Optimization methods for ℓ_0-problems

Théo Guyard

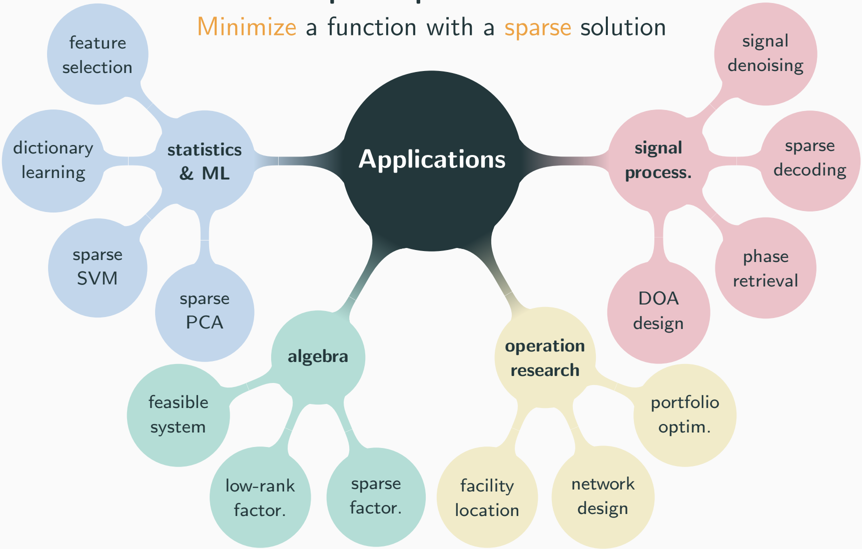
ML-MTP workshop – December 9th, 2024

Sparse optimization

Minimize a function with a sparse solution

Sparse optimization

Minimize a function with a sparse solution

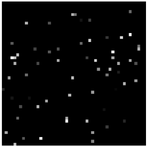


A. Tillmann *et. al* (2024)

Compressed sensing

Sparse signal

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



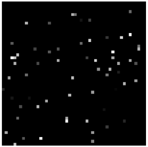
n pixels

Compressed sensing

Linear compression

Sparse signal

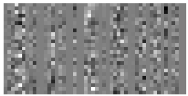
$\mathbf{x} \in \mathbf{R}^n$



n pixels

Observation

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

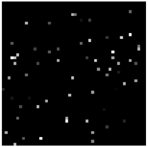


m pixels, $m \ll n$

Compressed sensing

Sparse signal

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



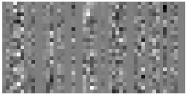
n pixels

Linear compression

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

Observation

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



m pixels, $m \ll n$

Compressed sensing

Linear compression

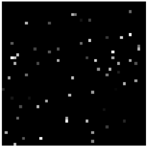
Sparse signal

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

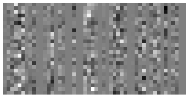
Observation

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

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



n pixels



m pixels, $m \ll n$

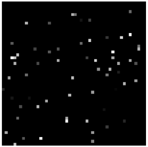
Goal

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

Compressed sensing

Sparse signal

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



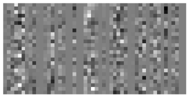
n pixels

Linear compression

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

Observation

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



m pixels, $m \ll n$

Goal

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

no unique solution

Goal

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

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]{\qquad\qquad} $ 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 explicability

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 explicability

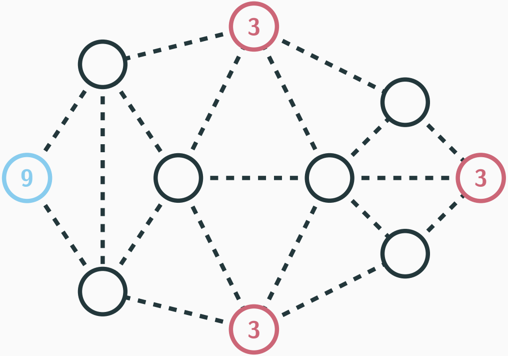
Use few features

$\downarrow$ $\downarrow$

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 ?

construction cost c per edge
transportation cost $Q(\mathbf{x})$
capacity $\mathbf{0} \leq \mathbf{x} \leq \mathbf{x}_{\text{ub}}$
flow conservation $\mathbf{D}\mathbf{x} \leq \mathbf{d}$

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a function with a sparse solution

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a function with a sparse solution

quantify cost

$f(\mathbf{x})$

thing to minimize

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a function 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 function with a sparse solution

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

$f(\mathbf{x})$

thing to minimize

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

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

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 function with a sparse solution

quantify cost

$f(\mathbf{x})$

thing to minimize

quantify sparsity

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

counts non-zeros

Constrained version

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

Minimized version

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

Minimized, constrained, or regularized problem ?

Sparse optimization

Minimize a function with a sparse solution

quantify cost

$f(\mathbf{x})$

thing to minimize

quantify sparsity

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

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

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

Regularized version

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

A bit of history

Problem

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

NP-hard to solve

Problem

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

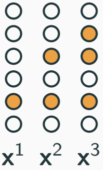
NP-hard to solve

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

1995

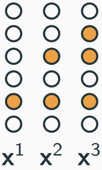
Heuristics

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

2000

Recovery cond.

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



OMP solves
ℓ_0-problem
under RIP

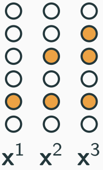
A bit of history

Problem

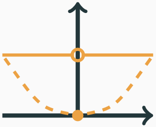
$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$

NP-hard to solve

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



OMP solves
ℓ_0-problem
under RIP



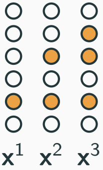
A bit of history

Problem

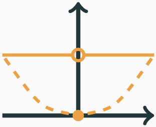
$$\min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

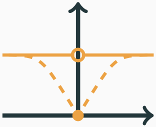
NP-hard to solve

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
ℓ_0-problem
under RIP





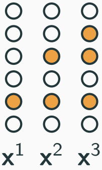
A bit of history

Problem

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

NP-hard to solve

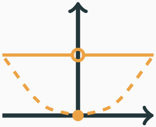
1995	2000	2005	2010	2015
Heuristics	Recovery cond.	Convex approx.	Concave approx.	Exact methods
MP, OMP, ...	RIP, NSP, ...	Lasso, Elastic-Net, ...	SCAD, MCP, ...	MIP, BnB, ...
S. Mallat (1993)	E. Candes (2004)	R. Tibshirani (2005)	C. Zhang (2010)	D. Bertsimas (2016)

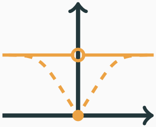


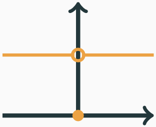
OMP solves

ℓ_0-problem

under RIP







Complexity-tractability balance

[Slide, tell what's my point with this talk]

Mixed-Integer Programming

MIP – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

$\downarrow$

Problem

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

MIP – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

$\downarrow$

Problem

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

$\downarrow$

MIP formulation

Standardized expressions

MIP – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

↓

Problem

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

↓

MIP formulation

Standardized expressions

Modelling framework

Python → `cvxpy`

Julia → `JuMP`

C++ → `CMPL`

Matlab → `YALMIP`

...

MIP – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

Problem

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

Modelling framework

Python $\rightarrow$ `cvxpy`

Julia $\rightarrow$ `JuMP`

C++ $\rightarrow$ `CMPL`

Matlab $\rightarrow$ `YALMIP`

...

MIP formulation

Standardized expressions

Problem solutions

MIP – Pipeline

Application

ML, Stats, Signal, Operation Research, ...

↓

Problem

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

↓

Modelling framework

Python → `cvxpy`

Julia → `JuMP`

C++ → `CMPL`

Matlab → `YALMIP`

...

MIP formulation

Standardized expressions

↓

Generic MIP solver

Gurobi, CPLEX, SCIP,

Mosek, CbC, ...

Problem solutions

MIP – Formulation

Problem

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

MIP formulation

Use standardized expressions

linear, quadratic, conic, ...

MIP – Formulation

Problem

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

$\downarrow$ linearize ℓ_0-norm

Lifted formulation

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

MIP formulation

Use standardized expressions

linear, quadratic, conic, ...

Lifted ℓ_0-norm formulation

$\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$ with $\mathbf{x} \in \mathbf{R}^n$ and $\mathbf{z} \in \{0,1\}^n$

if $x_i = 0 \iff z_i = 0$ for all $i \in [1, n]$

MIP – Formulation

Problem

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

$\downarrow$ linearize ℓ_0-norm

Lifted formulation

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

$\downarrow$ avoid logical cstr.

MIP formulation

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

MIP formulation

Use standardized expressions

linear, quadratic, conic, ...

Lifted ℓ_0-norm formulation

$\|\mathbf{x}\|_0 = \mathbf{1}^{\mathrm{T}}\mathbf{z}$ with $\mathbf{x}\in\mathbf{R}^n$ and $\mathbf{z}\in\{0,1\}^n$

if $x_i = 0 \iff z_i = 0$ for all $i\in[1,n]$

Logical constraint standardization

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

MIP – Hands-on with cvxpy

Sparse regression

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

$\downarrow$

Optimization problem

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

- $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- $h(\mathbf{x}) = \mathrm{Ind}(-M \leq \mathbf{x} \leq M)$

$\downarrow$

MIP formulation

$$\begin{cases} \min\ f(\mathbf{x}) + \lambda\mathbf{1}^{\mathrm{T}}\mathbf{z} \\ \text{s.t.}\ \ -M\mathbf{z} \leq \mathbf{x} \leq M\mathbf{z} \\ \qquad \mathbf{x} \in \mathbf{R}^n, \mathbf{z} \in \{0,1\}^n \end{cases}$$

```
$ pip install cvxpy
```

```
import cvxpy as cp
from sklearn.datasets import make_regres

# Generate sparse regression data
A, y = make_regression()

# Define variables
n = A.shape[1]
x = cp.Variable(n)
z = cp.Variable(n, boolean=True)

# Define objective and constraints
objective = cp.Minimize(
    cp.sum_squares(A @ x - y) + 10 * cp.
        )
)
constraints = [-M * z <= x, x <= M * z]

# Solve the problem using Gurobi
problem = cp.Problem(objective, constrai
problem.solve(solver=cp.GUROBI)
```

MIP – Let's sum up

Problem

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

Pipeline

1) Introduce binary variable

2) Establish MIP formulation

3) Use generic MIP solvers

Pros

✓ Rich MIP literature

✓ Black-box solvers

✓ Convenient for practitioners

Cons

✗ Mostly commercial solvers

✗ Unable to exploit structure

✗ Performance issues

Question time !

🤓