First-Order Optimal Sequential Subspace Change-Point Detection

Liyan Xie*

Joint work with **Yao Xie*** and **George Moustakides**[†]

*Georgia Institute of Technology, [†]Rutgers University

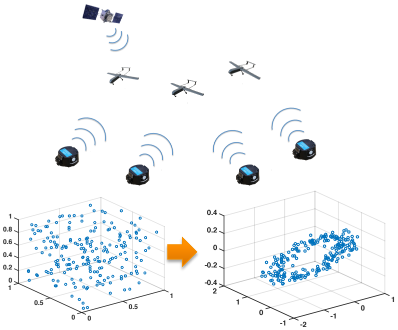
November 29, 2018

Present at GlobalSIP, 2018

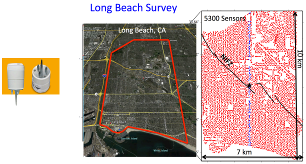
Outline

- Motivation
- Optimal: Exact CUSUM procedure
- Practical: Subspace-CUSUM procedure
- Theoretical result: first-order optimality
- Summary

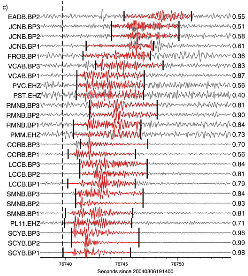
Motivation



Swarm behavior change



dense geophysical sensor array

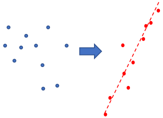


Seismic tremor signal detection

Problem Setup

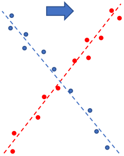
- The *emerging subspace* problem:

$$x_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I_p), \qquad t = 1, 2, \ldots, \tau,$$
$$x_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I_p + \theta u u^\intercal), \quad t = \tau + 1, \ldots$$



- The *switching subspace* problem:

$$x_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I_p + \theta u_1 u_1^\intercal), \quad t = 1, 2, \ldots, \tau,$$
$$x_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I_p + \theta u_2 u_2^\intercal), \quad t = \tau + 1, \ldots$$



- known noise level σ^2
- unknown change-point location τ
- unknown θ, u

Equivalence

Switching subspace is equivalent to emerging subspace.

$\exists Q \in \mathbb{R}^{(p-1)\times p}$ in the orthogonal space to u_1, s.t.,

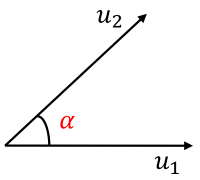
$$Qu_1 = 0, \quad QQ^\intercal = I_{p-1}.$$

Let $y_t = Qx_t$, and $\tilde{u} = Qu_2 / \|Qu_2\|$:

$$\begin{aligned} y_t &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I_{p-1}), & t &= 1, 2, \ldots, \tau, \\ y_t &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I_{p-1} + \tilde{\theta}\tilde{u}\tilde{u}^\intercal), & t &= \tau+1, \tau+2, \ldots \end{aligned}$$

where

$$\begin{aligned} \tilde{\theta} &= \theta \|Qu_2\|^2 = \theta[1 - (u_1^\intercal u_2)^2] \\ &= \theta \sin^2 \alpha. \end{aligned}$$



Related work

- Spiked covariance model (Johnstone, 2001)

$$\Sigma = I + \theta V V^\intercal$$

a small number of directions explain most of the variance.

- Fixed-sample test (Berthet and Rigollet, 2013a, 2013b)

$$\begin{aligned} H_0 : &\quad x_i \overset{iid}{\sim} \mathcal{N}(0, I) \\ H_1 : &\quad x_i \overset{iid}{\sim} \mathcal{N}(0, I + \theta u u^\intercal) \end{aligned}$$

prove sparse eigenvalue statistic is minimax optimal test.

Outline

- Motivation
- Optimal: Exact CUSUM procedure
- Practical: Subspace-CUSUM procedure
- Theoretical result: first-order optimality
- Summary

Optimal: Exact CUSUM based on likelihood ratio

- CUSUM procedure:

$$T_{\mathrm{C}} = \inf \left\{ t : \max_{k \le t} \sum_{i=k+1}^{t} \log \frac{f_0(x_i)}{f_\infty(x_i)} > b \right\}.$$

- Recursive implementation of detection statistic

$$S_t = \max \left\{ S_{t-1} + \log \frac{f_0(x_t)}{f_\infty(x_t)}, 0 \right\}, \quad S_0 = 0.$$

- Given **all parameters known**, CUSUM is **optimal** (Lorden 1971) (Moustakides 1986).

Exact CUSUM

- Derive the log-likelihood ratio:

$$\log \frac{f_0(x_t)}{f_\infty(x_t)} = \frac{1}{2\sigma^2}\frac{\rho}{1+\rho}\Big\{(u^\intercal x_t)^2 - \sigma^2(1+\frac{1}{\rho})\log(1+\rho)\Big\}.$$

  SNR: $\rho = \theta/\sigma^2$.

- Known subspace vector u and SNR θ.
- CUSUM statistics

$$S_t = (S_{t-1})^+ + (u^\intercal x_t)^2 - \sigma^2(1+\frac{1}{\rho})\log(1+\rho).$$

- Exact CUSUM procedure

$$T_{\mathrm{C}} = \inf\{t : S_t > b\}.$$

Practical: Subspace-CUSUM

- Parameters need to be estimated:

subspace $\hat{u}_t$, and drift d.

$$\mathcal{S}_t = (\mathcal{S}_{t-1})^+ + (\hat{u}_t{}^\intercal x_t)^2 - \underbrace{\sigma^2(1+\frac{1}{\rho})\log(1+\rho)}_{d}.$$

- Subspace-CUSUM procedure

$$\mathcal{T}_{\mathrm{C}} = \inf\{t : \mathcal{S}_t > b\}.$$

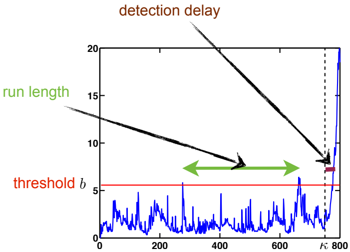
- where $\hat{u}_t$ is estimated sequentially: $\hat{u}_t \leftarrow \hat{u}_{t-1}$.
- time window $(t+1, t+w)$ (independence between $\hat{u}_t$ and x_t).
- Subspace tracking: (Balzano et al. 2010) (Chi et al. 2011).

Performance metrics

- average run length (ARL):

$$\mathbb{E}_\infty(T)$$

- worst-case expected detection delay (Lorden, 1971):

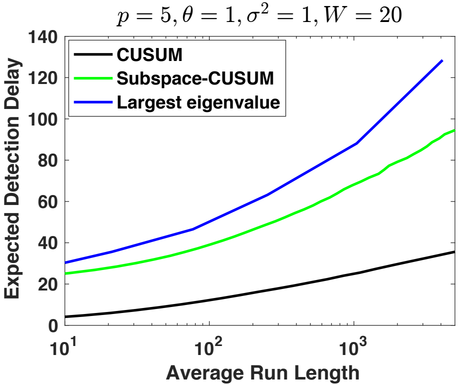


$$\text{EDD} = \sup_k \text{ess sup } \mathbb{E}_k((T-k+1)^+ | \mathcal{F}_{k-1})$$

Commonly used approximation (when the change happens at the first moment)

$$\text{EDD} = \mathbb{E}_0(T)$$

Numerical comparison



exact CUSUM $>$ Subspace-CUSUM $\gg$ largest eigenvalue

Main theoretical result

Theorem 1 (Xie, Moustakides, X., 2018)

The "practical" subspace CUSUM is nearly optimal.

- Subspace-CUSUM procedure

$$\mathcal{S}_t = (\mathcal{S}_{t-1})^+ + (\hat{u}_t{}^\intercal x_t)^2 - d$$

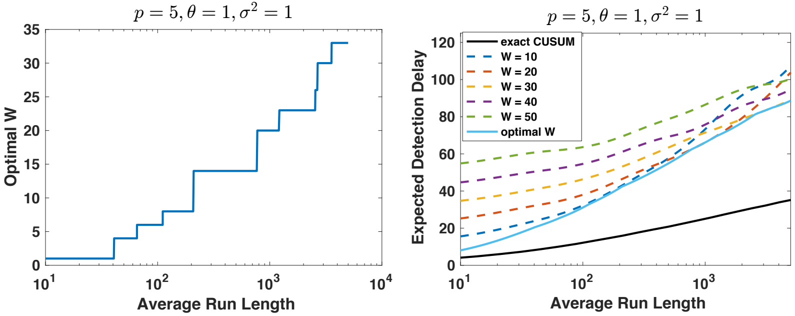
$$\mathcal{T}_{\rm C} = \inf\{t : \mathcal{S}_t > b\}$$

where $\hat{u}_t$ is estimated using samples in a window $(t, t+w)$.

- Proof sketch: find optimal d^* and window w^* to minimize EDD given constant ARL. Then the resulted Subspace CUSUM is nearly optimal.

Optimal w^*

- Choice of w involves a tradeoff in the estimation accuracy and the detection delay.



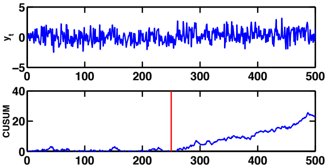
Optimal d

- Important property of exact CUSUM statistic:

$$\underbrace{\mathbb{E}_\infty\left[\log\frac{f_0(x_i)}{f_\infty(x_i)}\right] < 0}_{\text{No change}} \quad vs. \quad \underbrace{\mathbb{E}_0\left[\log\frac{f_0(x_i)}{f_\infty(x_i)}\right] > 0}_{\text{Exist change}}$$

- choice of drift:

$$\underbrace{\mathbb{E}_\infty[(\hat{u}_t{}^\intercal x_t)^2] - d < 0}_{\text{No change}} \quad vs. \quad \underbrace{\mathbb{E}_0[(\hat{u}_t{}^\intercal x_t)^2] - d > 0}_{\text{Exist change}}$$



- Using CLT

$$\mathbb{E}_\infty[(\hat{u}_t^\intercal x_t)^2] = \sigma^2, \quad \mathbb{E}_0[(\hat{u}_t^\intercal x_t)^2] = \sigma^2(1+\rho)[1 - \frac{p-1}{w\rho}]$$

ARL and EDD

Subspace-CUSUM:

$$S_t = (S_{t-1})^+ + (\hat{u}_t^\intercal x_t)^2 - d$$

Goal: Minimize EDD given constant ARL.

- ARL and EDD of **exact CUSUM**

Lemma 1 (Siegmund 1985)

$$\mathbb{E}_\infty(T) = \frac{e^b}{I_\infty}(1 + o(1)), \quad \mathbb{E}_0(T) = \frac{b}{I_0}(1 + o(1))$$

K-L divergences:

$$I_\infty = \mathbb{E}_\infty \left[\frac{\log f_\infty(x)}{\log f_0(x)} \right], \quad I_0 = \mathbb{E}_0 \left[\frac{\log f_0(x)}{\log f_\infty(x)} \right]$$

- However, $(\hat{u}_t^\intercal x_t)^2 - d$ is not exactly a log-likelihood ratio.

Proof to Theorem 1

- Equalizer trick
- Introducing "equalizer" δ_∞

$$\mathbb{E}_\infty[e^{\delta_\infty[(\hat{u}_t^\intercal x_t)^2 - d]}] = 1.$$

After equalizing, red term $\approx$ a log-likelihood ratio.

- Represent d using δ_∞

$$d = -\frac{1}{2\delta_\infty}\log(1 - 2\sigma^2\delta_\infty).$$

Proof to Theorem 1

- Set constant ARL $= \gamma$, the detection delay

$$\mathbb{E}_0\left[T\right] = \frac{\log(\gamma)(1+o(1))}{\tilde{\delta}_\infty(1+\rho)\left(1-\frac{p-1}{w\rho}\right)+\frac{1}{2}\log(1-2\tilde{\delta}_\infty)} + w.$$

- For each window size w, the optimal drift d which minimizes the EDD is

$$d^* = \sigma^2\left[\frac{(1+\rho)\left(1-\frac{p-1}{w\rho}\right)}{(1+\rho)\left(1-\frac{p-1}{w\rho}\right)-1}\log\left[(1+\rho)\left(1-\frac{p-1}{w\rho}\right)\right]\right]$$

- Plug d^* back to EDD, we can derive the optimal window size w

$$w^* = \sqrt{\log\gamma}\cdot\frac{\sqrt{2(k-1)}}{\rho-\log(1+\rho)}(1+o(1)).$$

First-order optimality of Subspace-CUSUM

Theorem 1 (Xie, Moustakides, X., 2018)

When ρ, σ^2 known, u unknown, for any ARL $= \gamma$ ($\gamma > 0$), EDD of Subspace-CUSUM using the optimal drift and optimal window size is

$$\mathbb{E}_0\left[T\right] = \frac{2\log(\gamma)}{\rho - \log(1+\rho)}\left(1 + o(1)\right),$$

which matches the first-order EDD of the exact CUSUM.

Summary

- Exact and practical approaches to detecting low-rank changes
- Subspace-CUSUM is **first-order optimal**
  - Optimal dift
  - Optimal window $w^* = \mathcal{O}(\sqrt{\log \gamma})$
- Ongoing: First-order optimality in **general** settings

arXiv:1806.10760. First-order optimal sequential subspace change-point detection. Liyan Xie, George V. Moustakides, Yao Xie.