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

• Title: Sequential change-point detection for structured high-dimensional streaming data
• Objective: Developing new computationally efficient and statistically powerful algorithms to detecting changes online.
  - Subspace structure
  - Partially observable data
  - Data dynamic
  - Distributed processing
• Collaborators: Dr. Matthew Berger, Dr. Lee Seversky, Dr. Lauren Hue-Seversky
  e.g. Swarm behavior change detection

Swarm behavior change

Seismic tremor signal detection

dense geophysical sensor array
Problem Setup

- **The emerging subspace problem:**
  
  \[ \begin{align*}
  x_t & \sim \mathcal{N}(0, \sigma^2 I_p), \quad t = 1, 2, \ldots, \tau, \\
  x_t & \sim \mathcal{N}(0, \sigma^2 I_p + \theta u u^\top), \quad t = \tau + 1, \ldots
  \end{align*} \]

- **The switching subspace problem:**
  
  \[ \begin{align*}
  x_t & \sim \mathcal{N}(0, \sigma^2 I_p + \theta u_1 u_1^\top), \quad t = 1, 2, \ldots, \tau, \\
  x_t & \sim \mathcal{N}(0, \sigma^2 I_p + \theta u_2 u_2^\top), \quad t = \tau + 1, \ldots
  \end{align*} \]

- known noise level \(\sigma^2\)
- unknown change-point location \(\tau\)
- unknown \(\theta, u\)
Equivalence

Switching subspace is equivalent to emerging subspace.

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

\[ Qu_1 = 0, \quad QQ^\top = I_{p-1}. \]

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

\[
y_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I_{p-1}), \quad t = 1, 2, \ldots, \tau,
\]

\[
y_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I_{p-1} + \tilde{\theta} \tilde{u} \tilde{u}^\top), \quad t = \tau + 1, \tau + 2, \ldots
\]

where

\[
\tilde{\theta} = \theta \|Qu_2\|^2 = \theta [1 - (u_1^\top u_2)^2]
\]

\[ = \theta \sin^2 \alpha. \]
Related work

- **Spiked covariance model** (Johnstone, 2001)

  \[ \Sigma = I + \theta V V^\top \]

  a small number of directions explain most of the variance.

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

  \[
  \begin{align*}
  H_0 & : \quad x_i \overset{iid}{\sim} \mathcal{N}(0, I) \\
  H_1 & : \quad x_i \overset{iid}{\sim} \mathcal{N}(0, I + \theta uu^\top)
  \end{align*}
  \]

  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:

\[
T_C = \inf \left\{ t : \max_{k \leq t} \sum_{i=k+1}^{t} \log \frac{f_0(x_i)}{f_\infty(x_i)} > b \right\}.
\]

- 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:
  \[
  \log \frac{f_0(x_t)}{f_\infty(x_t)} = \frac{1}{2\sigma^2} \frac{\rho}{1 + \rho} \left\{ (u^T x_t)^2 - \sigma^2 \left(1 + \frac{1}{\rho}\right) \log(1 + \rho) \right\}.
  \]

  SNR: \( \rho = \theta/\sigma^2 \).

- Known subspace vector \( u \) and SNR \( \theta \).

- CUSUM statistics
  \[
  S_t = (S_{t-1})^+ + (u^T x_t)^2 - \sigma^2 \left(1 + \frac{1}{\rho}\right) \log(1 + \rho).
  \]

- Exact CUSUM procedure
  \[
  T_C = \inf\{t : S_t > b\}.
  \]
Practical: Subspace-CUSUM

- Parameters need to be estimated:

  \[ S_t = (S_{t-1})^+ + (\hat{u}_t^T x_t)^2 - \sigma^2 (1 + \frac{1}{\rho}) \log(1 + \rho). \]

- Subspace-CUSUM procedure

  \[ \mathcal{T}_C = \inf\{t : S_t > b\}. \]

- 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):**

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

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

  \[ \text{EDD} = \mathbb{E}_0(T) \]
Numerical comparison

$p = 5, \theta = 1, \sigma^2 = 1, W = 20$

exact CUSUM $> \text{Subspace-CUSUM} \gg \text{largest eigenvalue}$
Main theoretical result

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

The “practical” subspace CUSUM is nearly optimal.

Subspace-CUSUM procedure

\[
S_t = (S_{t-1})^+ + (\hat{u}_t^T x_t)^2 - d
\]

\[
T_C = \inf\{t : S_t > b\}
\]

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.

\[ p = 5, \theta = 1, \sigma^2 = 1 \]
Optimal $d$

- Important property of exact CUSUM statistic:
  \[
  \mathbb{E}_{\infty} \left[ \log \frac{f_0(x_i)}{f_{\infty}(x_i)} \right] < 0 \quad \text{vs.} \quad \mathbb{E}_0 \left[ \log \frac{f_0(x_i)}{f_{\infty}(x_i)} \right] > 0
  \]
  No change \quad \text{Exist change}

- Choice of drift:
  \[
  \mathbb{E}_{\infty}[(\hat{u}_t^T x_t)^2] - d < 0 \quad \text{vs.} \quad \mathbb{E}_0[(\hat{u}_t^T x_t)^2] - d > 0
  \]
  No change \quad \text{Exist change}

- Using CLT
  \[
  \mathbb{E}_{\infty}[(\hat{u}_t^T x_t)^2] = \sigma^2, \quad \mathbb{E}_0[(\hat{u}_t^T x_t)^2] = \sigma^2(1 + \rho)[1 - \frac{p - 1}{w\rho}]
  \]
ARL and EDD

Subspace-CUSUM:

\[ S_t = (S_{t-1})^+ + (\hat{u}_t^T x_t)^2 - d \]

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, \((\hat{u}_t^T x_t)^2 - d\) is not exactly a log-likelihood ratio.
Proof to Theorem 1

- Equalizer trick
- Introducing “equalizer” $\delta_\infty$

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

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

- Represent $d$ using $\delta_\infty$

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

Set constant \( ARL = \gamma \), the detection delay

\[
E_0[T] = \frac{\log(\gamma)(1 + o(1))}{\tilde{\delta}_\infty(1 + \rho) \left(1 - \frac{p-1}{w\rho}\right) + \frac{1}{2} \log(1 - 2\tilde{\delta}_\infty)} + w.
\]

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 $\rho, \sigma^2$ known, $\mu$ unknown, for any $\text{ARL} = \gamma$ ($\gamma > 0$), EDD of Subspace-CUSUM using the optimal drift and optimal window size is

$$E_0[T] = \frac{2 \log(\gamma)}{\rho - \log(1 + \rho)} \left(1 + o(1)\right),$$

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 drift
  - Optimal window $w^* = O(\sqrt{\log \gamma})$
  - 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.