Segmentation bidimensionnelle rapide pour l'étude des données Hi-C

Vincent Brault, Julien Chiquet et Céline Lévy-Leduc

Mardi 29 Septembre

AgroParisTech
INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT
PARIS INSTITUTE OF TECHNOLOGY FOR LIFE, FOOD AND ENVIRONMENTAL SCIENCES

INRA
SCIENCE & IMPACT

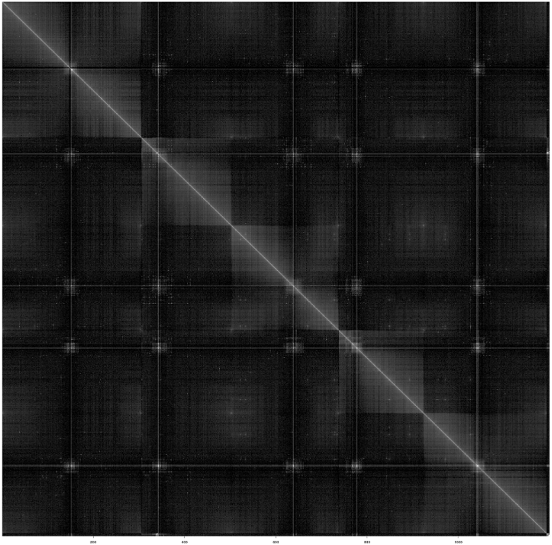
Hi-C data (High Chromosome Contact map)

- To better understand the organisation of a cell (Lieberman-Aiden et al. [2009]).
- To quantify the interaction between two positions of the genome (intra-chromosome and inter-chromosome).
- Each entry (i, j) : Number of interactions between the loci i and j [1].

1. A locus is a particular and invariable location on the chromosome.

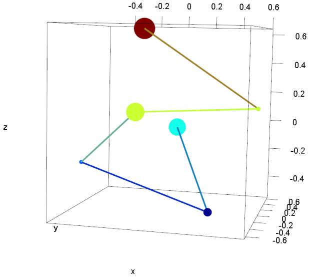
Motivations

Hi-C data of 5 chromosomes of the Arabidopsis Thaliana ; collaboration with M. Benhamed of the *institut de biologie des plantes (UMR 8618)*.



Goal

- To form group without permutations.
- To obtain a grid panel.
- To study matrices 10 000$\times$10 000.



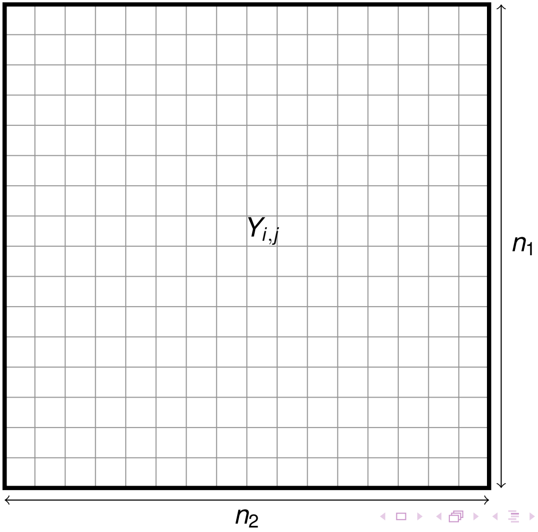
Plan

1. Model
2. Least Absolute Shrinkage eStimatOr
3. Theoretical results
4. Model selection
5. Numerical experiments

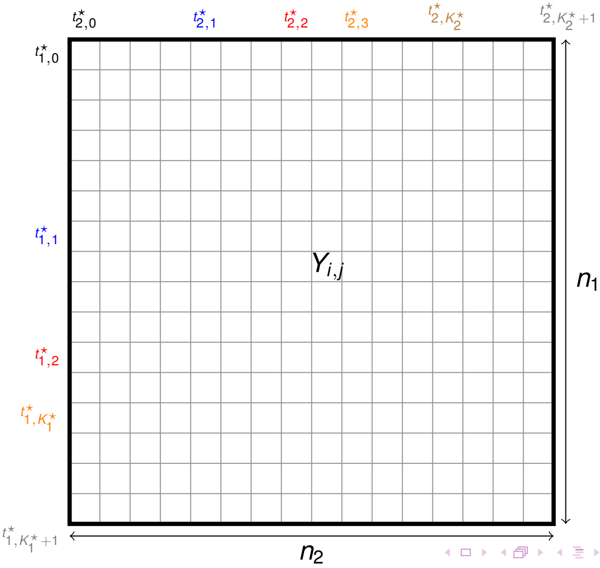
Plan

1. Model
2. Least Absolute Shrinkage eStimatOr
3. Theoretical results
4. Model selection
5. Numerical experiments

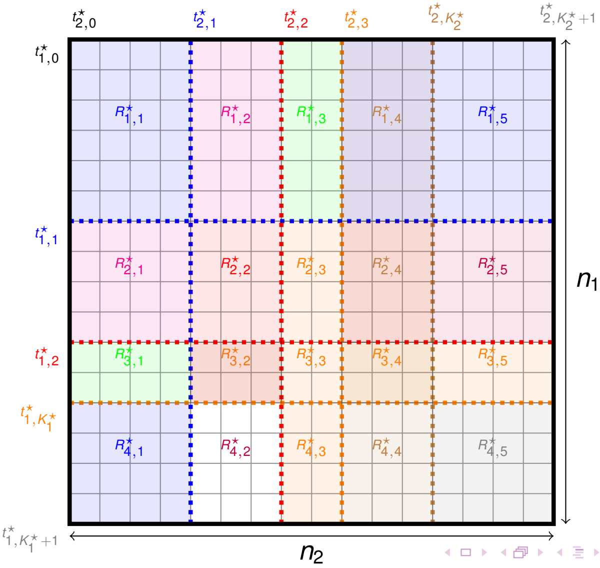
Notations



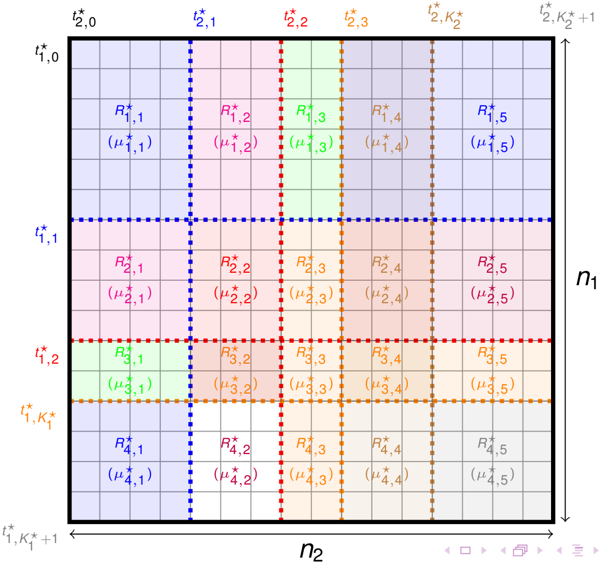
Notations


$t^\star_{2,0}$
$t^\star_{2,1}$
$t^\star_{2,2}$
$t^\star_{2,3}$
$t^\star_{2,K_2^\star}$
$t^\star_{2,K_2^\star+1}$
$t^\star_{1,0}$
$t^\star_{1,1}$
$t^\star_{1,2}$
$t^\star_{1,K_1^\star}$
$t^\star_{1,K_1^\star+1}$
$Y_{i,j}$
n_1
n_2

Notations


$t^\star_{2,0}$
$t^\star_{2,1}$
$t^\star_{2,2}$
$t^\star_{2,3}$
$t^\star_{2,K^\star_2}$
$t^\star_{2,K^\star_2+1}$
$t^\star_{1,0}$
$t^\star_{1,1}$
$t^\star_{1,2}$
$t^\star_{1,K^\star_1}$
$t^\star_{1,K^\star_1+1}$
$R^\star_{1,1}$ $R^\star_{1,2}$ $R^\star_{1,3}$ $R^\star_{1,4}$ $R^\star_{1,5}$
$R^\star_{2,1}$ $R^\star_{2,2}$ $R^\star_{2,3}$ $R^\star_{2,4}$ $R^\star_{2,5}$
$R^\star_{3,1}$ $R^\star_{3,2}$ $R^\star_{3,3}$ $R^\star_{3,4}$ $R^\star_{3,5}$
$R^\star_{4,1}$ $R^\star_{4,2}$ $R^\star_{4,3}$ $R^\star_{4,4}$ $R^\star_{4,5}$
n_1
n_2

Notations


$t^\star_{2,0}$
$t^\star_{2,1}$
$t^\star_{2,2}$
$t^\star_{2,3}$
$t^\star_{2,K^\star_2}$
$t^\star_{2,K^\star_2+1}$
$t^\star_{1,0}$
$t^\star_{1,1}$
$t^\star_{1,2}$
$t^\star_{1,K^\star_1}$
$t^\star_{1,K^\star_1+1}$
$R^\star_{1,1}$ $(\mu^\star_{1,1})$
$R^\star_{1,2}$ $(\mu^\star_{1,2})$
$R^\star_{1,3}$ $(\mu^\star_{1,3})$
$R^\star_{1,4}$ $(\mu^\star_{1,4})$
$R^\star_{1,5}$ $(\mu^\star_{1,5})$
$R^\star_{2,1}$ $(\mu^\star_{2,1})$
$R^\star_{2,2}$ $(\mu^\star_{2,2})$
$R^\star_{2,3}$ $(\mu^\star_{2,3})$
$R^\star_{2,4}$ $(\mu^\star_{2,4})$
$R^\star_{2,5}$ $(\mu^\star_{2,5})$
$R^\star_{3,1}$ $(\mu^\star_{3,1})$
$R^\star_{3,2}$ $(\mu^\star_{3,2})$
$R^\star_{3,3}$ $(\mu^\star_{3,3})$
$R^\star_{3,4}$ $(\mu^\star_{3,4})$
$R^\star_{3,5}$ $(\mu^\star_{3,5})$
$R^\star_{4,1}$ $(\mu^\star_{4,1})$
$R^\star_{4,2}$ $(\mu^\star_{4,2})$
$R^\star_{4,3}$ $(\mu^\star_{4,3})$
$R^\star_{4,4}$ $(\mu^\star_{4,4})$
$R^\star_{4,5}$ $(\mu^\star_{4,5})$
n_1
n_2

Model

Let $Y = (Y_{i,j})_{1 \leq i,j \leq n}$ be the random matrix defined by

$$\mathbf{Y} = \mathbf{U} + \mathbf{E},$$

where $\mathbf{U} = (U_{i,j})$ is a blockwise constant matrix such that

$$U_{i,j} = \mu^\star_{k,\ell} \quad \text{if } t^\star_{1,k-1} \leq i \leq t^\star_{1,k} - 1 \text{ and } t^\star_{2,\ell-1} \leq j \leq t^\star_{2,\ell} - 1,$$

with the convention $t^\star_{1,0} = t^\star_{2,0} = 1$ and $t^\star_{1,K^\star_1+1} = t^\star_{2,K^\star_2+1} = n + 1$. The entries $E_{i,j}$ of the matrix $E = (E_{i,j})_{1 \leq i,j \leq n}$ are iid zero-mean random variables.

Trick

$$\mathbf{U} = \mathbf{T}\mathbf{B}\mathbf{T}^\top$$

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{array}{c|cccccccccccccccc|c}
 & t^\star_{2,0} & & & & t^\star_{2,1} & & & t^\star_{2,2} & & t^\star_{2,3} & & & t^\star_{2,K_2^\star} & & & & t^\star_{2,K_2^\star+1} \\
\hline
t^\star_{1,0} & B_{1,1} & 0 & 0 & 0 & B_{1,5} & 0 & 0 & B_{1,8} & 0 & B_{1,10} & 0 & 0 & B_{1,13} & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
t^\star_{1,1} & B_{7,1} & 0 & 0 & 0 & B_{7,5} & 0 & 0 & B_{7,8} & 0 & B_{7,10} & 0 & 0 & B_{7,13} & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
t^\star_{1,2} & B_{11,1} & 0 & 0 & 0 & B_{11,5} & 0 & 0 & B_{11,8} & 0 & B_{11,10} & 0 & 0 & B_{11,13} & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
t^\star_{1,K_1^\star} & B_{13,1} & 0 & 0 & 0 & B_{13,5} & 0 & 0 & B_{13,8} & 0 & B_{13,10} & 0 & 0 & B_{13,13} & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
\hline
t^\star_{1,K_1^\star+1} & & & & & & & & & & & & & & & & &
\end{array}$$

$$\text{with } \mu^\star_{k,\ell} = \sum_{i=1}^{k} \sum_{j=1}^{\ell} B_{t^\star_{1,i},t^\star_{2,j}}.$$

Vectorisation

$$\mathbf{Y} = \mathbf{T}\mathbf{B}\mathbf{T}^\top + \mathbf{E}$$

is equivalent to

$$\mathrm{Vec}(\mathbf{Y}) = \mathrm{Vec}(\mathbf{T}\mathbf{B}\mathbf{T}^\top) + \mathrm{Vec}(\mathbf{E})$$

with

$$\mathrm{Vec}(\mathbf{T}\mathbf{B}\mathbf{T}^\top) = \left(\mathbf{T}^{\top^\top} \otimes \mathbf{T}\right) \mathrm{Vec}(\mathbf{B}) = (\mathbf{T} \otimes \mathbf{T}) \mathrm{Vec}(\mathbf{B})$$

and we obtain

$$\underbrace{\mathcal{Y}}_{n^2 \times 1} = \underbrace{\mathcal{X}}_{n^2 \times n^2} \underbrace{\mathcal{B}}_{n^2 \times 1} + \underbrace{\mathcal{E}}_{n^2 \times 1}.$$

Kronecker product

Vectorisation

$$\mathbf{Y} = \mathbf{T}\mathbf{B}\mathbf{T}^\top + \mathbf{E}$$

is equivalent to

$$\mathrm{Vec}(\mathbf{Y}) = \mathrm{Vec}(\mathbf{T}\mathbf{B}\mathbf{T}^\top) + \mathrm{Vec}(\mathbf{E})$$

with

$$\mathrm{Vec}(\mathbf{T}\mathbf{B}\mathbf{T}^\top) = \left(\mathbf{T}^{\top^\top} \otimes \mathbf{T}\right) \mathrm{Vec}(\mathbf{B}) = (\mathbf{T} \otimes \mathbf{T}) \mathrm{Vec}(\mathbf{B})$$

and we obtain

$$\underbrace{\mathcal{Y}}_{n^2 \times 1} = \underbrace{\mathcal{X}}_{n^2 \times n^2} \underbrace{\mathcal{B}}_{n^2 \times 1} + \underbrace{\mathcal{E}}_{n^2 \times 1}.$$

Kronecker product

Vectorisation

$$\mathbf{Y} = \mathbf{T}\mathbf{B}\mathbf{T}^\top + \mathbf{E}$$

is equivalent to

$$\mathrm{Vec}(\mathbf{Y}) = \mathrm{Vec}(\mathbf{T}\mathbf{B}\mathbf{T}^\top) + \mathrm{Vec}(\mathbf{E})$$

with

$$\mathrm{Vec}(\mathbf{T}\mathbf{B}\mathbf{T}^\top) = \left(\mathbf{T}^{\top^\top} \otimes \mathbf{T}\right) \mathrm{Vec}(\mathbf{B}) = (\mathbf{T} \otimes \mathbf{T}) \mathrm{Vec}(\mathbf{B})$$

and we obtain

$$\underbrace{\mathcal{Y}}_{n^2 \times 1} = \underbrace{\mathcal{X}}_{n^2 \times n^2} \underbrace{\mathcal{B}}_{n^2 \times 1} + \underbrace{\mathcal{E}}_{n^2 \times 1}.$$

Kronecker product

Plan

1 Model

2 Least Absolute Shrinkage eStimatOr

3 Theoretical results

4 Model selection

5 Numerical experiments

Least Absolute Shrinkage eStimatOr (LASSO)

For all $\lambda_n \geq 0$, we define

$$\widehat{\mathcal{B}}(\lambda_n) = \underset{\mathcal{B}\in\mathbb{R}^{n^2}}{\text{Argmin}} \left\{ \|\mathcal{Y} - \mathcal{X}\mathcal{B}\|_2^2 + \lambda_n \|\mathcal{B}\|_1 \right\}$$

and the active set

$$\widehat{\mathcal{A}}(\lambda_n) = \left\{ j \in \{1, \ldots, n^2\} : \widehat{\mathcal{B}}_j(\lambda_n) \neq 0 \right\}$$

A reminder on the norm

Least Absolute Shrinkage eStimatOr (LASSO)

For all $\lambda_n \geq 0$, we define

$$\widehat{\mathcal{B}}(0) = \underset{\mathcal{B} \in \mathbb{R}^{n^2}}{\operatorname{Argmin}} \left\{ \|\mathcal{Y} - \mathcal{X}\mathcal{B}\|_2^2 + \lambda_n \|\mathcal{B}\|_1 \right\}$$

and the active set

$$\widehat{\mathcal{A}}(0) = \left\{ j \in \{1, \ldots, n^2\} : \widehat{\mathcal{B}}_j(\lambda_n) \neq 0 \right\} \approx \{1, \ldots, n^2\}$$

A reminder on the norm

Least Absolute Shrinkage eStimatOr (LASSO)

For all $\lambda_n \geq 0$, we define

$$\widehat{\mathcal{B}}(+\infty) = \underset{\mathcal{B} \in \mathbb{R}^{n^2}}{\operatorname{Argmin}} \left\{ \|\mathcal{Y} - \mathcal{X}\mathcal{B}\|_2^2 + \lambda_n \|\mathcal{B}\|_1 \right\}$$

and the active set

$$\widehat{\mathcal{A}}(+\infty) = \left\{ j \in \{1, \ldots, n^2\} : \widehat{\mathcal{B}}_j(\lambda_n) \neq 0 \right\} = \emptyset$$

A reminder on the norm

Least Absolute Shrinkage eStimatOr (LASSO)

For all $\lambda_n \geq 0$, we define

$$\widehat{\mathcal{B}}(\lambda_n) = \underset{\mathcal{B} \in \mathbb{R}^{n^2}}{\text{Argmin}} \left\{ \|\mathcal{Y} - \mathcal{X}\mathcal{B}\|_2^2 + \lambda_n \|\mathcal{B}\|_1 \right\}$$

and the active set

$$\widehat{\mathcal{A}}(\lambda_n) = \left\{ j \in \{1, \ldots, n^2\} : \widehat{\mathcal{B}}_j(\lambda_n) \neq 0 \right\}$$

A reminder on the norm

Estimation of break change-point

$$\begin{pmatrix} 1\\2\\3\\4\\5\\6\\7\\8\\9\\10\\11\\12\\13\\14\\15\\16 \end{pmatrix} \Leftrightarrow \begin{array}{cc} & q_a+1 \\ \begin{pmatrix} 1 & 5 & 9 & 13 \\ 2 & 6 & 10 & 14 \\ 3 & 7 & 11 & 15 \\ 4 & 8 & 12 & 16 \end{pmatrix} & r_a+1 \end{array}$$

$\forall a \in \widehat{\mathcal{A}}(\lambda_n)$, we define (q_a, r_a) as the Euclidean division of $(a-1)$ by n, namely $(a-1) = nq_a + r_a$ then

$$\widehat{\mathbf{t}}_1 = (\widehat{t}_{1,k})_{1\leq k\leq |\widehat{\mathcal{A}}_1(\lambda_n)|} \in \widehat{\mathcal{A}}_1(\lambda_n) = \{r_a+1 : a \in \widehat{\mathcal{A}}(\lambda_n)\},$$

$$\widehat{\mathbf{t}}_2 = (\widehat{t}_{2,\ell})_{1\leq \ell\leq |\widehat{\mathcal{A}}_2(\lambda_n)|} \in \widehat{\mathcal{A}}_2(\lambda_n) = \{q_a+1 : a \in \widehat{\mathcal{A}}(\lambda_n)\}$$

where $\widehat{t}_{1,1} < \widehat{t}_{1,2} < \cdots < \widehat{t}_{1,|\widehat{\mathcal{A}}_1(\lambda_n)|}$,

and $\widehat{t}_{2,1} < \widehat{t}_{2,2} < \cdots < \widehat{t}_{2,|\widehat{\mathcal{A}}_2(\lambda_n)|}$.

Estimation of break change-point

$$\begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \Leftrightarrow \begin{array}{cc} \begin{array}{cccc} & & q_a+1 & \end{array} & \\ \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & a & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} & \begin{array}{c} \\ r_a+1 \\ \\ \\ \end{array} \end{array}$$

For example, a=10

$10 - 1 = 9 = 4 \times 2 + 1$

2+1=3 and 1+1=2

$\forall a \in \widehat{\mathcal{A}}(\lambda_n)$, we define (q_a, r_a) as the Euclidean division of $(a-1)$ by n, namely $(a-1) = nq_a + r_a$ then

$$\widehat{\mathbf{t}}_1 = (\widehat{t}_{1,k})_{1 \leq k \leq |\widehat{\mathcal{A}}_1(\lambda_n)|} \in \widehat{\mathcal{A}}_1(\lambda_n) = \{r_a + 1 : a \in \widehat{\mathcal{A}}(\lambda_n)\},$$

$$\widehat{\mathbf{t}}_2 = (\widehat{t}_{2,\ell})_{1 \leq \ell \leq |\widehat{\mathcal{A}}_2(\lambda_n)|} \in \widehat{\mathcal{A}}_2(\lambda_n) = \{q_a + 1 : a \in \widehat{\mathcal{A}}(\lambda_n)\}$$

where $\widehat{t}_{1,1} < \widehat{t}_{1,2} < \cdots < \widehat{t}_{1,|\widehat{\mathcal{A}}_1(\lambda_n)|}$,

and $\widehat{t}_{2,1} < \widehat{t}_{2,2} < \cdots < \widehat{t}_{2,|\widehat{\mathcal{A}}_2(\lambda_n)|}$.

Estimation of break change-point

$$\begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \Leftrightarrow \begin{array}{cc} \begin{array}{cccc} & & q_a+1 & \end{array} & \\ \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & a & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} & \begin{array}{c} \\ r_a+1 \\ \\ \\ \end{array} \end{array}$$

For example, a=10
$10 - 1 = 9 = 4 \times 2 + 1$
2+1=3 and 1+1=2

$\forall a \in \widehat{\mathcal{A}}(\lambda_n)$, we define (q_a, r_a) as the Euclidean division of $(a-1)$ by n, namely $(a-1) = nq_a + r_a$ then

$$\widehat{\mathbf{t}}_1 = (\widehat{t}_{1,k})_{1 \le k \le |\widehat{\mathcal{A}}_1(\lambda_n)|} \in \widehat{\mathcal{A}}_1(\lambda_n) = \{r_a + 1 : a \in \widehat{\mathcal{A}}(\lambda_n)\},$$

$$\widehat{\mathbf{t}}_2 = (\widehat{t}_{2,\ell})_{1 \le \ell \le |\widehat{\mathcal{A}}_2(\lambda_n)|} \in \widehat{\mathcal{A}}_2(\lambda_n) = \{q_a + 1 : a \in \widehat{\mathcal{A}}(\lambda_n)\}$$

$$\text{where} \quad \widehat{t}_{1,1} < \widehat{t}_{1,2} < \cdots < \widehat{t}_{1,|\widehat{\mathcal{A}}_1(\lambda_n)|},$$

$$\text{and} \quad \widehat{t}_{2,1} < \widehat{t}_{2,2} < \cdots < \widehat{t}_{2,|\widehat{\mathcal{A}}_2(\lambda_n)|}.$$

Estimation of break change-point

$$\begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \Leftrightarrow \begin{array}{cc} \begin{array}{cccc} & & q_a+1 & \end{array} & \\ \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & a & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} & r_a+1 \end{array}$$

For example, a=10
$10-1=9=4\times 2+1$
2+1=3 and 1+1=2

$\forall a \in \widehat{\mathcal{A}}(\lambda_n)$, we define (q_a, r_a) as the Euclidean division of $(a-1)$ by n, namely $(a-1)=nq_a+r_a$ then

$$\widehat{\mathbf{t}}_1 = (\widehat{t}_{1,k})_{1\le k\le |\widehat{\mathcal{A}}_1(\lambda_n)|} \in \widehat{\mathcal{A}}_1(\lambda_n) = \{r_a+1 : a\in\widehat{\mathcal{A}}(\lambda_n)\},$$

$$\widehat{\mathbf{t}}_2 = (\widehat{t}_{2,\ell})_{1\le \ell\le |\widehat{\mathcal{A}}_2(\lambda_n)|} \in \widehat{\mathcal{A}}_2(\lambda_n) = \{q_a+1 : a\in\widehat{\mathcal{A}}(\lambda_n)\}$$

where $\widehat{t}_{1,1} < \widehat{t}_{1,2} < \cdots < \widehat{t}_{1,|\widehat{\mathcal{A}}_1(\lambda_n)|}$,

and $\widehat{t}_{2,1} < \widehat{t}_{2,2} < \cdots < \widehat{t}_{2,|\widehat{\mathcal{A}}_2(\lambda_n)|}$.

Estimation of break change-point

$$\begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \Leftrightarrow \begin{array}{c} \begin{array}{cccc} & & q_a+1 & \end{array} \\ \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & a & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} \end{array} \begin{array}{l} \\ r_a+1 \\ \\ \\ \end{array}$$

For example, a=10
$10-1=9=4\times 2+1$
2+1=3 and 1+1=2

$\forall a \in \widehat{\mathcal{A}}(\lambda_n)$, we define (q_a, r_a) as the Euclidean division of $(a-1)$ by n, namely $(a-1)=nq_a+r_a$ then

$$\widehat{\mathbf{t}}_1 = (\widehat{t}_{1,k})_{1\leq k\leq |\widehat{\mathcal{A}}_1(\lambda_n)|} \in \widehat{\mathcal{A}}_1(\lambda_n) = \{r_a+1 : a \in \widehat{\mathcal{A}}(\lambda_n)\},$$

$$\widehat{\mathbf{t}}_2 = (\widehat{t}_{2,\ell})_{1\leq \ell\leq |\widehat{\mathcal{A}}_2(\lambda_n)|} \in \widehat{\mathcal{A}}_2(\lambda_n) = \{q_a+1 : a \in \widehat{\mathcal{A}}(\lambda_n)\}$$

$$\text{where} \quad \widehat{t}_{1,1} < \widehat{t}_{1,2} < \cdots < \widehat{t}_{1,|\widehat{\mathcal{A}}_1(\lambda_n)|},$$

$$\text{and} \quad \widehat{t}_{2,1} < \widehat{t}_{2,2} < \cdots < \widehat{t}_{2,|\widehat{\mathcal{A}}_2(\lambda_n)|}.$$

Standard complexity : [2]

$$\mathcal{O}\left(|\mathcal{A}|\, mp + p\,|\mathcal{A}|^2 + |\mathcal{A}|^3\right).$$

In our case, we have :

$$\mathcal{O}\left(|\mathcal{A}|\, n^4\right).$$

$$\mathcal{X} = \mathbf{T} \otimes \mathbf{T}$$

2. see for example Bach et al. [2011].

Standard complexity :[2]

$$\mathcal{O}\left(|\mathcal{A}|\, mp + p\,|\mathcal{A}|^2 + |\mathcal{A}|^3\right).$$

In our case, we have :

$$\mathcal{O}\left(|\mathcal{A}|\, n^4\right).$$

$$\mathcal{X} = \mathbf{T} \otimes \mathbf{T}$$

2. see for example Bach et al. [2011].

Standard complexity :[2]

$$\mathcal{O}\left(|\mathcal{A}|\, mp + p\,|\mathcal{A}|^2 + |\mathcal{A}|^3\right).$$

In our case, we have :

$$\mathcal{O}\left(|\mathcal{A}|\, n^4\right).$$

$$\mathcal{X} = \mathbf{T} \otimes \mathbf{T}$$

2. see for example Bach et al. [2011].

Fast LARS for two-dimensional change-point detection :

Input : data matrix Y, maximal number of active variables s.

// Initialisation

Start with no change-point $\mathcal{A} \leftarrow \emptyset$, $\hat{\mathcal{B}} = \mathbf{0}$

Compute current correlations $\hat{\mathbf{c}} = \mathcal{X}^\top \mathcal{Y}$ $\mathcal{O}(n^2)$

While $\lambda > 0$ or $|\mathcal{A}| < s$

// Update the set of active variables

Determine next change-point(s) by setting $\lambda \leftarrow \|\hat{\mathbf{c}}\|_\infty$ and $\mathcal{A} \leftarrow \left\{j : \hat{\mathbf{c}}_j = \lambda\right\}$

Update the Cholesky factorization of $\mathcal{X}_\mathcal{A}^\top \mathcal{X}_\mathcal{A}$ $\mathcal{O}(|\mathcal{A}|^2)$

// Compute the direction of descent

Get the unormalized direction $\tilde{w}_\mathcal{A} \leftarrow \left(\mathcal{X}_{\cdot\mathcal{A}}^\top \mathcal{X}_{\cdot\mathcal{A}}\right)^{-1} \mathrm{sign}(\hat{c}_\mathcal{A})$ $\mathcal{O}(|\mathcal{A}|^2)$

Normalize $w_\mathcal{A} \leftarrow \alpha \tilde{w}_\mathcal{A}$ with $\alpha \leftarrow 1/\sqrt{\tilde{w}_\mathcal{A}^\top \mathrm{sign}(\hat{c}_\mathcal{A})}$

Compute the equiangular vector $u_\mathcal{A} = \mathcal{X}_\mathcal{A} w_\mathcal{A}$ and $\mathbf{a} = \mathcal{X}^\top u_\mathcal{A}$ $\mathcal{O}(n^2)$

// Compute the direction step

Find the maximal step preserving equicorrelation $\gamma_{\mathrm{in}} \leftarrow \min^+_{j\in\mathcal{A}^c} \left\{ \frac{\lambda - \mathbf{c}_j}{\alpha - a_j}, \frac{\lambda + \mathbf{c}_j}{\alpha + a_j} \right\}$

Find the maximal step preserving the signs $\gamma_{\mathrm{out}} \leftarrow \min^+_{j\in\mathcal{A}} \left\{-\hat{\mathcal{B}}_\mathcal{A}/w_\mathcal{A}\right\}$

The direction step that preserves both is $\hat{\gamma} \leftarrow \min(\gamma_{\mathrm{in}}, \gamma_{\mathrm{out}})$

Update the correlations $\hat{\mathbf{c}} \leftarrow \hat{\mathbf{c}} - \hat{\gamma}\mathbf{a}$ and $\hat{\mathcal{B}}_\mathcal{A} \leftarrow \hat{\mathcal{B}}_\mathcal{A} + \hat{\gamma} w_\mathcal{A}$ accordingly $\mathcal{O}(n)$

// Drop variable crossing the zero line

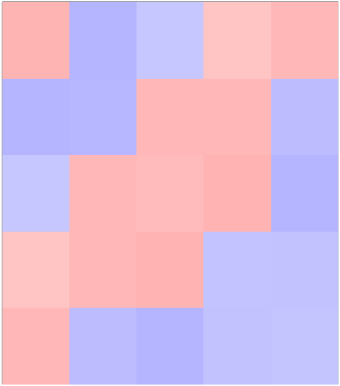
If $\gamma_{\mathrm{out}} < \gamma_{\mathrm{in}}$

Remove existing change-point(s) $\mathcal{A} \leftarrow \mathcal{A} \setminus \left\{j \in \mathcal{A} : \hat{\mathcal{B}}_j = 0\right\}$

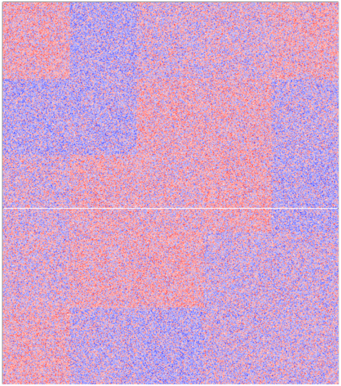
Downdate the Cholesky factorization of $\mathcal{X}_\mathcal{A}^\top \mathcal{X}_\mathcal{A}$ $\mathcal{O}(|\mathcal{A}|)$

Output : Sequence of triplet $(\mathcal{A}, \lambda, \hat{\mathcal{B}})$ recorded at each iteration.

Mu matrix



Orignal matrix



Gray version

Plan

1. Model
2. Least Absolute Shrinkage eStimatOr
3. Theoretical results
4. Model selection
5. Numerical experiments

Proposition

Let $(Y_{i,j})_{\substack{1\leq i\leq n_1\\ 1\leq j\leq n_2}}$ a data matrix and $\widehat{t}_{1,k}$, $\widehat{t}_{2,k}$ the estimators obtained by the LASSO.
Under some assumptions and assume that $|\widehat{\mathcal{A}}_1(\lambda_n)| = K_1^\star$ and that $|\widehat{\mathcal{A}}_2(\lambda_n)| = K_2^\star$ then for all $a \in \{1, \ldots, n^2\}$,

$$\mathbb{P}\left(\left\{\max_{1\leq k\leq K_1^\star}\left|\widehat{t}_{1,k} - t_{1,k}^\star\right| \leq n_1\delta_{n_1,n_2}\right\} \cap \left\{\max_{1\leq k\leq K_2^\star}\left|\widehat{t}_{2,k} - t_{2,k}^\star\right| \leq n_2\delta_{n_1,n_2}\right\}\right) \underset{n\to\infty}{\longrightarrow} 1.$$

$$\frac{n_1}{\log n_2} \underset{n_1,n_2\to+\infty}{\longrightarrow} +\infty$$

Key of proof

Proposition

Let $(Y_{i,j})_{\substack{1 \le i \le n_1 \\ 1 \le j \le n_2}}$ a data matrix and $\widehat{t}_{1,k}$, $\widehat{t}_{2,k}$ the estimators obtained by the LASSO.
Under some assumptions and assume that $|\widehat{\mathcal{A}}_1(\lambda_n)| = K_1^\star$ and that $|\widehat{\mathcal{A}}_2(\lambda_n)| = K_2^\star$ then for all $a \in \{1, \dots, n^2\}$,

$$\mathbb{P}\left(\left\{\max_{1 \le k \le K_1^\star} \left|\widehat{t}_{1,k} - t_{1,k}^\star\right| \le n_1 \delta_{n_1,n_2}\right\} \cap \left\{\max_{1 \le k \le K_2^\star} \left|\widehat{t}_{2,k} - t_{2,k}^\star\right| \le n_2 \delta_{n_1,n_2}\right\}\right) \underset{n \to \infty}{\longrightarrow} 1.$$

$$\frac{n_1}{\log n_2} \underset{n_1, n_2 \to +\infty}{\longrightarrow} +\infty$$

Key of proof

Plan

1. Model
2. Least Absolute Shrinkage eStimatOr
3. Theoretical results
4. Model selection
5. Numerical experiments

Stability selection

Stability selection :

Input : data vector $\mathcal{Y} \in \mathcal{M}_{n^2 \times 1}$, an integer $M \in \mathbb{N}^\star$, a pair of numbers $(K_1^\star, K_2^\star) \in \{1, \ldots, n\}^2$.

For $\underline{iter \in \{1, \ldots, M\}}$

Chose randomly $ind^{(iter)} = \{i_1, \ldots, i_{n^2/2}\} \subset \{1, \ldots, n^2\}$.

Use the procedure with $(K_1^\star, K_2^\star)$ change-points on the data $\mathcal{Y}_{ind^{(iter)}}$ to obtain $\left(\widehat{\mathbf{t}}_1^{(iter)}, \widehat{\mathbf{t}}_2^{(iter)}\right)$.

Output : Sequence of couples $\left(\widehat{\mathbf{t}}_1^{(iter)}, \widehat{\mathbf{t}}_2^{(iter)}\right)$ recorded at each iteration or only the couple of change-points appearing a number of times larger than a given threshold.

Adaptation

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 5 & 9 & 13 \\ 2 & 6 & 10 & 14 \\ 3 & 7 & 11 & 15 \\ 4 & 8 & 12 & 16 \end{pmatrix}$$

Adaptation

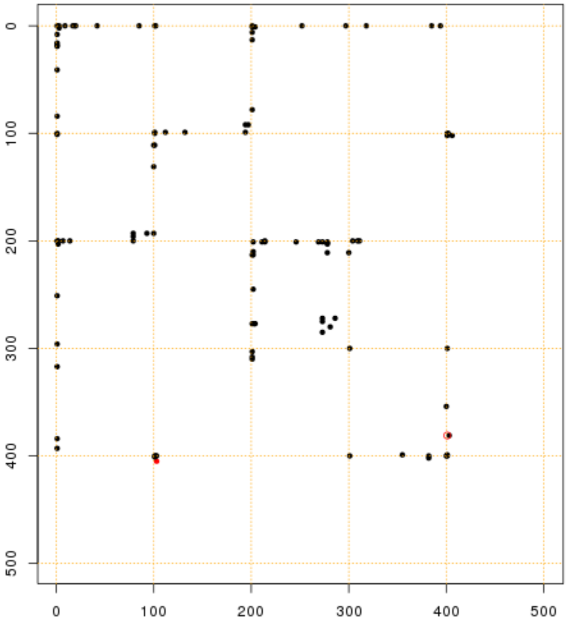
$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ . \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 5 & 9 & 13 \\ 2 & 6 & . & 14 \\ 3 & 7 & 11 & 15 \\ 4 & 8 & 12 & 16 \end{pmatrix}$$

Adaptation

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 13 \\ 14 \\ 15 \\ 16 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 5 & \cdot & 13 \\ 2 & 6 & \cdot & 14 \\ 3 & 7 & \cdot & 15 \\ 4 & 8 & \cdot & 16 \end{pmatrix}$$

Adaptation

$$\begin{pmatrix} 1 \\ 2 \\ . \\ 4 \\ 5 \\ 6 \\ . \\ 8 \\ . \\ . \\ . \\ . \\ 13 \\ 14 \\ . \\ 16 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 5 & . & 13 \\ 2 & 6 & . & 14 \\ . & . & . & . \\ 4 & 8 & . & 16 \end{pmatrix}$$


0
100
200
300
400
500
0
100
200
300
400
500

Adaptation stability selection

Stability selection :

Input : data matrix $Y \in \mathcal{M}_{n\times n}$, an integer $M \in \mathbb{N}^\star$, a pair of numbers $(K_1^\star, K_2^\star) \in \{1, \dots, n\}^2$.

For $\underline{iter \in \{1, \dots, M\}}$

Choose randomly $ind_1^{(iter)} = \left\{i_1^{(1)}, \dots, i_{n/2}^{(1)}\right\} \subset \{1, \dots, n\}$ and $ind_2^{(iter)} = \left\{i_1^{(2)}, \dots, i_{n/2}^{(2)}\right\} \subset \{1, \dots, n\}$.

Use the procedure with $(K_1^\star, K_2^\star)$ change-points on the data $Y_{ind_1^{(iter)}, ind_2^{(iter)}}$ to obtain $\left(N_1^{(iter)}, N_2^{(iter)}\right)$ the number of times that each change-point of $\{1, \dots, n\}^2$ was selected.

Output : Sequence of couple of numbers $\left(N_1^{(iter)}, N_2^{(iter)}\right)$ recorded at each iteration.

Plan

1 Model

2 Least Absolute Shrinkage eStimatOr

3 Theoretical results

4 Model selection

5 Numerical experiments

Experimental design

- $K_1^\star = K_2^\star = 4$.
- $\left(\mu_{k,\ell}^\star\right)_{k\in\{1,\ldots K_1^\star+1\},\ell\in\{1,\ldots K_2^\star+1\}} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}$.
- $(t_{1,k}^\star)_{1\leq k\leq K_1^\star} = ([nk/(K_1^\star + 1)] + 1)_{1\leq k\leq K_1^\star}$ and $(t_{2,k}^\star)_{1\leq k\leq K_2^\star} = ([nk/(K_2^\star + 1)] + 1)_{1\leq k\leq K_2^\star}$.
- $E_{i,j} \sim \mathcal{N}\left(0, \sigma^2\right)$.
- 1000 matrices simulated for each case.

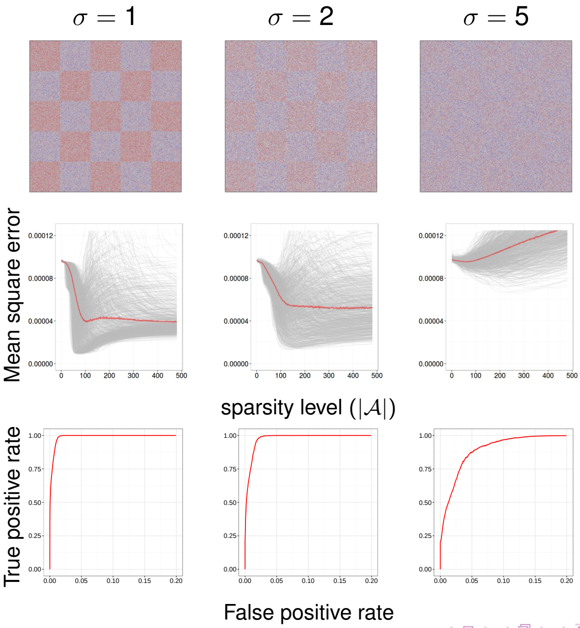
Statistical performances

Parameters :

- $n = 500$.
- $\sigma \in \{1, 2, 5\}$.

Evaluation :

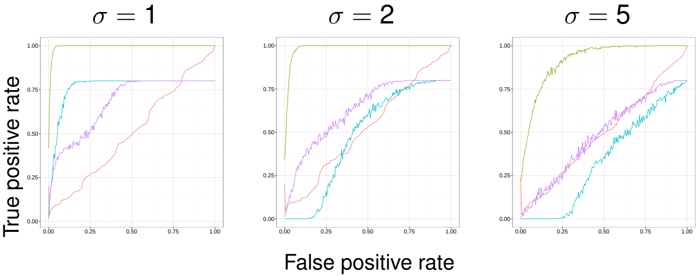
- Mean square error $n^{-2}\|\mathcal{B} - \widehat{\mathcal{B}}\|_2^2$ as a function of the number of nonzero elements in $\widehat{\mathcal{B}}$ for each scenario.
- ROC curves for the estimated change-points in rows.



$\sigma = 1$
$\sigma = 2$
$\sigma = 5$
Mean square error
0.00012
0.00008
0.00004
0.00000
0 100 200 300 400 500
sparsity level ($|\mathcal{A}|$)
True positive rate
1.00
0.75
0.50
0.25
0.00
0.00 0.05 0.10 0.15 0.20
False positive rate

- Green : Our method.
- Blue : One-dimensional LASSO[3] with at least in one row.
- Purple : One-dimensional LASSO[3] with at least in $([n/2]+1)$ rows.
- Red : Extension of CART.

$n = 250$ with 100 matricies



3. Using the procedure of Harchaoui and Lévy-Leduc [2010]

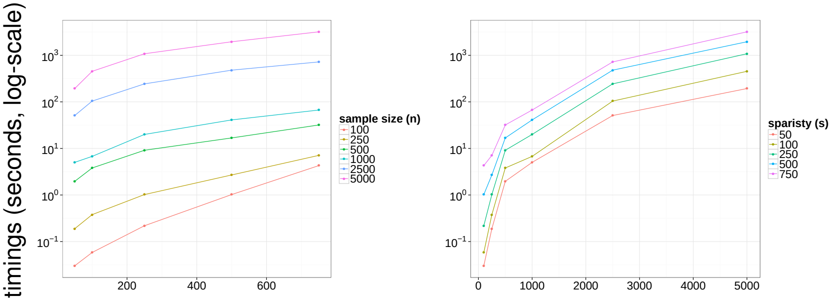
Statistical performances

Parameters :

- $n \in \{100, 250, 500, 1000, 2500, 5000\}$.
- $|\mathcal{A}| \in \{50, 100, 250, 500, 750\}$.
- $\sigma = 10$.
- Linux workstation with Intel Xeon 2.4 GHz processor and 8 GB of memory

Evaluation :

- The median runtimes.



sparsity level ($|\mathcal{A}|$) sample size (n)

1 000 $\times$ 1 000 $=$ 1 000 000 less than 2 minutes.

Model selection

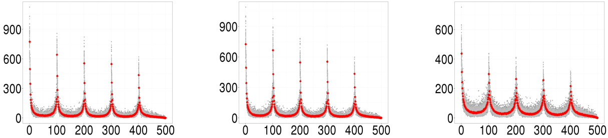
Parameters :

- $n = 500$.
- $\sigma \in \{1, 2, 5\}$.
- $M = 100$.

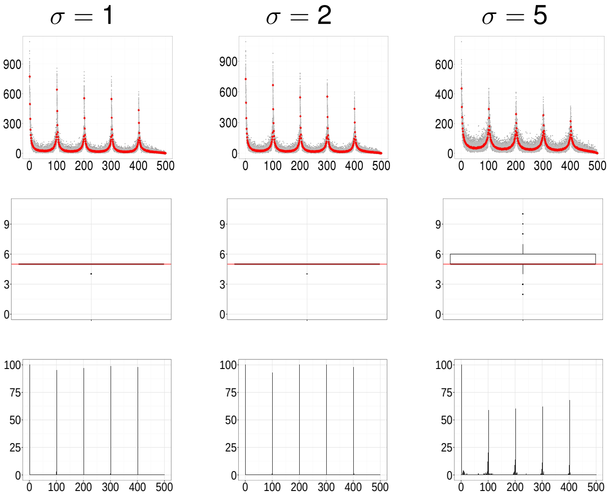
Evaluation :

- Boxplots of the estimation of $K_1^\star$.
- Histograms of the estimated change-points in rows.

$\sigma = 1$ $\sigma = 2$ $\sigma = 5$



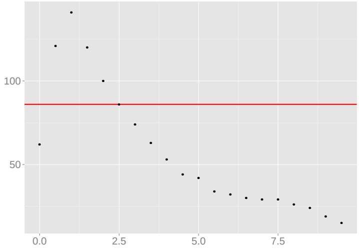
Colour Film Gray Film

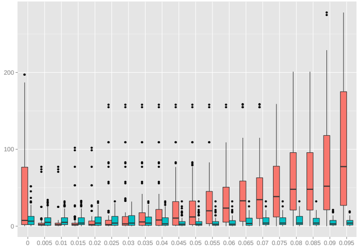


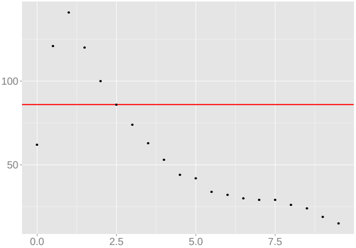
σ = 1
σ = 2
σ = 5
900
600
300
0
0
100
200
300
400
500
400
200
9
6
3
100
75
50
25

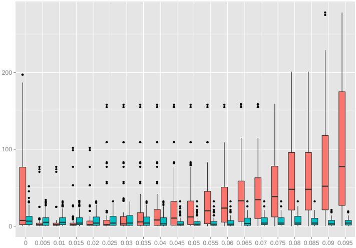
Real data

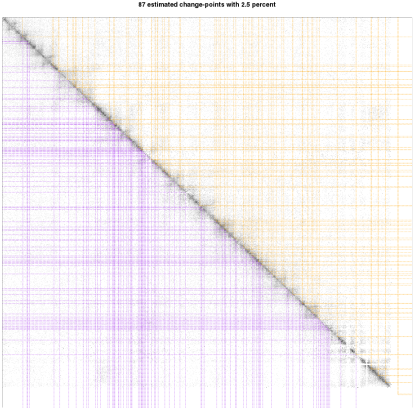
- Chromosome 19 of the mouse cortex at a resolution 40 kb.
- Comparison with Dixon et al. [2012].

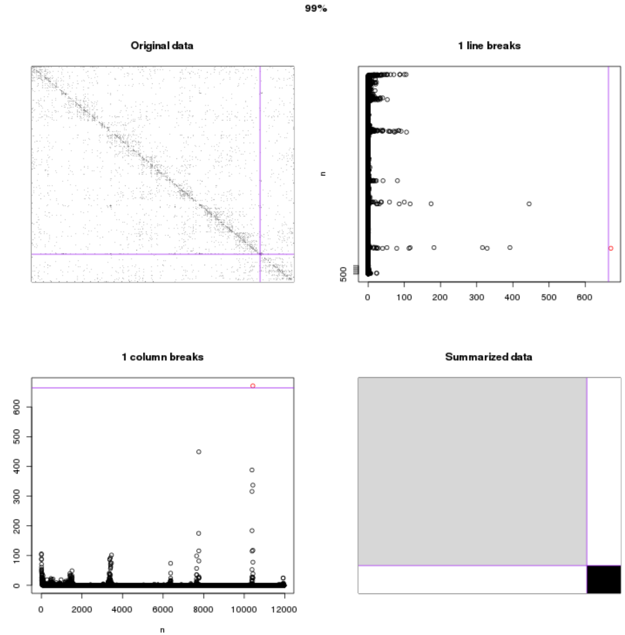

100
50
0.0
2.5
5.0
7.5


200
100
0
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035
0.04
0.045
0.05
0.055
0.06
0.065
0.07
0.075
0.08
0.085
0.09
0.095

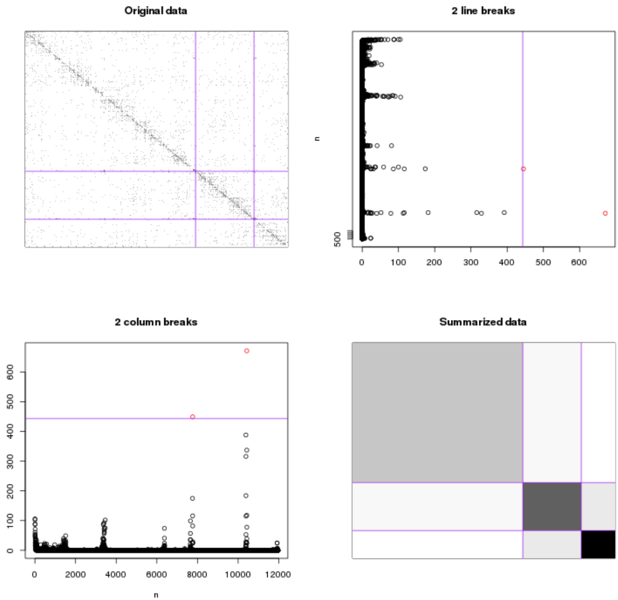

100
50
0.0
2.5
5.0
7.5


200
100
0
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035
0.04
0.045
0.05
0.055
0.06
0.065
0.07
0.075
0.08
0.085
0.09
0.095

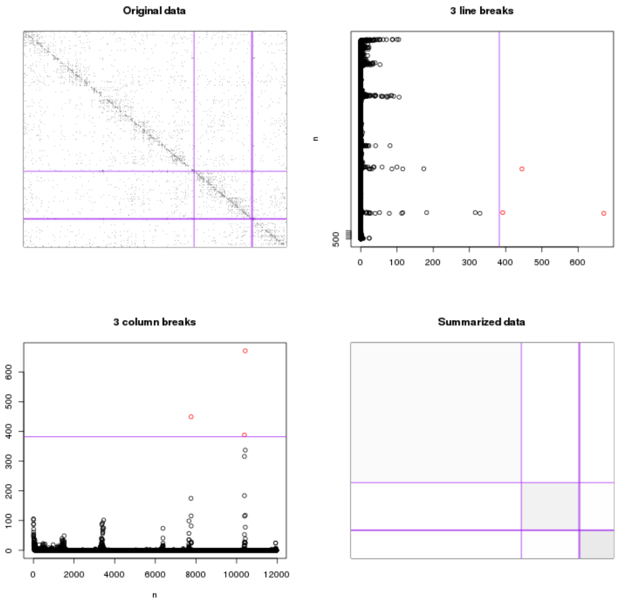

87 estimated change-points with 2.5 percent

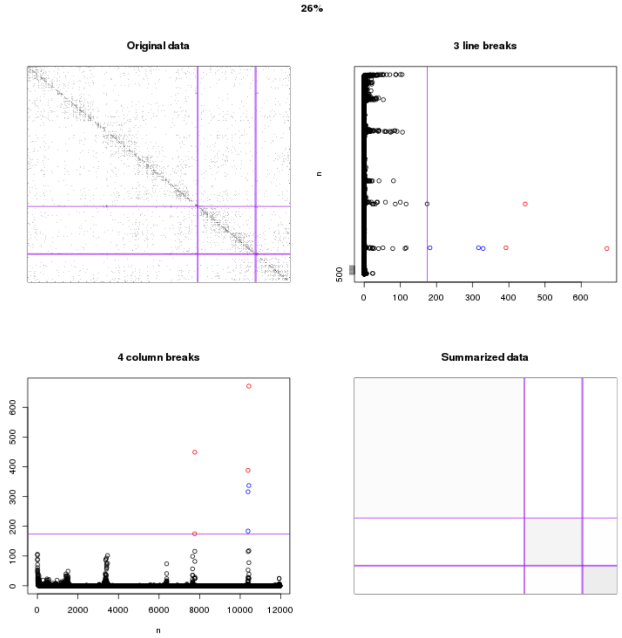

99%
Original data
1 line breaks
n
500
0
100
200
300
400
500
600
1 column breaks
600
500
400
300
200
100
0
2000
4000
6000
8000
10000
12000
n
Summarized data

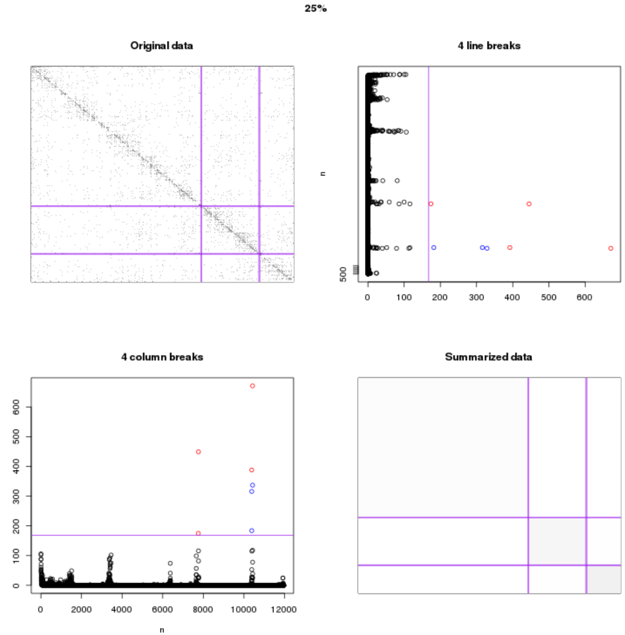
66%


Original data
2 line breaks
n
500
0 100 200 300 400 500 600
2 column breaks
0 100 200 300 400 500 600
0 2000 4000 6000 8000 10000 12000
n
Summarized data

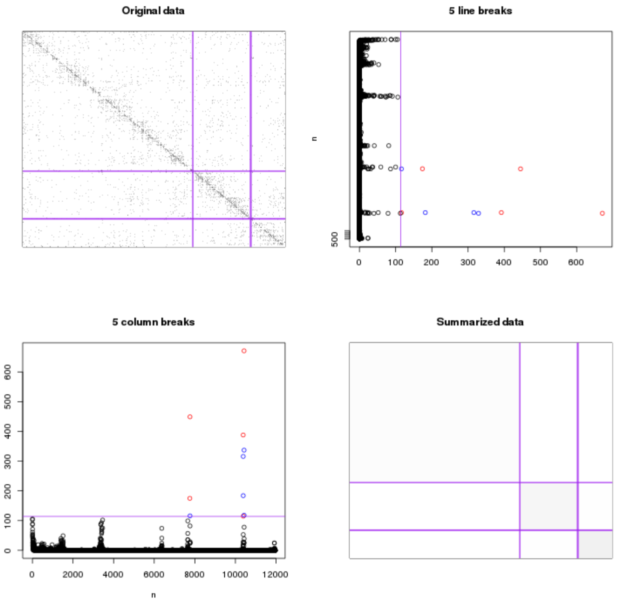
57%

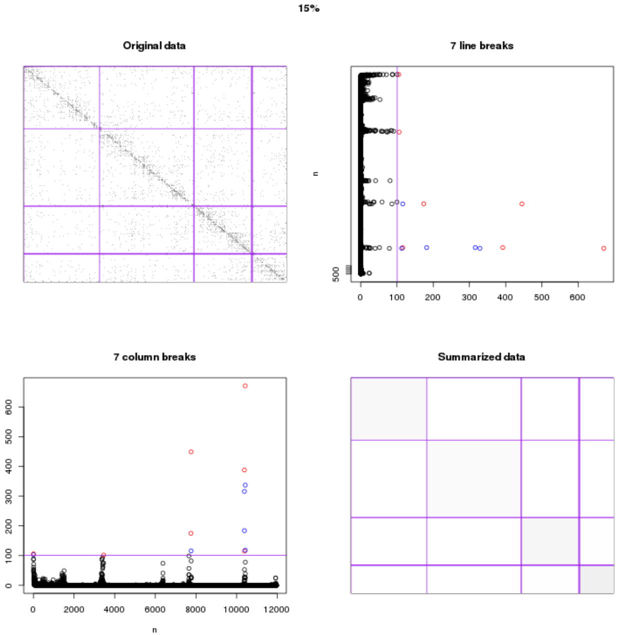

Original data
3 line breaks
n
500
0
100
200
300
400
500
600
3 column breaks
0
100
200
300
400
500
600
0
2000
4000
6000
8000
10000
12000
n
Summarized data


26%
Original data
3 line breaks
n
500
0 100 200 300 400 500 600
4 column breaks
0 100 200 300 400 500 600
0 2000 4000 6000 8000 10000 12000
n
Summarized data

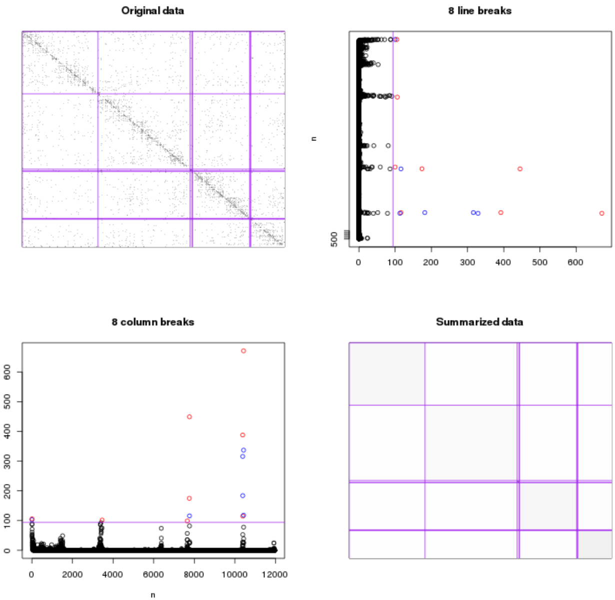

25%
Original data
4 line breaks
n
500
0
100
200
300
400
500
600
4 column breaks
0
100
200
300
400
500
600
0
2000
4000
6000
8000
10000
12000
n
Summarized data

17%


Original data
5 line breaks
n
500
0
100
200
300
400
500
600
5 column breaks
0
100
200
300
400
500
600
0
2000
4000
6000
8000
10000
12000
n
Summarized data

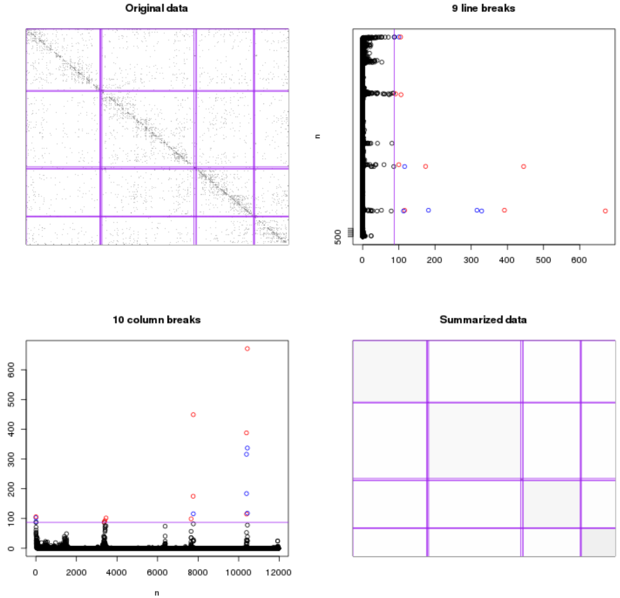

15%
Original data
7 line breaks
n
500
0 100 200 300 400 500 600
7 column breaks
600 500 400 300 200 100 0
0 2000 4000 6000 8000 10000 12000
n
Summarized data

14%

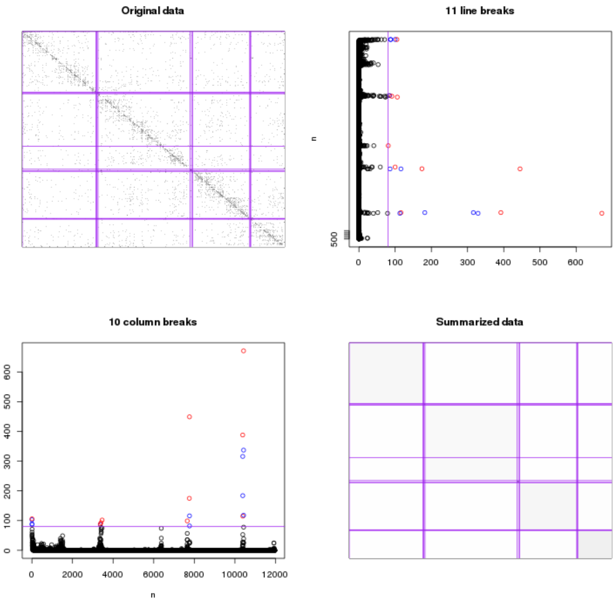


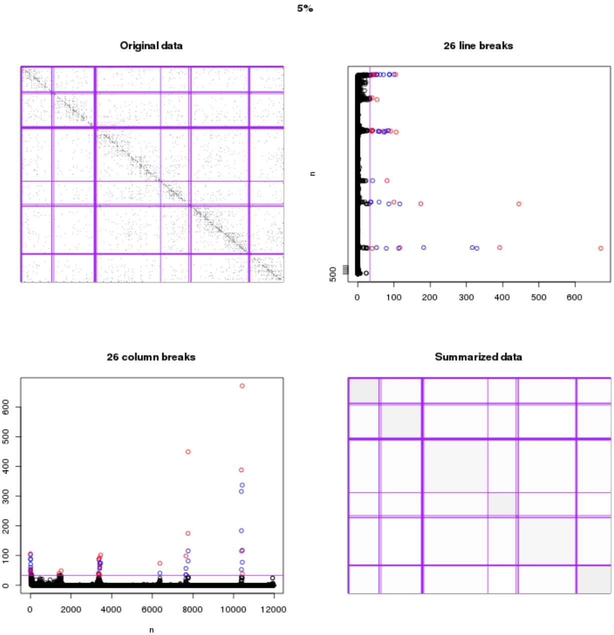
Original data
8 line breaks
n
500
0
100
200
300
400
500
600
8 column breaks
0
100
200
300
400
500
600
0
2000
4000
6000
8000
10000
12000
n
Summarized data

13%

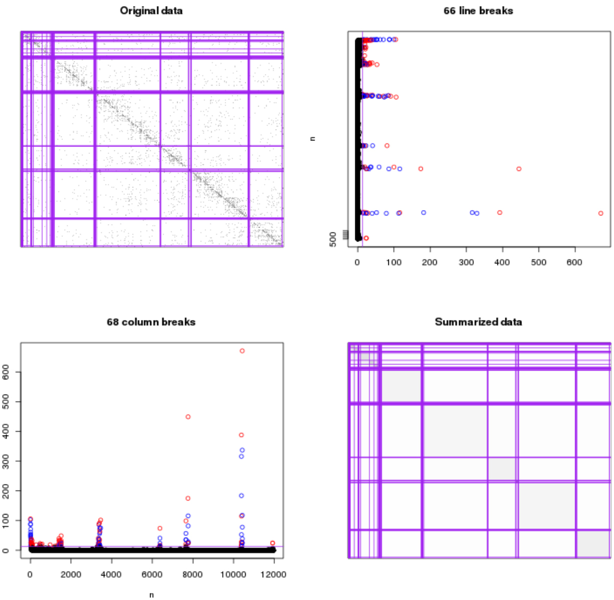

Original data
9 line breaks
n
500
0
100
200
300
400
500
600
10 column breaks
0
100
200
300
400
500
600
0
2000
4000
6000
8000
10000
12000
n
Summarized data

12%


Original data
11 line breaks
n
500
0
100
200
300
400
500
600
10 column breaks
0
100
200
300
400
500
600
0
2000
4000
6000
8000
10000
12000
n
Summarized data


5%
Original data
26 line breaks
n
500
0
100
200
300
400
500
600
26 column breaks
0
100
200
300
400
500
600
0
2000
4000
6000
8000
10000
12000
n
Summarized data

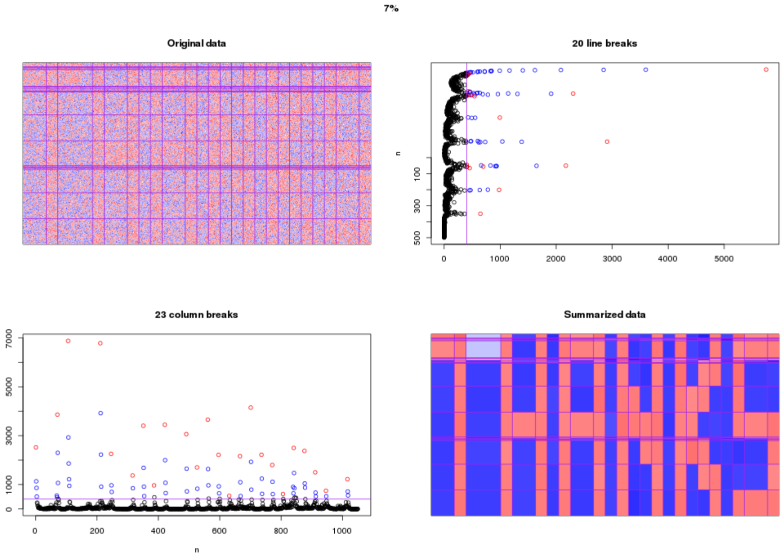
2%


Original data
66 line breaks
n
500
0
100
200
300
400
500
600
68 column breaks
0
2000
4000
6000
8000
10000
12000
n
Summarized data

Perspectives

- Theoretical result for model selection.
- Improvement of our package : *Blockseg*.
- Improvement 3D representation.
- Adaptation to symmetric matrices.
- More real datas.

Thank you for your attention



Bibliographie

F. Bach, R. Jenatton, J. Mairal, G. Obozinski, et al. Convex optimization with sparsity-inducing norms. *Optimization for Machine Learning*, pages 19–53, 2011.

J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398) :376–380, 2012.

Z. Harchaoui and C. Lévy-Leduc. Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, 105(492), 2010.

H. Hoefling. A path algorithm for the fused lasso signal approximator. *Journal of Computational and Graphical Statistics*, 19(4) :984–1006, 2010.

C. Lévy-Leduc, M. Delattre, T. Mary-Huard, and S. Robin. Two-dimensional segmentation for analyzing hic data. *Bioinformatics*, 30(17) :386–392, 2014.

E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326(5950) :289–293, 2009.

R. J. Tibshirani. *The solution path of the generalized lasso*. Stanford University, 2011.

Plan

6 Usual notations

7 Gray film

8 Optimization of the algorithm

9 Proof

Let $\mathbf{A} \in \mathcal{M}_{n\times m}(\mathbb{R})$ and $\mathbf{B} \in \mathcal{M}_{p\times q}(\mathbb{R})$ two matrices, the kronecker product of $\mathbf{A}$ and $\mathbf{B}$ is a matrix $(np) \times (mq)$ satisfying :

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & a_{nm}\mathbf{B} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1q} & a_{12}b_{11} & \cdots & \cdots & a_{1m}b_{11} & \cdots & a_{1m}b_{1q} \\ a_{11}b_{21} & a_{11}b_{22} & \cdots & a_{11}b_{2q} & a_{12}b_{21} & \cdots & \cdots & a_{1m}b_{21} & \cdots & a_{1m}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & \vdots & & & \vdots & \ddots & \vdots \\ a_{11}b_{p1} & a_{11}b_{p2} & \cdots & a_{11}b_{pq} & a_{12}b_{p1} & \cdots & \cdots & a_{1m}b_{p1} & \cdots & a_{1m}b_{pq} \\ \vdots & \vdots & & \vdots & \vdots & \ddots & & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & & \ddots & \vdots & & \vdots \\ a_{n1}b_{11} & a_{n1}b_{12} & \cdots & a_{n1}b_{1q} & a_{n2}b_{11} & \cdots & \cdots & a_{nm}b_{11} & \cdots & a_{nm}b_{1q} \\ \vdots & \vdots & \ddots & \vdots & \vdots & & & \vdots & \ddots & \vdots \\ a_{n1}b_{p1} & a_{n1}b_{p2} & \cdots & a_{n1}b_{pq} & a_{n2}b_{p1} & \cdots & \cdots & a_{nm}b_{p1} & \cdots & a_{nm}b_{pq} \end{pmatrix}.$$

Return vectorisation

$$
\begin{aligned}
\mathcal{X} &= \mathbf{T} \otimes \mathbf{T} \\
&= \begin{pmatrix}
\mathbf{T} & 0 & 0 & \cdots & 0 \\
\mathbf{T} & \mathbf{T} & 0 & \cdots & 0 \\
\vdots & & \ddots & \ddots & \vdots \\
\vdots & & & \ddots & 0 \\
\mathbf{T} & \cdots & \cdots & & \mathbf{T}
\end{pmatrix}
\end{aligned}
$$

Return vectorisation

$\|u\|_2^2$ is defined for a vector u in $\mathbb{R}^N$ by

$$\|u\|_2^2 = \sum_{i=1}^{N} u_i^2$$

and $\|u\|_1$ is defined for a vector u in $\mathbb{R}^N$ by

$$\|u\|_1 = \sum_{i=1}^{N} |u_i|.$$

Return LASSO

Plan

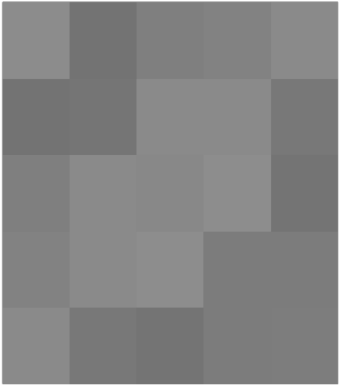
6 Usual notations

7 Gray film

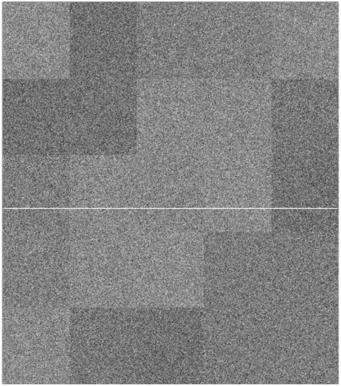
8 Optimization of the algorithm

9 Proof

Mu matrix



Orignal matrix



Colour version

Plan

6 Usual notations

7 Gray film

8 Optimization of the algorithm

9 Proof

By the form of $\mathcal{X}$, we have for all $\mathcal{V} \in \mathbb{R}^{n^2}$ with $\mathbf{v}$ the associated matrix

$$\mathcal{X}\mathcal{V} = \mathrm{Vec}\left[\left(\sum_{i'=1}^{i}\sum_{j'=1}^{j} v_{i',j'}\right)_{1\leq i,j\leq n}\right]$$

$$= \mathrm{Vec}\left[\begin{pmatrix} v_{1,1} & v_{1,1}+v_{1,2} & \cdots & \sum_{j=1}^{n} v_{1,j} \\ v_{1,1}+v_{2,1} & v_{1,1}+v_{2,1}+v_{1,2} & & \sum_{j=1}^{n} v_{1,j}+\sum_{j=1}^{n} v_{2,j} \\ \vdots & & \ddots & \vdots \\ \sum_{i=1}^{n} v_{i,1} & \sum_{i=1}^{n} v_{i,1}+\sum_{i=1}^{n} v_{i,2} & \cdots & \sum_{i=1}^{n}\sum_{j=1}^{n} v_{i,j} \end{pmatrix}\right]$$

Lemma :

For any vector $\mathcal{V} \in \mathbb{R}^{n^2}$, computing $\mathcal{X}\mathcal{V}$ and $\mathcal{X}^\top\mathcal{V}$ requires at worse $2n^2$ operations.

Return algo

By the form of $\mathcal{X}$, we have for all $\mathcal{V} \in \mathbb{R}^{n^2}$ with $\mathbf{v}$ the associated matrix

$$\mathcal{X}\mathcal{V} = \mathrm{Vec}\left[\left(\sum_{i'=1}^{i}\sum_{j'=1}^{j} v_{i',j'}\right)_{1\leq i,j\leq n}\right]$$

$$= \mathrm{Vec}\left[\begin{pmatrix} v_{1,1} & v_{1,1}+v_{1,2} & \cdots & \sum_{j=1}^{n} v_{1,j} \\ v_{1,1}+v_{2,1} & v_{1,1}+v_{2,1}+v_{1,2} & & \sum_{j=1}^{n} v_{1,j}+\sum_{j=1}^{n} v_{2,j} \\ \vdots & & \ddots & \vdots \\ \sum_{i=1}^{n} v_{i,1} & \sum_{i=1}^{n} v_{i,1}+\sum_{i=1}^{n} v_{i,2} & \cdots & \sum_{i=1}^{n}\sum_{j=1}^{n} v_{i,j} \end{pmatrix}\right]$$

Lemma :

For any vector $\mathcal{V} \in \mathbb{R}^{n^2}$, computing $\mathcal{X}\mathcal{V}$ and $\mathcal{X}^\top\mathcal{V}$ requires at worse $2n^2$ operations.

Return algo

Lemma :

Let $\mathcal{A} = \{a_1, \ldots, a_K\}$ and for each a in $\mathcal{A}$ let us consider the Euclidean division of $a - 1$ by n given by $a - 1 = nq_a + r_a$, then

$$\left((\mathcal{X}^\top \mathcal{X})_{\mathcal{A},\mathcal{A}}\right)_{1\leq k,\ell\leq K} = ((n - (q_{a_k} \vee q_{a_\ell})) \times (n - (r_{a_k} \vee r_{a_\ell})))_{1\leq k,\ell\leq K} .$$

Moreover, for any non empty subset $\mathcal{A}$ of distinct indices in $\{1, \ldots, n^2\}$, the matrix $\mathcal{X}_\mathcal{A}^\top \mathcal{X}_\mathcal{A}$ is invertible.

In some cases, we have the explicit form of $\left(\mathcal{X}_\mathcal{A}^\top \mathcal{X}_\mathcal{A}\right)^{-1}$.

Return algo

Lemma :

Let $\mathcal{A} = \{a_1, \ldots, a_K\}$ and for each a in $\mathcal{A}$ let us consider the Euclidean division of $a - 1$ by n given by $a - 1 = nq_a + r_a$, then

$$\left((\mathcal{X}^\top \mathcal{X})_{\mathcal{A},\mathcal{A}}\right)_{1\le k,\ell\le K} = ((n - (q_{a_k} \vee q_{a_\ell})) \times (n - (r_{a_k} \vee r_{a_\ell})))_{1\le k,\ell\le K} \,.$$

Moreover, for any non empty subset $\mathcal{A}$ of distinct indices in $\{1, \ldots, n^2\}$, the matrix $\mathcal{X}_{\mathcal{A}}^\top \mathcal{X}_{\mathcal{A}}$ is invertible.

In some cases, we have the explicit form of $\left(\mathcal{X}_{\mathcal{A}}^\top \mathcal{X}_{\mathcal{A}}\right)^{-1}$.

Return algo

Lemma :

Assume that we have at our disposal the Cholesky factorization of $\mathcal{X}_{\mathcal{A}}^{\top}\mathcal{X}_{\mathcal{A}}$.
The updated factorization on the extended set $\mathcal{A} \cup \{j\}$ only requires solving a $|\mathcal{A}|$-size triangular system, with complexity $\mathcal{O}(|\mathcal{A}|^2)$.
Moreover, the downdated factorization on the restricted set $\mathcal{A} \setminus \{j\}$ requires a rotation with negligible cost to preserve the triangular form of the Cholesky factorization after a column deletion.

Cholesky factorization Return algo

Cholesky factorization

Every positive-definite matrix $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ can be decompose in the product

$$\mathbf{A} = \mathbf{L}\mathbf{L}^\top$$

with $\mathbf{L}$ is a lower triangular matrix.

Return lemma

Plan

6 Usual notations

7 Gray film

8 Optimization of the algorithm

9 Proof

Notations

$$I^\star_{\min} = \min_{0\leq k\leq K_1^\star} |t^\star_{1,k+1} - t^\star_{1,k}| \wedge \min_{0\leq k\leq K_2^\star} |t^\star_{2,k+1} - t^\star_{2,k}|,$$
$$J^\star_{\min} = \min_{1\leq k\leq K_1^\star, 1\leq \ell\leq K_2^\star+1} |\mu^\star_{k+1,\ell} - \mu^\star_{k,\ell}| \wedge \min_{1\leq k\leq K_1^\star+1, 1\leq \ell\leq K_2^\star} |\mu^\star_{k,\ell+1} - \mu^\star_{k,\ell}|,$$

which corresponds to the smallest length between two consecutive change-points and to the smallest jump size between two consecutive blocks, respectively.

Return theorem

Assumption

(A1) The random variables $(E_{i,j})_{\substack{1\leq i\leq n_1 \\ 1\leq j\leq n_2}}$ are iid zero mean random variables such that there exists a positive constant β such that for all ν in $\mathbb{R}$, $\mathbb{E}[\exp(\nu E_{1,1})] \leq \exp(\beta\nu^2)$.

(A2) The sequence (δ_{n_1,n_2}) is a non increasing and positive sequence tending to zero such that $n_1\delta_{n_1,n_2}{J^\star_{\min}}^2/\log(n_2) \to \infty$ and $n_2\delta_{n_1,n_2}{J^\star_{\min}}^2/\log(n_1) \to \infty$, as n_1 and n_2 tends to infinity.

(A3) The sequence (λ_{n_1,n_2}) is such that $(n_1\delta_{n_1,n_2}J^\star_{\min})^{-1}\lambda_{n_1,n_2} \to 0$ and $(n_2\delta_{n_1,n_2}J^\star_{\min})^{-1}\lambda_{n_1,n_2} \to 0$, as n_1 and n_2 tends to infinity.

(A4) $I^\star_{\min} \geq n_1\delta_{n_1,n_2}$ and $I^\star_{\min} \geq n_2\delta_{n_1,n_2}$.

Return theorem

Lemma

Let $(Y_{i,j})_{\substack{1\leq i\leq n_1\\1\leq j\leq n_2}}$ the data matrix. Then, $\widehat{\mathcal{U}} = \mathcal{X}\widehat{\mathcal{B}}$ is such that

$$\sum_{k=r_a+1}^{n_1}\sum_{\ell=q_a+1}^{n_2} Y_{k,\ell} - \sum_{k=r_a+1}^{n_1}\sum_{\ell=q_a+1}^{n_2} \widehat{\mathcal{U}}_{k,\ell} = \frac{\lambda_{n_1,n_2}}{2}\mathrm{sign}(\widehat{\mathcal{B}}_a), \text{ if } \widehat{\mathcal{B}}_a \neq 0,$$

$$\left|\sum_{k=r_a+1}^{n_1}\sum_{\ell=q_a+1}^{n_2} Y_{k,\ell} - \sum_{k=r_a+1}^{n_1}\sum_{\ell=q_a+1}^{n_2} \widehat{\mathcal{U}}_{k,\ell}\right| \leq \frac{\lambda_{n_1,n_2}}{2}, \text{ if } \widehat{\mathcal{B}}_a = 0,$$

where $(a-1) = nq_a + r_a$.

Return theorem

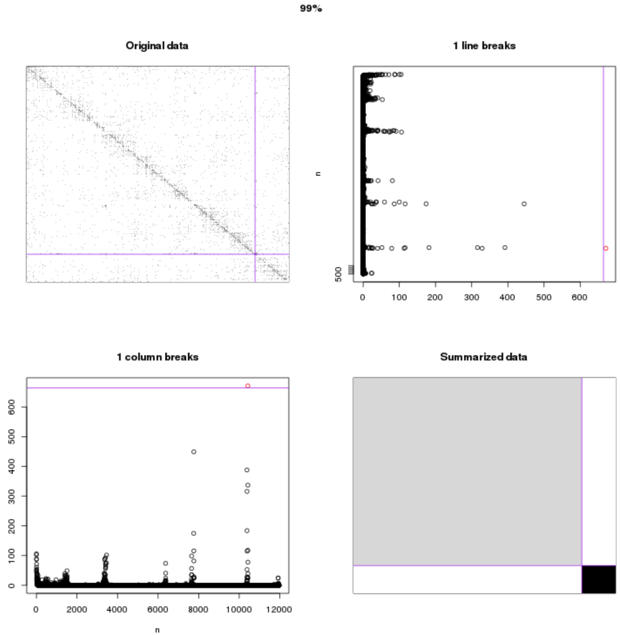
Lemma

Let $(E_{i,j})_{\substack{1\le i\le n_1\\1\le j\le n_2}}$ be random variables satisfying (A1). Let also (v_{n_1,n_2}) and (x_{n_1,n_2}) be two positive sequences such that $v_{n_1,n_2}x^2_{n_1,n_2}/\log(n_2)\to\infty$, then

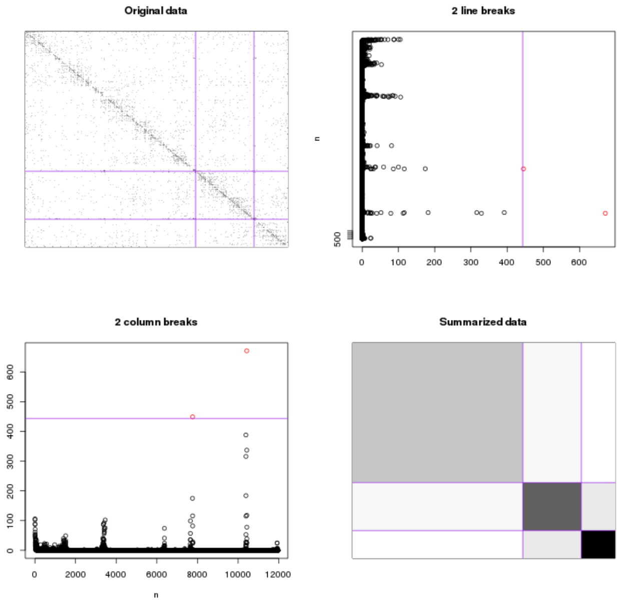
$$\mathbb{P}\left(\max_{\substack{1\le r_{n_1,n_2}<s_{n_1,n_2}\le n_2\\|r_{n_1,n_2}-s_{n_1,n_2}|\ge v_{n_1,n_2}}}\left|(s_{n_1,n_2}-r_{n_1,n_2})^{-1}\sum_{j=r_n}^{s_n-1}E_{n,j}\right|\ge x_{n_1,n_2}\right)\underset{n_1,n_2\to\infty}{\longrightarrow}0,$$

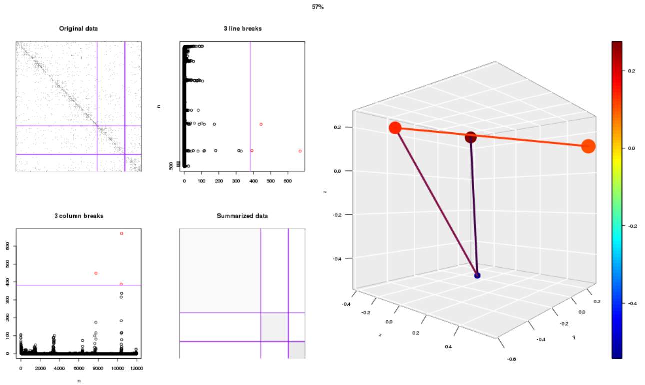
the result remaining valid if $E_{n,j}$ is replaced by $E_{j,n}$.

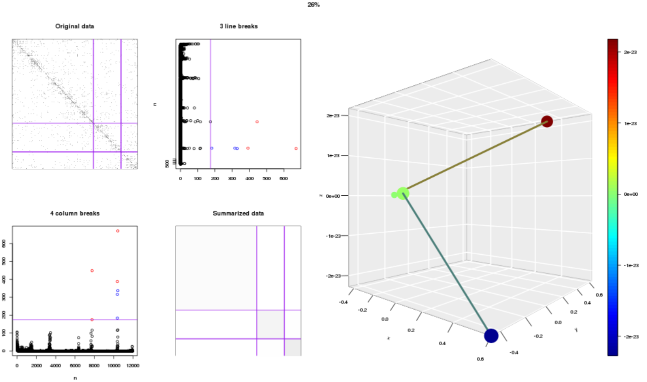
Return theorem

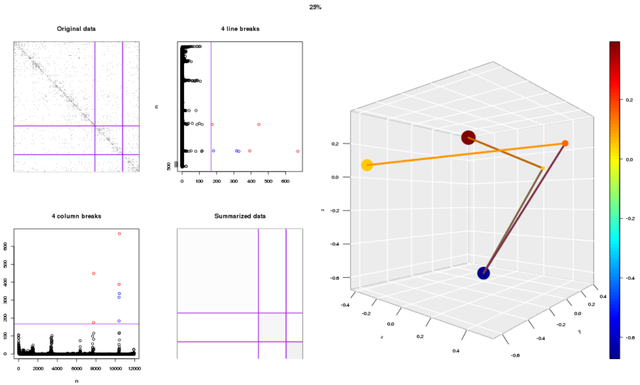

99%
Original data
1 line breaks
n
500
0
100
200
300
400
500
600
1 column breaks
600
500
400
300
200
100
0
0
2000
4000
6000
8000
10000
12000
n
Summarized data

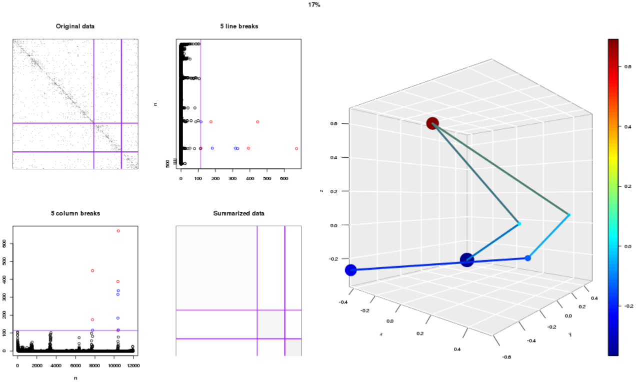
66%

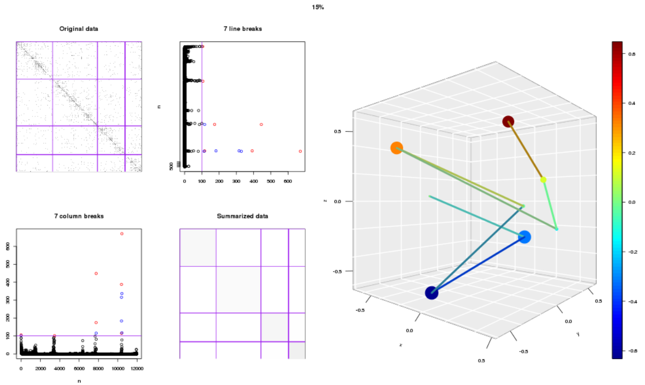

Original data
2 line breaks
n
500
0
100
200
300
400
500
600
2 column breaks
0
100
200
300
400
500
600
0
2000
4000
6000
8000
10000
12000
n
Summarized data

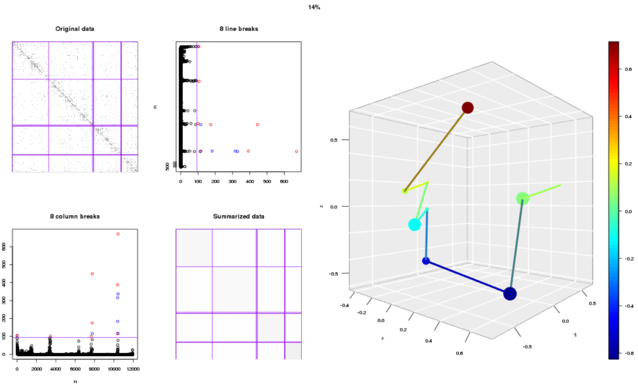

57%
Original data
3 line breaks
3 column breaks
Summarized data
n

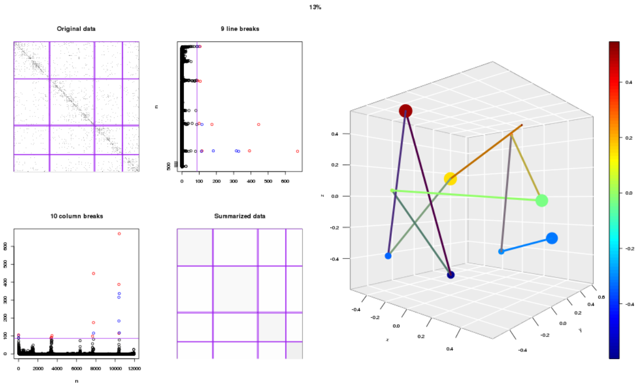

26%
Original data
3 line breaks
4 column breaks
Summarized data

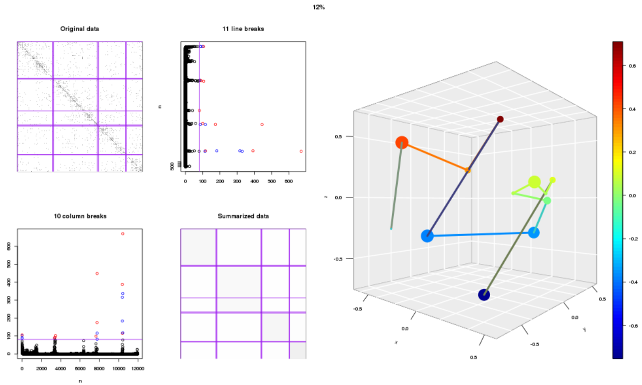

25%
Original data
4 line breaks
4 column breaks
Summarized data

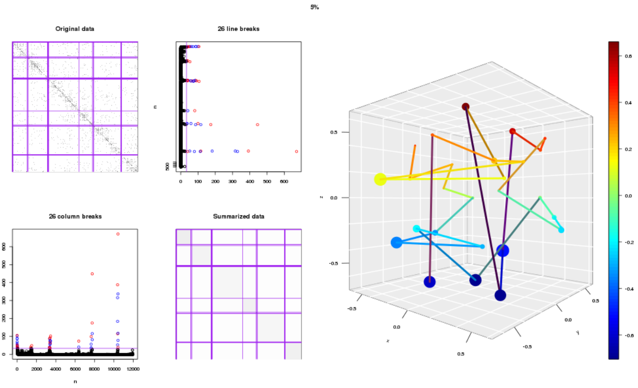

17%
Original data
5 line breaks
5 column breaks
Summarized data
n

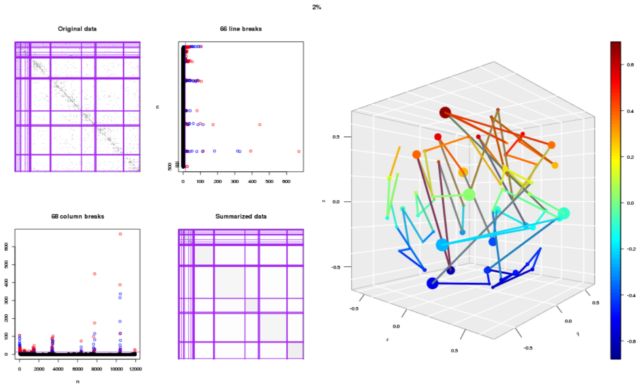

15%
Original data
7 line breaks
7 column breaks
Summarized data


14%
Original data
8 line breaks
8 column breaks
Summarized data
n


13%
Original data
9 line breaks
10 column breaks
Summarized data
n


12%
Original data
11 line breaks
10 column breaks
Summarized data
n


5%
Original data
26 line breaks
26 column breaks
Summarized data


2%
Original data
66 line breaks
68 column breaks
Summarized data