Module overview

Information geometry is a relatively new mathematical field that seeks to apply the methods of differential geometry to statistical manifolds, i.e., manifolds of probability distributions. The field was mostly developed by S-I. Amari, and the monograph *Methods of Information Geometry* contains a concise if rather dense presentation of the theory.

- Differential geometry of curves
- Differential geometry of surfaces
- Differential geometry of manifolds
- Geodesic transport
- Statistical manifolds
- Fisher information metric

A *C^r curve* in $\mathbb{R}^n$ is a vector-valued function $\boldsymbol{\gamma}\colon [a,b] \to \mathbb{R}^n$, $\boldsymbol{\gamma} \in C^r[a,b]$, i.e., γ is r-times differentiable. For $t \in [a,b]$, $\boldsymbol{\gamma}(t)$ is a position vector in $\mathbb{R}^n$

The matrix-valued function, $\mathbf{B}\colon [a,b] \to \mathbb{R}$, defines a set of column vectors $\mathbf{B}(t) = \begin{pmatrix} \boldsymbol{\gamma}'(t) & \boldsymbol{\gamma}''(t) & \ldots & \boldsymbol{\gamma}^{(r)}(t) \end{pmatrix}$, and let $m = \min_{a \leqslant t \leqslant b} \operatorname{rank} \mathbf{B}(t)$. The curve is said to be m-regular.

The *length* of a curve is $l = \int_a^b \|\boldsymbol{\gamma}'(t)\| \, \mathrm{d}t$, usually using a 2-norm (Euclidean)

Orthonormalization (through Gram-Schmidt) of $\mathbf{B}(t) = \mathbf{Q}(t)\mathbf{R}(t)$ gives the Frenet reference frame $\mathbf{Q}_m(t) = \begin{pmatrix} \mathbf{q}_1(t) & \ldots & \mathbf{q}_m(t) \end{pmatrix}$.

In particular $\mathbf{q}_1(t) = \boldsymbol{\gamma}'(t) / \|\boldsymbol{\gamma}'(t)\|$ is the *unit tangent vector*, and

$$\mathbf{q}_j(t) = \mathbf{u}_j(t) / \|\mathbf{u}_j(t)\|, \mathbf{u}_j(t) = [\mathbf{I}_n - \mathbf{Q}_j(t)\mathbf{Q}_j^T(t)]\boldsymbol{\gamma}^{(j)}(t)$$

defines generalized curvature vectors, and $\kappa_i(t) = (\mathbf{q}_i'(t))^T \mathbf{q}_{i+1}(t) / \|\boldsymbol{\gamma}'(t)\|$ is the generalized curvature of order i.

It is convenient to introduce a change of variable $s(t)$ such that $\|\boldsymbol{\gamma}'(s)\| = 1$, called the *natural parametrization* of a curve, i.e., the parameter is the current arc length.

The vectors $\mathbf{Q}_3(s) = \begin{pmatrix} \mathbf{q}_1(s) & \mathbf{q}_2(s) & \mathbf{q}_3(s) \end{pmatrix}$ define the *Frenet triad*, usually denoted as $\begin{pmatrix} \boldsymbol{t}(s) & \boldsymbol{n}(s) & \boldsymbol{b}(s) \end{pmatrix}$, the *tangent*, *normal*, and *binormal* unit vectors. The first two generalized curvatures are usually denoted as

$$\kappa(s) = \kappa_1(s) = (\mathbf{q}_1'(t))^T \mathbf{q}_2(t), \tau(s) = (\mathbf{q}_2'(t))^T \mathbf{q}_3(t)$$

are known as the *curvature* and *torsion* of a curve in $\mathbb{R}^3$.

A curve can be reconstructed from knowledge of its generalized curvatures. In $\mathbb{R}^3$:

$$\begin{pmatrix} \mathbf{q}_1'(t) \\ \mathbf{q}_2'(t) \\ \mathbf{q}_3'(t) \end{pmatrix} = \|\boldsymbol{\gamma}'(t)\| \begin{pmatrix} 0 & \kappa(t) & 0 \\ -\kappa(t) & 0 & \tau(t) \\ 0 & -\tau(t) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{q}_1(t) \\ \mathbf{q}_2(t) \\ \mathbf{q}_3(t) \end{pmatrix}$$

Note that a curve in $\mathbb{R}^n$ is simply a univariate function. A C^r *surface* in $\mathbb{R}^n$ is a bivariate function $\boldsymbol{\sigma}: [a,b] \times [c,d] \to \mathbb{R}^n$, with $\boldsymbol{\sigma}(u,v)$ a position vector to a point on the surface.

A curve within the surface is a restriction $\gamma(t) = \boldsymbol{\sigma}(u(t), v(t)) = \boldsymbol{\sigma}(\boldsymbol{w}(t))$. The length traversed on the curve during step $\mathrm{d}t$ is $\|\gamma'(t)\|$, with

$$\gamma'(t) = \frac{\partial \boldsymbol{\sigma}}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}t} + \frac{\partial \boldsymbol{\sigma}}{\partial v}\frac{\mathrm{d}v}{\mathrm{d}t} = u'\boldsymbol{\sigma}_u + v'\boldsymbol{\sigma}_v$$

$$\|\gamma'(t)\|^2 = (\boldsymbol{w}')^T \begin{pmatrix} \boldsymbol{\sigma}_u^T\boldsymbol{\sigma}_u & \boldsymbol{\sigma}_u^T\boldsymbol{\sigma}_v \\ \boldsymbol{\sigma}_v^T\boldsymbol{\sigma}_u & \boldsymbol{\sigma}_v^T\boldsymbol{\sigma}_v \end{pmatrix} \boldsymbol{w}' = (\boldsymbol{w}')^T \begin{pmatrix} E & F \\ F & G \end{pmatrix} \boldsymbol{w}' = (\boldsymbol{w}')^T \mathbf{G}\, \boldsymbol{w}'$$

The matrix that arises is known as the metric tensor $\mathbf{G}$, and contains the first fundamental form of the surface $\mathrm{d}s^2 = E\,\mathrm{d}u^2 + 2F\,\mathrm{d}u\,\mathrm{d}v + G\,\mathrm{d}v^2$. The area element is $\mathrm{d}A = \|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|\mathrm{d}u\,\mathrm{d}v = \sqrt{EG - F^2}\,\mathrm{d}u\,\mathrm{d}v = \det(\mathbf{G})^{1/2}\,\mathrm{d}u\,\mathrm{d}v$.

First derivatives furnish tangent vectors. Second derivatives furnish information on surface curvature. In particular, the second fundamental form is

$$\mathbb{I} = L\,\mathrm{d}u^2 + 2M\,\mathrm{d}u\,\mathrm{d}v + N\,\mathrm{d}v^2$$

$$\begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}_{uu}^T \boldsymbol{n} & \boldsymbol{\sigma}_{uv}^T \boldsymbol{n} \\ \boldsymbol{\sigma}_{vu}^T \boldsymbol{n} & \boldsymbol{\sigma}_{vv}^T \boldsymbol{n} \end{pmatrix}$$

with

$$\boldsymbol{n} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|}$$

the unit normal vector to the surface element

Though linear spaces (e.g., the vector space $\mathbb{R}^n$) dominate most of mathematical approximation, it is recognized that more insightful models can sometimes be constructed on non-linear objects. Simplicial complexes were one such construct, piecewise linear, but not overall. Manifolds are another such concept that features a local structure similar to $\mathbb{R}^n$.

An n-dimensional *topological manifold* is a topological space that is locally homeomorphic to the n-dimensional Euclidean space $\mathbb{E}^n$ (e.g., $\mathbb{R}^n$).

Recall that topological spaces S, T are homeomorphic if there exists a map $f: S \to T$ that is continuous, and with continuous inverse. Two topological manifolds M, N are *diffeomorphic* if there exists $f: M \to N$ differentiable, and with a differentiable inverse.

A *differentiable manifold* is a topological manifold that has overlapping diffeomorphic neighborhoods around every point on the manifold.

An *affine connection* $\mathcal{A}$ on a differentiable manifold connects nearby tangent spaces.

Denote by $\mathcal{T}_M(P), \mathcal{T}_M(P')$ the tangent spaces at infinitesimally close points P, P' on manifold M of dimension n on which are defined coordinates $\boldsymbol{\xi} \in \mathbb{R}^n$. Let $\mathbf{B}^P = \left(\begin{array}{ccc} \boldsymbol{e}_1^P \equiv \partial_1 & \ldots & \boldsymbol{e}_n^P \equiv \partial_n \end{array} \right)$, $\partial_i \equiv \partial / \partial \xi_i$, be a basis for $\mathcal{T}_M(P)$, with $\mathbf{B}^{P'}$ a basis for $\mathcal{T}_M(Q)$. Define $\mathrm{d}\boldsymbol{\xi}_i = \boldsymbol{\xi}_i(P') - \boldsymbol{\xi}_i(P)$. Consider action of affine connection $\mathcal{A}_{P,P'}: \mathcal{T}_M(P) \to \mathcal{T}_M(P')$ on a basis vector $\boldsymbol{e}_j^P$

$$\mathcal{A}_{P,P'} \boldsymbol{e}_j^P = \boldsymbol{e}_j^{P'} - (\delta \boldsymbol{e})_j = \boldsymbol{e}_j^{P'} - \delta_j^k \boldsymbol{e}_k^{P'} = \boldsymbol{e}_i^{P'} - \delta_i^j \boldsymbol{e}_j^{P'}$$

The coefficients Γ_{ij}^k of the series expansion $\delta_j^k = \Gamma_{ij}^k \, \mathrm{d}\xi^i$ are the *connection coefficients* of the affine connection $\mathcal{A}$. These are also known as *Christoffel symbols of the second kind*.

Integration of infinitesimal connections along a curve $\boldsymbol{\xi}(t)$ on M allows correspondence between $\mathcal{T}_M(P), \mathcal{T}_M(Q)$, with P, Q at finite distance from one another.

The length of a curve on an n-dimensional manifold is

$$\mathrm{d}s^2 = g_{ij}\,\mathrm{d}\xi^i\,\mathrm{d}\xi^j$$

where $\mathbf{g} = (g_{ij})_{1 \leqslant i,j \leqslant n}$ is the *metric tensor*. Whereas in flat space the scalar product of vectors $\mathbf{u} = (u_i)$, $\mathbf{v} = (v_i)$, is $\mathbf{u} \cdot \mathbf{v} = u_i\, v_i$, on a manifold the scalar product becomes $\mathbf{u} \cdot \mathbf{v} = g_{ij} u^i\, v^i$.

The connection coefficients can also be written in terms of the metric tensor

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl} (g_{li,j} + g_{lj,i} - g_{ij,l}).$$

The geodesic equation on a manifold is

$$\ddot{\xi}^k + \Gamma^k_{ij}\, \dot{\xi}^i \dot{\xi}^j = 0$$

Fisher information metric

Of particular interest in data analysis is the Fisher information metric

$$g_{ij} = \mathbb{E}\left[\frac{\partial \log p(x;\xi)}{\partial \xi_i}\frac{\partial \log p(x;\xi)}{\partial \xi_j}\right] = \mathbb{E}\left[\frac{\partial^2 \log p(x;\xi)}{\partial \xi_i \, \partial \xi_j}\right]$$

For example, for univariate Gaussian distributions $p(x;\mu,\sigma) = (2\pi\sigma)^{-1/2}\mathrm{exp}[-(x-\mu)^2/(2\sigma^2)]$, $\log p(x;\mu,\sigma) = -\frac{1}{2}\log(2\pi\sigma) - (x-\mu)^2/(2\sigma^2)$

$$g_{\mu\mu} = -\frac{1}{\sigma^2}\int_{-\infty}^{+\infty} p(x;\mu,\sigma)\,\mathrm{d}x = -\sigma^{-3/2}$$