Orientation-Preserving Vectorized Distance Between Curves

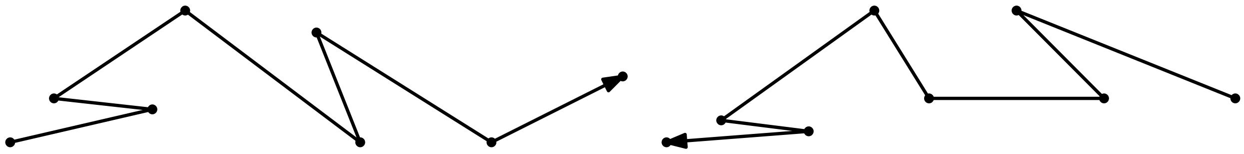
Jeff M. Phillips and Hasan Pourmahmood-Aghababa

School of Computing, University of Utah

August 2021

Distances on Trajectories and Their Efficiency

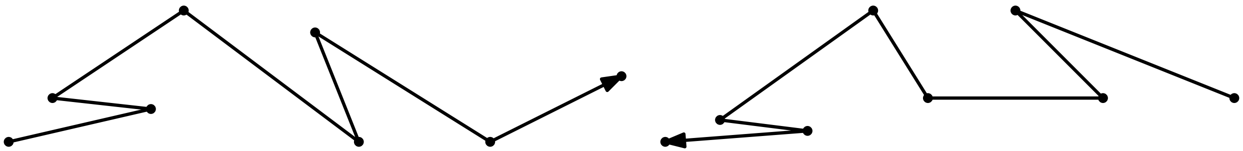
Consider two trajectories with m waypoints.



How can we measure their distance?

Distances on Trajectories and Their Efficiency

Consider two trajectories with m waypoints.



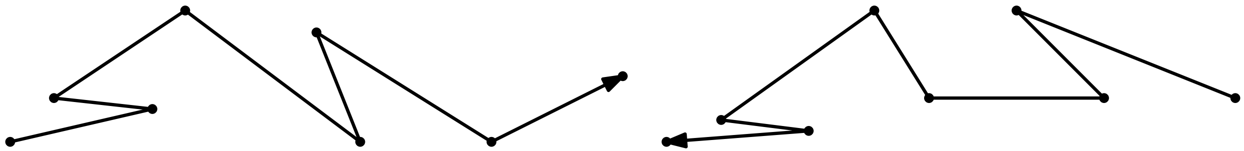
How can we measure their distance?

Popular Distances

- Hausdorff distance

Distances on Trajectories and Their Efficiency

Consider two trajectories with m waypoints.



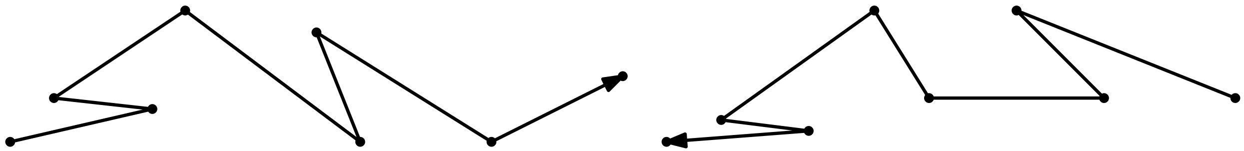
How can we measure their distance?

Popular Distances

- Hausdorff distance
  Does not preserve orientation; Complexity $\sim O(m)$.

Distances on Trajectories and Their Efficiency

Consider two trajectories with m waypoints.



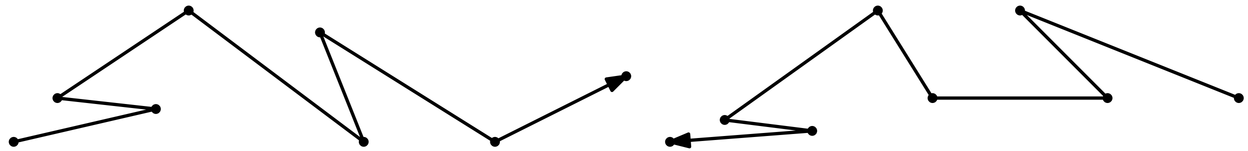
How can we measure their distance?

Popular Distances

- Hausdorff distance
  Does not preserve orientation; Complexity $\sim O(m)$.
- Dynamic Time Warping distance
- Fréchet distance
- Discrete Fréchet distance

Distances on Trajectories and Their Efficiency

Consider two trajectories with m waypoints.



How can we measure their distance?

Popular Distances

- Hausdorff distance
  Does not preserve orientation; Complexity $\sim O(m)$.
- Dynamic Time Warping distance
- Fréchet distance
- Discrete Fréchet distance
  All above 3 preserve orientation; Complexity $\sim O(m^2)$.

When m is large, all above distances are slow and so, e.g.

When m is large, all above distances are slow and so, e.g.

- KNN and almost all clustering algorithms with these distances are expensive.

When m is large, all above distances are slow and so, e.g.

- KNN and almost all clustering algorithms with these distances are expensive.
- More importantly, these metrics do not allow us to take the advantage of other ML algorithms as most of them need a vector/tensor/... as an input.

When m is large, all above distances are slow and so, e.g.

- KNN and almost all clustering algorithms with these distances are expensive.
- More importantly, these metrics do not allow us to take the advantage of other ML algorithms as most of them need a vector/tensor/... as an input.

Properties of a distance we are interested in:

When m is large, all above distances are slow and so, e.g.

- KNN and almost all clustering algorithms with these distances are expensive.
- More importantly, these metrics do not allow us to take the advantage of other ML algorithms as most of them need a vector/tensor/... as an input.

Properties of a distance we are interested in:

- It does not depend on m.
- It is as fast as calculating dot products in Euclidean spaces.
- It provides an embedding for curves in a Euclidean space in order to enable the use of ML algorithms.

- MinDist Vectorization [Phillips-Tang 2019]

Let γ be a curve and $q \in \mathbb{R}^2$. Then

▶ MinDist Vectorization [Phillips-Tang 2019]

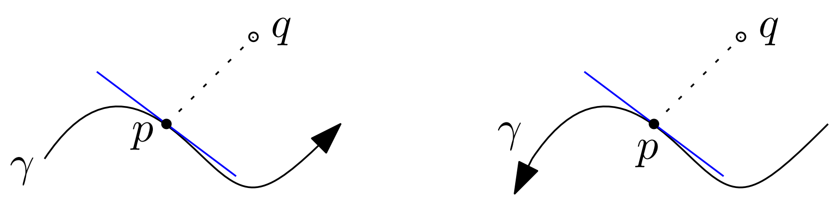
Let γ be a curve and $q \in \mathbb{R}^2$. Then

$$v_q^{\mathrm{mD}}(\gamma) = \mathrm{dist}(q, \gamma) = \|q - p\|.$$

▶ MinDist Vectorization [Phillips-Tang 2019]

Let γ be a curve and $q \in \mathbb{R}^2$. Then

$$v_q^{\mathrm{mD}}(\gamma) = \mathrm{dist}(q, \gamma) = \|q - p\|.$$

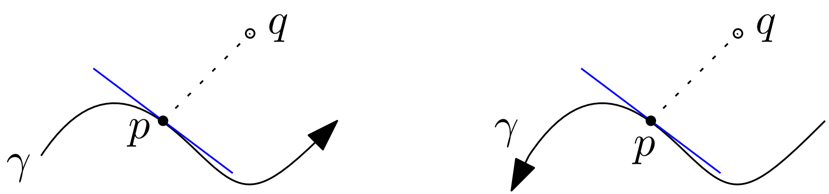


$p = \mathrm{argmin}_{p' \in \gamma} \|q - p'\|$

▶ MinDist Vectorization [Phillips-Tang 2019]

Let γ be a curve and $q \in \mathbb{R}^2$. Then

$$v_q^{\mathrm{mD}}(\gamma) = \mathrm{dist}(q, \gamma) = \|q - p\|.$$



$$p = \mathrm{argmin}_{p' \in \gamma} \|q - p'\|$$

For $Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2$, we get $v_Q^{\mathrm{mD}} : \{\mathrm{curves}\} \to \mathbb{R}^n$ by

$$v_Q^{\mathrm{mD}}(\gamma) = (v_{q_1}^{\mathrm{mD}}(\gamma), \cdots, v_{q_n}^{\mathrm{mD}}(\gamma)).$$

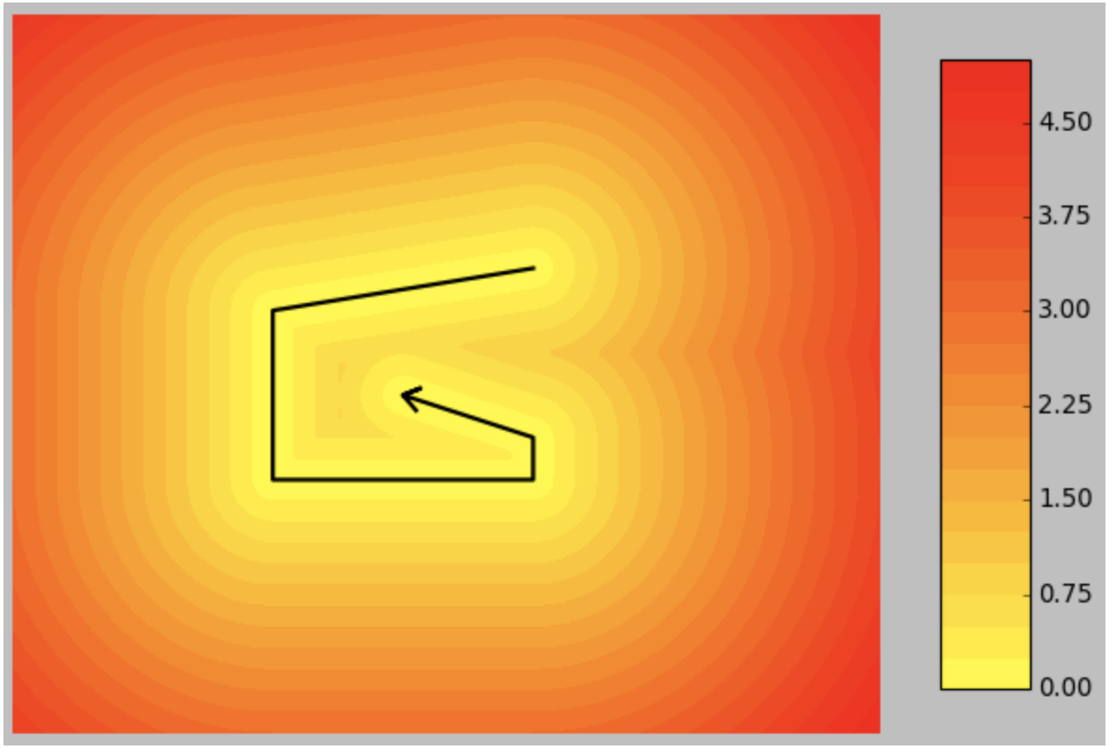


Figure: An example of MinDist function v_q^{mD} for a curve.

MinDist Distance

Let γ, γ' be two curves and $Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2$. Then

$$\mathrm{d}_Q^{\mathrm{mD}}(\gamma, \gamma') = \frac{1}{\sqrt{n}} \| v_Q^{\mathrm{mD}}(\gamma) - v_Q^{\mathrm{mD}}(\gamma') \|.$$

MinDist Distance

Let γ, γ' be two curves and $Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2$. Then

$$\mathrm{d}_Q^{\mathrm{mD}}(\gamma, \gamma') = \frac{1}{\sqrt{n}} \| v_Q^{\mathrm{mD}}(\gamma) - v_Q^{\mathrm{mD}}(\gamma') \|.$$

Complexity: $O(|Q|)$

MinDist Distance

Let γ, γ' be two curves and $Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2$. Then

$$\mathrm{d}_Q^{\mathrm{mD}}(\gamma, \gamma') = \frac{1}{\sqrt{n}} \| v_Q^{\mathrm{mD}}(\gamma) - v_Q^{\mathrm{mD}}(\gamma') \|.$$

Complexity: $O(|Q|)$

Question

How can we encode orientation preserving property into $\mathrm{d}_Q^{\mathrm{mD}}$?

Definitions and Notation

- Simple Curve

A non-self-crossing (possibly closed) curve.

Definitions and Notation

- Simple Curve

  A non-self-crossing (possibly closed) curve.

- Γ

  The class of all a.e. differentiable curves γ in $\mathbb{R}^2$ that have countably many number of self-crossings.

Definitions and Notation

- Simple Curve

A non-self-crossing (possibly closed) curve.

- Γ

The class of all a.e. differentiable curves γ in $\mathbb{R}^2$ that have countably many number of self-crossings.

- Γ'

The subset of Γ containing all simple curves.

Definitions and Notation

- Simple Curve

A non-self-crossing (possibly closed) curve.

- Γ

The class of all a.e. differentiable curves γ in $\mathbb{R}^2$ that have countably many number of self-crossings.

- Γ'

The subset of Γ containing all simple curves.

- n_p

Considering the direction of curve, n_p is the unique normal at p.

▶ SignedDist Function

Let $\gamma \in \Gamma$, $q \in \mathbb{R}^2$ and $\sigma > 0$ and set $p = \operatorname{argmin}_{p' \in \gamma} \|q - p'\|$.

▶ SignedDist Function

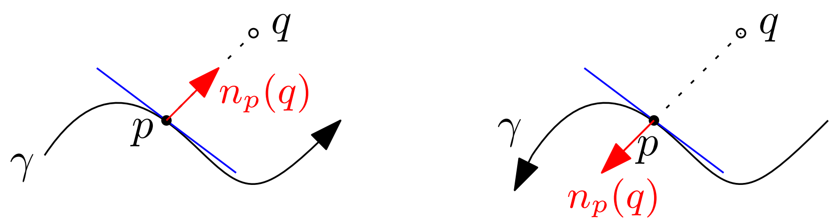
Let $\gamma \in \Gamma$, $q \in \mathbb{R}^2$ and $\sigma > 0$ and set $p = \operatorname{argmin}_{p' \in \gamma} \|q - p'\|$. If p is not an endpoint of γ, we define

$$v_q^\sigma(\gamma) = \frac{1}{\sigma}\langle n_p(q), q - p\rangle e^{-\frac{\|q-p\|^2}{\sigma^2}}.$$

▶ SignedDist Function

Let $\gamma \in \Gamma$, $q \in \mathbb{R}^2$ and $\sigma > 0$ and set $p = \mathrm{argmin}_{p' \in \gamma} \|q - p'\|$. If p is not an endpoint of γ, we define

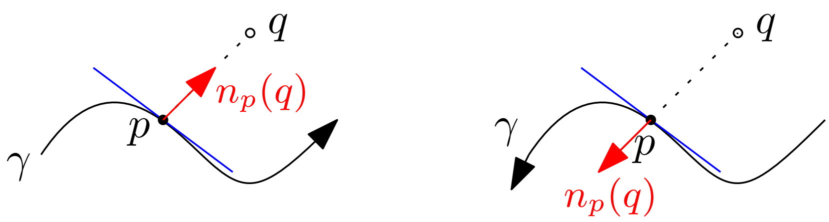
$$v_q^\sigma(\gamma) = \frac{1}{\sigma} \langle n_p(q), q - p \rangle e^{-\frac{\|q-p\|^2}{\sigma^2}}.$$



▶ SignedDist Function

Let $\gamma \in \Gamma$, $q \in \mathbb{R}^2$ and $\sigma > 0$ and set $p = \operatorname{argmin}_{p' \in \gamma} \|q - p'\|$. If p is not an endpoint of γ, we define

$$v_q^\sigma(\gamma) = \frac{1}{\sigma} \langle n_p(q), q - p \rangle e^{-\frac{\|q-p\|^2}{\sigma^2}}.$$



For endpoints we set

$$v_q^\sigma(\gamma) = \frac{1}{\sigma} \langle n_p, q - p \rangle \frac{\|q\|_{\infty,p}}{\|q-p\|} e^{-\frac{\|q-p\|^2}{\sigma^2}},$$

where $\|q\|_{\infty,p}$ is the l^∞-norm of q in the coordinate system with axis parallel to n_p and L (tangent line at p) and origin at p.

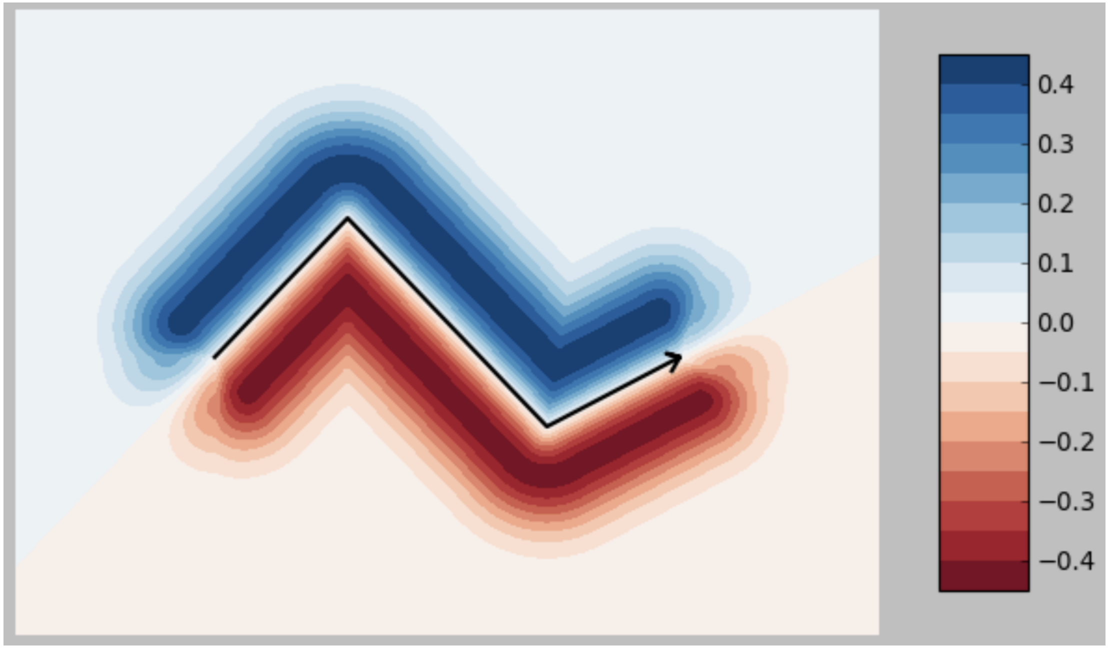


Figure: An example of v_q^σ function for a curve with sidedness encoded by positive/negative values.

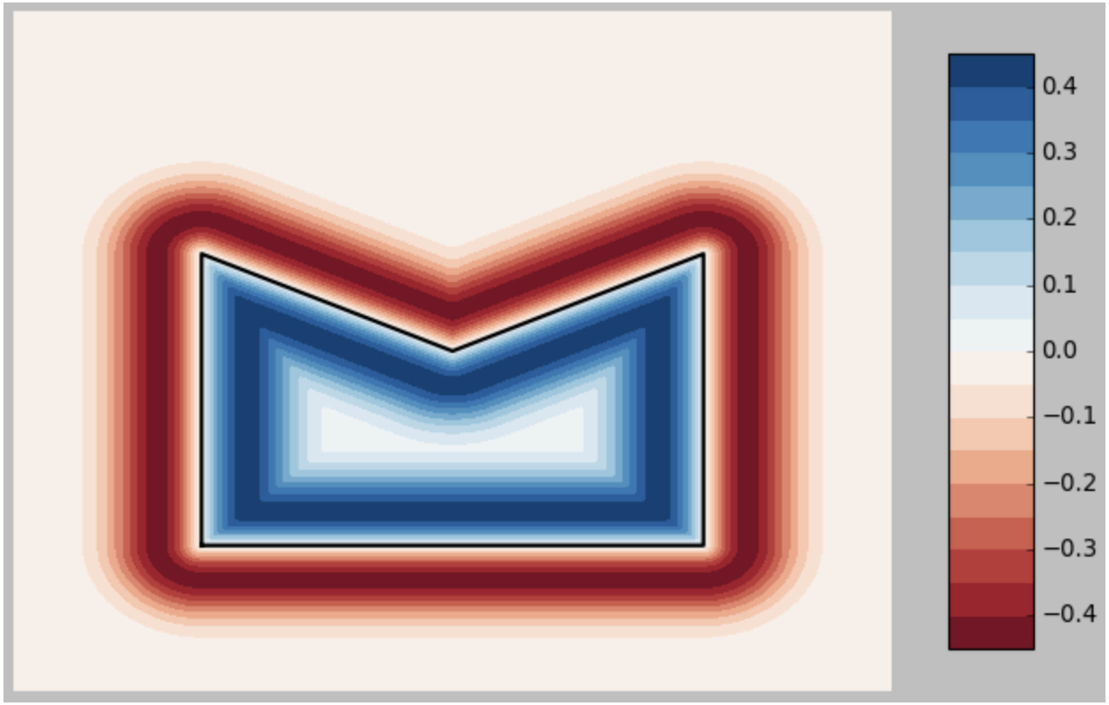


Figure: An example of v_q^σ function for a closed curve encoded by positive/negative values.

- SignedDist Vectorization

- SignedDist Vectorization
  - $Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2$,
  - $\sigma > 0$,
  - Define $v_Q^\sigma : \Gamma \to \mathbb{R}^n$ by

$$v_Q^\sigma(\gamma) = (v_{q_1}^\sigma(\gamma), \cdots, v_{q_n}^\sigma(\gamma)).$$

- SignedDist Vectorization
  - $Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2$,
  - $\sigma > 0$,
  - Define $v_Q^\sigma : \Gamma \to \mathbb{R}^n$ by

$$v_Q^\sigma(\gamma) = (v_{q_1}^\sigma(\gamma), \cdots, v_{q_n}^\sigma(\gamma)).$$

This embedding enables using ML algorithms, which is the biggest advantage of our work.

Orientation Preserving Distance

Orientation Preserving Distance

- $\gamma, \gamma' \in \Gamma$,
- $Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2$,
- $\sigma > 0$,

Orientation Preserving Distance

- $\gamma, \gamma' \in \Gamma$,
- $Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2$,
- $\sigma > 0$,

$$\mathrm{d}_Q^\sigma(\gamma, \gamma') = \frac{1}{\sqrt{n}} \| v_Q^\sigma(\gamma) - v_Q^\sigma(\gamma') \|.$$

Orientation Preserving Distance

- $\gamma, \gamma' \in \Gamma$,
- $Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2$,
- $\sigma > 0$,

$$\mathrm{d}_Q^\sigma(\gamma, \gamma') = \frac{1}{\sqrt{n}} \| v_Q^\sigma(\gamma) - v_Q^\sigma(\gamma') \|.$$

Complexity: $O(|Q|)$

Orientation Preserving Distance

- $\gamma, \gamma' \in \Gamma$,
- $Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2$,
- $\sigma > 0$,

$$\mathrm{d}_Q^\sigma(\gamma, \gamma') = \frac{1}{\sqrt{n}} \| v_Q^\sigma(\gamma) - v_Q^\sigma(\gamma') \|.$$

Complexity: $O(|Q|)$

Metric Property

Assuming Q is dense enough, d_Q^σ is a metric on Γ.

Orientation Preserving Distance

- $\gamma, \gamma' \in \Gamma$,
- $Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2$,
- $\sigma > 0$,

$$\mathtt{d}_Q^\sigma(\gamma, \gamma') = \frac{1}{\sqrt{n}} \| v_Q^\sigma(\gamma) - v_Q^\sigma(\gamma') \|.$$

Complexity: $O(|Q|)$

Metric Property

Assuming Q is dense enough, $\mathtt{d}_Q^\sigma$ is a metric on Γ.

In Practice

In practice, however, we found that $|Q| = 20$ is usually enough.

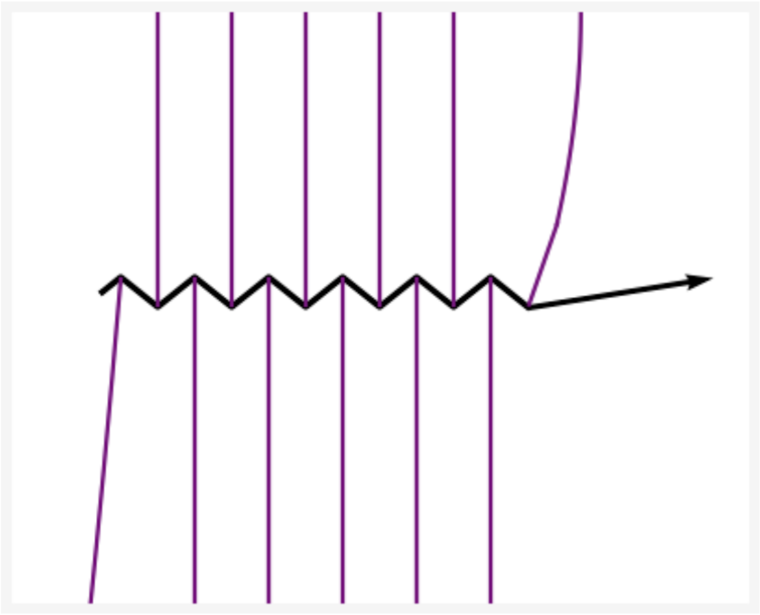


Figure: Medial axis in purple

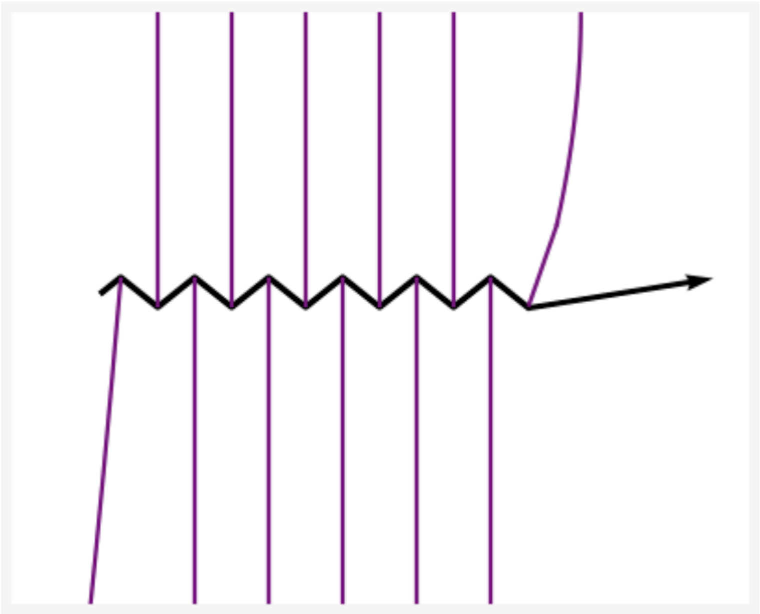


Figure: Medial axis in purple

As we can see, the MA captures the noise in a curve.

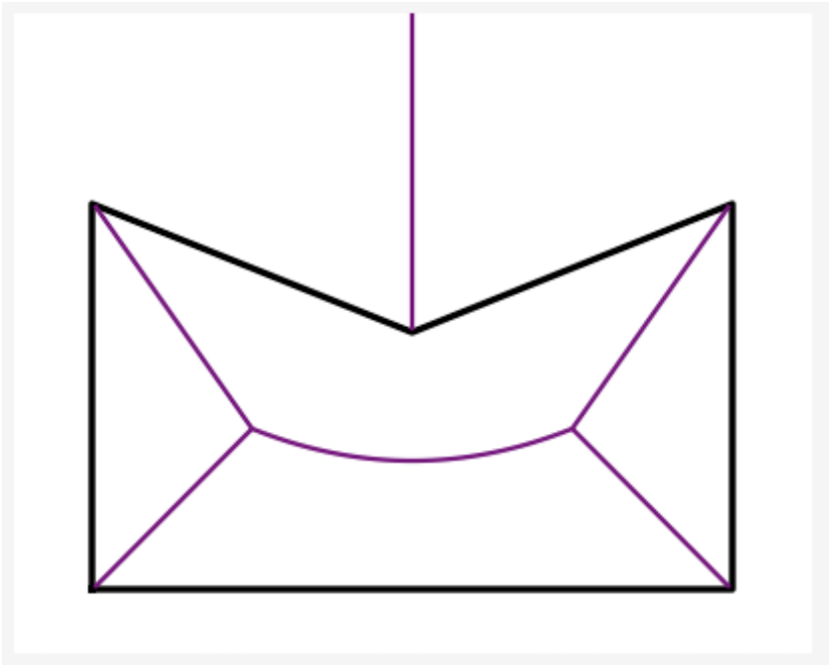


Figure: Medial axis in purple

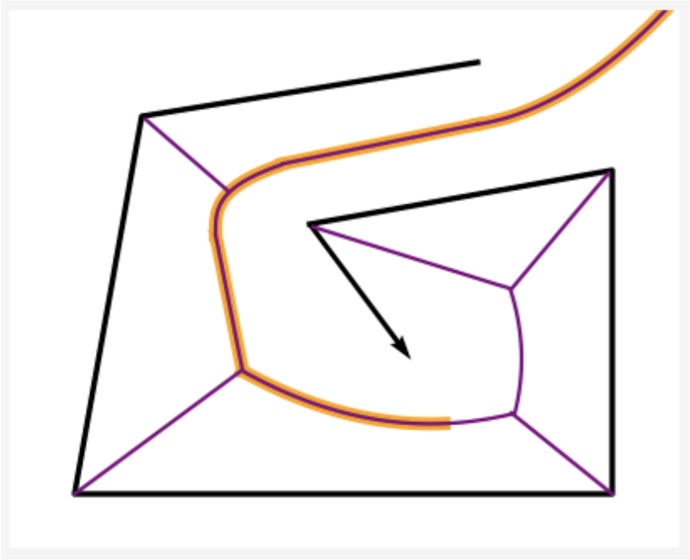


Figure: Medial axis in purple and Signed medial axis in orange

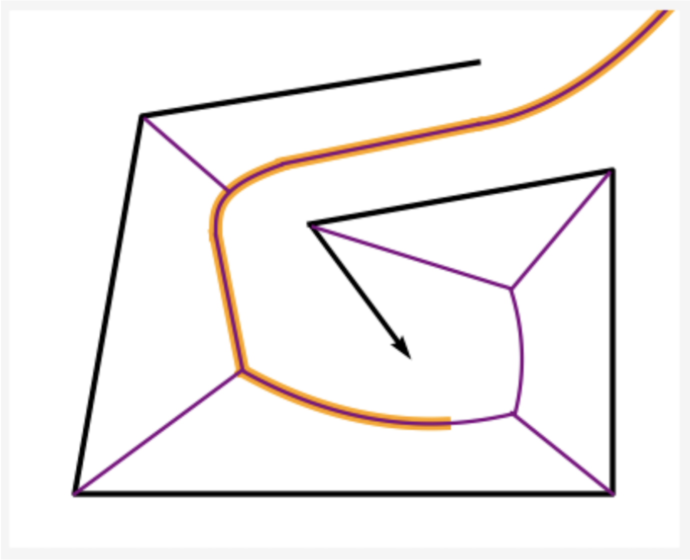


Figure: Medial axis in purple and Signed medial axis in orange

The notion of SMA comes to play to capture the sidedness of curve, i.e. capturing equidistance points with different v_q^σ signs.

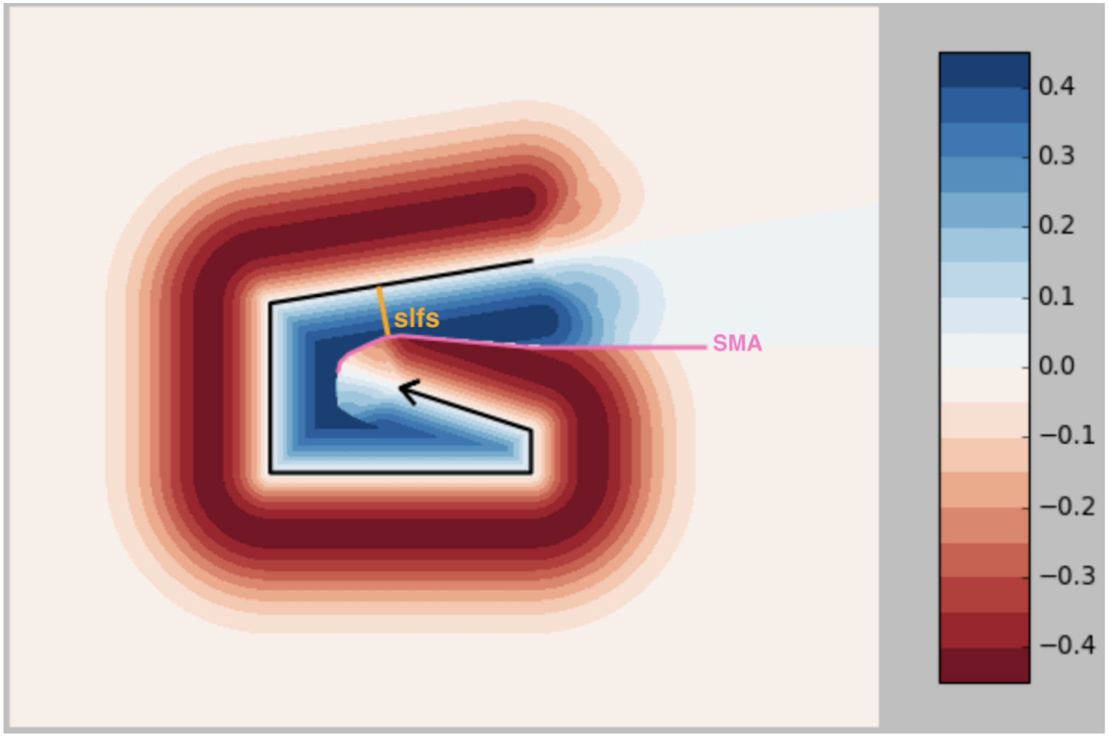


Figure: Signed medial axis in pink and Signed local feature size in orange

Landmark stability

Landmark stability

Under some slfs-related mild conditions on q and q', for $\gamma \in \Gamma'$, we have

$$|v_q^\sigma(\gamma) - v_{q'}^\sigma(\gamma)| \leq \frac{1}{\sigma}\|q - q'\|.$$

Landmark stability

Under some slfs-related mild conditions on q and q', for $\gamma \in \Gamma'$, we have

$$|v_q^\sigma(\gamma) - v_{q'}^\sigma(\gamma)| \le \frac{1}{\sigma}\|q - q'\|.$$

Curve Stability

Landmark stability

Under some slfs-related mild conditions on q and q', for $\gamma \in \Gamma'$, we have

$$|v_q^\sigma(\gamma) - v_{q'}^\sigma(\gamma)| \leq \frac{1}{\sigma}\|q - q'\|.$$

Curve Stability

Under some slfs-related conditions on the place of q_i's, for $\gamma, \gamma' \in \Gamma'$ we have

$$\sigma \mathtt{d}_Q^\sigma(\gamma, \gamma') \leq \mathtt{d}_F(\gamma, \gamma').$$

Landmark stability

Under some slfs-related mild conditions on q and q', for $\gamma \in \Gamma'$, we have

$$|v_q^\sigma(\gamma) - v_{q'}^\sigma(\gamma)| \leq \frac{1}{\sigma}\|q - q'\|.$$

Curve Stability

Under some slfs-related conditions on the place of q_i's, for $\gamma, \gamma' \in \Gamma'$ we have

$$\sigma \mathtt{d}_Q^\sigma(\gamma, \gamma') \leq \mathtt{d}_F(\gamma, \gamma').$$

In contrast, $\mathtt{d}_Q^{\mathrm{mD}}$ relates to Hausdorff distance $\mathtt{d}_H$:

Landmark stability

Under some slfs-related mild conditions on q and q', for $\gamma \in \Gamma'$, we have

$$|v_q^\sigma(\gamma) - v_{q'}^\sigma(\gamma)| \leq \frac{1}{\sigma}\|q - q'\|.$$

Curve Stability

Under some slfs-related conditions on the place of q_i's, for $\gamma, \gamma' \in \Gamma'$ we have

$$\sigma \mathtt{d}_Q^\sigma(\gamma, \gamma') \leq \mathtt{d}_F(\gamma, \gamma').$$

In contrast, $\mathtt{d}_Q^{\mathrm{mD}}$ relates to Hausdorff distance $\mathtt{d}_H$:

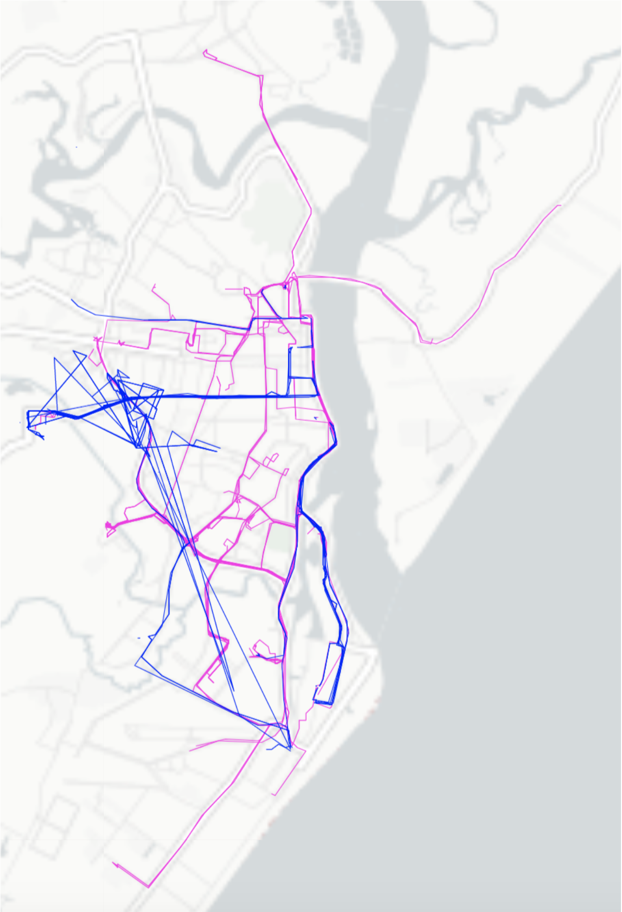
Theorem

$$\mathtt{d}_Q^{\mathrm{mD}}(\gamma, \gamma') \leq \mathtt{d}_H(\gamma, \gamma').$$

Average test errors with v_Q^{σ} and v_Q^{mD} vectorizations

	Feature Mapping	v_Q^{σ}	v_Q^{mD}
	Classifier	**Test Error**	**Test Error**
Car-Bus	Linear SVM	0.361	0.361
	Gaussian SVM	0.225	0.302
	Decision Tree	0.230	0.212
	Random Forest	**0.157**	0.183
Characters	Linear SVM	0.018	0.040
	Gaussian SVM	0.012	0.038
	Decision Tree	0.018	0.074
	Random Forest	**0.010**	0.049
Pigeons	Linear SVM	**0.003**	0.506
	Gaussian SVM	0.006	0.517
	Decision Tree	0.016	0.516
	Random Forest	0.006	0.524

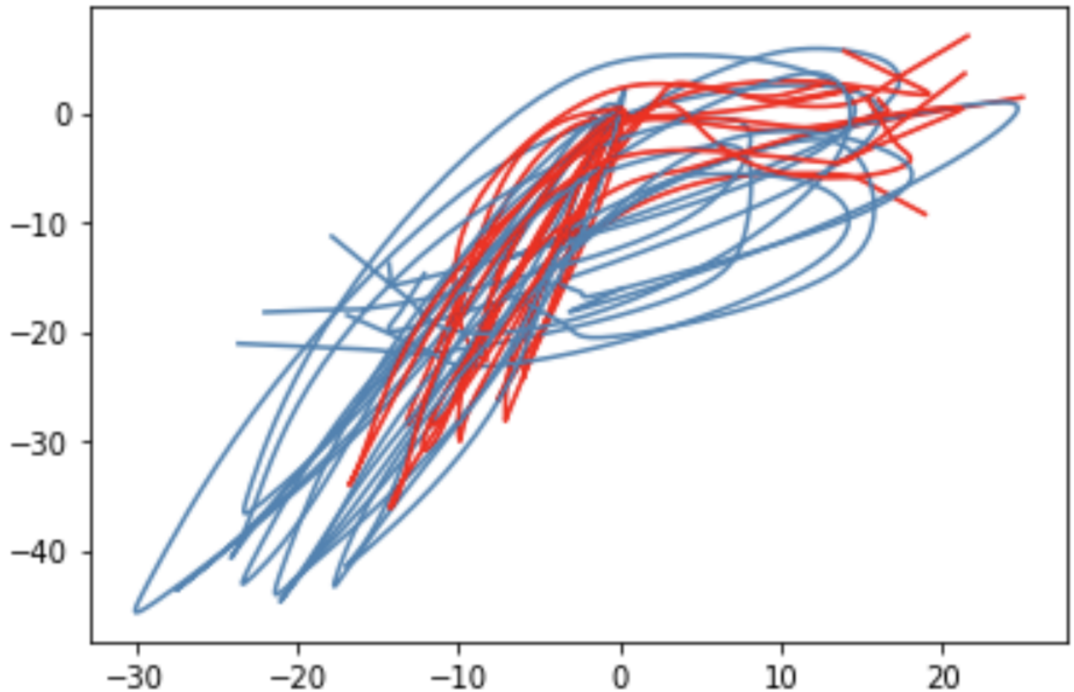
Buses: blue, Cars: pink



Average test errors with v_Q^{σ} and v_Q^{mD} vectorizations (car-bus)

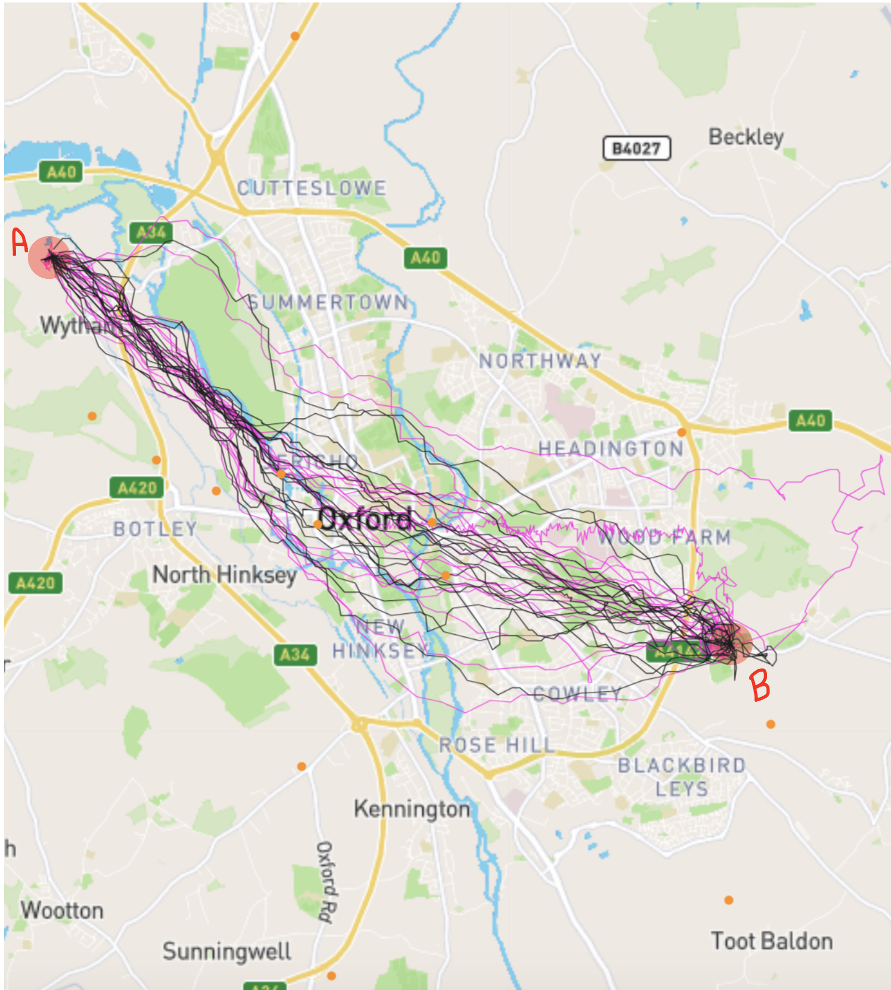
	Feature Mapping	v_Q^{σ}	v_Q^{mD}
	Classifier	**Test Error**	**Test Error**
Car-Bus	Linear SVM	0.361	0.361
	Gaussian SVM	0.225	0.302
	Decision Tree	0.230	0.212
	Random Forest	**0.157**	0.183

Letters p and r



Average test errors with v_Q^{σ} and v_Q^{mD} vectorizations (Characters)

	Feature Mapping	v_Q^{σ}	v_Q^{mD}
	Classifier	**Test Error**	**Test Error**
Characters	Linear SVM	0.018	0.040
	Gaussian SVM	0.012	0.038
	Decision Tree	0.018	0.074
	Random Forest	**0.010**	0.049


A
B
A40
A34
CUTTESLOWE
A40
B4027
Beckley
SUMMERTOWN
NORTHWAY
A40
HEADINGTON
A420
BOTLEY
Oxford
WOOD FARM
A420
North Hinksey
NEW HINKSEY
A34
COWLEY
ROSE HILL
BLACKBIRD LEYS
Kennington
Oxford Rd
Wootton
Sunningwell
A34
Toot Baldon

Average test errors with v_Q^σ and v_Q^{mD} vectorizations (Pigeons)

	Feature Mapping	v_Q^σ	v_Q^{mD}
	Classifier	**Test Error**	**Test Error**
Pigeons	Linear SVM	**0.003**	0.506
	Gaussian SVM	0.006	0.517
	Decision Tree	0.016	0.516
	Random Forest	0.006	0.524

Using the implemented codes

The related package is available in Python package index (PyPI) via the package

trjtrypy

Just: pip install trjtrypy

Jeff M. Phillips and Hasan Pourmahmood-Aghababa, Orientation-Preserving Vectorized Distance Between Curves, Mathematical and Scientific Machine Learning (MSML), August 2021.

Jeff M. Phillips and Pingfan Tang, Simple Distances for Trajectories via Landmarks, ACM GIS SIGSPATIAL, 2019.

Jeff M. Phillips and Pingfan Tang, Sketched MinDist, International Symposium on Computational Geometry, 2020.

Thank You!