Received: 25 September 2022 | Revised: 4 January 2023 | Accepted: 27 January 2023

DOI: 10.1002/wics.1608

ADVANCED REVIEW

WIREs COMPUTATIONAL STATISTICS WILEY

Sampling constrained continuous probability distributions: A review

Shiwei Lan[1] | **Lulu Kang**[2]

[1]School of Mathematical and Statistical Sciences, Arizona State University, Tempe, Arizona, USA

[2]Department of Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois, USA

Correspondence
Lulu Kang, Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616, USA.
Email: lkang2@iit.edu

Funding information
NSF, Division of Mathematical Sciences, Grant/Award Numbers: 1916467, 2134256, 2153029

Edited by: Henry Lu, Commissioning Editor and David Scott, Review Editor and Co-Editor-in-Chief

Abstract

The problem of sampling constrained continuous distributions has frequently appeared in many machine/statistical learning models. Many Markov Chain Monte Carlo (MCMC) sampling methods have been adapted to handle different types of constraints on random variables. Among these methods, Hamilton Monte Carlo (HMC) and the related approaches have shown significant advantages in terms of computational efficiency compared with other counterparts. In this article, we first review HMC and some extended sampling methods, and then we concretely explain three constrained HMC-based sampling methods, reflection, reformulation, and spherical HMC. For illustration, we apply these methods to solve three well-known constrained sampling problems, truncated multivariate normal distributions, Bayesian regularized regression, and nonparametric density estimation. In this review, we also connect constrained sampling with another similar problem in the statistical design of experiments with constrained design space.

This article is categorized under:

Applications of Computational Statistics > Computational Mathematics

Statistical and Graphical Methods of Data Analysis > Bayesian Methods and Theory

Statistical and Graphical Methods of Data Analysis > Markov Chain Monte Carlo (MCMC)

Statistical and Graphical Methods of Data Analysis > Sampling

KEYWORDS

constrained sampling, Hamilton Monte Carlo, regularized regression, Riemannian Monte Carlo, truncated multivariate Gaussian

1 | INTRODUCTION

In many machine learning applications, it is necessary to sample from distributions with various types of constraints. For example, the truncated multivariate normal distribution can be difficult to sample from, especially for high-dimensional cases. Even using the leave-one-out type of Gibbs sampling scheme (Held & Holmes, 2006; Kang et al., 2021), the algorithms can still be computationally costly. Another common example is the regression model with norm constraint on the parameters, $\|\boldsymbol{\beta}\|_q \leq C$, such as Lasso ($l_1$ norm, $q = 1$) (Tibshirani, 1996) or bridge estimator (l_q norm, $q \geq 0$) (Frank & Friedman, 1993; Fu, 1998). Other examples include copula models, latent Dirichlet allocation,

covariance matrix estimation, and nonparametric density function estimation. Often, the resulting models are intractable, and thus sampling from these constrained distributions is a challenging task (Brubaker et al., 2012; Neal et al., 2012; Neal & Roberts, 2008; Pakman & Paninski, 2014; Sherlock & Roberts, 2009). In this article, we give an overview of the statistical sampling methods for constrained distributions. Specifically, we focus on distributions of continuous variables and provide a more detailed explanation of methods based on Hamiltonian Monte Carlo (HMC) and some related approaches. Sampling constrained discrete distributions (Chewi et al., 2022; Jacob et al., 2021; Jessen, 1970) is not included in this review due to their different nature from continuous distributions.

There are various types of boundary constraints on the parameters of many statistical models, such as positive requirement, linear (summation) constraint, upper bound on a vector norm of the parameters, and so on. Sometimes, the constraints can be considered as certain general manifolds, for example, sphere, positive definite matrices, and Stiefel manifold. Specifically, the norm constraint imposed on the regression coefficients in regularized regression can be considered a sphere manifold (Lan et al., 2014; Lan & Shahbaba, 2016). Manifold methods are also used for sampling from a positive definite matrix in Lan et al. (2020) and Holbrook et al. (2018). Many sampling approaches have been introduced to tackle one or several kinds of constraints. Based on the nature of these methods, we roughly categorize them into three groups.

1. *Rejection Type*. These methods simply discard samples that violate the constraints or keep trying until the proposed sample satisfies the constraint. Most Markov Chain Monte Carlo (MCMC) or other statistical sampling algorithms can be easily modified to achieve this goal. For example, Lang et al. (2007) proposed a rejection-based sequential Monte Carlo for Bayesian estimation of constrained dynamic systems, and Li and Ghosh (2015) also developed methods based on rejection for truncated multivariate normal and student-t distributions subject to linear inequality constraints. Since these approaches do not directly address the constraints, they can be computationally inefficient for complicated constraints and high-dimensional problems.
2. *Reflection Type*. These approaches consider the boundary of the constrained domain as an (energy) wall and make a reflection (hit-and-bounce) for the sampler to move inward the constrained domain whenever it hits the (energy) wall. For example, Neal (2011) suggested modifying the standard HMC by setting the potential energy to infinity for parameter values that violate the constraints. Following this idea, Pakman and Paninski (2014) proposed an exact HMC for truncated multivariate Gaussian distributions. Betancourt (2011) and Olander (2020) applied such an HMC-based reflection idea to nested sampling which requires likelihood-restricted prior sampling (Skilling, 2006).
3. *Reformulation Type*. These techniques transform the constrained sampling problem or the constrained domain into something easier to work with. Motivated by the constrained optimization methods, Brubaker et al. (2012) proposed a family of HMC-based MCMC methods, which incorporated the constraint on parameters $c(\boldsymbol{\theta}) = 0$ using Lagrange multipliers. Ahn and Chewi (2021) recently derived another optimization-motivated algorithm using mirror-Langevin dynamics. In many cases, distributions with constraints in $\mathbb{R}^d$ can be transformed into distributions on manifolds. Some HMC-related algorithms have been created to sample distributions on manifolds. For example, Kook et al. (2022) used Riemannian manifold HMC instead of the original HMC with a Lagrange multiplier. Byrne and Girolami (2013) showed how HMC methods can be designed for and applied to the distribution defined on manifolds embedded in Euclidean space by the explicit forms for geodesics if they exist. In particular, motivated by Byrne and Girolami (2013), Spherical HMC (Lan et al., 2014; Lan & Shahbaba, 2016) does not require any manifold embedding. It focuses on constraints that can be transformed into vector norms and eventually mapped onto a hypersphere. Spherical HMC can be viewed as a more efficient special case of Brubaker et al. (2012) and Kook et al. (2022). A related work SPInS (Chaudhry et al., 2021) maps a hyper-ball containing the constrained domain inside out so that sampler can be defined on a larger unbounded space.

In general, algorithms of the rejection type tend to be inefficient because frequent rejected attempts cause a significant waste of computation. The reflection type is intuitive but its efficiency usually depends on the specific constraints. Certain constrained domains may need too many reflections in high-dimensional space and hence the excessive computational time. The reformulation type is more intrinsic and sophisticated in incorporating the constraints in the step of proposing new samples. More importantly, they usually can be easily scaled to large dimensions and handle more varieties of constraints. Therefore, we are going to focus on the reflection and reformulation types of methods in the following review.

In recent years, optimal transport (OT) and the more general variational inference methods have gained much attention from the machine learning community and have been adapted to many statistical and machine learning models. Although in theory variational inference methods, including OT, can deal with distributions with any compact support

regions, in numerical implementation it is much more challenging when the support region is not the entire $\mathbb{R}^d$ with d being the dimension of the variables. Das and Bhattacharya (2020) addressed the issue of state-dependent nonlinear equality-constrained state estimation using Bayesian filtering based on OT. Ahn and Chewi (2021) proposed using the mirror-Langevin algorithm, which is a discretization of the mirror-Langevin diffusion, for constrained sampling.

Constrained sampling is naturally connected with numerical integration in constrained domains. Quasi-Monte Carlo methods deal with numerical integration over rectangular-constrained domains (Leobacher & Pillichshammer, 2014; Owen, 2013). Other quadrature methods have to be modified for non-rectangular constraints (Gessner et al., 2020; Legrain, 2021; Olshanskii & Safin, 2016; Saye, 2022). Constrained sampling is also related to the statistical design of experiments. Draguljić et al. (2012), Pratola et al. (2017), and Huang et al. (2021) proposed different methods to generate space-filling designs, which essentially approximate the uniform distribution, in various irregular-constrained domains. Kang (2019) developed algorithms to generate different optimal designs, including a distance-criterion-based space-filling design, in complicated design regions.

In the remaining article, we first review the necessary background on HMC and its related methods in Section 2. Next, we review in detail three reflection and reformulation approaches in Section 3. Three constrained sampling problems are illustrated in Section 4, including truncated multivariate Gaussian distributions, Bayesian regularized regression, and non-parametric density estimation. In Section 5, we connect constrained sampling with the constrained design of experiments and review some existing methods for constructing different types of constrained designs. The article concludes in Section 6.

2 | HAMILTONIAN MONTE CARLO AND EXTENSIONS

MCMC can be inherently inefficient due to its random walk nature. Different from the Gibbs sampler and Metropolis algorithms, HMC simulates Hamiltonian dynamics to propose new states and reduce the local random walk behavior, and thus moves more rapidly toward the target distribution. The proposed states of HMC are significantly distant from the current states and yet still have a high acceptance probability. Neal (2011) recalled the origin and history of HMC and detailed the HMC method and its appealing properties. Here we briefly review the HMC algorithm and its manifold extensions such as the Riemannian/Lagrangian Monte Carlos, geodesic Monte Carlo, and so on.

2.1 | Hamiltonian Monte Carlo

We begin with an introduction to Hamiltonian dynamics, which is the basis of HMC. It consists of a d-dimensional vector $\boldsymbol{\theta}$, called *position* state, and a d-dimensional vector ϕ, called *momentum* state. For illustration, Neal (2011) used a simple physical example, the dynamics of a frictionless puck that slides over a surface of varying height. The *potential energy* of the puck, denoted by $U(\boldsymbol{\theta})$, is proportional to the height given position $\boldsymbol{\theta}$. The *kinetic energy* of the puck, denoted by $K(\boldsymbol{\phi})$, is equal to $|\boldsymbol{\phi}|^2/m$, with m the *mass* of the puck. The total energy of the dynamic, called *Hamiltonian function* and denoted by $H(\boldsymbol{\theta},\boldsymbol{\phi})$, is the sum of the potential energy and kinetic energy of the puck, that is,

$$H(\boldsymbol{\theta},\boldsymbol{\phi}) = U(\boldsymbol{\theta}) + K(\boldsymbol{\phi}). \tag{1}$$

As the puck slides over the surface, the potential energy increases as it moves over a rising slope and the kinetic energy decreases as the velocity of the puck, $\boldsymbol{\phi}/m$, decreases. The two energies change in opposite directions when the puck moves over a descending slope. Due to frictionless assumption, the total energy remains the same as the initial state of the system. The system of $(\boldsymbol{\theta},\boldsymbol{\phi})$ evolves following the *Hamilton's equations*

$$\bar{\boldsymbol{\theta}} = \frac{d\boldsymbol{\theta}}{dt} = \frac{\partial H(\boldsymbol{\theta},\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} = \nabla_{\boldsymbol{\phi}} K(\boldsymbol{\phi}), \tag{2}$$

$$\bar{\boldsymbol{\phi}} = \frac{d\boldsymbol{\phi}}{dt} = -\frac{\partial H(\boldsymbol{\theta},\boldsymbol{\phi})}{\partial \boldsymbol{\theta}} = -\nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta}). \tag{3}$$

HMC applies Hamiltonian dynamics to MCMC sampling. The position vector $\boldsymbol{\theta}$ is the random variable of interest. The potential energy is $U(\boldsymbol{\theta}) = -\log p(\boldsymbol{\theta})$, where $p(\boldsymbol{\theta})$ is the density function of the target distribution to be sampled from. In the Bayesian framework, $p(\boldsymbol{\theta})$ is the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$, where $\boldsymbol{y}$ represents data and $\boldsymbol{\theta}$ the unknown parameters. The momentum vector $\boldsymbol{\phi}$ can be considered as an auxiliary random variable that is usually assumed to follow a multivariate normal distribution $\mathcal{N}(0,\boldsymbol{M})$, where $\boldsymbol{M}$ is a user-specified covariance matrix (often chosen to be identity matrix $\boldsymbol{I}$), also known as a *mass matrix*. Thus the kinetic energy becomes $K(\boldsymbol{\phi}) = -\log N(\boldsymbol{\phi}\,|\,0,\boldsymbol{M}) = \boldsymbol{\phi}^T\boldsymbol{M}^{-1}\boldsymbol{\phi}/2 + \text{constant}$. To numerically solve (2) and (3) with the above potential and kinetic energies, the *leapfrog* method (Verlet, 1967) is commonly used to approximate the Hamilton's equations by discretizing time. The HMC uses the leapfrog method to simulate Hamiltonian dynamics for some time horizon $\tau = L\epsilon$ to propose new samples that are further accepted or rejected according to certain probability as the next state. Algorithm 1 shows the HMC procedure.

Following either the Hamilton's equations (2) and (3) or the leapfrog procedure, the intuition of HMC is straightforward as explained in Gelman et al. (2013). When the current value of $\boldsymbol{\theta}$ is at a flat region of $p(\boldsymbol{\theta})$, similar to the situation when the puck is on a flat surface, the velocity of the puck, and thus its momentum $\boldsymbol{\phi}$, becomes close to a constant. Therefore, the position $\boldsymbol{\theta}$ would move at a constant speed exploring the flat region. If the position $\boldsymbol{\theta}$ moves to a region with decreasing density of $p(\boldsymbol{\theta})$, which is not favorable, then $\nabla_{\boldsymbol{\theta}}\log p(\boldsymbol{\theta})$ is negative, and thus the momentum $\boldsymbol{\phi}$ would decrease in the direction of movement. Next, the position $\boldsymbol{\theta}$ would move in this unfavorable direction with a reduced velocity. The trends reverse if $\boldsymbol{\theta}$ value moves to a region with an increasing density of $p(\boldsymbol{\theta})$. The system possesses three important properties for the proof of ergodicity: (i) time-reversibility (going from the end of a trajectory with the reversed momentum takes the sampler back to the starting point); (ii) volume-preservation (volume moving along the flow $T_t : (\boldsymbol{\theta},\boldsymbol{\phi}) \mapsto (\boldsymbol{\theta}^*,\boldsymbol{\phi}^*)$ does not change); and (iii) energy conservation (approximately under discretized system). The exact proof of convergence of HMC can be found in Neal (1994) and Neal (2011). The choice of ϵ should be sufficiently small so that the acceptance rate is high but not too small so that the computation is still efficient. The length of trajectory is also a crucial parameter for HMC and it can be varied from 20 to 1000, depending on the complexity and dimension of the problem. Trial and error can be used for setting both ϵ and L. More discussion can be found in Neal (2011).

2.2 | Riemannian and Lagrangian Monte Carlo

Girolami and Calderhead (2011) extended HMC to Riemannian HMC (RHMC) by defining Hamiltonian dynamics on a Riemannian manifold of distributions. Compared with HMC, RHMC can exploit fully the geometric properties of the parameter space of $\boldsymbol{\theta}$ using a position-specific mass matrix, that is, $\boldsymbol{M} = \boldsymbol{G}(\boldsymbol{\theta})$, where $\boldsymbol{G}(\boldsymbol{\theta})$ is usually set as the Fisher information matrix of $p(\boldsymbol{\theta})$. The distribution of momentum vector ϕ is $\mathcal{N}(\boldsymbol{0},\boldsymbol{G}(\boldsymbol{\theta}))$, which is no longer independent of $\boldsymbol{\theta}$. The Hamiltonian function (1) becomes

$$H(\boldsymbol{\theta},\boldsymbol{\phi}) = \psi(\boldsymbol{\theta}) + K(\boldsymbol{\theta},\boldsymbol{\phi}) = -\log p(\boldsymbol{\theta}) + \frac{1}{2}\log\det(\boldsymbol{G}(\boldsymbol{\theta})) + \frac{1}{2}\boldsymbol{\phi}^T\boldsymbol{G}(\boldsymbol{\theta})^{-1}\boldsymbol{\phi}. \tag{4}$$

ALGORITHM 1 Hamiltonian Monte Carlo (HMC)

1: Initialize $\boldsymbol{\theta}^{(0)}$ at current $\boldsymbol{\theta}_{t-1}$, and randomly sample $\boldsymbol{\phi}^{(0)} \sim \mathcal{N}(0,\boldsymbol{M})$
2: **for** $\ell = 0$ to $L-1$ **do**
3: Update $\boldsymbol{\phi}$ by a half-step of ϵ: $\boldsymbol{\phi}^{\left(\ell+\frac{1}{2}\right)} = \boldsymbol{\phi}^{(\ell)} + \frac{1}{2}\epsilon\nabla_{\boldsymbol{\theta}}\log p\left(\boldsymbol{\theta}^{(\ell)}\right)$.
4: Update $\boldsymbol{\theta}$ by a full-step of ϵ: $\boldsymbol{\theta}^{(\ell+1)} = \boldsymbol{\theta}^{(\ell)} + \epsilon\boldsymbol{M}^{-1}\boldsymbol{\phi}^{\left(\ell+\frac{1}{2}\right)}$.
5: Update $\boldsymbol{\phi}$ by another half-step of ϵ: $\boldsymbol{\phi}^{(\ell+1)} = \boldsymbol{\phi}^{\left(\ell+\frac{1}{2}\right)} + \frac{1}{2}\epsilon\nabla_{\boldsymbol{\theta}}\log p\left(\boldsymbol{\theta}^{(\ell+1)}\right)$.
6: **end for**
7: Set $(\boldsymbol{\theta}^*,\boldsymbol{\phi}^*) = \left(\boldsymbol{\theta}^{(L)},\boldsymbol{\phi}^{(L)}\right)$ and compute the accept rate $r = \frac{p(\boldsymbol{\theta}^*)N(\boldsymbol{\phi}^*|0,\boldsymbol{M})}{p(\boldsymbol{\theta}_{t-1})N(\boldsymbol{\phi}_{t-1}|0,\boldsymbol{M})}$.
8: Set $\boldsymbol{\theta}_t = \boldsymbol{\theta}^*$ with probability $\min(r,1)$ and $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$ otherwise.

where $\psi(\boldsymbol{\theta}) = -\log p(\boldsymbol{\theta}) + \frac{1}{2}\log\det(G(\boldsymbol{\theta}))$, and $K(\boldsymbol{\theta},\boldsymbol{\phi}) = \frac{1}{2}\boldsymbol{\phi}^T \boldsymbol{G}(\boldsymbol{\theta})^{-1}\boldsymbol{\phi}$. Due to dependence of $\boldsymbol{\phi}$ on $\boldsymbol{\theta}$, the dynamics of $\boldsymbol{\theta}$ and ϕ become non-separable. The previous version of the leapfrog is not applicable. Instead, the *generalized leapfrog* (Iserles, 1986) is used, which is an implicit scheme of fixed-point iterations.

To avoid the time-consuming iterations, Lan et al. (2015) introduced the *Lagrangian Monte Carlo* (LMC). It uses a variable transformation approach that changes Hamiltonian dynamics to Lagrangian dynamics. Specifically, let $\boldsymbol{v} = \boldsymbol{G}(\boldsymbol{\theta})^{-1}\boldsymbol{\phi}$ and its distribution is $\boldsymbol{v} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{G}(\boldsymbol{\theta})^{-1}\right)$. As $\boldsymbol{v}$ is the momentum divided by mass, it can be considered as velocity intuitively. The original Hamilton's equations in HMC (2) and (3) become the following Lagrangian dynamics (a.k.a. Euler–Lagrange equation):

$$\frac{d\boldsymbol{\theta}}{dt} = \boldsymbol{v}, \tag{5}$$

$$\frac{d\boldsymbol{v}}{dt} = -\boldsymbol{\eta}(\boldsymbol{\theta},\boldsymbol{v}) - \boldsymbol{G}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}\psi(\boldsymbol{\theta}), \tag{6}$$

where $\boldsymbol{\eta}(\boldsymbol{\theta},\boldsymbol{v})$ is a vector whose kth element is $\boldsymbol{v}^T\boldsymbol{\Gamma}^k(\boldsymbol{\theta})\boldsymbol{v}$. Here $\boldsymbol{\Gamma}^k_{i,j}(\boldsymbol{\theta}) := \frac{1}{2}\sum_l g^{k,l}\left(\partial_i g_{l,j} + \partial_j g_{i,l} - \partial_l g_{i,j}\right)$ are the Christoffel symbols, where $g_{i,j} = [\boldsymbol{G}(\boldsymbol{\theta})]_{i,j}$ and $g^{i,j} = \left[\boldsymbol{G}(\boldsymbol{\theta})^{-1}\right]_{i,j}$ and ∂_i means partial derivative with respect to θ_i. Based on the Lagrangian dynamics, Lan et al. (2015) proposed an explicit integrator, which is time reversible but not volume preserving. This is different from the HMC and RHMC. However, one can adjust the acceptance probability with the Jacobian determinant to satisfy the detailed balance condition. The LMC algorithm in Lan et al. (2015) is shown to be computationally more stable and efficient than RHMC.

Byrne and Girolami (2013) developed these manifold HMC algorithms for a class of problems where the geodesic equation (the system (5) and (6) without $\boldsymbol{G}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}\psi(\boldsymbol{\theta})$ term) can be analytically solved. Lan et al. (2014); Lan and Shahbaba (2016) proposed a specific geodesic MC on a hyper-sphere (whose geodesic is a great circle) and applied it to sample from distributions with constraints defined by vector norm.

3 | CONSTRAINED SAMPLING BASED ON HMC

In this section, we explain three different constrained sampling methods which are adapted from the original HMC. They are *Wall HMC*, *Constrained HMC*, and *Sphere HMC*, which are among the most representative ones in the reflection and reformulation types of algorithms.

3.1 | Constrained HMC by reflection

Neal (2011) discussed a method of handling the constraint $c(\boldsymbol{\theta}) \geq 0$ by modifying the original potential energy $U(\boldsymbol{\theta})$ to create a "soft" wall:

$$U_r(\boldsymbol{\theta}) = U(\boldsymbol{\theta}) + w(\boldsymbol{\theta}), \quad w(\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } c(\boldsymbol{\theta}) \geq 0, \\ r^{r+1}|c(\boldsymbol{\theta})|^r, & \text{else.} \end{cases} \tag{7}$$

Such constraint becomes a "hard" wall with infinite barrier

$$\widetilde{U}(\boldsymbol{\theta}) = \lim_{r\to\infty} U_r(\boldsymbol{\theta}) = \begin{cases} U(\boldsymbol{\theta}), & \text{if } c(\boldsymbol{\theta}) \geq 0, \\ \infty, & \text{else.} \end{cases} \tag{8}$$

Suppose the sampler just hits the wall, that is, $c(\boldsymbol{\theta}) < 0$. According to (3) with U_r, we have the momentum $\boldsymbol{\phi}$ updated as

$$\boldsymbol{\phi}_{t+1} = \boldsymbol{\phi}_t - \nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta})\Delta t - r^{r+2}\frac{|c(\boldsymbol{\theta})|^r}{c(\boldsymbol{\theta})}\Delta t \nabla_{\boldsymbol{\theta}} c(\boldsymbol{\theta}) \tag{9}$$

We can choose the time step $\Delta t = C(r) \to 0$ as $r \to \infty$ such that we have the following perfect reflection for updating momentum (Betancourt, 2011)

$$\boldsymbol{\phi}_{t+1} = \boldsymbol{\phi}_t - 2\langle \boldsymbol{\phi}_t, \boldsymbol{n}\rangle \boldsymbol{n}, \quad \boldsymbol{n} = \nabla_{\boldsymbol{\theta}} c(\boldsymbol{\theta}) / \| \nabla_{\boldsymbol{\theta}} c(\boldsymbol{\theta}) \| . \tag{10}$$

Such reflection-based HMC (named as "Wall HMC" in Lan et al. (2014)) proceeds with $\widetilde{U}(\boldsymbol{\theta})$ replacing $U(\boldsymbol{\theta})$ in Algorithm 1 where the half-step momentum updates (lines 3 and 5) are replaced by the above reflection (10) once hitting the wall (constraint violated). Betancourt (2011) and Olander (2020) successfully applied this algorithm to nest sampling. Pakman and Paninski (2014) developed a more sophisticated exact HMC algorithm for truncated multivariate Gaussian distributions based on a similar idea of reflection.

3.2 | Constrained HMC by reformulation

Brubaker et al. (2012) considered HMC for a general constraint $c(\boldsymbol{\theta}) = 0$ that defines a connected, differentiable submanifold of $\mathbb{R}^d$, denoted as $\mathscr{M} = \{\boldsymbol{\theta} \in \mathbb{R}^d | c(\boldsymbol{\theta}) = 0\}$. The constraint determines the *tangent bundle* of $\mathscr{M}$, $\mathcal{T}\mathscr{M} = \{(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}) | c(\boldsymbol{\theta}) = 0 \text{ and } \frac{\partial c}{\partial \boldsymbol{\theta}}\dot{\boldsymbol{\theta}} = 0\}$ where $\frac{\partial c}{\partial \boldsymbol{\theta}}$ is the Jacobian of the constraints. Now we have the new Hamiltonian as

$$H(\boldsymbol{\theta}, \boldsymbol{\phi}, \lambda) = \widehat{H}(\boldsymbol{\theta}, \boldsymbol{\phi}) + \lambda^T c(\boldsymbol{\theta}), \quad \widehat{H}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \psi(\boldsymbol{\theta}) + K(\boldsymbol{\theta}, \boldsymbol{\phi}) \tag{11}$$

where λ is the Lagrange multiplier. Then the Hamiltonian dynamics with the above guided Hamiltonian $\widehat{H}(\boldsymbol{\theta}, \boldsymbol{\phi})$ become (2) and (3) with extra an equation $c(\boldsymbol{\theta}) = 0$ and are defined on the *cotangent bundle* $\mathcal{T}^*\mathscr{M} = \left\{(\boldsymbol{\theta}, \boldsymbol{\phi}) | c(\boldsymbol{\theta}) = 0 \text{ and } \frac{\partial c}{\partial \boldsymbol{\theta}}\frac{\partial \widehat{H}}{\partial \boldsymbol{\phi}} = 0\right\}$. Brubaker et al. (2012) used a consistent integrator called RATTLE (Andersen, 1983; Leimkuhler & Reich, 1994) to solve the constrained Hamiltonian dynamics in their proposed constrained HMC (CHMC).

$$\begin{aligned}
\boldsymbol{\phi}^{\left(\ell+\frac{1}{2}\right)} &= \boldsymbol{\phi}^{(\ell)} - \frac{\epsilon}{2}\left[\frac{\partial \widehat{H}\left(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\phi}^{\left(\ell+\frac{1}{2}\right)}\right)}{\partial \boldsymbol{\theta}} + \frac{\partial c}{\partial \boldsymbol{\theta}}\left(\boldsymbol{\theta}^{(\ell)}\right)\lambda\right] \\
\boldsymbol{\theta}^{(\ell+1)} &= \boldsymbol{\theta}^{(\ell)} + \frac{\epsilon}{2}\left[\frac{\partial \widehat{H}\left(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\phi}^{\left(\ell+\frac{1}{2}\right)}\right)}{\partial \phi} + \frac{\partial \widehat{H}\left(\boldsymbol{\theta}^{(\ell+1)}, \boldsymbol{\phi}^{\left(\ell+\frac{1}{2}\right)}\right)}{\partial \phi}\right] \\
0 &= c\left(\boldsymbol{\theta}^{(\ell+1)}\right) \\
\boldsymbol{\phi}^{(\ell+1)} &= \boldsymbol{\phi}^{\left(\ell+\frac{1}{2}\right)} - \frac{\epsilon}{2}\left[\frac{\partial \widehat{H}\left(\boldsymbol{\theta}^{(\ell+1)}, \boldsymbol{\phi}^{\left(\ell+\frac{1}{2}\right)}\right)}{\partial \boldsymbol{\theta}} + \frac{\partial c}{\partial \boldsymbol{\theta}}\left(\boldsymbol{\theta}^{(\ell+1)}\right)\mu\right] \\
0 &= \frac{\partial c}{\partial \boldsymbol{\theta}}\left(\boldsymbol{\theta}^{(\ell+1)}\right)\frac{\partial \widehat{H}\left(\boldsymbol{\theta}^{(\ell+1)}, \boldsymbol{\phi}^{(\ell+1)}\right)}{\partial \boldsymbol{\phi}}
\end{aligned} \tag{12}$$

where λ and μ are Lagrange multipliers associated with the state and momentum constraints (the third and the fifth equations of (12)). This generalizes the leapfrog method to handle the manifold constraints. The proposed state $\boldsymbol{\theta}^*$ is accepted with probability defined by H in (11).

3.3 | Spherical HMC

Lan et al. (2014) considered HMC defined on a special manifold, hyper-sphere, denoted as $\mathcal{S}^d = \left\{\boldsymbol{\theta} \in \mathbb{R}^{d+1} | \| \boldsymbol{\theta}\|_2^2 = \sum_{i=1}^{d+1} \theta_i^2 = 1\right\}$. This algorithm is particularly useful to handle a class of constraints defined by the following vector ($\boldsymbol{\beta} \in \mathbb{R}^d$) q-norm:

$$\| \boldsymbol{\beta}\|_q = \begin{cases} \left(\sum_{i=1}^{d} |\beta_i|^q\right)^{1/q}, & q \in (0, +\infty) \\ \max_{1 \le i \le d} |\beta_i|, & q = +\infty \end{cases} \tag{13}$$

The q-norm domain, $\mathcal{Q}^d \coloneqq \left\{\boldsymbol{\beta} \in \mathbb{R}^d | \| \boldsymbol{\beta}\|_q \le 1\right\}$, can be transformed to the unit ball $\mathcal{B}_0^d(1) \coloneqq \left\{\boldsymbol{\theta} \in \mathbb{R}^d | \| \boldsymbol{\theta}\|_2 \le 1\right\}$ by either $\beta_i \mapsto \theta_i = \text{sgn}(\beta_i)|\beta_i|^{q/2}$ (shown the left panel of Figure 1) or $\boldsymbol{\beta} \mapsto \boldsymbol{\theta} = \boldsymbol{\beta} \frac{\|\boldsymbol{\beta}\|_\infty}{\|\boldsymbol{\beta}\|_2}$ (shown in the right panel of Figure 1).

To define HMC for these constrained distributions, Lan and Shahbaba (2016) proposed an idea of *spherical augmentation* to further map the unit ball $\mathcal{B}_{\mathbf{0}}^d(1)$ to the hyper-sphere $\mathcal{S}^d$ by appending an auxiliary variable θ_{d+1} to the original vector $\boldsymbol{\theta} \in \mathcal{B}_{\mathbf{0}}^d(1)$ such that the extended parameter $\widetilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}, \theta_{d+1}) \in \mathcal{S}^d$. The lower hemisphere ($\theta_{d+1} = -\sqrt{1 - \| \boldsymbol{\theta}\|_2^2}$) is also identified with the upper hemisphere ($\theta_{d+1} = \sqrt{1 - \| \boldsymbol{\theta}\|_2^2}$) by ignoring the sign of θ_{d+1}. After collecting samples $\left\{\widetilde{\boldsymbol{\theta}}\right\}$ using spherical HMC defined on the sphere, $\mathcal{S}^d$, the last component θ_{d+1} is discarded and the obtained samples $\{\boldsymbol{\theta}\}$ automatically satisfy the constraint $\| \boldsymbol{\theta}\|_2 \le 1$. As illustrated in Figure 2, the boundary of the constraint, that is, $\| \boldsymbol{\theta}\|_2 = 1$, corresponds to the equator on the sphere $\mathcal{S}^d$. Therefore, as the sampler moves on the sphere, for example, from A to B, passing across the equator from one hemisphere to the other translates to "bouncing back" off the boundary in the original parameter space.

On the Riemannian manifold $\left(\mathcal{S}^d, \mathbf{G}(\boldsymbol{\theta})\right)$ where $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{I}_d + \boldsymbol{\theta}\boldsymbol{\theta}^T / \left(1 - \| \boldsymbol{\theta}\|_2^2\right)$ is the *canonical spherical metric*, the tangent space at $\widetilde{\boldsymbol{\theta}}$ is defined as $\mathcal{T}_{\widetilde{\boldsymbol{\theta}}}\mathcal{S}^d = \left\{\widetilde{\boldsymbol{\nu}} = (\boldsymbol{\nu}, \nu_{d+1}) \in \mathbb{R}^{d+1} | \widetilde{\boldsymbol{\theta}}^T \widetilde{\boldsymbol{\nu}} = 0\right\}$. We have the Hamiltonian (4) redefined on the tangent bundle $\mathcal{TS}^d = \left\{\left(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\nu}}\right) | \| \widetilde{\boldsymbol{\theta}}\|_2 = 1 \text{ and } \widetilde{\boldsymbol{\theta}}^T \widetilde{\boldsymbol{\nu}} = 0\right\}$:

$$H\left(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\nu}}\right) = H^*\left(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\nu}}\right) + \frac{1}{2}\log\det(\boldsymbol{G}(\boldsymbol{\theta})), \quad H^*\left(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\nu}}\right) = U\left(\widetilde{\boldsymbol{\theta}}\right) + K(\widetilde{\boldsymbol{\nu}}) \tag{14}$$

where the potential energy $U\left(\widetilde{\boldsymbol{\theta}}\right) = U(\boldsymbol{\theta})$, and the kinetic energy $K(\widetilde{\boldsymbol{\nu}}) = \frac{1}{2}\boldsymbol{\nu}^T \mathbf{G}(\boldsymbol{\theta})\boldsymbol{\nu} = \frac{1}{2}\| \widetilde{\boldsymbol{\nu}}\|_2^2$ which is defined for the velocity random variable $\widetilde{\boldsymbol{\nu}} \sim \mathcal{N}\left(0, \mathcal{P}\left(\widetilde{\boldsymbol{\theta}}\right)\right)$ with $\mathcal{P}\left(\widetilde{\boldsymbol{\theta}}\right) = \mathbf{I}_{d+1} - \widetilde{\boldsymbol{\theta}}\widetilde{\boldsymbol{\theta}}^T$ being the projection matrix.

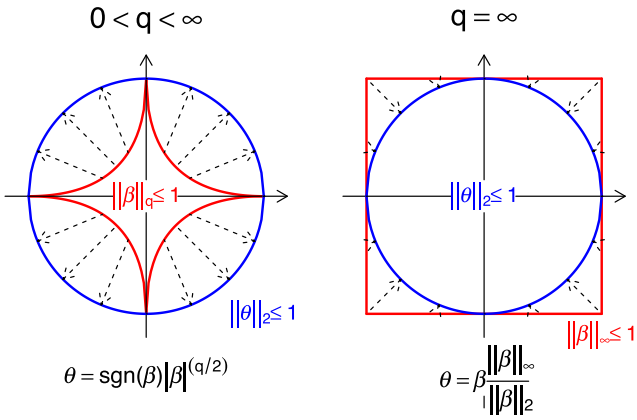

FIGURE 1 Transforming q-norm constrained domain to the unit ball. Left: from general q-norm domain $\mathcal{Q}^d$ to unit ball $\mathcal{B}_0^d(1)$; Right: from the unit cube $\mathcal{C}^d$ to the unit ball $\mathcal{B}_0^d(1)$.

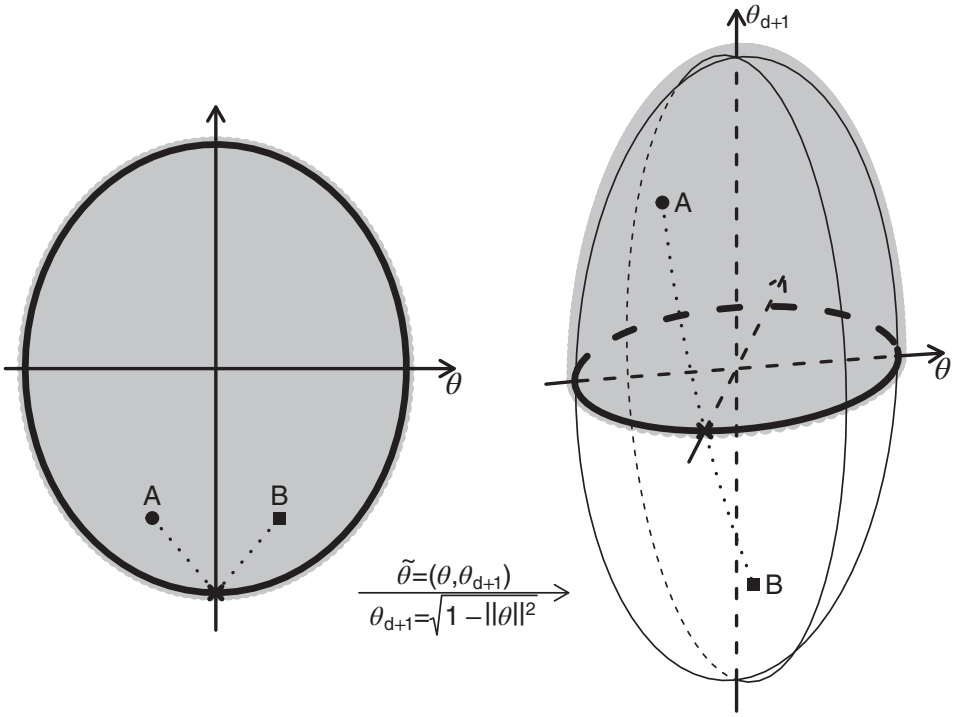

FIGURE 2 Transforming the unit ball $\mathscr{B}_{\mathbf{0}}^{d}(1)$ to the sphere $\mathcal{S}^{d}$.

With such guided Hamiltonian function H^{*} in (14), the Hamiltonian dynamics can be defined on the Riemannian manifold $\left(\mathcal{S}^{d},\mathbf{G}(\boldsymbol{\theta})\right)$ in terms of $(\boldsymbol{\theta},\mathbf{p})$, or equivalently as the following Lagrangian dynamics in terms of $(\boldsymbol{\theta},\widetilde{\boldsymbol{\nu}})$ (Lan et al., 2015):

$$\begin{aligned}\dot{\widetilde{\boldsymbol{\theta}}} &= \widetilde{\boldsymbol{\nu}}\\ \dot{\widetilde{\boldsymbol{\nu}}} &= -\|\widetilde{\boldsymbol{\nu}}\|_2^2\widetilde{\boldsymbol{\theta}} - \mathcal{P}\left(\widetilde{\boldsymbol{\theta}}\right)\nabla_{\widetilde{\boldsymbol{\theta}}}U(\boldsymbol{\theta})\end{aligned} \tag{15}$$

where the projection matrix $\mathcal{P}\left(\widetilde{\boldsymbol{\theta}}\right)$ maps the directional derivative $\nabla_{\widetilde{\boldsymbol{\theta}}}U(\boldsymbol{\theta})$ onto the tangent space $\mathcal{T}_{\widetilde{\boldsymbol{\theta}}}\mathcal{S}^{d}$. The dynamics (15) can be split into two dynamics

$$\begin{cases}\dot{\widetilde{\boldsymbol{\theta}}} &= \mathbf{0}\\ \dot{\widetilde{\boldsymbol{\nu}}} &= -\mathcal{P}\left(\widetilde{\boldsymbol{\theta}}\right)\nabla_{\widetilde{\boldsymbol{\theta}}}U(\boldsymbol{\theta})\end{cases} \tag{16a}$$

$$\begin{cases}\dot{\widetilde{\boldsymbol{\theta}}} &= \widetilde{\boldsymbol{\nu}}\\ \dot{\widetilde{\boldsymbol{\nu}}} &= -\mathcal{P}\left(\widetilde{\boldsymbol{\theta}}\right)\nabla_{\widetilde{\boldsymbol{\theta}}}U(\boldsymbol{\theta})\end{cases} \tag{16b}$$

where the solution to (16a) only involves updating velocity and (16b) is the geodesic equation on $\mathcal{S}^{d}$ with the solution being the big circle. *Spherical HMC* (*SphHMC*) proceeds by proposing a joint state by alternate updates according to (16a) (lines 5 and 8 in Algorithm 2) and (16b) (lines 6–7 in Algorithm 2) and accepting or rejecting the proposal based on the acceptance probability defined by H. Algorithm 2 summarizes the details of spherical HMC (SphHMC).

ALGORITHM 2 **Spherical HMC (SphHMC)**

1: Initialize $\widetilde{\boldsymbol{\theta}}^{(0)}$ at current $\widetilde{\boldsymbol{\theta}}$.
2: Sample a new velocity value $\widetilde{\mathbf{v}}^{(0)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{d+1})$, and set $\widetilde{\mathbf{v}}^{(0)} \leftarrow \mathcal{P}\left(\widetilde{\boldsymbol{\theta}}^{(0)}\right)\widetilde{\mathbf{v}}^{(0)}$.
3: Calculate $H\left(\widetilde{\boldsymbol{\theta}}^{(0)}, \widetilde{\mathbf{v}}^{(0)}\right) = U\left(\boldsymbol{\theta}^{(0)}\right) + K\left(\widetilde{\mathbf{v}}^{(0)}\right)$.
4: **for** $\ell = 0$ to $L-1$ **do**
5: $\quad \widetilde{\mathbf{v}}^{\left(\ell+\frac{1}{2}\right)} = \widetilde{\mathbf{v}}^{(\ell)} - \frac{\epsilon}{2}\mathcal{P}\left(\widetilde{\boldsymbol{\theta}}^{(\ell)}\right)\nabla_{\boldsymbol{\theta}} U\left(\boldsymbol{\theta}^{(\ell)}\right)$
6: $\quad \widetilde{\boldsymbol{\theta}}^{(\ell+1)} = \widetilde{\boldsymbol{\theta}}^{(\ell)}\cos\left(\| \widetilde{\mathbf{v}}^{\left(\ell+\frac{1}{2}\right)} \| \epsilon\right) + \frac{\widetilde{\mathbf{v}}^{\left(\ell+\frac{1}{2}\right)}}{\| \widetilde{\mathbf{v}}^{\left(\ell+\frac{1}{2}\right)} \|}\sin\left(\| \widetilde{\mathbf{v}}^{\left(\ell+\frac{1}{2}\right)} \| \epsilon\right)$
7: $\quad \widetilde{\mathbf{v}}^{\left(\ell+\frac{1}{2}\right)} \leftarrow -\widetilde{\boldsymbol{\theta}}^{(\ell)} \| \widetilde{\mathbf{v}}^{\left(\ell+\frac{1}{2}\right)} \| \sin\left(\| \widetilde{\mathbf{v}}^{\left(\ell+\frac{1}{2}\right)} \| \epsilon\right) + \widetilde{\mathbf{v}}^{\left(\ell+\frac{1}{2}\right)}\cos\left(\| \widetilde{\mathbf{v}}^{\left(\ell+\frac{1}{2}\right)} \| \epsilon\right)$
8: $\quad \widetilde{\mathbf{v}}^{(\ell+1)} = \widetilde{\mathbf{v}}^{\left(\ell+\frac{1}{2}\right)} - \frac{\epsilon}{2}\mathcal{P}\left(\widetilde{\boldsymbol{\theta}}^{(\ell+1)}\right)\nabla_{\boldsymbol{\theta}} U\left(\boldsymbol{\theta}^{(\ell+1)}\right)$
9: **end for**
10: Calculate $H\left(\widetilde{\boldsymbol{\theta}}^{(L)}, \widetilde{\mathbf{v}}^{(L)}\right) = U\left(\boldsymbol{\theta}^{(L)}\right) + K\left(\widetilde{\mathbf{v}}^{(L)}\right)$.
11: Calculate the acceptance probability $\alpha = \min\left\{1, \exp\left[-H\left(\widetilde{\boldsymbol{\theta}}^{(L)}, \widetilde{\mathbf{v}}^{(L)}\right) + H\left(\widetilde{\boldsymbol{\theta}}^{(0)}, \widetilde{\mathbf{v}}^{(0)}\right)\right]\right\}$.
12: Accept or reject the proposal according to α for the next state $\widetilde{\boldsymbol{\theta}}'$.

4 | APPLICATIONS OF CONSTRAINED SAMPLING

4.1 | Truncated multivariate Gaussian

For illustration purposes, we start with a truncated bivariate Gaussian distribution,

$$\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}1 & 0.5\\ 0.5 & 1\end{bmatrix}\right), \quad 0 \le \beta_1 \le 5, \;\; 0 \le \beta_2 \le 1$$

This is a rectangular type constraint with the lower and upper limits as $\mathbf{l} = (0,0)$ and $\mathbf{u} = (5,1)$ respectively. The original rectangle domain can be mapped to 2d unit disc $\mathscr{B}_0^2(1)$ to use c-SphHMC, or mapped to 2d rectangle $\mathscr{R}_0^2$ where s-SphHMC can be directly applied (Lan & Shahbaba, 2016).

The upper leftmost panel of Figure 3 shows the heatmap based on the exact density function, and the other panels show the corresponding heatmaps based on MCMC samples from RWM, Wall HMC, exact HMC, c-SphHMC, and s-SphHMC, respectively. All algorithms generate probability density estimates that visually match the true density. Table 1 compares the true mean and covariance of the above truncated bivariate Gaussian distribution with the point estimates using 2×10^5 (2×10^4 for each of 10 repeated experiments with different random seeds) MCMC samples in each method. Overall, all methods estimate the mean and covariance reasonably well.

To evaluate the efficiency of the above-mentioned methods, we repeat this experiment for higher dimensions, $D = 10$, and $D = 100$. As before, we set the mean to zero and set the (i,j)th element of the covariance matrix to $\Sigma_{ij} = 1/(1+|i-j|)$. Further, we impose the following constraints on the parameters,

$$0 \le \beta_i \le u_i$$

where u_i (i.e., the upper bound) is set to 5 when $i = 1$; otherwise, it is set to 0.5.

For each method, we obtain 10^5 MCMC samples after discarding the initial 10^4 samples. We set the tuning parameters of algorithms such that their overall acceptance rates are within a reasonable range. As shown in Table 2, Spherical HMC algorithms are substantially more efficient than RWM and Wall HMC. For RWM, the proposed states are rejected

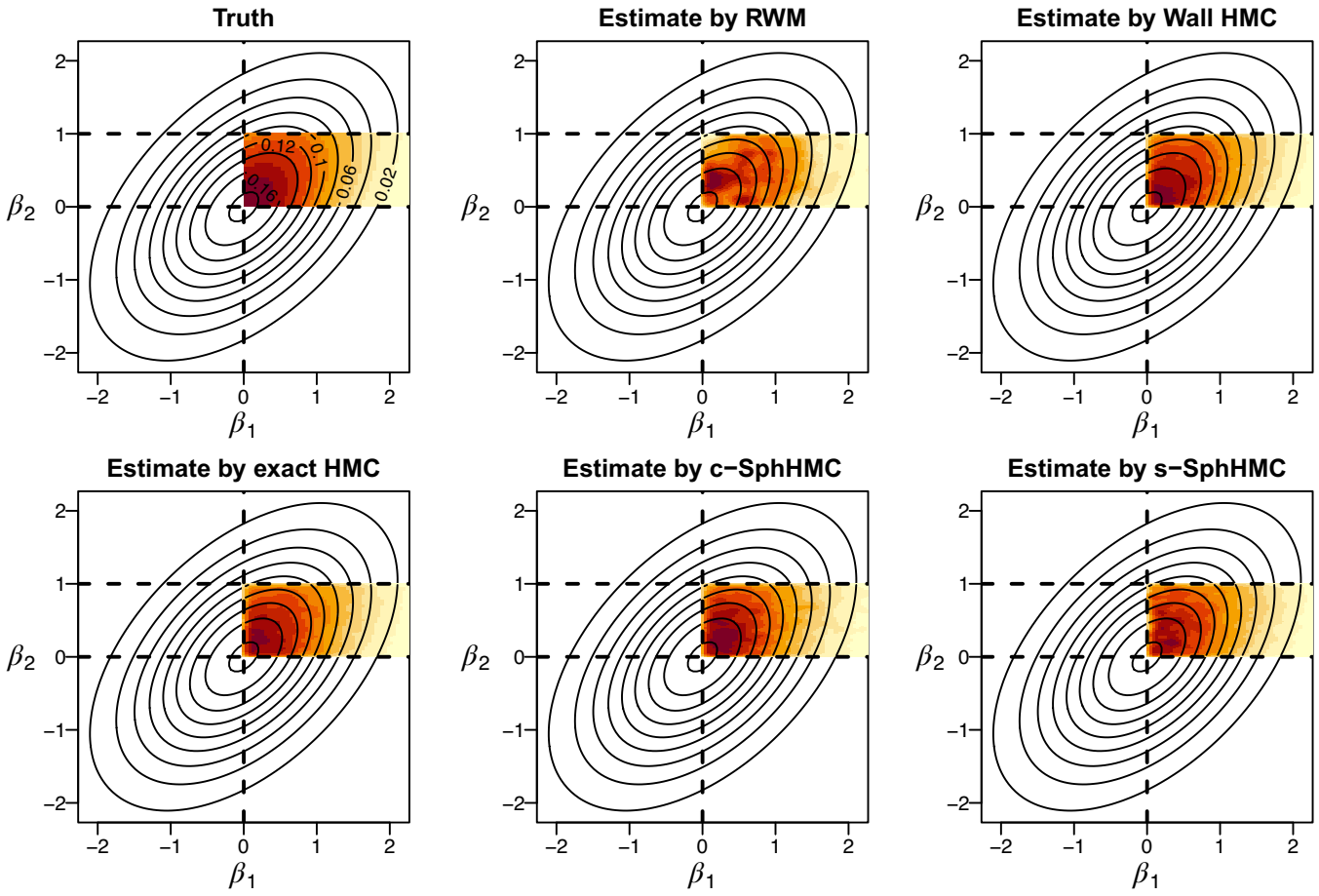

FIGURE 3 Density plots of a truncated bivariate Gaussian using exact density function (upper leftmost) and MCMC samples from RWM, Wall HMC, exact HMC, c-SphHMC, and s-SphHMC, respectively. Solid elliptical curves always show true unconstrained probability density contours. Dashed lines define linear constrained domains. Colored heatmaps indicate constrained probability density based on truth or estimation from MCMC samples.

TABLE 1 Comparing the point estimates for the mean and covariance of a bivariate truncated Gaussian distribution using RWM, Wall HMC, exact HMC, and SphHMC.

Method	Mean	Covariance
Truth	$\begin{bmatrix}0.7906\\0.4889\end{bmatrix}$	$\begin{bmatrix}0.3269 & 0.0172\\0.0172 & 0.08\end{bmatrix}$
RWM	$\begin{bmatrix}0.7796\pm0.0088\\0.4889\pm0.0034\end{bmatrix}$	$\begin{bmatrix}0.3214\pm0.009 & 0.0158\pm0.001\\0.0158\pm0.001 & 0.0798\pm5e-04\end{bmatrix}$
Wall HMC	$\begin{bmatrix}0.7875\pm0.0049\\0.4884\pm8e-04\end{bmatrix}$	$\begin{bmatrix}0.3242\pm0.0043 & 0.017\pm0.001\\0.017\pm0.001 & 0.08\pm3e-04\end{bmatrix}$
exact HMC	$\begin{bmatrix}0.7909\pm0.0025\\0.4885\pm0.001\end{bmatrix}$	$\begin{bmatrix}0.3272\pm0.0026 & 0.0174\pm7e-04\\0.0174\pm7e-04 & 0.08\pm3e-04\end{bmatrix}$
SphHMC	$\begin{bmatrix}0.79\pm0.005\\0.4864\pm0.0016\end{bmatrix}$	$\begin{bmatrix}0.3249\pm0.0045 & 0.0172\pm0.0012\\0.0172\pm0.0012 & 0.0801\pm0.001\end{bmatrix}$

about 95% of times due to violation of the constraints. On average, Wall HMC bounces off the wall at around 3.81 ($L=2$) and 6.19 ($L=5$) times per iteration for $D=10$ and $D=100$, respectively. Exact HMC is quite efficient for relatively low dimensional truncated Gaussian ($D=10$); however, it becomes very slow for higher dimensions ($D=100$). In contrast, by augmenting the parameter space, Spherical HMC algorithms handle the constraints more efficiently. Since

TABLE 2 Comparing the efficiency of RWM, Wall HMC, exact HMC, and SphHMC in terms of sampling from truncated Gaussian distributions.

Dimension	Method	AP[a]	s/iter[b]	ESS (min, med, max)[c]	Min(ESS)/s[d]	Speedup
$D = 10$	RWM	0.62	5.72E−05	(48, 691, 736)	7.58	1.00
	Wall HMC	0.83	1.19E−04	(31,904, 86,275, 87,311)	2441.72	322.33
	Exact HMC	1.00	7.60E−05	(1e+05, 1e+05, 1e+05)	11,960.29	1578.87
	SphHMC	0.82	2.53E−04	(62,658, 85,570, 86,295)	2253.32	297.46
$D = 100$	RWM	0.81	5.45E−04	(1, 4, 54)	0.01	1.00
	Wall HMC	0.74	2.23E−03	(17,777, 52,909, 55,713)	72.45	5130.21
	Exact HMC	1.00	4.65E−02	(97,963, 1e+05, 1e+05)	19.16	1356.64
	SphHMC	0.73	3.45E−03	(55,667, 68,585, 72,850)	146.75	10,390.94

[a]Acceptance probability.
[b]Seconds per iteration.
[c](Minimum, median, maximum) effective sample size.
[d]Minimal ESS per second.

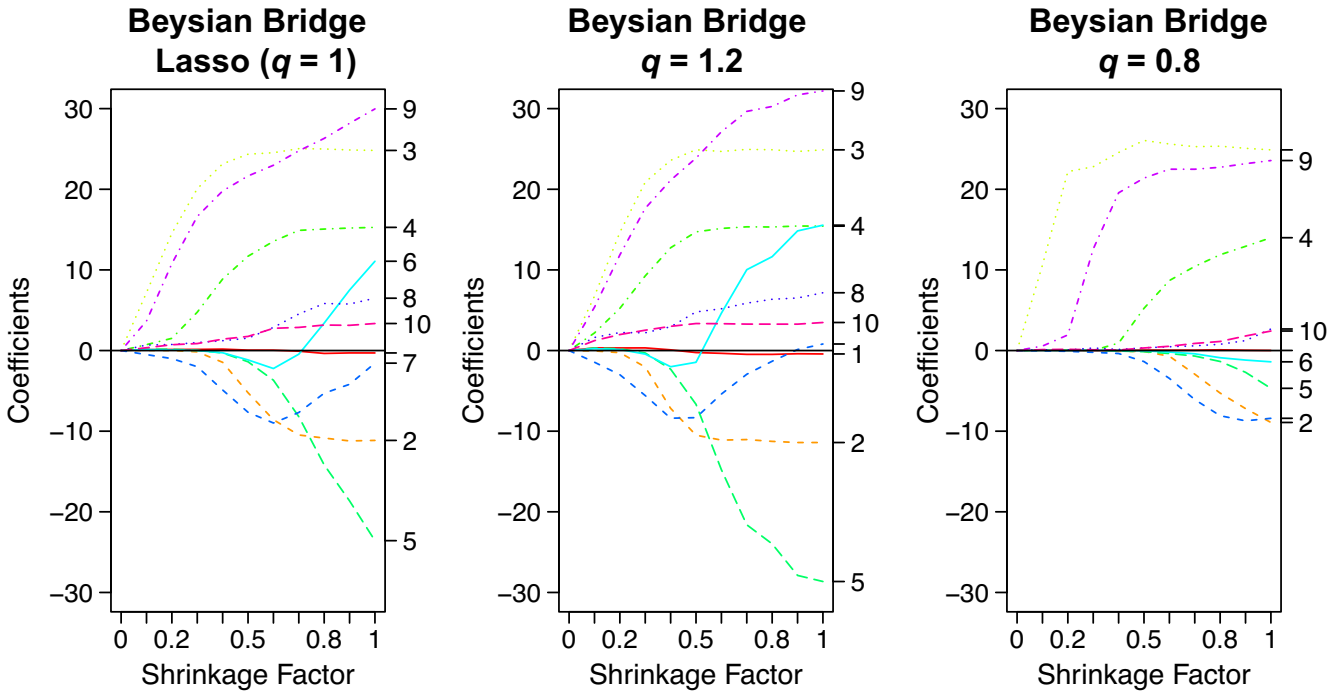

FIGURE 4 Bayesian bridge regression by spherical HMC: Lasso ($q = 1$, left), $q = 1.2$ (middle), and $q = 0.8$ (right).

s-SphHMC is more suited for rectangular-type constraints, it is substantially more efficient than c-SphHMC in this example.

It is worth mentioning that some new non-HMC methods have been developed to sample from the truncated multivariate normal distributions. Examples include the bouncy particle sampler in Zhang et al. (2021) and zig–zag HMC in Nishimura et al. (2021); Zhang et al. (2022) which use the non-reversible zig–zag process in Bierkens et al. (2019). But since they are non-HMC methods or designed for truncated Gaussians, we skip reviewing them in more details.

4.2 | Bayesian regularized regression

In regression analysis, overly complex models tend to overfit the data. Regularized regression models control complexity by imposing a penalty on model parameters. *Bridge regression* (Frank & Friedman, 1993) is a family of regression

models where the coefficients are obtained by minimizing the residual sum of squares subject to a constraint on the magnitude of regression coefficients as follows:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - X_i\boldsymbol{\beta})^2 \quad \text{subject to} \quad \|\boldsymbol{\beta}\|_q \le r \tag{17}$$

When $q=1$, this corresponds to *Lasso* (least absolute shrinkage and selection operator) proposed by Tibshirani (1996) which allows the model to force some of the coefficients to become exactly zero (i.e., become excluded from the model). When $q=2$, this model is known as *ridge regression*. Bridge regression is more flexible by allowing different q norm constraints for different effects on shrinking the magnitude of parameters (see Figure 4).

Park and Casella (2008) and Hans (2009) proposed *Bayesian Lasso* by employing a conjugate prior distribution of the form $P(\beta) \propto \exp(-\lambda|\beta|)$. Frank and Friedman (1993) also constructed a complicated prior for the Bayesian inference of bridge regression. With spherical HMC, there is much flexibility in choosing priors with q-norm constraints. We can define the following *Bayesian regularized linear regression model*:

$$\begin{aligned} \boldsymbol{y}|\boldsymbol{X},\boldsymbol{\beta},\sigma_\epsilon^2 &\sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{\beta},\sigma_\epsilon^2\mathbf{I}\right) \\ \boldsymbol{\beta} &\sim p(\boldsymbol{\beta})\mathbf{1}\left(\|\boldsymbol{\beta}\|_q \le r\right) \end{aligned} \tag{18}$$

Figure 4 compares the parameter estimates of Bayesian Lasso to the estimates obtained from two Bridge regression models with $q=1.2$ and $q=0.8$ for the diabetes data set ($N=442$, $D=10$) studied in Park and Casella (2008). Truncated Gaussian prior $\boldsymbol{\beta} \sim \mathcal{N}(0,\sigma^2\mathbf{I})\mathbf{1}\left(\|\boldsymbol{\beta}\|_q \le r\right)$ is considered and posterior samples are collected using spherical HMC algorithm. Figure 4 illustrates the parameter estimates $\widehat{\boldsymbol{\beta}}$ with respect to the shrinkage factor $s := \|\widehat{\boldsymbol{\beta}}^{\text{Lasso}}\|_1 / \|\widehat{\boldsymbol{\beta}}^{\text{OLS}}\|_1$ varying from 0 to 1, where $\widehat{\boldsymbol{\beta}}^{\text{OLS}}$ denotes the estimates obtained by ordinary least squares (OLS) regression. As expected, tighter constraints (e.g., $q=0.8$) would lead to faster shrinkage of regression parameters as we decrease s. Note, the model (18) can easily be generalized to generalized linear models or nonparametric models such as Gaussian process regression.

4.3 | Non-parametric density estimation

In this example of non-parametric density estimation, we show how a density function $p(x)$ can be modeled on an infinite dimensional sphere and how the spherical HMC can be applied to the Bayesian inference to efficiently obtain the posterior estimate. We briefly explain the method in the following. More details can be found in Holbrook et al. (2020).

Suppose we want to attribute a smooth density function $p(x)$ to observed data $\{x_n\}_{n=1}^N$ on finite domain $\mathcal{D} \subset \mathbb{R}^d$. Define the space of density functions $p(x)$ and the space of square-root density functions, $q(x) = \sqrt{p(x)}$:

$$\mathcal{P} := \left\{ p : \mathcal{D} \to \mathbb{R} \mid p \ge 0, \int_{\mathcal{D}} p(x)\mu(dx) = 1 \right\}$$

$$\mathcal{Q} := \left\{ q : \mathcal{D} \to \mathbb{R} \mid \int_{\mathcal{D}} q(x)^2\mu(dx) = 1 \right\},$$

respectively. Although the space $\mathcal{P}$ contains the functions of interest, we instead opt to deal with the space $\mathcal{Q}$ of square-root densities, which can be viewed as the unit sphere in the infinite-dimensional Hilbert space $L^2(\mathcal{D})$. We model the square-root density with a GP prior multiplied by the Dirac measure restricting the function to the unit sphere:

$$q(\cdot) \sim \mathcal{GP}(0,\mathcal{C}) \times \delta(q(\cdot) \in \mathcal{Q}). \tag{19}$$

where $\mathcal{C}$ has Eigen-pairs $\{\lambda_\ell, \phi_\ell(x)\}$ such that $\mathcal{C}\phi_\ell(x) = \lambda_\ell\phi_\ell(x)$ for $\ell \in \mathbb{N}$.

Based on the Karhunen–Loève representation (Wang, 2008) of the Gaussian random function, we have the following expansion of q:

$$q(\cdot) = \sum_{\ell=1}^{\infty} q_\ell \phi_i(\cdot), q_\ell \overset{ind}{\sim} N\left(0, \lambda_\ell^2\right) \tag{20}$$

If we let $\{\phi_\ell(x)\}$ be an orthonormal basis on $L^{(\mathcal{D})}$, then the unit sphere restriction, $\delta(q \in \mathcal{Q})$, translates to following requirement on the infinite sequence $q := \{q_\ell\}$:

$$q \in \mathcal{S}^\infty = \left\{ q \in \ell^2 | \langle q, q \rangle_{\ell^2} = \sum_{i=1}^{\infty} q_i^2 = 1 \right\}.$$

In practice, we truncate the K–L expansion (20) at $L > 0$ terms and have $q = \{q_\ell\}_{\ell=1}^L \in \mathcal{S}^{L-1}$.
Now we have the prior for q and the likelihood of the data $x := \{x_n\}_{n=1}^N$ given q as follows

$$\begin{aligned} \pi(q) \propto \quad & \delta\left(q \in \mathcal{S}^{L-1}\right) \prod_{i=1}^{L} \exp\left(-q_i^2/\left(2\lambda_i^2\right)\right), \\ \pi(x|q) = \quad & \prod_{n=1}^{N} q^2(x_n) = \prod_{n=1}^{N} |\sum_{\ell=1}^{L} q_\ell \phi_\ell(x_n)|^2, \end{aligned}$$

We can then apply the spherical HMC to sample from the posterior $\pi(q|x) \propto \pi(q)\pi(x|q)$ which is naturally defined on the sphere $\mathcal{S}^{L-1}$.

Figure 5 depicts 1000 data points (red) drawn from four different distributions on the unit square along with the contours of the pointwise median of 1000 posterior draws from the model $\pi(q|x)$ as defined above. The data in the first three plots are generated using truncated Gaussians and mixtures of truncated Gaussians. The data for the last plot is

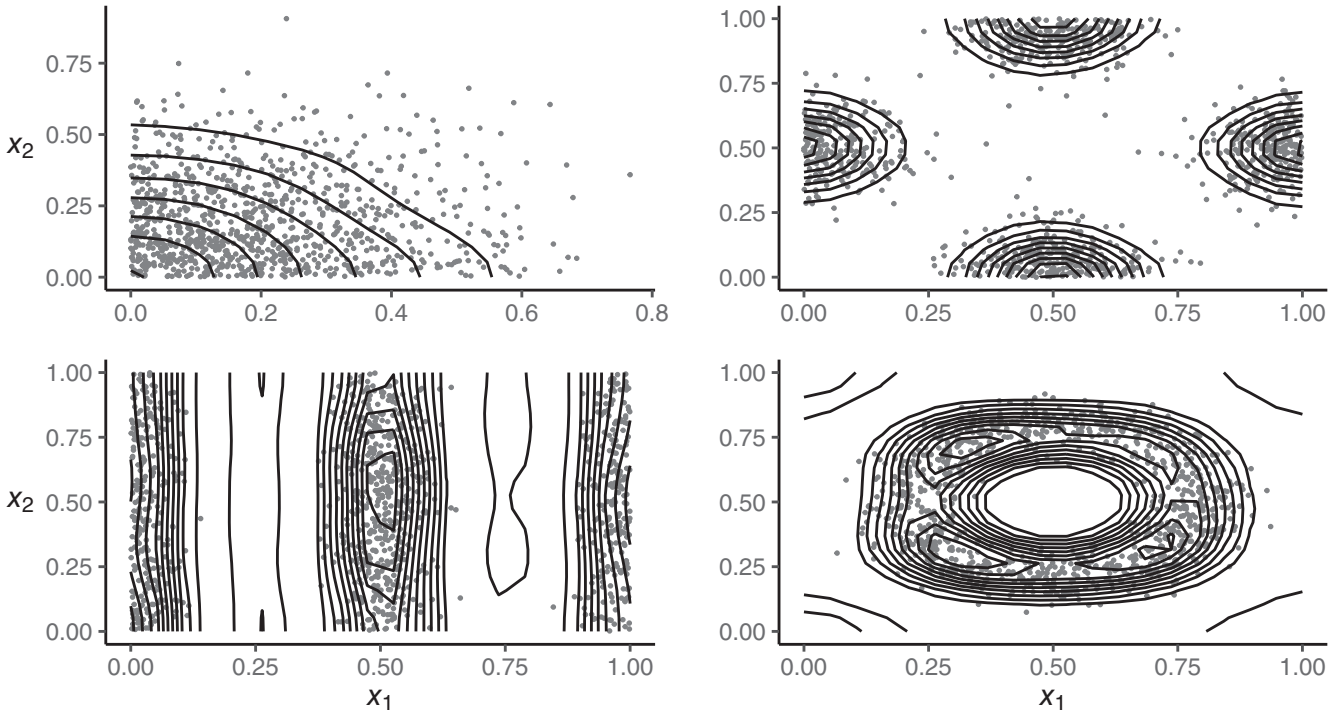

FIGURE 5 The contours (black) of the posterior median from 1000 draws of the χ^2-process density sampler. Each posterior is conditioned on 1000 data points (red).

generated by Gaussian noise added to the uniform distribution on the circle. The model adapts easily to multimodal and patterned data samples.

5 | CONSTRAINED DESIGN OF EXPERIMENTS

The statistical design of experiments is naturally related to statistical sampling, although they achieve different goals. Driven by practical needs, sometimes the design variables have to satisfy certain constraints, which make the design space irregular.

A common type of constrained physical experiment is the mixture experiment, which is widely used in the chemical, pharmaceutical, and food industries. The input variables of a mixture experiment are the proportions of different ingredients of the whole mixture. Thus the values of all the variables are between 0 and 1 (without scaling or other transformation), and the sum of them must be equal to 1. In Piepel et al. (2005), the experiment contains 21 ingredients including both rectangular and linear constraints. Mixture experiments can also have layered mixture structures which makes both design and modeling more complicated (Kang et al., 2011; Shen et al., 2020).

Many computer experiments also involve irregular constraints, and space-filling design methods have to be adapted to deal with these constraints. Stinstra et al. (2003) showed a large-scale computer experiment for television tube design that involves 23 input variables and 44 non-rectangular constraints. Draguljić et al. (2012) illustrated a Total Elbow Replacement computer simulation with four input variables and two linear non-rectangular constraints.

There have been many design methods proposed to handle the constraints. Draguljić et al. (2012) developed a column-wise construction method for space-filling design. In Pratola et al. (2017), a dense candidate set is generated and then the unfeasible candidate points are removed and the optimal design points are selected from the feasible ones. Maximin distance and the (robust) IMSPE criteria are used for the space-filling design. Huang et al. (2021) proposed a constrained minimum energy design method for constructing space-filling designs in any non-regular bounded space, and its key idea is to apply the minimum energy design (a deterministic sampling algorithm) on the target distribution using the probabilistic constraints proposed in sequentially constrained Monte Carlo.

The aforementioned methods are only for constrained space-filling designs, which can be considered as sampling from uniform distributions with constraints. Kang (2019) proposed a generic method to construct optimal designs for an irregular constrained design space. A stochastic coordinate-exchange (SCE) method is developed. In each iteration of coordinate exchange, the multi-dimensional constraints are projected into the dimension of the coordinate to be exchanged (or improved) with the other coordinates fixed at the current values. Therefore, the multi-dimensional constraints are reduced to one-dimensional rectangular constraints. The generic method can be adapted for different design criteria, including the D- and linear-optimal design for physical experiments, and the ϕ_p-optimal space-filling design for computer experiments. Figure 6 shows three optimal designs constructed by the SCE method for different kinds of constraints.

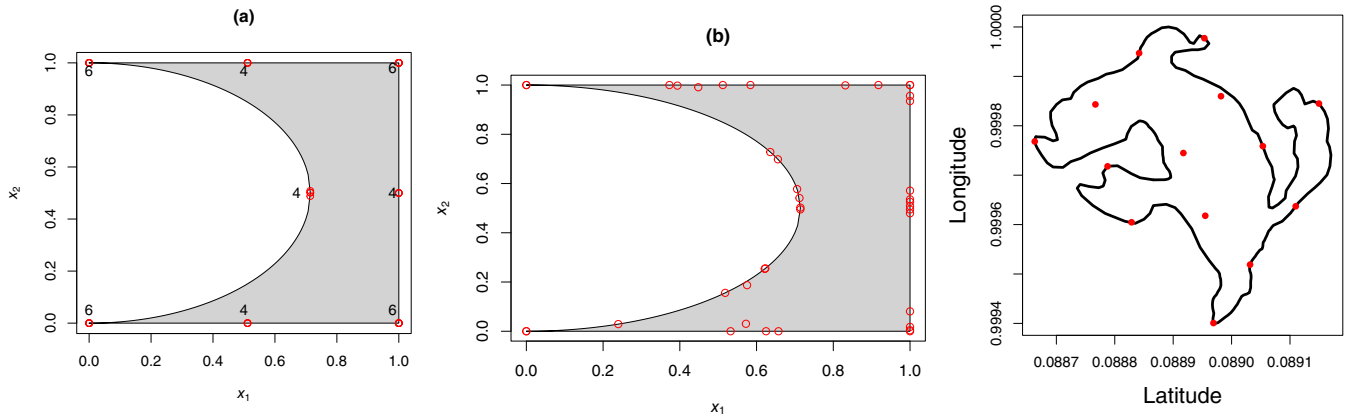

FIGURE 6 Different constrained designs returned by SCE in Kang (2019): (a) D-optimal design; (b) I-optimal design; and (c) space-filling design for the glacier area in the case study of Pratola et al. (2017).

At last, we want to point out a close connection between design of experiments and approximate inference in the general sense, meaning that the design space may be regular or irregular. Specifically, this connection is between Bayesian optimal design and MCMC sampling. Based on Bayesian framework, Bayesian optimal design $\boldsymbol{d}^*$ maximizes the expected utility function $U(\boldsymbol{d})$ over the design space $\mathcal{D}$ with respect to future data $\boldsymbol{y}$ and model parameters $\boldsymbol{\theta}$, according to the notation from the review article by Ryan et al. (2016). The utility function is defined by

$$U(\boldsymbol{d}) = \int_{\mathbf{Y}} \int_{\Theta} U(\boldsymbol{d},\boldsymbol{\theta},\boldsymbol{y}) p(\boldsymbol{\theta} \mid \boldsymbol{d},\boldsymbol{y}) p(\boldsymbol{y}|\boldsymbol{d}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{y},$$

where $p(\boldsymbol{\theta} \mid \boldsymbol{d},\boldsymbol{y})$ is the posterior distribution of the parameters and $p(\boldsymbol{y}|\boldsymbol{d})$ is the marginal sampling distribution of the future observations given the design. The utility function $U(\boldsymbol{d},\boldsymbol{\theta},\boldsymbol{y})$ is a user-specified design criterion. For example, $U(\boldsymbol{d},\boldsymbol{\theta},\boldsymbol{y}) = [\det(\mathrm{cov}(\boldsymbol{\theta} \mid \boldsymbol{d},\boldsymbol{y}))]^{-1}$ of the Bayesian D-posterior precision criterion. Since the utility function $U(\boldsymbol{d})$ is based on the posterior distribution, many simulation based methods, such as MCMC, sequential MC, approximate Bayesian computation, and so on, can be used to provide good approximation to the posterior distribution (Amzal et al., 2006; Drovandi & Pettitt, 2013; Drovandi & Tran, 2018; Ryan et al., 2016). HMC-based methods can certainly be a helpful tool for Bayesian optimal design, and constrained HMC methods can be used when $\boldsymbol{\theta}$ have to meet certain constraints.

6 | CONCLUSION

In this article, we review in detail Hamilton Monte Carlo sampling and its variants, and more importantly, how they are modified to overcome various constraints. Based on the nature of the algorithms, we categorize them into three groups, rejection, reflection, and reformulation. Specifically, we explain three constrained HMC-based sampling algorithms, Wall HMC, Constrained HMC, and Sphere HMC. Wall HMC is a reflection-type algorithm and the other two belong to the reformulation group. Three important applications of constrained sampling algorithms are illustrated. They are truncated multivariate Gaussian, Bayesian regularized regression, and non-parametric density estimation.

Constrained sampling is an important problem and has broad applications in many statistical/machine learning models. Some models that are typically not considered to be constrained sampling problems can be solved by constrained sampling algorithms, such as regularized regression/classification and density estimation. With the rapid development of AI and big data technologies, future research on this topic is most likely in the direction of creating fast algorithms for high-dimensional distributions with complex constraints, their theoretical foundations, and software implementations.

AUTHOR CONTRIBUTIONS

Shiwei Lan: Conceptualization (equal); investigation (equal); methodology (equal); project administration (equal); software (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal). **Lulu Kang:** Conceptualization (equal); investigation (equal); methodology (equal); project administration (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

We thank the editor, the associate editor, and reviewers in advance for reviewing this article. We look forward to their comments and suggestions.

FUNDING INFORMATION

Shiwei Lan's work is supported by NSF grant DMS-2134256. Lulu Kang's work is supported by NSF grants DMS-1916467 and DMS-2153029.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Lulu Kang https://orcid.org/0000-0002-6000-3436

RELATED WIREs ARTICLES

Shrinkage and absolute penalty estimation in linear regression models
A personal history of Bayesian statistics
An overview of reciprocal L1-regularization for high dimensional regression data

REFERENCES

Ahn, K., & Chewi, S. (2021). Efficient constrained sampling via the mirror-langevin algorithm. *Advances in Neural Information Processing Systems*, *34*, 28405–28418.

Amzal, B., Bois, F. Y., Parent, E., & Robert, C. P. (2006). Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical Association*, *101*(474), 773–785.

Andersen, H. C. (1983). Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, *52*(1), 24–34.

Betancourt, M. (2011). Nested sampling with constrained hamiltonian Monte Carlo. *AIP Conference Proceedings*, *1305*, 165–172.

Bierkens, J., Fearnhead, P., & Roberts, G. (2019). The zig-zag process and super-efficient sampling for bayesian analysis of big data. *The Annals of Statistics*, *47*(3), 1288–1320.

Brubaker, M. A., Salzmann, M., & Urtasun, R. (2012). A family of mcmc methods on implicitly defined manifolds. In N. D. Lawrence & M. A. Girolami (Eds.), *Proceedings of the fifteenth international conference on artificial intelligence and statistics (AISTATS-12)* (pp. 161–172). ACM.

Byrne, S., & Girolami, M. (2013). Geodesic Monte Carlo on embedded manifolds. *Scandinavian Journal of Statistics*, *40*(4), 825–845.

Chaudhry, S., Lautzenheiser, D., & Ghosh, K. (2021). An efficient scheme for sampling in constrained domains. *arXiv Preprints*, arXiv:2110.10840. https://doi.org/10.48550/ARXIV.2110.10840

Chewi, S., Gerber, P. R., Lu, C., Le Gouic, T., & Rigollet, P. (2022). Rejection sampling from shape-constrained distributions in sublinear time. In International conference on artificial intelligence and statistics, pp. 2249–2265.

Das, N., & Bhattacharya, R. (2020). Optimal transport based filtering with nonlinear state equality constraints. *IFAC-PapersOnLine*, *53*(2), 2373–2378.

Draguljić, D., Santner, T. J., & Dean, A. M. (2012). Noncollapsing space-filling designs for bounded nonrectangular regions. *Technometrics*, *54*(2), 169–178.

Drovandi, C. C., & Pettitt, A. N. (2013). Bayesian experimental design for models with intractable likelihoods. *Biometrics*, *69*(4), 937–948.

Drovandi, C. C., & Tran, M.-N. (2018). Improving the efficiency of fully bayesian optimal design of experiments using randomised quasi-Monte Carlo. *Bayesian Analysis*, *13*(1), 139–162.

Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, *35*(2), 109–135.

Fu, W. J. (1998). Penalized regressions: The bridge versus the lasso. *Journal of Computational and Graphical Statistics*, *7*(3), 397–416.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Gessner, A., Kanjilal, O., & Hennig, P. (2020). Integrals over Gaussians under linear domain constraints. In International conference on artificial intelligence and statistics, pages 2764–2774.

Girolami, M., & Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, *73*(2), 123–214.

Hans, C. (2009). Bayesian lasso regression. *Biometrika*, *96*(4), 835–845.

Held, L., & Holmes, C. C. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, *1*(1), 145–168.

Holbrook, A., Lan, S., Streets, J., & Shahbaba, B. (2020). Nonparametric fisher geometry with application to density estimation. In J. Peters & D. Sontag (Eds.), *Proceedings of the 36th conference on uncertainty in artificial intelligence (UAI), volume 124 of proceedings of machine learning research* (pp. 101–110). PMLR.

Holbrook, A., Lan, S., Vandenberg-Rodes, A., & Shahbaba, B. (2018). Geodesic lagrangian Monte Carlo over the space of positive definite matrices: With application to bayesian spectral density estimation. *Journal of Statistical Computation and Simulation*, *88*(5), 982–1002.

Huang, C., Joseph, V. R., & Ray, D. M. (2021). Constrained minimum energy designs. *Statistics and Computing*, *31*(6), 1–15.

Iserles, A. (1986). Generalized leapfrog methods. *IMA Journal of Numerical Analysis*, *6*(4), 381–392.

Jacob, P. E., Gong, R., Edlefsen, P. T., & Dempster, A. P. (2021). A gibbs sampler for a class of random convex polytopes. *Journal of the American Statistical Association*, *116*(535), 1181–1192.

Jessen, R. J. (1970). Probability sampling with marginal constraints. *Journal of the American Statistical Association*, *65*(330), 776–796.

Kang, L. (2019). Stochastic coordinate-exchange optimal designs with complex constraints. *Quality Engineering*, *31*(3), 401–416.

Kang, L., Roshan Joseph, V., & Brenneman, W. A. (2011). Design and modeling strategies for mixture-of-mixtures experiments. *Technometrics*, *53*(2), 125–136.

Kang, X., Ranganathan, S., Kang, L., Gohlke, J., & Deng, X. (2021). Bayesian auxiliary variable model for birth records data with qualitative and quantitative responses. *Journal of Statistical Computation and Simulation*, *91*(16), 3283–3303.

Kook, Y., Lee, Y. T., Shen, R., & Vempala, S. S. (2022). Sampling with riemannian Hamiltonian Monte Carlo in a constrained space. *arXiv Preprints*, arXiv:2202.01908. https://doi.org/10.48550/ARXIV.2202.01908

Lan, S., Holbrook, A., Elias, G. A., Fortin, N. J., Ombao, H., & Shahbaba, B. (2020). Flexible Bayesian dynamic modeling of correlation and covariance matrices. *Bayesian Analysis*, *15*(4), 1199–1228.

Lan, S., & Shahbaba, B. (2016). *Algorithmic advances in Riemannian geometry and applications*, chapter 2-sampling constrained probability distributions using spherical augmentation. In S. B. Kang (Ed.), *Advances in computer vision and pattern recognition* (1st ed., pp. 25–71). Springer International Publishing.

Lan, S., Stathopoulos, V., Shahbaba, B., & Girolami, M. (2015). Markov chain Monte Carlo from lagrangian dynamics. *Journal of Computational and Graphical Statistics*, *24*(2), 357–378.

Lan, S., Zhou, B., & Shahbaba, B. (2014). Spherical hamiltonian Monte Carlo for constrained target distributions. In E. P. Xing (Ed.), *The 31st international conference on machine learning* (pp. 629–637). PMLR.

Lang, L., Shiang Chen, W., Bakshi, B. R., Goel, P. K., & Ungarala, S. (2007). Bayesian estimation via sequential Monte Carlo sampling—Constrained dynamic systems. *Automatica*, *43*(9), 1615–1622.

Legrain, G. (2021). Non-negative moment fitting quadrature rules for fictitious domain methods. *Computers & Mathematics with Applications*, *99*, 270–291.

Leimkuhler, B., & Reich, S. (1994). Symplectic integration of constrained hamiltonian systems. *Mathematics of Computation*, *63*(208), 589.

Leobacher, G., & Pillichshammer, F. (2014). *Introduction to quasi-Monte Carlo integration and applications*. Springer.

Li, Y., & Ghosh, S. K. (2015). Efficient sampling methods for truncated multivariate normal and student-t distributions subject to linear inequality constraints. *Journal of Statistical Theory and Practice*, *9*(4), 712–732.

Neal, P., & Roberts, G. O. (2008). Optimal scaling for random walk metropolis on spherically constrained target densities. *Methodology and Computing in Applied Probability*, *10*, 277–297.

Neal, P., Roberts, G. O., & Yuen, W. K. (2012). Optimal scaling of random walk metropolis algorithms with discontinuous target densities. *Annals of Applied Probability*, *22*(5), 1880–1927.

Neal, R. M. (1994). An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics*, *111*(1), 194–203.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. Jones, & X. L. Meng (Eds.), *Handbook of Markov chain Monte Carlo* (pp. 113–162). Chapman and Hall/CRC.

Nishimura, A., Zhang, Z., & Suchard, M. A. (2021). Hamiltonian zigzag sampler got more momentum than its markovian counterpart: Equivalence of two zigzags under a momentum refreshment limit. *arXiv Preprints*, arXiv:2104.07694.

Olander, J. (2020). Constrained space mcmc methods for nested sampling Bayesian computations (Master's thesis). Chalmers University of Technology, Gothenburg, Sweden.

Olshanskii, M. A., & Safin, D. (2016). Numerical integration over implicitly defined domains for higher order unfitted finite element methods. *Lobachevskii Journal of Mathematics*, *37*(5), 582–596.

Owen, A. B. (2013). *Monte Carlo theory, methods and examples*. Springer.

Pakman, A., & Paninski, L. (2014). Exact hamiltonian Monte Carlo for truncated multivariate gaussians. *Journal of Computational and Graphical Statistics*, *23*(2), 518–542.

Park, T., & Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, *103*(482), 681–686.

Piepel, G., Cooley, S., & Jones, B. (2005). Construction of a 21-component layered mixture experiment design using a new mixture coordinate-exchange algorithm. *Quality Engineering*, *17*(4), 579–594.

Pratola, M. T., Harari, O., Bingham, D., & Flowers, G. E. (2017). Design and analysis of experiments on nonconvex regions. *Technometrics*, *59*(1), 36–47.

Ryan, E. G., Drovandi, C. C., McGree, J. M., & Pettitt, A. N. (2016). A review of modern computational algorithms for bayesian optimal design. *International Statistical Review*, *84*(1), 128–154.

Saye, R. I. (2022). High-order quadrature on multi-component domains implicitly defined by multivariate polynomials. *Journal of Computational Physics*, *448*, 110720.

Shen, S., Kang, L., & Deng, X. (2020). Additive heredity model for the analysis of mixtureof- mixtures experiments. *Technometrics*, *62*(2), 265–276.

Sherlock, C., & Roberts, G. O. (2009). Optimal scaling of the random walk metropolis on elliptically symmetric unimodal targets. *Bernoulli*, *15*(3), 774–798.

Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis*, *1*(4), 833–859.

Stinstra, E., den Hertog, D., Stehouwer, P., & Vestjens, A. (2003). Constrained maximin designs for computer experiments. *Technometrics*, *45*(4), 340–346.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Verlet, L. (1967). Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physics Review*, *159*(1), 98–103.

Wang, L. (2008). Karhunen-Loeve expansions and their applications (PhD thesis). London School of Economics and Political Science, UK.

Zhang, Z., Nishimura, A., Bastide, P., Ji, X., Payne, R. P., Goulder, P., Lemey, P., & Suchard, M. A. (2021). Large-scale inference of correlation among mixed-type biological traits with phylogenetic multivariate probit models. *The Annals of Applied Statistics*, *15*(1), 230–251.

Zhang, Z., Nishimura, A., trov˜ao, N. S., Cherry, J. L., Holbrook, A. J., Ji, X., Lemey, P., & Suchard, M. A. (2022). Accelerating Bayesian inference of dependency between complex biological traits. *arXiv Preprints*, arXiv:2201.07291.

How to cite this article: Lan, S., & Kang, L. (2023). Sampling constrained continuous probability distributions: A review. *WIREs Computational Statistics*, e1608. https://doi.org/10.1002/wics.1608