A New Branch-and-Bound Pruning Framework for L0-Regularized Problems

Théo Guyard, Cédric Herzet, Clément Elvira and Ayşe-Nur Arslan
Inria and Insa Rennes, Ensai Rennes, CentraleSupélec Rennes, Inria Bordeaux

Context

Sparse optimization problems aim to minimize a **loss** function and enforce **sparsity** via a regularization.

Optimization problem

$$p = \min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{A}\mathbf{x}) + \underbrace{\lambda\|\mathbf{x}\|_0 + h(\mathbf{x})}_{g(\mathbf{x})}$$

- Loss function f tied to the application at hand
- Matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ encoding a linear model
- Regularization involving an ℓ_0-norm

Numerical Considerations

- Non-convex and NP-hard problem
- Can be modeled as a Mixed Integer Program
- **Large-scale instances reputed intractable**

Recent Advances

- Specialized **Branch-and-Bound** solvers
  - Hazimeh et al., 2022: BnB for $h(\mathbf{x}) = \|\mathbf{x}\|_2^2$
  - Samain et al., 2023: BnB for $h(\mathbf{x}) = \eta_M(\mathbf{x})$
- Dimensionality reduction
  - Pilanci et al., 2015: Randomized rounding
  - Atamtürk et al., 2020: Screening for $h(\mathbf{x}) = \|\mathbf{x}\|_2^2$
  - Guyard et al., 2022: Screening for $h(\mathbf{x}) = \eta_M(\mathbf{x})$
- **Restricted to some problem instances**

Contributions

Generic Framework

- f is proper, closed, convex
- h is proper, closed, convex, separable
- $h(\mathbf{0}) = 0$, $\mathbf{x} = \mathbf{0}$ is an accumulation point of dom(h)

Simultaneous Pruning

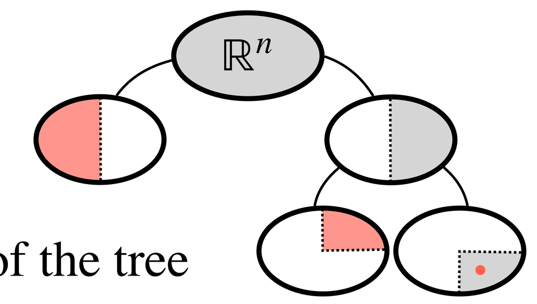
- Acceleration of Branch-and-Bound solvers
- Significant computational gains
- Address instances out of computational reach so far

Branch-and-Bound

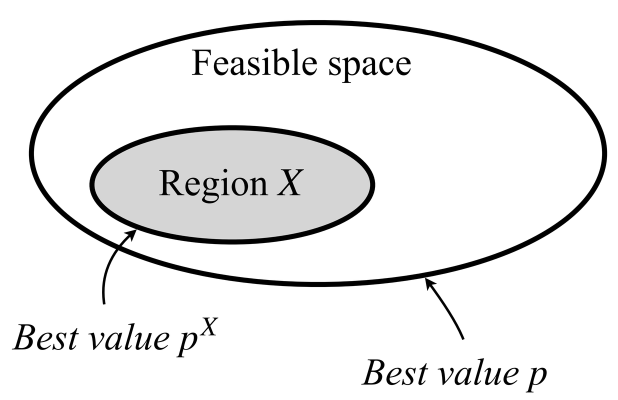
Partition the feasible space into **regions** and discard those that cannot contain global solutions with a **pruning test**.

Feasible Space Exploration

- Implicit enumeration of all possible supports
- Split the feasible space
- Drive the sparsity of $\mathbf{x}$
- Total of 2^n **regions**
- Decision tree exploration
- Pruning test to discard parts of the tree



Pruning Test



$$p_X > p \rightarrow \text{prune } X$$

lower bound ↓ upper bound ↓

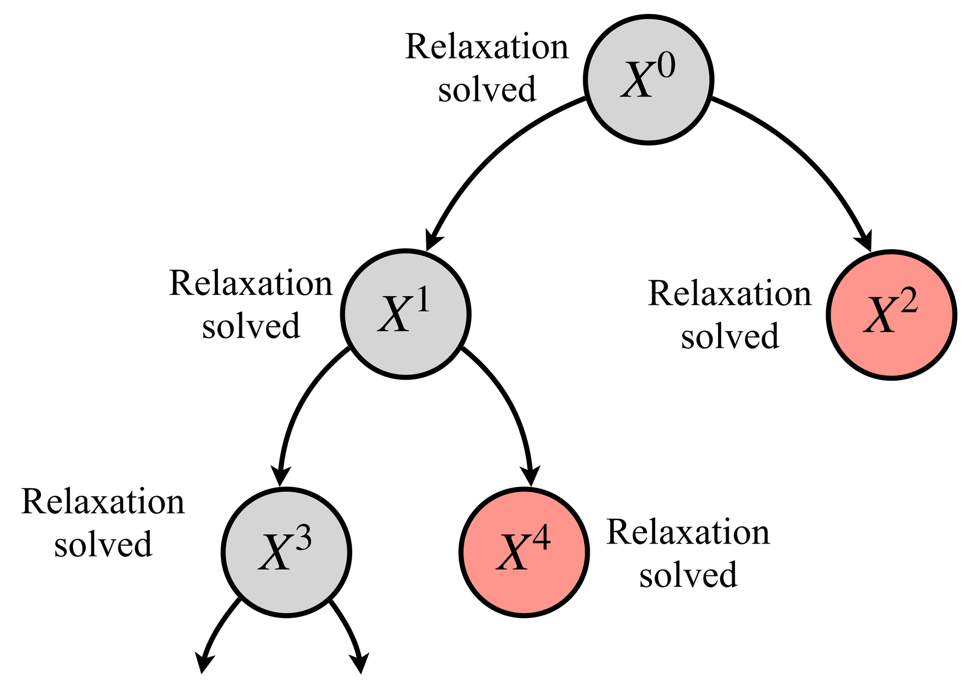
$$\check{p}_X > \hat{p} \rightarrow \text{prune } X$$

Lower Bound Computation

Relaxation for region X

$$\check{p}_X = \min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{A}\mathbf{x}) + \check{g}_X(\mathbf{x})$$

- Lower approximation $\check{g}_X(\mathbf{x})$ of $g(\mathbf{x}) + \eta_X(\mathbf{x})$
- Convex optimization problem
- First-order method (proximal gradient, admm, …)
- **One relaxation solved per pruning test**



Simultaneous Pruning

Trade some **tightness** of the lower bounding process for a more **tractable** implementation.

Another Family of Lower Bounds

Fenchel-Rockafellar dual for region X

$$\check{d}_X = \max_{\mathbf{u}\in\mathbb{R}^m} \underbrace{-f^\star(-\mathbf{u}) + \check{g}_X^\star(-\mathbf{A}^\top\mathbf{u})}_{D_X(\mathbf{u})}$$

- $f^\star$ and $\check{g}_X^\star$ are the convex conjugates of f and $\check{g}_X$
- Dual lower bound

$$D_X(\mathbf{u}) > \hat{p} \rightarrow \text{prune } X$$

- Works with any $\mathbf{u} \in \mathbb{R}^m$
- No need to solve the dual relaxation
- Already used for some problem instances but we provide a **generic characterization of $\check{g}_X$**

Further Pruning Opportunities

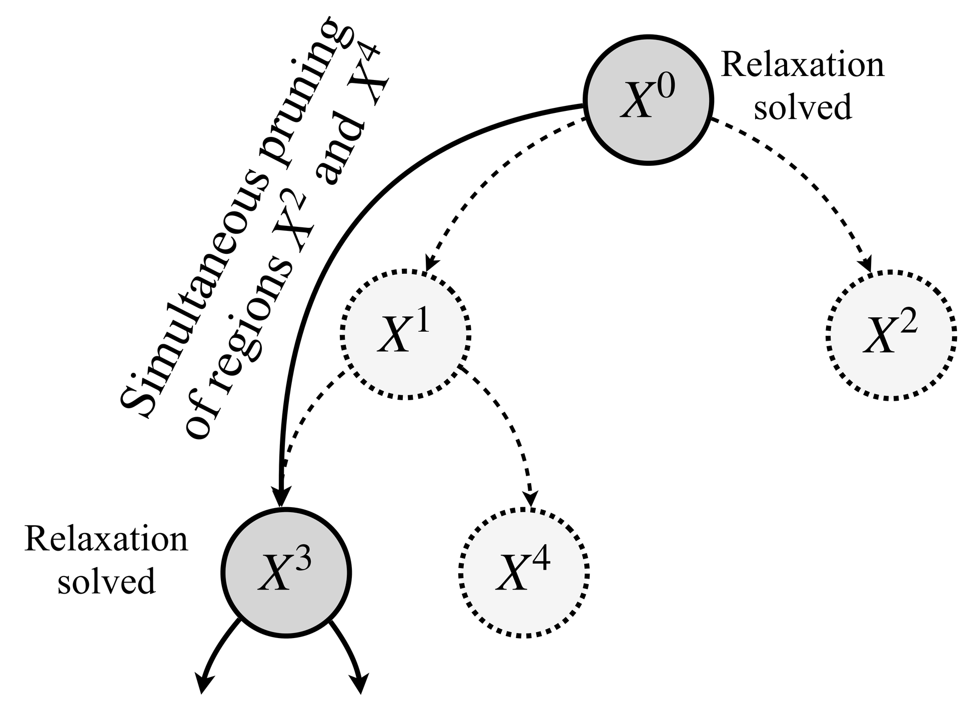
- Nested region $X' \subseteq X$
- Link between objective functions of dual problems

$$D_{X'}(\mathbf{u}) = D_X(\mathbf{u}) + \Delta_{X,X'}(\mathbf{u})$$

- Term $\Delta_{X,X'}(\mathbf{u})$ evaluated in $\mathcal{O}(1)$
- Pruning test for several nested regions

$$D_X(\mathbf{u}) + \Delta_{X,X'}(\mathbf{u}) > \hat{p} \rightarrow \text{prune } X'$$

- Only requires one evaluation of the function D_X
- **Simultaneous pruning of regions at low cost**



Numerics

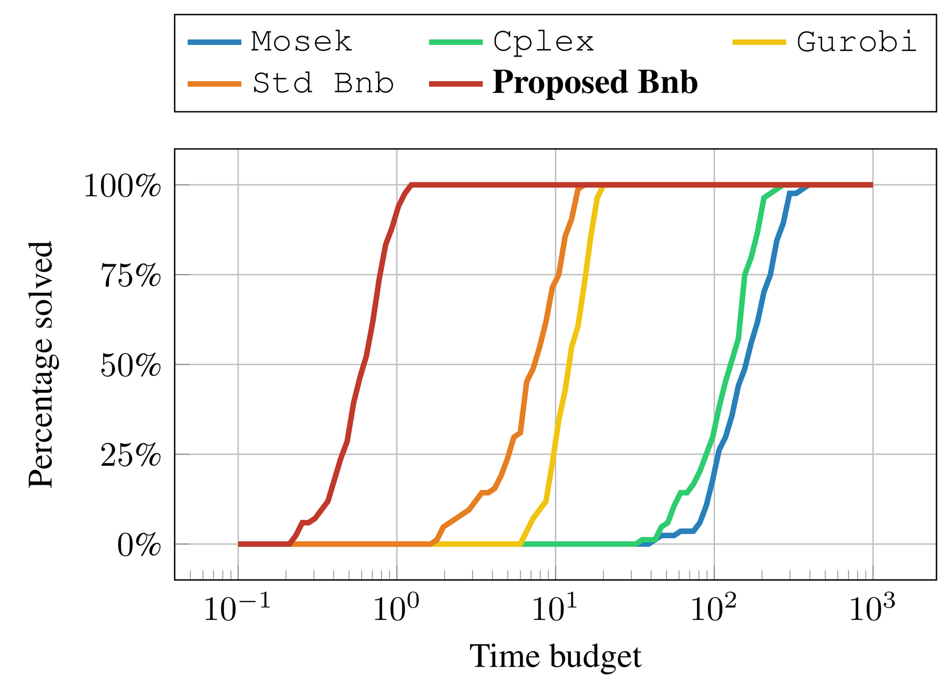
How can our simultaneous pruning strategy accelerate Branch-and-Bound solvers?

```
~ pip install el0ps
```

Setup: Comparison against generic MIP solvers (`Mosek`, `Cplex` and `Gurobi`) and a state-of-the-art Branch-and-Bound implementation (`L0bnb` from Hazimeh et al.).

Performance Profiles

- f : Least-squares loss
- h : Squared L2-norm with Big-M constraint
- Synthetic sparse regression data



Large Scale Datasets

- f : Least-squares loss or Logistic loss
- h : L1-norm / squared L2-norm with Big-M constraint
- Fit a regularization path