Fast and Exact Simulation of Multivariate Normal and Wishart Random Variables with Box Constraints

Hillary Koch∗
Department of Statistics, Pennsylvania State University and
Gregory P. Bopp†
Department of Statistics, Pennsylvania State University

July 2, 2019

Abstract
Models which include domain constraints occur in myriad contexts such as econometrics, genomics, and environmetrics, though simulating from constrained distributions can be computationally expensive. In particular, repeated sampling from constrained distributions is a common task in Bayesian inferential methods, where coping with these constraints can cause troublesome computational burden. Here, we introduce computationally efficient methods to make exact and independent draws from both the multivariate normal and Wishart distributions with box constraints. In both cases, these variables are sampled using a direct algorithm. By substantially reducing computing time, these new algorithms improve the feasibility of Monte Carlo-based inference for box-constrained, multivariate normal and Wishart distributions.

Keywords: box constraints, Monte Carlo, truncated

∗The author gratefully acknowledges NIH pre-doctoral fellowship F31HG010574.
†The author gratefully acknowledges the NSF (DMS-1752280).
1 Introduction

Multivariate normal (MVN) and Wishart random variables with box constraints arise frequently in practice. The truncated MVN distribution appears in many Bayesian models, such as Bayesian linear regression (Geweke 1996, Kato & Hoijtink 2006), multinomial probit (Albert & Chib 1993) and logit (O’Brien & Dunson 2004) models, isotonic regression (Neelon & Dunson 2004) and non-parametrics (Kottas et al. 2005). Structured covariance and precision matrices, which are typically modelled using, respectively, the Wishart and inverse-Wishart distributions, likewise show up in many contexts including Gaussian graphical models (Snoussi & Mohammad-Djafari 2007, Dobra et al. 2011, Mohammadi et al. 2015), MVN hierarchical models (Everson & Morris 2000a), and longitudinal data analysis (Daniels & Pourahmadi 2002, Quintana et al. 2016).

A naive approach to drawing from a constrained distribution whose unconstrained distribution is easy to sample from is with rejection sampling. Rejection sampling retains draws from corresponding unconstrained distributions that satisfy the desired constraints. While this may be satisfactory in low dimensions or when the truncated region is small, it becomes computationally impractical as the dimension of the problem grows and rejection sampling becomes highly inefficient. Because many modern Bayesian inferential methods involve repeated sampling, inefficient sampling methods can render such approaches computationally prohibitive.

Consequently, much effort has been given to studying sampling methods for the truncated MVN distribution, most notably using a Gibbs sampling approach (Geweke 1991, Gelfand et al. 1992, Kotecha & Djurić 1999), though other Monte Carlo-based alternatives have also been studied (Pakman & Paninski 2014, Li & Ghosh 2015, Cong et al. 2017). For structured matrices, the task of sampling from the matrix-valued G-Wishart distribution (Roverato 2002, Atay-Kayis & Massam 2005, Lenkoski 2013), which fixes certain off-diagonal elements to zero, has garnered much attention. Everson & Morris (2000b) develop an approach to simulating Wishart matrices with eigenvalue constraints. However, an efficient procedure for generating matrices with off-diagonal box constraints (e.g., sign constraints on the covariance terms), which occur in contexts such as multivariate meta-
analysis (Hurtado Rua et al. 2015), covariance selection modeling (Wong et al. 2003), and mixture modeling (Ingrassia & Rocci 2007, Li et al. 2011), has not been developed.

In this article, we propose two algorithms for making exact and independent draws – one for the truncated MVN distribution, and another for the Wishart distribution with off-diagonal constraints. These algorithms can be used to impose box inequality constraints of the form

\[ r_i \leq x_i \leq s_i \quad \text{for} \quad i \in \{1, \ldots, D\} \tag{1} \]

and

\[ r_{pm} \leq \Sigma_{pm} \leq s_{pm} \quad \text{for} \quad p \neq m, \text{ and } p, m \in \{1, \ldots, D\} \tag{2} \]

where \( r \) and \( s \) are constants, \( x_i \) is the \( i \)th element of \( D \)-dimensional MVN vector \( \mathbf{x} \) and \( \Sigma_{pm} \) is the \((p,m)\)th off-diagonal element of \( D \times D \) Wishart matrix \( \Sigma \). We note that Equations 1 and 2 also encompass equality constraints (e.g., \( \Sigma_{pm} = 0 \) for some \( p, m \)). It is worth emphasizing that, while our algorithms are useful in many contexts, they cannot be used to impose arbitrary polytope constraints.

We demonstrate through simulations that our algorithm for simulating from the truncated MVN outpaces the current state-of-the-art sampling technique. For the Wishart, on the other hand, our simulation scheme provides a tool for sampling from a constrained distribution, which has until now remained unavailable.

This article is outlined as follows. In Section 2 we review standard algorithms for simulating unconstrained MVN and Wishart random variables. In Section 3 we show how these algorithms can be modified to accommodate box parameter constraints. Finally, we demonstrate our algorithms and numerically evaluate their computational complexity with simulations in Section 4.

## 2 Background

Our sampling scheme builds directly upon standard procedures for simulating MVN and Wishart random variables. Thus, to clarify the steps in our sampling scheme, we briefly review the simulation procedures for unconstrained MVN and Wishart random variables.
Let $\mathbf{x} \sim \mathcal{N}_D(\boldsymbol{\mu}, \Sigma)$, a $D$-dimensional MVN distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$. To draw a random sample $\mathbf{x}$, one computes the lower Cholesky factor $L$ of $\Sigma$, and sets

$$\mathbf{x} = L\mathbf{z} + \boldsymbol{\mu}$$  \hspace{1cm} (3)

where $\mathbf{z} = (z_1, \ldots, z_D)^T$ and $z_i \overset{iid}{\sim} \mathcal{N}_1(0,1)$ for $i \in \{1, \ldots, D\}$.

Wijsman [1957] proposed a method for sampling Wishart random matrices based on the Bartlett decomposition. Consider $\Sigma \sim \mathcal{W}_D(\nu, \Psi)$, a Wishart distribution parameterized by $D \times D$ positive definite scale matrix $\Psi$ and degrees of freedom $\nu$, $\nu > D - 1$. To draw a random sample from this distribution, one computes $U$, the upper-triangular Cholesky factor of $\Psi$, and simulates a $D \times D$ lower-triangular matrix $A$ as

$$A = \begin{pmatrix}
c_1 & 0 & \ldots & 0 \\
z_{21} & c_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
z_{D1} & z_{D2} & \ldots & c_D
\end{pmatrix}$$

where $c_i^2 \sim \chi_\nu^2_{\nu-i+1}$ and $z_{ij} \sim \mathcal{N}(0,1)$. Then

$$\Sigma = U^TAA^TU$$  \hspace{1cm} (4)

is a Wishart-distributed random matrix.

### 3 Direct sampling schemes

In what follows, we adapt these simulation procedures to establish direct, sequential algorithms which generate draws from constrained MVN and Wishart distributions. Both algorithms have a similar flavor, and we make the same heuristic argument for their validity in the main text. Formal proofs that the algorithms produce draws from their target densities are provided in the Appendix.
3.1 Simulating from a Multivariate Truncated Normal Distribution

Our approach is to modify the MVN simulation procedure to incorporate box constraints. Equation 3 contains just three terms to work with: $L$, $z$, and $\mu$. We cannot tamper with $L$ or $\mu$ as this would destroy the desired structure of the MVN distribution that we wish to preserve in its truncated analog, thus we turn to the random vector $z$.

The general idea of our algorithm for simulating truncated MVN data is to translate constraints on random vector $x$ into a more manageable task of constraining the elements of $z$ individually. To do this, first expand Equation 3 to obtain the following system of equations

$$
L_{11}z_1 + \mu_1 = x_1 \\
L_{21}z_1 + L_{22}z_2 + \mu_2 = x_2 \\
\vdots \\
L_{D1}z_1 + \ldots + L_{DD}z_D + \mu_D = x_D
$$

(5)

Exploitation of this decomposition has in fact been discussed by several others in the context of approximating the normalizing constant or cumulative distribution of the MVN distribution (Genz 1992, Botev 2017). Complimentary to this, we seek to make explicit the utility of these equations in the context of simulating truncated MVN data.

The decomposition reveals two key facts: First, due to the lower-triangular form of $L$, it is clear that each $x_i$ is a linear combination of all $z_j$ where $j \leq i$. Second, $x_i$ is monotone non-decreasing in $z_i$ for all $i$, since $L_{ii} > 0$, $\forall i$. The former reveals that if it is possible to find a value for $z_1$ which guarantees that $x_1$ satisfies its constraint, then conditional on this value, it is possible to find a value for $z_2$ such that $x_2$ satisfies its constraint; this logic iterates across the entire random vector $z$. The latter’s relevance is easily illustrated by the following example.

Consider the case of sampling $x$ from a truncated MVN distribution with support on the positive orthant. The monotonicity in $z_i$ of the $i$th equation in (5) and the positivity of the diagonal terms of $L$ ensures that for fixed $z_j$, $j < i$, and some truncation point $a_i$, we will have $x_i < 0$ whenever $z_i < a_i$. Therefore, by simulating $z_i \sim \text{TN}(0, 1, a_i, +\infty)$,
$x_i$ will satisfy the target constraints, where $\text{TN}(\mu, \sigma^2, a, b)$ denotes a truncated univariate normal distribution whose non-truncated analog has mean $\mu$, variance $\sigma^2$, and whose lower and upper truncation points are $a$ and $b$.

It follows from these two facts that we can re-express the constraints on $x$ in terms of the constraints on $z$. First, let the constraints on the MVN, described in Equation 1, be stacked into two vectors of lower and upper bounds on the truncated MVN; term these vectors $r = (r_1, \ldots, r_D)^T$ and $s = (s_1, \ldots, s_D)^T$, respectively. The lower and upper truncation points $a_i$ and $b_i$ for each $z_i$ are found in sequence by solving the $i$th equation in the system of equations (5) for $z_i$ and conditioning on $z_j$, $j < i$. Solving for $z_i$ in these inequalities gives

$$
\begin{align*}
\frac{r_1 - \mu_1}{L_{11}} \leq z_1 \leq \frac{s_1 - \mu_1}{L_{11}} \\
\frac{r_2 - \mu_2 - L_{21}z_1}{L_{22}} \leq z_2 \leq \frac{s_2 - \mu_2 - L_{21}z_1}{L_{22}} \\
\vdots \\
\frac{r_D - \mu_D - \left[ \sum_{j=1}^{D-1} L_{Dj}z_j \right]}{L_{DD}} \leq z_D \leq \frac{s_D - \mu_D - \left[ \sum_{j=1}^{D-1} L_{Dj}z_j \right]}{L_{DD}}
\end{align*}
$$

(6)

such that we can obtain

$$
z_i \sim \text{TN}(0, 1, a_i, b_i), \quad i = 1, \ldots, D
$$

(7)

where $a_i$ and $b_i$ are the left- and right-hand sides of the $i$th inequality in Equation 6. This gives rise to a simple, sequential, and approximation-free method for simulating iid draws from a truncated MVN distribution that guarantees the specified truncations while also preserving the desired underlying MVN distribution’s structure. This is made explicit in Algorithm 1.

### 3.2 Simulating from a Constrained Wishart Distribution

The proposed algorithm for sampling from a constrained Wishart distribution follows the same logic as for the truncated MVN, though the necessary calculations are more involved. In Equation 4, changing the matrix $U$ or the diagonal of $A$ used to generate the Wishart
Algorithm 1: Sampling $D$-dimensional mean vector from a truncated MVN distribution

1: Define fixed parameters $\Sigma$, $\mu$, $r$ and $s$, where $r$ and $s$ are vectors of lower and upper truncation points in each of the $D$ dimensions of a MVN distribution.

2: Compute $L$, the lower-triangular Cholesky factor of $\Sigma$

3: for $i$ in 1 to $D$ do

4: if $i = 1$ then

5: $z_1 \sim \text{TN}(0, 1, \frac{r_1 - \mu_1}{L_{11}}, \frac{s_1 - \mu_1}{L_{11}})$

6: else

7: $z_i \sim \text{TN}(0, 1, \frac{r_i - \mu_i - \sum_{j=1}^{i-1} L_{ij} z_j}{L_{ii}}, \frac{s_i - \mu_i - \sum_{j=1}^{i-1} L_{ij} z_j}{L_{ii}})$

8: return $Lz + \mu$

random variable will destroy the desired distributional covariance structure within the non-truncated region, and may even result in a sampled matrix which is not positive-definite.

Looking thus to the off-diagonal elements of $A$, we observe that for $m < p$, any element $\Sigma_{pm}$ and $\Sigma_{mp}$ in the matrix $\Sigma = U^TAA^T U$ can be expanded as

$$\Sigma_{pm} = \Sigma_{mp} = \begin{cases} U_{i1} A_{11} (\sum_{i=1}^{p} U_{ip} A_{ip}), & \text{for } m = 1 \\ \sum_{k=1}^{m} U_{km} \left\{ \sum_{j=1}^{k} A_{kj} \left[ \sum_{i=j}^{p} (A_{ij} U_{ip}) \right] \right\}, & \text{for } m > 1 \end{cases}. \tag{8}$$

This representation shows that, due to the triangular structure of both $U$ and $A$, each off-diagonal element $\Sigma_{pm}$ computed using Equation 8 depends only on certain off-diagonal elements of $A$. This mirrors how, in Equation 5 for a truncated MVN, each random element $x_i$ only depended on $z_j$ for $j \leq i$. Further, each $\Sigma_{pm}$ is a monotonic function of $A_{pm}$. We thus arrive at a natural, prescribed order to simulate the off-diagonal elements of $A$, beginning in the top-left corner of the matrix, and moving progressively towards the bottom-right in a row-wise manner (that is, $A_{21}, A_{31}, A_{32}, A_{41}, A_{42}, A_{43}, \ldots, A_{D,D-2}, A_{D,D-1}$).

Again as with the truncated MVN, in order to satisfy the required constraints, the off-diagonal elements of $A$ can be simulated from a truncated standard normal distribution. For the $(p,m)$th off-diagonal element, $m < p$, solving Equation 8 for $A_{pm}$ reveals the equations for the lower and upper truncation points on the univariate standard normal
distribution from which to sample. After some algebra, $A_{pm}$ is expressed as

$$A_{pm} = \begin{cases} \frac{\Sigma_{p1} - U_{11}A_{11}(\Sigma_{p1-1}U_{ip}A_{i1})}{U_{pp}U_{11}A_{11}} & \text{for } m = 1 \\ \frac{\Sigma_{pm} - \sum_{k=1}^{m-1} U_{km} \left\{ \sum_{j=1}^{k} A_{kj} \left[ \sum_{i=j}^{p} (A_{ij}U_{ip}) \right] \right\}}{U_{pp}U_{mm}A_{mm}} & \\ - \frac{\sum_{i=m}^{p-1} A_{im}U_{ip}}{U_{pp}} & \text{for } m > 1 \end{cases}$$

(9)

Similar to Equation 6, this formula for $A_{pm}$ is used to find upper and lower truncation points for the univariate standard normal random variables which comprise the off-diagonal elements of $A$ such that the resulting Wishart matrix satisfies the desired constraints. The off-diagonal element $A_{21}$ must be simulated first; conditioning on the observation of $A_{21}$, a satisfactory value for $A_{31}$ can be obtained, followed by $A_{32}, A_{41},$ and so forth until matrix $A$ has been populated. Details are made explicit in Algorithm 2.

Algorithm 2: Sampling $D \times D$ matrix from a constrained Wishart distribution

1: Define fixed parameters $\Psi$, $\nu$, $R$, and $S$, where $R$ and $S$ are symmetric $D \times D$ matrices of lower and upper truncation points for each of the off-diagonal elements of a Wishart distribution

2: Compute $U$, the upper-triangular Cholesky factor of $\Psi$

3: for $i$ in $1$ to $D$

4: $A_{ii} \sim \sqrt{\chi^2_{\nu_i-1}}$

5: for $p$ in $2$ to $D$

6: for $m$ in $1$ to $p-1$

7: $a_{pm} \leftarrow$ solution to Equation 9 setting $\Sigma_{pm} = R_{pm}$

8: $b_{pm} \leftarrow$ solution to Equation 9 setting $\Sigma_{pm} = S_{pm}$

9: $A_{pm} \sim \text{TN}(0, 1, a_{pm}, b_{pm})$

10: return $U^TAA^TU$
4 Simulations

We perform simulations to assess the accuracy and speed of our algorithms. To visualize how simulated data from Algorithm 1 compare to data from their unconstrained counterpart, we first make draws of a four-dimensional random variable \( \mathbf{x} \) coming from a truncated MVN distribution with parameters

\[
\mu = (1, 2, -2, -1), \quad \Sigma = \begin{pmatrix}
2 & 0.3 & -0.6 & -0.8 \\
0.3 & 1.5 & -0.75 & -0.1 \\
-0.6 & -0.75 & 1.5 & 0.4 \\
-0.8 & -0.1 & 0.4 & 2
\end{pmatrix}
\]

with \( r = (0, 0, -\infty, -\infty) \) and \( s = (\infty, \infty, 0, 0) \). We generated 100,000 realizations of this random variable using our approach, as well as 100,000 realizations from the analogous unconstrained MVN distribution. The results show (Figure 1) that our truncated MVN draws reflect the ordinary MVN’s distributional structure within the truncated distribution’s region of support.

For the constrained Wishart, we let \( \Psi = \Sigma \) from Equation 10, and set \( \nu = 25 \). We define constraints

\[
R_{mp} = R_{pm} = \begin{cases}
0, & \text{for } \Sigma_{mp} > 0 \\
-\infty, & \text{for } \Sigma_{mp} < 0
\end{cases}, \quad S_{mp} = S_{pm} = \begin{cases}
\infty, & \text{for } \Sigma_{mp} > 0 \\
0, & \text{for } \Sigma_{mp} < 0
\end{cases}.
\]

where \( m < p \). We make 100,000 draws from the constrained Wishart distribution using Algorithm 2. As seen in Figure 2, our draws preserve the structure observed in the ordinary Wishart distribution within the constrained distribution’s region of support.

We tested the speed of our exact truncated MVN sampler at up to 500 dimensions. For each dimension size \( D \), we let the mean of the MVN equal 1 in each dimension, and generated \( \Sigma \sim \mathcal{W}_D(D + 100, I)/(D + 100) \). We restricted all draws to the positive orthant. For each simulation setting, we generated 100 draws from the truncated MVN distribution, and considered the average performance across all iterations. For comparison, we followed this same procedure with the Gibbs sampling approach of Geweke (1991), which is implemented in the R package \texttt{tmvtnorm} (Wilhelm & Manjunath 2015) and the non-truncated case as
Figure 1: (Non-)Truncated MVN draws plotted in pairs of dimensions. The columns correspond to given pairs. Row A shows truncated MVN samples generated with Algorithm 1 while row B shows the corresponding ordinary MVN samples. Comparing across rows, we see that the truncated MVN data share the structure of their non-truncated counterpart data.
Figure 2: (Un)constrained Wishart draws plotted in pairs of off-diagonal elements. Row A shows constrained Wishart samples generated with Algorithm 2, while row B shows the corresponding ordinary Wishart samples. Comparing within columns, one observes that data generated with Algorithm 2 match their unconstrained counterparts.

Figure 3: Average performance on the $\log_{10}$ scale across 100 replications of our A) truncated MVN and B) constrained Wishart simulation algorithms against comparable alternatives.
a baseline measure. Results in Figure 3A show that the proposed algorithm out-performs
the alternative.

We proceeded similarly with the evaluation of our exact sampler for the constrained
Wishart distribution. As a baseline measure, we compared the speed of our method against
an unconstrained Wishart sampler. Since we are unaware of a direct comparison (that is,
an algorithm for simulating random Wishart matrices satisfying box constraints on the off-
диagonals), we compared against the G-Wishart sampler of Lenkoski (2013), implemented
in the R package BDgraph (Mohammadi & Ernst 2019). In Bayesian Gaussian graphical
models, the G-Wishart distribution is the conjugate prior to a precision matrix which en-
codes conditional independence among nodes in a graph (Roverato 2002, Atay-Kayis &
Massam 2005). This requires that the \((m, p)\)th element of this matrix be fixed to zero
when there is no edge connecting nodes \(m\) and \(p\). Like the work by Lenkoski (2013), given
a node adjacency matrix, one can generate a matrix using Algorithm 2 with the neces-
sary elements correctly fixed at zero. We note, however, that the comparison between
these two algorithms is not the most direct, in that the G-Wishart sampler produces ran-
dom matrices in the form of an inverse-Wishart, while our approach produces constrained
Wishart-distributed matrices. There is no simple extension of our algorithm to produce
inverse-Wishart matrices with off-diagonal constraints.

For each simulation at a given dimension, we generated a \(D\)--dimensional adjacency
matrix \(E\) establishing the connectedness of the graph, with the probability of an edge con-
necting two nodes equal to 0.75. We then drew \(\Psi \sim \mathcal{W}_D(2D, I)\), and made 100 G-Wishart
proposals with degrees of freedom \(2D\), scale matrix \(\Psi\), and adjacency matrix \(E\). The av-
erage performance of our algorithm, the direct G-Wishart sampler of Lenkoski (2013), and
the unconstrained Wishart distribution are displayed in Figure 3B. It is clear that simu-
lating such a constrained matrix is a computationally demanding task, but our approach
yields speeds competitive with the implementation in the BDgraph package. Moreover, our
algorithm opens up the possibility of efficiently sampling Wishart covariance matrices with
box constraints on the off-diagonal matrix elements.
5 Discussion

We have proposed two simple, sequential algorithms to simulate from the MVN and Wishart distributions with box inequality constraints. These algorithms are attractive because they yield independent draws, are easy to implement, and are highly scalable. Algorithm 1 introduces a direct approach to generating independent truncated MVN samples, which has computational complexity on the order of ordinary Cholesky factor-based MVN sampling methods. Algorithm 2 introduces a novel means to simulate Wishart matrices with off-diagonal constraints, a task which has heretofore been practically infeasible. Moreover, the proposed procedures have the potential to be modified or extended to simulate from constrained distributions not addressed here, such as truncated scale mixtures of the MVN like the truncated multivariate Student’s $t$, Cauchy, and Laplace distributions, as well as the constrained matrix normal distribution. The algorithms presented here have been implemented in an R package which will be made available on CRAN.

APPENDIX

A Technical Proofs

Proposition 1: Algorithm 1 produces draws $x \sim \text{truncated MVN} (\mu, \Sigma, r, s)$

Proof. Algorithm 1 sequentially simulates $x_1, x_2, \ldots, x_D$ conditionally as $x_1, x_2 | x_1, x_3 | x_1, x_2$ etc. (i.e. $x_j | x_1, \ldots, x_{j-1}$, for $j = 2, \ldots, D$) to construct a sample that jointly follows the target truncated multivariate normal distribution. We show that Algorithm 1 produces samples from the desired distribution by building up these conditional distributions, and making a connection to those of the ordinary multivariate normal distribution. Throughout, we assume without loss of generality that the location parameter $\mu = 0$. We begin by examining the non-truncated case.

Letting $\tilde{z}_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 1), i = 1, \ldots, D$, and $L$ denote the lower Cholesky factor of $\Sigma$, s.t. $\Sigma = LL^T$, then for $\tilde{x} = L\tilde{z}$, $\tilde{x} \sim \text{MVN}(0, \Sigma)$, with density

$$f(\tilde{x}) \propto \exp\left\{-\frac{1}{2} \tilde{x}\Sigma^{-1}\tilde{x}\right\}, \quad \tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_D)^T.$$ (12)
The multivariate transformation $\tilde{x} = L\tilde{z}$ can equivalently be re-expressed in terms of $D$ univariate transformations. Define $g_1(\tilde{z}_1) = L_{11}\tilde{z}_1$, (i.e., $g_1^{-1}(\tilde{x}_1) = \frac{\tilde{x}_1}{L_{11}}$) and conditional on $\tilde{z}_i$ for $i < j$, define $g_j(\tilde{z}_j; \tilde{z}_1, \ldots, \tilde{z}_{j-1}) = L_{jj}\tilde{z}_j + \sum_{i=1}^{j-1} L_{ji}\tilde{z}_i$, (i.e., $g_j^{-1}(\tilde{x}_j; \tilde{z}_1, \ldots, \tilde{z}_{j-1}) = \tilde{x}_j + \sum_{i=1}^{j-1} \frac{L_{ji}}{L_{jj}}\tilde{z}_i$) for $j = 2, \ldots, D$. For simplicity of notation, we express $g_j(\tilde{z}_j; \tilde{z}_1, \ldots, \tilde{z}_{j-1})$ as $g_j(\tilde{z}_j)$ and $g_j^{-1}(\tilde{x}_j; \tilde{z}_1, \ldots, \tilde{z}_{j-1})$ as $g_j^{-1}(\tilde{x}_j)$. Then for $\tilde{x}_j = g_j(\tilde{z}_j)$, by straightforward transformation of variables,

$$f(\tilde{x}_j) \propto \exp\left\{ -\frac{g_j^{-1}(\tilde{x}_j)^2}{2} \right\}$$

$$f(\tilde{x}_j | \tilde{x}_1, \ldots, \tilde{x}_{j-1}) \propto \exp\left\{ -\frac{g_j^{-1}(\tilde{x}_j)^2}{2} \right\} \text{ for } j = 2, \ldots, D,$$  

(13)

since $\left| \frac{dg_j^{-1}(\tilde{x}_j)}{d\tilde{x}_j} \right| = \left| \frac{1}{L_{jj}} \right| \forall j$.

The univariate $\text{TN}(0, 1, a, b)$ density has the form $f(z) \propto \exp\left\{ -\frac{z^2}{2} \right\} \mathbb{1}\{ a \leq z \leq b \}$. Take $z_1 \sim \text{TN}(0, 1, a_1, b_1)$ and $z_j | z_1, \ldots, z_{j-1} \sim \text{TN}(0, 1, a_j, b_j)$, $j = 2, \ldots, D$, where $a_j$ and $b_j$ are both functions of $z_1, \ldots, z_{j-1}$ according to the bounds in Equation 6. Set $x_i = g_i(z_i)$ as in the unconstrained case. Again, by straightforward transformation of variables, the conditional densities are

$$f(x_1) \propto \exp\left\{ -\frac{g_1^{-1}(x_1)^2}{2} \right\} \mathbb{1}\{ g_1(a_1) \leq x_1 \leq g_1(b_1) \}$$

$$f(x_j | x_1, \ldots, x_{j-1}) \propto \exp\left\{ -\frac{g_j^{-1}(x_j)^2}{2} \right\} \mathbb{1}\{ g_j(a_j) \leq x_j \leq g_j(b_j) \} \text{ for } j = 2, \ldots, D,$$  

(14)

which are identical to the densities in Equation 13 up to the indicator functions. Hence, by the equivalence between Equations 12 and 13, taking the product of conditional densities in Equation 14 gives the desired truncated multivariate normal density

$$f(x) \propto \exp\left\{ -\frac{1}{2}x\Sigma^{-1}x \right\} \mathbb{1}\{ r \leq x \leq s \}, \quad x = (x_1, \ldots, x_D)^T$$

(15)

where the inequalities apply componentwise.

□

**Proposition 2**: Algorithm 2 produces draws $\Sigma \sim \text{constrained Wishart}(\nu, \Psi, R, S)$
Proof. Proof of the validity of Algorithm 2 follows the same line of argument as above. The algorithm simulates $A_{ii}, i = 1, \ldots, D$ independently followed by $A_{21}, A_{31}, A_{32}, \ldots, A_{D,D-1}$ (in top-down, row-wise order) conditionally as $[A_{ij} | \{A_{kl} : (i > k) \cup (i = k \cap j > l)\}]$. We build up these conditional distributions to show that Algorithm 2 produces draws from the target constrained Wishart distribution by likening the product of conditional distributions to that of the joint distribution of the matrix elements in the unconstrained case. Throughout, we assume without loss of generality that $\Psi$ is the identity matrix $I$.

We begin by considering the unconstrained Wishart distribution. Let $	ilde{A}_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$, $i \in \{2, \ldots, D\}$, $j \in \{1, \ldots, i-1\}$, $A_{dd} \sim \chi_{{\nu} - d + 1}^2$, $d \in \{1, \ldots, D\}$, all independent. Since all elements of $	ilde{A}$ are independent, their joint density is

$$f(\text{diag}(\tilde{A})_d, \tilde{A}_{ij}, d \in \{1, \ldots, D\}, i \in \{2, \ldots, D\}, j \in \{1, \ldots, i-1\}) \propto \exp\left\{ -\frac{1}{2} \sum_{i=2}^{D} \sum_{j=1}^{i-1} \tilde{A}_{ij}^2 \right\} \prod_{d=1}^{D} \left( \text{diag}(\tilde{A})_d^{(\nu-d-1)/2} \exp\left\{ -\frac{1}{2} \text{diag}(\tilde{A})_d \right\} \right),$$

(16)

where $\text{diag}(A)$ indicates the diagonal of matrix $A$ and $\text{diag}(A)_d$ refers to its $d^{th}$ element. Treating this joint density as a starting point, Kshirsagar (1959) demonstrates that the transformation $\tilde{\Sigma} = \tilde{A} \tilde{A}^T$ follows a $\mathcal{W}_D(\nu, I)$ distribution (Note that since $\Psi = I$, the upper Cholesky factor $U$ of $\Psi$ is also equal to $I$).

The Wishart density can equivalently be hierarchically expressed as the product of the conditional densities of elements $\Sigma_{ij}$, where the conditioning occurs in top-down, row-wise order:

$$f(\tilde{\Sigma})[\Sigma_{ij}, i \in \{1, \ldots, D\}, j \in \{1, \ldots, i\}] = f_{11}^{(\tilde{\Sigma})}[\tilde{\Sigma}_{11}] \times f_{21}^{(\tilde{\Sigma})}[\tilde{\Sigma}_{21} | \tilde{\Sigma}_{11}] \times f_{22}^{(\tilde{\Sigma})}[\tilde{\Sigma}_{22} | \tilde{\Sigma}_{21}, \tilde{\Sigma}_{11}] \times f_{31}^{(\tilde{\Sigma})}[\tilde{\Sigma}_{31} | \tilde{\Sigma}_{22}, \tilde{\Sigma}_{21}, \tilde{\Sigma}_{11}] \times \cdots \times f_{DD}^{(\tilde{\Sigma})}[\tilde{\Sigma}_{DD} | \tilde{A}_{ij}, i \in \{1, \ldots, D\}, j \in \{1, \ldots, i-1\}] ]$$

(17)

This hierarchical representation is useful because $\tilde{\Sigma}$ can also be expressed in terms of $\binom{D+1}{2}$ univariate transformations. Define for $m \leq p$

$$g_{pm}[\tilde{A}_{pm}; \tilde{A}_{mi}, \tilde{A}_{pj}, i = 1, \ldots, m, j = 1, \ldots, m - 1] = \sum_{k=1}^{m} \tilde{A}_{mk} \tilde{A}_{pk}$$

(18)
such that

\[ g_{pm}^{-1}[\tilde{\Sigma}_{pm}, \tilde{A}_{mi}, \tilde{A}_{pj}, i = 1, \ldots, m, j = 1, \ldots, m-1] = \begin{cases} \tilde{\Sigma}_{11}^{1/2}, & \text{for } 1 = m = p, \\ \tilde{\Sigma}_{pl}/\tilde{A}_{11}, & \text{for } m = 1, p > m. \end{cases} \quad (19) \]

Again for simplicity of notation, we refer to the functions in Equations 18 and 19 as \( g_{pm}(\tilde{A}_{pm}) \) and \( g_{pm}^{-1}(\tilde{\Sigma}_{pm}) \). In terms of densities of the constituent \( \tilde{A} \) terms \( f^{(A)}_{pm} \), each of the conditional densities \( f^{(\Sigma)}_{pm} \) in Equation 17 is of the form

\[ f^{(\Sigma)}_{pm}[\tilde{\Sigma}_{pm} \mid \{\tilde{A}_{kl} : (k < p) \cup (k = p \cap l < m)\}] = \begin{cases} f^{(A)}_{11}[g_{11}^{-1}(\tilde{\Sigma}_{11})] \cdot \frac{1}{2} \tilde{\Sigma}_{11}^{-1/2}, & \text{for } 1 = m = p \\ f^{(A)}_{pm}[g_{pm}^{-1}(\tilde{\Sigma}_{pm}) \mid \{\tilde{A}_{kl} : (k < p) \cup (k = p \cap l < m)\}] \cdot \frac{1}{\tilde{A}_{mm} - \tilde{A}_{mm}}, & \text{for } m < p, \\ f^{(A)}_{mm}[g_{mm}^{-1}(\tilde{\Sigma}_{mm}) \mid \{\tilde{A}_{kl} : (k < m) \cup (k = m \cap l < m)\}] \times \frac{1}{2}(\tilde{\Sigma}_{mm} - \sum_{k=1}^{m-1} \tilde{A}_{mk}^2)^{-1/2}, & \text{for } 1 < m = p \end{cases} \quad (20) \]

In the case of the constrained Wishart distribution, we again proceed in top-down, row-wise order, but now taking \( A_{ij} \{A_{kl} : (k < i) \cup (i = k \cap l < j)\} \sim \text{TN}(0, 1, a_{ij}, b_{ij}), i \in \{2, \ldots, D\}, j \in \{1, \ldots, i-1\} \), where \( a_{ij} \) and \( b_{ij} \) are both functions of \( \{A_{kl} : (k < i) \cup (i = k \cap l < j)\} \) given by \( a_{ij} = g_{ij}^{-1}(R_{ij}) \) and \( b_{ij} = g_{ij}^{-1}(S_{ij}) \). Applying the same transformations in Equation 18 replacing \( \tilde{A} \) terms by \( A \) terms, the corresponding conditional densities are

\[ f^{(\Sigma)}_{pm}[\Sigma_{pm} \mid \{A_{kl} : (k < p) \cup (k = p \cap l < m)\}] = \begin{cases} f^{(A)}_{11}[g_{11}^{-1}(\Sigma_{11})] \cdot \Sigma_{11}^{-1/2}, & \text{for } 1 = m = p \\ f^{(A)}_{pm}[g_{pm}^{-1}(\Sigma_{pm}) \mid \{A_{kl} : (k < p) \cup (k = p \cap l < m)\}] \times 1\{g_{pm}(a_{pm}) \leq \Sigma_{pm} \leq g_{pm}(b_{pm})\}, & \text{for } m < p, \\ f^{(A)}_{mm}[g_{mm}^{-1}(\Sigma_{mm}) \mid \{A_{kl} : (k < m) \cup (k = m \cap l < m)\}] \times (\Sigma_{mm} - \sum_{k=1}^{m-1} A_{mk}^2)^{-1/2}, & \text{for } 1 < m = p \end{cases} \quad (21) \]
hence each of the conditional densities has the same kernel as those in Equation 20 up to indicator functions constraining the domain of the off-diagonal terms. Because of the equivalence of the kernels of each the conditional densities in Equations 20 and 21 up to indicator functions, setting $\Sigma = AA^T$ gives a random matrix with the desired $\Sigma \sim$ constrained Wishart($\nu, I, R, S$) distribution.

References

Albert, J. H. & Chib, S. (1993), ‘Bayesian analysis of binary and polychotomous response data’, *J Am Stat Assoc* **88**(422), 669–679.

Atay-Kayis, A. & Massam, H. (2005), ‘A monte carlo method for computing the marginal likelihood in nondecomposable gaussian graphical models’, *Biometrika* **92**(2), 317–335.

Botev, Z. I. (2017), ‘The normal law under linear restrictions: simulation and estimation via minimax tilting’, *J Roy Stat Soc B* **79**(1), 125–148.

Cong, Y., Chen, B., Zhou, M. et al. (2017), ‘Fast simulation of hyperplane-truncated multivariate normal distributions’, *Bayesian Anal* **12**(4), 1017–1037.

Daniels, M. J. & Pourahmadi, M. (2002), ‘Bayesian analysis of covariance matrices and dynamic models for longitudinal data’, *Biometrika* **89**(3), 553–566.

Dobra, A., Lenkoski, A. & Rodriguez, A. (2011), ‘Bayesian inference for general gaussian graphical models with application to multivariate lattice data’, *J Am Stat Assoc* **106**(496), 1418–1433.

Everson, P. J. & Morris, C. N. (2000a), ‘Inference for multivariate normal hierarchical models’, *J Roy Stat Soc B* **62**(2), 399–412.

Everson, P. J. & Morris, C. N. (2000b), ‘Simulation from wishart distributions with eigenvalue constraints’, *J Comput Graph Stat* **9**(2), 380–389.
Gelfand, A. E., Smith, A. F. & Lee, T.-M. (1992), ‘Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling’, *J Am Stat Assoc* **87**(418), 523–532.

Genz, A. (1992), ‘Numerical computation of multivariate normal probabilities’, *J Comput Graph Stat* **1**(2), 141–149.

Geweke, J. (1991), ‘Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities’.

Geweke, J. F. (1996), Bayesian inference for linear models subject to linear inequality constraints, in ‘Modelling and Prediction Honoring Seymour Geisser’, Springer, pp. 248–263.

Hurtado Rua, S. M., Mazumdar, M. & Strawderman, R. L. (2015), ‘The choice of prior distribution for a covariance matrix in multivariate meta-analysis: a simulation study’, *Stat Med* **34**(30), 4083–4104.

Ingrassia, S. & Rocci, R. (2007), ‘Constrained monotone em algorithms for finite mixture of multivariate gaussians’, *Comput Stat Data An* **51**(11), 5339–5351.

Kato, B. S. & Hoijtink, H. (2006), ‘A bayesian approach to inequality constrained linear mixed models: estimation and model selection’, *Statistical Model* **6**(3), 231–249.

Kotecha, J. H. & Djuric, P. M. (1999), Gibbs sampling approach for generation of truncated multivariate gaussian random variables, in ‘1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)’, Vol. 3, IEEE, pp. 1757–1760.

Kottas, A., Müller, P. & Quintana, F. (2005), ‘Nonparametric bayesian modeling for multivariate ordinal data’, *J Comput Graph Stat* **14**(3), 610–625.

Kshirsagar, A. M. (1959), ‘Bartlett decomposition and wishart distribution’, *Ann Math Stat* **30**(1), 239–241.
Lenkoski, A. (2013), ‘A direct sampler for g-wishart variates’, *arXiv preprint arXiv:1304.1350* 2(1), 119–128.

Li, Q., Brown, J. B., Huang, H., Bickel, P. J. et al. (2011), ‘Measuring reproducibility of high-throughput experiments’, *Ann Appl Stat* 5(3), 1752–1779.

Li, Y. & Ghosh, S. K. (2015), ‘Efficient sampling methods for truncated multivariate normal and student-t distributions subject to linear inequality constraints’, *J Stat Theory Pract* 9(4), 712–732.

Mohammadi, A., Wit, E. C. et al. (2015), ‘Bayesian structure learning in sparse gaussian graphical models’, *Bayesian Anal* 10(1), 109–138.

Mohammadi, R. & Ernst, W. C. (2019), ‘BDgraph: An R package for Bayesian structure learning in graphical models’, *J Stat Softw* 89(3), 1–30.

Neelon, B. & Dunson, D. B. (2004), ‘Bayesian isotonic regression and trend analysis’, *Biometrics* 60(2), 398–406.

O’Brien, S. M. & Dunson, D. B. (2004), ‘Bayesian multivariate logistic regression’, *Biometrics* 60(3), 739–746.

Pakman, A. & Paninski, L. (2014), ‘Exact hamiltonian monte carlo for truncated multivariate gaussians’, *J Comput Graph Stat* 23(2), 518–542.

Quintana, F. A., Johnson, W. O., Waetjen, L. E. & B. Gold, E. (2016), ‘Bayesian nonparametric longitudinal data analysis’, *J Am Stat Assoc* 111(515), 1168–1181.

Roverato, A. (2002), ‘Hyper inverse wishart distribution for non-decomposable graphs and its application to bayesian inference for gaussian graphical models’, *Scand J Stat* 29(3), 391–411.

Snoussi, H. & Mohammad-Djafari, A. (2007), ‘Estimation of structured gaussian mixtures: the inverse em algorithm’, *IEEE T Signal Proces* 55(7), 3185–3191.
Wijsman, R. A. (1957), ‘Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis’, *Ann Math Stat* p. 415-423.

Wilhelm, S. & Manjunath, B. (2015), *tmvtnorm: Truncated Multivariate Normal and Student t Distribution*. R package version 1.4-10.

**URL**: http://CRAN.R-project.org/package=tmvtnorm

Wong, F., Carter, C. K. & Kohn, R. (2003), ‘Efficient estimation of covariance selection models’, *Biometrika* **90**(4), 809–830.