Ch3. Power-spectrum estimation for sensing the environment (1/2)
in Cognitive Dynamic Systems, S. Haykin

Course: Autonomous Machine Learning

Soojeong Kim

Cloud and Mobile Systems Laboratory
School of Computer Science and Engineering
Seoul National University
http://cmslab.snu.ac.kr
CONTENTS

Part A
  Review: Sensing the environment

Part B
  Power spectrum

Part C
  Power spectrum estimation
    C1. Parametric methods
    C2. Nonparametric methods

Part D
  Multitaper method

Part E
  Multitaper method in space
Part A
Review: Sensing the environment
Ill-posed inverse problem

- Two views of perception: ill-posed inverse problem, Bayesian inference

- Ill-posed problem: have lack of well-posed conditions
  (existence, uniqueness, continuity)

- Inverse problem: uncover underlying physical laws from the stimuli
  = find a mapping from stimuli to the state
Sensing the environment

- Solving an ill-posed inverse problem

- Power spectrum estimation is an example of sensing the environment
Part B

Power spectrum
What is power spectrum?

- The average power of the incoming signal expressed as a function of frequency

- Signal is measured by sensors as a time series (time domain)

- Power spectrum focus on signal power on frequency (frequency domain)
Why power spectrum estimation is useful?

- Impossible to predict future signal
- Estimate the shape of the signal by only finite set of data
- Ex: Radar, Radio, Speech Recognition
Power spectrum in a theory

\[ S(f) = \int_{-\infty}^{\infty} s(t) \cdot e^{-i2\pi ft} \, dt. \]

- Incoming signals are picked up as a stochastic (random) process
- Only finite set of data is used (applying theory is impossible)
- Power spectrum is an ill-posed inverse problem

Fourier transform
Part C

Power spectrum estimation
Power spectrum Estimation

- Parametric methods (model-based)
- Non-parametric methods (model-free)
Parametric method example

Build a model $H(e^{j2\pi f})$

- $H$ is a function to convert time domain signals to frequency domain signals

- Apply prior knowledge of Fourier transform (assume $e^{j2\pi f}$ relationship)

- $H$ contains some parameters/coefficients
Parametric method example

Controlled input signal

- Controlled signal is given -> input signal power on frequency is known value

- Model output is only depends on the model, $H$

White noise of zero mean and variance $\sigma^2$

Stochastic model of process $x(n)$, characterized by transfer function $H(e^{j2\pi f})$

Parameterized process with a rational power spectrum equal to $\sigma^2 |H(e^{j2\pi})|^2$
Parametric method example

Problem: Finding parameters/coefficient of $H$

- Finding parameters of $H$ to produce acceptable output

- Finding parameters is a typical machine learning problem
Parametric method example

After finding good parameters of H

- Model is built!, problem is solved

- Use the model as an estimator
Power spectrum Estimation

- Parametric methods (model-based)
  
  If model is good: produce good quality estimates with less sample data
  
  If model is bad: mislead conclusion

- Non-parametric methods (model-free)
Nonparametric method (periodogram)

\[ x(n) \rightarrow X_N(f) \rightarrow S(f) \quad \text{(Fourier transform)} \]

\[
S(f) = \lim_{n \to \infty} \frac{1}{N} E[|X_N(f)|^2]
\]

Only finite set of signal data is used in practice

Use estimates of \( S(f) \), \( \hat{S}(f) = \frac{1}{N} |X_N(f)|^2 \)
Nonparametric method (periodogram)

\[ \hat{S}(f) = \frac{1}{N} |X_N(f)|^2 \]

Same effect as

Define. \( a(n) = 1 \) if \( n \) is in bound of \( N \), \( 0 \) otherwise

\[ x(n) \times a(n) \rightarrow x^*(n) \rightarrow X_N^*(f) \rightarrow \hat{S}(f) \text{ (Fourier transform)} \]

\( a(n) \) is taper
Nonparametric method (periodogram)

Not require any model to develop

Directly calculate the value using **Fourier transform with taper**
Bias-variance dilemma with taper

2 Conditions for good estimator

1. Low bias

2. Low variance (not noisy)
Bias-variance dilemma with taper

Bias from spectral leakage

Cause of spectral leakage
Bias-variance dilemma with taper

Spectral leakage ↓ by taper (not rectangle) -> bias ↓
Bias-variance dilemma with taper

bias↓ -> variance↑

Taper-> reduce effective sample size -> variance ↑

Bias-variance trade-off!
Part D

Multitaper method
Multitaper method

Address the trade-off using multiple tapers instead of one

K tapers are used on the same sample data

K set of estimates are created

Average K set of estimates

$$\bar{S}(f) = \frac{1}{K} \sum_{k=1}^{K} \hat{S}_k(f).$$
Tapers are Slepian sequences

Tapers following Slepian sequences

$k$th taper: $k \uparrow \rightarrow \text{bias} \uparrow, \text{variance} \downarrow$

\[ \hat{S}_1(f) \]

\[ \hat{S}_7(f) \]
Multitaper method

All K estimates are averaged

Average of $\hat{S}_1(f), \hat{S}_2(f), \hat{S}_3(f)$,
Multitaper method

All $K$ estimates are averaged

Find appropriate $K$ to compromise bias and variance
Why multitaper method is good?

Reduce bias by tapers

Reduce variance by increasing effective sample size ($N \times K$)

Easy to solve bias-variance trade-off by choosing $K$
Part E
Multitaper method in space
Multitaper method in space

Previous methods consider power of signal on frequency.

Extended multitaper method to get power of signal on frequency and space.
Multitaper method in space

\[
A(f) = \begin{bmatrix}
a_0^{(0)} X_0^{(0)}(f) & a_1^{(0)} X_1^{(0)}(f) & \cdots & a_{K-1}^{(0)} X_{K-1}^{(0)}(f) \\
a_0^{(1)} X_0^{(1)}(f) & a_1^{(1)} X_1^{(1)}(f) & \cdots & a_{K-1}^{(1)} X_{K-1}^{(1)}(f) \\
\vdots & \vdots & \ddots & \vdots \\
a_0^{(M-1)} X_0^{(M-1)}(f) & a_1^{(M-1)} X_1^{(M-1)}(f) & \cdots & a_{K-1}^{(M-1)} X_{K-1}^{(M-1)}(f)
\end{bmatrix},
\]

\[X_{m,k}(f)\] : Fourier transform of input signal by m-th sensor

\[a_{m,k}^{m}\] : coefficients accounting for different localized area around the gridpoints
Multitaper method in space

\[ A(f) = \begin{bmatrix}
    a_0^{(0)} X_0^{(0)}(f) & a_1^{(0)} X_1^{(0)}(f) & \ldots & a_{K-1}^{(0)} X_{K-1}^{(0)}(f) \\
    a_0^{(1)} X_0^{(1)}(f) & a_1^{(1)} X_1^{(1)}(f) & \ldots & a_{K-1}^{(1)} X_{K-1}^{(1)}(f) \\
    \vdots & \vdots & \ddots & \vdots \\
    a_0^{(M-1)} X_0^{(M-1)}(f) & a_1^{(M-1)} X_1^{(M-1)}(f) & \ldots & a_{K-1}^{(M-1)} X_{K-1}^{(M-1)}(f)
\end{bmatrix}, \]

Apply singular value decomposition to matrix A

\[ A(f) = U \Sigma V^* \]

Sensor related (M x M)  \hspace{5cm} \text{Power spectrum estimates (Diagonal)} \hspace{5cm} \text{Taper related (K x K)}
Thank You!