Fast Detection of Multiple Change-Points Shared by Many Signals

Jean-Philippe Vert and Kevin Bleakley

Mines ParisTech CBIO - Institut Curie - INSERM U900, Paris, France

Abstract

We present a fast algorithm for the detection of multiple change-points when each is frequently shared by members of a set of co-occurring one-dimensional signals. Conditions are given for consistency of both this algorithm and a fast approximation to it when the number of signals increases. Empirical evidence is provided to support the result.

Method

- Y a $n \times p$ matrix of signals (p signals of length n)
- For a single profile ($p = 1$), total variation (TV) denoising [1] or fused Lasso signal approximator (FLSA) [2] solve

$$\min_{U \in \mathbb{R}^n} \frac{1}{2} \| Y - U \|^2 + \lambda \sum_{i=1}^{n-1} | U_{i+1} - U_i | \,. \qquad (1)$$

- For multiple profiles ($p > 1$) we propose the following extension to find shared change-points:

$$\min_{U \in \mathbb{R}^{n \times p}} \frac{1}{2} \| Y - U \|^2 + \lambda \sum_{i=1}^{n-1} \| U_{i+1,\bullet} - U_{i,\bullet} \| \,. \qquad (2)$$

- The ℓ_1/ℓ_2 penalty on the increments enforces the same change-points for all profiles.
- More generally we consider the following weighted version:

$$\min_{U \in \mathbb{R}^{n \times p}} \frac{1}{2} \| Y - U \|^2 + \lambda \sum_{i=1}^{n-1} \sqrt{i(n-i)} \| U_{i+1,\bullet} - U_{i,\bullet} \| \,. \qquad (3)$$

- This is equivalent to a group Lasso regression problem [3] of the form

$$\min_{\beta \in \mathbb{R}^{(n-1) \times p}} \frac{1}{2} \| \bar{Y} - \bar{X}\beta \|^2 + \lambda \sum_{i=1}^{n-1} \| \beta_{i,\bullet} \| \,, \qquad (4)$$

for a particular design matrix $\bar{X}$, after the change of variable $\beta_{i,\bullet} = \sqrt{i(n-i)}\,(U_{i+1,\bullet} - U_{i,\bullet})$.

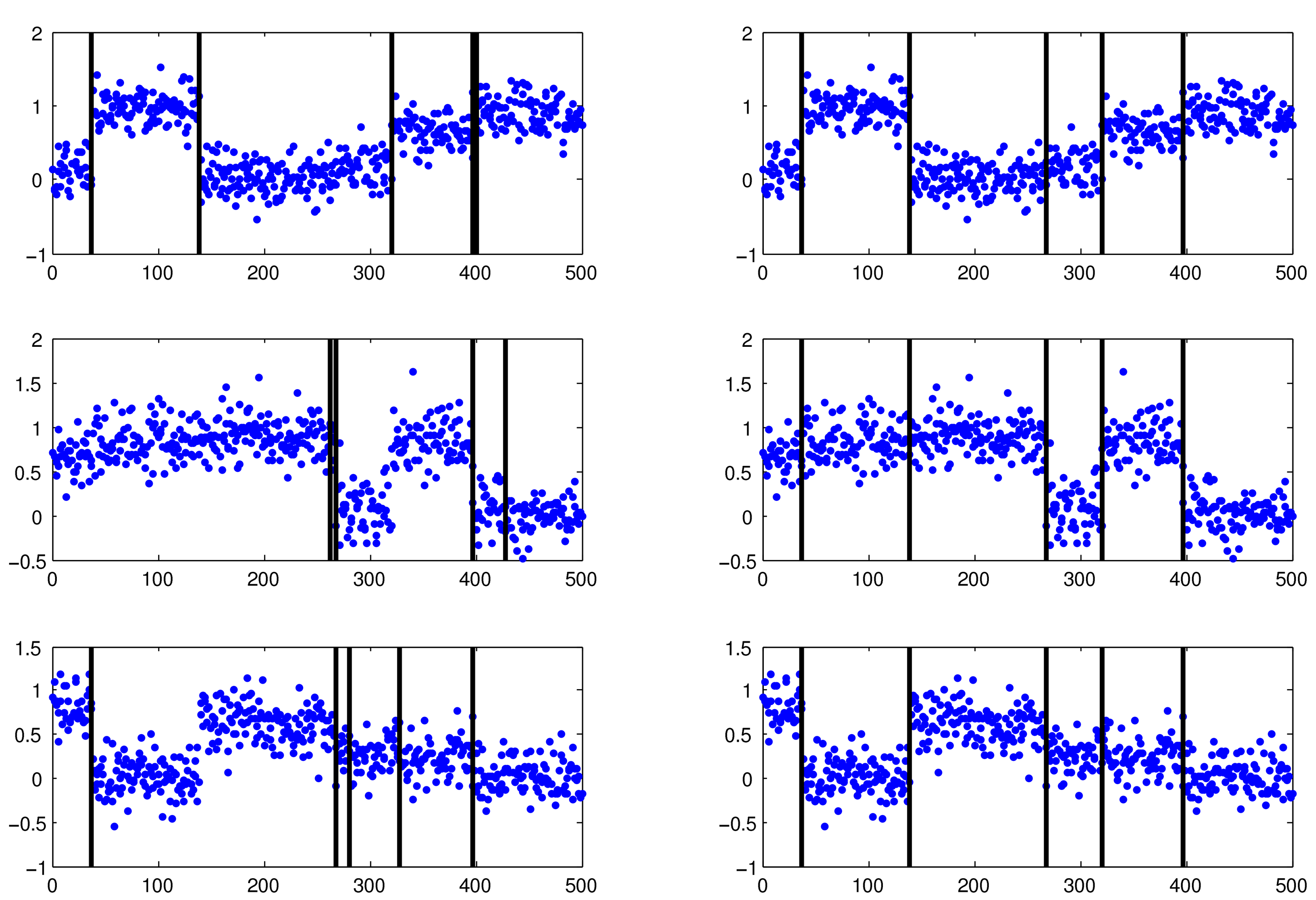

Figure 1: Change-point detection in 3 signals with the individual fused Lasso (left) and group fused Lasso (right).

Implementation

In spite of its size $n \times (n-1)$, the design matrix $\bar{X}$ leads to efficient computations thanks to the following technical results (in the spirit of [4]):

Lemma 1 • *For any $R \in \mathbb{R}^{n \times p}$, we can compute $C = \bar{X}^\top R$ in $O(np)$.*

- *For any two subset of indices $A = \left(a_1, \ldots, a_{|A|}\right)$ and $B = \left(b_1, \ldots, b_{|B|}\right)$ in $[1, n-1]$, the matrix $\bar{X}_{\bullet,A}^\top \bar{X}_{\bullet,B}$ can be computed in $O\left(|A||B|\right)$.*
- *For any $R \in \mathbb{R}^{(n-1) \times p}$, we can compute $C = \bar{X}^\top \bar{X} R$ in $O(np)$.*
- *For any $A = \left(a_1, \ldots, a_{|A|}\right)$, set of distinct indices with $1 \le a_1 < \ldots < a_{|A|} \le n-1$, the matrix $\left(\bar{X}_{\bullet,A}^\top \bar{X}_{\bullet,A}\right)$ is invertible, and for any $|A| \times p$ matrix R we can compute $\left(\bar{X}_{\bullet,A}^\top \bar{X}_{\bullet,A}\right)^{-1} R$ in $O(|A|p)$.*

This allows us to implement efficiently two methods to solve (4) efficiently, following [3]:

1. **Fast approximate solution with a group LARS**, where we approximate the regularization path by a piecewise affine path. The first k change-points are found in $O(npk)$.
2. **Exact solution by block coordinate descent**, which combined with an active set strategies leads to a complexity at least $O(npk + pk^3)$.

Consistency

While most theoretical work on change-point detection focuses on p fixed and increasing n, we ask the opposite question: for a given profile length n, does it help to increase the number of profiles p? Let us assume that there exists a single change-point at position u, and

$$\bar{Y} = \bar{X}\beta^* + W \,,$$

where W is an additive i.i.d. Gaussian noise of variance σ^2, β has a single non-zero row, and $\bar{\beta}^2 = \lim_{p \to \infty} \frac{1}{p} \sum_{i=1}^{p} (\beta^*_{u,i})^2$ exists. We can then characterize the first selected change-point:

Lemma 2 *Assume, without loss of generality, that $u \ge n/2$, and let*

$$G_i = d_i^2 \frac{i(n-i)}{n} \sigma^2 + \frac{\bar{\beta}^2 d_i^2 d_u^2}{n^2} \times \begin{cases} i^2 (n-u)^2 & \text{if } i \le u \,, \\ u^2 (n-i)^2 & \text{otherwise.} \end{cases} \qquad (5)$$

When $p \to +\infty$, the first change-point selected is in $\operatorname{argmax}_{i \in [1, n-1]} G_i$ with probability $\to 1$.

We can then deduce conditions for the unweighted and weighted fused Lasso to find the first change-point correctly

Theorem 3 • *Let $\alpha = u/n$ and*

$$\tilde{\sigma}_\alpha^2 = n\bar{\beta}^2 \frac{(1-\alpha)^2(\alpha - \frac{1}{2n})}{\alpha - \frac{1}{2} - \frac{1}{2n}} \,. \qquad (6)$$

When $\sigma^2 < \tilde{\sigma}_\alpha^2$, the probability that the first change-point selected by the unweighted group fused Lasso (2) is the correct one tends to 1 as $p \to +\infty$. When $\sigma^2 > \tilde{\sigma}_\alpha^2$, it is not the correct one with probability tending to 1.

- *The weighted group fused Lasso (3) correctly finds the first change-point with probability tending to 1 as $p \to +\infty$.*

References

[1] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[2] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(1):91–108, 2005.

[3] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B*, 68(1):49–67, 2006.

[4] Z. Harchaoui and C. Levy-Leduc. Catching change-points with lasso. In *NIPS 20*, 2008.

Experiments

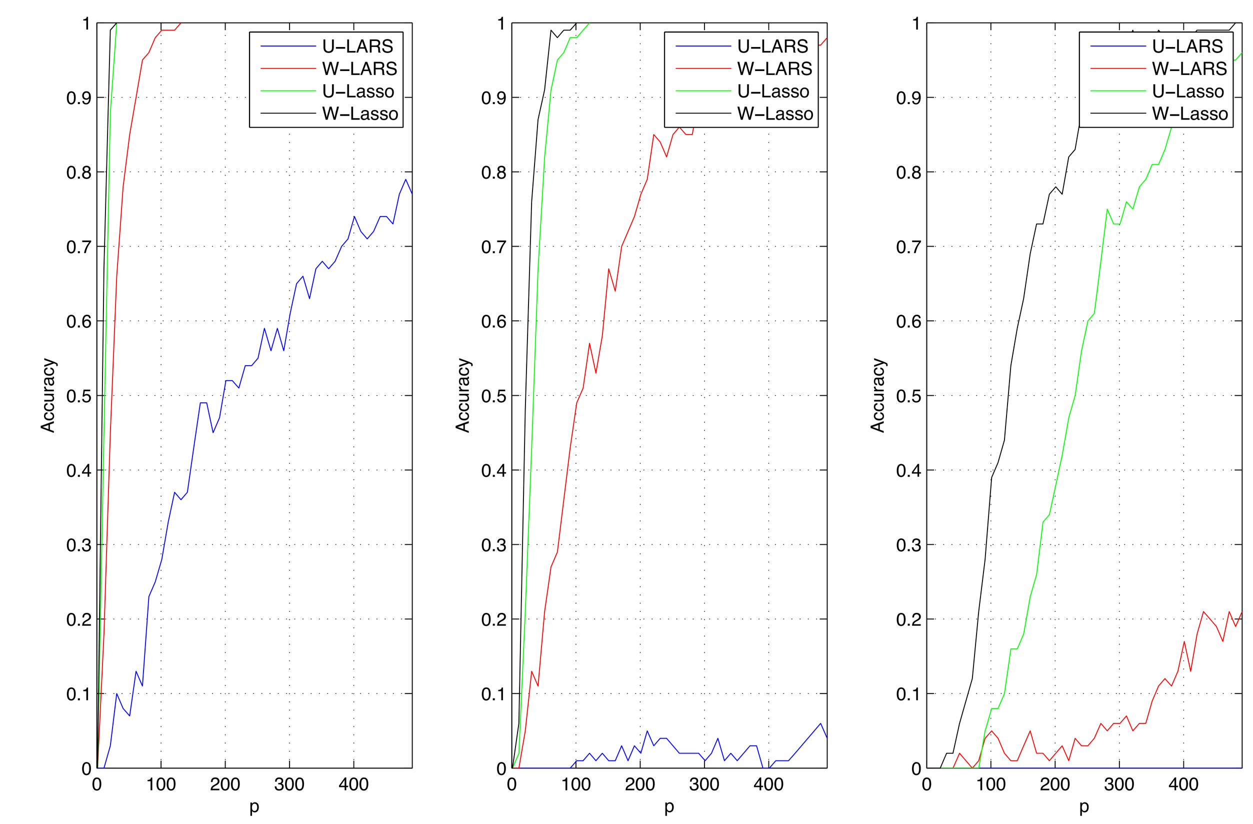

Figure 2: Probability to correctly identify the first 10 change-points on simulated data, as a function of p. The three pictures (from left to right) correspond to increasing noise. We compare approximate (LARS) vs exact (group Lasso) optimization, and weighted vs unweighted formulation.

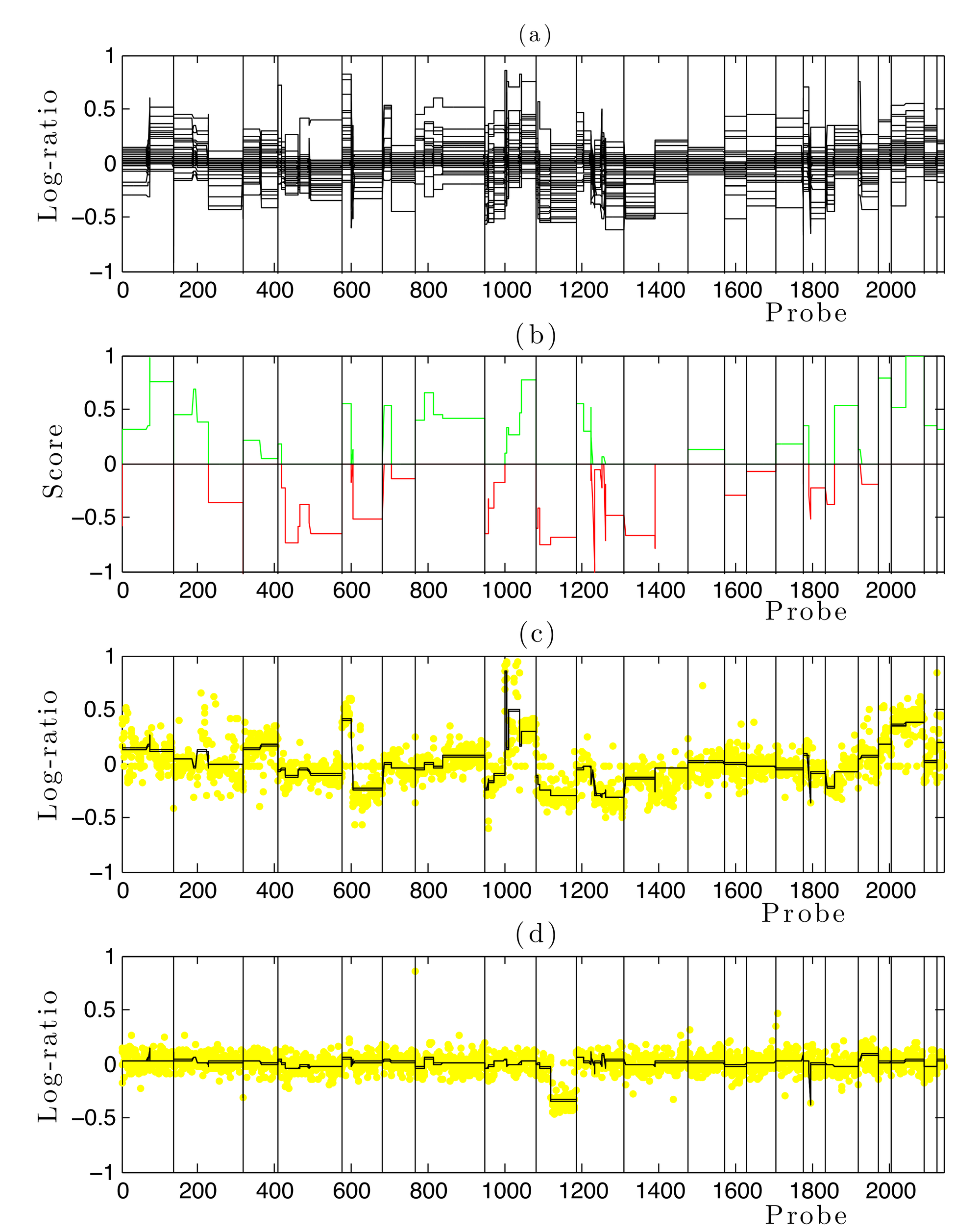

Figure 3: Joint segmentation of DNA copy number profiles (array CGH) of 57 bladder tumors, and detection of ammplified/deleted regions.