# Non-parametric spectral estimation

Roland Badeau,
roland.badeau@telecom-paris.fr

Linear time series (part 2)
TSIA202b

## Program of the course unit

- Lesson + tutorial "Non-parametric spectral estimation"
- Lesson + tutorial "Parametric estimation of rational spectra"
- Lesson + tutorial "Filter bank methods"
- Practical work "Spectral estimation"
- Lesson + tutorial "Parametric estimation of line spectra"
- Practical work "High resolution methods"
- Lesson "Kalman filtering"
- Revision tutorial
- Examination



## Applications

- Spectral analysis is a universal signal processing tool that is widely used in various fields (audio processing, biomedical data analysis (e.g. MRI), seismic signal processing, mechanics (e.g. for modal analysis), astronomy, etc.).
- It is the prerequisite of many applications, e.g. audio analysis/synthesis, audio coding, denoising, source separation, pitch estimation, speech recognition... and even detecting which atoms compose a celestial body far from earth.
- In addition, high resolution methods are used not only for spectral analysis, but also for solving source localization problems (DOA estimation for "direction of arrival") and in digital communications.



## Educational resources

- Textbook "Spectral analysis of signals", Stoica and Moses, 2005
- HR methods : course handout "High resolution methods"
- Kalman filter : course handout "Linear time series" (TSIA202a)
- Website `https://ecampus.paris-saclay.fr/` (English/French)
  - Links to textbook + course handouts + slides
  - Subjects of tutorials and practical works + data
  - Solutions available online after every tutorial
  - Online submission of the practical works reports
- Grading (20 points)
  - Practical works reports (3 points for each practical work)
  - Examination (14 points)

# Part I

# Reminder: WSS processes



## Wide Sense Stationary (WSS) processes

- Definition : sequence of random variables $X_t$ such that
  - $\mathbb{E}[|X_t|^2] < +\infty$
  - $\mathbb{E}[X_t] = \mu_X$ does not depend on $t$,
  - $\forall k \in \mathbb{Z}$, $\mathrm{cov}[X_{t+k}, X_t] = \mathbb{E}[X^c_{t+k}\overline{X^c_t}]$ (where $X^c_t = X_t - \mu_X$) does not depend on $t$
- Properties
  - If $\mathbb{E}[|X_t|^2] < +\infty$, strict stationarity $\Rightarrow$ WSS
  - If $X_t$ is Gaussian, strict stationarity $\Leftrightarrow$ WSS



## Autocovariance function

- Definition : $r_{XX}(k) = \mathrm{cov}[X_{t+k}, X_t] = \mathbb{E}[X^c_{t+k}\overline{X^c_t}]$
- Properties :
  - $\mathrm{Var}[X] = r_{XX}(0) \geq 0$
  - Hermitian symmetry : $r_{XX}(-k) = \overline{r_{XX}(k)}$
  - Positive semi-definiteness : $\forall k$, $\forall t_1 \dots t_k$, $\forall \lambda_1 \dots \lambda_k \in \mathbb{C}$,
  $$\sum_{i=1}^{k}\sum_{j=1}^{k} \lambda_i \overline{\lambda_j} r_{XX}(t_i - t_j) \geq 0$$
  - Boundedness : $|r_{XX}(k)| \leq r_{XX}(0)$
    - Proof : $|\mathbb{E}[X^c_{t+k}\overline{X^c_t}]|^2 \leq \mathbb{E}[|X^c_{t+k}|^2]\mathbb{E}[|X^c_t|^2]$ (Schwarz inequality)
  - Remark : power of a WSS process
  $$P_X = \mathbb{E}[|X_t|^2] = r_{XX}(0) + |\mu_X|^2$$



## Power spectral density (PSD)

- If $r_{XX}(k) \in l^1(\mathbb{Z})$, let $\forall \nu \in \mathbb{R}$, $S_{XX}(\nu) = \sum_{k=-\infty}^{+\infty} r_{XX}(k) e^{-2i\pi\nu k}$
- Inversion : $r_{XX}(k) = \int_{-1/2}^{+1/2} S_{XX}(\nu) e^{+2i\pi\nu k} d\nu$
- Continuity : $\nu \mapsto S_{XX}(\nu)$ is a continuous function
- Herglotz theorem : $S_{XX}(\nu) \geq 0 \ \forall \nu \in \mathbb{R}$
- Power of a WSS process :

$$P_X = r_{XX}(0) + |\mu_X|^2 = \int_{-1/2}^{+1/2} S_{XX}(\nu) d\nu + |\mu_X|^2$$

# Filtering theorem for WSS processes

- Let $h_k$ be the impulse response of a stable filter of frequency response $H(\nu)$, $X$ be a WSS process, and $Y = h * X$.
- Then $Y$ is a WSS process :
  - Mean :
  $$\mu_Y = \mu_X H(0) = \mu_X \sum_{k \in \mathbb{Z}} h_k$$
  - Autocovariance function :
  $$r_{YY} = h * \widetilde{h} * r_{XX}$$
  where $\widetilde{h}_t = \overline{h_{-t}}$
  - If in addition $r_{XX} \in l^1(\mathbb{Z})$, the power spectral density is defined as :
  $$S_{YY}(\nu) = |H(\nu)|^2 S_{XX}(\nu)$$



Part II

Reminder: estimation of the mean and of the autocovariance function



# Parametric estimation

- Let $X$ be a random variable parameterized by $\theta$.
- An estimator $\widehat{\theta}$ of $\theta$ is a function of $X$
- Bias : $b(\theta, \widehat{\theta}) = \mathbb{E}_\theta[\widehat{\theta}(X) - \theta]$
- Mean Square Error (MSE) :
  $R(\theta, \widehat{\theta}) = \mathbb{E}_\theta[|\widehat{\theta}(X) - \theta|^2] = \mathrm{Var}[\widehat{\theta}(X)] + |b(\theta, \widehat{\theta})|^2$
  - Existence of a lower bound for $R$ called Cramer-Rao lower bound for unbiased estimators
- Asymptotic approach of estimation :
  - Observation vector : $X = [X_1, \ldots, X_N]^\top$
  - Asymptotic unbiasedness : $\lim_{N \to +\infty} b(\theta, \widehat{\theta}_N) = 0$
  - Mean square consistency :
  $$\lim_{N \mapsto +\infty} R(\theta, \widehat{\theta}_N) = 0$$



# Estimation of the mean

- Let $X_t$ be a WSS process of mean $\mu_X$ and autocovariance function $r_{XX}(k)$
- Empirical mean : $\hat{\mu}_X = \frac{1}{N} \sum_{t=1}^{N} X_t$
- Unbiased estimator : $\mathbb{E}[\hat{\mu}_X] = \mu_X$
- Variance : $\mathrm{Var}[\hat{\mu}_X] = \frac{1}{N} \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N}\right) r_{XX}(k)$
- If in addition $r_{XX} \in l^1(\mathbb{Z})$ :
  - Mean square consistency : $\mathrm{Var}[\hat{\mu}_X] \sim \frac{1}{N} S_{XX}(0)$

# Estimation of the autocovariance function

- $X_t$ is a centered WSS process of autocovariance function $r_{XX}(k)$
- Empirical autocovariance function : $\hat{r}_{XX}(k) = \frac{1}{N}\sum_{t=1}^{N-k} X_{t+k}\overline{X_t}$ if $0 \leq k < N$, $\hat{r}_{XX}(k) = 0$ if $|k| \geq N$, and $\hat{r}_{XX}(-k) = \overline{\hat{r}_{XX}(k)}$ $\forall k \in \mathbb{Z}$
- Property : $\hat{r}_{XX}$ is positive semi-definite
- Asymptotically unbiased estimator :

$$\mathbb{E}[\hat{r}_{XX}(k)] = \left(1 - \frac{|k|}{N}\right) r_{XX}(k)$$

- If in addition $X_t$ is a strong linear process : $X_t = \sum_{k\in\mathbb{Z}} h_k Z_{t-k}$ where $h_k \in l^1(\mathbb{Z})$ and $Z_t \sim \mathrm{IID}(0,\sigma^2)$ with $\mathbb{E}\left[Z_t^4\right] < +\infty$ :
  - Mean square consistency : $\mathrm{Var}[\hat{r}_{XX}(k)] = O\left(\frac{1}{N}\right)$



# Part III

# Non-parametric spectral estimation



# Estimation of the PSD

- Let $X_t$ be a centered WSS process such that $r_{XX} \in l^1(\mathbb{Z})$
- Periodogram : $\hat{S}_{P,XX}(\nu) = \frac{1}{N}\left|\sum_{t=1}^{N} X_t e^{-2i\pi\nu t}\right|^2$
- Correlogram : $\hat{S}_{C,XX}(\nu) = \sum_{k=-(N-1)}^{N-1} \hat{r}_{XX}(k) e^{-2i\pi\nu k}$ where $\hat{r}_{XX}(k) = 0$ if $|k| \geq N$, $\hat{r}_{XX}(-k) = \overline{\hat{r}_{XX}(k)}$ $\forall k \in \mathbb{Z}$, and $\forall k \in [0, N[$,
  - Unbiased estimator of the autocovariance function :
    $$\hat{r}_{XX}(k) = \frac{1}{N-k}\sum_{t=1}^{N-k} X_{t+k}\overline{X_t}$$
  - Biased estimator of the autocovariance function :
    $$\hat{r}_{XX}(k) = \frac{1}{N}\sum_{t=1}^{N-k} X_{t+k}\overline{X_t} \quad \text{(positive semi-definite)}$$
- Proposition : $\hat{S}_{P,XX}(\nu) = \hat{S}_{C,XX}(\nu)$ if $\hat{r}_{XX}$ is the biased estimator

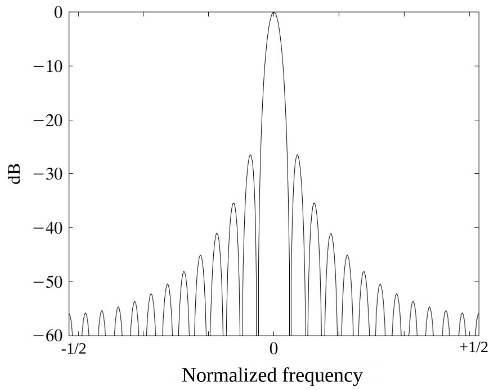


# Bias analysis of the periodogram

- Mean : $\mathbb{E}\left[\hat{S}_{P,XX}(\nu)\right] = \sum_{k=-(N-1)}^{N-1}(1 - \frac{|k|}{N})\, r_{XX}(k)\, e^{-2i\pi\nu k}$
- Let $w_B(k) = 1 - \frac{|k|}{N}$ if $|k| < N$ and $w_B(k) = 0$ if $|k| \geq N$
- $\mathbb{E}\left[\hat{S}_{P,XX}(\nu)\right] = \sum_{k\in\mathbb{Z}} w_B(k) r_{XX}(k) e^{-2i\pi\nu k} = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{XX}(\nu - \xi) W_B(\xi) d\xi$
- Fejer kernel : $W_B(\xi) = \frac{1}{N}\left|\frac{\sin(\pi\xi N)}{\sin(\pi\xi)}\right|^2$
- Consequences :
  - main lobe $\Rightarrow$ smearing (the width is $\frac{2}{N}$)
  - side lobes $\Rightarrow$ leakage
  - loss of resolution (do not confuse with the precision)
- Since $r_{XX} \in l^1(\mathbb{Z})$,, $\hat{S}_{P,XX}(\nu)$ is asymptotically unbiased : $\lim_{N\to+\infty} \mathbb{E}\left[\hat{S}_{P,XX}(\nu)\right] = S_{XX}(\nu)$

# Fejer kernel

$$W_B(\xi) = \frac{1}{N}\left|\frac{\sin(\pi\xi N)}{\sin(\pi\xi)}\right|^2$$




Fejer kernel, $W_B(\nu)/W_B(0)$, for $N = 25$



# Variance analysis of the periodogram

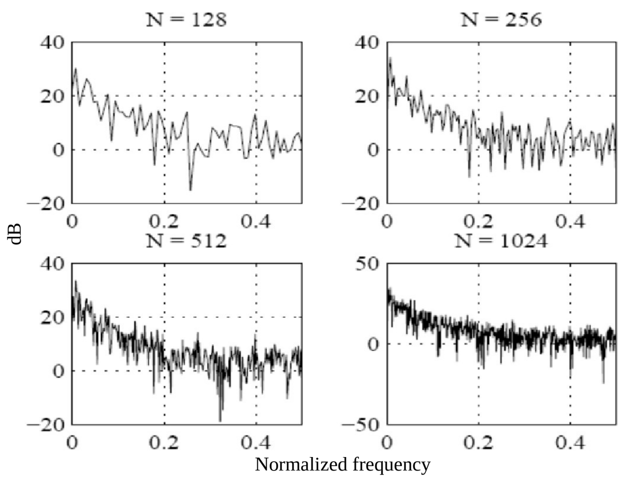
- Definition : complex (or circular) white noise :

$$\begin{cases} \mathbb{E}[Z_t \overline{Z_s}] = \sigma^2 \delta_{t,s} \\ \mathbb{E}[Z_t Z_s] = 0 \ \forall t,s \end{cases} \Leftrightarrow \begin{cases} \mathbb{E}[Re(Z_t)Re(Z_s)] = \frac{1}{2}\sigma^2\delta_{t,s} \\ \mathbb{E}[Im(Z_t)Im(Z_s)] = \frac{1}{2}\sigma^2\delta_{t,s} \\ \mathbb{E}[Re(Z_t)Im(Z_s)] = 0 \ \forall t,s \end{cases}$$

- Proposition : if $Z_t$ is Gaussian white noise, then
$\lim_{N\to+\infty} \mathrm{cov}\left[\hat{S}_{P,ZZ}(\nu), \hat{S}_{P,ZZ}(\xi)\right] = \begin{cases} S_{ZZ}(\nu)^2 & \forall \nu = \xi \\ 0 & \forall \nu \neq \xi \end{cases}$
  - Lemma : if $a, b, c, d$ are jointly Gaussian random variables, then
  $$\mathbb{E}[abcd] = \mathbb{E}[ab]\mathbb{E}[cd] + \mathbb{E}[ac]\mathbb{E}[bd] + \mathbb{E}[ad]\mathbb{E}[bc] - 2\mathbb{E}[a]\mathbb{E}[b]\mathbb{E}[c]\mathbb{E}[d]$$
  - Then $\mathbb{E}\left[\hat{S}_{P,ZZ}(\nu)\hat{S}_{P,ZZ}(\xi)\right] = \sigma^4 + \sigma^4 \left|\frac{\sin(\pi(\nu-\xi)N)}{N\sin(\pi(\nu-\xi))}\right|^2$
- The proposition also holds for $X_t = \sum_{k\in\mathbb{Z}} h_k Z_{t-k}$ where $h_k \in l^1(\mathbb{Z})$
- $\hat{S}_{P,XX}$ is not even asymptotically mean square consistent



# Variance analysis of the periodogram




Periodogram of a WSS process for $N = 128$, 256, 512, 1024.



# Blackman-Tukey method

- Truncation of the empirical autocovariance function ($M < N$) :

$$\hat{S}_{BT,XX}(\nu) = \sum_{k=-M+1}^{M-1} \hat{r}_{XX}(k)\, e^{-2i\pi\nu k}$$

- Properties :
  - If $M \to +\infty$, $\hat{S}_{BT,XX}$ is asymptotically unbiased
  - If $M/N \to 0$, $\mathrm{Var}\left(\hat{S}_{BT,XX}(\nu)\right) = O\left(\frac{M}{N}\right) \to 0$
  - e.g. if $M = N^\alpha$ with $0 < \alpha < 1$, $\hat{S}_{BT,XX}$ is mean square consistent
- Trade-off between spectral resolution ($O(\frac{1}{M})$) and variance ($O\left(\frac{M}{N}\right)$)

## Blackman-Tukey method

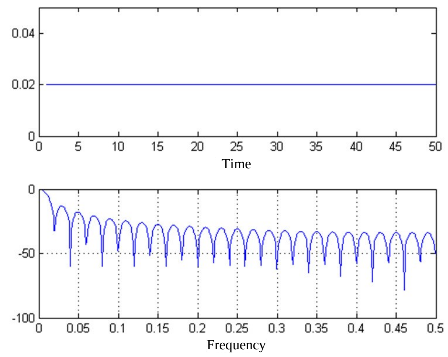
- Windowed periodogram : with $M < N$, $w(-k) = w(k)$, $w(0) = 1$ :
$$\hat{S}_{BT,XX}(\nu) = \sum_{k=-(M-1)}^{M-1} w(k)\widehat{r}_{XX}(k)e^{-2i\pi\nu k}$$
- Then $\hat{S}_{BT,XX}(\nu) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{S}_{P,XX}(\nu - \xi)W(\xi)d\xi$ where $W(\xi) = \sum_{k=-(M-1)}^{M-1} w_k e^{-2i\pi\xi k} \in \mathbb{R} \to$ local weighted average
- If $w(k)$ is positive semidefinite, then $\hat{S}_{BT,XX}(\nu) \geq 0 \ \forall \nu \in \mathbb{R}$
- The choice of the window's length is based on a trade-off between spectral resolution and variance
- The selection of the window's shape is based on a trade-off between smearing and leakage effects



## Rectangular window

$$w(k) = \mathbf{1}_{[-(M-1)\ldots M-1]}(k)$$

- Width : $1/M$, second lobe : -13 dB, decrease : -6 dB / octave

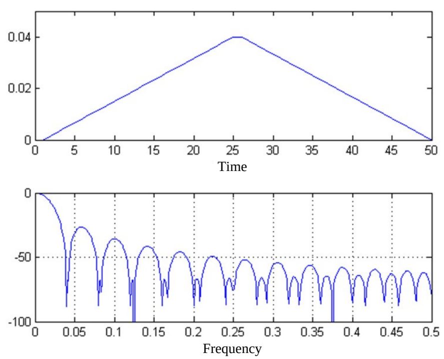





## Bartlett window

$$w(k) = \left(1 - \frac{|k|}{M}\right) \mathbf{1}_{[-(M-1)\ldots M-1]}(k)$$

- Width : $2/M$, second lobe : -26 dB, decrease : -12 dB / octave

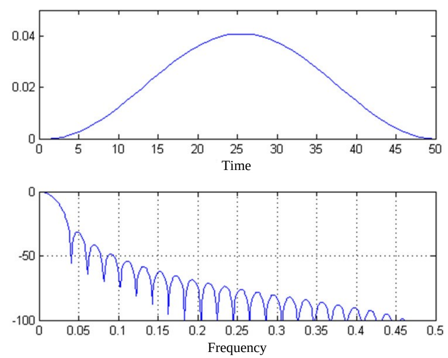





## Hann window

$$w(k) = \left(0.5 + 0.5\cos\left(\frac{\pi k}{M}\right)\right) \mathbf{1}_{[-(M-1)\ldots M-1]}(k)$$

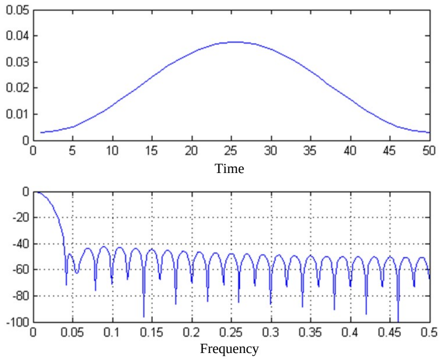
- Width : $2/M$, second lobe : -31 dB, decrease : -18 dB / octave



# Hamming window

$$w(k) = \left(0.54 + 0.46\cos\left(\frac{\pi k}{M-1}\right)\right) \mathbf{1}_{[-(M-1)\ldots M-1]}(k)$$

- Width : $2/M$, second lobe : -41 dB, decrease : -6 dB / octave

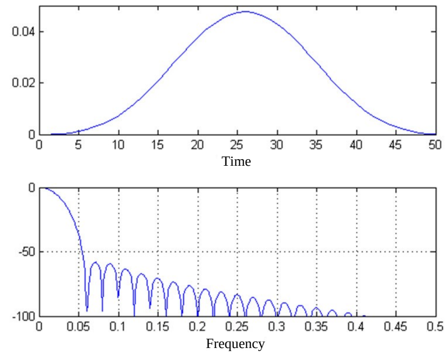





# Blackman window

$$w(k) = \left(0.4266 + 0.4965\cos\left(\frac{\pi k}{M-1}\right) + 0.076\cos\left(\frac{2\pi k}{M-1}\right)\right) \mathbf{1}_{[-(M-1)\ldots M-1]}(k)$$

- Width : $3/M$, second lobe : -57 dB, décroissance : -18 dB / octave






# Bartlett method

- Idea : compared to the original periodogram, split up the data and average several periodograms in order to reduce the variance
- Segment $N$ samples into $L$ sub-samples of size $M = \frac{N}{L}$
- $\hat{S}_{B,XX}(\nu) = \frac{1}{L}\sum_{i=1}^{L} \frac{1}{M}\left|\sum_{t=1}^{M} \tilde{X}_{i,t} e^{-2i\pi\nu t}\right|^2$ where $\tilde{X}_{i,t} = X_{(i-1)M+t}$ for $t \in [1, M]$ and $i \in [1, L]$
- The spectral resolution is $O(\frac{1}{M})$ and the variance is $O(\frac{M}{N})$
- Same trade-off between spectral resolution and variance as the Blackman-Tukey estimate with a rectangular window



# Welch method

- Refinement of the Bartlett method :
  - data segments overlap
  - each data segment is windowed
- $\tilde{X}_{i,t} = X_{(i-1)K+t}$ for $t \in [1, M]$ and $i \in [1, S]$
- If $K = M \Rightarrow$ Bartlett : $S = L = \frac{N}{M}$
- Recommended : $K = \frac{M}{2}$, $S \approx \frac{2N}{M}$
- $\hat{S}_{W,XX}(\nu) = \frac{1}{S}\sum\limits_{i=1}^{S} \hat{S}^{(i)}_{P,XX}(\nu)$ and $\hat{S}^{(i)}_{P,XX}(\nu) = \frac{1}{MP}\left|\sum\limits_{t=1}^{M} v(t)\tilde{X}_{i,t} e^{-2i\pi\nu t}\right|^2$ where $P = \frac{1}{N}\sum_{t=1}^{M} |v(t)|^2$ in order to normalize every periodogram
- Better control of smearing and leakage, variance similar to Bartlett

# Daniell method

- Idea : reduce the variance by smoothing the periodogram :

$$\hat{S}_{D,XX}(\nu) = \frac{1}{2J+1} \sum_{j=-J}^{J} \hat{S}_{P,XX}\left(\nu + \frac{j}{\tilde{N}}\right)$$

  where $\tilde{N} = N$ without *zero-padding*, or $\tilde{N} > N$ with *zero-padding*.
- The continuous version of the Daniell method is $\hat{S}_{D,XX}(\nu) = \frac{1}{\beta} \int_{\nu-\frac{\beta}{2}}^{\nu+\frac{\beta}{2}} \hat{S}_{P,XX}(\nu+\xi)\, d\xi$ with $\beta = \frac{2J}{\tilde{N}}$
- It can be seen as a particular case of the Blackman-Tukey method, with $W(\xi) = 1/\beta$ if $\xi \in [-\frac{\beta}{2}, \frac{\beta}{2}]$, or $W(\xi) = 0$ otherwise.