SAMPLING FROM A MULTIVARIATE GAUSSIAN DISTRIBUTION TRUNCATED ON A SIMPLEX: A REVIEW

Yoann Altmann[*]*, Steve McLaughlin*

Heriot-Watt University
School of Engineering and Physical Sciences
Edinburgh, United Kingdom

Nicolas Dobigeon

University of Toulouse
IRIT/INP-ENSEEIHT
Toulouse, France

ABSTRACT

In many Bayesian models, the posterior distribution of interest is a multivariate Gaussian distribution restricted to a specific domain. In particular, when the unknown parameters to be estimated can be considered as proportions or probabilities, they must satisfy positivity and sum-to-one constraints. This paper reviews recent Monte Carlo methods for sampling from multivariate Gaussian distributions restricted to the standard simplex. First, a classical Gibbs sampler is presented. Then, two Hamiltonian Monte Carlo methods are described and analyzed. In a similar fashion to the Gibbs sampler, the first method has a acceptance rate equal to one whereas the second requires an accept/reject procedure. The performance of the three methods are compared through the use of a few examples.

Index Terms— Markov Chain Monte Carlo methods, truncated multivariate Gaussian distributions, Constrained Hamiltonian Monte Carlo.

1. INTRODUCTION

Markov chain Monte Carlo (MCMC) methods are powerful tools for sampling complex multivariate distributions. These methods have been intensively studied for the last twenty years and have proven their efficiency in solving numerous inference problems in the signal processing and machine learning communities. In many applications (e.g., linear mixtures corrupted by additive Gaussian noise), the full posterior distributions or some conditional distributions of interest are multivariate Gaussian distributions (MGDs). However, when the unknown parameters (e.g., the mixing parameters) satisfy inequality and/or equality constraints introduced through prior distributions, the posterior distribution is a MGD restricted to a constrained parameter space, or truncated MGD (TMGD). Efficient sampling from TMGDs, is a challenging problem and classical simulation methods often reduce to a Gibbs sampler which samples sequentially from univariate truncated Gaussian distributions. Each step of the Gibbs sampler can be achieved using simple and efficient methods [1–3]. However, the Gibbs sampler can be inaccurate for large vectors (large number of Gibbs steps) and for correlated variables (slow exploration of the target density).

In this paper, we focus on the special case where the parameters to be sampled belong to a standard simplex, i.e., are positive and sum to one. This situation typically occurs whenever the unknown parameters can be interpreted as fractions or probabilities, for instance when analyzing chemical mixtures [4], hyperspectral data [5] or gene expression microarrays [6]. Efficient sampling from a simplex can be challenging since the variables can be *a posteriori* highly correlated, mainly because of the sum-to-one constraint. In this paper, we review Hamiltonian Monte Carlo (HMC) [7] methods which sample according to TMGDs restricted to the standard simplex. HMCs are powerful simulation strategies based on Hamiltonian dynamics which can improve the convergence and mixing properties of classical MCMC methods (such as the Gibbs sampler and the Metropolis-Hastings algorithm) [8, 9]. These methods have received growing interest in many applications, especially when the number of parameters to be estimated is large [10, 11]. Classical HMC methods can only be used for unconstrained variables. However, new HMC methods have been recently proposed to handle constrained variables [8, Chap. 5] [12–14] which allow HMCs to sample according to TMGDs restricted to the standard simplex.

The paper is organized as follows. Section 2 introduces the TMGD to be sampled from. The classical Gibbs sampler is presented in Section 3. Section 4 recalls the principle of HMC methods and studies two methods for sampling according to the distribution of interest. The performance of the HMC methods is evaluated using a few examples in Section 5 and conclusions are finally reported in Section 6.

2. PROBLEM

Consider the vector $\boldsymbol{a} = [a_1, \ldots, a_R]^T \in \mathbb{R}^R$, $R > 1$ belonging to the standard $(R-1)$-simplex defined by

$$\mathcal{S}_{R-1} = \left\{ \boldsymbol{a} \middle| \sum_{r=1}^{R} a_r = 1, \quad a_r > 0, \forall r \right\}. \tag{1}$$

The problem addressed in this paper is the simulation of samples distributed according to a multivariate Gaussian distribution restricted to the standard simplex (1) defined by

$$f(\boldsymbol{a}|\mathbf{m}, \boldsymbol{\Sigma}) = \begin{cases} \beta_{\mathbf{m},\boldsymbol{\Sigma}}^{-1} f_0(\boldsymbol{a}|\mathbf{m}, \boldsymbol{\Sigma}) & \text{if } \boldsymbol{a} \in \mathcal{S}_{R-1} \\ 0 & \text{else.} \end{cases} \tag{2}$$

where $\beta_{\mathbf{m},\boldsymbol{\Sigma}}$ is a normalisation constant and $f_0(\boldsymbol{a}|\mathbf{m}, \boldsymbol{\Sigma})$ is the probability density function of the multivariate Gaussian distribution

$$f_0(\boldsymbol{a}|\mathbf{m}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (\boldsymbol{a} - \mathbf{m})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{a} - \mathbf{m}) \right]$$

with mean $\mathbf{m} \in \mathbb{R}^R$ and positive definite covariance matrix $\boldsymbol{\Sigma}$.

[*]Part of this work was supported by the Direction Générale de l'Armement, French Ministry of Defence, by ANR-11-LABX-0040-CIMI within the program ANR-11-IDEX-0002-02 and by the Hypanema ANR Project, ANR Project n°$12 - BS03 - 003$ in addition the support of the EPSRC via grant EP/J015180/1.

In this paper, we will focus on sampling vectors on the standard simplex defined in (1). For more general simplexes, an affine transformation can be applied to scale the parameters to the standard simplex. The density in (2) is known up to the scaling factor

$$\beta_{\mathbf{m},\mathbf{\Sigma}} = \int_{\mathcal{S}_{R-1}} f_0(\boldsymbol{a}|\mathbf{m},\mathbf{\Sigma})\mathrm{d}\boldsymbol{a} \tag{3}$$

which is generally not easy to compute [15]. Since $f(\boldsymbol{a}|\mathbf{m},\mathbf{\Sigma})$ is a non-standard distribution, MCMC methods are used to generate samples asymptotically distributed according to (2). The next section recalls the main steps of the Gibbs sampler usually considered to sample from the multivariate Gaussian distribution restricted to the standard simplex in (2).

3. GIBBS SAMPLER

The first sampling strategy investigated in this paper is a classical Gibbs sampler which consists of a sequential sampling of the elements of $\boldsymbol{a}$. However, because of the sum-to-one constraint for the elements of $\boldsymbol{a}$, at least two components must be updated simultaneously. The vector $\boldsymbol{a}$ can be rewritten

$$\boldsymbol{a} \quad = \quad \begin{bmatrix} \mathbf{c} \\ 1-\mathbf{c}^T\mathbf{1}_{R-1} \end{bmatrix} = \mathbf{J}\mathbf{c}+\mathbf{b}$$

where $\mathbf{c} = [c_1,\ldots,c_{R-1}]^T = [a_1,\ldots,a_{R-1}]^T$ is an $(R-1)\times 1$ vector, $\mathbf{J} = [\mathbf{I}_{R-1}, -\mathbf{1}_{R-1}]^T$ is an $(R-1)\times R$ matrix and $\mathbf{b} = [\mathbf{0}_{R-1}^T, 1]^T$ is an $R\times 1$ vector. Moreover, $\mathbf{1}_{R-1}$ (resp. $\mathbf{0}_{R-1}$) stands for a vector of $(R-1)$ ones (resp. zeros). Note that this reparametrization is chosen for notation simplicity. However, the component of $\boldsymbol{a}$ to be discarded can be randomly chosen. Sampling from (2) can be achieved by sampling from

$$f(\mathbf{c}|\boldsymbol{\mu},\mathbf{S}) = \begin{cases} \beta_{\boldsymbol{\mu},\mathbf{S}}^{-1} f_0(\mathbf{c}|\boldsymbol{\mu},\mathbf{S}) & \text{if } \mathbf{J}\mathbf{c}+\mathbf{b}\in\mathcal{S}_{R-1} \\ 0 & \text{else.} \end{cases} \tag{4}$$

where

$$\begin{cases} \mathbf{S}^{-1} &= \mathbf{J}^T\mathbf{\Sigma}^{-1}\mathbf{J} \\ \boldsymbol{\mu} &= \mathbf{S}\mathbf{J}^T\mathbf{\Sigma}^{-1}(\mathbf{m}+\mathbf{b}) \end{cases}. \tag{5}$$

The easiest way to sample from (4) is probably to use a standard rejection sampling method [9, p. 49] that consists of sampling from $f_0(\mathbf{c}|\boldsymbol{\mu},\mathbf{S})$ until the generated sample satisfies $\mathbf{J}\mathbf{c}+\mathbf{b}\in\mathcal{S}_{R-1}$. However, this can lead to very low acceptance rate if the probability $\mathrm{P}\left[\mathbf{J}\mathbf{c}+\mathbf{b}\in\mathcal{S}_{R-1}|\mathbf{c}\sim\mathcal{N}(\boldsymbol{\mu},\mathbf{S})\right]$ is low.

The classical Gibbs sampler consists of drawing sequentially the elements of $\mathbf{c}$ from their full conditional distributions. As it [16], it can be shown that the distributions $f(c_r|\mathbf{c}_{\setminus r},\boldsymbol{\mu},\mathbf{\Sigma})$ are doubly truncated Gaussian distributions from which it can be easily sampled using the algorithm described in [3]. Though the convergence of this sampler to the target distribution (2) is ensured, the intrinsic limitations of the Gibbs sampler can lead to poor mixing properties, especially for high dimensional simplexes and/or correlated variables. Note that Block-Gibbs samplers cannot be used here since they also rely on sampling from multivariate Gaussian distributions restricted to simplexes. The next section presents new sampling strategies to sample from (2) using simultaneous updates of all the elements of $\boldsymbol{a}$.

4. HAMILTONIAN MONTE CARLO METHODS

Hamiltonian Monte Carlo (HMC) methods are powerful tools for sampling from many continuous distributions by introducing fictitious momentum variables. Let $\mathbf{q}\in\mathbb{R}^D$ be the parameter of interest and $\pi(\mathbf{q})$ its corresponding distribution to be sampled from. From statistical mechanics, the distribution $\pi(\mathbf{q})$ can be related to a potential energy function $U(\mathbf{q}) = -\log[\pi(\mathbf{q})] + c$ where c is a positive constant such that $\int \exp(-U(\mathbf{q})+c)\,\mathrm{d}\mathbf{q} = 1$. The Hamiltonian of $\pi(\mathbf{q})$ is a function of the energy $U(\mathbf{q})$ and of an additional momentum vector $\mathbf{p}\in\mathbb{R}^D$ defined as

$$H(\mathbf{q},\mathbf{p}) = U(\mathbf{q}) + K(\mathbf{p}) \tag{6}$$

where $K(\mathbf{p})$ is an arbitrary kinetic energy function. Usually, a quadratic kinetic energy is chosen, i.e. , $K(\mathbf{p}) = \mathbf{p}^T\mathbf{M}\mathbf{p}/2$ where $\mathbf{M}$ is a positive definite symmetric matrix. The Hamiltonian (6) defines the distribution

$$\begin{aligned} f(\mathbf{q},\mathbf{p}) &\propto \exp[-H(\mathbf{q},\mathbf{p})] \\ &\propto \pi(\mathbf{q})\exp\left(-\frac{1}{2}\mathbf{p}^T\mathbf{M}\mathbf{p}\right) \end{aligned} \tag{7}$$

for $(\mathbf{q},\mathbf{p})$ which shows that $\mathbf{q}$ and $\mathbf{p}$ are independent and that the marginal distribution of $\mathbf{p}$ is a $\mathcal{N}(\mathbf{0}_D,\mathbf{M}^{-1})$ distribution. HMC algorithms generally allow samples to be asymptotically generated according to $f(\mathbf{q},\mathbf{p})$. The ith HMC iteration starts with an initial pair of vectors $(\mathbf{q}^{(i)},\mathbf{p}^{(i)})$ and consists of two steps. The first step resamples the initial momentum $\tilde{\mathbf{p}}^{(i)}$ according to $\mathcal{N}(\mathbf{0}_D,\mathbf{M})$. The second step uses Hamiltonian dynamics to propose a candidate $(\mathbf{q}^*,\mathbf{p}^*)$ which is accepted with the following probability

$$\rho = \min\left\{\exp\left[-H(\mathbf{q}^*,\mathbf{p}^*)+H(\mathbf{q}^{(i)},\tilde{\mathbf{p}}^{(i)})\right],1\right\}. \tag{8}$$

If the vector $\mathbf{q}$ satisfies inequality constraints, the dynamics can be modified to ensure $\mathbf{q}^*$ belongs to its admissible set. Hamiltonian dynamics are usually simulated by discretization methods such as Euler or leapfrog methods (as will be shown in Section 4.2). However, when $U(\mathbf{q})$ is quadratic, it has been shown [14] that it is possible to simulate exactly from $f(\mathbf{q},\mathbf{p})$. This method is summarized below.

4.1. Exact Hamiltonian Monte Carlo

In this section, we summarize the Exact Hamiltonian Monte Carlo (EHMC) method investigated in [14] to sample from (4). Let $\mathbf{q} = \mathbf{c}\in\mathbb{R}^{R-1}$, $U(\mathbf{q}) = (\mathbf{q}-\boldsymbol{\mu})^T\mathbf{S}^{-1}(\mathbf{q}-\boldsymbol{\mu})/2$ and $\mathbf{M} = \mathbf{S}^{-1}$. Hamilton's equations,

$$\frac{\partial\mathbf{q}}{\partial t} = \left(\frac{\partial H}{\partial\mathbf{p}}\right)^T, \quad \frac{\partial\mathbf{p}}{\partial t} = -\left(\frac{\partial H}{\partial\mathbf{q}}\right)^T \tag{9}$$

can be combined, leading to the solution

$$\begin{cases} \mathbf{q}(t) &= \mathbf{a}\cos(t)+\mathbf{b}\sin(t) \\ \mathbf{p}(t) &= \mathbf{S}^{-1}(\mathbf{a}\sin(t)-\mathbf{b}\cos(t)) \end{cases} \tag{10}$$

where $\mathbf{a}$ and $\mathbf{b}$ are $(R-1)\times 1$ vectors related to the initial couple $(\mathbf{q}^{(i)},\mathbf{p}^{(i)})$ (see [14] for details). Moreover, it can be shown that $H(\mathbf{q}(t),\mathbf{p}(t)) = H(\mathbf{q}^{(i)},\mathbf{p}^{(i)}),\forall t$ and that (10) can be used to sample from $f_0(\mathbf{c}|\boldsymbol{\mu},\mathbf{S})$. However, since $\mathbf{q}(t)$ satisfies the constraints $\mathbf{J}\mathbf{q}(t)-\mathbf{b}\succ 0$, a reflection-based dynamic is proposed in [14] to ensure that $\mathbf{J}\mathbf{q}(t)-\mathbf{b}\in\mathcal{S}_{R-1}$. Precisely, the trajectory described by (10) is used until $\mathbf{q}$ reaches a boundary of its admissible set. The active constraint is considered as a wall on which $\mathbf{q}$ is reflected. The trajectory is then modified until the next wall while preserving the Hamiltonian (6).

The EHMC method allows all the elements of $\mathbf{c}$ to be updated simultaneously with a unitary acceptation ratio. The performance

of this sampler depends on the trajectory length, i.e., the maximum value T of the parameter $t \in]0, 2\pi[$ in (10). Small values can lead to slow distribution exploration whereas large values lead to long trajectories and potentially more bounces onto the constraints. In [14], a tradeoff is proposed by setting $T = \pi/2$. However, computing the bouncing time instants and the new trajectories are the most computationally intensive steps of the method. If the EHMC can be very accurate (in terms of mixing properties and computational cost) for small vectors, it is important to note that the computational burden can be prohibitive for large vectors since the probability of hitting a wall generally grows with the number of constraints (equal to R here). Indeed, it is difficult to predict the number of bounces and thus the algorithm running time in advance. The next paragraph studies a second HMC method, which relies on an accept/reject procedure but whose complexity can be more easily tuned.

4.2. Constrained Hamiltonian Monte Carlo

Efficient sampling from (2) is difficult, mainly because the sum-to-one constraint involves all the elements of $\boldsymbol{a}$. In [17] was proposed the following change of variables

$$a_r = \left(\prod_{k=1}^{r-1} z_k\right) \times \begin{cases} 1 - z_r & \text{if } r < R \\ 1 & \text{if } r = R \end{cases} . \tag{11}$$

One motivation for using the latent variables z_r instead of a_r is the fact that the constraints associated with (1) for $\boldsymbol{a}$ translate for $\boldsymbol{z} = [z_1, \ldots, z_{R-1}]^T$ as $0 < z_r < 1, \quad \forall r \in \{1, \ldots, R-1\}$. As a consequence, the constraints for $\boldsymbol{z}$ are much easier to handle for the sampling procedure than those for the original vector $\boldsymbol{a}$. Straightforward computations lead to

$$f(\boldsymbol{z}|\mathbf{m}, \boldsymbol{\Sigma}) \propto \begin{cases} f(\boldsymbol{a}|\mathbf{m}, \boldsymbol{\Sigma}) f(\mathbf{z}) & \text{if } \boldsymbol{z} \in (0,1)^{R-1} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where $f(\mathbf{z}) \propto \prod_{r=1}^{R-1} z_r^{R-r-1}$ and $\boldsymbol{a}$ is given by (11). In this paragraph, we investigate a constrained HMC (CHMC) method to sample from (12). In contrast to the exact HMC method described in the previous paragraph, a discretization method must be used to simulate Hamiltonian dynamics associated with (12). In this paper, we consider a constrained leapfrog method which consists of N_{LF} steps with a discretization stepsize ϵ. Let $\mathbf{q} = \boldsymbol{z} \in \mathbb{R}^{R-1}$, $\mathbf{M} = \mathbf{I}_{R-1}$ and

$$U(\mathbf{q}) = \frac{1}{2} (\boldsymbol{a} - \mathbf{m})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{a} - \mathbf{m}) - \sum_{r}^{R-1} (R - r - 1) \log(q_r),$$

where $\boldsymbol{a}$ is given by (11). The nth classical leapfrog step can be expressed as

$$\mathbf{p}^{(i,n\epsilon+\epsilon/2)} = \mathbf{p}^{(i,n\epsilon)} - \frac{\epsilon}{2} \frac{\partial U}{\partial \mathbf{q}^T} \left[\mathbf{q}^{(i,n\epsilon)}\right] \tag{13a}$$

$$\mathbf{q}^{(i,(n+1)\epsilon)} = \mathbf{q}^{(i,n\epsilon)} + \epsilon \mathbf{p}^{(i,n\epsilon+\epsilon/2)} \tag{13b}$$

$$\mathbf{p}^{(i,(n+1)\epsilon)} = \mathbf{p}^{(i,n\epsilon+\epsilon/2)} - \frac{\epsilon}{2} \frac{\partial U}{\partial \mathbf{q}^T} \left[\mathbf{q}^{(i,(n+1)\epsilon)}\right]. \tag{13c}$$

The leapfrog method starts with $(\mathbf{q}^{(i,0)}, \tilde{\mathbf{p}}^{(i)}) = (\mathbf{q}^{(i)}, \tilde{\mathbf{p}}^{(i)})$ and the candidate is set after N_{LF} steps to $(\mathbf{q}^*, \mathbf{p}^*) = (\mathbf{q}^{(i,\epsilon N_{\text{LF}})}, \tilde{\mathbf{p}}^{(i,\epsilon N_{\text{LF}})})$.

However, since $\mathbf{q}$ is subject to constraints, more sophisticated discretization methods must be used. In this paper we propose the use of the constrained leapfrog scheme studied in [8, Chap. 5]. Each CHMC iteration starts in a similar way to the classical leapfrog method, with the sequential sampling of the momentum $\mathbf{p}$ (13a) and the vector $\mathbf{q}$ (13b). However, if the generated vector $\mathbf{q}$ violates the constraints, it is modified depending on the violated constraints and the momentum is negated (see [8, Chap. 5] for more details). This step is repeated until each component of the generated $\mathbf{q}$ satisfies the constraints. The CHMC ends with the update of the momentum $\mathbf{p}$ (13c). The performance of the HMC mainly relies on the values of the parameters N_{LF} and ϵ_q which can be tuned using preliminary runs or during the burn-in period (see [8, Chap. 5] and [18]).

5. SIMULATIONS

The performance of the Gibbs sampler introduced in Section 3 and the two HMC methods described in Section 4 are evaluated using the three following scenarios

$$\begin{cases} \text{Scenario \#1:} & R = 3, \mathbf{m} = R^{-1}\mathbf{1}_R, \boldsymbol{\Sigma} = 10^{-2}\boldsymbol{\Sigma}_0 \\ \text{Scenario \#2:} & R = 20, \mathbf{m} = R^{-1}\mathbf{1}_R, \boldsymbol{\Sigma} = 10^{-2}\boldsymbol{\Sigma}_0 \\ \text{Scenario \#3:} & R = 20, \mathbf{m} = R^{-1}\mathbf{1}_R, \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 \end{cases}$$

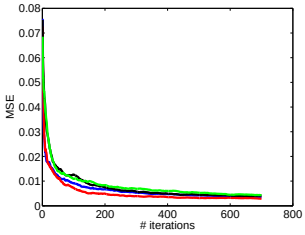
characterised by different vector sizes $R \in \{3, 20\}$ and covariance matrices, where

$$[\boldsymbol{\Sigma}_0]_{i,j} = \exp\left(\frac{|i-j|}{R}\right), \quad \forall i, j \in \{1, \ldots, R\}. \tag{14}$$

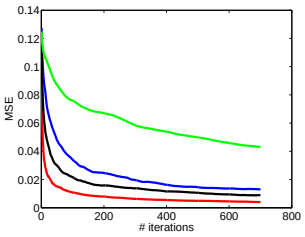
This choice of covariance matrix is mainly motivated by the resulting high correlations between the components of $\boldsymbol{a}$. Note that the mean $\mathbf{m}$ belongs to the simplex $\mathcal{S}_{R-1}$ for each scenario. The number of leapfrog steps for the CHMC sampler studied in paragraph 4.2 has been arbitrarily set to $N_{\text{LF}} = 1$ (constrained Langevin move denoted CL in the sequel) and $N_{\text{LF}} = 25$. The stepsizes have been fixed from preliminary runs to obtain an acceptance rate around 55% for each scenario. The mixing properties of the Gibbs, EHMC, CL and CHMC samplers are evaluated using the root mean squared errors (RMSEs) defined as $RMSE(t) = \sqrt{\|\hat{\boldsymbol{a}}_{\text{MMSE}}(t) - \hat{\boldsymbol{a}}_{\text{MMSE}}(\infty)\|^2}$ where $\hat{\boldsymbol{a}}_{\text{MMSE}}(\infty) = \text{E}[f(\boldsymbol{a}|\mathbf{m}, \boldsymbol{\Sigma})]$ and

$$\hat{\boldsymbol{a}}_{\text{MMSE}}(t) = \frac{1}{t} \sum_{i=N_{\text{bi}}+1}^{N_{\text{bi}}+t} \boldsymbol{a}^{(i)} \tag{15}$$

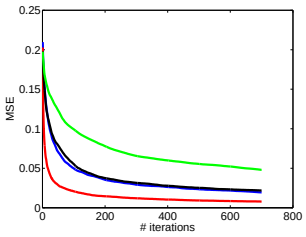
is the minimum mean square error (MMSE) estimator of $\boldsymbol{a}$ after $t + N_{\text{bi}}$ iterations. Moreover, N_{bi} stands for the number of burn-in iterations and $\boldsymbol{a}^{(i)}$ is the vector generated at the ith iteration. In this paper, N_{bi} has been set to $N_{\text{bi}} = 300$ for all simulations and samplers. The lower the RMSE, the better the mixing properties of the sampler. Fig. 1 shows the RMSEs of the three samplers for the first $N_{\text{r}} = 700$ iterations (after removing the burn-in iterations) averaged over 50 random initializations. Table 1 provides the average time required to achieve the $N_{\text{r}} + N_{\text{bi}} = 1000$ first iterations. For the first scenario, the results show that the four samplers mix in a similar fashion. However, for scenarios #2 and #3 ($R = 20$), the CL is less accurate than the Gibbs sampler and the other HMC methods (mainly because of a too small N_{LF}) and the EHMC mixes better than the CHMC. From a computational point of view, the Gibbs and EHMC samplers have similar behavior and their complexity increases with the number of parameters R. For the former, the processing time increases mainly because of the number of Gibbs steps (equal to $(R-1)$) while the main computational cost of the latter consists of computing the derivatives in (13a) and (13b). As mentioned above, the complexity of the EHMC is difficult to predict and relies on the parameters (mean and covariance) of the distribution, as well as the sample size R.



(a) Scenario #1



(b) Scenario #2



(c) Scenario #3

Fig. 1. RMSEs of the Gibbs (blue), EHMC (red) and CHMC with $N_{\text{LF}} = 1$ (green) and $N_{\text{LF}} = 25$ (black) for the scenarios #1 to #3.

	Gibbs	EHMC	CHMC	
			$N_{\text{LF}} = 1$	$N_{\text{LF}} = 25$
Scenario #1	0.20	**0.04**	0.28	1.50
Scenario #2	2.12	**0.29**	0.68	6.69
Scenario #3	2.20	7.57	**0.69**	6.76

Table 1. Processing time for $N = 1000$ it. (in s):

6. CONLUSION

In this paper, we have reviewed two recent HMC methods to sample from multivariate Gaussian distributions restricted to a simplex. The EHMC is a powerful sampler with unitary acceptance ratio which can drastically improve the mixing properties when compared with the classical Gibbs sampler. However, its computational complexity can be prohibitive, especially for large dimensional spaces. The CHMC, which relies on an accept/reject procedure, can also improve the mixing properties of the chains with a controled computational cost. However, the gain obtained with this sampler depends on the leapfrog parameters which must be tuned carefully from preliminary runs. Both HMC methods can be considered as interesting alternatives to the Gibbs sampler but the choice of the method can depend on the application and relies on a tradeoff between good mixing properties and computational considerations. Finally, it is important to note that the EHMC can be applied more general TMGDs with linear and/or quadratic constraints [14], whereas the CHMC can be easily applied to non-Gaussian distributions restricted to the simplex $\mathcal{S}_{R-1}$ [18].

7. REFERENCES

[1] C. P. Robert, "Simulation of truncated normal variables," *Statistics and Computing*, vol. 5, pp. 121–125, 1995.

[2] P. Damien and S. G. Walker, "Sampling truncated normal, beta, and gamma densities," *Journal of Computational and Graphical Statistics*, vol. 10, no. 2, pp. 206–215, 2001.

[3] N. Chopin, "Fast simulation of truncated Gaussian distributions," *Statistics and Computing*, vol. 21, no. 2, pp. 275–288, 2011.

[4] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, "Bayesian separation of spectral sources under non-negativity and full additivity constraints," *Signal Processing*, vol. 89, no. 12, pp. 2657–2669, Dec. 2009.

[5] N. Dobigeon, J.-Y. Tourneret, and C.-I. Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, July 2008.

[6] C. Bazot, N. Dobigeon, J.-Y. Tourneret, A. K. Zaas, G. S. Ginsburg, and A. O. Hero, "Unsupervised Bayesian linear unmixing of gene expression microarrays," *BMC Bioinformatics*, vol. 14, no. 99, March 2013.

[7] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid monte carlo," *Physics Letters B*, vol. 195, pp. 216–222, Sept. 1987.

[8] S. Brooks, *Handbook of Markov Chain Monte Carlo*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis, 2011.

[9] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, New York, 1999.

[10] R. M. Neal, *Bayesian Learning for Neural Networks (Lecture Notes in Statistics)*, Springer, Secaucus, NJ, USA, 1 edition, Aug. 1996.

[11] M. N. Schmidt, "Function factorization using warped Gaussian processes.," in *Proc. Int. Conf. Machine Learning*, 2009, vol. 382, p. 116.

[12] C. Hartmann and C. Schuette, "A constrained hybrid Monte-Carlo algorithm and the problem of calculating the free energy in several variables," *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift fr Angewandte Mathematik und Mechanik*, vol. 85, no. 10, pp. 700–710, 2005.

[13] M. A. Brubaker, M. Salzmann, and R. Urtasun, "A family of MCMC methods on implicitly defined manifolds.," *J. Machine Learning Research - Proceedings Track*, vol. 22, pp. 161–172, 2012.

[14] A. Pakman and L. Paninski, "Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians," *ArXiv e-prints*, Aug. 2012.

[15] N. Dobigeon and J.-Y. Tourneret, "Recursive computation of the normalization constant of a multivariate Gaussian distribution truncated on a simplex," Tech. Rep., University of Michigan, USA, Jan. 2008.

[16] Nicolas Dobigeon and Jean-Yves Tourneret, "Efficient sampling according to a multivariate Gaussian distribution truncated on a simplex," Tech. Rep., IRIT/ENSEEIHT/TéSA, March 2007.

[17] M. J. Betancourt, "Cruising The Simplex: Hamiltonian Monte Carlo and the Dirichlet Distribution," *ArXiv e-prints*, Oct. 2010.

[18] Y. Altmann, N. Dobigeon, and J. Tourneret, "Unsupervised post-nonlinear unmixing of hyperspectral images using a Hamiltonian Monte Carlo algorithm," *IEEE Trans. Image Process.*, 2014, to appear.