HOOKE'S LAW IN STATISTICAL MANIFOLDS AND DIVERGENCES

MASAYUKI HENMI AND RYOICHI KOBAYASHI

Abstract. The concept of the canonical divergence is defined for dually flat statistical manifolds in terms of the Legendre transform between dual affine coordinates. In this article, we introduce a new two point function defined for any triple (g, ∇, ∇^*) of a Riemannian metric g and two affine connections ∇ and ∇^*. We show that this interprets the canonical divergence without refering to the existence of special coordinates (dual affine coordinates) but in terms of only classical mechanics concerning ∇- and ∇^*-geodesics. We also discuss the properties of the two point function and show that this shares some important properties with the canonical divergence defined on dually flat statistical manifolds.

§0. Introduction

Let (M, g, ∇, ∇^*) be a Riemannian manifold with two affine connections (not necessarily metric preserving and possibly with torsion) ∇ and ∇^*. In this article we introduce a two point function $W(p\|q)$ and a "two point function" $W_c(p\|q)$ depending on paths connecting p and q (see Definition 1.1). The two point function $W(p\|q)$ is defined by connecting q and p by a unique ∇-geodesic and computing the necessary work to move a unit mass from q to p against the force field induced by the "Hooke's law" applied to ∇^*-geodesics (considered as a spring) connecting points of the ∇-geodesic and q. On the other hand, the paths depending "two point function" $W_c(p\|q)$ is defined by the same procedure (applying Hooke's law to ∇^*-springs) along any path c connecting p and q. We study fundamental properties of these two point functions. To formulate a potential theoretic property of W_c, we need to introduce a class of torsion-free[1] statistical manifolds satisfying the

---

Received March 4, 1999.
Revised June 2, 1999.
2000 Mathematics Subject Classification: 53A15, 53B05.

[1] A statistical manifold (M, g, ∇, ∇^*) is said to be torsion-free iff both ∇ and ∇^* are torsion-free.

condition (S) or $(S)^*$ (see Definition 4.1)[2].

THEOREM 0.1.   (Theorem 4.3 (i)) *If* (M, g, ∇, ∇^*) *is a torsion-free statistical manifold satisfying the condition* (S), *then* $W_c(p\|q) \equiv W(p\|q)$ *holds, i.e.,* $W_c(p\|q)$ *does not depend on the choice of paths connecting* p *and* q.

Suppose that (M, g, ∇, ∇^*) is a torsion-free statistical manifold satisfying the condition (S) or $(S)^*$. Then the two point function $W(q\|p)$ coincides with the two point function $W^*(p\|q)$ defined by interchanging the role of ∇- and ∇^*-geodesics.

THEOREM 0.2.   (Theorem 5.4) *Suppose that* (M, g, ∇, ∇^*) *is a torsion-free statistical manifold satisfying the condition* $(S)^*$ *(resp.* (S)*). Then there exists a function* f *(resp.* f^**) satisfying the condition*

$$f(0) = 0, \ \ f'(0) = 1 \ \ \ (resp. \ f^*(0) = 0, \ \ (f^*)'(0) = 1)$$

such that

$$W(q\|p) = f(W^*(p\|q)) \ \ \ (resp. \ W^*(q\|p) = f^*(W(p\|q)))$$

holds.

Moreover we will show that the two point function $W_c(p\|q) \equiv W(p\|q)$ on torsion-free statistical manifolds generalizes the concept of the canonical divergence defined originally on dually flat statistical manifolds in terms of the Legendre transform between dual affine coordinates. Moreover if the statistical mamifold under consideration satisfies the condition (S), two point function $W_c(p\|q) \equiv W(p\|q)$ on such torsion-free statistical manifolds shares the orthogonality property (potential property) with the canonical divergence and Theorem 0.1 for dually flat case is a consequence of this property:

THEOREM 0.3.   (Theorem 6.3 and Theorem 4.2 (ii)) (i) *For dually flat statistical manifolds, the two point function* $W(p\|q)$ *coincides with the canonical divergence* $D(p\|q)$.
(ii) *Let* (M, g, ∇, ∇^*) *be a torsion-free statistical manifold satisfying the condition* (S). *Then each* ∇^*-*geodesic emanating from* q *is perpendicular to the level hypersurfaces of the function* $W(\,\cdot\,\|q)$.

---

[2]The condition (S) is a statistical geometric analogue of the concept of symmetric spaces in Riemannian geometry.

It follows from Theorems 0.1 and 0.3 (ii) that if a torsion-free statistical manifold under consideration satisfies the condition (S), then $W(\cdot\|q)$ is the potential function for the conservative force field induced from Hooke's law applied to ∇^*-geodesics emanating from q.

Finally we would like to mention a conjectural meaning of the two point function $W(p\|q)$ in the asymptotic theory of statistical estimation. Let (M, g, ∇, ∇^*) be an n-parameter statistical model (i.e., a family of everywhere positive absolutely continuous probability distributions on a fixed probability space $(X, \sigma(X))$), where g is the Fisher metric, $\nabla = \nabla^{(1)}$ and $\nabla^* = \nabla^{(-1)}$. The relative entropy $D(p\|q)$ $(p, q \in M)$ is defined by

$$D(p\|q) = \int_X \log \frac{dp}{dq}(\omega)p(d\omega),$$

(dp/dq being the Radon-Nikodym derivative of mutually absolutely continuous probability measures p and q) which coincides with the canonical divergence $((-1)$-divergence) if (M, g, ∇, ∇^*) is an exponential family which is dually-flat. In general case, we have two two point functions $W(p\|q)$ and $D(p\|q)$. The function $W(p\|q)$ depends essentially on the geometry of the parameter space (M, g, ∇, ∇^*) and the function $D(p\|q)$ is completely independent of the parameter space to which p and q belong. So, to understand the difference of these two point functions would be interesting. It would roughly be understood as follows. Let $M^{(k)}$ consist of k points corresponding to zero-simplices of a simplicial decomposition of M. We then try to find a finite dimensional exponential family $P(k)$ on $(X, \sigma(X))$ together with an embedding $i^{(k)} : M^{(k)} \to P^{(k)}$. We then approximate M by constructing (from $i^{(k)}(M^{(k)})$) a piecewise ∇^*-linear simplicial complex in $P^{(k)}$ corresponding to the simplicial decomposition of M. The limit of taking finer and finer simplicial decompositions of M would produce an embedding $i^{(\infty)}$ of the given statistical model M into "the maximal exponential family" $P^{(\infty)}$ on $(X, \sigma(X))$, which is "dually-flat". The two point function $D(p\|q)$ is now the canonical divergence with respect to the ambient "maximal exponential family" and the difference of two point functions $D(p\|q)$ and $W(p\|q)$ explains the extrinsic-geometrical nature (the second fundamental form) of the virtual embedding $i^{(\infty)} : M \to P^{(\infty)}$.

§1. Hooke's law in Riemannian manifolds with two affine connections

Throughout this article, we always restrict our attention to such re-

gions in a given Riemannian manifold any pair of two points of which is joined by a unique geodesic with respect to connections under consideration. Moreover we use Einstein's convention, e.g., a metric tensor is written as $ds^2 = g_{ij}\, dx^i dx^j$ instead of $ds^2 = \sum_{i,j} g_{ij}\, dx^i dx^j$. All parametric curves $c(t)$ in this article are parameterized by a parameter t in the unit interval $[0,1]$ with $c(0)$ and $c(1)$ their end points, unless otherwise specified.

Let (M, g, ∇, ∇^*) be a Riemannian manifold with two affine connections ∇ and ∇^* (not necessarily metric preserving and possibly with torsion) and p, q two points of M which are connected by a unique ∇- (resp. ∇^*-) geodesic c (resp. c^*). Now we consider the ∇-geodesic $c(t)$ starting at q and ending at p ($t \in [0,1]$, $c(0) = q$ and $c(1) = p$). Assume that for all $t \in [0,1]$ there exists a unique ∇^*-geodesic $c_t^*(s)$ ($s \in [0,1]$) joining $c_t^*(0) = c(t)$ and $c_t^*(1) = q$. Let a point of unit mass be located at $c(t)$. Now we consider the ∇^*-geodesic $c_t^*(s)$ as a "spring" which is stretched from the "equilibrium" state q. Appling *Hooke's law* to the ∇^*-springs $c_t^*(s)$ ($t \in [0,1]$) at each $c(t) = c_t^*(0)$, we get a force field (the stress)

$$F(c(t)) = \frac{d}{ds}\Big|_{s=0} c_t^*(s).$$

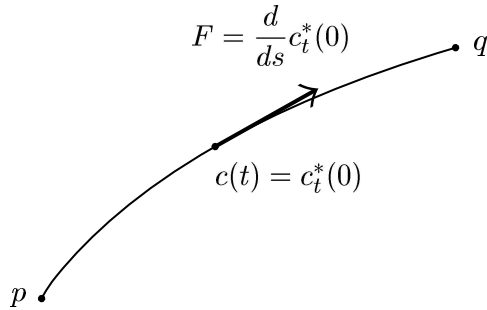along the ∇-geodesic $c(t)$ ($t \in [0,1]$) (see Figure 1).

Figure 1: Hooke's law in Riemannian manifolds with two connections.

The work $W(p\|q)$ which is necessary to move a point of unit mass from q to p along the ∇-geodesic $c(t)$ against the force field $F(c(t))$ is given by the integral

$$(1) \qquad W(p\|q) := -\int_0^1 dt \left[g_{ij}(c(t)) \frac{d}{dt} c^i(t) \frac{d}{ds} c_t^{*j}(0) \right].$$

On the other hand, we introduce the quantity $W_c(p\|q)$ (which depends on the Riemannian metric g and the connection ∇^* and on the choice of paths c joining q and p) by the same integral as in (1) performed along any path $c(t)$ ($0 \leq t \leq 1$) with $c(0) = q$ and $c(1) = p$:

$$(2) \qquad W_c(p\|q) := -\int_0^1 dt \left[g_{ij}(c(t)) \frac{d}{dt} c^i(t) \frac{d}{ds} c_t^{*j}(0) \right],$$

which is the work necessary to move a point of unit mass along a path c from q to p against the force field induced by the ∇^*-spring.

DEFINITION 1.1. (i) (Two point function) Let (g, ∇, ∇^*) be any triple of a Riemannian metric g and any two affine connections ∇ and ∇^*. For any two points p and q such that there is a unique ∇-geodesic $c(t)$ ($c(0) = q$ and $c(1) = p$) and such that for any $t \in [0, 1]$ there is a unique ∇^*-geodesic $c_t^*(s)$ ($c_t^*(0) = c(t)$ and $c_t^*(1) = q$), the two point function $W(p\|q)$ is defined by the integral (1).

(ii) ("Two point function" depending on paths) Let $c(t)$, $0 \leq t \leq 1$, be any piecewise smooth path such that $c(0) = q$, $c(1) = p$ and such that for any $t \in [0, 1]$ there is a unique ∇^*-geodesic $c_t^*(s)$ ($c_t^*(0) = c(t)$ and $c_t^*(1) = q$). The "two point function" $W_c(p\|q)$ depending on paths c is defined by the integral (2).

§2. Simple examples of Hooke's law

EXAMPLE 2.1. The stress which is caused by small deformations of a spring obeys Hooke's law

$$F = -kx$$

where F is the magnitude of the stress, x stands for the displacement compared to its natural length and k is its elastic constant. It follows that the work which is necessary to move a point of unit mass attached to the spring by a small distance x from the equilibrium state is given by the integral

$$\int_0^x kx \, dx = \frac{1}{2} kx^2.$$

It is therefore natural to think that, forgetting the spring, the "potential energy" which exists between two points p and q of the real line $\mathbb{R}$ (with its natural metric) is equal to

$$\frac{1}{2} k(p - q)^2.$$

EXAMPLE 2.2. (Hooke's Law w.r.t. the Levi-Civita connection) Let (M, g) be a Riemannian manifold ($g = g_{ij} \, dx^i dx^j$ being its Riemannian metric) equipped with the Levi-Civita connection. We consider a parametric curve $c(t)$ ($t \in [0, 1]$) in M. Set $c(0) = q$, $c(1) = p$. Let $c_t^*(s)$ (s is proportional to the arclength and $s \in [0, 1]$) be the geodesic connecting $c(t)$ and q such that $c_t^*(0) = c(t)$ and $c_t^*(1) = q$. We think of the geodesic arc $c_t^*(s)$ ($s \in [0, 1]$) as a "small displacement" in the positive direction of the "spring" from its "equilibrium" state q. And we consider that a point of unit mass located at $c(t)$ is attached to the "spring" c_t^*. Applying Hooke's law in this situation yields the force field

$$F(c(t)) := \frac{d}{ds}\Big|_{s=0} c_t^*(s) = -\operatorname{grad} \rho_q(c(t))$$

along the curve c, where ρ_q is the half of the squared Riemannian distance function from q (see Figure 2):

$$\rho_q(\,\cdot\,) = \rho(\,\cdot\,, q) := \frac{1}{2} \operatorname{dist}(\,\cdot\,, q)^2.$$
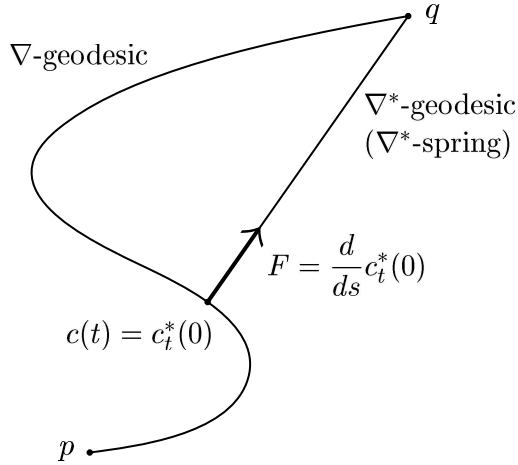

Figure 2: Hooke's law w.r.t. the Levi-Civita connection.

The work $W_c(p\|q)$ which is necessary to move a point of unit mass from q to p along the curve c against this force field is given by the half of

the squared Riemannian distance $\frac{1}{2}\operatorname{dist}(p,q)^2$ between p and q as is shown below:

$$-\int_0^1 dt\left[g_{ij}(c(t))\frac{d}{dt}c^i(t)\frac{d}{ds}c_t^{*j}(0)\right] = \int_0^1 dt\left\langle\frac{d}{dt}c(t),\operatorname{grad}\rho_q(c(t))\right\rangle_{c(t)}$$

$$= \int_0^1 dt\,\frac{d}{dt}\rho_q(c(t))$$

$$= \rho_q(p) = \frac{1}{2}\operatorname{dist}(p,q)^2.$$

Therefore the integral $W_c(p\|q) := -\int_0^1 g(c(t))_{ij}\frac{d}{dt}c^i(t)\frac{d}{ds}c_t^{*j}(0)\,dt$ depends only on end points q and p and does not depend on the choice of paths c connectiong these (hence $W_c(p\|q) \equiv W(p\|q)$). It is therefore natural to think, forgetting the spring, the "potential energy" which exists between two points p and q in M is

$$\frac{1}{2}\operatorname{dist}(p,q)^2.$$

§3. The concept of statistical manifolds ([N-A], [A], [A-N])

Suppose that we are given a Riemannian metric g and two affine connections ∇ and ∇^* on a smooth manifold M.

DEFINITION 3.1. (cf. [N-A], [A], [A-N]) By a *dual structure* on M, we mean a triple (g,∇,∇^*) on M such that

$$(3) \qquad Zg(X,Y) = g(\nabla_Z X,Y) + g(X,\nabla_Z^* Y).$$

A *statistical manidold* is a Riemannian manifold M equipped with a dual structure (g,∇,∇^*).

Torsion freeness for both ∇ and ∇^* is assumed very often in the definition of statistical manifolds. However, as we are interested in the effect of torsion-freeness itself in this article, we do not consider torsion free-ness as an a priori condition. Instead, we define that a statistical manifold (M,g,∇,∇^*) is *torsion-free* iff both ∇ and ∇^* are torsion-free. For any Riemannian manifold (M,g) with an affine connection ∇, there exists a unique affine connection ∇^* (the dual connection) satisfying the condition (3). Thus we can make a statistical manifold (M,g,∇,∇^*) from a given Riemannian manifold with an affine connection ∇.

In terms of local coordinates (ξ^i), the above condition (3) is equivalent to

$$\partial_i g_{kj} = \Gamma_{ik,j} + \Gamma^*_{ij,k}$$

where $\Gamma_{ik,j} := g(\nabla_{\partial_i}\partial_k, \partial_j)$ and $\Gamma^*_{ij,k} := g(\nabla^*_{\partial_i}\partial_j, \partial_k)$. We note that the symmetry $g_{ij} = g_{ji}$ implies

(4) $\qquad \Gamma_{ik,j} + \Gamma^*_{ij,k} = \Gamma_{ij,k} + \Gamma^*_{ik,j} \Longleftrightarrow \Gamma^*_{ij,k} - \Gamma_{ij,k} = \Gamma^*_{ik,j} - \Gamma_{ik,j}.$

In general there is no relation between the torsion tensor of ∇ and its dual ∇^*. Suppose now that, for a statistical manifold (M, g, ∇, ∇^*), both ∇ and ∇^* are torsion-free. Then we have

$$\partial_j g_{ik} = \begin{cases} \Gamma^*_{ji,k} + \Gamma_{jk,i} = \Gamma^*_{ij,k} + \Gamma_{jk,i} \\ \Gamma^*_{jk,i} + \Gamma_{ji,k} = \Gamma^*_{jk,i} + \Gamma_{ij,k} \end{cases}$$

which implies

(5) $\qquad \Gamma^*_{ij,k} + \Gamma_{jk,i} = \Gamma^*_{jk,i} + \Gamma_{ij,k} \Longleftrightarrow \Gamma^*_{ij,k} - \Gamma_{ij,k} = \Gamma^*_{jk,i} - \Gamma_{jk,i}.$

EXAMPLE 3.2. (α-connections ([A], [A-N])) Let

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

be a statistical model of (absolutely continuous) probability distributions on a fixed probability space $(X, \sigma(X))$, where Θ is a parameter space which is an open subset of $\mathbb{R}^n$. Assume that $p_\xi(x) > 0$ everywhere on X. The Fisher metric g on the parameter space Θ is defined by

$$g_{ij} := E_\xi[\partial_i l_\xi \partial_j l_\xi],$$

where $\partial_i = \partial/\partial\xi^i$, $l_\xi = \log p_\xi$. For any $\alpha \in \mathbb{R}$, put

$$(\Gamma^{(\alpha)}_{ij,k})_\xi := E_\xi\left[\left(\partial_i \partial_j l_\xi + \frac{1-\alpha}{2}\partial_i l_\xi \partial_j l_\xi\right)(\partial_k l_\xi)\right].$$

Then, using the Fisher metric, we set

$$g\left(\nabla^{(\alpha)}_{\partial i}\partial_j, \partial_k\right) = \Gamma^{(\alpha)}_{ij,k}.$$

It turns out that this defines a torsion-free affine conection called the α-connection. Then we have

$$\partial_i g_{kj} = \Gamma^{(\alpha)}_{ik,j} + \Gamma^{(-\alpha)}_{ij,k},$$

i.e., the α and $(-\alpha)$-connections are dual to each other.

§4. $W(p\|q)$ and ∇^*-geodesics in statistical manifolds

First of all we recall that we always work on regions any two points of which are jouned by a unique geodesic with respect to connections under consideration.

Let (M, g, ∇, ∇^*) be a statistical manifold and q, p two points which are joined by unique ∇- and ∇^*-geodesics. Let $c(t)$ be the ∇-geodesic such that $c(0) = q$ and $c(1) = p$. Assume moreover that there is a unique ∇^*-geodesic $c_t^*(s)$ joining $c_t^*(0) = c(t)$ and $c_t^*(1) = q$. Recall that

$$W(p\|q) = -\int_0^1 dt\left[g_{ij}(c(t))\frac{d}{dt}c^i(t)\frac{d}{ds}c_t^{*j}(0)\right],$$

i.e., $W(p\|q)$ is equal to the work necessary to move a point of unit mass from q to p against the force field defined by the tangent vector $\frac{d}{ds}c_t^*(0)$ of the ∇^*-geodesic joining $c(t)$ to q. Therefore, under a variation $c(\varepsilon, t)$ of ∇-geodesics emanating from q ending on a fixed level-hypersurface $W(\,\cdot\,\|q) = c$, the work which is necesarry to move a point of unit mass, against the force field defined by the tangent vector (at $c(t)$) of ∇^*-geodesics $c_t^*(s)$, from q to the endpoint of $c(\varepsilon, t)$ on $W(\,\cdot\,\|q) = c$ does not change. It therefore might be expected that, if the statistical manifold (M, g, ∇, ∇^*) under consideration is torsion-free, at each point $c(t)$ the ∇^*-geodesic $c_t^*(s)$ is orthogonal to the level-hypersurface of $W(\,\cdot\,\|q)$ defined by $W(\,\cdot\,\|q) = W(c(t)\|q)$. We prove this under certain assumption on (M, g, ∇, ∇^*).

DEFINITION 4.1. An n-dimensional torsion-free statistical manifold (M, g, ∇, ∇^*) satisfies the condition (S) iff:

(i) $\langle R^*(V, T)T, T\rangle = 0\ (\forall V, T)$.

(ii) $\nabla^* R^* = 0$ (R^* being the curvature tensor of ∇^*).

We say that (M, g, ∇, ∇^*) satisfies the condition (S) iff (i) and (ii) hold for ∇ and R (R being the curvature tensor of ∇).

The concept of torsion-free statistical manifolds satisfying the condition (S) is an analogue of that of symmetric spaces in Riemannian geometry. Moreover the condition (S)-(i) is a necessary condition for the function $W_c(p\|q)$ to be a two-point function (see the remark after the proof of Proposition 4.2).

We introduce an auxiliary two point function $W_*(p\|q)$ by choosing a path c in Definition 1.1 (ii) to be a ∇^*-geodesic c_* connecting q and p. Then we have:

PROPOSITION 4.2. *Suppose that a torsion-free statistical manifold* $(M,$ $g, \nabla, \nabla^*)$ *satisfies the condition* (S). *Then each* ∇^*-*geodesic emanating from* q *is perpendicular to the level-hypersurfaces of the function* $W_*(\,\cdot\,\|q)$.

Proof. Let $c(\tau, t)$ $(-\varepsilon < \tau < \varepsilon,\ 0 \leq t \leq 1)$ be a variation of ∇^*-geodesics such that

$$\begin{cases} c(0, t) = c_*(t) \\ c(\tau, 0) = q \quad (\forall\, \tau) \\ \langle V, T \rangle(1) = 0 \end{cases}$$

where $T := \frac{\partial}{\partial t} c(\tau, t)$ (the velocity vector field along $c_*(t)$) and $V := \left(\frac{\partial}{\partial \tau}\right)_{\tau=0} c_{\tau,t}$ (the variation vector field, i.e., a ∇^*-Jacobi field).

Let $c_{\tau,t}^*(s)$ $(0 \leq s \leq 1)$ be a ∇^*-geodesic such that $c_{\tau,t}^*(0) = q$ and $c_{\tau,t}^*(1) = c(\tau, t)$. Then, if we set $S := \left(\frac{\partial}{\partial s}\right)_{s=1} c_{\tau,t}^*(s)$, we have the proportionality relation $S = tT$. The vector fields T and V are considered to be the image of coordinate vector fields under some smooth map $A :$ $[0,1]_t \times (-\varepsilon, \varepsilon)_\tau \to M$, where $[0,1]_t$ denotes the interval $[0,1]$ coordinatized by t, *etc.* In this setting, we have:

$$\left\langle \operatorname{grad} W_*(\,\cdot\,\|q), \frac{d}{d\tau}\Big|_{\tau=0} c(\tau, 1) \right\rangle = \frac{d}{d\tau}\Big|_{\tau=0} W_*(c(\tau, 1)\|q) = \int_0^1 V\langle T, S\rangle \, dt$$

$$= \int_0^1 \left(\langle \nabla_V T, S \rangle + \langle T, \nabla_V^* S \rangle \right) dt$$

$$= \int_0^1 \left(\langle \nabla_T V, S \rangle + \langle T, \nabla_V^* S \rangle \right) dt \quad \text{(because } \nabla \text{ is torsion-free)}$$

$$= \int_0^1 \left(T\langle V, S \rangle - \langle V, \nabla_T^* S \rangle + \langle T, \nabla_V^* S \rangle \right) dt$$

$$= [\langle V, S \rangle]_0^1 + \int_0^1 \left(\langle T, \nabla_V^* S \rangle - \langle V, \nabla_T^* S \rangle \right) dt$$

$$= \langle V, T \rangle(1) - \int_0^1 \langle T, V \rangle \, dt + \int_0^1 t\langle T, \nabla_V^* T \rangle \, dt$$
$$\text{(because } c_{0,t}^* \text{ is a } \nabla^*\text{-geodesic and } S = tT)$$

$$= -a + x,$$

where

$$a := \int_0^1 \langle T, V \rangle \, dt \quad \text{and} \quad x := \int_0^1 t\langle T, \nabla_T^* V \rangle \, dt.$$

Now we would like to show that $-a + x = 0$ if the condition (S) is fulfilled.

Let $\mathcal{J}$ (resp. $\mathcal{J}_0$) be the space of all ∇^*-Jacobi field along the ∇^*-geodesic $c_*(t)$ (resp. vanishing at the "origin" q). Set

$$V_0 := \{J \in \mathcal{J}_0 \; ; \; \exists k \in \mathbb{Z}_{>0}, \; (\nabla_T^*)^k J \text{ is } \nabla^*\text{-parallel.}\}.$$

The condition (S)-(ii) and (∇^*)-Jacobi's differential equation imply that the composite $(\nabla_T^*)^2$ induces a linear map of $\mathcal{J}_0$ to itself. It then follows from the definition of V_0 that $(\nabla_T^*)^2$ induces an isomorphism

$$(\nabla_T^*)^2 : \mathcal{J}_0/V_0 \longrightarrow \mathcal{J}_0/V_0.$$

Therefore we may assume that one of the following (a) or (b) occurs:

(a) the ∇^*-Jacobi field V (which we started with) is contained in the image of $(\nabla_T^*)^2 : \mathcal{J}_0 \to \mathcal{J}_0$,

(b) $V(1)$ is the boundary value of a ∇^*-Jacobi field J in V_0.

Suppose first that (a) occurs. Then we may write $V = \nabla_T^* W$ for some $W \in \mathcal{J}$. Then we have

$$x = -\int_0^1 t \langle R^*(W,T)T, T \rangle \, dt = 0$$

because the integrand vanishes identically by the condition (S)-(i). Moreover we have $V = (\nabla_T^*)^2 U$ for some $U \in \mathcal{J}_0$ and therefore

$$a = -\int_0^1 \langle R^*(U,T)T, T \rangle \, dt = 0$$

by the condition (S)-(i).

Suppose next that (b) occurs. Then J is written as $tJ_1 + t^2 J_2 + \cdots + t^k J_k$ where $0 \le t \le 1$, $0 < k < n$ and J_i are all parallel Jacobi fields along $c_*(t)$. As $\langle V, T \rangle(1) = 0$, we may assume that each J_i is perpendicular to T at $t = 1$. Indeed, there exist constants α_i such that $\sum_{i=1}^k \alpha_i = 0$ and each $J_i + \alpha_i T$ is perpendicular to T at $t = 1$. For each J_i we consider a coordinate system (x) defined in a tubular neighborhood of $c_*(t)$ with respect to which the connection ∇^* is written as a sum of a flat trivial connection ∇_0 plus an error term $O(\varepsilon)$:

$$\nabla^* = \nabla_0 + O(\varepsilon),$$

where ε is the Euclidean distance from $c_*(t)$. As we are working along $c_*(t)$, allowing the error of order ε, we may pretend to work on a 2-dimensional dually flat space (and let $\varepsilon \to 0$ later). In fact, we prove a result (Theorem 6.3) from which Proposition 4.2 follows in dually flat situation. Then, letting $\varepsilon \to 0$ at each i yields the desired result. $\qquad\square$

We remark here that the condition (S)-(i), i.e., the property

$$\langle R^*(V,T)T,T \rangle = 0$$

for the curvature tensor R^* of ∇^*, is a necessary condition for the validity of Proposition 4.2. It is easy to check this. Indeed, suppose that $V = \nabla_T^* W$. Then we consider the Taylor expansion at q of the quantity

$$-a + x = -\int_0^1 \langle T, \nabla_T^* W \rangle \, dt - \int_0^1 t \langle R^*(W,T)T,T \rangle \, dt$$

under the condition that the distance of q and p is very small.

In the Riemannian case, the integral (2) defined for any curve $c(t)$ $(c(0) = q$ and $c(1) = p)$ depends only on end points p and q. The following Theorem 4.3, which follows from Proposition 4.2, asserts that this is the case for any torsion-free statistical manifold (M, g, ∇, ∇^*) satisfying the condition (S).

THEOREM 4.3. *Let (M, g, ∇, ∇^*) be a torsion-free statistical manifold satisfying the condition (S). Then:*

(i) The quantity $W_c(p\|q)$ (defined for any curve $c(t)$ with end points $c(0) = q$ and $c(1) = p$) does not depend on the choice of paths c and depends only on end points q and p. In other words, $W_c(p\|q) \equiv W(p\|q)$ holds.

(ii) The two point function $W(p\|q)$ is the potential energy for the conservative force field defined by the ∇^*-spring: $\frac{d}{ds}c_p^*(0)$ where $c_p^*(s)$ is the ∇^*-geodesic such that $c_p^*(0) = p$ and $c_p^*(1) = q$.

Proof. Let $c_*(t)$ be a ∇^*-geodesic such that $c_*(0) = q$ and $c_*(1) = p$. For a fixed t let $c_t^*(s)$ be a unique ∇^*-geodesic such that $c_t^*(0) = q$ and $c_t^*(1) = c_*(t)$. Then

$$W_*(p\|q) = \int_0^1 dt \left\langle \frac{d}{dt}c_*(t), \frac{d}{ds}c_t^*(1) \right\rangle_{c_*(t)}.$$

We extend a ∇^*-geodesic $c_1^*(s)$ by $\varepsilon > 0$ across $c_1^*(1) = c_*(1) = p$ and consider $p_\tau := c_1^*(1 + \tau)$ $(-\varepsilon < \tau < \varepsilon)$ as a one-parameter variation of the endpoint p. Let $c_*(\tau, t)$ be a one-parameter variation of ∇^*-geodesics such that $c_*(\tau, 0) = q$ for all τ and $c_*(\tau, 1) = p_\tau$. Repeating the same argument as in the proof of Proposition 4.2, we have

$$
(6) \qquad \left\langle \operatorname{grad} W_*(\cdot \,\|q), \frac{d}{d\tau}\Big|_{\tau=0} p_\tau \right\rangle = \frac{d}{d\tau}\Big|_{\tau=0} W_*(p_\tau\|q)
$$

$$
= \left\langle \frac{d}{ds}\Big|_{s=1} c_1^*(s), \frac{d}{d\tau}\Big|_{\tau=0} p_\tau \right\rangle.
$$

By Proposition 4.2 all ∇^*-geodesics emanating from q are perpendicular to the level hypersurfaces of $W_*(\,\cdot\,\|q)$. Therefore (6) implies

$$\operatorname{grad} W_*(\,\cdot\,|q)(c_1^*(1)) = \frac{d}{ds}\Big|_{s=1} c_1^*(s).$$

It follows that the integral curves of $\operatorname{grad} W_*(\,\cdot\,\|q)$ coincides with the ∇^*-geodesics. Therefore, if $c(t)$ is any curve such that $c(0) = q$ and $c(1) = p$, we have

$$W_c(p\|q) = \int_0^1 dt \left\langle \frac{d}{dt}c(t), \frac{d}{ds}c_t^*(1) \right\rangle_{c(t)} = \int_0^1 dt \frac{d}{dt} W_*(c(t)\|q)$$

$$= W_*(c(1)\|q) = W_*(p\|q).$$

Therefore the integral (2) depends only on end points q and p. In particular, we have $W_c(p\|q) = W_*(p\|q) = W(p\|q)$. □

COROLLARY 4.4. *Let M be as in Proposition 4.2. Let S be a smooth submanifold of M of codimension ≥ 1. Let q be a point of M such that $q \notin S$. Then the function $r \mapsto W(r\|q)$ defined on S takes its minimum at a point $p \in S$ iff the ∇^*-geodesic joining q and p is orthogonal to the submanifold S at p.*

DEFINITION 4.5. *Let (g, ∇, ∇^*) be a triple of a Riemannian metric g and two affine connections ∇ and ∇^*. For any two points p and q such that there is a unique ∇^*-geodesic $c^*(t)$ $(c^*(0) = q$ and $c^*(1) = p)$ and such that for any $t \in [0,1]$ there is a unique ∇-geodesic $c_t(s)$ $(c_t(0) = c^*(t)$ and $c_t(1) = q)$, the two point function $W^*(p\|q)$ is defined by the integral*

$$(1^*) \qquad W^*(p\|q) := -\int_0^1 dt \left[g_{ij}(c^*(t)) \frac{d}{dt} c^{*i}(t) \frac{d}{ds} c_t^j(0) \right].$$

The "two point function" $W_{c^*}^*(p\|q)$ depending on paths c^* can be defined for any path $c^*(t)$ $(0 \leq t \leq 1)$ connecting q and p by performing the same integral as above along c

$$(2^*) \qquad W_{c^*}^*(p\|q) := -\int_0^1 dt \left[g_{ij}(c^*(t)) \frac{d}{dt} c^{*i}(t) \frac{d}{ds} c_t^j(0) \right],$$

c_t being the ∇-geodesic such that $c_t(0) = c^*(t)$ and $c_t(1) = q$ as above. Finally we can define the dual version $(S)^*$ of the condition (S) by interchanging the role of ∇ and ∇^*.

Interchanging the role of ∇- and ∇^*-geodesics in the definition of $W(p\|q)$ (Definition 1.1) yields the above definition of $W^*(p\|q)$.

COROLLARY 4.6. *If (g, ∇, ∇^*) is a torsion-free statistical manifold satisfying the condition $(S)^*$. Then $W^*_{c^*}(p\|q) = W^*(p\|q)$ holds and the function $W^*(p\|q)$ is the potential energy for the conservative force field defined by the ∇-spring: $\frac{d}{ds}c_p(0)$ where $c_p(s)$ is the ∇-geodesic such that $c_p(0) = p$ and $c_p(1) = q$.*

§5. Contrast function defined by $W(p\|q)$

Eguchi [E] introduced a construction of statistical structures on a given manifold using the concept of the *contrast function*. In this section we will compute derivatives of $W(p\|q)$ and we will follow notations in [E] (see also [A-N]). Let M be a smooth manifold. A contrast function $\rho(p\|q)$ is defined in a neighborhood of the diagonal set of $M \times M$. For a function $\rho(p\|q)$ to be a contrast function, we require

$$\rho(p, q) \geq 0, \quad \rho(p, q) = 0 \Longleftrightarrow p = q$$

and

$$\rho(\partial_i \partial_j \| \cdot) = -\rho(\partial_i \| \partial_j) = g_{ij}$$

is positive definite along the diagonal. Here, we define

$$\rho(\partial_i \partial_j \| \cdot) = (\partial_i)_p (\partial_j)_p \rho(p\| \cdot), \quad \rho(\partial_i \| \partial_j) = (\partial_i)_p (\partial_j)_q \rho(p\|q)$$

and so forth where $(\partial_i)_p$ and $(\partial_i)_q$ are $\partial/\partial \xi^i$ operating on the variables p and q, respectively. If $\rho(p, q)$ is a contrast function, then

$$\Gamma^\rho_{ij,k} := -\rho(\partial_i \partial_j \| \partial_k)$$

defines a torsion-free affine conection and

$$\Gamma^{\rho^*}_{ij,k} := -\rho(\partial_k \| \partial_i \partial_j)$$

is its dual connection (also torsion-free) with respect to the Riemannian metric g_{ij}. Now let p and q are sufficiently close to each other so that both ∇- and ∇^*-geodesics combining q and p are sufficiently close to a line. Let fix a coordinate system so that (p^i) and (q^i) are coordinate expressions of points p and q.

In the rest of this section, we assume that (M, g, ∇, ∇^*) is a statistical manifold (not necessarily torsion-free).

LEMMA 5.1. *Let $c(t)$ $(t \in [0,1])$ be the ∇-geodesic combining $q = c(0)$ and $p = c(1)$ (where q and p are sufficiently close to each other). Then we have*

$$\frac{d}{dt}c^i(t) = (p^i - q^i) + \frac{1}{2}\Gamma^i_{jk}(q)(p^j - q^j)(p^k - q^k)$$
$$- t\Gamma^i_{jk}(q)(p^j - p^j)(p^k - q^k) + O(\|p - q\|^3).$$

Proof. We consider a map

$$\frac{d}{dt}\Big|_{t=0}c^i(t) \longmapsto c^i(1)$$

which sends the initial vector at $t = 0$ of ∇-geodesics emanating from q to the end point $c^i(1)$. Introduce indeterminate coefficients x^i_{jk} by setting

$$\frac{d}{dt}\Big|_{t=0}c^i(t) = (p^i - q^i) + x^i_{jk}(p^j - q^j)(p^k - q^k) + O(\|p - q\|^3).$$

As $t \mapsto c(t)$ is a ∇-geodesic, we have

$$\frac{d^2}{dt^2}c^i(t) + \Gamma^i_{jk}(c(t))\frac{d}{dt}c^j(t)\frac{d}{dt}c^k(t) = 0.$$

It follows that the Taylor expansion of $c^i(t)$ in t around $t = 0$ is

$$c^i(t) = c^i(0) + t\{(p^i - q^i) + x^i_{jk}(p^j - q^j)(p^k - q^k)\}$$
$$- \frac{t^2}{2}\Gamma^i_{jk}(q)\{(p^j - q^j) + x^j_{ab}(p^a - q^a)(p^b - q^b)\}$$
$$\times \{(p^k - q^k) + x^k_{cd}(p^c - q^c)(p^d - q^d)\}$$
$$+ O(\|p - q\|^3).$$

Therefore we have

$$c^i(1) = c^i(0) + (p^i - q^i) + x^i_{jk}(p^j - q^j)(p^k - q^k)$$
$$- \frac{1}{2}\Gamma^i_{jk}(q)(p^j - q^j)(p^k - q^k) + O(\|p - q\|^3).$$

Putting $c^i(1) = p^i$, we have

$$x^i_{jk} = \frac{1}{2}\Gamma^i_{jk}(q).$$

□

We substitute Taylor expansions

$$g_{ij}(c(t)) = g_{ij}(q) + t\partial_k g_{ij}(q)(p^k - q^k) + O(\|p - q\|^2),$$

$$\frac{d}{dt}c^i(t) = (p^i - q^i) + \frac{1}{2}\Gamma^i_{jk}(q)(p^j - q^j)(p^k - q^k)$$
$$- t\Gamma^i_{jk}(q)(p^j - q^j)(p^k - q^k) + O(\|p - q\|^3),$$

$$\frac{d}{ds}c^{*j}_t(0) = (q^j - c^j(t)) + \frac{1}{2}\Gamma^{*j}_{kl}(c(t))(q^k - c^k(t))(q^l - c^l(t))$$
$$+ O(\|q - c(t)\|^3)$$

of all factors involved in the integral (1). As

$$q^j - c^j(t) = q^j - \left[q^j + t\left\{(p^j - q^j) + \frac{t}{2}\Gamma^j_{kl}(q)(p^k - q^k)(p^l - q^l)\right\}\right.$$
$$\left. - \frac{t^2}{2}\Gamma^j_{kl}(q)(p^k - q^k)(p^l - q^l)\right] + O(\|p - q\|^3)$$

$$= -t(p^j - q^j) - \frac{t}{2}\Gamma^j_{kl}(q)(p^k - q^k)(p^l - q^l)$$
$$+ \frac{t^2}{2}\Gamma^j_{kl}(q)(p^k - q^k)(p^l - q^l) + O(\|p - q\|^3),$$

we have

$$\frac{d}{ds}c^{*j}_t(0) = (q^j - c^j(t)) + \frac{1}{2}\Gamma^{*j}_{kl}(c(t))(q^k - c^k(t))(q^l - c^l(t))$$
$$+ O(\|q - c(t)\|^3)$$
$$= -t(p^j - q^j) - \frac{t}{2}\Gamma^j_{kl}(q)(p^k - q^k)(p^l - q^l)$$
$$+ \frac{t^2}{2}\Gamma^j_{kl}(q)(p^k - q^k)(p^l - q^l) + \frac{t^2}{2}\Gamma^{*j}_{kl}(c(t))(q^k - p^k)(q^l - p^l)$$
$$+ O(\|q - c(t)\|^3).$$

Therefore we have

(7)
$$W(p\|q) = -\int_0^1 dt\left[g_{ij}(c(t))\frac{d}{dt}c^i(t)\frac{d}{ds}c^{*j}_t(0)\right]$$
$$= -\int_0^1 dt\left[\{g_{ij}(q) + t\partial_k g_{ij}(q)(p^k - q^k)\}\right.$$
$$\left. \times \left\{(p^i - q^i) + \frac{1}{2}\Gamma^i_{ab}(q)(p^a - q^a)(p^b - q^b) - t\Gamma^i_{ab}(q)(p^a - q^a)(p^b - q^b)\right\}\right.$$

$$\times \left\{ -t(p^j - q^j) - \frac{t}{2}\Gamma_{cd}^j(q)(p^c - q^c)(p^d - q^d) \right.$$

$$\left. + \frac{t^2}{2}\Gamma_{cd}^j(q)(p^c - q^c)(p^d - q^d) + \frac{t^2}{2}\Gamma_{cd}^{*j}(q)(p^c - q^c)(p^d - q^d) \right\} \right]$$

$$+ O(\|p - q\|^4).$$

THEOREM 5.2. *The integral $W(p\|q)$ is a contrast function and induces the original Riemannian metric:*

$$(8) \qquad\qquad\qquad -W(\partial_i\|\partial_j) = g_{ij}.$$

Proof. It follows from (7) that

$$W(p\|q) \geq 0, \quad W(p\|q) = 0 \iff p = q$$

if p and q are sufficiently close to each other. This implies that $W(p\|q)$ is a contrast function. Applying $(\partial_i)_p(\partial_j)_q$ to (7) and putting $p = q$, we have (8). $\qquad\square$

THEOREM 5.3. *The connection induced from the contrast function $W(p\|q)$ is given by*

$$(9) \quad -W(\partial_i\partial_j\|\partial_k) = \frac{1}{3}(\Gamma_{ij,k} + \Gamma_{ji,k} + \Gamma_{jk,i} + \Gamma_{kj,i} + \Gamma_{ki,j} + \Gamma_{ik,j})$$

$$+ \frac{1}{6}(\Gamma_{ij,k}^* + \Gamma_{ji,k}^* + \Gamma_{jk,i}^* + \Gamma_{kj,i}^* + \Gamma_{ki,j}^* + \Gamma_{ik,j}^*)$$

$$- \Gamma_{ki,j} - \Gamma_{kj,i}^*.$$

In particular, if both ∇ and ∇^* are torsion-free, the connection (9) coincides with the original connection ∇.

Proof. Direct computation shows that

$$-W(\partial_i\partial_j\|\partial_k) = -\partial_k g_{ij} + \frac{2}{3}(\partial_k g_{ij} + \partial_i g_{jk} + \partial_j g_{ki})$$

$$- \frac{1}{6}\left\{ g_{ak}(\Gamma_{ij}^{*a} + \Gamma_{ji}^{*a}) + g_{aj}(\Gamma_{ki}^{*a} + \Gamma_{ik}^{*a}) + g_{ai}(\Gamma_{jk}^{*a} + \Gamma_{kj}^{*a}) \right\}$$

$$= -\Gamma_{ki,j} - \Gamma_{ik,j}^* + \frac{2}{3}(\Gamma_{ki,j} + \Gamma_{kj,i}^* + \Gamma_{ij,k} + \Gamma_{ik,j}^* + \Gamma_{jk,i} + \Gamma_{ji,k}^*)$$

$$- \frac{1}{6}(\Gamma_{ij,k}^* + \Gamma_{ji,k}^* + \Gamma_{ki,j}^* + \Gamma_{ik,j}^* + \Gamma_{jk,i}^* + \Gamma_{kj,i}^*).$$

Now we use (4) to conclude

$$
\begin{aligned}
-W&(\partial_i\partial_j\|\partial_k)\\
&= \frac{1}{3}(\Gamma_{ij,k} + \Gamma_{ji,k} + \Gamma_{jk,i} + \Gamma_{kj,i} + \Gamma_{ki,j} + \Gamma_{ik,j})\\
&\quad + \frac{1}{6}(\Gamma^*_{ij,k} + \Gamma^*_{ji,k} + \Gamma^*_{jk,i} + \Gamma^*_{kj,i} + \Gamma^*_{ki,j} + \Gamma^*_{ik,j}) - \Gamma_{ki,j} - \Gamma^*_{kj,i}\\
&\quad + \frac{1}{3}\{(\Gamma_{ki,j} - \Gamma_{ik,j}) - (\Gamma^*_{ki,j} - \Gamma^*_{ik,j})\}\\
&\quad + \frac{1}{3}\{(\Gamma_{ij,k} - \Gamma_{ji,k}) - (\Gamma^*_{ij,k} - \Gamma^*_{ji,k})\}\\
&\quad + \frac{1}{3}\{(\Gamma_{jk,i} - \Gamma_{kj,i}) - (\Gamma^*_{jk,i} - \Gamma^*_{kj,i})\}\\
&= \frac{1}{3}(\Gamma_{ij,k} + \Gamma_{ji,k} + \Gamma_{jk,i} + \Gamma_{kj,i} + \Gamma_{ki,j} + \Gamma_{ik,j})\\
&\quad + \frac{1}{6}(\Gamma^*_{ij,k} + \Gamma^*_{ji,k} + \Gamma^*_{jk,i} + \Gamma^*_{kj,i} + \Gamma^*_{ki,j} + \Gamma^*_{ik,j}) - \Gamma_{ki,j} - \Gamma^*_{kj,i}.
\end{aligned}
$$

This implies (9). Now suppose that both ∇ and ∇^* are torsion-free. Then we have

$$
\begin{aligned}
\Gamma^*_{ij,k} - \Gamma_{ij,k} &= \Gamma^*_{ik,j} - \Gamma_{ik,j} \quad \text{(from (4))}\\
&= \Gamma^*_{jk,i} - \Gamma_{jk,i} \quad \text{(from (5))}.
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
-W&(\partial_i\partial_j\|\partial_k)\\
&= \frac{1}{3}(\Gamma_{ij,k} + \Gamma_{ji,k} + \Gamma_{jk,i} + \Gamma_{kj,i} + \Gamma_{ki,j} + \Gamma_{ik,j})\\
&\quad + \frac{1}{6}(\Gamma^*_{ij,k} + \Gamma^*_{ji,k} + \Gamma^*_{jk,i} + \Gamma^*_{kj,i} + \Gamma^*_{ki,j} + \Gamma^*_{ik,j}) - \Gamma_{ki,j} - \Gamma^*_{kj,i}\\
&= \Gamma_{ij,k} + \frac{1}{3}(\Gamma^*_{ij,k} - \Gamma_{ij,k}) + \frac{1}{3}(\Gamma^*_{ki,j} - \Gamma_{ki,j}) - \frac{2}{3}(\Gamma^*_{jk,i} - \Gamma_{jk,i})\\
&= \Gamma_{ij,k} + \frac{1}{3}(\Gamma^*_{jk,i} - \Gamma_{jk,i}) + \frac{1}{3}(\Gamma^*_{kj,i} - \Gamma_{kj,i}) - \frac{2}{3}(\Gamma^*_{jk,i} - \Gamma_{jk,i})\\
&= \Gamma_{ij,k}.
\end{aligned}
$$

This implies

$$
-W(\partial_i\partial_j\|\partial_k) = \Gamma_{ij,k}.
$$

Thus, if ∇ and ∇^* are torsion-free, then the connection induced from the contrast function $W(p\|q)$ coincides with the original connection ∇. $\qquad\square$

Let ∇ be any affine connection (possibly with torsion) on a Riemannian manifold (M, g). Let Φ be the map which associates to the affine connection ∇ the torsion-free affine connection induced from the contrast function $W(p\|q)$ defined in terms of ∇^*-geodesics where ∇^* is the dual connection of ∇. Then Φ is a map from the space of all affine connections of M to the subspace of torsion-free affine connections. If both ∇ and ∇^* are torsion-free, then $\Phi(\nabla)$ coincides with ∇. If ∇ preserves the Riemannian metric g (i.e., $\nabla^* = \nabla$), then $\Phi(\nabla)$ coincides with the Levi-Civita connection of (M, g).

By Theorem 4.3, if (M, g, ∇, ∇^*) is a torsion-free statistical manifold satisfying the condition (S) and $(S)^*$, then $W_c(p\|q) \equiv W(p\|q)$ and $W_{c^*}^*(p\|q) \equiv W^*(p\|q)$. In general the two-point function $W_c(p\|q) \equiv W(p\|q)$ defined on any torsion-free statistical manifold satisfying the condition (S) and $(S)^*$ is not symmetric: $W(q\|p) \neq W(p\|q)$. For more general torsion-free statistical manifolds, the function $W(\,\cdot\,\|q)$ might no longer be the potential function whose trajectories are ∇^*-geodesics emanating from q. However, the functions $W(p\|q)$ is defined without condition (S) and it is natural to ask what the relation of $W(p\|q)$ and $W(q\|p)$ is. The following Theorem 5.4 is not a satisfactory result but a proposal of a problem (see a remark after the proof of Theorem 5.4).

THEOREM 5.4. *For any torsion-free statistical manifold (M, g, ∇, ∇^*), the two point functions $W^*(p\|q)$ and $W(q\|p)$ (resp. $W(p\|q)$ and $W^*(q\|p)$) are contrast functions which induce the connection ∇^* (resp. ∇) on (M, g).*

If moreover (M, g, ∇, ∇^*) satisfies the condition $(S)^*$ (resp. (S)), then there exists a function f (resp. f^*) satisfying the condition

$$f(0) = 0, \;\; f'(0) = 1 \;\;\; (resp. \; f^*(0) = 0, \;\; (f^*)'(0) = 1)$$

such that

$$W(q\|p) = f(W^*(p\|q)) \;\;\; (resp. \; W^*(q\|p) = f^*(W(p\|q)))$$

holds.

Proof. Put $W'(p\|q) := W(q\|p)$. Then $W'(p\|q)$ is a contrast function and as ∇ is torsion-free, repeating the argument in the proof of Theorem 5.3 implies that the connection induced by $W'(p\|q)$ is

$$-W'(\partial_i \partial_j \| \partial_k) = \Gamma_{ij,k}^*,$$

i.e., the connection ∇^* dual to ∇ with respect to the metric g (moreover by computing $W'(\partial_i \partial_j \partial_k \| \partial_l)$, one gets the curvature tensor of the connection ∇^* ([E])). Indeed, interchanging p and q in (7) and applying $(\partial_i)_p (\partial_j)_p (\partial_k)_q$ yields the result:

$$
\begin{aligned}
-W'(\partial_i \partial_j \| \partial_k) &= \partial_i g_{kj} + \partial_j g_{ki} - \frac{2}{3}(\partial_k g_{ij} + \partial_i g_{jk} + \partial_j g_{kj}) \\
&\quad + \frac{1}{6}(\Gamma^*_{ij,k} + \Gamma^*_{ji,k} + \Gamma^*_{jk,i} + \Gamma_{kj,i} + \Gamma^*_{ki,j} + \Gamma^*_{ik,j}) \\
&= W(\partial_i \partial_j \| \partial_k) + (\partial_i g_{kj} + \partial_j g_{ki} - \partial_k g_{ij}) \\
&= -\Gamma_{ij,k} + (\Gamma_{ik,j} + \Gamma^*_{ij,k} + \Gamma_{jk,i} + \Gamma^*_{ji,k} - \Gamma_{ki,j} - \Gamma^*_{kj,i}) \\
&= \Gamma^*_{ij,k}.
\end{aligned}
$$

On the other hand, by interchanging the role of two connections ∇ and ∇^* in Definition 1.1, we get the quantity $W^*(p\|q)$ defined by the integral (1*). Now we compare the Taylor expansion of $W(p\|q)$ and $W^*(p\|q)$ along the diagonal. It follows from the definition of these contrast functions that replacing the coefficients of the Taylor expansion of $W(p\|q)$ along the diagonal by the corresponding dual quantities yields the Taylor expansion of $W^*(p\|q)$.

Thus we get two contrast functions $W'(p\|q)$ and $W^*(p\|q)$ which induce the same Riemannian metric g, the same dual connection ∇^*.

We here examine the relationship between the condition $(S)^*$ and Theorem 5.4. Consider a torsion-free statistical manifold (M, g, ∇, ∇^*) (not necessarily satisfying the condition $(S)^*$) a family of the level hypersurfaces of the function $W'(\cdot \| q) = W(q \| \cdot)$. Let S be one of such hypersurfaces. For a point $u \in S$ we define

$$
L_u := \left\{ x \in M \; ; \; W'(x\|u) = \min_{v \in S} W'(x\|v) \right\}.
$$

In [E, Section 4] Eguchi proved that (locally) $\{L_u\}_{u \in S}$ is a foliation of M with 1-dimensional leaves such that

(i) each leaf L_u is orthogonal to S at u, and

(ii) the second fundamental form w.r.t. ∇ of L_u at u vanishes.

Therefore, the family of all curves which are orthogonal to the level hypersurfaces of $W'(\cdot \| q)$ are all ∇-geodesics emanating from the origin q (with a suitable choice of the parameter).

We note that the result in [E, Section 4] is complementary to our Proposition 4.2, Theorem 4.3, Corollary 4.4 and Corollary 4.6.

Suppose now that the statistical manifold (M, g, ∇, ∇^*) satisfies the condition $(S)^*$. It then follows from the above argument and Corollary 4.6 that the level hypersurfaces of $W'(\,\cdot\,\|q)$ and $W^*(\,\cdot\,\|q)$ coincide as families and in particular the gradient flow of these functions are identical as a family of curves. This implies that there exists a function $f : [0, K) \to \mathbb{R}_{\geq 0}$ such that $W'(p\|q) = f(W^*(p\|q))$, where K is a positive constant (possibly ∞). □

§6. Hooke's law in dually flat statistical manifolds

In this section we show that the two point function $W_c(p\|q) \equiv W(p\|q)$ defined on torsion-free statistical manifolds generalizes the concept of the canonical divergence originally defined on dually flat statistical manifolds. Suppose that, for a statistical manifold (M, g, ∇, ∇^*), both ∇ and ∇^* are torsion-free. Then it is clear that ∇ is flat iff so is ∇^*.

DEFINITION 6.1. A statistical manifold (M, g, ∇, ∇^*) is *dually flat* iff both ∇ and ∇^* are torsion-free and flat.

For a dually flat statistical manifold (M, g, ∇, ∇^*), let (θ^i) (resp. (η_i)) be the ∇-affine (resp. ∇^*-affine) coordinates such (θ^i) and (η_i) are dual with respect to the Riemannian metric g, i.e.,

$$g(\partial_i, \partial^j) = \delta_i^j \quad \left(\partial_i = \frac{\partial}{\partial \theta^i}, \ \partial^j = \frac{\partial}{\partial \eta_j}\right).$$

Let ψ (resp. ϕ) be the potential in the *Legendre transform* connecting (θ^i) and (η_j). This means that

$$(10) \qquad \partial_i \psi = \eta_i, \quad \partial^i \phi = \theta^i, \quad \psi + \phi - \theta^i \eta_i = 0.$$

DEFINITION 6.2. Let (M, g, ∇, ∇^*) be a dually flat statistical manifold and (θ^i) and (η_i) are as above. For points p and q in the domain where both (θ^i) and (η_i) are defined, the ∇-*divergence* $D(p\|q)$ is defined by

$$D(p\|q) := \psi(p) + \phi(q) - \theta^i(p)\eta_i(q).$$

Interchanging the role of ∇ and ∇^* yields the definition of the ∇^*-divergence $D^*(p\|q)$. The Legendre transform implies

$$(11) \qquad D(p\|q) = D^*(q\|p).$$

It is well known that, in the exponential family, the (-1)-divergence coincides with the relative entropy (= the Kullback divergence) (see, for instance, [A] or [A-N]).

Let (M, g, ∇, ∇^*) is a dually flat statistical manifold.

THEOREM 6.3. *For dually flat statistical manifolds, the two point function $W(p\|q)$ coincides with the canonical divergence $D(p\|q)$.*

Proof. Let (θ^i) (resp. (η_i)) be an ∇-affine (resp. ∇^*-affine) system of local coordinates such that (θ^i) and (η_i) are dual to each other with respect to the metric g, i.e., $\langle \partial_i, \partial^j \rangle = \delta_i^j$, where $\partial_i = \partial/\partial\theta^i$ and $\partial^j = \partial/\partial\eta_j$. Assume that these coordinates are defined in a commomn domain of M. In terms of these coordinates, a ∇-geodesics $c(t)$ and ∇^*-geodesics $c_t^*(s)$ (which appear in (1)) are written as

$$\theta(c(t)) = \theta(q) + t(\theta(p) - \theta(q)), \quad (t \in [0,1]),$$
$$\eta(c_t^*(s)) = \eta(c(t)) + s(\eta(q) - \eta(c(t))), \quad (s \in [0,1]).$$

Therefore we have

(12)
$$W(p\|q) = -\int_0^1 dt \left\langle (\theta^i(p) - \theta^i(q))(\partial_i)_{c(t)}, (\eta_j(q) - \eta_j(c(t)))(\partial^j)_{c(t)} \right\rangle_{c(t)}$$
$$= \int_0^1 dt \, (\theta^i(p) - \theta^i(q))(\eta_i(c(t)) - \eta_i(q)) \quad \text{(because } \langle \partial_i, \partial^j \rangle = \delta_i^j)$$
$$= (\theta^i(q) - \theta^i(p))\eta_i(q) + \int_0^1 dt \, \eta_i(c(t))(\theta^i(p) - \theta^i(q)).$$

Let ψ be the potential in the Legendre transform defined in (10). Then we have $\partial_i\psi = \eta_i$ and so

$$\frac{d}{dt}\psi(c(t)) = \partial_i\psi(c(t))\frac{d}{dt}\theta^i(c(t)) = \eta_i(c(t))(\theta^i(p) - \theta^i(q)).$$

Substituting this into (12) and applying the Legendre transform $\psi + \phi - \theta^i\eta_i = 0$, we have

(13)
$$W(p\|q) = (\theta^i(q) - \theta^i(p))\eta_i(q) + \psi(p) - \psi(q)$$
$$= (\psi(q) + \phi(q)) - \theta^i(p)\eta_i(q) + \psi(p) - \psi(q)$$
$$= \psi(p) + \phi(q) - \theta^i(p)\eta_i(q)$$
$$= D(p\|q) \quad \text{(from Definition 6.2)}.$$

□

The divergence on dually flat statistical manifolds is originally defined in terms of mutually dual affine coordinates and the potentials of their Legendre transform (10). On the other hand, (13) means that the divergence is equal to the integral (1) which is defined completely in terms of geodesics (without using dually flatness). Therefore it is natural to think that the integral (1) generalizes the concept of the divergence in general statistical manifolds. More generally, (1) is defined, provided a Riemannian metric g and any two affine connections ∇ and ∇^* (not necessarily dual) are given.

Now let examine the validity of Theroem 4.3 for dually flat statistical manifolds. Suppose that a curve $c(t)$ ($c(0) = q$ and $c(1) = p$) is not necessarily a ∇-geodesic and let $c_t^*(s)$ ($c_t^*(0) = c(t)$, $c_t^*(1) = q$) be the ∇^*-geodesic. If (g, ∇, ∇^*) is a dually flat structure, then ∇^*-geodesics emanating from q coincide with the image of integral curves of $\operatorname{grad} D(\,\cdot\,\|q)$. Therefore we have

$$-\int_0^1 dt \left[g_{ij}(c(t)) \frac{d}{dt} c^i(t) \frac{d}{ds} c_t^{*j}(0) \right] = \int_0^1 dt \left\langle \frac{d}{dt} c(t), \operatorname{grad} D(\,\cdot\,\|q)(c(t)) \right\rangle_{c(t)}$$

$$= \int_0^1 dt \frac{d}{dt} D(c(t)\|q) = D(p\|q),$$

which implies, in simply connected dually flat statistical manifolds, the integral (2) (defined in general for any curve $c(t)$ such that $c(0) = q$ and $c(1) = p$) depends only on end points q and p and is equal to the canonical divergence $D(p\|q)$.

References

[A]   S. Amari, Differential-Geometrical Methods in Statistics, Springer Lecture Notes in Statistics **28**, 1985.

[A-N] S. Amari and H. Nagaoka, Method of Information Geometry, to appear, AMS.

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

[N-A] H. Nagaoka and S. Amari, *Differential geometry of smooth families of probability distributions*, METR, **82-7**, University of Tokyo.

Masayuki Henmi
Sūgaku Kōbō
3-2 Surugadai
Kanda, Chiyoda-ku
Tokyo, 101-0062
Japan
`m96005b@math.nagoya-u.ac.jp`

Ryoichi Kobayashi
Graduate School of Mathematices
Nagoya University
Chikusa-ku
Nagoya, 464-8602
Japan
ryoichi@math.nagoya-u.ac.jp