Truncated multivariate normal distribution under nonlinear constraints

Hassan Maatouk, Didier Rullière, Xavier Bay

Truncated multivariate normal distribution under nonlinear constraints

Hassan Maatouk[†1], Didier Rullière[‡2] and Xavier Bay[‡3]

(†) LAMPS, Université de Perpignan via Domitia, 52 av. Paul Alduy, 66860 Cedex 9 Perpignan, France
(‡) Mines Saint-Étienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, Saint-Étienne, F-42023, France

Abstract

Generating truncated multivariate normal distributions is widely used in Bayesian constrained statistical modeling. This technique is applied in various fields, including ecology, economics, physics, computer science, biology, geosciences, and machine learning. In this paper, the efficient approach developed by [37] for generating positive Gaussian vectors is considered. Their main idea is to incorporate a *smooth relaxation* of the complex constraints appearing in the constrained density function into the likelihood and to employ a highly efficient Markov Chain Monte Carlo (MCMC) sampler. Our contributions are fourfold. First, we extend this approach to address linear, quadratic, and nonlinear inequality constraints, which can be applied individually or in combination. The functions generating the nonlinear inequality constraints can be piecewise continuous or continuously differentiable of any order. Second, we propose updating the approximate parameter in the likelihood at each MCMC iteration to enhance the stability and ensure the convergence of the algorithm. This allows the proposed approach to handle extreme cases that are beyond the reach of existing samplers. Third, for boundedness constraints with constant bounds, we develop an efficient formula for the log-likelihood function, which reduces computational complexity and improves efficiency in high-dimensional settings with respect to computational running time. Fourth, we explore flexibility and performance of the proposed approach through studies on both synthetic and real data within the context of Bayesian shape-restricted function estimation. A comparison with the efficient Hamiltonian Monte Carlo (HMC) sampler is included. In contrast to the HMC sampler, the starting point of the proposed MCMC does not need to satisfy the inequality constraints.

Keywords Truncated Gaussian · Elliptical slice sampling · Smooth relaxation · Nonlinear constraints · MCMC

1 Introduction

Sampling truncated multivariate normal (tMVN) distribution is commonly used in Bayesian constraints statistics [15, 24, 40, 47] and is applied in many fields such as econometrics [5, 8, 9], nuclear

[1]hassan.maatouk@univ-perp.fr
[2]drulliere@emse.fr
[3]bay@emse.fr

physics [46], and machine learning [17, 44]. For instance, in the context of shape-restricted function estimation, structured shape constraints such as monotonicity, boundedness, and convexity are often addressed by expanding the function space using an appropriate basis, where the shape constraints are enforced by imposing linear inequality constraints on the basis coefficients. Examples of such bases include constrained splines [4, 30], Bernstein polynomials [11, 43], piecewise linear functions [13], and compactly supported basis [24, 26]. Other settings where the tMVN distribution has been investigated include multinomial probit and logit models [1, 29], Bayesian bridge [36], and unmixing of hyperspectral data [12].

Generating high-dimensional tMVN distribution is an active area of research [20], for which existing efficient algorithms include Gibbs sampling [18, 21, 38, 41], Metropolis-Hastings (MH) [7, 32], the minimax tilting method [2], and the highly efficient Hamiltonian Monte Carlo (HMC) sampler [35]. All of these algorithms address only linear inequality constraints except for the HMC sampler implemented in the `R`-package *tmg*, which handles both linear and quadratic inequality constraints. We refer the reader to [14] for statistical applications in the cases involving quadratic constraints.

The Markov Chain Monte Carlo (MCMC) approach developed in the present paper is general in the sense that it can generate a Gaussian vector subject to linear, quadratic, and nonlinear inequality constraints. This technique is based on sampling from the prior, which can offer significant computational advantages, particularly when the covariance structure of the prior respects certain properties such as stationarity, sparsity or Toeplitz. For example, the authors in [6] proposed a fast algorithm for generating a hyperplane-truncated multivariate normal (MVN) distribution when the covariance matrix can be expressed as a positive-definite matrix minus (or plus) a low-rank symmetric matrix. The main idea of the approach developed in the present paper is to approximate the indicator function of the set of hard constraints appearing in the posterior probability density function (pdf) by a smooth function and use efficient Elliptical Slice Sampling (ESS) [31] to generate samples from the resulting posterior distribution. It allows incorporating linear, quadratic, and nonlinear inequality constraints, either applied individually or in combination. Furthermore, it handles functions that are piecewise continuous and of any order continuously differentiable, which generate nonlinear inequality constraints. Finally, it demonstrates high flexibility and successfully generates samples with complex constraints in extreme cases—cases that are impossible for all other samplers. This is due to the approximate parameter, which can be updated at each MCMC iteration. As the proposed approach is based on sampling before conditioning rather than after, we refer the reader to [42], where the authors provide a complete review and a comparison of computational approaches for generating high-dimensional Gaussian vectors proposed in various communities, ranging from iterative numerical linear algebra to MCMC approaches.

In this paper, we consider the methodology for generating the tMVN distribution using ESS and a smooth relaxation of the set of constraints originally proposed by [37]. This methodology has been employed by [28, 27] in the contexts of large-scale and high-dimensional shape-restricted function estimation, respectively. We extend this methodology in four different directions:

- First, we generalize this methodology to address linear, quadratic, and nonlinear inequality constraints, either applied individually or in combination. To the best of our knowledge, this is the only approach in the literature that handles general nonlinear constraints. The functions appearing in the nonlinear constraints can be piecewise continuous and continuously differentiable to any order.

- Second, we update the approximate parameter of the set of constraints at each MCMC iteration, which allows us to handle extreme cases that are impossible for all other existing approaches. Furthermore, we propose projecting the samples onto the set of linear constraints in order to strictly enforce them.

- Third, we develop a new efficient formula for the log-likelihood function for boundedness constraints, which reduces the computational complexity of the proposed approach in high-dimensions. Additionally, we compare its efficiency with the state-of-the-art HMC approach in terms of computational runtime.

- Fourth, we explore the applicability and flexibility of the proposed MCMC approach in the context of Bayesian estimation for shape-restricted functions.

The rest of the article is structured as follows. Section 2 is dedicated to the proposed general methodology for sampling from the tMVN distribution, constrained by linear, quadratic, and nonlinear constraints. We also highlight the challenges involved in sampling a tMVN distribution. In Section 3, the challenges associated with sampling tMVN distribution are investigated and an updating strategy of the approximate parameter in the likelihood is proposed. Section 4 is devoted to several numerical illustrations for both fixed and updated η. In Section 5, we explore the performance of the proposed approach within the context of Bayesian shape-restricted function estimation. Additionally, we provide a comparison with the HMC sampler in terms of computational running time and prediction accuracy.

2 Truncated Gaussian vectors under nonlinear constraints

Let $\boldsymbol{\zeta} \in \mathbb{R}^d$ be a d-dimensional Gaussian vector with mean vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and symmetric, positive-definite covariance matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{d\times d}$, i.e., $\boldsymbol{\zeta} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Gamma})$. In this section, we are interested in sampling from a constrained pdf with the following form:

$$f(\boldsymbol{\zeta}|\boldsymbol{\zeta} \in \mathcal{C}) \propto \exp\left(-\frac{1}{2}[\boldsymbol{\zeta}-\boldsymbol{\mu}]^\top \boldsymbol{\Gamma}^{-1}[\boldsymbol{\zeta}-\boldsymbol{\mu}]\right)\mathbf{1}_{\mathcal{C}}(\boldsymbol{\zeta}), \quad \boldsymbol{\zeta} \in \mathbb{R}^d, \tag{1}$$

where $\mathcal{C}$ is any convex or non-convex set in $\mathbb{R}^d$ representing the set of $m \in \mathbb{N}^*$ linear, quadratic and/or nonlinear inequality constraints on the Gaussian vector $\boldsymbol{\zeta}$. This problem appears in many Bayesian constrained statistical modeling. The indicator function in the above pdf, as shown in Equation (1), represents the restricted set of inequality constraints. The density f in (1) corresponds to a tMVN distribution $\mathcal{TN}_d(\boldsymbol{\mu}, \boldsymbol{\Gamma}; \mathcal{C})$, where the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$, and $\mathcal{C}$ represent the mean, covariance, and restricted convex and non-convex sets, respectively. Here, d represents the dimension of the prior random vector $\boldsymbol{\zeta}$. Let us consider the following notations:

$$\mathcal{C} = \begin{cases} \mathcal{C}_{\text{lin}} := \left\{\boldsymbol{\zeta} \in \mathbb{R}^d \ : \ \boldsymbol{a}_\kappa^\top \boldsymbol{\zeta} + b_\kappa \geq 0, \ \kappa = 1, \ldots, m\right\} \\ \mathcal{C}_{\text{quad}} := \left\{\boldsymbol{\zeta} \in \mathbb{R}^d \ : \ \boldsymbol{\zeta}^\top \boldsymbol{C}_\kappa \boldsymbol{\zeta} + \boldsymbol{d}_\kappa^\top \boldsymbol{\zeta} + e_\kappa \geq 0, \ \kappa = 1, \ldots, m\right\} \\ \mathcal{C}_{\text{nlin}} := \left\{\boldsymbol{\zeta} \in \mathbb{R}^d \ : \ g_\kappa(\boldsymbol{\zeta}) \geq 0, \ \kappa = 1, \ldots, m\right\} \end{cases} \tag{2}$$

which corresponds to m linear, quadratic and nonlinear inequality constraints, respectively. Let us mention that the m linear inequality constraints, given by $\boldsymbol{a}_\kappa^\top \boldsymbol{\zeta} + b_\kappa \geq 0$ for $\kappa = 1, \ldots, m$ in (2), can be expressed in matrix form as follows:

$$\boldsymbol{A}\boldsymbol{\zeta} + \boldsymbol{b} \geq \mathbf{0},$$

where $\boldsymbol{a}_\kappa \in \mathbb{R}^d$ represents the κ^{th} row of the $m \times d$ matrix of constraints $\boldsymbol{A}$, b_κ represents the κ^{th} component of the vector $\boldsymbol{b} \in \mathbb{R}^m$, and $\mathbf{0}$ represents the m-dimensional zero vector. For the quadratic inequality constraints $\mathcal{C}_{\text{quad}}$ in (2), $\boldsymbol{C}_\kappa$ is a $d \times d$ matrix, $\boldsymbol{d}_\kappa$ is a d-dimensional vector and e_κ is a real number. Finally, for the nonlinear inequality constraints $\mathcal{C}_{\text{nlin}}$, g_κ represents any nonlinear piecewise continuous or any order continuously differentiable function, for any $\kappa = 1, \ldots, m$.

Let us mention that sampling tMVN distribution is an active area of research, for which existing algorithms include Gibbs sampling [41], MH [7, 32], the minimax tilting method accept-reject

sampler [2], the exact rejection sampling from the mode [23], and the highly efficient HMC sampler [35]. The approach developed in the present paper is quite different, as sampling is performed before conditioning rather than after. This can offer significant computational advantages, particularly when the covariance matrix of the prior respects certain properties such as sparsity or block circulant (Toeplitz), or when it can be decomposed as a positive-definite matrix minus (plus) a low-rank symmetric matrix [6]. These advantages will be highlighted in the numerical examples of the present paper, within the context of Bayesian shape-restricted function estimation.

2.1 Methodology development

The methodology developed in the present paper draws inspiration from the recent efficient approach developed by [37] for generating samples from the MVN distribution $\boldsymbol{\zeta} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ restricted to the positive orthant (i.e., $\boldsymbol{\zeta} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ such that $\boldsymbol{\zeta} \in [0, +\infty[^d$ or $\boldsymbol{\zeta} \sim \mathcal{TN}_d(\boldsymbol{\mu}, \boldsymbol{\Gamma}; \mathcal{C})$, where $\mathcal{C}$ is the positive orthant $[0, +\infty[^d)$. Their main interesting idea is to approximate the indicator function $\mathbf{1}_{[0,+\infty[}(\cdot)$ appearing in the constrained pdf (1) with a smooth function and incorporate it into a likelihood function. They used the logistic sigmoid function $1/[1 + \exp(-x)]$, which represents the cumulative distribution function (cdf) of the logistic distribution. To be more precise, they used a scaled logistic sigmoid function:

$$\mathbf{1}_{[0,+\infty[}(\zeta) \approx \frac{1}{1 + e^{-\eta\zeta}}, \quad \forall \zeta \in \mathbb{R}, \tag{3}$$

where $\eta > 0$ is a parameter controlling the quality of the approximation. The approximation in (3) is enhanced with a large value of η. The parameter η will play a key role in investigating the flexibility of the proposed approach. Furthermore, in the context of shape-restricted function estimation, the parameter η can be seen as a trade-off between the quality of the prediction and mitigating the *mass-shifting* phenomenon described in [47]. From (3), one can deduce the following:

$$\mathbf{1}_{[0,+\infty[^d}(\boldsymbol{\zeta}) \approx \prod_{\kappa=1}^{d} \frac{1}{1 + e^{-\eta\zeta_\kappa}}, \quad \forall \boldsymbol{\zeta} \in \mathbb{R}^d,$$

where ζ_κ is the κ^{th} component of the vector $\boldsymbol{\zeta}$.

Our aim in this paper is to generalize this approach to handle sets of $m \in \mathbb{N}^*$ linear, quadratic, and nonlinear inequality constraints given by (2), either applied individually or in combination. This enables the proposed approach to handle both convex and non-convex restricted sets. Moreover, it broadens the applicability of the proposed approach to include more general Bayesian constrained models and to address more complex problems, such as nonlinear optimization and multiple and high-dimensional shape restricted function estimation. We pay carefully attention on the choice of the approximate parameter η in order to ensure stability and convergence of the proposed approach. We show later in this paper how this parameter, η, can be updated at each MCMC iteration.

First, let us present the following result, which generalizes the findings in [37] for nonlinear inequality constraints:

Proposition 1 (Approximation of indicator function of inequality constraints)**.** *Let g be any piecewise continuous function defined on $\mathbb{R}^d$. Then, the indicator function $\mathbf{1}_{\{g(\cdot)\geq 0\}}$ can be approximated by a scaled logistic sigmoid function as follows:*

$$\mathbf{1}_{\{g(\cdot)\geq 0\}}(\boldsymbol{x}) \approx \frac{1}{1 + \exp[-\eta g(\boldsymbol{x})]}, \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{4}$$

where the parameter $\eta > 0$ controls the quality of the approximation. As η increases, the quality of the approximation improves. The approximate function in (4) possesses the same degree of differentiability as g and we have

$$\left| \mathbf{1}_{\{g(\cdot)\geq 0\}}(\boldsymbol{x}) - \frac{1}{1+e^{-\eta g(\boldsymbol{x})}} \right| \leq \frac{1}{1+e^{\eta |g(\boldsymbol{x})|}},$$

for any $\boldsymbol{x} \in \mathbb{R}^d$.

Proof. Let us consider two cases. When $g(\boldsymbol{x}) < 0$, then $\mathbf{1}_{\{g(\boldsymbol{x})\geq 0\}} = 0$, for any $\boldsymbol{x} \in \mathbb{R}^d$. Thus,

$$\left| \mathbf{1}_{\{g(\cdot)\geq 0\}}(\boldsymbol{x}) - \frac{1}{1+e^{-\eta g(\boldsymbol{x})}} \right| \leq \frac{1}{1+e^{-\eta |g(\boldsymbol{x})|}} \iff \frac{1}{1+e^{-\eta g(\boldsymbol{x})}} \leq \frac{1}{1+e^{\eta |g(\boldsymbol{x})|}}.$$

The second case, when $g(\boldsymbol{x}) \geq 0$, holds since $1 - 1/[1+\exp\{-\eta g(\boldsymbol{x})\}] = 1/[1+\exp\{\eta g(\boldsymbol{x})\}]$, and thus completing the proof of the proposition. □

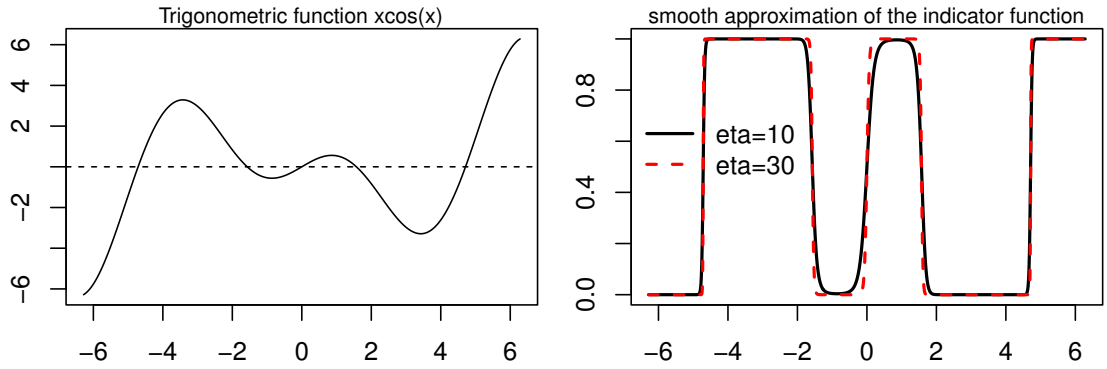

Figure 1: Smooth approximations of the indicator function $\mathbf{1}_{\{g(\cdot)\geq 0\}}(x)$ on $[-2\pi, 2\pi]$ for different values of η, where $g(x) = x\cos(x)$ (black curve in the left panel).

Figure 1 displays smooth approximations of the indicator function $\mathbf{1}_{\{g(\cdot)\geq 0\}}(x)$ on the interval $[-2\pi, 2\pi]$ for different values of η using (4), where $g(x) = x\cos(x)$ is the nonlinear continuous function represented by the black curve in the left panel.

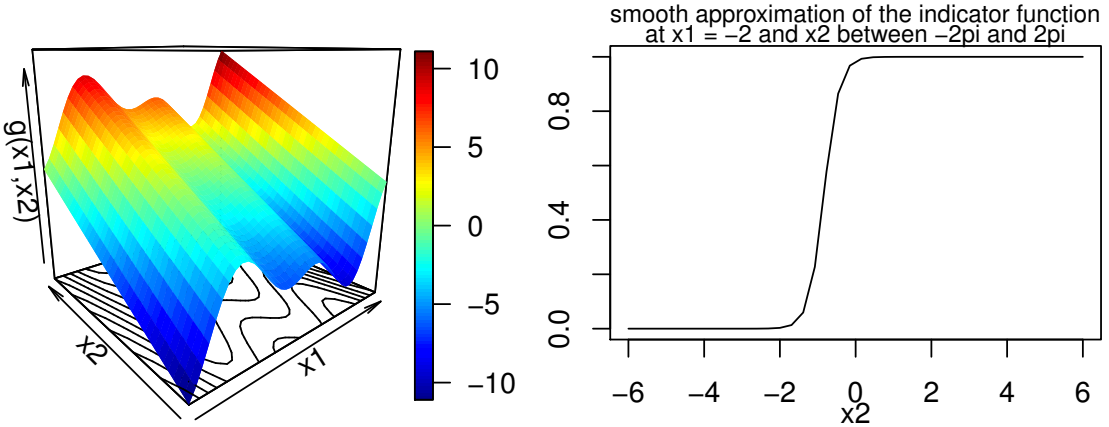

Figure 2: The two-dimensional continuous function $g(x_1, x_2) = x_1\cos(x_1) + x_2$ on $[-2\pi, 2\pi]^2$ (left panel). Smooth approximation of the indicator function $\mathbf{1}_{\{g(\cdot,\cdot)\geq 0\}}(x_1, x_2)$ at $x_1 = -2$ and $x_2 \in [-2\pi, 2\pi]$ for $\eta = 5$ (right panel).

Figure 2 displays a smooth approximation of the indicator function $\mathbf{1}_{\{g(\cdot,\cdot)\geq 0\}}(-2, x_2)$, for $x_2 \in [-2\pi, 2\pi]$ using (4) with $\eta = 5$, where $g(x_1, x_2) = x_1\cos(x_2) + x_2$ is the two-dimensional nonlinear continuous function shown in the left panel.

Now, when applying m linear, quadratic or nonlinear inequality constraints, the associated indicator function, i.e., $\mathbf{1}_{\mathcal{C}_{\text{lin}}}(\cdot)$, $\mathbf{1}_{\mathcal{C}_{\text{quad}}}(\cdot)$ and $\mathbf{1}_{\mathcal{C}_{\text{nlin}}}(\cdot)$, can be approximated as follows:

$$\begin{aligned}
\mathbf{1}_{\mathcal{C}_{\text{lin}}}(\boldsymbol{\zeta}) &\approx \prod_{\kappa=1}^{m} \frac{1}{1+\exp[-\eta(\boldsymbol{a}_\kappa^\top \boldsymbol{\zeta} + b_\kappa)]} = \gamma_\eta(\boldsymbol{\zeta}), \quad \forall \boldsymbol{\zeta} \in \mathbb{R}^d, \\
\mathbf{1}_{\mathcal{C}_{\text{quad}}}(\boldsymbol{\zeta}) &\approx \prod_{\kappa=1}^{m} \frac{1}{1+\exp[-\eta(\boldsymbol{\zeta}^\top \boldsymbol{C}_\kappa \boldsymbol{\zeta} + \boldsymbol{d}_\kappa^\top \boldsymbol{\zeta} + e_\kappa)]} = \gamma_\eta(\boldsymbol{\zeta}), \quad \forall \boldsymbol{\zeta} \in \mathbb{R}^d, \\
\mathbf{1}_{\mathcal{C}_{\text{nlin}}}(\boldsymbol{\zeta}) &\approx \prod_{\kappa=1}^{m} \frac{1}{1+\exp[-\eta g_\kappa(\boldsymbol{\zeta})]} = \gamma_\eta(\boldsymbol{\zeta}), \quad \forall \boldsymbol{\zeta} \in \mathbb{R}^d,
\end{aligned} \tag{5}$$

where γ_η will act as a likelihood function.

Let us first focus on the nonlinear inequality constraints $\mathcal{C}_{\text{nlin}}$, which can be seen as a generalization of other inequality constraints, such as linear and quadratic constraints. Consequently, the constrained pdf in (1) can be expressed as follows:

$$f(\boldsymbol{\zeta}|\boldsymbol{\zeta} \in \mathcal{C}_{\text{nlin}}) \propto \exp\left(-\frac{1}{2}[\boldsymbol{\zeta}-\boldsymbol{\mu}]^\top \boldsymbol{\Gamma}^{-1}[\boldsymbol{\zeta}-\boldsymbol{\mu}]\right) \mathbf{1}_{\mathcal{C}_{\text{nlin}}}(\boldsymbol{\zeta}), \quad \boldsymbol{\zeta} \in \mathbb{R}^d. \tag{6}$$

By substituting the above approximate γ_η from Equation (5) into Equation (6), we obtain the following density approximation $\tilde{f}_\eta$ of f:

$$\begin{aligned}
\tilde{f}_\eta(\boldsymbol{\zeta}|\boldsymbol{\zeta} \in \mathcal{C}_{\text{nlin}}) &\propto \exp\left(-\frac{1}{2}[\boldsymbol{\zeta}-\boldsymbol{\mu}]^\top \boldsymbol{\Gamma}^{-1}[\boldsymbol{\zeta}-\boldsymbol{\mu}]\right) \gamma_\eta(\boldsymbol{\zeta}) \\
&\propto \underbrace{\left[\prod_{\kappa=1}^{m} \frac{1}{1+\exp[-\eta g_\kappa(\boldsymbol{\zeta})]}\right]}_{L_\eta(\boldsymbol{\zeta}) \,:\, \text{likelihood function}} \underbrace{\exp\left(-\frac{1}{2}[\boldsymbol{\zeta}-\boldsymbol{\mu}]^\top \boldsymbol{\Gamma}^{-1}[\boldsymbol{\zeta}-\boldsymbol{\mu}]\right)}_{\text{MVN prior}},
\end{aligned} \tag{7}$$

for any $\boldsymbol{\zeta} \in \mathbb{R}^d$. By a simple change of variable $\boldsymbol{\xi} = \boldsymbol{\zeta} - \boldsymbol{\mu}$, the expression in (7) become

$$\underbrace{\left[\prod_{\kappa=1}^{m} \frac{1}{1+\exp[-\eta g_\kappa(\boldsymbol{\xi}+\boldsymbol{\mu})]}\right]}_{L_\eta(\boldsymbol{\xi}) \,:\, \text{likelihood function}} \underbrace{\exp\left(-\frac{1}{2}\boldsymbol{\xi}^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\xi}\right)}_{\text{zero mean MVN}}. \tag{8}$$

The approximate pdf in (8) is proportional to the product of a likelihood function, denoted $L_\eta(\boldsymbol{\xi})$, and an untruncated prior, which is a zero-mean Gaussian vector $\boldsymbol{\xi} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Gamma})$. It is worth noting that the change of variables in Equation (8) enables the use of the ESS approach and ensures that the term within the square brackets is a likelihood function.

Proposition 2 below extends the result in [37] for nonlinear inequality constraints.

Proposition 2. *Assume $\lambda_d(\cap_{\kappa=1}^m \{g_\kappa = 0\}) = 0$, where λ_d is the Lebesgue-measure on $\mathbb{R}^d$. Let f and $\tilde{f}_\eta$ be respectively defined as in (6) and (7). Then,*

$$\lim_{\eta \to +\infty} \int_{\mathbb{R}^d} \left| f(\boldsymbol{\zeta}) - \tilde{f}_\eta(\boldsymbol{\zeta}) \right| \mathrm{d}\boldsymbol{\zeta} = 0.$$

This means that the L_1 distance between $\tilde{f}_\eta$ and f converges to 0 as η goes to infinity.

Proof. Suppose first that $m = 1$, then

$$\begin{aligned}\int_{\mathbb{R}^d} \left| f(\boldsymbol{\zeta}) - \tilde{f}_\eta(\boldsymbol{\zeta}) \right| \mathrm{d}\boldsymbol{\zeta} &\propto \int_{\mathbb{R}^d} e^{-\frac{1}{2}[\boldsymbol{\zeta}-\boldsymbol{\mu}]^\top \boldsymbol{\Gamma}^{-1}[\boldsymbol{\zeta}-\boldsymbol{\mu}]} \left| \mathbf{1}_{g_1(\cdot)(\boldsymbol{\zeta})} - \frac{1}{1+e^{-\eta g_1(\boldsymbol{\zeta})}} \right| \mathrm{d}\boldsymbol{\zeta} \\ &\leq \int_{\mathbb{R}^d} e^{-\frac{1}{2}[\boldsymbol{\zeta}-\boldsymbol{\mu}]^\top \boldsymbol{\Gamma}^{-1}[\boldsymbol{\zeta}-\boldsymbol{\mu}]} \frac{1}{1+e^{\eta|g_1(\boldsymbol{\zeta})|}} \mathrm{d}\boldsymbol{\zeta}.\end{aligned}$$

The last inequality holds by Proposition 1. Since $\lambda_d(g_1 = 0)$ equals zero, we can apply the dominated convergence theorem to get the result, as

$$\forall \eta > 0, \quad e^{-\frac{1}{2}[\boldsymbol{\zeta}-\boldsymbol{\mu}]^\top \boldsymbol{\Gamma}^{-1}[\boldsymbol{\zeta}-\boldsymbol{\mu}]} \frac{1}{1+e^{\eta|g_1(\boldsymbol{\zeta})|}} \leq e^{-\frac{1}{2}[\boldsymbol{\zeta}-\boldsymbol{\mu}]^\top \boldsymbol{\Gamma}^{-1}[\boldsymbol{\zeta}-\boldsymbol{\mu}]},$$

and

$$e^{-\frac{1}{2}[\boldsymbol{\zeta}-\boldsymbol{\mu}]^\top \boldsymbol{\Gamma}^{-1}[\boldsymbol{\zeta}-\boldsymbol{\mu}]} \frac{1}{1+e^{\eta|g_1(\boldsymbol{\zeta})|}} \underset{\eta \to 0}{\longrightarrow} 0,$$

λ_d-almost everywhere, since $\lambda_d(g_1 = 0)$ is equal to zero. The result holds for $m \geq 2$ by using the fact that

$$\left| \mathbf{1}_{\{g_\kappa(\cdot) \geq 0,\ \kappa = 1,\ldots,m\}}(\boldsymbol{\zeta}) - \prod_{\kappa=1}^{m} \frac{1}{1+e^{-\eta g_\kappa(\boldsymbol{\zeta})}} \right| \leq 1, \tag{9}$$

and observing that the left-hand side of the inequality (9) tends to zero almost everywhere with respect to λ_d, as η tends to infinity. □

2.2 Algorithm development

The approximate density function $\tilde{f}_\eta$ in (8) is proportional to a product of a likelihood function $L_\eta(\cdot)$ and a zero-mean prior Gaussian vector $\boldsymbol{\xi}$:

$$\tilde{f}_\eta(\boldsymbol{\xi}) \propto L_\eta(\boldsymbol{\xi}) \mathcal{N}_d(\boldsymbol{\xi}; \mathbf{0}, \boldsymbol{\Gamma}). \tag{10}$$

In this context, sampling from (10) can be performed using the efficient ESS developed by [31] which can be seen as a parametrization of the MH proposals [7, 32]:

$$\boldsymbol{\xi}' = \sin(\theta)\boldsymbol{\nu} + \cos(\theta)\boldsymbol{\xi}, \quad \boldsymbol{\nu} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Gamma}), \tag{11}$$

where $\boldsymbol{\xi}$ is the current state, $\boldsymbol{\xi}'$ is the proposal state, and $\boldsymbol{\nu}$ is an auxiliary variable. The ESS is a variant form of slice sampling [33] that leverages the Gaussian prior to enhance mixing efficiency and remove the need for parameter tuning. Let us recall that the ESS acceptance ratio $\alpha = \min\{1, L_\eta(\boldsymbol{\xi}')/L_\eta(\boldsymbol{\xi})\}$ relies only on the likelihood ratio and is independent of θ. The angle θ is uniformly generated from a $[\theta_{\min}, \theta_{\max}]$ interval which is shrunk exponentially fast until an acceptable state is reached. The authors in [31] provide details instructions on how to shrink the bracket. The new state $\boldsymbol{\xi}'$ is accepted if it satisfies $U L_\eta(\boldsymbol{\xi}) \leq L_\eta(\boldsymbol{\xi}')$, where $U \sim \mathcal{U}([0,1])$. Hence, the acceptance ratio is

$$\mathrm{P}\left(U \leq \frac{L_\eta(\boldsymbol{\xi}')}{L_\eta(\boldsymbol{\xi})} \right) = \min\left\{1, L_\eta(\boldsymbol{\xi}')/L_\eta(\boldsymbol{\xi})\right\}.$$

From (8), the associated logarithm function of $L_\eta(\boldsymbol{\xi})$ is given by

$$\log(L_\eta(\boldsymbol{\xi})) = -\sum_{\kappa=1}^{m} \log\left(1 + \exp[-\eta g_\kappa(\boldsymbol{\xi} + \boldsymbol{\mu})]\right), \quad \forall \boldsymbol{\xi} \in \mathbb{R}^d. \tag{12}$$

Now, sampling from the approximate density in (10) using ESS involves sampling from a Gaussian prior distribution $\boldsymbol{\nu} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Gamma})$ (auxiliary variable), which admits a computational complexity of order $\mathcal{O}(d^3)$ in general. We refer the reader to [42] for a complete review on various approaches for generating high-dimensional MVN distribution. However, in some situations, such as in high-dimensions, the covariance matrix of the prior, $\boldsymbol{\Gamma}$, exhibits particular structures like sparsity or stationarity. In that case, existing highly efficient algorithm can be employed such as the Fast Fourier Transform [45] and the recent large-scale approach [25] (see Section 5 for more details).

Lemma 1 (Boundedness constraints)**.** *Let us suppose that the Gaussian vector $\boldsymbol{\zeta} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ is bounded between lower and upper vectors $\boldsymbol{l} \in \mathbb{R}^d$ and $\boldsymbol{u} \in \mathbb{R}^d$, respectively. In that case, the set of linear inequality constraints is $\mathcal{C}_{lin} = \{\boldsymbol{\zeta} \in \mathbb{R}^d \ : \ \boldsymbol{l} \leq \boldsymbol{\zeta} \leq \ \boldsymbol{u}\}$. In matrix form, we obtain: $\boldsymbol{A}\boldsymbol{\zeta} + \boldsymbol{b} \geq \mathbf{0}$, where $\boldsymbol{A} = \begin{pmatrix} \mathbf{I}_d \\ -\mathbf{I}_d \end{pmatrix} \in \mathbb{R}^{2d \times d}$ and $\boldsymbol{b} = \begin{pmatrix} -\boldsymbol{l} \\ \boldsymbol{u} \end{pmatrix} \in \mathbb{R}^{2d}$, with $\mathbf{I}_d$ the $d \times d$ identity matrix. Furthermore, we have*

$$\mathbf{1}_{\mathcal{C}_{lin}}(\boldsymbol{\zeta}) \approx \prod_{\kappa=1}^{2d} \frac{1}{1 + \exp(-\eta[\boldsymbol{A}\boldsymbol{\zeta} + \boldsymbol{b}]_\kappa)}, \quad \forall \boldsymbol{\zeta} \in \mathbb{R}^d,$$

where $[\boldsymbol{A}\boldsymbol{\zeta} + \boldsymbol{b}]_\kappa$ represents the κ^{th} component of the $2d$-dimensional vector $\boldsymbol{A}\boldsymbol{\zeta} + \boldsymbol{b}$ and $\eta > 0$ is a given approximate parameter. Then, according to (7), the logarithm of the associated likelihood function has the following form:

$$\log(L_\eta(\boldsymbol{\zeta})) = -\sum_{\kappa=1}^{m} \log\left[1 + e^{-\eta(\boldsymbol{\zeta} - \boldsymbol{l})} + e^{-\eta(\boldsymbol{u} - \boldsymbol{\zeta})} + e^{-\eta(\boldsymbol{u} - \boldsymbol{l})}\right], \tag{13}$$

which eliminates the necessity of performing matrix products and hence, reducing the computational complexity of evaluating the log-likelihood in (13) to $\mathcal{O}(d)$ instead of $\mathcal{O}(d^2)$.

The significant advantages of the results in Lemma 1 have been explored in high-dimensions (Table 1 below). It is worth noting that these results remain valid when the lower and upper bound vectors, $\boldsymbol{l}$ and $\boldsymbol{u}$, take on infinite values.

Proof of Lemma 1. The idea of the proof is based on the following: suppose that $\zeta \in \mathbb{R}$ is bounded between $l_b \in \mathbb{R}$ and $u_b \in \mathbb{R}$, i.e., $l_b \leq \zeta \leq u_b$. Then,

$$\begin{aligned} \mathbf{1}_{\{l_b \leq \zeta \leq u_b\}} &\approx \frac{1}{1 + \exp(-\eta[\zeta - l_b])} \times \frac{1}{1 + \exp(-\eta[u_b - \zeta])} \\ &= \frac{1}{1 + \exp(-\eta[\zeta - l_b]) + \exp(-\eta[u_b - \zeta]) + \exp(-\eta[u_b - l_b])}, \end{aligned}$$

for a given approximate parameter $\eta > 0$ and for any $\zeta \in \mathbb{R}$. The two real numbers l_b and u_b represent the lower and upper bounds, respectively. □

Figure 3 display contour plots of the pdf f in (1) of a zero-mean bivariate normal distribution truncated to the positive orthant (last column) along with those for $\tilde{f}_\eta$ for various values of η, with η increasing from left to right. Each row corresponds to a different value of the correlation ρ. Figure 4 shows the same contour plots in the situation where the mean of the prior bivariate normal distribution $\boldsymbol{\mu} = [-2, -1]^\top$ lies outside the positive orthant. It is clear that the approximation improves quickly as η increases, and $\eta = 50$ provides a *good* approximation and numerical stability, as we will show in the simulation studies of this paper. Furthermore, we develop a strategy to update the approximate parameter η at each MCMC iteration, ensuring the stability and convergence of the proposed algorithm, and addressing extreme cases. Additionally, we propose a projection strategy to ensure that the resulting posterior samples strictly respect the linear inequality constraints.

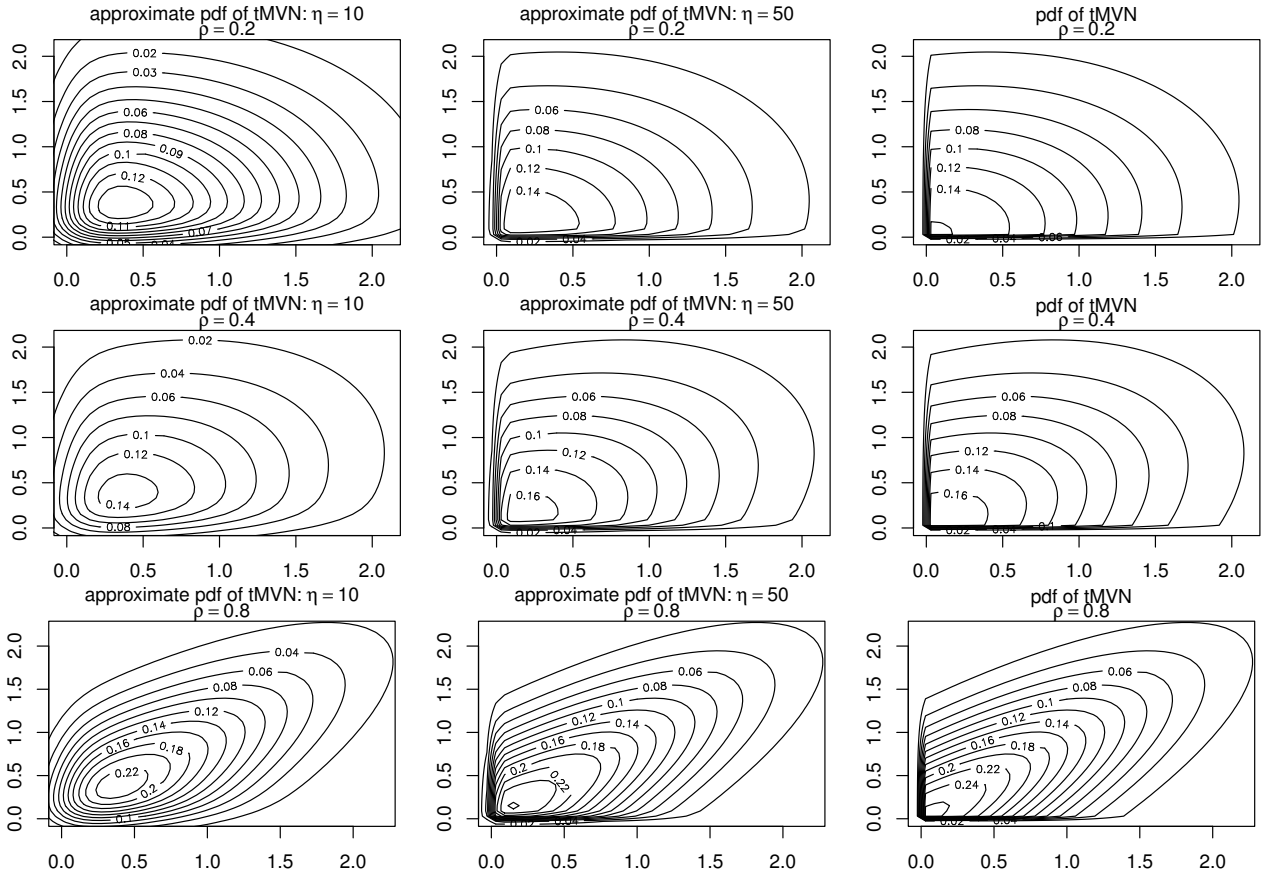

Figure 3: Contour plots of the pdf f in (1) of a zero-mean bivariate normal distribution truncated to the positive orthant (third column) and its approximate $\tilde{f}_\eta$ for $\eta = 10$ and $\eta = 50$ (first and second columns).

Remark 1 (Posterior Mode)**.** *In most real-world applications, the mean of the prior does not belong to the restricted domain (set of inequality constraints). As mentioned previously in this section, the starting point of the proposed MCMC approach does not necessitate the verification of inequality constraints. However, we recommend using the posterior mode, denoted as $\boldsymbol{\mu}^*$, as the initial point for the MCMC sampler, especially when the probability of the MVN distribution in the restricted domain is too low (i.e., the mean of the prior $\boldsymbol{\mu}$ lies outside the restricted domain). This is an interesting strategy since we start with the higher-probability point. In that case, the proposed MCMC sampler converges faster and avoids numerical instability. This is because when using the posterior mode as a starting point in the MCMC sampler, the first proposal $\boldsymbol{\xi}'$ satisfying the inequality $UL_\eta(\boldsymbol{\xi}) \leq L_\eta(\boldsymbol{\xi}')$ is closer to the restricted domain than when starting with the mean of the prior, $\boldsymbol{\mu}$. Hereafter, we describe how the posterior mode $\boldsymbol{\mu}^*$ can be computed.*

Let $\boldsymbol{\mu}^*$ denotes the posterior mode that maximizes the constrained pdf (1)*. Thus, maximizing* (1) *is equivalent to solving the quadratic optimization problem:*

$$\boldsymbol{\mu}^* := \underset{\boldsymbol{\zeta} \in \mathcal{C}}{\arg\max} \left\{ -\frac{1}{2} [\boldsymbol{\zeta} - \boldsymbol{\mu}]^\top \boldsymbol{\Gamma}^{-1} [\boldsymbol{\zeta} - \boldsymbol{\mu}] \right\}, \tag{14}$$

with $\mathcal{C}$ the set of inequality constraints (2)*. For instance, when $\mathcal{C}$ is the set of linear inequality constraints $\mathcal{C}_{lin}$ in* (2)*, the optimization problem in* (14) *becomes a quadratic optimization problem subject to linear inequality constraints [3, 19], known as quadratic programming (QP):*

$$\boldsymbol{\mu}^* := \underset{\boldsymbol{\zeta} \ s.t. \ \boldsymbol{A}\boldsymbol{\zeta} + \boldsymbol{b} \geq \boldsymbol{0}}{\arg\min} \left\{ \frac{1}{2} \boldsymbol{\zeta}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\zeta} - \boldsymbol{\mu}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\zeta} \right\}. \tag{15}$$

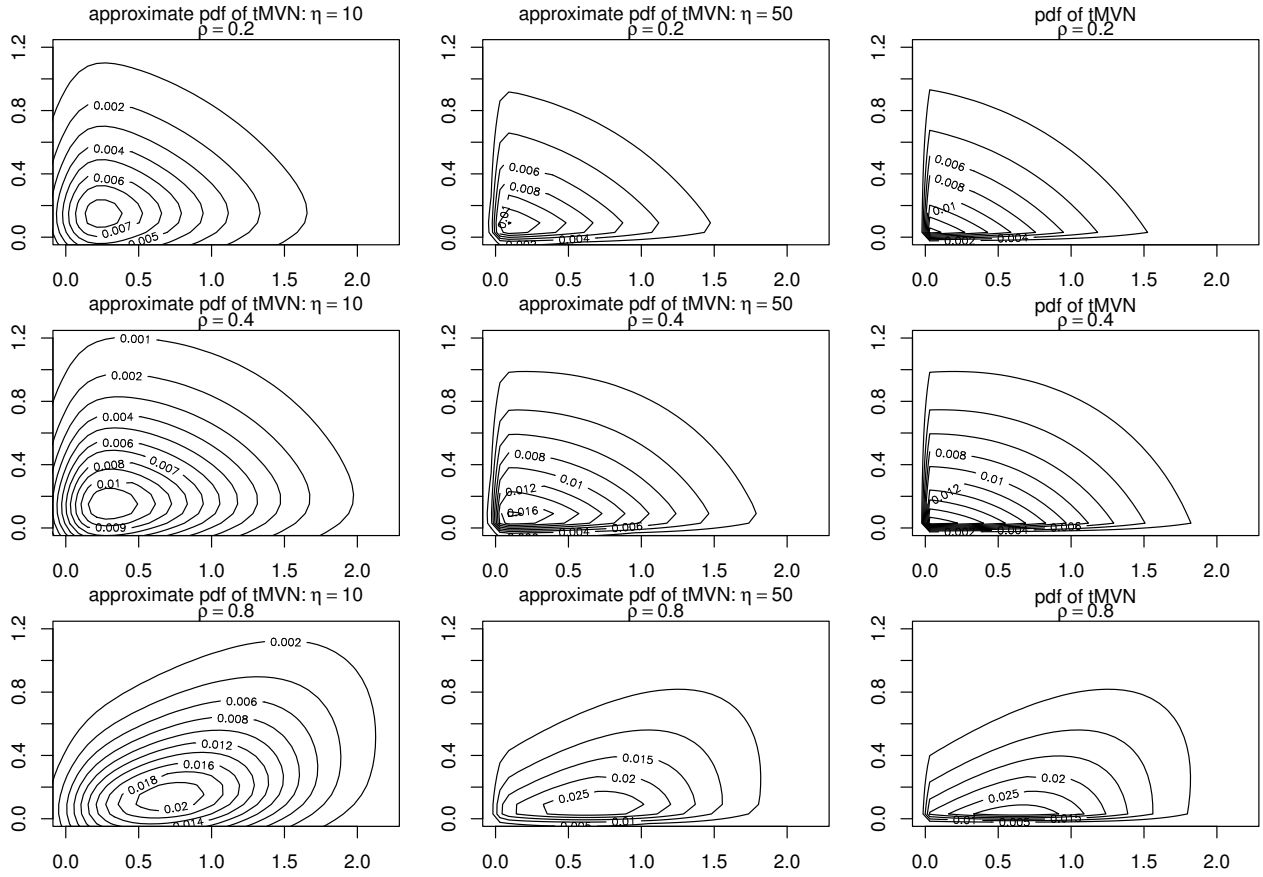

Figure 4: Same as Figure 3, with a mean of the prior $\boldsymbol{\mu} = [-2, -1]^\top$ lying outside the positive orthant.

In the numerical examples of this paper, the efficient R*-package* quadprog *has been used to compute the posterior mode* $\boldsymbol{\mu}^*$ *in* (15)*. Furthermore, since the proposed approach is based on an approximation, some of the generated samples may not strictly satisfy the linear inequality constraints. To address this, we propose projecting those samples onto the set of linear constraints at each MCMC iteration.*

Algorithm 1 outlines the scheme for generating n samples from the tMVN $\boldsymbol{\zeta} \sim \mathcal{TN}_d(\mathbf{0}, \boldsymbol{\Gamma}; \mathcal{C}_{\text{lin}})$, where $\mathcal{C}_{\text{lin}} = \{\boldsymbol{\zeta} \in \mathbb{R}^d \ : \ \boldsymbol{A}\boldsymbol{\zeta} + \boldsymbol{b} \geq \mathbf{0}\}$, using the proposed MCMC approach.

3 Sampling challenges for tMVN distributions

The aim of this section is to highlight the challenges associated with sampling from tMVN distributions and to propose an improved solution. Two cases are considered. The first is referred to as the 'extreme' case, in which the mean of the prior is located far outside the restricted domain. This results in a very low probability for the MVN distribution within the restricted domain. The second is called the 'normal' case, where the mean of the prior distribution is near or within the restricted domain.

3.1 Low or high value of the approximate parameter η

As mentioned previously, the approximate parameter η, which appears in Equation (3), plays a crucial role in the proposed approach, as it ensures the stability of the developed algorithm, the convergence of the posterior distribution, and addresses extreme cases. As proved in Proposition 2, the L_1 distance between the resulting approximated posterior pdf $\tilde{f}_\eta$ and the target posterior pdf

Algorithm 1: Sampling scheme of the tMVN $\boldsymbol{\zeta} \sim \mathcal{TN}_d(\mathbf{0}, \boldsymbol{\Gamma}; \mathcal{C}_{\text{lin}})$, where $\mathcal{C}_{\text{lin}} = \{\boldsymbol{\zeta} \in \mathbb{R}^d : \boldsymbol{A}\boldsymbol{\zeta} + \boldsymbol{b} \geq \mathbf{0}\}$ using the proposed MCMC approach for n samples.

Input: current state $\boldsymbol{\zeta} = \boldsymbol{\mu}^*$ in (15) for $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Gamma} \in \mathbb{R}^{d\times d}$, $\boldsymbol{A} \in \mathbb{R}^{m\times d}$, $\boldsymbol{b} \in \mathbb{R}^m$, $\eta \in \mathbb{R}_+^*$, and $n \in \mathbb{N}^*$ MCMC iterations.

- For i from 1 to n, **do**
  1. Generate $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$.
  2. Compute the new state $\boldsymbol{\zeta}'_{(i)}$ in (11) using the ESS in [31]:
     - Compute the log-likelihood functions:

$$\log(L(\boldsymbol{\zeta})) = -\sum_{\kappa=1}^{m} \log\left(1 + \exp\{-\eta[\boldsymbol{A}\boldsymbol{\zeta} + \boldsymbol{b}]_\kappa\}\right)$$
$$\log y = \log(L(\boldsymbol{\zeta})) + \log u, \quad u \sim \mathcal{U}([0,1]).$$

     - Define an initial bracket $[\theta_{\min}, \theta_{\max}]$:

$$\theta \sim \mathcal{U}([0, 2\pi]);$$
$$\theta_{\min} = \theta - 2\pi \quad \text{and} \quad \theta_{\max} = \theta.$$

     - Compute the proposal

$$\boldsymbol{\zeta}'_{(i)} = \sin(\theta)\boldsymbol{\nu} + \cos(\theta)\boldsymbol{\zeta}.$$

     - **While** $\log(L(\boldsymbol{\zeta}'_{(i)})) < \log y$;
       - shrink the bracket $[\theta_{\min}, \theta_{\max}]$ as in [31] and generate $\theta \sim \mathcal{U}([\theta_{\min}, \theta_{\max}])$;
       - compute:

$$\boldsymbol{\zeta}'_{(i)} = \sin(\theta)\boldsymbol{\nu} + \cos(\theta)\boldsymbol{\zeta}.$$

     if $[\boldsymbol{A}\boldsymbol{\zeta}'_{(i)} + \boldsymbol{b}]_\kappa < 0$, for any $\kappa \in \{1, \ldots, m\}$ **then**
     project $\boldsymbol{\zeta}'_{(i)}$ on $\mathcal{C}_{\text{lin}}$.
     else
     accept $\boldsymbol{\zeta}'_{(i)}$.
     end if
  3. Reinitialization: $\boldsymbol{\zeta} = \boldsymbol{\zeta}'_{(i)}$.
  4. Update η.
- **EndFor**

Output: return $\left[\boldsymbol{\zeta}'_{(1)}, \ldots, \boldsymbol{\zeta}'_{(n)}\right] \in \mathbb{R}^{d\times n}$.

f converges to zero as η goes to infinity.

On one hand, a *low* value of the approximate parameter η involves two problems:

- First, a *low* value of η involves a poor approximation of the indicator function, hence, a poor approximation of the likelihood.
- Second, a *low* value of η results in a poor approximation of the posterior truncated distribution, as shown in Figures 3, 4, and 5 (left panel), and leads to poor convergence of the proposed MCMC algorithm (left panel in Figure 13).

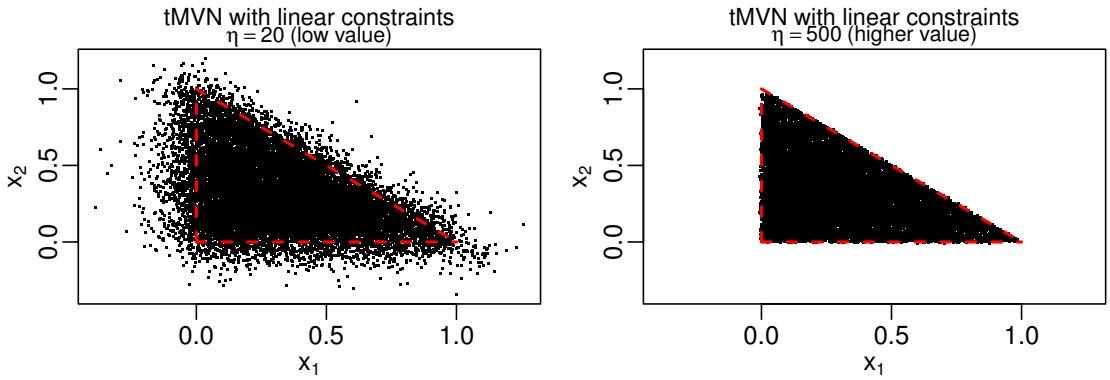

Figure 5: *Low* & *high* η (normal case): 20,000 samples (black dots) from $\mathcal{TN}_2(\boldsymbol{\mu}, \boldsymbol{\Gamma}; \mathcal{C})$ using the proposed approach, where $\mathcal{C} = \mathcal{C}_{\text{lin}}$ (red dashed triangle). The first 5,000 samples are discarded as burn-in. The mean of the prior is $\boldsymbol{\mu} = [0, 0]^\top$, and the correlation is $\rho = 0.4$. A *low* value of the approximate parameter η is employed in the left panel, while a *high* value is employed in the right.

In Figure 5, the effect of the approximate parameter, η, on the posterior truncated distribution is highlighted. The aim in this example is to generate a zero-mean bivariate normal distribution, $\boldsymbol{\zeta}$, with covariance matrix $\boldsymbol{\Gamma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where the correlation ρ is fixed at 0.4, and restrict it to a triangle with vertices (0,0), (1,0) and (0,1) (red dashed triangle). The black dots represent 20,000 samples generated using the proposed approach, with η fixed at 20 in the left panel and 500 in the right panel. The first 5,000 samples are discarded as burn-in. As expected, we observe that a low value of the approximate parameter, η, results in a poor approximation of the posterior truncated distribution. It is worth noting that some samples are located outside the triangle even when η is very high. This is because the projection strategy into the restricted domain at each MCMC iteration (see the second step in Algorithm 1) is not employed in this example. Finally, let us mention that this numerical example addresses a 'normal' case, and a high value of the approximate parameter η has no impact on the stability of the proposed algorithm (see the first point below for *high* value of η).

On the other hand, when η is *high*, two problems occur:

- First, in the tMVN distribution framework, a *high* value of η results in instability in the proposed algorithm for extreme cases, as the log-likelihood function approaches infinity. The extreme case occurs when the mean of the prior is located far outside the restricted domain (right panel in Figure 6).
- Second, in Bayesian shape-restricted function estimation, a *high* value of η accentuates the *mass-shifting* phenomenon quantified in [47] (see Sections 3.3 and 5.2 below), as tMVN priors lead to biased posterior inference when the underlying function contains flat regions.

In Figure 6, the same settings as in Figure 5 are considered, except that the mean of the prior $\boldsymbol{\mu}$ is fixed at $[-6, -2]^\top$. This is referred to as an 'extreme' case. As expected, due to the divergence of the logarithm of the likelihood function in (12), a high value of the approximate parameter η (right panel) leads to instability in the proposed MCMC algorithm, causing the samples to fail to respect the constraints. It is worth noting that a low value of η (left panel) effectively ensures the stability and convergence of the proposed algorithm; however, it results in a poor approximation of the posterior distribution.

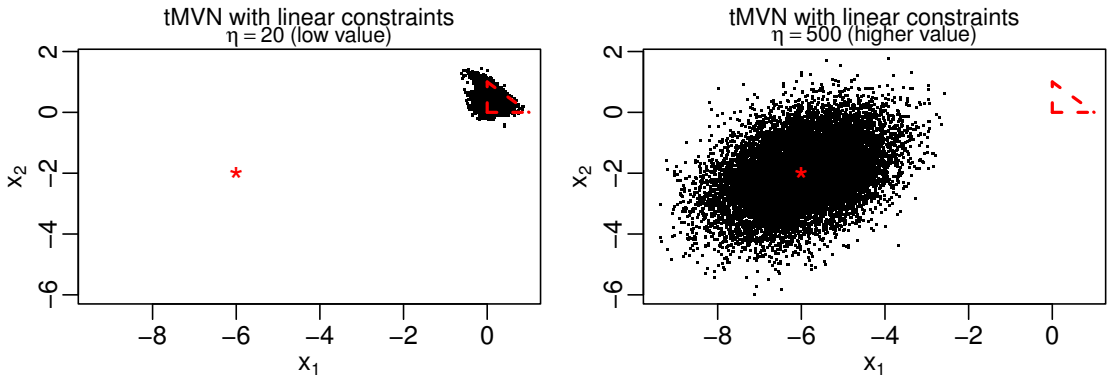

Figure 6: *Low & high* η (extreme case): the caption is the same as in Figure 5, except that the mean of the prior is fixed at $\boldsymbol{\mu} = [-6, -2]^\top$ (red cross).

3.2 Updating η at each MCMC iteration

On one hand, a *high* value of the approximate parameter η improves the precision of the proposed approach in approximating the target posterior distribution. However, it accentuates the *mass-shifting* phenomenon (see Sections 3.3 and Section 5) and fails to ensure the convergence in extreme cases. On the other hand, a *low* value of η in the tMVN distribution avoids numerical instability and addresses extreme cases, such as when the mean of the prior distribution lies outside the acceptance region. However, it produces a poor approximation of the target truncated posterior distribution (left panel of Figure 5). As a result, we develop a strategy that updates the approximate parameter η at each MCMC iteration in order to take advantage of both *low* and *high* values: numerical stability when the parameter is *low*, and precision when the parameter is *high*. The main idea is to start the MCMC algorithm with a *low* value and increase it at each iteration to ensure convergence to the target posterior distribution, as the initial iterations can be discarded as burn-in. For example, the approximate parameter η can be multiplied by $1 + \epsilon$ at each MCMC iteration (see the right panel of Figure 13, where ϵ is fixed at 0.01%). This leads to a robust approach that handles extreme cases while ensuring the convergence of the proposed posterior distribution (see Figures 13 and 14). This strategy also works well and is coherent with projecting the samples into the linear restricted domain. The numerical efficiency of the updating strategy is shown in Section 4.2.

3.3 *Mass-shifting* phenomenon of tMVN

In this section, we explore the *mass-shifting* phenomenon of the tMVN distribution highlighted by [47]. In fact, the *mass-shifting* phenomenon appears in the low-dimensional marginal densities of a dependent zero-mean multivariate normal distribution restricted to the positive orthant. The marginal density of a tMVN distribution having the mode at the origin exhibits small mass near the origin as the dimension increases. Furthermore, this phenomenon becomes more pronounced with stronger correlations between the random variables. This unexpected behavior has significant implications for Bayesian shape-restricted function estimation. As empirically observed by [11, 26, 34], tMVN priors lead to biased posterior inference when the underlying function contains flat regions. Before presenting the *mass-shifting* phenomenon, let us recall the following well-known result: suppose that X_1 and X_2 are two normal random variables with means μ_1 and μ_2, respectively, and variances $\Sigma_{11} = \text{Var}(X_1)$ and $\Sigma_{22} = \text{Var}(X_2)$, respectively. Then,

$$\{X_1|X_2\} \sim \mathcal{N}(\mu, \Sigma), \quad \text{where,}$$
$$\begin{cases} \mu = \mu_1 + \Sigma_{11}\Sigma_{12}^{-1}(X_2 - \mu_2); \\ \Sigma = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}; \end{cases}$$

with $\Sigma_{12} = \Sigma_{21} = \text{Cov}(X_1, X_2)$. Now, we consider a zero-mean bivariate Gaussian vector $X = [X_1, X_2]^\top$ truncated within the positive orthant. Suppose that X admits a covariance matrix $\boldsymbol{\Gamma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $\rho \in [-1, 1]$ is the correlation parameter. This means that $\text{Var}(X_1) = \text{Var}(X_2) = 1$ and $\text{Cov}(X_1, X_2) = \text{Cor}(X_1, X_2) = \rho$.

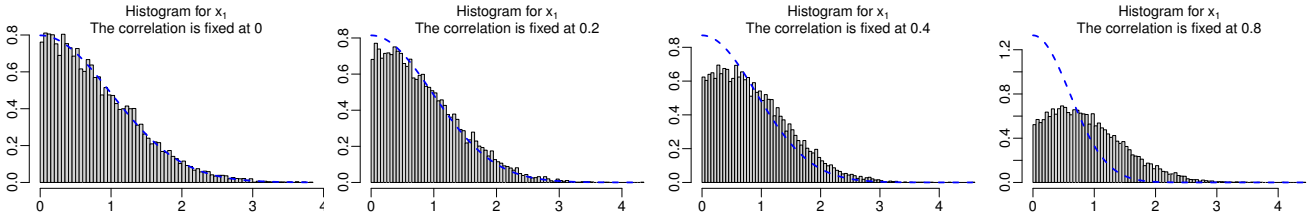

Figure 7: Histogram of the marginal distribution of a two-dimensional zero-mean TMVN restricted to the positive orthant at different correlation. The approximate parameter η is updated at each MCMC iteration. The blue dashed curve represents the pdf of the true truncated marginal distribution.

In Figure 7, we illustrate the histograms of the marginal distribution of the two-dimensional zero-mean Gaussian vector $X = [X_1, X_2]^\top$ restricted to the positive orthant for different values of the correlation parameter ρ. As expected, a *mass-shifting* phenomenon as highlighted by [47] of the marginal distribution of a zero-mean tMVN distribution is observed. The histograms of the first variable x_1 are based on 20,000 MCMC iterations of the proposed approach developed in the present paper, where the first 5,000 as discarded as burn-in. The approximate parameter η is updated at each MCMC iteration, which increases fast to infinity. The blue dashed curve represents the pdf of the true truncated marginal distribution. One can observe that for high correlations, the histograms shift to the right compared to the true truncated marginal distribution, a phenomenon known as *mass-shifting*. This effect becomes more pronounced as correlations increase. The impact of the *mass-shifting* phenomenon is highlighted in Section 5 for shape-restricted function estimation, when the unknown target function admits a flat region. In this context, this phenomenon involves a biased Bayesian estimator.

4 Numerical illustrations

The aim of this section is to provide performance illustrations of the approach developed in the present paper (Section 2.1) for a combination of linear, quadratic, and nonlinear inequality constraints, both when η is fixed and when it is updated. In all the numerical examples in this section, except where mentioned otherwise, $\boldsymbol{\zeta}$ is a bivariate Gaussian vector with mean vector $\boldsymbol{\mu} = [\mu_1, \mu_2]^\top$ and covariance matrix $\boldsymbol{\Gamma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $\rho \in [-1, 1]$ is the correlation parameter.

4.1 Illustrations with fixed η

In this section, the parameter η in (7) is fixed at 50, providing a *good* approximation (Figures 3 and 4) of the posterior pdf (1) and avoiding numerical instability, especially when the logarithm of the likelihood function $L_\eta(\cdot)$ tends to infinity. This is referred to as the 'normal' case, which occurs in most applications. It is worth noting that in this case, a fixed large value of η is enough to ensure stability and convergence of the proposed MCMC approach.

Example 1 (Nonlinear inequality constraints)**.** *In this example, the set of inequality constraints* $\mathcal{C}$ *is the non-convex set* $\mathcal{C}_{nlin} = \{\boldsymbol{x} \in \mathbb{R}^2 \ : \ g_\kappa(\boldsymbol{x}) \geq 0, \ \kappa = 1, 2\}$*, with* $g_1(\boldsymbol{x}) = \cos(x_1) + x_2 - 1$

and $g_2(\boldsymbol{x}) = -\cos(x_1) - x_2 + 2$ *two nonlinear continuous functions, and* $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$. *The aim is to generate a bivariate normal distribution* $\boldsymbol{\zeta} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ *restricted to* $\mathcal{C}_{nlin}$, *where* $d = 2$, *i.e.,* $\boldsymbol{\zeta} \sim \mathcal{TN}_2(\boldsymbol{\mu}, \boldsymbol{\Gamma}; \mathcal{C}_{nlin})$. *This is the so-called tMVN distribution subject to nonlinear inequality constraints.*

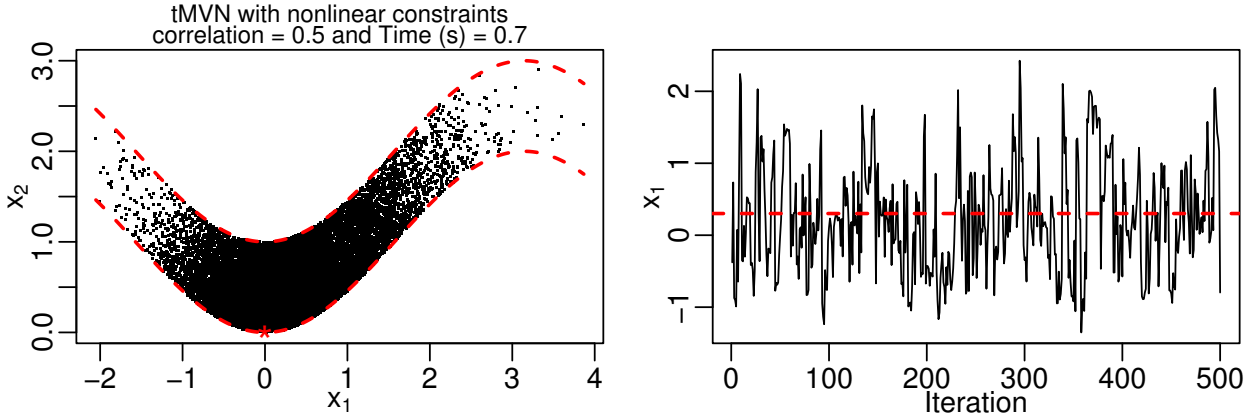

Figure 8: 20,000 samples (black dots) from $\mathcal{TN}_2(\boldsymbol{\mu}, \boldsymbol{\Gamma}; \mathcal{C})$ using the proposed approach, where $\mathcal{C} = \mathcal{C}_{\text{nlin}}$. The mean of the prior $\boldsymbol{\mu} = [0, 0]^\top$ (red star) and the correlation is $\rho = 0.5$. The two dashed curves form the set of nonlinear inequality constraints $\mathcal{C}$. The right panel displays the values of x_1 for the first 500 MCMC iterations.

Figure 8 shows a performance illustration of the proposed approach for handling nonlinear inequality constriants. The black dots in the left panel represent the 20,000 samples from a two-dimensional normal distribution $\mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ *constrained by the non-convex set* $\mathcal{C} = \mathcal{C}_{nlin}$ *formed by the two dashed curves (lower and upper bound functions, respectively). The mean of the prior is* $\boldsymbol{\mu} = [0, 0]^\top$ *indicated by the red star, and the correlation parameter* ρ *is fixed at* 0.5. *The computational running time of generating 20,000 MCMC samples is displayed in main of the left panel. The right panel displays the values of the first variable* x_1 *for the first 500 MCMC iterations which oscillate rapidly around the posterior mean (red dashed line), as desired.*

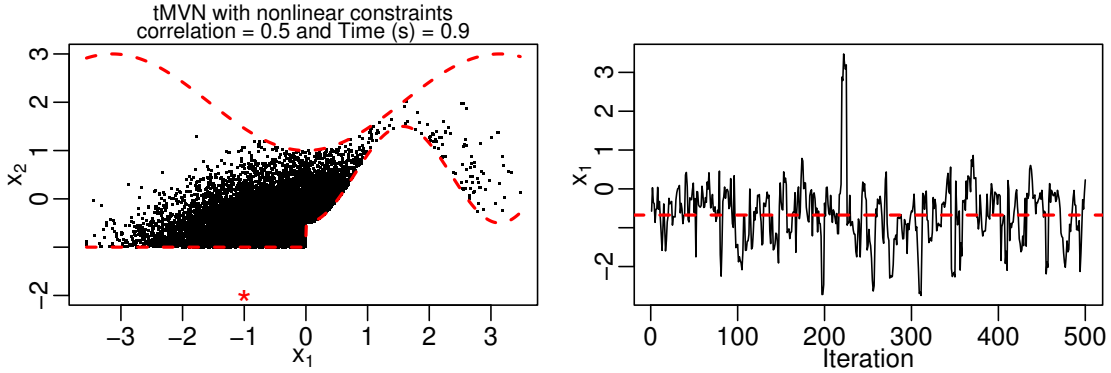

Figure 9: Same as Figure 8 for a different lower piecewise continuous function. The mean of the prior $\boldsymbol{\mu}$ is fixed at $[-1, -2]^\top$.

Similar to Figure 8, the left panel of Figure 9 displays 20,000 samples (black dots) from a two-dimensional normal distribution constrained by nonlinear inequality constraints. The correlation and the computational running time are displayed in the main panel. The mean of the prior is fixed at $\boldsymbol{\mu} = [-1, -2]^\top$ *and is represented by the red star. The first function* $g_1(\cdot)$ *of the nonlinear inequality constraints is defined as follows:*

$$g_1(\boldsymbol{x}) = p(x_1) + x_2, \quad \boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2,$$

where $p(\cdot)$ is the piecewise nonlinear continuous function defined as $\cos(2x_1) - 0.5$ if $x_1 > 0$ and 1 otherwise (the lower dashed curve). In the right panel, we illustrate the values of x_1 for the first 500 MCMC iterations, which rapidly oscillate around the posterior mean indicated by the red dashed line.

Example 2 (Linear and quadratic constraints)**.** *In this example, the set of inequality constraints $\mathcal{C}$ is the intersection between linear and quadratic inequality constraints, i.e., $\mathcal{C} = \mathcal{C}_{lin} \cap \mathcal{C}_{quad}$, where*

$$\mathcal{C}_{lin} = \{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1 + x_2 \geq 0 \text{ and } x_2 \geq 0\}; \tag{16}$$

$$\mathcal{C}_{quad} = \{\boldsymbol{x} \in \mathbb{R}^2 \ : \ \boldsymbol{x}^\top \boldsymbol{C}_\kappa \boldsymbol{x} + \boldsymbol{d}_\kappa^\top \boldsymbol{x} + e_\kappa \geq 0, \ \kappa = 1, 2\}; \tag{17}$$

where $\boldsymbol{C}_1 = \begin{pmatrix} -1/8 & 0 \\ 0 & -1/2 \end{pmatrix}$, $\boldsymbol{C}_2 = \begin{pmatrix} 4 & -1 \\ -1 & 8 \end{pmatrix}$, $\boldsymbol{d}_1 = [0.5, 0.5]^\top$, $\boldsymbol{d}_2 = [0, 5]^\top$, $e_1 = 0.75$, and $e_2 = -1$. Let us mention that the linear inequality constraints in (16) can be expression as $\boldsymbol{A}\boldsymbol{\zeta} + \boldsymbol{b} \geq \boldsymbol{0}$, where

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This numerical example involves two linear and two quadratic inequality constraints.

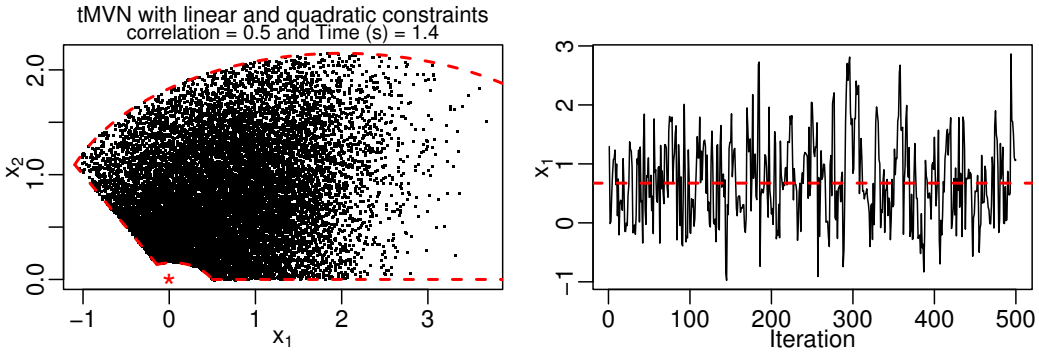

Figure 10: Left: 20,000 samples (black dots) from a zero-mean bivariate normal distribution restricted to linear and quadratic inequality constraints (16)-(17). The correlation parameter ρ is fixed at 0.5. Right: the first 500 values of x_1 oscillating around the posterior mean (red horizontal dashed line), as desired.

The left panel of Figure 10 displays 20,000 samples (black dots) from the zero-mean two-dimensional normal distribution $\mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ restricted to the linear and quadratic inequality constraints given in (16) and (17). The correlation and the computational running time of generating 20,000 MCMC samples are displayed in the main panel. The red star represents the mean of the prior $\boldsymbol{\mu} = [0, 0]^\top$. In the right panel, we illustrate the first 500 values of x_1 which oscillate rapidly around the posterior mean (represented by the red dashed line), as desired.

Example 3 (Extreme case)**.** *The aim of this example is to highlight the efficiency of the proposed approach through an extreme case example that is impossible for other MCMC samplers. To do this, we fix the mean vector $\boldsymbol{\mu}$ of the bivariate normal distribution at $[5, 13]^\top$ and the correlation parameter ρ at 0.5. The set of linear constraints is $\mathcal{C}_{lin} = \{\boldsymbol{x} \in \mathbb{R}^2 \ : \ 10 \leq x_1 \leq 13$ and $8 \leq x_2 \leq 11\}$ (dashed square in Figure 11). This is a difficult situation because the probability of the MVN distribution restricted to the rectangle $\mathcal{C}_{lin}$ is too low.*

In Figure 11, we generate 20,000 samples (black dots) from a bivariate normal distribution with mean $\boldsymbol{\mu} = [5, 13]^\top$ (red star in the left panel) restricted by linear constraints represented by the dashed square. The first 5,000 samples are discarded as burn-in. The correlation between the variables x_1 and x_2 is fixed at 0.5. The approximate parameter η is fixed at 50. The right panel is a zoom-in of the left one within the restricted domain. This represents a very difficult situation (an

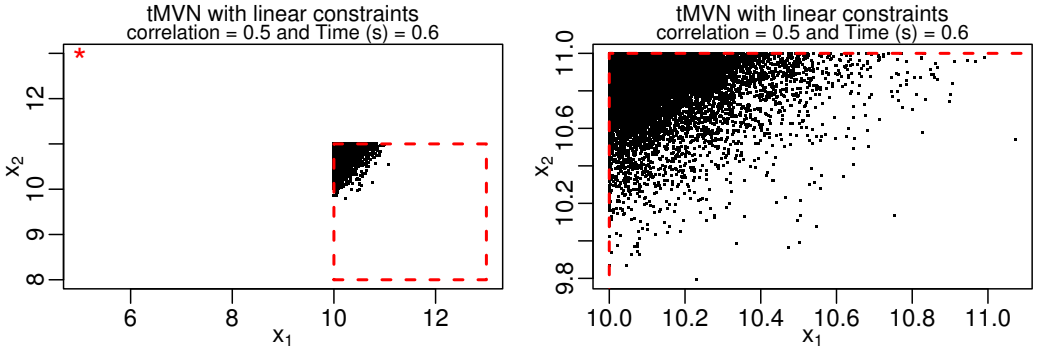

Figure 11: 20,000 samples (black dots) from a bivariate normal distribution with mean $\boldsymbol{\mu} = [5, 13]^\top$ restricted to linear inequality constraints (dashed square). The first 5,000 samples are discarded as burn-in. The correlation parameter ρ is fixed at 0.5. The right panel is a zoom-in of the left panel within the acceptance region.

extreme case) because the mean of the prior $\boldsymbol{\zeta}$ does not belong and is far from this restricted domain. Using the proposed approach, generating 20,000 samples takes 0.6 seconds. It is worth noting that the HMC sampler fails to generate samples in this extreme case.

Example 4 (High-dimensions)**.** *In this example, we investigate the effectiveness of the proposed approach in high-dimensions. To do this, we consider a Gaussian process (GP) denoted by $(Y(x))_{x\in\mathcal{D}}$ with mean function $m(\cdot)$ and covariance function $k(\cdot,\cdot)$, i.e., $Y \sim \mathcal{GP}(m(\cdot), k(\cdot,\cdot))$. Without loss of generality, we assume the domain $\mathcal{D}$ is the unit interval $[0, 1]$. Let $\{t_j\}$, $j = 1, \ldots, d$ be a uniform discretization of $\mathcal{D}$. Then, $\boldsymbol{\zeta} = [Y(t_1), \ldots, Y(t_d)]^\top \in \mathbb{R}^d$ is a Gaussian vector with mean vector $\boldsymbol{\mu} = [m(t_1), \ldots, m(t_d)]^\top$ and covariance matrix $\boldsymbol{\Gamma} = (k(t_j, t_l))_{1\le j,l\le d}$. Let*

$$\mathcal{C}_{lin} = \left\{\boldsymbol{x} \in \mathbb{R}^d \ : \ 0 \le x_i \le 10, \ i = 1, \ldots, d\right\} = [0, 10]^d$$

be a set of linear inequality constraints, where d is fixed at 1,000.

Figure 12 displays the first two coordinates and the first and the last coordinates (top and bottom panels, respectively) of 20,000 samples generated from a tMVN distribution with mean $\boldsymbol{\mu} = [-5, \ldots, -5]^\top \in \mathbb{R}^d$. The covariance matrix $\boldsymbol{\Gamma}$ is extracted from a Matérn covariance function:

$$k(x, x') = \tau^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{\ell}|x - x'|\right)^{\nu} B_\nu \left(\frac{\sqrt{2\nu}}{\ell}|x - x'|\right), \quad x, x' \in \mathcal{D}, \tag{18}$$

where $\Gamma(\cdot)$ is the Gamma function and $B_\nu(\cdot)$ denotes the modified Bessel function of the second kind of order ν [39]. The smoothness parameter ν regulates the degree of smoothness of the GP sample paths/functions. Let us recall that a process with the Matérn kernel of order ν is $\lceil \nu - 1 \rceil$ times differentiable (see Section 4.2.1 in [44]). The parameters τ^2 and ℓ, which are positive, are commonly known as the signal variance and correlation length-scale, respectively. The smoothness parameter ν and the length-scale parameter ℓ in this example are fixed at 1.5 and 0.4, respectively. The first 5,000 samples are discarded as burn-in. The red star in the left panels represents the mean of the prior and the dashed square represents the restricted domain. The dimension of the Gaussian vector d is fixed at 1,000. Using the proposed approach and Lemma 1, generating 20,000 samples with dimension $d = 1,000$ takes approximately 11 seconds. In the top panels, the correlation between the first two coordinates is close to 1 (i.e., $k(t_1, t_2) \approx 1$), while in the bottom panels, the correlation $k(t_1, t_{1000}) = 0.07$ is close to 0.

Table 1 shows the computational running time in seconds for generating 20,000 samples from a tMVN distribution as a function of the dimension d. The proposed ESS sampler is faster than the HMC sampler for both low and high-dimensions. It is worth noting that the Gibbs sampling approach from the `R`*-package restrictedMVN suffers from the converge for $d \in \{100, 250, 500, 1000\}$.*

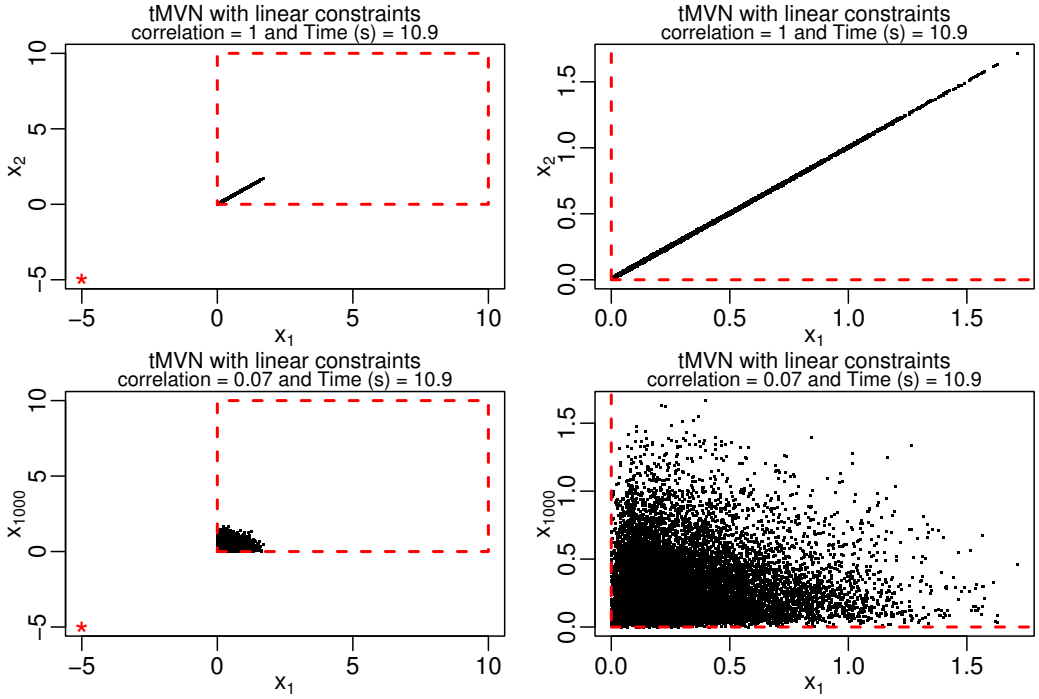

Figure 12: 20,000 samples of a MVN distribution $\boldsymbol{\zeta} \in \mathbb{R}^d$ restricted to $[0, 10]^d$, where d is fixed at 1,000 and the first 5,000 samples are discarded as burn-in. The mean of the prior is $\boldsymbol{\mu} = [-5, \ldots, -5]^\top$. The right panels are the zoom-in of the left panels. The red star represents the mean of the prior and the dashed square represents the restricted domain. The top panels illustrate the first two coordinates, where the correlation $k(t_1, t_2)$ is close to 1. The bottom panels display the first and the last coordinates, where the correlation $k(t_1, t_d) = 0.07$ is close to 0.

	Computational running time (s)				
	$d = 50$	$d = 100$	$d = 250$	$d = 500$	$d = 1000$
ESS sampler	0.90	1.17	2.09	4.50	11.10
HMC sampler	1.51	3.82	12.90	61.60	210.93

Table 1: The computational running time in seconds for generating 20,000 MCMC samples as a function of the dimension d for the two competing approaches, ESS and HMC.

4.2 Illustrations with updated η

The aim of this section is to investigate the case when the approximate parameter η is updated at each MCMC iteration.

We consider an extreme case where the mean of the prior, $\boldsymbol{\mu}$, is fixed at $[-31, -10]^\top$, which is located far from the restricted domain $\mathcal{C}_{\text{lin}} = \{\boldsymbol{x} \in \mathbb{R}^2 \ : \ 10 \leq x_1 \leq 13 \text{ and } 8 \leq x_2 \leq 11\}$, considered in Example 3. We fix the correlation parameter at $\rho = 0.5$. Two cases are considered: the first is when the approximate parameter η is fixed at 20 and the second is when the approximate parameter η is updated at each MCMC iteration. It is worth noting that the parameter η is fixed at 20 to avoid numerical instability. Larger values provide instability, and the log-likelihood goes to $-\infty$ in (12).

Figure 13 shows the evolution of the proposed ESS sampler in an extreme case with two different strategies (fixed and updated η). The aim is to generate a bivariate normal distribution with mean $\boldsymbol{\mu} = [-31, -10]^\top$ restricted to a linear set $\mathcal{C}_{\text{lin}} = \{\boldsymbol{x} \in \mathbb{R}^2 \ : \ 10 \leq x_1 \leq 13 \text{ and } 8 \leq x_2 \leq 11\}$ (dashed square in Figure 11). We generate 15,000 samples (black dots) using the proposed approach without discarding any samples. In the left panel, the approximate parameter η is fixed at 20 to avoid numerical instability. In the right panel, η is updated at each MCMC iteration, starting at

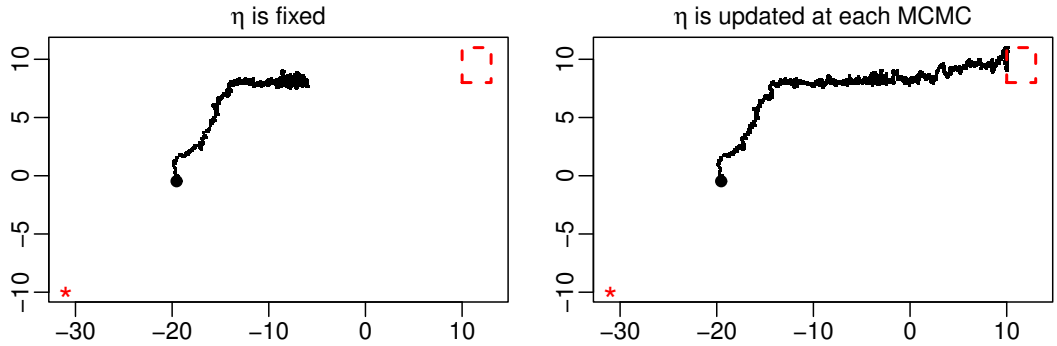

Figure 13: Evolution of convergence for fixed (left) and updated (right) η: 15,000 samples (black dots) from a bivariate normal distribution with mean $\boldsymbol{\mu} = [-31, -10]^\top$ (red star) restricted to linear inequality constraints (dashed square) without any discarded samples. The black ball represents the first accepted sample. The correlation parameter ρ is fixed at 0.5. The proposed approach has been employed with a fixed η at 20 in the left panel and updated η at each MCMC iteration in the right panel.

20 and increasing by 0.01%, so that it converges to infinity. The red star represents the mean of the prior and the black ball represents the first accepted proposal. We observe that when η is fixed, the proposed MCMC sampler fails to reach the restricted domain. However, when η is updated, the proposed MCMC successfully reaches the restricted domain. This occurs because, with the updating strategy, the parameter η can start with low values and gradually converge to infinity. This ensures the convergence and stability of the proposed ESS approach, effectively addressing extreme cases. It is worth noting that in both cases, the proposal sample is not projected into the restricted domain at each MCMC iteration, in order to display the scheme of the proposed MCMC sampler. Finally, let us recall that this is an extreme case, and it is evident that the HMC sampler and other existing samplers, such as those cited in the introduction, fail to generate samples from the corresponding posterior distribution.

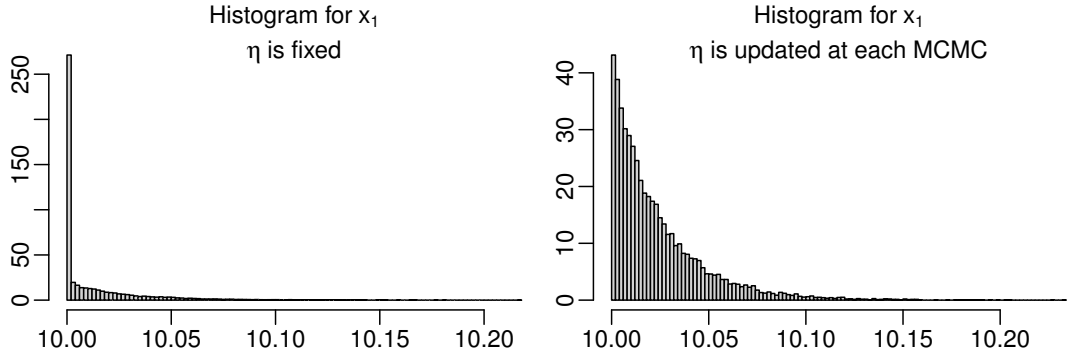

Figure 14: Histogram of the proposed posterior distribution for the first variable x_1 when the projection strategy is applied at each MCMC iteration. The approximate parameter η is fixed at 20 in the left panel and updated at each MCMC iteration in the right panel.

Figure 14 shows the histogram of the proposed posterior distribution based on 20,000 samples, where the first 5,000 samples are discarded as burn-in. The projection into the restricted domain is applied at each MCMC iteration. In the left panel, the approximate parameter η is fixed at 20, while in the right panel, it is updated at each MCMC iteration, as shown in the right panel of Figure 13. Unlike the case when η is updated, we observe a mass of probability at the lower bound of the restricted domain, $x_1 = 10$, when η is fixed at 20. This is due to the low value of the approximate parameter η and the projection phenomenon employed at each MCMC iteration to obtain samples that verify the linear constraints. It is worth noting that this mass of probability

at $x_1 = 10$ is interesting in situations such as approximating a monotone function with a flat region within the context of Bayesian shape restricted function estimation, as well highlighted in [47].

4.3 Measuring independence of samples

MCMC techniques remain the gold standard for approximate Bayesian inference, but they are only approximate methods, which may suffer from several problems such as convergence and correlation among samples [10]. In this section, a comparison between the efficient HMC and the proposed approach based on the ESS, developed in Section 2.1, is investigated. The HMC sampler [35] has been very successful in recent years due to its efficiency in simulating high-dimensional tMVN distributions. As HMC is limited to linear and quadratic constraints, only linear inequality constraints are considered in this section. The efficiency of the two competing approaches, HMC and ESS, can be quantified via the effective sample factor (ESF) [22]. If we denote $\boldsymbol{\zeta}^{(j)}$ the j^{th} sample, then the variance a function $f(\boldsymbol{\zeta})$ using n_s samples is

$$\text{Var}\left(\frac{f(\boldsymbol{\zeta}^{(1)}) + \ldots + f(\boldsymbol{\zeta}^{(n_s)})}{n_s}\right) = \frac{\text{Var}(f(\boldsymbol{\zeta}))}{n_s}\left[1 + \sum_{j=1}^{n_s-1}\left(1 - \frac{j}{n_s}\right)\rho_j\right], \tag{19}$$

where ρ_j is the autocorrelation function (ACF) with lag j and n_s is the number of samples. The ESF is defined as follows:

$$\text{ESF} = \left[1 + \sum_{j=1}^{n_s-1}\left(1 - \frac{j}{n_s}\right)\rho_j\right]^{-1}.$$

A sampling scheme is considered more efficient when it has a higher ESF, as this results in lower variance in (19). The ESF indicator depends on several parameters such as the position of the mean of the prior $\boldsymbol{\mu}$, the probability of the MVN distribution in the restricted domain, the set of inequality constraints $\mathcal{C}$, and the covariance structure $\boldsymbol{\Gamma}$ of the prior $\boldsymbol{\zeta}$.

As in Section 4, we consider here a bivariate Gaussian vector $\boldsymbol{\zeta}$ with mean vector $\boldsymbol{\mu} = [-2, 1]^\top$ and covariance matrix $\boldsymbol{\Gamma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $\rho \in [-1, 1]$ is the correlation parameter. Let $\mathcal{C} = \mathcal{C}_{\text{lin}} = \{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_1 \geq 0, \ x_2 \geq 0\}$, the positive orthant. The aim is to investigate the performance of the two competing approaches, ESS and HMC, when generating $\boldsymbol{\zeta}$ restricted to $\mathcal{C}$, i.e., $\boldsymbol{\zeta} \sim \mathcal{TN}_2(\boldsymbol{\mu}, \boldsymbol{\Gamma}; \mathcal{C})$ using the ESF indicator.

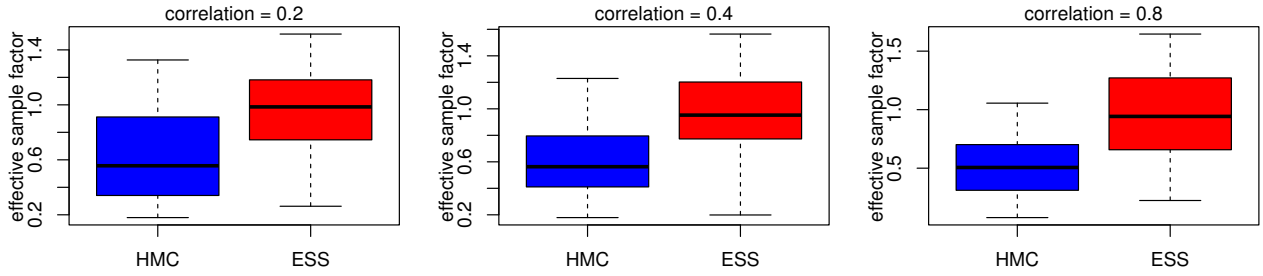

Figure 15: Performance illustration of the effective sample factor (ESF) for the two competing approaches (HMC blue and ESS red) based on 20,000 samples, with the first 5,000 discarded as burn-in. The sampling scheme is repeated 50 times for different values of the correlation parameter ρ. The mean of the prior $\boldsymbol{\mu}$ is fixed at $[-2, 1]^\top$ and the set of constraints is the positive orthant. The proposed ESS approach outperforms HMC as it admits a higher ESF.

Figure 15 shows the ESF boxplots of the first variable x_1 for the two competing approaches (HMC blue and ESS red) for different correlation parameter and based on 20,000 samples, where

the first 5,000 are discarded as burn-in. The sampling scheme is repeated 50 times. In that case, the proposed ESS approach outperforms the HMC sampler as it admits a higher ESF for different values of the correlation parameter ρ ($\{\rho = 0.2,\ \rho = 0.4,\ \rho = 0.8\}$, representing low, moderate, and high correlation, respectively). As in Section 4.1, the approximate parameter η for the proposed approach is fixed at 50.

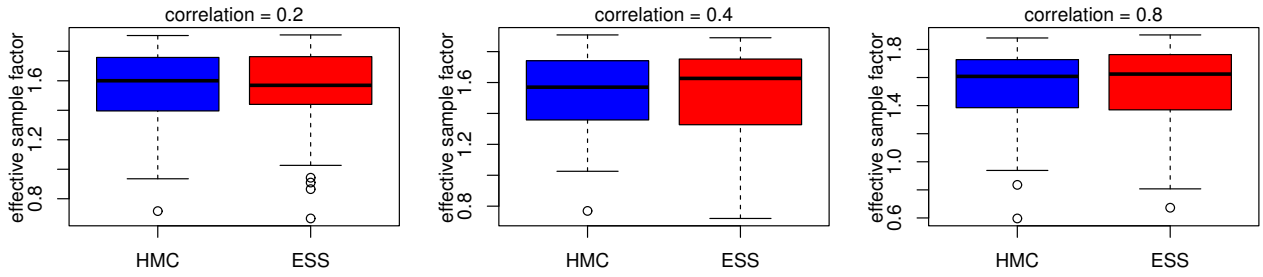

Figure 16: Performance illustration of the effective sample factor (ESF) for the two competing approaches (HMC blue and ESS red) based on 20,000 samples, with the first 5,000 discarded as burn-in. The sampling scheme is repeated 50 times for different values of the correlation parameter ρ. The mean of the prior $\boldsymbol{\mu}$ is fixed at $[5, 5]^\top$ and the set of constraints is the positive orthant.

Now, we investigate the numerical convergence of the two competing approaches (ESS and HMC). To do this, we consider in Figure 16 the case where the mean of the prior $\boldsymbol{\mu}$ lies inside the positive orthant such that the linear inequality constraints have no impact on the prior distribution. As with Figure 15, Figure 16 displays the ESF for the two competing approaches when the mean of the prior $\boldsymbol{\mu}$ is fixed at $[5, 5]^\top$. As expected, both approaches provide equivalently high values of the ESF.

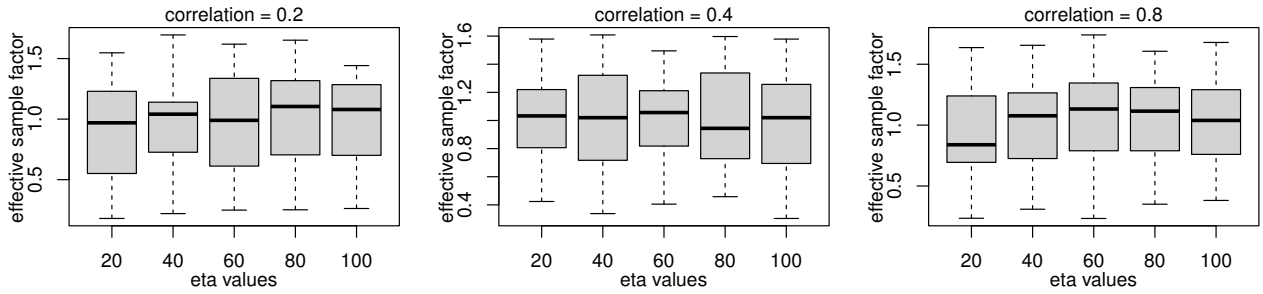

Figure 17: Performance illustration of the ESF using the proposed ESS approach for different values of η and based on 20,000 samples, where the first 5,000 are discarded as burn-in. The sampling scheme is repeated 50 times for different values of the correlation parameter ρ. The mean of the prior $\boldsymbol{\mu}$ is fixed at $[-2, 1]^\top$ and the set of constraints is the positive orthant.

Figure 17 displays the evolution of the ESF of the proposed ESS approach as a function of the approximate parameter η for different values of the correlation parameter ρ. Under the same settings as Figure 15, the ESF of the proposed approach remains stable for different values of η.

Based on the results obtained in this section, the proposed ESS algorithm can be seen as a general, efficient, and flexible approach for generating tMVN distributions in both extreme and non-extreme cases. As discussed, the approximate parameter η should be updated at each MCMC iteration to ensure the stability, precision, and convergence of the proposed approach.

5 Application to constrained Gaussian process regression

The aim of this section is to demonstrate the performance of the algorithm developed in Section 2.1 within the context of Bayesian shape-restricted function estimation. A comparison with the highly

efficient HMC sampler in terms of computational running time is included. To accomplish this, we consider the nonparametric function estimation through GP regression, where the unknown function satisfies structured shape constraints such as monotonicity, boundedness, or convexity, or any combination of these. Let $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ be a set of n noisy samples. The following regression problem is considered

$$y_i = f(\boldsymbol{x}_i) + \epsilon_i, \quad \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \tag{20}$$

$i = 1, \ldots, n$, where f represents an unknown function that generates the data $\boldsymbol{y} = [y_1, \ldots, y_n]^\top$ and respects shape constraints such as monotonicity, convexity, and boundedness, either applied individually or in combination. The d-dimensional vector $\boldsymbol{x}_i \in \mathbb{R}^d$ is a covariate, and ϵ_i is an additive, independent and identically distributed (i.i.d.) zero-mean Gaussian noise with constant variance σ^2. Throughout this section, the noise variance σ^2 is estimated at each MCMC iteration using an inverse Gamma distribution. Gaussian processes (GPs) are powerful Bayesian models widely employed in machine learning for solving regression problems. This is due to efficient sampling algorithms, a rich methodological literature, and a strong theoretical foundation [44]. It is based on assuming a GP prior distribution $(Y(x))_{x\in\mathcal{D}}$ on the unknown function f, where we consider $\mathcal{D}$ to be the unit interval for simplicity. We also assume, without loss of generality, that Y is a zero-mean GP with covariance function k, denoted as $Y \sim \mathcal{GP}(0, k(\cdot,\cdot))$. A GP, in general, is characterized by its covariance function k, which plays a crucial role in incorporating assumptions such as differentiability, sparsity, and periodicity. In this section, the Matérn family of covariance functions (18) is employed, which is widely used in machine learning community. The parameters τ^2 and ℓ, which are positive, are commonly known as the signal variance and correlation length-scale, respectively. Since these parameters are typically unknown, they must be estimated from available data. For further details on parameter estimation techniques, we refer the reader to [16, 44].

Structural constraints, such as monotonicity, boundedness, and convexity, appear in many real-world data applications [9, 26] and are commonly enforced by expanding the function in a suitable basis. These constraints can be induced by imposing linear inequality constraints on the basis coefficients. Examples of such bases include piecewise linear functions [13], splines [4, 30], and Bernstein polynomials [11, 43]. In this section, our focus is on the compactly supported bases developed in [24, 26]:

$$y_i = \sum_{j=1}^{N} \zeta_j \phi_j(x_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{21}$$

where $\zeta_j = Y(t_j)$ represents the value of the parent GP Y at a uniform grid point $\{t_j\}$ for $j = 1, \ldots, N$ and $\{\phi_j\}$ are the compactly supported basis functions of class $C^0(\mathcal{D}, \mathbb{R})$[4], see [24] for more details on the properties of these basis functions. Since Y is a zero-mean GP with covariance function k, we deduce that $\boldsymbol{\zeta} = [\zeta_1, \ldots, \zeta_N]^\top$ is a zero-mean Gaussian vector with covariance matrix $\tau^2\boldsymbol{\Gamma}$ such that $\boldsymbol{\Gamma}_{j,l} = \frac{1}{\tau^2} k(t_j, t_l)$, for any $j, l = 1, \ldots, N$. Since k is a stationary covariance function and $\{t_j\}$ are equally spaced, the covariance $\boldsymbol{\Gamma}$ exhibits Toeplitz structure. This property is explored in the numerical examples of this section. In [24], the authors propose several bases where various structured constraints are equivalently translated into linear inequality constraints on the basis coefficients $\{\zeta_j\}$. Furthermore, in [26], additional smoothness basis functions are introduced with the aim of generalizing the approach in [24] to deal with higher orders of smoothness paths. Through Model (21) and the compactly supported bases from [24], monotonicity, boundedness, and convexity

[4] $C^0(\mathcal{D}, \mathbb{R})$ represents the set of continuous functions from $\mathcal{D}$ into $\mathbb{R}$.

$$\mathcal{E} = \begin{cases} \mathcal{E}_m = \{f \in C^0(\mathcal{D}, \mathbb{R}) \ : \ f(x) \leq f(y), \ \forall x < y\} \quad \text{(monotonicity)} \\ \mathcal{E}_b = \{f \in C^0(\mathcal{D}, \mathbb{R}) \ : \ l_b \leq f(x) \leq u_b, \ \forall x \in \mathcal{D}\} \quad \text{(boundedness)} \\ \mathcal{E}_c = \left\{f \in C^0(\mathcal{D}, \mathbb{R}) \ : \ \frac{f(x')-f(x)}{x'-x} \leq \frac{f(x'')-f(x')}{x''-x'}, \ x < x' < x''\right\} \quad \text{(convexity)} \end{cases} \tag{22}$$

are equivalent to the following linear inequality constraints on the basis coefficients ζ_j:

$$\mathcal{C} = \begin{cases} \mathcal{C}_m = \left\{\boldsymbol{\zeta} \in \mathbb{R}^N \ : \ \zeta_{j-1} \leq \zeta_j, \ j = 2, \ldots, N\right\}; \\ \mathcal{C}_b = \left\{\boldsymbol{\zeta} \in \mathbb{R}^N \ : \ l_b \leq \zeta_j \leq u_b, \ j = 1, \ldots, N\right\}; \\ \mathcal{C}_c = \left\{\boldsymbol{\zeta} \in \mathbb{R}^N \ : \ \frac{\zeta_{j-1}-\zeta_{j-2}}{t_{j-1}-t_{j-2}} \leq \frac{\zeta_j-\zeta_{j-1}}{t_j-t_{j-1}}, \ j = 3, \ldots, N\right\}; \end{cases} \tag{23}$$

which corresponds to monotonicity, boundedness, and convexity constraints, respectively. It is worth noting that the above linear inequality constraints on the basis coefficients can be expressed in matrix form $\boldsymbol{A}\boldsymbol{\zeta} + \boldsymbol{b} \geq \boldsymbol{0}$, where $\boldsymbol{A}$ is an $m \times N$ matrix and $\boldsymbol{b}$ is an m-dimensional vector. Thus,

$$Y^N(x) := \sum_{j=1}^{N} \zeta_j \phi_j(x) \in \mathcal{E} \quad \Leftrightarrow \quad \boldsymbol{\zeta} \in \mathcal{C},$$

where $\mathcal{E}$ and $\mathcal{C}$ are given in (22) and in (23), respectively. Given the noisy observations $\boldsymbol{y} = [y_1, \ldots, y_n]^\top$, sampling Y^N conditionally on $\boldsymbol{y}$ and respecting shape constraints is equivalent to generating $\{\boldsymbol{\zeta} | \boldsymbol{X}\boldsymbol{\zeta} + \boldsymbol{\epsilon} = \boldsymbol{y}, \ \boldsymbol{\zeta} \in \mathcal{C}\}$, where $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_n]^\top$ is a zero-mean Gaussian noise vector with covariance matrix $\sigma^2 \mathbf{I}_n$, and $\boldsymbol{X}$ is an $n \times N$ matrix such that $\boldsymbol{X}_{i,j} = (\phi_j(x_i))$. In that case, we have

$$\{\boldsymbol{\zeta} | \boldsymbol{y}, \ \boldsymbol{\zeta} \in \mathcal{C}\} \sim \mathcal{TN}_N(\boldsymbol{m}, \boldsymbol{\Sigma}; \mathcal{C}), \quad \text{where,} \tag{24}$$

$$\begin{aligned} \boldsymbol{m} &= \left(\boldsymbol{X}^\top \boldsymbol{X}/\sigma^2 + \boldsymbol{\Gamma}^{-1}/\tau^2\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}/\sigma^2; \\ \boldsymbol{\Sigma} &= \left(\boldsymbol{X}^\top \boldsymbol{X}/\sigma^2 + \boldsymbol{\Gamma}^{-1}/\tau^2\right)^{-1}; \end{aligned}$$

and $\mathcal{C}$ is the set of linear inequality constraints (23). Let us mention that sampling from the posterior tMVN distribution in (24) can be achieved through the highly efficient HMC sampler developed by [35] and implemented in the `R`-package *tmg*. However, in situations where N is large, $(\boldsymbol{X}^\top \boldsymbol{X}/\sigma^2 + \boldsymbol{\Gamma}^{-1}/\tau^2)$ changes during each MCMC iteration due to updates in σ^2 and τ^2, necessitating an $N \times N$ matrix inversion at every iteration. Moreover, within a large MCMC algorithm, updating the unknown covariance function parameters involves computing the inversion of an $N \times N$ matrix at each step. To apply the proposed approach developed in Section 2 which avoids matrix inversion, we first compute the pdf of the tMVN distribution (24):

$$p(\boldsymbol{\zeta} | \boldsymbol{y}, \boldsymbol{\zeta} \in \mathcal{C}) \propto \underbrace{\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\zeta}\|^2\right)}_{\text{likelihood function}} \underbrace{\exp\left(-\frac{1}{2\tau^2}\boldsymbol{\zeta}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\zeta}\right)}_{\text{Gaussian prior}} \underbrace{\mathbf{1}_{\mathcal{C}}(\boldsymbol{\zeta})}_{\text{constraint}}, \quad \boldsymbol{\zeta} \in \mathbb{R}^N. \tag{25}$$

The above constrained posterior pdf (25) is proportional to a likelihood function derived from the data, a prior MVN distribution, and an indicator function representing the set of linear inequality constraints (23). Using the approximate $\gamma_\eta(\cdot)$ of $\mathbf{1}_{\mathcal{C}}$ proposed in Section 2.1, the above posterior pdf (25) can be approximated as follows:

$$\begin{aligned} \tilde{p}_\eta(\boldsymbol{\zeta} | \boldsymbol{y}, \boldsymbol{\zeta} \in \mathcal{C}) \ &\propto \ \underbrace{\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\zeta}\|^2\right) \gamma_\eta(\boldsymbol{\zeta})}_{\textit{new} \text{ likelihood function}} \underbrace{\exp\left(-\frac{1}{2\tau^2}\boldsymbol{\zeta}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\zeta}\right)}_{\text{Gaussian prior}}, \quad \boldsymbol{\zeta} \in \mathbb{R}^N \\ &= \ \underbrace{\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\zeta}\|^2\right) \prod_{\kappa=1}^{m} \frac{1}{1 + \exp(-\eta[\boldsymbol{a}_\kappa^\top \boldsymbol{\zeta} + b_\kappa])}}_{L_\eta(\boldsymbol{\zeta})} \underbrace{\exp\left(-\frac{1}{2\tau^2}\boldsymbol{\zeta}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\zeta}\right)}_{\text{Gaussian prior}}, \end{aligned} \tag{26}$$

where m represents the number of linear inequality constraints on the basis coefficients (23). For instance, when boundedness constraints are imposed (i.e., $\boldsymbol{\zeta} \in \mathcal{C}_b$), we obtain $m = 2N$ linear constraints. Furthermore, if monotonicity non-decreasing and upper bound constraints are imposed together (i.e., $\boldsymbol{\zeta} \in \mathcal{C}_m \cap \mathcal{C}_b$), we obtain only $m = N$ linear inequality constraints (i.e., $\zeta_1 \leq \zeta_2 \leq \ldots \leq \zeta_N \leq u_b$, where $u_b \in \mathbb{R}$ is the upper bound). Now, the approximate posterior density in (26) is proportional to a *new* likelihood function $L_\eta(\boldsymbol{\zeta})$ and a zero-mean MVN prior. In that case, the efficient ESS described in Section 2.1 can be employed. From (26), the logarithm of the *new* likelihood function $L_\eta(\boldsymbol{\zeta})$ is given as follows:

$$\log(L_\eta(\boldsymbol{\zeta})) = -\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\zeta}\|^2 - \sum_{\kappa=1}^{m} \log(1 + \exp\{-\eta[\boldsymbol{a}_\kappa^\top \boldsymbol{\zeta} + b_\kappa]\}).$$

Hence, it admits a computational complexity of order $\mathcal{O}(nN)$, where n represents the number of samples and N represents the dimension of the Gaussian prior $\boldsymbol{\zeta}$. Consequently, this approach which avoid matrix inversion, can handle large data sets. Furthermore, with this approach, there is no need to compute the posterior distribution as in (24), as sampling is performed before conditioning rather than after. This can be an advantage, as it simplifies the computational process, reduces the overall complexity, and can lead to more efficient sampling, particularly in high-dimensional spaces where evaluating the posterior distribution directly can be computationally prohibitive. For example, in our case, the covariance matrix of the prior $\boldsymbol{\zeta}$ exhibits Toeplitz structure. Additionally, $\boldsymbol{\zeta}$ is extracted from a stationary GP. In that case, several highly efficient samplers can be employed such as the Fast Fourier Transform developed by [45] and the recently fast large-scale sampler developed by [25].

5.1 Performance illustration (multiple shape-constraints)

The aim of this example is to investigate the performance of the approach developed in Section 2 for multiple shape-restricted function estimation using Model (21). To do this, we consider the target function $x \mapsto x^2$, $x \in \mathcal{D}$, which is increasing, bounded between 0 and 1, and convex. This is an interesting case since one can incorporate multiple shape constraints into Model (21) in order to improve the prediction accuracy and obtain more realistic confidence intervals.

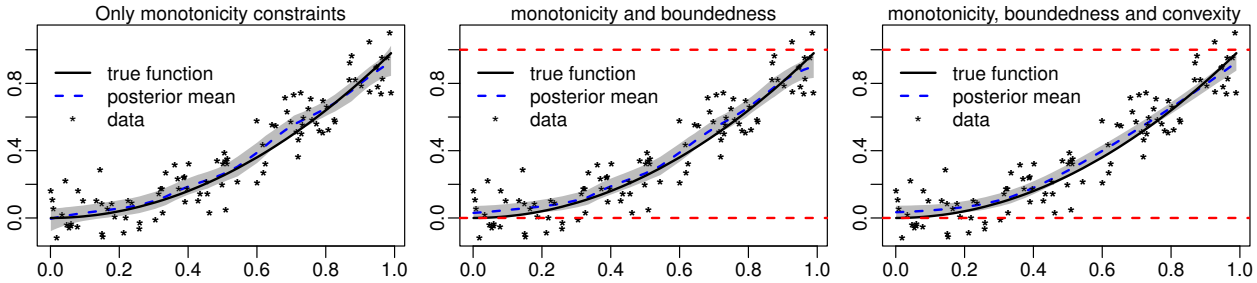

Figure 18: Performance illustration of the proposed approach through Model (21) under only monotonicity constraints (left), under both monotonicity and boundedness (middle), and under monotonicity, boundedness and convexity (right). The gray shaded area represents the 95% credible interval based on 15,000 MCMC iterations, where the first 5,000 are discarded as burn-in. The dashed horizontal lines represent the lower and upper bound constraints.

In Figure 18, we randomly generate $n = 100$ samples (black stars) from (20) using the target function $f(x) = x^2$ (black curve) and a noise standard deviation $\sigma = 0.1$. The covariate $\{x_i\}$, $i = 1, \ldots, n$ are generated uniformly between 0 and 1. The Matérn covariance function (18) is employed with a smoothness parameter $\nu = 2.5$. The length-scale parameter ℓ has been chosen such that the correlation at the maximum possible separation between the covariates equals 0.05.

The noise and signal variance parameters σ^2 and τ^2 are updated at each MCMC iteration using an inverse Gamma distribution. Figure 18 illustrates the performance of the proposed approach developed in Section 2 using Model (21) with multiple shape constraints and $N = n/4 = 25$ basis functions to avoid overfitting [26, 34]. Only monotonicity (non-decreasing) constraints are imposed in the left panel; monotonicity and boundedness between 0 and 1 are imposed in the middle; and monotonicity, boundedness and convexity are imposed in the right panel. The blue dashed line represents the mean of the posterior paths, which is computed numerical based on 15,000 MCMC iterations. While, the black solid curve represents the target function. The red horizontal lines represent the lower and upper bound constraints. The algorithm developed in Section 2.1 is employed with $\eta = 50$. The average of the noise variance parameter is 0.013 for the left and middle panel situations, and 0.014 for the right panel situation. It is visually clear that the incorporation of multiple shape constraints improves the prediction accuracy and reduces the confidence interval.

5.2 Numerical example: flat problem

The aim of this section is to investigate the prediction accuracy of the proposed ESS approach when applied to a Bayesian shape restricted function estimation. To do this, we consider the following monotone (nondecreasing) function:

$$f(x) = \frac{3}{1 + \exp\{-10x + 2.1\}}, \quad x \in [0, 1]. \tag{27}$$

This is a difficult situation because the function f is nondecreasing and approximately flat on $[0.7, 1]$.

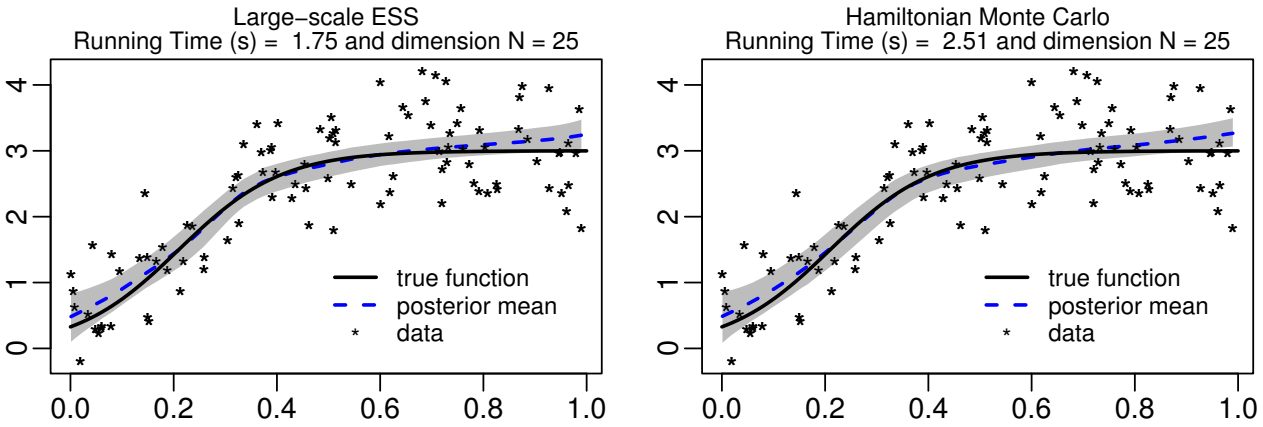

Figure 19: Prediction accuracy of the proposed ESS approach (left) and HMC (right) when using Model (21) under monotonicity constraints. The gray shaded area represents the 95% credible interval based on 15,000 MCMC iterations, where the first 5,000 are discarded as burn-in.

As in Section 5.1, the black stars in Figure 19 represents the $n = 100$ training data, which are randomly generated from (20) using the target function f in (27) (black curve) and a noise standard deviation $\sigma = 0.5$. Figure 19 shows a performance illustration of the two competing approaches, ESS (left panel) and HMC (right panel), when using Model (21) under monotonicity constraints with $N = n/4 = 25$ basis functions to avoid overfitting [26, 34]. Both approaches fit the data well on the interval where the function is increasing. However, the HMC sampler fails to capture the flat region on the interval $[0.7, 1]$. Furthermore, the running time in seconds of generating 15,000 monotonic curves is displayed in the main of each panel.

Table 2 summarizes the average RMSE ($\times 10^2$) over 50 replicates of the posterior mean and the running time in seconds of generating 15,000 monotonic curves. The simulation study is based on a dataset of size 150 generated from (20) using the target function f (27) and a noise standard deviation $\sigma = 0.5$. The dataset is split into a training set of size 100 and a testing set of size 50.

	Average RMSE	Running time (s)
Posterior mean (ESS sampler)	**58.71**	**1.75**
Posterior mean (HMC sampler)	59.33	2.51

Table 2: Average RMSE ($\times 10^2$) over 50 replicates of the posterior mean and the computational running time in seconds for generating 15,000 MCMC monotone sample paths for the two competing approaches, ESS and HMC.

As expected and based on 50 replicates, the proposed approach outperforms the HMC sampler in terms of prediction accuracy and computational running time. This is due to the *mass-shifting* phenomenon, which is more pronounced when the HMC sampler is employed.

5.3 Real application

The aim of this section is to demonstrate the performance of the algorithm developed in Section 2.1 in terms of computational running time through a real world data application.

Electricity usage and temperature data

The data set used has 55 observations on monthly electricity usage (kilowatt-hours) and average temperature (degrees Fahrenheit) for a house in Westchester County, New York, USA. The aim is to estimate the electricity usage as a function of temperature. The data suggests that the true underlying function exhibits a monotone (non-increasing) constraints. This dataset serves to highlight the superiority of the proposed approach, denoted as LS-ESS, over the HMC sampler in terms of computational running time.

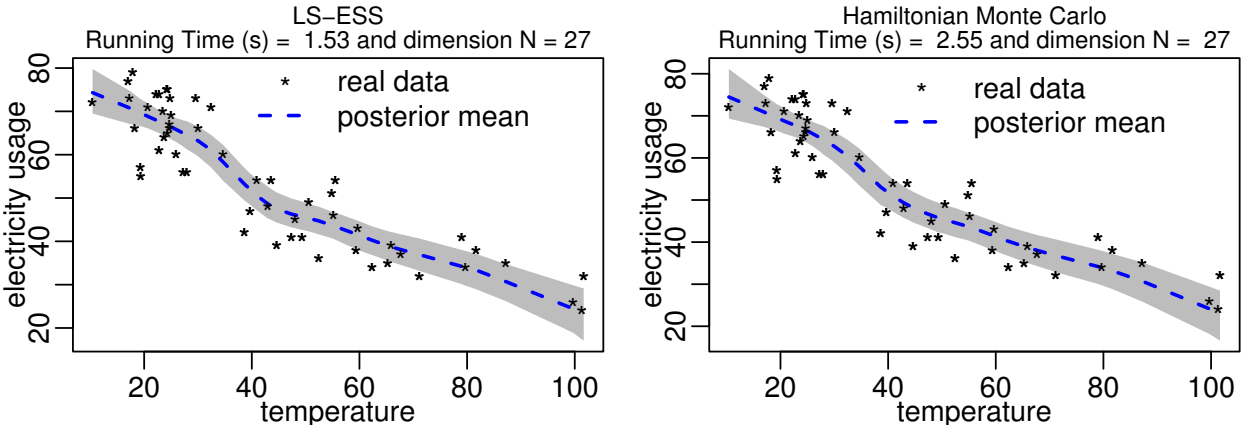

Figure 20: Estimation accuracy of the two competing methods applied on the electricity usage and temperature data. The computational running time for generating 15,000 monotonic trajectories is displayed for both proposed and HMC approaches. The dimension of the prior Gaussian vector in Model (21) is fixed at $N = 27$. The gray shaded area represents the 95% credible interval based on 15,000 MCMC iterations, where the first 5,000 are discarded as burn-in. The WAIC criterion is equal to 912.84 and 912.60 for LS-ESS and HMC approaches, respectively.

Figure 20 demonstrates the performance of the proposed approach (left panel) compared to the HMC sampler (right panel) in terms of computational running time. The proposed approach is denoted as LS-ESS. The same parameters as in Figure 18 are used, except for the number of basis function, which is set to $N = 27$, approximately half the number of samples. The average of the noise variance parameter σ^2 is 41.91 and 42.56 for the LS-ESS and the HMC sampler, respectively. The gray shaded area represents the 95% credible interval based on 15,000 MCMC iterations, where the first 5,000 are discarded as burn-in. The proposed LS-ESS approach outperforms the HMC

sampler in terms of computational running time. It is worth noting that the computational running time of the HMC sampler depends on the smoothness parameter ν of the Matérn covariance kernel. It performs poorly for small values of ν. Furthermore, with the proposed approach, sampling is performed through an approximate posterior distribution. This leads to high flexibility and enables handling tasks that are impossible for the HMC sampler. Finally, let us mention that the proposed approach avoids matrix inversion, as sampling is performed before conditioning rather than after, allowing us to handle large data sets.

6 Conclusion

In this paper, we develop an efficient approach for generating multivariate normal distribution restricted to linear, quadratic, and nonlinear inequality constraints. It is based on incorporating a smooth relaxation of the set of complex constraints from the constrained density function into a likelihood function, and then applying a highly efficient Markov Chain Monte Carlo (MCMC) approach to sample from the resulting distribution. The theoretical convergence of the proposed approach is provided. The proposed approach extends the recent literature on nonlinear inequality constraints, including both convex and non-convex sets. We pay careful attention to updating the approximate parameter at each MCMC iteration, allowing us to effectively address very extreme cases while ensuring the precision and convergence of the proposed approach. Additionally, we developed an efficient formula for the log-likelihood function when the restricted domain is defined by constant bounds, resulting in a significant reduction in computational complexity in high-dimensions. The excellent effectiveness and adaptability of the proposed approach have been showcased within the framework of Bayesian shape-restricted function estimation.

Acknowledgements

This research was conducted with the support of the consortium in Applied Mathematics CIROQUO (`https://doi.org/10.5281/zenodo.6581217`), gathering partners in technological and academia in the development of advanced methods for Computer Experiments.

References

[1] J. H. Albert and S. Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.

[2] Z. I. Botev. The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):125–148, 2017.

[3] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[4] B. Cai and D. B. Dunson. Bayesian multivariate isotonic regression splines: applications to carcinogenicity studies. *Journal of the American Statistical Association*, 102(480):1158–1171, 2007.

[5] M. Chataigner, A. Cousin, S. Crépey, M. Dixon, and D. Gueye. Beyond surrogate modeling: Learning the local volatility via shape constraints. *SIAM Journal on Financial Mathematics*, 12(3):SC58–SC69, 2021.

[6] Y. Cong, B. Chen, and M. Zhou. Fast simulation of hyperplane-truncated multivariate normal distributions. *Bayesian Analysis*, 12(4):1017–1037, 2017.

[7] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, 28(3):424–446, 2013.

[8] A. Cousin, A. Deleplace, and A. Misko. Gaussian process regression for swaption cube construction under no-arbitrage constraints. *Risks*, 10(12):232, 2022.

[9] A. Cousin, H. Maatouk, and D. Rullière. Kriging of financial term-structures. *European Journal of Operational Research*, 255(2):631–648, 2016.

[10] M. Cowles and B. Carlin. Markov Chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91:883–904, 1996.

[11] S. M. Curtis and S. K. Ghosh. A variable selection approach to monotonic regression with Bernstein polynomials. *Journal of Applied Statistics*, 38(5):961–976, 2011.

[12] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero. Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery. *IEEE Transactions on Signal Processing*, 57(11):4355–4368, 2009.

[13] D. B. Dunson and B. Neelon. Bayesian inference on order-constrained parameters in generalized linear models. *Biometrics*, 59(2):286–295, 2003.

[14] N. Ellis and R. Maitra. Multivariate Gaussian simulation outside arbitrary ellipsoids. *Journal of Computational and Graphical Statistics*, 16(3):692–708, 2007.

[15] A. E. Gelfand, A. F.M. Smith, and T.-M. Lee. Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, 87(418):523–532, 1992.

[16] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 3 edition, 2013.

[17] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

[18] J. F. Geweke. Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pages 571–578, 1991.

[19] D. Goldfarb and A. Idnani. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, 27(1):1–33, 1983.

[20] W. Horrace. Some results on the multivariate truncated normal distribution. *Journal of Multivariate Analysis*, 94:209–221, 2005.

[21] J. H. Kotecha and P. M. Djuric. Gibbs sampling approach for generation of truncated multivariate Gaussian random variables. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 3, pages 1757–1760. IEEE, 1999.

[22] J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer, New York, Berlin, Heidelberg, 2008.

[23] H. Maatouk and X. Bay. *A new rejection sampling method for truncated multivariate Gaussian random variables restricted to convex sets*, volume 163, pages 521–530. Springer International Publishing, Cham, 2016.

[24] H. Maatouk and X. Bay. Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences*, 49(5):557–582, 2017.

[25] H. Maatouk, D. Rullière, and X. Bay. Sampling large hyperplane-truncated multivariate normal distributions. *Computational Statistics*, 39:1779–1806, 2023.

[26] H. Maatouk, D. Rullière, and X. Bay. Bayesian analysis of constrained Gaussian processes. *Bayesian Analysis*, pages 1–30, 2024.

[27] H. Maatouk, D. Rullière, and X. Bay. Efficient constrained Gaussian process approximation using elliptical slice sampling. working paper or preprint, March 2024.

[28] H. Maatouk, D. Rullière, and X. Bay. Large-scale constrained Gaussian processes for shape-restricted function estimation. *To appear in Statistics and Computing*, 2024.

[29] R. E. McCulloch, N. G. Polson, and P. E. Rossi. A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of econometrics*, 99(1):173–193, 2000.

[30] M. C. Meyer, A. J. Hackstadt, and J. A. Hoeting. Bayesian estimation and inference for generalised partial linear models using shape-restricted splines. *Journal of Nonparametric Statistics*, 23(4):867–884, 2011.

[31] I. Murray, R. Adams, and D. MacKay. Elliptical slice sampling. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 541–548. JMLR Workshop and Conference Proceedings, 2010.

[32] R. M. Neal. Regression and classification using Gaussian process priors. *Bayesian statistics 6*, pages 475–501, 1999.

[33] R. M. Neal. Slice sampling. *The annals of statistics*, 31(3):705–767, 2003.

[34] B. Neelon and D. B. Dunson. Bayesian isotonic regression and trend analysis. *Biometrics*, 60(2):398–406, 2004.

[35] A. Pakman and L. Paninski. Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *Journal of Computational and Graphical Statistics*, 23(2):518–542, 2014.

[36] N. G. Polson, J. G. Scott, and J. Windle. The Bayesian bridge. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(4):713–733, 2014.

[37] P. Ray, D. Pati, and A. Bhattacharya. Efficient Bayesian shape-restricted function estimation with constrained Gaussian process priors. *Statistics and Computing*, 30(4):839–853, 2020.

[38] G. Rodriguez-Yam, R. A. Davis, and L. L. Scharf. Efficient Gibbs sampling of truncated multivariate normal with application to constrained linear regression. *Technical repport*, 2004.

[39] M. Stegun and Abramowitz I. A. Handbook of Mathematical Functions. Dover, New York. 1965.

[40] L. P. Swiler, M. Gulian, A. L. Frankel, C. Safta, and J. D. Jakeman. A survey of constrained Gaussian process regression: Approaches and implementation challenges. *Journal of Machine Learning for Modeling and Computing*, 1(2), 2020.

[41] J. Taylor and Y. Benjamini. RestrictedMVN: multivariate normal restricted by affine constraints. *R package version*, 1, 2016.

[42] M. Vono, N. Dobigeon, and P. Chainais. High-dimensional Gaussian sampling: a review and a unifying approach based on a stochastic proximal point algorithm. *SIAM Review*, 64(1):3–56, 2022.

[43] J. Wang and S. Ghosh. Shape restricted nonparametric regression with Bernstein polynomials. *Computational Statistics & Data Analysis*, 56(9):2729–2741, 2012.

[44] C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[45] A. TA Wood and G. Chan. Simulation of stationary Gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics*, 3(4):409–432, 1994.

[46] S. Zhou, P. Giulani, J. Piekarewicz, A. Bhattacharya, and D. Pati. Reexamining the proton-radius problem using constrained Gaussian processes. *Physical Review C*, 99:055202, May 2019.

[47] S. Zhou, P. Ray, D. Pati, and A. Bhattacharya. A mass-shifting phenomenon of truncated multivariate normal priors. *Journal of the American Statistical Association*, 119(545):582–596, 2024.