Complexity Analysis of the Lasso Regularization Path

Isoform Discovery from RNA-Seq Data

Julien Mairal, Inria, Grenoble

Statslab seminar, Cambridge, 2015
Collaborators

Bin Yu  Elsa Bernard  Laurent Jacob  Jean-Philippe Vert

First part: a curiosity


Second part: a useful application of sparsity

Part I: Complexity Analysis of the Lasso Regularization Path

joint work with Bin Yu from UC Berkeley

Early thoughts about parsimony

(a) Dorothy Wrinch 1894–1980
(b) Harold Jeffreys 1891–1989

The existence of simple laws is, then, apparently, to be regarded as a quality of nature; and accordingly we may infer that it is justifiable to prefer a simple law to a more complex one that fits our observations slightly better.

[Wrinch and Jeffreys, 1921]. Philosophical Magazine Series.
Historical overview of parsimony

- 14th century: Ockham’s razor;
- 1921: Wrinch and Jeffreys’ simplicity principle;
- 1952: Markowitz’s portfolio selection;
- 60 and 70’s: best subset selection in statistics;
- 70’s: use of the $\ell_1$-norm for signal recovery in geophysics;
- 90’s: wavelet thresholding in signal processing;
- 1996: Olshausen and Field’s dictionary learning;
- 1996–1999: Lasso (statistics) and basis pursuit (signal processing);
- 2006: compressed sensing (signal processing) and Lasso consistency (statistics);
What this work is about

- another paper about the Lasso/Basis Pursuit [Tibshirani, 1996, Chen et al., 1999]:

\[
\min_{w \in \mathbb{R}^p} \frac{1}{2} \| y - Xw \|_2^2 + \lambda \| w \|_1; \tag{1}
\]

- the first complexity analysis of the homotopy method [Ritter, 1962, Osborne et al., 2000, Efron et al., 2004] for solving (1);

A story similar to

- the simplex algorithm for linear programs [Klee and Minty, 1972];
- the SVM regularization path [Gärtner, Jaggi, and Maria, 2010].
Regularizing with the $\ell_1$-norm

The projection onto a convex set is “biased” towards singularities.
Regularizing with the $\ell_2$-norm

$\|w\|_2 \leq T$

The $\ell_2$-norm is isotropic.
The Lasso Regularization Path and the Homotopy

Under uniqueness assumption of the Lasso solution, the regularization path is piecewise linear:
Our Main Results

Theorem - worst case analysis

*In the worst-case, the regularization path of the Lasso has exactly $(3^p + 1)/2$ linear segments.*

Proposition - approximate analysis

*There exists an $\varepsilon$-approximate path with $O(1/\sqrt{\varepsilon})$ linear segments.*
Piecewise linearity

Under uniqueness assumptions of the Lasso solution, the regularization path $\lambda \mapsto w^*(\lambda)$ is continuous and piecewise linear.
Brief Introduction to the Homotopy Algorithm

Piecewise linearity

Under uniqueness assumptions of the Lasso solution, the regularization path $\lambda \mapsto w^*(\lambda)$ is continuous and piecewise linear.

Recipe of the homotopy method - main ideas

1. finds a trivial solution $w^*(\lambda_\infty) = 0$ with $\lambda_\infty = \|X^T y\|_\infty$;
2. compute the direction of the current linear segment of the path;
3. follow the direction of the path by decreasing $\lambda$;
4. stop at the next “kink” and go back to 2.
Brief Introduction to the Homotopy Algorithm

Piecewise linearity

Under uniqueness assumptions of the Lasso solution, the regularization path \( \lambda \mapsto \mathbf{w}^*(\lambda) \) is continuous and piecewise linear.

Recipe of the homotopy method - main ideas

1. finds a trivial solution \( \mathbf{w}^*(\lambda_\infty) = 0 \) with \( \lambda_\infty = \|\mathbf{X}^\top \mathbf{y}\|_\infty \);
2. compute the direction of the current linear segment of the path;
3. follow the direction of the path by decreasing \( \lambda \);
4. stop at the next “kink” and go back to 2.

Caveats

- kinks can be very close to each other;
- the direction of the path can involve ill-conditioned matrices;
- worst-case exponential complexity (main result of this work).
Theorem - worst case analysis

In the worst-case, the regularization path of the Lasso has exactly \((3^p + 1)/2\) linear segments.
Worst case analysis

Consider a Lasso problem \((y \in \mathbb{R}^n, \mathbf{X} \in \mathbb{R}^{n \times p})\).

Define the vector \(\tilde{y}\) in \(\mathbb{R}^{n+1}\) and the matrix \(\tilde{\mathbf{X}}\) in \(\mathbb{R}^{(n+1) \times (p+1)}\) as follows:

\[
\tilde{y} \triangleq \begin{bmatrix} y \\ y_{n+1} \end{bmatrix}, \quad \tilde{\mathbf{X}} \triangleq \begin{bmatrix} \mathbf{X} & 2\alpha \mathbf{y} \\ 0 & \alpha y_{n+1} \end{bmatrix},
\]

where \(y_{n+1} \neq 0\) and \(0 < \alpha < \lambda_1/(2\mathbf{y}^\top \mathbf{y} + y_{n+1}^2)\).

Adverserial strategy

If the regularization path of the Lasso \((y,\mathbf{X})\) has \(k\) linear segments, the path of \((\tilde{y},\tilde{\mathbf{X}})\) has \(3k - 1\) linear segments.
Worst case analysis

\[ \tilde{y} \triangleq \begin{bmatrix} y \\ y_{n+1} \end{bmatrix}, \quad \tilde{X} \triangleq \begin{bmatrix} X & 2\alpha y \\ 0 & \alpha y_{n+1} \end{bmatrix}, \]

Let us denote by \( \{\eta^1, \ldots, \eta^k\} \) the sequence of \( k \) sparsity patterns in \( \{-1, 0, 1\}^p \) encountered along the path of the Lasso \((y, X)\).

The new sequence of sparsity patterns for \((\tilde{y}, \tilde{X})\) is

\[
\begin{cases}
\text{first } k \text{ patterns} \\
\begin{bmatrix} \eta^1 = 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \eta^2 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} \eta^k \\ 0 \end{bmatrix}, \\
\text{middle } k \text{ patterns} \\
\begin{bmatrix} \eta^k \\ 1 \end{bmatrix}, \begin{bmatrix} \eta^{k-1} \\ 1 \end{bmatrix}, \ldots, \begin{bmatrix} \eta^1 = 0 \\ 1 \end{bmatrix}, \\
\text{last } k-1 \text{ patterns} \\
\begin{bmatrix} -\eta^2 \\ 1 \end{bmatrix}, \begin{bmatrix} -\eta^3 \\ 1 \end{bmatrix}, \ldots, \begin{bmatrix} -\eta^k \\ 1 \end{bmatrix}
\end{cases}
\]
Worst case analysis

We are now in shape to build a pathological path with \((3^p + 1)/2\) linear segments. Note that this lower-bound complexity is tight.

\[
\begin{align*}
\mathbf{y} & \triangleq \begin{bmatrix}
1 \\
1 \\
1 \\
\vdots \\
1
\end{bmatrix}, & \mathbf{x} & \triangleq \begin{bmatrix}
\alpha_1 & 2\alpha_2 & 2\alpha_3 & \ldots & 2\alpha_p \\
0 & \alpha_2 & 2\alpha_3 & \ldots & 2\alpha_p \\
0 & 0 & \alpha_3 & \ldots & 2\alpha_p \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & \alpha_p
\end{bmatrix},
\end{align*}
\]
Approximate Complexity
Refinement of Giesen, Jaggi, and Laue [2010] for the Lasso

Strong Duality

Strong duality means that $\max_\kappa g(\kappa) = \min_w f(w)$
Strong duality means that \( \max_{\kappa} g(\kappa) = \min_{\mathbf{w}} f(\mathbf{w}) \)

The duality gap guarantees us that \( 0 \leq f(\tilde{\mathbf{w}}) - f(\mathbf{w}^*) \leq \delta(\tilde{\mathbf{w}}, \tilde{\kappa}) \).
Approximate Complexity

\[
\min \limits_w \left\{ f_\lambda(w) \triangleq \frac{1}{2} \| y - Xw \|^2_2 + \lambda \| w \|_1 \right\}, \quad \text{(primal)}
\]

\[
\max \limits_\kappa \left\{ g_\lambda(\kappa) \triangleq -\frac{1}{2} \kappa^T \kappa - \kappa^T y \quad \text{s.t.} \quad \| X^T \kappa \|_\infty \leq \lambda \right\}. \quad \text{(dual)}
\]

\(\varepsilon\)-approximate solution

\(w\) satisfies \(\text{APPROX}_\lambda(\varepsilon)\) when there exists a dual variable \(\kappa\) s.t.

\[
\delta_\lambda(w, \kappa) = f_\lambda(w) - g_\lambda(\kappa) \leq \varepsilon f_\lambda(w).
\]

\(\varepsilon\)-approximate path

A path \(\mathcal{P} : \lambda \mapsto w(\lambda)\) is an approximate path if it always contains \(\varepsilon\)-approximate solutions.

(see Giesen et al. [2010] for generic results on approximate paths)
Approximate Complexity

Main relation

\[ \text{APPROX}_\lambda(0) \iff \text{APPROX}_{\lambda(1 - \sqrt{\varepsilon})}(\varepsilon) \]

Key: find an appropriate dual variable \( \kappa(w) \) + simple calculation;

Proposition - approximate analysis

there exists an \( \varepsilon \)-approximate path with at most \( \left\lceil \frac{\log(\lambda_{\infty}/\lambda_1)}{\sqrt{\varepsilon}} \right\rceil \) segments.

Approximate homotopy - main ideas

- Maintain approximate optimality conditions along the path;
- Make steps in \( \lambda \) greater than or equal to \( \lambda(1 - \theta\sqrt{\varepsilon}) \);
- When the kinks are too close to each other, make a large step and switch to first-order method;
A Few Messages to Conclude

- Despite its exponential complexity, the homotopy algorithm remains extremely powerful in practice;
- numerical stability is still an issue of the homotopy algorithm;
- when one does not care about precision, the worst-case complexity of the path can be significantly reduced.
Part II: Isoform Discovery from RNA-Seq Data with Network Flows

joint work with Elsa Bernard (Institut Curie), Laurent Jacob (CNRS) and Jean-Philippe Vert (Institut Curie)

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;

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...
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

**Transcripts Quantification using annotations**
- RQuant (Böhmert 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)
- Kisssplice (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 (Mezini et al. 2012)
- MiTe (Behr et al. 2013)
- FlipFlop
De Novo methods

De Novo approaches
- OASES (Schultz et al. 2012)
- Trinity (Grabherr et al. 2011)
- Kissplice (Sacomoto et al. 2012)
Genome-Based Methods

RNA-Seq reads

Align reads to genome

Genome

Assemble transcripts from spliced alignments

More abundant

Less abundant

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)
- FlipFlop
Genome-Based Isoforms Reconstruction

Input:
spliced alignment of reads against reference genome

Job:
reconstruct transcripts multi-assembly problem
Place in the literature

RNA sample

transcripts

reads
50-200pb

library preparation

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 (Mezitini et al. 2012)
- MiTe (Behr et al. 2013)
- FlipFlop

What is new?
Contributions

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

Bioconductor

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 at mines-paristech.fr>

To install this package, start R and enter:

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

1 → 4 → 5
2 → 3

- FlipFlop graph:

s → 1 → t
Graph adapted to long reads

- Splicing graph for a gene with 5 exons:

  ![Splicing graph diagram]

- FlipFlop graph:

  ![FlipFlop graph diagram]
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 $\beta_1$
Graph adapted to long reads

- Splicing graph for a gene with 5 exons:
  
  ![Splicing Graph](image)

- FlipFlop graph: another path with abundance $\beta_2$ ...
Select a Small Number of Paths?

\[ n \text{ exons} \rightarrow \sim 2^n \text{ paths/candidate isoforms} \]

feature selection problem with \( \sim 1000 \) candidates for 10 exons and \( \sim 1000000 \) 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 \textit{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 \( \beta \) 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 \( \beta \) the unknown \( c \)-dimensional vector of abundance
3. Let \( \mathcal{L}(\beta) \) quantify whether \( \beta \) explains the observed read counts
   - e.g., Poisson negative log-likelihood:
     \[
     \mathcal{L}(\beta) = \sum_{\text{node } u} -\log p(X_u) \quad \text{with} \quad X_u \sim \mathcal{P}(\delta_u) \quad \text{and} \quad \delta_u \propto l_u \sum_{\text{path } p \ni u} \beta_p
     \]
Select a small number of paths?

Regularization approach

1. Suppose there are \( c \) candidate isoforms (\( c \) large).
2. Let \( \beta \) the unknown \( c \)-dimensional vector of abundance.
3. Let \( L(\beta) \) quantify whether \( \beta \) explains the observed read counts.
   - e.g., Poisson negative log-likelihood:
     \[
     L(\beta) = \sum_{\text{node } u} - \log p(X_u) \quad \text{with} \quad X_u \sim \mathcal{P}(\delta_u) \quad \text{and} \quad \delta_u \propto l_u \sum_{\text{path } p \ni u} \beta_p
     \]
4. Regularization-based approaches try to solve:
   \[
   \min_{\beta \in \mathbb{R}^{c}_+} L(\beta) \quad \text{such that} \quad \beta \text{ is sparse}
   \]
Isoform Deconvolution with the $\ell_1$-norm

$\ell_1$-regularization

Estimate $\beta$ sparse by solving:

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

with $\mathcal{L}$ a convex loss function.

**Computationally challenging:**

$\rightarrow$ IsoLasso: strong filtering

$\rightarrow$ NSMAP, SLIDE: number of exons cut-off

**FlipFlop:** Fast Lasso-based Isoform Prediction as a FLOw Problem

$\rightarrow$ no filtering

$\rightarrow$ no exons restrictions
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

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

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

- Linear combinations of isoforms $\Rightarrow$ Flow value on every edges
- Flow value on every edges $\Rightarrow$ Flow Decomposition (linear time algorithm)

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 $\mathbf{U}$):

$$
\mathcal{L}(\mathbf{U}^T \mathbf{\beta}) = \sum_{u \in V} \left[ \mathcal{N}_{lu}(\mathbf{U}^T \mathbf{\beta})_u - y_u \log(\mathcal{N}_{lu}(\mathbf{U}^T \mathbf{\beta})_u) \right]
$$

Flow Decomposition:

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

Convex Cost Flow:

$$
\min_{f_{\text{flow}}} \sum_{u \in V} \left[ \mathcal{N}_{lu} f_u - y_u \log(f_u) \right] + \lambda f_t
$$

Solved using $\varepsilon$-relaxation method (Bertsekas 1998).
Summary

Isoform Detection = Path Selection Problem
\sim 2^n \text{ variables (all paths in the splicing graph)}

\iff 

Equivalent Network Flow Problem
\sim \frac{n^2}{2} \text{ variables (all exons and exon-exon junctions in the splicing graph)}

\downarrow

Network Flow Algorithms
Efficient Algorithms ! Polynomial Time.
Performance increases with read length

- Human Simulation: hg19, 1137 genes on chr1, 1 million reads by transcript levels.
Performance increases with coverage

- Human Simulation: hg19, 1137 genes on chr1, 1 million reads by transcript levels.

![Graph showing performance increases with coverage]

<table>
<thead>
<tr>
<th>Coverage</th>
<th>1 M (150bp)</th>
<th>5 M (150bp)</th>
<th>10 M (150bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>1 transcript</td>
<td>2 transcripts</td>
<td>3–4 transcripts</td>
</tr>
<tr>
<td></td>
<td>PRECISION</td>
<td>PRECISION</td>
<td>PRECISION</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>80</td>
</tr>
<tr>
<td>Isolasso</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Cufflinks</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>FlipFlop</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>NSMAP</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Julien Mairal, Inria

Two talks related to sparsity
Extension to paired-end reads

- Human Simulation: hg19, 1137 genes on chr1, 1 million reads by transcript levels.

![Graph showing precision and recall for different transcript numbers and read lengths]
Speed Trial

- Human Simulation: hg19, 1137 genes on chr1, 1 million reads by transcript levels.

<table>
<thead>
<tr>
<th>Exon Range</th>
<th>CPU Time (ms) by gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>2–5 exons</td>
<td></td>
</tr>
<tr>
<td>5–10 exons</td>
<td></td>
</tr>
<tr>
<td>10–20 exons</td>
<td></td>
</tr>
<tr>
<td>20–116 exons</td>
<td></td>
</tr>
</tbody>
</table>

- IsoLasso
- Cufflinks
- FlipFlop
- NSMAP
- SLIDE

Two talks related to sparsity
Human Simulation: hg19, chr1, 150bp single-end reads, 2 million, 4140 transcripts.


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

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

<table>
<thead>
<tr>
<th>PRECISION</th>
<th>RECALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>30</td>
<td>40</td>
</tr>
</tbody>
</table>

Legend:
- FPKM>1
- FPKM>3
- FPKM>5
- IsoLasso
- Cufflinks
- FlipFlop

Two talks related to sparsity
Conclusion/Discussion

FlipFlop $\rightarrow$ 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!

C++ interfaced with Matlab, R, Python.

proximal gradient methods for $\ell_0$, $\ell_1$, elastic-net, fused-Lasso, group-Lasso, tree group-Lasso, tree-$\ell_0$, sparse group Lasso, overlapping group Lasso...

...for square, logistic, multi-class logistic loss functions.

handles sparse matrices, provides duality gaps.

fast implementations of OMP and LARS - homotopy.

dictionary learning and matrix factorization (NMF, sparse PCA).

coordinate descent, block coordinate descent algorithms.

fast projections onto some convex sets.

Try it! http://www.di.ens.fr/willow/SPAMS/
References

- http://cbio.ensmp.fr/flipflop/
- SParse Modelling Software SPAMS
  http://lear.inrialpes.fr/people/mairal/software.php
Precision-Recall curves on real data

SRR065504 PAIRED-END

ERR361241 SINGLE-END

Methods
- IsoLasso
- Cufflinks
- FlipFlop
Speed comparison on real data

![Bar chart showing speed comparison between different tools for SRR065504 PAIRED-END and ERR361241 SINGLE-END datasets. The tools compared are IsoLasso, Cufflinks, and FlipFlop. The chart displays the speed in minutes, with IsoLasso generally outperforming Cufflinks and FlipFlop.]
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, 1 million 75 bp single-end reads by transcript levels.

<table>
<thead>
<tr>
<th>1 transcript</th>
<th>2 transcripts</th>
<th>3−5 transcripts</th>
<th>5−7 transcripts</th>
</tr>
</thead>
</table>

Error in % deviation from true value

- IsoLasso
- Cufflinks
- FlipFlop
- SLIDE

Julien Mairal, Inria

Two talks related to sparsity


References II


Worst case analysis - Backup Slide

\[ \tilde{y} \triangleq \begin{bmatrix} y \\ y_{n+1} \end{bmatrix}, \quad \tilde{X} \triangleq \begin{bmatrix} X & 2\alpha y \\ 0 & \alpha y_{n+1} \end{bmatrix}, \]

Some intuition about the adverserial strategy:

1. the patterns of the new path must be \([\eta^T, 0]^T\) or \([\pm \eta^T, 1]^T\);
2. the factor \(\alpha\) ensures the \((p + 1)\)-th variable to enter late the path;
3. after the \(k\) first kinks, we have \(y \approx Xw^*(\lambda)\) and thus

\[ \tilde{X} \begin{bmatrix} w^*(\lambda) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ y_{n+1} \end{bmatrix} \approx \tilde{y} \approx \tilde{X} \begin{bmatrix} -w^*(\lambda) \\ 1/\alpha \end{bmatrix}. \]
Worst case analysis - Backup Slide 2

\[
\min_{\tilde{w} \in \mathbb{R}^p, \tilde{w} \in \mathbb{R}^1} \frac{1}{2} \|\tilde{y} - \tilde{X} \begin{bmatrix} \tilde{w} \\ \tilde{w} \end{bmatrix} \|^2_2 + \lambda \| \begin{bmatrix} \tilde{w} \\ \tilde{w} \end{bmatrix} \|_1 = \frac{1}{2} \|(1 - 2\alpha \tilde{w})y - X\tilde{w} \|^2_2 + \frac{1}{2}(y_{n+1} - \alpha y_{n+1} \tilde{w})^2 + \lambda \|\tilde{w}\|_1 + \lambda |\tilde{w}|.
\]

is equivalent to

\[
\min_{\tilde{w}' \in \mathbb{R}^p} \frac{1}{2} \|y - X\tilde{w}'\|^2_2 + \frac{\lambda}{|1 - 2\alpha \tilde{w}^*|} \|\tilde{w}'\|_1,
\]

and then

\[
\tilde{w}^* = \begin{cases} 
(1 - 2\alpha \tilde{w}^*)w^* \left( \frac{\lambda}{|1 - 2\alpha \tilde{w}^*|} \right) & \text{if } \tilde{w}^* \neq \frac{1}{2\alpha} \\
0 & \text{otherwise}
\end{cases}
\]