Optimal multiple change-point detection and Localization

Nicolas Verzelen
INRAE, MISTEA, Montpellier

Joint work with A. Carpentier, M. Fromont,
M. Lerasle, E. Pilliat, and P. Reynaud-Bouret

https://arxiv.org/abs/2010.11470
https://arxiv.org/abs/2011.07818

RMR 2024 - June 20th

Offline Change-point Analysis

First Aim

In the mean change-point univariate setting :

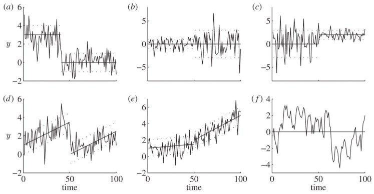
1. understand the **information-theoretical** scalings.
2. set up **specifications** for change-point procedures
3. exhibit several procedures achieving them.

Offline Change-point Analysis

First Aim

In the mean change-point univariate setting :

1. understand the **information-theoretical** scalings.
2. set up **specifications** for change-point procedures
3. exhibit several procedures achieving them.



Beaulieu et al.('12)

Second Aim

Start a **roadmap** towards more general models with e.g. **sparse multivariate change point** models.

(Sub)-Gaussian univariate mean change-point Model

Data : Time series $\mathbf{Y} \in \mathbb{R}^n$

$$y_i = \theta_i + \epsilon_i, \qquad \text{where } \epsilon_i \overset{ind.}{\sim} \mathcal{SG}(1) \ ,$$

where we assume that $\boldsymbol{\theta} \in \mathbb{R}^n$ is piece-wise constant.

We leave aside possible time dependencies

(Sub)-Gaussian univariate mean change-point Model

Data : Time series $\mathbf{Y} \in \mathbb{R}^n$

$$y_i = \theta_i + \epsilon_i, \qquad \text{where } \epsilon_i \overset{ind.}{\sim} \mathcal{SG}(1) \ ,$$

where we assume that $\boldsymbol{\theta} \in \mathbb{R}^n$ is piece-wise constant.

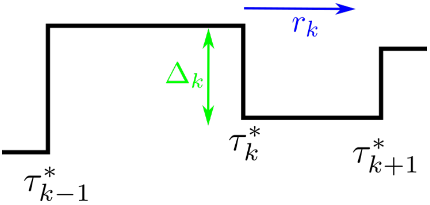
We leave aside possible time dependencies

Notation : change-point vector $\boldsymbol{\tau}^*$
$1 < \tau_1^* < \ldots < \tau_K^* \leq n$
s.t. $\boldsymbol{\theta}$ is constant over $[\tau_k^*, \tau_{k+1}^*)$.

Height $\Delta_k = \theta_{\tau_k^*} - \theta_{\tau_k^*-1}$

Radius $r_k = \frac{(\tau_{k+1}^* - \tau_k^*)(\tau_k^* - \tau_{k-1}^*)}{\tau_{k+1}^* - \tau_{k-1}^*}$

$\asymp (\tau_{k+1}^* - \tau_k^*) \wedge (\tau_k^* - \tau_{k-1}^*)$.



(Sub)-Gaussian univariate mean change-point Model

Data : Time series $\mathbf{Y} \in \mathbb{R}^n$

$$y_i = \theta_i + \epsilon_i, \qquad \text{where } \epsilon_i \overset{ind.}{\sim} \mathcal{SG}(1) \ ,$$

where we assume that $\boldsymbol{\theta} \in \mathbb{R}^n$ is piece-wise constant.

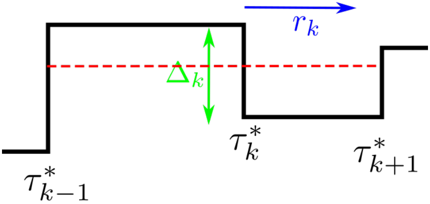
We leave aside possible time dependencies

Notation : change-point vector $\boldsymbol{\tau}^*$
$1 < \tau_1^* < \ldots < \tau_K^* \le n$
s.t. $\boldsymbol{\theta}$ is constant over $[\tau_k^*, \tau_{k+1}^*)$.

Height $\Delta_k = \theta_{\tau_k^*} - \theta_{\tau_k^*-1}$

Radius $r_k = \frac{(\tau_{k+1}^* - \tau_k^*)(\tau_k^* - \tau_{k-1}^*)}{\tau_{k+1}^* - \tau_{k-1}^*}$

$\asymp (\tau_{k+1}^* - \tau_k^*) \wedge (\tau_k^* - \tau_{k-1}^*)$.



Definition of the Energy of τ_k^*

The **Square Energy** of τ_k^* is $\qquad E_k^2 = r_k \Delta_k^2$

l_2 distance between $\boldsymbol{\theta}$ and best approximation by a piece-wise constant vector on $\boldsymbol{\tau}^{(-k)} = (\tau_1^*, \ldots, \tau_{k-1}^*, \tau_{k+1}^*, \ldots)$.

Two mathematical perspectives on change-point detection

- Denoising/Estimation : Estimating $\boldsymbol{\theta} \rightsquigarrow$ small risk $\mathbb{E}[\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2]$
- Clustering/Segmentation : Recover the change-points $\boldsymbol{\tau}^*$.

Two mathematical perspectives on change-point detection

- Denoising/Estimation : Estimating $\boldsymbol{\theta} \rightsquigarrow$ small risk $\mathbb{E}[\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2]$
- Clustering/Segmentation : Recover the change-points $\boldsymbol{\tau}^*$.

Denoising perspective :
Minimax-Optimal rates (for $K \geq 2$) $\qquad K\left[1 + \log\left(\frac{n}{K}\right)\right]$

achieved e.g. by penalized least-squares [Birgé and Massart, 2001, Gao et al., 2020]

Quadratic computational complexity by dynamic programming...

Change-point detection as a clustering problem

Several lines of literature :

- At Most One Change-point (AMOC) $[K \leq 1]$. **Least-square** estimator detects $\widehat{K} = 1$ if $E_1 \gg \sqrt{\log\log(n)}$ and $|\widehat{\tau}_1 - \tau_1^*| = O(\Delta_1^{-2})$ [Csorgo and Horváth, 1997]. V.et al.('20) $\rightsquigarrow$ detection iif $E_1 \geq \sqrt{2\log\log(\frac{n}{r_1})}$.

Change-point detection as a clustering problem

Several lines of literature :

- At Most One Change-point (AMOC) $[K \leq 1]$. **Least-square** estimator detects $\widehat{K} = 1$ if $E_1 \gg \sqrt{\log\log(n)}$ and $|\widehat{\tau}_1 - \tau_1^*| = O(\Delta_1^{-2})$ [Csorgo and Horváth, 1997]. V.et al.('20) $\rightsquigarrow$ detection iif $E_1 \geq \sqrt{2\log\log(\frac{n}{r_1})}$.
- Penalized Least-square Estimator. BIC penalty [Yao and Au, 1989, Wang et al., 2020].

Change-point detection as a clustering problem

Several lines of literature :

- At Most One Change-point (AMOC) $[K \leq 1]$. **Least-square** estimator detects $\widehat{K} = 1$ if $E_1 \gg \sqrt{\log\log(n)}$ and $|\widehat{\tau}_1 - \tau_1^*| = O(\Delta_1^{-2})$ [Csorgo and Horváth, 1997].
  V.et al.('20) $\rightsquigarrow$ detection iif $E_1 \geq \sqrt{2\log\log(\frac{n}{r_1})}$.
- Penalized Least-square Estimator. BIC penalty
  [Yao and Au, 1989, Wang et al., 2020].
- Greedy or Aggregation methods
  Binary segmentation [Scott and Knott, 1974] = iterative bisection.
  Many recent variants
  [Fryzlewicz, 2014, Fryzlewicz, 2018, Wang and Samworth, 2018]
  [Wang et al., 2020, Kovács et al., 2020, Cho and Kirch, 2019]

Change-point detection as a clustering problem

Several lines of literature :

- At Most One Change-point (AMOC) $[K \leq 1]$. **Least-square** estimator detects $\widehat{K} = 1$ if $E_1 \gg \sqrt{\log\log(n)}$ and $|\widehat{\tau}_1 - \tau_1^*| = O(\Delta_1^{-2})$ [Csorgo and Horváth, 1997]. V.et al.('20) $\rightsquigarrow$ detection iif $E_1 \geq \sqrt{2\log\log(\frac{n}{r_1})}$.
- Penalized Least-square Estimator. BIC penalty [Yao and Au, 1989, Wang et al., 2020].
- Greedy or Aggregation methods
  Binary segmentation [Scott and Knott, 1974] = iterative bisection.
  Many recent variants
  [Fryzlewicz, 2014, Fryzlewicz, 2018, Wang and Samworth, 2018]
  [Wang et al., 2020, Kovács et al., 2020, Cho and Kirch, 2019]

computational complexity $O(n\log(n))$.

Typical Results in the literature

Theorem (Typical modern result. *sloppy version* ; [Wang et al., 2020, Fryzlewicz, 2018, Kovács et al., 2020])

If $\min_k E_k^2 \gtrsim \log(n)$, *then whp* $\widehat{K} = K$ *and*

$$d_H(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*) = \max_{k=1,\dots,K} |\widehat{\tau}_k - \tau_k^*| \lesssim \frac{\log(n)}{\min_k \Delta_k^2}$$

But see [Frick et al., 2014] and [Cho and Kirch, 2019] for tighter results in different senses.

Typical Results in the literature

Theorem (Typical modern result. *sloppy version* ; [Wang et al., 2020, Fryzlewicz, 2018, Kovács et al., 2020])

If $\min_k E_k^2 \gtrsim \log(n)$, *then whp* $\widehat{K} = K$ *and*

$$d_H(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*) = \max_{k=1,\dots,K} |\widehat{\tau}_k - \tau_k^*| \lesssim \frac{\log(n)}{\min_k \Delta_k^2}$$

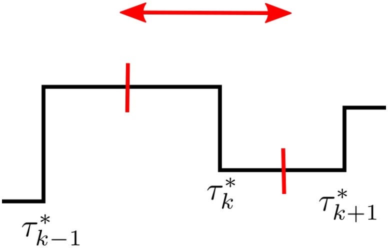
But see [Frick et al., 2014] and [Cho and Kirch, 2019] for tighter results in different senses.

Questions :

- Is $\min_k E_k^2 \gtrsim \log(n)$ really necessary ?
- What if a few change-points have a small energy ?
- Is the second $\log(n)$ necessary ?

Two sub-problems

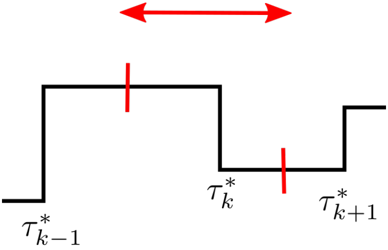
Change-Point Detection
= Detecting the existence of the change-point



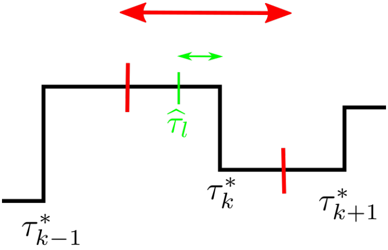
Objectives

Two sub-problems

Change-Point Detection
= Detecting the existence of the change-point



Change-Point Localization
= small estimation error
$d_{H,1}(\widehat{\boldsymbol{\tau}}, \tau^*_k) = \min_l |\widehat{\tau}_l - \tau^*_k|$

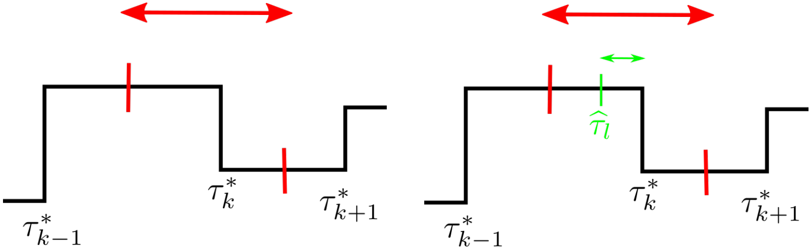


Objectives

Two sub-problems

Change-Point Detection
= Detecting the existence of the change-point

Change-Point Localization
= small estimation error
$d_{H,1}(\widehat{\boldsymbol{\tau}}, \tau_k^*) = \min_l |\widehat{\tau}_l - \tau_k^*|$

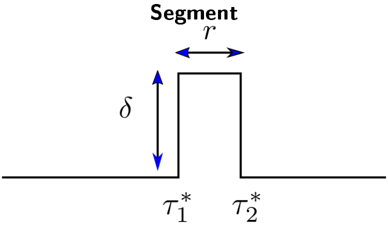


Questions

- What is the energy requirement for detection ?
- How is the transition between detection and localization ?
- Is penalized least-square optimal ? For which penalty ?

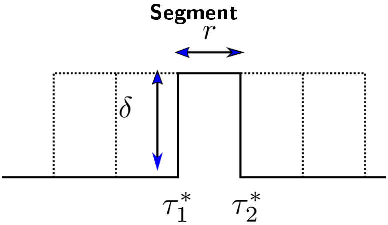
Simpler problem : testing $\boldsymbol{\theta} = 0$ versus

$$\boldsymbol{\theta} \in \Theta[r, \delta] = \left\{ \boldsymbol{\theta} \in \mathbb{R}^n : \exists \tau \text{ such that } \theta_i = \delta \mathbb{1}_{i \in [\tau, \tau + r)} \right\} .$$



Simpler problem : testing $\boldsymbol{\theta} = 0$ versus

$$\boldsymbol{\theta} \in \Theta[r,\delta] = \left\{\boldsymbol{\theta} \in \mathbb{R}^n : \exists \tau \in \{n/4, n/4+r, n/4+2r, \ldots, 3n/4\} \text{ such that } \theta_i = \delta \mathbb{1}_{i\in[\tau,\tau+r)}\right\} .$$



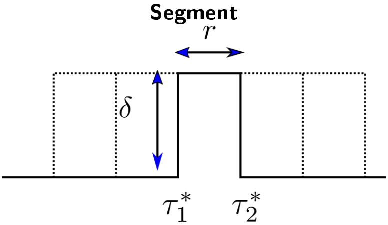
$\lfloor \frac{n}{2r} \rfloor$ possible positions

For each τ, sufficient statistic
$Z_\tau = r^{-1/2} \sum_{i=\tau}^{\tau+r-1} y_i \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(r^{1/2}\delta, 1) \end{cases}$

Gaussian Change-point Detection

Simpler problem : testing $\boldsymbol{\theta} = 0$ versus

$$\boldsymbol{\theta} \in \Theta[r,\delta] = \left\{\boldsymbol{\theta} \in \mathbb{R}^n : \exists \tau \in \{n/4, n/4{+}r, n/4{+}2r, \ldots, 3n/4\} \text{ such that } \theta_i = \delta \mathbb{1}_{i \in [\tau, \tau+r)}\right\} .$$



$\lfloor \frac{n}{2r} \rfloor$ possible positions

For each τ, sufficient statistic

$$Z_\tau = r^{-1/2} \sum_{i=\tau}^{\tau+r-1} y_i \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(r^{1/2}\delta, 1) \end{cases}$$

If $r \ll n$, then $r = r_1(1 + o(1)) = r_2(1 + o(1))$.

Proposition (Segment Detection ≈ [Arias-Castro et al., 2011])

If $\delta\sqrt{r} \le \sqrt{2(1 - o(1)) \log[n/(2r)]}$, *then testing* **better than random guess** *is impossible,*

$$\inf_T \mathbb{P}_0[T = 1] + \sup_{\boldsymbol{\theta} \in \Theta[\delta, r]} \mathbb{P}_{\boldsymbol{\theta}}[T = 0] \ge 1 - o(1) .$$

High-energy Change-points

$\kappa > 1$; $q > 0$.

Definition

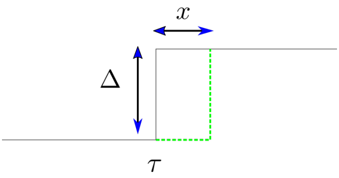
τ_k^* is a (κ, q)-**high-energy** change-point if $\mathbf{E}_k(\boldsymbol{\theta}) > \kappa\sqrt{2\log\left(\frac{n}{r_k}\right) + q}$.

Remarks :

- For small r_k, then $\log(n/r_k) \asymp \log(n)$.
- For small $r_k \asymp n$, then $\log(n/r_k) \asymp 1$
- The additive term q will play the role of a global probability.

Simplified setting :

- one change-point with known means $\boldsymbol{\mu} = (\mu_1, \mu_2)$
- Two possible positions for τ^* : τ or $\tau + x$.

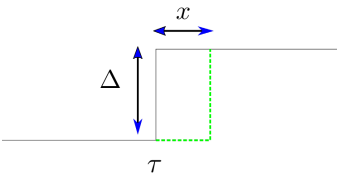


Sufficient statistic

$$Z = x^{-1/2} \sum_{i=\tau}^{\tau+x-1} (y_i - \mu_1) \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(x^{1/2}\Delta, 1) \end{cases}$$

Simplified setting :

- one change-point with known means $\boldsymbol{\mu} = (\mu_1, \mu_2)$
- Two possible positions for τ^* : τ or $\tau + x$.



Sufficient statistic

$$Z = x^{-1/2} \sum_{i=\tau}^{\tau+x-1} (y_i - \mu_1) \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(x^{1/2}\Delta, 1) \end{cases}$$

Lemma (Lower bound for Localization ≈ [Wang and Samworth, 2018])

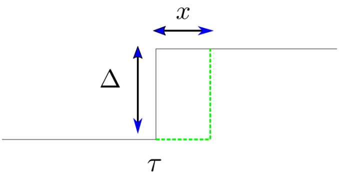
Write $\Delta = \mu_2 - \mu_1$. *For any* $r \geq 2$,

$$\inf_{\widehat{\tau}} \sup_{\tau^* \in \{2,\dots,n\}} \mathbb{P}_{\boldsymbol{\theta}(\tau^*, \boldsymbol{\mu})} \left(|\widehat{\tau} - \tau^*| \geq \frac{r}{\Delta^2} \right) \gtrsim e^{-cr} \quad ,$$

Gaussian Change-point Localization

Simplified setting :

- one change-point with known means $\boldsymbol{\mu} = (\mu_1, \mu_2)$
- Two possible positions for τ^* : τ or $\tau + x$.



Sufficient statistic

$Z = x^{-1/2} \sum_{i=\tau}^{\tau+x-1} (y_i - \mu_1) \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(x^{1/2}\Delta, 1) \end{cases}$

Lemma (Lower bound for Localization ≈ [Wang and Samworth, 2018])

Write $\Delta = \mu_2 - \mu_1$. *For any* $r \geq 2$,

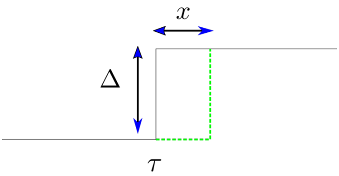
$$\inf_{\hat{\tau}} \sup_{\tau^* \in \{2,\dots,n\}} \mathbb{P}_{\boldsymbol{\theta}(\tau^*, \boldsymbol{\mu})} \left(|\hat{\tau} - \tau^*| \geq \frac{r}{\Delta^2} \right) \gtrsim e^{-cr} ,$$

Small Δ : At best, $|\hat{\tau} - \tau^*| \asymp \Delta^{-2}$ and has a sub-exponential tail.

Gaussian Change-point Localization

Simplified setting :

- one change-point with known means $\boldsymbol{\mu} = (\mu_1, \mu_2)$
- Two possible positions for τ^* : τ or $\tau + x$.



Sufficient statistic

$$Z = x^{-1/2} \sum_{i=\tau}^{\tau+x-1} (y_i - \mu_1) \sim \begin{cases} \mathcal{N}(0,1) \\ \mathcal{N}(x^{1/2}\Delta, 1) \end{cases}$$

Lemma (Lower bound for Localization ≈ [Wang and Samworth, 2018])

Write $\Delta = \mu_2 - \mu_1$. *For any* $r \geq 2$,

$$\inf_{\widehat{\tau}} \sup_{\tau^* \in \{2,\dots,n\}} \mathbb{P}_{\boldsymbol{\theta}(\tau^*, \boldsymbol{\mu})} \left(|\widehat{\tau} - \tau^*| \geq \frac{r}{\Delta^2} \right) \gtrsim e^{-cr} \quad ,$$

Small Δ : At best, $|\widehat{\tau} - \tau^*| \asymp \Delta^{-2}$ and has a sub-exponential tail.

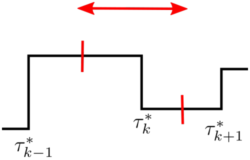
Large Δ : At best, $\widehat{\tau} = \tau^*$ with proba higher than $1 - c' e^{-c\Delta^2}$.

Desiderata for a suitable change-point procedure

Under an event $\mathcal{A}$ of high (*to be discussed*) probability, then $\widetilde{\boldsymbol{\tau}}$

(NoSp). No spurious change-point is detected :

$$\begin{cases} \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau^*_{k-1}+\tau^*_k}{2}, \frac{\tau^*_k+\tau^*_{k+1}}{2}\right]\right| \leq 1 \ , \ \text{for all } k \text{ in } \{2,\ldots,K-1\} \ ; \\ \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left[2, \frac{\tau^*_1+\tau^*_2}{2}\right]\right| \leq 1 \ ; \ \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau^*_{K-1}+\tau^*_K}{2}, n\right]\right| \leq 1 \ . \end{cases}$$



Desiderata for a suitable change-point procedure

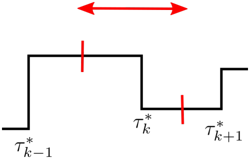
Under an event $\mathcal{A}$ of high (*to be discussed*) probability, then $\widetilde{\boldsymbol{\tau}}$

(NoSp). No spurious change-point is detected :

$$\begin{cases} \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau^*_{k-1}+\tau^*_k}{2}, \frac{\tau^*_k+\tau^*_{k+1}}{2}\right]\right| \leq 1 \;\; , \text{ for all } k \text{ in } \{2, \dots, K-1\} \;\; ; \\ \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left[2, \frac{\tau^*_1+\tau^*_2}{2}\right]\right| \leq 1 \;\; ; \;\; \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau^*_{K-1}+\tau^*_K}{2}, n\right]\right| \leq 1 \;\; . \end{cases}$$

(Detec). All high-energy change-points are detected.
For all k in $[K]$, if τ^*_k is a (κ, q)-high-energy change-point then

$$d_{H,1}(\widetilde{\boldsymbol{\tau}}, \tau^*_k) \leq \min\left\{\frac{\tau^*_{k+1} - \tau^*_k}{2}, \frac{\tau^*_k - \tau^*_{k-1}}{2}, c\frac{\log\left(1 \vee n\Delta_k^2\right) + q}{\Delta_k^2}\right\} \;\; .$$



Desiderata for a suitable change-point procedure

Under an event $\mathcal{A}$ of high (*to be discussed*) probability, then $\widetilde{\boldsymbol{\tau}}$

(NoSp). No spurious change-point is detected :

$$\begin{cases} \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau_{k-1}^*+\tau_k^*}{2}, \frac{\tau_k^*+\tau_{k+1}^*}{2}\right]\right| \leq 1 \ , \text{ for all } k \text{ in } \{2,\dots,K-1\} \ ; \\ \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left[2, \frac{\tau_1^*+\tau_2^*}{2}\right]\right| \leq 1 \ ; \ \left|\{\widetilde{\boldsymbol{\tau}}\} \cap \left(\frac{\tau_{K-1}^*+\tau_K^*}{2}, n\right]\right| \leq 1 \ . \end{cases}$$

(Detec). All high-energy change-points are detected.
For all k in $[K]$, if τ_k^* is a (κ, q)-high-energy change-point then

$$d_{H,1}(\widetilde{\boldsymbol{\tau}}, \tau_k^*) \leq \min\left\{\frac{\tau_{k+1}^* - \tau_k^*}{2}, \frac{\tau_k^* - \tau_{k-1}^*}{2}, c\frac{\log\left(1 \vee n\Delta_k^2\right) + q}{\Delta_k^2}\right\} \ .$$

(Loc). High-energy change-points are localized at the optimal rate.
Any high-energy change-point τ_k^* satisfies

$$\mathbb{P}\left(d_{H,1}(\widetilde{\boldsymbol{\tau}}, \tau_k^*)\mathbb{1}_{\mathcal{A}} \geq c\frac{x}{\Delta_k^2}\right) \lesssim e^{-x}, \quad \forall x \geq 1 \ .$$

Penalized least-square estimator

$\boldsymbol{\tau}$= vector of tentative change-points
$\boldsymbol{\Pi}_{\boldsymbol{\tau}}$ = projector onto the space of piece-wise constant vectors with changes at $\boldsymbol{\tau}$

$$\widehat{\tau} = \arg\min_{\tau} \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}) = \arg\min_{\boldsymbol{\tau}} \|\mathbf{Y} - \boldsymbol{\Pi}_{\boldsymbol{\tau}} \mathbf{Y}\|^2 + L \; \mathrm{pen}_0(\boldsymbol{\tau}, q) \quad ,$$

BIC Penalty $\quad \mathrm{pen}_{BIC}(\boldsymbol{\tau}, q) = 2|\boldsymbol{\tau}| \log(n)$

Penalized least-square estimator

$\boldsymbol{\tau}$= vector of tentative change-points
$\boldsymbol{\Pi_\tau}$ = projector onto the space of piece-wise constant vectors with changes at $\boldsymbol{\tau}$

$$\widehat{\tau} = \arg\min_{\tau} \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}) = \arg\min_{\boldsymbol{\tau}} \|\mathbf{Y} - \boldsymbol{\Pi_\tau} \mathbf{Y}\|^2 + L \; \mathrm{pen}_0(\boldsymbol{\tau}, q) \;\; ,$$

BIC Penalty $\mathrm{pen}_{BIC}(\boldsymbol{\tau}, q) = 2|\boldsymbol{\tau}| \log(n)$

Multi-scale penalty $\mathrm{pen}_0(\boldsymbol{\tau}, q) = q|\boldsymbol{\tau}| + 2 \sum_{k=1}^{|\boldsymbol{\tau}|+1} \log\left(\frac{n}{\tau_k - \tau_{k-1}}\right).$

Penalized least-square estimator

$\boldsymbol{\tau}$= vector of tentative change-points
$\boldsymbol{\Pi}_{\boldsymbol{\tau}}$ = projector onto the space of piece-wise constant vectors with changes at $\boldsymbol{\tau}$

$$\widehat{\tau} = \arg\min_{\tau} \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}) = \arg\min_{\boldsymbol{\tau}} \|\mathbf{Y} - \boldsymbol{\Pi}_{\boldsymbol{\tau}}\mathbf{Y}\|^2 + L\ \mathrm{pen}_0(\boldsymbol{\tau}, q)\ ,$$

BIC Penalty $\mathrm{pen}_{BIC}(\boldsymbol{\tau}, q) = 2|\boldsymbol{\tau}|\log(n)$

Multi-scale penalty $\mathrm{pen}_0(\boldsymbol{\tau}, q) = q|\boldsymbol{\tau}| + 2\sum_{k=1}^{|\boldsymbol{\tau}|+1} \log\left(\frac{n}{\tau_k - \tau_{k-1}}\right).$

Remarks :

- Additive Penalty $\rightsquigarrow$ dynamic programming (and its refinements [Killick et al., 2012])
- Over-penalizes **small** segments.
- Differs from complexity penalties $\mathrm{pen}_{BM}(\boldsymbol{\tau}, q) = (|\boldsymbol{\tau}| + 1)(1 + \log(n/|\boldsymbol{\tau}|))$.

Connection between CUSUM and Least-square penalty

Definition (CUSUM Statistic)

For $\mathbf{t} = (t_1, t_2, t_3)$, $\mathbf{C}(\mathbf{Y}, \mathbf{t}) = \left[\overline{\mathbf{Y}}_{[t_2, t_3)} - \overline{\mathbf{Y}}_{[t_1, t_2)}\right] \sqrt{\frac{(t_2 - t_1)(t_3 - t_2)}{t_3 - t_1}}$

Connection between CUSUM and Least-square penalty

Definition (CUSUM Statistic)

For $\mathbf{t} = (t_1, t_2, t_3)$, $\mathbf{C}(\mathbf{Y}, \mathbf{t}) = \left[\overline{\mathbf{Y}}_{[t_2,t_3)} - \overline{\mathbf{Y}}_{[t_1,t_2)}\right] \sqrt{\frac{(t_2-t_1)(t_3-t_2)}{t_3-t_1}}$

Lemma (deletion of a change-point e.g. [Wang et al., 2020])

$\boldsymbol{\tau}^{(-l)} = (\tau_1, \ldots, \tau_{l-1}, \tau_{l+1}, \ldots)$

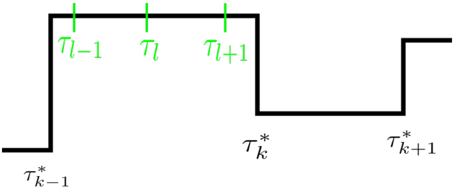
$$\|\mathbf{Y} - \mathbf{\Pi}_{\boldsymbol{\tau}} \mathbf{Y}\|^2 - \|\mathbf{Y} - \mathbf{\Pi}_{\boldsymbol{\tau}^{(-l)}} \mathbf{Y}\|^2 = -\mathbf{C}^2[\mathbf{Y}, (\tau_{l-1}, \tau_l, \tau_{l+1})] \ .$$

$$\begin{aligned} \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}) - \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}^{(-l)}) \quad &= \quad -\mathbf{C}^2\big(\mathbf{Y}, (\tau_{l-1}, \tau_l, \tau_{l+1})\big) \\ &\quad +L\left[2\log\left(\frac{n(\tau_{l+1} - \tau_{l-1})}{(\tau_{l+1} - \tau_l)(\tau_l - \tau_{l-1})}\right) + q\right] \ . \end{aligned}$$

Local Optimality and uniform Control of the CUSUM

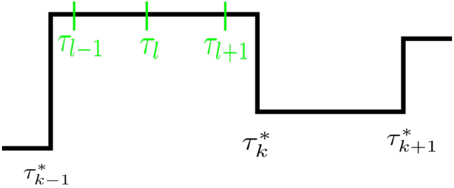
Consider $\boldsymbol{\tau}$ such that
$\boldsymbol{\theta}$ is constant on $[\tau_{l-1}, \tau_{l+1})$

Goal : show that $\boldsymbol{\tau} \neq \widehat{\boldsymbol{\tau}}$?



Consider $\boldsymbol{\tau}$ such that
$\boldsymbol{\theta}$ is constant on $[\tau_{l-1}, \tau_{l+1})$

Goal : show that $\boldsymbol{\tau} \neq \widehat{\boldsymbol{\tau}}$?



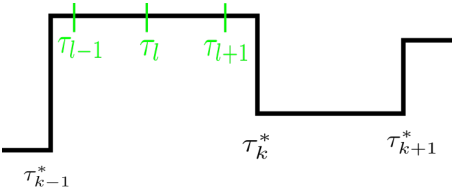
$$\mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}) - \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}^{(-l)}) = -\mathbf{C}^2\big(\boldsymbol{\epsilon}, (\tau_{l-1}, \tau_l, \tau_{l+1})\big) + L\left[2\log\left(\frac{n(\tau_{l+1} - \tau_{l-1})}{(\tau_{l+1} - \tau_l)(\tau_l - \tau_{l-1})}\right) + q\right] .$$

$\boldsymbol{\tau} \neq \widehat{\boldsymbol{\tau}}$ as long as $\mathbf{C}^2\big(\boldsymbol{\epsilon}, (\tau_{l-1}, \tau_l, \tau_{l+1})\big)$ small enough.

Local Optimality and uniform Control of the CUSUM

Consider $\boldsymbol{\tau}$ such that
$\boldsymbol{\theta}$ is constant on $[\tau_{l-1}, \tau_{l+1})$

Goal : show that $\boldsymbol{\tau} \neq \widehat{\boldsymbol{\tau}}$?



$$\mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}) - \mathrm{Cr}_0(\mathbf{Y}, \boldsymbol{\tau}^{(-l)}) = -\mathbf{C}^2(\boldsymbol{\epsilon}, (\tau_{l-1}, \tau_l, \tau_{l+1})) + L\left[2\log\left(\frac{n(\tau_{l+1} - \tau_{l-1})}{(\tau_{l+1} - \tau_l)(\tau_l - \tau_{l-1})}\right) + q\right] .$$

$\boldsymbol{\tau} \neq \widehat{\boldsymbol{\tau}}$ as long as $\mathbf{C}^2(\boldsymbol{\epsilon}, (\tau_{l-1}, \tau_l, \tau_{l+1}))$ small enough.

Local Optimality $\rightsquigarrow$ Uniform bound for the CUSUM

Lemma (Multi-scale chaining ; in the spirit of [Dumbgen and Spokoiny, 2001])

$$\mathcal{A}_q = \left\{ |\mathbf{C}(\boldsymbol{\epsilon}, \mathbf{t})| \leq 2\sqrt{2\log\left(\frac{n(t_3 - t_1)}{(t_3 - t_2)(t_2 - t_1)}\right) + q}, \quad \forall \mathbf{t} = (t_1, t_2, t_3) \right\} .$$

We have $\mathbb{P}[\mathcal{A}_q] \geq 1 - ce^{-c'q}$.

First Analysis of Penalized Least-square

High Energy condition : $E_k(\theta) \gtrsim \sqrt{\log(\frac{n}{r_k}) + q}$

Proposition (V. et al. ('20))

For any L and q large enough, under $\mathcal{A}_q$, the penalized least-square estimator $\widehat{\tau}$ satisfies

(a) **(NoSp)** *No Spurious Jump is detected.*

(b) **(Detec)** *All high-energy change-points τ_k^* are detected*

$$d_{H,1}(\widehat{\boldsymbol{\tau}}, \tau_k^*) \leq \min \left\{ \frac{\tau_{k+1}^* - \tau_k^*}{2}, \frac{\tau_k^* - \tau_{k-1}^*}{2}, \kappa_L \frac{\log\left(n\Delta_k^2\right) + q}{\Delta_k^2} \right\}$$

First Analysis of Penalized Least-square

High Energy condition : $E_k(\theta) \gtrsim \sqrt{\log(\frac{n}{r_k}) + q}$

Proposition (V. et al. ('20))

For any L and q large enough, under $\mathcal{A}_q$, the penalized least-square estimator $\widehat{\tau}$ satisfies

(a) **(NoSp)** *No Spurious Jump is detected.*

(b) **(Detec)** *All high-energy change-points τ_k^* are detected*

$$d_{H,1}(\widehat{\boldsymbol{\tau}}, \tau_k^*) \leq \min\left\{ \frac{\tau_{k+1}^* - \tau_k^*}{2}, \frac{\tau_k^* - \tau_{k-1}^*}{2}, \kappa_L \frac{\log\left(n\Delta_k^2\right) + q}{\Delta_k^2} \right\}$$

Remarks :

- Allow arbitrarily low-energy jumps.
- Local condition for high energy $\rightsquigarrow \log(n/r_k)$ (see also [Frick et al., 2014, Chan and Chen, 2017])
- Dependency in q is optimal with respect to the probability $1 - ce^{-c'q}$
- Complexity-based penalties are highly suboptimal.

Localization **(Loc)** by Penalized Least-squares

Proposition (V. et al. ('20))

Fix any L and q large enough. For any high-energy change-point τ_k^*, we have

$$\mathbb{P}\left(d_{H,1}(\widehat{\boldsymbol{\tau}}, \tau_k^*)\mathbb{1}_{\mathcal{A}_q} \geq c\frac{x}{\Delta_k^2}\right) \lesssim e^{-x} \quad \forall x \geq 1 \ .$$

Remarks :

- Recovers the optimal subexponential rate of order Δ_k^{-2} for a specific change-point
- **Regional** to **Local** phenomenon :
  Detection= High-Energy **Localization** only depends on Δ_k !

Localization **(Loc)** by Penalized Least-squares

Proposition (V. et al. ('20))

Fix any L and q large enough. For any high-energy change-point τ_k^*, we have

$$\mathbb{P}\left(d_{H,1}(\widehat{\boldsymbol{\tau}}, \tau_k^*)\mathbb{1}_{\mathcal{A}_q} \geq c\frac{x}{\Delta_k^2}\right) \lesssim e^{-x} \quad \forall x \geq 1 \ .$$

Remarks :

- Recovers the optimal subexponential rate of order Δ_k^{-2} for a specific change-point
- **Regional** to **Local** phenomenon :
  Detection= High-Energy   **Localization** only depends on Δ_k !
- Localization errors of high-energy change-points behave nearly **independently**.

Hausdorff and Wasserstein Loss

If $|\widehat{\boldsymbol{\tau}}| = |\boldsymbol{\tau}^*|$, define

$$d_W(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*) = \sum_{k=1}^{K} |\widehat{\tau}_k - \tau_k^*|$$

$$d_H(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*) = \max_{k=1}^{K} |\widehat{\tau}_k - \tau_k^*|$$

Corollary

Assuming that all change-points have high-energy, we deduce

$$\mathbb{E}\left[d_W\left(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*\right)\mathbb{1}_{\mathcal{A}_q}\right] \lesssim \sum_{k=1}^{K}\left(e^{-c''\Delta_k^2} \wedge \frac{1}{\Delta_k^2}\right) ,$$

$$\mathbb{E}\left[d_H\left(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*\right)\mathbb{1}_{\mathcal{A}_q}\right] \lesssim \max_{k\in\{1,\dots,K\}}\left(K e^{-c''\Delta_k^2} \wedge \frac{\log K}{\Delta_k^2}\right) .$$

Remark : Hausdorff and Wasserstein rates are minimax optimal.

Summary

Wrap-up :

- Regional to Local phenomenon.
- Low-energy change-points are (almost) unharmful.
- Localization errors behave almost independently.

Summary

Wrap-up :

- Regional to Local phenomenon.
- Low-energy change-points are (almost) unharmful.
- Localization errors behave almost independently.

One versus Multiple change-points.
When $K = 1$, $\log(n/r_k)$ conditions are replaced by $\log\log(n/r_1)$ conditions.

Summary

Wrap-up :

- Regional to Local phenomenon.
- Low-energy change-points are (almost) unharmful.
- Localization errors behave almost independently.

One versus Multiple change-points.
When $K = 1$, $\log(n/r_k)$ conditions are replaced by $\log\log(n/r_1)$ conditions.

Possible Extensions/ Open Questions :

- Heavier **tail** distribution, **time dependencies** :
  $\rightsquigarrow$ uniform control of the CUSUM (e.g.[Cho and Kirch, 2019])
- **Exact constant** for detection ?
- Similar results for two-steps bottom-up approach.

A general change-point framework

Data : Random sequence $Y = (y_1, y_2, \dots, y_n)$ in some measured space $\mathcal{Y}^n$.
Notation : $\mathbb{P}_i \in \mathcal{P}$ marginal distribution of y_i.
Change-point Functional : $\Gamma : \mathcal{P} \to \mathcal{V}$.

Change-Points : changes in the sequence $(\Gamma(\mathbb{P}_1), \Gamma(\mathbb{P}_2), \dots, \Gamma(\mathbb{P}_n))$

A general change-point framework

Data : Random sequence $Y = (y_1, y_2, \dots, y_n)$ in some measured space $\mathcal{Y}^n$.
Notation : $\mathbb{P}_i \in \mathcal{P}$ marginal distribution of y_i.
Change-point Functional : $\Gamma : \mathcal{P} \to \mathcal{V}$.

Change-Points : changes in the sequence $(\Gamma(\mathbb{P}_1), \Gamma(\mathbb{P}_2), \dots, \Gamma(\mathbb{P}_n))$

Examples :

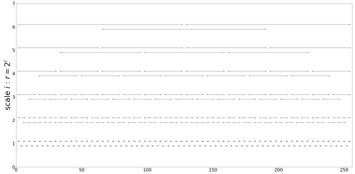
- **Gaussian mean Univariate change-point** :
  $\mathcal{Y} = \mathbb{R}$, $\mathcal{P} = \{\mathcal{N}(\theta, \sigma^2), \theta \in \mathbb{R}\}$, $\Gamma(\mathbb{P}) = \int x\mathbb{P}(dx)$
- **Gaussian mean multivariate change-point** [Chan and Chen, 2017, Wang and Samworth, 2018] :
  $\mathcal{Y} = \mathbb{R}^p$, $\mathcal{P} = \{\mathcal{N}(\theta, \sigma^2 \mathbf{I}_p), \theta \in \mathbb{R}^p\}$, $\Gamma(\mathbb{P}) = \int x\mathbb{P}(dx)$
- **Semi-parametric median (or quantile) univariate change-point** : [Jula Vanegas et al., 2021]
  $\mathcal{Y} = \mathbb{R}$, $\mathcal{P} = \{\text{Probability measure on } \mathbb{R}\}$, $\Gamma(\mathbb{P}) = median(\mathbb{P})$
- **Non-parametric univariate change-point** : [Padilla et al., 2019]
  $\mathcal{Y} = \mathbb{R}$, $\mathcal{P} = \{\text{Probability measure on } \mathbb{R}\}$, $\Gamma(\mathbb{P}) = \mathbb{P}$
- **Gaussian Covariance multivariate change-point** : [Wang et al., 2017]
  $\mathcal{Y} = \mathbb{R}^p$, $\mathcal{P} = \{\mathcal{N}(0, \Sigma), \Sigma \in \mathcal{S}_p^+\}$, $\Gamma(\mathbb{P}) = \int xx^T\mathbb{P}(dx)$

Goal : Detecting the change-points $\tau_1^*, \dots, \tau_K^*$
(leave aside the problem of localization)

A generic bottom-up algorithm

We are given :

- A collection (called a grid) $\mathcal{G}$ of (l, r) (location, scale) corresponding to segments $(l-r, l+r) \subset [1, n+1]$
  E.g. Complete Grid $\mathcal{G}_F = \{(l, r) : \ (l-r, l+r) \subset [1, n+1]\}$;
  Dyadic Grid $\mathcal{G}_D$.

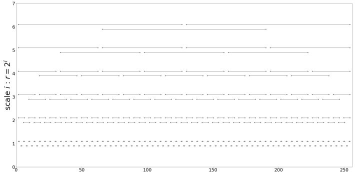


Dyadic grid

A generic bottom-up algorithm

We are given :

- A collection (called a grid) $\mathcal{G}$ of (l, r) (location, scale) corresponding to segments $(l-r, l+r) \subset [1, n+1]$
  E.g. Complete Grid $\mathcal{G}_F = \{(l,r) : (l-r, l+r) \subset [1, n+1]\}$;
  Dyadic Grid $\mathcal{G}_D$.
- A collection of local homogeneity tests $\mathcal{T} = (T_{l,r})$
  ($H_{0,l,r}$: no change-point in $(l-r, l+r)$)
  equivalent to $\Gamma(\mathbb{P}_{l-r}) = \Gamma(\mathbb{P}_{l-r+1}) = \ldots = \Gamma(\mathbb{P}_{l+r-1})$



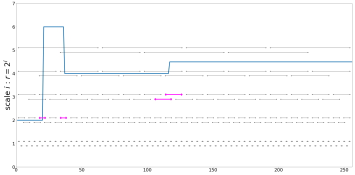
Dyadic grid

A generic bottom-up algorithm

We are given :

- A collection (called a grid) $\mathcal{G}$ of (l, r) (location, scale) corresponding to segments $(l-r, l+r) \subset [1, n+1]$
  E.g. Complete Grid $\mathcal{G}_F = \{(l,r) : \ (l-r, l+r) \subset [1, n+1]\}$;
  Dyadic Grid $\mathcal{G}_D$.
- A collection of local homogeneity tests $\mathcal{T} = (T_{l,r})$
  ($H_{0,l,r}$: no change-point in $(l-r, l+r)$)
  equivalent to $\Gamma(\mathbb{P}_{l-r}) = \Gamma(\mathbb{P}_{l-r+1}) = \ldots = \Gamma(\mathbb{P}_{l+r-1})$

Data: Local test $(T_{l,r})$
$\mathcal{CI} = \varnothing$; $\mathcal{CP} = \varnothing$;
For $r \in$ Scales
For $l \in$ Locations s.t. $T_{l,r} = 1$
if $[l-r+1, l+r-1] \cap \mathcal{CI} = \varnothing$
then
  $\mathcal{CI} \leftarrow \mathcal{CI} \cup [l-r+1, l+r-1]$;
  $\mathcal{CP} \leftarrow \mathcal{CP} \cup \{l\}$;
end
return $\widehat{\tau}_{ag} = \mathcal{CP}$

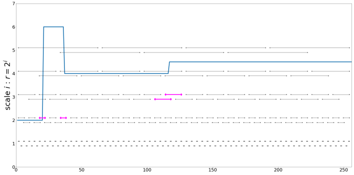


Dyadic grid

We are given :

- A collection (called a grid) $\mathcal{G}$ of (l, r) (location, scale) corresponding to segments $(l - r, l + r) \subset [1, n+1]$
  E.g. Complete Grid $\mathcal{G}_F = \{(l, r) : \; (l - r, l + r) \subset [1, n+1]\}$;
  Dyadic Grid $\mathcal{G}_D$.
- A collection of local homogeneity tests $\mathcal{T} = (T_{l,r})$
  ($H_{0,l,r}$: no change-point in $(l - r, l + r)$)
  equivalent to $\Gamma(\mathbb{P}_{l-r}) = \Gamma(\mathbb{P}_{l-r+1}) = \ldots = \Gamma(\mathbb{P}_{l+r-1})$

Data: Local test $(T_{l,r})$
$\mathcal{CI} = \varnothing$; $\mathcal{CP} = \varnothing$;
For $r \in$ Scales
For $l \in$ Locations s.t. $T_{l,r} = 1$
if $\underline{[l - r + 1, l + r - 1] \cap \mathcal{CI} = \varnothing}$ **then**
  $\mathcal{CI} \leftarrow \mathcal{CI} \cup [l - r + 1, l + r - 1]$;
  $\mathcal{CP} \leftarrow \mathcal{CP} \cup \{l\}$;
end
return $\widehat{\tau}_{ag} = \mathcal{CP}$



Dyadic grid

$O(n)$ tests for dyadic grids

Differs e.g. from [Chan and Chen, 2017, Kovács et al., 2020] because **intervals** of detected change-points are not allowed to intersect.

From multiple tests properties to detection properties

Proposition

1. *If* $FWER(\mathcal{T}) \leq \delta$*, then* $\widehat{\tau}_{ag}$ *satisfies* **(NoSp)** *with probability higher than* $1 - \delta$*.*
2. *Any change-point* τ_k^* *detected by a local test* $T_{l,\overline{r}_k}$ *with* $\overline{r}_k < r_k/4$*, is detected by* $\widehat{\tau}_{ag}$ *and*

$$d_{H,1}(\widehat{\tau}_{ag}, \tau_k^*) \leq \overline{r}_k - 1 \ .$$

From multiple tests properties to detection properties

Proposition

1. *If* $FWER(\mathcal{T}) \leq \delta$*, then* $\widehat{\tau}_{ag}$ *satisfies* $\mathbf{(NoSp)}$ *with probability higher than* $1 - \delta$.
2. *Any change-point* τ_k^* *detected by a local test* $T_{l,\overline{r}_k}$ *with* $\overline{r}_k < r_k/4$*, is detected by* $\widehat{\boldsymbol{\tau}}_{ag}$ *and*
$$d_{H,1}(\widehat{\tau}_{ag}, \tau_k^*) \leq \overline{r}_k - 1 \ .$$

$\mathcal{K}^*$ = collection of **significant** change-points
$\asymp$ "with proba higher than $1 - \delta$, all $\tau_k^* \in \mathcal{K}^*$ detected by a suitable local test"

$\leadsto$ With probability higher than $1 - 2\delta$, $\widehat{\tau}_{ag}$ satisfies **NoSp** and **Detects** significant change-points.

From multiple tests properties to detection properties

Proposition

1. *If* $FWER(\mathcal{T}) \leq \delta$, *then* $\widehat{\tau}_{ag}$ *satisfies* **(NoSp)** *with probability higher than* $1-\delta$.
2. *Any change-point* τ_k^* *detected by a local test* $T_{l,\bar{r}_k}$ *with* $\bar{r}_k < r_k/4$, *is detected by* $\widehat{\tau}_{ag}$ *and*
$$d_{H,1}(\widehat{\tau}_{ag}, \tau_k^*) \leq \bar{r}_k - 1 \ .$$

$\mathcal{K}^*$ = collection of **significant** change-points
$\asymp$ "with proba higher than $1-\delta$, all $\tau_k^* \in \mathcal{K}^*$ detected by a suitable local test"

$\rightsquigarrow$ With probability higher than $1-2\delta$, $\widehat{\tau}_{ag}$ satisfies **NoSp** and **Detects** significant change-points.

General recipe :

- Introducing a sensible notion of Energy
- Optimal testing with respect to that energy.
- Proper multiple testing correction to account for all tests in $\mathcal{G}_D$.

Gaussian Multivariate Change-point Model

$$y_i = \boldsymbol{\theta}_i + \epsilon_i, \qquad \text{where } \boldsymbol{\theta}_i \in \mathbb{R}^p \text{ and } \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I}_p).$$

Objective : Detecting **times** $\tau_1^*, \ldots, \tau_K^*$ such that $\boldsymbol{\theta}_{\tau_k^*} \neq \boldsymbol{\theta}_{\tau_k^*-1}$
(side information $\boldsymbol{\theta}_{\tau_k^*} - \boldsymbol{\theta}_{\tau_k^*-1}$ is possibly sparse)

[Wang and Samworth, 2018, Chan and Chen, 2017, Enikeeva and Harchaoui, 2019, Liu et al., 2019]

Energy of a Change-Point

$$E_k^2 = r_k \frac{\|\boldsymbol{\theta}_{\tau_k^*} - \boldsymbol{\theta}_{\tau_k^*-1}\|^2}{\sigma^2}$$

Local Homogeneity tests on $[l-r, l+r)$
1st Simplification : two-sample tests over data in $[l-r, l)$ versus $[l, l+r)$.

2nd Simplification : (possibly-sparse) signal detection test with multivariate CUSUM statistics

$$\mathbf{C}_{l,r} = \left[\overline{\mathbf{Y}}_{[l,l+r)} - \overline{\mathbf{Y}}_{[l-r,l)}\right] \frac{\sqrt{2r}}{\sigma} \sim \mathcal{N}\left[\left(\overline{\boldsymbol{\theta}}_{[l,l+r)} - \overline{\boldsymbol{\theta}}_{[l-r,l)}\right) \frac{\sqrt{2r}}{\sigma}, \mathbf{I}_p\right]$$

Energy and Optimal Tests for sparse high-dimensional data

Energy of a Change-Point

$$E_k^2 = r_k \frac{\|\boldsymbol{\theta}_{\tau_k^*} - \boldsymbol{\theta}_{\tau_k^*-1}\|^2}{\sigma^2}$$

Local Homogeneity tests on $[l-r, l+r)$
1st Simplification : two-sample tests over data in $[l-r, l)$ versus $[l, l+r)$.

2nd Simplification : (possibly-sparse) signal detection test with multivariate CUSUM statistics

$$\mathbf{C}_{l,r} = \left[\overline{\mathbf{Y}}_{[l,l+r)} - \overline{\mathbf{Y}}_{[l-r,l)}\right] \frac{\sqrt{2r}}{\sigma} \sim \mathcal{N}\left[\left(\overline{\boldsymbol{\theta}}_{[l,l+r)} - \overline{\boldsymbol{\theta}}_{[l-r,l)}\right) \frac{\sqrt{2r}}{\sigma}, \mathbf{I}_p\right]$$

Old toy detection Problem : [Baraud, 2002, Donoho and Jin, 2004, Collier et al., 2015]
$\rightsquigarrow$ Higher-Criticism + χ^2 type statistics (minimax optimal wrt sparsity s and p)

Energy and Optimal Tests for sparse high-dimensional data

Energy of a Change-Point

$$E_k^2 = r_k \frac{\|\boldsymbol{\theta}_{\tau_k^*} - \boldsymbol{\theta}_{\tau_k^*-1}\|^2}{\sigma^2}$$

Local Homogeneity tests on $[l-r, l+r)$
1st Simplification : two-sample tests over data in $[l-r, l)$ versus $[l, l+r)$.

2nd Simplification : (possibly-sparse) signal detection test with multivariate CUSUM statistics

$$\mathbf{C}_{l,r} = \left[\overline{\mathbf{Y}}_{[l,l+r)} - \overline{\mathbf{Y}}_{[l-r,l)}\right] \frac{\sqrt{2r}}{\sigma} \sim \mathcal{N}\left[\left(\overline{\boldsymbol{\theta}}_{[l,l+r)} - \overline{\boldsymbol{\theta}}_{[l-r,l)}\right) \frac{\sqrt{2r}}{\sigma}, \mathbf{I}_p\right]$$

Old toy detection Problem : [Baraud, 2002, Donoho and Jin, 2004, Collier et al., 2015]
$\rightsquigarrow$ Higher-Criticism + χ^2 type statistics (minimax optimal wrt sparsity s and p)

Not Sufficient : $\Omega(n)$ tests are considered
$\rightsquigarrow$ one also needs optimal dependencies wrt Types I and II error probabilities :
e.g. variants of HC [Liu et al., 2019] ; see also Pilliat et al.('20).

Optimal Detection

$\delta \in (0,1)$; s_k sparsity of change-point τ_k^*.

High-energy change-point

τ_k^* is a high-energy change-point if $E_k^2 \geq c\psi_{n,p,s_k,\delta}$ where

$$\psi_{n,p,s_k,\delta} = s_k \log\left(1 + \frac{\sqrt{p}}{s_k}\sqrt{\log\left(\frac{n}{r_k\delta}\right)}\right) + \log\left(\frac{n}{r_k\delta}\right) .$$

Optimal Detection

$\delta \in (0,1)$; s_k sparsity of change-point τ_k^*.

High-energy change-point

τ_k^* is a high-energy change-point if $E_k^2 \geq c\psi_{n,p,s_k,\delta}$ where

$$\psi_{n,p,s_k,\delta} = s_k \log\left(1 + \frac{\sqrt{p}}{s_k}\sqrt{\log\left(\frac{n}{r_k\delta}\right)}\right) + \log\left(\frac{n}{r_k\delta}\right) \ .$$

Theorem (Pilliat et al.('20))

With probability higher than $1-\delta$, $\widehat{\boldsymbol{\tau}}_{ag}$ *achieves* **(NoSp)** *and* **(Detects)** *all* **high-energy** *change-points* τ_k^* *with* $E_k^2 \geq c_+\psi_{n,p,s_k,\delta}$.

Optimal Detection

$\delta \in (0,1)$; s_k sparsity of change-point τ_k^*.

High-energy change-point

τ_k^* is a high-energy change-point if $E_k^2 \geq c\psi_{n,p,s_k,\delta}$ where

$$\psi_{n,p,s_k,\delta} = s_k \log\left(1 + \frac{\sqrt{p}}{s_k}\sqrt{\log\left(\frac{n}{r_k\delta}\right)}\right) + \log\left(\frac{n}{r_k\delta}\right) .$$

Theorem (Pilliat et al.('20))

With probability higher than $1-\delta$, $\widehat{\boldsymbol{\tau}}_{ag}$ achieves **(NoSp)** *and* **(Detects)** *all* **high-energy** *change-points τ_k^* with $E_k^2 \geq c_+\psi_{n,p,s_k,\delta}$.*

Conversely, no procedure achieving **(NoSp)** *is able to* **(Detect)** *high-energy change-points τ_k^* with $E_k^2 \geq c_-\psi_{n,p,s_k,\delta}$*

Remark : Generalizes asymptotic result of [Chan and Chen, 2017]. For $K \leq 1$, see [Liu et al., 2019].

Application 2 : Covariance change-point detection

Gaussian covariance Change-point Model [Wang et al., 2017]

$$y_i \sim \mathcal{N}(0, \Sigma_i) \ , \quad i = 1, \ldots, n$$

Objective : Detecting **times** $\tau_1^*, \ldots, \tau_K^*$ such that $\Sigma_{\tau_k^*} \neq \Sigma_{\tau_k^*-1}$
Side information : $\max_i \|\Sigma_i\|_{op} \leq B^2$ (for some known B)

Application 2 : Covariance change-point detection

Gaussian covariance Change-point Model [Wang et al., 2017]

$$y_i \sim \mathcal{N}(0, \Sigma_i) \ , \quad i = 1, \ldots, n$$

Objective : Detecting **times** $\tau_1^*, \ldots, \tau_K^*$ such that $\Sigma_{\tau_k^*} \neq \Sigma_{\tau_k^*-1}$

Side information : $\max_i \|\Sigma_i\|_{op} \leq B^2$ (for some known B)

Energy $E_k^2 := r_k \|\Sigma_{\tau_k^*} - \Sigma_{\tau_{k-1}^*}\|_{op}^2$.

Local test $T_{l,r} = \mathbb{1}\left\{\|\widehat{\Sigma}_{l,r} - \widehat{\Sigma}_{l,-r}\|_{op} \geq c_0 B^2 \left[\sqrt{\frac{p}{r}} + \frac{p}{r} + \sqrt{\frac{\log(\frac{2n}{\delta r})}{r}} + \frac{\log(\frac{2n}{\delta r})}{r}\right]\right\}$,

Optimal Detection

$\delta \in (0,1)$.

High-energy change-point

τ_k^* is a high-energy change-point if $E_k^2 \geq c\psi_{n,p,r_k}$ where

$$\psi_{n,p,r_k} := B^4 \left[\left(p + \log\left(\frac{2n}{r_k \delta} \right) \right) \wedge r_k \right]$$

Theorem (Pilliat et al.('20))

With probability higher than $1-\delta$, $\widehat{\boldsymbol{\tau}}_{ag}$ *achieves* **(NoSp)** *and* **(Detects)** *all* **high-energy** *change-points* τ_k^*, *with* $E_k^2 \geq c_+ \psi_{n,p,r_k}$

Optimal Detection

$\delta \in (0,1)$.

High-energy change-point

τ_k^* is a high-energy change-point if $E_k^2 \geq c\psi_{n,p,r_k}$ where

$$\psi_{n,p,r_k} := B^4\left[\left(p + \log\left(\frac{2n}{r_k\delta}\right)\right) \wedge r_k\right]$$

Theorem (Pilliat et al.('20))

With probability higher than $1-\delta$, $\widehat{\boldsymbol{\tau}}_{ag}$ *achieves* **(NoSp)** *and* **(Detects)** *all* **high-energy** *change-points* τ_k^*, *with* $E_k^2 \geq c_+\psi_{n,p,r_k}$

Conversely, no procedure achieving **(NoSp)** *is able to* **(Detect)** *high-energy change-points* τ_k^* *with* $E_k^2 \geq c_-\psi_{n,p,r_k}$

(Improves over the $p\log(n)$ condition of [Wang et al., 2017])

Discussion

Main message

A simple reduction of change-point detection problems to multiple testing problems

Discussion

Main message

A simple reduction of change-point detection problems to multiple testing problems

Sloppy optimality statement

If each test $T_{l,r}$ is **minimax optimal** (in some sense), then $\widehat{\boldsymbol{\tau}}_{ag}$ should (almost) optimally **detect** (in some sense).

Discussion

Main message

A simple reduction of change-point detection problems to multiple testing problems

Sloppy optimality statement

If each test $T_{l,r}$ is **minimax optimal** (in some sense), then $\widehat{\boldsymbol{\tau}}_{ag}$ should (almost) optimally **detect** (in some sense).

Open Questions

Localization rates require model-specific techniques.

For (sparse) high-dimensional change-points, there seem to exist several phase transitions from regional to local

References I

Arias-Castro, E., Candès, E. J., and Durand, A. (2011).
Detection of an anomalous cluster in a network.
Ann. Statist., 39(1) :278–304.

Baraud, Y. (2002).
Non-asymptotic minimax rates of testing in signal detection.
Bernoulli, 8(5) :577–606.

Birgé, L. and Massart, P. (2001).
Gaussian model selection.
J. Eur. Math. Soc. (JEMS), 3(3) :203–268.

Chan, H.-P. and Chen, H. (2017).
Multi-sequence segmentation via score and higher-criticism tests.
arXiv preprint arXiv :1706.07586.

Cho, H. and Kirch, C. (2019).
Localised pruning for data segmentation based on multiscale change point procedures.
arXiv preprint arXiv :1910.12486.

Collier, O., Comminges, L., and Tsybakov, A. B. (2015).
Minimax estimation of linear and quadratic functionals on sparsity classes.
arXiv preprint arXiv :1502.00665.

Csorgo, M. and Horváth, L. (1997).
Limit theorems in change-point analysis.
John Wiley & Sons Chichester.

Donoho, D. and Jin, J. (2004).
Higher criticism for detecting sparse heterogeneous mixtures.
Ann. Statist., 32(3) :962–994.

Dumbgen, L. and Spokoiny, V. G. (2001).
Multiscale testing of qualitative hypotheses.
Annals of Statistics, pages 124–152.

Enikeeva, F. and Harchaoui, Z. (2019).
High-dimensional change-point detection under sparse alternatives.
Ann. Statist., 47(4) :2051–2079.

Frick, K., Munk, A., and Sieling, H. (2014).
Multiscale change point inference.
Journal of the Royal Statistical Society : Series B (Statistical Methodology), 76(3) :495–580.

Fryzlewicz, P. (2014).
Wild binary segmentation for multiple change-point detection.
The Annals of Statistics, 42(6) :2243–2281.

References III

Fryzlewicz, P. (2018).
Tail-greedy bottom-up data decompositions and fast multiple change-point detection.
The Annals of Statistics, 46(6B) :3390–3421.

Gao, C., Han, F., and Zhang, C.-H. (2020).
On estimation of isotonic piecewise constant signals.
Ann. Statist., 48(2) :629–654.

Jula Vanegas, L., Behr, M., and Munk, A. (2021).
Multiscale quantile segmentation.
Journal of the American Statistical Association, pages 1–14.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012).
Optimal detection of changepoints with a linear computational cost.
Journal of the American Statistical Association, 107(500) :1590–1598.

Kovács, S., Li, H., Bühlmann, P., and Munk, A. (2020).
Seeded binary segmentation : A general methodology for fast and optimal change point detection.
arXiv preprint arXiv :2002.06633.

Liu, H., Gao, C., and Samworth, R. J. (2019).
Minimax rates in sparse, high-dimensional changepoint detection.
arXiv preprint arXiv :1907.10012.

Padilla, O. H. M., Yu, Y., Wang, D., and Rinaldo, A. (2019).
Optimal nonparametric multivariate change point detection and localization.
arXiv preprint arXiv :1910.13289.

Scott, A. J. and Knott, M. (1974).
A cluster analysis method for grouping means in the analysis of variance.
Biometrics, pages 507–512.

Wang, D., Yu, Y., and Rinaldo, A. (2017).
Optimal covariance change point localization in high dimension.
arXiv preprint arXiv :1712.09912.

Wang, D., Yu, Y., and Rinaldo, A. (2020).
Univariate mean change point detection : Penalization, CUSUM and optimality.
Electron. J. Stat., 14(1) :1917–1961.

Wang, T. and Samworth, R. J. (2018).
High dimensional change point estimation via sparse projection.
J. R. Stat. Soc. Ser. B. Stat. Methodol., 80(1) :57–83.

Yao, Y.-C. and Au, S. T. (1989).
Least-squares estimation of a step function.
Sankhyā Ser. A, 51(3) :370–381.