Sekine, J.
Osaka J. Math.
32 (1995), 71–95

THE HILBERT RIEMANNIAN STRUCTURE OF EQUIVALENT GAUSSIAN MEASURES ASSOCIATED WITH THE FISHER INFORMATION

JUN SEKINE

(Received January 11, 1993)

1. Introduction.

Rao has firstly introduced the Riemannian structure associated with the Fisher information matrix over a finite dimensional parametrized statistical model. He proposed the Riemannian distance as a measure of dissimilality between two probability measures. (cf. [2], for example.) In [1], Amari introduced a pair of dual affine connections with respect to the metric and discussed of the differential geometry of the space of a finite dimensional parametrized statistical model. It provides a differential geometrical meaning to statistical inference.

In the present paper, we realize the above idea for a family of equivalent (i.e.,mutually absolute continuous) Gaussian measures on a Banach space. Our main result is as follows.

Let B be a real separable Banach space and P be a centered gaussian measure on B', the topological dual of B (cf. [6]). The covariance of P naturally determines the Hilbert space H. i.e., for arbitrary x_1 and $x_2 \in B$, let

$$C^P(x_1, x_2) = \int_{B'} \langle x_1, \xi \rangle \langle x_2, \xi \rangle P(d\xi)$$

be the covariance operator of P where $\langle \cdot, \cdot \rangle$ denotes the natural pairing between B and B'. The completion of $(B, C(\cdot, \cdot))$ is a separable Hilbert space. We denote it by $(H, (\cdot, \cdot))$. The space B is continuously embedded in H, so the following relation is satisfied

$$B \underset{\rightarrow}{\subset} H \cong H' \underset{\rightarrow}{\subset} B'.$$

Let us denote

$$\bar{\Theta}_1 = \{A \in L^s_{(2)}(H);\ (I + A) \text{ is positive definite}\} \tag{1.1}$$

where $L^s_{(2)}(H)$ is the totality of symmetric Hilbert-Schmidt operators on H. It is a well-known fact that a Gaussian measure Q which is equivalent to P has a mean vector $b \in H$, i.e.,

$$\int_{B'} \langle x, \xi \rangle Q(d\xi) = (x, b)_H \qquad \text{for all } x \in B$$

and has a covariance operator $(I+A)^{-1}$ with $A \in \bar{\Theta}_1$, i.e.,

$$C^Q(x_1, x_2) = \int_{B'} \langle x_1, \xi \rangle \langle x_2, \xi \rangle Q(d\xi) = ((I+A)^{-1} x_1, x_2)_H$$

for all $x_1, x_2 \in B$. Let

$$\Theta = \bar{\Theta}_1 \times H \tag{1.2}$$

and denote the above Q by P_θ with $\theta := (A, b) \in \Theta$. The totality of Gaussian measures on B' which are equivalent to P, say S, is parametrized by Θ, i.e.,

$$S = \{P_\theta; \theta \in \Theta\}. \tag{1.3}$$

Since we obtain the explicit formula for Radon-Nikodym derivative dP_θ/dP for arbitrary $\theta \in \Theta$, we can introduce the Fisher information at $\theta \in \Theta$. Let

$$\mathscr{H} = L^s_{(2)}(H) \times H \tag{1.4}$$

and call the following symmetric and nonnegative definite bilinear form on $\mathscr{H}$ the Fisher information on S at $\theta \in \Theta$

$$\mathscr{G}_\theta[u, v] = E_\theta[D_u l(\theta) D_v l(\theta)] \qquad \text{with} \quad l(\theta) = \log \frac{dP_\theta}{dP}.$$

Here, $u, v \in \mathscr{H}$, $E_\theta[\cdot]$ denotes the expectation with respect to P_θ and $D.l(\theta)$ denotes the Frechét derivative of $l(\theta)$ which will be defined in Section 4. In our case, because $\mathscr{G}_\theta[\cdot, \cdot]$ is strictly positive definite, it can be interpreted as a Riemennian metric on S. First, we establish the following:

Theorem 1. *The set S is a Hilbert-Riemannian manifold with a global chart*

$$\Phi : \Theta \ni \theta \mapsto \Phi(\theta) = P_\theta \in S$$

and with the Riemannian metric

$$\mathscr{G}_\theta[u, v] = ((I+A)u_2, v_2)_H + \frac{1}{2} T_r((I+A)^{-1} U_1 (I+A)^{-1} V_1) \tag{1.5}$$

at $\theta=(A,b)\in\Theta$ and for $u=(U_1,u_2)$, $v=(V_1,v_2)\in\mathscr{H}$ under the identification between the Hilbert space $\mathscr{H}$ and the tangent space $T_\theta S$.

REMARK: The Fisher information is exactly the hessian of $H(P|P_\theta)$, the relative entropy of P with respect to P_θ. In fact,

$$H(P|P_\theta)=\frac{1}{2}|(I+A)^{1/2}b|_H^2-\frac{1}{2}\log\det{}_2(I+A). \tag{1.6}$$

Next, for $\alpha\in\boldsymbol{R}$ and u, v and $w\in\mathscr{H}$, we will set

$$\overset{(\alpha)}{\Gamma}_\theta[w;u,v]=E_\theta[D_wl(\theta)(D_{uv}^2l(\theta)+\frac{1-\alpha}{2}D_ul(\theta)D_vl(\theta))]$$

and, with this trilinear form on $\mathscr{H}$, we will define a 1-parameter family of affine connections on S. We obtain the following:

Theorem 2. (i) *For $\alpha\in\boldsymbol{R}$, $\theta\in\Theta$ and $u=(U_1,u_2)$, $v=(V_1,v_2)$ and $w=(W_1,w_2)\in\mathscr{H}$,*

$$\begin{aligned}\overset{(\alpha)}{\Gamma}_\theta[w;u,v]&=(U_1v_2+V_1u_2,w_2)_H\\&-\frac{1-\alpha}{2}((V_1w_2,u_2)_H+(W_1u_2,v_2)_H+(U_1v_2,w_2)_H)\\&-\frac{1-\alpha}{2}T_r((I+A)^{-1}U_1(I+A)^{-1}V_1(I+A)^{-1}W_1).\end{aligned} \tag{1.7}$$

The corresponding affine connection on S, say $\overset{(\alpha)}{\nabla}$, is torsion free and satisfies

$$Z(\mathscr{G}_\theta[X,Y])=\mathscr{G}_\theta[\overset{(\alpha)}{\nabla}_ZX,Y]+\mathscr{G}_\theta[X,\overset{(-\alpha)}{\nabla}_ZY] \tag{1.8}$$

for any vector fields X, Y and Z on Θ under the identification of S and Θ. So, $\overset{(-\alpha)}{\nabla}$ is called the dual connection of $\overset{(\alpha)}{\nabla}$ with respect to the metric. Especially, $\overset{(0)}{\nabla}$ is the Levi-Civita connection.

(ii) *$\overset{(1)}{\nabla}$ and $\overset{(-1)}{\nabla}$ are flat connections on S. In fact, for $\theta=(A,b)\in\Theta$, set*

$$\tilde{\theta}=\tilde{\theta}(\theta)=(\tilde{\theta}_1(\theta),\tilde{\theta}_2(\theta))=(-\frac{1}{2}A,(I+A)b) \tag{1.9}$$

$$\text{and } \eta=\eta(\theta)=(\eta_1(\theta),\eta_2(\theta))=((I+A)^{-1}+b\otimes b,b). \tag{1.10}$$

Then, (1.9) *is a* $\overset{(1)}{\nabla}$-*flat,* (1.10) *is a* $\overset{(-1)}{\nabla}$ *flat coordinate, respectively.*

REMARK: (i) For θ, $\theta' \in \Theta$, let us denote

$$\tilde{\theta} = \tilde{\theta}(\theta), \quad \tilde{\theta}' = \tilde{\theta}(\theta'), \quad \eta = \eta(\theta) \quad \text{and} \quad \eta' = \eta(\theta').$$

Then, from the assertion (ii), the $\overset{(1)}{\nabla}$-geodesic between $\tilde{\theta}$ and $\tilde{\theta}'$ is given by

$$(1-t)\tilde{\theta} + t\tilde{\theta}' \qquad (0 \le t \le 1)$$

and the $\overset{(-1)}{\nabla}$-geodesic between η and η' is given by

$$(1-t)\eta + t\eta' \qquad (0 \le t \le 1),$$

respectively.

(ii) Theorem 2 is the extension of the results in finite dimensional statistical model cases. Especially, the assertion (ii) is the analogy of finite dimensional exponential families cases (cf. Section 2 or [1]), since our Gaussian family can be interpreted as "an infinite dimensinal exponential family".

The Gaussian family on a finite dimensional space $\boldsymbol{R}^d$ has non-positive sectional curvatures with respect to the Levi-Civita connection. Analogous result is obtained in our infinite dimensional case.

Theorem 3. *The Riemann-Christoffel curvature tensors* $R^{\alpha}_{\theta}[\cdot;\cdot,\cdot,\cdot]$ *on* S *is equal to*

$$R^{\alpha}_{\theta}[z; u, v, w] = -\frac{1-\alpha^2}{4}((I+A)(U_1 v_2 - V_1 u_2),\ (Z_1 w_2 - W_1 z_2))_H \tag{1.11}$$

for $\alpha \in \boldsymbol{R}$, $\theta = (A, b) \in \Theta$ *and* $u = (U_1, u_2)$, $v = (V_1, v_2)$, $w = (W_1, w_2)$ *and* $z = (Z_1, z_2) \in \mathscr{H}$. *When* $\alpha = 0$, *the sectional curvatures are non-positive. Therefore the manifold* S *has non-positive curvature with respect to the Levi-Civita connection.*

In Section 5, we will deal with an example of the family of linear Gaussian diffusions $X = \{X_t\}_{0 \le t \le T}$ defined by the following stochastic differential equation

$$\begin{cases} dX_t(\omega) = dB_t(\omega) + (a(t)Y_t(\omega) + b(t))dt \quad (0 \le t \le T) \\ \quad X_0 = x \in \boldsymbol{R} \end{cases} \tag{1.12}$$

where $dB_t(\omega)$ is the Wiener integral. We will regard (a, b) as parameters, so will denote the law of the above X by $P_T^{(a,b)}$ and will use the notation P_T instead of

$P_T^{(0,0)}$, which is the Wiener measure on $C([0,T],\boldsymbol{R})$. We obtain that

$$P_T^{(a,b)} \sim P_T \Leftrightarrow (a,b) \in L^2([0,T],tdt) \times L^2([0,T],dt)$$

where "$\sim$" denotes the equivalence (i.e., mutual absolute continuity) between probability measures. Let us set

$$S_T = \{P_T^{(a,b)}; (a,b) \in L^2([0,T],tdt) \times L^2([0,T],dt)\}. \tag{1.13}$$

Further, in this example, by Itô formula, we get the explicit expression of $X_t(\omega)$ and by Cameron-Martin-Maruyama-Girsanov formula, the Radon-Nikodym derivative is written as

$$\frac{dP_T^{(a,b)}}{dP_T}(B(\omega)) = \exp\{\int_0^T (a(t)B_t(\omega) + b(t))dB_t(\omega) - \frac{1}{2}\int_0^T (a(t)B_t(\omega) + b(t))^2 dt\}$$

We will take

$$\rho(t) = (m(t), \sigma^2(t)) = (E[X_t], E[(X_t - m(t))^2])$$

for $0 \leq t \leq T$ as a new coordinate and will compute the Fisher information and the α-affine connections ($\alpha \in \boldsymbol{R}$) on S with respect to this coordinate.

Theorem 4. *Let* $u=(u_1,u_2)$, $v=(v_1,v_2)$ *and* $w=(w_1,w_2) \in (C^1[0,T])^2$ *and for* $\rho(t)=(m(t),\sigma^2(t)) \in (C^1[0,T])^2$, *set*

$$\begin{cases} K_\rho u_1 = \dfrac{d}{dt}u_1 - \dfrac{\frac{d}{dt}\sigma^2(t) - 1}{2\sigma^2(t)}u_1 \\ L_\rho u_2 = \dfrac{1}{2\sigma^2(t)}\left(\dfrac{d}{dt}u_2 - \dfrac{\frac{d}{dt}\sigma^2(t) - 1}{\sigma^2(t)}u_2\right). \end{cases} \tag{1.14}$$

The Fisher information and the α-affine connections on S in ρ-coordinate are equal to

$$\mathscr{G}_\rho^T[u,v] = \int_0^T \{(L_\rho u_2)(L_\rho v_2)\sigma^2(t) + (K_\rho u_1)(K_\rho v_1)\}dt \tag{1.15}$$

(1.16)

$$\overset{(\alpha)}{\Gamma}_{\rho,T}[w;u,v] = -\int_0^T \{(u_2(L_\rho v_2) + v_2) + v_2(L_\rho u_2))(L_\rho w_2)$$

$$+ (u_1(L_\rho v_2) + v_1(L_\rho u_2))(K_\rho w_1)\}dt$$

$$
\begin{aligned}
&-\int_0^T \{(L_\rho u_2)(L_\rho v_2)w_2 \\
&\quad +((K_\rho u_1)(L_\rho v_2)+(K_\rho v_1)(L_\rho u_2))w_1\}dt \\
&+\frac{1-\alpha}{2}\int_0^T [\{(L_\rho u_2)(L_\rho v_2)w_2 \\
&\quad +(L_\rho v_2)(L_\rho w_2)u_2+(L_\rho w_2)(L_\rho u_2)v_2\} \\
&\quad +\{(K_\rho u_1)(L_\rho v_2)+(K_\rho v_1)(L_\rho u_2)\}w_1 \\
&\quad +\{(K_\rho v_1)(L_\rho w_2)+(K_\rho w_1)(L_\rho v_2)\}u_1 \\
&\quad +\{(K_\rho w_1)(L_\rho u_2)+(K_\rho u_1)(L_\rho w_2)\}v_1]dt.
\end{aligned}
$$

In Section 2, we give a brief survey of the differential geometrical structure associated with the Fisher information of an exponential family and mention that it appears naturally in the limit and the rate of convergence in the central limit theorem.

In Section 5, we compute the covariance of the functional $\int_0^T f(t)dX_t^\varepsilon$ of the scaled diffusion $X^\varepsilon(t) := \varepsilon X(\frac{t}{\varepsilon^2})$ $(\varepsilon > 0)$ and observe the behavior as $\varepsilon \to 0$. The Fisher information, 1-and (-1)-connections naturally appear in the asymptotics.

ACKNOWLEDGEMENT. This paper originates in my Master Thesis. I would like to express my sincerely gratitude to my supervisor, Proof. Y. Takahashi, for all his encouragements. I also thank to the referee for his careful reading and his valuable suggestions.

2. A Differential Geometrical Structure of Exponential Families.

Let us consider a family S of mutually absolutely continuous probability measures. Such a family is usually called a statistical model in statistics. We fix a measure space $(\Omega, \mathscr{F}, m)$ so each $P \in S$ can be written as $P(d\omega) = \frac{dP}{dm}(\omega)m(d\omega)$. Sometimes we identify S with the space of Radon-Nikodym derivatives

$$
\{p(\omega) = \frac{dP}{dm}(\omega) : P \in S\}.
$$

A special type of statistical model called an exponential family plays an important role in statiatical inference.Gaussian family and Poissonian family are typical examples of exponential families. In this section, we restrict our statistical models to finite dimensional exponential families and give a brief survey of differential geometry on them. We call it, following Amari(cf. [1]), the information geometry of exponential families. Simple explicit expressions can be obtained on them.

Let $X(\omega)=(X^1(\omega),\cdots,X^n(\omega))$ be an $\boldsymbol{R}^n$-valued random variable on $(\Omega,\mathscr{F},m)$ and $(\cdot,\cdot)$ be the standard inner product on $\boldsymbol{R}^n$. Assume that there existsw a domainΘ in $\boldsymbol{R}^n$ such that

$$\int_\Omega \exp\{(X(\omega),\theta)\}m(d\omega)<\infty$$

for all $\theta\in\Theta$. We set

$$\psi(\theta)=\log\int_\Omega \exp\{(\mathrm{X}(\omega),\theta)\}m(d\omega)$$

for $\theta\in\Theta$. By definition, $\psi(\theta)$ is an analytic function on Θ. Now, we will call the following statistical model

$$S=\{\exp\{l(\omega;\theta)\};\theta\in\Theta\} \tag{2.1}$$

$$\text{where } l(\omega;\theta)=(X(\omega),\theta)-\psi(\theta),$$

an (n-dimensional) exponential family. The function $l(\omega;\cdot)$ on Θ is called the loglikelihood and the parameter $\theta=(\theta_1,\cdots,\theta_n)$ is called a natural parameter of the exponential family S in statistical inference.

The Fisher information matrix is the following symmetric nonnegative matrix

$$G_\theta=(g_{ij}(\theta))_{i\le i,j\le n}=E_\theta[\frac{\partial l(\omega;\theta)}{\partial\theta_i}\frac{\partial l(\omega;\theta)}{\partial\theta_j}] \tag{2.2}$$

where $E_\theta[\cdot]$ denotes the expectation with respect to $P_\theta(d\omega):=e^{l(\theta)}m(d\omega)$. If we assume it is strictly positive definite, it gives the Riemannian metric $(\cdot,\cdot)$ on the tangent space $T_\theta S$ by

$$(u,v)_\theta=\sum_{1\le i,j\le n} g_{ij}(\theta)u^iv^j \tag{2.3}$$

where $u=\sum_{i=1}^n u^i(\frac{\partial}{\partial\theta_j})_\theta$ and $v=\sum_{j=1}^n v^j(\frac{\partial}{\partial\theta_j})_\theta\in T_\theta S$.

Next, we will define a 1-parameter family of torsion free affine connections $\overset{(\alpha)}{\nabla}$ $(\alpha\in\boldsymbol{R})$ on S by

$$((\frac{\partial}{\partial^\theta{}_k})_\theta,\ \overset{(\alpha)}{\nabla}_{(\frac{\partial}{\partial\theta_i})_\theta}(\frac{\partial}{\partial\theta_j})_\theta)_\theta=E_\theta[\frac{\partial l(\theta)}{\partial\theta_k}(\frac{\partial^2 l(\theta)}{\partial\theta_i\partial\theta_j}+\frac{1-\alpha}{2}\frac{\partial l(\theta)}{\partial\theta_i}\frac{\partial l(\theta)}{\partial\theta_j})]. \tag{2.4}$$

By definition, $\overset{(\alpha)}{\nabla}$ and $\overset{(-\alpha)}{\nabla}$ satisfy

$$A(B, C)_\theta = (\overset{(\alpha)}{\nabla}_A B, C)_\theta + (B, \overset{(-\alpha)}{\nabla}_A C)_\theta \tag{2.5}$$

for any smooth vector fields A, B and C. So $\overset{(-\alpha)}{\nabla}$ is called the dual connection of $\overset{(\alpha)}{\nabla}$with respect to the Riemannian metric. The connection $\overset{(0)}{\nabla}$ is actually the Levi-Civita connection(cf. [5]).

Furthemore, we can observe that $\overset{(1)}{\nabla}$ and $\overset{(-1)}{\nabla}$ are flat affine connections on S. In fact, we can choose $\overset{(1)}{\nabla}$ and $\overset{(-1)}{\nabla}$ flat coordinates as follows:

Since

$$\begin{aligned} ((\frac{\partial}{\partial\theta_k})_\theta, \overset{(1)}{\nabla}(\frac{\partial}{\partial\theta_i})_\theta(\frac{\partial}{\partial\theta_j})_\theta)_\theta &= E_\theta[\frac{\partial l(\theta)}{\partial\theta_k}\frac{\partial^2 l(\theta)}{\partial\theta_i\partial\theta_j}] \\ &= -\frac{\partial^2\psi(\theta)}{\partial\theta_i\partial\theta_j}E_\theta[\frac{\partial l(\theta)}{\partial\theta_k}] \\ &= -\frac{\partial^2\psi(\theta)}{\partial\theta_i\partial\theta_j}E[\frac{\partial}{\partial\theta_k}(e^{l(\theta)})] = 0 \end{aligned} \tag{2.6}$$

for $1 \leq i,j,k \leq n$, so the natural parameter θ *is a* $\overset{(1)}{\nabla}$-flat coordinate. A $\overset{(-1)}{\nabla}$-flat coordinate is defined by

$$\eta = \eta(\theta) = (\eta^1(\theta), \cdots, \eta^n(\theta)) = E_\theta[X] \tag{2.7}$$

and called the expectation parameter of S. It is easy to see that

$$(\frac{\partial\eta_i(\theta)}{\partial\theta_i})_{1 \leq i,j \leq n} = G_\theta > 0,$$

which shows η is actually another coordinate of S. It follows from

$$((\frac{\partial}{\partial\theta_i})_\theta, (\frac{\partial}{\partial\eta_j})_{\eta(\theta)})_\theta = \delta_{ij}$$

that

$$(2.8)\qquad ((\frac{\partial}{\partial\theta_i})_\theta, \overset{(-1)}{\nabla}_{A_\theta}(\frac{\partial}{\partial\eta_j})_{\eta(\theta)})_\theta$$

$$=A_\theta((\frac{\partial}{\partial\theta_i})_\theta, (\frac{\partial}{\partial\eta_i})_{\eta(\theta)})_\theta-(\overset{(1)}{\nabla}_{A_\theta}(\frac{\partial}{\partial\theta_i})_\theta, (\frac{\partial}{\partial\eta_j})_{\eta(\theta)})_\theta=0.$$

is satisfied for arbitrary $1\leq i, j\leq n$ and a smooth vector field A. Hence, η is a $\overset{(-1)}{\nabla}$-flat coordinate.

These geometrical quantities appear in a natural manner when we compute the rate of convergence in the central limit theorem as follows.

Let $\{X_i(\cdot)\}_{j\in N}$ be the independent copies of $X(\cdot)$ and let

$$\tilde{X}_N=\frac{1}{\sqrt{N}}\sum_{j=1}^{N}(X_j-E_\theta[X_j]).$$

Of course, $\tilde{X}_N$ goes to the n-dimensional centered Gaussian random variable with covariance matrix G_θ in law when N goes to infinity by the central limit theorem. For arbitrary $u\in \boldsymbol{R}$, we obtain the followings.

$$(2.9)\qquad \begin{cases} E_\theta^N[(\tilde{X}_N, u)]=0 \\ E_\theta^N[(\tilde{X}_N, u)^2]=(u,u)_\theta \\ E_\theta^N[(\tilde{X}_N, u)^3]=\dfrac{1}{\sqrt{N}}(\overset{(-1)}{\nabla}_u u, u)_\theta \\ E_\theta^N[(\tilde{X}_N, u)^4]=3(u,u)_\theta^2) \\ \qquad +\dfrac{1}{N}((\overset{(-1)}{\nabla}_u \overset{(-1)}{\nabla}_u u, u)_\theta+3(u,u)_\theta^2) \end{cases}$$

et cetera where $E_\theta^N[\cdot]$ denotes the expectation with respect to $P_\theta^N:=\prod^N P_\theta(d\omega)$. The k-th iterates of $\overset{(-1)}{\nabla}$ appears in the k-th moment of $\tilde{X}_N$. These formulas are understood in a more natural manner when we look at the cumulant expansions:

$$(2.10)\qquad \int_{(\boldsymbol{R}^n)N}\exp\{\sqrt{-1}(\tilde{X}_N, u)\}P_\theta^N(d\omega)$$

$$=\exp\{-\frac{1}{2}(u,u)_\theta+\sum_{k=3}^{\infty}\frac{1}{N^{\frac{k-2}{2}}}\frac{\sqrt{-1}^k}{k!}(\underbrace{\overset{(-1)}{\nabla}_u\cdots\overset{(-1)}{\nabla}_u}_{(k-2)times}u,u)_\theta\}.$$

In the next section, we will deal with the family of equivalent Gaussian measures on a Banach space. It is an example of "infinite dimensional version of exponential family".

3. Gaussian Measures on a Banach space.

Let B a real separable Banach space. A centered Gaussian measure P on B' is characterized as follows. Since its one dimensional marginal $\langle x,\cdot\rangle$, is a Gaussian distribution with mean 0, so $\langle x,\cdot\rangle$, is square integrable, we have a natural embedding j: $B\ni x\mapsto\langle x,\cdot\rangle\in L^2(B',P)$. We denote the completion of the range of j in $L^2(B',P)$ by H, which is a separable Hilbert space. We regard j as a map from B to H, so we obtain the relation as $B\underset{\rightarrow}{\subset}H\cong H'\underset{\rightarrow}{\subset}B'$. This structure (j,H,B',P) is called as abstract Wiener space and P is called as abstract Wiener (the standard Gaussian) measure. For details, see [6] for example. The characteristic function of P is expressed as

$$\int_{B'}e^{\sqrt{-1}\langle x,\xi\rangle}\,P(d\xi)=e^{-\frac{1}{2}|x|_H^2}$$

where $x\in B$ and $|\cdot|_H=\sqrt{(\cdot,\cdot)_H}$ is the Hilbert norm on H.

We choose our statistical model as the set of Gaussian measures on B' which are equivalent to P and denote it by S through this paper.

Theorem 0. *Let P be the standard Gaussian measure on (j,H,B'). Then, a Gaussian measure on B', say Q, is equivalent to P, if and only if there exist $A\in\bar{\Theta}_1$ and $b\in H$, and the characteristic function of Q is given by the following*:

$$\int_{B'}\exp\{\sqrt{-1}\langle x,\xi\rangle\}G(d\xi)=\exp\{\sqrt{-1}(x,b)_H-\frac{1}{2}|(I+A)^{-1/2}x|_H^2\} \tag{3.1}$$

for $x\in B$.

Sketch of the Proof: Let

$$\begin{cases}\mu_k(dx)=\dfrac{1}{\sqrt{2\pi}}\exp\{-\dfrac{x^2}{2}\}dx\\[2ex] \nu_k(dx)=\sqrt{\dfrac{(1+\alpha_k)}{2\pi}}\exp\{-\dfrac{(1+\alpha_k)(x-b_k)^2}{2}\}dx\end{cases}$$

be Gaussian measures on $\boldsymbol{R}$ for $k \in \boldsymbol{N}$ and consider two Gaussian measures on $\boldsymbol{R}$ such that

$$\begin{cases} \mu = \prod_{k=1}^{\infty} \mu_k \\ \nu = \prod_{k=1}^{\infty} \nu_k \end{cases}$$

where $\prod^{\infty}(\cdot)$ denotes the infinite products of probability measures. Then, by using Kakutani's theorem (cf. [6]), straightfoward caluculations tell us that

$$\mu \sim \nu \Leftrightarrow \sum_k \alpha_k^2 < +\infty \text{ and } \sum_k b_k^2 < +\infty.$$

The proof of general Banach space's case reduces to the proof of the above $\boldsymbol{R}^{\boldsymbol{N}}$'s case. For details, see [9] for example. ■

Set

$$\Theta = \bar{\Theta}_1 \times H$$

and write the above Gaussian measure Q as P with $\theta = (A, b) \in \Theta$. The set S is expressed as

$$S = \{P_\theta : \theta \in \Theta\}.$$

Next we need to express the Radon-Nikodym derivatives. When A is a trace class operator and b balongs to B, we can write down dP_θ/dP immedeately. Let

$$\Theta_1 = \{A \in L^s_{(1)}(H);\ I + A \text{ is positive definite}\}$$

where $L^s_{(1)}(H)$ is the totality of symmetric trace class operators on H.

Lemma 1. *Let $\theta = (A, b) \in \Theta_1 \times B$. Then,*

$$\frac{dP_\theta}{dP}(\xi) = \exp\{l(\xi; \theta)\} \tag{3.2}$$

where $l(\xi; \theta) = \frac{1}{2} \log \det(I + A) - \frac{1}{2} \langle A(\xi - b), \xi - b \rangle + \langle b, \xi \rangle - \frac{1}{2} |b|_H^2.$

REMARK: The quantity $\langle A(\xi - b), \xi\text{-}b \rangle$ will be defined in the proof below as the random variable on B'.

Proof: Take eigenfunctions of A consisting of an orthonormal basis of H, say $\{e_n\}_{n\in N}$, and corresponding eigenvalues, say $\{\alpha_n\}_{n\in N}$. Then, $\alpha_n > -1$ for all $n\in N$, $\Sigma_n|\alpha_n| < +\infty$ and for any $h\in H$,

$$Ah=\sum_n \alpha_n h_n e_n \quad \text{with} \quad h_n=(h,e_n)_H.$$

Let us denote the projection on H to the span $\{e_1,\cdots,e_n\}$ by Q_n and write $\theta_n=((Q_nA, Q_nb)$ for $\theta=(A, b)\in\Theta_1\times B$. When $\xi\in H$, the Radon-Nikodym derivative is expressed as

$$\frac{dP_{\theta_n}}{dP}(\xi)=\sqrt{\det(I+Q_nA)}\ \exp\{-\frac{1}{2}(Q_nA(\xi-b),\ \xi-b)_H+(Q_nb,\ \xi)_H-\frac{1}{2}|Q_nb|_H^2\}$$

or equivalently

$$\frac{dP_{\theta_n}}{dP}(\xi)=(\prod_{k=1}^{n}(1+\alpha_k)^{1/2})\exp\{-\frac{1}{2}\sum_{k=1}^{n}\alpha_k((\xi,e_k)_H-b_k)^2$$

$$+\sum_{k=1}^{n}b_k(\xi,e_k)_H-\frac{1}{2}|Q_nb|_H^2\}.$$

Here, (ξ,e) can be extended to $\xi\in B'$ as the Gaussian random variable on B' with mean 0 and variance 1, which we will denote $e_k(\xi)$ $(\xi\in B')$. (cf. [6])

So, for $\xi\in B'$, Radon-Nikodym derivative is written as

$$\frac{dP_{\theta_n}}{dP}(\xi)=(\prod_{k=1}^{n}(1+\alpha_k)^{1/2})\exp\{-\frac{1}{2}\sum_{k=1}^{n}\alpha_k(e_k(\xi)-b_k)^2+\sum_{k=1}^{n}b_ke_k(\xi)-\frac{1}{2}|Q_nb|_H^2\}.$$

Martingale convergence theorem (cf. [3], for example) assures the existence of the limit

$$\frac{dP_\theta}{dP}=\lim_{n\to\infty}\frac{dP_{\theta_n}}{dP}$$

and its integrability

$$E[\frac{dP_\theta}{dP}]=\lim_{n\to\infty}E[\frac{dP_{\theta_n}}{dP}]=1,$$

hence follows the lemma. ∎

When we will write down $\frac{dP_\theta}{dP}(\xi)$ for $A \in \bar{\Theta}_1 \backslash \Theta_1$ and $b \in H \backslash B$, since the expression of the above lemma lose its mean so we have to *renormalize* it. In fact, we can express it as the certain limit as stated below.

Lemma 2. *Let $\{P_n\}_{n\in N}$ be a sequence of finite dimensional projection on H such that $P_n \to I$ as $n \to \infty$ with respect to the operator norm. Then, the limits of*

$$\langle P_n(I+A)b, \xi\rangle$$
$$\text{and } \langle P_n A\xi, \xi\rangle - \mathrm{Tr}(P_n A)$$

in both L^2-sense and almost sure-sense as $n \to \infty$ exist.

Proof: If $P_n = Q_n$, which was defined in the proof of Lemma 1, we have

$$\langle Q_n(I+A)b, \xi\rangle = \sum_{k=1}^{n} (1+\alpha_k) b_k e_k(\xi)$$

$$\text{and } \langle Q_n A\xi, \xi\rangle - \mathrm{Tr}(Q_n A) = \sum_{k=1}^{n} \alpha_k (e_k(\xi)^2 - 1).$$

Khinchin-Kolmogorov's theorem tells us that if independent random variables $\{f_k\}_{k\in N}$ satisfy $E[f_k]=0$ for all $k \in N$ and $\sum_k E[f_k{}^2] < \infty$, then the limit $\lim_n \sum_{k=1}^{n} f_k$ in both L^2-sense and almost sure-sense exists (cf. [3]). Since $\{e_k(\cdot)\}_{k\in N}$ is a sequence of independent identically distributed Gaussian random variables with means 0 and variances 1, the assertion is proved. It is an easy step to complete the proof for general $\{P_n\}_{n\in N}$ (cf. [6]). ■

NOTATION 1. *We denote that*

$$:\langle (I+A)b, \xi\rangle: = \lim_{n\to\infty} \langle P_n(I+A)b, \xi\rangle \tag{3.3}$$

$$:\langle A\xi, \xi\rangle - \mathrm{Tr}(A): = \lim_{n\to\infty} \{\langle P_n A\xi, \xi\rangle - \mathrm{Tr}(P_n A)\}. \tag{3.4}$$

DEFINITION 1. *For a Hilbert-Schmidt operator A on H, let us define*

$$\det{}_2(I+A) = \det((I+A)e^{-A})$$

(*cf.* [8]).

REMARK: In particular, if A is a trace class operator on H,

$$\det_2(I+A)=\prod_{n=1}^{\infty}\{(1+\alpha_n)e^{-\alpha_n}\}=(\det(I+A))e^{-\mathrm{Tr}(A)}$$

where $\{\alpha_n\}_{n\in N}$ are eigenvalues of A.

Lemma 3. *For arbitrary $\theta=(A,b)\in\Theta$, let*

$$(3.5)\qquad :l(\xi;\,\theta):=-\frac{1}{2}:\langle A\xi,\xi\rangle-\mathrm{Tr}(A):+:\langle(I+A)b,\xi\rangle:$$

$$-\frac{1}{2}|(I+A)^{1/2}b|_H^2+\frac{1}{2}\log\det_2(I+A)$$

and let $\theta_n=(P_nA,\ P_nb)$ where $\{P_n\}_{n\in N}$ is as same as in Lemma 2. Then,

$$(3.6)\qquad :l(\xi;\,\theta):=\lim_{n\to\infty}l(\xi\ \theta_n)$$

in both $L^2(P)$-sense and almost sure-sense and it satisfies

$$(3.7)\qquad E[\exp\{:l(\xi;\,\theta):\}]=1.$$

Therefore,

$$(3.8)\qquad \frac{dP_\theta}{dP}(\xi)=\exp\{:l(\xi;\,\theta):\}$$

for almost everywhere $\xi\in B'$.

Proof: We will only show in the case of $P_n=Q_n$ for all $n\in N$. By Lemma 1, we see that

$$l((\xi;\,\theta_n)=-\frac{1}{2}\sum_{k=1}^{n}\alpha_k e_k(\xi)^2+\sum_{k=1}^{n}(1+\alpha_k)b_k e_k(\xi)$$

$$-\frac{1}{2}|Q_n(I+A)^{1/2}b|_H^2+\frac{1}{2}\log(\prod_{k=1}^{n}(1+\alpha_k))$$

$$=-\frac{1}{2}\sum_{k=1}^{n}\alpha_k(e_k(\xi)^2-1)+\sum_{k=1}^{n}(1+\alpha_k)b_k e_k(\xi)$$

$$-\frac{1}{2}|Q_n(I+A)^{1/2}b|_H^2+\frac{1}{2}\log\{\prod_{k=1}^{n}(1+\alpha_k)e^{-\alpha_k})\},$$

so, (3.6) is obtained by Lemma 2. Further, (3.7) comes from martingale convergence theorem, hence follows the lemma. ■

REMARK: Let ν and λ be probability measures on a mesurable space and let $H(\nu|\lambda)$ be the relative entropy of ν with respect to λ which is given by the formula

$$H(\nu|\lambda)=\begin{cases}\int \log \frac{d\nu}{d\lambda}d\nu & \text{if } \nu \ll \lambda, \\ +\infty & \text{otherwise}\end{cases}$$

In our Gaussian case, $P \ll P_\theta$ is equivalent to $P \sim P_\theta$ ($\sim$ denotes the mutual absolute continuity between probability measures) and for $\theta \in \Theta$, we have

$$H(P|P_\theta)=-E[:l(\theta):]=\frac{1}{2}|(I+A)^{1/2}b|_H^2-\frac{1}{2}\log \det_2(I+A),$$

hence, we observe that

$$P \sim P_\theta \Leftrightarrow H(P|P_\theta)<\infty.$$

4. Gaussian Statistical Manifold.

In this Section, we will construct an infinite dimensional version of *information geometry* for Gaussian measures. We will start with the following:

Lemma 4. *The function*

$$\log \det_2(I+\cdot)|_{\bar{\Theta}_1}:\bar{\Theta}_1 \mapsto \boldsymbol{R}$$

which is restricted to $\bar{\Theta}_1$*, an open set of* $L^s_{(2)}(H)$*, is* C^∞*-Fréchet-differentiable. i.e., for instance, the continuous linear function* $D.(\log \det_2(I+A))$ *on* $L^s_{(2)}(H)$ *which satisfies*

$$|\log \det_2(I+(A+sU_1))-\log \det_2(I+A)-sD_{U_1}(\log \det_2(I+A))|=o(s)$$

for $U_1 \in L^s_{(2)}(H)$ *as* $s \to 0$ *exists for each* $A \in \bar{\Theta}_1$ *and it is computed as*

$$D_{U_1}(\log \det_2(I+A))=-\mathrm{Tr}(A(I+A)^{-1}U_1). \tag{4.1}$$

The second and the third derivative are equal to

$$\begin{aligned} D^2_{U_1V_1}(\log \det_2(I+A))&=D_{U_1}(D_{V_1}(\log \det_2(I+A))) \\ &=-\mathrm{Tr}((I+A)^{-1}U_1(I+A)^{-1}V_1) \qquad (4.2) \end{aligned}$$

$$D^3_{U_1V_1W_1}(\log \det_2(I+A))=D_{U_1}(D^2_{V_1W_1}(\log \det_2(I+A)))$$

(4.3) $$=2\mathrm{Tr}((I+A)^{-1}U_1(I+A)^{-1}V_1(I+A)^{-1}W_1)$$

for $V_1,\ W_1\in L^s_{(2)}(H)$.

REMARK: When $A\in\bar{\Theta}_1$, since A is a compact operator, 0 is the only accumulation point of eigenvalues, so $(I+A)^{-1}$ is a bounded operator.

Proof: The analyticity is obvious from the definition, and to obtain the derivatives we use the following relation. (cf. [8])

$$\det{}_2(I+A+B+AB)=\det{}_2(I+A)\det{}_2(I+B)e^{-\mathrm{Tr}(AB)}$$

for Hilbert-Schmidt operators A, B. Then, for arbitrary $U\in L^s_{(2)}(H)$ and $s\in\boldsymbol{R}$,

$$\begin{aligned}
&\det{}_2(I+A+sU)\\
&\quad=\det{}_2(I+A)\det{}_2(I+s(I+A)^{-1}U)e^{-s\mathrm{Tr}(A(I+A)^{-1}U)}\\
&\quad=\det{}_2(I+A)\{1-s\mathrm{Tr}(A(I+A)^{-1}U)\\
&\qquad+\frac{s^2}{2}(-\mathrm{Tr}((I+A)^{-1}U(I+A)^{-1}U)+(\mathrm{Tr}(A(I+A)^{-1}U))^2\\
&\qquad+\frac{s^3}{6}(2\mathrm{Tr}((I+A)^{-1}U(I+A)^{-1}U(I+A)^{-1}U)-(\mathrm{Tr}(A(I+A)^{-1}U))^3)\\
&\qquad\quad+3\mathrm{Tr}((I+A)^{-1}U(I+A)^{-1}U)\mathrm{Tr}(A(I+A)^{-1}U)+o(s^3)\}.
\end{aligned}$$

Hence follows the lemma. ■

Lemma 5. *Let* $\theta=(A,b)\in\Theta$ *and* $\mathscr{U}$ *be an open neighborhood of* 0 *in* $\bar{\Theta}_1$ *such that* $A+\mathscr{U}\in\bar{\Theta}_1$. *Then, the function*

$$(:l(\theta+\cdot):)|_{U\times H}:\mathscr{U}\times H\mapsto L^2(P_\theta)$$

which is retricted to $\mathscr{U}\times H$ *is* C^∞*-Frechét-differentiable. Especially, first two derivatives are given as follows*

(4.4)
$$\begin{aligned}
D_u(:l(\xi;\theta):)=&-\frac{1}{2}(:\langle U\xi,\xi\rangle-\mathrm{Tr}(U):)+\langle(I+A)u,\xi\rangle:+:\langle Ub,\xi\rangle:\\
&-((I+A)b,u)_H-\frac{1}{2}(Ub,b)_H-\frac{1}{2}\mathrm{Tr}(A(I+A)^{-1}U)
\end{aligned}$$

$$D^2_{uv}(:l(\xi;\theta):)=:\langle U_1v_2+V_1u_2,\xi\rangle:-(U_1v_2+V_1u_2,b)_H$$

$$-((I+A)u_2,v_2)_H-\frac{1}{2}\mathrm{Tr}((I+A)^{-1}U_1(I+A)^{-1}V_1) \tag{4.5}$$

where $u=(U_1,u_2), v=(V_1,v_2)\in\mathscr{H}:=L^s_{(2)}(H)\times H$.

Moreover, the derivatives belong to $L^k(P_\theta)$ *for any* $k\in\boldsymbol{N}$.

Proof: Differentiability is obvious, because $:l(\xi;\cdot):$ is the sum of continuous (multi) linear functions. By direct computations, we get (4.4) and (4.5). Higher derivatives $D^k(:l(\theta):)$ for $k\geq 3$ are also obtained directly. Integrability of the derivatives is straightfoward from the expressions of them. ■

Now, we can introduce the Fisher information and the α-affine connections on S.

DEFINITION 2. (i) *The Fisher Information on* S *at* $\theta\in\Theta$ *is the following symmetric, nonnegative definite bilinear form on* $\mathscr{H}$

$$\mathscr{G}_\theta[u,v]=E_\theta[D_u(:l(\theta):)D_v(:l(\theta):)] \tag{4.6}$$

for $u,v\in\mathscr{H}$, *where* $E_\theta[\cdot]$ *denotes the expectation with respect to* P_θ.

(ii) *For* $\alpha\in\boldsymbol{R}$, *the* α-*affine connection on* S *at* $\theta\in\Theta$ *is the following trilinear form on* $\mathscr{H}$

$$\overset{(\alpha)}{\Gamma}_\theta[w;u,v]=E_\theta[D_w(:l(\theta):)(D^2_{uv}(:l(\theta):)+\frac{1-\alpha}{2}D_u(:l(\theta):)D_v(:l(\theta):))]. \tag{4.7}$$

They are well-defined by Lemma 5, and are computed as follows.

Lemma 6. *Let* $\theta=(A,b)\in\Theta$, $u=(U_1,u_2), v=(V_1,v_2)$ *and* $w=(W_1,w_2)\in\mathscr{H}$ *and* $\alpha\in\boldsymbol{R}$. *Then,*

(i) $\mathscr{G}_\theta[u,v]=((I+A)u_2,v_2)_H+\dfrac{1}{2}\mathrm{Tr}((I+A)^{-1}U_1(I+A)^{-1}V_1)$

(ii) $\overset{(\alpha)}{\Gamma}_\theta[w;u,v]=(U_1v_2+V_1u_2,w_2)_H$

$$-\frac{1-\alpha}{2}\{(V_1w_2,u_2)_H+(W_1u_2,v_2)_H+(U_1v_2,w_2)_H\}$$

$$-\frac{1-\alpha}{2}\mathrm{Tr}((I+A)^{-1}U_1(I+A)^{-1}V_1(I+A)^{-1}W_1).$$

Proof: (i) For any $\theta=(A,b)\in\Theta$, take Q_n, the projection on H which was defined in the proof of Lemma 1, and set $\theta_n=(Q_nA,Q_nb)$. It is easy to see that

$$E[D^k(e^{l(\theta_n)})]=D^k(E[e^{l(\theta_n)}])=0$$

for $k\geq 1$. In particular, we have

$$E[D^2_{u_nv_n}(e^{l(\theta_n)})]=E[(D^2_{u_nv_n}l(\theta_n)+D_{u_n}l(\theta_n))D_{v_n}l(\theta_n))e^{l(\theta_n)}]=0$$

for any $u,v\in\mathscr{H}$ with $u_n=(Q_nU_1,Q_nu_2)$, $v_n=(Q_nV_1,Q_nv_2)$. So,

$$\begin{aligned}\mathscr{G}_{\theta_n}[u_n,v_n]&=-E_{\theta_n}[D^2_{u_nv_n}l(\theta_n)]\\&=((I+Q_nA)Q_nu_2,v_2)_H+\frac{1}{2}\mathrm{Tr}((I+Q_nA)^{-1}U_1(I+Q_nA)^{-1}V_1)\end{aligned}$$

and by using Lebesgue's dominated convergence theorem, we get

$$\begin{aligned}\mathscr{G}_\theta[u,v]&=\lim_{n\to\infty}E(D_{u_n}l(\theta_n)D_{v_n}l(\theta_n)e^{l(\theta_n)}]\\&=\lim_{n\to\infty}(\mathscr{G}_{\theta_n}[u_n,v_n])\\&=((I+A)u_2,v_2)_H+\frac{1}{2}\mathrm{Tr}((I+A)^{-1}U_1(I+A)^{-1}V_1),\end{aligned}$$

hence the assertion (i) follows.

(ii) Note that

$$E_{\theta_n}[D_{w_n}l(\theta_n)D^2_{u_nv_n}l(\theta_n)]=(Q_n(U_1v_2+V_1u_2),Q_nw_2)_H$$

holds for $w=(W_1,w_2)\in\mathscr{H}$ with $w_n=(Q_nW_1,Q_nw_2)$. So, using Lebesgue's dominated convergence theorem, we get that

$$E_\theta[D_w(:l(\theta):)D^2_{uv}(:l(\theta):)]=(U_1v_2+V_1u_2,w_2)_H. \tag{4.8}$$

In the similar way, we see that

$$\begin{aligned}D_w(\mathscr{G}_\theta[u,v])&=D_w(E[D_u(:l(\theta):)D_v(:l(\theta):)e^{:l(\theta):}])\\&=E_\theta[D^2_{wu}(:l(\theta):)D_v(:l(\theta):)]+E_\theta[D_u(:l(\theta):)D^2_{wv}(:l(\theta):)]\\&\quad+E_\theta[D_u(:l(\theta):)D_v(:l(\theta):)D_w(:l(\theta):)].\end{aligned} \tag{4.9}$$

The left hand side of (4.9) is computed as

$$(W_1 u_2, v_2)_H - \mathrm{Tr}((I+A)^{-1} U_1 (I+A)^{-1} V_1 (I+A)^{-1} W_1).$$

Now, it is easy to deduce the assertion (ii) from (4.8) and (4.9). ■

Proof of Theorem 1: We can take a map

$$\Phi : \Theta \ni \theta \mapsto \Phi(\theta) = P_\theta \in S$$

from Θ, an open set of $\mathscr{H}$, to S as a global chart of S. Therefore, S is a Hilbert manifold with the model space Θ (cf.[7]). Further, since the Fisher information $\mathscr{G}_\theta[\cdot,\cdot]$ is strictly positive definite for any $\theta \in \Theta$, it defines the Riemannian metric $((\cdot,\cdot))_\theta$ on the tangent space $T_\theta S$ by the formula

$$((U, V))_\theta = \mathscr{G}_\theta[(\Phi_*)_\theta^{-1} U, (\Phi_*)_\theta^{-1} V]$$

for U, $V \in T_\theta S$ with

$$(\Phi_*)_\theta : \mathscr{H} \mapsto T_\theta S,$$

the differential of Φ at θ, which is a linear isomorphism between the Hilbert space $\mathscr{H}$ and the tangent space $T_\theta S$. Therefore, Theorem 1 is established. ■

Proof of theorem 2: (i) It is obvious that $\overset{(\alpha)}{\nabla}$ is torsion free, so let us observe the relation (1.8). Take a smooth curve $c = \{\theta(t); t \in (-\varepsilon, \varepsilon)\}$ on Θ for some $\varepsilon > 0$. For arbitrary

$$X = \{(\theta(t), u(t)); t \in (-\varepsilon, \varepsilon)\} \text{ and } Y = \{(\theta(t), u(t)); t \in (-\varepsilon, \varepsilon)\} \subset \Theta \times \mathscr{H},$$

smooth vector fields along the curve c, we see that

$$\begin{aligned}
\frac{d}{dt}(\mathscr{G}_{\theta(t)}[u(t), v(t)]) &= \frac{d}{dt}(E[D_{u(t)}(:l(\theta(t)):) D_{v(t)}(:l(\theta(t)):) e^{:l(\theta(t)):}]) \\
&= (\mathscr{G}_{\theta(t)}[\dot{u}(t), v(t)] + \overset{(\alpha)}{\Gamma}_{\theta(t)}[v(t): \dot{\theta}(t), u(t)]) \\
&\quad + (\mathscr{G}_{\theta(t)}[u(t), \dot{v}(t)] + \overset{(-\alpha)}{\Gamma}_{\theta(t)}[u(t); \dot{\theta}(t), v(t)]) \\
&= \mathscr{G}_{\theta(t)}[\overset{(\alpha)}{\nabla}_{\theta(t)} u(t), v(t)] + \mathscr{G}_{\theta(t)}[u(t), \overset{(-\alpha)}{\nabla}_{\theta(t)} v(t)]
\end{aligned}$$

where

$$\dot{\theta}(t)=\frac{d}{dt}\theta(t), \dot{u}(t)=\frac{d}{dt}u(t) \text{ and } \dot{v}(t)=\frac{d}{dt}v(t).$$

Hence, (1.8) follows. Obviously, $\overset{(0)}{\nabla}$ is the Levi-Civita connection.

(ii) If we take

$$\tilde{\theta}=(\tilde{\theta}_1,\tilde{\theta}_2)=(\tilde{\theta}_1(\theta),\tilde{\theta}_2(\theta))=(-\frac{1}{2}A,(I+A)b)$$

as another coordinate of S, then, the log-likelihood function is written as

$$(:l(\tilde{\theta}):)=(:l(\theta(\tilde{\theta})):)=(:\langle\tilde{\theta}_1,\xi\otimes\xi\rangle-\mathrm{Tr}(\tilde{\theta}_1):)+(:\langle\tilde{\theta}_2,\xi\rangle:)-\psi(\tilde{\theta})$$

where $\psi(\tilde{\theta})$ is a non-random function of $\tilde{\theta}$. So, $\tilde{\theta}$ is "a natural parameter" of our "infinite dimensional exponential family" (cf. Section 2). Therefore, we get

$$E_\theta[D_{\tilde{w}}(:l(\tilde{\theta}):)D^2_{\tilde{u},\tilde{v}}(:l(\tilde{\theta}):)]=-D^2_{\tilde{u},\tilde{v}}\psi(\tilde{\theta})E_\theta[D_{\tilde{w}}(:l(\tilde{\theta}):)]=0$$

for $\tilde{u}$, $\tilde{v}$ and $\tilde{w}\in\mathscr{H}$ where $E_{\tilde{\theta}}[\cdot]$ denotes the expectation with respect to $P_{\theta(\tilde{\theta})}$. Hence, $\tilde{\theta}$ is a $\overset{(1)}{\nabla}$flat coordinate.

Furthermore, if we take

$$\begin{aligned}\eta=(\eta_1,\eta_2)=(\eta_1(\theta),\eta_2(\theta))&=(E_\theta[\xi\otimes\xi],E_\theta[\xi])\\&=((I+A)^{-1}+b\otimes b,b)\end{aligned}$$

as a new coordinate of S, it is "an expectation parameter" of S (cf. Section 2), so, by the similar way in Section 2, we see that η is a $\overset{(-1)}{\nabla}$ flat coordinate. ■

Proof of Theorem 3: The Riemann-Christoffel curvature tensors (cf. [5]) $R^\alpha_\theta[\cdot;\cdot,\cdot,\cdot]$ at $\theta=(A,b)\in\Theta$ is

$$\begin{aligned}R^\alpha_\theta[z;u,v,w]&=\mathscr{G}_\theta[(\overset{(\alpha)}{\nabla}_u\overset{(\alpha)}{\nabla}_v-\overset{(\alpha)}{\nabla}_v\overset{(\alpha)}{\nabla}_u)w,z]\\&=D_u(\mathscr{G}_\theta[\overset{(\alpha)}{\nabla}_v w,z])-D_v(\mathscr{G}_\theta[\overset{(\alpha)}{\nabla}_u w,z])\\&\quad-\{\mathscr{G}_\theta[\overset{(\alpha)}{\nabla}_v w,\overset{(-\alpha)}{\nabla}_u z]-\mathscr{G}_\theta[\overset{(\alpha)}{\nabla}_u w,\overset{(-\alpha)}{\nabla}_v z])\end{aligned}\tag{4.10}$$

for $u=(U_1,u_2)$, $v=(V_1,v_2)$, $w=(W_1,w_2)$ and $z=(Z_1,z_2)\in\mathscr{H}$. Note that

$$
\begin{aligned}
D_u(\mathscr{G}_\theta[\overset{(\alpha)}{\nabla}_v w, z]) &= D_u\{E_\theta[(D^2_{vw}(:l(\theta):) + \frac{1-\alpha}{2}D_v(:l(\theta):)D_w(:l(\theta):))D_z(:l(\theta):)]\} \\
&= E_\theta[D^3_{uvw}(:l(\theta):)D_z(:l(\theta):) + D^2_{vw}(:l(\theta):)D^2_{uz}(:l(\theta):) \\
&\qquad + D_u(:l(\theta):)D^2_{vw}(:l(\theta):)D_z(:l(\theta):)] \\
&\quad + \frac{1-\alpha}{2}E_\theta[D^2_{uv}(:l(\theta):)D_w(:l(\theta):)D_z(:l(\theta):) \\
&\qquad + D_v(:l(\theta):)D^2_{uw}(:l(\theta):)D_z(:l(\theta):) \\
&\qquad + D_v(:l(\theta):)D_w(:l(\theta):)D^2_{uz}(:l(\theta):) \\
&\qquad + D_u(:l(\theta):)D_v(:l(\theta):)D_w(:l(\theta):)D_z(:l(\theta):)].
\end{aligned}
$$

A little tedious computations tells that

$$
D_u(\mathscr{G}_\theta[\overset{(\alpha)}{\nabla}_v w, z]) - D_v(\mathscr{G}_\theta[\overset{(\alpha)}{\nabla}_u w, z]) = 0. \tag{4.11}
$$

Let us denote

$$
\begin{cases}
\tilde{e}_j = (I+A)^{1/2}e_j \\
\tilde{E}_{ij} = \dfrac{1}{\sqrt{2}}(I+A)^{-1}\{e_i \otimes e_j + e_j \otimes e_i\},
\end{cases}
$$

then, we get that

$$
\begin{cases}
\mathscr{G}_\theta[0, \tilde{e}_i), (0, \tilde{e}_j)] = \delta_{ij} \\
\mathscr{G}_\theta[(\tilde{E}_{ij}, 0), (\tilde{E}_{kl}, 0)] = 1 \quad \text{when } \{i,j\} = \{k,l\} \\
\qquad\qquad\qquad\qquad\quad = 0 \quad \text{otherwise}
\end{cases}
$$

With this orthonormal basis of $\mathscr{H}$, we can observe that

$$
\begin{aligned}
\mathscr{G}_\theta[\overset{(\alpha)}{\nabla}_v w, \overset{(-\alpha)}{\nabla}_u z] &= \sum_{i \le j} \mathscr{G}_\theta[\overset{(\alpha)}{\nabla}_v w, (\tilde{E}_{ij}, 0)]\mathscr{G}_\theta[\overset{(-\alpha)}{\nabla}_u z, (\tilde{E}_{ij}, 0)] \\
&\quad + \sum_j \mathscr{G}_\theta[\overset{(\alpha)}{\nabla}_v w, (0, \tilde{e}_j)]\mathscr{G}_\theta[\overset{(-\alpha)}{\nabla}_u z, (0, \tilde{e}_j)].
\end{aligned} \tag{4.12}
$$

Now, it is easy to deduce that

$$
R^\alpha_\theta[z; u, v, w] = -\frac{1-\alpha^2}{4}((I+A)(U_1 v_2 - V_1 u_2), (Z_1 w_2 - W_1 z_2))_H \tag{4.13}
$$

from (4.10–12) and (1.7). Of course, (4.13) vanishes when $\alpha \pm 1$. Let us compute the sectional curvature $K_\theta^{(0)}[\cdot]$ for the Levi-Civita connection $\overset{(0)}{\nabla}$ (cf. [5]). We have

$$K_\theta^{(0)}[\{\tilde{e}_i, \tilde{e}_j\}] = R_\theta^{(0)}[\tilde{e}_i; \tilde{e}_i, \tilde{e}_j, \tilde{e}_j] = 0$$

$$K_\theta^{(0)}[\{\tilde{E}_{ij}, \tilde{e}_k\}] = R_\theta^{(0)}[\tilde{e}_k; \tilde{e}_k, \tilde{E}_{ij}, \tilde{E}_{ij}] = -\frac{1}{2} \quad \text{when } k \in \{i, j\}$$

$$= 0 \quad \text{otherwise}$$

$$K_\theta^{(0)}[\{\tilde{E}_{ij}, \tilde{E}_{kl}\}] = R_\theta^{(0)}[\tilde{E}_{ij}; \tilde{E}_{ij}, \tilde{E}_{kl}, \tilde{E}_{kl}] = 0,$$

hence, Theorem 3 is established. ■

5. An Example —Linear Gaussian Diffusions—.

Let $(\Omega, \mathscr{F}, \mu, \{\mathscr{F}_t\}_{0 \le t \le T})$ be a probability space with $\{\mathscr{F}_t\}_{0 \le t \le T}$ increasing, right continuous family of sub σ-algebras and consider a 1-dimensional linear Gaussian diffusion $X = \{X_t\}_{0 \le t \le T}$ on it which is defined by the following stochastic differential equation.

$$\begin{cases} dX_t(\omega) = dB_t(\omega) + (a(t)X_t(\omega) + b(t))dt & (0 \le t \le T) \\ X_0(\omega) = x \in \boldsymbol{R} \end{cases}$$

where $dB_t(\omega)$ denotes the Wiener integral. We will deal with only 1-dimensional diffusion to simplify the discussion and notations. Of course, similar results are obtained in general multi dimensional cases. By Itô formula, the explicit expression of X is given by

$$X_t(\omega) = \Phi(t)\{x + \int_0^t \Phi^{-1}(s)b(s)ds + \int_0^t \Phi^{-1}(s)dB_s(\omega)\} \tag{5.1}$$

where $\Phi(t) = \exp\{\int_0^t a(s)ds\}$. As stated in Introduction, we will regard (a, b) as parameters, so let us denote the law of the above $X = X(a, b)$ on the space $C([0, T], \boldsymbol{R})$ by $P_T^{(a,b)}$. When $a = b = 0$, we will use the notation P_T instead of $P_T^{(0,0)}$. If $a(t)$ and $b(t) \in C([0, T], \boldsymbol{R})$, immedeately by Cameron-Martin-Maruyama-Girsanov formula, we obtain the absolute continuity of $P_T^{(a,b)}$ with respect to P_T and the formula of the Radon-Nikodym derivative (cf. [4])

$$\frac{dP_T^{(a,b)}}{dP_T}(B(\omega)) = \exp\{l_T(B(\omega); a, b)\}$$

where $l_T(B(\omega)); a, b) = \int_0^T (a(t)B_t(\omega) + b(t))dB_t(\omega) - \frac{1}{2}\int_0^T (a(t)B_t(\omega) + b(t))^2 dt$. By the Remark of Lemma 3, we have

$$H(P_T | P_T^{(a,b)}) < \infty \Leftrightarrow P_T \sim P_T^{(a,b)},$$

so, it is easy to see that

$$P_T^{(a,b)} \sim P_T \Leftrightarrow (a, b) \in L^2([0, T], tdt) \times L^2([0, T], dt).$$

We will set

$$S_T = \{P_T^{(a,b)}; (a, b) \in L^2([0, T], tdt) \times L^2([0,T], dt)\},$$

then, we can introduce the Hilbert-Riemannian structure associated with the Fisher information on S_T in the similar way as Section 4. Let us take

$$\rho(t) = (m(t), \sigma^2(t)) = (E[X_t], E[(X_t - m(t))^2])$$

as a new coordinate of S_T and compute the Fisher information and α-affine connections in ρ-coordinate. By (5.1), we get

$$\begin{cases} m(t) = \Phi(t)\{x + \int_0^t \Phi^{-1}(s)b(s)ds\} \\ \sigma^2(t) = \Phi(t)^2\{\int_0^t \Phi^{-2}(s)ds\}, \end{cases}$$

so, $m(t)$ and $\sigma^2(t)$ satisfy the following ordinary differential equations

$$\begin{cases} \dfrac{d}{dt}m(t) = a(t)m(t) + b(t) \\ \dfrac{d}{dt}\sigma^2(t) = 2a(t)\sigma^2(t) + 1. \end{cases}$$

Therefore,

$$\begin{cases} a(t) = a(\rho(t)) = \dfrac{\frac{d}{dt}(\sigma^2(t)) - 1}{2\sigma^2(t)} \\ b(t) = b(\rho(t)) = \dfrac{d}{dt}(m(t)) - \dfrac{\frac{d}{dt}(\sigma^2(t)) - 1}{2\sigma^2(t)} m(t). \end{cases}$$

Let us denote $l_T(\rho) = l_T(a(\rho), b(\rho))$ and define the Fisher information and the

α-affine connections by the formulas:

$$\mathscr{G}_\rho^T[u,v]=E_\rho[D_u l_T(\rho)D_v l_T(\rho)]$$

$$\overset{(\alpha)}{\Gamma}_{\rho,T}[w;u,v]=E_\rho[D_w l_T(\rho)(D^2_{uv}l_T(\rho)+\frac{1-\alpha}{2}D_u l_T(\rho)D_v l_T(\rho))]$$

for $u=(u_1,u_2)$, $v=(v_1,v_2)$ and $w=(w_1,w_2)\in(C^1[0,T])^2$ and $D.l(\rho)$ denotes the Gateaux derivative of $l(\rho)$. Straightfoward calculations lead us to Theorem 4 in Introduction.

Finally, let us observe the behavior of the scaled diffusion $X^\varepsilon(t):=\varepsilon X(\frac{t}{\varepsilon^2})$ ($\varepsilon>0$) as $\varepsilon\to 0$. We set

$$\begin{cases} a_1(t)=t^{-k} & \text{where } k>1,\\ b_1(t)=t^{-1} & \text{where } l>1/2\end{cases}$$

and consider the Gaussian diffusion $X(a_1,b_1)$. Note that $P_T^{(a_1,b_1)}\in S_T$ for any $T>0$, since $(a_1,b_1)\in L^2(\boldsymbol{R},tdt)\times L^2(\boldsymbol{R},dt)$. For $\varepsilon>0$, the scaled diffusion $X^\varepsilon(a_1,b_1):=\{\varepsilon X(\frac{t}{\varepsilon^2})\}_{t\geq 0}$ satisfies

$$\begin{cases} dX_t^\varepsilon(\omega)=\varepsilon dB_{t/\varepsilon^2}(\omega)+(\varepsilon^{k-1}a_1(t)X_t^\varepsilon(\omega)+\varepsilon^{2l-1}b_1(t))dt & (0\leq t\leq T)\\ X_0^\varepsilon=x, \end{cases}$$

so, the law of $X^\varepsilon(a_1,b_1)$ is equal to $P_T^{(\varepsilon^{k-1}a_1,\varepsilon^{2l-1}b_1)}$. Obviously, $P_T^{(\varepsilon^{k-1}a_1,\varepsilon^{2l-1}b_1)}$ goes to P weakly as $\varepsilon\to 0$ and if we compute the covariance of the functional $\int_0^T f(t)dX_t^\varepsilon$ for any test function $f\in C^\infty([0,T])$, we get

$$E[(\int_0^T f(t)dX_t^\varepsilon)^2]=E[(\int_0^T f(t)dX_t^\varepsilon)^2\exp\{l_T(\varepsilon^{k-1}a_1,\varepsilon^{2l-1}b_1)\}]. \tag{5.2}$$

In (a,b)-coordinate on S_T, we can observe that

$$D_{(u_1,u_2)}l_T(a,b)=\int_0^T(u_1(t)B_t+u_2(t))(dB_t-(a(t)B_t+b(t))dt)$$

$$D^2_{(u_1,u_2)(v_1,v_2)}l_T(a,b)=\int_0^T(u_1(t)B_t+u_2(t))(v_1(t)B_t+v_2(t))dt$$

et cetera for any (a,b), (u_1,u_2) and $(v_1,v_2)\in L^2([0,T],tdt)\times L^2([0,T],dt)$. Therefore, (5.2) is equal to

$$E[(D_{(0,f)}l_T(0,0))^2\exp\{l_T(\varepsilon^{k-1}a^1,\varepsilon^{2l-1}b_1)\}]$$

and it is easy to observe that

$$E[(\int_0^T f(t)dX_t^\varepsilon)^2] = E[(D_{(0,f)}l_T(0,0))^2\{1+\varepsilon^{k-1}D_{(a_1,0)}l_T(0,0)$$

$$+\varepsilon^{2l-1}D_{(0,b_1)}l_T(0,0)\}] + o(\varepsilon^{((k-1)\vee(2l-1))})$$

$$= \mathscr{G}^T_{(0,0)}[(0,f),(0,f)]$$

$$+\varepsilon^{k-1}\{\overset{(-1)}{\Gamma}_{(0,0),T}[(a_1,0);(0,f),(0,f)]$$

$$-\overset{(1)}{\Gamma}_{(0,0),T}[(a_1,0);(0,f),(0,f)]\}$$

$$+\varepsilon^{2l-1}\{\overset{(-1)}{\Gamma}_{(0,0),T}[(0,b_1);(0,f),(0,f)]$$

$$-\overset{(1)}{\Gamma}_{(0,0),T}[(0,b_1);(0,f),(0,f)]\}$$

$$+o(\varepsilon^{((k-1)\vee(2l-1))}).$$

References

[1] S-I. Amari: Differential Geometrical Method in Statistics, Lecture Note in Statistics, Vol.**28**, Springer, 1985.

[2] S-I. Amari, O.E. Barndorff-Nielsen, R.E. Kass, S.L. Lauritzen and C.R. Rao: Differential Geometry in Statistical Inference, I.M.S.Lecture Note, Vol.**10**, 1987.

[3] J.L. Doob: Stochastic Processes, John Wiley & Sons, 1953.

[4] N. Ikeda and S. Watanabe: Stochastic Differential Equations and Diffusion Processes, North-Holland, Kodansha, 1981.

[5] S. Kobayashi and K. Nomizu: Foundations of Differential Geometry .1, Interscience Press, 1963.

[6] H.H. Kuo: Gaussian Measures in Banach Spaces, Lecture Note in Mathematics, Vol.463, Springer, 1975.

[7] S. Lang: Differential Manifolds, Addison-Wesley, 1972.

[8] B. Simon: Trace Ideals and Its Applications, London Mathematics Society Lecture Note Series **35**, Cambridge University Press, 1979.

[9] Y. Yamazaki: Measures on Infinite Dimensional Spaces 2, (in Japanese), Kinokuniya-Shoten, 1978.

Department of Pure and Applied Sciences
University of Tokyo
Komaba, Meguro-ku, Tokyo 153, Japan