Supplementary materials: Fast detection of multiple change-points shared by many signals using group LARS

Jean-Philippe Vert and Kevin Bleakley
Mines ParisTech CBIO, Institut Curie, INSERM U900
{firstname.lastname}@mines-paristech.fr

1 Proof of Lemma 1

In order to show the complexity claimed in Lemma 1, let us rewrite it with an explicit description of the computation of $C = \bar{X}^\top R$:

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

1. *Compute the $n \times p$ matrix r of cumulative sums $r_{i,\bullet} = \sum_{j=1}^{i} R_{j\bullet}$ by the induction:*
   - $r_{1,\bullet} = R_{1,\bullet}$.
   - *For $i = 1, \ldots, n$, $r_{i,\bullet} = r_{i-1,\bullet} + R_{i,\bullet}$.*
2. *For $i = 1, \ldots, n-1$, compute $C_{i,\bullet} = i r_{n,\bullet}/n - r_{i,\bullet}$.*

Proof. $\bar{X}$ being by definition obtained by mean centering each column of X, its i-th column is given by

$$\forall i = 1, \ldots, n-1, \quad \bar{X}_{\bullet,i} = \Bigg(\underbrace{\frac{i}{n} - 1, \ldots, \frac{i}{n} - 1}_{i}, \underbrace{\frac{i}{n}, \ldots, \frac{i}{n}}_{n-i} \Bigg)^\top . \tag{1}$$

We therefore obtain the i-th row of $C = \bar{X}^\top R$, for $i = 1, \ldots, n-1$, with:

$$\begin{aligned} C_{i,\bullet} &= \bar{X}_{\bullet,i}^\top R \\ &= \left(\frac{i}{n} - 1\right)\left(\sum_{j=1}^{i} R_{j,\bullet}\right) + \frac{i}{n}\left(\sum_{j=i+1}^{n} R_{j,\bullet}\right) \\ &= \frac{i}{n} r_{n,\bullet} - r_{i,\bullet} . \end{aligned}$$

□

2 Proof of Lemma 2

In order to show the complexity claimed in Lemma 2, let us also rewrite it with an explicit algorithm:

Lemma 2. *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, the matrix*

$$C = \left(\bar{X}_{\bullet,A}^\top \bar{X}_{\bullet,A}\right)^{-1} R$$

can be computed in $O(|A|p)$ by

1. For $i = 1, \dots, |A|-1$, compute $\Delta_i = \frac{R_{i+1,\bullet} - R_{i,\bullet}}{a_{i+1} - a_i}$.

2. Compute the successive rows of C according to:

$$\begin{aligned} C_{1,\bullet} &= \frac{R_{1,\bullet}}{a_1} - \Delta_1 \,, \\ C_{i,\bullet} &= \Delta_{i-1} - \Delta_i \quad \textit{for } i = 2, \dots, |A|-1 \,, \\ C_{|A|,\bullet} &= \Delta_{|A|-1} + \frac{R_{|A|,\bullet}}{n - a_{|A|}} \,. \end{aligned} \tag{2}$$

Proof. Let us denote $V = \bar{X}_{\bullet,A}^\top \bar{X}_{\bullet,A}$. By (1) we have an explicit formula for $V_{i,j}$, namely, for $1 \le i \le j \le |A|$,

$$\begin{aligned} V_{i,j} &= \bar{X}_{\bullet,a_i}^\top \bar{X}_{\bullet,a_j} \\ &= a_i \left(\frac{a_i}{n} - 1\right)\left(\frac{a_j}{n} - 1\right) + (a_j - a_i)\frac{a_i}{n}\left(\frac{a_j}{n} - 1\right) + (n - a_j)\frac{a_i}{n}\frac{a_j}{n} \\ &= \frac{a_i(n - a_j)}{n} \,. \end{aligned}$$

From this explicit formula, one can easily check that V is invertible and admits as inverse a tri-diagonal matrix with the following entries:

$$\begin{aligned} V_{i,i}^{-1} &= \frac{1}{a_i - a_{i-1}} + \frac{1}{a_{i+1} - a_i} \quad \text{for } i = 1, \dots, |A|, \\ V_{i,i+1}^{-1} = V_{i+1,i}^{-1} &= -\frac{1}{a_{i+1} - a_i} \quad \text{for } i = 1, \dots, |A|-1, \end{aligned} \tag{3}$$

where by convention we define $a_0 = 0$ and $a_{|A|+1} = n$. This tri-diagonal structure allows successive rows of C to be expressed as a sum of just a few terms. More precisely, for $1 < i < |A|$, we obtain:

$$\begin{aligned} C_{i,\bullet} &= -\frac{R_{i-1,\bullet}}{a_i - a_{i-1}} + R_{i,\bullet}\left(\frac{1}{a_i - a_{i-1}} + \frac{1}{a_{i+1} - a_i}\right) - \frac{R_{i+1,\bullet}}{a_{i+1} - a_i} \\ &= \frac{R_{i,\bullet} - R_{i-1,\bullet}}{a_i - a_{i-1}} + \frac{R_{i,\bullet} - R_{i+1,\bullet}}{a_{i+1} - a_i} \\ &= \Delta_{i-1} - \Delta_i \,. \end{aligned}$$

Similarly, for $i = 1$ and $i = |A|$ we easily recover (2). □

3 Proof of Lemma 3

The first breakpoint is the row with the largest Euclidean norm of the matrix:

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

The entries of the matrix $\hat{c}$ are therefore jointly Gaussian. Since only the u-th row $\beta_{u,\bullet}$ of β is non-zero, we get

$$\bar{X}^\top \bar{X}\beta^* = \bar{X}^\top \bar{X}_{\bullet,u}\beta^*_{u,\bullet} \,.$$

A simple computation (see proof of Lemma 2) then shows that:

$$\left[\bar{X}^\top \bar{X}\beta^*\right]_{i,\bullet} = \begin{cases} \frac{i(n-u)}{n}\beta^*_{u,\bullet} & \text{for } 1 \le i \le u \,, \\ \frac{u(n-i)}{n}\beta^*_{u,\bullet} & \text{for } u \le i \le n-1 \,. \end{cases} \tag{4}$$

On the other hand, by the definition of $\bar{X}$ we have for any $i \in [1, n-1]$,

$$\left[\bar{X}^\top W\right]_{i,\bullet} = \sum_{j=1}^{i}\left(\frac{i}{n} - 1\right)W_{j,\bullet} + \sum_{j=i+1}^{n}\frac{i}{n}W_{j,\bullet} \,.$$

Since

$$E\left(W_{i,\bullet}^\top W_{j,\bullet}\right) = \delta_{i,j}\sigma^2 \mathbf{I}_p \,,$$

where $\delta_{i,j}$ is the Dirac function, we have for $1 \leq i \leq j \leq n-1$:

$$\begin{aligned} & E\left(\left[\bar{X}^\top W\right]_{i,\bullet}^\top \left[\bar{X}^\top W\right]_{j,\bullet}\right) \\ & = \left[i\left(\frac{i}{n}-1\right)\left(\frac{j}{n}-1\right) + (j-i)\frac{i}{n}\left(\frac{j}{n}-1\right) + (n-j)\frac{i}{n}\frac{j}{n}\right]\sigma^2\mathbf{I}_p \\ & = \frac{i\,(n-j)}{n}\sigma^2\mathbf{I}_p\,. \end{aligned} \tag{5}$$

In summary, we have shown that $\hat{c}$ is jointly Gaussian with $E\left(\hat{c}_{i,\bullet}\right)$ given by (9) and covariance between $\hat{c}_{i,\bullet}$ and $\hat{c}_{j,\bullet}$ given by (5).

In particular, if we denote $F_i = \parallel \hat{c}_{i,\bullet} \parallel^2$, then, for $i \leq u$, $F_i n / \left(i(n-i)\sigma^2\right)$ follows a non-central χ^2 distribution with p degrees of freedom and non-centrality parameter $p\bar{\beta}_p^2 i(n-u)^2 / \left[n(n-i)\sigma^2\right]$. In particular,

$$EF_i = \bar{\beta_p}^2 p \frac{i^2\,(n-u)^2}{n^2} + p\frac{i(n-i)}{n}\sigma^2\,,$$

and since $\lim_{p\rightarrow+\infty} \bar{\beta_p}^2 = \bar{\beta}^2$, we get that F_i/p converges in probability to

$$G_i = \frac{EF_i}{p} = \bar{\beta}^2\frac{i^2\,(n-u)^2}{n^2} + \frac{i(n-i)}{n}\sigma^2\,. \tag{6}$$

Similarly, for $i \geq u$ we get that F_i/p converges in probability to

$$G_i := \bar{\beta}^2\frac{u^2\,(n-i)^2}{n^2} + \frac{i(n-i)}{n}\sigma^2\,.$$

Let us assume without loss of generality that $u \geq n/2$. Then G_i is decreasing on $[u, n]$, as it is a sum of two decreasing functions of i on this interval, and therefore

$$\max_{i\in[1,n]} G_i = \max_{i\in[1,u]} G_i\,.$$

To conclude the proof it suffices to observe that, assuming the uniqueness of

$$\hat{u} = \operatorname*{argmax}_{i\in[1,n]} G_i = \operatorname*{argmax}_{i\in[1,u]} \bar{\beta}^2\frac{i^2\,(n-u)^2}{n^2} + \sigma^2\frac{i\,(n-i)}{n}\,, \tag{7}$$

the probability of the event $F_{\hat{u}} > F_i$ tends to 1 for any $i \neq \hat{u}$ as $p \rightarrow +\infty$ because $G_{\hat{u}} > G_i$, and by the union bound the probability of the event $F_{\hat{u}} = \max_{i\in[1,u]} F_i$ also converges to 1. This shows that the probability to select $\hat{u}$ converges to 1 as $p \rightarrow +\infty$. □

4 Proof of Theorem 4

By Lemma 3, if $u = \hat{u}$ then we will select the correct change-point with probability tending to 1 with p, otherwise we will choose $\hat{u}$ instead of u with probability also increasing to 1 with p. To prove Theorem 4 we therefore need to study whether or not $u = \hat{u}$, i.e., whether or not $G_u = \max_{i\in[1,u]} G_i$, where G_i is given by (6) (since we assume $u \geq n/2$).

G_i is a second-order polynomial of i, which is equal to 0 at $i = 0$ and strictly positive for $i = u$. Therefore $G_u = \max_{i\in[1,u]} G_i$ if and only if $G_u > G_{u-1}$. Let us therefore compute:

$$\begin{aligned} G_u - G_{u-1} & = \bar{\beta}^2\frac{(n-u)^2}{n^2}\left[u^2 - (u-1)^2\right] + \frac{\sigma^2}{n}\left[u(n-u) - (u-1)(n-u+1)\right] \\ & = \frac{\bar{\beta}^2(2u-1)(n-u)^2}{n^2} + \frac{\sigma^2(n-2u+1)}{n} \\ & = 2\left[\bar{\beta}^2 n\,(1-\alpha)^2\left(\alpha - \frac{1}{2n}\right) + \sigma^2\left(\frac{1}{2} - \alpha + \frac{1}{2n}\right)\right] \\ & = 2\left(\tilde{\sigma}^2 - \sigma^2\right)\left(\alpha - \frac{1}{2} - \frac{1}{2n}\right)\,, \end{aligned} \tag{8}$$

where $\alpha = u/n$ and

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

This shows that, when $\alpha > 1/2 + 1/(2n)$, $G_u > G_{u-1}$ if and only if $\sigma < \tilde{\sigma}$. On the other hand, when $\alpha = 1/2$ or $1/2 + 1/(2n)$, we have always that $G_u > G_{u-1}$. □

5 Proof of Theorem 5

Here we suppose that the i-th profile has a change-point at the random position U_i of random value β_i, assumed to be independent from both each other and between different profiles. Each U_i follows a distribution P_U on $[1, n-1]$ and β_i is distributed according to a distribution P_β on $\mathbb{R}$. We denote $\bar{\beta}^2 = E_{P_\beta}\beta^2$ and $p_i = P_U(U = i)$ for $i \in [1, n-1]$.

Following the proof of Lemma 3, let us estimate $F_i = \| \hat{c}_{i,\bullet} \|^2$ for $i \in [1, n-1]$. For any $j \in [1, p]$, we first observe that

$$\left[\bar{X}^\top \bar{X}\beta\right]_{i,j} = \begin{cases} \frac{i(n-U_j)}{n}\beta_j & \text{if } i \leq U_j\,, \\ \frac{U_j(n-i)}{n}\beta_j & \text{otherwise. } . \end{cases} \tag{9}$$

Therefore,

$$\sum_{j=1}^{p} \left[\bar{X}^\top \bar{X}\beta\right]_{i,j}^2 = \frac{i^2(n-i)^2}{n^2} \sum_{j=1}^{p} \beta_j^2 \left[\left(\frac{n-U_j}{n-i}\right)^2 \mathbf{1}(i \leq U_j) + \left(\frac{U_j}{i}\right)^2 \mathbf{1}(i > U_j) \right] , \tag{10}$$

and by independence of β_i and U_i:

$$\frac{1}{p} E \sum_{j=1}^{p} \left[\bar{X}^\top \bar{X}\beta\right]_{i,j}^2 = \bar{\beta}^2 \frac{i^2(n-i)^2}{n^2} \left[\sum_{u=1}^{i} p_u \left(\frac{u}{i}\right)^2 + \sum_{u=i+1}^{n-1} p_u \left(\frac{n-u}{n-i}\right)^2 \right] .$$

Since $(\beta_i, U_i)_{i=1,\dots,p}$ are independent of the noise, we obtain that F_i/p converges in probability to

$$G_i := \bar{\beta}^2 \frac{i^2(n-i)^2}{n^2} \left[\sum_{u=1}^{i} p_u \left(\frac{u}{i}\right)^2 + \sum_{u=i+1}^{n-1} p_u \left(\frac{n-u}{n-i}\right)^2 \right] + \frac{i(n-i)}{n}\sigma^2 . \tag{11}$$

As in Lemma 3 we can conclude that the method will select the position

$$\hat{u} = \underset{u \in [1, n-1]}{\operatorname{argmax}}\ G_i$$

with probability tending to 1 as p increases.

Let us now assume that the support of P_U is an interval $[a, b]$ (corresponding to a possible range of fluctuation of a change-point). Then, we observe that for $i \leq a$, the first term in (11) reduces to

$$\bar{\beta}^2 \frac{i^2(n-i)^2}{n^2} \left[0 + \sum_{u=a}^{b} p_u \left(\frac{n-u}{n-i}\right)^2 \right] = \bar{\beta}^2 \frac{i^2 E(n-U)^2}{n^2} ,$$

showing that it is strictly increasing on $[0, a]$. Similarly, we can show that it is strictly decreasing on $[b, n-1]$, from which we deduce that $\hat{u} \in [a, b]$ when the second term in (11) is negligible compared to the first one, e.g., when $\sigma^2 = 0$.

When $\sigma^2 > 0$, the second term in (11) moves the maximum of G_i towards $n/2$. When $n/2 \in [a, b]$, we deduce that for any $\sigma^2 > 0$, $\hat{u} \in [a, b]$. Otherwise, let us suppose without lack of generality that $n/2 < a \leq b$. Then, F_i being quadratic on $[0, a]$ and equal to 0 at 0, the maximum of G_i will not occur before a if and only if $G_{a-1} < G_a$. A computation similar to the one in the proof of Theorem 4 shows that this is the case if and only if

$$\sigma^2 < \tilde{\sigma}_{P_U} = n\bar{\beta}^2 \frac{E(1-\alpha)^2(\alpha_m - \frac{1}{2n})}{\alpha_m - \frac{1}{2} - \frac{1}{2n}} ,$$

where $\alpha = U/n$ and $\alpha_m = a$. To conclude the proof it suffices to observe that

$$E(1-\alpha)^2 = (1 - E\alpha)^2 + \text{var}(\alpha) .$$

□