IMMERSION OF STATISTICAL MANIFOLDS

HÔNG VÂN LÊ
INSTITUTE OF MATHEMATICS OF ASCR
ZITNA 25, 11567 PRAHA 1, CZECH REPUBLIC
AND
RIKEN BRAIN SCIENCE INSTITUTE
LABORATORY FOR MATHEMATICAL NEUROSCIENCE
WAKO-SHI HIROSAWA 2-1, SAITAMA 351-0198, JAPAN

I. Statistical models and statistical manifolds

In mathematical statistics the data from a sample are interpreted as realizations of stochastically independent random variables. Furthermore one assumes that the data to be analyzed are distributed according to a probability measure in a given statistical model. A statistical model $M := \{(p_x, x \in M)\}$ is a family in the set $Cap(\Omega, \omega)$ of all probability measures on a sample space $(\Omega, \mathcal{A}, d\omega)$. For many reasons we assume that both M and Ω are differentiable manifolds, moreover $d\omega$ is a given Borel measure on Ω. Under this assumption we identify elements of $Cap(\Omega, \omega)$ with density functions $p(x, \omega) \in C^\infty(M \times \Omega)$ which satisfy the following conditions

$$p(x, \omega) \geq 0 \;\forall x \in M, \omega \in \Omega,$$
$$\int_\Omega p(x, \omega)\, d\omega = 1.$$

Around 60-70th years in the last century Nikolai Chentsov in Moscow and Shun-ichi Amari in Tokyo independently discovered an important natural geometric structure on statistical models. This geometric structure consists of *the Fisher metric* g^F and a family ∇^α of torsion-free α-connections, which are called *Amari-Chentsov connections*. The Fisher metric on a statistical manifold $M \subset Cap(\Omega, \omega)$ is defined by

$$g^F(V, W)_x := \int_\Omega (\partial_V \ln p(x, \omega))(\partial_W \ln p(x, \omega)) dp_x. \tag{0.1}$$

The family of Amari-Chentsov connections ∇^α is defined as follows

$$(\nabla^\alpha - \nabla^{LV})(X, Y, Z) = \alpha T^{AC}, \tag{0.2}$$

where ∇^{LV} denotes the Levi-Civita connection of the Riemannian metric g^F, and

$$T^{AC} := \int_\Omega (\partial_X \ln p(x, \omega))(\partial_Y \ln p(x, \omega))(\partial_Z \ln p(x, \omega))\, dp_x. \tag{0.3}$$

Date: August 29.

The 3-symmetric tensor T^{AC} is called *the Amari-Chentsov tensor.*

Example 1. *The space Cap^N* of all probability measures on a finite sample space Ω_N of N elementary events is a statistical model. Let us denote by f_i the probability distribution on Ω_N such that $f_i(\omega_j) = \delta_{ij}$. Each element $f \in Cap^N$ can be written as $\sum_{i=1}^N p_i f_i$, where $p_i \geq 0$ und $\sum p_i = 1$. The interior part Cap_+^N of Cap^N can be described by the following equation

$$Cap_+^N = \{(p_1, \cdots, p_N, \mid p_i > 0 \, \forall i = \overline{1, N} \, \& \sum p_i = 1\}.$$

Clearly Cap_+^N is an open manifold of dimension $N-1$. The following map

$$\phi : Cap_+^N \to S_+^{N-1}(2),$$
$$(p_1, \cdots, p_N) \mapsto (q_1 = 2\sqrt{p_1}, \cdots, q_N = 2\sqrt{p_N}).$$

is an isometry between Riemannian manifolds (Cap_+^N, g^F) and $(S_+^{N-1}(2), g_0)$, where g_0 is the metric of constant curvature, induced from the Euclidean metric on $\mathbb{R}^N$. Now let us denote by T_0 the following tensor on $\mathbb{R}_+^N$

$$T_0(x_1, \cdots, x_N) = \sum_{i=1}^{N} \frac{(dx_i^3)}{x_i}.$$

Then the Amari-Chentsov tensor T^{AC} on Cap_+^N is equal to $\phi^*(T_0)_{|S_+^{N-1}}$.

Example 2. *The 2-dimensional Gaussian statistical model* $(\mathcal{N}(\mu, \sigma^2), g^F, T^{CA})$ consists of normal distributions $\{p(\mu, \sigma, \omega) | \, \mu \in \mathbb{R}, \sigma > 0\}$ on $(\mathbb{R}, d\omega)$ defined by

$$p(\mu, \sigma, \omega) := \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(\omega - \mu)^2}{2\sigma^2}).$$

Its Fisher metric and Amari-Chentsov tensor are defined by

$$g^F(\partial_\mu, \partial_\mu)_{(\mu,\sigma)} = \frac{1}{\sigma^2},$$
$$g^F(\partial_\mu, \partial_\sigma)_{(\mu,\sigma)} = 0,$$
$$g^F(\partial_\sigma, \partial_\sigma)_{(\mu,\sigma)} = \frac{2}{\sigma^2}.$$
$$T^{CA}(\partial_\mu, \partial_\mu, \partial_\mu)_{(\mu,\sigma)} = 0 = T^{CA}(\partial_\mu, \partial_\sigma, \partial_\sigma)_{(\mu,\sigma)},$$
$$T^{CA}(\partial_\mu, \partial_\mu, \partial_\sigma)_{(\mu,\sigma)} = \frac{2}{\sigma^3} = \frac{1}{4} T^{CA}(\partial_\sigma, \partial_\sigma, \partial_\sigma)_{(\mu,\sigma)}.$$

We note that $(\mathcal{N}(\mu, \sigma^2), g^F)$ is a hyperbolic half-plane.

Because of the importance of the Fisher metric and the Amari-Chentsov connections, Stephen Lauritzen introduced the notion of a statistical manifold.

Definition 1. (Lauritzen, 1987). A statistical manifold (M^m, g, T) ist a Riemannian manifold (M^m, g) provided with a 3-symmetric tensor $T \in \Gamma(S^3(T^*M))$.

There are many important examples of statistical manifolds except statistical models provided with the Fisher metric g^F and the Amari-Chentsov tensor. For example, Riemannian manifolds $(M, g, T = 0)$, and Lagrangian submanifolds L in a Kähler manifold (M, J, g), where $g = g_{|L}$, $T(X, Y, Z) = \langle \nabla_X Y, JZ \rangle$ are statistical manifolds.

II Immersion problems of statistical manifolds

To make sure that the notion of a statistical manifold is a right concept, Lauritzen asked if any statistical manifold is a statistical model. Lauritzen's question is equivalent to the following problem. Let (M, g, T) be a statistical manifold. Are there a probability space $(\Omega, \mathcal{A}, d\omega)$ and an immersion of M in the space $Cap(\Omega, d\omega)$ of all probability distributions on $(\Omega, d\omega)$ such that g and T are induced from the Fisher metric and the Chentsov-Amari tensor on the space $Cap(\Omega, d\omega)$?

Definition 2. Assume that (M_1, g_1, T_1) and (M_2, g_2, T_2) are statistical manifolds. An immersion $f : M_1 \to M_2$ is called *statistical*, if $f^*(g_2) = g_1$ and $f^*(T_2) = T_1$.

Motivated by the problem of parameter estimation in mathematical statistics Amari also asked if any statistical model can be regarded as a subfamily in some space Cap^N for $N < \infty$.

We note that the immersion problem for statistical manifolds can be considered as an immersion problem for Riemannian manifolds provided with a constraint - a 3-symmetric tensor.

In this lecture we will explain the following theorem, which gives a positive answer to the Lauritzen and Amari questions.

Main Theorem. [6] *Each statistical manifold (M^m, g, T) is a statistical model. More precisely, for $N = 4(m+1)[2(m(m+1)-1)m(m+1)+2+m)+(m+2)(m+3)]$ there exists a statistical immersion of (M^m, g, T) into the statistical model (Cap^N, g^F, T^{CA}).*

As a direct consequence of the Main Theorem we obtain

Corollary 1. [3] *On each statistical manifold (M, g, T) there exists a contrast function $f \in C^\infty(M \times M)$ which generates the Riemannian metric g and the 3-symmetric tensor T, i.e.*

$$g(X, Y)_{(x,x)} = Hess(K)(i_1(X), i_1(Y)),$$

$$T(X, Y, Z)_x = -\partial_{i_2(Z)} Hess(\rho)(i_1(X), i_1(Y))_{(x,x)} + \partial_{i_1(Z)} Hess(\rho)(i_2(X), i_2(Y))_{(x,x)},$$

where the embedding $i_1, i_2 : T_x M \to T_{(x,x)} M \times M$ are defined by the following formula

$$T_{(x,x)}(M \times M) = (T_x M, 0) \oplus (0, T_x M) = i_1(T_x M) \oplus i_2(T_x M).$$

III Monotone invariants and statistical immersions

In the study of statistical immersions we discover an important class of invariants of statistical manifolds, which we call *monotone invariants*. A monotone invariant is a functor from the category of statistical manifolds whose morphisms are statistical immersions to another category, e.g. the category $(\mathbb{R}, \le)$ of real numbers whose morphisms are the relation $(\le)$.

Clearly monotone invariants are obstructions to the existence of statistical immersions.

We begin our systematical study of monotone invariants with the subcategory of linear statistical manifolds $(\mathbb{R}^n, g_0, T)$, where $g_0 = \sum_{i=1}^n (dx^i)^2$, T is a constant 3-symmetric tensor on $\mathbb{R}^n$, and morphisms are linear immersions of statistical manifolds. A simple way to find monotone invariants for this subcategory is to use the representation theory.

Let us look at the decomposition of the space $S^3(\mathbb{R}^n)^*$ into irreducible $O(n)$-modules.

$$(1) \qquad S^3(\mathbb{R}^n)^* = \mathcal{R}(3\pi_1) \oplus \mathcal{R}^n.$$

Thus each tensor $T \in S^3(\mathbb{R}^n)^*$ can be decomposed into two irreducible components, one of them is defined explicitly as follows

$$(2) \qquad Tr : S^3(\mathbb{R}^n)^* \to \mathcal{R}^n \stackrel{\phi}{=} (\mathbb{R}^n)^*, \ \langle Tr(T), v\rangle := Tr(v\rfloor T),$$

where $v\rfloor T$ is a quadratic form on $\mathbb{R}^n$ and $Tr(v\rfloor T)$ is its trace. The linear map $\phi : (\mathbb{R}^n)^* \to \mathcal{R}^n$ is defined by $v^* \mapsto T^{v^*} := v^* \bullet g_0 \in S^3(\mathbb{R}^n)^*$. Set

$$(3) \qquad \pi_1 : S^3(\mathbb{R}^n)^* \to \mathcal{R}(3\pi_1), \ T \mapsto T - Tr(T).$$

Using (2) and (3) we define the following functor - monotone invariant

$$\hat{T}r : \{(\mathbb{R}^n, g, T)\} \to \{\mathbb{R}\times\mathbb{R}, \ (a, b) \stackrel{m}{\mapsto} (a', b') \text{ iff } 0 < a \le a', \text{ otherwise if } a' = 0 \,\&b \le b'\},$$

$$T \mapsto (||\pi_1(T)||, ||Tr(T)||).$$

Here are some other examples of monotone invariants for linear statistical manifolds. For $(\mathbb{R}^n, g_0, T)$ and $1 \le k \le n$ set

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

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

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

$$\lambda_k(T) := \min_{\mathbb{R}^k \subset \mathbb{R}^n} \mathcal{M}^1(T_{|\mathbb{R}^k}).$$

Monotone invariants also provide sufficient conditions for the existence of statistical immersions.

Proposition 1. 1. $(\mathbb{R}^k, g_0, T)$ can be statistically embedded into $(\mathbb{R}^N, g_0, T^{v^*})$, if and only if $N \ge k$ and $T = T^{w^*}$ with $||w^*|| \le ||v^*||$.
2. A statistical line $(\mathbb{R}, g_0, T)$ can be statistically embedded into $(\mathbb{R}^N, g_0, T')$,

if and only if $\mathcal{M}^1(T) \le \mathcal{M}^1(T')$.
3. We can always statistically embed the 2-dimensional statistical space $(\mathbb{R}^2, g_0, 0)$ into any linear statistical space $(\mathbb{R}^n, g_0, T)$, if $n \ge 7$.
4. Any statistical space $(\mathbb{R}^n, g_0, T)$ can be statistically embedded into the statistical space $(\mathbb{R}^{n(n+1)}, g_0, T' = 2||T|| \sum_{i=1}^{N(n)} x_i^3)$, where x_i are the canonical Euclidean coordinates on $\mathbb{R}^{n(n+1)}$.
5. The trivial space $(\mathbb{R}^n, g_0, 0)$ can be statistically embedded into $(\mathbb{R}^{2n}, g_0, \sum_{i=1}^{2n}(dx^i)^3)$ for all n.

To get monotone invariants for non-linear statistical manifolds we combine the min-max argument and monotone invariants for linear statistical manifolds. Here is an simple example of a monotone invariant of statistical manifolds obtained in this way.

$$\mathcal{M}^1(M) := \sup_{x \in M} \max_{v \in T_x M, |v|_g = 1} T(v, v, v).$$

We have also another type of monotone invariants. For $\rho > 0$ let

$$d_\rho(M, g, T) := \sup\{l \in \mathbb{R}^+ \cup \infty \mid \exists \text{ a statistical immersion of } ([0, l], dx^2, \rho(dx)^3) \text{ to } (M, g, T)\}.$$

$d_\rho(M, g, T)$ is called *the diameter with weight* ρ of (M, g, T). Clearly d_ρ are monotone invariants for all ρ.

Proposition 2. *For any given ρ the diameter with weight ρ of Cap^N is equal to infinity, if $N \ge 4$.*

As a corollary of Proposition 2 we obtain

Proposition 3. *The statistical model Cap^N cannot be statistical immersed into a direct product of m copies of the normal Gaussian statistical manifold $\mathbb{R}^2$ for any $N \ge 3$ and any finite number m.*

An outline of the proof of the Main Theorem using monotone invariants will be given in the lecture.

At the end of the lecture we will discuss the relations between statistical immersions and statistical embeddings, as well as some open problems.

References

[1] S. Amari, H. Nagaoka, Methods of information geometry, Translations of mathematical monographs; v. 191, American Mathematical Society, 2000.

[2] , N. N. Chentsov, Statistical decision rules and optimal inferences, AMR (1982), originally published in Russian, Nauka, Moscow (1972).

[3] T. Matumoto, Any statistical manifold has a contrast function - On the C^3-functions taking the minimum at the diagonal of the product manifolds, Hiroshima Math. J. 23 (1993), 327-332.

[4] E. Morozova and N. Chentsov, Markov invariant geometry on manifolds of states, Itogi Nauki i Techniki, Current problems of mathematics, Newest achievements 6 (1990), 69-102.

[5] N. Morse and R. Sacksteder, Statistical isomorphism, Annals of Math. Statistics, 37 (1966), 203-214.

[6] H. V. Le, Statistical manifolds are statistical models, J. Geom. 84 (2005), 83-93.

[7] H. V. Le, Monotone Invariants and Embeddings of Statistical Manifolds, Advances in Deterministic and Stochastic Analysis, World Scientific (2007), 231-254.