Proofs of "An Information Geometry of Statistical Manifold Learning"

Ke Sun Stéphane Marchand-Maillet
Viper Group, Computer Vision & Multimedia Laboratory
University of Geneva, Switzerland
{Ke.Sun,Stephane.Marchand-Maillet}@unige.ch

Notations

$\delta_{i_1,\ldots,i_n} = 1$ if $i_1 = \cdots = i_n$; $\delta_{i_1,\ldots,i_n} = 0$ if otherwise.

Proof of Lemma 3

Proof. Consider the coordinate transformation $\theta_i = \log(p_i/p_0)$ introduced in section 1. The reverse mapping $\boldsymbol{\theta} \to (p_1, \ldots, p_n)$ is given by

$$p_\xi(\boldsymbol{\theta}) = \frac{\exp\theta_\xi}{\sum_{\imath=1}^n \exp\theta_\imath}.$$

Then,

$$\log p_\xi(\boldsymbol{\theta}) = \sum_{i=1}^n \theta_i \cdot \delta_{i\xi} - \log\left(\sum_{\imath=1}^n \exp\theta_\imath\right).$$

Hence,

$$\frac{\partial}{\partial\theta_i}\log p_\xi(\boldsymbol{\theta}) = \delta_{i\xi} - \frac{\exp\theta_i}{\sum_{\imath=1}^n \exp\theta_\imath} = \delta_{i\xi} - p_i(\boldsymbol{\theta}). \tag{21}$$

By definition of Fisher information metric (see section 1) and Eq.(21),

$$\begin{aligned}
\mathfrak{G}_{ij}(\boldsymbol{\theta}) &= \sum_{\xi=1}^n p_\xi(\boldsymbol{\theta})\left[\frac{\partial}{\partial\theta_i}\log p_\xi(\boldsymbol{\theta}) \cdot \frac{\partial}{\partial\theta_j}\log p_\xi(\boldsymbol{\theta})\right] = \sum_{\xi=1}^n p_\xi(\boldsymbol{\theta})\Big[\,(\delta_{i\xi} - p_i(\boldsymbol{\theta}))\cdot(\delta_{j\xi} - p_j(\boldsymbol{\theta}))\,\Big]\\
&= \sum_{\xi=1}^n p_\xi(\boldsymbol{\theta})\delta_{ij\xi} - \sum_{\xi=1}^n p_\xi(\boldsymbol{\theta})\delta_{i\xi}p_j(\boldsymbol{\theta}) - \sum_{\xi=1}^n p_\xi(\boldsymbol{\theta})\delta_{j\xi}p_i(\boldsymbol{\theta}) + \sum_{\xi=1}^n p_\xi(\boldsymbol{\theta})p_i(\boldsymbol{\theta})p_j(\boldsymbol{\theta})\\
&= \delta_{ij}p_i(\boldsymbol{\theta}) - p_i(\boldsymbol{\theta})p_j(\boldsymbol{\theta}) - p_j(\boldsymbol{\theta})p_i(\boldsymbol{\theta}) + p_i(\boldsymbol{\theta})p_j(\boldsymbol{\theta}) = \delta_{ij}p_i(\boldsymbol{\theta}) - p_i(\boldsymbol{\theta})p_j(\boldsymbol{\theta}).
\end{aligned}$$

□

Proof of Theorem 4

We only prove the pullback metric with respect to eq. (1), the sample-wise normalization. The proof of the metric with respect to eq. (2), the matrix-wise normalization, is similar.

Proof. Consider a differentiable mapping $\tilde{\boldsymbol{\theta}}(\boldsymbol{x}) = (\theta_1(\boldsymbol{x}), \ldots, \theta_{n-2}(\boldsymbol{x}), \varepsilon)$ from a smooth manifold $\mathcal{C}$ to $\mathcal{S}^{n-1}$, where $\mathcal{C}$ is parametrized by a coordinate system $\{\phi_i\}$, and $\varepsilon > 0$ is constant. The pullback metric $\tilde{g}$ is given by

$$\left\langle \frac{\partial}{\partial \phi_i}, \frac{\partial}{\partial \phi_j} \right\rangle_{\tilde{g}} = \left\langle \sum_{k=1}^{n-2} \frac{\partial \theta_k}{\partial \phi_i} \frac{\partial}{\partial \theta_k}, \sum_{l=1}^{n-2} \frac{\partial \theta_l}{\partial \phi_j} \frac{\partial}{\partial \theta_l} \right\rangle_{\mathfrak{G}^{n-1}(\tilde{\boldsymbol{\theta}})} = \sum_{k=1}^{n-2} \sum_{l=1}^{n-2} \frac{\partial \theta_k}{\partial \phi_i} \frac{\partial \theta_l}{\partial \phi_j} \mathfrak{G}_{kl}^{n-1}(\tilde{\boldsymbol{\theta}}), \tag{22}$$

where $\mathfrak{G}^{n-1}$ denotes FIM with respect to the canonical parameters $\theta_1, \ldots, \theta_{n-1}$ on $\mathcal{S}^{n-1}$. On the other hand, consider the mapping $\boldsymbol{\theta}(\boldsymbol{x}) = (\theta_1(\boldsymbol{x}), \ldots, \theta_{n-2}(\boldsymbol{x}))$ from $\mathcal{C}$ to $\mathcal{S}^{n-2}$. The corresponding pullback metric g is

$$\left\langle \frac{\partial}{\partial \phi_i}, \frac{\partial}{\partial \phi_j} \right\rangle_{g} = \left\langle \sum_{k=1}^{n-2} \frac{\partial \theta_k}{\partial \phi_i} \frac{\partial}{\partial \theta_k}, \sum_{l=1}^{n-2} \frac{\partial \theta_l}{\partial \phi_j} \frac{\partial}{\partial \theta_l} \right\rangle_{\mathfrak{G}^{n-2}(\boldsymbol{\theta})} = \sum_{k=1}^{n-2} \sum_{l=1}^{n-2} \frac{\partial \theta_k}{\partial \phi_i} \frac{\partial \theta_l}{\partial \phi_j} \mathfrak{G}_{kl}^{n-2}(\boldsymbol{\theta}), \tag{23}$$

where $\mathfrak{G}^{n-2}$ is FIM with respect to $\theta_1, \ldots, \theta_{n-2}$ on one of $\mathcal{S}^{n-1}$'s faces where $\theta_{n-1} \to -\infty$ and $p_{n-1} \to 0$. Note, from lemma 3, the information geometry on the statistical simplex $\mathcal{S}^{n-1}$ is continuous on its boundaries, in the sense that

$$\lim_{\theta_{n-1} \to -\infty} \mathfrak{G}^{n-1}(\tilde{\boldsymbol{\theta}}) = \begin{bmatrix} \mathfrak{G}^{n-2}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}. \tag{24}$$

Therefore,

$$\lim_{\varepsilon \to -\infty} \left\langle \frac{\partial}{\partial \phi_i}, \frac{\partial}{\partial \phi_j} \right\rangle_{\tilde{g}} = \left\langle \frac{\partial}{\partial \phi_i}, \frac{\partial}{\partial \phi_j} \right\rangle_{g}. \tag{25}$$

This allows the following strategy to derive the pullback metric with respect to eq. (1). We first map $\forall \boldsymbol{X} \in \mathcal{M}^r$ into $\mathcal{S}^{n-1}$ following the mapping

$$\tilde{p}_{j|i}(\boldsymbol{X}) = \frac{\exp(-s_{ij}(\boldsymbol{X}))}{\epsilon + \sum_{j \neq i} \exp(-s_{ij}(\boldsymbol{X}))} \ (\forall j \neq i), \quad \tilde{p}_{i|i} = \frac{\epsilon}{\epsilon + \sum_{j \neq i} \exp(-s_{ij}(\boldsymbol{X}))}, \tag{26}$$

where $\epsilon > 0$ is a small constant. Then, we pullback FIM along the above $\tilde{p}$. Finally, we take the limit $\epsilon \to 0$ on this pullback metric. According to eq. (25), the resulting metric is the same as pulling-back p with respect to eq. (1). Such a strategy makes the derivation simpler, because one only has to apply FIM inside $\mathcal{S}^{n-1}$.

Eq. (26) features a mapping from $\mathcal{M}^r$ to the product manifold $(\mathcal{S}^{n-1})^n$. For the i'th simplex in this product, we choose $\tilde{p}_{i|i}$ to be the "reference probability" (see section 2). Note, the pullback metric does not vary with such a specific choice of coordinate system. Then, the canonical parameters of the i'th simplex are given by

$$\tilde{\theta}_{j|i} = \log\left(\tilde{p}_{j|i}/\tilde{p}_{i|i}\right) = -s_{ij}(\boldsymbol{X}) - \log \epsilon \quad (\forall j \neq i). \tag{27}$$

By Lemma 3, the corresponding FIM is

$$\mathfrak{G}_{kl}^{(i)} = \tilde{p}_{k|i}\delta_{kl} - \tilde{p}_{k|i} \cdot \tilde{p}_{l|i},$$

where the superscript "(i)" means FIM is computed with respect to $\tilde{\theta}_{j|i}, \forall j \neq i$ on the i'th simplex. Then, $\forall \boldsymbol{u} = (u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n)$, $\boldsymbol{v} = (v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n) \in \Re^{n-1}$, we have

$$\boldsymbol{u}^T \mathfrak{G}^{(i)} \boldsymbol{v} = \sum_{j:j \neq i} \tilde{p}_{j|i} u_j v_j - \sum_{j:j \neq i} \tilde{p}_{j|i} u_j \sum_{j:j \neq i} \tilde{p}_{j|i} v_j. \tag{28}$$

Note, the Riemannian metric on the product manifold $(\mathcal{S}^{n-1})^n$ is a product metric [Jost(2008)] in a block diagonal form. Then, we have

$$\begin{aligned}
\tilde{g}_{ij}(\boldsymbol{X}) &= \begin{pmatrix} \frac{\partial \tilde{\boldsymbol{\theta}}_{\cdot|1}}{\partial \varphi_i} \\ \frac{\partial \tilde{\boldsymbol{\theta}}_{\cdot|2}}{\partial \varphi_i} \\ \vdots \\ \frac{\partial \tilde{\boldsymbol{\theta}}_{\cdot|n}}{\partial \varphi_i} \end{pmatrix}^T \begin{bmatrix} \mathfrak{G}^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathfrak{G}^{(2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathfrak{G}^{(n)} \end{bmatrix} \begin{pmatrix} \frac{\partial \tilde{\boldsymbol{\theta}}_{\cdot|1}}{\partial \varphi_j} \\ \frac{\partial \tilde{\boldsymbol{\theta}}_{\cdot|2}}{\partial \varphi_j} \\ \vdots \\ \frac{\partial \tilde{\boldsymbol{\theta}}_{\cdot|n}}{\partial \varphi_j} \end{pmatrix} \\
&= \sum_{k=1}^{n} \left(\frac{\partial \tilde{\boldsymbol{\theta}}_{\cdot|k}}{\partial \varphi_i} \right)^T \mathfrak{G}^{(k)} \left(\frac{\partial \tilde{\boldsymbol{\theta}}_{\cdot|k}}{\partial \varphi_j} \right) \\
&= \sum_{k=1}^{n} \left[\sum_{l:l\neq k} \tilde{p}_{l|k} \left(\frac{\partial \tilde{\theta}_{l|k}}{\partial \varphi_i} \cdot \frac{\partial \tilde{\theta}_{l|k}}{\partial \varphi_j} \right) - \left(\sum_{l:l\neq k} \tilde{p}_{l|k} \frac{\partial \tilde{\theta}_{l|k}}{\partial \varphi_i} \right) \left(\sum_{l:l\neq k} \tilde{p}_{l|k} \frac{\partial \tilde{\theta}_{l|k}}{\partial \varphi_j} \right) \right] \\
&= \sum_{k=1}^{n} \left[\sum_{l:l\neq k} \tilde{p}_{l|k} \left(\frac{\partial s_{kl}}{\partial \varphi_i} \cdot \frac{\partial s_{kl}}{\partial \varphi_j} \right) - \left(\sum_{l:l\neq k} \tilde{p}_{l|k} \frac{\partial s_{kl}}{\partial \varphi_i} \right) \left(\sum_{l:l\neq k} \tilde{p}_{l|k} \frac{\partial s_{kl}}{\partial \varphi_j} \right) \right].
\end{aligned}$$

Then, we let $\epsilon \to 0$ and get exactly the metric in theorem 4. □

Proof of Corollary 5

Proof. By theorem 4,

$$\begin{aligned}
\left\langle \frac{\partial}{\partial \boldsymbol{x}_i}, \frac{\partial}{\partial \boldsymbol{x}_i} \right\rangle_{\tilde{g}(\boldsymbol{X})} &= \sum_{k=1}^{n} \left[\sum_{l=1}^{n} p_{l|k} \left(\frac{\partial ||\boldsymbol{x}_k - \boldsymbol{x}_l||^2}{\partial \boldsymbol{x}_i} \right) \left(\frac{\partial ||\boldsymbol{x}_k - \boldsymbol{x}_l||^2}{\partial \boldsymbol{x}_i} \right)^T \right. \\
&\quad \left. - \left(\sum_{l=1}^{n} p_{l|k} \frac{\partial ||\boldsymbol{x}_k - \boldsymbol{x}_l||^2}{\partial \boldsymbol{x}_i} \right) \left(\sum_{l=1}^{n} p_{l|k} \frac{\partial ||\boldsymbol{x}_k - \boldsymbol{x}_l||^2}{\partial \boldsymbol{x}_i} \right)^T \right] \\
&= 4 \sum_{l=1}^{n} p_{l|i} \left(\boldsymbol{x}_i - \boldsymbol{x}_l\right) \left(\boldsymbol{x}_i - \boldsymbol{x}_l\right)^T + 4 \sum_{k=1}^{n} p_{i|k} \left(\boldsymbol{x}_i - \boldsymbol{x}_k\right) \left(\boldsymbol{x}_i - \boldsymbol{x}_k\right)^T \\
&\quad - 4 \left(\sum_{l=1}^{n} p_{l|i} \left(\boldsymbol{x}_i - \boldsymbol{x}_l\right) \right) \left(\sum_{l=1}^{n} p_{l|i} \left(\boldsymbol{x}_i - \boldsymbol{x}_l\right) \right)^T - 4 \sum_{k:k\neq i} \left(p_{i|k}(\boldsymbol{x}_i - \boldsymbol{x}_k) \cdot p_{i|k}(\boldsymbol{x}_i - \boldsymbol{x}_k)^T \right) \\
&= 4 \sum_{j=1}^{n} p_{j|i} \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right) \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)^T - 4 \left(\boldsymbol{x}_i - \sum_{j=1}^{n} p_{j|i} \boldsymbol{x}_j \right) \left(\boldsymbol{x}_i - \sum_{j=1}^{n} p_{j|i} \boldsymbol{x}_j \right)^T \\
&\quad + 4 \sum_{j=1}^{n} (p_{i|j} - p_{i|j}^2) \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right) \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)^T \\
&= 4 \left\{ \sum_{j=1}^{n} p_{j|i} \boldsymbol{x}_j \boldsymbol{x}_j^T - \left(\sum_{j=1}^{n} p_{j|i} \boldsymbol{x}_j \right) \left(\sum_{j=1}^{n} p_{j|i} \boldsymbol{x}_j \right)^T \right\} + 4 \sum_{j=1}^{n} p_{i|j} (1 - p_{i|j}) \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right) \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)^T.
\end{aligned}$$

□

Proof of Corollary 6

Proof. $\mathcal{O}^n_{\boldsymbol{X},\mathfrak{k}}(\boldsymbol{\tau})$ is equipped with a global coordinate system $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)$. By theorem 4, the Riemannian metric corresponding to eq. (3) is given by

$$
\begin{aligned}
g_{ij}(\boldsymbol{X}) &= \sum_{k=1}^{n} \left[\sum_{l \in \mathfrak{k}\mathrm{NN}_k} p_{l|k} \left(\frac{\partial(\tau_k s^{\boldsymbol{X}}_{kl})}{\partial \tau_i} \cdot \frac{\partial(\tau_k s^{\boldsymbol{X}}_{kl})}{\partial \tau_j} \right) - \sum_{l \in \mathfrak{k}\mathrm{NN}_k} p_{l|k} \frac{\partial(\tau_k s^{\boldsymbol{X}}_{kl})}{\partial \tau_i} \cdot \sum_{l \in \mathfrak{k}\mathrm{NN}_k} p_{l|k} \frac{\partial(\tau_k s^{\boldsymbol{X}}_{kl})}{\partial \tau_j} \right] \\
&= \sum_{k=1}^{n} \left[\delta_{ijk} \sum_{l \in \mathfrak{k}\mathrm{NN}_k} p_{l|k} \left(s^{\boldsymbol{X}}_{kl}\right)^2 - \delta_{ijk} \left(\sum_{l \in \mathfrak{k}\mathrm{NN}_k} p_{l|k} s^{\boldsymbol{X}}_{kl} \right)^2 \right] \\
&= \delta_{ij} \left[\sum_{l \in \mathfrak{k}\mathrm{NN}_i} p_{l|i} \left(s^{\boldsymbol{X}}_{il}\right)^2 - \left(\sum_{l \in \mathfrak{k}\mathrm{NN}_i} p_{l|i} s^{\boldsymbol{X}}_{il} \right)^2 \right]. \qquad (29)
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&-\frac{\partial}{\partial \tau_i} \left(\sum_{j \in \mathfrak{k}\mathrm{NN}_i} p_{j|i}(\tau_i) s^{\boldsymbol{X}}_{ij} \right) = -\frac{\partial}{\partial \tau_i} \sum_{j \in \mathfrak{k}\mathrm{NN}_i} \frac{\exp(-\tau_i s^{\boldsymbol{X}}_{ij})}{\sum_{\xi \in \mathfrak{k}\mathrm{NN}_i} \exp(-\tau_i s^{\boldsymbol{X}}_{i\xi})} s^{\boldsymbol{X}}_{ij} \\
&= -\sum_{j \in \mathfrak{k}\mathrm{NN}_i} \left[\frac{\exp(-\tau_i s^{\boldsymbol{X}}_{ij})}{\sum_{\xi \in \mathfrak{k}\mathrm{NN}_i} \exp(-\tau_i s^{\boldsymbol{X}}_{i\xi})} (-s^{\boldsymbol{X}}_{ij}) s^{\boldsymbol{X}}_{ij} - \frac{\exp(-\tau_i s^{\boldsymbol{X}}_{ij}) \cdot s^{\boldsymbol{X}}_{ij}}{\left(\sum_{\xi \in \mathfrak{k}\mathrm{NN}_i} \exp(-\tau_i s^{\boldsymbol{X}}_{i\xi})\right)^2} \sum_{\xi \in \mathfrak{k}\mathrm{NN}_i} \exp(-\tau_i s^{\boldsymbol{X}}_{i\xi})(-s^{\boldsymbol{X}}_{i\xi}) \right] \\
&= \sum_{j \in \mathfrak{k}\mathrm{NN}_i} p_{j|i}(\tau_i) \left(s^{\boldsymbol{X}}_{ij}\right)^2 - \sum_{j \in \mathfrak{k}\mathrm{NN}_i} p_{j|i}(\tau_i) s^{\boldsymbol{X}}_{ij} \sum_{\xi \in \mathfrak{k}\mathrm{NN}_i} p_{\xi|i}(\tau_i) s^{\boldsymbol{X}}_{i\xi}. \qquad (30)
\end{aligned}
$$

From eq. (29) and eq. (30),

$$
g_{ij}(\boldsymbol{X}) = -\delta_{ij} \frac{\partial}{\partial \tau_i} \left(\sum_{j \in \mathfrak{k}\mathrm{NN}_i} p_{j|i}(\tau_i) s^{\boldsymbol{X}}_{ij} \right).
$$

The proof for the metric corresponding to eq. (4) is similar. Note,

$$
\frac{\partial}{\partial \tau_i} s^t_{kl} = \begin{cases} \delta_{ki} s^{\boldsymbol{X}}_{kl} / \left(\tau_k s^{\boldsymbol{X}}_{kl} + 1\right) & \text{if } l \in \mathfrak{k}\mathrm{NN}_k; \\ 0 & \text{if otherwise.} \end{cases} \qquad (31)
$$

Denote $\mathfrak{d}_{kl} = s^{\boldsymbol{X}}_{kl} / (\tau_k s^{\boldsymbol{X}}_{kl} + 1)$. Then, by theorem 4,

$$
\begin{aligned}
g^t_{ij}(\boldsymbol{X}) &= \sum_{k=1}^{n} \left[\sum_{l \in \mathfrak{k}\mathrm{NN}_k} p_{l|k} \left(\frac{\partial s^t_{kl}}{\partial \tau_i} \cdot \frac{\partial s^t_{kl}}{\partial \tau_j} \right) - \sum_{l \in \mathfrak{k}\mathrm{NN}_k} p_{l|k} \frac{\partial s^t_{kl}}{\partial \tau_i} \cdot \sum_{l \in \mathfrak{k}\mathrm{NN}_k} p_{l|k} \frac{\partial s^t_{kl}}{\partial \tau_j} \right] \\
&= \sum_{k=1}^{n} \left[\delta_{ijk} \sum_{l \in \mathfrak{k}\mathrm{NN}_k} p_{l|k} \mathfrak{d}^2_{kl} - \delta_{ijk} \left(\sum_{l \in \mathfrak{k}\mathrm{NN}_k} p_{l|k} \mathfrak{d}_{kl} \right)^2 \right] \\
&= \delta_{ij} \left[\sum_{l \in \mathfrak{k}\mathrm{NN}_i} p_{l|i} \mathfrak{d}^2_{il} - \left(\sum_{l \in \mathfrak{k}\mathrm{NN}_i} p_{l|i} \mathfrak{d}_{il} \right)^2 \right]. \qquad (32)
\end{aligned}
$$

□

Proof of Proposition 8

Proof. By eq. (3) and corollary 6,

$$\int_0^\infty \sqrt{g_i(t)}dt = \int_0^\infty dt \sqrt{\sum_{j\in\mathfrak{k}\mathrm{NN}_i}\left[\frac{\exp(-ts_{ij}^{\boldsymbol{X}})}{\sum_{j\in\mathfrak{k}\mathrm{NN}_i}\exp(-ts_{ij}^{\boldsymbol{X}})}\left(s_{ij}^{\boldsymbol{X}}\right)^2\right] - \left[\sum_{j\in\mathfrak{k}\mathrm{NN}_i}\frac{\exp(-ts_{ij}^{\boldsymbol{X}})}{\sum_{j\in\mathfrak{k}\mathrm{NN}_i}\exp(-ts_{ij}^{\boldsymbol{X}})}s_{ij}^{\boldsymbol{X}}\right]^2}$$
$$= \int_0^\infty \sqrt{\frac{\sum_{j\in\mathfrak{k}\mathrm{NN}_i}\exp(-ts_{ij}^{\boldsymbol{X}})\cdot\left(d(ts_{ij}^{\boldsymbol{X}})\right)^2}{\sum_{j\in\mathfrak{k}\mathrm{NN}_i}\exp(-ts_{ij}^{\boldsymbol{X}})} - \left[\frac{\sum_{j\in\mathfrak{k}\mathrm{NN}_i}\exp(-ts_{ij}^{\boldsymbol{X}})\cdot d(ts_{ij}^{\boldsymbol{X}})}{\sum_{j\in\mathfrak{k}\mathrm{NN}_i}\exp(-ts_{ij}^{\boldsymbol{X}})}\right]^2}. \tag{33}$$

Note, in the above integration, t and $s_{ij}^{\boldsymbol{X}}$ always appear together in the term $(t\cdot s_{ij}^{\boldsymbol{X}})$. Assume $\bar{\boldsymbol{X}} = \lambda\boldsymbol{X}$. Then $s_{ij}^{\boldsymbol{X}}$ is multiplied by λ^2. We let $\bar{t} = \lambda^2 t$. Then the LAI integration of $\bar{\boldsymbol{X}}$ changes to $\int_0^\infty \sqrt{g_i(\bar{t})}d\bar{t}$, which has the same value as eq. (33). By definition 7,

$$|\mathcal{O}_{\mathfrak{k}}|(\lambda\boldsymbol{X}) = |\mathcal{O}_{\mathfrak{k}}|(\boldsymbol{X}). \tag{34}$$

□

Proof of Proposition 9

Proof. Note, $\int_0^\infty \sqrt{g_i(t)}dt$ is the length of a curve connecting the uniform distribution

$$p_{j|i} = \begin{cases} 1/\mathfrak{k} & \text{if } j\in\mathfrak{k}\mathrm{NN}_i; \\ 0 & \text{if otherwise,}\end{cases}$$

and the deterministic distribution

$$p_{j|i} = \begin{cases} 1 & \text{if } j \text{ is the nearest neighbour of } i; \\ 0 & \text{if otherwise.}\end{cases}$$

This length is lower bounded by the geodesic distance [Lebanon(2003)]. Therefore

$$\forall i,\quad \int_0^\infty \sqrt{g_i(t)}\,dt \geq \arccos(1/\sqrt{\mathfrak{k}}). \tag{35}$$

By eq. (5),

$$|\mathcal{O}_{\mathfrak{k}}|(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^n \int_0^\infty \sqrt{g_i(t)}dt \geq \arccos(1/\sqrt{\mathfrak{k}}). \tag{36}$$

In the following, we prove that LAI always has a finite value. Let $s_{i0}^{\boldsymbol{X}} = \min_{j:j\neq i} s_{ij}^{\boldsymbol{X}}$ be the square distance between $\boldsymbol{x}_i$ and its 1NN, $r_i^{\boldsymbol{X}} = \min_{j:s_{ij}^{\boldsymbol{X}}>s_{i0}^{\boldsymbol{X}}} |s_{ij}^{\boldsymbol{X}} - s_{i0}^{\boldsymbol{X}}|$, and $R_i^{\boldsymbol{X}} = \max_j |s_{ij}^{\boldsymbol{X}} - s_{i0}^{\boldsymbol{X}}|$. By corollary 6,

$$g_i(t) \leq \sum_{j\in\mathfrak{k}\mathrm{NN}_i} p_{j|i}(t)\left(s_{ij}^{\boldsymbol{X}} - s_{i0}^{\boldsymbol{X}}\right)^2 = \sum_{j\in\mathfrak{k}\mathrm{NN}_i}\frac{\exp(-ts_{ij}^{\boldsymbol{X}})}{\sum_{j\in\mathfrak{k}\mathrm{NN}_i}\exp(-ts_{ij}^{\boldsymbol{X}})}\left(s_{ij}^{\boldsymbol{X}} - s_{i0}^{\boldsymbol{X}}\right)^2$$
$$= \sum_{j\in\mathfrak{k}\mathrm{NN}_i}\frac{\exp(-t(s_{ij}^{\boldsymbol{X}} - s_{i0}^{\boldsymbol{X}}))}{\sum_{j\in\mathfrak{k}\mathrm{NN}_i}\exp(-t(s_{ij}^{\boldsymbol{X}} - s_{i0}^{\boldsymbol{X}}))}\left(s_{ij}^{\boldsymbol{X}} - s_{i0}^{\boldsymbol{X}}\right)^2$$
$$\leq \sum_{j\in\mathfrak{k}\mathrm{NN}_i}\exp(-t(s_{ij}^{\boldsymbol{X}} - s_{i0}^{\boldsymbol{X}}))\left(s_{ij}^{\boldsymbol{X}} - s_{i0}^{\boldsymbol{X}}\right)^2 \leq (\mathfrak{k}-2)\exp(-tr_i^{\boldsymbol{X}})\left(R_i^{\boldsymbol{X}}\right)^2.$$

Hence,

$$\int_0^\infty \sqrt{g_i(t)}dt \leq \sqrt{\mathfrak{k}-2}R_i^{\boldsymbol{X}} \int_0^\infty \exp\left(-tr_i^{\boldsymbol{X}}/2\right)dt = 2\sqrt{\mathfrak{k}-2}R_i^{\boldsymbol{X}}/r_i^{\boldsymbol{X}}.$$

By definition 7, $|\mathcal{O}_{\mathfrak{k}}|(\boldsymbol{X}) < \infty$. ☐

Proof of Corollary 10

Proof. $\mathcal{H}^{2n}_{\boldsymbol{Y},\boldsymbol{Z},\mathfrak{k}}$ has a global coordinate system $(a_1, b_1, \ldots, a_n, b_n)$. From eq. (6),

$$\forall i = 1, \ldots, n, \quad \frac{\partial s_{kl}(\boldsymbol{a}, \boldsymbol{b})}{\partial a_i} = \begin{cases} \delta_{ki} s_{kl}^{\boldsymbol{Y}} & \text{if } l \in \mathfrak{k}\mathtt{NN}_k; \\ 0 & \text{if otherwise,} \end{cases} \tag{37}$$

and

$$\forall i = 1, \ldots, n, \quad \frac{\partial s_{kl}(\boldsymbol{a}, \boldsymbol{b})}{\partial b_i} = \begin{cases} \delta_{ki} s_{kl}^{\boldsymbol{Z}} & \text{if } l \in \mathfrak{k}\mathtt{NN}_k; \\ 0 & \text{if otherwise.} \end{cases} \tag{38}$$

By theorem 4, the Riemannian metric g of $\mathcal{H}^{2n}_{\boldsymbol{Y},\boldsymbol{Z},\mathfrak{k}}$ with respect to eq. (6) is given by

$$\begin{aligned} \langle \partial a_i, \partial a_j \rangle_g &= \sum_{k=1}^n \Bigg[\sum_{l:l\in\mathfrak{k}\mathtt{NN}_k} p_{l|k} \left(\delta_{ki} s_{kl}^{\boldsymbol{Y}} \cdot \delta_{kj} s_{kl}^{\boldsymbol{Y}} \right) \\ &\qquad - \left(\sum_{l:l\in\mathfrak{k}\mathtt{NN}_k} p_{l|k} \delta_{ki} s_{kl}^{\boldsymbol{Y}} \right) \left(\sum_{l:l\in\mathfrak{k}\mathtt{NN}_k} p_{l|k} \delta_{kj} s_{kl}^{\boldsymbol{Y}} \right) \Bigg] \\ &= \delta_{ij} \sum_{l:l\in\mathfrak{k}\mathtt{NN}_i} p_{l|i} \left(s_{il}^{\boldsymbol{Y}} \right)^2 - \delta_{ij} \left(\sum_{l:l\in\mathfrak{k}\mathtt{NN}_i} p_{l|i} s_{il}^{\boldsymbol{Y}} \right)^2, \end{aligned} \tag{39}$$

$$\begin{aligned} \langle \partial b_i, \partial b_j \rangle_g &= \sum_{k=1}^n \Bigg[\sum_{l:l\in\mathfrak{k}\mathtt{NN}_k} p_{l|k} \left(\delta_{ki} s_{kl}^{\boldsymbol{Z}} \cdot \delta_{kj} s_{kl}^{\boldsymbol{Z}} \right) \\ &\qquad - \left(\sum_{l:l\in\mathfrak{k}\mathtt{NN}_k} p_{l|k} \delta_{ki} s_{kl}^{\boldsymbol{Z}} \right) \left(\sum_{l:l\in\mathfrak{k}\mathtt{NN}_k} p_{l|k} \delta_{kj} s_{kl}^{\boldsymbol{Z}} \right) \Bigg] \\ &= \delta_{ij} \sum_{l:l\in\mathfrak{k}\mathtt{NN}_i} p_{l|i} \left(s_{il}^{\boldsymbol{Z}} \right)^2 - \delta_{ij} \left(\sum_{l:l\in\mathfrak{k}\mathtt{NN}_i} p_{l|i} s_{il}^{\boldsymbol{Z}} \right)^2, \end{aligned} \tag{40}$$

$$\begin{aligned} \langle \partial a_i, \partial b_j \rangle_g &= \sum_{k=1}^n \Bigg[\sum_{l:l\in\mathfrak{k}\mathtt{NN}_k} p_{l|k} \left(\delta_{ki} s_{kl}^{\boldsymbol{Y}} \cdot \delta_{kj} s_{kl}^{\boldsymbol{Z}} \right) \\ &\qquad - \left(\sum_{l:l\in\mathfrak{k}\mathtt{NN}_k} p_{l|k} \delta_{ki} s_{kl}^{\boldsymbol{Y}} \right) \left(\sum_{l:l\in\mathfrak{k}\mathtt{NN}_k} p_{l|k} \delta_{kj} s_{kl}^{\boldsymbol{Z}} \right) \Bigg] \\ &= \delta_{ij} \sum_{l:l\in\mathfrak{k}\mathtt{NN}_i} p_{l|i} s_{il}^{\boldsymbol{Y}} s_{il}^{\boldsymbol{Z}} \\ &\quad - \delta_{ij} \left(\sum_{l:l\in\mathfrak{k}\mathtt{NN}_i} p_{l|i} s_{il}^{\boldsymbol{Y}} \right) \left(\sum_{l:l\in\mathfrak{k}\mathtt{NN}_i} p_{l|i} s_{il}^{\boldsymbol{Z}} \right). \end{aligned} \tag{41}$$

The proof of the Riemannian metric corresponding to eq. (7) is similar and is omitted. ☐

Proof of Proposition 11

Proof. The Jacobi matrix of the mapping $(a, b) \to (a_1, b_1, \ldots, a_n, b_n)$ is given by

$$J = \begin{bmatrix} \tau_1(\boldsymbol{Y}, \kappa) & 0 \\ 0 & \tau_1(\boldsymbol{Z}, \kappa) \\ \vdots & \vdots \\ \tau_n(\boldsymbol{Y}, \kappa) & 0 \\ 0 & \tau_n(\boldsymbol{Z}, \kappa) \end{bmatrix}.$$

Therefore, the Riemannian metric of the submanifold $\bar{\mathcal{H}}_{\boldsymbol{Y},\boldsymbol{Z},\mathfrak{k},\kappa}$ is

$$J^T g(\boldsymbol{c}) J = \sum_{i=1}^{n} \begin{bmatrix} \tau_i^2(\boldsymbol{Y}, \kappa) g_{aa}^i & \tau_i(\boldsymbol{Y}, \kappa)\tau_i(\boldsymbol{Z}, \kappa) g_{ab}^i \\ \tau_i(\boldsymbol{Y}, \kappa)\tau_i(\boldsymbol{Z}, \kappa) g_{ba}^i & \tau_i^2(\boldsymbol{Z}, \kappa) g_{bb}^i \end{bmatrix}.$$

The corresponding Riemannian volume element is given by

$$\mathtt{v}(a, b) = \sqrt{\left(\sum_{i=1}^{n} \tau_i^2(\boldsymbol{Y}, \kappa) g_{aa}^i\right)\left(\sum_{i=1}^{n} \tau_i^2(\boldsymbol{Z}, \kappa) g_{bb}^i\right) - \left(\sum_{i=1}^{n} \tau_i(\boldsymbol{Y}, \kappa)\tau_i(\boldsymbol{Z}, \kappa) g_{ab}^i\right)^2}. \tag{42}$$

The volume (area) of some Ω on $\bar{\mathcal{H}}_{\boldsymbol{Y},\boldsymbol{Z},\mathfrak{k},\kappa}$ is

$$|\bar{\mathcal{H}}_{\mathfrak{k},\kappa,\Omega}|(\boldsymbol{Y}, \boldsymbol{Z}) = \iint_{\Omega} \mathtt{v}(a, b)\, da\, db$$

$$= \iint_{\Omega} \sqrt{\left(\sum_{i=1}^{n} \tau_i^2(\boldsymbol{Y}, \kappa) g_{aa}^i\right)\left(\sum_{i=1}^{n} \tau_i^2(\boldsymbol{Z}, \kappa) g_{bb}^i\right) - \left(\sum_{i=1}^{n} \tau_i(\boldsymbol{Y}, \kappa)\tau_i(\boldsymbol{Z}, \kappa) g_{ab}^i\right)^2}\, da\, db.$$

□

Proof of Proposition 12

Proof. We only show that $\forall \boldsymbol{Y}, \forall \boldsymbol{Z}, \forall \lambda_y$, $|\bar{\mathcal{H}}_{\lambda_y \boldsymbol{Y},\boldsymbol{Z},\mathfrak{k},\kappa}| = |\bar{\mathcal{H}}_{\boldsymbol{Y},\boldsymbol{Z},\mathfrak{k},\kappa}|$. The proof of the invariance of $|\bar{\mathcal{H}}_{\boldsymbol{Y},\boldsymbol{Z},\mathfrak{k},\kappa}|$ with respect to the scalling of $\boldsymbol{Z}$ is similar. By the definition of $\tau_i(\boldsymbol{Y}, \kappa)$,

$$\tau_i(\lambda_y \boldsymbol{Y}, \kappa) s_{ij}^{\lambda_y \boldsymbol{Y}} = \tau_i(\boldsymbol{Y}, \kappa) s_{ij}^{\boldsymbol{Y}}.$$

Therefore, the term

$$\sum_{i=1}^{n} \tau_i^2(\boldsymbol{Y}, \kappa) g_{aa}^i = \sum_{i=1}^{n} \sum_{j \in \mathfrak{k}\mathtt{NN}_i} \frac{\exp\left(-a\tau_i(\boldsymbol{Y}, \kappa) s_{ij}^{\boldsymbol{Y}} - b_i s_{ij}^{\boldsymbol{Z}}\right)}{\sum_{j \in \mathfrak{k}\mathtt{NN}_i} \exp\left(-a\tau_i(\boldsymbol{Y}, \kappa) s_{ij}^{\boldsymbol{Y}} - b_i s_{ij}^{\boldsymbol{Z}}\right)} \left(\tau_i(\boldsymbol{Y}, \kappa) s_{ij}^{\boldsymbol{Y}}\right)^2$$
$$- \sum_{i=1}^{n} \left(\sum_{j \in \mathfrak{k}\mathtt{NN}_i} \frac{\exp\left(-a\tau_i(\boldsymbol{Y}, \kappa) s_{ij}^{\boldsymbol{Y}} - b_i s_{ij}^{\boldsymbol{Z}}\right)}{\sum_{j \in \mathfrak{k}\mathtt{NN}_i} \exp\left(-a\tau_i(\boldsymbol{Y}, \kappa) s_{ij}^{\boldsymbol{Y}} - b_i s_{ij}^{\boldsymbol{Z}}\right)} \tau_i(\boldsymbol{Y}, \kappa) s_{ij}^{\boldsymbol{Y}}\right)^2$$

is invariant to the scaling of $\boldsymbol{Y}$. Similarly, the values of $\sum_{i=1}^{n} \tau_i^2(\boldsymbol{Z}, \kappa) g_{bb}^i$ and $\sum_{i=1}^{n} \tau_i(\boldsymbol{Y}, \kappa)\tau_i(\boldsymbol{Z}, \kappa) g_{ab}^i$ do not change when $\boldsymbol{Y}$ is scaled up or down. According to eq. (42), the Riemannian volume element is invariant to the scalling of $\boldsymbol{Y}$. Hence, $\forall \boldsymbol{Y}, \forall \boldsymbol{Z}, \forall \lambda_y$, $|\bar{\mathcal{H}}_{\mathfrak{k},\kappa,\Omega}|(\lambda_y \boldsymbol{Y}, \boldsymbol{Z}) = |\bar{\mathcal{H}}_{\mathfrak{k},\kappa,\Omega}|(\boldsymbol{Y}, \boldsymbol{Z})$. □

Proof of Proposition 13

Proof. The "only if" part of the proof is trivial (zero volume implies zero density somewhere). We prove the "if" part as follows. By noting that each 2×2 diagonal block of $g(\boldsymbol{c})$ is a covariance matrix, we have the inequality

$$\sum_{i=1}^{n} \tau_i^2(\boldsymbol{Y}, \kappa) g_{aa}^i \sum_{i=1}^{n} \tau_i^2(\boldsymbol{Z}, \kappa) g_{bb}^i \geq \left(\sum_{i=1}^{n} \tau_i(\boldsymbol{Y}, \kappa) \tau_i(\boldsymbol{Z}, \kappa) \sqrt{g_{aa}^i g_{bb}^i} \right)^2 \geq \left(\sum_{i=1}^{n} \tau_i(\boldsymbol{Y}, \kappa) \tau_i(\boldsymbol{Z}, \kappa) g_{ab}^i \right)^2 . \tag{43}$$

Assume $\exists a, b$, such that $vol(a, b) = 0$. According to the definition of $vol(a, b)$ in eq. (9), the inequality in eq. (43) must be tight. To make the second "$\geq$" in eq. (43) tight, we have $\forall i$, the two vectors $(s_{i1}^Y, \ldots, s_{in}^Y)$ and $(s_{i1}^Z, \ldots, s_{in}^Z)$ are linearly dependant. Assume that $\forall j$, $s_{ij}^Y = \lambda_i s_{ij}^Z$. In this case, it is straightforward that

$$\forall a, \forall b, \forall i, \quad \tau_i^2(\boldsymbol{Y}, \kappa) g_{aa}^i = \tau_i^2(\boldsymbol{Z}, \kappa) g_{bb}^i \quad \text{and} \quad g_{aa}^i g_{bb}^i = \left(g_{ab}^i \right)^2 .$$

This makes the inequality in eq. (43) always tight for any a and b. Therefore the Riemmainian volume element in eq. (42) is everywhere zero, which leads to $|\bar{\mathcal{H}}_\Omega| = 0$. $\square$

References

[Jost(2008)] Jost, J. *Riemannian Geometry and Geometric Analysis*. Springer, 5th edition, 2008.

[Lebanon(2003)] Lebanon, G. Learning Riemannian metrics. In *UAI*, pp. 362–369, 2003.