

# Un algorithme EM pour une version parcimonieuse de l'analyse en composantes principales probabiliste

Charles Bouveyron\*, Julien Jacques\*\*

\*Université Paris Descartes - Laboratoire MAP5  
charles.bouveyron@parisdescartes.fr,  
<http://w3.mi.parisdescartes.fr/~cbouveyr/>

\*\*Université Lumière Lyon 2 - Laboratoire ERIC  
julien.jacques@univ-lyon2.fr  
<http://eric.univ-lyon2.fr/~jjacques/>

**Résumé.** Nous considérons une version parcimonieuse de l'analyse en composantes principales probabiliste. La pénalité  $\ell_1$  imposée sur les composantes principales rend leur interprétation plus aisée en ne faisant dépendre ces dernières que d'un nombre restreint de variables initiales. Un algorithme EM, simple de mise en œuvre, est proposé pour l'estimation des paramètres du modèle. La méthode de l'heuristique de pente est finalement utilisée pour choisir le coefficient de pénalisation.

## 1 Introduction

L'analyse en composantes principales (ACP, Jolliffe (1986)) est une des méthodes, si ce n'est la méthode, d'analyse exploratoire les plus couramment utilisées. Elle a été ré-interprétée sous un formalisme probabiliste par Tipping et Bishop (1999), montrant que les composantes principales pouvaient être estimées par maximum de vraisemblance dans le cadre d'un modèle à variables latentes. Avec l'avènement des données de grande dimension, la problématique consistant à sélectionner un petit nombre de variables d'intérêt parmi l'ensemble des variables disponibles est devenue primordiale. Un des soucis majeurs de l'ACP dans cette optique est que les composantes principales sont définies comme une combinaison linéaire de l'ensemble des variables initiales. Des versions parcimonieuses de l'ACP (Zou et al., 2004) ainsi que de sa version probabiliste (Guan et Dy, 2009) ont été proposées récemment. La version parcimonieuse de Zou et al. (2004) repose sur l'ajout d'une pénalisation de type  $\ell_1$  au problème des moindres carrés, qui nécessite le choix du coefficient de pénalisation de façon heuristique. Dans Guan et Dy (2009), une version sparse bayésienne de l'ACP probabiliste est proposée. Nous proposons dans ce travail une alternative fréquentiste utilisant un algorithme EM pour l'inférence. La procédure d'estimation obtenue à l'avantage d'être particulièrement simple, et ne nécessite pas le choix de loi a priori. Elle offre en outre la possibilité de considérer le problème du choix de la pénalité comme un problème de choix de modèles.