Automated and Unbiased Coefficients Clustering

Anne Gagneux[1], Mathurin Massias[1], Emmanuel Soubies[2]
[1]OCKHAM, Inria Lyon, LIP, ENS Lyon [2]IRIT, CNRS, Toulouse,

Motivation: introduce the **unbiased** counterparts of the Sorted L1 penalty (**Slope**, [1]) for **sparse structured generalized linear models.**

Composite penalized problems

$$\arg\min_{\mathbf{x}\in\mathbb{R}^p} f(\mathbf{x}) + \Psi(\mathbf{x})$$

datafit ↗ ↖ penalty

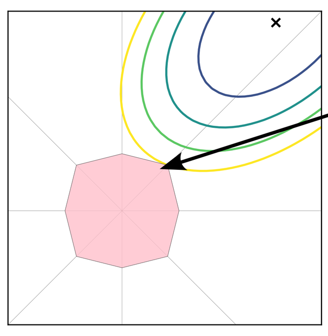
Q: which **penalty** should be used to automatically group features ?

Proximal algorithms

$$\mathbf{x}^{k+1} = \mathrm{prox}_{\eta\Psi}(\dots)$$

Sorted Nonconvex Penalties [2]

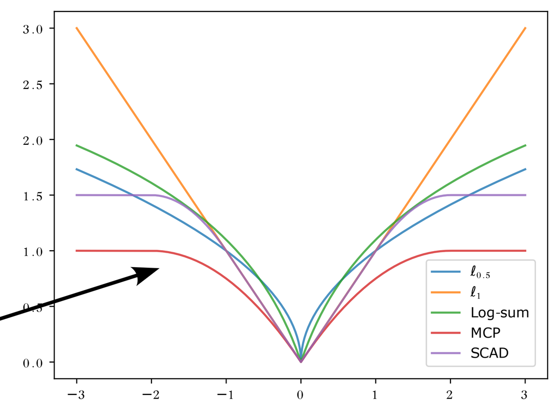
Sorted L1 (Slope)



Sorting promotes pairwise equality of coefficients

⊕

Nonconvexity penalizes less large coefficients



$$\Psi(\mathbf{x}) = \sum_i \psi(x_{(i)}; \lambda_i)$$

$|x_{(1)}| \geq \dots \geq |x_{(p)}|$ $\lambda_1 \geq \dots \geq \lambda_p \geq 0$

→ $\psi = \lambda|\cdot|$ → Sorted L1 (Slope, [1])

→ ψ nonconvex → Our work

Using **a non-smooth penalty** requires knowing its **proximal operator**. Computing the **prox of a sorted penalty** comes down to an **isotonic minimization** problem.

$$\mathrm{prox}_{\eta\Psi}(\mathbf{y}) = \arg\min_{\mathbf{x}:|x_1|\geq\dots\geq|x_p|} \frac{1}{2\eta}\|\mathbf{x}-\mathbf{y}\|^2 + \sum_i \psi(x_i; \lambda_i)$$

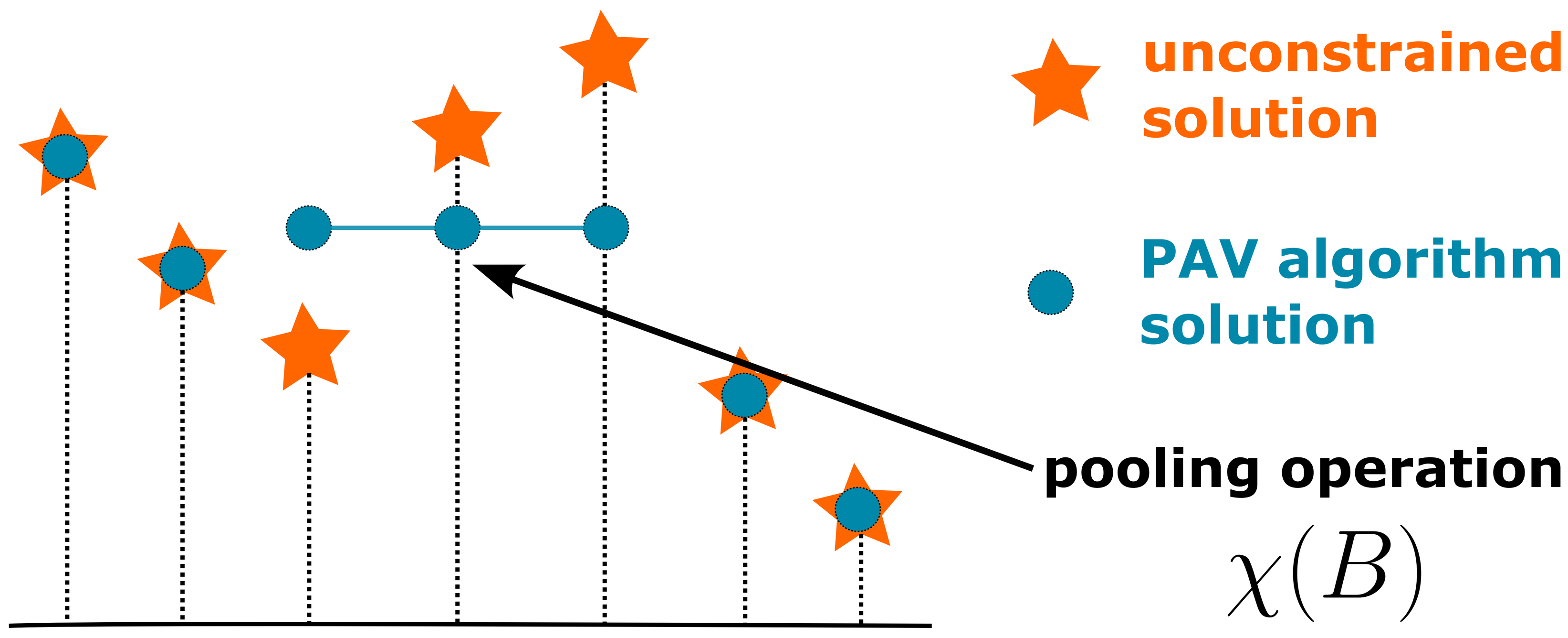
Sorted *weakly-convex* penalties

Result:

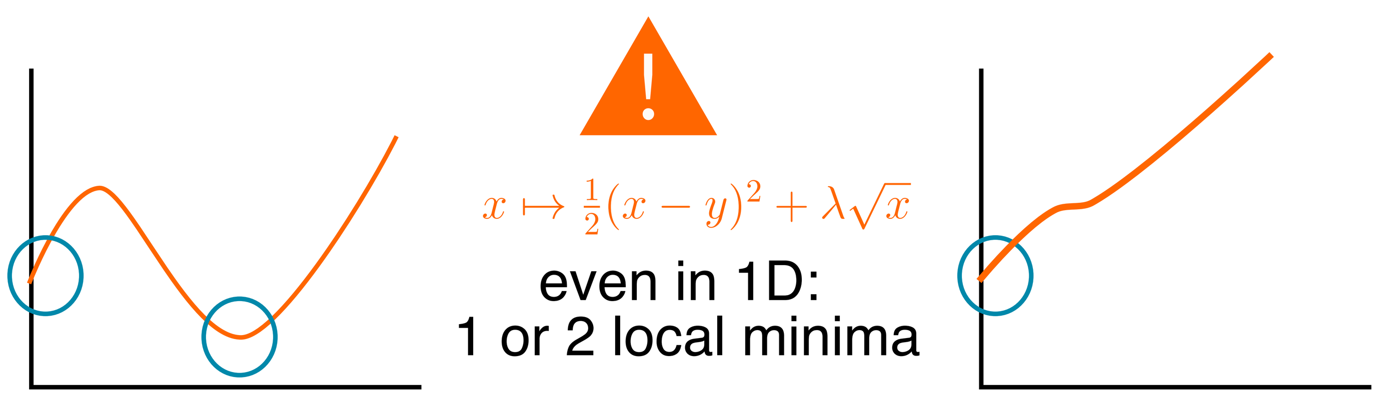
The **PAV algorithm solves** the isotonic minimization problem: it computes a partition of the vector and performs a scalar proximal operation block-wise.

$$\chi(B) = \mathrm{prox}_{\frac{\eta}{|B|}\sum_{i\in B}\psi(\cdot;\lambda_i)}(\bar{y}_B)$$

Pool Adjacent Violators algorithm [3]



Sorted L_q



Result 1:

Necessary and sufficient conditions to be ***local*** minimizers of the **prox** problem.

Result 2:

Necessary conditions to be ***global*** minimizers of the **prox** problem.

In practice:

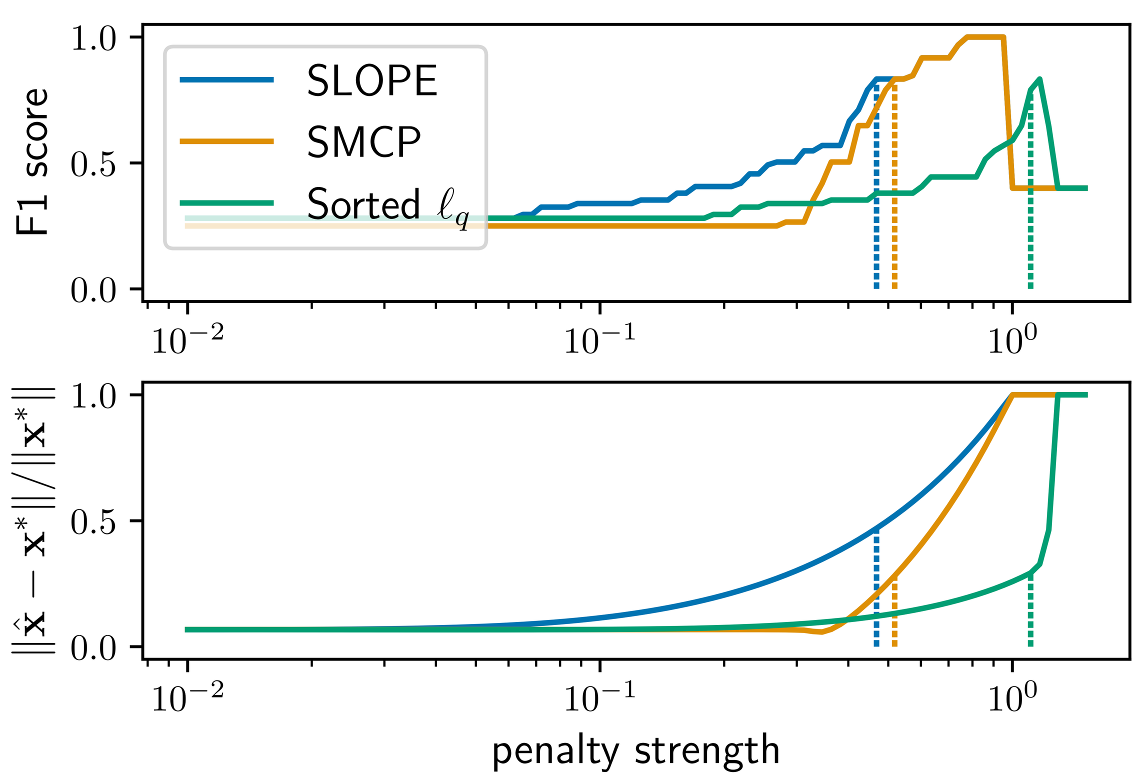
A **Decomposed PAV** algorithm:

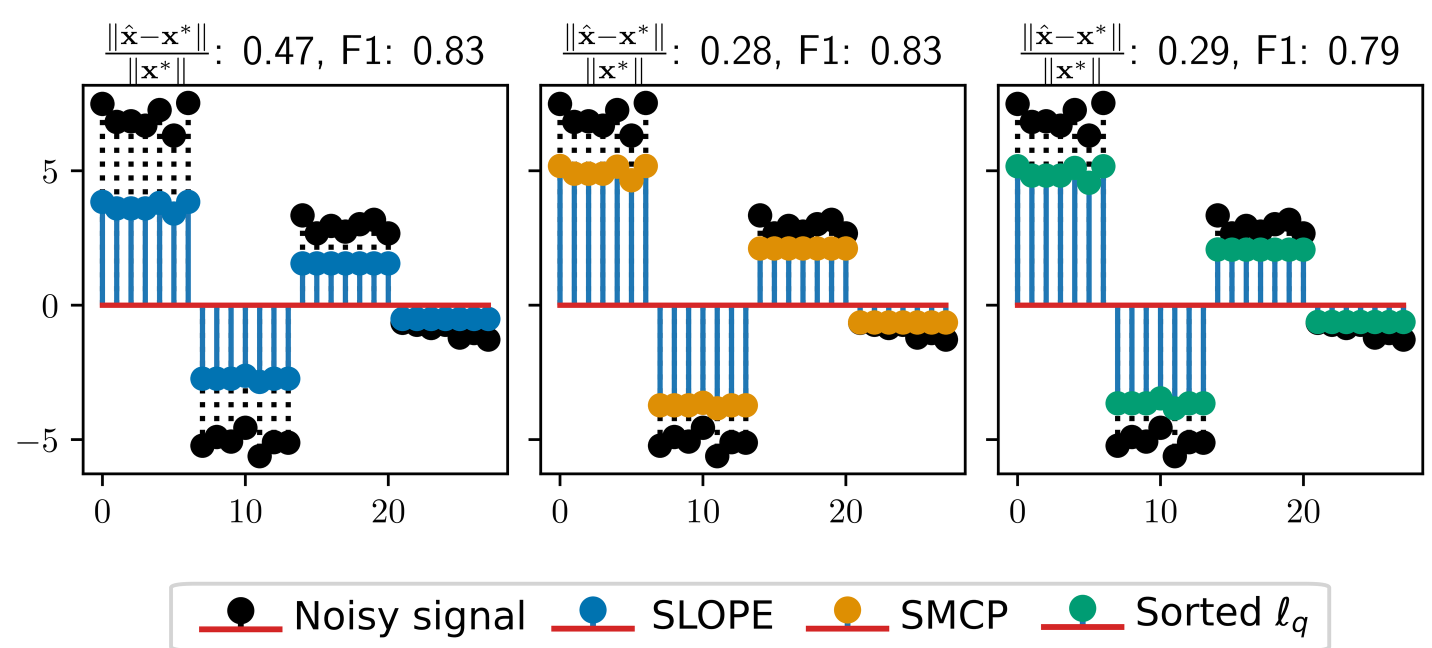
For $k \in \{1, \dots, p\}$
Apply PAV algorithm on problem of size k : $\mathbf{u}^k$
Complete $\mathbf{u}^k$ with 0 : $\mathbf{z}^k = (\mathbf{u}^k, 0, \dots, 0)$
Return $\arg\min_{\mathbf{z}\in\{\mathbf{z}^1,\dots,\mathbf{z}^p\}} P(\mathbf{z})$
where P is the prox objective

→ Converges to a local minimizer which satisfies necessary conditions on global minimizers.

Experiments

Signal denoising: noisy observation of a clustered signal





[1] Bogdan, M., Van Den Berg, E., Sabatti, C., Su, W., & Candès, E. J. (2015). SLOPE—adaptive variable selection via convex optimization.
[2] Feng, L., & Zhang, C. H. (2019). Sorted concave penalized regression..
[3] Best, M. J., Chakravarti, N., & Ubhaya, V. A. (2000). Minimizing separable convex functions subject to simple chain constraints.