

# ALGORITHMES GRADIENT PROXIMAUX PERTURBÉS POUR L'ESTIMATION DE MODLES À EFFETS MIXTES PÉNALISÉS

Gersende Fort <sup>1</sup> & Edouard Ollier <sup>2,3</sup> & Adeline Samson <sup>3</sup> & Eric Moulines <sup>4</sup>

<sup>1</sup> *IMT UMR5219, Université de Toulouse, CNRS; F-31062 Toulouse Cedex 9, France.*

<sup>2</sup> *INSERM, U1059, Dysfonction Vasculaire et Hémostase, Saint Etienne, France.*

<sup>3</sup> *U.M.P.A., Ecole Normale Supérieure de Lyon, CNRS UMR 5669; INRIA, Project-team NUMED. 46 Allée d'Italie, 69364 Lyon Cedex 07, France.*

<sup>4</sup> *Université Grenoble-Alpes, Laboratoire Jean Kuntzmann, UMR CNRS 5224.*

<sup>5</sup> *CMAP, Ecole Polytechnique, Palaiseau, France.*

**Résumé.** Les modèles à données cachées sont largement utilisés de nos jours notamment pour modéliser des données longitudinales comme c'est le cas en pharmacocinétique avec les modèles non linéaires à effets mixtes. La sélection de tels modèles peut faire appel à des problèmes de maximum de vraisemblance pénalisée pour lesquels des versions pénalisées de l'algorithme SAEM ont été développées. Bien que fonctionnant très bien en pratique, il n'existe pour l'instant aucune analyse théorique de la convergence de tels algorithmes. Dans ce travail, nous présenterons donc des résultats de convergence théorique dans le cas où la vraisemblance est concave. Nous verrons de plus qu'il est possible de réduire l'étape M à une application d'un opérateur gradient proximal tout en conservant la convergence. Ces résultats seront illustrés sur données simulées ainsi que sur un exemple de pharmacogénomique. **Mots-clés.** Modèles à effet mixtes, vraisemblance pénalisée, algorithme EM, gradient proximal perturbé.

**Abstract.** Latent variable models are classical tools to model longitudinal data such as in population pharmacokinetic with non linear mixed effects models. Selection of such model may rely on the use of penalized maximum likelihood estimator for which penalized version of the SAEM algorithm have been developed. Even if these algorithm perform well in practice, there is a lack of theoretical information concerning their convergence. In this work, we will present convergence results in the case of a concave likelihood. Moreover we will show that the M step could be reduced to the application of a proximal gradient operator. This result will be illustrated on simulated data and on a real example from pharmacogenomics.

**Keywords.** Mixed effect models, Penalized likelihood, EM algorithm, Perturbed proximal gradient.

# 1 Introduction

Les modèles à effets mixtes sont des outils classiques pour modéliser des données de nature longitudinales. C'est notamment le cas des données de pharmacocinétique pour lesquels les modèles non linéaires à effet mixtes sont devenus le principal outil d'analyse. Des travaux récents (Ollier (2016), Bertrand (2015)) se sont intéressés à la sélection de ces modèles en utilisant un problème de vraisemblance pénalisée par des normes induisant de la sparsité. Les algorithmes développés dans ces travaux, sont basés sur des versions pénalisées de l'algorithme SAEM (Kuhn (2004)). Ils possèdent de bonnes propriétés empiriques cependant aucune preuve théorique de leur convergence n'a été réalisée.

## 2 Modèles à variables cachées

Dans ce travail, nous nous intéressons à des modèles à variables cachées dont la vraisemblance complète, paramétrée par  $\theta \in \mathbb{R}^d$ , appartient à la famille exponentielle. Dans ce cas, la log-vraisemblance  $\ell$  est de la forme:

$$\theta \mapsto \ell(\theta) := \log \int_{\mathcal{Z}} \exp(\phi(\theta) + \langle S(z), \psi(\theta) \rangle) \mu(dz), \quad (1)$$

où  $\langle a, b \rangle$  correspond au produit scalaire entre deux vecteurs  $a, b \in \mathbb{R}^l$ ,  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^q$  et  $S : \mathcal{Z} \rightarrow \mathbb{R}^q$  sont des fonctions mesurables avec  $\mu$  une mesure positive  $\sigma$ -finie sur  $\mathcal{Z}$ . La quantité  $\theta \mapsto \phi(\theta) + \langle S(z), \psi(\theta) \rangle$  correspond à la log vraisemblance des données complètes,  $z$  est le vecteur de données cachées. De plus, on a sous condition de régularité:

$$\nabla \ell(\theta) = \nabla \phi(\theta) + \mathbf{J} \psi(\theta) \int_{\mathcal{Z}} S(z) \pi_{\theta}(z) \mu(dz), \quad \text{with } \pi_{\theta}(z) := \frac{\exp(\langle S(z), \psi(\theta) \rangle)}{\int_{\mathcal{Z}} \exp(\langle S(u), \psi(\theta) \rangle) \mu(du)}, \quad (2)$$

avec  $\mathbf{J} \psi(\theta)$  la transposé de la matrice jacobienne des fonctions  $\phi$  et  $\psi$  en  $\theta$ .

Nous nous intéresserons aux problèmes de maximum de vraisemblance pénalisée de la forme:

$$\operatorname{argmax}_{\theta \in \mathbb{R}^d} F(\theta) = \ell(\theta) - g(\theta), \quad (3)$$

avec  $g$  est fonction convexe possiblement non dérivable, par exemple  $g(\theta) = \|\theta\|_1$ .

## 3 Algorithmes EM pénalisé et gradient proximal perturbé

Dans le cadre de ce travail, nous étudierons deux algorithmes. Le premier étant un algorithme de type EM dans lequel l'étape de maximisation correspond à un problème d'optimisation

non lisse du fait de la pénalité. Une solution explicite n'étant pas toujours disponible, ce problème est donc résolu numériquement à chaque itération de l'algorithme EM.

**input** : The initial value  $\theta_0 \in \Theta$   
**output**: The sequence  $\{\theta_n, n \geq 0\}$

- 1 **for**  $n \geq 0$  **do**
- 2     *E-step* ;
- 3     Compute an approximation  $S_{n+1}$  of  $\bar{S}(\theta_n)$  ;
- 4     *M-step* ;
- 5     Set  $\theta_{n+1} = \operatorname{argmax}_{\theta} \{\phi(\theta) + \langle S_{n+1}, \psi(\theta) \rangle - g(\theta)\}$

**Algorithm 1:** Algorithme de type EM pénalisé.

Le second algorithme est de type gradient proximal perturbé (Atchadé (2016)) dans lequel l'étape M correspond juste à l'application de l'opérateur proximal associé à la fonction  $g$ .

**input** : The initial values  $\theta_0 \in \Theta$  and  $Z_{m_0, -1} := z_*$ , a  $(0, 1/L]$ -valued sequence  $\{\gamma_n, n \geq 0\}$  and an integer valued sequence  $\{m_n, n \geq 0\}$   
**output**: The sequence  $\{\theta_n, n \geq 0\}$

- 1 **for**  $n \geq 0$  **do**
- 2     *Simulation-step* ;
- 3     sample a path  $Z_{1,n}, \dots, Z_{m_{n+1}, n}$  of a Markov chain with invariant distribution  $\pi_{\theta_n} d\mu$  and started from  $Z_{m_n, n-1}$ ;
- 4     *Expectation step* ;
- 5     Compute an approximation  $S_{n+1}$  of  $\bar{S}(\theta_n)$  ;
- 6     *Maximization step* ;
- 7     Set  $\theta_{n+1} = \operatorname{Prox}_{\gamma_{n+1}, g}(\theta_n + \gamma_{n+1} \{\nabla \phi(\theta_n) + \Psi(\theta_n) S_{n+1}\})$

**Algorithm 2:** Algorithme de type gradient proximal perturbé

Différentes stratégies pour l'approximation de  $\bar{S}(\theta_n)$  seront étudiées, notamment l'utilisation de schémas d'approximation stochastique ou encore de méthode de Monte Carlo.

## 4 Analyse théorique de la convergence des algorithmes et illustration sur données simulées et réelles

Nous présenterons tout d'abord des résultats théoriques de convergence dans le cas où  $\ell$  et  $-g$  sont concaves et discuterons la mise en oeuvre des différentes techniques d'approximation de  $\bar{S}(\theta)$ . Ces résultats seront ensuite illustrés numériquement dans un cas où  $\ell$  concave et un autre non-concave. Pour finir, nous illustrerons le comportement de ses algorithmes sur un jeu de données de pharmacocinétique en grande dimension ( en nombre de covariables).

## Bibliographie

- [1] Ollier, E., Samson, A., Delavenne, X. et Viallon, V. (2016), A SAEM algorithm for fused lasso penalized non linear mixed effect models: Application to group comparison in pharmacokinetic, *Comput. Stat. Data Anal.*, **95**, 207–21.
- [2] Kuhn, E. et Lavielle, M. (2004), Coupling a stochastic approximation version of EM with an MCMC procedure, *ESAIM: Probability and Statistics*, 8, 115-131.
- [3] Bertrand, J., De Iorio, M., et Balding, D. J. (2015), Integrating dynamic mixed-effect modelling and penalized regression to explore genetic association with pharmacokinetics, *Pharmacogenetics and genomics*, 25(5), 231.
- [4] Atchade, Y. F., Fort, G. et Moulines, E. (2016), On perturbed proximal gradient algorithms, *Accepté pour publication dans Journal of Machine Learning Research*.