



# A geometric framework for asymptotic inference of principal subspaces in PCA

Dimbihery Rabenoro, Xavier Pennec

## ► To cite this version:

Dimbihery Rabenoro, Xavier Pennec. A geometric framework for asymptotic inference of principal subspaces in PCA. 2022. hal-03842125

**HAL Id: hal-03842125**

**<https://inria.hal.science/hal-03842125>**

Preprint submitted on 7 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A geometric framework for asymptotic inference of principal subspaces in PCA

Dimbihery Rabenoro<sup>†</sup> and Xavier Pennec<sup>†</sup>

<sup>†</sup>Université Côte d’Azur and INRIA, Epione team

September 5, 2022

## Abstract

In this article, we develop an asymptotic method for testing hypothesis on the set of all linear subspaces arising from PCA and for constructing confidence regions for this set. This procedure is derived from intrinsic estimation in each Grassmannian, endowed with a structure of Riemannian manifold, to which each of these subspaces belong.

## 1 Introduction

Let  $X$  be Gaussian a random vector valued in  $\mathbb{R}^d$ ,  $d \geq 1$ , with covariance matrix  $\Sigma$ . Denote by  $\lambda_1 > \dots > \lambda_r$  the *different* eigenvalues of  $\Sigma$  and by  $\mathbf{I} = (q_i)_{1 \leq i \leq r}$  their respective multiplicities. The sequence  $\mathbf{I}$  induces a partition of the indices  $(1, \dots, d)$  into  $r$  parts  $(B_i)_{1 \leq i \leq r}$ , of respective sizes  $(q_i)_{1 \leq i \leq r}$ . For example,  $B_1 = (1, \dots, q_1)$  and  $B_r = (q_{r-1} + 1, \dots, d)$ . In PCA, the principal subspaces (PS’s) are defined as the eigenspaces of  $\Sigma$ , denoted by  $(P_i(\Sigma))_{1 \leq i \leq r}$ . Now,  $\Sigma$  is estimated by the sample covariance matrices  $(\hat{\Sigma}_n)_{n \geq 1}$  from iid samples of  $X$ . Recall that almost surely, for all  $n \geq 1$ ,  $\hat{\Sigma}_n$  has  $d$  distinct (random) eigenvalues  $\mu_1(\hat{\Sigma}_n) > \dots > \mu_d(\hat{\Sigma}_n)$ . To estimate the eigenspaces of  $\Sigma$ , these eigenvalues need to be regrouped according to the sequence  $\mathbf{I}$ : the PS’s are estimated by the eigenspaces  $(P_i(\hat{\Sigma}_n))_{1 \leq i \leq r}$  which are the linear subspaces spanned by the eigenvectors of  $\hat{\Sigma}_n$  associated to its eigenvalues in  $M_n^i := \{\mu_k(\hat{\Sigma}_n) : k \in B_i\}$ .

The landmark papers (Anderson, 1963) and (Tyler, 1981) investigated the inference of a *single* PS from the asymptotic distribution of statistics involving the eigenvectors of  $\hat{\Sigma}_n$ . The inference obtained for a PS of dimension  $q = 1$  in (Anderson, 1963) was generalized to the case of one PS of any dimension  $q \geq 1$  in (Tyler, 1981) using linear algebraic methods. In contrast, we introduce in this paper a *geometric setting* which allow us to derive the inference of *all PS’s together*, for any sequence of dimensions. Namely, a given PS belongs to a Grassmannian, while the collection  $F^{\mathbf{I}}(\Sigma) := (P_i(\Sigma))_{1 \leq i \leq r}$  of them lies in a Flag manifold. Here, a flag is viewed as a collection  $\mathcal{P} = (P_i)_{1 \leq i \leq r}$  of mutually orthogonal subspaces spanning  $\mathbb{R}^d$ , indexed by its type, that is the sequence  $\mathbf{I} = (q_i)_{1 \leq i \leq r}$  of dimensions of the  $P_i$ ’s. This incremental subspaces representation is equivalent to the more classical one as properly nested sequence of linear subspaces. We denote by  $\mathcal{F}^{\mathbf{I}}$  the set of all flags of type  $\mathbf{I}$ . In order to estimate the PS’s, the idea is then to endow Grassmannians and flags of given type with a structure of Riemannian manifold, allowing to perform *intrinsic* estimation, based on Central Limit Theorems (CLT’s), as in (Bhattacharya and Patrangenaru, 2005). It is well-known that  $\mathcal{F}^{\mathbf{I}}$  is diffeomorphic to  $O(d)/O(\mathbf{I})$ , where  $O(d)$  is the orthogonal group of  $\mathbb{R}^d$  and  $O(\mathbf{I}) := O(q_1) \times \dots \times O(q_r)$ . While in (Tyler, 1981) each PS is estimated separately, our geometric approach enables us to estimate the whole flag  $F^{\mathbf{I}}(\Sigma)$  from the limiting distribution of the sample flag  $F^{\mathbf{I}}(\hat{\Sigma}_n) := (P_i(\hat{\Sigma}_n))_{1 \leq i \leq r}$ .

In order to obtain *Confidence Regions* for  $F^{\mathbf{I}}(\Sigma)$  in  $\mathcal{F}^{\mathbf{I}}$ , a first method could be to establish a CLT in  $\mathcal{F}^{\mathbf{I}}$  in normal coordinates, following (Bhattacharya and Patrangenaru, 2005). However, the Riemannian Logarithm and even the geodesic distance are *unknown* in closed form in  $\mathcal{F}^{\mathbf{I}}$ . The representation of a flag by its incremental subspaces is a key point of our method: it allows embedding the flag space  $\mathcal{F}^{\mathbf{I}}$  in the product of Grassmannians  $G^{\mathbf{I}}$ , for which the geometry is much simpler and the Logarithm is available in closed form. We establish CLT’s

in Grassmannians for each  $P_i(\widehat{\Sigma}_n)$  and we prove that their rate of convergence to a Gaussian distribution are of order  $\frac{1}{\sqrt{n}}$ . Thanks to the fact that the tangent space to all Grassmannians can be embedded in the same symmetric matrix space with the projector representation, we then derive the estimation of the whole flag  $F^I(\Sigma)$ . Namely, we prove a result of the form

$$n\Delta_{\text{Spec}(\Sigma)}\left(F^I(\Sigma), F^I(\widehat{\Sigma}_n)\right) \xrightarrow[n \rightarrow \infty]{d} \chi_D^2, \quad (1.1)$$

for some  $D > 0$ . Here,  $\Delta_{\text{Spec}(\Sigma)}(\cdot, \cdot)$  can be viewed as a discrepancy function on  $\mathcal{F}^I$ , indexed by the spectrum of  $\Sigma$ . The convergence in (1.1) is proved rigorously and *checked numerically* on synthetic experiments. In an estimation setting, the unknown eigenvalues of  $\Sigma$  are replaced by the block-means of eigenvalues of the empirical matrix  $\widehat{\Sigma}_n$ . Thanks to the consistency of this estimator, the convergence to the  $\chi_D^2$  distribution is preserved, and this provides *confidence regions* for  $F^I(\Sigma)$ .

This article is organized as follows. In Section 2, we develop the background of geometry of Grassmannians and Flag manifolds needed for the asymptotic inference methods derived in Section 3.

## 2 Grassmannian and Flags

For Riemannian manifolds, we refer to (Lee, 1963). Our reference for the geometry of Grassmannians is (Bendokat Zimmermann and Absil, 2020) and for that of Flag manifolds is (Ye Wong and Lim, 2022). The key formula for the closed form for the Riemannian Logarithm for Grassmannians (Theorem 1 below) is due to (Batzies Huper Machado and Silva Leite, 2015).

### 2.1 The Grassmannian manifold

For  $1 \leq q < d$ , the Grassmannian  $G(q, d)$  is by definition the set of all linear subspaces of dimension  $q$  of  $\mathbb{R}^d$ . In order to perform calculations, it is identified with the set of orthogonal projectors of rank  $q$ :

$$G(q, d) \simeq \{P \in \mathbb{R}^{d \times d} : P' = P, P^2 = P, \text{rank}(P) = q\}. \quad (2.1)$$

Then, we introduce the Stiefel manifold  $\text{St}(q, d)$ . As a set, it is the set of all  $d \times q$  matrices whose columns are orthonormal vectors. Denoting by  $I_d$  the identity matrix of order  $d$ ,

$$\text{St}(q, d) = \{Y \in \mathbb{R}^{d \times q} : Y'Y = I_d\}.$$

The sets  $G(q, d)$  and  $\text{St}(q, d)$  are linked by the map  $\pi_q^{SG} : \text{St}(q, d) \longrightarrow G(q, d)$ , which associates to any set of  $q$  orthonormal vectors, the projector whose range is the subspace which they span, i.e.:

$$\pi_q^{SG}(Y) = YY'.$$

#### 2.1.1 Action of the orthogonal group

The orthogonal group  $O(d)$  acts on  $G(q, d)$  by

$$(Q, P) \longmapsto QPQ', \quad \text{for } Q \in O(d) \text{ and } P \in G(q, d). \quad (2.2)$$

Let  $P_0 \in G(q, d)$  whose range is spanned by the  $q$  first vectors of the standard basis of  $\mathbb{R}^d$ . Let  $\pi_q^{OG}$  be the map defined by

$$\pi_q^{OG} : O(d) \longrightarrow G(q, d) \quad , \quad \pi_q^{OG}(Q) = QP_0Q'.$$

Fix  $P \in G(q, d)$ . Then, for any  $Q \in O(d)$  whose  $q$  first columns span the range of  $P$ , we have that  $\pi_q^{OG}(Q) = P$ . So, the orbit of  $P_0$  is the whole  $G(q, d)$ . On the other hand, the stabilizer of  $P_0$  is identified with  $O(q) \times O(d - q)$ , so that  $\pi_q^{OG}$  induces a bijection

$$\widetilde{\pi}_q^{OG} : O(d)/(O(q) \times O(d - q)) \simeq G(q, d).$$

Now,  $O(d)/(O(q) \times O(d - q))$  is endowed with the manifold structure such that the canonical quotient map from  $O(d)$  to  $O(d)/(O(q) \times O(d - q))$  is a smooth submersion, i.e. with surjective differential at every point. Then, the manifold structure on  $G(q, d)$  is the one for which  $\widetilde{\pi}_q^{OG}$  is a diffeomorphism.

### 2.1.2 Tangent space

By definition, the Lie algebra of  $O(d)$  is the tangent space  $T_{I_d}O(d)$ . It is the set  $\mathfrak{so}(d)$  of skew-symmetric matrices. Now, for any  $Q \in O(d)$ ,

$$T_Q O(d) = \{Q\tilde{\Omega} : \tilde{\Omega} \in \mathfrak{so}(d)\}.$$

Then, the tangent space at  $P \in G(q, d)$  is the linear subspace of  $\text{Sym}_d$  defined by

$$T_P G(q, d) = \{\Delta \in \text{Sym}_d : \Delta P + P\Delta = \Delta\}.$$

## 2.2 Riemannian geometry of the Grassmannian

### 2.2.1 Metric

A *metric* on a smooth manifold  $M$  is a collection  $g = (g_p)_{p \in M}$  of inner products on the tangent spaces  $T_p M$ , varying smoothly wrt  $p$ . We say that  $(M, g)$  is a *Riemannian manifold* and we assume in the sequel that  $M$  is *connected*. For  $v \in T_p M$ , its length is  $\|v\|_p := (g_p(v, v))^{1/2}$ . For  $p, q \in M$ , a  $C^1$  curve between  $p$  and  $q$  is a  $C^1$  map  $\gamma : [a, b] \rightarrow M$  such that  $\gamma(a) = p$  and  $\gamma(b) = q$ . Its length is  $\int_a^b \left\| \frac{d}{dt} \gamma(t) \right\|_{\gamma(t)} dt$ . Then,  $g$  defines the *geodesic distance*  $d_g$ , where for any  $p, q \in M$ ,  $d_g(p, q)$  is the infimum of lengths of all  $C^1$  curves between  $p$  and  $q$ .

Let  $\pi : M \rightarrow N$  be a smooth submersion. A metric  $g$  on  $M$  allows to split any tangent space  $T_p M$  into a vertical part  $\text{Ver}_p^\pi M := \ker(d\pi_p)$  and a horizontal part  $\text{Hor}_p^\pi M$ , which is the orthogonal complement of  $\text{Ver}_p^\pi M$  in  $T_p M$  wrt  $g_p$ . We say that  $\pi$  is a *Riemannian submersion* when, for all  $p \in M$ , the restriction of  $d\pi_p$  to  $\text{Hor}_p^\pi M$  is an isometry.

As a compact Lie group,  $O(d)$  is endowed with its canonical metric  $g^O$  defined by

$$g_Q^O(\Omega_1, \Omega_2) = \frac{1}{2} \text{tr}(\Omega_1' \Omega_2), \quad \text{for } Q \in O(d), \Omega_i \in T_Q O(d).$$

We endow  $G(q, d)$  with the unique metric  $g^G$  for which  $\pi^{OG}$  is a *Riemannian submersion*. Thus, for  $P \in G(q, d)$  and  $\Delta_1, \Delta_2 \in T_P G(q, d)$ ,

$$g_P^G(\Delta_1, \Delta_2) = g_Q^O(\Delta_{1,Q}^h, \Delta_{2,Q}^h)$$

where  $\Delta_{i,Q}^h$  is the horizontal lift at  $Q$  of  $\Delta_i$ , i.e.  $\Delta_{i,Q}^h \in \text{Hor}_Q^O O(d)$  and  $d\pi_Q^{OG}(\Delta_{i,Q}^h) = \Delta_i$ .

**Lemma 1.** *The action of  $O(d)$  on  $G(q, d)$  defined above is an isometry. This means that, for any  $P, R \in G(q, d)$  and  $Q \in O(d)$ ,*

$$d_g^G(QPQ', QRQ') = d_g^G(P, R),$$

where  $d_g^G$  denotes the geodesic distance on  $G(q, d)$  associated to  $g^G$ .

### 2.2.2 Geodesics

Let  $\gamma$  be a smooth curve in  $M$ . If  $M$  is a submanifold of  $\mathbb{R}^n$ , then the acceleration  $\ddot{\gamma}(t)$  is not always a tangent vector at  $\gamma(t)$ . On a Riemannian manifold  $(M, g)$ , one builds from the metric an operation of derivation of vector fields whose outputs are tangent vectors, called the *covariant derivative* and denoted by  $\nabla$ . Thus, if  $\gamma$  is in  $(M, g)$ , the acceleration vector of  $\gamma$  at  $\gamma(t)$  is the vector of  $T_{\gamma(t)} M$  equals to the covariant derivative of the velocity  $\dot{\gamma}$  at  $\gamma(t)$  in the direction of  $\dot{\gamma}(t)$ , denoted by  $(\nabla_{\dot{\gamma}} \dot{\gamma})_{\gamma(t)}$ . Then, a *geodesic* is a smooth curve  $\gamma$  on  $M$  of zero acceleration, i.e. such that

$$\nabla_{\dot{\gamma}} \dot{\gamma} = 0.$$

This equation is locally an ODE, so that for any  $p \in M$  and  $v \in T_p M$ , there exists a unique geodesic  $\gamma_{p,v}$  such that  $\gamma_{p,v}(0) = p$  and  $\dot{\gamma}_{p,v}(0) = v$ . By the *Hopf-Rinow theorem*, the metric space  $(M, d_g)$  is complete if and only if all geodesics are defined on  $\mathbb{R}$ . In that case, for any  $p, q \in M$ , there exists at least one geodesic of minimal length between  $p$  and  $q$ . Given a Riemannian submersion  $\pi : M \rightarrow N$ , the geodesics in  $N$  are the images by  $\pi$  of the geodesics with horizontal tangent vectors in  $M$ . Since  $\pi^{OG}$  is a Riemannian submersion, the geodesics in  $G(q, d)$  are given by those in  $O(d)$ . The geodesic in  $O(d)$  starting at  $Q$  with initial velocity  $Q\tilde{\Omega}$  is

$$t \mapsto Q \exp_m(t\tilde{\Omega}),$$

where  $\exp_m$  is the matrix exponential. So, the geodesic in  $G(q, d)$  starting at  $P$  with initial velocity  $\Delta$  is

$$t \mapsto \pi^{OG} \left( Q \exp_m(t\tilde{\Omega}) \right) = Q \exp_m(t\tilde{\Omega}) P_0 \exp_m(-t\tilde{\Omega}) Q', \quad (2.3)$$

for any  $Q \in (\pi^{OG})^{-1}(P)$  and where  $Q\tilde{\Omega} = \Delta_Q^h$ . Therefore,  $(G(q, d), g^O)$  is *complete*.

### 2.2.3 Exponential map

In the sequel,  $(M, g)$  is a complete Riemannian manifold. For any  $p \in M$ , the Riemannian exponential at  $p$  is the map  $\text{Exp}_p^M$  from  $T_p M$  to  $M$  defined by

$$\text{Exp}_p^M(v) = \gamma_{p,v}(1).$$

**Lemma 2.** *By (2.3), for any  $P \in G(q, d)$  and  $\Delta \in T_P G(q, d)$ ,*

$$\text{Exp}_P^{G(q,d)}(\Delta) = \exp_m([\Delta, P]) P \exp_m(-[\Delta, P]).$$

### 2.2.4 Cut locus

It is known that, for all  $p \in M$ , any geodesic starting at  $p$  is locally length-minimizing. The *cut locus* of  $p \in M$  is the set  $\text{Cut}(p)$  of points  $q$  such that the geodesics starting at  $p$  cease to be length-minimizing beyond  $q$ . It is established that if  $q \notin \text{Cut}(p)$ , then there exists a unique length-minimizing geodesic between  $p$  and  $q$ . For example, on a sphere, the geodesics are the great circles, so that two antipodal points are each other's cut locus.

**Lemma 3.** *Let  $P \in G(q, d)$ . Then, writing  $P = YY'$ ,*

$$\text{Cut}(P) = \{R = ZZ' \in G(q, d) : \text{rank}(Y'Z) < q\}.$$

*This means that this characterization is independent of  $Y \in \text{St}(q, d)$  such that  $P = YY'$ .*

### 2.2.5 Riemannian Logarithm

Let  $p \in M$  and  $q \in M \setminus \text{Cut}(p)$ . Then, there exists a unique shortest geodesic between  $p$  and  $q$ , whose initial velocity is thus the smallest tangent vector  $v \in T_p M$  such that  $\text{Exp}_p^M(v) = q$ . It is called the *Riemannian Logarithm* of  $q$  at  $p$ , denoted by  $\text{Log}_p^M(q)$ . We have that

$$d_g(p, q) = \|\text{Log}_p^M(q)\|_p. \quad (2.4)$$

For  $G(q, d)$ , combining (2.4), Lemmas 3 and 1 yield the following result.

**Proposition 1.** *For all  $P, R \in G(q, d)$  and  $Q \in O(d)$ ,*

$$Q R Q' \notin \text{Cut}(Q P Q') \iff R \notin \text{Cut}(P), \quad (2.5)$$

*and in that case,*

$$\left\| \text{Log}_{Q P Q'}^{G(q,d)}(Q R Q') \right\|_{Q P Q'} = \left\| \text{Log}_P^{G(q,d)}(R) \right\|_P. \quad (2.6)$$

Finally, the next result provides the closed form for the Riemannian Logarithm for  $G(q, d)$ .

**Theorem 1.** *Let  $P \in G(q, d)$  and  $R \in G(q, d) \setminus \text{Cut}(P)$ . Then,  $\text{Log}_P^{G(q,d)}(R) = [\Omega, P]$ , where  $\Omega$  is determined by*

$$\Omega = \frac{1}{2} \log_m((I_d - 2R)(I_d - 2P)),$$

*and  $\log_m$  denotes the matrix logarithm.*

## 2.3 The set of flags of fixed type

Let  $\mathbf{I} = (q_i)_{1 \leq i \leq r}$  be a sequence of positive integers such that  $\sum_{i=1}^r q_i = d$ .

### 2.3.1 Notations

The sequence  $I$  induces a partition of the indices  $\{1, \dots, d\}$  into  $r$  parts  $(B_i)_{1 \leq i \leq r}$ , i.e.  $B_1 = (1, \dots, q_1)$ ,  $B_r = (q_{r-1} + 1, \dots, d)$  and for  $1 < i < r$ ,  $B_i = (\bar{q}_{i-1} + 1, \dots, \bar{q}_i)$ , where  $\bar{q}_i := \sum_{j=1}^i q_j$ . Let  $A$  be a  $d \times d$  matrix. If  $A$  is defined by its columns, i.e.  $A = [A_1, \dots, A_d]$ , then, for all  $1 \leq i \leq r$ , set

$$A^{(i)} := [A_k : k \in B_i].$$

If  $A$  is defined by its entries, i.e.  $A = (a_{k\ell})_{1 \leq k \leq \ell \leq d}$ , then, for all  $1 \leq i \leq j \leq r$ , set

$$A^{(ij)} := (a_{k\ell})_{\substack{k \in B_i \\ \ell \in B_j}}.$$

Then, setting  $I_d^{(i)} := (I_d)^{(i)}$ , we have that  $A^{(i)} = AI_d^{(i)}$  and  $A^{(ij)} = (I_d^{(i)})' AI_d^{(j)}$ .

### 2.3.2 Definitions

We recall that a flag of type  $I$  is a collection  $\mathcal{V} = (V_i)_{1 \leq i \leq r}$  of  $r$  mutually orthogonal linear subspaces of  $\mathbb{R}^d$  such that for all  $1 \leq i \leq r$ ,  $\dim(V_i) = q_i$ . Thus, the set  $\mathcal{F}^I$  of flags of type  $I$  is a subset of

$$G^I := \prod_{i=1}^r G^i \quad \text{where} \quad G^i := G(q_i, d).$$

Identifying Grassmannians with projectors as in (2.1),

$$\mathcal{F}^I \simeq \{ \mathcal{P} = (P_i)_{1 \leq i \leq r} \in G^I : P_i P_j = 0, i \neq j \}.$$

Introduce the standard flag  $\mathcal{P}_0^I = (P_0^i)_{1 \leq i \leq r}$ , where  $P_0^i \in G^i$  is the block-diagonal matrix defined by

$$P_0^i := \text{Diag}[0_{q_1}, \dots, I_{q_i}, \dots, 0_{q_r}],$$

and  $0_{q_i}$  is the null matrix of order  $q_i$ . Then,

$$P_0^i = I_d^{(i)} (I_d^{(i)})'. \quad (2.7)$$

### 2.3.3 Action of the orthogonal group

The action of  $O(d)$  on Grassmannians induces an action of  $O(d)$  on  $\mathcal{F}^I$ , defined by

$$(Q, \mathcal{P}) \mapsto (QP_i Q')_{1 \leq i \leq r} \in \mathcal{F}^I,$$

for  $Q \in O(d)$  and  $\mathcal{P} = (P_i)_{1 \leq i \leq r} \in \mathcal{F}^I$ . Indeed,  $O(d)$  acts naturally on  $G^I$  and this action preserves the mutual orthogonality of linear subspaces. Then, define the map  $\pi^I : O(d) \rightarrow \mathcal{F}^I$  by

$$\pi^I(Q) = (QP_0^i Q')_{1 \leq i \leq r}.$$

**Proposition 2.** *The orbit of  $\mathcal{P}_0^I$  under this action is the whole  $\mathcal{F}^I$  and the stabilizer of  $\mathcal{P}_0^I$  is the group  $O(I)$ , where*

$$O(I) := \{ \text{Diag}[H_1, \dots, H_r] : H_i \in O(q_i), 1 \leq i \leq r \} = \prod_{i=1}^r O(q_i).$$

*In other words, the map  $\pi^I$  is surjective and for all  $Q \in O(d)$  and  $H \in O(I)$ ,  $\pi^I(QH) = \pi^I(Q)$ . Therefore,  $\pi^I$  induces a bijection*

$$\tilde{\pi}^I : O(d)/O(I) \simeq \mathcal{F}^I.$$

*such that  $\pi^I = \tilde{\pi}^I \circ \bar{p}^I$ , where  $\bar{p}^I$  is the canonical quotient map from  $O(d)$  to  $O(d)/O(I)$ .*

Now, let  $\pi_I^{SG}$  be the map from  $\text{St}^I := \prod_{i=1}^r \text{St}(q_i, d)$  to  $G^I$  defined by

$$\mathcal{Y} = (Y_i)_{1 \leq i \leq r} \mapsto (\pi_{q_i}^{SG}(Y_i))_{1 \leq i \leq r}.$$

Set  $\text{St}_\perp^I = \{\mathcal{Y} = (Y_i)_{1 \leq i \leq r} \in \text{St}^I : (Y_i)'Y_j = 0, i \neq j\}$ . Then, the restriction of  $\pi_I^{SG}$  to  $\text{St}_\perp^I$  is valued in  $\mathcal{F}^I$ .

**Lemma 4.** *Let  $\pi_I^{OS}$  be the bijection between  $O(d)$  and  $\text{St}_\perp^I$  defined by  $\pi_I^{OS}(Q) = (Q^{(i)})_{1 \leq i \leq r}$ . Then, for all  $Q \in O(d)$ ,*

$$\pi_I^{SG}(\pi_I^{OS}(Q)) = \pi^I(Q).$$

*Proof.* For  $Q \in O(d)$ ,  $\pi_I^{SG}(\pi_I^{OS}(Q)) = (\pi_{q_i}^{SG}(Q^{(i)}))_{1 \leq i \leq r}$  and for all  $1 \leq i \leq r$ , by (2.7),

$$\pi_{q_i}^{SG}(Q^{(i)}) = Q^{(i)}(Q^{(i)})' = (QI_d^{(i)})(QI_d^{(i)})' = QP_0^i Q'.$$

□

**Corollary 1.** *The triangles of the following diagram commute.*

$$\begin{array}{ccc} O(d) & \xrightarrow{\pi_I^{OS}} & \text{St}_\perp^I \\ \bar{P}^I \downarrow & \searrow \pi^I & \downarrow \pi_I^{SG} \\ O(d)/O(I) & \xrightarrow{\bar{\pi}^I} & \mathcal{F}^I \end{array}$$

## 2.4 Flag of eigenspaces

For  $S \in \text{Sym}_d$ , its spectrum is denoted by

$$\text{Spec}(S) = \{\mu_1(S) \geq \dots \geq \mu_d(S)\}.$$

We denote by  $\text{Sym}_d^\neq$  the set of  $S \in \text{Sym}_d$  such that  $\mu_1(S) > \dots > \mu_d(S)$ . For all  $1 \leq i \leq r$ , set  $\bar{q}_i := \sum_{j=1}^i q_j$ . Let

$\overline{\text{Sym}}_d^{(I)}$  be the set of  $S \in \text{Sym}_d$  such that for all  $1 \leq i \leq r-1$ ,  $\mu_{\bar{q}_i}(S) > \mu_{\bar{q}_i+1}(S)$ . Denote by  $\text{Sym}_d^I$  the set of  $S \in \overline{\text{Sym}}_d^{(I)}$  such that for all  $1 \leq i \leq r$  and  $k, \ell$  in  $B^i$ ,  $\mu_k(S) = \mu_\ell(S)$ . Then,

$$\text{Sym}_d^\neq \cup \text{Sym}_d^I \subset \overline{\text{Sym}}_d^{(I)}.$$

Let  $S \in \overline{\text{Sym}}_d^{(I)}$ . For  $1 \leq i \leq r$ , let  $\{u_k : 1 \leq k \leq d\}$  be eigenvectors respectively associated to the eigenvalues  $\{\mu_k(S) : k \in B_i\}$ . Then, the linear subspace  $\mathcal{V}_i$  spanned by  $\{u_k : 1 \leq k \leq d\}$  is independent of the choice of these eigenvectors. We define the *eigenprojection* associated to  $\{\mu_k(S) : k \in B_i\}$  as the map  $P_i : \overline{\text{Sym}}_d^{(I)} \rightarrow G^i$  such that  $P_i(S)$  is the projector onto  $\mathcal{V}_i$ . Then, define the map  $F^I : \overline{\text{Sym}}_d^{(I)} \rightarrow \mathcal{F}^I$  by

$$F^I(S) = (P_i(S))_{1 \leq i \leq r}.$$

$F^I(S)$  is called the *flag of eigenspaces* of  $S$ .

**Lemma 5.** *Let  $S \in \overline{\text{Sym}}_d^{(I)}$  and  $Q \in O(d)$ . Then,  $\pi^I(Q) = F^I(S)$  if and only if  $Q$  is a matrix of eigenvectors of  $S$  such that for all  $1 \leq i \leq r$ , the columns of  $Q^{(i)}$  are eigenvectors associated to  $\{\mu_k(S) : k \in B_i\}$  if and only if*

$$QP_0^i Q' = P_i(S), \quad 1 \leq i \leq r. \quad (2.8)$$

### 3 Estimation of $F^I(\Sigma)$

Recall that  $X$  is a random vector valued in  $\mathbb{R}^d$ ,  $d \geq 1$ , with covariance matrix  $\Sigma$  whose different eigenvalues  $\lambda_1 > \dots > \lambda_r$  have respective multiplicities  $\mathbf{I} = (q_i)_{1 \leq i \leq r}$ . We aim at estimating  $F^I(\Sigma)$ , through  $F^I(\hat{\Sigma}_n)$ , where  $\hat{\Sigma}_n$  is a sample covariance matrix from an iid sample of size  $n$  of  $X$ . In the preceding section, we have not described the Riemannian geometry of  $\mathcal{F}^I$ , since the Riemannian logarithm is not available in closed form. Instead, when  $X$  is normally distributed in  $\mathbb{R}^d$ , we derive the estimation of  $F^I(\Sigma)$  from the collection of CLT's for all  $P_i(\hat{\Sigma}_n)$ ,  $1 \leq i \leq r$ , established below. Throughout the sequel,  $\Gamma$  is a matrix of eigenvectors of  $\Sigma$  such that  $\pi^I(\Gamma) = F^I(S)$ . Such a  $\Gamma$  is described by Lemma 5. So, for all  $1 \leq i \leq r$ ,  $P_i(\Sigma) = \Gamma P_0^i \Gamma'$ .

#### 3.1 Review of Anderson's results

The following CLT provides the uncertainty of the estimation of  $\hat{\Sigma}_n$  by  $\Sigma$ , which one wishes to propagate, in order to derive that of flags of eigenspaces.

**Theorem 2.** *Let  $\Delta$  be the diagonal matrix of eigenvalues of  $\Sigma$  defined by  $\Delta = \Gamma' \Sigma \Gamma$ . Then,*

$$U_n := \sqrt{n} \left( \Gamma' \hat{\Sigma}_n \Gamma - \Delta \right) \xrightarrow[n \rightarrow \infty]{d} U,$$

where  $U$  is a random matrix whose distribution is characterized as follows:  $U \in \text{Sym}_d$ , the blocks  $\{U^{(i,j)} : 1 \leq i < j \leq r\}$  are mutually independent and for  $i \neq j$ , the entries of  $U^{(i,j)}$  are iid rv's  $\mathcal{N}(0, s_{i,j}^2)$ , of standard deviation  $s_{i,j} := \sqrt{\lambda_i \lambda_j}$ .

Let  $\psi$  be the map from  $\text{Sym}_d^\neq$  to  $O(d)$  such that for  $S \in \text{Sym}_d^\neq$  and  $1 \leq k \leq d$ , the  $k$ -th column of  $\psi(S)$  is the eigenvector associated to  $\mu_k(S)$  whose  $k$ -th entry is non-negative. Then, for all  $n \geq 1$ , set

$$C_n := \psi(\hat{\Sigma}_n) \quad \text{and} \quad E_n := \Gamma' C_n.$$

Thus,  $C_n$  is a matrix of eigenvectors of  $\hat{\Sigma}_n$  such that  $\pi^I(C_n) = F^I(\hat{\Sigma}_n)$ . The following Theorem is the main result of (Anderson, 1963).

**Theorem 3.** *For  $1 \leq i \leq r$ ,*

$$E_n^{(i,i)} \xrightarrow[n \rightarrow \infty]{d} E^{(i,i)},$$

where  $E^{(i,i)}$  is uniformly distributed over the set of orthogonal matrices with non-negative diagonal entries. For all  $i \neq j$  and  $n \geq 1$ , set  $F_n^{(i,j)} := \sqrt{n} E_n^{(i,j)}$ . Then,

$$F_n^{(i,j)} \xrightarrow[n \rightarrow \infty]{d} F^{(i,j)},$$

where  $F^{(i,j)}$  is a random matrix whose entries are iid rv's  $\mathcal{N}(0, \sigma_{i,j}^2)$ , of standard deviation

$$\sigma_{i,j} := \frac{\sqrt{\lambda_i \lambda_j}}{|\lambda_i - \lambda_j|}.$$

Furthermore, the blocks  $\{E^{(i,i)}, F^{(i,j)} : 1 \leq i \leq r, 1 \leq j \leq r, i \neq j\}$  are mutually independent.

The strategy for the proof of Theorem 3 is to express  $E_n$  in function of  $U_n$  and then to apply Theorem 4 below. Recall that  $U_n = \sqrt{n}(\Gamma' \hat{\Sigma}_n \Gamma - \Delta)$ , so that  $\hat{\Sigma}_n = \phi_n(U_n)$  where  $\phi_n$  is the map from  $\text{Sym}_d$  to  $\text{Sym}_d$  defined by

$$\phi_n(u) = \Gamma \left( \Delta + \frac{u}{\sqrt{n}} \right) \Gamma'.$$

Then,  $E_n = f_n(U_n)$ , where  $f_n$  is the map from  $(\phi_n)^{-1}(\text{Sym}_d^\neq)$  to  $O(d)$  defined by

$$f_n : u \mapsto \Gamma' \psi(\phi_n(u)) \quad \text{that is,} \quad f_n(u) = \psi \left( \Delta + \frac{u}{\sqrt{n}} \right).$$



**Theorem 4.** Let  $\mathcal{F}, \mathcal{T}$  be metric spaces and  $(\mathcal{F}_n)_{n \geq 1}$  a sequence of subsets of  $\mathcal{F}$ . Let  $(V_n)_{n \geq 1}$  be a sequence of rv's valued in  $(\mathcal{F}_n)_{n \geq 1}$ . Assume that

$$V_n \xrightarrow[n \rightarrow \infty]{d} V,$$

where  $V$  is a rv valued in  $\mathcal{F}$ . For all  $n \geq 1$ , let  $g_n$  be a map from  $\mathcal{F}_n$  to  $\mathcal{T}$  and  $g$  a map from  $\mathcal{F}$  to  $\mathcal{T}$ . Let  $\mathcal{F}_0$  be a subset of  $\mathcal{F}$  such that  $g$  is continuous on  $\mathcal{F}_0$  and

$$\Pr(V \in \mathcal{F}_0) = 1.$$

Assume that for all  $u \in \mathcal{F}_0$ , and all sequence  $(u_n)_{n \geq 1}$  valued in  $(\mathcal{F}_n)_{n \geq 1}$ ,

$$u_n \xrightarrow[n \rightarrow \infty]{} u \implies g_n(u_n) \xrightarrow[n \rightarrow \infty]{} g(u).$$

Then,

$$g_n(V_n) \xrightarrow[n \rightarrow \infty]{d} g(V).$$

*Proof.* See APPENDIX D in (Anderson, 1963) or section 18.11 in (Van der Vaart, 2000).  $\square$

### 3.2 CLT for $P_i(\hat{\Sigma}_n)$

For any  $u \in \text{Sym}_d$ , set  $M_{<}^i(u) = \left[ \frac{1}{\lambda_i - \lambda_j} u^{(i,j)} : i < j \right]$  and  $M_{>}^i(u) = \left[ \frac{1}{\lambda_i - \lambda_j} u^{(i,j)} : i > j \right]$ .

For all  $1 \leq i \leq r$ , define the map  $g^i : \text{Sym}_d \longrightarrow T_{P_0^i} G^i$  by

$$g^1(u) = \begin{pmatrix} 0_{q_1} & M_{>}^1(u) \\ (M_{>}^1(u))' & 0 \end{pmatrix}, \quad g^r(u) = \begin{pmatrix} 0 & (M_{<}^r(u))' \\ M_{<}^r(u) & 0_{q_r} \end{pmatrix}$$

$$g^i(u) = \begin{pmatrix} 0 & (M_{<}^i(u))' & 0 \\ M_{<}^i(u) & 0_{q_i} & M_{>}^i(u) \\ 0 & (M_{>}^i(u))' & 0 \end{pmatrix}, \quad i \neq 1, r.$$

Hereabove, when we write 0 as a block matrix, it is implicit that its size is the suitable one. In the sequel, given rv's  $X : \Omega \rightarrow V$  and  $Y : \Omega \rightarrow (E, \mathcal{B})$ , where  $V$  is a vector space and  $(E, \mathcal{B})$  a measurable space, for  $B \in \mathcal{B}$ , we denote by  $X \mathbf{1}_{\{Y \in B\}}$  the rv valued in  $V$  equals to  $X$  if  $Y \in B$  and to 0 else.

**Theorem 5.** For all  $1 \leq i \leq r$ ,

$$\sqrt{n} \text{Log}_{P_0^i}^{G^i} \left( \Gamma' P_i(\hat{\Sigma}_n) \Gamma \right) \mathbf{1}_{\mathcal{T}_n^i} = g^i(U) + O_{\Pr}(1/\sqrt{n}), \quad (3.1)$$

where  $\mathcal{T}_n^i := \left\{ \Gamma' P_i(\hat{\Sigma}_n) \Gamma \notin \text{Cut}(P_0^i) \right\}$ .

The proof of Theorem 5 is based on Theorem 4. We introduce below some sequences of different natures satisfying the assumptions of Theorem 4. For all  $1 \leq i \leq r$  and  $n \geq 1$ , set

$$\mathcal{F}_n^i := \left\{ u \in \phi_n^{-1}(\text{Sym}_d^{\neq}) : \Gamma' P_i(\phi_n(u)) \Gamma \notin \text{Cut}(P_0^i) \right\}.$$

Then, consider the rv  $V_n^i$  valued in  $\mathcal{F}_n^i$ , defined by

$$V_n^i := U_n \mathbf{1}_{\{U_n \in \mathcal{F}_n^i\}}$$

Define the map  $g_n^i : \mathcal{F}_n^i \longrightarrow T_{P_0^i} G^i$  by

$$g_n^i(u) = \sqrt{n} \log_{P_0^i}^{G^i} \left( \Gamma' P_i(\hat{\Sigma}_n) \Gamma \right).$$

Since  $\phi_n(0) = \Gamma \Delta \Gamma' = \Sigma$  and  $\Gamma^t P_i(\Sigma) \Gamma = P_0^i$ , we have that  $0 \in \mathcal{F}_n^i$  and  $g_n^i(0) = 0$ . Therefore,

$$g_n^i(V_n^i) = \sqrt{n} \log_{P_0^i}^{G^i} \left( \Gamma' P_i(\hat{\Sigma}_n) \Gamma \right) \mathbf{1}_{\mathcal{T}_n}.$$

The following series of results imply that the assumptions of Theorem 4 are fulfilled.

**Proposition 3.** For all  $1 \leq i \leq r$ ,

$$V_n^i \xrightarrow[n \rightarrow \infty]{d} U.$$

*Proof.* By the CLT for  $\hat{\Sigma}_n$ ,  $U_n \xrightarrow[n \rightarrow \infty]{d} U$ , which combined to Lemma 6 below, concludes the proof.  $\square$

**Lemma 6.** For all  $1 \leq i \leq r$ ,

$$\Pr(U_n \in \mathcal{F}_n^i) \xrightarrow[n \rightarrow \infty]{} 0.$$

*Proof.* For all  $1 \leq i \leq r$  and  $n \geq 1$ ,

$$U_n \in \mathcal{F}_n^i \iff E_n^{(i)} (E_n^{(i)})' \notin \text{Cut}(P_0^i) \iff \text{rk}(E_n^{(ii)}) = q_i.$$

By Theorem 3,  $E_n^{ii} \xrightarrow[n \rightarrow \infty]{d} E^{ii}$  and  $E^{ii}$  is uniformly distributed on the orthogonal group  $O(q_i)$ , so that, a.s.,  $\text{rk}(E^{ii}) = q_i$ . Now, for all  $n \geq 1$ ,  $\text{rk}(E_n^{ii}) \leq q_i$ . We conclude by Lemma 7 below.  $\square$

**Lemma 7.** Let  $B_n$  and  $B$  be random matrices of same size. Assume that  $B_n \xrightarrow[n \rightarrow \infty]{d} B$ , that, a.s.,  $\text{rk}(B) = b$  and that  $\Pr(\text{rk}(B_n) \leq b) \xrightarrow[n \rightarrow \infty]{} 1$ . Then,  $\Pr(\text{rk}(B_n) = b) \xrightarrow[n \rightarrow \infty]{} 1$ .

*Proof.* See Lemma 2.6. in (Tyler, 1981)  $\square$

**Remark 1.** For all  $1 \leq i \leq r$ ,  $g^i$  is continuous on  $\text{Sym}_d^\neq$  and

$$\Pr(U \in \text{Sym}_d^\neq) = 1$$

**Proposition 4.** Fix  $1 \leq i \leq r$ . Let  $v \in \text{Sym}_d^\neq$ . Let  $(v_n)_{n \geq 1}$  be a sequence in  $(\mathcal{F}_n^i)_{n \geq 1}$  such that  $v_n \xrightarrow[n \rightarrow \infty]{} v$ . Then,  $g_n^i(v_n)$  is of the form

$$g_n^i(v_n) = [\Lambda_n^i(v_n), P_0^i] + (\sqrt{n})^{-1} [\Phi_n^i(v_n), P_0^i] + o((\sqrt{n})^{-1}),$$

where for all  $n \geq 1$ ,  $\Delta_n^i(v_n)$  is a matrix such that

$$[\Lambda_n^i(v_n), P_0^i] \in T_{P_0^i} G^i \quad \text{with} \quad [\Lambda_n^i(v_n), P_0^i] \xrightarrow[n \rightarrow \infty]{} g^i(v),$$

and the sequence  $(\Phi_n^i(v_n))_{n \geq 1}$  converges to a matrix  $\Phi^i(v)$ , as  $n \rightarrow \infty$ .

*Proof.* We use the proof of Theorem 3 and expand the matrix logarithm involved in the closed form of the Riemannian Logarithm for Grassmannians.  $\square$

### 3.3 A pivotal statistic

We aim at deriving a *pivotal statistic* of  $F^I(\Sigma)$ , i.e. which depends only on  $F^I(\Sigma)$  and the sample, and whose asymptotic distribution does not depend on any unknown parameter.

In the sequel, we denote by  $\mathcal{D}(I)$  the set of  $d \times d$  diagonal matrices  $D$  of the form

$$D = \text{diag}[\beta_i I_{q_i} : 1 \leq i \leq r].$$

**Remark 2.** For all  $H \in O(I)$  and  $K \in \mathcal{D}(I)$ ,  $[H, K] = 0$ .

For any  $1 \leq i \leq r$ , we aim at normalizing the entries of  $g^i(U)$ . By Theorem 2, for  $i \neq j$ , the entries of  $\frac{1}{\lambda_i - \lambda_j} U^{(i,j)}$  are real i.i.d rv's  $\mathcal{N}(0, \sigma_{i,j}^2)$ , of standard deviation  $\sigma_{i,j} = \frac{\sqrt{\lambda_i \lambda_j}}{|\lambda_i - \lambda_j|}$ . For any  $1 \leq i \leq r$ , define  $K^i \in \mathcal{D}(\mathbf{I})$  by

$$K^i := \text{diag} \left[ \frac{1}{\sigma_{i,1}} I_{q_1}, \dots, I_{q_i}, \dots, \frac{1}{\sigma_{i,r}} I_{q_r} \right]. \quad (3.2)$$

Now, consider the rv  $\bar{g}^i(U)$  defined by

$$\bar{g}^i(U) := K^i g^i(U) K^i.$$

Then,  $\bar{g}^i(U)$  is a rv valued in  $T_{P_0^i} G^i$ , with the same stochastic dependence between and within blocks as  $g^i(U)$ . Now, left and right multiplications of  $g^i(U)$  by  $K^i$  imply that  $\frac{1}{\lambda_i - \lambda_j} U^{(i,j)}$  is replaced by

$$W^{(i,j)} := \frac{1}{\sigma_{i,j}} \left( \frac{1}{\lambda_i - \lambda_j} U^{(i,j)} \right),$$

whose entries are real **iid** rv's  $\mathcal{N}(0, 1)$ . Since  $U$  is a symmetric random matrix,

$$\sum_{i=1}^r \|\bar{g}^i(U)\|_F^2 = 4 \sum_{1 \leq i < j \leq r} \|W^{(i,j)}\|_F^2.$$

Now, the blocks  $\{W^{(i,j)} : 1 \leq i < j \leq r\}$  are mutually independent. So,

$$\frac{1}{4} \sum_{i=1}^r \|\bar{g}^i(U)\|_F^2 = \sum_{1 \leq i < j \leq r} \|W^{(i,j)}\|_F^2 \sim \chi_{D^{\mathbf{I}}}^2,$$

where

$$D^{\mathbf{I}} := \frac{1}{2} \left( d^2 - \sum_{i=1}^r q_i^2 \right). \quad (3.3)$$

By Theorem 5, for all  $1 \leq i \leq r$ ,

$$\sqrt{n} \left[ K^i \left( \text{Log}_{\Pi_0^i}^{G^i} \left( \Gamma' P^j \left( \hat{\Sigma}_n \right) \Gamma \right) K^i \right) \mathbf{1}_{\mathcal{T}_n^i} \xrightarrow[n \rightarrow \infty]{d} \bar{g}^i(U) \right].$$

Therefore, we have proved the following result.

**Proposition 5.** For  $(K^i)_{1 \leq i \leq r}$  defined in (3.2), and  $D^{\mathbf{I}}$  defined in (3.3),

$$\frac{n}{4} \sum_{i=1}^r \left\| K^i \text{Log}_{P_0^i}^{G^i} \left( \Gamma' P^i \left( \hat{\Sigma}_n \right) \Gamma \right) K^i \right\|_F^2 \mathbf{1}_{\mathcal{T}_n^i} \xrightarrow[n \rightarrow \infty]{d} \chi_{D^{\mathbf{I}}}^2.$$

**Corollary 2.** The statistic  $\hat{T}_n$  defined below is a a pivotal statistic of  $F^{\mathbf{I}}(\Sigma)$ :

$$\hat{T}_n := \frac{n}{4} \sum_{i=1}^r \left\| \hat{K}_n^i \text{Log}_{P_0^i}^{G^i} \left( \Gamma' P^i \left( \hat{\Sigma}_n \right) \Gamma \right) \hat{K}_n^i \right\|_F^2 \mathbf{1}_{\mathcal{T}_n^i} \xrightarrow[n \rightarrow \infty]{d} \chi_{D^{\mathbf{I}}}^2, \quad (3.4)$$

where, for  $1 \leq i \leq r$ ,  $\hat{K}_n^i$  is derived from  $K^i$  by replacing  $\lambda_i$  by any consistent estimator  $\hat{\lambda}_n^i$ , for example

$$\hat{\lambda}_n^i = \frac{1}{q_i} \sum_{k \in B_i} \mu_k \left( \hat{\Sigma}_n \right).$$

### 3.4 Simulation

Figure 1 illustrates the convergence in distribution of  $\hat{T}_n$  to a  $\chi^2$  distribution. The parameters for this simulation are  $d = 4$  and  $\mathbf{I} = (1, 1, 1, 1)$ , so that  $\Sigma \in \text{Sym}_d^{\neq}$ . Then  $D^{\mathbf{I}} = d(d-1)/2 = 6$ . For the sample size, we take  $n = 10000$ . The histogram in blue represents the probability distribution of  $\hat{T}_n$  and the curve in red is that of the probability distribution function of the  $\chi_6^2$  distribution. We see that the empirical distribution, i.e. that of  $\hat{T}_n$  is indeed very close to that of the  $\chi_6^2$  one.

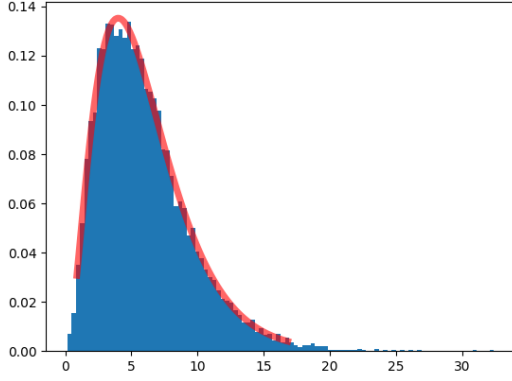


Figure 1: Illustration of the convergence of  $\hat{T}_n$  to the  $\chi_6^2$  distribution

### 3.5 Confidence regions and Tests

Let  $\mathbb{K} = (K^i)_{1 \leq i \leq r} \in (\mathcal{D}(\mathbb{I}))^r$ . For  $Q \in O(d)$  and  $\mathcal{R} = (R_i)_{1 \leq i \leq r} \in \mathcal{F}^I$ , set

$$\Delta_{\mathbb{K}}(Q, \mathcal{R}) = \frac{1}{4} \sum_{i=1}^r \left\| K^i \text{Log}_{P_0^i}^{G^i}(Q' R_i Q) K^i \right\|_F^2 \mathbf{1}_{\{Q' R_i Q \notin \text{Cut}(P_0^i)\}}.$$

**Proposition 6.** For all  $H \in O(\mathbb{I})$ ,

$$\Delta_{\mathbb{K}}(QH, \mathcal{R}) = \Delta_{\mathbb{K}}(Q, \mathcal{R}). \quad (3.5)$$

*Proof.* First, we prove that for all  $1 \leq i \leq r$  and  $H \in O(\mathbb{I})$ ,

$$(QH)' R_i (QH) \notin \text{Cut}(P_0^i) \iff Q' R_i Q \notin \text{Cut}(P_0^i) \quad (3.6)$$

Indeed, writing  $R_i = C_i(C_i)'$  with  $C_i \in \text{St}(q_i, d)$ ,

$$\begin{aligned} (QH)' R_i (QH) \notin \text{Cut}(P_0^i) &\iff (H' Q' C_i)(H' Q' C_i)' \notin \text{Cut}(P_0^i) \\ &\iff \text{rank} \left( (I_d^{(i)})' (H' Q' C_i) \right) = q_i \\ &\iff \text{rank} \left( (H I_d^{(i)})' (Q' C_i) \right) = q_i \end{aligned}$$

Now, for all  $1 \leq i \leq r$ ,  $(H I_d^{(i)}) (H I_d^{(i)})' = H P_0^i H' = P_0^i$ . Therefore,

$$\text{rank} \left( (H I_d^{(i)})' (Q' C_i) \right) = q_i \iff \text{rank} \left( (I_d^{(i)})' (Q' C_i) \right) = q_i \iff Q' R_i Q \notin \text{Cut}(P_0^i).$$

This proves (3.6). Now, when  $(QH)' R_i (QH) \notin \text{Cut}(P_0^i)$ , the closed formula for  $\text{Log}_{P_0^i}^{G^i}(\cdot)$  yields that

$$\text{Log}_{P_0^i}^{G^i}((QH)' R_i (QH)) = \text{Log}_{P_0^i}^{G^i}(H' (Q' R_i Q) H) = H' \text{Log}_{P_0^i}^{G^i}(Q' R_i Q) H.$$

Now, Remark 2 combined to the properties of invariance of the Frobenius norm imply that for all  $1 \leq i \leq r$ ,

$$\left\| K^i \left( H' \text{Log}_{P_0^i}^{G^i}(Q' R_i Q) H \right) K^i \right\|_F = \left\| H' \left( K^i \text{Log}_{P_0^i}^{G^i}(Q' R_i Q) K^i \right) H \right\|_F = \left\| K^i \text{Log}_{P_0^i}^{G^i}(Q' R_i Q) K^i \right\|_F$$

□

For fixed  $n \geq 1$ , set  $\hat{\mathbb{K}}_n := (\hat{K}_n^i)_{1 \leq i \leq r} \in (\mathcal{D}(\mathbb{I}))^r$  and define the map  $\hat{T}_n : O(d) \rightarrow [0, \infty)$  by

$$\hat{T}_n(Q) := n \Delta_{\hat{\mathbb{K}}_n} \left( Q, F^I(\hat{\Sigma}_n) \right).$$

Then, for any matrix  $\Gamma$  of eigenvectors of  $\Sigma$ ,  $\hat{T}_n(\Gamma)$  is the pivotal statistic  $\hat{T}_n$  defined in (3.4). Furthermore, Proposition 6 implies that the value of  $\hat{T}_n(Q)$  is independent of the class of  $Q$  modulo  $O(\mathbb{I})$ .

**Corollary 3.** For any  $\alpha \in (0, 1)$ , a confidence region for  $F^I(\Sigma)$  of asymptotic level  $(1 - \alpha)$  is given by

$$R_{n,\alpha} := \left\{ \pi^I(Q) \in \mathcal{F}^I : \widehat{T}_n(Q) \leq \chi_{DI}^2(1 - \alpha) \right\},$$

where  $\chi_{DI}^2(1 - \alpha)$  is the quantile of order  $(1 - \alpha)$  of the  $\chi_{DI}^2$  distribution.

**Corollary 4.** Consider the following null hypothesis assumption.

$$H_0 : \pi^I(Q_0) = F^I(\Sigma), \text{ for } Q_0 \in O(d).$$

For any  $\alpha \in (0, 1)$ , consider the test which accepts  $H_0$  when  $\widehat{T}_n(Q_0) \leq \chi_{DI}^2(1 - \alpha)$  and rejects  $H_0$  else. Then, this test is of asymptotic level  $\alpha$ .

### 3.6 Conclusion

Given a normally distributed random vector  $X$  valued in  $\mathbb{R}^d$ ,  $d \geq 1$ , a geometric framework allows us to develop an asymptotic procedure to infer the set of all its PS's. In addition, we provide easily implementable tests concerning the complete collection of principal subspaces and confidence regions for them. These results opens many questions which could lead to useful extensions. Among them, two of the most striking questions are:

- How to estimate or learn the type I?
- Is it possible to relax the Gaussian assumption?

### Acknowledgements

The authors have received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement G-Statistics No 786854). It was also supported by the French government through the 3IA Côte d'Azur Investments ANR-19-P3IA-0002 managed by the National Research Agency.

### References

- Anderson, T.W. (1963). *Asymptotic theory for principal components analysis*. Ann. Math. Stat. (1963), Vol. 34, 122-148.
- Batzies, E., Huper, K., Machado, L. and Silva Leite, F. (2015). *Geometric mean and geodesic regression on Grassmannians*. Linea Algebra Appl., 466:83-101, 2015.
- Bhattacharya, R. and Patrangenaru, V. (2005). *Large sample theory of intrinsic and extrinsic sample means on manifolds-II*. The Annals of Statistics (2005), Vol. 33, No. 3, 1225-1259.
- Bendokat, T., Zimmermann, R., Absil, P.A. (2020). *A Grassmann Manifold Handbook: Basic Geometry and Computational Aspects*. arXiv preprint arXiv:2011.13699, 2020.
- Lee, J.M. (2018). *Introduction to Riemannian manifolds*. volume 176 of Graduate Texts in Mathematics. Springer, Cham, 2018.
- Tyler, D.E. (1981). *Asymptotic inference for eigenvectors*. The Annals of Statistics 9 (4), 725-736, 1981.
- Van der Vaart, A.W. (2000). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, Volume 3.
- Ye, K., Wong, K.S.W., Lim, L.H. (2022). *Optimization on flag manifolds*. Mathematical Programming (Series A), 2022, 194(1-2), pp. 621-660.