Geometry-Aware Hierarchical Bayesian Learning on Manifolds Supplementary

1. Theorem1 and Proof

Theorem 1. *A real-valued function $T(v,t)$ on $\mathbb{R}^d$ is a spatial-temporal kernel function if it is a linear/non-linear diffusion process: $\frac{\partial T}{\partial t} = \alpha \Delta T + P(t)\delta(v)$, where α is a positive constant, $P(t)$ is a periodic function, $\delta(v)$ is the Dirac delta function, and Δ is the Laplace operator.*

Proof. As known, there exists a corresponding Green's function $G(v, v', t, t')$ for a parabolic partial differential equation so that the diffusion process expressed by this parabolic partial differential equation has the form [3, 7]:

$$T(v, v', t) = \int_0^t G(v, v', t-s)P(s)ds \tag{1}$$

In the main submission we have proved that the analytical solution to the following equation is PSD:

$$T = \int_0^t G(v, v', t-s)cos\omega(t-s)ds \tag{2}$$

Since the spatial variable mainly exists in the Green's function and the Green's function is spatial stationary in a $\mathcal{R}^d$ diffusion process, the primary task is to prove any choices of periodic function $P(t)$ can lead to the same conclusion. Because Eq. (2) has been prove to be PSD, we can draw the same conclusion if Eq. (1) has a similar form with Eq. (2). Therefore, the main idea is using cosine function to generalize a periodic function $P : \mathbb{R} \to \mathbb{R}$. As known, a periodic function can be estimated with the Fourier series expansion:

$$P(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}[a_n cos(nt) + b_n sin(nt)] \tag{3}$$

where a, b are arbitrary real numbers. Assuming there exists a cosine function $x_n cos(nt + y_n)$ that is equal to $a_n cos(nt) + b_n sin(nt)$, n is a constant. By using the trigonometric sum formulae, we get: $x_n = \pm\sqrt{a_n^2 + b_n^2}$ and $y = arctan(-\sqrt{\frac{b_n}{a_n}})$. Eq. (3) is then transformed to:

$$P(t) = \frac{1}{2}a_0 \pm \sum_{n=1}^{\infty}\sqrt{a_n^2 + b_n^2}cos\left[nt + arctan(-\sqrt{\frac{b_n}{a_n}})\right] \tag{4}$$

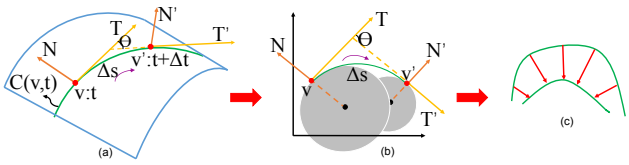

Figure 1. Sketch plots of (a) a planar curve example; (b) the curve in a 2D plane. The shaded circles are inscribed circles on v and v'; (c) the curve moves from top to down by curvature flow.

Eq. (4) shows that any periodic functions can be approximated by the linear combination of cosine functions. By substituting Eq. (4) into Eq. (1), we see that the diffusion process is estimated to be the integral of Green's function times the combination of cosine functions. Applying the PSD summation identity, we can draw the conclusion that the solution to Eq. (1) is also PSD. Therefore, it is a valid kernel function. □

2. Lemma 1 and Proof

Lemma 1. *The GAC Kernel embeds the mean curvature flow in $\mathbb{R}^3$, which enables it to be geometry-aware.*

Proof. In differential geometry, a curvature flow numerically links intrinsic geometric features and extrinsic flows together [4]. We take the proof on a planar curve by assuming manifold $\mathcal{M}$ is a two dimensional manifold in $\mathbb{R}^3$ as an example for convenience. In this case, the GAC kernel is actually equivalent to a curve-shortening flow which can be considered as a one dimension mean curvature flow [1]. Figure 1 shows sketch plots of the symbols used in this proof. Suppose v is a point on the manifold. $C(v)$ is the intersection between the manifold and the normal plane on v. As known, $C(v)$ is a 1-dimensional smooth curve. Assume one point moves along C from v to v'. Let Δs be the arc length of this movement and θ be the rotation angle of the tangent vector, then we can define the following concepts:
(i) the velocity vector at v is $\frac{dC}{dv}$;
(ii) the velocity is the magnitude of the velocity vector, which is $\left|\frac{dC}{dv}\right| = \frac{ds}{dv}$;
(iii) the unit tangent vector $T = \frac{dC}{ds} / \left|\frac{dC}{ds}\right|$ and the unit normal vector $N = \mathcal{R}T$. $\mathcal{R}$ is a $\pi/2$ rotation matrix;

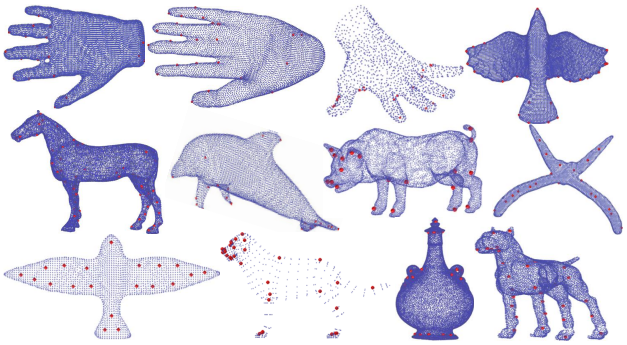

Figure 2. Illustration of saliency selection on McGill 3D shape benchmark [6]. The salient points are marked by red spheres.

(iv) the curvature κ, which measures how fast the unit tangent vector rotates relative to the arc length: $\kappa = \lim_{\Delta s \to 0} \left| \frac{\Delta \theta}{\Delta s} \right|$. And we can further get $\frac{dT}{ds} = (-sin\theta, cos\theta)\frac{d\theta}{ds} = \kappa N$ and similarly $\frac{dN}{ds} = -\kappa T$.

Assume all points on the curve start to move along their normal directions at a velocity of $\kappa(v)$ during time t, we have the curvature flow: $\frac{dC}{dt} = \kappa N$. With the equation (iii) and (iv), we write the curvature flow as: $\frac{dC}{dt} = \frac{d}{ds}\frac{dC}{ds}$, which is clearly a diffusion process with a zero reaction function. This equivalence proves that the GAC kernel can theoretically reflect the geometric features. From the perspective of physical meanings, $G(v, v')$ means how much curvature changes from point v to its neighborhood v' during a period of time. Therefore, the physical meanings also support that the GAC kernel embeds the geometric information of the manifolds. Lemma 1 is proved. □

3. Lemma 2 and Proof

Lemma 2. *The GAC Kernel embeds a convolution filtering within the kernel structure, called intra-kernel convolution.*

Proof. According to the reaction diffusion theory [5, 7], Eq. (1) can be expressed as:

$$T = \int_0^t \int_{-\infty}^{\infty} G(v - v', t - s)P(v', s)dv'ds \qquad (5)$$

If integrating along the temporal variable, then the result has the form $\int P(v')G(v, v')dv'$, which matches with the definition of a convolutional filtering $\int f(v')h_t(v, v')dv' = (f_0 * h_t)(v)$. Our kernel derivation also indicates the existence of a convolution on manifolds. Reminding that we estimate the integral in Eq. (1) as the summation of a sine Fourier transform and a cosine Fourier transform (Eq.14 in the main submission). Each term implements the transform from time domain to frequency domain. According to the Convolution Theorem, we can draw the same conclusion. The similar theory has also been applied in geometric deep learning to realize the convolution on manifolds [2]. Lemma 2 is proved. □

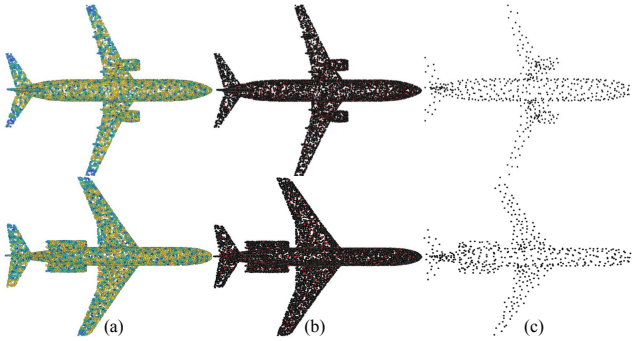

Figure 3. Illustration of saliency selection on Modelnet40 [8]. (a)Demonstrations of saliency maps after selecting 200 salient points. (b)Demonstrations of 200 salient points. The salient points are marked by red spheres. (c) The selected 200 salient points out of 10000 vertices can generally represent the original shape.

4. Unsupervised Salient Point Selection

We provide more salient point selection results on point clouds here. Figure. 2 shows twelve examples of salient point selection on McGill 3D shape benchmark [6]. The shapes have different sampling densities. The original data type is the triangle mesh. We remove their edge connections and only use the vertex coordinates as inputs. The results show that the GAC-GP can learn the geometric property of the inputs in the prior and the selected salient points are representative in distinguishing the shapes.

Figure. 3 illustrates two examples of salient point selection on Modelnet40 dataset. Figure. 3 (a) demonstrates the saliency maps after selecting 200 salient points. Figure. 3 (b) shows salient points on the original point clouds. Figure. 3 (c) is 200 salient points. We can see that 200 salient points can generally depict the original shapes. Noting that we did not use salient point selection algorithm in the point cloud classifications in the main submission. This is mainly because the comparison methods used the whole data as inputs, for fairness and convenience, we also use 10000 vertices as inputs in the experiments.

References

[1] Steven J Altschuler. Shortening space curves. *Differential geometry: Riemannian geometry (Los Angeles, CA, 1990)*, 54:45–51, 1993.

[2] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[3] Gert Ehrlich and Kaj Stolt. Surface diffusion. *Annual Review of Physical Chemistry*, 31(1):603–637, 1980.

[4] Satyanad Kichenassamy, Arun Kumar, Peter Olver, Allen Tannenbaum, and Anthony Yezzi. Gradient flows and geometric active contour models. In *Proceedings of IEEE International Conference on Computer Vision*, pages 810–815. IEEE, 1995.

[5] Christina Kuttler. Reaction-diffusion equations with applications. In *Internet seminar*, 2011.

[6] Kaleem Siddiqi, Juan Zhang, Diego Macrini, Ali Shokoufandeh, Sylvain Bouix, and Sven Dickinson. Retrieving articulated 3-d models using medial surfaces. *Machine vision and applications*, 19(4):261–275, 2008.

[7] Walter A Strauss. *Partielle Differentialgleichungen: eine Einführung*. Springer-Verlag, 2013.

[8] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015.