Ph.D. proposal:
"Avoiding false detection in high dimensional regression thanks to non-convex penalties"

"Éviter les fausses détections en régression de grande dimension grâce à des pénalités non convexes"

Keywords:
Optimisation, logiciel libre, fausses détections, passage de message approximatif (AMP), régression parsimonieuse, non-convexité.

1 Présentation scientique du projet

Présentation scientifique du projet de recherche

Les statistiques en grande dimension ont fait l'objet d'un travail considérable depuis leur avènement dans les années 90, principalement régies par les applications en génomique. Après l'introduction du Lasso [13, 5], les pénalités parcimonieuses ont joué un rôle majeur, en particulier la pénalité convexe (le cas ℓ_1) pour découvrir par exemple l'activité de certains gènes et confirmer des phénomènes biologiques (un autre exemple étant la localisation d'activité neuronales en neuro-imagerie [1]).

Pourtant, le Lasso, malgré des performances empiriques prometteuses, souffre de détections précoces sur le chemin de régularisation: même pour de grands paramètres de régularisation, des détections erronées peuvent déjà apparaître [12], et ce même si le nombre d'observation est grand ou que le signal est très puissant. Cela peut alors être dramatique, en créant de fausses détections pour le praticien parmi certaines variables explicatives.

Suite à des travaux récents décrivant les propriétés attrayantes de certaines alternatives non convexes, notamment MCP (Minimax Concave Penalty) [15], également connu sous le nom de CEL0 [11], nous analyserons ce type d'estimateur pour montrer leur supériorité sur les contreparties convexes.

Description du travail

Techniquement, les résultats développés dans [12] ont été prouvés à l'aide de la théorie AMP (Approximate Message Passing) [6]. Pour ce projet, l'objectif du candidat sera d'adapter la méthodologie AMP et d'autres théories alternative, comme celle "min-max" de Gordon à la pénalité MCP [15].

Rappelons la définition de l'estimateur Minimax Concave Penalty (MCP). Tout d'abord, pour un paramètre $\gamma > 1$ et $\lambda \geq 0$, et pour tout $t \in \mathbb{R}$, nous définissons la pénalité 1D comme suit :

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

L'opérateur proximal[1] de $p_{\lambda,\gamma}$ pour des paramètres $\lambda > 0$ et $\gamma > 1$ est défini comme suit (voir [4, Sec. 2.1]) :

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

où $\text{ST}(t, \lambda) = \text{sign}(t) \cdot (|t| - \lambda)_+$ pour tout $t \in \mathbb{R}$ et $\lambda \geq 0$ (resp. $\text{HT}(t, \lambda) = t \cdot \mathbb{1}_{\{|t|>\lambda\}}$), correspond au cas limite lorsque $\gamma \to \infty$ (resp. lorsque $\gamma \to 1$). L'opérateur proximal défini dans l'équation (2) est également connu sous le nom de "rétrécissement ferme" [8]. Pour $\lambda \in \mathbb{R}$ et $\gamma > 1$, l'estimateur MCP est alors défini par :

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

où $X \in \mathbb{R}^{n\times p}$ est la matrice de conception et $y \in \mathbb{R}^n$ le signal observé.

Des illustrations de la pénalité et de l'opérateur proximal sont fournies dans la Figure 1 pour certains paramètres, y compris les comportements limites.

L'objectif du doctorat est de montrer que l'utilisation de MCP réduira les (éventuellement nombreux) faux positifs dus au biais de contraction dont souffre le Lasso, voir [12]. Le choix de MCP est dû aux propriétés d'optimisation attrayantes prouvées pour cette pénalité (en termes de minima locaux partagés avec la pénalité ℓ_0, conduisant hélas à des problèmes NP dur), voir en particulier [11].

La théorie s'appuie sur des développements récents de techniques d'analyse d'algorithmes d'apprentissage statistique de haute dimension basées sur l'asymptotique en champ moyen, comme par exemple la théorie asymptotique des algorithmes de passage de messages approximatifs (AMP) (voir [10], [7], [3]). Certaines modifications/extensions seront nécessaires pour traiter la pénalité MCP non convexe.

Compétences requises

- Maîtrise de l'anglais
- Formation en statistique et apprentissage automatique
- Multidisciplinaire (interaction avec les praticiens du côté polonais)
- Optimisation
- `Python`
- `Git`
- `R` (pas obligatoire, mais pourrait être utile)

Encadrants

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

[1]Pour une fonction p cet opérateur est défini par $\text{prox}(x) = \arg\min_{x'} p(x') + \|x - x'\|^2/2$.

2 Liens sujet de thèse et adéquation avec la politique du laboratoire

L'unité de recherche d'accueil sera l'IMAG (Institut Montpelliérain Alexander Grothendieck), au sein de l'Univ. Montpellier. Le sujet de thèse est au coeur de la politique de l'IMAG qui vise à développer l'apprentissage automatique (*machine learning*) et les problématiques d'optimisation numériques afférentes. De plus les retombées en génomiques pourraient être un plus pour asseoir la place centrale de l'IMAG au sein de la communauté de recherche de Montpellier. L'engagement de l'IMAG est fort depuis 2018 (année du recrutement de J. Salmon sur poste de PR) sur la thématique de l'IA. Cet effort a été couronné de succès notamment grâce à l'obtention de la chaire IA: ANR CaMeLOt ANR-20-CHIA-0001-01. En particulier les recherches permettront de renforcer les compétences en optimisation de l'équipe probabilités et statistiques (EPS) de l'IMAG.

3 Valorisation du projet

Parallèlement à l'étude théorique, une étude numérique approfondie est attendue. Le candidat devra donc fournir une adaptation pour le cas MCP des expériences réalisées dans l'article [12] mais pour cette régularisation non convexe. Les variations dues à des difficultés algorithmiques (*e.g.,* minima locaux) pourraient être intéressantes. La liste des solveurs possibles inclut par exemple :

- Descente par coordonnées [14, 4]
- Descente de gradient proximal [2]
- Programmation par différence convexes (DC) [9]
- etc.

Un objectif ambitieux de ce projet est de fournir une solide boîte à outils Python pour mettre en œuvre la nouvelle méthode algorithmique analysée durant le doctorat. Cela passe par un travail important sur la documentation, l'intégration continue, la dissémination, etc.

Enfin les retours applicatifs sur les aspects génomiques et en neuro-imagerie pourraient permettre d'améliorer les logiciels états de l'art dans ces deux domaines.

Salaire

Salaire mensuel brut : environ 1 975 euros.
Ce travail sera partiellement financé par l'ANR CaMeLOt ANR-20-CHIA-0001-01.

Durée

Le financement du programme de doctorat est d'une durée de 3 ans.

References

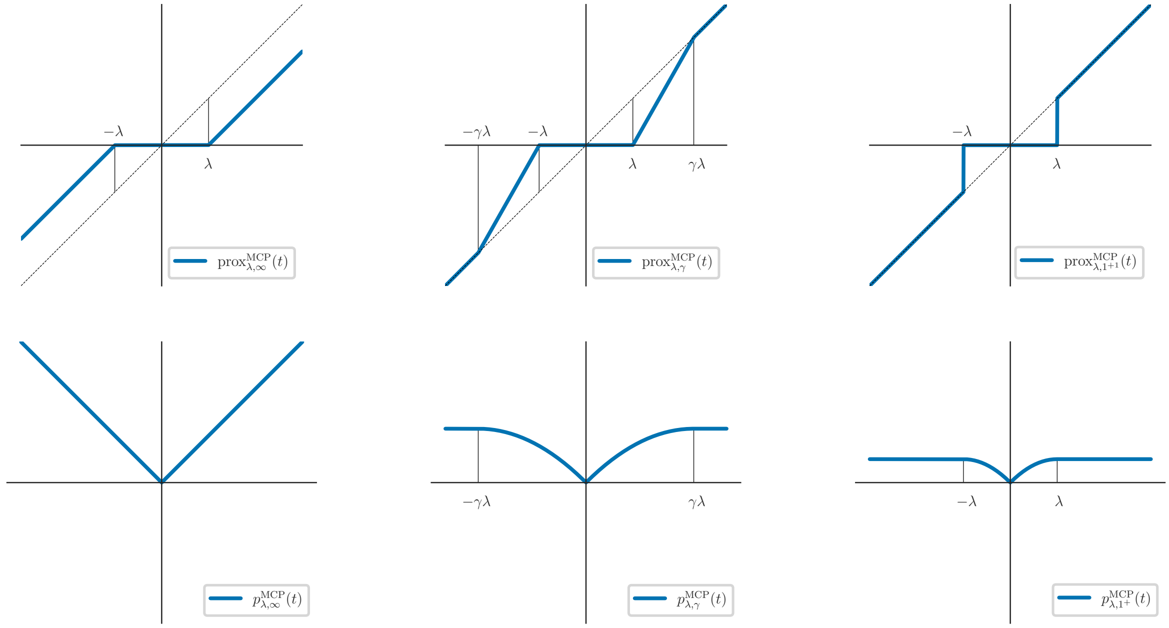
[1] P.-A. Bannier, Q. Bertrand, J. Salmon, and A. Gramfort. "Electromagnetic neural source imaging under sparsity constraints with SURE-based hyperparameter tuning". *Medical imaging meets NeurIPS (Workshop)*. 2021.

[2] 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.

[3] P. C. Bellec, Y. Shen, and C.-H. Zhang. "Asymptotic normality of robust M-estimators with convex penalty". *arXiv preprint arXiv:2107.03826* (2021).

[4] 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.

[5] 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.



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

Figure 1: Pénalité (en bas) et opérateur proximal associé (en haut) pour un λ fixe, avec $\gamma \to +\infty$ dans (a), pour $\gamma = 2.3$ dans (b) et $\gamma \to 1^+$ dans (c).

[6] 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.

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

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

[9] 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.

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

[11] 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.

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

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

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

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