Expanding the Family of Grassmannian Kernels: An Embedding Perspective

Mehrtash T. Harandi, Mathieu Salzmann, Sadeep Jayasumana,
Richard Hartley, and Hongdong Li

Australian National University, Canberra, ACT 0200, Australia
NICTA, Locked Bag 8001, Canberra, ACT 2601, Australia⋆

1 Proof of Length Equivalence

Here, we prove Theorem 2 from Section 3, *i.e.*, the equivalence up to a scale of $\sqrt{2}$ of the length of any given curve under the Binet-Cauchy distance δ_{bc} derived from the Plücker embedding and the geodesic distance δ_g. The proof of this theorem follows several steps. We start with the definition of curve length and intrinsic metric. Without any assumption on differentiability, let $(\mathcal{M}, d)$ be a metric space. A curve in $\mathcal{M}$ is a continuous function $\gamma : [0, 1] \to \mathcal{M}$ and joins the starting point $\gamma(0) = x$ to the end point $\gamma(1) = y$.

Definition 1. *The length of a curve γ is the supremum of $L(\gamma; \{t_i\})$ over all possible partitions $\{t_i\}$, where $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1$ and $L(\gamma; \{t_i\}) = \sum_i d\left(\gamma(t_i), \gamma(t_{i-1})\right)$.*

Definition 2. *The intrinsic metric $\widehat{\delta}(x, y)$ on $\mathcal{M}$ is defined as the infimum of the lengths of all paths from x to y.*

Theorem 1 ([2]). *If the intrinsic metrics induced by two metrics d_1 and d_2 are identical up to a scale ξ, then the length of any given curve is the same under both metrics up to ξ.*

Theorem 2 ([2]). *If $d_1(x, y)$ and $d_2(x, y)$ are two metrics defined on a space $\mathcal{M}$ such that*

$$\lim_{d_1(x,y)\to 0} \frac{d_2(x, y)}{d_1(x, y)} = 1. \tag{1}$$

uniformly (with respect to x and y), then their intrinsic metrics are identical.

Therefore, here, we need to study the behavior of

$$\lim_{\delta_g^2(\mathbf{X},\mathbf{Y})\to 0} \frac{\delta_{bc}^2(\mathbf{X}, \mathbf{Y})}{\delta_g^2(\mathbf{X}, \mathbf{Y})}$$

to prove our theorem on curve length equivalence.

⋆ NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the ARC through the ICT Centre of Excellence program.

Proof. Since $\sin(\theta_i) \to \theta_i$ for $\theta_i \to 0$, we can see that

$$\begin{aligned}\lim_{\delta_g(\mathbf{X},\mathbf{Y})\to 0} \frac{\delta_{bc}^2(\mathbf{X},\mathbf{Y})}{\delta_g^2(\mathbf{X},\mathbf{Y})} &= \lim_{\theta_i \to 0} \frac{2 - 2\prod_{i=1}^{p}(1-\sin^2\theta_i)}{\sum_{i=1}^{p}\theta_i^2} \\ &= \lim_{\theta_i \to 0} \frac{2 - 2\prod_{i=1}^{p}(1-\theta_i^2)}{\sum_{i=1}^{p}\theta_i^2} = \lim_{\theta_i \to 0} \frac{2 - 2(1-\sum_{i=1}^{p}\theta_i^2)}{\sum_{i=1}^{p}\theta_i^2} = 2 \ .\end{aligned}$$

This, in conjunction with Theorem 1, concludes the proof. □

2 Additional Experiment

We performed an additional experiment on body-gesture recognition using the UMD Keck dataset [3]. To this end, we consider the problem of kernel sparse coding on the Grassmannian which can be formulated as

$$\min_{\boldsymbol{y}} \left\| \phi(\mathbf{X}) - \sum\nolimits_{j=1}^{N} y_j \phi(\mathbf{D}_j) \right\|^2 + \lambda \|\boldsymbol{y}\|_1 \ , \tag{2}$$

where $\mathbf{D}_j \in \mathcal{G}(p,d)$ is a dictionary atom, $\mathbf{X} \in \mathcal{G}(p,d)$ is the query and $\boldsymbol{y}$ is the vector of sparse codes. In practice, we used each training sample as an atom in the dictionary. Note that, as shown in [1], (2) only depends on the kernel values computed between the dictionary atoms, as well as between the query point and the dictionary. Classification is then performed by assigning the label of the dictionary element $\mathbf{D}_i$ with strongest response $\boldsymbol{y}_i$ to the query.

The UMD Keck dataset [3] comprises 14 body gestures with static and dynamic backgrounds (see examples in Figure 1). The dataset contains 126 videos from static scenes and 168 ones from dynamic environments. Following the experimental protocol used in [4], we first extracted the region of interest around each gesture and resized it to 32×32 pixels. We then represented each video by a subspace of order 6, thus yielding points on $\mathcal{G}(6, 1024)$.

Table 1 compares the performance of our kernels with that of k_{bc}^2 and k_p. Note that our kernels outperform the baselines in both the static and dynamic settings. The maximum accuracy is obtained by $k_{l,p}$ for the static scenario (99.2%), and by $k_{bi,p}$ for the dynamic one (99.1%). For the same experiments, the state-of-the-art solution using product manifolds [4] achieves 94.4% and 92.3%, respectively.

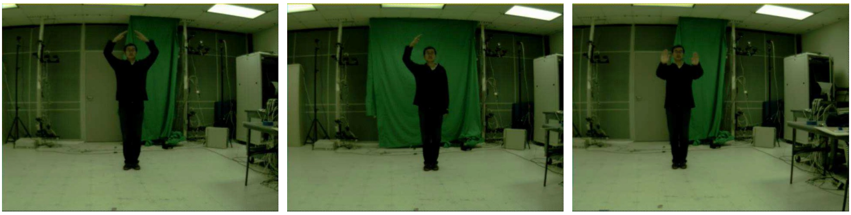

Fig. 1: Sample images from the UMD Keck body-gesture dataset [3].

Table 1: **Body-gesture recognition.** Accuracies on UMD Keck [3] using kernel sparse coding.

kernel	$\mathbf{k_{bc}^2}$	$\mathbf{k_{p,bc}}$	$\mathbf{k_{r,bc}}$	$\mathbf{k_{l,bc}}$	$\mathbf{k_{bi,bc}}$	$\mathbf{k_{log,bc}}$
Static	84.1% ± 3.6	88.1% ± 6.0	90.5% ± 4.7	91.3% ± 7.7	**93.7% ± 3.6**	87.3% ± 3.6
Dynamic	90.2%	92.0%	93.8%	**94.6%**	**94.6%**	92.9%
kernel	$\mathbf{k_{p}}$	$\mathbf{k_{p,p}}$	$\mathbf{k_{r,p}}$	$\mathbf{k_{l,p}}$	$\mathbf{k_{bi,p}}$	$\mathbf{k_{log,p}}$
Static	88.9% ± 8.4	94.4% ± 6.0	97.6% ± 4.1	**99.2% ± 1.3**	96.0% ± 3.6	92.9% ± 2.4
Dynamic	91.1%	92.0%	98.2%	98.2%	**99.1%**	97.3%

References

1. Gao, S., Tsang, I.W.H., Chia, L.T.: Kernel sparse representation for image classification and face recognition. In: Proc. European Conference on Computer Vision (ECCV). pp. 1–14 (2010)
2. Hartley, R., Trumpf, J., Dai, Y., Li, H.: Rotation averaging. Int. Journal of Computer Vision 103(3), 267–305 (2013)
3. Lin, Z., Jiang, Z., Davis, L.S.: Recognizing actions by shape-motion prototype trees. In: Proc. Int. Conference on Computer Vision (ICCV). pp. 444–451. IEEE (2009)
4. Lui, Y.M.: Human gesture recognition on product manifolds. Journal of Machine Learning Research 13(1), 3297–3321 (2012)