

Comparing Hamiltonian Monte Carlo and Elliptical Slice Sampling for constrained Gaussian distributions

732A76 Research Project Report

Bayu Brahmantio (baybr878)

2021-01-26

1. Background

High-dimensional multivariate gaussian distribution is used in various models and applications. In some cases, we need to generate from a certain distribution which applies constraints to a multivariate gaussian distribution (Gelfand et al. (1992) and Rodriguez-Yam et al. (2004)). Sampling from this distribution is still a challenging issue, particularly because it is not straightforward to compute the normalizing constant for the density function.

The gibbs sampler has proven to be a suitable choices to sample from truncated mutivariate gaussian distributions (Gelfand et al. (1992)). Recently, more sophisticated methods have been developed to generate samples from truncated mutivariate gaussian distributions. In this research project, two methods, namely Exact Hamiltonian Monte Carlo (Pakman et al. (2013)) and Analytic Elliptical Slice Sampling (Fagan et al. (2016)), will be compared.

2. Definitions

2.1. Truncated Multivariate Gaussian Distribution

The truncated multivariate gaussian distribution is a probability distribution obtained from multivariate gaussian random variable by bounding it under some linear (or quadratic) constraints.

Let \mathbf{W} be a d -dimensional gaussian random variable with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The corresponding truncated multivariate gaussian distribution can be defined as:

$$p(\mathbf{x}) = \frac{\exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}}{\int_{\mathbf{F}\mathbf{x} + \mathbf{g} \geq 0} \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\} d\mathbf{x}} I(\mathbf{F}\mathbf{x} + \mathbf{g} \geq 0)$$

Where X is a d -dimensional truncated gaussian random variable, I is an indicator function, and \mathbf{F} is an $m \times d$ matrix, which, together with the $m \times 1$ vector of \mathbf{g} , defines m constraints.

We can rewrite $f_{\mathbf{X}}$ as:

$$p(\mathbf{x}) = \frac{1}{Z} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \boldsymbol{\Lambda} \mathbf{x} + \boldsymbol{\nu}^T \mathbf{x} \right\} I(\mathbf{F}\mathbf{x} + \mathbf{g} \geq 0)$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$, $\boldsymbol{\nu} = \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, and Z is the normalizing constant.

2.2. Exact Hamiltonian Monte Carlo for Truncated Multivariate Gaussians

Exact Hamiltonian Monte Carlo (HMC) for Truncated Multivariate Gaussians (TMG) (Pakman et al. (2013)) considers the exact paths of particle trajectories that follow Hamiltonian equations of motion:

$$\begin{aligned}\frac{\partial x_i}{\partial t} &= \frac{\partial H}{\partial s_i} = s_i \\ \frac{\partial s_i}{\partial t} &= -\frac{\partial H}{\partial x_i} = -x_i\end{aligned}$$

To do this, let us sample from a standard normal distribution

$$p(\mathbf{x}) = \frac{1}{Z} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{x} \right\} I(\mathbf{F}\mathbf{x} + \mathbf{g} \geq 0)$$

which can be obtained from () through linear change of variables.

It takes the Hamiltonian function

$$H = \frac{1}{2} \mathbf{x}^T \mathbf{x} + \frac{1}{2} \mathbf{s}^T \mathbf{s}$$

in which the joint distribution is $P(\mathbf{X}, \mathbf{S}) = \exp(-H)$. With this form of H , we can derive the Hamilton's equations as

$$\begin{aligned}\frac{\partial x_i}{\partial t} &= \frac{\partial H}{\partial s_i} = s_i \\ \frac{\partial s_i}{\partial t} &= -\frac{\partial H}{\partial x_i} = -x_i\end{aligned}$$

for $i = 1, \dots, d$. The solution to those equations is

$$x_i(t) = s_i(0) \sin(t) + x_i(0) \cos(t)$$

which we can use as the trajectory of the particle along the state space.

To sample \mathbf{X} , we first sample values of \mathbf{S} from $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ and use them with previous values, \mathbf{x} , as the initial conditions. We will then move along this trajectory for a time T until we reach a new point \mathbf{x}^* . This sequence forms a markov chain that converges to the distribution $P(\mathbf{X})$ (Pakman et al. (2013)).

2.3. Analytic Elliptical Slice Sampling

Elliptical Slice Sampling (ESS, Murray et al. (2010)) aims to sample from posterior of the form

$$P^*(\mathbf{X}) = \frac{1}{Z} \mathcal{N}(\mathbf{X}; \mathbf{0}, \mathbf{\Sigma}) L(\mathbf{X})$$

where Z is the normalization constant, $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ is a multivariate gaussian prior, and L is a likelihood function. The algorithm begins by sampling values of $\boldsymbol{\nu}$ which, together with the values of current state \mathbf{x} , make an ellipse of possible new values

$$\mathbf{x}' = \mathbf{x} \cos(\theta) + \boldsymbol{\nu} \sin(\theta)$$

.

We then sample y uniformly between $[0, L(\mathbf{x})]$ to determine the height of the slice. To sample a new state from the ellipse, we sample values from the bracket $[0, 2\pi)$. A proposed point is accepted if it satisfies $L(\mathbf{x}') > y$. Otherwise, we shrink the bracket towards $\theta = 0$ until the proposed state is accepted.

Analytic Elliptical Slice Sampling (Fagan et al. (2016)) is an extension to ESS where the possible sampling bracket is pre-determined. To do this, we define

$$\mathcal{E} = \{\mathbf{x}' : \mathbf{x}'(\theta) = \mathbf{x} \cos(\theta) + \boldsymbol{\nu} \sin(\theta)\}$$

as the set of possible new states and define

$$\mathcal{S}(y, \mathcal{E}) = \{\mathbf{x}' \in \mathcal{E} : L(\mathbf{x}') > y\}$$

as the slice of acceptable proposed states. To form a markov chain that converges to the posterior distribution, we need to sample uniformly from the analytically pre-determined $\mathcal{S}(y, \mathcal{E})$ (Fagan et al. (2016)).

3. Results

In this section we are going to compare both methods (Exact HMC for TMG and Analytic ESS) under different settings of truncated multivariate gaussians. For both of them, we use linear change of variables to transform the distribution into a whitened frame such that the generated points are from $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$.

Figure 1 shows how both algorithms sample a two-dimensional gaussian distribution with some constraints that results in a narrow support, similar to Figure 1 in Pakman et al. (2013). From both cases, we can see that they worked as expected. The generated samples rapidly converged to the mean while also complying the constraints. The autocorrelation function plots also show that the samples are not highly correlated. However, the Exact HMC tends to be spread more across the support than Analytic ESS.

Notice that the maximum traveling time (T) and number of samples per iteration (J) affect the computation time for Exact HMC and Analytic ESS respectively. The higher the maximum traveling time (T), the longer it takes for a particle in an HMC system to move while also increasing the possibility of bouncing. At the same time, an increase in the value of J lets us sample multiple points from the same ellipse without a significant increase in computation time.

From the experiments, we observed a large disparity between the runtimes of both algorithms. As seen in the Table 1, Exact HMC performs much faster than the fastest Analytic ESS.

Table 1: Table 1: Average computation time per sample for both models under different settings for 10-dimensional TMG under the constraints $-1 \leq x_i \leq 1, i = 1, \dots, 10$ using CPU. For HMC, T is the maximum travel time. For ESS, J is the number of points sampled in each iteration. The computation time for Exact HMC is overall much faster than Analytic ESS.

| Model | Avg. computation time per sample (second) |
|--------------------|---|
| HMC($T=\pi/2$) | 0.0001728 |
| HMC($T=3*\pi/4$) | 0.000173 |
| ESS($J=1$) | 0.0271275 |
| ESS($J=5$) | 0.0250084 |
| ESS($J=10$) | 0.0236001 |

4. Discussion

In this research project, we have compared Exact HMC and Analytic ESS. In terms of the generated samples, they are not significantly different. Exact HMC also performs much faster than Analytic ESS. More rigorous measurements can be made to evaluate whether the samples are autocorrelated within a chain or not by

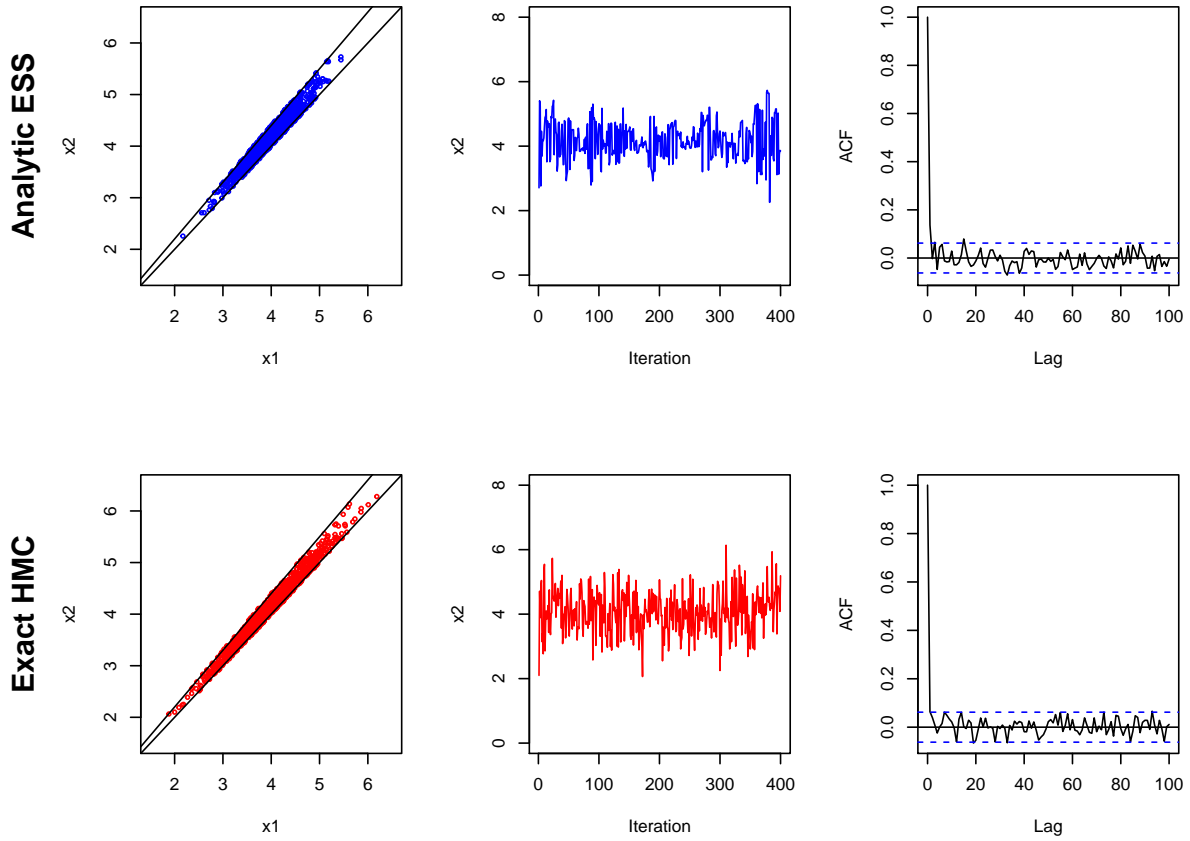


Figure 1: Comparison of generated points from $\mathcal{N}((4,4)^T, \mathbf{I}_2)$ under the constraints $X_1 \leq X_2 \leq 1.1X_1$ and $X_1, X_2 \geq 0$ with initial points $(2, 2.1)$. First column: 1000 iterations. Second column: trace plots of the first 400 iterations of X_2 . Third column: Autocorrelation of X_2 . In this example we used $T = \pi/2$ for HMC and $J = 1$ for ESS. Both methods converges around $X_2 = 4$ and have relatively low autocorrelation scores but HMC appears to explore the possible space more.

measuring their effective sample size.

There are, however, some caveats to the computation of Analytic ESS in this case. We expected the runtime to be somewhat similar to Exact HMC, but it runs more than 100 times slower. This could be to the fact that external packages are used for the computation of the intervals and slices which could have been replaced by customized functions that are more specific to the task, hence the faster runtimes. The programming language used (R) is also not popularly chosen because of its speed. Thus, the algorithm might need to be rewritten in a language that is proven to be computationally faster.

5. References

- Fagan, F., Bhandari, J., and Cunningham, J. P. (2016), “Elliptical slice sampling with expectation propagation,” in *Proceedings of the thirty-second conference on uncertainty in artificial intelligence*, UAI’16, Arlington, Virginia, USA: AUAI Press, pp. 172–181.
- Gelfand, A. E., Smith, A. F. M., and Lee, T.-M. (1992), “Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling,” *Journal of the American Statistical Association*, [American Statistical Association, Taylor & Francis, Ltd.], 87, 523–532.
- Murray, I., Adams, R. P., and MacKay, D. J. C. (2010), “Elliptical slice sampling,” *JMLR: W&CP*, 9, 541–548.
- Pakman, A., and Paninski, L. (2013), “Exact hamiltonian monte carlo for truncated multivariate gaussians.”
- Rodríguez-Yam, G. A., Davis, R. I. A., and Scharf, L. (2004), “Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression.”

6. Appendix

All codes are available on <https://github.com/bayubeta/research-project>.