Peeling for L0-Regularized Least-Squares

Théo Guyard[⋆], Gilles Monnoyer[◇], Cédric Herzet[†] and Clément Elvira[‡]

[⋆]Inria, Rennes, France
[◇]UCLouvain, Louvain, Belgium
[†]Ensai, Rennes, France
[‡]CentraleSupelec, Rennes, France

PGMO days
EDF Lab, Saclay, France
November 29th, 2023

L0-Regularized Least-Squares

Framework

Sparse linear models

- Linear regression with a sparse optimizer
- Applications in signal processing, machine learning, statistics, etc...

Framework

Sparse linear models

- Linear regression with a sparse optimizer
- Applications in signal processing, machine learning, statistics, etc...

ℓ_0-regularized least-squares

$$\min_x \ \frac{1}{2}\|\mathrm{y} - \mathrm{A}\mathrm{x}\|_2^2 + \lambda\|x\|_0$$

Ingredients

- Least-squares loss to ensure the linear model fitting
- ℓ_0-norm that counts the non-zeros in x
- Tradeoff parameter $\lambda > 0$ to control the sparsity

Solving ℓ_0-regularized problems

Initial problem

$$\min_x \ \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0$$

Initial problem

$$\min_x \; \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0$$

Non-linear, non-convex,
non-smooth, NP-hard, ...

😵

Solving ℓ_0-regularized problems

Initial problem

$$\min_x \ \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0$$

Non-linear, non-convex, non-smooth, NP-hard, ...

Big-M method

Reformulation

$$\begin{aligned} \min_x \ & \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} \ & -M \leq x \leq M \end{aligned}$$

Solving ℓ_0-regularized problems

Initial problem

$$\min_x \ \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0$$

Non-linear, non-convex, non-smooth, NP-hard, ... 😵

$\boxed{\text{Big-M method}}$

Reformulation

$$\min_x \ \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0$$
$$\text{s.t.} \quad -M \leq x \leq M$$

Add binary variables to linearize the ℓ_0-norm

Solving ℓ_0-regularized problems

Initial problem

$$\min_{x} \ \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0$$

Non-linear, non-convex, non-smooth, NP-hard, ...

Big-M method

Add binary variables to linearize the $\boldsymbol{\ell_0}$-norm

Reformulation

$$\begin{aligned} \min_{x} \quad & \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} \quad & -M \leq x \leq M \end{aligned}$$

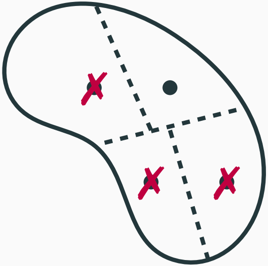
Generic Mixed-Integer Program solvers

Specialized Branch-and-Bound solvers

Branch-and-Bound

Branch-and-Bound

"Enumerate all candidate solutions and discard sub-optimal ones."



Main principles

Branching:	Divide the search space
Bounding:	Test whether a region can contain optimal solutions
Pruning:	Discard regions without optimal solutions

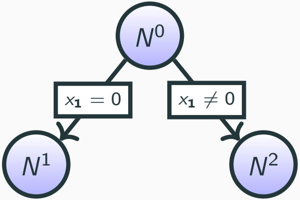
Tree search

$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$

N^0

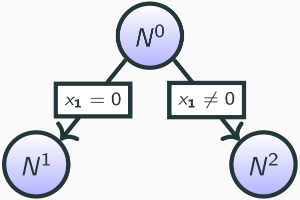
Tree search

$$\begin{cases} \min & \frac{1}{2}\|\mathrm{y} - \mathrm{A}\mathrm{x}\|_2^2 + \lambda\|\mathrm{x}\|_0 \\ \text{s.t.} & -M \leq \mathrm{x} \leq M \end{cases}$$



Tree search

$$\begin{cases} \min & \frac{1}{2}\|\mathrm{y} - \mathrm{Ax}\|_2^2 + \lambda\|\mathrm{x}\|_0 \\ \text{s.t.} & -M \leq \mathrm{x} \leq M \end{cases}$$

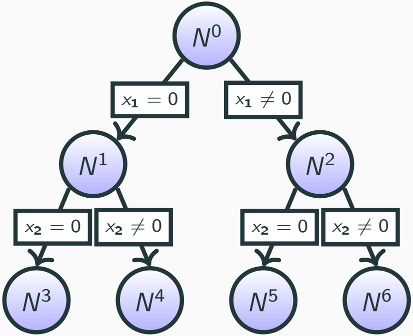


Same problem + constraint $x_1 = 0$

Same problem + constraint $x_1 \neq 0$

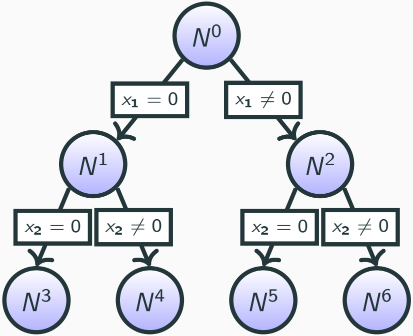
Tree search

$$\begin{cases} \min & \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0 \\ \text{s.t.} & -M \leq \mathbf{x} \leq M \end{cases}$$



Tree search

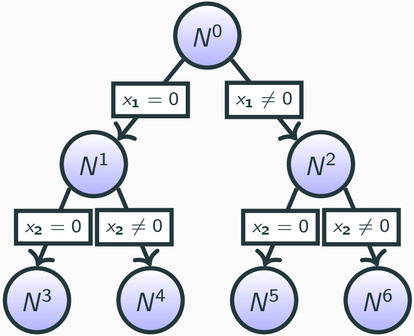
$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$

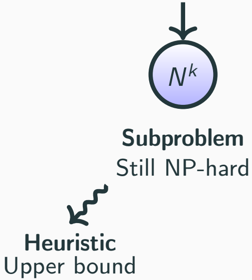


Subproblem
Still NP-hard

Tree search

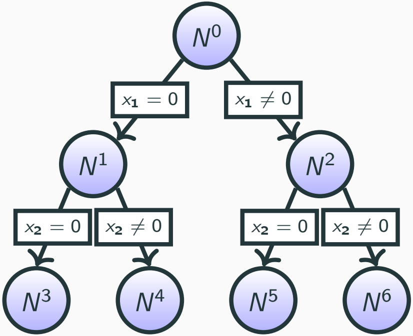
$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$

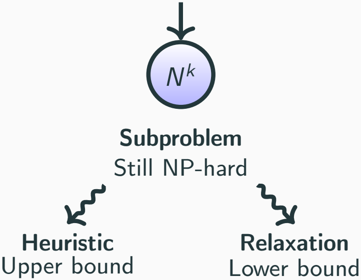




Tree search

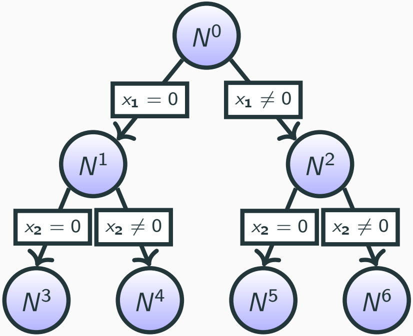
$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & -M \le x \le M \end{cases}$$

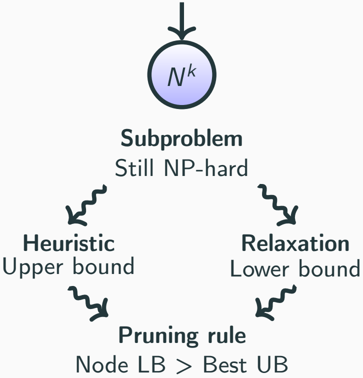




Tree search

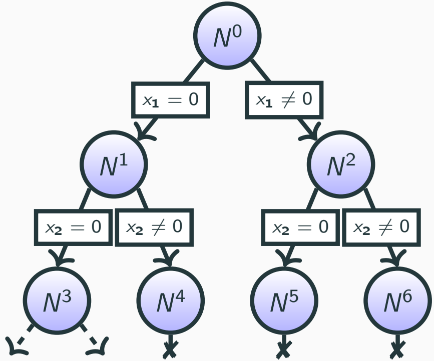
$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$

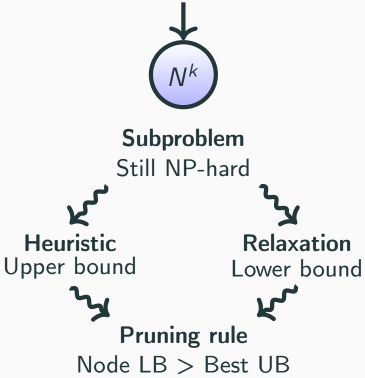




Tree search

$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$





Relaxation construction

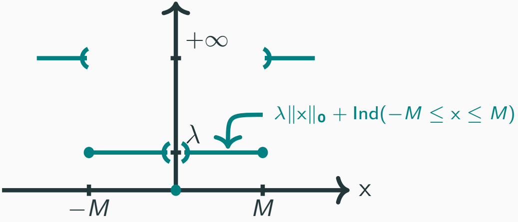
Node problem

$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$

Relaxation construction

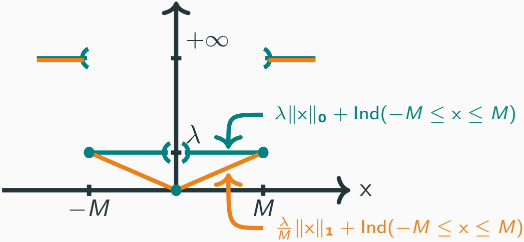
Node problem

$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$



Node problem

$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$



Relaxation construction

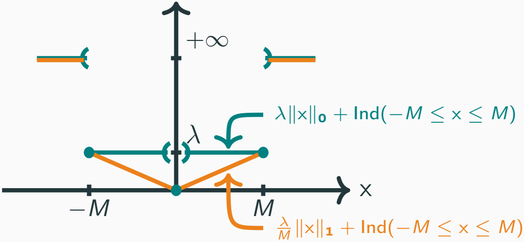
Node problem

$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$

↓

Relaxation

$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \frac{\lambda}{M}\|x\|_1 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$



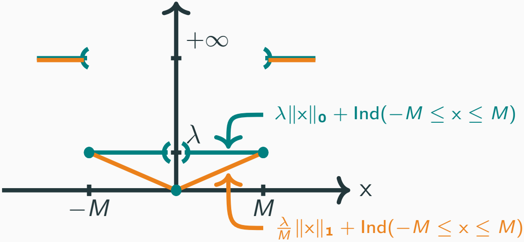
Relaxation construction

Node problem

$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$

Relaxation

$$\begin{cases} \min & \frac{1}{2}\|y - Ax\|_2^2 + \frac{\lambda}{M}\|x\|_1 \\ \text{s.t.} & -M \leq x \leq M \end{cases}$$



Dilemma in the choice of M

Large value of M to recover interesting solutions

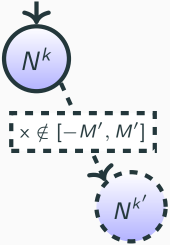
Small value of M to build strong relaxations

Our solution: Peeling

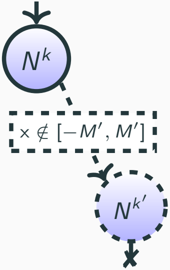
Start with M ***large and refine it locally along branches***

Peeling

Bound tightening

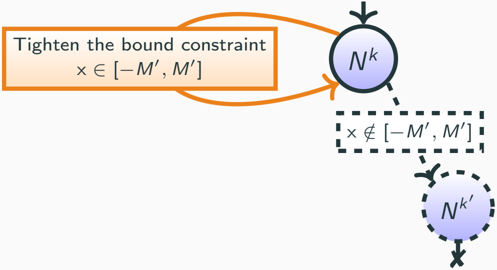


Bound tightening



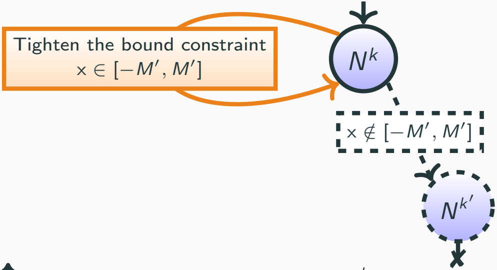
Find some
$M' \leq M$
such that $N^{k'}$ is pruned

Bound tightening



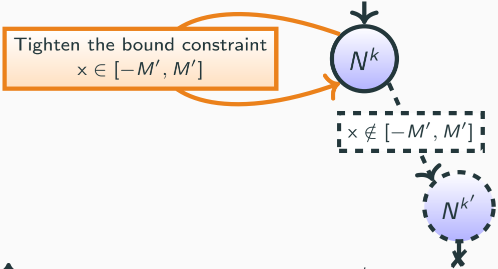
Find some
$M' \leq M$
such that $N^{k'}$ is pruned

Bound tightening



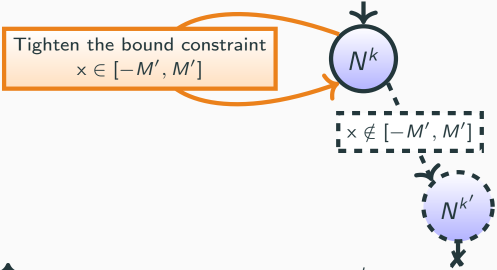
Find some
$M' \leq M$
such that $N^{k'}$ is pruned

Bound tightening



Find some
$M' \le M$
such that $N^{k'}$ is pruned

Bound tightening

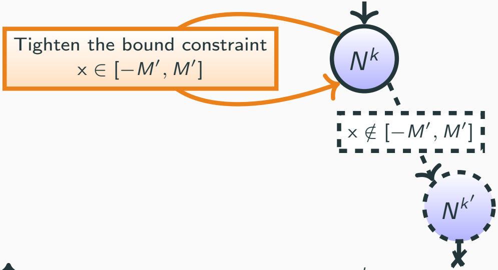


Find some
$M' \leq M$
such that $N^{k'}$ is pruned

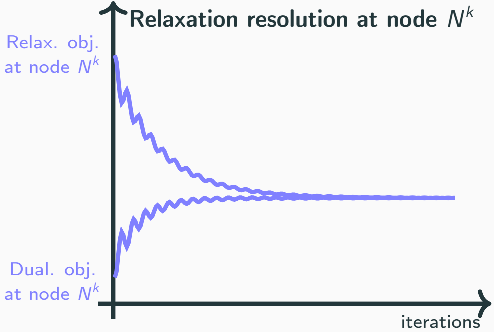
Ingredients

- Already-computed quantities
- Cost-free dual evaluation

Bound tightening



Find some
$M' \leq M$
such that $N^{k'}$ is pruned

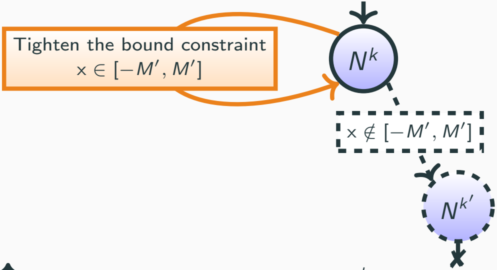


Ingredients

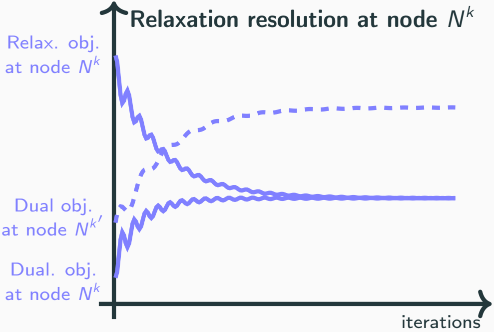
- Already-computed quantities
- Cost-free dual evaluation
- Dual link between nodes

$$D_{N^{k'}}(u) = D_{N^k}(u) + \gamma(u)$$

Bound tightening



Find some
$M' \leq M$
such that $N^{k'}$ is pruned

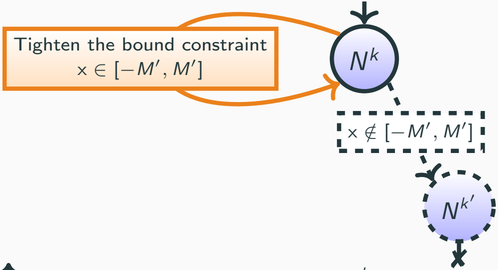


Ingredients

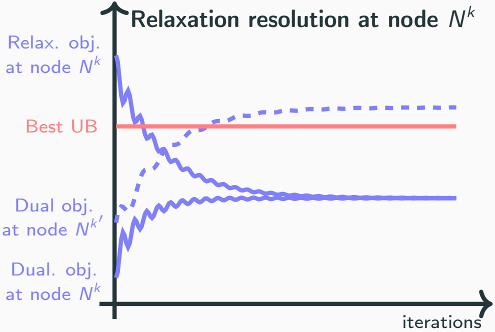
- Already-computed quantities
- Cost-free dual evaluation
- Dual link between nodes

$$D_{N^{k'}}(\mathrm{u}) = D_{N^k}(\mathrm{u}) + \gamma(\mathrm{u})$$

Bound tightening



Find some
$M' \leq M$
such that $N^{k'}$ is pruned

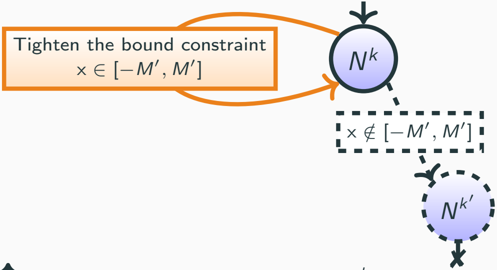


Ingredients

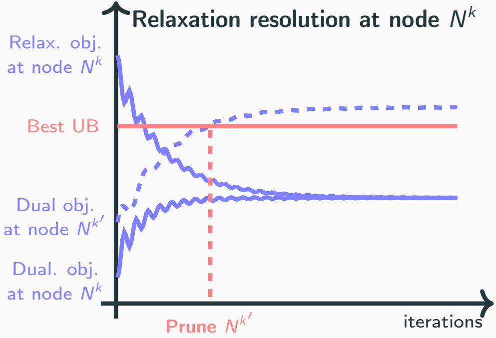
- Already-computed quantities
- Cost-free dual evaluation
- Dual link between nodes

$$D_{N^{k'}}(\mathrm{u}) = D_{N^k}(\mathrm{u}) + \gamma(\mathrm{u})$$

Bound tightening



Find some
$M' \leq M$
such that $N^{k'}$ is pruned

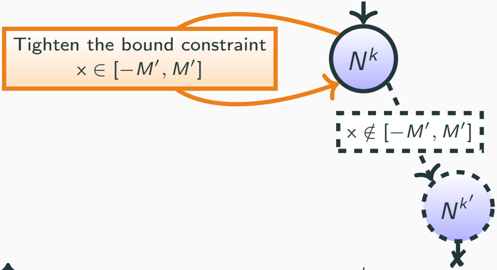


Ingredients

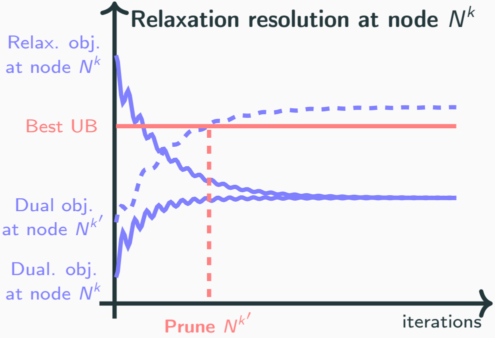
- Already-computed quantities
- Cost-free dual evaluation
- Dual link between nodes

$$D_{N^{k'}}(\mathrm{u}) = D_{N^k}(\mathrm{u}) + \gamma(\mathrm{u})$$

Bound tightening



Find some
$M' \leq M$
such that $N^{k'}$ is pruned

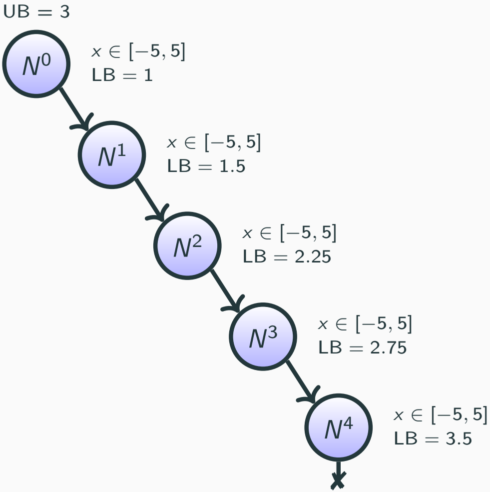


Ingredients

- Already-computed quantities
- Cost-free dual evaluation
- Dual link between nodes
- Coordinate-wise tightening
- Non-symmetric tightening

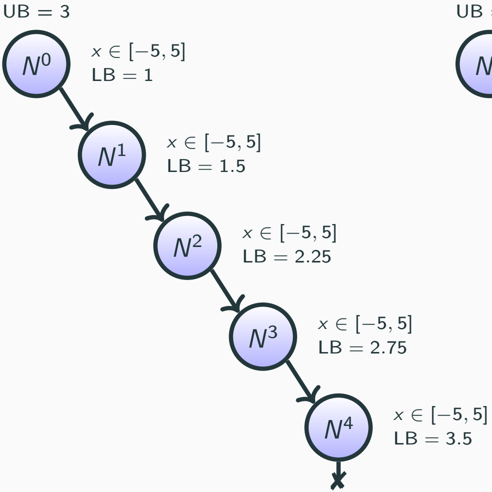
$$D_{N^{k'}}(\mathrm{u}) = D_{N^k}(\mathrm{u}) + \gamma(\mathrm{u})$$

Standard BnB

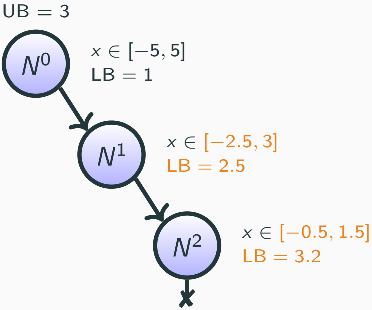


Propagation along branches

Standard BnB

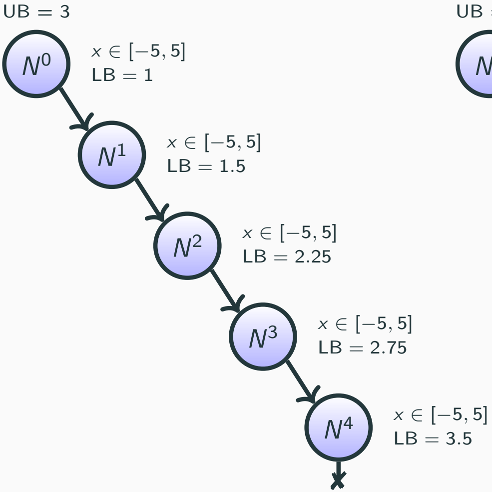


BnB with peeling

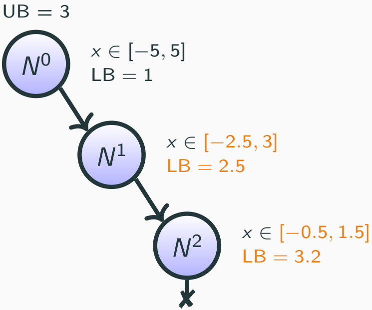


Propagation along branches

Standard BnB

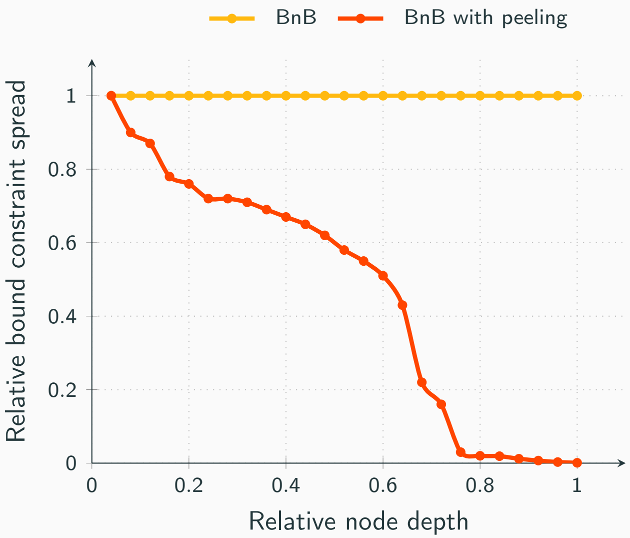


BnB with peeling



- ✔ Branches pruned early-on
- ✔ Less nodes explored
- ✔ Reduce solve time

Bound spread reduction



Numerical results

Numerical results

$$\min_{\mathsf{x}\in[-M,M]} \ \frac{1}{2}\|\mathsf{y}-\mathsf{A}\mathsf{x}\|_2^2+\lambda\|\mathsf{x}\|_0$$

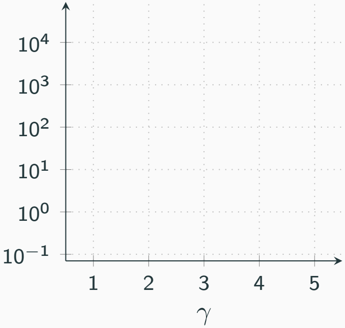
Data A **and** y : Sparse regression
Parameter λ : Set statistically
Bound M : $M=\gamma\|\mathsf{x}^\star\|_\infty$

Numerical results

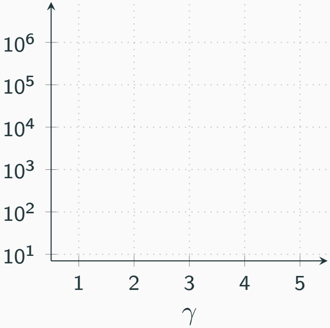
$$\min_{x \in [-M, M]} \ \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0$$

Data A **and** y : Sparse regression
Parameter λ : Set statistically
Bound M : $M = \gamma\|x^\star\|_\infty$

Solve time (sec.)



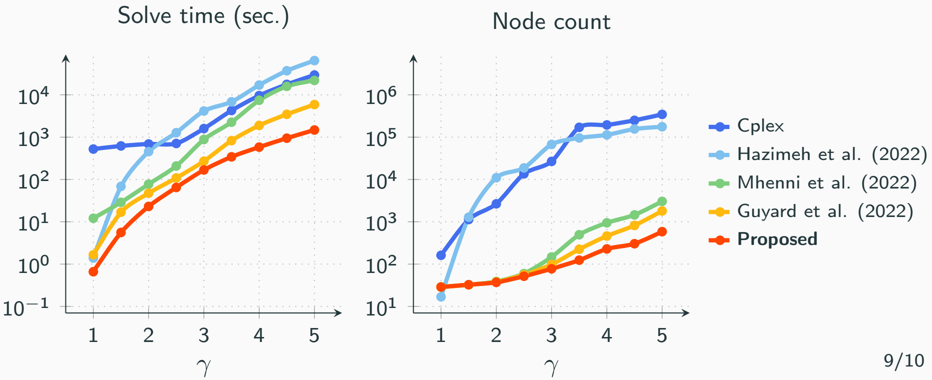
Node count



Numerical results

$$\min_{x \in [-M,M]} \ \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda \|x\|_0$$

Data A **and** y : Sparse regression
Parameter λ : Set statistically
Bound M : $M = \gamma \|x^\star\|_\infty$



Take-home message

ℓ_0-regularized least-squares

$$\min_{x} \quad \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0$$
$$\text{s.t.} \quad -M \leq x \leq M$$

Wide interval $[-M, M]$ to recover the solution

Tight interval $[-M, M]$ to build strong relaxations

Peeling

Start with wide interval $[-M, M]$ and refine it along the BnB branches

Question time

🤓