First-order optimality of Subspace-CUSUM

Liyan Xie*, George V. Moustakides†, and Yao Xie*

*H. Milton Stewart School of Industrial & Systems Engineering, Georgia Institute of Technology
†Department of Computer Science, Rutgers University

Overview

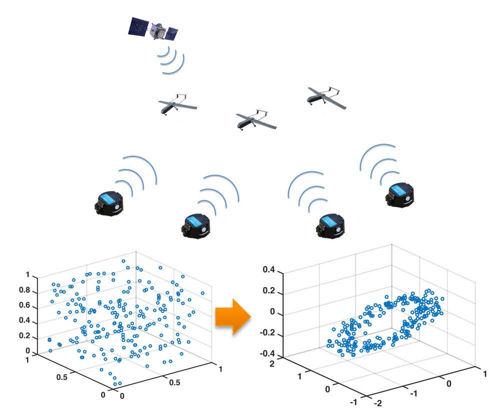
- Consider sequential change-point detection for detecting covariance changes characterized by the subspace.
- Propose the Subspace-CUSUM procedure, which is first-order asymptotic optimal.
- Develop a analytical methodology that includes proper parameter optimizations for the proposed detection scheme.

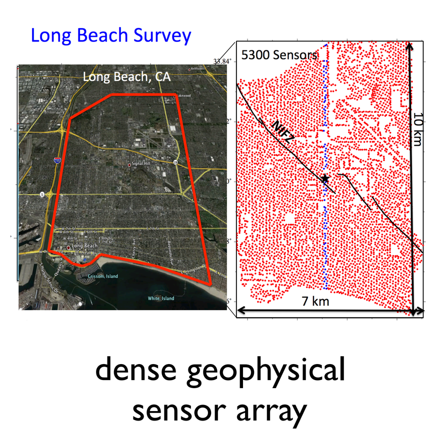
Introduction

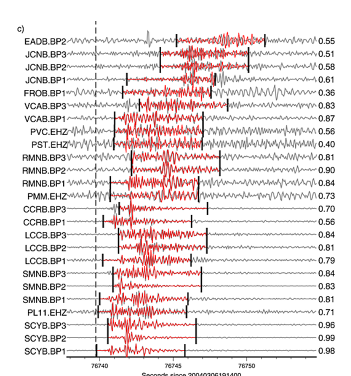
- Given a sequence of samples

$$x_1, x_2, \dots, x_t \dots,$$

there may be a change-point time τ where the distribution of the data stream changes.
- **Goal**: Detect the change as quickly as possible from sequential data.
- **Applications**: swarm behavior monitoring, seismic signal detection, power network anomaly detection, etc.



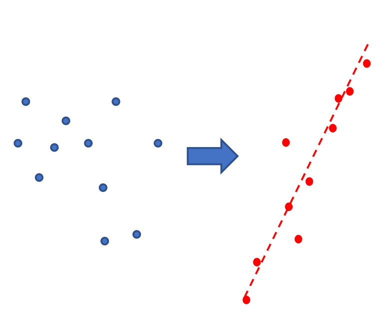




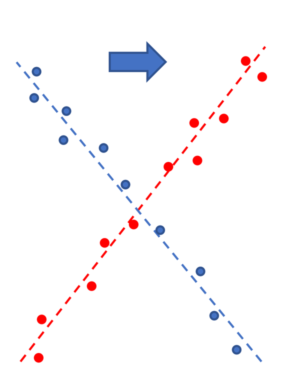
Problem setup

- The *emerging subspace* problem:

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



- The *switching subspace* problem:



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

- Equivalence: $\exists Q \in \mathbb{R}^{(k-1)\times k}$ s.t. $Qu_1 = 0$ and $QQ^\intercal = I_{k-1}$.

$$y_t = Qx_t \quad \Longrightarrow \quad \begin{array}{ll} y_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I_{k-1}), & t = 1, 2, \dots, \tau, \\ y_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I_{k-1} + \tilde{\theta}\tilde{u}\tilde{u}^\intercal), & t = \tau+1, \tau+2, \dots \end{array}$$

Detection procedure: Subspace CUSUM

- Log-likelihood ratio:

$$\log\frac{f_1(x_t)}{f_0(x_t)} = (u^\intercal x_t)^2 - \underbrace{\sigma^2\left(1+\frac{1}{\rho}\right)\log(1+\rho)}_{drift},$$

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

Subspace-CUSUM detection scheme

- At each time t, form the sample covariance matrix using a sliding window of size w:

$$\Sigma_t = \sum_{i=t+1}^{t+w} x_i x_i^\intercal,$$

$\hat{u}_t$ is the unit-norm eigenvector corresponding to the largest eigenvalue of Σ_t.
- Update the detection statistic:

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

- The stopping time:

$$\mathcal{T}_{\mathrm{C}} = \inf\{t > 0 : \mathcal{S}_t \geq b\}.$$

- Quantity d is a constant satisfying

$$\mathbb{E}_\infty[(\hat{u}_t^\intercal x_t)^2] < d < \mathbb{E}_0[(\hat{u}_t^\intercal x_t)^2].$$

- Leveraging the independence between $\hat{u}_t$ and x_t, we have

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

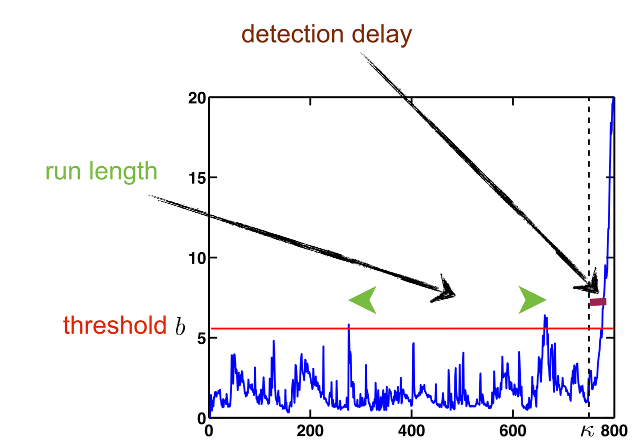
$$\Longrightarrow \quad \sigma^2 < d < \sigma^2(1+\rho)\left(1 - \frac{k-1}{w\rho}\right)$$

Performance metrics

- average run length (ARL):

$$\mathbb{E}_\infty(\mathcal{T}_{\mathrm{C}})$$

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



$$\sup_{\tau\geq 0} \operatorname{ess\,sup} \mathbb{E}_\tau[(T-\tau)^+ | T > \tau, x_1, \dots, x_\tau].$$

- Approximation: EDD $= \mathbb{E}_0(\mathcal{T}_{\mathrm{C}})$.

Asymptotic analysis

Optimality

The Subspace-CUSUM is asymptotically first-order optimal.

Proof Sketch

- **Equalizer trick**: Introducing an "equalizer" δ_∞ satisfying

$$\mathbb{E}_\infty[e^{\delta_\infty[(\hat{u}_t^\intercal x_t)^2 - d]}] = 1, \left(d = -\frac{1}{2\delta_\infty}\log(1 - 2\sigma^2\delta_\infty)\right)$$

after equalizing, $(\hat{u}_t^\intercal x_t)^2 - d \approx$ a log-likelihood ratio;
- Set constant ARL $= \gamma$;
- $\forall w$, the optimal drift d which minimizes the EDD is

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

- Substitute d^* and derive the optimal w which minimizes the EDD:

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

Numerical examples

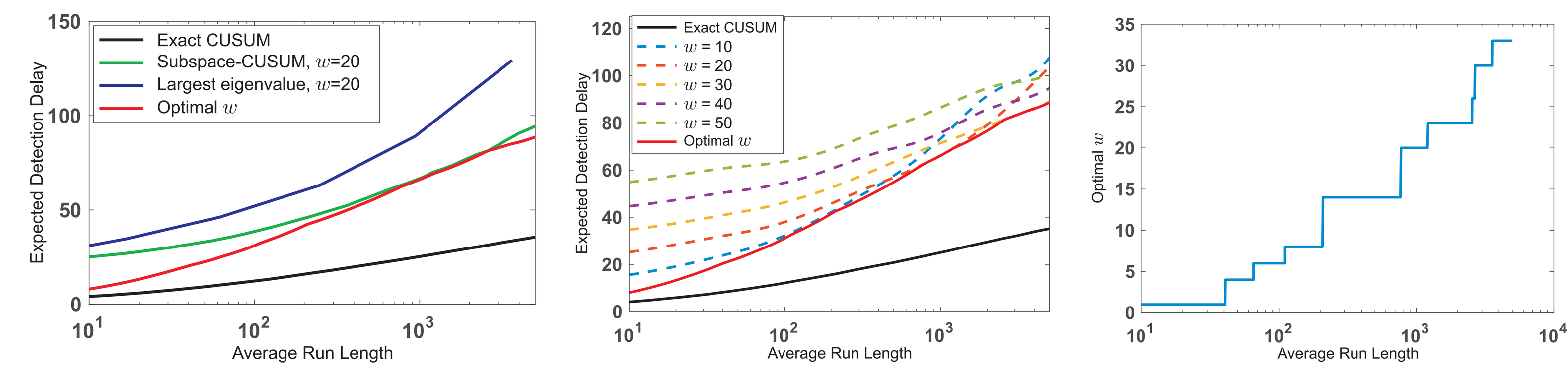


Figure 1: Left: Minimal EDD (red) among window sizes w from 1 to 50; Middle: Various EDD for different window sizes; Right: Corresponding optimal window size w.

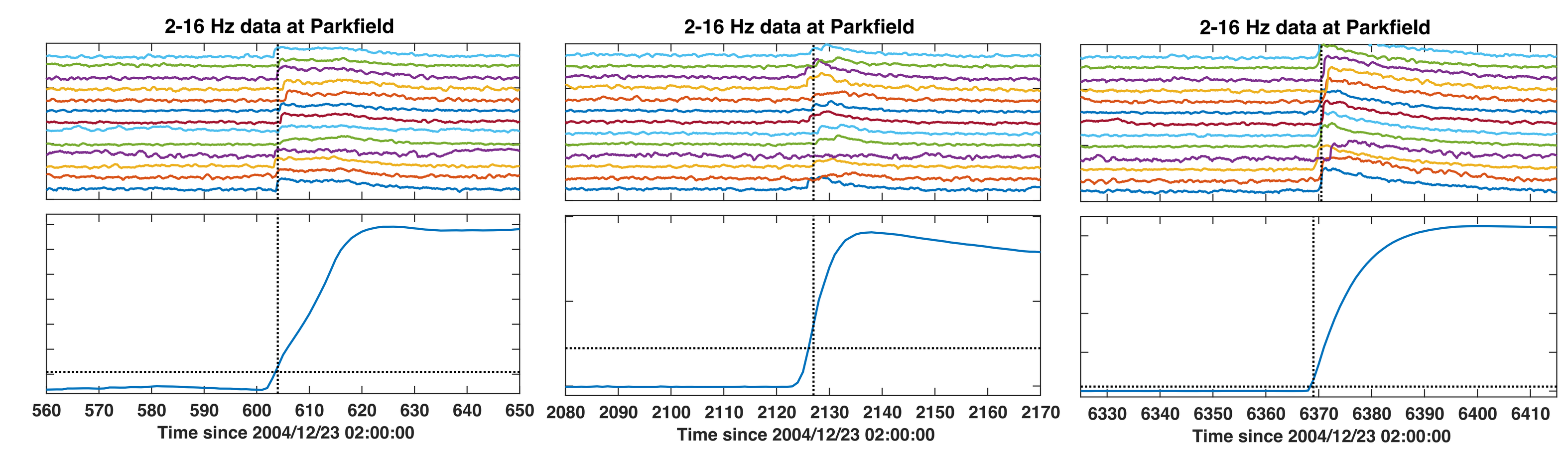


Figure 2: Real seismic data example: left, middle, and right figures correspond to the seismic event at time 605, 2127, and 6370 respectively.

This work has been accepted for oral presentation at GlobalSIP 2018. The full paper can be found at `https://arxiv.org/pdf/1806.10760.pdf`.

Founded by NSF CCF-1650913, CMMI-1538746, and CCF-1442635 of Yao Xie, and NSF CIF-1513373 of George V. Moustakides.