Published in Journal of Multivariate Analysis 93 (2005) 180–197

Angular Gaussian and Cauchy estimation

Claude Auderset,[a] Christian Mazza,[b,*] and Ernst A. Ruh[a,1]

[a] *Department of Mathematics, University of Fribourg, CH-1700 Fribourg, Switzerland*
[b] *Section de Mathématiques, 2-4 Rue du Lièvre, CP 240, CH-1211 Genève 24, Switzerland*

Abstract

This paper proposes a unified treatment of maximum likelihood estimates of angular Gaussian and multivariate Cauchy distributions in both the real and the complex case. The complex case is relevant in shape analysis. We describe in full generality the set of maxima of the corresponding log-likelihood functions with respect to an arbitrary probability measure. Our tools are the convexity of log-likelihood functions and their behaviour at infinity.

AMS 2000 subject classifications: 62H11; 62H12; 62F10

Keywords: Multivariate Cauchy; Angular Gaussian; Directional and shape analysis; Maximum likelihood; Differential geometry; Equivariance; Geodesics; Symmetric spaces

1. Introduction

Real angular Gaussian distributions are studied in directional analysis (cf. [11,16,15, section 3.4.7, section 3.6]), while real multivariate Cauchy distributions can be viewed as *t*-distributions with one degree of freedom (see e.g. [8]). In fact, as observed by Knight and Meyer [10], these two apparently unrelated statistical models are essentially identical. On the other hand, complex angular Gaussian distributions are used in shape analysis (see e.g. [7] or [13]); they provide an interesting alternative to the Bingham distribution since their densities do not contain involved parameter-dependent normalization (see [6] or [2]). They relate, in a

*Corresponding author.
E-mail addresses: claude.auderset@unifr.ch (C. Auderset), christian.mazza@math.unige.ch (C. Mazza), ernst.ruh@unifr.ch (E.A. Ruh).
[1] Supported by Swiss National Science Foundation Grant 20-05811.99 and 20-67619.02

0047-259X
doi:10.1016/j.jmva.2004.01.007

similar way as in the real case, to complex multivariate Cauchy distributions, which can be viewed as t-distributions with two degrees of freedom.

Existence and uniquess of angular Gaussian and Cauchy maximum likelihood estimates (MLE's) have been intensively studied, at least in the real case. Tyler [15] has shown that the q-variate angular Gaussian MLE is almost surely well-defined for an i.i.d. random sample of size $n > q+1$. Kent and Tyler [8] and Kent et al. [9] study MLE's of the more general t-distributions. Arslan and Kent (1998) [1] show that the maxima of the q-variate Cauchy likelihood function of a sample of size $q+1$ in general position form a manifold of dimension q. Corresponding results for the complex case appear not to have been published yet, notwithstanding their importance in shape analysis (see e.g. [12,13]).

We give a necessary and sufficient condition for existence and uniquess of angular Gaussian and Cauchy MLE's in both the real and complex case. More precisely, our main result (Theorem 1, Section 3) describes in full generality the set of maxima of the corresponding log-likelihood function for an arbitrary probability distribution, in particular for the empirical distribution of a sample.

Before presenting it, we recall in Section 2 how these various models can be unified by reducing them to normal laws. Section 4 is on the convexity of log-likelihood functions and Section 6 on their behaviour at infinity. We present the angular Gaussian maximum likelihood equation in Section 5 and prove our main result in Sections 7 and 8. Appendix A is for readers interested in groups and differential geometry and Appendix B describes projective subspaces of plane shapes.

2. Reduction of angular Gaussian and Cauchy models to normal laws

Let $X = (X_1, \dots, X_{q+1})'$ be a random vector of central normal law $\mathcal{N}(0, \theta)$. In order to treat the real and the complex case in parallel, let us assume that $X \in \mathbb{F}^{q+1}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Up to a constant factor, the density of X is $\exp(-x^*\theta^{-1}x/2)$, where x^* denotes the adjoint of x, i.e. the transpose x' of x in the real case, and the conjugate of x' in the complex case. The covariance matrix θ of X is self-adjoint, i.e. symmetric when $\mathbb{F} = \mathbb{R}$, and Hermitian when $\mathbb{F} = \mathbb{C}$. The complex normal distribution $\mathcal{N}(0, \theta)$ in $\mathbb{C}^{q+1}$ can be viewed as the usual normal distribution $\mathcal{N}(0, \theta_{\mathbb{R}})$ in $\mathbb{R}^{2(q+1)}$ with the real covariance matrix $\theta_{\mathbb{R}} = \begin{pmatrix} \theta_1 & -\theta_2 \\ \theta_2 & \theta_1 \end{pmatrix}$ of order $2(q+1)$, where $\theta = \theta_1 + i\theta_2$ with real matrices θ_1 and θ_2 of order $q+1$.

The angular Gaussian model is obtained from the normal vector X by retaining only its *axis* (or unoriented direction) $[X] = \{\lambda X | \lambda \in \mathbb{F}\}$ and forgetting anything else. The law of $[X]$ is called the (*real or complex*) *q-variate angular Gaussian distribution of parameter* θ. We denote it by $\mathcal{G}_\theta$.

The sample space of the angular Gaussian model is the set of axes in $\mathbb{F}^{q+1}$. It is, in fact, the *projective space* $\mathbb{F}\mathrm{P}^q = \{[x] | x \in \mathbb{F}^{q+1}, x \neq 0\}$ of dimension q. A data point in $\mathbb{F}\mathrm{P}^q$ can also be viewed as a pair of opposite unit vectors $\pm x \in \mathbb{R}^{q+1}$ in the real case, and a one-dimensional family $e^{i\varphi}x \in \mathbb{C}^{q+1}$ $(\varphi \in \mathbb{R})$ of unit vectors in the complex case.

The complex case has gained much interest after an important discovery by Kendall [5]: the manifold Σ_2^{q+2} of similarity shapes of configurations of $q+2$ non-identical points in the plane can be identified with the complex projective space $\mathbb{C}\mathrm{P}^q$. For more information concerning shape analysis and the relevance of the complex angular Gaussian distribution, see e.g. [7,13].

As $[\lambda X]=[X]$ for $\lambda \in \mathbb{F}$, $\lambda \neq 0$, the covariance matrix θ of the central normal vector X is determined up to a positive constant only. We remove this indeterminacy by requiring that $\det \theta = 1$. So, we parametrize the angular Gaussian distributions $\mathscr{G}_\theta$ by the space $\Theta_q^{\mathbb{F}}$ of positive definite self-adjoint matrices of order $q+1$ and determinant 1. In the special case of the unit matrix $\theta = I \in \Theta_q^{\mathbb{F}}$, we get the *uniform distribution* $\mathscr{G}_I$ *on* $\mathbb{F}\mathrm{P}^q$.

A computation shows that the density of the unit vector $X/||X||$ with respect to the uniform probability distribution of the unit sphere in $\mathbb{F}^{q+1}$ is $(x^*\theta^{-1}x)^{-i_{\mathbb{F}}(q+1)}$ $(||x||=1)$, where $i_{\mathbb{R}}=\frac{1}{2}$ and $i_{\mathbb{C}}=1$. Thus, the density of the angular Gaussian distribution $\mathscr{G}_\theta$ with respect to the uniform distribution $\mathscr{G}_I$ of $\mathbb{F}\mathrm{P}^q$ is

$$f_\theta^{\mathscr{G}}([x]) = [(x^*x)/(x^*\theta^{-1}x)]^{i_{\mathbb{F}}(q+1)} \quad (\theta \in \Theta_q^{\mathbb{F}}, [x] \in \mathbb{F}\mathrm{P}^q), \tag{1}$$

since the uniform distribution $\mathscr{G}_I$ of $\mathbb{F}\mathrm{P}^q$ is the image measure of the uniform distribution on the unit sphere in $\mathbb{F}^{q+1}$ under the projection $x \mapsto [x]$ $(||x||=1)$ onto $\mathbb{F}\mathrm{P}^q$.

We arrive at the Cauchy model by observing that the axis $[X] \in \mathbb{F}\mathrm{P}^q$ of the $\mathscr{N}(0,\theta)$ distributed random vector $X=(X_1,\dots,X_{q+1})'$ is determined by the vector $Y=(X_1,\dots,X_q)'/X_{q+1} \in \mathbb{F}^q$, at least in the almost sure case where $X_{q+1} \neq 0$. A computation shows that the density of Y with respect to the Lebesgue measure of $\mathbb{F}^q$ is

$$g_{\mu,\Sigma}(y) = \frac{c_q^{\mathbb{F}}(\det \Sigma)^{-i_{\mathbb{F}}}}{[1+(y-\mu)^*\Sigma^{-1}(y-\mu)]^{i_{\mathbb{F}}(q+1)}} \quad (y \in \mathbb{F}^q), \tag{2}$$

where $c_q^{\mathbb{F}}$ is a constant and

$$\theta = (\det \Sigma)^{-1/(q+1)} \begin{pmatrix} \Sigma + \mu\mu^* & \mu \\ \mu^* & 1 \end{pmatrix} \in \Theta_q^{\mathbb{F}}. \tag{3}$$

Here, $\mu \in \mathbb{F}^q$ and Σ is a positive definite self-adjoint matrix of order q. The law of the random vector $Y \in \mathbb{F}^q$ is called the (*real or complex*) *q-variate Cauchy distribution* $\mathscr{C}(\mu,\Sigma)$ *of location-scatter parameters* μ, Σ.

Let us imbed, as usual, the affine space $\mathbb{F}^q$ into the projective space $\mathbb{F}\mathrm{P}^q$ by identifying the point $y=(y_1,\dots,y_q)' \in \mathbb{F}^q$ to the axis $[(y_1,\dots,y_q,1)'] \in \mathbb{F}\mathrm{P}^q$. Then, by definition, restricting the angular Gaussian distribution $\mathscr{G}_\theta$ to $\mathbb{F}^q$ yields the Cauchy distribution $\mathscr{C}(\mu,\Sigma)$, where the parameters θ, μ and Σ are related by Eq. (3). So, these two statistical models are essentially identical. However, the sample space of Cauchy distributions is slightly smaller than the sample space of angular Gaussian distributions since the affine space $\mathbb{F}^q$ does not encompass the so-called points at infinity $[(x_1,\dots,x_q,0)']$ of the projective space $\mathbb{F}\mathrm{P}^q$.

3. Maximum likelihood estimates, main result

Let P be an arbitrary Borel probability distribution on $\mathbb{F}\mathrm{P}^q$, typically (but not necessarily) the empirical distribution of a sample. Recall that $f_\theta^{\mathscr{G}}$ denotes the density of the angular Gaussian distribution $\mathscr{G}_\theta$ with respect to the uniform distribution $\mathscr{G}_I$ (cf. Eq. (1)). The *angular Gaussian log-likelihood function* for P is the expectation of the logarithm of $f_\theta^{\mathscr{G}}$ evaluated at a random point $[x] \in \mathbb{F}\mathrm{P}^q$ of law P:

$$\ell_P^{\mathscr{G}}(\theta) = \int_{\mathbb{F}\mathrm{P}^q} \log(f_\theta^{\mathscr{G}}([x])\, dP([x]) \quad (\theta \in \Theta_q^{\mathbb{F}}). \tag{4}$$

The *angular Gaussian maximum likelihood estimate* of P is the maximum of $\ell_P^{\mathscr{G}}$ (if it exists and is unique). We denote it by $\mathrm{MLE}^{\mathscr{G}}(P) \in \Theta_q^{\mathbb{F}}$.

The Cauchy analogue of the density $f_\theta^{\mathscr{G}}$ of $\mathscr{G}_\theta$ with respect to the uniform distribution $\mathscr{G}_I$ is not $g_{\mu,\Sigma}$, which is the density (2) of $\mathscr{C}(\mu, \Sigma)$ with respect to the Lebesgue measure, but

$$f_{\mu,\Sigma}^{\mathscr{C}}(y) = g_{\mu,\Sigma}(y)/g_{0,I}(y) \quad (y \in \mathbb{F}^q), \tag{5}$$

which is the density of $\mathscr{C}(\mu, \Sigma)$ with respect to the *standard Cauchy distribution* $\mathscr{C}(0, I)$. So, given an arbitrary Borel probability distribution P on $\mathbb{F}^q$, we define the *Cauchy log-likelihood function* $\ell_P^{\mathscr{C}}(\mu, \Sigma)$ for P as the expectation of the logarithm of the density $f_{\mu,\Sigma}^{\mathscr{C}}$ evaluated at a random vector $y \in \mathbb{F}^q$ of law P

$$\ell_P^{\mathscr{C}}(\mu, \Sigma) = \int_{\mathbb{R}^q} \log f_{\mu,\Sigma}^{\mathbb{C}}(y)\, dP(y). \tag{6}$$

The *Cauchy maximum likelihood estimate* of P is the maximum of $\ell_P^{\mathscr{C}}$ (if it exists and is unique). We denote it by $\mathrm{MLE}^{\mathscr{C}}(P)$.

Of course, if μ, Σ and θ are related by Eq. (3), then $\ell_P^{\mathscr{C}}(\mu, \Sigma) = \ell_P^{\mathscr{G}}(\theta)$, thus $\mathrm{MLE}^{\mathscr{C}}(P) = (\mu, \Sigma)$ if and only if $\mathrm{MLE}^{\mathscr{G}}(P) = \theta$.

Remark 1. The usual Cauchy log-likelihood function is

$$l_P(\mu, \Sigma) = \int_{\mathbb{R}^q} \log g_{\mu,\Sigma}(y)\, dP(y).$$

We prefer Definition (6) because it applies to all probability measures, while the integral $l_P(\mu, \Sigma)$ does not always exist. Anyhow, when $\log g_{0,I}$ is P-integrable, the two log-likelihood functions differ by a constant only since

$$l_P(\mu, \Sigma) = \ell_P^{\mathscr{C}}(\mu, \Sigma) + \int_{\mathbb{R}^q} \log g_{0,I}(y)\, dP(y)$$

in view of Eq. (5).

After these preliminaries, we come to our existence and uniqueness criterion for MLE's. We need some definitions to formulate it.

A *projective subspace of dimension k* of the projective space $\mathbb{F}\mathrm{P}^q$ is the set of axes $[x] \in \mathbb{F}\mathrm{P}^q$ of the non-zero vectors x lying in a linear subspace of dimension $k+1$ of $\mathbb{F}^{q+1}$. Given a Borel probability measure P on $\mathbb{F}\mathrm{P}^q$, call for short a non-trivial projective subspace V of $\mathbb{F}\mathrm{P}^q$ ($V \neq \emptyset$, $V \neq \mathbb{F}\mathrm{P}^q$)

$$\begin{array}{lccc} P\text{-elliptic} & & < & \\ P\text{-parabolic} & \text{if } P(V) & = & \dfrac{\dim V + 1}{q+1}. \\ P\text{-hyperbolic} & & > & \end{array}$$

A P-parabolic projective subspace is *minimal* if it contains no proper P-parabolic projective subspaces.

Let $\mathbb{F}^{q+1} = E_1 \oplus \cdots \oplus E_s$ be a direct sum decomposition of the linear space $\mathbb{F}^{q+1}$ into linear subspaces E_k and let $V_k = \{[x] | x \in E_k, x \neq 0\}$ be the corresponding projective subspaces of $\mathbb{F}\mathrm{P}^q$. We say that $V_1, \ldots, V_s$ is a *direct sum decomposition of the projective space* $\mathbb{F}\mathrm{P}^q$. For example, $\mathbb{F}\mathrm{P}^3$ decomposes into the direct sum of two skew lines, and $\mathbb{F}\mathrm{P}^q$ into the direct sum of $q+1$ points not lying in a proper projective subspace of $\mathbb{F}\mathrm{P}^q$.

Theorem 1. *Let P be an arbitrary Borel probability measure on the* (*real or complex*) *projective space* $\mathbb{F}\mathrm{P}^q$.

(1) *If every non-trivial projective subspace of* $\mathbb{F}\mathrm{P}^q$ *is P-elliptic*, *then the* log-*likelihood function* $\ell_P^{\mathscr{G}}$ *has a unique maximum*: $\mathrm{MLE}^{\mathscr{G}}(P)$ *is well-defined.*
(2) *If* $\mathbb{F}\mathrm{P}^q$ *contains a P-hyperbolic projective subspace*, *then* $\sup_{\theta \in \Theta_q^{\mathbb{F}}} \ell_P^{\mathscr{G}}(\theta) = \infty$: $\mathrm{MLE}^G(P)$ *does not exist.*
(3) *If* $\mathbb{F}\mathrm{P}^q$ *contains no P-hyperbolic subspaces and at least one P-parabolic projective subspace*, *then*
    (a) *if* $\mathbb{F}\mathrm{P}^q$ *decomposes into the direct sum of minimal P-parabolic projective subspaces $V_1, \ldots, V_s$, then the maxima of $\ell_P^{\mathscr{G}}$ form a submanifold of dimension $s-1$ of $\Theta_q^{\mathbb{F}}$*: $\mathrm{MLE}^G(P)$ *is not well-defined*;
    (b) *otherwise* $\ell_P^{\mathscr{G}}$ *admits no maximum*: $\mathrm{MLE}^{\mathscr{G}}(P)$ *does not exist.*

Section 8 presents an explicit description of the set of maxima of $\ell_P^{\mathscr{G}}$ in case 3 (a), by means of the MLE's of the restrictions of P to the parabolic subspaces $V_1, \ldots, V_s$ of the decomposition of $\mathbb{F}\mathrm{P}^q$.

Some special cases are worth mentioning

- If the distribution P is absolutely continuous with respect to the uniform distribution on $\mathbb{F}\mathrm{P}^q$, then $\mathrm{MLE}^{\mathscr{G}}(P)$ is well-defined.
- If the P-probability of some point $[x] \in \mathbb{F}\mathrm{P}^q$ is larger than $1/(q+1)$, then $\mathrm{MLE}^{\mathscr{G}}(P)$ does not exist. In particular, consider the level α contamination $P_\alpha = (1-\alpha)P + \alpha\delta_{[x]}$ of an arbitrary distribution P and let α^* be the breakdown point of $\mathrm{MLE}^{\mathscr{G}}(P_\alpha)$. Then $\alpha^* \leqslant 1/(q+1)$.

- Let $P_n = (\delta_{[x_1]} + \cdots + \delta_{[x_n]})/n$ be the empirical measure of a sample $[x_1], \ldots, [x_n]$ in $\mathbb{F}\mathrm{P}^q$.

  Suppose that $n > q+1$ and that the sample is in general position, i.e., that any non-trivial projective subspace of dimension k of $\mathbb{F}\mathrm{P}^q$ contains at most $k+1$ points of the sample. Then $\mathrm{MLE}^{\mathscr{G}}(P_n)$ is well-defined (see Kent and Tyler [8] for the real case).

  If $n < q+1$, then $\mathrm{MLE}^{\mathscr{G}}(P_n)$ does not exist.

  If $n = q+1$ and the sample is in general position, then the maxima of $\ell^{\mathscr{G}}_{P_n}$ form a submanifold of dimension q of $\Theta^{\mathbb{F}}_q$ (see Olcay and Kent (1998) for the real case).

  If $q = 3$, $n \geqslant 6$ is even and all (distinct) data lie on two skew lines of $\mathbb{F}\mathrm{P}^3$ with half of them on each line, then the maxima of $\ell^{G}_{P_n}$ form a one-dimensional submanifold of $\Theta^{\mathbb{F}}_q$.

Theorem 1 also holds with slight changes for real or complex Cauchy distributions. In this case, P is a Borel probability distribution on $\mathbb{F}^q$ and projective subspaces of $\mathbb{F}\mathrm{P}^q$ should be replaced by affine subspaces of $\mathbb{F}^q$, i.e. translates of linear subspaces.

In the case of the shape space $\Sigma^{q+2}_2 \cong \mathbb{C}\mathrm{P}^q$, the complex field $\mathbb{F} = \mathbb{C}$ is relevant. To apply Theorem 1, we need a convenient description of projective subspaces of $\mathbb{C}\mathrm{P}^q$ in terms of shapes. We propose a construction of such subspaces by means of barycentres in Appendix B.

4. Convexity of log-likelihood

Every parameter $\theta \in \Theta^{\mathbb{F}}_q$ of the angular Gaussian model $(\mathscr{G}_\theta)_{\theta \in \Theta^{\mathbb{F}}_q}$ defines a scalar product and a norm

$$(x|y)_\theta = x^* \theta^{-1} y \quad \text{and} \quad ||x||_\theta = \sqrt{(x|x)_\theta} \;\; (\theta \in \Theta^{\mathbb{F}}_q, x, y \in \mathbb{F}^{q+1}) \tag{7}$$

on the linear space $\mathbb{F}^{q+1}$. Given a probability distribution P on $\mathbb{F}\mathrm{P}^q$, we put

$$\rho_P(\theta) = \int_{\mathbb{F}\mathrm{P}^q} \rho_{[x]}(\theta)\, dP([x]), \quad (\theta \in \Theta^{\mathbb{F}}_q). \tag{8}$$

where

$$\rho_{[x]}(\theta) = \log(||x||^2_\theta / ||x||^2) \quad ([x] \in \mathbb{F}\mathrm{P}^q)$$

In view of Eqs. (1) and (4), ρ_P is the angular Gaussian log-likelihood function $\ell^{\mathscr{G}}_P$ up to a constant negative factor.

We study the minima of ρ_P, i.e. the maxima of $\ell^{\mathscr{G}}_P$, by restricting the function ρ_P to certain curves $\gamma : \mathbb{R} \to \Theta^{\mathbb{F}}_q$, which should be sufficiently general for catching all minima of the function ρ_P.

Given a square matrix v of order $q+1$ with coefficients in $\mathbb{F}$, we call a curve $\gamma : \mathbb{R} \to \Theta^{\mathbb{F}}_q$ a *geodesic of velocity* v if it satisfies the differential equation $\dot{\gamma}(t) = v\gamma(t)$.

The solution $\gamma(t) = (\exp tv)\gamma(0)$ must lie in $\Theta_q^{\mathbb{F}}$, i.e., the matrices $\gamma(t)$ must be self-adjoint of determinant 1 for all $t \in \mathbb{R}$. This imposes some conditions on the matrix v, besides the obvious condition $\gamma(0) \in \Theta_q^{\mathbb{F}}$ on the starting point.

Given $\theta \in \Theta_q^{\mathbb{F}}$, call a matrix A of order $q+1$ *self-θ-adjoint* if $(Ax|y)_\theta = (x|Ay)_\theta$ for all $x, y \in \mathbb{F}^{q+1}$, i.e., if it coincides with its *θ-adjoint* $\theta A^* \theta^{-1}$. Denote by S_θ the linear space of self-θ-adjoint matrices of order $q+1$ and trace 0. Then, the condition on v we are looking for is

$$(\exp tv)\theta \in \Theta_q^{\mathbb{F}} \text{ for all } t \in \mathbb{R} \;\Leftrightarrow\; \theta \in \Theta_q^{\mathbb{F}} \text{ and } v \in S_\theta.$$

In fact, if the matrix $\gamma(t) = (\exp tv)\theta$ is self-adjoint for all $t \in \mathbb{R}$, then its derivative $\dot{\gamma}(0) = v\theta$ must be self-adjoint too, so $\theta v^* \theta^{-1} = v$, i.e., v is self-θ-adjoint. Conversely, if $\theta v^* \theta^{-1} = v$, then $\theta (v^k)^* \theta^{-1} = v^k$ for all non-negative integers k, so $\theta(\exp tv)^* \theta^{-1} = \exp tv$, hence $\gamma(t)^* = \gamma(t)$. Moreover, the determinant of the matrix $\gamma(t)\theta^{-1} = \exp tv$ is 1 for all $t \in \mathbb{R}$ if and only if the trace of v is 0.

Remark 2. As will be shown in Appendix A, the curves $\gamma(t) = (\exp tv)\theta$ $(t \in \mathbb{R})$ with $\theta \in \Theta_q^{\mathbb{F}}$ and $v \in \mathrm{S}_\theta$ are exactly the geodesics—in the sense of differential geometry—of the parameter space $\Theta_q^{\mathbb{F}}$ endowed with its symmetric space structure.

A real-valued function f on $\Theta_q^{\mathbb{F}}$ is called *convex* if its restriction $f(\gamma(t))$ $(t \in \mathbb{R})$ to any geodesic γ is convex in the usual sense (see e.g. [3, 1.6.4 p.24]).

Theorem 2. *For any Borel probability measure P on the (real or complex) projective space $\mathbb{F}\mathrm{P}^q$, the angular Gaussian log-likelihood ρ_P is a convex function on the parameter space $\Theta_q^{\mathbb{F}}$. More precisely, its restriction to a geodesic of velocity v is strictly convex if the measure P does not concentrate on the eigenaxes of the matrix v, and affine linear otherwise.*

Proof. In view of Eqs. (8) the restriction of the log-likelihood function ρ_P along a geodesic γ of velocity v is

$$\rho_P^\gamma(t) = \rho_P(\gamma(t)) = \int_{\mathbb{F}\mathrm{P}^q} \rho_{[x]}^\gamma(t)\, dP([x]), \quad (t \in \mathbb{R}). \tag{9}$$

where

$$\rho_{[x]}^\gamma(t) = \rho_{[x]}(\gamma(t)) = \log(||x||_{\gamma(t)}^2 / ||x||^2) \quad ([x] \in \mathbb{F}\mathrm{P}^q)$$

In order to compute the derivatives of $\rho_{[x]}^\gamma$, put $g(t) = ||x||_{\gamma(t)}^2$. Then

$$\dot{\rho}_{[x]}^\gamma(t) = \frac{\dot{g}(t)}{g(t)} \quad \text{and} \quad \ddot{\rho}_{[x]}^\gamma(t) = \frac{\ddot{g}(t)g(t) - \dot{g}(t)^2}{g(t)^2}.$$

Given arbitrary vectors x and $y \in \mathbb{F}^{q+1}$,

$$\frac{d}{dt}(x|y)_{\gamma(t)} = \frac{d}{dt}(x^*\gamma(t)^{-1}y) = -x^*\gamma(t)^{-1}\dot{\gamma}(t)\gamma(t)^{-1}y = -x^*\gamma(t)^{-1}vy$$
$$= -(x|vy)_{\gamma(t)}.$$

Thus $\dot{g}(t) = -(x|vx)_{\gamma(t)} = -(vx|x)_{\gamma(t)}$ and $\ddot{g}(t) = (vx|vx)_{\gamma(t)}$, hence

$$\dot{\rho}^{\gamma}_{[x]}(t) = -\frac{(x|vx)_{\gamma(t)}}{(x|x)_{\gamma(t)}} \tag{10}$$

and

$$\ddot{\rho}^{\gamma}_{[x]}(t) = \frac{(vx|vx)_{\gamma(t)}(x|x)_{\gamma(t)} - (x|vx)^2_{\gamma(t)}}{(x|x)^2_{\gamma(t)}}. \tag{11}$$

By the Cauchy–Schwarz inequality, $\ddot{\rho}^{\gamma}_{[x]}(t) \geqslant 0$ for all $t \in \mathbb{R}$. Moreover, $\ddot{\rho}^{\gamma}_{[x]}(t) = 0$ for some $t \in \mathbb{R}$ if and only if vx is a multiple of x, i.e., iff $[x]$ is an eigenaxis of v.

Let $||v||_{\mathrm{sp}}$ be the spectral norm of the matrix v, i.e. the largest eigenvalue of v in absolute value. Then $||vx||_{\gamma(t)} \leqslant ||v||_{\mathrm{sp}}||x||_{\gamma(t)}$, hence $|\dot{\rho}^{\gamma}_{[x]}(t)| \leqslant ||v||_{\mathrm{sp}}$ and $|\ddot{\rho}^{\gamma}_{[x]}(t)| \leqslant ||v||^2_{\mathrm{sp}}$ for all $[x] \in \mathbb{F}\mathrm{P}^q$ and $t \in \mathbb{R}$. So, by Lebesgue's dominated convergence theorem, the order of integration and differentiation may be interchanged:

$$\dot{\rho}^{\gamma}_P(t) = \int_{\mathbb{F}\mathrm{P}^q} \dot{\rho}^{\gamma}_{[x]}(t)\, dP([x]) \text{ and } \ddot{\rho}^{\gamma}_P(t) = \int_{\mathbb{F}\mathrm{P}^q} \ddot{\rho}^{\gamma}_{[x]}(t)\, dP([x]). \tag{12}$$

This shows that $\ddot{\rho}^{\gamma}_P(t) \geqslant 0$ for all $t \in \mathbb{R}$ and that $\ddot{\rho}^{\gamma}_P(t) = 0$ for some $t \in \mathbb{R}$ if and only if the measure P concentrates on the eigenaxes of v, hence the theorem. □

5. Maximum likelihood equation

Theorem 3. *Let P be a Borel probability measure on $\mathbb{F}\mathrm{P}^q$. A parameter $\theta \in \Theta^{\mathbb{F}}_q$ is a* MLE *of P, i.e. minimizes ρ_P, if and only if*

$$\int_{\mathbb{F}\mathrm{P}^q} \left(\frac{|(a|x)_\theta|}{||a||_\theta ||x||_\theta}\right)^2 dP([x]) = \frac{1}{q+1} \quad \textit{for all } [a] \in \mathbb{F}\mathrm{P}^q, \tag{13}$$

or, equivalently, if and only if θ solves the equation

$$\frac{\theta}{q+1} = \int_{\mathbb{F}\mathrm{P}^q} \frac{xx^*}{x^*\theta^{-1}x}\, dP([x]). \tag{14}$$

In the real case, (14) gives Eq. (2) obtained by Tyler [15, p. 580], in the case of the empirical measure of a sample. We need the following basic property of geodesics.

Proposition 1. *Given* $\theta_0, \theta_1 \in \Theta_q^{\mathbb{F}}$, *there is a unique geodesic* $\gamma : \mathbb{R} \to \Theta_q^{\mathbb{F}}$ *such that* $\gamma(0) = \theta_0$ *and* $\gamma(1) = \theta_1$. *In other words, the* geodesic exponential map $\mathrm{Exp}_{\theta_0} : \mathrm{S}_{\theta_0} \to \Theta_q^{\mathbb{F}}$, *defined by* $\mathrm{Exp}_{\theta_0} h = (\exp h)\theta_0$, *is one-to-one and onto.*

A direct verification is not difficult. In fact, this result holds for any Hadamard manifold, i.e. any simply connected complete Riemannian manifold of non-positive sectional curvature (see e.g. [3, 1.1.4, p. 19]).

The proof of Theorem 3 uses another ingredient.

Lemma 1. *Every matrix* $v \in \mathrm{S}_\theta$ $(\theta \in \Theta_q^{\mathbb{F}})$ *is a linear combination*

$$v = \sum_{k=1}^{q+1} \lambda_k w_{a_k} \quad (\lambda_k \in \mathbb{R}),$$

where $a_1, \ldots, a_{q+1}$ *are non-zero vectors of* $\mathbb{F}\mathrm{P}^q$ *and* $w_a \in \mathrm{S}_\theta$ *is defined by*

$$w_a x = \frac{(a|x)_\theta}{(a|a)_\theta} a - \frac{1}{q+1} x \quad (a, x \in \mathbb{F}^{q+1}, a \neq 0). \tag{15}$$

Being self-θ-adjoint, the matrix $v \in \mathrm{S}_\theta$ is diagonalizable and its eigenvalues are real. Let $a_1, \ldots, a_{q+1} \in \mathbb{F}^{q+1}$ be eigenvectors of v with $(a_i|a_j)_\theta = 0$ for $i \neq j$, and $\lambda_1, \ldots, \lambda_{q+1}$ the corresponding eigenvalues.

Let π_a denote the θ-orthogonal projector onto an axis $[a] \in \mathbb{F}\mathrm{P}^q$. It is given by $\pi_a x = (a|a)_\theta^{-1}(a|x)_\theta a$ $(x \in \mathbb{F}^{q+1})$ i.e. $\pi_a = (a^*\theta^{-1}a)^{-1}aa^*\theta^{-1}$ in matrix notation. As v is diagonal with respect to the θ-orthogonal base $a_1, \ldots, a_{q+1}$, we can write it as $v = \sum_{k=1}^{q+1} \lambda_k \pi_{a_k}$. But $\sum_{k=1}^{q+1} \lambda_k = 0$ since the trace of v is 0. So, $v = \sum_{k=1}^{q+1} \lambda_k w_{a_k}$, where $w_a = \pi_a - (q+1)^{-1} I \in \mathrm{S}_\theta$.

Proof of Theorem 3. If ρ_P has a minimum at $\theta \in \Theta_q^{\mathbb{F}}$ then $\dot{\rho}_P^\gamma(0) = 0$ for all geodesics γ with $\gamma(0) = \theta$. We observe that the converse also holds. In fact, given any $\theta_1 \in \Theta_q^{\mathbb{F}}$, let γ be the unique geodesic with $\gamma(0) = \theta$ and $\gamma(1) = \theta_1$ (Proposition 1). By hypothesis, $\dot{\rho}_P^\gamma(0) = 0$. Thus, by convexity of ρ_P^γ (Theorem 2), $\rho_P(\theta) = \rho_P^\gamma(0) \leqslant \rho_P^\gamma(1) = \rho_P(\theta_1)$, which shows that θ is a minimum of ρ_P.

On the other hand, by Eqs. (10) and (12), the condition $\dot{\rho}_P^\gamma(0) = 0$ means

$$\int_{\mathbb{F}\mathrm{P}^q} \frac{(x|vx)_\theta}{(x|x)_\theta} dP([x]) = 0 \quad \text{for all } v \in \mathrm{S}_\theta.$$

According to Lemma 1, this amounts to

$$\int_{\mathbb{F}\mathrm{P}^q} \frac{(x|w_a x)_\theta}{(x|x)_\theta} dP([x]) = 0 \quad \text{for all } a \in \mathbb{F}^{q+1} \backslash \{0\},$$

which is condition (13) by taking account of Definition 15. Eq. (14) is just a rewriting of (13): $(a|x)_\theta^2 = a^*\theta^{-1}xx^*\theta^{-1}a$, and (13) becomes

$$\frac{a^*\theta^{-1}a}{q+1} = a^*\theta^{-1}\left(\int_{\mathbb{F}\mathrm{P}^q} \frac{xx^*}{(x|x)_\theta^2}\, dP([x])\right)\theta^{-1}a \quad \text{for all } [a]\in\mathbb{F}\mathrm{P}^q,$$

giving the result. □

6. Behaviour of log-likelihood at infinity

Let $\gamma(t) = (\exp tv)\theta$ $(t\in\mathbb{R})$ be the geodesic of non-zero velocity $v\in \mathrm{S}_\theta$ issuing from $\theta = \gamma(0)$. We are interested in the limit of the log-likelihood $\rho_P^\gamma(t)$ along γ, or rather of its derivative $\dot\rho_P^\gamma(t)$, as $t\to\infty$. The result depends on the spectral decomposition of the matrix v.

As the matrix v is self-θ-adjoint, its eigenvalues are real. We also call them the *eigenvalues* $\lambda_1 > \lambda_2 > \cdots > \lambda_s$ *of the geodesic* γ. The corresponding eigenspaces $E_1, \dots, E_s$ are pairwise θ-orthogonal, i.e., $(x_i|x_j)_\theta = 0$ for $x_i\in E_i$, $x_j\in E_j$ and $i\neq j$. Moreover, $\mathbb{F}^{q+1} = E_1\oplus\cdots\oplus E_s$. Let π_k be the θ-orthogonal projector onto E_k, i.e. the unique self-θ-adjoint matrix with range E_k, such that $\pi_k\pi_k = \pi_k$. Note that $\pi_i\pi_j = 0$ for $i\neq j$ and that $\pi_1+\cdots+\pi_s = I$. According to the spectral theorem, $v = \lambda_1\pi_1+\cdots+\lambda_s\pi_s$ and $\exp tv = e^{\lambda_1 t}\pi_1+\cdots+e^{\lambda_s t}\pi_s$. Given $[x]\in\mathbb{F}\mathrm{P}^q$, it follows according to Definition 7 that

$$(x|x)_{\gamma(t)} = (x|(\exp - tv)x)_\theta = e^{-\lambda_1 t}||\pi_1 x||_\theta^2 + \cdots + e^{-\lambda_s t}||\pi_s x||_\theta^2$$

and

$$(x|vx)_{\gamma(t)} = (x|(\exp - tv)vx)_\theta = \lambda_1 e^{-\lambda_1 t}||\pi_1 x||_\theta^2 + \cdots + \lambda_s e^{-\lambda_s t}||\pi_s x||_\theta^2.$$

Thus, by Eq. (10),

$$\dot\rho_{[x]}^\gamma(t) = -\frac{\lambda_1 e^{-\lambda_1 t}||\pi_1 x||_\theta^2 + \cdots + \lambda_s e^{-\lambda_s t}||\pi_s x||_\theta^2}{e^{-\lambda_1 t}||\pi_1 x||_\theta^2 + \cdots + e^{-\lambda_s t}||\pi_s x||_\theta^2}.$$

Putting $m([x]) = \max\{k = 1,\dots,s|\pi_k x\neq 0\}$, we get

$$\lim_{t\to+\infty} \dot\rho_{[x]}^\gamma(t) = -\lambda_{m([x])}.$$

This is a simple function of $[x]$. So we may interchange limit and integration with respect to the probability measure P. Taking Eq. (12) into account, we obtain

$$\lim_{t\to+\infty} \dot\rho_P^\gamma(t) = -\int_{\mathbb{F}\mathrm{P}^q} \lambda_{m([x])}\, dP([x]).$$

In order to put this result into a neater form, consider the projective subspaces

$$\begin{aligned} F_k &= \{[x]\in\mathbb{F}\mathrm{P}^q|\pi_{k+1}x = \pi_{k+2}x = \cdots = \pi_s x = 0\} \\ &= \{[x]\in\mathbb{F}\mathrm{P}^q|x\in E_1+E_2+\cdots+E_k\} \quad (k = 1,\dots,s-1). \end{aligned}$$

The sequence $F_1 \subset \cdots \subset F_{s-1}$ is an important characteristic of the geodesic γ, called its *flag* (see [3, 2.12.8] for the real case). As $m([x]) = k$ if and only if $[x] \in = F_k \backslash F_{k-1}$ (where F_s is defined as $\mathbb{F}\mathrm{P}^q$),

$$\begin{aligned}&\int_{\mathbb{F}\mathrm{P}^q} \lambda_{m([x])}\, dP([x]) \\ &\quad = \lambda_1 P(F_1) + \lambda_2 P(F_2 \backslash F_1) + \cdots + \lambda_{s-1} P(F_{s-1} \backslash F_{s-2}) + \lambda_s P(\mathbb{F}\mathrm{P}^q \backslash F_{s-1}) \\ &\quad = (\lambda_1 - \lambda_2) P(F_1) + \cdots + (\lambda_{s-1} - \lambda_s) P(F_{s-1}) + \lambda_s.\end{aligned}$$

With the *eigenvalue differences* $\alpha_k = \lambda_k - \lambda_{k+1}$ $(k = 1, \ldots, s-1)$,

$$\lim_{t \to +\infty} \dot{\rho}_P^\gamma(t) = -\lambda_s - \sum_{k=1}^{s-1} \alpha_k P(F_k).$$

But the trace of the matrix $v \in \mathrm{S}_0$ is zero, so $\lambda_1 \dim E_1 + \cdots + \lambda_s \dim E_s = 0$. Taking the equalities $\dim F_k + 1 = \dim E_1 + \cdots + \dim E_k$ into account, we find

$$\lambda_s = -\frac{1}{q+1}[\alpha_1(\dim F_1 + 1) + \cdots + \alpha_{s-1}(\dim F_{s-1} + 1)]. \tag{16}$$

Replacing this value of λ_s into the previous equation yields

$$\lim_{t \to +\infty} \dot{\rho}_P^\gamma(t) = \sum_{k=1}^{s-1} \alpha_k \left(\frac{\dim F_k + 1}{q+1} - P(F_k) \right). \tag{17}$$

From this equation and Theorem 2, some conclusions can be drawn on the global behaviour of the functions ρ_P^γ, which we sum up in the following theorem. We denote by $V_1, \ldots, V_s$ the projective subspaces of $\mathbb{F}\mathrm{P}^q$ corresponding to the eigenspaces $E_1, \ldots, E_s$ of the velocity v and call them *projective eigenspaces of the geodesic* γ.

Theorem 4. *Let γ be a non-constant geodesic in $\Theta_q^{\mathbb{F}}$ of flag $F_1 \subset \cdots \subset F_{s-1}$ and projective eigenspaces $V_1, \ldots, V_s$.*

(1) *If no F_k is P-hyperbolic and at least one is P-elliptic, then* $\lim_{t \to \infty} \rho_P^\gamma(t) = +\infty$.
(2) *If no F_k is P-elliptic and at least one is P-hyperbolic, then ρ_P^γ is strictly decreasing from $+\infty$ to $-\infty$.*
(3) *If every F_k is P-parabolic and $P(V_1 \cup \cdots \cup V_s) < 1$, then ρ_P^γ is strictly decreasing.*
(4) *The function ρ_P^γ is constant if and only if every V_k is P-parabolic.*

Proof. Suppose that no F_k is P-hyperbolic and at least one is P-elliptic. Then $\lim_{t \to +\infty} \dot{\rho}_P^\gamma(t) > 0$ in view of Eq. (17), thus $\lim_{t \to +\infty} \rho_P^\gamma(t) = +\infty$.

If no F_k is P-elliptic and at least one is P-hyperbolic, then $\lim_{t \to +\infty} \dot{\rho}_P^\gamma(t) < 0$ in view of Eq. (17). As the function ρ_P^γ is convex by Theorem 2, it must be strictly decreasing from $+\infty$ to $-\infty$.

If every F_k is P-parabolic, then $\lim_{t \to +\infty} \dot{\rho}_P^\gamma(t) = 0$ by Eq. (17). Moreover, if $P(V_1 \cup \cdots \cup V_s) < 1$, the measure P does not concentrate on the eigenaxes of the velocity of γ. By Theorem 2, the function ρ_P^γ is strictly convex, thus strictly decreasing.

If every V_k is P-parabolic, then $P(V_1 \cup \cdots \cup V_s) = P(V_1) + \cdots + P(V_s) = [(\dim V_k + 1) + \cdots + (\dim V_s + 1)]/(q+1) = 1$. Thus the function ρ_P^γ is affine linear by Theorem 2. On the other hand, $V_1 \cup \cdots V_k \subseteq F_k$ and $F_k \cap (V_{k+1} \cup \cdots V_s) = \emptyset$, so $P(F_k) = P(V_1) + \cdots + P(V_k) = [(\dim V_1 + 1) + \cdots + (\dim V_k + 1)]/(q+1) = (\dim F_k + 1)/(q+1)$. By Eq. (17), $\lim_{t\to+\infty} \dot{\rho}_P^\gamma(t) = 0$, thus the function ρ_P^γ must be constant.

Suppose conversely that the function ρ_P^γ is constant. Then $P(V_1 \cup \cdots \cup V_s) = 1$ by Theorem 2 and each F_k is P-parabolic by Eq. (17). From the relations $P(F_k \cup V_{k+1} \cup \cdots \cup V_s) = 1$ and $F_k \cap (V_{k+1} \cup \cdots \cup V_s) = \emptyset$ follows $P(F_k) + P(V_{k+1} + \cdots + P(V_s)) = 1$. By recursion on $k = s, s-1, \ldots, 1$, we find that each V_k is P-parabolic. □

7. Existence and uniqueness of MLE's

This section and the next one are devoted to the proof of Theorem 1. We need the following property of geodesics:

Proposition 2. *Let $F_1 \subset \cdots \subset F_{s-1}$ be distinct non-trivial projective subspaces of $\mathbb{F}\mathrm{P}^q$ and $\alpha_1, \ldots, \alpha_{s-1} > 0$ positive real numbers ($s \geqslant 2$). For every parameter $\theta \in \Theta_q^{\mathbb{F}}$, there is a unique geodesic γ issuing from $\gamma(0) = \theta$, of flag $F_1 \subset \cdots \subset F_{s-1}$ and eigenvalue differences $\alpha_1, \ldots, \alpha_{s-1}$ (i.e., $\alpha_k = \lambda_k - \lambda_{k+1}$ where $\lambda_1 > \cdots > \lambda_s$ are the eigenvalues of the velocity of γ).*

Proof. Consider the linear subspaces $G_1 \subset \cdots \subset G_{s-1}$ corresponding to $F_1 \subset \cdots \subset F_{s-1}$ ($F_k = \{[x] \in \mathbb{F}\mathrm{P}^q | x \in G_k, x \neq 0\}$) and put $G_s = \mathbb{F}^{q+1}$. Let $E_1 = G_1$ and define recursively $E_{k+1} = \{y \in G_{k+1} | (x|y)_\theta = 0 \text{ for all } x \in E_k\}$. Then $E = E_1 \oplus \cdots \oplus E_s$. Let π_k be the θ-orthogonal projector onto E_k and define λ_s by equation (16) and $\lambda_k = \lambda_{k+1} + \alpha_k$ ($k = s-1, \ldots, 1$). The geodesic we are looking for is $\gamma(t) = e^{\lambda_1 t}\pi_1 + \cdots + e^{\lambda_s t}\pi_s$. It is unique because the last equation must be the spectral decomposition of γ. □

Proof of Theorem 1. *Part* 1: Suppose that every non-trivial projective subspace of $\mathbb{F}\mathrm{P}^q$ is P-elliptic. We show that the log-likelihood function ρ_P admits a unique minimum.

For the existence proof, choose an arbitrary parameter $\theta_0 \in \Theta_q^{\mathbb{F}}$ and consider the composite function $f : \mathrm{S}_{\theta_0} \to \mathbb{R}$ given by $f(h) = \rho_P(\mathrm{Exp}_{\theta_0}(h))$, where Exp_{θ_0} is the geodesic exponential defined in Proposition 1. For any geodesic $\gamma(t) = \mathrm{Exp}_{\theta_0}\, tv$ ($v \in \mathrm{S}_\theta$, $v \neq 0$), $\lim_{t\to+\infty} \rho_P^\gamma(t) = \lim_{t\to+\infty} f(tv) = +\infty$ according to the first part of Theorem 3. This ensures the existence of a minimum of f, hence of ρ_P.

For the uniqueness of the minimum θ_0 of ρ_P, consider a parameter $\theta_1 \neq \theta_0$ and let γ be the geodesic with $\gamma(0) = \theta_0$ and $\gamma(1) = \theta_1$ (Proposition 1). Let $V_1, \ldots, V_s$ be the projective eigenspaces of γ. Then $P(V_1 \cup \cdots \cup V_s) = P(V_1) + \cdots + P(V_s) < [(\dim V_k + 1) + \cdots + (\dim V_s + 1)]/(q+1) = 1$ according to the P-ellipticity condition. By Theorem 2, ρ_P^γ is strictly convex. So, $\rho_P(\theta_0) = \rho_P^\gamma(0) < \rho_P^\gamma(1) = \rho_P(\theta_1)$.

Part 2: Suppose that $\mathbb{F}\mathrm{P}^q$ contains a P-hyperbolic projective subspace V. Let γ be a geodesic of flag reduced to V (cf. Proposition 2). According to the second point of Theorem 4, ρ_P^γ decreases from $+\infty$ to $-\infty$. So $\inf_{\theta \in \Theta_q^{\mathbb{F}}} \rho_P(\theta) = -\infty$ and ρ_P has no minimum.

Part 3: Suppose that $\mathbb{F}\mathrm{P}^q$ contains a P-parabolic projective subspace. We first prove the alternative: either ρ_P has no minimum or $\mathbb{F}\mathrm{P}^q$ decomposes into the direct sum of minimal P-parabolic projective subspaces. This requires a preparation.

Lemma 2. *Let F be a P-parabolic projective subspace of $\mathbb{F}\mathrm{P}^q$. If $\mathbb{F}\mathrm{P}^q$ does not decompose into the direct sum of F and another projective subspace $\tilde{F}$, then ρ_P has no minimum.*

Proof. Let θ be an arbitrary parameter and let γ be a geodesic issuing from $\gamma(0) = \theta$ of flag reduced to F (cf. Proposition 2). The eigenaxes of the velocity matrix $v \in \mathrm{S}_\theta$ lie either in F or in the complemetary subspace

$$\tilde{F} = \{[y] \in \mathbb{F}\mathrm{P}^q | (x|y)_\theta = 0 \text{ for all } [x] \in F\}.$$

By hypothesis, $\tilde{F}$ is not P-parabolic, hence $P(F \cup \tilde{F}) = P(F) + P(\tilde{F}) < 1$. According to the third part of Theorem 4, the function ρ_P^γ is strictly monotone decreasing. Thus the parameter θ does not minimize ρ_P. □

Now, let $V_1 = F_1$ be a minimal P-parabolic subspace of $\mathbb{F}\mathrm{P}^q$. Suppose that $\mathbb{F}\mathrm{P}^q$ decomposes into the direct sum of F_1 and some projective subspace $\tilde{F}_1$. In the opposite case, ρ_P has no minimum according to Lemma 2. Let V_2 be a minimal P-parabolic subspace of $\tilde{F}_2$. If $V_2 = \tilde{F}_2$, then $\mathbb{F}\mathrm{P}^q$ is the direct sum of the minimal P-parabolic subspaces V_1 and V_2. Otherwise, we apply the same process to the smallest projective subspace F_2 containing V_1 and V_2, and so on. At the end, either we find that ρ_P has no minimum or we get a direct sum decomposition of $\mathbb{F}\mathrm{P}^q$ into minimal P-parabolic projective subspaces $V_1, \ldots, V_s$. □

8. The decomposable case

In this section, we make precise and prove part 3(a) of Theorem 1. It is our purpose to describe the set of all angular Gaussian maximum likelihood estimates of P, i.e. of all minima of ρ_P. Perhaps the neatest way to achieve this goal is to work directly with positive definite quadratic forms on linear spaces, rather than representing them by positive definite matrices.

Let E be a linear space over $\mathbb{F}$. A *quadratic form* $p : E \to \mathbb{R}$ can be written in a unique way as $p(x) = (x|x)_p$ $(x \in E)$, where $(x|y)_p \in \mathbb{F}$ $(x, y \in E)$ is a symmetric bilinear form in the real case, and a Hermitian form in the complex case. Let $\mathrm{PD}(E)$ be the set of positive definite quadratic forms on E.

We denote by $\Theta(E) = \{[p] | p \in \mathrm{PD}(E)\}$ $([p] = \{\lambda p | \lambda > 0\})$ the space of positive quadratic forms on E, up to a positive factor. The parameter space $\Theta_q^{\mathbb{F}}$ can be identified with $\Theta(\mathbb{F}^{q+1})$ by associating to the positive definite self-adjoint matrix $\theta \in \Theta_q^{\mathbb{F}}$ the class $[p] \in \Theta(\mathbb{F}^{q+1})$ of the quadratic form $p(x) = x^* \theta^{-1} x$ $(x \in \mathbb{F}^{q+1})$.

Let $\theta = [p] \in \Theta(E)$. Given a linear subspace E_1 of E, we call $\theta_1 = [p_1] \in \Theta(E_1)$ the *restriction* of θ to E_1 if $p_1(x) = p(x)$ for all $x \in E_1$. We say that two linear subspaces E_1 and E_2 of E are θ-*orthogonal* if $(x_1|x_2)_p = 0$ for all $x_1 \in E_1$ and $x_2 \in E_2$.

Theorem 5. *Let P be a Borel probability measure on $\mathbb{F}\mathrm{P}^q$ and suppose that $\mathbb{F}\mathrm{P}^q$ contains no P-hyperbolic projective subspaces. Let $\mathbb{F}^{q+1} = E_1 \oplus \cdots \oplus E_s$ be a direct sum decomposition into linear subspaces and let $V_k = \{[x] | x \in E_k, x \neq 0\}$ be the projective space corresponding to E_k. Suppose that all V_k are minimal P-parabolic. Let P_k be the Borel probability measure on V_k defined by*

$$P_k(A) = \frac{q+1}{\dim V_k + 1} P(A) \quad \text{for every Borel subset } A \text{ of } V_k. \tag{18}$$

Then, every P_k has a unique angular Gaussian maximum likelihood estimate $\theta_k = [p_k] = \mathrm{MLE}^{\mathscr{G}}(P_k) \in \Theta(E_k)$.

Moreover, a parameter $\theta \in \Theta(\mathbb{F}^{q+1}) \cong \Theta_q^{\mathbb{F}}$ is an angular Gaussian maximum likelihood estimate of P if and only if $E_1, \ldots, E_s$ are pairwise θ-orthogonal and the restriction of θ to E_k is θ_k for $k = 1, \ldots, s$.

Corollary 1. *Under the hypotheses of the preceding theorem, put $\theta_k = [p_k]$. Then, the set of angular Gaussian maximum likelihood estimates of P, i.e. the set of minima of ρ_P, consists of those $\theta = [p] \in \Theta(\mathbb{F}^{q+1}) \cong \Theta_q^{\mathbb{F}}$ that can be represented by a positive definite quadratic form p on E defined by*

$$p(x_1 + \cdots + x_s) = \lambda_1 p_1(x_1) + \cdots + \lambda_s p_s(x_s) \quad \text{for all } x_k \in E_k, \tag{19}$$

where $\lambda_1, \ldots, \lambda_s$ are positive real numbers. It is a submanifold of dimension $s-1$ of $\Theta_q^{\mathbb{F}}$ since we can choose $\lambda_k = e^{\alpha_k}$ for $k = 1, \ldots, s-1$ and $\lambda_s = 1$ with arbitrary $(\alpha_1, \ldots, \alpha_{s-1}) \in \mathbb{R}^{s-1}$.

Proof of Corollary 1. Suppose, as in the conclusion of Theorem 5, that $\theta_k = [p_k]$ is the restriction of $\theta = [p]$ to E_k and that $(x_j|x_k)_p = 0$ for all $x_j \in E_j$, $x_k \in E_k$ and $j \neq k$. Then, there is a positive real number λ_k such that $p(x_k) = \lambda_k p_k(x_k)$ for any $x_k \in E_k$ and

$$\begin{aligned} p(x_1 + \cdots + x_s) &= (x_1 + \cdots + x_s | x_1 + \cdots + x_s)_p = (x_1|x_1)_p + \cdots + (x_s|x_s)_p \\ &= p(x_1) + \cdots + p(x_s) = \lambda_1 p_1(x_1) + \cdots + \lambda_s p_s(x_s). \end{aligned}$$

Suppose conversely that condition (19) holds. In particular, $p(x_k) = \lambda_k p_k(x_k)$ for any $x_k \in E_k$. Thus $[\lambda_k p_k] = [p_k] = \theta_k$ is the restriction of $[p] = \theta$ to E_k. Condition (19) also implies $p(x_j + x_k) = \lambda_j p_j(x_j) + \lambda_k p_k(x_k) = p(x_j) + p(x_k)$ for $x_j \in E_j$, $x_k \in E_k$ and $j \neq k$. Then

$$\begin{aligned}(x_j|x_j)_p + (x_k|x_k)_p &= (x_j + x_k|x_j + x_k)_p \\ &= (x_j|x_j)_p + (x_k|x_k)_p + (x_j|x_k)_p + (x_k|x_j)_p,\end{aligned}$$

hence $(x_j|x_k)_p + (x_k|x_j)_p = 0$. In the real case, $(x_k|x_j)_p = (x_j|x_k)_p$ so $(x_j|x_k)_p = 0$. In the complex case $(x_k|x_j)_p$ is the conjugate of $(x_j|x_k)_p$, thus the real part of $(x_j|x_k)_p$ is zero. The imaginary part of $(x_j|x_k)_p$ is also zero since it is the real part of $(ix_j|x_k)$ ($i = \sqrt{-1}$). In both cases, the linear subspaces $E_1, \ldots, E_s$ are pairwise θ-orthogonal. $\square$

Proof of Theorem 5. By hypothesis, V_k contains neither P-hyperbolic nor P-parabolic proper projective subspaces. In other words, taking Eq. (18) into account, all non-trivial projective subspaces of V_k are P_k-elliptic. Thus, by the first point of Theorem 1, $\theta_k = \mathrm{MLE}^{\mathscr{G}}(P_k) \in \Theta(V_k)$ is well-defined.

We come to the main part of Theorem 5. Suppose first that the restriction of $\theta \in \Theta(\mathbb{F}^{q+1})$ to E_k is θ_k for $k = 1, \ldots, s$ and that $(a_i|a_j)_p = 0$ for all $a_i \in E_i$, $a_j \in E_j$ and $i \neq j$. We prove that $\theta = [p]$ satisfies the maximum likelihood equation of Theorem 3. As $P(V_1 \cup \cdots \cup V_s) = 1$,

$$\begin{aligned}(q+1) \int_{\mathbb{F}\mathrm{P}^q} \frac{|(a|x)_p|^2}{(x|x)_p} dP([x]) &= \sum_{k=1}^{s} (q+1) \int_{V_k} \frac{|(a|x)_p|^2}{(x|x)_p} dP([x]) \\ &= \sum_{k=1}^{s} (dim\ V_k + 1) \int_{V_k} \frac{|(a|x)_p|^2}{(x|x)_p} dP_k([x])\end{aligned}$$

for any $a \in \mathbb{F}^{q+1}$. We can write $a = a_1 + \cdots + a_s$ with $a_k \in E_k$ since $\mathbb{F}^{q+1} = E_1 \oplus \cdots \oplus E_s$. As the restriction θ_k of $\theta = [p]$ to E_k is the MLE of P_k, it satisfies the maximum likelihood equation (13)

$$(\dim V_k + 1) \int_{V_k} \frac{|(a_k|x)_p|^2}{(x|x)_p} dP_k([x]) = (a_k|a_k)_p$$

of Theorem 3. In this equation, $(a_k|x)_p = (a|x)_p$ since $x \in E_k$, hence

$$\begin{aligned}(q+1) \int_{\mathbb{F}\mathrm{P}^q} \frac{|(a|x)_p|^2}{(x|x)_p} dP([x]) &= \sum_{k=1}^{s} (\dim V_k + 1) \int_{V_k} \frac{|(a|x)_p|^2}{(x|x)_p} dP_k([x]) \\ &= \sum_{k=1}^{s} (a_k|a_k)_p = (a|a)_p.\end{aligned}$$

So, by Theorem 3, θ is a MLE of P. This proves the 'if' implication of Theorem 5.

Conversely, let $\tilde{\theta} \in \Theta_q^{\mathbb{F}}$ be an arbitrary angular Gaussian maximum likelihood estimate of P. Choose a parameter $\theta \in \Theta(\mathbb{F}^{q+1}) \cong \Theta_q^{\mathbb{F}}$ such that $E_1, \ldots, E_s$ are pairwise θ-orthogonal and the restriction of θ to E_k is θ_k for $k = 1, \ldots, s$. Let γ be the geodesic with $\gamma(0) = \theta$ and $\gamma(1) = \tilde{\theta}$ (Proposition 1). As both θ and $\tilde{\theta}$ minimize ρ_P, the function ρ_P^γ has a minimum at 0 and 1. According to Theorem 2, ρ_P^γ is convex, thus $\rho_P^\gamma(t)$ is constant for $0 \leqslant t \leqslant 1$. But ρ_P^γ is either strictly convex or affine linear, so it must be constant.

Let $\tilde{E}_1, \ldots, \tilde{E}_r$ be the eigenspaces of the velocity $v \in \mathrm{S}_\theta$ of γ and let $\tilde{V}_k$ be the projective subspace corresponding to $\tilde{E}_k$. As ρ_P^γ is constant, the projective eigenspaces $\tilde{V}_1, \ldots, \tilde{V}_r$ are P-parabolic by the fourth point of Theorem 4. The projective space $\mathbb{F}\mathrm{P}^q$ has two direct sum decompositions $V_1, \ldots, V_s$ and $\tilde{V}_1, \ldots, \tilde{V}_r$ into P-parabolic subspaces. As each V_k is minimal P-parabolic, it can be proven that each $\tilde{E}_j$ is a sum of some E_k's. In other words, the velocity of γ can be written as $v = \lambda_1 \pi_1 + \cdots + \lambda_s \pi_s$, where π_k is the θ-orthogonal projector onto E_k and $\lambda_1, \ldots, \lambda_s$ are (non-necessarily distinct) real numbers such that $\sum_{k=1}^s \lambda_k \dim E_k = 0$. It follows that $\gamma(t) = [e^{\lambda_1 t}\pi_1 + \cdots + e^{\lambda_s t}\pi_s]\theta$, in particular $\tilde{\theta} = \gamma(1) = [e^{\lambda_1}\pi_1 + \cdots + e^{\lambda_s}\pi_s]\theta$. Given $x_i \in E_i$ and $x_j \in E_j$ with $i \neq j$, $(x_i|x_j)_{\tilde{\theta}} = (x_i|[e^{-\lambda_1}\pi_1 + \cdots + e^{-\lambda_s}\pi_s]x_j)_\theta = 0$ since $\pi_i x_j = 0$ for $i \neq j$. This proves that the subspaces $E_1, \ldots, E_s$ are $\tilde{\theta}$-orthogonal.

It remains to show that the restriction of $\tilde{\theta} = [\tilde{p}] \in \Theta(\mathbb{F}^{q+1})$ to E_k is the maximum likelihood estimate $\theta_k = \mathrm{MLE}^{\mathscr{G}}(P_k) \in \Theta(E_k)$. By Theorem 3,

$$(q+1) \int_{\mathbb{F}\mathrm{P}^q} \frac{|(a|x)_{\tilde{p}}|^2}{\tilde{p}(a)\tilde{p}(x)} dP([x]) = \sum_{j=1}^{s} (\dim V_j + 1) \int_{V_j} \frac{|(a|x)_{\tilde{p}}|^2}{\tilde{p}(a)\tilde{p}(x)} dP_k([x]) = 1$$

for all $a \in \mathbb{F}^{q+1}$. In particular, for $a = a_k \in E_k$ and $[x] \in V_j$ with $j \neq k$, $(a_k|x)_{\tilde{p}} = 0$ since E_k and E_j are $\tilde{\theta}$-orthogonal. Thus

$$(\dim V_k + 1) \int_{V_k} \frac{|(a_k|x)_{\tilde{p}}|^2}{\tilde{p}(a_k)\tilde{p}(x)} dP_k([x]) = 1$$

for all $a_k \in E_k$. So, by Theorem 3, the restriction of $\tilde{\theta}$ to E_k is a MLE of P_k. As θ_k is the unique MLE of P_k, the restriction of $\tilde{\theta}$ to E_k must be θ_k. This completes the proof of Theorem 5. □

Acknowledgments

The authors would like to thank the referees for their helpful comments on an early version of the paper.

Appendix A. Equivariance and symmetric spaces

Let $g \in \mathrm{SL}(q+1, \mathbb{F})$ be a square matrix of order $q+1$ and determinant 1, and $X \in \mathbb{F}^{q+1}$ a random vector of law $\mathcal{N}(0, \theta)$. The law of the image vector $gX \in \mathbb{F}^{q+1}$ under g is $\mathcal{N}(0, g\theta g^*)$. So, letting g act on axes $[x] \in \mathbb{F}\mathrm{P}^q$ by $g[x] = [gx] \in \mathbb{F}\mathrm{P}^q$, we find that the image measure of the angular Gaussian distribution $\mathcal{G}_\theta$ under g is

$$g\mathcal{G}_\theta = \mathcal{G}_{g\theta g^*} \quad (g \in \mathrm{SL}(q+1, \mathbb{F}),\ \theta \in \Theta_q^{\mathbb{F}}). \tag{A.1}$$

As a consequence, the maximum likelihood estimator is equivariant under the action of $\mathrm{SL}(q+1, \mathbb{F})$: given a probability distribution P on $\mathbb{F}\mathrm{P}^q$,

$$\mathrm{MLE}^{\mathcal{G}}(gP) = g\mathrm{MLE}^{\mathcal{G}}(P)g^* \quad (g \in \mathrm{SL}(q+1, \mathbb{F})), \tag{A.2}$$

where gP is the image measure of P under g. Equivariance has been exploited by McCullagh [14] in the real case of dimension 1.

The isotropy group of the uniform distribution $\mathcal{G}_I$ is the special orthogonal group $\mathrm{SO}(q+1) = \{g \in \mathrm{SL}(q+1, \mathbb{R}) | gg' = I\}$ in the real case, and the special unitary group $\mathrm{SU}(q+1) = \{g \in \mathrm{SL}(q+1, \mathbb{C}) | gg^* = I\}$ in the complex case. Moreover, the group $\mathrm{SL}(q+1, \mathbb{F})$ acts transitively on $\Theta_q^{\mathbb{F}}$ since any matrix $\theta \in \Theta_q^{\mathbb{F}}$ can be decomposed as $\theta = gg^*$, where $g \in \mathrm{SL}(q+1, \mathbb{F})$ is e.g. a triangular matrix (Cholesky decomposition). It follows that the parameter space $\Theta_q^{\mathbb{F}}$ is the homogeneous space

$$\Theta_q^{\mathbb{R}} \cong \mathrm{SL}(q+1, \mathbb{R})/\mathrm{SO}(q+1) \quad \text{in the real case} \tag{A.3}$$

and

$$\Theta_q^{\mathbb{C}} \cong \mathrm{SL}(q+1, \mathbb{C})/\mathrm{SU}(q+1) \quad \text{in the complex case.}$$

These are, in fact, symmetric spaces of non-compact type (see [4, Chapters IV, VI]). More details on $\mathrm{SL}(q+1, \mathbb{R})/\mathrm{SO}(q+1)$ can be found in [3].

Appendix B. Barycentric combinations of shapes

Let $(x^{(1)}, \dots, x^{(n)})$ be a configuration of $n \geqslant 2$ labelled points in the plane $\mathbb{R}^2 \cong \mathbb{C}$, not degenerating into a single point. A *plane n-shape* is given by such a configuration, up to an orientation preserving similarity. By a translation, we can represent a shape by a *central configuration* $(x^{(1)}, \dots, x^{(n)})$ with $x^{(1)} + \cdots + x^{(n)} = 0$.

Consider the hyperplane $H_n = \{(x^{(1)}, \dots, x^{(n)}) \in \mathbb{C}^n | x^{(1)} + \cdots + x^{(n)} = 0\}$. The *shape manifold* is the complex projective space $\Sigma_2^n = \{[x] | x \in H_n, x \neq 0\} \cong \mathbb{C}\mathrm{P}^{n-2}$. Here, $[x] = \{\lambda x | \lambda \in \mathbb{C}\} \in \Sigma_2^n$ denotes the shape of a central configuration $x = (x^{(1)}, \dots, x^{(n)}) \in H_n \backslash \{0\}$ of n points in the plane.

Let V be the projective span of given shapes $[x_1], \dots, [x_r] \in \Sigma_2^n$, i.e. the smallest projective subspace of $\Sigma_2^n \cong \mathbb{C}\mathrm{P}^{n-2}$ containing them. It consists of the shapes $[x]$ of all non-zero linear combinations $x = \lambda_1 x_1 + \cdots + \lambda_r x_r$ with $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$. We can also represent the shapes $[x_k]$ by the configurations $y_k = r\lambda_k x_x$ (if some y_k is zero, then cross it). Then $x = (y_1 + \cdots + y_r)/r$. We call the configuration x the *barycentre* of the central configurations $y_1, \ldots, y_r$.

Let us say that a shape $[x] \in \Sigma_2^n$ is a *barycentric combination* of a set S of shapes if the central configuration x can be obtained as the barycentre of a finite set of central configurations $y_1, \ldots, y_r$ with $[y_1], \ldots, [y_r] \in S$. With this terminology,

Proposition 3. *The projective span of a set of shapes in Σ_2^n consists of their barycentric combinations. Consequently, a set of shapes is a projective subspace of the shape manifold Σ_2^n if and only if it is closed under barycentric combinations.*

References

[1] O. Arslan, J.T. Kent, A note on the maximum likelihood estimators for the location and scatter parameters of a multivariate Cauchy distribution, Comm. Statist.—Theory Methods 27 (1998) 3007–3014.
[2] I.L. Dryden, K.V. Mardia, Statistical Shape Analysis, Wiley, Chichester, 1998.
[3] P.B. Eberlein, Geometry of Nonpositively Curved Manifolds, The University of Chicago Press, Chicago, 1996.
[4] S.H. Helgason, Differential Geometry, Lie Groups, and Symmetric Spaces, American Mathematical Society, Providence, RI, 2001.
[5] D.G. Kendall, Shape manifolds, procrustean metrics and complex projective spaces, Bull. Lond. Math. Soc. 16 (1984) 81–121.
[6] J.T. Kent, New directions in shape analysis, in: K.V. Mardia (Ed.), The Art of Statistical Science, Wiley, Chichester, 1992, pp. 115–127.
[7] J.T. Kent, The complex Bingham distribution and shape analysis, J. Roy. Statist. Soc. B 56 (1994) 285–299.
[8] J.T. Kent, D.E. Tyler, Redescending M-estimators of multivariate location and scatter, Ann. Statist. 19 (1991) 2102–2119.
[9] J.T. Kent, D.E. Tyler, Y. Vardi, A curious likelihood identity for the multivariate *t*-distribution, Comm. Statist. Simulation 23 (1994) 441–453.
[10] F.B. Knight, P.A. Meyer, Une caractérisation de la loi de Cauchy, Z. Wahrscheinlichkeit. Verw. Gebiete 34 (1976) 129–134.
[11] K.V. Mardia, Statistics of Directional Data, Academic Press, New York, 1972.
[12] K.V. Mardia, Shape advances and future perspectives, in: K.V. Mardia, C.A. Gill (Eds.), Current Issues in Statistical Shape Analysis, University of Leeds Press, Leeds, 1995, pp. 57–75.
[13] K.V. Mardia, Directional statistics and shape analysis, J. Appl. Statist. 8 (1999) 949–957.
[14] P. McCullagh, Möbius transformations and Cauchy parameter estimation, Ann. Statist. 24 (1996) 787–808.
[15] D.E. Tyler, Statistical analysis for the angular central Gaussian distributions on the sphere, Biometrika 74 (1987) 579–589.
[16] G.S. Watson, Statistics on Spheres, Wiley, New York, 1983.