SYMPLECTIC STRUCTURES ON STATISTICAL MANIFOLDS

TOMONORI NODA

Abstract

A relationship between symplectic geometry and information geometry is studied. The square of a dually flat space admits a natural symplectic structure that is the pullback of the canonical symplectic structure on the cotangent bundle of the dually flat space via the canonical divergence. With respect to the symplectic structure, there exists a moment map whose image is the dually flat space. As an example, we obtain a duality relation between the Fubini–Study metric on a projective space and the Fisher metric on a statistical model on a finite set. Conversely, a dually flat space admitting a symplectic structure is locally symplectically isomorphic to the cotangent bundle with the canonical symplectic structure of some dually flat space. We also discuss nonparametric cases.

1. Introduction

Information geometry is the study of probability and information from a differential geometric viewpoint. On the space of probability measures, there exist a natural Riemannian metric, called the Fisher metric, and a family of affine connections, called α-connections. In this paper we see that symplectic structures are very natural for the elementary spaces appearing in information geometry.

Let M be a smooth manifold and ω be a 2-form on M. Denote by $\omega^\flat$ the linear map from TM to T^*M determined by ω. Then ω is a symplectic structure on M if ω is closed and $\omega^\flat$ is an isomorphism. If N is a smooth manifold, then the cotangent bundle T^*N of N admits a natural symplectic structure ω_0. In particular, by identifying $T^*N = T^*\mathbb{R}^n$ with $\mathbb{R}^{2n}$ and taking $(q^1, \ldots, q^n, p_1, \ldots, p_n)$ as coordinates on $\mathbb{R}^{2n}$, we have $\omega_0 = \sum_i dp_i \wedge dq^i$. Conversely, the Darboux theorem states that any symplectic structure has this form locally.

On a symplectic manifold (M, ω), the equations for the Hamiltonian are given by $\omega^\flat(X_H) = dH$, where $H \in C^\infty(M)$. In the case of $(\mathbb{R}^{2n}, \omega_0)$, the equations for the

---

Hamiltonian reduce to the classical Hamilton equations

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i},$$
$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}.$$

There have been a number of attempts to apply symplectic geometry to information geometry. See, for example, Friedrich [5], Shishido [11], Nakamura [9, 10], and Barndorff-Nielsen and Jupp [2]. Friedrich [5] and Shishido [11] studied symplectic structures on the spaces of all probability measures on compact Riemannian manifolds by taking suitable model spaces. In Barndorff-Nielsen and Jupp [2] the notion of a yoke is used to define a symplectic structure on the square of a statistical model. In Nakamura [9, 10] the resemblance between completely integrable systems and statistical models is studied. In the case of the dually flat spaces, defined in Section 2, the relationships between these independently obtained results become very clear. Based on these results, we prove part (i) of the following theorem in Section 3 and part (ii) in Section 4.

THEOREM 1.1. (i) *Let (S, g, ∇, ∇^*) be a dually flat space and D be the canonical divergence on S. Then $S \times S$ admits a natural symplectic structure which is the pullback of the canonical symplectic structure on T^*S via D. With respect to the symplectic structure on $S \times S$, the Hamiltonian flow of D induces geodesic flows for ∇ and ∇^*.*

(ii) *Let $(M, g, J, \omega, \nabla, \nabla^*)$ be a (dually) flat symplectic statistical manifold. Then there exists a dually flat space (S, g_S, ∇^S) such that the symplectic structure ω on M is (locally) isomorphic to the canonical symplectic structure on T^*S.*

Indeed, if we take the canonical divergence on a dually flat space S as a contrast function and the pullback of the canonical symplectic structure on T^*S to be the square $S \times S$ of the dually flat space, then we obtain a 'canonical' symplectic structure on $S \times S$. This construction corresponds to the 'canonical' symplectic structure on the tangent bundle of a manifold obtained by a regular Lagrangian. If we consider the case of the space of all probability distributions on a finite set, then we obtain a duality between Fisher metrics and Fubini–Study metrics (Example 3.1).

If we have a contrast function on a manifold S, then, by taking derivatives, we obtain the statistical structure as a Riemannian metric and connections (see, for example, Eguchi [4], Matsuzoe [8]). In particular, if we take the canonical divergence as a contrast function, then we obtain the canonical symplectic structure on $S \times S$. Note that general contrast functions induce general symplectic structures on $S \times S$ (Barndorff-Nielsen and Jupp [2]).

This construction of symplectic structures on statistical manifolds also works in nonparametric cases. As an analogue of part (i) of Theorem 1.1, we can construct a symplectic structure on the square of maximal exponential manifolds. We also obtain a local isomorphism from $\mathrm{Diff}_k(S^1)/S^1$, the space of L_k^2 probability densities on the

unit circle S^1, to the cotangent bundle of a Lagrangian submanifold of $\text{Diff}_k(S^1)/S^1$. This is an analogue of part (ii) of Theorem 1.1 in the nonparametric case.

2. Statistical models and statistical manifolds

In this section we review some definitions and results from information geometry to be used later. A good general reference for this area is Amari and Nagaoka [1].

Let (X, B, μ) be a measure space. Here, B is a σ-algebra over X and μ is a measure on X. We denote the set of all positive probability density functions on X by

$$\mathcal{P}(X) = \left\{ p : X \to \mathbb{R} \ \middle|\ p > 0, \ \int_X p \, d\mu = 1 \right\}.$$

(The positivity of p is to be interpreted as μ-almost everywhere.) We define a statistical model to be a submanifold of $\mathcal{P}(X)$.

Let $\Xi \subset \mathbb{R}^n$ be open. A collection

$$S := \{ p_\xi = p(x; \xi) \in \mathcal{P}(X) \mid \xi \in \Xi \}$$

of probability distributions on X, parameterized by Ξ, is called an *n-dimensional statistical model* on X.

Throughout this paper, we assume that the map from Ξ to S is one-to-one and is an element of C^∞. We also assume that the derivatives with respect to ξ and the integrals over X of p_ξ in S commute with each other. Moreover, we assume that any statistical model and the dually flat space defined below admit global coordinate systems.

A statistical model

$$S = \{ p_\theta \mid \theta \in \Theta \}$$

is an *exponential family* if there exist functions C and $F_1, \ldots, F_n$ on X and ψ on Θ such that

$$p_\theta = \exp\{ C(x) + \theta^i F_i(x) - \psi(\theta) \}.$$

Here, we use the conventional notation for an exponential family of θ, Θ, instead of ξ, Ξ. Normal distributions, Poisson distributions, and $\mathcal{P}(X)$ for finite X are examples of exponential families.

By the definition, any statistical model is isomorphic to an open subset of $\mathbb{R}^n$ as a point set. It admits a natural Riemannian metric and a family of affine connections as follows.

For a statistical model

$$S = \{ p_\xi = p(x; \xi) \mid \xi \in \Xi \},$$

set $l_\xi(x) := \log p(x; \xi)$. We define

$$g_{ij}(\xi) := E_\xi[\partial_i l_\xi \partial_j l_\xi] = \int_X (\partial_i l_\xi)(\partial_j l_\xi) p_\xi \, dx = -E[\partial_i \partial_j l_\xi],$$

where $\partial_i = \partial/\partial \xi^i$. We call $g = [g_{ij}]$ the *Fisher information matrix* of S. The matrix g is symmetric and positive semidefinite. If g is positive definite, then g defines a Riemannian metric on S and is called the *Fisher metric* on S. The Cramér–Rao inequality asserts that the inverse matrix of the Fisher metric gives a lower bound for the mean squared error of an unbiased estimator.

EXAMPLE 2.1. Let $X = \{x_0, x_1, \ldots, x_n\}$ and

$$\Xi = \left\{ (\xi^1, \ldots, \xi^n) \,\middle|\, \xi^i > 0, 1 \le i \le n, \sum_i \xi^i < 1 \right\}.$$

If we define

$$p_\xi = p(x_i; \xi) = \begin{cases} \xi^i & \text{if } 1 \le i \le n, \\ 1 - \sum_i \xi^i & \text{if } i = 0, \end{cases}$$

then $\mathcal{P}_n := \{p_\xi \mid \xi \in \Xi\}$ is a statistical model and the Fisher metric on $\mathcal{P}_n$ is given by

$$g_{ij}(\xi) = \frac{\delta_{ij}}{\xi^i} + \frac{1}{1 - \sum_i \xi^i}.$$

Note that $\mathcal{P}_n$ is an exponential family for any n.

For a statistical model S with Fisher metric g, we denote the Christoffel symbols of the Levi-Civita connection of g by $\Gamma^{(0)}_{ij,k}$, where $1 \le i, j, k \le n$. For any real number α, set

$$T_{ijk} := E_\xi[\partial_i l_\xi \partial_j l_\xi \partial_k l_\xi], \quad \Gamma^{(\alpha)}_{ij,k} := \Gamma^{(0)}_{ij,k} - \frac{\alpha}{2} T_{ijk}, \tag{2.1}$$

and define an affine connection $\nabla^{(\alpha)}$ by

$$g(\nabla^{(\alpha)}_{\partial_i} \partial_j, \partial_k) = \Gamma^{(\alpha)}_{ij,k}. \tag{2.2}$$

Note that $\nabla^{(\alpha)}$ is torsion-free for all α. We call $\nabla^{(\alpha)}$ the α-connection and the symmetric 3-tensor T the *Chentsov–Amari tensor*. In general, $\nabla^{(\alpha)}$ is not a metric connection unless $\alpha = 0$.

In information geometry, the cases where $\alpha = \pm 1$ are important. The connections $\nabla^{(1)}$ and $\nabla^{(-1)}$ are called the *e-connection* and the *m-connection*, respectively. For all $\alpha \in \mathbb{R}$ and vector fields $X, Y, Z \in \mathfrak{X}(S)$ on S,

$$Zg(X, Y) = g(\nabla^{(\alpha)}_Z X, Y) + g(X, \nabla^{(-\alpha)}_Z Y).$$

This shows the duality between $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$.

Taking into account the properties of statistical models, Lauritzen [7] called a triple (S, g, T) a statistical manifold, where S is a manifold, g is a Riemannian metric and T is a symmetric 3-tensor on S. In this paper we adopt the following definition.

Let (S, g) be a Riemannian manifold and ∇ be an affine connection on S. There exists a unique affine connection ∇^*, called the *dual connection* of ∇, satisfying

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

for any $X, Y, Z \in \mathfrak{X}(S)$. In this situation, the following conditions are equivalent.

(i) ∇ is torsion-free and ∇g is symmetric.

(ii) ∇ and ∇^* are torsion-free.

A triple (S, g, ∇), or equivalently a quadruple (S, g, ∇, ∇^*), satisfying condition (i) or (ii) is called a *statistical manifold*.

It follows from equations (2.1) and (2.2) that any statistical model is a statistical manifold. Conversely, any statistical manifold can be embedded into $\mathcal{P}(X)$ for suitable X (see Vân [12]).

Let (S, g, ∇, ∇^*) be a statistical manifold and R^∇ denote the curvature tensor of ∇. That is,

$$R^\nabla(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$$

for all $X, Y, Z \in \mathfrak{X}(S)$. For any (S, g, ∇, ∇^*), we see that the conditions $R^\nabla = 0$ and $R^{\nabla^*} = 0$ are equivalent. A statistical manifold (S, g, ∇, ∇^*) is called a *dually flat space* if $R^\nabla = 0$. Every exponential family endowed with the e-connection and the m-connection is an example of a dually flat space.

Given a dually flat space (S, g, ∇, ∇^*), there exist affine coordinate systems $(\theta^1, \ldots, \theta^n)$ with respect to ∇ and $(\eta_1, \ldots, \eta_n)$ with respect to ∇^* such that

$$g\left(\frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \eta_j}\right) = \delta^i_j.$$

These coordinate systems are called *dual coordinate systems*. In this case, since

$$\frac{\partial}{\partial \theta^i} \eta_j = g_{ij} = \frac{\partial}{\partial \theta^j} \eta_i,$$

there exists a function $\psi : S \to \mathbb{R}$ such that $d\psi = \eta_i \, d\theta^i$. Similarly, there exists a function $\varphi : S \to \mathbb{R}$ such that $d\varphi = \theta^i \, d\eta_i$. For these φ and ψ,

$$\varphi + \psi = \theta^i \eta_i.$$

The correspondence between $(\theta^1, \ldots, \theta^n)$ and $(\eta_1, \ldots, \eta_n)$ is called the *Legendre transformation*. We define $D : S \times S \to \mathbb{R}$ by

$$D(p, q) := \varphi(p) + \psi(q) - \eta_i(p)\theta^i(q) \quad \forall p, q \in S,$$

and call D the *canonical divergence* on S. In the case of exponential families, D is the Kullback–Leibler divergence (or the relative entropy). In general, we may consider D as the square of a 'distance-like function'. Indeed, the Pythagorean relation holds, but D is not symmetric.

We now summarize some facts about the components of the Fisher metric g on a dually flat space.

LEMMA 2.2. *Let $\{\theta^1, \ldots, \theta^n\}$ and $\{\xi_1, \ldots, \xi_n\}$ be a dual coordinate system on a dually flat space (S, g, ∇, ∇^*). Set $\partial_i := \partial/\partial\theta^i$ and $\partial^j := \partial/\partial\eta_j$. Then the following relations hold:*

$$\theta^i = \frac{\partial \varphi}{\partial \eta_i}, \quad \eta_j = \frac{\partial \psi}{\partial \theta^j},$$

$$d\theta^i = \frac{\partial \theta^i}{\partial \eta_j} \, d\eta_j = g^{ij} \, d\eta_j, \quad d\eta_j = \frac{\partial \eta_j}{\partial \theta^i} \, d\theta^i = g_{ij} \, d\theta^j,$$

$$g_{ij} = g(\partial_i, \partial_j) = \frac{\partial \eta_j}{\partial \theta^i} = \frac{\partial \eta_i}{\partial \theta^j} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j},$$

$$g^{ij} = g(\partial^i, \partial^j) = \frac{\partial \theta^i}{\partial \eta_j} = \frac{\partial \theta^j}{\partial \eta_i} = \frac{\partial^2 \varphi}{\partial \eta_i \partial \eta_j}.$$

3. Dynamics of dually flat spaces: proof of part (i) of Theorem 1.1

Let (S, g, ∇, ∇^*) be a dually flat space. Then there exists a dual coordinate system $\{\theta^1, \ldots, \theta^n\}$ and $\{\eta_1, \ldots, \eta_n\}$. Although these coordinate systems are local coordinate systems on S, we shall consider $\{\theta^1, \ldots, \theta^n\}$ as fiber coordinates on TS and $\{\eta_1, \ldots, \eta_n\}$ as fiber coordinates on T^*S. Then the potential function φ is the Legendre transformation of ψ in the sense of analytical mechanics.

Let S_1 and S_2 be copies of S. The 2-form ω defined by

$$\omega = d\eta_i^* \wedge d\theta^i$$

is a symplectic structure on $S_1 \times S_2$. Here $\{\eta_1^*, \ldots, \eta_n^*\}$ and $\{\theta^1, \ldots, \theta^n\}$ are local coordinates on S_1 and S_2, respectively. By Lemma 2.2,

$$\omega = d\eta_i^* \wedge d\theta^i = g_{ij}^* \, d\theta^{*i} \wedge d\theta^j = g_{ik}^* g^{kj} \, d\theta^{*i} \wedge d\eta_j = g^{ij} \, d\eta_i^* \wedge d\eta_j.$$

The symplectic structure ω is also obtained as follows. For the canonical divergence $D : S_1 \times S_2 \to \mathbb{R}$, given by

$$D(p, q) = \varphi^*(p) - \psi(q) - \eta_i^* \theta^i(q) \quad \forall (p, q) \in S_1 \times S_2,$$

define a map $D_1 : S_1 \times S_2 \to T^*S$ by $(p, q) \mapsto (p, -d_1 D)$, where $d_1 D$ denotes the exterior derivative of D along the first component. By taking a pullback of the canonical symplectic structure ω_0 on T^*S to $S_1 \times S_2$ via D_1, we obtain a symplectic structure ω on $S_1 \times S_2$, that is, $\omega = (D_1)^* \omega_0$.

The Hamiltonian vector field X_D of the canonical divergence D is given by

$$X_D = (\eta_i(q) - \eta_i^*(p)) \frac{\partial}{\partial \eta_i^*} + (\theta^i(q) - \theta^{*i}(p)) \frac{\partial}{\partial \theta^i}.$$

If we restrict the first and the second components of X_D to the diagonal of $S_1 \times S_2$, then we have tangent vectors of ∇^*-geodesics and ∇-geodesics on S, respectively. The geodesic flows for ∇ and ∇^* are obtained by Hamiltonian flows for the canonical divergence. This is Hamilton's principle or the principle of stationary action in physics. We have thus proved part (i) of Theorem 1.1.

Let (M, ω) be a symplectic manifold and G be a Lie group with Lie algebra $\mathfrak{g}$. Assume that G acts on M in Hamiltonian fashion; that is, for any $\xi \in \mathfrak{g}$ there exists $\mu_\xi \in C^\infty(M)$ such that $\omega^\flat(\xi_M) = d\mu_\xi$, where ξ_M denotes the infinitesimal generator associated to $\xi \in \mathfrak{g}$. In this situation, we call the G-equivariant map $\mu : M \to \mathfrak{g}^*$ a *moment map*. Angular and linear momentum are examples of moment maps.

Let (S, g, ∇, ∇^*) be a dually flat space. For simplicity, assume that S is isomorphic to $\mathbb{R}^n$. Let $\{\theta^1, \ldots, \theta^n\}$ and $\{\eta_1, \ldots, \eta_n\}$ be a dual coordinate system on S, and let G_+ be the additive group $(\mathbb{R}^n, +)$. Define the action of G_+ on $S_1 \times S_2$ by

$$(a_1, \ldots, a_n) \cdot (\eta_1^*, \ldots, \eta_n^*, \theta^1, \ldots, \theta^n) := (\eta_1^* + a_1, \ldots, \eta_n^* + a_n, \theta^1, \ldots, \theta^n).$$

For an infinitesimal generator $\partial^{*i} = \partial / \partial \eta_i^*$,

$$\omega^\flat(\partial^{*i}) = d\theta^i.$$

Since $\{\theta^1, \ldots, \theta^n\}$ is a coordinate system on S, we can reconstruct S as the image of the moment map associated to the action of G_+.

In this construction each coordinate function θ^i is a linear momentum. We can also reconstruct S by using angular momentum as follows. Let

$$\mathbb{C}^n = \{(w_1, \ldots, w_n) \mid w_i \in \mathbb{C}, 1 \le i \le n\},$$

and define a map $S_1 \times S_2 \to \mathbb{C}^n$ by

$$w_i = e^{\theta^i + \sqrt{-1}\eta_i^*},$$

where $1 \le i \le n$. The Kähler form

$$\omega_{\mathbb{C}^n} = -\sqrt{-1}g_{ij}\frac{dw_i}{w_i} \wedge \frac{d\bar{w}_j}{\bar{w}_j}$$

corresponds to the symplectic form ω on $S_1 \times S_2$. For the natural action of $T^n = (S^1)^n$ on $\mathbb{C}^n$, the infinitesimal generators are given by

$$v_i = \frac{\sqrt{-1}}{2}\left(w_i\frac{\partial}{\partial w_i} - \bar{w}_i\frac{\partial}{\partial \bar{w}_i}\right),$$

where $1 \le i \le n$. Since $\theta^i = \log |w_i|^2/2$,

$$d\theta^i = \frac{1}{2}\left(\frac{dw_i}{w_i} + \frac{d\bar{w}_i}{\bar{w}_i}\right).$$

Hence

$$\omega^\flat(v_i) = \frac{1}{2}g_{ij}\left(\frac{dw_j}{w_j} + \frac{d\bar{w}_j}{\bar{w}_j}\right) = g_{ij}\,d\theta^i = d\eta_i,$$

and S is obtained as the moment image. If we take S to be the space of all nonnegative probability densities on a finite set and M to be the complex projective space, then we obtain the following example.

EXAMPLE 3.1. Let M be the two-dimensional complex projective space $\mathbb{C}P^2$, that is,

$$M = \{[z_0 : z_1 : z_2] \mid (z_0, z_1, z_2) \in \mathbb{C}^3 \setminus \{(0, 0, 0)\}\}.$$

We write w_1 and w_2 for the inhomogeneous coordinates z_1/z_0 and z_2/z_0 on $\mathbb{C}P^2$. Set $\theta^i := \log |w_i|^2$ and define a Hermitian metric on the anticanonical bundle K_M^{-1} on M by

$$h = e^{-\psi}\left(\sqrt{-1}\frac{dw_1 \wedge d\bar{w}_1}{|w_1|^2}\right) \wedge \left(\sqrt{-1}\frac{dw_2 \wedge d\bar{w}_2}{|w_2|^2}\right),$$

$$\psi(\theta^1, \theta^2) = -\log 9 - \theta^1 - \theta^2 + 3\log(1 + e^{\theta^1} + e^{\theta^2}).$$

The Chern form $c_1(K_M^{-1}; h)$ of this metric is

$$\omega_M = \frac{\sqrt{-1}}{2\pi}\bar{\partial}\partial h = \frac{\sqrt{-1}}{2\pi}\partial\bar{\partial}\psi = \frac{\sqrt{-1}}{2\pi}\frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}\frac{dw_i}{w_i} \wedge \frac{d\bar{w}_j}{\bar{w}_j}. \tag{3.1}$$

Note that ω_M is an Einstein–Kähler metric and is a constant multiple of the Fubini–Study metric on M.

The moment map $\mu : \mathbb{C}P^2 \to \mathbb{R}^2$ for the natural T^2 action on $\mathbb{C}P^2$ is given by

$$\mu = \left(\frac{\partial \psi}{\partial \theta^1}, \frac{\partial \psi}{\partial \theta^2}\right) = \left(-1 + \frac{3|w_1|}{1 + |w_1| + |w_2|}, -1 + \frac{3|w_2|}{1 + |w_1| + |w_2|}\right).$$

The image $\mu(\mathbb{C}P^2)$ is the triangle Δ_2 in $\mathbb{R}^2$ with vertices $(-1, -1)$, $(2, -1)$ and $(-1, 2)$.

Let $\mathcal{P}_3$ be the space of all nonnegative probability densities on the finite set with three elements. We have a parameter space

$$\Xi = \{(\xi^1, \xi^2) \in \mathbb{R}^2 \mid \xi^1 > 0, \xi^2 > 0, \xi^1 + \xi^2 < 1\}$$

of $\mathcal{P}_3$. By setting

$$\eta_1 = 3\xi^1 - 1,$$
$$\eta_2 = 3\xi^2 - 1,$$

we have an isomorphism from $\mathcal{P}_3$ onto Δ_2. Note that (η_1, η_2) is a ∇^*-affine coordinate system on $\mathcal{P}_3$. It follows from Lemma 2.2 and the formula $\eta_i = \partial \psi / \partial \theta^i$ (where $i = 1, 2$), that (θ^1, θ^2) is a ∇-affine coordinate system on $\mathcal{P}_3$. For the Legendre transform $\varphi : \mathbb{R}^2 \to \mathbb{R}$ of ψ,

$$\text{Hess}(\varphi) = \frac{1}{3}\left[\frac{\delta_{ij}}{\xi_i} + \frac{1}{1 - \xi_1 - \xi_2}\right]_{1 \le i, j \le 2}.$$

We see that φ is a potential function of the Fisher metric on $\mathcal{P}_3$ with respect to (η_1, η_2). We obtain the Fisher metric on $\mathcal{P}_3$ as the Hessian matrix of φ. On the other hand, the Hessian matrix of ψ induces the Fubini–Study metric on $M = \mathbb{C}P^2$ by (3.1). In this sense, we obtain a duality between Fisher metrics and Fubini–Study metrics. Note that this duality holds in general dimensions.

In Example 3.1, ψ satisfies a Monge–Ampère equation. The duality correspondence above shows that the Kähler–Einstein metric on $\mathbb{C}P^n$ corresponds to the Fisher metric on the convex polytope as the moment image.

4. Symplectic structures on statistical manifolds: proof of Theorem 1.1(ii)

In this section we give conditions under which a statistical manifold admits a symplectic structure. On statistical manifolds, we already have Riemannian metrics. We may also use almost complex structures to define symplectic structures.

Let (S, g, ∇, ∇^*) be a statistical manifold and J be an almost complex structure on S. If we define the $(1, 1)$ tensor J^* by $g(JX, Y) = -g(X, J^*Y)$, then

$$(J^*)^* = J, \quad (J^*)^2 = -I, \quad g(JX, J^*Y) = g(X, Y).$$

Note that these relations hold for general almost complex structures on Riemannian manifolds, but not necessarily on statistical manifolds. In the case of statistical manifolds, we have the following result.

LEMMA 4.1. *Let (S, g, ∇, ∇^*) be a statistical manifold and J be an almost complex structure on S.*

(i)   *For all $X, Y \in \mathfrak{X}(S)$,*

$$g((\nabla_Z J)X, Y) = -g(X, (\nabla_Z^* J^*)Y).$$

(ii) $g(JX, JY) = g(X, Y)$ *for all $X, Y \in \mathfrak{X}(S)$ if and only if $J = J^*$.*

PROOF. Part (i) follows by direct calculation:

$$
\begin{aligned}
g((\nabla_Z J)X, Y) &= g(\nabla_Z(JX), Y) - g(J\nabla_Z X, Y) \\
&= Zg(JX, Y) - g(JX, \nabla_Z^* Y) + g(\nabla_Z X, J^* Y) \\
&= Zg(JX, Y) + g(X, J^* \nabla_Z^* Y) + Zg(X, J^* Y) - g(X, \nabla_Z^*(J^* Y)) \\
&= -g(X, (\nabla_Z^* J^*)Y).
\end{aligned}
$$

To prove part (ii), suppose that $g(JX, JY) = g(X, Y)$ for all X and Y. By definition, $g(JX, Y) = -g(X, J^* Y)$, and by assumption, $g(JX, Y) = g(J^2 X, JY) = -g(X, JY)$. It follows that $g(X, (J - J^*)Y) = 0$. Since X and Y are arbitrary, $J - J^* = 0$.

Conversely, if $J = J^*$, then

$$g(JX, JY) = -g(X, J^* JY) = -g(X, J^2 Y) = g(X, Y) \quad \forall X, Y \in \mathfrak{X}(S),$$

as required.   □

The next lemma gives a sufficient condition for statistical manifolds to admit symplectic structures.

LEMMA 4.2. *Let (S, g, ∇, ∇^*) be a statistical manifold and J be an almost complex structure on S such that $g(JX, JY) = g(X, Y)$ for all $X, Y \in \mathfrak{X}(S)$. If*

$$\nabla_X^* Y = \nabla_X Y - J(\nabla_X J)Y \quad and \quad (\nabla_X J)Y = (\nabla_Y J)X$$

for all $X, Y \in \mathfrak{X}(S)$, then the 2-form ω on S defined by $\omega(X, Y) := g(JX, Y)$ for all $X, Y \in \mathfrak{X}(S)$ is closed and is parallel with respect to ∇. In particular, ω is a symplectic structure on S and ∇ is a symplectic connection.

PROOF. Since $g(JX, JY) = g(X, Y)$ for all $X, Y \in \mathfrak{X}(S)$,

$$
\begin{aligned}
d\omega(X, Y, Z) &= g(J\nabla_X Y - J[X, Y], Z) + g(J\nabla_Y Z - J[Y, Z], X) \\
&\quad + g(J\nabla_Z X - J[Z, X], Y) \\
&\quad + g(JY, \nabla_X^* Z) + g(JZ, \nabla_Y^* X) + g(JX, \nabla_Z^* Y) \\
&\quad + g((\nabla_X J)Y, Z) + g((\nabla_Y J)Y, X) + g((\nabla_Z J)Y, Y) \\
&= g(JX, \nabla_Z^* Y - \nabla_Z Y) + g(JY, \nabla_X^* Z - \nabla_X Z) \\
&\quad + g(JZ, \nabla_Y^* X - \nabla_Y X) \\
&\quad + g((\nabla_X J)Y, Z) + g((\nabla_Y J)Y, X) + g((\nabla_Z J)Y, Y)
\end{aligned}
$$

for all $X, Y, Z \in \mathfrak{X}(S)$. By using the formulas $\nabla J = -J(\nabla - \nabla^*)$ and $\nabla^* = \nabla - J(\nabla J)$, we obtain

$$d\omega(X, Y, Z) = g(X, (\nabla_Y J)Z - (\nabla_Z J)Y) + g(Y, (\nabla_Z J)X - (\nabla_X J)Z)$$
$$+ g(Z, (\nabla_X J)Y - (\nabla_Y J)X) = 0.$$

In the last equality, we use the fact that $(\nabla_X J)Y - (\nabla_Y J)X = 0$. Therefore $d\omega = 0$.

We next prove that $\nabla \omega = 0$. On the one hand,

$$\nabla_X \omega(Y, JZ) = (\nabla_X \omega)(Y, JZ) + \omega(\nabla_X Y, JZ) + \omega(Y, (\nabla_X J)Z) + \omega(Y, J\nabla_X Z)$$
$$= (\nabla_X \omega)(Y, JZ) + g(\nabla_X Y, Z) - g(Y, J(\nabla_X J)Z) + g(Y, \nabla_X Z).$$

On the other hand,

$$\nabla_X \omega(Y, JZ) = \nabla_X g(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z - J(\nabla_X J)Z)$$
$$= g(\nabla_X Y, Z) + g(Y, \nabla_X Z) - g(Y, J(\nabla_X J)Z).$$

It follows that $\nabla \omega = 0$. $\square$

The converse of Lemma 4.2 also holds.

LEMMA 4.3. *Let (M, ω) be a symplectic manifold and ∇ be a symplectic connection on M. That is, ∇ is torsion-free and ω is parallel with respect to ∇. Let J be an almost complex structure on M adapted to ω. That is, $\omega(JX, JY) = \omega(X, Y)$ for all $X, Y \in \mathfrak{X}(M)$, and the 2-tensor g defined by $g(X, Y) := \omega(X, JY)$ is a Riemannian metric on M. Then $\nabla^* = \nabla - J(\nabla J)$. Moreover, (M, g, ∇) is a statistical manifold if and only if $(\nabla_X J)Y = (\nabla_Y J)X$ for all $X, Y \in \mathfrak{X}(M)$.*

PROOF. It is easy to see that $\nabla - J(\nabla J)$ is an affine connection on S. By hypothesis,

$$X\omega(Y, JZ) = \omega(\nabla_X Y, JZ) + \omega(Y, J(\nabla_X Z - J(\nabla_X J)Z))$$

for all $X, Y, Z \in \mathfrak{X}(M)$, and it follows that

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z - J(\nabla_X J)Z).$$

Hence $\nabla^* = \nabla - J(\nabla J)$. Denote the torsion of ∇^* by T^{∇^*}; then

$$T^{\nabla^*}(X, Y) = (\nabla_X Y - J(\nabla_X J)Y) - (\nabla_Y X - J(\nabla_Y J)X) - [X, Y]$$
$$= \nabla_X Y - \nabla_Y X - [X, Y] - J((\nabla_X J)Y - (\nabla_Y J)X).$$

Therefore $T^{\nabla^*} = 0$ if and only if $(\nabla_X J)Y = (\nabla_Y J)X$ for all $X, Y \in \mathfrak{X}(M)$, as claimed. $\square$

A quintuple $(M, g, J, \omega, \nabla)$ is a *symplectic statistical manifold* if (M, g, ∇) is a statistical manifold, (M, g, J, ω) is an almost Kähler manifold, and ω is parallel with respect to ∇.

LEMMA 4.4. *Let $(M, g, J, \omega, \nabla)$ be a symplectic statistical manifold. Then $\nabla^* \omega = 0$. In particular, the dual connection ∇^* of ∇ is also a symplectic connection.*

PROOF. We first calculate

$$X\omega(Y, JZ) = (\nabla_X^* \omega)(Y, JZ) + \omega(\nabla_X^* Y, JZ) + \omega(Y, (\nabla_X^* J)Z) + \omega(Y, J\nabla_X^* Z)$$
$$= (\nabla_X^* \omega)(Y, JZ) + g(\nabla_X^* Y, z) + \omega(Y, (\nabla_X^* J)Z) + g(Y, \nabla_X^* Z).$$

It then follows that

$$(\nabla_X^* \omega)(Y, JZ) = X\omega(Y, JZ) - g(\nabla_X^* Y, Z) - g(Y, \nabla_X^* Z) - \omega(Y, (\nabla_X^* J)Z)$$
$$= g(\nabla_X Y, Z) - g(\nabla_X Y - J(\nabla_X J)Y, Z) - \omega(Y, (\nabla_X^* J)Z)$$
$$= g(J(\nabla_X J)Y, JZ) - \omega(Y, (\nabla_X^* J)Z)$$
$$= -g((\nabla_X J)Y, JZ) - \omega(Y, (\nabla_X^* J)Z)$$
$$= -g(\nabla_X(JY), JZ) + g(J\nabla_X Y, JZ) - \omega(Y, (\nabla_X^* J)Z)$$
$$= -Xg(JY, JZ) + g(JY, \nabla_X^*(JZ))$$
$$\quad + g(\nabla_X Y, Z) - \omega(Y, (\nabla_X^* J)Z)$$
$$= -Xg(Y, Z) + g(Y, \nabla_X^* Z) + g(JY, (\nabla_X^* J)Z)$$
$$\quad + g(\nabla_X Y, Z) - \omega(Y, (\nabla_X^* J)Z) = 0$$

for all $X, Y, Z \in \mathfrak{X}(M)$, establishing our result.   □

Next we assume that $(M, g, J, \omega, \nabla)$ is a symplectic statistical manifold and $R^\nabla = 0$. Then there exists a (locally) ∇-flat coordinate system $\{\theta^1, \ldots, \theta^{2n}\}$, where $2n$ denotes the dimension of M. Set

$$\omega = \omega_{ij} \, d\theta^i \wedge d\theta^j.$$

Since

$$\nabla\omega = d\omega_{ij} \wedge d\theta^i \wedge d\theta^j = 0,$$

it follows that

$$0 = (d\omega_{ij} \wedge d\theta^i \wedge d\theta^j)(\partial_i, \partial_j, \partial_k) = d\omega_{ij}(\partial_k),$$

where $\partial_i = \partial/\partial\theta^i$. Hence any ω_{ij} is a (locally) constant function. Normalizing ω by an element in $\mathrm{Sp}(2n; \mathbb{R})$, we may assume that, in terms of $\{\theta^1, \ldots, \theta^{2n}\}$,

$$\omega = d\theta^1 \wedge d\theta^{n+1} + \cdots + d\theta^n \wedge d\theta^{2n}.$$

Here, $\{\theta^1, \ldots, \theta^{2n}\}$ is a (locally) ∇-flat Darboux coordinate system.

If $\{\theta^1, \ldots, \theta^{2n}\}$ is a ∇-flat Darboux coordinate system, then the dual coordinate system $\{\eta_1, \ldots, \eta_{2n}\}$ is a ∇^*-flat Darboux coordinate system. Moreover,

$$J\frac{\partial}{\partial\eta_i} = \begin{cases} \partial/\partial\theta^{n+i} & \text{if } 1 \leq i \leq n, \\ -\partial/\partial\theta^{i-n} & \text{if } n+1 \leq i \leq 2n. \end{cases} \tag{4.1}$$

Indeed, $d\theta^{n+i}(J\partial^j) = \delta_i^j$ when $i \in \{1, \ldots, n\}$, since

$$\omega(\partial_i, J\partial^j) = \delta_i^j, \quad \omega = \sum_k d\theta^k \wedge d\theta^{n+k}, \quad J\partial^j = \partial_{n+i}.$$

The case where $i \in \{n+1 \ldots 2n\}$ can be established similarly. Note that $J\partial/\partial\theta^i$ cannot coincide with $\partial/\partial\theta^j$ if i and j are distinct.

Consider the integral submanifold L defined by $\{\theta^1 = 0, \ldots, \theta^n = 0\}$ in a flat Darboux coordinate system $\{\theta^1, \ldots, \theta^{2n}\}$. We see that ω is just the canonical symplectic structure on T^*L. This proves part (ii) of Theorem 1.1.

5. Concluding remarks: nonparametric cases

In this section, we see that Theorem 1.1 holds in infinite dimensions. We first review the definitions of symplectic structures on Banach manifolds. Let M be a Banach manifold and ω be a 2-form on M. Let $\omega^\flat$ denote the linear map from TM to T^*M induced by ω. We say that ω is a *weak symplectic structure* if ω is closed and $\omega^\flat$ is injective. If in addition $\omega^\flat$ is an isomorphism, then ω is called a *strong symplectic structure*. It is clear that any strong symplectic structure is weak symplectic, but the converse does not hold in general. If M is finite-dimensional, then any weak symplectic structure is also a strong symplectic structure.

EXAMPLE 5.1. Let B be a Banach space and B^* be its dual. Set $E := B \times B^*$ and define a skew symmetric bilinear form on E by

$$\omega((u, v), (u', v')) = v'(u) - v(u') \quad \forall (u, v), (u', v') \in B \times B^*.$$

Then ω is a weak symplectic structure on E. In this case, $\omega^\flat : B \times B^* \to B^* \times B^{**}$ is given by $\omega^\flat((u, v)) = (v, u)$ for all $(u, v) \in B \times B^*$. It is easily seen that ω is a strong symplectic structure if and only if B is reflexive.

Let (Ω, χ, μ) be a probability space and denote the space of all probability distributions on Ω by

$$\mathcal{M} := \left\{ p \;\middle|\; p > 0, \int_X p \, d\mu = 1 \right\}.$$

(The positivity of p is to be interpreted as μ-almost everywhere.) Cena and Pistone [3] show that $\mathcal{M}$ admits a smooth manifold structure whose connected component is of the form

$$\mathcal{E}(p) = \{e^{u - K_p(u)} p \mid u \in L^{\Phi_1}(p), E_p[u] = 0, K_p(u) < \infty\}.$$

Here $L^{\Phi_1}(p)$ denotes the Orlicz space

$$L^{\Phi_1}(p) = \{u \mid \inf_{\alpha > 0} E_p[\Phi_1(\alpha u)] < +\infty\}$$

for the Young's function $\Phi_1(x) = \cosh x - 1$, and K is the cumulative generating function. Note that $L^{\Phi_1}(p)$ is a Banach space, but it is not reflexive.

By taking the pullback of the canonical symplectic structure on T^*S via the canonical divergence D as in the finite-dimensional case, we obtain a weak symplectic structure on $\mathcal{E}(p) \times \mathcal{E}(p)$, given by

$$\omega((v_1, v_2), (w_1, w_2)) = -E_{p_1}[v_1 w_2] + E_{p_1}[w_1 v_2]$$

for all $(v_1, v_2), (w_1, w_2) \in T_{p_1}\mathcal{E}(p) \times T_{p_2}\mathcal{E}(p)$. With respect to the symplectic structure ω, if we formally consider the Hamiltonian vector field of D, then the second component of the Hamiltonian vector field induces an exponential arc as in the finite-dimensional case. Although the first component of the Hamiltonian vector field is a constant function that corresponds to a mixture arc, it is not tangent to $\mathcal{E}(p_1)$.

Finally, we see an analogue of part (ii) of Theorem 1.1 in infinite dimensions. Let dV be the normalized volume form $dt/2\pi$ on the unit circle S^1, and consider the space of all L_k^2 probability distributions on S^1 defined by

$$\mathcal{P}^k(S^1) = \left\{ f \in L_k^2 \,\middle|\, f > 0, \int_{S^1} f \, dV = 1 \right\}.$$

Here $L_2^k(X)$ denotes the Sobolev space consisting of functions on X of finite L_k^2-norm.

It is known that $\mathcal{P}^k(S^1)$ is a Hilbert manifold that admits a strong symplectic structure (see Friedrich [5] and Shishido [11]). Further, it is known that $\mathcal{P}^k(S^1)$ admits a Kähler structure (see Kirillov and Yur'ev [6]). Since the symplectic structure on the Hilbert manifold $\mathcal{P}^k(S^1)$ is strong, every maximal isotropic submanifold is Lagrangian. There exists a Lagrangian submanifold $\mathcal{L}$ in $\mathcal{P}^k(S^1)$ such that $\mathcal{P}^k(S^1)$ is locally symplectically isomorphic to the cotangent bundle $T^*\mathcal{L}$ of $\mathcal{L}$ (see, for example, Weinstein [13]).

Acknowledgement

References

[1]   S. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs, 191 (American Mathematical Society, Providence, RI, and Oxford University Press, Oxford, 2000), translated from the 1993 Japanese original by Daishi Harada.

[2]   O. E. Barndorff-Nielsen and P. E. Jupp, 'Statistics, yokes and symplectic geometry', *Ann. Fac. Sci. Toulouse Math.* **6** (1997), 389–427.

[3]   A. Cena and G. Pistone, 'Exponential statistical manifold', *Ann. Inst. Statist. Math.* **59** (2007), 27–56.

[4]   S. Eguchi, 'Geometry of minimum contrast', *Hiroshima Math. J.* **22** (1992), 631–647.

[5]   T. Friedrich, 'Die Fisher-Information und symplektische Strukturen', *Math. Nachr.* **153** (1991), 273–296.

[6]   A. A. Kirillov and D. V. Yur'ev, 'Kähler geometry of the infinite-dimensional homogeneous space $M = \text{Diff}_+(S^1)/\text{Rot}(S^1)$', *Funktsional. Anal. i Prilozhen.* **20** (1986), 79–80 (in Russian).

[7]   S. Lauritzen, 'Statistical manifolds', in: *Differential Geometry in Statistical Inference*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 10 (Inst. Math. Statist., Hayward, CA, 1987), pp. 163–216.

[8]   H. Matsuzoe, 'Geometry of contrast functions and conformal geometry', *Hiroshima Math. J.* **29** (1999), 175–191.

[9] Y. Nakamura, 'Completely integrable gradient systems on the manifolds of Gaussian and multinomial distributions', *Japan J. Indust. Appl. Math.* **10** (1993), 179–189.

[10] Y. Nakamura, 'Gradient system associated with probability distributions', *Japan J. Indust. Appl. Math.* **11** (1994), 21–30.

[11] Y. Shishido, 'Strong symplectic structures on spaces of probability measures with positive density function', *Proc. Japan Acad.* **81** (2005), 134–136.

[12] L. H. Vân, 'Monotone invariants and embeddings of statistical manifolds', in: *Advances in Deterministic and Stochastic Analysis* (World Scientific, Hackensack, NJ, 2007), pp. 231–254.

[13] A. Weinstein, 'Symplectic manifolds and their Lagrangian submanifolds', *Adv. Math.* **6** (1971), 329–346.

TOMONORI NODA, Department of Mathematics, Osaka Dental University, 8-1, Kuzuhahanazono-cho, Hirakata-shi, Osaka 573-1121, Japan
e-mail: noda-t@cc.osaka-dent.ac.jp