# Embeddings of statistical manifolds

Hông Vân Lê
Institute of Mathematics, CAS

Conference in honor of Shun-ichi Amari
Liblice, June 2016

1. Statistical manifold and embedding of statistical manifolds.

2. Obstructions to the existence of an isostatistical immersion.

3. Outline the proof of the existence of an isostatistical embedding.

4. Final remarks and related open problems.

## 1. Statistical manifold and embedding of statistical manifolds

**Definition.** (Lauritzen, 1987) A statistical manifold $(M, g, T)$ is a manifold $M$ equipped with a Riemannian metric $g$ and a 3-symmetric tensor $T$.

- We assume here that $\dim M < \infty$.

**Examples.**

- Statistical model $(M, \mathfrak{g}, \mathbf{T})$

$$(1) \qquad \mathfrak{g}(\xi; V_1, V_2) = \mathbb{E}_{p(\cdot;\xi)}\left(\frac{\partial \log p}{\partial V_1}\frac{\partial \log p}{\partial V_2}\right),$$

(2)

$$\mathbf{T}(\xi; V_1, V_1, V_1) = \mathbb{E}_{p(\cdot;\xi)}\left(\frac{\partial \log p}{\partial V_1} \frac{\partial \log p}{\partial V_2} \frac{\partial \log p}{\partial V_3}\right).$$

- Manifolds $(M, \rho)$ where $\rho \in C^\infty(M \times M)$ is a divergence (contrast function). (Eguchi) $(M, \rho) \Longrightarrow (g, \nabla, \nabla^*)$, a torsion-free dualistic structure.

**Remarks** 1. $(g, \nabla, \nabla^*) \Longleftrightarrow (g, T)$:

$$T(A, B, C) := g(\nabla_A B - \nabla^*_A B, C).$$

2. Why $(M, g, T)$? : $T$ is "simpler" than $\nabla$.

Lauritzen's question: some Riemannian manifolds with a symmetric 3-tensor $T$ might not correspond to a particular statistical model. If there exist $(\Omega, \mu)$ and $p : \Omega \times M \to R$ conditions hold, we shall call the function $p(x; \xi)$ a probability density for $g$ and $T$. $\Longrightarrow$ Lauritzen question $\Longleftrightarrow$ the existence question of a probability density for the tensors $g$ and $T$ on a statistical manifold $(M, g, T)$.

Lauritzen's question leads to the immersion problem of statistical manifolds.

**Definition.** An immersion $h : (M_1, g_1, T_1) \to (M_2, g_2, T_2)$ will be called isostatistical, if $g_1 = h^*(g_2)$, $T_1 = h^*(T_2)$.

**Lemma.** Assume $h : (M_1, g_1, T_1) \to (M_2, g_2, T_2)$ is an isostatistical immersion. If there exist $\Omega$ and $p(x; \xi_2) : \Omega \times M_2 \to \mathbf{R}$ such that $p$ is a probability density for $g_2$ and $T_2$ then $h^*(p)(x; \xi_1) := p(x; h(\xi_1))$ is a probability density for $g_1$ and $T_1$.

• $(\mathcal{P}_+(\Omega_n), \mathfrak{g}, \mathbf{T})$, where $\#(\Omega_n) = n$, has a natural probability density $p \in C^\infty(\Omega_n \times \mathcal{P}_+(\Omega_n))$, $p(x; \xi) := \xi(x)$.

• Let $g_0 = \sum dx_i^2 \in S^2T^*(\mathbf{R}^n_+)$ be the restriction of the Euclidean metric,

$$T^* := \sum_{i=1}^{n} 2(x_i)^{-1} dx_i^3 \in S^3T^*(\mathbf{R}^n_+).$$

$$\pi^{1/2} : \mathcal{P}_+(\Omega_n) \to \mathbf{R}^n_+$$

$$\xi = \sum_{i=1}^{n} p(i;\xi)\,\delta^i \mapsto 2\sum_{i=1}^{n} \sqrt{p(i;\xi)}\, e_i,$$

is a statistical embedding

$$\pi^{1/2}(g_0) = \mathfrak{g},\ \pi^{1/2}(T^*) = \mathbf{T}.$$

**Main Theorem** (2005/2016) Any smooth ($C^1$ resp.) compact statistical manifold $(M, g, T)$ (possibly with boundary) admits an isostatistical embedding into the statistical manifold $(\mathcal{P}_+(\Omega_N), \mathfrak{g}, \mathbf{T})$ for some finite number $N$. Any finite dimensional noncompact statistical manifold $(M, g, T)$ admits an embedding $I$ into the space $\mathcal{P}_+(\Omega_{\mathbf{N}^+})$ of all positive probability measures on the set $\mathbf{N}^+$ of all natural numbers such that $g$ is equal to the Fisher metric defined on $I(M)$ and $T$ is equal to the Amari-Chentsov tensor on $I(M)$.

**Corollaries**

- Any statistical structure on a manifold is induced from the canonical structure on a statistical model.

- A new proof of Matumoto's theorem asserting that any statistical manifold is generated by a divergence function. Hence $\alpha$-geodesics can be described in terms of gradient flow of relative entropy (Nihat Ay).

## 2. Obstruction to the existence of an isostatistical immersion

**Definition** (Le2007) Let $K(M, e)$ denote the category of statistical manifolds $M$, $e$ - embeddings. A functor of $K(M, e)$ is called a monotone invariant of statistical manifolds.

• Any monotone invariant is an invariant of statistical manifolds.

• Let $f : (M_1, g_1, T_1) \to (M_2, g_2, T_2)$ be a statistical immersion. Then $\forall\ x \in M_1$

$$Df : T_x M_1 \to T_{f(x)} M_2$$

is an isostatistical embedding.

• A statistical manifold $(\mathbf{R}^m, g, T)$ is called a linear statistical manifold, if $g$ and $T$ are constant tensors.

• Functors of the subcategory $K_l(M, e)$ of linear statistical manifolds will be called linear monotone invariants.

Given a linear statistical manifold $M = (\mathbf{R}^n, g, T)$ we set

$$\mathcal{M}^3(T) := \max_{|x|=1,|y|=1,|z|=1} T(x, y, z),$$

$$\mathcal{M}^2(T) := \max_{|x|=1,|y|=1} T(x, y, y),$$

$$\mathcal{M}^1(T) := \max_{|x|=1} T(x, x, x).$$

Clearly we have

$$0 \leq \mathcal{M}^1(T) \leq \mathcal{M}^2(T) \leq \mathcal{M}^3(T).$$

**Proposition 1.** The comasses $\mathcal{M}^i$, $i \in [1,3]$, are nonnegative linear monotone invariants, which vanish if and only if $T = 0$.

$$\mathcal{M}^1(M) := \sup_{x \in M} \mathcal{M}^1(T(x)).$$

**Proposition 2** The comass $\mathcal{M}^1(M)$ is a nonnegative monotone invariant, which vanishes if and only if $T = 0$.

**Proposition 3.** A statistical line $(\mathbf{R}, g_0, T)$ can be embedded into a linear statistical manifold $(\mathbf{R}^N, g_0, T')$, if and only if $\mathcal{M}^1(T) \leq \mathcal{M}^1(T')$.

- Let $(\Gamma^2, \mathfrak{g}, \mathbf{T})$ be the Gaussian model.

$$p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right),\ x \in \mathbf{R}.$$

$$\mathfrak{g}\left(\frac{\partial}{\partial\mu},\frac{\partial}{\partial\mu}\right) = \frac{1}{\sigma^2},\ \mathfrak{g}\left(\frac{\partial}{\partial\mu},\frac{\partial}{\partial\sigma}\right) = 0,$$

$$\mathfrak{g}\left(\frac{\partial}{\partial\sigma},\frac{\partial}{\partial\sigma}\right) = \frac{2}{\sigma^2}.$$

$$\mathbf{T}\left(\frac{\partial}{\partial\mu},\frac{\partial}{\partial\mu},\frac{\partial}{\partial\mu}\right) = 0 = \mathbf{T}\left(\frac{\partial}{\partial\mu},\frac{\partial}{\partial\sigma},\frac{\partial}{\partial\sigma}\right),$$

$$\mathbf{T}\left(\frac{\partial}{\partial\mu},\frac{\partial}{\partial\mu},\frac{\partial}{\partial\sigma}\right) = \frac{2}{\sigma^3},\ \mathbf{T}\left(\frac{\partial}{\partial\sigma},\frac{\partial}{\partial\sigma},\frac{\partial}{\partial\sigma}\right) = \frac{8}{\sigma^3}.$$

$$\mathcal{M}^1(\mathbf{R}^2(\mu, \sigma)) < \infty.$$

$$\mathcal{M}^1(\mathcal{P}_+(\Omega_N), \mathfrak{g}, \mathbf{T}) = \infty.$$

**Proposition 4.** The statistical manifold $(\mathcal{P}_+(\Omega_N), \mathfrak{g}, \mathbf{T})$ cannot be embedded into the Cartesian product of $m$ copies of the normal Gaussian statistical manifold $(\Gamma^2, \mathfrak{g}, \mathbf{T})$ for any $N \geq 4$ and finite $m$.

## 3. Outline the proof of the existence of a isostatistical embedding

Step 1. Prove the existence of an isostatistical immersion.

Step 2. Modify the obtained immersion to get an embedding.

Step 1.

$$T_0 := \sum_{i=1}^{m} dx_i^3 \in S^3(T^*\mathbf{R}^n).$$

**Proposition 1a** Let $(M^m, g, T)$ be compact. Then there exist numbers $N \in \mathbf{N}^+$ and $A > 0$ and a smooth ($C^1$ resp. ) immersion $f : (M^m, g, T) \to (\mathbf{R}^N, g_0, A \cdot T_0)$ s.t. $f^*(g_0) = g$ and $f^*(A \cdot T_0) = T$.

**Nash's embedding theorem.** Any smooth (resp. $C^1$) Riemannian manifold $(M^n, g)$ can be isometrically embedded into $(\mathbf{R}^{N(n)}, g_0)$ for some $N(n)$.

**Gromov's immersion theorem.** Suppose that $T \in \Gamma(S^3 T^* M^m)$. There exists a smooth immersion $f : M^m \to \mathbf{R}^{N_1(m)}$ such that $f^*(T_0) = T$.

**Lemma 1b.** For all $N$ there is a linear isometric embedding $L_N : (\mathbf{R}^N, g_0) \to (\mathbf{R}^{2N}, g_0)$ such that $L_N^*(T_0) = 0$.

**Proposition 1c.** For any $(\mathbf{R}^n, g_0, A \cdot T_0)$ there exists an isostatistical immersion of $(\mathbf{R}^n, g_0, A \cdot T_0)$ into $(\mathcal{P}_+([4n]), \mathfrak{g}, \mathbf{T})$.

• $U(\bar{A}, r)$ - the ball of radius $r$ in the sphere $(S^3, 2\sqrt{n})$ of radius $2/\sqrt{n}$ that centered at $(\lambda(\bar{A}), (2\bar{A})^{-1}, (2\bar{A})^{-1}, (2\bar{A})^{-1}) \subset (S^3, 2\sqrt{n})$.

**Lemma 1d.** For $A > 0$ there exist $\bar{A} > 0$ that depends only on $n$ and $A$, $0 < r$ arbitrarily small and an isostatistical immersion $h$ from $(\mathbf{R}^n, g_0, A{\cdot}T_0)$ into $(\mathcal{P}_+([4n]), \mathfrak{g}, \mathbf{T})$ s.t. $h(\mathbf{R}^n, g_0, A{\cdot}T_0) \subset U(\bar{A}, r) \times_{n \text{ times}} \times U(\bar{A}, r)$.

**Lemma 1e.** There exist a positive number $\bar{A} = \bar{A}(n, A)$ and an embedded torus $T^2$ in $U(\bar{A}, r)$ which is provided with a unit vector field $V$ on $T^2$ such that $T^*(V, V, V) = A$.

• We reduce the existence of an immersion of a *noncompact* of $(M^m, g, T)$ into $(\mathcal{P}_+(\mathbf{N}^+)$ satisfying the condition of the Main Theorem to Case I, using partition of unity and a Nash's trick.

Step 2. To prove the Main Theorem we repeat the proof of the existence of isostatistical immersion, replacing the Nash immersion theorem by the Nash embedding theorem.

The proof is reduced to the proof of the existence of an isostatistical immersion of a bounded statistical interval $([0,R], dt^2, A \cdot dt^3)$ into a torus $T^2$ of a small domain in $(S^7_{2/\sqrt{n},+}, \mathfrak{g}, T^*) \subset (\mathbf{R}^8, g_0, T^*)$. Detail will be in our book “Information Geometry”.

## 4. Final remarks and related problems

• We can replace the compactness of $(M, g, T)$ in the Main Theorem by the boundedness of $\mathcal{M}^3(M, g, T)$.

**Problem**. Find a general setting of differentiable stratified statistical manifolds that are suitable for parameter estimation problems and gradient flow methods.

Motivations:

- S. Amari, Information geometry & Applications, Chapter 12 Natural Gradient Learning and Its Dynamics in Singular Regions,
-D. Geiger, C. Meek, B. Sturmfels, On the toric algebra of graphical models, The Annals of Statistics (2006),
- J. Rauh, T. Kahle, N. Ay, Support sets in exponential families and oriented matroid theory, International Journal of Approximate Reasoning (2011).

How to do? - Apply general Grothendieck abstract ideas in algebraic geometry to differential geometry.

- A. Navarro González and J.B. Sancho de Salas, $C^\infty$-differentiable spaces, volume 1824 of Lecture Notes in Mathematics, appeared in 2003 (excellent for general finite dimension setting).

- H. V. Lê, P. Somberg and J. Vanžura, Poisson smooth structures on stratified symplectic spaces, Springer Proceeding in Mathematics and Statistics, Volume 98, (2015), chapter 7, p. 181-204.

- H.V. Lê, P. Somberg, and J. Vanžura, Smooth structures on pseudomanifolds with isolated conical singularities. Acta Math. Vietnam., 38(1):33-54, 2013.

Problem 1. How to put compatible statistical (geometric) structures on differentiable stratified manifolds?

- Mather (1973), Cheeger (1979-1983), Melrose (1992) etc. proposed different frameworks of singular Riemannian manifolds.

- We have different frameworks for symplectic singular spaces.
- My thesis: we need to focus and pose the condition on the inverse of the Fisher metric, also called the covariance matrix. The covariance matrix is smoothly extended to the boundary of $\mathcal{P}(\Omega_n)$ provided with the canonical smooth structure. Hence the gradient flow is well-behaved.

Problem 2. How far we can extend the setting by Navarro González and Sancho de Salas to nondiscrete sample spaces $\Omega$ but finite (or infinite) dimensional parameter space. (Kriegl-Michor?)

THANK YOU !