Comments

- The derivation of the Wishart in the standard basis seems to work just fine. If you don't care about it you can skip to 1.2.
- For simplicity I just replaced all X with X^2 for the transformed case. Even though this is not clean I found it the easiest way to calculate with.
- The derivations for the first and second derivative for the sqrtm-transformed case are correct. I have double-checked them numerically. The Problem is just to get the mode by solving for X.

Interlude: Definition of the Box-product compared to the Kronecker-Product:

Kronecker-product: $A \otimes B \in \mathbb{R}^{(m_1 m_2) \times (n_1 n_2)}$ is defined by $(A \otimes B)_{(i-1)m_2+j,(k-1)n_2+l} = a_{il} b_{jk} = (A \otimes B)_{(ij)(kl)}$.

Box-product: $A \boxtimes B \in \mathbb{R}^{(m_1 m_2) \times (n_1 n_2)}$ is defined by $(A \boxtimes B)_{(i-1)m_2+j,(k-1)n_1+l} = a_{ik} b_{jl} = (A \boxtimes B)_{(ij)(kl)}$.

I found this box-product only in two sources, one of which is this: `https://researcher.watson.ibm.com/researcher/files/us-pederao/ADTalk.pdf` but it generally seems to be very helpful for matrix derivations with transposed matrices.

1 Wishart Distribution

1.1 Wishart distribution in Standard Basis

the pdf of the Wishart is

$$f(X; n, p, V) = \frac{1}{2^{np/2} \left|\mathbf{V}\right|^{n/2} \Gamma_p\left(\frac{n}{2}\right)} \left|\mathbf{X}\right|^{(n-p-1)/2} e^{-(1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{X})} \tag{1}$$

which can be written as

$$f(X; n, p, V) = \exp\left[(n-p-1)/2 \log(|X|) - (1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{X}) - \log\left(2^{np/2} \left|\mathbf{V}\right|^{n/2} \Gamma_p\left(\frac{n}{2}\right)\right)\right] \tag{2}$$

with $T = (\log(X), X), \eta = ((n-p-1)/2, V^{-1})$ and $A(n, p, V) = \log\left(2^{np/2} \left|\mathbf{V}\right|^{n/2} \Gamma_p\left(\frac{n}{2}\right)\right)$

1.1.1 Laplace Approximation of the standard Wishart distribution

Using $\frac{\partial \det(X)}{\partial X} = \det(X)(X^{-1})^\top$ and $\frac{\partial}{\partial X} Tr(AX^\top) = A$ we can calculate the mode by setting the first derivative of the log-pdf to zero

$$\begin{aligned}
\frac{\partial \log f(X; n, p, V)}{\partial X} &= \frac{(n-p-1)\det(X)(X^{-\top})}{2\det(X)} - \frac{V^{-1}}{2} \\
\Rightarrow 0 &= \frac{(n-p-1)X^{-1}}{2} - \frac{V^{-1}}{2} \\
\Leftrightarrow \frac{(n-p-1)X^{-1}}{2} &= \frac{V^{-1}}{2} \\
\Leftrightarrow X &= (n-p-1)V
\end{aligned}$$

Using the fact that $\frac{\partial X^{-T}}{\partial X} = X^{-T} \boxtimes X^{-1}$ where $\boxtimes$ is the Box-product we compute the second derivative as

$$\frac{\partial^2 \log f(X;n,p,V)}{\partial^2 X} = -\frac{(n-p-1)}{2} X^{-\top} \boxtimes X^{-1}$$

Using $(\alpha A)^{-1} = \alpha^{-1} A^{-1}$, the linearity of the Kronecker product to pull out scalars and $X^{-1} \boxtimes X^{-1} = (X \boxtimes X)^{-1}$ to insert the mode and invert we get:

$$\begin{aligned}
-\frac{(n-p-1)}{2} X^{-1} \boxtimes X^{-1} &= -\frac{(n-p-1)}{2} \frac{1}{(n-p-1)} V^{-1} \otimes \frac{1}{(n-p-1)} V^{-1} \\
&= -\frac{1}{2(n-p-1)} (V \boxtimes V)^{-1} \\
\Rightarrow \Sigma &= 2(n-p-1)(V \boxtimes V)
\end{aligned}$$

In summary, the Laplace approximation of a Wishart distribution in the standard basis is $\mathcal{N}\left(X; (n-p-1)V, 2(n-p-1)(V \boxtimes V)\right)$, where the representation of the symmetric positive definite matrices has been changed from $\mathbb{R}^{n \times n}$ to $\mathbb{R}^{n^2}$.

1.2 Sqrtm-Transformed Wishart distribution

we transform the distribution with $g(X) = \text{sqrtm}(X) = X^{\frac{1}{2}}$, i.e. $g^{-1}(X) = X^2$, where $\text{sqrtm}(X)$ is the square root of the matrix. The new pdf becomes

$$\begin{aligned}
f_t(X;n,p,V) &= \frac{1}{2^{np/2} \left|\mathbf{V}\right|^{n/2} \Gamma_p\left(\frac{n}{2}\right)} \left|\mathbf{X^2}\right|^{(n-p-1)/2} e^{-(1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{X^2})} \cdot |2X| &(3)\\
&= \frac{1}{2^{np/2} \left|\mathbf{V}\right|^{n/2} \Gamma_p\left(\frac{n}{2}\right)} \left|\mathbf{X}\right|^{2(n-p-1)/2} e^{-(1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{X^2})} \cdot 2^p|X| &(4)\\
&= \frac{1}{2^{np/2} \left|\mathbf{V}\right|^{n/2} \Gamma_p\left(\frac{n}{2}\right)} \left|\mathbf{X}\right|^{(n-p)} e^{-(1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{X^2})} &(5)
\end{aligned}$$

where we drop the 2^p in line (4) because there are 2^p matrices that are a root of X (I have explained this in more detailed in another version of the current draft). This can be rewritten as

$$f_t(X;n,p,V) = \exp\left[(n-p)\log(|X|) - (1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{X^2}) - \log\left(2^{np/2}\left|\mathbf{V}\right|^{n/2}\Gamma_p\left(\frac{n}{2}\right)\right)\right] \tag{6}$$

with $T = (\log(X), X^2), \eta = ((n-p), V^{-1})$ and $A(n,p,V) = \log\left(2^{np/2}\left|\mathbf{V}\right|^{n/2}\Gamma_p\left(\frac{n}{2}\right)\right)$

1.2.1 Laplace Approximation of the sqrtm-transformed Wishart distribution

Using $\frac{\partial \det(X)}{\partial X} = \det(X)(X^{-1})^\top$ and $\frac{\partial}{\partial X} Tr(AX^2) = (AX + XA)^T$ we can calculate the mode by setting the first derivative of the log-pdf to zero

$$\begin{aligned}
\frac{\partial \log f_t(X;n,p,V)}{\partial X} &= \frac{(n-p)\det(X)(X^{-\top})}{\det(X)} - \frac{(V^{-1}X + XV^{-1})^\top}{2} \\
\Rightarrow 0 &= (n-p)X^{-\top} - \frac{(V^{-1}X + XV^{-1})^\top}{2} \\
\Leftrightarrow (n-p)X^{-\top} &= \frac{(V^{-1}X + XV^{-1})^\top}{2} \\
\Leftrightarrow (n-p)X^{-1} &= \frac{(V^{-1}X + XV^{-1})}{2} \\
\Leftrightarrow X &= ???
\end{aligned}$$

THIS IS WHERE SOLVING FOR X GETS COMPLICATED. Maybe we can rewrite it with Kronecker products and vectorized matrices like for the Sylvester equation and these laws `https://en.wikipedia.org/wiki/Vectorization_(mathematics)#Compatibility_with_Kronecker_products`.

So far I have found the following relationships that don't get me any further to the solution of X:

$$
\begin{aligned}
(n-p)X^{-1} &= \frac{(V^{-1}X + XV^{-1})}{2} \\
&\Leftrightarrow C = BXX + XBX \\
&\Leftrightarrow C = (I_p \otimes BX)\vec{X} + (B^T X^T \otimes I_p)\vec{X} \\
&\Leftrightarrow C = (B^T X^T \oplus BX)\vec{X} \\
&\Leftrightarrow C = (BX \oplus BX)\vec{X}
\end{aligned}
$$

Computing the second derivative by using $\frac{\partial}{\partial X} X^{-\top} = -X^{-\top} \boxtimes X^{-1}$, $\frac{\partial}{\partial X}(AX + XA)^{\top} = I \boxtimes A + A \boxtimes I$:

$$
\begin{aligned}
\frac{\partial^2 \log f_t(X; n, p, V)}{\partial^2 X} &= \frac{\partial}{\partial X}\left[(n-p)X^{-\top} - \frac{(V^{-1}X + XV^{-1})^{\top}}{2}\right] \\
&= -(n-p)(X^{-\top} \boxtimes X^{-1}) - \frac{1}{2}(I_p \boxtimes V^{-1} + V^{-1} \boxtimes I_p)
\end{aligned}
$$

Now we would want to multiply with -1 and insert the mode for X. However, this hinges on the problem of actually solving the first derivative for X.