Régression en grande dimension et épistasie par blocs pour les études d’association

V. Stanislas, C. Dalmasso, C. Ambroise

Laboratoire de Mathématiques et Modélisation d’Évry

"Statistique et Génomé"
Summary

1. GWAS and Block of linkage desequilibrium
   - Genome Wide Association Studies
   - Blocks of linkage disequilibrium
   - Hierarchical Clustering with Adjacency Constraints
   - How to improve?
   - Some computation times

2. Epistasis
   - The Gene-Gene Eigen Epistasis Modeling approach
   - Simulations

3. Application
   - Ankylosing Spondylitis
   - First results
Sommaire

1 GWAS and Block of linkage desequilibrium
   • Genome Wide Association Studies
   • Blocks of linkage desequilibrium
   • Hierachical Clustering with Adjacency Constraints
   • How to improve?
   • Some computation times

2 Epistasis
   • The Gene-Gene Eigen Epistasis Modeling approach
   • Simulations

3 Application
   • Ankylosing Spondylitis
   • First results
Single-Nucleotide Polymorphism Data

- 90% of human genetic variation,
- In human genóm, SNP with allelic frequency greater than 1% are present every 300 base pairs (in average)
- 2 SNP among 3 substitute cytosine with thymine

Figure: SNP (wikipedia)
Four possible C/T configurations

\[ X_{ij} \in \{0, 1, 2\} \text{ (Individual } i \text{ at locus } j) \]

<table>
<thead>
<tr>
<th></th>
<th>Father</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>Mother</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

High-dimension

\[ n \text{ individuals} \begin{pmatrix} \mathbf{X}_{11} & \cdots & \cdots & \mathbf{X}_{1p} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{X}_{n1} & \cdots & \cdots & \mathbf{X}_{np} \end{pmatrix} \approx \text{up to few million of SNPs can be genotyped!} \]

\[ \Rightarrow \text{number of variables } \gg \text{number of individuals } (p \gg n) \]
Genome-Wide Association Studies

GWAS characteristics:

- **Objective**: find associations between genetic markers \((SNP_{i,j} \in \{0, 1, 2\})\) and a phenotypic trait \((Y_i \in \{0, 1\}\) or \(Y_i \in \mathbb{R}\))

- **Genetic markers**: SNP

![Diagram showing Patient DNA and Non-patient DNA with Disease-specific SNPs and Non-disease SNPs](http://www.siriusgenomics.com/technology/)
**Generalized Linear Model**

\[
g(E[Y_i|x_i]) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} , i = 1, \ldots, n
\]

- \(n\) : number of individuals
- \(p\) : number of covariates
- \(Y_i\) : response for the individual \(i\)
- \(x_{ij}\) : observations for covariate \(j\) (coded in 0, 1 or 2)
Genome-Wide Association Studies

- **SNP analysis**
  Differences between cases and controls at a specific SNP

- **GWAS limits:**
  - Reproductibility
  - Heritability

- **Data particularities:**
  - Structuration
  - High dimension (p » n)
  - Small effects
The LD measures

Linkage Desequilibrium
- non-random association of alleles at two or more loci
- depends on the difference between observed allelic frequencies and those expected from a independent randomly distributed model.

Computation
- $Z_j$ the indicator of the presence of minor allele for SNP $j$.
- $Z_j \sim \text{B}(p_j)$
- $D(j, k) = p_{jk} - p_j p_k = E[Z_j Z_k] - p_j p_k = \text{cov}(Z_j, Z_k)$
- $r^2(j, k) = \text{corr}(Z_j, Z_k)$
- ou
- $D'(j, k) = D(j, k)/D_{\text{max}}(j, k)$
How to estimate LD?

<table>
<thead>
<tr>
<th>snp</th>
<th>vv</th>
<th>vV</th>
<th>VV</th>
</tr>
</thead>
<tbody>
<tr>
<td>uu</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>uU</td>
<td>d</td>
<td>e</td>
<td>f</td>
</tr>
<tr>
<td>UU</td>
<td>g</td>
<td>h</td>
<td>i</td>
</tr>
</tbody>
</table>

⚠️ Only the genotype data table is observed

- $\alpha$, $\beta$, $\gamma$, $\delta$ are estimated
- a system of equations. e.g. $\alpha = 2a + b + d + pe$

with $p$ the "probability" of the haplotype $(uv, UV)$.

$\Rightarrow$ estimating $p$, then $(\alpha, \beta, \gamma, \delta)$ and finally

$D = p_{uv} - p_{upv}$. 
The LD block structure

- the $r^2$ coefficients among the 50 first SNPs of the Chromosome 22 (Dalmasso et al. 2008)
- LD structured in blocks
Hierarchical Clustering with Adjacency Constraints

Cluster Dendrogram

LD
corr
Block-Wise Approach using Linkage Disequilibrium (BALD)

1. Hierarchical clustering of the SNPs with adjacency constraint and using the LD similarity.
2. Estimation of the optimal number of groups using the Gap statistic (Tibshirani et. al., 2001).
The h-band

- All coefficients outside the band “h” are null
- a $p \times h$ similarity matrix

$\Rightarrow$ a hierarchical clustering with adjacency constraint
A pseudocode

**Data:** $\mathbf{X} \in \{0, 1, 2\}^{n \times p}$, $\text{Sim}$

$\mathcal{C} \leftarrow \{ C_i = \{ \mathbf{X}_{.i} \}, i \in 1, \ldots, p \}$ /* clusters = singletons */ ;

$D \leftarrow \{ 1 - \text{Sim}(\mathbf{X}_{.i}, \mathbf{X}_{.(i+1)}) \}, i \in 1, \ldots, p - 1 \}$ ;

**for** step $= 1$ **to** $p - 1$ **do**

\[
i^* \leftarrow \arg\min_{i \in \{1, \ldots, p-\text{step}\}} D(C_i, C_{i+1}) ;
\]

$\mathcal{C} \leftarrow \mathcal{C} \setminus \{ C_{i^*}, C_{i^*+1} \} \cup \{ C_{i^*} \cup C_{i^*+1} \} ;$

$d_1 \leftarrow D(C_{i^*-1}, C_{i^*} \cup C_{i^*+1}) ;$

$d_2 \leftarrow D(C_{i^*} \cup C_{i^*+1}, C_{i^*+2}) ;$

$D \leftarrow D \setminus \{ D(C_{i^*-1}, C_{i^*}), D(C_{i^*}, C_{i^*+1}) \} \cup \{ d_1, d_2 \} ;$

**end**
Ward Constrained Hierarchical Clustering

\[ d(A, B) = \frac{n_A n_B}{n_A + n_B} \left( \frac{1}{n_A^2} S_{A,A} + \frac{1}{n_B^2} S_{B,B} - \frac{2}{n_A n_B} S_{A,B} \right) \]
The pencils’ trick: Calculating $S_{AA}$ and $S_{AB}$
The “pencils”

Assessing $S_{AA}$, $S_{BB}$ and $S_{AB}$ requires the calculation of sums of LD measures within *pencil-shaped areas* defined by:

- direction: right or left
- depth: $h_{\text{Loc}}$
- end point: $\text{lim}$

$\Rightarrow$ Two arrays of sizes $p \times h$ for storing the pencils sums.
The binary min-heap

- All nodes are either **less than or equal** to each of its children.

- Uniquely represented by storing its level order traversal in an array. Given a position $i$:
  - Parent($i$) = $\lfloor i/2 \rfloor$
  - Left($i$) = $2i$
  - Right($i$) = $2i + 1$
DeleteMin

Time complexity : $O(\log(p))$
InsertHeap

Time complexity: $O(\log(p))$
BuildHeap

**Data:** An array $A$

**Result:** A min-heap $H$

for $i = \lfloor length(A)/2 \rfloor$ down to 1

| PercolDown($A, i$); |

end

Time complexity: $O(p \log(p))$
<table>
<thead>
<tr>
<th></th>
<th>findMin</th>
<th>insert</th>
<th>deleteMin</th>
</tr>
</thead>
<tbody>
<tr>
<td>unordered array</td>
<td>$O(p)$</td>
<td>$O(1)$</td>
<td>$O(p)$</td>
</tr>
<tr>
<td>binary heap</td>
<td>$O(1)$</td>
<td>$O(\log(p))$</td>
<td>$O(\log(p))$</td>
</tr>
</tbody>
</table>
Figure: The mean computation time $t$ versus the number of markers $p$ for the cWard algorithm applied to randomly sampled SNP matrices. $N = 100$, $h = 30$ and $t$ is averaged across 50 simulation runs.
Compared to a former implementation

Figure: The mean computation time $t$ versus the number of markers $p$ for the cWard algorithm and an implementation without heaps. $t$ is averaged across 20 simulation runs.
Scalable Hierarchical Clustering with pencils and binary heap

To sum up:

1. A $\mathcal{O}(p^2)$ algorithm does not scale for GWA studies.
2. The Ward distance written in a simple way.
3. Space complexity of $\mathcal{O}(ph)$ by using the pencils’ trick.
4. Time complexity of:

\[
\mathcal{O}(ph) + \mathcal{O}(p\log(p))
\]

2 $p \times h$ arrays of pencils’ sums

building the heap and insert/delete heaps’ operations within the loop

Ongoing work:

- Currently implemented with a genotype matrix as input.
  - can be generalized to any band similarity matrix.
Sommaire

1 GWAS and Block of linkage desequilibrium
   - Genome Wide Association Studies
   - Blocks of linkage desequilibrium
   - Hierarchical Clustering with Adjacency Constraints
   - How to improve?
   - Some computation times

2 Epistasis
   - The Gene-Gene Eigen Epistasis Modeling approach
   - Simulations

3 Application
   - Ankylosing Spondylitis
   - First results
Epistasis

Definition
Interaction of alleles effects from different markers

Existing methods
- mainly SNP x SNP
- some at the block (gene) scale

Advantages of gene (or block) scale approaches
- results biologically interpretable
- genetic effects may be easier to detect
- reduce the number of variables
Epistasis - Gene scale methods

Existing gene scale methods:

Two or few genes
- PCA + logistic regression \((\text{He et al. 2011, Li et al. 2009, Zhang et al. 2008})\)
- PLS + logistic regression \((\text{Wang T et al. 2009})\)

For a larger number of genes
- PCA + LASSO \((\text{D’Angelo et al. 2009})\)
- PCA + pathway-guided penalized regression \((\text{Wang X et al. 2014})\)
Epistasis - Gene scale methods

Existing gene scale methods:

- Two or few genes
  - PLS + logistic regression (*Wang T et al.* 2009)

- For a larger number of genes
  - PCA + LASSO (*D’Angelo et al.* 2009)
  - PCA + pathway-guided penalized regression (*Wang X et al.* 2014)

Objectives:

- To develop a new gene scale method:
  - considers a more accurate definition of interaction variables,
  - is applicable with with genes,
  - takes into account the group structure
Group modeling approach

<table>
<thead>
<tr>
<th></th>
<th>SNP_{1,1}</th>
<th>..</th>
<th>SNP_{1,p_1}</th>
<th>..</th>
<th>SNP_{G,1}</th>
<th>..</th>
<th>SNP_{G,p_G}</th>
<th>Pheno</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind_1</td>
<td>1</td>
<td>..</td>
<td>0</td>
<td>..</td>
<td>0</td>
<td>..</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ind_2</td>
<td>0</td>
<td>..</td>
<td>0</td>
<td>..</td>
<td>2</td>
<td>..</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>..</td>
<td>2</td>
<td>..</td>
<td>1</td>
<td>..</td>
<td>1</td>
<td>..</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Ind_i</td>
<td>0</td>
<td>..</td>
<td>1</td>
<td>..</td>
<td>0</td>
<td>..</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ \text{gene}_1 \]
\[ \text{gene}_G \]

We note \( SNP_{1,1} = X_{1,1} \)
We note $SNP_{1,1} = X_{1,1}$, $r, s$ two genes.

<table>
<thead>
<tr>
<th></th>
<th>$SNP_{1,1}$</th>
<th>$SNP_{1,p_1}$</th>
<th>$SNP_{G,1}$</th>
<th>$SNP_{G,p_G}$</th>
<th>Pheno</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind_1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ind_2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ind_i</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

$\text{gene}_1$

$\text{gene}_G$
Group modeling approach

**Model**

\[
Y_i = \beta_0 + \sum_g \beta_g \left( \sum_{k \in C} X_{ik}^g \right) + \epsilon_i
\]

**Main effects**

\[
\beta = \begin{pmatrix} \beta_{1,1}, \beta_{1,2}, \ldots, \beta_{1,p_1}, \ldots, \beta_{G,1}, \ldots, \beta_{G,p_G} \end{pmatrix}^T
\]
Group modeling approach

Model

$$Y_i = \beta_0 + \sum_{g} \beta_g \left( \sum_{k \in C} X_{ik}^g \right) + \sum_{r,s} \gamma_{r,s} Z \_{r,s}^i + \epsilon_i$$

Main effects

$$\beta = \begin{pmatrix} \beta_{1,1}, \beta_{1,2}, \ldots, \beta_{1,p_1}, \ldots, \beta_{G,1}, \ldots, \beta_{G,p_G} \end{pmatrix}_{\text{gene}_1}^{\text{gene}_G}$$

Interaction effects

$$\gamma = \begin{pmatrix} \gamma_{12}, \ldots, \gamma_{1G}, \ldots, \gamma_{(G-1)G} \end{pmatrix}$$

$$q : \text{# of interaction variables for a couple}$$
We consider $f_u(X^r_i, X^s_i)$ to represent the interaction between genes $r, s$.

$$\hat{u} = \arg \max_{u, \|u\|=1} \text{cor}(y, f_u(X^r, X^s))$$

**Eigen Epistasis**

$$f_u(X^r, X^s) = W^{rs} u$$

with

$$W^{rs} = \{ X_{ij} X_{ik} \}^{j=1\ldots,p_r, k=1\ldots,p_s}_{i=1\ldots,n}$$

$$\max_{u, \|u\|=1} \| c\text{or}[W^{rs} u, y] \|^2 = \max_{u, \|u\|=1} u^T W^{rs} y y^T W^{rs} u$$

$u$ : only eigen vector associated of $W^{rs^T} y y^T W^{rs}$

For each couple $(r, s)$ \( Z^{rs} = W^{rs^T} u \).
Coefficients estimation

**Group LASSO regression**

\[
(\hat{\beta}, \hat{\gamma}) = \arg\min_{\beta, \gamma} \sum_i (y_i - X_i \beta - Z_i \gamma)^2 + \\
\lambda \left( \sum_g \sqrt{p_g} \| \beta^g \|_2^2 + \sum_{rs} \sqrt{p_r p_s} \| \gamma^{rs} \|_2^2 \right)
\]

Limits of the group LASSO regression:

- Difficult to compute p-value or confidence interval
Coefficients estimation

**Group LASSO regression**

\[
(\hat{\beta}, \hat{\gamma}) = \arg\min_{\beta, \gamma} \sum_i (y_i - X_i \beta - Z_i \gamma)^2 + \\
\lambda \left( \sum_g \sqrt{p_g} \| \beta^g \|_2 + \sum_{rs} \sqrt{p_r p_s} \| \gamma^{rs} \|_2 \right)
\]

Limits of the group LASSO regression:

- Difficult to compute p-value or confidence interval

**Adaptive-Ridge Cleaning**  
*Becu JM, 2015*

- Use of a specific penalty for group LASSO
- Permutation test based on Fisher test approach for each group

\[
P_k = \frac{1}{B} \#\{F^*_k \geq F_k\}\]
## Interaction variable modeling approaches comparison

<table>
<thead>
<tr>
<th>Methods</th>
<th>Criteria</th>
<th>$Z_i^{rs} \gamma^{rs}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-GEE</td>
<td>$\text{cor}(Y, f_u(X^r, X^s))$</td>
<td>$W^{rs} u \gamma^{rs}$</td>
</tr>
<tr>
<td>PCA</td>
<td>$\text{var}(G_r v)$ and $\text{var}(G_s v)$</td>
<td>$\sum_{j=1}^{q} \sum_{k=1}^{q} \gamma^{rs}_{jk} C^r_j C^s_k$</td>
</tr>
<tr>
<td>CCA</td>
<td>$\text{cor}(G_r a, G_s b)$</td>
<td>$\sum_{j=1}^{q} \gamma^{rs}_{j} A^r_j B^s_j$</td>
</tr>
<tr>
<td>PLS</td>
<td>$\text{cov}(YG_r c, G_s w)$</td>
<td>$\sum_{j=1}^{q} \gamma^{rs}_{j} T^{rs}_j$</td>
</tr>
</tbody>
</table>
Simulation design

Genotype:
\( X_i \sim \mathcal{N}_p(0, \Sigma) \) with \( \Sigma \) a block diagonal correlation matrix
\((\rho = 0.8 \text{ for two SNPs in the same gene})\)

\( MAF_j \sim \mathcal{U}[0.05, 0.5] \) with \( MAF_j = 0.2 \) if \( j \) causal SNP

Continuous phenotype simulated under two different schemes:

\( \rightarrow \) from Wang X et al., 2014:

\[
Y_i = \beta_0 + \sum_g \beta_g \left( \sum_{k \in C} X_{ik}^g \right) + \sum_{rs} \gamma_{rs} \left( \sum_{(j,k) \in C^2} X_{ij}^r X_{ik}^s \right) + \epsilon_i \quad (1)
\]

\( \rightarrow \) PCA model:

\[
Y_i = \beta_0 + \sum_g \beta_g \left( \sum_{k \in C} X_{ik}^g \right) + \sum_{rs} \gamma_{rs} C_{i1}^r C_{i1}^s + \epsilon_i. \quad (2)
\]
Simulations design

Scenarios:

We consider 600 subjects and 6 SNPs by gene

→ First scenario on 6 genes, two settings:
  - same genes for main and interaction effects,
  - different genes for main and interaction effects.

→ Second scenario on 25 genes, one setting:
  - different genes for main and interaction effects.
Simulations results - First scenario on 6 genes

→ Main effects:
   - gene 1
   - gene 2

→ Interaction effects:
   - gene 1 x gene 2

→ Main effects:
   - gene 1
   - gene 2

→ Interaction effects:
   - gene 3 x gene 4
Simulations results - First scenario on 6 genes

Wang X et al. model

- Main effects: gene 1, gene 2
- Interaction effects: gene 1 x gene 2

PCA model

- Main effects: gene 1, gene 2
- Interaction effects: gene 3 x gene 4
Simulations results - Second scenario on 25 genes

Figure: Wang X et al. model, $r^2 = 0.7$

→ Main effects:
  - gene 1
  - gene 2

→ Interaction effects:
  - gene 3 x gene 4
  - gene 5 x gene 6
  - gene 7 x gene 8
  - gene 9 x gene 10
1. GWAS and Block of linkage disequilibrium
   - Genome Wide Association Studies
   - Blocks of linkage disequilibrium
   - Hierarchical Clustering with Adjacency Constraints
   - How to improve?
   - Some computation times

2. Epistasis
   - The Gene-Gene Eigen Epistasis Modeling approach
   - Simulations

3. Application
   - Ankylosing Spondylitis
   - First results
Ankylosing Spondylitis

Chronic inflammatory disease of the axial skeleton

Epidemiology:
- Age at first symptoms: 20 - 30 years
- Sex: predominance for men (sex ratio 2M:1W)
- Prevalence: depend on populations (0.1% - 1.4%)

Right etiology unknown:
- Environmental factors?
- Genetic factors?
  - Importance of HLA complex

HLA complex:
- Localized on chromosome 6
- Regroup about 200 genes
- Coding the immunity system
- Antigen HLA-B27: associated to SPA
Known genes

<table>
<thead>
<tr>
<th>Known genes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RUNX3</td>
<td>Runt-related transcription factor 3</td>
</tr>
<tr>
<td>IL23R</td>
<td>Interleukin 23 receptor</td>
</tr>
<tr>
<td>IL12Rβ2</td>
<td>Interleukin 12 receptor, β2</td>
</tr>
<tr>
<td>GRP25</td>
<td>G-protein-coupled receptor 25</td>
</tr>
<tr>
<td>KIF21B</td>
<td>Kinesin family member 21B</td>
</tr>
<tr>
<td>PTGER4</td>
<td>Prostaglandin E receptor 4 (subtype EP3)</td>
</tr>
<tr>
<td>ERAP1</td>
<td>Endoplasmic reticulum aminopeptidase 1</td>
</tr>
<tr>
<td>ERAP2</td>
<td>Endoplasmic reticulum aminopeptidase 2</td>
</tr>
<tr>
<td>LNPEP</td>
<td>Leucyl/cystyl aminopeptidase</td>
</tr>
<tr>
<td>IL12B</td>
<td>Interleukin 12B</td>
</tr>
<tr>
<td>CARD9</td>
<td>Caspase recruitment-domain family member 9</td>
</tr>
<tr>
<td>LTαR</td>
<td>Lymphotixin β-receptor (TNFR superfamily, member 3)</td>
</tr>
<tr>
<td>TNFRSF1A</td>
<td>Tumor-necrosis factor-receptor superfamily member 1A</td>
</tr>
<tr>
<td>NPEPS</td>
<td>Aminopeptidase puromycin-sensitive</td>
</tr>
<tr>
<td>TBx2P1</td>
<td>TNFR-associated factor family member-associated nuclear factor-κB-binding kinase 1-binding protein</td>
</tr>
<tr>
<td>TBX2</td>
<td>T-box 21</td>
</tr>
</tbody>
</table>

IL6R        : Interleukin 6 receptor
FCGR2A      : Fc fragment of immunoglobulin G, low-affinity IIa, receptor (CD32)
UBE2E3      : Ubiquitin-conjugating enzyme E2E 3
GPR35       : G-protein-coupled receptor 35
NKX2-3      : NK2 homeobox 3
ZMIZ1       : Zinc finger, MIZ type-containing 1
SH2B3       : Src homology 2B adaptor protein 3
GPR65       : G-protein-coupled receptor 65
IL27        : Interleukin 27
SULT1A1     : Sulphotransferase family cytosolic 1A
TYK2        : Tyrosine kinase 2
ICOSLG      : Inducible T-cell costimulator ligand
EOMES       : Eomesoderm
IL7R        : Interleukin 7 receptor
BACH2       : BTB and CNC homology 1, basic leucine-zipper transcription-factor 2

Abbreviation: CD, classification determinant.

→ 29 susceptibility genes identified by GWAS
<table>
<thead>
<tr>
<th>Method</th>
<th>Significant results</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-GEE</td>
<td>HLA-B x SULT1A1</td>
</tr>
<tr>
<td></td>
<td>IL23R x ERAP2</td>
</tr>
<tr>
<td>PLS</td>
<td>HLA-B</td>
</tr>
<tr>
<td></td>
<td>EOMES x BACH2</td>
</tr>
<tr>
<td>PCA</td>
<td>HLA-B</td>
</tr>
<tr>
<td>CCA</td>
<td>-</td>
</tr>
</tbody>
</table>
Conclusions and perspectives

The G-GEE method:
- Takes into account the gene structure of data
- Can be applied on a large number of genes
- Uses a specific interaction modeling approach

Ankylosing Spondylitis:
- Identification of potential interactions to discuss with doctors
- HLA-B effect

Perspectives:
- Explore new $f_u(X^r_i, X^s_i)$ definition
- Additional simulations on larger data set
- New applications on other data set
Thank you for your attention!
Adaptive-Ridge Cleaning

specific penalty for group LASSO: \[ \lambda \frac{1}{\sqrt{|k(j)|} \sum_{m \in k(j)} \hat{\theta}_m^2} \]