An SDP relaxation for computing distances between metric spaces

Soledad Villar, UT Austin

Ongoing work with:
Afonso Bandeira (MIT, NYU)
Andrew Blumberg (UT Austin)
Rachel Ward (UT Austin)

SIAM Imaging

May 23, 2016

Outline and References

- Gromov-Hausdorff distance.
  - Facundo Mémoli. On the use of Gromov-Hausdorff Distances for Shape Comparison.
- SDP relaxation of Gromov-Hausdorff distance.
  - Ongoing work with Afonso Bandeira, Andrew Blumberg and Rachel Ward.
- Numerical performance on real data.
  - Boyer, Lipman, Daubechies, et al. Algorithms to automatically quantify the geometric similarity of anatomical surfaces.
  - Yang, Sun and Toh. SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints.
- Future work.

Outline and References

- Gromov-Hausdorff distance.
  - Facundo Mémoli. On the use of Gromov-Hausdorff Distances for Shape Comparison.
- SDP relaxation of Gromov-Hausdorff distance.
  - Ongoing work with Afonso Bandeira, Andrew Blumberg and Rachel Ward.
- Numerical performance on real data.
  - Boyer, Lipman, Daubechies, et al. Algorithms to automatically quantify the geometric similarity of anatomical surfaces.
  - Yang, Sun and Toh. SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints.
- Future work.

Outline and References

- Gromov-Hausdorff distance.
  - Facundo Mémoli. On the use of Gromov-Hausdorff Distances for Shape Comparison.
- SDP relaxation of Gromov-Hausdorff distance.
  - Ongoing work with Afonso Bandeira, Andrew Blumberg and Rachel Ward.
- Numerical performance on real data.
  - Boyer, Lipman, Daubechies, et al. Algorithms to automatically quantify the geometric similarity of anatomical surfaces.
  - Yang, Sun and Toh. SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints.
- Future work.

Outline and References

- Gromov-Hausdorff distance.
  - Facundo Mémoli. On the use of Gromov-Hausdorff Distances for Shape Comparison.
- SDP relaxation of Gromov-Hausdorff distance.
  - Ongoing work with Afonso Bandeira, Andrew Blumberg and Rachel Ward.
- Numerical performance on real data.
  - Boyer, Lipman, Daubechies, et al. Algorithms to automatically quantify the geometric similarity of anatomical surfaces.
  - Yang, Sun and Toh. SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints.
- Future work.

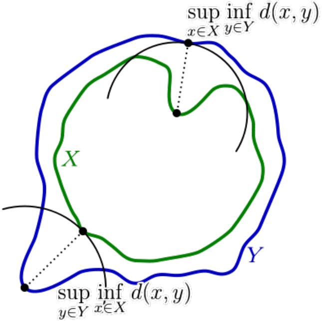
Outline and References

- Gromov-Hausdorff distance.
  - Facundo Mémoli. On the use of Gromov-Hausdorff Distances for Shape Comparison.
- SDP relaxation of Gromov-Hausdorff distance.
  - Ongoing work with Afonso Bandeira, Andrew Blumberg and Rachel Ward.
- Numerical performance on real data.
  - Boyer, Lipman, Daubechies, et al. Algorithms to automatically quantify the geometric similarity of anatomical surfaces.
  - Yang, Sun and Toh. SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints.
- Future work.

The Hausdorff distance

Let X, Y compact sets of a metric space. Define

$$\begin{aligned} d(x, Y) &= \inf\{d(x,y) : y \in Y\} \\ d(X, Y) &= \sup\{d(x,Y) : x \in X\} \end{aligned}$$



$$d_H(X, Y) = \max\{d(X, Y), d(Y, X)\}$$

The Gromov-Hausdorff distance

Let X, Y compact metric spaces. Define

$$d_{GH}(X, Y) = \inf_{Z,f,g} d_H(f(X), g(Y))$$

where $f : X \to Z, \quad g : Y \to Z$ are isometric embeddings.

Fact: $d_{GH}(X, Y) = 0$ if and only if X, Y are isometric

The Gromov-Hausdorff distance

Let X, Y compact metric spaces. Define

$$d_{GH}(X, Y) = \inf_{Z,f,g} d_H(f(X), g(Y))$$

where $f : X \to Z, \quad g : Y \to Z$ are isometric embeddings.

Fact: $d_{GH}(X, Y) = 0$ if and only if X, Y are isometric

Gromov-Hausdorff distance in point clouds

Motivations

- Shape comparison of geometric objects (like surfaces).
- Applications to spaces where geometry is less apparent (like tree spaces, dna sequences, etc).

Quadratic assignment formulation (Mémoli 2007)

Let $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_m\}$.

Let $R \subset X \times Y$ and $\delta_{ij} = \begin{cases} 1 & \text{if } (x_i, y_j) \in R \\ 0 & \text{otherwise} \end{cases}$

Let $\Gamma_{ik,jl} = |d_X(x_i, x_k) - d_Y(y_j, y_l)|$

$$d_{GH}(X, Y) = \frac{1}{2} \quad \min_R \quad \max_{ik,jl} \Gamma_{ik,jl} \delta_{ij} \delta_{kl}$$

$$\text{subject to} \quad \delta_{ij} \in \{0, 1\}, \ \sum_{j=1}^{m} \delta_{ij} \geq 1, \ \sum_{i=1}^{n} \delta_{ij} \geq 1$$

Fact: Computing GH distance is NP-hard and also, to approximate it better than a factor of 3 is NP-hard (Agarwal et al 2015).

Quadratic assignment formulation (Mémoli 2007)

Let $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_m\}$.

Let $R \subset X \times Y$ and $\delta_{ij} = \begin{cases} 1 & \text{if } (x_i, y_j) \in R \\ 0 & \text{otherwise} \end{cases}$

Let $\Gamma_{ik,jl} = |d_X(x_i, x_k) - d_Y(y_j, y_l)|$

$$d_{GH}(X, Y) = \frac{1}{2} \qquad \min_R \qquad \max_{ik,jl} \Gamma_{ik,jl}\delta_{ij}\delta_{kl}$$

$$\text{subject to} \quad \delta_{ij} \in \{0, 1\}, \ \sum_{j=1}^{m} \delta_{ij} \geq 1, \ \sum_{i=1}^{n} \delta_{ij} \geq 1$$

Fact: Computing GH distance is NP-hard and also, to approximate it better than a factor of 3 is NP-hard (Agarwal et al 2015).

Quadratic assignment formulation (Mémoli 2007)

Let $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_m\}$.

Let $R \subset X \times Y$ and $\delta_{ij} = \begin{cases} 1 & \text{if } (x_i, y_j) \in R \\ 0 & \text{otherwise} \end{cases}$

Let $\Gamma_{ik,jl} = |d_X(x_i, x_k) - d_Y(y_j, y_l)|$

$$d_{GH}(X, Y) = \frac{1}{2} \qquad \min_R \qquad \max_{ik,jl} \Gamma_{ik,jl} \delta_{ij} \delta_{kl}$$

$$\text{subject to} \quad \delta_{ij} \in \{0, 1\}, \ \sum_{j=1}^{m} \delta_{ij} \geq 1, \ \sum_{i=1}^{n} \delta_{ij} \geq 1$$

Fact: Computing GH distance is NP-hard and also, to approximate it better than a factor of 3 is NP-hard (Agarwal et al 2015).

Gromov-Wasserstein distance (Mémoli 2007)

$$D_{GW,p}(X,Y) = \frac{1}{2}\left(\inf_{\delta}\sum_{i,j}\sum_{k,l}\Gamma^{p}_{ik,jl}\delta_{ij}\delta_{kl}\right)^{1/p} \quad \sum_{i}\delta_{ij}=1, \sum_{j}\delta_{ij}=1$$

- In this formulation δ's are thought as probability measures on the set of points ($\delta' s$ do not come from a map).
- The max is changed for a sum.
- Mémoli considers a spectral relaxation of the Gromov-Wasserstein distance using heat kernels.

An SDP relaxation of the Gromov-Hausdorff distance

Focus on case $|X| = |Y|$.

$$\tilde{d}(X,Y) = \quad \min_Z \quad \sum_{i,j,k,l} \Gamma_{ik,jl} Z_{ij,kl}$$

$$\text{subject to} \quad \sum_{i=1} X_{ij,ij} = 1 \text{ for all } j$$

$$\sum_{j=1} Z_{ij,ij} = 1 \;\text{ for all } i$$

$$Z_{ij,il} = 0, \text{ for all } i,j,l \text{ with } l \neq j,$$

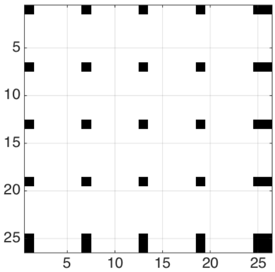
$$Z_{ij,kj} = 0, \text{ for all } i,j,k \text{ with } i \neq k,$$

$$Z_{ij,kl} \geq 0, \quad Z \succeq 0$$

$$\sum_i Z_{ij,N^2+1} = 1 \text{ for all } j,$$

$$\sum_j Z_{ij,N^2+1} = 1 \text{ for all } i,$$

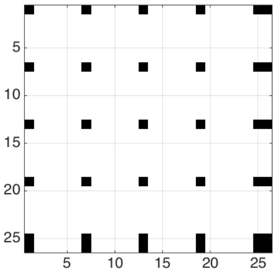
$$Z_{N^2+1,N^2+1} = 1$$



An SDP relaxation of the Gromov-Hausdorff distance

Focus on case $|X| = |Y|$.

$$
\begin{aligned}
\tilde{d}(X,Y) = \quad \min_Z \quad & \sum_{i,j,k,l} \Gamma_{ik,jl} Z_{ij,kl} \\
\text{subject to} \quad & \sum_{i=1} X_{ij,ij} = 1 \text{ for all } j \\
& \sum_{j=1} Z_{ij,ij} = 1 \text{ for all } i \\
& Z_{ij,il} = 0, \text{ for all } i,j,l \text{ with } l \neq j, \\
& Z_{ij,kj} = 0, \text{ for all } i,j,k \text{ with } i \neq k, \\
& Z_{ij,kl} \geq 0, \quad Z \succeq 0 \\
& \sum_{i} Z_{ij,N^2+1} = 1 \text{ for all } j, \\
& \sum_{j} Z_{ij,N^2+1} = 1 \text{ for all } i, \\
& Z_{N^2+1,N^2+1} = 1
\end{aligned}
$$



Some properties of $\tilde{d}$

- $\tilde{d}(X, Y)$ is a lower bound for a distance between metric spaces.
- $\tilde{d}(X, Y) \leq \tilde{d}(X, W) + \tilde{d}(W, Y)$
- $\tilde{d}(X, Y) = 0$ if X and Y are isometric, and the SDP finds the isometry.
  - Numerically we observe stability with respect to noise.
- $\tilde{d}(X, Y)$ may be 0 for non isometric X and Y.
  - Graph isomorphism problem can be posed as deciding whether GH distance between graphs is zero.

Some properties of $\tilde{d}$

- $\tilde{d}(X, Y)$ is a lower bound for a distance between metric spaces.
- $\tilde{d}(X, Y) \leq \tilde{d}(X, W) + \tilde{d}(W, Y)$
- $\tilde{d}(X, Y) = 0$ if X and Y are isometric, and the SDP finds the isometry.
  - Numerically we observe stability with respect to noise.
- $\tilde{d}(X, Y)$ may be 0 for non isometric X and Y.
  - Graph isomorphism problem can be posed as deciding whether GH distance between graphs is zero.

Some properties of $\tilde{d}$

- $\tilde{d}(X,Y)$ is a lower bound for a distance between metric spaces.
- $\tilde{d}(X,Y) \leq \tilde{d}(X,W) + \tilde{d}(W,Y)$
- $\tilde{d}(X,Y) = 0$ if X and Y are isometric, and the SDP finds the isometry.
  - Numerically we observe stability with respect to noise.
- $\tilde{d}(X,Y)$ may be 0 for non isometric X and Y.
  - Graph isomorphism problem can be posed as deciding whether GH distance between graphs is zero.

Some properties of $\tilde{d}$

- $\tilde{d}(X, Y)$ is a lower bound for a distance between metric spaces.
- $\tilde{d}(X, Y) \leq \tilde{d}(X, W) + \tilde{d}(W, Y)$
- $\tilde{d}(X, Y) = 0$ if X and Y are isometric, and the SDP finds the isometry.
  - Numerically we observe stability with respect to noise.
- $\tilde{d}(X, Y)$ may be 0 for non isometric X and Y.
  - Graph isomorphism problem can be posed as deciding whether GH distance between graphs is zero.

Some properties of $\tilde{d}$

- $\tilde{d}(X, Y)$ is a lower bound for a distance between metric spaces.
- $\tilde{d}(X, Y) \leq \tilde{d}(X, W) + \tilde{d}(W, Y)$
- $\tilde{d}(X, Y) = 0$ if X and Y are isometric, and the SDP finds the isometry.
  - Numerically we observe stability with respect to noise.
- $\tilde{d}(X, Y)$ may be 0 for non isometric X and Y.
  - Graph isomorphism problem can be posed as deciding whether GH distance between graphs is zero.

Some properties of $\tilde{d}$

- $\tilde{d}(X, Y)$ is a lower bound for a distance between metric spaces.
- $\tilde{d}(X, Y) \leq \tilde{d}(X, W) + \tilde{d}(W, Y)$
- $\tilde{d}(X, Y) = 0$ if X and Y are isometric, and the SDP finds the isometry.
  - Numerically we observe stability with respect to noise.
- $\tilde{d}(X, Y)$ may be 0 for non isometric X and Y.
  - Graph isomorphism problem can be posed as deciding whether GH distance between graphs is zero.

Numerical considerations

Computing $\tilde{d}$ involves solving a big SDP!

- Improving SDP solvers is an active research area.
- Work around with sampling, good initializations, etc.
- SDP's and dual certificates can be used to obtain fast algorithms (see Dustin's talk tomorrow).

Numerical considerations

Computing $\tilde{d}$ involves solving a big SDP!

- Improving SDP solvers is an active research area.
- Work around with sampling, good initializations, etc.
- SDP's and dual certificates can be used to obtain fast algorithms (see Dustin's talk tomorrow).

Numerical considerations

Computing $\tilde{d}$ involves solving a big SDP!

- Improving SDP solvers is an active research area.
- Work around with sampling, good initializations, etc.
- SDP's and dual certificates can be used to obtain fast algorithms (see Dustin's talk tomorrow).

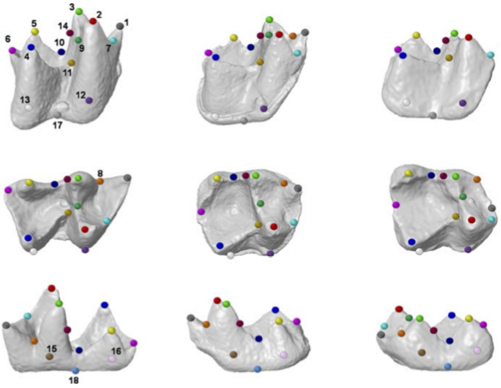
Numerical considerations

Computing $\tilde{d}$ involves solving a big SDP!

- Improving SDP solvers is an active research area.
- Work around with sampling, good initializations, etc.
- SDP's and dual certificates can be used to obtain fast algorithms (see Dustin's talk tomorrow).

Real data application

Boyer, Lipman, Daubechies, et al. Algorithms to automatically quantify the geometric similarity of anatomical surfaces.



Real data application

Objective: teeth classification.
Two methods:

1. Lipman and Daubechies map the teeth surfaces to the hyperbolic disk and consider a Wesserstein distance that is invariant under conformal transformations.
2. Boyer labels 18 landmarks on each teeth. Then they find the best rigid transformation to match the labeled landmarks.

Assessment: They consider 116 teeth. For each teeth find the closest teeth according to each distance, and see whether they are in the same category.

Real data application

Objective: teeth classification.
Two methods:

1. Lipman and Daubechies map the teeth surfaces to the hyperbolic disk and consider a Wesserstein distance that is invariant under conformal transformations.
2. Boyer labels 18 landmarks on each teeth. Then they find the best rigid transformation to match the labeled landmarks.

Assessment: They consider 116 teeth. For each teeth find the closest teeth according to each distance, and see whether they are in the same category.

Real data application

Objective: teeth classification.
Two methods:

1. Lipman and Daubechies map the teeth surfaces to the hyperbolic disk and consider a Wesserstein distance that is invariant under conformal transformations.
2. Boyer labels 18 landmarks on each teeth. Then they find the best rigid transformation to match the labeled landmarks.

Assessment: They consider 116 teeth. For each teeth find the closest teeth according to each distance, and see whether they are in the same category.

Real data application

Our experiments

- Consider $X_i = \{p^i_1, \dots p^i_{18}\}$ $i = 1 \dots 116$.
- Find $\tilde{d}(X_i, X_j)$
- Use their classification scheme

Future work

- Understand topological properties of this SDP-induced distance on the set of finite metric spaces.
  - Convergence
  - Compactness
- Applications to datasets that are not surfaces.
- Understand how the distance behaves with respect to sampling under geometric assumptions.
- Compare with other lower bounds available in the literature.

Future work

- Understand topological properties of this SDP-induced distance on the set of finite metric spaces.
  - Convergence
  - Compactness
- Applications to datasets that are not surfaces.
- Understand how the distance behaves with respect to sampling under geometric assumptions.
- Compare with other lower bounds available in the literature.

Future work

- Understand topological properties of this SDP-induced distance on the set of finite metric spaces.
  - Convergence
  - Compactness
- Applications to datasets that are not surfaces.
- Understand how the distance behaves with respect to sampling under geometric assumptions.
- Compare with other lower bounds available in the literature.

Future work

- Understand topological properties of this SDP-induced distance on the set of finite metric spaces.
  - Convergence
  - Compactness
- Applications to datasets that are not surfaces.
- Understand how the distance behaves with respect to sampling under geometric assumptions.
- Compare with other lower bounds available in the literature.

Questions?

Tomorrow:
MS22: Convex Signal Recovery from Pairwise Measurements

10:30-10:55. Dustin Mixon. *Probably Certifiably Correct K-Means Clustering*.

11:00-11:25. Soledad Villar. *Efficient Global Solutions to K-Means Clustering Via Semidefinite Relaxation*