

RNA isoforms discovery from RNA-seq data

FlipFlop: Fast Lasso based Isoform Prediction as a FLOW Problem

Elsa Bernard¹, Laurent Jacob², Julien Mairal³, Jean-Philippe Vert¹

¹Mines ParisTech, Institut Curie, INSERM U900

²CNRS - LBBE Laboratory, Lyon

³Inria - LEAR, Grenoble

Persyvact workshop on probabilistic graphical models - July 2014



Team Players



Elsa
Bernard



Laurent
Jacob



Jean-Philippe
Vert

- E. Bernard, L. Jacob, J. Mairal, and J-P. Vert. Efficient RNA Isoform Identification and Quantification from RNA-Seq Data with Network Flows Bioinformatics. 2014.
- (J. Mairal and B. Yu. Supervised Feature Selection in Graphs with Path Coding Penalties. Journal of Machine Learning Research. 2013).

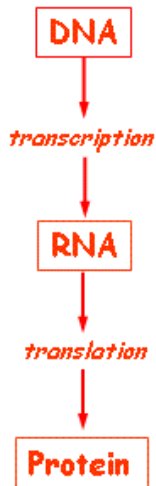
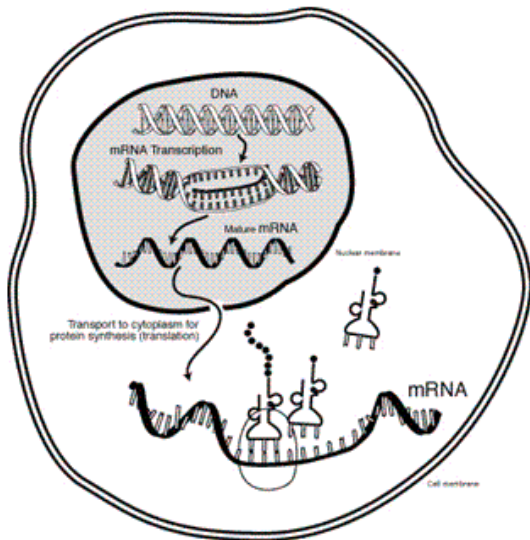
There will be a probabilistic model and a graph

\neq

Bayesian graphical model or

Markov Random Field

DNA Transcription/Translation (Central Dogma, 1958)



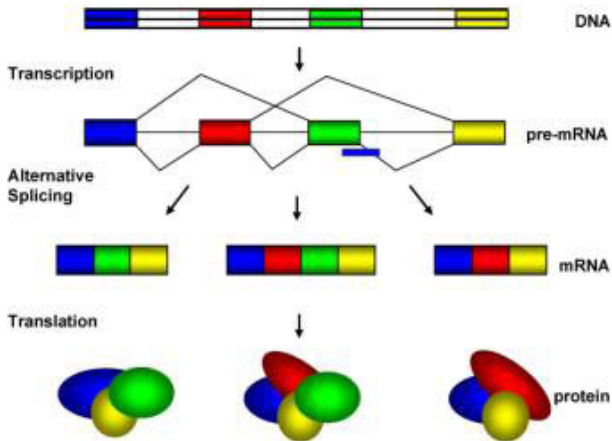
Modern Biology and Challenges



DOE Joint Genome institute

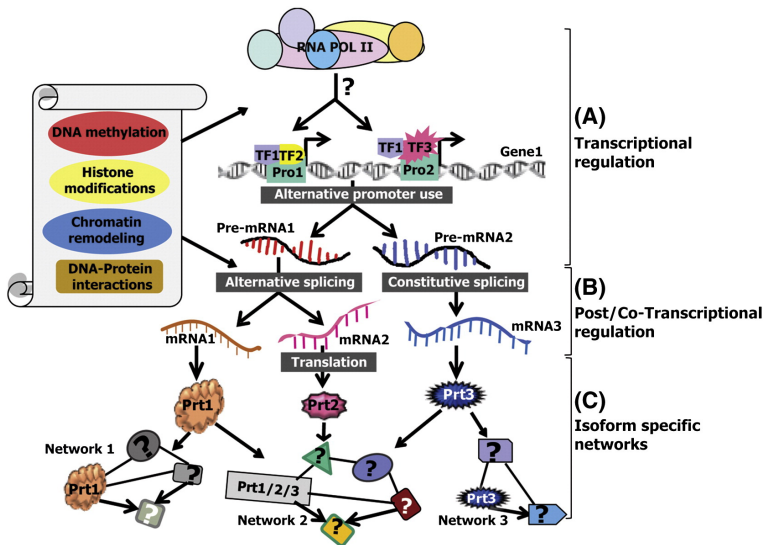
- biology is producing massive amount of data;
- sequencing one genome now costs about 1000\$ (vs 0.1 billion \$ in 2001), and produces about a few gigabytes of data;
- prediction from DNA data.

Alternative Splicing: 1 Gene = Many Proteins



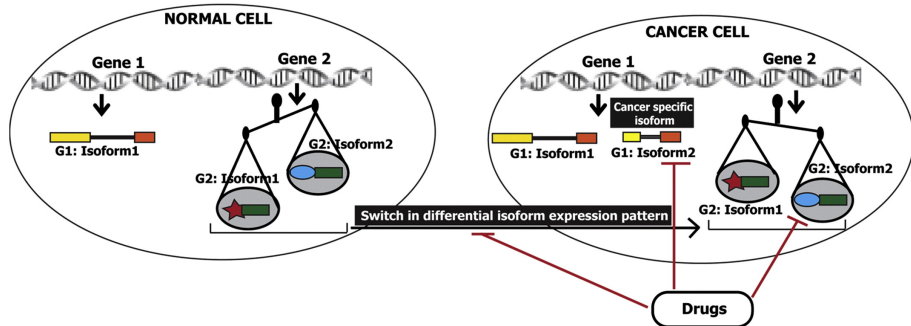
In human, 28k genes give 120k known transcripts (*Pal et al., 2012*)

Importance of Alternative Splicing



(Pal et al., 2012)

Opportunities for Drug Developments...

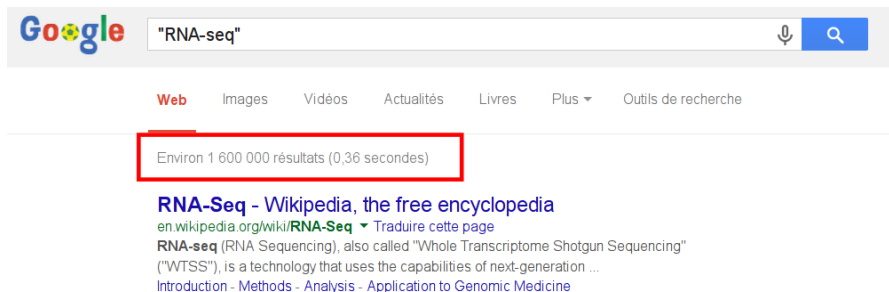


(Pal et al., 2012)



RNA-Seq or Next-Generation Sequencing

What is RNA-Seq?

- RNA-Seq measures abundance of RNA;



The image shows a Google search interface. The search bar contains the text "RNA-seq". Below the search bar, there are navigation tabs for "Web", "Images", "Vidéos", "Actualités", "Livres", "Plus", and "Outils de recherche". A red box highlights the search results summary: "Environ 1 600 000 résultats (0,36 secondes)". Below this, the first search result is for "RNA-Seq - Wikipedia, the free encyclopedia", with the URL "en.wikipedia.org/wiki/RNA-Seq" and a link to "Traduire cette page". The snippet for this result reads: "RNA-seq (RNA Sequencing), also called "Whole Transcriptome Shotgun Sequencing" ("WTSS"), is a technology that uses the capabilities of next-generation sequencing ... Introduction - Methods - Analysis - Application to Genomic Medicine".

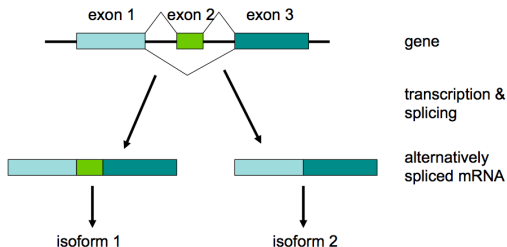
Google "RNA-seq"  

Web Images Vidéos Actualités Livres Plus ▾ Outils de recherche

Environ 1 600 000 résultats (0,36 secondes)

RNA-Seq - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/RNA-Seq ▾ Traduire cette page
RNA-seq (RNA Sequencing), also called "Whole Transcriptome Shotgun Sequencing" ("WTSS"), is a technology that uses the capabilities of next-generation sequencing ...
[Introduction](#) - [Methods](#) - [Analysis](#) - [Application to Genomic Medicine](#)

The Isoform Identification and Quantification Problem



Given a biological sample can we:

- 1 identify the isoform(s) of each gene present in the sample?
- 2 quantify their abundance?

From RNA-Seq Reads to Isoforms

**RNA sample
transcripts**



library preparation

**reads
50-200pb**



Transcripts Quantification using annotations

- RQuant (Bohner et al. 2009)
- FluxCapacitor (Montgomery et al. 2010)
- IsoEM (Nicolae et al. 2011)
- eXpress (Roberts et al. 2013)

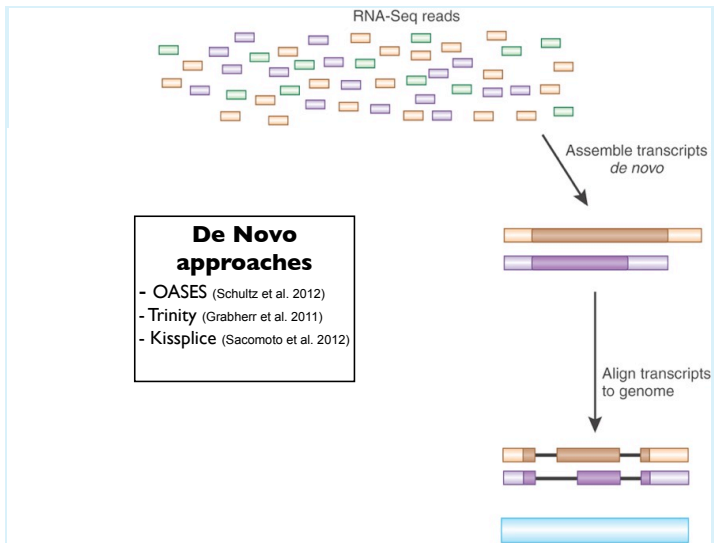
De Novo approaches

- Trinity (Grabherr et al. 2011)
- OASES (Schultz et al. 2012)
- Kissplice (Sacomoto et al. 2012)

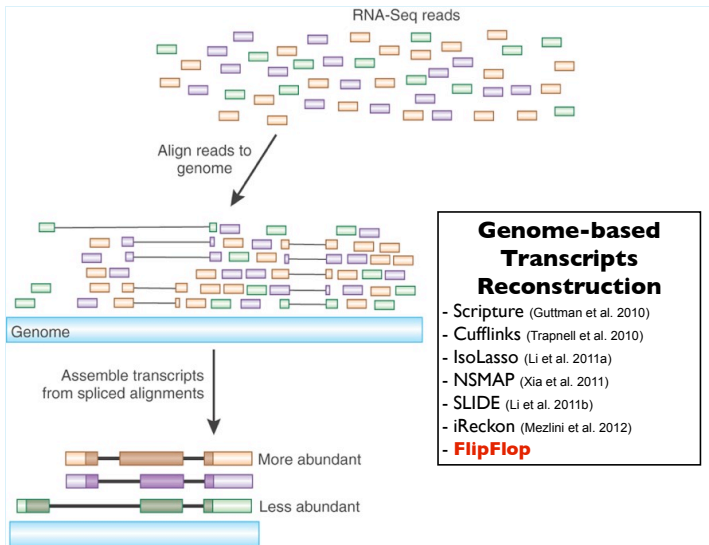
Genome-based Transcripts Reconstruction

- Scripture (Guttman et al. 2010)
- Cufflinks (Trapnell et al. 2010)
- IsoLasso (Li et al. 2011a)
- NSMAP (Xia et al. 2011)
- SLIDE (Li et al. 2011b)
- iReckon (Mezlini et al. 2012)
- MiTie (Behr et al. 2013)
- **FlipFlop**

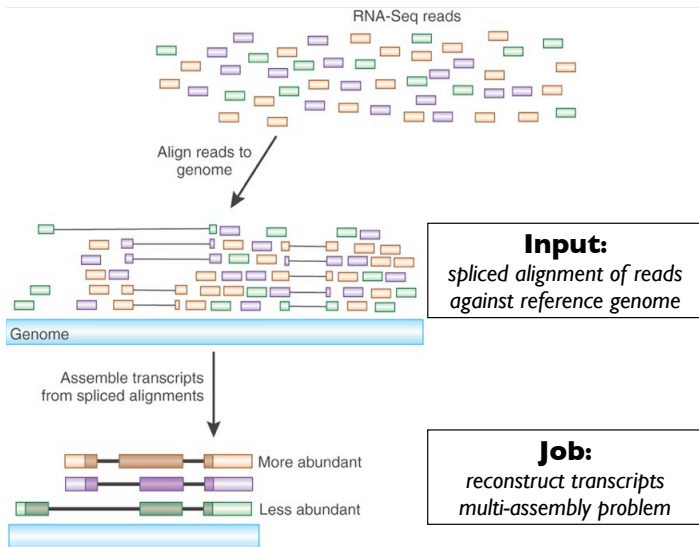
De Novo methods



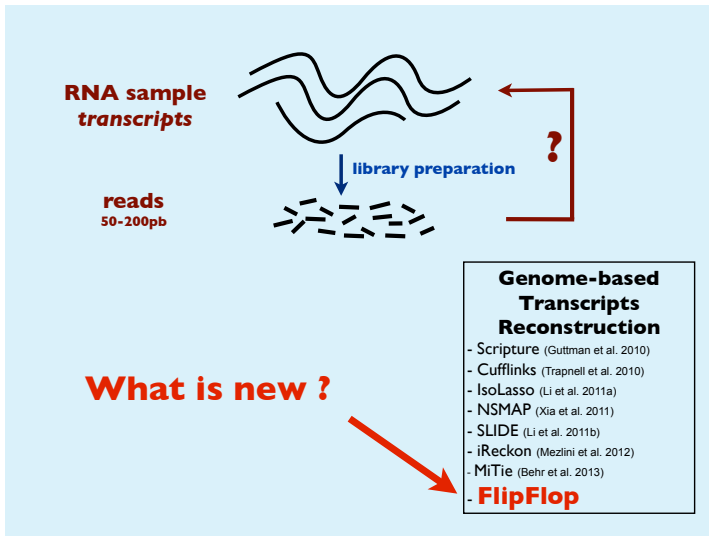
Genome-Based Methods



Genome-Based Isoforms Reconstruction



Place in the literature



Contributions

- **NO NEED** for **FILTERING** of candidate isoforms
- **FASTER** than existing methods that solve the same problem
- adapted to **LONG READS**
- **R package**

*flow
method*

Contributions

- **NO NEED** for **FILTERING** of candidate isoforms
 - **FASTER** than existing methods that solve the same problem
 - **adapted to LONG READS**
 - **R package**
- } *particular splicing graph*

Contributions

- **NO NEED** for **FILTERING** of candidate isoforms
- **FASTER** than existing methods that solve the same problem
- adapted to long reads
- **R package**

[Home](#)[Install](#)[Help](#)

[Home](#) » [Bioconductor 2.13](#) » [Software Packages](#) » [flipflop](#)

flipflop

Fast lasso-based isoform prediction as a flow problem

Bioconductor version: Release (2.13)

Flipflop discovers which isoforms of a gene are expressed in a given sample together with their abundances, based on RNA-Seq read data.

Author: Elsa Bernard, Laurent Jacob, Julien Mairal and Jean-Philippe Vert

Maintainer: Elsa Bernard <elsa.bernard@mines-paristech.fr>

To install this package, start R and enter:

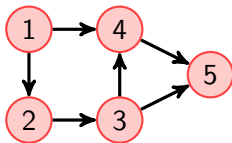
```
source("http://bioconductor.org/biocLite.R")
biocLite("flipflop")
```

Outline

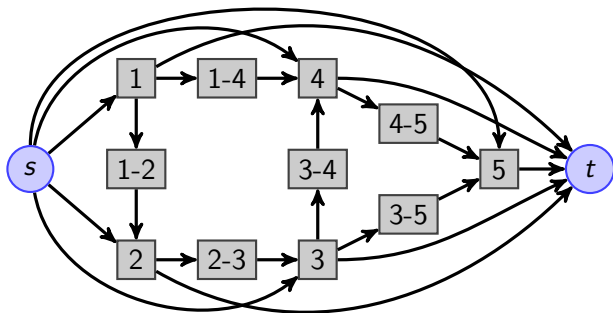
- 1 Formulation as a Path Selection Problem
- 2 Sparse Probabilistic Model and Optimization: FlipFlop
- 3 Results and Perspectives

Isoforms are Paths in a Graph

- Splicing graph for a gene with 5 exons:

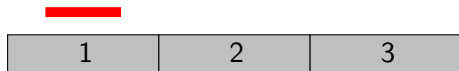
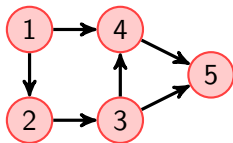


- FlipFlop graph: **1 type of read** \leftrightarrow **1 node**

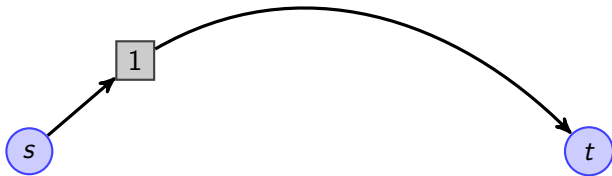


Graph adapted to long reads

- Splicing graph for a gene with 5 exons:

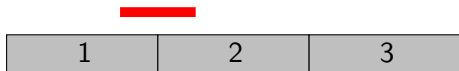
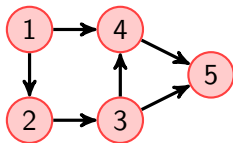


- FlipFlop graph:

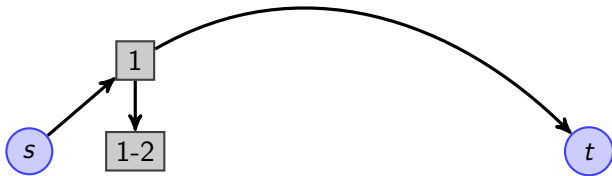


Graph adapted to long reads

- Splicing graph for a gene with 5 exons:

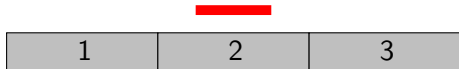
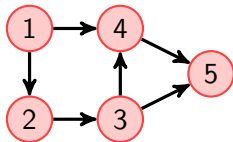


- FlipFlop graph:

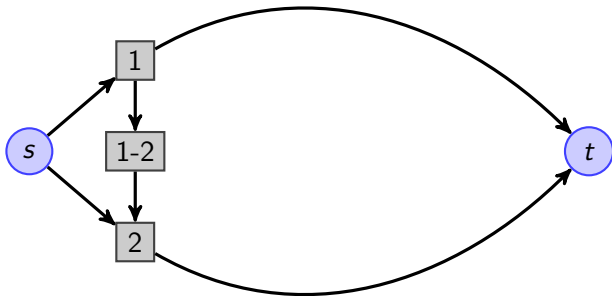


Graph adapted to long reads

- Splicing graph for a gene with 5 exons:

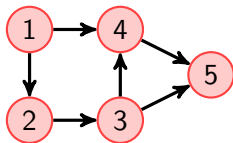


- FlipFlop graph:

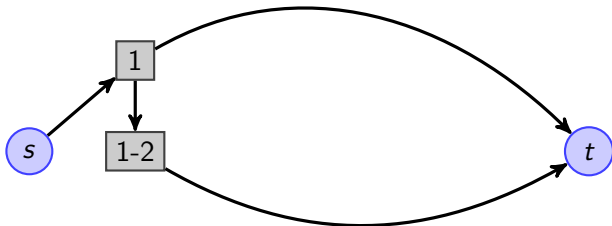


Graph adapted to long reads

- Splicing graph for a gene with 5 exons:

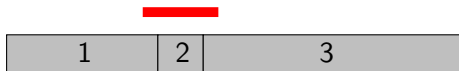
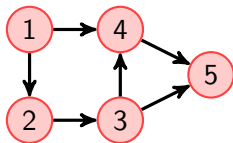


- FlipFlop graph:

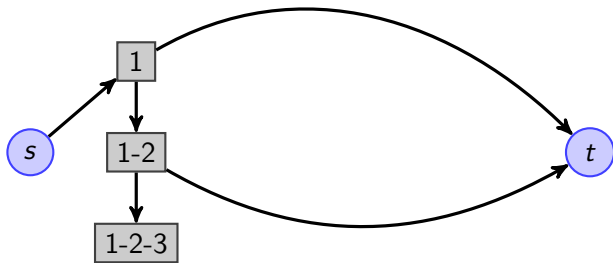


Graph adapted to long reads

- Splicing graph for a gene with 5 exons:

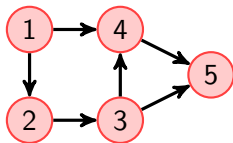


- FlipFlop graph:

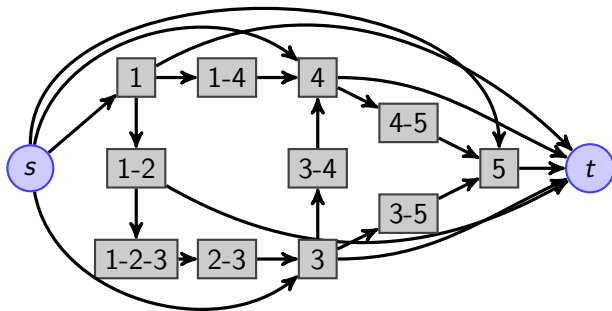


Graph adapted to long reads

- Splicing graph for a gene with 5 exons:

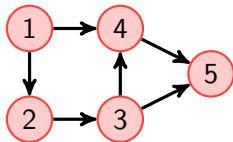


- FlipFlop graph:

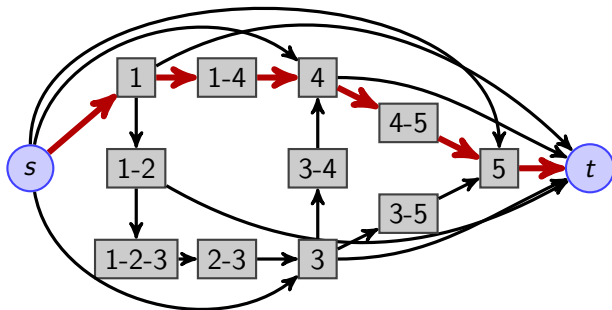


Graph adapted to long reads

- Splicing graph for a gene with 5 exons:

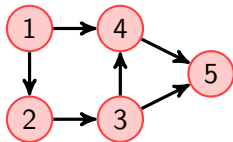


- FlipFlop graph: **one path with abundance β_1**

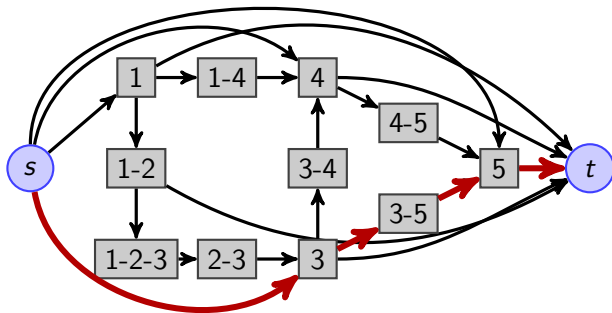


Graph adapted to long reads

- Splicing graph for a gene with 5 exons:



- FlipFlop graph: another path with abundance $\beta_2 \dots$



Select a Small Number of Paths?

n exons $\rightarrow \sim 2^n$ paths/candidate isoforms

feature selection problem with ~ 1000 candidates for 10 exons and ~ 1000000 for 20 exons

Minimal path cover

- Cufflinks

Regularization approach

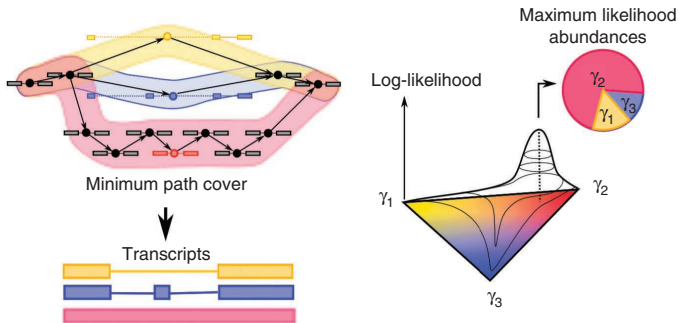
- IsoLasso, NSMAP, SLIDE, iReckon, MiTie, FlipFlop

Select a Small Number of Paths?

Cufflinks strategy

A two-step approach

- 1 find a set of *minimal paths* to explain read positions (independent from read counts)
- 2 estimate isoform abundances using read counts



Select a small number of paths?

Regularization approach

- 1 Suppose there are c candidate isoforms (c large)
- 2 Let β the unknown c -dimensional vector of abundance

Select a small number of paths?

Regularization approach

- 1 Suppose there are c candidate isoforms (c large)
- 2 Let β the unknown c -dimensional vector of abundance
- 3 Let $\mathcal{L}(\beta)$ quantify whether β explains the observed read counts
 - e.g., Poisson negative log-likelihood:

$$\mathcal{L}(\beta) = \sum_{\text{node } u} -\log p(X_u) \text{ with } X_u \sim \mathcal{P}(\delta_u) \text{ and } \delta_u \propto l_u \sum_{\text{path } p \ni u} \beta_p$$

Select a small number of paths?

Regularization approach

- 1 Suppose there are c candidate isoforms (c large)
- 2 Let β the unknown c -dimensional vector of abundance
- 3 Let $\mathcal{L}(\beta)$ quantify whether β explains the observed read counts
 - e.g., Poisson negative log-likelihood:

$$\mathcal{L}(\beta) = \sum_{\text{node } u} -\log p(X_u) \text{ with } X_u \sim \mathcal{P}(\delta_u) \text{ and } \delta_u \propto l_u \sum_{\text{path } p \ni u} \beta_p$$

- 4 Regularization-based approaches try to solve:

$$\min_{\beta \in \mathbb{R}_+^c} \mathcal{L}(\beta) \text{ such that } \beta \text{ is sparse}$$

Isoform Deconvolution with the ℓ_1 -norm

ℓ_1 -regularization

Estimate β sparse by solving:

$$\min_{\beta \in \mathbb{R}_+^c} \mathcal{L}(\beta) + \lambda \|\beta\|_1 ,$$

with \mathcal{L} a convex loss function.

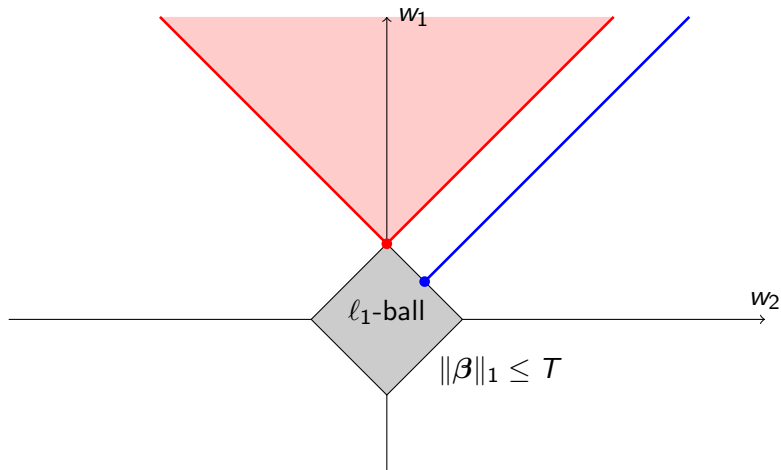
Computationally challenging:

- IsoLasso: strong filtering
- NSMAP, SLIDE: number of exons cut-off

FlipFlop: Fast Lasso-based Isoform Prediction as a FLOW Problem

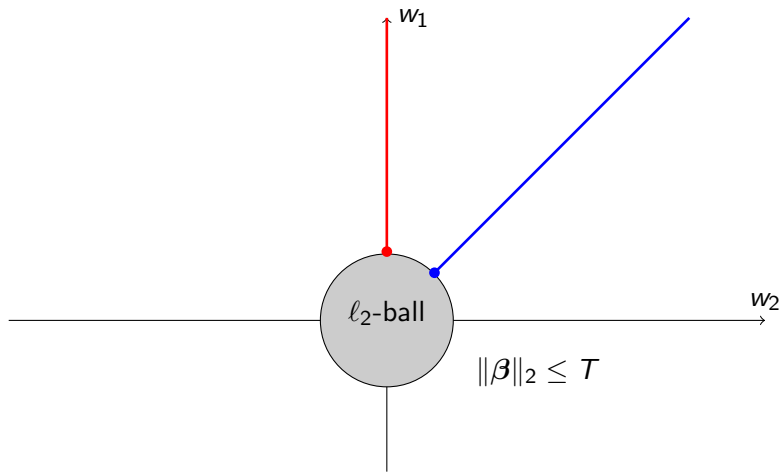
- no filtering
- no exons restrictions

Regularizing with the ℓ_1 -norm



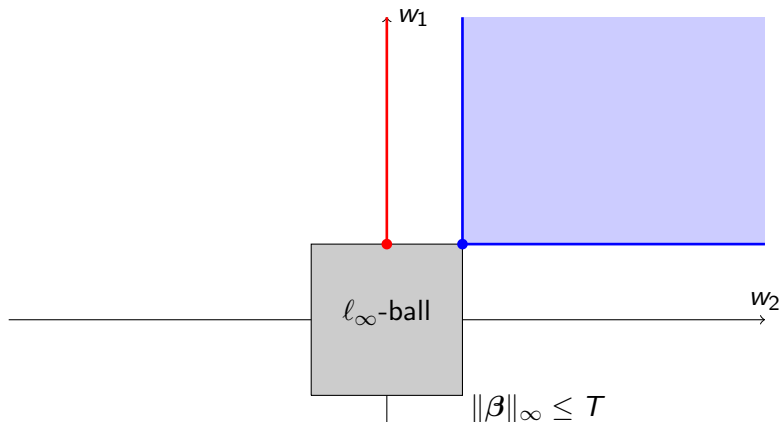
The projection onto a convex set is “biased” towards singularities.

Regularizing with the ℓ_2 -norm



The ℓ_2 -norm is isotropic.

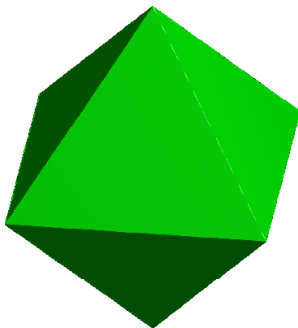
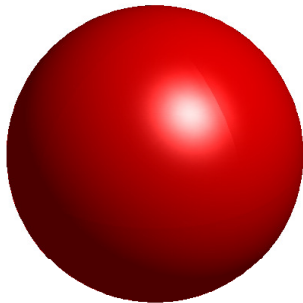
Regularizing with the ℓ_∞ -norm



The ℓ_∞ -norm encourages $|\beta_1| = |\beta_2|$.

In 3D.

Copyright G. Obozinski



Fast Isoform Deconvolution with the lasso

Theoretical (practical) result

The isoform deconvolution problem

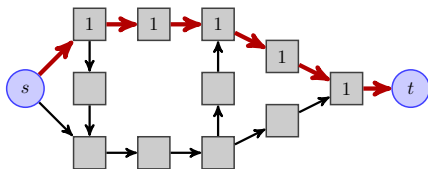
$$\min_{\beta \in \mathbb{R}_+^c} \mathcal{L}(\beta) + \lambda \|\beta\|_1 ,$$

can be solved in **polynomial time** with the number of nodes of the splicing graph.

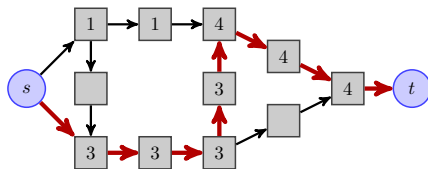
Ideas:

- 1 the sum of isoform abundances correspond to a **flow** on the graph
- 2 reformulation as a **convex cost flow problem** (Mairal and Yu, 2012)
- 3 recover isoforms by flow decomposition algorithm

Combinations of isoforms are flows



(a) Reads at every node corresponding to one isoform.

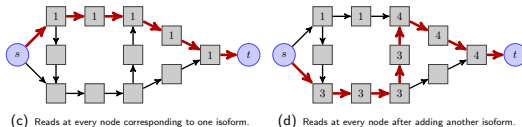


(b) Reads at every node after adding another isoform.

- **Linear combinations of isoforms** \Rightarrow **Flow value on every edges**
 - **Flow value on every edges** \Rightarrow **Paths with given value/abundance**
- Flow Decomposition
(linear time algorithm)**

Flux Capacitor. 2008. A Novel Min-Cost Flow Method for Estimating Transcript Expression with RNA-Seq. RECOMB-2013.

Equivalent flow problem (simpler!)



- For each edge sum abundances of isoforms that include the edge :

$$f_{uv} = \sum_{\text{path } p \ni (u,v)} \beta_p \quad \text{is a flow}$$

- Moreover

$$\|\beta\|_1 = \sum_{\text{path } p} \beta_p = f_t$$

- Therefore

$$\min_{\beta \in \mathbb{R}_+^c} \mathcal{L}(\beta) + \lambda \|\beta\|_1 \quad \text{is equivalent to} \quad \min_{f \text{ flow}} \tilde{\mathcal{L}}(f) + \lambda f_t$$

Technical details

Poisson Loss (with binary matrix \mathbf{U}):

$$\mathcal{L}(\mathbf{U}^T \boldsymbol{\beta}) = \sum_{u \in V} \left[Nl_u(\mathbf{U}^T \boldsymbol{\beta})_u - \mathbf{y}_u \log(Nl_u(\mathbf{U}^T \boldsymbol{\beta})_u) \right]$$

Flow Decomposition:

$$f_{uv} = \sum_{p \in \mathcal{P}'} \beta_p \mathbf{1}_{\{(u,v) \in p\}}$$
$$\Rightarrow f_v = \sum_{u \in V'} f_{uv} = (\mathbf{U}^T \boldsymbol{\beta})_v$$

Convex Cost Flow:

$$\min_{f_{\text{flow}}} \sum_{u \in V} [Nl_u f_u - \mathbf{y}_u \log(f_u)] + \lambda f_t$$

Solved using ε -relaxation method (Bertsekas 1998).

Summary

Isoform Detection=Path Selection Problem

$\sim 2^n$ variables (all paths in the splicing graph)



Equivalent Network Flow Problem

$\sim \frac{n^2}{2}$ variables (all exons and exon-exon junctions in the splicing graph)

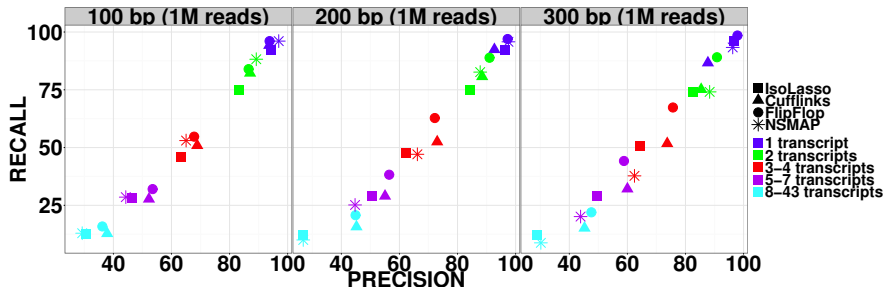


Network Flow Algorithms

Efficient Algorithms ! Polynomial Time.

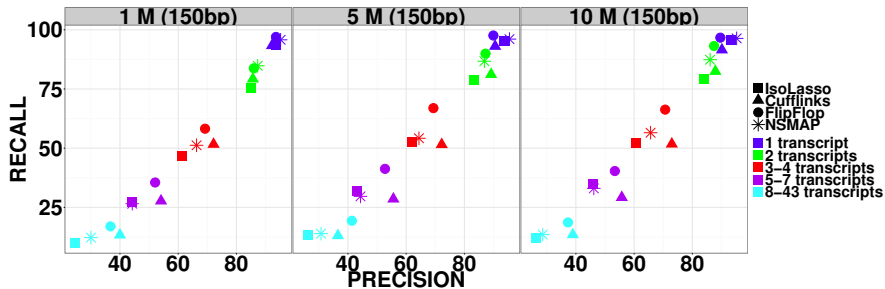
Performance increases with read length

- Human Simulation: hg19, 1137 genes on chr1, 1million reads by transcript levels.
- Simulator: <http://alumni.cs.ucr.edu/~liw/rnaseqreadsimulator.html>



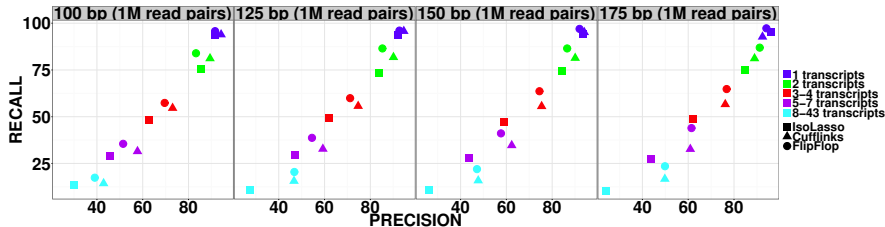
Performance increases with coverage

- Human Simulation: hg19, 1137 genes on chr1, 1million reads by transcript levels.
- Simulator: <http://alumni.cs.ucr.edu/~liw/rnaseqreadsimulator.html>



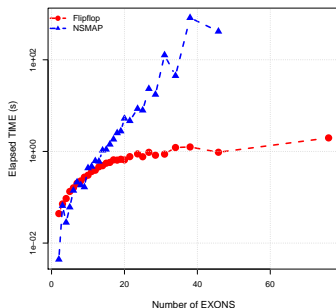
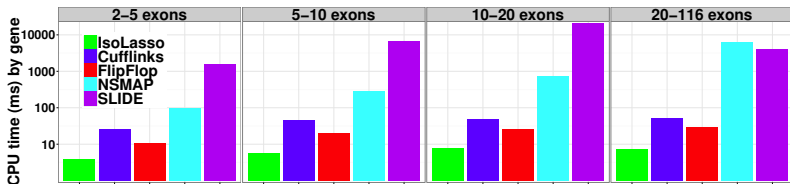
Extension to paired-end reads OK

- Human Simulation: hg19, 1137 genes on chr1, 1 million reads by transcript levels.
- Simulator: <http://alumni.cs.ucr.edu/~liw/rnaseqreadsimulator.html>



Speed Trial

- Human Simulation: hg19, 1137 genes on chr1, 1 million reads by transcript levels.
- Simulator: <http://alumni.cs.ucr.edu/~liw/rnaseqreadsimulator.html>

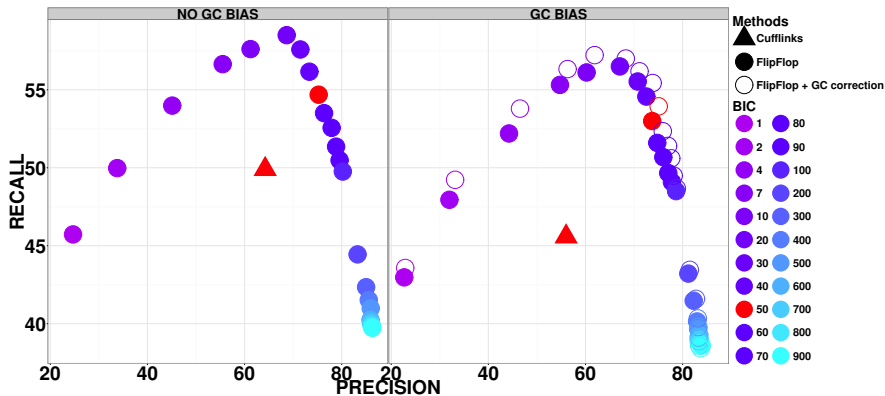


GC bias - Precision-Recall curve

● Human Simulation: hg19, chr1, 150bp single-end reads, 2 million, 4140 transcripts.

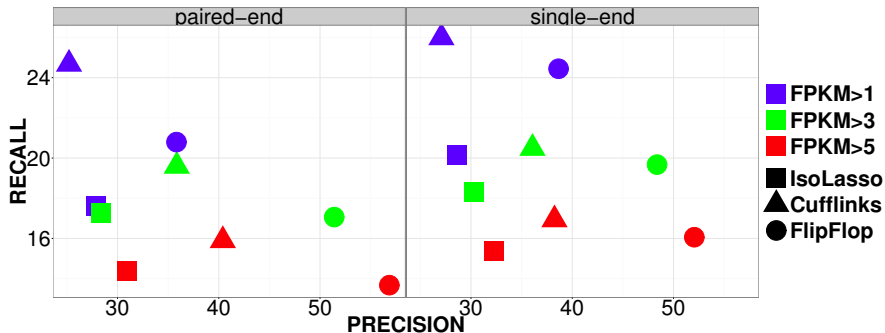
FluxSimulator, Griebel et al, 2012.

Model selection: set of solutions minimizing $\mathcal{L}(\beta) + \lambda \|\beta\|_1$ for different values of $\lambda \rightarrow$ BIC criteria



Real Data OK

● Human: 50 million 75bp paired-end reads.



FlipFlop → transcripts reconstruction over an exponential number of candidates in polynomial time

- ① **Hard combinatorial ill-posed** prediction problem !
- ② **Model Selection:** Cross Validation, Stability Selection?
- ③ **Multiple-samples:** on-going work with promising preliminary results.
- ④ **Differential Expression** testing at the isoform level ?

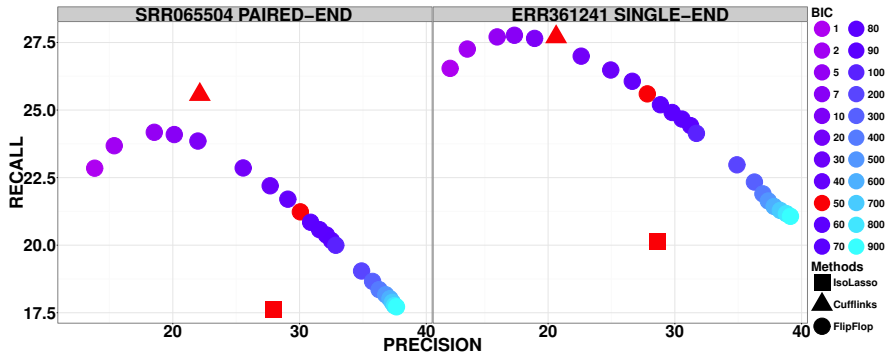
Conclusion/Discussion: get FlipFlop for free!



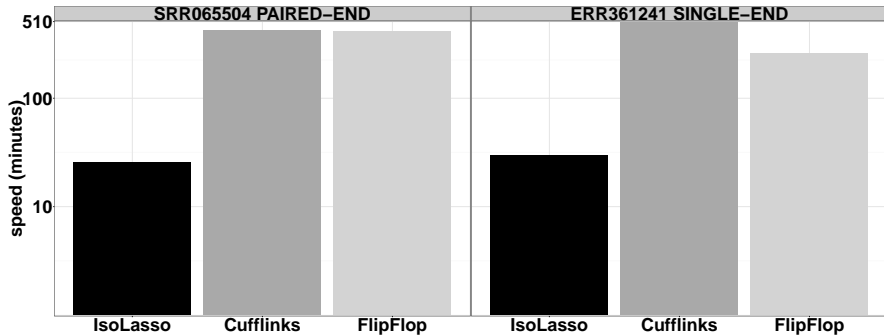
References

- <http://cbio.ensmp.fr/flipflop/>
- SParse Modelling Software SPAMS
<http://lear.inrialpes.fr/people/mairal/software.php>
- H. Jiang and W. H. Wong. *Bioinformatics*, 25(8):1026–1032, 2009.
- C. Trapnell et al. *Nat Biotechnol*, 28(5):511–515, 2010.
- Z. Xia et al. *BMC Bioinformatics*, 12:162, 2011.
- W. Li et al. *J Comput Biol*, 18:1693–1707, 2011.
- J.J. Li et al. *P Natl Acad Sci USA*, 108(50):19867–19872, 2011.
- R. K. Ahuja et al. *Prentice Hall*, 1993.
- D. P. Bertsekas. *Athena Scientific*, 1998.
- J. Mairal and B. Yu. *JMLR*, 2013.

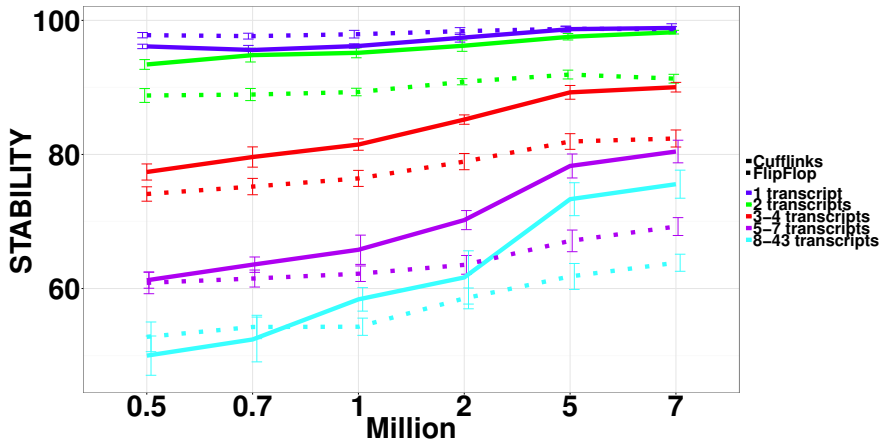
Precision-Recall curves on real data



Speed comparison on real data

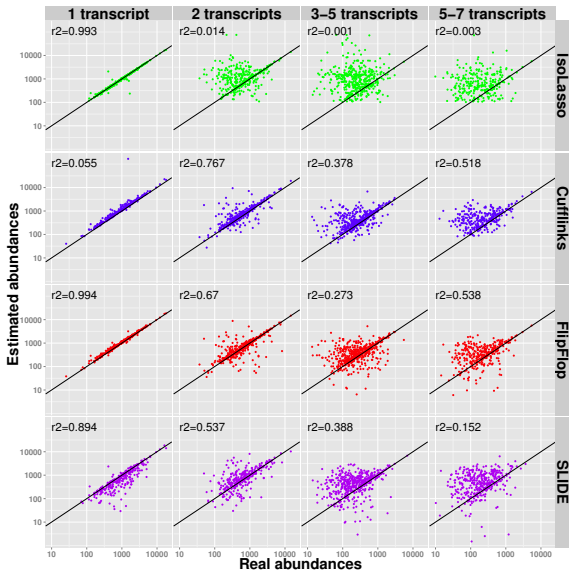


Stability study



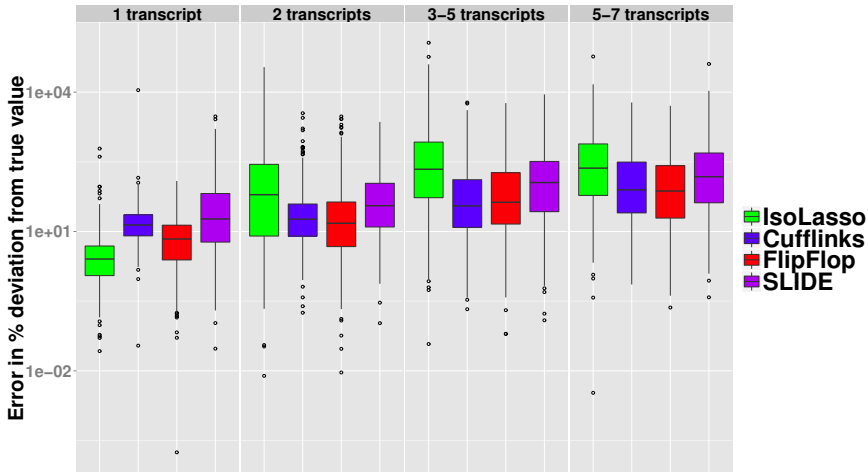
Human Simulation: Abundances

hg19, 1137 genes on chr1, 1million 75 bp single-end reads by transcript levels.



Simulation: Deviation

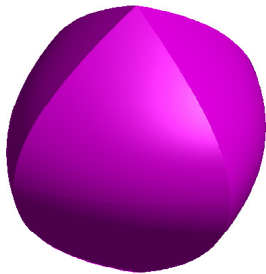
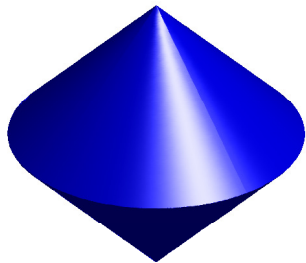
hg19, 1137 genes on chr1, 1million 75 bp single-end reads by transcript levels.



Part IV: Back to Structured Sparsity (depending on time)

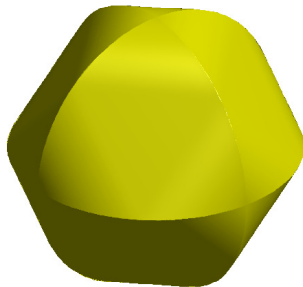
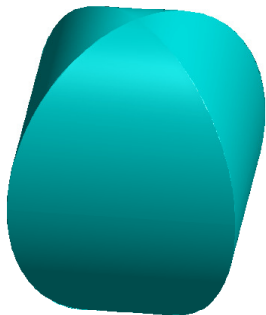
What about more complicated norms?

Copyright G. Obozinski



What about more complicated norms?

Copyright G. Obozinski

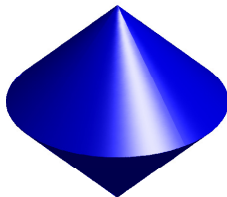


Group Lasso

[Turlach et al., 2005, Yuan and Lin, 2006]

the l_1/l_q -norm :
$$\Omega(\beta) = \sum_{g \in \mathcal{G}} \|\beta_g\|_q.$$

- \mathcal{G} is a **partition** of $\{1, \dots, p\}$;
- $q = 2$ or $q = \infty$ in practice;
- can be interpreted as the l_1 -norm of $[\|\beta_g\|_q]_{g \in \mathcal{G}}$.



$$\Omega(\beta) = \|\beta_{\{1,2\}}\|_2 + |\beta_3|.$$

Structured sparsity with overlapping groups

Warning: Under the name “structured sparsity” appear in fact significantly different formulations!

① non-convex

- zero-tree wavelets [Shapiro, 1993];
- predefined collection of sparsity patterns: [Baraniuk et al., 2010];
- **select a union of groups: [Huang et al., 2009];**
- structure via Markov Random Fields: [Cehver et al., 2008];

② convex (norms)

- **tree-structure: [Zhao et al., 2009];**
- **select a union of groups: [Jacob et al., 2009];**
- **zero-pattern is a union of groups: [Jenatton et al., 2009];**
- other norms: [Micchelli et al., 2010].

Group Lasso with overlapping groups

[Jenatton, Audibert, and Bach, 2009]

$$\Omega(\beta) = \sum_{g \in \mathcal{G}} \|\beta_g\|_q.$$

What happens when the groups overlap?

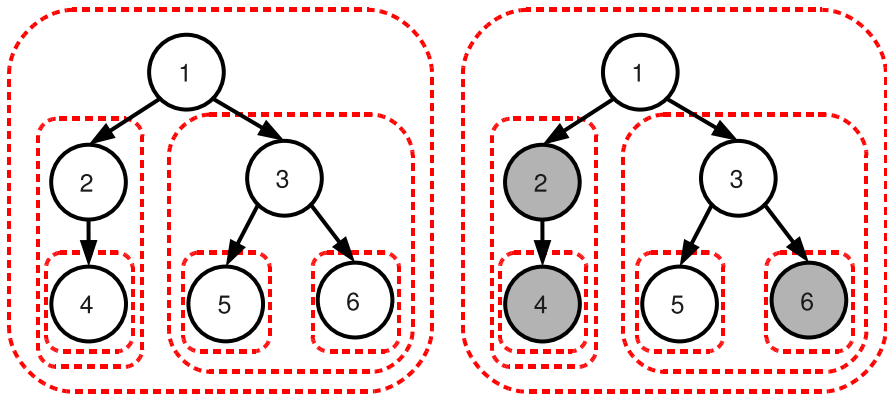
- the pattern of non-zero variables is an intersection of groups;
- the **zero pattern is a union of groups.**



$$\Omega(\beta) = \|\beta\|_2 + |\beta_2| + |\beta_3|.$$

Hierarchical Norms

[Zhao, Rocha, and Yu, 2009]



A node can be active only if its **ancestors are active**.
The selected patterns are **rooted subtrees**.

Modelling Patterns as Unions of Groups

the non-convex penalty of Huang, Zhang, and Metaxas [2009]

Warning: different point of view than the two previous slides

$$\varphi(\beta) \triangleq \min_{\mathcal{J} \subseteq \mathcal{G}} \left\{ \sum_{g \in \mathcal{J}} \eta_g \text{ s.t. } \text{Supp}(\beta) \subseteq \bigcup_{g \in \mathcal{J}} g \right\}.$$

- the penalty is **non-convex**.
- is **NP-hard** to compute (set cover problem).
- The pattern of non-zeroes in β is a **union** of (a few) groups.

It can be rewritten as a boolean linear program:

$$\varphi(\beta) = \min_{\mathbf{x} \in \{0,1\}^{|\mathcal{G}|}} \left\{ \boldsymbol{\eta}^\top \mathbf{x} \text{ s.t. } \mathbf{N}\mathbf{x} \geq \text{Supp}(\beta) \right\}.$$

Modelling Patterns as Unions of Groups

convex relaxation and the penalty of Jacob, Obozinski, and Vert [2009]

The penalty of Huang et al. [2009]:

$$\varphi(\beta) = \min_{\mathbf{x} \in \{0,1\}^{|\mathcal{G}|}} \left\{ \boldsymbol{\eta}^\top \mathbf{x} \text{ s.t. } \mathbf{N}\mathbf{x} \geq \text{Supp}(\beta) \right\}.$$

A convex LP-relaxation:

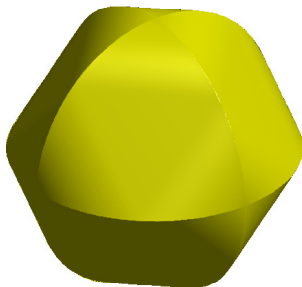
$$\psi(\beta) \triangleq \min_{\mathbf{x} \in \mathbb{R}_+^{|\mathcal{G}|}} \left\{ \boldsymbol{\eta}^\top \mathbf{x} \text{ s.t. } \mathbf{N}\mathbf{x} \geq |\beta| \right\}.$$

Lemma: ψ is the penalty of Jacob et al. [2009] with the ℓ_∞ -norm:

$$\psi(\beta) = \min_{(\xi^g \in \mathbb{R}^p)_{g \in \mathcal{G}}} \sum_{g \in \mathcal{G}} \eta_g \|\xi^g\|_\infty \text{ s.t. } \beta = \sum_{g \in \mathcal{G}} \xi^g \text{ and } \forall g, \text{Supp}(\xi^g) \subseteq g,$$

Modelling Patterns as Unions of Groups

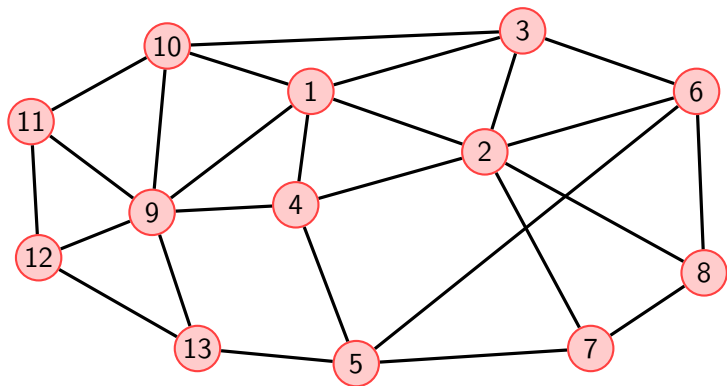
The norm of Jacob et al. [2009] in 3D



$\psi(\beta)$ with $\mathcal{G} = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$.

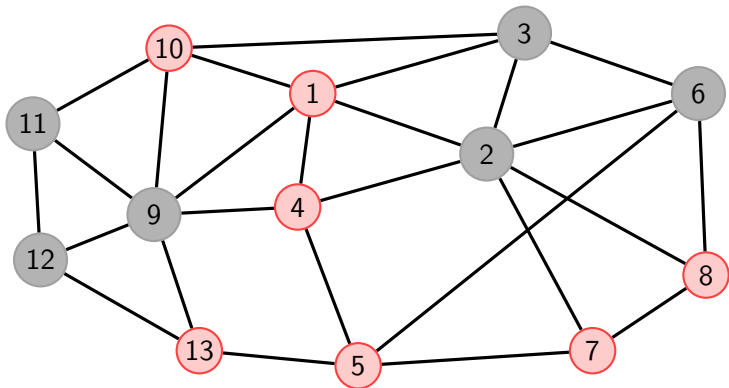
Graph sparsity

$G = (V, E)$, with $V = \{1, \dots, p\}$



Graph sparsity

Encouraging patterns with a small number of connected components



Formulation

$$\min_{\beta \in \mathbb{R}^p} \underbrace{R(\beta)}_{\text{convex, smooth}} + \underbrace{\lambda \Omega(\beta)}_{\text{regularization}},$$

Ω should encourage connected patterns in the graph.

- the penalty of Huang et al. [2009]:

$$\varphi(\beta) = \min_{\mathbf{x} \in \{0,1\}^{|\mathcal{G}|}} \left\{ \boldsymbol{\eta}^\top \mathbf{x} \text{ s.t. } \mathbf{N}\mathbf{x} \geq \text{Supp}(\beta) \right\}.$$

- a convex LP-relaxation (penalty of Jacob et al. [2009]):

$$\psi(\beta) \stackrel{\Delta}{=} \min_{\mathbf{x} \in \mathbb{R}_+^{|\mathcal{G}|}} \left\{ \boldsymbol{\eta}^\top \mathbf{x} \text{ s.t. } \mathbf{N}\mathbf{x} \geq |\beta| \right\}.$$

Structured sparsity for graphs

Group structure for graphs.

Natural choices to encourage connectivity in the graph is to define \mathcal{G} as

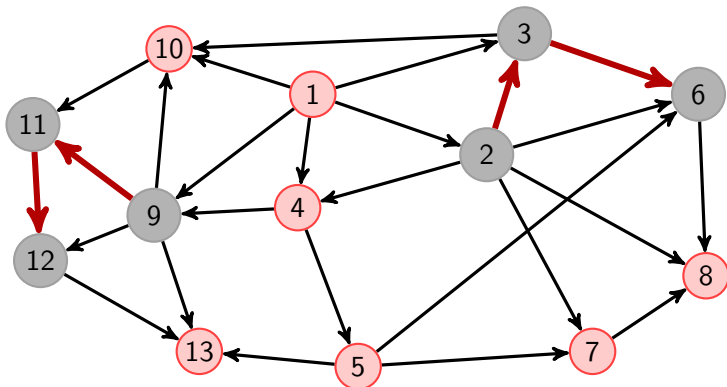
- 1 pairs of vertices linked by an arc. **only models local interactions;**
- 2 all connected subgraphs up to a size L . **cumbersome/intractable;**
- 3 all connected subgraphs. **intractable.**

Question

Can we replace connected subgraphs by another structure which (i) is rich enough to model long-range interactions in the graph, and (ii) leads to computationally feasible penalties?

A solution when the graph is a DAG (Mairal and Yu, 2012)

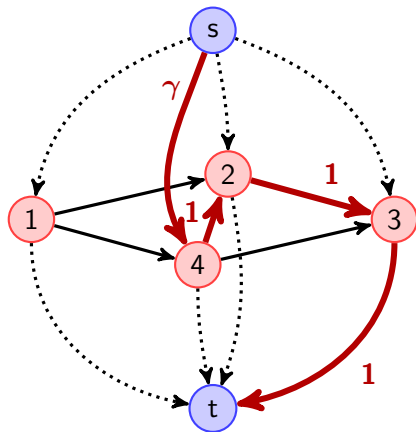
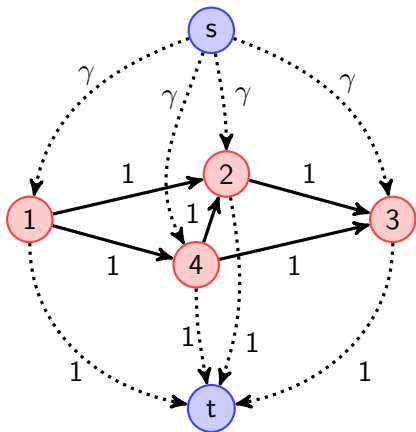
- 1 Define \mathcal{G} to be the **set of all paths in the DAG**.
- 2 Define η_g to be $\gamma + |g|$ (the cost of selecting a path g).



$$\varphi(\beta) = (\gamma + 3) + (\gamma + 3)$$

Graph sparsity for DAGs

Decomposability of the weights $\eta_g = \gamma + |g|$



Equivalence to network flows

An optimization problem on paths might be transformed into an equivalent flow problem.

Proposition 1

$$\varphi(\beta) = \min_{f \in \mathcal{F}} \sum_{(u,v) \in E'} f_{uv} c_{uv} \quad \text{s.t.} \quad s_j(f) \geq 1, \quad \forall j \in \text{Supp}(\beta),$$

Proposition 2

$$\psi(\beta) = \min_{f \in \mathcal{F}} \sum_{(u,v) \in E'} f_{uv} c_{uv} \quad \text{s.t.} \quad s_j(f) \geq |\beta_j|, \quad \forall j \in \{1, \dots, p\},$$

$\varphi(\beta)$, $\psi(\beta)$ and similarly the proximal operators, the dual norm of ψ **can be computed in polynomial time** using network flow optimization.

References I

- R. G. Baraniuk, V. Cevher, M. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 2010. to appear.
- V. Cehver, M. F. Duarte, C. Hegde, and R. G. Baraniuk. Sparse signal recovery using markov random fields. In *Advances in Neural Information Processing Systems*, 2008.
- J. Huang, Z. Zhang, and D. Metaxas. Learning with structured sparsity. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2009.
- L. Jacob, G. Obozinski, and J.-P. Vert. Group Lasso with overlap and graph Lasso. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2009.
- R. Jenatton, J-Y. Audibert, and F. Bach. Structured variable selection with sparsity-inducing norms. Technical report, 2009. preprint arXiv:0904.3523v1.

References II

- J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, January 2008.
- C. A. Micchelli, J. M. Morales, and M. Pontil. A family of penalty functions for structured sparsity. In *Advances in Neural Information Processing Systems*, 2010.
- J.M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462, 1993.
- B. A. Turlach, W. N. Venables, and S. J. Wright. Simultaneous variable selection. *Technometrics*, 47(3):349–363, 2005.
- M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B*, 68:49–67, 2006.

References III

- P. Zhao, G. Rocha, and B. Yu. The composite absolute penalties family for grouped and hierarchical variable selection. 37(6A):3468–3497, 2009.