A Shrinkage-Thresholding Metropolis adjusted Langevin algorithm for Bayesian Variable Selection

Authors: Amandine Schreck (Télécom ParisTech), Gersende Fort (Télécom ParisTech), Éric Moulines (Télécom ParisTech), Sylvain Le Corff (Université d'Orsay)

Abstract

We consider the long-standing problem of Bayesian variable selection in a linear regression model. Variable selection is a complicated task in high dimensional settings where the number of regression parameters P is much larger than the number of observations N. In this context, it is crucial to introduce sparsity assumptions based on the prior knowledge that only a few number of regression parameters are significant. Using a sequence of observations from a linear regression model, the aims are *(i)* to determine which components of the regression vector are active and explain the observations and *(ii)* to estimate the regression vector.

In this work, we introduce a new MCMC algorithm, called Shrinkage-Thresholding MALA (STMALA), designed to sample sparse regression vectors by jointly sampling a model and a regression vector in this model. This algorithm, which is a transdimensional MCMC method, relies on MALA (see [Roberts and Tweedie, 1996]). The proposal distribution of MALA is based on the computation of the gradient of the logarithm of the target distribution. In order to both deal with a non-differentiable target posterior distribution and to actually set some components to zero, we propose to combine MALA with a shrinkage-thresholding operator:

- first compute a noisy gradient step involving the term of the logarithm of the target distribution which is continuously differentiable;

- then a shrinkage-thresholding operator is applied to ensure sparsity and shrink small values of the regression parameters toward zero.

Such an algorithm is motivated by Bayesian variable selection with non-smooth priors. This algorithm can perform global moves from one model to a rather distant other one, which allows to explore efficiently high dimensional spaces in comparison to local move algorithms, like reversible jump MCMC (RJMCMC - see [Green 1995]). The geometric ergodicity of this new algorithm is proved for a large class of target distributions.

Joint work with Gersende Fort, Sylvain Le Corff and Éric Moulines.

References:

P. Green, *Reversible jump Markov chain Monte Carlo computation and Bayesian model determination*, Biometrika 82(4) 1995.

G.O. Roberts, R.L. Tweedie, *Exponential convergence of Langevin distributions and their discrete approximations*, Bernouilli 2(4) 1996.

Keywords: Markov Chain Monte Carlo, Proximal operators, Bayesian variable selection.