Optimization methods for ℓ_0-problems

Théo Guyard

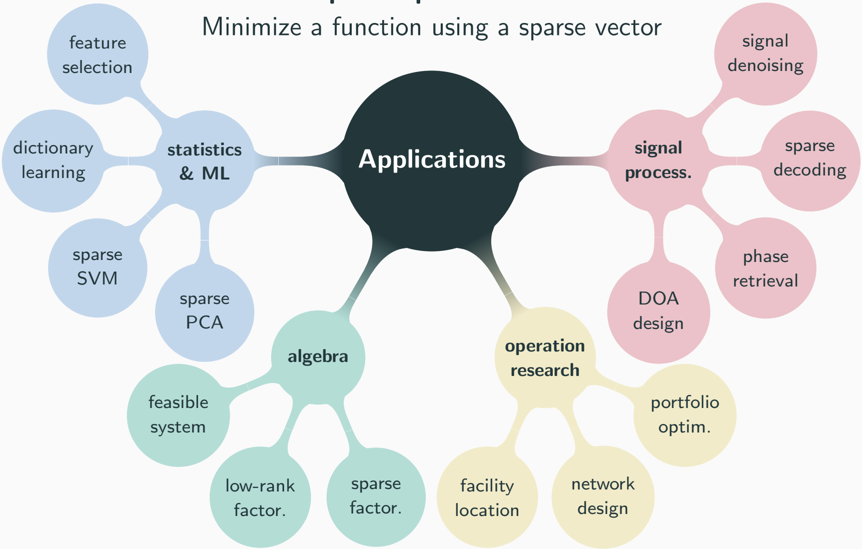
JOPT, HEC Montréal, Canada – May 12th, 2025

Sparse optimization

Minimize a function using a sparse vector

Sparse optimization

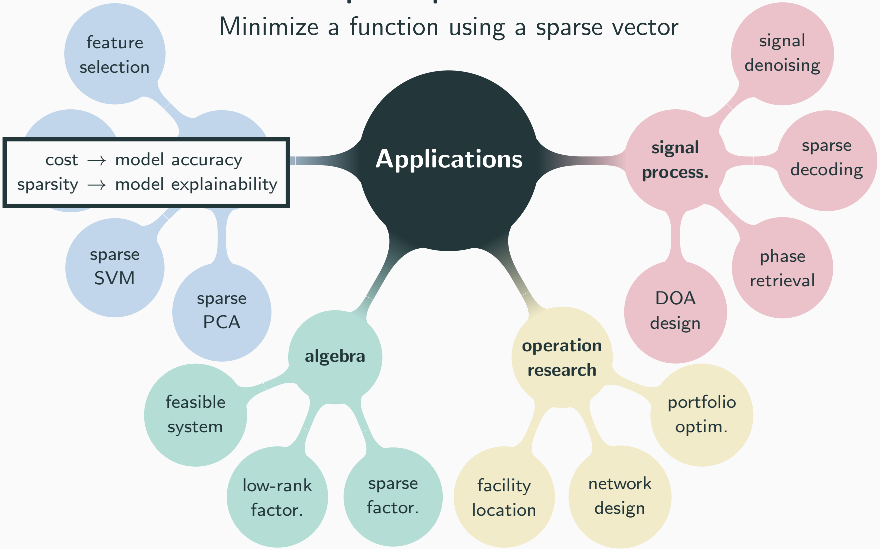
Minimize a function using a sparse vector



A. Tillmann *et al.* (2024)

Sparse optimization

Minimize a function using a sparse vector

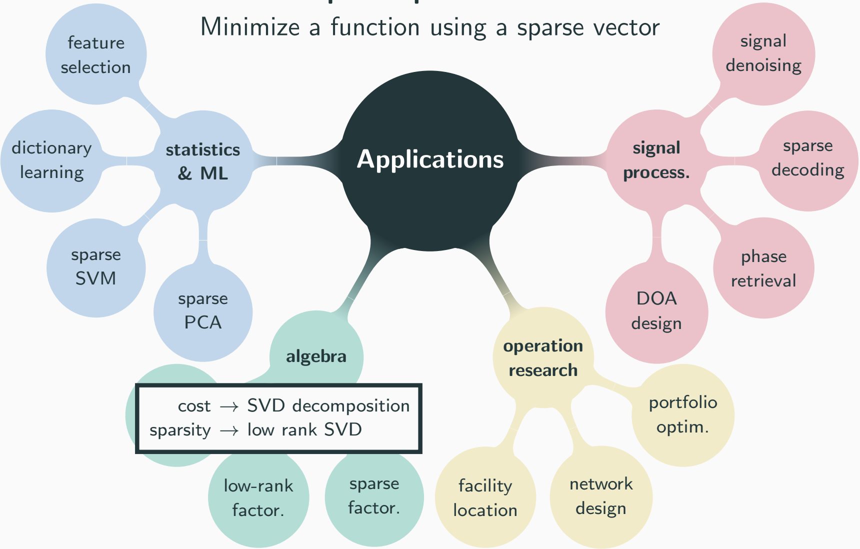


cost $\longrightarrow$ model accuracy
sparsity $\longrightarrow$ model explainability

A. Tillmann *et al.* (2024)

Sparse optimization

Minimize a function using a sparse vector



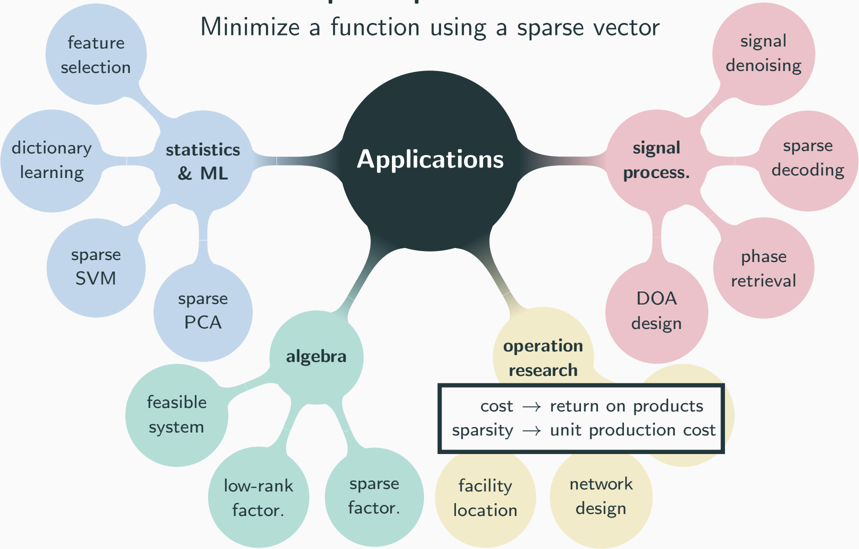
cost → SVD decomposition
sparsity → low rank SVD

A. Tillmann *et al.* (2024)

Sparse optimization

Sparse optimization

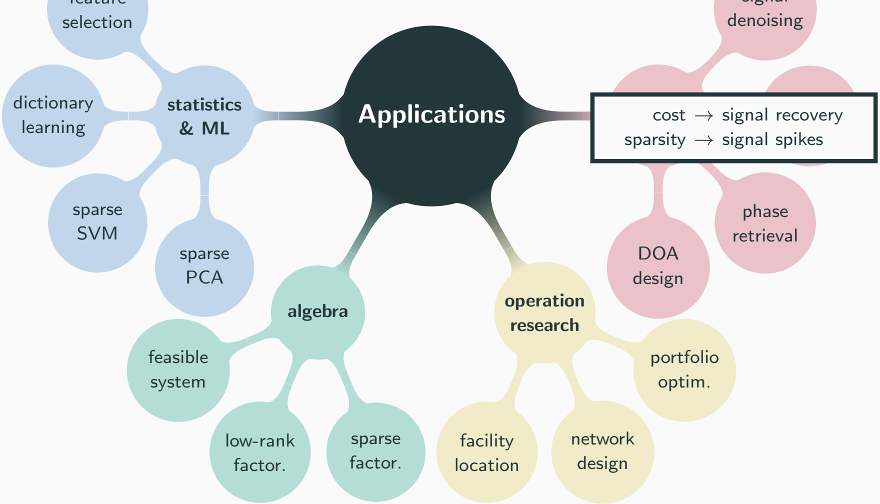
Minimize a function using a sparse vector



A. Tillmann *et al.* (2024)

Sparse optimization

Minimize a function using a sparse vector



A. Tillmann *et al.* (2024)

Sparse optimization

Minimize a function using a sparse vector

ℓ_0-regularized problems

Sparse optimization

Minimize a function using a sparse vector

cost

$f(\mathbf{x})$

loss function

ℓ_0-regularized problems

Sparse optimization

Minimize a function using a sparse vector

cost

$f(\mathbf{x})$

loss function

sparsity

$\|\mathbf{x}\|_0$

ℓ_0-norm

counts number of non-zeros in $\mathbf{x}$

ℓ_0-regularized problems

Sparse optimization

Minimize a function using a sparse vector

cost

sparsity

counts number of non-zeros in $\mathbf{x}$

$f(\mathbf{x})$
loss function

$\|\mathbf{x}\|_0$
ℓ_0-norm

ℓ_0-regularized problem

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

ℓ_0-regularized problems

Sparse optimization

Minimize a function using a sparse vector

$\boxed{\text{cost}}$

$\boxed{\text{sparsity}}$

counts number of non-zeros in $\mathbf{x}$

$f(\mathbf{x})$
loss function

$\longleftrightarrow$

$\|\mathbf{x}\|_0$
ℓ_0-norm

ℓ_0-regularized problem

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

ℓ_0-regularized problems

Sparse optimization

Minimize a function using a sparse vector

cost

sparsity

counts number of non-zeros in $\mathbf{x}$

$f(\mathbf{x})$
loss function

$\|\mathbf{x}\|_0$
ℓ_0-norm

ℓ_0-regularized problem

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

NP-hard

Sparse optimization

Minimize a function using a sparse vector

$\boxed{\text{cost}}$

$\boxed{\text{sparsity}}$

counts number of non-zeros in $\mathbf{x}$

$f(\mathbf{x})$

loss function

$\|\mathbf{x}\|_0$

ℓ_0-norm

ℓ_0-regularized problem

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

NP-hard

Off-the-shelf MIP solvers

✗ Poor numerical performances

✗ Need standardized expressions

ℓ_0-regularized problems

Sparse optimization

Minimize a function using a sparse vector

cost

sparsity

counts number of non-zeros in $\mathbf{x}$

$f(\mathbf{x})$

loss function

$\|\mathbf{x}\|_0$

ℓ_0-norm

ℓ_0-regularized problem

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

NP-hard

Off-the-shelf MIP solvers

✗ Poor numerical performances

✗ Need standardized expressions

Specialized Branch-and-Bound

✓ Better numerical performances

✓ Greater flexibility

ℓ_0-regularized problems

Sparse optimization

Minimize a function using a sparse vector

cost

sparsity

counts number of non-zeros in $\mathbf{x}$

$f(\mathbf{x})$
loss function

$\|\mathbf{x}\|_0$
ℓ_0-norm

ℓ_0-regularized problem

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

NP-hard

Off-the-shelf MIP solvers

✗ Poor numerical performances
✗ Need standardized expressions

Topic of this talk

Specialized Branch-and-Bound

✓ Better numerical performances
✓ Greater flexibility

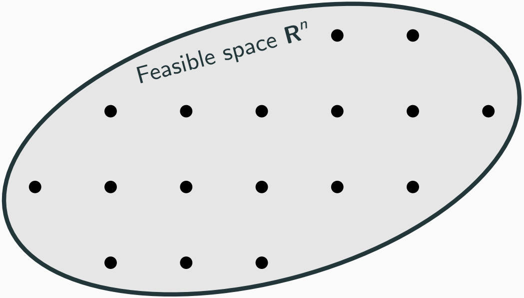
Branch-and-Bound Algorithms

BnB – Algorithmic principle

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

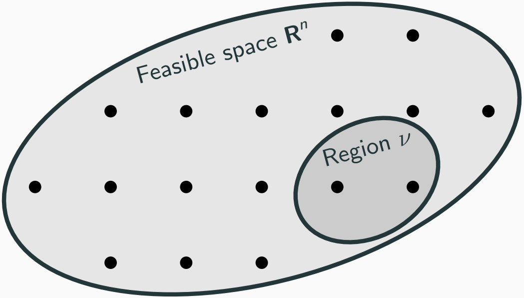
BnB – Algorithmic principle

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



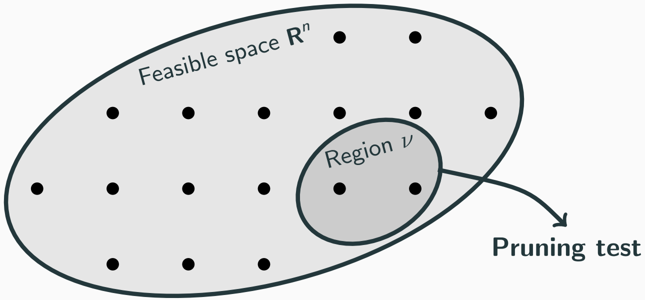
BnB – Algorithmic principle

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



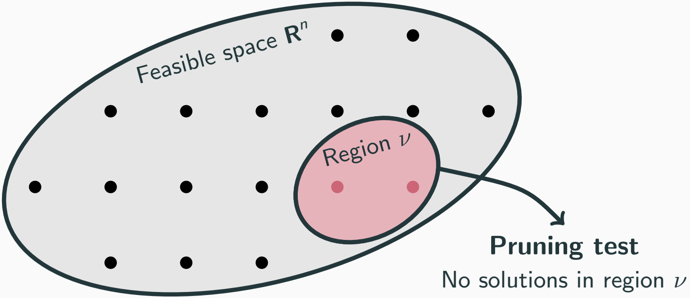
BnB – Algorithmic principle

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



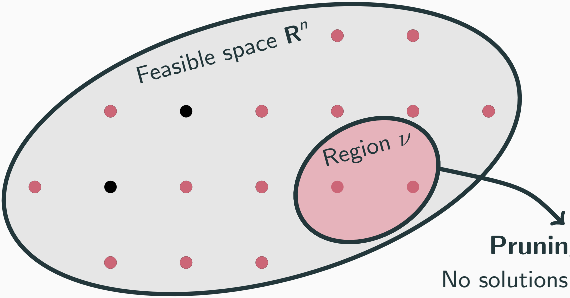
BnB – Algorithmic principle

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.

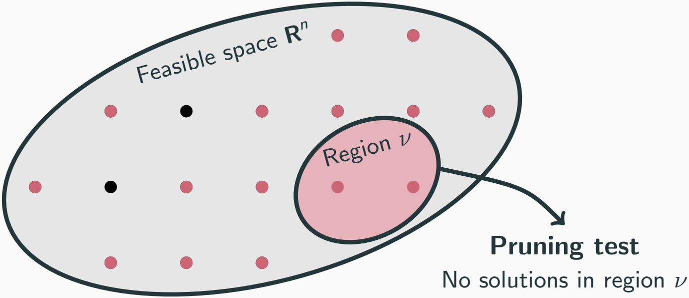


Pruning test

No solutions in region ν

BnB – Algorithmic principle

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



Branching step – Region design and exploration
Bounding step – Pruning test evaluation

BnB – Branching step

Problem

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

BnB – Branching step

Problem

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

Observation

Solutions are expected to be sparse

BnB – Branching step

Problem

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

Observation

Solutions are expected to be sparse

Method

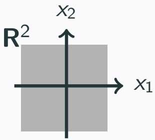
Drive the sparsity of the optimization variable

Problem

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

Observation

Solutions are expected to be sparse



Method

Drive the sparsity of the optimization variable

BnB – Branching step

Problem

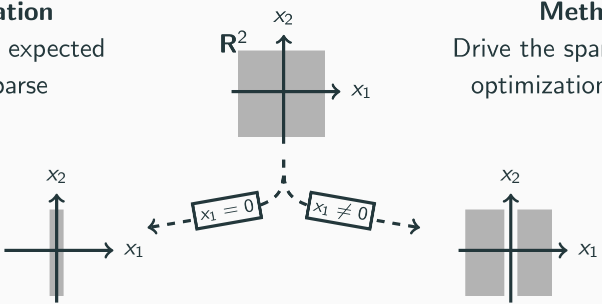
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



BnB – Branching step

Problem

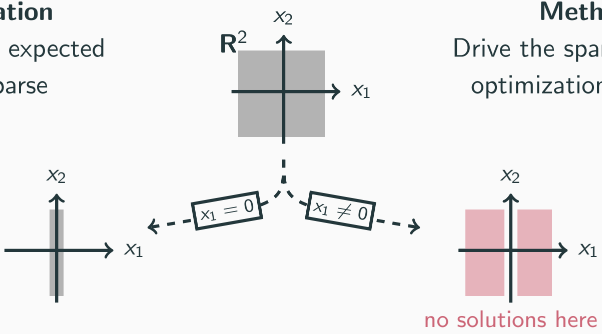
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



BnB – Branching step

Problem

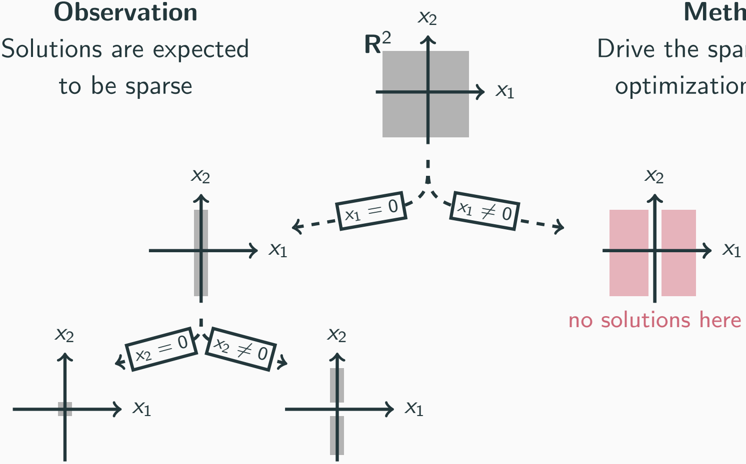
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable

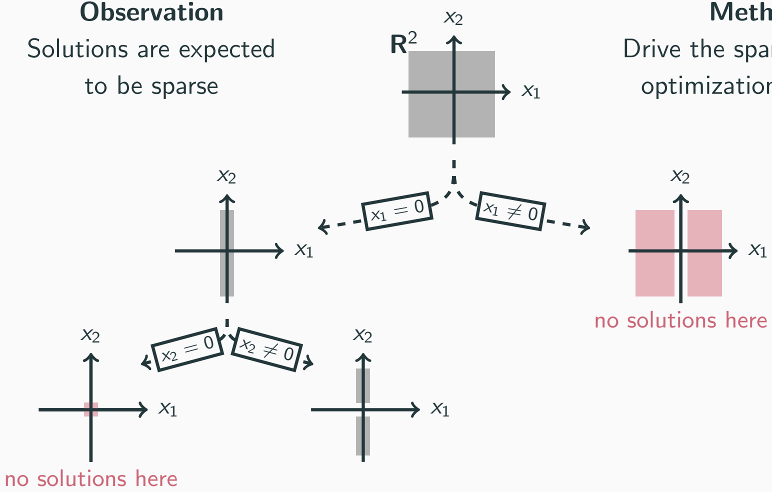


Problem

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

Observation

Solutions are expected to be sparse



Method

Drive the sparsity of the optimization variable

BnB – Branching step

Problem

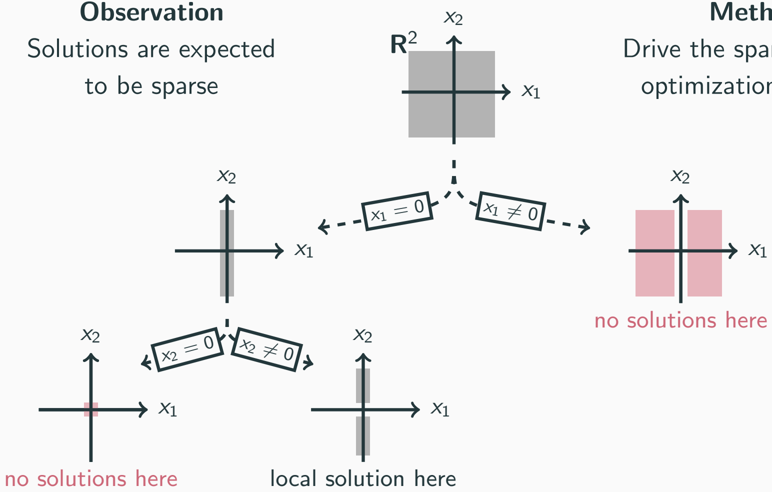
$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



Problem

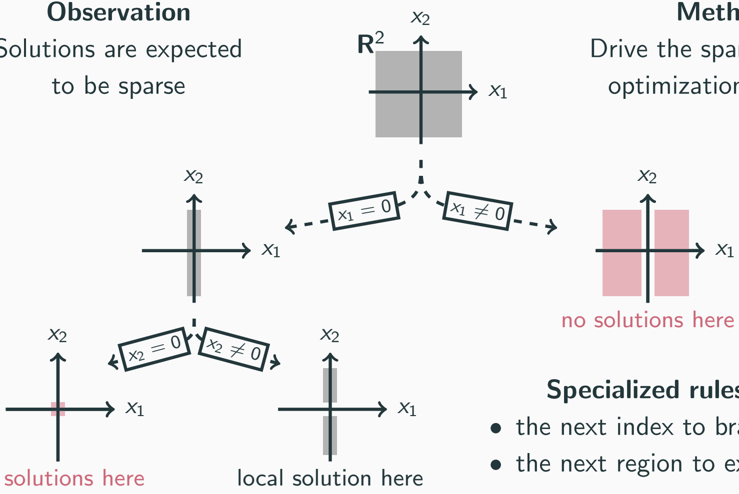
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

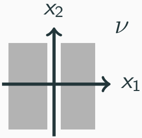
Drive the sparsity of the optimization variable



Specialized rules for

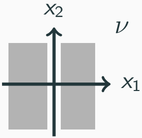
- the next index to branch on
- the next region to explore

BnB – Bounding step



Does region ν contains optimal solutions ?

BnB – Bounding step

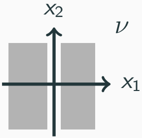


Does region ν contains optimal solutions ?

Problem

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

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

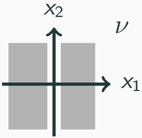
$p^\star = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$ restrict to ν

Restriction to region ν

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

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

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

$\downarrow$ restrict to ν

Restriction to region ν

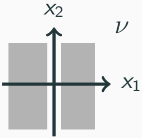
$$p^\nu = \min_{\mathbf{x}\in\nu} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

$\downarrow$ compare

Pruning test

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

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

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

restrict to ν

Restriction to region ν

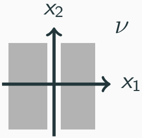
$p^\nu = \min_{\mathbf{x} \in \nu} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

compare

Pruning test

$p^\nu > p^\star$ $\longrightarrow$ prune ν

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

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

$\downarrow$ restrict to ν

Restriction to region ν

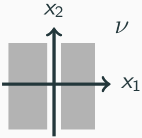
$p^\nu = \min_{\mathbf{x}\in\nu} \; f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$ compare

Pruning test

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

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

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

$\downarrow$ restrict to ν

Restriction to region ν

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

$\downarrow$ compare

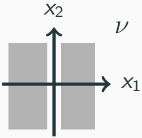
Pruning test

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

Easy task

Compute an upper bound on $p^\star$

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

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

restrict to ν

Restriction to region ν

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

compare

Pruning test

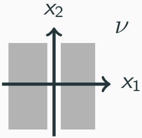
$p^\nu_{\mathrm{lb}} > p^\star_{\mathrm{ub}}$ $\longrightarrow$ prune ν

Easy task

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

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

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

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

restrict to ν

Restriction to region ν

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

compare

Pruning test

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

Easy task

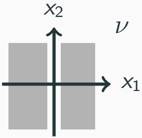
Compute an upper bound on $p^\star$

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

Main challenge

Compute a lower bound on p^ν

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

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

restrict to ν

Restriction to region ν

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

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

Construct and
solve a relaxation

BnB – Building relaxations

Restriction to region ν

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

seek tight/tractable lower bound on p^ν

BnB – Building relaxations

Restriction to region ν

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

seek tight/tractable lower bound on p^ν

reformulation

Restriction to region ν

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

with g^ν proper and closed

BnB – Building relaxations

Restriction to region ν

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

seek tight/tractable lower bound on p^{ν}

reformulation

Restriction to region ν

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

with g^{ν} proper and closed

$g^{\nu}_{\mathrm{lb}} \leq g^{\nu}$, g^{ν}_{lb} convex

Relaxation for region ν

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

set g^{ν}_{lb} set as the convex envelope of g^{ν}

BnB – Geometrical intuition

Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ with $\mathbf{x} \in \mathbf{R}^n$

Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ **with** $\mathbf{x} \in \mathbf{R}^n$

$\uparrow$

h separable

$\downarrow$

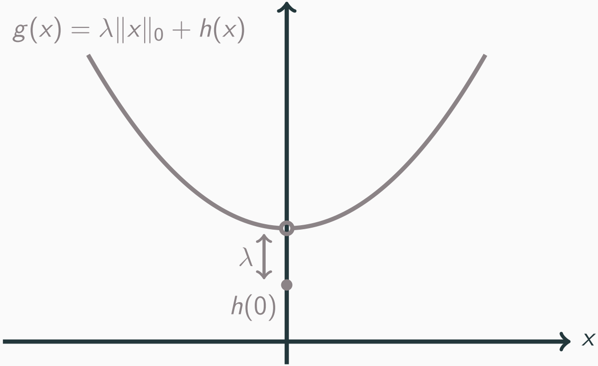
Convex envelope of $g(x) = \lambda\|x\|_0 + h(x)$ **with** $x \in \mathbf{R}$

BnB – Geometrical intuition

Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ **with** $\mathbf{x} \in \mathbf{R}^n$

h separable

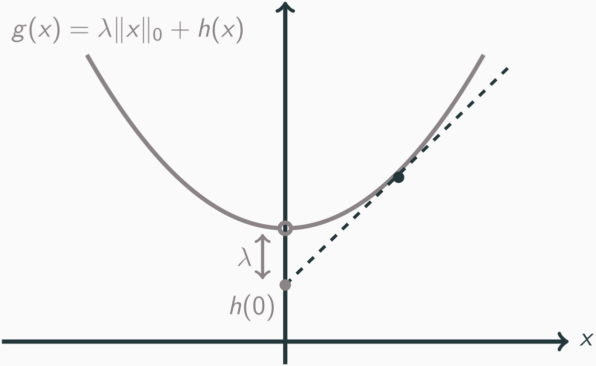
Convex envelope of $g(x) = \lambda\|x\|_0 + h(x)$ **with** $x \in \mathbf{R}$



Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ with $\mathbf{x} \in \mathbf{R}^n$

$\boxed{h \text{ separable}}$

Convex envelope of $g(x) = \lambda\|x\|_0 + h(x)$ with $x \in \mathbf{R}$



BnB – Geometrical intuition

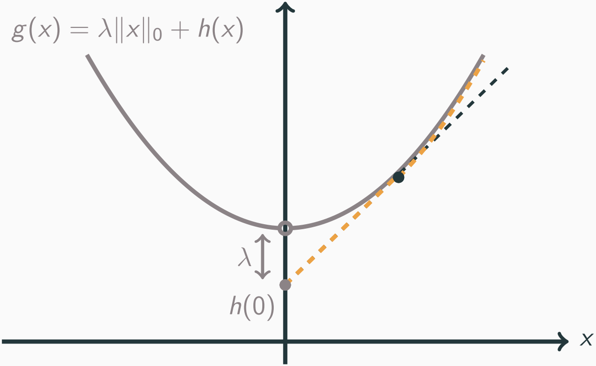
Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ **with** $\mathbf{x} \in \mathbf{R}^n$

$\uparrow$

$\boxed{h \text{ separable}}$

$\downarrow$

Convex envelope of $g(x) = \lambda\|x\|_0 + h(x)$ **with** $x \in \mathbf{R}$

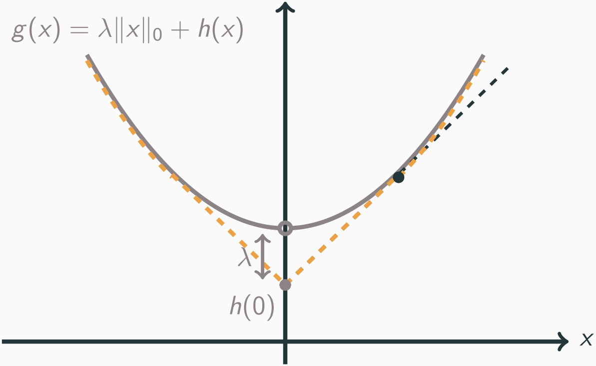


BnB – Geometrical intuition

Convex envelope of $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$ **with** $\mathbf{x} \in \mathbf{R}^n$

h separable

Convex envelope of $g(x) = \lambda\|x\|_0 + h(x)$ **with** $x \in \mathbf{R}$



BnB – The secrete sauce

Overall solve time

pruning test time $\times$ number of regions processed

BnB – The secrete sauce

Overall solve time

$\underbrace{\text{pruning test time}} \times$ number of regions processed

Relaxation for region ν

$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$

Overall solve time

$\underbrace{\text{pruning test time}} \times$ number of regions processed

Relaxation for region ν

$p_{\text{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\text{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

BnB – The secrete sauce

Overall solve time

$\underbrace{\text{pruning test time}}$ $\times$ number of regions processed

Relaxation for region ν

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

g^{ν}_{lb} is proper, closed, convex,
separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

BnB – The secrete sauce

Overall solve time

$\underbrace{\text{pruning test time}} \times$ number of regions processed

Relaxation for region ν

$p_{\text{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\text{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex,
separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

$\rightarrow$ acceleration strategies

working set, screening tests, ...

BnB – The secrete sauce

Overall solve time

$\underbrace{\text{pruning test time}} \times \underbrace{\text{number of regions processed}}$

Relaxation for region ν

$p_{\text{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\text{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning

BnB – The secrete sauce

Overall solve time

$$\underbrace{\text{pruning test time}} \times \underbrace{\text{number of regions processed}}$$

Relaxation for region ν

$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

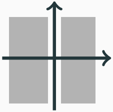
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...

Overall solve time

pruning test time $\times$ number of regions processed

Relaxation for region ν

$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

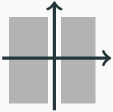
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

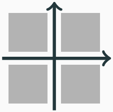
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...



perform degraded but low-cost pruning test

BnB – The secrete sauce

Overall solve time

pruning test time $\times$ number of regions processed

Relaxation for region ν

$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

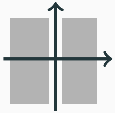
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

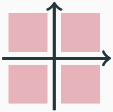
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...



perform degraded but low-cost pruning test

BnB – The secrete sauce

Overall solve time

pruning test time × number of regions processed

Relaxation for region ν

$$p_{\mathsf{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathsf{lb}}^{\nu}(\mathbf{x})$$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

$\downarrow$

This is a convex sparse optimization problem

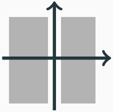
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

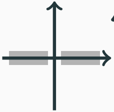
$\rightarrow$ acceleration strategies

working set, screening tests, ...

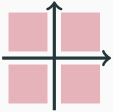
Simultaneous pruning



processing region ...



continue processing



perform degraded but low-cost pruning test

Overall solve time

$\underbrace{\text{pruning test time}} \times \underbrace{\text{number of regions processed}}$

Relaxation for region ν

$p_{\mathrm{lb}}^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g_{\mathrm{lb}}^{\nu}(\mathbf{x})$

g_{lb}^{ν} is proper, closed, convex, separable, and non-smooth at $\mathbf{x} = \mathbf{0}$

This is a convex sparse optimization problem

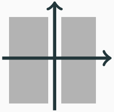
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

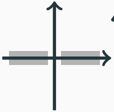
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning

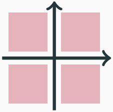


processing region ...



continue processing

bypass region



perform degraded but low-cost pruning test

Numerical Illustration

Numerics – Feature Selection Problem

Problem

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

MIP solvers: `cplex` `mosek` **BnB solver:** `el0ps`

Numerics – Feature Selection Problem

Problem

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

MIP solvers: `cplex` `mosek` **BnB solver:** `el0ps`

Instance 1

- f : LeastSquares$(\mathbf{y}, \mathbf{A}\mathbf{x})$
- h : ℓ_2-norm + bound constraint on $\mathbf{x}$
- `riboflavin`: $\mathbf{y} \in \mathbf{R}^{71}$ and $\mathbf{A} \in \mathbf{R}^{71\times 4088}$

Numerics – Feature Selection Problem

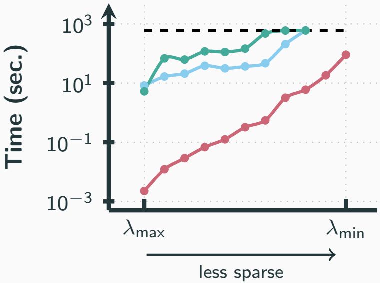
Problem

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

MIP solvers: cplex mosek **BnB solver:** el0ps

Instance 1

- f : LeastSquares$(\mathbf{y}, \mathbf{Ax})$
- h : ℓ_2-norm + bound constraint on $\mathbf{x}$
- `riboflavin`: $\mathbf{y} \in \mathbf{R}^{71}$ and $\mathbf{A} \in \mathbf{R}^{71\times 4088}$



Numerics – Feature Selection Problem

Problem

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

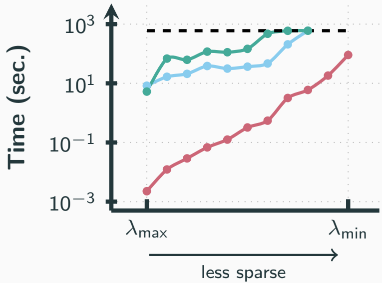
MIP solvers: `cplex` `mosek` **BnB solver:** `el0ps`

Instance 1

- f : LeastSquares$(\mathbf{y}, \mathbf{Ax})$
- h : ℓ_2-norm + bound constraint on $\mathbf{x}$
- `riboflavin`: $\mathbf{y} \in \mathbf{R}^{71}$ and $\mathbf{A} \in \mathbf{R}^{71\times 4088}$

Instance 2

- f : Logistic$(\mathbf{y}, \mathbf{Ax})$
- h : ℓ_1-norm + bound constraint on $\mathbf{x}$
- `leukemia`: $\mathbf{y} \in \mathbf{B}^{38}$ and $\mathbf{A} \in \mathbf{R}^{38\times 7129}$



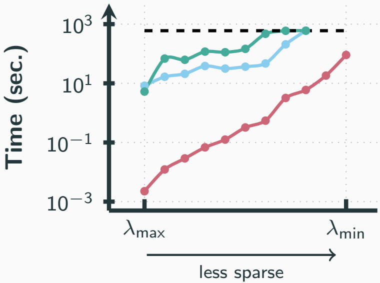
Problem

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

MIP solvers: `cplex` `mosek` **BnB solver:** `el0ps`

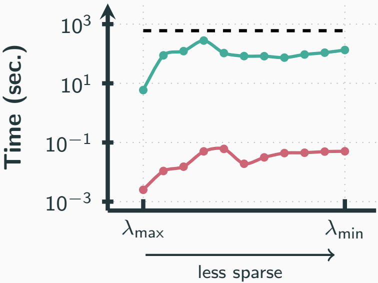
Instance 1

- f : LeastSquares$(\mathbf{y}, \mathbf{Ax})$
- h : ℓ_2-norm + bound constraint on $\mathbf{x}$
- `riboflavin`: $\mathbf{y} \in \mathbf{R}^{71}$ and $\mathbf{A} \in \mathbf{R}^{71\times 4088}$



Instance 2

- f : Logistic$(\mathbf{y}, \mathbf{Ax})$
- h : ℓ_1-norm + bound constraint on $\mathbf{x}$
- `leukemia`: $\mathbf{y} \in \mathbf{B}^{38}$ and $\mathbf{A} \in \mathbf{R}^{38\times 7129}$



Take-home messages

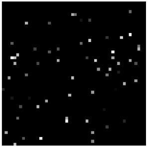
- ℓ_0-problems are of great practical interest but NP-hard
- Off-the-shelf MIP solvers
  - ✗ Poor numerical performances
  - ✗ Need standardized expressions
- Specialized BnB solvers
  - ✓ Better numerical performances
  - ✓ Greater flexibility
- Structure-exploitation is key
  - $\rightarrow$ Sparsity-driven branching strategy
  - $\rightarrow$ Convexification-based bounding strategy
  - $\rightarrow$ Efficient convex solver for lower bounds
  - $\rightarrow$ Dedicated simultaneous pruning strategy

Question time !

🤓

Compressed sensing

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

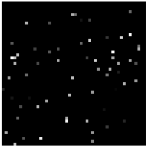


n pixels

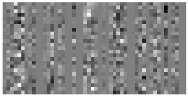
Compressed sensing

Sparse signal $\xrightarrow{\boxed{\text{Linear compression}}}$ Compressed signal

$\mathbf{x} \in \mathbf{R}^n$ $\qquad$ $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbf{R}^m$



n pixels



m pixels, $m \ll n$

Compressed sensing

Sparse signal

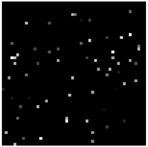
$\mathbf{x} \in \mathbf{R}^n$

Linear compression →

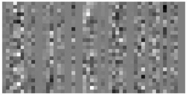
Compressed signal

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

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



n pixels



m pixels, $m \ll n$

Compressed sensing

Sparse signal

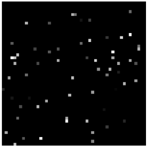
$\mathbf{x} \in \mathbf{R}^n$

Linear compression

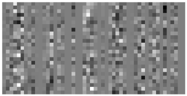
Recover $\mathbf{x}$ from $\mathbf{A}$ and $\mathbf{y}$

Compressed signal

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



n pixels



m pixels, $m \ll n$

Goal

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

Compressed sensing

Linear compression

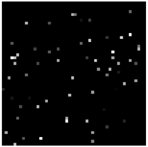
Sparse signal

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

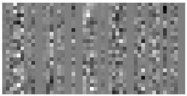
Compressed signal

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

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



n pixels



m pixels, $m \ll n$

Goal

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

no unique solution

Goal

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

Feature selection

	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

Feature selection

	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 $\mathbf{y} = \phi(\mathbf{A}\mathbf{x})$

Feature selection

	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 explainability

Use few features

Feature selection

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

$\downarrow$

Model explainability

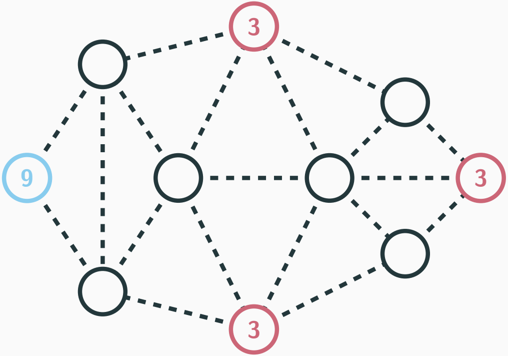
Use few features

$\downarrow$

Goal

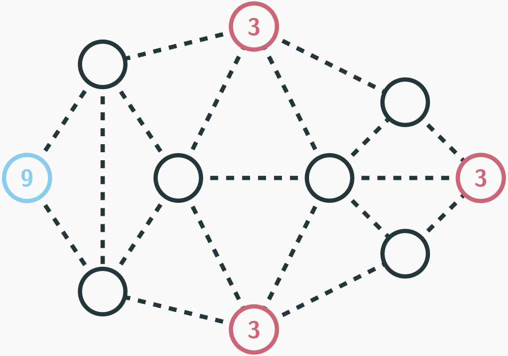
Find $\mathbf{x}$ sparse such that $\mathcal{L}_\phi(\mathbf{A}\mathbf{x}, \mathbf{y})$ is small

Network design



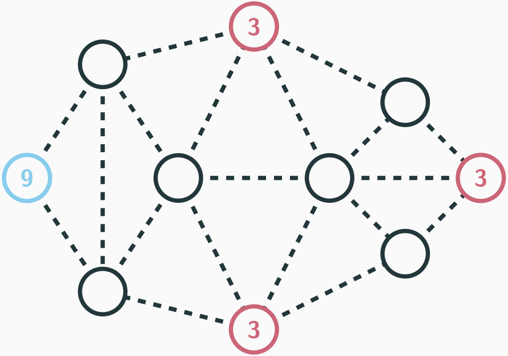
Which edges to build to transport
products from source to sink nodes ?

Network design



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

Network design

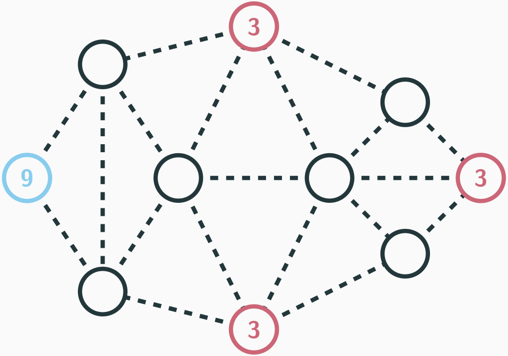


Which edges to build to transport products from source to sink nodes ?

construct edge $i \in I$ if $x_i > 0$

pay construction cost c

Network design



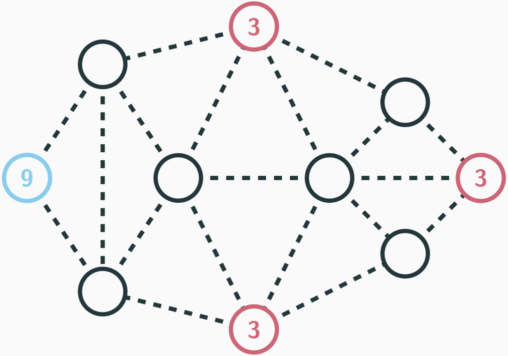
Which edges to build to transport products from source to sink nodes ?

construct edge $i \in I$ if $x_i > 0$

pay construction cost c

Question

How to construct the least number of edges to satisfy transportation needs ?



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

construct edge $i \in I$ if $x_i > 0$

pay construction cost c

Question

How to construct the least number of edges to satisfy transportation needs ?

$\updownarrow$

Find $\mathbf{x} \in \mathbf{R}^{\mathrm{card}(I)}$ sparse such that $\mathcal{Q}(\mathbf{x}) = 0$

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**

$\underbrace{\hspace{10em}}_{\text{4,088 genes}}$

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