Dans de nombreux problèmes d'apprentissage, on est confronté à la résolution d'un problème d'optimisation (P) : minimisation d'une fonction de la forme f+g où f est une fonction régulière, g est une fonction convexe non régulière, et ni f+g ni le gradient de f ne sont calculables explicitement.

Typiquement, f traduit une mesure d'adéquation d'une famille de modèles aux observations, et g traduit des contraintes sur le paramètre du modèle. Par exemple, en inférence dans les modèles statistiques à données cachées, f est la log-vraisemblance (négative), définie par une intégrale sur l'espace des données cachées, et son gradient s'écrit comme une espérance (incalculable) sous la loi conditionnelle des données cachées sachant les observations. En apprentissage, f exprimera une fonction de perte empirique calculée sur un très grand nombre d'exemples.

Dans ces deux exemples, si l'on ne peut disposer du gradient de f ou si l'on doit y renoncer du fait d'un coût de calcul prohibitif, on peut au moins en calculer une approximation par une méthode de Monte Carlo, approximation éventuellement biaisée - par exemple lorsqu'elle repose sur des échantillons d'une chaîne de Markov.

Les algorithmes gradient-proximaux sont des algorithmes itératifs pour la résolution de tels problèmes d'optimisation (P). Chaque itération combine une étape de gradient relative à la partie régulière de la fonction objectif, et une itération d'un opérateur proximal relatif à la partie non régulière.

Motivés par des applications à l'inférence paramétrique dans les modèles à données cachées, nous avons étudié la convergence d'algorithmes de type gradient-proximaux perturbés, la perturbation portant sur l'étape de gradient. L'originalité de nos résultats est de donner des conditions réalistes sur la perturbation pour traiter notamment le cas d'approximations stochastiques biaisées du gradient.

Je présenterai les résultats de (vitesse de) convergence obtenus dans le cas où f et g sont convexes, et discuterai de la mise en oeuvre des approximations stochastiques : doit-on choisir un pas constant ou un pas décroissant dans l'étape de gradient ? doit-on construire des approximations Monte Carlo avec des sommes à nombre croissant ou constant de termes ?

Nous illustrerons ces résultats par des applications à l'inférence par maximum de vraisemblance (MV) pénalisé dans des modèles à effets mixtes.

Dans ces modèles, l'estimateur MV est habituellement calculé par l'algorithme Expectation Maximization (EM) ou ses versions stochastiques. Nous ferons le lien entre l'approche EM et l'approche gradient-proximal et montrerons que les résultats de convergence des algorithmes gradient-proximaux perturbés permettent d'établir la convergence d'algorithmes de type Monte Carlo-EM et de type Stochastic Approximation-EM pour le calcul du MV pénalisé.

Travaux en collaboration avec Yves Atchadé (Univ. Michigan, USA), Eric Moulines (CMAP, Ecole Polytechnique), Edouard Ollier (UMPA, ENS Lyon) et Adeline Leclercq-Samson (LJK, Univ. Grenoble Alpes).