

Information geometry in a global setting

Atsuhide MORI

(Received January 6, 2017)

(Revised June 8, 2018)

ABSTRACT. We begin a global study of information geometry. In this article, we describe the geometry of normal distributions by means of positive and negative contact structures associated to the suspension Anosov flows on Sol^3 -manifolds.

1. Introduction

The information geometry is basically a local study of a parameter space of a family of probability distributions by means of differential geometry (see [1]). We begin a global study of this object. In this article, we restrict ourselves to the case of normal distributions. Then the parameter space is the upper half-plane $\mathbb{H} = \{(m, s) \in \mathbb{R}^2 \mid s > 0\}$ with a certain non-Kähler metric and a non-metric connection, where m denotes the mean and s the standard deviation. These local structures are derived from a certain global function D on $\mathbb{H} \times \mathbb{H}$, which is fundamental in information theory. On the other hand, we know that the product $\mathbb{H} \times \mathbb{H}$ parametrizes abelian surfaces in moduli theory. Precisely, the quotient of the product $\mathbb{H} \times \mathbb{H}$ under the action of a Hilbert modular group is a non-compact singular complex surface which parametrizes the isomorphism classes of complex abelian surfaces with a real multiplication structure. Its topology was studied by Hirzebruch [5]. He described each end of a Hilbert modular surface as a cusp, namely the neighborhood of ∞ of the quotient M of $\mathbb{H} \times \mathbb{H}$ under a certain action of a semi-direct product $\mathbb{Z} \ltimes \mathbb{Z}^2$ of lattices of \mathbb{R} and \mathbb{R}^2 . From the contact topological point of view, $M \cup \{\infty\}$ is the cone with positive and negative symplectic structures whose base N is a Sol^3 -manifold with the positive and negative contact structures associated to the suspension Anosov flow (see [7] and §4 below for the precise meaning). In this article, we reorganize the information geometry of normal distributions so that it fits with the contact topology of a cusp. The result of this article can be summarized as follows.

- i) (The model in §3.) We describe the information geometry of \mathbb{H} by introducing a self-correspondence $F \subset \mathbb{H} \times \mathbb{H}$ which identifies the

2010 *Mathematics Subject Classification.* 57R17, 57R30, 62B10.

Key words and phrases. Information geometry, contact structure, foliation.

convolution operation for the probability densities in each factor of $\mathbb{H} \times \mathbb{H}$ to the operation of Bayesian learning for those in the other factor.

- ii) (The propositions in §5.) The action of $\mathbb{Z} \times \mathbb{Z}^2$ preserves the metric and the connection on each factor of $\mathbb{H} \times \mathbb{H}$. The sum (resp. the difference) of the natural area forms descends to the positive (resp. negative) symplectic structure on the quotient M . The surface F is a Lagrangian correspondence which descends to a densely immersed Lagrangian submanifold $L(\subset N) \subset M$ with respect to the negative symplectic structure. Further L is decomposed into Legendrian submanifolds of N with respect to the positive contact structure.
- iii) (Theorem 1 in §5.) We can take a contact Hamiltonian flow on the universal cover $\tilde{N}(\subset \mathbb{H} \times \mathbb{H})$ of the Sol^3 -manifold $N(\subset M)$ such that
 - a) the surface $F \subset \tilde{N}(\subset \mathbb{H} \times \mathbb{H})$ is an invariant submanifold,
 - b) the projection of the induced flow on F to each factor of $\mathbb{H} \times \mathbb{H}$ is tangent to a foliation whose leaves are geodesics for the non-metric connection in the information geometry, and
 - c) the iterations of convolutions along the foliation of the first factor and a certain Bayesian learning process along the foliation of the second factor are identified via the correspondence F .

Note that we could take a similar contact flow just partially on the quotient. Indeed the contact Hamiltonian function for the flow in c) is the inverse coefficient of variance m/s on the first factor which is not preserved under the \mathbb{Z}^2 -action (but preserved under the \mathbb{Z} -action). We raise open problems concerning the relation between abelian varieties and pairs of normal distributions (§6).

2. Information geometry

2.1. Smooth parametric statistics. In the smooth setting of parametric statistics, we consider an open set $U \subset \mathbb{R}^n$ and a positive function $p(x, X) > 0$ on $\mathbb{R} \times U$ ($X = (X^1, \dots, X^n)$) such that the normality condition

$$\int_{-\infty}^{\infty} p(x, X) dx = 1 \quad (\forall X \in U)$$

is satisfied. Here we consider that X is a parameter of a random variable x with probability density $p_X(x) = p(x, X)$. We notice that the absolute entropy

$$-\int_{-\infty}^{\infty} p_X(x) \log p_X(x) dx$$

is possibly negative. That is why, in information theory, the relative entropy

$$D(X, Y) = - \int_{-\infty}^{\infty} p_X(x) \log \frac{p_Y(x)}{p_X(x)} dx \left(\geq -\log \int_{-\infty}^{\infty} p_X(x) \frac{p_Y(x)}{p_X(x)} dx = 0 \right)$$

for $(X, Y) \in U \times U$ is fundamental. It defines a separating premetric¹ on U , which is called the Kullback-Leibler divergence in information theory. Using the tensor notation only in this section, we define the Fisher metric $g = \sum_{i,j} g_{ij} dX^i dX^j$ by the quadratic approximation

$$D(X, X + \Delta X) \approx \frac{1}{2!} \sum_{i,j} g_{ij} \Delta X^i \Delta X^j (\approx D(X + \Delta X, X))$$

and the e(xponential or Efron)-connection ∇ by the cubic approximation

$$\begin{aligned} D(X + \Delta X, X) - D(X, X + \Delta X) \\ \approx \frac{1}{3!} \sum_{i,j,k} (\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij} - 2\Gamma_{ij,k}) \Delta X^i \Delta X^j \Delta X^k, \end{aligned}$$

where

$$\begin{aligned} |\Delta X| &= \sqrt{(\Delta X^1)^2 + \dots + (\Delta X^n)^2} \ll 1, \\ \nabla_{\partial_i} \partial_j &= \sum_k \Gamma_{ij}^k \partial_k, \quad \Gamma_{ij,k} = \sum_l g_{kl} \Gamma_{ij}^l, \quad g_{ij} = g_{ji}, \quad \text{and} \\ \partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij} - 2\Gamma_{ij,k} &= \partial_j g_{ki} + \partial_k g_{ij} - \partial_i g_{jk} - 2\Gamma_{jk,i}. \end{aligned}$$

The coefficients of the e-connection all vanish for an exponential family

$$p_X(x) = \exp \left(\sum_{i=1}^n X^i f_i(x) + f_0(x) - v(X) \right),$$

where $f_i(x)$ are preferable functions and $v(X)$ the normalizer.

The space of torsion-free connections on U is an affine space containing the Levi-Civita connection, and the underlying vector space consists of symmetric bilinear map from $\mathfrak{X}(U) \times \mathfrak{X}(U)$ to $\mathfrak{X}(U)$ ($= \{\text{vector fields on } U\}$). We may regard any torsion-free connection as a vector starting at the Levi-Civita connection. Then we define the α -connection as the multiplication of the vector going to the above e-connection ∇ by a scalar $\alpha \in \mathbb{R}$. This draws

¹A premetric or prametric on a set U is a non-negative function on $U \times U$ which vanishes along the diagonal set. If it is positive elsewhere, we say that it is separating. The value of a premetric is called a distance.

the line through ∇ and the Levi-Civita connection unless they coincide, i.e., unless all of the equations $\Gamma_{ij,k} = \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij})$ hold. Particularly, the Levi-Civita connection is the 0-connection, and ∇ is the 1-connection. Intuitively, the difference between these connections linearizes the asymmetry of $D(X, Y)$ along the diagonal set of $U \times U$. On the other hand, the (-1)-connection is called the m(ixture)-connection since its coefficients vanish for any mixture $p_X(x) = \sum_{l=0}^n X^l q_l(x)$ of probability densities $q_0(x), \dots, q_n(x)$, where $X = (X^1, \dots, X^n)$ and

$$U = \{X^0 := 1 - X^1 - \dots - X^n > 0, X_1 > 0, \dots, X_n > 0\} \subset \mathbb{R}^n.$$

It is remarkable that an α -connection is flat (meaning that its coefficients vanish with respect to some coordinate system on U) if and only if the $(-\alpha)$ -connection is flat. In particular, the m-connection for the exponential family vanishes with respect to a certain coordinate system. (The expected values of $f_1(x), \dots, f_n(x)$ provide such coordinates.) See the book [1] for this fact and other fundamental results in the information geometry.

2.2. Normal distributions. Hereafter we restrict ourselves to the case of the normal distributions, namely we consider the probability density

$$p_{(m,s)}(x) = \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{(x-m)^2}{2s^2}\right)$$

for $(m, s) \in \mathbb{H} = \mathbb{R} \times \mathbb{R}_{>0}$ and put $U = \mathbb{H}$ and $(X^1, X^2) = (m, s)$. Then the Kullback-Leibler divergence $D(X, Y)$ is expressed as

$$\begin{aligned} D((m, s), (m', s')) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{(x-m)^2}{2s^2}\right) \left\{ -\frac{(x-m)^2}{2s^2} + \frac{(x-m')^2}{2(s')^2} - \log \frac{s}{s'} \right\} dx \\ &= -\frac{1}{2} + \frac{s^2}{2(s')^2} + 0 + \frac{(m' - m)^2}{2(s')^2} - \log \frac{s}{s'}. \end{aligned}$$

From this we can see that the Fisher metric g and the e-connection ∇ satisfy

$$g = \frac{dm^2 + 2 ds^2}{s^2}, \quad \Gamma_{12,1} = \Gamma_{21,1} = \frac{-2}{s^3}, \quad \text{and} \quad \Gamma_{22,2} = \frac{-6}{s^3},$$

and the other coefficients $\Gamma_{ij,k}$ vanish. Writing the probability density as

$$p_{(m,s)}(x) = \exp\left(\frac{m}{s^2}x + \frac{-1}{2s^2}x^2 - \frac{m^2}{2s^2} - \log(\sqrt{2\pi s})\right),$$

we see that the normal distributions form an exponential family parametrized by the coefficients $\left(\frac{m}{s^2}, \frac{-1}{2s^2}\right)$ of (x, x^2) . The e-connection vanishes with respect to this parametrization. Thus a geodesic for the e-connection is one of the following: a horizontal line with affine parameter m , a vertical half-line with s^{-2} , or an upper semi-parabola $s = \sqrt{am + b}$ ($a, b \in \mathbb{R}$) with s^{-2} . On the other hand, the m-connection vanishes with respect to the other coordinate system $(m, m^2 + s^2)$ consisting of expected values of (x, x^2) . This implies that a geodesic for the m-connection is either a vertical half-line with affine parameter s^2 or an upper semi-circle with m . Whereas its image appears as that of a geodesic for the Poincaré metric $\frac{dm^2 + 1 ds^2}{s^2}$ ($\neq g$), the affine parametrization is different.

3. The basic model

Let (m, s, M, S) denote the coordinate system of $\mathbb{H} \times \mathbb{H}$. Take the correspondence

$$F: M = -\frac{m}{s^2} \quad \text{and} \quad S = \frac{1}{s}$$

$$\left(\Leftrightarrow \frac{m}{s} + \frac{M}{S} = 0 \quad \text{and} \quad sS = 1 \right)$$

and regard it as a submanifold of $\mathbb{H} \times \mathbb{H}$. For any point (m, s, M, S) in $\mathbb{H} \times \mathbb{H}$, we put

$$f(s, m, S, M) = \frac{\left(\frac{M}{S} + \exp(-h)\frac{m}{s}\right)^2 + \exp(-2h) - 1 + 2h}{2},$$

where $h = -\log(sS)$.

PROPOSITION 1. *Suppose that under the correspondence F a point (m', s') of the first factor of $\mathbb{H} \times \mathbb{H}$ is identified with the point (M, S) of the second factor, i.e., $(m', s', M, S) \in F$. Then the function $f(m, s, M, S)$ presents the Kullback-Leibler divergence $D((m, s), (m', s'))$.*

PROOF.

$$2f\left(s, m, \frac{1}{s'}, \frac{-m'}{(s')^2}\right) = \left(\frac{-m'}{s'} + \frac{s}{s'} \cdot \frac{m}{s}\right)^2 + \frac{s^2}{(s')^2} - 1 - 2 \log \frac{s}{s'}$$

$$= 2D((m, s), (m', s')) \quad \square$$

We define the product $(m, s) * (m', s')$ on the first factor of $\mathbb{H} \times \mathbb{H}$ by putting

$$(m, s) * (m', s') = (m + m', \sqrt{s^2 + (s')^2}).$$

It represents the convolution of probability density functions

$$P_{(m,s)*(m',s')}(x) = (P_{(m,s)} * P_{(m',s')})(x).$$

We define another product $(M, S) \cdot (M', S')$ on the second factor by putting

$$(M, S) \cdot (M', S') = \left(\frac{M(S')^2 + (M')S^2}{S^2 + (S')^2}, \sqrt{\frac{S^2(S')^2}{S^2 + (S')^2}} \right).$$

This represents the normalized pointwise product

$$P_{(M,S)\cdot(M',S')} = P_{(M,S)}(x) \cdot P_{(M',S')}(x) / \int_{-\infty}^{\infty} P_{(M,S)}(x) \cdot P_{(M',S')}(x) dx,$$

which can be interpreted as a Bayesian learning as follows: In Bayesian statistics, the observation is an event, and the parameter x of the population distribution is a random variable. Since the population is unknown, the distribution of x is also unknown. Thus we have to fix a prior probability density $p(x)$ before the observation. Since the observation is an event, it has the probability r . It also has the conditional probability for the population at each x . This is not a probability density of x , but the non-negative function of x called the likelihood. Suppose that we obtain a probability density $q(x)$ from the likelihood by multiplying with it a positive constant. Then we obtain the posterior probability density $p'(x) = p(x)q(x)/r$ from Bayes' rule. The posterior probability density $p'(x)$ presents the updated prior probability density learned from the present observation. We will use $p'(x) = p(x) \cdot q(x)$ as the prior probability density in the next observation. Although $p(x)$ is arbitrary and subjective, we can update it to $p(x) \cdot q(x)$ by the observation. We call this procedure a Bayesian learning.

PROPOSITION 2. *The above correspondence F on $\mathbb{H} \times \mathbb{H}$ identifies the product $*$ on the first factor (resp. the second factor) with the product \cdot on the second factor (resp. the first factor).*

PROOF. Let (m, s, M, S) and (m', s', M', S') be two points on F . Then we have

$$\sqrt{\frac{S^2(S')^2}{S^2 + (S')^2}} = \frac{1}{\sqrt{s^2 + (s')^2}}$$

and

$$\frac{M(S')^2 + (M')S^2}{S^2 + (S')^2} = \frac{-\frac{m}{s^2} \frac{1}{(s')^2} - \frac{m'}{(s')^2} \frac{1}{s^2}}{\frac{1}{s^2} + \frac{1}{(s')^2}} = \frac{-(m + m')}{s^2 + (s')^2},$$

which prove the identification between $*$ on the first factor and \cdot on the second factor. They also prove the other identification because F is invariant under the interchange of the order of the product $\mathbb{H} \times \mathbb{H}$. \square

Thus the correspondence F defines a Fourier-like transformation on \mathbb{H} , hence the notation.

4. Contact/foliation topology of Sol^3 -manifolds

4.1. Topology of contact structures. Contact geometry is soft enough to interest topologists since Gray’s stability [4] says that any smooth homotopy of contact structures on a closed manifold (i.e. a compact manifold without boundary) can be traced by an isotopy of the ambient manifold. However, the isotopy class of a contact structure is not determined by the homotopy class of almost contact structures containing it. Among the isotopy classes, geometrically interesting “tight” ones e.g., the examples in the next subsection are known to be rather exceptional. See [2] and [8] to see their importance in current contact topology.

On the other hand, contact geometry and symplectic geometry are deeply related. To see this, let ξ be a co-oriented positive contact structure on an oriented $(2n + 1)$ -manifold N . Then we can take a 1-form α on N which satisfies $\xi = \ker \alpha$ and $\alpha \wedge (d\alpha)^n > 0$. Here we consider that $-\alpha$ does not define ξ in that co-orientation while $e^\varphi \alpha$ does ($\forall \varphi \in C^\infty(N)$). Note that $e^\varphi \alpha$ satisfies $(e^\varphi \alpha) \wedge (d(e^\varphi \alpha))^n = e^{(n+1)\varphi} \alpha \wedge (d\alpha)^n > 0$. Let us temporarily fix a contact form α presenting ξ . Then the cylinder $M = \mathbb{R} \times N$ carries the symplectic form $d(e^z \alpha)$, where z is the coordinate on \mathbb{R} . We call this symplectic manifold the symplectization of (N, α) , and $\mathbb{R} \times \{*\}$ its \mathbb{R} -fiber. A section of the \mathbb{R} -fibration can be considered as the graph $z = \varphi$ of a function $\varphi \in C^\infty(N)$. Then considering $z' = z - \varphi$ as a new coordinate, we get another description of the same \mathbb{R} -fibration where the section $z = \varphi$ becomes the zero-section. Then from $d(e^{z'} \alpha) = d(e^{z'} (e^\varphi \alpha))$ we see that another contact form $e^\varphi \alpha$ defines the same symplectization up to diffeomorphism preserving the \mathbb{R} -fibration (but changing the zero section). For a fixed contact form α and a function H on N , we can define the contact Hamiltonian vector field Y for H as the well-defined push-forward of the usual Hamiltonian vector field \tilde{Y} for the func-

tion $\tilde{H} = e^z H$ on the symplectization of (N, α) to the base space N of the \mathbb{R} -fibration. Then we have $\alpha(Y) = H$. Note that Y is contact, i.e., it preserves the contact structure $\ker \alpha$ since

$$\mathcal{L}_{\tilde{Y}} e^z \alpha = \iota_{\tilde{Y}} d(e^z \alpha) + dt_{\tilde{Y}}(e^z \alpha) = -d(e^z H) + d(e^z H) = 0.$$

From the non-integrability of ξ , we see that a contact vector field Y' tangent to ξ (i.e., $\alpha(Y') = 0$) must be zero. This implies that any contact vector field is a contact Hamiltonian vector field with $H = \alpha(Y)$. If we fix \tilde{H} and change the section of the \mathbb{R} -fibration of M , we obtain another pair of contact form $e^\varphi \alpha$ and contact Hamiltonian function $e^{-\varphi} H$ on the base manifold N which defines the same contact vector field Y . Thus we can say that a contact form fixes an isomorphism between the space of functions and the space of contact vector fields, while we need no functions to characterize contact Hamiltonian system.

4.2. On Sol^3 -manifolds. Honda [6] completed the isotopy classification of contact structures on (the total spaces of) T^2 -bundles over the circle. On the other hand, given an element $A \in SL(2, \mathbb{Z})$ with hyperbolicity $\text{tr } A > 2$, the T^2 -bundle with monodromy A possesses a canonical pair (ξ_+, ξ_-) of positive and negative contact structures associated to the suspension Anosov flow. Several authors studied the isotopy classes of these contact structures in terms of Honda's classification. Kasuya [7], without appealing to the classification, related the contact structures ξ_\pm to Hirzebruch's construction [5] of Hilbert modular cusp. Hereafter we use relevant part of Kasuya's description of ξ_\pm .

It is well-known that any element A of $SL(2, \mathbb{Z})$ with $\text{tr } A > 2$ is conjugate to a positive word of $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ which uses both letters. Note that the conjugacy class is not determined by the trace unlike in the case of $SL(2, \mathbb{R})$. Then since $\text{tr}(A^{-1}) > 2$, we can take a map

$$\mathbb{Z}/r\mathbb{Z} \ni k \mapsto b_k \in \mathbb{Z}_{\geq 2}$$

not identically 2 such that A^{-1} is conjugate to

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{b_{k+r-1}-2} \cdots \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{b_k-2}.$$

Then from the formula

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{-1} \left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{b_k-2} \right) \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} b_k & 1 \\ -1 & 0 \end{bmatrix},$$

we see that A^{-1} is conjugate to

$$\begin{bmatrix} s_k & t_k \\ u_k & v_k \end{bmatrix} := \begin{bmatrix} b_{k+r-1} & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} b_{k+r-2} & 1 \\ -1 & 0 \end{bmatrix} \cdots \begin{bmatrix} b_k & 1 \\ -1 & 0 \end{bmatrix}.$$

One may start with a given recurring sequence $b_k \geq 2$ other than the constant sequence $b_k = 2$ since the trace of the above composition map is greater than 2 unless identically $b_k = 2$. On the other hand the continued fractions

$$w_k = b_k - \frac{1}{b_{k+1} - \frac{1}{b_{k+2} + \cdots}} \quad (k \in \mathbb{Z}/r\mathbb{Z})$$

and the sum $c = \log(w_1) + \cdots + \log(w_r)$ of their logarithms satisfy

$$\begin{bmatrix} b_k & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ w_k \end{bmatrix} = \frac{1}{w_{k+1}} \begin{bmatrix} -1 \\ w_k \end{bmatrix}$$

and

$$\begin{bmatrix} s_k & t_k \\ u_k & v_k \end{bmatrix} \begin{bmatrix} -1 \\ w_k \end{bmatrix} = e^{-c} \begin{bmatrix} -1 \\ w_k \end{bmatrix}.$$

Putting $k = 1$, we see that the irrational number w_1 and its irrational conjugate \bar{w}_1 are the solutions of the quadratic equation

$$t_1 w_1^2 - (s_1 - v_1)w_1 - u_1 = 0$$

which satisfy $0 < \bar{w}_1 < 1 < w_1$. This implies the formula

$$\begin{bmatrix} s_1 & t_1 \\ u_1 & v_1 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ w_1 & \bar{w}_1 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ w_1 & \bar{w}_1 \end{bmatrix} \begin{bmatrix} e^{-c} & 0 \\ 0 & e^c \end{bmatrix}.$$

Now we take any $K > 0$ to fix the vectors

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ w_1 & \bar{w}_1 \end{bmatrix}^{-1} \begin{bmatrix} K \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ w_1 & \bar{w}_1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ K \end{bmatrix},$$

which satisfy

$$\begin{bmatrix} e^{-c} & 0 \\ 0 & e^c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = s_1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + u_1 \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

and

$$\begin{bmatrix} e^{-c} & 0 \\ 0 & e^c \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = t_1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + v_1 \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

We define the quotient manifold $M(x, y, c) = \mathbb{H} \times \mathbb{H} / \sim$ by the equivalences

$$\begin{aligned} (m, s, M, S) &\sim (e^c m, e^c s, e^{-c} s, e^{-c} S), \\ (m, s, M, S) &\sim (m + x_1, s, M + x_2, S), \end{aligned}$$

and

$$(m, s, M, S) \sim (m + y_1, s, M + y_2, S)$$

for any $(m, s, M, S) \in \mathbb{H} \times \mathbb{H}$. Note that the map

$$(m, s, M, S) \mapsto (e^c m, e^c s, e^{-c} s, e^{-c} S)$$

preserves the lattice on the mM -plane generated by $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Thus $M(x, y, c)$ is a 4-manifold diffeomorphic to a T^2 -bundle over the open annulus $S^1 \times \mathbb{R}$.

Next we take the function $h = -\log(sS)$ on $\mathbb{H} \times \mathbb{H} \subset \mathbb{C} \times \mathbb{C}$. It is strictly plurisubharmonic since the Hessian $\text{diag}\left(\frac{1}{4s^2}, \frac{1}{4S^2}\right)$ is clearly positive definite. Now we shift our ground from the complex structure J_{std} to the exact symplectic structure $d\lambda_+$ with fixed primitive

$$\lambda_+ = -J_{\text{std}}^* dh = \frac{dm}{s} + \frac{dM}{S} = \exp\left(\frac{h+t}{2}\right) dm + \exp\left(\frac{h-t}{2}\right) dM,$$

where $t = \log \frac{S}{s}$. The plurisubharmonicity can also be expressed as $dh(X) > 0$

by means of a Liouville vector field X . Here we choose X so that $\iota_X d\lambda_+ = \lambda_+$ holds. Note that this equation uniquely determines the vector field X since the non-degeneracy of $d\lambda_+$.

The flow foliation of X carries the holonomy invariant transverse contact structure which is defined by the 1-form λ_+ . Conversely the symplectic manifold $\mathbb{H} \times \mathbb{H}$ is the symplectization of the section $\{h=0\}$, where the flow lines of X are the \mathbb{R} -fibers with fiber coordinate $z = h/2$. The function h descends to the quotient $M(x, y, c)$, so that the 0-level set is a closed contact 3-manifold $N(x, y, c)$. Then $M(x, y, c)$ is the symplectization of $N(x, y, c)$. The function $t \pmod{2c\mathbb{Z}}$ also descends to $M(x, y, c)$, so that its restriction to the section $N(x, y, c)$ defines a T^2 -bundle projection to the circle $\mathbb{R}/2c\mathbb{Z}$. The T^2 -fiber is the quotient of the mM -plane by the lattice generated by x and y . The

monodromy map $\begin{bmatrix} e^c & 0 \\ 0 & e^{-c} \end{bmatrix}$ with respect to the fundamental basis of the mM -plane is written as $\begin{bmatrix} s_1 & t_1 \\ u_1 & v_1 \end{bmatrix}^{-1} \in SL(2, \mathbb{Z})$ with respect to the basis $\{x, y\}$.

This finally leads us to consider the contact structure ζ_+ on the T^2 -bundle with monodromy A instead of (the primitive λ_+ of) the symplectic structure $d\lambda_+$.

We can take the global frame (e_1, e_2, e_3) of $TN(x, y, c)$ by putting

$$e_1 = e^{-t/2} \partial_m, \quad e_2 = e^{t/2} \partial_M, \quad \text{and} \quad e_3 = 2\partial_t.$$

It satisfies the sol^3 -relations

$$[e_1, e_2] = 0, \quad [e_3, e_1] = -e_1, \quad \text{and} \quad [e_3, e_2] = e_2.$$

The dual coframe (e_1^*, e_2^*, e_3^*) satisfies the corresponding relations

$$de_3^* = 0, \quad de_1^* = -e_1^* \wedge e_3, \quad \text{and} \quad de_2^* = e_2^* \wedge e_3^*,$$

where

$$e_1^* = e^{t/2} dm, \quad e_2^* = e^{-t/2} dM, \quad \text{and} \quad e_3^* = dt/2.$$

We call $N(x, y, c)$ the Sol^3 -manifold associated to A . The suspension Anosov flow is the flow generated by e_3 . Since it expands e_1 and contracts e_2 , the 1-forms $\alpha_{\pm} = e_1^* \pm e_2^*$ define a pair of positive and negative contact structures ζ_{\pm} , which we call the bi-contact structure associated to the flow. (This ζ_+ coincides with the above one, hence we use the same notation.) The suspension Anosov flow drifts the positive and negative contact structures toward the unstable Anosov foliation defined by the equation $e_2^* = 0$. Further we have the negative symplectic form $d\lambda_-$ on $M(x, y, c)$ which is the symplectization of the negative contact structure ζ_- , where

$$\lambda_- = e^{h/2}(e_1^* - e_2^*) = \exp\left(\frac{h+t}{2}\right)dm - \exp\left(\frac{h-t}{2}\right)dM.$$

5. Results

Now we investigate the model in §3 in the light of the description in §4.

PROPOSITION 3. *The correspondence F and the function f are invariant under the monodromy map $(m, s, M, S) \mapsto (e^c m, e^c s, e^{-c} S, e^{-c} S)$.*

PROOF. The monodromy map preserves the inverse coefficients of variance $\frac{m}{s}$ and $\frac{M}{S}$, and the strictly plurisubharmonic function $h = -\log(sS)$. \square

PROPOSITION 4. *The symplectic structure*

$$d\lambda_+ = \frac{dm \wedge ds}{s^2} + \frac{dM \wedge dS}{S^2},$$

the sum of Fisher metrics

$$g = \frac{dm^2 + 2 ds^2}{s^2} + \frac{dM^2 + 2 dS^2}{S^2},$$

and the almost complex structure

$$J : \partial_m \mapsto \frac{1}{\sqrt{2}} \partial_s, \quad \partial_M \mapsto \frac{1}{\sqrt{2}} \partial_S$$

on $\mathbb{H} \times \mathbb{H}(\ni (m, s, M, S))$ satisfy $g(\cdot, \cdot) = \sqrt{2} d\lambda_+(\cdot, J\cdot)$ and descend to the quotient $M(x, y, c)$.

PROOF. We have

$$\sqrt{2} d\lambda_+\left(\partial_m, \frac{1}{\sqrt{2}} \partial_s\right) = \frac{1}{s^2}$$

and

$$\sqrt{2} d\lambda_+(\partial_s, -\sqrt{2} \partial_m) = \frac{2}{s^2}.$$

The rest is clear. \square

PROPOSITION 5. *The ‘‘Fourier’’ correspondence $F \subset (\mathbb{H} \times \mathbb{H}, d\lambda_+)$ is a smooth surface contained in the contact-type hypersurface $\tilde{N} = \{h = 0\}$ with positive contact structure $\ker \alpha_+$. It is the union of Legendrian lines*

$$\{e^{T/2}m + e^{-T/2}M = 0 \text{ and } t = T\} \quad (T \in \mathbb{R}).$$

These lines descend to the quotient $N(x, y, c)$ of \tilde{N} as Legendrian curves which form a dense immersion of the surface F .

PROOF. We see that

$$e^{t/2}m + e^{-t/2}M = 0 \quad \Leftrightarrow \quad \frac{m}{s} + \frac{M}{S} = 0.$$

Together with $sS = 1$ ($\Leftrightarrow h = 0$), this defines the surface F . The Legendrian lines for $T = T_0 + 2c$ and $T = T_0$ descend to the same Legendrian immersed curve on $(N(x, y, c), \ker \alpha_+)$, which is either closed or dense in the toral fiber $\{t = T\} \subset N(x, y, c)$ depending on whether the slope $-e^T$ is rational or irrational with respect to the basis $\{x, y\}$. \square

PROPOSITION 6. *For the negative symplectic structure*

$$d\lambda_- = \frac{dm \wedge ds}{s^2} - \frac{dM \wedge dS}{S^2}$$

on $\mathbb{H} \times \mathbb{H}$, the surface F is a Lagrangian correspondence.

PROOF. The tangent space of F is expressed as

$$TF = \text{span}\{s\hat{\partial}_m - S\hat{\partial}_M, m\hat{\partial}_m + s\hat{\partial}_s - M\hat{\partial}_M - S\hat{\partial}_S\}.$$

Then we have

$$\frac{dm \wedge ds(s\hat{\partial}_m, m\hat{\partial}_m + s\hat{\partial}_s)}{s^2} - \frac{dM \wedge dS(-S\hat{\partial}_M, -M\hat{\partial}_M - S\hat{\partial}_S)}{S^2} = 1 - 1 = 0. \quad \square$$

The correspondence F also satisfies the following interesting property. We call the image of a geodesic for the e-connection an e-geodesic.

PROPOSITION 7. *Any e-geodesic on the first factor of $\mathbb{H} \times \mathbb{H}$ corresponds to an e-geodesic on the second factor via F , and vice-versa.*

PROOF. (The image of) the e-geodesics on the first factor becomes straight if we replace the vertical coordinate function s with its square $v = s^2$. Namely they are straight on sv -half plane. Similarly we can take $V = S^2$ as the vertical coordinate of the second factor, so that the e-geodesics of the second factor are also straight. On the other hand, the correspondence F induces a connection on the second factor from the e-connection on the first factor. In the above mentioned expression $p_{(m,s)}(x) = \exp(X^1 f_1(x) + X^2 f_2(x) - v(X))$ as an exponential family with $f_1(x) = x$ and $f_2(x) = x^2$, we have the coordinates $X^1 = \frac{m}{s^2} = \frac{m}{v}$ and $X^2 = \frac{-1}{2s^2} = \frac{-1}{2v}$. With respect to these coordinates the coefficients of the e-connection vanish. These coordinates can also be written as $X^1 = M$ and $X^2 = -S^2 = -V$ via the correspondance F . Thus the geodesics of the induced metric is straight (and further their affine parameters are affine) on the SV -half plane. Turning to the MS -half plane, we see that the proposition holds. \square

Here we would like to point out that the vertical half-line on the MS -half plane corresponding to an e-geodesic on the ms -half plane with end $(0,0)$ can also be considered as the image of a vertical geodesic for the m-connection on the MS -half plane. Further the affine parametrization of the vertical geodesic for the m-connection on the MS -half plane matches with that of the corresponding geodesic for the e-connection on the mv -half

plane with end $(0, 0)$ via F . The contact Hamiltonian flow in the following theorem describes this phenomena by adding an extra structure to the correspondence F .

THEOREM 1. *The contact Hamiltonian vector field Y of the restriction of the inverse coefficient of variation $\frac{m}{s}$ to the hypersurface $\tilde{N} = \{h = 0\} \subset \mathbb{H} \times \mathbb{H}$ with respect to α_+ is expressed as*

$$Y = \frac{m}{s}e_1 - \frac{1}{2}e_3 = m\partial_m - \frac{1}{2}(S\partial_S - s\partial_s)|_{\tilde{N}}.$$

It is tangent to the surface F . Let Y_i denote the push-forward of the restriction $Y|_F$ to the i -th factor of $\mathbb{H} \times \mathbb{H}$ ($i = 1, 2$). Then the flow of Y_1 comes out of the origin $(0, 0) \in \overline{\mathbb{H}}$ along the geodesics for the e -connection, and the flow of Y_2 goes into the M -axis $\{S = 0\} \subset \overline{\mathbb{H}}$ along the vertical geodesics for the m -connection. The former can be discretized into the iterations of convolutions and therefore the latter into those of Bayesian learnings.

PROOF. The interior product $\iota_Y d\left(\frac{m}{s} + \frac{M}{S}\right) = \frac{1}{2}\left(\frac{m}{s} + \frac{M}{S}\right)$ vanishes along F . The Lie derivative of α_+ is $\mathcal{L}_Y \alpha_+ = \frac{1}{2}\alpha_+$. The Hamiltonian function is $\iota_Y \alpha_+ = \frac{m}{s}$. The rest is easy. \square

6. Further discussions

Since the information geometry concerns parameter spaces with geometric structures, it would have some relation to moduli theory. This was one of the starting points of this research. In the present, we have no intrinsic relation between the pairs of normal distributions and abelian surfaces.

PROBLEM 1. Is there any number theoretical relation between two normal distributions which enables us to relate the pair to an abelian surface with real multiplication?

Another starting point was the following splitting result proved in [9] (see also [3]): A non-singular flow on a closed 3-manifold admits a projectively Anosov splitting (or a dominated splitting) of the normal bundle if and only if it is simultaneously tangent to a mutually transverse pair of positive and negative contact structures, i.e., to a bi-contact structure. Since information theory concerns the distance of the pair of probability distributions rather than

the entropy of a single distribution, we want to know how to split a parameter space (with non-trivial topology) into the pair of parameter spaces. Thus projectively Anosov flows and bi-contact structures would be helpful for topological understanding of the information geometry.

References

- [1] S. Amari, Information geometry and its applications, App. Math. Sci., Springer, Tokyo, 2016.
- [2] Matthew Strom Borman, Yakov Eliashberg, and Emmy Murphy, Existence and classification of overtwisted contact structures in all dimensions, Acta Math. **215**(2) (2015), 281–361.
- [3] Y. Eliashberg and W. Thurston, Confoliations, A. M. S. University Lecture Series **13** (1997).
- [4] John W. Gray, Some global properties of contact structures, Ann. of Math. **69**(2) (1959), 421–450.
- [5] Friedrich Hirzebruch, Hilbert modular surfaces, L'Enseignement Math. **19** (1973), 183–281.
- [6] Ko Honda, On the classification of tight contact structures II, J. Diff. Geom. **55** (2000), 83–143.
- [7] Naohiko Kasuya, The canonical contact structure on the link of a cusp singularity, Tokyo J. Math. **37** (2014), 1–20.
- [8] Patrick Massot, Klaus Niederkrüger, and Chris Wendl, Weak and strong fillability of higher dimensional contact manifolds, Invent. Math. **192**(2) (2013), 287–373.
- [9] Yoshihiko Mitsumatsu, Anosov flows and non-Stein symplectic manifolds, Ann. Inst. Fourier **45** (1995), 1407–1421.

Atsuhide Mori
Department of Mathematics
Osaka Dental University
8-1 Kuzuha-hanazono, Hirakata
Osaka 573-1121 Japan
E-mail: mori-a@cc.osaka-dent.ac.jp
ka-mori@ares.eonet.ne.jp