SUBMANIFOLDS OF STATISTICAL MANIFOLDS AND DATA ANALYSIS

Olena Chepurna, Yevhenia Kuleshova, Irena Hinterleitner

Communicated by Josef Mikeš

June 4 – 11, 2026

XXVI International Conference
Geometry, Integrability and Quantization

Varna, Bulgaria

Submanifolds of Statistical Manifolds

The present study is devoted to the geometry of curved exponential models, hierarchical models, and higher-order interactions from the viewpoint of information geometry.
Curved exponential families are treated as statistical submanifolds embedded in full exponential families, and their geometric properties are analysed via the induced Fisher metric and the family of α-connections introduced by Amari.
Second fundamental forms of such embeddings are computed and related to the statistical concept of model curvature. Hierarchical models for multivariate distributions are revisited as flat submanifolds in an ambient exponential family, with higher-order interaction parameters interpreted as components of the embedding curvature.
Connection properties of statistical submanifolds — including α-flatness, dually flat structures, and integrability conditions — are discussed within a unified geometric framework.

The Main New Results

The main new result is a local quadratic approximation of the KL divergence near a hierarchical model (Theorem 5.4) under a block-orthogonality condition on the Fisher metric, which identifies higher-order interaction parameters as components of a normal vector to the hierarchical submanifold in the ambient Fisher–Rao geometry.

As an explicit illustration of the general theory, the α-connection structure of Gaussian location families is analysed: an explicit verification shows that the Amari–Chentsov tensor vanishes identically for such families, causing all α-connections to collapse to the Levi-Civita connection (Proposition 7.5), with the consequence that the information-geometric curvature of the parabolic submanifold $\mu_2 = \mu_1^2$ is the same for all α (Proposition 7.6, Corollary 7.7).

A corrected and fully detailed version of the constant-variance example (Example 4.4) is also included. The results offer a coherent geometric interpretation of parameter estimation, model selection, and the structure of interaction terms.

The geometry of families of probability distributions has been a central theme in mathematical statistics since Rao [13] recognised that the Fisher information defines a Riemannian metric on a parametric model.
The systematic development of this idea into a full differential-geometric theory, centred on the notion of a statistical manifold equipped with a dualistic connection structure, is due principally to Amari [1] and was given its most transparent axiomatic formulation by Lauritzen [9].
Modern comprehensive treatments are provided by Amari [3] and by Ay, Jost, Lê Hông Vân, and Schwachhöfer [4].

Within this framework, a parametric family of probability distributions $\{p(x;\theta) \colon \theta \in \Theta \subset \mathbb{R}^n\}$ is treated as an n-dimensional manifold M, the Fisher information tensor serves as a Riemannian metric g, and a one-parameter family of torsion-free affine connections $\{\nabla^{(\alpha)}\}_{\alpha\in\mathbb{R}}$ — the α-connections of Amari and Chentsov [2] — encodes the higher-order geometry of the model.

A submodel, that is, a subfamily $N \subset M$ parametrised by fewer parameters, inherits from M an induced metric and induced connections; it thus becomes a statistical submanifold in the sense of Opozda [12], Vos [14], and Matsuzoe [10]. The geometry of this embedding—captured by the second fundamental form and the normal curvature—carries direct statistical meaning: it controls the bias of maximum-likelihood estimates, the Cramér–Rao bound on constrained models, and the information loss incurred by restricting attention to the submodel.

Two particularly important classes of submanifolds arise naturally in data analysis:

(i) *Curved exponential families*, introduced by Efron [7] and studied geometrically by Amari, which are low-dimensional smooth submanifolds of a full exponential family and are generically curved with respect to the mixture (-1)-connection.

(ii) *Hierarchical models* for multivariate discrete distributions, which are defined by setting interaction parameters of order $k+1$ and higher to zero; these submanifolds turn out to be e-flat (geodesically complete with respect to the exponential connection) and carry a natural product structure.

The present paper brings together these two threads and continues the line of investigation initiated in [6, 5].
Section 2 recalls the algebraic and geometric structure of a statistical manifold.
Section 3 develops the theory of statistical submanifolds and their curvature properties.
Sections 4 and 5 treat curved exponential families and hierarchical models, respectively.
Section 6 draws the consequences for data analysis.
Section 7 works through an explicit example, including new results on α-connection collapse and asymptotic estimation.
Section 8 collects conclusions.

Fisher Information Metric

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and
let $M = \{p(\cdot;\theta) \colon \theta \in \Theta\}$ be a smooth n-dimensional family of probability densities on Ω, with $\Theta \subset \mathbb{R}^n$ open and convex.
We assume standard regularity: each $p(\cdot;\theta)$ is positive μ-almost everywhere, smooth in θ, and all integrals may be differentiated under the sign of integration.

Definition

The *score function* in the direction $\partial_i = \partial/\partial\theta^i$ is

$$\ell_i(\theta) := \partial_i \log p(x;\theta), \qquad i = 1, \ldots, n.$$

The *Fisher information matrix* (Fisher–Rao metric) is the $n \times n$ symmetric positive-definite tensor

$$g_{ij}(\theta) = E_\theta[\ell_i \ell_j] = \int_\Omega (\partial_i \log p)(\partial_j \log p)\, p \,\mathrm{d}\mu.$$

Since $E_\theta[\ell_i] = 0$, one has the identity $g_{ij} = -E_\theta[\partial_i \partial_j \log p]$. The metric g defines a Riemannian structure on M; by Chentsov's theorem it is, up to a positive scalar multiple, the unique Riemannian metric on the simplex of finite probability measures that is monotone under Markov morphisms [2].

Amari–Chentsov Tensor

Beyond the metric, the statistical structure of M is further characterised by a completely symmetric cubic tensor.

Definition

The *Amari–Chentsov tensor* (or skewness tensor) is

$$T_{ijk}(\theta) = E_\theta[\ell_i \ell_j \ell_k] = \int_\Omega (\partial_i \log p)(\partial_j \log p)(\partial_k \log p)\, p \,\mathrm{d}\mu.$$

This tensor is the third central moment of the score vector and plays the role of the Christoffel difference between the two fundamental connections.

α-Connections

Definition

For $\alpha \in \mathbb{R}$, the *α-connection* $\nabla^{(\alpha)}$ is the torsion-free affine connection on M whose Christoffel symbols in the coordinate system (θ^i) are

$$\Gamma^{(\alpha)}_{ij,k}(\theta) := E_\theta\left[\left(\partial_i\partial_j \log p + \tfrac{1-\alpha}{2}\,\partial_i \log p \cdot \partial_j \log p\right)\partial_k \log p\right].$$

Equivalently,

$$\Gamma^{(\alpha)}_{ij,k} = \Gamma^{(m)}_{ij,k} + \tfrac{1+\alpha}{2}\,T_{ijk} = \Gamma^{(e)}_{ij,k} - \tfrac{1-\alpha}{2}\,T_{ijk},$$

where $\Gamma^{(m)}$ and $\Gamma^{(e)}$ denote the mixture ($\alpha = -1$) and exponential ($\alpha = 1$) connection Christoffel symbols, respectively, and $T_{ijk} = \Gamma^{(e)}_{ij,k} - \Gamma^{(m)}_{ij,k}$ [2]. The Levi-Civita connection of g corresponds to $\alpha = 0$.

Dual Connections

Definition

Two connections ∇ and ∇^* on a Riemannian manifold (M, g) are *dual* (or conjugate) with respect to g if, for all vector fields X, Y, Z,

$$Z\, g(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y).$$

The α- and $(-\alpha)$-connections are dual: $(\nabla^{(\alpha)})^* = \nabla^{(-\alpha)}$.
The triple $(g, \nabla^{(1)}, \nabla^{(-1)})$ – a *dualistic structure* – constitutes the central object of information geometry.

Proposition

The covariant curvature tensors $R^{(\alpha)}_{ijkl}$ and $R^{(-\alpha)}_{ijkl}$ of the dual connections satisfy $R^{(\alpha)}_{ijkl} + R^{(-\alpha)}_{ijlk} = 0$.

In particular, M is α-flat if and only if it is $(-\alpha)$-flat.

Induced Metric and Connections

Let $\iota : N \hookrightarrow M$ be a smooth immersion of an m-dimensional manifold N ($m < n$) into the statistical manifold M.

Definition

The *induced Fisher metric* on N is the pullback $h = \iota^* g$, given in local coordinates (ξ^μ) on N by

$$h_{\mu\nu}(\xi) = g_{ij}(\theta(\xi)) \frac{\partial \theta^i}{\partial \xi^\mu} \frac{\partial \theta^j}{\partial \xi^\nu}.$$

The *induced α-connection* $\hat{\nabla}^{(\alpha)}$ on N is defined by projecting the ambient α-covariant derivative onto the tangent bundle of N:

$$\hat{\nabla}^{(\alpha)}_X Y = \tan\big(\nabla^{(\alpha)}_{\tilde{X}} \tilde{Y}\big)\big|_N,$$

where $\tilde{X}, \tilde{Y}$ are local extensions of $X, Y \in TN$ to vector fields on M.

Second Fundamental Form

Definition

The α-*second fundamental form* of the embedding $\iota : N \hookrightarrow M$ is the normal-valued symmetric bilinear form

$$B^{(\alpha)}(X, Y) = \nabla^{(\alpha)}_{\tilde{X}} \tilde{Y} - \hat{\nabla}^{(\alpha)}_{X} Y, \qquad X, Y \in TN,$$

measuring the deviation of the ambient connection from the induced one.

The *Gauss equation* for statistical connections (Voss, Opozda [14, 12]) expresses the induced curvature in terms of the ambient curvature and the second fundamental forms. For $X, Y, Z, W \in TN$,

$$\begin{aligned} g\big(\hat{R}^{(\alpha)}(X, Y)Z, W\big) &= g\big(R^{(\alpha)}(\tilde{X}, \tilde{Y})\tilde{Z}, \tilde{W}\big) + \\ g\big(B^{(-\alpha)}(Y, Z), A^{(\alpha)}_{W} X\big) &- g\big(B^{(-\alpha)}(X, Z), A^{(\alpha)}_{W} Y\big), \end{aligned} \tag{1}$$

where $A^{(\alpha)}_{W} X := -\tan(\nabla^{(\alpha)}_{\tilde{X}} \tilde{W})$ is the shape operator,
and $R^{(\alpha)}$, $\hat{R}^{(\alpha)}$ are the curvature tensors of the ambient and induced α-connections, respectively.

Proposition

(Fundamental equation for statistical submanifolds.)
The α-second fundamental form and the $(-\alpha)$-shape operator of $N \subset M$ satisfy

$$g\big(B^{(\alpha)}(X, Y), \xi\big) = g\big(A_{\xi}^{(-\alpha)} X, Y\big)$$

for all $X, Y \in TN$ and $\xi \in T^{\perp}N$.

Proof.

Since $Y \in TN$ and $\xi \in T^{\perp}N$, we have $g(\tilde{Y}, \tilde{\xi}) = 0$ on N. Differentiating along $\tilde{X}$ and applying the duality relation gives

$$g(\nabla^{(\alpha)}_{\tilde{X}} \tilde{Y}, \tilde{\xi}) + g(\tilde{Y}, \nabla^{(-\alpha)}_{\tilde{X}} \tilde{\xi}) = 0.$$

Decomposing each covariant derivative into tangential and normal parts and using $g(\hat{\nabla}^{(\alpha)}_{X} Y, \xi) = 0$ yields the result; see Voss [14, p. 432]. $\square$

Curvature Properties

Definition

A submanifold $N \subset M$ is called *α-flat* if the induced connection $\hat{\nabla}^{(\alpha)}$ has vanishing curvature. It is called *auto-parallel* with respect to $\nabla^{(\alpha)}$ if $B^{(\alpha)} \equiv 0$, i.e. if $\nabla^{(\alpha)}_{\tilde{X}} \tilde{Y} \in TN$ for all $X, Y \in TN$.

Since the α-connections of Definition 3 are torsion-free, the condition $B^{(\alpha)} \equiv 0$ is the natural statistical submanifold notion of auto-parallelism throughout this paper; it coincides with total geodesy in the Riemannian sense when $\alpha = 0$, i.e. for the Levi-Civita connection of g (see Nomizu, Sasaki[11]).
In particular, auto-parallelism with respect to the e-connection ($\alpha = 1$) means that every e-geodesic in N is also an e-geodesic in M.

Geometric Formulation of Exponential Families

Definition

An *exponential family* with natural parameters $\theta = (\theta^1, \dots, \theta^n)$ and sufficient statistic $t(x) = (t_1(x), \dots, t_n(x))$ is the manifold

$$\mathcal{E} = \left\{ p(x;\theta) = \exp\left(\theta^i t_i(x) - \psi(\theta)\right) h(x) \colon \theta \in \Theta \right\},$$

where $\psi(\theta) = \log \int \exp(\theta^i t_i) h \, \mathrm{d}\mu$ is the log-partition function.

Exponential families are *e*-flat: the Christoffel symbols of $\nabla^{(1)}$ vanish in natural coordinates θ. They are simultaneously *m*-flat in the dual expectation coordinates $\eta_i = E_\theta[t_i] = \partial_i \psi$.

Curved Exponential Families as Submanifolds

Definition ([7])

A *curved exponential family* is a smooth m-dimensional submanifold $N \subset \mathcal{E}$ defined by a smooth embedding $\theta : \Xi \to \Theta$, $\xi \mapsto \theta(\xi)$, where $\Xi \subset \mathbb{R}^m$, $m < n$. The family is *curved* if the induced (-1)-connection $\hat{\nabla}^{(-1)}$ has nonzero curvature, i.e. $B^{(-1)} \not\equiv 0$.

Proposition

Let $N \subset \mathcal{E}$ be a curved exponential family with embedding $\theta(\xi)$. Then:
(i) *the Fisher metric on N is $h_{\mu\nu}(\xi) = g_{ij}(\theta(\xi))\,\partial_\mu\theta^i\,\partial_\nu\theta^j$, where $g_{ij}{=}\partial_i\partial_j\psi$;*
(ii) *the (1)-second fundamental form is $B^{(1)}(\partial_\mu, \partial_\nu) = \sum_i(\partial_\mu\partial_\nu\theta^i)\,\mathrm{nor}(\partial_i)$;*
(iii) *N is e-auto-parallel (i.e. $B^{(1)} \equiv 0$) if and only if the embedding $\xi \mapsto \theta(\xi)$ is affine; in this case N inherits e-flatness from the ambient e-flat structure of $\mathcal{E}$ [3, Amari, Chap. 3].*

Proof.

Part (i) follows by direct pullback.

For (ii), since e-connection symbols vanish on $\mathcal{E}$, one has $\nabla^{(1)}_{\partial_\mu}\partial_\nu = 0$ in $\mathcal{E}$; hence $B^{(1)}(\partial_\mu, \partial_\nu) = \sum_i (\partial_\mu \partial_\nu \theta^i)\,\mathrm{nor}(\partial_i)$ by the Gauss decomposition (see Voss [14]).

Part (iii): $B^{(1)} = 0$ if and only if $\partial_\mu \partial_\nu \theta^i = 0$ for all μ, ν, i, i.e. the map is affine. An affine embedding maps N onto a coordinate affine subspace of Θ in natural coordinates; such a subspace is a coordinate submanifold of the e-flat manifold $\mathcal{E}$ and therefore e-flat. □

Examples

Example (Normal family with constrained variance)

Consider the full normal exponential family $\mathcal{E}$ on $\mathbb{R}$ with natural parameters $\theta^1 = \mu/\sigma^2$, $\theta^2 = -1/(2\sigma^2)$, sufficient statistics $t_1(x) = x$, $t_2(x) = x^2$, and log-partition function

$$\psi(\theta) = -(\theta^1)^2/(4\theta^2) + \tfrac{1}{2}\log(-\pi/\theta^2).$$

The Fisher metric in (μ, σ) coordinates is $g = \mathrm{diag}(1/\sigma^2, 2/\sigma^2)$, and in natural coordinates at (μ, σ):

$$G(\mu,\sigma) = \begin{pmatrix} \sigma^2 & \mu\sigma^2 \\ \mu\sigma^2 & \mu^2\sigma^2 + 2\sigma^4 \end{pmatrix}, \qquad \det G = 2\sigma^6. \tag{2}$$

The Amari–Chentsov tensor has non-vanishing components $T_{112} = 2\sigma^4$ and $T_{222} = 8\sigma^6$ (all other independent components vanish by odd-moment symmetry of the Gaussian).

Sub-case $\gamma = 0$: constant-variance family $N_0 = \{N(\mu, c^2) \colon \mu \in \mathbb{R}\}$.
The subfamily, parametrised by $\xi = \mu$, maps to $(\xi/c^2, -1/(2c^2))$ in natural coordinates – an affine embedding. By Proposition 3(iii), N_0 is *e*-flat: $B^{(1)} \equiv 0$. The induced Fisher metric is $h(\xi) = 1/c^2$ (constant). For the (-1)-second fundamental form, using $\Gamma^{(-1)}_{ij,k} = -T_{ijk}$, with $\dot\theta = (1/c^2, 0)$ and $\ddot\theta = (0,0)$, the *m*-covariant acceleration has lowered components

$$[\nabla^{(-1)}_{\dot\theta}\dot\theta]_k = \Gamma^{(-1)}_{11,k}(1/c^2)^2 = \begin{cases} 0 & k = 1, \\ -T_{112}/c^4 = -2 & k = 2. \end{cases}$$

Raising with G^{-1} and projecting onto the normal bundle of N_0 gives

$$B^{(-1)}(\partial_\xi, \partial_\xi) = (\xi/c^4, -1/c^4),$$

which satisfies $g(B^{(-1)}, \dot\theta) = 0$ (normality check).
The squared g-norm is $\|B^{(-1)}\|^2 = 2/c^4$ (constant, independent of ξ).

The geodesic curvature and Efron statistical curvature of N_0 are therefore

$$\kappa_0 = \frac{\|B^{(-1)}\|}{h(\xi)} = \frac{\sqrt{2}/c^2}{1/c^2} = \sqrt{2}, \qquad \gamma_0 = \frac{\|B^{(-1)}\|^2}{h(\xi)} = \frac{2}{c^2}. \tag{3}$$

Thus N_0 is *e*-flat but *m*-curved, with geodesic curvature $\kappa_0 = \sqrt{2}$ independent of c and statistical curvature $\gamma_0 = 2/c^2$ inversely proportional to the variance. The MLE bias on N_0 is $O(n^{-1})$ with coefficient proportional to γ_0.

Remark. The subfamily with $\gamma = 0$ (constant variance) is *e*-flat but not *m*-flat; the two notions are distinct.

Section 4. Hierarchical Models and Higher-Order Interactions

Geometric Representation

Let $\mathrm{X} = (X_1, \dots, X_k)$ be a random vector taking values in a finite set $\mathcal{X} = \prod_{i=1}^{k} \mathcal{X}_i$. The set of all strictly positive probability distributions on $\mathcal{X}$ forms an exponential family $\mathcal{E}$ with natural parameters indexed by interaction terms. For any subset $A \subseteq [k]$, define the A-interaction sufficient statistic

$$t_A(x) := \prod_{i \in A} t_i^{(x_i)},$$

where $t_i^{(a)}$ are centred indicator functions satisfying $\sum_a t_i^{(a)} = 0$. Every distribution in $\mathcal{E}$ can be written $\log p(x;\theta) = \sum_{\emptyset \neq A \subseteq [k]} \theta^A t_A(x)$.

Definition

The *hierarchical model of order r* is the exponential submanifold

$$H_r := \{p \in \mathcal{E} \colon \theta^A = 0 \text{ for all } |A| > r\}.$$

Proposition

H_r is an e-flat submanifold of $\mathcal{E}$; in particular, it is both auto-parallel with respect to $\nabla^{(1)}$ and metrically complete.

Proof.

In natural coordinates θ, H_r is a coordinate subspace of Θ; coordinate subspaces of an e-flat manifold inherit e-flatness. Since the coordinate slice $\{\theta^A = 0, |A| > r\}$ is defined by linear equalities in the natural coordinates, the e-geodesics of $\mathcal{E}$ through points of H_r that remain in H_r are simply lines in the coordinate space, which are complete. Auto-parallelism follows from $\partial_\mu \partial_\nu \theta^i = 0$ for an affine embedding (Proposition 3(iii)). □

Remark

The m-projection of $p \in \mathcal{E}$ onto H_r yields the distribution in H_r whose moments $E_p[t_A]$, $|A| \leq r$, match those of p; it is thus the maximum-entropy distribution consistent with the low-order interaction structure. For $r = 1$ this coincides with the product distribution having the same marginals. This is the information-geometric interpretation of hierarchical modelling.

Statistical Interpretation of Higher-Order Terms

The parameters θ^A for $|A| = r + 1$ measure the discrepancy between $p \in \mathcal{E}$ and its *m*-projection onto H_r; they are precisely the $(r+1)$-th order interaction contrasts of classical log-linear analysis. The magnitude of this discrepancy is captured by the KL divergence

$$D_{\mathrm{KL}}(p \,\|\, \pi_r(p)) = \psi(\theta^{H_r}) - \psi(\theta) + (\theta - \theta^{H_r})^\top \eta, \tag{4}$$

where θ^{H_r} denotes the natural parameters of $\pi_r(p)$ and $\eta = \nabla\psi(\theta)$ the expectation parameters.

Local KL Divergence Near H_r and Interaction Geometry

The formula (4) is exact but global. The following theorem provides a local, geometrically transparent version in terms of the ambient Fisher metric.

Theorem (Local KL expansion near H_r)

Let $p_0 \in H_r$ and suppose that at p_0 the parameter blocks $\{\theta^A \colon |A| \le r\}$ and $\{\theta^A \colon |A| > r\}$ are g-orthogonal: $g_{AB}(p_0) = \mathrm{Cov}_{p_0}(t_A, t_B) = 0$ for all $|A| \le r < |B|$. Write $\varepsilon = (\theta^A)_{|A|>r}$ for the higher-order interaction parameters of p near p_0. Then

$$D_{\mathrm{KL}}(p \,\|\, \pi_r(p)) = \frac{1}{2} \sum_{|A|,|B|>r} g_{AB}(p_0)\, \theta^A \theta^B + O(|\varepsilon|^3), \tag{5}$$

where $g_{AB}(p_0) = \partial_A \partial_B \psi|_{p_0} = \mathrm{Cov}_{p_0}(t_A, t_B)$ is the restriction of the Fisher metric of $\mathcal{E}$ to the interaction directions of order greater than r.

Proof.

Set $F(\varepsilon) = D_{\mathrm{KL}}(p \,\|\, \pi_r(p))$ for p with parameters $(\theta^{\le r}(p_0), \varepsilon)$, where $\theta^{\le r}(p_0)$ are fixed at the base point. Since $\pi_r(p_0) = p_0$, we have $F(0) = 0$. The map $\varepsilon \mapsto F(\varepsilon)$ is C^∞ in a neighbourhood of 0 by smoothness of ψ and of the projection π_r. By the variational characterisation of m-projection, $\nabla_\varepsilon F|_{\varepsilon=0} = 0$.

To determine the Hessian we use the *envelope theorem* for parametric minima. Write $F(\varepsilon) = \min_{q \in H_r} D_{\mathrm{KL}}(p(\varepsilon) \,\|\, q)$. Because $\pi_r(p(\varepsilon))$ realises the minimum for each ε, the first-order optimality condition gives $\partial D_{\mathrm{KL}}(p\|q)/\partial q\big|_{q=\pi_r(p)} = 0$. The envelope theorem for smooth parametric minima therefore states:

$$\frac{\partial^2 F}{\partial \varepsilon^A \partial \varepsilon^B}\bigg|_0 = \frac{\partial^2 D_{\mathrm{KL}}(p(\varepsilon) \,\|\, q)}{\partial \varepsilon^A \partial \varepsilon^B}\bigg|_{\varepsilon=0,\, q=p_0},$$

where differentiation acts only on the p-argument (the minimiser is held fixed at p_0 when evaluating the Hessian in ε). □

Since $D_{\mathrm{KL}}(p\|q) = \psi(\theta_q) - \psi(\theta_p) - (\theta_q - \theta_p)^\top \eta(\theta_p)$ on the exponential family [2, Amari, Nagaoka, Prop. 1.5], and only the $> r$ components of θ_p vary with ε (the $\leq r$ components are fixed at $\theta_0^{\leq r}$), the partial second derivatives in the ε^A, ε^B directions at $\varepsilon = 0$ (with $\theta_q = \theta_p = \theta_0$) are:

$$\left.\frac{\partial^2 D_{\mathrm{KL}}(p\|q)}{\partial\varepsilon^A \partial\varepsilon^B}\right|_0 = \left.\frac{\partial^2 \psi(\theta_p)}{\partial\theta^A \partial\theta^B}\right|_{\theta_0} = g_{AB}(p_0), \quad |A|, |B| > r.$$

(The block-orthogonality hypothesis $g_{AB}(p_0) = 0$ for $|A| \leq r < |B|$ ensures that perturbations in the ε directions are normal to H_r in the Fisher metric, so no cross-term between the $\leq r$ and $> r$ parameter blocks contributes.) Taylor's theorem then gives (5), where $|\varepsilon|$ denotes any fixed norm on the finite-dimensional parameter space $(\theta^A)_{|A|>r}$.

Remark

The leading term in (5) equals half the squared Fisher–Rao normal distance from p to H_r: $D_{\mathrm{KL}}(p \,\|\, \pi_r(p)) \approx \frac{1}{2}\, d_g^{\perp}(p, H_r)^2 + O(|\varepsilon|^3)$. Consequently the interaction parameters θ^A with $|A| = r + 1$ are, to leading order, the components of a normal vector from H_r to p in the ambient statistical manifold, connecting the embedding geometry to classical log-linear model selection. The block-orthogonality condition is satisfied, for instance, when p_0 is a product distribution ($r = 1$), since in that case $\mathrm{Cov}_{p_0}(t_A, t_B) = 0$ for $|A| = 1$ and $|B| > 1$.

Geometric Interpretation of Parameter Estimation

Maximum-likelihood estimation on a curved exponential family $N \subset \mathcal{E}$ is geometrically an m-projection: the MLE $\hat{\xi}$ is the parameter value minimising the KL divergence from the empirical distribution $\hat{p}$ to the model, $\hat{\xi} = \arg\min_{\xi} D_{\mathrm{KL}}(\hat{p} \,\|\, p(\cdot;\theta(\xi)))$. The asymptotic bias of the MLE, at order n^{-1}, is controlled by $B^{(-1)}$: Amari [1] and Kass–Vos [8] show that $\mathrm{Bias}(\hat{\xi})$ is $O(n^{-1})$ and is expressible as a contraction of $B^{(-1)}$ with the m-connection Christoffel symbols and the inverse induced metric h^{-1}. The bias vanishes to this order if and only if $B^{(-1)} \equiv 0$, i.e. when N is m-auto-parallel.

Model Complexity

The Fisher information volume element $\sqrt{\det h(\xi)}\,\mathrm{d}\xi$ of the submanifold is a natural measure of model complexity, appearing in the MDL criterion and in the Jeffreys prior. The geometric complexity is larger than that of an e-flat model of the same dimension precisely because of the curvature of the embedding.

Information-Theoretic Viewpoint

The *Pythagorean theorem* for KL divergences on dually flat manifolds [2] states: if $N \subset \mathcal{E}$ is e-flat and $p \in \mathcal{E}$, then for any $q \in N$,

$$D_{\mathrm{KL}}(p \,\|\, q) = D_{\mathrm{KL}}(p \,\|\, \pi(p)) + D_{\mathrm{KL}}(\pi(p) \,\|\, q),$$

where $\pi(p)$ is the e-projection of p onto N. This decomposition underlies the geometry of iterative proportional fitting, expectation-maximisation algorithms, and hierarchical independence testing.

Bivariate Gaussian Exponential Family

Let $\mathcal{E}$ be the exponential family of bivariate Gaussian distributions on $\mathbb{R}^2$ with mean $\mu \in \mathbb{R}^2$ and covariance $\Sigma = I_2$. The log-density is

$$\log p(x;\mu) = \mu_1 x_1 + \mu_2 x_2 - \tfrac{1}{2}x_1^2 - \tfrac{1}{2}x_2^2 - \tfrac{1}{2}(\mu_1^2 + \mu_2^2) - \log(2\pi).$$

The natural parameters are $\theta^1 = \mu_1$, $\theta^2 = \mu_2$, $\theta^{11} = \theta^{22} = -\frac{1}{2}$, $\theta^{12} = 0$ ($n = 5$). The log-partition function $\psi(\theta^1, \theta^2) = \frac{1}{2}[(\theta^1)^2 + (\theta^2)^2] + \log(2\pi)$ has Hessian $g_{ij} = \delta_{ij}$, so the Fisher metric is the identity matrix in mean coordinates μ.

A Curved Submanifold

Consider the one-parameter subfamily $N \subset \mathcal{E}$ defined by $\mu_1(\xi) = \xi$, $\mu_2(\xi) = \xi^2$. The embedding in natural coordinates is the parabolic arc $\xi \mapsto (\theta^1, \theta^2, \theta^{11}, \theta^{12}, \theta^{22}) = (\xi, \xi^2, -\frac{1}{2}, 0, -\frac{1}{2})$.

Fisher Metric Calculation

Since $g_{ij} = \delta_{ij}$ in the (θ^1, θ^2) directions, pulling back to N via $\dot{\theta} = (\dot{\theta}^1, \dot{\theta}^2) = (1, 2\xi)$:

$$h(\xi) = \left(\frac{\mathrm{d}\theta^1}{\mathrm{d}\xi}\right)^2 + \left(\frac{\mathrm{d}\theta^2}{\mathrm{d}\xi}\right)^2 = 1 + 4\xi^2. \tag{6}$$

This is the arc-length element of the parabola (ξ, ξ^2) in $\mathbb{R}^2$.

Second Fundamental Form and Curvature

The velocity and acceleration vectors of the embedding are $\dot{\theta}(\xi) = (1, 2\xi, 0, 0, 0)$ and $\ddot{\theta}(\xi) = (0, 2, 0, 0, 0)$. Their g-inner product is $\dot{\theta} \cdot \ddot{\theta} = 4\xi$, and $|\dot{\theta}|^2 = 1 + 4\xi^2$. The normal component of $\ddot{\theta}$ is $(0, 2, 0, 0, 0) - \frac{4\xi}{1+4\xi^2}(1, 2\xi, 0, 0, 0)$, with squared norm $\frac{4}{1+4\xi^2}$. Following Amari [1], we define the *geodesic curvature* of N at ξ as $\kappa(\xi) = \|B^{(-1)}(\partial_\xi, \partial_\xi)\|/h(\xi)$, where $h(\xi) = |\dot{\theta}|^2$ is the metric coefficient; this is the information-geometric curvature in the sense of [1, 7] and differs from the standard Riemannian geodesic curvature by a factor of $h(\xi)^{1/2}$. Direct computation gives

$$\kappa(\xi) = \frac{|\ddot{\theta} - (\dot{\theta} \cdot \ddot{\theta}/|\dot{\theta}|^2)\dot{\theta}|}{|\dot{\theta}|^2} = \frac{\sqrt{4/(1+4\xi^2)}}{1+4\xi^2} = \frac{2}{(1+4\xi^2)^{3/2}}, \tag{7}$$

which is maximal at $\xi = 0$ and decays to zero as $|\xi| \to \infty$. This curvature coincides with the norm of $B^{(-1)}$ (proved in Proposition 6 below): the MLE on N has an $O(n^{-1})$ bias proportional to κ.

For hierarchical models of bivariate Gaussians the interaction parameter is θ^{12}, measuring the strength of linear dependence. Setting $\theta^{12} = 0$ defines the product-Gaussian submanifold H_1, which is e-flat (Proposition 4), so the MLE on H_1 is asymptotically unbiased.

α-Connection Collapse and Self-Duality of Second Fundamental Forms

The claim in Section 7.4 that $\kappa(\xi) = \|B^{(-1)}\|/h$ requires a verification that $B^{(-1)} = B^{(1)}$. The following proposition establishes this for all Gaussian location families and for all α. The key observation is that for such families the Amari–Chentsov tensor T vanishes identically—a direct consequence of the vanishing of all odd-order moments of a centred Gaussian vector—so that the formula $\Gamma^{(\alpha)} = \Gamma^{(e)} - (1-\alpha)/2 \cdot T$ reduces to $\Gamma^{(\alpha)} = 0$ for all α. This structural collapse of the α-connection family to a single Riemannian connection distinguishes Gaussian location families from all exponential families with nonzero third cumulants.

Proposition (α-Connection collapse for Gaussian location families)

Let $\mathcal{E} = \{N(\mu, \Sigma) \colon \mu \in \mathbb{R}^n\}$ be the exponential family of n-dimensional Gaussian distributions with fixed positive-definite covariance Σ. Then:

(i) *The Amari–Chentsov tensor T vanishes identically: $T_{ijk}(\theta) = 0$ for all i, j, k.*

(ii) *For every $\alpha \in \mathbb{R}$, the α-connection coincides with the Levi-Civita connection of g: $\nabla^{(\alpha)} = \nabla^{(0)}$.*

(iii) *For every smooth immersed submanifold $N \subset \mathcal{E}$, the α-second fundamental form is independent of α: $B^{(\alpha)} = B^{(0)}$ for all $\alpha \in \mathbb{R}$.*

(iv) *N is e-auto-parallel in $\mathcal{E}$ if and only if it is m-auto-parallel, if and only if $B^{(\alpha)} \equiv 0$ for all $\alpha \in \mathbb{R}$.*

Proof.

(i) The natural sufficient statistics are linear: $t_i(x) = (\Sigma^{-1}x)_i$, so the score functions $\ell_i = t_i(x) - \eta_i$ form a centred Gaussian vector with covariance Σ^{-1}. All odd-order moments of a centred Gaussian vanish, so $T_{ijk} = E[\ell_i \ell_j \ell_k] = 0$ for all i, j, k.
(ii) By Definition 3: $\Gamma^{(\alpha)}_{ij,k} = \Gamma^{(1)}_{ij,k} - \frac{1-\alpha}{2} T_{ijk} = 0 - 0 = 0$, using e-flatness ($\Gamma^{(1)} = 0$ in natural coordinates) and $T \equiv 0$ from (i). Since $g_{ij} = (\Sigma^{-1})_{ij}$ is constant, $\nabla^{(0)}$ also has vanishing Christoffel symbols, so $\nabla^{(\alpha)} = \nabla^{(0)}$ for all α.
(iii) By Definition 6, $B^{(\alpha)}(X, Y) = \nabla^{(\alpha)}_{\tilde{X}} \tilde{Y} - \hat{\nabla}^{(\alpha)}_{X} Y$; both sides are independent of α by (ii), so $B^{(\alpha)} = B^{(0)}$.
(iv) Auto-parallelism for the α-connection is $B^{(\alpha)} \equiv 0$, which by (iii) is independent of α; when $\alpha = 0$ this is the condition for total geodesy in the Riemannian manifold $(\mathcal{E}, g)$. $\square$

Remark

The non-trivial dualistic hierarchy $\nabla^{(1)} \neq \nabla^{(-1)}$ arises precisely when $T \not\equiv 0$, i.e. when the model has nonzero third cumulants. Among exponential families this occurs for all non-Gaussian models (discrete, gamma, inverse-Gaussian, etc.). The Gaussian location family with fixed covariance is a canonical example in which the entire α-connection family degenerates to a single Riemannian connection, making the parabolic submanifold of Sections 7.2–7.6 geometrically self-dual.

Explicit Curvature and Asymptotic Estimation

Proposition

For the parabolic submanifold $N : \xi \mapsto (\xi, \xi^2)$ *in* $\mathcal{E} = \{N(\mu, I_2) \colon \mu \in \mathbb{R}^2\}$*:*

(i) *For every* $\alpha \in \mathbb{R}$*, the* α*-second fundamental form is*

$$B^{(\alpha)}(\partial_\xi, \partial_\xi) = \Big(\frac{-4\xi}{1+4\xi^2}, \frac{2}{1+4\xi^2}\Big), \qquad \|B^{(\alpha)}\|^2 = \frac{4}{1+4\xi^2}.$$

(ii) *The information-geometric curvature of* N *in the sense of Amari [1] and Efron [7] (defined as* $\kappa = \|B^{(-1)}\|/h$*, cf. Section 7.4) equals* $\kappa(\xi) = 2/(1+4\xi^2)^{3/2}$*, and this value is the same for all* α*-connections.*

(iii) *Efron's statistical curvature at* ξ *is* $\gamma(\xi) = \|B^{(-1)}\|^2/h(\xi) = 4/(1+4\xi^2)^2$.

Proof.

By Theorem 5(iii), $B^{(\alpha)}$ is independent of α, so it suffices to compute $B^{(1)}$. Since $\Gamma^{(1)}_{ij,k} = 0$ on $\mathcal{E}$, the ambient e-covariant derivative of $\dot{\theta} = (1, 2\xi)$ along N equals $\ddot{\theta} = (0, 2)$. With $g = I_2$:

$$\tan(\ddot{\theta}) = \frac{g(\ddot{\theta}, \dot{\theta})}{g(\dot{\theta}, \dot{\theta})}\dot{\theta} = \frac{4\xi}{1 + 4\xi^2}(1, 2\xi),$$

so $B^{(1)}(\partial_\xi, \partial_\xi) = \ddot{\theta} - \tan(\ddot{\theta}) = \big(-4\xi/(1 + 4\xi^2),\ 2/(1 + 4\xi^2)\big)$, with squared norm $4/(1 + 4\xi^2)$, establishing (i). Parts (ii) and (iii) follow by direct computation with $h = 1 + 4\xi^2$. $\square$

Corollary (Curvature, variance inflation, and information loss)

Let $\hat{\xi}_n$ denote the MLE of ξ on N from n i.i.d. observations. Then:

(i) $n \operatorname{Var}(\hat{\xi}_n) \to 1/h(\xi) = (1 + 4\xi^2)^{-1}$.

(ii) *The statistical curvature* $\gamma(\xi) = 4/(1 + 4\xi^2)^2$ *is strictly decreasing in* $|\xi|$, *with maximum* $\gamma(0) = 4$ *at the vertex and* $\gamma(\xi) \sim 16/\xi^4$ *as* $|\xi| \to \infty$.

(iii) *The expected KL divergence satisfies* $E[D_{\mathrm{KL}}(\hat{p}_n \,\|\, \pi(\hat{p}_n))] \sim \gamma(\xi)/(2n) = 2/[n(1 + 4\xi^2)^2]$ *as* $n \to \infty$.

Proof.

(i) The parabolic submanifold N is a smooth regular curved exponential family satisfying the Fisher regularity conditions (positive definite induced Fisher information $h(\xi) = 1 + 4\xi^2 > 0$ for all ξ, and standard integrability of the log-likelihood). Under these conditions the MLE $\hat{\xi}_n$ is consistent and asymptotically efficient [3, Chap. 4], so $n\,\mathrm{Var}(\hat{\xi}_n) \to 1/h(\xi)$ is the Cramér–Rao bound [1, p. 57]. (ii): $\gamma'(\xi) = -64\xi(1 + 4\xi^2)^{-3} < 0$ for $\xi > 0$, giving strict monotonicity. (iii) follows from the second-order asymptotic theory for curved exponential families [1, Chap. 4], which shows that the expected KL divergence between the empirical distribution and its m-projection onto the model is asymptotically $\gamma(\xi)/(2n)$, where $\gamma(\xi) = \|B^{(-1)}\|^2/h$ is Efron's statistical curvature [7]; combined with Proposition 6(iii) this gives the stated formula. □

Remark

Equations (7) and Corollary 13 together quantify the geometric cost of restricting the bivariate Gaussian to the parabolic constraint $\mu_2 = \mu_1^2$. By Proposition 5, this curvature is intrinsic to the Riemannian structure of $(\mathcal{E}, g)$ and is invisible to the α-connection hierarchy—a feature of Gaussian location families that follows from the vanishing of the Amari–Chentsov tensor T.

Section 8. Conclusions

We have studied statistical submanifolds from the perspective of information geometry, with emphasis on two geometrically and statistically motivated classes: curved exponential families and hierarchical models. The central findings are as follows. The (-1)-second fundamental form of an embedding $N \hookrightarrow \mathcal{E}$ measures statistical curvature in the sense of Efron and Amari, controlling the $O(n^{-1})$ bias of the MLE and the excess KL divergence between the model and its ambient family. Hierarchical models of order r are e-flat coordinate submanifolds of the full exponential family; their higher-order interaction parameters θ^A ($|A| > r$) are the natural parameters of the ambient family missing from the model, and the KL distance to H_r measures the total information contributed by interactions of order $r+1$ and above; near H_r, this distance is approximated to leading order by half the squared Fisher–Rao normal distance (Theorem 12).

Proposition 5 verifies that for Gaussian location families with fixed covariance, the Amari–Chentsov tensor T vanishes identically—a direct consequence of the vanishing of all odd-order moments of the centred Gaussian score vector—causing the entire α-connection hierarchy to collapse to the Riemannian Levi-Civita connection. As a consequence, the α-second fundamental form of any submanifold is independent of α, the notions of *e*- and *m*-auto-parallelism (in the sense of $B^{(\alpha)} \equiv 0$) coincide, and the information-geometric curvature is simultaneously the norm of all $B^{(\alpha)}$. This provides a canonical illustration of how the vanishing of T simplifies the dualistic geometry of a statistical manifold.

The Pythagorean decomposition of KL divergences provides a unified geometric account of projection algorithms in statistics (EM, IPF), and the curvature of the embedding gives a geometric measure of model complexity that refines the dimension-based notions used in AIC and BIC. The explicit Gaussian example of Sections 7.1–7.6 illustrates all these concepts concretely, with fully computed curvature, statistical curvature, and asymptotic KL formula.

Future directions include a systematic study of higher-codimension embeddings for non-Gaussian exponential families (where $T \not\equiv 0$ and the α-connection hierarchy is non-trivial) in the spirit of affine differential geometry [11], the extension of the Gauss–Codazzi equations (1) to non-parametric models via Orlicz spaces [4], and applications to deep latent-variable models viewed as statistical submanifolds of a nonparametric ambient family.

Acknowledgements

The author thanks the organisers of the XXVI International Conference on Geometry, Integrability and Quantization (Varna, 2026) for the invitation to contribute.

Later, Levi-Civita (1896) found the basic equations of geodesic mappings of Riemannian spaces, and also showed their application to modeling mechanical processes.

In the twenties, Weyl and other authors found new results for more general types of spaces, especially spaces with affine connection. In addition, they pointed out the connection between these questions and the then new theory of relativity founded by Einstein.

Bibliography

Amari S., *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics **28**, Springer-Verlag, Berlin 1985.

Amari S. and Nagaoka H., *Methods of Information Geometry*, Translations of Mathematical Monographs **191**, AMS/Oxford University Press, Providence 2000.

Amari S., *Information Geometry and Its Applications*, Applied Mathematical Sciences **194**, Springer, Tokyo 2016.

Ay N., Jost J., Lê Hông Vân and Schwachhöfer L., *Information Geometry*, Ergebnisse der Mathematik und ihrer Grenzgebiete **64**, Springer, Cham 2017.

Cherevko Y., Chepurna O. and Kuleshova Y., *Information Geometry and Some Diffeomorphisms*, Geom. Integr. Quantization **33** (2026) 41–56.

- Cherevko Y., Chepurna O. and Kuleshova Y., *On Information Geometry Methods for Data Analysis*, Geom. Integr. Quantization **29** (2024) 11–22. DOI: 10.7546/giq-29-2024-11-22.
- Efron B., *Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency)*, Ann. Stat. **3** (1975) 1189–1242.
- Kass R. E. and Vos P. W., *Geometrical Foundations of Asymptotic Inference*, Wiley Series in Probability and Statistics, Wiley-Interscience, New York 1997.
- Lauritzen S. L., *Statistical Manifolds*, in: Differential Geometry in Statistical Inference, IMS Lecture Notes – Monograph Series **10**, Institute of Mathematical Statistics, Hayward 1987, pp. 163–216.
- Matsuzoe H., *Statistical manifolds and affine differential geometry*, in: Probabilistic Approach to Geometry, Advanced Studies in Pure Mathematics **57**, Mathematical Society of Japan, Tokyo 2010, pp. 303–321.

- Nomizu K. and Sasaki T., *Affine Differential Geometry*, Cambridge Tracts in Mathematics **111**, Cambridge University Press, Cambridge 1994.
- Opozda B., *Bochner's Technique for Statistical Structures*, Ann. Glob. Anal. Geom. **48** (2016) 357–395.
- Rao C. R., *Information and the Accuracy Attainable in the Estimation of Statistical Parameters*, Bull. Calcutta Math. Soc. **37** (1945) 81–89.
- Vos P. W., *Fundamental Equations for Statistical Submanifolds with Applications to the Bartlett Correction*, Ann. Inst. Stat. Math. **41** (1989) 429–450.

Thank you very much for you attention!