Screen & Relax : Accelerating the resolution of the Elastic-Net

Paper ID : 2487

Théo Guyard*, Cédric Herzet†, Clément Elvira‡

*Applied Mathematics Department, INSA Rennes, France | †SIMSMART team, INRIA Rennes-Bretagne Atlantique, France | ‡SCEE team, CentraleSupelec Rennes, France

Objectives

Accelerate the resolution of the Elastic-Net :

- Identification of *zeros* in the optimizer
- Identification of *non-zeros* in the optimizer
- Reduction of the problem dimension
- Reduction of the complexity burden

Problem of interest

- Sparse decomposition aims at finding some approximation of a vector $\mathbf{y}$ as the linear combination of a few columns of a dictionary $\mathbf{A}$. The Elastic-Net is one way to achieve this :

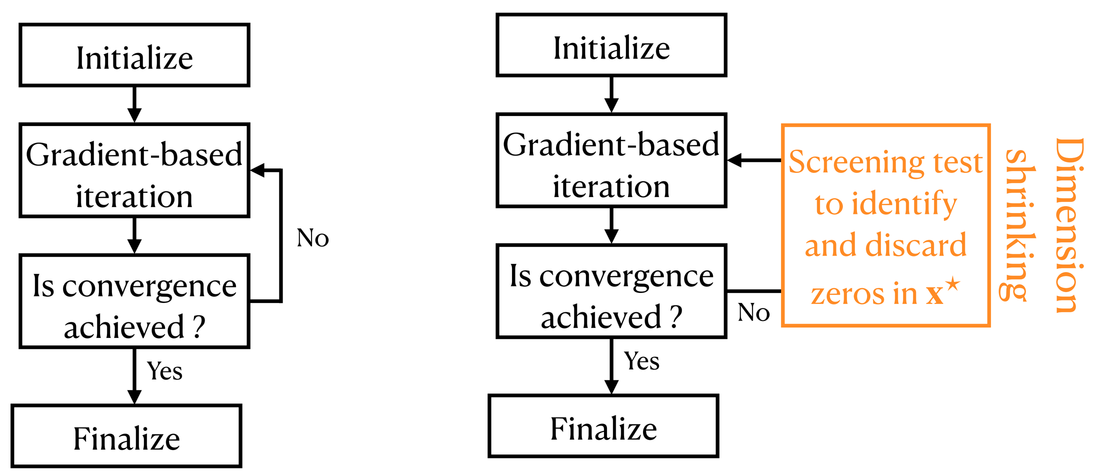
$\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2$	$\|\mathbf{x}\|_1$	$\|\mathbf{x}\|_2^2$
Ensures data fidelity	Promotes sparsity	Promotes a grouping effect

$$\mathbf{x}^\star = \arg\min \left\{ \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 + \tfrac{\gamma}{2}\|\mathbf{x}\|_2^2 \right\}$$

Elastic-Net problem

- The Elastic-Net is a convex problem so gradient-based methods are particularity well suited to solve it :

Solution methods backbone | Acceleration exploiting sparsity



- Our initial idea :
  - Why not identifying *non-zeros* in $\mathbf{x}^\star$?
  - We could break the non-differentiability of the ℓ_1-norm at zero ...
  - This potentially allows to shrink even more the problem dimension !

Let's play with duality !

Fenchel dual problem

$$\mathbf{u}^\star = \arg\max \left\{ \tfrac{1}{2}\|\mathbf{y}\|_2^2 - \tfrac{1}{2}\|\mathbf{y}-\mathbf{u}\|_2^2 - \tfrac{1}{2\gamma}\|[\,|\mathbf{A}^\top\mathbf{u}| - \lambda]_+\|_2^2 \right\}$$

A « different parametrization » of the Elastic-Net

Optimality conditions

$$\mathbf{u}^\star = \mathbf{y} - \mathbf{A}\mathbf{x}^\star$$
$$\gamma\mathbf{x}^\star = [\,|\mathbf{A}^\top\mathbf{u}^\star| - \lambda]_+$$

Links the primal and the dual optimizers

Screening tests

(already existing)

Goal : Identification of *zeros* in $\mathbf{x}^\star$.

Let $\mathcal{S}(\mathbf{u}, r)$ be a sphere containing $\mathbf{u}^\star$, then

$$\forall i, \quad |\mathbf{a}_i^\top\mathbf{u}| + r < \lambda \implies \mathbf{x}_i^\star = 0 \qquad (1)$$

Elements that have passed the screening test can be discarded safely from the problem, as well as the corresponding columns in $\mathbf{A}$.

Relaxing tests

(our contribution)

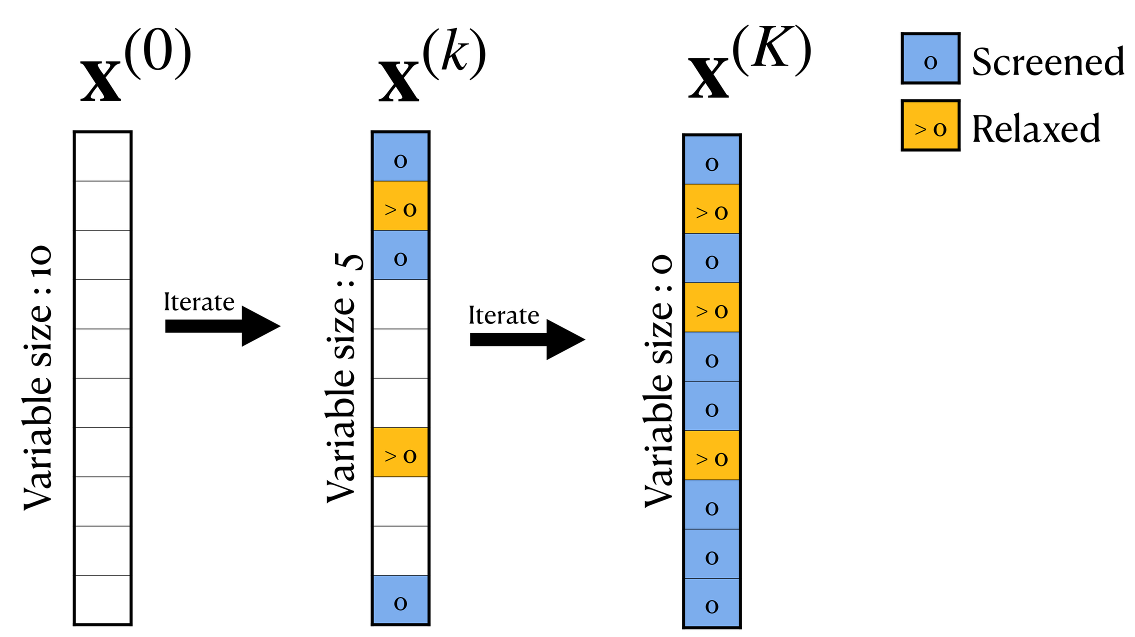
Goal : Identification of *non-zeros* in $\mathbf{x}^\star$.

Let $\mathcal{S}(\mathbf{u}, r)$ be a sphere containing $\mathbf{u}^\star$, then

$$\forall i, \quad |\mathbf{a}_i^\top\mathbf{u}| - r > \lambda \implies \mathbf{x}_i^\star \neq 0 \qquad (2)$$

Elements that have passed the relaxing test can be expressed as a linear combination of all the other elements of $\mathbf{x}$ in the problem.

Screen & Relax strategy



- Benefits of the Screen & Relax strategy :
  - Dimension shrinking
  - Iteration complexity reduction
  - Conditioning improvement
  - Closed form solution when all elements have been either screened or relaxed

Pseudo-code

Algorithm 1: Iterative method for the Elastic-Net problem enhanced with a "Screen & Relax" strategy.

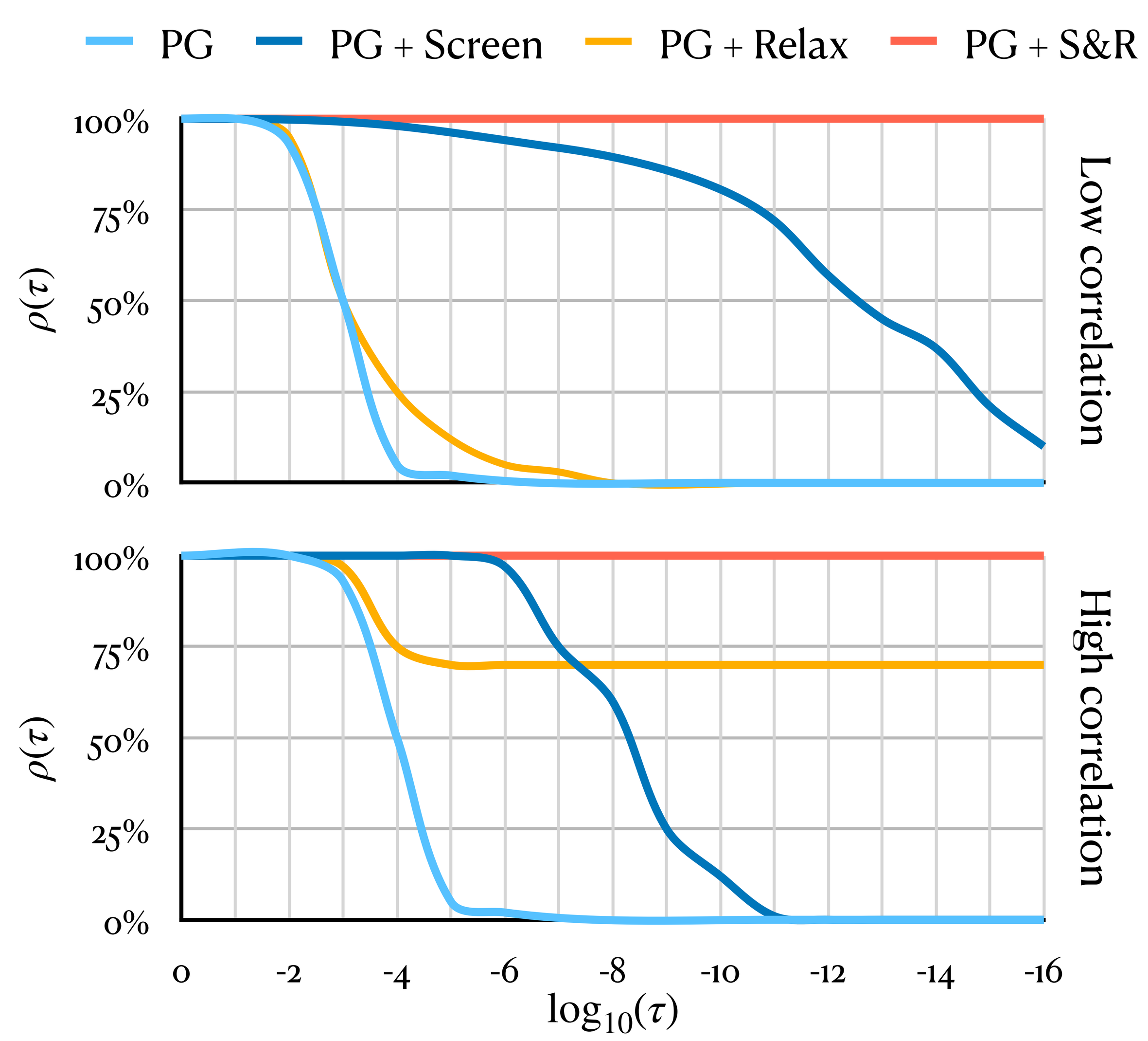
Input: Problem data $(\mathbf{A}, \mathbf{y}, \lambda, \gamma)$

while *convergence is not met* **do**
- Update the current iterate $\mathbf{x}^{(t)}$
- Construct a new safe sphere $\mathcal{S}(\mathbf{u}^{(t)}, r^{(t)})$
- Perform the screening and relaxing tests
- If new elements have been *screened*, discard them from the problem
- If new elements have been *relaxed*, express them as a function of the others. This requires a *modification of the problem data* that can be done efficiently using *rank-one rules*.

end

Numerical results

- Data generation :
  1) Generate a random matrix $\mathbf{A} \in \mathbf{R}^{m\times n}$ with either a low or an high correlation between the columns
  2) Generate a sparse vector $\mathbf{x}^\dagger \in \mathbf{R}^n$
  3) Set $\mathbf{y} = \mathbf{A}\mathbf{x}^\dagger$ + noise with 10dB SNR
  4) Calibrate λ and γ statistically
- Concurrent methods :
  - Proximal-Gradient (PG) algorithm
  - PG algorithm with *screening*
  - PG algorithm with *relaxing*
  - PG algorithm with *screening and relaxing*



- Observations :
  - Screening only : efficient when correlation between the columns is *low*.
  - Relaxing only : efficient when correlation between the columns is *high*.
  - Screening and Relaxing : allows convergence up to machine precision in all the setups tested.

Take home message : Both the identification of *zero* and *non-zero* elements in the Elastic-Net solution allows to enhance its resolution.