



HAL
open science

A note on simulating hyperplane-truncated multivariate normal distributions

Hassan Maatouk, Xavier Bay, Didier Rullière

► **To cite this version:**

Hassan Maatouk, Xavier Bay, Didier Rullière. A note on simulating hyperplane-truncated multivariate normal distributions. *Statistics and Probability Letters*, 2022, 191, 10.1016/j.spl.2022.109650 . hal-03581252v2

HAL Id: hal-03581252

<https://hal.science/hal-03581252v2>

Submitted on 26 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A note on simulating hyperplane-truncated multivariate normal distributions

Hassan Maatouk^{†1}, Xavier Bay^{‡2} and Didier Rullière^{‡3}

(†) CY Tech Cergy Paris University, Site du Parc, 95011 Cergy, France

(‡) Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, F - 42023 Saint-Etienne France

Abstract Statistical researchers have shown increasing interest in generating conditional multivariate normal distributions. In this paper, we discuss several existing methods for the simulation of multivariate normal distribution truncated on the intersection of a set of hyperplanes. We also propose an approach based on the consideration of an orthonormal basis on the set of constraints. Contrarily to the standard approaches, we do not need to compute the covariance matrix of the posterior distribution and its decomposition. The interest of the proposed approach is shown through numerical examples.

Keywords Conditional Gaussian vectors · hyperplanes · projection · orthonormal basis.

1 Introduction

The simulation of conditional multivariate normal (MVN) distributions with linear equality constraints is widely used in many spatial statistics problems for computing Monte-Carlo estimators (see [1, 3, 4, 7]). As conditional simulations are time and memory expensive, an important challenge is the reduction of computation time through efficient algorithms. For instance, in

¹hmk@cy-tech.fr

²bay@emse.fr

³drulliere@emse.fr

[1], the authors study the settings where conditioning observations are assimilated batch-sequentially, i.e. one point or batch of points at each stage. To the best of our knowledge, the Matheron's update rule has first appeared in geostatistics [6] and then in astrophysics [5]. The idea is to simulate a random variable from the unconstrained distribution and then map it to the set of constraints. In [2], the authors provide a fast and easy-to-implement simulation algorithm for hyperplane-truncated multivariate normal distributions. This method can be seen as a Matheron's update rule. The authors in [2] generalize the Matheron's update rule algorithm to efficiently simulate random variables from a multivariate normal distribution whose covariance (precision) matrix can be decomposed as a positive-definite matrix minus (plus) a low-rank symmetric matrix. Their idea is to simulate from a block diagonal covariance matrix and use the Matheron's update rule. Recently, the authors in [8] use the Matheron's update rule to simulate conditional Gaussian processes. This method is so-called *pathwise conditioning*, which has been applied to global optimization problems. By the Matheron's update rule, the simulation is based on the full prior distribution without take into account the reduction of the dimension. Additionally, the researchers usually use the standard approaches to simulate from the conditional distribution by computing the covariance matrix of the posterior distribution. In that case, once the conditional covariance matrix calculated, the sampling procedure does not use the information about the unconditional covariance matrix: indeed, the existing sampling programs usually take as input the desired mean and covariance matrix. Furthermore, sampling procedures do not always consider the reduced rank of the conditional (posterior) covariance matrix.

In the present paper, simulating conditional multivariate normal distribution under linear equality constraints is investigated. We provide a new formula for simulating conditional Gaussian vectors. The main idea of the proposed method is to compute the orthonormal basis on the set of constraints. Contrarily to the standard approaches, we do not need to compute the covariance matrix of the posterior distribution and its decomposition.

The article is structured as follows: In Section 2, in the first part the standard approaches are briefly recalled. In the second part, the theoretical results of the proposed approach are given. In Section 3, the efficiency and the performance of the proposed method are investigated. A comparison with different approaches in terms of stability is included.

2 Hyperplane-truncated MVN distributions

2.1 Framework

Let X be a vector following a N -dimensional multivariate normal (MVN) distribution with mean μ and covariance matrix Γ , i.e., $X \sim \mathcal{N}(\mu, \Gamma)$, where Γ is symmetric and positive-definite. We consider the problem of simulating the random vector X truncated on the intersection of $n < N$ hyperplanes

$$\{X|AX = b\},$$

where $A \in \mathbb{R}^{n \times N}$, $b \in \mathbb{R}^n$ and $\text{Rank}(A) = n$.

The following proposition presents the well-known result to compute the posterior distribution of conditional Gaussian vectors with linear equality constraints.

Proposition 1 (Posterior distribution). *Let X be a vector following a N -dimensional MVN distribution with mean μ and covariance matrix Γ . Then, the conditional distribution of X given $AX = b$ is multivariate normal, i.e.,*

$$\{X|AX = b\} \sim \mathcal{N}(\mu_c, C), \quad \text{where}$$

$$\begin{aligned} \mu_c &= \mu + \Gamma A^\top (A\Gamma A^\top)^{-1}(b - A\mu) \\ C &= \Gamma - (A\Gamma)^\top (A\Gamma A^\top)^{-1} A\Gamma, \end{aligned}$$

with A^\top the transpose of A .

Proof. The posterior mean μ_c is done as follows:

$$\begin{aligned} \mu_c &= \mathbb{E}[X|AX = b] = \mathbb{E}[X] + \Sigma_{AX,X}^\top \Sigma_{AX,AX}^{-1}(b - \mathbb{E}[AX]) \\ &= \mu + (A\Gamma)^\top (A\Gamma A^\top)^{-1}(b - A\mu), \end{aligned}$$

where $\Sigma_{AX,X} = \text{Cov}(AX, X) = A\Gamma$ is the covariance between the vectors AX and X . Applying Bayes' rule (see Section 2.1 in [8]), we get

$$\begin{aligned} C &= \text{Cov}(X|AX = b) = \Sigma_{X,X} - \Sigma_{AX,X}^\top \Sigma_{AX,AX}^{-1} \Sigma_{AX,X} \\ &= \Gamma - (A\Gamma)^\top (A\Gamma A^\top)^{-1} A\Gamma. \end{aligned}$$

□

The standard approach to simulate from $\mathcal{N}(\mu_c, C)$ consists of finding a positive integer k and a *scaling* matrix $S \in \mathbb{R}^{N \times k}$ such that $SS^\top = C$ and using

$$\{X|AX = b\} \stackrel{d}{=} \mu_c + S\epsilon, \tag{1}$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}_k, \mathbf{I}_k)$ is a k -dimensional standard normal vector with \mathbf{I}_k the $k \times k$ identity matrix. Many *scaling* matrices can be proposed: as an example, let O in $\mathbb{R}^{k \times k}$ be an orthogonal matrix, it is easy to see that if S is a suitable *scaling* matrix, then $\tilde{S} = SO$ is also suitable. Some suitable classical choices for the *scaling* matrix are given below. Next section gives the most standard sampling methods to simulate from the conditional distribution $\mathcal{N}(\mu_c, C)$.

2.2 Standard sampling methods

Proposition 2 (Cholesky decomposition). *According to previous notations,*

$$\{X|AX = b\} \stackrel{d}{=} \mu_c + L\epsilon, \quad (2)$$

where $\mu_c = \mathbb{E}[X|AX = b]$, L is the $N \times N$ lower triangle matrix with the property $LL^\top = C$ and ϵ is the N -dimensional standard normal vector. The lower triangle matrix L is the *scaling* matrix S in (1).

The truncated version of the Cholesky decomposition (2) can be written as

$$\{X|AX = b\} \stackrel{d}{=} \mu_c + L_p \tilde{\epsilon}, \quad (3)$$

where L_p is the $p \times N$ truncated Cholesky lower triangle and $\tilde{\epsilon}$ is the $p = N - n$ dimensional standard normal vector. In fact, L_p corresponds to the $p = N - n$ non-zero columns of the lower triangle matrix L .

Proposition 3 (eigendecomposition). *According to previous notations,*

$$\{X|AX = b\} \stackrel{d}{=} \mu_c + V\Lambda^{\frac{1}{2}}\epsilon, \quad (4)$$

where $\mu_c = \mathbb{E}[X|AX = b]$, ϵ is a N -dimensional standard normal vector, $V = (v_1, \dots, v_N)$ is the $N \times N$ matrix such that the columns are the eigenvectors of C associated to the eigenvalues $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$ and $\Lambda = \text{diag}(\boldsymbol{\lambda})$ is a diagonal matrix. Again, $V\Lambda^{\frac{1}{2}}$ is the *scaling* matrix S in (1).

The truncated version of the eigendecomposition (4) can be written as

$$\{X|AX = b\} \stackrel{d}{=} \mu_c + V_p \Lambda_p^{\frac{1}{2}} \tilde{\epsilon}, \quad (5)$$

where $V_p = (v_1, \dots, v_p)$ is the truncated $N \times p$ matrix such that the columns are the eigenvectors of C associated to the $p = N - n$ non-zero eigenvalues $\boldsymbol{\lambda}_p = (\lambda_1, \dots, \lambda_p)$, $\Lambda_p = \text{diag}(\boldsymbol{\lambda}_p)$ and $\tilde{\epsilon}$ is the $p = N - n$ dimensional standard normal vector.

The previous approaches presented in this section (eigendecomposition and Cholesky) used a *scaling* matrix S of the posterior covariance matrix C .

Now, we will see how simulating $\{X|AX = b\}$ is possible without computing the posterior covariance matrix C and its decomposition. The Matheron's update rule (MUR), see Theorem 1 in [8], is investigated. Let us briefly recall this method. Let Z_1 and Z_2 be jointly Gaussian random variables. Then, the random variable Z_1 conditional on $Z_2 = \beta$ can be expressed as

$$\{Z_1|Z_2 = \beta\} \stackrel{d}{=} Z_1 + \Sigma_{Z_1, Z_2} \Sigma_{Z_2, Z_2}^{-1} (\beta - Z_2),$$

where $\Sigma_{Z_1, Z_2} = \text{Cov}(Z_1, Z_2)$ the covariance between Z_1 and Z_2 . As mentioned in [8], a key difference with the standard approach is that we now sample before conditioning, rather than after.

Applying the previous result to simulate $\{X|AX = b\}$, where $Z_1 = X$ and $Z_2 = AX$, we get the following result.

Proposition 4 (Matheron's update rule). *Let $X \sim \mathcal{N}(\mu, \Gamma)$, then*

$$\{X|AX = b\} \stackrel{d}{=} X + \Gamma A^\top (A \Gamma A^\top)^{-1} (b - AX). \quad (6)$$

Algorithm 1: Sampling scheme by Matheron's update rule of $X \sim \mathcal{N}(\mu, \Gamma)$ given $AX = b$.

- sample $w \sim \mathcal{N}(\mu, \Gamma)$;
 - return $X = w + \Gamma A^\top (A \Gamma A^\top)^{-1} (b - Aw)$.
-

2.3 Alternative method to the spectral decomposition

Let $F = \{x \in \mathbb{R}^N : Ax = b\}$ be the intersection set of n affine hyperplanes and directed by the vector space $F_0 = \{x \in \mathbb{R}^N : Ax = \mathbf{0}_n\}$, where $\mathbf{0}_n = (0, \dots, 0)^\top \in \mathbb{R}^n$.

Since the matrix A is of rank n , we need $p = N - n$ linear coordinates to describe F and to represent the conditional vector $X \in \mathbb{R}^N$ given $AX = b$. The simple idea is to choose the Euclidean coordinates associated to the eigendecomposition or spectral representation of the truncated MVN distribution of X given $AX = b$. To be more precise, let us consider the truncated density function f_T of the random vector $X \in \mathbb{R}^N$ given $AX = b$. From Proposition 1, we already known the posterior mean $\mu_c = \mathbb{E}[X|AX = b]$ of this distribution and we can write f_T as

$$f_T(x) = k \exp\left(-\frac{1}{2}(x - \mu_c)^\top \Gamma^{-1}(x - \mu_c)\right), \quad \text{for } x \in F,$$

where k is a normalization constant (on F). Note that the quadratic term in f_T is given by the same covariance matrix Γ of the unconditional distribution of $X \sim \mathcal{N}(\mu, \Gamma)$.

Now, let B be the (symmetric) matrix associated to the orthogonal projection on F_0 . As $x - \mu_c \in F_0$, we get $B(x - \mu_c) = x - \mu_c$, for any $x \in F$ and the alternative formula

$$f_T(x) = k \exp\left(-\frac{1}{2}(x - \mu_c)^\top B\Gamma^{-1}B(x - \mu_c)\right),$$

for the truncated density function. By the spectral decomposition of the (symmetric) matrix $B\Gamma^{-1}B$, we get

$$B\Gamma^{-1}B = \Omega\Delta\Omega^\top,$$

where $\Omega^\top\Omega = \mathbf{I}_p$, the $(p \times p)$ identity matrix, and $\Delta = \text{diag}(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_p^2})$ the diagonal matrix of non zero eigenvalues.

As a result, we obtain the nice representation for the conditional vector X given $AX = b$:

$$\{X|AX = b\} \stackrel{d}{=} \mu_c + \sum_{k=1}^p \sigma_k \epsilon_k \Omega_k,$$

where $\Omega_1, \dots, \Omega_p$ are the column vectors of Ω or the first p eigenvectors of $B\Gamma^{-1}B$ and $\epsilon_1, \dots, \epsilon_p$ are independent normal $\mathcal{N}(0, 1)$ coordinates. This result is summarized in the following theorem.

Theorem 2.1 (Proposed sampling scheme). *According to previous notations, define*

$$B = \mathbf{I}_N - A^\top(A^\top A)^{-1}A$$

to be the matrix of the orthogonal projection on $F_0 = \ker(A)$. The (symmetric) semi-positive matrix $B\Gamma^{-1}B$ is of rank p and admits the following eigendecomposition

$$B\Gamma^{-1}B = \Omega\Delta\Omega^\top,$$

where $\Omega^\top\Omega = \mathbf{I}_p$ and $\Delta = \text{diag}(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_p^2})$ with $\sigma_p^2 \geq \dots \geq \sigma_1^2 > 0$. The conditional vector $X \sim \mathcal{N}(\mu, \Gamma)$ given $AX = b$ is given by

$$\{X|AX = b\} \stackrel{d}{=} \mu_c + \Omega\Sigma\tilde{\epsilon}, \tag{7}$$

where $\Omega = (\Omega_1, \dots, \Omega_p) \in \mathbb{R}^{N \times p}$, $\tilde{\epsilon} = (\epsilon_1, \dots, \epsilon_p)^\top \in \mathbb{R}^p$ is the p dimensional standard normal vector and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p)$ is the $p \times p$ diagonal matrix. The matrix $\Omega\Sigma$ is the scaling matrix S in (1).

Proof. We have $B^2 = B$ and $B^\top = B$. For any $x \in \ker(A)$,

$$Bx = (\mathbf{I}_N - A^\top(A^\top A)^{-1}A)x = x.$$

Additionally, for any $x \in \ker(A)^\perp = \text{Im}(A^\top)$, there exists $u \in \mathbb{R}^n$ such that $x = A^\top u$ and

$$Bx = (\mathbf{I}_N - A^\top(A^\top A)^{-1}A)(A^\top u) = A^\top u - A^\top(A^\top A)^{-1}AA^\top u = A^\top u - A^\top u = 0.$$

Thus, B is the orthogonal projection on the $\ker(A) = F_0$. The matrix $B\Gamma^{-1}B$ is symmetric semi-positive and of rank $p = N - n$ since B is the orthogonal projector on $\ker(A)$ which is of dimension $p = N - n$.

The probability density function (pdf) of the vector $X \sim \mathcal{N}(\mu, \Gamma)$ is

$$f(x) = k \exp\left(-\frac{1}{2}(x - \mu)^\top \Gamma^{-1}(x - \mu)\right), \quad \forall x \in \mathbb{R}^N.$$

For any $x \in F = \{x \in \mathbb{R}^N : Ax = b\}$ and $\mu_c = \mathbb{E}[X|AX = b]$, we have $x - \mu_c \in F_0$ since x and μ_c are in F (i.e., $Ax = A\mu_c = b$). Let $q(x) = (x - \mu)^\top \Gamma^{-1}(x - \mu)$, thus

$$\begin{aligned} q(x) &= (x - \mu_c + \mu_c - \mu)^\top \Gamma^{-1}(x - \mu_c + \mu_c - \mu) \\ &= (x - \mu_c)^\top \Gamma^{-1}(x - \mu_c) + 2(x - \mu_c)^\top \Gamma^{-1}(\mu_c - \mu) + (\mu_c - \mu)^\top \Gamma^{-1}(\mu_c - \mu). \end{aligned}$$

From Proposition 1, we have

$$\begin{aligned} \Gamma^{-1}(\mu_c - \mu) &= \Gamma^{-1}(\Gamma A^\top (A\Gamma A^\top)^{-1}(b - A\mu)) \\ &= A^\top (A\Gamma A^\top)^{-1}(b - A\mu). \end{aligned}$$

Thus,

$$\langle x - \mu_c, A^\top (A\Gamma A^\top)^{-1}(b - A\mu) \rangle = \langle A(x - \mu_c), (A\Gamma A^\top)^{-1}(b - A\mu) \rangle = 0,$$

since $x - \mu_c \in F_0$, where $\langle \cdot, \cdot \rangle$ represents the scalar product. Finally, we get the pdf of $\{X|AX = b\}$ as follows

$$f_T(x) = k' \exp\left(-\frac{1}{2}(x - \mu_c)^\top \Gamma^{-1}(x - \mu_c)\right), \quad \forall x \in F,$$

where k' is a normalization constant. Now, for any $x \in F$, we have $(x - \mu_c) \in F_0$. Thus, $B(x - \mu_c) = (x - \mu_c)$. Therefore, the truncated pdf can be

reformulated as

$$\begin{aligned}
f_T(x) &= k' \exp \left(-\frac{1}{2} (x - \mu_c)^\top B \Gamma^{-1} B (x - \mu_c) \right) \\
&= k' \exp \left(-\frac{1}{2} (x - \mu_c)^\top \Omega \begin{pmatrix} \frac{1}{\sigma_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_p^2} \end{pmatrix} \Omega^\top (x - \mu_c) \right) \\
&= k' \exp \left(-\frac{1}{2} \sum_{k=1}^p \frac{y_k^2}{\sigma_k^2} \right),
\end{aligned}$$

where $(y_1, \dots, y_p)^\top = \Omega^\top (x - \mu_c)$ the euclidean coordinates of $(x - \mu_c)$ in the basis $(\Omega_1, \dots, \Omega_p)$ of F_0 . Finally, the pdf f_T corresponds to the vector

$$X = \mu_c + \sum_{k=1}^p \sigma_k \epsilon_k \Omega_k = \mu_c + \Omega \Sigma \epsilon,$$

where $\Omega = (\Omega_1, \dots, \Omega_p) \in \mathbb{R}^{N \times p}$, $\epsilon = (\epsilon_1, \dots, \epsilon_p)^\top \in \mathbb{R}^p$ is the p dimensional standard normal vector and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p)$. \square

The result proved in the Theorem 2.1 is very interesting in terms of simulation. It can be seen as an efficient and exact way to simulate hyperplane-truncated multivariate normal distributions.

Remark 1. *The proposed approach is easy to implement and can be extended to simulate truncated Gaussian vectors restricted to convex sets. Contrary to the standard approaches, we do not need to compute the covariance matrix C of the posterior distribution. Additionally, it is specially efficient when the unconditional matrix Γ admits some special structure that makes it easy to invert.*

Remark 2. *In fact, the proposed approach is the eigendecomposition of the hyperplane-truncated MVN distribution. From Equations (5) and (7), we get respectively*

$$\begin{aligned}
C &= \left(V_p \Lambda_p^{\frac{1}{2}} \right) \left(V_p \Lambda_p^{\frac{1}{2}} \right)^\top = V_p \Lambda_p V_p^\top, \quad \text{where } V_p^\top V_p = \mathbf{I}_p \text{ and} \\
C &= (\Omega \Sigma) (\Omega \Sigma)^\top = \Omega \Sigma^2 \Omega^\top, \quad \text{where } \Omega^\top \Omega = \mathbf{I}_p.
\end{aligned}$$

Thus, as eigenvalues are ordered here, we get $\Sigma^2 = \Lambda_p$.

In Algorithm 2, the different steps to simulate the hyperplane-truncated MVN distribution $\{X | AX = b\}$ using the proposed method are shown.

Algorithm 2: Proposed sampling scheme of $X \sim \mathcal{N}(\mu, \Gamma)$ given $AX = b$.

Initialization:

$\mu; A; b; \Gamma; N; n; p$

Projection on F_0 :

$$B = \mathbf{I}_N - A^\top(AA^\top)^{-1}A$$

Spectral decomposition of $B\Gamma^{-1}B$:

$\tilde{\Omega}$ is the matrix of p eigenvectors of $B\Gamma^{-1}B$

$$\Omega = B\tilde{\Omega}$$

$\frac{1}{\sigma^2} = (1/\sigma_1^2, \dots, 1/\sigma_p^2)$ the p non-zero eigenvalues of $B\Gamma^{-1}B$

$$\mu_c = \mu + \Gamma A^\top(A\Gamma A^\top)^{-1}(b - A\mu)$$

Simulation:

$$\epsilon \sim \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$$

$$X = \mu_c + \Omega\Sigma\epsilon, \text{ avec } \Sigma = \text{diag}(\sigma).$$

Illustrative example in two-dimensional case We consider the case when $N = 2$ and $n = 1$. Let $X = (X_1 \ X_2)^\top$ be a zero-mean bivariate Gaussian vector with covariance matrix the identity (i.e., $\Gamma = \mathbf{I}_2$). The aim is to simulate $X \sim \mathcal{N}(\mathbf{0}, \Gamma)$ given $AX = b$. Let $b = 1$ and $A = (1 \ 1)$. In that case, we get

$$\mu_c = \mathbb{E}[X|AX = b] = (0.5 \ 0.5)^\top$$

$$C = \text{Cov}(X|AX = b) = \Gamma - (A\Gamma)^\top(A\Gamma A^\top)^{-1}A\Gamma = \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix}$$

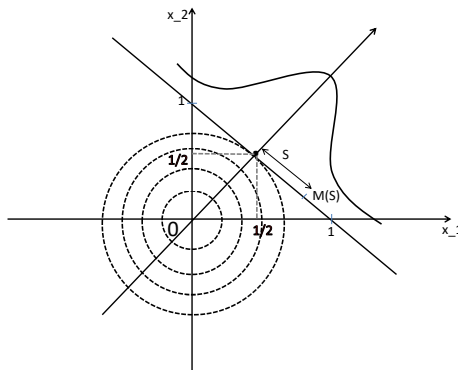


Figure 1: Illustration of the proposed method in two dimensional case.

From Theorem 2.1, we deduce that $\{X|AX = b\} \stackrel{d}{=} \mu_c + \Omega\tilde{\epsilon}$, where $\tilde{\epsilon} \sim \mathcal{N}(0, 1)$ and Ω is the orthonormal basis of the straight line $x_1 + x_2 = 1$.

We deduce that $\Omega = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}^\top$ (see Figure 1). Thus $X = (X_1 \ X_2)^\top = \begin{pmatrix} 0.5 + \frac{\sqrt{2}}{2}\epsilon & 0.5 - \frac{\sqrt{2}}{2}\epsilon \end{pmatrix}^\top$.

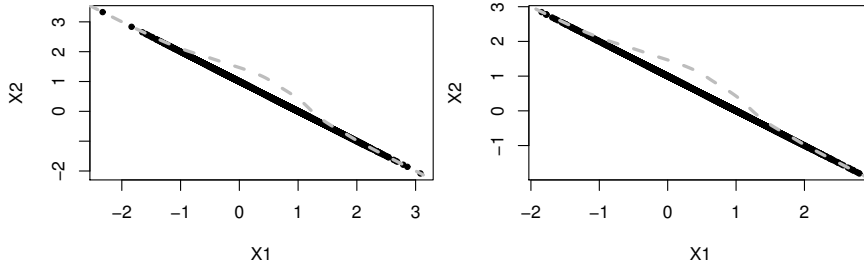


Figure 2: The density function together with the conditional bivariate normal distribution using the proposed approach (*left*) and the standard one (*right*).

In Figure 2, we illustrate the density function together with the conditional bivariate normal distribution using the proposed approach (*left*) and the standard one via the eigendecomposition (*right*). The black points represent 5,000 random variables $X = (X_1 \ X_2)^\top$ verifying $X_1 + X_2 = 1$ and the gray dashed-lines represent the associated probability density functions.

3 Simulation study

The aim of this section is to study the performance of the proposed method. We illustrate the theoretical results on numerical experiments. In this section, the elements of μ , A and b are sampled from $\mathcal{N}(0, 1)$. The unconditional covariance matrix Γ is generated as follows: let c be the Matérn covariance function with regularity parameter $\nu = 5/2$:

$$c(x, y) = \eta^2 \left(1 + \frac{\sqrt{5}|x - y|}{\theta} + \frac{5(x - y)^2}{3\theta^2} \right) \exp \left(-\sqrt{5} \frac{|x - y|}{\theta} \right),$$

where θ and η are fixed to 0.2 and 10 respectively. Let u_1, \dots, u_N be an uniform subdivision of $[0, 1]$ such that $u_1 = 0 < u_2 < \dots < u_N = 1$. Then $\Gamma_{j,\ell} = c(u_j, u_\ell)$, for all $j, \ell = 1, \dots, N$.

In all graphics, Cholesky and Eigen refer to the naive Cholesky decomposition (2) and eigendecomposition (4) respectively. However, T Chol and T Eigen refer to the truncated version of the Cholesky decomposition (3) and eigendecomposition (5) respectively. Finally, MUR represents the Mathéron's update rule (6).

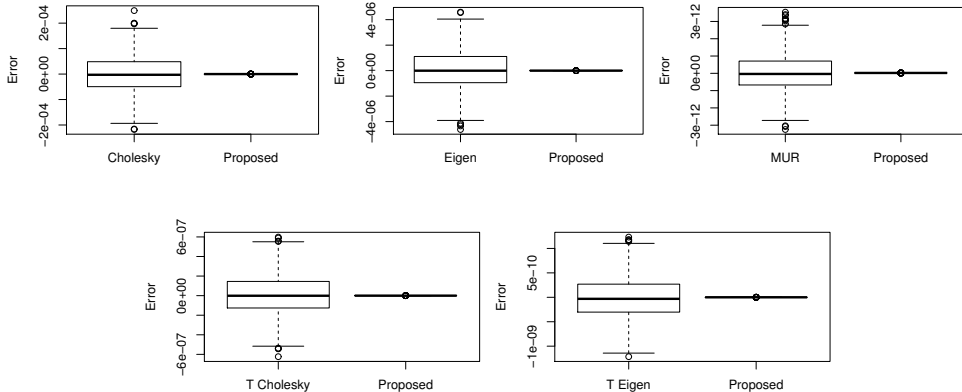


Figure 3: The numerical error $AX - b$ using different approaches.

Let us first consider the numerical errors obtained by the different sampling methods. In Figure 3, the boxplot of the numerical errors $AX - b$ using the proposed method and the standard approaches are shown for one hundred simulations. The data dimension N is fixed to 50 and the dimension of the set of constraints $n = 8$. On our experiments, despite numerical errors are really small, the proposed method is always numerically more precise. Let us mention that the MUR is more stable than eigen and Cholesky decompositions with and without truncation. In the Cholesky decomposition, a nugget effect of order 10^{-10} has been added for computing the lower triangles L and L_p in order to avoid computation problems with the software and program used (*chol* function in R).

A comparison between the different approaches in terms of computational time is now investigated. The unconditional covariance matrix Γ is the $N \times N$ identity matrix. A variety of settings is considered, the variation of sample size as well as the dimension of the set of constraints $n = N - p$. To do this, let $X \in \mathbb{R}^N$ be a Gaussian vector with mean μ and covariance matrix Γ generated as previously. The goal is to simulate $X \sim \mathcal{N}(\mu, \Gamma)$ given $AX = b$.

The computation time of simulating hyperplane truncated MVN distributions averaged over ten random trials is shown in Figure 4. The number of hyperplane constraints and data dimension N are fixed at $n = 300$ and $N = 500$ respectively. The number of samples increases from 5,000 to 50,000. It is clear that, when the number of samples is high, the proposed approach has a clear advantage over other naive approaches (left panel). In the right panel of Figure 4, the proposed approach has been compared to the truncated version of Cholesky and eigendecomposition. The MUR has been removed

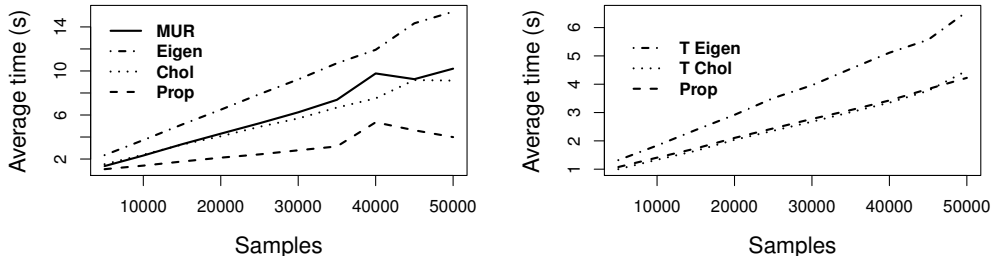


Figure 4: Average time of simulating hyperplane-truncated MVN samples over ten random trials, where the number of constraints and the data dimension are fixed to $n = 300$ and $N = 500$ respectively.

to see clearly the slight difference between the truncated eigen (T Eigen), the truncated Cholesky (T Chol) and the proposed approach. So, in terms of computational time, on this experiment, T Eigen, T Chol and proposed exhibit a good performance.

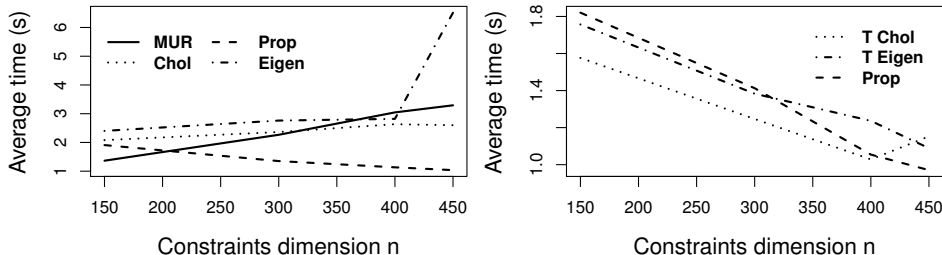


Figure 5: Average time of simulating hyperplane-truncated MVN samples over ten random trials, where the number of constraints n increases.

Now, we investigate the case when the number of constraints increases. In Figure 5, the data dimension N and the samples are fixed to 500 and 10,000 respectively. The dimension of the constraints set is equal to 30%, 60%, 80% and 90% of the data dimension. In the left panel, the proposed approach has been compared to standard naive approaches. As mentioned in [2], when the covariance matrix of X is non-diagonal, the MUR has no advantage compared to Cholesky composition. This fact has been confirmed in Figure 5 (left panel). Contrary to Matheron's update rules, naive Cholesky and naive eigendecomposition, the computation time decreases when the dimension of set of constraints n increases. This is because, we simulate from $p = N - n$ dimensional standard multivariate normal distributions instead of the full

N -dimensional MVN distribution. In the right panel of Figure 5, the proposed approach has been compared to the truncated version of Cholesky and eigendecomposition. The MUR has been removed to see clearly the slight difference between the truncated eigen (T Eigen), the truncated Cholesky (T Chol) and the proposed approach. Again, the proposed approach outperforms the naive approaches in terms of computational time. However, in this experiment, there is no large difference by comparison with the truncated version of eigendecomposition and Cholesky.

4 Conclusion

In this paper, simulating Gaussian vectors truncated on the intersection of a set of hyperplanes is investigated. First, the standard naive approaches and their truncated version are briefly recalled. Second, a new approach is developed: we provide an exact method based on constructing the projection of an orthonormal basis on the set of constraints. This is done using a new formula for simulating conditional Gaussian vectors. The performance of the proposed algorithm has been investigated using numerical simulations. The proposed model gives precise results in terms of stability compared to standard approaches. It is specially efficient when the number of samples and/or the dimension of constraints n is high as well as when the unconditional matrix Γ admits some special structure that makes it easy to invert.

References

- [1] C. Chevalier, X. Emery, and D. Ginsbourger. Fast update of conditional simulation ensembles. *Math. Geosc.*, 47(7):771–789, 2015.
- [2] Y. Cong, B. Chen, and M. Zhou. Fast simulation of hyperplane-truncated multivariate normal distributions. *Bayesian Anal.*, 12(4):1017 – 1037, 2017.
- [3] A. Genz and F. Bretz. *Computation of multivariate normal and t probabilities*, volume 195. Springer Science & Business Media, 2009.
- [4] A. Genz and K.-S. Kwong. Numerical evaluation of singular multivariate normal distributions. *J. Stat. Comput. Simul.*, 68(1):1–21, 2000.
- [5] Y. Hoffman and E. Ribak. Constrained realizations of Gaussian fields: A simple algorithm. *The Astrophysical Journal*, 380:L5, October 1991.

- [6] A. G. Journel and C. J. Huijbregts. *Mining geostatistics*. Academic Press, 1976.
- [7] H. Maatouk and X. Bay. A new rejection sampling method for truncated multivariate Gaussian random variables restricted to convex sets. In Ronald Cools and Dirk Nuyens, editors, *Monte Carlo and Quasi-Monte Carlo Methods*, pages 521–530. Springer International Publishing, Cham, 2016.
- [8] J. T. Wilson, V. Borovitskiy, A. Terenin, P. Mostowsky, and M. P. Deisenroth. Pathwise conditioning of Gaussian processes. *J. Mach. Learn. Res.*, 22(105):1–47, 2021.