Group lasso for genomic data

Jean-Philippe Vert

Mines ParisTech and Curie Institute

Machine learning: Theory and Computation workshop,
IMA, Minneapolis, March 26-30, 2012

Outline

1. Motivations
2. Finding multiple change-points in a single profile
3. Finding multiple change-points shared by many signals
4. Learning molecular classifiers with network information
5. Conclusion

Outline

1. Motivations

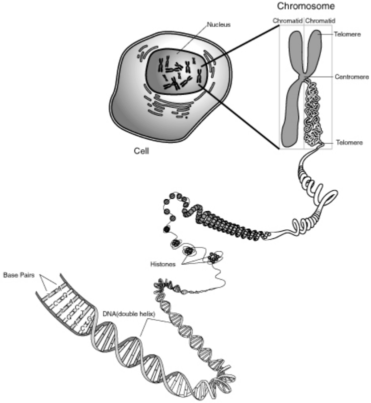
2. Finding multiple change-points in a single profile

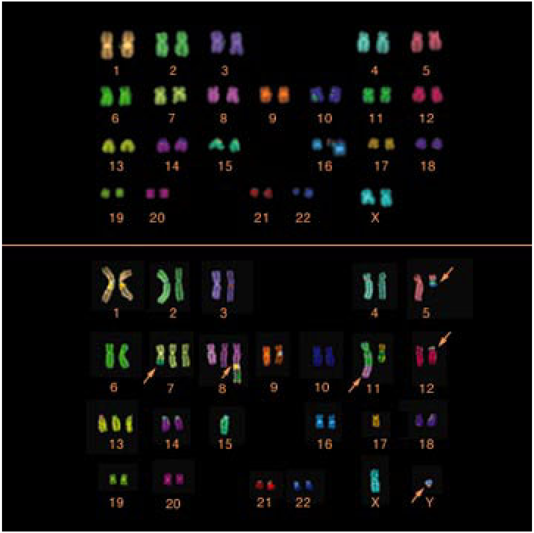
3. Finding multiple change-points shared by many signals

4. Learning molecular classifiers with network information

5. Conclusion

Chromosomic aberrations in cancer

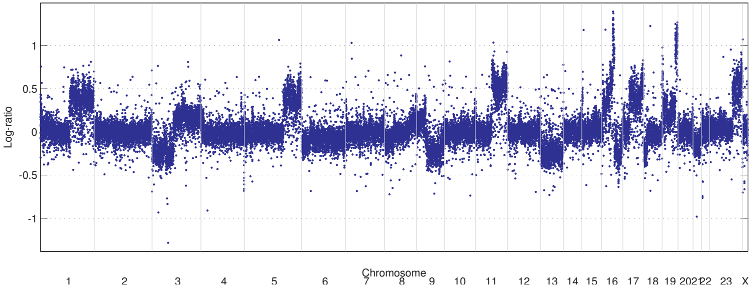




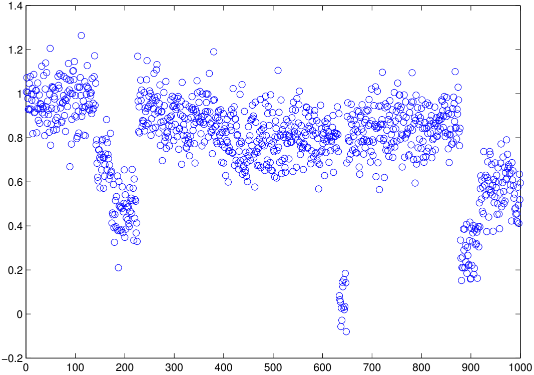
Comparative Genomic Hybridization (CGH)

Motivation

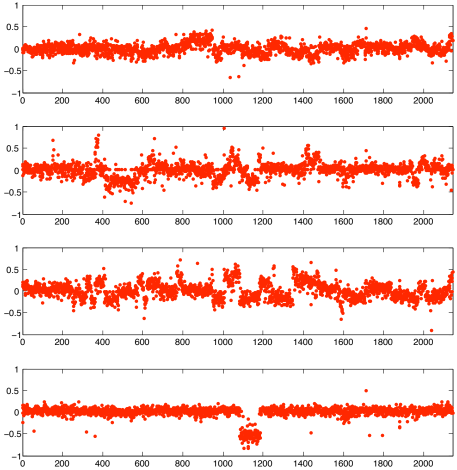
- Comparative genomic hybridization (CGH) data measure the DNA copy number along the genome
- Very useful, in particular in cancer research to observe systematically variants in DNA content



Can we identify breakpoints and "smooth" each profile?

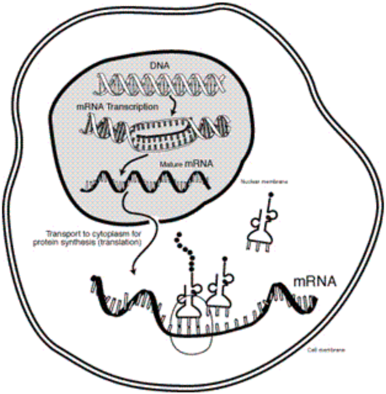


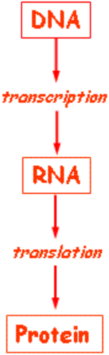
Can we detect frequent breakpoints?



A collection of bladder tumour copy number profiles.

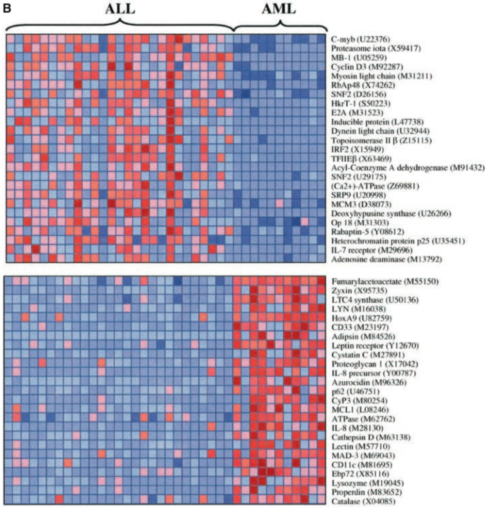
DNA → RNA → protein



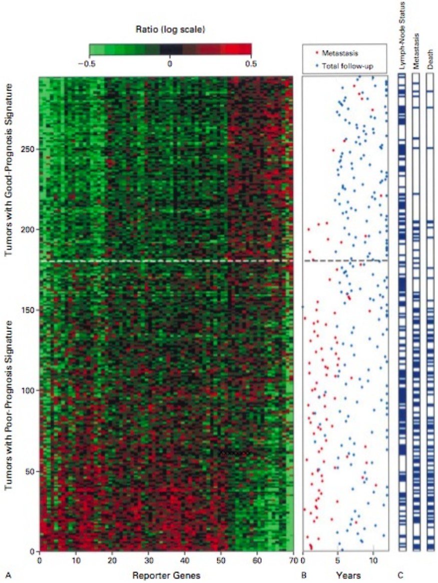


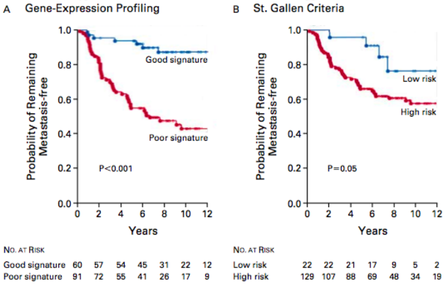
- CGH shows the (static) DNA
- Cancer cells have also abnormal (dynamic) gene expression (= transcription)

Can we identify the cancer subtype? (diagnosis)



Can we predict the future evolution? (prognosis)

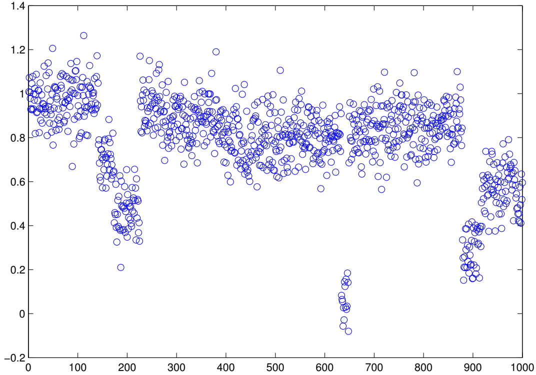




Outline

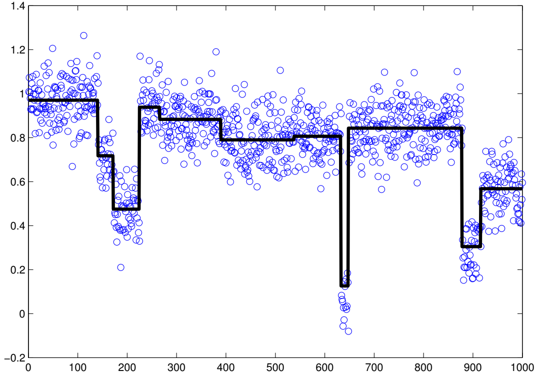
1. Motivations
2. Finding multiple change-points in a single profile
3. Finding multiple change-points shared by many signals
4. Learning molecular classifiers with network information
5. Conclusion

The problem



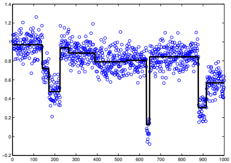
- Let $Y \in \mathbb{R}^p$ the signal
- We want to find a piecewise constant approximation $\hat{U} \in \mathbb{R}^p$ with at most k change-points.

The problem



- Let $Y \in \mathbb{R}^p$ the signal
- We want to find a piecewise constant approximation $\hat{U} \in \mathbb{R}^p$ with at most k change-points.

An optimal solution?

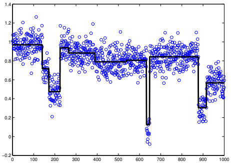


- We can define an "optimal" piecewise constant approximation $\hat{U} \in \mathbb{R}^p$ as the solution of

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

- This is an optimization problem over the $\binom{p}{k}$ partitions...
- Dynamic programming finds the solution in $O(p^2k)$ in time and $O(p^2)$ in memory
- But: does not scale to $p = 10^6 \sim 10^9$...

An optimal solution?

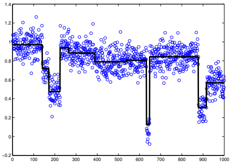


- We can define an "optimal" piecewise constant approximation $\hat{U} \in \mathbb{R}^p$ as the solution of

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

- This is an optimization problem over the $\binom{p}{k}$ partitions...
- Dynamic programming finds the solution in $O(p^2 k)$ in time and $O(p^2)$ in memory
- But: does not scale to $p = 10^6 \sim 10^9$...

An optimal solution?

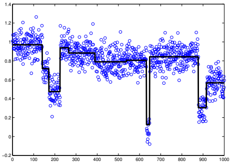


- We can define an "optimal" piecewise constant approximation $\hat{U} \in \mathbb{R}^p$ as the solution of

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

- This is an optimization problem over the $\binom{p}{k}$ partitions...
- Dynamic programming finds the solution in $O(p^2 k)$ in time and $O(p^2)$ in memory
- But: does not scale to $p = 10^6 \sim 10^9$...

An optimal solution?



- We can define an "optimal" piecewise constant approximation $\hat{U} \in \mathbb{R}^p$ as the solution of

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

- This is an optimization problem over the $\binom{p}{k}$ partitions...
- Dynamic programming finds the solution in $O(p^2 k)$ in time and $O(p^2)$ in memory
- But: does not scale to $p = 10^6 \sim 10^9$...

Promoting sparsity with the ℓ_1 penalty

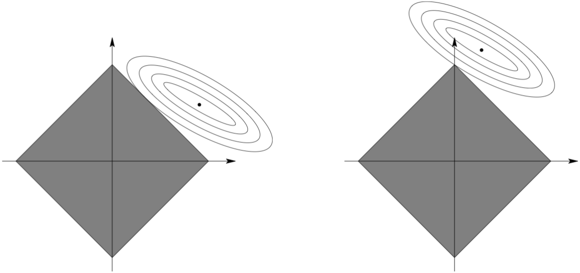
The ℓ_1 penalty (Tibshirani, 1996; Chen et al., 1998)

If $R(\beta)$ is convex and "smooth", the solution of

$$\min_{\beta \in \mathbb{R}^p} R(\beta) + \lambda \sum_{i=1}^{p} |\beta_i|$$

is usually sparse.

Geometric interpretation with $p = 2$



Promoting piecewise constant profiles penalty

The total variation / variable fusion penalty

If $R(\beta)$ is convex and "smooth", the solution of

$$\min_{\beta \in \mathbb{R}^p} R(\beta) + \lambda \sum_{i=1}^{p-1} |\beta_{i+1} - \beta_i|$$

is usually piecewise constant (Rudin et al., 1992; Land and Friedman, 1996).

Proof:

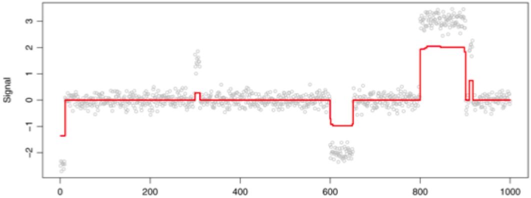
- Change of variable $u_i = \beta_{i+1} - \beta_i$, $u_0 = \beta_1$
- We obtain a Lasso problem in $u \in \mathbb{R}^{p-1}$
- u sparse means β piecewise constant

TV signal approximator

$$\min_{\beta \in \mathbb{R}^p} \| Y - \beta \|^2 \quad \text{such that} \quad \sum_{i=1}^{p-1} | \beta_{i+1} - \beta_i | \leq \mu$$

Adding additional constraints does not change the change-points:

- $\sum_{i=1}^{p} | \beta_i | \leq \nu$ (Tibshirani et al., 2005; Tibshirani and Wang, 2008)
- $\sum_{i=1}^{p} \beta_i^2 \leq \nu$ (Mairal et al. 2010)

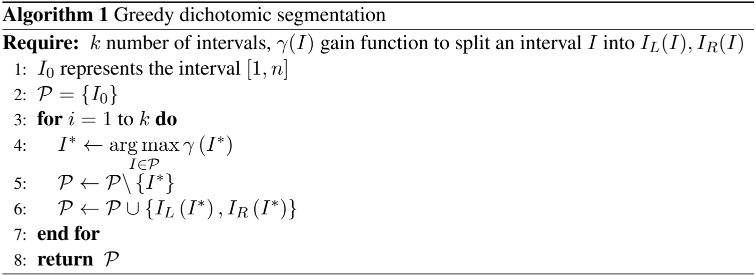


Solving TV signal approximator

$$\min_{\beta \in \mathbb{R}^p} \| Y - \beta \|^2 \quad \text{such that} \quad \sum_{i=1}^{p-1} | \beta_{i+1} - \beta_i | \leq \mu$$

- QP with sparse linear constraints in $O(p^2)$ -> 135 min for $p = 10^5$ (Tibshirani and Wang, 2008)
- Coordinate descent-like method $O(p)$? -> 3s s for $p = 10^5$ (Friedman et al., 2007)
- For all μ with the LARS in $O(pK)$ (Harchaoui and Levy-Leduc, 2008)
- For all μ in $O(p \ln p)$ (Hoefling, 2009)
- For the first K change-points in $O(p \ln K)$ (Bleakley and V., 2010)

TV signal approximator as dichotomic segmentation



Algorithm 1 Greedy dichotomic segmentation

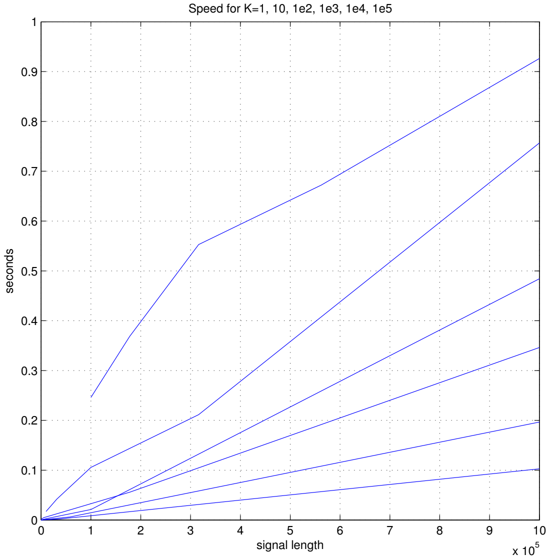
```
Require: $k$ number of intervals, $\gamma(I)$ gain function to split an interval $I$ into $I_L(I), I_R(I)$
$I_0$ represents the interval $[1, n]$
$\mathcal{P} = \{I_0\}$
for $i = 1$ to $k$ do
    $I^* \leftarrow \arg\max_{I \in \mathcal{P}} \gamma(I^*)$
    $\mathcal{P} \leftarrow \mathcal{P} \backslash \{I^*\}$
    $\mathcal{P} \leftarrow \mathcal{P} \cup \{I_L(I^*), I_R(I^*)\}$
end for
return $\mathcal{P}$
```

Theorem

TV signal approximator performs "greedy" dichotomic segmentation

(V. and Bleakley, 2010; see also Hoefling, 2009)

Speed trial : 2 s. for $K = 100$, $p = 10^7$



Outline

1 Motivations

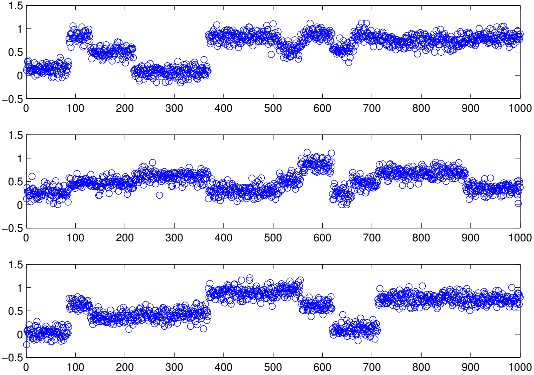
2 Finding multiple change-points in a single profile

3 Finding multiple change-points shared by many signals

4 Learning molecular classifiers with network information

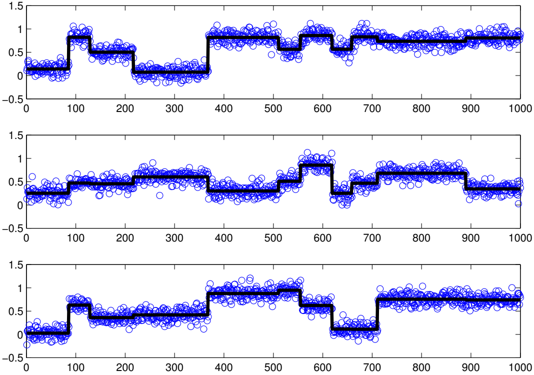
5 Conclusion

The problem



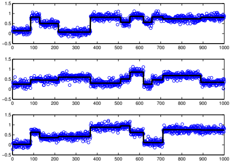
- Let $Y \in \mathbb{R}^{p \times n}$ the n signals of length p
- We want to find a piecewise constant approximation $\hat{U} \in \mathbb{R}^{p \times n}$ with at most k change-points.

The problem



- Let $Y \in \mathbb{R}^{p \times n}$ the n signals of length p
- We want to find a piecewise constant approximation $\hat{U} \in \mathbb{R}^{p \times n}$ with at most k change-points.

"Optimal" segmentation by dynamic programming



- Define the "optimal" piecewise constant approximation $\hat{U} \in \mathbb{R}^{p \times n}$ of Y as the solution of

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

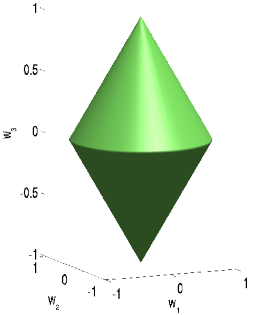
- DP finds the solution in $O(p^2 kn)$ in time and $O(p^2)$ in memory
- But: does not scale to $p = 10^6 \sim 10^9$...

Selecting pre-defined groups of variables

Group lasso (Yuan & Lin, 2006)

If groups of covariates are likely to be selected together, the ℓ_1/ℓ_2-norm induces sparse solutions *at the group level*:

$$\Omega_{group}(w) = \sum_g \|w_g\|_2$$



$$\begin{aligned}\Omega(w_1, w_2, w_3) &= \|(w_1, w_2)\|_2 + \|w_3\|_2 \\ &= \sqrt{w_1^2 + w_2^2} + \sqrt{w_3^2}\end{aligned}$$

TV approximator for many signals

- Replace

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

by

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

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:

$$\begin{aligned} \gamma &= U_{1,\bullet}\,, \\ \beta_{i,\bullet} &= w_i\left(U_{i+1,\bullet} - U_{i,\bullet}\right) \quad \text{for } i = 1, \ldots, p-1\,. \end{aligned}$$

- TV approximator is then equivalent to the following group Lasso problem (Yuan and Lin, 2006):

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

where $\bar{Y}$ is the centered signal matrix and $\bar{X}$ is a particular $(p-1) \times (p-1)$ design matrix.

TV approximator implementation

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

Theorem

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 $(p-1)\times(p-1)$:

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

Speed trial

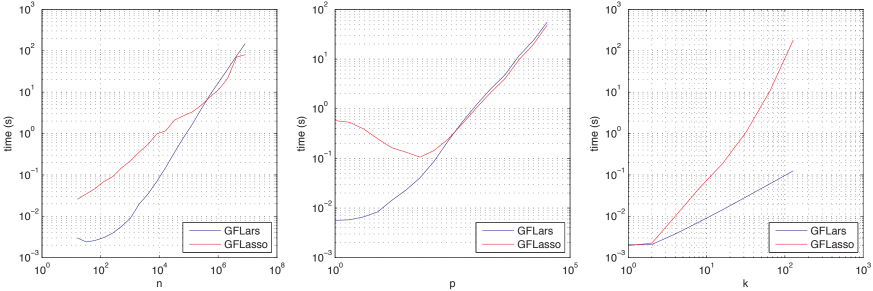
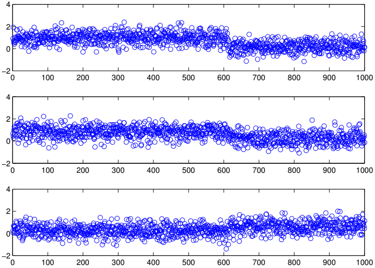


Figure 2: **Speed trials for group fused LARS (top row) and Lasso (bottom row).** *Left column:* varying n, with fixed $p = 10$ and $k = 10$; *center column:* varying p, with fixed $n = 1000$ and $k = 10$; *right column:* varying k, with fixed $n = 1000$ and $p = 10$. Figure axes are log-log. Results are averaged over 100 trials.

Consistency for a single change-point

Suppose a single change-point:

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



Does the TV approximator correctly estimate the first change-point as p increases?

Consistency of the unweighted TV approximator

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

Theorem

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

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

- correct estimation on $[p\epsilon, p(1-\epsilon)]$ with $\epsilon = \sqrt{\frac{\sigma^2}{2p\bar{\beta}^2}} + o(p^{-1/2})$.
- wrong estimation near the boundaries

Consistency of the weighted TV approximator

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

Theorem

The weighted TV approximator with weights

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

correctly finds the first change-point with probability tending to 1 *as* $n \to +\infty$.

- we see the benefit of increasing n
- we see the benefit of adding weights to the TV penalty

Proof sketch

- The first change-point $\hat{i}$ found by TV approximator maximizes $F_i = \| \hat{c}_{i,\bullet} \|^2$, 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 χ^2 distribution with

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

- We then just check when $G_u = \max_i G_i$

Consistency for a single change-point

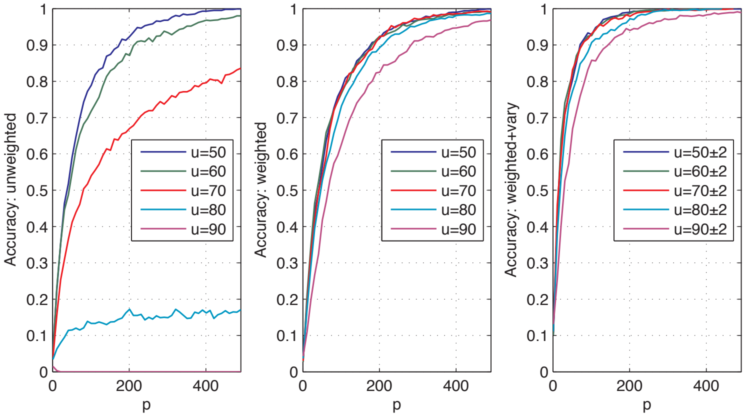


Figure 3: **Single change-point accuracy for the group fused Lasso.** Accuracy as a function of the number of profiles p when the change-point is placed in a variety of positions $u = 50$ to $u = 90$ (left and centre plots, resp. unweighted and weighted group fused Lasso), or: $u = 50 \pm 2$ to $u = 90 \pm 2$ (right plot, weighted with varying change-point location), for a signal of length 100.

Estimation of more change-points?

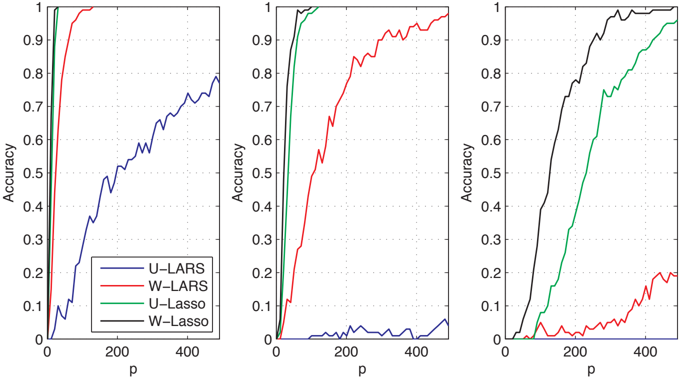
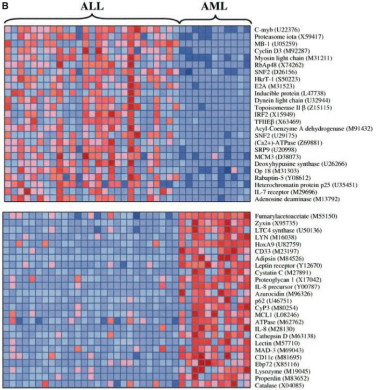


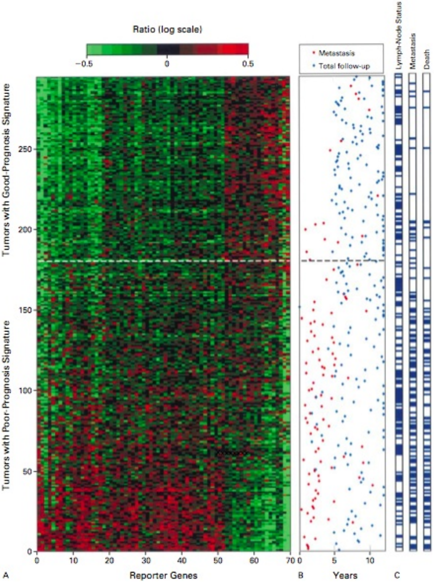
Figure 4: **Multiple change-point accuracy.** Accuracy as a function of the number of profiles p when change-points are placed at the nine positions $\{10, 20, \ldots, 90\}$ and the variance σ^2 of the centered Gaussian noise is either 0.05 (left), 0.2 (center) and 1 (right). The profile length is 100.

Outline

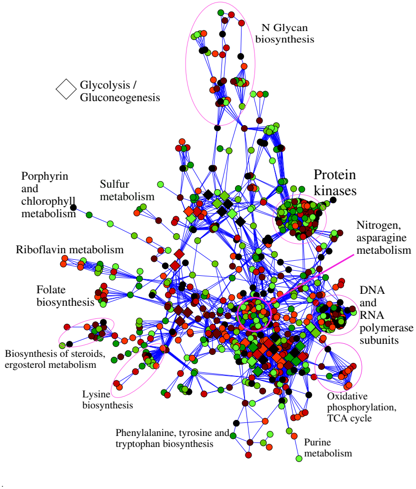
1. Motivations
2. Finding multiple change-points in a single profile
3. Finding multiple change-points shared by many signals
4. Learning molecular classifiers with network information
5. Conclusion

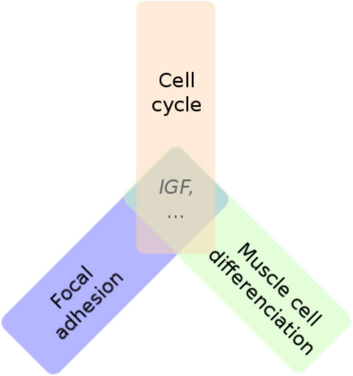
Molecular diagnosis / prognosis / theragnosis





Gene networks, gene groups


N Glycan biosynthesis
Glycolysis / Gluconeogenesis
Protein kinases
Porphyrin and chlorophyll metabolism
Sulfur metabolism
Nitrogen, asparagine metabolism
Riboflavin metabolism
Folate biosynthesis
DNA and RNA polymerase subunits
Biosynthesis of steroids, ergosterol metabolism
Lysine biosynthesis
Oxidative phosphorylation, TCA cycle
Phenylalanine, tyrosine and tryptophan biosynthesis
Purine metabolism


Cell cycle

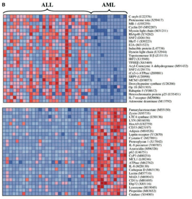
IGF, ...

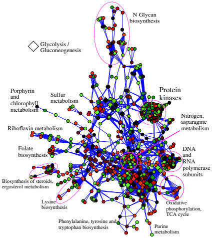
Focal adhesion

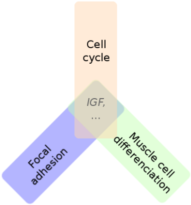
Muscle cell differenciation

Structured feature selection

- Basic biological functions usually involve the coordinated action of several proteins:
  - Formation of protein complexes
  - Activation of metabolic, signalling or regulatory pathways
- How to perform structured feature selection, such that selected genes
  - belong to only a few groups?
  - form a small number of connected components on the graph?




N Glycan biosynthesis
Glycolysis / Gluconeogenesis
Porphyrin and chlorophyll metabolism
Sulfur metabolism
Protein kinases
Nitrogen, asparagine metabolism
Riboflavin metabolism
Folate biosynthesis
DNA and RNA polymerase subunits
Biosynthesis of steroids, ergosterol metabolism
Lysine biosynthesis
Oxidative phosphorylation, TCA cycle
Phenylalanine, tyrosine and tryptophan biosynthesis
Purine metabolism

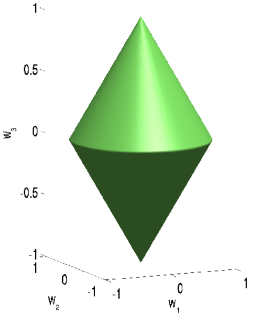


Selecting pre-defined groups of variables

Group lasso (Yuan & Lin, 2006)

If groups of covariates are likely to be selected together, the ℓ_1/ℓ_2-norm induces sparse solutions *at the group level*:

$$\Omega_{group}(w) = \sum_g \| w_g \|_2$$

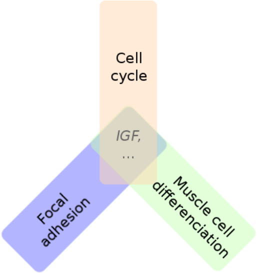


$$\Omega(w_1, w_2, w_3) = \|(w_1, w_2)\|_2 + \|w_3\|_2$$

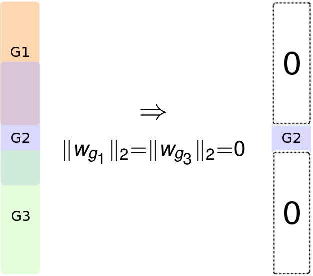
Group lasso with overlapping groups

Idea 1: shrink groups to zero (Jenatton et al., 2009)

- $\Omega_{group}(w) = \sum_g \|w_g\|_2$ sets groups to 0.
- One variable is selected ⇔ all the groups to which it belongs are selected.



IGF selection ⇒ selection of unwanted groups



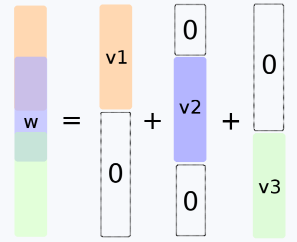
$\|w_{g_1}\|_2 = \|w_{g_3}\|_2 = 0$

Removal of *any* group containing a gene ⇒ the weight of the gene is 0.

Group lasso with overlapping groups

Idea 2: latent group Lasso (Jacob et al., 2009)

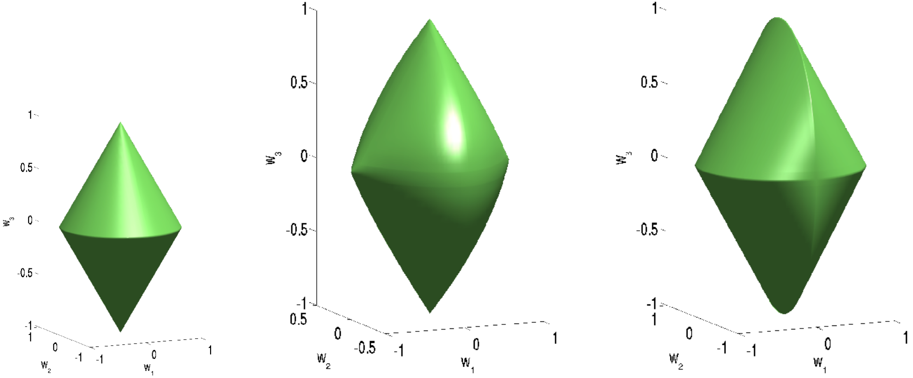
$$\Omega^{\mathcal{G}}_{\text{latent}}(w) \triangleq \begin{cases} \min_v \sum_{g\in\mathcal{G}} \|v_g\|_2 \\ w = \sum_{g\in\mathcal{G}} v_g \\ \operatorname{supp}(v_g) \subseteq g. \end{cases}$$



Properties

- Resulting support is a *union* of groups in $\mathcal{G}$.
- Possible to select one variable without selecting all the groups containing it.
- Equivalent to group lasso when there is no overlap

Overlap and group unity balls



Balls for $\Omega^{\mathcal{G}}_{\text{group}}(\cdot)$ (middle) and $\Omega^{\mathcal{G}}_{\text{latent}}(\cdot)$ (right) for the groups $\mathcal{G} = \{\{1,2\},\{2,3\}\}$ where w_2 is represented as the vertical coordinate. Left: group-lasso ($\mathcal{G} = \{\{1,2\},\{3\}\}$), for comparison.

Theoretical results

Consistency in group support (Jacob et al., 2009)

- Let $\bar{w}$ be the true parameter vector.
- Assume that there exists a unique decomposition $\bar{v}_g$ such that $\bar{w} = \sum_g \bar{v}_g$ and $\Omega^{\mathcal{G}}_{\text{latent}}(\bar{w}) = \sum \|\bar{v}_g\|_2$.
- Consider the regularized empirical risk minimization problem $L(w) + \lambda\Omega^{\mathcal{G}}_{\text{latent}}(w)$.

Then

- under appropriate mutual incoherence conditions on X,
- as $n \to \infty$,
- with very high probability,

the optimal solution $\hat{w}$ admits a unique decomposition $(\hat{v}_g)_{g\in\mathcal{G}}$ such that

$$\left\{g \in \mathcal{G} | \hat{v}_g \neq 0\right\} = \left\{g \in \mathcal{G} | \bar{v}_g \neq 0\right\}.$$

Theoretical results

Consistency in group support (Jacob et al., 2009)

- Let $\bar{w}$ be the true parameter vector.
- Assume that there exists a unique decomposition $\bar{v}_g$ such that $\bar{w} = \sum_g \bar{v}_g$ and $\Omega_{\text{latent}}^{\mathcal{G}}(\bar{w}) = \sum \|\bar{v}_g\|_2$.
- Consider the regularized empirical risk minimization problem $L(w) + \lambda \Omega_{\text{latent}}^{\mathcal{G}}(w)$.

Then

- under appropriate mutual incoherence conditions on X,
- as $n \to \infty$,
- with very high probability,

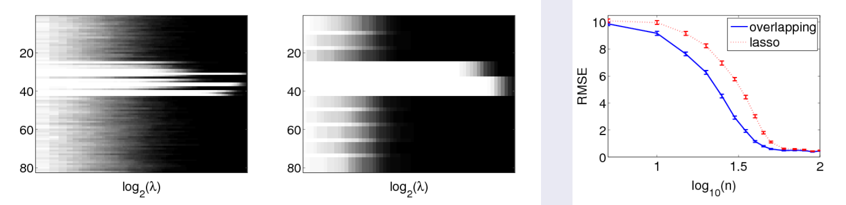
the optimal solution $\hat{w}$ admits a unique decomposition $(\hat{v}_g)_{g \in \mathcal{G}}$ such that

$$\left\{ g \in \mathcal{G} | \hat{v}_g \neq 0 \right\} = \left\{ g \in \mathcal{G} | \bar{v}_g \neq 0 \right\} .$$

Experiments

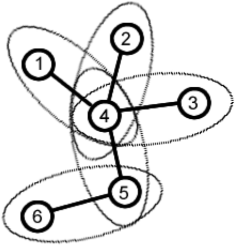
Synthetic data: overlapping groups

- 10 groups of 10 variables with 2 variables of overlap between two successive groups :$\{1,\ldots,10\},\{9,\ldots,18\},\ldots,\{73,\ldots,82\}$.
- Support: union of $4th$ and $5th$ groups.
- Learn from 100 training points.



Frequency of selection of each variable with the lasso (left) and $\Omega^{\mathcal{G}}_{\text{latent}}(.)$ (middle), comparison of the RMSE of both methods (right).

Graph lasso



Two solutions

$$\Omega^{\mathcal{G}}_{\text{group}}(\beta) = \sum_{i \sim j} \sqrt{\beta_i^2 + \beta_j^2} \,,$$

$$\Omega^{\mathcal{G}}_{\text{latent}}(\beta) = \sup_{\alpha \in \mathbb{R}^p : \forall i \sim j, \|\alpha_i^2 + \alpha_j^2\| \leq 1} \alpha^\top \beta \,.$$

Preliminary results

Breast cancer data

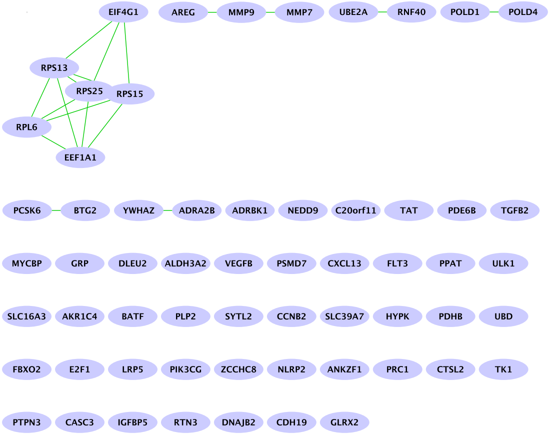
- Gene expression data for 8,141 genes in 295 breast cancer tumors.
- Canonical pathways from MSigDB containing 639 groups of genes, 637 of which involve genes from our study.

METHOD	ℓ_1	$\Omega^{\mathcal{G}}_{\text{LATENT}}(.)$
ERROR	0.38 ± 0.04	0.36 ± 0.03
MEAN ♯ PATH.	130	30

- Graph on the genes.

METHOD	ℓ_1	$\Omega_{graph}(.)$
ERROR	0.39 ± 0.04	0.36 ± 0.01
AV. SIZE C.C.	1.03	1.30

Lasso signature



EIF4G1 AREG MMP9 MMP7 UBE2A RNF40 POLD1 POLD4

RPS13 RPS25 RPS15 RPL6 EEF1A1

PCSK6 BTG2 YWHAZ ADRA2B ADRBK1 NEDD9 C20orf11 TAT PDE6B TGFB2

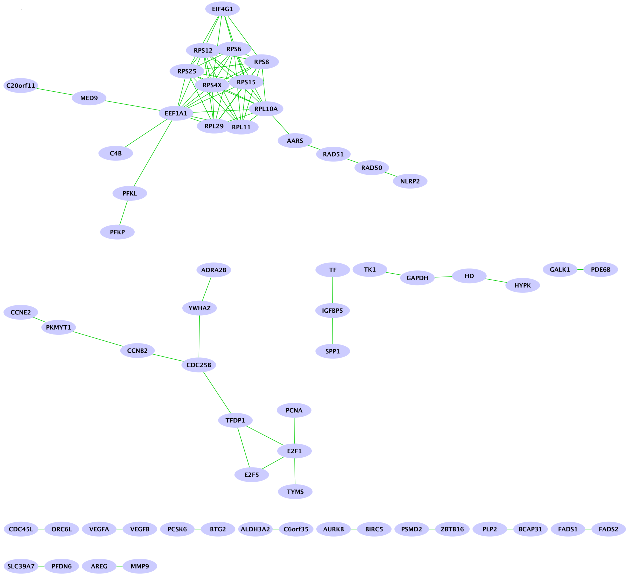
MYCBP GRP DLEU2 ALDH3A2 VEGFB PSMD7 CXCL13 FLT3 PPAT ULK1

SLC16A3 AKR1C4 BATF PLP2 SYTL2 CCNB2 SLC39A7 HYPK PDHB UBD

FBXO2 E2F1 LRP5 PIK3CG ZCCHC8 NLRP2 ANKZF1 PRC1 CTSL2 TK1

PTPN3 CASC3 IGFBP5 RTN3 DNAJB2 CDH19 GLRX2

Graph Lasso signature



Outline

1. Motivations
2. Finding multiple change-points in a single profile
3. Finding multiple change-points shared by many signals
4. Learning molecular classifiers with network information
5. Conclusion

Conclusions

- Penalty design as a way to incorporate prior knowledge
- Convex sparsity-inducing penalties are useful; efficient implementations + consistency results

Kevin Bleakley (INRIA), Laurent Jacob (UC Berkeley) Guillaume Obozinski (INRIA)

Post-docs available in Paris!