Bayesian Analysis (0000) **00**, Number 0, pp. 1

Bayesian Inference over the Stiefel Manifold via the Givens Representation: Appendix

Arya A. Pourzanjani*, Richard M. Jiang*, Brian Mitchell*, Paul J. Atzberger† and Linda R. Petzold *

Appendix A: Deriving the Change-of-Measure Term

We derive the simplified form of the differential form showing that

$$\bigwedge_{i=1}^{p}\bigwedge_{j=i+1}^{n} G_j^T\, dY_i = \prod_{i=1}^{p}\prod_{j=i+1}^{n} \cos^{j-i-1}\theta_{ij}. \tag{A.1}$$

We point out that Khatri and Mardia (1977) provide a similar expression for a slightly different representation, but do not offer a derivation. We start by considering the determinant of the matrix form of the change-of-measure term:

$$\bigwedge_{i=1}^{p}\bigwedge_{j=i+1}^{n} G_j^T\, J_{Y_i(\Theta)}(\Theta)d\Theta = \begin{pmatrix} G_{2:n}^T J_{Y_1(\Theta)}(\Theta) \\ G_{3:n}^T J_{Y_2(\Theta)}(\Theta) \\ \vdots \\ G_{p:n}^T J_{Y_p(\Theta)}(\Theta) \end{pmatrix} \tag{A.2}$$

For $l = 1, \cdots, n$, we define the following shorthand notation

$$\partial_{i,i+l} Y_k := \frac{\partial}{\partial\theta_{i,i+l}} Y_k \tag{A.3}$$

and

$$\partial_i Y_k := \begin{pmatrix} \partial_{i,i+1}Y_k & \partial_{i,i+2}Y_k & \cdots & \partial_{in}Y_k . \end{pmatrix} \tag{A.4}$$

In the new notation Equation can be written in the following block matrix form:

$$\begin{pmatrix} G_{2:n}^T\partial_1 Y_1 & G_{2:n}^T\partial_2 Y_1 & \cdots & G_{2:n}^T\partial_p Y_1 \\ G_{3:n}^T\partial_1 Y_2 & G_{3:n}^T\partial_2 Y_2 & \cdots & G_{3:n}^T\partial_p Y_2 \\ \vdots & \vdots & \ddots & \vdots \\ G_{p:n}^T\partial_1 Y_p & G_{p:n}^T\partial_2 Y_p & \cdots & G_{p:n}^T\partial_p Y_p \end{pmatrix}. \tag{A.5}$$

*Computer Science Department, University of California, Santa Barbara arya@ucsb.edu
†Mathematics Department, University of California, Santa Barbara

 DOI: 0000

Note that the block matrices above the diagonal are all zero. This can be seen by observing that the rotations in the Givens representation involving elements greater than i will not affect e_i, i.e. letting $R_i := R_{i,i+1} \cdots R_{in}$,

$$Y_i = R_1 R_2 \cdots R_p e_i = R_1 \cdots R_i e_i. \tag{A.6}$$

Thus for $j > i$, $\partial_j Y_i = 0$ and the determinant of Expression A.5 simplifies to the product of the determinant of the matrices on the diagonal i.e. the following expression:

$$\prod_{i=1}^{p} \det\left(G_{i+1:n}^T \partial_i Y_i\right). \tag{A.7}$$

A.1 Simplifying Diagonal Block Terms

Let I_i denote the first i columns of the $n \times n$ identity matrix and let I_{-i} represent the last $n-i$ columns. The term $G_{i+1:n}^T$ in expression A.7 can be written as

$$G_{i+1:n}^T = I_{-i}^T G^T = I_{-i}^T R_p^T \cdots R_1^T. \tag{A.8}$$

To simplify the diagonal block determinant terms in Expression A.7 we take advantage of the following fact

$$\det\left(G_{i+1:n}^T \partial_i Y_i\right) \quad = \quad \det\left(I_{-i}^T R_p^T \cdots R_1^T\right) = \det\left(I_{-i}^T R_i^T \cdots R_1^T \partial_i Y_i\right). \tag{A.9}$$

In other words, the terms $R_p^T \cdots R_{i+1}^T$ have no effect on the determinant. This can be seen by first separating terms so that

$$\begin{aligned} \det\left(G_{i+1:n}^T \partial_i Y_i\right) &= \det\left(\underbrace{I_{-i}^T}_{(n-i)\times n} R_p^T \cdots R_1^T \underbrace{\partial_i Y_i}_{n\times(n-i)}\right) && \text{(A.10)} \\ &= \det\left(I_{-i}^T \left[R_p^T \cdots R_{i+1}^T\right] \left[R_i^T \cdots R_1^T \partial_i Y_i\right]\right), && \text{(A.11)} \end{aligned}$$

and then noticing that $R_{i+1} \cdots R_p$ affects only the first i columns of the identity matrix so

$$I_{-i}^T \left[R_p^T \cdots R_{i+1}^T\right] \quad = \quad \left(R_{i+1} \cdots R_p \, I_{-i}\right)^T = \left(I_{-i}\right)^T. \tag{A.12}$$

Thus Expression A.7 is equivalent to

$$\prod_{i=1}^{p} \det\left(I_{-i}^T R_i^T \cdots R_1^T \partial_i Y_i\right). \tag{A.13}$$

Now consider the k, l element of the $(n-i)\times(n-i)$ block matrix $I_{-i}^T R_i^T \cdots R_1^T \partial_i Y_i$. This can be written as

$$\begin{aligned} e_{i+k}^T R_i^T \cdots R_1^T \partial_{i,i+l} Y_i &= e_{i+k}^T R_i^T \cdots R_1^T \partial_{i,i+l}(R_1 \cdots R_i e_i) \\ &= e_{i+k}^T R_i^T \cdots R_1^T R_1 \cdots R_{i-1}(\partial_{i,i+l} R_i e_i) \\ &= e_{i+k}^T R_i^T (\partial_{i,i+l} R_i e_i). \end{aligned} \tag{A.14}$$

Since $e_{i+k}^T R_i^T R_i e_i = 0$, taking the derivatives of both sides and applying the product rule yields

$$\begin{aligned} & \partial_{i,i+l}(e_{i+k}^T R_i^T R_i e_i) = \partial_{i,i+l} 0 \\ \Rightarrow \quad & (\partial_{i,i+l} e_{i+k}^T R_i^T) R_i e_i + e_{i+k}^T R_i^T (\partial_{i,i+l} R_i e_i) = 0 \\ \Rightarrow \quad & e_{i+k}^T R_i^T (\partial_{i,i+l} R_i e_i) = -(\partial_{i,i+l} e_{i+k}^T R_i^T) R_i e_i. \end{aligned} \tag{A.15}$$

Combining expression A.15 this fact with expression A.14, the expression for the k, l element of $I_{-i}^T R_i^T \cdots R_1^T \partial_i Y_i$ becomes $-(\partial_{i,i+l} e_{i+k}^T R_i^T) R_i e_i$.

However, note that

$$e_{i+k}^T R_i^T \quad = \quad e_{i+k}^T R_{in}^T \cdots R_{i,i+1}^T = e_{i+k}^T R_{i,i+k}^T \cdots R_{i,i+1}^T, \tag{A.16}$$

and the partial derivative of this expression with respect to $i, i+l$ is zero when $k > l$. Thus it is apparent that $I_{-i}^T R_i^T \cdots R_1^T \partial_i Y_i$ contains zeros above the diagonal and that $\det\left(I_{-i}^T R_i^T \cdots R_1^T \partial_i Y_i\right)$ is simply the product of the diagonal elements of the matrix.

A.2 Diagonal Elements of the Block Matrices

To obtain the diagonal terms of the block matrices, we directly compute $-\partial_{i,i+l} e_{i+k}^T R_i^T$ for $l = k$, $R_i e_i$, and their inner-product. Defining $D_{ij} := \partial_{ij} R_{ij}$,

$$\begin{aligned} -\partial_{i,i+k} R_i e_{i+k} &= -\partial_{i,i+k}(R_{i,i+1} \cdots R_{i,i+k} e_{i+k}) && \text{(A.17)} \\ &= -R_{i,i+1} \cdots R_{i,i+k-1} D_{i,i+k} e_{i+k} && \text{(A.18)} \\ & && \text{(A.19)} \end{aligned}$$

$$= \quad R_{i,i+1} \cdots R_{i,i+k-1} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \cos\theta_{i,i+k} \\ 0 \\ \vdots \\ 0 \\ \sin\theta_{i,i+k} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{A.20}$$

$$= \quad R_{i,i+1} \cdots R_{i,i+k-2} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \cos\theta_{i,i+k-1}\cos\theta_{i,i+k} \\ 0 \\ \vdots \\ 0 \\ \sin\theta_{i,i+k-1}\cos\theta_{i,i+k} \\ \sin\theta_{i,i+k} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{A.21}$$

$$= \quad \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \cos\theta_{i,i+1}\cos\theta_{i,i+2}\cdots\cos\theta_{i,i+k-1}\cos\theta_{i,i+k} \\ \sin\theta_{i,i+1}\cos\theta_{i,i+2}\cdots\cos\theta_{i,i+k-1}\cos\theta_{i,i+k} \\ \vdots \\ \sin\theta_{i,i+k-1}\cos\theta_{i,i+k} \\ \sin\theta_{i,i+k} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{A.22}$$

$$\tag{A.23}$$

which is zero up to the ith spot. After the $i+k$th spot,

$$R_i e_i \quad = \quad R_{i,i+1} \cdots R_{in} e_i \tag{A.24}$$

$$\tag{A.25}$$

$$= \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \cos\theta_{i,i+1}\cos\theta_{i,i+2}\cdots\cos\theta_{i,n-1}\cos\theta_{in} \\ \sin\theta_{i,i+1}\cos\theta_{i,i+2}\cdots\cos\theta_{i,n-1}\cos\theta_{in} \\ \vdots \\ \sin\theta_{i,n-1}\cos\theta_{in} \\ \sin\theta_{in} \end{pmatrix}. \tag{A.26}$$

Finally, directly computing the inner-product of $-\partial_{i,i+l}e_{i+k}^T R_i^T$ and $R_i e_i$ yields

$$\begin{aligned} -(\partial_{i,i+l}e_{i+k}^T R_i^T)(R_i e_i) &= \cos^2\theta_{i,i+1}\cos^2\theta_{i,i+2}\cdots\cos^2\theta_{i,i+k}\cos\theta_{i,i+k+1}\cdots\cos\theta_{in} \\ &+ \sin^2\theta_{i,i+1}\cos^2\theta_{i,i+2}\cdots\cos^2\theta_{i,i+k}\cos\theta_{i,i+k+1}\cdots\cos\theta_{in} \\ &+ \sin^2\theta_{i,i+2}\cos^2\theta_{i,i+3}\cdots\cos^2\theta_{i,i+k}\cos\theta_{i,i+k+1}\cdots\cos\theta_{in} \\ &\vdots \\ &+ \sin^2\theta_{i,i+k}\cos\theta_{i,i+k+1}\cdots\cos\theta_{in} \\ &= \cos^2\theta_{i,i+2}\cos^2\theta_{i,i+3}\cdots\cos^2\theta_{i,i+k}\cos\theta_{i,i+k+1}\cdots\cos\theta_{in} \\ &+ \sin^2\theta_{i,i+2}\cos^2\theta_{i,i+3}\cdots\cos^2\theta_{i,i+k}\cos\theta_{i,i+k+1}\cdots\cos\theta_{in} \\ &\vdots \\ &+ \sin^2\theta_{i,i+k}\cos\theta_{i,i+k+1}\cdots\cos\theta_{in} \\ &= \cdots \\ &= \cos\theta_{i,i+k+1}\cdots\cos\theta_{in} \\ &= \prod_{k=i+1}^{n}\cos\theta_{ik}. \end{aligned} \tag{A.27}$$

Thus the determinant of the entire block matrix $I_{-i}^T R_i^T \cdots R_1^T \partial_i Y_i$ simplifies to

$$\prod_{k=i+1}^{n}\left(\prod_{j=k+1}^{n}\cos\theta_{ik}\right) = \prod_{j=i+1}^{n}\cos^{j-i-1}\theta_{ij}. \tag{A.28}$$

Combining this with Expression A.13 yields

$$\prod_{i=1}^{p}\det\left(I_{-i}^T R_i^T\cdots R_1^T\partial_i Y_i\right) = \prod_{i=1}^{p}\prod_{j=i+1}^{n}\cos^{j-i-1}\theta_{ij}. \tag{A.29}$$

References

Khatri, C. and Mardia, K. (1977). "The von Mises-Fisher matrix distribution in orientation statistics." *Journal of the Royal Statistical Society. Series B (Methodological)*, 95–106.