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

732A76 Research Project Report

Bayu Brahmantio (baybr878)

February 8, 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

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

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

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

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

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

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

The target distribution is related 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 $\mathbf{s}$.

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:

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

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 $\mathbf{F}_h \cdot \mathbf{x}(t_h) + \mathrm{g}_h = 0$. 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 velocity

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.