Bornes non-asymptotiques optimales pour la moyennisation de Ruppert-Polyak

Sébastien Gadat, Toulouse School of Economics

Mots-clés : Optimisation stochastique, moyennisation, inégalité de Kurdyka-Łojasiewicz

Nous considérons l'algorithme de Ruppert-Polyak des algorithmes stochastiques introduits par Robbins et Monro:

$$\theta_{n+1} = \theta_n - \gamma_{n+1}\nabla f(\theta_n) + \gamma_{n+1}\Delta M_{n+1},$$

où $-\nabla f(\theta_n) + \Delta M_{n+1}$ est une évaluation non biaisée du gradient d'une fonction f à minimiser, perturbée par un incrément de martingale.
La moyennisation introduite dans [R88, PJ92] consiste à opérer une moyennisation au sens de Cesaro de l'algorithme initial:

$$\bar{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}\theta_i$$

Dans notre travail, nous présentons des bornes non-asymptotiques optimales (au sens de Cramer-Rao) dans un cadre très général qui permet de gérer à la fois le cas de l'uniforme forte convexit de [BM11] mais aussi les situations plus générales où f sastisfait une condition faible de type Kurdyka-Łojiasewicz (voir [K98, L63]).
En particulier, cela nous permet de considérer les exemples problmatiques de type régression logistique séquentielle (voir [B14]) ou estimation récursive du quantile. Nous démontrons qu'un schéma optimal est obtenu pour une suite de pas $(\gamma_n)_{n\geq 1}$ donn par $\gamma_n = \gamma_1 n^{-\beta}$ avec $\beta = 3/4$, obtenant ainsi un terme de second ordre en $O(n^{-5/4})$.

Références

[B14] F. Bach, Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression, Journal of Machine Learning Research, vol. 15, 595-627 (2014).

[BM11] F. Bach and E. Moulines, Non-Asymptotic Analysis of Stochastic Approximation Algorithms for Machine Learning, Neural Information Processing Systems (NIPS) (2011).

[K98] K. Kurdyka, On gradients of functions definable in o-minimal structures, Ann. Inst. Fourier, Universit de Grenoble. Annales de lInstitut Fourier, vol. 48 (3), 769-783 (1998).

[L63] S. Łojasiewicz, Une propriété topologique des sous-ensembles analytiques réels, Editions du CNRS, Paris, Les Équations aux Dérivées Partielles, 87-89 (1963).

[R88] D. Ruppert, Efficient estimations from a slowly convergent Robbins-Monro process, Technical Report, 781, Cornell University Operations Research and Industrial Engineering, (1988).

[PJ92] B. T. Polyak and A. Juditsky, Acceleration of Stochastic Approximation by Averaging, SIAM Journal on Control and Optimization, vol. 30 (4), 838-855 (1992).

Sébastien Gadat, Toulouse School of Economics, Universit Toulouse I Capitole, 21 alles de Brienne, 31000 Toulouse