Branch-and-Bound Algorithms for ℓ_0-Regularized Problems

Théo Guyard

Inria and Insa Rennes

PhD defense - November 27th, 2024

ℓ_0-regularized problems

Problem

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

ℓ_0-regularized problems

Problem

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

1) Loss $f(\mathbf{Ax})$

Models the quantity to minimize in the application at hand

ℓ_0-regularized problems

Problem

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

1) Loss $f(\mathbf{A}\mathbf{x})$

Models the quantity to minimize in the application at hand

2) ℓ_0-norm $\|\mathbf{x}\|_0$

Counts non-zeros in its input to promote sparse solutions

ℓ_0-regularized problems

Problem

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

1) Loss $f(\mathbf{A}\mathbf{x})$

Models the quantity to minimize in the application at hand

2) ℓ_0-norm $\|\mathbf{x}\|_0$

Counts non-zeros in its input to promote sparse solutions

3) Penalty $h(\mathbf{x})$

Linked to the application or artificially introduced to enable solution methods

ℓ_0-regularized problems

Problem

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

1) Loss $f(\mathbf{A}\mathbf{x})$

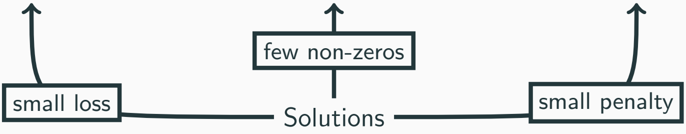
Models the quantity to minimize in the application at hand

2) ℓ_0-norm $\|\mathbf{x}\|_0$

Counts non-zeros in its input to promote sparse solutions

3) Penalty $h(\mathbf{x})$

Linked to the application or artificially introduced to enable solution methods



ℓ_0-regularized problems

Problem

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

tradeoff parameter $\lambda > 0$

1) Loss $f(\mathbf{A}\mathbf{x})$

Models the quantity to minimize in the application at hand

2) ℓ_0-norm $\|\mathbf{x}\|_0$

Counts non-zeros in its input to promote sparse solutions

3) Penalty $h(\mathbf{x})$

Linked to the application or artificially introduced to enable solution methods

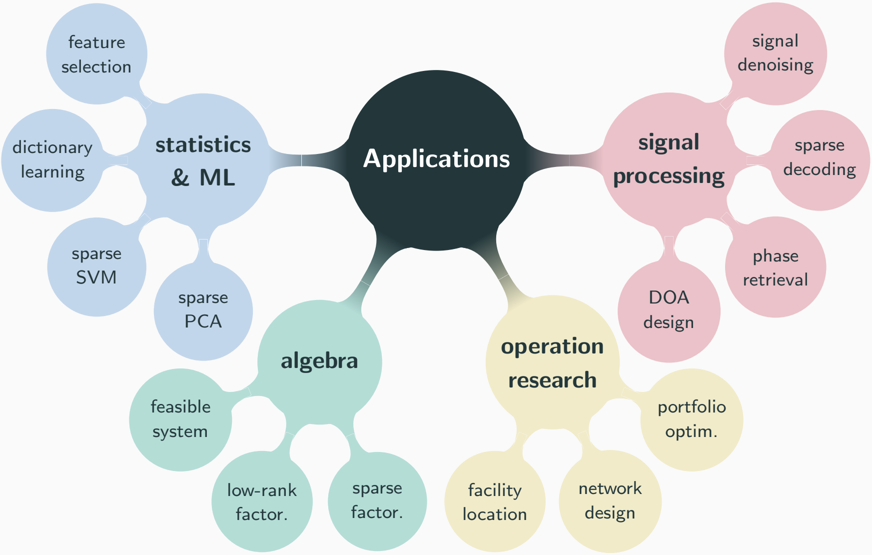
small loss

few non-zeros

small penalty

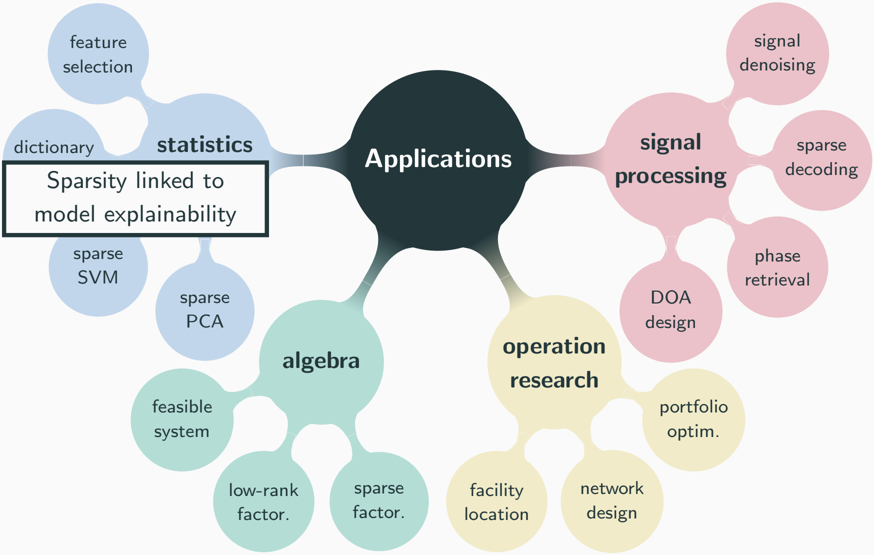
Solutions

One problem to rule them all



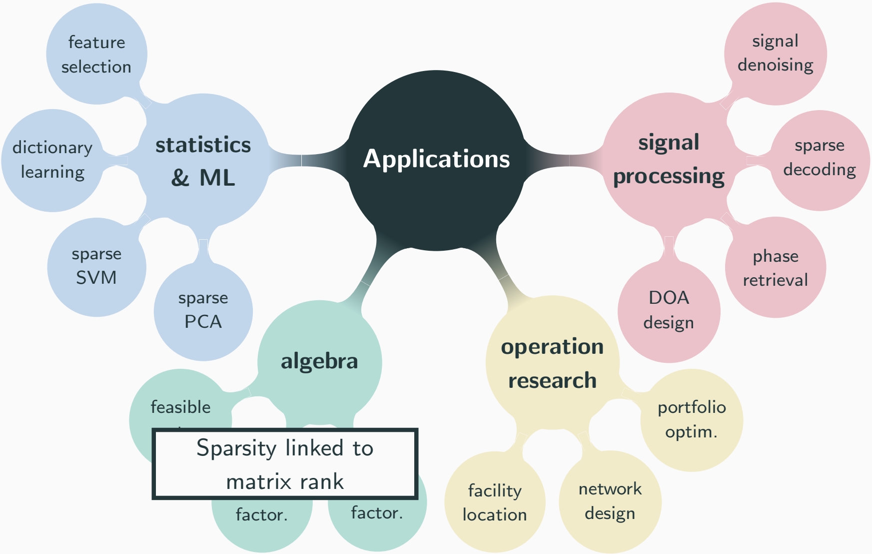
A. Tillmann *et. al* (2024)

One problem to rule them all



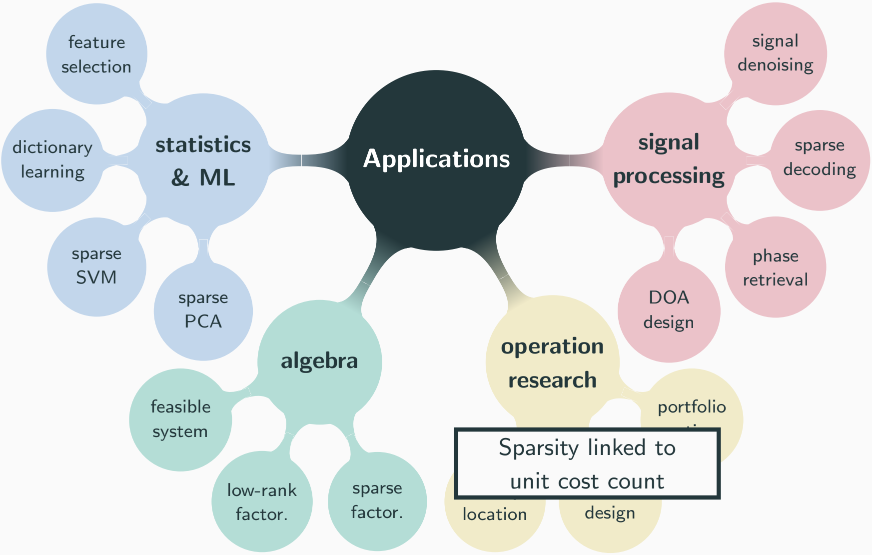
A. Tillmann *et. al* (2024)

One problem to rule them all



A. Tillmann *et. al* (2024)

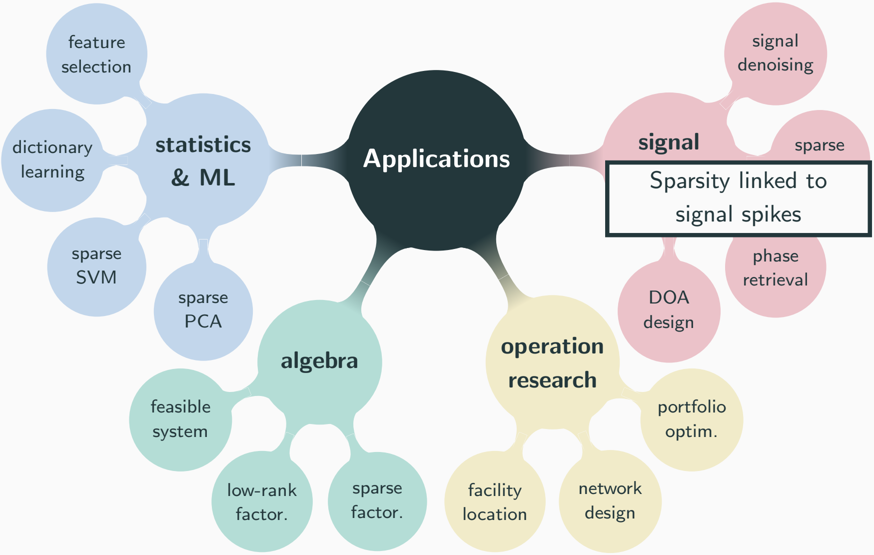
One problem to rule them all



Sparsity linked to unit cost count

A. Tillmann *et. al* (2024)

One problem to rule them all



A. Tillmann *et. al* (2024)

Solution methods

Problem

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

Question

$\boxed{\text{How to design generic and efficient solution methods ?}}$

Solution methods

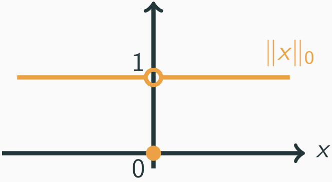
Problem

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

Question

How to design generic and efficient solution methods ?

ℓ_0-norm in 1d



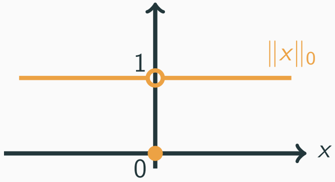
Problem

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

Question

How to design generic and efficient solution methods ?

ℓ_0-norm in 1d



Properties

- non-convex
- non-smooth
- non-continuous
- ...

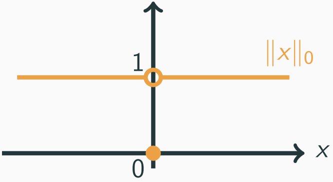
Problem

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

Question

How to design generic and efficient solution methods ?

ℓ_0-norm in 1d



Properties

- non-convex
- non-smooth
- non-continuous
- ...

$\downarrow$

Can be tackled using generic MIP solvers

D. Bertsimas *et. al* (2016)

Numerics

Problem

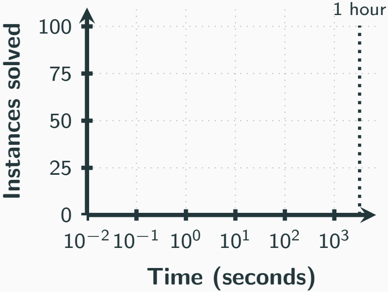
$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

$\mathbf{A} \in \mathbf{R}^{100 \times 300}$ / f : Quadratic / h : Bound constraint

Problem

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

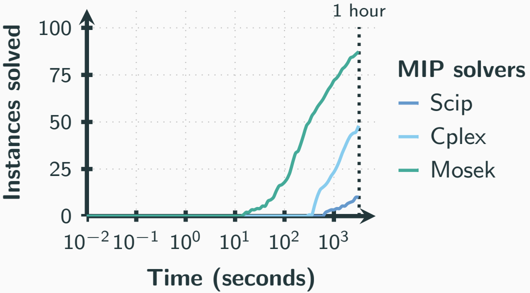
$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Quadratic / h : Bound constraint



Problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Quadratic / h : Bound constraint

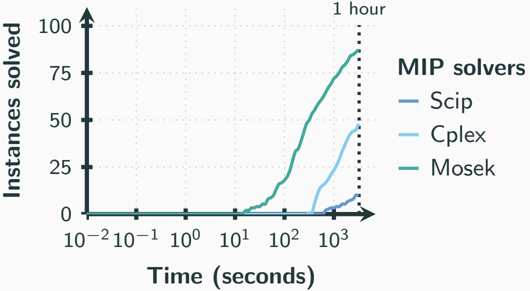


Numerics

Problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Quadratic / h : Bound constraint

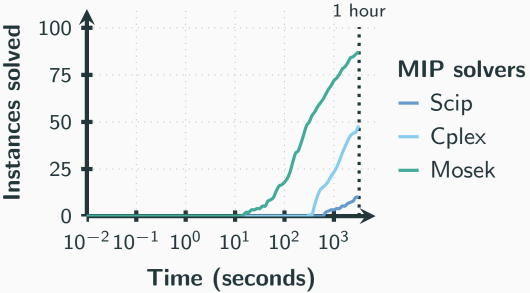


Can we do better ?

Problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Quadratic / h : Bound constraint



Can we do better ?

↓

BnB solvers

Specialized for the problem at hand

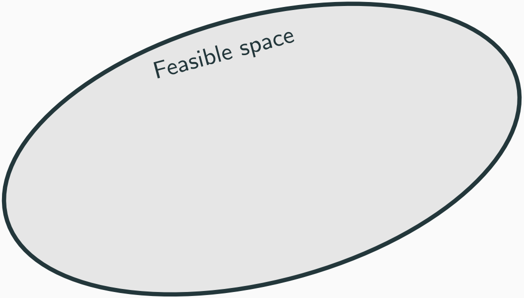
Branch-and-Bound Algorithms

BnB – Algorithmic principle

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

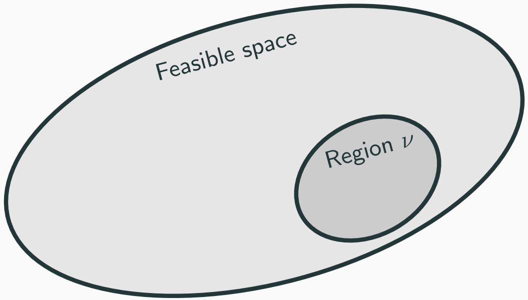
BnB – Algorithmic principle

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



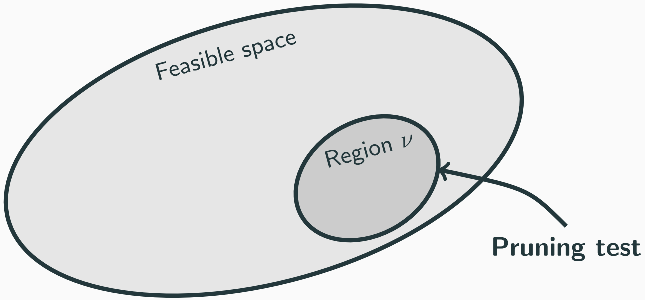
BnB – Algorithmic principle

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



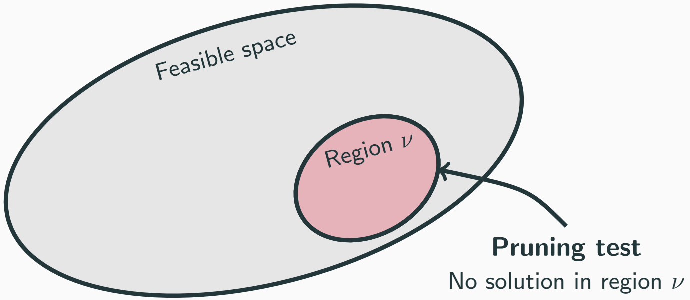
BnB – Algorithmic principle

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



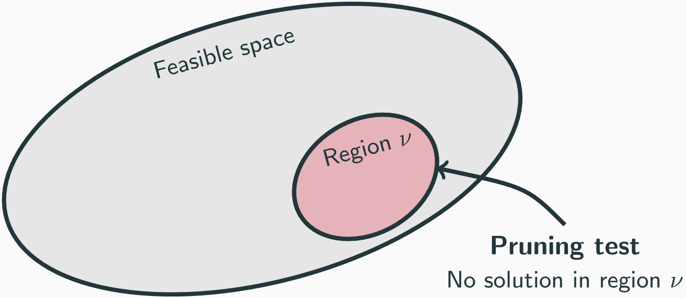
BnB – Algorithmic principle

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



BnB – Algorithmic principle

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



Branching – Region management
Bounding – Pruning test evaluation

BnB – How to construct regions in the feasible space ?

Problem

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

BnB – How to construct regions in the feasible space ?

Problem

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

Observation

Solutions expected to be sparse

BnB – How to construct regions in the feasible space ?

Problem

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

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable

BnB – How to construct regions in the feasible space ?

Problem

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

Observation

Solutions expected to be sparse

$\mathbf{R}^n$

$\downarrow$

ν_0

Method

Drive the sparsity of the optimization variable

BnB – How to construct regions in the feasible space ?

Problem

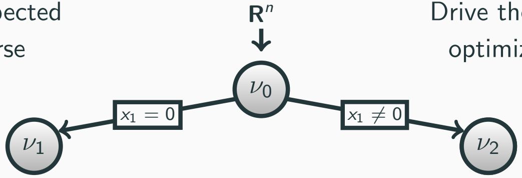
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable



BnB – How to construct regions in the feasible space ?

Problem

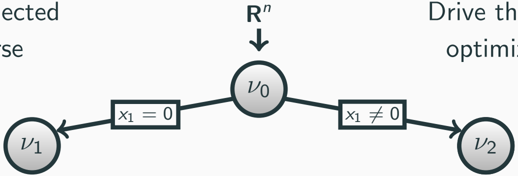
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable



Pruning test
Solutions in ν_1 ?

Pruning test
Solutions in ν_2 ?

Problem

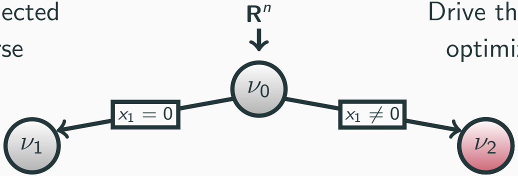
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable



Pruning test
Solutions in ν_1 ?

Pruning test
Solutions in ν_2 ?
No

BnB – How to construct regions in the feasible space ?

Problem

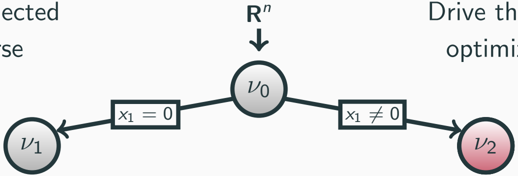
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable



Pruning test
Solutions in ν_1 ?
Maybe

Pruning test
Solutions in ν_2 ?
No

Problem

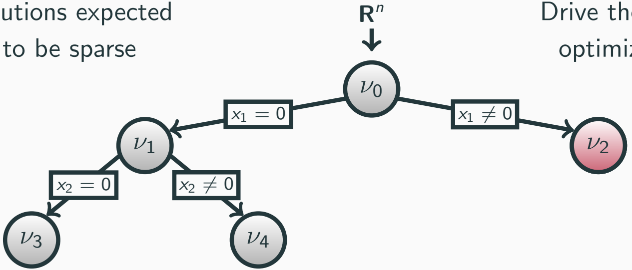
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable



Problem

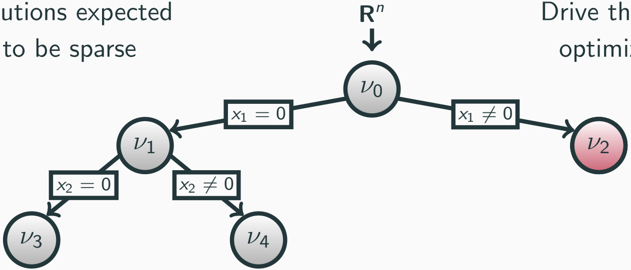
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable



Pruning test
Solutions in ν_3 ?

Pruning test
Solutions in ν_4 ?

Problem

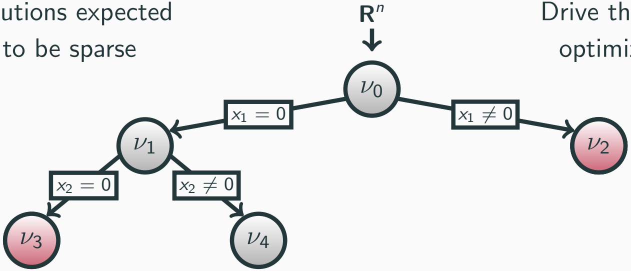
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable



Pruning test
Solutions in ν_3 ?
No

Pruning test
Solutions in ν_4 ?
Maybe

BnB – How to construct regions in the feasible space ?

Problem

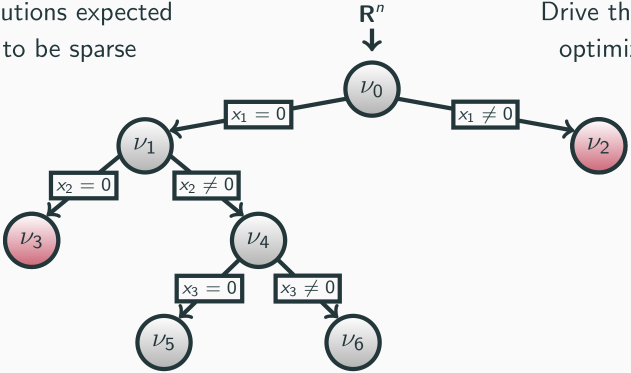
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable



BnB – How to construct regions in the feasible space ?

Problem

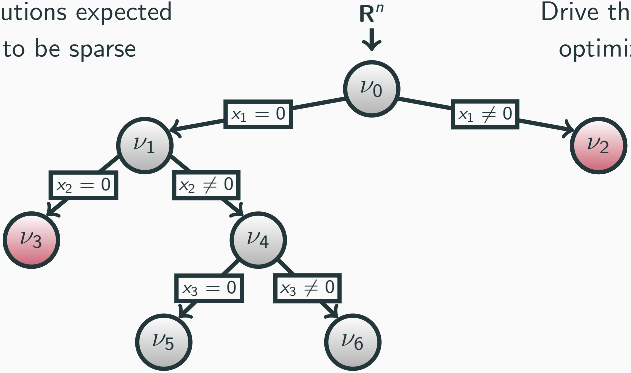
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable



Pruning test
Solutions in ν_5 ?

Pruning test
Solutions in ν_6 ?

BnB – How to construct regions in the feasible space ?

Problem

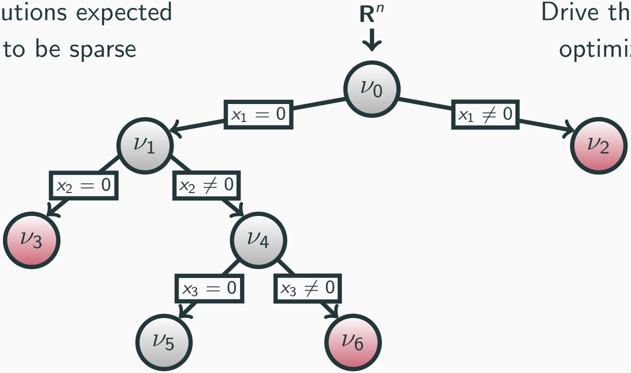
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable



Pruning test
Solutions in ν_5 ?
Maybe

Pruning test
Solutions in ν_6 ?
No

BnB – How to construct regions in the feasible space ?

Problem

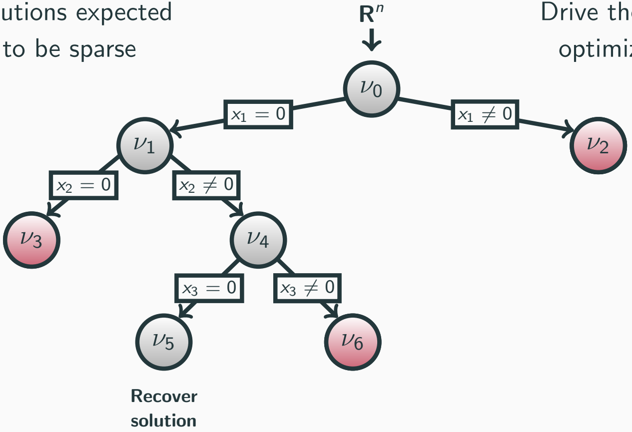
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions expected to be sparse

Method

Drive the sparsity of the optimization variable

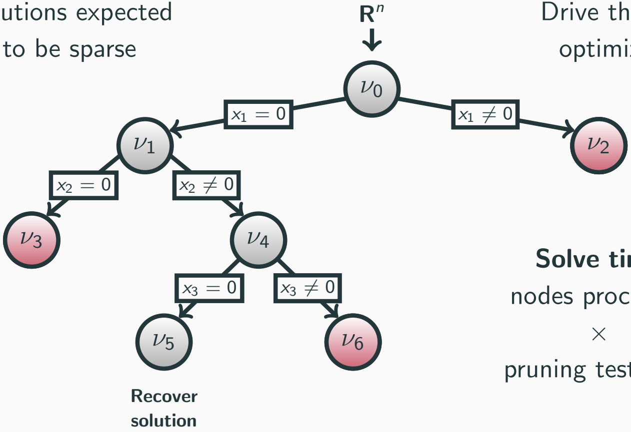


Problem

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

Observation

Solutions expected to be sparse



Method

Drive the sparsity of the optimization variable

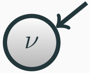
Solve time

nodes processed

$\times$

pruning test time

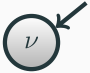
BnB – How to evaluate a pruning test ?



Pruning test

Solutions in region ν ?

BnB – How to evaluate a pruning test ?



Pruning test

Solutions in region ν ?

Problem

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

BnB – How to evaluate a pruning test ?

ν

Pruning test

Solutions in region ν ?

Problem

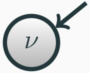
$$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

restrict to ν

Restriction to region ν

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

BnB – How to evaluate a pruning test ?



Pruning test

Solutions in region ν ?

Problem

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

restrict to ν

Restriction to region ν

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

compare

Pruning test

$$p^\nu > p^\star$$

ν

Pruning test

Solutions in region ν ?

Problem

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

$\downarrow$ restrict to ν

Restriction to region ν

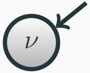
$$p^\nu = \min_{\mathbf{x} \in \nu} \ f(\mathbf{Ax}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$$

$\downarrow$ compare

Pruning test

$$p^\nu > p^\star$$

$\longrightarrow$ prune ν



Pruning test

Solutions in region ν ?

Problem

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

$\downarrow$ $\boxed{\text{restrict to } \nu}$

Restriction to region ν

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

$\downarrow$ $\boxed{\text{compare}}$

Pruning test

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

$\longrightarrow$ prune ν

BnB – How to evaluate a pruning test ?

ν

Pruning test

Solutions in region ν ?

Problem

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

restrict to ν

Restriction to region ν

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

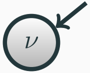
compare

Pruning test

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

Easy task

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



Pruning test

Solutions in region ν ?

Problem

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

$\downarrow$ restrict to ν

Restriction to region ν

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

$\downarrow$ compare

Pruning test

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

Easy task

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

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

BnB – How to evaluate a pruning test ?

ν

Pruning test

Solutions in region ν ?

Problem

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

restrict to ν

Restriction to region ν

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

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

BnB – Standard lower bounding strategy

Restriction to region ν

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

BnB – Standard lower bounding strategy

Restriction to region ν

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

Requirement 1 – Tight lower bound on p^ν
Requirement 2 – Tractable lower bound on p^ν

BnB – Standard lower bounding strategy

Restriction to region ν

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

Requirement 1 – Tight lower bound on p^{ν}

Requirement 2 – Tractable lower bound on p^{ν}

$\downarrow$

Standard strategy

Construct and solve a relaxation

BnB – Standard lower bounding strategy

Restriction to region ν

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

Requirement 1 – Tight lower bound on p^ν

Requirement 2 – Tractable lower bound on p^ν

$\downarrow$

Standard strategy

Construct and solve a relaxation

BnB – Standard lower bounding strategy

Restriction to region ν

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

Requirement 1 – Tight lower bound on p^{ν}

Requirement 2 – Tractable lower bound on p^{ν}

$\downarrow$

Standard strategy

Construct and solve a relaxation

Relaxation for region ν

$$p^{\nu}_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\mathrm{lb}}(\mathbf{x})$$

BnB – Standard lower bounding strategy

Restriction to region ν

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

Requirement 1 – Tight lower bound on p^{ν}

Requirement 2 – Tractable lower bound on p^{ν}

↓

Standard strategy

Construct and solve a relaxation

Relaxation for region ν

$$p^{\nu}_{\text{lb}} = \min_{\mathbf{x} \in \nu} f(\mathbf{A}\mathbf{x}) + g_{\text{lb}}(\mathbf{x})$$

$g_{\text{lb}} \leq g$

g_{lb} convex

BnB – Standard lower bounding strategy

Restriction to region ν

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

Requirement 1 – Tight lower bound on p^ν

Requirement 2 – Tractable lower bound on p^ν

$\downarrow$

Standard strategy

Construct and solve a relaxation

Relaxation for region ν

$$p^\nu_{\text{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\text{lb}}(\mathbf{x})$$

$g_{\text{lb}} \leq g$

g_{lb} convex

Lower bound

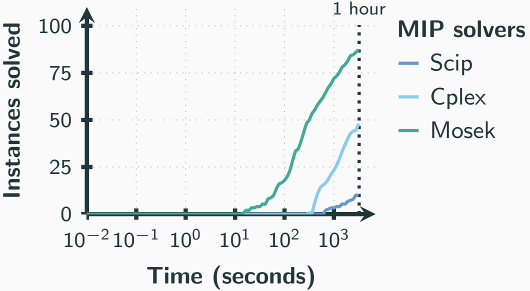
$$\boxed{p^\nu_{\text{lb}} \leq p^\nu}$$

BnB – Numerics

Problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Quadratic / h : Bound constraint

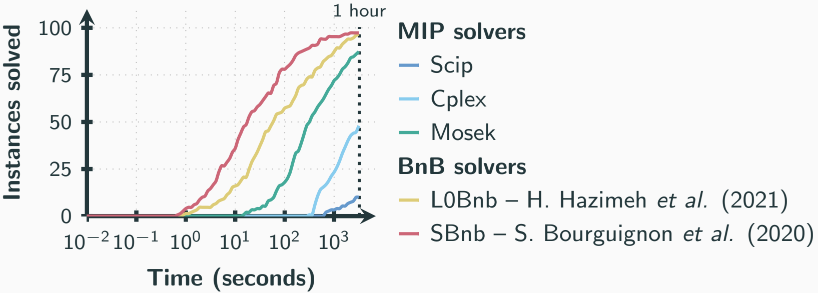


BnB – Numerics

Problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Quadratic / h : Bound constraint

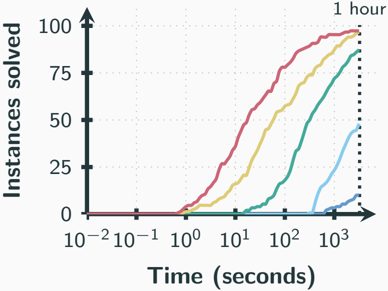


BnB – Numerics

Problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Quadratic / h : Bound constraint



MIP solvers

- Scip
- Cplex
- Mosek

BnB solvers

- L0Bnb – H. Hazimeh *et al.* (2021)
- SBnb – S. Bourguignon *et al.* (2020)

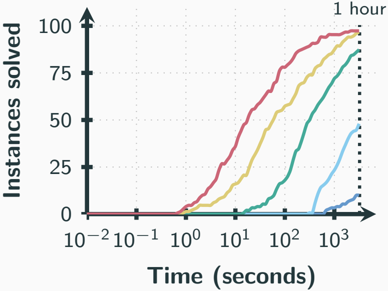
Observation

Better performance with BnB solvers

BnB – Numerics

Problem

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



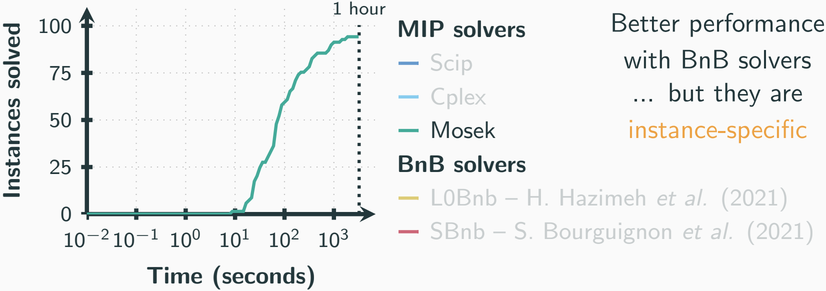
Observation

Better performance with BnB solvers ... but they are instance-specific

Problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Logistic / h : Bound cstr. + ℓ_1-norm



Observation

Better performance with BnB solvers ... but they are instance-specific

Let's recap

Problem

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

$\rightarrow g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\rightarrow$ many applications

Let's recap

Problem

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

BnB algorithm

Restriction to region ν

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

→ $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

→ many applications

→ pruning test

→ lower bound on p^ν

Let's recap

Problem

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

→ $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

→ many applications

↓ BnB algorithm

Restriction to region ν

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

→ pruning test

→ lower bound on p^ν

↓ standard strategy

Relaxation for region ν

$p^\nu_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\mathrm{lb}}(\mathbf{x})$

→ instance-specific construction of g_{lb}

Let's recap

Problem

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

→ $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

→ many applications

↓ BnB algorithm

Restriction to region ν

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

→ pruning test

→ lower bound on p^ν

↓ standard strategy

Relaxation for region ν

$p^\nu_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\mathrm{lb}}(\mathbf{x})$

→ instance-specific construction of g_{lb}

Axis 1

How to construct relaxations generically ?

Manuscript – Chap. 3

→ Journal paper (202x)

Axis 1 – How to construct relaxations generically ?

Axis 1 – Generic relaxation construction

Restriction to region ν

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

$g(\mathbf{x}) = \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

lower bound on p^{ν}

Axis 1 – Generic relaxation construction

Restriction to region ν

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

$g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$
lower bound on p^{ν}

$\downarrow$ standard strategy

Relaxation for region ν

$p^{\nu}_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\mathrm{lb}}(\mathbf{x})$

$g_{\mathrm{lb}} \leq g$
g_{lb} convex

Axis 1 – Generic relaxation construction

Restriction to region ν

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

$g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

lower bound on p^ν

$\boxed{\text{standard strategy}}$

Relaxation for region ν

$p^\nu_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\mathrm{lb}}(\mathbf{x})$

$g_{\mathrm{lb}} \leq g$

g_{lb} convex

h : ℓ_2-norm	h : bound cstr.	h : ℓ_2-norm + bound cstr.
M. Pilanci *et al.* (2016)	S. Bourguignon *et al.* (2020)	H. Hazimeh *et al.* (2021)

Axis 1 – Generic relaxation construction

Restriction to region ν

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

$g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$
lower bound on p^ν

$\boxed{\text{standard strategy}}$

Relaxation for region ν

$p^\nu_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\mathrm{lb}}(\mathbf{x})$

$g_{\mathrm{lb}} \leq g$
g_{lb} convex

h : ℓ_2-norm	h : bound cstr.	h : ℓ_2-norm + bound cstr.
M. Pilanci *et al.* (2016)	S. Bourguignon *et al.* (2020)	H. Hazimeh *et al.* (2021)

Generalization by setting $g_{\mathrm{lb}} = g_{\mathrm{cvx}}$

Restriction to region ν

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

$g(\mathbf{x}) = \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$
lower bound on p^{ν}

standard strategy

Relaxation for region ν

$p^{\nu}_{\mathrm{lb}} = \min_{\mathbf{x} \in \nu} f(\mathbf{A}\mathbf{x}) + g_{\mathrm{lb}}(\mathbf{x})$

$g_{\mathrm{lb}} \leq g$
g_{lb} convex

h : ℓ_2-norm	h : bound cstr.	h : ℓ_2-norm + bound cstr.
M. Pilanci *et al.* (2016)	S. Bourguignon *et al.* (2020)	H. Hazimeh *et al.* (2021)

Generalization by setting $g_{\mathrm{lb}} = g_{\mathrm{cvx}}$

$\rightarrow$ h proper, closed, convex

$\rightarrow$ h separable, even, supercoercive, $h \geq h(\mathbf{0}) = 0$

Axis 1 – Convex envelope characterization

Spotlight result

The convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ admits a closed-form expression.

Axis 1 – Convex envelope characterization

Spotlight result

The convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ admits a closed-form expression.

Theorem (1d version) – Let $g(x) = \lambda\|x\|_0 + h(x)$, one has

$$g_{\mathsf{cvx}}(x) = \begin{cases} \tau|x| & \text{if } |x| \leq \mu \\ h(x) + \lambda & \text{otherwise} \end{cases}$$

where (τ, μ) are some "easy-to-compute" quantities.

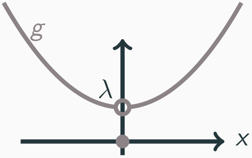
Axis 1 – Convex envelope characterization

Spotlight result

The convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ admits a closed-form expression.

Theorem (1d version) – Let $g(x) = \lambda\|x\|_0 + h(x)$, one has

$$g_{\mathrm{cvx}}(x) = \begin{cases} \tau|x| & \text{if } |x| \leq \mu \\ h(x) + \lambda & \text{otherwise} \end{cases}$$



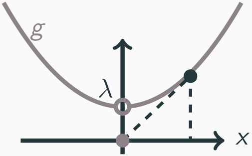
where (τ, μ) are some "easy-to-compute" quantities.

Spotlight result

The convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ admits a closed-form expression.

Theorem (1d version) – Let $g(x) = \lambda\|x\|_0 + h(x)$, one has

$$g_{\text{cvx}}(x) = \begin{cases} \tau|x| & \text{if } |x| \leq \mu \\ h(x) + \lambda & \text{otherwise} \end{cases}$$



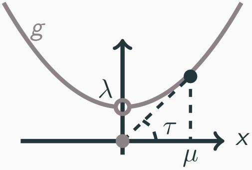
where (τ, μ) are some "easy-to-compute" quantities.

Spotlight result

The convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ admits a closed-form expression.

Theorem (1d version) – Let $g(x) = \lambda\|x\|_0 + h(x)$, one has

$$g_{\mathrm{cvx}}(x) = \begin{cases} \tau|x| & \text{if } |x| \leq \mu \\ h(x) + \lambda & \text{otherwise} \end{cases}$$



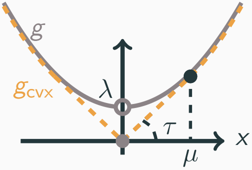
where (τ, μ) are some "easy-to-compute" quantities.

Spotlight result

The convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ admits a closed-form expression.

Theorem (1d version) – Let $g(x) = \lambda\|x\|_0 + h(x)$, one has

$$g_{\mathrm{cvx}}(x) = \begin{cases} \tau|x| & \text{if } |x| \leq \mu \\ h(x) + \lambda & \text{otherwise} \end{cases}$$



where (τ, μ) are some "easy-to-compute" quantities.

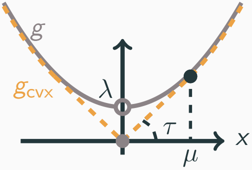
Axis 1 – Convex envelope characterization

Spotlight result

The convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ admits a closed-form expression.

Theorem (1d version) – Let $g(x) = \lambda\|x\|_0 + h(x)$, one has

$$g_{\mathsf{cvx}}(x) = \begin{cases} \tau|x| & \text{if } |x| \leq \mu \\ h(x) + \lambda & \text{otherwise} \end{cases}$$



where (τ, μ) are some "easy-to-compute" quantities.

↓

Generic relaxation construction

Characterize $g_{\mathsf{lb}} = g_{\mathsf{cvx}}$

Encompasses prior contributions

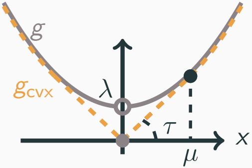
Axis 1 – Convex envelope characterization

Spotlight result

The convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ admits a closed-form expression.

Theorem (1d version) – Let $g(x) = \lambda\|x\|_0 + h(x)$, one has

$$g_{\mathsf{cvx}}(x) = \begin{cases} \tau|x| & \text{if } |x| \leq \mu \\ h(x) + \lambda & \text{otherwise} \end{cases}$$



where (τ, μ) are some "easy-to-compute" quantities.

↓

Generic relaxation construction

Characterize $g_{\mathsf{lb}} = g_{\mathsf{cvx}}$

Encompasses prior contributions

↓

Practical relaxation construction

Closed-form for $\partial g_{\mathsf{cvx}}$ and $\mathrm{prox}_{g_{\mathsf{cvx}}}$

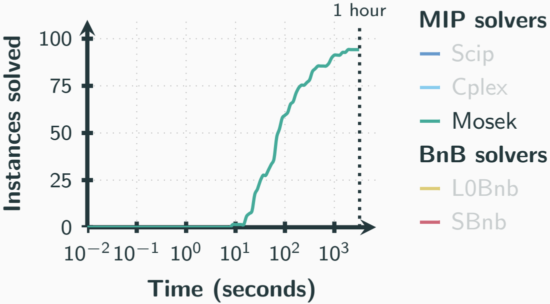
Enables standard solution methods

Axis 1 – Numerics

Problem

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

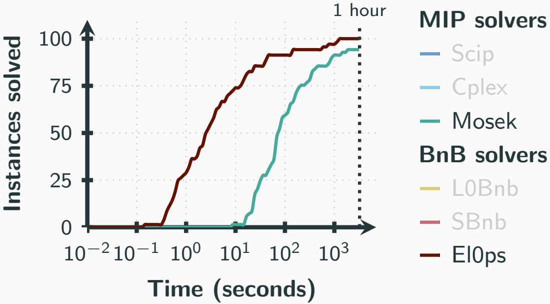
$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Logistic / h : Bound cstr. + ℓ_1-norm



Problem

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

$\mathbf{A} \in \mathbf{R}^{100 \times 300}$ / f : Logistic / h : Bound cstr. + ℓ_1-norm

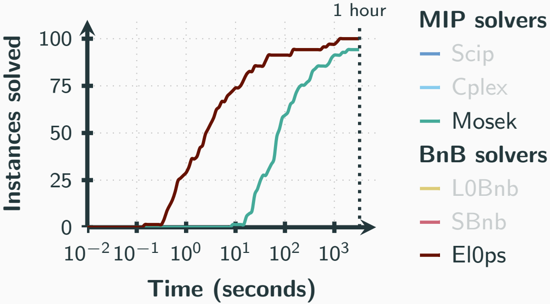


Axis 1 – Numerics

Problem

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

$\mathbf{A} \in \mathbf{R}^{100\times300}$ / f : Logistic / h : Bound cstr. + ℓ_1-norm



El0ps is a generic BnB solver with state-of-the-art performance

Let's recap

Problem

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

→ $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

→ many applications

↓ BnB algorithm

Restriction to region ν

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

→ pruning test

→ lower bound on p^ν

↓ standard strategy

Relaxation for region ν

$p^\nu_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\mathrm{lb}}(\mathbf{x})$

→ instance-specific construction of g_{lb}

Axis 1

How to construct relaxations generically ?

1) Set $g_{\mathrm{lb}} = g_{\mathrm{cvx}}$
2) Closed-form expression
3) Generalize BnB method

Let's recap

Problem

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

→ $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

→ many applications

↓ BnB algorithm

Restriction to region ν

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

→ pruning test

→ lower bound on p^ν

↓ standard strategy

Relaxation for region ν

$p^\nu_{\text{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\text{lb}}(\mathbf{x})$

→ instance-specific construction of g_{lb}

Axis 1

How to construct relaxations generically ?

1) Set $g_{\text{lb}} = g_{\text{cvx}}$
2) Closed-form expression
3) Generalize BnB method

Axis 2

How to solve relaxations efficiently ?

Manuscript – Chap. 6

→ ICASSP (2022)

→ Journal paper (202x)

Axis 2 – How to solve relaxations efficiently ?

Axis 2 – Convex optimization

Relaxation for region ν

$$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\mathsf{cvx}}(\mathbf{x})$$

Relaxation for region ν

$$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\mathrm{cvx}}(\mathbf{x})$$

$\updownarrow$ $\boxed{\text{reformulation}}$

Convex problem

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

Axis 2 – Convex optimization

Relaxation for region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{cvx}}(\mathbf{x})$

$\updownarrow$ $\boxed{\text{reformulation}}$

Convex problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax}) + \tilde{g}(\mathbf{x})$

- $\tilde{g}$ proper, closed, convex
- $\tilde{g}$ separable
- $\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

Axis 2 – Convex optimization

Relaxation for region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\text{cvx}}(\mathbf{x})$

remind g_{cvx} acts as an ℓ_1-norm near $\mathbf{x} = \mathbf{0}$

$\updownarrow$ $\boxed{\text{reformulation}}$

Convex problem

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

$\tilde{g}$ proper, closed, convex

$\tilde{g}$ separable

$\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

Relaxation for region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{cvx}}(\mathbf{x})$

remind g_{cvx} acts as an ℓ_1-norm near $\mathbf{x} = \mathbf{0}$

$\uparrow$ $\boxed{\text{reformulation}}$ $\downarrow$

Applications beyond the BnB scope

Convex problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax}) + \tilde{g}(\mathbf{x})$

$\tilde{g}$ proper, closed, convex

$\tilde{g}$ separable

$\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

Axis 2 – Convex optimization

Relaxation for region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\text{cvx}}(\mathbf{x})$

remind g_{cvx} acts as an ℓ_1-norm near $\mathbf{x} = \mathbf{0}$

$\Big\updownarrow$ $\boxed{\text{reformulation}}$

Applications beyond the BnB scope $\longrightarrow$

Convex problem

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

$\tilde{g}$ proper, closed, convex

$\tilde{g}$ separable

$\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

1st-order methods

Proximal gradient

Coordinate descent

...

iteration

$\mathbf{x}^0 \longrightarrow \boxed{\mathbf{x}^k \to \mathbf{x}^{k+1}} \longrightarrow \mathbf{x}^\star$

Axis 2 – Convex optimization

Relaxation for region ν

$\min_{\mathbf{x} \in \nu} f(\mathbf{A}\mathbf{x}) + g_{\text{cvx}}(\mathbf{x})$

remind g_{cvx} acts as an ℓ_1-norm near $\mathbf{x} = \mathbf{0}$

$\updownarrow$ reformulation

Applications beyond the BnB scope $\rightarrow$

Convex problem

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

$\tilde{g}$ proper, closed, convex

$\tilde{g}$ separable

$\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

1^{st}-order methods

Proximal gradient

Coordinate descent

...

iteration

$\mathbf{x}^0 \longrightarrow \boxed{\mathbf{x}^k \rightarrow \mathbf{x}^{k+1}} \longrightarrow \mathbf{x}^\star$

Solving cost

cost per iteration

$\times$

number of iterations

Axis 2 – Convex optimization

Relaxation for region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\text{cvx}}(\mathbf{x})$

remind g_{cvx} acts as an ℓ_1-norm near $\mathbf{x} = \mathbf{0}$

$\boxed{\text{reformulation}}$

Applications beyond the BnB scope

Convex problem

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

$\tilde{g}$ proper, closed, convex

$\tilde{g}$ separable

$\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

1st-order methods

Proximal gradient

Coordinate descent

...

iteration

$\mathbf{x}^0 \longrightarrow \boxed{\mathbf{x}^k \to \mathbf{x}^{k+1}} \longrightarrow \mathbf{x}^\star$

Solving cost

cost per iteration

$\times$

number of iterations

dimension of $\mathbf{x}$

Relaxation for region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\text{cvx}}(\mathbf{x})$

remind g_{cvx} acts as an ℓ_1-norm near $\mathbf{x} = \mathbf{0}$

$\uparrow$ reformulation $\downarrow$

Applications beyond the BnB scope $\rightarrow$

Convex problem

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

- $\tilde{g}$ proper, closed, convex
- $\tilde{g}$ separable
- $\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

1st-order methods

Proximal gradient

Coordinate descent

...

iteration

$\mathbf{x}^0 \longrightarrow \boxed{\mathbf{x}^k \to \mathbf{x}^{k+1}} \longrightarrow \mathbf{x}^\star$

Solving cost

cost per iteration

$\times$

number of iterations

dimension of $\mathbf{x}$

regularity of $f/\tilde{g}$

Axis 2 – Sparse structure exploitation

Convex problem

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

Task

$\boxed{\text{Reduce the solving cost}}$

Axis 2 – Sparse structure exploitation

Convex problem

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

Task

Reduce the solving cost

Option 1

Reduce the dimension of $\mathbf{x}$

Axis 2 – Sparse structure exploitation

Convex problem

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

$\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

$\Downarrow$

sparse solution $\mathbf{x}^\star$

Task

Reduce the solving cost

Option 1

Reduce the dimension of $\mathbf{x}$

Axis 2 – Sparse structure exploitation

Convex problem

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

$\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

sparse solution $\mathbf{x}^\star$

Task

Reduce the solving cost

$\downarrow$

Option 1

Reduce the dimension of $\mathbf{x}$

$\downarrow$

Screening tests

Identify zeros in $\mathbf{x}^\star$

L. El Ghaoui *et al.* (2011)

E. Ndiaye *et al.* (2020)

Axis 2 – Sparse structure exploitation

Convex problem

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

$\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

sparse solution $\mathbf{x}^\star$

Task

Reduce the solving cost

Option 1

Reduce the dimension of $\mathbf{x}$

$\downarrow$

Screening tests

Identify zeros in $\mathbf{x}^\star$

L. El Ghaoui *et al.* (2011)

E. Ndiaye *et al.* (2020)

Option 2

Improve the regularity of $f/\tilde{g}$

Axis 2 – Sparse structure exploitation

Convex problem

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

$\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

sparse solution $\mathbf{x}^\star$

Task

Reduce the solving cost

Option 1

Reduce the dimension of $\mathbf{x}$

$\downarrow$

Screening tests

Identify zeros in $\mathbf{x}^\star$

L. El Ghaoui *et al.* (2011)

E. Ndiaye *et al.* (2020)

Option 2

Improve the regularity of $f/\tilde{g}$

$\downarrow$

Smoothing tests

Identify non-zeros in $\mathbf{x}^\star$

Axis 2 – Sparse structure exploitation

Convex problem

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

$\tilde{g}$ non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

sparse solution $\mathbf{x}^\star$

Task

Reduce the solving cost

Option 1

Reduce the dimension of $\mathbf{x}$

$\downarrow$

Screening tests

Identify zeros in $\mathbf{x}^\star$

L. El Ghaoui *et al.* (2011)

E. Ndiaye *et al.* (2020)

Option 2

Improve the regularity of $f/\tilde{g}$

$\downarrow$

Smoothing tests

Identify non-zeros in $\mathbf{x}^\star$

Axis 2 – Screening and smoothing tests

Convex problem

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

Axis 2 – Screening and smoothing tests

Convex problem

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

Dual problem

$\max_{\mathbf{u}\in\mathbf{R}^m} D(\mathbf{u})$

Axis 2 – Screening and smoothing tests

Convex problem

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

solution $\mathbf{x}^\star$

if strong duality holds

$\mathbf{A}^{\mathrm{T}}\mathbf{u}^\star \in \partial\tilde{g}(\mathbf{x}^\star)$

solution $\mathbf{u}^\star$

Dual problem

$\max_{\mathbf{u}\in\mathbf{R}^m} D(\mathbf{u})$

Convex problem

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

$\downarrow$ solution $\mathbf{x}^\star$

if strong duality holds

$\mathbf{A}^{\mathrm{T}}\mathbf{u}^\star \in \partial\tilde{g}(\mathbf{x}^\star)$

$\uparrow$ solution $\mathbf{u}^\star$

Dual problem

$\max_{\mathbf{u}\in\mathbf{R}^m} D(\mathbf{u})$

Intermediate result

Some zeros and non-zeros in $\mathbf{x}^\star$ can be identified from a dual solution $\mathbf{u}^\star$.

Axis 2 – Screening and smoothing tests

Convex problem

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

solution $\mathbf{x}^\star$

if strong duality holds

$\mathbf{A}^{\mathrm{T}}\mathbf{u}^\star \in \partial\tilde{g}(\mathbf{x}^\star)$

solution $\mathbf{u}^\star$

Dual problem

$\max_{\mathbf{u}\in\mathbf{R}^m} D(\mathbf{u})$

Intermediate result

Some zeros and non-zeros in $\mathbf{x}^\star$ can be identified from a dual solution $\mathbf{u}^\star$.

Safe region

$\mathcal{R} \subseteq \mathbf{R}^m$ with $\mathbf{u}^\star \in \mathcal{R}$

Axis 2 – Screening and smoothing tests

Convex problem

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

solution $\mathbf{x}^\star$

if strong duality holds

$\mathbf{A}^{\mathrm{T}}\mathbf{u}^\star \in \partial\tilde{g}(\mathbf{x}^\star)$

solution $\mathbf{u}^\star$

Dual problem

$\max_{\mathbf{u}\in\mathbf{R}^m} D(\mathbf{u})$

Intermediate result

Some zeros and non-zeros in $\mathbf{x}^\star$ can be identified from a dual solution $\mathbf{u}^\star$.

Safe region

$\mathcal{R} \subseteq \mathbf{R}^m$ with $\mathbf{u}^\star \in \mathcal{R}$

Spotlight result

Some zeros and non-zeros in $\mathbf{x}^\star$ can be identified from a safe region $\mathcal{R}$.

Axis 2 – Screening and smoothing tests

remind $\tilde{g}(\mathbf{x}) = \sum_{i=1}^{n} \tilde{g}_i(x_i)$

Convex problem

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

solution $\mathbf{x}^\star$

if strong duality holds

$\mathbf{A}^{\mathrm{T}}\mathbf{u}^\star \in \partial\tilde{g}(\mathbf{x}^\star)$

solution $\mathbf{u}^\star$

Dual problem

$\max_{\mathbf{u}\in\mathbf{R}^m} D(\mathbf{u})$

Intermediate result

Some zeros and non-zeros in $\mathbf{x}^\star$ can be identified from a dual solution $\mathbf{u}^\star$.

Safe region

$\mathcal{R} \subseteq \mathbf{R}^m$ with $\mathbf{u}^\star \in \mathcal{R}$

Spotlight result

Some zeros and non-zeros in $\mathbf{x}^\star$ can be identified from a safe region $\mathcal{R}$.

remind $\tilde{g}(\mathbf{x}) = \sum_{i=1}^{n} \tilde{g}_i(x_i)$

Convex problem

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

$\downarrow$ solution $\mathbf{x}^\star$

if strong duality holds

$\mathbf{A}^{\mathrm{T}}\mathbf{u}^\star \in \partial\tilde{g}(\mathbf{x}^\star)$

$\uparrow$ solution $\mathbf{u}^\star$

Dual problem

$\max_{\mathbf{u} \in \mathbf{R}^m} D(\mathbf{u})$

Intermediate result

Some zeros and non-zeros in $\mathbf{x}^\star$ can be identified from a dual solution $\mathbf{u}^\star$.

Safe region

$\mathcal{R} \subseteq \mathbf{R}^m$ with $\mathbf{u}^\star \in \mathcal{R}$

Spotlight result

Some zeros and non-zeros in $\mathbf{x}^\star$ can be identified from a safe region $\mathcal{R}$.

Theorem – Given a safe region $\mathcal{R}$, note $\mathbf{a}^{\mathrm{T}}\mathcal{R} = \left\{\mathbf{a}^{\mathrm{T}}\mathbf{u} \mid \mathbf{u} \in \mathcal{R}\right\}$, one has

$$\text{Screening test:} \quad \mathbf{a}_i^{\mathrm{T}}\mathcal{R} \subseteq \mathrm{int}(\partial\tilde{g}_i(0)) \implies x_i^\star = 0$$

$$\text{Smoothing test:} \quad \mathbf{a}_i^{\mathrm{T}}\mathcal{R} \subseteq \mathrm{cmpl}(\partial\tilde{g}_i(0)) \implies x_i^\star \neq 0$$

remind $\tilde{g}(\mathbf{x}) = \sum_{i=1}^{n} \tilde{g}_i(x_i)$

Convex problem

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

solution $\mathbf{x}^\star$

if strong duality holds

$$\mathbf{A}^{\mathrm{T}}\mathbf{u}^\star \in \partial\tilde{g}(\mathbf{x}^\star)$$

solution $\mathbf{u}^\star$

Dual problem

$\max_{\mathbf{u}\in\mathbf{R}^m} D(\mathbf{u})$

Intermediate result

Some zeros and non-zeros in $\mathbf{x}^\star$ can be identified from a dual solution $\mathbf{u}^\star$.

Safe region

$\mathcal{R} \subseteq \mathbf{R}^m$ with $\mathbf{u}^\star \in \mathcal{R}$

Spotlight result

Some zeros and non-zeros in $\mathbf{x}^\star$ can be identified from a safe region $\mathcal{R}$.

Theorem – Given a safe region $\mathcal{R}$, note $\mathbf{a}^{\mathrm{T}}\mathcal{R} = \{\mathbf{a}^{\mathrm{T}}\mathbf{u} \mid \mathbf{u} \in \mathcal{R}\}$, one has

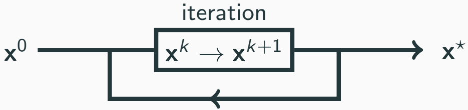
$$\text{Screening test:} \quad \mathbf{a}_i^{\mathrm{T}}\mathcal{R} \subseteq \operatorname{int}(\partial\tilde{g}_i(0)) \implies x_i^\star = 0$$

$$\text{Smoothing test:} \quad \underbrace{\mathbf{a}_i^{\mathrm{T}}\mathcal{R} \subseteq \operatorname{cmpl}(\partial\tilde{g}_i(0))}_{\text{easy to evaluate if } \mathcal{R} \text{ has a simple shape}} \implies x_i^\star \neq 0$$

Convex problem

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

1^{st}-order methods



Convex problem

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

Solving cost

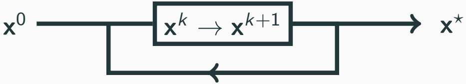
cost per iteration

$\times$

number of iterations

1st-order methods

iteration



Convex problem

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

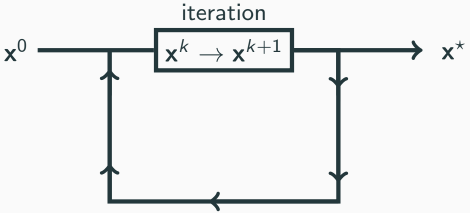
Solving cost

cost per iteration

$\times$

number of iterations

1^{st}-order methods



Axis 2 – Dynamic identification

Convex problem

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

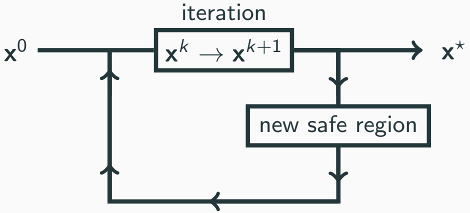
Solving cost

cost per iteration

$\times$

number of iterations

1^{st}-order methods



Axis 2 – Dynamic identification

Convex problem

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

Solving cost

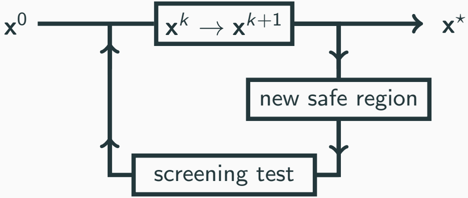
cost per iteration

$\times$

number of iterations

1^{st}-order methods

iteration



Identify some $x_i^\star = 0$

Axis 2 – Dynamic identification

Convex problem

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

Solving cost

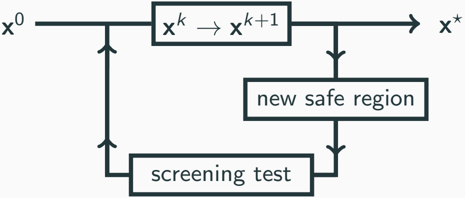
cost per iteration

$\times$

number of iterations

1^{st}-order methods

iteration



Identify some $x_i^\star = 0$

$\downarrow$

Set $x_i = 0$

Reduce cost

per iteration

Convex problem

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

Solving cost

cost per iteration

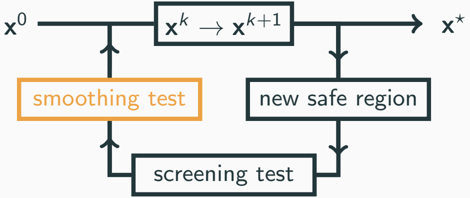
$\times$

number of iterations

1^{st}-order methods

iteration

Identify some $x_i^\star \neq 0$



Identify some $x_i^\star = 0$

$\downarrow$

Set $x_i = 0$

Reduce cost

per iteration

Axis 2 – Dynamic identification

Convex problem

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

Solving cost

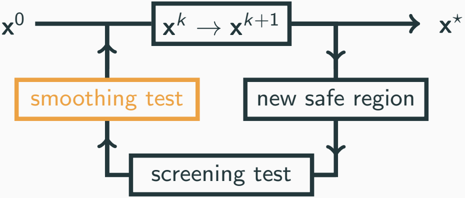
cost per iteration

$\times$

number of iterations

1^{st}-order methods

iteration



Identify some $x_i^\star \neq 0$

$\downarrow$

Smooth $\tilde{g}_i$, use

2^{nd}-order iterations

Reduce number

of iterations

Identify some $x_i^\star = 0$

$\downarrow$

Set $x_i = 0$

Reduce cost

per iteration

Axis 2 – Dynamic identification

Convex problem

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

Solving cost

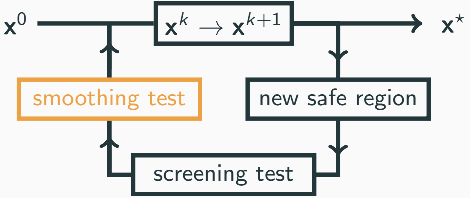
cost per iteration

$\times$

number of iterations

1^{st}-order methods

iteration



Identify some $x_i^\star \neq 0$

$\downarrow$

Smooth $\tilde{g}_i$, use

2nd-order iterations

Reduce number

of iterations

$\tilde{g}(\mathbf{x}) = \|\mathbf{x}\|_1 + \|\mathbf{x}\|_2^2$

$\downarrow$

$\tilde{g}(\mathbf{x}) = \mathbf{s}^{\mathrm{T}}\mathbf{x} + \|\mathbf{x}\|_2^2$

Identify some $x_i^\star = 0$

$\downarrow$

Set $x_i \ = \ 0$

Reduce cost

per iteration

Axis 2 – Dynamic identification

Convex problem

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

Solving cost

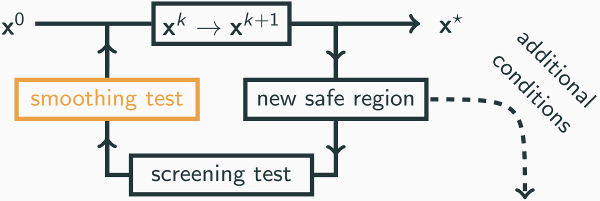
cost per iteration

$\times$

number of iterations

1^{st}-order methods

iteration



Identify some $x_i^\star \neq 0$

$\downarrow$

Smooth $\tilde{g}_i$, use

2^{nd}-order iterations

Reduce number

of iterations

$\tilde{g}(\mathbf{x}) = \|\mathbf{x}\|_1 + \|\mathbf{x}\|_2^2$

$\downarrow$

$\tilde{g}(\mathbf{x}) = \mathbf{s}^{\mathrm{T}}\mathbf{x} + \|\mathbf{x}\|_2^2$

Identify some $x_i^\star = 0$

$\downarrow$

Set $x_i = 0$

Reduce cost

per iteration

All zeros and non-zeros

identified after a finite

number of iterations

Axis 2 – Dynamic identification

Convex problem

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

Solving cost

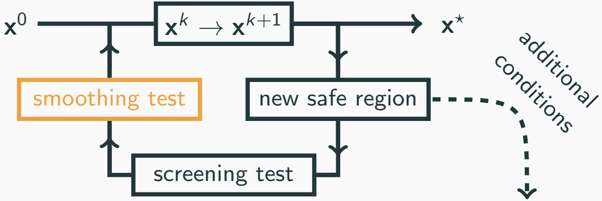
cost per iteration

$\times$

number of iterations

1^{st}-order methods

iteration



Identify some $x_i^\star \neq 0$

$\downarrow$

Smooth $\tilde{g}_i$, use

2^{nd}-order iterations

Reduce number

of iterations

$\tilde{g}(\mathbf{x}) = \|\mathbf{x}\|_1 + \|\mathbf{x}\|_2^2$

$\downarrow$

$\tilde{g}(\mathbf{x}) = \mathbf{s}^{\mathrm{T}}\mathbf{x} + \|\mathbf{x}\|_2^2$

Identify some $x_i^\star = 0$

$\downarrow$

Set $x_i = 0$

Reduce cost

per iteration

All zeros and non-zeros

identified after a finite

number of iterations

$\downarrow$

Smooth problem of

reduced dimension

Axis 2 – Numerics

Convex problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Logistic / $\tilde{g}$: Elastic-net

Convex problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Logistic / $\tilde{g}$: Elastic-net

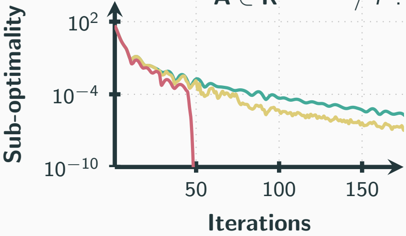
Accelerated proximal gradient

- Vanilla method
- With screening
- With screening and smoothing

Convex problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Logistic / $\tilde{g}$: Elastic-net



Accelerated proximal gradient

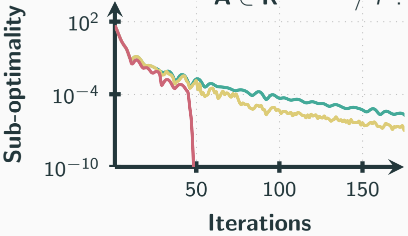
- Vanilla method
- With screening
- With screening and smoothing

Axis 2 – Numerics

Convex problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Logistic / $\tilde{g}$: Elastic-net



Accelerated proximal gradient

- Vanilla method
- With screening
- With screening and smoothing

1^{st}**-order** $\rightarrow$ 2^{nd}**-order**

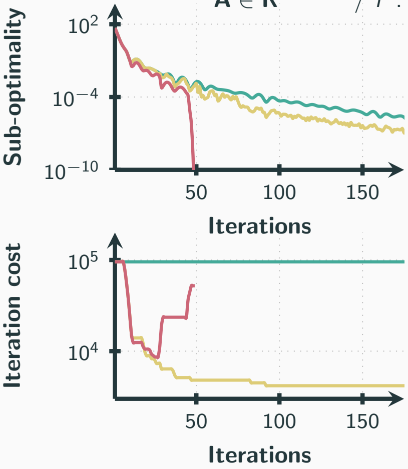
Faster convergence

Axis 2 – Numerics

Convex problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Logistic / $\tilde{g}$: Elastic-net



Accelerated proximal gradient

- Vanilla method
- With screening
- With screening and smoothing

1st-order → 2nd-order

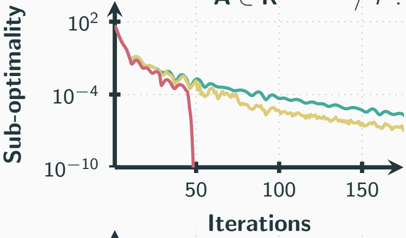
Faster convergence

More expensive iterations

Convex problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Logistic / $\tilde{g}$: Elastic-net



Accelerated proximal gradient

- Vanilla method
- With screening
- With screening and smoothing

1^{st}**-order** $\rightarrow$ 2^{nd}**-order**

Faster convergence

More expensive iterations

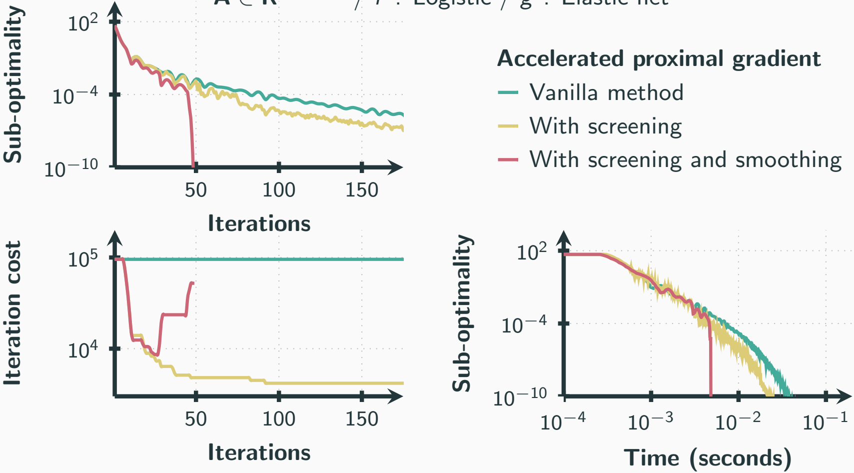
What about the solving time ?

Axis 2 – Numerics

Convex problem

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

$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Logistic / $\tilde{g}$: Elastic-net



Let's recap

Problem

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

→ $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

→ many applications

BnB algorithm

Restriction to region ν

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

→ pruning test

→ lower bound on p^ν

standard strategy

Relaxation for region ν

$p^\nu_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\mathrm{lb}}(\mathbf{x})$

→ instance-specific construction of g_{lb}

Axis 1

How to construct relaxations generically ?

1) Set $g_{\mathrm{lb}} = g_{\mathrm{cvx}}$
2) Closed-form expression
3) Generalize BnB method

Axis 2

How to solve relaxations efficiently ?

1) Cast as convex problem
2) Screening/smoothing
3) Reduce solving cost

Let's recap

Problem

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

→ $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

→ many applications

BnB algorithm

Restriction to region ν

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

→ pruning test

→ lower bound on p^ν

standard strategy

Relaxation for region ν

$p^\nu_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\mathrm{lb}}(\mathbf{x})$

→ instance-specific construction of g_{lb}

Axis 1

How to construct relaxations generically ?

1) Set $g_{\mathrm{lb}} = g_{\mathrm{cvx}}$
2) Closed-form expression
3) Generalize BnB method

Axis 2

How to solve relaxations efficiently ?

1) Cast as convex problem
2) Screening/smoothing
3) Reduce solving cost

Axis 3

How to improve the standard strategy ?

Manuscript – Chap. 4-5

→ ICASSP (2022)

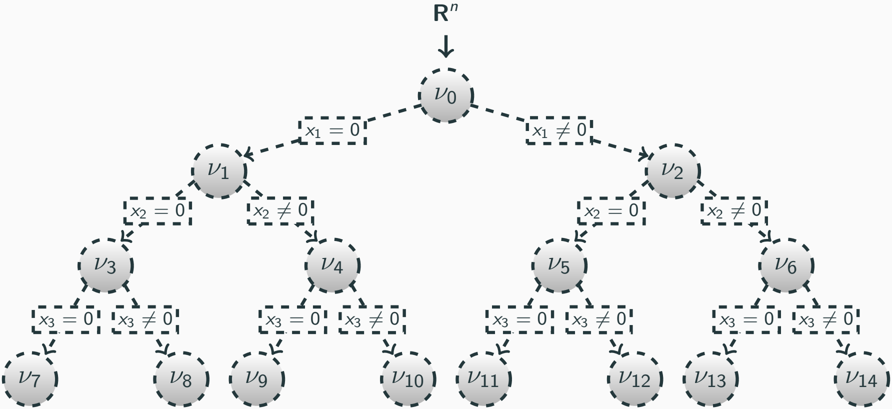
→ EUSIPCO (2023)

→ ICML (2024)

Axis 3 – How to improve the standard strategy ?

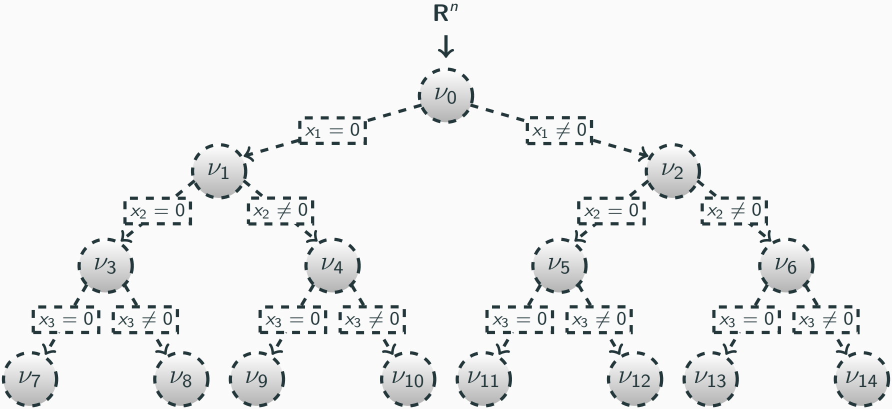
Problem

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



Problem

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

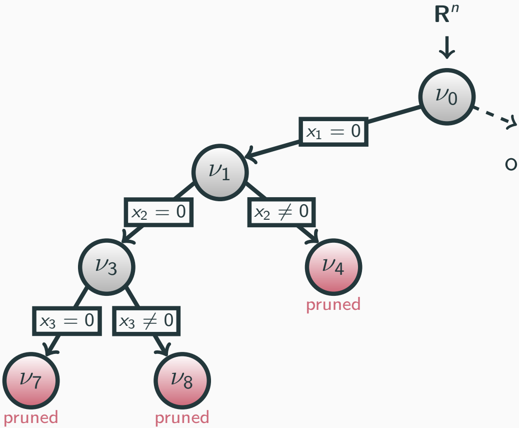


Solving time

$\boxed{\text{number of nodes explored} \times \text{pruning test time}}$

Problem

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



Standard strategy

one relaxation solved per node

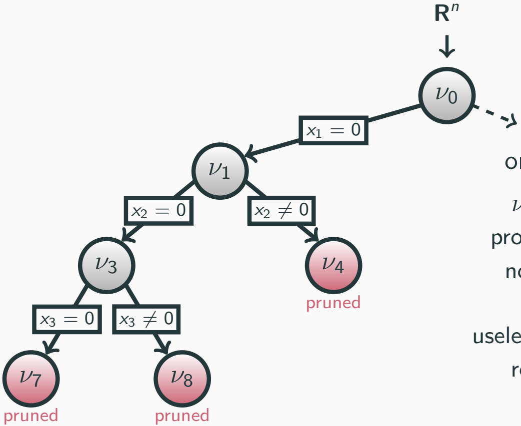
Solving time

$$\boxed{\text{number of nodes explored} \times \text{pruning test time}}$$

Axis 3 – Algorithmic complexity

Problem

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



Standard strategy

one relaxation solved per node

ν_1 and ν_3 processed but not pruned

↓

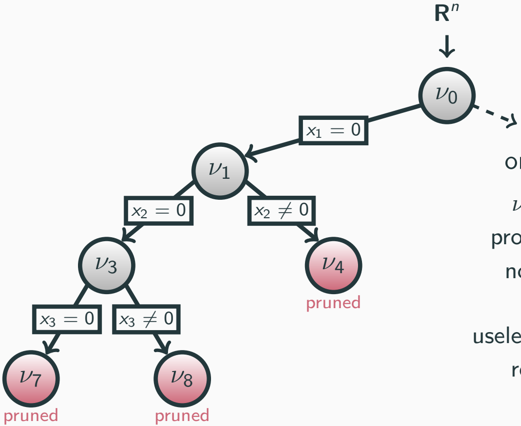
useless relaxation resolution

Solving time

$$\boxed{\text{number of nodes explored} \times \text{pruning test time}}$$

Problem

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



Standard strategy

one relaxation solved per node

ν_1 and ν_3 processed but not pruned

↓

useless relaxation resolution

ν_4, ν_7 and ν_8 pruned with rough lower bounds

↓

unnecessarily tight lower bounds

Solving time

$\boxed{\text{number of nodes explored} \times \text{pruning test time}}$

Axis 3 – Dual bounds

Question

Can we balance the complexity/tightness tradeoff when computing the lower bound on p^ν ?

Restriction to region ν

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

$\rightarrow g(\mathbf{x}) = \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

Axis 3 – Dual bounds

Question

Can we balance the complexity/tightness tradeoff
when computing the lower bound on p^ν ?

standard strat.

Restriction to region ν

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

→ $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\vee|$

Relaxation for region ν

$p^\nu_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\mathrm{lb}}(\mathbf{x})$

→ $p^\nu_{\mathrm{lb}} \leq p^\nu$

Axis 3 – Dual bounds

Question

Can we balance the complexity/tightness tradeoff when computing the lower bound on p^ν ?

standard strat.

dual problem

Restriction to region ν

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

$\rightarrow g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\geq$

Relaxation for region ν

$$p^\nu_{\text{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{lb}}(\mathbf{x})$$

$\rightarrow p^\nu_{\text{lb}} \leq p^\nu$

$\geq$

Dual relaxation for region ν

$$p^\nu_{\text{dual}} = \max_{\mathbf{u}\in\mathbf{R}^m} \; D^\nu(\mathbf{u})$$

Axis 3 – Dual bounds

Question

Can we balance the complexity/tightness tradeoff when computing the lower bound on p^ν ?

standard strat.

dual problem

Restriction to region ν

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

→ $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\vee|$

Relaxation for region ν

$$p^\nu_{\text{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\text{lb}}(\mathbf{x})$$

→ $p^\nu_{\text{lb}} \leq p^\nu$

$\vee|$

Dual relaxation for region ν

$$p^\nu_{\text{dual}} = \max_{\mathbf{u}\in\mathbf{R}^m} \; D^\nu(\mathbf{u})$$

→ $p^\nu_{\text{dual}} \leq p^\nu_{\text{lb}} \leq p^\nu$

Axis 3 – Dual bounds

Question

Can we balance the complexity/tightness tradeoff when computing the lower bound on p^ν ?

Restriction to region ν

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

$\rightarrow g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

standard strat.

$\vee\!|$

Relaxation for region ν

$$p^\nu_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\mathrm{lb}}(\mathbf{x})$$

$\rightarrow p^\nu_{\mathrm{lb}} \leq p^\nu$

dual problem

$\vee\!|$

Dual relaxation for region ν

$$p^\nu_{\mathrm{dual}} = \max_{\mathbf{u}\in\mathbf{R}^m} \; D^\nu(\mathbf{u})$$

$\rightarrow p^\nu_{\mathrm{dual}} \leq p^\nu_{\mathrm{lb}} \leq p^\nu$

Any evaluation of D^ν gives a lower-bound on p^ν

Axis 3 – Successor link

Standard lower bound

$$p^{\nu} \geq p^{\nu}_{\mathrm{lb}}$$

Tight but expensive

Dual lower bound

$$p^{\nu} \geq D^{\nu}(\mathbf{u})$$

Rough but economical

Axis 3 – Successor link

Standard lower bound	**Dual lower bound**
$p^\nu \geq p^\nu_{\text{lb}}$	$p^\nu \geq D^\nu(\mathbf{u})$
Tight but expensive	Rough but economical

A. Atamtürk *et al.* (2020) / G. Samain *et al.* (2023)

Axis 3 – Successor link

Standard lower bound	Dual lower bound
$p^\nu \geq p^\nu_{\text{lb}}$	$p^\nu \geq D^\nu(\mathbf{u})$
Tight but expensive	Rough but economical

A. Atamtürk *et al.* (2020) / G. Samain *et al.* (2023)

Spotlight result

Successor nodes in the BnB tree share similar dual lower bounds.

Axis 3 – Successor link

Standard lower bound

$p^\nu \geq p^\nu_{\mathrm{lb}}$

Tight but expensive

Dual lower bound

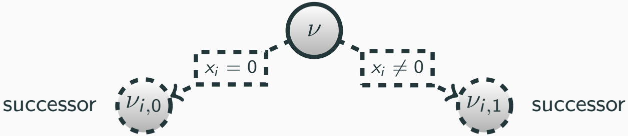
$p^\nu \geq D^\nu(\mathbf{u})$

Rough but economical

A. Atamtürk *et al.* (2020) / G. Samain *et al.* (2023)

Spotlight result

Successor nodes in the BnB tree share similar dual lower bounds.



Axis 3 – Successor link

Standard lower bound	Dual lower bound
$p^{\nu} \geq p^{\nu}_{\mathsf{lb}}$	$p^{\nu} \geq D^{\nu}(\mathbf{u})$

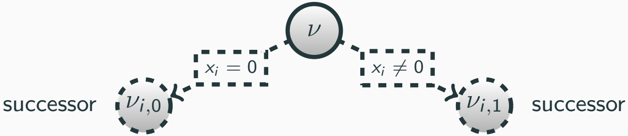
Tight but expensive

Rough but economical

A. Atamtürk *et al.* (2020) / G. Samain *et al.* (2023)

Spotlight result

Successor nodes in the BnB tree share similar dual lower bounds.



Theorem – Let $\nu_{i,b}$ with $b \in \{0, 1\}$ be a successor of node ν, then

$$D^{\nu_{i,b}}(\mathbf{u}) = D^{\nu}(\mathbf{u}) + \Delta^{i,b}(\mathbf{u})$$

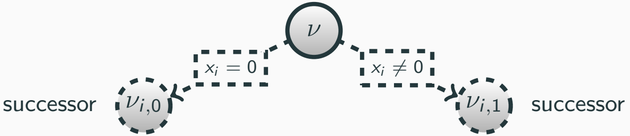
Axis 3 – Successor link

Standard lower bound	**Dual lower bound**
$p^\nu \geq p^\nu_{\mathrm{lb}}$	$p^\nu \geq D^\nu(\mathbf{u})$
Tight but expensive	Rough but economical

A. Atamtürk *et al.* (2020) / G. Samain *et al.* (2023)

Spotlight result

$\boxed{\text{Successor nodes in the BnB tree share similar dual lower bounds.}}$



Theorem – Let $\nu_{i,b}$ with $b \in \{0,1\}$ be a successor of node ν, then

$$D^{\nu_{i,b}}(\mathbf{u}) = \underbrace{D^\nu(\mathbf{u})}_{\text{Independent of } \nu_{i,b}} + \underbrace{\Delta^{i,b}(\mathbf{u})}_{\text{Virtually cost-free}}$$

Standard lower bound

$p^{\nu} \geq p^{\nu}_{\text{lb}}$

Tight but expensive

Dual lower bound

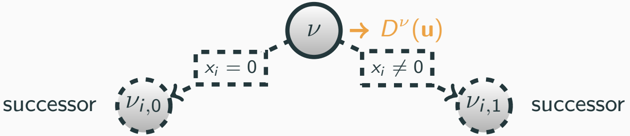
$p^{\nu} \geq D^{\nu}(\mathbf{u})$

Rough but economical

A. Atamtürk *et al.* (2020) / G. Samain *et al.* (2023)

Spotlight result

Successor nodes in the BnB tree share similar dual lower bounds.



Theorem – Let $\nu_{i,b}$ with $b \in \{0,1\}$ be a successor of node ν, then

$$D^{\nu_{i,b}}(\mathbf{u}) = \underbrace{D^{\nu}(\mathbf{u})}_{\text{Independent of } \nu_{i,b}} + \underbrace{\Delta^{i,b}(\mathbf{u})}_{\text{Virtually cost-free}}$$

Axis 3 – Successor link

Standard lower bound	**Dual lower bound**
$p^\nu \geq p^\nu_{\mathsf{lb}}$	$p^\nu \geq D^\nu(\mathbf{u})$

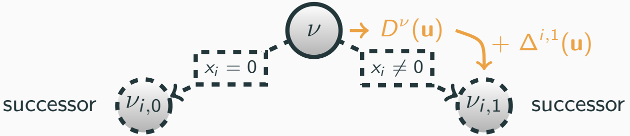
Tight but expensive

Rough but economical

A. Atamtürk *et al.* (2020) / G. Samain *et al.* (2023)

Spotlight result

Successor nodes in the BnB tree share similar dual lower bounds.



Theorem – Let $\nu_{i,b}$ with $b \in \{0,1\}$ be a successor of node ν, then

$$D^{\nu_{i,b}}(\mathbf{u}) = \underbrace{D^\nu(\mathbf{u})}_{\text{Independent of } \nu_{i,b}} + \underbrace{\Delta^{i,b}(\mathbf{u})}_{\text{Virtually cost-free}}$$

Standard pruning

Solve one relaxation per node

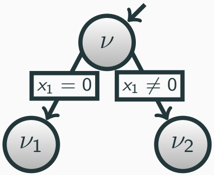
Select two successors to test next

Axis 3 – Paradigm shift

Standard pruning

Solve one relaxation per node

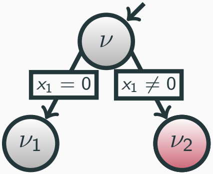
Select two successors to test next



Standard pruning

Solve one relaxation per node

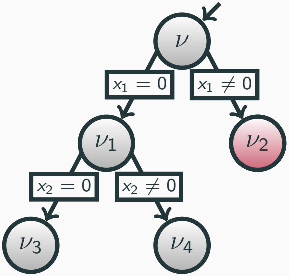
Select two successors to test next



Standard pruning

Solve one relaxation per node

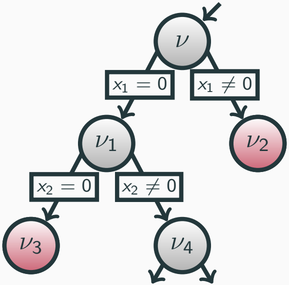
Select two successors to test next



Standard pruning

Solve one relaxation per node

Select two successors to test next

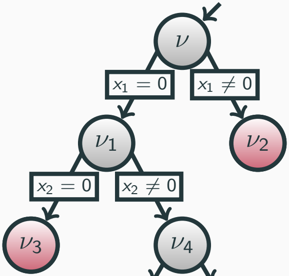


Axis 3 – Paradigm shift

Standard pruning

Solve one relaxation per node

Select two successors to test next



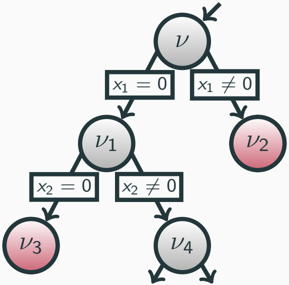
Slow and costly
tree expansion

Axis 3 – Paradigm shift

Standard pruning

Solve one relaxation per node

Select two successors to test next



Slow and costly

tree expansion

Simultaneous pruning

Obtain $D^\nu(\mathbf{u})$ during node processing

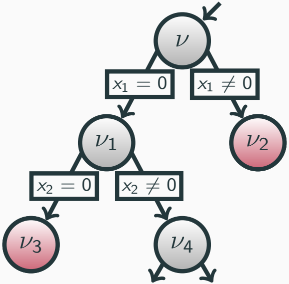
Test all successors with dual bounds

Axis 3 – Paradigm shift

Standard pruning

Solve one relaxation per node

Select two successors to test next

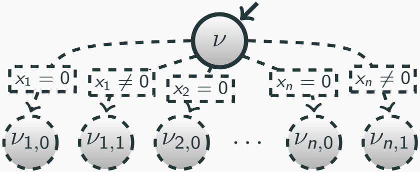


Slow and costly
tree expansion

Simultaneous pruning

Obtain $D^\nu(\mathbf{u})$ during node processing

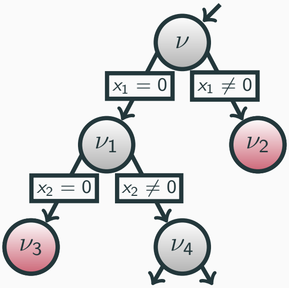
Test all successors with dual bounds



Standard pruning

Solve one relaxation per node

Select two successors to test next



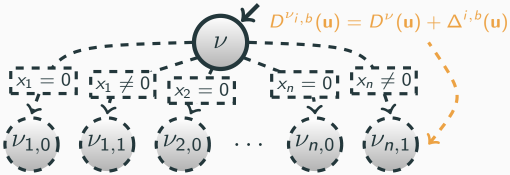
Slow and costly

tree expansion

Simultaneous pruning

Obtain $D^{\nu}(\mathbf{u})$ during node processing

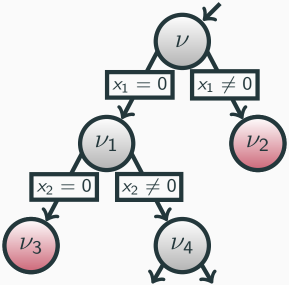
Test all successors with dual bounds



Standard pruning

Solve one relaxation per node

Select two successors to test next

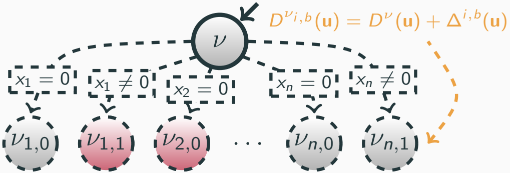


Slow and costly
tree expansion

Simultaneous pruning

Obtain $D^{\nu}(\mathbf{u})$ during node processing

Test all successors with dual bounds

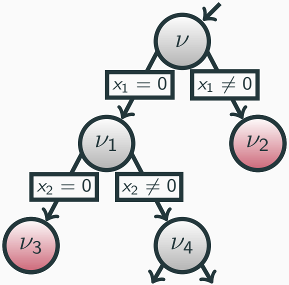


Axis 3 – Paradigm shift

Standard pruning

Solve one relaxation per node

Select two successors to test next

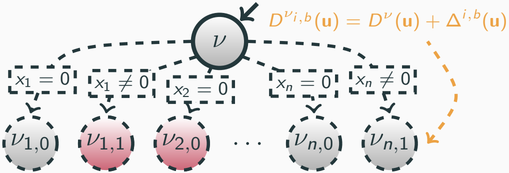


Slow and costly
tree expansion

Simultaneous pruning

Obtain $D^{\nu}(\mathbf{u})$ during node processing

Test all successors with dual bounds



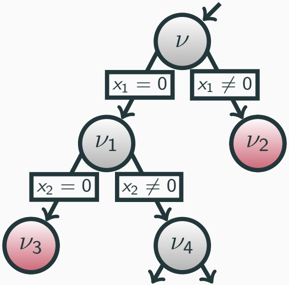
Tree expansion

Axis 3 – Paradigm shift

Standard pruning

Solve one relaxation per node

Select two successors to test next

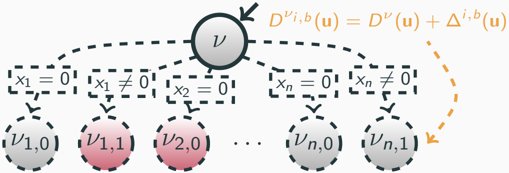


Slow and costly tree expansion

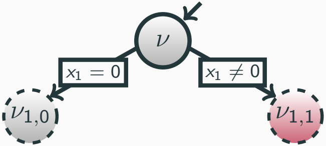
Simultaneous pruning

Obtain $D^{\nu}(\mathbf{u})$ during node processing

Test all successors with dual bounds



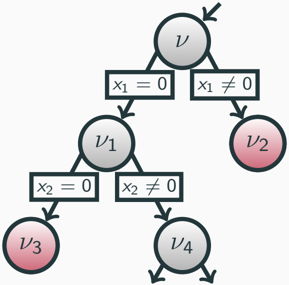
Tree expansion



Standard pruning

Solve one relaxation per node

Select two successors to test next

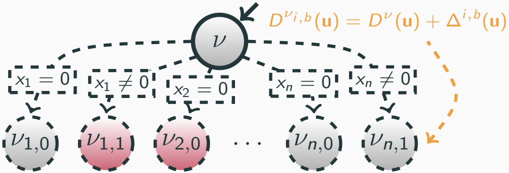


Slow and costly tree expansion

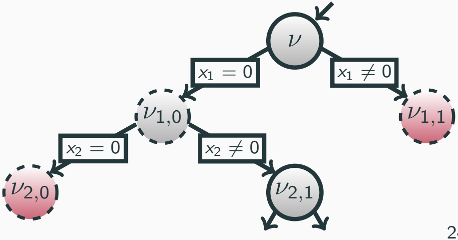
Simultaneous pruning

Obtain $D^{\nu}(\mathbf{u})$ during node processing

Test all successors with dual bounds



Tree expansion

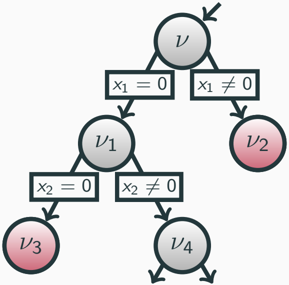


Axis 3 – Paradigm shift

Standard pruning

Solve one relaxation per node

Select two successors to test next

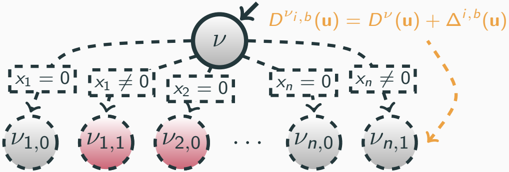


Slow and costly
tree expansion

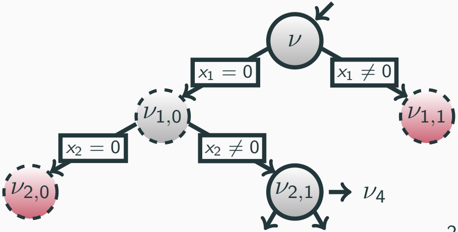
Simultaneous pruning

Obtain $D^{\nu}(\mathbf{u})$ during node processing

Test all successors with dual bounds



Tree expansion

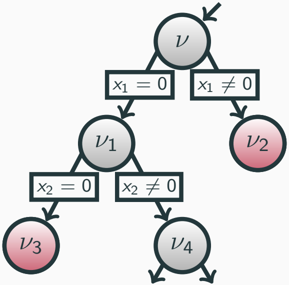


Axis 3 – Paradigm shift

Standard pruning

Solve one relaxation per node

Select two successors to test next



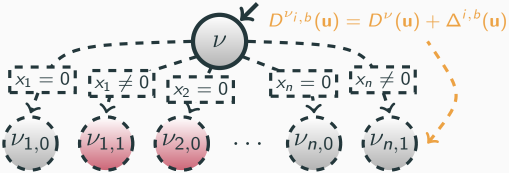
Slow and costly

tree expansion

Simultaneous pruning

Obtain $D^{\nu}(\mathbf{u})$ during node processing

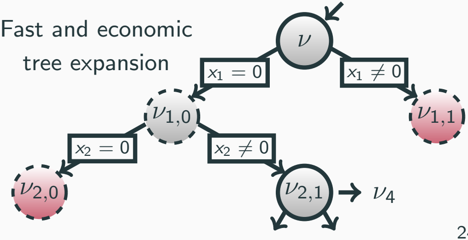
Test all successors with dual bounds



Tree expansion

Fast and economic

tree expansion



Axis 3 – Numerics

Problem

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

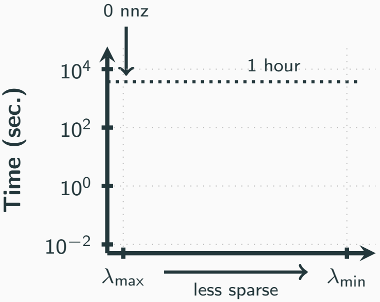
$\mathbf{A} \in \mathbf{R}^{62\times 2000}$ from ML dataset / f : Logistic / h : ℓ_2-norm

Axis 3 – Numerics

Problem

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

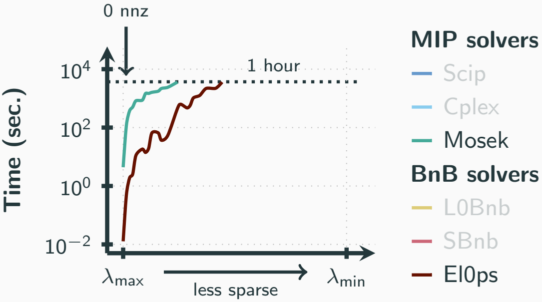
$\mathbf{A} \in \mathbf{R}^{62 \times 2000}$ from ML dataset / f : Logistic / h : ℓ_2-norm



Problem

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

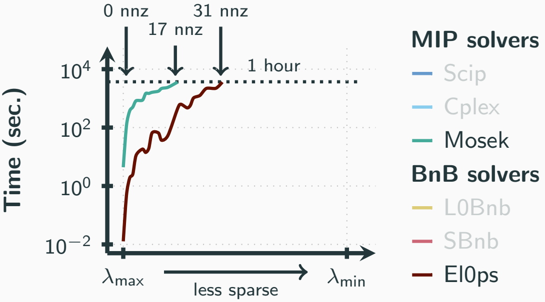
$\mathbf{A} \in \mathbf{R}^{62\times 2000}$ from ML dataset / f : Logistic / h : ℓ_2-norm



Problem

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

$\mathbf{A} \in \mathbf{R}^{62\times 2000}$ from ML dataset / f : Logistic / h : ℓ_2-norm

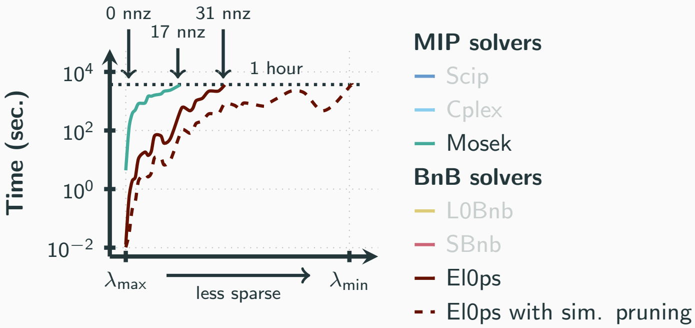


Axis 3 – Numerics

Problem

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

$\mathbf{A} \in \mathbf{R}^{62\times 2000}$ from ML dataset / f : Logistic / h : ℓ_2-norm

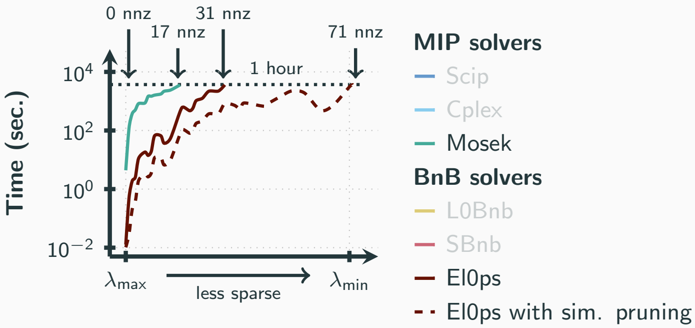


Axis 3 – Numerics

Problem

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

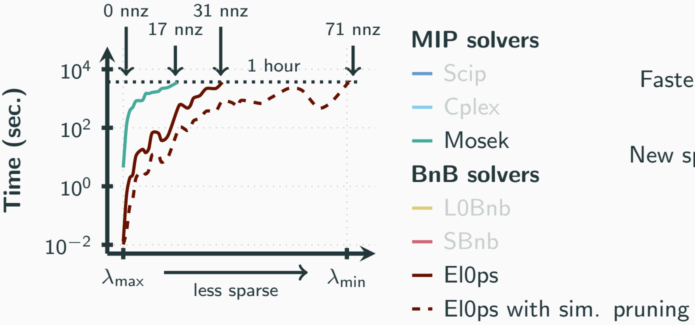
$\mathbf{A} \in \mathbf{R}^{62\times 2000}$ from ML dataset / f : Logistic / h : ℓ_2-norm



Problem

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

$\mathbf{A} \in \mathbf{R}^{62\times 2000}$ from ML dataset / f : Logistic / h : ℓ_2-norm



Faster solving time

↓

New sparsity regimes

Let's recap

Problem

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

→ $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

→ many applications

↓ BnB algorithm

Restriction to region ν

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

→ pruning test

→ lower bound on p^ν

↓ standard strategy

Relaxation for region ν

$p^\nu_{\mathrm{lb}} = \min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\mathrm{lb}}(\mathbf{x})$

→ instance-specific construction of g_{lb}

Axis 1

How to construct relaxations generically ?

1) Set $g_{\mathrm{lb}} = g_{\mathrm{cvx}}$
2) Closed-form expression
3) Generalize BnB method

Axis 2

How to solve relaxations efficiently ?

1) Cast as convex problem
2) Screening/smoothing
3) Accelerate solution

Axis 3

How to improve the standard strategy ?

1) Dual bound $D^\nu(\mathbf{u}) \leq p^\nu$
2) Link between successors
3) Change of paradigm

Conclusion

Contributions

Problem

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

minimize loss / sparse solutions

Contributions

Problem

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

minimize loss / sparse solutions

Question

How to design generic and efficient solution methods ?

Contributions

Problem

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

minimize loss / sparse solutions

Question

How to design generic and efficient solution methods ?

1) Generic solver

Axis 1

BnB solver with
generic framework

Contributions

Problem

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

minimize loss / sparse solutions

Question

How to design generic and efficient solution methods ?

1) Generic solver

Axis 1

BnB solver with generic framework

2) Efficient solver

Axis 2 & 3

Efficient relaxation solution, simultaneous pruning

Contributions

Problem

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

minimize loss / sparse solutions

Question

How to design generic and efficient solution methods ?

1) Generic solver

Axis 1

BnB solver with generic framework

2) Efficient solver

Axis 2 & 3

Efficient relaxation solution, simultaneous pruning

3) Practical solver

El0ps

Flexible with state-of-the-art performance

Perspectives

Perspectives

Extension to other formulations

Minimize loss $f(\mathbf{Ax})$

Force sparsity with $\|\mathbf{x}\|_0$

$\downarrow$ $\boxed{\text{contributions}}$

Regularized version

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

Perspectives

Extension to other formulations

Minimize loss $f(\mathbf{A}\mathbf{x})$

Force sparsity with $\|\mathbf{x}\|_0$

contributions

Regularized version

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

connections

Constrained version

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x})$ s.t. $\|\mathbf{x}\|_0 \leq k$

Extension to other formulations

Minimize loss $f(\mathbf{A}\mathbf{x})$

Force sparsity with $\|\mathbf{x}\|_0$

$\downarrow$ contributions $\downarrow$

Regularized version

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

$\uparrow$ connections $\downarrow$

Constrained version

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x})$ s.t. $\|\mathbf{x}\|_0 \leq k$

$\rightarrow$ non-separability of the ℓ_0-norm constraint

Perspectives

Extension to other formulations

Minimize loss $f(\mathbf{Ax})$

Force sparsity with $\|\mathbf{x}\|_0$

$\boxed{\text{contributions}}$

Regularized version

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

$\boxed{\text{connections}}$

Constrained version

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax})$ s.t. $\|\mathbf{x}\|_0 \leq k$

$\rightarrow$ non-separability of the ℓ_0-norm constraint

Towards stronger relaxations

Restriction to region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g(\mathbf{x})$

$\boxed{\text{lower bound}}$

Convex relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{cvx}}(\mathbf{x})$

Perspectives

Extension to other formulations

Minimize loss $f(\mathbf{Ax})$

Force sparsity with $\|\mathbf{x}\|_0$

$\downarrow$ contributions

Regularized version

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

$\updownarrow$ connections

Constrained version

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax})$ s.t. $\|\mathbf{x}\|_0 \leq k$

$\rightarrow$ non-separability of the ℓ_0-norm constraint

Towards stronger relaxations

Restriction to region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g(\mathbf{x})$

$\downarrow$ lower bound

Convex relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{cvx}}(\mathbf{x})$

$\downarrow$ improve

Non-cvx relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{non-cvx}}(\mathbf{x})$

Perspectives

Extension to other formulations

Minimize loss $f(\mathbf{Ax})$

Force sparsity with $\|\mathbf{x}\|_0$

$\boxed{\text{contributions}}$

Regularized version

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

$\boxed{\text{connections}}$

Constrained version

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax})$ s.t. $\|\mathbf{x}\|_0 \leq k$

$\rightarrow$ non-separability of the ℓ_0-norm constraint

Towards stronger relaxations

Restriction to region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g(\mathbf{x})$

$\boxed{\text{lower bound}}$

Convex relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{cvx}}(\mathbf{x})$

$\boxed{\text{improve}}$

Non-cvx relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{non-cvx}}(\mathbf{x})$

$\rightarrow$ tune $g_{\text{non-cvx}}$ to preserve the overall convexity

Perspectives

Extension to other formulations

Minimize loss $f(\mathbf{Ax})$

Force sparsity with $\|\mathbf{x}\|_0$

contributions

Regularized version

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

connections

Constrained version

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax})$ s.t. $\|\mathbf{x}\|_0 \leq k$

$\rightarrow$ non-separability of the ℓ_0-norm constraint

Towards stronger relaxations

Restriction to region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g(\mathbf{x})$

lower bound

Convex relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{cvx}}(\mathbf{x})$

improve

Non-cvx relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{non-cvx}}(\mathbf{x})$

$\rightarrow$ tune $g_{\text{non-cvx}}$ to preserve the overall convexity

Broader exploitation of smoothing tests

Convex problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax}) + \tilde{g}(\mathbf{x})$

accelerate

Screen/smooth tests

$x_i^\star = 0$ or $x_i^\star \neq 0$

Perspectives

Extension to other formulations

Minimize loss $f(\mathbf{Ax})$

Force sparsity with $\|\mathbf{x}\|_0$

contributions ↓

Regularized version

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

↑ connections ↓

Constrained version

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax})$ s.t. $\|\mathbf{x}\|_0 \leq k$

$\rightarrow$ non-separability of the ℓ_0-norm constraint

Towards stronger relaxations

Restriction to region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g(\mathbf{x})$

lower bound ↓

Convex relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{cvx}}(\mathbf{x})$

improve ↓

Non-cvx relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{Ax}) + g_{\text{non-cvx}}(\mathbf{x})$

$\rightarrow$ tune $g_{\text{non-cvx}}$ to preserve the overall convexity

Broader exploitation of smoothing tests

Convex problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{Ax}) + \tilde{g}(\mathbf{x})$

accelerate ↓

Screen/smooth tests

$x_i^\star = 0$ or $x_i^\star \neq 0$

↓

Set $x_i = 0$

Tailored to any instance

Perspectives

Extension to other formulations

Minimize loss $f(\mathbf{A}\mathbf{x})$
Force sparsity with $\|\mathbf{x}\|_0$

contributions

Regularized version

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

connections

Constrained version

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x})$ s.t. $\|\mathbf{x}\|_0 \leq k$

$\rightarrow$ non-separability of the ℓ_0-norm constraint

Towards stronger relaxations

Restriction to region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g(\mathbf{x})$

lower bound

Convex relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\text{cvx}}(\mathbf{x})$

improve

Non-cvx relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\text{non-cvx}}(\mathbf{x})$

$\rightarrow$ tune $g_{\text{non-cvx}}$ to preserve the overall convexity

Broader exploitation of smoothing tests

Convex problem

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

accelerate

Screen/smooth tests

$x_i^\star = 0$ or $x_i^\star \neq 0$

Set $x_i = 0$
Tailored to any instance

Smooth $\tilde{g}_i$
Depends on the instance

Perspectives

Extension to other formulations

Minimize loss $f(\mathbf{A}\mathbf{x})$

Force sparsity with $\|\mathbf{x}\|_0$

contributions

Regularized version

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

connections

Constrained version

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x})$ s.t. $\|\mathbf{x}\|_0 \leq k$

$\rightarrow$ non-separability of the ℓ_0-norm constraint

Towards stronger relaxations

Restriction to region ν

$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g(\mathbf{x})$

lower bound

Convex relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\text{cvx}}(\mathbf{x})$

improve

Non-cvx relaxation

$\min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g_{\text{non-cvx}}(\mathbf{x})$

$\rightarrow$ tune $g_{\text{non-cvx}}$ to preserve the overall convexity

Broader exploitation of smoothing tests

Convex problem

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

accelerate

Screen/smooth tests

$x_i^\star = 0$ or $x_i^\star \neq 0$

Set $x_i = 0$ — Tailored to any instance

Smooth $\tilde{g}_i$ — Depends on the instance

$\rightarrow$ Newton accel. for proximal identification

G. Bareilles *et al.* (2022)

Question time !

🤓

Context – Machine learning application

Tabular ML dataset

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

Model explicability

Sparsity-inducing $\mathcal{R}(\mathbf{x})$

Optimization problem

$\min_{\mathbf{x}\in\mathbf{R}^n} \mathcal{L}_\phi(\mathbf{A}\mathbf{x}, \mathbf{y}) + \lambda\mathcal{R}(\mathbf{x})$

Context – Algebra application

Sparse Component Analysis

Goal

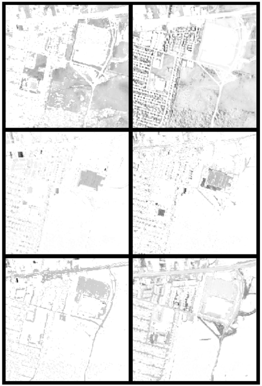
Given $\mathbf{M} \in \mathbf{R}^{m\times n}$, find $\mathbf{D} \in \mathbf{R}^{m\times r}$ and $\mathbf{B} \in \mathbf{R}^{r\times n}$ such that $\mathbf{M} \simeq \mathbf{DB}$ with sparse columns in $\mathbf{B}$.

$\downarrow$

Optimization problem

$$\min_{\mathbf{D}\in\mathbf{R}^{m\times r},\mathbf{B}\in\mathbf{R}^{r\times n}} \frac{1}{2}\|\mathbf{M}-\mathbf{DB}\|_F^2 + \lambda \sum_{i=1}^{n} \|\mathbf{b}_i\|_0$$

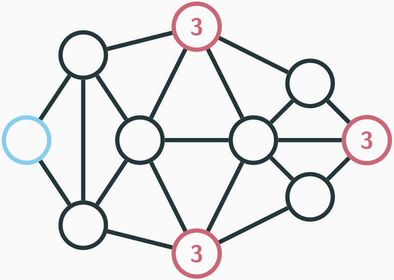
J. Cohen, N. Gillis (2019)



Extract material abundance map from hyperspectral image

Context – Operation research application

Max. capacity per edge: 10
Edge construction cost: 5



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

Network design

$$\begin{cases} \min\ Q(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 \\ \text{s.t. } \mathbf{D}\mathbf{x} \leq \mathbf{d},\ \ \mathbf{x} \leq \mathbf{c} \\ \mathbf{x} \in \mathbf{R}_+^{\mathrm{card}(E)} \end{cases}$$

Q : transportation cost

λ : unit construction cost

$\mathbf{D}\mathbf{x} \leq \mathbf{d}$: flow conservation

$\mathbf{x} \leq \mathbf{c}$: capacity constraint

Context – Signal processing application

Compressed sensing

Sparse signal

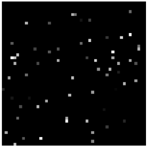
$\mathbf{x} \in \mathbf{R}^n$

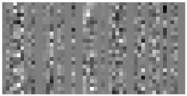
Linear measurement

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

Observation

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





Goal

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

$m \ll n$: no unique solution

Goal (with sparse prior)

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

Optimization problem

$\min_{\mathbf{x}\in\mathbf{R}^n} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$

sparsity-inducing function

Sparse optimization problem

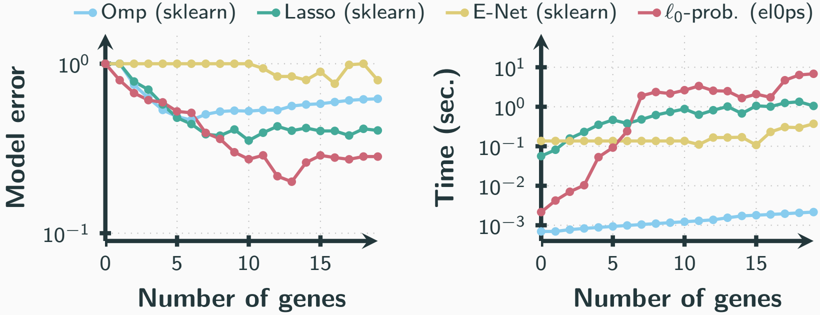
$\min_{\mathbf{x}\in\mathbf{R}^n} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0$

Context – 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



Context – Balancing solution quality and problem hardness

Sparse signal

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

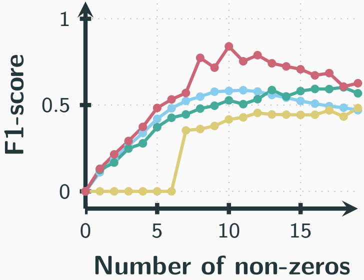
Observation

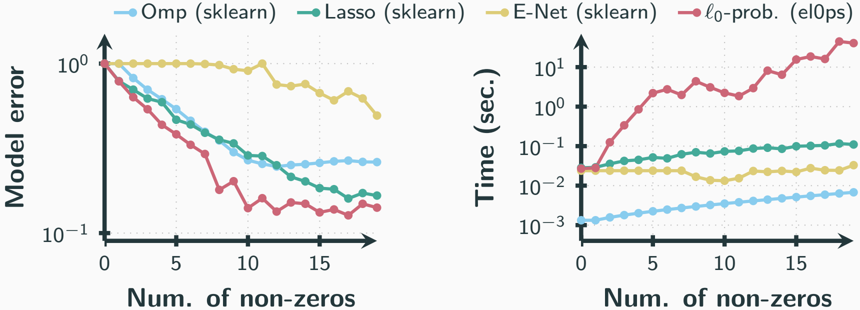
$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon} \in \mathbf{R}^m$

Setup

$\mathbf{A} \in \mathbf{R}^{100\times 200}$ highly correlated, $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma\mathbf{I})$,

$\mathbf{x}$ with 10 unit spikes of random location





Context – Balancing solution quality and problem hardness

Sparse signal

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

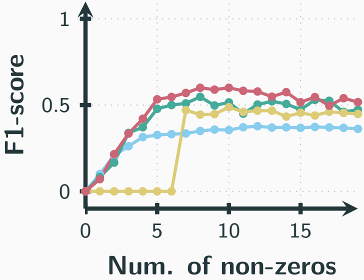
Observation

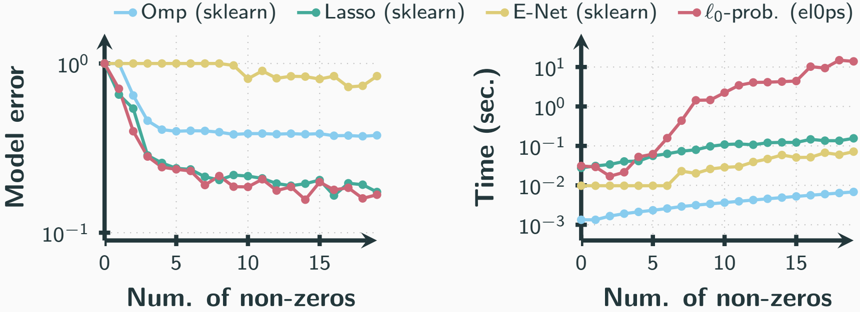
$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon} \in \mathbf{R}^m$

Setup

$\mathbf{A} \in \mathbf{R}^{100 \times 200}$ highly correlated, $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma\mathbf{I})$,

$\mathbf{x}$ with 10 unit and evenly-spaced spikes



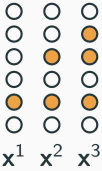


Context – A bit of history

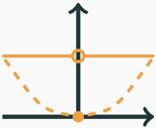
Problem

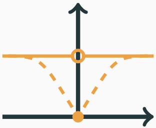
$$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda \|\mathbf{x}\|_0$$

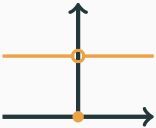
1995	2000	2005	2010	2015
Heuristics	Recovery cond.	Convex approx.	Concave approx.	Exact methods
MP, OMP, ...	RIP, NSP, ...	Lasso, Elastic-Net, ...	SCAD, MCP, ...	MIP, BnB, ...
S. Mallat (1993)	E. Candes (2004)	R. Tibshirani (2005)	C. Zhang (2010)	D. Bertsimas (2016)



OMP solves ℓ_0-problem under RIP







Problem

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

linearize ℓ_0-norm

MIP formulation

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

avoid logical cstr.

Practical MIP formulation

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

Use generic MIP solvers

Need standardized expressions

linear/quadratic/conic/...

Lifted formulation

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

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

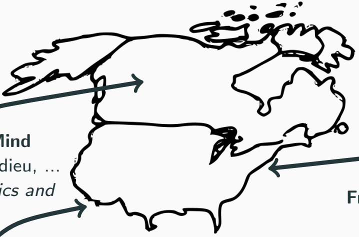
if $x_i = 0 \implies z_i = 0, \ \forall i$

Construct h_{mip} depending on h

$h(\mathbf{x})$	$h_{\mathrm{mip}}(\mathbf{x},\mathbf{z})$
$\eta(\|\mathbf{x}\|_\infty \leq M)$	$\eta(-M\mathbf{z} \leq \mathbf{x} \leq M\mathbf{x})$
$\alpha\|\mathbf{x}\|_2^2$	$\sum_{i=1}^n \alpha\frac{x_i^2}{z_i}$

Context – Research community

Non-exhaustive list



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

Google Deep Mind
H. Hazimeh, A. Dedieu, ...
MIO-based heuristics and softwares

Lund University
M. Carlsson, C. Olsson...
Quadratic envelope

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

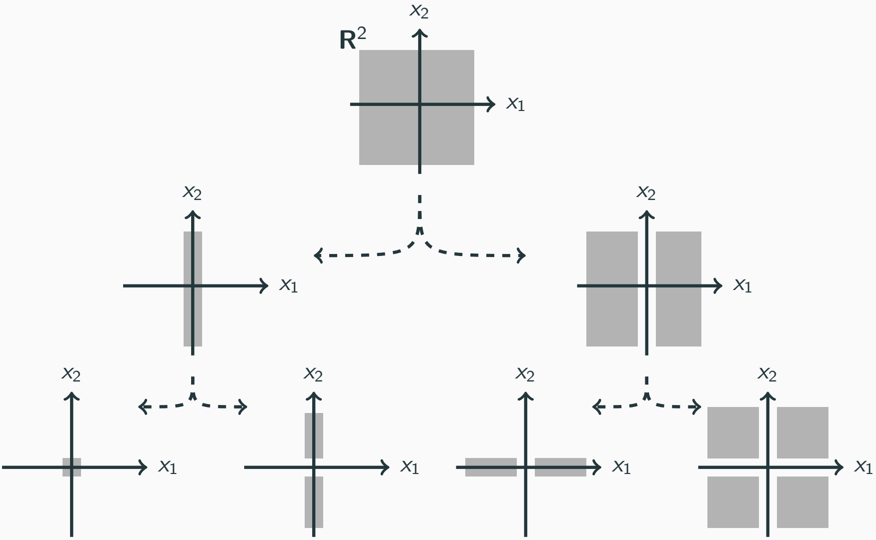
Ponts ParisTech
M. De Lara, P. Chancelier, A. Parmentier, ...
Non-convex analysis for ℓ_0-norm, ML appli.

Centrale Nantes / ENSTA Bretagne
S. Bourguignon, J. Ninin, ...
Branch-and-Bound for ℓ_0-problems

Inria / CentraleSupélec
C. Herzet, C. Elvira, A. Arslan, ...
Generalization, acceleration

IRIT / I3S
E. Soubies, L. Blanc-Féraud, ...
Strong relax. of ℓ_0-norm

BnB – Region separation



Axis 1 – Relaxation construction

Region $\nu \equiv (\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_\bullet)$ with $\begin{cases} x_i = 0 & \text{if } i \in \mathcal{S}_0 \\ x_i \neq 0 & \text{if } i \in \mathcal{S}_1 \\ x_i \in \mathbf{R} & \text{if } i \in \mathcal{S}_\bullet \end{cases}$

Restriction to region ν

$$p^\nu = \min_{\mathbf{x} \in \nu} f(\mathbf{A}\mathbf{x}) + g(\mathbf{x}) \quad \text{with} \quad g_i(x_i) = \lambda \|x_i\|_0 + h_i(x_i)$$

$\boxed{\text{reformulation}}$

Restriction to region ν

$$p^\nu = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + g^\nu(\mathbf{x}) \quad \text{with} \quad g_i^\nu(x_i) = \begin{cases} g_i(x_i) + \eta(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ g_i(x_i) + \eta(x_i \neq 0) & \text{if } i \in \mathcal{S}_1 \\ g_i(x_i) & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

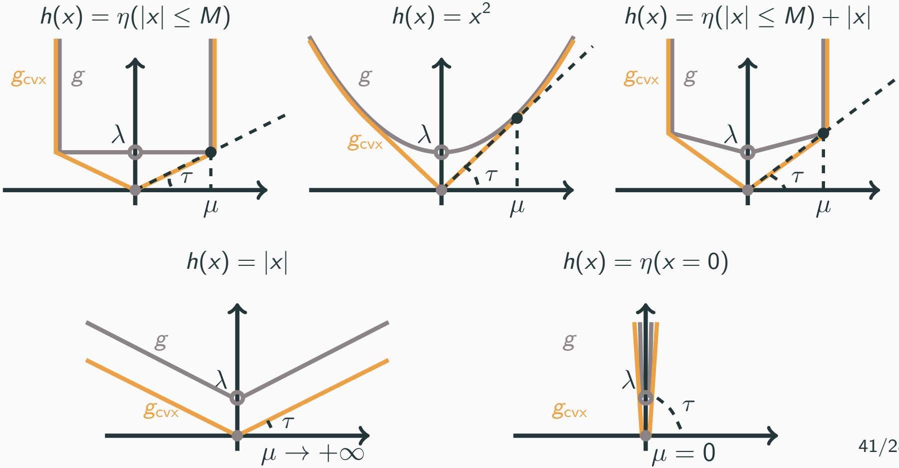
$\boxed{g_{\text{lb}}^\nu \leq g,\ g_{\text{lb}}^\nu \text{ convex}}$

Relaxation for region ν

$$p_{\text{lb}}^\nu = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + g_{\text{lb}}^\nu(\mathbf{x}) \quad \text{with} \quad g_{i,\text{lb}}^\nu(x_i) = \begin{cases} \eta(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ g_{i,\text{cvx}}(x_i) & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

Axis 1 – Graphical interpretation

$$g(x) = \lambda\|x\|_0 + h(x) \xrightarrow{\text{convexify}} g_{\text{cvx}}(x) = \begin{cases} \tau|x| & \text{if } |x| \leq \mu \\ \lambda + h(x) & \text{otherwise} \end{cases}$$



Axis 2 – Reduced and smoothed formulation

Convex problem

$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda\|\mathbf{x}\|_1 + h(\mathbf{x})$

$\downarrow$

Knowledge of zeros/positive/negative entries in the solutions

$\downarrow$

Reduced/smoothed formulation

$\min_{\tilde{\mathbf{x}}\in\mathbf{R}^{\tilde{n}}} f(\tilde{\mathbf{A}}\tilde{\mathbf{x}}) + \lambda\boldsymbol{\theta}^{\mathrm{T}}\tilde{\mathbf{x}} + h(\tilde{\mathbf{x}})$

Reduced dimension $\tilde{n} \ll n$

Smooth objective if f/h smooth

1st-order methods

Proximal gradient

Coordinate descent

$\downarrow$

Sub-linear/linear convergence rate

Cost $\mathcal{O}(nm)$ per iteration

2nd-order methods

Newton's method

LBFGS

$\downarrow$

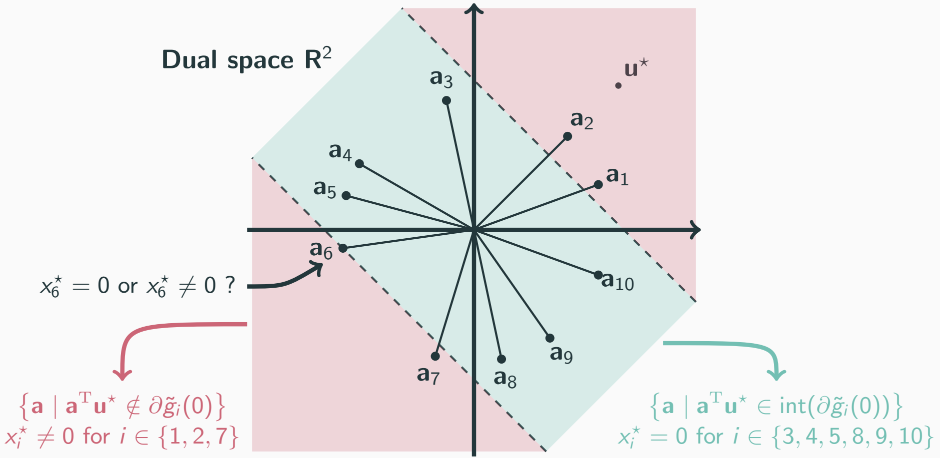
Super-linear convergence rate

Cost $\mathcal{O}(\tilde{n}m)$ per iteration

Axis 2 – Graphical interpretation (dual)

$$\text{Screening:} \quad \mathbf{a}_i^{\mathrm{T}}\mathbf{u}^\star \in \mathrm{int}(\partial\tilde{g}_i(0)) \implies x_i^\star = 0$$

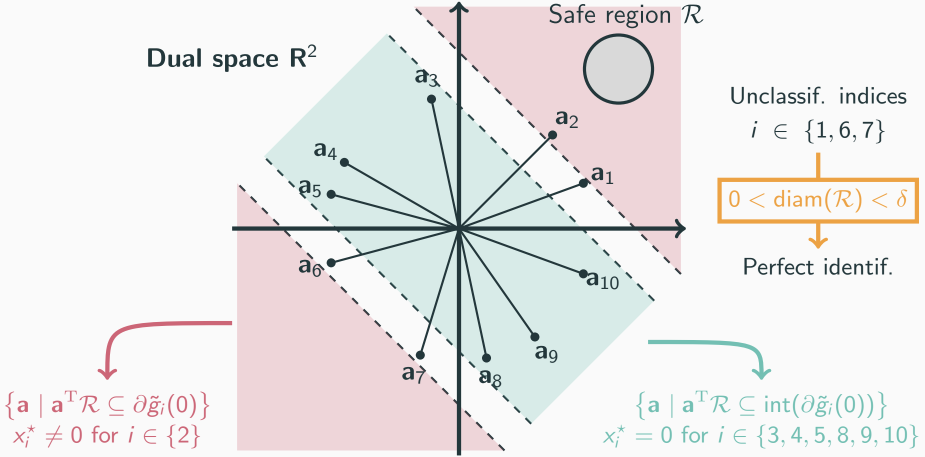
$$\text{Smoothing:} \quad \mathbf{a}_i^{\mathrm{T}}\mathbf{u}^\star \in \mathrm{cmpl}(\partial\tilde{g}_i(0)) \implies x_i^\star \neq 0$$



Axis 2 – Graphical interpretation (safe)

$$\text{Safe screening:} \quad \mathbf{a}_i^{\mathrm{T}}\mathcal{R} \subseteq \mathrm{int}(\partial \tilde{g}_i(0)) \implies x_i^\star = 0$$

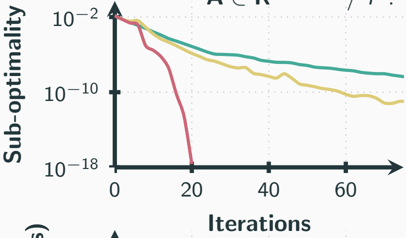
$$\text{Safe smoothing:} \quad \mathbf{a}_i^{\mathrm{T}}\mathcal{R} \subseteq \mathrm{cmpl}(\partial \tilde{g}_i(0)) \implies x_i^\star \neq 0$$



Convex problem

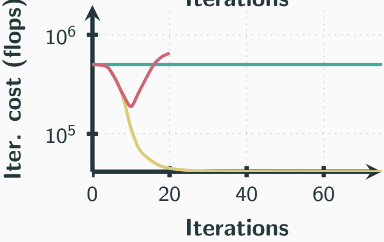
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \tilde{g}(\mathbf{x})$$

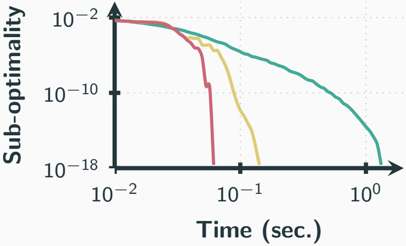
$\mathbf{A} \in \mathbf{R}^{100\times 300}$ / f : Quadratic / $\tilde{g}$: $\ell_1\ell_2$-norm



Accelerated proximal gradient

- Vanilla method
- With screening
- With screening and smoothing





ν Region $\nu \equiv (\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_\bullet)$ with $\begin{cases} x_i = 0 & \text{if } i \in \mathcal{S}_0 \\ x_i \neq 0 & \text{if } i \in \mathcal{S}_1 \\ x_i \in \mathbf{R} & \text{if } i \in \mathcal{S}_\bullet \end{cases}$

Restriction to region ν

$$p^\nu \;=\; \min_{\mathbf{x}\in\nu} f(\mathbf{A}\mathbf{x}) + g(\mathbf{x}) \qquad \text{with} \quad g_i(x_i) = \lambda\|x_i\|_0 + h_i(x_i)$$

$\boxed{g_{\text{lb}}^\nu \leq g,\ g_{\text{lb}}^\nu \text{ convex}}$

Relaxation for region ν

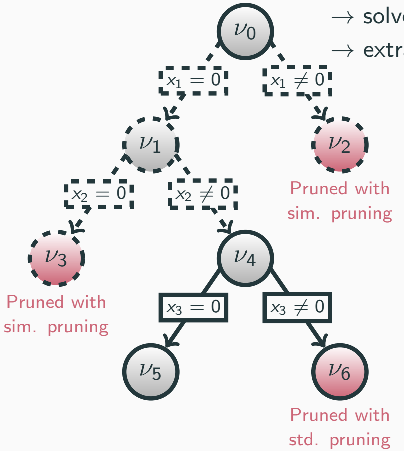
$$p_{\text{lb}}^\nu \;=\; \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + g_{\text{lb}}^\nu(\mathbf{x}) \qquad \text{with} \quad g_{i,\text{lb}}^\nu(x_i) = \begin{cases} \eta(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ g_{i,\text{cvx}}(x_i) & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

$\boxed{\text{dualize}}$

Dual relaxation for region ν

$$p_{\text{dual}}^\nu = \max_{\mathbf{u}\in\mathbf{R}^m} -f^*(-\mathbf{u}) - (g_{\text{lb}}^\nu)^*(\mathbf{A}^{\mathrm{T}}\mathbf{u}) \quad \text{with} \quad (g_{i,\text{lb}}^\nu)^*(\mathbf{a}_i^{\mathrm{T}}\mathbf{u}) = \begin{cases} 0 & \text{if } i \in \mathcal{S}_0 \\ h_i^*(\mathbf{a}_i^{\mathrm{T}}\mathbf{u}) - \lambda & \text{if } i \in \mathcal{S}_1 \\ [h_i^*(\mathbf{a}_i^{\mathrm{T}}\mathbf{u}) - \lambda]_+ & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

Axis 3 – Combining both paradigms



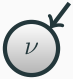
→ solve standard relaxation
→ extract dual lower bound

Solve relaxations in nodes ν_0, ν_4, ν_5 and ν_6

Bypass nodes ν_1, ν_2 and ν_3 using simultaneous pruning

Problem

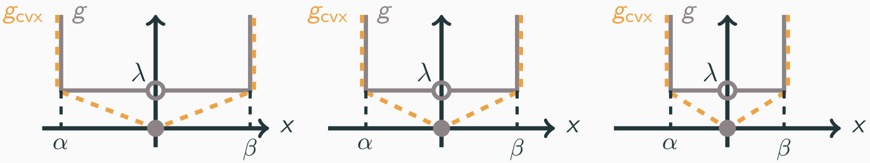
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + \eta(\boldsymbol{\alpha} \leq \mathbf{x} \leq \boldsymbol{\beta})$



Pruning test

Construct and solve a relaxation

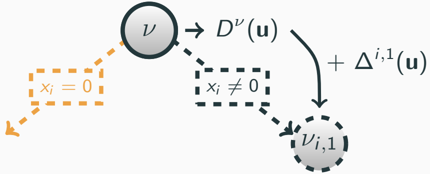
$g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + \eta(\boldsymbol{\alpha} \leq \mathbf{x} \leq \boldsymbol{\beta}) \to g_{\mathrm{cvx}}$



Practical side – Large interval $[\alpha, \beta]$ to obtain relevant solutions.
Numerical side – Small interval $[\alpha, \beta]$ to obtain strong relaxations.

Axis 4 – Peeling tests

Simultaneous pruning test



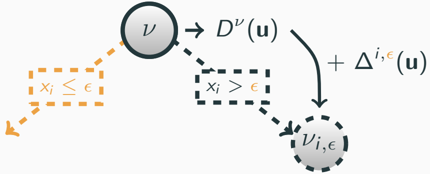
$$D^{\nu_{i,1}}(\mathbf{u}) = D(\mathbf{u}) + \Delta^{i,1}(\mathbf{u})$$

Result 1

$\rightarrow$ peeling test gives weaker conclusions but is easier to pass

$$D^{\nu_{i,\epsilon}}(\mathbf{u}) \ \geq \ D^{\nu_{i,1}}(\mathbf{u})$$

Peeling test



$$D^{\nu_{i,\epsilon}}(\mathbf{u}) = D(\mathbf{u}) + \Delta^{i,\epsilon}(\mathbf{u})$$

Result 2

$\rightarrow$ we can find the smallest $\epsilon > 0$ such that the peeling test is passed

$$p^{\star}_{\mathrm{ub}} \ < \ D^{\nu_{i,\epsilon}}(\mathbf{u})$$

Numerics – El0ps

Problem

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

El0ps

Problem management

Builtin instances of f/h

User-defined instances of f/h

- $\rightarrow$ function value
- $\rightarrow$ convex conjugate
- $\rightarrow$ subdifferential
- $\rightarrow$ proximal operator

Branch-and-Bound solver

Generic backbone (Axis 1)

$\rightarrow$ generic convex relaxation

Relaxation solution (Axis 2)

- $\rightarrow$ convex optim. solver
- $\rightarrow$ accel. with screen/smooth

Reduced complexity (Axis 3)

- $\rightarrow$ dual bounds
- $\rightarrow$ simultaneous pruning

Perspectives – Stronger relaxations

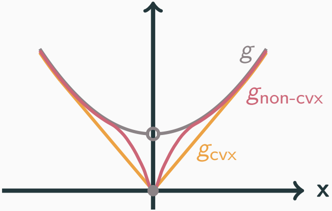
$$
\begin{aligned}
p^{\nu} &= \min_{\mathbf{x} \in \nu} \; f(\mathbf{A}\mathbf{x}) + g(\mathbf{x}) \\
&\vee\!| \\
p^{\nu}_{\text{non-cvx}} &= \min_{\mathbf{x} \in \nu} \; f(\mathbf{A}\mathbf{x}) + g_{\text{non-cvx}}(\mathbf{x}) \\
&\vee\!| \\
p^{\nu}_{\text{cvx}} &= \min_{\mathbf{x} \in \nu} \; f(\mathbf{A}\mathbf{x}) + g_{\text{cvx}}(\mathbf{x})
\end{aligned}
$$

→ Need valid and tractable relaxation

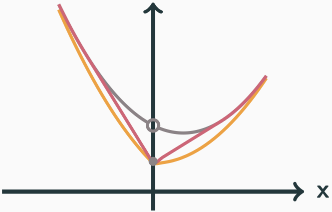
$$
\begin{aligned}
g(\mathbf{x}) &= \lambda \|\mathbf{x}\|_0 + \tfrac{\gamma}{2}\|\mathbf{x}\|_2^2 \\
&\vee\!| \\
g_{\text{non-cvx}}(\mathbf{x}) &= \mathrm{Mcp}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{x}) + \tfrac{\gamma}{2}\|\mathbf{x}\|_2^2 \\
&\vee\!| \\
g_{\text{cvx}}(\mathbf{x}) &= \mathrm{Berhu}_{\lambda,\gamma}(\mathbf{x})
\end{aligned}
$$

→ Tune $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ depending on f and $\mathbf{A}$ to ensure the overall objective convexity

Regularization functions



Objective functions



Perspectives – Proximal identification

Convex problem

$$\mathbf{x}^\star \in \operatorname{argmin}_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{A}\mathbf{x}) + \tilde{g}(\mathbf{x})$$

Screening/smoothing

Optimality conditions

$$\mathbf{A}^{\mathrm{T}}\mathbf{u}^\star \in \partial\tilde{g}(\mathbf{x}^\star)$$

$$\boxed{\tilde{g}(\mathbf{x}^\star) = \textstyle\sum_{i=1}^n \tilde{g}_i(x_i^\star)}$$

Identif. from dual solution

$$\mathbf{a}_i^{\mathrm{T}}\mathbf{u}^\star \to x_i^\star = 0 \text{ or } x_i^\star \neq 0$$

$$\boxed{\mathcal{R} \subseteq \mathbf{R}^m \text{ with } \mathbf{u}^\star \in \mathcal{R}}$$

Identif. from safe region

$$\mathbf{a}_i^{\mathrm{T}}\mathcal{R} \to x_i^\star = 0 \text{ or } x_i^\star \neq 0$$

Safe but can miss indices

$$\mathbf{w} = \operatorname{prox}_{\tilde{g}}(\mathbf{x}) \longleftrightarrow \mathbf{x} - \mathbf{w} \in \partial\tilde{g}(\mathbf{w})$$

unsafe exploit.

make safe

Proximal identification

Optimality conditions

$$\mathbf{x}^\star = \operatorname{prox}_{\tilde{g}}(\mathbf{x}^\star)$$

$$\boxed{\tilde{g}(\mathbf{x}^\star) = \textstyle\sum_{i=1}^n \tilde{g}_i(x_i^\star)}$$

Identif. from prox. operator

$$\operatorname{prox}_{\tilde{g}_i}(x_i^\star) \to x_i^\star = 0 \text{ or } x_i^\star \neq 0$$

$$\boxed{\mathbf{x} \in \mathbf{R}^n \text{ near } \mathbf{x}^\star}$$

Identif. from arbitrary point

$$\operatorname{prox}_{\tilde{g}_i}(x_i) \to x_i^\star = 0 \text{ or } x_i^\star \neq 0$$

Unsafe but classify all indices

Perspectives – Newton acceleration

Algorithm 1: Our approach

Input: $\mathbf{x}^0 \in \mathbf{R}^n$

Initialize $(\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_\bullet) = (\emptyset, \emptyset, [1, n])$
for $k = 1, 2, \dots, k_{\max}$ **do**
  `// Update iterate`
  $\mathbf{x}^k_{\mathcal{S}_0} \leftarrow \mathbf{0}$
  $\mathbf{x}^k_{\mathcal{S}_1} \leftarrow 2^{\text{nd}}\text{OrderIteration}(\mathbf{x}^{k-1}_{\mathcal{S}_1})$
  $\mathbf{x}^k_{\mathcal{S}_\bullet} \leftarrow 1^{\text{st}}\text{OrderIteration}(\mathbf{x}^{k-1}_{\mathcal{S}_\bullet})$
  `// Update structure knowledge`
  $\mathcal{R} \leftarrow \text{SafeRegion}(\mathbf{x}^k)$
  $\mathcal{S}_0 \leftarrow \mathcal{S}_0 \cup \text{ScreeningTest}(\mathcal{R})$
  $\mathcal{S}_1 \leftarrow \mathcal{S}_1 \cup \text{SmoothingTest}(\mathcal{R})$
  $\mathcal{S}_\bullet \leftarrow [1, n] \setminus (\mathcal{S}_0 \cup \mathcal{S}_1)$
end

- $\mathcal{S}_0 \subseteq \mathcal{S}_0^\star$ and $\mathcal{S}_1 \subseteq \mathcal{S}_1^\star$ at any iter.
- $\mathcal{S}_\bullet \neq \emptyset$ until $k \geq k_0$

$\rightarrow$ Safe but uncomplete identification

$$\boxed{\begin{aligned} \mathcal{S}_0^\star &= \{i \mid x_i^\star = 0\} \\ \mathcal{S}_1^\star &= \{i \mid x_i^\star \neq 0\} \end{aligned}}$$

Algorithm 2: G. Bareilles *et al.*

Input: $\mathbf{x}^0 \in \mathbf{R}^n$

for $k = 1, 2, \dots, k_{\max}$ **do**
  `// Update iterate`
  $\tilde{\mathbf{x}}^k \leftarrow \text{ProxIteration}(\mathbf{x}^{k-1})$
  `// Get local structure`
  $(\mathcal{S}_0, \mathcal{S}_1) \leftarrow \text{ProxIdentification}(\tilde{\mathbf{x}}^k)$
  `// Follow local structure`
  $\mathbf{x}^k \leftarrow \text{StructureUpdate}_{(\mathcal{S}_0, \mathcal{S}_1)}(\tilde{\mathbf{x}}^k)$
end

- $\mathcal{S}_\bullet = \emptyset$ at any iter.
- $\mathcal{S}_0 \neq \mathcal{S}_0^\star$ and $\mathcal{S}_1 \neq \mathcal{S}_1^\star$ until $k \geq k_0$

$\rightarrow$ Unsafe but complete identification