Optimization methods for ℓ_0-problems

Théo Guyard

ML MTP – December 9th, 2024

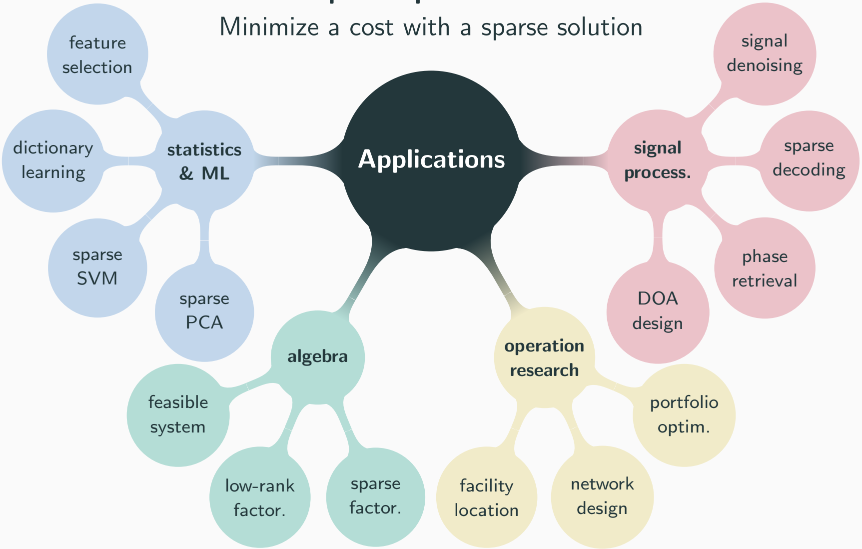
Sparse optimization

Minimize a cost with a sparse solution

Sparse optimization

Sparse optimization

Minimize a cost with a sparse solution

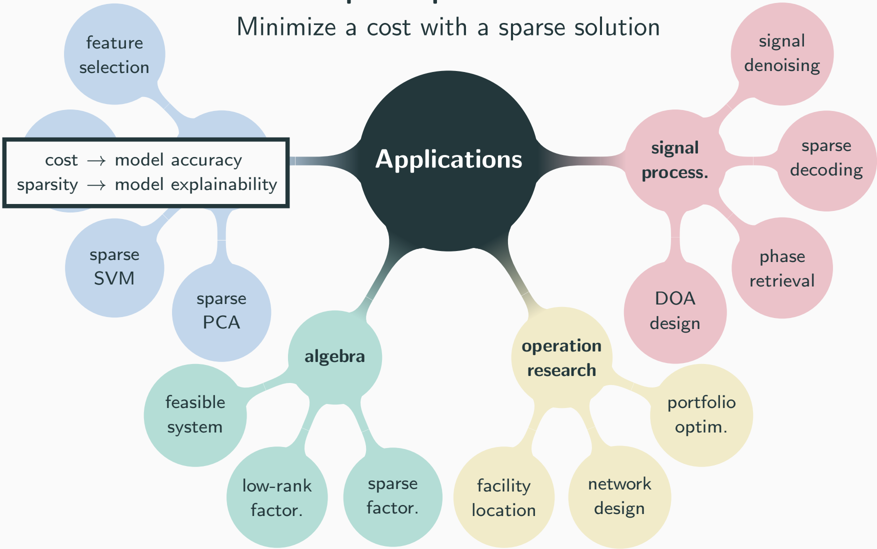


A. Tillmann *et. al* (2024)

Sparse optimization

Sparse optimization

Minimize a cost with a sparse solution

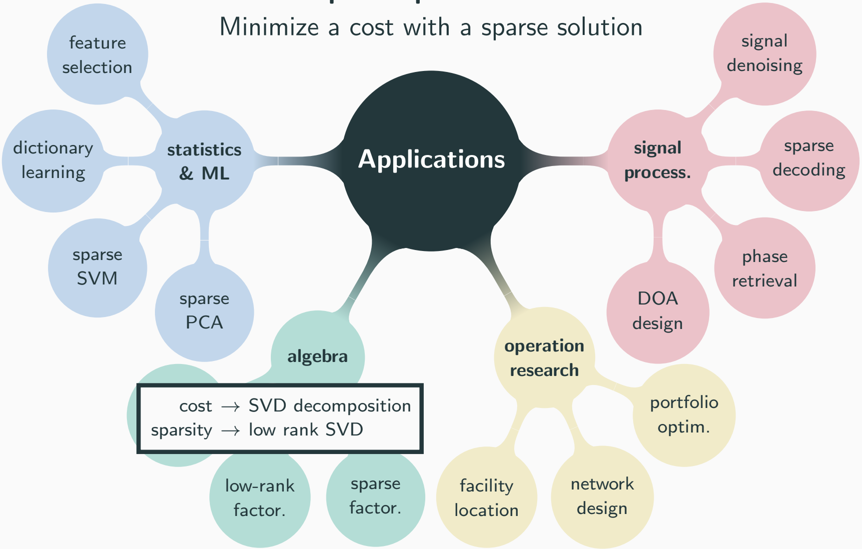


cost $\longrightarrow$ model accuracy
sparsity $\longrightarrow$ model explainability

A. Tillmann *et. al* (2024)

Sparse optimization

Minimize a cost with a sparse solution

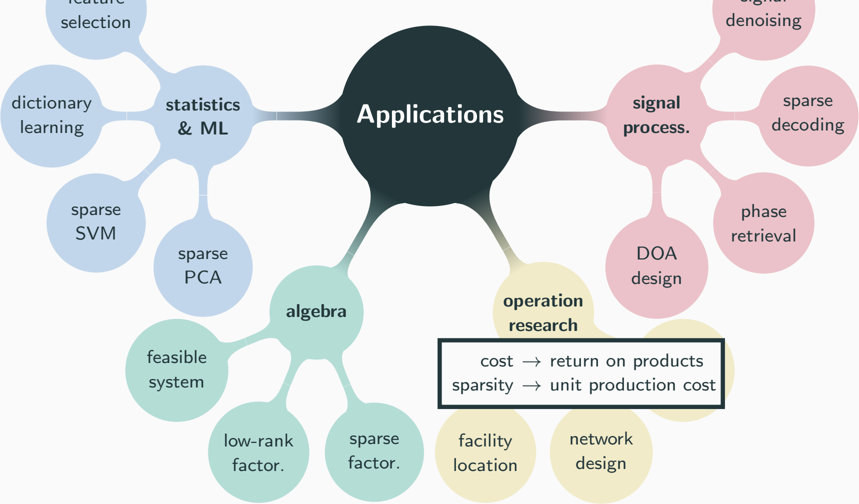


A. Tillmann *et. al* (2024)

Sparse optimization

Sparse optimization

Minimize a cost with a sparse solution

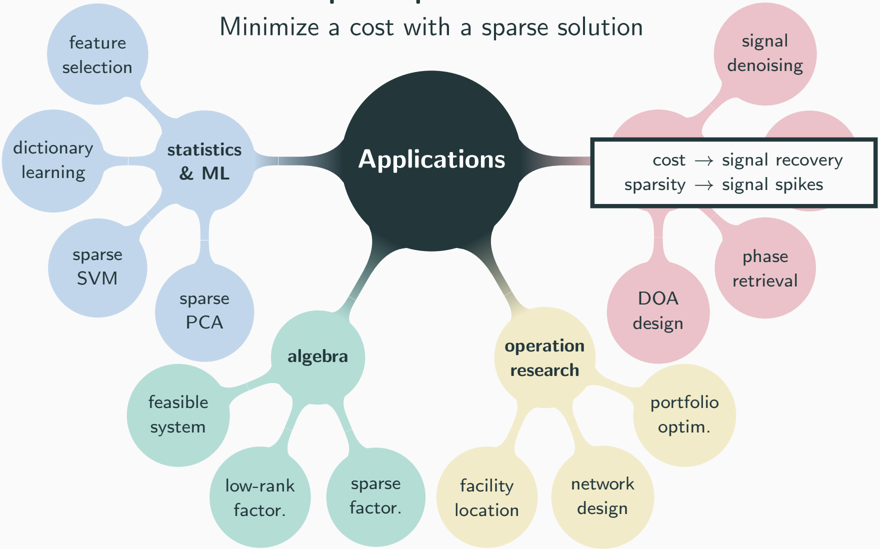


A. Tillmann *et. al* (2024)

Sparse optimization

Sparse optimization

Minimize a cost with a sparse solution



A. Tillmann *et. al* (2024)

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a cost with a sparse solution

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a cost with a sparse solution

quantify cost

$f(\mathbf{x})$

thing to minimize

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a cost with a sparse solution

quantify cost

$f(\mathbf{x})$

thing to minimize

quantify sparsity

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

counts non-zeros

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a cost with a sparse solution

quantify cost

quantify sparsity

$f(\mathbf{x})$

thing to minimize

optim. problem

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

counts non-zeros

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a cost with a sparse solution

quantify cost

quantify sparsity

$f(\mathbf{x})$

thing to minimize

optim. problem

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

counts non-zeros

Constrained version

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

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a cost with a sparse solution

quantify cost

quantify sparsity

$f(\mathbf{x})$

thing to minimize

optim. problem

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

counts non-zeros

Constrained version

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

Minimized version

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

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a cost with a sparse solution

quantify cost

quantify sparsity

$f(\mathbf{x})$

thing to minimize

optim. problem

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

counts non-zeros

Constrained version

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

Minimized version

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

Regularized version

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

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a cost with a sparse solution

quantify cost

quantify sparsity

$f(\mathbf{x})$

thing to minimize

optim. problem

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

counts non-zeros

Constrained version

$$\min_{\mathbf{x}\in\mathbf{R}^n} \quad f(\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{x}\|_0 \leq s$$

Minimized version

$$\min_{\mathbf{x}\in\mathbf{R}^n} \quad \|\mathbf{x}\|_0 \quad \text{subject to} \quad f(\mathbf{x}) \leq \epsilon$$

Regularized version

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

$+\ h(\mathbf{x})$ separable

A bit of history

Problem

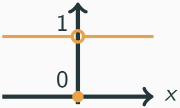
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

A bit of history

Problem

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

1d graph

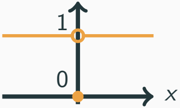


A bit of history

Problem

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

1d graph



non-convex, non-diff., non-continuous, ...

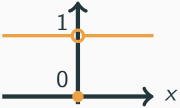
A bit of history

Problem

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

NP-hard to solve

1d graph



non-convex, non-diff.,
non-continuous, ...

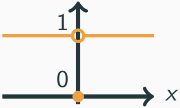
A bit of history

Problem

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

NP-hard to solve

1d graph

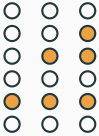


non-convex, non-diff.,
non-continuous, ...

1995

Heuristics

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



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

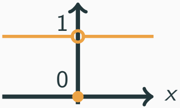
A bit of history

Problem

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

NP-hard to solve

1d graph

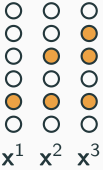


non-convex, non-diff.,
non-continuous, ...

1995 | **2000**

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

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



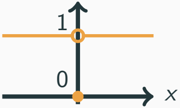
OMP solves
ℓ_0-problem
under RIP

Problem

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

NP-hard to solve

1d graph

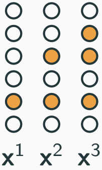


non-convex, non-diff., non-continuous, ...

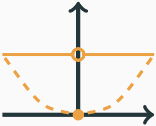
1995 — Heuristics
MP, OMP, ...
S. Mallat (1993)

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

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



OMP solves
ℓ_0-problem
under RIP



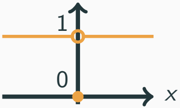
A bit of history

Problem

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

NP-hard to solve

1d graph



non-convex, non-diff.,
non-continuous, ...

1995

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

2000

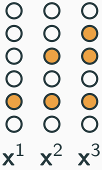
Recovery cond.
RIP, NSP, ...
E. Candes (2004)

2005

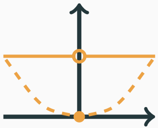
Convex approx.
Lasso, Elastic-Net, ...
R. Tibshirani (2005)

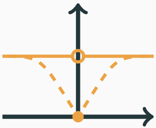
2010

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



OMP solves
ℓ_0-problem
under RIP





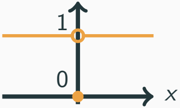
A bit of history

Problem

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

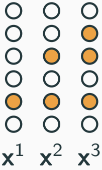
NP-hard to solve

1d graph

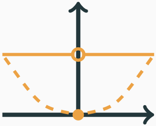


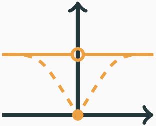
non-convex, non-diff.,
non-continuous, ...

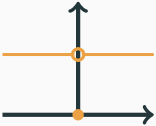
1995	2000	2005	2010	2015
Heuristics	Recovery cond.	Convex approx.	Concave approx.	Exact methods
MP, OMP, ...	RIP, NSP, ...	Lasso, Elastic-Net, ...	SCAD, MCP, ...	MIP, BnB, ...
S. Mallat (1993)	E. Candes (2004)	R. Tibshirani (2005)	C. Zhang (2010)	D. Bertsimas (2016)



OMP solves
ℓ_0-problem
under RIP







Topic of this talk

Ok, ℓ_0-problems can be of practical interest !

Topic of this talk

Ok, ℓ_0-problems can be of practical interest !

🤓

What are the solution methods then ?

🤔

Topic of this talk

Ok, ℓ_0-problems can be of practical interest !

🤓

What are the solution methods then ?

🤔

1) MIP-based methods

Convenient for practitioners

Poor numerical performances

Topic of this talk

Ok, ℓ_0-problems can be of practical interest !

What are the solution methods then ?

1) MIP-based methods

Convenient for practitioners

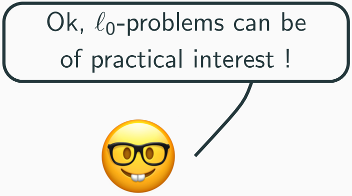
Poor numerical performances

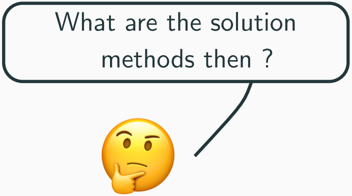
2) Specialized Branch-and-Bound

More sophisticated mechanism

Better numerical performances

Topic of this talk





1) MIP-based methods

Convenient for practitioners

Poor numerical performances

2) Specialized Branch-and-Bound

More sophisticated mechanism

Better numerical performances

High-level concepts and practical tools

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, ...

↓

Problem

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

↓

MIP formulation

Standardized expressions

↓

Problem solutions

Modelling framework

Python $\rightarrow$ `cvxpy`

Julia $\rightarrow$ `JuMP`

C++ $\rightarrow$ `CMPL`

Matlab $\rightarrow$ `YALMIP`

...

MIP – Pipeline

Application

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

$\downarrow$

Problem

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

$\downarrow$

Modelling framework

Python $\rightarrow$ `cvxpy`

Julia $\rightarrow$ `JuMP`

C++ $\rightarrow$ `CMPL`

Matlab $\rightarrow$ `YALMIP`

...

MIP formulation

Standardized expressions

$\downarrow$

Generic MIP 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 ℓ_0-norm

$\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$ when

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

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

MIP – Formulation

Problem

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

$\downarrow$ linearize ℓ_0-norm $\downarrow$

Linearized formulation

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

MIP formulation

Use standardized expressions

linear, quadratic, conic, ...

Linearize ℓ_0-norm

$\|\mathbf{x}\|_0 = \mathbf{1}^\mathrm{T} \mathbf{z}$ when

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

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

MIP – Formulation

Problem

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

linearize ℓ_0-norm

Linearized formulation

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

MIP formulation

Use standardized expressions

linear, quadratic, conic, ...

Linearize ℓ_0-norm

$\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$ when

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

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

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

Avoid logical constraint

$$h_{\mathrm{mip}}(\mathbf{x},\mathbf{z}) = \begin{cases} h(\mathbf{x}) & \text{if } x_i = 0 \iff z_i = 0 \\ +\infty & \text{otherwise} \end{cases}$$

MIP – Formulation

Problem

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

linearize ℓ_0-norm

Linearized formulation

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

avoid logical cstr.

MIP formulation

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

MIP formulation

Use standardized expressions

linear, quadratic, conic, ...

Linearize ℓ_0-norm

$\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$ when

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

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

Avoid logical constraint

$$h_{\mathrm{mip}}(\mathbf{x},\mathbf{z}) = \begin{cases} h(\mathbf{x}) & \text{if } x_i = 0 \iff z_i = 0 \\ +\infty & \text{otherwise} \end{cases}$$

Sparse regression

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

Sparse regression

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

$\downarrow$

Optimization problem

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

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

Sparse regression

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

$$\downarrow$$

Optimization problem

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

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

$$\downarrow$$

MIP formulation

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

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
```

```
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 $\mathbf{y} \simeq \mathbf{Ax}$

$\downarrow$

Optimization problem

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

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

$\downarrow$

MIP formulation

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

```
$ 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]

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

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

2) Establish MIP formulation

3) 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

2) Establish MIP formulation

3) Use generic MIP solvers

Pros

✓ Rich MIP literature

✓ Black-box solvers

✓ Convenient for practitioners

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

2) Establish MIP formulation

3) Use generic MIP solvers

Pros

✓ Rich MIP literature

✓ Black-box solvers

✓ Convenient for practitioners

Cons

✗ Mostly commercial solvers

✗ Unable to exploit structure

✗ 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

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

$\downarrow$

Design BnB solver

Specialized mechanisms

BnB – Pipeline

Application

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

$\downarrow$

Problem

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

$\downarrow$

Design BnB solver

Specialized mechanisms

$\downarrow$

Problem solutions

BnB – Pipeline

Application

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

↓

Problem

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

↓

~~**Design 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)

BnB – Pipeline

Application

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

↓

Problem

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

↓

~~**Design 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 fast and flexible !

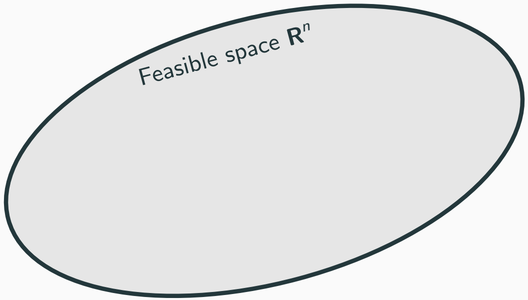
😎

BnB – Algorithmic principle

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

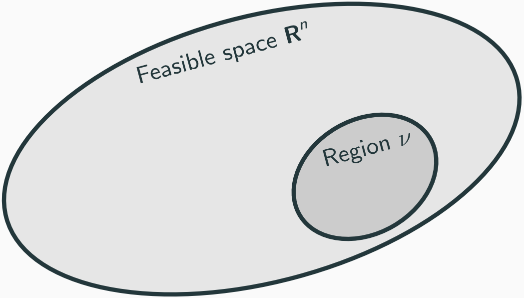
BnB – Algorithmic principle

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



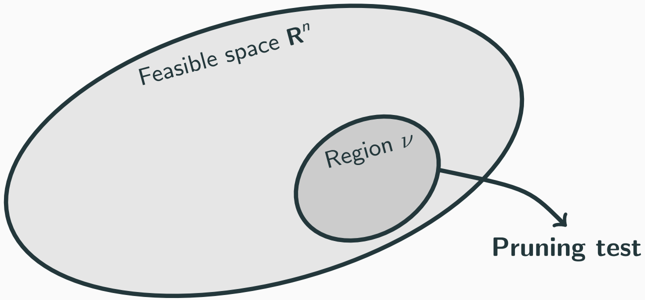
BnB – Algorithmic principle

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



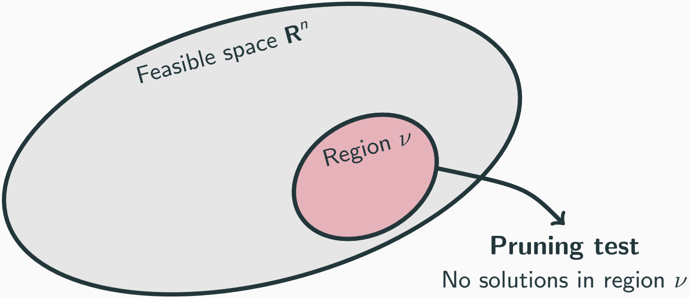
BnB – Algorithmic principle

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

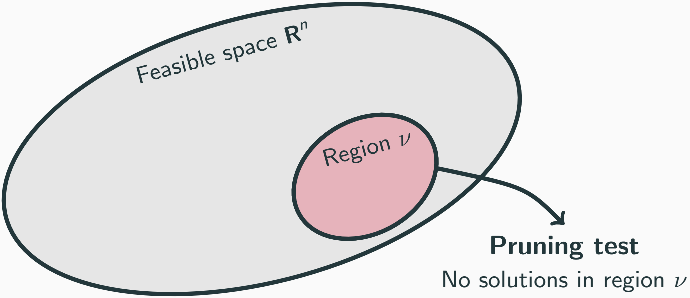


BnB – Algorithmic principle

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



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



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

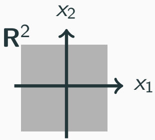
BnB – Branching step

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

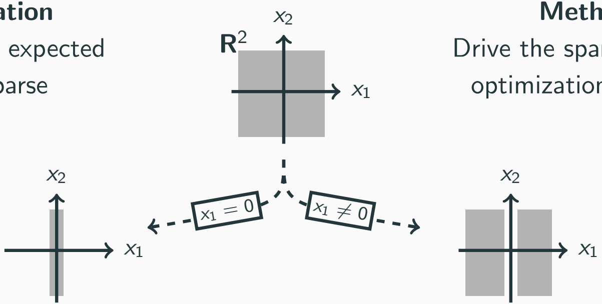
BnB – Branching step

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

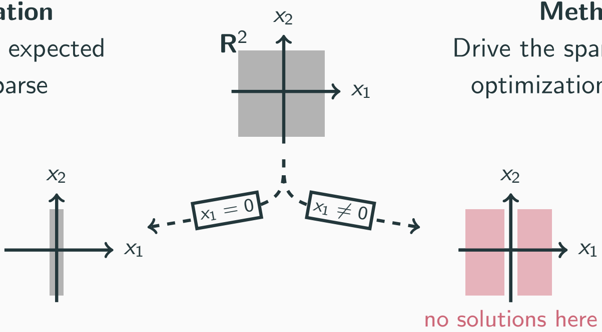
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



Problem

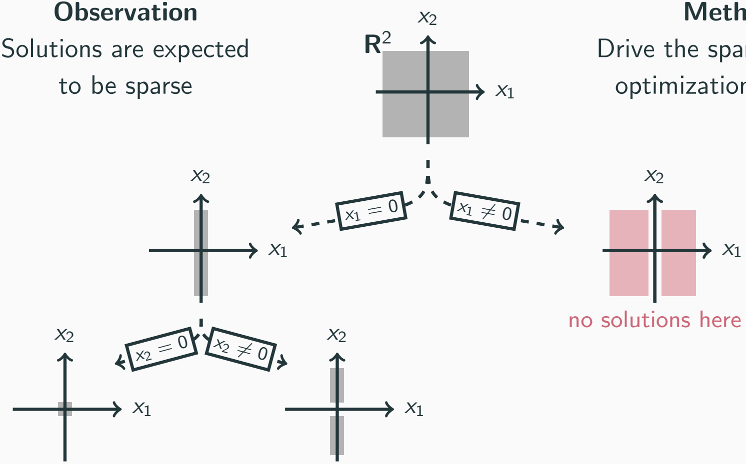
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



Problem

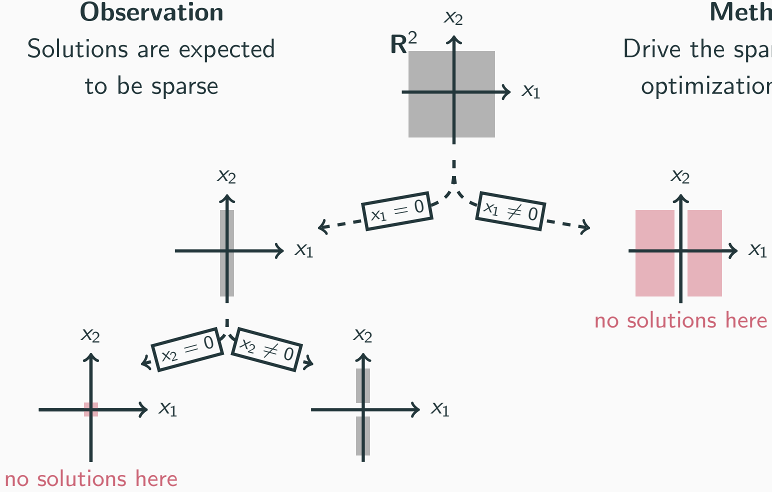
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



Problem

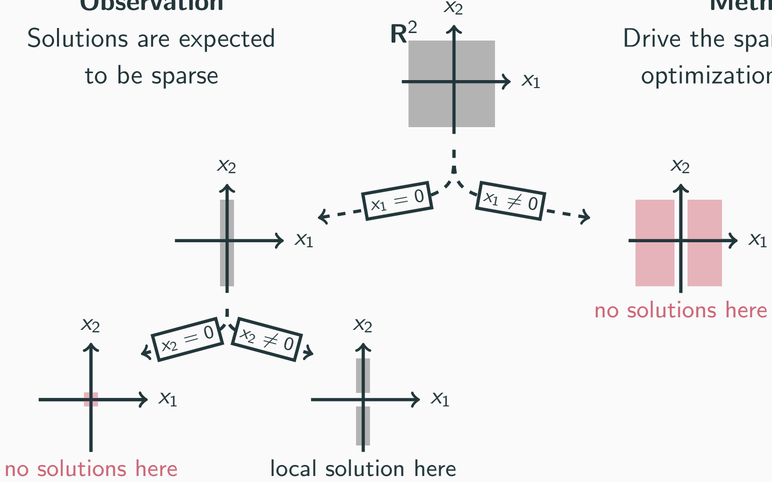
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



BnB – Branching step

Problem

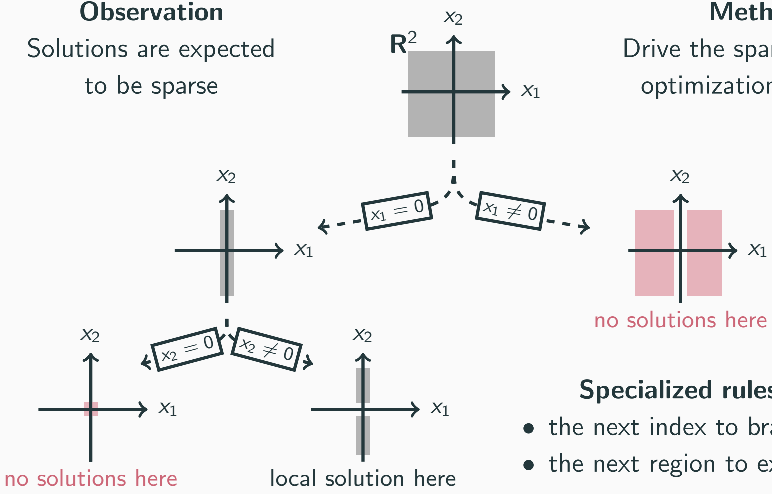
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

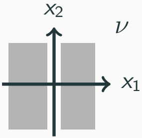
Drive the sparsity of the optimization variable



Specialized rules for

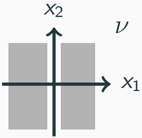
- the next index to branch on
- the next region to explore

BnB – Bounding step



Does region ν contains optimal solutions ?

BnB – Bounding step

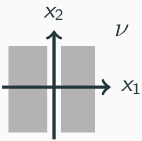


Does region ν contains optimal solutions ?

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

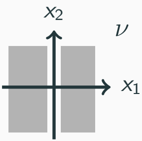
$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$ restrict to ν

Restriction to region ν

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

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

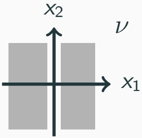
$$p^\nu = \min_{\mathbf{x}\in\nu} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

$\downarrow$ 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

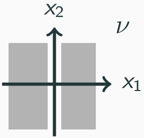
$p^\nu = \min_{\mathbf{x} \in \nu} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

compare

Pruning test

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

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

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

restrict to ν

Restriction to region ν

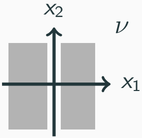
$p^\nu = \min_{\mathbf{x}\in\nu} \; f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

compare

Pruning test

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

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

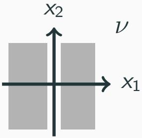
Pruning test

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

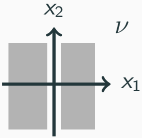
$p^\nu_{\text{lb}} > p^\star_{\text{ub}}$ $\longrightarrow$ prune ν

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

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

$\downarrow$ 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

Easy task

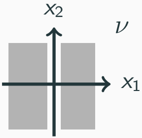
$\boxed{\text{Compute an upper bound on } p^\star}$

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

Main challenge

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

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

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

Easy task

Compute an upper bound on $p^\star$

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

Main challenge

Compute a lower bound on p^ν

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

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

Relaxation for region ν

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

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

BnB – The full picture

Problem

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

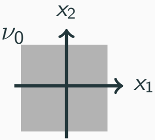
BnB – The full picture

Problem

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

Best upper bound

$p^\star_{\mathrm{ub}} = +\infty$



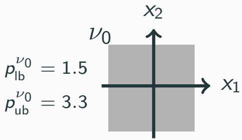
BnB – The full picture

Problem

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

Best upper bound

$p^\star_{\mathrm{ub}} = +\infty$



BnB – The full picture

Problem

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

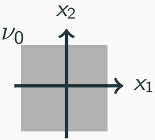
Best upper bound

$\cancel{p^\star_{\text{ub}} = +\infty}$

$p^\star_{\text{ub}} = 3.3$

$p^{\nu_0}_{\text{lb}} = 1.5$

$p^{\nu_0}_{\text{ub}} = 3.3$



BnB – The full picture

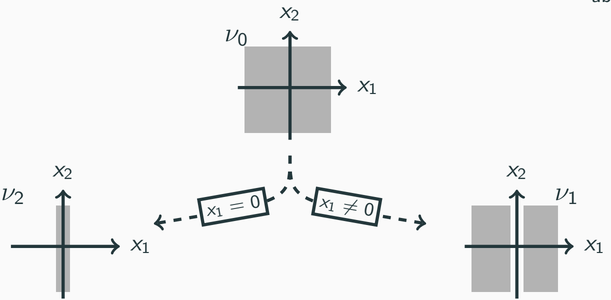
Problem

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

Best upper bound

$\cancel{p^\star_{\mathrm{ub}} = +\infty}$

$p^\star_{\mathrm{ub}} = 3.3$



BnB – The full picture

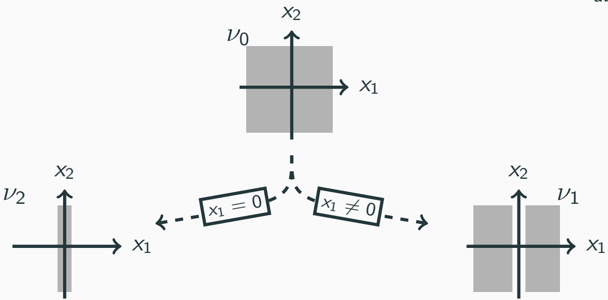
Problem

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

Best upper bound

$\cancel{p^\star_{\mathrm{ub}} = +\infty}$

$p^\star_{\mathrm{ub}} = 3.3$



$p^{\nu_1}_{\mathrm{lb}} = 4.5$

BnB – The full picture

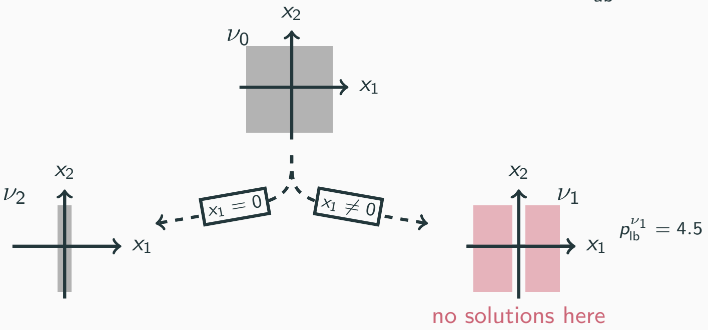
Problem

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

Best upper bound

$\cancel{p^\star_{\mathrm{ub}} = +\infty}$

$p^\star_{\mathrm{ub}} = 3.3$



BnB – The full picture

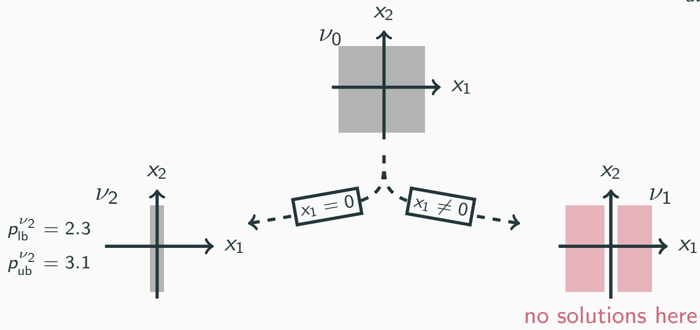
Problem

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

Best upper bound

$\cancel{p^\star_{\text{ub}} = +\infty}$

$p^\star_{\text{ub}} = 3.3$



BnB – The full picture

Problem

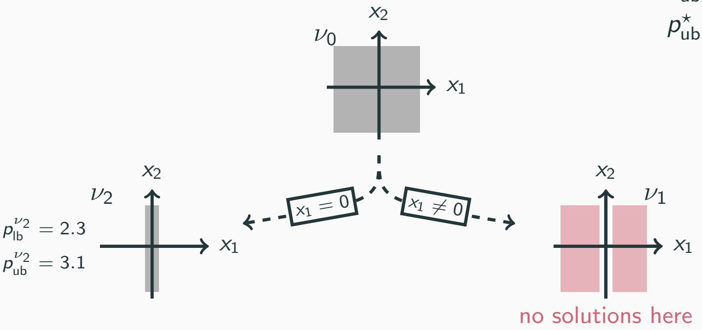
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Best upper bound

$\cancel{p^\star_{\text{ub}} = +\infty}$

$\cancel{p^\star_{\text{ub}} = 3.3}$

$p^\star_{\text{ub}} = 3.1$



BnB – The full picture

Problem

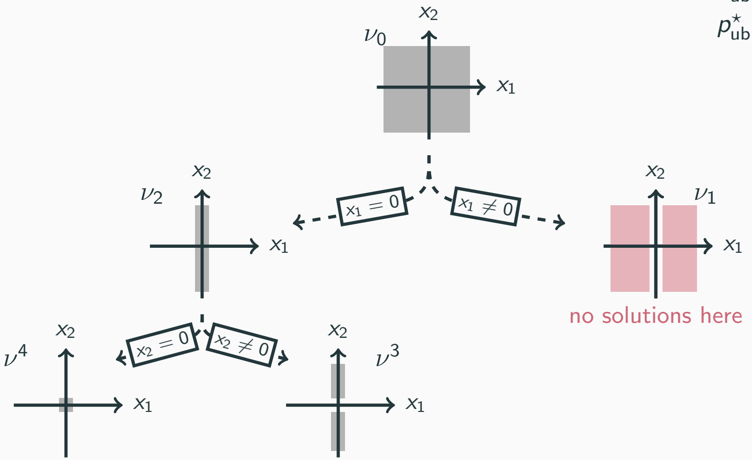
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Best upper bound

$\cancel{p^\star_{\text{ub}} = +\infty}$

$\cancel{p^\star_{\text{ub}} = 3.3}$

$p^\star_{\text{ub}} = 3.1$



BnB – The full picture

Problem

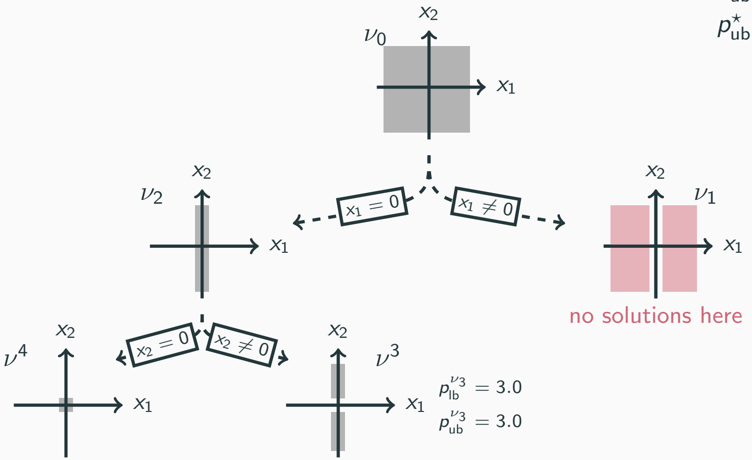
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Best upper bound

$\cancel{p^\star_{\text{ub}} = +\infty}$

$\cancel{p^\star_{\text{ub}} = 3.3}$

$p^\star_{\text{ub}} = 3.1$



$p^{\nu_3}_{\text{lb}} = 3.0$

$p^{\nu_3}_{\text{ub}} = 3.0$

BnB – The full picture

Problem

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

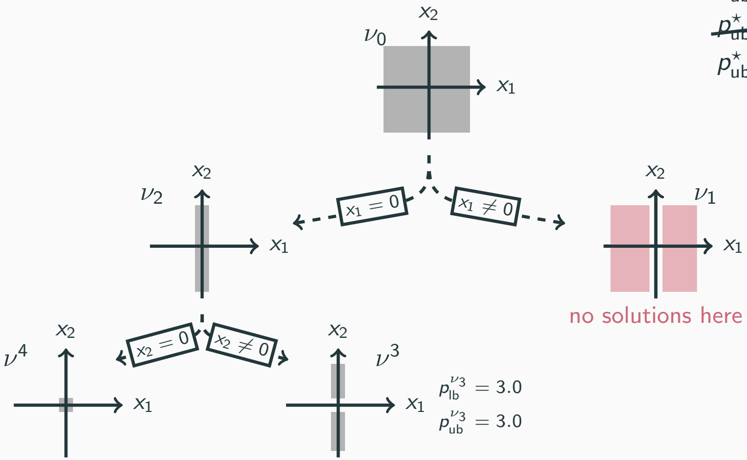
Best upper bound

$\cancel{p^\star_{\text{ub}} = +\infty}$

$\cancel{p^\star_{\text{ub}} = 3.3}$

$\cancel{p^\star_{\text{ub}} = 3.1}$

$p^\star_{\text{ub}} = 3.0$



BnB – The full picture

Problem

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

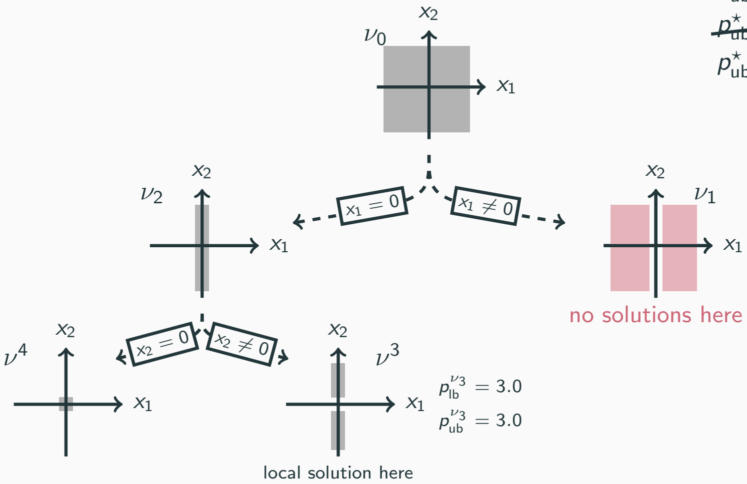
Best upper bound

$\cancel{p^\star_{\text{ub}} = +\infty}$

$\cancel{p^\star_{\text{ub}} = 3.3}$

$\cancel{p^\star_{\text{ub}} = 3.1}$

$p^\star_{\text{ub}} = 3.0$



BnB – The full picture

Problem

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

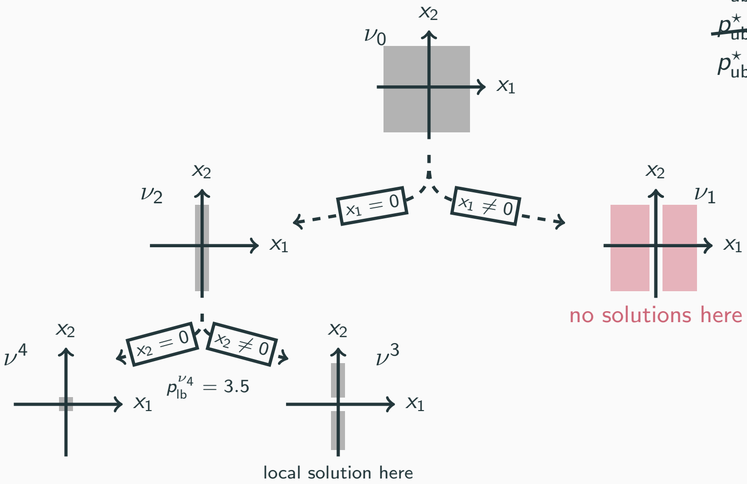
Best upper bound

~~$p^\star_{\mathrm{ub}} = +\infty$~~

~~$p^\star_{\mathrm{ub}} = 3.3$~~

~~$p^\star_{\mathrm{ub}} = 3.1$~~

$p^\star_{\mathrm{ub}} = 3.0$



BnB – The full picture

Problem

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

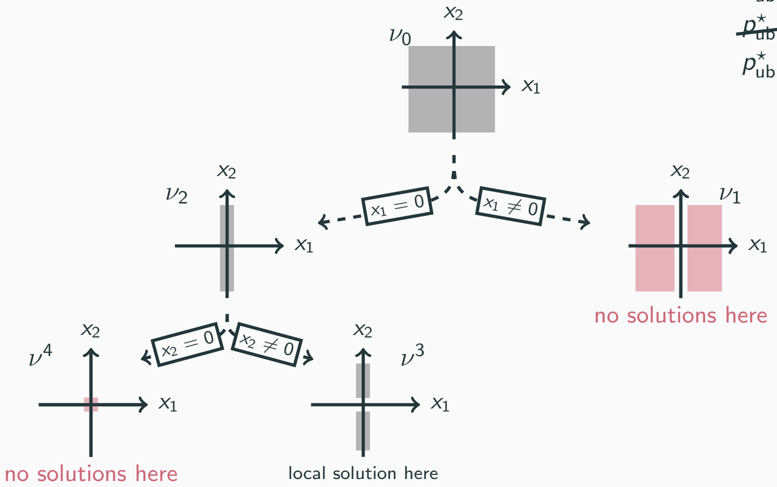
Best upper bound

$\cancel{p^\star_{\text{ub}} = +\infty}$

$\cancel{p^\star_{\text{ub}} = 3.3}$

$\cancel{p^\star_{\text{ub}} = 3.1}$

$p^\star_{\text{ub}} = 3.0$



BnB – The secrete sauce

Solve time

region processing time $\times$ number of regions processed

BnB – The secrete sauce

Solve time

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

Relaxation for region ν

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

BnB – The secrete sauce

Solve time

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

Relaxation for region ν

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

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

BnB – The secrete sauce

Solve time

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

Relaxation for region ν

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

g_{lb}^{ν} is proper, closed, convex, separable, and non-diff. 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

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

g_{lb}^{ν} is proper, closed, convex, separable, and non-diff. 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

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

g^{ν}_{lb} is proper, closed, convex, separable, and non-diff. 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

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

Relaxation for region ν

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

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

$\downarrow$

This is a convex sparse optimization problem !

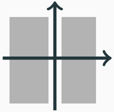
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...

BnB – The secrete sauce

Solve time

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

Relaxation for region ν

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

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

$\downarrow$

This is a convex sparse optimization problem !

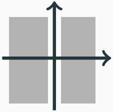
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

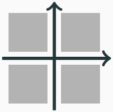
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...



perform degraded but low-cost pruning test

Solve time

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

Relaxation for region ν

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

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

$\downarrow$

This is a convex sparse optimization problem !

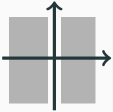
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

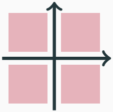
$\rightarrow$ acceleration strategies

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

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

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

$\downarrow$

This is a convex sparse optimization problem !

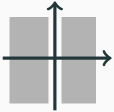
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

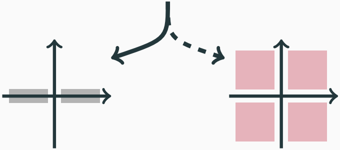
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...



continue processing

perform degraded but low-cost pruning test

BnB – The secrete sauce

Solve time

region processing time $\times$ number of regions processed

Relaxation for region ν

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

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

This is a convex sparse optimization problem !

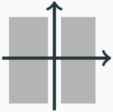
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

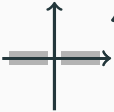
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning

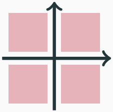


processing region ...



continue processing

bypass region



perform degraded but low-cost pruning test

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

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

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

```
$ 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 $\mathbf{y} \simeq \mathbf{A}\mathbf{x}$

$\downarrow$

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()
```

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)
```

```
$ pip install el0ps
```

Sparse regression

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

$$\downarrow$$

Optimization problem

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

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

```
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)
```

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 specialized BnB

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 specialized BnB

1b) Use an existing BnB solver

2) Solve the problem

Pros

✓ Numerical efficiency

✓ Open-source softwares

✓ Convenient for practitioners

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 specialized BnB

1b) Use an existing BnB solver

2) Solve the problem

Pros

✓ Numerical efficiency

✓ Open-source softwares

✓ Convenient for practitioners

Cons

✗ Assumptions on f/h

✗ f/h proper, closed, convex

✗ f smooth, h coercive

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

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

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

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

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

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

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

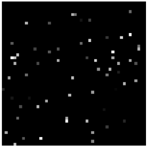
- Although NP-hard, ℓ_0-problems are of practical interest
- There exists methods to tackle them
  - MIP-based formulation and generic solvers
  - BnB algorithms with specialized steps
  - Structure-exploitation is key
- It's an active research area
  - Theoretical and methodological developments still needed
  - Need to reach the applicative world

Question time !

🤓

Compressed sensing

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



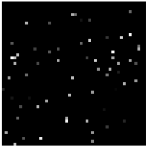
n pixels

Compressed sensing

Linear compression

Sparse signal

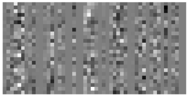
$\mathbf{x} \in \mathbf{R}^n$



n pixels

Compressed signal

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

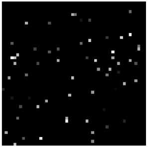


m pixels, $m \ll n$

Compressed sensing

Sparse signal

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



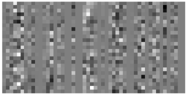
n pixels

Linear compression

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

Compressed signal

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



m pixels, $m \ll n$

Compressed sensing

Sparse signal

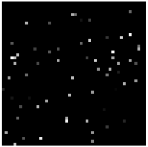
$\mathbf{x} \in \mathbf{R}^n$

Linear compression

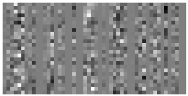
Recover $\mathbf{x}$ from $\mathbf{A}$ and $\mathbf{y}$

Compressed signal

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



n pixels



m pixels, $m \ll n$

Goal

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

Compressed sensing

Sparse signal

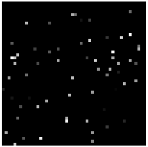
$\mathbf{x} \in \mathbf{R}^n$

Linear compression

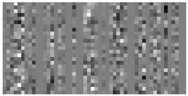
Recover $\mathbf{x}$ from $\mathbf{A}$ and $\mathbf{y}$

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 $\mathbf{y} = \phi(\mathbf{Ax})$

Model accuracy

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

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

$\downarrow$

Model explainability

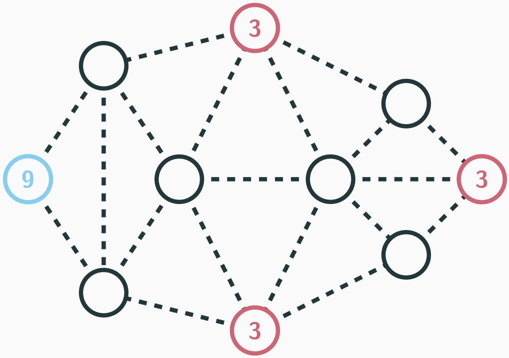
Use few features

$\downarrow$

Goal

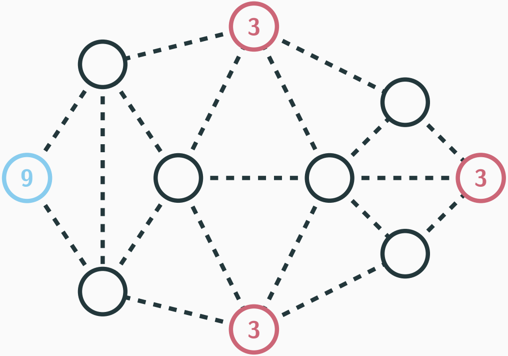
Find $\mathbf{x}$ sparse such that $\mathcal{L}_\phi(\mathbf{A}\mathbf{x}, \mathbf{y})$ is small

Network design



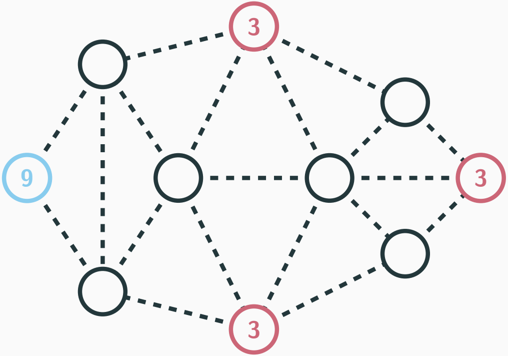
Which edges to build to transport products from source to sink nodes ?

Network design



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

Network design

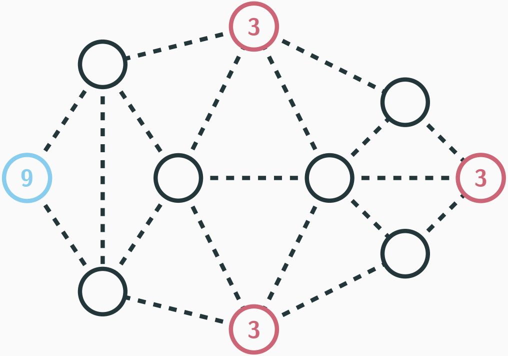


Which edges to build to transport products from source to sink nodes ?

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

pay construction cost c

Network design



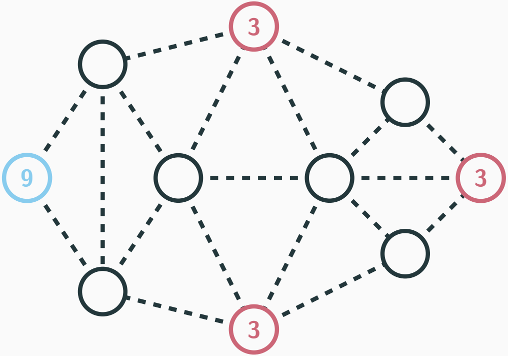
Which edges to build to transport products from source to sink nodes ?

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 ?

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**

$\underbrace{\hspace{6em}}_{\text{4,088 genes}}$