Peeling for L0-Regularized Least-Squares

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

[⋆]ICTEAM, UCLouvain, Belgium
[†]SIMSMART, Inria, France
[‡]IETR, CentraleSupelec, France

EUSIPCO - European Signal Processing Conference
Helsinki, Finland
September 4-8, 2023

L0-Regularized Least-Squares

Sparse linear model

Linear approximation ⟷ **Sparse linear model** ⟷ **Sparse solution**

Machine Learning
Signal processing
Network design
...

Sparse linear model

Linear approximation $\longleftrightarrow$ **Sparse linear model** $\longleftrightarrow$ **Sparse solution**

Machine Learning
Signal processing
Network design
...

ℓ_0-regularized least-squares

$$\begin{array}{ll} \min_x & \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0 \\ \text{s.t.} & L \leq x \leq U \end{array}$$

Sparse linear model

Linear approximation $\longleftrightarrow$ **Sparse linear model** $\longleftrightarrow$ **Sparse solution**

Machine Learning
Signal processing
Network design
...

ℓ_0-regularized least-squares

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

Bounds needed for numerical tractability

Sparse linear model

Linear approximation ⟷ **Sparse linear model** ⟷ **Sparse solution**

Machine Learning
Signal processing
Network design
...

ℓ_0-regularized least-squares

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

Bounds needed for numerical tractability

Generic Mixed-Integer Program solvers

Specialized Branch-and-Bound solvers

Branch-and-Bound

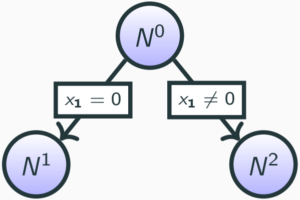
Tree search

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

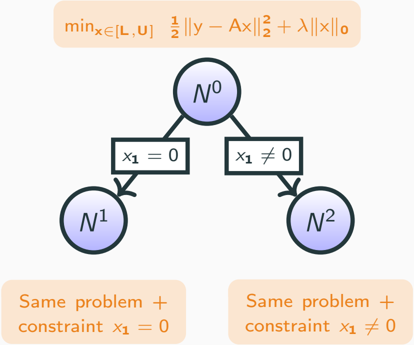
N^0

Tree search

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

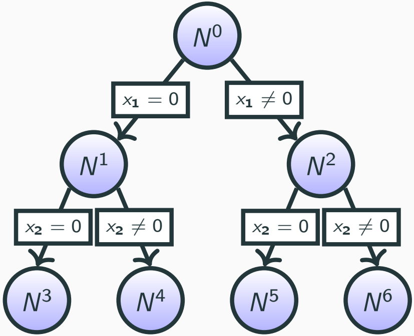


Tree search


$\min_{\mathbf{x} \in [\mathbf{L}, \mathbf{U}]} \ \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0$
N^0
$x_1 = 0$
$x_1 \neq 0$
N^1
N^2
Same problem + constraint $x_1 = 0$
Same problem + constraint $x_1 \neq 0$

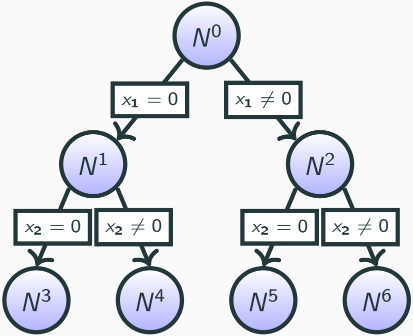
Tree search

$\min_{\mathbf{x}\in[\mathbf{L},\mathbf{U}]} \quad \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0$



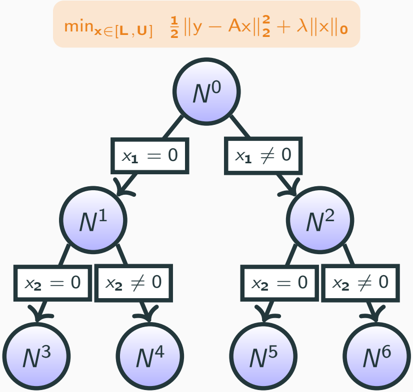
Tree search

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

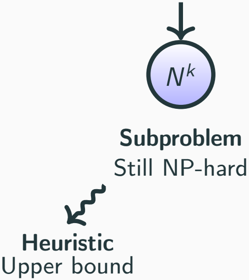


Subproblem
Still NP-hard

Tree search



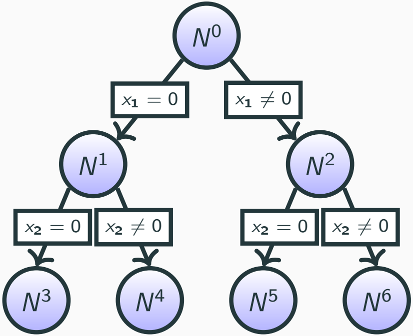
$\min_{\mathbf{x} \in [\mathbf{L}, \mathbf{U}]} \ \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0$
N^0
$x_1 = 0$
$x_1 \neq 0$
N^1
N^2
$x_2 = 0$
$x_2 \neq 0$
$x_2 = 0$
$x_2 \neq 0$
N^3
N^4
N^5
N^6

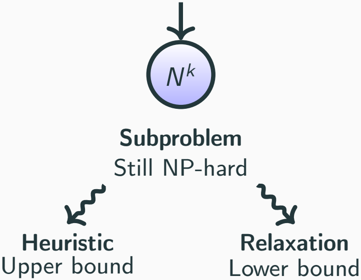


N^k
Subproblem
Still NP-hard
Heuristic
Upper bound

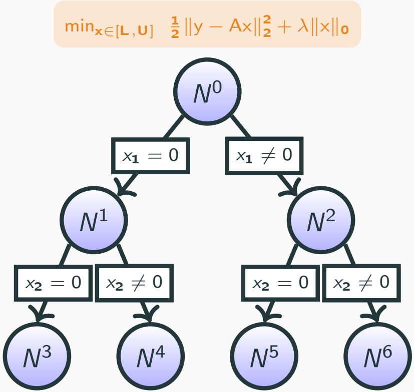
Tree search

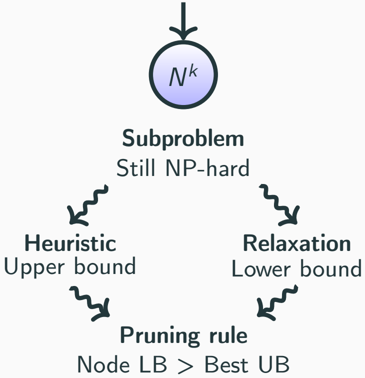
$\min_{\mathbf{x}\in[\mathbf{L},\mathbf{U}]} \ \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0$



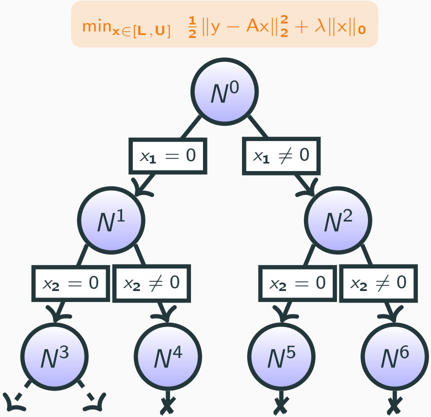


Tree search

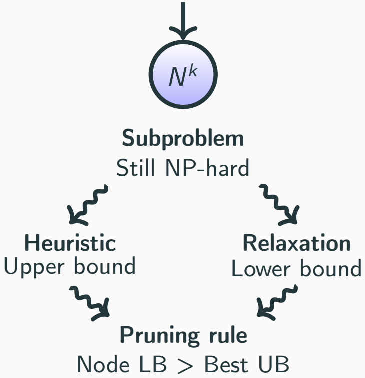

$\min_{\mathbf{x}\in[\mathbf{L},\mathbf{U}]} \ \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0$
N^0
$x_1 = 0$
$x_1 \neq 0$
N^1
N^2
$x_2 = 0$
$x_2 \neq 0$
$x_2 = 0$
$x_2 \neq 0$
N^3
N^4
N^5
N^6


N^k
Subproblem
Still NP-hard
Heuristic
Upper bound
Relaxation
Lower bound
Pruning rule
Node LB > Best UB

Tree search



$\min_{\mathbf{x} \in [\mathbf{L}, \mathbf{U}]} \ \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0$
N^0
$x_1 = 0$
$x_1 \neq 0$
N^1
N^2
$x_2 = 0$
$x_2 \neq 0$
$x_2 = 0$
$x_2 \neq 0$
N^3
N^4
N^5
N^6



N^k
Subproblem
Still NP-hard
Heuristic
Upper bound
Relaxation
Lower bound
Pruning rule
Node LB > Best UB

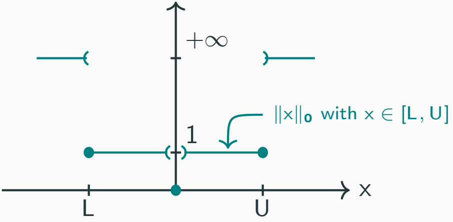
Relaxation construction

Node problem

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

Node problem

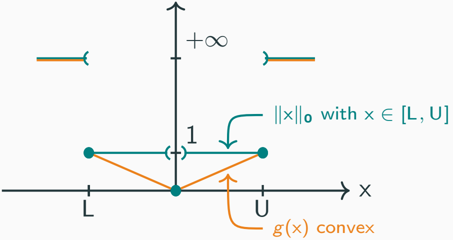
$$\min_{\mathbf{x}\in[\mathbf{L},\mathbf{U}]} \ \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0$$



Relaxation construction

Node problem

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



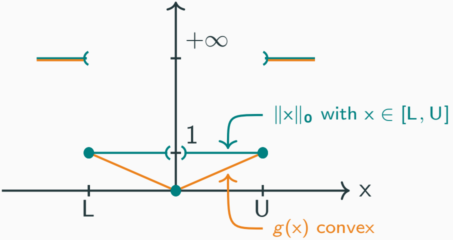
Node problem

$\min_{\mathbf{x}\in[\mathbf{L},\mathbf{U}]} \; \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0$

$\downarrow$

Relaxation

$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda g(\mathbf{x})$



Relaxation construction

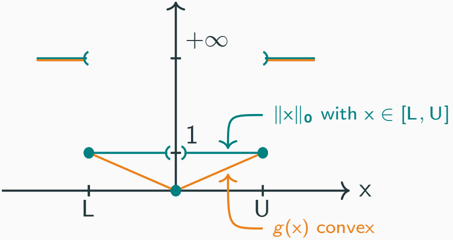
Node problem

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

$\downarrow$

Relaxation

$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda g(\mathbf{x})$



Dilemma in the choice of $[L, U]$

- Wide interval $[L, U]$ to recover the solution
- Tight interval $[L, U]$ to build strong relaxations

Relaxation construction

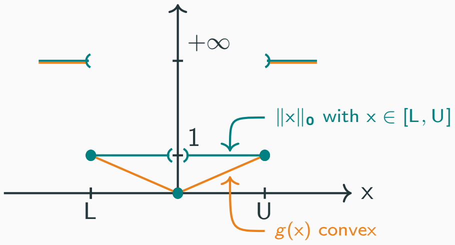
Node problem

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

↓

Relaxation

$\min_{x} \frac{1}{2}\|y - Ax\|_2^2 + \lambda g(x)$



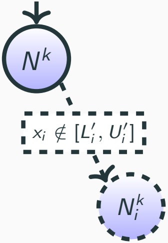
Dilemma in the choice of $[L, U]$

- Wide interval $[L, U]$ to recover the solution
- Tight interval $[L, U]$ to build strong relaxations

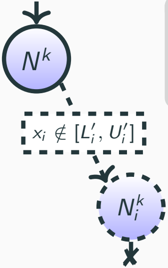
Our solution : Peeling

Peeling

Bound tightening

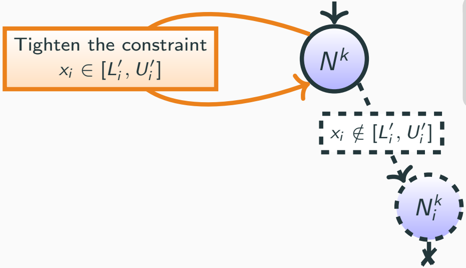


Bound tightening



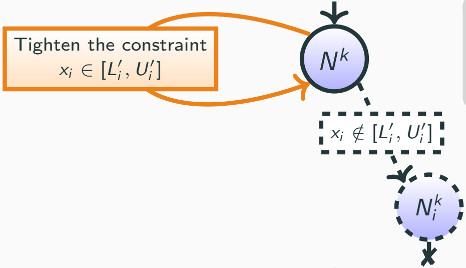
Find some
$[L_i', U_i'] \subseteq [L_i, U_i]$
such that N_i^k is pruned

Bound tightening



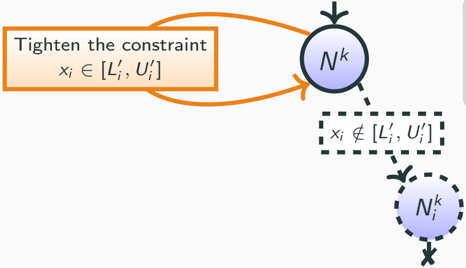
Find some
$[L'_i, U'_i] \subseteq [L_i, U_i]$
such that N_i^k is pruned

Bound tightening



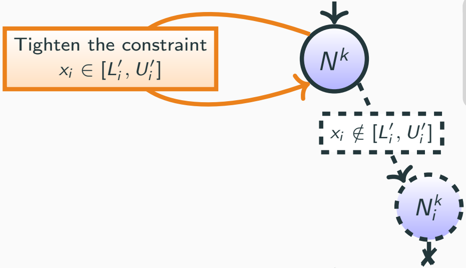
Find some
$[L_i', U_i'] \subseteq [L_i, U_i]$
such that N_i^k is pruned

Bound tightening



Find some
$[L'_i, U'_i] \subseteq [L_i, U_i]$
such that N_i^k is pruned

Bound tightening

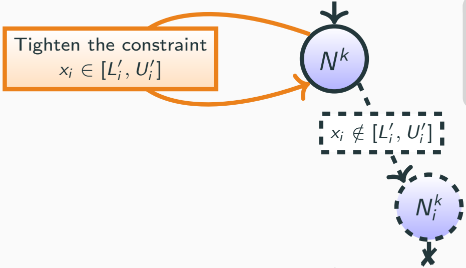


Find some
$[L'_i, U'_i] \subseteq [L_i, U_i]$
such that N^k_i is pruned

Ingredients

- Already-computed quantities
- Cost-free dual evaluation

Bound tightening

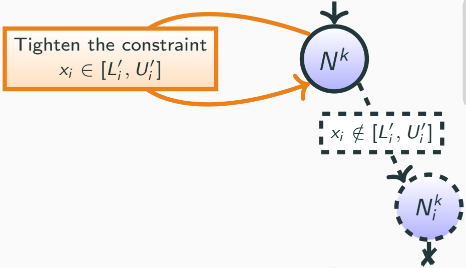


Find some
$[L'_i, U'_i] \subseteq [L_i, U_i]$
such that N^k_i is pruned

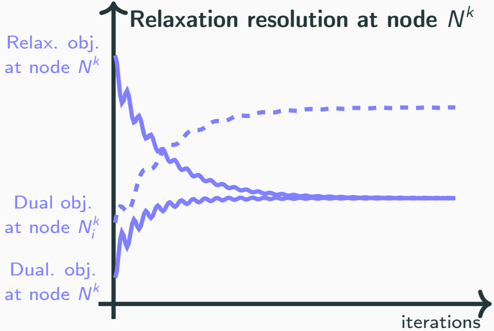
Ingredients

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

Bound tightening



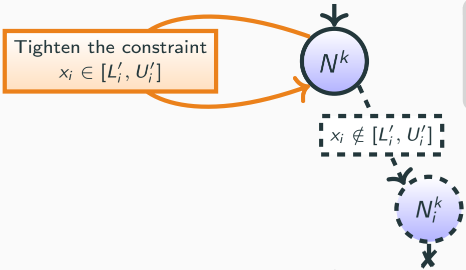
Find some
$[L'_i, U'_i] \subseteq [L_i, U_i]$
such that N^k_i is pruned



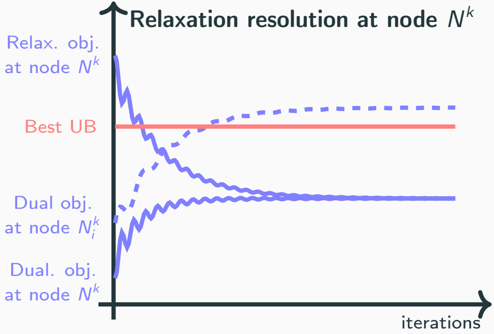
Ingredients

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

Bound tightening

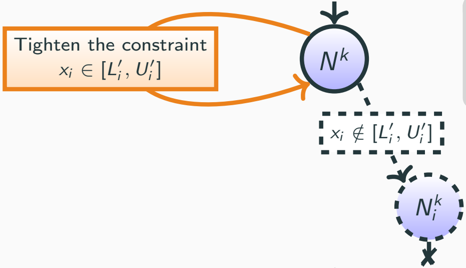


Find some
$[L'_i, U'_i] \subseteq [L_i, U_i]$
such that N_i^k is pruned

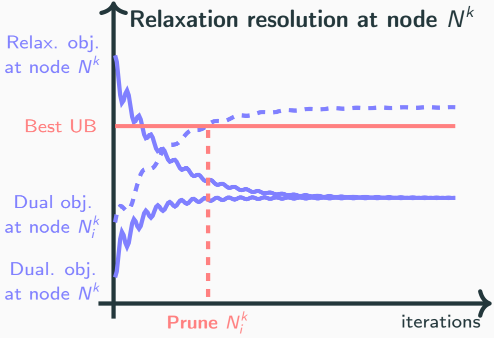


Ingredients
- Already-computed quantities
- Cost-free dual evaluation
- Dual link between nodes

Bound tightening



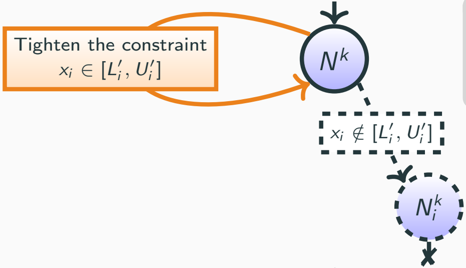
Find some
$[L'_i, U'_i] \subseteq [L_i, U_i]$
such that N_i^k is pruned



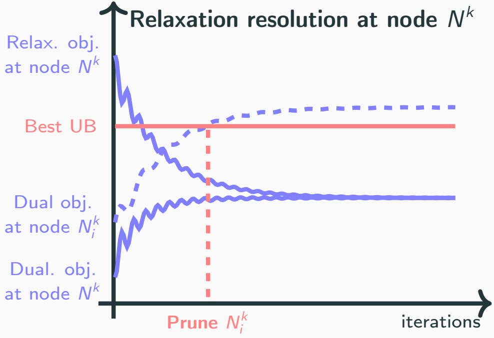
Ingredients

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

Bound tightening



Find some
$[L'_i, U'_i] \subseteq [L_i, U_i]$
such that N_i^k is pruned

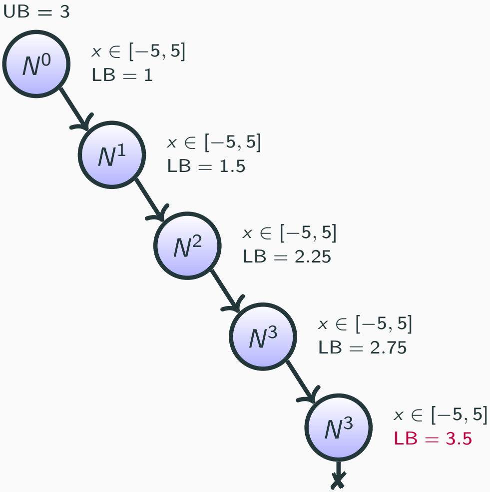


Ingredients

- Already-computed quantities
- Cost-free dual evaluation
- Dual link between nodes
- Can tighten **multiple** bounds

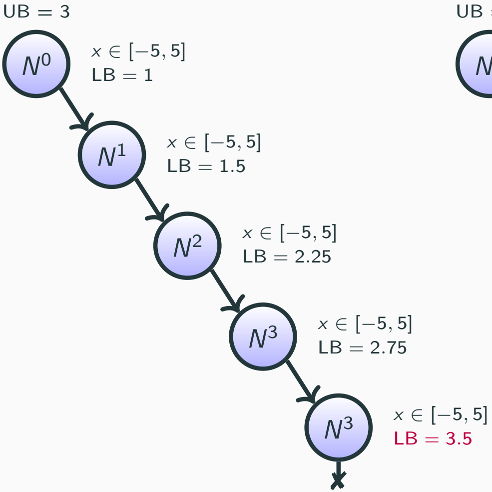
Propagation along branches

Standard BnB

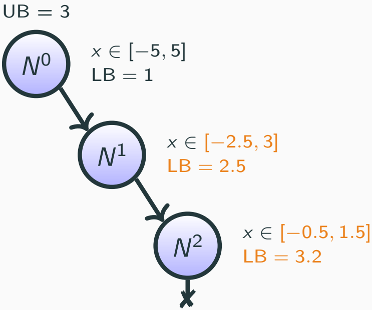


Propagation along branches

Standard BnB

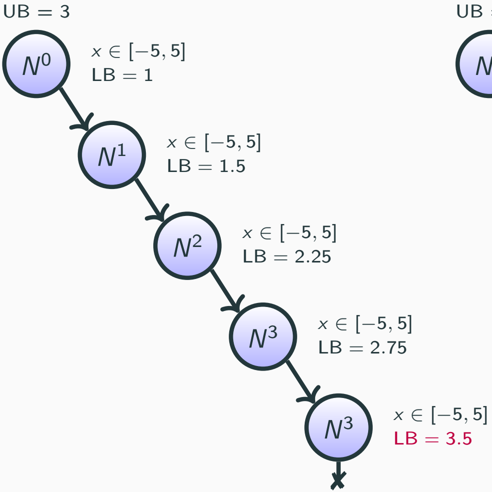


BnB with peeling

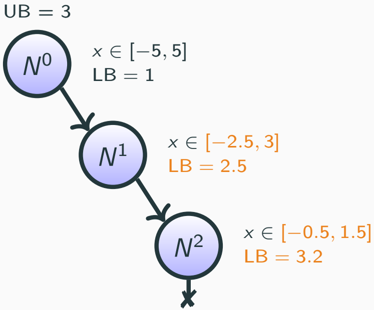


Propagation along branches

Standard BnB



BnB with peeling



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

Numerical results

Numerical results

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

Data A **and** y : Sparse regression

Bounds L **and** U : $-M1 \leq x \leq M1$

Parameter λ : Set statistically

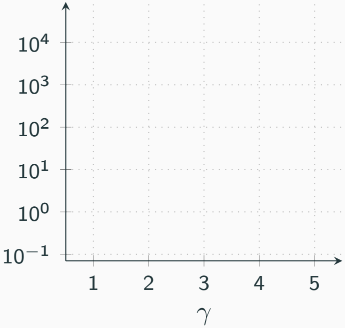
Parameter γ : $M = \gamma\|x^\star\|_\infty$

Numerical results

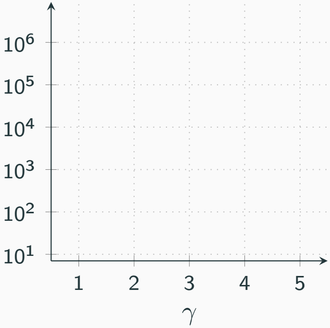
$$\min_{x \in [L,U]} \ \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_0$$

Data A **and** y : Sparse regression
Bounds L **and** U : $-M1 \leq x \leq M1$
Parameter λ : Set statistically
Parameter γ : $M = \gamma\|x^\star\|_\infty$

Solve time (sec.)



Node count



Numerical results

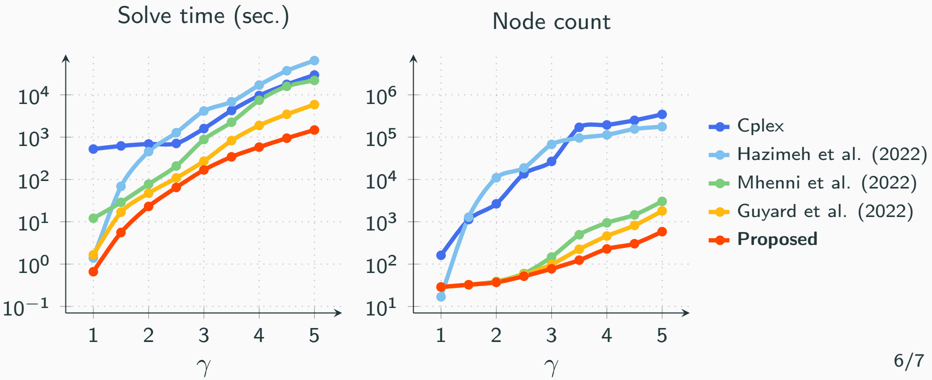
$$\min_{\mathrm{x}\in[\mathrm{L},\mathrm{U}]} \ \tfrac{1}{2}\|\mathrm{y}-\mathrm{A}\mathrm{x}\|_2^2 + \lambda\|\mathrm{x}\|_0$$

Data A **and** y : Sparse regression

Bounds L **and** U : $-M\mathbf{1} \leq \mathrm{x} \leq M\mathbf{1}$

Parameter λ : Set statistically

Parameter γ : $M = \gamma\|\mathrm{x}^\star\|_\infty$



Take-home message

ℓ_0-regularized least-squares

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

Wide interval $[L, U]$ to recover the solution

Tight interval $[L, U]$ to build strong relaxations

Peeling

Start with wide interval $[L, U]$ and refine it along the BnB branches.

Question time

🤓