# Optimization methods for $\ell_0$-problems

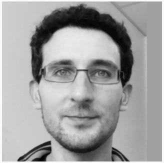
Théo Guyard

ML MTP – December 9th, 2024

# Optimization methods for $\ell_0$-problems


Théo Guyard


ML MTP – December 9th, 2024




**Cédric Herzet**
Inria/Ensai Rennes

**Clément Elvira**
CentraleSupélec Rennes

**Ayșe-Nur Arslan**
Inria Bordeaux

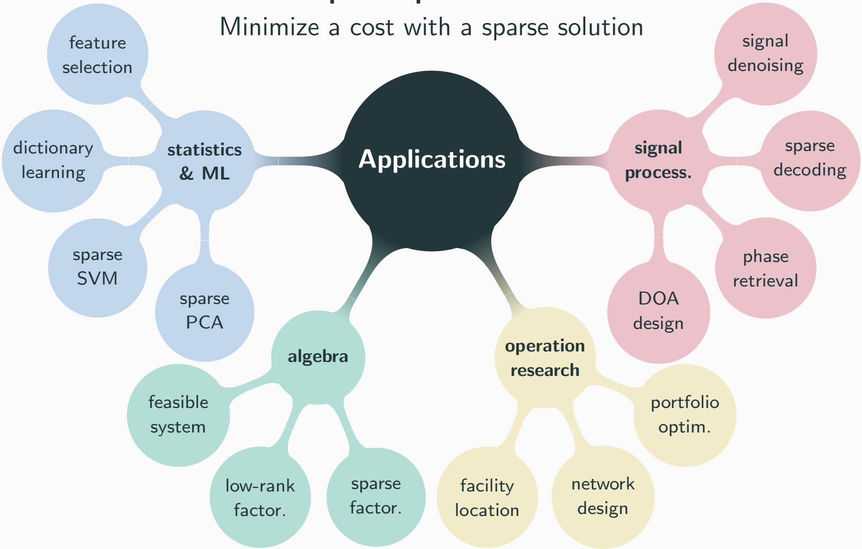
**Sparse optimization**

Minimize a cost with a sparse solution

# Sparse optimization

**Sparse optimization**

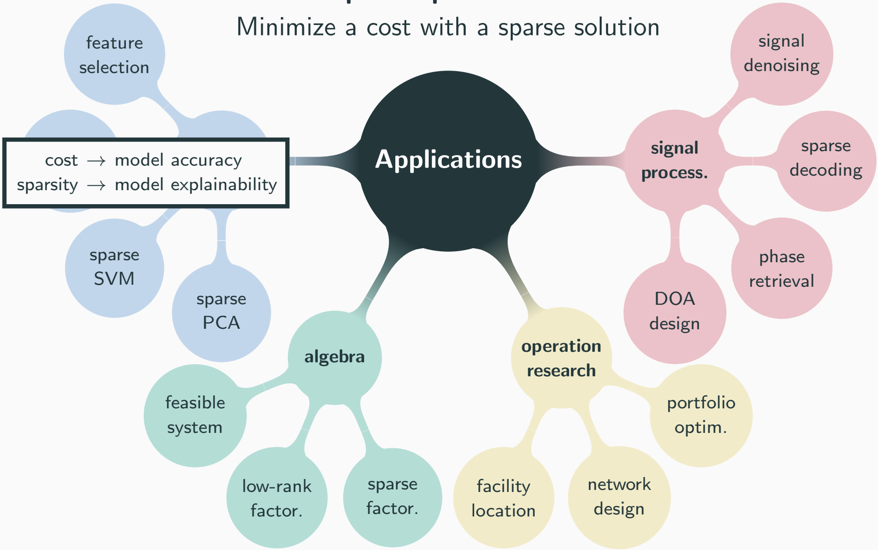
Minimize a cost with a sparse solution




A. Tillmann *et. al* (2024)

## Sparse optimization

Minimize a cost with a sparse solution

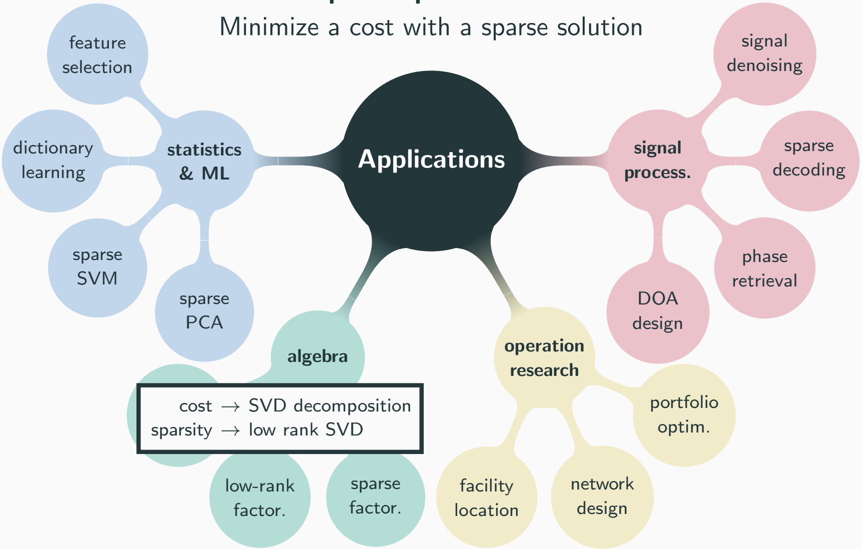



A. Tillmann *et. al* (2024)

# Sparse optimization

**Sparse optimization**

Minimize a cost with a sparse solution



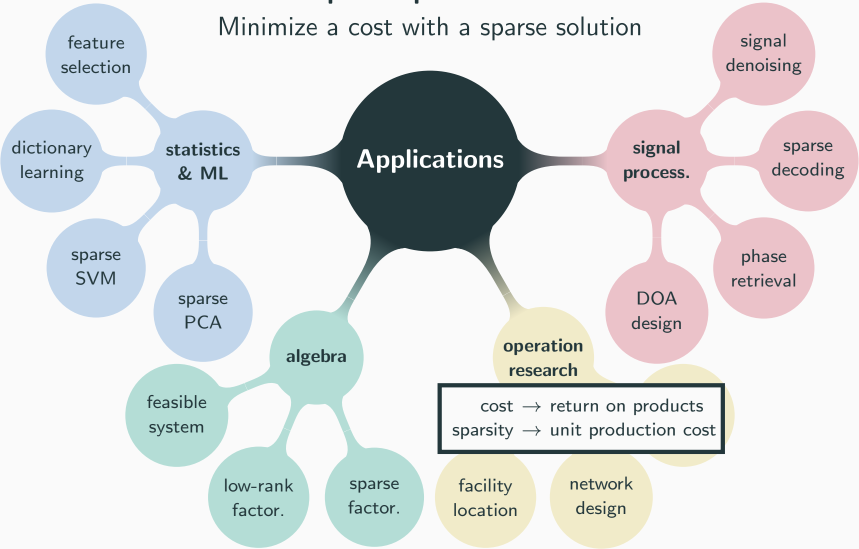

cost $\rightarrow$ SVD decomposition
sparsity $\rightarrow$ low rank SVD

A. Tillmann *et. al* (2024)

# Sparse optimization

**Sparse optimization**

Minimize a cost with a sparse solution

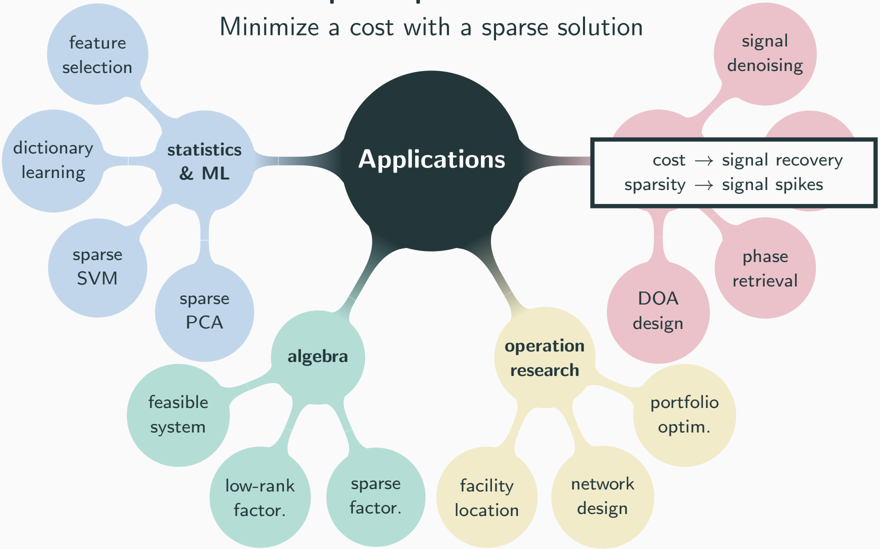



A. Tillmann *et. al* (2024)

# Sparse optimization

**Sparse optimization**

Minimize a cost with a sparse solution




A. Tillmann *et. al* (2024)

# Minimized, constrained, or regularized problem ?

**Sparse optimization**

Minimize a cost with a sparse solution

# Minimized, constrained, or regularized problem ?

**Sparse optimization**

Minimize a cost with a sparse solution

quantify cost

$f(\mathbf{x})$

thing to minimize

# Minimized, constrained, or regularized problem ?

**Sparse optimization**

Minimize a cost with a sparse solution

quantify cost

$f(\mathbf{x})$

thing to minimize

quantify sparsity

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

counts non-zeros

# Minimized, constrained, or regularized problem ?

**Sparse optimization**

Minimize a cost with a sparse solution

quantify cost

quantify sparsity

$f(\mathbf{x})$

thing to minimize

optim. problem

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

counts non-zeros

# Minimized, constrained, or regularized problem ?

**Sparse optimization**

Minimize a cost with a sparse solution

$\boxed{\text{quantify cost}}$

$\boxed{\text{quantify sparsity}}$

$f(\mathbf{x})$

thing to minimize

optim. problem

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

counts non-zeros

**Constrained version**

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbf{R}^n} & f(\mathbf{x}) \\ \text{subject to} & \|\mathbf{x}\|_0 \leq s \end{array}$$

# Minimized, constrained, or regularized problem ?

**Sparse optimization**

Minimize a cost with a sparse solution

quantify cost

quantify sparsity

$f(\mathbf{x})$

thing to minimize

optim. problem

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

counts non-zeros

**Constrained version**

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbf{R}^n} & f(\mathbf{x}) \\ \text{subject to} & \|\mathbf{x}\|_0 \leq s \end{array}$$

**Minimized version**

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbf{R}^n} & \|\mathbf{x}\|_0 \\ \text{subject to} & f(\mathbf{x}) \leq \epsilon \end{array}$$

# Minimized, constrained, or regularized problem ?

**Sparse optimization**

Minimize a cost with a sparse solution

$\boxed{\text{quantify cost}}$ $\boxed{\text{quantify sparsity}}$

$f(\mathbf{x})$ $\|\mathbf{x}\|_0$

thing to minimize $\xleftrightarrow{\text{optim. problem}}$ counts non-zeros

**Constrained version**

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbf{R}^n} & f(\mathbf{x}) \\ \text{subject to} & \|\mathbf{x}\|_0 \leq s \end{array}$$

**Minimized version**

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbf{R}^n} & \|\mathbf{x}\|_0 \\ \text{subject to} & f(\mathbf{x}) \leq \epsilon \end{array}$$

**Regularized version**

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

# Minimized, constrained, or regularized problem ?

**Sparse optimization**

Minimize a cost with a sparse solution

quantify cost

quantify sparsity

$f(\mathbf{x})$

thing to minimize

optim. problem

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

counts non-zeros

**Constrained version**

$$\begin{aligned} \min_{\mathbf{x}\in\mathbf{R}^n} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \|\mathbf{x}\|_0 \le s \end{aligned}$$

**Minimized version**

$$\begin{aligned} \min_{\mathbf{x}\in\mathbf{R}^n} \quad & \|\mathbf{x}\|_0 \\ \text{subject to} \quad & f(\mathbf{x}) \le \epsilon \end{aligned}$$

**Regularized version**

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

# A bit of history

**Problem**

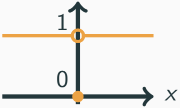
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

# A bit of history

**Problem**

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

**1d graph**

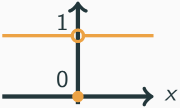


# A bit of history

**Problem**

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

**1d graph**




non-convex, non-diff.,
non-continuous, ...

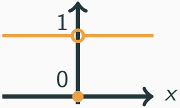
# A bit of history

**Problem**

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

**NP-hard to solve**

**1d graph**




non-convex, non-diff.,
non-continuous, ...

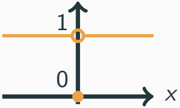
# A bit of history

**Problem**

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

**NP-hard to solve**

**1d graph**

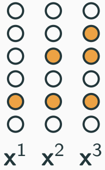



non-convex, non-diff.,
non-continuous, ...

**1995**

Heuristics

MP, OMP, ...
S. Mallat (1993)



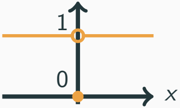
# A bit of history

**Problem**

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

**NP-hard to solve**

**1d graph**




non-convex, non-diff.,
non-continuous, ...

**1995**

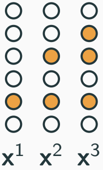
Heuristics

MP, OMP, ...
S. Mallat (1993)

**2000**

Recovery cond.

RIP, NSP, ...
E. Candes (2004)




OMP solves
$\ell_0$-problem
under RIP

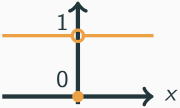
## A bit of history

**Problem**

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

**NP-hard to solve**

**1d graph**

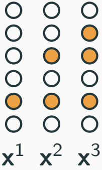



non-convex, non-diff., non-continuous, ...

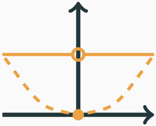
**1995** — Heuristics
MP, OMP, ...
S. Mallat (1993)

**2000** — Recovery cond.
RIP, NSP, ...
E. Candes (2004)

**2005** — Convex approx.
Lasso, Elastic-Net, ...
R. Tibshirani (2005)




OMP solves $\ell_0$-problem under RIP



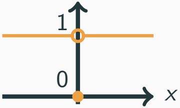
# A bit of history

**Problem**

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

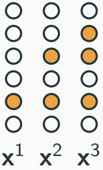
**NP-hard to solve**

**1d graph**

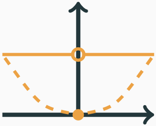


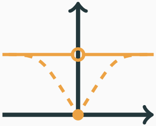

non-convex, non-diff.,
non-continuous, ...

| **1995** | **2000** | **2005** | **2010** |
|---|---|---|---|
| Heuristics | Recovery cond. | Convex approx. | Concave approx. |
| MP, OMP, ... | RIP, NSP, ... | Lasso, Elastic-Net, ... | SCAD, MCP, ... |
| S. Mallat (1993) | E. Candes (2004) | R. Tibshirani (2005) | C. Zhang (2010) |




OMP solves
$\ell_0$-problem
under RIP





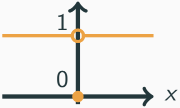
# A bit of history

**Problem**

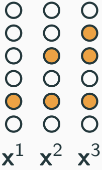
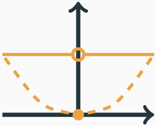
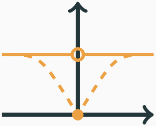
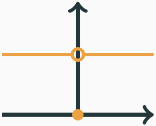
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

**NP-hard to solve**

**1d graph**




non-convex, non-diff.,
non-continuous, ...

| **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 $\ell_0$-problem under RIP |  |  |  |

# Topic of this talk

Ok, $\ell_0$-problems can be of practical interest !

# Topic of this talk

Ok, $\ell_0$-problems can be of practical interest !

What are the solution methods then ?

# Topic of this talk

Ok, $\ell_0$-problems can be of practical interest !

What are the solution methods then ?

**1) MIP-based methods**

Convenient for practitioners

Poor numerical performances

# Topic of this talk

Ok, $\ell_0$-problems can be of practical interest !

What are the solution methods then ?

**1) MIP-based methods**

Convenient for practitioners

Poor numerical performances

**2) Specialized Branch-and-Bound**

More sophisticated mechanism

Better numerical performances

# Topic of this talk

Ok, $\ell_0$-problems can be of practical interest !

What are the solution methods then ?

**1) MIP-based methods**

Convenient for practitioners

Poor numerical performances

**2) Specialized Branch-and-Bound**

More sophisticated mechanism

Better numerical performances

**High-level concepts and practical tools**

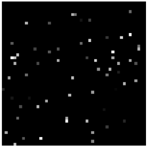
**Question time !**

🤓

# Compressed sensing

Sparse signal

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

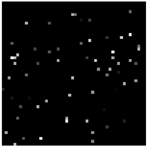


$n$ pixels

# Compressed sensing

Sparse signal

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

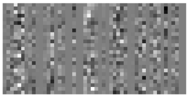


$n$ pixels

Linear compression

Compressed signal

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



$m$ pixels, $m \ll n$

# Compressed sensing

Sparse signal

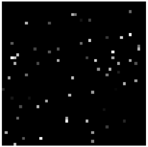
$\mathbf{x} \in \mathbf{R}^n$

Linear compression

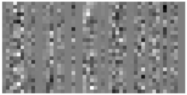
Recover $\mathbf{x}$ from $\mathbf{A}$ and $\mathbf{y}$

Compressed signal

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



$n$ pixels

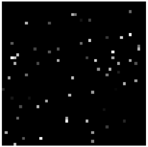


$m$ pixels, $m \ll n$

# Compressed sensing

Sparse signal

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



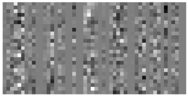
$n$ pixels

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, $m \ll n$

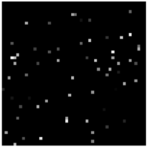
**Goal**

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

# Compressed sensing

Sparse signal

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



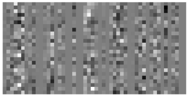
$n$ pixels

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** | $a_{1,1}$ | $a_{1,2}$ | ... | $a_{1,n}$ | $y_1$ |
| **Sample 2** | $a_{2,1}$ | $a_{2,2}$ | ... | $a_{2,n}$ | $y_2$ |
| **Sample 3** | $a_{3,1}$ | $\boxed{\mathbf{A} \in \mathbf{R}^{m \times n}}$ | ... | $a_{3,n}$ | $\boxed{\mathbf{y} \in \mathbf{R}^{m}}$ |
| ... | ... | ... | ... | ... | ... |
| **Sample m** | $a_{m,1}$ | $a_{m,2}$ | ... | $a_{m,n}$ | $y_m$ |

# 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** | $a_{m,1}$ | $a_{m,2}$ | $\cdots$ | $a_{m,n}$ | $y_m$ |

$$\text{Features } \mathbf{A} \in \mathbf{R}^{m \times n} \xleftrightarrow[\text{weights } \mathbf{x} \in \mathbf{R}^{n}]{} \text{Target } \mathbf{y} = \phi(\mathbf{A}\mathbf{x})$$

**Model accuracy**

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

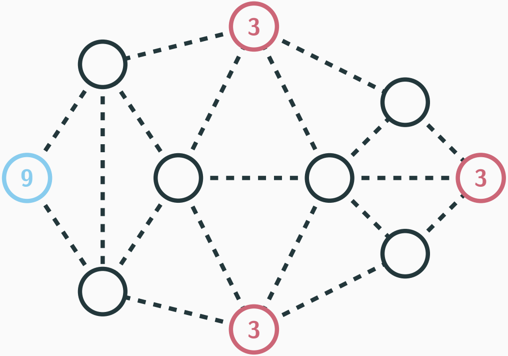
**Model explainability**

Use few features

**Goal**

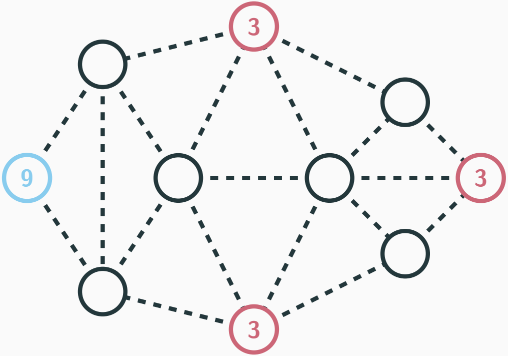
Find $\mathbf{x}$ sparse such that $\mathcal{L}_\phi(\mathbf{A}\mathbf{x}, \mathbf{y})$ is small

# Network design




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

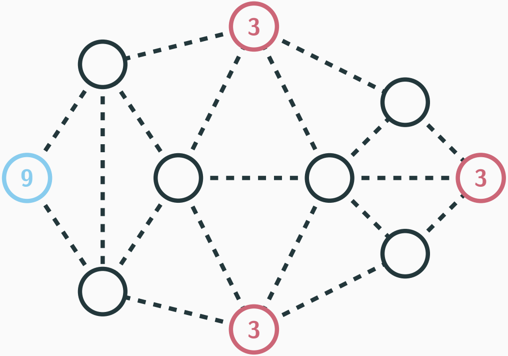
# Network design





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

# Network design

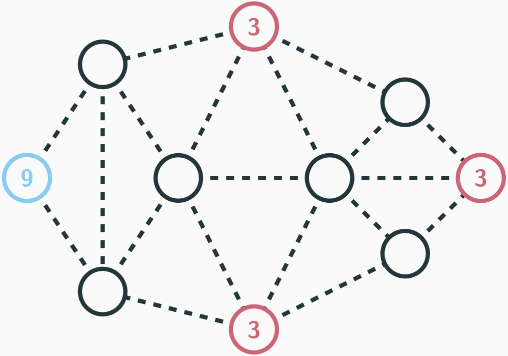



Which edges to build to transport products from source to sink nodes ?


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 ?

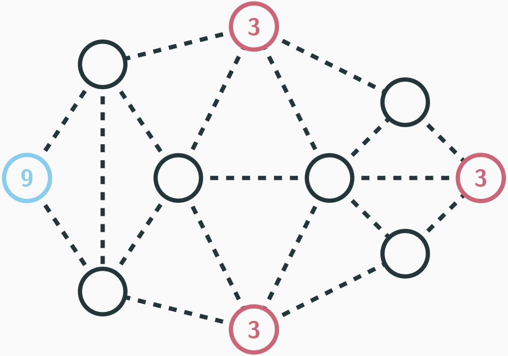

construct edge $i \in I$ if $x_i > 0$

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