Advanced Studies in Pure Mathematics 57, 2010
Probabilistic Approach to Geometry
pp. 463–472

On Wasserstein geometry of Gaussian measures

Asuka Takatsu

Abstract.

The space of Gaussian measures on a Euclidean space is geodesically convex in the L^2-Wasserstein space. This is a finite dimensional manifold since Gaussian measures are parameterized by means and covariance matrices. By restricting to the space of Gaussian measures inside the L^2-Wasserstein space, we manage to provide detailed descriptions of the L^2-Wasserstein geometry from a Riemannian geometric viewpoint. We obtain a formula for the sectional curvatures of the space of Gaussian measures, which is written out in terms of the eigenvalues of the covariance matrix.

§1. Introduction

For a vector m in $\mathbb{R}^d$ and a symmetric positive definite matrix V of size d, a Gaussian measure $N(m,V)$ with mean m and covariance matrix V is an absolutely continuous probability measure on $\mathbb{R}^d$ with respect to the Lebesgue measure dx whose Radon–Nikodym derivative is given by

$$\frac{dN(m,V)}{dx} = \frac{1}{\sqrt{\det(2\pi V)}} \exp\left[-\frac{1}{2}\langle x-m, V^{-1}(x-m)\rangle\right].$$

We denote by $\mathcal{N}^d$ the space of Gaussian measures on $\mathbb{R}^d$. Since Gaussian measures are completely determined by the means and the covariance matrices, $\mathcal{N}^d$ is identified with $\mathbb{R}^d \times \mathrm{Sym}^+(d,\mathbb{R})$, where $\mathrm{Sym}^+(d,\mathbb{R})$ is the set of symmetric positive definite matrices of size d.

Received January 15, 2009.
Revised May 1, 2009.
2000 *Mathematics Subject Classification.* 60D05, 28A33.
Key words and phrases. Wasserstein space, Gaussian measures.
This work was partially supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

Let $\mathcal{P}_2^{\mathrm{ac}}$ be the set of absolutely continuous probability measures on $\mathbb{R}^d$ whose second moments are finite. Then for $\mu,\ \nu \in \mathcal{P}_2^{\mathrm{ac}}$, L^2-Wasserstein distance between μ and ν is defined by

$$W_2(\mu,\nu)^2 = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} d(x,y)^2 d\pi(x,y),$$

where the infimum is taken over all Borel probability measures on $\mathbb{R}^d\times\mathbb{R}^d$ whose marginals are μ and ν. Then $(\mathcal{P}_2^{\mathrm{ac}}, W_2)$ is a geodesic space and all geodesics are given by push-forward measures (see [4]). Although it is usually difficult to obtain the concrete value of the L^2-Wasserstein distance, the L^2-Wasserstein distance between Gaussian measures can be explicitly computed by several authors; Dowson–Landau [1], Givens–Short [3], Knott–Smith [5] and Olkin–Pukelsheim [6]: For $N(m,V)$ and $N(n,U)$, we get

$$W_2(N(m,V),N(n,U))^2 = |m-n|^2 + \mathrm{tr}V + \mathrm{tr}U - 2\mathrm{tr}\left(U^{\frac{1}{2}}VU^{\frac{1}{2}}\right)^{\frac{1}{2}}.$$

It implies that variations of mean and covariance do not interact, and the geometry on mean variations is trivial. Then it suffices to consider the geometry on covariance matrix variations. We use $\mathcal{N}_0^d$ for the set of all Gaussian measures with mean 0. We denote by $N(V)$ the Gaussian measure with mean 0 and covariance matrix V.

The tangent space to $\mathcal{N}_0^d$ at each point can be regarded as $\mathrm{Sym}(d,\mathbb{R})$, where $\mathrm{Sym}(d,\mathbb{R})$ is the set of symmetric matrices of size d. McCann [4] showed that $\mathcal{N}_0^d$ is geodesically convex in $\mathcal{P}_2^{\mathrm{ac}}$ and a geodesic $\exp_{N(V)} tX$ from $N(V)$ with direction $X \in \mathrm{Sym}(d,\mathbb{R})$ is given by

$$\exp_{N(V)} tX = N(U_t), \text{ where } U_t = [(1-t)E + tX]V[(1-t)E + tX].$$

By restricting to a geodesically convex submanifold $\mathcal{N}_0^d$ of $\mathcal{P}_2^{\mathrm{ac}}$, we obtain a formula for sectional curvatures of $\mathcal{N}_0^d$. This coincides with a formal expressions of sectional curvatures of $\mathcal{P}_2^{\mathrm{ac}}$ given by Otto [7].

Theorem 1.1. [8, Theorem1.1] *For an orthogonal matrix P and positive numbers $\{\lambda_i\}_{i=1}^d$, we set $V = P\mathrm{diag}[\lambda_1,\dots,\lambda_d]\,{}^{\mathsf{T}}P$, where ${}^{\mathsf{T}}P$ is the transpose matrix of P. Then the tangent space to $\mathcal{N}_0^d$ at $N(V)$ is spanned by*

$$\left\{ e_+ = \frac{P\,(E_{11}+E_{dd})\,{}^{\mathsf{T}}P}{\sqrt{\lambda_1+\lambda_d}},\ e_{ij} = \frac{P\,(E_{ii}-E_{jj})\,{}^{\mathsf{T}}P}{\sqrt{\lambda_i+\lambda_j}},\ f_{ij} = \frac{P\,(E_{ij}+E_{ji})\,{}^{\mathsf{T}}P}{\sqrt{\lambda_i+\lambda_j}} \right\},$$

where E_{ij} is an (i,j)-matrix unit, whose (i,j)-component is 1, 0 elsewhere. Then we obtain the following expressions of the sectional curvatures K with respect to the vectors:

$$
\begin{array}{lll}
(1) & K(e_+, e_{ij}) = 0 & \\
(2) & K(e_+, f_{1d}) = 0 & \\
(3) & K(e_+, f_{ij}) = \dfrac{3\lambda_i\lambda_j}{(\lambda_i+\lambda_j)^2(\lambda_1+\lambda_d)} & (i = 1 \text{ or } j = d) \\
(4) & K(e_+, f_{kl}) = 0 & (1 < k < l < d) \\
(5) & K(e_{ij}, e_{kl}) = 0 & \\
(6) & K(e_{ij}, f_{kl}) = 0 & (\{i,j\} \cap \{k,l\} = \emptyset) \\
(7) & K(e_{ik}, f_{ij}) = \dfrac{3\lambda_i\lambda_j}{(\lambda_i+\lambda_j)^2(\lambda_i+\lambda_k)} & (j \neq k) \\
(8) & K(e_{ij}, f_{ij}) = \dfrac{12\lambda_i\lambda_j}{(\lambda_i+\lambda_j)^3} & \\
(9) & K(f_{ij}, f_{kl}) = 0 & (\{i,j\} \cap \{k,l\} = \emptyset) \\
(10) & K(f_{ij}, f_{ik}) = \dfrac{3\lambda_j\lambda_k}{(\lambda_i+\lambda_j)(\lambda_j+\lambda_k)(\lambda_k+\lambda_i)} & (j \neq k).
\end{array}
$$

This shows that the sectional curvatures with respect to the vectors $\{e_+, e_{ij}, f_{ij}\}$ are non-negative and depend only on the eigenvalues of the covariance matrix. Moreover, the author [9] proved that $\mathcal{N}_0^d$ is really a space of non-negative curvature in a different method and the metric completion of $\mathcal{N}_0^d$ has a cone structure.

§2. Proof of Theorem 1.1

The author [8] has explicitly constructed a metric g on $\mathcal{N}_0^d$, which induces the L^2-Wasserstein distance. In the following, we proceed to calculate the sectional curvatures using the following lemmas. We omit some calculations and the comprehensive proof can be found in [8].

Lemma 2.1. [2, Theorem 3.68] *For a Riemannian manifold (M, g), suppose that $\{u, v\}$ is an orthonormal basis of a 2-plane in the tangent space at $p \in M$. Let $C_r(\theta) = \exp_p r(u\cos\theta + v\sin\theta)$, and $L(r)$ be the length of the curve C_r. Then the function $L(r)$ admits an asymptotic expansion*

$$L(r) = 2\pi r\left(1 - \frac{K(u,v)}{6}r^2 + o(r^2)\right), \quad \text{as } r \searrow 0.$$

Lemma 2.2. [8, Lemma 3.2] *For $A, B \in \{e_+, e_{ij}, f_{ij}\}$, $0 < r \ll 1$ and $\theta \in [0, 2\pi]$, $C_r(\theta) = \exp_{N(V)} r(\cos\theta \cdot A + \sin\theta \cdot B)$ is a Gaussian measure $N(X)$, where $X = X(r,\theta) = (x_{\alpha\beta})$ is given by*

$$X = [E + r(\cos\theta \cdot A + \sin\theta \cdot B)] \cdot V \cdot [E + r(\cos\theta \cdot A + \sin\theta \cdot B)], \tag{2.1}$$

where E is the identity matrix.

Lemma 2.3. [8, Corollary 2.5] *For an orthogonal matrix P, let $\mathcal{N}_0^d(P)$ be a subset of $\mathcal{N}_0^d$ whose covariance matrices are diagonalized by P. Then $\mathcal{N}_0^d(P)$ is a geodesically convex and flat submanifold of $\mathcal{N}_0^d$.*

Lemma 2.4. [8, Lemma 3.3] *Let $M \in \mathrm{Sym}(2, \mathbb{R})$, then we obtain $(\mathrm{tr}M)^2 = \mathrm{tr}M^2 + 2\det M$.*

Proof of (1) and (5)

If we choose $A = e_+, B = e_{ij}$ or $A = e_{ij}, B = e_{kl}$ in (2.1), then $N(X)$ belongs to $\mathcal{N}_0^d(P)$. By Lemma 2.3, the curvatures vanish.

A strategy for proving the remaining case is as follows. We first calculate

$$W(\theta_0, \theta) = W_2(C_r(\theta_0), C_r(\theta))^2 \text{ and } W(\theta_0) = \lim_{\theta \to \theta_0} \frac{W(\theta_0, \theta)}{\theta^2}$$

by using Lemma 2.2 and Lemma 2.4. Then we get

$$L(r) = \int_0^{2\pi} W(\theta)^{\frac{1}{2}} d\theta.$$

Finally we use Lemma 2.1 to obtain the expression of the sectional curvatures. Without loss of generality, we may assume $P = E$, because $W(\theta_0, \theta)$ is invariant under taking conjugation with P.

For $1 \le i, j \le d$, $\theta \in [0, 2\pi]$ and sufficiently small $r > 0$, we set

$$c_{ij}(r, \theta) = \frac{r\cos\theta}{\sqrt{\lambda_i + \lambda_j}}, \quad s_{ij}(r, \theta) = \frac{r\sin\theta}{\sqrt{\lambda_i + \lambda_j}}.$$

Proof of (2) and (8)

For (2), we take $A = f_{1d}$, $B = e_+$ and $I = \{1, d\}$, whereas, for (8), take $A = f_{ij}$, $B = e_{ij}$ and $I = \{i, j\}$. Then we notice that for any $\alpha, \beta \notin I$, (α, β)-components of X are independent of the variables r and θ. If we set

$$\widetilde{X}(\theta) = \begin{pmatrix} x_{\alpha\alpha} & x_{\alpha\beta} \\ x_{\beta\alpha} & x_{\beta\beta}, \end{pmatrix}$$

we obtain

$$W(\theta_0, \theta) = \mathrm{tr}\widetilde{X}(\theta_0) + \mathrm{tr}\widetilde{X}(\theta) - 2\mathrm{tr}\left(\widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}(\theta)\widetilde{X}(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}},$$

where $\{\alpha, \beta\} = I$. For (2), using Lemma 2.4, we conclude

$$W(\theta_0, \theta) = 4r^2\sin^2(\theta - \theta_0) \text{ and } \lim_{\theta \to \theta_0} \frac{W(\theta_0, \theta)}{(\theta - \theta_0)^2} = r^2.$$

It follows that $L(r) = 2\pi r$, proving $K(e_+, f_{1d}) = 0$.

For (8), in a similar way, we have

$$W(\theta_0, \theta) = 4r^2 \sin^2 \frac{1}{2}(\theta - \theta_0) - \frac{4r^4 \lambda_i \lambda_j \sin^2(\theta - \theta_0)}{(\lambda_i + \lambda_j)^2 a_r(\theta_0, \theta)} + o(|\theta - \theta_0|^2),$$

where

$$\begin{aligned} a_r(\theta_0, \theta) =& \lambda_i \left[(1 + c_{ij}(r, \theta_0))(1 + c_{ij}(r, \theta)) + s_{ij}(r, \theta_0) s_{ij}(r, \theta)\right] \\ &+ \lambda_j \left[(1 - c_{ij}(r, \theta_0))(1 - c_{ij}(r, \theta)) + s_{ij}(r, \theta_0) s_{ij}(r, \theta)\right]. \end{aligned}$$

Since the limit of $a_r(\theta_0, \theta)$ exists as $\theta \to \theta_0$ and

$$a_r(\theta_0, \theta_0) = (\lambda_i + \lambda_j)(1 + r^2) + 2(\lambda_i - \lambda_j) r \cos\theta_0,$$

we have

$$\lim_{\theta \to \theta_0} \frac{W(\theta_0, \theta)}{(\theta - \theta_0)^2} = r^2 - \frac{4r^4 \lambda_i \lambda_j}{(\lambda_i + \lambda_j)^2 a_r(\theta_0, \theta_0)}.$$

It follows that

$$L(r) = \int_0^{2\pi} r \left(1 - \frac{1}{2} \frac{4r^2 \lambda_i \lambda_j}{(\lambda_i + \lambda_j)^2 a_r(\theta, \theta)} + o(r^2)\right) d\theta.$$

Because $a_0(\theta, \theta) = \lambda_i + \lambda_j$, using Lemma 2.1 and the bounded convergence theorem, we obtain

$$K(e_{ij}, f_{ij}) = \frac{12 \lambda_i \lambda_j}{(\lambda_i + \lambda_j)^3}.$$

Proof of (3) and (7)

For (3), assuming $i = 1$, take $A = e_+$, $B = f_{1j}$ and $I = \{1, j, d\}$, whereas, for (7), assuming $j < k$, take $A = e_{ik}$, $B = f_{ij}$ and $I = \{i, j, k\}$. Because for any $\alpha, \beta \notin I$, (α, β)-components of X are independent of the variables r and θ, we obtain

$$\begin{aligned} &W(\theta_0, \theta) \\ =& \mathrm{tr}\widetilde{X}(\theta_0) + \mathrm{tr}\widetilde{X}(\theta) - 2\mathrm{tr}\left(\widetilde{X}(\theta_0)^{\frac{1}{2}} \widetilde{X}(\theta) \widetilde{X}(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}} \\ =& \mathrm{tr}\widetilde{Y}(\theta_0) + \mathrm{tr}\widetilde{Y}(\theta) - 2\mathrm{tr}\left(\widetilde{Y}(\theta_0)^{\frac{1}{2}} \widetilde{Y}(\theta) \widetilde{Y}(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}} + \frac{r^2 \lambda_\gamma (\cos\theta - \cos\theta_0)^2}{\lambda_\alpha + \lambda_\gamma}, \end{aligned}$$

where

$$\widetilde{X}(\theta) = \begin{pmatrix} x_{\alpha\alpha} & x_{\alpha\beta} & x_{\alpha\gamma} \\ x_{\beta\alpha} & x_{\beta\beta} & x_{\beta\gamma} \\ x_{\gamma\alpha} & x_{\gamma\beta} & x_{\gamma\gamma} \end{pmatrix} = \begin{pmatrix} \widetilde{Y}(\theta) & {}^\top\mathbf{0} \\ \mathbf{0} & \lambda_\gamma (1 + c_{\alpha\gamma}(r, \theta))^2 \end{pmatrix}, \quad \mathbf{0} = (0, 0)$$

and $\{\alpha,\beta,\gamma\}=I$. Using Lemma 2.4, we conclude

$$W(\theta_0,\theta)=4r^2\sin^2\frac{1}{2}(\theta-\theta_0)-\frac{r^4}{a_r(\theta,\theta_0)}\frac{\lambda_\alpha\lambda_\beta\sin^2(\theta-\theta_0)}{(\lambda_\alpha+\lambda_\beta)(\lambda_\alpha+\lambda_\gamma)}+o(\theta^2),$$

where

$$a_r(\theta,\theta_0)=\lambda_\alpha(1+c_{\alpha\gamma}(r,\theta_0))(1+c_{\alpha\gamma}(r,\theta))+r^2\sin\theta_0\sin\theta+\lambda_\beta.$$

Since the limit of $a_r(\theta_0,\theta)$ exists as $\theta\to\theta_0$, we have

$$\lim_{\theta\to\theta_0}\frac{W(\theta_0,\theta)}{(\theta-\theta_0)^2}=r^2-\frac{r^4}{a_r(\theta_0,\theta_0)}\frac{\lambda_\alpha\lambda_\beta}{(\lambda_\alpha+\lambda_\beta)(\lambda_\alpha+\lambda_\gamma)}.$$

It follows that

$$L(r)=\int_0^{2\pi}r\left(1-\frac{1}{2}\frac{r^2}{a_r(\theta,\theta)}\frac{\lambda_\alpha\lambda_\beta}{(\lambda_\alpha+\lambda_\beta)(\lambda_\alpha+\lambda_\gamma)}+o(r^2)\right)d\theta.$$

Because $a_0(\theta,\theta)=(\lambda_\alpha+\lambda_\beta)$, using Lemma 2.1 and the bounded convergence theorem, we obtain

$$K(A,B)=\frac{3\lambda_\alpha\lambda_\beta}{(\lambda_\alpha+\lambda_\beta)^2(\lambda_\alpha+\lambda_\gamma)}.$$

We can prove the case of $i\neq 1$ and $j=d$ in a similar way.

Proof of (4), (6) and (9)

We take (A,B) in (2.1) as (e_+,f_{kl}) ($\{1,d\}\cap\{k,l\}=\emptyset$), (e_{ij},f_{kl}) ($\{i,j\}\cap\{k,l\}=\emptyset$) and (f_{ij},f_{kl}) ($\{i,j\}\cap\{k,l\}=\emptyset$) in this order. Moreover we set $I=\{1,d\}$ in the case (4) and $I=\{i,j\}$ in the case of (6) and (9). We notice that for any $\alpha,\beta\notin I$, (α,β)-components of X are independent of the variables r and θ. If we set

$$\widetilde{X}_c(\theta)=\begin{pmatrix}x_{\alpha\alpha} & x_{\alpha\beta}\\ x_{\beta\alpha} & x_{\beta\beta}\end{pmatrix},\ \widetilde{X}_s(\theta)=\begin{pmatrix}x_{kk} & x_{kl}\\ x_{lk} & x_{ll}\end{pmatrix},$$

we obtain

$$\begin{aligned}W(\theta_0,\theta)=&\operatorname{tr}\widetilde{X}_c(\theta_0)+\operatorname{tr}\widetilde{X}_c(\theta)-2\operatorname{tr}\left(\widetilde{X}_c(\theta_0)^{\frac12}\widetilde{X}_c(\theta)\widetilde{X}_c(\theta_0)^{\frac12}\right)^{\frac12}\\&+\operatorname{tr}\widetilde{X}_s(\theta_0)+\operatorname{tr}\widetilde{X}_s(\theta)-2\operatorname{tr}\left(\widetilde{X}_s(\theta_0)^{\frac12}\widetilde{X}_s(\theta)\widetilde{X}_s(\theta_0)^{\frac12}\right)^{\frac12},\end{aligned}$$

where $\{\alpha,\beta\}=I$. Using Lemma 2.4, we conclude

$$\lim_{\theta\to\theta_0}\frac{W(\theta_0,\theta)}{(\theta-\theta_0)^2}=\left(\lim_{\theta\to\theta_0}\frac{r\sin(\theta-\theta_0)}{\theta-\theta_0}\right)^2=r^2.$$

It follows that $L(r)=2\pi r$ and $K(A,B)=0$.

Proof of (10)

Without loss of generality, we may assume $j < k$. Taking A and B as f_{ij} and f_{ik} in (2.1) respectively. We notice that for any $\alpha, \beta \notin \{i, j, k\}$, (α, β)-components of X are independent of the variables r and θ. If we set

$$\widetilde{X}(\theta) = \begin{pmatrix} x_{ii} & x_{ij} & x_{ik} \\ x_{ji} & x_{jj} & x_{jk} \\ x_{ki} & x_{kj} & x_{kk} \end{pmatrix},$$

we obtain

$$W(\theta_0, \theta) = \mathrm{tr}\widetilde{X}(\theta_0) + \mathrm{tr}\widetilde{X}(\theta) - 2\mathrm{tr}\left(\widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}(\theta)\widetilde{X}(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}}. \tag{2.2}$$

For the value of the last term in (2.2), Lemma 2.4 can not be used as the size of matrices is 3×3.

We define some notations:

$$\begin{aligned}
A &= A_{\theta_0}(\theta) = \widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}(\theta)\widetilde{X}(\theta_0)^{\frac{1}{2}} \\
B &= B_{\theta_0}(\theta) = \left(\widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}(\theta)\widetilde{X}(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}} \\
&\{\sigma_\alpha = \sigma_{\theta_0}(\theta)_\alpha\}_{\alpha=1}^3 : \text{eigenvalues of } B \\
f_{\theta_0}(\theta) &= \mathrm{tr}B = \sigma_1 + \sigma_2 + \sigma_3 \\
g_{\theta_0}(\theta) &= \mathrm{tr}A = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 \\
h_{\theta_0}(\theta) &= \sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_3\sigma_1 \\
\varphi_{\theta_0}(\theta) &= \sigma_1^2\sigma_2^2 + \sigma_2^2\sigma_3^2 + \sigma_3^2\sigma_1^2 \\
D_{\theta_0}(\theta) &= \det B = (\det A)^{\frac{1}{2}} = \sigma_1\sigma_2\sigma_3
\end{aligned}$$

Rewriting (2.2) with the Taylor approximation of $f_{\theta_0}(\cdot)$ at θ_0, we obtain

$$W(\theta_0, \theta) = -2f'_{\theta_0}(\theta_0)(\theta - \theta_0) - f''_{\theta_0}(\theta_0)(\theta - \theta_0)^2 + o(|\theta - \theta_0|^2).$$

Since we can get the values of g, φ and D without information of $X^{1/2}$, we compute f' and f'' by using these values.

We first calculate $f'_{\theta_0}(\theta_0)$. Differentiating $B_{\theta_0}(\theta) \cdot B_{\theta_0}(\theta) = A_{\theta_0}(\theta)$ with respect to θ, we have

$$B'_{\theta_0}(\theta)B_{\theta_0}(\theta) + B_{\theta_0}(\theta)B'_{\theta_0}(\theta) = A'_{\theta_0}(\theta) = \widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}'(\theta)\widetilde{X}(\theta_0)^{\frac{1}{2}}.$$

After multiplying $B_{\theta_0}(\theta)^{-1}$ from the left, taking the trace gives

$$\mathrm{tr}B'_{\theta_0}(\theta_0) + \mathrm{tr}(B_{\theta_0}(\theta_0)B'_{\theta_0}(\theta_0)B_{\theta_0}(\theta_0)^{-1}) = 2f'_{\theta_0}(\theta_0)$$

at $\theta = \theta_0$. Because $\mathrm{tr}\widetilde{X}(\theta)$ is constant, at $\theta = \theta_0$ the right hand side is equal to

$$\mathrm{tr}(\widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}'(\theta_0)\widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}(\theta_0)^{-1}) = \mathrm{tr}\widetilde{X}'(\theta_0) = \left(\mathrm{tr}\widetilde{X}(\theta)\right)'\Big|_{\theta=\theta_0} = 0.$$

Therefore we conclude

$$f'_{\theta_0}(\theta_0) = 0. \tag{2.3}$$

Next we compute $f''_{\theta_0}(\theta_0)$. Differentiating $f^2 = g + 2h$ at $\theta = \theta_0$, we have

$$2f_{\theta_0}(\theta_0)f'_{\theta_0}(\theta_0) = g'_{\theta_0}(\theta_0) + 2h'_{\theta_0}(\theta_0),$$

proving $2h'_{\theta_0}(\theta_0) = -g'_{\theta_0}(\theta_0)$. Differentiating once more,

$$f''_{\theta_0}(\theta) = -\frac{f'_{\theta_0}(\theta)}{2f_{\theta_0}(\theta)^2}\left(g'_{\theta_0}(\theta) + 2h'_{\theta_0}(\theta)\right) + \frac{g''_{\theta_0}(\theta) + 2h''_{\theta_0}(\theta)}{2f_{\theta_0}(\theta)}.$$

Because of (2.3), we get at $\theta = \theta_0$

$$f''_{\theta_0}(\theta_0) = \frac{g''_{\theta_0}(\theta_0) + 2h''_{\theta_0}(\theta_0)}{2f_{\theta_0}(\theta_0)}. \tag{2.4}$$

We compute directly

$$g_{\theta_0}(\theta) = \sum_{\alpha,\beta\in\{i,j,k\}} x_{\alpha\beta}(\theta_0)x_{\beta\alpha}(\theta).$$

This enables us to get the derivatives of $g_{\theta_0}(\theta)$. Because $B_{\theta_0}(\theta_0) = X(\theta_0)$, using the relation

$$\det(tE - B) = t^3 - t^2\cdot f + t\cdot h - D,$$

we have

$$h_{\theta_0}(\theta_0) = \sum_{\substack{\alpha,\beta\in\{i,j,k\}\\ \alpha\neq\beta}} \left(x_{\alpha\alpha}(\theta_0)x_{\beta\beta}(\theta_0) - x_{\alpha\beta}(\theta_0)^2\right).$$

While it is hard to compute $B_{\theta_0}(\theta)$ directly, it is also hard to know the values of $h_{\theta_0}(\theta)$. We want to derive $h''_{\theta_0}(\theta)$ without the information of $B_{\theta_0}(\theta)$. So differentiating $h^2 = \varphi + 2Df$ twice at $\theta = \theta_0$, we have

(2.5)

$$h''_{\theta_0}(\theta_0) = -\frac{g'_{\theta_0}(\theta_0)^2}{4h_{\theta_0}(\theta_0)} + \frac{\varphi''_{\theta_0}(\theta_0) + 2D''_{\theta_0}(\theta_0)f_{\theta_0}(\theta_0) + 2D_{\theta_0}(\theta_0)f''_{\theta_0}(\theta_0)}{2h_{\theta_0}(\theta_0)}.$$

In order to analyze (2.5), we consider $D_{\theta_0}(\theta)$ and $\varphi_{\theta_0}(\theta)$. From the definition, we can compute $D_{\theta_0}(\theta)$ directly:

$$D_{\theta_0}(\theta) = \lambda_i\lambda_j\lambda_k\left[1 - c_{ij}(r,\theta_0)^2 - s_{ik}(r,\theta_0)^2\right]\left[1 - c_{ij}(r,\theta)^2 - s_{ik}(r,\theta)^2\right].$$

We next consider $\varphi_{\theta_0}(\theta)$. Using the equation

$$\begin{aligned}\det(tE - A) &= t^3 - t^2\cdot g(\theta) + t\cdot\varphi - D^2\\ &= \det\widetilde{X}(\theta_0)\cdot\det[t\widetilde{X}(\theta_0)^{-1} - \widetilde{X}(\theta)],\end{aligned}$$

we conclude

$$\varphi_{\theta_0}(\theta) = \det(\widetilde{X}(\theta_0)\widetilde{X}(\theta))\cdot\mathrm{tr}(Y(\theta_0)Y(\theta)),\ \text{where } Y(\theta) = \widetilde{X}(\theta)^{-1}.$$

We can obtain the value of $\varphi_{\theta_0}(\theta)$ since it depends only on $\widetilde{X}(\theta)$. Therefore we can now specify the value of $h''_{\theta_0}(\theta_0)$ in (2.5).

Inserting (2.5) into (2.4), we obtain

$$W(\theta_0,\theta) = -f''_{\theta_0}(\theta_0) + o(|\theta-\theta_0|^2) = -\frac{\beta_r(\theta_0)}{\alpha_r(\theta_0)} + o(|\theta-\theta_0|^2)$$

where

$$\begin{aligned}\alpha_r(\theta_0) =& 2\left[f_{\theta_0}(\theta_0)h_{\theta_0}(\theta_0) - D_{\theta_0}(\theta_0)\right]\\ \beta_r(\theta_0) =& h_{\theta_0}(\theta_0)g''_{\theta_0}(\theta_0) - \frac{1}{2}g'_{\theta_0}(\theta_0)^2 + \varphi''_{\theta_0}(\theta_0) + 2D''_{\theta_0}(\theta_0).\end{aligned}$$

Therefore we have

$$W(\theta_0) = \lim_{\theta\to\theta_0}\frac{W(\theta,\theta_0)}{\theta^2} = -\frac{\beta_r(\theta_0)}{\alpha_r(\theta_0)}.$$

If we set

$$\begin{aligned}L &= 2(\lambda_i+\lambda_j)(\lambda_j+\lambda_k)(\lambda_k+\lambda_i),\\ a &= \lambda_j^2 + \lambda_k^2 + 4\lambda_j\lambda_k + \lambda_i\lambda_j + \lambda_i\lambda_k + (\lambda_j-\lambda_k)(\lambda_j+\lambda_k+3\lambda_i)\cos\theta,\\ b &= (\lambda_j+\lambda_k)(\lambda_i+\lambda_j+\lambda_k) + (\lambda_j-\lambda_k)(\lambda_j+\lambda_k+3\lambda_i)\cos\theta,\end{aligned}$$

we have

$$L(r) = \int_0^{2\pi} r\left(1 + \frac{r^2(b-a)}{2(L+r^2a)} + o(r^2)\right)d\theta.$$

Using Lemma 2.1 and the bounded convergence theorem, we obtain

$$2\pi\frac{K(u,v)}{6} = \int_0^{2\pi}\lim_{r\searrow 0}\frac{a-b}{2(L+r^2a)}d\theta = 2\pi\frac{a-b}{2L},$$

which implies that

$$K(f_{ij}, f_{ik}) = \frac{3\lambda_k\lambda_j}{(\lambda_i + \lambda_j)(\lambda_j + \lambda_k)(\lambda_k + \lambda_i)}.$$

This completes the proof of Theorem 1.1. Q.E.D.

Acknowledgments. The author would like to express her gratitude to Professor Sumio Yamada, for his advice, support and encouragement.

References

[1] D. C. Dowson and B. V. Landau, The Fréchet distance between multivariate normal distributions, J. Multivariate Anal., **12** (1982), 450–455.

[2] S. Gallot, D. Hulin and J. Lafontaine, Riemannian Geometry. Third ed., Universitext, Springer-Verlag, Berlin, 2004.

[3] C. R. Givens and R. M. Shortt, A class of Wasserstein metrics for probability distributions, Michigan Math. J., **31** (1984), 231–240.

[4] R. J. McCann, A convexity principle for interacting gases, Adv. Math., **128** (1997), 153–179.

[5] M. Knott and C. S. Smith, On the optimal mapping of distributions, J. Optim. Theory Appl., **43** (1984), 39–49.

[6] I. Olkin and F. Pukelsheim, The distance between two random vectors with given dispersion matrices, Linear Algebra Appl., **48** (1982), 257–263.

[7] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations, **26** (2001), 101–174.

[8] A. Takatsu, On Wasserstein geometry of the space of Gaussian measures, preprint.

[9] A. Takatsu, Wasserstein geometry of Gaussian measures, preprint.

Mathematical Institute
Tohoku University
Sendai 980-8578
Japan

E-mail address: `sa6m21@math.tohoku.ac.jp`