Change-Point Detection of Gaussian Graph Signals with Partial Information

Yanxi Chen, Xianghui Mao, Dan Ling, Yuantao Gu

Department of Electronic Engineering
Tsinghua University

April 19, 2018

Outline

Backgrounds

Change-Point Detection

- Observe a sequence of signals $\mathbf{x}^t \in \mathbb{R}^N, t = 1, 2, \ldots$
- Change-point $T_c \geq 1$
- $t < T_c, \mathcal{H}_0 : \mathbf{x}^t \sim P_0$, with p.d.f. $f_0(\mathbf{x^t})$
- $t \geq T_c, \mathcal{H}_1 : \mathbf{x}^t \sim P_1$, with p.d.f. $f_1(\mathbf{x^t})$

Backgrounds

Change-Point Detection

- Observe a sequence of signals $\mathbf{x}^t \in \mathbb{R}^N, t = 1, 2, \ldots$
- Change-point $T_c \geq 1$
- $t < T_c, \mathcal{H}_0 : \mathbf{x}^t \sim P_0$, with p.d.f. $f_0(\mathbf{x^t})$
- $t \geq T_c, \mathcal{H}_1 : \mathbf{x}^t \sim P_1$, with p.d.f. $f_1(\mathbf{x^t})$

Cumulative Sum (CUSUM)

- Score $L^t : \mathbb{E}[L^t|\mathcal{H}_0] < 0, \mathbb{E}[L^t|\mathcal{H}_1] > 0$
- Log-likelihood ratio (LLR) $L^t = \log(f_1(\mathbf{x}^t)/f_0(\mathbf{x}^t))$
- Stopping time $T_s = \inf\{t > 0 : \max_{1 \leq i \leq t} \sum_{k=i}^{t} L^k \geq b\}$
- Recursive: $y^0 = 0,\ y^t = \max\{y^{t-1} + L^t, 0\},\ T_s = \inf\{t > 0 : y^t \geq b\}$

Backgrounds

Change-Point Detection

- Observe a sequence of signals $\mathbf{x}^t \in \mathbb{R}^N, t = 1, 2, \dots$
- Change-point $T_c \geq 1$
- $t < T_c, \mathcal{H}_0 : \mathbf{x}^t \sim P_0$, with p.d.f. $f_0(\mathbf{x^t})$
- $t \geq T_c, \mathcal{H}_1 : \mathbf{x}^t \sim P_1$, with p.d.f. $f_1(\mathbf{x^t})$

Cumulative Sum (CUSUM)

- Score $L^t : \mathbb{E}[L^t|\mathcal{H}_0] < 0, \mathbb{E}[L^t|\mathcal{H}_1] > 0$
- Log-likelihood ratio (LLR) $L^t = \log(f_1(\mathbf{x}^t)/f_0(\mathbf{x}^t))$
- Stopping time $T_s = \inf\{t > 0 : \max_{1\leq i\leq t} \sum_{k=i}^{t} L^k \geq b\}$
- Recursive: $y^0 = 0$, $y^t = \max\{y^{t-1} + L^t, 0\}$, $T_s = \inf\{t > 0 : y^t \geq b\}$

Performance: Average Running Length (ARL)

- $\mathrm{ARL}_0 = \mathbb{E}[T_s|T_c = \infty]$: false-alarm rate $1/\mathrm{ARL}_0$
- $\mathrm{ARL}_1 = \mathbb{E}[T_s|T_c = 1]$: detection delay

Backgrounds

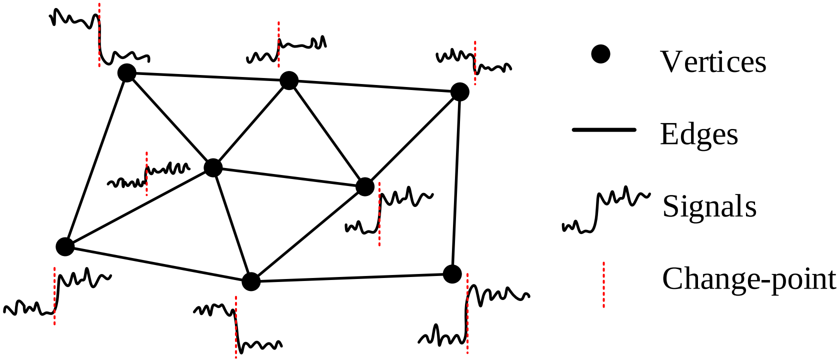


Figure: Change-point detection of Gaussian graph signals.

Backgrounds

Graph Signal Processing (GSP)

- Graph $G = (V, E), |V| = N$; signal $\mathbf{x} \in \mathbb{R}^N$
- Adjacent matrix $\mathbf{A} \in \{0, 1\}^{N \times N}$
- Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is diagonal, $D_{ii} = \sum_{j=1}^{N} A_{ji}$
- Eigen-decomposition $\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$, $\mathrm{diag}\{\mathbf{\Lambda}\}$ sorted in ascend
- Fourier transform $\hat{\mathbf{x}} = \mathbf{V}^{\mathrm{T}}\mathbf{x}$, inverse transform $\mathbf{x} = \mathbf{V}\hat{\mathbf{x}}$
- K-bandlimited (smoothness): $\hat{x}_i = 0, \forall i \in \{K + 1, \dots, N\}$

Problem Formulation and Algorithms

Problem Formulation

- $t < T_C$: $\mathbf{x}^t \sim \mathcal{N}(\boldsymbol{\mu}_0, \sigma^2 \mathbf{I}_N)$, $\mathbf{x}^t = \boldsymbol{\mu}_0 + \mathbf{e}^t$
- $t \geq T_C$: $\mathbf{x}^t \sim \mathcal{N}(\boldsymbol{\mu}_1, \sigma^2 \mathbf{I}_N)$, $\mathbf{x}^t = \boldsymbol{\mu}_1 + \mathbf{e}^t$

Problem Formulation and Algorithms

Problem Formulation

- $t < T_C$: $\mathbf{x}^t \sim \mathcal{N}(\boldsymbol{\mu}_0, \sigma^2 \mathbf{I}_N)$, $\mathbf{x}^t = \boldsymbol{\mu}_0 + \mathbf{e}^t$
- $t \geq T_C$: $\mathbf{x}^t \sim \mathcal{N}(\boldsymbol{\mu}_1, \sigma^2 \mathbf{I}_N)$, $\mathbf{x}^t = \boldsymbol{\mu}_1 + \mathbf{e}^t$

Assumptions

- $\boldsymbol{\mu}_0, \sigma^2$ are known, but $\boldsymbol{\mu}_1$ is unknown
- $\boldsymbol{\mu}_0$ is K-bandlimited, while $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \mathbf{V}\hat{\boldsymbol{\mu}}_h$ is full-band

Problem Formulation and Algorithms

Problem Formulation

- $t < T_C$: $\mathbf{x}^t \sim \mathcal{N}(\boldsymbol{\mu}_0, \sigma^2 \mathbf{I}_N)$, $\mathbf{x}^t = \boldsymbol{\mu}_0 + \mathbf{e}^t$
- $t \geq T_C$: $\mathbf{x}^t \sim \mathcal{N}(\boldsymbol{\mu}_1, \sigma^2 \mathbf{I}_N)$, $\mathbf{x}^t = \boldsymbol{\mu}_1 + \mathbf{e}^t$

Assumptions

- $\boldsymbol{\mu}_0, \sigma^2$ are known, but $\boldsymbol{\mu}_1$ is unknown
- $\boldsymbol{\mu}_0$ is K-bandlimited, while $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \mathbf{V}\hat{\boldsymbol{\mu}}_h$ is full-band

Log-likelihood ratio $L^t = \log \frac{f_1(\mathbf{x}^t)}{f_0(\mathbf{x}^t)} = \frac{\|\mathbf{e}^t\|_2^2 - \|\mathbf{e}^t - \mathbf{V}\hat{\boldsymbol{\mu}}_h\|_2^2}{2\sigma^2}$

- Problem: $\hat{\boldsymbol{\mu}}_h$ is unknown

Problem Formulation and Algorithms

Problem Formulation

- $t < T_C$: $\mathbf{x}^t \sim \mathcal{N}(\boldsymbol{\mu}_0, \sigma^2 \mathbf{I}_N)$, $\mathbf{x}^t = \boldsymbol{\mu}_0 + \mathbf{e}^t$
- $t \geq T_C$: $\mathbf{x}^t \sim \mathcal{N}(\boldsymbol{\mu}_1, \sigma^2 \mathbf{I}_N)$, $\mathbf{x}^t = \boldsymbol{\mu}_1 + \mathbf{e}^t$

Assumptions

- $\boldsymbol{\mu}_0, \sigma^2$ are known, but $\boldsymbol{\mu}_1$ is unknown
- $\boldsymbol{\mu}_0$ is K-bandlimited, while $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \mathbf{V}\hat{\boldsymbol{\mu}}_h$ is full-band

Log-likelihood ratio $L^t = \log \frac{f_1(\mathbf{x}^t)}{f_0(\mathbf{x}^t)} = \frac{\|\mathbf{e}^t\|_2^2 - \|\mathbf{e}^t - \mathbf{V}\hat{\boldsymbol{\mu}}_h\|_2^2}{2\sigma^2}$

- Problem: $\hat{\boldsymbol{\mu}}_h$ is unknown

Solution

- Maximization: $L^t = \max_{\hat{\boldsymbol{\mu}}_h} \frac{\|\mathbf{e}^t\|_2^2 - \|\mathbf{e}^t - \mathbf{V}\hat{\boldsymbol{\mu}}_h\|_2^2}{2\sigma^2} = \frac{\|\hat{\mathbf{e}}_h^t\|_2^2}{2\sigma^2}$
- Correction: $L^t = \frac{\|\hat{\mathbf{e}}_h^t\|_2^2}{2\sigma^2} - \frac{N-K}{2} - \delta$, $\mathbb{E}[L^t | \mathcal{H}_0] = -\delta < 0$

Problem Formulation and Algorithms

Algorithm

- Parameters: $N, K, \boldsymbol{\mu}_0, \sigma^2, b, \delta$
- Initialize: $y^0 = 0$; estimate $\boldsymbol{\mu}_0$ and σ^2 from historical data
- Projection: $\mathbf{r} \leftarrow \mathbf{x}^t - \boldsymbol{\mu}_0$, $\hat{\mathbf{r}} \leftarrow \mathbf{V}^T\mathbf{r}$, $\hat{\mathbf{r}}_h \leftarrow \hat{\mathbf{r}}_{K+1:N}$
- Maximization and correction: $L^t \leftarrow \frac{\|\hat{\mathbf{r}}_h\|_2^2}{2\sigma^2} - \frac{N-K}{2} - \delta$
  - $\mathbb{E}[L^t|\mathcal{H}_0] = -\delta < 0, \mathbb{E}[L^t|\mathcal{H}_1] = \frac{\|\hat{\boldsymbol{\mu}}_h\|_2^2}{2\sigma^2} - \delta > 0$
- Recursive CUSUM: $y^t \leftarrow \max\{y^{t-1} + L^t, 0\}$
- Inference: if $y^t \geq b$, then stopping time $T_s \leftarrow t$, detection is done

Problem Formulation and Algorithms

Algorithm

- Parameters: $N, K, \boldsymbol{\mu}_0, \sigma^2, b, \delta$
- Initialize: $y^0 = 0$; estimate $\boldsymbol{\mu}_0$ and σ^2 from historical data
- Projection: $\mathbf{r} \leftarrow \mathbf{x}^t - \boldsymbol{\mu}_0$, $\hat{\mathbf{r}} \leftarrow \mathbf{V}^T\mathbf{r}$, $\hat{\mathbf{r}}_h \leftarrow \hat{\mathbf{r}}_{K+1:N}$
- Maximization and correction: $L^t \leftarrow \frac{\|\hat{\mathbf{r}}_h\|_2^2}{2\sigma^2} - \frac{N-K}{2} - \delta$
  - $\mathbb{E}[L^t|\mathcal{H}_0] = -\delta < 0, \mathbb{E}[L^t|\mathcal{H}_1] = \frac{\|\hat{\boldsymbol{\mu}}_h\|_2^2}{2\sigma^2} - \delta > 0$
- Recursive CUSUM: $y^t \leftarrow \max\{y^{t-1} + L^t, 0\}$
- Inference: if $y^t \geq b$, then stopping time $T_s \leftarrow t$, detection is done

Extension: both $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1$ are arbitrary (no *a priori* on bandwidth)

- Trick: w.l.o.g. set $\boldsymbol{\mu}_0 = \mathbf{0}$, then bandwidth $K = 0$
- $L^t = \frac{\|\hat{\mathbf{r}}_h\|_2^2}{2\sigma^2} - \frac{N-K}{2} - \delta = \frac{\|\mathbf{r}\|_2^2}{2\sigma^2} - \frac{N}{2} - \delta$

Problem Formulation and Algorithms

Algorithm

- Parameters: $N, K, \boldsymbol{\mu}_0, \sigma^2, b, \delta$
- Initialize: $y^0 = 0$; estimate $\boldsymbol{\mu}_0$ and σ^2 from historical data
- Projection: $\mathbf{r} \leftarrow \mathbf{x}^t - \boldsymbol{\mu}_0$, $\hat{\mathbf{r}} \leftarrow \mathbf{V}^T\mathbf{r}$, $\hat{\mathbf{r}}_h \leftarrow \hat{\mathbf{r}}_{K+1:N}$
- Maximization and correction: $L^t \leftarrow \frac{\|\hat{\mathbf{r}}_h\|_2^2}{2\sigma^2} - \frac{N-K}{2} - \delta$
  - $\mathbb{E}[L^t|\mathcal{H}_0] = -\delta < 0, \mathbb{E}[L^t|\mathcal{H}_1] = \frac{\|\hat{\boldsymbol{\mu}}_h\|_2^2}{2\sigma^2} - \delta > 0$
- Recursive CUSUM: $y^t \leftarrow \max\{y^{t-1} + L^t, 0\}$
- Inference: if $y^t \geq b$, then stopping time $T_s \leftarrow t$, detection is done

Extension: both $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1$ are arbitrary (no *a priori* on bandwidth)

- Trick: w.l.o.g. set $\boldsymbol{\mu}_0 = \mathbf{0}$, then bandwidth $K = 0$
- $L^t = \frac{\|\hat{\mathbf{r}}_h\|_2^2}{2\sigma^2} - \frac{N-K}{2} - \delta = \frac{\|\mathbf{r}\|_2^2}{2\sigma^2} - \frac{N}{2} - \delta$

Further extension: noise variance σ_i^2 for i-th vertex

- $L^t = \sum_{i=1}^N \frac{r_i^2}{2\sigma_i^2} - \frac{N}{2} - \delta, \mathbb{E}[L^t|\mathcal{H}_1] = \sum_{i=1}^N \frac{\mu_{1i}^2}{2\sigma_i^2} - \delta$

Problem Formulation and Algorithms

Distributed algorithm
(no fusion center; each vertex only communicates with its neighbors)

- Parameters: N, σ^2, b, δ; $\mathbf{W} : \sum_{u \in N(v)} W_{vu} = 1, \forall v$
- Initialize: $y_v^0 = 0, z_v^0 = 0, \forall v \in \{1, 2, \ldots, N\}$
- Maximization and correction: $L_v^t \leftarrow \frac{|x_v^t|^2}{2\sigma^2} - \frac{1}{2} - \delta$
- Local CUSUM: $o_v^t \leftarrow y_v^{t-1}$, $y_v^t \leftarrow \max\{y_v^{t-1} + L_v^t, 0\}$
- Communication: $z_v^t \leftarrow \sum_{u \in N(v)} W_{vu}(z_u^{t-1} + y_u^t - o_u^t)$
- Inference: if $\max_v z_v^t \geq b$, then $T_s \leftarrow t$, detection is done

Problem Formulation and Algorithms

Discussions

- Performance analysis: sub-exponential distribution of L^t under both $\mathcal{H}_0$ and $\mathcal{H}_1 \Rightarrow \mathrm{ARL}_0 \sim \exp(b), \mathrm{ARL}_1 \approx \frac{b}{\mathbb{E}[L^t|\mathcal{H}_1]}$

Discussions

- Performance analysis: sub-exponential distribution of L^t under both $\mathcal{H}_0$ and $\mathcal{H}_1 \Rightarrow \mathrm{ARL}_0 \sim \exp(b), \mathrm{ARL}_1 \approx \frac{b}{\mathbb{E}[L^t|\mathcal{H}_1]}$
- Advantages of algorithm: as efficient as pure CUSUM (detection *without* estimation); can handle time-varying post-change distribution parameters

Problem Formulation and Algorithms

Discussions

- Performance analysis: sub-exponential distribution of L^t under both $\mathcal{H}_0$ and $\mathcal{H}_1 \Rightarrow \mathrm{ARL}_0 \sim \exp(b), \mathrm{ARL}_1 \approx \frac{b}{\mathbb{E}[L^t|\mathcal{H}_1]}$
- Advantages of algorithm: as efficient as pure CUSUM (detection *without* estimation); can handle time-varying post-change distribution parameters
- Role of GSP: *a priori* knowledge of $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1$ can improve performance in ARL

Problem Formulation and Algorithms

Discussions

- Performance analysis: sub-exponential distribution of L^t under both $\mathcal{H}_0$ and $\mathcal{H}_1 \Rightarrow \mathrm{ARL}_0 \sim \exp(b), \mathrm{ARL}_1 \approx \frac{b}{\mathbb{E}[L^t|\mathcal{H}_1]}$
- Advantages of algorithm: as efficient as pure CUSUM (detection *without* estimation); can handle time-varying post-change distribution parameters
- Role of GSP: *a priori* knowledge of $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1$ can improve performance in ARL
- Gaussian assumption: not necessary for algorithms to work $\left(L^t = \frac{\|\mathbf{r}\|_2^2}{2\sigma^2} - \frac{N}{2} - \delta\right)$

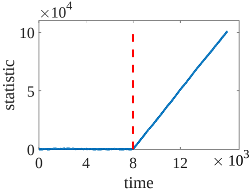
Problem Formulation and Algorithms

Discussions

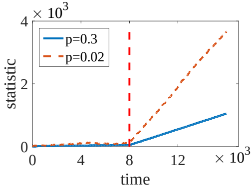
- Performance analysis: sub-exponential distribution of L^t under both $\mathcal{H}_0$ and $\mathcal{H}_1 \Rightarrow \mathrm{ARL}_0 \sim \exp(b), \mathrm{ARL}_1 \approx \frac{b}{\mathbb{E}[L^t|\mathcal{H}_1]}$
- Advantages of algorithm: as efficient as pure CUSUM (detection *without* estimation); can handle time-varying post-change distribution parameters
- Role of GSP: *a priori* knowledge of $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1$ can improve performance in ARL
- Gaussian assumption: not necessary for algorithms to work ($L^t = \frac{\|\mathbf{r}\|_2^2}{2\sigma^2} - \frac{N}{2} - \delta$)
- Reflection: obtain a qualified CUSUM score L^t based on (but beyond) log-likelihood ratio
  - Maximization and correction: for unknown post-change parameter θ_1, $L^t = \max_{\theta_1} \log \frac{f_1(\mathbf{x}^t|\theta_1)}{f_0(\mathbf{x}^t)} - C$ such that $\mathbb{E}[L^t|\mathcal{H}_0] < 0$ and $\mathbb{E}[L^t|\mathcal{H}_1] > 0$

Experiments

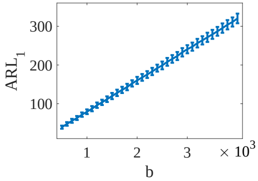
Synthetic data: random graph with $N = 100$, edge probability $p = 0.3$, $\boldsymbol{\mu}_0 = \mathbf{0}, \|\boldsymbol{\mu}_1\| = 1, \sigma = 0.2$



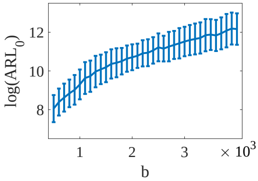
(a) y^t (centralized)



(b) $\max_v z_v^t$ (distributed)



(c) ARL_1-b (centralized)

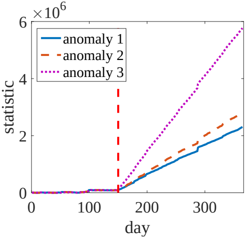


(d) $\log(\mathrm{ARL}_0)$-b (centralized)

Real-world data: Manhattan taxi pickup in 2014 and 2015, $N = 13679$

- Estimate $\boldsymbol{\mu}_0$ and $\{\sigma_i^2\}$ from data of 2014
- Simulate small, additive anomalies in data of 2015 after $T_c = 150$
  1. Add a constant 5 to the 112 green vertices
  2. Add an increament $\sim \mathrm{Uniform}\{1, 2, \ldots, 9\}$ to green vertices, *i.i.d.* among time steps and vertices
  3. Add an increament $\sim \mathrm{Uniform}\{1, 2, 3, 4\}$ to 112 randomly chosen vertices, *i.i.d.* among time steps and vertices

(a) Manhattan road map.



(b) Statistic y^t.

Summary

Contributions

- Obtain a qualified CUSUM score L^t via maximization and correction of log-likelihood ratio: efficient and practical; can handle time-varying post-change distribution parameter
- Centralized and distributed algorithms
- Utilize graph structure to improve performance

Summary

Contributions

- Obtain a qualified CUSUM score L^t via maximization and correction of log-likelihood ratio: efficient and practical; can handle time-varying post-change distribution parameter
- Centralized and distributed algorithms
- Utilize graph structure to improve performance

Thank you for your attention!