# Regression Models on Riemannian Symmetric Spaces (Supplementary Report)

Emil Cornea[*], Hongtu Zhu[*] †, Peter Kim[**], and Joseph G. Ibrahim [*]

[*] *Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA*

[**] *Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada*

## 1. Appendix: Proofs

PROOF (PROOF OF THEOREM 4.1). Recall the definition of $\mathcal{Q}_n(\mathrm{q}, \boldsymbol{\beta})$ and define

$$\mathcal{Q}(\mathrm{q}, \boldsymbol{\beta}) := (E[h(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}, \mathrm{q}, \boldsymbol{\beta})])^T W(E[h(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}, \mathrm{q}, \boldsymbol{\beta})]). \quad (1)$$

It follows from (C1) and (C4) that $\sup_{(\mathrm{q},\boldsymbol{\beta})\in\Theta} |\mathcal{Q}_n(\mathrm{q}, \boldsymbol{\beta}) - \mathcal{Q}(\mathrm{q}, \boldsymbol{\beta})| \xrightarrow{p} 0$, while it follows from (C2), (C3), and (C5) that for any $\epsilon > 0$, we have

$$\mathcal{Q}(\mathrm{q}_*, \boldsymbol{\beta}_*) = 0 < \inf_{(\mathrm{q},\boldsymbol{\beta}):\mathrm{dist}_{\mathcal{M}}(\mathrm{q},\mathrm{q}_*)+\|\boldsymbol{\beta}-\boldsymbol{\beta}_*\|\geq\epsilon} \mathcal{Q}(\mathrm{q}, \boldsymbol{\beta}).$$

Thus, consistency of $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$ in (16) follows from Theorem 5.7 in (van der Vaart, 1998).

The proof of Theorem 4.1 (b) consists of two parts as follows.

- Part 1 is to show that Theorem 4.1 (b) is valid when $\mathcal{M}$ is an open subset of the Euclidean space $R^d$ and $\phi = \mathrm{id}$. Since this result is a classical result (Newey, 1993), we omit it for simplicity.

- Part 2 is to show that Theorem 4.1 (b) is valid when $\mathcal{M}$ is a RSS. We focus on part 2 below.

†*Address for correspondence and reprints:* Hongtu Zhu, Ph.D., Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA, hzhu@bios.unc.edu.

When $\mathcal{M}$ is a RSS, we consider a chart $(U, \phi)$ on $\mathcal{M}$ near $\mathrm{q}_*$ with $\mathbf{t}_* = \phi(\mathrm{q}_*)$. As $(\hat{\mathrm{q}}, \hat{\boldsymbol{\beta}})$ is a consistent estimator for $(\mathrm{q}_*, \boldsymbol{\beta}_*)$, $\hat{\mathrm{q}} \in U$ holds with probability approaching one, as $n \to \infty$. When $\hat{\mathrm{q}} \in U$, we define $\hat{\mathbf{t}} = \phi(\hat{\mathrm{q}})$. If follows from the continuous mapping theorem that $(\phi(\hat{\mathrm{q}})^T, \hat{\boldsymbol{\beta}}^T)^T$ is a consistent estimator for $(\phi(\mathrm{q}_*)^T, \boldsymbol{\beta}_*^T)^T$ in $R^{(d_{\mathcal{M}}+d_{\beta})}$. Conditions (C6)-(C10) hold for $(\phi(\mathrm{q}_*)^T, \boldsymbol{\beta}_*^T)^T$ and functions $(\mathbf{t}^T, \boldsymbol{\beta}^T)^T \mapsto h(\mathbf{x}, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})$. Let $\mathcal{H}_n(\mathrm{q}, \boldsymbol{\beta}) = 0.5\partial_{(\mathrm{q},\boldsymbol{\beta})}\mathcal{Q}_n(\mathrm{q}, \boldsymbol{\beta})$. To establish Theorem 4.1 (b), we can apply the proof of part 1 in (Newey, 1993) to the following function

$$(\mathbf{t}^T, \boldsymbol{\beta}^T)^T \mapsto \mathcal{H}_n(\phi^{-1}(\mathbf{t}), \boldsymbol{\beta}).$$

Compared with Euclidean case, although $\mathcal{H}_n(\phi^{-1}(\mathbf{t}), \boldsymbol{\beta})$ is a function of random variables $(y_i, \mathbf{x}_i)$ with $y_i$ being $\mathcal{M}$-valued, $\mathcal{E}(y_i, \mathbf{x}_i; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})$ and $\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y_i, \mathbf{x}_i; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})$ are real vector-valued variables for all $(\mathbf{t}, \boldsymbol{\beta})$ and $i$. Thus, all arguments in part 1 still hold for part 2.

By using the chain rule and $\phi' = (\phi' \circ \phi^{-1}) \circ \phi$ near $\mathrm{q}_*$, we can establish the last statement of Theorem 4.1 (b).

Proof (Proof of Theorem 4.2 ). Theorem 4.2 (i) directly follows from Theorem 4.1 with $\Sigma_\phi^* = (G_\phi^{*\top} W_\phi^* G_\phi^*)^{-1}$, where $G_\phi^*$ is given by

$$E\left[h_\phi^*(\mathbf{x})\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}_*)\big|_{\mathbf{t}=\phi(\mathrm{q}_*)}\,\middle|\,\mathbf{x}\right] = E[D_\phi(\mathbf{x})\Omega(\mathbf{x})^{-1}D_\phi(\mathbf{x})^T].$$

and $W_\phi^* = \left(\mathrm{Var}[h_\phi^*(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)]\right)^{-1} = G_\phi^{*-1}$. Thus, we have

$$\Sigma_\phi^* = (G_\phi^* G_\phi^{*-1} G_\phi^*)^{-1} = G_\phi^{*-1} = \left(E[D_\phi(\mathbf{x})\Omega(\mathbf{x})^{-1}D_\phi(\mathbf{x})^T]^T\right)^{-1}.$$

To show Theorem 4.2 (ii), it is sufficient to show that $(\Sigma_\phi^*)^{-1} - (\Sigma_{\phi,h}^{opt})^{-1}$ is non-negative for any $s \times d_{\mathcal{M}}$ matrix-valued function $h(\mathbf{x}; \mathrm{q}, \boldsymbol{\beta})$. With some simple calculations, we can show that $(\Sigma_\phi^*)^{-1} - (\Sigma_{\phi,h}^{opt})^{-1}$ is equal to

$$E[(\Omega^{-1/2}\{D_\phi(\mathbf{x}) - \Omega(\mathbf{x})h(\mathbf{x})^T E[h(\mathbf{x})\Omega(\mathbf{x})h(\mathbf{x})^T]^{-1}E[h(\mathbf{x})D_\phi(\mathbf{x})^T]\})^{\otimes 2}],$$

which is non-negative definite.

Theorem 4.2 (iii) is derived in the main paper right after the theorem's statement.

Proof (Proof of Theorem 4.3). The proof consists of two parts as follows.

- Part 1 is to show $(\phi(\tilde{\mathrm{q}}_E), \tilde{\boldsymbol{\beta}}_E) \to^p (\phi(\mathrm{q}_*), \boldsymbol{\beta}_*)$ as $n \to \infty$;
- Part 2 is to show that $\sqrt{n}[(\phi(\tilde{\mathrm{q}}_E)^T, \tilde{\boldsymbol{\beta}}_E^T)^T - (\phi(\mathrm{q}_*)^T, \boldsymbol{\beta}_*^T)^T]$ is asymptotically normally distributed.

To prove Part 1, we proceed as follows. Recall the definitions of $\hat{h}_{E,\phi}(\mathbf{x}_i)$, $\hat{W}_{E,\phi}$, and $(\phi(\tilde{\mathrm{q}}_E), \tilde{\boldsymbol{\beta}}_E)$. We need to prove two sufficient results as follows:

(i) $\sup_{(\mathrm{q},\boldsymbol{\beta})} ||\mathbb{P}_n\{[\hat{h}_{E,\phi}(\mathbf{x}_i) - h^*_{E,\phi}(\mathbf{x})]\mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\}|| \to^p 0$;

(ii) $||\hat{W}_{E,\phi} - W^*_{E,\phi}|| \to^p 0$.

Based on the results (i) and (ii), we can show that in probability, $\mathcal{Q}_n(\mathrm{q}, \boldsymbol{\beta})$ based on $\hat{h}_{E,\phi}(\mathbf{x}_i)$ and $\hat{W}_{E,\phi}$ converges uniformly to

$$\mathcal{Q}_1(\mathrm{q}, \boldsymbol{\beta}) := (E[h(\mathbf{x}, \mathrm{q}_*, \boldsymbol{\beta}_*)\mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})])^T W^*_{E,\phi}(E[h(\mathbf{x}, \mathrm{q}_*, \boldsymbol{\beta}_*)\mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})]).$$

Then, we can apply the same arguments of Theorem 4.1 to finish the proof of Part 1.

We prove Part 1 (i) as follows. It follows from the triangle inequality, the Cauchy-Schwarz inequality, the trace inequality, and (C16) that

$$\begin{aligned} &||\mathbb{P}_n\{[\hat{h}_{E,\phi}(\mathbf{x}) - h^*_{E,\phi}(\mathbf{x})]\mathcal{E}(y, \mathbf{x}; \mathrm{q}, \boldsymbol{\beta})\}||^2 \\ &\quad \le [\mathbb{P}_n\{||\hat{h}_{E,\phi}(\mathbf{x}) - h^*_{E,\phi}(\mathbf{x})||^2\}][\mathbb{P}_n\{f_0(y, \mathbf{x})^2\}] \qquad (2) \\ &\quad \le O_p(1)\Big[\mathbb{P}_n||\hat{D}_\phi(\mathbf{x}) - D_\phi(\mathbf{x})||^2||\hat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}||^2 \\ &\quad +\mathbb{P}_n||D_\phi(\mathbf{x})||^2||\hat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1} - V_{E*}^{-1}||^2\Big]. \end{aligned}$$

It follows from (C3) and (C6) that $\hat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}$ converges to $V_{E*}^{-1}$ in probability. Combining this result with (C19) leads to Part 1 (i).

We prove Part 1 (ii) as follows. We first prove $||(\hat{W}_{E,\phi})^{-1} - (W^*_{E,\phi})^{-1}|| \to^p 0$. Now $(\hat{W}_{E,\phi})^{-1} - (W^*_{E,\phi})^{-1}$ can be decomposed as the sum of three terms given by

$$\begin{aligned} \text{(ii.1)} &= \mathbb{P}_n\{[h^*_{E,\phi}(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)]^{\otimes 2}\} - (W^*_{E,\phi})^{-1}, \\ \text{(ii.2)} &= \mathbb{P}_n\{[h^*_{E,\phi}(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)]^{\otimes 2} - [h^*_{E,\phi}(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)]^{\otimes 2}\}, \\ \text{(ii.3)} &= \mathbb{P}_n\{[\hat{h}_{E,\phi}(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)]^{\otimes 2} - [h^*_{E,\phi}(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)]^{\otimes 2}\}. \end{aligned}$$

It follows from (C15) that (ii.1) converges to zero in probability. It follows from (C17) and a Taylor's series expansion that

$$\text{(ii.2)} \le o_p(1)E[f_0(y, \mathbf{x})^2 f_1(\mathbf{x})^2] = o_p(1).$$

To prove (ii.3), we define

$$\hat{B}(y, \mathbf{x}) = \hat{h}_{E,\phi}(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I) \text{ and } \tilde{B}(y, \mathbf{x}) = h^*_{E,\phi}(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I).$$

Then, similar to the arguments in (2), we have

$$\begin{aligned}
|(\text{ii.3})| &= \|\mathbb{P}_n[\hat{B}(y, \mathbf{x})\hat{B}(y, \mathbf{x})^T - \tilde{B}(y, \mathbf{x})\tilde{B}(y, \mathbf{x})^T]\| \\
&\leq \mathbb{P}_n\{\|\hat{B}(y, \mathbf{x}) - \tilde{B}(y, \mathbf{x})\|^2\} + 2\mathbb{P}_n\{\|\hat{B}(y, \mathbf{x}) - \tilde{B}(y, \mathbf{x})\|\|\tilde{B}(y, \mathbf{x})\|\} \\
&\leq O_p(1)[\mathbb{P}_n\{||\hat{h}_{E,\phi}(\mathbf{x}) - h^*_{E,\phi}(\mathbf{x})||^2\}]^{1/2}[\mathbb{P}_n\{f_0(y, \mathbf{x})^4 f_1(\mathbf{x})^2\}]^{1/2}.
\end{aligned}$$

Therefore, it follows from (C3), (C6), (C17), (C19), and (2) that (ii.3) converges to zero in probability. Therefore, this completes the proof of Part 1.

To prove Part 2, we proceed as follows. We define $\mathcal{H}_n(\mathbf{t}, \boldsymbol{\beta})$ to be

$$(\mathbb{P}_n[\hat{h}_{E,\phi}(\mathbf{x})\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})])^\top \hat{W}_{E,\phi}(\mathbb{P}_n[\hat{h}_{E,\phi}(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta})]).$$

It follows from a Taylor's series expansion and some simple calculations that as $n \to \infty$, we have

$$\sqrt{n}[(\phi(\tilde{\mathrm{q}}_E)^\top, \tilde{\boldsymbol{\beta}}_E^\top)^\top - (\mathbf{t}_*^\top, \boldsymbol{\beta}_*^\top)^\top] = [-\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{H}_n(\overline{\mathbf{t}}, \overline{\boldsymbol{\beta}})]^{-1}\sqrt{n}\mathcal{H}_n(\mathbf{t}_*, \boldsymbol{\beta}_*),$$

where $(\phi^{-1}(\overline{\mathbf{t}}), \overline{\boldsymbol{\beta}}) \in B((\mathbf{t}_*, \boldsymbol{\beta}_*), \delta)$ for any $\delta > 0$ in probability. We need to prove two results as follows:

(i) $\sqrt{n}\mathcal{H}_n(\mathbf{t}_*, \boldsymbol{\beta}_*)$ is asymptotically normal;

(ii) $-\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{H}_n(\overline{\mathbf{t}}, \overline{\boldsymbol{\beta}})$ converges to a positive definite matrix in probability.

Combining Part 2 (i) and (ii) finishes the proof of Theorem 4.3.

To prove Part 2 (i), we need to show three results as follows:

(i.1) $\sqrt{n}\mathbb{P}_n\{h^*_{E,\phi}(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\} \to^L N(0, (W^*_{E,\phi})^{-1})$;

(i.2) $\mathbb{P}_n\{[\hat{h}_{E,\phi}(\mathbf{x}) - h^*_{E,\phi}(\mathbf{x})]\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\} \to^p 0$;

(i.3) $\sqrt{n}\mathbb{P}_n\{[\hat{h}_{E,\phi}(\mathbf{x}) - h^*_{E,\phi}(\mathbf{x})]\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\} \to^p 0$.

Part 2 (i.1) follows from the standard central limit theorem. Part 2 (i.2) follows from the same arguments used in (2). We prove Part 2 (i.3) as follows. Note that the left term in (i.3) can be written as the sum of two terms, given by

$$\sqrt{n}\mathbb{P}_n\{[\hat{D}_\phi(\mathbf{x}) - D_\phi(\mathbf{x})]\hat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\}, \tag{3}$$

$$\sqrt{n}\mathbb{P}_n\{D_\phi(\mathbf{x})[\hat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1} - V_{E*}^{-1}]\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\}. \tag{4}$$

Let $\mathbf{e}_j$ be the $(d_{\mathcal{M}} + d_\beta) \times 1$ vector with a 1 at the $j$-th component and a 0 otherwise for $j = 1, \ldots, d_{\mathcal{M}} + d_\beta$. It follows from (C18) that

$$\begin{aligned}
& \left|\sqrt{n}\mathbb{P}_n\{\mathbf{e}_j^\top[\hat{D}_\phi(\mathbf{x}) - D_\phi(\mathbf{x})]\hat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\}\right| \\
= & \left|\mathrm{tr}\left(\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}\sqrt{n}\mathbb{P}_n\{\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\mathbf{e}_j^\top[\hat{D}_\phi(\mathbf{x}) - D_\phi(\mathbf{x})]\}\right)\right| \\
\leq & \|\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1}\| \, \|\sqrt{n}\mathbb{P}_n\{\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\mathbf{e}_j^\top[\hat{D}_\phi(\mathbf{x}) - D_\phi(\mathbf{x})]\}\| \\
\leq & O_p(1)o_p(1) = o_p(1).
\end{aligned}$$

It follows from (C15) and (C16) and standard central limit theory that

$$\sqrt{n}\mathbb{P}_n\{D_\phi(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\} = O_p(1).$$

Therefore, after some simple calculations, we have

$$\begin{aligned}
& \left|\sqrt{n}\mathbb{P}_n\{\mathbf{e}_j^\top D_\phi(\mathbf{x})[\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1} - V(\mathrm{q}_*, \boldsymbol{\beta}_*)^{-1}]\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\}\right| \leq \\
& \|\widehat{V}(\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)^{-1} - V(\mathrm{q}_*, \boldsymbol{\beta}_*)^{-1}\| \, \|\sqrt{n}\mathbb{P}_n\{\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\mathbf{e}_j^\top D_\phi(\mathbf{x})\}\| = o_p(1).
\end{aligned}$$

Based on these results, we obtain

$$\begin{aligned}
\sqrt{n}\mathcal{H}_n(\mathbf{t}_*, \boldsymbol{\beta}_*) \quad &= \quad G_{\phi, h_{E,\phi}^*}W_{E,\phi}^*\sqrt{n}\mathbb{P}_n\{h_{E,\phi}^*(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\} \\
&\to^L \quad N(0, G_{\phi, h_{E,\phi}^*}W_{E,\phi}^*G_{\phi, h_{E,\phi}^*}).
\end{aligned}$$

To prove Part 2 (ii), we need to show two results as follows:

(ii.1) as $n \to \infty$, $\mathbb{P}_n\{h_{E,\phi}^*(\mathbf{x})\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}_*), \boldsymbol{\beta}_*)\} \to^p 0$;

(ii.2) as $\delta_n \to 0$,

$$\sup_{(\phi^{-1}(\mathbf{t}), \boldsymbol{\beta}) \in B((\mathrm{q}_*, \boldsymbol{\beta}_*), \delta_n)} \mathbb{P}_n\{\hat{h}_{E,\phi}(\mathbf{x})\partial_{(\mathbf{t}, \boldsymbol{\beta})}^l\mathcal{E}(y, \mathbf{x}; \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}) - h_{E,\phi}^*(\mathbf{x})\partial_{(\mathbf{t}, \boldsymbol{\beta})}^l\mathcal{E}(y, \mathbf{x}; \mathrm{q}_*, \boldsymbol{\beta}_*)\}$$

converges to zero in probability for $l = 0, 1$, and 2.

Part 2 (ii.1) follows from the law of large numbers. We prove Part 2 (ii.2) as follows. It can be shown that $\hat{h}_{E,\phi}(\mathbf{x})\partial^l_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y,\mathbf{x};\phi^{-1}(\mathbf{t}),\boldsymbol{\beta}) - h^*_{E,\phi}(\mathbf{x})\partial^l_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y,\mathbf{x};\mathrm{q}_*,\boldsymbol{\beta}_*)$ can be written as the sum of three terms, given by

$$\begin{aligned}
T_1(y,\mathbf{x};\mathbf{t},\boldsymbol{\beta}) &= [\widehat{D}_\phi(\mathbf{x}) - D_\phi(\mathbf{x})]\widehat{V}(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{-1}\partial^l_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y,\mathbf{x};\phi^{-1}(\mathbf{t}),\boldsymbol{\beta}),\\
T_2(y,\mathbf{x};\mathbf{t},\boldsymbol{\beta}) &= D_\phi(\mathbf{x})[\widehat{V}(\hat{\mathrm{q}}_I,\hat{\boldsymbol{\beta}}_I)^{-1} - V^{-1}_{E,*}]\partial^l_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y,\mathbf{x};\phi^{-1}(\mathbf{t}),\boldsymbol{\beta}),\\
T_3(y,\mathbf{x};\mathbf{t},\boldsymbol{\beta}) &= h^*_{E,\phi}(\mathbf{x})[\partial^l_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y,\mathbf{x};\phi^{-1}(\mathbf{t}),\boldsymbol{\beta}) - \partial^l_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{E}(y,\mathbf{x};\mathrm{q}_*,\boldsymbol{\beta}_*)].
\end{aligned}$$

By using the same reasoning as in (3), (4), and (C16), we have

$$\sup_{(\phi^{-1}(\mathbf{t}),\boldsymbol{\beta})\in B((\mathrm{q}_*,\boldsymbol{\beta}_*),\delta_n)} \sum_{j=1}^{2} ||\mathbb{P}_n\{T_j(y,\mathbf{x};\mathbf{t},\boldsymbol{\beta})\}|| \to^p 0.$$

It follows from (C17) and the law of large numbers that

$$\begin{aligned}
&\sup_{(\phi^{-1}(\mathbf{t}),\boldsymbol{\beta})\in B((\mathrm{q}_*,\boldsymbol{\beta}_*),\delta_n)} |\mathbb{P}_n\{T_3(y,\mathbf{x};\mathbf{t},\boldsymbol{\beta})\}| \le o_p(1) =\\
&\sup_{(\phi^{-1}(\mathbf{t}),\boldsymbol{\beta})\in B((\mathrm{q}_*,\boldsymbol{\beta}_*),\delta_n)} |E\{T_3(y,\mathbf{x};\mathbf{t},\boldsymbol{\beta})\}|\\
&+ \sup_{(\phi^{-1}(\mathbf{t}),\boldsymbol{\beta})\in B((\mathrm{q}_*,\boldsymbol{\beta}_*),\delta_n)} |\mathbb{P}_n\{T_3(y,\mathbf{x};\mathbf{t},\boldsymbol{\beta})\} - E\{T_3(y,\mathbf{x};\mathbf{t},\boldsymbol{\beta})\}|.
\end{aligned}$$

Based on these results, we obtain

$$\sup_{(\bar{\mathbf{t}},\overline{\boldsymbol{\beta}})\in B((\mathbf{t}_*,\boldsymbol{\beta}_*),\delta_n)} || -\partial_{(\mathbf{t},\boldsymbol{\beta})}\mathcal{H}_n(\bar{\mathbf{t}},\overline{\boldsymbol{\beta}}) - G_{\phi,h^*_{E,\phi}}W^*_{E,\phi}G_{\phi,h^*_{E,\phi}}|| \to^p 0$$

as $\delta_n \to 0$ and $n \to \infty$.

Proof (Proof of Theorem 4.4). The proof follows similar steps as in Theorem 4.3 with straightforward modifications, which for brevity are omitted.

Proof (Proof of Theorem 4.5). Let $(U,\phi)$ be a chart on $\mathcal{M}$ near $\mathrm{q}_*$.

(i) We only prove the result for $W^{(2)}_{n,\phi}$ as follows. Under $H^{(2)}_0$, the true value $\mathrm{q}_*$ equals $\mathrm{q}_0$ and $(U,\phi)$ is a chart near $\mathrm{q}_0$. As $\tilde{\mathrm{q}}_E$ is a consistent estimator for q, it follows that $\tilde{\mathrm{q}}_E \in U$, for $n$ large enough, with probability approaching 1. From the CLT for $\tilde{\mathrm{q}}_E$, we have that, under $H^{(2)}_0$,

$$\sqrt{n}\,(\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)) \xrightarrow{d} N_{d_\mathcal{M}}(\mathbf{0}, (I_{d_\mathcal{M}}\ \mathbf{0})\Sigma_{E,\phi}(I_{d_\mathcal{M}}\ \mathbf{0})^T)$$

As $n\widehat{\Sigma}_{E,\phi} \xrightarrow{p} \Sigma_{E,\phi}$, by the continuous mapping theorem, we get

$$\frac{1}{\sqrt{n}}\left[(I_{d_{\mathcal{M}}}\ \mathbf{0})\widehat{\Sigma}_{E,\phi}(I_{d_{\mathcal{M}}}\ \mathbf{0})^T\right]^{-1/2} \xrightarrow{p} \left[(I_{d_{\mathcal{M}}}\ \mathbf{0})\Sigma_{E,\phi}(I_{d_{\mathcal{M}}}\ \mathbf{0})^T\right]^{-1/2}.$$

Then, using Slutzky's theorem, we have

$$\left[(I_{d_{\mathcal{M}}}\ \mathbf{0})\widehat{\Sigma}_{E,\phi}(I_{d_{\mathcal{M}}}\ \mathbf{0})^T\right]^{-1/2}(\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, I_{d_{\mathcal{M}}})$$

which implies $W^{(2)}_{n,\phi} \xrightarrow{d} \chi^2_{d_{\mathcal{M}}}$.

(ii) Since $\tilde{\boldsymbol{\beta}}_E$ and the lower-right $d_{\boldsymbol{\beta}} \times d_{\boldsymbol{\beta}}$ submatrix of $\widehat{\Sigma}_{E,\phi}$ are independent of the chart $(U, \phi)$, so is $W^{(1)}_{n,\phi}$.

(iii) Let $(U', \phi')$ be another chart near $\mathrm{q}_0$ with $\hat{\mathrm{q}}_E \in U'$. A Taylor's series expansion of the transition function $\phi' \circ \phi$ about $\phi(\mathrm{q}_0)$ shows that $\phi'(\mathrm{q}_E) - \phi'(\mathrm{q}_0) = (J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1))(\phi(\mathrm{q}_E) - \phi(\mathrm{q}_0))$. Let $\hat{\mathrm{q}}$ be the consistent estimator of q that the asymptotic covariance estimator $\widehat{\Sigma}_{E,\phi}$ is based on. As $\widehat{\Sigma}_{E,\phi}$ is compatible with the manifold structure of $\mathcal{M}$ and $J(\phi' \circ \phi)_{\phi(\hat{\mathrm{q}})} = J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1)$, we have

$$\begin{aligned}
W^{(2)}_{n,\phi'} &= [\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)]^T[J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1)]^T \\
&\quad \times \left[\left(J(\phi' \circ \phi)_{\phi(\hat{\mathrm{q}})}\ \mathbf{0}\right)\widehat{\Sigma}_{E,\phi}\left(J(\phi' \circ \phi)_{\phi(\hat{\mathrm{q}})}\ \mathbf{0}\right)^T\right]^{-1} \\
&\quad \times [J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1)][\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)] \\
&= [\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)]^T[J(\phi' \circ \phi)^{-1}_{\phi(\mathrm{q}_0)} + o_p(1)]^{-\top} \\
&\quad \times \left[\left(J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1)\ \ \mathbf{0}\right)\widehat{\Sigma}_{E,\phi}\left(J(\phi' \circ \phi)_{\phi(\mathrm{q}_0)} + o_p(1)\ \ \mathbf{0}\right)^T\right]^{-1} \\
&\quad \times [J(\phi' \circ \phi)^{-1}_{\phi(\mathrm{q}_0)} + o_p(1)]^{-1}[\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)] \\
&= [\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)]^T\left[(I_{d_{\mathcal{M}}}\ \mathbf{0})\,\widehat{\Sigma}_{E,\phi}\,(I_{d_{\mathcal{M}}}\ \mathbf{0})^T + o_p(1)\right]^{-1}[\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)] \\
&= W^{(2)}_{n,\phi} + o_p(1).
\end{aligned}$$

Thus, $W^{(2)}_{n,\phi'}$ and $W^{(2)}_{n,\phi}$ are asymptotically equivalent.

(iv) Let $\phi$ and $\phi'$ be two normal charts on $\mathcal{M}$ centered at $\tilde{\mathrm{q}}_E$. Thus, $\phi(\cdot) = A \circ \mathrm{Log}_{\tilde{\mathrm{q}}_E}(\cdot)$ and $\phi'(\cdot) = A' \circ \mathrm{Log}_{\tilde{\mathrm{q}}_E}(\cdot)$, where $A, A' : T_{\tilde{\mathrm{q}}_E}\mathcal{M} \to R^{d_{\mathcal{M}}}$ are two isomorphisms of linear spaces induced by the coordinates with respect to two orthonormal bases of $T_{\tilde{\mathrm{q}}_E}\mathcal{M}$. Therefore, $\phi'(\cdot) = O\phi(\cdot)$, where $O = A'A^{-1}$ corresponds to an orthonormal matrix, and $\widehat{\Sigma}_{E,\phi';11} = O\widehat{\Sigma}_{E,\phi;11}O^T$. Thus, $\widehat{\Sigma}_{E;11} := A^{-1} \circ \widehat{\Sigma}_{E,\phi;11} \circ A$ is independent of the chart $\phi$ and is a 1-1 linear map from $T_{\tilde{\mathrm{q}}_E}\mathcal{M}$ onto itself. Since $A$ preserves the inner product, we have

$$\begin{aligned} W_{n,\phi}^{(2)} &= \operatorname{tr}\{[\widehat{\Sigma}_{E,\phi;11}^{-1} A(\mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0))]^\top A(\mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0))\} \\ &= \operatorname{tr}\{[A((\widehat{\Sigma}_{E;11})^{-1}\mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0))]^\top A(\mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0))\} \\ &= \mathrm{m}_{\tilde{\mathrm{q}}_E}((\widehat{\Sigma}_{E;11})^{-1}\mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0), \mathrm{Log}_{\tilde{\mathrm{q}}_E}(\mathrm{q}_0)) = W_{M,n}^{(2)}. \end{aligned}$$

Proof (Proof of Theorem 4.6). The proof follows from a straightforward application of a Taylor's series expansion and Slutzky's theorem. We only prove (ii). We have that, under $H_{1,n}^{(2)}$,

$$\sqrt{n}\,(\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_n)) \xrightarrow{d} N_{d_\mathcal{M}}(\mathbf{0}, (I_{d_\mathcal{M}}\ \mathbf{0})\Sigma_{E,\phi}(I_{d_\mathcal{M}}\ \mathbf{0})^T),$$

where $\mathrm{q}_n = \mathrm{Exp}_{\mathrm{q}_0}(\mathbf{v}/\sqrt{n} + o(1/\sqrt{n}))$. As $n\widehat{\Sigma}_{E,\phi} \xrightarrow{p} \Sigma_{E,\phi}$, by the continuous mapping theorem, we get $n\widehat{\Sigma}_{E,\phi;11} \xrightarrow{p} \Sigma_{E,\phi;11}$ Then, using Slutzky's theorem, we have

$$\left[\widehat{\Sigma}_{E,\phi;11}\right]^{-1/2}(\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_n)) \xrightarrow{d} N_{d_\mathcal{M}}(\mathbf{0}, I_{d_\mathcal{M}})$$

From Taylor's series expansion of the map $\phi \circ \mathrm{Exp}_{\mathrm{q}_0}$ at $\mathbf{0}$, we have

$$\sqrt{n}(\phi(\mathrm{q}_n) - \phi(\mathrm{q}_0)) = J(\phi \circ \mathrm{Exp}_{\mathrm{q}_0})_{\mathbf{0}}(\mathbf{v}) + o(1).$$

Thus, again using Slutzky's theorem, we obtain that

$$\left[\widehat{\Sigma}_{E,\phi;11}\right]^{-1/2}(\phi(\tilde{\mathrm{q}}_E) - \phi(\mathrm{q}_0)) \xrightarrow{d} N_{d_\mathcal{M}}([\Sigma_{E,\phi;11}]^{-1/2} J(\phi \circ \mathrm{Exp}_{\mathrm{q}_0})_{\mathbf{0}}(\mathbf{v}), I_{d_\mathcal{M}}),$$

which implies that, under $H_{1,n}^{(2)}$, $W_{n,\phi}^{(2)}$ converges in distribution to a noncentral $\chi^2_{d_\mathcal{M}}$ with noncentrality parameter

$$J(\phi \circ \mathrm{Exp}_{\mathrm{q}_0})_0(\mathbf{v})^T \left[\hat{\Sigma}_{E,\phi;11}\right]^{-1} J(\phi \circ \mathrm{Exp}_{\mathrm{q}_0})_0(\mathbf{v}).$$

(iii) It follows from (ii) applied to a normal chart $\phi = \mathrm{Log}_{\tilde{\mathrm{q}}_E}$ near $\tilde{\mathrm{q}}_E$.

Proof (Proof of Theorem 4.7). Consider a Taylor's series expansion of the real-valued function $\mathrm{dist}_\mathcal{M}(\mathrm{q}, \mathrm{q}_0)^2$ around the point $\mathrm{q}_*$

$$\begin{aligned} \mathrm{dist}_\mathcal{M}(\mathrm{q}, \mathrm{q}_0)^2 = \mathrm{dist}_\mathcal{M}(\mathrm{q}_*, \mathrm{q}_0)^2 &+ \mathrm{grad}_{\mathrm{q}_*}(\mathrm{dist}_\mathcal{M}(\cdot, \mathrm{q}_0)^2)(\mathrm{Log}_{\mathrm{q}_*}(\mathrm{q})) \\ &+ \frac{1}{2}\mathrm{Hess}_{\mathrm{q}_*}(\mathrm{dist}_\mathcal{M}(\cdot, \mathrm{q}_0)^2)(\mathrm{Log}_{\mathrm{q}_*}(\mathrm{q}), \mathrm{Log}_{\mathrm{q}_*}(\mathrm{q})) \\ &+ O(\|\mathrm{Log}_{\mathrm{q}_*}(\mathrm{q})\|^3), \end{aligned}$$

for any q in a normal chart centered at $q_*$ with $\text{dist}_{\mathcal{M}}(q, q_*) < \rho^*_{\mathcal{M}}$. The result depending on which method is used, implies

$$\sqrt{n}\,\text{Log}_{q_*}(\tilde{q}_E) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, \Sigma_{E,\text{Log}_{q_*;11}}), \tag{5}$$

where $\Sigma_{E,\text{Log}_{q_*}}$ is the matrix representation of the asymptotic covariance matrix of $\tilde{q}_E$ with respect to the orthonormal basis in $T_{q_*}\mathcal{M}$ associated with the normal chart under consideration. The squared distance function becomes $\text{dist}_{\mathcal{M}}(\text{Exp}_{q_*}(\mathbf{v}), q_*)^2 = \|\mathbf{v}\|^2_{T_{q_*}M}$, for any $\mathbf{v} \in T_{q_*}\mathcal{M}$ with $\|\mathbf{v}\|_{T_{q_*}M} < \rho^*_{\mathcal{M}}$, and the matrix representation of its Hessian at the chart center $q_*$ is the identity matrix $I_{d_{\mathcal{M}}}$, with respect to any orthonormal basis of $T_{q_*}\mathcal{M}$.

(a) Under the null hypothesis $H_0^{(2)}$, $\tilde{q}_E$ belongs to a normal chart centered at $q_0$ with probability approaching one, and

$$W_{dist} = m_{q_0}(\text{Log}_{q_0}(\hat{q}_E), \text{Log}_{q_0}(\hat{q}_E)) = (\text{Log}_{q_0}(\hat{q}_E))^T(\text{Log}_{q_0}(\hat{q}_E)),$$

when $\text{Log}_{q_0}(\hat{q}_E)$ is expressed in the orthonormal basis of $T_{q_0}\mathcal{M}$ associated with the normal chart centered at $q_* = q_0$. From this and (5), it follows that $nW_{dist} \xrightarrow{d} \chi^2(\lambda_1, \ldots, \lambda_{d_{\mathcal{M}}})$, where $\lambda_1, \ldots, \lambda_{d_{\mathcal{M}}}$ are the eigenvalues of the matrix $\Sigma_{E,\text{Log}_{q_0},11}$. Let $\Sigma_{E,\text{Log}_{q_0}}$ and $\Sigma'_{E,\text{Log}_{q_0}}$ be the matrix representations of the asymptotic covariance matrix of $\tilde{q}_E$ in two normal charts centered at $q_0$. Then $\Sigma'_{E,\text{Log}_{q_0},11} = O\Sigma_{E,\text{Log}_{q_0},11}O^T$, for some $d_{\mathcal{M}} \times d_{\mathcal{M}}$ orthogonal matrix $O$, so the eigenvalues $\lambda_1, \ldots, \lambda_{d_{\mathcal{M}}}$ are independent of the normal chart.

(b) Under the alternative hypothesis $H_1^{(2)}$, from the Taylor's series expansion above, we have

$$\begin{aligned} W_{dist} - \text{dist}_{\mathcal{M}}(q_*, q_0)^2 &= \text{grad}_{q_*}(\text{dist}_{\mathcal{M}}(\cdot, q_0)^2)[\text{Log}_{q_*}(\tilde{q}_E) + O_p(\|\text{Log}_{q_*}\tilde{q}_E\|^2)] \\ &= [D^T_{dist} + o_p(1)]\text{Log}_{q_*}(\tilde{q}_E). \end{aligned}$$

Using Slutzky's theorem, we get

$$\sqrt{n}(W_{dist} - \text{dist}_{\mathcal{M}}(q_*, q_0)^2) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, D^T_{dist}\Sigma_{E,\text{Log}_{q_*},11}\, D_{dist}).$$

In the case when $q_0$ is close to $q_*$ so that $q_0$ is in a normal chart centered at $q_*$, then $D_{dist} = \text{grad}_{q_*}(\text{dist}_{\mathcal{M}}(\cdot, q_0)^2) = -2\text{Log}_{q_*}q_0$ and we have

$$\sqrt{n}(W_{dist} - \text{dist}_{\mathcal{M}}(q_*, q_0)^2) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, 4[\text{Log}_{q_*}q_0]^T\Sigma_{E,\text{Log}_{q_*},11}\,[\text{Log}_{q_*}q_0],$$

which completes the proof.

PROOF (PROOF OF THEOREM 4.8). We introduce some notation. For any chart $(U,\phi)$ on $\mathcal{M}$ with $\mathrm{q}_0 \in U$, we define $F^*_{\phi i}$ and $\mathrm{U}^*$ in a similar way as $F_{\phi i}$ and $\mathrm{U}^*$, respectively, by replacing $(\mathrm{q}_0, \tilde{\boldsymbol{\beta}}_I)$ with $(\mathrm{q}_*, \boldsymbol{\beta}_*)$. That is,

$$F^*_{\phi i} = (F^{*\top}_{\phi i,1}, F^{*\top}_{\phi i,2})^\top = \partial_{(\mathbf{t},\boldsymbol{\beta})}\mathrm{dist}_{\mathcal{M}}(f(\mathbf{x}_i, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}), y_i)^2\big|_{\mathbf{t}=\phi(\mathrm{q}_*),\boldsymbol{\beta}_*},$$

$$\mathrm{U}^* = \begin{pmatrix} \mathrm{U}^*_{\mathbf{tt}} & \mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}} \\ \mathrm{U}^*_{\boldsymbol{\beta}\mathbf{t}} & \mathrm{U}^*_{\boldsymbol{\beta}\boldsymbol{\beta}} \end{pmatrix} = \sum_{i=1}^n \partial^2_{(\mathbf{t},\boldsymbol{\beta})}\mathrm{dist}_{\mathcal{M}}(f(\mathbf{x}_i, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}), y_i)^2\Big|_{\mathbf{t}=\phi(\mathrm{q}_*),\boldsymbol{\beta}_*},$$

where the subcomponents $F^{*\top}_{\phi i,1}$ and $F^{*\top}_{\phi i,2}$ correspond to $\mathbf{t}$ and $\boldsymbol{\beta}$, respectively.

(i) The key idea in deriving the asymptotic distribution of $W_{SC,\phi}$ consists of two steps. In Step 1, using a Taylor's series expansion of $\sum_{i=1}^n \tilde{\mathrm{U}}_{i,2}$ at $(\phi(\mathrm{q}_*), \boldsymbol{\beta}_*)$, we can show that, under the *null* hypothesis $H_0^{(2)}$,

$$\tilde{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_* = (-\mathrm{U}^*_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1}\sum_{i=1}^n F^*_{\phi i,2} + O_p(n^{-1}).$$

In Step 2, under $H_0^{(2)}$, we expand $\sum_{i=1}^n F_{\phi i,1}$ at $(\phi(\mathrm{q}_*), \boldsymbol{\beta}_*)$ to get

$$\begin{aligned}\sum_{i=1}^n F_{\phi i,1} &= \sum_{i=1}^n F^*_{\phi i,1} + \mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_*)[1+o_p(1)] \\ &= \sum_{i=1}^n F^*_{\phi i,1} - \mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}}\mathrm{U}^{*-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\sum_{i=1}^n F^*_{\phi i,2})[1+o_p(1)] \\ &= (I_{d_{\mathcal{M}}}, -\mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}}(\mathrm{U}^*_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1})\left(\sum_{i=1}^n F^*_{\phi i}\right)[1+o_p(1)].\end{aligned}$$

Thus, by using Slutzky's theorem, we have

$$(I_{d_{\mathcal{M}}}, -\mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}}(\mathrm{U}^*_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1})\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n F^*_{\phi i}\right) \xrightarrow{d} N_{d_{\mathcal{M}}}(\mathbf{0}, \Sigma_{\phi,\mathrm{q}_*}),$$

where $\Sigma_{\phi,\mathrm{q}_*}$ is given by

$$E\Big\{[(I_{d_{\mathcal{M}}}, -\mathrm{U}^*_{\mathbf{t}\boldsymbol{\beta}}(\mathrm{U}^*_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1})\partial_{\mathbf{t},\boldsymbol{\beta}}\mathrm{dist}_{\mathcal{M}}(f(\mathbf{x}, \phi^{-1}(\mathbf{t}), \boldsymbol{\beta}_*), y)^2\big|_{\mathbf{t}=\phi(\mathrm{q}_*)}]^{\otimes 2}\Big\}.$$

Since $\tilde{\Sigma}_{\phi,\mathrm{q}} = [n^{-1}\sum_{i=1}^n[(I_{d_{\mathcal{M}}}, -\mathrm{U}_{\mathbf{t}\boldsymbol{\beta}}\mathrm{U}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}})(F_{\phi i} - \overline{F}_\phi)]^{\otimes 2}] \xrightarrow{p} \Sigma_\phi$, it follows from the continuous mapping theorem and Slutzky's theorem that under $H_0^{(2)}$, the score test statistic $W_{SC,\phi} = (\sum_{i=1}^n F_{\phi i,1})^\top \tilde{\Sigma}^{-1}_{\phi,\mathrm{q}}(\sum_{i=1}^n F_{\phi i,1}) \xrightarrow{d} \chi^2_{d_{\mathcal{M}}}$.

(ii) Let $(U', \phi')$ be another chart on $\mathcal{M}$ with $\mathrm{q}_0 \in U'$. Under $H_0^{(2)}$, by the chain rule, we have $F_{\phi',i} = \mathrm{diag}(J(\phi \circ \phi'^{-1})_{\phi'(\mathrm{q}_0)}, I_{d_{\boldsymbol{\beta}}})^T F_{\phi,i}$ and $\mathrm{U}_{\phi',\mathbf{t}'\boldsymbol{\beta}} = J(\phi \circ \phi'^{-1})^T_{\phi'(\mathrm{q}_0)} \mathrm{U}_{\phi,\mathbf{t}\boldsymbol{\beta}}$. It immediately follows that

$$\begin{aligned} F_{\phi' i,1} &= J(\phi \circ \phi'^{-1})^T_{\phi'(\mathrm{q}_0)} F_{\phi i,1}, \\ \tilde{\Sigma}_{\phi',\mathrm{q}} &= J(\phi \circ \phi'^{-1})^T_{\phi'(\mathrm{q}_0)} \tilde{\Sigma}_{\phi,\mathrm{q}} J(\phi \circ \phi'^{-1})_{\phi'(\mathrm{q}_0)}, \end{aligned}$$

which implies $W_{SC,\phi'} = W_{SC,\phi}$. Thus, the score test statistic $W_{SC,\phi}$ is independent of the chart $(U, \phi)$ near $\mathrm{q}_0$.

## 2. Intrinsic Regression Model - Multicenter link functions

In the paper, we mainly discussed single-center link functions, as defined (1) of Section 3.1. We may also consider a multicenter link function to account for the presence of multiple discrete covariates, such as gender and diagnostic group. Let $\mathbf{x}_i = (\mathbf{x}_{i,C}, \mathbf{x}_{i,D})$, where $\mathbf{x}_{i,D}$ and $\mathbf{x}_{i,C}$ are, respectively, a $d_{\mathbf{x},D} \times 1$ vector of all the discrete covariates and a $d_{\mathbf{x},C} \times 1$ vector of all the continuous covariates and their potential interactions with $\mathbf{x}_{i,D}$. We may introduce a center for each covariate class based on $\mathbf{x}_{i,D}$ (McCullagh and A.Nelder, 1989). In this case, we may define the multicenter link function as follows:

$$\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}(\mathbf{x}_D), \boldsymbol{\beta}) : R^{d_{\mathbf{x}}} \times \mathcal{M}^{d_D} \times R^{d_{\boldsymbol{\beta}}} \to M, \tag{6}$$

where $d_D$ is an integer associated with the number of covariate classes and $\boldsymbol{\beta}$ is primarily associated with continuous covariates. Moreover, it is assumed that $\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta})$ satisfies a multicenter property as follows:

$$\boldsymbol{\mu}((\mathbf{0}, \mathbf{x}_D), \mathrm{q}(\mathbf{x}_D), \boldsymbol{\beta}) = \boldsymbol{\mu}(\mathbf{x}, \mathrm{q}(\mathbf{x}_D), \mathbf{0}) = \mathrm{q}(\mathbf{x}_D). \tag{7}$$

When the regression coefficients vector $\boldsymbol{\beta}$ equals $\mathbf{0}$, the link function is independent of continuous covariates and reduces to $\mathrm{q}(\mathbf{x}_D)$ in $\mathcal{M}$. When all continuous covariates are equal to zero, the link function is independent of the regression coefficients and reduces to the center $\mathrm{q}(\mathbf{x}_D)$ in $\mathcal{M}$. For instance, we may extend the geodesic link function (3), Section 3.1 of the paper, to the scenarios with multiple discrete covariates by assuming

$$\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}(\mathbf{x}_D), \boldsymbol{\beta}) = \mathrm{Exp}^{\mathcal{M}}_{\mathrm{q}(\mathbf{x}_D)}(\sum_{k=1}^{d_{\mathbf{x}}} x_{ik} V_{\mathrm{q}(\mathbf{x}_D),k}), \tag{8}$$

where $V_{\mathrm{q}(\mathbf{x}_D),k}$'s are tangent vectors in $T_{\mathrm{q}(\mathbf{x}_D)}\mathcal{M}$ for all possible $\mathbf{x}_D$.

More generally, we will consider a general link function defined as

$$\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta}) : R^{d_{\mathrm{x}}} \times \Theta \to M, \tag{9}$$

where $\boldsymbol{\theta}$ is a vector of unknown parameters in a parameter space $\Theta$. For the multicenter link function (6), $\boldsymbol{\theta}$ contains all unknown parameters in $\mathrm{q}(\mathbf{x}_D)$'s and $\boldsymbol{\beta}$ and $\Theta$ equals $\mathcal{M}^{d_D} \times R^{d_\beta}$. However, for notational simplicity, throughout the paper we focus on single-center functions (1), as defined in Section 3.1 of the paper, since the extension to (9) is trivial.

## 3. Differential Geometry - Technical Details

### *3.1. Riemannian Metric, Distance, and Geodesics*

A Riemannian manifold $(\mathcal{M}, \mathrm{m})$ is a smooth manifold $\mathcal{M}$ together with a metric m. The $\mathrm{m} = (\mathrm{m}_\mathrm{p})_{\mathrm{p}\in\mathcal{M}}$ is a family of inner products $m_\mathrm{p}$ on the tangent space $T_\mathrm{p}\mathcal{M}$ of $\mathcal{M}$ at $\mathrm{p} \in \mathcal{M}$, and for any smooth vector fields $X = (X_\mathrm{p})_{\mathrm{p}\in\mathcal{M}}$ and $Y = (Y_\mathrm{p})_{\mathrm{p}\in\mathcal{M}}$ on an open set $U \subset \mathcal{M}$, the real valued map $\mathrm{p} \mapsto m_\mathrm{p}(X_\mathrm{p}, Y_\mathrm{p})$ is smooth on $U$. Let $d_\mathcal{M}$ be the dimension of $\mathcal{M}$. The tangent space $T_\mathrm{p}\mathcal{M}$ is isomorphic to $R^{d_\mathcal{M}}$. For a local chart $(U, \phi)$, $U$ is an open subset of $\mathcal{M}$ and there is a homeomorphism $\phi : U \to \phi(U) \subset R^{d_\mathcal{M}}$, where $\phi(U)$ is an open set containing $\phi(\mathrm{p}) = \mathbf{t} = (t^1, \ldots, t^{d_\mathcal{M}})^T$. Let $\partial_j$ denote the tangent vector with respect to the coordinate curves $\partial/\partial t^j$ for $j = 1, \ldots, d_\mathcal{M}$. The vector fields $\frac{\partial}{\partial \mathbf{t}} = (\partial_1, \ldots, \partial_{d_\mathcal{M}})^T$ induce a basis at each of the tangent spaces $T_{\phi^{-1}(\mathbf{t})}\mathcal{M}$ for $\mathbf{t} \in \phi(U)$. In this basis, the metric can be expressed by a symmetric positive definite matrix $\mathrm{M}_\phi(\mathbf{t}) = [\mathrm{m}_{jk}(\mathbf{t})]$, where $\mathrm{m}_{j,k}(\mathbf{t}) = m_{\phi^{-1}(\mathbf{t})}(\partial_j, \partial_k)$. The matrix $\mathrm{M}_\phi(\mathbf{t})$ is called the local representation of the Riemannian metric in the chart $(U, \phi)$, and for any $\mathrm{p} \in U$, the inner product of $\mathbf{v}$ and $\mathbf{w} \in T_\mathrm{p}\mathcal{M}$ is given by $m_p(\mathbf{v}, \mathbf{w}) = \tilde{\mathbf{v}}^T \mathrm{M}_\phi(\phi(\mathrm{p}))\tilde{\mathbf{w}}$, where $\tilde{\mathbf{v}} = (v^1, \ldots, v^{d_\mathcal{M}})^\top$ and $\tilde{\mathbf{w}} = (w^1, \ldots, w^{d_\mathcal{M}})^\top$ are the representations of $\mathbf{v}$ and $\mathbf{w}$, respectively, in the chart $(U, \phi)$, i.e., $\mathbf{v} = \sum_{j=1}^{d_\mathcal{M}} v^j \partial_j$.

The length $\ell(\gamma)$ of a $C^1-$curve $\gamma : [t_0, t_1] \to \mathcal{M}$ on a Riemannian manifold $\mathcal{M}$ is defined by $\ell(\gamma) = \int_{t_0}^{t_1} \sqrt{\mathrm{m}_{\gamma(t)}(\gamma'(t), \gamma'(t))}\, dt$. The length of a continuous, piecewise smooth curve on $\mathcal{M}$ is defined as the sum of the lengths of its smooth pieces. The *geodesic distance* $\mathrm{dist}_\mathcal{M}(\mathrm{p}, \mathrm{q})$ between p and $\mathrm{q} \in \mathcal{M}$ is defined as the infimum of $L(\gamma)$ taken over all continuous, piecewise smooth curves $\gamma : [a, b] \to \mathcal{M}$ with $\gamma(a) = \mathrm{p}$ and

$\gamma(b) = \mathrm{q}$. The Riemannian manifold $(\mathcal{M}, \mathrm{dist}_{\mathcal{M}})$ is a metric space and geodesics are then, by definition, the locally distance-minimizing paths. The geodesics are the curves satisfying the second order differential system in the chart $(U, \phi)$ given by

$$\ddot{\gamma}^j + \sum_{l',l} \Gamma^j_{l'l} \dot{\gamma}^{l'} \dot{\gamma}^l = 0,$$

where $\Gamma^j_{l'l} = 0.5 \sum_{j'} \mathrm{m}^{j'j}(\partial_{l'} \mathrm{m}_{li} + \partial_{j'} \mathrm{m}_{ll'} - \partial_l \mathrm{m}_{j'l'})$ are the Christoffel symbols of the first kind.

For $\mathrm{p} \in \mathcal{M}$ and $\mathbf{v}$ in $T_{\mathrm{p}}\mathcal{M}$, there exists a unique geodesic $\gamma = \gamma(\cdot; \mathrm{p}, \mathbf{v}) : I \to \mathcal{M}$ satisfying $\gamma(0) = \mathrm{p}$ and $\gamma'(0) = \mathbf{v}$, where $I$ is a maximal open interval in $R$ containing 0. Moreover, $\gamma$ depends smoothly on both p and $\mathbf{v}$. In general, $I$ may not be all of $R$. The manifold is said to be *geodesically complete* if the maximal interval $I$ is the entire real line $R$ for all geodesics. For example, the Euclidean space $R^n$ and the unit sphere $S^n$ are geodesically complete manifolds, while $R \setminus \{0\}$ is not. The Hopf-Rinow-De Rham theorem states that a geodesically complete Riemannian manifold is complete as a metric space with the distance induced by the Riemannian metric, and that there always exists at least one distance minimizing geodesic between any two points of the manifold.

## 3.2. Exponential and Logarithmic Maps

For a general Riemannian manifold, given a vector $\mathbf{v}$ in $T_{\mathrm{p}}\mathcal{M}$ and a real number $\tau \in R$, we have that $\gamma(t; \mathrm{p}, \tau\mathbf{v}) = \gamma(t\tau; \mathrm{p}, \mathbf{v})$, for all $t \in R$ with $t\tau$ in the definition domain of $\gamma(\cdot; \mathrm{p}, \mathbf{v})$. Therefore, for a tangent vector $\mathbf{v} \in T_{\mathrm{p}}\mathcal{M}$ with $\|\mathbf{v}\|_{\mathrm{p}} := (m_{\mathrm{p}}(\mathbf{v}, \mathbf{v}))^{1/2} < r(\mathrm{p})$ for some small $r(\mathrm{p}) > 0$, the geodesic starting from p and with initial velocity $\mathbf{v}$ is defined on an interval containing $[0, 1]$. The manifold *exponential map* at a point $\mathrm{p} \in \mathcal{M}$, $\mathrm{Exp}^{\mathcal{M}}_{\mathrm{p}} : B_{\mathrm{p}}(\mathbf{0}, r(\mathrm{p})) \to \mathcal{M}$ is defined by $\mathrm{Exp}^{\mathcal{M}}_{\mathrm{p}}(V) = \gamma(1; \mathrm{p}, V)$ for $V \in B_{\mathrm{p}}(0, r(\mathrm{p}))$, where, $B_{\mathrm{p}}(\mathbf{0}, r(\mathrm{p}))$ denotes the ball of radius $r(\mathrm{p})$ centered at the origin in $T_{\mathrm{p}}\mathcal{M}$. The exponential map $\mathrm{Exp}^{\mathcal{M}}_{\mathrm{p}}$ is a locally smooth diffeomorphism around $\mathbf{0} \in T_{\mathrm{p}}\mathcal{M}$, i.e. there is a $r_*(\mathrm{p}) \in (0, r(\mathrm{p}))$ such that $\mathrm{Exp}^{\mathcal{M}}_{\mathrm{p}}$ is a diffeomorphism from $B_{\mathrm{p}}(\mathbf{0}, r_*(\mathrm{p}))$ into $\mathcal{M}$. The inverse map is denoted by $\mathrm{Log}^{\mathcal{M}}_{\mathrm{p}}$ and it provides normal coordinates on $\mathcal{M}$ around p.

For $\mathrm{q} \in \mathrm{Exp}^{\mathcal{M}}_{\mathrm{p}}(B_{\mathrm{p}}(\mathbf{0}, r_*(\mathrm{p})))$, the geodesic distance from p to q can be expressed as $\mathrm{dist}_{\mathcal{M}}(\mathrm{p}, \mathrm{q}) = \|\mathrm{Log}_{\mathrm{p}}\mathrm{q}\|_{\mathrm{p}}$, and thus $\mathrm{Exp}^{\mathcal{M}}_{\mathrm{p}}(B_{\mathrm{p}}(\mathbf{0}, r_*(\mathrm{p})))$ is the ball $B^{\mathcal{M}}(\mathrm{p}, r^*(\mathrm{p}))$ in $\mathcal{M}$, with the induced distance, centered at p of radius $r^*(\mathrm{p})$. As the tangent space $T_\mathrm{p}\mathcal{M}$ is isomorphic to $R^{d_\mathcal{M}}$, the logarithmic map $\mathrm{Log}_\mathrm{p}$ provides a local chart near p. If the tangent space of $\mathcal{M}$ at p is endowed with an orthonormal basis, then such a chart is called a *normal chart* and the coordinates are called *normal coordinates.*

### 3.3. Cut Locus and Radius of Injectivity

From now on, we will assume that the manifold $\mathcal{M}$ is geodesically complete, and thus the exponential map $\mathrm{Exp}_p$ is defined on the entire tangent space $T_\mathrm{p}\mathcal{M}$. A geodesic $\gamma(t;\mathrm{p},\mathbf{v})$ is either always minimizing the distance from p to $\gamma(t;\mathrm{p},\mathbf{v})$ from $t=0$ to $\infty$, or it is minimizing up to a finite point $t_0$ and no more thereafter. In the latter case, the point $\gamma(t_0;\mathrm{p},\mathbf{v})$ is called a *cut point* for the geodesic $\gamma(\cdot;\mathrm{p},\mathbf{v})$ and the tangent vector $t_0\mathbf{v}$ is called a *tangential cut point.* The set of cut points of all geodesics starting from p is called the *cut locus of* p and denoted by $C(\mathrm{p})\subset\mathcal{M}$. The set of corresponding tangent vectors is called the *tangential cut locus of* p and denoted by $\mathcal{C}(\mathrm{p})\subset T_\mathrm{p}\mathcal{M}$. We have $C(\mathrm{p})=\mathrm{Exp}_\mathrm{p}(\mathcal{C}(\mathrm{p}))$ and thus, the maximal definition domain of the normal chart centered at p is the domain $\mathcal{D}(\mathrm{p})\subset T_\mathrm{p}\mathcal{M}$ containing $\mathbf{0}$ and bounded by $\mathcal{C}(\mathrm{p})$. The domain $\mathcal{D}(\mathrm{p})$ is connected and star-shaped with respect to the origin and its image via $\mathrm{Exp}_\mathrm{p}$ is the entire manifold except the cut locus of p (Pennec, 2006). Hence the normal chart centered at p is given by

$$\mathrm{Log}_\mathrm{p}\,:\,D(\mathrm{p})=M\setminus C(\mathrm{p})\to\mathcal{D}(\mathrm{p})\subset R^{d_\mathcal{M}}.$$

Here $T_\mathrm{p}\mathcal{M}$ is endowed with an orthonormal basis and identified with $R^{d_\mathcal{M}}$. The size of this chart is quantified by the *radius of injectivity of* $\mathcal{M}$ *at* p, $\rho^*(\mathcal{M},\mathrm{p})=\mathrm{dist}_{T_\mathrm{p}}(\mathbf{0},\mathcal{C}(\mathrm{p}))$, which is the maximal radius of origin centered balls in $T_\mathrm{p}\mathcal{M}$ on which the exponential map is one-to-one. The radius of injectivity $\rho^*_\mathcal{M}$ of the manifold $\mathcal{M}$ is the infimum of the radii of injectivity at all points over the manifold. For example, in the case of Euclidean space $R^d$, the maximal definition domain of the normal chart is $D(\mathbf{t})=R^d$, for all $\mathbf{t}\in R^d$, and therefore the radius of injectivity is $\rho^*_{R^d}=\infty$. In the case of the unit sphere $S^k$, the Riemannian metric induced by the canonical inner product on $R^{k+1}$, the cut locus of a point $\mathrm{p}\in S^k$ is $C(\mathrm{p})=\{-\mathrm{p}\}$, and the tangential cut locus is $\mathcal{C}(\mathrm{p})=S^{k-1}(\pi)\subset T_\mathrm{p}S^k$. Therefore, we have $\mathcal{D}(\mathrm{p})=B(\mathbf{0},\pi)\subset T_\mathrm{p}S^k$, $D(\mathrm{p})=S^k\setminus\{-\mathrm{p}\}$, and $\rho^*(S^k,\mathrm{p})=\pi$ for all points p on $S^k$. Thus, the radius of injectivity of $S^k$ is $\rho^*_{S^k}=\pi$.

### 3.4. Taylor's Series Expansion of Real Functions on Riemannian Manifolds

Let $f : M \to R$ be a smooth real-valued function. The gradient $\text{grad}_{\text{p}} f$ of $f$ at point p is the linear form on $T_{\text{p}}\mathcal{M}$. Thus, it can be uniquely identified with a vector in $T_{\text{p}}\mathcal{M}$ via the inner product $\text{m}_{\text{p}}(\cdot,\cdot)$ such that $\text{grad}_{\text{p}} f(\mathbf{v})$ corresponds to the directional derivative $\partial_{\mathbf{v}} f$. In a local chart $(U, \phi)$ near p with $\phi(\text{p}) = \mathbf{0}$, the expression of the gradient is

$$\text{grad}_{\phi^{-1}(\mathbf{t})} f = \text{M}_{\phi}^{-1}(\mathbf{t}) \frac{\partial (f \circ \phi^{-1})^T}{\partial \mathbf{t}} = \sum_{l=1}^{d_{\mathcal{M}}} \text{m}^{jl}(\mathbf{t}) \partial_l (f \circ \phi^{-1}).$$

The *Hessian* of $f$ at p in a local chart $(U, \phi)$ near p is given by

$$\text{Hess}_{\phi^{-1}(\mathbf{t})} f = \sum_{j,j'=1}^{d_{\mathcal{M}}} \{\partial_{jj'}(f \circ \phi^{-1}) - \sum_{l=1}^{d_{\mathcal{M}}} \Gamma_{jj'}^{l} \partial_l (f \circ \phi^{-1})\} dt^j dt^{j'}.$$

Let $\phi_{\text{p}}$ be a normal chart at p, i.e. $\phi_{\text{p}}(\text{q}) = \text{Log}_{\text{p}}(\text{q})$, and $f_{\text{p}} = f \circ \text{Exp}_{\text{p}}$. Thus, $f_{\text{p}}(0) = f(\text{p})$. The Taylor's series expansion of $f_{\text{p}}(\mathbf{v})$ around $\mathbf{0}$ is given by

$$f_{\text{p}}(\mathbf{v}) = f_{\text{p}}(0) + J_{f_{\text{p}},0}\mathbf{v} + \frac{1}{2}\mathbf{v}^T H_{f_{\text{p}},0}\mathbf{v} + O(\|\mathbf{v}\|^3),$$

where $J_{f_{\text{p}},0} = [\partial_j f_{\text{p}}(0)]$ and $H_{f_{\text{p}},0} = [\partial_{jj'} f_{\text{p}}(0)]$. In a normal chart, $J_{f_{\text{p}},0}$ reduces to $\text{grad}_{\text{p}} f^T$, and the Christoffel symbols vanishes at the origin such that $H_{f_{\text{p}},0}$ corresponds to the Hessian $\text{Hess}_{\text{p}} f$ of $f$ at p. Thus, for all $\mathbf{v} \in \mathcal{D}(\text{p})$, we have

$$f(\text{Exp}_{\text{p}}(\mathbf{v})) = f(\text{p}) + \text{grad}_{\text{p}} f(\mathbf{v}) + \frac{1}{2}\text{Hess}_{\text{p}} f(\mathbf{v},\mathbf{v}) + O(\|\mathbf{v}\|^3). \tag{10}$$

### 3.5. Lie Groups

A Lie group $G$ is a group together with a smooth manifold structure such that the group operations are compatible with the smooth structure, that is, the operations of multiplication $(a, b) \mapsto ab$ and inversion $a \mapsto a^{-1}$ are smooth maps. Let $G$ be a $C^\infty$ Lie group of dimension $d_G$ and with the identity element $e$. Let $T_a G$ be the tangent space of $G$ at $a \in G$, which is a $d_G$ dimensional linear space, and let $TG$ be the tangent bundle on $G$, which itself is a $2d_G$ dimensional manifold. For $a \in G$, let $L_a$ and $R_a$ be, respectively, the *left* and *right multiplications* by $a$, which are defined by

$$L_a : G \to G, \quad L_a(b) = ab, \quad b \in G,$$
$$R_a : G \to G, \quad R_a(b) = ba, \quad b \in G.$$

These maps are $C^\infty$-diffeomorphisms and the inverses are $L_a^{-1} = L_{a^{-1}}$ and $R_a^{-1} = R_{a^{-1}}$, respectively. They include maps of the tangent bundle to itself given by $L_{a*} : T_hG \to T_{ah}G$ and $R_{a*} : T_hG \to T_{ha}G$ for $a, h \in G$. They are $C^\infty$-diffeomorphisms and their inverses are, respectively, $L_{a*}^{-1} = L_{a^{-1}*}$ and $R_{a*}^{-1} = R_{a^{-1}*}$. Moreover, for any $b \in G$, we have $T_{ab}G = L_{a*}(T_bG)$ and $T_{ba}G = R_{a*}(T_bG)$. The fiber map $L_{a*,b}$ (or $R_{a*,b}$) is the restriction and corestriction of $L_{a*}$ (or $R_{a*}$) to $T_bG$ and is a linear isomorphism from $T_bG$ onto $T_{ab}G$ (or $T_{ba}G$) with their inverse $L_{a*,b}^{-1} = L_{a^{-1}*,ab}$ (or $R_{a*,b}^{-1} = R_{a^{-1}*,ba}$).

A Lie group is equipped with a canonical vector-valued one form, the so called *Maurer-Cartan form* $\omega(X_a) = L_{a^{-1}*}(X_a)$ for $X_a \in T_aG$. Thus, the tangent bundle to $G$ is trivial $TG \cong G \times T_eG$. A *left-invariant* vector is completely defined by its value at the group unity $e$. In particular, there is an isomorphism between the tangent space at the origin and left-invariant vector fields. Since the Lie bracket of such fields is again a left-invariant vector field, the Lie algebra structure on vector fields is inherited by the tangent space at the origin $T_eG$. This algebra is called the Lie algebra of the group $G$ and it is denoted by $\mathfrak{g}$. We also have $T_aG = L_{a*}(\mathfrak{g})$, for any $a \in G$.

The *exponential map of* $G$ *at the unity* $e$ is the map $\mathrm{Exp}_e^G : \mathfrak{g} \to G$ defined as follows. For $\mathbf{v} \in \mathfrak{g}$, the exponential of $\mathbf{v}$ is defined by $\mathrm{Exp}_e^G(\mathbf{v}) = \gamma^G(1; \mathbf{v})$, where $\gamma^G(\cdot; \mathbf{v}) : R \to G$ is the unique one-parameter subgroup of $G$ with $\gamma^G(0; \mathbf{v}) = e$ and $\frac{d}{dt}\gamma^G(0; \mathbf{v}) = \mathbf{v}$. It follows easily from the chain rule that $\mathrm{Exp}_e^G(t\mathbf{v}) = \gamma^G(t; \mathbf{v})$. The map $\gamma^G(\cdot; \mathbf{v})$ may be constructed as the integral curve of either the left- or right-invariant vector field associated with $\mathbf{v}$. The integral curve exists for all real parameters followed by left- or right-translation of the solution near zero. Therefore, $\mathrm{Exp}_e^G$ is globally defined on $\mathfrak{g}$ with $\mathrm{Exp}_e^G(0) = e$, and $\mathrm{Exp}_e^G(-\mathbf{v}) = (\mathrm{Exp}_e^G(\mathbf{v}))^{-1}$ for $\mathbf{v} \in \mathfrak{g}$. Moreover, the exponential map $\mathrm{Exp}_e^G$ is a local $C^\infty$-diffeomorphism around $\mathbf{0} \in \mathfrak{g} = T_eG$.

For $a \in G$, the *exponential map of* $G$ *at* $a$, $\mathrm{Exp}_a^G$, is the unique map from $T_aG$ into $G$ that satisfies the following condition

$$\mathrm{Exp}_a^G \circ L_{a*} = L_a \circ \mathrm{Exp}_e^G \tag{11}$$

on $\mathfrak{g}$. Therefore, $\mathrm{Exp}_a^G$ is globally defined on $T_aG$, $\mathrm{Exp}_a^G(0) = a$, and $\mathrm{Exp}_a^G$ is a local $C^\infty$-diffeomorphism around $0 \in T_aG$. Assume that $X_1, \ldots, X_{d_G}$ is a given basis for

$\mathfrak{g} = T_eG$. Any $\mathbf{v} \in T_eG$ can be uniquely written as

$$\mathbf{v} = \sum_{\ell=1}^{d_G} v^\ell X_\ell.$$

Let $X_{1,a}, \ldots, X_{d_G,a}$ be the (unique) left invariant tangent vector fields with values $X_1, \ldots, X_{d_G}$ at $e$, i.e $X_{\ell,a} = L_{a*}(X_\ell)$ for $\ell = 1, \ldots, d_G$, and $a \in G$. Then, $X_{1,a}, \ldots, X_{d_G,a}$ at $a$ form a basis for $T_aG$, for all $a \in G$, so they define a *trivialization of the tangent bundle* of $G$ as follows:

$$f : TG \to G \times R^{d_G}, \qquad f(\sum_{\ell=1}^{d_G} c^\ell \mathbf{X}_{\ell,a}) = (a, (c^1, \ldots c^{d_G})). \tag{12}$$

Let $< \cdot, \cdot >_e$ be an inner product on $T_eG$ and $< \cdot, \cdot >$ be the Riemannian metric defined as in (13). A *Riemannian* or *pseudo-Riemannian metric* on a Lie group $G$ is *left invariant* if it is preserved under every left multiplication $L_a$, that is,

$$< \mathbf{v}, \mathbf{w} >_b = < L_{a,*}(\mathbf{v}), L_{a,*}(\mathbf{w}) >_{ab}, \quad \text{for} \ \ \mathbf{v}, \mathbf{w} \in T_bG, \ \ \text{and} \ \ b, a \in G.$$

A left-invariant metric is uniquely defined by its restriction to the tangent space to the group at unity, hence by an inner product on $\mathfrak{g}$. Therefore, any inner product $< \cdot, \cdot >_e$ on $T_eG$ can be extended to a (unique) left invariant Riemannian metric $< \cdot, \cdot >= \{< \cdot, \cdot >_a\}_{a \in G}$ on $G$, namely

$$< X, Y >_a := < L_{a^{-1}*}(X), L_{a^{-1}*}(Y) >_e, \quad X, Y \in T_aG, \ a \in G. \tag{13}$$

The associated norm is denoted by $\| \cdot \|_a$ and $\|X\|_a = < X, X >_a^{1/2} = \|L_{a^{-1}*}(X)\|_e$.

It is easy to see that the exponential maps $\mathrm{Exp}_e^G$ and $\mathrm{Exp}_a^G$ defined as above using the algebraic structure of $G$ coincide with the manifold exponential maps defined when $G$ is viewed as a Riemannian manifold (with a left-invariant metric). Moreover, the maximal domain on which the exponential map $\mathrm{Exp}_a^G$ is one-to-one is $\mathcal{D}(a) = L_{a,*}(\mathcal{D}(e))$ and so $\rho^*(G, e) = \rho^*(G, a)$, for all $a \in G$. Therefore, the radius of injectivity of $G$ is $\rho_G^* = \rho^*(G, e)$. Let $\mathrm{Log}_e^G$ and $\mathrm{Log}_a^G$ be the inverse maps of $\mathrm{Exp}_e^G$ and $\mathrm{Exp}_a^G$. We have that $\mathrm{Log}_e^G(b^{-1}) = -\mathrm{Log}_e^G(b)$ provided $b \in D(e)$. For $b \in D(a) = \mathrm{Exp}_a^G(\mathcal{D}(a))$, the geodesic distance from $b$ to $a$ can be expressed as

$$\mathrm{dist}_G(b, a) = \|\mathrm{Log}_a^G(b)\|_a = \|\mathrm{Log}_e^G(a^{-1}b)\|_e. \tag{14}$$

### 3.6. Riemannian Symmetric Spaces

A map $f : M \to \mathcal{M}$ defined on a neighborhood of $\mathrm{p} \in \mathcal{M}$ is said to be a *geodesic symmetry* if it fixes the point p and reverses geodesics through that point, i.e. if $\gamma(\cdot)$ is a geodesic with $\gamma(0) = \mathrm{p}$ and $f(\gamma(t)) = \gamma(-t)$, for any $t$. A Riemannian symmetric space RSS is a connected Riemannian manifold $\mathcal{M}$ with the property that at each point, geodesic symmetries are isometric or distance preserving (Boothby, 1986; Helgason, 1978). They arise in a wide variety of situations in both mathematics and physics. Basic examples of RSS's are Euclidean spaces, $R^d$, spheres, $S^k$, projective spaces, $PR^d$, and hyperbolic spaces, $H^d$, each with their standard Riemannian metric. Symmetric spaces arise naturally from Lie group actions on manifolds. Many common geometric transformations of Euclidean spaces - rotations, translations, dilations, and affine transformations on $R^d$ - form Lie groups. In general, Lie groups can be used to describe transformations of smooth manifolds.

Given a smooth manifold $\mathcal{M}$ and a Lie group $G$, a *smooth group action of $G$ on $\mathcal{M}$* is a smooth mapping $G \times \mathcal{M} \to \mathcal{M}$, $(a, y) \mapsto a \cdot y$, such that $e \cdot y = y$ and $(ab) \cdot y = a \cdot (b \cdot y)$ for all $a, b \in G$ and all $y \in \mathcal{M}$, where $e$ is the identity element of $G$. The group action should be interpreted as a group of transformations of the manifold $\mathcal{M}$, namely, $\{L_a\}_{a \in G}$, where $L_a$ is the action of the group element $a$ on $\mathcal{M}$, $L_a : \mathcal{M} \to \mathcal{M}$, $L_a(y) = a \cdot y$ for $y \in \mathcal{M}$ and $a \in G$. $L_a$ is a smooth diffeomorphism on $\mathcal{M}$ and its inverse is $L_{a^{-1}}$. Given $y \in \mathcal{M}$ a point on $\mathcal{M}$, let $\iota_y$ denote the action of $G$ on the point $y$, i.e. $\iota_y : G \to \mathcal{M}$, $\iota_y(a) = a \cdot y = L_a(y)$ for all $a \in G$. The $\iota_y$ is a smooth map from $G$ into $\mathcal{M}$. For example, for any Lie group $G$, the group multiplication defines a group action of $G$ on itself, and the action of an element $a$ on the group itself is exactly the left-multiplication by $a$. Another example is $\mathrm{SO}(d)$, which is a Lie group and it acts on $R^d$ as rotations, i. e. $R \cdot \mathbf{y} = R\mathbf{y}$ for all $R \in \mathrm{SO}(d)$ and $\mathbf{y} \in R^d$.

We now introduce some common concepts related to group actions. The *orbit* of a point $y \in \mathcal{M}$ is defined as $G(y) = \{a \cdot y \,|\, a \in G\}$. The orbits form a partition of $\mathcal{M}$, and we say that two points $y, y' \in \mathcal{M}$ are equivalent if they belong to the same orbit. In the case that $\mathcal{M}$ consists of a single orbit, we say that the group action is *transitive* or $G$ acts *transitively* on $\mathcal{M}$, and we call $\mathcal{M}$ a *homogeneous* space. The *isotropy* subgroup of a point $y$ of $\mathcal{M}$ is defined as $G_y = \{a \in G \,|\, a \cdot y = y\}$. For example, for the action of the group, $\mathrm{SO}(2)$, the isotropy subgroup of the zero vector is $G_{\mathbf{0}} = \mathrm{SO}(2)$ and for any non-zero vector $\mathbf{y} \in R^2$, the isotropy group $G_{\mathbf{y}}$ reduces to the trivial subgroup $\{I_2\}$.

Let $H$ be a closed Lie subgroup of the Lie group $G$. Then the *left coset* of an element $a \in G$ is defined by $aH = \{ah \mid h \in H\}$. The space of all such cosets is called a quotient space of the group $G$ with respect to the subgroup $H$, denoted by $G/H$, and it is a smooth manifold with the quotient topology. When a Lie group $G$ acts smoothly on a smooth manifold $\mathcal{M}$ for any $y \in \mathcal{M}$, there is a natural bijection from the orbit $G(y)$ onto the quotient manifold given by the mapping $a \cdot y \mapsto aG_y$, which is well-defined and smooth, so $G(y) \cong G/G_y$.

Now let $\mathcal{M}$ be a symmetric space and choose an arbitrary base point $\mathrm{p} \in \mathcal{M}$. We can always view $\mathcal{M}$ as a homogeneous space $\mathcal{M} \cong G/G_{\mathrm{p}}$, where $G$ is a connected group of isometries of $\mathcal{M}$ and the isotropy subgroup $G_{\mathrm{p}}$ is compact. We call $G$ a *group of isometries* of $\mathcal{M}$ if for all $a \in G$, $\mathrm{dist}_{\mathcal{M}}(y, z) = \mathrm{dist}_{\mathcal{M}}(a \cdot y, a \cdot z)$ for all $y, z \in \mathcal{M}$. Any Lie group $G$ can be viewed as a symmetric space with a Riemannian structure induced by an inner product on $T_eG$, and $G$ acting on itself by left multiplication. Obviously, this action is transitive and the isotropy subgroups are trivial, i.e. $G_a = \{e\}$, for all $a \in G$.

A very common example of a symmetric space is $S^2$, which is a 2-dimensional compact Riemannian manifold. The Lie group, SO(3), of all rotations in $R^3$ acts smoothly and transitively on $S^2$. For example, let us choose the north pole $\mathrm{p} = (0, 0, 1) \in S^2$ as the base point. It is easy to see that the orbit of p is the entire sphere and thus $S^2$ is a homogeneous space. The isotropy subgroup of p is the group of all rotations about the $z$-axis in $R^3$, which can be identified with the group of 2D rotations, SO(2). Hence, $S^2$ can be naturally identified with the quotient space SO(3)/SO(2). Similarly, the $k$-dimensional unit sphere, $S^k$, can be identified as the quotient space $\mathrm{SO}(k+1)/\mathrm{SO}(k)$. The sphere $S^k$ is a compact Riemannian manifold.

Other examples of symmetric spaces can be obtained by taking Cartesian products of symmetric spaces. Consider two manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$ and two Lie groups $G_1$ and $G_2$ so that $G_j$ acts transitively on $\mathcal{M}_j$ for $j = 1, 2$. Thus, the group $G = G_1 \times G_2$ is a Lie group and acts transitively on the manifold $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$. Given a base point $\mathrm{p} = (\mathrm{p}_1, \mathrm{p}_2)$ in $\mathcal{M}$, the isotropy subgroup of p in $G$ is $G_{\mathrm{p}} = G_{1,\mathrm{p}_1} \times G_{2,\mathrm{p}_2}$. Thus, we can write $\mathcal{M}_1 \times \mathcal{M}_2$ as a homogeneous space $G/G_{\mathrm{p}} = (G_1/G_{1,\mathrm{p}_1}) \times (G_2/G_{2,\mathrm{p}_2})$.

An example of a symmetric space used in the study of 3D geometric objects is the

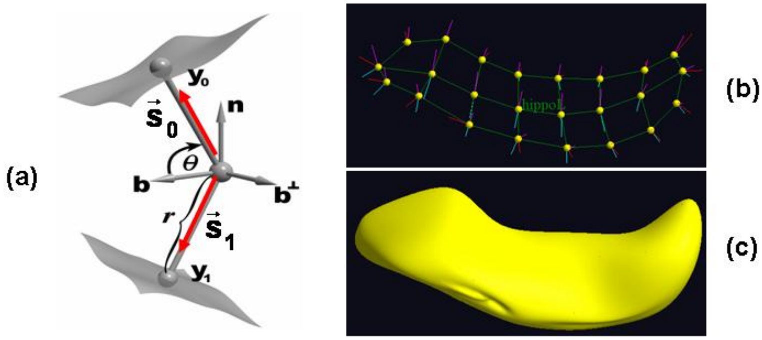


**Figure 3.** (a) A medial representation model $\mathbf{m} = (\mathbf{O}, r, \mathbf{s}_0, \mathbf{s}_1)$ at an atom, where $\mathbf{O}$ is the center of the inscribed sphere, $r$ is the common spoke length, abd $\{\mathbf{s}_0, \mathbf{s}_1\}$ are the two unit spoke directions; (b) a skeleton of a hippocampus with 24 medial atoms,; (c) the smoothed surface of the hippocampus.

space of medial atoms, $\mathcal{M} = R^3 \times R^+ \times S^2 \times S^2$ (Shi et al., 2012). See Figure 3 for an illustration. The group $G = R^3 \times R^+ \times \mathrm{SO}(3) \times \mathrm{SO}(3)$ acts smoothly on $\mathcal{M}$. For an element $a = (O', r', R_0, R_1) \in G$ and an medial atom $\mathrm{q} = (O, r, \mathbf{s}_0, \mathbf{s}_1) \in \mathcal{M}$, the group action is defined by

$$a \cdot \mathrm{q} = (O + O', rr', R_0\mathbf{s}_0, R_1\mathbf{s}_1),$$

which is a transitive action. Consider the atom p located at $O = (0, 0, 0)$ with radius $r = 1$ and spokes $\mathbf{s}_0 = \mathbf{s}_1 = (0, 0, 1)$. Then, the isotropy subgroup of p is $G_{\mathrm{p}} = \{0\} \times \{1\} \times \mathrm{SO}(2) \times \mathrm{SO}(2)$, and we can write the medial atom space as the quotient space $\mathcal{M} = R^3 \times R^+ \times (\mathrm{SO}(3)/\mathrm{SO}(2)) \times (\mathrm{SO}(3)/\mathrm{SO}(2))$.

From now on, it is assumed that the manifold $\mathcal{M}$ is a symmetric space, $\mathcal{M} = G/G_{\mathrm{p}}$ with $G$ being a Lie group of isometries acting transitively on $\mathcal{M}$. Geodesics on $\mathcal{M}$ are computed through the action of $G$ on $\mathcal{M}$. Due to the transitive action of the group $G$ of isometries on $\mathcal{M}$, it suffices to consider only the geodesic starting at the base point p. For an arbitrary point $y \in \mathcal{M}$, geodesics starting from $y$ are of the form $a \cdot \gamma(\cdot)$, where $\gamma(\cdot)$ is a geodesic starting from p with $\gamma(0) = \mathrm{p}$ and $y = a \cdot \mathrm{p}$ for some $a \in G$. Due to the local uniqueness of geodesics, if $y = a' \cdot \mathrm{p}$ for some other $a' \in G$, then $a \cdot \gamma(\cdot) = a' \cdot \gamma(\cdot)$.

Geodesics on $\mathcal{M}$ starting from p are the images of the action of a 1-parameter subgroup of $G$ acting on the base point p. In other words, for any geodesic $\gamma$ on $\mathcal{M}$, $\gamma(\cdot) : I \to \mathcal{M}$, starting from p, there exists a 1-parameter subgroup $c(\cdot) : R \to G$ such that $\gamma(t) = c(t) \cdot \mathrm{p}$ for all $t \in I$. The manifold exponential map $\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}$ at the base point p is defined by

$$\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(t\mathbf{v}) = \gamma(t; \mathrm{p}, \mathbf{v}) = c(t; e, \mathbf{u}) \cdot \mathrm{p},$$

where $\gamma(0; \mathrm{p}, \mathbf{v}) = \mathrm{p}$, $\frac{d}{dt}\gamma(0; \mathrm{p}, \mathbf{v}) = \mathbf{v} \in T_{\mathrm{p}}\mathcal{M}$, $c(0; e, \mathbf{u}) = e$, and $\frac{d}{dt}c(0; \mathrm{p}, \mathbf{u}) = \mathbf{u} \in T_e G$ with $\mathbf{u}$ so that $\iota_{\mathrm{p}*,e}(\mathbf{u}) = \mathbf{v}$ for small $t \in R$. That is,

$$\mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(t\, \iota_{\mathrm{p}*,e}(\mathbf{u})) = \mathrm{Exp}_{\mathrm{p}}^{G}(t\mathbf{u}) \cdot \mathrm{p},$$

for all $\mathbf{u} \in T_e G$ and $t \in R$ with small $\|t\mathbf{u}\|$.

Moreover, the manifold exponential map $\mathrm{Exp}_{\mathrm{q}}^{\mathcal{M}}$ of $\mathcal{M}$ at a point q is defined by

$$\mathrm{Exp}_{\mathrm{q}}^{\mathcal{M}}(L_{a*,\mathrm{p}}\mathbf{v}) = a \cdot \mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(\mathbf{v})$$

for any $a \in G$ with $\mathrm{q} = a \cdot \mathrm{p}$ and any small $\mathbf{v} \in T_{\mathrm{p}}\mathcal{M}$, where $L_a$ is the action of the element $a$ on the points of $\mathcal{M}$. Due to the uniqueness of geodesics, if $\mathrm{q} = a_1 \cdot \mathrm{p} = a_2 \cdot \mathrm{p}$ with $a_1, a_2 \in G$ and $\mathbf{w} = L_{a_1*,\mathrm{p}}(\mathbf{v}_1) = L_{a_2*,\mathrm{p}}(\mathbf{v}_2)$ with $\mathbf{v}_1, \mathbf{v}_2 \in T_{\mathrm{p}}\mathcal{M}$, then $a_1 \cdot \mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(\mathbf{v}_1) = a_2 \cdot \mathrm{Exp}_{\mathrm{p}}^{\mathcal{M}}(\mathbf{v}_2)$. Since $G$ is a group of isometries on $\mathcal{M}$, the radius of injectivity $\mathrm{Exp}_{\mathrm{q}}^{\mathcal{M}}$ of $\mathcal{M}$ at q is independent of the point q, so $\rho_{\mathcal{M}}^* = \rho^*(\mathcal{M}, \mathrm{p})$.

The unit sphere $S^k$ is a compact Riemannian manifold of dimension $d$ and injectivity radius $\rho = \pi$. The tangent space at $\mathrm{q} \in S^k$ is

$$T_{\mathrm{q}}S^k = \{\mathbf{v} \in R^{k+1} : \mathbf{v}^T\mathrm{q} = 0\}.$$

The tangent space is endowed with the metric tensor from $R^{k+1}$, $m_{\mathrm{q}}(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{v}_1^T\mathbf{v}_2$ for all $\mathbf{v}_1, \mathbf{v}_2 \in T_{\mathrm{q}}S^k$. The geodesic distance between two points $\mathrm{q}_1, \mathrm{q}_2 \in S^k$ is given by $\mathrm{dist}_{\mathcal{M}}(\mathrm{q}_1, \mathrm{q}_2) = \arccos(\mathrm{q}_1^T\mathrm{q}_2)$, which lies between 0 and $\pi$. The exponential map takes the form

$$\mathrm{Exp}_{\mathrm{q}} : T_{\mathrm{q}}S^k \to S^k,\ \mathrm{Exp}_{\mathrm{q}}(\mathbf{v}) = \cos(\|\mathbf{v}\|)\mathrm{q} + \frac{\sin(\|\mathbf{v}\|)}{\|\mathbf{v}\|}\mathbf{v}.$$

It is a diffeomorphism from $B(0, \pi) \subset T_{\mathrm{q}}S^k$ onto $S^k \setminus \{-\mathrm{q}\}$, and the logarithmic map is given by

$$\mathrm{Log}_{\mathrm{q}} : S^k \setminus \{-\mathrm{q}\} \to B(0, \pi),\ \mathrm{Exp}_{\mathrm{q}}(\mathrm{q}_1) = \frac{\arccos(\mathrm{q}_1^T\mathrm{q})}{\sqrt{1 - (\mathrm{q}_1^T\mathrm{q})^2}}\left(\mathrm{q}_1 - (\mathrm{q}_1^T\mathrm{q})\mathrm{q}\right),$$

for all $\mathrm{q}_1 \in S^k$ with $\mathrm{q}_1 \neq -\mathrm{q}$.

### *3.7. Symmetric Positive-definite Matrices*

We review some basic facts about the geometric structure of $\mathrm{Sym}^+(k)$ (Schwartzman, 2006; Lang, 1999; Terras, 1988; Fletcher et al., 2004; Batchelor et al., 2005; Zhu et al., 2009; Yuan et al., 2012; Osborne et al., 2013). Let $\mathrm{Sym}(k)$ be the set of $k \times k$ symmetric matrices with real entries, which is a topological linear space of dimension $k(k+1)/2$. The $\mathrm{Sym}^+(k)$ is an open subset of $\mathrm{Sym}(k)$ and $T_\mathrm{q}\mathrm{Sym}^+(k)$ is a copy of $\mathrm{Sym}(k)$ for $\mathrm{q} \in \mathrm{Sym}^+(k)$. Let $\mathrm{q} = C_\mathrm{q} C_\mathrm{q}^T$ be the Cholesky decomposition of q, where $C_q$ is a lower triangular matrix with strictly positive diagonal entries. Then, for $\mathrm{q}, \mathrm{q}' \in \mathrm{Sym}^+(k)$, the map $(\mathrm{q}, \mathrm{q}') \to \mathrm{q} \circ \mathrm{q}' := C_\mathrm{q} \mathrm{q}' C_\mathrm{q}^T$ induces a (non-commutative) Lie group structure on $\mathrm{Sym}^+(k)$, denoted by $G$. The unit element of $G$ is the identity matrix $I_k$ and the inverse of a matrix $\mathrm{q} \in G$ with respect to the operation on $G$ is $\mathrm{q}^{\sim 1} = C_\mathrm{q}^{-1} C_\mathrm{q}^{-\top}$. The Lie group $G$ can be entirely covered with a single chart. We also have $L_\mathrm{q}(\mathrm{q}') = C_\mathrm{q} \mathrm{q}' C_\mathrm{q}^T$ and $L_{\mathrm{q}*}(A) = C_\mathrm{q} A C_\mathrm{q}^T$ for $\mathrm{q}, \mathrm{q}' \in \mathrm{Sym}^+(k)$ and $A \in \mathrm{Sym}(k)$. The associated Lie algebra is $\mathfrak{sym}(k) = \mathrm{Sym}(k)$ with the bracket map being $[A_1, A_2] = A_1 A_2 - A_2 A_1$ for $A_1, A_2 \in \mathrm{Sym}(k)$. Let $\exp(\cdot)$ and $\log(\cdot)$ be, respectively, the matrix exponential and logarithm. The manifold exponential at $I_k$, $\mathrm{Exp}_{I_k}$, is the matrix exponential $\exp(\cdot)$ and its inverse map is $\mathrm{Log}_{I_k} = \mathrm{Exp}_{I_k}^{-1} = \log(\cdot)$. For $A \in \mathrm{Sym}(k)$ and $\mathrm{q}' \in \mathrm{Sym}^+(k)$, we have

$$\begin{aligned}
\mathrm{Exp}_\mathrm{q}(A) &= (L_\mathrm{q} \circ \mathrm{Exp}_{I_k} \circ L_{\mathrm{q}^{\sim 1}*})(A) = C_\mathrm{q} \exp(C_\mathrm{q}^{-1} A C_\mathrm{q}^{-\top}) C_\mathrm{q}^T, \\
\mathrm{Log}_\mathrm{q}(\mathrm{q}') &= \mathrm{Exp}_\mathrm{q}^{-1}(\mathrm{q}') = C_\mathrm{q} \log(C_\mathrm{q}^{-1} \mathrm{q}' C_\mathrm{q}^{-\top}) C_\mathrm{q}^T.
\end{aligned}$$

We consider the trace norm $\|A\| = \sqrt{\mathrm{tr}(A^2)}$ on $Sym(k)$, identified as $T_{I_k}\mathrm{Sym}^+(k)$. This norm is actually the 2-norm of the in $R^{k^2}$ of the vectorized form of the matrix. This allows to introduce the following metric on $\mathrm{Sym}^+(k)$

$$< A_1, A_2 >_\mathrm{q} := < L_{\mathrm{q}^{\sim 1}} * (A_1), L_{\mathrm{q}^{\sim 1}} * (A_2) >_{I_k} = \mathrm{tr}(A_1 \mathrm{q}^{-1} A_2 \mathrm{q}^{-1}),$$

for $A_1, A_2 \in T_\mathrm{q}\mathrm{Sym}^+(k)$ and $\mathrm{q} \in \mathrm{Sym}^+(k)$. This metric induces a Riemannian structure on the group $\mathrm{Sym}^+(k)$, and the above $\mathrm{Exp}_\mathrm{q}$ and $\mathrm{Log}_\mathrm{q}$ are the Riemannian exponential and logarithmic maps, respectively. The curve $t \to \gamma(t; \mathrm{q}, A) := \mathrm{Exp}_\mathrm{q}(tA)$ is the geodesic curve starting from q with initial tangent vector $A \in T_{\rm q}{\rm Sym}^+(k)$. The radius of injectivity is $\rho^*({\rm Sym}(k)) = \rho^*({\rm Sym}(k), I_k) = \infty$.

We introduce the intrinsic regression model for ${\rm Sym}^+(k)-$valued responses. Suppose that we observe $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where $y_i \in {\rm Sym}^+(k)$ for all $i$. We define a function $f(\mathbf{x}, \boldsymbol{\beta})$ given by

$$f(\cdot,\cdot) : R^{d_{\mathbf{x}}} \times R^{d_\beta} \to R^{k(k+1)/2} \ \text{ with } \ f(\mathbf{0}, \cdot) = f(\cdot, \mathbf{0}) = \mathbf{0}.$$

An example of $f(\cdot,\cdot)$ is $f(\mathbf{x}_i, \boldsymbol{\beta}) = B\mathbf{x}_i$ (Zhu et al., 2009), where $B$ is an $k(k+1)/2 \times d_{\mathbf{x}}$ matrix of regression coefficients and $\boldsymbol{\beta}$ includes all components of $B$. Let $\{E_{j\ell} : 1 \le \ell \le j \le k\}$ be the canonical basis of ${\rm Sym}(k)$, where $E_{j\ell}$ is the $m \times m$ matrix with the $(j, \ell)$ and $(\ell, j)$ entries being 1 and 0 otherwise; let $f(\mathbf{x}_i, \boldsymbol{\beta})_{j(j-1)/2+\ell}$ be the $j(j-1)/2 + \ell$-th component of $f(\mathbf{x}_i, \boldsymbol{\beta})$. We consider a single-center link function given by

$$\boldsymbol{\mu}(\mathbf{x}, {\rm q}, \boldsymbol{\beta}) = {\rm Exp}_{\rm q}(\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta})) = C_{\rm q} \exp(C_{\rm q}^{-1}\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta})C_{\rm q}^{-\top})C_{\rm q}^T,$$

where $\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta}) = \sum_{j=1}^k \sum_{\ell=1}^j f(\mathbf{x}_i, \boldsymbol{\beta})_{j(j-1)/2+\ell} E_{j\ell}$ and ${\rm q} = C_{\rm q} C_{\rm q}^T \in {\rm Sym}^+(k)$ is the 'center'. The *rotated residual* is given by

$$\mathcal{E}(y_i, \mathbf{x}_i, {\rm q}, \boldsymbol{\beta}) = \log(C_i({\rm q}, \boldsymbol{\beta})^{-1} y_i C_i({\rm q}, \boldsymbol{\beta})^{-\top}),$$

where $C_i({\rm q}, \boldsymbol{\beta}) C_i({\rm q}, \boldsymbol{\beta})^T$ is the Cholesky decomposition of $f(\mathbf{x}_i, {\rm q}, \boldsymbol{\beta})$.

### 3.8. *Special Orthogonal Group* ${\rm SO}(k)$

We review some basic facts about the geometric structure of ${\rm SO}(k)$ (Grenander et al., 1998; Moakher, 2002; Gallier and Xu, 2002). This is a compact ($C^\infty$) submanifold of $R^{k\times k}$ of dimension $k(k-1)/2$ as well as a Lie group with respect to matrix multiplication. The unit element of ${\rm SO}(k)$ is the identity matrix $I_k$ and its associated Lie algebra $\mathfrak{so}(k) = T_{I_k}{\rm SO}(k)$ is the linear space of all $k \times k$ skew-symmetric matrices q, i.e. ${\rm q}^T = -{\rm q}$, denoted by ${\rm SkewSym}(k)$. For ${\rm q} \in {\rm SO}(k)$, $T_{\rm q}{\rm SO}(k)$ is given by

$$T_{\rm q}{\rm SO}(k) = \{A \in R^{k\times k} : A^T = -{\rm q}^T A {\rm q}^T\} = {\rm q}\ \ {\rm SkewSym}(k).$$

We consider the trace metric on $T_{\rm q}{\rm SO}(k)$. The trace metric is also a left-invariant Riemannian metric on ${\rm SO}(k)$. Specifically, since ${\rm q}\,{\rm q}^T = I_k$, for $A_1, A_2 \in T_{I_m}{\rm SO}(k)$, we have

$$< {\rm q}A_1, {\rm q}A_2 >_{\rm q} = {\rm tr}[({\rm q}A_1)^T({\rm q}A_2)] = {\rm tr}(A_1^T A_2) = < A_1, A_2 >_{I_k}.$$

The Lie exponential map at $I_k$ is given by the usual matrix exponentiation. Although the Lie logarithmic map at $I_k$ has a closed form, the formula for a general $k$ is quite complicated. We present the Lie logarithmic map for $k = 2$ and 3 in the supplementary document. Generally, the Lie exponential map of $A \in T_{\mathrm{q}}\mathrm{SO}(k)$ at $\mathrm{q} \in \mathrm{SO}(k)$ and its corresponding Lie logarithmic map are, respectively, given by

$$\mathrm{Exp}_{\mathrm{q}}(A) = \mathrm{q}\,\mathrm{Exp}_{I_k}(\mathrm{q}^T A) = \mathrm{q}\exp(\mathrm{q}^T A) \text{ and } \mathrm{Log}_{\mathrm{q}}(\mathrm{q}') = \mathrm{q}\,\mathrm{Log}_{I_k}(\mathrm{q}^T \mathrm{q}').$$

We introduce the intrinsic regression model for $\mathrm{SO}(k)-$valued responses. Suppose that we observe $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where $y_i \in \mathrm{SO}(k)$ for all $i$. We define a function $f(\mathbf{x}, \boldsymbol{\beta})$ given by

$$f(\cdot,\cdot) : R^{d_{\mathbf{x}}} \times R^{d_\beta} \to R^{k(k-1)/2} \quad \text{with} \quad f(\mathbf{0},\cdot) = f(\cdot,\mathbf{0}) = \mathbf{0}.$$

An example of $f(\cdot,\cdot)$ is $f(\mathbf{x}_i, \boldsymbol{\beta}) = B_1\mathbf{x}_i$, where $B_1$ is a $k(k-1)/2 \times d_{\mathbf{x}}$ matrix of regression coefficients and $\boldsymbol{\beta}$ includes all components of $B_1$. Let $\{\tilde{E}_{j\ell} : 1 \leq \ell \leq j \leq k\}$ be the basis of SkewSym$(k)$, where $\tilde{E}_{j\ell}$ is a $k \times k$ matrix with the $(j, \ell)$ and $(\ell, j)$ entries being $(-1)^{j+\ell-1}$ and $(-1)^{j+\ell}$, respectively, and 0 otherwise. Let $\mathrm{q} \in \mathrm{SO}(k)$ be the 'center', and we consider a single-center link function given by

$$\boldsymbol{\mu}(\mathbf{x}, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Exp}_{\mathrm{q}}(\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta})) = \mathrm{q}\exp(\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta})),$$

where $\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta}) = \sum_{j=2}^{k}\sum_{\ell=1}^{j-1} f(\mathbf{x}_i, \boldsymbol{\beta})_{(j-2)(j-1)/2+\ell}\tilde{E}_{j\ell} \in \mathrm{SkewSym}(k)$. The *rotated residual* is given by

$$\mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Log}_{I_k}(\exp(-\mathbf{u}(\mathbf{x}_i, \boldsymbol{\beta}))\mathrm{q}^T y_i).$$

The explicit form of $\mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}, \boldsymbol{\beta})$ for $k = 2, 3$ can be found in the supplementary document.

### 3.9. *Unit circle $S^1$ in the complex plane*

Let $S^1 = \{z = \cos(\phi) + j\sin(\phi) : \phi \in R\}$ be the unit circle in the complex plane $\mathbf{C}$, where $j = \sqrt{-1}$. The $S^1$ with the usual multiplication of complex numbers forms a compact 1-dimensional $C^\infty$ Lie group with 1 as the unity. The tangent space of $S^1$ at $a = \cos(\theta_0) + j\sin(\theta_0) \in S^1$ is given by $T_a(S^1) = \{t\,(-\sin(\theta_0) + j\cos(\theta_0)) : t \in R\}$, which is a 1-dimensional real linear subspace of $\mathbf{C}$ formed by all $z = z_x + jz_y$'s that are orthogonal to $a$ as vectors in $R^2$. The Lie algebra of $S^1$ is $T_1S^1 = \{jt : t \in R\}$ and the exponential map at unity is given by $\mathrm{Exp}_1(jt) = e^{jt} = \cos t + j \sin t$. Thus, we have

$$\mathrm{Exp}_a\left(t\left(-\sin(\theta_0) + j\cos(\theta_0)\right)\right) = \cos(t+\theta_0) + j\sin(t+\theta_0).$$

Geometrically, $\mathrm{Exp}_a$ "wraps" the tangent line at $a$ around the circle, and thus the injectivity radius $\rho$ equals $\pi$.

Suppose that we observe $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where $y_i = \cos(\phi_i) + j\sin(\phi_i) \in S^1$ for all $i$. We define

$$I(\cdot,\cdot) : R^{d_{\mathbf{x}}} \times R^{d_\beta} \to R \text{ with } I(\mathbf{0}, \cdot) = I(\cdot, \mathbf{0}) = 0.$$

For an $a \in S^1$, we consider a single-center link function and its corresponding *rotated residual*, which are, respectively, given by

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{x}_i, a, \boldsymbol{\beta}) &= a\, e^{jI(\mathbf{x}_i,\boldsymbol{\beta})} = e^{j(\theta_0 + I(\mathbf{x}_i,\boldsymbol{\beta}))}, \\ \mathcal{E}_i(a, \boldsymbol{\beta}) &= j(\phi_i - \theta_0 - I(\mathbf{x}_i, \boldsymbol{\beta}))_{\mathrm{mod}\, 2\pi}, \end{aligned}$$

where $t_{\mathrm{mod}\,2\pi}$ is the unique number in $(-\pi, \pi]$ so that $t - t_{\mathrm{mod}\,2\pi} \in 2\pi\mathbb{Z}$. Thus, the intrinsic regression model is written as

$$E[\mathcal{E}_i(a, \boldsymbol{\beta}) |\, \mathbf{x}_i] = 0, \qquad i = 1, \ldots, n. \tag{15}$$

### *3.10. Lie Logarithmic Maps of SO(2) and SO(3)*

When $k = 2$, $SO(2)$ is the set of all $2\times2$ matrices of the form $\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$ with $x^2+y^2 = 1$ for $x, y \in R$. The group $SO(2)$ of rotations in $R^2$ is isomorphic with $S^1$. The canonical isomorphism is $\begin{pmatrix} x & -y \\ y & x \end{pmatrix} \to z = x + jy$. A $2 \times 2$ skew-symmetric matrix $B$ can be written as $B = \lambda J$, where $\lambda \in R$ and

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

There is a canonical isomorphism from the linear space $T_{I_2}SO(2)$ into the space of pure imaginary numbers, $jR$, namely, $\lambda J \mapsto j\lambda$ for $\lambda \in R$. It can be shown that

$$e^B = e^{\lambda J} = \cos(\lambda) I_2 + \sin(\lambda) J.$$

and, since $J^2 = -I_2$, it follows

$$\cos(\lambda) = \frac{1}{2}\text{tr}(e^B) \qquad \sin(\lambda) = -\frac{1}{2}\text{tr}(e^B J).$$

Thus $e^B$ determines $\lambda$ uniquely up to an additive multiple of $2\pi$.

Given a rotation matrix $O \in SO(2)$, the Lie logarithmic map at $I_2$ of $O$ is given by

$$\text{Log}_{I_2}(O) = \lambda J, \tag{16}$$

where $\cos(\lambda) = 0.5\text{tr}(R)$ and $\sin(\lambda) = -0.5\text{tr}(RJ)$ for $\lambda \in (-\pi, \pi]$. Thus, when $SO(2)$ is endowed with the trace metric, it follows immediately that the radius of injectivity of $SO(2)$ is $\rho^*_{SO(2)} = \sqrt{2}\,\pi$.

When $k = 3$, a $3 \times 3$ skew-symmetric matrix $B$ is of the form

$$B = \begin{pmatrix} 0 & -c_1 & c_2 \\ c_1 & 0 & -c_3 \\ -c_2 & c_3 & 0 \end{pmatrix},$$

and letting $\lambda = \sqrt{c_1^2 + c_2^2 + c_3^2}$, we have the well-known Rodrigues formula

$$e^B = I_3 + \frac{\sin(\lambda)}{\lambda} B + \frac{[1 - \cos(\lambda)]}{\lambda^2} B^2.$$

It may be more convenient to normalize $B$ such that one can write $B = \lambda B_1$ (or, equivalently, $B_1 = B/\lambda$, assuming $\lambda \neq 0$). In this case, $e^B$ can be written as

$$e^B = e^{\lambda B_1} = I_3 + \sin\lambda\, B_1 + (1 - \cos\lambda) B_1^2.$$

Observing that $\text{tr}(e^B) = 1 + 2\cos(\lambda)$ and $0.5[e^B - (e^B)^T] = \sin(\lambda)\, B_1$, the logarithmic map at $I_3$ of a rotation $O \in SO(3)$ is given by

$$\text{Log}_{I_3}(O) = \lambda B_1, \tag{17}$$

where $\lambda = \arccos((\text{tr}(O) - 1)/2)$ and $B_1 = (O - O^T)/(2\sin\lambda)$. When $\lambda = 0$ or $\lambda = \pi$, the above formulae cannot be used. When $\lambda = 0$, we have $O = I_3$ and $B_1 = 0$, so $\text{Log}_I(O) = \text{Log}_I(I_3) = O_3$, and Eq. (17) still holds. When $\lambda = \pi$, we need to find $B_1$ such that $B_1^2 = \frac{1}{2}(O - I_3)$. As $B_1$ is a skew-symmetric matrix, this amounts to solving a simple system of equations with three unknowns. When $SO(3)$ is endowed with the trace metric, elementary calculations on the Rodrigues formula yield that $\mathrm{Exp}_I$ is one-to-one on the ball $B(\mathbf{0}, \sqrt{2}\,\pi)$ in $T_{I_2}SO(3)$, but not on any ball $B(\mathbf{0}, \rho)$ with $\rho > \sqrt{2}\,\pi$. Therefore, the radius of injectivity of $SO(3)$ is $\rho^*_{SO(3)} = \sqrt{2}\,\pi$.

Suppose we observe an element $\mathrm{q}_i \in SO(k)$ and a $d_x \times 1$ covariate vector $\mathbf{x}_i$ for $i = 1, \ldots, n$. We consider an intercept rotation matrix $O$. Then, for a given map $f(\mathbf{x}_i, \boldsymbol{\beta})$ with $f(\cdot,\cdot) : R^{d_x} \times R^{d_\beta} \to R^{k(k-1)/2}$ with $f(\mathbf{0}, \cdot) = \mathbf{0}$, let $\Lambda(\mathbf{x}_i, \boldsymbol{\beta}) = \sum_{k=2}^{m} \sum_{\ell=1}^{k-1} f(\mathbf{x}_i, \boldsymbol{\beta})_{(k-1)(k-2)/2+\ell} X_{k\ell} \in \mathrm{SkewSym}(k)$, and consider the "directional" matrix $\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{q}\Lambda(\mathbf{x}_i, \boldsymbol{\beta})$ as a tangent vector to $SO(k)$ at $\mathrm{q} \in SO(k)$. By considering the "conditional mean"

$$\mu(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Exp}_{\mathrm{q}}(\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})) = \mathrm{q}\exp(\Lambda(\mathbf{x}_i, \boldsymbol{\beta})),$$

the intrinsic residual is given by

$$\mathcal{E}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathcal{E}_i(\mathrm{q}, \boldsymbol{\beta}) = \mathrm{Log}(e^{-\Lambda(\mathbf{x}_i, \boldsymbol{\beta})}\mathrm{q}^T \mathrm{q}_i).$$

When $k = 2$, both the manifold $SO(2)$ and the linear space SkewSym(2) have dimension 1, so $f(\mathbf{x}_i, \boldsymbol{\beta})$ is a scalar map. We have

$$\begin{aligned} \Lambda(\mathbf{x}_i, \boldsymbol{\beta}) &= f(\mathbf{x}_i, \boldsymbol{\beta})J = \begin{pmatrix} 0 & -f(\mathbf{x}_i, \boldsymbol{\beta}) \\ f(\mathbf{x}_i, \boldsymbol{\beta}) & 0 \end{pmatrix} \\ \mathrm{q} &= \begin{pmatrix} \cos(\theta_0) & -\sin(\theta_0) \\ \sin(\theta_0) & \cos(\theta_0) \end{pmatrix}, \quad \mathrm{q}_i = \begin{pmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{pmatrix} \end{aligned}$$

where $\theta_i$ are in $(-\pi, \pi]$ for $i = 0, \ldots, n$. The "conditional mean" becomes

$$\mu(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \begin{pmatrix} \cos(\theta_0 + f(\mathbf{x}_i, \boldsymbol{\beta})) & -\sin(\theta_0 + f(\mathbf{x}_i, \boldsymbol{\beta})) \\ \sin(\theta_0 + f(\mathbf{x}_i, \boldsymbol{\beta})) & \cos(\theta_0 + f(\mathbf{x}_i, \boldsymbol{\beta})) \end{pmatrix}$$

and the "intrinsic residual" is

$$\mathcal{E}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = (\phi_i - \theta_0 - \mathrm{I}(\mathbf{x}_i, \boldsymbol{\beta}))_{\mathrm{mod}\ 2\pi} J.$$

We observe that we recapture, via the canonical isomorphism between $SO(2)$ and $S^1$, the intrinsic model presented in Example 2. When $SO(2)$ is endowed with the trace metric, the Riemannian distance on $SO(2)$ between two rotations is a constant multiple of the Riemannian distance on $S^1$ between their counterparts in $S^1$; the multiplicative factor is $\sqrt{2}$.

When $k=3$, the manifold $SO(3)$ and the linear space SkewSym(3) have dimension 3, and the above "conditional mean" becomes

$$\mu(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{q}\left(I_3 + \frac{\sin\lambda(\mathbf{x}_i,\boldsymbol{\beta})}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})}\Lambda(\mathbf{x}_i,\boldsymbol{\beta}) + \frac{(1-\cos\lambda(\mathbf{x}_i,\boldsymbol{\beta}))}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})^2}\Lambda(\mathbf{x}_i,\boldsymbol{\beta})^2\right),$$

where $\lambda(\mathbf{x}_i,\boldsymbol{\beta}) = \|f(\mathbf{x}_i,\boldsymbol{\beta})\|$, and the "intrinsic residual" is

$$\begin{aligned}\mathcal{E}(\mathbf{x}_i,\mathrm{q},\boldsymbol{\beta}) \;=\; & \frac{\theta_i(\mathrm{q},\boldsymbol{\beta})}{2\sin(\theta_i(\mathrm{q},\boldsymbol{\beta}))}\times \\ & \Big[(\mathrm{q}^T\mathrm{q}_i - \mathrm{q}_i^T\mathrm{q}) \\ & - \frac{\sin(\lambda(\mathbf{x}_i,\boldsymbol{\beta}))}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})}\left(\Lambda(\mathbf{x}_i,\boldsymbol{\beta})\mathrm{q}^T\mathrm{q}_i + \mathrm{q}_i^T\mathrm{q}\Lambda(\mathbf{x}_i,\boldsymbol{\beta})\right) \\ & + \frac{[1-\cos(\lambda(\mathbf{x}_i,\boldsymbol{\beta}))]}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})^2}\left(\Lambda(\mathbf{x}_i,\boldsymbol{\beta})^2\mathrm{q}^T\mathrm{q}_i - \mathrm{q}_i^T\mathrm{q}\Lambda(\mathbf{x}_i,\boldsymbol{\beta})^2\right)\Big],\end{aligned}$$

where $\theta_i(\mathrm{q},\boldsymbol{\beta})$ is given by

$$\begin{aligned}&\arccos\left\{\frac{1}{2}\left[\mathrm{tr}\,(\mathrm{q}^T\mathrm{q}_i) - \frac{\sin(\lambda(\mathbf{x}_i,\boldsymbol{\beta}))}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})}\mathrm{tr}\,(\Lambda(\mathbf{x}_i,\boldsymbol{\beta})\mathrm{q}^T\mathrm{q}_i)\right.\right. \\ &+ \left.\left.\frac{(1-\cos(\lambda(\mathbf{x}_i,\boldsymbol{\beta})))}{\lambda(\mathbf{x}_i,\boldsymbol{\beta})^2}\mathrm{tr}\,(\Lambda(\mathbf{x}_i,\boldsymbol{\beta})^2\mathrm{q}^T\mathrm{q}_i)\right] - 1\right\}.\end{aligned}$$

### 3.11. Kendall's Planar Shape Space $\Sigma_2^k$

We review the definition and some basic facts about the geometric structure of the shape space $\Sigma_2^k$ formed by $k$ landmarks in $R^2$, $k>2$ (Kendall, 1984; Kendall et al., 1999; Dryden and Mardia, 1998; Huckemann et al., 2010; Su et al., 2012). Geometrical planar objects are studied by placing $k>2$ landmarks at specific locations of each object, usually on the boundary of the object. Then each object is described by a $k\times 2$ matrix $y\in R^{k\times 2}$, each row $y^m$ denoting the coordinates of a point in $R^2$ for $m=1,\ldots,k$. It is often convenient to identify points in $R^2$ with complex numbers, i.e. $y^m=(y^{m,1},y^{m,2})\equiv z^m = y^{m,1}+jy^{m,2}\in C$, where $j=\sqrt{-1}$. In this representation, a configuration $y$ with $k$ landmarks is an element $z\in C^k$. We remove the translations by restricting to those elements of $C^k$ whose average is zero, $\sum_{m=1}^k z^m=0$, and the scale variability by rescaling the matrix to have norm one, $\|z\|_2^2=\bar{z}^Tz=\sum_{m=1}^k z^m\overline{z^m}=1$, where the "overline" denotes complex conjugation. Thus, we obtain a set $\mathcal{D}^k=\{z=(z^1,\ldots,z^k)^T\in C^k \,|\, k^{-1}\sum_{m=1}^k z^m=0, \|z\|_2=1\}$ called the pre-shape space. Here, $\mathcal{D}^k$ is a unit sphere and we can utilize the geometry of a sphere to analyze points on it.

Thus, $\mathcal{D}^k$ has the canonical structure of a real Riemannian manifold of real dimension $(2k-3)$, with the metric induced by the standard inner product on $R^{k\times 2}$ which is the real part of the complex inner product on $C^k$. The tangent space of $\mathcal{D}^k$ at a point $z$ is $T_z\mathcal{D}^k = \{\mathbf{v} = (v^1,\dots,v^k)^T \in C^k \,|\, Re(\bar{z}^T v) = 0,\, k^{-1}\sum_{m=1}^k v^m = 0\}$ and the geodesic distance on $\mathcal{D}^k$ is the spherical distance $d_{\mathcal{D}^k}(z,z') = \arccos(Re(\bar{z'}^T z))$. The special unitary group $G = \mathrm{SU}(\mathcal{V}) \cong \mathrm{SU}(k-1) \subset \mathrm{SU}(k)$ acts transitively on $\mathcal{D}^k$, where $\mathcal{V}$ is the complex-orthogonal complement of $span_{C^k}\{(1,\dots,1)^T\}$ in $C^k$. $\mathcal{V}$ has complex dimension $k-1$. $\mathrm{SU}(k-1)$ is a real Lie group of dimension $(k-1)^2-1$. The isotropy subgroup of $z$ is $G_z \cong \mathrm{SU}(k-2)$. Thus, $\mathcal{D}^k$ is a Riemannian symmetric space.

To obtain the shape space, we remove the planar rotations of pre-shapes. For $z \in \mathcal{D}^k$, let $[z]$ be the set of all planar rotations of a configuration $z$ according to $[z] = \{z' = e^{j\theta}z | \theta \in S^1\}$. One defines an equivalence relation on $\mathcal{D}^k$ by setting all elements of the set $[z]$ as equivalent, i.e. $z \sim z'$ if there is an angle $\theta$ such that $z' = e^{j\theta}z$. The set of all such equivalence classes is the quotient space $\mathcal{D}^k/S^1$. This space is called *Kendall's planar shape space* and is denoted by $\Sigma_2^k$. Since $S^1$ acts freely on $\mathcal{D}^k$, i.e. the only element of $S^1$ whose action has fixed points is the unit element of $S^1$, then the quotient space $\Sigma_2^k$ is a $(2k-4)$-dimensional real Riemannian manifold. In fact, this space can be identified with a complex projective space $CP^{k-2}$. Since, $z \sim z'$ implies $Uz \sim Uz'$, for any $z, z' \in \mathcal{D}^k$ and any $U \in \mathrm{SU}(\mathcal{V})$, the group $G$ acts transitively on $\Sigma_e^k$ as well, and the isotropy subgroup is $G_{[z]} \cong \mathrm{SU}(k-2) \times S^1$. The natural Riemannian structure on $\Sigma_2^k$ (as $CP^{k-2}$) is given by the Fubini-Study metric, which is defined as follows.

The tangent space of $\Sigma_2^k$ at a point $\mathrm{q} = [z_\mathrm{q}]$, with $z_\mathrm{q} \in \mathcal{D}^k$, is

$$\begin{aligned} T_\mathrm{q}\Sigma_2^k &= \{\mathbf{v} = (v^1,\dots,v^k)^T \in C^k \,|\, Re\left(\overline{(e^{j\theta}z_\mathrm{q})}^T\mathbf{v}\right) = 0,\, \theta \in S^1,\, k^{-1}\sum_{i=1}^k v^i = 0\} \\ &= \{\mathbf{v} = (v^1,\dots,v^k)^T \in C^k \,|\, \overline{z_\mathrm{q}}^T\mathbf{v} = 0,\, k^{-1}\sum_{i=1}^k v^i = 0\}, \end{aligned}$$

and it is equipped with the complex inner product induced from $C^k$, that is, $< \mathbf{v}, \mathbf{w} >_\mathrm{q} := \overline{\mathbf{w}}^T\mathbf{v}$, for $\mathbf{v}, \mathbf{w} \in T_\mathrm{q}\Sigma_2^k$, which is well-defined.

A geodesic between two elements $\mathrm{q}_1, \mathrm{q}_2 \in \Sigma_2^k$, with $\mathrm{q}_l = [z_{\mathrm{q}_l}]$, $l = 1,2$, is given by a spherical geodesic on $\mathcal{D}^k$ between $z_{\mathrm{q}_1}$ and $z^*_{\mathrm{q}_2}$, where $z^*_{\mathrm{q}_2} = e^{j\theta^*}z_{\mathrm{q}_2}$ and $\theta^*$ is the optimal rotational alignment of $z_{\mathrm{q}_2}$ to $z_{\mathrm{q}_1}$ given by $\bar{z}_{\mathrm{q}_2}^T z_{\mathrm{q}_1} = e^{j\theta^*}|\bar{z}_{\mathrm{q}_2}^T z_{\mathrm{q}_1}|$. The geodesic distance on $\Sigma_2^k$ between $\mathrm{q}_1, \mathrm{q}_2$, $d_{\Sigma_2^k}(\mathrm{q}_1, \mathrm{q}_2)$, is the spherical distance $d_{\mathcal{D}_m^k}(z_{\mathrm{q}_1}, z_{\mathrm{q}_2}^*) = \arccos(\overline{z_{\mathrm{q}_2}^*}^T z_{\mathrm{q}_1}) = \arccos(|\overline{z_{\mathrm{q}_2}}^T z_{\mathrm{q}_1}|)$. The definitions of both the geodesics and geodesic distance are independent of the choice of representatives for the equivalence classes $\mathrm{q}_1$ and $\mathrm{q}_2$. For $\mathbf{v} \in T_{\mathrm{q}}\Sigma_2^k$, the Riemannian Exponential map is given by $\mathrm{Exp}_{\mathrm{q}}(\mathbf{v}) = \cos(\|\mathbf{v}\|)z_q + \sin(\|\mathbf{v}\|)\frac{\mathbf{v}}{\|\mathbf{v}\|}$. The exponential map is well-defined and it is a bijection on the set of $[(z_{\mathrm{q}}, \mathbf{v})]$ so that $\|\mathbf{v}\| \in [0, \frac{\pi}{2})$. The Riemannian Logarithmic map is given by $\mathrm{Log}_{\mathrm{q}}(\mathrm{q}') = \arccos(|\overline{z_{\mathrm{q}'}}^T z_{\mathrm{q}}|)\mathbf{v}/\|\mathbf{v}\| = \frac{r}{\sin(r)}\mathbf{v}$, where $\mathbf{v} = z_{\mathrm{q}'}^* - |\overline{z_{\mathrm{q}'}}^T z_{\mathrm{q}}|z_{\mathrm{q}}$, $r = d_{\Sigma_2^k}(\mathrm{q}, \mathrm{q}')$, and $z_{\mathrm{q}'}^*$ is the optimal alignment of $z_{\mathrm{q}'}$ to $z_{\mathrm{q}}$. It is easy to check that all the definitions above are independent of the choice of representatives for the corresponding equivalence classes.

Note that with respect to a chosen complex orthonormal basis $\{Z_1, \ldots, Z_{k-2}\}$ for $T_{\mathrm{p}}\Sigma_2^k$, the normal chart $\phi$ centered at p has the expression

$$\begin{aligned} \phi(\mathrm{q}) &= \boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{k-2})^T \in C^{k-2}, \\ &\cong \mathbf{t} = (t_1, \ldots, t_{2k-4})^T \in R^{2k-4} \end{aligned} \tag{18}$$

where $\zeta_\ell = t_{2\ell-1} + jt_{2\ell}$ and

$$\zeta_\ell = \frac{r}{\sin(r)} e^{j\theta}\, \overline{Z_\ell}^T z_{\mathrm{q}},\ r = d_{\Sigma_2^k}(\mathrm{q}, \mathrm{p}) = \arccos(|\overline{z_{\mathrm{q}}}^T z_{\mathrm{p}}|),\ e^{j\theta} = \frac{\overline{z_{\mathrm{q}}}^T z_{\mathrm{p}}}{|\overline{z_{\mathrm{q}}}^T z_{\mathrm{p}}|} \tag{19}$$

for $\ell = 1, \ldots, k-2$ (Bhattacharya and Bhattacharya, 2008).

We introduce the intrinsic regression model for $\Sigma_2^k$-valued responses. Suppose that we observe $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where $y_i \in \Sigma_2^k$ and $\mathbf{x}_i \in R^{d_{\mathbf{x}}}$, for all $i$. We define a function $f(\mathbf{x}, \boldsymbol{\beta})$ given by

$$f(\cdot, \cdot) : R^{d_{\mathbf{x}}} \times R^{d_\beta} \to R^{2k-4} \ \text{ with } \ f(\mathbf{0}, \cdot) = f(\cdot, \mathbf{0}) = \mathbf{0}. \tag{20}$$

An example of $f(\cdot, \cdot)$ is $f(\mathbf{x}_i, \boldsymbol{\beta}) = B\mathbf{x}_i$, where $B$ is a $(2k-4) \times d_{\mathbf{x}}$ matrix of regression coefficients and $\boldsymbol{\beta}$ includes all components of $B$. We fix a point $\mathrm{p} \in \Sigma_2^k$, as the base point, and $\{Z_1, \ldots, Z_{k-2}\}$ an orthonormal basis for $T_{\mathrm{p}}\Sigma_2^k/\sqrt{(\ell+1)(\ell+2)}$. Thus, $\{Z_1, \ldots, Z_{k-2}\}$ forms a complex orthonormal basis of $T_{\mathrm{p}}\Sigma_2^k$ when viewed as a complex linear space, and, equivalently, $\{Z_1, jZ_1, \ldots, Z_{k-2}, jZ_{k-2}\}$ forms a real orthonormal basis of $T_{\mathrm{p}}\Sigma_2^k$ when viewed as a real linear space. Letting $\mathrm{q} \in \Sigma_2^k$ be the 'center', we consider a single-center link function given by

$$\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \mathrm{Exp}_{\mathrm{q}}(\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})) \in \Sigma_2^k, \tag{21}$$

where $\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) \in T_{\mathrm{q}}\Sigma_2^k$. An example of $\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})$ is given by

$$\mathbf{u}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) = \left[\sum_{\ell=1}^{k-2} (f(\mathbf{x}_i, \boldsymbol{\beta})_{2\ell-1} + j f(\mathbf{x}_i, \boldsymbol{\beta})_{2\ell}) U_{z_{\mathrm{p}}, z_{\mathrm{q}}^*} Z_\ell\right] \in T_{\mathrm{q}}\Sigma_2^k, \tag{22}$$

where $\mathrm{p} = [z_{\mathrm{p}}]$, $\mathrm{q} = [z_{\mathrm{q}}]$, with $z_{\mathrm{p}}, z_{\mathrm{q}} \in \mathcal{D}^k$. Here, for $z_1, z_2 \in \mathcal{D}^k$, $U_{z_1,z_2} \in \mathrm{SU}(\mathcal{V}) \subset SU(k)$ denotes the unique special unitary map in the subspace generated by $z_1$ and $z_2$ that maps $z_1$ onto $z_2$. The map $U_{z_1,z_2}$ takes the form

$$\begin{aligned} U_{z_1,z_2}\mathbf{v} \quad &= \quad \mathbf{v} - (\overline{z_1}^T \mathbf{v}) z_1 - (\overline{\widetilde{z}_2}^T \mathbf{v}) \widetilde{z}_2 \\ &\quad + \left( (\overline{z_1}^T z_2)(\overline{z_1}^T \mathbf{v}) - \sqrt{1 - |\overline{z_1}^T z_2|^2}\, (\overline{\widetilde{z}_2}^T \mathbf{v}) \right) z_1 \\ &\quad + \left( \sqrt{1 - |\overline{z_1}^T z_2|^2}\, (\overline{z_1}^T \mathbf{v}) + \overline{(\overline{z_1}^T z_2)} (\overline{\widetilde{z}_2}^T \mathbf{v}) \right) \widetilde{z}_2, \end{aligned} \tag{23}$$

for $\mathbf{v} \in C^k$, where $\widetilde{z}_2 = \frac{z_2 - (\overline{z_1}^T z_2) z_1}{\sqrt{1 - |\overline{z_1}^T z_2|^2}}$. Thus, $U_{\mathrm{q}_1,\mathrm{q}_2}\mathbf{v} := U_{z_{\mathrm{q}_1}, z_{\mathrm{q}_2}^*}\mathbf{v} \in \mathcal{V}$, $\mathbf{v} \in \mathcal{V}$, and $U_{\mathrm{q}_1,\mathrm{q}_2}\mathrm{q} := [U_{z_{\mathrm{q}_1}, z_{\mathrm{q}_2}^*} z_{\mathrm{q}}] \in \Sigma_2^k$, $\mathrm{q} \in \Sigma_2^k$, are well defined, independently of the choice of representatives $z_{\mathrm{q}_1}, z_{\mathrm{q}_2}$, and $z_{\mathrm{q}}$ for $\mathrm{q}_1, \mathrm{q}_2$, and q, respectively. The *rotated residual* is given by

$$\mathcal{E}(y_i, \mathbf{x}_i; \mathrm{q}, \boldsymbol{\beta}) \quad = \quad U^{-1}_{\mathrm{p}_0, \boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})} (\mathrm{Log}_{\boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})} y_i) = \mathrm{Log}_{\mathrm{p}_0} (\overline{U_{\mathrm{p}_0, \boldsymbol{\mu}(\mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta})}}^T y_i). \tag{24}$$

We consider the *intrinsic* model

$$E(\mathcal{E}(y_i, \mathbf{x}_i, \mathrm{q}, \boldsymbol{\beta}) \,|\, \mathbf{x}_i) = 0, \quad i = 1, \ldots, n. \tag{25}$$

## 4. Annealing evolutionary stochastic approximation Monte Carlo

We now develop an annealing evolutionary stochastic approximation Monte Carlo algorithm for computing $\hat{\boldsymbol{\theta}}_I = (\hat{\mathrm{q}}_I, \hat{\boldsymbol{\beta}}_I)$ and $\hat{\boldsymbol{\theta}}_E = (\hat{\mathrm{q}}_E, \hat{\boldsymbol{\beta}}_E)$. Quite recently, the stochastic approximation Monte Carlo algorithm (Liang et al., 2010) has been proposed in the literature as a general simulation technique, which possesses a nice feature in that the moves are self-adjustable and thus not likely to get trapped by local energy minima. The annealing evolutionary SAMC algorithm (Liang et al., 2010) represents a further improvement of stochastic approximation Monte Carlo for optimization problems by incorporating some features of simulated annealing and the genetic algorithm into its search process.

Like the genetic algorithm, annealing evolutionary stochastic approximation Monte Carlo works on a population of samples. Let $\boldsymbol{\theta}^l = (\boldsymbol{\theta}_{(1)}, \ldots, \boldsymbol{\theta}_{(l)})$ denote the population, where $l$ is the population size, and $\boldsymbol{\theta}_{(k)} = (\theta_{k1}, \ldots, \theta_{kp_\theta})$ is a $p_\theta$-dimensional vector called an individual or chromosome in terms of genetic algorithms. Thus, the minimum of the objective function $\mathcal{Q}_n(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, can be obtained by minimizing the function $U(\boldsymbol{\theta}^l) = \sum_{k=1}^{l} \mathcal{Q}_n(\boldsymbol{\theta}_{(k)})$. An unnormalized Boltzmann density can be defined for the population as follows,

$$\psi(\boldsymbol{\theta}^l) = \exp\left\{-U(\boldsymbol{\theta}^l)/\tau\right\}, \quad \boldsymbol{\theta}^l \in \Theta^l, \tag{26}$$

where $\tau = 1$ is called the temperature, and $\Theta^l = \Theta \times \cdots \times \Theta$ is a product sample space. The sample space can be partitioned according to the function $U(\boldsymbol{\theta}^l)$ into $b$ subregions: $\mathbb{E}_1 = \{\boldsymbol{\theta}^l : U(\boldsymbol{\theta}^l) \leq \delta_1\}$, $\mathbb{E}_2 = \{\boldsymbol{\theta}^l : \delta_1 < U(\boldsymbol{\theta}^l) \leq \delta_2\}$, $\cdots$, $\mathbb{E}_{b-1} = \{\boldsymbol{\theta}^l : \delta_{b-2} < U(\boldsymbol{\theta}^l) \leq \delta_{b-1}\}$, and $\mathbb{E}_b = \{\boldsymbol{\theta}^l : U(\boldsymbol{\theta}^l) > \delta_{b-1}\}$, where $\delta_1 < \delta_2 < \ldots < \delta_{b-1}$ are $b-1$ known real numbers. We note that here the sample space is not necessarily partitioned according to the function $U(\boldsymbol{\theta}^l)$, for example, the function $\lambda(\boldsymbol{\theta}^l) = \min\{\mathcal{Q}_n(\boldsymbol{\theta}_{(1)}), \ldots, \mathcal{Q}_n(\boldsymbol{\theta}_{(l)})\}$ also works.

Let $\varpi(\delta)$ denote the index of the subregion that a sample with energy $U(\boldsymbol{\theta}^l)$ belongs to. For example, if $\boldsymbol{\theta}^l \in E_j$, then $\varpi(U(\boldsymbol{\theta}^l)) = j$. Let $\mathcal{B}^{(t)}$ denote the sample space at iteration $t$. The algorithm initiates its search in the entire sample space $\mathcal{B}_0 = \bigcup_{j=1}^{b} E_j$, and then iteratively searches in the set

$$\mathcal{B}_t = \bigcup_{j=1}^{\varpi(U_{\min}^{(t)}+\aleph)} \mathbb{E}_j, \quad t = 1, 2, \ldots, \tag{27}$$

where $U_{\min}^{(t)}$ is the best function value obtained until iteration $t$, and $\aleph > 0$ is a user specified parameter which determines the broadness of the sample space at each iteration. Note that in this method, the sample space shrinks iteration by iteration. To ensure the convergence of the algorithm to the set of global minima, the moves at each itertaion are required to admit the following distribution as the invariant distribution,

$$f_{w^{(t)}}(\boldsymbol{\theta}^l) \propto \sum_{j=1}^{\varpi(U_{\min}^{(t)}+\aleph)} \frac{\psi(\boldsymbol{\theta}^l)}{e^{w_j^{(t)}}} I(\boldsymbol{\theta}^l \in \mathbb{E}_j), \tag{28}$$

where $w_j^{(t)}$ are the working parameters which will be updated from itertaion to iteration as described in the algorithm below.

The annealing evolutionary stochastic approximation Monte Carlo includes five types of moves, the MH-Gibbs mutation, $K$-point mutation, $K$-point crossover, snooker crossover, and linear crossover operators. See Liang et al. (2010) for the details of the moves. Let $\rho_1, \ldots, \rho_5$, $0 < \rho_k < 1$ and $\sum_{k=1}^5 \rho_k = 1$, denote the respective working probabilities of the five types of moves. The algorithm can be summarized as follows.

*The algorithm:*

(a) (Initialization) Partition the sample space $\mathcal{B}^l$ into $b$ disjoint subregions $\mathbb{E}_1, \ldots, \mathbb{E}_b$; choose the threshold value $\aleph$ and the working probabilities $\rho_1, \ldots, \rho_5$; initialize a population $\boldsymbol{\theta}^{l(0)}$ at random; and set $w^{(0)} = (w_1^{(0)}, \ldots, w_b^{(0)}) = (0, 0, \ldots, 0)$, $\mathcal{B}_0^l = \bigcup_{j=1}^b \mathbb{E}_j$, $U_{\min}^{(0)} = U(\boldsymbol{\theta}^{l(0)})$ and $t = 0$. Let $\mathcal{W}$ be a compact set in $R^b$.

(b) (Sampling) Update the current population $\boldsymbol{\theta}^{l(t)}$ using the MH-Gibbs mutation, $K$-point mutation, $K$-point crossover, snooker crossover, and linear crossover operators according to the respective working probabilities.

(c) (Working weight updating) Update the working weight $w^{(t)}$ by setting

$$w_j^* = w_j^{(t)} + \gamma_{t+1} H_j(w^{(t)}, \boldsymbol{\theta}^{l(t+1)}), \quad j = 1, \ldots, \varpi(U_{\min}^{(t)} + \aleph),$$

where $H_j(w^{(t)}, \boldsymbol{\theta}^{l(t+1)}) = I(\boldsymbol{\theta}^{l(t+1)} \in \mathbb{E}_j)$ for the crossover operators, $H_j(w^{(t)}, \boldsymbol{\theta}^{l(t+1)}) = \sum_{k=1}^l I(\boldsymbol{\theta}^{l(t+1,k)} \in \mathbb{E}_j)/l$ for the mutation operators, and $\gamma_{t+1}$ is called the gain factor. If $w^* \in \mathcal{W}$, set $w^{(t+1)} = w^*$; otherwise, set $w^{(t+1)} = w^* + c^*$, where $c^* = (c^*, \ldots, c^*)$ and $c^*$ is chosen such that $w^* + c^* \in \mathcal{W}$.

(d) (Termination Checking) Check the termination condition, e.g., whether a fixed number of iterations has been reached. Otherwise, set $t \to t + 1$ and go to step (b).

In this article, we follow Liang et al. (2010) to set $\rho_1 = \rho_2 = 0.05$, $\rho_3 = \rho_4 = \rho_5 = 0.3$, and the gain factor sequence

$$\gamma_t = \frac{t_0}{\max(t_0, t)}, \quad t = 0, 1, 2, \ldots, \tag{29}$$

with $t_0 = 5000$. In general, a large value of $t_0$ will allow the sampler to reach all the subregions very quickly even for a large system. As shown in Liang et al. (2010), it can converge weakly toward a neighboring set of global minima of $U(\boldsymbol{\theta}^l)$ in the space of energy. More precisely, the sample $\boldsymbol{\theta}^{l(t)}$ converges in distribution to a random population with the density function

$$f_w(\boldsymbol{\theta}^l) \propto \sum_{j=1}^{\varpi(U_{\min}+\aleph)} \frac{\psi(\boldsymbol{\theta}^l)}{\int_{\mathbb{E}_j} \psi(\boldsymbol{\theta}^l) d\boldsymbol{\theta}^l} I(x \in \mathbb{E}_j), \tag{30}$$

where $U_{\min}$ is the global minimum value of $U(\boldsymbol{\theta})$,

Regarding the setting of other parameters, we have the following suggestions. In the algorithm, the moves are reduced to the Metropolis-Hastings moves (Metropolis et al., 1953; Hastings, 1970) within the same subregions. Hence, the sample space should be partitioned such that the MH moves within the same subregion have a reasonable acceptance rate. In this article, we set $\delta_{j+1} - \delta_j \equiv 0.2$ for $j = 1, \ldots, b - 1$.

The crossover operator has been modified to serve as a proposal for the moves, and it is no longer as critical as to the genetic algorithm. Hence, the population size $l$ is usually set to a moderate number, ranging from 10 to 100. Since $\aleph$ determines the size of the neighboring set toward which the method converges, $\aleph$ should be chosen carefully for efficiency of the algorithm. If $\aleph$ is too small, it may take a long time for the algorithm to locate the global minima. In this case, the sample space may contain a lot of separated regions, and most of the proposed transitions will be rejected if the proposal distribution is not spread out enough. If $\aleph$ is too large, it may also take a long time for the algorithm to locate the global energy minimum due to the broadness of the sample space. In practice, the values of $l$ and $\aleph$ can be determined through a trial and error process based on the diagnosis for the convergence of the algorithm. If it fails to converge, the parameters should be tuned to larger values. The convergence of the method can be diagnosed by examining the difference of the patterns of the working weights obtained in multiple runs. In this article, we set $l = 50$ and $\aleph = 50$.

## References

Batchelor, P. G., M. Moakher, D. Atkinson, F. Calamante, and A. Connelly (2005). A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine 53*(1), 221–225.

Bhattacharya, A. and R. Bhattacharya (2008). Statistics on Riemannian manifolds: Asymptotic distribution and curvature. *Proceedings of the American Mathematical Society 136*(8), pp. 2959–2967.

Boothby, W. M. (1986). *An Introduction to Differentiable Manifolds and Riemannian Geometry* (Second ed.), Volume 120 of *Pure and Applied Mathematics.* Orlando, FL: Academic Press Inc.

Dryden, I. L. and K. V. Mardia (1998). *Statistical Shape Analysis.* Chichester: John Wiley & Sons Ltd.

Fletcher, P. T., C. Lu, S. Pizer, and S. Joshi (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *Medical Imaging, IEEE Transactions on 23*(8), 995 –1005.

Gallier, J. and D. Xu (2002). Computing exponentials of skew-symmetric matrices and logarithms of orthogonal matrices. *Int. J. Robot. Autom. 17*(4), 1–11.

Grenander, U., M. Miller, and A. Srivastava (1998). Hilbert-Schmidt lower bounds for estimators on matrix Lie groups for atr. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 20*(8), 790–802.

Helgason, S. (1978). *Differential Geometry, Lie Groups, and Symmetric Spaces*, Volume 80 of *Pure and Applied Mathematics.* New York, NY: Academic Press Inc.

Huckemann, S., T. Hotz, and A. Munk (2010). Intrinsic manova for Riemannian manifolds with an application to Kendall's space of planar shapes. *IEEE Trans. Patt. Anal. Mach. Intell. 32*, 593–603.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc. 16*(2), 81–121.

Kendall, D. G., D. Barden, T. K. Carne, and H. Le (1999). *Shape and Shape Theory.* Chichester: John Wiley & Sons Ltd.

Lang, S. (1999). *Fundamentals of Differential Geometry*, Volume 191 of *Graduate Texts in Mathematics.* New York: Springer-Verlag.

Liang, F., C. Liu, and R. J. Carroll (2010). *Advanced Markov Chain Monte Carlo: Learning from Past Samples.* New York: Wiley.

McCullagh, P. and J. A.Nelder (1989). *Generalized Linear Models (2nd edition).* London: Chapman and Hall.

Moakher, M. (2002). Means and averaging in the group of rotations. *SIAM J. Matrix Anal. Appl. 24*, 1–16.

Newey, W. K. (1993). Efficient estimation of models with conditional moment restrictions. In *Econometrics*, Volume 11 of *Handbook of Statist.*, pp. 419–454. Amsterdam: North-Holland.

Osborne, D., V. Patrangenaru, L. Ellingson, D. Groisser, and A. Schwartzman (2013). Nonparametric two-sample tests on homogeneous Riemannian manifolds, cholesky decompositions and diffusion tensor image analysis. *Journal of Multivariate Analysis 119*, 163 – 175.

Pennec, X. (2006). Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *J. Math. Imaging Vision 25*(1), 127–154.

Schwartzman, A. (2006). *Random ellipsoids and false discovery rates: Statistics for diffusion tensor imaging data.* Ph. D. thesis. Stanford University.

Shi, X., H. Zhu, J. G. Ibrahim, F. Liang, J. Liberman, and M. Styner (2012). Intrinsic regression models for median representation of subcortical structures. *Journal of American Statistical Association 107*, 12–23.

Su, J., I. Dryden, E. Klassen, H. Le, and A. Srivastava (2012). Fitting smoothing splines to time-indexed, noisy points on nonlinear manifolds. *Image and Vision Computing 30*(6–7), 428 – 442.

Terras, A. (1988). *Harmonic Analysis on Symmetric Spaces and Applications. II.* Berlin: Springer-Verlag.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, Volume 3. Cambridge: Cambridge University Press.

Yuan, Y., H. Zhu, W. Lin, and J. S. Marron (2012). Local polynomial regression for symmetric positive definite matrices. *Journal of Royal Statistical Society B 74*, 697–719.

Zhu, H., Y. Chen, J. G. Ibrahim, Y. Li, C. Hall, and W. Lin (2009). Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *J. Amer. Statist. Assoc. 104* (487), 1203–1212.