EFFICIENTLY GENERATING INVERSE-WISHART MATRICES AND THEIR CHOLESKY FACTORS

SETH D. AXEN

University of Tübingen, Germany

Abstract. This paper presents a new algorithm for generating random inverse-Wishart matrices that directly generates the Cholesky factor of the matrix without computing the factorization. Whenever parameterized in terms of a precision matrix $\Omega = \Sigma^{-1}$, or its Cholesky factor, instead of a covariance matrix Σ, the new algorithm is more efficient than the current standard algorithm.

1. Introduction

The inverse-Wishart distribution is the conjugate prior for the covariance matrix of a multivariate normal distribution if the mean is known. As a result, it is often used in Bayesian analyses (**gelman˙bayesian˙2013**; **hoff˙multivariate˙2009**). It is derived from the Wishart distribution, which is the conjugate prior for the precision matrix of a multivariate normal distribution if the mean is known and is a key object of interest in random matrix theory.

The standard approach for generating a random draw from the inverse-Wishart distribution, termed an inverse-Wishart matrix, first generates a draw from the Wishart distribution, termed a Wishart matrix, and inverts it. Variants of this approach are implemented in popular Python[1], R[2], and Julia[3] packages. We present an alternative algorithm that, for certain parameterizations of the inverse-Wishart distribution, is more efficient than the standard approach.

Key definitions and notation. Given a positive definite matrix X, the upper Cholesky factor of X is the unique upper triangular matrix U with positive diagonal elements such that $X = U^\top U$. $|X|$ indicates the absolute value of the determinant of X, while $\|X\|$ is the Frobenius norm of X, i.e. $\|X\| = \sqrt{\operatorname{tr}(X^\top X)}$.

E-mail address: `seth@sethaxen.com`.

[1] `scipy.stats.invwishart` (v1.11.3)

[2] `LaplacesDemon::rinvwishart` (v16.1.6), `MCMCpack::riwish` (v1.6-3), and `mniw::riwish` (v1.0.1)

[3] `Distributions.InverseWishart`(v0.25.102)

To perform changes of variables, we will use the exterior product of differentials to derive absolute determinants of Jacobians. Following **muirhead˙aspects˙1982**, $\mathrm{d}X$ refers to the matrix of differentials $\mathrm{d}x_{ij}$ of an $m \times m$ positive definite or upper triangular matrix X, and $(\mathrm{d}X)$ refers to the differential form of the Lebesgue measure on the distinct elements of $\mathrm{d}X$, that is,

$$(\mathrm{d}X) := \bigwedge_{j=1}^{m} \bigwedge_{i=1}^{j} \mathrm{d}x_{ij} = \mathrm{d}x_{11} \wedge \mathrm{d}x_{12} \wedge \mathrm{d}x_{22} \wedge \ldots \wedge \mathrm{d}x_{mm},$$

where $\wedge$ is the exterior (or wedge) product with the antisymmetric property $\mathrm{d}a \wedge \mathrm{d}b = -(\mathrm{d}b \wedge \mathrm{d}a)$ so that $\mathrm{d}a \wedge \mathrm{d}a = 0$.

Change of variables. Let $X = f(Y)$, where f is a diffeomorphic (smooth, one-to-one function) function. Then $(\mathrm{d}X) = |J_f|\,(\mathrm{d}Y)$, where J_f is the Jacobian of f at Y. If $p_X(X)$ is an absolutely continuous density with respect to $(\mathrm{d}X)$, then the change of variables formula gives the corresponding density of Y with respect to $(\mathrm{d}Y)$ as

$$p_Y(Y) = p_X(f(Y))\,|J_f|\,.$$

The Wishart distribution. Let Z be an $m \times n$ matrix whose ith column $Y_i \sim \mathcal{N}(0, \Sigma)$, where the scale matrix Σ is an $m \times m$ positive definite matrix. Then the $m \times m$ positive definite matrix $A = YY^\top$ is distributed according to the Wishart distribution $\mathcal{W}_m(n, \Sigma)$, where n is termed the degrees of freedom. This paper considers only the case where $n > m-1$, so that A is nonsingular. The density of this Wishart distribution with respect to $(\mathrm{d}A)$ is (**muirhead˙aspects˙1982**)

$$p_A(A \mid n, \Sigma) \propto \exp\left(-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}A)\right) |A|^{(n-m-1)/2}\,. \tag{1.1}$$

The normalization constant is known but unimportant for this paper.

The inverse-Wishart distribution. If $B = A^{-1}$, then B follows the inverse-Wishart distribution with degrees of freedom n and scale matrix $\Omega = \Sigma^{-1}$, written $\mathcal{W}_m^{-1}(n, \Omega)$. The density of this distribution with respect to $(\mathrm{d}B)$ is

$$p_B(B \mid n, \Omega) \propto \exp\left(-\frac{1}{2}\operatorname{tr}(\Omega B^{-1})\right) |B|^{-(n+m+1)/2}\,. \tag{1.2}$$

The inverse-Wishart distribution may also be parameterized in terms of Σ. While the two parameterizations are equivalent, the choice of parameterization can affect the efficiency of both density computation and sampling algorithms.

Summary of contributions. This paper's main theoretical contribution is Theorem 3.2, which is applied in Algorithm 4 for efficiently generating a random Cholesky factor of a inverse-Wishart matrix without performing a Cholesky factorization. Algorithm 5 uses this algorithm to efficiently generate an inverse-Wishart matrix. In Section 4 we compare the new and standard algorithm and recommend which to use depending on the parameterization of the inverse-Wishart distribution.

2. The Cholesky parameterizations of the Wishart and inverse-Wishart distributions

It is frequently more convenient to work with Cholesky factors of Wishart- and inverse-Wishart-distributed matrices than the matrices themselves, since from such factors, the matrix inverse and determinant are easily computed. For example, if the inverse-Wishart distribution is used to draw random covariance matrices to parameterize a multivariate normal distribution, then computing the density of the multivariate normal would generally require a Cholesky factor. Thus, it is useful to define the corresponding distributions of Cholesky factors.

The Cholesky-Wishart distribution. Let U_A and U_Σ be the upper Cholesky factors of A and Σ, respectively. Then U_A is distributed according to the Cholesky-Wishart distribution $\mathrm{chol}\mathcal{W}_m(n, U_\Sigma)$. Theorem 2.1 allows us to derive the density of this distribution in Theorem 2.2.

Theorem 2.1 (muirhead˙aspects˙1982). *Let X be an $m \times m$ positive definite matrix and T be its upper Cholesky factor. Then*

$$(\mathrm{d}X) = 2^m \left(\prod_{j=1}^{m} t_{jj}^{m+1-j} \right) (\mathrm{d}T) . \tag{2.1}$$

Theorem 2.2 (Cholesky-Wishart density)**.** *The density of* $\mathrm{chol}\mathcal{W}_m(n, U_\Sigma)$ *with respect to* $(\mathrm{d}U_A)$ *is*

$$p_{U_A}(U_A \mid n, U_\Sigma) \propto \exp\left(-\frac{1}{2} \|U_A U_\Sigma^{-1}\|^2 \right) \prod_{j=1}^{m} (u_A)_{jj}^{n-j} . \tag{2.2}$$

Proof. From Equation (1.1) and Equation (2.1), and using the change of variables formula,

$$p_{U_A}(U_A \mid n, U_\Sigma) \propto \exp\left(-\frac{1}{2} \operatorname{tr}((U_\Sigma^\top U_\Sigma)^{-1} U_A^\top U_A) \right) |U_A|^{n-m-1} \prod_{j=1}^{m} (u_A)_{jj}^{m+1-j}$$

$$= \exp\left(-\frac{1}{2} \operatorname{tr}(U_\Sigma^{-1} U_\Sigma^{-\top} U_A^\top U_A) \right) \prod_{j=1}^{m} (u_A)_{jj}^{n-j} .$$

The cyclic property of the trace implies

$$\operatorname{tr}(U_\Sigma^{-1} U_\Sigma^{-\top} U_A^\top U_A) = \operatorname{tr}((U_A U_\Sigma^{-1})^\top U_A U_\Sigma^{-1}) = \|U_A U_\Sigma^{-1}\|^2 .$$

Substituting this back into the previous equation completes the proof. □

The Cholesky-inverse-Wishart distribution. Similarly, if U_B and U_Ω are the upper Cholesky factors of B and Ω, respectively, then U_B is distributed according to the Cholesky-inverse-Wishart distribution $\mathrm{chol}\mathcal{W}_m^{-1}(n, U_\Omega)$.

Theorem 2.3 (Cholesky-inverse-Wishart density)**.** *The density of* $\mathrm{chol}\mathcal{W}_m^{-1}(n, U_\Omega)$ *with respect to the measure* $(\mathrm{d}U_B)$ *is*

$$p_{U_B}(U_B \mid n, U_\Omega) \propto \exp\left(-\frac{1}{2}\|U_\Omega U_B^{-1}\|^2\right) \prod_{j=1}^{m} (u_B)_{jj}^{-(n+j)}. \tag{2.3}$$

Proof. Starting from Equation (1.2) and Equation (2.1), the proof follows the same steps as that of Theorem 2.2. □

One could alternatively parameterize the Cholesky-inverse-Wishart distribution in terms of U_Σ rather than U_Ω, though this would increase the cost of computing the density.

3. Generating random inverse-Wishart matrices

3.1. The standard approach.

Random Wishart matrices are usually generated using the so-called Bartlett decomposition of the Cholesky factor of a Wishart matrix into the product of two Cholesky factors (**bartlett˙xxtheory˙1934**):

Theorem 3.1 (**muirhead˙aspects˙1982**)**.** *Let* $U_A = ZU_\Sigma$*. Then*

$$z_{ij} \sim \begin{cases} \mathcal{N}(0,1) & i < j \\ \chi_{n+1-j} & i = j \end{cases} \iff U_A \sim \mathrm{chol}\mathcal{W}_m(n, U_\Sigma). \tag{3.1}$$

The resulting algorithm from **smith˙algorithm˙1972** is summarized in Algorithm 1.

Algorithm 1 Generate a Cholesky-Wishart matrix.

```
function RWISHART_CHOL($m, n, U_\Sigma$)                ▷ $\mathrm{chol}\mathcal{W}_m(n, U_\Sigma)$
    $Z \leftarrow 0_{m,m}$
    for $j \in 1 \ldots m$ do
        for $i \in 1 \ldots j-1$ do
            Generate $z_{ij} \sim \mathcal{N}(0,1)$
        Generate $z_{jj} \sim \chi_{n+1-j}$               ▷ $Z \sim \mathrm{chol}\mathcal{W}_m(n, I_m)$
    $U_A \leftarrow ZU_\Sigma$
    return $U_A$
```

The standard algorithm for generating an inverse-Wishart matrix is to first generate a Wishart matrix using Algorithm 1 and then compute the matrix inverse. As noted by **jones˙generating˙1985**, this is most efficiently computed by inverting the Cholesky factor generated by Algorithm 1. The resulting sampling procedure is summarized in Algorithm 3, where cholU(X) computes the upper Cholesky factor of the positive definite matrix X. It uses Algorithm 2 to support all common parameterizations of the inverse-Wishart distribution.

Algorithm 2 Compute the upper Cholesky factor of a positive definite matrix or its inverse from the matrix or its upper Cholesky factor.

```
function CHOLESKY_UPPER($S$, invert, ischolU)
    if invert then
        if ischolU then
            $U \leftarrow S$
        else
            $U \leftarrow \mathrm{cholU}(S)$
        $C \leftarrow U^{-1}$
        $P \leftarrow CC^\top$
        return cholU($P$)
    else
        if ischolU then
            return $S$
        else
            return cholU($S$)
```

Algorithm 3 Generate an inverse-Wishart matrix using Algorithm 1.

```
function RINVWISHART_INDIRECT($m$,$n$,$S$,iscov,ischolU, retcholU)
    $U_\Sigma \leftarrow$ CHOLESKY_UPPER($S$, !iscov, ischolU)
    Generate $U_A \sim$ RWISHART_CHOL($m, n, U_\Sigma$)
    $V \leftarrow U_A^{-1}$
    $B \leftarrow VV^\top$
    if retcholU then
        $T \leftarrow \mathrm{cholU}(B)$
        return $T$
    else
        return $B$
```

3.2. **A new approach.** We present a similar decomposition to Bartlett's for the inverse-Wishart distribution:

Theorem 3.2. *Let $U_B = Z^{-1}U_\Omega$. Then*

$$z_{ij} \sim \begin{cases} \mathcal{N}(0,1) & i<j \\ \chi_{n-m+j} & i=j \end{cases} \iff U_B \sim \mathrm{chol}\mathcal{W}_m^{-1}(n, U_\Omega). \tag{3.2}$$

To prove this result, we first work out the necessary Jacobian determinants in Theorem 3.3 and Lemma 3.1.

Theorem 3.3. *Let T be an $m \times m$ invertible triangular matrix and $R = T^{-1}$. Then $(\mathrm{d}T) = |R|^{-(m+1)}(\mathrm{d}R)$.*

Proof. We first differentiate the constraint $RT = I_m$:

$$\mathrm{d}RT + R\mathrm{d}T = 0 \implies \mathrm{d}T = -T\mathrm{d}RT$$

Now we inspect the elements of $\mathrm{d}T$:

$$\begin{aligned}
-\mathrm{d}t_{ij} &= \sum_{k=1}^{m}\sum_{l=1}^{m} t_{ik}\mathrm{d}r_{kl}t_{lj} = \sum_{k=1}^{i}\sum_{l=k}^{j} t_{ik}t_{lj}\mathrm{d}r_{kl}\\
-\mathrm{d}t_{11} &= t_{11}t_{11}\mathrm{d}r_{11}\\
-\mathrm{d}t_{12} &= t_{11}t_{22}\mathrm{d}r_{12} + t_{11}t_{12}\mathrm{d}r_{11}\\
-\mathrm{d}t_{22} &= t_{22}t_{22}\mathrm{d}r_{22} + t_{21}t_{22}\mathrm{d}r_{12} + t_{21}t_{12}\mathrm{d}r_{11}\\
&\vdots\\
-\mathrm{d}t_{ij} &= t_{ii}t_{jj}\mathrm{d}r_{ij} + \text{terms involving } \mathrm{d}r_{kl} \text{ for } l < j \text{ or } (l = j \text{ and } k < i)
\end{aligned}$$

By the antisymmetric property of the exterior product, $\mathrm{d}t_{ij} \wedge \mathrm{d}t_{ij} = 0$, so if we take the exterior product of all expressions for $\mathrm{d}t_{ij}$ given above starting with $\mathrm{d}t_{11}$ and working down, only the leading terms $t_{ii}t_{jj}\mathrm{d}r_{ij}$ remain:

$$\begin{aligned}
(\mathrm{d}T) &= \bigwedge_{j=1}^{m}\bigwedge_{i=1}^{j}\mathrm{d}t_{ij} = \bigwedge_{j=1}^{m}\bigwedge_{i=1}^{j} t_{ii}t_{jj}\mathrm{d}r_{ij} = \left(\prod_{j=1}^{m} t_{jj}\prod_{i=1}^{j} t_{ii}\right)(\mathrm{d}R)\\
&= \left(\prod_{j=1}^{m} t_{jj}^{m+1}\right)(\mathrm{d}R) = |T|^{m+1}(\mathrm{d}R) = |R|^{-m-1}(\mathrm{d}R)
\end{aligned}$$

□

Lemma 3.1. *Let $Y = CX$, where Y, C, and X are $m \times m$ triangular matrices, and C is constant. Then $(\mathrm{d}Y) \propto (\mathrm{d}X)$.*

Proof. Since the operation is linear, its differential is the same, and its Jacobian depends only on the constant matrix C. □

Corollary 3.1. *Let $U_B = Z^{-1}U_\Omega$. Then Theorem 3.3 and Lemma 3.1 imply*

$$(\mathrm{d}U_B) \propto |Z|^{-m-1}(\mathrm{d}Z). \tag{3.3}$$

Proof of Theorem 3.2. First we note that $(u_\Omega)_{jj} = z_{jj}(u_B)_{jj}$. Then we use the change of variables formula with the density of Equation (2.3) and Jacobian determinant of Equation (3.3) and simplify:

$$\begin{aligned}
p_Z(Z \mid n) &= p_{U_B}(Z^{-1}U_\Omega \mid n, U_\Omega)\,|Z|^{-m-1} \\
&\propto \exp\left(-\frac{1}{2}\|Z\|^2\right)|Z|^{-m-1}\prod_{j=1}^{m}(u_B)_{jj}^{-(n+j)} \\
&= \exp\left(-\frac{1}{2}\|Z\|^2\right)\prod_{j=1}^{m} z_{jj}^{-m-1} z_{jj}^{n+j}(u_\Omega)_{jj}^{n+j} \\
&\propto \left(\prod_{j=1}^{m} z_{jj}^{n-m+j-1}e^{-z_{jj}^2/2}\right)\left(\prod_{j=2}^{m}\prod_{i=1}^{j-1} e^{-z_{ij}^2/2}\right) \\
&\propto \left(\prod_{j=1}^{m} \chi_{n-m+j}(z_{jj})\right)\left(\prod_{j=2}^{m}\prod_{i=1}^{j-1}\mathcal{N}(z_{ij} \mid 0,1)\right)
\end{aligned}$$

□

Theorem 3.2 motivates Algorithm 4 for generating a random inverse-Wishart Cholesky factor, which we then use in Algorithm 5 for generating random inverse-Wishart matrices.

Algorithm 4 Generate a Cholesky-inverse-Wishart matrix.

```
function RINVWISHART_CHOL($m, n, U_\Omega$)        ▷ chol$\mathcal{W}_m^{-1}(n, U_\Omega)$
    $Z \leftarrow 0_{m,m}$
    for $j \in 1 \ldots m$ do
        for $i \in 1 \ldots j-1$ do
            Generate $z_{ij} \sim \mathcal{N}(0,1)$
        Generate $z_{jj} \sim \chi_{n-m+j}$
    $C \leftarrow Z^{-1}$        ▷ $C \sim$ chol$\mathcal{W}_m^{-1}(n, I_m)$
    $U_B \leftarrow C U_\Omega$
    return $U_B$
```

Algorithm 5 Generate an inverse-Wishart-distributed matrix.

```
function RINVWISHART_DIRECT($m$,$n$,$S$,iscov,ischolU,retcholU)
    $U_\Omega \leftarrow$ CHOLESKY_UPPER($S$, iscov, ischolU)
    Generate $U_B \sim$ RINVWISHART_CHOL($m, n, U_\Omega$)
    if retcholU then
        return $U_B$
    else
        $B \leftarrow U_B^\top U_B$
        return $B$
```

Reference implementations of all algorithms are available at `https://github.com/sethaxen/rand_invwishart`.

4. Comparison of algorithms

Parameter (S)	Algorithm 3			Algorithm 5		
	TRTRI	TRMM	POTRF	TRTRI	TRMM	POTRF
Σ	**1**	**2**	**1(2)**	2	3(2)	2
U_Σ	**1**	**2**	**0(1)**	2	3(2)	1
Ω	2	3	2(3)	**1**	**2(1)**	**1**
U_Ω	2	3	1(2)	**1**	**2(1)**	**0**

Table 1. Number of $\mathcal{O}(m^3)$ BLAS and LAPACK operations required by each algorithm for each parameterization. The corresponding count to generate just the Cholesky factor of the random matrix, when different, is given in parentheses. For each parameterization the counts for the most efficient algorithm are bolded.
TRMM (BLAS): multiplication by a triangular matrix
TRTRI (LAPACK): inversion of a triangular matrix
POTRF (LAPACK): Cholesky factorization

Both Algorithm 1 and Algorithm 4 require the same number and type of PRNG calls, so Algorithm 3 and Algorithm 5 primarily differ in the number of $\mathcal{O}(m^2)$ allocations (which can be eliminated by accepting and reusing storage) and in the number of $\mathcal{O}(m^3)$ linear algebraic operations. Table 1 gives the number of $\mathcal{O}(m^3)$ operations required by each algorithm, depending on the parameterization.

If the parameter is specified as Σ or U_Σ, then Algorithm 3 is the most efficient, while Algorithm 5 is more efficient if the parameter is specified as Ω or U_Ω. The difference is more pronounced when generating random Cholesky factors, since for the standard algorithm, more work is necessary to compute the factor, while for the new algorithm the Cholesky factor is already available.

We therefore recommend Algorithm 3 if the inverse-Wishart distribution is parameterized by Σ or U_Σ and Algorithm 5 if parameterized by Ω or U_Ω.

Acknowledgments. Seth Axen is a member of the Cluster of Excellence Machine Learning: New Perspectives for Science, University of Tübingen and is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645.