Discrete optimization methods for sparse problems

Théo Guyard

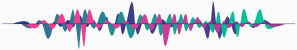
Inria, Centre de l'Université de Rennes, France

Journée Doctorant·e·s Rennais·e·s en Statistique
18th of March, 2024
Rennes, France

Sparse problems

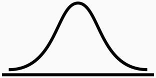
Two goals, one problem

Optimize a function $\xleftrightarrow{\textbf{Sparse problems}}$ **Sparse solution**



Signal processing

Machine learning



High-dim. statistics

And many others

Example

Sparse GLM

Features $A \in \mathbb{R}^{m \times n}$

Targets $y \in \mathbb{R}^{m}$

$\downarrow$

$\max_x \mathcal{L}(Ax, y)$

No unique solution when $m \ll n$

$\downarrow$

$\max_x \mathcal{L}(Ax, y)$ with x sparse

Example

Sparse GLM

Features $A \in \mathbb{R}^{m \times n}$

Targets $y \in \mathbb{R}^{m}$

$\downarrow$

$\max_x \ \mathcal{L}(Ax, y)$

No unique solution when $m \ll n$

$\downarrow$

$\max_x \ \mathcal{L}(Ax, y)$ with x sparse

Sparse PCA

Features $A \in \mathbb{R}^{m \times n}$

Covariance $\Sigma = A^{\mathrm{T}} A$

$\downarrow$

$\max_{\|x\|_2 = 1} \ x^{\mathrm{T}} \Sigma x$

Not relevant when $m \ll n$

$\downarrow$

$\max_{\|x\|_2 = 1} \ x^{\mathrm{T}} \Sigma x$ with x sparse

Example

Heart disease dataset (LIBSVM)

Age	Sex	Cholesterol	Blood pressure	...	**Disease**
31	M	50.3 mg/dl	95 mm/hg	...	**No**
35	F	54.9 mg/dl	98 mm/hg	...	**Yes**
42	F	49.8 mg/dl	92 mm/hg	...	**Yes**
37	M	59.1 mg/dl	89 mm/hg	...	**No**
...	...	...	...	...	...

Example

Heart disease dataset (LIBSVM)

Age	Sex	Cholesterol	Blood pressure	...	**Disease**
31	M	50.3 mg/dl	95 mm/hg	...	**No**
35	F	54.9 mg/dl	98 mm/hg	...	**Yes**
42	F	49.8 mg/dl	92 mm/hg	...	**Yes**
37	M	59.1 mg/dl	89 mm/hg	...	**No**
...	...	...	...	...	...

Data $\longrightarrow$ **Logistic regression** $\longrightarrow$ **Estimator** x

Example

Heart disease dataset (LIBSVM)

Age	Sex	Cholesterol	Blood pressure	...	**Disease**
31	M	50.3 mg/dl	95 mm/hg	...	**No**
35	F	54.9 mg/dl	98 mm/hg	...	**Yes**
42	F	49.8 mg/dl	92 mm/hg	...	**Yes**
37	M	59.1 mg/dl	89 mm/hg	...	**No**
...	...	...	...	...	...

Data $\longrightarrow$ **Logistic regression** $\longrightarrow$ **Estimator** x

Involves all the features
Explainability and robustness issues

Example

Heart disease dataset (LIBSVM)

Age	Sex	Cholesterol	Blood pressure	...	**Disease**
31	M	50.3 mg/dl	95 mm/hg	...	**No**
35	F	54.9 mg/dl	98 mm/hg	...	**Yes**
42	F	49.8 mg/dl	92 mm/hg	...	**Yes**
37	M	59.1 mg/dl	89 mm/hg	...	**No**
...	...	...	...	...	...

Data $\longrightarrow$ **Logistic regression** $\longrightarrow$ **Estimator** x

Force the use of only few features via sparsity in x

Involves all the features
Explainability and robustness issues

Objective, constraint or both ?

Optimize a function $\longleftrightarrow$ **Sparse optimization** $\longleftrightarrow$ **Sparse solution**

Loss function $F(x)$

Counting function $\|x\|_0$

Objective, constraint or both ?

Optimize a function $\longleftrightarrow$ **Sparse optimization** $\longleftrightarrow$ **Sparse solution**

Loss function $F(\mathrm{x})$

Counting function $\|\mathrm{x}\|_0$

Constrainted version

$\min_{\mathrm{x}} F(\mathrm{x})$ s.t. $\|\mathrm{x}\|_0 \leq \kappa$

Minimized version

$\min_{\mathrm{x}} \|\mathrm{x}\|_0$ s.t. $F(\mathrm{x}) \leq \epsilon$

Penalized version

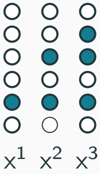
$\min_{\mathrm{x}} F(\mathrm{x}) + \lambda\|\mathrm{x}\|_0$

A bit of history

1990

Sparse heuristics

MP, OMP, IHT and co.



A bit of history

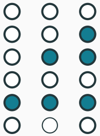
1990 **2000**

Sparse heuristics

MP, OMP, IHT and co.

Formulation with ℓ_0-norm

RIP, NSP, EkR and co.



x^1 x^2 x^3

Heuristic

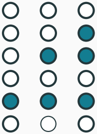
$\equiv$

ℓ_0-prob.

A bit of history

1990 — Sparse heuristics

MP, OMP, IHT and co.



x^1 x^2 x^3

2000 — Formulation with ℓ_0-norm

RIP, NSP, EkR and co.

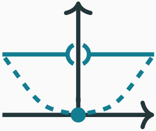
Heuristic

$|||$

ℓ_0-prob.

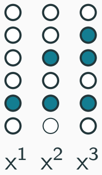
2005 — Convex approx. of ℓ_0-norm

Lasso, Elastic-Net and co.



A bit of history

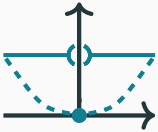
1990	2000	2005	2013
Sparse heuristics	Formulation with ℓ_0-norm	Convex approx. of ℓ_0-norm	Concave approx. of ℓ_0-norm
MP, OMP, IHT and co.	RIP, NSP, EkR and co.	Lasso, Elastic-Net and co.	SCAD, MCP CEL0 and co.

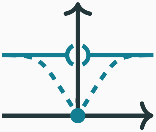


Heuristic

|||

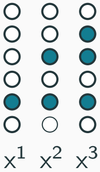
ℓ_0-prob.



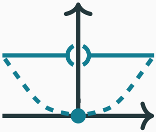


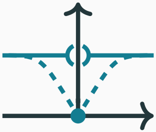
A bit of history

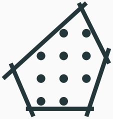
1990	2000	2005	2013	2016
Sparse heuristics	Formulation with ℓ_0-norm	Convex approx. of ℓ_0-norm	Concave approx. of ℓ_0-norm	Exact methods for ℓ_0-problems
MP, OMP, IHT and co.	RIP, NSP, EkR and co.	Lasso, Elastic-Net and co.	SCAD, MCP CEL0 and co.	MIP, BnB CP and co.



Heuristic
$\equiv$
ℓ_0-prob.







Mixed-Integer Optimization

Handeling the L0-norm with MIO tools

Sparse problem

$\min_x F(x) + \lambda\|x\|_0$

$\downarrow$ Solver

Solution

Handeling the L0-norm with MIO tools

Sparse problem

$\min_{x} F(x) + \lambda \|x\|_0$

Solver

Solution

ℓ_0-norm is non-linear, non-convex, non-smooth, NP-hard, ...

Handeling the L0-norm with MIO tools

Sparse problem

$\min_x F(x) + \lambda \|x\|_0$

Solver

Solution

ℓ_0-norm is non-linear, non-convex, non-smooth, NP-hard, ...

😵

The ℓ_0-norm counts the number of non-zeros in x

Encode the nullity in some binary vector z

We know how to deal with this in MIO !

Fitting the MIO formalism

Linearizing the ℓ_0-norm

Real vector $\mathrm{x} \in \mathrm{R}^n$ and binary vector $\mathrm{z} \in \mathrm{B}^n$:

$$\|\mathrm{x}\|_0 = 1^{\mathrm{T}}\mathrm{z} \quad \text{if} \quad \mathrm{x} \odot (1 - \mathrm{z}) = 0$$

Fitting the MIO formalism

Linearizing the ℓ_0-norm

Real vector $x \in R^n$ and binary vector $z \in B^n$:

$$\|x\|_0 = 1^{\mathrm{T}} z \quad \text{if} \quad x \odot (1 - z) = 0$$

$$\boxed{\min_x F(x) + \lambda \|x\|_0}$$

Fitting the MIO formalism

Linearizing the ℓ_0-norm

Real vector $x \in R^n$ and binary vector $z \in B^n$:

$$\|x\|_0 = 1^T z \quad \text{if} \quad x \odot (1 - z) = 0$$

$$\boxed{\min_x F(x) + \lambda \|x\|_0}$$

$$\downarrow$$

$$\boxed{\min_{x,z} F(x) + \lambda 1^T z + H(x, z)}$$

Fitting the MIO formalism

Linearizing the ℓ_0-norm

Real vector $x \in \mathbb{R}^n$ and binary vector $z \in \mathbb{B}^n$:

$$\|x\|_0 = 1^{\mathrm{T}} z \quad \text{if} \quad x \odot (1 - z) = 0$$

$$\boxed{\min_x F(x) + \lambda \|x\|_0}$$

$$\downarrow$$

$$\boxed{\min_{x,z} F(x) + \lambda 1^{\mathrm{T}} z + H(x, z)}$$

$$\downarrow$$

$$\boxed{\begin{array}{rl} \min_{x,z} & F(x) + \lambda 1^{\mathrm{T}} z \\ \text{s.t.} & -Mz \leq x \leq Mz \end{array}}$$

Fitting the MIO formalism

Linearizing the ℓ_0-norm

Real vector $x \in \mathbb{R}^n$ and binary vector $z \in \mathbb{B}^n$:

$$\|x\|_0 = 1^{\mathrm{T}} z \quad \text{if} \quad x \odot (1 - z) = 0$$

$$\boxed{\min_x F(x) + \lambda \|x\|_0}$$

$$\downarrow$$

$$\boxed{\min_{x,z} F(x) + \lambda 1^{\mathrm{T}} z + H(x, z)}$$

$$\downarrow$$

$$\boxed{\begin{array}{rl} \min_{x,z} & F(x) + \lambda 1^{\mathrm{T}} z \\ \text{s.t.} & -Mz \leq x \leq Mz \end{array}}$$

Generic MIO solvers
Branch-and-bound
Cutting planes
...

Overview of solution methods

Branch-and-Bound

"Explore regions in the feasible space and discard those that cannot contain solutions."

Cutting Planes

"Add valid cuts to restrict the feasible space until the optimum is isolated."

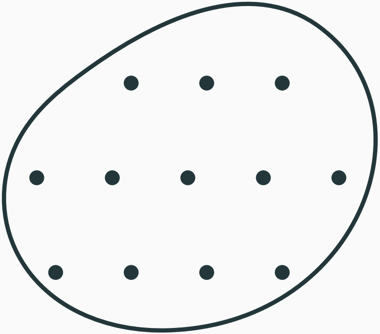
Overview of solution methods

Branch-and-Bound

"Explore regions in the feasible space and discard those that cannot contain solutions."

Cutting Planes

"Add valid cuts to restrict the feasible space until the optimum is isolated."



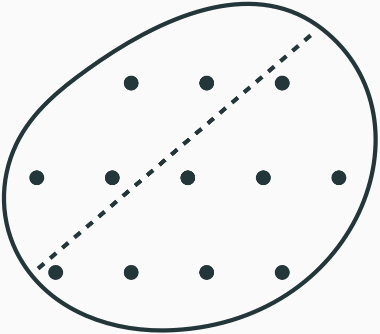
Overview of solution methods

Branch-and-Bound

"Explore regions in the feasible space and discard those that cannot contain solutions."

Cutting Planes

"Add valid cuts to restrict the feasible space until the optimum is isolated."



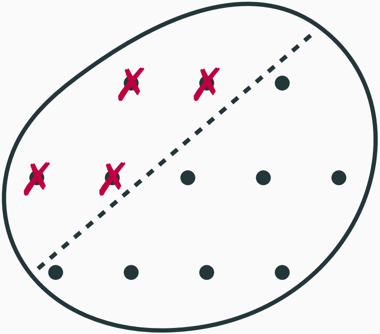
Overview of solution methods

Branch-and-Bound

"Explore regions in the feasible space and discard those that cannot contain solutions."

Cutting Planes

"Add valid cuts to restrict the feasible space until the optimum is isolated."



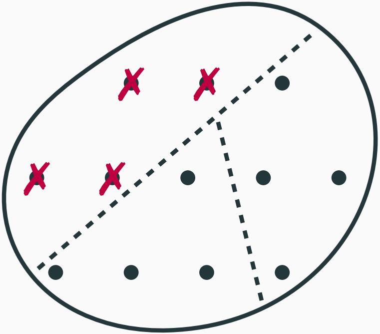
Overview of solution methods

Branch-and-Bound

"Explore regions in the feasible space and discard those that cannot contain solutions."

Cutting Planes

"Add valid cuts to restrict the feasible space until the optimum is isolated."



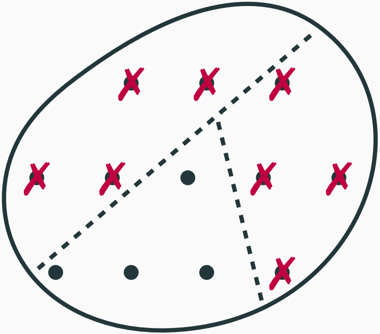
Overview of solution methods

Branch-and-Bound

"Explore regions in the feasible space and discard those that cannot contain solutions."

Cutting Planes

"Add valid cuts to restrict the feasible space until the optimum is isolated."



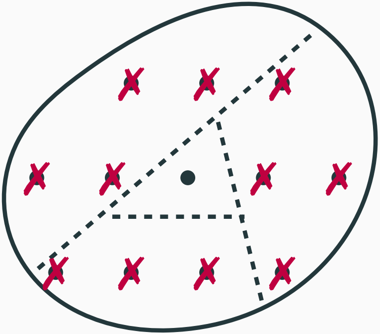
Overview of solution methods

Branch-and-Bound

"Explore regions in the feasible space and discard those that cannot contain solutions."

Cutting Planes

"Add valid cuts to restrict the feasible space until the optimum is isolated."



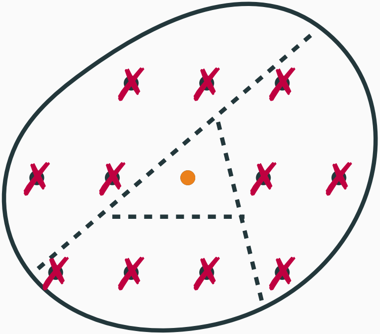
Overview of solution methods

Branch-and-Bound

"Explore regions in the feasible space and discard those that cannot contain solutions."

Cutting Planes

"Add valid cuts to restrict the feasible space until the optimum is isolated."



Ongoing Research Directions

Non-exhaustive list

Inria / Insa / Ensai / CentraleSupélec
T. Guyard, C. Herzet, C. Elvira, A. Arslan
Generalization, acceleration

Ongoing Research Directions

Non-exhaustive list

Lund University
M. Carlsson, C. Olsson...
Quadratic envelope

Frankfurt / Wurzburg Universities
C. Kanzow, A. Tillmann, ...
Optimality conditions

London Business School
J. Pauphilet, R. Cory-Wright, ...
Healthcare applications

Ponts ParisTech
M. De Lara, P. Chancelier, A. Parmentier, ...
Non-convex analysis for ℓ_0-norm, ML appli.

Centrale Nantes / ENSTA Bretagne
S. Bourguignon, J. Ninin, ...
Branch-and-Bound for ℓ_0-problems

Inria / Insa / Ensai / CentraleSupélec
T. Guyard, C. Herzet, C. Elvira, A. Arslan
Generalization, acceleration

IRIT / I3S
E. Soubies, L. Blanc-Féraud, ...
Tight relax. of ℓ_0-norm

Ongoing Research Directions

Non-exhaustive list

MIT
D. Bertsimas, R. Mazmuder, ...
MIO theory for ℓ_0-problems

Lund University
M. Carlsson, C. Olsson...
Quadratic envelope

Google Deep Mind
H. Hazimeh, A. Dedieu, ...
MIO-based heuristics

Frankfurt / Wurzburg Universities
C. Kanzow, A. Tillmann, ...
Optimality conditions

London Business School
J. Pauphilet, R. Cory-Wright, ...
Healthcare applications

Berkley
A. Atamtürk, A. Gomès, ...
Convex-based acceleration

Ponts ParisTech
M. De Lara, P. Chancelier, A. Parmentier, ...
Non-convex analysis for ℓ_0-norm, ML appli.

Centrale Nantes / ENSTA Bretagne
S. Bourguignon, J. Ninin, ...
Branch-and-Bound for ℓ_0-problems

Inria / Insa / Ensai / CentraleSupélec
T. Guyard, C. Herzet, C. Elvira, A. Arslan
Generalization, acceleration

IRIT / I3S
E. Soubies, L. Blanc-Féraud, ...
Tight relax. of ℓ_0-norm

Take-home message

- ℓ_0-norm problems arise in many applied mathematical fields
- Mixed-integer optimization tools to address them
- Structure exploitation is the key to achieve competitive performances
- Active research area
  - $\rightarrow$ Theoretical results
  - $\rightarrow$ Efficiency, flexibility and accessibility of solution methods
  - $\rightarrow$ Software development
  - $\rightarrow$ Diffusion to other communities

Question time

🤓