High-dimensional, multiscale online changepoint detection

Tengyao Wang

University College London

Joint Statistics Seminar, Fudan University

18 June 2020

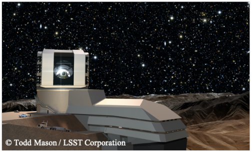
Collaborators

Yudong Chen

Richard Samworth

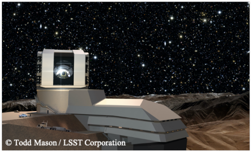
Changepoint problems

- Modern technology has facilitated the real-time monitoring of many types of evolving processes.



Changepoint problems

- Modern technology has facilitated the real-time monitoring of many types of evolving processes.



- Very often, a key feature of interest for data streams is a **changepoint**.

From offline to online

- The vast majority of the literature concerns the offline problem (Killick et al., 2012; W. and Samworth, 2018; Wang et al., 2018; Baranowski et al., 2019; Liu et al., 2019).
- Univariate online changepoints have been studied within the well-established field of *statistical process control* (Duncan, 1952; Page, 1954; Barnard, 1959; Fearnhead and Liu, 2007; Oakland, 2007).
- Much less work on multivariate, online changepoint problems (Tartakovsky et al., 2006; Mei, 2010; Zou et al., 2015).
  Several methods involve scanning a moving window of fixed size (Xie and Siegmund, 2013; Soh and Chandrasekaran, 2017; Chan, 2017).

Online algorithm

Key definition of an **online algorithm** for a data stream:

Definition. An algorithm is online if both its storage requirements and the computational complexity for processing a new observation depend only on **the number of bits needed to represent the new data**.

Online algorithm

Key definition of an **online algorithm** for a data stream:

Definition. An algorithm is online if both its storage requirements and the computational complexity for processing a new observation depend only on **the number of bits needed to represent the new data**.

- For the purposes of this definition, all real numbers are considered as floating point numbers.
- Importantly, we can only track a finite amount of summary statistics and are not allowed to store all historical data.

Problem setting

We consider a high-dimensional online changepoint detection problem:

- **Data**: for some unknown, deterministic time $z \in \mathbb{N} \cup \{0\}$, we have

$$X_1, \ldots, X_z \sim N_p(0, I_p) \quad \text{and} \quad X_{z+1}, X_{z+2}, \ldots \sim N_p(\theta, I_p).$$

- $\theta = 0$: data generated **under the null**, i.e. no change.
- $\theta \neq 0$: data generated **under the alternative**, i.e. there exists a change.

- Assume $\vartheta := \|\theta\|_2$ is at least a known lower bound $\beta > 0$.

Performance measures

A **sequential changepoint procedure** is an extended stopping time N (w.r.t. the natural filtration) taking values in $\mathbb{N} \cup \{\infty\}$.

- **patience**: $\mathbb{E}_0(N)$;
- Two types of **response delays**:
  - Average case response delay

$$\bar{\mathbb{E}}_\theta(N) := \sup_{z \in \mathbb{N}} \mathbb{E}_{z,\theta}\big\{(N - z) \vee 0\big\};$$

  - Worst case response delay

$$\bar{\mathbb{E}}_\theta^{\mathrm{wc}}(N) := \sup_{z \in \mathbb{N}} \operatorname{ess\,sup} \mathbb{E}_{z,\theta}\big\{(N - z) \vee 0 \mid X_1, \ldots, X_z\big\}.$$

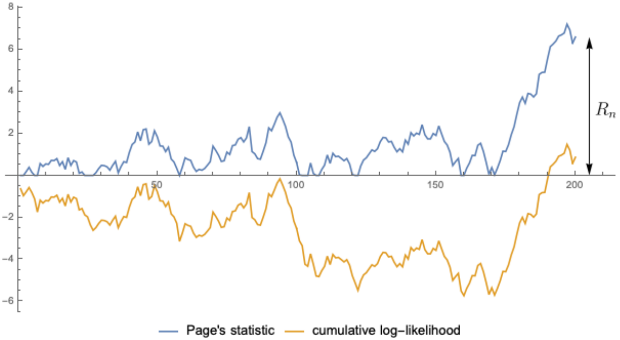
A warm-up: univariate online changepoint detection

Example of an online algorithm

Let $p = 1$ and assume $\theta > 0$. Page's procedure (Page, 1954):

$$R_n := \max_{0 \leq h \leq n} \sum_{i=n-h+1}^{n} \beta(X_i - \beta/2) = \max\{R_{n-1} + \beta(X_n - \beta/2), 0\}.$$

Threshold $T \equiv T_\beta$ for changepoint declaration.

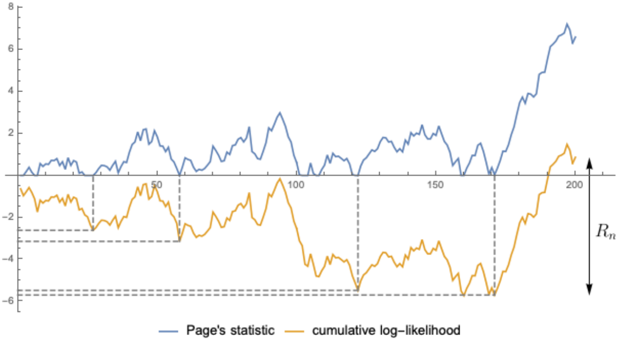


Example of an online algorithm

Let $p = 1$ and assume $\theta > 0$. Page's procedure (Page, 1954):

$$R_n := \max_{0 \le h \le n} \sum_{i=n-h+1}^{n} \beta(X_i - \beta/2) = \max\{R_{n-1} + \beta(X_n - \beta/2), 0\}.$$

Threshold $T \equiv T_\beta$ for changepoint declaration.

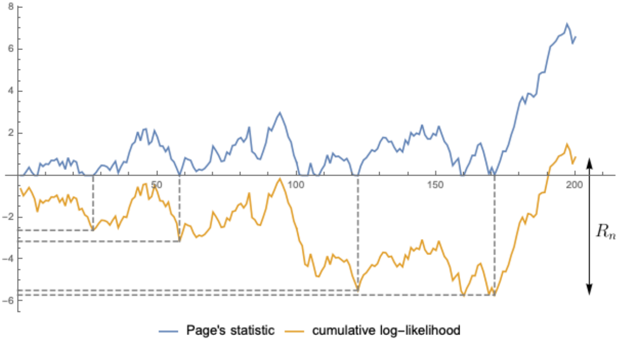


Example of an online algorithm

Let $p = 1$ and assume $\theta > 0$. Page's procedure (Page, 1954):

$$R_n := \max_{0 \le h \le n} \sum_{i=n-h+1}^{n} \beta(X_i - \beta/2) = \max\{R_{n-1} + \beta(X_n - \beta/2), 0\}.$$

Threshold $T \equiv T_\beta$ for changepoint declaration.



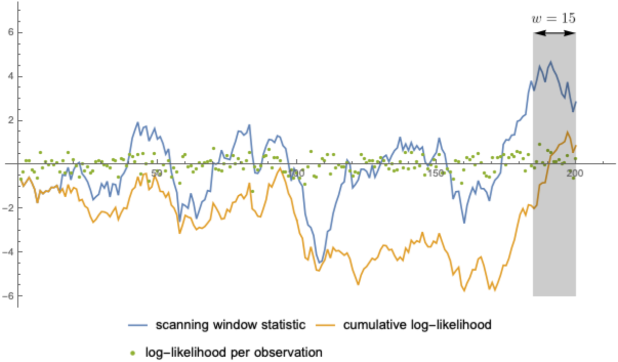
Page's procedure has asymptotically optimal worst case response delay under a patience constraint (Lorden, 1971).

Example of an online algorithm?

Let $p = 1$ and assume $\theta > 0$. Scanning window-based method with window width $w > 0$:

$$W_n := \sum_{i=n-w+1}^{n} \beta(X_i - \beta/2).$$

- Window size w needs to increase when β decreases.
- Storage requirement depends on β.



Example of a non-online algorithm

Let $p = 1$ and assume $\theta > 0$. Shiryaev–Roberts procedure (Shiryaev, 1963; Roberts, 1966):

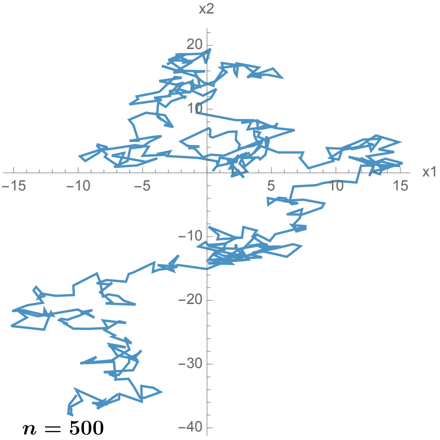
$$SR_n := \sum_{i=1}^{n} \prod_{h=i}^{n} e^{b(X_h - b/2)}.$$

- The statistics cannot be defined recursively
- A sequential but not online algorithm

A high-dimensional, multiscale online algorithm: `ocd`

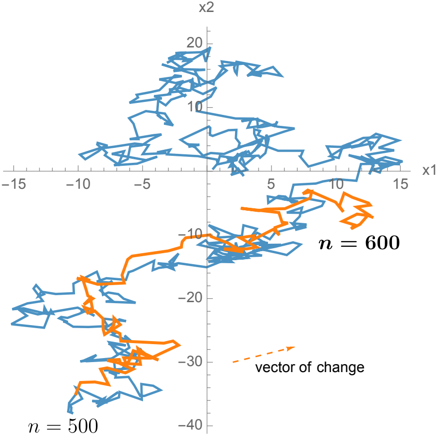
Curse of dimensionality

- Page's procedure in 1-d relies on the well-ordering of $\mathbb{R}$.
- Generalising to high dimensions:



Curse of dimensionality

- Page's procedure in 1-d relies on the well-ordering of $\mathbb{R}$.
- Generalising to high dimensions:

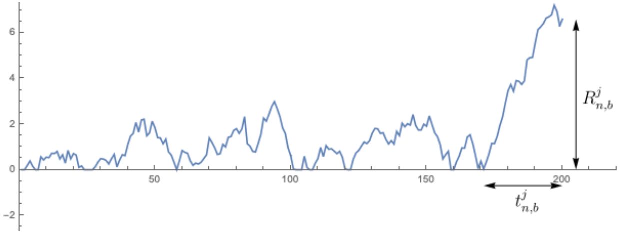


  - If we know the direction of θ, Page's procedure can still be used.
  - Infeasible to examine all possible directions for a change for large p.

- Write $X_i = (X_i^1, \dots, X_i^p)^\top \in \mathbb{R}^p$. Fix $n \in \mathbb{N}$ and $b \in \mathbb{R}\backslash\{0\}$. For each $j \in [p]$, define (we have suppressed n and b dependence)

$$R^j := \max_{0 \le h \le n} \sum_{i=n-h+1}^{n} b(X_i^j - b/2)$$

$$t^j := \operatorname*{argmax}_{0 \le h \le n} \sum_{i=n-h+1}^{n} b(X_i^j - b/2).$$

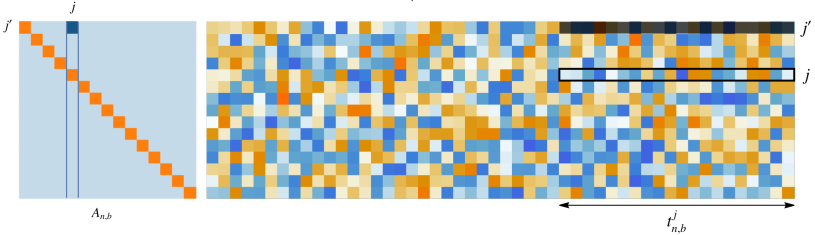


- $(R^j)_{j\in[p]}$ are called the **diagonal statistics**.

Off-diagonal statistics

- For each j, compute normalised tail partial sums of length t^j in all coordinates $j' \in [p]$:

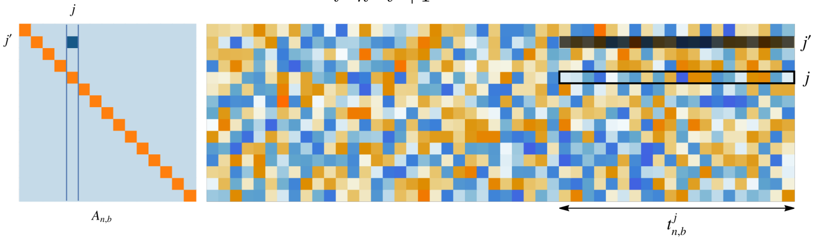
$$A^{j',j} := \frac{1}{\sqrt{t^j}} \sum_{i=n-t^j+1}^{n} X_i^{j'} \sim_{\text{null}} N(0,1).$$



Off-diagonal statistics

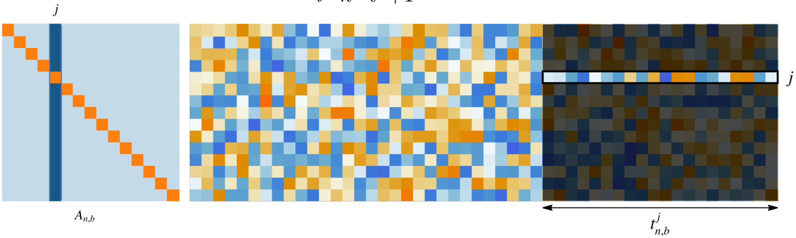
- For each j, compute normalised tail partial sums of length t^j in all coordinates $j' \in [p]$:

$$A^{j',j} := \frac{1}{\sqrt{t^j}} \sum_{i=n-t^j+1}^{n} X_i^{j'} \sim_{\text{null}} N(0,1).$$



- For each j, compute normalised tail partial sums of length t^j in all coordinates $j' \in [p]$:

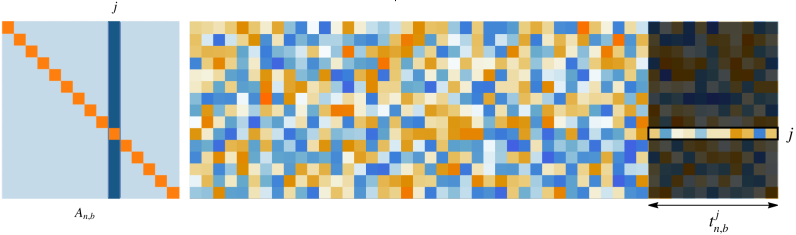
$$A^{j',j} := \frac{1}{\sqrt{t^j}} \sum_{i=n-t^j+1}^{n} X_i^{j'} \sim_{\text{null}} N(0,1).$$



Off-diagonal statistics

- For each j, compute normalised tail partial sums of length t^j in all coordinates $j' \in [p]$:

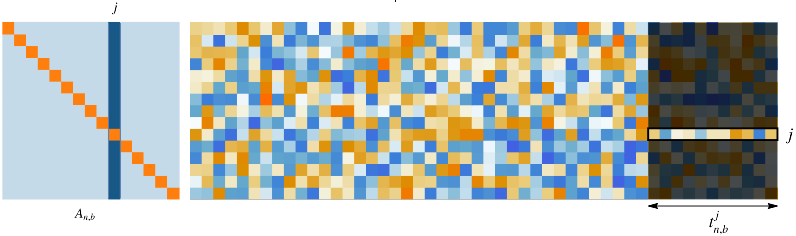
$$A^{j',j} := \frac{1}{\sqrt{t^j}} \sum_{i=n-t^j+1}^{n} X_i^{j'} \sim_{\text{null}} N(0,1).$$



Off-diagonal statistics

- For each j, compute normalised tail partial sums of length t^j in all coordinates $j' \in [p]$:

$$A^{j',j} := \frac{1}{\sqrt{t^j}} \sum_{i=n-t^j+1}^{n} X_i^{j'} \sim_{\text{null}} N(0,1).$$



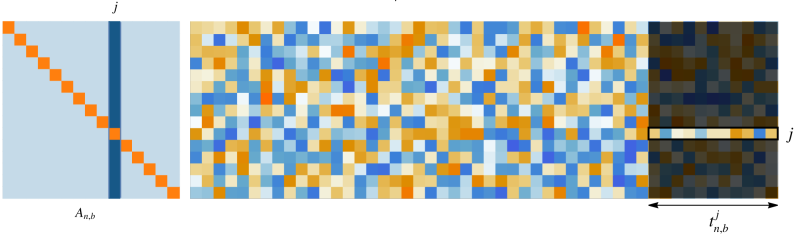
- We aggregate to form an **off-diagonal statistic** anchored at coordinate j:

$$Q^j := \sum_{j':j'\neq j} \left(A^{j',j}\right)^2 \sim_{\text{null}} \chi^2_{p-1}.$$

Off-diagonal statistics

- For each j, compute normalised tail partial sums of length t^j in all coordinates $j' \in [p]$:

$$A^{j',j} := \frac{1}{\sqrt{t^j}} \sum_{i=n-t^j+1}^{n} X_i^{j'} \sim_{\text{null}} N(0,1).$$



- We aggregate to form an **off-diagonal statistic** anchored at coordinate j:

$$Q^j := \sum_{j':j'\neq j} \left(A^{j',j} \mathbb{1}_{\{|A^{j',j}| \geq a\}} \right)^2 \quad \text{for some } a > 0.$$

Putting it together

UCL

- Vary the **scale** parameter b over a (signed) dyadic grid

$$\mathcal{B} := \left\{ \pm \frac{\beta}{\sqrt{2^{\ell} \log_2(2p)}} : \ell = 0, \ldots, \lfloor \log_2(2p) \rfloor \right\}.$$

- Aggregate diagonal and off-diagonal statistics from different coordinates and at different scales (recall R^j and Q^j both have n and b dependence):

$$S_n^{\text{diag}} := \max_{(j,b)\in[p]\times\mathcal{B}} R_{n,b}^j,$$
$$S_n^{\text{off}} := \max_{(j,b)\in[p]\times\mathcal{B}} Q_{n,b}^j.$$

- Declare change when either S_n^{diag} or S_n^{off} is large.

Algorithm 1: Pseudo-code of the `ocd` algorithm

Input: $X_1, X_2 \ldots \in \mathbb{R}^p$ observed sequentially, $\beta > 0$, $a \geq 0$, $T^{\mathrm{diag}} > 0$ and $T^{\mathrm{off}} > 0$

Set: $\mathcal{B} = \left\{\pm\frac{\beta}{\sqrt{2^\ell \log_2(2p)}} : \ell = 0, \ldots, \lfloor \log_2 p \rfloor\right\}$, $\mathcal{B}_0 = \left\{\pm\frac{\beta}{\sqrt{2^{\lfloor \log_2 p\rfloor+1}\log_2(2p)}}\right\}$, $n = 0$,

$A_b = \mathbf{0} \in \mathbb{R}^{p\times p}$ and $t_b = 0 \in \mathbb{R}^p$ for all $b \in \mathcal{B}\cup\mathcal{B}_0$

repeat

  $n \leftarrow n + 1$

  observe new data vector X_n

  for $(j, b) \in [p] \times (\mathcal{B}\cup\mathcal{B}_0)$ **do**

    $t_b^j \leftarrow t_b^j + 1$

    $A_b^{\cdot,j} \leftarrow A_b^{\cdot,j} + X_n$

    if $bA_b^{j,j} - b^2 t_b^j/2 \leq 0$ **then**

      $t_b^j \leftarrow 0$ and $A_b^{\cdot,j} \leftarrow 0$

    compute $Q_b^j \leftarrow \sum_{j'\in[p]:j'\neq j} \frac{(A_b^{j',j})^2}{t_b^j \vee 1} \mathbb{1}_{\left\{|A_b^{j',j}| \geq a\sqrt{t_b^j}\right\}}$

  $S^{\mathrm{diag}} \leftarrow \max_{(j,b)\in[p]\times(\mathcal{B}\cup\mathcal{B}_0)} \left(bA_b^{j,j} - b^2 t_b^j/2\right)$

  $S^{\mathrm{off}} \leftarrow \max_{(j,b)\in[p]\times\mathcal{B}} Q_b^j$

until $S^{\mathrm{diag}} \geq T^{\mathrm{diag}}$ or $S^{\mathrm{off}} \geq T^{\mathrm{off}}$;

Output: $N = n$

UCL

Algorithm 1: Pseudo-code of the `ocd` algorithm

Input: $X_1, X_2 \ldots \in \mathbb{R}^p$ observed sequentially, $\beta > 0$, $a \geq 0$, $T^{\mathrm{diag}} > 0$ and $T^{\mathrm{off}} > 0$

Set: $\mathcal{B} = \left\{ \pm \frac{\beta}{\sqrt{2^\ell \log_2(2p)}} : \ell = 0, \ldots, \lfloor \log_2 p \rfloor \right\}$, $\mathcal{B}_0 = \left\{ \pm \frac{\beta}{\sqrt{2^{\lfloor \log_2 p \rfloor + 1} \log_2(2p)}} \right\}$, $n = 0$,

$A_b = \mathbf{0} \in \mathbb{R}^{p \times p}$ and $t_b = 0 \in \mathbb{R}^p$ for all $b \in \mathcal{B} \cup \mathcal{B}_0$

Summary statistics stored

repeat

  $n \leftarrow n + 1$

  observe new data vector X_n

  for $(j, b) \in [p] \times (\mathcal{B} \cup \mathcal{B}_0)$ **do**

    $t_b^j \leftarrow t_b^j + 1$

    $A_b^{\cdot,j} \leftarrow A_b^{\cdot,j} + X_n$

    if $b A_b^{j,j} - b^2 t_b^j / 2 \leq 0$ **then**

      $t_b^j \leftarrow 0$ and $A_b^{\cdot,j} \leftarrow 0$

    compute $Q_b^j \leftarrow \sum_{j' \in [p] : j' \neq j} \frac{(A_b^{j',j})^2}{t_b^j \vee 1} \mathbb{1}_{\left\{ |A_b^{j',j}| \geq a \sqrt{t_b^j} \right\}}$

  $S^{\mathrm{diag}} \leftarrow \max_{(j,b) \in [p] \times (\mathcal{B} \cup \mathcal{B}_0)} \left(b A_b^{j,j} - b^2 t_b^j / 2 \right)$

  $S^{\mathrm{off}} \leftarrow \max_{(j,b) \in [p] \times \mathcal{B}} Q_b^j$

until $S^{\mathrm{diag}} \geq T^{\mathrm{diag}}$ or $S^{\mathrm{off}} \geq T^{\mathrm{off}}$;

Output: $N = n$

Complexity $O(p^2 \log_2(2p))$

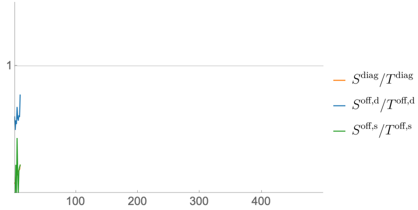
Dense, sparse and adaptive versions

- Choose $a = 0$ to detect a **dense** change.
- Choose $a = \sqrt{8 \log p}$ to detect a **sparse** change.

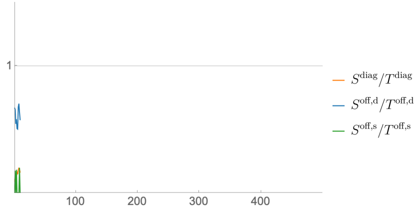
- The **adaptive** version runs two ocd algorithms with $a = 0$ and $a = \sqrt{8 \log p}$ in parallel and declares when either detects a change:

$$N := \inf\left\{n : \frac{S_n^{\mathrm{diag}}}{T^{\mathrm{diag}}} \vee \frac{S_n^{\mathrm{off,d}}}{T^{\mathrm{off,d}}} \vee \frac{S_n^{\mathrm{off,s}}}{T^{\mathrm{off,s}}} \geq 1\right\}.$$

Setting: $p = 100, z = 900, \vartheta = \beta = 1$ and $\gamma = 5000$



$s = 3$



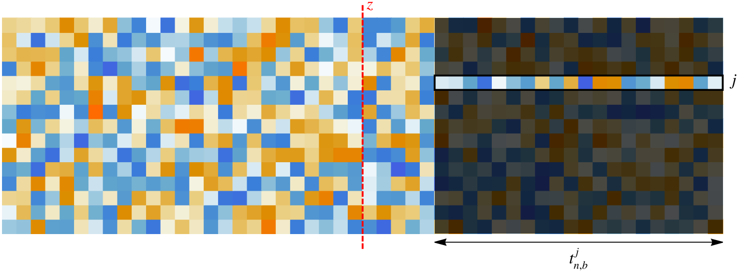
$s = 100$

Theoretical analysis

Why does ocd work?

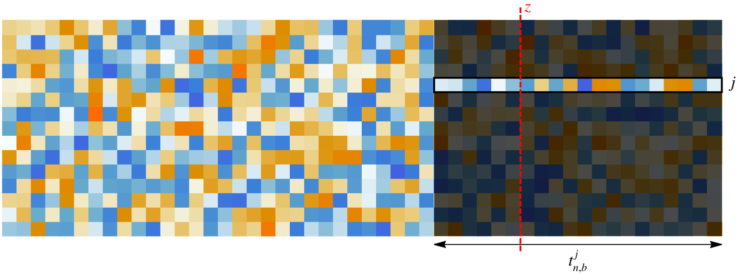
- **Patience**: guaranteed by choosing thresholds appropriately.
- **Response delay**:
  - S^{diag} detect changes that are concentrated in a single coordinate.
  - S^{off} aggregates signal across many coordinates.
  - If the tail partial sum consists of post-change data only, then

$$Q^j := \sum_{j':j' \neq j} \left(A^{j',j}\right)^2 \sim_{\text{altern.}} \chi^2_{p-1}(t^j \|\theta^{-j}\|_2^2).$$



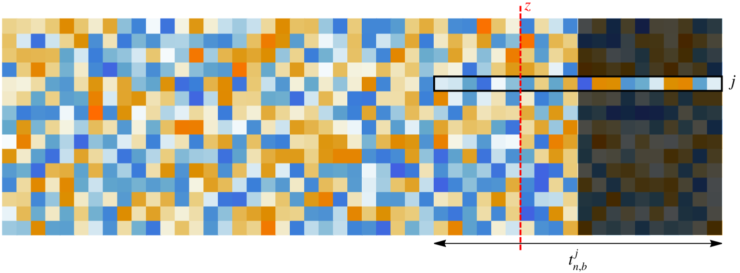
Halving the tail length

- What if The last t^j points contain some pre-change data?



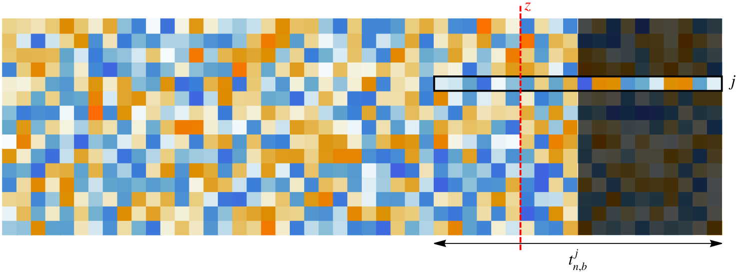
Halving the tail length

- What if The last t^j points contain some pre-change data?



- We would like to aggregate over $\approx t^j/2$ points, so that Q^j is eventually formed using post-change data only.

- What if The last t^j points contain some pre-change data?



- We would like to aggregate over $\approx t^j/2$ points, so that Q^j is eventually formed using post-change data only.

  How can we achieve this in an **online** manner?

A toy problem

Given $X_1, X_2, \ldots \in \mathbb{R}$, how can we keep track of the sum of the final $\tau \approx t/2$ observations at time t in an online way?

A toy problem

Given $X_1, X_2, \ldots \in \mathbb{R}$, how can we keep track of the sum of the final $\tau \approx t/2$ observations at time t in an online way?

$\mathbf{t}$	1	2	3	4	5	6	7	8	...
$\boldsymbol{\tau}$	1	1	2	2	3	4	5	4	...
$\boldsymbol{\Lambda}$	X_1	X_2	$X_2 + X_3$	$X_3 + X_4$	$X_3 + X_4 + X_5$	$X_3 + \cdots + X_6$	$X_3 + \cdots + X_7$	$X_5 + \cdots + X_8$	...
$\widetilde{\boldsymbol{\Lambda}}$	0	0	X_3	0	X_5	$X_5 + X_6$	$X_5 + X_6 + X_7$	0	...

$t/2 \leq \tau < 3t/4$ for $t \geq 2$.

A slight variant: ocd′

Part of the modified algorithm, ocd′, using ‘halved’ tail lengths:

for $(j, b) \in [p] \times (\mathcal{B} \cup \mathcal{B}_0)$ **do**

$t_b^j \leftarrow t_b^j + 1$ and $A_b^{\cdot,j} \leftarrow A_b^{\cdot,j} + X_n$

set $\delta = 0$ if t_b^j is a power of 2 and $\delta = 1$ otherwise.

$\tau_b^j \leftarrow \tau_b^j \delta + \tilde{\tau}_b^j (1 - \delta) + 1$ and $\Lambda_b^{\cdot,j} \leftarrow \Lambda_b^{\cdot,j} \delta + \tilde{\Lambda}_b^{\cdot,j} (1 - \delta) + X_n$

$\tilde{\tau}_b^j \leftarrow (\tilde{\tau}_b^j + 1)\delta$ and $\tilde{\Lambda}_b^{\cdot,j} \leftarrow (\tilde{\Lambda}_b^{\cdot,j} + X_n)\delta$.

track approx. half tail lengths

if $b A_b^{j,j} - b^2 t_b^j / 2 \leq 0$ **then**

$t_b^j \leftarrow \tau_b^j \leftarrow \tilde{\tau}_b^j \leftarrow 0$

$A_b^{\cdot,j} \leftarrow \Lambda_b^{\cdot,j} \leftarrow \tilde{\Lambda}_b^{\cdot,j} \leftarrow 0$

compute $Q_b^j \leftarrow \sum_{j' \in [p] : j' \neq j} \frac{(\Lambda_b^{j',j})^2}{\tau_b^j \vee 1} \mathbb{1}_{\left\{ |\Lambda_b^{j',j}| \geq a \sqrt{\tau_b^j} \right\}}$

use Λ to compute off-diag stats

Theoretical guarantees: patience

Choose thresholds

$$T^{\mathrm{diag}} = \log\{24p\gamma \log_2(4p)\}$$
$$T^{\mathrm{off,d}} = \psi\bigl(2\log\{24p\gamma \log_2(2p)\}\bigr)$$
$$T^{\mathrm{off,s}} = 8\log\{24p\gamma \log_2(2p)\}$$

where $\psi(x) = p - 1 + x + \sqrt{2(p-1)x}$ and $\gamma \geq 1$ is a user-specified **desired patience level**.

Theorem. Assume there is no change. Then, the adaptive version of $\mathtt{ocd}'$ with the above choice of thresholds satisfies $\mathbb{E}_0(N) \geq \gamma$.

Theoretical guarantees: response delay

The **effective sparsity** of $\theta \in \mathbb{R}^p$ is

$$s \equiv s(\theta) := \left| \left\{ j \in [p] : |\theta^j| \geq \frac{\|\theta\|_2}{\sqrt{s(\theta)\log_2(2p)}} \right\} \right|.$$

Theorem. Assume that the post-change signal θ satisfies $\|\theta\|_2 = \vartheta \geq \beta > 0$ with effective sparsity s. Then, the adaptive version of ocd′ with the same choice of thresholds satisfies:

(a) (Worst case response delay)

$$\bar{\mathbb{E}}_\theta^{\mathrm{wc}}(N) \lesssim \frac{s\log(ep\gamma)\log(ep)}{\beta^2} \vee 1;$$

(b) (Average case response delay)

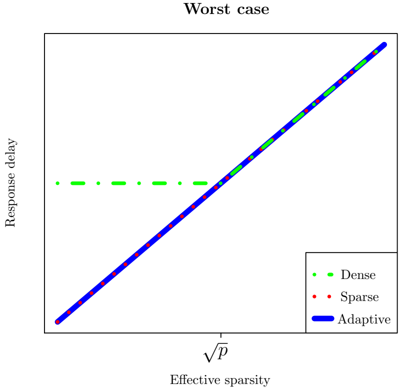
$$\bar{\mathbb{E}}_\theta(N) \lesssim \left(\frac{\sqrt{p}\log(ep\gamma)}{\vartheta^2} \vee \frac{\sqrt{s}\log(ep/\beta)\log(ep)}{\beta^2} \right) \wedge \frac{s\log(ep\gamma)\log(ep)}{\beta^2},$$

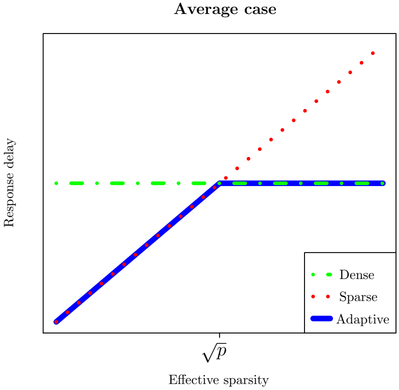
for all sufficiently small $\beta < \beta_0(s)$.

Response delays vs. sparsity

Assume that $\vartheta \asymp \beta \lesssim 1$ and $\log(\gamma/\beta) \lesssim \log p$. Then

$$\bar{\mathbb{E}}_\theta^{\mathrm{wc}}(N) \lesssim \frac{s \log^2(ep)}{\vartheta^2} \quad \text{and} \quad \bar{\mathbb{E}}_\theta(N) \lesssim \frac{(s \wedge \sqrt{p}) \log^2(ep)}{\vartheta^2}.$$





Comparison with other methods

We compare `ocd` with other recently proposed methods:

- `Mei`: ℓ_1 and ℓ_∞ aggregation of likelihood ratio tests in each coordinate. (Mei, 2010)
- `XS`: Use window-based method to aggregate statistics for testing the null against a normal mixture in each coordinate. (Xie and Siegmund, 2013)
- `Chan`: Similar to `XS`, but with an improved choice of tuning parameters. (Chan, 2017)

Simulation settings: $p \in \{100, 2000\}$, $s \in \{5, \lfloor\sqrt{p}\rfloor, p\}$, $\vartheta \in \{1, 0.5, 0.25\}$ and θ is generated as ϑU, where U is uniformly distributed on the union of all s sparse unit spheres in $\mathbb{R}^p$.

- All thresholds are determined using Monte Carlo simulation.

p	s	ϑ	ocd	Mei	XS	Chan
100	5	1	46.9	125.9	47.3	**42.0**
100	5	0.5	174.8	383.1	194.3	**163.7**
100	5	0.25	**583.5**	970.4	2147	1888.8
100	10	1	53.8	150.1	52.9	**51.5**
100	10	0.5	**194.4**	458.2	255.8	245.6
100	10	0.25	**629.7**	1171.3	2730.7	2484.9
100	100	1	**74.4**	268.3	89.6	102.1
100	100	0.5	**287.9**	834.9	526.8	756.0
100	100	0.25	**1005.8**	1912.9	3598.3	3406.6
2000	5	1	67.3	316.7	79.5	**59.5**
2000	5	0.5	**247.3**	680.2	607.7	285.0
2000	5	0.25	**851.3**	1384.8	4459.2	3856.9
2000	44	1	**136.0**	596.1	149.1	145.0
2000	44	0.5	**479.1**	1270.8	2945.5	2751.4
2000	44	0.25	**1584.2**	2428.8	4457.8	5049.7
2000	2000	1	**360.7**	2126.5	1020.0	2074.7
2000	2000	0.5	**1296.0**	3428.1	4669.3	4672.7
2000	2000	0.25	**3436.7**	4140.4	5063.7	5233.5

Table: Estimated response delay for ocd, Mei, XS and Chan over 200 repetitions, with $z = 0$ and $\gamma = 5000$.

Summary

- We propose a new, multiscale method for high-dimensional online changepoint detection.
- We perform likelihood ratio tests against simple alternatives of different scales in each coordinate, and aggregate these statistics.
- **R** package **ocd** is available on CRAN.

Main reference

- Chen, Y., Wang, T. and Samworth, R. J. (2020) High-dimensional, multiscale online changepoint detection. *Preprint*, arxiv:2003.03668.

References

- Barnard, G. A. (1959) Control charts and stochastic processes. *J. Roy. Statist. Soc., Ser. B*, **21**, 239–271.
- Baranowski, R., Chen, Y. and Fryzlewicz, P. (2019) Narrowest-Over-Threshold detection of multiple change points and change-point-like Features. *J. Roy. Statist. Soc., Ser. B*, **81**, 649–672.
- Chan, H. P. (2017) Optimal sequential detection in multi-stream data. *Ann. Statist.*, **45**, 2736–2763.
- Duncan, A. J. (1952) *Quality Control and Industrial Statistics*, Richard D. Irwin Professional Publishing Inc., Chicago.
- Fearnhead, P. and Liu, Z. (2007) On-line inference for multiple changepoint problems. *J. Roy. Statist. Soc., Ser. B*, **69**, 589–605.
- Killick, R., Fearnhead, P. and Eckley, I. A. (2012) Optimal detection of changepoints with a linear computational cost. *J. Amer. Stat. Assoc.*, **107**, 1590–1598.
- Mei, Y. (2010) Efficient scalable schemes for monitoring a large number of data streams. *Biometrika*, **97**, 419–433.
- Liu, H., Gao, C. and Samworth, R. J. (2019) Minimax rates in sparse, high-dimensional changepoint detection. *Preprint*, arxiv:1907.10012.
- Lorden, G. (1971) Procedures for reacting to a change in distribution. *Ann. Math. Statist.*, **42**, 1897–1908.

References

- Oakland, J. S. (2007) *Statistical Process Control* (6th ed.). Routledge, London.
- Page, E. S. (1954) Continuous inspection schemes. *Biometrika*, **41**, 100–115.
- Roberts, S. W. (1966) A comparison of some control chart procedures. *Technometrics*, **8**, 411–430.
- Shiryaev, A. N. (1963) On optimum methods in quickest detection problems. *Theory Probab. Appl.*, **8**, 22–46.
- Soh, Y. S. and Chandrasekaran, V. (2017) High-dimensional change-point estimation: Combining filtering with convex optimization. *Appl. Comp. Harm. Anal.*, **43**, 122–147.
- Tartakovsky, A., Nikiforov, I. and Basseville, M. (2014) *Sequential Analysis: Hypothesis testing and Changepoint Detection.* Chapman and Hall, London.
- Wang, T. and Samworth, R. J. (2018) High dimensional change point estimation via sparse projection. *J. Roy. Statist. Soc., Ser. B*, **80**, 57–83.
- Wang, D., Yu, Y. and Rinaldo, A. (2018) Univariate mean change point detection: penalization, CUSUM and optimality. *Preprint*, arxiv:1810.09498v4.
- Xie, Y. and Siegmund, D. (2013) Sequential multi-sensor change-point detection. *Ann. Statist.*, **41**, 670–692.
- Zou, C., Wang, Z., Zi, X. and Jiang, W. (2015) An efficient online monitoring method for high-dimensional data streams. *Technometrics*, **57**, 374–387.

Thank you!