Intrinsic Statistical Models on Symmetric Spaces

Dustin Pluta

UC-Irvine

Table of contents

Background and Motivation

Motivating Scientific Question

Dynamic Connectivity Models for Neuroimaging

Given a stochastic multivariate time series $Y \in \mathbb{R}^{q \times T}$ with time-dependent covariance $\Sigma_t, t = 1, \cdots, T$, dynamic connectivity models seek to describe and estimate characteristics of the matrix-variate time-series Σ_t.

Some Existing Methods

- Log-euclidean regression
- Hidden Markov models
- Dynamic graph models

Contribution

We instead propose a novel *intrinsic* modeling approach for time-series of SPD matrices, and for time-series on symmetric spaces more generally.

Symmetric Spaces

- *Symmetric spaces* are an important class of Riemannian manifolds with a special structure
- Many real-world phenomena can be viewed as measurements on a symmetric space
- Examples of symmetric spaces are
  - $\mathbb{R}^n$: Euclidean space
  - $\mathbb{S}^{n-1}$: Sphere
  - $\mathbb{H}$: Poincaré upper half plane
  - $\mathbb{P}^n$: space of $n \times n$ symmetric positive definite matrices

Definition: Symmetric Space

A *symmetric space* $(X, d\mu)$ is a Riemannian manifold (with metric $d\mu$) for which a geodesic reversing isometry $f : X \to X$ exists.

That is, for a symmetric space $(X, d\mu)$ with geodesic distance function d, and a geodesic $\gamma(t)$, the function f satisfies

$$f(\gamma(t)) = \gamma(-t),$$

$$d(f(A), f(B)) = d(A, B).$$

Lie Group Perspective

A symmetric space X can also be identified with the quotient of a Lie group G by a normal subgroup H.

Scientific Motivation

Scientific data often live in a symmetric space

- Angular measurements (sphere)
- Measurements affected by general relativity (Minkowski space-time)
- Evolution of quantum states (unitary matrices)
- Covariance matrices (symmetric positive definite matrices).

Modeling these phenomena requires accounting for the special structure of the observational space.

Statistical Motivation

Intrinsic Methods for Symmetric Spaces

- The tail behavior of the intrinsic Gaussian may be more desirable or easier to characterize compared to other approaches.
- In the case of SPD matrices, the log-Euclidean framework has often been adopted because it is computationally convenient. However, little work has been done on comparing the performance of the affine-invariant metric with the log-Euclidean.
- A version of the Central Limit Theorem holds for symmetric spaces, with the asymptotic distribution given as a function of the intrinsic Gaussian.
- The development of this theory is related to many practical problems in engineering and science, but applications in statistical modeling have been limited.

Statistics on Manifolds

In modeling manifold-valued data, the constraints and geometry of the manifold must be accounted for in order to obtain meaningful results.

Some approaches are:

1. *Ad hoc* adjustments to classical (Euclidean) models.
2. *Transformation* of a Euclidean model to the space of interest.
3. *Intrinsic* models built directly on the manifold.

Example: Estimating the Center of a Sample of SPD Matrices

Let $(X_1, \ldots, X_m) \in \mathcal{P}_n^m$ be a sample of SPD matrices.

Approach 1: Ad Hoc

An *ad hoc* method could be to calculate the usual mean $\bar{X} = \frac{1}{n} \sum_i X_i$ and project the result to the closest SPD.

- $\bar{X}$ preserves the trace but inflates the determinant, which is undesirable because it does not preserve volumes.
- It may be possible for the closest SPD to be non-unique.

Approach 2: Transformation

A possible *transformation* method is to map $\mathcal{P}_n$ to the space of symmetric matrices (which is isomorphic to $\mathbb{R}^{n^2}$) by $X \to \log(X)$, where log is the matrix logarithm. The mean of the image can then be calculated and mapped back to $\mathcal{P}_n$ by exp.

$$\tilde{X} = \exp\left\{\frac{1}{m}\sum_{i=1}^{m} \log X_i\right\} = \left(\prod_{i=1}^{m} X_i\right)^{1/m}$$

- This yields the geometric mean of the sample.
- This estimate preserves the determinant.

Approach 3: Intrinsic

An example *intrinsic* approach here is to calculate the Frechét mean X^* wrt the Riemannian metric on $\mathcal{P}_n$ (defined as $d^2(X,Y) = \mathrm{tr}\left\{\log^2\left[X^{-1/2}YX^{-1/2}\right]\right\}$).

$$X^* = \arg\min_{X\in\mathcal{P}_n} \sum_{i=1}^{m} d^2(X, X_i).$$

- This will yield a SPD matrix.
- The formula generalizes the variance-minimizing formula of the usual sample mean.
- More difficult to compute than other methods, requires and iterative optimization routine.

Comparing the Log-Euclidean and Affine-Invariant Metrics

The log-Euclidean (LE) metric is closely related to the affine-invariant (AI) metric on $\mathcal{P}_n$, but a thorough comparison of the intrinsic Gaussians resulting from the two metrics in the context of statistical modeling has not been undertaken, as models using the intrinsic Gaussian for the AI metric have not yet been developed or implemented (as far as I know).

Here are a few observations on differences between the two metrics. Whether these are bugs or features of the AI metric is not clear.

Comparing the Log-Euclidean and Affine-Invariant Metrics

- If $XY = YX$ then $d^2_{AI}(X, Y) = d^2_{LE}(X, Y)$.
- In general, $d^2(X, Y)_{AI} \leq d^2_{LE}(X, Y)$, i.e. the AI metric contracts distances relative to the LE.
- The log-Euclidean metric is not affine-invariant, meaning that for $g \in GL(n, \mathbb{R})$ and $X, Y \in \mathcal{P}_n$, $d^2_{LE}(g^T X g, g^T Y g) \neq d^2_{LE}(X, Y)$ in general, while $d^2_{AI}(g^T X g, g^T Y g) = d^2_{AI}(X, Y)$
- Because $(\mathcal{P}_n, d_{LE}) \cong \mathbb{R}^{\binom{n}{2}}$, the corresponding Lie algebra is commutative, whereas the Lie algebra for $(\mathcal{P}_n, d_{AI})$ is noncommutative.
- The tail behavior of the intrinsic Gaussians for the two metrics will be different, which should result in different inferences in some situations. It's unknown whether these differences are practically significant in application settings.

The Heat Equation on $\mathbb{R}^n$

Definition: The Heat Equation on $\mathbb{R}^n$

$$\begin{cases} \frac{\partial E}{\partial t} - a^2 \Delta E = \delta(x,t) & t > 0, x \in \mathbb{R}^n \\ \Delta = \sum_{j=1}^{n} \frac{\partial^2}{\partial x_j^2} \end{cases}$$

- Δ is the Laplacian operation on $\mathbb{R}^n$,
- a is a constant related to the diffusion rate.

Definition: Fourier Transform

$$\hat{f}(t) = \int_{-\infty}^{\infty} \exp\{-2\pi i t x\} f(x)\, dx.$$

Theorem. Properties of the Fourier Transform

1. $D^a \hat{f} = \widehat{(-2\pi i x)^a f}$
2. $\widehat{D^a f} = (2\pi i x)^a \hat{f}$
3. Convolution: $\widehat{f * g} = \hat{f} \cdot \hat{g}$
4. Fourier Inversion: $\hat{\hat{f}}(x) = f(-x)$

To solve the heat equation, apply the Fourier transform to change the problem into an ODE:

$$\frac{d\hat{E}}{dt} + (2\pi a x)^2 \hat{E} = 1$$
$$\hat{E}(t) = H(t)\exp\{-(2\pi a x)^2 t\}$$
$$E(x,t) = \hat{\hat{E}}(-x,t) = H(t)\exp\{\widehat{-(2\pi a} x)^2 t\}$$
$$E(x,t) = H(t)(2a\sqrt{\pi t})^{-1}\exp\left\{-\frac{1}{2}\frac{\|x\|^2}{4a^2 t}\right\}$$

Fundamental Solution to the Heat Equation on $\mathbb{R}^n$

Fixing $t = 1$ and setting $v = 2a^2$, we write

$$G_v(x) := H(t)(\sqrt{2\pi v})^{-1}\exp\left\{-\frac{1}{2}\frac{\|x\|^2}{2v}\right\}$$

The Heat Equation on Symmetric Spaces

Solving the Heat Equation on Symmetric Spaces

1. Construct the Haar measure of $X = G/K$ and corresponding Riemannian metric invariant under group action of G.
2. Compute the intrinsic Laplacian Δ^* wrt $(X, d\mu)$ by applying the change of variables from $\mathbb{R}^m$ to X by $x \to u$, with Jacobian $A = \partial x/\partial u$.
3. Construct the Fourier transform and corresponding zonal polynomials for $(X, d\mu)$.
4. Derive the fundamental solution of the heat kernel on $(X, d\mu)$, and use this to define the intrinsic Gaussian on $(X, d\mu)$.

Let $X = G/K$ be a symmetric space with G-invariant Riemannian metric $d\mu$.

Definition: The Intrinsic Laplacian

The *intrinsic Laplacian operator* Δ^* on X is the operator obtained by transforming Δ according to the change of coordinates $\mathbb{R}^n \to X$.

Definition: The Intrinsic Gaussian

The *intrinsic Gaussian* $G_v(x)$ for $(X, d\mu)$ is the fundamental solution to the heat equation for the intrinsic Laplacian.

Definition: The Intrinsic Normal Distribution

$\mathcal{N}^*(\alpha, v)$ is the distribution with density $f(x) = G_v(x[\alpha])$, where $[\cdot]$ represents the action of the group of Haar measure isometries of X.

Spherical polar coordinates are given by the following transformations:

$$x = r \sin\theta \cos\phi$$

$$y = r \sin\theta \sin\phi$$

$$z = r \cos\theta$$

The transformed Laplacian on the sphere is

$$\Delta^* = \frac{1}{\sin\theta} \left(\frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} \right) + \frac{1}{\sin\theta} \frac{\partial^2}{\partial\phi^2} \right)$$

The heat equation on the sphere is

$$\Delta^*(\theta, \phi, t) = \frac{\partial u}{\partial t}, \quad u(\theta, \phi, 0) = f(\theta), \quad t > 0,$$

where $f(\theta)$ is a given initial heat distribution.

Definition: Convolution on the sphere

$$(f * g)(x) = \int_{SO(3)} f(u)g(xu^{-1})\,du \qquad \text{for } x \text{ in } SO(3).$$

Definition: The Legendre Polynomials

The *Legendre Polynomials* are

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n$$

$$P_n^m(x) = (1-x^2)^{m/2}\frac{d^m P_n(x)}{dx^m}.$$

The Heat Equation on the Sphere

$$\Delta^*(\theta,\phi,t) = \frac{\partial u}{\partial t}, \quad u(\theta,\phi,0) = f(\theta), \quad t > 0,$$

where $f(\theta)$ is a given initial heat distribution.

Solution

$u = G_t * f$, where

$$G_t(\theta,\phi) = G_t(\theta) = \sum_{n>0} c_n P_n(\cos\theta) \exp[-n(n+1)t],$$

with constants c_n chosen so that $G_t \to \delta$ as $t \to 0^+$.

Definition: Poincare Upper Half-plane

The *Poincaré* upper half-plane is the space

$$H = \{x + iy \mid x, y \in \mathbb{R}, y > 0\}, \quad \text{where} \quad i = \sqrt{-1}.$$

- Note that H can be identified with the homogeneous (symmetric) space $SL(2, \mathbb{R})/SO(2)$.
- We can also identify H with the complex unit disc $U = \{z \in \mathbb{C} \mid |z| \leq 1\}$ via the Cayley transform $z \to (z - i)(z + i)^{-1}$.

Definition: Arc Length on H

The *arc length* on H is given by

$$ds^2 = y^{-2}(dx^2 + dy^2)$$

Note that this arc length measure is invariant under the action of $g \in SL(2, \mathbb{R})$ on $z \in \mathbb{H}$ defined by the fractional linear transformation:

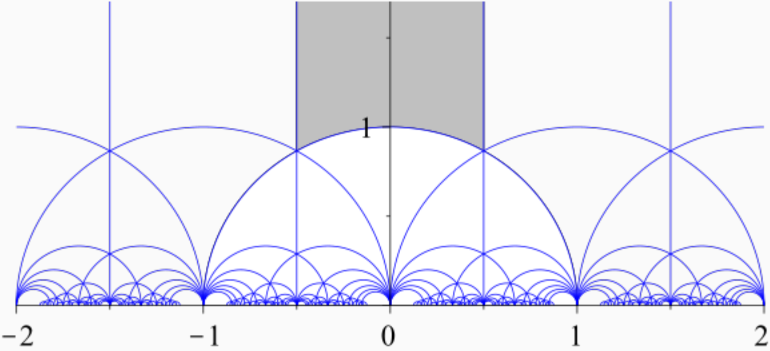
$$gz \equiv g(z) \equiv \frac{az + b}{cz + d},$$

for

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \text{with} \quad a, b, c, d \in \mathbb{R}, ad - bc = 1.$$

Geodesics in H

Geodesics in H are lines and semi-circles orthogonal to the real axis.



Geodesic distance is invariant under the action of $SL(2,\mathbb{R})$.

Definition: Laplacian on H

$$\Delta^*_{\mathbb{H}} \equiv \Delta^* = y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right).$$

Note that Δ^* is invariant under the action of $SL(2, \mathbb{R})$.

Definition: Geodesic polar coordinates on H

$$x = y \sinh r \sin(2u), \quad y = (\cosh r + \cos(2u) \sinh r)^{-1}$$

Definition: Heat Equation on H

$$\Delta^* u = \frac{\partial u}{\partial t} = y^2 \Delta u(x + iy, t), \quad u(z, 0) = \delta(z)$$

Solution

$$G_t(e^{-r} i) = \frac{\sqrt{2}}{(4\pi t)^{3/2} e^{t/4}} \int_r^\infty \frac{b e^{-b^2/4t} \, db}{\sqrt{\cosh b - \cosh r}}$$

Application to the Mantel Test

The **weighted Mantel test** for a sample $(X_i, Y_i) \in \mathbb{R}^{m\times n} \times \mathbb{R}^m$ and SPD weight matrix W has test statistic

$$T = \operatorname{tr}(YY^TX^TWX).$$

The effect of W on the test is through its eigenspace and the relative magnitudes of the eigenvalues. That is, any test T can be reduced to a unit determinant test $\tilde{T} = \operatorname{tr}(YY^TX^T\tilde{W}X)$, where $\tilde{W} = W/\sqrt[n]{|W|}$.

Application to the Mantel Test and Metric Learning

- The determinant one surface $\mathcal{SP}_n$ is the submanifold of $\mathcal{P}_n$ of matrices with determinant one, and can be identified with $\mathcal{SP}_n = SL(n, \mathbb{R})/SO(n)$. The $SL(n, \mathbb{R})$-invariant Riemannian metric is given by the restriction of the Riemannian metric on $\mathcal{P}_n$ to $\mathcal{SP}_n$.
- In particular, $H = SL(2, \mathbb{R})/SO(2)$.
- Since $\mathcal{N}_H^*$ is defined intrinsically over $\mathcal{SP}_n \cong H$, this places a distribution over all weighted Mantel tests of the form given by T, with mode at I, and equal density on equivalence classes defined by action of $SO(2)$.

Application to the Mantel Test and Metric Learning

Using the intrinsic Gaussian on H (or $SL(n,\mathbb{R})/SO(n)$ for $n > 2$), we can state a Bayesian model for the observed data

$$
\begin{aligned}
&Y \sim \mathcal{N}(0, \sigma^2 X^T W X + \sigma_\varepsilon^2 I_m) \\
&W \sim \mathcal{N}^*_{\mathcal{SP}_n}(I, t) \\
&\sigma^2 \sim Gamma(a, b) \\
&\sigma_\varepsilon^2 \sim Gamma(c, d),
\end{aligned}
$$

where values are chosen for t (prior variance of W) and a, b, c, d (according to prior beliefs about variance of Y and $h^2 = \sigma^2/(\sigma^2 + \sigma_\varepsilon^2)$).

The Helgason-Fourier Transform on $\mathcal{P}_n$

$$\mathcal{H}f(s,k) = \int_{Y \in \mathcal{P}_n} f(Y)\overline{p_s(Y[k])}d\mu_n(Y),$$

where $p_s(Y) = \prod_{j=1}^{n} |Y_j|^{s_j}$ for $s = (s_1, \ldots, s_n) \in \mathbb{C}^n$, and $Y[k] = k^T Y k$ is the $GL(n, \mathbb{R})$ group action.

Theorem. Properties of the Helgason-Fourier Transform

Suppose that $f : \mathcal{P}_n \to \mathbb{C}$ is infinitely differentiable with compact support.

(1) **Inversion Formula**

$$f(Y) = \omega_n \int_{s \in \mathbb{C}^n, \quad \mathrm{Re} s = -\rho} \mathcal{H}f(s,k) p_s(Y[k]) d\bar{k} |c_n(s)|^{-2} \, ds,$$

where

$$\rho = \left(\frac{1}{2}, \cdots, \frac{1}{2}, \frac{1}{4}(1-n)\right), \qquad \omega_n = \prod_{j=1}^{n} \frac{\Gamma(j/2)}{j(2\pi i)\pi^{j/2}},$$

$$c_n(s) = \prod_{1 \le i \le j \le n-1} \frac{B\left(\frac{1}{2}, s_i + \cdots + s_j + \frac{1}{2}(j-i+1)\right)}{B\left(\frac{1}{2}, \frac{1}{2}(j-i+1)\right)}.$$

Theorem. Properties of the Helgason-Fourier Transform

Suppose that $f : \mathcal{P}_n \to \mathbb{C}$ is infinitely differentiable with compact support.

(2) **Convolution Property**

If either f is a K-invariant ($K = O(n)$) function on $\mathcal{P}_n$, then for $g \in L^1(\mathcal{P}_n, d\mu_n)$,

$$\mathcal{H}(f * g) = \mathcal{H}f \cdot \mathcal{H}g,$$

where convolution is defined as

$$(f * g)(a) = \int_{GL(n,\mathbb{R})} f(b)g(ab^{-1})\, db.$$

Theorem. Properties of the Helgason-Fourier Transform

Suppose that $f : \mathcal{P}_n \to \mathbb{C}$ is infinitely differentiable with compact support.

(3) **G-invariant Differential Operators Changed to Multiplaction by a Polynomial**

If $L \in D(\mathcal{P}_n)$, then

$$\mathcal{H}(Lf)(s,k) = \overline{\lambda_{L^*}(s)}\mathcal{H}f(s,k), \quad \text{where } L_{p_s}(Y) = \lambda_L(s)p_s(Y).$$

Here L^* denotes the adjoint of L. Note that the eigenvalue $\lambda_{L^*}(s)$ is a polynomial in s.

Theorem. Properties of the Helgason-Fourier Transform

Suppose that $f : \mathcal{P}_n \to \mathbb{C}$ is infinitely differentiable with compact support.

(4) **Plancherel Theorem in K-invariant case**

Let $\alpha(s) = \omega_n |c_n(s)|^{-2}$. If f is K-invariant,

$$\int_{\mathcal{P}_n} |f(Y)|^2 \, d\mu_n(Y) = \int_{\mathrm{Re}s=-\rho} \left|\hat{f}(s)\right|^2 \alpha(s) \, ds.$$

Spherical Function on $\mathcal{P}_n$

$$h_s(Y) = \int_{k \in K} p_s(Y[k])\, dk, \quad \text{for } \; Y \in \mathcal{P}_n, s \in \mathbb{C}^n$$

The Heat Equation on $\mathcal{P}_n$

$$u_t - \Delta^* u = \delta(Y, t), \quad \Delta^* = \mathrm{tr}((Y\partial/\partial Y)^2).$$

Fundamental Solution to the Heat Equation on $\mathcal{P}_n$

$$G_t(Y) = \omega_n \int_{\mathrm{Re}s=-\rho} \exp[\lambda_2(s)t] h_s(Y) |c_n(s)|^{-2}\, ds.$$

- $h_s(Y)$ is the spherical function determined by s
- $\Delta p_s = \lambda_2(s) p_s$, with $\lambda_2^n(r) = r_1^2 + \cdots + r_n^2 + (n - n^3)/48$
- $s_j + \cdots + s_n = r_j + \frac{2j-n-1}{4}, j = 1, \ldots, n,$
- $r_j - r_{j+1} = s_j - \frac{1}{2}, j = 1, \ldots, n-1; s_n = r_n - \frac{n-1}{4}.$

Central Limit Theorems on Symmetric Spaces

Classical Central Limit Theorem

$$\int_{a\sqrt{n}}^{b\sqrt{n}} (f * \cdots * f)(x)\, dx \to \frac{1}{2\pi} \int_a^b \exp\{-x^2/2\}\, dx, \quad \text{as } \ n \to \infty$$

CLT on H

Suppose that $\{Z_n\}_{n\geq 1}$ is a sequence of iid $SO(2)$-invariant random variables in $\mathbb{H}$ with density $f(z)$. Let $S_n = Z_1 \circ \cdots \circ Z_n$ be normalized as $S_n^{\#}$. Then for measurable sets $A \subset H$ we have

$$\int_A f_n^{\#}(z)\, d\mu \to \exp(d/4) \int_A G_{d/4}(z)\, d\mu \quad \text{as} n \to \infty,$$

where G_c is the Gaussian on $\mathbb{H}$, and with $d = 2\pi \int_{r>0} f_Z(e^{-r} i) r^2 \sinh r\, dr$.

CLT on $\mathcal{P}_n$

Suppose that $\{Y_m\}_{m\geq 1}$ is a sequence of iid, $O(n)$-invariant random variables in $\mathcal{P}_n$ with density $f(Y)$ satisfying (1.157) and (1.158). Let S_m be the normalized form of $Y_1 \circ \cdots \circ Y_m$, with density f_m. For measurable sets $S \in \mathcal{P}_n$, as $m \to \infty$ we have

$$\int_S f_m(Y)\, d\mu(Y) \sim \exp\left(\frac{n^2 - n}{96(n+2)}\right) \int_S G_{3/(2(n+2))} * F_{1/(n+2)}(Y)\, d\mu(Y),$$

Central Limit Theorem for $\mathcal{P}_n$

In the preceding theorem, G_t is the fundamental solution of the heat equation on $\mathcal{P}_n$, with Helgason-Fourier transform

$$\widehat{G}_t(s(r)) = \exp\left\{ t\left(\sum_{i=1}^{n} r_i^2 - \frac{n^3 - n}{48} \right) \right\},$$

and F_t defined by its Helgason-Fourier transform

$$\widehat{F}_t(s(r)) = \exp\left\{ t\left(\sum_{1 \le i < j \le n} r_i r_j \right) \right\}.$$

Intrinsic Models for Manifold-Valued Data

Best Fit Geodesic Regression

Let $\gamma_M(t; T_M), t \in \mathbb{R}$ be the geodesic in $(X, d\mu)$ through $M \in X$ in the direction of $T_M \in \mathcal{T}_M X$. Consider a sample $(t_i, Y_i) \in (\mathbb{R} \times X)^n$, and assume the model

$$Y_i \sim \mathcal{N}^*(\gamma_M(t_i; T_M), v),$$

where $M \in X, T_M \in \mathcal{T}X$ are parameters determining the geodesic. The MLEs $\hat{M}, \hat{T}_M$ can (in theory) then be computed from the likelihood

$$\mathcal{L}(M, T_M | t, Y) = \prod_{i=1}^{n} f_i(Y_i \mid t_i, M, T_M),$$

where f_i is the density of Y_i.

Autoregressive Model for Time-series on $\mathcal{P}_n$

For an observed time series $Y_1, \ldots, Y_n$ on $\mathcal{P}_n$, an intrinsic first-order autoregressive model can be specified by

$$Y_t | Y_{t-1} \sim \mathcal{N}^*(g(Y_{t-1}), v), \quad t = 2, \cdots, n,$$

where $g : \mathcal{P}_n \to \mathcal{P}_n$. The likelihood can then be written as a product of conditional densities, and parameters in g estimated numerically.

Next Steps

- Work out computational methods for intrinsic Gaussian on $\mathcal{P}_n$, including optimization and sampling in order to run simulations.
- Derive properties of estimators from intrinsic models.
- Develop joint distributions of manifold-valued random variables.