Multiple change-points detection in multiple signal

Kevin Bleakley and Jean-Philippe Vert

Mines ParisTech / Curie Institute / Inserm

Multiple change-points detection in 1 signal
Multiple change-points detection in 1 signal

![Graph of a signal with multiple change-points detected](image)
Multiple change-points detection in many signals

K. Bleakley and J.P Vert (ParisTech)
Why we care?

- Joint segmentation should increase the statistical power
- Applications:
  - multi-dimensional signals (multimedia, sensors...)
  - genomic profiles
Chromosomal aberrations in cancer
Comparative Genomic Hybridization (CGH)

Jain et al. Genome research 2002 12:325-332
A collection of bladder tumours
Typical applications

- Find **frequent breakpoints** in a collection of tumours (fusion genes...)
- **Low-dimensional summary and visualization** of the set of profiles

Detection of **frequently altered regions**
What we want

1. An algorithm that scales in time and memory to
   - Profiles length: \( n = 10^6 \sim 10^9 \)
   - Number of profiles (dimension): \( p = 10^2 \sim 10^3 \)
   - Number of change-points: \( k = 10^2 \sim 10^3 \)

2. A method with good statistical properties when \( p \) increases for \( n \) fixed (opposite to most existing literature).
Y ∈ ℝⁿ×ᵖ the signals

Define a piecewise constant approximation ̂U ∈ ℝⁿ×ᵖ of Y with k change-points as the solution of

\[ \min_{U \in \mathbb{R}^{n \times p}} \| Y - U \|_2^2 \quad \text{such that} \quad \sum_{i=1}^{n-1} 1(U_{i+1,\bullet} \neq U_{i,\bullet}) \leq k \]

DP finds the solution in \(O(n^2kp)\) in time and \(O(n^2)\) in memory

Does not scale to \(n = 10^6 \sim 10^9\)...
TV approximator for a single signal ($p = 1$)

- Replace

$$\min_{U \in \mathbb{R}^n} \| Y - U \|^2 \quad \text{such that} \quad \sum_{i=1}^{n-1} 1(U_{i+1} \neq U_i) \leq k$$

by

$$\min_{U \in \mathbb{R}^n} \| Y - U \|^2 \quad \text{such that} \quad \sum_{i=1}^{n-1} |U_{i+1} - U_i| \leq \mu$$

- An instance of total variation penalty (Rudin et al., 1992)
- Convex problem, fast implementations in $O(nK)$ or $O(n \log n)$ (Friedman et al., 2007; Harchaoui and Levy-Leduc, 2008; Hoefling, 2009)
TV approximator for many signals

- Replace

$$\min_{U \in \mathbb{R}^{n \times p}} \| Y - U \|^2 \text{ such that } \sum_{i=1}^{n-1} \mathbf{1}(U_{i+1,\bullet} \neq U_{i,\bullet}) \leq k$$

by

$$\min_{U \in \mathbb{R}^{n \times p}} \| Y - U \|^2 \text{ such that } \sum_{i=1}^{n-1} w_i \| U_{i+1,\bullet} - U_{i,\bullet} \| \leq \mu$$

Questions

- Practice: can we solve it efficiently?
- Theory: does it benefit from increasing $p$ (for $n$ fixed)?
TV approximator as a group Lasso problem

- Make the change of variables:
  \[ \gamma = U_{1,:} , \]
  \[ \beta_{i,:} = w_i \left( U_{i+1,:} - U_{i,:} \right) \quad \text{for } i = 1, \ldots, n - 1 . \]

- TV approximator is then equivalent to the following group Lasso problem (Yuan and Lin, 2006):
  \[
  \min_{\beta \in \mathbb{R}^{(n-1) \times p}} \| \bar{Y} - \bar{X} \beta \|^2 + \lambda \sum_{i=1}^{n-1} \| \beta_{i,:} \| ,
  \]
  where \( \bar{Y} \) is the centered signal matrix and \( \bar{X} \) is a particular \( (n - 1) \times (n - 1) \) design matrix.
The TV approximator can be solved efficiently:

- **approximately** with the group LARS in $O(npk)$ in time and $O(np)$ in memory
- **exactly** with a block coordinate descent + active set method in $O(np)$ in memory
Proof: computational tricks...

Although $\bar{X}$ is $(n - 1) \times (n - 1)$:

- For any $R \in \mathbb{R}^{n \times p}$, we can compute $C = \bar{X}^\top R$ in $O(np)$ operations and memory.
- For any two subset of indices $A = (a_1, \ldots, a_{|A|})$ and $B = (b_1, \ldots, b_{|B|})$ in $[1, n - 1]$, we can compute $\bar{X}_{\cdot, A}^\top \bar{X}_{\cdot, B}$ in $O(|A||B|)$ in time and memory.
- For any $A = (a_1, \ldots, a_{|A|})$, set of distinct indices with $1 \leq a_1 < \ldots < a_{|A|} \leq n - 1$, and for any $|A| \times p$ matrix $R$, we can compute $C = \left(\bar{X}_{\cdot, A}^\top \bar{X}_{\cdot, A}\right)^{-1} R$ in $O(|A|p)$ in time and memory.
Consistency for a single change-point

Suppose a single change-point:

- at position $u = \alpha n$
- with increments $(\beta_i)_{i=1,...,p}$ s.t. $\overline{\beta}^2 = \lim_{k \to \infty} \frac{1}{p} \sum_{i=1}^{k} \beta_i^2$
- corrupted by i.i.d. Gaussian noise of variance $\sigma^2$

Does the TV approximator correctly estimate the first change-point as $p$ increases?
Consistency of the unweighted TV approximator

$$\min_{U \in \mathbb{R}^{n \times p}} \| Y - U \|^2 \quad \text{such that} \quad \sum_{i=1}^{n-1} \| U_{i+1,\bullet} - U_{i,\bullet} \| \leq \mu$$

**Theorem**

The unweighted TV approximator finds the correct change-point with probability tending to 1 (resp. 0) as $p \to +\infty$ if $\sigma^2 < \tilde{\sigma}_\alpha^2$ (resp. $\sigma^2 > \tilde{\sigma}_\alpha^2$), where

$$\tilde{\sigma}_\alpha^2 = n\beta^2 \frac{(1 - \alpha)^2(\alpha - \frac{1}{2n})}{\alpha - \frac{1}{2} - \frac{1}{2n}}.$$ 

- correct estimation on $[n\epsilon, n(1 - \epsilon)]$ with $\epsilon = \sqrt{\frac{\sigma^2}{2n\beta^2}} + o(n^{-1/2}).$
- wrong estimation near the boundaries
Consistency of the weighted TV approximator

\[
\min_{U \in \mathbb{R}^{n \times p}} \| Y - U \|^2 \quad \text{such that} \quad \sum_{i=1}^{n-1} w_i \| U_{i+1} - U_i \| \leq \mu
\]

**Theorem**

The weighted TV approximator with weights

\[
\forall i \in [1, n-1], \quad w_i = \sqrt{\frac{i(n-i)}{n}}
\]

correctly finds the first change-point with probability tending to 1 as \( p \to +\infty \).

- we see the benefit of increasing \( p \)
- we see the benefit of adding weights to the TV penalty
The first change-point \( \hat{i} \) found by TV approximator maximizes

\[
F_i = \| \hat{c}_i \cdot \|_2^2,
\]

where

\[
\hat{c} = \bar{X}^\top \bar{Y} = \bar{X}^\top \bar{X} \beta^* + \bar{X}^\top W.
\]

\( \hat{c} \) is Gaussian, and \( F_i \) is follows a non-central \( \chi^2 \) distribution with

\[
G_i = EF_i \frac{p}{\sigma^2} = \frac{i(n - i)}{nw_i^2} \sigma^2 + \frac{\bar{\beta}^2}{w_i^2 w_u^2 n^2} \times \begin{cases} 
  i^2 (n - u)^2 & \text{if } i \leq u, \\
  u^2 (n - i)^2 & \text{otherwise}.
\end{cases}
\]

We then just check when \( G_u = \max_i G_i \)
Consistent estimation of more change-points?

\[ n = 100, \ k = 10, \bar{\beta}^2 = 1, \sigma^2 \in \{0.05; 0.2; 1\} \]

\[ U - \text{LARS} \quad W - \text{LARS} \quad U - \text{Lasso} \quad W - \text{Lasso} \]
Conclusion

- A new convex formulation for multiple change-point detection in multiple signals
- Better estimation with more signals
- Importance of weights
- Efficient approximate (gLARS) and exact (gLASSO) implementations; GLASSO more expensive but more accurate
- Consistency for the first $K > 1$ change-points observed experimentally but technically tricky to prove.