

Elastic Net avec gestion des interactions et débiaisage

Florent Bascou*, Sophie Lèbre**
Joseph Salmon***

*IMAG, Univ. Montpellier, CNRS Montpellier, France
florent.bascou@umontpellier.fr,

**IMAG, Univ. Montpellier, CNRS Montpellier, France
Univ. Paul-Valéry-Montpellier 3, Montpellier, France
sophie.lebre@umontpellier.fr

***IMAG, Univ. Montpellier, CNRS Montpellier, France
joseph.salmon@umontpellier.fr

Résumé. Nous présentons quelques résultats statistiques pour un modèle de régression pénalisée et dé-biaisée pour ajuster, en grande dimension, un modèle linéaire parcimonieux avec interactions. L'analyse statistique s'intéresse à l'approche proposée par Bascou et al. (2020), notamment afin d'illustrer le fonctionnement du débiaisage. De plus, on montre qu'il permet de sélectionner moins de variables que le modèle sans débiaisage sur données simulées et réelles.

1 Introduction

Grâce à leur interprétabilité, les modèles linéaires sont populaires, néanmoins, le nombre de variables explicatives est souvent supérieur au nombre d'observations, de sorte qu'une régularisation est nécessaire. Des techniques de régularisation exploitant la norme ℓ_1 ont conduit à la création de nombreux estimateurs, dont les plus connus sont le Lasso (Tibshirani, 1996) et l'Elastic Net (Zou et Hastie, 2005). Pour traiter les interactions entre variables, la parcimonie est cruciale, car même limité aux d'ordre deux, le nombre de variables grandit déjà de manière quadratique. Nous estimons les coefficients à l'aide de l'estimateur Elastic Net qui permet de réduire le nombre de variables grâce à la pénalité ℓ_1 , tout en tenant compte des fortes corrélations entre variables grâce à la pénalité ℓ_2 (Tikhonov, 1943; Hoerl et Kennard, 1970). Dans Bascou et al. (2020), nous avons adapté un algorithme de descente de coordonnées (popularisé par `glmnet` (Friedman et al., 2007, 2010)) pour que la matrice d'interaction n'ait pas besoin d'être stockée en mémoire. Par ailleurs, sachant que l'Elastic Net contracte les grands coefficients vers zéro, dans Bascou et al. (2020) nous suggérons de calculer une version non dé-biaisée des coefficients (Deledalle et al., 2017) afin de proposer un algorithme approchant le LS Elastic Net (Elastic Net suivi des moindres carrés sur le support). Dans ce travail, nous comparons l'Elastic Net et sa version débiaisée sur données simulées avec différents scénarios d'hérédité pour montrer que le débiaisage améliore la sélection de variables. Nous terminons par une étude sur données réelles, pour confirmer ce comportement.

2 Elastic Net pour interactions et débiaisage

On note n le nombre d'observations, p le nombre de variables, et $q = p(p+1)/2$ (ou $q = p(p-1)/2$ si l'on exclut les effets quadratiques¹) le nombre de variables d'interactions d'ordre 2. On note $y \in \mathbb{R}^n$ le vecteur réponse, $X = [x_1, \dots, x_p] \in \mathbb{R}^{n \times p}$ la matrice de design, et $Z = [z_1, \dots, z_q] \in \mathbb{R}^{n \times q}$ celle des interactions. L'estimateur Elastic Net est alors :

$$(\hat{\beta}, \hat{\Theta}) \in \arg \min_{\substack{\beta \in \mathbb{R}^p \\ \Theta \in \mathbb{R}^q}} \frac{1}{2n} \|y - X\beta - Z\Theta\|^2 + \alpha_{1,1} \|\beta\|_1 + \alpha_{1,2} \|\Theta\|_1 + \alpha_{2,1} \frac{\|\beta\|^2}{2} + \alpha_{2,2} \frac{\|\Theta\|^2}{2} . \quad (1)$$

où $\alpha_{1,1} > 0, \alpha_{1,2} > 0, \alpha_{2,1} > 0$ et $\alpha_{2,2} > 0$ sont des paramètres à ajuster, par exemple par validation croisée. L'adaptation d'un algorithme de descente par coordonnées pour la minimisation du problème (1) permet d'éviter de stocker la matrice des interactions, et repose sur les mises à jours suivantes (voir détails dans Bascou et al. (2020)) :

Proposition 2.1. Soient $\hat{\beta}^k$ et $\hat{\Theta}^k$ les coefficients calculés par l'algorithme de descente par coordonnées à la k^e passe sur les données, et notons $r^k = y - X\hat{\beta}^k - Z\hat{\Theta}^k$ les résidus associés. Les mises à jour des j_0^e et $j_{j_0}^e$ coordonnées sont données par :

$$\hat{\beta}_{j_0}^{k+1} = \frac{1}{\|x_{j_0}\|^2 + n\alpha_{2,1}} \text{ST} \left(x_{j_0}^\top (r^k + \hat{\beta}_{j_0}^k x_{j_0}), n\alpha_{1,1} \right) , \quad (2)$$

$$\hat{\Theta}_{j_{j_0}}^{k+1} = \frac{1}{\|z_{j_{j_0}}\|^2 + n\alpha_{2,2}} \text{ST} \left(z_{j_{j_0}}^\top (r^k + \hat{\Theta}_{j_{j_0}}^k z_{j_{j_0}}), n\alpha_{1,2} \right) , \quad (3)$$

où ST représente la fonction définie pour tout $x \in \mathbb{R}$ par : $\text{ST}(x, \alpha) = (|x| - \alpha)_+ \text{sign}(x)$.

Preuve. Voir par exemple (Friedman et al., 2007). □

Les coefficients estimés par l'Elastic Net sont biaisés Salmon (2017), et l'on peut faire une étape de moindres carrés sur les coefficients non-nuls obtenus (Naive-LSEnet, Bascou et al. (2020)) pour contrecarrer cet effet. Néanmoins, cette approche est limitée car elle nécessite de nouveau la construction de la matrice des interactions sur le support obtenu, ce qui peut poser le même problème de mémoire qu'évoqué précédemment. Pour palier à ces limitations, nous avons adapté CLEAR (Covariant LEAst-square Refitting Deledalle et al. (2017)) - une méthode reposant sur une correction de premier ordre de l'estimateur initial (voir aussi Tukey (1977) ou Osher et al. (2005)) - à l'estimateur Elastic Net dans (Bascou et al., 2020).

Définition 2.1. L'estimateur CLEAR associé à un estimateur différentiable presque partout $\mathbb{R}^n \ni y \rightarrow \hat{\beta}(y) \in \mathbb{R}^p$ est, pour tout $y \in \mathbb{R}^n$, donné par :

$$\mathcal{R}_{\hat{\beta}}(y) := \hat{\beta}(y) + \rho J \cdot (y - X\hat{\beta}(y)) \text{ avec } \rho := \begin{cases} \frac{\langle XJ\delta | \delta \rangle}{\|XJ\delta\|^2} & , \text{ si } XJ\delta \neq 0, \\ 1 & , \text{ sinon,} \end{cases} \quad (4)$$

où $\delta = y - X\hat{\beta}(y)$ et $J = J_{\hat{\beta}}(y) \in \mathbb{R}^{p \times n}$ est la matrice Jacobienne de $\hat{\beta}$ au point y .

Nous calculons la matrice Jacobienne avec un schéma de différentiation automatique, menant à la proposition suivante et à l'Algorithme 2 de Bascou et al. (2020).

1. $q = 0$ si l'on ne tient pas compte des interactions

Proposition 2.2. *Supposons que les coefficients $\hat{\beta}^k$ et $\hat{\Theta}^k$ soient itérativement mis-à-jour par les Équations (2) et (3). On définit la Jacobienne de $\hat{\beta}^k$ (resp. $\hat{\Theta}^k$) appliquée aux résidus par $J_{\hat{\beta}^{k+1}} r^k$ (resp. $J_{\hat{\Theta}^{k+1}} r^k$) où e_j (resp. e_{jj}) est la base canonique de \mathbb{R}^p (resp. \mathbb{R}^q) :*

$$(J_{\hat{\beta}^{k+1}} r^{k+1})_j = \frac{x_j^\top \left((x_j e_j^\top - X) J_{\hat{\beta}^k} + (\text{Id}_n - Z J_{\hat{\Theta}^k}) \right) r^k}{\|x_j\|^2 + n\alpha_{2,1}} \mathbb{1}_{\{|x_j^\top (r^k + \hat{\beta}_j^k x_j)| \geq n\alpha_{1,1}\}} \quad (5)$$

$$(J_{\hat{\Theta}^{k+1}} r^{k+1})_{jj} = \frac{z_{jj}^\top \left((z_{jj} e_{jj}^\top - Z) J_{\hat{\Theta}^k} + (\text{Id}_n - X J_{\hat{\beta}^k}) \right) r^k}{\|z_{jj}\|^2 + n\alpha_{2,2}} \mathbb{1}_{\{|z_{jj}^\top (r^k + \hat{\Theta}_{jj}^k z_{jj})| \geq n\alpha_{1,2}\}} \quad (6)$$

et $\rho^{k+1} = \frac{\langle [X, Z][J_{\hat{\beta}^{k+1}} r^{k+1}, J_{\hat{\Theta}^{k+1}} r^{k+1}]^\top; r^{k+1} \rangle}{\|[X, Z][J_{\hat{\beta}^{k+1}} r^{k+1}, J_{\hat{\Theta}^{k+1}} r^{k+1}]^\top\|^2}$. En considérant l'Équation (1), l'estimateur CLEAR associé s'écrit :

$$\hat{\beta}^{k+1} = \hat{\beta}^{k+1} + \rho^{k+1} J_{\hat{\beta}^{k+1}} r^{k+1} \quad \text{et} \quad \hat{\Theta}^{k+1} = \hat{\Theta}^{k+1} + \rho^{k+1} J_{\hat{\Theta}^{k+1}} r^{k+1} \quad (7)$$

Preuve : Notons : $t_j = x_j^\top (r^k + \hat{\beta}_j^k x_j)$ et rappelons que $(J_{\hat{\beta}^{k+1}})_j = \frac{\partial \hat{\beta}_j^{k+1}}{\partial y}$, alors :

$$\begin{aligned} \frac{\partial \hat{\beta}_j^{k+1}}{\partial y} &= \frac{1}{\|x_j\|^2 + n\alpha_{2,1}} \frac{\partial \text{ST}(t_j, n\alpha_{1,1})}{\partial y} \\ &= \frac{1}{\|x_j\|^2 + n\alpha_{2,1}} \left(\frac{\partial \text{ST}(t_j, n\alpha_{1,1})}{\partial \beta} \frac{\partial \hat{\beta}^k}{\partial y} + \frac{\partial \text{ST}(t_j, n\alpha_{1,1})}{\partial \Theta} \frac{\partial \hat{\Theta}^k}{\partial y} + \frac{\partial \text{ST}(t_j, n\alpha_{1,1})}{\partial y} \right) \\ &= \frac{1}{\|x_j\|^2 + n\alpha_{2,1}} \left(\frac{\partial \text{ST}(t_j, n\alpha_{1,1})}{\partial \beta} J_{\hat{\beta}^k} + \frac{\partial \text{ST}(t_j, n\alpha_{1,1})}{\partial \Theta} J_{\hat{\Theta}^k} + \frac{\partial \text{ST}(t_j, n\alpha_{1,1})}{\partial y} \right) \\ &= \frac{1}{\|x_j\|^2 + n\alpha_{2,1}} \left((e_j \|x_j\|^2 - x_j^\top X)^\top J_{\hat{\beta}^k} - x_j^\top Z^\top J_{\hat{\Theta}^k} + x_j^\top \right) \mathbb{1}_{\{|t_j| \geq n\alpha_{1,1}\}} \end{aligned}$$

Finalement, en factorisant par x_j^\top à gauche et en appliquant à r^k de l'autre :

$$(J_{\hat{\beta}^{k+1}} r^{k+1})_j = \frac{x_j^\top \left((x_j e_j^\top - X) J_{\hat{\beta}^k} + (\text{Id}_n - Z J_{\hat{\Theta}^k}) \right) r^k}{\|x_j\|^2 + n\alpha_{2,1}} \mathbb{1}_{\{|x_j^\top (r^k + \hat{\beta}_j^k x_j)| \geq n\alpha_{1,1}\}} \quad \cdot$$

Le résultat concernant les interactions s'établit de manière similaire. \square

3 Expérimentations numériques

On se limitera ici au cas $\alpha_{1,1} = \alpha_{1,2}$ et $\alpha_{2,1} = \alpha_{2,2}$: on ne différencie pas les pénalités des effets simples et des interactions, et ces coefficients encodé par ratio noté `ll_ratio`, afin de déterminer un compromis entre les normes ℓ_1 et ℓ_2 , comme dans le module `sklearn` (Pedregosa et al., 2011).

3.1 Étude sur données simulées

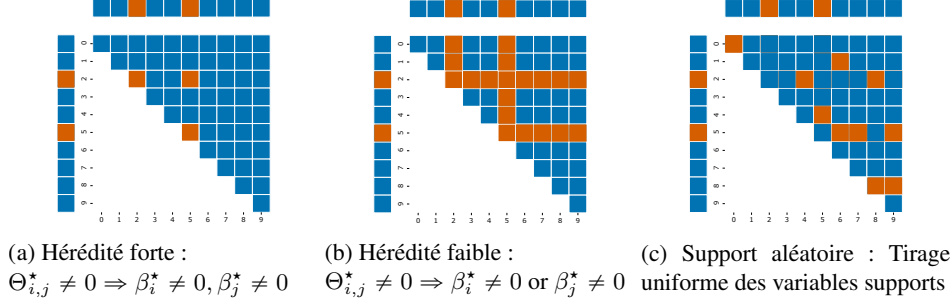


FIG. 1: Comparaison des scénarios d'hérédité où seuls β_3^* et β_6^* sont des coefficients actifs (orange). À gauche, on illustre trois indices pour les coefficients d'interactions actifs $\Theta_{3,6}^*$, $\Theta_{3,3}^*$ et $\Theta_{6,6}^*$. Au centre, on observe l'ensemble des possibilités de coefficients actifs pour l'hérédité faible et à droite le cas aléatoire, sans lien entre les variables actives de β^* et de Θ^* .

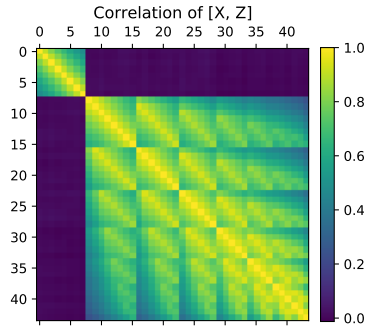


FIG. 2: Structure de la corrélation de $[X, Z]$, selon la corrélation Toeplitz de X : $\text{corr}(x_i, x_j) = 0.9^{|i-j|}$, avec $p = 8$ et $q = 36$. La partie correspondant à X peut-être observée en haut à gauche, celle concernant Z en bas à droite.

Plan d'expérience : pour comparer Elastic Net with Interaction avec sa version débiaisée, on étudie plusieurs scénarios d'hérédités entre les effets simples et les interactions : forte, faible ou aléatoire (Figure 1), en s'inspirant de Bien et al. (2013). Nous générons X comme une matrice de taille $n \times p$ issue du tirage d'une loi normale multivariée centrée mais dont la variance induit une corrélation Toeplitz (Figure 2). Les vecteurs des coefficients des effets simples et des interactions β^* et Θ^* sont artificiels : les coefficients non nuls sont choisis aléatoirement (selon les contraintes liées à l'hérédité) parmi β^* (resp. Θ^*) = ± 1 . Enfin, la variable du bruit ε , est simulée selon $\mathcal{N}(0, \sigma^2 \text{Id}_n)$, où σ contrôle le niveau de bruit en fonction du rapport signal sur bruit SNR fixé à 2, qu'on définit Équation (8) d'après Bühlmann et Mandozzi (2013) :

$$\text{SNR} = \frac{\sqrt{\beta^{*\top} X^\top X \beta^* + \Theta^{*\top} Z^\top Z \Theta^*}}{\sqrt{n} \sigma} \iff \sigma = \frac{\sqrt{\beta^{*\top} X^\top X \beta^* + \Theta^{*\top} Z^\top Z \Theta^*}}{\sqrt{n} \text{SNR}} . \quad (8)$$

La réponse est obtenue par $y = X\beta^* + Z\Theta^* + \varepsilon$, avec $n = 100$ observations, $p = 50$ variables simples ($q = 1275$ interactions). Nos résultats sont moyennés sur 100 répétitions aléatoires de validations croisées 5-blocs (`l1_ratio` ∈ {1, 0.99, 0.95, 0.9}) avec X et Z standardisées à la volée pour chaque estimateur. La tolérance du saut de dualité est fixée à 10^{-4} , la grille d'alpha est telle que $\alpha_{\min} = \frac{\alpha_{\max}}{100}$ (en gardant la convention de grille de `sklearn`).

Critère de comparaison : on veut montrer que le débiaisage améliore la sélection de variables. Dans notre cas, les vrais positifs VP (resp. négatifs TN) sont les coefficients qui sont

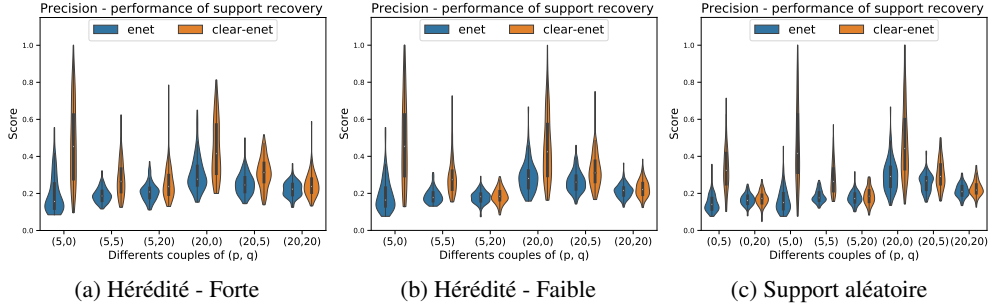


FIG. 3: Précision : résultats moyennés sur 100 répétitions aléatoires de validations croisées 5-blocs ($l1_ratio \in \{1, 0.99, 0.95, 0.9\}$). CLEAR Elastic Net with Interaction obtient de meilleurs résultats surtout quand le support est réduit, et inclut moins de faux positifs.

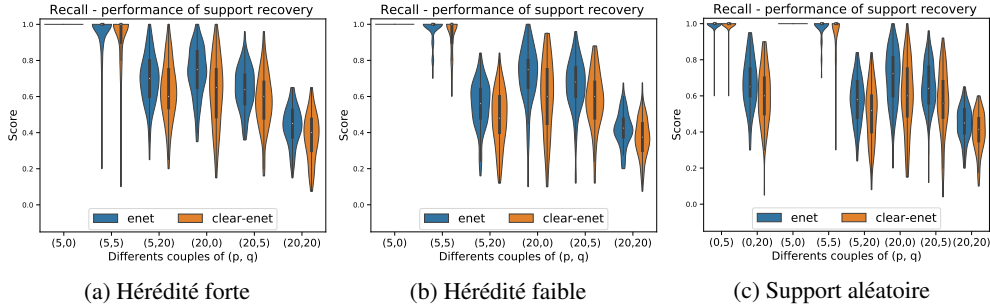


FIG. 4: Rappel : résultats moyennés sur 100 répétitions aléatoires de validations croisées 5-blocs ($l1_ratio \in \{1, 0.99, 0.95, 0.9\}$). Les deux méthodes ont des résultats du même ordre, mais l'estimateur débiaisé a manqué de vrais positifs par rapport à la version originale et a donc de moins bons résultats.

non-nuls (resp. nuls) pour l'estimation et les vrais coefficients. Les faux négatifs FN (resp. positifs FP) sont les coefficients estimés nuls (resp. non-nuls) mais en réalité non-nuls (resp. nuls). On a la précision $\frac{VP}{VP+FP}$ (resp. rappel $\frac{VP}{VP+FN}$) qui mesure à quel point un modèle inclut (resp. oublie) de faux (resp. vrais) positifs. Enfin, le F score : $2 \times \frac{\text{précision} \times \text{rappel}}{\text{précision} + \text{rappel}}$: mesure un équilibre entre les scores précédent et prend en compte le grand nombre de vrais négatifs.

Résultats : on voit Figure 3 que peu importe le scénario, CLEAR Elastic Net with Interaction obtient une précision plus élevée que la version originale. Mais Figure 4, on constate que la version originale a des scores de rappel meilleures que la version débiaisée, cette dernière oubliant quelques vraies variables trouvées par la version originale. Enfin, Figure 5 on résume les figures précédentes : dans la majorité des cas, le F score de la version débiaisée est légèrement meilleur que celui de Elastic Net with Interaction. Aussi, Figure 6 on compare l'erreur quadratique moyenne : elles sont du même ordre, même si celle de la version débiaisée est légèrement plus élevée. On conclut que CLEAR Elastic Net with Interaction sélectionne moins de variables que la version originale : il inclura moins de faux positifs (Figure 3) mais

ElasticNet avec gestion des interactions et débiaisage

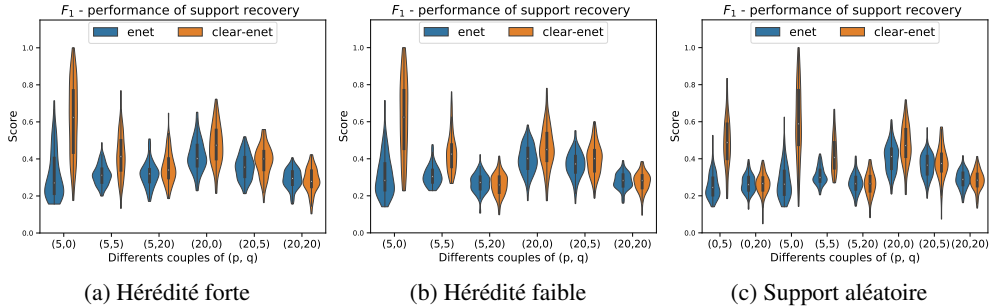


FIG. 5: F_1 : résultats moyennés sur 100 répétitions aléatoires de validations croisées 5-blocs ($ll_ratio \in \{1, 0.99, 0.95, 0.9\}$) avec X et Z standardisées à la volée pour chaque méthode. On voit que CLEAR-Enet a en général des scores plus élevés que l'estimateur original, ainsi le débiaisage améliore la sélection de variable, notamment quand la taille de (β^*, Θ^*) est faible.

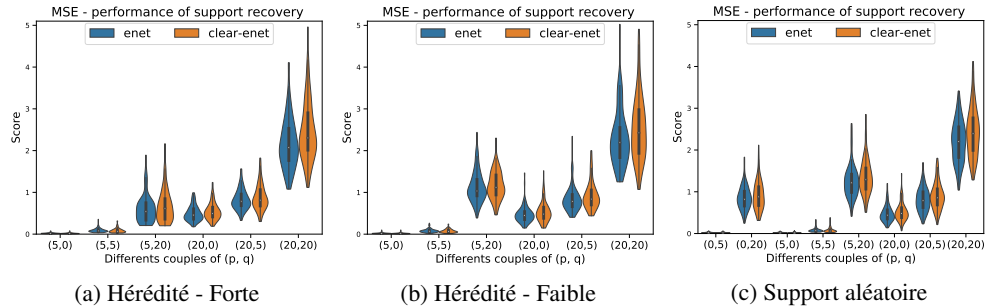
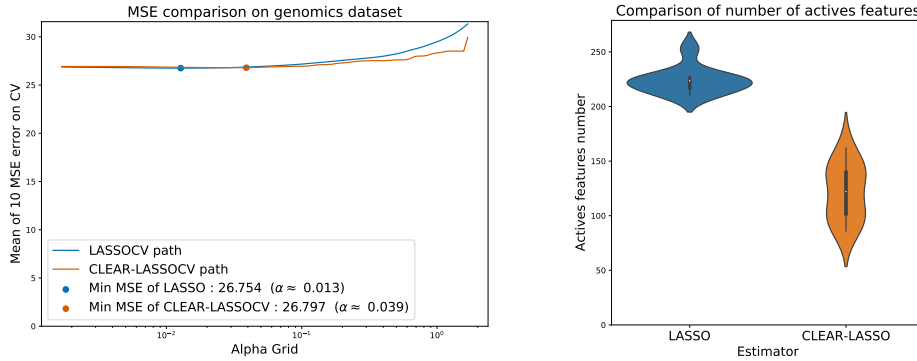


FIG. 6: Erreur quadratique moyenne : résultats moyennés sur 100 répétitions aléatoires de validations croisées 5-blocs ($ll_ratio \in \{1, 0.99, 0.95, 0.9\}$). Les méthodes ont une erreur du même ordre, même si celle de CLEAR Elastic Net with Interaction est un peu plus élevée.

manquera certains coefficients non nuls (Figure 4). Néanmoins, le compromis illustré Figure 5 montre que CLEAR Elastic Net with Interaction est en général plus efficace en sélection de variables, pour une erreur quadratique moyenne du même ordre.

3.2 Étude sur données réelles : données génomiques

Présentation du jeu de données : on s'intéresse à des données génomiques où les vecteurs réponses sont des mesures d'expressions génétiques (RNA-seq) pour des patients atteints de cancers. Ils partagent la même matrice de design où chaque ligne représente un unique gène de l'ADN et où les colonnes correspondent à différentes mesures de l'ADN (détaillées dans (Bessière et al., 2018)). Nous gardons uniquement $p = 530$ variables pour $n = 19393$ observations et utilisons comme vecteur réponse y la première colonne de la matrice des 241 réponses. Notre objectif sur ces données est de construire un modèle de régression linéaire prenant en compte les interactions. Mais, dû au temps de calcul, on s'intéresse dans ce travail à l'effet du débiaisage sur les effets principaux, la prise en compte des interactions étant lente.

(a) Erreur quadratique moyenne en fonction de α .

(b) Nombre de variables sélectionnées.

FIG. 7: Comparaison de CLEAR LASSO et de l'estimateur original sans interaction sur données génomiques ($n = 19393$, $p = 530$, $q = 0$). Résultats moyennés sur 10 répétitions aléatoires de validations croisées 5-blocs. À gauche, on observe que l'erreur minimum mesurée pour la version débiaisée est pour un α plus grand que pour la version originale. À droite, on voit que le nombre de variables sélectionnées par la version débiaisée est plus faible.

Méthodologie et Resultats : pour avoir des résultats stables, on effectue 10 validations croisées 5-blocs sur des données mélangées aléatoirement. Les paramètres de l'expérience sont : la tolérance du saut de dualité est 10^{-4} , la grille d'alpha est telle que $\alpha_{\min} = \frac{\alpha_{\max}}{1000}$, enfin le `l1_ratio` est fixé à un, c'est à dire que le cas étudié est un LASSO. On observe Figure 7a que le minimum de CLEAR Elastic Net with Interaction est pour une valeur d'alpha plus élevée que pour Elastic Net with Interaction, ce qui se traduit par un nombre de coefficients actifs estimés plus faible que la version biaisée ce qu'on observe Figure 7b. Enfin, l'erreur quadratique moyenne est du même ordre sans et avec débiaisage (plus élevée de l'ordre de 0.1%).

4 Conclusion

Nous avons présenté les résultats statistiques de notre estimateur CLEAR Elastic Net with Interaction. Il apparaît sur données simulées et réelles que le débiaisage permet d'obtenir un nombre de variables actives plus faible pour une erreur quadratique moyenne du même ordre, rendant le modèle débiaisé meilleur en sélection de variables. Les prochains travaux porteront sur des données de tailles plus importantes, l'objectif étant d'inclure les interactions dans notre analyse des données réelles. Le temps de calcul étant une limite pour ces cas, l'adaptation d'une méthode d'ensemble actif est à considérer, en partant par exemple de Massias et al. (2018).

Références

Bascou, F., S. Lèbre, et J. Salmon (2020). Debiasing the elastic net for models with interactions. In *Journées de Statistique*.

- Bessière, C., M. Taha, F. Petitprez, J. Vandel, J.-M. Marin, L. Bréhélin, S. Lèbre, et C.-H. Lecellier (2018). Probing instructions for expression regulation in gene nucleotide compositions. *PLOS Computational Biology* 14(1), 1–28.
- Bien, J., J. Taylor, et R. Tibshirani (2013). A lasso for hierarchical interactions. *Ann. Statist.* 41(3), 1111–1141.
- Bühlmann, P. et J. Mandozzi (2013). High-dimensional variable screening and bias in subsequent inference, with an empirical comparison.
- Deledalle, C.-A., N. Papadakis, J. Salmon, et S. Vaiter (2017). CLEAR : Covariant LEAST-square Re-fitting with applications to image restoration. *SIAM J. Imaging Sci.* 10(1), 243–284.
- Friedman, J., T. J. Hastie, H. Höfling, et R. Tibshirani (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.* 1(2), 302–332.
- Friedman, J., T. J. Hastie, et R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33(1), 1–22.
- Hoerl, A. E. et R. W. Kennard (1970). Ridge regression : Biased estimation for nonorthogonal problems. *Technometrics* 12(1), 55–67.
- Massias, M., A. Gramfort, et J. Salmon (2018). Celer : a Fast Solver for the Lasso with Dual Extrapolation. In *ICML*.
- Osher, S., M. Burger, D. Goldfarb, J. Xu, et W. Yin (2005). An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.* 4(2), 460–489.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, et E. Duchesnay (2011). Scikit-learn : Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Salmon, J. (2017). *On high dimensional regression : computational and statistical perspectives*. Habilitation à diriger des recherches, ENS Paris-Saclay.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 58(1), 267–288.
- Tikhonov, A. N. (1943). On the stability of inverse problems. *Dokl. Akad. Nauk SSSR* 39, 176–179.
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Publishing Company.
- Zou, H. et T. J. Hastie (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67(2), 301–320.

Summary

We present some statistical results for a penalized and de-biased regression estimator, to handle in high dimension interactions and sparsity. The statistical analysis focuses on the approach proposed in (Bascou et al., 2020), in particular to illustrate the functioning of debiasing. Moreover, it is shown that it allows the selection of fewer variables than the model without debiasing on simulated and real data.