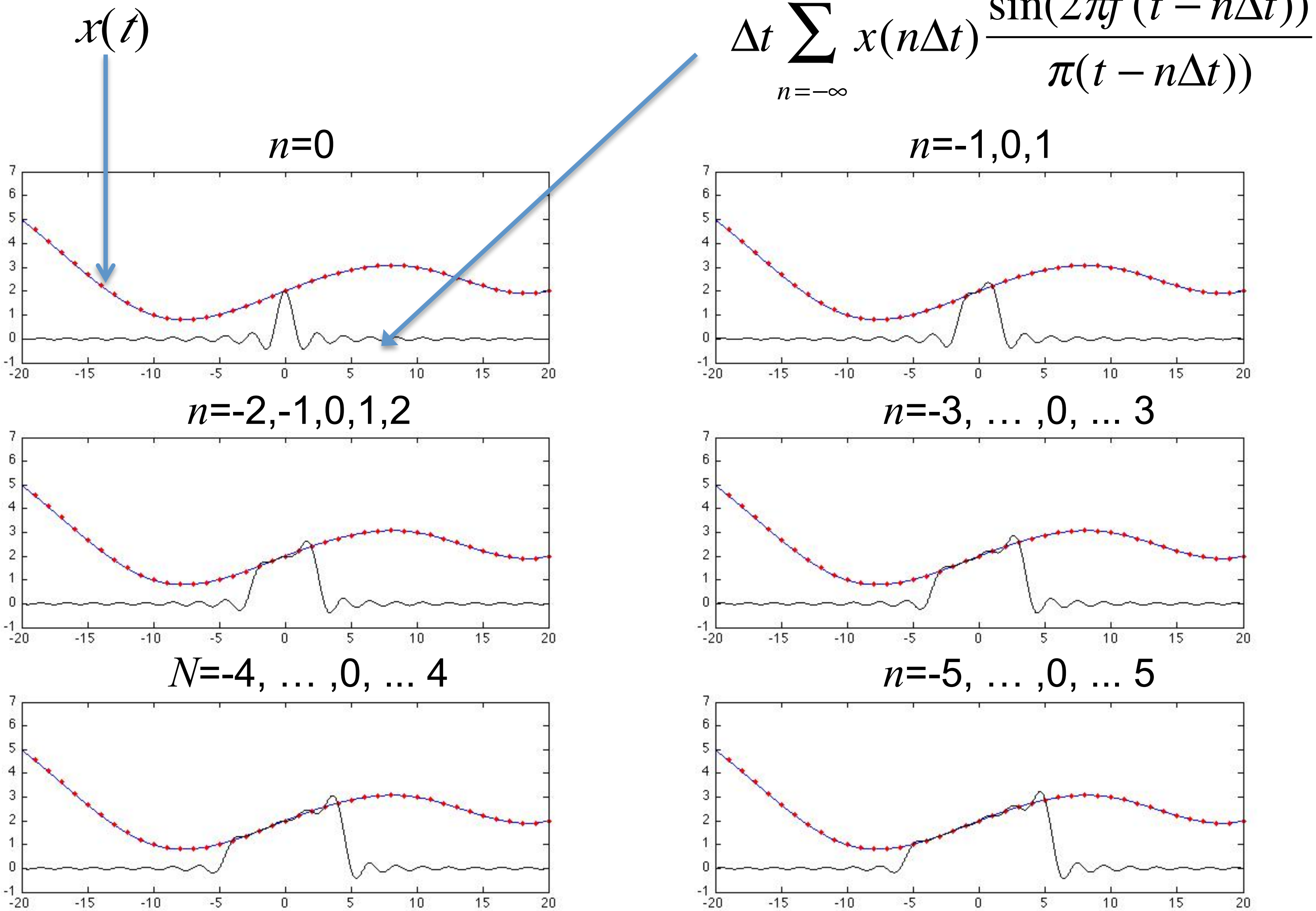
(Nyquist-Shannon sampling theorem)

Which in turn means that a band-limited function can be perfectly reconstructed by samples of the function at discrete times $n\Delta t$.

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

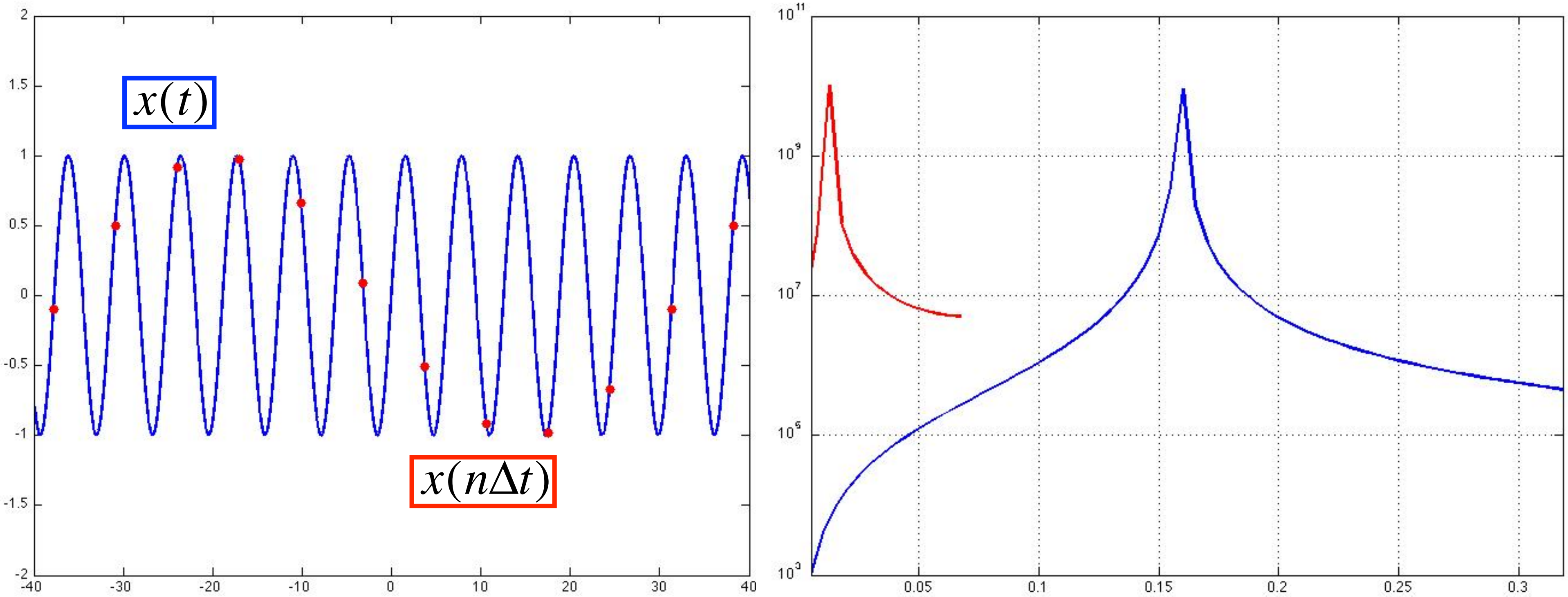
is the *Nyquist frequency.*

$$\Delta t \sum_{n=-\infty}^{\infty} x(n\Delta t) \frac{\sin(2\pi f(t-n\Delta t))}{\pi(t-n\Delta t)}$$



Aliasing. If *x(t)* is not band-limited, then the sampling theorem is not satisfied, and there might be an aliasing effect.

The frequencies higher than the Nyquist frequency will be squashed on to the frequencies resolved.



Discrete Fourier analysis: sampling in the frequency domain.

If we have a discrete and limited timeseries

$$x(t) = x(n\Delta t) = x_n, \quad n = 0,1,2,3......,N-1$$

We seek its discrete Fourier transform in the frequency range $-1/2\Delta t \leq s \leq 1/2\Delta t$, that is $-f \leq s \leq f$. Sampling in regular intervals we get the frequencies

$$s_k = \frac{k}{N\Delta t} \qquad k = -\frac{N}{2},......,\frac{N}{2}$$

In this cases it is:

$$\hat{x}(s_k) = \int_{-\infty}^{\infty} x(t)e^{-i2\pi st}dt = \sum_{n=0}^{N-1} x_n e^{-i2\pi s_k t_n}\Delta t = \Delta t \sum_{n=0}^{N-1} x_n e^{-i2\pi(n\Delta t)(k/N\Delta t)}$$

$$\hat{x}(s_k) = \Delta t \sum_{n=0}^{N-1} x_n e^{-i2\pi\frac{nk}{N}}$$

$$\hat{x}_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi\frac{nk}{N}}$$ is the discrete Fourier transform

The inverse transform is

$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} \hat{x}_k e^{i2\pi\frac{nk}{N}}$$

It is the exact reconstruction of the time domain discrete function via the Fourier coefficients and the Fourier basis.

Exercise: prove the above definition of the inverse transform.

The Parseval theorem takes the form:

$$\sum_{n=0}^{N-1} x_n^{\,2} = \frac{1}{N}\sum_{k=0}^{N-1} \left|\hat{x}_k\right|^2$$

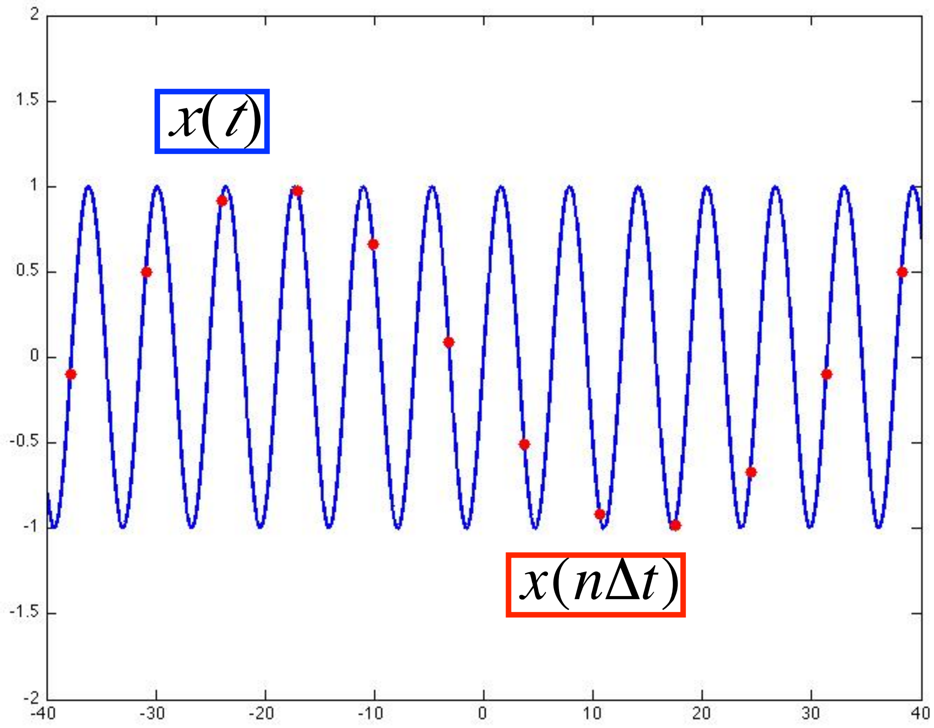
Exercise: prove it!

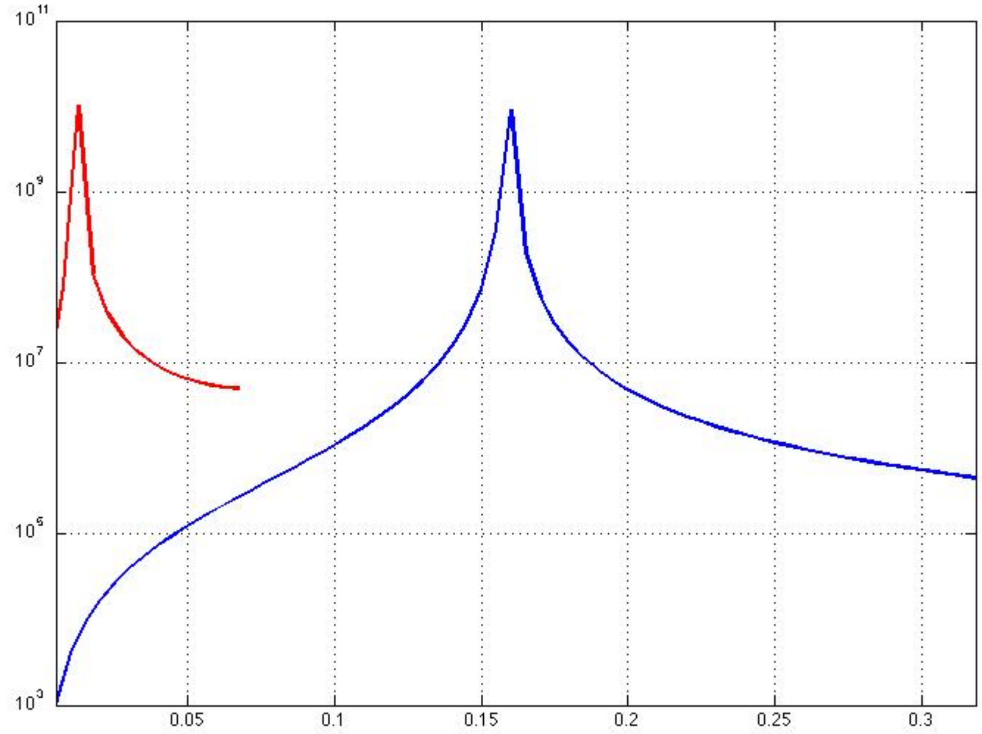
Some symmetry properties, they hold for continuous and discrete FT

If ...	then ...
$h(t)$ is real	$H(-f) = [H(f)]^*$
$h(t)$ is imaginary	$H(-f) = -[H(f)]^*$
$h(t)$ is even	$H(-f) = H(f)$ [i.e., $H(f)$ is even]
$h(t)$ is odd	$H(-f) = -H(f)$ [i.e., $H(f)$ is odd]
$h(t)$ is real and even	$H(f)$ is real and even
$h(t)$ is real and odd	$H(f)$ is imaginary and odd
$h(t)$ is imaginary and even	$H(f)$ is imaginary and even
$h(t)$ is imaginary and odd	$H(f)$ is real and odd

Aliasing. If *x(t)* is not band-limited, then the sampling theorem is not satisfied, and there might be an aliasing effect.

The frequencies higher than the Nyquist frequency will be squashed on to the frequencies resolved.





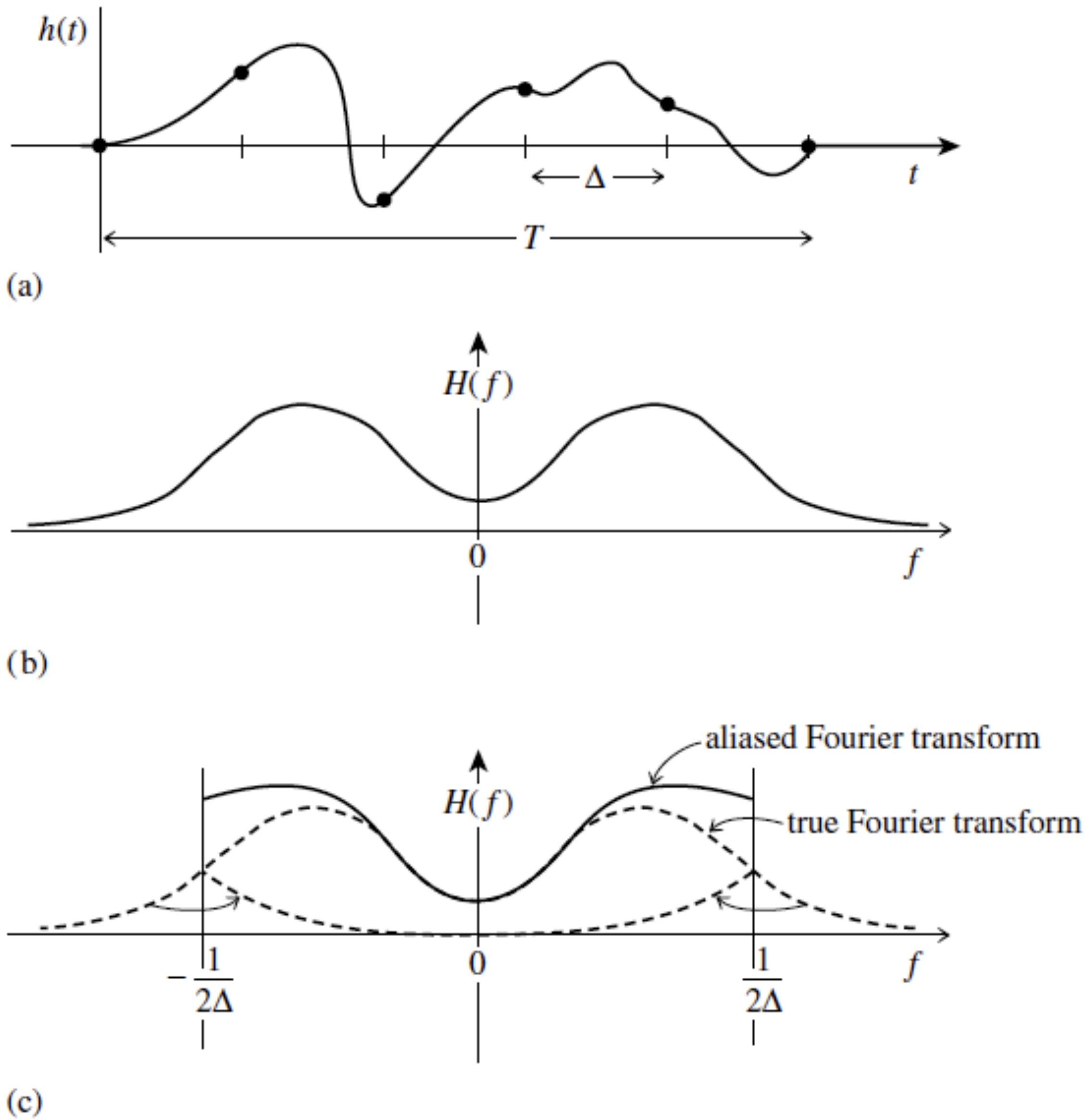


Figure 12.1.1. The continuous function shown in (a) is nonzero only for a finite interval of time T. It follows that its Fourier transform, whose modulus is shown schematically in (b), is not bandwidth limited but has finite amplitude for all frequencies. If the original function is sampled with a sampling interval Δ, as in (a), then the Fourier transform (c) is defined only between plus and minus the Nyquist critical frequency. Power outside that range is folded over or "aliased" into the range. The effect can be eliminated only by low-pass filtering the original function *before sampling*.

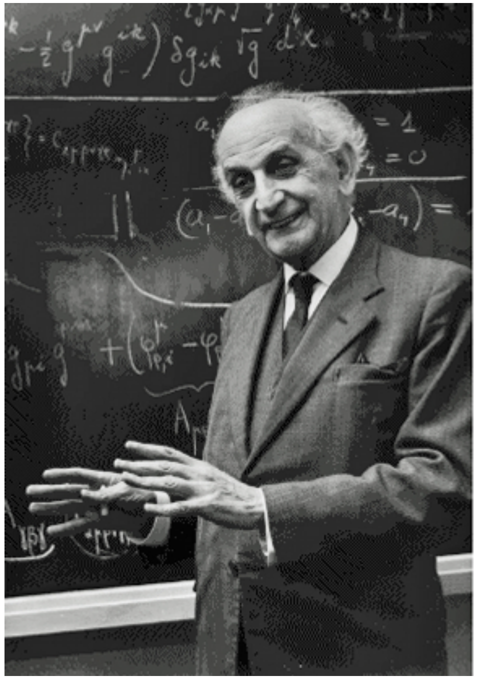
Time aliasing
“the cartwheel effect”

Wagon Wheel (stroboscopic) Effect

Fast Fourier Transform (FFT)

One of the most common methods of spectral analysis used is the Fast Fourier Transform.

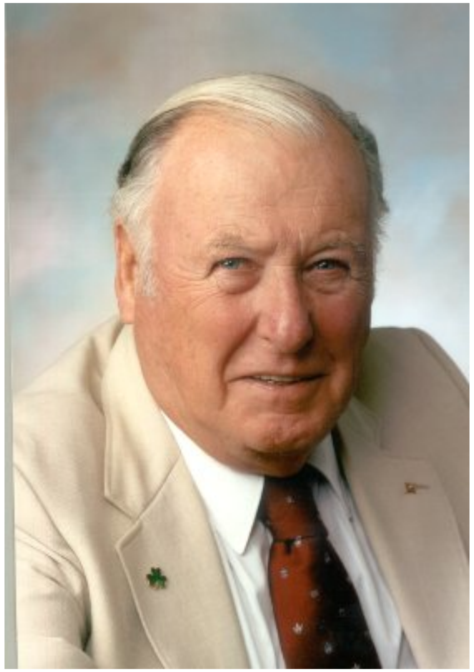
The FFT reduces the number of operation to N log N.



Cornelius Lanczos
(1893-1974)

Gordon Danielson
(1912-1983)

John Tukey
(1915-2000)



James Cooley
(1926-)

The FFT is an efficient algorithm that works best when the length of the time series has been chosen to be an integer power of two $N=2^n$.

Defining $W = e^{-i\frac{2\pi}{N}}$, the Fourier transform becomes: $\hat{x}_k = \sum_{n=0}^{N} W^{nk} x_n$.

Hence the transform requires N^2 operations.
The FFT reduces the number of operation to $N \log_2 N$.

An example with N=4 just to understand the idea
Note that $W^2 = -1$ and $W^4 = 1$

$$\hat{x}_0 = x_0 + x_1 + x_2 + x_3 = \left(x_0 + x_2\right) + \left(x_1 + x_3\right)$$

$$\hat{x}_1 = x_0 + W^1 x_1 + W^2 x_2 + W^3 x_3 = \left(x_0 - x_2\right) + W^1\left(x_1 - x_3\right)$$

$$\hat{x}_0 = x_0 + W^2 x_1 + W^4 x_2 + W^6 x_3 = \left(x_0 + x_2\right) - \left(x_1 + x_3\right)$$

$$\hat{x}_0 = x_0 + W^3 x_1 + W^6 x_2 + W^9 x_3 = \left(x_0 - x_2\right) - W^1\left(x_1 - x_3\right)$$

It is based on the idea of cutting the transform in two, then again in two, etc etc.

At the end the transform is cut into pieces of length 1 tranforms. A length1 transform is just the Identity. In general:

$$
\begin{aligned}
F_k &= \sum_{j=0}^{N-1} e^{2\pi i jk/N} f_j \\
&= \sum_{j=0}^{N/2-1} e^{2\pi ik(2j)/N} f_{2j} + \sum_{j=0}^{N/2-1} e^{2\pi ik(2j+1)/N} f_{2j+1} \\
&= \sum_{j=0}^{N/2-1} e^{2\pi ikj/(N/2)} f_{2j} + W^k \sum_{j=0}^{N/2-1} e^{2\pi ikj/(N/2)} f_{2j+1} \\
&= F_k^e + W^k \, F_k^o
\end{aligned}
$$

$$= F_k^{ee} + W^k F_k^{eo} + W^k F_k^{oe} + W^{2k} F_k^{oo}$$

So we are reduced to N identities, the trick consist in finding which n corresponds to a given pattern of *o's* and *e's.*

$$F_k^{eoeeoeo\cdots oee} = f_n \qquad \text{for some } n$$

As it happens, if you put *e*=0 and *o*=1, the patterns *« …eooeoeo... »*, reversed, give the value of n in binary!!

The existences of the FFT Algorithm is one of the reasons Why Fourier analysis is so popular.

Spectra.

The simplest estimate of spectral power is the periodogram.
One does the Fourier transform of the data series:

$$\hat{x}_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi \frac{nk}{N}}$$

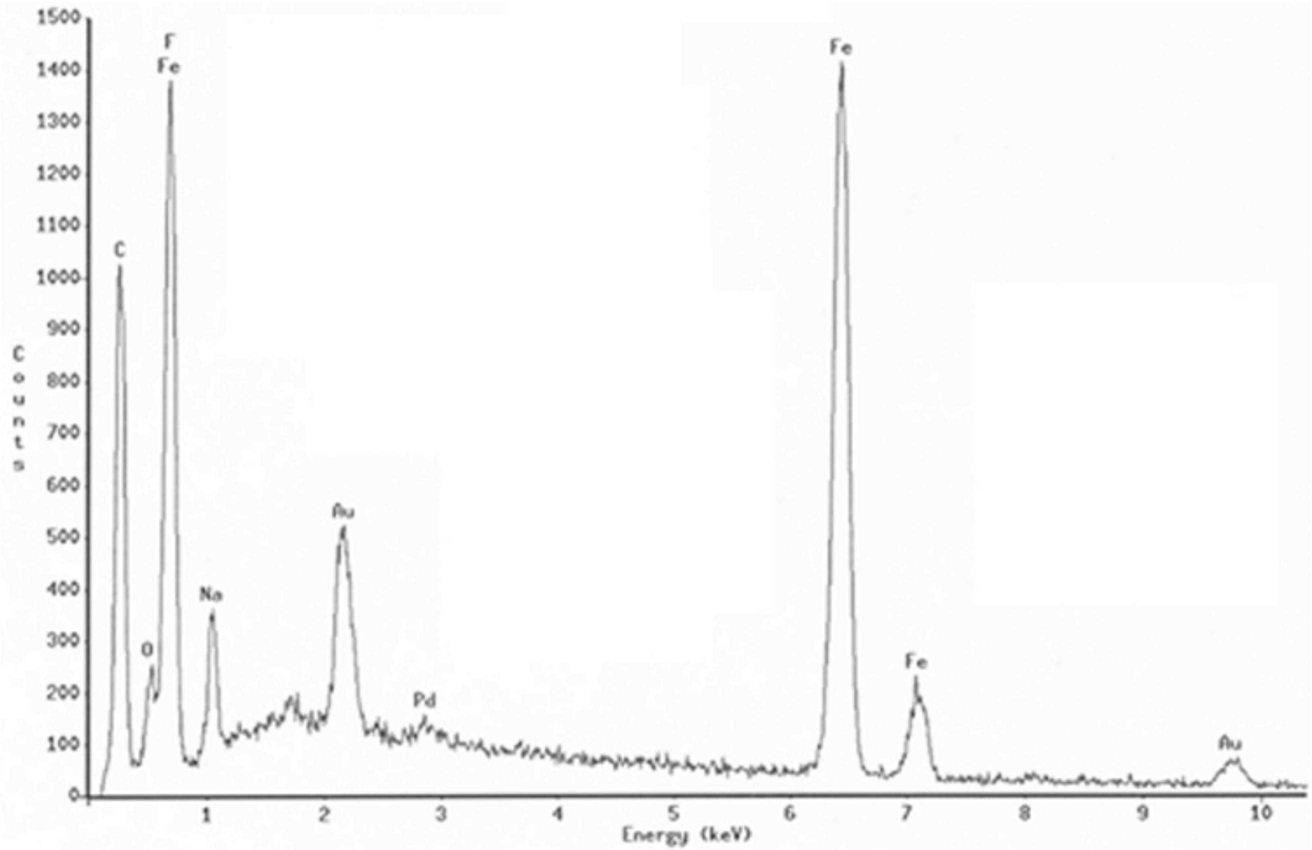
And then plots:

$$P(0) = \frac{1}{N^2} \left| \hat{x}_0 \right|^2$$

$$P(f_k) = \frac{1}{N^2} \left(\left| \hat{x}_k \right|^2 + \left| \hat{x}_{N-k} \right|^2 \right)$$

Remembering the Parseval theorem, one can see the periodogram is the variance of the signal per frequency. The sum of all $P(f_k)$ is the total variance.

$$\sum_{n=0}^{N-1} x_n{}^2 = \frac{1}{N} \sum_{k=0}^{N-1} \left| \hat{x}_k \right|^2 \quad \text{(Parseval)}$$



1500
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
Counts
C
F
Fe
O
Na
Au
Pd
Fe
Fe
Au
1
2
3
4
5
6
7
8
9
10
Energy (keV)

Leakage, tapering etc…

Some further considerations on computing spectra. An example to understand. Let us compute the Fourier transform of a cosine:

$$x(t) = \cos\left(2\pi p_1 t\right)$$

$$\hat{x}(s) = \int_{-\infty}^{\infty} \cos(2\pi p_1 t) e^{i2\pi st}\, dt =$$

$$= \frac{1}{2}\left(\delta(s - p_1) + \delta(s + p_1)\right)$$

Exercise: do it at the blackboard

This is the “real” continuous Fourier transform.
Now suppose that you only observe the original cosine signal for a limited amount of time [0,T]. How does the Fourier transform change?

This problem becomes the problem of computing the Fourier transform of

$$x_\Pi(t) = \Pi\left(\frac{t}{T}\right)\cos(2\pi p_1 t)$$, where $$\Pi(x) = \begin{cases} 0 & if \quad |x| > 1 \\ 1 & if \quad |x| \leq 1 \end{cases}$$

With some algebra one finds:

$$\hat{x}_\Pi(s) = ???$$

Exercise: Who does it at
the blackboard?

This problem becomes the problem of computing the Fourier transform of

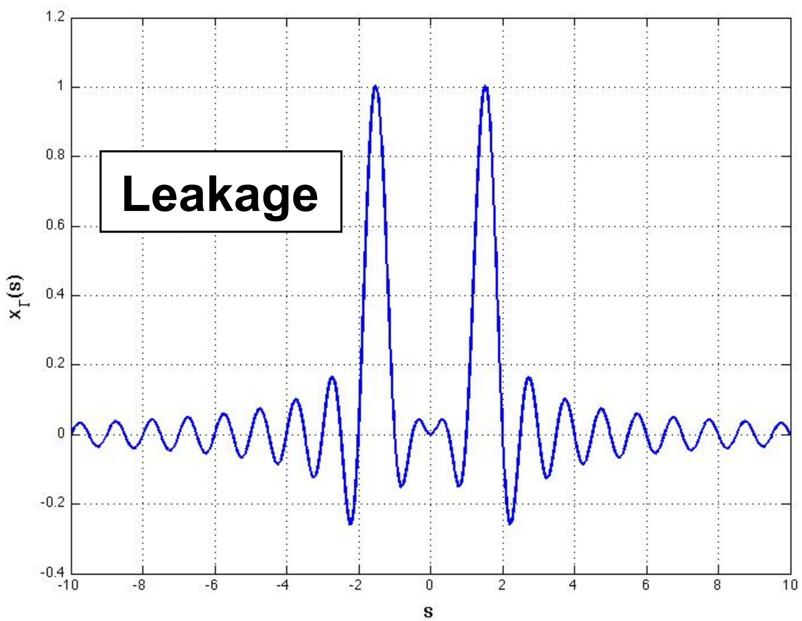
$$x_{\Pi}(t) = \Pi\left(\frac{t}{T}\right)\cos\left(2\pi p_1 t\right) \text{ , where } \quad \Pi(x) = \begin{cases} 0 & if \quad |x| > 1 \\ 1 & if \quad |x| \le 1 \end{cases}$$

With some algebra one finds:

$$\hat{x}_{\Pi}(s) = \left\{ \frac{\sin(2\pi T(s - p_1)}{2\pi(s - p_1)} + \frac{\sin(2\pi T(s + p_1)}{2\pi(s + p_1)} \right\}$$

Exercise: Who does it at the blackboard?

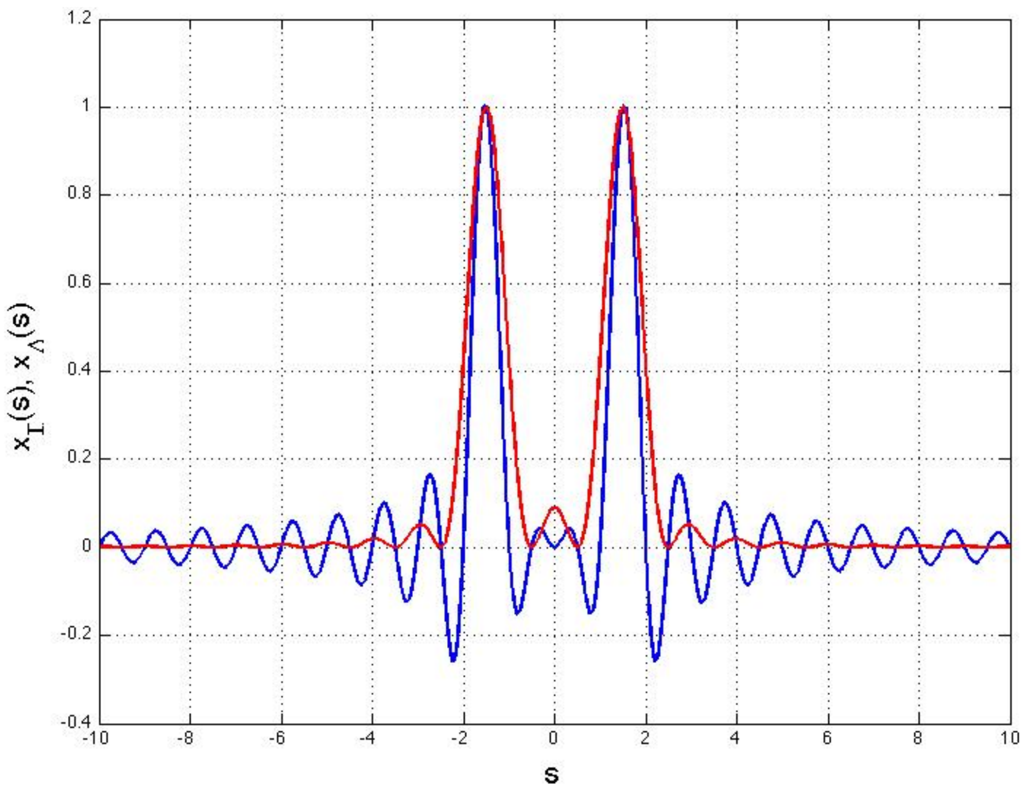
for $T = 1, \; p_1 = 1.5$



Can we reduce this problem? Change the Π function.

For example, instead of a square function, take a triangle:

$$\Lambda(x) = \begin{cases} 0 & if \quad |x| > 1 \\ 1-|x| & if \quad |x| \le 1 \end{cases}$$



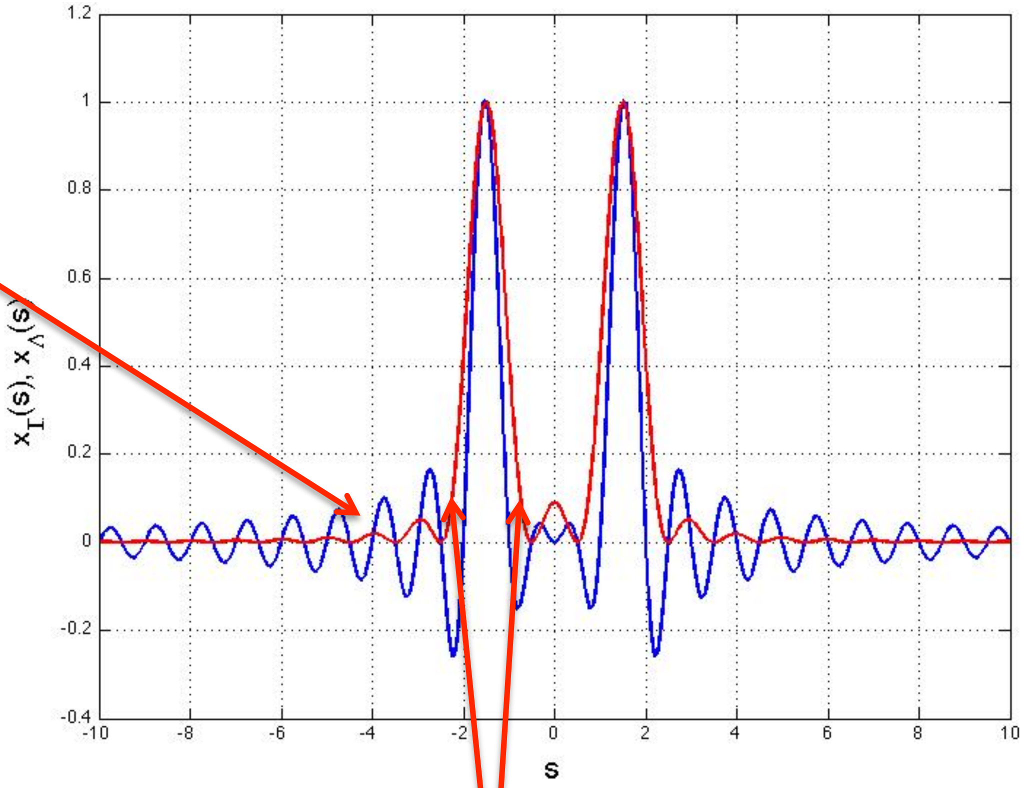
The Fourier transform of

$$x_\Lambda(t) = \Lambda\left(\frac{t}{T}\right)\cos(2\pi p_1 t)$$

gives:

$$\hat{x}_\Lambda(s) = \left\{ \frac{\sin^2(2\pi T(s-p_1))}{[2\pi(s-p_1)]^2} + \frac{\sin^2(2\pi T(s+p_1))}{[2\pi(s+p_1)]^2} \right\}$$

Smaller leakage



There is always a trade off between leakage and resolution

Larger peaks

More sophisticated windows and associated leaking function.
Multiplying the data timeseries by a window is called "tapering".

From the "Numerical Recipes":

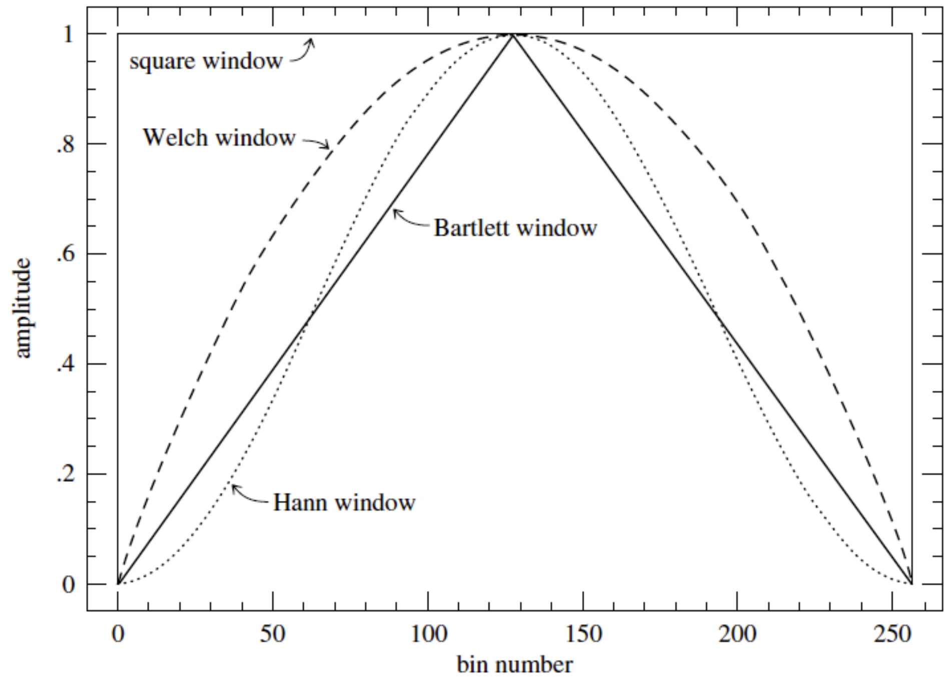


Figure 13.4.1. Window functions commonly used in FFT power spectral estimation. The data segment, here of length 256, is multiplied (bin by bin) by the window function before the FFT is computed. The square window, which is equivalent to no windowing, is least recommended. The Welch and Bartlett windows are good choices.

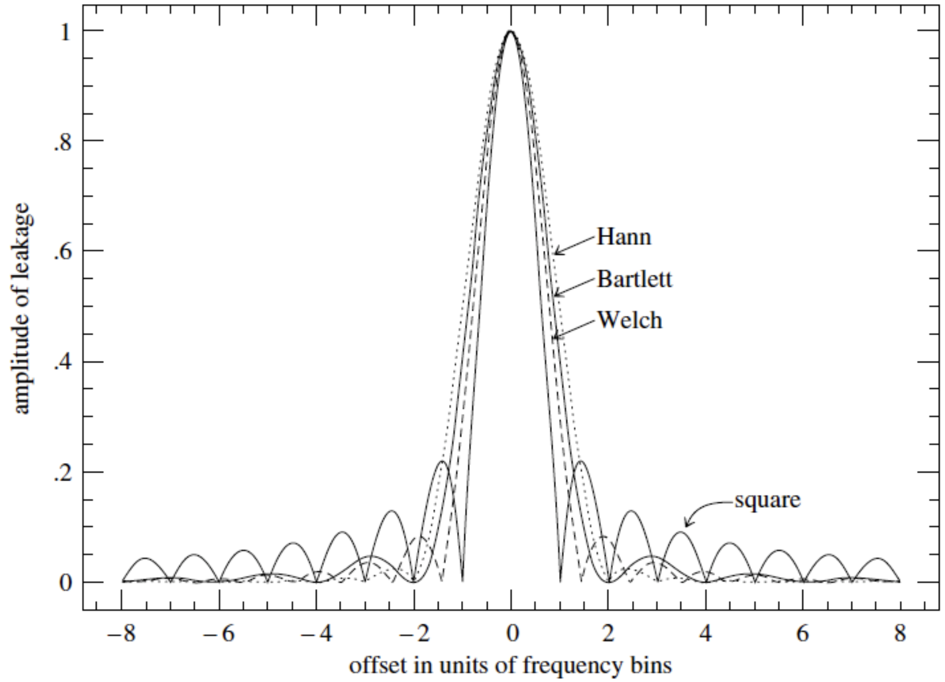


Figure 13.4.2. Leakage functions for the window functions of Figure 13.4.1. A signal whose frequency is actually located at zero offset "leaks" into neighboring bins with the amplitude shown. The purpose of windowing is to reduce the leakage at large offsets, where square (no) windowing has large sidelobes. Offset can have a fractional value, since the actual signal frequency can be located between two frequency bins of the FFT.

Data windowing.

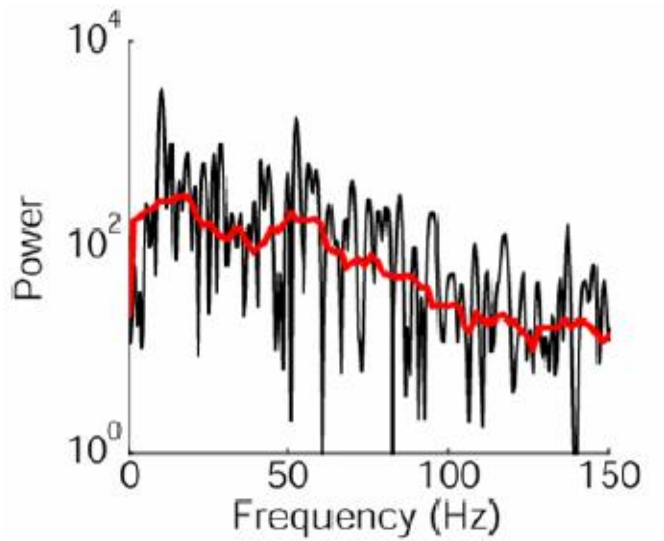
According to the considerations that we made above on spectral leakage, before computing the spectrum, one can multiply the data by one of the well conceived windows that we have seen. (Tapering)

Most software packages propose pre-computed window tapering.

The **periodogram values are subject to substantial random fluctuation**, because they are one realization of an underlying random process. We are thus faced with the problem of very many "chaotic" periodogram spikes. Every determination of the spectral power at a given frequency is affected by an error.

A by-the-eye estimate of the error – the variance – of the power estimates is the width of the wiggles in the periodogram.

How can we reduce the error of estimation of spectral density?



First consideration: **having a longer timeseries does not reduce the error in the estimate of the spectral powe**r at a given frequency, i.e. it doesn't increase the statistical significance of the spectrum found.
- If we take a longer sample at the same sampling rate, we increase the resolution.
- If we take a higher sampling rate, we have a larger spectrum of frequencies.
Remind that the frequencies are given by:

$$f_k = \frac{k}{N\Delta t} \qquad\qquad k = -\frac{N}{2}, \ldots\ldots, \frac{N}{2}$$

Where for $k=-N/2$ and $k=N/2$ we get the Nyquist frequency $\dfrac{1}{2\Delta t}$.

How is this possible? **All the additional information goes into computing more frequency points, not into computing more accurately the same frequencies**.

Instead, we may want to find frequency regions, consisting of many adjacent frequencies, we want to **trade spectral resolution versus statistical significance**.

Two possible strategies to do so:
1) Band averaging
2) Subsampling.

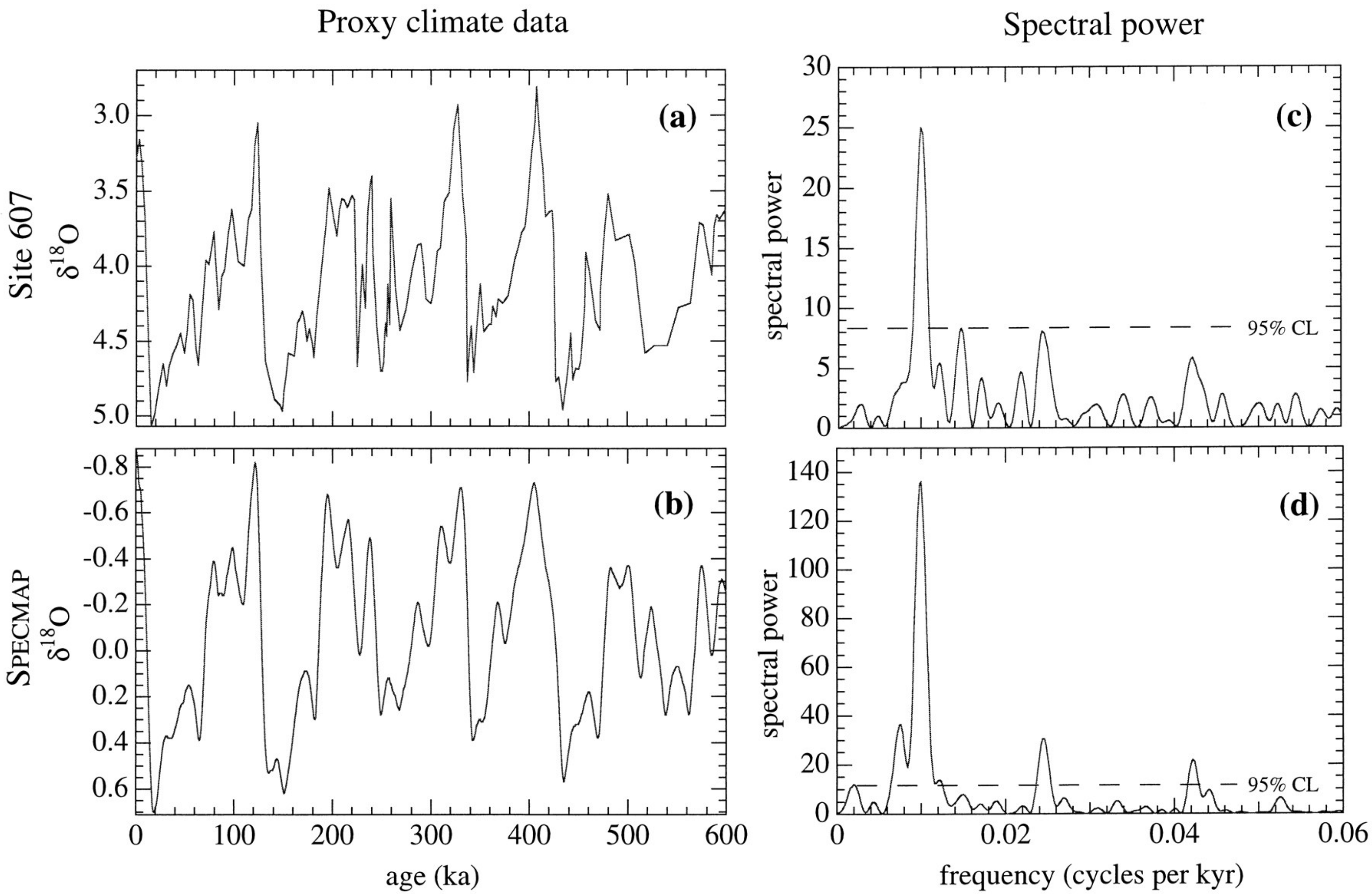
1) Band averaging consists simply in computing a periodogram estimate with finer discrete frequency spacing than you really need, and then to sum the periodogram estimates at K consecutive discrete frequencies to get one “smoother” estimate at the mid-frequency of those K. The variance of that summed estimate will be smaller than the estimate itself by a factor of exactly K.

2) A second technique is to partition the original sampled data into K segments each of M consecutive sampled points. Each segment is separately Fourier-transformed to produce a periodogram. Finally, the K periodogram estimates are averaged at each frequency. It is this final averaging that reduces the variance of the estimate by a factor K (the standard deviation by $K^{1/2}$). This technique is the natural choice for processing long runs of data.
Instead of just cutting the series, one can taper each chunk with a non-square window, in this case, some overlap of the chunk edges is possible.

There is a technique that optimizes the trade-off between reduction of spectral leakage, reduction of error in the estimate of the spectrum and loss of resolution.

The **Multitaper method** computes spectra tapering with windows that are orthogonal to one-another, and that minimize the leakage.

δ18O for past 800 kyr.



Richard A. Muller, and Gordon J. MacDonald PNAS
1997;94:8329-8334

Examples of power spectra

Cullen, D'Arrigo, Cook, Mann. Multiproxy reconstructions of the North Atlantic Oscillation.

Multitaper spectral analysis results of a). the Hurrell [1995] DJFM NAOSLP index over the periods 1874-1979 and b). 1874-1995; c). R4 over the periods 1874-1979 and d). 1750-1979 Figure

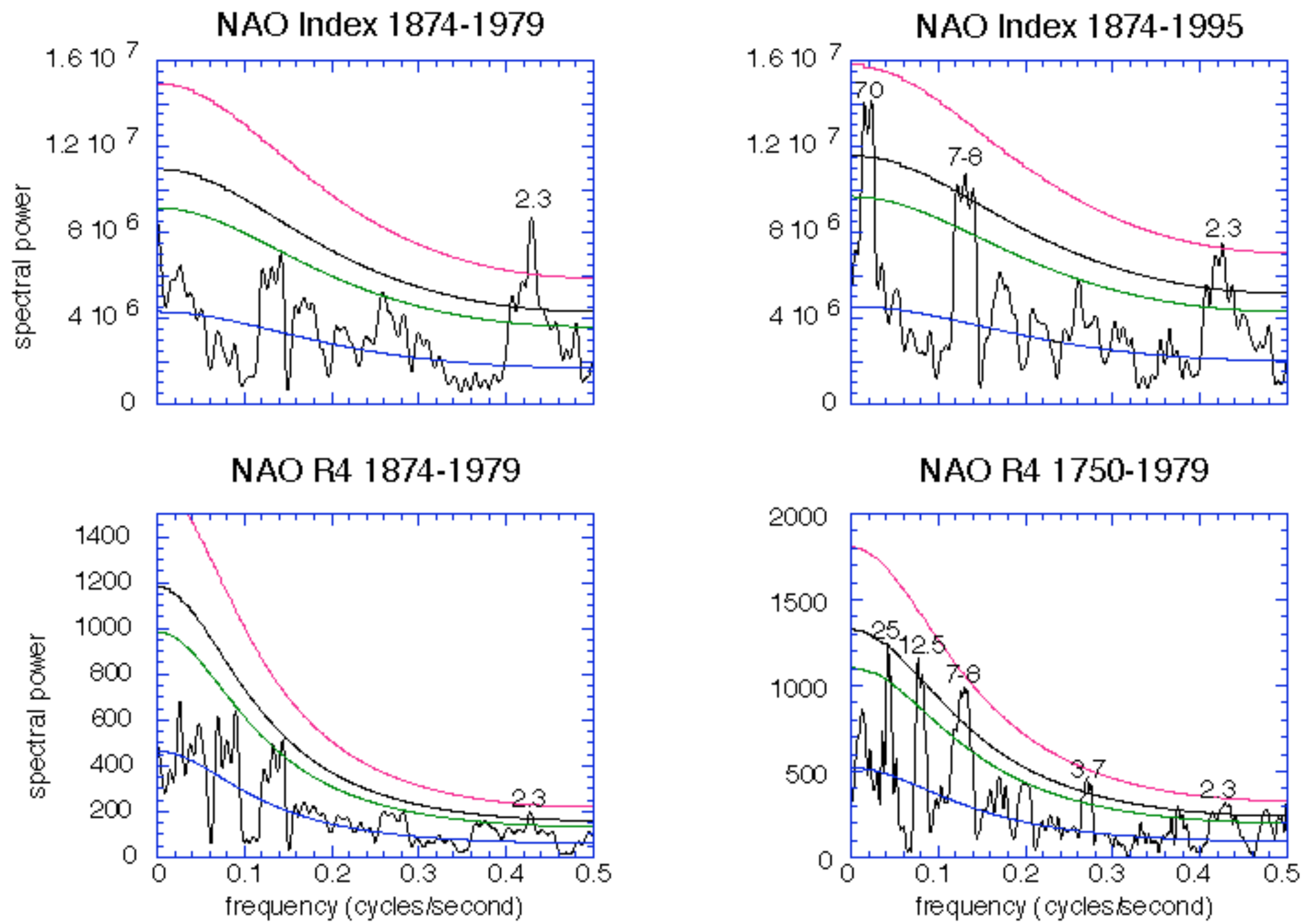


Figure 5

Examples of power spectra

Braganza, K., J. L. Gergis, S. B. Power, J. S. Risbey, and A. M. Fowler (2009), A multiproxy index of the El Niño–Southern Oscillation, A.D. 1525–1982, J. Geophys. Res., 114, D05106, doi:10.1029/2008JD010896.

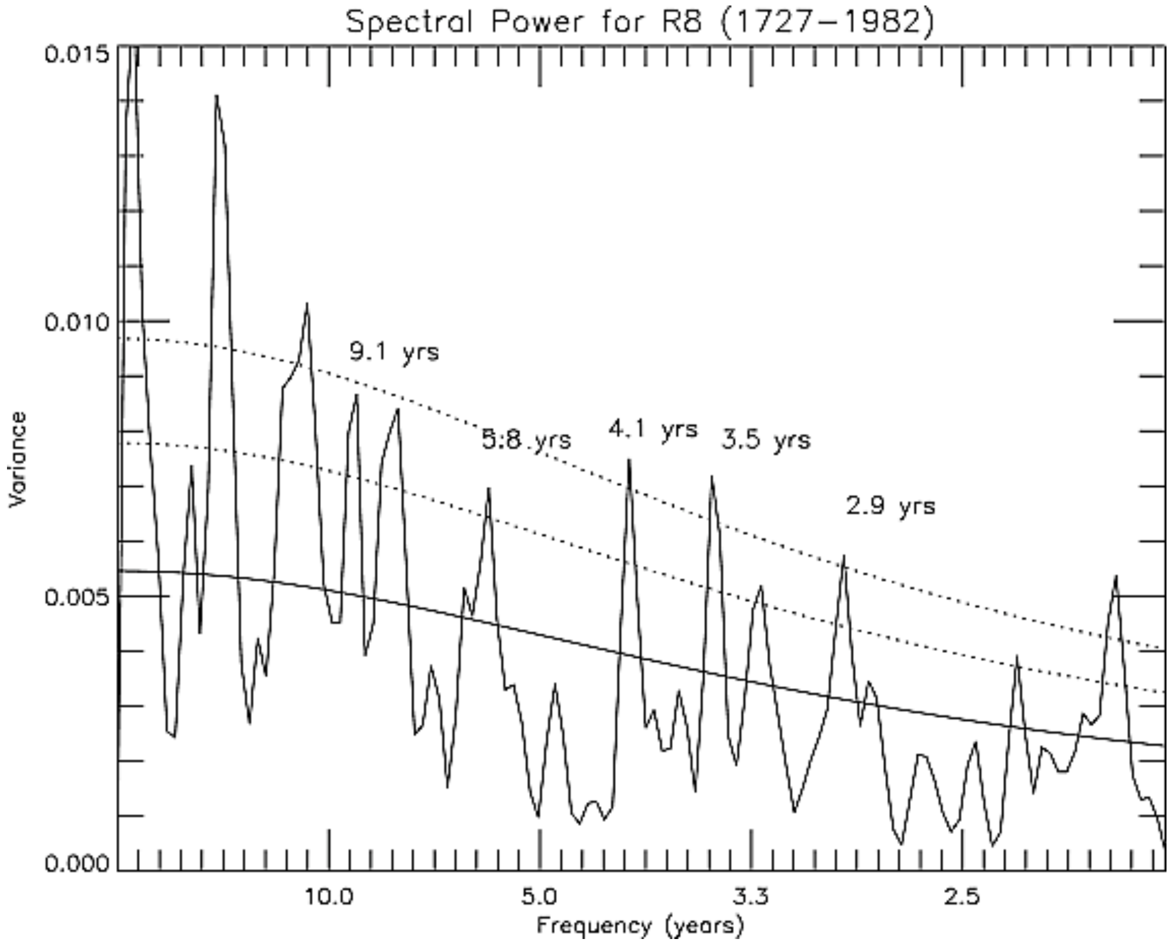
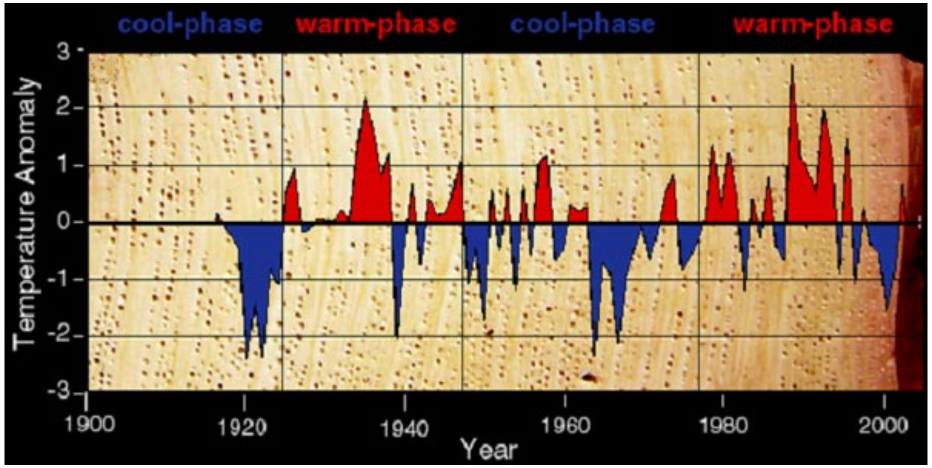
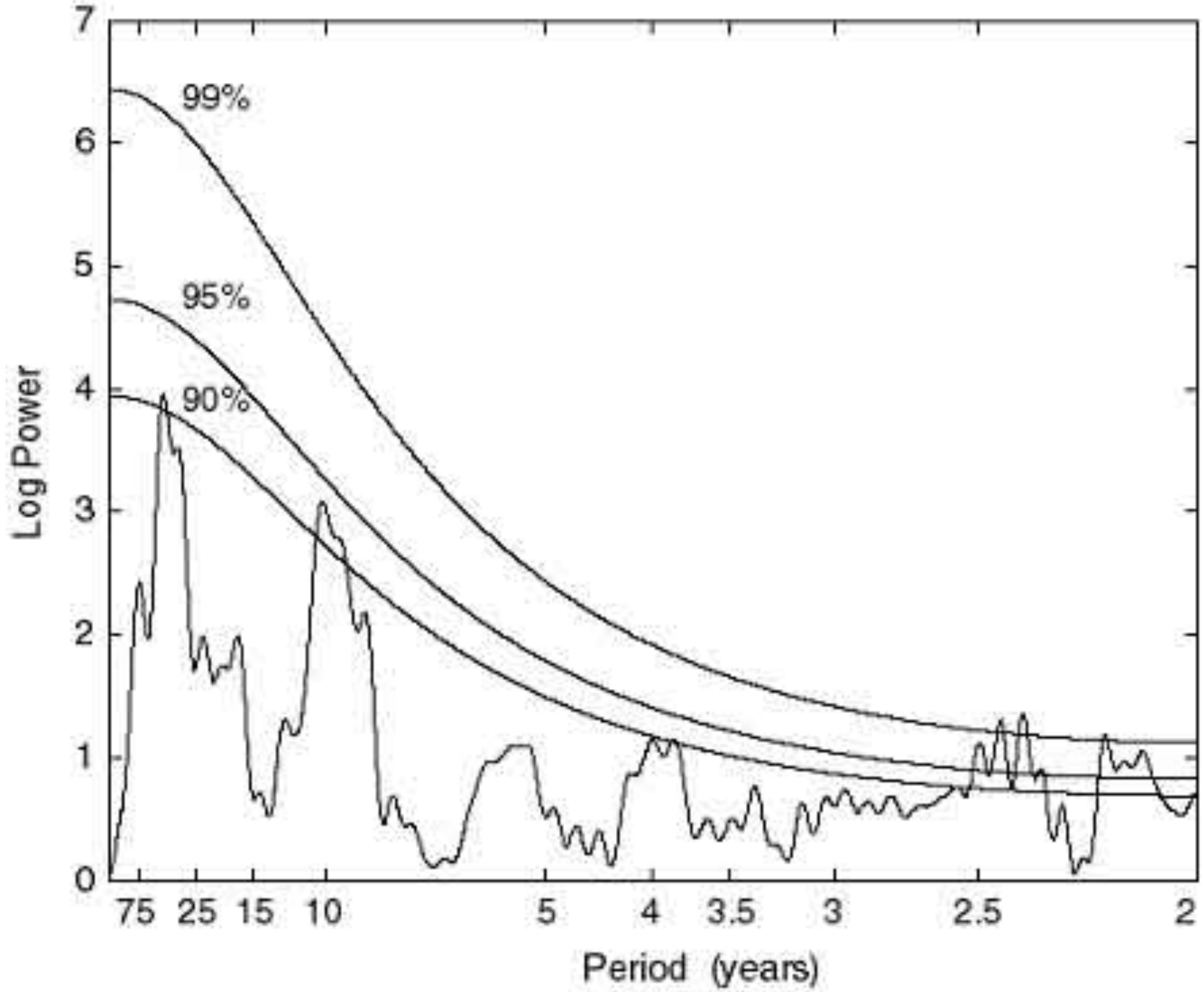


Figure 4. Power spectrum (unnormalized variance) for R8 proxy ENSO index (EOF1) for the period 1727–1982. Significance at the 90% and 95% (dotted lines) level is indicated relative to estimated background AR1 noise (solid line). Effective bandwidth after smoothing = 3.2/N cycles/a.
Enhanced EPS [1.0 MB]

Examples of power spectra





LEFT: Multi-taper power spectrum of the above coralline red algal time series showing significant power at ~60, 10, 4, 2.5 and 2.25 years. The power spectrum corresponds to cyclic climate patterns associated with the El Nino-Southern Oscillation and the Pacific Decadal Oscillation.

What are the statistical significance lines that you see in these figures?

It is the spectrum of the “null hypothesis”, i.e. the hypothesis that the spectrum is the result of a random time series, or the spectrum of noise.

So, one can build a number of rednoise time series that has the same autocorrelation at a given lag τ, as the data sample, as well as the same variance. Then one computes the spectra and extract a PDF, plotting the – say – 95^{th} centile as a function of the frequency.

The spectrum of a rednoise process with a given autocorrelation and variance can also be computed theoretically.

2D application

Figure 12.6.2. Fourier processing of an image. Upper left: Original image. Upper right: Blurred by low-pass filtering. Lower left: Sharpened by enhancing high frequency components. Lower right: Magnitude of the derivative operator as computed in Fourier space.

END

ROADMAP for exercise 2.

1) Get Southern Oscillation Index (SOI) data from the class website.

What's SOI? Check it out: http://www.cgd.ucar.edu/cas/catalog/climind/soi.html.

Transform the data from an array to a vector (reshape…)

2) Compute Fourier transform of the data (commands: fftshift, fft)

3) Define the frequencies.

4) Compute the spectrum (periodogram) and plot it as a function of frequency

Try different graphical representations, log-log, log-lin, lin-lin…

5) Band-average for smoothing the result

Define a window length, sum the spectral coefficient and plot as a function of the central or mean frequency.

6) Define a rednoise model of your data timeseries

The rednoise model is defined by the recursive formula $x_{n+1}=ax_n+w_n$, *where the variance of* x_n *and the autocorrelation at lag 1 are the same as the ones of the data.*

7) Use the rednoise model to build a test of statistical significance of your spectrum.

build 1000 rednoise timeseries, compute the spectrum for each of them, with the same band averaging as the data. Estimate the 95% and 99% centile of the rednoise spectra. Plot on top of the data spectrum. You may also compare with the thoeretical spectrum of the rednoise.

From: Labat et al, 2005. Journal of Hydrology, 314, pp. 289-311

SOI

NAO