Spectral Estimation & Adaptive SP

Nonparametric Spectrum Estimation

Danilo Mandic

room 813, ext: 46271

Department of Electrical and Electronic Engineering
Imperial College London, UK

d.mandic@imperial.ac.uk, URL: www.commsp.ee.ic.ac.uk/∼mandic

Aims

- To introduce methods for **nonparametric** spectrum estimation, which are based on the Fourier transform
- To analyse the periodogram as an estimator and to understand its properties in terms of the mean squared error (MSE) performance
- To derive expressions for the **bias** and **variance** of the periodogram
- To introduce variants of the periodogram and their analyse their properties as estimators of power spectrum
- To understand the trade-off between the periodogram resolution, bias, variance, window function, and data length
- To illustrate practical applications of periodogram (narrowband signal estimation, multiple harmonic components, brain computer interface)

Problem Statement

Estimate Power Spectral Density (PSD) of a wide-sense stationary signal

Recall that PSD $= \mathcal{F}(ACF)$.

Therefore, estimating the power spectrum is equivalent to estimating the autocorrelation.

Recall that for an autocorrelation ergodic process,

$$\lim_{N\to\infty}\left\{\frac{1}{2N+1}\sum_{n=-N}^{N}x(n+k)x(n)\right\}=r_{xx}(k)$$

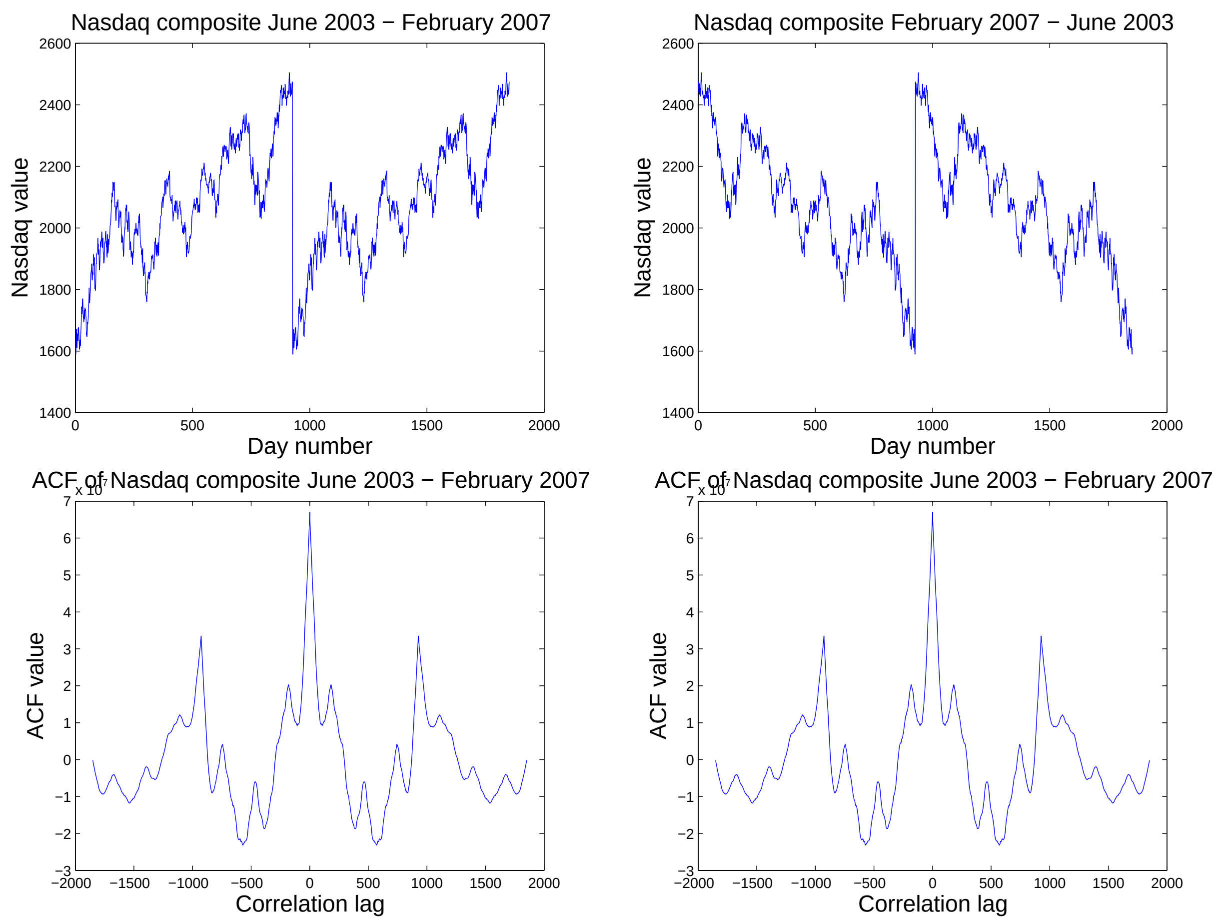
If $x(n)$ is known for all n, estimating the power spectrum is straightforward

- **Difficulty 1:** the amount of data is **always limited**, and may be very small (genomics, biomedical)
- **Difficulty 2:** real world data is **almost invariably corrupted by noise**, or contaminated with an interfering signal

Modelling a financial time series

From: `http://finance.yahoo.com/q/ta?s=%5EIXIC&t=1d&l=on&z=m&q=b&p=v&a=&c=`

Nasdaq ascending **Nasdaq descending**



ACF does not model 'direction'. Also remember the Wold decomp. th.

Physical intuition: Connecting PSD and ACF
positive (semi)-definiteness

Recall: $\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^T\} = \begin{bmatrix} r(0) & r(1) & \cdots & r(N-1) \\ r(1) & r(0) & \cdots & r(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(N-1) & r(N-2) & \cdots & r(0) \end{bmatrix}$

Then, for a linear system with input sequence $\{x\}$, output $\{y\}$, and the vector of coefficients $\mathbf{a}$, the output has the form

$$y(n) = \sum_{k=0}^{N-1} a(k)x(n-k) = \mathbf{x}^T\mathbf{a} = \mathbf{a}^T\mathbf{x} \quad \text{where} \quad \mathbf{a} = [a(0), \ldots, a(N-1)]^T$$

The power $P_y = E\{y^2\}$ is **always** positive, and thus $((\mathrm{a}^T\mathrm{b})^T = \mathrm{b}^T\mathrm{a}^T)$

$$E\{y^2[n]\} = E\{y[n]y^T[n]\} = E\{\mathbf{a}^T\mathbf{x}\mathbf{x}^T\mathbf{a}\} = \mathbf{a}^T E\{\mathbf{x}\mathbf{x}^T\}\mathbf{a} = \mathbf{a}^T\mathbf{R}_{xx}\mathbf{a}$$

$\Rightarrow$ to maintain positive power, the autocorrelation matrix R_{xx} must be positive semidefinite

In other words: a positive semidefinite R_{xx} will alway produce positive power spectrum!

But, is our estimate of ACF always positive definite?

From the last lecture: Two ways to estimate the ACF

For an **autocorrelation ergodic** process with an unlimited amount of data, the ACF may be determined:

1) Using the time–average

$$r_{xx}(k) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n+k)x(n)$$

If $x(n)$ is measured over a finite time interval, $n = 0, 1, \ldots, N-1$ then we need to *estimate* the ACF from a finite sum

$$\hat{r}_{xx}(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n+k)x(n)$$

2) In order to ensure that the values of $x(n)$ that fall outside interval $[0, N-1]$ are excluded from the sum, we have (**biased estimator**)

$$\hat{r}_{xx}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n+k)x(n), \quad k = 0, 1, \ldots, N-1$$

Cases 1) and 2) are equivalent for small lags and a fast decaying ACF

Case 1) gives positive semidefinite ACF, this is not guaranteed for Case 2)

Let us re–define the problem

Therefore, the above problem should be re–defined as **estimating $P_x(e^{\jmath\omega})$ from a finite number of noisy measurements of $x(n)$.**

☺ in some applications, estimating the power spectrum may be facilitated by having prior knowledge about how the process is generated. (e.g. $AR(p,q)$), **revise from Lecture 1.**

In that case, for example, for an AR(p) process we have:

$$y[n] = \sum_{i=1}^{p} a_i y[n-i] + w[n] \qquad P_y = \frac{\sigma_w^2}{|1 - \sum_{i=1}^{p} a_i e^{-\jmath\theta i}|^2}$$

Since estimating ACF $\Leftrightarrow$ estimating PSD:

- **Nonparametric** methods rely on the **direct use** of the available data;
- **Parametric** methods **rely on a model** for the signal generation.

Our choice depends whether we estimate full spectrum or some spectral lines

Choice of the spectrum estimation method

Trade–off between **simple but limited accuracy** nonparametric and **demanding but possibly more accurate** parametric PSD estimates.

Assume that the data sequence is:

- **Ergodic** $\looparrowright$ so that the statistical expectations are replaced by time–averages;
- **Stationary** $\looparrowright$ so that infinite averages are replaced by averages over a finite time interval;
- **Approximation** $\looparrowright$ in terms of some sort of windowing is often necessary.

The periodogram (Schuster 1898):

$$\hat{\mathbf{P}}_{per}(f) = \frac{1}{N}\left|\sum_{k=0}^{N-1} x[k]e^{-\jmath 2\pi f k}\right|^2$$

So, the periodogram $\hat{P}_{per} \sim X(f)X^*(f)$ is Fourier-based, hence non-parametric

Periodogram based estimation of power spectrum
more intuition $\looparrowright$ connection with DFT

A nonparametric estimator the power spectrum – **the periodogram**

$$\hat{P}_{per}(e^{\jmath\omega}) = \sum_{k=-N+1}^{N+1} \hat{r}_{xx}(k)e^{-\jmath k\omega}$$

It is, however, more convenient to express the periodogram in terms of the process $x[n]$ (alternative derivation):

- Notice that $\hat{r}_{xx}(k) = \frac{1}{N}x(k) * x(-k)$
- Apply the FT to obtain

$$\hat{P}_{per}(e^{\jmath\omega}) = \frac{1}{N}X(e^{\jmath\omega})X^{*}(e^{\jmath\omega}) = \frac{1}{N}\left|X(e^{\jmath\omega})\right|^{2}$$

where $X(e^{\jmath\omega}) = \sum_{n=0}^{N-1} x(n)e^{-\jmath\omega n}$. (this is a DTFT of $x(n)$).

Periodogram and Matlab

```
Px=abs(fft(x(n1:n2))).^2/(n2-n1-1)
```

or the direct command '**periodogram**'

- `Pxx = PERIODOGRAM(X)`
  returns the PSD estimate of the signal specified by vector X in the vector Pxx. By default, the signal X is windowed with a BOXCAR window of the same length as X;

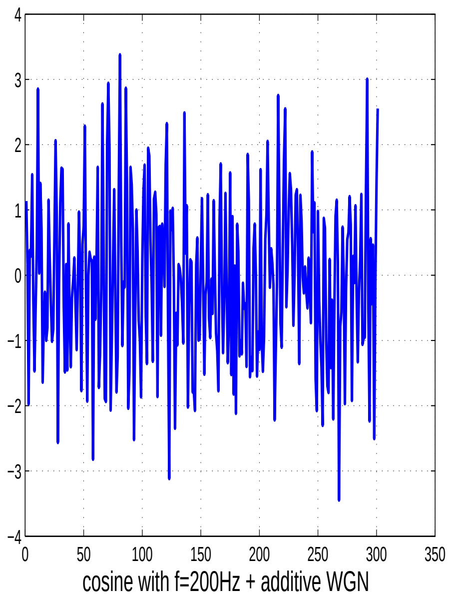
- `PERIODOGRAM(X,WINDOW)`
  specifies a window to be applied to X. WINDOW must be a vector of the same length as X;

- `[Pxx,W] = PERIODOGRAM(X,WINDOW,NFFT)`
  specifies the number of FFT points used to calculate the PSD estimate.

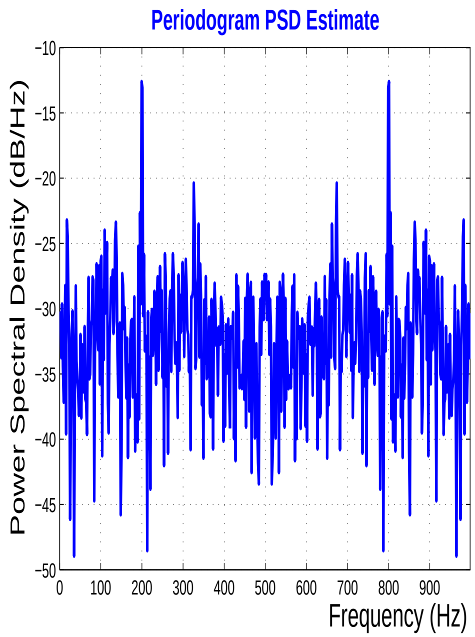
Example: PSD of a cosine in WGN

Calculate the PSD of a signal given by

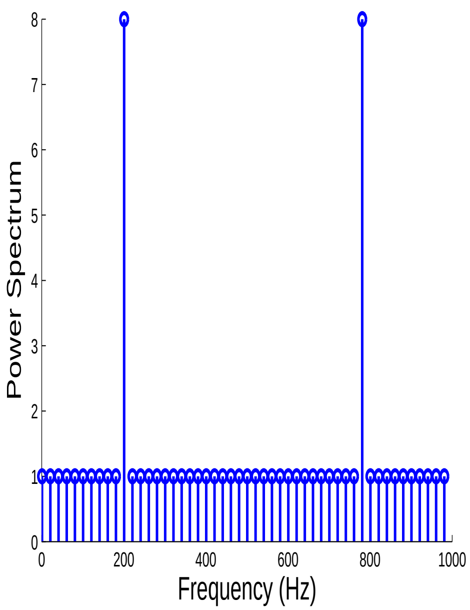
```
x = cos(2*pi*t*200)+randn(size(t))
```



Signal x



Periodogram of x



Ideal PSD

Homework: i) Plot PSDs for varios signal to noise ratios (SNR)

ii) Use both the linear and logarithmic scale (in [dB])

Example: WGN of short duration

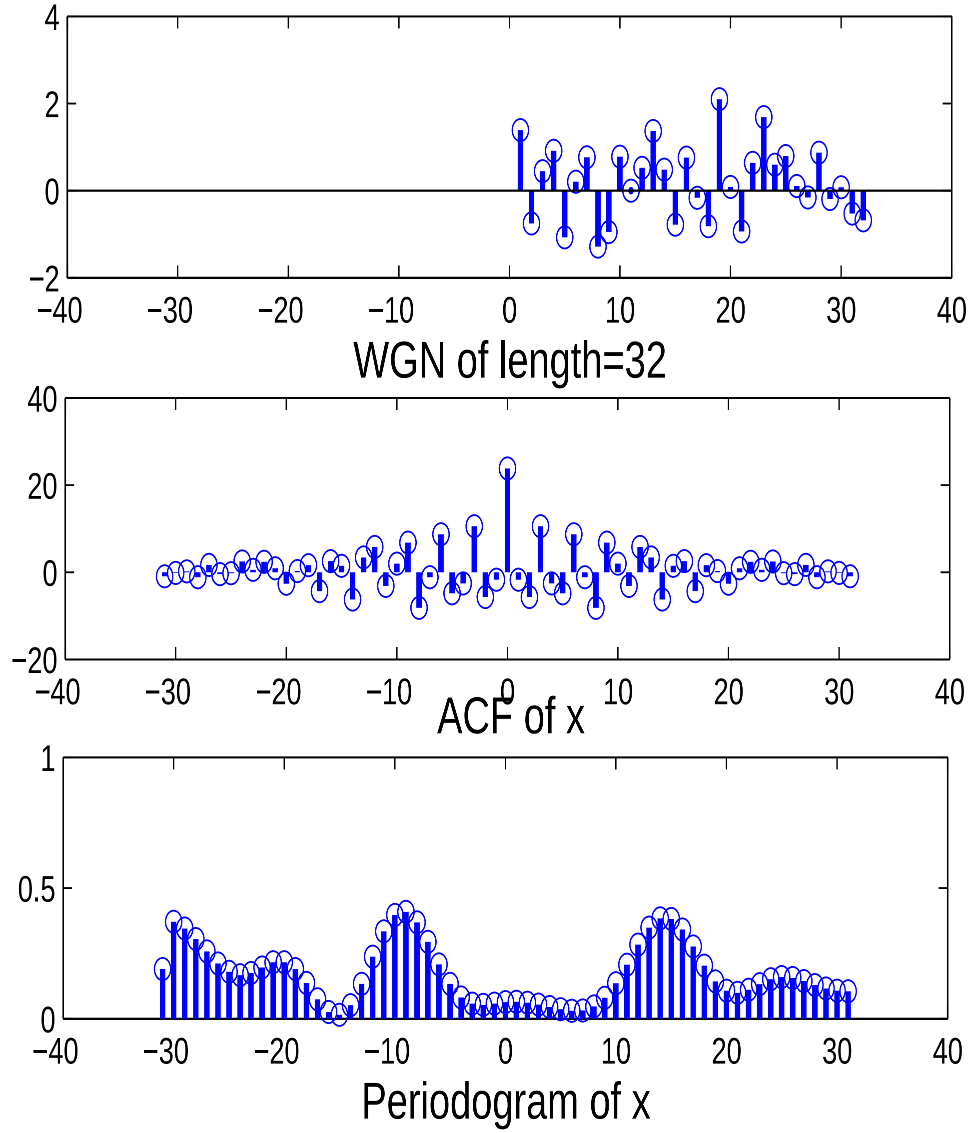
Consider a realisation of WGN, with $n = 0, \ldots, 32$ (file f82.m)

The waveform (top), autocorrelation function (middle) and periodogram (bottom) are shown on the right hand side.

Homework: Try to reproduce these diagrams for N larger. What can you conclude?

Homework: Suggest some ways to generate WGN with flat spectral properties.

Homework: Elaborate any difference in the behaviour of PSD for small N, when considering other pdf's (e.g. uniform, Laplacian, ...)



The bias–variance dilemma
a tool for the performance evaluation of periodograms

The mean square error (MSE) of an estimate $\hat{\theta}$ of a parameter θ is given by

$$MSE(\hat{\theta}) = E\{(\hat{\theta}-\theta)^2\} \quad \text{average deviation from the true value}$$

For every estimator: **Bias:** $B = E\{\hat{\theta}\} - \theta$ **Variance:** $\text{var} = E\{(\hat{\theta} - E\{\hat{\theta}\})^2\}$

Therefore:

$$\begin{aligned}
\text{MSE} &= E\{(\hat{\theta}-\theta)^2\} = E\{[\hat{\theta} - E\{\hat{\theta}\} + \underbrace{E\{\hat{\theta}\} - \theta}_{\text{bias} \;\; \mathrm{B}(\hat{\theta})}]^2\} \\
&= E\{[\hat{\theta} - E\{\hat{\theta}\}]^2\} + E\{B^2(\hat{\theta})\} + 2E\{[\hat{\theta} - E\{\hat{\theta}\}]B(\hat{\theta})\} \\
&\quad \text{due to the linearity of the } E\{\cdot\} \text{ operator and that } E\{B(\hat{\theta})\} = B(\hat{\theta}) \\
&= E\{[\hat{\theta} - E\{\hat{\theta}\}]^2\} + B^2(\hat{\theta}) + \underbrace{2E\{[\hat{\theta} - E\{\hat{\theta}\}]\}}_{=0, \text{ the } E\{\hat{\theta}\} \text{ are equal}} B(\hat{\theta}) \\
&= \text{var}(\hat{\theta}) + B^2(\hat{\theta})
\end{aligned}$$

Performance of the periodogram
(we desire a minimum variance unbiased (MVU) est.)

Its performance is analysed in the same was as the performance of any other estimator:

- **Bias**, that is, whether

$$\lim_{N \to \infty} E\left\{\hat{P}_{per}(f)\right\} = P_x(f)$$

- **Variance**

$$\lim_{N \to \infty} Var\left\{\hat{P}_{per}(f)\right\} = 0$$

- **Mean square convergence**

$$\mathrm{MSE} = \mathrm{bias}^2 + \mathrm{variance} = E\left\{\left[\hat{P}_{per}(f) - P_x(f)\right]^2\right\}$$

$$\text{we} \quad \text{desire} \qquad \lim_{N \to \infty} E\left\{\left[\hat{P}_{per}(f) - P_x(f)\right]^2\right\} = 0$$

☞ we need to check $\hat{P}_{per}(f)$ is a **consistent** estimator of the true PSD.

Bias of the periodogram as an estimator

We can calculate this by finding the expected value of $\hat{\mathbf{r}}_{xx}(k) = \frac{1}{N}\sum_{n=0}^{N-1-|k|} x(n)x(n+|k|)$. Thus (**biased estimate**)

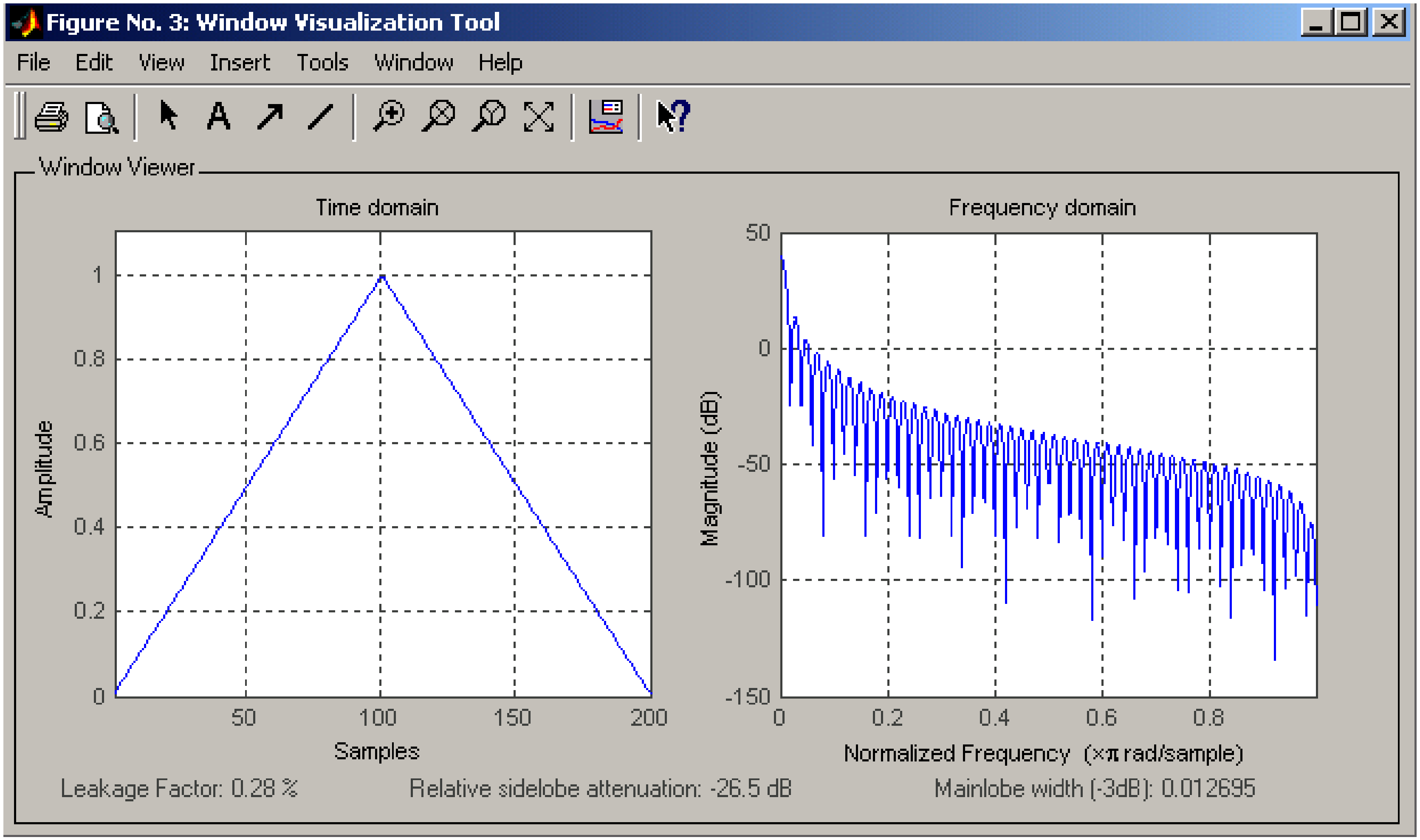
$$
\begin{aligned}
E\{P_{per}(f)\} &= \sum_{k=-(N-1)}^{N-1} E\{\hat{\mathbf{r}}_{xx}(k)\}e^{-\jmath 2\pi f k} \\
&= \sum_{k=-(N-1)}^{N-1} \frac{N-|k|}{N}\,\mathbf{r}_{xx}(k)e^{-\jmath 2\pi f k} = \text{``}\mathbf{w}_B(k) \;\times\; \mathbf{r}_{xx}(k)\text{''}
\end{aligned}
$$

where $\mathbf{r}_{xx}$ is the **true ACF** and the Bartlett (triangular) window is defined by

$$
\mathbf{w}_B(k) = \begin{cases} 1-\frac{|k|}{N}; & |k| \le N \\ 0; & |k| > N-1 \end{cases}
$$

Notice the maximum at n=0, and a slow decay towards the end of the sequence

Effects of the Bartlett window on resolution



Behaves as $sinc^2$

Periodogram bias – continued

From the previous observation, we have

$$E\left\{\hat{P}_{per}(f)\right\} \quad = \quad \sum_{k=-\infty}^{\infty} \mathbf{r}_{xx}(k)\mathbf{w}_B(k)e^{-\jmath 2\pi k f} \Leftrightarrow W_B(f) \; * \; P_{xx}(f)$$

where

$$W_B(f) = \frac{1}{N} \left[\frac{\sin \pi f N}{\sin \pi f} \right]^2.$$

In words, the expected value of the periodogram is the **convolution** of the power spectrum $P_{xx}(f)$ with the Fourier transform of the Bartlett window, and therefore, the periodogram is a **biased** estimate.

Since when $N \rightarrow \infty$, $W_B(f) \rightarrow \delta(0)$, the periodogram is **asymptotically unbiased**

$$\lim_{N\rightarrow\infty} E\left\{\hat{P}_{per}(f)\right\} = P_{xx}(f)$$

Example: Periodogram of WGN

For WGN

$$P_{xx}(f) = \sigma_N^2$$

and

$$E\left\{\hat{P}_{per}(f)\right\} = \sigma_N^2 \int_{-1/2}^{1/2} \frac{1}{N} \left(\frac{\sin \pi (f-\xi) N}{\sin \pi (f-\xi)} \right)^2 d\xi$$

Using the result that

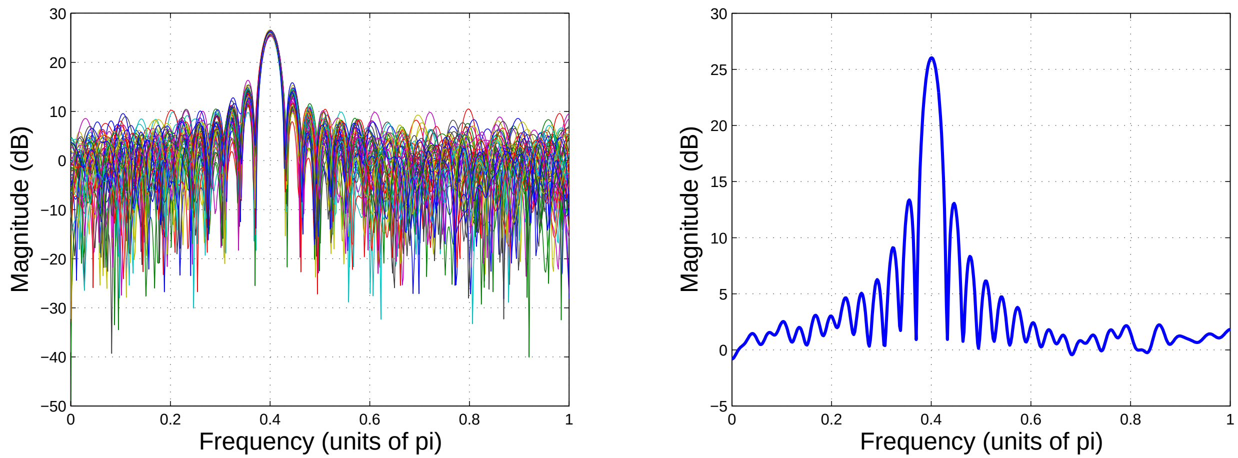
$$w_B(k) = \mathcal{F}^{-1}\left[W_B(f)\right] = \int_{-1/2}^{1/2} \frac{1}{N} \left(\frac{\sin \pi f N}{\sin \pi f} \right)^2 e^{\jmath 2\pi f k} df$$

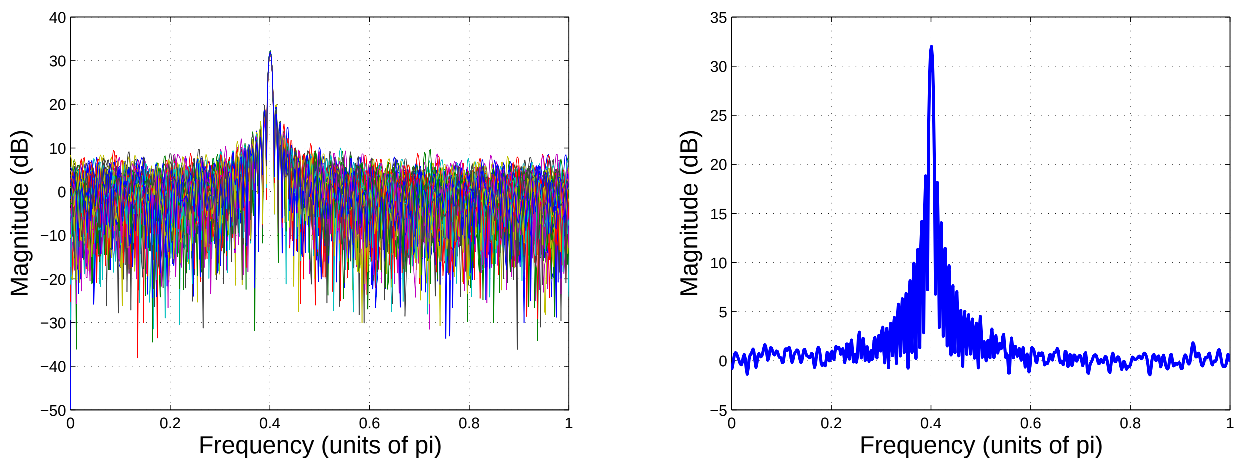
we have $w_B(0) = 1$.

Example: Sinusoid in WGN

$x(n) = A\sin(n\omega_0 + \Phi) + w(n)$, $A = 5, \omega_0 = 0.4\pi$

N=64: Overlay of 50 periodograms   periodogram average





N=256: Overlay of 50 periodograms   periodogram average

Periodogram resolution: Two sinusoids in white noise

This is a random process ($\Phi_1 \perp \Phi_2$, $w(n) \sim \mathcal{U}(0, \sigma_w^2)$) described by :

$$x(n) = A_1 \sin(n\omega_1 + \Phi_1) + A_2 \sin(n\omega_2 + \Phi_2) + w(n)$$

The true PSD is

$$P_{xx}(\omega) = \sigma_w^2 + \frac{1}{2}\pi A_1^2 \left[\delta(\omega - \omega_1) + \delta(\omega + \omega_1)\right] + \frac{1}{2}\pi A_2^2 \left[\delta(\omega - \omega_2) + \delta(\omega + \omega_2)\right]$$

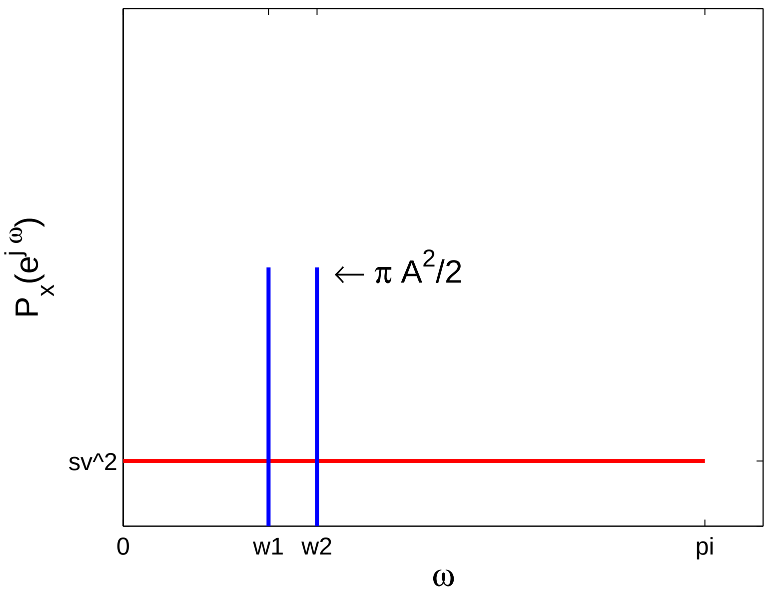
The expected PSD $E\left\{\hat{P}_{per}(\omega)\right\}$ $(P_x * W_B)$ becomes

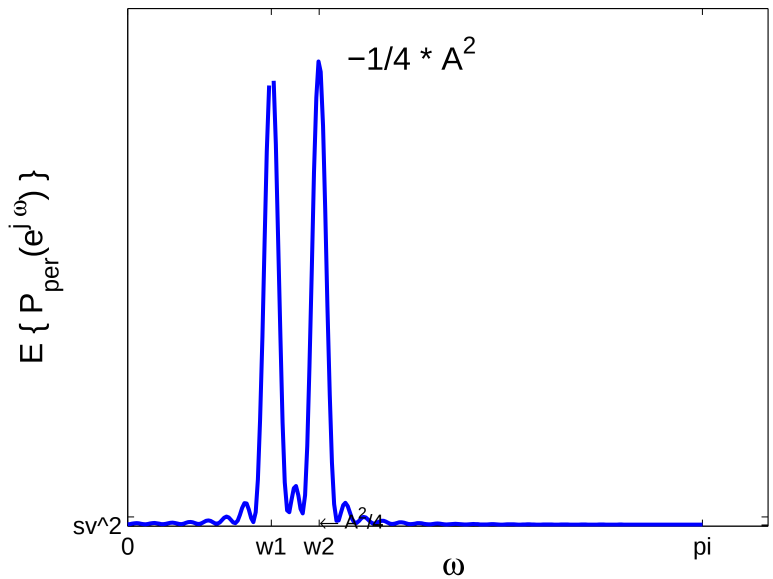
$$\sigma_w^2 + \frac{1}{4}A_1^2 \left[W_B(\omega - \omega_1) + W_B(\omega + \omega_1)\right] + \frac{1}{4}A_2^2 \left[W_B(\omega - \omega) + W_B(\omega + \omega_2)\right]$$

☞ **there is a limit on how closely two sinusoids or two narrowband processes may be located before they can no longer be resolved.**

Sketch of power spectra: Two Sinusoids in WGN

$A_1 = A_2 \qquad N = 64$





N=256: Power spectrum

Expected value of periodogram

☞ **There is a limit on how closely two narrowband processes may be located before they can no longer be resolved**

Homework: Explain why the two peaks are not of the same height

Example: Estimation of two sinusoids in WGN

Based on previous example, try to generate these yourselves

$$x(n) = A_1 \sin(n\omega_1 + \Phi_1) + A_2 \sin(n\omega_2 + \Phi_2) + w(n)$$

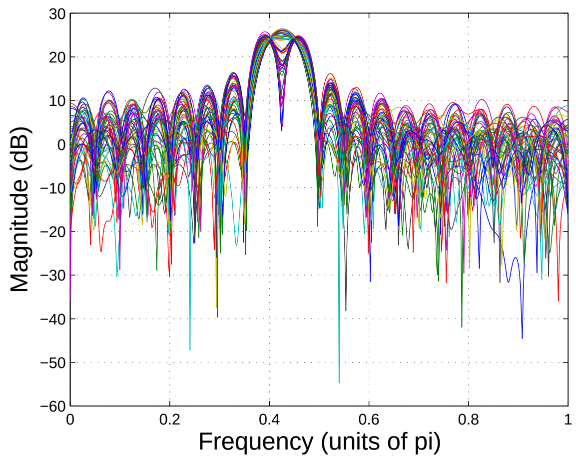
where

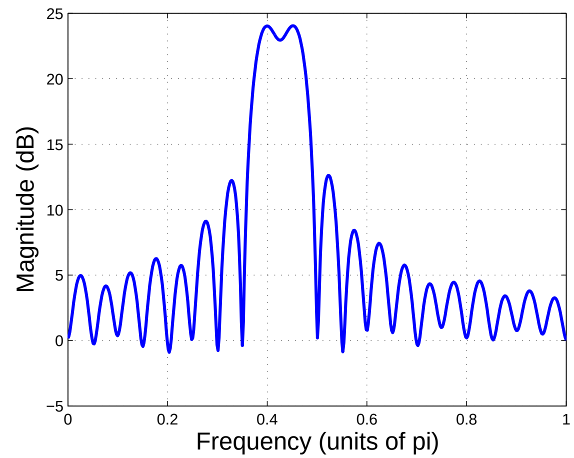
- datalength $N = 40$, $N = 64$, $N = 256$
- $A_1 = A_2$, $\omega_1 = 0.4\pi$, $\omega_2 = 0.45\pi$
- $A_1 \neq A_2$, $\omega_1 = 0.4\pi$, $\omega_2 = 0.45\pi$
- produce overlay plots of 50 periodograms and also averaged periodograms

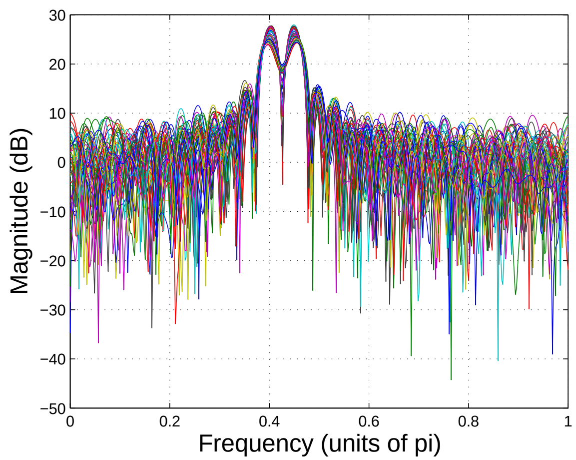
Example: Periodogram resolution $\looparrowright$ two sinusoids

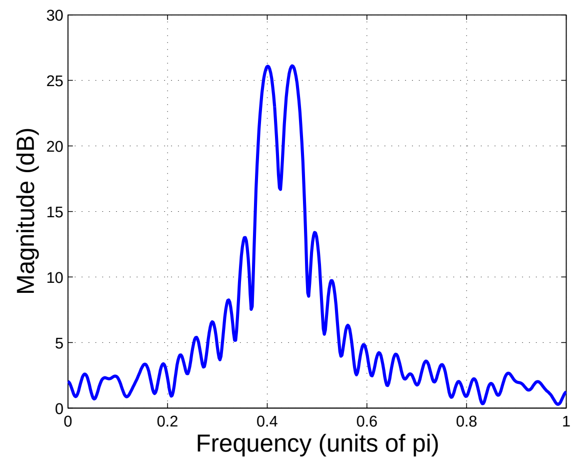
see also Problem 4.6 in your Problem/Answer set

N=40: Overlay of 50 periodograms   periodogram average









N=64: Overlay of 50 periodograms   periodogram average

Effects of the Window Choice

Recall: The spectrum of the (rectangular) window is a *sinc* which has a main lobe and sidelobes

All the other window functions (addressed later) also have the mainlobe and sidelobes.

- The effect of the main lobe (its width) is to **smear** or **smooth** the estimated spectrum shape
- From the previous slide: the width of the mainlobe causes the next peak in the spectrum to be masked if the two peaks are not separated by $1/N$ - the spectral resolution
- The sidelobes cause **spectral leakage** $\looparrowright$ transferring power from the correct frequency bin into the frequency bins which contain no signal power

These effects are dangerous, e.g. when estimating peaky spectra

Some observations

- The Bartlett window **biases** the periodogram;

- It also introduces **smoothing**, which **limits** the ability of the periodogram to resolve closely–spaced narrowband components in $x(n)$;

- This is due to the width of the main lobe of $W_B(f)$;

- Periodogram **averaging would reduce the variance** (remember MVU estimators!)

- **Resolution of the periodogram**

  - set $\Delta\omega =$ width of the main lobe of spectral window, at its "half power"
  - for Bartlett window $\Delta\omega \sim 0.89(2\pi/N) =$ periodogram resolution!
  - notice that the resolution is inversely proportional to the amount of data N

Variance of the periodogram

☹ it is difficult to evaluate the variance of the periodogram of an arbitrary process $x(n)$ since the variance depends on the fourth–order moments of the process.

☺ the variance may be evaluated in the special case of WGN $\longrightarrow$

$$E\left\{\hat{P}_{per}(f_1)\hat{P}_{per}(f_2)\right\} = \left(\frac{1}{N}\right)^2 \sum_k \sum_l \sum_m \sum_n E\left\{x(k)x(l)x(m)x(n)\right\} \quad \times$$
$$\times \quad e^{-\jmath 2\pi[f_1(k-l)+f_2(m-n)]}$$

For WGN, these fourth–order moments become

$$E\left\{x(k)x(l)x(m)x(n)\right\} \quad =$$
$$E\{x(k)x(l)\}E\{x(m)x(n)\} + E\{x(k)x(m)\}E\{x(l)x(n)\} + E\{x(k)x(n)\}E\{x(l)x(m)\}$$
$$= \sigma_x^4\left[\delta(k-l)\delta(m-n) + \delta(k-m)\delta(l-n) + \delta(k-n)\delta(l-m)\right]$$

This is = σ_x^4 if k=l, m=n, or k=m, l=n, or k=n, l=m, or otherwise 0

Variance of the periodogram – contd.

After some simplifications, and recognising

$$\frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{m=0}^{N-1}\sigma_x^4 = \sigma_x^4$$

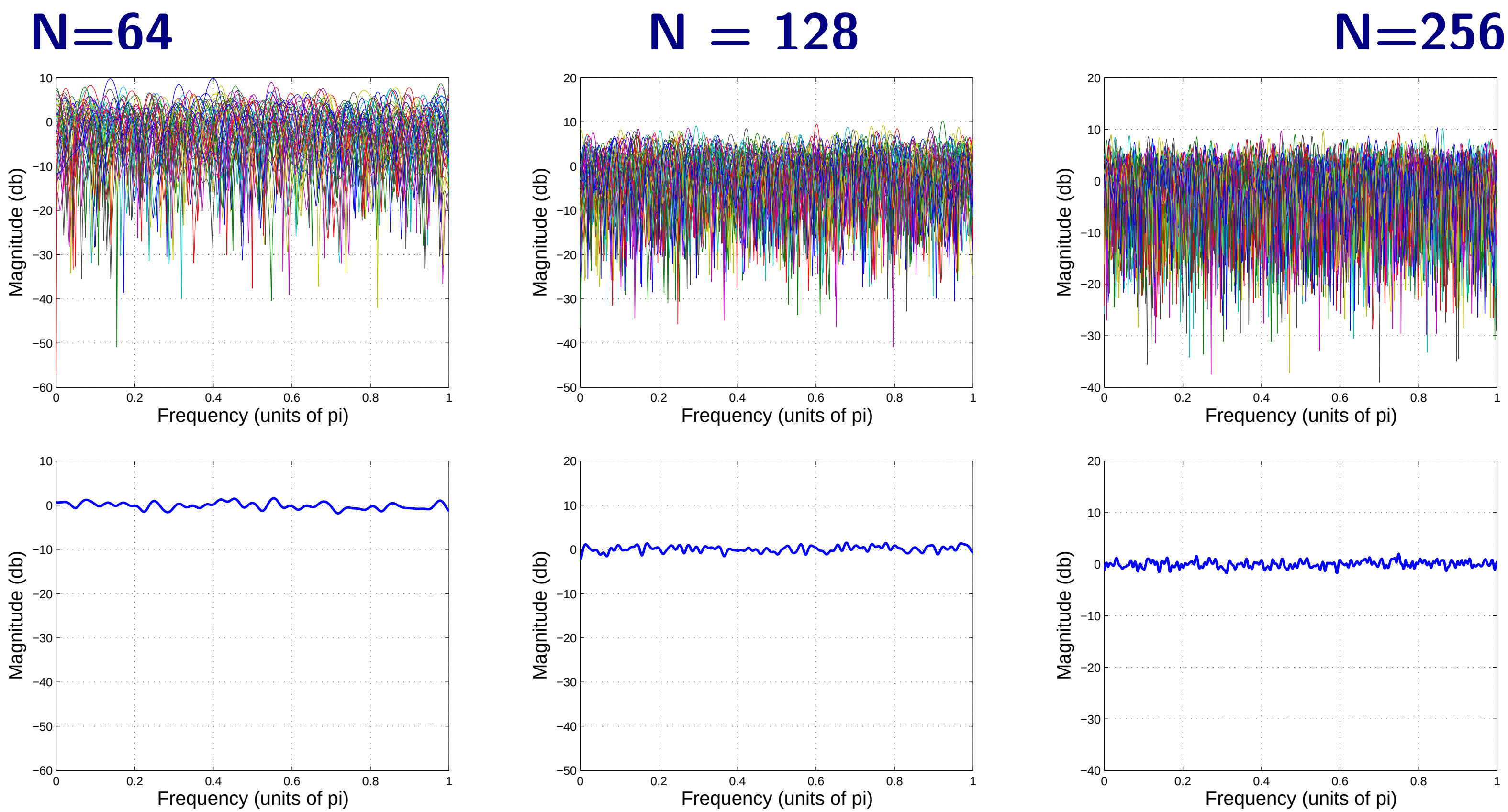
we have the variance of the periodogram for a given frequency:

$$var\left\{\hat{P}_{per}(f)\right\} = P_{xx}^2(f)\left[1 + \left(\frac{\sin 2\pi N f}{N \sin 2\pi f}\right)^2\right]$$

For the periodogram to be consistent, $var(P_{per}) \rightarrow 0 \ as \ N \rightarrow \infty$.

From the above, this is **not** the case $\Rightarrow$ the **periodogram estimator is inconsistent.** In fact, $var(P_{per}(f)) = P_x^2(f) \quad \looparrowright \quad$ quite large

Example: Periodogram of white noise



$$P_{xx} = 1, \quad E\{\hat{P}_{per}(e^{\jmath\omega})\} = 1, \quad var\left[\hat{P}_{per}(e^{\jmath\omega})\right] = 1$$

Although the periodogram is unbiased, the variance is equal to a constant, that is, independent of the data length N

Bias vs variance

Recall that for any estimator, its mean square error (MSE) is given by:

$$\mathrm{MSE} = \mathrm{bias}^2 + \mathrm{variance}$$

A way to overcome periodogram limitations:

- bias performance must be traded for variance performance
- the dataset is divided up into independent blocks
- the periodograms for every block may be averaged
- the resultant estimator is termed the **averaged periodogram**

$$\hat{P}_{aver,per} = \frac{1}{L}\sum_{m=0}^{L-1}\hat{P}_{per}^{(m)}(f)$$

From Estimation Theory: averaging of random trials reduces noise power!

Bias vs variance – recap

- **Bias** pertains to the question: "**Does the estimate approach the correct value as $N \to \infty$**".
  - If yes then the estimator is unbiased, else it is biased
  - Notice that the main lobe of the window has a width of $2\pi/N$ and hence when $N \to \infty$ we have $\lim_{N\to\infty} \hat{P}_{per}(f) = P_{xx}(f) \Rightarrow$ periodogram is an **asymptotically unbiased** estimator of true PSD.
  - **For the window to yield an unbiased estimator:**

$$\sum_{n=0}^{N-1} w^2(n) = N \quad \& \quad \text{the mainlobe width} \sim \frac{1}{N}$$

- **Variance** refers to the "goodness" of the estimate, that is, whether the power of the estimation error tend to zero when $N \to \infty$.
  - We have shown that even for a very large window the variance of the estimate is as large as the true PSD
  - This means that the periodogram **is not a consistent** estimator of true PSD.

Properties of the standard periodogram

Functional relationship:

$$\hat{P}_{per}(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-\jmath n\omega} \right|^2$$

- **Bias**

$$E\left\{\hat{P}_{per}(\omega)\right\} = \frac{1}{2\pi} P_x(\omega) \ * \ W_B(\omega)$$

- **Resolution**

$$\Delta\omega = 0.89 \frac{2\pi}{N}$$

- **Variance**

$$Var\left\{\hat{P}_{per}(\omega)\right\} \approx P_x^2(\omega)$$

Periodogram modifications $\looparrowright$ some intuition

Clearly, we need to reduce the variance of the periodogram, since in general they are not adequate for precise estimation of PSD.

We can think of several modifications:

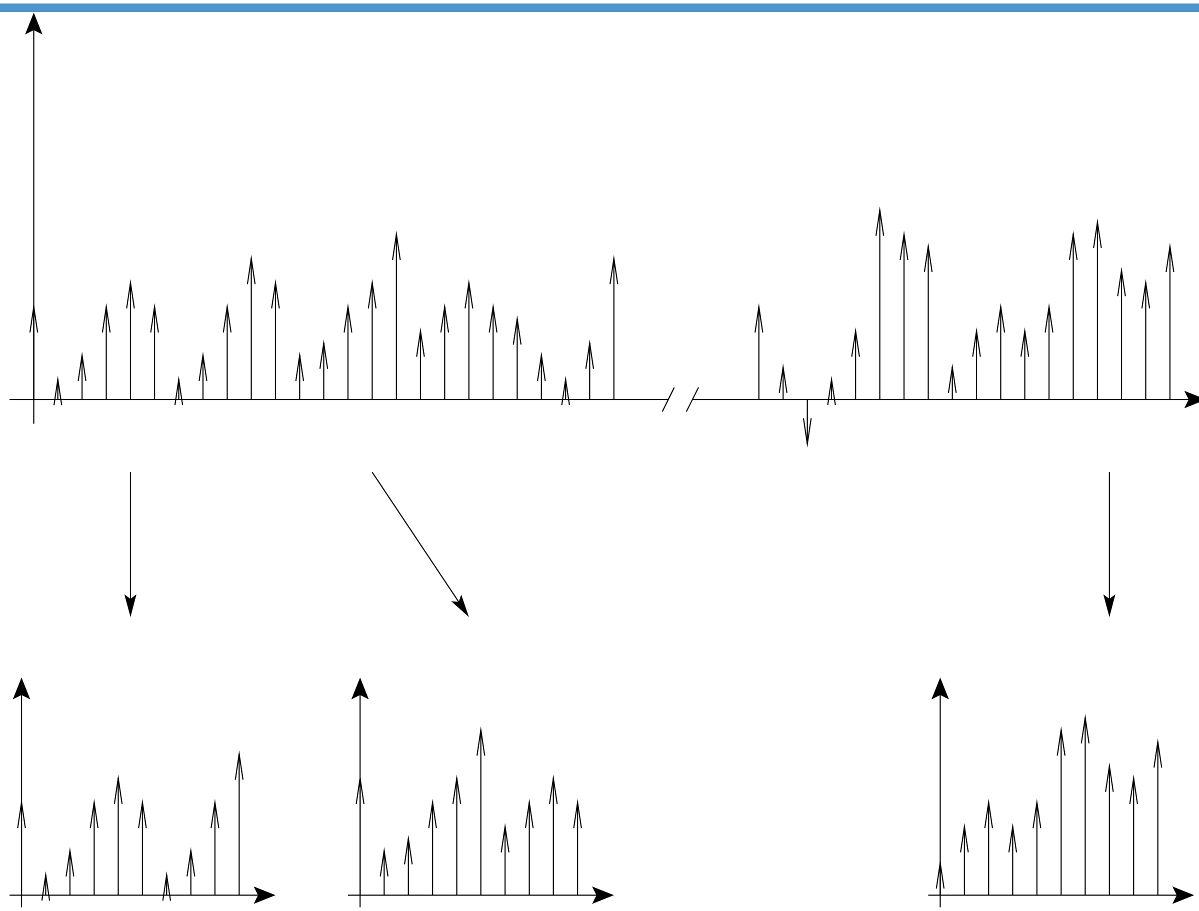
1) **averaging over a set of periodograms** (we have already seen the effect of this in some simulations).

   Recall that from the general estimation theory, by averaging M times we have the effect of $var \rightarrow var/M$.

2) **applying different windows** $\looparrowright$ it is possible to choose or design a window which will have a narrow mainlobe

3) **overlapping windowed segments** for additional variance reduction $\looparrowright$ averaging periodograms along one realisation of a random process (instead of across the ensemble)

Partitioning the data set (K segments of length L each)



Partitioning $x(n)$ into K non–overlapping segments

This way, the total length $N = K \times L$

© D. P. Mandic

Bartlett's method: Averaging periodograms

The **averaged** periodogram can be expressed as:

$$\hat{P}_{aver,per}(f) = \frac{1}{K}\sum_{m=1}^{K}\hat{P}_{per}^{(m)}(f)$$

where for each of the K segments, the segment-wise PSD estimate $P_{per}^{(i)}, \; i = 1, \dots, K$ is given by

$$P_{per}^{(i)}(\omega) = \frac{1}{L}\left|\sum_{n=0}^{L-1} x_i[n]e^{-\jmath n\omega}\right|^2$$

- Idea: to reduce the variance by the factor of "K" $=$ total number of blocks
- Therefore: provided that the blocks are statistically independent (not often the case in practice) we desire to have

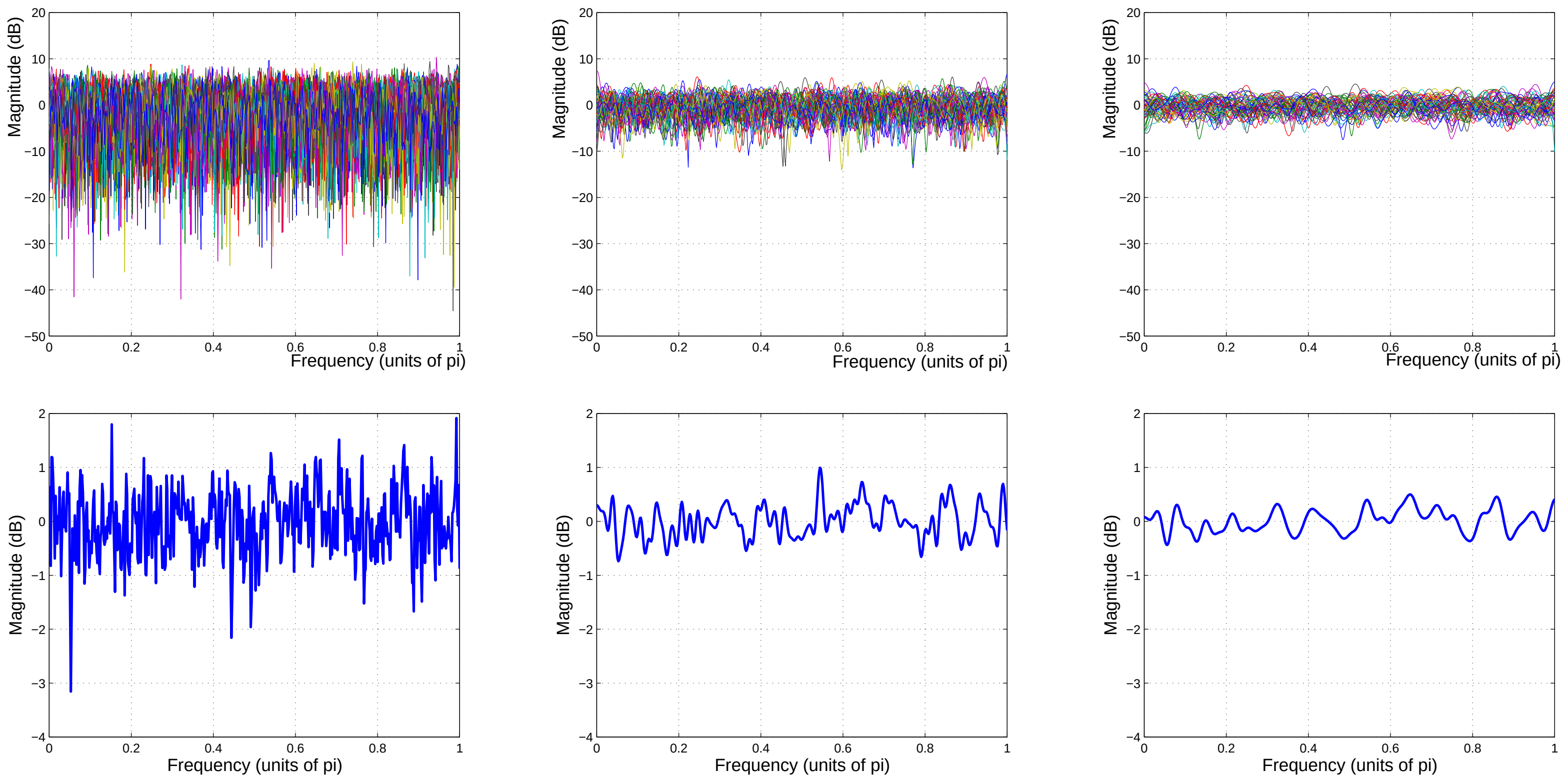
$$var\left\{\hat{P}_{aver,per}(f)\right\} = \frac{1}{K} var\left\{\hat{P}_{per}(f)\right\}$$

Example: Estimation of WGN spectrum using Bartlett's method

50 periodograms with $N = 512$

50 Bartlett estimates $K = 4, \; L = 128$

50 Bartlett estimates $K = 8, \; L = 64$



Ensemble averages

Variance of Bartlett's method

Averaging a set of uncorrelated measurements of a random variable x yields a consistent estimate of the mean $E\{x\}$.

Consider $x_i[n], \quad i = 1, 2, \ldots, K$ be K uncorrelated realisations of a random process $x[n]$ over the interval $0 \leq n < L$. Then

$$\hat{P}^{(i)}_{per}(\omega) = \frac{1}{L} \left| \sum_{n=0}^{L-1} x_i[n] e^{-\jmath n \omega} \right|^2 \quad \Rightarrow \quad \hat{P}_{aver,per} = \frac{1}{K} \sum_{i=1}^{K} \hat{P}^{(i)}_{per}(\omega)$$

Evaluating $\hat{P}_{aver,per}(\omega)$, we have (**asymptotically unbiased**)

$$E\left\{\hat{P}_{aver,per}(\omega)\right\} = E\left\{\hat{P}^{(i)}_{per}(\omega)\right\} = \frac{1}{2\pi} P_x(\omega) \ * \ W_B(\omega)$$

and for the **variance:**

$$Var\left\{\hat{P}_{aver,per}(\omega)\right\} = \frac{1}{K}\, Var\left\{\hat{P}^{(i)}_{per}(\omega)\right\} \approx \frac{1}{K}\, P_x^2(\omega)$$

$\Rightarrow$ Var $\rightarrow 0$ as $K \rightarrow \infty$ **consistent estimator** ☺

Pros and cons of Bartlett's method

NB: Bartlett's method gives a consistent estimate of the power spectrum provided that both K and L are allowed to go to ∞.

Problem: uncorrelated realisations of a process generally not available!

Bartlett's solution: $x[n]$ be partitioned into K non–overlapping sequences of length $L = \frac{N}{K}$, with

$$x_i[n] = x[n + iL], \qquad n = 0, 1, \ldots, L-1, \quad i = 0, 1, \ldots, K-1$$

$$\hat{P}_B(\omega) = \frac{1}{N} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} x[n + iL] e^{-\jmath n\omega} \right|^2$$

In Matlab

```
L=floor(length(x)/nsect);
Px=0; n1=1;
for i=1:nsect
Px=Px+periodogram(x(n1:n1+L-1))/nsect;
end
```

Performance evaluation of Bartlett's method

- **Bias:** The expected value of Bartlett's estimate

$$E\left\{\hat{P}_B(\omega)\right\} = \frac{1}{2\pi} P_x(\omega) \ * \ W_B(\omega)$$

  $\Rightarrow$ **asymptotically unbiased**.

- **Resolution:** Due to K segments of length L, as a consequence we have that Res$(P_B) <$ Res(P_{per}), that is

$$Res\left[\hat{P}_B(\omega)\right] = 0.89\,\frac{2\pi}{L} = 0.89\ K\,\frac{2\pi}{N}$$

- **Variance:**

$$Var\left\{\hat{P}_B(\omega)\right\} \approx \frac{1}{K}\,Var\left\{\hat{P}_{per}^{(i)}(\omega)\right\} \approx \frac{1}{K}\,P_x^2(\omega)$$

  For non–white data, variance reduction is not as large as K times!

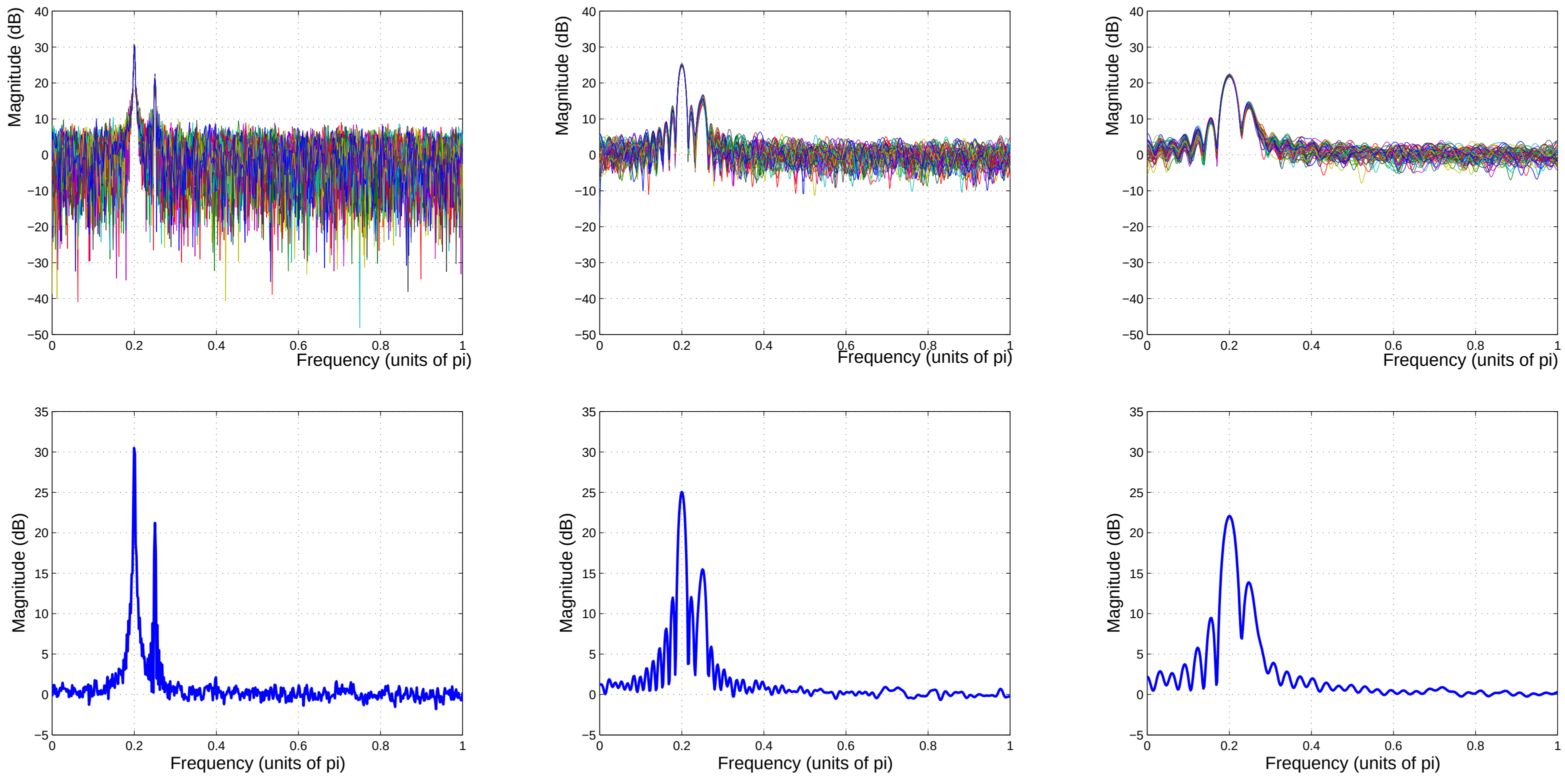
By changing the values of L and K, Bartlett's method allows us to:

trade a reduction in spectral resolution for a reduction in variance

Example: Estimation of two sinewaves in white noise

$$x[n] = \sqrt{10} sin(n * 0.2\pi + \Phi_1) + sin(n * 0.25\pi + \Phi_2) + w[n]$$

50 periodograms with $N = 512$ | **50 Bartlett estimates $K = 4, \; L = 128$** | **50 Bartlett estimates $K = 8, \; L = 64$**



Ensemble averages

Notice the variance – resolution trade–off!

The Modified Periodogram

The periodogram of a process that is windowed with a suitable general window $w[n]$ is called a **modified periodogram** and is given by:

$$\hat{P}_M(\omega) = \frac{1}{NU} \left| \sum_{n=-\infty}^{\infty} x[n]w[n]e^{-\jmath n\omega} \right|^2$$

where N is the window length and $U = \frac{1}{N}\sum_{n=0}^{N-1} |w[n]|^2$ is a constant, **and is defined so that $\hat{P}_M(\omega)$ is asymptotically unbiased.**

In Matlab:

```
xw=x(n1:n2).*w/norm(w);
Pm=N * periodogram(xw);
```

where, for different windows

```
w=hanning(N); w=bartlett(N);w=blackman(n);
```

The Modified Periodogram – "Windowing"

Recall that

$$Periodogram \quad \sim \quad \mathcal{F}\left(|x[n]w_r[n]|^2\right)$$

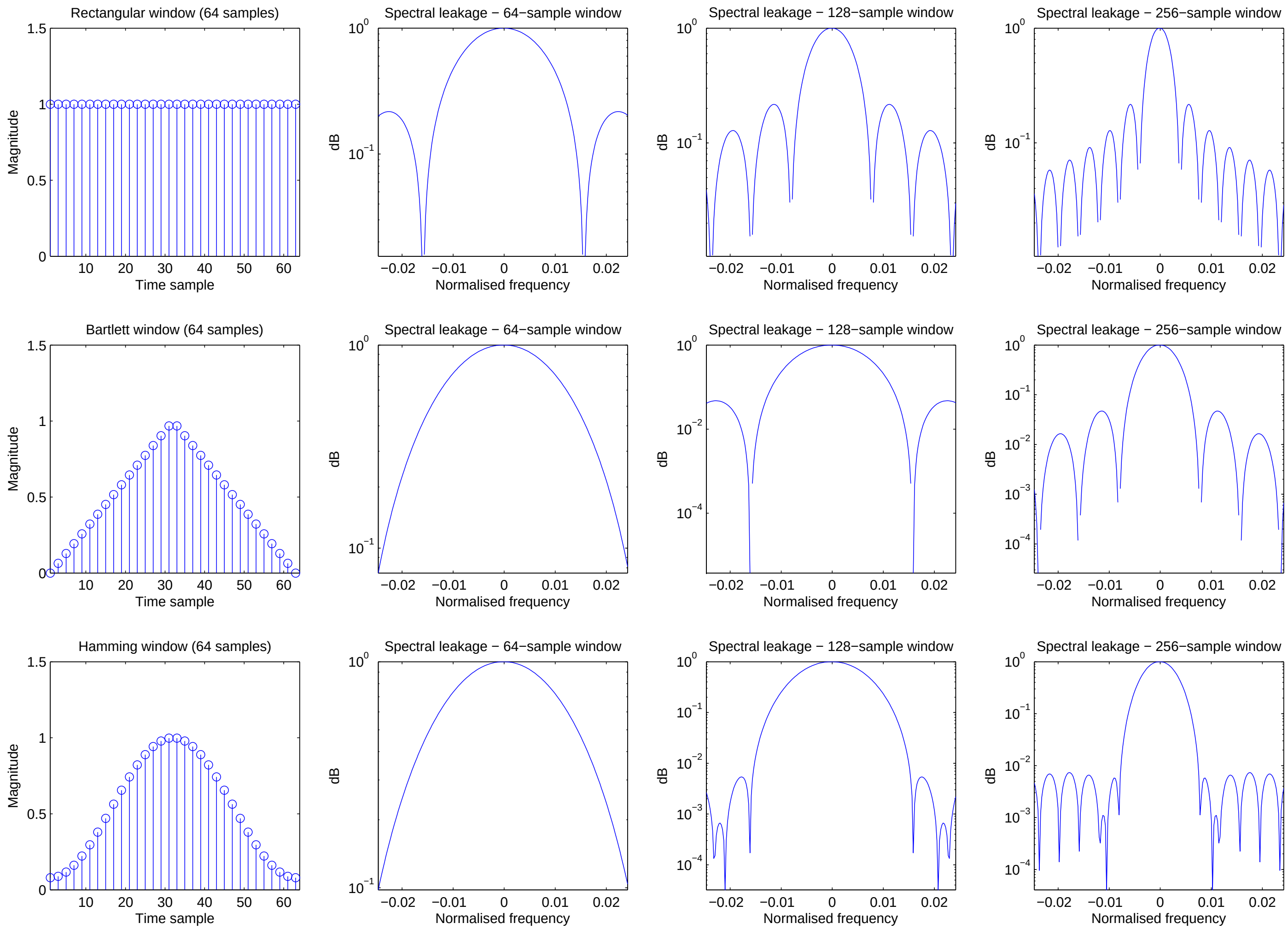
Therefore: The amount of smoothing in the periodogram is determined by the window that is applied to the data. For instance, a rectangular window has a narrow main lobe (and hence least amount of spectral smoothing), but its relatively large sidelobes may lead to masking of weak narrowband components.

Question: Would there be any benefit of using a different data window on the bias and resolution of the periodogram.

Example: can we differentiate between the following two sinusoids for $\omega_1 = 0.2\pi, \omega_2 = 0.3\pi, N = 128$

$$x[n] = 0.1\sin(n\omega_1 + \Phi_1) + \sin(n\omega_2 + \Phi_2) + w[n]$$

More on the properties of window functions



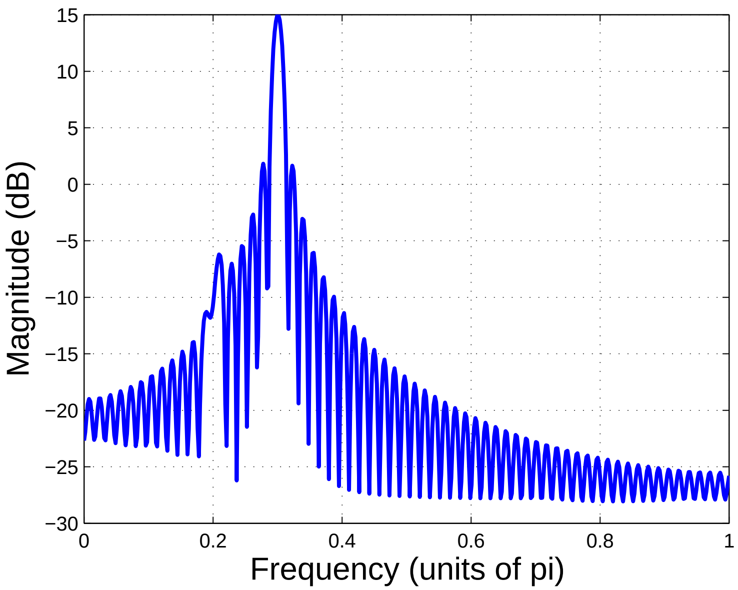
The level of sidelobes does not depend on the amount of data N!

Example: Estimation of two sinusoids in WGN
Modified periodogram using Hamming window

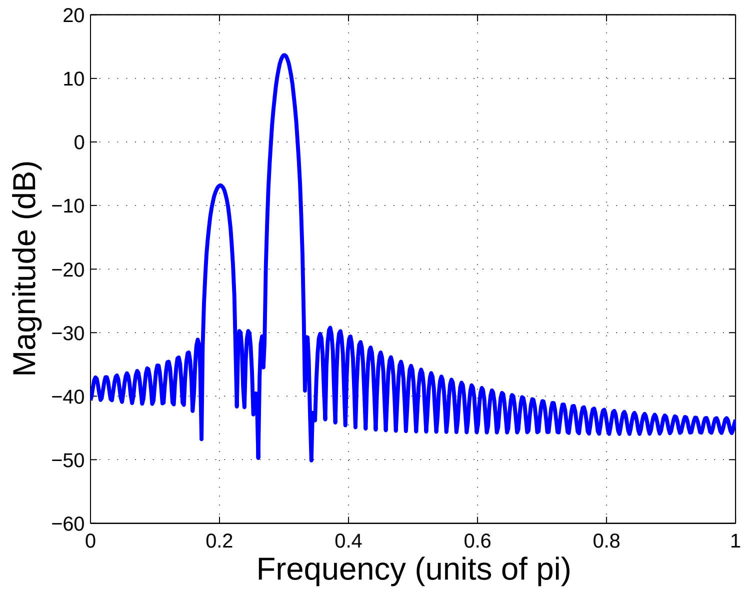
Problem: Estimate spectra of the following two sinusoids using: (a) The standard periodogram; (b) Hamming-windowed periodogram

$$x[n] = 0.1\sin(n*0.2\pi + \Phi_1) + \sin(n*0.3\pi + \Phi_2) + w[n] \qquad N = 128$$

Hamming window $\qquad w[n] = 0.54 - 0.46\cos\left(2\pi\frac{n}{N}\right)$



Expected value of periodogram



Periodogram using Hamming window

Properties of an ideal window function

Consider a window sequence $w(n)$ whose DFT is a **squared magnitude of another sequence** $v(n)$, that is

$$V(\omega) = \sum_{k=0}^{M-1} v(k)e^{-\jmath\omega k} \quad \looparrowright \quad W(\omega) = |V(\omega)|^2 \quad (\textbf{positive} \;\; \textbf{definite})$$

Then

$$\sum_{k=-(M-1)}^{M-1} w(k)e^{-\jmath\omega k} = \sum_{n=0}^{M-1}\sum_{p=0}^{M-1} v(n)v(p)e^{-\jmath(n-p)}$$

$$= \sum_{k=-(M-1)}^{M-1} \Big[\sum_{n=0}^{M-1} v(n)v(n-k)\Big] e^{-\jmath k}, \quad \text{for } v(k) = 0,\;\; k \notin [0, M-1]$$

This gives

$$w(k) = \sum_{n=0}^{M-1} v(n)v(n-k) = v(k) * v(k) \quad \Leftrightarrow \quad W(\omega) \geq 0 \quad \textbf{pos. semidefinit.}$$

A window design should trade-off between smearing and leakage

For instance: weak sinewave + strong narrowband interference $\rightarrow$ leakage more detrimental than smearing

Homework: can we use optimisation to balance between smearing and leakage

Several frequently used "cosine–type windows"

Idea: suppress sidelobes, perhaps sacrifice the width of mainlobe

- **Hann** window

```
w = 0.5 * (1 - cos(2*pi*(0:m-1)'/(n-1)));
```

- **Hamming** window

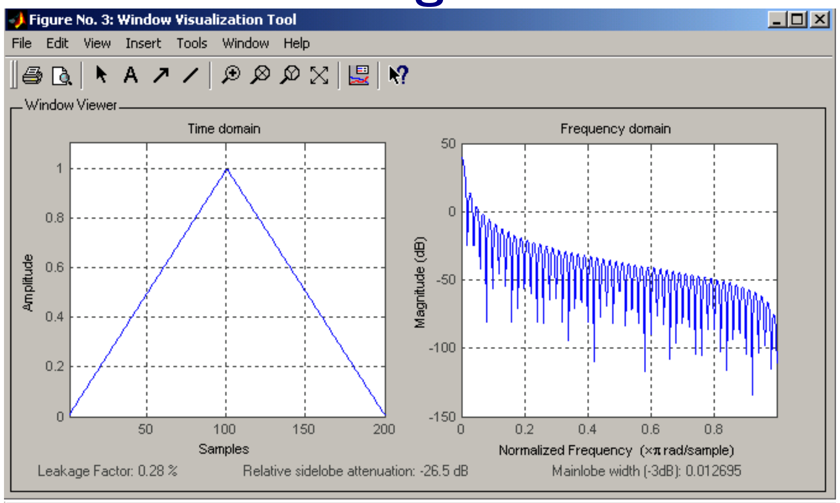
```
w = (54 - 46*cos(2*pi*(0:m-1)'/(n-1)))/100;
```

- **Blackman** window

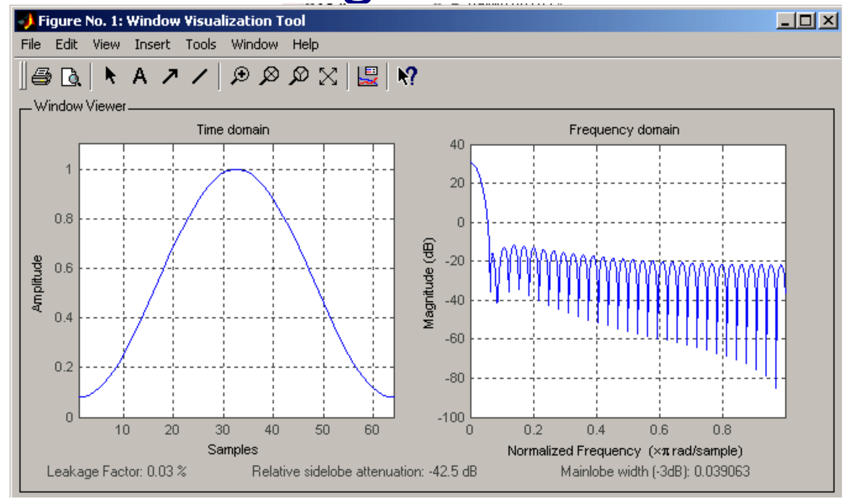
```
w = (42 - 50*cos(2*pi*(0:m-1)/(n-1)) +
                   + 8*cos(4*pi*(0:m-1)/(n-1)))'/100;
```

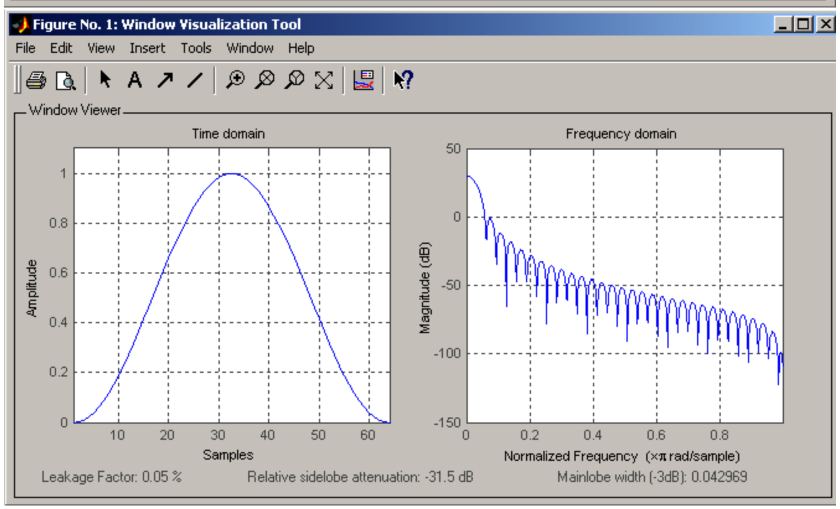
Example: Properties of standard window functions

Triangular window

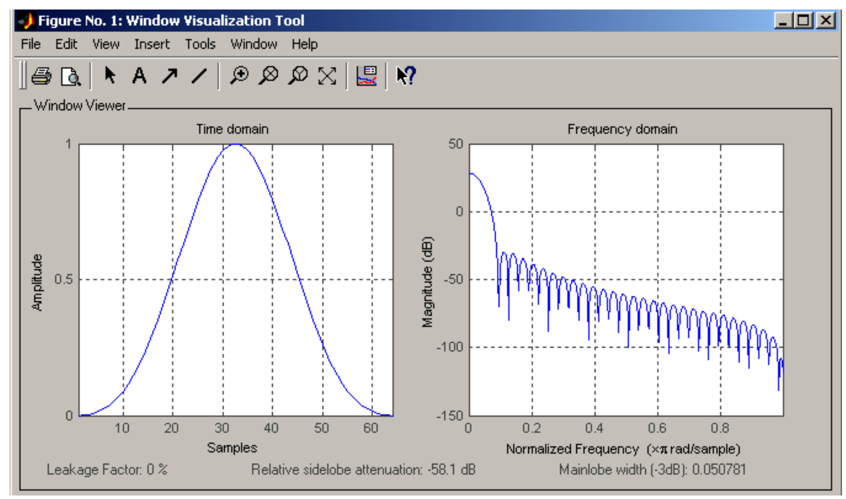


Hamming window





Hann window



Blackman window

Performance of the modified periodogram

- **Bias:** Since

$$U = \frac{1}{N}\sum_{n=0}^{N-1} |w[n]|^2 = \frac{1}{N}\int_{-\pi}^{\pi} |W(e^{\jmath\omega})|^2 \, d\omega \quad \Rightarrow \quad \frac{1}{2\pi NU}\int_{-\pi}^{\pi} |W(e^{\jmath\omega})|^2 \, d\omega = 1$$

  for $N \to \infty$ the modified periodogram is asymptotically unbiased.

- **Variance:** Since $\hat{P}_M$ is simply $\hat{P}_{per}$ of a windowed data sequence

$$Var\left\{\hat{P}_M(\omega)\right\} \approx P_{xx}^2(\omega)$$

  $\Rightarrow$ **not a consistent estimate** of the power spectrum, and the data window offers no benefit in terms of reducing the variance

- **Resolution:** Data window provides a trade–off between spectral resolution (main lobe width) and spectral masking (sidelobe amplitude).

Periodogram modifications: Effects of different windows

Properties of several commonly used windows with length N:

- **Rectangular** – Sidelobe level = -13 [dB], $3\ dB$ BW $\rightarrow 0.89(2\pi/N)$
- **Bartlett** – Sidelobe level = -27 [dB], $3\ dB$ BW $\rightarrow 1.28(2\pi/N)$
- **Hanning** – Sidelobe level = -32 [dB], $3\ dB$ BW $\rightarrow 1.44(2\pi/N)$
- **Hamming** – Sidelobe level = -43 [dB], $3\ dB$ BW $\rightarrow 1.30(2\pi/N)$
- **Blackman** – Sidelobe level = -58 [dB], $3\ dB$ BW $\rightarrow 1.68(2\pi/N)$

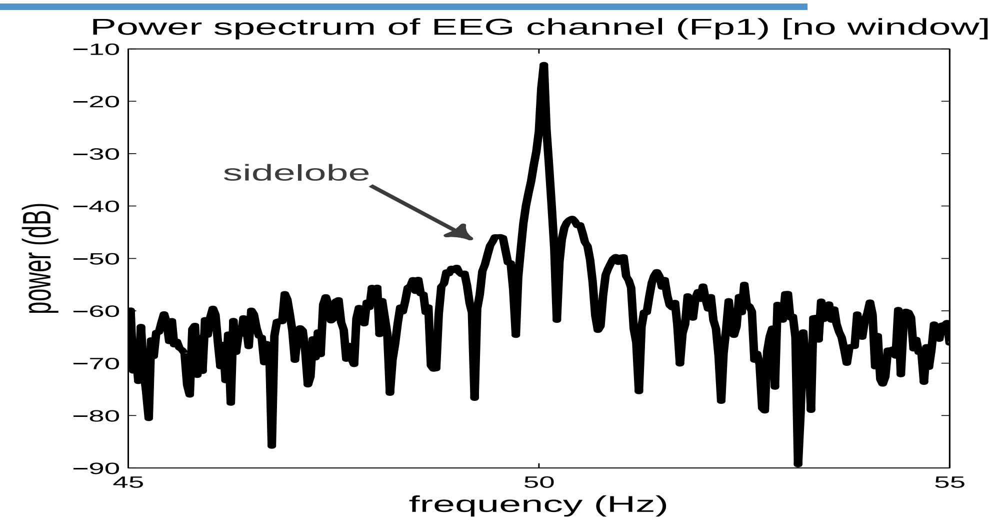
Notice the relationship between the sidelobe level and bandwidth!

Example: FFT leakage $\looparrowright$ EEG power spectrum

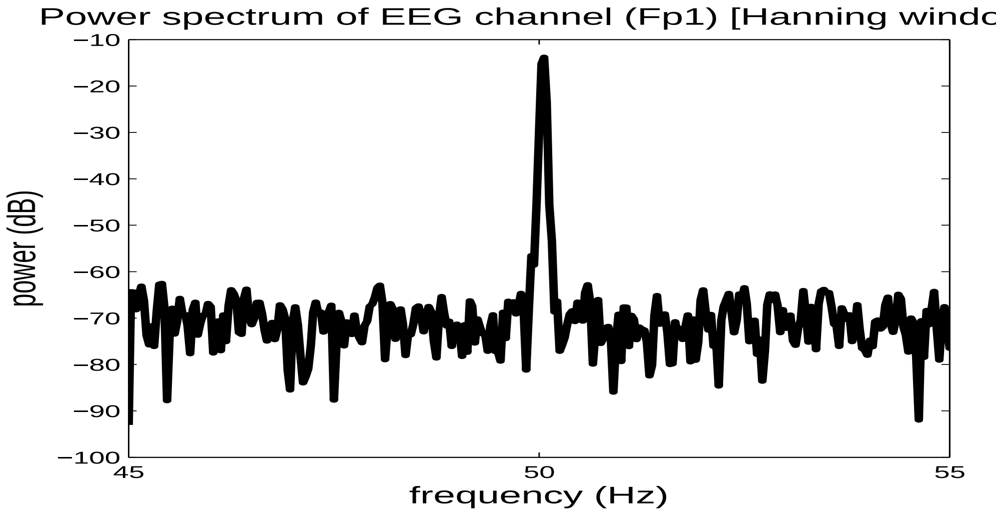
we record $\approx 10\mu V$ signals in the presence of external noise

Problem: estimate power of the 50Hz artefact picked up by EEG leads

- Using the standard periodogram - the resolution is good but the artefact is partially masked
- **Remedy:** Use a windowing function (e.g. Hanning window).
  - Note that sidelobes are reduced, energy over narrow frequency range around 50Hz.
- Window value is zero at the beginning and end of a segment
  - Multiply with the signal with a window that has small sidelobes to reduce leakage
- **Windows reduce, but do not eliminate leakage completely!**



`periodogram(x,[],N,Fs,'onesided');`



`periodogram(x,hann(length(x)),N,Fs,'onesided');`

Welch's method: Averaging modified periodograms

In 1967, Welch proposed two modifications to Bartlett's method:

- allow the sequences $x_i[n]$ to overlap
- to allow data window $w[n]$ to be applied to each sequence $\Rightarrow$ averaging modified periodograms

This way, successive segments are offset by D points and each segment is L points long

$$x_i[n] = x[n + iD] \qquad n = 0, 1, \ldots, L - 1$$

The amount of overlap between $x_i[n]$ and $x_{i+1}[n]$ is $L - D$ points and

$$N = L + D(K - 1)$$

N- total number of points, L- length of segments, D- amount of overlap, K- number of sequences

Variations on the theme

We may vary between **no overlap D=L** and say 50 % overlap **D = L/2** or anything else.

☺ we can trade a reduction in the variance for a reduction in the resolution, since

$$\hat{P}_W(\omega) = \frac{1}{KLU} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} w[n]x[n+iD]e^{-\jmath n\omega} \right|^2$$

or in terms of modified periodograms

$$\hat{P}_W(\omega) = \frac{1}{K} \sum_{i=0}^{K-1} \hat{P}_M^{(i)}(\omega)$$

$\rightsquigarrow$ **asymptotically unbiased** (follows from the bias of the modified periodogram)

Welch vs. Bartlett

- the amount of overlap between $x_i[n]$ and $x_{i+1}[n]$ is $L - D$ points, and if K sequences cover the entire N data points, then

$$N = L + D(K + 1)$$

- If there is no overlap, $(D = L)$ we have $K = \frac{N}{L}$ sections of length L as in Bartlett's method
- Of the sequences are overlapping by 50 % $D = \frac{L}{2}$ then we may form $K = 2\frac{N}{L} - 1$ sections of length L. thus maintaining the same resolution as Bartlett's method while doubling the number of modified periodograms that are averaged, thereby reducing the variance.
- With 50% overlap we could also form $K = \frac{N}{L} - 1$ sequences of length $2L$, thus increasing the resolution while maintaining the same variance as Bartlett's method.

Therefore, by allowing sequences to overlap, it is possible to increase the number and/or length of the sequences that are averaged, thereby trading a reduction in variance for a reduction in resolution.

Properties of Welch's method

- **Functional relationship:**

$$\hat{P}_W(\omega) = \frac{1}{KLU} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} w[n] x[n+iD] e^{-\jmath n\omega} \right|^2 \qquad U = \frac{1}{L} \sum_{n=0}^{L-1} |w[n]|^2$$

- **Bias**

$$E\left\{ \hat{P}_W(\omega) \right\} = \frac{1}{2\pi LU} P_x(\omega) \; * \; |W(\omega)|^2$$

- **Resolution** $\looparrowright$ window dependent

- **Variance** (assuming 50 % overlap and Bartlett window)

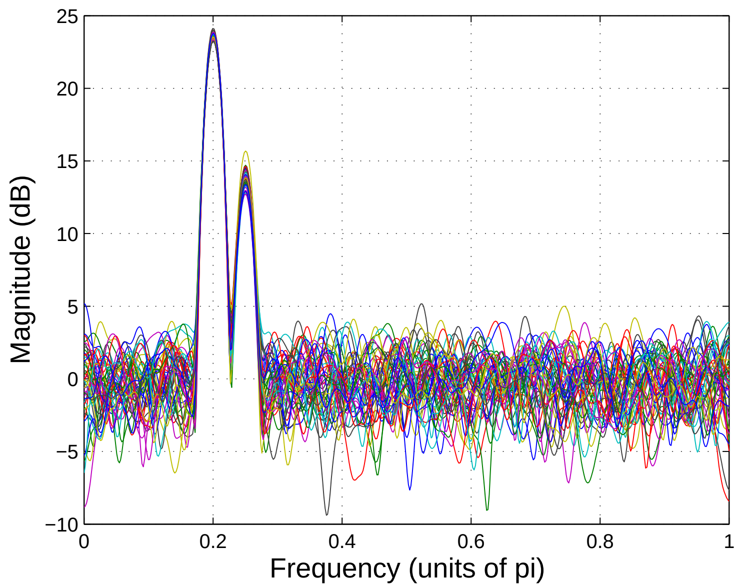
$$Var\left\{ \hat{P}_W(\omega) \right\} \approx \frac{9}{16} \frac{L}{N} P_x^2(\omega)$$

Example: Two sinusoids in noise $\looparrowright$ Welch estimates

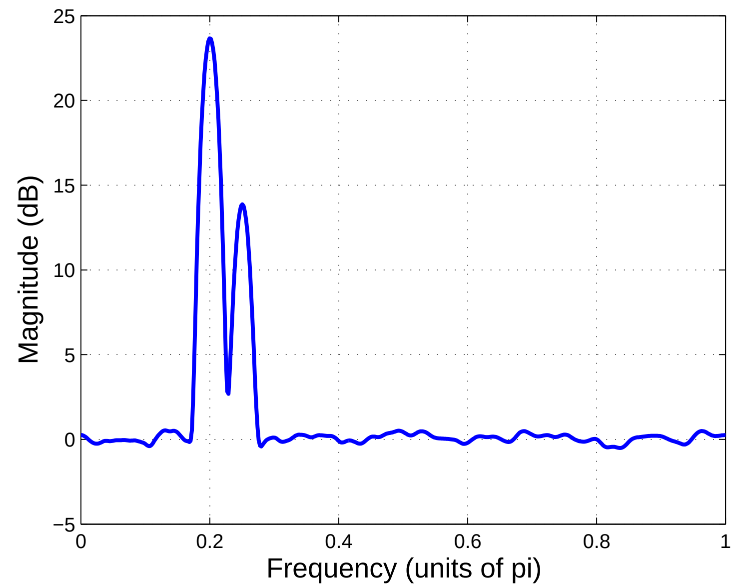
Problem: Estimate the spectra of the following two sinewaves using Welch's method

$$x[n] = \sqrt{10}\sin(n * 0.2\pi + \Phi_1) + \sin(n * 0.3\pi + \Phi_2) + w[n]$$

Unit noise variance, $N = 512$, $L = 128$, 50 % overlap (7 sections)



Overlay of 50 estimates

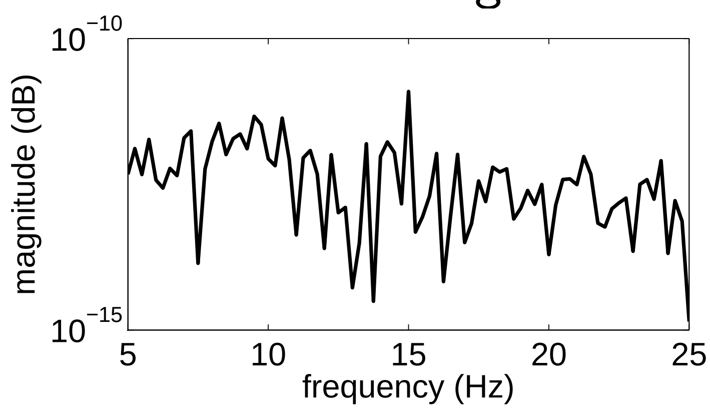


Periodogram using Welch's method

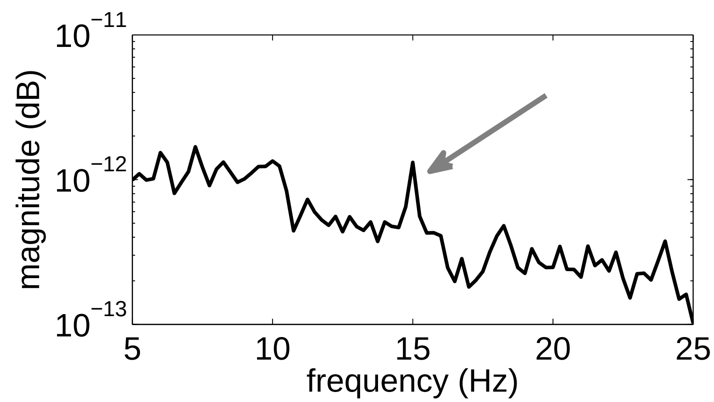
EEG feature estimation ssvep movie.wmv

Subject presented with flashing visual stimulus (15 Hz) which causes a response in EEG at same frequency

- Applying the periodogram to data gives a noisy estimate of PSD
  - **stimulus response not clearly visible**
- The averaged periodogram reduces the level of noise
  - **stimulus response at 15 Hz !**
  - The total signal length was $N = 48000$, the averaged periodogram used a window length of $L = 3600$, and (with overlap) the total number of averaged windows was $K = 102$



Periodogram.



Averaged Periodogram.

More on EEG spectral estimation: EEG propagation

only for illustration, not examinable

EEG recording principle

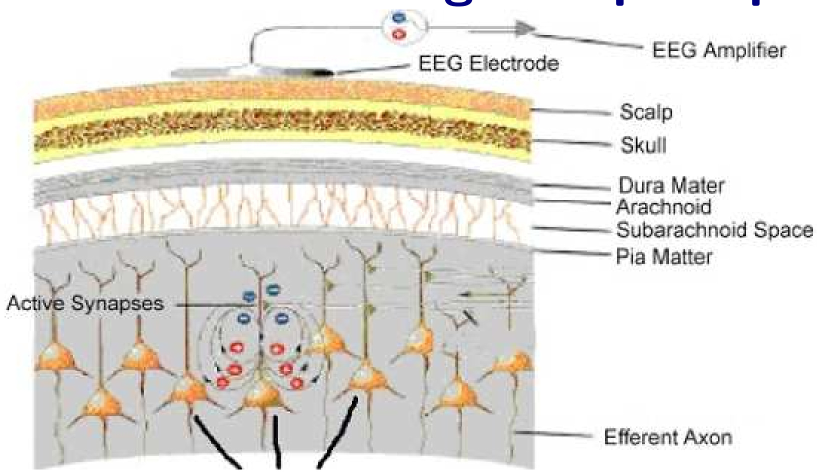


Figure 1: [**Left**] The left Ear-EEG earplug with electrode positions visible (grey dots) and an arrow indicating the direction in which it enters the ear canal. [**Right**] Recording setup: joint recording of Ear-EEG and on-scalp EEG for comparative analysis.

Principle of in-the-ear EEG

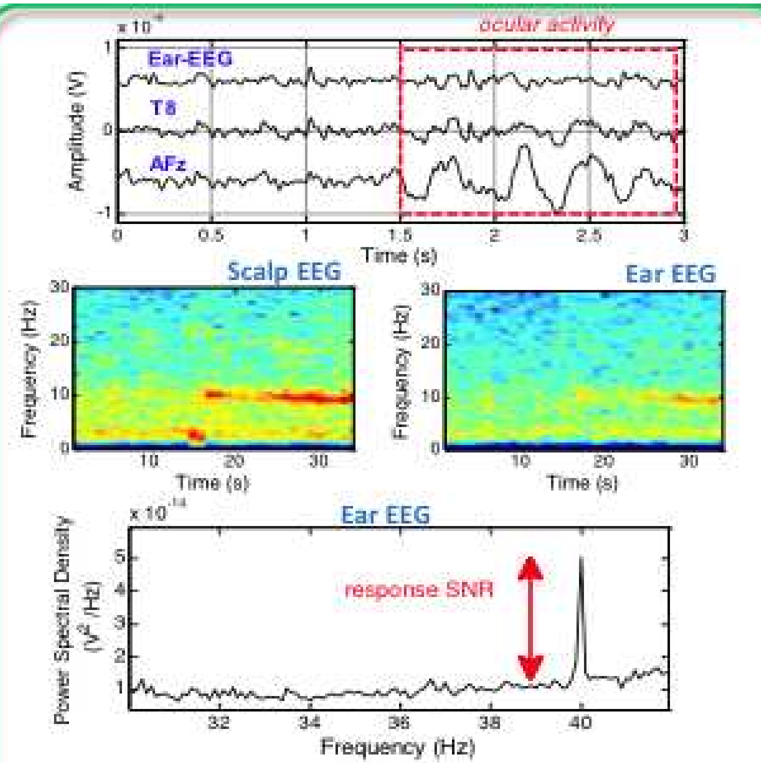


Figure 2: [**Upper**] Time waveforms for scalp and Ear-EEG over 3s with consecutive eye blinking starting at 1.5s, Ear-EEG exhibits a suppression of ocular artifacts. [**Centre**] Time-frequency plots as subject closes eyes from 15-35s, with increased activity visible for scalp [**Centre, left**] and Ear-EEG [**Centre, right**] in the alpha range (8-12 Hz). [**Lower**] The auditory steady state response for Ear EEG (40 Hz stimulus). The SNR (ratio of the response peak to background EEG) matches that of temporal scalp electrodes [4].

Ear-EEG vs on-scalp EEG

More on EEG spectral estimation: Some applications

only for illustration, not examinable

EEG brain computer interface

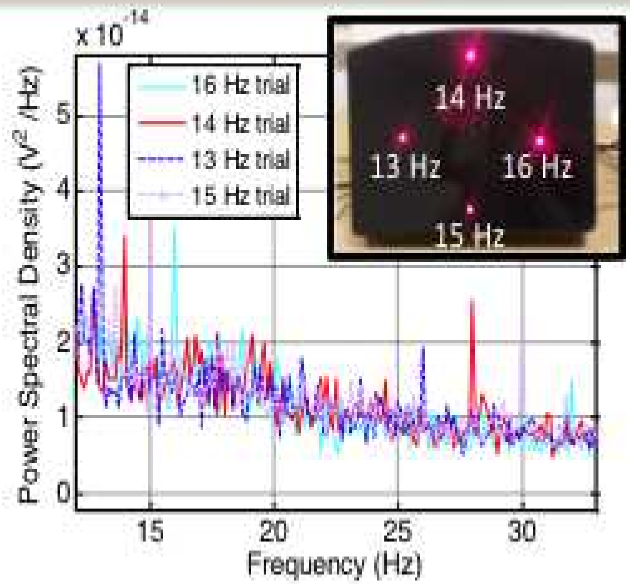
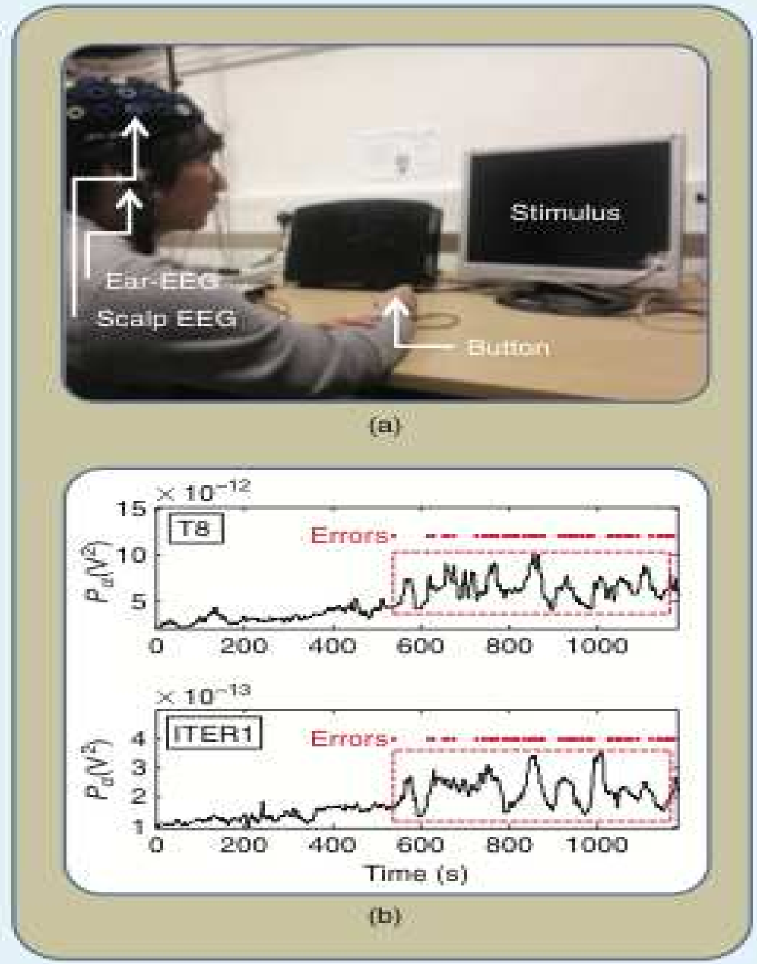


Figure 3: The average Ear-EEG power spectral density (PSD) for four 120s trials as the user attended 13, 14, 15 and 16 Hz stimuli. [**Insert: upper right**] The SSVEP stimulus interface with four LEDs.

The subject focuses on flashing diodes of different frequencies (13, 14, 15, 16 Hz), and this is reflected in EEG. Our task is to perform good spectrum estimation.

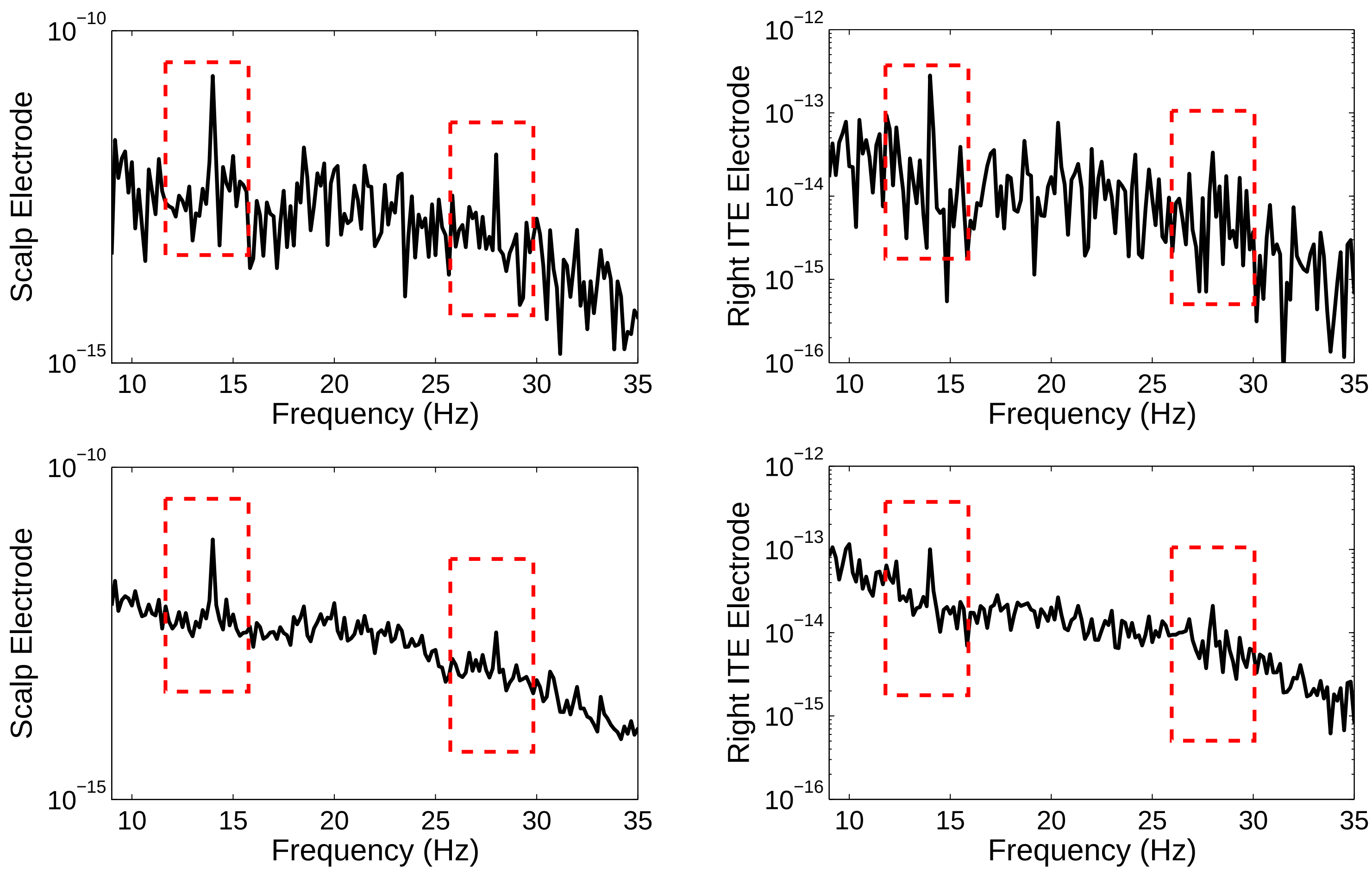


An interactive vigilance test

SSVEP - we look for a 14 Hz stimulus in a 50s recording using Welch's method

Standard: A 50s EEG from scalp (Oz) and right ear (ITE). Averaged: 27 segments of 12s.

Top: no window **Bottom: Hann window**



Blackman–Tukey method: Periodogram smoothing

Recall that the methods by Bartlett and Welch are designed to reduce the variance of the periodogram by averaging periodograms and modified periodograms, respectively.

Another possibility is "periodogram smoothing" often called the Blackman–Tukey method.

Let us identify the problem ☹

$$\hat{r}_x[N-1] = \frac{1}{N} x[N-1]x[0]$$

$\Rightarrow$ there is little averaging when calculating the estimates of $\hat{r}_x[k]$ for $|k| \approx N$.

These estimates will be **unreliable** no matter how large N. We have two choices:

- reduce the variance of those unreliable estimates
- reduce the contribution these unreliable estimates make to the periodogram

Blackman–Tukey Method: Resolution vs. Variance

The variance of the periodogram is decreased by reducing the variance of the ACF estimate by calculating more robust ACF estimates over fewer data points ($M < N$).

$\Rightarrow$ Apply a window to $\hat{r}_x[k]$ to decrease the contribution of unreliable estimates and obtain the Blackman–Tukey estimate:

$$\hat{P}_{BT}(\omega) = \sum_{k=-M}^{M} \hat{r}_x[k] w[k] e^{-\jmath k\omega}$$

where $w[k]$ is a **lag window** applied to the ACF estimate.

$$\hat{P}_{BT}(\omega) = \frac{1}{2\pi}\hat{P}_{per}(\omega) \ * \ W(\omega) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{P}_{per}(e^{\jmath u}) W(e^{\jmath(\omega-u)}) du$$

that is, **we trade the reduction in the variance for a reduction in the resolution** (smaller number of ACF estimates used to calculate the PSD)

Properties of the Blackman–Tukey method

- **Functional relationship:**

$$\hat{P}_{BT}(\omega) = \sum_{k=-M}^{M} \hat{r}_x[k]w[k]e^{-\jmath k\omega}$$

- **Bias**

$$E\left\{\hat{P}_{BT}(\omega)\right\} \approx \frac{1}{2\pi}P_x(\omega) \ * \ W(\omega)$$

- **Resolution**– window dependent (window – conjugate symmetric and with non–negative FT)
- **Variance:** Generally, it is recommended $M < N/5$.

$$Var\left\{\hat{P}_{BT}(\omega)\right\} \approx P_x^2(\omega)\frac{1}{N}\sum_{k=-M}^{M} w^2[k]$$

Trade–off: for a small bias M needs to be large to minimize the width of the mainlobe of $W(\omega)$, whereas M should be small in order to minimize the variance.

Non-negative definiteness of the BT spectrum estimator

see also Problem 4.9 in your Problem/Answer set

The main problem with periodogram is its high statistical variability. This arises from:

- Poor accuracy of the autocorrelation estimate for large lags m
- Accumulating of these errors in the spectrum estimate

These effects can be mitigated by taking fewer points (M instead of N) in ACF estimation.

Observe that the Blackman–Tukey spectral estimator corresponds to a locally weighted average of the periodogram.

Roughly speaking:

⊛ the resolution of the BT estimator is $\sim 1/M$

⊛ the variance of the BT estimator is $\sim M/N$

Performance comparison of periodogram–based methods

Let us introduce criteria for performance comparison:

- **Variability of the estimate**

$$\nu = \frac{var\left\{\hat{P}_x(\omega)\right\}}{E^2\left\{\hat{P}_x(\omega)\right\}}$$

which is effectively **normalised variance**

- **Figure of merit**

$$\mathcal{M} = \nu \times \Delta\omega$$

that is, **product of variability and resolution.**

$\mathcal{M}$ **should be as small as possible.**

Performance measures for the Nonparametric methods of Spectrum Estimation

Method	Variability ν	Resolution $\Delta\omega$	Figure of merit $\mathcal{M}$
Periodogram	1	$0.89\frac{2\pi}{N}$	$0.89\frac{2\pi}{N}$
Bartlett	$\frac{1}{K}$	$0.89K\frac{2\pi}{N}$	$0.89\frac{2\pi}{N}$
Welch	$\frac{9}{8}\frac{1}{K}$	$1.28\frac{2\pi}{L}$	$0.72\frac{2\pi}{N}$
Blackman–Tukey	$\frac{2}{3}\frac{M}{N}$	$0.64\frac{2\pi}{M}$	$0.43\frac{2\pi}{N}$

- Observe that each method has a Figure of Merit which is approximately the same
- Figure of merit are inversely proportional to N
- Although each method differs in its resolution and variance, **the overall performance is fundamentally limited by the amount of data that is available.**

© D. P. Mandic

Conclusions

FFT based spectral estimation is limited by:

- correlation assumed to be zero beyond N - biased/unbiased estimates
- resolution limited by the DFT "baggage"
- if two frequencies are separated by Δf, then we need $N \geq \frac{1}{\Delta f}$ data points to separate them
- limitations for spectra with narrow peaks (resonances, speech, sonar)
- limit on the resolution imposed by N also causes bias
- variance of the periodogram is almost independent of data length
- the derived variance formulae are only illustrative for real–world signals

But also many opportunities: spectral coherency, spectral entropy, TF, ...

Next time: model based spectral estimation for discrete spectral lines

Opportunities: Spectral Coherence and LS Periodogram

see also Problem 4.7 in your P/A sets

The **spectral coherence** shows similarity between two spectra

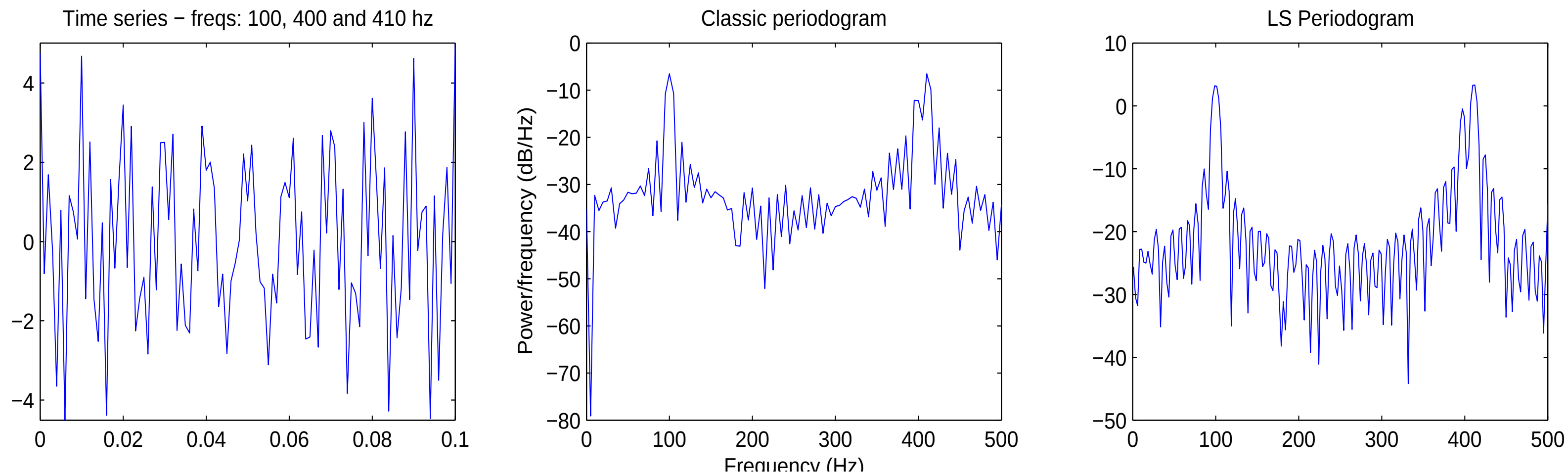
$$C_{xy}(\omega) = \frac{P_{xy}(\omega)}{\left[P_{xx}(\omega)P_{yy}(\omega)\right]^{1/2}}$$

It is invariant to linear filtering of x and y (even with different filters)

The periodogram $P_{per}(\omega)$ can be seen as a **Least Squares** solution to

$$P_{per}(\omega) = \|\hat{\beta}(\omega)\|^2, \;\; \hat{\beta} = \underset{\beta(\omega)}{\operatorname{argmin}} \sum_{n=1}^{N} \|y(n) - \beta e^{j\omega n}\|^2,$$

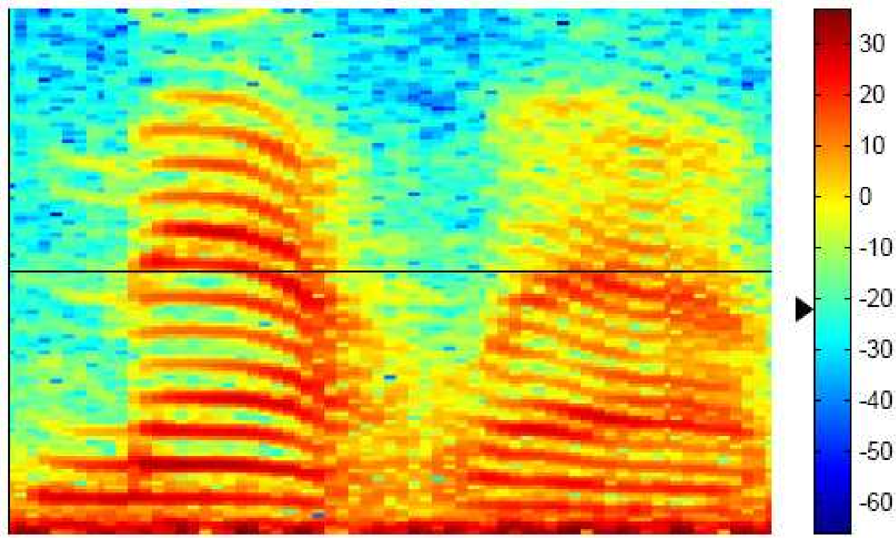
Periodogram and LS periodog. for a sinewave mixture (100, 400, 410) Hz



Opportunities: Time-Frequency estimation

time–frequency spectrogram of “Matlab” $\looparrowright$ ‘specgramdemo‘

Frequency



time

M aaa t l aaa b

For every time instant “t”, the PSD is plotted along the vertical axis

Darker areas: higher magnitude of PSD

Time-Frequency (TF) analysis: Principles

Assume $x(n)$ has a Fourier transform $X(\omega)$ and power spectrum $|X(\omega)|^2$.

The function $TF(n,\omega)$ determines how the energy is distributed in time-frequency, and it satisfies the following **marginal properties:**

$$\sum_{n=-\infty}^{\infty} TF(n,\omega) \quad = \quad |X(\omega)|^2 \quad \text{energy in the signal at frequency } \omega$$

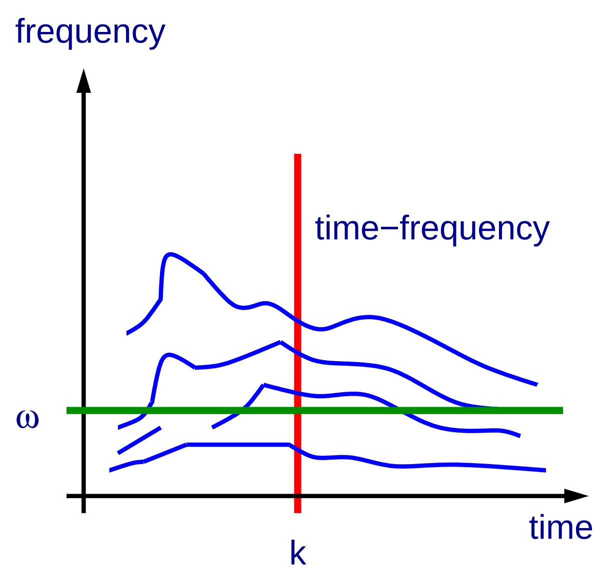
$$\frac{1}{2\pi}\int_{-\pi}^{\pi} TF(n,\omega)d\omega \quad = \quad |x(n)|^2 \quad \text{energy at time instant `k}' \text{ due to all } \omega$$

Then

$$\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\int_{\infty}^{\infty} TF(n,\omega)d\omega = \sum_{n=-\infty}^{\infty} |x(n)|^2$$

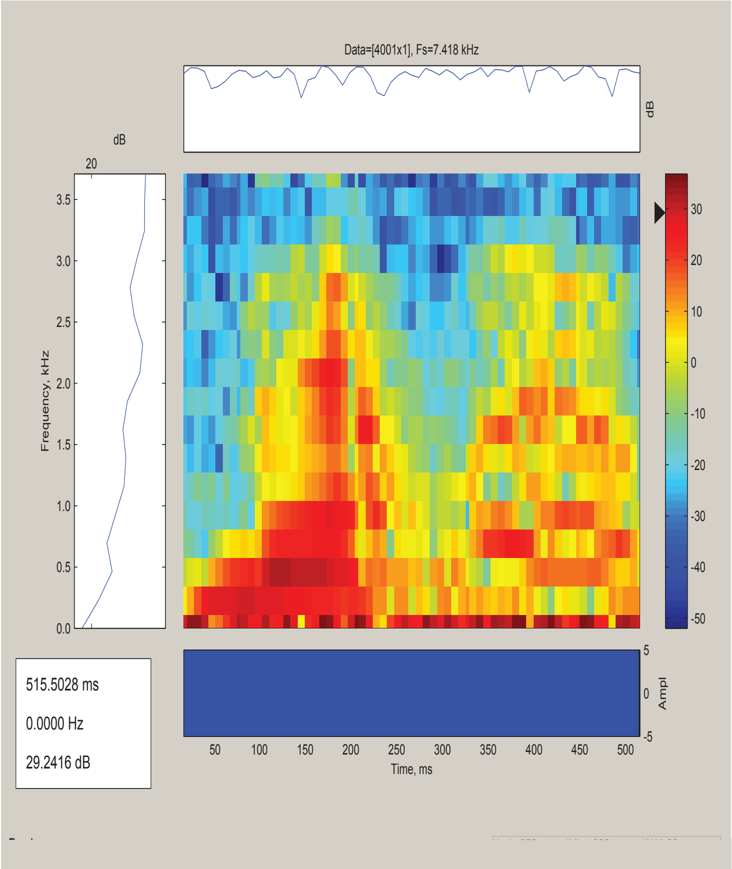
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} |X(\omega)|^2 d\omega$$

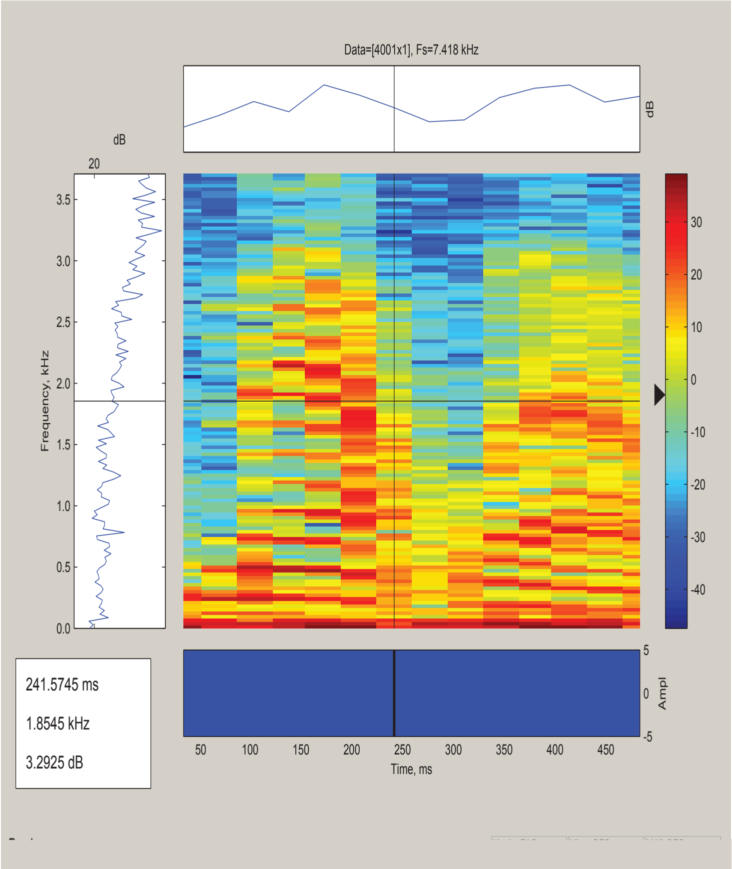
giving the **total energy** (all frequencies and samples) of a signal.



Time–frequency spectrogram of a speech signal

(wide band spectrogram) **(narrow band spectrogram)**





(win-len=256, overlap=200, ftt-len=32) **(win-len=512, overlap=200, ftt-len=256)**

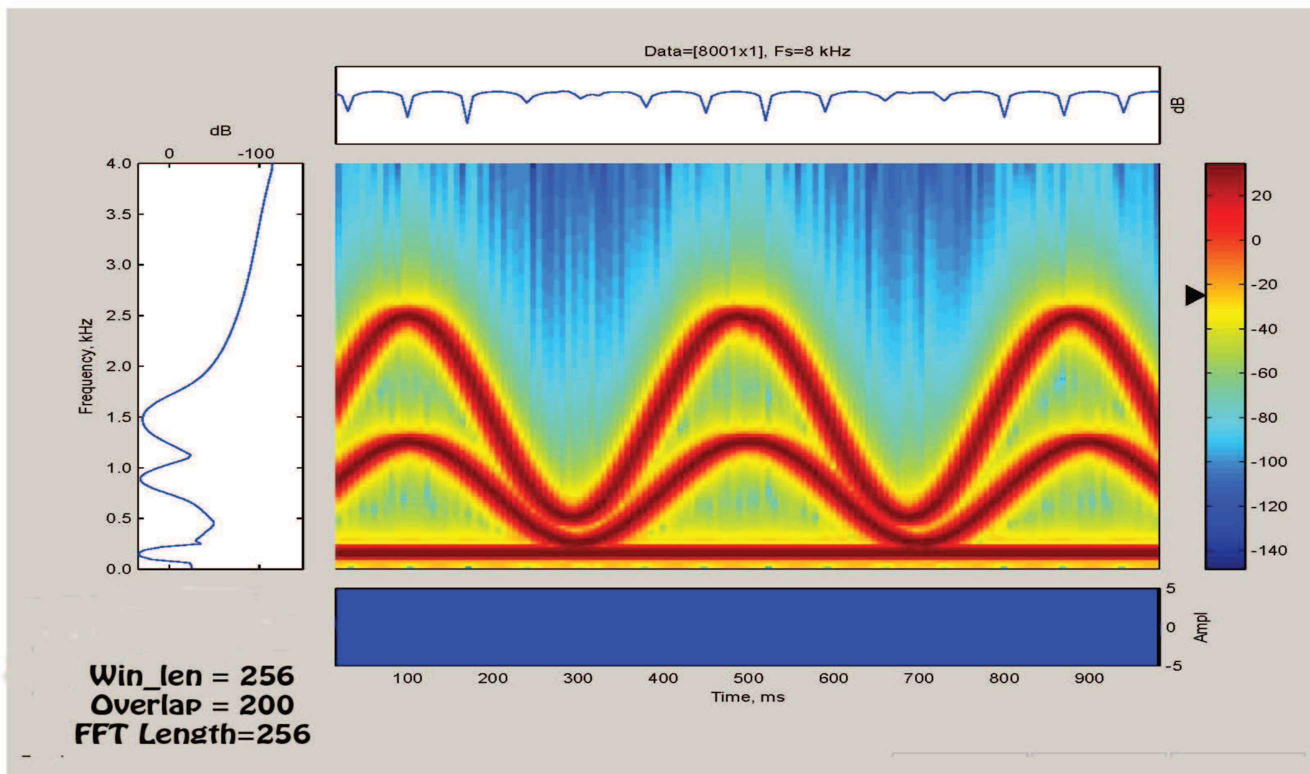
Homework: **evaluate all the methods from the lecture for this T-F spectrogram**

TF spectrogram of a frequency-modulated signal (check also your coursework)

The time-frequency spectrogram of a frequency modulated (FM) signal

$$y(t) = A\cos\left[\omega_0 t + k_f \int_{-\infty}^{t} x(\alpha)d\alpha\right]$$

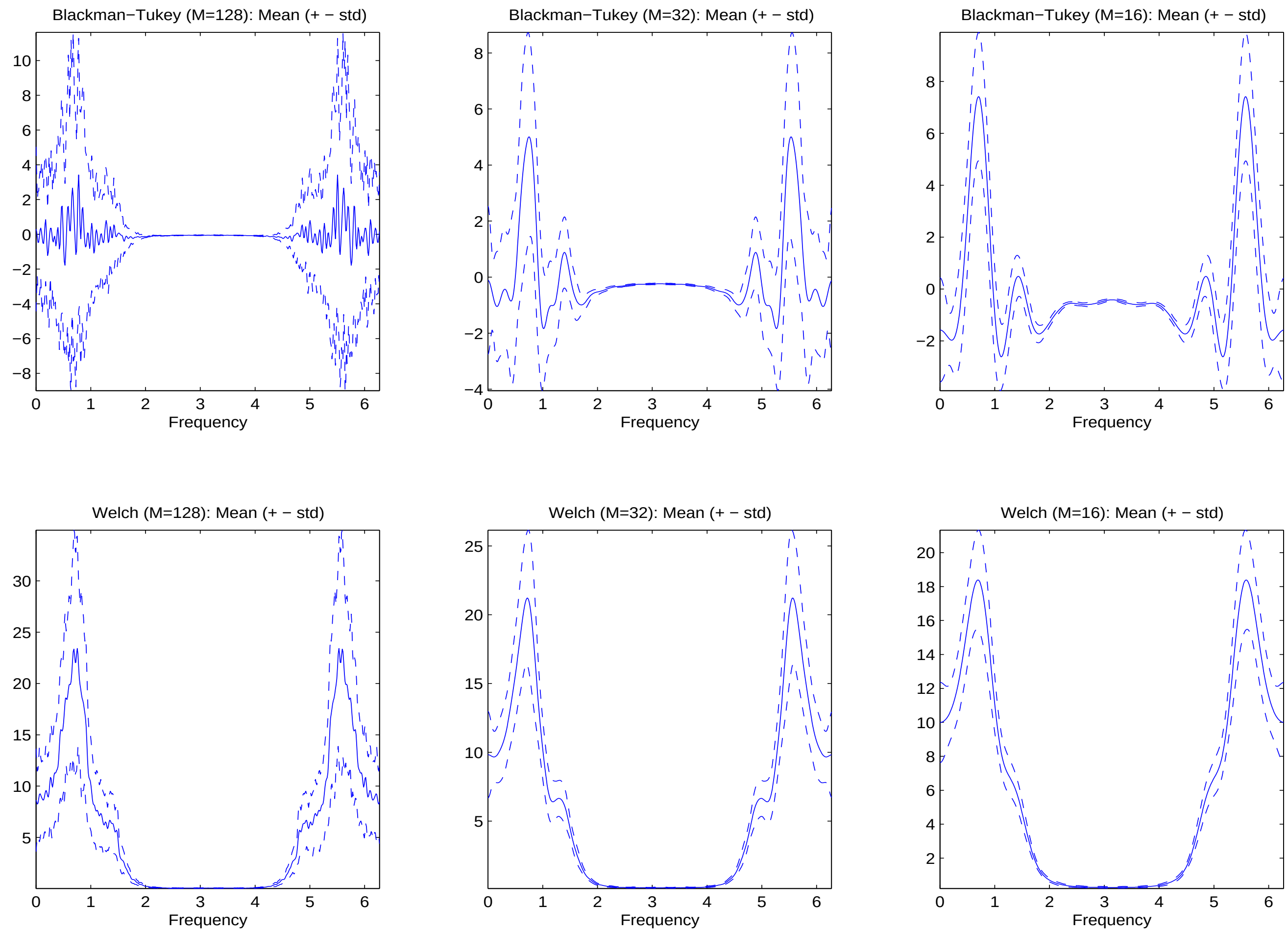
frequency



time

Opportunities: ARMA spectrum

N=512 samples, freq. res=1/500



Signal: ARMA(4,4), b=[1, 0.3544, 0.3508, 0.1736, 0.2401] a=[1, -1.3817, 1.5632, -0.8843, 0.4096]

Sometimes we only desire the correct position of the peaks $\looparrowright$ **ARMA Spectrum Estimation**

A note on positive-semidefiniteness of the R_{xx}

The autocorrelation matrix $\mathbf{R}_{xx} = E\big[\mathbf{x}\mathbf{x}^T\big]$

where $\mathbf{x} = \big[x[0], \ldots, x[N-1]\big]^T$. It is symmetric and of size $N \times N$.

There are four ways to define positive semidefiniteness: (see also your Problem-Answer sets)

1. All the eigenvalues of the autocorrelation matrix $\mathbf{R}$ are such that $\lambda_i \geq 0$, for i=1,. . . ,N
2. For any nonzero vector $\mathbf{a} \in \mathbb{R}^{N\times 1}$ we have $\mathbf{a}^T\mathbf{R}\mathbf{a} \geq 0$. For complex valued matrices, the condition becomes $\mathbf{a}^H\mathbf{R}\mathbf{a}$
3. There exists a matrix $\mathbf{U}$ such that $\mathbf{R} = \mathbf{U}\mathbf{U}^T$, where the matrix $\mathbf{U}$ is called a root of $\mathbf{R}$
4. All the principal submatrices of $\mathbf{R}$ are positive semidefinite. A principal submatrix is formed by removing $i = 1, \ldots, N$ rows and columns of $\mathbf{R}$

For positive definiteness conditions, replace $\geq$ with $>$

Opportunities: Spectral Entropy

Spectral entropy can be used to measure the peakiness of the spectrum.

This is achieved via the probability mass function (PMF) (normalised PSD) given by

$$\eta[i] = \frac{P_{per}[i]}{\sum_{l=0}^{N-1} P_{per}[l]} \quad \rightarrow \quad H_{sp} = -\sum_{i=0}^{N-1} \eta[i] \log_2 \eta[i] = \sum_{i=0}^{N-1} \eta[i] \log_2 \frac{1}{\eta[i]}$$

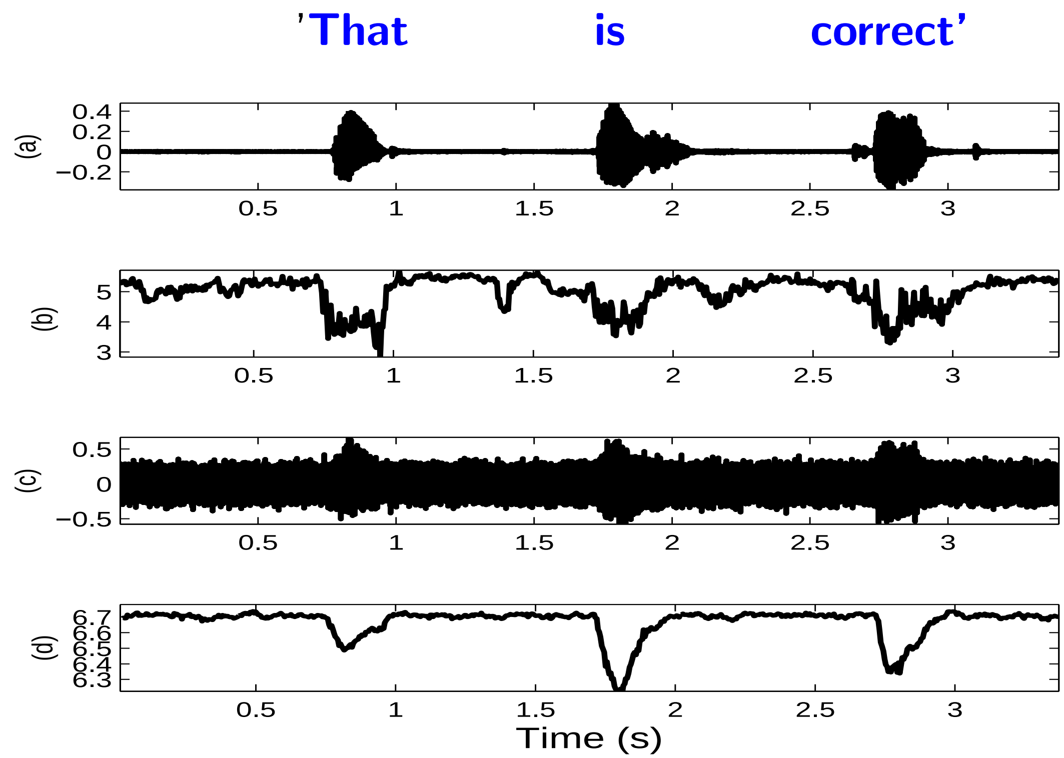
Intuition:

- peaky spectrum (e.g. $\sin(x)$) $\looparrowright$ low spectral entropy

- flat spectrum (e.g. WGN) $\looparrowright$ high spectral entropy

Figure on the right:

From top to bottom: a) clean speech, b) spectral entropy, c) speech + noise, d)spectral entropy of (speech+noise)



Notes

Notes

Notes