M2 internship offer: Improving sparse penalties with non convexity and coefficient clustering

Supervision: Emmanuel Soubies, IRIT, CNRS (emmanuel.soubies@irit.fr)
Mathurin Massias, Inria Lyon (mathurin.massias@inria.fr)

Location: Depending on the candidate's interest, this internship can take place either at

- INRIA Lyon (located at ENS de Lyon),
- IRIT laboratory in Toulouse.

Moreover, work visits from one place to another will be possible.

Profile: We are looking for a highly motivated student with a background in mathematics (optimization, probability an statistics, geometry) and/or electrical engineering (signal/image processing). Strong abilities in computer sciences are required. Experience with Python is a plus. If the candidate is successful, this internship may be pursued by a PhD.

Salary: The intern will be granted the usual stipend of ∼600 euros/month.

Keywords: Optimization, non-convexity, high dimension, sparsity, minimax concave penalty, sorted L1 norm.

1 Context

Sparse optimization problems have become ubiquitous in high dimensional regression and classification problems. A very common formulation is given by

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2}\|y - X\beta\|^2 + \lambda P(\beta)$$

where $X \in \mathbb{R}^{m \times p}$, $y \in \mathbb{R}^m$, and P is a penalization term that promotes sparse solutions (i.e., with many zeros). For instance, a popular penalization is the ℓ_1-norm ($P = \| \cdot \|_1$) (Tibshirani, 1996), which can be seen as the best compromise between sparsity and convexity.

On the other hand, non-convex penalties such as ℓ_p ($0 < p < 1$), SCAD or MCP have been shown to mitigate the well-known amplitude bias of their convex counterparts. Amongst these, the Minimax Concave Penalty (MCP) (Zhang, 2010) optimally satisfies the necessary conditions for unbiased recovery of sparse signals (Soubies et al., 2017). It is defined by

$$\mathrm{MCP}(x; \lambda, \gamma) = \begin{cases} \lambda|x| - x^2/2\gamma \ , & \text{if } |x| \le \lambda\gamma \ , \\ \lambda^2\gamma/2 & \text{otherwise} \ . \end{cases} \tag{1}$$

where $\lambda > 0$ and $\gamma \geq 1$ are two parameters controlling the penalization strength and the non-convexity, respectively.

2 Subject

The supervisors have begun investigating various open questions in this field. The intern will be able to explore one (or both) of the following research directions, depending on his/her interests.

Improving the minimization of MCP Computing the solution of MCP regularized problems is a challenge, due to the non convexity of the latter, that leads to existence of local minima that are not global.

The first topic of this internship will be to improve the robustness of numerical solvers to spurious local minimizers. To that end, the intern will build upon *SparseNet* (Mazumder et al., 2011) that combines graduated non-convexity (Mobahi and Fisher, 2015) together with warm-start strategies so as to compute the full MCP regularization path (i.e., all possibles solution from a large λ to a small one). In particular, the goal is to exploit known properties of MCP (Soubies et al., 2020), as well as fast solvers (Bertrand et al., 2022) to improve and accelerate the search strategy.

SLOPE MCP The Sorted L1 Penalization Estimator (SLOPE) is a convex generalization of the Lasso (Bogdan et al., 2013). It is defined as the solution of:

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2}\|y - X\beta\|^2 + \sum_{j=1}^{p} \lambda_j |\beta_{(j)}| \tag{2}$$

where $\lambda_1 \geq \cdots \geq \lambda_p$ and $(\cdot)$ is a permutation that reorders β by decreasing magnitude: $|\beta_{(1)}| \geq \cdots \geq |\beta_{(p)}|$. It enjoys better statistical properties than the Lasso (Bogdan et al., 2015), and automatically clusters coefficients associated to correlated features (Figueiredo and Nowak, 2016).

Similar to (3), a Sorted MCP (S-MCP) could be defined as:

$$\min_{\beta \in \mathbb{R}^d} \frac{1}{2}\|y - X\beta\|^2 + \sum_{j=1}^{d} \mathrm{MCP}(|\beta_{(j)}|; \gamma, \lambda_j) \ . \tag{3}$$

We expect such a penalty to leverage both the advantages of SLOPE (clusturing of coefficients) and MCP (sparse solution without bias).

The second topic of this internship will be to study the numerical feasability of solvers for S-MCP regression problems. SLOPE's practical success relies on the efficient computation of its proximal operator (Zeng and Figueiredo, 2014) that allows the use of fast first order algorithms (Parikh and Boyd, 2014). As a start, the inter will thus work on the computation of the S-MCP proximal operator so as to efficiently deploy first order algorithms. Then, we shall also consider the use of iteratively reweighed algorithms.

References

R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 58(1):267–288, 1996.

C.-H. Zhang. Nearly unbiased variable selection under minimax concave penalty. 38(2):894–942, 2010.

E. Soubies, L. Blanc-Féraud, and G. Aubert. A unified view of exact continuous penalties for ℓ_2-ℓ_0 minimization. 27(3):2034–2060, 2017.

Rahul Mazumder, Jerome H Friedman, and Trevor Hastie. Sparsenet: Coordinate descent with nonconvex penalties. *Journal of the American Statistical Association*, 106(495):1125–1138, 2011.

Hossein Mobahi and John W Fisher. On the link between gaussian homotopy continuation and convex envelopes. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 43–56. Springer, 2015.

Emmanuel Soubies, Laure Blanc-Féraud, and Gilles Aubert. New insights on the optimality conditions of the l2 - l0 minimization problem. *Journal of Mathematical Imaging and Vision*, 62(6):808–824, 2020.

Quentin Bertrand, Quentin Klopfenstein, Pierre-Antoine Bannier, Gauthier Gidel, and Mathurin Massias. Beyond l1: Faster and better sparse models with skglm. *arXiv preprint arXiv:2204.07826*, 2022.

Malgorzata Bogdan, Ewout van den Berg, Weijie Su, and Emmanuel Candes. Statistical estimation and testing via the sorted l1 norm. *arXiv preprint arXiv:1310.1969*, 2013.

Małgorzata Bogdan, Ewout van den Berg, Chiara Sabatti, Weijie Su, and Emmanuel Candès. SLOPE - adaptive variable selection via convex optimization. 9(3):1103–1140, 2015.

Mario Figueiredo and Robert Nowak. Ordered weighted L1 regularized regression with strongly correlated covariates: Theoretical aspects. In *AISTATS*, pages 930–938, 2016.

Xiangrong Zeng and Mario Figueiredo. The ordered weighted ℓ_1 norm: Atomic formulation, projections, and algorithms. *arXiv preprint arXiv:1409.4271*, 2014.

N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Machine Learning*, 1(3): 127–239, 2014.