Parametric estimation of line spectra

Linear time series (part 2)
TSIA202b

Roland Badeau

Sinusoidal modeling of audio signals

- Sounds that generate pitch perception have a quasi-periodic waveform

Sinusoidal modeling of audio signals

- Sounds that generate pitch perception have a quasi-periodic waveform
- Spectrum made of harmonic multiples of the fundamental frequency :

Sinusoidal modeling of audio signals

- Sounds that generate pitch perception have a quasi-periodic waveform
- Spectrum made of harmonic multiples of the fundamental frequency :
  - voiced speech sounds, produced by quasi-periodic vibration of the vocal cords

Sinusoidal modeling of audio signals

- Sounds that generate pitch perception have a quasi-periodic waveform
- Spectrum made of harmonic multiples of the fundamental frequency :
  - voiced speech sounds, produced by quasi-periodic vibration of the vocal cords
  - sounds produced by string or wind instruments

Sinusoidal modeling of audio signals

- Sounds that generate pitch perception have a quasi-periodic waveform
- Spectrum made of harmonic multiples of the fundamental frequency :
  - voiced speech sounds, produced by quasi-periodic vibration of the vocal cords
  - sounds produced by string or wind instruments
- The harmonicity property does not always hold :

Sinusoidal modeling of audio signals

- Sounds that generate pitch perception have a quasi-periodic waveform
- Spectrum made of harmonic multiples of the fundamental frequency :
  - voiced speech sounds, produced by quasi-periodic vibration of the vocal cords
  - sounds produced by string or wind instruments
- The harmonicity property does not always hold :
  - Some instruments are slightly inharmonic

Sinusoidal modeling of audio signals

- Sounds that generate pitch perception have a quasi-periodic waveform
- Spectrum made of harmonic multiples of the fundamental frequency :
  - voiced speech sounds, produced by quasi-periodic vibration of the vocal cords
  - sounds produced by string or wind instruments
- The harmonicity property does not always hold :
  - Some instruments are slightly inharmonic
  - Polyphony : overlap of harmonic combs

Sinusoidal modeling of audio signals

- Sounds that generate pitch perception have a quasi-periodic waveform
- Spectrum made of harmonic multiples of the fundamental frequency :
  - voiced speech sounds, produced by quasi-periodic vibration of the vocal cords
  - sounds produced by string or wind instruments
- The harmonicity property does not always hold :
  - Some instruments are slightly inharmonic
  - Polyphony : overlap of harmonic combs
  - Presence of pairs or triplets of close frequencies :

Sinusoidal modeling of audio signals

- Sounds that generate pitch perception have a quasi-periodic waveform
- Spectrum made of harmonic multiples of the fundamental frequency :
  - voiced speech sounds, produced by quasi-periodic vibration of the vocal cords
  - sounds produced by string or wind instruments
- The harmonicity property does not always hold :
  - Some instruments are slightly inharmonic
  - Polyphony : overlap of harmonic combs
  - Presence of pairs or triplets of close frequencies :
    - asymmetry in a bell geometry

Sinusoidal modeling of audio signals

- Sounds that generate pitch perception have a quasi-periodic waveform
- Spectrum made of harmonic multiples of the fundamental frequency :
  - voiced speech sounds, produced by quasi-periodic vibration of the vocal cords
  - sounds produced by string or wind instruments
- The harmonicity property does not always hold :
  - Some instruments are slightly inharmonic
  - Polyphony : overlap of harmonic combs
  - Presence of pairs or triplets of close frequencies :
    - asymmetry in a bell geometry
    - coupling between the strings and bridge (*chevalet*) in a guitar

Sinusoidal modeling of audio signals

- Sounds that generate pitch perception have a quasi-periodic waveform
- Spectrum made of harmonic multiples of the fundamental frequency :
  - voiced speech sounds, produced by quasi-periodic vibration of the vocal cords
  - sounds produced by string or wind instruments
- The harmonicity property does not always hold :
  - Some instruments are slightly inharmonic
  - Polyphony : overlap of harmonic combs
  - Presence of pairs or triplets of close frequencies :
    - asymmetry in a bell geometry
    - coupling between the strings and bridge (*chevalet*) in a guitar
    - pairs or triplets of strings in a piano, plus coupling of the vertical and horizontal vibration modes

Part I

Parametric signal model

Exponential Sinusoidal Model (ESM)

- Exponential amplitude modulation to model the natural damping of free vibrating systems

Exponential Sinusoidal Model (ESM)

- Exponential amplitude modulation to model the natural damping of free vibrating systems
- Real model : $s_t = \sum_{k=0}^{K-1} a_k \, e^{\delta_k t} \cos\left(2\pi f_k t + \phi_k\right)$

Exponential Sinusoidal Model (ESM)

- Exponential amplitude modulation to model the natural damping of free vibrating systems
- Real model : $s_t = \sum_{k=0}^{K-1} a_k \, e^{\delta_k t} \, \cos\left(2\pi f_k t + \phi_k\right)$
- Complex model : $s_t = \sum_{k=0}^{K-1} a_k \, e^{\delta_k t} \, e^{i(2\pi f_k t + \phi_k)}$

Exponential Sinusoidal Model (ESM)

- Exponential amplitude modulation to model the natural damping of free vibrating systems
- Real model : $s_t = \sum_{k=0}^{K-1} a_k \, e^{\delta_k t} \cos\left(2\pi f_k t + \phi_k\right)$
- Complex model : $s_t = \sum_{k=0}^{K-1} a_k \, e^{\delta_k t} \, e^{i(2\pi f_k t + \phi_k)}$
- Compact form : $s_t = \sum_{k=0}^{K-1} \alpha_k \, {z_k}^t$ where

Exponential Sinusoidal Model (ESM)

- Exponential amplitude modulation to model the natural damping of free vibrating systems
- Real model : $s_t = \sum_{k=0}^{K-1} a_k\, e^{\delta_k t} \cos\left(2\pi f_k t + \phi_k\right)$
- Complex model : $s_t = \sum_{k=0}^{K-1} a_k\, e^{\delta_k t}\, e^{i(2\pi f_k t + \phi_k)}$
- Compact form : $s_t = \sum_{k=0}^{K-1} \alpha_k\, {z_k}^t$ where
  - $\alpha_k = a_k e^{i\phi_k}$ is a *complex amplitude*,

Exponential Sinusoidal Model (ESM)

- Exponential amplitude modulation to model the natural damping of free vibrating systems
- Real model : $s_t = \sum_{k=0}^{K-1} a_k\, e^{\delta_k t} \cos\left(2\pi f_k t + \phi_k\right)$
- Complex model : $s_t = \sum_{k=0}^{K-1} a_k\, e^{\delta_k t}\, e^{i(2\pi f_k t + \phi_k)}$
- Compact form : $s_t = \sum_{k=0}^{K-1} \alpha_k\, {z_k}^t$ where
  - $\alpha_k = a_k e^{i\phi_k}$ is a *complex amplitude*,
  - $z_k = e^{\delta_k + i2\pi f_k}$ is a *complex pole*.

Exponential Sinusoidal Model (ESM)

- Exponential amplitude modulation to model the natural damping of free vibrating systems
- Real model : $s_t = \sum_{k=0}^{K-1} a_k \, e^{\delta_k t} \, \cos(2\pi f_k t + \phi_k)$
- Complex model : $s_t = \sum_{k=0}^{K-1} a_k \, e^{\delta_k t} \, e^{i(2\pi f_k t + \phi_k)}$
- Compact form : $s_t = \sum_{k=0}^{K-1} \alpha_k \, {z_k}^t$ where
  - $\alpha_k = a_k e^{i\phi_k}$ is a *complex amplitude*,
  - $z_k = e^{\delta_k + i2\pi f_k}$ is a *complex pole*.
- Hypotheses : for all $k \in \{0 \ldots K-1\}$, $\alpha_k \neq 0$, $z_k \neq 0$, and all poles z_k are pairwise distinct

Exponential Sinusoidal Model (ESM)

- Exponential amplitude modulation to model the natural damping of free vibrating systems
- Real model : $s_t = \sum_{k=0}^{K-1} a_k \, e^{\delta_k t} \cos(2\pi f_k t + \phi_k)$
- Complex model : $s_t = \sum_{k=0}^{K-1} a_k \, e^{\delta_k t} \, e^{i(2\pi f_k t + \phi_k)}$
- Compact form : $s_t = \sum_{k=0}^{K-1} \alpha_k \, {z_k}^t$ where
  - $\alpha_k = a_k e^{i\phi_k}$ is a *complex amplitude*,
  - $z_k = e^{\delta_k + i 2\pi f_k}$ is a *complex pole*.
- Hypotheses : for all $k \in \{0 \ldots K-1\}$, $\alpha_k \neq 0$, $z_k \neq 0$, and all poles z_k are pairwise distinct
- The observed signal x_t is modeled as the signal s_t plus a complex Gaussian white noise b_t of variance σ^2 (sequence of complex IID r.v. of PDF $p(b) = \frac{1}{\pi\sigma^2} \, e^{-\frac{|b|^2}{\sigma^2}}$)

Spectral estimation by Fourier analysis

- Peak detection in the Fourier transform

Spectral estimation by Fourier analysis

- Peak detection in the Fourier transform
- Advantages

Spectral estimation by Fourier analysis

- Peak detection in the Fourier transform
- Advantages
  - existence of a fast algorithm (FFT)

Spectral estimation by Fourier analysis

- Peak detection in the Fourier transform
- Advantages
  - existence of a fast algorithm (FFT)
  - robust estimation method

Spectral estimation by Fourier analysis

- Peak detection in the Fourier transform
- Advantages
  - existence of a fast algorithm (FFT)
  - robust estimation method
- Drawbacks

Spectral estimation by Fourier analysis

- Peak detection in the Fourier transform
- Advantages
  - existence of a fast algorithm (FFT)
  - robust estimation method
- Drawbacks
  - spectral resolution limited by the window length

Spectral estimation by Fourier analysis

- Peak detection in the Fourier transform
- Advantages
  - existence of a fast algorithm (FFT)
  - robust estimation method
- Drawbacks
  - spectral resolution limited by the window length
  - spectral precision limited by the length of the transform

Spectral estimation by Fourier analysis

- Peak detection in the Fourier transform
- Advantages
  - existence of a fast algorithm (FFT)
  - robust estimation method
- Drawbacks
  - spectral resolution limited by the window length
  - spectral precision limited by the length of the transform
  - trade-off between the width of the principal lobe and the height of the secondary lobes induced by the window shape

Spectral estimation by Fourier analysis

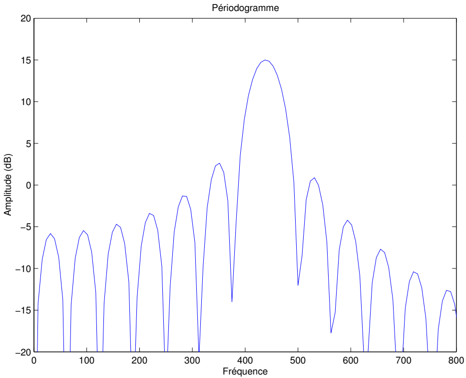
- Peak detection in the Fourier transform
- Advantages
  - existence of a fast algorithm (FFT)
  - robust estimation method
- Drawbacks
  - spectral resolution limited by the window length
  - spectral precision limited by the length of the transform
  - trade-off between the width of the principal lobe and the height of the secondary lobes induced by the window shape
  - widening of the peak in case of exponential damping

Resolution problems

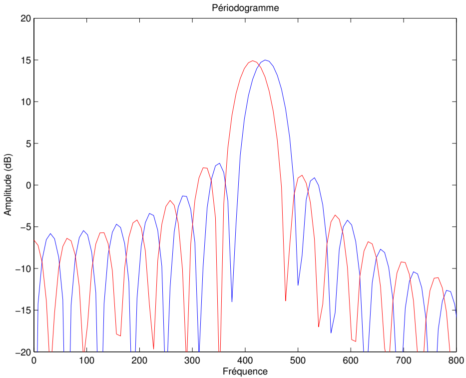
Test signal :

- Sampling frequency : 8000 Hz
- First sinusoid : 440 Hz (A)
- Second sinusoid : 415,3 Hz (G#)
- No damping, all amplitudes equal to 1
- Length of the rectangular window : $N = 128$ (16 ms)
- Length of the transform : 1024 samples

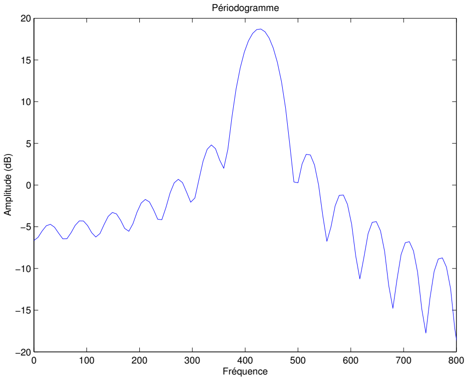
Resolution problems



Resolution problems



Resolution problems



Part II

Maximum Likelihood Method

Definitions

- $\boldsymbol{s}(t) = [s_{t-l+1}, \ldots, s_{t+n-1}]^\top$, of dimension $N = n + l - 1$

Definitions

- $\boldsymbol{s}(t) = [s_{t-l+1}, \ldots, s_{t+n-1}]^\top$, of dimension $N = n + l - 1$
- Let $\boldsymbol{D} = \mathrm{diag}(z_0, \ldots, z_{K-1})$ and $\forall z \in \mathbb{C}$, $\boldsymbol{v}(z) = [1, z, \ldots, z^{N-1}]^\top$

Definitions

- $\boldsymbol{s}(t) = [s_{t-l+1}, \ldots, s_{t+n-1}]^\top$, of dimension $N = n + l - 1$
- Let $\boldsymbol{D} = \text{diag}(z_0, \ldots, z_{K-1})$ and $\forall z \in \mathbb{C}$, $\boldsymbol{v}(z) = [1, z, \ldots, z^{N-1}]^\top$
- $\boldsymbol{V}^N = [\boldsymbol{v}(z_0), \ldots, \boldsymbol{v}(z_{K-1})]$ is a Vandermonde matrix :

$$\boldsymbol{V}^N = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ {z_0}^{N-1} & {z_1}^{N-1} & \ldots & {z_{K-1}}^{N-1} \end{bmatrix}$$

Definitions

- $\boldsymbol{s}(t) = [s_{t-l+1}, \dots, s_{t+n-1}]^\top$, of dimension $N = n + l - 1$
- Let $\boldsymbol{D} = \mathrm{diag}(z_0, \dots, z_{K-1})$ and $\forall z \in \mathbb{C}$, $\boldsymbol{v}(z) = [1, z, \dots, z^{N-1}]^\top$
- $\boldsymbol{V}^N = [\boldsymbol{v}(z_0), \dots, \boldsymbol{v}(z_{K-1})]$ is a Vandermonde matrix :

$$\boldsymbol{V}^N = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_0 & z_1 & \dots & z_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ {z_0}^{N-1} & {z_1}^{N-1} & \dots & {z_{K-1}}^{N-1} \end{bmatrix}$$

- $\boldsymbol{V}^N$ is of full-rank if and only if all the z_k are distinct

Definitions

- $\boldsymbol{s}(t) = [s_{t-l+1}, \ldots, s_{t+n-1}]^\top$, of dimension $N = n + l - 1$
- Let $\boldsymbol{D} = \mathrm{diag}(z_0, \ldots, z_{K-1})$ and $\forall z \in \mathbb{C}$, $\boldsymbol{v}(z) = [1, z, \ldots, z^{N-1}]^\top$
- $\boldsymbol{V}^N = [\boldsymbol{v}(z_0), \ldots, \boldsymbol{v}(z_{K-1})]$ is a Vandermonde matrix :

$$\boldsymbol{V}^N = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ {z_0}^{N-1} & {z_1}^{N-1} & \ldots & {z_{K-1}}^{N-1} \end{bmatrix}$$

- $\boldsymbol{V}^N$ is of full-rank if and only if all the z_k are distinct
- Let $\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_{K-1}]^\top$ and $\boldsymbol{\alpha}(t) = \boldsymbol{D}^{t-l+1}\boldsymbol{\alpha}$

Definitions

- $\boldsymbol{s}(t) = [s_{t-l+1}, \ldots, s_{t+n-1}]^\top$, of dimension $N = n + l - 1$
- Let $\boldsymbol{D} = \mathrm{diag}(z_0, \ldots, z_{K-1})$ and $\forall z \in \mathbb{C}$, $\boldsymbol{v}(z) = [1, z, \ldots, z^{N-1}]^\top$
- $\boldsymbol{V}^N = [\boldsymbol{v}(z_0), \ldots, \boldsymbol{v}(z_{K-1})]$ is a Vandermonde matrix :

$$\boldsymbol{V}^N = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ {z_0}^{N-1} & {z_1}^{N-1} & \ldots & {z_{K-1}}^{N-1} \end{bmatrix}$$

- $\boldsymbol{V}^N$ is of full-rank if and only if all the z_k are distinct
- Let $\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_{K-1}]^\top$ and $\boldsymbol{\alpha}(t) = \boldsymbol{D}^{t-l+1}\boldsymbol{\alpha}$
- $s_t = \sum_{k=0}^{K-1} \alpha_k \, {z_k}^t \Rightarrow \boldsymbol{s}(t) = \boldsymbol{V}^N \boldsymbol{\alpha}(t) \Rightarrow \forall t, \; \boldsymbol{s}(t) \in \mathrm{Span}(\boldsymbol{V}^N)$

Definitions

- $\boldsymbol{s}(t) = [s_{t-l+1}, \ldots, s_{t+n-1}]^\top$, of dimension $N = n + l - 1$
- Let $\boldsymbol{D} = \operatorname{diag}(z_0, \ldots, z_{K-1})$ and $\forall z \in \mathbb{C}$, $\boldsymbol{v}(z) = [1, z, \ldots, z^{N-1}]^\top$
- $\boldsymbol{V}^N = [\boldsymbol{v}(z_0), \ldots, \boldsymbol{v}(z_{K-1})]$ is a Vandermonde matrix :

$$\boldsymbol{V}^N = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ {z_0}^{N-1} & {z_1}^{N-1} & \ldots & {z_{K-1}}^{N-1} \end{bmatrix}$$

- $\boldsymbol{V}^N$ is of full-rank if and only if all the z_k are distinct
- Let $\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_{K-1}]^\top$ and $\boldsymbol{\alpha}(t) = \boldsymbol{D}^{t-l+1}\boldsymbol{\alpha}$
- $s_t = \sum_{k=0}^{K-1} \alpha_k {z_k}^t \Rightarrow \boldsymbol{s}(t) = \boldsymbol{V}^N \boldsymbol{\alpha}(t) \Rightarrow \forall t,\ \boldsymbol{s}(t) \in \operatorname{Span}(\boldsymbol{V}^N)$
- $\boldsymbol{b}(t) = [b_{t-l+1} \ldots b_{t+n-1}]^\top \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_{bb})$ with $\boldsymbol{R}_{bb} = \sigma^2 \boldsymbol{I}_N$

Definitions

- $\boldsymbol{s}(t) = [s_{t-l+1}, \dots, s_{t+n-1}]^\top$, of dimension $N = n + l - 1$
- Let $\boldsymbol{D} = \mathrm{diag}(z_0, \dots, z_{K-1})$ and $\forall z \in \mathbb{C}$, $\boldsymbol{v}(z) = [1, z, \dots, z^{N-1}]^\top$
- $\boldsymbol{V}^N = [\boldsymbol{v}(z_0), \dots, \boldsymbol{v}(z_{K-1})]$ is a Vandermonde matrix :

$$\boldsymbol{V}^N = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_0 & z_1 & \dots & z_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ {z_0}^{N-1} & {z_1}^{N-1} & \dots & {z_{K-1}}^{N-1} \end{bmatrix}$$

- $\boldsymbol{V}^N$ is of full-rank if and only if all the z_k are distinct
- Let $\boldsymbol{\alpha} = [\alpha_0, \dots, \alpha_{K-1}]^\top$ and $\boldsymbol{\alpha}(t) = \boldsymbol{D}^{t-l+1}\boldsymbol{\alpha}$
- $s_t = \sum_{k=0}^{K-1} \alpha_k \, {z_k}^t \Rightarrow \boldsymbol{s}(t) = \boldsymbol{V}^N \boldsymbol{\alpha}(t) \Rightarrow \forall t,\ \boldsymbol{s}(t) \in \mathrm{Span}(\boldsymbol{V}^N)$
- $\boldsymbol{b}(t) = [b_{t-l+1} \dots b_{t+n-1}]^\top \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_{bb})$ with $\boldsymbol{R}_{bb} = \sigma^2 \boldsymbol{I}_N$
- $\boldsymbol{x}(t) = [x_{t-l+1}, \dots, x_{t+n-1}]^\top = \boldsymbol{s}(t) + \boldsymbol{b}(t)$

Maximum likelihood method

- Objective : estimate parameters σ^2, $z_0, \ldots, z_{K-1}$ and $\alpha(t)$

Maximum likelihood method

- Objective : estimate parameters σ^2, $z_0, \dots, z_{K-1}$ and $\alpha(t)$
- General parametric estimation principle, asymptotically unbiased, consistent and efficient

Maximum likelihood method

- Objective : estimate parameters σ^2, $z_0, \dots, z_{K-1}$ and $\alpha(t)$
- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- Probability density function (PDF) of the random vector $\boldsymbol{x}(t) = \boldsymbol{s}(t) + \boldsymbol{b}(t)$, where $\boldsymbol{s}(t) = \boldsymbol{V}^N \alpha(t)$ is deterministic and $\boldsymbol{b}(t) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_{bb})$:

$$p(\boldsymbol{x}(t)) = \frac{1}{\pi^N \det(\boldsymbol{R}_{bb})} e^{-(\boldsymbol{x}(t)-\boldsymbol{s}(t))^H \boldsymbol{R}_{bb}^{-1} (\boldsymbol{x}(t)-\boldsymbol{s}(t))}$$

Maximum likelihood method

- Objective : estimate parameters σ^2, $z_0, \dots, z_{K-1}$ and $\alpha(t)$
- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- Probability density function (PDF) of the random vector $\boldsymbol{x}(t) = \boldsymbol{s}(t) + \boldsymbol{b}(t)$, where $\boldsymbol{s}(t) = \boldsymbol{V}^N \alpha(t)$ is deterministic and $\boldsymbol{b}(t) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_{bb})$:

$$p(\boldsymbol{x}(t)) = \frac{1}{\pi^N \det(\boldsymbol{R}_{bb})} e^{-(\boldsymbol{x}(t)-\boldsymbol{s}(t))^H \boldsymbol{R}_{bb}^{-1} (\boldsymbol{x}(t)-\boldsymbol{s}(t))}$$

- Log-likelihood of the observations :
$L(\sigma^2, z_0 \dots z_{K-1}, \alpha(t)) = -N \ln(\pi\sigma^2) - \frac{1}{\sigma^2} g(z_0 \dots z_{K-1}, \alpha(t))$
where $g(z_0 \dots z_{K-1}, \alpha(t)) = \left(\boldsymbol{x}(t) - \boldsymbol{V}^N \alpha(t)\right)^H \left(\boldsymbol{x}(t) - \boldsymbol{V}^N \alpha(t)\right)$

Maximum likelihood method

- Objective : estimate parameters σ^2, $z_0,\dots,z_{K-1}$ and $\alpha(t)$
- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- Probability density function (PDF) of the random vector $\boldsymbol{x}(t) = \boldsymbol{s}(t) + \boldsymbol{b}(t)$, where $\boldsymbol{s}(t) = \boldsymbol{V}^N\alpha(t)$ is deterministic and $\boldsymbol{b}(t) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_{bb})$:

$$p(\boldsymbol{x}(t)) = \frac{1}{\pi^N \det(\boldsymbol{R}_{bb})} e^{-(\boldsymbol{x}(t)-\boldsymbol{s}(t))^H \boldsymbol{R}_{bb}^{-1} (\boldsymbol{x}(t)-\boldsymbol{s}(t))}$$

- Log-likelihood of the observations :
$L(\sigma^2, z_0 \dots z_{K-1}, \alpha(t)) = -N\ln(\pi\sigma^2) - \frac{1}{\sigma^2} g(z_0 \dots z_{K-1}, \alpha(t))$
where $g(z_0 \dots z_{K-1}, \alpha(t)) = \left(\boldsymbol{x}(t) - \boldsymbol{V}^N\alpha(t)\right)^H \left(\boldsymbol{x}(t) - \boldsymbol{V}^N\alpha(t)\right)$
- Maximization w.r.t. σ^2 : $\sigma^2 = \frac{1}{N}\left\|\boldsymbol{x}(t) - \boldsymbol{V}^N\alpha(t)\right\|^2$

Maximum likelihood method

- Matrix $\boldsymbol{V}^{N^H}\boldsymbol{V}^N$ is invertible because $\boldsymbol{V}^N$ is full-rank

Maximum likelihood method

- Matrix $\boldsymbol{V}^{N^H}\boldsymbol{V}^N$ is invertible because $\boldsymbol{V}^N$ is full-rank
- We finally have to minimize function

$$g(z_0 \dots z_{K-1}, \alpha(t)) = \boldsymbol{x}(t)^H \boldsymbol{x}(t) - \boldsymbol{x}(t)^H \boldsymbol{V}^N \left(\boldsymbol{V}^{N^H}\boldsymbol{V}^N\right)^{-1} \boldsymbol{V}^{N^H}\boldsymbol{x}(t)$$
$$+\left(\alpha(t) - \left(\boldsymbol{V}^{N^H}\boldsymbol{V}^N\right)^{-1} \boldsymbol{V}^{N^H}\boldsymbol{x}(t)\right)^H \left(\boldsymbol{V}^{N^H}\boldsymbol{V}^N\right) \left(\alpha(t) - \left(\boldsymbol{V}^{N^H}\boldsymbol{V}^N\right)^{-1} \boldsymbol{V}^{N^H}\boldsymbol{x}(t)\right)$$

Maximum likelihood method

- Matrix ${\boldsymbol{V}^N}^H \boldsymbol{V}^N$ is invertible because $\boldsymbol{V}^N$ is full-rank
- We finally have to minimize function

$$g(z_0 \dots z_{K-1}, \alpha(t)) = \boldsymbol{x}(t)^H \boldsymbol{x}(t) - \boldsymbol{x}(t)^H \boldsymbol{V}^N \left({\boldsymbol{V}^N}^H \boldsymbol{V}^N\right)^{-1} {\boldsymbol{V}^N}^H \boldsymbol{x}(t)$$
$$+ \left(\alpha(t) - \left({\boldsymbol{V}^N}^H \boldsymbol{V}^N\right)^{-1} {\boldsymbol{V}^N}^H \boldsymbol{x}(t)\right)^H \left({\boldsymbol{V}^N}^H \boldsymbol{V}^N\right) \left(\alpha(t) - \left({\boldsymbol{V}^N}^H \boldsymbol{V}^N\right)^{-1} {\boldsymbol{V}^N}^H \boldsymbol{x}(t)\right)$$

- Minimization w.r.t. $\alpha(t)$: $\alpha(t) = \left({\boldsymbol{V}^N}^H \boldsymbol{V}^N\right)^{-1} {\boldsymbol{V}^N}^H \boldsymbol{x}(t)$

Maximum likelihood method

- Matrix ${\boldsymbol{V}^N}^H \boldsymbol{V}^N$ is invertible because $\boldsymbol{V}^N$ is full-rank
- We finally have to minimize function

$$g(z_0 \dots z_{K-1}, \alpha(t)) = \boldsymbol{x}(t)^H \boldsymbol{x}(t) - \boldsymbol{x}(t)^H \boldsymbol{V}^N \left({\boldsymbol{V}^N}^H \boldsymbol{V}^N\right)^{-1} {\boldsymbol{V}^N}^H \boldsymbol{x}(t)$$
$$+ \left(\alpha(t) - \left({\boldsymbol{V}^N}^H \boldsymbol{V}^N\right)^{-1} {\boldsymbol{V}^N}^H \boldsymbol{x}(t)\right)^H \left({\boldsymbol{V}^N}^H \boldsymbol{V}^N\right) \left(\alpha(t) - \left({\boldsymbol{V}^N}^H \boldsymbol{V}^N\right)^{-1} {\boldsymbol{V}^N}^H \boldsymbol{x}(t)\right)$$

- Minimization w.r.t. $\alpha(t)$: $\alpha(t) = \left({\boldsymbol{V}^N}^H \boldsymbol{V}^N\right)^{-1} {\boldsymbol{V}^N}^H \boldsymbol{x}(t)$
- g is minimal when $(z_0 \dots z_{K-1})$ maximize function
  $\mathcal{J}(z_0, \dots, z_{K-1}) = \boldsymbol{x}(t)^H \boldsymbol{V}^N \left({\boldsymbol{V}^N}^H \boldsymbol{V}^N\right)^{-1} {\boldsymbol{V}^N}^H \boldsymbol{x}(t)$

Maximum likelihood method

- Matrix $\boldsymbol{V^N}^H \boldsymbol{V^N}$ is invertible because $\boldsymbol{V^N}$ is full-rank
- We finally have to minimize function

$$g(z_0 \dots z_{K-1}, \alpha(t)) = \boldsymbol{x}(t)^H \boldsymbol{x}(t) - \boldsymbol{x}(t)^H \boldsymbol{V^N} \left(\boldsymbol{V^N}^H \boldsymbol{V^N} \right)^{-1} \boldsymbol{V^N}^H \boldsymbol{x}(t)$$
$$+ \left(\alpha(t) - \left(\boldsymbol{V^N}^H \boldsymbol{V^N} \right)^{-1} \boldsymbol{V^N}^H \boldsymbol{x}(t) \right)^H \left(\boldsymbol{V^N}^H \boldsymbol{V^N} \right) \left(\alpha(t) - \left(\boldsymbol{V^N}^H \boldsymbol{V^N} \right)^{-1} \boldsymbol{V^N}^H \boldsymbol{x}(t) \right)$$

- Minimization w.r.t. $\alpha(t)$: $\alpha(t) = \left(\boldsymbol{V^N}^H \boldsymbol{V^N} \right)^{-1} \boldsymbol{V^N}^H \boldsymbol{x}(t)$
- g is minimal when $(z_0 \dots z_{K-1})$ maximize function
  $\mathcal{J}(z_0, \dots, z_{K-1}) = \boldsymbol{x}(t)^H \boldsymbol{V^N} \left(\boldsymbol{V^N}^H \boldsymbol{V^N} \right)^{-1} \boldsymbol{V^N}^H \boldsymbol{x}(t)$
- This optimization problem has to be solved numerically

Summary

- General parametric estimation principle, asymptotically unbiased, consistent and efficient

Summary

- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- It leads to a 3-step estimation :

Summary

- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- It leads to a 3-step estimation :
  - Estimation of the complex poles : numerical optimization of a function of K complex variables

Summary

- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- It leads to a 3-step estimation :
  - Estimation of the complex poles : numerical optimization of a function of K complex variables
  - Estimation of the complex amplitudes : by means of the least squares method

Summary

- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- It leads to a 3-step estimation :
  - Estimation of the complex poles : numerical optimization of a function of K complex variables
  - Estimation of the complex amplitudes : by means of the least squares method
  - Estimation of the variance : power of the residual signal

Summary

- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- It leads to a 3-step estimation :
  - Estimation of the complex poles : numerical optimization of a function of K complex variables
  - Estimation of the complex amplitudes : by means of the least squares method
  - Estimation of the variance : power of the residual signal
- Difficulties of the first step :

Summary

- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- It leads to a 3-step estimation :
  - Estimation of the complex poles : numerical optimization of a function of K complex variables
  - Estimation of the complex amplitudes : by means of the least squares method
  - Estimation of the variance : power of the residual signal
- Difficulties of the first step :
  - computational complexity

Summary

- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- It leads to a 3-step estimation :
  - Estimation of the complex poles : numerical optimization of a function of K complex variables
  - Estimation of the complex amplitudes : by means of the least squares method
  - Estimation of the variance : power of the residual signal
- Difficulties of the first step :
  - computational complexity
  - presence of many local maxima

Summary

- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- It leads to a 3-step estimation :
  - Estimation of the complex poles : numerical optimization of a function of K complex variables
  - Estimation of the complex amplitudes : by means of the least squares method
  - Estimation of the variance : power of the residual signal
- Difficulties of the first step :
  - computational complexity
  - presence of many local maxima
- Need for specific methods for the complex poles

Summary

- General parametric estimation principle, asymptotically unbiased, consistent and efficient
- It leads to a 3-step estimation :
  - Estimation of the complex poles : numerical optimization of a function of K complex variables
  - Estimation of the complex amplitudes : by means of the least squares method
  - Estimation of the variance : power of the residual signal
- Difficulties of the first step :
  - computational complexity
  - presence of many local maxima
- Need for specific methods for the complex poles
- High resolution parametric estimation methods overcome the limits of Fourier analysis

Fourier resolution

- Particular case : $\forall k, \; \delta_k = 0$

Fourier resolution

- Particular case : $\forall k,\ \delta_k = 0$
- Closed-form solution if $K = 1$, and approximated solution if $K > 1$

Fourier resolution

- Particular case : $\forall k,\ \delta_k = 0$
- Closed-form solution if $K = 1$, and approximated solution if $K > 1$
- If $K = 1$, $\mathcal{J}(z_0) = \widehat{S}_{P,xx}(f_0)$, where $\widehat{S}_{P,xx}$ is the periodogram

$$\widehat{S}_{P,xx}(f_0) = \frac{1}{N}\left|X(f_0)\right|^2$$

where $X(f_0) = \boldsymbol{v}(e^{i2\pi f_0})^H \boldsymbol{x}(t) = \sum_{\tau=0}^{N-1} x_{t-l+1+\tau}\, e^{-i2\pi f_0 \tau}$

Fourier resolution

- Particular case : $\forall k, \delta_k = 0$
- Closed-form solution if $K = 1$, and approximated solution if $K > 1$
- If $K = 1$, $\mathcal{J}(z_0) = \widehat{S}_{P,xx}(f_0)$, where $\widehat{S}_{P,xx}$ is the periodogram

$$\widehat{S}_{P,xx}(f_0) = \frac{1}{N} \left|X(f_0)\right|^2$$

where $X(f_0) = \boldsymbol{v}(e^{i2\pi f_0})^H \boldsymbol{x}(t) = \sum_{\tau=0}^{N-1} x_{t-l+1+\tau}\, e^{-i2\pi f_0 \tau}$

- In the same way, $\alpha_0(t) = \frac{1}{N} X(f_0)$ and $\sigma^2 = \frac{1}{N}\left(\|\boldsymbol{x}(t)\|^2 - \widehat{S}_{P,xx}(f_0)\right)$

Fourier resolution

- Particular case : $\forall k,\ \delta_k = 0$
- Closed-form solution if $K = 1$, and approximated solution if $K > 1$
- If $K = 1$, $\mathcal{J}(z_0) = \widehat{S}_{P,xx}(f_0)$, where $\widehat{S}_{P,xx}$ is the periodogram

$$\widehat{S}_{P,xx}(f_0) = \frac{1}{N}\left|X(f_0)\right|^2$$

where $X(f_0) = \boldsymbol{v}(e^{i2\pi f_0})^H \boldsymbol{x}(t) = \sum_{\tau=0}^{N-1} x_{t-l+1+\tau}\, e^{-i2\pi f_0 \tau}$

- In the same way, $\alpha_0(t) = \frac{1}{N} X(f_0)$ and $\sigma^2 = \frac{1}{N}\left(\|\boldsymbol{x}(t)\|^2 - \widehat{S}_{P,xx}(f_0)\right)$
- The maximum likelihood principle results in detecting the frequency at which the periodogram reaches its maximum.

Fourier resolution

- Particular case : $\forall k,\ \delta_k = 0$
- Closed-form solution if $K = 1$, and approximated solution if $K > 1$
- If $K = 1$, $\mathcal{J}(z_0) = \widehat{S}_{P,xx}(f_0)$, where $\widehat{S}_{P,xx}$ is the periodogram

$$\widehat{S}_{P,xx}(f_0) = \frac{1}{N}\left|X(f_0)\right|^2$$

where $X(f_0) = \boldsymbol{v}(e^{i2\pi f_0})^H \boldsymbol{x}(t) = \sum\limits_{\tau=0}^{N-1} x_{t-l+1+\tau}\, e^{-i2\pi f_0 \tau}$

- In the same way, $\alpha_0(t) = \frac{1}{N} X(f_0)$ and $\sigma^2 = \frac{1}{N}\left(\|\boldsymbol{x}(t)\|^2 - \widehat{S}_{P,xx}(f_0)\right)$
- The maximum likelihood principle results in detecting the frequency at which the periodogram reaches its maximum.
- The complex amplitude is proportional to the value of the TFD.

Fourier resolution

- Particular case : $\forall k,\ \delta_k = 0$
- Closed-form solution if $K = 1$, and approximated solution if $K > 1$
- If $K = 1$, $\mathcal{J}(z_0) = \widehat{S}_{P,xx}(f_0)$, where $\widehat{S}_{P,xx}$ is the periodogram

$$\widehat{S}_{P,xx}(f_0) = \frac{1}{N} |X(f_0)|^2$$

where $X(f_0) = \boldsymbol{v}(e^{i2\pi f_0})^H \boldsymbol{x}(t) = \sum_{\tau=0}^{N-1} x_{t-l+1+\tau}\, e^{-i2\pi f_0 \tau}$

- In the same way, $\alpha_0(t) = \frac{1}{N} X(f_0)$ and $\sigma^2 = \frac{1}{N}\left(\|\boldsymbol{x}(t)\|^2 - \widehat{S}_{P,xx}(f_0)\right)$
- The maximum likelihood principle results in detecting the frequency at which the periodogram reaches its maximum.
- The complex amplitude is proportional to the value of the TFD.
- The noise variance is the residual power.

Fourier resolution

- If $K \geq 1$, we assume that $N >> \dfrac{1}{\min\limits_{k_1 \neq k_2} \left| f_{k_2} - f_{k_1} \right|}$

Fourier resolution

- If $K \geq 1$, we assume that $N >> \frac{1}{\min\limits_{k_1 \neq k_2} \left| f_{k_2} - f_{k_1} \right|}$
- We have $\left\{ \boldsymbol{V}^{N^H} \boldsymbol{V}^N \right\}_{(k_1,\, k_2)} = \sum_{\tau=0}^{N-1} (z_{k_1}^* z_{k_2})^\tau$.

Fourier resolution

- If $K \geq 1$, we assume that $N >> \frac{1}{\min\limits_{k_1 \neq k_2} \left| f_{k_2} - f_{k_1} \right|}$
- We have $\left\{ \boldsymbol{V}^{N^H} \boldsymbol{V}^N \right\}_{(k_1, k_2)} = \sum_{\tau=0}^{N-1} (z_{k_1}^* z_{k_2})^\tau$.
- Then $\frac{1}{N} \boldsymbol{V}^{N^H} \boldsymbol{V}^N = \boldsymbol{I}_K + O\left(\frac{1}{N}\right)$, thus $\left(\boldsymbol{V}^{N^H} \boldsymbol{V}^N \right)^{-1} = \frac{1}{N} \boldsymbol{I}_K + O\left(\frac{1}{N^2}\right)$

Fourier resolution

- If $K \geq 1$, we assume that $N >> \frac{1}{\min\limits_{k_1 \neq k_2} \left| f_{k_2} - f_{k_1} \right|}$
- We have $\left\{ \boldsymbol{V}^{N^H} \boldsymbol{V}^N \right\}_{(k_1, k_2)} = \sum_{\tau=0}^{N-1} (z_{k_1}^* z_{k_2})^\tau$.
- Then $\frac{1}{N} \boldsymbol{V}^{N^H} \boldsymbol{V}^N = \boldsymbol{I}_K + O\left(\frac{1}{N}\right)$, thus $\left(\boldsymbol{V}^{N^H} \boldsymbol{V}^N \right)^{-1} = \frac{1}{N} \boldsymbol{I}_K + O\left(\frac{1}{N^2}\right)$
- $\mathcal{J}(z_0 \dots z_{K-1}) = \frac{1}{N} \left\| \boldsymbol{V}^{N^H} \boldsymbol{x}(t) \right\|^2 + O\left(\frac{1}{N^2}\right) = \sum\limits_{k=0}^{K-1} \widehat{S}_{P,xx}(f_k) + O\left(\frac{1}{N^2}\right)$

Fourier resolution

- If $K \geq 1$, we assume that $N >> \frac{1}{\min\limits_{k_1 \neq k_2} \left| f_{k_2} - f_{k_1} \right|}$
- We have $\left\{ \boldsymbol{V}^{N^H} \boldsymbol{V}^N \right\}_{(k_1, k_2)} = \sum_{\tau=0}^{N-1} (z_{k_1}^* z_{k_2})^\tau$.
- Then $\frac{1}{N} \boldsymbol{V}^{N^H} \boldsymbol{V}^N = \boldsymbol{I}_K + O\left(\frac{1}{N}\right)$, thus $\left(\boldsymbol{V}^{N^H} \boldsymbol{V}^N \right)^{-1} = \frac{1}{N} \boldsymbol{I}_K + O\left(\frac{1}{N^2}\right)$
- $\mathcal{J}(z_0 \dots z_{K-1}) = \frac{1}{N} \left\| \boldsymbol{V}^{N^H} \boldsymbol{x}(t) \right\|^2 + O\left(\frac{1}{N^2}\right) = \sum\limits_{k=0}^{K-1} \widehat{S}_{P,xx}(f_k) + O\left(\frac{1}{N^2}\right)$
- $\alpha(t) = \frac{1}{N} \boldsymbol{V}^{N^H} \boldsymbol{x}(t) + O\left(\frac{1}{N^2}\right)$, hence $\alpha_k(t) = \frac{1}{N} X(f_k) + O\left(\frac{1}{N^2}\right)$

Fourier resolution

- If $K \geq 1$, we assume that $N >> \frac{1}{\min\limits_{k_1 \neq k_2} \left| f_{k_2} - f_{k_1} \right|}$
- We have $\left\{ \boldsymbol{V^N}^H \boldsymbol{V^N} \right\}_{(k_1, k_2)} = \sum_{\tau=0}^{N-1} (z_{k_1}^* z_{k_2})^\tau$.
- Then $\frac{1}{N} \boldsymbol{V^N}^H \boldsymbol{V^N} = \boldsymbol{I}_K + O\left(\frac{1}{N}\right)$, thus $\left(\boldsymbol{V^N}^H \boldsymbol{V^N} \right)^{-1} = \frac{1}{N} \boldsymbol{I}_K + O\left(\frac{1}{N^2}\right)$
- $\mathcal{J}(z_0 \dots z_{K-1}) = \frac{1}{N} \left\| \boldsymbol{V^N}^H \boldsymbol{x}(t) \right\|^2 + O\left(\frac{1}{N^2}\right) = \sum\limits_{k=0}^{K-1} \widehat{S}_{P,xx}(f_k) + O\left(\frac{1}{N^2}\right)$
- $\alpha(t) = \frac{1}{N} \boldsymbol{V^N}^H \boldsymbol{x}(t) + O\left(\frac{1}{N^2}\right)$, hence $\alpha_k(t) = \frac{1}{N} X(f_k) + O\left(\frac{1}{N^2}\right)$
- $\sigma^2 = \frac{1}{N} \left(\|\boldsymbol{x}(t)\|^2 - \sum_{k=0}^{K-1} \widehat{S}_{P,xx}(f_k) \right) + O\left(\frac{1}{N^2}\right)$.

Fourier resolution

- If $K \geq 1$, we assume that $N >> \frac{1}{\min\limits_{k_1 \neq k_2} \left| f_{k_2} - f_{k_1} \right|}$
- We have $\left\{ \boldsymbol{V}^{N^H} \boldsymbol{V}^N \right\}_{(k_1,\, k_2)} = \sum_{\tau=0}^{N-1} (z_{k_1}^* z_{k_2})^\tau$.
- Then $\frac{1}{N} \boldsymbol{V}^{N^H} \boldsymbol{V}^N = \boldsymbol{I}_K + O\left(\frac{1}{N}\right)$, thus $\left(\boldsymbol{V}^{N^H} \boldsymbol{V}^N \right)^{-1} = \frac{1}{N} \boldsymbol{I}_K + O\left(\frac{1}{N^2}\right)$
- $\mathcal{J}(z_0 \dots z_{K-1}) = \frac{1}{N} \left\| \boldsymbol{V}^{N^H} \boldsymbol{x}(t) \right\|^2 + O\left(\frac{1}{N^2}\right) = \sum\limits_{k=0}^{K-1} \widehat{S}_{P,xx}(f_k) + O\left(\frac{1}{N^2}\right)$
- $\alpha(t) = \frac{1}{N} \boldsymbol{V}^{N^H} \boldsymbol{x}(t) + O\left(\frac{1}{N^2}\right)$, hence $\alpha_k(t) = \frac{1}{N} X(f_k) + O\left(\frac{1}{N^2}\right)$
- $\sigma^2 = \frac{1}{N} \left(\|\boldsymbol{x}(t)\|^2 - \sum_{k=0}^{K-1} \widehat{S}_{P,xx}(f_k) \right) + O\left(\frac{1}{N^2}\right)$.
- Find the K greatest values of the periodogram

Fourier resolution

- If $K \geq 1$, we assume that $N >> \frac{1}{\min\limits_{k_1 \neq k_2} \left| f_{k_2} - f_{k_1} \right|}$
- We have $\left\{ \boldsymbol{V}^{N^H} \boldsymbol{V}^N \right\}_{(k_1,\, k_2)} = \sum_{\tau=0}^{N-1} (z_{k_1}^* z_{k_2})^\tau$.
- Then $\frac{1}{N} \boldsymbol{V}^{N^H} \boldsymbol{V}^N = \boldsymbol{I}_K + O\left(\frac{1}{N}\right)$, thus $\left(\boldsymbol{V}^{N^H} \boldsymbol{V}^N \right)^{-1} = \frac{1}{N} \boldsymbol{I}_K + O\left(\frac{1}{N^2}\right)$
- $\mathcal{J}(z_0 \dots z_{K-1}) = \frac{1}{N} \left\| \boldsymbol{V}^{N^H} \boldsymbol{x}(t) \right\|^2 + O\left(\frac{1}{N^2}\right) = \sum\limits_{k=0}^{K-1} \widehat{S}_{P,xx}(f_k) + O\left(\frac{1}{N^2}\right)$
- $\boldsymbol{\alpha}(t) = \frac{1}{N} \boldsymbol{V}^{N^H} \boldsymbol{x}(t) + O\left(\frac{1}{N^2}\right)$, hence $\alpha_k(t) = \frac{1}{N} X(f_k) + O\left(\frac{1}{N^2}\right)$
- $\sigma^2 = \frac{1}{N} \left(\|\boldsymbol{x}(t)\|^2 - \sum_{k=0}^{K-1} \widehat{S}_{P,xx}(f_k) \right) + O\left(\frac{1}{N^2}\right)$.
- Find the K greatest values of the periodogram
- Limit of Fourier analysis : $\min\limits_{k_1 \neq k_2} \left| f_{k_2} - f_{k_1} \right| >> \frac{1}{N}$

Part III

High resolution methods based on linear prediction

Jean-Baptiste Joseph Fourier (1768-1830)

Gaspard-Marie Riche of Prony (1755-1839)

Linear prediction methods

- Principle : any signal such that $s_t - z_0\, s_{t-1} = 0$ is of the form $s_t = \alpha_0\, {z_0}^t$

Linear prediction methods

- Principle : any signal such that $s_t - z_0\, s_{t-1} = 0$ is of the form $s_t = \alpha_0\, {z_0}^t$
- General case : let $P[z] \triangleq \prod_{k=0}^{K-1} (z - z_k) = \sum_{\tau=0}^{K} p_\tau\, z^{K-\tau}$.

Linear prediction methods

- Principle : any signal such that $s_t - z_0\, s_{t-1} = 0$ is of the form $s_t = \alpha_0\, {z_0}^t$
- General case : let $P[z] \triangleq \prod_{k=0}^{K-1} (z - z_k) = \sum_{\tau=0}^{K} p_\tau\, z^{K-\tau}$.
- *A discrete signal* $\{s_t\}_{t\in\mathbb{Z}}$ *is solution of the recursion* $\sum_{\tau=0}^{K} p_\tau\, s_{t-\tau} = 0$ *if and only if it is of the form* $s_t = \sum_{k=0}^{K-1} \alpha_k\, {z_k}^t$

Linear prediction methods

- Principle : any signal such that $s_t - z_0\, s_{t-1} = 0$ is of the form $s_t = \alpha_0\, {z_0}^t$
- General case : let $P[z] \triangleq \prod_{k=0}^{K-1} (z - z_k) = \sum_{\tau=0}^{K} p_\tau\, z^{K-\tau}$.
- *A discrete signal $\{s_t\}_{t\in\mathbb{Z}}$ is solution of the recursion $\sum_{\tau=0}^{K} p_\tau\, s_{t-\tau} = 0$ if and only if it is of the form $s_t = \sum_{k=0}^{K-1} \alpha_k\, {z_k}^t$*
- Prony and Pisarenko methods :

Linear prediction methods

- Principle : any signal such that $s_t - z_0\, s_{t-1} = 0$ is of the form $s_t = \alpha_0\, {z_0}^t$
- General case : let $P[z] \triangleq \prod_{k=0}^{K-1} (z - z_k) = \sum_{\tau=0}^{K} p_\tau\, z^{K-\tau}$.
- *A discrete signal $\{s_t\}_{t \in \mathbb{Z}}$ is solution of the recursion $\sum_{\tau=0}^{K} p_\tau\, s_{t-\tau} = 0$ if and only if it is of the form $s_t = \sum_{k=0}^{K-1} \alpha_k\, {z_k}^t$*
- Prony and Pisarenko methods :
    - Estimate polynomial $P[z]$ by means of linear prediction

Linear prediction methods

- Principle : any signal such that $s_t - z_0\, s_{t-1} = 0$ is of the form $s_t = \alpha_0\, {z_0}^t$
- General case : let $P[z] \triangleq \prod_{k=0}^{K-1} (z - z_k) = \sum_{\tau=0}^{K} p_\tau\, z^{K-\tau}$.
- *A discrete signal $\{s_t\}_{t\in\mathbb{Z}}$ is solution of the recursion $\sum_{\tau=0}^{K} p_\tau\, s_{t-\tau} = 0$ if and only if it is of the form $s_t = \sum_{k=0}^{K-1} \alpha_k\, {z_k}^t$*
- Prony and Pisarenko methods :
  - Estimate polynomial $P[z]$ by means of linear prediction
  - Extract the roots of this polynomial

Linear prediction methods

- Principle : any signal such that $s_t - z_0\, s_{t-1} = 0$ is of the form $s_t = \alpha_0\, {z_0}^t$
- General case : let $P[z] \triangleq \prod_{k=0}^{K-1} (z - z_k) = \sum_{\tau=0}^{K} p_\tau\, z^{K-\tau}$.
- *A discrete signal $\{s_t\}_{t\in\mathbb{Z}}$ is solution of the recursion $\sum_{\tau=0}^{K} p_\tau\, s_{t-\tau} = 0$ if and only if it is of the form $s_t = \sum_{k=0}^{K-1} \alpha_k\, {z_k}^t$*
- Prony and Pisarenko methods :
  - Estimate polynomial $P[z]$ by means of linear prediction
  - Extract the roots of this polynomial
- Drawback : mediocre performance in presence of noise

Prony method

- Let $\varepsilon_t \triangleq \sum_{\tau=0}^{K} p_\tau \, x_{t-\tau} = \sum_{\tau=0}^{K} p_\tau \, b_{t-\tau}$ be the prediction error

Prony method

- Let $\varepsilon_t \triangleq \sum_{\tau=0}^{K} p_\tau \, x_{t-\tau} = \sum_{\tau=0}^{K} p_\tau \, b_{t-\tau}$ be the prediction error
- We let $n = K + 1$, and we assume that $l \geq K + 1$

Prony method

- Let $\varepsilon_t \triangleq \sum_{\tau=0}^{K} p_\tau\, x_{t-\tau} = \sum_{\tau=0}^{K} p_\tau\, b_{t-\tau}$ be the prediction error
- We let $n = K + 1$, and we assume that $l \geq K + 1$
- Then $\boldsymbol{p}^H \boldsymbol{X}(t) = \boldsymbol{\varepsilon}(t)^H$ where $\boldsymbol{p} = [p_K, p_{K-1}, \ldots, p_0]^H$, $\boldsymbol{\varepsilon}(t) = [\varepsilon_{t-l+K+1}, \varepsilon_{t-l+K+2}, \ldots, \varepsilon_{t+K}]^H$ and

$$\boldsymbol{X}(t) = \begin{bmatrix} x_{t-l+1} & \cdots & x_{t-1} & x_t \\ x_{t-l+2} & \cdots & x_t & x_{t+1} \\ \vdots & \cdots & \vdots & \vdots \\ x_{t-l+K+1} & \cdots & x_{t+K-1} & x_{t+K} \end{bmatrix}$$

Prony method

- Let $\varepsilon_t \triangleq \sum_{\tau=0}^{K} p_\tau \, x_{t-\tau} = \sum_{\tau=0}^{K} p_\tau \, b_{t-\tau}$ be the prediction error
- We let $n = K+1$, and we assume that $l \geq K+1$
- Then $\boldsymbol{p}^H \boldsymbol{X}(t) = \boldsymbol{\varepsilon}(t)^H$ where $\boldsymbol{p} = [p_K, p_{K-1}, \ldots, p_0]^H$, $\boldsymbol{\varepsilon}(t) = [\varepsilon_{t-l+K+1}, \varepsilon_{t-l+K+2}, \ldots, \varepsilon_{t+K}]^H$ and

$$\boldsymbol{X}(t) = \begin{bmatrix} x_{t-l+1} & \cdots & x_{t-1} & x_t \\ x_{t-l+2} & \cdots & x_t & x_{t+1} \\ \vdots & \cdots & \vdots & \vdots \\ x_{t-l+K+1} & \cdots & x_{t+K-1} & x_{t+K} \end{bmatrix}$$

- We minimize $\frac{1}{l} \|\boldsymbol{\varepsilon}\|^2$ w.r.t. $\boldsymbol{p}$, under the constraint $p_0 = 1$

Prony method

- Let $\varepsilon_t \triangleq \sum_{\tau=0}^{K} p_\tau\, x_{t-\tau} = \sum_{\tau=0}^{K} p_\tau\, b_{t-\tau}$ be the prediction error
- We let $n = K + 1$, and we assume that $l \geq K + 1$
- Then $\boldsymbol{p}^H \boldsymbol{X}(t) = \boldsymbol{\varepsilon}(t)^H$ where $\boldsymbol{p} = [p_K, p_{K-1}, \ldots, p_0]^H$, $\boldsymbol{\varepsilon}(t) = [\varepsilon_{t-l+K+1}, \varepsilon_{t-l+K+2}, \ldots, \varepsilon_{t+K}]^H$ and

$$\boldsymbol{X}(t) = \begin{bmatrix} x_{t-l+1} & \cdots & x_{t-1} & x_t \\ x_{t-l+2} & \cdots & x_t & x_{t+1} \\ \vdots & \cdots & \vdots & \vdots \\ x_{t-l+K+1} & \cdots & x_{t+K-1} & x_{t+K} \end{bmatrix}$$

- We minimize $\frac{1}{l}\|\boldsymbol{\varepsilon}\|^2$ w.r.t. $\boldsymbol{p}$, under the constraint $p_0 = 1$
- However $\frac{1}{l}\|\boldsymbol{\varepsilon}\|^2 = \boldsymbol{p}^H \widehat{\boldsymbol{R}}_{xx}(t)\, \boldsymbol{p}$, where $\widehat{\boldsymbol{R}}_{xx}(t) = \frac{1}{l}\, \boldsymbol{X}(t)\, \boldsymbol{X}(t)^H$

Prony method

- Let $\varepsilon_t \triangleq \sum_{\tau=0}^{K} p_\tau\, x_{t-\tau} = \sum_{\tau=0}^{K} p_\tau\, b_{t-\tau}$ be the prediction error
- We let $n = K+1$, and we assume that $l \geq K+1$
- Then $\boldsymbol{p}^H \boldsymbol{X}(t) = \boldsymbol{\varepsilon}(t)^H$ where $\boldsymbol{p} = [p_K, p_{K-1}, \ldots, p_0]^H$, $\boldsymbol{\varepsilon}(t) = [\varepsilon_{t-l+K+1}, \varepsilon_{t-l+K+2}, \ldots, \varepsilon_{t+K}]^H$ and
$$\boldsymbol{X}(t) = \begin{bmatrix} x_{t-l+1} & \cdots & x_{t-1} & x_t \\ x_{t-l+2} & \cdots & x_t & x_{t+1} \\ \vdots & \cdots & \vdots & \vdots \\ x_{t-l+K+1} & \cdots & x_{t+K-1} & x_{t+K} \end{bmatrix}$$
- We minimize $\frac{1}{l}\,\|\boldsymbol{\varepsilon}\|^2$ w.r.t. $\boldsymbol{p}$, under the constraint $p_0 = 1$
- However $\frac{1}{l}\|\boldsymbol{\varepsilon}\|^2 = \boldsymbol{p}^H \widehat{\boldsymbol{R}}_{xx}(t)\,\boldsymbol{p}$, where $\widehat{\boldsymbol{R}}_{xx}(t) = \frac{1}{l}\,\boldsymbol{X}(t)\,\boldsymbol{X}(t)^H$
- The solution is $\boldsymbol{p} = \frac{1}{\boldsymbol{e}_1^H \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1}\, \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1$ where $\boldsymbol{e}_1 \triangleq [1,\, 0 \ldots 0]^\top$

Prony and Pisarenko methods

- Prony method :
  - Construct matrix $\boldsymbol{X}(t)$ and compute $\widehat{\boldsymbol{R}}_{xx}(t)$
  - Compute $\boldsymbol{p} = \frac{1}{\boldsymbol{e}_1^H \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1} \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1$
  - Determine the z_k's as the roots of $P[z] = \sum_{k=0}^{K} p_k \, z^{K-k}$

Prony and Pisarenko methods

- Prony method :
  - Construct matrix $\boldsymbol{X}(t)$ and compute $\widehat{\boldsymbol{R}}_{xx}(t)$
  - Compute $\boldsymbol{p} = \frac{1}{\boldsymbol{e}_1^H \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1} \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1$
  - Determine the z_k's as the roots of $P[z] = \sum_{k=0}^{K} p_k \, z^{K-k}$
- Method of Pisarenko

Prony and Pisarenko methods

- Prony method :
  - Construct matrix $\boldsymbol{X}(t)$ and compute $\widehat{\boldsymbol{R}}_{xx}(t)$
  - Compute $\boldsymbol{p} = \frac{1}{\boldsymbol{e}_1^H \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1} \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1$
  - Determine the z_k's as the roots of $P[z] = \sum_{k=0}^{K} p_k \, z^{K-k}$
- Method of Pisarenko
  - Minimize $\frac{1}{l} \|\varepsilon\|^2 = \boldsymbol{p}^H \widehat{\boldsymbol{R}}_{xx}(t) \, \boldsymbol{p}$ under the constraint $\|\boldsymbol{p}\|_2 = 1$

Prony and Pisarenko methods

- Prony method :
  - Construct matrix $\boldsymbol{X}(t)$ and compute $\widehat{\boldsymbol{R}}_{xx}(t)$
  - Compute $\boldsymbol{p} = \frac{1}{\boldsymbol{e}_1^H \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1} \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1$
  - Determine the z_k's as the roots of $P[z] = \sum_{k=0}^{K} p_k \, z^{K-k}$
- Method of Pisarenko
  - Minimize $\frac{1}{l} \|\varepsilon\|^2 = \boldsymbol{p}^H \widehat{\boldsymbol{R}}_{xx}(t) \, \boldsymbol{p}$ under the constraint $\|\boldsymbol{p}\|_2 = 1$
  - Solution : $\boldsymbol{p}$ = eigenvector of $\widehat{\boldsymbol{R}}_{xx}(t)$ of lowest eigenvalue

Prony and Pisarenko methods

- Prony method :
  - Construct matrix $\boldsymbol{X}(t)$ and compute $\widehat{\boldsymbol{R}}_{xx}(t)$
  - Compute $\boldsymbol{p} = \frac{1}{\boldsymbol{e}_1^H \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1} \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1$
  - Determine the z_k's as the roots of $P[z] = \sum_{k=0}^{K} p_k \, z^{K-k}$
- Method of Pisarenko
  - Minimize $\frac{1}{l} \|\varepsilon\|^2 = \boldsymbol{p}^H \widehat{\boldsymbol{R}}_{xx}(t) \, \boldsymbol{p}$ under the constraint $\|\boldsymbol{p}\|_2 = 1$
  - Solution : $\boldsymbol{p}$ = eigenvector of $\widehat{\boldsymbol{R}}_{xx}(t)$ of lowest eigenvalue
  - Pisarenko method :
    - Construct the matrix $\boldsymbol{X}(t)$ and compute $\widehat{\boldsymbol{R}}_{xx}(t)$
    - Diagonalize $\widehat{\boldsymbol{R}}_{xx}(t)$
    - $\boldsymbol{p}$ = eigenvector of $\widehat{\boldsymbol{R}}_{xx}(t)$ of lowest eigenvalue
    - Determine the z_k's as the roots of $P[z] = \sum_{k=0}^{K} p_k \, z^{K-k}$

Part IV

Subspace-based HR methods

Matrix representation of the signal

- Observation horizon : $t \in \{0 \dots N-1\}$, where $N > 2K$

Matrix representation of the signal

- Observation horizon : $t \in \{0 \dots N-1\}$, where $N > 2K$
- Data matrix ($n > K$, $l > K$ and $N = n + l - 1$) :

$$\boldsymbol{S} = \begin{bmatrix} s_0 & s_1 & \dots & s_{l-1} \\ s_1 & s_2 & \dots & s_l \\ \vdots & \vdots & \vdots & \vdots \\ s_{n-1} & s_n & \dots & s_{N-1} \end{bmatrix}$$

Matrix representation of the signal

- Observation horizon : $t \in \{0 \dots N-1\}$, where $N > 2K$
- Data matrix ($n > K$, $l > K$ and $N = n + l - 1$) :

$$\boldsymbol{S} = \begin{bmatrix} s_0 & s_1 & \dots & s_{l-1} \\ s_1 & s_2 & \dots & s_l \\ \vdots & \vdots & \vdots & \vdots \\ s_{n-1} & s_n & \dots & s_{N-1} \end{bmatrix}$$

- Factorization of matrix $\boldsymbol{S}$: $\boldsymbol{S} = \boldsymbol{V}^n \, \boldsymbol{A} \, {\boldsymbol{V}^l}^\top$, where

Matrix representation of the signal

- Observation horizon : $t \in \{0 \dots N-1\}$, where $N > 2K$
- Data matrix ($n > K$, $l > K$ and $N = n + l - 1$) :

$$\boldsymbol{S} = \begin{bmatrix} s_0 & s_1 & \dots & s_{l-1} \\ s_1 & s_2 & \dots & s_l \\ \vdots & \vdots & \vdots & \vdots \\ s_{n-1} & s_n & \dots & s_{N-1} \end{bmatrix}$$

- Factorization of matrix $\boldsymbol{S}$: $\boldsymbol{S} = \boldsymbol{V}^n \boldsymbol{A} \boldsymbol{V}^{l\top}$, where
  - $\boldsymbol{V}^n$ is the Vandermonde matrix of dimension $n \times K$,

$$\boldsymbol{V}^n = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_0 & z_1 & \dots & z_{K-1} \\ {z_0}^2 & {z_1}^2 & \dots & {z_{K-1}}^2 \\ \vdots & \vdots & \vdots & \vdots \\ {z_0}^{n-1} & {z_1}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}$$

Matrix representation of the signal

- Observation horizon : $t \in \{0 \dots N-1\}$, where $N > 2K$
- Data matrix ($n > K$, $l > K$ and $N = n + l - 1$) :

$$\boldsymbol{S} = \begin{bmatrix} s_0 & s_1 & \dots & s_{l-1} \\ s_1 & s_2 & \dots & s_l \\ \vdots & \vdots & \vdots & \vdots \\ s_{n-1} & s_n & \dots & s_{N-1} \end{bmatrix}$$

- Factorization of matrix $\boldsymbol{S}$: $\boldsymbol{S} = \boldsymbol{V}^n \, \boldsymbol{A} \, \boldsymbol{V}^{l\top}$, where
  - $\boldsymbol{V}^n$ is the Vandermonde matrix of dimension $n \times K$,
  - $\boldsymbol{V}^l$ is the Vandermonde matrix of dimension $l \times K$,

Matrix representation of the signal

- Observation horizon : $t \in \{0 \dots N-1\}$, where $N > 2K$
- Data matrix ($n > K$, $l > K$ and $N = n + l - 1$) :

$$\boldsymbol{S} = \begin{bmatrix} s_0 & s_1 & \dots & s_{l-1} \\ s_1 & s_2 & \dots & s_l \\ \vdots & \vdots & \vdots & \vdots \\ s_{n-1} & s_n & \dots & s_{N-1} \end{bmatrix}$$

- Factorization of matrix $\boldsymbol{S}$: $\boldsymbol{S} = \boldsymbol{V}^n \boldsymbol{A} \boldsymbol{V}^{l\top}$, where
  - $\boldsymbol{V}^n$ is the Vandermonde matrix of dimension $n \times K$,
  - $\boldsymbol{V}^l$ is the Vandermonde matrix of dimension $l \times K$,
  - $\boldsymbol{A} = \mathrm{diag}(\alpha_0, \alpha_1 \dots \alpha_{K-1})$ is a diagonal matrix of dimension $K \times K$.

Empirical covariance matrix

- Let us define the empirical covariance matrix $\boldsymbol{R}_{ss} = \frac{1}{l}\,\boldsymbol{S}\,\boldsymbol{S}^H$

Empirical covariance matrix

- Let us define the empirical covariance matrix $\boldsymbol{R}_{ss} = \frac{1}{l}\,\boldsymbol{S}\,\boldsymbol{S}^H$
- Then $\boldsymbol{R}_{ss} = \boldsymbol{V}^n\,\boldsymbol{P}\,\boldsymbol{V}^{n^H}$, where $\boldsymbol{P} = \frac{1}{l}\,\boldsymbol{A}\,\boldsymbol{V}^{l^\top}\,\boldsymbol{V}^{l^*}\boldsymbol{A}^H$

Empirical covariance matrix

- Let us define the empirical covariance matrix $\boldsymbol{R}_{ss} = \frac{1}{l}\, \boldsymbol{S}\, \boldsymbol{S}^H$
- Then $\boldsymbol{R}_{ss} = \boldsymbol{V}^n\, \boldsymbol{P}\, \boldsymbol{V}^{nH}$, where $\boldsymbol{P} = \frac{1}{l}\, \boldsymbol{A}\, \boldsymbol{V}^{l\top}\, \boldsymbol{V}^{l*} \boldsymbol{A}^H$
- Matrix $\boldsymbol{R}_{ss}$ has rank K

Empirical covariance matrix

- Let us define the empirical covariance matrix $\boldsymbol{R}_{ss} = \frac{1}{l}\,\boldsymbol{S}\,\boldsymbol{S}^H$
- Then $\boldsymbol{R}_{ss} = \boldsymbol{V}^n\,\boldsymbol{P}\,\boldsymbol{V}^{n^H}$, where $\boldsymbol{P} = \frac{1}{l}\,\boldsymbol{A}\,\boldsymbol{V}^{l^\top}\,\boldsymbol{V}^{l^*}\boldsymbol{A}^H$
- Matrix $\boldsymbol{R}_{ss}$ has rank K
- $\boldsymbol{R}_{ss}$ is diagonalizable in an orthonormal basis $\{\boldsymbol{w}_0 \dots \boldsymbol{w}_{n-1}\}$

Empirical covariance matrix

- Let us define the empirical covariance matrix $\boldsymbol{R}_{ss} = \frac{1}{l}\,\boldsymbol{S}\,\boldsymbol{S}^H$
- Then $\boldsymbol{R}_{ss} = \boldsymbol{V}^n\,\boldsymbol{P}\,\boldsymbol{V}^{nH}$, where $\boldsymbol{P} = \frac{1}{l}\,\boldsymbol{A}\,\boldsymbol{V}^{l^\top}\,\boldsymbol{V}^{l^*}\boldsymbol{A}^H$
- Matrix $\boldsymbol{R}_{ss}$ has rank K
- $\boldsymbol{R}_{ss}$ is diagonalizable in an orthonormal basis $\{\boldsymbol{w}_0 \ldots \boldsymbol{w}_{n-1}\}$
- Its eigenvalues $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{n-1} \geq 0$ are such that

Empirical covariance matrix

- Let us define the empirical covariance matrix $\boldsymbol{R}_{ss} = \frac{1}{l}\, \boldsymbol{S}\, \boldsymbol{S}^H$
- Then $\boldsymbol{R}_{ss} = \boldsymbol{V}^n\, \boldsymbol{P}\, \boldsymbol{V}^{nH}$, where $\boldsymbol{P} = \frac{1}{l}\, \boldsymbol{A}\, \boldsymbol{V}^{l\top}\, \boldsymbol{V}^{l*} \boldsymbol{A}^H$
- Matrix $\boldsymbol{R}_{ss}$ has rank K
- $\boldsymbol{R}_{ss}$ is diagonalizable in an orthonormal basis $\{\boldsymbol{w}_0 \dots \boldsymbol{w}_{n-1}\}$
- Its eigenvalues $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{n-1} \geq 0$ are such that
  - $\forall i \in \{0 \dots K-1\}$, $\lambda_i > 0$;

Empirical covariance matrix

- Let us define the empirical covariance matrix $\boldsymbol{R}_{ss} = \frac{1}{l}\,\boldsymbol{S}\,\boldsymbol{S}^H$
- Then $\boldsymbol{R}_{ss} = \boldsymbol{V}^n\,\boldsymbol{P}\,\boldsymbol{V}^{nH}$, where $\boldsymbol{P} = \frac{1}{l}\,\boldsymbol{A}\,\boldsymbol{V}^{l\top}\,\boldsymbol{V}^{l*}\boldsymbol{A}^H$
- Matrix $\boldsymbol{R}_{ss}$ has rank K
- $\boldsymbol{R}_{ss}$ is diagonalizable in an orthonormal basis $\{\boldsymbol{w}_0 \dots \boldsymbol{w}_{n-1}\}$
- Its eigenvalues $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{n-1} \geq 0$ are such that
  - $\forall i \in \{0 \dots K-1\}$, $\lambda_i > 0$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda_i = 0$.

Empirical covariance matrix

- Let us define the empirical covariance matrix $\boldsymbol{R}_{ss} = \frac{1}{l}\,\boldsymbol{S}\,\boldsymbol{S}^H$
- Then $\boldsymbol{R}_{ss} = \boldsymbol{V}^n\,\boldsymbol{P}\,\boldsymbol{V}^{nH}$, where $\boldsymbol{P} = \frac{1}{l}\,\boldsymbol{A}\,\boldsymbol{V}^{l\top}\,\boldsymbol{V}^{l*}\boldsymbol{A}^H$
- Matrix $\boldsymbol{R}_{ss}$ has rank K
- $\boldsymbol{R}_{ss}$ is diagonalizable in an orthonormal basis $\{\boldsymbol{w}_0 \dots \boldsymbol{w}_{n-1}\}$
- Its eigenvalues $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{n-1} \geq 0$ are such that
  - $\forall i \in \{0 \dots K-1\}$, $\lambda_i > 0$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda_i = 0$.
- Let $\widehat{\boldsymbol{R}}_{bb} = \frac{1}{l}\boldsymbol{B}\,\boldsymbol{B}^H$ and $\boldsymbol{R}_{bb} = \mathbb{E}\left[\widehat{\boldsymbol{R}}_{bb}\right] = \sigma^2\boldsymbol{I}_n$.

Empirical covariance matrix

- Let us define the empirical covariance matrix $\boldsymbol{R}_{ss} = \frac{1}{l}\,\boldsymbol{S}\,\boldsymbol{S}^H$
- Then $\boldsymbol{R}_{ss} = \boldsymbol{V}^n\,\boldsymbol{P}\,\boldsymbol{V}^{nH}$, where $\boldsymbol{P} = \frac{1}{l}\,\boldsymbol{A}\,\boldsymbol{V}^{l\top}\,\boldsymbol{V}^{l*}\boldsymbol{A}^H$
- Matrix $\boldsymbol{R}_{ss}$ has rank K
- $\boldsymbol{R}_{ss}$ is diagonalizable in an orthonormal basis $\{\boldsymbol{w}_0 \dots \boldsymbol{w}_{n-1}\}$
- Its eigenvalues $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{n-1} \geq 0$ are such that
  - $\forall i \in \{0 \dots K-1\}$, $\lambda_i > 0$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda_i = 0$.
- Let $\widehat{\boldsymbol{R}}_{bb} = \frac{1}{l}\boldsymbol{B}\,\boldsymbol{B}^H$ and $\boldsymbol{R}_{bb} = \mathbb{E}\left[\widehat{\boldsymbol{R}}_{bb}\right] = \sigma^2\boldsymbol{I}_n$.
- In the same way, let $\widehat{\boldsymbol{R}}_{xx} = \frac{1}{l}\boldsymbol{X}\,\boldsymbol{X}^H$ and $\boldsymbol{R}_{xx} = \mathbb{E}\left[\widehat{\boldsymbol{R}}_{xx}\right]$.

Empirical covariance matrix

- Let us define the empirical covariance matrix $\boldsymbol{R}_{ss} = \frac{1}{l}\,\boldsymbol{S}\,\boldsymbol{S}^H$
- Then $\boldsymbol{R}_{ss} = \boldsymbol{V}^n\,\boldsymbol{P}\,\boldsymbol{V}^{nH}$, where $\boldsymbol{P} = \frac{1}{l}\,\boldsymbol{A}\,\boldsymbol{V}^{l\top}\,\boldsymbol{V}^{l*}\boldsymbol{A}^H$
- Matrix $\boldsymbol{R}_{ss}$ has rank K
- $\boldsymbol{R}_{ss}$ is diagonalizable in an orthonormal basis $\{\boldsymbol{w}_0 \dots \boldsymbol{w}_{n-1}\}$
- Its eigenvalues $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{n-1} \geq 0$ are such that
  - $\forall i \in \{0 \dots K-1\}$, $\lambda_i > 0$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda_i = 0$.
- Let $\widehat{\boldsymbol{R}}_{bb} = \frac{1}{l}\boldsymbol{B}\,\boldsymbol{B}^H$ and $\boldsymbol{R}_{bb} = \mathbb{E}\left[\widehat{\boldsymbol{R}}_{bb}\right] = \sigma^2\boldsymbol{I}_n$.
- In the same way, let $\widehat{\boldsymbol{R}}_{xx} = \frac{1}{l}\boldsymbol{X}\,\boldsymbol{X}^H$ and $\boldsymbol{R}_{xx} = \mathbb{E}\left[\widehat{\boldsymbol{R}}_{xx}\right]$.
- Then $\boldsymbol{R}_{xx} = \boldsymbol{R}_{ss} + \sigma^2\boldsymbol{I}_n$

Signal subspace and noise subspace

- For all $i \in \{0 \dots n-1\}$, $\boldsymbol{w}_i$ is also an eigenvector of $\boldsymbol{R}_{xx}$ corresponding to the eigenvalue $\lambda'_i = \lambda_i + \sigma^2$. Therefore,

Signal subspace and noise subspace

- For all $i \in \{0 \dots n-1\}$, $\boldsymbol{w}_i$ is also an eigenvector of $\boldsymbol{R}_{xx}$ corresponding to the eigenvalue $\lambda'_i = \lambda_i + \sigma^2$. Therefore,
  - $\forall i \in \{0 \dots K-1\}$, $\lambda'_i > \sigma^2$;

Signal subspace and noise subspace

- For all $i \in \{0 \dots n-1\}$, $\boldsymbol{w}_i$ is also an eigenvector of $\boldsymbol{R}_{xx}$ corresponding to the eigenvalue $\lambda_i' = \lambda_i + \sigma^2$. Therefore,
  - $\forall i \in \{0 \dots K-1\}$, $\lambda_i' > \sigma^2$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda_i' = \sigma^2$.

Signal subspace and noise subspace

- For all $i \in \{0 \dots n-1\}$, $\boldsymbol{w}_i$ is also an eigenvector of $\boldsymbol{R}_{xx}$ corresponding to the eigenvalue $\lambda'_i = \lambda_i + \sigma^2$. Therefore,
  - $\forall i \in \{0 \dots K-1\}$, $\lambda'_i > \sigma^2$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda'_i = \sigma^2$.
- Let $\boldsymbol{W} = [\boldsymbol{w}_0 \dots \boldsymbol{w}_{K-1}]$, and $\boldsymbol{W}_\perp = [\boldsymbol{w}_K \dots \boldsymbol{w}_{n-1}]$

Signal subspace and noise subspace

- For all $i \in \{0 \dots n-1\}$, $\boldsymbol{w}_i$ is also an eigenvector of $\boldsymbol{R}_{xx}$ corresponding to the eigenvalue $\lambda'_i = \lambda_i + \sigma^2$. Therefore,
  - $\forall i \in \{0 \dots K-1\}$, $\lambda'_i > \sigma^2$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda'_i = \sigma^2$.
- Let $\boldsymbol{W} = [\boldsymbol{w}_0 \dots \boldsymbol{w}_{K-1}]$, and $\boldsymbol{W}_\perp = [\boldsymbol{w}_K \dots \boldsymbol{w}_{n-1}]$
- Then $\mathrm{Span}(\boldsymbol{W}) = \mathrm{Span}(\boldsymbol{V}^n)$ is referred to as the signal subspace

Signal subspace and noise subspace

- For all $i \in \{0 \dots n-1\}$, $\boldsymbol{w}_i$ is also an eigenvector of $\boldsymbol{R}_{xx}$ corresponding to the eigenvalue $\lambda'_i = \lambda_i + \sigma^2$. Therefore,
  - $\forall i \in \{0 \dots K-1\}$, $\lambda'_i > \sigma^2$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda'_i = \sigma^2$.
- Let $\boldsymbol{W} = [\boldsymbol{w}_0 \dots \boldsymbol{w}_{K-1}]$, and $\boldsymbol{W}_\perp = [\boldsymbol{w}_K \dots \boldsymbol{w}_{n-1}]$
- Then $\mathrm{Span}(\boldsymbol{W}) = \mathrm{Span}(\boldsymbol{V}^n)$ is referred to as the signal subspace
- In the same way, $\mathrm{Span}(\boldsymbol{W}_\perp)$ is referred to as the noise subspace

Signal subspace and noise subspace

- For all $i \in \{0 \dots n-1\}$, $\boldsymbol{w}_i$ is also an eigenvector of $\boldsymbol{R}_{xx}$ corresponding to the eigenvalue $\lambda'_i = \lambda_i + \sigma^2$. Therefore,
  - $\forall i \in \{0 \dots K-1\}$, $\lambda'_i > \sigma^2$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda'_i = \sigma^2$.
- Let $\boldsymbol{W} = [\boldsymbol{w}_0 \dots \boldsymbol{w}_{K-1}]$, and $\boldsymbol{W}_\perp = [\boldsymbol{w}_K \dots \boldsymbol{w}_{n-1}]$
- Then $\mathrm{Span}(\boldsymbol{W}) = \mathrm{Span}(\boldsymbol{V}^n)$ is referred to as the signal subspace
- In the same way, $\mathrm{Span}(\boldsymbol{W}_\perp)$ is referred to as the noise subspace
- The poles $\{z_k\}_{k\in\{0\dots K-1\}}$ are the solutions of equation $\left\|\boldsymbol{W}_\perp^H \boldsymbol{v}(z)\right\|^2 = 0$, where $\boldsymbol{v}(z) = [1, z, \dots, z^{n-1}]$

Signal subspace and noise subspace

- For all $i \in \{0 \dots n-1\}$, $\boldsymbol{w}_i$ is also an eigenvector of $\boldsymbol{R}_{xx}$ corresponding to the eigenvalue $\lambda'_i = \lambda_i + \sigma^2$. Therefore,
  - $\forall i \in \{0 \dots K-1\}$, $\lambda'_i > \sigma^2$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda'_i = \sigma^2$.
- Let $\boldsymbol{W} = [\boldsymbol{w}_0 \dots \boldsymbol{w}_{K-1}]$, and $\boldsymbol{W}_\perp = [\boldsymbol{w}_K \dots \boldsymbol{w}_{n-1}]$
- Then $\mathrm{Span}(\boldsymbol{W}) = \mathrm{Span}(\boldsymbol{V}^n)$ is referred to as the signal subspace
- In the same way, $\mathrm{Span}(\boldsymbol{W}_\perp)$ is referred to as the noise subspace
- The poles $\{z_k\}_{k \in \{0 \dots K-1\}}$ are the solutions of equation $\left\| \boldsymbol{W}_\perp^H \boldsymbol{v}(z) \right\|^2 = 0$, where $\boldsymbol{v}(z) = [1, z, \dots, z^{n-1}]$
- The MUSIC method consists in solving this equation

Signal subspace and noise subspace

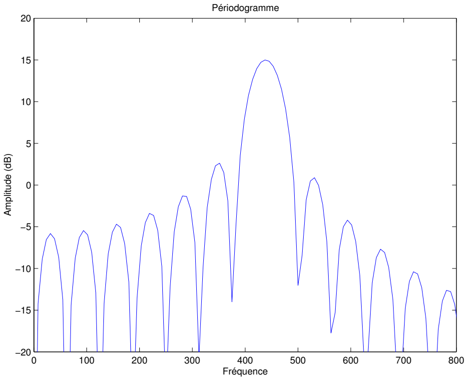
- For all $i \in \{0 \dots n-1\}$, $\boldsymbol{w}_i$ is also an eigenvector of $\boldsymbol{R}_{xx}$ corresponding to the eigenvalue $\lambda'_i = \lambda_i + \sigma^2$. Therefore,
  - $\forall i \in \{0 \dots K-1\}$, $\lambda'_i > \sigma^2$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda'_i = \sigma^2$.
- Let $\boldsymbol{W} = [\boldsymbol{w}_0 \dots \boldsymbol{w}_{K-1}]$, and $\boldsymbol{W}_\perp = [\boldsymbol{w}_K \dots \boldsymbol{w}_{n-1}]$
- Then $\mathrm{Span}(\boldsymbol{W}) = \mathrm{Span}(\boldsymbol{V}^n)$ is referred to as the signal subspace
- In the same way, $\mathrm{Span}(\boldsymbol{W}_\perp)$ is referred to as the noise subspace
- The poles $\{z_k\}_{k\in\{0\dots K-1\}}$ are the solutions of equation $\left\|\boldsymbol{W}_\perp^H \boldsymbol{v}(z)\right\|^2 = 0$, where $\boldsymbol{v}(z) = [1, z, \dots, z^{n-1}]$
- The MUSIC method consists in solving this equation
- The Spectral-MUSIC method consists in detecting the K highest peaks in function $z \mapsto \dfrac{1}{\left\|\boldsymbol{W}_\perp^H \boldsymbol{v}(z)\right\|^2}$.

Spectral MUSIC method

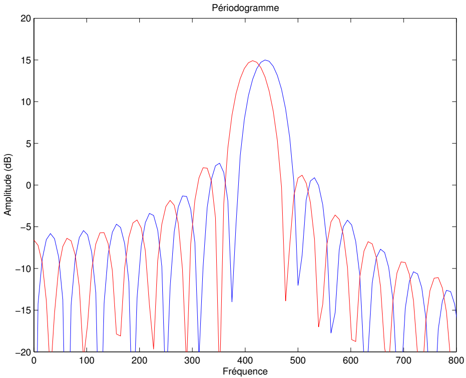
Test signal :

- Sampling frequency : 8000 Hz
- First sinusoid : 440 Hz (A)
- Second sinusoid : 415,3 Hz (G#)
- No damping, all amplitudes equal to 1
- Length of the rectangular window : $N = 128$ (16 ms)
- Length of the transform : 1024 samples

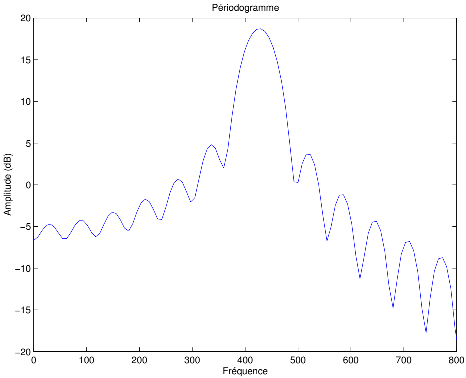
Spectral *MUSIC* method



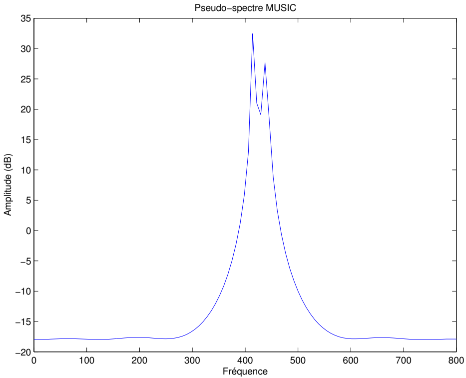
Spectral *MUSIC* method



Spectral *MUSIC* method



Spectral MUSIC method



ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$:

$$\underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\substack{\boldsymbol{V}^n \\ n \times K}}$$

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$:

$$\underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\substack{\boldsymbol{V}^n{}_{\uparrow} \\ (n-1) \times K}}$$

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$:

$$\underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\substack{\boldsymbol{V}^n{}_{\uparrow} \\ (n-1)\times K}} \quad \underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\substack{\boldsymbol{V}^n \\ n\times K}}$$

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$:

$$\underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\substack{\boldsymbol{V}^n{}_{\uparrow} \\ (n-1)\times K}} \quad \underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\substack{\boldsymbol{V}^n{}_{\downarrow} \\ (n-1)\times K}}$$

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$:

$$\underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\substack{\boldsymbol{V}^n{}_{\uparrow} \\ (n-1)\times K}} = \underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\substack{\boldsymbol{V}^n{}_{\downarrow} \\ (n-1)\times K}} \underbrace{\begin{bmatrix} z_0 & & (0) \\ & \ddots & \\ (0) & & z_{K-1} \end{bmatrix}}_{\substack{\boldsymbol{D} \\ K\times K}}$$

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$: $\boldsymbol{V}^n{}_{\uparrow} = \boldsymbol{V}^n{}_{\downarrow}\, \boldsymbol{D}$

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$: $\boldsymbol{V}^n{}_{\uparrow} = \boldsymbol{V}^n{}_{\downarrow}\, \boldsymbol{D}$
- Change of basis : $\boldsymbol{V}^n = \boldsymbol{W}\, \boldsymbol{G}$

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$: $\boldsymbol{V}^n{}_{\uparrow} = \boldsymbol{V}^n{}_{\downarrow}\, \boldsymbol{D}$
- Change of basis : $\boldsymbol{V}^n = \boldsymbol{W}\,\boldsymbol{G}$
- Rotational invariance of $\boldsymbol{W}$: $\boldsymbol{W}_{\uparrow} = \boldsymbol{W}_{\downarrow}\,\boldsymbol{\Phi}$
  where $\boldsymbol{\Phi} = \boldsymbol{G}\,\boldsymbol{D}\,\boldsymbol{G}^{-1}$ is referred to as the spectral matrix

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$: $\boldsymbol{V}^n{}_\uparrow = \boldsymbol{V}^n{}_\downarrow \boldsymbol{D}$
- Change of basis : $\boldsymbol{V}^n = \boldsymbol{W}\,\boldsymbol{G}$
- Rotational invariance of $\boldsymbol{W}$: $\boldsymbol{W}_\uparrow = \boldsymbol{W}_\downarrow \boldsymbol{\Phi}$
  where $\boldsymbol{\Phi} = \boldsymbol{G}\,\boldsymbol{D}\,\boldsymbol{G}^{-1}$ is referred to as the spectral matrix
- The eigenvalues of $\boldsymbol{\Phi}$ are the poles $\{z_k\}_{k \in \{0 \ldots K-1\}}$

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$: $\boldsymbol{V}^n{}_{\uparrow} = \boldsymbol{V}^n{}_{\downarrow}\,\boldsymbol{D}$
- Change of basis : $\boldsymbol{V}^n = \boldsymbol{W}\,\boldsymbol{G}$
- Rotational invariance of $\boldsymbol{W}$: $\boldsymbol{W}_{\uparrow} = \boldsymbol{W}_{\downarrow}\,\boldsymbol{\Phi}$
  where $\boldsymbol{\Phi} = \boldsymbol{G}\,\boldsymbol{D}\,\boldsymbol{G}^{-1}$ is referred to as the spectral matrix
- The eigenvalues of $\boldsymbol{\Phi}$ are the poles $\{z_k\}_{k\in\{0\ldots K-1\}}$
- Matrix $\boldsymbol{\Phi}$ is such that $\boldsymbol{\Phi} = \left(\boldsymbol{W}_{\downarrow}^{H}\boldsymbol{W}_{\downarrow}\right)^{-1}\,\boldsymbol{W}_{\downarrow}^{H}\boldsymbol{W}_{\uparrow}$

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$: $\boldsymbol{V}^n{}_\uparrow = \boldsymbol{V}^n{}_\downarrow \boldsymbol{D}$
- Change of basis : $\boldsymbol{V}^n = \boldsymbol{W}\,\boldsymbol{G}$
- Rotational invariance of $\boldsymbol{W}$: $\boldsymbol{W}_\uparrow = \boldsymbol{W}_\downarrow \boldsymbol{\Phi}$
  where $\boldsymbol{\Phi} = \boldsymbol{G}\,\boldsymbol{D}\,\boldsymbol{G}^{-1}$ is referred to as the spectral matrix
- The eigenvalues of $\boldsymbol{\Phi}$ are the poles $\{z_k\}_{k\in\{0\dots K-1\}}$
- Matrix $\boldsymbol{\Phi}$ is such that $\boldsymbol{\Phi} = \left(\boldsymbol{W}_\downarrow^H \boldsymbol{W}_\downarrow\right)^{-1} \boldsymbol{W}_\downarrow^H \boldsymbol{W}_\uparrow$
- ESPRIT algorithm :

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$: $\boldsymbol{V^n}_{\uparrow} = \boldsymbol{V^n}_{\downarrow}\,\boldsymbol{D}$
- Change of basis : $\boldsymbol{V}^n = \boldsymbol{W}\,\boldsymbol{G}$
- Rotational invariance of $\boldsymbol{W}$: $\boldsymbol{W}_{\uparrow} = \boldsymbol{W}_{\downarrow}\,\boldsymbol{\Phi}$
  where $\boldsymbol{\Phi} = \boldsymbol{G}\,\boldsymbol{D}\,\boldsymbol{G}^{-1}$ is referred to as the spectral matrix
- The eigenvalues of $\boldsymbol{\Phi}$ are the poles $\{z_k\}_{k\in\{0\ldots K-1\}}$
- Matrix $\boldsymbol{\Phi}$ is such that $\boldsymbol{\Phi} = \left(\boldsymbol{W}_{\downarrow}^H \boldsymbol{W}_{\downarrow}\right)^{-1} \boldsymbol{W}_{\downarrow}^H \boldsymbol{W}_{\uparrow}$
- ESPRIT algorithm :
  - compute the estimator $\widehat{\boldsymbol{R}}_{xx}$ of matrix $\boldsymbol{R}_{xx}$,

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$: $\boldsymbol{V}^n{}_\uparrow = \boldsymbol{V}^n{}_\downarrow \boldsymbol{D}$
- Change of basis : $\boldsymbol{V}^n = \boldsymbol{W}\,\boldsymbol{G}$
- Rotational invariance of $\boldsymbol{W}$: $\boldsymbol{W}_\uparrow = \boldsymbol{W}_\downarrow \boldsymbol{\Phi}$
  where $\boldsymbol{\Phi} = \boldsymbol{G}\,\boldsymbol{D}\,\boldsymbol{G}^{-1}$ is referred to as the spectral matrix
- The eigenvalues of $\boldsymbol{\Phi}$ are the poles $\{z_k\}_{k\in\{0\ldots K-1\}}$
- Matrix $\boldsymbol{\Phi}$ is such that $\boldsymbol{\Phi} = \left(\boldsymbol{W}_\downarrow^H \boldsymbol{W}_\downarrow\right)^{-1} \boldsymbol{W}_\downarrow^H \boldsymbol{W}_\uparrow$
- ESPRIT algorithm :
  - compute the estimator $\widehat{\boldsymbol{R}}_{xx}$ of matrix $\boldsymbol{R}_{xx}$,
  - diagonalize it and extract matrix $\boldsymbol{W}$,

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$: $\boldsymbol{V}^n{}_\uparrow = \boldsymbol{V}^n{}_\downarrow \boldsymbol{D}$
- Change of basis : $\boldsymbol{V}^n = \boldsymbol{W}\,\boldsymbol{G}$
- Rotational invariance of $\boldsymbol{W}$: $\boldsymbol{W}_\uparrow = \boldsymbol{W}_\downarrow \boldsymbol{\Phi}$
  where $\boldsymbol{\Phi} = \boldsymbol{G}\,\boldsymbol{D}\,\boldsymbol{G}^{-1}$ is referred to as the spectral matrix
- The eigenvalues of $\boldsymbol{\Phi}$ are the poles $\{z_k\}_{k\in\{0\ldots K-1\}}$
- Matrix $\boldsymbol{\Phi}$ is such that $\boldsymbol{\Phi} = \left(\boldsymbol{W}_\downarrow^H \boldsymbol{W}_\downarrow\right)^{-1} \boldsymbol{W}_\downarrow^H \boldsymbol{W}_\uparrow$
- ESPRIT algorithm :
  - compute the estimator $\widehat{\boldsymbol{R}}_{xx}$ of matrix $\boldsymbol{R}_{xx}$,
  - diagonalize it and extract matrix $\boldsymbol{W}$,
  - compute $\boldsymbol{\Phi} = \left(\boldsymbol{W}_\downarrow^H \boldsymbol{W}_\downarrow\right)^{-1} \boldsymbol{W}_\downarrow^H \boldsymbol{W}_\uparrow$,

ESPRIT method

- Rotational invariance property of $\boldsymbol{V}^n$: $\boldsymbol{V}^n{}_\uparrow = \boldsymbol{V}^n{}_\downarrow \boldsymbol{D}$
- Change of basis : $\boldsymbol{V}^n = \boldsymbol{W}\,\boldsymbol{G}$
- Rotational invariance of $\boldsymbol{W}$: $\boldsymbol{W}_\uparrow = \boldsymbol{W}_\downarrow \boldsymbol{\Phi}$
  where $\boldsymbol{\Phi} = \boldsymbol{G}\,\boldsymbol{D}\,\boldsymbol{G}^{-1}$ is referred to as the spectral matrix
- The eigenvalues of $\boldsymbol{\Phi}$ are the poles $\{z_k\}_{k\in\{0\ldots K-1\}}$
- Matrix $\boldsymbol{\Phi}$ is such that $\boldsymbol{\Phi} = \left(\boldsymbol{W}_\downarrow^H \boldsymbol{W}_\downarrow\right)^{-1} \boldsymbol{W}_\downarrow^H \boldsymbol{W}_\uparrow$
- ESPRIT algorithm :
  - compute the estimator $\widehat{\boldsymbol{R}}_{xx}$ of matrix $\boldsymbol{R}_{xx}$,
  - diagonalize it and extract matrix $\boldsymbol{W}$,
  - compute $\boldsymbol{\Phi} = \left(\boldsymbol{W}_\downarrow^H \boldsymbol{W}_\downarrow\right)^{-1} \boldsymbol{W}_\downarrow^H \boldsymbol{W}_\uparrow$,
  - diagonalize $\boldsymbol{\Phi}$ and get the poles $\{z_k\}_{k\in\{0\ldots K-1\}}$.

Part V

Estimation of the other parameters

Estimation of the modeling order

- Information Theoretic Criteria (ITC) : we minimize

$$\mathrm{ITC}(p) = -(n-p)\,l\,\ln\left(\frac{\left(\prod_{q=p+1}^{n}\sigma_q^2\right)^{\frac{1}{n-p}}}{\frac{1}{n-p}\sum_{q=p+1}^{n}\sigma_q^2}\right) + p\,(2n-p)\,C(l)$$

where σ_q^2 are the eigenvalues of $\widehat{\boldsymbol{R}}_{xx}$ by decreasing order

Estimation of the modeling order

- Information Theoretic Criteria (ITC) : we minimize

$$\mathrm{ITC}(p) = -(n-p)\, l \ln\left(\frac{\left(\prod_{q=p+1}^{n} \sigma_q^2\right)^{\frac{1}{n-p}}}{\frac{1}{n-p}\sum_{q=p+1}^{n} \sigma_q^2}\right) + p\,(2n-p)\, C(l)$$

where σ_q^2 are the eigenvalues of $\widehat{\boldsymbol{R}}_{xx}$ by decreasing order

  - The criterion AIC is given by $C(l) = 1$, and MDL by $C(l) = \frac{1}{2}\ln(l)$

Estimation of the modeling order

- Information Theoretic Criteria (ITC) : we minimize

$$\mathrm{ITC}(p) = -(n-p)\, l \ln\left(\frac{\left(\prod_{q=p+1}^{n} \sigma_q^2\right)^{\frac{1}{n-p}}}{\frac{1}{n-p}\sum_{q=p+1}^{n} \sigma_q^2}\right) + p\,(2n-p)\, C(l)$$

where σ_q^2 are the eigenvalues of $\widehat{\boldsymbol{R}}_{xx}$ by decreasing order

  - The criterion AIC is given by $C(l) = 1$, and MDL by $C(l) = \frac{1}{2}\ln(l)$
  - The criteria EDC are such that $\lim\limits_{l\to+\infty} \frac{C(l)}{l} = 0$ and $\lim\limits_{l\to+\infty} \frac{C(l)}{\ln(\ln(l))} = +\infty$

Estimation of the modeling order

- Information Theoretic Criteria (ITC) : we minimize

$$\mathrm{ITC}(p) = -(n-p)\, l\, \ln\left(\frac{\left(\prod_{q=p+1}^{n} \sigma_q^2\right)^{\frac{1}{n-p}}}{\frac{1}{n-p}\sum_{q=p+1}^{n} \sigma_q^2}\right) + p\,(2n-p)\, C(l)$$

where σ_q^2 are the eigenvalues of $\widehat{\boldsymbol{R}}_{xx}$ by decreasing order

  - The criterion AIC is given by $C(l) = 1$, and MDL by $C(l) = \frac{1}{2}\ln(l)$
  - The criteria EDC are such that $\lim\limits_{l \to +\infty} \frac{C(l)}{l} = 0$ and $\lim\limits_{l \to +\infty} \frac{C(l)}{\ln(\ln(l))} = +\infty$
  - Maximization of the ratio between the geometrical mean of the eigenvalues of the noise subspace and their arithmetical mean.

Estimation of the modeling order

- Information Theoretic Criteria (ITC) : we minimize

$$\mathrm{ITC}(p) = -(n-p)\, l \ln\left(\frac{\left(\prod\limits_{q=p+1}^{n} \sigma_q^2\right)^{\frac{1}{n-p}}}{\frac{1}{n-p}\sum\limits_{q=p+1}^{n} \sigma_q^2}\right) + p\,(2n-p)\, C(l)$$

where σ_q^2 are the eigenvalues of $\widehat{\boldsymbol{R}}_{xx}$ by decreasing order

- The criterion AIC is given by $C(l) = 1$, and MDL by $C(l) = \frac{1}{2}\ln(l)$
- The criteria EDC are such that $\lim\limits_{l\to+\infty} \frac{C(l)}{l} = 0$ and $\lim\limits_{l\to+\infty} \frac{C(l)}{\ln(\ln(l))} = +\infty$
- Maximization of the ratio between the geometrical mean of the eigenvalues of the noise subspace and their arithmetical mean.
- The maximum is reached when all these eigenvalues are equal. The criterion thus measures the noise whiteness.

Estimation of the modeling order

- Information Theoretic Criteria (ITC) : we minimize

$$\mathrm{ITC}(p) = -(n-p)\, l \ln\left(\frac{\left(\prod_{q=p+1}^{n} \sigma_q^2\right)^{\frac{1}{n-p}}}{\frac{1}{n-p}\sum_{q=p+1}^{n} \sigma_q^2}\right) + p\,(2n-p)\, C(l)$$

where σ_q^2 are the eigenvalues of $\widehat{\boldsymbol{R}}_{xx}$ by decreasing order

  - The criterion AIC is given by $C(l) = 1$, and MDL by $C(l) = \frac{1}{2}\ln(l)$
  - The criteria EDC are such that $\lim\limits_{l\to+\infty} \frac{C(l)}{l} = 0$ and $\lim\limits_{l\to+\infty} \frac{C(l)}{\ln(\ln(l))} = +\infty$
  - Maximization of the ratio between the geometrical mean of the eigenvalues of the noise subspace and their arithmetical mean.
  - The maximum is reached when all these eigenvalues are equal. The criterion thus measures the noise whiteness.
  - The penalty term $C(l)$ avoids over-estimating p.

Estimation of the amplitudes and phases

- Let $\boldsymbol{x}$ be the vector $[x_0, x_1, \ldots, x_{N-1}]^\top$ of dimension N

Estimation of the amplitudes and phases

- Let $\boldsymbol{x}$ be the vector $[x_0, x_1, \ldots, x_{N-1}]^\top$ of dimension N
- Let $\boldsymbol{V}^N$ denote the Vandermonde matrix with N rows

Estimation of the amplitudes and phases

- Let $\boldsymbol{x}$ be the vector $[x_0, x_1, \ldots, x_{N-1}]^\top$ of dimension N
- Let $\boldsymbol{V}^N$ denote the Vandermonde matrix with N rows
- Let $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \ldots, \alpha_{K-1}]^\top$ denote the vector of complex amplitudes that we aim to estimate

Estimation of the amplitudes and phases

- Let $\boldsymbol{x}$ be the vector $[x_0, x_1, \ldots, x_{N-1}]^\top$ of dimension N
- Let $\boldsymbol{V}^N$ denote the Vandermonde matrix with N rows
- Let $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \ldots, \alpha_{K-1}]^\top$ denote the vector of complex amplitudes that we aim to estimate
- The maximum likelihood principle leads to using the lest squares method : $\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\| \boldsymbol{x} - \boldsymbol{V}^N \boldsymbol{\beta} \right\|^2$

Estimation of the amplitudes and phases

- Let $\boldsymbol{x}$ be the vector $[x_0, x_1, \ldots, x_{N-1}]^\top$ of dimension N
- Let $\boldsymbol{V}^N$ denote the Vandermonde matrix with N rows
- Let $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \ldots, \alpha_{K-1}]^\top$ denote the vector of complex amplitudes that we aim to estimate
- The maximum likelihood principle leads to using the lest squares method : $\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\| \boldsymbol{x} - \boldsymbol{V}^N \boldsymbol{\beta} \right\|^2$
- The solution is $\widehat{\boldsymbol{\alpha}} = \left(\boldsymbol{V}^{N^H} \boldsymbol{V}^N \right)^{-1} \boldsymbol{V}^{N^H} \boldsymbol{x}$

Estimation of the amplitudes and phases

- Let $\boldsymbol{x}$ be the vector $[x_0, x_1, \ldots, x_{N-1}]^\top$ of dimension N
- Let $\boldsymbol{V}^N$ denote the Vandermonde matrix with N rows
- Let $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \ldots, \alpha_{K-1}]^\top$ denote the vector of complex amplitudes that we aim to estimate
- The maximum likelihood principle leads to using the lest squares method : $\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\| \boldsymbol{x} - \boldsymbol{V}^N \boldsymbol{\beta} \right\|^2$
- The solution is $\widehat{\boldsymbol{\alpha}} = \left(\boldsymbol{V}^{N^H} \boldsymbol{V}^N \right)^{-1} \boldsymbol{V}^{N^H} \boldsymbol{x}$
- We finally get $\widehat{a}_k = |\widehat{\alpha}_k|$ and $\widehat{\phi}_k = \arg(\widehat{\alpha}_k)$

Part VI

Performance of the estimators

Cramér-Rao bounds

- Regular statistical model
  - Consider a statistical model $p(\boldsymbol{x};\,\boldsymbol{\theta})$ parameterized by $\boldsymbol{\theta}$
  - Score function : $\boldsymbol{l}(\boldsymbol{x};\,\boldsymbol{\theta}) \triangleq \boldsymbol{\nabla}_{\boldsymbol{\theta}} \ln p(\boldsymbol{x};\,\boldsymbol{\theta})\, \mathbf{1}_{p(\boldsymbol{x};\,\boldsymbol{\theta})>0}$
  - The parameterization is said *regular* if :
    1. $p(\boldsymbol{x};\,\boldsymbol{\theta})$ is continuously differentiable w.r.t. $\boldsymbol{\theta}$.
    2. $\boldsymbol{F}(\boldsymbol{\theta}) \triangleq \int_H \boldsymbol{l}(\boldsymbol{x};\,\boldsymbol{\theta})\, \boldsymbol{l}(\boldsymbol{x};\,\boldsymbol{\theta})^\top p(\boldsymbol{x};\,\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}$ (Fisher information matrix) is positive definite for all $\boldsymbol{\theta}$ and continuous w.r.t. $\boldsymbol{\theta}$

Cramér-Rao bounds

- Regular statistical model
  - Consider a statistical model $p(\boldsymbol{x};\,\boldsymbol{\theta})$ parameterized by $\boldsymbol{\theta}$
  - Score function : $\boldsymbol{l}(\boldsymbol{x};\,\boldsymbol{\theta}) \triangleq \boldsymbol{\nabla}_{\boldsymbol{\theta}} \ln p(\boldsymbol{x};\,\boldsymbol{\theta})\,\mathbf{1}_{p(\boldsymbol{x};\,\boldsymbol{\theta})>0}$
  - The parameterization is said *regular* if :
    1. $p(\boldsymbol{x};\,\boldsymbol{\theta})$ is continuously differentiable w.r.t. $\boldsymbol{\theta}$.
    2. $\boldsymbol{F}(\boldsymbol{\theta}) \triangleq \int_H \boldsymbol{l}(\boldsymbol{x};\,\boldsymbol{\theta})\,\boldsymbol{l}(\boldsymbol{x};\,\boldsymbol{\theta})^\top\,p(\boldsymbol{x};\,\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x}$ (Fisher information matrix) is positive definite for all $\boldsymbol{\theta}$ and continuous w.r.t. $\boldsymbol{\theta}$
- Cramér-Rao bounds
  - Consider a regular statistic model parameterized by $\boldsymbol{\theta}$
  - Let $\widehat{\boldsymbol{\theta}}$ be an unbiased estimator of $\boldsymbol{\theta}$ ($\forall \boldsymbol{\theta} \in \Theta,\, \mathbb{E}_{\boldsymbol{\theta}}[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$)
  - Then the dispersion matrix $\boldsymbol{D}(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}) \triangleq \mathbb{E}_{\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^\top\right]$ is such that matrix $\boldsymbol{D}(\boldsymbol{\theta},\,\widehat{\boldsymbol{\theta}}) - \boldsymbol{F}(\boldsymbol{\theta})^{-1}$ is positive semidefinite.

Cramér-Rao bounds

- For a family of complex Gaussian distributions of covariance $\boldsymbol{R}_{bb}(\boldsymbol{\theta}) \in \mathcal{C}^1(\Theta, \mathbb{C}^{N\times N})$ and of mean $\boldsymbol{s}(\boldsymbol{\theta}) \in \mathcal{C}^1(\Theta, \mathbb{C}^{N})$,
$$\boldsymbol{F}_{(i,j)}(\boldsymbol{\theta}) = \text{trace}\left(\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial \theta_j}\right) + 2\mathcal{R}e\left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \theta_i}^H \boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \theta_j}\right)$$

Cramér-Rao bounds

- For a family of complex Gaussian distributions of covariance $\boldsymbol{R}_{bb}(\boldsymbol{\theta}) \in \mathcal{C}^1(\Theta, \mathbb{C}^{N\times N})$ and of mean $\boldsymbol{s}(\boldsymbol{\theta}) \in \mathcal{C}^1(\Theta, \mathbb{C}^N)$,
$$\boldsymbol{F}_{(i,j)}(\boldsymbol{\theta}) = \text{trace}\left(\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial\theta_i}\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial\theta_j}\right) + 2\mathcal{R}e\left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial\theta_i}^H \boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial\theta_j}\right)$$
- The Cramér-Rao bounds for the parameters (ϕ_k, δ_k, f_k) are independent of $a_{k'}$ for all $k' \neq k$, and proportional to $1/a_k^2$

Cramér-Rao bounds

- For a family of complex Gaussian distributions of covariance $\boldsymbol{R}_{bb}(\boldsymbol{\theta}) \in \mathcal{C}^1(\Theta, \mathbb{C}^{N\times N})$ and of mean $\boldsymbol{s}(\boldsymbol{\theta}) \in \mathcal{C}^1(\Theta, \mathbb{C}^N)$,
$$\boldsymbol{F}_{(i,j)}(\boldsymbol{\theta}) = \text{trace}\left(\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial \theta_j}\right) + 2\mathcal{R}e\left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \theta_i}^H \boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \theta_j}\right)$$
- The Cramér-Rao bounds for the parameters (ϕ_k, δ_k, f_k) are independent of $a_{k'}$ for all $k' \neq k$, and proportional to $1/a_k^2$
- The bound for parameter a_k is independent of all $a_{k'}$

Cramér-Rao bounds

- For a family of complex Gaussian distributions of covariance $\boldsymbol{R}_{bb}(\boldsymbol{\theta}) \in \mathcal{C}^1(\Theta, \mathbb{C}^{N\times N})$ and of mean $\boldsymbol{s}(\boldsymbol{\theta}) \in \mathcal{C}^1(\Theta, \mathbb{C}^{N})$,
$$\boldsymbol{F}_{(i,j)}(\boldsymbol{\theta}) = \text{trace}\left(\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial \theta_j}\right) + 2\mathcal{R}e\left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \theta_i}^H \boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \theta_j}\right)$$
- The Cramér-Rao bounds for the parameters (ϕ_k, δ_k, f_k) are independent of $a_{k'}$ for all $k' \neq k$, and proportional to $1/a_k^2$
- The bound for parameter a_k is independent of all $a_{k'}$
- All bounds are independent of the phases ϕ_k and are unchanged by any translation of the set of frequencies f_k

Cramér-Rao bounds

- For a family of complex Gaussian distributions of covariance $\boldsymbol{R}_{bb}(\boldsymbol{\theta}) \in \mathcal{C}^1(\Theta, \mathbb{C}^{N\times N})$ and of mean $\boldsymbol{s}(\boldsymbol{\theta}) \in \mathcal{C}^1(\Theta, \mathbb{C}^N)$,
$$\boldsymbol{F}_{(i,j)}(\boldsymbol{\theta}) = \mathrm{trace}\left(\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial \theta_j}\right) + 2\mathcal{R}e\left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \theta_i}^H \boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \theta_j}\right)$$
- The Cramér-Rao bounds for the parameters (ϕ_k, δ_k, f_k) are independent of $a_{k'}$ for all $k' \neq k$, and proportional to $1/a_k^2$
- The bound for parameter a_k is independent of all $a_{k'}$
- All bounds are independent of the phases ϕ_k and are unchanged by any translation of the set of frequencies f_k
- If $\forall k$, $\delta_k = 0$ and if $N \to +\infty$, then :
  - $\mathrm{CRB}\{\sigma\} = \frac{\sigma^2}{4N} + O\left(\frac{1}{N^2}\right)$
  - $\mathrm{CRB}\{f_k\} = \frac{6\sigma^2}{4\pi^2 N^3 a_k^2} + O\left(\frac{1}{N^4}\right)$
  - $\mathrm{CRB}\{a_k\} = \frac{2\sigma^2}{N} + O\left(\frac{1}{N^2}\right)$
  - $\mathrm{CRB}\{\phi_k\} = \frac{2\sigma^2}{Na_k^2} + O\left(\frac{1}{N^2}\right)$

Performance of HR methods

- Performance of an estimator
  - Performance expressed in terms of bias and variance
  - Efficiency : ratio between variance and Cramér-Rao bound
  - An estimator is said *efficient* if its efficiency is equal to 1.

Performance of HR methods

- Performance of an estimator
  - Performance expressed in terms of bias and variance
  - Efficiency : ratio between variance and Cramér-Rao bound
  - An estimator is said *efficient* if its efficiency is equal to 1.
- Maximum likelihood : unbiased and asymptotically efficient ($N \to +\infty$)

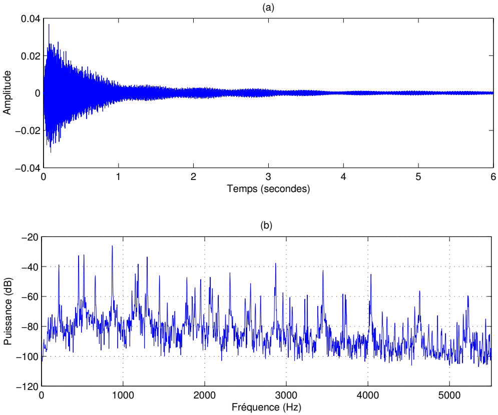
Performance of HR methods

- Performance of an estimator
  - Performance expressed in terms of bias and variance
  - Efficiency : ratio between variance and Cramér-Rao bound
  - An estimator is said *efficient* if its efficiency is equal to 1.
- Maximum likelihood : unbiased and asymptotically efficient ($N \to +\infty$)
- HR methods : results based on the perturbation theory
  - Assumptions : $N \to +\infty$ or SNR$\to +\infty$
  - All the HR methods are asymptotically unbiased
  - The Prony and Pisarenko methods are very inefficient : their variances are significantly greater than the Cramér-Rao bounds.
  - MUSIC and ESPRIT have an asymptotic efficiency close to 1

Part VII

Signals to be processed

Bell sound



(a) Signal waveform
(b) Power spectral density