A PROOF FOR INVARIANCE OF THE INVARIANT KERNEL

Lemma A.1 *The kernel* $\kappa^{\max}(\cdot,\cdot)$ *as defined in equation (2.2), is invariant, i.e.* $\kappa^{max}(\boldsymbol{v},\boldsymbol{v}') = \kappa^{max}(\boldsymbol{v}\circ\boldsymbol{\tau},\boldsymbol{v}'\circ\boldsymbol{\tau}')$ *for some data points* $\boldsymbol{v},\boldsymbol{v}'$ *and transformations* $\boldsymbol{\tau},\boldsymbol{\tau}'\in\mathbb{T}$.

Proof Let $\boldsymbol{\tau}_\star,\boldsymbol{\tau}_\star'$ be optimal for the problem in (2.2):

$$(\boldsymbol{\tau}_\star,\boldsymbol{\tau}_\star') = \arg\max_{\boldsymbol{\tau},\boldsymbol{\tau}'\in\mathbb{T}} \kappa\Big(\boldsymbol{v}\circ\boldsymbol{\tau},\boldsymbol{v}'\circ\boldsymbol{\tau}'\Big). \tag{A.1}$$

If $\bar{\boldsymbol{v}}$ and $\bar{\boldsymbol{v}}'$ are transformed versions of $\boldsymbol{v}$ and $\boldsymbol{v}'$, respectively, i.e., $\bar{\boldsymbol{v}} = \boldsymbol{v}\circ\boldsymbol{\sigma}$ and $\bar{\boldsymbol{v}}' = \boldsymbol{v}'\circ\boldsymbol{\sigma}'$, for some $\boldsymbol{\sigma},\boldsymbol{\sigma}'\in\mathbb{T}$, then we have

$$\kappa^{\max}(\bar{\boldsymbol{v}},\bar{\boldsymbol{v}}') = \max_{\boldsymbol{\tau}\in\mathbb{T},\boldsymbol{\tau}'\in\mathbb{T}} \kappa\Big(\bar{\boldsymbol{v}}\circ\boldsymbol{\tau},\bar{\boldsymbol{v}}'\circ\boldsymbol{\tau}'\Big) \tag{A.2}$$

$$= \max_{\boldsymbol{\tau}\in\mathbb{T},\boldsymbol{\tau}'\in\mathbb{T}} \kappa\Big(\boldsymbol{v}\circ\boldsymbol{\sigma}\circ\boldsymbol{\tau},\boldsymbol{v}'\circ\boldsymbol{\sigma}'\circ\boldsymbol{\tau}'\Big). \tag{A.3}$$

Because $\mathbb{T}$ is a group, it is closed under inversion and composition, and so

$$\boldsymbol{\sigma}^{-1}\circ\boldsymbol{\tau}_\star\in\mathbb{T}, \quad \text{and} \quad {\boldsymbol{\sigma}'}^{-1}\circ\boldsymbol{\tau}_\star'\in\mathbb{T}, \tag{A.4}$$

and $(\boldsymbol{\sigma}^{-1}\circ\boldsymbol{\tau}_\star,{\boldsymbol{\sigma}'}^{-1}\circ\boldsymbol{\tau}_\star')$ is optimal for (A.3). This implies that

$$\kappa^{\max}(\bar{\boldsymbol{v}},\bar{\boldsymbol{v}}') = \kappa^{\max}(\boldsymbol{v},\boldsymbol{v}'). \tag{A.5}$$

■

B OVERVIEW OF THEORETICAL ANALYSIS

Data and Assumptions. Our theory pertains to a continuum model for images. We consider images $\boldsymbol{v}$ as functions $\boldsymbol{v}(\boldsymbol{u})$ of a two-dimensional pixel location $\boldsymbol{u}\in\mathbb{R}^2$, which is (i) square integrable, i.e., $\boldsymbol{v}\in L^2(\mathbb{R}^2)$ and (ii) smooth, i.e., $\boldsymbol{v}\in C^\infty$. Our theory[4] considers (special) Euclidean transformations $\boldsymbol{\tau}\in\mathbb{T}=\mathrm{SE}(2)$, which act on images via

$$[\boldsymbol{v}\circ\boldsymbol{\tau}](\boldsymbol{u}) = \boldsymbol{v}(\boldsymbol{\tau}(\boldsymbol{u})), \tag{B.1}$$

where $\boldsymbol{u}\in\mathbb{R}^2$ is represents an arbitrary location in the image plane. The mapping $\boldsymbol{v}\mapsto\boldsymbol{v}\circ\boldsymbol{\tau}$ is an isometry, in the sense that $\|\boldsymbol{v}\circ\boldsymbol{\tau}\|_{L^2} = \|\boldsymbol{v}\|_{L^2}$ for any $\boldsymbol{v}\in\mathcal{I}$; similarly, for any $\boldsymbol{v},\boldsymbol{v}'\in\mathcal{I}$, $\|\boldsymbol{v}\circ\boldsymbol{\tau}-\boldsymbol{v}'\circ\boldsymbol{\tau}\|_{L^2} = \|\boldsymbol{v}-\boldsymbol{v}'\|_{L^2}$.

Image Manifolds. For a given image $\boldsymbol{v}$, let

$$\mathcal{S}_{\boldsymbol{v}} = \{\boldsymbol{v}\circ\boldsymbol{\tau} \mid \boldsymbol{\tau}\in\mathbb{T}\}\subset L^2(\mathbb{R}^2). \tag{B.2}$$

If the mapping

$$\boldsymbol{\tau}\mapsto\boldsymbol{v}\circ\boldsymbol{\tau} \tag{B.3}$$

is injective, i.e., $\boldsymbol{v}$ is not self-similar under any subgroup of $\mathrm{SE}(2)$, then $S_{\boldsymbol{v}}$ is a 3-dimensional compact Riemannian submanifold of L^2, which is diffeomorphic to $\mathrm{SE}(2)$.

Tangent and Normal Spaces. For any point $\boldsymbol{v}'\in\mathcal{S}_{\boldsymbol{v}}$, we let $T_{\boldsymbol{v}'}\mathcal{S}_{\boldsymbol{v}}$ denote the tangent space to $\mathcal{S}_{\boldsymbol{v}}$. This is a 3-dimensional linear subspace of $T_{\boldsymbol{v}}L^2$; we identify it with a 3-dimensional linear subspace of L^2 in a natural fashion. Because our objective functions are defined extrinsically, our analysis will occasionally need to compare tangent vectors $\boldsymbol{\xi}'\in T_{\boldsymbol{v}'}\mathcal{S}_{\boldsymbol{v}}$ and $\boldsymbol{\xi}''\in T_{\boldsymbol{v}''}\mathcal{S}_{\boldsymbol{v}}$ extrinsically, i.e., writing $\|\boldsymbol{\xi}'-\boldsymbol{\xi}''\|_{L^2}$.

At point $\boldsymbol{v}'\in\mathcal{S}_{\boldsymbol{v}}$, we let $N_{\boldsymbol{v}'}\mathcal{S}_{\boldsymbol{v}}$ denote the normal space. This is the orthogonal complement of $T_{\boldsymbol{v}'}\mathcal{S}_{\boldsymbol{v}}$ in $T_{\boldsymbol{v}'}\mathcal{I}$.

[4]The computational approach in this paper accommodates far broader classes of transformations, including similarity, affine, perspective and beyond. Our theory focuses on Euclidean transformations. We believe that similar results can be obtained for other transformation groups, albeit with certain modifications to account for scale changes.

Riemannian Structure. The manifold $\mathcal{S}_{\boldsymbol{v}}$ inherits the Riemannian metric and connection from $\mathcal{I}$. We let $\exp_{\boldsymbol{v}'} : T_{\boldsymbol{v}'}\mathcal{S}_{\boldsymbol{v}} \to \mathcal{S}_{\boldsymbol{v}}$ denote the exponential map; because $\mathcal{S}_{\boldsymbol{v}}$ is compact, it is geodesically complete, and the domain of definition of $\exp_{\boldsymbol{v}'}$ is the entirety of the tangent space.

Reach, Tubular Neighborhoods and Projections. The distance function is given by

$$d(\boldsymbol{u}, S_{\boldsymbol{v}}) = \inf_{\boldsymbol{v}' \in S_{\boldsymbol{v}}} \|\boldsymbol{u} - \boldsymbol{v}'\|_2 \tag{B.4}$$

$$= \inf_{\boldsymbol{\tau} \in \mathrm{SE}(2)} \|\boldsymbol{u} - \boldsymbol{v} \circ \boldsymbol{\tau}\|_2. \tag{B.5}$$

Because $S_{\boldsymbol{v}}$ is compact, there always exists at least one minimizing point $\boldsymbol{v}' \in S_{\boldsymbol{v}}$. The *reach* $\rho(S_{\boldsymbol{v}})$ of the manifold $S_{\boldsymbol{v}}$ is maximum $\rho \in \mathbb{R}$ such that for all points $\boldsymbol{u}$ satisfying $d(\boldsymbol{u}, S_{\boldsymbol{v}}) < \rho$, the minimizing point $\boldsymbol{v}'$ is unique. We write $\mathcal{P}_{S_{\boldsymbol{v}}}[\boldsymbol{u}]$ for this unique closest point $\boldsymbol{v}'$. By the tubular neighborhood theorem, $\rho(S_{\boldsymbol{v}}) > 0$, i.e., the reach is positive.

Write $\Gamma_{\boldsymbol{v}} = \{\boldsymbol{u} \mid d(\boldsymbol{u}, S_{\boldsymbol{v}}) < \rho(S_{\boldsymbol{v}})\}$, i.e., this is the tubular neighborhood of points which are within distance ρ of $S_{\boldsymbol{v}}$. As described above, for $\boldsymbol{u} \in \Gamma_{\boldsymbol{v}}$, the projection is unique; for $\boldsymbol{u} \in \Gamma_{\boldsymbol{v}} \setminus S_{\boldsymbol{v}}$, write

$$\boldsymbol{\eta}(\boldsymbol{u}) = \frac{\boldsymbol{u} - \mathcal{P}_{S_{\boldsymbol{v}}}\boldsymbol{u}}{\|\boldsymbol{u} - \mathcal{P}_{S_{\boldsymbol{v}}}\boldsymbol{u}\|_2} \in N_{\mathcal{P}_{S_{\boldsymbol{v}}}\boldsymbol{u}} S_{\boldsymbol{v}}. \tag{B.6}$$

Reach and Curvature for Convex Combinations. Our proposed invariant attention mechanism forms new data points as weighted combinations of transformed versions of the observed data $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. Our analysis is phrased in terms of a summary parameter, which bounds the reach of transformation manifolds generated by this operation. Let $\Delta = \{\boldsymbol{\alpha} \in \mathbb{R}^n \mid \boldsymbol{\alpha} \geq \boldsymbol{0}, \langle \boldsymbol{1}, \boldsymbol{\alpha} \rangle = 1\}$. For $\boldsymbol{\alpha} \in \Delta$ and $\boldsymbol{T} = (\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n) \in \mathbb{T}^n$, write

$$\boldsymbol{v}_{\boldsymbol{\alpha}, \boldsymbol{T}} = \sum_{i=1}^{n} \boldsymbol{\alpha}_i \boldsymbol{v}_i \circ \boldsymbol{\tau}_i, \tag{B.7}$$

and

$$\mathcal{S}_{\boldsymbol{\alpha}, \boldsymbol{T}} = \{\boldsymbol{v}_{\boldsymbol{\alpha}, \boldsymbol{T}} \circ \boldsymbol{\tau} \mid \boldsymbol{\tau} \in \mathbb{T}\}. \tag{B.8}$$

Then we define the *infimal convex combination reach* as

$$\rho_{\min}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) = \inf_{\boldsymbol{\alpha} \in \Delta, \boldsymbol{T} \in \mathbb{T}^n} \operatorname{reach}\Big(\mathcal{S}_{\boldsymbol{\alpha}, \boldsymbol{T}}\Big). \tag{B.9}$$

The reach controls the extrinsic geodesic curvature – for a given manifold $\mathcal{S}$, $\kappa(S) \leq \operatorname{reach}(\mathcal{S})^{-1}$. The convex combination reach controls the maximum curvature of any transformation manifold $\mathcal{S}_{\boldsymbol{\alpha}, \boldsymbol{T}}$: write

$$\kappa_{\max}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) = \sup_{\boldsymbol{\alpha} \in \Delta, \boldsymbol{T} \in \mathbb{T}^n} \kappa(\mathcal{S}_{\boldsymbol{\alpha}, \boldsymbol{T}}). \tag{B.10}$$

Then

$$\kappa_{\max}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) \leq \Big(\rho_{\min}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)\Big)^{-1}. \tag{B.11}$$

C Differentiating Projections

The following lemma shows that the projection of a point onto $S_{\boldsymbol{v}}$ is differentiable, and gives an expression for its derivative:

Lemma C.1 *On $\Gamma_{\boldsymbol{v}}$, the projection $\mathcal{P}_{S_{\boldsymbol{v}}}\boldsymbol{u}$ is a differentiable function of $\boldsymbol{u}$. Moreover, setting $\boldsymbol{v}' = \mathcal{P}_{S_{\boldsymbol{v}}}\boldsymbol{u}$ and $T = T_{\boldsymbol{v}'}S_{\boldsymbol{v}}$, we have that*

$$\frac{d}{dt}\Big|_0 \mathcal{P}_{S_{\boldsymbol{v}}}[\boldsymbol{u} + t\boldsymbol{r}] \tag{C.1}$$

is the unique solution $\boldsymbol{h} \in T$ to the equation

$$\Big(\boldsymbol{I} - \mathrm{I\!I}^*(\boldsymbol{u} - \boldsymbol{v}')\Big)\boldsymbol{h} = \mathcal{P}_T \boldsymbol{r}, \tag{C.2}$$

where $\mathrm{I\!I}$ *is the second fundamental form at* $\boldsymbol{v}'$*, and* $\mathrm{I\!I}^*$ *is the symmetric operator defined by*

$$\Big(\mathrm{I\!I}^*\boldsymbol{\eta}\Big)[\boldsymbol{p},\boldsymbol{q}] = \langle \mathrm{I\!I}(\boldsymbol{p},\boldsymbol{q}),\boldsymbol{\eta}\rangle\,, \tag{C.3}$$

which for $d(\boldsymbol{u},S_{\boldsymbol{v}}) < \rho(S_{\boldsymbol{v}})$ *satisfies*

$$\|\mathrm{I\!I}^*(\boldsymbol{u}-\boldsymbol{v}')\| < 1. \tag{C.4}$$

In particular, this implies that the derivative can be expressed in a convergent Neumann series

$$\boldsymbol{h}_\star = \sum_{k=0}^{\infty}\Big(\mathrm{I\!I}^*(\boldsymbol{u}-\boldsymbol{v}')\Big)^k\mathcal{P}_T\boldsymbol{r}. \tag{C.5}$$

Here, the $k=0$ term is the projection of $\boldsymbol{r}$ onto the tangent space at $\boldsymbol{v}'$; the higher powers account for changes in the normal direction as one moves across the manifold. Interestingly, although this formula takes into account changes in the normal direction, it only depends on $\boldsymbol{r}$ through its projection onto the tangent space, $\mathcal{P}_T\boldsymbol{r}$.

Proof Because $S_{\boldsymbol{v}}$ is geodesically complete, any point $\boldsymbol{v}'' \in S_{\boldsymbol{v}}$ can be expressed as $\exp_{\boldsymbol{v}'}(\boldsymbol{h})$ for some $\boldsymbol{h}\in T$. The projection problem is equivalent to solving

$$\min_{\boldsymbol{h}\in T}\frac{1}{2}\Big\|\boldsymbol{u}+t\boldsymbol{r}-\exp_{\boldsymbol{v}'}(\boldsymbol{h})\Big\|_2^2 \equiv \varphi(\boldsymbol{h}) \tag{C.6}$$

in the sense that

$$\mathcal{P}_{S_{\boldsymbol{v}}}[\boldsymbol{u}+t\boldsymbol{r}] = \exp_{\boldsymbol{v}'}(\boldsymbol{h}_\star), \tag{C.7}$$

where $\boldsymbol{h}_\star$ is the unique optimal solution to (C.6). The exponential map $\exp_{\boldsymbol{v}'}(\boldsymbol{h})$ is a smooth[5] function of $\boldsymbol{h}$, and so φ is a smooth (and hence, differentiable) function of $\boldsymbol{h}$. The solution $\boldsymbol{h}_\star$ is a critical point of φ over T, and so

$$\mathcal{P}_T\nabla_{\boldsymbol{h}}\varphi(\boldsymbol{h}_\star) = \boldsymbol{0}. \tag{C.8}$$

Differentiating the objective function φ, we have

$$\mathcal{P}_T\left(\nabla_{\boldsymbol{h}}\exp_{\boldsymbol{v}'}(\boldsymbol{h})\Big|_{\boldsymbol{h}=\boldsymbol{h}_\star}\right)^*\Big(\exp_{\boldsymbol{v}'}(\boldsymbol{h}_\star)-(\boldsymbol{u}+t\boldsymbol{r})\Big) = \boldsymbol{0}, \tag{C.9}$$

where $\cdot^*$ represents the adjoint of the Jacobian $\nabla_{\boldsymbol{h}}\exp_{\boldsymbol{v}'}(\boldsymbol{h})$. The exponential map can be expanded as

$$\begin{aligned} \exp_{\boldsymbol{v}'}(\boldsymbol{h}) &= \boldsymbol{v}'+\boldsymbol{h}+\tfrac{1}{2}\mathrm{I\!I}(\boldsymbol{h},\boldsymbol{h})+\boldsymbol{\xi}(\boldsymbol{h}), &\text{(C.10)}\\ \nabla_{\boldsymbol{h}}\exp_{\boldsymbol{v}'}(\boldsymbol{h}) &= \boldsymbol{I}+\mathrm{I\!I}(\boldsymbol{h},\cdot)+\nabla_{\boldsymbol{h}}\boldsymbol{\xi}(\boldsymbol{h}), &\text{(C.11)}\end{aligned}$$

where the residual term $\boldsymbol{\xi}(\boldsymbol{h})$ satisfies

$$\boldsymbol{\xi}(\boldsymbol{h}) \le C\|\boldsymbol{h}\|_2^3, \qquad \|\nabla_{\boldsymbol{h}}\boldsymbol{\xi}(\boldsymbol{h})\|_{\ell^2\to\ell^2} \le C\|\boldsymbol{h}\|_2^2 \tag{C.12}$$

for some $C\in\mathbb{R}_+$ which does not depend on $\boldsymbol{h}$. Plugging in, we obtain the equation

$$\mathcal{P}_T\Big(\boldsymbol{I}+\mathrm{I\!I}(\boldsymbol{h}_\star,\cdot)+\nabla_{\boldsymbol{h}}\boldsymbol{\xi}(\boldsymbol{h}_\star)\Big)^*\Big(\boldsymbol{v}'+\boldsymbol{h}_\star+\tfrac{1}{2}\mathrm{I\!I}(\boldsymbol{h}_\star,\boldsymbol{h}_\star)+\boldsymbol{\xi}(\boldsymbol{h}_\star)-(\boldsymbol{u}+t\boldsymbol{r})\Big) = \boldsymbol{0}, \tag{C.13}$$

whence

$$\begin{aligned}\mathcal{P}_T\Big(\boldsymbol{I}-\mathrm{I\!I}^*\Big(\boldsymbol{u}-\boldsymbol{v}'\Big)\Big)\boldsymbol{h}_\star \;&=\; \mathcal{P}_T\Big(\boldsymbol{u}+t\boldsymbol{r}-\boldsymbol{v}'\Big) \;+\; t\,\mathrm{I\!I}^*(\boldsymbol{r})\boldsymbol{h}_\star - \;\mathcal{P}_T\Big(\tfrac{1}{2}\mathrm{I\!I}(\boldsymbol{h}_\star,\boldsymbol{h}_\star)+\boldsymbol{\xi}(\boldsymbol{h}_\star)\Big)\\ &\quad - \mathcal{P}_T\mathrm{I\!I}^*\Big(\boldsymbol{h}_\star+\tfrac{1}{2}\mathrm{I\!I}(\boldsymbol{h}_\star,\boldsymbol{h}_\star)+\boldsymbol{\xi}(\boldsymbol{h}_\star)\Big)\boldsymbol{h}_\star\\ &\quad - \mathcal{P}_T(\nabla_{\boldsymbol{h}}\boldsymbol{\xi}(\boldsymbol{h}_\star))^*\Big(\boldsymbol{v}'+\boldsymbol{h}_\star+\tfrac{1}{2}\mathrm{I\!I}(\boldsymbol{h}_\star,\boldsymbol{h}_\star)+\boldsymbol{\xi}(\boldsymbol{h}_\star)-(\boldsymbol{u}+t\boldsymbol{r})\Big).\end{aligned}$$

[5]Smoothness follows from smooth dependence of solutions of ordinary differential equations on their initial conditions – see e.g., Lee (2006) Proposition 5.7.

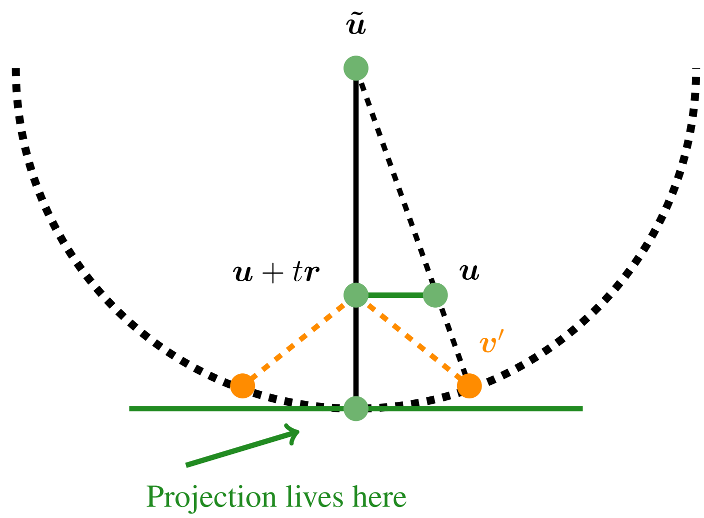

Figure 6: Proof of Lemma C.2

By Lemma C.3,

$$\left\| \mathrm{II}^* \Big(\boldsymbol{u} - \boldsymbol{v}' \Big) \right\| \le \kappa \| \boldsymbol{u} - \boldsymbol{v}' \| < 1. \tag{C.14}$$

This implies that the operator

$$\mathcal{P}_T \left(\boldsymbol{I} - \mathrm{II}^* \Big(\boldsymbol{u} - \boldsymbol{v}' \Big) \right) \mathcal{P}_T \tag{C.15}$$

is stably invertible over T, and its inverse is given by the Neumann series

$$\left(\mathcal{P}_T \left(\boldsymbol{I} - \mathrm{II}^* \Big(\boldsymbol{u} - \boldsymbol{v}' \Big) \right) \mathcal{P}_T \right)^{-1} = \sum_k \left(\mathrm{II}^* \Big(\boldsymbol{u} - \boldsymbol{v}' \Big) \right)^k, \tag{C.16}$$

whence

$$\left\| \left(\mathcal{P}_T \left(\boldsymbol{I} - \mathrm{II}^* \Big(\boldsymbol{u} - \boldsymbol{v}' \Big) \right) \mathcal{P}_T \right)^{-1} \right\|_{\ell^2 \to \ell^2} \le \sum_k \left\| \mathrm{II}^* \Big(\boldsymbol{u} - \boldsymbol{v}' \Big) \right\|^k \le \frac{1}{1 - \kappa \| \boldsymbol{u} - \boldsymbol{v}' \|}. \tag{C.17}$$

Using this and the fact that $\boldsymbol{u} - \boldsymbol{v}' \in T^\perp$ (which implies that $\mathcal{P}_T(\boldsymbol{u} + t\boldsymbol{r} - \boldsymbol{v}')$) we obtain that

$$\left\| \boldsymbol{h}_\star - t \left[\mathcal{P}_T \left(\boldsymbol{I} - \mathrm{II}^* \Big(\boldsymbol{u} - \boldsymbol{v}' \Big) \right) \mathcal{P}_T \right]^{-1} \mathcal{P}_T \boldsymbol{r} \right\| \le C t^2, \tag{C.18}$$

where we have used Lemma C.2, which shows that $\| \boldsymbol{h}_\star \| \le Ct$. This implies that

$$\lim_{t \to 0} \frac{\boldsymbol{h}_\star(t) - \boldsymbol{h}_\star(0)}{t} = \left[\mathcal{P}_T \left(\boldsymbol{I} - \mathrm{II}^* \Big(\boldsymbol{u} - \boldsymbol{v}' \Big) \right) \mathcal{P}_T \right]^{-1} \mathcal{P}_T \boldsymbol{r}, \tag{C.19}$$

whence $\boldsymbol{h}_\star$ is differentiable with respect to t at $t = 0$, and its derivative is given by the right hand side of (C.19). This implies that $\mathcal{P}_{S_{\boldsymbol{v}}}[\boldsymbol{u} + t\boldsymbol{r}]$ is differentiable at $t = 0$, and

$$\frac{d}{dt}\Big|_0 \mathcal{P}_{S_{\boldsymbol{v}}}[\boldsymbol{u} + t\boldsymbol{r}] = \frac{d}{dt}\Big|_0 \exp_{\boldsymbol{v}'}(\boldsymbol{h}_\star(t)) = \frac{d}{dt}\Big|_0 \boldsymbol{h}_\star(t), \tag{C.20}$$

giving the claimed result. ■

The following lemma argues that for small t, the projection stays local.

Lemma C.2 *Let $\boldsymbol{u} \in \Gamma_{\boldsymbol{v}}$, and consider a perturbation $\boldsymbol{u} + t\boldsymbol{r}$, with $\|\boldsymbol{r}\| = 1$. Then there exist positive numbers c, C, which do not depend on t, such that for $t < c$,*

$$d_{\mathcal{S}_{\boldsymbol{v}}}\Big(\mathcal{P}_{S_{\boldsymbol{v}}}[\boldsymbol{u} + t\boldsymbol{r}], \mathcal{P}_{S_{\boldsymbol{v}}}[\boldsymbol{u}]\Big) < Ct. \tag{C.21}$$

Proof Let $\boldsymbol{u}(t) = \boldsymbol{u} + t\boldsymbol{r}$. For appropriately chosen $c_1 > 0$, $\boldsymbol{u}(t)$ lies in $\Gamma_{\boldsymbol{v}}$ for $0 \le t < c_1$. Let

$$\boldsymbol{v}'(t) = \mathcal{P}_{\mathcal{S}_{\boldsymbol{v}}}[\boldsymbol{u}(t)] \tag{C.22}$$

be projection onto $\mathcal{S}_{\boldsymbol{v}}$. Because $\boldsymbol{u} \in \Gamma_{\boldsymbol{v}}$, $\|\boldsymbol{u} - \boldsymbol{v}'(0)\| < \rho$. Choose $\varepsilon > 0$ such that $(1-\varepsilon)\rho > d(\boldsymbol{u}, \mathcal{S}_{\boldsymbol{v}})$. Set

$$\tilde{\boldsymbol{u}} = \boldsymbol{v}'(0) + (1-\varepsilon)\rho \frac{\boldsymbol{u} - \boldsymbol{v}'(0)}{\|\boldsymbol{u} - \boldsymbol{v}'(0)\|} \in \Gamma_{\boldsymbol{v}}. \tag{C.23}$$

Figure 6 illustrates this construction.

Notice that $\|\boldsymbol{u} - \tilde{\boldsymbol{u}}\| = \xi \equiv (1-\varepsilon)\rho - \|\boldsymbol{u} - \boldsymbol{v}'\| > 0$. Hence, for $t < \xi/2\|\boldsymbol{r}\|$, $\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\| \ge \xi/2$, which is constant with respect to t. For $t < c_2 \equiv \xi/2\|\boldsymbol{r}\|$, set

$$\boldsymbol{w}(t) = \tilde{\boldsymbol{u}} + (1-\varepsilon)\rho \frac{\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}}{\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|_2}. \tag{C.24}$$

Setting

$$\boldsymbol{z}(t) = \frac{\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}}{\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|_2}, \tag{C.25}$$

we have

$$\begin{aligned}
\|\boldsymbol{z}(t) - \boldsymbol{z}(0)\|_2 &= \left\| \frac{\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}}{\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|_2} - \frac{\boldsymbol{u}(0) - \tilde{\boldsymbol{u}}}{\|\boldsymbol{u}(0) - \tilde{\boldsymbol{u}}\|_2} \right\| && \text{(C.26)}\\
&\le \left\| \frac{\boldsymbol{u}(t) - \boldsymbol{u}(0)}{\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|_2} \right\| + \left\| \frac{\boldsymbol{u}(0) - \tilde{\boldsymbol{u}}}{\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|_2} - \frac{\boldsymbol{u}(0) - \tilde{\boldsymbol{u}}}{\|\boldsymbol{u}(0) - \tilde{\boldsymbol{u}}\|_2} \right\| && \text{(C.27)}\\
&= \frac{\|\boldsymbol{u}(t) - \boldsymbol{u}(0)\|}{\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|_2} + \|\boldsymbol{u}(0) - \tilde{\boldsymbol{u}}\| \left| \frac{1}{\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|} - \frac{1}{\|\boldsymbol{u}(0) - \tilde{\boldsymbol{u}}\|} \right| && \text{(C.28)}\\
&\le \frac{\|\boldsymbol{u}(t) - \boldsymbol{u}(0)\|}{\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|_2} + \|\boldsymbol{u}(0) - \tilde{\boldsymbol{u}}\| \left| \frac{\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\| - \|\boldsymbol{u}(0) - \tilde{\boldsymbol{u}}\|}{\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\| \|\boldsymbol{u}(0) - \tilde{\boldsymbol{u}}\|} \right| && \text{(C.29)}\\
&\le 2\frac{\|\boldsymbol{u}(t) - \boldsymbol{u}(0)\|}{\|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|} < \frac{4t\|\boldsymbol{r}\|}{\xi} \equiv C_1 t. && \text{(C.30)}
\end{aligned}$$

Then

$$\|\boldsymbol{w}(t) - \boldsymbol{v}'(0)\| = \|\boldsymbol{w}(t) - \boldsymbol{w}(0)\|_2 < (1-\varepsilon)\rho C_1 t \equiv C_2 t. \tag{C.31}$$

We make two geometric observations, which together constrain $\boldsymbol{v}'(t)$ to be within distance Ct of $\boldsymbol{w}(t)$, and hence within distance $C't$ of $\boldsymbol{w}(0) = \boldsymbol{v}'(0)$. First, the projection of $\tilde{\boldsymbol{u}}$ onto $\mathcal{S}_{\boldsymbol{v}}$ is unique, and equal to $\boldsymbol{v}'(0)$. Hence

$$\|\boldsymbol{v}'(t) - \tilde{\boldsymbol{u}}\| \ge d(\tilde{\boldsymbol{u}}, \mathcal{S}_{\boldsymbol{v}}) = (1-\varepsilon)\rho. \tag{C.32}$$

Second, because $\boldsymbol{v}'(t)$ is the unique closest point to $\boldsymbol{u}(t)$ in $\mathcal{S}_{\boldsymbol{v}}$, we have

$$\|\boldsymbol{v}'(t) - \boldsymbol{u}(t)\| \le \|\boldsymbol{v}'(0) - \boldsymbol{u}(t)\| \tag{C.33}$$

In other words, setting $r = \|\boldsymbol{v}'(0) - \boldsymbol{u}(t)\|$, we have

$$\boldsymbol{v}'(t) \in Q \equiv B(\boldsymbol{u}(t), r) \cap \Big(\mathrm{int}\,[B(\tilde{\boldsymbol{u}}, (1-\varepsilon)\rho)]\Big)^c. \tag{C.34}$$

This set is in turn contained in the intersection of $B(\boldsymbol{u}(t), r)$ with the halfspace of points $\boldsymbol{q}$ satisfying $\langle \boldsymbol{q} - \tilde{\boldsymbol{u}}, \boldsymbol{z}(t)\rangle \ge \langle \boldsymbol{v}'(0) - \tilde{\boldsymbol{u}}, \boldsymbol{z}(t)\rangle$:

$$Q \subseteq Q' \equiv B(\boldsymbol{u}(t), r) \cap \Big\{\boldsymbol{q} \mid \langle \boldsymbol{q} - \boldsymbol{u}(t), \boldsymbol{z}(t)\rangle \ge \langle \boldsymbol{v}'(0) - \boldsymbol{u}(t), \boldsymbol{z}(t)\rangle\Big\}. \tag{C.35}$$

We demonstrate this by showing that every point $\boldsymbol{q}$ satisfying

$$\|\boldsymbol{q} - \tilde{\boldsymbol{u}}\| \ge (1-\varepsilon)\rho = \|\boldsymbol{v}'(0) - \tilde{\boldsymbol{u}}\| \tag{C.36}$$

and

$$\langle \boldsymbol{q} - \tilde{\boldsymbol{u}}, \boldsymbol{z}(t) \rangle < \langle \boldsymbol{v}'(0) - \tilde{\boldsymbol{u}}, \boldsymbol{z}(t) \rangle \tag{C.37}$$

must satisfy $\|\boldsymbol{q} - \boldsymbol{u}(t)\| > r = \|\boldsymbol{v}'(0) - \boldsymbol{u}(t)\|$. First, note that

$$\begin{aligned} \|\boldsymbol{q} - \tilde{\boldsymbol{u}}\|_2^2 &= \|\boldsymbol{q} - \boldsymbol{u}(t)\|_2^2 + \|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|_2^2 + 2\langle \boldsymbol{q} - \boldsymbol{u}(t), \boldsymbol{u}(t) - \tilde{\boldsymbol{u}} \rangle \\ &= \|\boldsymbol{q} - \boldsymbol{u}(t)\|_2^2 - \|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|_2^2 + 2\langle \boldsymbol{q} - \tilde{\boldsymbol{u}}, \boldsymbol{z}(t) \rangle \end{aligned} \tag{C.38}$$

and similarly

$$\|\boldsymbol{v}'(0) - \tilde{\boldsymbol{u}}\|_2^2 = \|\boldsymbol{v}'(0) - \boldsymbol{u}(t)\|_2^2 - \|\boldsymbol{u}(t) - \tilde{\boldsymbol{u}}\|_2^2 + 2\langle \boldsymbol{v}'(0) - \boldsymbol{u}(t), \boldsymbol{z}(t) \rangle . \tag{C.39}$$

Comparing (C.38) and (C.39) via (C.36), we see that

$$\begin{aligned} & \|\boldsymbol{q} - \boldsymbol{u}(t)\|_2^2 + 2(1-\varepsilon)\rho \langle \boldsymbol{q} - \tilde{\boldsymbol{u}}, \boldsymbol{z}(t) \rangle \\ \geq \; & \|\boldsymbol{v}'(0) - \boldsymbol{u}(t)\|_2^2 + 2(1-\varepsilon)\rho \langle \boldsymbol{v}'(0) - \boldsymbol{u}(t), \boldsymbol{z}(t) \rangle . \end{aligned} \tag{C.40}$$

Together with (C.37), this implies that $\|\boldsymbol{q} - \boldsymbol{u}(t)\|_2 > \|\boldsymbol{v}'(0) - \boldsymbol{u}(t)\|_2 = r$. This establishes (C.35).

The set Q' is a solid spherical cap, whose central axis is along the direction $\boldsymbol{w}(t) - \boldsymbol{u}(t)$. Set

$$\bar{\boldsymbol{w}}(t) = \frac{\boldsymbol{w}(t) - \boldsymbol{u}(t)}{\|\boldsymbol{w}(t) - \boldsymbol{u}(t)\|} \langle \boldsymbol{v}'(0) - \boldsymbol{u}(t), \boldsymbol{z}(t) \rangle . \tag{C.41}$$

The spherical cap Q' has diameter

$$2\|\bar{\boldsymbol{w}}(t) - \boldsymbol{v}'(0)\| \leq 2\|\boldsymbol{w}(t) - \boldsymbol{v}'(0)\|, \tag{C.42}$$

Since $\boldsymbol{v}'(t) \in Q'$,

$$\|\boldsymbol{v}'(t) - \boldsymbol{v}'(0)\| \leq 2\|\boldsymbol{w}(t) - \boldsymbol{v}'(0)\| \leq 2C_2 t. \tag{C.43}$$

By Theorem 1 of Boissonnat et al. (2019), provided $2C_2 t < 2\rho$, we have

$$\begin{aligned} d_{\mathcal{S}_{\boldsymbol{v}}}\Big(\boldsymbol{v}'(t), \boldsymbol{v}'(0)\Big) &\leq 2\rho \sin^{-1}\left(\frac{\|\boldsymbol{v}'(t) - \boldsymbol{v}'(0)\|}{2\rho}\right), && \text{(C.44)} \\ &\leq \frac{\pi}{2}\|\boldsymbol{v}'(t) - \boldsymbol{v}'(0)\|, && \text{(C.45)} \\ &\leq C_2 \pi t, && \text{(C.46)} \end{aligned}$$

as claimed. ■

Lemma C.3 *For any $\boldsymbol{\eta}$, the linear operator $\mathrm{I\!I}^*(\boldsymbol{\eta}) : T_{\boldsymbol{v}'}\mathcal{S}_{\boldsymbol{v}} \to T_{\boldsymbol{v}'}\mathcal{S}_{\boldsymbol{v}}$ satisfies*

$$\|\mathrm{I\!I}^*(\boldsymbol{\eta})\|_{\ell^2 \to \ell^2} \leq \kappa \|\boldsymbol{\eta}\| \tag{C.47}$$

Proof Because $\mathrm{I\!I}^*(\boldsymbol{\eta})$ is a symmetric linear operator,

$$\|\mathrm{I\!I}^*(\boldsymbol{\eta})\| = \left| \max_{\|\boldsymbol{h}\|=1} \mathrm{I\!I}^*(\boldsymbol{\eta})[\boldsymbol{h}, \boldsymbol{h}] \right| = \max_{\|\boldsymbol{h}\|=1} |\langle \boldsymbol{\eta}, \mathrm{I\!I}[\boldsymbol{h}, \boldsymbol{h}] \rangle| \leq \kappa \|\boldsymbol{\eta}\|_2, \tag{C.48}$$

where we have used the Cauchy-Schwarz inequality and the fact that for $\|\boldsymbol{h}\| = 1$, $\|\mathrm{I\!I}[\boldsymbol{h}, \boldsymbol{h}]\| \leq \kappa$. ■

D Proof of Uniqueness of the Invariant Mean

In this section, we prove that under our hypotheses, the invariant mean is unique up to transformations. We reproduce the theorem statement here:

Theorem D.1 (Uniqueness of the Invariant Mean) *Consider data points $\{\boldsymbol{v}_i\}_{i=1}^n$ and their corresponding transformation manifolds $\mathcal{S}_i = \{\boldsymbol{v}_i \circ \boldsymbol{\tau} \mid \boldsymbol{\tau} \in \mathbb{T} = \mathrm{SE}(2)\}$, and let $\rho_{\min}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ denote the infimal convex combination reach, defined in (B.9). Consider the optimization problem*

$$\min_{\boldsymbol{\mu} \in L^2(\mathbb{R}^2)} \varphi(\boldsymbol{\mu}) \equiv \sum_{i=1}^n \min_{\boldsymbol{\tau}_i \in \mathbb{T}} \boldsymbol{W}_{ij} \|\boldsymbol{v}_i \circ \boldsymbol{\tau}_i - \boldsymbol{\mu}\|_{L^2}^2, \tag{D.1}$$

with $\boldsymbol{W}_{ij} \geq 0$ and $\sum_i \boldsymbol{W}_{ij} = 1$. There exists a numerical constant $c > 0$ such that if

$$\max_{i,j} d(\mathcal{S}_i, \mathcal{S}_j) \leq c\, \rho_{\min}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n), \tag{D.2}$$

then the solution to (D.1) is unique up to transformation, in the sense that for any pair of solutions $\boldsymbol{\mu}^\star$, $\boldsymbol{\mu}^{\star\prime}$, we have $\boldsymbol{\mu}^{\star\prime} = \boldsymbol{\mu}^\star \circ \boldsymbol{\tau}$ for some $\boldsymbol{\tau} \in \mathbb{T}$.

Proof Set

$$\mathcal{I} = \{\boldsymbol{\mu} \in L^2 \mid \min_j d(\boldsymbol{\mu}, \mathcal{S}_j) \le R\}. \tag{D.3}$$

By Lemma D.2, every minimizer $\boldsymbol{\mu}^\star$ of (2.3) satisfies $\boldsymbol{\mu}^\star \in \mathcal{I}$.

Fix an arbitrary minimizer $\boldsymbol{\mu}^\star$ and consider any $\boldsymbol{\mu} \in \mathcal{I} \setminus \mathcal{S}_{\boldsymbol{\mu}^\star}$. Let $\widehat{\boldsymbol{\mu}}$ be the (unique) closest point on $\mathcal{S}_{\boldsymbol{\mu}^\star}$. Then $\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}} \in N_{\widehat{\boldsymbol{\mu}}} S_{\boldsymbol{\mu}_j^\star}$, and $D = \|\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}\| \le 2R$. Letting

$$\boldsymbol{\nu} = \frac{\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}}{\|\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}\|_2}, \tag{D.4}$$

we have

$$\begin{aligned}
\varphi(\boldsymbol{\mu}) &= \varphi(\widehat{\boldsymbol{\mu}} + D\boldsymbol{\nu}) && \text{(D.5)}\\
&= \varphi(\widehat{\boldsymbol{\mu}}) + \int_{t=0}^{D} \tfrac{d}{ds}\varphi(\widehat{\boldsymbol{\mu}} + s\boldsymbol{\nu})\Big|_{s=t} dt && \text{(D.6)}\\
&= \varphi(\widehat{\boldsymbol{\mu}}) + D\tfrac{d}{ds}\varphi(\widehat{\boldsymbol{\mu}} + s\boldsymbol{\nu})\Big|_{s=0} + \int_{t=0}^{D}\int_{s=0}^{t} \tfrac{d^2}{dr^2}\varphi(\widehat{\boldsymbol{\mu}} + r\boldsymbol{\nu})\Big|_{r=s} ds\, dt && \text{(D.7)}\\
&> \varphi(\widehat{\boldsymbol{\mu}}) = \varphi(\boldsymbol{\mu}^\star). && \text{(D.8)}
\end{aligned}$$

where we have applied Theorem D.3 to show that the integrand in (D.7) is positive, and used the fact that $\widehat{\boldsymbol{\mu}}$ is a critical point of φ to conclude that

$$\tfrac{d}{ds}\varphi(\widehat{\boldsymbol{\mu}} + s\boldsymbol{\nu})\Big|_{s=0} = 0. \tag{D.9}$$

Hence, for any $\boldsymbol{\mu} \in \mathcal{I} \setminus \mathcal{S}_{\boldsymbol{\mu}^\star}$, $\varphi(\boldsymbol{\mu}) > \varphi(\boldsymbol{\mu}^\star)$, and so every optimal solution to (D.1) lies on $\mathcal{S}_{\boldsymbol{\mu}^\star}$, and hence is of the form $\boldsymbol{\mu}^\star \circ \boldsymbol{\tau}$ for some $\boldsymbol{\tau} \in \mathbb{T}$. ■

D.1 Objective function value in the Outside Region

The following lemma shows that any minimizer to (D.1) is close to at least one of the manifolds $\mathcal{S}_i$:

Lemma D.2 (Objective Value in Outside Region) *Under the hypotheses of Theorem D.1, if $\boldsymbol{\mu}^\star$ is any minimizer of (D.1), and*

$$\boldsymbol{\mu} \in \mathcal{O} = \left\{\boldsymbol{\mu} \mid \min_i d(\boldsymbol{\mu}, \mathcal{S}_i) > R\right\}, \tag{D.10}$$

where

$$R = \max_{k,\ell} d(\mathcal{S}_k, \mathcal{S}_\ell), \tag{D.11}$$

then $\varphi(\boldsymbol{\mu}) > \varphi(\boldsymbol{\mu}^\star)$, for any minimizer $\boldsymbol{\mu}^\star$ of (D.1).

Proof Using the closed form expression for $\boldsymbol{\mu}^\star$ in (2.5), we have

$$\begin{aligned}
d(\boldsymbol{\mu}^\star, \mathcal{S}_i) &= \min_{\boldsymbol{\tau}_i} \left\| \boldsymbol{v}_i \circ \boldsymbol{\tau}_i - \frac{1}{\sum_{\ell'} \boldsymbol{W}_{\ell' j}} \sum_{\ell} \boldsymbol{W}_{\ell j} \left(\boldsymbol{v}_\ell \circ \boldsymbol{\tau}_\ell^\star\right) \right\|_2 \\
&= \frac{1}{\sum_{\ell'} \boldsymbol{W}_{\ell' j}} \min_{\boldsymbol{\tau}_i} \left\| \sum_{\ell j} \boldsymbol{W}_{\ell j} (\boldsymbol{v}_i \circ \boldsymbol{\tau}_i - \boldsymbol{v}_\ell \circ \boldsymbol{\tau}_\ell^\star) \right\|_2 \\
&\le \frac{1}{\sum_{\ell'} \boldsymbol{W}_{\ell' j}} \sum_{\ell} \boldsymbol{W}_{\ell j} \min_{\boldsymbol{\tau}_i} \|\boldsymbol{v}_i \circ \boldsymbol{\tau}_i - \boldsymbol{v}_\ell \circ \boldsymbol{\tau}_\ell^\star\|_2 \\
&\le \frac{1}{\sum_{\ell'} \boldsymbol{W}_{\ell' j}} \sum_{\ell} \boldsymbol{W}_{\ell j} d(\mathcal{S}_i, \mathcal{S}_\ell) \\
&\le \frac{1}{\sum_{\ell'} \boldsymbol{W}_{\ell' j}} \sum_{\ell} \boldsymbol{W}_{\ell j} R \\
&\le R,
\end{aligned} \tag{D.12}$$

and so

$$\begin{aligned}\varphi(\boldsymbol{\mu}^\star) &= \sum_{i=1}^n \boldsymbol{W}_{ij} \min_{\boldsymbol{\tau}_i} \|\boldsymbol{v}_i \circ \boldsymbol{\tau}_i - \boldsymbol{\mu}^\star\|_2^2, \\ &= \sum_{i=1}^n W_{ij} d^2(\boldsymbol{\mu}^\star, \mathcal{S}_i), \\ &\le \sum_{i=1}^n W_{ij} R^2.\end{aligned}$$

Conversely, for $\boldsymbol{\mu} \in \mathcal{O}$,

$$\varphi(\boldsymbol{\mu}) = \sum_{i=1}^n \boldsymbol{W}_{ij} d^2(\boldsymbol{\mu}, \mathcal{S}_i) > \sum_{i=1}^n \boldsymbol{W}_{ij} R^2 \ge \varphi(\boldsymbol{\mu}^\star), \tag{D.13}$$

as claimed. ∎

D.2 Positive Curvature in the Inside Region

Theorem D.3 (Positive Curvature in the Inside Region) *Let*

$$R = \max_{k,\ell} d(\mathcal{S}_k, \mathcal{S}_\ell). \tag{D.14}$$

There is a positive numerical constant $c > 0$ such that if

$$R < c\, \rho_{\min}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n), \tag{D.15}$$

then for any optimal solution $\boldsymbol{\mu}_j^\star$ to (2.3) with corresponding transformation manifold

$$\mathcal{S}_{\boldsymbol{\mu}_j^\star} = \left\{ \boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau} \mid \boldsymbol{\tau} \in \mathbb{T} \right\}, \tag{D.16}$$

and any unit vector $\boldsymbol{r} \in \mathcal{N}_{\boldsymbol{\mu}_j^\star} \mathcal{S}_{\boldsymbol{\mu}_j^\star}$, for $t \in [0, 2R]$, $\varphi(\boldsymbol{\mu}_j^\star + t\boldsymbol{r})$ is a twice-differentiable function of t, and

$$\frac{d^2}{dt^2} \varphi\big(\boldsymbol{\mu}_j^\star + t\boldsymbol{r}\big) \ge \frac{1}{3}. \tag{D.17}$$

Proof Lemma D.6 implies that for $t \in [0, 2R]$, $\varphi(\boldsymbol{\mu}_j^\star + t\boldsymbol{r})$ is twice differentiable, and

$$\frac{d^2}{dt^2} \varphi(\boldsymbol{\mu}_j^\star + t\boldsymbol{r}) \ge 2 \sum_{i=1}^n \boldsymbol{W}_{ij} \left(\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]} \mathcal{S}_i} [\boldsymbol{r}]\|_2^2 - \frac{3\kappa_{\max} R}{1 - 3\kappa_{\max} R} \right).$$

By Lemma D.4, when (D.15) is satisfied for $c > 0$ sufficiently small, for all i, j,

$$\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]} \mathcal{S}_i} [\boldsymbol{r}]\|_2^2 \ge \tfrac{1}{2}. \tag{D.18}$$

Using that $\sum_{i=1}^n \boldsymbol{W}_{ij} = 1$, and $\kappa_{\max} \le 1/\rho_{\min}$, for $c < \frac{1}{12}$, we have

$$\frac{d^2}{dt^2} \varphi(\boldsymbol{\mu}_j^\star + t\boldsymbol{r}) \ge 2 \left(\sum_{i=1}^n \boldsymbol{W}_{ij} \right) \times \left(\frac{1}{2} - \frac{3c}{1 - 3c} \right) \ge \frac{1}{3}, \tag{D.19}$$

as claimed. ∎

Lemma D.4 *There exists a numerical constant $c > 0$ such that if*

$$R = \max_{k,\ell} d(\mathcal{S}_k, \mathcal{S}_\ell) < c\, \rho_{\min}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n), \tag{D.20}$$

then for any unit normal vector $\boldsymbol{r} \in \mathcal{N}_{\boldsymbol{\mu}_j^\star} \mathcal{S}_{\boldsymbol{\mu}_j^\star}$ and $t \in [0, 2R]$,

$$\|P_{\mathcal{N}_{\mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]} \mathcal{S}_i} [\boldsymbol{r}]\|_2^2 \ge \tfrac{1}{2}.$$

Proof Let

$$\Gamma_A = \mathcal{N}_{\mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}\mathcal{S}_i \tag{D.21}$$

$$\Gamma_B = \mathcal{N}_{\mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star]}\mathcal{S}_i \tag{D.22}$$

and let P_A, P_B denote the projection matrices onto Γ_A and Γ_B, respectively. Similarly, let $P_A^\perp, P_B^\perp$ to be the projection matrices onto the orthogonal complements $\Gamma_A^\perp$ and $\Gamma_B^\perp$. Then,

$$\begin{aligned}
\left\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{\mathcal{S}_i}[\mu_j^\star + t\boldsymbol{r}]}\mathcal{S}_i}[\boldsymbol{r}]\right\|^2 &= \|P_A[\boldsymbol{r}]\|_2^2 \\
&= 1 - \left\|P_A^\perp[\boldsymbol{r}]\right\|_2^2 \\
&= 1 - \left\|P_A^\perp[P_B\boldsymbol{r} + P_B^\perp\boldsymbol{r}]\right\|_2^2 \\
&\geq 1 - 2\left\|P_A^\perp P_B\boldsymbol{r}\right\|_2^2 - 2\left\|P_A^\perp P_B^\perp\boldsymbol{r}\right\|_2^2 \\
&\geq 1 - 2\left\|P_A^\perp P_B\right\|_2^2 - 2\left\|P_B^\perp\boldsymbol{r}\right\|_2^2 \\
&= 1 - 2d^2(\Gamma_A, \Gamma_B) - 2\Big(1 - \|P_B\boldsymbol{r}\|_2^2\Big),
\end{aligned}$$

where $d(\Gamma_A, \Gamma_B) = \|P_A^\perp P_B\|$ denotes the subspace distance (sine of the maximum subspace angle). Applying Lemmas D.9, D.8 to bound $d(\Gamma_A, \Gamma_B)$ and $\|P_B\boldsymbol{r}\|_2$, we have

$$\begin{aligned}
\left\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{\mathcal{S}_i}[\mu_j^\star + t\boldsymbol{r}]}\mathcal{S}_i}[\boldsymbol{r}]\right\|^2 &\geq 1 - 2(6\kappa_{\max}t)^2 - 2\Big(1 - \big(1 - \kappa_{\max}R\big)^2\Big) \\
&\geq 1 - 72\kappa_{\max}^2R^2 - 4\kappa_{\max}R \\
&> 1 - 70c^2 - 4c,
\end{aligned}$$

where we have used that $\kappa_{\max} \leq 1/\rho_{\min}$. For sufficiently small $c > 0$ (say, $c < \frac{1}{10}$), this is strictly larger than $\frac{1}{2}$, as claimed. ■

D.2.1 Differentiating curvature in terms of projections of normal vectors

Lemma D.5 *For any $\boldsymbol{\mu}, \boldsymbol{r} \in L^2(\mathbb{R}^2)$ such that $d(\boldsymbol{\mu} + t\boldsymbol{r}, \mathcal{S}_i) < \rho(\mathcal{S}_i)$, the squared distance function $d^2(\boldsymbol{\mu} + t\boldsymbol{r}, \mathcal{S}_i)$ is a differentiable function of t, and*

$$\frac{d}{dt}\Big\{d^2(\boldsymbol{\mu} + t\boldsymbol{r}, \mathcal{S}_i)\Big\} = 2\left\langle \boldsymbol{r}, \boldsymbol{\mu} + t\boldsymbol{r} - \boldsymbol{v}_j \circ \boldsymbol{\tau}_\star \right\rangle. \tag{D.23}$$

Proof We have

$$\begin{aligned}
d^2(\boldsymbol{\mu} + t\boldsymbol{r}, \mathcal{S}_i) &= \min_{\boldsymbol{\tau} \in \mathbb{T}} \|\boldsymbol{\mu} + t\boldsymbol{r} - \boldsymbol{v}_i \circ \boldsymbol{\tau}\|_2^2 \\
&= \|\boldsymbol{\mu} + t\boldsymbol{r}\|_2^2 + \psi(\boldsymbol{\mu} + t\boldsymbol{r}),
\end{aligned} \tag{D.24}$$

with

$$\psi(\boldsymbol{v}) = \min_{\boldsymbol{\tau} \in \mathbb{T}}\Big\{ -2\left\langle \boldsymbol{v}, \boldsymbol{v}_j \circ \boldsymbol{\tau} \right\rangle + \|\boldsymbol{v}_j \circ \boldsymbol{\tau}\|_2^2 \Big\}. \tag{D.25}$$

This is a pointwise minimum of linear functions of $\boldsymbol{v}$, and hence is concave. By Danskin's theorem, ψ is differentiable at any $\boldsymbol{v}$ for which the minimizing $\boldsymbol{\tau}_\star$ is unique, and

$$\nabla_{\boldsymbol{v}}\psi(\boldsymbol{v}) = -2\boldsymbol{v}_j \circ \boldsymbol{\tau}_\star. \tag{D.26}$$

Hence, $\min_{\boldsymbol{\tau} \in \mathbb{T}} \|\boldsymbol{v} - \boldsymbol{v}_j \circ \boldsymbol{\tau}\|_2^2 = \psi(\boldsymbol{v}) + \|\boldsymbol{v}\|_2^2$ is a differentiable function of $\boldsymbol{v}$, and its gradient is given by

$$\nabla_{\boldsymbol{v}}\Big\{\min_{\boldsymbol{\tau} \in \mathbb{T}} \|\boldsymbol{v} - \boldsymbol{v}_j \circ \boldsymbol{\tau}\|_2^2\Big\} = 2\boldsymbol{v} - 2\boldsymbol{v}_j \circ \boldsymbol{\tau}_\star. \tag{D.27}$$

In particular, when $d(\boldsymbol{\mu} + t\boldsymbol{r}, \mathcal{S}_i) < \rho(\mathcal{S}_i)$, the minimizing $\boldsymbol{\tau}_\star$ is unique, d^2 is differentiable, and (D.27) holds. Applying the chain rule, $d^2(\boldsymbol{\mu} + t\boldsymbol{r}, \mathcal{S}_i)$ is differentiable function of t, and

$$\frac{d}{dt}\Big\{d^2(\boldsymbol{\mu} + t\boldsymbol{r}, \mathcal{S}_i)\Big\} = 2\left\langle \boldsymbol{r}, \boldsymbol{\mu} + t\boldsymbol{r} - \boldsymbol{v}_j \circ \boldsymbol{\tau}_\star \right\rangle, \tag{D.28}$$

as claimed. ■

Lemma D.6 *There exists a positive numerical constant $c > 0$ such that if*

$$R = \max_{k,\ell} d(\mathcal{S}_k, \mathcal{S}_\ell) \le c\,\rho_{\min}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n), \tag{D.29}$$

then for any $t \in [0, 2R]$, and for any unit vector $\boldsymbol{r} \in \mathcal{N}_{\boldsymbol{\mu}_j^\star} S_{\boldsymbol{\mu}_j^\star}$, $\varphi(\boldsymbol{\mu}_j^\star + t\boldsymbol{r})$ is a twice differentiable function of t, and

$$\frac{d^2}{dt^2}\varphi(\boldsymbol{\mu}_j^\star + t\boldsymbol{r}) \ge 2\sum_{i=1}^n \boldsymbol{W}_{ij}\left(\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]} S_i}[\boldsymbol{r}]\|_2^2 - \frac{3\kappa_{\max} R}{1 - 3\kappa_{\max} R} \right). \tag{D.30}$$

Proof Recall that

$$\varphi(\boldsymbol{\mu}) = \sum_{i=1}^n \boldsymbol{W}_{ij} d^2(\boldsymbol{\mu}, \mathcal{S}_i). \tag{D.31}$$

Because

$$d(\boldsymbol{\mu}_j^\star + t\boldsymbol{r}, \mathcal{S}_i) \le d(\boldsymbol{\mu}_j^\star, \mathcal{S}_i) + t \le 3R. \tag{D.32}$$

When $R < \frac{1}{3}\rho_{\min}$, $d(\boldsymbol{\mu}_j^\star + t\boldsymbol{r}, \mathcal{S}_i) < \rho(\mathcal{S}_i)$. By Lemma D.5, $d^2(\boldsymbol{\mu} + t\boldsymbol{r}, \mathcal{S}_i)$ is a differentiable function of t, and

$$\frac{d}{dt}\left\{ d^2(\boldsymbol{\mu} + t\boldsymbol{r}, \mathcal{S}_i) \right\} = 2\left\langle \boldsymbol{r}, \boldsymbol{\mu} + t\boldsymbol{r} - \boldsymbol{v}_j \circ \boldsymbol{\tau}_\star \right\rangle. \tag{D.33}$$

Noting that

$$\boldsymbol{v}_j \circ \boldsymbol{\tau}_\star = \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}], \tag{D.34}$$

by Lemma C.2, $\boldsymbol{v}_j \circ \boldsymbol{\tau}_\star$ is a differentiable function of t. Thus $d^2(\boldsymbol{\mu} + t\boldsymbol{r}, \mathcal{S}_i)$ is twice differentiable, with

$$\frac{d^2}{dt^2}\varphi(\boldsymbol{\mu}_j^\star + t\boldsymbol{r}) = 2\sum_{i=1}^n \boldsymbol{W}_{ij}\left\langle \boldsymbol{r}, \boldsymbol{r} - \frac{d}{dt}\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}] \right\rangle. \tag{D.35}$$

By Lemma C.2, writing

$$\boldsymbol{\delta}_{i,t} = \boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}] \tag{D.36}$$

and

$$\Gamma_{i,t} = \mathcal{T}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]} S_i, \tag{D.37}$$

we have

$$\frac{d}{dt}\mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}] = \sum_{k=0}^\infty \mathcal{P}_{\Gamma_{i,t}}\left(\mathrm{I\!I}^*(\boldsymbol{\delta}_{i,t})\right)^k \mathcal{P}_{\Gamma_{i,t}}\boldsymbol{r}. \tag{D.38}$$

Note that

$$\begin{aligned} \|\boldsymbol{\delta}_{i,t}\| &= \|\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]\| \\ &\le \|\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star]\| \\ &\le \|\boldsymbol{\mu}_j^\star - \mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star]\| + t \\ &\le 3R. \end{aligned} \tag{D.39}$$

Combining with (D.35), and using Lemma C.3 to bound the norm of $\mathrm{I\!I}^*$, we obtain

$$\begin{aligned} \frac{d^2}{dt^2}\varphi(\boldsymbol{\mu}_j^\star + t\boldsymbol{r}) &= 2\sum_{i=1}^n \boldsymbol{W}_{ij}\left\langle \boldsymbol{r}, \boldsymbol{r} - \sum_{k=0}^\infty \left(\mathcal{P}_{\Gamma_{i,t}}\mathrm{I\!I}^*(\boldsymbol{\delta}_{i,t})\mathcal{P}_{\Gamma_{i,t}}\right)^k \boldsymbol{r} \right\rangle \\ &= 2\sum_{i=1}^n \boldsymbol{W}_{ij}\left(\|\mathcal{P}_{\Gamma_{i,t}^\perp}\boldsymbol{r}\|_2^2 - \sum_{k=1}^\infty \left\langle \mathcal{P}_{\Gamma_{i,t}}\boldsymbol{r}, \left(\mathrm{I\!I}^*(\boldsymbol{\delta}_{i,t})\right)\mathcal{P}_{\Gamma_{i,t}}\boldsymbol{r} \right\rangle \right) \\ &\ge 2\sum_{i=1}^n \boldsymbol{W}_{ij}\left(\|\mathcal{P}_{\Gamma_{i,t}^\perp}\boldsymbol{r}\|_2^2 - \sum_{k=1}^\infty \|\mathrm{I\!I}^*(\boldsymbol{\delta}_{i,t})\|^k \right) \\ &\ge 2\sum_{i=1}^n \boldsymbol{W}_{ij}\left(\|\mathcal{P}_{\Gamma_{i,t}^\perp}\boldsymbol{r}\|_2^2 - \frac{3\kappa_{\max} R}{1 - 3\kappa_{\max} R} \right), \end{aligned} \tag{D.40}$$

completing the proof. ■

D.2.2 Tangent and Normal Variation Across Manifolds

Under our assumptions, the manifolds $\mathcal{S}_i$ (i) are close together, and (ii) have large reach, and hence small curvature. In this section, we prove two lemmas which use properties (i)-(ii) to show that at nearby points on $\mathcal{S}_i$ and $\mathcal{S}_j$, the tangent spaces to $\mathcal{S}_i$ and $\mathcal{S}_j$ are close together, and similarly for the normal spaces.

Lemma D.7 *There exists a positive numerical constant $c > 0$ such that if*

$$R = \max_{k,\ell} d(\mathcal{S}_k, \mathcal{S}_\ell) < c\,\rho_{\min}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n), \tag{D.41}$$

for any $\boldsymbol{x} \in \mathcal{S}_i$ and any unit vector $\boldsymbol{v} \in \mathcal{T}_{\boldsymbol{x}}\mathcal{S}_i$,

$$\|\mathcal{P}_{\mathcal{T}_{\mathcal{P}_{\mathcal{S}_j}[\boldsymbol{x}]}\mathcal{S}_j}[\boldsymbol{v}]\|_2 \geq \Big(1 - \kappa_{\max} R\Big)^2. \tag{D.42}$$

Proof Let $\gamma(t)$ be a unit-speed geodesic in $\mathcal{S}_i$ with $\gamma(0) = \boldsymbol{x}$ and $\dot\gamma(0) = \boldsymbol{v}$. Then there exist transformations $\boldsymbol{\tau}_t \in \mathbb{T}$ such that $\gamma(t) = \boldsymbol{v}_i \circ \boldsymbol{\tau}_t$. For $\boldsymbol{v} \in L^2$, write

$$\psi(\boldsymbol{v}) = \min_{\boldsymbol{\tau} \in \mathbb{T}} -2\left\langle \boldsymbol{v}, \boldsymbol{v}_j \circ \boldsymbol{\tau} \right\rangle + \|\boldsymbol{v}_j \circ \boldsymbol{\tau}\|_2^2. \tag{D.43}$$

This is a pointwise minimum of affine functions of $\boldsymbol{v}$, and hence is concave. Hence, by Danskin's theorem, ψ is differentiable at any $\boldsymbol{v}$ for which the minimizing $\boldsymbol{\tau}$ is unique, and

$$\nabla_{\boldsymbol{v}} \psi(\boldsymbol{v}) = -2\boldsymbol{v}_j \circ \boldsymbol{\tau}_\star. \tag{D.44}$$

Hence, $\min_{\boldsymbol{\tau} \in \mathbb{T}} \|\boldsymbol{v} - \boldsymbol{v}_j \circ \boldsymbol{\tau}\|_2^2 = \psi(\boldsymbol{v}) + \|\boldsymbol{v}\|_2^2$ is a differentiable function of $\boldsymbol{v}$, and its gradient is given by

$$\nabla_{\boldsymbol{v}} \Big\{ \min_{\boldsymbol{\tau} \in \mathbb{T}} \|\boldsymbol{v} - \boldsymbol{v}_j \circ \boldsymbol{\tau}\|_2^2 \Big\} = 2\boldsymbol{v} - 2\boldsymbol{v}_j \circ \boldsymbol{\tau}_\star. \tag{D.45}$$

Applying the chain rule, we have that

$$\frac{d}{dt}\Big\{ \min_{\boldsymbol{\tau}} \|\boldsymbol{v}_i \circ \boldsymbol{\tau}_t - \boldsymbol{v}_j \circ \boldsymbol{\tau}\|_2^2 \Big\} = 2\left\langle \boldsymbol{v}_i \circ \boldsymbol{\tau}_t - \boldsymbol{v}_j \circ \boldsymbol{\tau}^\star(t), \frac{d}{dt}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t \right\rangle \tag{D.46}$$

where $\boldsymbol{\tau}^\star(t)$ is the optimal transformation for the value of $\boldsymbol{v} = \boldsymbol{v}_i \circ \boldsymbol{\tau}_t$, which is unique when $d(\boldsymbol{v}_i \circ \boldsymbol{\tau}_t, \mathcal{S}_j) < \rho(\mathcal{S}_j)$, which is satisfied for $c < 1$. By Lemma C.2, $\boldsymbol{v}_j \circ \boldsymbol{\tau}^\star(t)$ is a differentiable function of t, and so the right hand side of (D.46) is a differentiable function of t. This implies that

$$d^2(\boldsymbol{v}_i \circ \boldsymbol{\tau}_t, \mathcal{S}_j) = \min_{\boldsymbol{\tau}} \|\boldsymbol{v}_i \circ \boldsymbol{\tau}_t - \boldsymbol{v}_j \circ \boldsymbol{\tau}\|_2^2$$

is twice differentiable, with

$$\begin{aligned} \frac{d^2}{dt^2}\Big\{ d^2(\boldsymbol{v}_i \circ \boldsymbol{\tau}_t, \mathcal{S}_j) \Big\} = 2\Bigg(&\left\langle \boldsymbol{v}_i \circ \boldsymbol{\tau}_t - \boldsymbol{v}_j \circ \boldsymbol{\tau}^\star(t), \frac{d^2}{dt^2}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t \right\rangle + \left\langle \frac{d}{dt}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t, \frac{d}{dt}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t \right\rangle \\ &- \left\langle \frac{d}{dt}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t, \frac{d}{dt}\boldsymbol{v}_j \circ \boldsymbol{\tau}^\star(t) \right\rangle \Bigg) \end{aligned} \tag{D.47}$$

On the other hand, because for any $\boldsymbol{\sigma} \in \mathbb{T}$,

$$\|\boldsymbol{v}_i \circ \boldsymbol{\sigma} - \boldsymbol{v}_j \circ \boldsymbol{\tau} \circ \boldsymbol{\sigma}\| = \|\boldsymbol{v}_i - \boldsymbol{v}_j \circ \boldsymbol{\tau}\|, \tag{D.48}$$

we have

$$\begin{aligned} d(\boldsymbol{v}_i \circ \boldsymbol{\sigma}, \mathcal{S}_j) &= \min_{\boldsymbol{\tau}} \|\boldsymbol{v}_i \circ \boldsymbol{\sigma} - \boldsymbol{v}_j \circ \boldsymbol{\tau}\| \\ &= \min_{\boldsymbol{\tau}} \|\boldsymbol{v}_i \circ \boldsymbol{\sigma} - \boldsymbol{v}_j \circ \boldsymbol{\tau} \circ \boldsymbol{\sigma}\| \\ &= \min_{\boldsymbol{\tau}} \|\boldsymbol{v}_i - \boldsymbol{v}_j \circ \boldsymbol{\tau}\| \\ &= d(\boldsymbol{v}_i, \mathcal{S}_j). \end{aligned} \tag{D.49}$$

This implies that $d^2(\boldsymbol{v}_i \circ \boldsymbol{\tau}_t, \mathcal{S}_j)$ is a constant function of t, and so

$$\frac{d^2}{dt^2}\Big\{ d^2(\boldsymbol{v}_i \circ \boldsymbol{\tau}_t, \mathcal{S}_j) \Big\} = 0. \tag{D.50}$$

Setting (D.47) equal to zero, we have that

$$\begin{aligned}\left\langle \frac{d}{dt}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t, \frac{d}{dt}\boldsymbol{v}_j \circ \boldsymbol{\tau}^\star(t) \right\rangle &= \left\langle \frac{d}{dt}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t, \frac{d}{dt}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t \right\rangle + \left\langle \boldsymbol{v}_i \circ \boldsymbol{\tau}_t - \boldsymbol{v}_j \circ \boldsymbol{\tau}^\star(t), \frac{d}{dt}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t \right\rangle \\ &\geq 1 - \kappa_{\max} R. \end{aligned} \tag{D.51}$$

where we have used that $\boldsymbol{v}_i \circ \boldsymbol{\tau}_t$ is a unit speed geodesic, and hence $\| \frac{d}{dt}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t \| = 1$ and $\| \frac{d^2}{dt^2}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t \|_2 \leq \kappa$, and used Cauchy-Schwarz to bound the second term.

For simplicity of notation, let $\Gamma_t = \mathcal{T}_{\mathcal{P}_{S_j}[\boldsymbol{v}_i \circ \boldsymbol{\tau}_t]}$ and $\boldsymbol{\delta}_t = \boldsymbol{v}_i \circ \boldsymbol{\tau}_t - \boldsymbol{v}_j \circ \boldsymbol{\tau}^\star(t)$. By Lemma C.2,

$$\frac{d}{dt}\boldsymbol{v}_j \circ \boldsymbol{\tau}^\star(t) = \sum_{k=0}^{\infty} \mathcal{P}_{\Gamma_t}(\mathrm{I\!I}^*(\boldsymbol{\delta}_t))^k \mathcal{P}_{\Gamma_t}[\dot{\boldsymbol{\gamma}}(t)], \tag{D.52}$$

and so

$$\begin{aligned}\left\langle \frac{d}{dt}\boldsymbol{v}_i \circ \boldsymbol{\tau}_t, \frac{d}{dt}\boldsymbol{v}_j \circ \boldsymbol{\tau}^\star(t) \right\rangle &= \sum_{k=0}^{\infty} \dot{\boldsymbol{\gamma}}(t)^* \mathcal{P}_{\Gamma_t}(\mathrm{I\!I}^*(\boldsymbol{\delta}_t))^k \mathcal{P}_{\Gamma_t} \dot{\boldsymbol{\gamma}}(t) \\ &\leq \|\mathcal{P}_{\Gamma_t}\dot{\boldsymbol{\gamma}}(t)\|_2^2 \sum_{k=0}^{\infty} \|\mathrm{I\!I}^*(\boldsymbol{\delta}_t)\|^k \\ &\leq \|\mathcal{P}_{\Gamma_t}\dot{\boldsymbol{\gamma}}(t)\|_2^2 \frac{1}{1-\kappa_{\max}R}, \end{aligned} \tag{D.53}$$

where we have used that $\|\mathrm{I\!I}(\boldsymbol{\delta}_t)\| \leq \kappa_{\max}\|\boldsymbol{\delta}_t\|$. Comparing to (D.51), we have

$$\|\mathcal{P}_{\Gamma_t}\dot{\boldsymbol{\gamma}}(t)\|_2^2 \geq \Big(1-\kappa_{\max}R\Big)^2, \tag{D.54}$$

as claimed. ■

Lemma D.8 *There exists a positive numerical constant $c > 0$ such that if*

$$R = \max_{k,\ell} d(\mathcal{S}_k, \mathcal{S}_\ell) < c\,\rho_{\min}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n), \tag{D.55}$$

for any $\boldsymbol{x} \in \mathcal{S}_i$ and any unit vector $\boldsymbol{v} \in \mathcal{N}_{\boldsymbol{x}}\mathcal{S}_i$,

$$\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{S_j}[\boldsymbol{x}]}S_j}[\boldsymbol{v}]\|_2 \geq \Big(1-\kappa_{\max}R\Big)^2. \tag{D.56}$$

Proof Let $\Gamma_i = \mathcal{T}_{\boldsymbol{x}}S_i$ and $\Gamma_j = \mathcal{T}_{\mathcal{P}_{S_j}[\boldsymbol{x}]}\mathcal{S}_j$. From Lemma D.7, we have

$$\begin{aligned} d^2(\Gamma_i, \Gamma_j) &= \left\|(I - P_{\Gamma_j})P_{\Gamma_i}\right\|_2^2 \\ &= \max_{\boldsymbol{r}\in\Gamma_i, \|\boldsymbol{r}\|_2 = 1} \left\|(I-P_{\Gamma_j})\boldsymbol{r}\right\|_2^2 \\ &= \max_{\boldsymbol{r}\in\Gamma_i, \|\boldsymbol{r}\|_2 = 1} 1 - \left\|P_{\Gamma_j}\boldsymbol{r}\right\|_2^2 \\ &\leq 1-(1-\kappa R)^2. \end{aligned} \tag{D.57}$$

We further have

$$\begin{aligned} d^2(\Gamma_i^\perp, \Gamma_j^\perp) &= d^2(\Gamma_i, \Gamma_j) \\ &\leq 1-(1-\kappa R)^2, \end{aligned} \tag{D.58}$$

and so

$$\begin{aligned} \min_{\boldsymbol{v}\in\Gamma_i^\perp,\, \|\boldsymbol{v}\|=1} \|P_{\Gamma_j^\perp}\boldsymbol{v}\|_2^2 &= 1 - \max_{\boldsymbol{v}\in\Gamma_i^\perp,\, \|\boldsymbol{v}\|=1} \|(I-P_{\Gamma_j^\perp})\boldsymbol{v}\|_2^2 \\ &= 1 - d^2(\Gamma_i^\perp, \Gamma_j^\perp) \\ &\geq 1 - \left(1 - \Big(1-\kappa R\Big)^2\right) \\ &= \Big(1-\kappa R\Big)^2. \end{aligned} \tag{D.59}$$

Since $\Gamma_i^\perp = \mathcal{N}_{\boldsymbol{x}}\mathcal{S}_i$, and $\Gamma_j^\perp = \mathcal{N}_{\mathcal{P}_{S_j}[\boldsymbol{x}]}\mathcal{S}_j$, this establishes the claim. ■

D.3 Normal Variation Along a Manifold

In this section, we prove that under our assumptions, for nearby points on $\mathcal{S}_i$, the normal spaces to $\mathcal{S}_i$ are close together. More precisely, we control the variation of the normal spaces to $\mathcal{S}_i$ between two points, which are projections of nearby points $\boldsymbol{\mu}$ and $\boldsymbol{\mu} + t\boldsymbol{r}$ in the ambient space L^2.

Lemma D.9 *Consider $\boldsymbol{\mu} \in L^2$ with $d(\boldsymbol{\mu}, \mathcal{S}_i) \le R$, and $t \in [0, 2R]$. Let*

$$\Gamma_A = \mathcal{N}_{\mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}]}\mathcal{S}_i, \tag{D.60}$$
$$\Gamma_B = \mathcal{N}_{\mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}+t\boldsymbol{r}]}\mathcal{S}_i. \tag{D.61}$$

There exists a numerical constant $c > 0$ such that if

$$R = \max_{k,\ell} d(\mathcal{S}_k, \mathcal{S}_\ell) \le c\,\rho_{\min}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n), \tag{D.62}$$

then

$$d(\Gamma_A, \Gamma_B) \le 6\kappa_{\max} t. \tag{D.63}$$

Proof Let P_A, P_B be the projection matrices onto Γ_A, Γ_B respectively, and let $P_A^\perp, P_B^\perp$ be the projection matrices onto the orthogonal complements $\Gamma_A^\perp$ and $\Gamma_B^\perp$. Then

$$\begin{aligned} d(\Gamma_A, \Gamma_B) &= d(\Gamma_B, \Gamma_A) \\ &= \|(I - P_A)P_B\|_2 \\ &= \|P_A^\perp P_B\|_2 \\ &= \|P_B P_A^\perp\|_2 \\ &= \|(I - P_B^\perp)P_A^\perp\|_2 \\ &= d(\Gamma_B^\perp, \Gamma_A^\perp) \\ &= \max_{\boldsymbol{x}} \|(I - P_B^\perp)P_A^\perp \boldsymbol{x}\|_2 \\ &= \max_{\boldsymbol{v}\in\mathcal{S}_A^\perp, \|\boldsymbol{v}\|_2=1} \|(I - P_B^\perp)\boldsymbol{v}\|_2 \\ &= \max_{\boldsymbol{v}\in\mathcal{S}_A^\perp, \|\boldsymbol{v}\|_2=1} \min_{\boldsymbol{u}\in\mathcal{S}_B^\perp} \|\boldsymbol{v} - \boldsymbol{u}\|_2 \end{aligned} \tag{D.64}$$

Let $\boldsymbol{a} = \mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}]$, and $\boldsymbol{b} = \mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu} + t\boldsymbol{r}]$ and let $\gamma(s)$ be a minimum length unit speed geodesic on $\mathcal{S}_i$, satisfying $\gamma(0) = \boldsymbol{a}$ and $\gamma(s_b) = \boldsymbol{b}$.

Notice that $\Gamma_A^\perp$ is the tangent space to Γ_A at $\boldsymbol{a}$. For any $\boldsymbol{v} \in \Gamma_A^\perp$, we generate a parallel vector field $\boldsymbol{f_v}(s)$ along $\gamma(s)$ with $\boldsymbol{f_v}(0) = \boldsymbol{v}$ and $\boldsymbol{f_v}(s_b) \in \Gamma_B^\perp$. Let $\Pi_{\boldsymbol{b},\boldsymbol{a}}[\boldsymbol{v}] = \boldsymbol{v_v}(s_b) \in \Gamma_B^\perp$ denote this parallel transport operator.

We control $\|\Pi_{\boldsymbol{b},\boldsymbol{a}}[\boldsymbol{v}] - \boldsymbol{v}\|$ as follows. The parallel vector field $\boldsymbol{v}(s)$ satisfies

$$\frac{d\boldsymbol{v}}{ds} = \mathrm{I\!I}(\dot{\boldsymbol{\gamma}}(s), \boldsymbol{v}(s)), \tag{D.65}$$

whence

$$\begin{aligned} \|\Pi_{\boldsymbol{b},\boldsymbol{a}}\boldsymbol{v} - \boldsymbol{v}\| &= \|\boldsymbol{v}(s_b) - \boldsymbol{v}(0)\|_2 \\ &= \left\| \int_0^{s_b} \mathrm{I\!I}(\dot{\boldsymbol{\gamma}}(s), \boldsymbol{v}(s)) ds \right\|_2 \\ &\le \int_0^{s_b} \|\mathrm{I\!I}(\dot{\boldsymbol{\gamma}}(s), \boldsymbol{v}(s)\|_2 ds \\ &\le 3\kappa s_b \end{aligned} \tag{D.66}$$

where in the final line we have used Lemma D.10.

From (D.64), we have

$$\begin{aligned} d(\Gamma_A, \Gamma_B) &= \max_{\boldsymbol{v} \in \mathcal{S}_A^{\perp}, \|\boldsymbol{v}\|_2=1} \min_{\boldsymbol{u} \in \mathcal{S}_B^{\perp}} \|\boldsymbol{v} - \boldsymbol{u}\| \\ &\leq \max_{\boldsymbol{v} \in \mathcal{S}_A^{\perp}, \|\boldsymbol{v}\|_2=1} \|\boldsymbol{v} - \Pi_{\boldsymbol{b},\boldsymbol{a}} \boldsymbol{v}\| \\ &\leq 3\kappa s_b. \end{aligned} \tag{D.67}$$

We are left to bound $s_b = d_{\mathcal{S}_i}(\boldsymbol{a}, \boldsymbol{b})$. This is the infimum of the lengths of all differentiable curves joining $\boldsymbol{a}$, $\boldsymbol{b}$. Taking a particular curve

$$\xi(u) = \mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star + u\boldsymbol{r}], \tag{D.68}$$

we have

$$d_{\mathcal{S}_i}(\boldsymbol{a}, \boldsymbol{b}) \leq \int_0^t \left\| \frac{d}{du} \xi(u) \right\| du. \tag{D.69}$$

We bound the integrand using Lemma C.2. We begin by noting that

$$\begin{aligned} \|\boldsymbol{\mu}_j^\star + u\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + u\boldsymbol{r}]\|_2 &\leq \|\boldsymbol{\mu}_j^\star + u\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star]\|_2 \\ &\leq \|\boldsymbol{\mu}_j^\star - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star]\|_2 + u \\ &\leq 3R. \end{aligned} \tag{D.70}$$

When $R < \frac{1}{3}\rho_{\min}$, $\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + u\boldsymbol{r}]$ is a differentiable function of u and by Lemma C.2, we have

$$\frac{d}{du} \mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star + u\boldsymbol{r}] = \sum_{k=0}^{\infty} \mathrm{I\!I}^*(\boldsymbol{\delta}_u) \mathcal{P}_{\mathcal{T}_{\xi(u)}\mathcal{S}_i} \boldsymbol{r}, \tag{D.71}$$

with

$$\boldsymbol{\delta}_u = \boldsymbol{\mu}_j^\star + u\boldsymbol{r} - \mathcal{P}_{\mathcal{S}_i}[\boldsymbol{\mu}_j^\star + u\boldsymbol{r}]. \tag{D.72}$$

We have $\|\boldsymbol{\delta}_u\| \leq 3R$, and so $\|\mathrm{I\!I}^*(\boldsymbol{\delta}_u)\| \leq 3R\kappa_{\max}$, and so

$$\left\| \frac{d}{du} \xi(u) \right\| \leq \sum_{k=0}^{\infty} \|\mathrm{I\!I}^*(\boldsymbol{\delta}_u)\|^k \|\mathcal{P}_{\mathcal{T}_{\xi(u)}\mathcal{S}_i} \boldsymbol{r}\| \leq \frac{\|\mathcal{P}_{\mathcal{T}_{\xi(u)}\mathcal{S}_i} \boldsymbol{r}\|}{1 - \kappa_{\max}\|\boldsymbol{\delta}_u\|} \leq \frac{1}{1 - 3\kappa_{\max} R} \leq 2, \tag{D.73}$$

where by $\kappa_{\max} < 1/\rho_{\min}$ the final bound holds provided $c < 1/6$. Plugging in to (D.69), we obtain $d_{\mathcal{S}_i}(\boldsymbol{a}, \boldsymbol{b}) \leq 2t$; combining with (D.66) gives the claimed bound. ■

The proof of Lemma D.9 relies on the following lemma, which controls the second fundamental form (and hence controls the rate of change of parallel vector fields):

Lemma D.10 *Let $\boldsymbol{x} \in \mathcal{S}$, where $\mathcal{S}$ is an embedded submanifold of $\mathbb{R}^D$. Let $\mathrm{I\!I}(\boldsymbol{u}, \boldsymbol{v}) : \mathcal{T}_{\boldsymbol{x}} S \times \mathcal{T}_{\boldsymbol{x}} S \to \mathcal{N}_{\boldsymbol{x}} S$ denote the second fundamental form of $\mathcal{S}$ at $\boldsymbol{x}$. Then*

$$\max_{\substack{\boldsymbol{u}, \boldsymbol{v} \in \mathcal{T}_{\boldsymbol{x}} \mathcal{S} \\ \|\boldsymbol{u}\|_2 = 1, \|_2 \boldsymbol{v}\|_2 = 1}} \|\mathrm{I\!I}(\boldsymbol{u}, \boldsymbol{v})\|_2 \leq 3\kappa, \tag{D.74}$$

where κ is the extrinsic geodesic curvature of $\mathcal{S}$.

Proof The tangent space of the manifold is d dimensional. Hence the normal space is $D - d$ dimensional.Therefore, $\mathrm{I\!I}$ is a $D - d$ dimensional vector extrinsically. Since the second fundamental form is symmetric and bilinear, the i^{th} coordinate has the form $\mathrm{I\!I}_i(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^T \boldsymbol{\Phi}_i \boldsymbol{v}$ for some symmetric $d \times d$ matrix $\boldsymbol{\Phi}_i$. Then, we have the following:

$$\begin{aligned} &\boldsymbol{u}^T \boldsymbol{\Phi}_i \boldsymbol{v} = \tfrac{1}{2}(\boldsymbol{u} + \boldsymbol{v})^T \boldsymbol{\Phi}_i (\boldsymbol{u} + \boldsymbol{v}) - \tfrac{1}{2}\boldsymbol{u}^T \boldsymbol{\Phi}_i \boldsymbol{u} - \tfrac{1}{2}\boldsymbol{v}^T \boldsymbol{\Phi}_i \boldsymbol{v} \\ \Rightarrow &|\boldsymbol{u}^T \boldsymbol{\Phi}_i \boldsymbol{v}| \leq |\tfrac{1}{2}(\boldsymbol{u} + \boldsymbol{v})^T \boldsymbol{\Phi}_i (\boldsymbol{u} + \boldsymbol{v})| + |\tfrac{1}{2}\boldsymbol{u}^T \boldsymbol{\Phi}_i \boldsymbol{u}| + |\tfrac{1}{2}\boldsymbol{v}^T \boldsymbol{\Phi}_i \boldsymbol{v}| \\ \Rightarrow &|\boldsymbol{u}^T \boldsymbol{\Phi}_i \boldsymbol{v}|^2 \leq 3(|\tfrac{1}{2}(\boldsymbol{u} + \boldsymbol{v})^T \boldsymbol{\Phi}_i (\boldsymbol{u} + \boldsymbol{v})|^2 + |\tfrac{1}{2}\boldsymbol{u}^T \boldsymbol{\Phi}_i \boldsymbol{u}|^2 + |\tfrac{1}{2}\boldsymbol{v}^T \boldsymbol{\Phi}_i \boldsymbol{v}|^2). \end{aligned}$$

Using $(a+b+c)^2 \leq 3a^2+3b^2+3c^2$ which follows from convexity of the square. Summing over i, we obtain:

$$\begin{aligned}\|\mathrm{I\!I}(\boldsymbol{u},\boldsymbol{v})\|_2^2 = \sum_i |\boldsymbol{u}^T\boldsymbol{\Phi}_i\boldsymbol{v}|^2 &\leq \tfrac{3}{4}\left(\|\mathrm{I\!I}(\boldsymbol{u}+\boldsymbol{v},\boldsymbol{u}+\boldsymbol{v})\|_2^2 + \|\mathrm{I\!I}(\boldsymbol{u},\boldsymbol{u})\|_2^2 + \|\mathrm{I\!I}(\boldsymbol{v},\boldsymbol{v})\|_2^2\right)\\
&\leq \tfrac{9}{4}\max\left\{\|\mathrm{I\!I}(\boldsymbol{u}+\boldsymbol{v},\boldsymbol{u}+\boldsymbol{v})\|_2^2, \|\mathrm{I\!I}(\boldsymbol{u},\boldsymbol{u})\|_2^2, \|\mathrm{I\!I}(\boldsymbol{v},\boldsymbol{v})\|_2^2\right\}\\
&= \tfrac{9}{4}\max\left\{\|\boldsymbol{u}+\boldsymbol{v}\|_2^4\left\|\mathrm{I\!I}\left(\frac{\boldsymbol{u}+\boldsymbol{v}}{\|\boldsymbol{u}+\boldsymbol{v}\|_2},\frac{\boldsymbol{u}+\boldsymbol{v}}{\|\boldsymbol{u}+\boldsymbol{v}\|_2}\right)\right\|_2^2, \|\mathrm{I\!I}(\boldsymbol{u},\boldsymbol{u})\|_2^2, \|\mathrm{I\!I}(\boldsymbol{v},\boldsymbol{v})\|_2^2\right\}\end{aligned}$$

Since $\|\mathrm{I\!I}(\boldsymbol{u},\boldsymbol{v})\|_2^2 = \|\mathrm{I\!I}(\boldsymbol{u},-\boldsymbol{v})\|_2^2$, we can choose $\boldsymbol{u},\boldsymbol{v}$ such that $\langle\boldsymbol{u},\boldsymbol{v}\rangle \leq 0$ to maximize the norm of the second fundamental form. Therefore, $\|\boldsymbol{u}+\boldsymbol{v}\|_2^2 = 2+2\langle\boldsymbol{u},\boldsymbol{v}\rangle \leq 2 \Rightarrow \|\boldsymbol{u}+\boldsymbol{v}\|_2 \leq \sqrt{2}$. Using this in the equation above,

$$\|\mathrm{I\!I}(\boldsymbol{u},\boldsymbol{v})\|_2^2 \leq \tfrac{9}{4}\max\{4\kappa^2,\kappa^2,\kappa^2\}$$

we complete the proof. ∎

E Proof of Convergence

E.1 Proof of the clustering theorem

Theorem E.1 *Let $\boldsymbol{v}_1^{(p)},\ldots,\boldsymbol{v}_n^{(p)}$ denote the features produced by the p-th iteration of invariant attendion, and*

$$\mathcal{S}_j^{(p)} = \left\{\boldsymbol{v}_j^{(p)}\circ\boldsymbol{\tau} \mid \boldsymbol{\tau}\in\mathbb{T}\right\} \tag{E.1}$$

the corresponding transformation manifolds. Let

$$R^{(p)} = \max_{m,l} d\Big(\mathcal{S}_m^{(p)},\mathcal{S}_l^{(p)}\Big).$$

There exist positive constants c, c', ε such that if

$$R^{(p)} < c\,\rho_{min}(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n)$$

and

$$\beta < c'/\big(R^{(p)}\big)^2.$$

Then

$$d\Big(\mathcal{S}_j^{(p+1)},\mathcal{S}_k^{(p+1)}\Big) \leq (1-\varepsilon)d\Big(\mathcal{S}_j^{(p)},\mathcal{S}_k^{(p)}\Big). \tag{E.2}$$

Whenever

$$R^{(0)} < c\rho_{\min}(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n), \text{ and } \beta < c'/\big(R^{(0)}\big)^2, \tag{E.3}$$

the conditions of the theorem hold for all iterations p, and so $\max_{j,k} d\left(\mathcal{S}_j^{(p)},\mathcal{S}_k^{(p)}\right)$ converges to zero at a linear rate:

$$\max_{j,k} d\left(\mathcal{S}_j^{(p)},\mathcal{S}_k^{(p)}\right) \leq (1-\varepsilon)^p \max_{j,k} d\left(\mathcal{S}_j^{(0)},\mathcal{S}_k^{(0)}\right). \tag{E.4}$$

Proof Sketch Let $\boldsymbol{W}_{:,k}$ denote the invariant weights associated with the image $\boldsymbol{v}_k^{(p)}$, and $\boldsymbol{W}_{:,j}$ the corresponding weights for image $\boldsymbol{v}_j^{(p)}$, at iteration p. Also define:

$$\begin{aligned}\psi(\boldsymbol{\mu}) &= \min_{\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n\in\mathbb{T}}\left\{\sum_{i=1}^n \boldsymbol{W}_{ik}\|\boldsymbol{v}_i^{(p)}\circ\boldsymbol{\tau}_i-\boldsymbol{\mu}\|_2^2\right\},\\
\varphi(\boldsymbol{\mu}) &= \min_{\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n\in\mathbb{T}}\left\{\sum_{i=1}^n \boldsymbol{W}_{ij}\|\boldsymbol{v}_i^{(p)}\circ\boldsymbol{\tau}_i-\boldsymbol{\mu}\|_2^2\right\}.\end{aligned}$$

The next-iteration features $\boldsymbol{v}_k^{(p+1)}$ and $\boldsymbol{v}_j^{(p+1)}$ are minimizers of these functions:

$$\boldsymbol{v}_k^{(p+1)} \in \arg\min_{\boldsymbol{\mu}} \psi(\boldsymbol{\mu})$$
$$\boldsymbol{v}_j^{(p+1)} \in \arg\min_{\boldsymbol{\mu}} \varphi(\boldsymbol{\mu}),$$

and the next-iteration distance is the distance between the transformation manifolds $\mathcal{S}_k^{(p+1)}$ and $\mathcal{S}_j^{(p+1)}$ generated by these features. We use Lemma E.2 to upper bound

$$d\Big(\mathcal{S}_k^{(p+1)}, \mathcal{S}_j^{(p+1)}\Big)$$

in terms of two geometric quantities:

1. Positive curvature of the invariant mean objective function $\varphi(\boldsymbol{\mu})$, quantified through a lower bound λ on the second derivative of φ in directions normal to the $\mathcal{S}_i^{(p)}$.
2. An upper bound ϵ on the gradient of the objective function $\nabla\varphi(\boldsymbol{\mu})$ at the optimizer $\boldsymbol{v}_k^{(p+1)}$ of the *other* objective $\psi(\boldsymbol{\mu})$.

Comparing these quantities allows us to bound the distance between $\boldsymbol{v}_k^{(p+1)}$ and the closest transformation $\boldsymbol{v}_j^{(p+1)} \circ \boldsymbol{\tau}$ of $\boldsymbol{v}_j^{(p+1)}$.

We use Theorem D.3 to quantify λ. For ϵ, we first use Lemma E.3 bound ϵ in terms of the difference between the weights associated with the two objectives $\varphi(\boldsymbol{\mu})$ and $\psi(\boldsymbol{\mu})$. We then use Lemma E.5 to bound this difference in weights in terms of the distances $d(\mathcal{S}_i^{(p)}, \mathcal{S}_j^{(p)})$. With these quantities, we prove the theorem. A formal proof follows below.

Proof For $\epsilon > 0, \lambda > 0$, if $\forall \boldsymbol{\mu} \in \mathcal{I}, \|\nabla\varphi(\boldsymbol{\mu}) - \nabla\psi(\boldsymbol{\mu})\|_2 \leq \epsilon$ and $\frac{d^2}{dt^2}\psi(\boldsymbol{\mu} + t\hat{\boldsymbol{v}})\big|_{t=0} \geq \lambda$ where $\hat{\boldsymbol{v}} \in \mathcal{N}_{\boldsymbol{\mu}} S_{\boldsymbol{\mu}}$, using lemma E.2, given the quantities ϵ, λ,

$$d(S_j^+, S_k^+) \leq \frac{2\epsilon}{\lambda}$$

Using lemma E.3 we can quantify ϵ. For $\boldsymbol{\mu} \in \mathcal{I}$,

$$\|\nabla\varphi(\boldsymbol{\mu}) - \nabla\psi(\boldsymbol{\mu})\|_2 \leq 2nR\Delta W$$

where $\Delta W = \max_i |W_{ij} - W_{ik}|$.
Using lemma E.5, we see that

$$\Delta W \leq 2\frac{\beta R}{n} e^{\beta R^2}(1 + e^{\beta R^2}) d(S_j, S_k)$$

Therefore,

$$\begin{aligned}
\|\nabla\varphi(\boldsymbol{\mu}) - \nabla\psi(\boldsymbol{\mu})\|_2 &\leq 2nR \times 2\frac{\beta R}{n} e^{\beta R^2}(1 + e^{\beta R^2}) d(S_j, S_k) \\
&= 4\beta R^2 e^{\beta R^2}(1 + e^{\beta R^2}) d(S_j, S_k)
\end{aligned}$$

Thus, we use $\epsilon = 4\beta R^2 e^{\beta R^2}(1 + e^{\beta R^2}) d(S_j, S_k)$ in lemma E.2.

Let $\boldsymbol{\tau}_j^\star, \boldsymbol{\tau}_k^\star$ be a pair of optimal transformations to the problem in $d(S_j^+, S_k^+)$.Also denote $\boldsymbol{v} = \boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star - \boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star$. Define $\bar{\boldsymbol{\mu}} = a\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star + (1-a)\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star, 0 \leq a \leq 1$. Note that $\bar{\boldsymbol{\mu}} \in \mathcal{I}$.

For quantifying λ in lemma E.2,we first note that $\boldsymbol{v} \in \mathcal{N}_{\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star} S_{\boldsymbol{\mu}_k^\star}$

We also note that $\bar{\boldsymbol{\mu}} = a\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star + (1-a)\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star = \boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star + a\boldsymbol{v}$. Therefore,

$$\boldsymbol{v}^T \nabla^2 \psi(\bar{\boldsymbol{\mu}}) \boldsymbol{v} = \|\boldsymbol{v}\|_2^2 \frac{d^2}{dt^2} \psi(\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star + (a\|\boldsymbol{v}\|_2 + t)\hat{\boldsymbol{v}})\big|_{t=0}$$

Since $a\|\boldsymbol{v}\|_2 \leq R$, choosing a sufficiently small $c < 1$ in the positive curvature result from Theorem D.3 we see that $\lambda = \frac{1}{3}$, thus completing the proof.

■

Lemma E.2 *Let*

$$\psi(\boldsymbol{\mu}) = \min_{\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n \in \mathbb{T}} \left\{ \sum_{i=1}^{n} \boldsymbol{W}_{ik} \left\| \boldsymbol{v}_i^{(p)} \circ \boldsymbol{\tau}_i - \boldsymbol{\mu} \right\|^2 \right\}$$
$$\varphi(\boldsymbol{\mu}) = \min_{\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n \in \mathbb{T}} \left\{ \sum_{i=1}^{n} \boldsymbol{W}_{ij} \left\| \boldsymbol{v}_i^{(p)} \circ \boldsymbol{\tau}_i - \boldsymbol{\mu} \right\|^2 \right\}$$

and $\boldsymbol{\mu}_j^\star = \arg\min_{\boldsymbol{\mu}} \varphi(\boldsymbol{\mu})$, $\boldsymbol{\mu}_k^\star = \arg\min_{\boldsymbol{\mu}} \psi(\boldsymbol{\mu})$.
If $\forall \boldsymbol{\mu} \in \mathcal{I}, \|\nabla\varphi(\boldsymbol{\mu}) - \nabla\psi(\boldsymbol{\mu})\|_2 \leq \epsilon$, $\frac{d^2}{dt^2}\psi(\boldsymbol{\mu} + t\hat{\boldsymbol{v}})\big|_{t=0} \geq \lambda$ *where* $\hat{\boldsymbol{v}} \in \mathcal{N}_{\boldsymbol{\mu}} S_{\boldsymbol{\mu}}$, *then*

$$\min_{\boldsymbol{\tau}_j,\boldsymbol{\tau}_k} \|\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j - \boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k\|_2 \leq \frac{2\epsilon}{\lambda} \tag{E.5}$$

Proof Let $\boldsymbol{\tau}_j^\star, \boldsymbol{\tau}_k^\star$ be a pair of optimal transformations to the problem in $d(S_j^+, S_k^+)$. Also denote $\boldsymbol{v} = \boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star - \boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star$. Define $\bar{\boldsymbol{\mu}} = a\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star + (1-a)\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star$ for some $0 \leq a \leq 1$. Note that $\bar{\boldsymbol{\mu}} \in \mathcal{I}$.
Note that since $\boldsymbol{\mu}_j^\star$ is optimal to $\varphi(\boldsymbol{\mu})$, any solution of the form $\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}, \boldsymbol{\tau}$ is also a solution. Therefore, $\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star$ is optimal solution to $\varphi(\boldsymbol{\mu})$. The same argument holds for $\boldsymbol{\mu}_k^\star$ and $\psi(\boldsymbol{\mu})$
Using the Taylor series, we see that

$$\begin{aligned}
\psi(\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star) &= \psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star) + \left\langle \nabla\psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star), \boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star - \mu_j^\star \circ \boldsymbol{\tau}_j^\star \right\rangle \\
&\quad + \frac{1}{2}(\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star - \boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_k^\star)^T \nabla^2\psi(\bar{\boldsymbol{\mu}}) \\
&\quad = \psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star) - \underbrace{\left\langle \nabla\psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star), \boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star - \mu_k^\star \circ \boldsymbol{\tau}_k^\star \right\rangle}_{T1} \\
&\quad + \frac{1}{2}\underbrace{(\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star - \boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star)^T \nabla^2\psi(\bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star - \boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star)}_{T2}
\end{aligned} \tag{E.6}$$

Consider T1:

$$\left\langle \nabla\psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star), \boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star - \mu_k^\star \circ \boldsymbol{\tau}_k^\star \right\rangle \leq \|\nabla\psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star)\|_2 \|\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star - \mu_k^\star \circ \boldsymbol{\tau}_k^\star\|_2$$

Consider T2:

$$\begin{aligned}
(\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star - \boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star)^T \nabla^2\psi(\bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star - \boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star) &= (\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star - \boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star)^T \nabla^2\psi(\bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star - \boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star) \\
&= \|\boldsymbol{v}\|_2^2 \frac{d^2}{dt^2}\psi(\bar{\boldsymbol{\mu}} + t\hat{\boldsymbol{v}})\big|_{t=0}
\end{aligned}$$

Given that $\psi(\bar{\boldsymbol{\mu}}) \geq \lambda$,
$T2 \geq \|\boldsymbol{v}\|_2^2 \lambda$.
Plugging the inequalities of T2 and T1 in E.6, we see that

$$\psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star) \geq \psi(\boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star) \geq \psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star) - \|\nabla\psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star)\|_2 \|\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star - \mu_k^\star \circ \boldsymbol{\tau}_k^\star\|_2 + \frac{1}{2}\|\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star - \boldsymbol{\mu}_k^\star\|_2^2 \lambda$$
$$\Rightarrow \|\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star - \boldsymbol{\mu}_k^\star \circ \boldsymbol{\tau}_k^\star\|_2 \leq \frac{2\|\nabla\psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star)\|_2}{\lambda}$$
$$\Rightarrow d(S_j^+, S_k^+) \leq \frac{2\|\nabla\psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star)\|_2}{\lambda}$$

Since $\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star$ is optimal to $\varphi(\boldsymbol{\mu})$, $\|\nabla\varphi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star) - \nabla\psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star)\|_2 = \|\nabla\psi(\boldsymbol{\mu}_j^\star \circ \boldsymbol{\tau}_j^\star)\|_2 \leq \epsilon$. This completes the proof. ■

Lemma E.3 *Let*

$$\psi(\boldsymbol{\mu}) = \min_{\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n \in \mathbb{T}} \left\{ \sum_{i=1}^{n} \boldsymbol{W}_{ik} \|\boldsymbol{v}_i^{(p)} \circ \boldsymbol{\tau}_i - \boldsymbol{\mu}\|^2 \right\}$$
$$\varphi(\boldsymbol{\mu}) = \min_{\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n \in \mathbb{T}} \left\{ \sum_{i=1}^{n} \boldsymbol{W}_{ij} \|\boldsymbol{v}_i^{(p)} \circ \boldsymbol{\tau}_i - \boldsymbol{\mu}\|^2 \right\}$$

and $\Delta W = \max_i |W_{ik} - W_{ij}|$. *Then,*

$$\forall \boldsymbol{\mu} \in \mathcal{I} = \Big\{ \boldsymbol{x} \ | \ \min_j d(\boldsymbol{x}, \mathcal{S}_j^{(d)}) \leq R^{(p)} \Big\}, \tag{E.7}$$

we have

$$\|\nabla\varphi(\boldsymbol{\mu}) - \nabla\psi(\boldsymbol{\mu})\|_2 \leq 2nR^{(p)}\Delta W$$

Proof For a given $\boldsymbol{\mu} \in \mathcal{I}$, for all i the solution $\boldsymbol{\tau}_i'$ to the optimization problem

$$\min_{\boldsymbol{\tau} \in \mathbb{T}} \|\boldsymbol{v}_i \circ \boldsymbol{\tau}_i - \boldsymbol{\mu}\| \tag{E.8}$$

is unique. We use Danskin's Theorem to differentiate $\varphi(\boldsymbol{\mu})$. Let $\boldsymbol{\tau}_i'$ be the following:

$$\boldsymbol{\tau}_i' = \arg\min_{\boldsymbol{\tau}_i} \|\boldsymbol{v}_i \circ \boldsymbol{\tau}_i - \boldsymbol{\mu}\|_2$$

With this,

$$\nabla\varphi(\boldsymbol{\mu}) = -2\sum_i W_{ij}(\boldsymbol{v}_i \circ \boldsymbol{\tau}_i' - \boldsymbol{\mu})$$

$$\nabla\psi(\boldsymbol{\mu}) = -2\sum_i W_{ik}(\boldsymbol{v}_i \circ \boldsymbol{\tau}_i' - \boldsymbol{\mu})$$

$$\Rightarrow \nabla\varphi(\boldsymbol{\mu}) - \nabla\psi(\boldsymbol{\mu}) = -2\sum_i (W_{ij} - W_{ik})(\boldsymbol{v}_i \circ \boldsymbol{\tau}_i' - \boldsymbol{\mu})$$

$\boldsymbol{\tau}_i'$ is optimal both for $\varphi(\boldsymbol{\mu})$ and $\psi(\boldsymbol{\mu})$ since it does not depend on the weights unlike $\boldsymbol{\mu}$. Denote the following quantities:

$$W_{ik} = W_{ik} + \Delta W_i$$
$$\Delta W = \max_i |\Delta W_i|$$

with these definitions,

$$\begin{aligned} \|\nabla\varphi(\boldsymbol{\mu}) - \nabla\psi(\boldsymbol{\mu})\|_2 &\leq 2\Delta W \sum_i \|\boldsymbol{v}_i \circ \boldsymbol{\tau}_i' - \boldsymbol{\mu}\|_2 \\ &\leq 2nR\Delta W \qquad \because \boldsymbol{\mu} \in \mathcal{I} \end{aligned} \tag{E.9}$$

■

Lemma E.4 (Triangle inequality) *For transformation manifolds* $\mathcal{S}_i$, $\mathcal{S}_j$, $\mathcal{S}_k$,

$$d(\mathcal{S}_i, \mathcal{S}_j) - d(\mathcal{S}_i, \mathcal{S}_k) \leq d(\mathcal{S}_j, \mathcal{S}_k) \tag{E.10}$$

Proof We have

$$d(\mathcal{S}_i, \mathcal{S}_k) = \min_{v \in \mathcal{S}_i,\, v' \in \mathcal{S}_k} \|v - v'\|_2.$$

For every $v'' \in \mathcal{S}_j$, $||v - v'||_2 \leq ||v - v''||_2 + ||v'' - v'||_2$. Therefore, for every $v'' \in \mathcal{S}_j$:

$$\min_{v \in \mathcal{S}_i,\, v' \in \mathcal{S}_k} ||v - v'||_2 \leq \min_{v \in \mathcal{S}_i,\, v' \in \mathcal{S}_k} ||v - v''||_2 + ||v'' - v'||_2$$

In particular, this inequality holds for the $v'' \in \mathcal{S}_j$ which minimizes $||v - v''||_2$ and so we obtain

$$d(\mathcal{S}_i, \mathcal{S}_k) \leq \min_{v \in \mathcal{S}_i} d(v, \mathcal{S}_j) + \min_{v' \in \mathcal{S}_k} d(v'', v') = d(\mathcal{S}_i, \mathcal{S}_j) + d(v'', \mathcal{S}_k)$$

Finally, using the fact that these are euclidean transformation manifolds, $d(v'', \mathcal{S}_k) = d(\mathcal{S}_j, \mathcal{S}_k)$ for every choice of $v'' \in \mathcal{S}_j$. Hence we obtain

$$d(\mathcal{S}_i, \mathcal{S}_k) \leq d(\mathcal{S}_i, \mathcal{S}_j) + d(\mathcal{S}_j, \mathcal{S}_k),$$

completing the proof. ■

Lemma E.5 *Let* $\Delta W = \max_i |W_{ik} - W_{ij}|$ *and* $R = \max_{m,l} d(S_m, S_l)$*, then*

$$\Delta W \leq 2\frac{\beta R}{n} e^{\beta R^2}(1 + e^{\beta R^2}) d(S_j, S_k) \tag{E.11}$$

Proof

$$\begin{aligned}
\Delta W &= \max_i |W_{ik} - W_{ij}| \\
&= \left\| \frac{\gamma_k}{\|\gamma_k\|_1} - \frac{\gamma_j}{\|\gamma_j\|_1} \right\|_\infty \\
&= \left\| \frac{\gamma_k}{\|\gamma_k\|_1} + \frac{\gamma_j}{\|\gamma_k\|_1} - \frac{\gamma_j}{\|\gamma_k\|_1} - \frac{\gamma_j}{\|\gamma_j\|_1} \right\|_\infty \\
&= \left\| \frac{\gamma_k - \gamma_j}{\|\gamma_k\|_1} + \gamma_j \left(\frac{1}{\|\gamma_k\|_1} - \frac{1}{\|\gamma_j\|_1} \right) \right\|_\infty \\
&\leq \frac{\|\gamma_k - \gamma_j\|_\infty}{\|\gamma_k\|_1} + \left| \frac{\|\gamma_j\|_1 - \|\gamma_k\|_1}{\|\gamma_j\|_1 \|\gamma_k\|_1} \right| \|\gamma_j\|_\infty
\end{aligned} \tag{E.12}$$

CLAIM: $\|\gamma_j\|_1 \geq n e^{-\beta R^2} = \alpha$

Proof:

$$\begin{aligned}
\gamma_{ij} &= e^{-\beta d^2(S_i, S_j)} \\
&\geq e^{-\beta R^2} && \because d(S_i, S_j) \leq R \\
\Rightarrow \|\gamma_j\|_1 &\geq n e^{-\beta R^2}
\end{aligned}$$

Using the result in Claim E.1 in equation E.12 and the reverse triangle inequality,

$$\Delta W \leq \underbrace{\frac{\|\gamma_k - \gamma_j\|_\infty}{\alpha}}_{T1} + \underbrace{\frac{\|\gamma_j - \gamma_k\|_1}{\alpha^2}}_{T2} \qquad \because \gamma_{ij} \leq 1 \tag{E.13}$$

CLAIM: If $x \geq 0, y \geq 0$, $|e^{-x} - e^{-y}| < |y - x|$

Proof: $f(x) = e^{-x}$ is a convex function. therefore

$$\begin{aligned}
f(y) &\geq f(x) + \nabla f(x)^T (y - x) \\
e^{-y} &\geq e^{-x} - e^{-x}(y - x) \\
e^{-x} - e^{-y} &\leq \frac{(y - x)}{e^x} \\
&\leq \frac{(y - x)}{1 + x} \\
&\leq (y - x) && \because x \geq 0
\end{aligned}$$

We have:

$$|e^{-x} - e^{-y}| = \begin{cases} e^{-x} - e^{-y} \text{ if } x \leq y \\ e^{-y} - e^{-x} \text{ if } x \geq y \end{cases}$$

This implies:

$$|e^{-x} - e^{-y}| \leq \begin{cases} y - x \text{ if } x \leq y \\ x - y \text{ if } x \geq y \end{cases} \quad \Rightarrow |e^{-x} - e^{-y}| \leq |y - x|$$

Using claim E.1, Lemma E.4, we have the following:

$$\begin{aligned}
|\gamma_{ik} - \gamma_{ij}| &= |e^{-\beta d^2(S_i, S_k)} - e^{-\beta d^2(S_i, S_j)}| \\
&\leq \beta |d^2(S_i, S_j) - d^2(S_i, S_k)| \\
&\leq \beta |d(S_i, S_j) - d(S_i, S_k)| \times |d(S_i, S_j) + d(S_i, S_k)| \\
&\leq 2\beta R d(S_j, S_k)
\end{aligned} \tag{E.14}$$

Using E.14 in E.13, we have:

$$\Delta W \leq 2\beta R \left(\frac{1}{\alpha} + \frac{n}{\alpha^2}\right) d(S_j, S_k)$$
$$= 2\frac{\beta R}{n} e^{\beta R^2}(1 + e^{\beta R^2}) d(S_j, S_k) \tag{E.15}$$

■

F APPENDIXNA

We note using lemma D.5 that

$$\frac{d^2}{dt^2} d^2(\boldsymbol{\mu}_j^\star + t\boldsymbol{r}, S_i) = 2\left\langle \boldsymbol{r}, \boldsymbol{r} - \frac{d}{dt}\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]\right\rangle$$

We note that $\frac{d}{dt}\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}] = \frac{d}{dt'}\Big|_{(t'=0)}\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r} + t'\boldsymbol{r}]$.
We use Lemma C.1 to get the value of the derivative and substitute above:

$$\begin{aligned}
\frac{d^2}{dt^2} d^2(\boldsymbol{\mu}_j^\star + t\boldsymbol{r}, S_i) &= 2\left\langle \boldsymbol{r}, \boldsymbol{r} - \sum_{k=0}^{\infty}\Big(\mathrm{II}^*(\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}])\Big)^k \mathcal{P}_{\mathcal{T}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\right\rangle \\
&= 2\Big(\Big\langle \boldsymbol{r}, \boldsymbol{r} - \mathcal{P}_{\mathcal{T}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\Big\rangle - \Big\langle \boldsymbol{r}, \sum_{k=1}^{\infty}\Big(\mathrm{II}^*(\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}])\Big)^k \mathcal{P}_{\mathcal{T}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\Big\rangle\Big) \\
&= 2\Big(\Big\langle \boldsymbol{r}, \mathcal{P}_{\mathcal{N}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\Big\rangle - \Big\langle \boldsymbol{r}, \sum_{k=1}^{\infty}\Big(\mathrm{II}^*(\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}])\Big)^k \mathcal{P}_{\mathcal{T}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\Big\rangle\Big) \\
&= 2\Big(\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\|_2^2 - \Big\langle \boldsymbol{r}, \sum_{k=1}^{\infty}\Big(\mathrm{II}^*(\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}])\Big)^k \mathcal{P}_{\mathcal{T}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\Big\rangle\Big) \\
&\geq 2\Big(\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\|_2^2 - \|\sum_{k=1}^{\infty}\Big(\mathrm{II}^*(\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}])\Big)^k \mathcal{P}_{\mathcal{T}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\|_2\Big) \\
&\geq 2(\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\|_2^2 - \sum_{k=1}^{\infty}\|\Big(\mathrm{II}^*(\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}])\Big)^k \mathcal{P}_{\mathcal{T}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\|_2) \\
&\geq 2\Big(\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\|_2^2 - \sum_{k=1}^{\infty}\|\Big(\mathrm{II}^*(\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}])\Big)^k\|_2 \|\mathcal{P}_{\mathcal{T}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\|_2\Big) \\
&\geq 2\Big(\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\|_2^2 - \sum_{k=1}^{\infty}\|\Big(\mathrm{II}^*(\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}])\Big)^k\|_2\Big) \\
&\geq 2\Big(\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\|_2^2 - \sum_{k=1}^{\infty}\|\mathrm{II}^*(\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}])\|_2^k\Big) \\
&\geq 2\Big(\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]}S_i}[\boldsymbol{r}]\|_2^2 - \sum_{k=1}^{\infty}\left(\kappa_i\|\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]\|_2\right)^k\Big) \qquad \text{(F.1)}
\end{aligned}$$

Here κ_i is the maximum curvature of the manifold S_i.
Given that $\frac{1}{\kappa_i} \geq \rho \geq 5R \Rightarrow \kappa_i \leq \frac{1}{5R}$. Also, $\|\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]\|_2 \leq 2R$. Therefore,

$$\kappa_i\|\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]\|_2 \leq \frac{2}{5} < 1.$$

Therefore (F.1) can be written as:

$$\begin{aligned}
\frac{d^2}{dt^2} d^2(\boldsymbol{\mu}_j^\star + t\boldsymbol{r}, S_i) &\geq 2\Big(\|\mathcal{P}_{\mathcal{N}_{\mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]} S_i}[\boldsymbol{r}]\|_2^2 - \frac{\kappa_i \|\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]\|_2}{1 - \kappa_i \|\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]\|_2}\Big) \\
&\geq 2\Big(1 - 72\kappa^2 R^2 - 2\kappa R - \frac{\kappa_i \|\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]\|_2}{1 - \kappa_i \|\boldsymbol{\mu}_j^\star + t\boldsymbol{r} - \mathcal{P}_{S_i}[\boldsymbol{\mu}_j^\star + t\boldsymbol{r}]\|_2}\Big) \\
&\geq 2(1 - 72\kappa^2 R^2 - 2\kappa R - \frac{2}{3}) \\
&= \frac{2}{3} - 144\kappa^2 R^2 - 4\kappa R
\end{aligned}$$