When Monte Carlo and Optimization met in a Markovian dance

Gersende Fort

CNRS

Institut de Mathématiques de Toulouse, France

ICTS "Advances in Applied Probability", Bengaluru, August 2019.

A dance, why ? New challenges in Data Science necessitate the combination of optimization techniques and Monte Carlo methods. Incorporating optimization tools in Monte Carlo methods is a known technique to improve their efficiency in order to learn on the fly a better value of some design parameters. Incorporating Monte Carlo techniques in Optimization algorithms allows the introduction of (stochastic) numerical approximations of intractable quantities. In both cases, this yields to algorithms which are perturbations of simpler ones whose asymptotic behavior is (usually) wll known. When introducing these intertwinings, it is fundamental to be sure that it will not destroy the convergence: convergence of the sampler to the target distribution, or convergence of the optimization method to the target set of solutions.

In the lectures, two points of view will be successively considered :

1- sufficient conditions for the adaptation mecanism of Monte Carlo algorithms, in order to identify the asymptotic distribution of the sampled points.

2- sufficient conditions on the stochastic perturbations, in order to ensure the correct convergence of the perturbed optimization algorithm. We will essentially consider the case of Stochastic Approximation (SA) algorithms (possibly also, Majorize-Minimization methods).

As a preliminary of these two parts, examples from Computational Statistics will be introduced. In both parts, we will consider the Markovian case: Monte Carlo methods based on Markov chains (MCMC); Stochastic Optimization fed with MCMC samples, which implies, as a unusual setting for SA, a biased stochastic approximation of intractable quantities.

The lectures will be concluded by a talk, presenting recent works on Stochastic Approximation combined with adaptive MCMC methods, and motivated by Computational Machine Learning problems. More precisely, we will consider the convergence of Proximal-Gradient based methods (also called "forward-backward" methods) for the optimization of a composite objective function i.e. a function defined as the sum a smooth part and of a non-smooth part. The convergence will be established, and we will discuss many implementation issues such as: choice of the design parameters, benefit of averaging techniques, benefit of Nesterov accelerations schemes, etc