

Approche bloc en ACP group-sparse: le package sparsePCA

Marie Chavent^{a,b} et Guy Chavent^c

^aIMB, UMR 5251, Université de Bordeaux, 33400 Talence

^bInria Bordeaux Sud-Ouest, équipe CQFD, 33405 Talence
marie.chavent@u-bordeaux.fr

^cInria-Paris, 75589 Paris
guy.chavent@inria.fr

Mots clefs : ACP sparse, sélection de variables groupe, optimisation en bloc.

La plupart des algorithmes développés ces dernières années en ACP sparse visent à déterminer une seule composante principale et utilisent le processus de déflation hérité de l'ACP non sparse lorsqu'il s'agit de calculer les composantes suivantes (voir par exemple [2], [5]). Cependant, l'utilisation de la déflation en ACP sparse où les composantes et les loadings ne sont plus nécessairement orthogonaux peut mener à des difficultés [4] et on peut s'attendre à ce que l'optimisation direct de la variance selon tous les loadings soit plus efficace qu'une approche séquentielle. Certains auteurs ont ainsi proposés des algorithmes déterminant tous les loadings simultanément : Zou et al. [6] résolvent l'ACP sparse comme un problème de type régression alternée et Journée et al. [3] utilisent une approche duale de type bloc.

Nous allons présenter dans cette communication une approche de type bloc en ACP group-sparse où les variables sont organisées en groupes et où la sparsité s'applique aux groupes de variables et non plus aux variables individuelles. Cette approche généralise au cas de variables groupe la formulation ℓ_1 -sparse de Journée et al. [3] et l'algorithme associé.

Ce nouvel algorithme d'ACP group-sparse a été implémenté dans le package R **sparsePCA** qui est disponible à l'adresse suivante : <https://github.com/chavent/sparsePCA>. Nous présenterons les principales fonctionnalités de ce package et nous comparerons sur une étude de simulation les approches bloc et déflation en ACP group-sparse [1]. Cette étude de simulation montrera également comment la prise en compte de l'information sur les groupes en ACP group-sparse peut permettre de mieux retrouver les patterns de sparsité qu'avec l'ACP sparse.

Références

- [1] Chavent, M., Chavent, G. (2017). Group-sparse bloc PCA and explained variance, *arXiv preprint arXiv:1705.00461 [stat.ML]*
- [2] D'Aspremont, A., Bach, F., El Ghaoui, L. (2008). Optimal solutions for sparse principal component analysis. *Journal of Machine Learning Research*, 9(Jul):1269-1294.
- [3] Journée, M., Nesterov, Y., Richtarik, P., Sepulchre, R. (2010). Generalized power method for sparse principal component analysis. *Journal of Machine Learning Research*, 11(Feb):517-553.
- [4] Mackey, L.W. (2009). Deflation methods for sparse pca. *In Advances in neural information processing systems*, 1017-1024.
- [5] Shen, H., Huang, J.Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of multivariate analysis*, 99(6):1015-1034, 2008.
- [6] Zou, H., Hastie, T., Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265-286.