Screen & Relax

Accelerating the resolution of the Elastic-Net by safe identification of the solution support

Theo Guyard[⋆], Cedric Herzet[†], Clement Elvira[‡]

ICASSP 2022

[⋆] Univ Rennes, INSA Rennes, CNRS, IRMAR-UMR 6625, F-35000 Rennes, France
[†] INRIA Rennes-Bretagne Atlantique, Campus de Beaulieu, 35000 Rennes, France
[‡] SCEE/IETR UMR CNRS 6164, CentraleSupélec, 35510 Cesson Sévigné, France

Table of contents

The Elastic-Net problem

Sparse-linear problem

Ingredients :

- An observation $y \in \mathrm{R}^m$
- A dictionary $A = [a_i]_{i=1}^n \in \mathrm{R}^{m \times n}$ (columns $\equiv$ atoms)

Sparse-linear problem

Ingredients :

- An observation $y \in \mathbb{R}^m$
- A dictionary $A = [a_i]_{i=1}^n \in \mathbb{R}^{m \times n}$ (columns $\equiv$ atoms)

Objectives :

- Approximate the observation as a linear combination of the atoms
- The linear combination must be *sparse*

Sparse-linear problem

Ingredients :

- An observation $y \in R^m$
- A dictionary $A = [a_i]_{i=1}^n \in R^{m \times n}$ (columns $\equiv$ atoms)

Objectives :

- Approximate the observation as a linear combination of the atoms
- The linear combination must be *sparse*

Problem

Find x *sparse* such that $y \simeq Ax$

Sparse-linear problem

Ingredients :

- An observation $y \in \mathbb{R}^m$
- A dictionary $A = [a_i]_{i=1}^n \in \mathbb{R}^{m \times n}$ (columns $\equiv$ atoms)

Objectives :

- Approximate the observation as a linear combination of the atoms
- The linear combination must be *sparse*

Problem

Find x *sparse* such that $y \simeq Ax$

The vector x weights each atom in the approximation.

The Elastic-net problem

Idea : Consider the problem

Elastic-net

$$x^\star = \operatorname{argmin}_x \left\{ P(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 + \frac{\gamma}{2}\|x\|_2^2 \right\} \qquad (\mathcal{P})$$

where $\lambda > 0$ and $\gamma > 0$ are two hyperparameters.

The Elastic-net problem

Idea : Consider the problem

Elastic-net

$$x^\star = \operatorname{argmin}_x \left\{ \mathrm{P}(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 + \frac{\gamma}{2}\|x\|_2^2 \right\} \qquad (\mathcal{P})$$

where $\lambda > 0$ and $\gamma > 0$ are two hyperparameters.

Properties :

- Ensures a good approximation
- Induces sparsity
- Good statistical properties
- Convex problem

Screening and relaxing tests

Main idea

Sparse problem :

- Where are zero entries of $x^\star$?
- Where are non-zero entries of $x^\star$?

Main idea

Sparse problem :

- Where are zero entries of $x^\star$?
- Where are non-zero entries of $x^\star$?
- Can we accelerate solution methods using this knowledge ?

Main idea

Sparse problem :

- Where are zero entries of $x^\star$?
- Where are non-zero entries of $x^\star$?
- Can we accelerate solution methods using this knowledge ?

$\rightarrow$ Spoiler alert : yes !

$\rightarrow$ We can leverage duality and optimality conditions

Main idea

Sparse problem :

- Where are zero entries of $x^\star$?
- Where are non-zero entries of $x^\star$?
- Can we accelerate solution methods using this knowledge ?

$\rightarrow$ Spoiler alert : yes !

$\rightarrow$ We can leverage duality and optimality conditions

Fenchel dual of $(\mathcal{P})$

$$u^\star = \mathrm{argmax}_u \left\{ D(u) = \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y-u\|_2^2 - \frac{1}{2\gamma}\|[|A^T u| - \lambda]_+\|_2^2 \right\} \quad (\mathcal{D})$$

Screening and relaxing tests

Optimality conditions :

$$\begin{aligned} |\mathsf{a}_i^{\mathrm{T}}\mathsf{u}^\star| \leq \lambda \quad &\Longleftrightarrow \quad \mathsf{x}^\star(i) = 0 \\ |\mathsf{a}_i^{\mathrm{T}}\mathsf{u}^\star| > \lambda \quad &\Longleftrightarrow \quad \mathsf{x}^\star(i) \neq 0 \end{aligned}$$

Screening and relaxing tests

Optimality conditions :

$$\begin{aligned} |\mathrm{a}_i^{\mathrm{T}}\mathrm{u}^\star| \leq \lambda \quad &\Longleftrightarrow \quad \mathrm{x}^\star(i) = 0 \\ |\mathrm{a}_i^{\mathrm{T}}\mathrm{u}^\star| > \lambda \quad &\Longleftrightarrow \quad \mathrm{x}^\star(i) \neq 0 \end{aligned}$$

$$\rightsquigarrow$$

Relaxed optimality condition : Let $\mathbb{S}(\mathrm{u}, r)$ be a spherical region containing $\mathrm{u}^\star$, then

$$\begin{aligned} |\mathrm{a}_i^{\mathrm{T}}\mathrm{u}| + r < \lambda \quad &\Longrightarrow \quad \mathrm{x}^\star(i) = 0 \qquad \text{(screening test)} \\ |\mathrm{a}_i^{\mathrm{T}}\mathrm{u}| - r > \lambda \quad &\Longrightarrow \quad \mathrm{x}^\star(i) \neq 0 \qquad \text{(relaxing test)} \end{aligned}$$

Dimensionality reduction

Problem reduction

With screening test

Zero entries of $x^\star$ can be discarded from the problem without changing the objective value.

Problem reduction

With screening test

Zero entries of $x^\star$ can be discarded from the problem without changing the objective value.

With relaxing test

Nonzero entries of $x^\star$ can be expressed as a linear combination of all the other entries.

Problem reformulation

Let $(\mathcal{S}_0, \mathcal{S}_\pm, \mathcal{S}_*)$ be subsets of zero, non-zero and unclassified indices of $\mathsf{x}^\star$:

$$\mathsf{x}^\star = \operatorname{argmin}_{\mathsf{x}} \left\{ \mathrm{P}(\mathsf{x}) = \tfrac{1}{2}\|\mathsf{y} - \mathsf{A}\mathsf{x}\|_2^2 + \lambda\|\mathsf{x}\|_1 + \tfrac{\gamma}{2}\|\mathsf{x}\|_2^2 \right\}$$

Problem reformulation

Let $(\mathcal{S}_0, \mathcal{S}_\pm, \mathcal{S}_*)$ be subsets of zero, non-zero and unclassified indices of $\mathsf{x}^\star$:

$$\mathsf{x}^\star = \operatorname{argmin}_{\mathsf{x}} \left\{ \mathrm{P}(\mathsf{x}) = \tfrac{1}{2}\|\mathsf{y} - \mathsf{A}\mathsf{x}\|_2^2 + \lambda\|\mathsf{x}\|_1 + \tfrac{\gamma}{2}\|\mathsf{x}\|_2^2 \right\}$$

Solve an *n* dimensional problem

Problem reformulation

Let $(\mathcal{S}_0, \mathcal{S}_\pm, \mathcal{S}_*)$ be subsets of zero, non-zero and unclassified indices of $\mathrm{x}^\star$:

$$\mathrm{x}^\star = \operatorname{argmin}_{\mathrm{x}} \left\{ \mathrm{P}(\mathrm{x}) = \tfrac{1}{2}\|\mathrm{y} - \mathrm{A}\mathrm{x}\|_2^2 + \lambda\|\mathrm{x}\|_1 + \tfrac{\gamma}{2}\|\mathrm{x}\|_2^2 \right\}$$

Solve an *n* dimensional problem

$$\rightsquigarrow$$

$$\begin{aligned} \mathrm{x}^\star_{\mathcal{S}_*} &= \operatorname{argmin}_{\mathrm{x}} \left\{ \tilde{\mathrm{P}}(\mathrm{x}) = \tfrac{1}{2}\|\tilde{\mathrm{y}} - \tilde{\mathrm{A}}\mathrm{x}\|_2^2 + \lambda\|\mathrm{x}\|_1 + \tfrac{\gamma}{2}\|\mathrm{x}\|_{\mathrm{M}}^2 \right\} \\ \mathrm{x}^\star_{\mathcal{S}_\pm} &= \mathrm{B}\mathrm{x}^\star_{\mathcal{S}_*} + \mathrm{b} \\ \mathrm{x}^\star_{\mathcal{S}_0} &= 0 \end{aligned}$$

Problem reformulation

Let $(\mathcal{S}_0, \mathcal{S}_\pm, \mathcal{S}_*)$ be subsets of zero, non-zero and unclassified indices of $x^\star$:

$$x^\star = \operatorname{argmin}_x \left\{ P(x) = \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 + \tfrac{\gamma}{2}\|x\|_2^2 \right\}$$

Solve an n dimensional problem

$$\rightsquigarrow$$

$$\begin{aligned} x^\star_{\mathcal{S}_*} &= \operatorname{argmin}_x \left\{ \tilde{P}(x) = \tfrac{1}{2}\|\tilde{y} - \tilde{A}x\|_2^2 + \lambda\|x\|_1 + \tfrac{\gamma}{2}\|x\|_M^2 \right\} \\ x^\star_{\mathcal{S}_\pm} &= Bx^\star_{\mathcal{S}_*} + b \\ x^\star_{\mathcal{S}_0} &= 0 \end{aligned}$$

- Compute $\tilde{y}$, $\tilde{A}$, M, B and b (linear algebra operations)
- Solve an $n - |\mathcal{S}_0| - |\mathcal{S}_\pm|$ dimensional problem

Dynamic Screen & Relax principle

Algorithm 1: "Screen & Relax" solving procedure

Input: $\mathsf{x}^{(0)}$, A, y, λ, γ

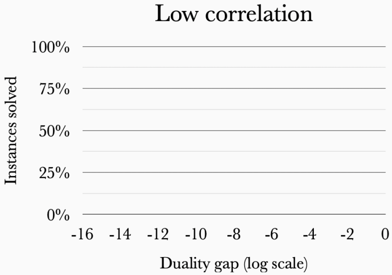
```
(S_0, S_±, S_*) ← (∅, ∅, {1, ..., n})
while convergence criterion is not met do
    Update the current iterate
    Compute a new safe sphere
    Update (S_0, S_±, S_*) with screening and relaxing tests
    Update the problem data
    if S_* = ∅ then
        The solution is available in closed form
    end
end
```

Some numerical results

Some numerical results

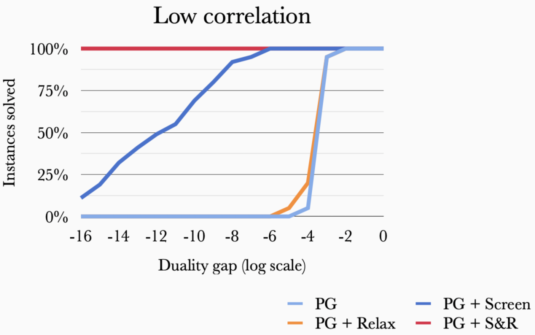
Setup : Percentage of instances solved up to a given accuracy for a fixed FLOPs budget.

Setup : Percentage of instances solved up to a given accuracy for a fixed FLOPs budget.



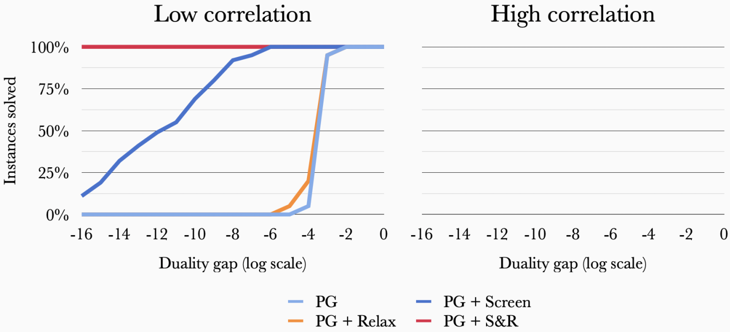
Some numerical results

Setup : Percentage of instances solved up to a given accuracy for a fixed FLOPs budget.



Some numerical results

Setup : Percentage of instances solved up to a given accuracy for a fixed FLOPs budget.



Some numerical results

Setup : Percentage of instances solved up to a given accuracy for a fixed FLOPs budget.