Internship proposal:
"AMP for MCP: Can early false detection be avoided with Minimax Concave Penalty?"

AGENCE NATIONALE DE LA RECHERCHE ANR Chair of research and teaching in artificial intelligence

IMAG INSTITUT MONTPELLIERAIN ALEXANDER GROTHENDIECK

Keywords: *Optimization, coding, approximate message passing, sparse regression, MCP*

Sparse regression and high dimensional statistics

High dimensional statistics has seen a tremendous amount of work since the 90's, mostly govern by the genomics applications. Following the Lasso [12, 4] introduction, sparsity enforcing penalties have played a major role, especially the convex one (the ℓ_1 case). Yet, the Lasso suffers from early detections on the path: even for large regularization parameters, wrong detections might already pop-up [11]. Technically, the results by these authors were proved using Approximate Message Passing (AMP) theory [5].

Internship description

For this project, the goal of the intern will be to adapt the AMP methodology for the MCP penalty [14].

Let us remind the definition of the Minimax Concave Penalty (MCP) estimator. First, for some parameter $\gamma > 1$ and $\lambda \geq 0$, and for any $t \in \mathbb{R}$ we define the 1D penalty as follows:

$$p_{\lambda,\gamma}^{\text{MCP}}(t) = \begin{cases} \lambda|t| - \frac{t^2}{2\gamma}, & \text{if } |t| \leq \gamma\lambda, \\ \frac{1}{2}\gamma\lambda^2, & \text{if } |t| > \gamma\lambda \enspace . \end{cases} \tag{1}$$

The proximity operator[1] of $p_{\lambda,\gamma}$ for parameters $\lambda > 0$ and $\gamma > 1$ is defined as follow (see [3, Sec. 2.1]):

$$\text{prox}_{\lambda,\gamma}^{\text{MCP}}(t) = \begin{cases} \frac{\text{ST}(t,\lambda)}{1-\frac{1}{\gamma}} & \text{if } |t| \leq \gamma\lambda \\ t & \text{if } |t| > \gamma\lambda \enspace , \end{cases} \tag{2}$$

where $\text{ST}(t,\lambda) = \text{sign}(t) \cdot (|t| - \lambda)_+$ for any $t \in \mathbb{R}$ and $\lambda \geq 0$ (resp. $\text{HT}(t,\lambda) = t \cdot \mathbb{1}_{\{|t|>\lambda\}}$), corresponds to the limiting case when $\gamma \to \infty$ (resp. when $\gamma \to 1$). The proximal operator defined in Eq.(2) is also known as the Firm Shrinkage [7]. For $\lambda \in \mathbb{R}$ and $\gamma > 1$ the MCP estimator is then defined by:

$$\hat{\beta}^{(\lambda,\gamma)}(y) \triangleq \mathop{\arg\min}_{\beta \in \mathbb{R}^p} \frac{1}{2n}\|y - X\beta\|_2^2 + \sum_{j=1}^{p} p_{\lambda,\gamma}^{\text{MCP}}(|\beta_j|) \enspace , \tag{3}$$

where $X \in \mathbb{R}^{n \times p}$ is the design matrix and $y \in \mathbb{R}^n$ the observed signal.

Illustration of the penalty and proximal operator are provided in Fig.1 for some parameters, including limiting behavior.

[1] For a function p this operator is defined by $\text{prox}(x) = \arg\min_{x'} p(x') + \|x - x'\|^2 / 2$.

The aim of the internship is to show that using MCP will reduce the (possibly many) false positive detections due to the contraction bias the Lasso is suffering from, see [11]. The choice of MCP is due to appealing optimization properties proved for this penalty (in terms of local minima shared with ℓ_0 penalty, unfortunately leading to NP hard problems), see in particular [10].

The theory relies on recent developments of techniques for the analyses of high-dimensional statistical learning algorithms based on the Mean Field Asymptotics, like *e.g.,* the asymptotic theory for Approximate Message Passing (AMP) Algorithms (see [9] , [6], [2]). Some modifications/extensions will be required to deal with the non-convex MCP penalty.

In parallel to the theoretical study, a thorough numerical investigation is expected. Hence, the intern would provide an adaptation for the MCP case of the experiments performed in the paper [11] but for this non-convex regularization. Variations due to algorithmic difficulties (*e.g.,* local minima) might be of interest and connected to the work of another intern on computational efficiency. The list of possible solvers includes for instance:

- Coordinate descent [13, 3]
- Proximal gradient descent [1]
- Difference of convex (DC) programming [8]
- etc.

Skills required

- `Python`
- `Git`
- `R` (not mandatory, but could come handy)

Supervision Team

- Joseph Salmon: joseph.salmon@umontpellier.fr
- Małgorzata Bogdan: Malgorzata.Bogdan@uwr.edu.pl
- Nicolas Meyer: nicolas.meyer@umontpellier.fr

Salary

Gross monthly salary: approx 550 Euros.
This work will be funded by the ANR CaMeLOt ANR-20-CHIA-0001-01.

Duration

The internship could last from 4 to 6 months.

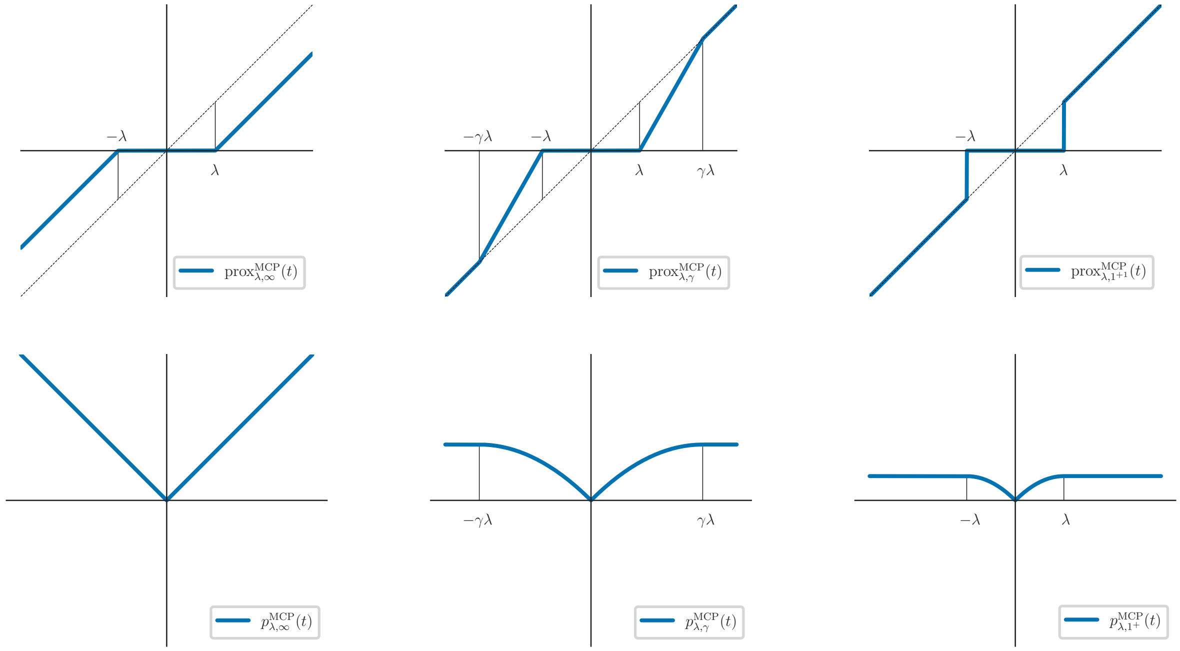
Location

The internship will be located in Montpellier (Univ. Montpellier), inside the mathematics department (IMAG).

References

[1] A. Beck and M. Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". *SIAM J. Imaging Sci.* 2.1 (2009), pp. 183–202.

[2] P. C. Bellec, Y. Shen, and C.-H. Zhang. "Asymptotic normality of robust M-estimators



(a) $\gamma \to +\infty$: recovers $\mathrm{ST}(\cdot, \lambda)$ (b) $\gamma = 2.3$ (c) $\gamma \to 1^+$: recovers $\mathrm{HT}(\cdot, \lambda)$

Figure 1: Penalty (bottom) and associated proximal operator (top) for a fixed λ, with $\gamma \to +\infty$ in (a), for $\gamma = 2.3$ in (b) and $\gamma \to 1^+$ in (c).

with convex penalty”. *arXiv preprint arXiv:2107.03826* (2021).

[3] P. Breheny and J. Huang. “Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection”. *Ann. Appl. Stat.* 5.1 (2011), p. 232.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders. “Atomic decomposition by basis pursuit”. *SIAM J. Sci. Comput.* 20.1 (1998), pp. 33–61.

[5] D. L. Donoho, A. Maleki, and A. Montanari. “Message-passing algorithms for compressed sensing”. *Proceedings of the National Academy of Sciences* 106.45 (2009), pp. 18914–18919.

[6] O. Y. Feng, R. Venkataramanan, C. Rush, and R. J. Samworth. “A unifying tutorial on Approximate Message Passing”. *arXiv preprint arXiv:2105.02180* (2021).

[7] H.-Y. Gao and A. G. Bruce. “WaveShrink with firm shrinkage”. *Statist. Sinica* (1997), pp. 855–874.

[8] G. Gasso, A. Rakotomamonjy, and S. Canu. “Recovering sparse signals with non-convex penalties and DC programming”. *IEEE Trans. Signal Process.* 57.12 (2009), pp. 4686–4698.

[9] A. Montanari. *Mean field asymptotics in high-dimensional statistics: A few references*. 2020.

[10] E. Soubies, L. Blanc-Féraud, and G. Aubert. “A Unified View of Exact Continuous Penalties for ℓ_2-ℓ_0 Minimization”. *SIAM J. Optim.* 27.3 (2017), pp. 2034–2060.

[11] W. Su, M. Bogdan, and E. J. Candès. “False discoveries occur early on the lasso path”. *Ann. Statist.* (2017), pp. 2133–2150.

[12] R. Tibshirani. “Regression Shrinkage and Selection via the Lasso”. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 58.1 (1996), pp. 267–288.

[13] T. T. Wu and K. Lange. “Coordinate descent algorithms for lasso penalized regression”. *Ann. Appl. Stat.* (2008), pp. 224–244.

[14] C.-H. Zhang. “Nearly unbiased variable selection under minimax concave penalty”. *Ann. Statist.* 38.2 (2010), pp. 894–942.