Bayesian Analysis (0000)

00, Number 0, pp. 1

Supplementary material: Conjugate Priors and Posterior Inference for the Matrix Langevin Distribution on the Stiefel Manifold

Subhadip Pal^{\dagger}, Subhajit Sengupta^{\ddagger}, Riten Mitra^{\dagger} and Arunava Banerjee^{*,§}

1 Properties of the Matrix Langevin distribution and ${}_0F_1\left(rac{n}{2},rac{D^2}{4} ight)$

We introduce a few lemmas. Readers may skip this section with no loss of understanding of subsequent sections in the paper.

Lemma 1. Let X be a random matrix taking values on the space $\mathcal{V}_{n,p}$. If $X \sim \mathcal{ML}(\cdot; M, d, V)$, then $E(X) = MD_{\mathbf{h}}V^T$. $D_{\mathbf{h}}$ is a diagonal matrix with diagonal entrees $\mathbf{h}(d) := (h_1(d), \ldots, h_p(d))$ where

$$h_j(\boldsymbol{d}) := \frac{\frac{\partial}{\partial d_j} {}_0 F_1\left(\frac{n}{2}, \frac{D^2}{4}\right)}{{}_0 F_1\left(\frac{n}{2}, \frac{D^2}{4}\right)} \text{ for } j = 1, 2, \cdots, p.$$

Proof of Lemma 1. Let $\Gamma_0 = [M, \overline{M}]$ be a $n \times n$ orthogonal matrix where the columns of the matrix \overline{M} comprise of a orthonormal basis for the orthogonal complement of the column space of M. Consider the random matrix $Y = \Gamma_0^T X V$. From Khatri and Mardia (1977) (see page 98) we know that

$$E(Y) = \begin{bmatrix} D_{\mathbf{h}} &, & \mathbf{0}_{n-p,p} \end{bmatrix}^T,$$
(1.1)

where $D_{\mathbf{h}}$ is a diagonal matrix with diagonal entrees $\mathbf{h}(\mathbf{d}) := (h_1(\mathbf{d}), \dots, h_p(\mathbf{d}))$ with

$$h_j(\boldsymbol{d}) := \frac{\frac{\partial}{\partial d_j} {}_0 F_1\left(\frac{n}{2}, \frac{D^2}{4}\right)}{{}_0 F_1\left(\frac{n}{2}, \frac{D^2}{4}\right)} \text{ for } j = 1, 2, \cdots, p.$$

Hence from Equation (1.1), it follows that

$$E(X) = \Gamma_0 \left[D_{\mathbf{h}} , \mathbf{0}_{n-p,p} \right]^T V^T = M D_{\mathbf{h}} V^T.$$

DOI: 0000

Center for Psychiatric Genetics, NorthShore University HealthSystem[‡]

arunava@ufl.edu

© 0000 International Society for Bayesian Analysis

Department of Bioinformatics and Biostatistics, University of Louisville[†]

Department of Computer & Information Science & Engineering, University of Florida § E-mail of the corresponding author

^{*}Corresponding author

Lemma 2. (*Chikuse, 2012; Hoff, 2009*) For any $p \times p$ diagonal matrix D with positive elements, $_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right) \leq etr(D)$ when $n \geq p$.

Proof of Lemma 2. From Equation (2.2) in the main article, we have

$$\int_{\mathcal{V}_{n,p}} f_{\mathcal{ML}}(X; (M, \boldsymbol{d}, V)) \, d\mu(X) = 1$$
$$\implies \ _{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right) = \int_{\mathcal{V}_{n,p}} etr(VDM^{T}X) \, d\mu(X).$$
(1.2)

We know that $f_{\mathcal{ML}}(X; (M, \boldsymbol{d}, V))$ has the unique modal orientation MV^T (page 32 in Chikuse (2012)). Hence it follows from Equation (1.2) that

$${}_{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right) \leq \int_{\mathcal{V}_{n,p}} etr(VDM^{T}MV^{T}) d\mu(X)$$

$$= etr(D) \int_{\mathcal{V}_{n,p}} d\mu(X) = etr(D), \qquad (1.3)$$

as μ is the normalized Haar measure, i.e. a probability measure on $\mathcal{V}_{n,p}$.

Lemma 3. Let A be a $n \times p$ real matrix with $n \geq p$, and $A_{j,j}$ be the (j, j)-th entry of the matrix A for j = 1, ..., p. Let $||A||_2$ denote the matrix operator norm (also known as spectral norm) of the matrix A. If $||A||_2 \leq \delta$ for some $\delta > 0$ then $|A_{j,j}| \leq \delta$ for j = 1, ..., p. Also, if $||A||_2 < \delta$ for some $\delta > 0$ then $|A_{j,j}| < \delta$ for j = 1, ..., p.

Proof of Lemma 3.

From the assumptions of Lemma 3 along with the definition of the spectral norm, it follows that $l^T A^T A \ l \leq \delta^2$ for all $l \in \mathbb{R}^p$ with $l^T l = 1$. In particular, $e_j^T A^T A \ e_j \leq \delta^2$ where $e_j \in \mathbb{R}^p$ such that its *j*-th entry equals 1 while rest of its entries are 0. Hence we have $\sum_{k=1}^n A_{k,j}^2 \leq \delta^2$ implying the fact that $|A_{j,j}| \leq \delta$. The assertion with strict inequality can also be shown in a similar fashion.

Lemma 4. Let D be a $p \times p$ diagonal matrix with positive diagonal elements $d = (d_1, d_2, \dots, d_p)$. Then for any $\delta > 0$ and $n \ge p$, there exists a positive constant, $K_{n,p,\delta}$, depending on n, p and δ , such that

$$_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right) > K_{n,p,\delta} \operatorname{etr}\left((1-\delta)D\right).$$

Proof of Lemma 4.

Note that D is a $p \times p$ diagonal matrix with positive diagonal elements $d_1, ..., d_p$. For the case $n \ge p$, define

$$\widetilde{M} = \begin{bmatrix} \mathbf{I}_p \\ \mathbf{0}_{n-p,p} \end{bmatrix}, \widetilde{V} = \mathbf{I}_p \text{ and } I^* := \begin{bmatrix} \mathbf{I}_p \\ \mathbf{0}_{n-p,p} \end{bmatrix},$$
(1.4)

where \mathbf{I}_p denotes the $p \times p$ identity matrix and $\mathbf{0}_{n-p,p}$ represents the zero matrix of dimension $(n-p) \times p$. For arbitrary given positive constant $\delta > 0$, consider

$$B_{\delta} := \{ X \in \mathcal{V}_{n,p}, \text{ such that } \| X - I^{\star} \|_2 < \delta \},\$$

where $\|\cdot\|_2$ denotes the spectral norm of a matrix. Let μ denotes the normalized Haar measure on the $\mathcal{V}_{n,p}$. Clearly, $0 < \mu(B_{\delta}) < \infty$, as B_{δ} is a non-empty open subset of $\mathcal{V}_{n,p}$. Now from Equation (2.2) we have,

$${}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right) = \int_{\mathcal{V}_{n,p}} etr\left(\widetilde{V}D\widetilde{M}^{T}X\right)d\mu(X).$$

$$\geq \int_{B_{\delta}} etr\left(\widetilde{V}D\widetilde{M}^{T}X\right)d\mu(X).$$
(1.5)

Using Lemma 3 we know that $X_{j,j} > (1-\delta)$ for j = 1, 2, ..., p where $X \in B_{\delta}$. Note that $X_{j,j}$ denotes the (j, j)-th entry of the matrix X. Hence from Equation (1.4) and (1.5) it follows that,

$${}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right) \geq \int_{B_{\delta}} \exp\left(\sum_{j=1}^{p} X_{j,j} d_{j}\right) d\mu(X),$$

$$> \mu(B_{\delta}) \operatorname{etr}\left((1-\delta)D\right), \qquad (1.6)$$

where the last inequality uses the fact that $d_j > 0$ for all j = 1, ..., p. Finally we denote $K_{n,p,\delta} := \mu(B_{\delta}) > 0$ as it depends on n, p along with δ , to conclude that

$$_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right) > K_{n,p,\delta} \ etr\left((1-\delta)D\right).$$

Lemma 5. For any $p \times p$ diagonal matrix D with positive elements $\mathbf{d} \in S_p$, the hypergeometric function of matrix argument denoted by ${}_0F_1\left(\frac{n}{2},\frac{D^2}{4}\right)$ is log-convex with respect to \mathbf{d} where $n \geq p$.

Proof of Lemma 5.

From Equation (2.2) in the main article, we have

$${}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right) = \int_{\mathcal{V}_{n,p}} etr(VDM^{T}X) d\mu(X), \qquad (1.7)$$

for arbitrary $M \in \widetilde{\mathcal{V}}_{n,p}$ and $V \in \mathcal{V}_{n,p}$ where $n \ge p$. Without loss of generality, we can take $M = \widetilde{M} = \begin{bmatrix} \mathbf{I}_p \\ \mathbf{0}_{(n-p),p} \end{bmatrix}$ and $V = \mathbf{I}_p$.

Let D_1 and D_2 be two $p \times p$ diagonal matrices with positive diagonal entries d_1 and d_2 , respectively, and $d_1 \neq d_2$. From Equation (1.7), we have

$${}_{0}F_{1}\left(\frac{n}{2},\frac{D_{1}^{2}}{4}\right) = \int_{\mathcal{V}_{n,p}} etr(D_{1}\widetilde{M}^{T}X) d\mu(X)$$

$${}_{0}F_{1}\left(\frac{n}{2},\frac{D_{2}^{2}}{4}\right) = \int_{\mathcal{V}_{n,p}} etr(D_{2}\widetilde{M}^{T}X) d\mu(X).$$
(1.8)

Let $\lambda \in (0, 1)$ be any real number, then

$${}_{0}F_{1}\left(\frac{n}{2},\frac{\left(\lambda D_{1}+\left(1-\lambda\right) D_{2}\right)^{2}}{4}\right)$$

$$=\int_{\mathcal{V}_{n,p}}etr(\left(\lambda D_{1}+\left(1-\lambda\right) D_{2}\right)\tilde{M}^{T}X)\left[dX\right]$$

$$=\int_{\mathcal{V}_{n,p}}\left(etr(D_{1}\tilde{M}^{T}X)\right)^{\lambda}\left(etr(D_{2}\tilde{M}^{T}X)\right)^{1-\lambda}d\mu(X)$$

$$<\left(\int_{\mathcal{V}_{n,p}}etr(D_{1}\tilde{M}^{T}X)d\mu(X)\right)^{\lambda}\left(\int_{\mathcal{V}_{n,p}}etr(D_{2}\tilde{M}^{T}X)d\mu(X)\right)^{1-\lambda}$$

$$=\left({}_{0}F_{1}\left(\frac{n}{2},\frac{D_{1}^{2}}{4}\right)\right)^{\lambda}\left({}_{0}F_{1}\left(\frac{n}{2},\frac{D_{2}^{2}}{4}\right)\right)^{1-\lambda}.$$
(1.9)

where the inequality is due to Hölder (Hardy et al., 1952; Billingsley, 1995) and we have strict inequality as $d_1 \neq d_2$.

Therefore from Equation (1.9), it follows that ${}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)$ is a log-convex function of the diagonal entries d of the matrix D. Note that, the properties of the exponential family of distributions have played a crucial role in establishing the result.

Lemma 6. For any $p \times p$ $(p \ge 2)$ diagonal matrix D with positive elements $d \in S_p$,

$$0 < \frac{\partial}{\partial d_i} \left[{}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right) \right] < {}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right),$$

for $i = 1, 2, \cdots, p$, where $n \ge p$.

Proof of Lemma 6.

Right hand side inequality Proceeding along similar lines as Lemma 5 we have

$${}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right) = \int_{\mathcal{V}_{n,p}} etr(D\widetilde{M}^{T}X) \, d\mu(X), \text{ where } \widetilde{M} = \begin{bmatrix} \mathbf{I}_{p} \\ \mathbf{0}_{(n-p),p} \end{bmatrix}.$$
(1.10)

From Equation (1.10), we have

$${}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right) = \int_{\mathcal{V}_{n,p}} \exp\left(\sum_{j=1}^{p} d_{j} X_{j,j}\right) d\mu(X)$$
(1.11)

Consider the set $\mathcal{V}_0 := \{X \in \mathcal{V}_{n,p} : X_{i,i} = 1\}$. Note that \mathcal{V}_0 is isomorphic to the lower dimensional Stiefel manifold, $\mathcal{V}_{n,p-1}$. \mathcal{V}_0 , being a lower dimensional subspace of $\mathcal{V}_{n,p}$, has measure zero i.e. $\int_{\mathcal{V}_{n,p}} \mathbb{I}(X \in \mathcal{V}_0) d\mu(X) = 0$, where $\mathbb{I}(X \in \mathcal{V}_0)$ is the indicator function for X to be in the set \mathcal{V}_0 . From Equation (1.11), we have

$${}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right) = \int_{\mathcal{V}_{n,p}} \exp\left(\sum_{j=1}^{p} d_{j} X_{j,j}\right) \mathbb{I}(X \in \mathcal{V}_{0}^{c}) d\mu(X), \quad (1.12)$$

where \mathcal{V}_0^c is the complement of \mathcal{V}_0 . Hence,

$$\frac{\partial}{\partial d_i} \left[{}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right) \right] = \int_{\mathcal{V}_{n,p}} X_{i,i} \,\mathbb{I}(X \in \mathcal{V}_0^c) \,\exp\left(\sum_{j=1}^p d_j \,X_{j,j}\right) \,d\mu(X).$$
(1.13)

Observe that $||X||_2 = 1$ on $\mathcal{V}_{n,p}$. Hence from Lemma 3 we have $|X_{i,i}| \leq 1$. Also, $X_{i,i} \neq 1$ when $X \in \mathcal{V}_0^c$. As a result, we conclude that $X_{i,i} < 1$ on $\mathcal{V}_{n,p} \cap \mathcal{V}_0^c$. Consequently, it follows from Equations (1.12) and (1.13) that,

$$\frac{\partial}{\partial d_i} \left[{}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right) \right] < \int_{\mathcal{V}_{n,p}} \exp\left(\sum_{j=1}^p d_j X_{j,j}\right) \mathbb{I}(X \in \mathcal{V}_0^c) \, d\mu(X)$$
$$= {}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right).$$
(1.14)

Left hand side inequality Consider $\mathcal{V}_{n,p}^{i,+} := \{X \in \mathcal{V}_{n,p} : X_{i,i} > 0\}, \mathcal{V}_{n,p}^{i,-} := \{X \in \mathcal{V}_{n,p} : X_{i,i} < 0\}$ and $\mathcal{V}_{n,p}^{i,0} := \{X \in \mathcal{V}_{n,p} : X_{i,i} = 0\}$. Clearly, $\mathcal{V}_{n,p}^{i,+}, \mathcal{V}_{n,p}^{i,0}$ and $\mathcal{V}_{n,p}^{i,-}$ forms a partition of $\mathcal{V}_{n,p}$. Hence from Equation (1.11) we have,

$$\begin{aligned} \frac{\partial}{\partial d_i} \left[{}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right) \right] \\ = & \int_{\mathcal{V}_{n,p}^{i,+}} X_{i,i} \exp\left(\sum_{j=1}^p d_j X_{j,j}\right) \, d\mu(X) + \int_{\mathcal{V}_{n,p}^{i,0}} X_{i,i} \exp\left(\sum_{j=1}^p d_j X_{j,j}\right) \, d\mu(X) \\ & + \int_{\mathcal{V}_{n,p}^{i,+}} X_{i,i} \exp\left(\sum_{j=1}^p d_j X_{j,j}\right) \, d\mu(X) \end{aligned}$$

$$= \int_{\mathcal{V}_{n,p}^{i,+}} X_{i,i} \exp\left(\sum_{j=1}^{p} d_{j} X_{j,j}\right) d\mu(X) + \int_{\mathcal{V}_{n,p}^{i,-}} X_{i,i} \exp\left(\sum_{j=1}^{p} d_{j} X_{j,j}\right) d\mu(X).$$
(1.15)

Let Γ be the $n \times n$ diagonal matrix such that $\Gamma_{j,j} = 1$ for $j = 1, \ldots, n, j \neq i$ and $\Gamma_{i,i} = -1$. Γ is an orthogonal matrix as $\Gamma^T \Gamma = \mathbf{I_n}$. It is easy to show that $\mathcal{V}_{n,p}^{i,+} = \{\Gamma X : X \in \mathcal{V}_{n,p}^{i,-}\}.$

Consider the change of variable $Y := \Gamma X$. Using standard algebra we can show that $X_{i,i} = -Y_{i,i}$ and $X_{j,j} = Y_{j,j}$ for $j = 1, \ldots, p, j \neq i$. As the normalized Haar measure on $\mathcal{V}_{n,p}$ is invariant under orthogonal transformation from Chikuse (2012), we get that

$$\int_{\mathcal{V}_{n,p}^{i,-}} X_{i,i} \exp\left(\sum_{j=1}^{p} d_{j} X_{j,j}\right) d\mu(X)$$

$$= -\int_{\mathcal{V}_{n,p}^{i,+}} Y_{i,i} \exp\left(-d_{i} Y_{i,i} + \sum_{j=1, j \neq i}^{p} d_{j} Y_{j,j}\right) d\mu(Y)$$

$$= -\int_{\mathcal{V}_{n,p}^{i,+}} X_{i,i} \exp\left(-d_{i} X_{i,i} + \sum_{j=1, j \neq i}^{p} d_{j} X_{j,j}\right) d\mu(X). \quad (1.16)$$

From Equations (1.15) and (1.16) we have,

$$\frac{\partial}{\partial d_{i}} \left[{}_{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right) \right]$$

$$= \int_{\mathcal{V}_{n,p}^{i,+}} X_{i,i} \exp\left(\sum_{j=1, j \neq i}^{p} d_{j} X_{j,j}\right) \left(\exp\left(d_{i} X_{i,i}\right) - \exp\left(-d_{i} X_{i,i}\right)\right) d\mu(X)$$

$$= \int_{\mathcal{V}_{n,p}^{i,+}} X_{i,i} \exp\left(\sum_{j=1, j \neq i}^{p} d_{j} X_{j,j}\right) 2 \sinh\left(d_{i} X_{i,i}\right) d\mu(X)$$
(1.17)

where sinh is the hyperbolic sin function. Note that $\sinh(d_i X_{i,i}) > 0$ as $d_i > 0$ and $X_{i,i} > 0$ on $\mathcal{V}_{n,p}^{i,+}$. Hence from Equation (1.17) it follows that,

$$\frac{\partial}{\partial d_i} \left[{}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right) \right] > 0.$$
(1.18)

From Equations (1.14) and (1.18), we have the result.

Lemma 7. Let $\Psi \in \mathbb{R}^{n \times p}$ and D be a diagonal matrix with positive diagonal entries. If $\|\Psi\|_2 < 1$, then for arbitrary $M \in \mathcal{V}_{n,p}, V \in \mathcal{V}_{p,p}$,

$$\frac{\operatorname{etr}\left(VDM^{T}\Psi\right)}{{}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)} < \frac{\operatorname{etr}\left(-\epsilon_{0}\,D\right)}{K_{n,p,\epsilon_{0}}},\tag{1.19}$$

where $\epsilon_0 = \frac{1}{2} (1 - \|\Psi\|_2)$ and $K_{n,p,\epsilon_0} > 0$ is a constant depending on n, p and ϵ_0 .

Proof of Lemma 7.

Note that $0 < \epsilon_0 < \frac{1}{2}$, as $\|\Psi\|_2 < 1$. Assume $Y_0 = M^T \Psi V \in \mathbb{R}^{p \times p}$. For arbitrary $l \in \mathbb{R}^p$ with $\|l\| = 1$, we have

$$l^{T}Y_{0}^{T}Y_{0}l = (V l)^{T}\Psi^{T}\Psi(V l) - l^{T}V^{T}\Psi^{T}(\mathbf{I}_{n} - MM^{T})\Psi V l$$

$$\leq (1 - 2\epsilon_{0})^{2}.$$
(1.20)

The last inequality follows as $\|\Psi\|_2 = 1 - 2\epsilon_0$ and $(\mathbf{I}_n - MM^T)$ is a non-negative definite matrix. From Equation (1.20) it follows that $\|Y_0\|_2 \leq 1 - 2\epsilon_0$. Hence, applying Lemma 3 we obtain that $|Y_{0j,j}| < 1 - 2\epsilon_0$ for $j = 1, \dots, p$, where $Y_{0,j}$ is the *j*-th diagonal element of the matrix Y_0 . Now applying Lemma 4 we have,

$$\frac{etr\left(VDM^{T}\Psi\right)}{_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)} < \frac{etr(DY_{0}-(1-\epsilon_{0})D)}{K_{n,p,\epsilon_{0}}} < \frac{etr(-\epsilon_{0}D)}{K_{n,p,\epsilon_{0}}}.$$

Lemma 8. Let R be a $p \times p$ symmetric positive definite matrix. Then for $a \ge p/2$,

$${}_{0}F_{1}(a,R) \ge \Gamma(a)(tr(R))^{\frac{1-a}{2}}I_{a-1}(\sqrt{4\ tr(R)}),$$
(1.21)

tr(R) denotes the trace of the matrix R.

Proof of Lemma 8. Let $\mathcal{D}_{k,p}$ denotes the set of all possible partitions of the integer k into no more than p parts, i.e.

$$\mathcal{D}_{k,p} = \{ (k_1, \dots, k_p) : k_1, \dots, k_p \in \mathbb{Z}, k_1 \ge \dots, k_p \ge 0, k_1 + \dots + k_p = k \}$$

where \mathbb{Z} denotes the set of non-negative integers. For a vector $\boldsymbol{\kappa} = (k_1, \ldots, k_p) \in \mathcal{D}_{k,p}$, we denote the quantity $\prod_{j=1}^{p} \frac{\Gamma(a-(j-1)/2+k_j)}{\Gamma(a-(j-1)/2)}$ by the notation $(a)_{\boldsymbol{\kappa}}$. Then from the Richards (2011) we get the representation

$${}_{0}F_{1}\left(a,R\right) = \sum_{k=0}^{\infty} \sum_{\boldsymbol{\kappa}\in\mathcal{D}_{k,p}} \frac{C_{\boldsymbol{\kappa}}(R)}{(a)_{\boldsymbol{\kappa}}k!},\tag{1.22}$$

where $C_{\kappa}(R)$ is the Zonal polynomial of the matrix argument R corresponding to the vector $\kappa \in \mathcal{D}_{k,p}$. More details about the Zonal polynomials can be found in Muirhead (2009), Gross and Richards (1987).

Note that, for $\boldsymbol{\kappa} = (k_1, \ldots, k_p) \in \mathcal{D}_{k,p}$, and $a \geq \frac{p}{2}$,

$$(a)_{\kappa} = \prod_{j=1}^{p} \frac{\Gamma(a - (j - 1)/2 + k_j)}{\Gamma(a - (j - 1)/2)}$$

$$\leq \frac{\Gamma(a + k)}{\Gamma(a)} \prod_{j=2}^{p} \frac{\Gamma(a - (j - 1)/2)}{\Gamma(a - (j - 1)/2)}$$

$$= \frac{\Gamma(a + k)}{\Gamma(a)}.$$
(1.23)

As a result, for $a \ge \frac{p}{2} \ge 1$, we get that

$${}_{0}F_{1}(a,R) \geq \sum_{k=0}^{\infty} \frac{\Gamma(a)}{k!\Gamma(a+k)} \sum_{\boldsymbol{\kappa}\in\mathcal{D}_{k,p}} C_{\boldsymbol{\kappa}}(R)$$

$$\stackrel{(**)}{=} \sum_{k=0}^{\infty} \frac{\Gamma(a)}{k!\Gamma(a+k)} tr(R)^{k}$$

$$= \Gamma(a)(tr(R))^{\frac{1-a}{2}} I_{a-1}(\sqrt{4} tr(R)), \qquad (1.24)$$

where the equality in (**) follows from Gross and Richards (1987) (See Equation(5) in Gross and Richards (1987)), while the last equality follows from the definition of $I_{a-1}(\cdot)$, the modified Bessel function of the first kind. We would like to point out that the result is motivated by a lower-bound developed in Sengupta (2013).

Lemma 9. Let $\nu \geq \frac{1}{2}$ then for M > 0,

$$I_{\nu}(x) \ge \frac{e^x}{\sqrt{x}} \left[\sqrt{M} e^{-M} I_{\nu}(M) \right], \qquad (1.25)$$

for all x > M.

Proof of Lemma 9

First we will show that the function $x \mapsto x^{\frac{1}{2}}e^{-x}I_{\nu}(x)$ is a non decreasing function for $\nu \geq \frac{1}{2}$ and x > 0. Consider that

$$\frac{\partial}{\partial x} (x^{\frac{1}{2}} e^{-x} I_{\nu}(x)) \\
= \frac{1}{2\sqrt{x}} e^{-x} I_{\nu}(x) - (x^{\frac{1}{2}} e^{-x} I_{\nu}(x)) + x^{\frac{1}{2}} e^{-x} (-\frac{\nu}{x} I_{\nu}(x) + I_{\nu-1}(x)) \\
= \sqrt{x} e^{-x} \left((\frac{1}{2} - \nu) \frac{I_{\nu}(x)}{x} + I_{\nu-1}(x) - I_{\nu}(x) \right) \\
= \sqrt{x} e^{-x} I_{\nu}(x) \left(\frac{0.5 - \nu}{x} - 1 + \frac{I_{\nu-1}(x)}{I_{\nu}(x)} \right).$$
(1.26)

From Segura (2011), we get that $\frac{I_{\nu}(x)}{I_{\nu-1}(x)} \leq \frac{x}{(\nu-0.5)+\sqrt{(\nu-0.5)^2+x^2}}$ for $\nu \geq \frac{1}{2}$ and x > 0. Hence, from (1.26), it follows that

$$\begin{aligned} &\frac{\partial}{\partial x} \left(x^{\frac{1}{2}} e^{-x} I_{\nu}(x) \right) \\ &\geq & \sqrt{x} e^{-x} I_{\nu}(x) \left(\frac{0.5 - \nu}{x} - 1 + \frac{(\nu - 0.5) + \sqrt{(\nu - 0.5)^2 + x^2}}{x} \right) \\ &= & \sqrt{x} e^{-x} I_{\nu}(x) \left(\frac{\sqrt{(\nu - 0.5)^2 + x^2}}{x} - 1 \right) \\ &= & \frac{\sqrt{x} e^{-x} I_{\nu}(x) (\nu - 0.5)^2}{x \left(x + \sqrt{(\nu - 0.5)^2 + x^2} \right)} > 0. \end{aligned}$$

As a result, the function $x \mapsto (x^{\frac{1}{2}}e^{-x}I_{\nu}(x))$ is a non-decreasing function for $\nu \geq \frac{1}{2}$. Hence, for M > 0 we have

$$x^{\frac{1}{2}}e^{-x}I_{\nu}(x) \ge M^{\frac{1}{2}}e^{-M}I_{\nu}(M)$$
$$\implies I_{\nu}(x) \ge \frac{e^{x}}{\sqrt{x}}\left[\sqrt{M}e^{-M}I_{\nu}(M)\right],$$
(1.27)

when x > M.

Lemma 10. Let $n > p \ge 2$ and M > 0, for all $d_1 > M$,

$$g_1(d_1) \leq K_{n,p,M} d_1^{\nu(n-1)/2} exp(-\nu(1-\eta_1) d_1),$$

(1.28)

where

$$g_1(d_1) = \frac{\exp(\nu \ \eta_1 \ d_1)}{{}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right)^{\nu}} \ \text{and} \ K_{n,p,M}^{\dagger} = \left[\frac{(p/4)^{\frac{n/2-1}{2}})}{\Gamma(n/2)\left\{\sqrt{M}e^{-M}I_{n/2-1}(M)\right\}}\right]^{\nu}.$$

Proof of Lemma 10 Let $a = \frac{n}{2}$. Note that $d_1 \ge d_2 \ge \ldots \ge d_p$ are the diagonal elements of the diagonal matrix D. From Lemma (8), we get that

$${}_{0}F_{1}\left(a,\frac{D^{2}}{4}\right) \geq \Gamma(a)\left(\frac{4}{tr(D^{2})}\right)^{\frac{a-1}{2}}I_{a-1}(\sqrt{tr(D^{2})})$$

$$\geq \Gamma(a)\left(\frac{4}{pd_{1}^{2}}\right)^{\frac{a-1}{2}}I_{a-1}(d_{1}),$$
(1.29)

As a result,

$$g_1(d_1) \leq \left[\frac{(pd_1^2/4)^{\frac{a-1}{2}}exp(\eta_1 d_1)}{\Gamma(a)I_{a-1}(d_1)}\right]^{\nu}.$$
(1.30)

With the help of Lemma (9), from Equation (1.30), we get that

$$g_{1}(d_{1}) \leq \left[\frac{(pd_{1}^{2}/4)^{\frac{a-1}{2}})exp(\eta_{1}d_{1})}{\Gamma(a)\left\{\frac{e^{d_{1}}}{\sqrt{d_{1}}}\left[\sqrt{M}e^{-M}I_{a-1}(M)\right]\right\}}\right]^{\nu}$$

$$= \left[\frac{(p/4)^{\frac{a-1}{2}})d_{1}^{a-0.5}exp(-(1-\eta_{1})d_{1})}{\Gamma(a)\left\{\left[\sqrt{M}e^{-M}I_{a-1}(M)\right]\right\}}\right]^{\nu}$$

$$= \left[\frac{(p/4)^{\frac{a-1}{2}}}{\Gamma(a)\left\{\sqrt{M}e^{-M}I_{a-1}(M)\right\}}\right]^{\nu}d_{1}^{\nu(a-0.5)}exp(-\nu(1-\eta_{1})d_{1}).$$
(1.31)

Note that $\lim_{M\to\infty}\sqrt{M}e^{-M}I_{a-1}(M) = \frac{1}{\sqrt{2\pi}}$ for all $a \geq \frac{3}{2}$. The upper bound is nontrivial in the sense that for larger values of M, the constant part involved in the inequality (1.31) does not approach infinity.

All the above lemmas are used for the theoretical development of the Bayesian analysis with \mathcal{ML} distributions.

2 Proofs of the Theorems

2.1 Proof of Theorem 1.

(a) When $\|\Psi\|_2 < 1$:

The function $g(M, \mathbf{d}, V; \nu, \Psi)$ can be normalized to construct a probability density function with respect to the product measure $\mu \times \mu_1 \times \mu_2$. Consider now

$$\begin{split} & \int_{\mathcal{V}_{n,p}} \int_{\mathcal{V}_{p,p}} \int_{\mathbb{R}^{p}_{+}} g(M, \boldsymbol{d}, V; \nu, \Psi) \ d\mu_{1}(\boldsymbol{d}) \ d\mu_{2}(V) \ d\mu(M) \\ = & \int_{\mathcal{V}_{n,p}} \int_{\mathcal{V}_{p,p}} \int_{\mathbb{R}^{p}_{+}} \frac{etr\left(\nu V D M^{T} \Psi\right)}{\left[_{0}F_{1}(\frac{n}{2}, \frac{D^{2}}{4})\right]^{\nu}} \ d\mu_{1}(\boldsymbol{d}) \ d\mu_{2}(V) \ d\mu(M) \\ \stackrel{(i)}{<} & \int_{\mathcal{V}_{n,p}} \int_{\mathcal{V}_{p,p}} \int_{\mathbb{R}^{p}_{+}} \frac{etr(-\nu\epsilon_{0}D)}{(K_{n,p,\epsilon_{0}})^{\nu}} \ d\mu_{1}(\boldsymbol{d}) \ d\mu_{2}(V) \ d\mu(M) \\ = & \int_{\mathcal{V}_{n,p}} d\mu(M) \int_{\mathcal{V}_{p,p}} d\mu_{2}(V) \ \int_{\mathbb{R}^{p}_{+}} \frac{etr(-\nu\epsilon_{0}D)}{(K_{n,p,\epsilon_{0}})^{\nu}} \ d\mu_{1}(\boldsymbol{d}) \end{split}$$

$$\stackrel{(ii)}{=} \frac{1}{K_{n,p,\epsilon_0}^{\nu}} \prod_{j=1}^p \int_{\mathbb{R}_+} \exp(-\nu\epsilon_0 d_j) \, dd_j < \infty,$$

where the inequality (i) is due to Lemma 7 while (ii) follows from μ and μ_2 being the normalized Haar measures.

(b) When $\|\Psi\|_2 > 1$: Let $\Psi := M_{\Psi} D_{\Psi} V_{\Psi}^T$ be the unique SVD (Chikuse, 2012) for the matrix Ψ . Note that, using sub-multiplicativity

$$\|\Psi\|_{2} \leq \|M_{\Psi}\|_{2} \|D_{\Psi}\|_{2} \|V_{\Psi}^{T}\|_{2} = \|D_{\Psi}\|_{2} = D_{\Psi,1}.$$

Hence there exists an $\epsilon_0 > 0$ such that, $D_{\Psi,1} > (1 + \epsilon_0)$ where $D_{\Psi,1}$ denotes the first diagonal element of the diagonal matrix D_{Ψ} . Now consider that

$$\int_{\mathcal{V}_{n,p}} \int_{\mathcal{V}_{p,p}} \int_{\mathbb{R}^{p}_{+}} g(M, \boldsymbol{d}, V; \nu, \Psi) \, d\mu_{1}(\boldsymbol{d}) \, d\mu_{2}(V) \, d\mu(M) \\
\geq \int_{\mathcal{V}_{n,p}} \int_{\mathcal{V}_{p,p}} \int_{\mathcal{S}_{p}} g(M, \boldsymbol{d}, V; \nu, \Psi) \, d\mu_{1}(\boldsymbol{d}) \, d\mu_{2}(V) \, d\mu(M) \\
= \int_{\mathcal{V}_{n,p}} \int_{\mathcal{V}_{p,p}} \int_{\mathcal{S}_{p}} \frac{etr\left(\nu \, VDM^{T}\Psi\right)}{\left[0F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right)\right]^{\nu}} \, d\mu_{1}(\boldsymbol{d}) \, d\mu_{2}(V) \, d\mu(M) \\
= \int_{\mathcal{V}_{n,p}} \int_{\mathcal{V}_{p,p}} \int_{\mathcal{S}_{p}} \frac{etr\left(\nu \, DM^{T}M_{\Psi}D_{\Psi}V_{\Psi}^{T}V\right)}{\left[0F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right)\right]^{\nu}} \, d\mu_{1}(\boldsymbol{d}) \, d\mu_{2}(V) \, d\mu(M). \quad (2.1)$$

Consider the change of variable via the following orthogonal transformations

$$M^* = \left[\begin{array}{cc} M_{\Psi} & , \end{array} \right] M, \qquad V^* = V_{\Psi}^T V,$$

where \overline{M}_{Ψ} is the matrix containing the orthonormal bases for the orthogonal com-plement of the column space of M_{Ψ} . Note that $\left[M_{\Psi} , \overline{M}_{\Psi} \right]^T M_{\Psi} = \left(I^{\star} \right)^T$ where $I^{\star} := \begin{bmatrix} \mathbf{I}_p & \mathbf{0}_{n-p,p} \end{bmatrix}^T$. As the Haar measure on the Stiefel manifold is invariant under orthogonal transformations (Chikuse, 2012), from Equation 2.1 we get that,

$$\int_{\mathcal{V}_{n,p}} \int_{\mathcal{V}_{p,p}} \int_{\mathbb{R}^{p}_{+}} g(M, \boldsymbol{d}, V; \nu, \Psi) \ d\mu_{1}(\boldsymbol{d}) \ d\mu_{2}(V) \ d\mu(M)$$

$$= \int_{\mathcal{V}_{n,p}} \int_{\mathcal{V}_{p,p}} \int_{\mathbb{R}^{p}_{+}} \frac{etr\left(\nu \ DM^{*T} I^{*} D_{\Psi} V^{*}\right)}{\left[_{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right)\right]^{\nu}} \ d\mu_{1}(\boldsymbol{d}) \ d\mu_{2}(V^{*}) \ d\mu(M^{*}).$$
(2.2)

Consider

$$\mathcal{V}_{n,p}^{\dagger} := \left\{ M \in \mathcal{V}_{n,p} : \| I^{\star} - M \|_{2} < \frac{\delta_{0}}{2} \right\}; \quad \mathcal{V}_{p,p}^{\dagger} := \left\{ V \in \mathcal{V}_{p,p} : \| \mathbf{I}_{p} - V \|_{2} < \frac{\delta_{0}}{2} \right\},$$

where $\delta_0 = \epsilon_0/(2 \|D_{\Psi}\|_2)$. Note that $\delta_0 > 0$ as $0 < \|D_{\Psi}\|_2 < \infty$. Clearly $\mathcal{V}_{n,p}^{\dagger}$ and $\mathcal{V}_{p,p}^{\dagger}$ are open subsets of $\mathcal{V}_{n,p}$ and $\mathcal{V}_{p,p}$, respectively. Hence, $\mu(\mathcal{V}_{n,p}^{\dagger}) > 0$ and $\mu_2(\mathcal{V}_{p,p}^{\dagger}) > 0$. If $M \in \mathcal{V}_{n,p}^{\dagger}$ and $V \in \mathcal{V}_{p,p}^{\dagger}$ then using sub-multiplicativity of $\|\cdot\|_2$ (Conway, 1990) and the triangle inequality, we get

$$\begin{split} \left\| M^{T} I^{*} D_{\Psi} V - D_{\Psi} \right\|_{2} &\leq \left\| M^{T} I^{*} D_{\Psi} V - D_{\Psi} V \right\|_{2} + \left\| D_{\Psi} V - D_{\Psi} \right\|_{2} \\ &\leq \left\| M^{T} I^{*} - \mathbf{I}_{p} \right\|_{2} \left\| D_{\Psi} V \right\|_{2} + \left\| D_{\Psi} \right\|_{2} \left\| V - \mathbf{I}_{p} \right\|_{2} \\ &= \left\| (M - I^{*})^{T} I^{*} \right\|_{2} \left\| D_{\Psi} V \right\|_{2} + \left\| D_{\Psi} \right\|_{2} \left\| V - \mathbf{I}_{p} \right\|_{2} \\ &\leq \left\| (M - I^{*})^{T} \right\|_{2} \left\| I^{*} \right\|_{2} \left\| D_{\Psi} \right\|_{2} \left\| V \right\|_{2} + \left\| D_{\Psi} \right\|_{2} \left\| V - \mathbf{I}_{p} \right\|_{2} \\ &\leq \left\| (M - I^{*})^{T} \right\|_{2} \left\| D_{\Psi} \right\|_{2} + \left\| D_{\Psi} \right\|_{2} \left\| V - \mathbf{I}_{p} \right\|_{2} \\ &\leq \delta_{0} \left\| D_{\Psi} \right\|_{2} \\ &= \frac{\epsilon_{0}}{2}. \end{split}$$

$$(2.3)$$

Let $\lambda_1, \ldots, \lambda_p$ be the diagonal elements of the matrix $M^T I^* D_{\Psi} V$. From Lemma 3 we get $|\lambda_j - D_{\Psi,j}| \leq \epsilon_0/2$ for $j = 1, \ldots, p$. Here $D_{\Psi,j}$ denotes the *j*-th diagonal element of the matrix D_{Ψ} . Hence for arbitrary $M \in \mathcal{V}_{n,p}^{\dagger}$ and $V \in \mathcal{V}_{n,p}^{\dagger}$, we have

$$tr\left(M^{T}I^{*}D_{\Psi}V\right) = \sum_{j=1}^{p} \lambda_{j} \ge \sum_{j=1}^{p} \left(D_{\Psi,j} - \frac{\epsilon_{0}}{2}\right), \qquad (2.4)$$

as $\lambda_j \ge \left(D_{\Psi,j} - \frac{\epsilon_0}{2}\right)$ for all $j = 1, 2, \cdots, p$. Now, from Equation 2.2, we have

$$\int_{\mathcal{V}_{n,p}} \int_{\mathcal{V}_{p,p}} \int_{\mathbb{R}^{p}_{+}} g(M, \boldsymbol{d}, V; \nu, \Psi) \, d\mu_{1}(\boldsymbol{d}) \, d\mu_{2}(V) \, d\mu(M)$$

$$\geq \int_{\mathcal{V}^{\dagger}_{n,p}} \int_{\mathcal{V}^{\dagger}_{p,p}} \int_{\mathbb{R}^{p}_{+}} \frac{\operatorname{etr}\left(\nu \, DM^{*T} I^{*} D_{\Psi} V^{*}\right)}{\left[_{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right)\right]^{\nu}} \, d\mu_{1}(\boldsymbol{d}) \, d\mu_{2}(V^{*}) \, d\mu(M^{*})$$

$$\stackrel{(iii)}{\geq} \int_{\mathcal{V}^{\dagger}_{n,p}} \int_{\mathcal{V}^{\dagger}_{p,p}} \int_{\mathbb{R}^{p}_{+}} \frac{\operatorname{exp}\left(\nu \, \sum_{j=1}^{p} d_{j}\left(D_{\Psi,j} - \frac{\epsilon_{0}}{2}\right)\right)}{\left[_{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right)\right]^{\nu}} \, d\mu_{1}(\boldsymbol{d}) \, d\mu_{2}(V^{*}) \, d\mu(M^{*}),$$

$$(2.5)$$

where (iii) follows from Equation 2.4. Using Lemma 2, we get that

$$\begin{split} & \int_{\mathcal{V}_{n,p}} \int_{\mathcal{V}_{p,p}} \int_{\mathbb{R}^{p}_{+}} g(M, \boldsymbol{d}, V; \nu, \Psi) \ d\mu_{1}(\boldsymbol{d}) \ d\mu_{2}(V) \ d\mu(M) \\ \stackrel{(iv)}{\geq} & \int_{\mathcal{V}^{\dagger}_{n,p}} \int_{\mathcal{V}^{\dagger}_{p,p}} \int_{\mathbb{R}^{p}_{+}} \frac{\exp\left(\nu \sum_{j=1}^{p} d_{j} \left(D_{\Psi,j} - \frac{\epsilon_{0}}{2}\right)\right)}{\left[etr(D)\right]^{\nu}} \ d\mu_{1}(\boldsymbol{d}) \ d\mu_{2}(V^{*}) \ d\mu(M^{*}), \\ \geq & \mu(\mathcal{V}^{\dagger}_{n,p}) \ \mu_{2}(\mathcal{V}^{\dagger}_{p,p}) \ \int_{\mathbb{R}^{p}_{+}} \exp\left(\nu \sum_{j=1}^{p} d_{j} \left(D_{\Psi,j} - 1 - \frac{\epsilon_{0}}{2}\right)\right) d\mu_{1}(\boldsymbol{d}), \end{split}$$

$$\stackrel{(v)}{\geq} \quad \mu(\mathcal{V}_{n,p}^{\dagger}) \ \mu_2(\mathcal{V}_{p,p}^{\dagger}) \ \int_{\mathbb{R}^p_+} \exp\left(\nu \ \frac{\epsilon_0}{2} d_1\right) \prod_{j=2}^p \exp\left(\nu \ d_j \left(D_{\Psi,j} - 1 - \ \frac{\epsilon_0}{2}\right)\right) d\mu_1(\boldsymbol{d}),$$

$$= \quad \infty,$$

where (v) follows as $D_{\Psi,1} > (1 + \epsilon_0)$.

2.2 Proof of Theorem 2.

Sufficient condition For any $\boldsymbol{\eta} := (\eta_1, \ldots, \eta_p) \in \mathbb{R}^p$, define $\boldsymbol{\eta}^+ := (\eta_1^+, \ldots, \eta_p^+)$ where η_j^+ equals η_j when $\eta_j > 0$ and zero otherwise. Define $D_{\boldsymbol{\eta}}$ to be the diagonal matrix with diagonal elements $\boldsymbol{\eta}^+$. Let us consider the following matrices

$$\Psi = \begin{bmatrix} D_{\boldsymbol{\eta}} \\ \mathbf{0}_{n-p,p} \end{bmatrix}, \quad M^{\star} = \begin{bmatrix} \mathbf{I}_{p,p} \\ \mathbf{0}_{n-p,p} \end{bmatrix} \text{ and } V^{\star} = \mathbf{I}_{p}.$$

Note that $\widetilde{M} \in \widetilde{\mathcal{V}}_{n,p}, \widetilde{V} \in \mathcal{V}_{p,p}$ and $D_{\eta} = \widetilde{M}^T \Psi \widetilde{V}$. Now from Definition 2, it follows that

$$\begin{split} \int_{\mathbb{R}_{+}^{p}} g(\boldsymbol{d};\boldsymbol{\nu},\boldsymbol{\eta},n) \, d\boldsymbol{d} &= \int_{\mathbb{R}_{+}^{p}} \frac{\exp(\boldsymbol{\nu}\sum_{j=1}^{p}\eta_{j}d_{j})}{\left[_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)\right]^{\boldsymbol{\nu}}} \, d\mu_{1}(\boldsymbol{d}) \\ &\leq \int_{\mathbb{R}_{+}^{p}} \frac{\exp(\boldsymbol{\nu}\sum_{j=1}^{p}\eta_{j}^{+}d_{j})}{\left[_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)\right]^{\boldsymbol{\nu}}} \, d\mu_{1}(\boldsymbol{d}) \\ &= \int_{\mathbb{R}_{+}^{p}} \frac{etr\left(\boldsymbol{\nu}DD\boldsymbol{\eta}\right)}{\left[_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)\right]^{\boldsymbol{\nu}}} \, d\mu_{1}(\boldsymbol{d}) \\ &= \int_{\mathbb{R}_{+}^{p}} \frac{etr\left(\boldsymbol{\nu}\overline{\boldsymbol{\nu}}D\widetilde{\boldsymbol{M}^{T}}\Psi\right)}{\left[_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)\right]^{\boldsymbol{\nu}}} \, d\mu_{1}(\boldsymbol{d}) \\ &\leq \int_{\mathbb{R}_{+}^{p}} \frac{etr\left(-\boldsymbol{\nu}\epsilon_{0}D\right)}{\left(K_{n,p,\epsilon_{0}}\right)^{\boldsymbol{\nu}}} \, d\mu_{1}(\boldsymbol{d}) \\ &\leq \frac{1}{\left(K_{n,p,\epsilon_{0}}\right)^{\boldsymbol{\nu}}} \prod_{j=1}^{p} \int_{\mathbb{R}_{+}} \exp(-\boldsymbol{\nu}\epsilon_{0}d_{j}) \, dd_{j} \\ &< \infty, \end{split}$$
(2.6)

where the inequality at step (vi) follows from Lemma 7 with an appropriate $\epsilon_0 > 0$.

Necessary condition Let $\eta \in \mathbb{R}^p$ be such that $\max_{j=1,\dots,p} \eta_j \ge 1$. There exists at least one $j \in \{1,\dots,p\}$ such that $\eta_j \ge 1$. Without loss of generality, let us assume that $\eta_1 \ge 1$. From Definition 2, we have

$$\int_{\mathbb{R}^p_+} g(\boldsymbol{d}\,;\,\nu,\boldsymbol{\eta},n) \ d\mu_1(\boldsymbol{d})$$

^

$$= \int_{\mathbb{R}^{+}} \frac{\exp(\nu \sum_{j=1}^{p} \eta_{j} d_{j})}{\left[_{0}F_{1}(\frac{n}{2}, \frac{D^{2}}{4})\right]^{\nu}} d\mu_{1}(\boldsymbol{d})$$

$$\geq \int_{\mathbb{R}^{+}_{+}} \frac{\exp(\nu \sum_{j=1}^{p} \eta_{j} d_{j})}{etr(\nu D)} d\mu_{1}(\boldsymbol{d})$$

$$= \prod_{j=1}^{p} \int_{\mathbb{R}_{+}} \exp(\nu(\eta_{j} - 1)d_{j}) dd_{j}$$

$$= \int_{\mathbb{R}_{+}} \exp(\nu(\eta_{1} - 1)d_{1}) dd_{1} \prod_{j=2}^{p} \int_{\mathbb{R}_{+}} \exp(\nu(\eta_{j} - 1)d_{j}) dd_{j}$$

$$= \infty,$$

where the inequality is due to Lemma 2.

2.3 Proof of Theorem **3**

Proofs of part(a) and part(b) of Theorem 3 follow immediately from Lemma 11 and Lemma 12, respectively.

Lemma 11. The probability density function for the prior distribution of $\mathbf{d} \sim CCPD(\mathbf{d}; \nu, \boldsymbol{\eta})$, denoted by $g(\mathbf{d}; \nu, \boldsymbol{\eta}) := \exp(\nu \, \boldsymbol{\eta}^T \mathbf{d}) / \left[{}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right) \right]^{\nu}$, is log-concave as a function of \mathbf{d} , where D is the diagonal matrix with diagonal elements \mathbf{d} , $\max_{1 \leq j \leq p} \eta_j < 1$, $\nu > 0$ and $n \geq p$.

Proof of Lemma 11.

From Definition 2 we have,

$$g(\boldsymbol{d};\nu,\boldsymbol{\eta}) = \frac{\exp(\nu\,\boldsymbol{\eta}^{T}\boldsymbol{d})}{\left[{}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)\right]^{\nu}},$$

$$\implies \log g(\boldsymbol{d};\nu,\boldsymbol{\eta}) = \nu\,\boldsymbol{\eta}^{T}\boldsymbol{d} - \nu\,\log\left({}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)\right)$$
(2.7)

From Lemma 5, it follows that $-\nu \log \left({}_{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4} \right) \right)$ is a concave function of d. Also, $\nu \eta^{T} d$ is a linear function of d. Therefore from Equation 2.7 it follows that $\log g(d; \nu, \eta)$ is a concave function of d.

Lemma 12. The distribution of d is unimodal if $0 < \eta_j < 1$ for all $j = 1, 2, \dots, p$. The mode of the distribution is characterized by the parameter η and it does not dependent on the parameter ν .

Proof of Lemma 12.

Let $l(\boldsymbol{d}, \nu, \boldsymbol{\eta}) = \log(g(\boldsymbol{d}; \nu, \boldsymbol{\eta}))$. If $\widehat{\boldsymbol{d}}$ is the mode of the distribution then

$$\begin{aligned} \frac{\partial}{\partial d} l(d,\nu,\eta) \Big|_{d=\hat{d}} &= 0, \\ \implies & \nu\eta - \nu \frac{\partial}{\partial d} \log \left({}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right) \right) \Big|_{d=\hat{d}} &= 0, \\ \implies & \frac{\partial}{\partial d} \log \left({}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right) \right) \Big|_{d=\hat{d}} &= \eta, \\ \implies & h(\hat{d}) = \eta, \end{aligned}$$
(2.8)

where $h(\boldsymbol{d}) := (h_1(\boldsymbol{d}), h_2(\boldsymbol{d}), \cdots, h_p(\boldsymbol{d}))$ with

$$h_j(\boldsymbol{d}) := \left(\frac{\partial}{\partial d_j} {}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right)\right) / {}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right),$$

for $j = 1, 2, \dots, p$. The function $h_j(d)$ is strictly increasing in each coordinate as the function ${}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right)$ is log-convex (see Lemma 5). Also, it follows from Lemma 6 that $0 < h_j(d) < 1$ for all $d \in S_p$. Hence Equation 2.8 has a unique solution when $0 < \eta_j < 1$ for all $j = 1, 2, \dots, p$. Also it is clear that the solution does not depend on ν . On the other hand, given any $\hat{d} \in S_p$ we can always find a η satisfying Equation 2.8 such that $0 < \max_{1 \le j \le p} \eta_j < 1$.

2.4 Proof of Theorem **4**

(a) From definitions of unimodality and level sets, we have

$$\left[\frac{g(\boldsymbol{y};\nu,\boldsymbol{\eta})}{g(\boldsymbol{x};\nu,\boldsymbol{\eta})}\right] > 1 \text{ for all } \boldsymbol{y} \in \mathcal{S} \text{ and for all } \boldsymbol{x} \in \mathcal{S}^{c}.$$
(2.9)

Consider the function

$$r(\nu, \boldsymbol{x}) := \int_{\mathcal{S}} \frac{g(\boldsymbol{y}; \nu, \boldsymbol{\eta})}{g(\boldsymbol{x}; \nu, \boldsymbol{\eta})} \, d\boldsymbol{y} = \int_{\mathcal{S}} \left[\frac{g(\boldsymbol{y}; 1, \boldsymbol{\eta})}{g(\boldsymbol{x}; 1, \boldsymbol{\eta})} \right]^{\nu} d\boldsymbol{y}, \tag{2.10}$$

where $\boldsymbol{x} \in \mathcal{S}^c$. Using Equation 2.9 it is easy to see that $\left[\frac{g(\boldsymbol{y};1,\boldsymbol{\eta})}{g(\boldsymbol{x};1,\boldsymbol{\eta})}\right]^{\nu}$ is monotonically increasing in ν for all $\boldsymbol{y} \in \mathcal{S}$. Hence $r(\nu, \boldsymbol{x})$ is an increasing function in ν for any $\boldsymbol{x} \in \mathcal{S}^c$. Note that,

$$\frac{P_{\nu}(\boldsymbol{d}\in\mathcal{S}^{c})}{P_{\nu}(\boldsymbol{d}\in\mathcal{S})} = \frac{\int_{\mathcal{S}^{c}}g(\boldsymbol{x};\nu,\boldsymbol{\eta})\,d\boldsymbol{x}}{\int_{\mathcal{S}}g(\boldsymbol{y};\nu,\boldsymbol{\eta})\,d\boldsymbol{y}} = \int_{\mathcal{S}^{c}}\frac{1}{\int_{\mathcal{S}}\frac{g(\boldsymbol{y};\nu,\boldsymbol{\eta})}{g(\boldsymbol{x};\nu,\boldsymbol{\eta})}\,d\boldsymbol{y}}\,d\boldsymbol{x} = \int_{\mathcal{S}^{c}}\frac{1}{r(\nu,\boldsymbol{x})}\,d\boldsymbol{x}.$$
(2.11)

Hence $P_{\nu}(\boldsymbol{d} \in S^c)/P_{\nu}(\boldsymbol{d} \in S)$ is a decreasing function of ν as $\frac{1}{r(\nu, \boldsymbol{x})}$ is a decreasing function in ν for every $\boldsymbol{x} \in S^c$. Equivalently, $P_{\nu}(\boldsymbol{d} \in S)$ is an increasing function in ν . \Box

(b) Let $\mathbf{d} \sim CCPD(\cdot; \nu, \eta)$ with $0 < \eta_j < 1$ for $j = 1, \dots p$. Let \mathbf{m}_{η} be the mode of the distribution. Note that the value of \mathbf{m}_{η} only depends on the parameter η and does not depend on the parameter ν . Let $f(\mathbf{d}; \nu, \eta)$ be the corresponding probability density function. Hence for the class of distribution functions defined in Definition 2, it follows that,

$$f(\boldsymbol{d};\nu,\boldsymbol{\eta}) = \frac{1}{K_{\nu,\boldsymbol{\eta}}} \frac{\exp(\nu \ \boldsymbol{\eta}^T \boldsymbol{d})}{\left[{}_0F_1\left(\frac{n}{2},\frac{D^2}{4}\right)\right]^{\nu}},\tag{2.12}$$

where $K_{\nu,\boldsymbol{\eta}}$ is the appropriate normalizing constant. Let us define the function $g(\boldsymbol{d};\boldsymbol{\eta}) = \exp(|\boldsymbol{\eta}^T \boldsymbol{d}|)/{}_0F_1\left(\frac{n}{2},\frac{D^2}{4}\right)$. Let $\mathbf{m}_{\boldsymbol{\eta}}$ be the unique mode of the density function $f(\boldsymbol{d};\nu,\boldsymbol{\eta})$ (See Lemma 12). If $\boldsymbol{d} \in \mathbb{R}^p_+$ such that $\boldsymbol{d} \neq \mathbf{m}_{\boldsymbol{\eta}}$, then for any $\lambda \in (0,1)$,

$$\frac{g(\boldsymbol{d};\boldsymbol{\eta})}{g(\lambda\mathbf{m}_{\boldsymbol{\eta}}+(1-\lambda)\boldsymbol{d};\boldsymbol{\eta})} = \frac{\exp(\boldsymbol{\eta}^{T}\boldsymbol{d})}{{}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)} \frac{{}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)}{\exp(\boldsymbol{\eta}^{T}(\lambda\mathbf{m}_{\boldsymbol{\eta}}+(1-\lambda)\boldsymbol{d}))} \\ \stackrel{(vii)}{\leq} \frac{\exp(\boldsymbol{\eta}^{T}\boldsymbol{d})}{{}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)} \frac{\left[{}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)\right]^{\lambda} \left[{}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)\right]^{1-\lambda}}{\exp(\boldsymbol{\eta}^{T}(\lambda\mathbf{m}_{\boldsymbol{\eta}}+(1-\lambda)\boldsymbol{d}))} \\ = \left[\frac{\exp(\boldsymbol{\eta}^{T}\boldsymbol{d})}{{}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)}\right]^{\lambda} \left[\frac{{}_{0}F_{1}\left(\frac{n}{2},\frac{D^{2}}{4}\right)}{\exp(\boldsymbol{\eta}^{T}\mathbf{m}_{\boldsymbol{\eta}})}\right]^{\lambda} \\ = \left[\frac{g(\boldsymbol{d};\boldsymbol{\eta})}{g(\mathbf{m}_{\boldsymbol{\eta}};\boldsymbol{\eta})}\right]^{\lambda}, \quad (2.13)$$

where D_m is the diagonal matrix with diagonal \mathbf{m}_{η} . Inequality (*vii*) follows from the log-convexity of $_0F_1(\cdot)$ (see Lemma 5). As a result,

$$\frac{f(\boldsymbol{d}\,;\boldsymbol{\nu},\boldsymbol{\eta})}{f\left(\lambda\mathbf{m}_{\boldsymbol{\eta}}+(1-\lambda)\boldsymbol{d};\boldsymbol{\nu},\boldsymbol{\eta}\right)} = \left[\frac{g(\boldsymbol{d}\,;\boldsymbol{\eta})}{g\left(\lambda\mathbf{m}_{\boldsymbol{\eta}}+(1-\lambda)\boldsymbol{d};\boldsymbol{\eta}\right)}\right]^{\nu} < \left[\frac{g(\boldsymbol{d},\boldsymbol{\eta})}{g(\mathbf{m}_{\boldsymbol{\eta}};\boldsymbol{\eta})}\right]^{\nu\lambda}.$$
(2.14)

If S is an open set containing \mathbf{m}_{η} then, there exists an open ball $B_{\epsilon} = \{ \boldsymbol{d} \in \mathbb{R}^{p}_{+} : \|\boldsymbol{d} - \mathbf{m}_{\eta}\| < \epsilon \}$, such that $B_{\epsilon} \subset S$ for some $\epsilon > 0$. Let S^{\star} be the complement of the set B_{ϵ} and \overline{B}_{ϵ} be the boundary of the open ball B_{ϵ} , i.e. $\overline{B}_{\epsilon} = \{ \boldsymbol{d} \in \mathbb{R}^{p}_{+} : \|\boldsymbol{d} - \mathbf{m}_{\eta}\| = \epsilon \}$. Note that, $\overline{B}_{\epsilon} \subset S^{\star}$.

Let $\zeta = \sup_{\boldsymbol{d} \in S^*} g(\boldsymbol{d}; \boldsymbol{\eta})$. If $\boldsymbol{d} \in S^* \setminus \overline{B}_{\epsilon}$ then $\|\boldsymbol{d} - \mathbf{m}_{\boldsymbol{\eta}}\| > \epsilon$. Consider the point $\boldsymbol{d}_0 = \lambda_0 \mathbf{m}_{\boldsymbol{\eta}} + (1 - \lambda_0) \boldsymbol{d}$ where $\lambda_0 = 1 - \epsilon / \|\boldsymbol{d} - \mathbf{m}_{\boldsymbol{\eta}}\|$. Observe that $\boldsymbol{d}_0 \in B_{\epsilon}$. As $\boldsymbol{d} \neq \mathbf{m}_{\boldsymbol{\eta}}$, from Equation 2.14, it follows that

$$rac{g(oldsymbol{d}\,;oldsymbol{\eta})}{g\left(oldsymbol{d}_0\,;oldsymbol{\eta}
ight)} < \left[rac{g(oldsymbol{d};oldsymbol{\eta})}{g(\mathbf{m}_{oldsymbol{\eta}};oldsymbol{\eta})}
ight]^{\lambda_0} < 1.$$

Hence for any $\mathbf{d} \in \mathcal{S}^* \setminus \overline{B}_{\epsilon}$, there is a point $\mathbf{d}_0 \in \overline{B}_{\epsilon}$ such that $g(\mathbf{d}; \boldsymbol{\eta}) < g(\mathbf{d}_0; \boldsymbol{\eta})$. Consequently, we get that $\zeta = \sup_{\mathbf{d} \in \mathcal{S}^*} g(\mathbf{d}; \boldsymbol{\eta}) = \sup_{\mathbf{d} \in \overline{B}_{\epsilon}} g(\mathbf{d}; \boldsymbol{\eta})$. As the set \overline{B}_{ϵ} is compact, there exist $\mathbf{d}^{\dagger} \in \overline{B}_{\epsilon}$ such that $\zeta = g(\mathbf{d}^{\dagger}; \boldsymbol{\eta})$. Therefore $\zeta < g(\mathbf{m}_{\boldsymbol{\eta}}; \boldsymbol{\eta})$ as $\mathbf{m}_{\boldsymbol{\eta}}$ is the unique maximizer of the function $g(\mathbf{d}, \boldsymbol{\eta})$ and $\mathbf{d}^{\dagger} \neq \mathbf{m}_{\boldsymbol{\eta}}$. Also,

$$P_{\nu}(\mathcal{S}^{\star}) = \int_{\mathcal{S}^{\star}} f(\boldsymbol{d} ; \nu, \boldsymbol{\eta}) d\mu_{1}(\boldsymbol{d})$$

$$= \int_{\mathcal{S}^{\star}} \frac{f(\boldsymbol{d} ; \nu, \boldsymbol{\eta})}{f(\lambda \mathbf{m}_{\boldsymbol{\eta}} + (1 - \lambda)\boldsymbol{d} ; \nu, \boldsymbol{\eta})} f(\lambda \mathbf{m}_{\boldsymbol{\eta}} + (1 - \lambda)\boldsymbol{d} ; \nu, \boldsymbol{\eta}) d\mu_{1}(\boldsymbol{d})$$

$$\leq \int_{\mathcal{S}^{\star}} \left[\frac{g(\boldsymbol{d}; \boldsymbol{\eta})}{g(\mathbf{m}_{\boldsymbol{\eta}}; \boldsymbol{\eta})} \right]^{\nu\lambda} f(\lambda \mathbf{m}_{\boldsymbol{\eta}} + (1 - \lambda)\boldsymbol{d} ; \nu, \boldsymbol{\eta}) d\mu_{1}(\boldsymbol{d})$$

$$\leq \int_{\mathcal{S}^{\star}} \left[\frac{\zeta}{g(\mathbf{m}_{\boldsymbol{\eta}}; \boldsymbol{\eta})} \right]^{\nu\lambda} f(\lambda \mathbf{m}_{\boldsymbol{\eta}} + (1 - \lambda)\boldsymbol{d} ; \nu, \boldsymbol{\eta}) d\mu_{1}(\boldsymbol{d})$$

$$= \left[\frac{\zeta}{g(\mathbf{m}_{\boldsymbol{\eta}}; \boldsymbol{\eta})} \right]^{\nu\lambda} \int_{\mathcal{S}^{\star}} f(\lambda \mathbf{m}_{\boldsymbol{\eta}} + (1 - \lambda)\boldsymbol{d} ; \nu, \boldsymbol{\eta}) d\mu_{1}(\boldsymbol{d})$$

$$\leq \left[\frac{\zeta}{g(\mathbf{m}_{\boldsymbol{\eta}}; \boldsymbol{\eta})} \right]^{\nu\lambda} \frac{1}{(1 - \lambda)^{p}}.$$

Hence

$$\lim_{\nu \to \infty} P_{\nu}(\mathcal{S}) \ge 1 - \lim_{\nu \to \infty} P_{\nu}(\mathcal{S}^{\star}) \ge 1 - \lim_{\nu \to \infty} \left[\frac{\zeta}{g(\mathbf{m}_{\eta}; \eta)} \right]^{\nu \lambda} \frac{1}{(1 - \lambda)^{p}} = 1.$$

2.5 Proof of Theorem **5**.

From Definition 1, we get that the joint density is proportional to

$$g(M, \boldsymbol{d}, V; \nu, \Psi) = \frac{etr(\nu VDM^{T}\Psi)}{\left[{}_{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right)\right]^{\nu}},$$
(2.15)

Consider the unique SVD of $\Psi = M_{\Psi} D_{\Psi} V_{\Psi}^T$.

We have,

$$etr(\nu VDM^{T}\Psi) = etr(\nu DM^{T}M_{\Psi}D_{\Psi}V_{\Psi}^{T}V)$$

$$= etr(\nu V_{\Psi}^{T}VDU_{M}D_{M}V_{M}^{T}D_{\Psi})$$

$$= etr(\nu V_{1}DU_{M}D_{M}V_{M}^{T}D_{\Psi})$$
(2.16)

where the SVD of $M^T M_{\Psi}$ is written as $M^T M_{\Psi} = U_M D_M V_M^T$ and $V_1 = V_{\Psi}^T V$ is an orthogonal matrix. Therefore,

$$etr(\nu VDM^T\Psi) = etr(\nu V_1D U_M D_M V_M^T D_\Psi)$$

$$\stackrel{(viii)}{\leq} etr(\nu DD_M D_\Psi), \tag{2.17}$$

where the inequality (*viii*) follows from Kristof (1969) (see Theorem on page 5) as V_1, U_M and V_M are orthogonal matrices while D, D_M and D_{Ψ} are diagonal matrices with nonnegative diagonal entries. Using sub-multiplicativity of $\|\cdot\|_2$ (Conway, 1990), we have

$$\|D_M\|_2 = \|U_M^T M^T M_{\Psi} V_M\|_2 \le \|U_M^T\|_2 \|M^T\|_2 \|M_{\Psi}\|_2 \|V_M\|_2 \le 1.$$

Therefore, using Lemma 3, we can infer that all the diagonal entries of D_M are less than or equal to 1. Hence from Equation 2.17, we get

$$etr(\nu VDM^T \Psi) \leq etr(\nu DD_{\Psi}).$$
 (2.18)

Therefore, it follows from Kristof (1969) that $M = M_{\Psi}$ and $V = V_{\Psi}$ are unique maximizers when $M_{\Psi} \in \widetilde{\mathcal{V}}_{n,p}$ and $V_{\Psi} \in \mathcal{V}_{p,p}$. Note that this does not depend on the choice of ν .

In Equation 2.15, replacing $M = M_{\Psi}$ and $V = V_{\Psi}$, we can maximize the function $etr(\nu DD_{\Psi}) / \left[{}_{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right) \right]^{\nu}$ with respect to the variable d. Note that the diagonal elements of D_{Ψ} are between 0 and 1 as $\|\Psi\|_{2} < 1$. Hence using part (b) of Theorem 3 we infer that $d \mapsto etr(\nu DD_{\Psi}) / \left[{}_{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right) \right]^{\nu}$ has a unique maximizer which does not depend on the value of ν .

2.6 Proof of Theorem **6**

(a) The argument is almost identical to part(a) of the proof of Theorem 4 (See Section 2.4).

(b) For any open set \mathcal{A} , there exist $\epsilon > 0$ such that $\mathcal{A}^* \subset \mathcal{A}$ where

$$\mathcal{A}^{\star} = \left\{ (M, \boldsymbol{d}, V) \in \mathcal{V}_{n, p} \times \mathbb{R}^{p}_{+} \times \mathcal{V}_{p, p} : \sqrt{\left\| M - \hat{M} \right\|_{2}^{2} + \left\| \boldsymbol{d} - \hat{\boldsymbol{d}} \right\|^{2} + \left\| V - \hat{V} \right\|_{2}^{2}} < \epsilon \right\}.$$

Let \hat{M} , \hat{V} and \hat{d} be the mode of the distribution and $\Psi = M_{\Psi} D_{\Psi} V_{\Psi}$ be the unique SVD of the matrix Ψ . Let η denotes the vector containing the diagonal elements of D_{Ψ} . From part(a) of Theorem 5 we get $\hat{M} = M_{\Psi}$, $\hat{V} = V_{\Psi}$, and $\hat{d} = \mathbf{m}_{\eta}$ where $\mathbf{h}(\mathbf{m}_{\eta}) = \eta$. Now consider

$$P_{\nu}(\|M - M_{\Psi}\|_{2} > \frac{\epsilon}{3})$$

$$= \iiint_{\{(M,d,V): \|M-M_{\Psi}\|_{2} > \epsilon\}} f(M,d,V;\nu,\Psi) \ d\mu(M) \ d\mu_{1}(d) \ d\mu_{2}(V)$$

$$= \iiint_{\{(M,d,V): \|M-M_{\Psi}\|_{2} > \epsilon\}} \frac{etr(\nu V D M^{T} \Psi)}{{}_{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right)^{\nu} K_{\nu,D_{\Psi}}} \ d\mu(M) \ d\mu_{1}(d) \ d\mu_{2}(V).$$

From Kristof (1969) (see Theorem on page 5) it follows that

$$etr(\nu VDM^T\Psi) = etr(\nu VDM^TM_{\Psi}D_{\Psi}V_{\Psi}^T) \le etr(\nu DD_{\delta}^{\star}D_{\Psi}),$$

where D^{\star}_{δ} is a diagonal matrix with all the diagonal entrees less than or equal to one and at least one of the diagonal elements is less than or equal to $(1 - \delta)$, where $\delta > 0$. Here $\delta > 0$ depends on the choice of $\epsilon > 0$ and Ψ . Without loss of generality, for the rest of the proof we assume that the first diagonal element of D^{\star}_{δ} is $1 - \delta$ and all the other diagonal elements are 1. From Equation 2.22, we get

$$P_{\nu}(\|M - M_{\Psi}\|_{2} \geq \frac{\epsilon}{3})$$

$$\leq \int_{\mathbb{R}^{p}_{+}} \frac{etr(\nu DD_{\delta}^{\star} D_{\Psi})}{{}_{0}F_{1}\left(\frac{n}{2}, \frac{D^{2}}{4}\right)^{\nu} K_{\nu, D_{\Psi}}} d\mu_{1}(\boldsymbol{d})$$

$$= \int_{\mathbb{R}^{p}_{+}} \frac{f(\boldsymbol{d} ; \nu, \boldsymbol{\eta}) exp(-\nu \delta d_{1})}{f(\lambda \mathbf{m}_{\boldsymbol{\eta}} + (1 - \lambda)\boldsymbol{d} ; \nu, \boldsymbol{\eta})} f(\lambda \mathbf{m}_{\boldsymbol{\eta}} + (1 - \lambda)\boldsymbol{d} ; \nu, \boldsymbol{\eta}) d\mu_{1}(\boldsymbol{d})$$

$$\stackrel{(xi)}{\leq} \int_{\mathbb{R}^{p}_{+}} \left[\frac{g(\boldsymbol{d} ; \boldsymbol{\eta}) exp(-\frac{\delta d_{1}}{\lambda})}{g(\mathbf{m}_{\boldsymbol{\eta}} ; \boldsymbol{\eta})} \right]^{\nu \lambda} f(\lambda \mathbf{m}_{\boldsymbol{\eta}} + (1 - \lambda)\boldsymbol{d} ; \nu, \boldsymbol{\eta}) d\mu_{1}(\boldsymbol{d})$$

$$(2.19)$$

where (xi) follows from Equation 2.14. If we denote $\boldsymbol{\eta}^{\star} = \boldsymbol{\eta} - (\frac{\delta}{\lambda}, 0, \dots, 0)^T$, then

$$g(\boldsymbol{d};\boldsymbol{\eta})exp(-\frac{\delta d_1}{\lambda}) = g(\boldsymbol{d};\boldsymbol{\eta}^{\star}) \le g(\mathbf{m}^{\star};\boldsymbol{\eta}^{\star}) = g(\mathbf{m}^{\star};\boldsymbol{\eta})exp(-\frac{\delta m_1^{\star}}{\lambda}),$$
(2.20)

where \mathbf{m}^{\star} is the unique mode of the $CCPD(\cdot; 1, \boldsymbol{\eta}^{\star})$ distribution (see Theorem 3) and m_1^{\star} is the element in the first coordinate of the vector \mathbf{m}^{\star} . m_1^{\star} depends on the value of λ, δ and Ψ . Note that, we can choose ϵ and λ in such a way that $\eta_1 - \delta/\lambda > 0$. Therefore, it can be made sure that $m_1^{\star} > 0$. From Equation 2.19 and Equation 2.20 we get that

$$P_{\nu}(\|M - M_{\Psi}\|_{2} \geq \frac{\epsilon}{3})$$

$$\leq \int_{\mathbb{R}^{p}_{+}} \left[\frac{g(\mathbf{m}_{\eta}^{\star}; \boldsymbol{\eta})}{g(\mathbf{m}_{\eta}; \boldsymbol{\eta})} \right]^{\nu \lambda} exp(-\nu \delta m_{1}^{\star}) f(\lambda \mathbf{m}_{\eta} + (1 - \lambda)d ; \nu, \boldsymbol{\eta}) d\mu_{1}(d).$$

Note that $g(\mathbf{m}_{\boldsymbol{\eta}}^{\star};\boldsymbol{\eta}) < g(\mathbf{m}_{\boldsymbol{\eta}};\boldsymbol{\eta})$ as $\mathbf{m}_{\boldsymbol{\eta}}$ is the unique maximizer of the function $\boldsymbol{d} \rightarrow g(\boldsymbol{d}; \boldsymbol{\eta})$. Also, if we denote $r_{\epsilon,\Psi,\lambda} = exp(-\delta m_1^{\star})$ then $0 < r_{\epsilon,\Psi,\lambda} < 1$ as $m_1^{\star} > 0$. Hence

$$\begin{aligned} P_{\nu}(\|M - M_{\Psi}\|_{2} \geq \frac{\epsilon}{3}) &\leq \int_{\mathbb{R}^{p}_{+}} r^{\nu}_{\epsilon,\Psi,\lambda} f(\lambda \mathbf{m}_{\eta} + (1-\lambda)d \; ; \; \nu, \eta) \; d\mu_{1}(d) \\ &= \frac{r^{\nu}_{\epsilon,\Psi,\lambda}}{(1-\lambda)^{p}}, \end{aligned}$$

where the last equality is obtained using a change of variable while using the fact that $f(\cdot)$ is a probability density function on \mathbb{R}^p_+ . In a similar fashion we obtain

$$P_{\nu}(\|V - V_{\Psi}\|_{2} \ge \frac{\epsilon}{3}) \le \frac{r_{1,\epsilon,\Psi,\lambda}^{\nu}}{(1-\lambda)^{p}}, \text{ for some } 0 < r_{1,\epsilon,\Psi,\lambda} < 1.$$
(2.21)

Additionally, from part(b) of Theorem 4, we get that

$$\lim_{\nu \to \infty} (P_{\nu}(\|\boldsymbol{d} - \mathbf{m}_{\boldsymbol{\eta}}\| \ge \frac{\epsilon}{3}) = 0.$$

Finally, we obtain

$$\lim_{\nu \to \infty} P_{\nu}(\mathcal{A})$$

$$= 1 - \lim_{\nu \to \infty} P_{\nu}(\mathcal{A}^{c})$$

$$\geq 1 - \lim_{\nu \to \infty} P_{\nu}(\mathcal{A}^{\star c})$$

$$\geq 1 - \lim_{\nu \to \infty} (P_{\nu}(\|\boldsymbol{d} - \mathbf{m}_{\boldsymbol{\eta}}\| \ge \frac{\epsilon}{3}) + P_{\nu}(\|V - V_{\Psi}\|_{2} \ge \frac{\epsilon}{3}) + P_{\nu}(\|M - M_{\Psi}\|_{2} \ge \frac{\epsilon}{3})$$

$$\geq 1 - \left(\lim_{\nu \to \infty} \left(\frac{r_{\epsilon,\Psi,\lambda}^{\nu}}{(1-\lambda)^{p}} + \frac{r_{1,\epsilon,\Psi,\lambda}^{\nu}}{(1-\lambda)^{p}}\right)\right)$$

$$= 1.$$

(c) For the *JCPD* distribution, the conditional distribution of M given (d, V) is proportional to

$$etr\left(\nu\left(\Psi VD\right)^{T}M\right)$$
.

This distribution is an \mathcal{ML} distribution with parameters $M_{\Psi}^M, D_{\Psi}^M, V_{\Psi}^M$ where the unique SVD of $\nu (\Psi VD) = M_{\Psi}^M D_{\Psi}^M (V_{\Psi}^M)^T$.

Similarly, the conditional distribution of V given M and d is proportional to

$$etr\left(\nu\left(\Psi^{T}MD\right)^{T}V\right)\right).$$

Therefore, it too is an \mathcal{ML} distribution with parameters $M_{\Psi}^V, D_{\Psi}^V, V_{\Psi}^V$ where the unique SVD of $\nu (\Psi^T M D) = M_{\Psi}^V D_{\Psi}^V (V_{\Psi}^V)^T$.

Finally, the conditional distribution of d given (M, V) is a distribution that belongs to the *CCPC* class of distributions with parameters ν and η_{Ψ} , where $\eta_{\Psi} = (\eta_{\Psi_1}, \eta_{\Psi_2}, \cdots, \eta_{\Psi_p})$. Here η_{Ψ_j} is the *j*-th diagonal element of the matrix $M^T \Psi V$ for $j = 1, \ldots p$.

2.7 Proof of Theorem 7

Proof of Theorem 7 follows immediately from Jupp and Mardia (1979) (see Proposition and Corollary on page 601 in Jupp and Mardia (1979)). For the sake of completeness, we include the arguments here.

Let W_1, \ldots, W_N be independent and identically distributed samples from an \mathcal{ML} distribution on the space $\mathcal{V}_{n,p}$. According to Proposition 2 in Jupp and Mardia (1979), \overline{W} has a density (i.e. absolutely continuous) with respect to Lebesgue measure on $\mathbb{R}^{n \times p}$ if $N \geq 2$, p < n or $N \geq 3$, $p = n \geq 3$. Consider that,

$$\left\|\overline{W}\right\|_{2} \le \frac{1}{n} \sum_{i=1}^{n} \|W_{i}\|_{2} = 1.$$

Hence $P(\overline{W} \in \{X \in \mathbb{R}^{n \times p} : ||X||_2 \le 1\}) = 1$. As Lebesgue measure on the set $\{X \in \mathbb{R}^{n \times p} : ||X||_2 = 1\}$ is zero, $P(\overline{W} \in \{X \in \mathbb{R}^{n \times p} : ||X||_2 = 1\} = 0$. As a result,

$$P(\overline{W} \in \{X \in \mathbb{R}^{n \times p} : \|X\|_2 < 1\}) = 1.$$

	. 1	

2.8 Proof of Theorem 8

Let $\overline{W} = \frac{1}{N} \sum_{i=1}^{N} W_i$ where W_1, \ldots, W_N are independent and identically distributed samples from an \mathcal{ML} -distribution on the space $\mathcal{V}_{n,p}$. If $Z_i = W_i - E(W_1)$ then

$$||Z_i||_2 = ||W_i - E(W_1)||_2 \le ||W_i||_2 + ||E(W_1)||_2 = 2.$$
(2.22)

Note that for all $i \in \{1, ..., N\}$ and for arbitrary $l \in \mathbb{R}^p$ such that $l^T l = 1$,

$$l^{T} E\left(Z_{i}^{T} Z_{i}\right) l = l^{T} E\left((W_{i} - E(W_{1}))^{T}(W_{i} - E(W_{1}))\right) l = E\left(l^{T} W_{i}^{T} W_{i} l\right) - l^{T} E(W_{i})^{T} E(W_{i}) l$$

$$= E\left(l^{T} I_{p \times p} l\right) - \|E(W_{i})l\|^{2}$$

$$= 1 - \|E(W_{i})l\|^{2} \leq 1,$$

as $||E(W_i)l||^2 > 0$. Consequently, for $i = 1, ..., N ||E(Z_i^T Z_i)||_2 < 1$, implying the fact that

$$\left\| E(\sum_{i=1}^{N} Z_{i}^{T} Z_{i}) \right\|_{2} \leq \sum_{i=1}^{N} \left\| E(Z_{i}^{T} Z_{i}) \right\|_{2} \leq N.$$
(2.23)

Similarly, we get that for all $i \in \{1, ..., N\}$ and for arbitrary $l_{\star} \in \mathbb{R}^n$ such that $l_{\star}^T l_{\star} = 1$,

$$l_{\star}^{T} E\left(Z_{i} Z_{i}^{T}\right) l_{\star} = E\left((W_{i} - E(W_{1}))(W_{i} - E(W_{1}))^{T}\right) = E\left(l^{T} W_{i} W_{i}^{T} l\right) - l^{T} E(W_{i}) E(W_{i})^{T} l_{\star}^{(\dagger\dagger)} \leq 1 - \|E(W_{i})^{T} l\|^{2} \leq 1.$$

Note that the (\dagger †) step of the inequality follows since the matrix $W_i W_i^T$ being real symmetric and idempotent, is a orthogonal projection matrix. Therefore, for i = 1, ..., N $\|E(Z_i Z_i^T)\|_2 < 1$, implying the fact that

$$\left\| E(\sum_{i=1}^{N} Z_{i} Z_{i}^{T}) \right\|_{2} \leq \sum_{i=1}^{N} \left\| E(Z_{i} Z_{i}^{T}) \right\|_{2} \leq N.$$
(2.24)

From Equation 2.23 and 2.24 we get that

$$\sigma_{\star}^{2} = max \left\{ \left\| E(\sum_{i=1}^{N} Z_{i}^{T} Z_{i}) \right\|_{2}, \left\| E(\sum_{i=1}^{N} Z_{i} Z_{i}^{T}) \right\|_{2} \right\} \le N.$$
(2.25)

Using Equations 2.22, 2.25, For arbitrary $\epsilon > 0$, we now apply the matrix Bernstein concentration inequality (see page 928 in Mackey et al. (2014)) to obtain that,

$$P\left(\left\|\overline{W} - E(W_{1})\right\|_{2} \ge \epsilon\right) = P\left(\left\|\sum_{i=1}^{N} Z_{i}\right\|_{2} \ge N\epsilon\right)$$

$$\leq (n+p) \exp\left(-\frac{\epsilon^{2} N^{2}}{3\sigma_{\star}^{2} + 4N\epsilon}\right),$$

$$\leq (n+p) \exp\left(-\frac{\epsilon^{2} N^{2}}{3N + 4N\epsilon}\right),$$

$$\leq (n+p) \exp\left(-\frac{\epsilon^{2} N}{3 + 4\epsilon}\right). \quad (2.26)$$

Using Borel-Cantelli Lemma (Billingsley, 1995), it follows that

$$\overline{W} \xrightarrow{a.s.} E(W_1)$$
 as $N \longrightarrow \infty$.

Consequently

$$\widehat{\Psi}_N = \left(\frac{\nu}{\nu+N}\Psi + \frac{N}{\nu+N}\overline{W}\right) \xrightarrow{a.s.} E(W_1) \text{ as } N \longrightarrow \infty.$$
(2.27)

Let $\hat{\eta}_{\Psi_N}$ be the diagonal elements of the diagonal matrix \hat{D}_{Ψ_N} , where $\hat{\Psi}_N = \hat{M}_N \hat{D}_{\Psi_N} \hat{V}_N$ is the unique SVD for $\hat{\Psi}_N$. Using Theorem 5, we get that the posterior mode for parameters M and V are \hat{M}_N and \hat{V}_N , respectively. For the parameter d, the posterior mode is \hat{d}_N where $\mathbf{h}(\hat{d}_N) = \hat{\eta}_{\Psi_N}$ and the function $\mathbf{h}(\cdot)$ is as defined in Equation 2.8. Note that the inverse function of $\mathbf{h}(\cdot)$ exists as $\mathbf{h}(\cdot)$ is strictly increasing in each coordinate. As a matter of convenience, we write $\hat{d}_N = \mathbf{h}^{-1}(\hat{\eta}_{\Psi_N})$. From Lemma 1, we get

$$E(W_1) = M D_{\mathbf{h}} V^T, \qquad (2.28)$$

where $D_{\mathbf{h}}$ is a diagonal matrix with diagonal entrees $\mathbf{h}(d)$. As the unique SVD (see Chikuse (2012)) is a continuous transformation, from Equation 2.27 and Equation 2.28, we get

$$(\hat{M}_N, \hat{\eta}_{\Psi_N}, \hat{V}_N) \xrightarrow{a.s.} (M, \mathbf{h}(d), V) \text{ as } N \longrightarrow \infty.$$
 (2.29)

Also, since $\mathbf{h}^{-1}(\cdot)$ is a continuous function, from Equation 2.29 we obtain that

$$\hat{\boldsymbol{d}}_N = \mathbf{h}^{-1} \left(\hat{\boldsymbol{\eta}}_{\Psi_N} \right) \xrightarrow{a.s.} \mathbf{h}^{-1} \left(\mathbf{h}(\boldsymbol{d}) \right) = \boldsymbol{d} \text{ as } N \longrightarrow \infty.$$
 (2.30)

As a result, the statistic \hat{M}_N , \hat{D}_N and \hat{V}_N are consistent estimators for the parameters M, d and V.

2.9 Proof of Theorem 9

Before the proof of Theorem 9, we establish Lemma 13 which is required for the proof.

Lemma 13. Let $\mathbf{d} \sim CCPD(\cdot;\nu,\eta)$ for some $\nu > 0$ and $\eta = (\eta_1,\ldots,\eta_p)$ where $\max_{1 \leq j \leq p} \eta_j < 1$. Let m be the mode of the $CCPD_1^*(\cdot;\mathbf{d}^{(-1)},\nu,\eta)$, the conditional distribution of d_1 given (d_2,\ldots,d_p) . If b > 0 then the function $Q(d_1) = g_1(d_1+b)/g_1(d_1)$ is strictly decreasing, where $g_1(\cdot) := g_1(\cdot;\mathbf{d}^{(-1)},\nu,\eta)$.

Proof of Lemma 13.

From Definition 5, we get that,

$$\log(g_1(d_1)) = \nu \eta_1 d_1 - \nu \log\left({}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right)\right)$$
$$\implies \frac{\partial^2}{\partial d_1^2} \left(\log g_1(d_1)\right) = -\nu \frac{\partial^2}{\partial d_1^2} \left(\log\left({}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right)\right)\right) < 0, \qquad (2.31)$$

as $\nu > 0$ and $\log \left({}_0F_1\left({n \over 2}, {D^2 \over 4} \right) \right)$ is a strictly convex function (from Lemma 5). Therefore ${\partial \over \partial d_1} \left(\log g_1(d_1) \right) = g'_1(d_1)/g_1(d_1)$ is a strictly decreasing function in d_1 . Consequently,

$$\frac{\partial}{\partial d_1} \left(\log Q(d_1) \right) = \frac{g_1'(d_1+b)}{g(d_1+b)} - \frac{g_1'(d_1)}{g_1(d_1)} < 0,$$

as b > 0. Therefore, $Q(d_1)$ is also a strictly decreasing function in d_1 .

Proof of Theorem 9.

The proofs of part (a) and part (b) follow immediately from the proof of Theorem 3.

(c) We use the notation $g_1(d_1) =: g_1(d_1; d^{(-1)}, \nu, \eta)$ for brevity.

Note that the unnormalized conditional density of the random variable d_1 given $d^{(-1)}$ is proportional to

$$g_1(d_1) = \frac{exp(\nu \eta_1 d_1)}{{}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right)^{\nu}}.$$

Let $f(d_1; \nu, \eta \mid (d_2, \ldots, d_p))$ be the density function for the conditional distribution of d_1 given (d_2, \ldots, d_p) . For notational convenience, for rest of this theorem we use $f_1(\cdot)$ as the conditional probability density function. Hence we have,

$$f_1(d_1) = \frac{1}{K_{\nu,\eta}^1} \frac{\exp\left(\nu \eta_1 d_1\right)}{{}_0F_1\left(\frac{n}{2}, \frac{D^2}{4}\right)^\nu},$$

where $K_{\nu,\eta}^1$ is an appropriate normalizing constant. From Lemma 13, it follows that $f_1(B+x)/f_1(m+x)$ is a decreasing function of x when B > m. Hence for all x > 0,

$$\frac{f_1(B+x)}{f_1(m+x)} = \frac{g_1(B+x)}{g_1(m+x)} < \frac{g_1(B)}{g_1(m)} \stackrel{(viii)}{<} \epsilon,$$

where the inequality at (viii) follows due to the assumption of the lemma. Therefore,

$$P(d_1 > B \mid (d_2, ..., d_p)) = \int_B^\infty f_1(y) dy$$

= $\int_0^\infty \frac{f_1(B+x)}{f_1(m+x)} f_1(m+x) dx$
< $\epsilon P (d_1 > m \mid (d_2, ..., d_p))$
< $\epsilon.$

(d) Proof of part(d) of the Theorem follows immediately from Lemma 10.

References

Billingsley, P. (1995). Probability and Measure. Wiley Series in Probability and Statistics. Wiley.

- Chikuse, Y. (2012). *Statistics on Special Manifolds*, volume 174. Springer Science & Business Media.
- Conway, J. B. (1990). A Course in Functional Analysis, volume 96 of Graduate Texts in Mathematics. New York: Springer-Verlag, 2nd ed edition.
- Gross, K. I. and Richards, D. S. P. (1987). "Special functions of matrix argument. I. Algebraic induction, zonal polynomials, and hypergeometric functions." *Transactions* of the American Mathematical Society, 301(2): 781–811.

- Hardy, G. H., Littlewood, J. E., and Pólya, G. (1952). *Inequalities*. Cambridge University Press.
- Hoff, P. D. (2009). "Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate and relational data." *Journal of Computational and Graphical Statistics*, 18(2): 438–456.
- Jupp, P. E. and Mardia, K. V. (1979). "Maximum likelihood estimators for the matrix von Mises-Fisher and Bingham distributions." The Annals of Statistics, 599–606.
- 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.
- Kristof, W. (1969). "A theorem on the trace of certain matrix products and some applications." *ETS Research Report Series*, 1969(1).
- Mackey, L., Jordan, M. I., Chen, R. Y., Farrell, B., and Tropp, J. A. (2014). "Matrix concentration inequalities via the method of exchangeable pairs." *The Annals of Probability*, 42(3): 906–945.
- Muirhead, R. J. (2009). Aspects of Multivariate Statistical Theory, volume 197. John Wiley & Sons.
- Richards, D. S. P. (2011). "High-dimensional random matrices from the classical matrix groups, and generalized hypergeometric functions of matrix argument." Symmetry, 3(3): 600–610.
- Segura, J. (2011). "Bounds for ratios of modified Bessel functions and associated Turántype inequalities." Journal of Mathematical Analysis and Applications, 374(2): 516– 528.
- Sengupta, S. (2013). "Two models involving Bayesian nonparametric techniques." Ph.D. thesis, University of Florida.