Supplementary Material for: Unsupervised Riemannian Metric Learning for Histograms Using Aitchison Transformations*

Tam Le
Graduate School of Informatics,
Kyoto University, Japan.
tam.le@iip.ist.i.kyoto-u.ac.jp

Marco Cuturi
Graduate School of Informatics,
Kyoto University, Japan.
mcuturi@i.kyoto-u.ac.jp

1 Proof for Proposition 1

The j^{th} component of the tangent vector of the positive sphere, mapped by a push-forward map F_* on a tangent vector $\mathbf{v} \in T_{\mathbf{x}}\mathbb{P}_n$

$$[F_*\mathbf{v}]_j = \frac{\mathrm{d}}{\mathrm{d}t}\sqrt{\frac{(\mathbf{x}_j + t\mathbf{v}_j)^{\boldsymbol{\alpha}_j}\boldsymbol{\lambda}_j}{\sum_{i=1}^{n+1}(\mathbf{x}_i + t\mathbf{v}_i)^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i}}\Bigg|_{t=0},$$

For simplify, let denote:

$$f_i(t) = (\mathbf{x}_i + t\mathbf{v}_i)^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i,$$
$$h_i(t) = \frac{\mathrm{d}f_i(t)}{\mathrm{d}t} = (\mathbf{x}_i + t\mathbf{v}_i)^{\boldsymbol{\alpha}_i - 1}\boldsymbol{\lambda}_i\boldsymbol{\alpha}_i\mathbf{v}_i.$$

So, we have:

$$[F_*\mathbf{v}]_j = \frac{h_j(t)\sum_{i=1}^{n+1} f_i(t) - f_j(t)\sum_{i=1}^{n+1} h_i(t)}{2\sqrt{\frac{f_j(t)}{\sum_{i=1}^{n+1} f_i(t)}}\left(\sum_{i=1}^{n+1} f_i(t)\right)^2}\Bigg|_{t=0}$$

Since at $t = 0$, $f_i(0) = \mathbf{x}_i^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i$ and $h_i(0) = \mathbf{x}_i^{\boldsymbol{\alpha}_i - 1}\boldsymbol{\lambda}_i\boldsymbol{\alpha}_i\mathbf{v}_i$,

$$[F_*\mathbf{v}]_j = \frac{1}{2}\mathbf{x}_j^{\frac{\boldsymbol{\alpha}_j}{2} - 1}\boldsymbol{\alpha}_j\mathbf{v}_j\boldsymbol{\lambda}_j^{\frac{1}{2}}\left(\sum_{\ell=1}^{n+1}\mathbf{x}_\ell^{\boldsymbol{\alpha}_\ell}\boldsymbol{\lambda}_\ell\right)^{-\frac{1}{2}} - \frac{1}{2}\mathbf{x}_j^{\frac{\boldsymbol{\alpha}_j}{2}}\boldsymbol{\lambda}_j^{\frac{1}{2}}\frac{\sum_{\ell=1}^{n+1}\mathbf{x}_\ell^{\boldsymbol{\alpha}_\ell - 1}\boldsymbol{\alpha}_\ell\mathbf{v}_\ell\boldsymbol{\lambda}_\ell}{\left(\sum_{\ell=1}^{n+1}\mathbf{x}_\ell^{\boldsymbol{\alpha}_\ell}\boldsymbol{\lambda}_\ell\right)^{\frac{3}{2}}}$$

Let apply $\mathbf{v} = \partial_i$, $1 \le i \le n$, the basis of the tangent space of the simplex $T_{\mathbf{x}}\mathbb{P}_n$

$$[F_*\partial_i]_j = \frac{1}{2}\mathbf{x}_j^{\frac{\boldsymbol{\alpha}_j}{2} - 1}\boldsymbol{\alpha}_j\boldsymbol{\lambda}_j^{\frac{1}{2}}\left(\sum_{\ell=1}^{n+1}\mathbf{x}_\ell^{\boldsymbol{\alpha}_\ell}\boldsymbol{\lambda}_\ell\right)^{-\frac{1}{2}}(\delta_{j,i} - \delta_{j,n+1}) - \frac{1}{2}\mathbf{x}_j^{\frac{\boldsymbol{\alpha}_j}{2}}\boldsymbol{\lambda}_j^{\frac{1}{2}}\frac{\mathbf{x}_i^{\boldsymbol{\alpha}_i - 1}\boldsymbol{\alpha}_i\boldsymbol{\lambda}_i - \mathbf{x}_{n+1}^{\boldsymbol{\alpha}_{n+1} - 1}\boldsymbol{\alpha}_{n+1}\boldsymbol{\lambda}_{n+1}}{\left(\sum_{\ell=1}^{n+1}\mathbf{x}_\ell^{\boldsymbol{\alpha}_\ell}\boldsymbol{\lambda}_\ell\right)^{\frac{3}{2}}},$$

**Proceedings of the* 32^{nd} *International Conference on Machine Learning*, Lille, France, 2015. JMLR: W&CP volume 37.

where $\delta_{j,i} = 1$ if $j = i$ and $\delta_{j,i} = 0$, otherwise.

Hence, we have $T = U(I - \boldsymbol{\beta}\boldsymbol{\eta}^T)D$.

Moreover, the metric on the positive sphere $\mathbb{S}_n^+$ is Euclidean. So, we have $J(\partial_i, \partial_j) = \langle F_*\partial_i, F_*\partial_j \rangle$. Consequently, we have the Gram matrix:

$$\mathcal{G} = TT^T = U(I - \boldsymbol{\beta}\boldsymbol{\eta}^T)D^2(I - \boldsymbol{\beta}\boldsymbol{\eta}^T)^T U^T.$$

2 Proof for Proposition 2

Let consider matrix $\boldsymbol{\beta}\boldsymbol{\eta}^T \in \mathbb{R}^{(n+1)\times(n+1)}$, and vector $\mathbf{v}$ such that $\boldsymbol{\eta}^T\mathbf{v} = 0$, so $\mathbf{v}$ is a eigenvector of $\boldsymbol{\beta}\boldsymbol{\eta}^T$ with eigenvalue 0. There are n independent vectors $\{\mathbf{v_i}\}_{1\le i\le n}$ such that $\boldsymbol{\eta}^T\mathbf{v_i} = 0$. Moreover, trace $\left(\boldsymbol{\beta}\boldsymbol{\eta}^T\right) = \sum_{i=1}^{n+1} \boldsymbol{\beta}_i\boldsymbol{\eta}_i = 1$, or sum of the eigenvalues of $\boldsymbol{\beta}\boldsymbol{\eta}^T$ is 1. So, the last of $(n+1)$ eigenvalues is 1. On the other hand, $\left(\boldsymbol{\beta}\boldsymbol{\eta}^T\right)\boldsymbol{\beta} = \boldsymbol{\beta}\left(\boldsymbol{\eta}^T\boldsymbol{\beta}\right) = \boldsymbol{\beta}$, or $\boldsymbol{\beta}$ is a eigenvector of $\boldsymbol{\beta}\boldsymbol{\eta}^T$ with eigenvalue 1. In summary, we have $\{(\mathbf{v_i}, 0)_{1\le i\le n}, (\boldsymbol{\beta}, 1)\}$ are eigenvectors and corresponding eigenvalues of $\boldsymbol{\beta}\boldsymbol{\eta}^T$. Let V be a matrix in $\mathbb{R}^{(n+1)\times(n+1)}$ whose columns are $\{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n}, \boldsymbol{\beta}\}$. So, we may express V as follow:

$$V = \begin{bmatrix} -\frac{\mathbf{x}_2\boldsymbol{\alpha}_1}{\boldsymbol{\alpha}_2\mathbf{x}_1} & \cdots & -\frac{\mathbf{x}_{n+1}\boldsymbol{\alpha}_1}{\boldsymbol{\alpha}_{n+1}\mathbf{x}_1} & \mathbf{x}_1^{\boldsymbol{\alpha}_1-1}\boldsymbol{\alpha}_1\boldsymbol{\lambda}_1 \\ 1 & \cdots & 0 & \mathbf{x}_1^{\boldsymbol{\alpha}_2-1}\boldsymbol{\alpha}_2\boldsymbol{\lambda}_2 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & \mathbf{x}_{n+1}^{\boldsymbol{\alpha}_{n+1}-1}\boldsymbol{\alpha}_{n+1}\boldsymbol{\lambda}_{n+1} \end{bmatrix}.$$

Let Λ is a diagonal matrix in $\mathbb{R}^{(n+1)\times(n+1)}$ where $\Lambda_{ii} = 0$, for all $1 \le i \le n$, and $\Lambda_{(n+1)(n+1)} = 1$. We have $\boldsymbol{\beta}\boldsymbol{\eta}^T = V\Lambda V^{-1}$. Consequently, we have $I - \boldsymbol{\beta}\boldsymbol{\eta}^T = V(I-\Lambda)V^{-1}$. Since $I - \Lambda = \text{diag}(1, 1, \cdots, 1, 0)$ and $I - \Lambda = (I-\Lambda)^2$, we may express $I - \boldsymbol{\beta}\boldsymbol{\eta}^T = \widetilde{V}\widetilde{V^{-1}}$, where $\widetilde{V} \in \mathbb{R}^{(n+1)\times n}$ is the matrix V whose last column is removed, and $\widetilde{V^{-1}} \in \mathbb{R}^{n\times(n+1)}$ is the matrix V^{-1} whose last row is removed.

Thus, we can express the Gram matrix $\mathcal{G}$ as follow:

$$\begin{aligned}\mathcal{G} &= U\widetilde{V}\widetilde{V^{-1}}D^2(\widetilde{V}\widetilde{V^{-1}})^T U^T \\ &= (U\widetilde{V})(\widetilde{V^{-1}}D^2\widetilde{V^{-1}}^T)(U\widetilde{V})^T\end{aligned}$$

We also note that $U\widetilde{V}$ and $\widetilde{V^{-1}}D^2\widetilde{V^{-1}}^T$ are matrices in $\mathbb{R}^{n\times n}$. So, we have $\det\mathcal{G} = \det^2(U\widetilde{V})\det(\widetilde{V^{-1}}D^2\widetilde{V^{-1}}^T)$.

Compute det$(U\widetilde{V})$: Since, we have

$$\begin{aligned} U\widetilde{V} &= \begin{pmatrix} -\frac{\mathbf{x}_2\boldsymbol{\alpha}_1}{\boldsymbol{\alpha}_2\mathbf{x}_1} & \cdots & -\frac{\mathbf{x}_n\boldsymbol{\alpha}_1}{\boldsymbol{\alpha}_n\mathbf{x}_1} & -\frac{\mathbf{x}_{n+1}\boldsymbol{\alpha}_1}{\boldsymbol{\alpha}_{n+1}\mathbf{x}_1} - 1 \\ 1 & \cdots & 0 & -1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & -1 \end{pmatrix} \\ &= \frac{\boldsymbol{\alpha}_1}{\mathbf{x}_1}\begin{pmatrix} -\sum_{i=1}^{n+1}\frac{\mathbf{x}_i}{\boldsymbol{\alpha}_i} & -\frac{\mathbf{x}_3}{\boldsymbol{\alpha}_3} & \cdots & -\frac{\mathbf{x}_n}{\boldsymbol{\alpha}_n} & -\frac{\mathbf{x}_{n+1}}{\boldsymbol{\alpha}_{n+1}} - \frac{\mathbf{x}_1}{\boldsymbol{\alpha}_1} \\ 0 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix}. \end{aligned}$$

Therefore, $\det(U\widetilde{V}) = (-1)^n\frac{\boldsymbol{\alpha}_1}{\mathbf{x}_1}\sum_{i=1}^{n+1}\frac{\mathbf{x}_i}{\boldsymbol{\alpha}_i}$.

Compute det$(\widetilde{V^{-1}}D^2\widetilde{V^{-1}}^T)$**:** Let consider a $(n+1)\times(n+1)$ matrix

$$W = \begin{pmatrix} -\frac{\mathbf{r}_2}{\mathbf{r}_1} & \cdots & -\frac{\mathbf{r}_{n+1}}{\mathbf{r}_1} & \mathbf{c}_1 \\ 1 & \cdots & 0 & \mathbf{c}_2 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & \mathbf{c}_{n+1} \end{pmatrix}$$

We have its inverse:

$$W^{-1} = \frac{1}{\langle \mathbf{r}, \mathbf{c}\rangle} \begin{pmatrix} -\mathbf{r}_1\mathbf{c}_2 & \langle \mathbf{r}, \mathbf{c}\rangle - \mathbf{r}_2\mathbf{c}_2 & -\mathbf{r}_3\mathbf{c}_2 & \cdots & -\mathbf{r}_{n+1}\mathbf{c}_2 \\ -\mathbf{r}_1\mathbf{c}_3 & -\mathbf{r}_2\mathbf{c}_3 & \langle \mathbf{r}, \mathbf{c}\rangle - \mathbf{r}_2\mathbf{c}_2 & \cdots & -\mathbf{r}_{n+1}\mathbf{c}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\mathbf{r}_1\mathbf{c}_{n+1} & -\mathbf{r}_2\mathbf{c}_{n+1} & -\mathbf{r}_2\mathbf{c}_{n+1} & \cdots & \langle \mathbf{r}, \mathbf{c}\rangle - \mathbf{r}_{n+1}\mathbf{c}_{n+1} \\ \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \cdots & \mathbf{r}_{n+1} \end{pmatrix},$$

where vector $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_{n+1})$ and vector $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_{n+1})$ are in $\mathbb{R}^{n+1}$.

Now, we apply for the matrix V where $\mathbf{r}_i = \frac{\mathbf{x}_i}{\boldsymbol{\alpha}_i}$ and $\mathbf{c}_i = \mathbf{x}_i^{\boldsymbol{\alpha}_i - 1}\boldsymbol{\alpha}_i\boldsymbol{\lambda}_i$, for all $1 \le i \le (n+1)$, and remove the last row to form $\widetilde{V^{-1}}$. For simplicity, we denote a diagonal matrix $P \in \mathbb{R}^{n\times n}$ where $P_{ii} = \mathbf{x}_{i+1}^{\boldsymbol{\alpha}_{i+1}-1}\boldsymbol{\alpha}_{i+1}\boldsymbol{\lambda}_{i+1}$, for all $1 \le i \le n$ and matrix $Q \in \mathbb{R}^{n\times(n+1)}$ as follow:

$$Q = \begin{pmatrix} -\frac{\mathbf{x}_1}{\boldsymbol{\alpha}_1} & \sum\limits_{\substack{1\le i\le n+1\\ i\neq 2}} \frac{\mathbf{x}_i^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i}{\mathbf{x}_2^{\boldsymbol{\alpha}_2-1}\boldsymbol{\alpha}_2\boldsymbol{\lambda}_2} & -\frac{\mathbf{x}_3}{\boldsymbol{\alpha}_3} & \cdots & -\frac{\mathbf{x}_{n+1}}{\boldsymbol{\alpha}_{n+1}} \\ -\frac{\mathbf{x}_1}{\boldsymbol{\alpha}_1} & -\frac{\mathbf{x}_2}{\boldsymbol{\alpha}_2} & \sum\limits_{\substack{1\le i\le n+1\\ i\neq 3}} \frac{\mathbf{x}_i^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i}{\mathbf{x}_3^{\boldsymbol{\alpha}_3-1}\boldsymbol{\alpha}_3\boldsymbol{\lambda}_3} & \cdots & -\frac{\mathbf{x}_{n+1}}{\boldsymbol{\alpha}_{n+1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\mathbf{x}_1}{\boldsymbol{\alpha}_1} & -\frac{\mathbf{x}_2}{\boldsymbol{\alpha}_2} & -\frac{\mathbf{x}_3}{\boldsymbol{\alpha}_3} & \cdots & \sum\limits_{i=1}^{n} \frac{\mathbf{x}_i^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i}{\mathbf{x}_{n+1}^{\boldsymbol{\alpha}_{n+1}-1}\boldsymbol{\alpha}_{n+1}\boldsymbol{\lambda}_{n+1}} \end{pmatrix}.$$

So, we have

$$\widetilde{V^{-1}} = \frac{1}{\sum\limits_{i=1}^{n+1}\mathbf{x}_i^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i} PQ.$$

Then, we can compute

$$\widetilde{V^{-1}}D = \frac{1}{2\left(\sum\limits_{i=1}^{n+1}\mathbf{x}_i^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i\right)^{\frac{3}{2}}} P \begin{pmatrix} -\sqrt{\mathbf{x}_1^{\boldsymbol{\alpha}_1}\boldsymbol{\lambda}_1} & \sum\limits_{\substack{1\le i\le n+1\\ i\neq 2}} \frac{\mathbf{x}_i^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i}{\sqrt{\mathbf{x}_2^{\boldsymbol{\alpha}_2}\boldsymbol{\lambda}_2}} & -\sqrt{\mathbf{x}_3^{\boldsymbol{\alpha}_3}\boldsymbol{\lambda}_3} & \cdots & -\sqrt{\mathbf{x}_{n+1}^{\boldsymbol{\alpha}_{n+1}}\boldsymbol{\lambda}_{n+1}} \\ -\sqrt{\mathbf{x}_1^{\boldsymbol{\alpha}_1}\boldsymbol{\lambda}_1} & \sqrt{\mathbf{x}_2^{\boldsymbol{\alpha}_2}\boldsymbol{\lambda}_2} & \sum\limits_{\substack{1\le i\le n+1\\ i\neq 3}} \frac{\mathbf{x}_i^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i}{\mathbf{x}_3^{\boldsymbol{\alpha}_3-1}\boldsymbol{\alpha}_3\boldsymbol{\lambda}_3} & \cdots & -\sqrt{\mathbf{x}_{n+1}^{\boldsymbol{\alpha}_{n+1}}\boldsymbol{\lambda}_{n+1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\sqrt{\mathbf{x}_1^{\boldsymbol{\alpha}_1}\boldsymbol{\lambda}_1} & \sqrt{\mathbf{x}_2^{\boldsymbol{\alpha}_2}\boldsymbol{\lambda}_2} & -\sqrt{\mathbf{x}_3^{\boldsymbol{\alpha}_3}\boldsymbol{\lambda}_3} & \cdots & \sum\limits_{i=1}^{n} \frac{\mathbf{x}_i^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i}{\sqrt{\mathbf{x}_{n+1}^{\boldsymbol{\alpha}_{n+1}}\boldsymbol{\lambda}_{n+1}}} \end{pmatrix}.$$

Consequently, we have:

$$\widetilde{V^{-1}}D^2\widetilde{V^{-1}}^T = \frac{1}{4\left(\sum\limits_{i=1}^{n+1}\mathbf{x}_i^{\boldsymbol{\alpha}_i}\boldsymbol{\lambda}_i\right)^2} P(\widehat{Q} - \mathbf{1}_{n\times n})P,$$

where $\widehat{Q}$ is a diagonal matrix in $\mathbb{R}^{n\times n}$, $\widehat{Q}_{ii} = \sum\limits_{j=1}^{n+1} \frac{\mathbf{x}_j^{\boldsymbol{\alpha}_j}\boldsymbol{\lambda}_j}{\mathbf{x}_{i+1}^{\boldsymbol{\alpha}_{i+1}}\boldsymbol{\lambda}_{i+1}}$ and $\mathbf{1}_{n\times n}$ is a matrix of 1 in $\mathbb{R}^{n\times n}$.

Moreover, $\det(\widehat{Q} - \mathbf{1}_{n\times n}) = \prod\limits_{i=1}^{n} Q_{ii} - \sum\limits_{i=1}^{n}\prod\limits_{j\neq i} Q_{jj}$, following Lemma 2 of Lebanon (2005).

Hence, we have:

$$\det(\widehat{Q} - \mathbf{1}_{n\times n}) = \frac{\left(\sum_{i=1}^{n+1} \mathbf{x}_i^{\boldsymbol{\alpha}_i} \boldsymbol{\lambda}_i\right)^{n-1}}{\prod_{j=1}^{n+1} \mathbf{x}_j^{\boldsymbol{\alpha}_j} \boldsymbol{\lambda}_j} \left(\mathbf{x}_1^{\boldsymbol{\alpha}_1} \boldsymbol{\lambda}_1\right)^2 .$$

Consequently, we have

$$\det\left(\widetilde{V^{-1}} D^2 \widetilde{V^{-1}}^T\right) = \frac{1}{4^n \left(\sum_{i=1}^{n+1} \mathbf{x}_i^{\boldsymbol{\alpha}_i} \boldsymbol{\lambda}_i\right)^{2n}} \det^2(P) \det(\widehat{Q} - \mathbf{1}_{n\times n}).$$

Since $\det(P) = \prod_{j=2}^{n+1} \mathbf{x}_j^{\boldsymbol{\alpha}_j - 1} \boldsymbol{\alpha}_j \boldsymbol{\lambda}_j$, we have:

$$\det\left(\widetilde{V^{-1}} D^2 \widetilde{V^{-1}}^T\right) = \frac{\left(\frac{\mathbf{x}_1}{\boldsymbol{\alpha}_1}\right)^2 \left(\prod_{i=1}^{n+1} \mathbf{x}_i^{\boldsymbol{\alpha}_i - 2} \boldsymbol{\alpha}_i^2 \boldsymbol{\lambda}_i\right)}{4^n \left(\sum_{i=1}^{n+1} \mathbf{x}_i^{\boldsymbol{\alpha}_i} \boldsymbol{\lambda}_i\right)^{n+1}}.$$

Hence, we have: $\det\mathcal{G} = \frac{\left(\sum_{i=1}^{n+1} \frac{\mathbf{x}_i}{\boldsymbol{\alpha}_i}\right)^2 \left(\prod_{i=1}^{n+1} \mathbf{x}_i^{\boldsymbol{\alpha}_i - 2} \boldsymbol{\alpha}_i^2 \boldsymbol{\lambda}_i\right)}{4^n \left(\sum_{i=1}^{n+1} \mathbf{x}_i^{\boldsymbol{\alpha}_i} \boldsymbol{\lambda}_i\right)^{n+1}}$.

3 Proof for Proposition 3

The partial derivative of the objective function $\mathcal{F}$ with respect to $\boldsymbol{\lambda}$ is:

$$\frac{\partial \mathcal{F}}{\partial \boldsymbol{\lambda}} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial \log \mathrm{dvol} g^{-1}(\mathbf{x_i})}{\partial \boldsymbol{\lambda}} - E\left(\frac{\partial \log \mathrm{dvol} g^{-1}(\mathbf{x})}{\partial \boldsymbol{\lambda}}\right)_{p(\mathbf{x})}$$

Since we have

$$\begin{aligned}
&\frac{\partial \log \int_{\mathbb{P}_n} \mathrm{dvol} J^{-1}(\mathbf{x}) \mathrm{d}\mathbf{x}}{\partial \boldsymbol{\lambda}} \\
&= \frac{1}{\int_{\mathbb{P}_n} \mathrm{dvol} J^{-1}(\mathbf{x}) \mathrm{d}\mathbf{x}} \frac{\partial \int_{\mathbb{P}_n} \mathrm{dvol} J^{-1}(\mathbf{x}) \mathrm{d}\mathbf{x}}{\partial \boldsymbol{\lambda}} \\
&= \frac{1}{\int_{\mathbb{P}_n} \mathrm{dvol} J^{-1}(\mathbf{x}) \mathrm{d}\mathbf{x}} \int_{\mathbb{P}_n} \frac{\partial \mathrm{dvol} J^{-1}(\mathbf{x})}{\partial \boldsymbol{\lambda}} \mathrm{d}\mathbf{x} \\
&= \frac{1}{\int_{\mathbb{P}_n} \mathrm{dvol} J^{-1}(\mathbf{x}) \mathrm{d}\mathbf{x}} \int_{\mathbb{P}_n} \mathrm{dvol} J^{-1}(\mathbf{x}) \frac{\partial \log \mathrm{dvol} g^{-1}(\mathbf{x})}{\partial \boldsymbol{\lambda}} \mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{P}_n} \frac{\mathrm{dvol} J^{-1}(\mathbf{x})}{\int_{\mathbb{P}_n} \mathrm{dvol} J^{-1}(\mathbf{z}) \mathrm{d}\mathbf{z}} \frac{\partial \log \mathrm{dvol} g^{-1}(\mathbf{x})}{\partial \boldsymbol{\lambda}} \mathrm{d}\mathbf{x} \\
&= E\left(\frac{\partial \log \mathrm{dvol} J^{-1}(\mathbf{x})}{\partial \boldsymbol{\lambda}}\right)_{p(\mathbf{x})}.
\end{aligned}$$

and

$$\frac{\partial \log \mathrm{dvol} J^{-1}(\mathbf{x})}{\partial \boldsymbol{\lambda}} = \frac{n+1}{2 \sum_{i=1}^{n+1} \mathbf{x}_i^{\boldsymbol{\alpha}_i} \boldsymbol{\lambda}_i} \left[\mathbf{x}_j^{\boldsymbol{\alpha}_j}\right]_{1 \le j \le n+1} .$$

So, we have the proof for $\frac{\partial \mathcal{F}}{\partial \boldsymbol{\lambda}}$.

Similarly, we also obtain the proof for $\frac{\partial \mathcal{F}}{\partial \boldsymbol{\alpha}}$.

References

G. Lebanon. *Riemannian Geometry and Statistical Machine Learning*. PhD thesis, CMU, 2005.