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

732A76 Research Project Report

Bayu Brahmantio (baybr878)

March 1, 2021

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 Rodríguez-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 multivariate Gaussian distributions (Gelfand et al. [1992]). Recently, more sophisticated methods have been developed to generate samples from truncated multivariate Gaussian distributions. In this research project, two methods, namely Exact Hamiltonian Monte Carlo (Pakman and Paninski [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 a 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

$$p(\mathbf{x}) = \frac{\exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}}{\int_{\mathbf{Fx+g} \geq 0} \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}d\mathbf{x}} \mathbb{1}(\mathbf{Fx} + \mathbf{g} \geq 0) \qquad (2.1)$$

1

where $\mathbf{x}$ is a d-dimensional truncated Gaussian random variable, $\mathbb{1}$ 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 all m constraints of $p(\mathbf{x})$. We denote this as $\mathbf{x} \sim TN(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{F}, \mathbf{g})$.

We can rewrite $p(\mathbf{x})$ as

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left\{ -\frac{1}{2}\mathbf{x}^\mathsf{T}\boldsymbol{\Lambda}\mathbf{x} + \boldsymbol{\nu}^\mathsf{T}\mathbf{x} \right\} \mathbb{1}(\mathbf{F}\mathbf{x} + \mathbf{g} \geq 0) \tag{2.2}$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$, $\boldsymbol{\nu} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, and Z is the normalizing constant. Through linear change of variables, (2.2) can be transformed into

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left\{ -\frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{x} \right\} \mathbb{1}(\mathbf{F}^*\mathbf{x} + \mathbf{g}^* \geq 0) \tag{2.3}$$

such that $\mathbf{x} \sim TN(\mathbf{0}, \mathbf{I}_d, \mathbf{F}^*, \mathbf{g}^*)$, for some values of $\mathbf{F}^*$ and $\mathbf{g}^*$.

2.2 Exact Hamiltonian Monte Carlo for Truncated Multivariate Gaussians

2.2.1 Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) is a Markov Chain Monte Carlo method that makes use of an ideal system that follow Hamiltonian dynamics. A particle in this system can be described by a Hamiltonian energy function

$$H(\mathbf{x}, \mathbf{s}) = U(\mathbf{x}) + K(\mathbf{s}) \tag{2.4}$$

where $U(\mathbf{x})$ is the potential energy term as a function of particle's position ($\mathbf{x}$) and $K(\mathbf{s})$ is the kinetic energy term as a function of particle's momentum ($\mathbf{s}$). Both $\mathbf{x}$ and $\mathbf{s}$ are of d-dimensions. The change of position and momentum over time t can be described by Hamilton's equations

$$\begin{aligned} \frac{\partial x_i}{\partial t} &= \frac{\partial H}{\partial s_i} \\ \frac{\partial s_i}{\partial t} &= -\frac{\partial H}{\partial x_i}, \qquad i = 1, ..., d. \end{aligned} \tag{2.5}$$

To sample from a distribution using HMC, we can relate the target distribution to the current energy state of the particle through canonical distribution:

$$p(\mathbf{x}) \propto \exp\{-E(\mathbf{x})\} \tag{2.6}$$

where the target distribution, $p(\mathbf{x})$, depends on the value of energy function $E(\mathbf{x})$. In a Hamiltonian system, we have $H(\mathbf{x}, \mathbf{s})$ as our energy function, which results in the canonical distribution:

$$\begin{aligned} p(\mathbf{x}, \mathbf{s}) &\propto \exp\{-H(\mathbf{x}, \mathbf{s})\} \\ &\propto \exp\{-U(\mathbf{x})\} \exp\{-K(\mathbf{s})\} \\ &\propto p(\mathbf{x})p(\mathbf{s}). \end{aligned} \tag{2.7}$$

Hence, $\mathbf{x}$ and $\mathbf{s}$ are independent. To sample from the target distribution $p(\mathbf{x})$, we can sample from the joint distribution $p(\mathbf{x}, \mathbf{s})$ and ignore the variable $\mathbf{s}$.

2.2.2 Exact HMC for Truncated Multivariate Gaussians

Exact Hamiltonian Monte Carlo (HMC) for Truncated Multivariate Gaussians (TMG) (Pakman and Paninski [2013]) considers the exact paths of particle trajectories in a Hamiltonian system that is proportional to the negative logarithm of TMG distributions density. Suppose our target distribution $p(\mathbf{x})$ is a truncated multivariate Gaussian distribution as in (2.3). We can set our momenta to be normally distributed, that is $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. Therefore, the Hamiltonian system can be described as:

$$H = U(\mathbf{x}) + K(\mathbf{s}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{x} + \frac{1}{2}\mathbf{s}^\mathsf{T}\mathbf{s} \tag{2.8}$$

subject to:

$$\mathbf{F}\mathbf{x} + \mathbf{g} \geq 0. \tag{2.9}$$

for some values of $\mathbf{F}$ and $\mathbf{g}$.

The equations of motion for the Hamiltonian system in (2.8) are:

$$\begin{aligned}
\frac{\partial x_i}{\partial t} &= \frac{\partial H}{\partial s_i} = s_i \\
\frac{\partial s_i}{\partial t} &= -\frac{\partial H}{\partial x_i} - x_i, \qquad i = 1, ..., d.
\end{aligned} \tag{2.10}$$

In this sense, we want the particles in Hamiltonian system to only move around inside the constrained space. The exact trajectory of a particle using the equations above is:

$$x_i(t) = s_i(0)\sin(t) + x_i(0)\cos(t). \tag{2.11}$$

A particle will follow the trajectory above until it hits a wall, or in other words, until $\mathbf{F}\mathbf{x} + \mathbf{g} = 0$. Let t_h be the time when the particle hits wall h, or when $\mathbf{F}_h \cdot \mathbf{x}(t_h) + \mathbf{g}_h = 0$. It will hit the wall with velocity $\dot{\mathbf{x}}(t_h)$ which can be decomposed into:

$$\dot{\mathbf{x}}(t_h) = proj_{\mathbf{n}}\dot{\mathbf{x}}(t_h) + proj_{\mathbf{F}_h}\dot{\mathbf{x}}(t_h) \tag{2.12}$$

where $proj_{\mathbf{n}}\dot{\mathbf{x}}(t_h)$ is the projection of $\dot{\mathbf{x}}(t_h)$ on the normal vector $\mathbf{n}$ perpendicular to $\mathbf{F}_h$ and

$$\begin{aligned}
proj_{\mathbf{F}_h}\dot{\mathbf{x}}(t_h) &= \frac{\mathbf{F}_h \cdot \dot{\mathbf{x}}(t_h)}{||\mathbf{F}_h||}\frac{\mathbf{F}_h}{||\mathbf{F}_h||} \\
&= \frac{\mathbf{F}_h \cdot \dot{\mathbf{x}}(t_h)}{||\mathbf{F}_h||^2}\mathbf{F}_h \\
&= \alpha_h\mathbf{F}_h.
\end{aligned} \tag{2.13}$$

By inverting the direction of $proj_{\mathbf{n}}\dot{\mathbf{x}}(t_h)$, we can obtain the reflected velocity as

$$\begin{aligned}
\dot{\mathbf{x}}_R(t_h) &= -proj_{\mathbf{n}}\dot{\mathbf{x}}(t_h) + proj_{\mathbf{F}_h}\dot{\mathbf{x}}(t_h) \\
&= -\dot{\mathbf{x}}(t_h) + 2\alpha_h\mathbf{F}_h
\end{aligned} \tag{2.14}$$

which can be used as the new initial velocity $(s_i(0))$ in (2.11) for the particle to continue its path. It can be shown that exact HMC follows the detailed balanced condition (Pakman and Paninski [2013]) which leaves the canonical distribution invariant.

The process of sampling using exact HMC can be summarized into two steps:

1. Sample $\mathbf{s}$ from $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$.

2. Use $\mathbf{s}$ and the last value of $\mathbf{x}$, and move the particle for a time T until we reach $\mathbf{x}^*$ and $\mathbf{s}^*$. Use $\mathbf{x}^*$ as initial positions for the next iteration.

2.3 Analytic Elliptical Slice Sampling

2.3.1 Slice Sampling

Slice sampling (Neal [2003]) is a method that tries to sample uniformly from the $(n+1)$-dimensional region under $f(x)$ that is proportional to density function $p(x)$. It is done by using Gibbs sampling to sample from the joint distribution (x, y) where y is an auxiliary variable. We do this by alternately sampling from $y|x \sim \mathrm{U}(0, f(x))$ and $x|y \sim \mathrm{U}(S)$ where $S = \{x : f(x) > y\}$. To obtain samples from x, we can sample from $p(x, y)$ and ignore the y later.

2.3.2 Elliptical Slice Sampling

Elliptical slice sampling (ESS, Murray et al. [2010]) is a form of slice sampling that uses Gaussian prior to sample from a posterior distribution of the form

$$p^*(\mathbf{x}) = \frac{1}{Z}\mathcal{N}(\mathbf{x}; \mathbf{0}, \boldsymbol{\Sigma})L(\mathbf{x}) \tag{2.15}$$

where Z is the normalization constant, $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is a multivariate Gaussian prior, and L is a likelihood function. To do this, we introduce an auxiliary variable $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ so that given an initial state $\mathbf{x}$, the update ellipse is

$$\mathbf{x}' = \mathbf{x}\cos(\theta) + \boldsymbol{\nu}\sin(\theta), \quad \theta \in [0, 2\pi] \tag{2.16}$$

The algorithm begins by sampling the slice height $y \sim \mathrm{U}(0, L(\mathbf{x}))$ and the auxiliary variable $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ that determines the ellipse in (2.16). We then sample θ uniformly from $[0, 2\pi]$ and shrink the interval towards $\theta = 0$ (initial state) until we found θ that satisfies $L(\mathbf{x}') > y$. The new point is accepted as the next state and we use it as the initial state for the next iteration.

ESS is analogous to slice sampling in the sense that it samples uniformly from the height of function and from the interval where the function value is more than the height. Furthermore, it can be shown that the transition operator of ESS is reversible which implies that the Markov chain converges to the target posterior distribution (Murray et al. [2010]).

2.3.3 Analytic Elliptical Slice Sampling

Analytic Elliptical Slice Sampling (Fagan et al. [2016]) tries to cut the process of shrinking the ellipse in ESS by pre-determining the possible sampling interval. First, we define

$$\mathcal{E} = \{\mathbf{x}' : \mathbf{x}'(\theta) = \mathbf{x}\cos(\theta) + \boldsymbol{\nu}\sin(\theta)\} \tag{2.17}$$

as the set of possible new states and define

$$\mathcal{S}(y, \mathcal{E}) = \{\mathbf{x}' \in \mathcal{E} : L(\mathbf{x}') > y\} \tag{2.18}$$

as the slice of acceptable proposed state. To form a markov chain that converges to the posterior distribution, we need to sample uniformly from the analytically pre-determined slice $\mathcal{S}(y, \mathcal{E})$ (Fagan et al. [2016]). That is, the algorithm is similar to ESS except that the process of shrinking the ellipse is replaced by determining $\mathcal{S}(y, \mathcal{E})$ so that sampling from θ guarantees to yield a new state $\mathbf{x}'$.

In a truncated multivariate Gaussian setting, we have

$$\begin{aligned} p^*(\mathbf{x}) &= \frac{1}{Z}\mathcal{N}(\mathbf{x}; \mathbf{0}, \boldsymbol{\Sigma})TN(\mathbf{x}; \mathbf{0}, \mathbf{I}_d, \mathbf{F}, \mathbf{g}) \\ &\propto \mathcal{N}(\mathbf{x}; \mathbf{0}, \boldsymbol{\Sigma})\mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{I}_d)\mathbb{1}(\mathbf{Fx} + \mathbf{g} \geq 0) \end{aligned} \tag{2.19}$$

Given $\mathbf{x}$, $\boldsymbol{\nu}$, and y, we can determine the slice of possible new states as

$$\begin{aligned} \mathcal{S}(y, \mathcal{E}) &= \{\theta \in [0, 2\pi] : L(\mathbf{x}') > y\} \\ &= \cap_{j=1}^m \{\theta \in [0, 2\pi] : F_j^\mathsf{T}\mathbf{x} + g_j \geq 0\} \\ &\quad \cap \{\theta \in [0, 2\pi] : \mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{I}_d) > y\}. \end{aligned} \tag{2.20}$$

Fagan et al. [2016] also demonstrates how to sample J points from a single ESS iteration. In an ESS setting, it is done by 'recycling' the interval that has been sampled until J points are accepted. One of the J points is then randomly selected as the next initial state. In the case of analytic ESS, J different values of y are sampled and for each of them we find the corresponding $\mathcal{S}(y, \mathcal{E})$ and uniformly sample one point.

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.

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 and Paninski [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.

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 increases 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.
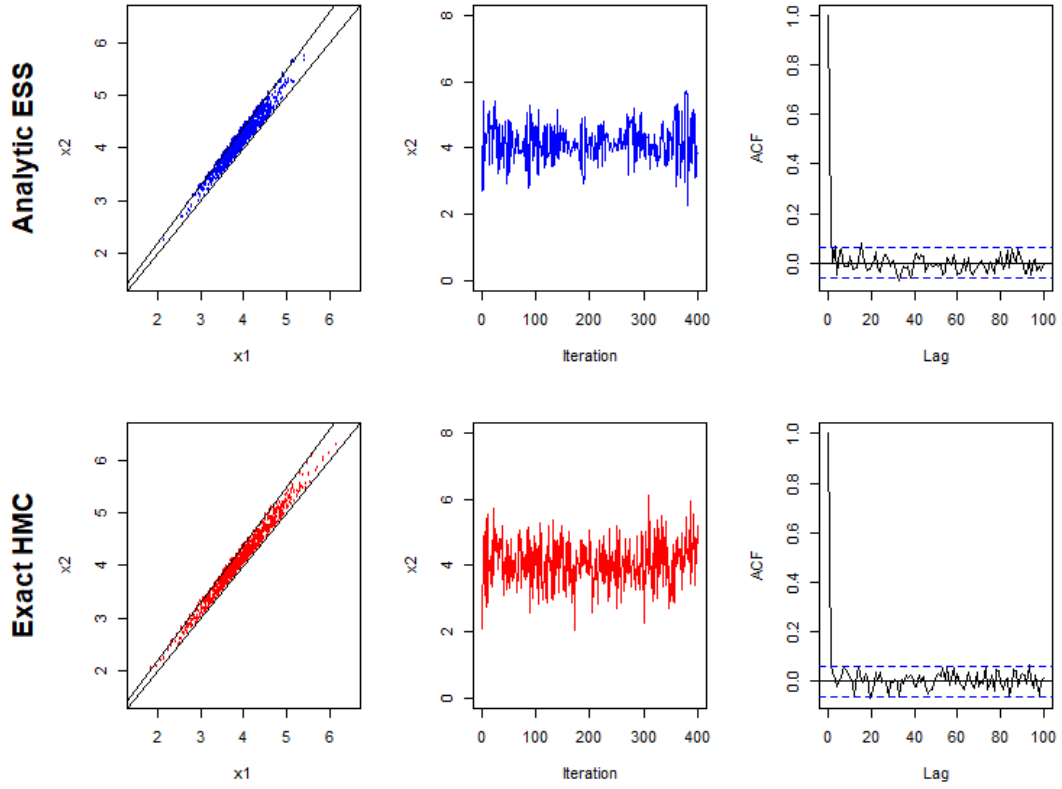
Figure 1: Comparison of generated points from $\mathcal{N}((4,4)^T, \mathbf{I}_2)$ under the constraints $x_1 \leq x_2 \leq 1.1x_1$ and $x_1, x_2 \geq 0$ with initial points (2, 2.1). First column: 1000 iterations. Second column: trace plots of the first 400 iterations of x_2. Third column: Autocorrelation 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.

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 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 runtime. The programming

Method	Avg. computation time per sample (second)
HMC($T = \frac{\pi}{2}$)	0.0001728
HMC($T = \frac{3\pi}{4}$)	0.000173
ESS($J = 1$)	0.0271275
ESS($J = 5$)	0.0250084
ESS($J = 10$)	0.0236001

Table 1: Average computation time per sample for both models under different settings for 10-dimensional TMG under the constraints $-1 \leq x_i \leq 1, i = 1, ..., 10$ 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.

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 (e.g. `Python`).

References

Alan E. Gelfand, Adrian F. M. Smith, and Tai-Ming Lee. Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association*, 87(418):523–532, 1992. ISSN 01621459. URL http://www.jstor.org/stable/2290286.

Gabriel A. Rodríguez-Yam, Richard I. A. Davis, and L. Scharf. Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression. 2004.

Ari Pakman and Liam Paninski. Exact hamiltonian monte carlo for truncated multivariate gaussians, 2013.

Francois Fagan, Jalaj Bhandari, and John P. Cunningham. Elliptical slice sampling with expectation propagation. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'16, page 172–181, Arlington, Virginia, USA, 2016. AUAI Press. ISBN 9780996643115.

Iain Murray, Ryan Prescott Adams, and David J. C. MacKay. Elliptical slice sampling. 9:541–548, 2010.

Radford M. Neal. Slice sampling. *The Annals of Statistics*, 31(3):705–741, 2003. ISSN 00905364. URL http://www.jstor.org/stable/3448413.