Stationary Graph Processes: Nonparametric Spectral Estimation

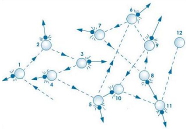
Santiago Segarra, Antonio G. Marques*, Geert Leus,
and Alejandro Ribeiro

*Dept. of Signal Theory and Communications
King Juan Carlos University - Madrid (SPAIN)
antonio.garcia.marques@urjc.es

ACK: Spanish MINECO TEC2013- 41604-R and USA NSF CCF-1217963.

IEEE SAM Workshop, Rio de Janeiro, July 11, 2016

- We frequently encounter stochastic processes
- Statistical signal processing
  - ⇒ Developed tools for their understanding



- Stationarity facilitates the analysis of random signals in time
  - ⇒ Statistical properties are time-invariant
- We seek to extend the concept of stationarity to graph processes
  - ⇒ Network data and irregular domains motivate this
  - ⇒ Lack of regularity lead to potentially multiple generalizations
- Classical SSP concepts and tools can be generalized: spectral estimation, average periodograms,...
  - ⇒ Better understanding and estimation of graph processes
  - ⇒ Related works: [Girault 15], [Perraudin 16]

Prelims 1: GSP Fundamentals

- Graph $G(\mathcal{V},\mathcal{E})$ with $N=|\mathcal{V}|$ nodes $\Rightarrow$ $\mathbf{A}$ and $\mathbf{L}=\mathbf{D}-\mathbf{A}$
- Graph signals mappings $x:\mathcal{V}\rightarrow\mathbb{R}$, represented as vectors $\mathbf{x}\in\mathbb{R}^N$
  - $\Rightarrow$ **As.:** Signal properties related to topology of G
- To understand GS $\Rightarrow$ Graph-shift operator $\mathbf{S}\in\mathbb{R}^{N\times N}$
  - $\Rightarrow$ Local $S_{ij}=0$ for $i\neq j$ and $(i,j)\notin\mathcal{E}$ $\Rightarrow$ Exs.: $\mathbf{A}$ or $\mathbf{L}$
  - $\Rightarrow$ Spectrum of $\mathbf{S}=\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$ (assume normality)
- Graph Fourier Transform (GFT) for signals: $\tilde{\mathbf{x}}=\mathbf{V}^H\mathbf{x}$
- Graph filters $H:\mathbb{R}^N\rightarrow\mathbb{R}^N$ are maps between graph signals
  - $\Rightarrow$ Polynomial in $\mathbf{S}$ with coefficients $\mathbf{h}\in\mathbb{R}^{L+1}$ $\Rightarrow$ $\mathbf{H}:=\sum_{l=0}^{L}h_l\mathbf{S}^l$
  - $\Rightarrow$ Diag. by $\mathbf{V}$ $\Rightarrow$ Freq. resp.: $\mathbf{H}=\mathbf{V}\mathrm{diag}(\tilde{\mathbf{h}})\mathbf{V}^H$
  - $\Rightarrow$ Local linear operators (good for modeling net dynamics)

Prelims 2: Weakly stationarity in time

(1) Correlation of stationary discrete time signals is invariant to shifts

$$\mathbf{C}_x := \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] = \mathbb{E}\left[x^H(n-l)_N \mathbf{x}(n-l)_N\right] = \mathbb{E}\left[\mathbf{S}^l\mathbf{x}(\mathbf{S}^l\mathbf{x})^H\right]$$

(2) Signal is the output of a LTI filter $\mathbf{H}$ excited with white noise $\mathbf{w}$

$$\mathbf{x} = \mathbf{H}\mathbf{w}, \quad \text{with } \mathbb{E}\left[\mathbf{w}\mathbf{w}^H\right] = \mathbf{I}$$

(3) The covariance matrix $\mathbf{C_x}$ is diagonalized by the Fourier matrix

$$\mathbf{C}_x = \mathbf{F}\text{diag}(\mathbf{p})\mathbf{F}^H$$

- The process has a power spectral density $\Rightarrow \mathbf{p} := \text{diag}\left(\mathbf{F}^H\mathbf{C}_x\mathbf{F}\right)$
- Each of these definitions can be generalized to graph signals

Stationary graph processes: Shifts

Definition (shift invariance)

Process $\mathbf{x}$ is weakly stationary with respect to $\mathbf{S}$ if and only if ($b > c$)

$$\mathbb{E}\left[\left(\mathbf{S}^{a}\mathbf{x}\right)\left(\left(\mathbf{S}^{H}\right)^{b}\mathbf{x}\right)^{H}\right] = \mathbb{E}\left[\left(\mathbf{S}^{a+c}\mathbf{x}\right)\left(\left(\mathbf{S}^{H}\right)^{b-c}\mathbf{x}\right)^{H}\right]$$

- Use a and b shifts as reference. Shift by c forward and backward
  - $\Rightarrow$ Signal is stationary if these shifts do not alter its covariance
- It reduces to $\mathbb{E}\left[\mathbf{x}\mathbf{x}^{H}\right] = \mathbb{E}\left[\mathbf{S}^{l}\mathbf{x}(\mathbf{S}^{l}\mathbf{x})^{H}\right]$ when $\mathbf{S}$ is a directed cycle
- Time shift is orthogonal, $\mathbf{S}^{H} = \mathbf{S}^{-1}$ ($a = 0$, $b = N$ and $c = l$)
- Need reference shifts because $\mathbf{S}$ can change energy of the signal

Definition (filtering of white noise)

Process $\mathbf{x}$ is weakly stationary with respect to $\mathbf{S}$ if it can be written as the output of linear shift invariant filter $\mathbf{H}$ with white input $\mathbf{w}$

$$\mathbf{x} = \mathbf{H}\mathbf{w}, \quad \text{with } \mathbb{E}\left[\mathbf{w}\mathbf{w}^H\right] = \mathbf{I}$$

- The filter $\mathbf{H}$ is linear shift invariant if $\Rightarrow \mathbf{H}(\mathbf{S}\mathbf{x}) = \mathbf{S}(\mathbf{H}\mathbf{x})$
- Equivalently, $\mathbf{H}$ polynomial on the shift operator $\Rightarrow \mathbf{H} = \sum_{l=0}^{L} h_l \mathbf{S}^l$
- Filter $\mathbf{H}$ determines color $\Rightarrow \mathbf{C}_\mathbf{x} = \mathbb{E}\left[(\mathbf{H}\mathbf{w})(\mathbf{H}\mathbf{w})^H\right] = \mathbf{H}\mathbf{H}^H$

Stationary graph processes: Diagonalization

Universidad Rey Juan Carlos

Definition (Simultaneous diagonalization)

Process $\mathbf{x}$ is weakly stationary with respect to $\mathbf{S}$ if the covariance $\mathbf{C_x}$ and the shift $\mathbf{S}$ are simultaneously diagonalizable

$$\mathbf{S} = \mathbf{V}\,\mathbf{\Lambda}\,\mathbf{V}^H \quad \Longrightarrow \quad \mathbf{C}_x = \mathbf{V}\,\text{diag}(\mathbf{p})\,\mathbf{V}^H$$

- Equivalent to time definition because $\mathbf{F}$ diagonalizes cycle graph
- The process has a power spectral density $\Rightarrow \mathbf{p} := \text{diag}\big(\mathbf{V}^H\mathbf{C}_x\mathbf{V}\big)$

Equivalence of definitions and PSD

- Have introduced three equally valid definitions of weak stationarity
  - ⇒ They are different but, pleasingly, equivalent

Theorem

Process $\mathbf{x}$ *has shift invariant correlation matrix* $\Leftrightarrow$ *it is the output of a linear shift invariant filter* $\Leftrightarrow$ *Covariance jointly diagonalizable with shift*

- Shift and Filtering ⇒ How stationary signals look like (local invariance)
- Simultaneous Diagonalization ⇒ A PSD exists ⇒ $\mathbf{p} := \text{diag}\big(\mathbf{V}^H \mathbf{C}_x \mathbf{V}\big)$
  - ⇒ The PSD collects the eigenvalues of $\mathbf{C}_x$.
  - ⇒ The PSD is nonnegative because $\mathbf{C}_\mathbf{x}$ is positive semidefinite

Weak stationary graph processes examples

Universidad Rey Juan Carlos

Example (White noise)

- White noise $\mathbf{w}$ is stationary in any graph shift $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$
- Covariance $\mathbf{C}_\mathbf{w} = \sigma^2\mathbf{I}$ simultaneously diagonalizable with all $\mathbf{S}$

Example (Covariance matrix graphs and Precision matrices)

- Every process is stationary in the graph defined by its covariance matrix
- If $\mathbf{S} = \mathbf{C}_\mathbf{x}$, shift $\mathbf{S}$ and covariance $\mathbf{C}_\mathbf{x}$ diagonalized by same basis
- Process is also stationary on precision matrix $\mathbf{S} = \mathbf{C}_\mathbf{x}^{-1}$

Example (Heat diffusion processes and ARMA processes)

- Heat diffusion process in a graph $\Rightarrow \mathbf{x} = \alpha_0(\mathbf{I} - \alpha\mathbf{L})^{-1}\mathbf{w}$
- Stationary in $\mathbf{L}$ since $\alpha_0(\mathbf{I} - \alpha\mathbf{L})^{-1}$ is a polynomial on $\mathbf{L}$
- Any autoregressive moving average (ARMA) process on a graph

Power spectral density examples

Universidad Rey Juan Carlos

Example (White noise)

- Power spectral density $\Rightarrow \mathbf{p} = \text{diag}\big(\mathbf{V}^H(\sigma^2\mathbf{I})\mathbf{V}\big) = \sigma^2\mathbf{1}$

Example (Covariance matrix graphs and Precision matrices)

- Power spectral density $\Rightarrow \mathbf{p} = \text{diag}\big(\mathbf{V}^H(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H)\mathbf{V}\big) = \text{diag}(\boldsymbol{\Lambda})$

Example (Heat diffusion processes and ARMA processes)

- Power spectral density $\Rightarrow \mathbf{p} = \text{diag}\Big[\alpha_0^2\,(\mathbf{I} - \alpha\boldsymbol{\Lambda})^{-2}\Big]$

PSD estimation with correlogram & periodogram

Universidad Rey Juan Carlos

- Given a process $\mathbf{x}$, the covariance of $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$ is given by

$$\mathbf{C}_{\tilde{\mathbf{x}}} := \mathbb{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\right] = \mathbb{E}\left[(\mathbf{V}^H\mathbf{x})(\mathbf{V}^H\mathbf{x})^H\right] = \text{diag}(\mathbf{p})$$

- Periodogram $\Rightarrow$ Given samples $\mathbf{x}_r$, average GFTs of samples

$$\hat{\mathbf{p}}_{\text{pg}} := \frac{1}{R}\sum_{r=1}^{R}\left|\tilde{\mathbf{x}}_r\right|^2 = \frac{1}{R}\sum_{r=1}^{R}\left|\mathbf{V}^H\mathbf{x}_r\right|^2$$

- Correlogram $\Rightarrow$ Replace $\mathbf{C}_x$ in PSD definition by sample covariance

$$\hat{\mathbf{p}}_{\text{cg}} := \text{diag}\left(\mathbf{V}^H\hat{\mathbf{C}}_x\mathbf{V}\right) := \text{diag}\left[\mathbf{V}^H\left[\frac{1}{R}\sum_{r=1}^{R}\mathbf{x}_r\mathbf{x}_r^H\right]\mathbf{V}\right]$$

- Periodogram and correlogram lead to identical estimates $\hat{\mathbf{p}}_{\text{pg}} = \hat{\mathbf{p}}_{\text{cg}}$

Accuracy of periodogram estimator

Theorem

If the process $\mathbf{x}$ *is Gaussian, periodogram estimates have bias and variance*

- *Bias* $\Rightarrow \mathbf{b}_{pg} := \mathbb{E}[\hat{\mathbf{p}}_{pg}] - \mathbf{p} = \mathbf{0}$
- *Variance* $\Rightarrow \mathbf{\Sigma}_{pg} := \mathbb{E}\left[(\hat{\mathbf{p}}_{pg} - \mathbf{p})(\hat{\mathbf{p}}_{pg} - \mathbf{p})^H\right] = \frac{2}{R}\text{diag}^2(\mathbf{p})$

- The periodogram is unbiased but the variance is not too good
  $\Rightarrow$ Quadratic in $\mathbf{p}$. Same as time processes
- Bias - Variance tradeoff $\Rightarrow$ Windows and Filterbanks

Windowed average periodogram

- Generalize Bartlett and Welch window methods to graphs
  - $\Rightarrow$ Consider windowed versions of a signal as separate realizations
- Given one realization $\mathbf{x}$ and bank of M windows $\mathcal{W} = \{\mathbf{w}_m\}_{m=1}^{M}$
  - $\Rightarrow$ Obtain the windowed average periodogram

$$\hat{\mathbf{p}}_{\mathcal{W}} := \frac{1}{M} \sum_{m=1}^{M} \left|\mathbf{V}^H \mathbf{x}_m\right|^2 = \frac{1}{M} \sum_{m=1}^{M} \left|\mathbf{V}^H \text{diag}(\mathbf{w}_m)\mathbf{x}\right|^2$$

- Similar to periodogram but windowed signals $\mathbf{x}_m$ are not independent
- Introduces bias but reduces variance compared to single observation

Accuracy of windowed periodogram

- Window PSD $\tilde{\mathbf{W}}_m := \mathbf{V}^H \text{diag}(\mathbf{w}_m)\mathbf{V}$ and spectrum mixing matrices

$$\tilde{\mathbf{W}}_{mm'} := \tilde{\mathbf{W}}_m \circ \tilde{\mathbf{W}}^*_{m'}$$

Theorem

For window bank $\mathcal{W} = \{\mathbf{w}_m\}_{m=1}^M$, the mean and variance of the windowed average periodogram $\hat{\mathbf{p}}_{\mathcal{W}}$ are

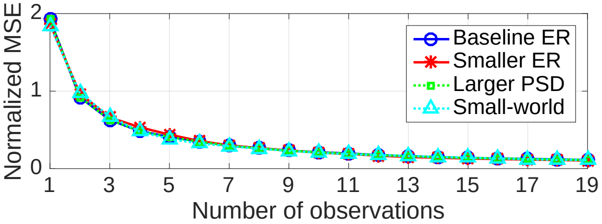
- *Mean* $\Rightarrow \mathbb{E}\left[\hat{\mathbf{p}}_{\mathcal{W}}\right] = \frac{1}{M}\sum_{m=1}^{M} \tilde{\mathbf{W}}_{mm}\mathbf{p}$
- *Variance* $\Rightarrow Tr[\mathbf{\Sigma}_{\mathcal{W}}] = \frac{2}{M^2}\sum_{m=1,m'=1}^{M} Tr\left[\left(\tilde{\mathbf{W}}_{mm'}\mathbf{p}\right)\left(\tilde{\mathbf{W}}_{mm'}\mathbf{p}\right)^H\right]$

- If windows are localized there is no (or minimal) spectrum mixing

Average periodogram

- MSE of periodogram as a function of the nr. of observations R

- Baseline ER random graph ($N = 100$ and $p = 0.05$) and $\mathbf{S} = \mathbf{A}$
- Observe filtered white Gaussian noise and estimate PSD



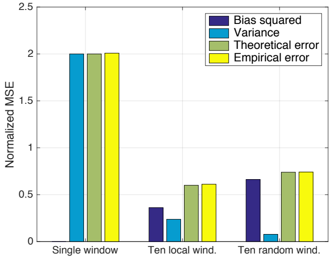
- Normalized MSE evolves as $2/R$ as expected

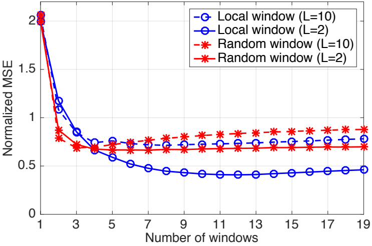
  $\Rightarrow$ Invariant to size, topology, and PSD

- Same behavior observed in non-Gaussian processes (theory not valid)

Windowed average periodogram

- Performance of local windows and random windows
- Block stochastic graph ($N = 100$, 10 communities) and small world
- Process filters white noise with different number of taps



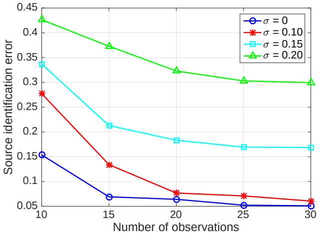


- The use of windows introduces bias but reduces total error (MSE)
- Local windows work better than random windows

  $\Rightarrow$ Advantage of local windows is larger for local processes

Opinion source identification

- Opinion diffusion in Zachary's karate club network ($N = 34$)
- Observed opinion $\mathbf{x}$ obtained by diffusing sparse white rumor $\mathbf{w}$
- Given $\{\mathbf{x}_r\}_{r=1}^R$ generated from unknown $\{\mathbf{w}_r\}_{r=1}^R$
  $\Rightarrow$ Diffused through filter of unknown nonnegative coefficients $\boldsymbol{\beta}$
- Goal $\Rightarrow$ Identify the support of each rumor $\mathbf{w}_r$
- First $\Rightarrow$ Estimate $\boldsymbol{\beta}$ from Moving Average PSD estimation
- Second $\Rightarrow$ Solve R sparse linear regressions to recover $\text{supp}(\mathbf{w}_r)$

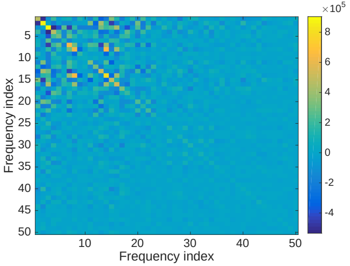


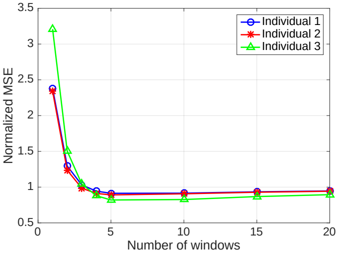
PSD of face images

Universidad Rey Juan Carlos

- PSD estimation for spectral signatures of faces of different people
- 100 grayscale face images $\{\mathbf{x}_i\}_{i=1}^{100} \in \mathbb{R}^{10304}$ (10 images $\times$ 10 people)

- Consider $\mathbf{x}_i$ as realization graph process that is Stationary on $\hat{\mathbf{C}}_{\mathbf{x}}$
- Construct $\hat{\mathbf{C}}_{\mathbf{x}}^{(j)} = \mathbf{V}^{(j)}\mathbf{\Lambda}_c^{(j)}\mathbf{V}^{H(j)}$ based on images of person j





- Process of person j approximately stationary in $\hat{\mathbf{C}}_{\mathbf{x}}$ (left)
- Use windowed average periodogram to estimate PSD of new face

Conclusion

Universidad Rey Juan Carlos

- Extended the notion of weak stationarity for graph processes
- Three definitions inspired in stationary time processes
  - ⇒ Shown all of them to be equivalent
- Defined power spectral density and studied its estimation
- Generalized classical non-parametric estimation methods
  - ⇒ Periodogram and correlogram where shown to be equivalent
  - ⇒ Windowed average periodogram leads to better estimate
- Extensions not described here
  - ⇒ Other non-parametric schemes: filter banks
  - ⇒ Parametric estimation: AR, MA, ARMA
  - ⇒ Space-time variation

Thanks!