Alternative Metrics for Riemann Manifold Monte Carlo

Vassilios Stathopoulos
Department of Computer Science
University of Glasgow
Glasgow, G12 8QQ
stathv@dcs.gla.ac.uk

Mark Girolami
Department of Statistical Science
University College London
London, WC1E 6BT
girolami@dcs.gla.ac.uk

Abstract

Recently [1] a new family of Markov Chain Monte Carlo algorithms which exploit the Riemann geometry of the parameter space of the target density have been proposed. The algorithms employ the manifold's metric tensor to automatically adapt the proposal mechanisms of Hamiltonian Monte Carlo (HMC) and Metropolis adjusted Langevin Algorithm (MALA) and achieve efficient sampling. For density functions a natural metric is the Fisher information which is derived by the KL divergence. The authors in [1] however, also stress the fact that the algorithms are applicable for any metric defining a manifold over the parameter space. Moreover, for many practical models the Fisher information matrix can be degenerate [2] or not analytically defined [3].

The Integrated Square Error (ISE) metric is a special case of the power divergence family [4] which has been used in several applications since it allows for robust estimation of parametric models [5] and induces sparsity in kernel density estimation [6]. Moreover, the ISE provides an analytic metric tensor for mixtures models. The ISE for a density $p(x|\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$ is defined as

$$\int_{\mathcal{X}} \nabla_{\boldsymbol{\theta}} p(x|\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}}^{T} p(x|\boldsymbol{\theta}) dx$$

In this work we study sampling efficiency in terms of Effective Sample Size (ESS) of Riemann manifold Metropolis adjusted Langevin (MMALA) and Hamiltonian Monte Carlo (RMHMC) algorithms defined on the manifolds induced by the Fisher information and Integrated Square Error metrics. We consider a logistic regression model and apply it on five real world datasets comparing our results with a component wise adaptive Metrpopolis-Hastings algorithm. Moreover, we experiment with mixtures of univariate Gaussians on five synthetically generated datasets and compare with a standard Gibbs sampler. For mixtures of Gaussians the Fisher information is not analytically defined and thus an empirical estimate is utilised.

Our results show that for the logistic regression model the ISE metric is not as efficient as the Fisher information although results are significantly better than adaptive component wise Metropolis-Hastings. For mixtures of Gaussians however, ISE allows more efficient sampling since its analytical expression is less computationally expensive as compared with the empirical estimate of the Fisher information.

Acknowledgments

This research is funded by the Engineering and Physical Sciences Council (EPSRC) grant number EP/H024875/1.

References

[1] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo. *To appear with discussion in Journal Of The Royal Statistical Society Series B*, ?(?):?, 2010.

[2] V. Stathopoulos and M. Filippone. Disussion on "Riemann manifold Langevin and Hamiltonian Monte Carlo" by Girolami, M. and Calderhead, B. *To appear in Journal Of The Royal Statistical Society Series B*, ?(?):?, 2010.

[3] V. Stathopoulos and M. Girolami. *Mixture Estimation and Applications*, chapter Manifold MCMC for Mixtures. To be published by Wiley series in probability and statistics. Wiley and Sons., 2010.

[4] A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559, 1998.

[5] D. W. Scott. Parametric statistical modeling by minimum integrated square error. *Technometrics*, 43(3):274–285, 2001.

[6] C. He and M. Girolami. Novelty detection employing an L2 optimal nonparametric density estimator. *Pattern Recognition Letters*, 25(12):1389–1397, 2004.