Thesis abstract

Efficient Hamiltonian Monte Carlo for large data sets by data subsampling

Doan Khue Dung Dang

Abstract of a thesis for a Doctorate of Philosophy submitted to The University of New South Wales, Sydney, Australia

Bayesian statistics carries out inference about the unknown parameters in a statistical model using their posterior distribution, which in many cases is computationally intractable. Therefore, simulation methods such as Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) are frequently used to approximate the posterior distribution. SMC has the attractive ability to accurately estimate the marginal likelihood, although it is computationally more expensive than MCMC. Nevertheless, both methods require efficient Markov moves to deal with complex, high-dimensional problems. While Hamiltonian Monte Carlo (HMC) is a remedy in many cases, it also increases the computational cost of the algorithms appreciably, especially for large data sets. This thesis presents some novel methods that focus on speeding up inference by combining HMC and data subsampling. The first contribution is a Metropolis-within-Gibbs algorithm that successfully speeds up standard HMC by orders of magnitude in two large data examples. I then show that the new approach can be incorporated into other HMC implementations such as the No-U-Turn sampler. The next contribution is an extension of the first method to SMC for Bayesian static models, which gives comparable results to full data SMC in terms of accuracy but is much faster in several model settings. The final contribution shows that the subsampling HMC scheme can also be applied to a thermodynamic integration method to estimate the marginal likelihood.

Dr. Doan Khue Dung Dang

E-mail: doankhuedung.dang@unsw.edu.au

URL: https://www.unsworks.unsw.edu.au/permalink/f/5gm2j3/unsworks_64135