Multiscale Covariance Tensors for Data on Riemannian Manifolds

Haibin Hang
hhang@math.fsu.edu

Supervisor: Dr. Washington Mio

Florida State University

Department of Mathematics

Covariance Matrix

Given a random vector $\mathbf{Y}$ in Euclidean space $\mathbb{R}^n$, with the probability measure α, we look at its projection $< \mathbf{Y}, v >= \mathbf{Y}^T v$ onto a one-dimensional vector space l_v spanned by unit vector $v \in S^{n-1}$.

Covariance Matrix

Given a random vector $\mathbf{Y}$ in Euclidean space $\mathbb{R}^n$, with the probability measure α, we look at its projection $< \mathbf{Y}, v >= \mathbf{Y}^T v$ onto a one-dimensional vector space l_v spanned by unit vector $v \in S^{n-1}$. We calculate its mean and variance:

$$\mathrm{E}[\mathbf{Y}^T v] = \int_{\mathbb{R}^n} y^T v d\alpha(y) = \left(\int_{\mathbb{R}^n} y d\alpha(y) \right)^T \cdot v = \mathrm{E}[\mathbf{Y}]^T \cdot v$$

Covariance Matrix

Given a random vector $\mathbf{Y}$ in Euclidean space $\mathbb{R}^n$, with the probability measure α, we look at its projection $< \mathbf{Y}, v >= \mathbf{Y}^T v$ onto a one-dimensional vector space l_v spanned by unit vector $v \in S^{n-1}$. We calculate its mean and variance:

$$\mathrm{E}[\mathbf{Y}^T v] = \int_{\mathbb{R}^n} y^T v d\alpha(y) = \left(\int_{\mathbb{R}^n} y d\alpha(y) \right)^T \cdot v = \mathrm{E}[\mathbf{Y}]^T \cdot v$$

$$\sigma^2(\mathbf{Y}^T v) = \int_{\mathbb{R}^n} (y^T v - \mathrm{E}[\mathbf{Y}^T v])^2 d\alpha(y)$$

Covariance Matrix

Given a random vector $\mathbf{Y}$ in Euclidean space $\mathbb{R}^n$, with the probability measure α, we look at its projection $< \mathbf{Y}, v >= \mathbf{Y}^T v$ onto a one-dimensional vector space l_v spanned by unit vector $v \in S^{n-1}$. We calculate its mean and variance:

$$\mathrm{E}[\mathbf{Y}^T v] = \int_{\mathbb{R}^n} y^T v d\alpha(y) = \left(\int_{\mathbb{R}^n} y d\alpha(y) \right)^T \cdot v = \mathrm{E}[\mathbf{Y}]^T \cdot v$$

$$\sigma^2(\mathbf{Y}^T v) = \int_{\mathbb{R}^n} (y^T v - \mathrm{E}[\mathbf{Y}^T v])^2 d\alpha(y)$$

Arrange the terms we get:

$$\sigma^2(\mathbf{Y}^T v) = v^T \cdot \left(\int_{\mathbb{R}^n} (y - \mathrm{E}[\mathbf{Y}]) \cdot (y - \mathrm{E}[\mathbf{Y}])^T d\alpha(y) \right) \cdot v$$

The **covariance matrix** of $\mathbf{Y}$ is defined to be

$$\begin{aligned}\Sigma :=& \int_{\mathbb{R}^n} (y - \mathrm{E}[\mathbf{Y}])(y - \mathrm{E}[\mathbf{Y}])^T d\alpha(y) \\ =& \mathrm{E}\left[(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^T\right]\end{aligned}$$

The **covariance matrix** of $\mathbf{Y}$ is defined to be

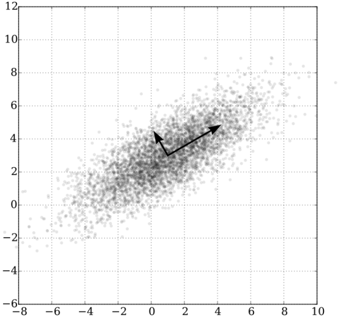
$$\begin{aligned}\Sigma :=& \int_{\mathbb{R}^n} (y - \mathrm{E}[\mathbf{Y}])(y - \mathrm{E}[\mathbf{Y}])^T d\alpha(y) \\ =& \mathrm{E}\left[(\mathbf{Y} - \mathrm{E}[\mathbf{Y}]) (\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^T \right]\end{aligned}$$

Then the variance of projection $\mathbf{Y}^T v$ is $\sigma^2(\mathbf{Y}^T v) = v^T \Sigma v$, which measures how far the data are spread out from their mean in direction v.

The **covariance matrix** of $\mathbf{Y}$ is defined to be

$$\begin{aligned}\Sigma :=& \int_{\mathbb{R}^n} (y - \mathrm{E}[\mathbf{Y}])(y - \mathrm{E}[\mathbf{Y}])^T d\alpha(y) \\ =& \mathrm{E}\left[(\mathbf{Y} - \mathrm{E}[\mathbf{Y}]) (\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^T \right]\end{aligned}$$

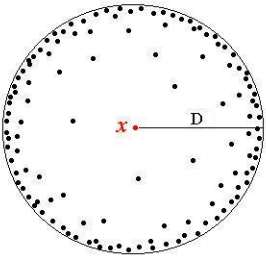
Then the variance of projection $\mathbf{Y}^T v$ is $\sigma^2(\mathbf{Y}^T v) = v^T \Sigma v$, which measures how far the data are spread out from their mean in direction v.
Hence covariance matrix contains information about how far the data are spread out from mean in each direction.



Sample points from a non-isotropic Gaussian distribution and its covariance matrix; the directions of the arrows correspond to the eigenvectors of this covariance matrix and their lengths to the square roots of the eigenvalues.
Covariance matrix is commonly used in principal component analysis.

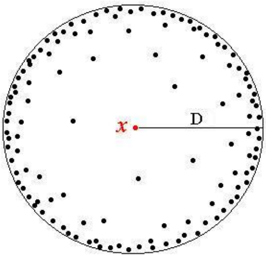
Why we look at local covariance matrix

We look at a very special distribution, with random points distributed evenly around a circle.



Why we look at local covariance matrix

We look at a very special distribution, with random points distributed evenly around a circle.



What shall the covariance matrix look like?

Kernel Function

$$K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

- Kernel function $K(x, y)$ measures the space scale dependence between data at points x and y.

Kernel Function

$$K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

- Kernel function $K(x, y)$ measures the space scale dependence between data at points x and y.
- When we fix (center) x, $K(x, y) \to 0$ as $y \to \infty$.

Kernel Function

$$K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

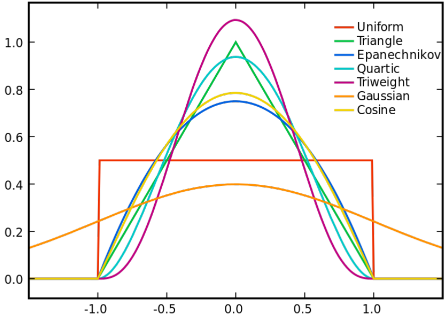
- Kernel function $K(x, y)$ measures the space scale dependence between data at points x and y.
- When we fix (center) x, $K(x, y) \to 0$ as $y \to \infty$.
- Kernel function $K(x, y)$ is symmetric on x and y variable.

Kernel Function

$$K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

- Kernel function $K(x, y)$ measures the space scale dependence between data at points x and y.
- When we fix (center) x, $K(x, y) \to 0$ as $y \to \infty$.
- Kernel function $K(x, y)$ is symmetric on x and y variable.

We can understand it as how clear we can see y from the standpoint x.



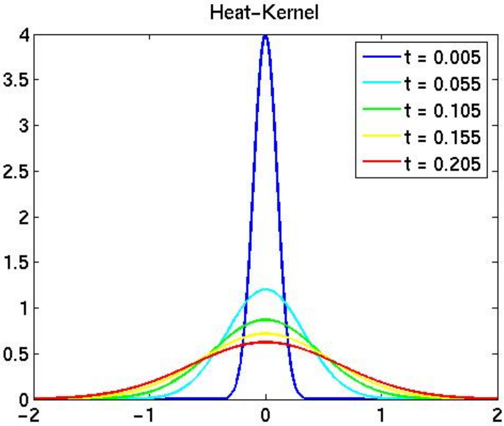
- A **multiscale kernel** function $K(x, y; t)$ is a sequence of kernel functions parametrized by $t \in [0, +\infty)$.

- A **multiscale kernel** function $K(x, y; t)$ is a sequence of kernel functions parametrized by $t \in [0, +\infty)$.
- As t goes to 0, $K(x, \cdot; t)$ converges to δ_x weakly.

- A **multiscale kernel** function $K(x, y; t)$ is a sequence of kernel functions parametrized by $t \in [0, +\infty)$.
- As t goes to 0, $K(x, \cdot; t)$ converges to δ_x weakly.

An important multiscale kernel: heat kernel

$$k(x, y, t) = \frac{1}{(4\pi t)^{n/2}} \exp(-\frac{\|x - y\|^2}{4t}).$$



Matrices and Tensors

- Given a vector space V, a multi linear map $V \times V \to \mathbb{R}$ is called ((0,2)-type) tensor.

Matrices and Tensors

- Given a vector space V, a multi linear map $V \times V \to \mathbb{R}$ is called ((0,2)-type) tensor.
- For any linear function $f, g \in V^*$, $f \otimes g : V \times V \to \mathbb{R}$, which sends (a, b) to $f(a)g(b)$, is called a pure tensor.

Matrices and Tensors

- Given a vector space V, a multi linear map $V \times V \to \mathbb{R}$ is called ((0,2)-type) tensor.
- For any linear function $f, g \in V^*$, $f \otimes g : V \times V \to \mathbb{R}$, which sends (a, b) to $f(a)g(b)$, is called a pure tensor.
- Each tensor is a linear combination of pure tensors.

Matrices and Tensors

- Given a vector space V, a multi linear map $V \times V \to \mathbb{R}$ is called ((0,2)-type) tensor.
- For any linear function $f, g \in V^*$, $f \otimes g : V \times V \to \mathbb{R}$, which sends (a, b) to $f(a)g(b)$, is called a pure tensor.
- Each tensor is a linear combination of pure tensors.
- If V is equipped with an inner product, we have an isomorphism $\alpha : V \to V^*$ which sends u to $u^* :=< u, \cdot >$.

Matrices and Tensors

- Given a vector space V, a multi linear map $V \times V \to \mathbb{R}$ is called ((0,2)-type) tensor.
- For any linear function $f, g \in V^*$, $f \otimes g : V \times V \to \mathbb{R}$, which sends (a, b) to $f(a)g(b)$, is called a pure tensor.
- Each tensor is a linear combination of pure tensors.
- If V is equipped with an inner product, we have an isomorphism $\alpha : V \to V^*$ which sends u to $u^* :=< u, \cdot >$.
- We write $u^* \otimes v^*$ simply as $u \otimes v$, hence $u \otimes v(a, b) =< u, a >< v, b >$, for $\forall u, v, a, b \in V$.

Matrices and Tensors

- Given a vector space V, a multi linear map $V \times V \to \mathbb{R}$ is called ((0,2)-type) tensor.
- For any linear function $f, g \in V^*$, $f \otimes g : V \times V \to \mathbb{R}$, which sends (a, b) to $f(a)g(b)$, is called a pure tensor.
- Each tensor is a linear combination of pure tensors.
- If V is equipped with an inner product, we have an isomorphism $\alpha : V \to V^*$ which sends u to $u^* :=< u, \cdot >$.
- We write $u^* \otimes v^*$ simply as $u \otimes v$, hence $u \otimes v(a, b) =< u, a >< v, b >$, for $\forall u, v, a, b \in V$.
- We can define an inner product between pure tensors $< u_1 \otimes v_1, u_2 \otimes v_2 >:=< u_1, u_2 >< v_1, v_2 >$, which expands linearly to the space of all (0,2) tensors.

Matrices and Tensors

- Given a vector space V, a multi linear map $V \times V \to \mathbb{R}$ is called ((0,2)-type) tensor.
- For any linear function $f, g \in V^*$, $f \otimes g : V \times V \to \mathbb{R}$, which sends (a, b) to $f(a)g(b)$, is called a pure tensor.
- Each tensor is a linear combination of pure tensors.
- If V is equipped with an inner product, we have an isomorphism $\alpha : V \to V^*$ which sends u to $u^* :=< u, \cdot >$.
- We write $u^* \otimes v^*$ simply as $u \otimes v$, hence $u \otimes v(a, b) =< u, a >< v, b >$, for $\forall u, v, a, b \in V$.
- We can define an inner product between pure tensors $< u_1 \otimes v_1, u_2 \otimes v_2 >:=< u_1, u_2 >< v_1, v_2 >$, which expands linearly to the space of all (0,2) tensors.
- If we fix a basis $\{e_1, e_2, \cdots, e_n\}$, then any tensor $\sum c^{ij} e_i \otimes e_j$ corresponds to a matrix (c^{ij}).

Matrices and Tensors

- Given a vector space V, a multi linear map $V \times V \to \mathbb{R}$ is called ((0,2)-type) tensor.
- For any linear function $f, g \in V^*$, $f \otimes g : V \times V \to \mathbb{R}$, which sends (a, b) to $f(a)g(b)$, is called a pure tensor.
- Each tensor is a linear combination of pure tensors.
- If V is equipped with an inner product, we have an isomorphism $\alpha : V \to V^*$ which sends u to $u^* :=< u, \cdot >$.
- We write $u^* \otimes v^*$ simply as $u \otimes v$, hence $u \otimes v(a, b) =< u, a >< v, b >$, for $\forall u, v, a, b \in V$.
- We can define an inner product between pure tensors $< u_1 \otimes v_1, u_2 \otimes v_2 >:=< u_1, u_2 >< v_1, v_2 >$, which expands linearly to the space of all (0,2) tensors.
- If we fix a basis $\{e_1, e_2, \cdots, e_n\}$, then any tensor $\sum c^{ij} e_i \otimes e_j$ corresponds to a matrix (c^{ij}).
- If the above basis is standard orthogonal basis, we have $\|\sum c^{ij} e_i \otimes e_i\| = \sqrt{\sum (c^{ij})^2}$.

Covariance Tensor Field on Euclidean Space

Definition (Martínez, Mémoli, Mio)

The multiscale covariance tensor field (CTF) of probability measure α on $\mathbb{R}^n$ associated with the kernel K is the one-parameter family of tensor fields, index by $t \in (0, \infty)$, given by

$$\Sigma_{\alpha,t}(x) := \int_{\mathbb{R}^n} (y - x) \otimes (y - x) K(x, y, t) d\alpha(y),$$

provided that the integral converges for each $x \in \mathbb{R}^d$ and $t > 0$.

Covariance Tensor Field on Euclidean Space

Definition (Martínez, Mémoli, Mio)

The multiscale covariance tensor field (CTF) of probability measure α on $\mathbb{R}^n$ associated with the kernel K is the one-parameter family of tensor fields, index by $t \in (0, \infty)$, given by

$$\Sigma_{\alpha,t}(x) := \int_{\mathbb{R}^n} (y - x) \otimes (y - x) K(x, y, t) d\alpha(y),$$

provided that the integral converges for each $x \in \mathbb{R}^d$ and $t > 0$.

It's important to show that this CTF is robust to noise and outliers. Therefore, they proved a stability theorem with respect to a proper distance between probability measures.

Wasserstein Distance

Definition

(Coupling) Let $\alpha, \beta \in \mathcal{P}(M)$ be two probability measures on M. Coupling measures α and β means constructing probability measure $\gamma \in \mathcal{P}(M \times M)$ on $M \times M$, in such a way that $p_{1*}(\gamma) = \alpha$, $p_{2*}(\gamma) = \beta$. Here $p_i : M \times M \to M$ is the projection onto i-th factor.
The set of all couplings of α, β is denoted by $\Gamma(\alpha, \beta)$.

Wasserstein Distance

Definition

(Coupling) Let $\alpha, \beta \in \mathcal{P}(M)$ be two probability measures on M. Coupling measures α and β means constructing probability measure $\gamma \in \mathcal{P}(M \times M)$ on $M \times M$, in such a way that $p_{1*}(\gamma) = \alpha$, $p_{2*}(\gamma) = \beta$. Here $p_i : M \times M \to M$ is the projection onto i-th factor.
The set of all couplings of α, β is denoted by $\Gamma(\alpha, \beta)$.

Definition

Let (M, d) be a Polish metric space, and let $p \in [1, \infty)$. For any two probability measures α and β on $\mathcal{X}$, the Wasserstein distance of order p between α and β is defined by the formula

$$W_p(\alpha, \beta) := \inf_{\gamma \in \Gamma(\alpha, \beta)} \left(\int \int d(z_1, z_2)^p \gamma(dz_1 \times dz_2) \right)^{1/p}$$

Stability Theorem of CTF on Euclidean Space

Definition

Let n be a positive integer and $f : [0, \infty) \to \mathbb{R}$ a bounded and measurable function satisfying:

1. $f(r) \geq 0$, $\forall r \in [0, \infty)$;
2. $M_n = \int_0^\infty r^{n/2-1} f(r) dr \leq \infty$;
3. There is $C > 0$ such that $rf(r) \leq C$, $\forall r \in [0, \infty)$.

The multiscale kernel $K : \mathbb{R}^n \times \mathbb{R}^n \times (0, \infty) \to \mathbb{R}$ associated with f is defined as

$$K(x, y, t) := \frac{1}{C_n(t)} f(\frac{\|y - x\|^2}{t^2}),$$

where $C_n(t) = \frac{1}{2} t^n M_n \omega_{n-1}$.

This is the stability theorem:

Theorem (Díaz Martínez, Mémoli, Mio)

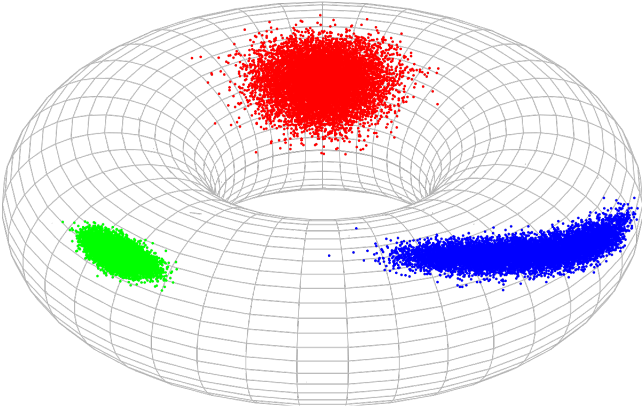
(Stability for Smooth Kernels) Let $f : [0, \infty) \to \mathbb{R}$ be as in definition 4 with multiscale kernel K. Suppose that f is differentiable and there exists a constant $A_1 > 0$ such that $r^{3/2}|f'(r)| \leq A_1$, $\forall r \geq 0$. Then, there is a constant $A_f > 0$, that depends only on f, such that

$$\sup_{x \in \mathbb{R}^n} \|\Sigma_{\alpha,t}(x) - \Sigma_{\alpha,t}(y)\| \leq \frac{tA_f}{C_n(t)} W_1(\alpha, \beta),$$

for any $\alpha, \beta \in \mathcal{P}_1(\mathbb{R}^n)$ and any $t > 0$.

Data on Riemannian Manifolds

How can we formulate a similar framework on more generalized Riemannian manifolds?



Riemannian Manifold

In our work, we are exploring extensions of CTF to Riemannian manifold (M, g).

In our work, we are exploring extensions of CTF to Riemannian manifold (M, g).

Definition

Let γ be a geodesic curve parametrized by arc length with $\gamma(0) = x$. If $t_0 = \sup\{t \in [0, \infty) | d(x, \gamma(s)) = s, \forall\ 0 < s \leq t\}$ is finite, we call $\gamma(t_0)$ the cut point of γ with respect to x. The union of all cut points is called the cut locus of x and is denoted by $C(x)$.

Riemannian Manifold

In our work, we are exploring extensions of CTF to Riemannian manifold (M, g).

Definition

Let γ be a geodesic curve parametrized by arc length with $\gamma(0) = x$. If $t_0 = \sup\{t \in [0, \infty) | d(x, \gamma(s)) = s, \forall\ 0 < s \leq t\}$ is finite, we call $\gamma(t_0)$ the cut point of γ with respect to x. The union of all cut points is called the cut locus of x and is denoted by $C(x)$.

Theorem

If there exist at least two different minimizing geodesics joining x to y, then $y \in C(x)$.

Riemannian Manifold

In our work, we are exploring extensions of CTF to Riemannian manifold (M, g).

Definition

Let γ be a geodesic curve parametrized by arc length with $\gamma(0) = x$. If $t_0 = \sup\{t \in [0, \infty) | d(x, \gamma(s)) = s, \forall\ 0 < s \leq t\}$ is finite, we call $\gamma(t_0)$ the cut point of γ with respect to x. The union of all cut points is called the cut locus of x and is denoted by $C(x)$.

Theorem

If there exist at least two different minimizing geodesics joining x to y, then $y \in C(x)$.

Theorem

$C(x)$ has measure zero with respect to the Riemannian measure μ on (M, g) for $\forall x \in M$.

Definition

For any $x \in M$, the largest r such that
$\exp_x : B(r) = \{v \in T_xM | \sqrt{< v, v >_g} < r\} \subset T_xM \to M$ is injective, is called the injectivity radius of x, denoted by $R(x)$.

Definition

For any $x \in M$, the largest r such that $\exp_x : B(r) = \{v \in T_xM | \sqrt{< v, v >_g} < r\} \subset T_xM \to M$ is injective, is called the injectivity radius of x, denoted by $R(x)$.

Definition

$diam(M) = \sup_{x,y \in M} d(x, y)$ is called the diameter of M.

Definition

For any $x \in M$, the largest r such that $\exp_x : B(r) = \{v \in T_xM | \sqrt{< v, v >_g} < r\} \subset T_xM \to M$ is injective, is called the injectivity radius of x, denoted by $R(x)$.

Definition

$diam(M) = \sup_{x,y \in M} d(x, y)$ is called the diameter of M.

Now we introduce the kernel functions on Riemannian manifold:

Definition

The multiscale kernel on M is a function $K : M \times M \times (0, \infty) \to \mathbb{R}$ which satisfies: (i) K(x,y,t)=K(y,x,t); (ii) As $t \to 0$, $K(\cdot, y, t) \to \delta_y$.

Heat Kernel

Definition

Heat kernel is a continuous function $k : M \times M \times (0, \infty) \to \mathbb{R}$ such that $\triangle_y k(x, y, t) + \frac{\partial}{\partial t} k(x, y, t) = 0$ and when we fix x, $k(x, \cdot, t)$ weakly converges to δ_x as $t \to 0$.

Heat Kernel

Definition

Heat kernel is a continuous function $k : M \times M \times (0, \infty) \to \mathbb{R}$ such that $\triangle_y k(x, y, t) + \frac{\partial}{\partial t} k(x, y, t) = 0$ and when we fix x, $k(x, \cdot, t)$ weakly converges to δ_x as $t \to 0$.

Theorem

Let (M, g) be a compact Riemannian manifold. Then the heat kernel function to the heat equation on (M, g) does exist. And the heat kernel is unique and symmetric in the two space variables.

Heat Kernel and Wasserstein Distance

Theorem

Given any Borel probability measure $\gamma \in \mathcal{P}(M)$ on compact Riemannian manifold (M, g), $u(x,t) = \int_M k(x,y,t)d\gamma(y)$ is a solution of the heat equation:

$$\begin{cases} \dfrac{\partial u}{\partial t} + \triangle u = 0 \\ u(t,\cdot) \xrightarrow[weakly]{t \to 0} \gamma \end{cases} \tag{1}$$

Heat Kernel and Wasserstein Distance

Theorem

Given any Borel probability measure $\gamma \in \mathcal{P}(M)$ on compact Riemannian manifold (M, g), $u(x,t) = \int_M k(x,y,t)d\gamma(y)$ is a solution of the heat equation:

$$\begin{cases} \dfrac{\partial u}{\partial t} + \triangle u = 0 \\ u(t,\cdot) \xrightarrow[weakly]{t\to 0} \gamma \end{cases} \tag{1}$$

Theorem

For any Borel probability measure $\gamma \in \mathcal{P}(M)$ on compact Riemannian manifold (M, g), the solution $u(x,t) = \int_M k(x,y,t)d\gamma(y)$ of heat equation 1 converges to γ in the sense of Wasserstein distance W_p $(p \geq 1)$ as $t \to 0$.

We review the definition on $\mathbb{R}^n$:

$$\Sigma_{\alpha,t}(x) := \int_{\mathbb{R}^n} (y-x) \otimes (y-x) K(x,y,t) d\alpha(y).$$

On a Riemannian manifold, generally we can't take the difference between two points. A natural idea is to replace $y - x$ by $\exp_x^{-1}(y) \in T_x M$.

Remark

$\exp_x^{-1}$ *may not be well-defined at the cut locus of* $x \in M$. *We set it to be* 0 *in that case.*

Definition

For any measure α on (M, g), we define the covariance tensor field associated with a kernel $K(x, y, t)$ to be

$$\Sigma_{\alpha,t}(x) := \int_M \exp_x^{-1}(y) \otimes \exp_x^{-1}(y) K(x, y, t) \alpha(dy),$$

provided that the integral converges for each $x \in M$ and $t > 0$.

Definition

For any measure α on (M, g), we define the covariance tensor field associated with a kernel $K(x, y, t)$ to be

$$\Sigma_{\alpha,t}(x) := \int_M \exp_x^{-1}(y) \otimes \exp_x^{-1}(y) K(x, y, t) \alpha(dy),$$

provided that the integral converges for each $x \in M$ and $t > 0$.

Remark

Suppose that (M, g) is a closed Riemannian manifold. $\alpha \in \mathcal{P}(M)$ is a probability measure. Then $\Sigma_{\alpha,t}$ is defined for $\forall x \in M$.

Since the behavior of $\exp_x^{-1}$ around the cut locus of x is hard to control, we did not find a way to prove a stability theorem like the Euclidean space.

Since the behavior of $\exp_x^{-1}$ around the cut locus of x is hard to control, we did not find a way to prove a stability theorem like the Euclidean space.
One way to fix this problem is to look at a class of very special kernel functions:

Definition

We say $h : M \times M \times (0, +\infty) \to \mathbb{R}$ is a kernel, away from cut locus, if h is a multiple kernel and for any fixed $x \in M$, the support of $h(x, \cdot)$ has no intersection with the cut locus $C(x)$.

Since the behavior of $\exp_x^{-1}$ around the cut locus of x is hard to control, we did not find a way to prove a stability theorem like the Euclidean space.
One way to fix this problem is to look at a class of very special kernel functions:

Definition

We say $h : M \times M \times (0, +\infty) \to \mathbb{R}$ is a kernel, away from cut locus, if h is a multiple kernel and for any fixed $x \in M$, the support of $h(x, \cdot)$ has no intersection with the cut locus $C(x)$.

Remark

For any $x \in M$ and $t > 0$, $V_{x,t}(y) := \exp_x^{-1}(y)\sqrt{h(x, y, t)}$ is continuous (or smooth) on M if $h(x, y, t)$ is continuous (or smooth) in the y variable. And we have $\Sigma_{\alpha,t}(x) = \int V_{x,t}(y) \otimes V_{x,t}(y) d\alpha(y)$.

Lemma

Suppose that M is compact and h is a continuous kernel away from cut locus. Then there exists $\kappa(t) > 0$ such that for any fixed $x \in M$, $\| \exp_x^{-1}(z_1)\sqrt{h(x, z_1, t)} - \exp_x^{-1}(z_2)\sqrt{h(x, z_2, t)}\| \leq \kappa(t) d(z_1, z_2)$.

Lemma

Suppose that M is compact and h is a continuous kernel away from cut locus. Then there exists $\kappa(t) > 0$ such that for any fixed $x \in M$, $\| \exp_x^{-1}(z_1)\sqrt{h(x, z_1, t)} - \exp_x^{-1}(z_2)\sqrt{h(x, z_2, t)}\| \leq \kappa(t)d(z_1, z_2)$.

Theorem

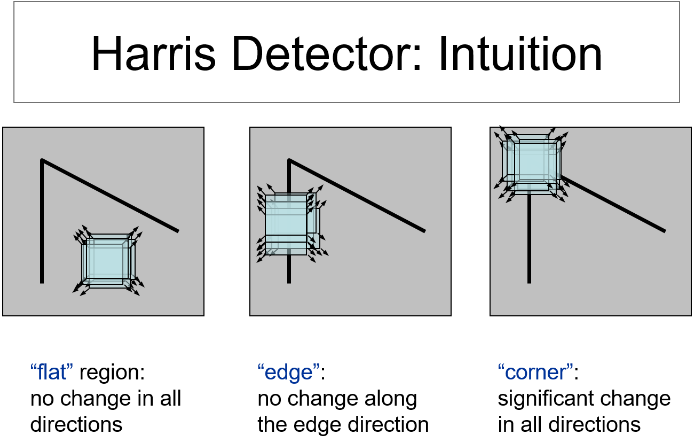
Let $d(M)$ be the diameter of compact Riemannian manifold (M, g). $h(x, y, t)$ is a continuous kernel away from the cut locus and $H(t) := \max_{x\in M, y\in M}\{h(x, y, t)\}$. If $\alpha, \beta \in \mathcal{P}_1(M)$, then the multiscale covariance tensor fields for α and β associated with the kernel h satisfy

$$\sup_{x\in M} \|\Sigma_{\alpha,t}(x) - \Sigma_{\beta,t}(x)\| \leq 2d(M)\kappa(t)\sqrt{H(t)}W_1(\alpha, \beta)$$

for any $t > 0$.

Structure CTFs

The structure tensor, also referred to as the second-moment matrix, is often used in computer vision and image processing.



- Given a random variable with smooth probability density function $f(x)$;

- Given a random variable with smooth probability density function $f(x)$;
- Shift f to the direction v, we get $v_*(f)(x) := f(x - v)$;

- Given a random variable with smooth probability density function $f(x)$;
- Shift f to the direction v, we get $v_*(f)(x) := f(x-v)$;
- Compare their square distance associated with kernel function $K(x,y,t)$:

$$\begin{aligned} d(f, v_*(f)) &= \int [f(y) - f(y-v)]^2 K(x,y,t) dy \\ &\approx \int < \nabla_x f(y), v > K(x,y,t) dy \\ &= \int \nabla_x f(y) \otimes \nabla_x f(y) K(x,y,t) dy (v,v) \end{aligned}$$

- Given a random variable with smooth probability density function $f(x)$;
- Shift f to the direction v, we get $v_*(f)(x) := f(x-v)$;
- Compare their square distance associated with kernel function $K(x,y,t)$:

$$
\begin{aligned}
d(f, v_*(f)) &= \int [f(y) - f(y-v)]^2 K(x,y,t)dy \\
&\approx \int < \nabla_x f(y), v > K(x,y,t)dy \\
&= \int \nabla_x f(y) \otimes \nabla_x f(y) K(x,y,t)dy(v,v)
\end{aligned}
$$

- We get a tensor $\Sigma := \int \nabla_x f(y) \otimes \nabla_x f(y) K(x,y,t)dy$

Definition

Let α be any distribution on Riemannian manifold M and $K(x, y, t)$ be a multiple kernel. We first convolve α with heat kernel to get a smooth heat solution $\alpha_s : M \to \mathbb{R}$. Then we have a covariance tensor field $\Sigma_{\alpha,s,t}(x) = \int_M \nabla\alpha_s(y) \otimes \nabla\alpha_s(y) K(x, y, t) d\mu(y)$.

Definition

Let α be any distribution on Riemannian manifold M and $K(x, y, t)$ be a multiple kernel. We first convolve α with heat kernel to get a smooth heat solution $\alpha_s : M \to \mathbb{R}$. Then we have a covariance tensor field $\Sigma_{\alpha,s,t}(x) = \int_M \nabla\alpha_s(y) \otimes \nabla\alpha_s(y) K(x, y, t) d\mu(y)$.

We need to work on the stability properties and some applications of this tensor field.

Some Thoughts on Our Definition

In some ways, these definitions are not very natural.

Some Thoughts on Our Definition

In some ways, these definitions are not very natural.

- $\int_M \exp_x^{-1}(y) \otimes \exp_x^{-1}(y) K(x, y, t) d\alpha(y)$

Some Thoughts on Our Definition

In some ways, these definitions are not very natural.

- $\int_M \exp_x^{-1}(y) \otimes \exp_x^{-1}(y) K(x, y, t) d\alpha(y)$
- $\int_M \nabla\alpha_s(y) \otimes \nabla\alpha_s(y) K(x, y, t) d\mu(y)$

Some Thoughts on Our Definition

In some ways, these definitions are not very natural.

- $\int_M \exp_x^{-1}(y) \otimes \exp_x^{-1}(y) K(x, y, t) d\alpha(y)$
- $\int_M \nabla\alpha_s(y) \otimes \nabla\alpha_s(y) K(x, y, t) d\mu(y)$

We were trying to relate CTF with the Hessian operator:

$$\Sigma_{\alpha,t}(x) := \int_M 4\,t^2\, Hess_x k(x, y, t) d\alpha(y) + 2\,t\,g \int_M k(x, y, t) d\alpha(y)$$

- The intuition behind the Harris detector suggests other ways of interpreting CTFs.

- The intuition behind the Harris detector suggests other ways of interpreting CTFs.
- Instead of moving distributions, why don't we move test functions?

- The intuition behind the Harris detector suggests other ways of interpreting CTFs.
- Instead of moving distributions, why don't we move test functions?

$$\begin{aligned} d(k, v_*(k)) &= \int_M (k(x,y,t) - k(x-v,y,t))^2 d\alpha(y) \\ &\approx \int_M < \nabla_x k(x,y,t), v > d\alpha(y) \\ &= \int_M \nabla_x k(x,y,t) \otimes \nabla_x k(x,y,t) d\alpha(y) \end{aligned}$$

In the Euclidean space, Gaussian kernel (heat kernel)

$$G(x, y, t) = \frac{1}{(4\pi t)^{n/2}} \exp\left(-\frac{\|x - y\|^2}{4t}\right)$$

In the Euclidean space, Gaussian kernel (heat kernel)

$$G(x, y, t) = \frac{1}{(4\pi t)^{n/2}} \exp(-\frac{\|x - y\|^2}{4t})$$

$$\nabla_x G(x, y, t) = \frac{1}{2t} G(x, y, t)(y - x)$$

In the Euclidean space, Gaussian kernel (heat kernel)

$$G(x,y,t) = \frac{1}{(4\pi t)^{n/2}} \exp(-\frac{\|x-y\|^2}{4t})$$

$$\nabla_x G(x,y,t) = \frac{1}{2t} G(x,y,t)(y-x)$$

$$\begin{aligned}\Sigma_{\alpha,t} &= \int \nabla_x G(x,y,t) \otimes \nabla_x G(x,y,t) d\alpha(y) \\ &= \int \frac{1}{4t^2}(y-x) \otimes (y-x) G^2(x,y,t) d\alpha(y)\end{aligned}$$

In the Riemannian manifold, Gaussian kernel

$$G(x, y, t) = \frac{1}{(4\pi t)^{n/2}} \exp(-\frac{d^2(x-y)}{4t})$$

In the Riemannian manifold, Gaussian kernel

$$G(x, y, t) = \frac{1}{(4\pi t)^{n/2}} \exp(-\frac{d^2(x-y)}{4t})$$

$$\nabla_x G(x, y, t) = \frac{1}{2t} G(x, y, t) \exp_x^{-1}(y)$$

In the Riemannian manifold, Gaussian kernel

$$G(x,y,t) = \frac{1}{(4\pi t)^{n/2}} \exp(-\frac{d^2(x-y)}{4t})$$

$$\nabla_x G(x,y,t) = \frac{1}{2t} G(x,y,t) \exp_x^{-1}(y)$$

$$\begin{aligned}\Sigma_{\alpha,t} &= \int \nabla_x G(x,y,t) \otimes \nabla_x G(x,y,t) d\alpha(y) \\ &= \int \frac{1}{4t^2} \exp_x^{-1}(y) \otimes \exp_x^{-1}(y) G^2(x,y,t) d\alpha(y)\end{aligned}$$

- More basically, we can show that the usual covariance tensor field may be expressed as

$$\Sigma_\alpha(x) = \int \nabla_x(\frac{d^2(x,y)}{2}) \otimes \nabla_x(\frac{d^2(x,y)}{2}) d\alpha(y),$$

where $d(x,y) = \|x - y\|$ is the Euclidean distance.

- More basically, we can show that the usual covariance tensor field may be expressed as

$$\Sigma_\alpha(x) = \int \nabla_x(\frac{d^2(x,y)}{2}) \otimes \nabla_x(\frac{d^2(x,y)}{2}) d\alpha(y),$$

  where $d(x,y) = \|x - y\|$ is the Euclidean distance.
- When we use diffusion distance $d_t(x,y) = \int (k(x,z,t) - k(y,z,t))^2 d\mu(z)$, and suppose that M is a homogeneous space, we get our multiscale covariance tensor filed.

Thank you for the attention!

Haibin Hang