Optimization methods for ℓ_0-problems

Théo Guyard

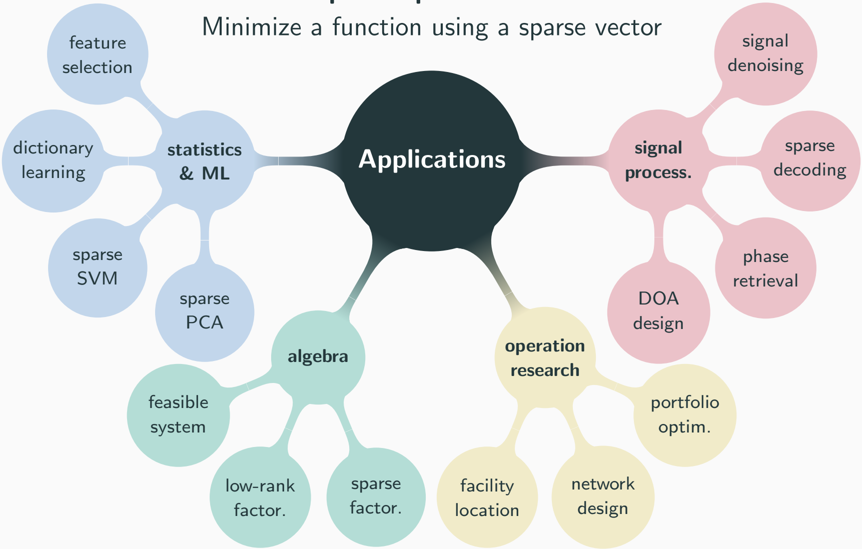
CIRRELT, Montréal, Canada – March 6th, 2025

Sparse optimization

Minimize a function using a sparse vector

Sparse optimization

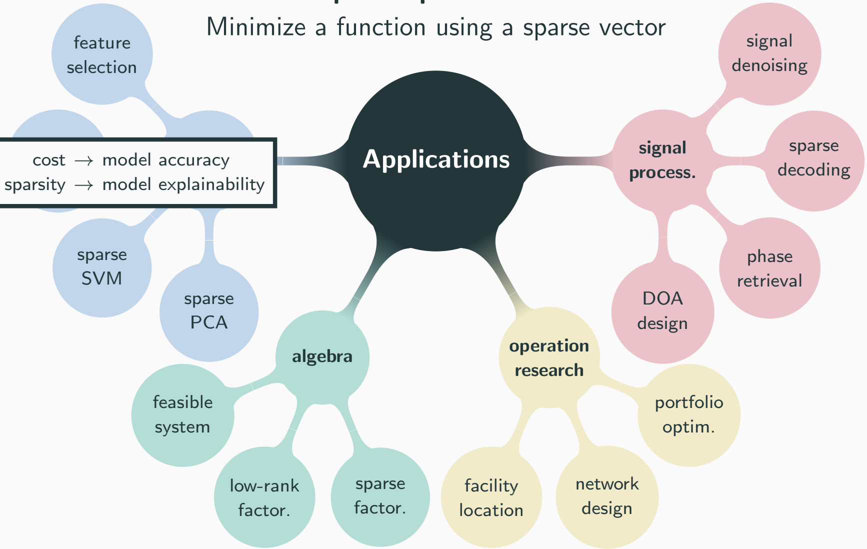
Minimize a function using a sparse vector



A. Tillmann *et. al* (2024)

Sparse optimization

Minimize a function using a sparse vector

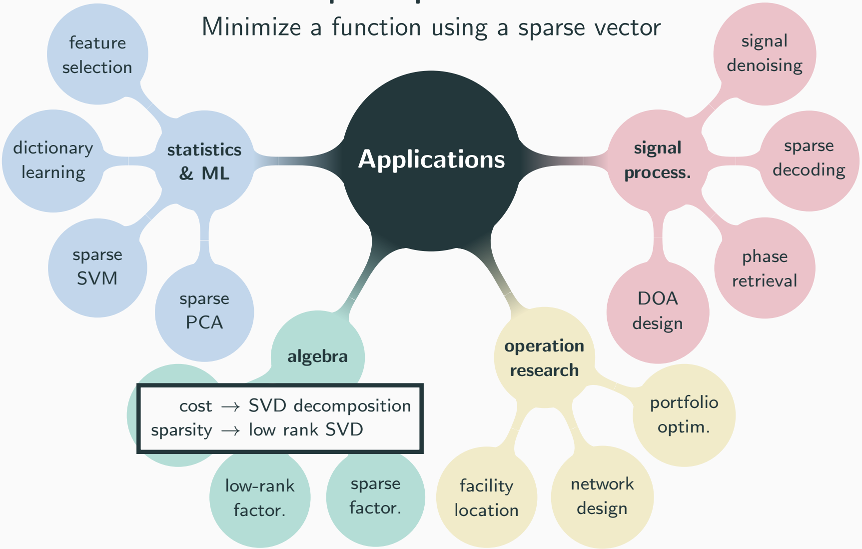


cost $\longrightarrow$ model accuracy
sparsity $\longrightarrow$ model explainability

A. Tillmann *et. al* (2024)

Sparse optimization

Minimize a function using a sparse vector



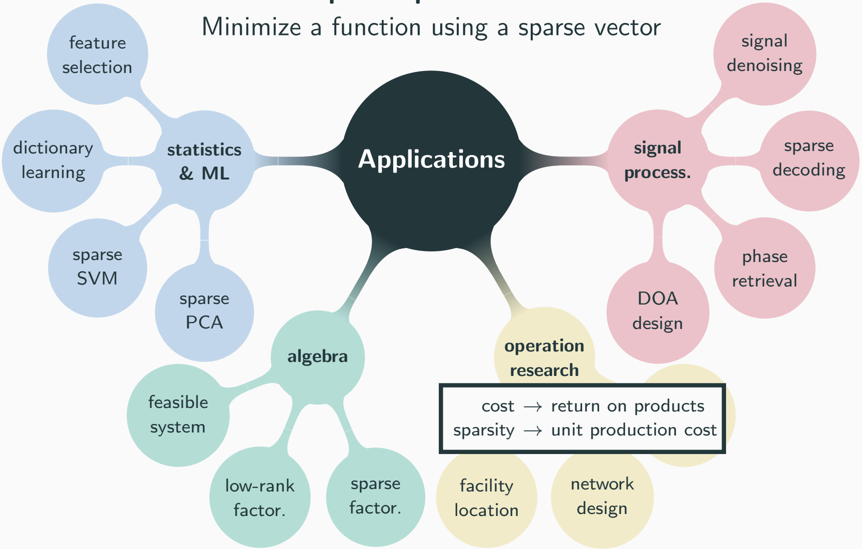
cost $\longrightarrow$ SVD decomposition
sparsity $\longrightarrow$ low rank SVD

A. Tillmann *et. al* (2024)

Sparse optimization

Sparse optimization

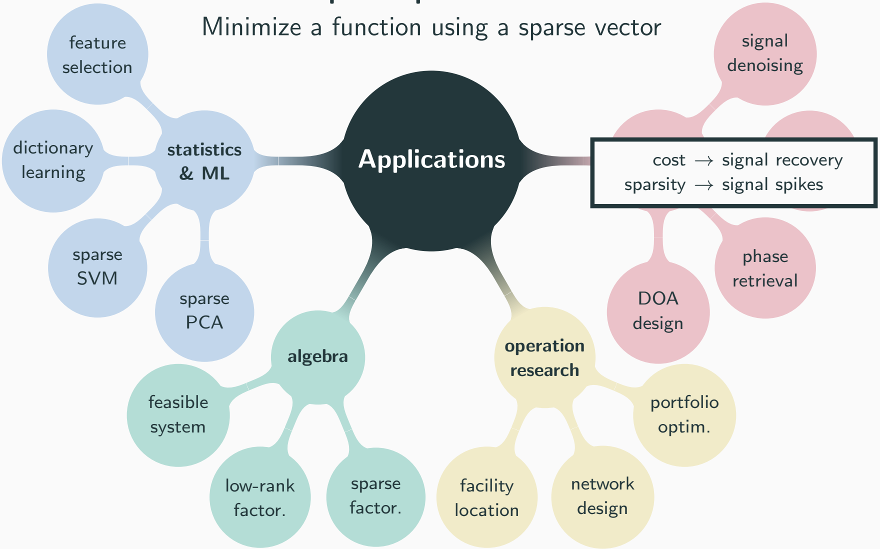
Minimize a function using a sparse vector



A. Tillmann *et. al* (2024)

Sparse optimization

Minimize a function using a sparse vector



A. Tillmann *et. al* (2024)

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a function using a sparse vector

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a function using a sparse vector

$\boxed{\text{cost}}$

$f(\mathbf{x})$

loss function

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a function using a sparse vector

cost

sparsity

number of

non-zeros in $\mathbf{x}$

$f(\mathbf{x})$

loss function

$\|\mathbf{x}\|_0$

ℓ_0-norm

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a function using a sparse vector

cost

sparsity

number of non-zeros in $\mathbf{x}$

$f(\mathbf{x})$ loss function

optimization problem

$\|\mathbf{x}\|_0$ ℓ_0-norm

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a function using a sparse vector

cost

sparsity

number of non-zeros in $\mathbf{x}$

$f(\mathbf{x})$
loss function

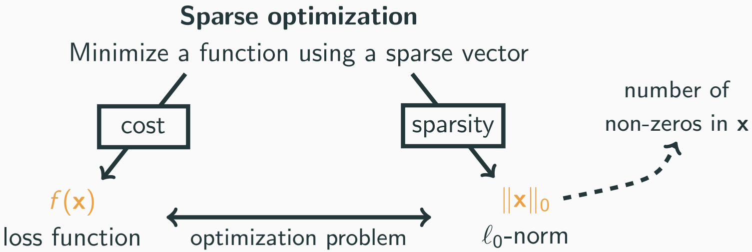
optimization problem

$\|\mathbf{x}\|_0$
ℓ_0-norm

Constrained problem

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbf{R}^n} & f(\mathbf{x}) \\ \text{subject to} & \|\mathbf{x}\|_0 \leq s \end{array}$$

Minimized, constrained, or regularized problem ?



Constrained problem

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbf{R}^n} & f(\mathbf{x}) \\ \text{subject to} & \|\mathbf{x}\|_0 \leq s \end{array}$$

Minimized problem

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbf{R}^n} & \|\mathbf{x}\|_0 \\ \text{subject to} & f(\mathbf{x}) \leq \epsilon \end{array}$$

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a function using a sparse vector

cost

sparsity

number of non-zeros in $\mathbf{x}$

$f(\mathbf{x})$

loss function

optimization problem

$\|\mathbf{x}\|_0$

ℓ_0-norm

Constrained problem

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbf{R}^n} & f(\mathbf{x}) \\ \text{subject to} & \|\mathbf{x}\|_0 \leq s \end{array}$$

Minimized problem

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbf{R}^n} & \|\mathbf{x}\|_0 \\ \text{subject to} & f(\mathbf{x}) \leq \epsilon \end{array}$$

Regularized problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0$$

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a function using a sparse vector

$\boxed{\text{cost}}$ $\boxed{\text{sparsity}}$

number of non-zeros in $\mathbf{x}$

$f(\mathbf{x})$ loss function $\longleftrightarrow$ optimization problem $\|\mathbf{x}\|_0$ ℓ_0-norm

Constrained problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbf{R}^n} \quad & f(\mathbf{x}) + h(\mathbf{x}) \\ \text{subject to} \quad & \|\mathbf{x}\|_0 \leq s \end{aligned}$$

Minimized problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbf{R}^n} \quad & \|\mathbf{x}\|_0 + h(\mathbf{x}) \\ \text{subject to} \quad & f(\mathbf{x}) \leq \epsilon \end{aligned}$$

Regularized problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

A bit of history

Problem

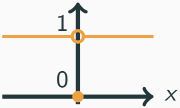
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

A bit of history

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

ℓ_0-norm in 1d

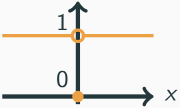


A bit of history

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

ℓ_0-norm in 1d



non-convex, non-smooth,
non-continuous, ...

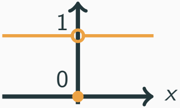
A bit of history

Problem

$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

NP-hard to solve

ℓ_0-norm in 1d



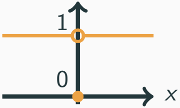
non-convex, non-smooth,
non-continuous, ...

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

NP-hard to solve

ℓ_0-norm in 1d

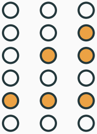


non-convex, non-smooth,
non-continuous, ...

1995

Heuristics

MP, OMP, ...
S. Mallat (1993)



$\mathbf{x}^1$ $\mathbf{x}^2$ $\mathbf{x}^3$

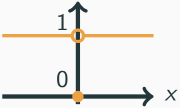
A bit of history

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

NP-hard to solve

ℓ_0-norm in 1d



non-convex, non-smooth, non-continuous, ...

1995

Heuristics

MP, OMP, ...

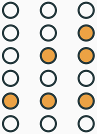
S. Mallat (1993)

2000

Recovery cond.

RIP, NSP, ...

E. Candes (2004)



$\mathbf{x}^1$ $\mathbf{x}^2$ $\mathbf{x}^3$

Heuristics solve ℓ_0-problem under some (strong) conditions

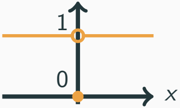
A bit of history

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

NP-hard to solve

ℓ_0-norm in 1d

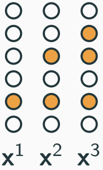


non-convex, non-smooth, non-continuous, ...

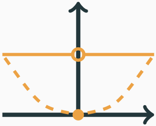
1995 — Heuristics
MP, OMP, ...
S. Mallat (1993)

2000 — Recovery cond.
RIP, NSP, ...
E. Candes (2004)

2005 — Convex approx.
Lasso, Elastic-Net, ...
R. Tibshirani (2005)



Heuristics solve ℓ_0-problem under some (strong) conditions



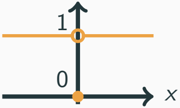
A bit of history

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

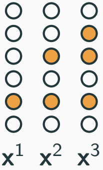
NP-hard to solve

ℓ_0-norm in 1d

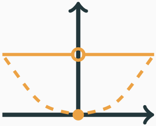


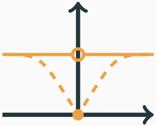
non-convex, non-smooth, non-continuous, ...

1995	2000	2005	2010
Heuristics	Recovery cond.	Convex approx.	Concave approx.
MP, OMP, ...	RIP, NSP, ...	Lasso, Elastic-Net, ...	SCAD, MCP, ...
S. Mallat (1993)	E. Candes (2004)	R. Tibshirani (2005)	C. Zhang (2010)



Heuristics solve ℓ_0-problem under some (strong) conditions





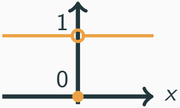
A bit of history

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

NP-hard to solve

ℓ_0-norm in 1d



non-convex, non-smooth, non-continuous, ...

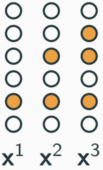
1995 — Heuristics
MP, OMP, ...
S. Mallat (1993)

2000 — Recovery cond.
RIP, NSP, ...
E. Candes (2004)

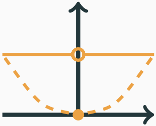
2005 — Convex approx.
Lasso, Elastic-Net, ...
R. Tibshirani (2005)

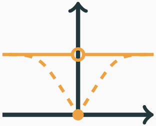
2010 — Concave approx.
SCAD, MCP, ...
C. Zhang (2010)

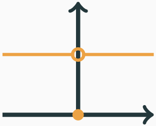
2015 — Exact methods
MIP, BnB, ...
D. Bertsimas (2016)



Heuristics solve ℓ_0-problem under some (strong) conditions







Ok, ℓ_0-problems can be addressed exactly!

(you)

Topic of this talk

Ok, ℓ_0-problems can be addressed exactly!

🤓

(you)

What are the solution methods then?

🤔

(also you if you're an optim. nerd)

Topic of this talk

Ok, ℓ_0-problems can be addressed exactly!

🤓

(you)

What are the solution methods then?

🤔

(also you if you're an optim. nerd)

1) MIP-based methods

Based on off-the-shelf solvers

Poor numerical performances

Topic of this talk

Ok, ℓ_0-problems can be addressed exactly!

(you)

What are the solution methods then?

(also you if you're an optim. nerd)

1) MIP-based methods

Based on off-the-shelf solvers

Poor numerical performances

2) BnB-based methods

Tailored solution method

Better numerical performances

Mixed-Integer Programming

MIP – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

$\downarrow$

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

MIP – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

$\downarrow$

Problem

$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$

MIP formulation

Standardized expressions

MIP – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

$\downarrow$

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$

MIP formulation

Standardized expressions

Modelling framework

Python $\rightarrow$ `cvxpy`

Julia $\rightarrow$ `JuMP`

C++ $\rightarrow$ `CMPL`

Matlab $\rightarrow$ `YALMIP`

...

MIP – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

$\downarrow$

Problem

$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$

MIP formulation

Standardized expressions

$\downarrow$

Problem solutions

Modelling framework

Python $\rightarrow$ `cvxpy`

Julia $\rightarrow$ `JuMP`

C++ $\rightarrow$ `CMPL`

Matlab $\rightarrow$ `YALMIP`

...

MIP – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Modelling framework

Python → `cvxpy`

Julia → `JuMP`

C++ → `CMPL`

Matlab → `YALMIP`

...

MIP formulation

Standardized expressions

Off-the-shelf solver

Gurobi, CPLEX, SCIP,

Mosek, CbC, ...

Problem solutions

MIP – Formulation

Problem

$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

MIP formulation

Use standardized expressions

linear, quadratic, conic, ...

MIP – Formulation

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

MIP formulation

Use standardized expressions

linear, quadratic, conic, ...

Linearize the ℓ_0-norm

We have $\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$ whenever

$$z_i = 0 \iff x_i = 0, \ \forall i$$
$$z_i = 1 \iff x_i \neq 0, \ \forall i$$

for all $\mathbf{x} \in \mathbf{R}^n$ and $\mathbf{z} \in \{0,1\}^n$

MIP – Formulation

Problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

$$\Big\downarrow \ \boxed{\text{linearize the } \ell_0\text{-norm}}$$

Linearized formulation

$$\begin{cases} \min & f(\mathbf{x}) + \lambda\mathbf{1}^{\mathrm{T}}\mathbf{z} + h(\mathbf{x}) \\ \text{s.t.} & z_i = 0 \implies x_i = 0, \ \forall i \\ & \mathbf{x} \in \mathbf{R}^n, \ \mathbf{z} \in \{0,1\}^n \end{cases}$$

MIP formulation

Use standardized expressions

linear, quadratic, conic, ...

Linearize the ℓ_0-norm

We have $\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$ whenever

$$z_i = 0 \iff x_i = 0, \ \forall i$$

$$z_i = 1 \iff x_i \neq 0, \ \forall i$$

for all $\mathbf{x} \in \mathbf{R}^n$ and $\mathbf{z} \in \{0,1\}^n$

MIP – Formulation

Problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

$$\Big\downarrow \boxed{\text{linearize the } \ell_0\text{-norm}}$$

Linearized formulation

$$\begin{cases} \min & f(\mathbf{x}) + \lambda\mathbf{1}^{\mathrm{T}}\mathbf{z} + h(\mathbf{x}) \\ \text{s.t.} & z_i = 0 \implies x_i = 0, \ \forall i \\ & \mathbf{x} \in \mathbf{R}^n, \ \mathbf{z} \in \{0,1\}^n \end{cases}$$

MIP formulation

Use standardized expressions

linear, quadratic, conic, ...

Linearize the ℓ_0-norm

We have $\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$ whenever

$$\begin{array}{ccccccc} z_i & = & 0 & \iff & x_i & = & 0, \ \forall i \\ z_i & = & 1 & \iff & x_i & \neq & 0, \ \forall i \end{array}$$

for all $\mathbf{x} \in \mathbf{R}^n$ and $\mathbf{z} \in \{0,1\}^n$

Avoid logical constraint

$$\tilde{h}(\mathbf{x},\mathbf{z}) = \begin{cases} h(\mathbf{x}) & \text{if } z_i = x_i = 0 \text{ or } z_i = 1, \forall i \\ +\infty & \text{otherwise} \end{cases}$$

MIP – Formulation

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\boxed{\text{linearize the } \ell_0\text{-norm}}$

Linearized formulation

$$\begin{cases} \min & f(\mathbf{x}) + \lambda\mathbf{1}^{\mathrm{T}}\mathbf{z} + h(\mathbf{x}) \\ \text{s.t.} & z_i = 0 \implies x_i = 0,\ \forall i \\ & \mathbf{x} \in \mathbf{R}^n,\ \mathbf{z} \in \{0,1\}^n \end{cases}$$

$\boxed{\text{avoid logical cstr.}}$

MIP formulation

$$\begin{cases} \min & f(\mathbf{x}) + \lambda\mathbf{1}^{\mathrm{T}}\mathbf{z} + \tilde{h}(\mathbf{x},\mathbf{z}) \\ \text{s.t.} & \mathbf{x} \in \mathbf{R}^n,\ \mathbf{z} \in \{0,1\}^n \end{cases}$$

MIP formulation

Use standardized expressions

linear, quadratic, conic, ...

Linearize the ℓ_0-norm

We have $\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$ whenever

$$z_i = 0 \iff x_i = 0,\ \forall i$$

$$z_i = 1 \iff x_i \neq 0,\ \forall i$$

for all $\mathbf{x} \in \mathbf{R}^n$ and $\mathbf{z} \in \{0,1\}^n$

Avoid logical constraint

$$\tilde{h}(\mathbf{x},\mathbf{z}) = \begin{cases} h(\mathbf{x}) & \text{if } z_i = x_i = 0 \text{ or } z_i = 1, \forall i \\ +\infty & \text{otherwise} \end{cases}$$

MIP – Penalty function

Two existing strategies to design $\tilde{h}(\mathbf{x}, \mathbf{z})$

$$\tilde{h}(\mathbf{x}, \mathbf{z}) = \begin{cases} h(\mathbf{x}) & \text{if } z_i = x_i = 0 \text{ or } z_i = 1, \forall i \\ +\infty & \text{otherwise} \end{cases}$$

MIP – Penalty function

Two existing strategies to design $\tilde{h}(\mathbf{x}, \mathbf{z})$

$$\tilde{h}(\mathbf{x}, \mathbf{z}) = \begin{cases} h(\mathbf{x}) & \text{if } z_i = x_i = 0 \text{ or } z_i = 1, \forall i \\ +\infty & \text{otherwise} \end{cases}$$

Big-M strategy

$h(\mathbf{x}) = \sum_{i=1}^{n} \operatorname{Ind}(|x_i| \leq M)$

with $M > 0$

MIP – Penalty function

Two existing strategies to design $\tilde{h}(\mathbf{x}, \mathbf{z})$

$$\tilde{h}(\mathbf{x}, \mathbf{z}) = \begin{cases} h(\mathbf{x}) & \text{if } z_i = x_i = 0 \text{ or } z_i = 1, \forall i \\ +\infty & \text{otherwise} \end{cases}$$

Big-M strategy

$h(\mathbf{x}) = \sum_{i=1}^{n} \mathrm{Ind}(|x_i| \leq M)$

with $M > 0$

$$\downarrow$$

$\tilde{h}(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \mathrm{Ind}(|x_i| \leq M z_i)$

MIP – Penalty function

Two existing strategies to design $\tilde{h}(\mathbf{x}, \mathbf{z})$

$$\tilde{h}(\mathbf{x}, \mathbf{z}) = \begin{cases} h(\mathbf{x}) & \text{if } z_i = x_i = 0 \text{ or } z_i = 1, \forall i \\ +\infty & \text{otherwise} \end{cases}$$

Big-M strategy

$$h(\mathbf{x}) = \sum_{i=1}^{n} \mathrm{Ind}(|x_i| \leq M)$$

with $M > 0$

$$\downarrow$$

$$\tilde{h}(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \mathrm{Ind}(|x_i| \leq M z_i)$$

ℓ_2-norm strategy

$$h(\mathbf{x}) = \sum_{i=1}^{n} \gamma x_i^2$$

with $\gamma > 0$

MIP – Penalty function

Two existing strategies to design $\tilde{h}(\mathbf{x}, \mathbf{z})$

$$\tilde{h}(\mathbf{x}, \mathbf{z}) = \begin{cases} h(\mathbf{x}) & \text{if } z_i = x_i = 0 \text{ or } z_i = 1, \forall i \\ +\infty & \text{otherwise} \end{cases}$$

Big-M strategy

$h(\mathbf{x}) = \sum_{i=1}^{n} \mathrm{Ind}(|x_i| \leq M)$

with $M > 0$

$$\downarrow$$

$\tilde{h}(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \mathrm{Ind}(|x_i| \leq M z_i)$

ℓ_2-norm strategy

$h(\mathbf{x}) = \sum_{i=1}^{n} \gamma x_i^2$

with $\gamma > 0$

$$\downarrow$$

$\tilde{h}(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \gamma \frac{x_i^2}{z_i}$

MIP – Penalty function

Two existing strategies to design $\tilde{h}(\mathbf{x}, \mathbf{z})$

$$\tilde{h}(\mathbf{x}, \mathbf{z}) = \begin{cases} h(\mathbf{x}) & \text{if } z_i = x_i = 0 \text{ or } z_i = 1, \forall i \\ +\infty & \text{otherwise} \end{cases}$$

Big-M strategy

$$h(\mathbf{x}) = \sum_{i=1}^{n} \mathrm{Ind}(|x_i| \leq M)$$

with $M > 0$

$$\downarrow$$

$$\tilde{h}(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \mathrm{Ind}(|x_i| \leq M z_i)$$

Linear expressions

ℓ_2-norm strategy

$$h(\mathbf{x}) = \sum_{i=1}^{n} \gamma x_i^2$$

with $\gamma > 0$

$$\downarrow$$

$$\tilde{h}(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \gamma \frac{x_i^2}{z_i}$$

Conic expressions

Standard for off-the-shelf MIP solvers

MIP – Let's sum up

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Pipeline

1) Introduce binary variable $\mathbf{z} \in \{0,1\}^n$

2) Linearize ℓ_0-norm as $\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$

3) Transform $h(\mathbf{x})$ into $\tilde{h}(\mathbf{x},\mathbf{z})$

4) Use generic MIP solvers

MIP – Let's sum up

Problem

$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

Pipeline

1) Introduce binary variable $\mathbf{z} \in \{0,1\}^n$

2) Linearize ℓ_0-norm as $\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$

3) Transform $h(\mathbf{x})$ into $\tilde{h}(\mathbf{x}, \mathbf{z})$

4) Use generic MIP solvers

Pros

✓ Rich MIP literature

✓ Black-box solvers

✓ Straightforward pipeline

MIP – Let's sum up

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Pipeline

1) Introduce binary variable $\mathbf{z} \in \{0,1\}^n$

2) Linearize ℓ_0-norm as $\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$

3) Transform $h(\mathbf{x})$ into $\tilde{h}(\mathbf{x},\mathbf{z})$

4) Use generic MIP solvers

Pros

✓ Rich MIP literature

✓ Black-box solvers

✓ Straightforward pipeline

Cons

✗ Mostly commercial solvers

✗ Only for $h =$ big-M/ℓ_2-norm

✗ Performance issues

Branch-and-Bound Algorithms

BnB – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

$\downarrow$

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

BnB – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

$\downarrow$

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$

Implement BnB solver

Specialized mechanisms

BnB – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

$\downarrow$

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$

Implement BnB solver

Specialized mechanisms

$\downarrow$

Problem solutions

BnB – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

$\downarrow$

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$

Implement BnB solver

Specialized mechanisms

$\downarrow$

Problem solutions

Existing BnB solver

`sbnb` $\rightarrow$ G. Samain *et al.* (2020)

`l0bnb` $\rightarrow$ H. Hazimeh *et al.* (2021)

`el0ps` $\rightarrow$ T. Guyard *et al.* (2024)

BnB – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Implement BnB solver

Specialized mechanisms

Problem solutions

Existing BnB solver

`sbnb` → G. Samain *et al.* (2020)

`l0bnb` → H. Hazimeh *et al.* (2021)

`el0ps` → T. Guyard *et al.* (2024)

Why using `el0ps`?

Is is free, fast and flexible!

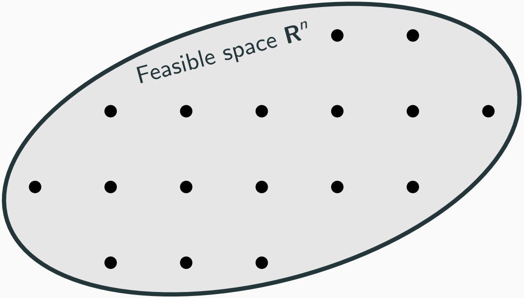
😎

BnB – Algorithmic principle

Explore regions in the feasible space and prune those that cannot contain any optimal solution.

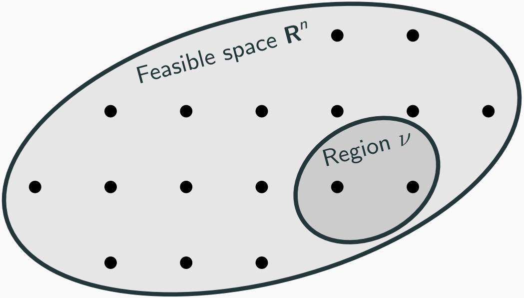
BnB – Algorithmic principle

Explore regions in the feasible space and prune those that cannot contain any optimal solution.



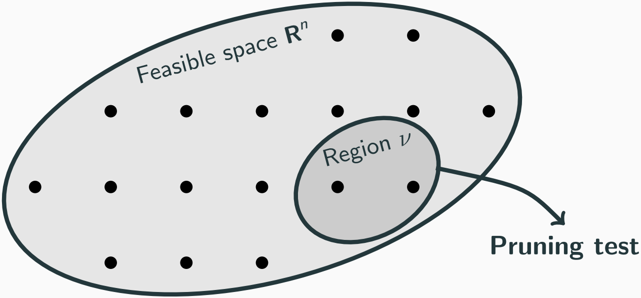
BnB – Algorithmic principle

Explore regions in the feasible space and prune those that cannot contain any optimal solution.



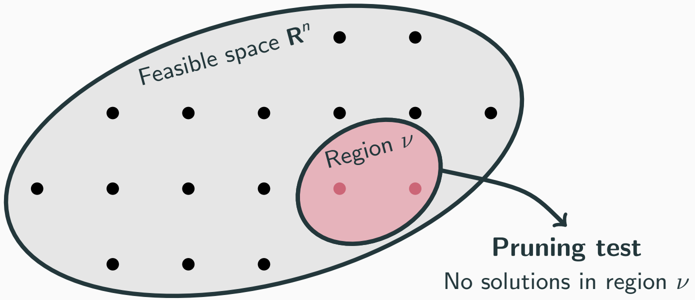
BnB – Algorithmic principle

Explore regions in the feasible space and prune those that cannot contain any optimal solution.



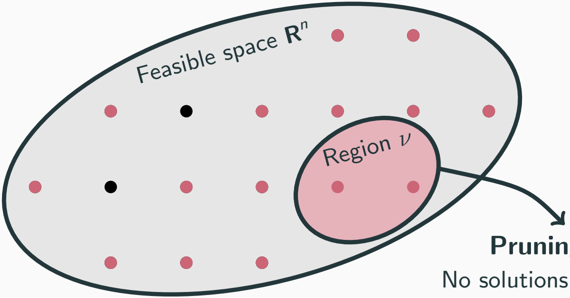
BnB – Algorithmic principle

Explore regions in the feasible space and prune those that cannot contain any optimal solution.



BnB – Algorithmic principle

Explore regions in the feasible space and prune those that cannot contain any optimal solution.

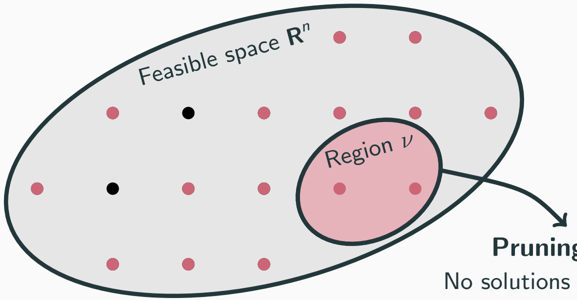


Pruning test

No solutions in region ν

BnB – Algorithmic principle

Explore regions in the feasible space and prune those that cannot contain any optimal solution.



Pruning test

No solutions in region ν

Branching step – Region design and exploration

Bounding step – Pruning test evaluation

BnB – Branching step

Problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

BnB – Branching step

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

BnB – Branching step

Problem

$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

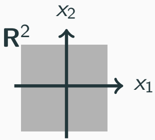
Drive the sparsity of the optimization variable

Problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

Observation

Solutions are expected to be sparse



Method

Drive the sparsity of the optimization variable

BnB – Branching step

Problem

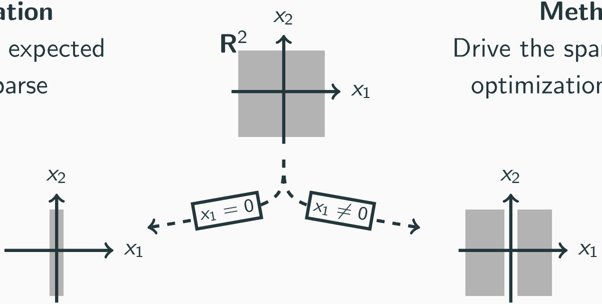
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



Problem

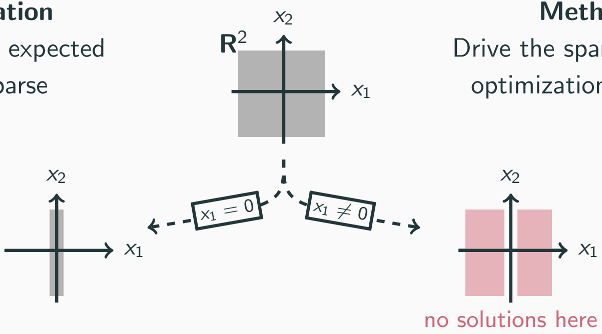
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



BnB – Branching step

Problem

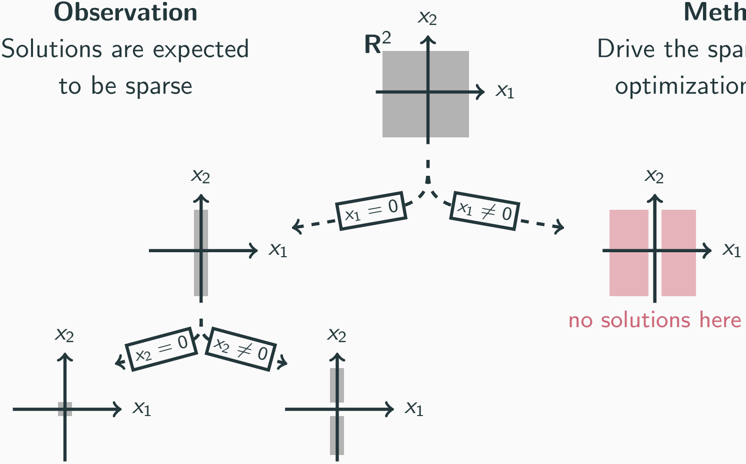
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



BnB – Branching step

Problem

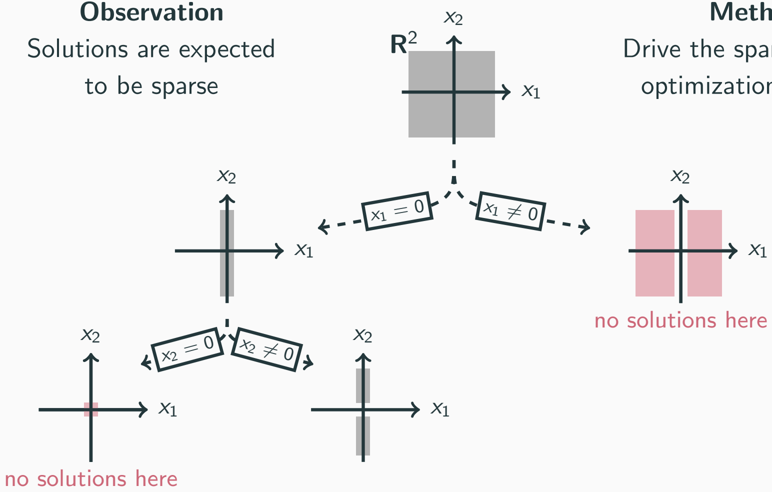
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



BnB – Branching step

Problem

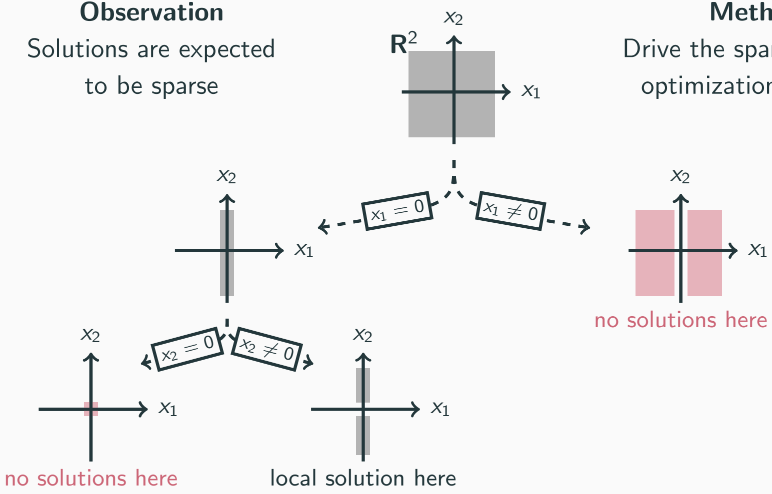
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable

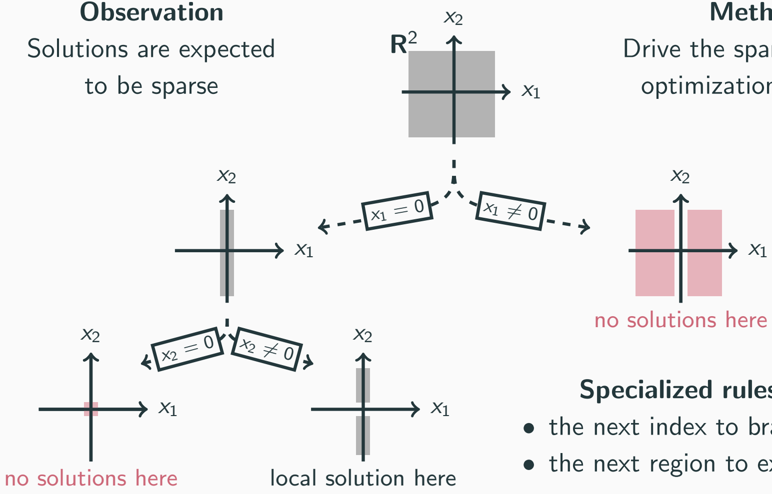


Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse



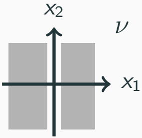
Method

Drive the sparsity of the optimization variable

Specialized rules for

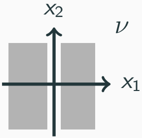
- the next index to branch on
- the next region to explore

BnB – Bounding step



Does region ν contains optimal solutions ?

BnB – Bounding step

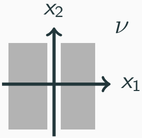


Does region ν contains optimal solutions ?

Problem

$p^\star = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

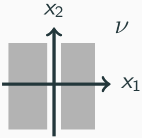
$$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

restrict to ν

Restriction to region ν

$$p^\nu = \min_{\mathbf{x}\in\nu} \; f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$ restrict to ν

Restriction to region ν

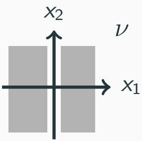
$p^\nu = \min_{\mathbf{x}\in\nu} \; f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$ compare

Pruning test

$p^\nu > p^\star$

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

↓ restrict to ν

Restriction to region ν

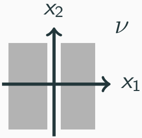
$p^\nu = \min_{\mathbf{x}\in\nu} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

↓ compare

Pruning test

$p^\nu > p^\star$ $\longrightarrow$ prune ν

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

$p^\star = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$ restrict to ν

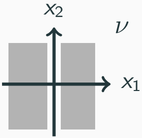
Restriction to region ν

$p^\nu = \min_{\mathbf{x} \in \nu} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$ compare

Pruning test

$p^\nu_{\mathrm{lb}} > p^\star_{\mathrm{ub}}$ $\longrightarrow$ prune ν



Does region ν contains optimal solutions ?

Problem

$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

restrict to ν

Restriction to region ν

$p^\nu = \min_{\mathbf{x}\in\nu} \; f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

compare

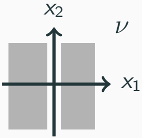
Pruning test

$p^\nu_{\mathrm{lb}} > p^\star_{\mathrm{ub}}$ $\longrightarrow$ prune ν

Easy task

Compute an upper bound on $p^\star$

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

restrict to ν

Restriction to region ν

$p^\nu = \min_{\mathbf{x}\in\nu} \; f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

compare

Pruning test

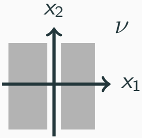
$p^\nu_{\text{lb}} > p^\star_{\text{ub}}$ $\longrightarrow$ prune ν

Easy task

$\boxed{\text{Compute an upper bound on } p^\star}$

Construct and evaluate
a feasible vector in each
region explored to refine $p^\star_{\text{ub}}$

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

restrict to ν

Restriction to region ν

$p^\nu = \min_{\mathbf{x}\in\nu}\ f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

compare

Pruning test

$p^\nu_{\text{lb}} > p^\star_{\text{ub}}$ $\longrightarrow$ prune ν

Easy task

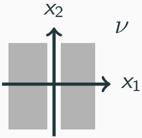
Compute an upper bound on $p^\star$

Construct and evaluate
a feasible vector in each
region explored to refine $p^\star_{\text{ub}}$

Main challenge

Compute a lower bound on p^ν

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$ restrict to ν

Restriction to region ν

$p^\nu = \min_{\mathbf{x}\in\nu} \; f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$ compare

Pruning test

$p^\nu_{\text{lb}} > p^\star_{\text{ub}}$ $\longrightarrow$ prune ν

Easy task

$\boxed{\text{Compute an upper bound on } p^\star}$

Construct and evaluate
a feasible vector in each
region explored to refine $p^\star_{\text{ub}}$

Main challenge

$\boxed{\text{Compute a lower bound on } p^\nu}$

Construct and
solve a relaxation

BnB – Building relaxations

Restriction to region ν

$p^\nu = \min_{\mathbf{x}\in\nu} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

seek tight/tractable lower bound on p^ν

BnB – Building relaxations

Restriction to region ν

$$p^\nu = \min_{\mathbf{x}\in\nu} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

seek tight/tractable lower bound on p^ν

$\downarrow$ reformulation

Restriction to region ν

$$p^\nu = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^\nu(\mathbf{x})$$

with g^ν proper and closed

BnB – Building relaxations

Restriction to region ν

$p^\nu = \min_{\mathbf{x} \in \nu} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

seek tight/tractable lower bound on p^ν

$\downarrow$ reformulation

Restriction to region ν

$p^\nu = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g^\nu(\mathbf{x})$

with g^ν proper and closed

$\downarrow$ $g^\nu_{\mathrm{lb}} \leq g^\nu$, g^ν_{lb} convex

Relaxation for region ν

$p^\nu_{\mathrm{lb}} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g^\nu_{\mathrm{lb}}(\mathbf{x})$

set g^ν_{lb} set as the convex envelope of g^ν

BnB – Geometrical intuition

Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ with $\mathbf{x} \in \mathbf{R}^n$

BnB – Geometrical intuition

Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ **with** $\mathbf{x} \in \mathbf{R}^n$

$\uparrow$

h separable and coercive

$\downarrow$

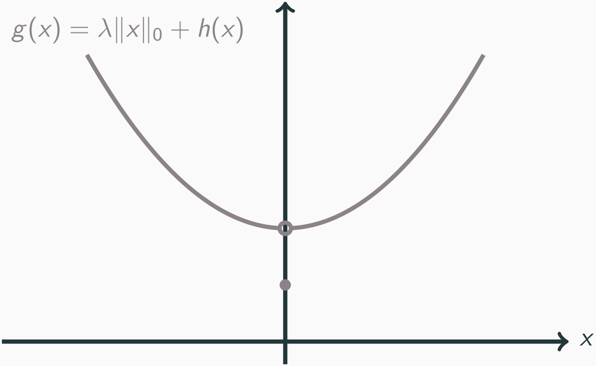
Convex envelope of $g(x) = \lambda\|x\|_0 + h(x)$ **with** $x \in \mathbf{R}$

BnB – Geometrical intuition

Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ with $\mathbf{x} \in \mathbf{R}^n$

h separable and coercive

Convex envelope of $g(x) = \lambda\|x\|_0 + h(x)$ with $x \in \mathbf{R}$



BnB – Geometrical intuition

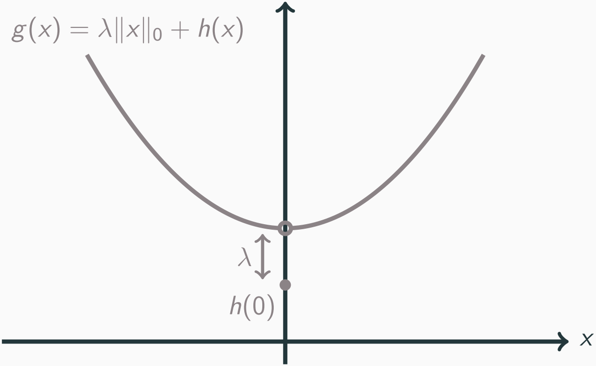
Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ with $\mathbf{x} \in \mathbf{R}^n$

$\uparrow$

$\boxed{h \text{ separable and coercive}}$

$\downarrow$

Convex envelope of $g(x) = \lambda\|x\|_0 + h(x)$ with $x \in \mathbf{R}$

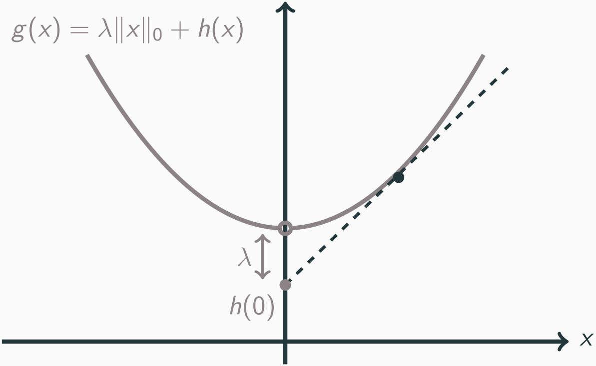


BnB – Geometrical intuition

Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ **with** $\mathbf{x} \in \mathbf{R}^n$

h separable and coercive

Convex envelope of $g(x) = \lambda\|x\|_0 + h(x)$ **with** $x \in \mathbf{R}$

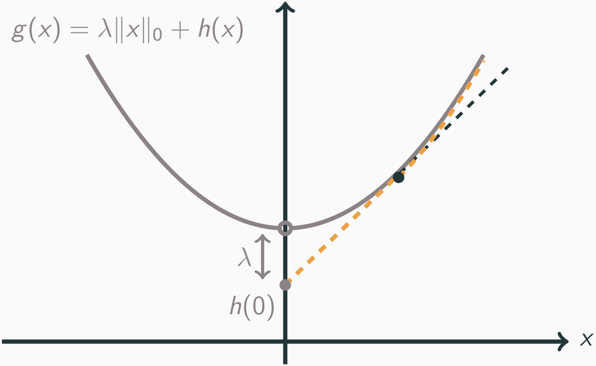


BnB – Geometrical intuition

Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ **with** $\mathbf{x} \in \mathbf{R}^n$

h separable and coercive

Convex envelope of $g(x) = \lambda\|x\|_0 + h(x)$ **with** $x \in \mathbf{R}$

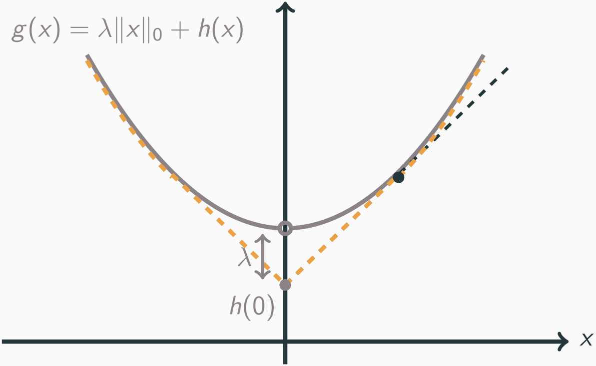


BnB – Geometrical intuition

Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ **with** $\mathbf{x} \in \mathbf{R}^n$

$\boxed{h \text{ separable and coercive}}$

Convex envelope of $g(x) = \lambda\|x\|_0 + h(x)$ **with** $x \in \mathbf{R}$



BnB – The secrete sauce

Solve time

region processing time $\times$ number of regions processed

BnB – The secrete sauce

Solve time

$\underbrace{\text{region processing time}}$ × number of regions processed

Relaxation for region ν

$p_{\text{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\text{lb}}^{\nu}(\mathbf{x})$

BnB – The secrete sauce

Solve time

$\underbrace{\text{region processing time}} \times$ number of regions processed

Relaxation for region ν

$p_{\text{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\text{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

Solve time

$\underbrace{\text{region processing time}} \times$ number of regions processed

Relaxation for region ν

$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

BnB – The secrete sauce

Solve time

$\underbrace{\text{region processing time}}$ $\times$ number of regions processed

Relaxation for region ν

$p^{\nu}_{\text{lb}} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g^{\nu}_{\text{lb}}(\mathbf{x})$

g^{ν}_{lb} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

$\rightarrow$ acceleration strategies

working set, screening tests, ...

BnB – The secrete sauce

Solve time

$\underbrace{\text{region processing time}} \times \underbrace{\text{number of regions processed}}$

Relaxation for region ν

$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning

BnB – The secrete sauce

Solve time

region processing time $\times$ number of regions processed

Relaxation for region ν

$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

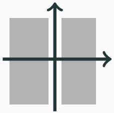
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...

BnB – The secrete sauce

Solve time

region processing time $\times$ number of regions processed

Relaxation for region ν

$$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

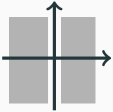
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

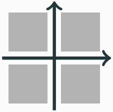
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...



perform degraded but low-cost pruning test

BnB – The secrete sauce

Solve time

$\underbrace{\text{region processing time}} \times \underbrace{\text{number of regions processed}}$

Relaxation for region ν

$$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

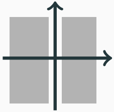
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

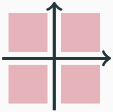
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...



perform degraded but low-cost pruning test

BnB – The secrete sauce

Solve time

$$\underbrace{\text{region processing time}} \times \underbrace{\text{number of regions processed}}$$

Relaxation for region ν

$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

This is a convex sparse optimization problem

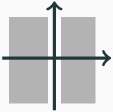
→ first-order methods

proximal gradient, coordinate descent, ...

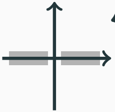
→ acceleration strategies

working set, screening tests, ...

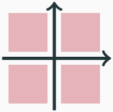
Simultaneous pruning



processing region ...



continue processing



perform degraded but low-cost pruning test

Solve time

$$\underbrace{\text{region processing time}} \times \underbrace{\text{number of regions processed}}$$

Relaxation for region ν

$$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

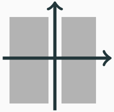
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

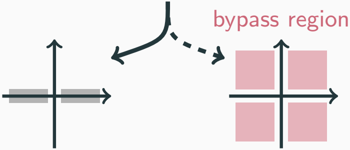
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...



continue processing

perform degraded but low-cost pruning test

BnB – Let's sum up

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Pipeline

1a) Implement a BnB solver

1b) Use an existing BnB solver

2) Solve the problem

BnB – Let's sum up

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Pipeline

1a) Implement a BnB solver

1b) Use an existing BnB solver

2) Solve the problem

Pros

✓ Numerical efficiency

✓ Open-source softwares

✓ Any h separable and coercive

BnB – Let's sum up

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Pipeline

1a) Implement a BnB solver

1b) Use an existing BnB solver

2) Solve the problem

Pros

✓ Numerical efficiency

✓ Open-source softwares

✓ Any h separable and coercive

Cons

✗ Less standard pipeline

Numerical Illustration

Numerics – Feature Selection Problem

Problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\mathbf{A} \in \mathbf{R}^{62\times 2000}$ from ML dataset / f : Logistic / h : ℓ_2-norm

Conclusion

A community with a bunch of people

Non-exhaustive list

A community with a bunch of people

Non-exhaustive list

MIT

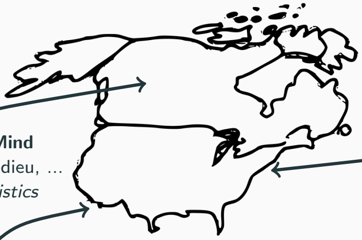
D. Bertsimas, R. Mazmuder, ...

Optimization tools for ℓ_0-problems

A community with a bunch of people

Non-exhaustive list

MIT
D. Bertsimas, R. Mazmuder, ...
Optimization tools for ℓ_0-problems



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

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

A community with a bunch of people

Non-exhaustive list

MIT
D. Bertsimas, R. Mazmuder, ...
Optimization tools for ℓ_0-problems

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

Lund University
M. Carlsson, C. Olsson...
Relaxation design

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

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

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

A community with a bunch of people

Non-exhaustive list

MIT
D. Bertsimas, R. Mazmuder, ...
Optimization tools for ℓ_0-problems

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

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

Lund University
M. Carlsson, C. Olsson...
Relaxation design

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, ML appli.

Centrale Nantes / ENSTA Bretagne
S. Bourguignon, J. Ninin, ...
BnB with quadratic loss

Inria / CentraleSupélec
C. Herzet, C. Elvira, A. Arslan, ...
Generic and efficient BnB

IRIT / I3S
E. Soubies, L. Blanc-Féraud, ...
Convex-concave relaxation

Take-home messages

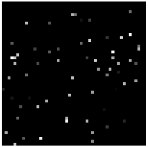
- Although NP-hard, ℓ_0-problems are of practical interest
- There exists methods to tackle them exactly
  - MIP-based formulation and off-the-shelf solvers
  - BnB-based specialized algorithms
  - Structure-exploitation is key for numerical efficiency
- It's an active research area
  - Theoretical and methodological developments still missing
  - Need to reach the application world

Question time !

Compressed sensing

Sparse signal

$\mathbf{x} \in \mathbf{R}^n$



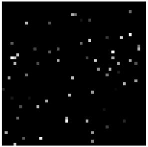
n pixels

Compressed sensing

Linear compression

Sparse signal

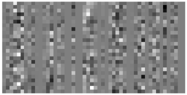
$\mathbf{x} \in \mathbf{R}^n$



n pixels

Compressed signal

$\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbf{R}^m$



m pixels, $m \ll n$

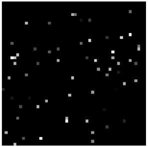
Compressed sensing

Sparse signal
$\mathbf{x} \in \mathbf{R}^n$

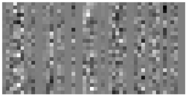
Linear compression

Recover $\mathbf{x}$ from $\mathbf{A}$ and $\mathbf{y}$

Compressed signal
$\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbf{R}^m$



n pixels



m pixels, $m \ll n$

Compressed sensing

Sparse signal

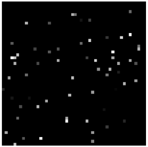
$\mathbf{x} \in \mathbf{R}^n$

Linear compression

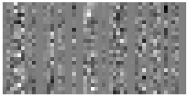
Recover $\mathbf{x}$ from $\mathbf{A}$ and $\mathbf{y}$

Compressed signal

$\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbf{R}^m$



n pixels



m pixels, $m \ll n$

Goal

Find $\mathbf{x}$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}$

Compressed sensing

Sparse signal

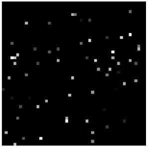
$\mathbf{x} \in \mathbf{R}^n$

Linear compression

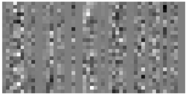
Recover $\mathbf{x}$ from $\mathbf{A}$ and $\mathbf{y}$

Compressed signal

$\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbf{R}^m$



n pixels



m pixels, $m \ll n$

Goal

Find $\mathbf{x}$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}$

no unique solution

Goal

Find $\mathbf{x}$ sparse such that $\mathbf{y} = \mathbf{A}\mathbf{x}$

Feature selection

	Feature 1	Feature 2	...	Feature n	Target
Sample 1	$a_{1,1}$	$a_{1,2}$	...	$a_{1,n}$	y_1
Sample 2	$a_{2,1}$	$a_{2,2}$	...	$a_{2,n}$	y_2
Sample 3	$a_{3,1}$	$\boxed{\mathbf{A} \in \mathbf{R}^{m \times n}}$	...	$a_{3,n}$	$\boxed{\mathbf{y} \in \mathbf{R}^m}$
...	...	...	...	...	...
Sample m	$a_{m,1}$	$a_{m,2}$	...	$a_{m,n}$	y_m

Feature selection

	Feature 1	Feature 2	...	Feature n	Target
Sample 1	$a_{1,1}$	$a_{1,2}$	$\cdots$	$a_{1,n}$	y_1
Sample 2	$a_{2,1}$	$a_{2,2}$	$\cdots$	$a_{2,n}$	y_2
Sample 3	$a_{3,1}$	$\boxed{\mathbf{A} \in \mathbf{R}^{m \times n}}$	$\cdots$	$a_{3,n}$	$\boxed{\mathbf{y} \in \mathbf{R}^{m}}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Sample m	$a_{m,1}$	$a_{m,2}$	$\cdots$	$a_{m,n}$	y_m

Features $\mathbf{A} \in \mathbf{R}^{m \times n}$ $\xleftrightarrow[\text{weights } \mathbf{x} \in \mathbf{R}^n]{}$ Target $\mathbf{y} = \phi(\mathbf{A}\mathbf{x})$

Feature selection

	Feature 1	Feature 2	...	Feature n	Target
Sample 1	$a_{1,1}$	$a_{1,2}$	$\cdots$	$a_{1,n}$	y_1
Sample 2	$a_{2,1}$	$a_{2,2}$	$\cdots$	$a_{2,n}$	y_2
Sample 3	$a_{3,1}$	$\boxed{\mathbf{A} \in \mathbf{R}^{m \times n}}$	$\cdots$	$a_{3,n}$	$\boxed{\mathbf{y} \in \mathbf{R}^{m}}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Sample m	$a_{m,1}$	$a_{m,2}$	$\cdots$	$a_{m,n}$	y_m

Features $\mathbf{A} \in \mathbf{R}^{m \times n}$ $\xleftrightarrow[\text{weights } \mathbf{x} \in \mathbf{R}^n]{}$ Target $\mathbf{y} = \phi(\mathbf{A}\mathbf{x})$

Model accuracy

Loss $\mathcal{L}_\phi(\mathbf{A}\mathbf{x}, \mathbf{y})$

Model explainability

Use few features

Feature selection

	Feature 1	Feature 2	...	Feature n	Target
Sample 1	$a_{1,1}$	$a_{1,2}$	$\cdots$	$a_{1,n}$	y_1
Sample 2	$a_{2,1}$	$a_{2,2}$	$\cdots$	$a_{2,n}$	y_2
Sample 3	$a_{3,1}$	$\boxed{\mathbf{A} \in \mathbf{R}^{m \times n}}$	$\cdots$	$a_{3,n}$	$\boxed{\mathbf{y} \in \mathbf{R}^{m}}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Sample m	$a_{m,1}$	$a_{m,2}$	$\cdots$	$a_{m,n}$	y_m

Features $\mathbf{A} \in \mathbf{R}^{m \times n}$ $\xleftrightarrow[\text{weights } \mathbf{x} \in \mathbf{R}^n]{}$ Target $\mathbf{y} = \phi(\mathbf{A}\mathbf{x})$

Model accuracy

Loss $\mathcal{L}_\phi(\mathbf{A}\mathbf{x}, \mathbf{y})$

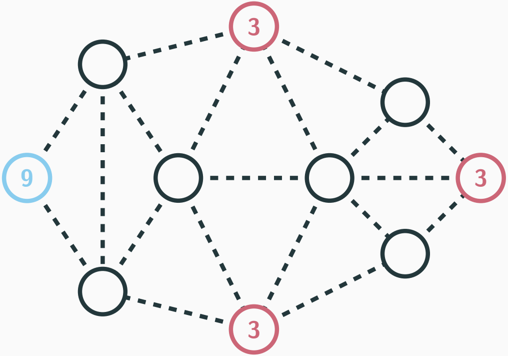
Model explainability

Use few features

Goal

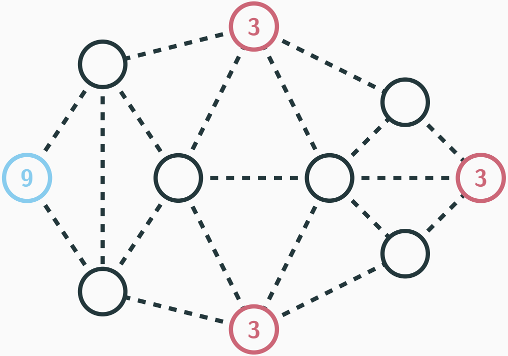
Find $\mathbf{x}$ sparse such that $\mathcal{L}_\phi(\mathbf{A}\mathbf{x}, \mathbf{y})$ is small

Network design



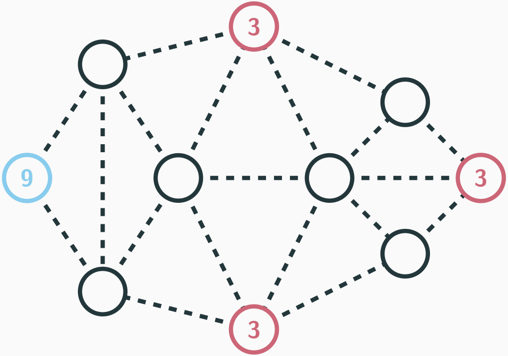
Which edges to build to transport products from source to sink nodes ?

Network design

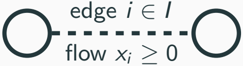


Which edges to build to transport products from source to sink nodes ?

Network design



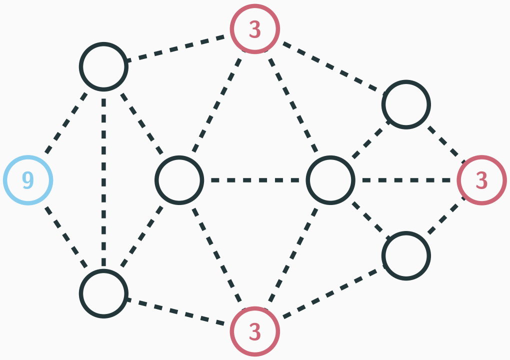
Which edges to build to transport products from source to sink nodes ?



construct edge $i \in I$ if $x_i > 0$

pay construction cost c

Network design



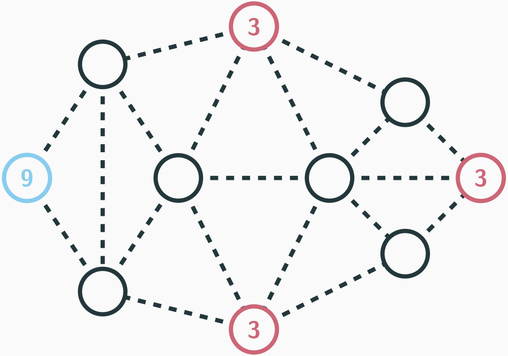
Which edges to build to transport products from source to sink nodes ?

construct edge $i \in I$ if $x_i > 0$

pay construction cost c

Question

How to construct the least number of edges to satisfy transportation needs ?



Which edges to build to transport products from source to sink nodes ?

construct edge $i \in I$ if $x_i > 0$

pay construction cost c

Question

How to construct the least number of edges to satisfy transportation needs ?

$\updownarrow$

Find $\mathbf{x} \in \mathbf{R}^{\mathrm{card}(I)}$ sparse such that $\mathcal{Q}(\mathbf{x}) = 0$

Balancing solution quality and problem hardness

Riboflavin dataset - P. Bühlmann *et al.* (2014)

Colony	AADK	AAPA	ABFA	ABH	...	ZUR	**B2 prod.**
#1	8.49	8.11	8.32	10.28	...	7.42	**-6.64**
#2	7.29	6.39	11.32	9.42	...	6.99	**-5.43**
...	...	...	...	...	...	...	...
#71	6.85	8.27	7.98	8.04	...	6.65	**-7.58**

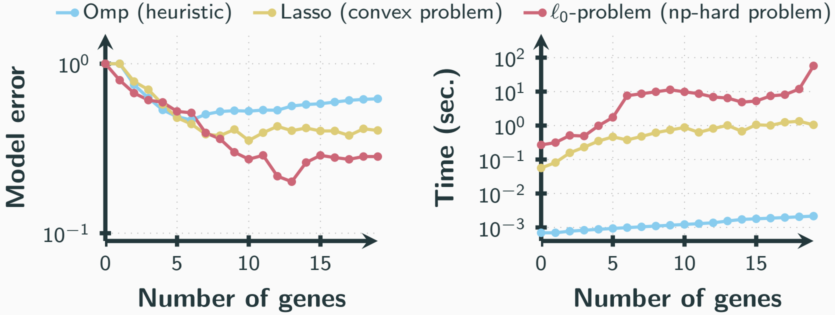
4,088 genes

Balancing solution quality and problem hardness

Riboflavin dataset - P. Bühlmann *et al.* (2014)

Colony	AADK	AAPA	ABFA	ABH	...	ZUR	**B2 prod.**
#1	8.49	8.11	8.32	10.28	...	7.42	**-6.64**
#2	7.29	6.39	11.32	9.42	...	6.99	**-5.43**
...	...	...	...	...	...	...	...
#71	6.85	8.27	7.98	8.04	...	6.65	**-7.58**

4,088 genes



MIP – Hands-on with `cvxpy`

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

$\downarrow$

Optimization problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

$\downarrow$

Optimization problem

$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

$\downarrow$

MIP formulation

$$\begin{cases} \min \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \mathbf{1}^{\mathrm{T}}\mathbf{z} \\ \text{s.t. } -M\mathbf{z} \leq \mathbf{x} \leq M\mathbf{z} \\ \qquad \mathbf{x} \in \mathbf{R}^n,\ \mathbf{z} \in \{0,1\}^n \end{cases}$$

MIP – Hands-on with `cvxpy`

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

$\downarrow$

Optimization problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

$\downarrow$

MIP formulation

$$\begin{cases} \min \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \mathbf{1}^{\mathrm{T}}\mathbf{z} \\ \text{s.t. } -M\mathbf{z} \leq \mathbf{x} \leq M\mathbf{z} \\ \qquad \mathbf{x} \in \mathbf{R}^n, \ \mathbf{z} \in \{0,1\}^n \end{cases}$$

```
$ pip install cvxpy
```

MIP – Hands-on with cvxpy

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

$\downarrow$

Optimization problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

$\downarrow$

MIP formulation

$$\begin{cases} \min \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\mathbf{1}^{\mathrm{T}}\mathbf{z} \\ \text{s.t. } -M\mathbf{z} \leq \mathbf{x} \leq M\mathbf{z} \\ \qquad \mathbf{x} \in \mathbf{R}^n,\ \mathbf{z} \in \{0,1\}^n \end{cases}$$

```
$ pip install cvxpy
```

```python
import cvxpy as cp

# Generate sparse regression data
A, y = make_regression()
```

MIP – Hands-on with cvxpy

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

$\downarrow$

Optimization problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

$\downarrow$

MIP formulation

$$\begin{cases} \min \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\mathbf{1}^{\mathrm{T}}\mathbf{z} \\ \text{s.t. } -M\mathbf{z} \leq \mathbf{x} \leq M\mathbf{z} \\ \quad \mathbf{x} \in \mathbf{R}^n,\ \mathbf{z} \in \{0,1\}^n \end{cases}$$

```
$ pip install cvxpy
```

```python
import cvxpy as cp

# Generate sparse regression data
A, y = make_regression()

# Define variables
n = A.shape[1]
x = cp.Variable(n)
z = cp.Variable(n, boolean=True)
```

MIP – Hands-on with cvxpy

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

$\downarrow$

Optimization problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

$\downarrow$

MIP formulation

$$\begin{cases} \min \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\mathbf{1}^{\mathrm{T}}\mathbf{z} \\ \text{s.t. } -M\mathbf{z} \leq \mathbf{x} \leq M\mathbf{z} \\ \qquad \mathbf{x} \in \mathbf{R}^n,\ \mathbf{z} \in \{0,1\}^n \end{cases}$$

```
$ pip install cvxpy
```

```
import cvxpy as cp

# Generate sparse regression data
A, y = make_regression()

# Define variables
n = A.shape[1]
x = cp.Variable(n)
z = cp.Variable(n, boolean=True)

# Define objective and constraints
obj = cp.Minimize(
    cp.sum_squares(A @ x - y) +
    0.01 * cp.sum(z)
)
cst = [-5.0 * z <= x, x <= 5.0 * z]
```

MIP – Hands-on with `cvxpy`

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

$\downarrow$

Optimization problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

$\downarrow$

MIP formulation

$$\begin{cases} \min \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \mathbf{1}^{\mathrm{T}}\mathbf{z} \\ \text{s.t. } -M\mathbf{z} \leq \mathbf{x} \leq M\mathbf{z} \\ \quad \mathbf{x} \in \mathbf{R}^n,\ \mathbf{z} \in \{0,1\}^n \end{cases}$$

```
$ pip install cvxpy
```

```
import cvxpy as cp

# Generate sparse regression data
A, y = make_regression()

# Define variables
n = A.shape[1]
x = cp.Variable(n)
z = cp.Variable(n, boolean=True)

# Define objective and constraints
obj = cp.Minimize(
    cp.sum_squares(A @ x - y) +
    0.01 * cp.sum(z)
)
cst = [-5.0 * z <= x, x <= 5.0 * z]

# Solve the problem using Gurobi
problem = cp.Problem(obj, cst)
problem.solve(solver=cp.GUROBI)
```

BnB – Hands-on with `el0ps`

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

BnB – Hands-on with `el0ps`

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

$$\downarrow$$

Optimization problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

BnB – Hands-on with `el0ps`

```
$ pip install el0ps
```

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

$\downarrow$

Optimization problem

$$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

```
$ pip install el0ps
```

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{Ax}$

$\downarrow$

Optimization problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{Ax}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

```python
from el0ps.datafits import Leastsquares
from el0ps.penalties import Bigm
from el0ps.solvers import BnbSolver

# Generate sparse regression data
A, y = make_regression()
```

BnB – Hands-on with `el0ps`

```
$ pip install el0ps
```

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

$\downarrow$

Optimization problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

```python
from el0ps.datafits import Leastsquares
from el0ps.penalties import Bigm
from el0ps.solvers import BnbSolver

# Generate sparse regression data
A, y = make_regression()

# Instantiate the loss and penalty
f = Leastsquares(A, y)
h = Bigm(M=5.0)
```

BnB – Hands-on with el0ps

```
$ pip install el0ps
```

Sparse regression

Find $\mathbf{x}$ sparse such that $\mathbf{y} \simeq \mathbf{Ax}$

$\downarrow$

Optimization problem

$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{Ax}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

```python
from el0ps.datafits import Leastsquares
from el0ps.penalties import Bigm
from el0ps.solvers import BnbSolver

# Generate sparse regression data
A, y = make_regression()

# Instantiate the loss and penalty
f = Leastsquares(A, y)
h = Bigm(M=5.0)

# Solve the problem with el0ps' solver
solver = BnbSolver()
solver.solve(f, h, lmbd=0.01)
```