Improving the Effectiveness of MCMC through Adaptation

Edna Milgo[1] EMILGO@VUB.AC.BE
AI Lab, Vrije Universiteit Brussel, Belgium

Nixon Ronoh[1] NRONOH@VUB.AC.BE
AI Lab, Vrije Universiteit Brussel, Belgium

Peter Waiganjo Wagacha WAIGANJO@UONBI.AC.KE
University of Nairobi, Kenya

Bernard Manderick BERNARD.MANDERICK@VUB.AC.BE
AI Lab, Vrije Universiteit Brussel, Belgium

Keywords: Adaptive MCMC, Covariance Matrix Adaptation

Abstract

The Metropolis-Hastings Markov Chain Monte Carlo algorithm uses a proposal distribution to sample from a given target distribution. The proposal distribution has to be tuned beforehand which is an expensive exercise. However, adaptive algorithms automatically tune the proposal distribution based on knowledge from past samples, reducing the tuning cost. In this study we first examine Adaptive Metropolis, Metropolis Gaussian Adaptation and Covariance Matrix Adaptation Evolutionary Strategies. The latter is usually used as an optimization strategy but it is feasible to incorporate it in MCMC.

1. Introduction

Markov chain Monte Carlo (MCMC) techniques create a Markov chain that has a target distribution π as a unique invariant distribution. The Metropolis-Hastings (MH) algorithm generates an ergodic Markov chain and uses an acceptance probability to ensure that its invariant distribution is the target distribution. MH algorithms require a proposal distribution to be tuned beforehand by an expert in the domain area which can be difficult and expensive. We look at three adaptive strategies that aim at overcoming the drawbacks of non-adaptive MH algorithm, Adaptive Metropolis (AM), Metropolis Gaussian Adaptation(M-GaA) and the Covariance Matrix Adaptation Evolutionary Strategies (CMA-ES) (Hansen, 2011).

2. Adaptation

The main drawback of the MH algorithm is that the correlation among the samples in the Markov chain can be very high when the acceptance rate is low. Adaptive proposals allow the sample to learn and remain longer in good states hence minimum variance and high acceptance rate. Although the adaptive proposal does not simulate exactly the target distribution, proofs show that it can be very close to the target in practice (Haario et al., 1999). Adaptation is achieved by either selecting a fixed number of previous states or by using all the previous states in the Markov chain (Haario et al., 2001).

3. Adaptive Methods

3.1. Metropolis Gaussian Adaptation

Metropolis Gaussian Adaptation (M-GaA) implements the MH algorithm within Gaussian Adaptation (Müller & Sbalzarini, 2010). M-GaA sets a threshold c_T and lowers it until a predefined convergence criteria is achieved. The *expansion* and *contraction* factors $f_e > 1$ and $f_c < 1$ of the covariance matrix

[1]Milgo and Ronoh are Lecturers at Moi University and are currently PhD scholars at Vrije Universiteit Brussel under the VLIR-UOS scholarship.

Appearing in *Proceedings of Benelearn 2016*.

are determined by the desired hitting probability. A scalar step size r is used to control the search. When the parameter vectors are accepted, r is increased and the mean is moved to the current state of the Markov chain. The iterative sampler adapts the orientation and scale of the covariance matrix under the constraint of the fixed hitting probability (Müller & Sbalzarini, 2011).

3.2. Adaptive Proposal

Adaptive Proposal (AP) samples a candidate point from the proposal distribution and accepts or rejects it based on a weighted average of the mean positions and their covariance matrix. After a burn-in period, the proposal distribution is centered at the current state, X_n of the Markov chain. The covariance matrix is set to $C_n = s_d Cov(X_0, \cdots, X_{n-1}) + s_d \epsilon I_d$, where s_d is a parameter that depends only on the dimension d of the state space, $\epsilon > 0$ is a very small constant and I_d denotes the d-dimensional identity matrix (Haario et al., 2006). A global scale factor is used to control the acceptance rate and the vanishing adaptation is applied to maintain ergodicity of the chain.

3.3. Covariance Matrix based Adaptation

CMA-ES is recognized as one of the most competitive evolutionary algorithms for real values optimization (Suttorp et al., 2009). CMA-ES was initially proposed as an optimization tool for non-separable and ill-conditioned objective functions. CMA-ES approximates the inverse Hessian matrix and maps the search distribution to the contour lines of the objective function. The search steps are taken by stochastic variation(mutation) of points which have been visited recently. The best points, determined using a fitness function, are allowed to evolve.

CMA-ES maintains ergordicity by adapting the chain only at the time of recurrence of the atom (Müller, 2010; Müller & Sbalzarini, 2010). CMA-ES also ensures that the invariance property is preserved. Evolution paths, the time evolution of the distribution mean, contain significant information about the correlation between consecutive steps. The covariance evolution path is used to facilitate a much faster variance increase of favorable directions. The step size path is used to conduct an additional step-size control. This step-size control aims to make consecutive movements of the mean orthogonal.

4. Conclusion

Adaptive MCMC overcomes the challenges faced by traditional MCMC by enabling the proposal distribution to learn online. In this paper we examined the Adaptive Metropolis, the Metropolis Gaussian Adaptation and CMA-ES. The later adapts well to complex distributions through recombination and mutation which allow variations in the sample place. By incorporating Metropolis-Hastings algorithm in CMA-ES, the effectiveness of adaptive MCMC can be further enhanced.

Acknowledgement

The authors wish to thank VLIR-UOS, Vrije Universiteit Brussel and Moi University without whose support this work would not have been possible.

References

Haario, H., Laine, M., Mira, A., & Saksman, E. (2006). Dram: Efficient adaptive mcmc. *Statistics and Computing*, *16-4*, 339–354.

Haario, H., Saksman, E., & Tamminen, J. (1999). Adaptive proposal distribution for random walk metropolis algorithm. *Computational Statistics*, *14-3*, 375–395.

Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, *7 no. 2*, 223–242.

Hansen, N. (2011). The cma evolution strategy: A tutorial.

Müller, C. L. (2010). Exploring the common concepts of adaptive mcmc and covariance matrix adaptation schemes.

Müller, C. L., & Sbalzarini, I. F. (2010). Gaussian adaptation as a unifying framework for continuous black-box optimization and adaptive monte carlo sampling. *Congress on Evolutionary Computation (CEC, IEEE)*, 1–8.

Müller, C. L., & Sbalzarini, I. F. (2011). Gaussian adaptation for robust design centering. *Evolutionary and deterministic methods for design, optimization and control, Capua, Italy*, 736–742.

Suttorp, T., Hansen, N., & Igel, C. (2009). Efficient covariance matrix update for variable metric evolution strategies. *Machine Learning*, *75-2*, 167–197.