Recent advances in high-dimensional changepoint problems

Richard J. Samworth

University of Cambridge

StatScale Industry Event
Wellcome Collection, London

20 April 2023

Changepoints – a hypercube lattice

Changepoints – a hypercube lattice

- Challenge $\in \{\text{Detection}, \text{Estimation}, \text{Confidence intervals}\}$

Changepoints – a hypercube lattice

- Challenge $\in \{\text{Detection}, \text{Estimation}, \text{Confidence intervals}\}$
- Number of changepoints $\in \{\text{One}, \text{Many}\}$

Changepoints – a hypercube lattice

- Challenge $\in \{\text{Detection}, \text{Estimation}, \text{Confidence intervals}\}$
- Number of changepoints $\in \{\text{One}, \text{Many}\}$
- Dimension $\in \{\text{Univariate}, \text{Multivariate}, \text{High}\}$

Changepoints – a hypercube lattice

- Challenge $\in \{\text{Detection}, \text{Estimation}, \text{Confidence intervals}\}$
- Number of changepoints $\in \{\text{One}, \text{Many}\}$
- Dimension $\in \{\text{Univariate}, \text{Multivariate}, \text{High}\}$
- Change type $\in \{\text{Mean}, \text{Covariance}, \text{Regression coefficient}, \text{Distribution}\}$

Changepoints – a hypercube lattice

- Challenge $\in$ {Detection, Estimation, Confidence intervals}
- Number of changepoints $\in$ {One, Many}
- Dimension $\in$ {Univariate, Multivariate, High}
- Change type $\in$ {Mean, Covariance, Regression coefficient, Distribution}
- Dependence $\in$ {None, Spatial, Temporal, ...}

Changepoints – a hypercube lattice

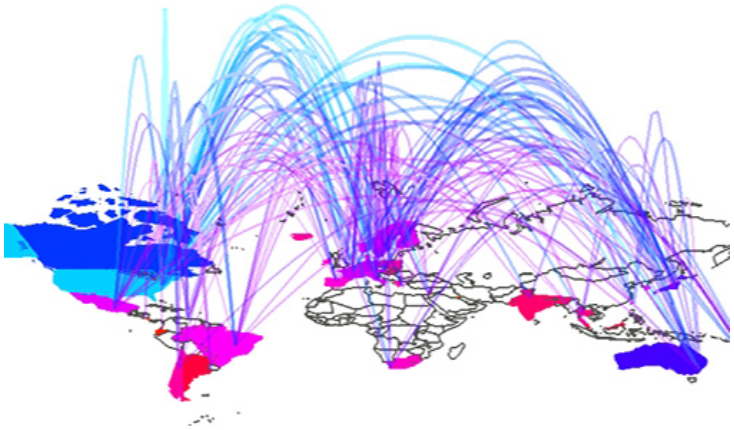
- Challenge $\in \{\text{Detection}, \text{Estimation}, \text{Confidence intervals}\}$
- Number of changepoints $\in \{\text{One}, \text{Many}\}$
- Dimension $\in \{\text{Univariate}, \text{Multivariate}, \text{High}\}$
- Change type $\in \{\text{Mean}, \text{Covariance}, \text{Regression coefficient}, \text{Distribution}\}$
- Dependence $\in \{\text{None}, \text{Spatial}, \text{Temporal}, \ldots\}$
- Observation format $\in \{\text{Offline}, \text{Online}\}$

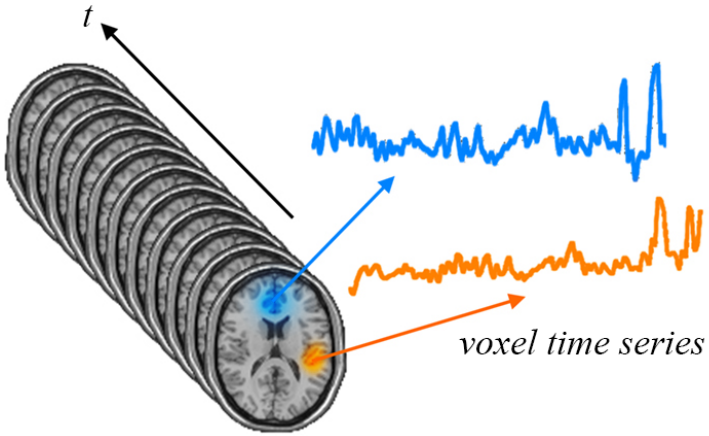
Changepoints – a hypercube lattice

- Challenge $\in$ {Detection, Estimation, Confidence intervals}
- Number of changepoints $\in$ {One, Many}
- Dimension $\in$ {Univariate, Multivariate, High}
- Change type $\in$ {Mean, Covariance, Regression coefficient, Distribution}
- Dependence $\in$ {None, Spatial, Temporal, ...}
- Observation format $\in$ {Offline, Online}
- Other structure $\in$ {Missingness, Network, Functional data, ...}

High-dimensional changepoint models

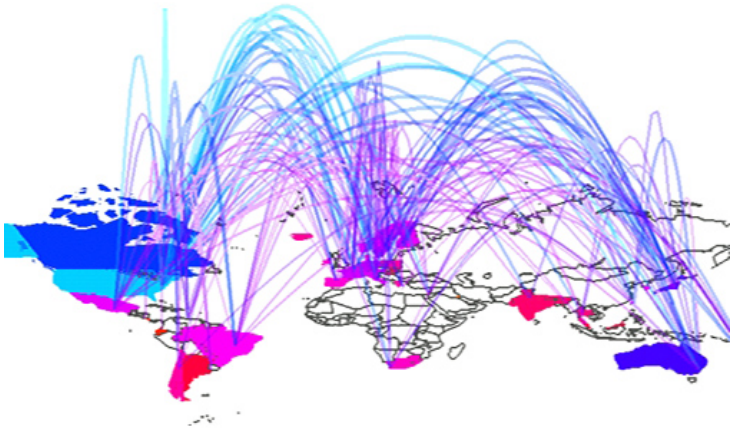
- The evolution of technology enables the collection of vast amounts of high-dimensional, time-ordered data:
  - Healthcare devices
  - Covid case numbers
  - Trading data of financial instruments

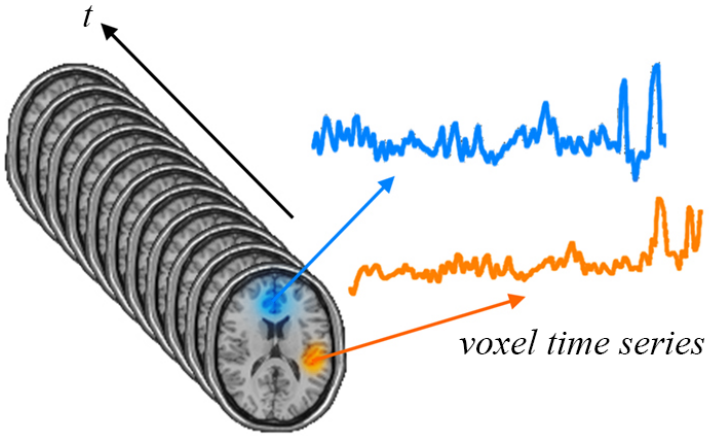




High-dimensional changepoint models

- The evolution of technology enables the collection of vast amounts of high-dimensional, time-ordered data:
  - Healthcare devices
  - Covid case numbers
  - Trading data of financial instruments





- Changes in the dynamics of the data streams are frequently of interest.

Recent trends in changepoints

- More specialised/realistic data generating mechanisms

Recent trends in changepoints

- More specialised/realistic data generating mechanisms
- Finite-sample bounds on performance, often uniform over classes

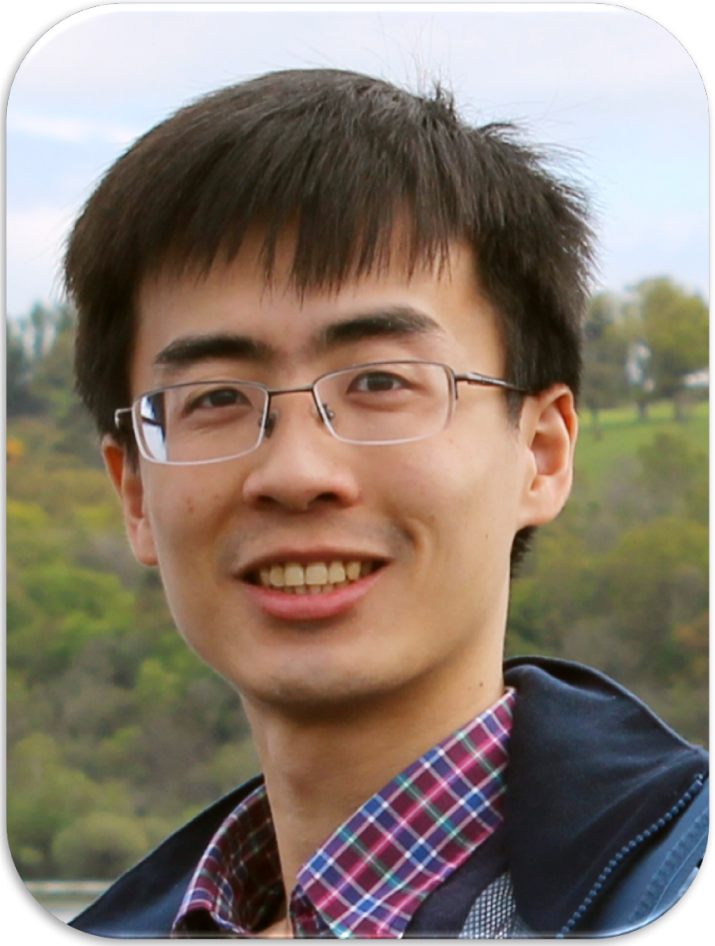
Recent trends in changepoints

- More specialised/realistic data generating mechanisms
- Finite-sample bounds on performance, often uniform over classes
- Optimality theory via minimax lower bounds

Recent trends in changepoints

- More specialised/realistic data generating mechanisms
- Finite-sample bounds on performance, often uniform over classes
- Optimality theory via minimax lower bounds
- Efficient computation.

inspect methodology



Tengyao Wang

Assume that $X_t = (X_{1,t}, \ldots, X_{p,t})^\top \sim N_p(\mu_t, \sigma^2 I_p)$ independently with

$$\mu_1 = \cdots = \mu_z = \mu^{(1)} \quad \text{and} \quad \mu_{z+1} = \cdots = \mu_n = \mu^{(2)}.$$

Assume that $X_t = (X_{1,t}, \ldots, X_{p,t})^\top \sim N_p(\mu_t, \sigma^2 I_p)$ independently with

$$\mu_1 = \cdots = \mu_z = \mu^{(1)} \quad \text{and} \quad \mu_{z+1} = \cdots = \mu_n = \mu^{(2)}.$$

We assume that the vector of change $\theta := \mu^{(2)} - \mu^{(1)}$ is sparse in the sense that $\|\theta\|_0 \leq s \ll p$.

Assume that $X_t = (X_{1,t}, \ldots, X_{p,t})^\top \sim N_p(\mu_t, \sigma^2 I_p)$ independently with

$$\mu_1 = \cdots = \mu_z = \mu^{(1)} \quad \text{and} \quad \mu_{z+1} = \cdots = \mu_n = \mu^{(2)}.$$

We assume that the vector of change $\theta := \mu^{(2)} - \mu^{(1)}$ is sparse in the sense that $\|\theta\|_0 \leq s \ll p$.

Goal: estimate the changepoint location z.

Further model assumptions

Assume that

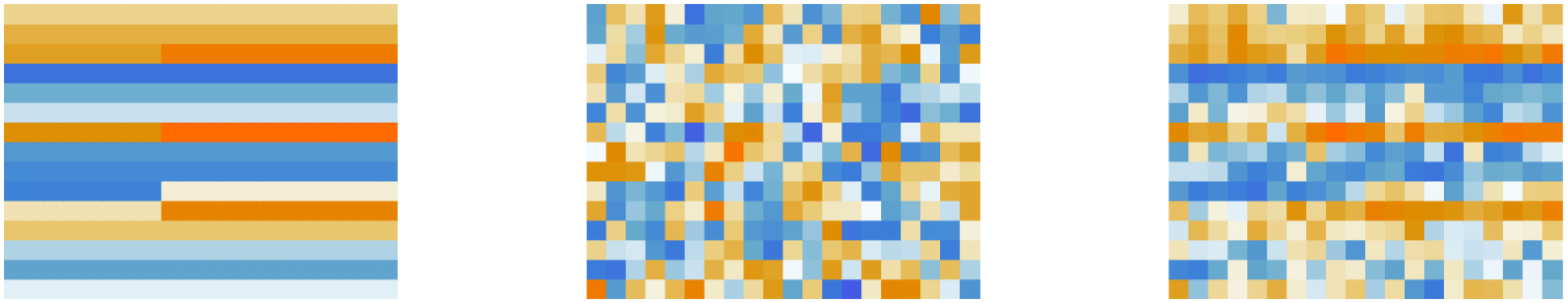
$$\frac{1}{n}\min(z, n-z) \geq \tau,$$

and the magnitude of the mean change is such that

$$\|\theta\|_2 \geq \vartheta.$$

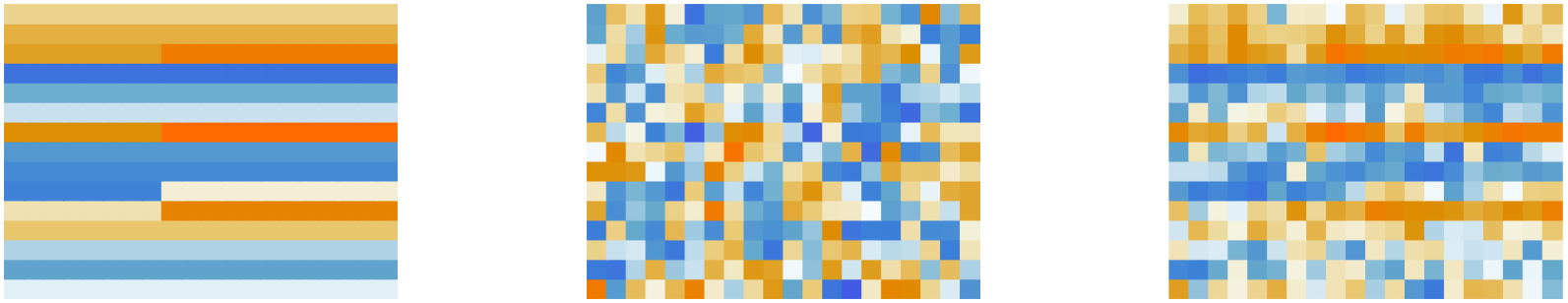
Let $\mathcal{P}(n, p, s, \vartheta, \tau, \sigma^2)$ be the set of distributions of such data matrices X.

Projection-based changepoint estimation



$$\boldsymbol{\mu} + W = X$$

Projection-based changepoint estimation



$\boldsymbol{\mu} \quad + \quad W \quad = \quad X$

For $a \in \mathbb{S}^{p-1}$,

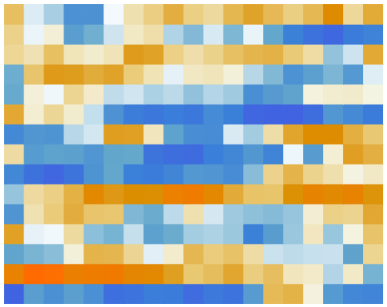
$$a^\top X_t \sim N(a^\top \boldsymbol{\mu}, \sigma^2).$$

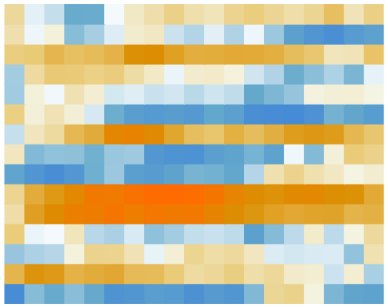
Optimal projection direction is $v := \theta / \|\theta\|_2$.

CUSUM transformation

Define CUSUM transformation $\mathcal{T}_{p,n} : \mathbb{R}^{p\times n} \to \mathbb{R}^{p\times(n-1)}$ by

$$[\mathcal{T}(M)]_{j,t} = [\mathcal{T}_{p,n}(M)]_{j,t} := \sqrt{\frac{t(n-t)}{n}}\left(\sum_{r=t+1}^{n}\frac{M_{j,r}}{n-t} - \sum_{r=1}^{t}\frac{M_{j,r}}{t}\right).$$





$$\begin{array}{ccccc} \mathcal{T}(\boldsymbol{\mu}) & + & \mathcal{T}(W) & = & \mathcal{T}(X) \\ A & + & E & = & T \end{array}$$

SVD of CUSUM transformation

The entries of A can be given explicitly:

$$A_{j,t} = \begin{cases} \sqrt{\frac{t}{n(n-t)}}(n-z)\theta_j, & \text{if } t \leq z \\ \sqrt{\frac{n-t}{nt}}z\theta_j, & \text{if } t > z \end{cases} =: (\theta\gamma^\top)_{j,t},$$

so the oracle projection direction is the leading left singular vector of the rank 1 matrix A.

SVD of CUSUM transformation

The entries of A can be given explicitly:

$$A_{j,t} = \begin{cases} \sqrt{\frac{t}{n(n-t)}}(n-z)\theta_j, & \text{if } t \leq z \\ \sqrt{\frac{n-t}{nt}}z\theta_j, & \text{if } t > z \end{cases} =: (\theta\gamma^\top)_{j,t},$$

so the oracle projection direction is the leading left singular vector of the rank 1 matrix A.

We could therefore consider estimating v by $\hat{v}_{\max,s} \in \operatorname{argmax}_{\tilde{v}\in\mathbb{S}^{p-1}(s)} \|T^\top \tilde{v}\|_2$, and indeed when $n \geq 6$, with probability at least $1 - 4(p\log n)^{-1/2}$,

$$\sin\angle(\hat{v}_{\max,s}, v) \leq \frac{16\sqrt{2}\sigma}{\tau\vartheta}\sqrt{\frac{s\log(p\log n)}{n}}.$$

A computationally efficient projection

Computing the s-sparse leading left singular vector of a matrix is NP-hard. However,

$$\max_{u \in \mathbb{S}^{p-1}(s)} \|u^\top T\|_2 = \max_{u \in \mathbb{S}^{p-1}(s), w \in \mathbb{S}^{n-2}} u^\top T w$$
$$= \max_{u \in \mathbb{S}^{p-1}, w \in \mathbb{S}^{n-2}, \|u\|_0 \leq s} \langle uw^\top, T \rangle = \max_{M \in \mathcal{M}} \langle M, T \rangle,$$

where $\mathcal{M} := \{M : \|M\|_* = 1, \mathrm{rank}(M) = 1, \mathrm{nnzr}(M) \leq s\}$.

A computationally efficient projection

Computing the s-sparse leading left singular vector of a matrix is NP-hard. However,

$$\max_{u \in \mathbb{S}^{p-1}(s)} \|u^\top T\|_2 = \max_{u \in \mathbb{S}^{p-1}(s), w \in \mathbb{S}^{n-2}} u^\top T w$$
$$= \max_{u \in \mathbb{S}^{p-1}, w \in \mathbb{S}^{n-2}, \|u\|_0 \leq s} \langle uw^\top, T \rangle = \max_{M \in \mathcal{M}} \langle M, T \rangle,$$

where $\mathcal{M} := \{M : \|M\|_* = 1, \mathrm{rank}(M) = 1, \mathrm{nnzr}(M) \leq s\}$.

For $\lambda > 0$, we therefore consider computing

$$\hat{M} \in \operatorname*{argmax}_{M \in \mathcal{S}} \big\{ \langle T, M \rangle - \lambda \|M\|_1 \big\},$$

where $\mathcal{S} := \{M \in \mathbb{R}^{p \times (n-1)} : \|M\|_* \leq 1\}$, using ADMM. We can then let $\hat{v}$ be a leading left singular vector of $\hat{M}$.

Projection direction estimator

Let $\hat{M} \in \operatorname{argmax}_{M \in \mathcal{S}} \{ \langle T, M \rangle - \lambda \|M\|_1 \}$ and $\hat{v} \in \operatorname{argmax}_{\tilde{v} \in \mathbb{S}^{p-1}} \|\hat{M}^\top \tilde{v}\|_2$. If $n \geq 6$ and $\lambda \geq 2\sigma\sqrt{\log(p \log n)}$, then with probability at least $1 - 4/\sqrt{p \log n}$,

$$\sin \angle(\hat{v}, v) \leq \frac{32\lambda\sqrt{s}}{\tau\vartheta\sqrt{n}}.$$

Projection direction estimator

Let $\hat{M} \in \mathrm{argmax}_{M \in \mathcal{S}} \{ \langle T, M \rangle - \lambda \|M\|_1 \}$ and $\hat{v} \in \mathrm{argmax}_{\tilde{v} \in \mathbb{S}^{p-1}} \|\hat{M}^\top \tilde{v}\|_2$. If $n \geq 6$ and $\lambda \geq 2\sigma\sqrt{\log(p \log n)}$, then with probability at least $1 - 4/\sqrt{p \log n}$,

$$\sin \angle(\hat{v}, v) \leq \frac{32\lambda\sqrt{s}}{\tau\vartheta\sqrt{n}}.$$

Let $\hat{z} \in \mathrm{argmax}_{t \in [n-1]} |\hat{v}^\top T_t|$, where T_t is the tth column of T.

Sample-splitting version performance

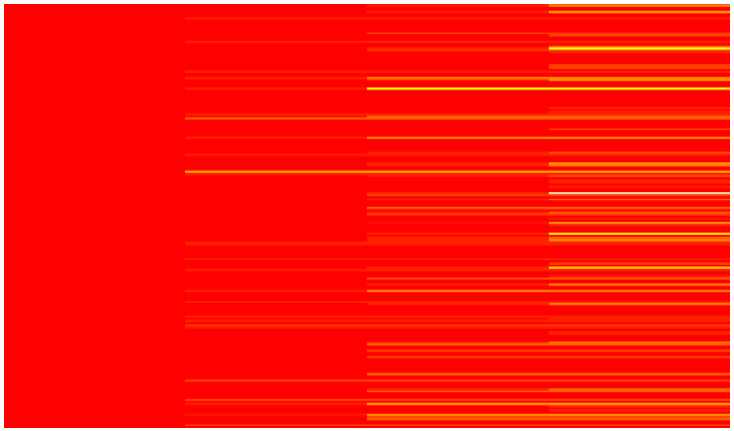
Let $\sigma > 0$ be known and $X \sim P \in \mathcal{P}(n, p, s, \vartheta, \tau, \sigma^2)$. Let $\hat{z}$ be the output of sample-splitting algorithm with input X, σ and $\lambda := 2\sigma\sqrt{\log(p \log n)}$. There exist universal constants $C, C' > 0$ such that if $n \geq 12$, z is even and

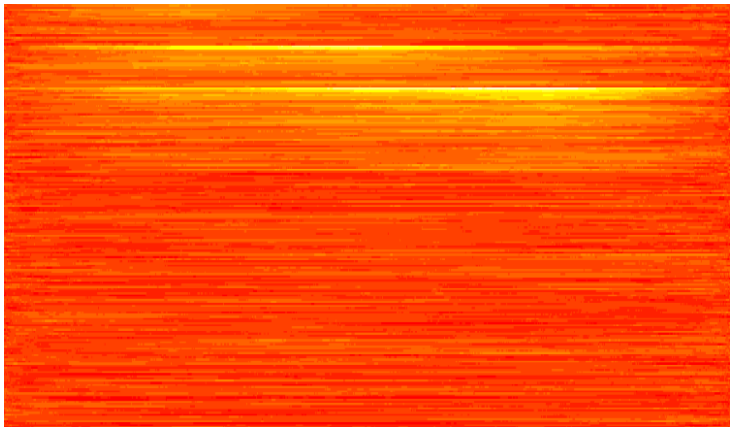
$$\frac{C\sigma}{\vartheta\tau}\sqrt{\frac{s\log(p\log n)}{n}} \leq 1,$$

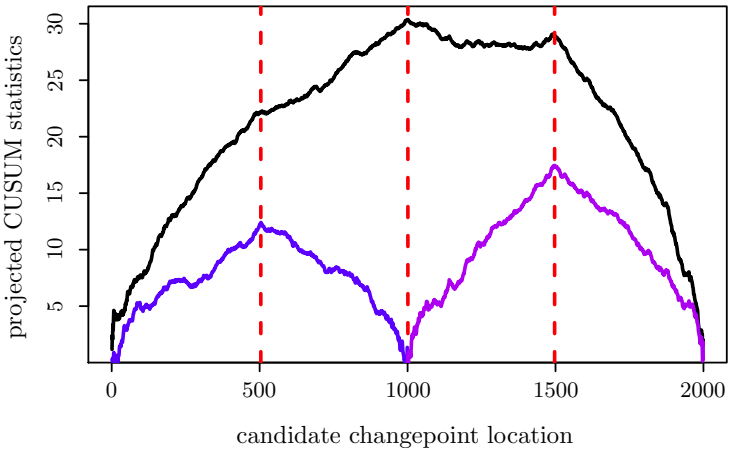
then with probability at least $1 - 4\{p\log(n/2)\}^{-1/2} - 17/\log(n/2)$,

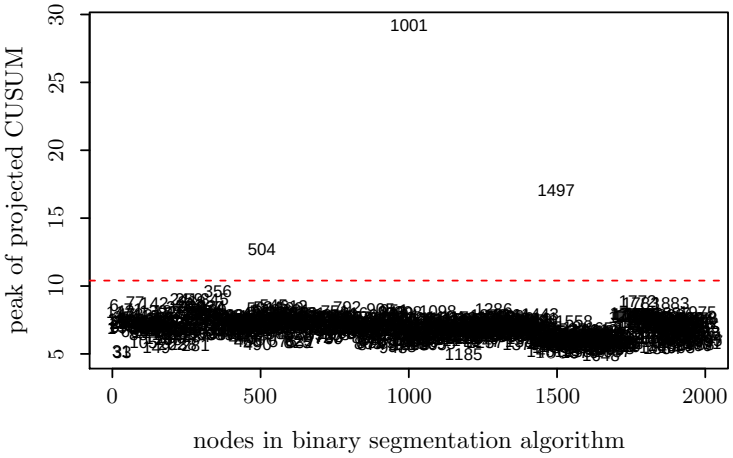
$$\frac{1}{n}|\hat{z} - z| \leq \frac{C'\sigma^2 \log\log n}{n\vartheta^2}.$$

Example



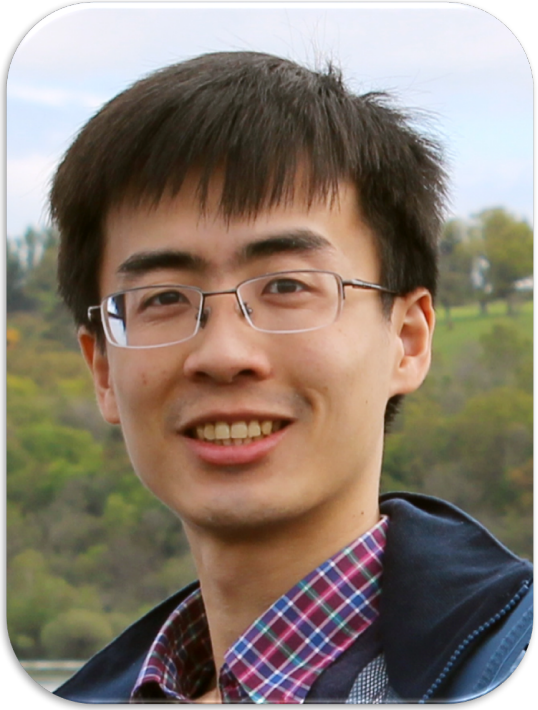



projected CUSUM statistics
5
10
15
20
25
30
0
500
1000
1500
2000
candidate changepoint location


peak of projected CUSUM
5
10
15
20
25
30
1001
1497
504
356
1185
0
500
1000
1500
2000
nodes in binary segmentation algorithm

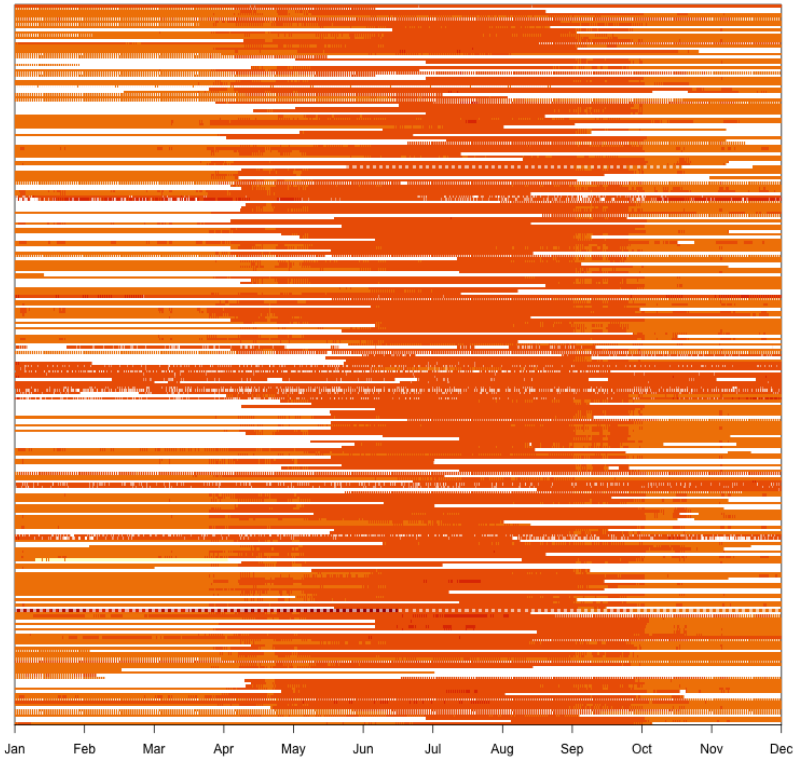
MissInspect methodology

Bertille Follain

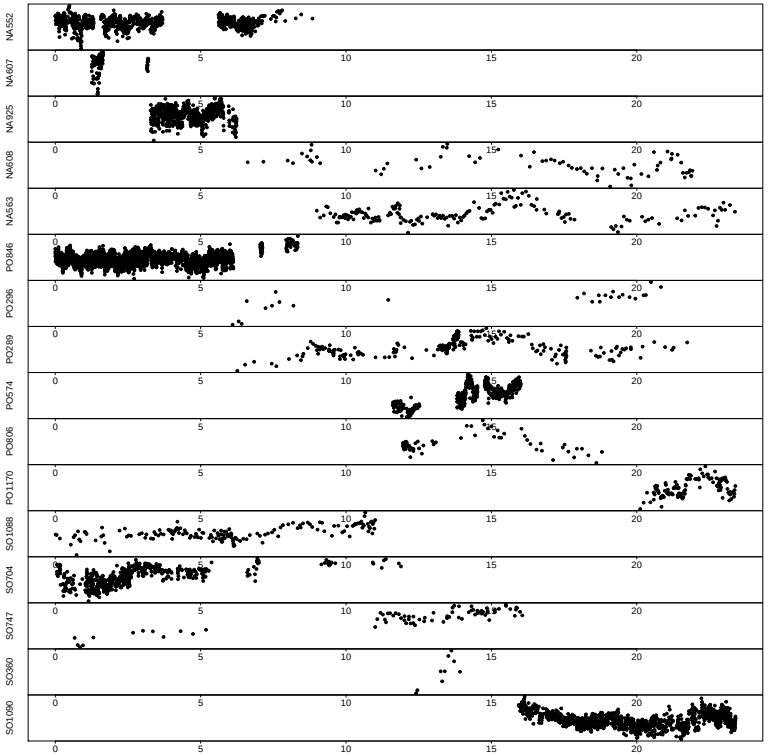


Tengyao Wang

Changepoint estimation with missing data



French river temperature in 2018



$^{13}C/^{12}C$ in ocean cores 0–23 Ma

Problem setup

- Observed data $(X \circ \Omega, \Omega)$
  - Full data matrix $X = (X_{j,t}) \in \mathbb{R}^{p \times n}$
  - Revelation matrix $\Omega = (\omega_{j,t}) \in \{0, 1\}^{p \times n}$: $\omega_{j,t} = 1$ if $X_{j,t}$ is observed and 0 otherwise.

Problem setup

- Observed data $(X \circ \Omega, \Omega)$
  - Full data matrix $X = (X_{j,t}) \in \mathbb{R}^{p \times n}$
  - Revelation matrix $\Omega = (\omega_{j,t}) \in \{0,1\}^{p \times n}$: $\omega_{j,t} = 1$ if $X_{j,t}$ is observed and 0 otherwise.
- Data distribution:
  - Assume $X_t = (X_{1,t}, \ldots, X_{p,t})^\top \sim N_p(\mu_t, \sigma^2 I_p)$ independently with

$$\mu_1 = \cdots = \mu_z = \mu^{(1)} \quad \text{and} \quad \mu_{z+1} = \cdots = \mu_n = \mu^{(2)}.$$

Problem setup

- Observed data $(X \circ \Omega, \Omega)$
  - Full data matrix $X = (X_{j,t}) \in \mathbb{R}^{p \times n}$
  - Revelation matrix $\Omega = (\omega_{j,t}) \in \{0,1\}^{p \times n}$: $\omega_{j,t} = 1$ if $X_{j,t}$ is observed and 0 otherwise.
- Data distribution:
  - Assume $X_t = (X_{1,t}, \ldots, X_{p,t})^\top \sim N_p(\mu_t, \sigma^2 I_p)$ independently with

$$\mu_1 = \cdots = \mu_z = \mu^{(1)} \quad \text{and} \quad \mu_{z+1} = \cdots = \mu_n = \mu^{(2)}.$$

  - Vector of change $\theta := \mu^{(2)} - \mu^{(1)}$ is s-sparse.

Problem setup

- Observed data $(X \circ \Omega, \Omega)$
  - Full data matrix $X = (X_{j,t}) \in \mathbb{R}^{p \times n}$
  - Revelation matrix $\Omega = (\omega_{j,t}) \in \{0,1\}^{p \times n}$: $\omega_{j,t} = 1$ if $X_{j,t}$ is observed and 0 otherwise.
- Data distribution:
  - Assume $X_t = (X_{1,t}, \ldots, X_{p,t})^\top \sim N_p(\mu_t, \sigma^2 I_p)$ independently with

$$\mu_1 = \cdots = \mu_z = \mu^{(1)} \quad \text{and} \quad \mu_{z+1} = \cdots = \mu_n = \mu^{(2)}.$$

  - Vector of change $\theta := \mu^{(2)} - \mu^{(1)}$ is s-sparse.
- Missingness mechanism:
  - $\omega_{j,t} \sim \mathrm{Bern}(q_j)$ independently, and independent of X.
- **Goal**: estimate the changepoint location z.

MissCUSUM transform

- Write

$$L_{j,t} := \sum_{r=1}^{t} \omega_{j,t}, \quad R_{j,t} := \sum_{j=n-t+1}^{n} \omega_{j,t}, \quad N_j := L_{j,n} + R_{j,n}.$$

- The MissCUSUM transformation $\mathcal{T}^{\mathrm{Miss}} : \mathbb{R}^{p\times n} \times \{0,1\}^{p\times n} \to \mathbb{R}^{p\times(n-1)}$ is defined such that $T_\Omega = \mathcal{T}^{\mathrm{Miss}}(X,\Omega)$ satisfies

$$(T_\Omega)_{j,t} := \sqrt{\frac{L_{j,t}R_{j,n-t}}{N_j}}\left(\frac{1}{R_{j,n-t}}\sum_{r=t+1}^{n}(X\circ\Omega)_{j,r} - \frac{1}{L_{j,t}}\sum_{r=1}^{t}(X\circ\Omega)_{j,r}\right),$$

  when $\min\{L_{j,t}, R_{j,t}\} > 0$ and 0 otherwise.
- When the data are fully-observed, i.e. Ω is an all-one matrix, $\mathcal{T}^{\mathrm{Miss}}$ reduces to the standard CUSUM transformation.

How to aggregate signal

- $T_\Omega = \mathcal{T}^{\mathrm{Miss}}(X \circ \Omega, \Omega)$ can be viewed as a perturbation of the rank one matrix $(\theta \circ \sqrt{\boldsymbol{q}})\gamma^\top$, where $\boldsymbol{q} := (q_1, \ldots, q_p)^\top$.

How to aggregate signal

- $T_\Omega = \mathcal{T}^{\mathrm{Miss}}(X \circ \Omega, \Omega)$ can be viewed as a perturbation of the rank one matrix $(\theta \circ \sqrt{\boldsymbol{q}})\gamma^\top$, where $\boldsymbol{q} := (q_1, \ldots, q_p)^\top$.
- We can estimate $\theta \circ \sqrt{\boldsymbol{q}}/\|\theta \circ \sqrt{\boldsymbol{q}}\|$ via the s-sparse leading left singular vector of T_Ω:

$$\max_{(v,w)\in\mathbb{B}^p\times\mathbb{B}^{n-1}} \quad v^\top T_\Omega w \qquad \text{subject to} \quad \|v\|_0 \leq s.$$

How to aggregate signal

- $T_\Omega = \mathcal{T}^{\mathrm{Miss}}(X \circ \Omega, \Omega)$ can be viewed as a perturbation of the rank one matrix $(\theta \circ \sqrt{\boldsymbol{q}})\gamma^\top$, where $\boldsymbol{q} := (q_1, \ldots, q_p)^\top$.
- We can estimate $\theta \circ \sqrt{\boldsymbol{q}}/\|\theta \circ \sqrt{\boldsymbol{q}}\|$ via the s-sparse leading left singular vector of T_Ω:

$$\max_{(v,w)\in\mathbb{B}^p\times\mathbb{B}^{n-1}} \quad v^\top T_\Omega w \qquad \text{subject to} \quad \|v\|_0 \leq s.$$

- Since this is non-convex and requires knowledge of s, we relax it into a bi-concave problem:

$$(\hat{v}, \hat{w}) \in \operatorname*{argmax}_{(v,w)\in\mathbb{B}^p\times\mathbb{B}^{n-1}} \left\{ v^\top T_\Omega w - \lambda\|v\|_1 \right\}$$

How to aggregate signal

- $T_\Omega = \mathcal{T}^{\mathrm{Miss}}(X \circ \Omega, \Omega)$ can be viewed as a perturbation of the rank one matrix $(\theta \circ \sqrt{\boldsymbol{q}})\gamma^\top$, where $\boldsymbol{q} := (q_1, \ldots, q_p)^\top$.
- We can estimate $\theta \circ \sqrt{\boldsymbol{q}}/\|\theta \circ \sqrt{\boldsymbol{q}}\|$ via the s-sparse leading left singular vector of T_Ω:

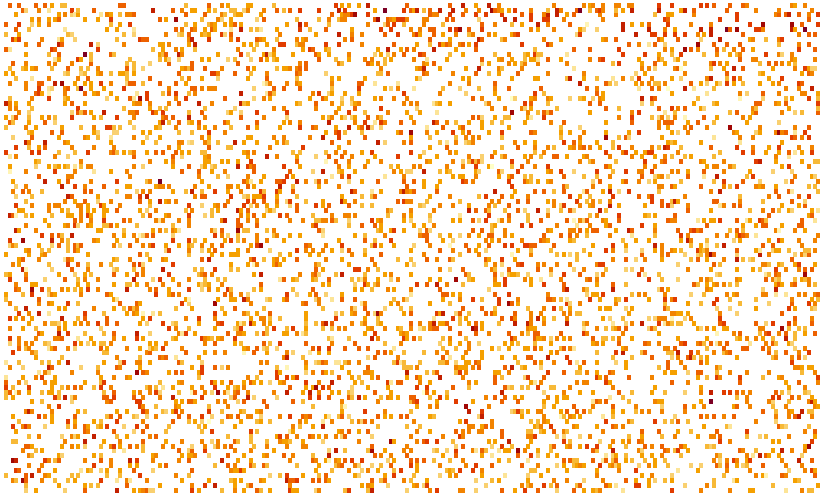
$$\max_{(v,w)\in\mathbb{B}^p\times\mathbb{B}^{n-1}} \quad v^\top T_\Omega w \qquad \text{subject to} \quad \|v\|_0 \leq s.$$

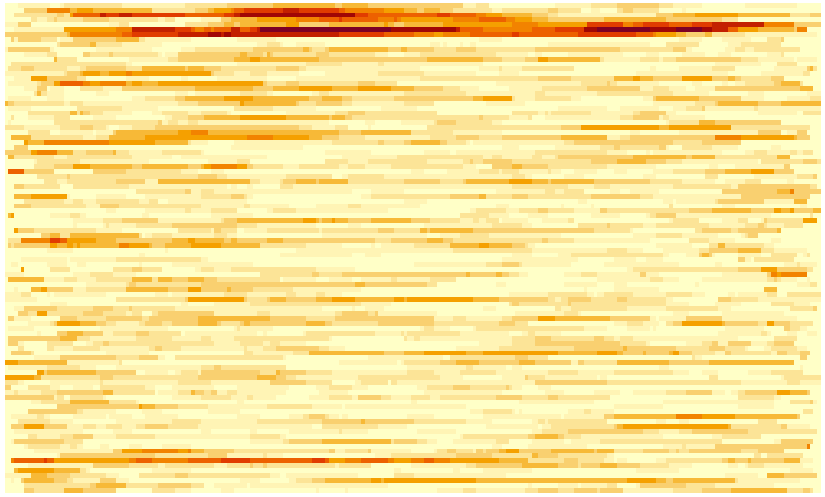
- Since this is non-convex and requires knowledge of s, we relax it into a bi-concave problem:

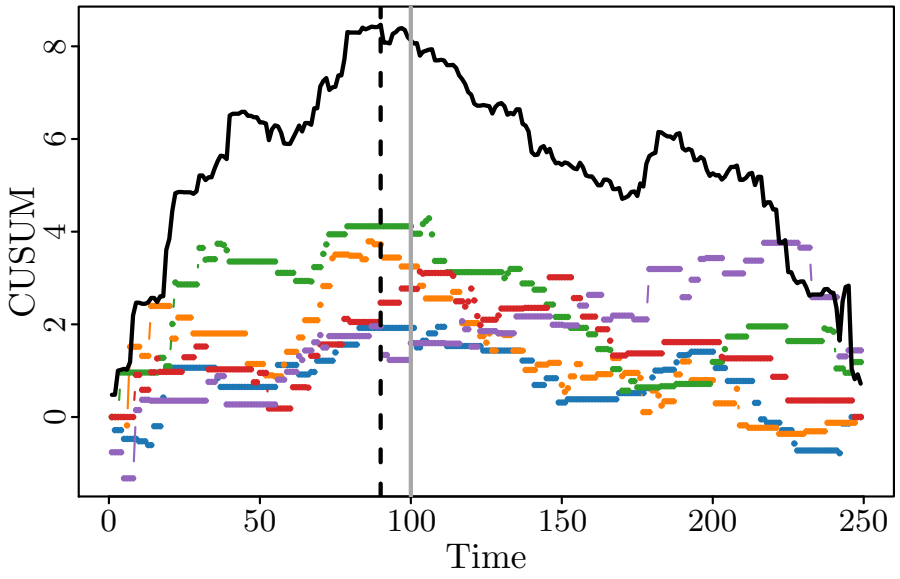
$$(\hat{v}, \hat{w}) \in \operatorname*{argmax}_{(v,w)\in\mathbb{B}^p\times\mathbb{B}^{n-1}} \left\{ v^\top T_\Omega w - \lambda\|v\|_1 \right\}$$

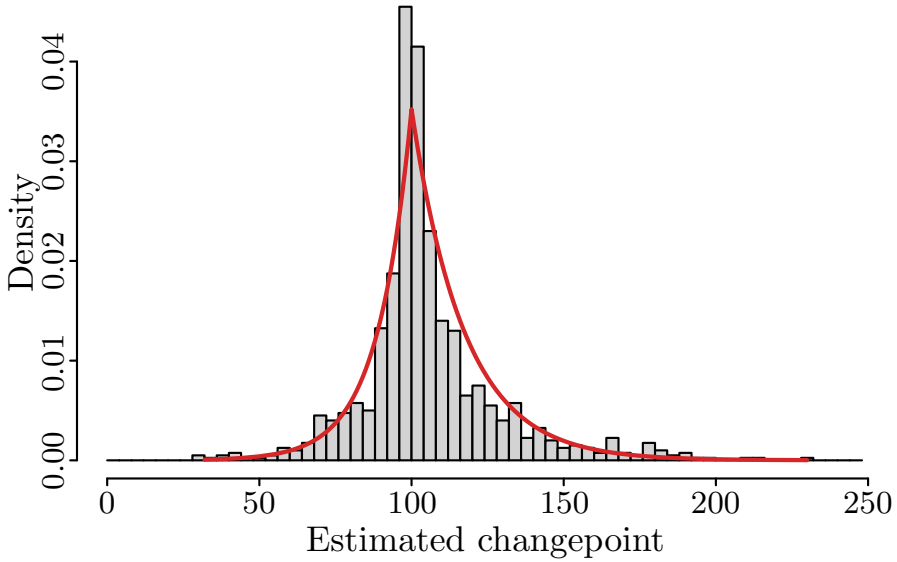
- Additional benefit: directly exploits the row sparsity pattern.

Illustration of the algorithm in action









Parameters: $p = 100$, $n = 250$, $z = 100$, $s = 10$, $\|\theta\|_2 = 2$, $q_j = 0.2 \; \forall \, j$

Summary

- We have made remarkable progress in understanding high-dimensional changepoints over the last 5-10 years!

Summary

- We have made remarkable progress in understanding high-dimensional changepoints over the last 5-10 years!

- Wang, T. and Samworth, R. J. (2018) High dimensional change point estimation via sparse projection. *J. Roy. Statist. Soc., Ser. B*, **80**, 57–83. R package: `InspectChangepoint`.
- Liu, H., Gao, C. and Samworth, R. J. (2021) Minimax rates in sparse, high-dimensional changepoint detection. *Ann. Statist.*, **49**, 1081–1112.
- Follain, B., Wang, T. and Samworth, R. J. (2022) High-dimensional changepoint estimation with heterogeneous missingness. *J. Roy. Statist. Soc., Ser. B*, **84**, 1023–1055.
- Chen, Y., Wang T. and Samworth, R. J. (2022) High-dimensional online changepoint detection. *J. Roy. Statist. Soc., Ser. B*, **84**, 234–266. R package: `ocd`.
- Chen, Y., Wang, T. and Samworth, R. J. (2023+) Inference in high-dimensional online changepoint detection. *J. Amer. Statist. Assoc.*, to appear.