Unifying Branch-and-Bound methods for ℓ_0-penalized problems

Théo Guyard*, Cédric Herzet†, Clément Elvira‡, Ayşe-Nur Arslan$^\diamond$

*Inria and Insa Rennes, †Ensai, ‡CentraleSupélec, $^\diamond$Inria Bordeaux

In short

Provide necessary ingredients to implement **generic** and **efficient** ℓ_0-problem solvers.

ℓ_0-problem

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

Generic solvers
✗ Slow
✓ Flexible w.r.t f/h

Specialized solvers
✓ Fast
✗ Only some f/h

Proposed solver
✓ Fast
✓ Flexible w.r.t f/h

Working hypotheses

- f simple to work with (convex, smooth, ...)
- h proper, closed, separable, cont. at $\mathbf{x} = \mathbf{0}$
- h even, convex, coercive, $h \geq h(\mathbf{0}) = 0$

Versatile toolbox

User-defined f
```
func_value
conj_value
prox_value
subd_value
```

User-defined h
```
func_value
conj_value
prox_value
subd_value
```

↓

`El0ps`
Branch-and-Bound solver
Tunable strategies
Acceleration methods

Branch-and-Bound algorithms

Principle

"Partition the feasible space into **regions** and **prune** those that cannot contain minimizers."

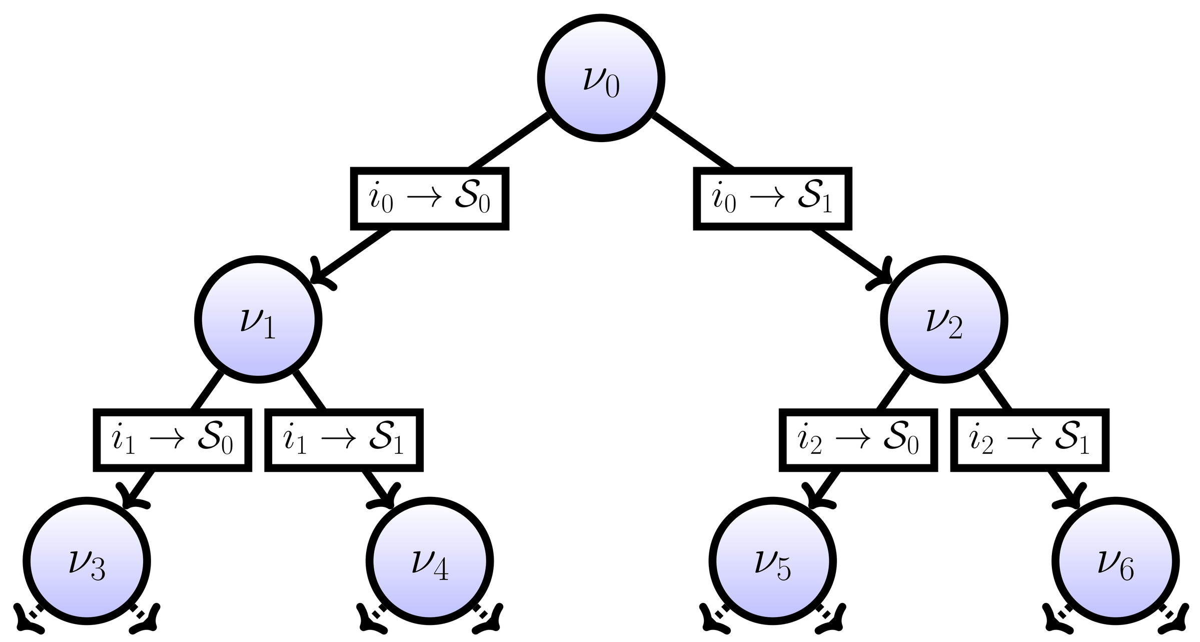
Regions construction

- $\nu = (\mathcal{S}_0, \mathcal{S}_1)$ disjoint subsets of $\{1,\dots,n\}$
- $\mathcal{S}_0$ entries fixed to zero
- $\mathcal{S}_1$ entries fixed to non-zero

$$\mathcal{X}^\nu = \{\mathbf{x}\in\mathbf{R}^n \mid \mathbf{x}_{\mathcal{S}_0} = \mathbf{0},\ \mathbf{x}_{\mathcal{S}_1} \neq \mathbf{0}\}$$

- All regions explored: problem solved

Feasible space exploration



Processing nodes

- Node problem

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

- Test if region $\mathcal{X}^\nu$ contains minimizers: $p^\nu > p^\star$
- Practical pruning test with bounds

$$p^\nu \geq \tilde{p}^\nu > \bar{p} \geq p^\star$$

- Upper bound $\bar{p}$ on $p^\star$
  → Evaluate objective at any point
  → Efficient heuristics

Generic relaxations

Constructing lower-bounds

Node problem reformulation

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

$$\text{with } g_i^\nu(x_i) = \begin{cases} \mathbf{I}(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ h_i(x_i) + \lambda\|x_i\|_0 & \text{otherwise} \end{cases}$$

Primal relaxation

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

Dual relaxation

$$\max_{\mathbf{u}\in\mathbf{R}^m} -f^\star(-\mathbf{u}) + (g^\nu)^\star(\mathbf{A}^{\mathrm{T}}\mathbf{u})$$

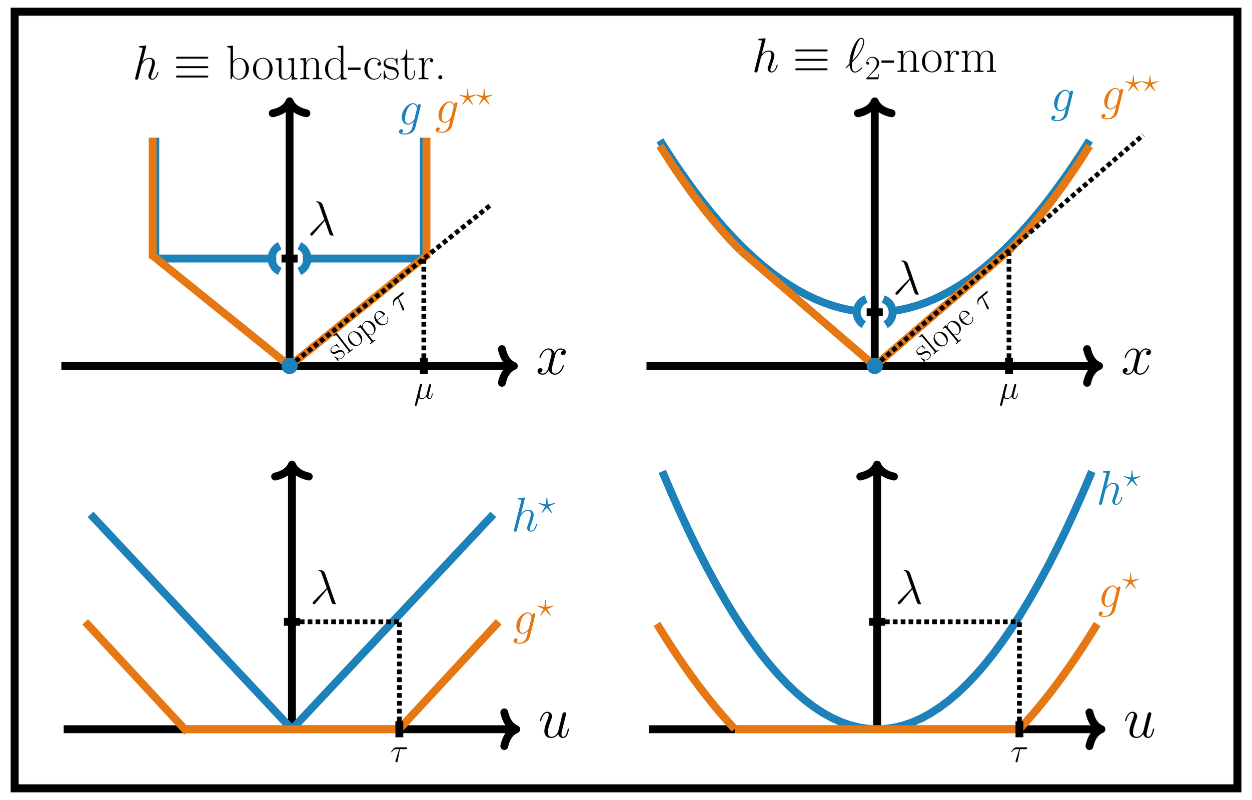
Closed-form expressions

- 1D-study of $g(x) = h(x) + \lambda\|x\|_0$
- Parameters (τ, μ) easily computable from $h^\star$
- Closed-form expressions

$$g^{\star\star}(x) = \begin{cases} \tau|x| & \text{if } |x| \leq \mu \\ h(x) + \lambda & \text{if } |x| > \mu \end{cases}$$

$$g^\star(u) = \begin{cases} 0 & \text{if } |u| \leq \tau \\ h^\star(u) - \lambda & \text{if } |u| > \tau \end{cases}$$

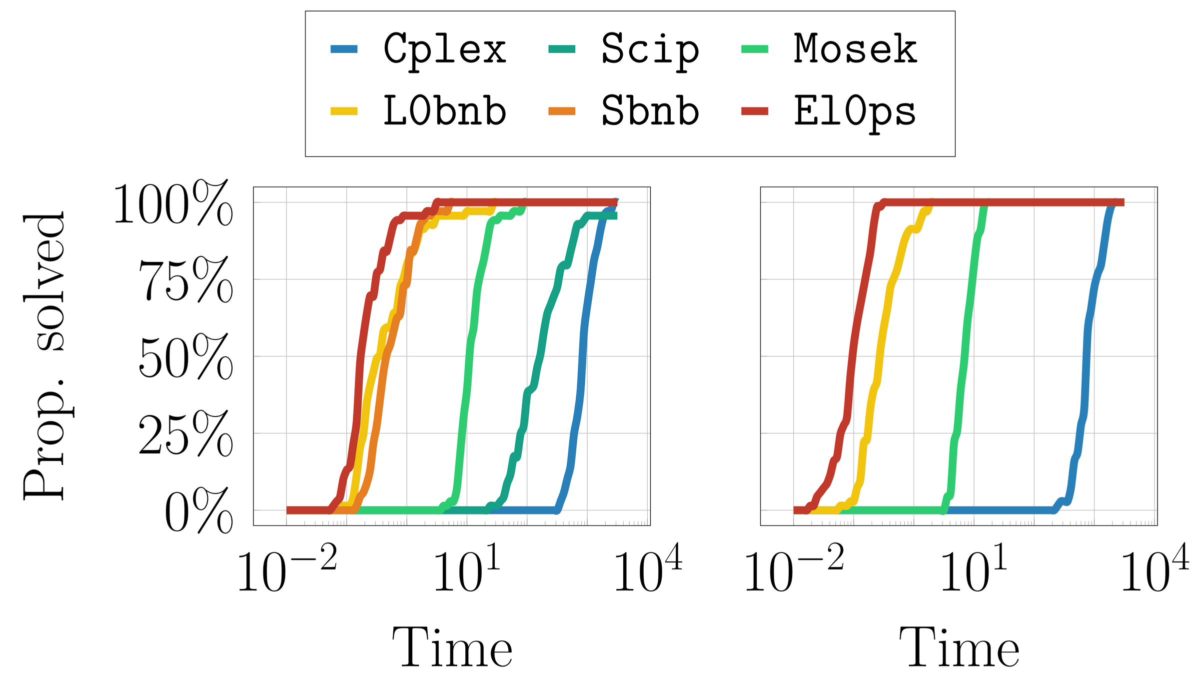
- Simple evaluation of prox, subdiff, ...



Numerical results

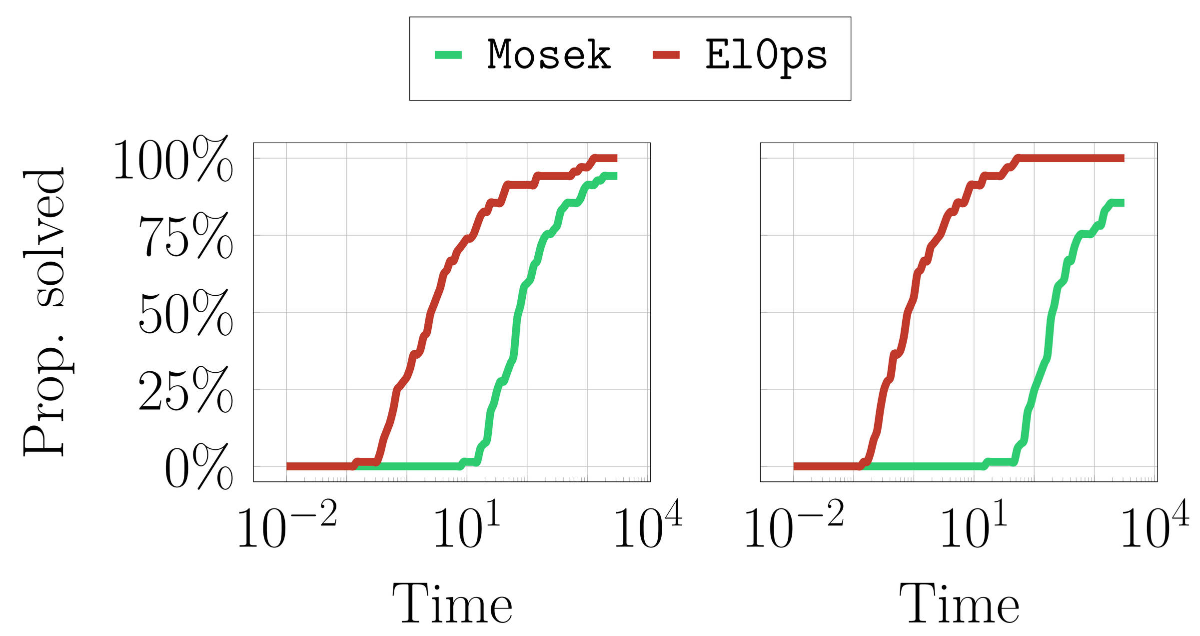
Standard instances

- f: least-squares
- h: bound-cstr. or ℓ_2-norm (left/right)
- λ: tuned statistically



New application opportunities

- f: logistic
- h: bound-cstr. + ℓ_1- or ℓ_2-norm (left/right)
- λ: tuned statistically



✓ Fast
✓ Flexible w.r.t f/h

Take home message

Opportunities to address new instances of ℓ_0-problems efficiently.