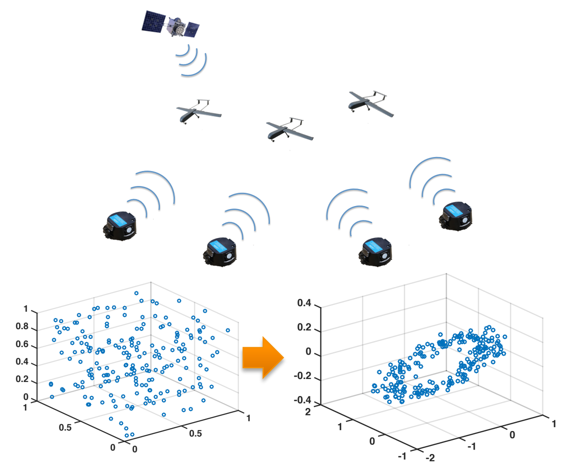
Robust Statistical Inference Through the Lens of Optimization

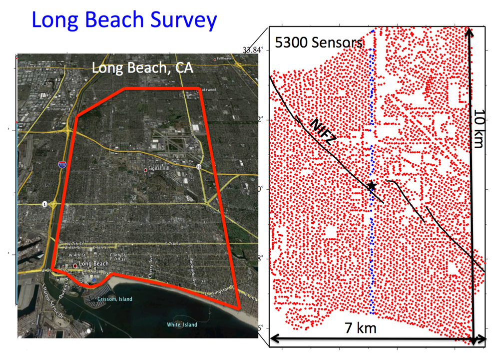
Liyan Xie

H. Milton Stewart School of Industrial & Systems Engineering, Georgia Institute of Technology

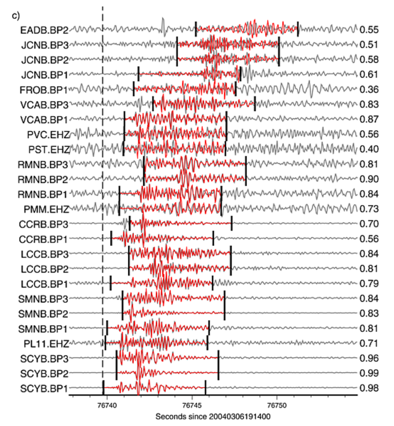
Overview and Motivating Examples

- Sequential change detection for multi-sensors based on subspace structures
- Robust hypothesis testing by hedging against data uncertainties





dense geophysical sensor array



Sensor/Social networks, Wearable sensors, Mobile health, etc.

Problem Setup

- Given a sequence of samples

$$\underbrace{x_1, x_2, \ldots, x_{\tau-1}}_{f_0}, \underbrace{x_\tau, x_{\tau+1}, \ldots, x_t \ldots}_{f_1}$$

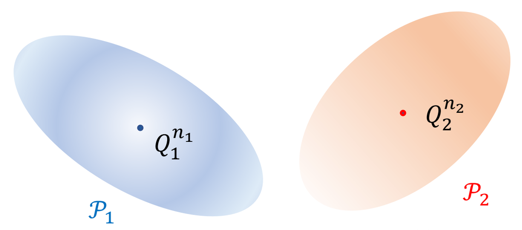
- **Goal**: Detect the change-point τ as quickly as possible.
- Metrics: **false-alarm rate** and **detection delay**.

1. The **subspace detection** problem:

$$f_0 = \mathcal{N}(0, \sigma^2 I_k), \quad f_1 = \mathcal{N}(0, \sigma^2 I_k + \theta u u^\intercal)$$

2. The **robust** hypothesis testing problem

$$H_0 : f_0 \in \mathcal{P}_1, \quad H_1 : f_1 \in \mathcal{P}_2$$



where Wasserstein uncertainty sets: $\mathcal{P}_k = \{P : \mathcal{W}(P, Q_k^{n_k}) \leq \theta_k\}$, $k = 1, 2$; $Q_k^{n_k}$ are empirical distributions

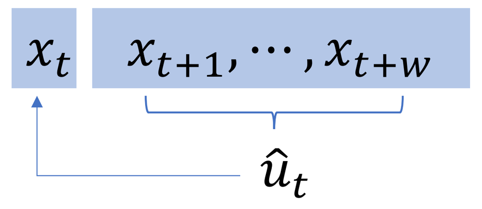
Subspace Detection Procedure

- Insight: log-likelihood ratio:

$$\log \frac{f_1(x_t)}{f_0(x_t)} = (u^\intercal x_t)^2 - \sigma^2 \left(1 + \frac{\sigma^2}{\theta}\right) \log(1 + \frac{\theta}{\sigma^2}).$$

- **Subspace-CUSUM** detection procedure:
  - At each time t, estimate $\hat{u}_t$ as the leading eigenvector of the sample covariance matrix $\Sigma_t = \sum_{i=t+1}^{t+w} x_i x_i^\intercal$;
  - Update the detection statistic recursively:

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



  - The stopping time:

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

- Optimality properties: the Subspace-CUSUM is asymptotically **first-order optimal** (minimize the detection delay subject to false-alarm constraint).

Robust Hypothesis Test

Robustness induces the minimax formulation:

$$\inf_T \sup_{P_1 \in \mathcal{P}_1, P_2 \in \mathcal{P}_2} \ell(T; P_1, P_2),$$

where test T can be induced by a randomized test; risk can be defined as $\ell(T; P_1, P_2) =$ Type-I + Type-II risks.

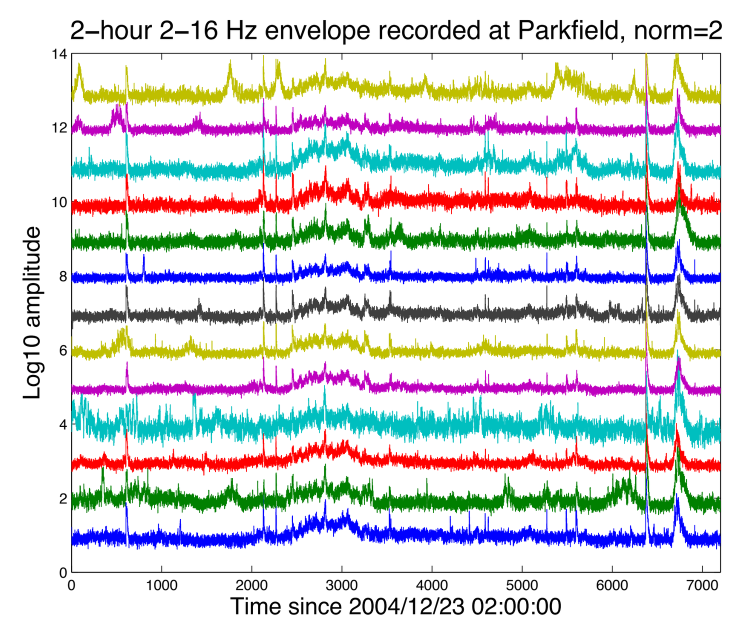
- Step I [**Strong duality**]

$$\inf_T \sup_{P_1 \in \mathcal{P}_1, P_2 \in \mathcal{P}_2} \ell(T; P_1, P_2) = \sup_{P_1 \in \mathcal{P}_1, P_2 \in \mathcal{P}_2} \inf_T \ell(T; P_1, P_2)$$

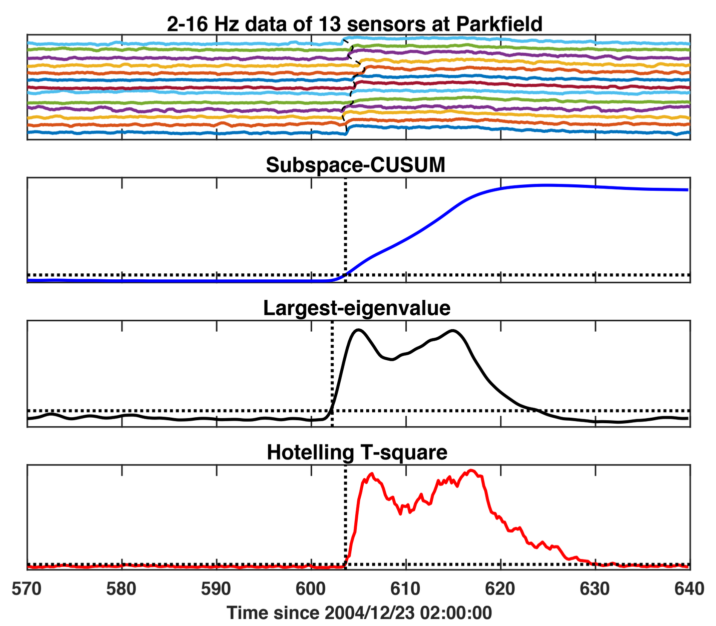
- Step II [**Least favorable distributions**] To find P_1^*, P_2^* closest to each other (statistical interpretation).

Minimax optimal test is induced by P_1^*, P_2^*.

Numerical Examples

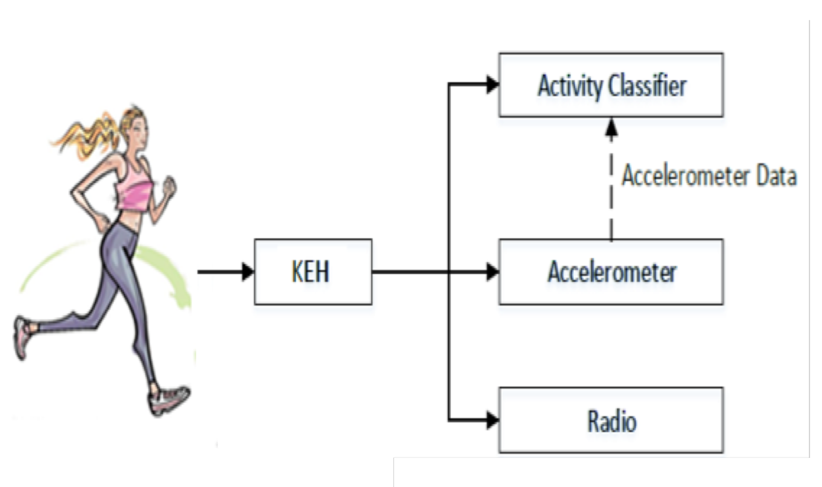


Raw data



Detection Statistics

Figure: **Seismic event detection**. Real data from long beach area.



Credit: CSIRO Research

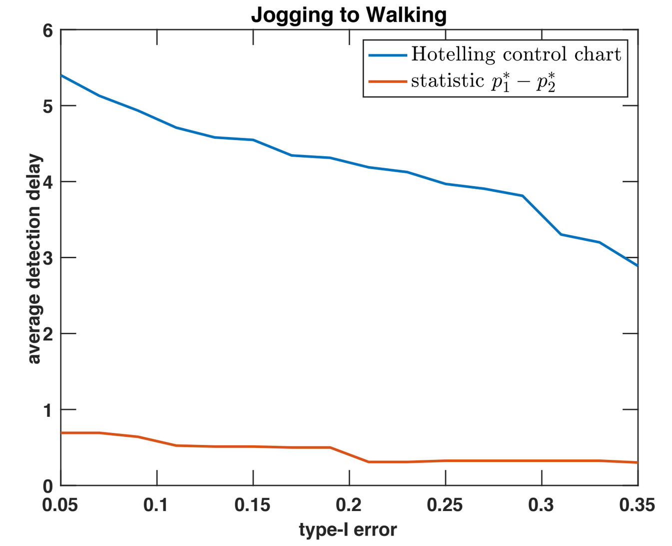


Figure: **Human activity change detection.** Jogging vs. Walking.

Discussion and References

- New applications: **mobile health** to monitor vital signs; **epidemic hot-spot detection**; limited training samples; etc.
- **Intersection of statistics and robust optimization**.

[1] Liyan Xie, Yao Xie, and George V. Moustakides. "Sequential subspace changepoint detection." arXiv:1811.03936 (2018), to appear in Sequential Analysis, 2020. [2] Rui Gao, Liyan Xie, Yao Xie, and Huan Xu. "Robust hypothesis testing using wasserstein uncertainty sets." NeurIPS, 2018.

Founded by NSF CCF-1650913, CMMI-1538746, CCF-1442635, CIF-1513373.