

Generating MCMC proposals by randomly rotating the regular simplex

Andrew J. Holbrook

Department of Biostatistics, University of California, Los Angeles, CA 90095, USA

ARTICLE INFO

Article history:

Received 13 October 2021

Received in revised form 13 September 2022

Accepted 15 September 2022

Available online 23 September 2022

AMS 2020 subject classifications:

primary 65C05

secondary 65C60

Keywords:

Haar measure

Markov chain Monte Carlo

Orthogonal group

Parallel MCMC

ABSTRACT

We present the *simplicial sampler*, a class of parallel MCMC methods that generate and choose from multiple proposals at each iteration. The algorithm's multiproposal randomly rotates a simplex connected to the current Markov chain state in a way that inherently preserves symmetry between proposals. As a result, the simplicial sampler leads to a simplified acceptance step: it simply chooses from among the simplex nodes with probability proportional to their target density values. We also investigate a multivariate Gaussian-based symmetric multiproposal mechanism and prove that it also enjoys the same simplified acceptance step. This insight leads to significant theoretical and practical speedups. While both algorithms enjoy natural parallelizability, we show that conventional implementations are sufficient to confer efficiency gains across an array of dimensions and a number of target distributions.

© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Generating random samples from probability distributions is a crucial task in many of the quantitative sciences. Markov chain Monte Carlo (MCMC) algorithms sample from target distributions by constructing Markov chains that maintain their targets as equilibrium distributions. The classic Metropolis–Hastings (MH) algorithm [13,21] builds such a Markov chain by iteratively generating a single proposal state conditioned on the current state and randomly accepting or rejecting that proposal based on a ratio involving the target and proposal densities evaluated at the current and proposed states. The majority of MCMC techniques share this general structure, randomly proposing and accepting a single proposal at each iteration, but some relatively recent methods make use of multiple proposals (or, multiproposals) at each step with the goals of, e.g., reducing autocorrelation between Markov chain states or making better use of parallel computing resources.

Liu et al. [19] develop the first such multiproposal technique, the multiple-try Metropolis algorithm (MTM). Each iteration of MTM starts by proposing multiple states θ_p^* , $p \in \{1, \dots, P\}$ from a proposal distribution $q(\theta^{(s)}, \cdot)$ and choosing a candidate θ^* from among them with probability proportional to their respective target density values $\pi(\theta_p^*)$. After this, the MTM algorithm generates an additional $P - 1$ states $\theta'_1, \dots, \theta'_{p-1}$ from the distribution $q(\theta^*, \cdot)$. In the case that $q(\cdot, \cdot)$ is symmetric, the MTM algorithm then accepts θ^* with probability

$$1 \wedge \frac{\pi(\theta_1^*) + \dots + \pi(\theta_p^*)}{\pi(\theta'_1) + \dots + \pi(\theta'_{p-1}) + \pi(\theta^{(s)})}. \quad (1)$$

Whereas MTM is the most widely used multiproposal MCMC algorithm, it has certain drawbacks. First, to use P proposals, one must generate $O(2P)$ states and evaluate the target density $\pi(\cdot)$ $O(2P)$ times. Second, after randomly selecting a

E-mail address: aholbroo@g.ucla.edu.

potentially good candidate state θ^* , one might nonetheless reject it based on the target density values of the other proposed states within the Metropolis step that uses (1). Third, Yang and Liu [32] rigorously prove superior mixing of MH over MTM albeit in the simplified setting of independence proposals. Perhaps most importantly, Tjelmeland [29] shows that MTM's additional Metropolis step and its extra noise are simply not necessary.

Next, Tjelmeland [29] and Frenkel [9] develop multiproposal MCMC algorithms that do not require the same binary accept-reject step as MTM. Tjelmeland [29] posits a collection of $P + 1$ random variables $\Theta^* = (\theta_1^*, \dots, \theta_{P+1}^*)$ and a probability distribution with density $q(\theta_p^*, \Theta^*)$ that describes the probability of proposing the entire set Θ^* given θ_p^* . Tjelmeland [29] augments the state space with a random variable $p \in \{1, \dots, P + 1\}$ indicating θ_p^* as the current state of the Markov chain and iterates between generating P proposals conditioned on the current state $\theta^{(s)} = \theta_{p+1}^*$ and randomly selecting a new value for p with probabilities

$$\left(\frac{\pi(\theta_1^*)q(\theta_1^*, \Theta^*)}{\sum_{p=1}^{P+1} \pi(\theta_p^*)q(\theta_p^*, \Theta^*)}, \frac{\pi(\theta_2^*)q(\theta_2^*, \Theta^*)}{\sum_{p=1}^{P+1} \pi(\theta_p^*)q(\theta_p^*, \Theta^*)}, \dots, \frac{\pi(\theta_{P+1}^*)q(\theta_{P+1}^*, \Theta^*)}{\sum_{p=1}^{P+1} \pi(\theta_p^*)q(\theta_p^*, \Theta^*)} \right). \quad (2)$$

In addition to these probabilities, Tjelmeland [29] advances two further transition strategies and two structured proposal strategies, which we discuss below. Finally, he proposes building estimators using a weighted average with weights equal to the probabilities (2) and proves that this estimator is consistent assuming the chain converges to equilibrium. Frenkel [9] also uses the probabilities (2) as weights to build the same estimator but in the context of statistical physics. Frenkel [9] provides an informal physics argument for the correctness of his approach, stating that it fails to maintain detailed balance but does maintain 'superdetailed' balance. More recently, Calderhead [6] adds an additional subroutine to the algorithm of Tjelmeland [29] so that, at each MCMC iteration, one simulates a random walk on the finite state space composed of the $P + 1$ states using transition probabilities closely related to (2). Ultimately, however, Yang et al. [31] and Schwedes and Calderhead [26] show that the weighted estimators of Tjelmeland [29] and Frenkel [9] constitute Rao-Blackwellizations over the index variable p within the framework of Calderhead [6] and thus enjoy reduced variance.

Despite theoretical results favoring the use of weighted estimators that feature all proposals at each step, there are practical downsides to this approach. Here, one must: either (1) save a massive $S \times D \times (P + 1)$ tensor to memory, where S is the total number of MCMC iterations; or (2) know what the estimand of interest is beforehand and maintain a running average of each MCMC iteration's contribution to the weighted estimator. Importantly, Schwedes and Calderhead [26] empirically demonstrate that the non-Rao-Blackwellized estimators of Calderhead [6] can perform poorly when a small number of jumps occur between the $P + 1$ states of the proposal set. On the one hand, this result is disconcerting because their experimental baseline of $1 \times P$ jumps for $P = 1000$ proposals performs much worse than $16 \times P$ jumps for the same number of proposals. Time spent iterating between proposed states is time not spent exploring the target distribution. On the other hand, these empirical results rely on a specific multiproposal scheme in which

$$\theta_1^*, \dots, \theta_P^* \stackrel{iid}{\sim} N_D(\theta_{P+1}^*, \Sigma) \quad (3)$$

and do not necessarily extend to other multiproposal strategies.

With this in mind, we consider it worthwhile to further develop structured multiproposal methods that maintain scalability while not requiring many jumps within the proposal set. As mentioned above, Tjelmeland [29] proposes two strategies. Proposal alternative 1, or 'P1', first generates a Gaussian random variable θ_0 centered at the current state θ_{P+1}^* and next generates P proposals centered at θ_0 :

$$\theta_1^*, \dots, \theta_P^* \stackrel{iid}{\sim} N_D(\theta_0, \Sigma), \quad \theta_0 \sim N_D(\theta_{P+1}^*, \Sigma). \quad (4)$$

Here, the two D -dimensional covariance matrices take the same form so that all $\Theta^* = (\theta_1^*, \dots, \theta_{P+1}^*)$ share the same expected distance from the center θ_0 . In some ways, the recent work of Luo and Tjelmeland [20] builds on this proposal strategy. They extend this multiproposal's simple star-shaped structure, where θ_0 is the center and $(\theta_1^*, \dots, \theta_{P+1}^*)$ are the flairs, to general acyclic graphs and tailor their proposals to models with certain discrete structures. Unlike his first proposal alternative, Tjelmeland [29]'s Proposal alternative 2, or 'P2', only applies to the scenario when $P = 2$ and constructs proposal sets $\Theta^* = (\theta_1^*, \theta_2^*, \theta_3^*)$ that are equidistant from a center θ_0 and 'maximally spread'. Setting θ_3^* to equal $\theta^{(s)}$, the current state of the Markov chain, this scheme satisfies the following relations:

$$\begin{aligned} \theta_1^* &= \theta_0 + n\mathbf{v}_1, \quad \theta_2^* = \theta_0 - \frac{n}{2}\mathbf{v}_1 + \frac{\sqrt{3}n}{2}\mathbf{v}_2, \quad \theta_0 = \theta_3^* + \frac{n}{2}\mathbf{v}_1 + \frac{\sqrt{3}n}{2}\mathbf{v}_2, \\ \mathbf{v}_1 &= \mathbf{u}_1, \quad \mathbf{v}_2 = \frac{(\mathbf{I}_D - \mathbf{u}_1\mathbf{u}_1^T)\mathbf{u}_2}{\|(\mathbf{I}_D - \mathbf{u}_1\mathbf{u}_1^T)\mathbf{u}_2\|_2}, \quad n \sim N(0, \sigma^2/3), \quad \mathbf{u}_1, \mathbf{u}_2 \stackrel{iid}{\sim} \text{Uniform}(\mathcal{S}^{D-1}). \end{aligned} \quad (5)$$

In words, one first generates two independent random variables that are each uniformly distributed on the sphere and forces them to be orthogonal to each other. Next, one generates the normally distributed n and obtains θ_0 followed by θ_1^* and θ_2^* . Straightforward calculations show that the proposals θ_1^* , θ_2^* and θ_3^* are indeed the same distance from the center θ_0 . Similar calculations show that the proposals θ_1^* , θ_2^* and θ_3^* are all equidistant from each other and together constitute an equilateral triangle with random edge lengths.

Tjelmeland [29] does not specify how one might extend his second proposal alternative to accommodate multiproposals when $P > 2$. In the following, we accomplish this multivariate generalization by first noting that the equilateral

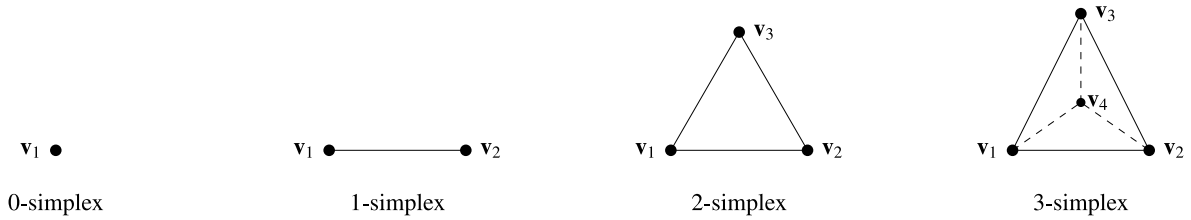


Fig. 1. Regular simplices. The 0-simplex is a point; the 1-simplex is a line connecting two points; the regular 2-simplex is an equilateral triangle; and the regular 3-simplex is a tetrahedron, the faces of which are equilateral triangles.

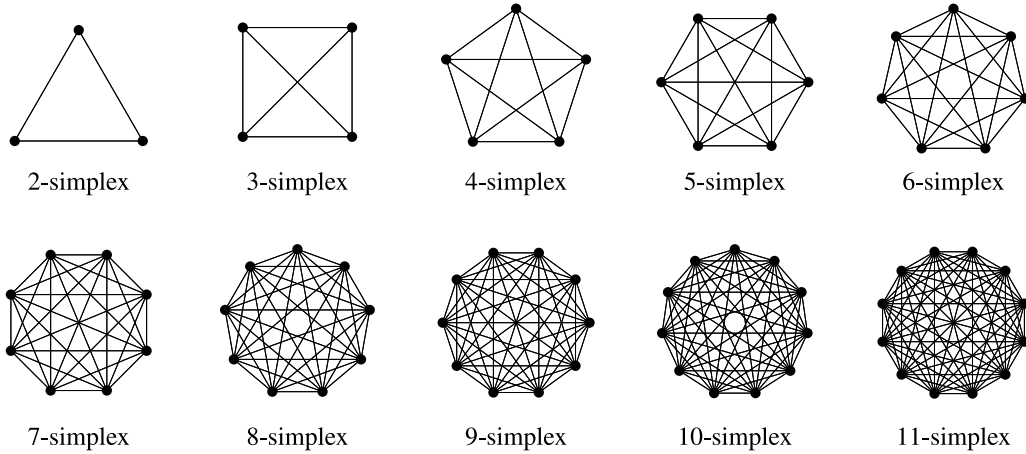


Fig. 2. Petrie polygon projections [7] of higher-dimensional, regular simplices preserve symmetry but not distances.

triangle is a regular 2-simplex. Generally, a D -dimensional space can hold as many as $D + 1$ points that are pairwise equidistant. If $\bar{D} \leq D + 1$, the points $\mathbf{v}_1, \dots, \mathbf{v}_{\bar{D}} \in \mathbb{R}^D$ that satisfy

$$\|\mathbf{v}_d - \mathbf{v}_{d'}\|_2 = \begin{cases} \lambda > 0, & d \neq d', \\ 0, & d = d', \end{cases}$$

identify a regular polytope called a $\bar{D}$ -simplex. Fig. 1 shows the regular simplices that fit in 3-dimensional space. Fig. 2 shows Petrie polygon projections [7] that are helpful for visualizing higher dimensional regular simplices. While we use the regular simplex as our starting point, the multistep nature of (5) makes it unclear how one might actually generate proposals when $P > 2$, an endeavor that doubtless involves more random variables with more complicated relationships. To preserve simplicity in this multivariate setting, in Section 3 we generate proposals using a matrix distribution over the orthogonal group $\mathcal{O}_D$, which consists of $D \times D$ matrices $\mathbf{Q}$ satisfying $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_D$. Because $\mathcal{O}_D$ is a compact topological group, it admits a unique normalized, left-invariant measure $\mathcal{H}(\cdot)$ called the Haar measure [8]. In symbols,

$$\mathcal{H}(\mathcal{O}_D) = 1, \quad \mathcal{H}(\mathbf{Q}\mathcal{B}) = \mathcal{H}(\mathcal{B})$$

for $\mathbf{Q} \in \mathcal{O}_D$ and $\mathcal{B}$ a Borel set on $\mathcal{O}_D$. The Haar measure on $\mathcal{O}_D$ is not only useful for constructing simple multivariate proposals: its uniformity helps maintain a simplified acceptance step as well (theorem 2).

Bearing in mind the poor performance of simple multiproposal strategies when the number of jumps between proposals is small [26], we are motivated to develop structured multiproposal strategies that do not require Rao-Blackwellization, or many inter-proposal jumps, and thus spend more time exploring the entire target distribution. We therefore develop a multivariate multiproposal framework that maintains the principles of Tjelmeland [29]'s second proposal strategy while also achieving greater simplicity. In the following, we develop the simplicial sampler and its extensions, prove algorithmic correctness and demonstrate usefulness of our approach on a range of target distributions. We start by deriving sufficient conditions for simplifying the acceptance probabilities (2).

2. Simplified acceptance probabilities

We wish to sample from the target distribution with density $\pi(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \mathbb{R}^D$. Given the current Markov chain state $\boldsymbol{\theta}^{(s)}$, we use the following steps to advance the Markov chain.

1. Propose states $\Theta^* = (\theta_1^*, \dots, \theta_p^*, \theta_{p+1}^*) = (\theta_1^*, \dots, \theta_p^*, \theta^{(s)})$ according to the distribution $q(\theta^{(s)}, \Theta^*)$;
2. Calculate densities $\pi(\theta_p^*)$ for $p \in \{1, \dots, P+1\}$;
3. Randomly select a single state θ_p^* from the set Θ^* according to the simplified selection probabilities

$$(\pi_1(\Theta^*), \dots, \pi_{P+1}(\Theta^*)) = \left(\frac{\pi(\theta_1^*)}{\sum_{p=1}^{P+1} \pi(\theta_p^*)}, \dots, \frac{\pi(\theta_{P+1}^*)}{\sum_{p=1}^{P+1} \pi(\theta_p^*)} \right); \quad (6)$$

4. Set $\theta^{(s+1)} = \theta_p^*$.

This transition rule uses a modified acceptance probability (2) that does not include the proposal density terms. As such, it rarely leaves the target distribution invariant. We require a further sufficient condition to guarantee that $\pi(\theta)$ is a stationary distribution for our Markov chain.

Proposition 1 (Tjelmeland 29). *The Markov chain with the above transition rule maintains detailed balance with respect to the target distribution $\pi(\theta)$ if $q(\cdot, \cdot)$ satisfies*

$$q(\theta_1^*, \Theta^*) = \dots = q(\theta_p^*, \Theta^*) = \dots = q(\theta_{P+1}^*, \Theta^*), \quad (7)$$

and so $\pi(\theta)$ is a stationary distribution of the chain.

Proof. Let $p(\theta, d\theta)$ be the transition kernel associated with the transition from an arbitrary state θ according to the above steps 1–4. Then

$$p(\theta, d\theta) = \sum_{p=1}^{P+1} \int_{(\mathbb{R}^D)^{\otimes P}} \pi_p(\Theta) \delta_{\theta_p}(d\theta) q(\theta, \Theta) d\theta_1 \dots d\theta_p.$$

For any two states $\theta, \tilde{\theta}$, the following holds:

$$\begin{aligned} \pi(\theta) d\theta p(\theta, d\tilde{\theta}) &= \pi(\theta) d\theta \sum_{p=1}^{P+1} \int_{(\mathbb{R}^D)^{\otimes P}} \pi_p(\Theta^*) \delta_{\theta_p^*}(d\tilde{\theta}) q(\theta, \Theta^*) d\theta_1^* \dots d\theta_p^* \\ &= \pi(\theta) d\theta \sum_{p=1}^{P+1} \int_{(\mathbb{R}^D)^{\otimes P}} \frac{\pi(\tilde{\theta})}{\sum_{p'=1}^{P+1} \pi(\theta_{p'}^*)} d\tilde{\theta} q(\theta, \Theta^*) \prod_{p' \neq p} d\theta_{p'}^* \\ &= \pi(\tilde{\theta}) d\tilde{\theta} \sum_{p=1}^{P+1} \int_{(\mathbb{R}^D)^{\otimes P}} \frac{\pi(\theta)}{\sum_{p'=1}^{P+1} \pi(\theta_{p'}^*)} d\theta q(\tilde{\theta}, \Theta^*) \prod_{p' \neq p} d\theta_{p'}^* \\ &= \pi(\tilde{\theta}) d\tilde{\theta} \sum_{p=1}^{P+1} \int_{(\mathbb{R}^D)^{\otimes P}} \pi_p(\Theta^*) \delta_{\theta_p^*}(d\theta) q(\tilde{\theta}, \Theta^*) d\theta_1^* \dots d\theta_p^* \\ &= \pi(\tilde{\theta}) d\tilde{\theta} p(\tilde{\theta}, d\theta), \end{aligned}$$

where we use assumption (7) in the third line. $\square$

Although the proof of Proposition 1 is new, the result itself is a special case of a result from Tjelmeland [29], who argues that his method constitutes a Gibbs sampler that leaves the joint and marginal distributions of $\pi(\theta)q(\theta, \Theta^*)$ invariant. Proposition 1 underlines what one may readily gather from the acceptance probabilities (2). Namely, it may sometimes prove useful to design multiproposals that maintain criterion (7) insofar as such proposals obviate the need for computing proposal densities. In the following, we show that the Gaussian centered multiproposal (4) provides just such a multiproposal mechanism. Under this scheme, consider the DP -vector $\theta_{-p}^* := \text{vec}(\Theta_{-p}^*)$ obtained by removing the state θ_p^* , or the p th column, from the proposal set Θ^* and applying the vectorization operator. Integrating over all possible θ_0 gives the conditional distribution of θ_{-p}^* given θ_p^* :

$$\theta_{-p}^* \sim N_{DP} \left(\mathbf{1}_p \otimes \theta_p^* = \begin{pmatrix} \theta_p^* \\ \theta_p^* \\ \vdots \\ \theta_p^* \end{pmatrix}, (\mathbf{1}_p \mathbf{1}_p^T + \mathbf{I}_p) \otimes \Sigma = \begin{pmatrix} 2\Sigma & \Sigma & \dots & \dots & \Sigma \\ \Sigma & 2\Sigma & \Sigma & \dots & \Sigma \\ \vdots & \Sigma & 2\Sigma & & \vdots \\ \vdots & \vdots & & \ddots & \Sigma \\ \Sigma & \Sigma & \dots & \Sigma & 2\Sigma \end{pmatrix} \right).$$

If one does not adapt the $D \times D$ proposal covariance Σ , then one may precompute the $DP \times DP$ covariance matrix inverse in $O(D^3 P^3)$ time and store it using $O(D^3 P^3)$ memory. In high-dimensional settings where such storage is not possible, one

may store only the $D \times D$ inverse covariance Σ^{-1} and fill the $DP \times DP$ inverse covariance at each step with complexity $O(D^2P^2)$ using the formula $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$. Once one has this inverse, the inner product term in each

$$q(\theta_p^*, \Theta^*) \propto \exp\left(-\frac{1}{2}(\theta_{-p}^* - \theta_p^*)^T ((\mathbf{1}_p \mathbf{1}_p^T + \mathbf{I}_p) \otimes \Sigma)^{-1} (\theta_{-p}^* - \theta_p^*)\right)$$

requires $O(D^2P^2)$ floating point operations. The upshot is an $O(D^2P^3)$ cost to compute all $P+1$ proposal densities $q(\theta_p^*, \Theta^*)$. But if one adapts Σ following Haario et al. [12], then this acceptance mechanism requires: an $O(D^3)$ matrix inversion for the updated Σ at each step; an $O(P^2D^2)$ Kronecker product to construct the $DP \times DP$ covariance matrix; and the $O(D^2P^3)$ floating point operations to compute the $P+1$ proposal densities $q(\theta_p^*, \Theta^*)$. Thankfully, it turns out, none of these computations are necessary.

Lemma 1. *The Gaussian centered multiproposal $q(\theta^{(s)}, \Theta^*)$ defined by*

$$\theta_1^*, \dots, \theta_p^* \stackrel{iid}{\sim} N_D(\theta_0, \Sigma), \quad \theta_0 \sim N_D(\theta_{p+1}^*, \Sigma).$$

satisfies symmetry relation (7), namely

$$q(\theta_1^*, \Theta^*) = \dots = q(\theta_p^*, \Theta^*) = \dots = q(\theta_{p+1}^*, \Theta^*).$$

Proof. Let $p(\cdot, \cdot)$ denote the generic probability density function evaluated at the right and conditioned on the left, and let $\phi(\cdot, \cdot)$ an analogous multivariate Gaussian density with dependence on the covariance suppressed. For two arbitrary indices $p, p' \in \{1, \dots, P+1\}$,

$$\begin{aligned} q(\theta_p^*, \Theta^*) - q(\theta_{p'}^*, \Theta^*) &= \int_{\mathbb{R}^D} p(\theta_p^*, \theta_0) p(\theta_0, \theta_{-p}^*) d\theta_0 - \int_{\mathbb{R}^D} p(\theta_{p'}^*, \theta_0) p(\theta_0, \theta_{-p'}^*) d\theta_0 \\ &= \int_{\mathbb{R}^D} (p(\theta_p^*, \theta_0) p(\theta_0, \theta_{-p}^*) - p(\theta_{p'}^*, \theta_0) p(\theta_0, \theta_{-p'}^*)) d\theta_0 \\ &= \int_{\mathbb{R}^D} \left(\phi(\theta_p^*, \theta_0) \prod_{p'' \neq p} \phi(\theta_0, \theta_{p''}^*) - \phi(\theta_{p'}^*, \theta_0) \prod_{p'' \neq p'} \phi(\theta_0, \theta_{p''}^*) \right) d\theta_0 \\ &= \int_{\mathbb{R}^D} \left(\prod_{p''=1}^{P+1} \phi(\theta_0, \theta_{p''}^*) - \prod_{p''=1}^{P+1} \phi(\theta_0, \theta_{p''}^*) \right) d\theta_0 = 0. \end{aligned}$$

In the third line, we use the Gaussianity and conditional independence structure of the multiproposal. In the fourth line, we use the symmetry of the Gaussian density function. $\square$

This simple result fills in a gap in Tjelmeland [29]: namely, it relates the symmetry in a proposal strategy to a simplified acceptance mechanism (6). Moreover, subsequent work [11] builds on this result by examining cases in which structured proposals lead to simplified acceptance probabilities. In the following, we develop a multivariate extension of the multiproposal (5) and show that it also satisfies symmetry relation (7) and thus also qualifies for the simplified acceptance mechanism (6).

3. Simplicial sampling

We develop a proposal mechanism and a few variations that satisfy symmetry relation (7) and help our basic algorithm preserve reversibility with respect to the target $\pi(\theta)$. This method uses certain mathematical objects introduced in Section 1: namely, the regular D -simplex, the orthogonal group $\mathcal{O}_D$ and the uniform Haar measure over the orthogonal group $\mathcal{H}(\mathcal{O}_D)$. Start with a regular simplex identified by a set of fixed vertices $\mathbf{v}_1, \dots, \mathbf{v}_{D+1} \in \mathbb{R}^D$ that satisfies $\mathbf{v}_{D+1} = \mathbf{0}$. Equating the number of new proposals P with the state space dimension D , the simplicial sampler generates and chooses from proposals $\theta_1^*, \dots, \theta_{D+1}^*$ as follows.

1. Sample $\mathbf{Q} \sim \mathcal{H}(\mathcal{O}_D)$.
2. Rotate and translate the simplicial vertices $(\mathbf{v}_1, \dots, \mathbf{v}_D, \mathbf{0}) \mapsto \mathbf{Q}(\mathbf{v}_1, \dots, \mathbf{v}_D, \mathbf{0}) + \theta^{(s)} =: (\theta_1^*, \dots, \theta_{D+1}^*)$.
3. Draw a single sample θ_d^* from $(\theta_1^*, \dots, \theta_{D+1}^*)$ with probability proportional to $\pi(\theta_d^*)$.
4. Set $\theta^{(s+1)} = \theta_d^*$.

In short, the simplicial sampler advances its Markov chain by randomly rotating a set of D equidistant vertices about the current state $\theta^{(s)}$ (Fig. 3) and randomly selecting the next state from among all $D+1$ vertices of the resulting simplex.

A modified simplicial sampler that only generates a single proposal reduces to Barker's acceptance criterion [3]

$$\pi(\theta^*) / (\pi(\theta^*) + \pi(\theta^{(s)}))$$

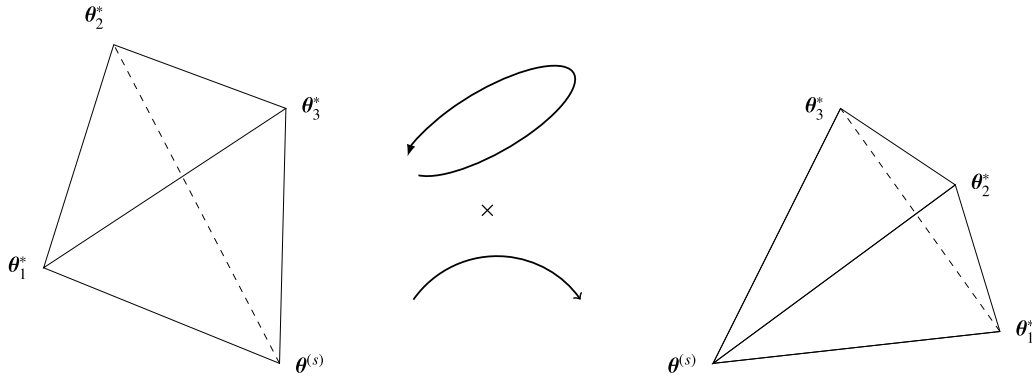


Fig. 3. A simplicial sampling multiproposal for $D = 3$. A proposal set is obtained by rotating three simplex vertices about current state $\theta^{(s)}$.

with a symmetric proposal distribution. Just as Barker's criterion is suboptimal with respect to the Peskun ordering [23], our criterion is also suboptimal. With this in mind, Tjelmeland [29] provides an algorithm that manipulates the acceptance probabilities (2), of which our simplified acceptance probabilities are a special case.

Importantly, the random rotation step satisfies symmetry relation (7).

Lemma 2. The simplicial sampler multiproposal $q(\theta^{(s)}, \Theta^*)$ defined by

$$(\mathbf{v}_1, \dots, \mathbf{v}_D, \mathbf{0}) \mapsto \mathbf{Q}(\mathbf{v}_1, \dots, \mathbf{v}_D, \mathbf{0}) + \theta^{(s)} =: (\theta_1^*, \dots, \theta_{D+1}^*) \text{ for } \mathbf{Q} \sim \mathcal{H}(\mathcal{O}_D)$$

satisfies symmetry relation (7), namely

$$q(\theta_1^*, \Theta^*) = \dots = q(\theta_d^*, \Theta^*) = \dots = q(\theta_{D+1}^*, \Theta^*).$$

Proof. Start by translating coordinates so that $\theta^{(s)} = \mathbf{0}$. We want to show that $q(\mathbf{0}, \Theta^*) = q(\theta_d^*, \Theta^*)$ for an arbitrary $\theta_d^* \in \Theta^*$. We focus on the case $d \leq D$, since the alternative case is trivial. Since the Haar measure $\mathcal{H}(\mathcal{O}_D)$ is uniform on $\mathcal{O}_D$, we only need to show the existence of a rotation $\mathbf{Q}_d \in \mathcal{O}_D$ that maps

$$(\mathbf{v}_1, \dots, \mathbf{v}_D) \mapsto (\theta_1^*, \dots, \theta_{d-1}^*, \mathbf{0}, \theta_{d+1}^*, \dots, \theta_D^*).$$

The key is recognizing that the unique reflection $\mathbf{R}$ that exchanges $\mathbf{0}$ and θ_d^* leaves all other vertices unchanged because they are equidistant from $\mathbf{0}$ and θ_d^* and inhabit the hyperplane that splits them evenly. In symbols, $\mathbf{R}\theta_{d'}^* = \theta_{d'}^*$ for $d' \neq d$. Thus

$$\mathbf{Q}^T \mathbf{R} \theta_{d'}^* = \mathbf{Q}^T \theta_{d'}^* = \mathbf{v}_{d'}, \quad \forall d' \neq d,$$

and

$$\mathbf{Q}^T \mathbf{R} \mathbf{0} = \mathbf{Q}^T \theta_d^* = \mathbf{v}_d.$$

It follows that $(\mathbf{Q}^T \mathbf{R})^T = \mathbf{R} \mathbf{Q}$ is the desired rotation matrix $\mathbf{Q}_d$. $\square$

Fig. 4 provides a stylized visualization of this argument in the 2D setting when $\mathbf{Q}$ represents a 60° rotation clockwise about $\theta^{(s)}$. Taken together, Lemma 2 and Proposition 1 assert that the simplicial sampler maintains detailed balance with respect to the target distribution.

Theorem 1. If the simplicial sampler's multiproposal is π -irreducible and absolutely continuous with respect to $\pi(\theta)$, then $\pi(\theta)$ is the unique stationary and limiting distribution of the Markov chain, and the strong law of large numbers holds.

Proof. The result is a simple combination of the sampler's reversibility with respect to $\pi(\theta)$ and the aperiodicity that arises from its self-loop, i.e., the ability to set $\theta^{(s+1)}$ to $\theta^{(s)}$ [28]. $\square$

While π -irreducibility is a major hypothesis of Theorem 1, we do not formally prove that the simplicial sampler exhibits generally irreducible behavior. That said, there are two intuitive reasons for thinking that the sampler produces generally irreducible Markov chains. First, the simplicial sampler can return to an arbitrarily small ball around a given state within just two steps with positive probability: the first iteration accepts a state λ away from the current state, and the second iteration accepts a state in the neighborhood of the preceding state. Second, the simplicial sampler can amble arbitrarily far away from a given state in the manner that a distracted geometry student can walk a compass clear off their desk. The need for a formally justified sampler leads us to randomly scale the proposal simplex in Section 3.1.1.

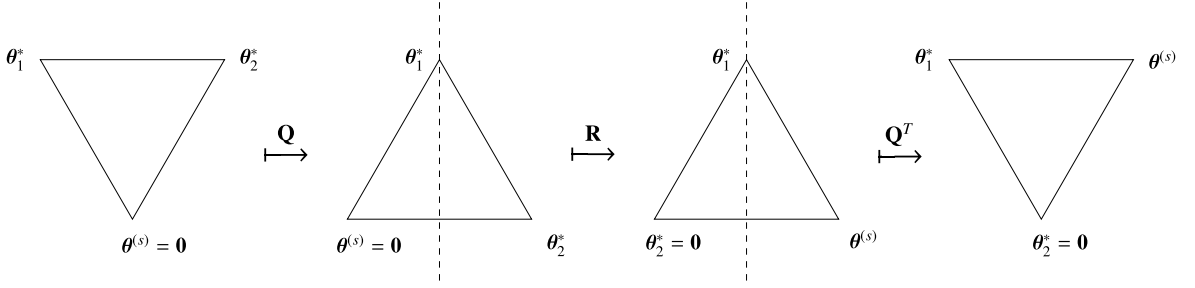


Fig. 4. Illustrating key argument of Lemma 2 for $D = 2$. $\mathbf{Q}$ is the simplicial sampler's random rotation about $\theta^{(s)}$ and $\mathbf{R}$ is a reflection that exchanges a given proposal θ_2^* with the current state $\theta^{(s)}$ while preserving all other proposals.

3.1. Variations

3.1.1. Gaussian simplicial sampling

If we scale the simplex edge lengths by the square root of a chi-square random variable, the proposals will have multivariate Gaussian marginal distributions. The orthogonal group $\mathcal{O}_D$ acts transitively on the sphere S^{D-1} , i.e., it takes any point on S^{D-1} to any other [17]. For this reason, left multiplying a single point on S^{D-1} by a Haar distributed rotation matrix $\mathbf{Q} \sim \mathcal{H}(\mathcal{O}_D)$ is one method for drawing a sample from the sphere's uniform distribution. On the other hand, it is a well known fact that one may obtain a standard D -dimensional Gaussian vector by uniformly sampling a single point on the sphere S^{D-1} and scaling that point by the square root of an independently sampled χ^2 random variable with D degrees of freedom. Randomly scaling the simplicial edge length in a similar manner provides the following variation that we call the Gaussian simplicial sampler.

1. Independently sample $r \sim \chi^2(D)$ and $\mathbf{Q} \sim \mathcal{H}(\mathcal{O}_D)$.
2. Rotate, scale and translate the vertices $(\mathbf{v}_1, \dots, \mathbf{v}_D, \mathbf{0}) \mapsto \sqrt{r}\mathbf{Q}(\mathbf{v}_1, \dots, \mathbf{v}_D, \mathbf{0}) + \theta^{(s)} =: (\theta_1^*, \dots, \theta_{D+1}^*)$.
3. Draw a single sample θ_d^* from $(\theta_1^*, \dots, \theta_{D+1}^*)$ with probability proportional to $\pi(\theta_d^*)$.
4. Set $\theta^{(s+1)} = \theta_d^*$.

By the above discussion, each of the proposals θ_d^* for $d \in \{1, \dots, D\}$ has a Gaussian marginal distribution with mean $\theta^{(s)}$ and covariance matrix $\lambda^2 \mathbf{I}_D$, where λ is still the simplicial edge length. Because of its random scaling, a stronger result than Theorem 1 applies to the Gaussian simplicial sampler.

Corollary 1. Any distribution $\pi(\theta)$ with support $\mathbb{R}^D$ is the unique stationary and limiting distribution of the Gaussian simplicial sampler that targets it, and the strong law of large numbers holds.

Proof. All that needs to be established is π -irreducibility and absolute continuity. The latter condition is accomplished by hypothesis. The former condition is accomplished by the sufficient condition [28] that

$$\int_A q(\theta_0, \Theta^*) d\theta_d^* > 0, \quad \text{if} \quad \int_A \pi(\theta) d\theta > 0, \quad d \in \{1, \dots, D\}, \quad A \subset \mathbb{R}^D.$$

This condition is itself accomplished by the unbounded nature of the multiproposal's χ^2 scaling. All edges are scaled by the same random variable, and the symmetry of the proposal simplex is therefore maintained. Because the random scaling is independent of the random rotation, the conditions within the proof of Lemma 2 remain exactly the same. $\square$

In Section 5, we show that its random scaling gives the Gaussian simplicial sampler improved performance for a multimodal target distribution.

3.1.2. Preconditioned simplicial sampling

For $\mathbf{C} \succ \mathbf{0}$, a $D \times D$ and real-valued matrix, the preconditioned simplicial sampler takes the following steps.

1. Sample $\mathbf{Q} \sim \mathcal{H}(\mathcal{O}_D)$.
2. Rotate, precondition and translate the vertices $(\mathbf{v}_1, \dots, \mathbf{v}_D, \mathbf{0}) \mapsto \mathbf{C}^{1/2}\mathbf{Q}(\mathbf{v}_1, \dots, \mathbf{v}_D, \mathbf{0}) + \theta^{(s)} =: (\theta_1^*, \dots, \theta_{D+1}^*)$.
3. Draw a single sample θ_d^* from $(\theta_1^*, \dots, \theta_{D+1}^*)$ with probability proportional to $\pi(\theta_d^*)$.
4. Set $\theta^{(s+1)} = \theta_d^*$.

The matrix $\mathbf{C}^{1/2}\mathbf{Q}$ is itself a rotation in the Euclidean space equipped with inner product $\langle \mathbf{v}, \mathbf{v}' \rangle_{\mathbf{C}^{-1}} := \mathbf{v}^T \mathbf{C}^{-1} \mathbf{v}'$ (Fig. 5).

Corollary 2. The guarantees of Theorem 1 apply to the preconditioned simplicial sampler as well.

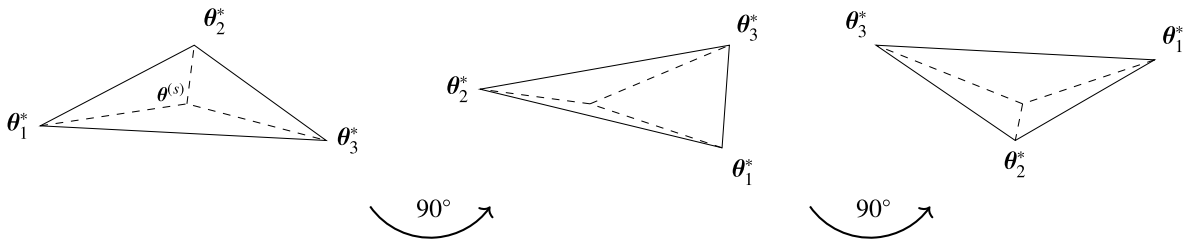


Fig. 5. The preconditioned simplicial sampler's rotations about $\theta^{(s)}$ with respect to a non-standard Euclidean metric. To a viewer relying on the standard Euclidean metric, fixed-length edges appear to lengthen and shorten as the simplex is rotated 90° according to the standard metric about the axis perpendicular to the page.

Proof. The preconditioned simplicial sampler maintains detailed balance with respect to its target distribution $\pi(\theta)$ because it satisfies symmetry relation (7). The proof takes the exact same form as that of Lemma 2, replacing $\mathbf{Q}_d$ with $\mathbf{C}^{1/2}\mathbf{Q}_d$. $\square$

Corollaries 1 and 2 combine to guarantee a unique stationary distribution and strong law of large numbers for a preconditioned Gaussian simplicial sampler, and we explore this algorithm in Section 5.

3.1.3. Adaptive simplicial sampling

As has been established for other MCMC algorithms, it is often useful to adapt the simplicial sampler multiproposal. We employ two distinct methods of adaptation in Section 5. Both methods use the diminishing adaptation strategy [24], which sacrifices guarantees of a strong law of large numbers but maintains a weak law. The first adapts all simplex edge lengths while keeping them equal so as to achieve a desired ‘acceptance’ rate. Although the simplicial sampler does not use a classical accept-reject step, we can check whether state $\theta^{(s+1)}$ equals $\theta^{(s)}$ and call this an acceptance or rejection accordingly. The second adapts the preconditioning matrix $\mathbf{C}$ following the proposal covariance adaptation outlined in Haario et al. [12]. In practice, we find it useful to combine these two adaptation schemes when implementing the preconditioned simplicial sampler.

3.1.4. Extra-dimensional simplicial sampling

The final variation for the simplicial sampler uses a number of proposals that is larger than the dimensionality of the target distribution, i.e., $P > D$. Let $\mathbf{W} = (\mathbf{I}_{D \times D}, \mathbf{0}_{D \times (P-D)})$ be the wide $D \times P$ matrix for which multiplication selects the first D elements of any P -vector. Letting each $\mathbf{v}_p$, $p \in \{1, \dots, P\}$, be a P -dimensional vector, the extra-dimensional simplicial sampler works as follows.

1. Sample $\mathbf{Q} \sim \mathcal{H}(\mathcal{O}_P)$.
2. Rotate, precondition and translate the vertices $(\mathbf{v}_1, \dots, \mathbf{v}_P, \mathbf{0}_P) \mapsto \mathbf{Q}(\mathbf{v}_1, \dots, \mathbf{v}_P, \mathbf{0}_P) + (\theta^{(s)T}, \mathbf{0}_{P-D}^T)^T =: (\theta_1^*, \dots, \theta_{P+1}^*)$.
3. Draw a single sample $\mathbf{W}\theta_p^*$ from $\mathbf{W}(\theta_1^*, \dots, \theta_{P+1}^*)$ with probability proportional to $\pi(\mathbf{W}\theta_p^*)$.
4. Set $\theta^{(s+1)} = \mathbf{W}\theta_p^*$.

One may interpret the application of $\mathbf{W}$ as a form of preconditioning with preconditioner $\mathbf{C}^{1/2} = \mathbf{W}$. In Section 3.1.2, we argue that the guarantees of Lemma 2 apply to the preconditioned simplicial sampler because $\mathbf{C}^{1/2}\mathbf{Q}$ is a random rotation in Euclidean space with inner product $\langle \mathbf{v}, \mathbf{v}' \rangle_{\mathbf{C}^{-1}} := \mathbf{v}^T \mathbf{C}^{-1} \mathbf{v}'$. Following this analogy, we have

$$\mathbf{C} = \mathbf{W}^T \mathbf{W} = \begin{pmatrix} \mathbf{I}_D & \mathbf{0}_{D \times (P-D)} \\ \mathbf{0}_{(P-D) \times D} & \mathbf{0}_{(P-D) \times (P-D)} \end{pmatrix} = \mathbf{C}^+,$$

for $\mathbf{C}^+$ the Moore–Penrose inverse [14]. Due to the degeneracy of $\mathbf{C}^+$, the resulting form $\langle \mathbf{v}, \mathbf{v}' \rangle_{\mathbf{C}^+} := \mathbf{v}^T \mathbf{C}^+ \mathbf{v}'$ is no longer an inner product, and we cannot assert that the extra-dimensional algorithm inherits the promises of Lemma 2. Nonetheless, limited empirical experiments do show the algorithm providing accurate results, and we find it helpful to visualize the sampler in Section 5.3.

4. Computational considerations

The real-world performance of an algorithm depends on its computational complexity, the language it is written in and the hardware it is implemented on. It also depends on how well one codes it up, e.g., efficient use of random access memory and parallelization when available. We separate these considerations into multiproposal and acceptance mechanisms, and the exact weight one assigns to the two steps will depend on the computational complexity of the target distribution.

4.1. Multiproposal mechanisms

4.1.1. The simplicial sampler multiproposal

The simplicial sampler requires the $O(D^3)$ generation of a random rotation $\mathbf{Q} \sim \mathcal{H}(\mathcal{O}_D)$. The most common way to perform this step is to generate a $D \times D$ matrix of standard normals and then perform the QR factorization. Once one has the orthogonal ‘Q’ matrix, one right multiplies it by a diagonal matrix consisting of the signs of the diagonal of the ‘R’ matrix, guaranteeing that the obtained matrix $\mathbf{Q}$ is the unique orthogonal matrix corresponding to the $\mathbf{R}$ with positive diagonal elements. Thus, one obtains $\mathbf{Q} \sim \mathcal{H}(\mathcal{O}_D)$ using the $O(D^3)$ QR factorization. There are additional algorithms for sampling from the Haar distribution over $\mathcal{O}_D$ that provide speedups while still maintaining cubic complexity [2,27]. In particular, Stewart [27] achieves a 50% speedup over the QR factorization based algorithm by iteratively pre-multiplying random Householder matrices of increasing dimensions. We are interested in the fact that, when one only requires the random rotation and not the matrix itself, Stewart [27]’s algorithm only requires $O(D^2)$ floating point operations to randomly rotate a D -vector. On the one hand, the simplicial sampler multiproposal still requires $O(D^3)$ floating point operations because it must apply this random rotation to the D D -vectors that identify the regular simplex. On the other hand, one could apply this algorithm in an embarrassingly parallel manner to the D non-zero D -vectors that contribute to the regular simplex. Furthermore, the entirety of the algorithm relies on matrix–matrix multiplications and should be extremely fast and scalable on a modern GPU: Li et al. [18] use a (now outdated) Tesla C1060 GPU to score 10,000-fold speedups over a single-core *IJK*-algorithm implementation written in C++. Such a fast GPU implementation of Stewart [27]’s algorithm requires significant engineering that is a contribution in its own right, so we focus on conventional implementations in the following.

Finally, even for serial implementations, the language of implementation can eclipse theoretical computational complexity. We find that an R based implementation of Stewart [27]’s algorithm is much slower than the QR factorization based method implemented in the PRACMA [5] R package. The QR decomposition in R calls the FORTRAN library LAPACK to perform the numerical linear algebra for the QR decomposition, and its precompiled and memory efficient implementation outperform interpreted R.

4.1.2. Other multiproposals

If one precomputes $\Sigma^{1/2}$, the multiproposals (3) and (4) are of complexity $O(D^2P)$, with the majority of time spent multiplying the $D \times D$ square-root covariance $\Sigma^{1/2}$ by a $D \times P$ matrix of standard normals. If one adopts the preconditioning strategy of Haario et al. [12], the complexity changes to $O(D^3 + D^2P)$ accounting for a Cholesky decomposition. If one specifies $\Sigma = \sigma^2 \mathbf{I}_D$, then the complexity drops to $O(DP)$, and the new bottleneck becomes generation of i.i.d. standard normals.

Just as for the simplicial sampler’s random rotations, one may parallelize the matrix multiplications for these multiproposals. One may also parallelize the standard normal generations using the parallel pseudo random number generators of Salmon et al. [25] that TENSORFLOW uses Abadi et al. [1] to great success.

4.2. Acceptance mechanisms

The simplified acceptance mechanism (6) requires D target density evaluations. These density evaluations are embarrassingly parallel, but there are some target distributions for which this kind of parallelization is not necessary. When one jointly infers a multivariate Gaussian random variable and its covariance matrix using Metropolis-within-Gibbs, one must invert the updated covariance every time one updates the Gaussian variable. Once one has inverted the covariance matrix, evaluating the multivariate Gaussian density over D proposals is almost as fast as evaluating the density over a single proposal. That said, this intuition also applies to Metropolis-within-Gibbs with a simple random walk Metropolis update: having updated and inverted the covariance matrix, one may perform many Metropolis updates of the Gaussian random variable at relatively little additional cost. In general, however, careful parallelization should help accelerate the acceptance step for multiproposal methods in a way unavailable to single-proposal methods. For example, Holbrook [15] uses a quantum computer to achieve quadratic speedups that reduce the number of target evaluations from $\mathcal{O}(P)$ to $\mathcal{O}(\sqrt{P})$.

The simplified acceptance mechanism also avoids potentially costly computations of the $P + 1$ proposal densities $q(\theta_p^*, \Theta^*)$ required for the general acceptance mechanism (2). For example, the i.i.d. multiproposal (3) requires an additional $O(P^2)$ computational cost to obtain the terms

$$q(\theta_p^*, \Theta^*) = \prod_{p' \neq p} \exp \left(-\frac{1}{2} (\theta_p^* - \theta_{p'}^*)^T \Sigma^{-1} (\theta_p^* - \theta_{p'}^*) \right), \quad p \in \{1, \dots, P + 1\}.$$

On the other hand, Lemma 1 shows that the centered Gaussian multiproposal (4) leads to a simplified acceptance mechanism that results in prodigious theoretical speedups thoroughly discussed in Section 3.

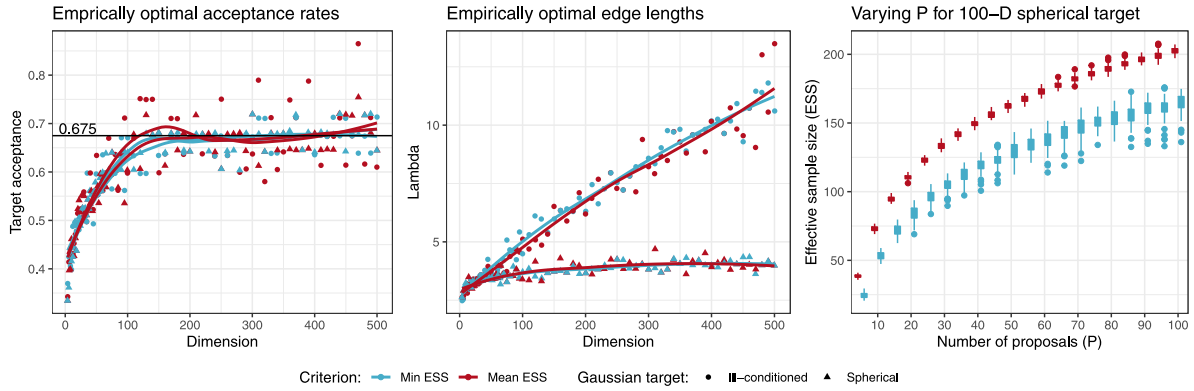


Fig. 6. The left two figures present acceptance rates and simplex edge lengths that empirically maximize mean and minimum effective sample sizes (ESS) across target dimensions. For spherical and ill-conditioned targets, we use vanilla and preconditioned simplicial sampler, respectively. The right figure presents ESS results for a sampler modified to use a lower dimensional simplex.

5. Results

In the following, we use (PC-)Simpl, (PC-)RWM and (PC-)MTM to indicate the (preconditioned) simplicial sampler, random walk and multiple-try Metropolis algorithms, respectively. Although PC-RWM is better known as the adaptive Metropolis algorithm of Haario et al. [12], we use this notation to clarify comparisons. ESS and ESSs denote effective sample size and effective sample size per second, respectively. All experiments use the R language [16] and the R package PRACMA [5] to produce random rotations, which itself uses the QR decomposition of a matrix with Gaussian entries (Section 4.1.1). All MTM simulations use D proposals for comparison. Finally, all results herein rely on a conventional single-core implementation.

5.1. Empirically optimal scaling

5.1.1. Acceptance rates and edge lengths

We first investigate Simpl and PC-Simpl performance for Gaussian targets of varying dimensions (4 through 500) as a function of acceptance rate and simplex edge length. For each dimension, we employ 20 independent MCMC runs of 10,000 iterations targeting one of 20 acceptance rates uniformly spaced between 0.2 and 0.95 (Section 3.1.3). The ‘ill-conditioned’ Gaussian target has diagonal covariance with elements 1 through D . The left two panels of Fig. 6 plot the acceptance rates or edge lengths that maximize minimum and mean ESS out of the 20 targeted. For both Simpl and PC-Simpl, optimal acceptance rates increase with dimensionality until flattening out at around 0.675. For Simpl, optimal edge lengths quickly flatten out at around 3. For PC-Simpl, optimal edge lengths grow with target dimension.

5.1.2. Number of proposals

In general, Simpl uses a simplex with $D + 1$ nodes, the maximal number available in D dimensions. The right panel of Fig. 6 shows ESS results when lower dimensional simplices (with fewer nodes) are used to generate fewer proposals ($P < D$). In practice, one may implement such a variation by replacing the initial set of pre-rotated vectors $(\mathbf{v}_1, \dots, \mathbf{v}_D, \mathbf{0})$ with $(\mathbf{v}_1, \dots, \mathbf{v}_P, \mathbf{0})$, again letting $P < D$. Algorithmic details otherwise remain the same. We use 100 independent MCMC runs of 10,000 iterations each to generate minimum and mean ESS across target dimensions. Both measures improve with greater numbers of proposals. We do not compare ESSs because the additional cost of a simplex node arises from a mere $\mathcal{O}(D^2)$ matrix–vector multiplication, compared to the $\mathcal{O}(D^3)$ generation of a random rotation.

5.2. Performance comparisons

5.2.1. Accelerated centered Gaussian multiproposal

Fig. 7 compares the simplicial sampler performance to that of the centered Gaussian sampler (4) using $P = D$ and $P = 2D$ proposals. The ‘slow’ Gaussian centered multiproposal calculates all $P + 1$ proposal probabilities $q(\theta_p^*, \Theta^*)$, whereas the ‘fast’ sampler uses the simplified acceptance probabilities (6) following Lemma 1. Samplers target a standard spherical normal distribution and a 50% acceptance rate for 10,000 iterations. In terms of ESS, doubling the number of proposals appears to benefit the centered Gaussian sampler, especially in lower dimensions. The fast implementation of the centered Gaussian sampler already starts to outperform the slow implementations in lower dimensions, and this relative performance gain only increase with dimensions. Across all dimensions, the ESSs trajectory of the simplicial sampler remains firmly in between those of the fast and slow centered Gaussian samplers.

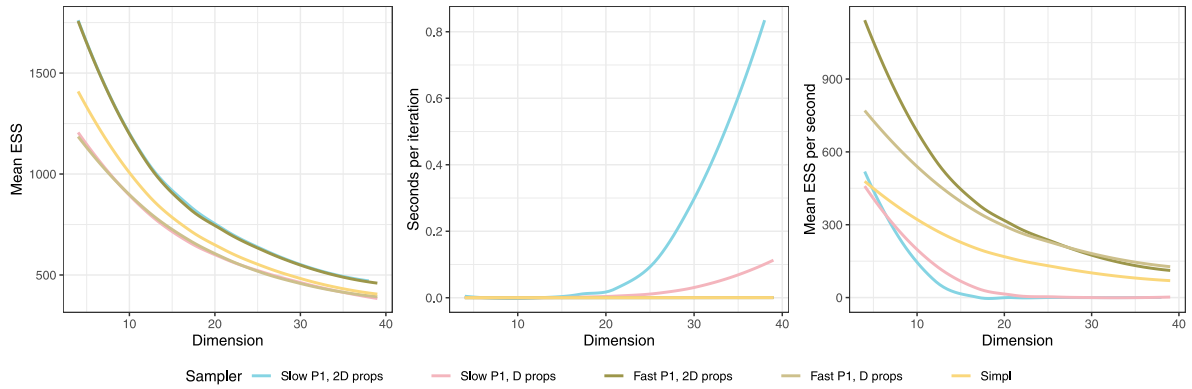


Fig. 7. Performance of the simplicial sampler (Simpl) and the 'P1' proposal mechanism (4), with both D and $2D$ proposals and both fast and slow implementations (Section 3). All samplers target a standard multivariate Gaussian distribution, run for 10,000 iterations and adapt to a 50% target acceptance rate.

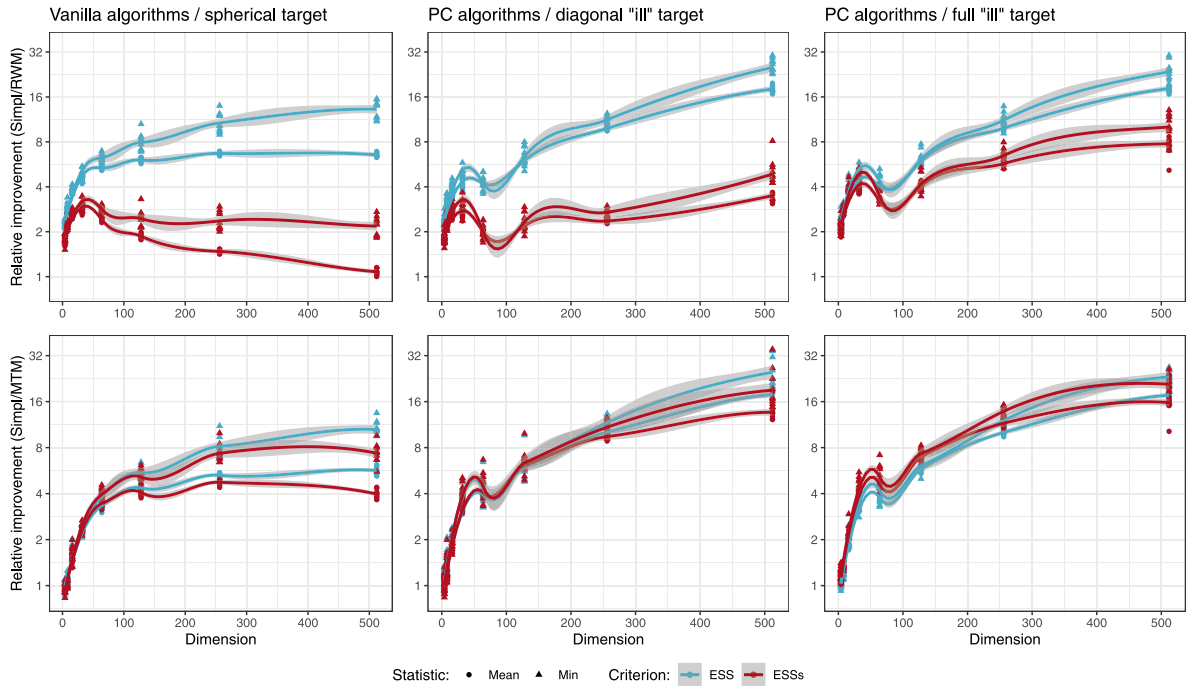


Fig. 8. Relative improvement of simplicial sampler (Simpl) against random walk Metropolis (RWM) and multiple-try Metropolis (MTM) as measured by effective sample size (ESS) and ESS per second (ESSs). We apply preconditioning (PC) in the exact same manner to all algorithms.

5.2.2. Gaussian targets

We compare vanilla algorithms on spherical targets of dimensions $D = 4, 8, 16, 32, 64, 128, 256$ and 512 , all with identity for covariance matrix, and preconditioned algorithms on targets with ill-conditioned diagonal and full covariance matrices. Ill-conditioned matrices all have spectra ranging uniformly from 1 to D . RWM and MTM both use optimal scaling of [10]. For the latter, this results in an asymptotic 30% acceptance rate as D gets large [4]. Simpl adapts to empirically optimal acceptance rates of Fig. 6. Fig. 8 shows relative ESS and ESSs results from 10 independent MCMC runs of 10,000 iterations each per algorithm. Since ESS is a univariate measure, we report means and minima with respect to target dimensions.

5.2.3. Bimodal target

In general, Gaussian simplicial sampling (G-Simpl) (Section 3.1.1) and its preconditioned form (PCG-Simpl) perform slightly worse (10% less ESS) than Simpl and PC-Simpl on unimodal targets, but G-Simpl and PCG-Simpl perform better on multimodal targets. We compare G-Simpl and PCG-Simpl to RWM and PC-RWM on mixture of Gaussian targets of

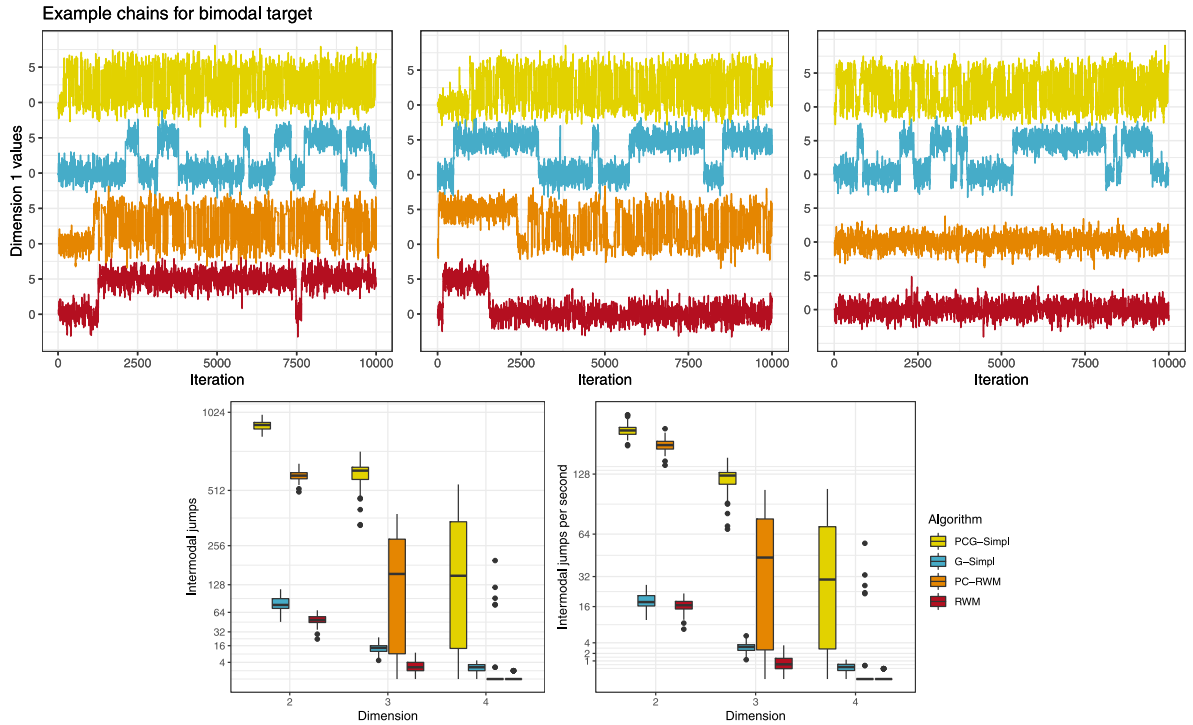


Fig. 9. The top figures show stacked traceplots for the first dimension of a 3D Gaussian mixture target. The bottom figures show the number of intermodal jumps for 100,000 iterations presented over 100 independent MCMC runs. PCG- and G-Simpl are preconditioned and non-preconditioned Gaussian simplicial samplers compared to the same for random walk Metropolis (RWM).

dimensions 2, 3 and 4. Mixture probabilities are 0.5 – 0.5 and covariances are identity, but modal means are $\mathbf{0}$ and $\mathbf{5} = (5, \dots, 5)$. Thus, modes become farther apart with increasing dimension. Fig. 9 shows stacked traceplots and boxplots for the number of intermodal jumps recorded from 100 independent MCMC runs of 100,000 iterations each.

5.2.4. Gaussian process classification

48 states use a winner-take-all rule when assigning electoral college votes during the U.S. presidential election. We use a Gaussian process (GP) with logit link [30] to regress binary outcomes (Trump or Clinton) over state center (latitude and longitude) and population data for the 2016 presidential election. For latent variables $\theta = (\theta_1, \dots, \theta_{48})$, the prior covariance of θ_i and θ_j is

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \xi^2 + \eta^2 \exp(-\rho^2 \|\mathbf{x}_i - \mathbf{x}_j\|_2^2) + \sigma^2 \delta_{ij},$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are their respective predictor vectors. We infer latent variables θ using (PC-)Simple, (PC-)RWM and (PC-)MTM. We let Simpl algorithms target an acceptance rate of 0.5 following the results of Fig. 6, while we let RWM, MTM and PC-MTM target 0.234, 0.3 and 0.4. The former value is optimal for RWM [24]. The last two values appear advantageous in preliminary tests. We infer other parameters using univariate slice samplers [22]. Table 1 shows results from 100 independent runs of 100,000 iterations each. We initialize all chains to misclassify all states and use a total misclassification of 10 as rough convergence threshold. While Simple and MTM require multiple target evaluations at each step, they only must compute the GP inverse covariance once for all such evaluations. These results only establish effectiveness of the simplicial sampler relative to a few basic samplers and do not establish superiority relative to all samplers. Gradient-based MCMC methods could likely outperform the simplicial sampler, but we abstain from such comparisons in order to focus on ‘apples-to-apples’ comparisons among non-gradient methods.

5.3. Visualizing the extra-dimensional simplicial sampler

Because of the diminishing returns (Fig. 6, right) and $O(D^3)$ computational cost associated with increasing the number of simplex vertices, the extra-dimensional simplicial sampler is not practical in the absence of advanced parallelization techniques. Moreover, the rotated vertices no longer maintain equal distances from each other in the low-dimensional space over which the target is defined. Fig. 10 shows three successive iterations of the extra-dimensional simplicial sampler with 1000 proposals targeting a 2D Gaussian distribution. The sampler successfully moves to the high-density region of the space in these three steps. We observe that rotating and projection the simplex vertices results in proposal point clouds that look plausibly Gaussian distributed, apparently lacking any geometric structure.

Table 1

Effective sample sizes (ESS) and ESS per second (ESSs) from 100 independent MCMC runs of 100,000 iterations each for Gaussian process classification of 48 state election results from 2016 U.S. presidential election. Standard errors of empirical mean estimates accompany in parentheses. Simpl, RWM and MTM are vanilla simplicial sampler, random walk and multiple-try Metropolis algorithms. PC implies preconditioning. Its and time to $err = 10$ represent iterations and seconds to achieve a misclassification total of 10, a rough measure of convergence speed. Winners in blue.

Algorithm	Mean ESS θ	Min ESS θ	Its to $err = 10$	ESS η^2	ESS ξ^2	ESS ρ^2	ESS σ^2
Simpl	1571.50 (15.6)	256.27 (6.4)	73.05 (1.8)	671.39 (10.2)	2463.76 (76.6)	3703.05 (50.0)	452.30 (6.7)
RWM	436.34 (2.3)	68.36 (2.3)	276.60 (8.3)	258.63 (5.7)	1407.88 (61.8)	1837.47 (31.3)	134.91 (2.3)
MTM	1195.22 (12.8)	252.13 (7.3)	1928.75 (20.5)	676.75 (12.8)	2332.81 (84.9)	3490.12 (58.41)	366.86 (6.3)
PC-Simpl	1865.18 (14.4)	1200.50 (20.4)	76.39 (2.2)	1312.98 (18.0)	4583.21 (95.5)	4222.10 (50.7)	532.49 (6.5)
PC-RWM	362.99 (4.5)	205.16 (5.3)	284.27 (10.5)	418.99 (10.9)	1381.00 (37.6)	1654.35 (29.8)	125.12 (3.0)
PC-MTM	1256.07 (15.9)	755.42 (16.4)	1873.89 (21.1)	951.97 (15.3)	3840.86 (88.3)	3736.27 (48.0)	272.08 (6.3)
	Mean ESSs θ	Min ESSs θ	Time to $err = 10$	ESSs η^2	ESSs ξ^2	ESSs ρ^2	ESSs σ^2
Simpl	5.62 (0.06)	0.92 (0.02)	15.87 (0.98)	2.40 (0.04)	8.81 (0.28)	13.24 (0.18)	1.62 (0.02)
RWM	2.50 (0.02)	0.39 (0.01)	145.41 (9.69)	1.48 (0.03)	8.09 (0.36)	10.55 (0.19)	0.77 (0.01)
MTM	3.85 (0.04)	0.81 (0.02)	11 674.88 (252.45)	2.18 (0.04)	7.52 (0.27)	11.25 (0.19)	1.18 (0.02)
PC-Simpl	5.09 (0.13)	3.30 (0.10)	23.99 (1.46)	3.58 (0.10)	12.59 (0.43)	11.53 (0.32)	1.45 (0.04)
PC-RWM	1.44 (0.04)	0.81 (0.03)	268.77 (39.96)	1.65 (0.06)	5.48 (0.22)	6.54 (0.21)	0.49 (0.02)
PC-MTM	3.87 (0.05)	2.33 (0.05)	11 540.43 (269.22)	2.93 (0.05)	11.85 (0.28)	11.53 (0.15)	0.84 (0.02)

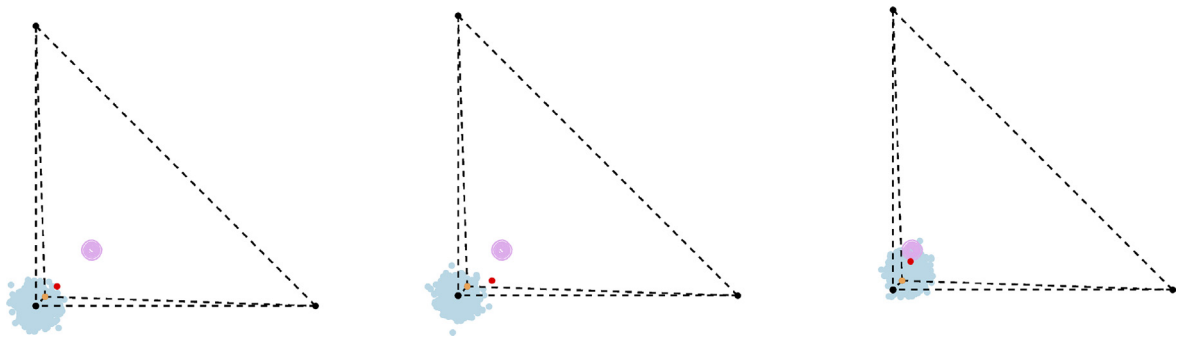


Fig. 10. Successive iterations of an extra-dimensional simplicial sampler using 1000 proposals to sample a bivariate Gaussian target. For each step, the yellow dot is the initial position, the black dots are the projected un-rotated simplex vertices, the blue dots are the proposals (or projected rotated vertices), and the red dots are the accepted proposals. The dashed lines represent simplex edges, and the purple contours represent the target distribution. Finally, 998 unrotated vertices project down to the same bottom-left black point on the plane.

6. Discussion

We have developed the simplicial sampler, a multivariate extension of the multiproposal mechanism (5) from Tjelmeland [29], and proven that this extension maintains detailed balance using simplified acceptance probabilities. We have shown that a variation on this algorithm, the Gaussian simplicial sampler, maintains a unique stationary distribution and an appropriate law of large numbers. Furthermore, we have proven that the centered Gaussian proposal mechanism (4) from Tjelmeland [29] enables the same simplified acceptance probability that leads to significant speedups. We have thoroughly discussed the computational issues that one must consider when implementing these algorithms. Among others, these considerations include: theoretical computational complexity, hardware, programming language, efficient use of memory and parallelization. Finally, we have thoroughly compared performance of sequential implementations on a collection of target distributions. That said, we note that these algorithms are well suited for parallelization enabled by, e.g., a GPU or a quantum computer [15].

We are interested in two lines of further investigation. First, we have shown that two structurally distinct multiproposal mechanisms have symmetries that enable simplified acceptance probabilities. Are these the only such mechanisms or are there others? Can one combine these in ways that preserve the simplified acceptance probabilities? Finally, do analogues for these methods exist in discrete settings, e.g., for lattice models? Second, we are interested in the quality of convergence for the broader class of multiproposal MCMC algorithms. Based on simulations, it would not be surprising for geometric ergodicity to hold for these algorithms. Moreover, one might expect for mixing speeds to improve as the number of proposals increases.

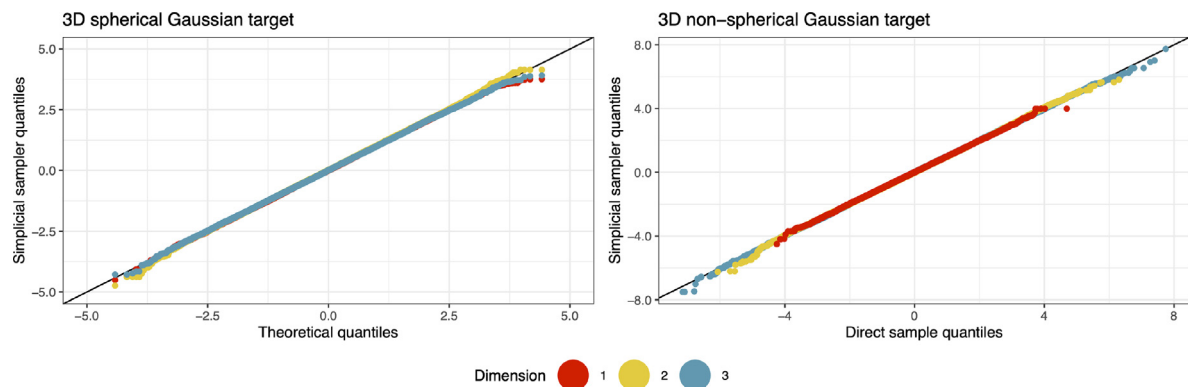


Fig. A.11. Quantile–quantile plots from 100,000 vanilla simplicial sampler iterations after removing 10,000 iterations as burn-in.

Acknowledgments

This work was supported by NIH, USA grant K25 AI153816 and NSF, USA grant DMS 2152774. We thank Nathan Glatt-Holtz, Cecilia Mondaini and Justin Krometis for helpful discussions and advice.

Appendix. Empirical validation

Fig. A.11 shows quantile–quantile (QQ) plots for the vanilla simplicial sampler for two different 3D Gaussian target distributions, one of which is spherical and the other of which has diagonal covariance with standard deviations 1, 2 and 3. The QQ plots demonstrate that the empirical quantiles of the simplicial sampler adhere closely to the quantiles of respective target distributions.

References

- [1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems, 2016, arXiv preprint [arXiv:1603.04467](https://arxiv.org/abs/1603.04467).
- [2] T.W. Anderson, I. Olkin, L.G. Underhill, Generation of random orthogonal matrices, *SIAM J. Sci. Statist. Comput.* 8 (4) (1987) 625–629.
- [3] A.A. Barker, Monte carlo calculations of the radial distribution functions for a proton? electron plasma, *Aust. J. Phys.* 18 (2) (1965) 119–134.
- [4] M. Bédard, R. Douc, E. Moulines, Scaling analysis of multiple-try MCMC methods, *Stochastic Process. Appl.* 122 (3) (2012) 758–786.
- [5] H.W. Borchers, *Pracma: Practical numerical math functions*, 2021, R package version 2.3.3.
- [6] B. Calderhead, A general construction for parallelizing Metropolis–Hastings algorithms, *Proc. Natl. Acad. Sci.* 111 (49) (2014) 17408–17413.
- [7] H.S.M. Coxeter, *Regular Polytopes*, Courier Corporation, New York, 1973.
- [8] G.B. Folland, *A Course in Abstract Harmonic Analysis*, Vol. 29, CRC Press, Boca Raton, Florida, 2016.
- [9] D. Frenkel, Speed-up of Monte Carlo simulations by sampling of rejected states, *Proc. Natl. Acad. Sci.* 101 (51) (2004) 17571–17575.
- [10] A. Gelman, W.R. Gilks, G.O. Roberts, Weak convergence and optimal scaling of random walk Metropolis algorithms, *Ann. Appl. Probab.* 7 (1) (1997) 110–120.
- [11] N.E. Glatt-Holtz, A.J. Holbrook, J.A. Krometis, C.F. Mondaini, *Parallel MCMC algorithms: Theoretical foundations, algorithm design, case studies*, 2022.
- [12] H. Haario, E. Saksman, J. Tamminen, An adaptive Metropolis algorithm, *Bernoulli* 7 (2) (2001) 223–242.
- [13] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* 57 (1) (1970) 97–109.
- [14] A. Holbrook, Differentiating the pseudo determinant, *Linear Algebra Appl.* 548 (2018) 293–304.
- [15] A.J. Holbrook, A quantum parallel Markov chain Monte Carlo, 2021, arXiv preprint [arXiv:2112.00212](https://arxiv.org/abs/2112.00212).
- [16] R. Ihaka, R. Gentleman, R: a language for data analysis and graphics, *J. Comput. Graph. Statist.* 5 (3) (1996) 299–314.
- [17] J.M. Lee, Quotient manifolds, in: *Introduction to Smooth Manifolds*, Springer New York, New York, NY, 2012, pp. 540–563.
- [18] J. Li, S. Ranka, S. Sahni, GPU matrix multiplication, *Multicore Comput.: Algorithms Archit. Appl.* 345 (2013).
- [19] J.S. Liu, F. Liang, W.H. Wong, The multiple-try method and local optimization in Metropolis sampling, *J. Amer. Statist. Assoc.* 95 (449) (2000) 121–134.
- [20] X. Luo, H. Tjelmeland, A multiple-try Metropolis–Hastings algorithm with tailored proposals, *Comput. Statist.* 34 (3) (2019) 1109–1133.
- [21] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* 21 (6) (1953) 1087–1092.
- [22] R.M. Neal, Slice sampling, *Ann. Statist.* 31 (3) (2003) 705–767.
- [23] P.H. Peskun, Optimum monte-carlo sampling using markov chains, *Biometrika* 60 (3) (1973) 607–612.
- [24] J.S. Rosenthal, Optimal proposal distributions and adaptive MCMC, *Handb. Markov Chain Monte Carlo* (2011) 119–138.
- [25] J.K. Salmon, M.A. Moraes, R.O. Dror, D.E. Shaw, Parallel random numbers: As easy as 1, 2, 3, in: *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '11, Association for Computing Machinery, New York, NY, USA, 2011.
- [26] T. Schwedes, B. Calderhead, Rao-blackwellised parallel MCMC, in: A. Banerjee, K. Fukumizu (Eds.), *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, in: *Proceedings of Machine Learning Research*, Vol. 130, PMLR, 2021, pp. 3448–3456.
- [27] G.W. Stewart, The efficient generation of random orthogonal matrices with an application to condition estimators, *SIAM J. Numer. Anal.* 17 (3) (1980) 403–409.
- [28] L. Tierney, Markov chains for exploring posterior distributions, *Ann. Statist.* 22 (4) (1994) 1701–1728.

- [29] H. Tjelmeland, Using All Metropolis-Hastings Proposals to Estimate Mean Values, Technical Report Tech Rep 4, Norwegian University of Science and Technology, Trondheim, Norway, 2004.
- [30] C.K. Williams, D. Barber, Bayesian classification with Gaussian processes, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (12) (1998) 1342–1351.
- [31] S. Yang, Y. Chen, E. Bernton, J.S. Liu, On parallelizable Markov chain Monte Carlo algorithms with waste-recycling, *Stat. Comput.* 28 (5) (2018) 1073–1081.
- [32] X. Yang, J.S. Liu, Convergence rate of multiple-try Metropolis independent sampler, 2021, arXiv preprint [arXiv:2111.15084](https://arxiv.org/abs/2111.15084).