Optimization methods for ℓ_0-problems

Théo Guyard

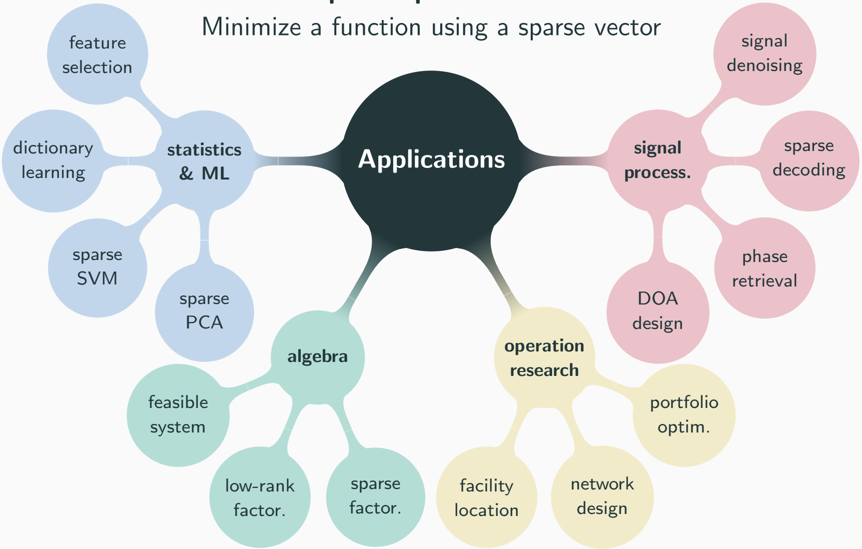
JOPT, HEC Montréal, Canada – May 12th, 2025

Sparse optimization

Minimize a function using a sparse vector

Sparse optimization

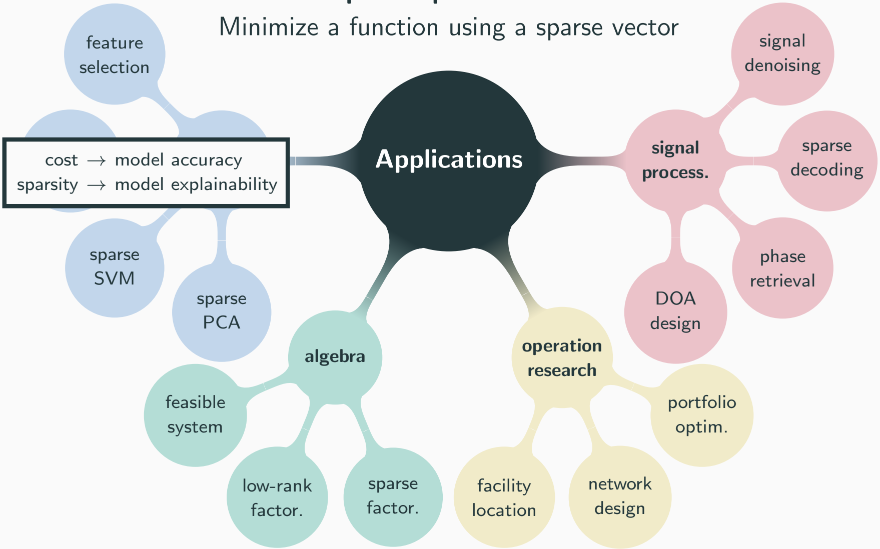
Minimize a function using a sparse vector



A. Tillmann *et al.* (2024)

Sparse optimization

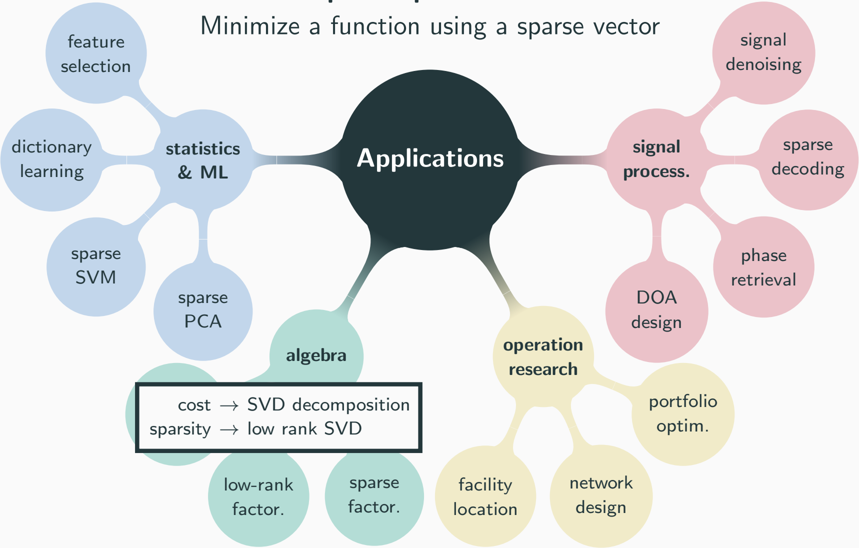
Minimize a function using a sparse vector



A. Tillmann *et al.* (2024)

Sparse optimization

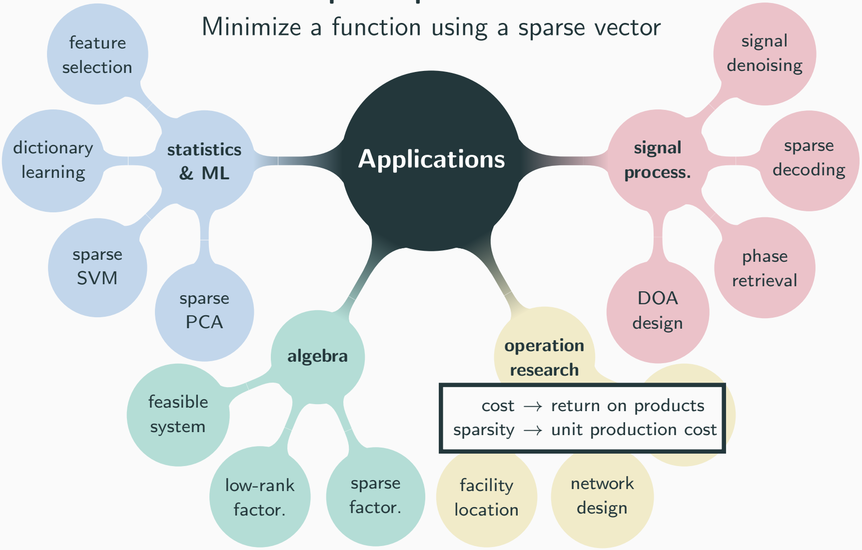
Minimize a function using a sparse vector



A. Tillmann *et al.* (2024)

Sparse optimization

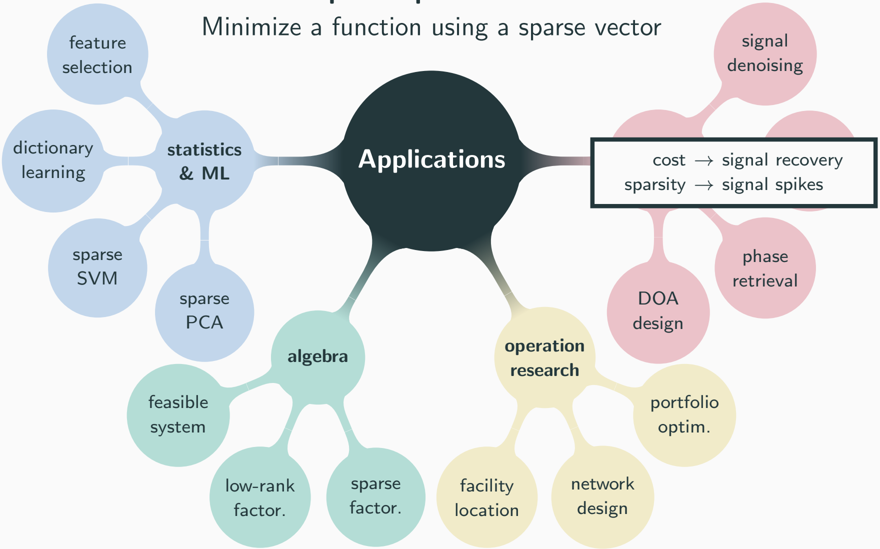
Minimize a function using a sparse vector



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

$\boxed{\text{cost}}$

$f(\mathbf{x})$

loss function

ℓ_0-regularized problems

Sparse optimization

Minimize a function using a sparse vector

$\boxed{\text{cost}}$

$f(\mathbf{x})$

loss function

$\boxed{\text{sparsity}}$

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

ℓ_0-norm

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

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

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

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

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

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

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

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

$\rightarrow$ NP-hard

MIP-based methods

Rely on off-the-shelf solvers

✗ Poor numerical performances

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

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

$\rightarrow$ NP-hard

MIP-based methods

Rely on off-the-shelf solvers

✗ Poor numerical performances

BnB-based methods

Tailored solution method

✓ Better numerical performances

Sparse optimization

Minimize a function using a sparse vector

cost

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

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

MIP-based methods

Rely on off-the-shelf solvers

✗ Poor numerical performances

Topic of this talk

BnB-based methods

Tailored solution method

✓ Better numerical performances

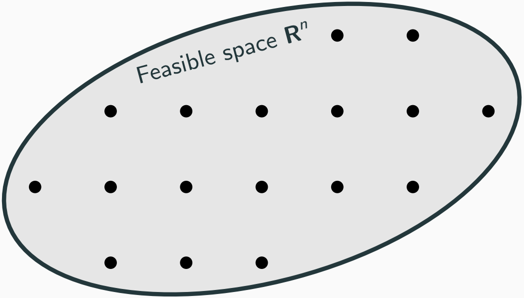
Branch-and-Bound Algorithms

BnB – Algorithmic principle

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

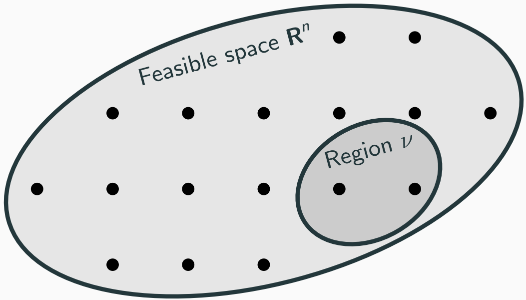
BnB – Algorithmic principle

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



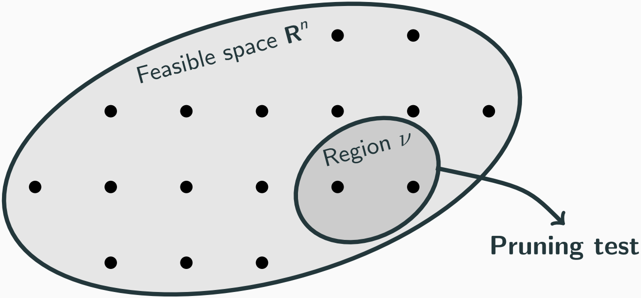
BnB – Algorithmic principle

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



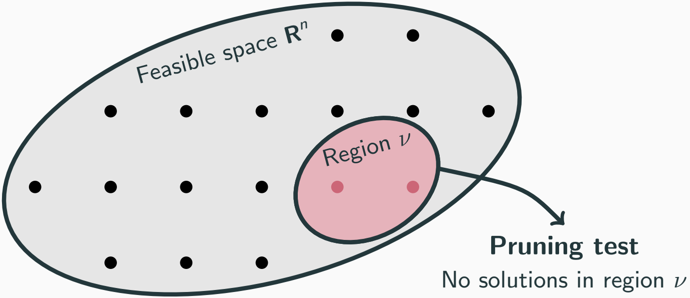
BnB – Algorithmic principle

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



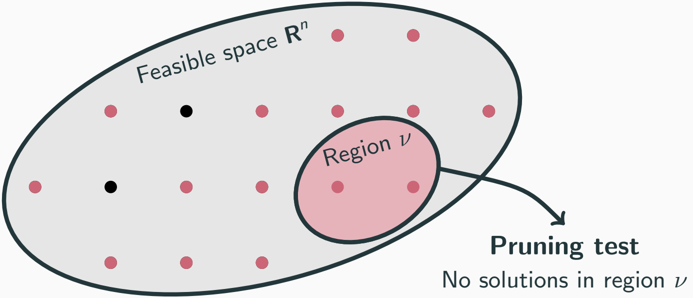
BnB – Algorithmic principle

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



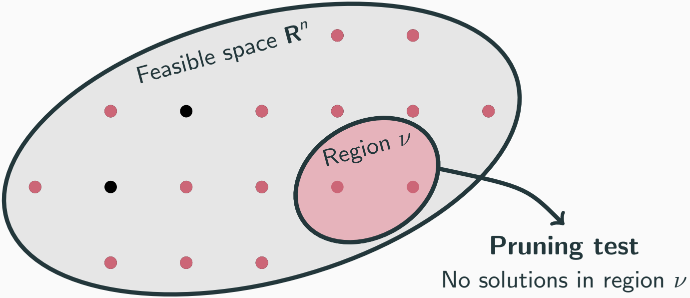
BnB – Algorithmic principle

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



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

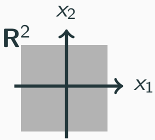
Drive the sparsity of the optimization variable

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

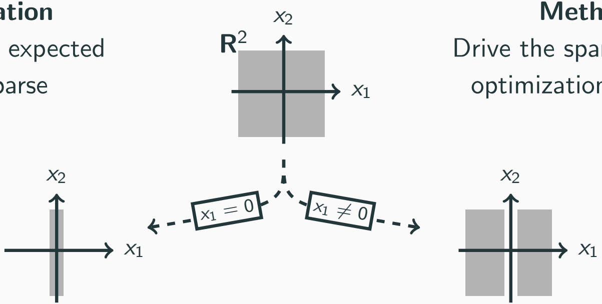
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



BnB – Branching step

Problem

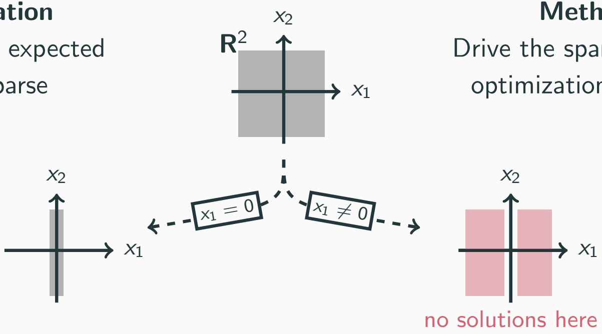
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



Problem

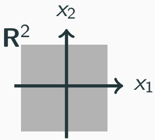
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

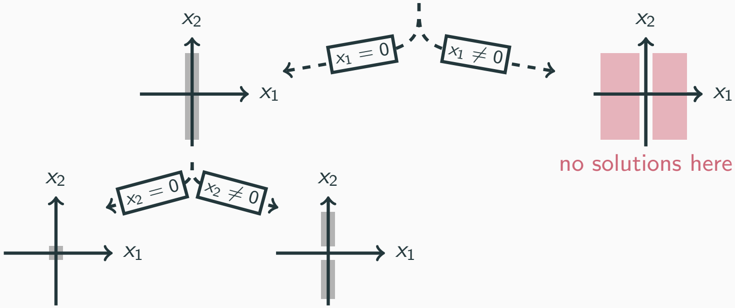
Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable





Problem

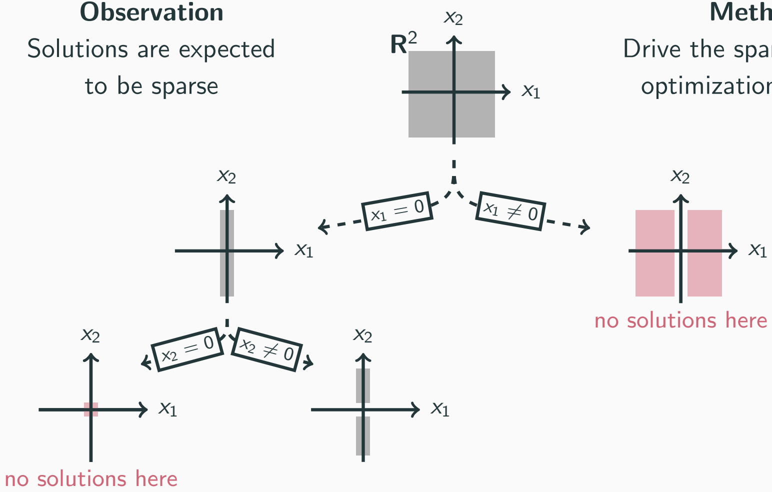
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



BnB – Branching step

Problem

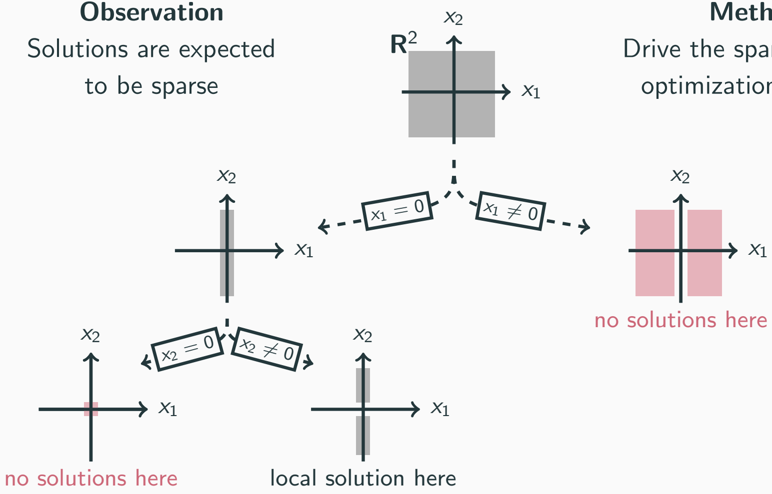
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

Method

Drive the sparsity of the optimization variable



Problem

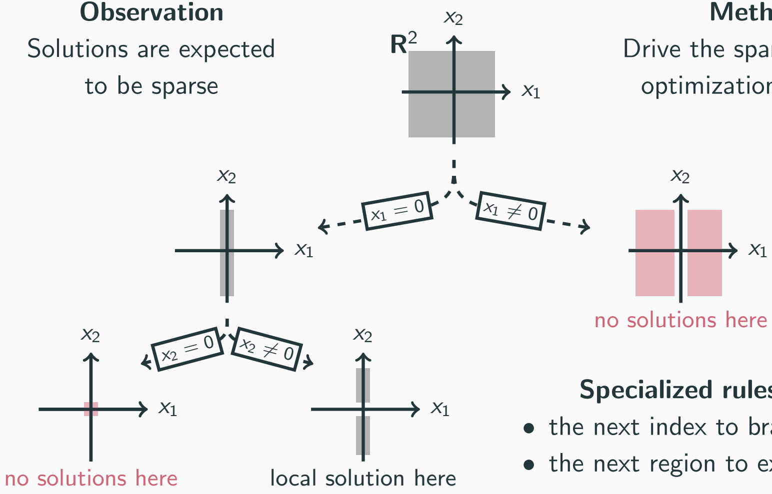
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

Observation

Solutions are expected to be sparse

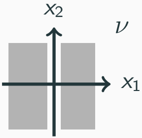
Method

Drive the sparsity of the optimization variable



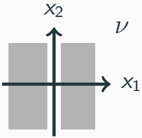
Specialized rules for

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



Does region ν contains optimal solutions ?

BnB – Bounding step

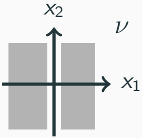


Does region ν contains optimal solutions ?

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

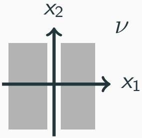
$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

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

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

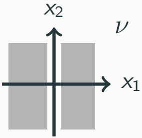
$p^\nu = \min_{\mathbf{x} \in \nu} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$ compare

Pruning test

$p^\nu > p^\star$

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

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

$\downarrow$ restrict to ν

Restriction to region ν

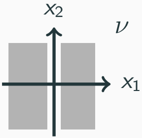
$p^\nu = \min_{\mathbf{x} \in \nu} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

$\downarrow$ compare

Pruning test

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

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

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

restrict to ν

Restriction to region ν

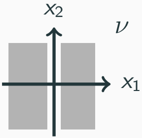
$p^\nu = \min_{\mathbf{x} \in \nu} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$

compare

Pruning test

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

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

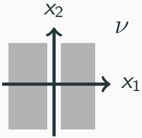
Pruning test

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

Easy task

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

BnB – Bounding step



Does region ν contains optimal solutions ?

Problem

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

$\downarrow$ 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

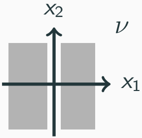
Pruning test

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

Easy task

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

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



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

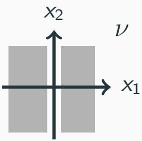
$\boxed{\text{Compute an upper bound on } p^\star}$

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

Main challenge

$\boxed{\text{Compute a lower bound on } p^\nu}$

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

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

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

$\downarrow$ 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

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

Relaxation for region ν

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

set g_{lb}^ν 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 $\mathbf{x} \in \mathbf{R}^n$

BnB – Geometrical intuition

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

$\uparrow$

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

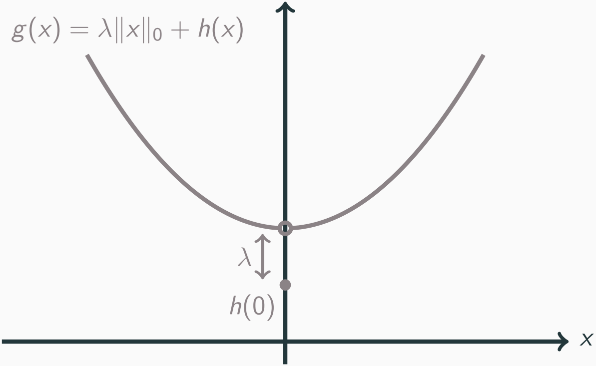
$\downarrow$

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

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** $x \in \mathbf{R}$

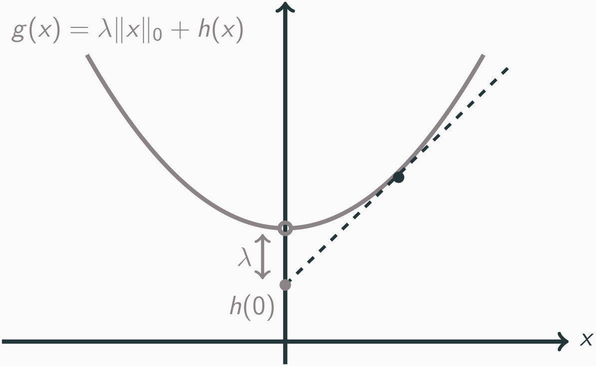


BnB – Geometrical intuition

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** $x \in \mathbf{R}$

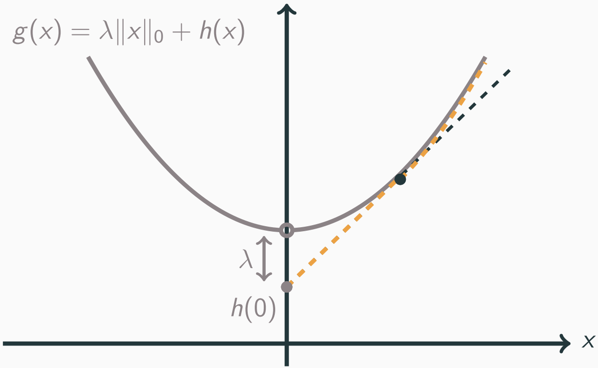


BnB – Geometrical intuition

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** $x \in \mathbf{R}$

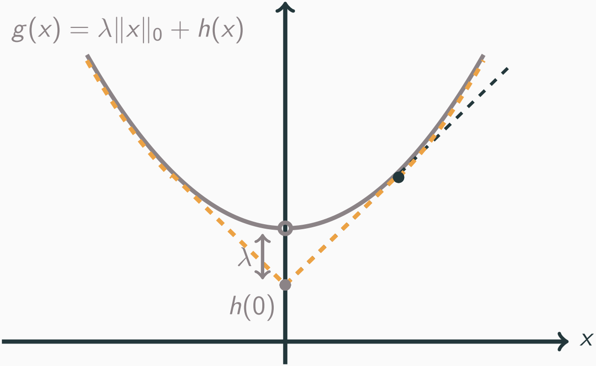


BnB – Geometrical intuition

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** $x \in \mathbf{R}$



Overall solve time

pruning test time $\times$ number of regions processed

BnB – The secrete sauce

Overall solve time

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

Relaxation for region ν

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

BnB – The secrete sauce

Overall solve time

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

Relaxation for region ν

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

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

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

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

$\downarrow$

This is a convex sparse optimization problem

BnB – The secrete sauce

Overall solve time

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

$\downarrow$

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, ...

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

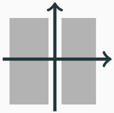
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...

Overall solve time

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

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

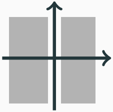
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

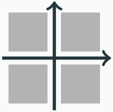
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...



perform degraded but low-cost pruning test

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

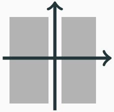
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

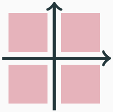
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...



perform degraded but low-cost pruning test

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

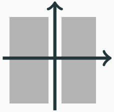
$\rightarrow$ first-order methods

proximal gradient, coordinate descent, ...

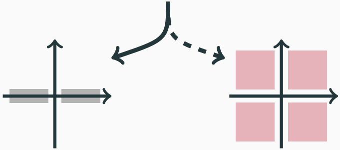
$\rightarrow$ acceleration strategies

working set, screening tests, ...

Simultaneous pruning



processing region ...



continue processing

perform degraded but low-cost pruning test

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 $\mathbf{x} = \mathbf{0}$

This is a convex sparse optimization problem

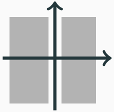
→ first-order methods

proximal gradient, coordinate descent, ...

→ acceleration strategies

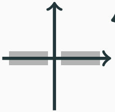
working set, screening tests, ...

Simultaneous pruning

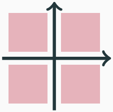


processing region ...

bypass region



continue processing



perform degraded but low-cost pruning test

BnB – Let's sum up

Problem

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

Pipeline

1a) Implement a BnB solver

1b) Use an existing BnB solver

2) Solve the problem

BnB – Let's sum up

Problem

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

Pipeline

1a) Implement a BnB solver

1b) Use an existing BnB solver

2) Solve the problem

Pros

✓ Numerical efficiency

✓ Open-source softwares

✓ Any h separable and coercive

BnB – Let's sum up

Problem

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

Pipeline

1a) Implement a BnB solver

1b) Use an existing BnB solver

2) Solve the problem

Pros

✓ Numerical efficiency

✓ Open-source softwares

✓ Any h separable and coercive

Cons

✗ Less standard pipeline

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 solvers:** `l0bnb` `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 solvers:** `l0bnb` `el0ps`

Instance 1

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \frac{\gamma}{2}\|\mathbf{x}\|_2^2 + \mathrm{Cstr}(\|\mathbf{x}\|_\infty \leq M)$
- `riboflavin` dataset with $\mathbf{A} \in \mathbf{R}^{71\times 4088}$

Numerics – Feature Selection Problem

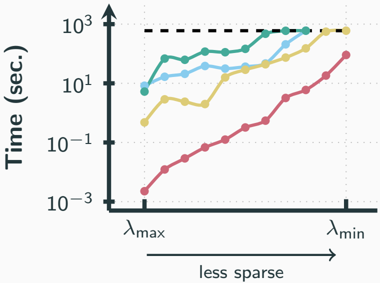
Problem

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

MIP solvers: `cplex` `mosek` **BnB solvers:** `l0bnb` `el0ps`

Instance 1

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \frac{\gamma}{2}\|\mathbf{x}\|_2^2 + \mathrm{Cstr}(\|\mathbf{x}\|_\infty \leq M)$
- `riboflavin` dataset with $\mathbf{A} \in \mathbf{R}^{71\times 4088}$



Numerics – Feature Selection Problem

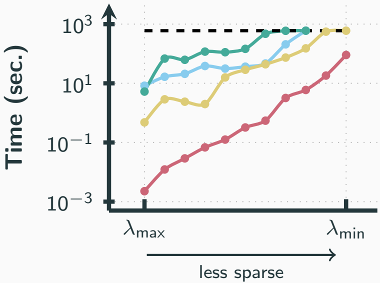
Problem

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

MIP solvers: `cplex` `mosek` **BnB solvers:** `l0bnb` `el0ps`

Instance 1

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \frac{\gamma}{2}\|\mathbf{x}\|_2^2 + \text{Cstr}(\|\mathbf{x}\|_\infty \leq M)$
- `riboflavin` dataset with $\mathbf{A} \in \mathbf{R}^{71\times 4088}$



Instance 2

- $f(\mathbf{x}) = \mathbf{1}^{\mathrm{T}} \log(\mathbf{1} + \exp(-\mathbf{y} \odot \mathbf{A}\mathbf{x}))$
- $h(\mathbf{x}) = \gamma\|\mathbf{x}\|_1 + \text{Cstr}(\|\mathbf{x}\|_\infty \leq M)$
- `leukemia` dataset with $\mathbf{A} \in \mathbf{R}^{38\times 7129}$

Numerics – Feature Selection Problem

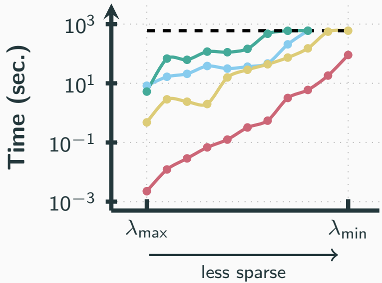
Problem

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

MIP solvers: `cplex` `mosek` **BnB solvers:** `l0bnb` `el0ps`

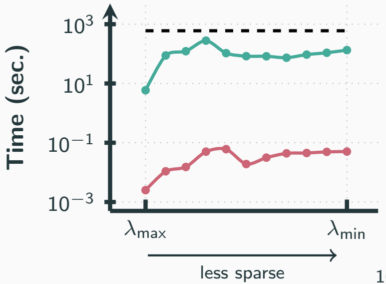
Instance 1

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \frac{\gamma}{2}\|\mathbf{x}\|_2^2 + \mathrm{Cstr}(\|\mathbf{x}\|_\infty \leq M)$
- `riboflavin` dataset with $\mathbf{A} \in \mathbf{R}^{71\times 4088}$



Instance 2

- $f(\mathbf{x}) = \mathbf{1}^{\mathrm{T}}\log(\mathbf{1} + \exp(-\mathbf{y} \odot \mathbf{A}\mathbf{x}))$
- $h(\mathbf{x}) = \gamma\|\mathbf{x}\|_1 + \mathrm{Cstr}(\|\mathbf{x}\|_\infty \leq M)$
- `leukemia` dataset with $\mathbf{A} \in \mathbf{R}^{38\times 7129}$



Conclusion

Take-home messages

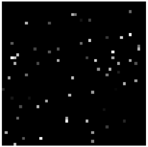
- Although NP-hard, ℓ_0-problems are of practical interest
- There exists methods to tackle them exactly
    - MIP-based off-the-shelf solvers
    - BnB-based specialized algorithms
    - Structure-exploitation is key for numerical efficiency

Question time !

🤓

Compressed sensing

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



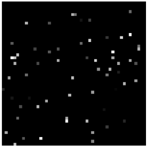
n pixels

Compressed sensing

Linear compression

Sparse signal

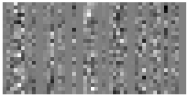
$\mathbf{x} \in \mathbf{R}^n$



n pixels

Compressed signal

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



m pixels, $m \ll n$

Compressed sensing

Sparse signal

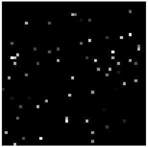
$\mathbf{x} \in \mathbf{R}^n$

Linear compression

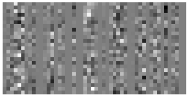
Recover $\mathbf{x}$ from $\mathbf{A}$ and $\mathbf{y}$

Compressed signal

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



n pixels



m pixels, $m \ll n$

Compressed sensing

Sparse signal

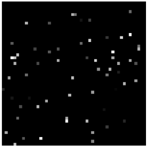
$\mathbf{x} \in \mathbf{R}^n$

Linear compression

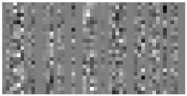
Recover $\mathbf{x}$ from $\mathbf{A}$ and $\mathbf{y}$

Compressed signal

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



n pixels



m pixels, $m \ll n$

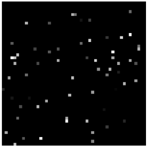
Goal

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

Compressed sensing

Sparse signal

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



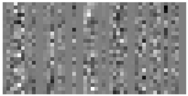
n pixels

Linear compression

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

Compressed signal

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



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

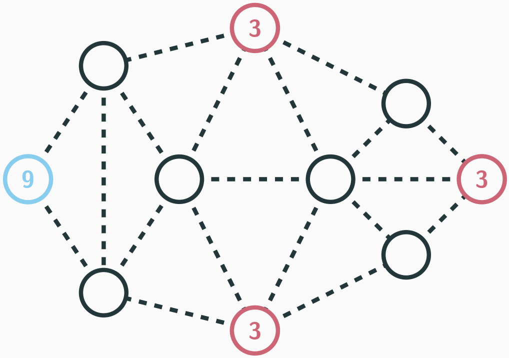
Model accuracy
Loss $\mathcal{L}_\phi(\mathbf{A}\mathbf{x}, \mathbf{y})$

Model explainability
Use few features

Goal

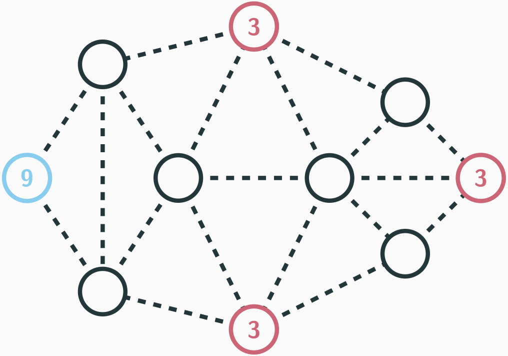
Find $\mathbf{x}$ sparse such that $\mathcal{L}_\phi(\mathbf{A}\mathbf{x}, \mathbf{y})$ is small

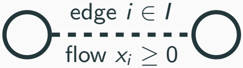
Network design



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

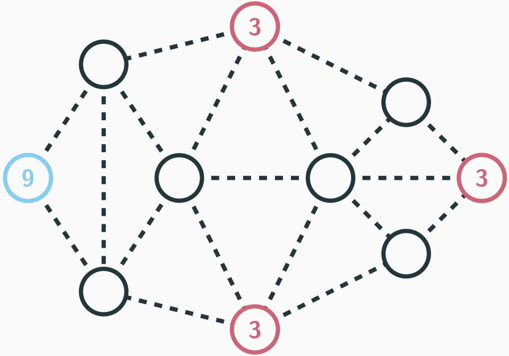
Network design





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

Network design

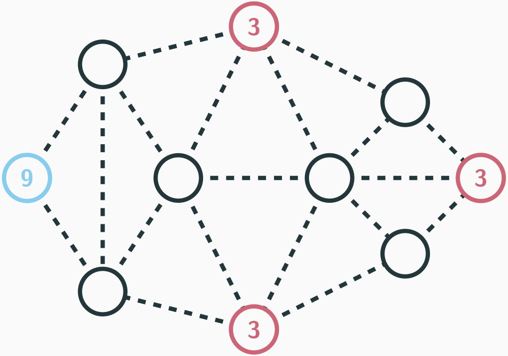


Which edges to build to transport products from source to sink nodes ?

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 ?

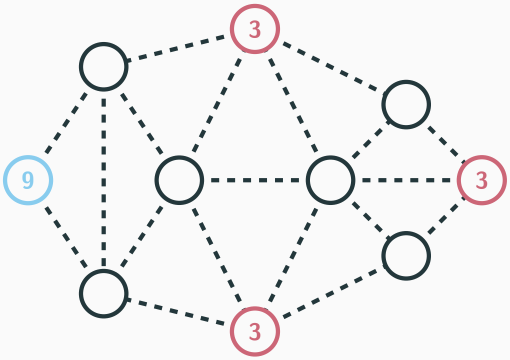
construct edge $i \in I$ if $x_i > 0$

pay construction cost c

Question

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

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 ?

$\updownarrow$

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**

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