Shape Matching, A Metric Geometry Approach
CS468-Fall-2008
Lecture-3

Facundo Mémoli
Department of Mathematics, Stanford University
`memoli@math.stanford.edu`

October 13th, 2008

Contents

1 Properties of the GH distance

From now on, let $\mathfrak{M}$ denote the collection of all compact metric spaces. We have the following

Proposition 1. *For all $A, B \in \mathfrak{M}$,*

$$\mathbf{dist}_1(A,B) = \mathbf{dist}_2(A,B) = \mathbf{dist}_3(A,B) \ \left(= d_{\mathcal{GH}}(A,B)\right).$$

- The GH distance satisfies other desirable properties. In particular, it is a metric.

  Theorem 1 (See **[BBI]** for a full proof of item 1.)**.** *1. The GH distance is a metric on $\mathfrak{M}$.*

  2. For all $X \in \mathfrak{M}$ and $\mathbb{X} \subset X$,

  $$d_{\mathcal{GH}}((X,d_X),(\mathbb{X},d_X|_{\mathbb{X}\times\mathbb{X}})) \leqslant d_{\mathcal{H}}^X(X,\mathbb{X})$$

  3. Bonus from triangle inequality: For all $X, Y \in \mathfrak{M}$

  – $\frac{1}{2}|\mathbf{diam}\,(X) - \mathbf{diam}\,(Y)\,| \leqslant d_{\mathcal{GH}}(X,Y)$.[1]

  – *Let $\mathbb{X} \subset X$ and $\mathbb{Y} \subset Y$ be ε-coverings of X and Y, respectively. Then,*

  $$|d_{\mathcal{GH}}(X,Y) - d_{\mathcal{GH}}(\mathbb{X},\mathbb{Y})| \leqslant 2\varepsilon.$$

[1] This is the most basic lower bound for the GH distance.

- Stability? Consistency?
- An expression for the GH distance involving maps. This expression is at the foundations of the computational ideas in [**MS05**] and then also [**BBK06**].
- Remember from Lecture-1 that given for some $\phi : X \to Y$ and $\psi : Y \to X$, a correspondence $R \in \mathcal{R}(X,Y)$ can be constructed as

$$R = \{(x, \phi(x)),\, x \in X\} \bigcup \{(\psi(y), y)\, y \in Y\}.$$

- Let $\overline{\mathcal{R}}(X,Y)$ be the set of all correspondences that can be formed via such construction. Then (Exercise!), it is enough to consider R in this class in $\mathbf{dist}_1$.
- Expanding expression of the GH distance given by $\mathbf{dist}_1$, for this map-based correspondences we find

$$(*) \qquad d_{\mathcal{GH}}(X,Y) = \frac{1}{2} \inf_{\phi,\psi} \max(\mathrm{dis}(\phi), \mathrm{dis}(\psi), C(\phi,\psi))$$

where $\mathrm{dis}(\phi) := \|d_X - d_Y \circ (\phi,\phi)\|_{L^\infty(X\times X)}$, $\mathrm{dis}(\psi) := \|d_X \circ (\psi,\psi) - d_Y\|_{L^\infty(Y\times Y}$ and

$$C(\phi,\psi) = \max_{x\in X, y\in Y} |d_X(x,\psi(y)) - d_X(y,\phi(x))|.$$

- What does the term $C(\phi,\psi)$ mean?
- Notice that $d_{\mathcal{GH}}(X,Y) < \eta$ means that there exist $\phi : X \to Y$ and $\psi : Y \to X$ s.t. in particular, $C(\phi,\psi) < \eta$. This means that for all $x \in X, y \in Y$,

$$|d_X(x,\psi(y)) - d_Y(y,\phi(x))| < \eta.$$

Fix $x \in X$. Then the above holds for $y = \phi(x)$. This means that

$$d_X(x, \psi\circ\phi(x)) < \eta \quad \text{for all } x \in X.$$

- Similarly, we obtain

$$d_Y(y, \phi\circ\psi(y)) < \eta \quad \text{for all } y \in Y.$$

- This can be interpreted as expressing the fact that ϕ *and* ψ *are* η*-almost inverses of eachother.*
- What about connections with the concept of ε-isometry? Remember: A map $f : X \to Y$ is called an ε-isometry between X and Y if (1) $\mathrm{dis}(f) < \varepsilon$ and (2) $f(X)$ is a ε-net for Y.

Proposition 2 ([**BBI**])**.** *Let $X, Y \in \mathfrak{M}$.*

1. *Assume $d_{\mathcal{GH}}(X,Y) < \varepsilon$. Then there exist a 2ε-isometry between X and Y.*
2. *Assume that there exist an ε-isometry between X and Y. Then, $d_{\mathcal{GH}}(X,Y) < 2\varepsilon$.*

- Let $N_{X,n}^{R,s}$ denote a set of n-points $\{x_1,\ldots,x_n\}$ in X, which are an R-net for X and are s-separated: $\min_{\alpha\neq\beta} d_X(x_\alpha, x_\beta) = s$.
- For finite metric spaces $\mathbb{X}$ and $\mathbb{Y}$ with the same cardinality n, let

$$d_{\mathcal{I}}(\mathbb{X},\mathbb{Y}) := \frac{1}{2} \min_{\pi\in\Pi_n} \|\Gamma_{\mathbb{X},\mathbb{Y}}\|_{L^\infty(R_\pi\times R_\pi)}$$

where for a permutation π (lecture-1), $R_\pi \in \mathcal{R}(\mathbb{X},\mathbb{Y})$ is given by $R_\pi = \{(x_i, y_{\pi_i}),\, i = 1,\ldots,n\}$.

- Obviously, $d_{\mathcal{GH}}(\mathbb{X},\mathbb{Y}) \leqslant d_{\mathcal{I}}(\mathbb{X},\mathbb{Y})$.
- We can obtain the following (proof is *Exercise*):

  Corollary 1 (check [**MS05**]). *Let (X,d_X) and (Y,d_Y) be in $\mathfrak{M}$ s.t. $d_{\mathcal{GH}}(X,Y)=\eta$. Let $\{x_i\} = N_{X,n}^{(R,s)}$ be given. Then, for all $\gamma > 0$, there exist points $\mathbb{Y}^\gamma := \{y_1^\gamma,\ldots,y_n^\gamma\} \subset Y$ s.t.*

  1. $d_{\mathcal{I}}\left(N_{X,n}^{(R,s)},\mathbb{Y}^\gamma\right) \leqslant \eta+\gamma$.
  2. *$\mathbb{Y}^\gamma$ is a $R+2(\eta+\gamma)$ cover of Y.*
  3. *sep*$(\mathbb{Y}^\gamma) \geqslant s-2(\eta+\gamma)$.

2 A relationship with the Elad-Kimmel approach

- The [**EK**] approach can be (loosely) interpreted as an attempt to compute something like the GH distance.
- See slides.

3 Computational ideas

- There are two approaches, to the best of my knowledge.
- Both of them assume X and Y are smooth Riemannian manifolds.
- The approach in the original papers [**MS04**], [**MS05**] and then the MDS-like approach of [**BBK06**].
- They both hinge on expression $(*)$.
- [**MS05**] provides probabilisitic bounds for a certain estimate of the GH distance. These bounds can be extended to more general metric spaces. The underlying algorithmic procedure is a search over points in each metric space.
- [**BBK06**] uses the same idea, but they use the smooth structure of X and Y to define a (potentially large) continuous optimization problem. The quantity being estimated is not the GH distance. No bounds relating to the GH distance are given nor desirable properties of this quantity are established.
- Neither method has theoretical guarantees for the estimate they ultimately compute in practice.
- This is a problem in all the methods I know. Either you have to solve a hard combinatorial problem, or you have a huge continuous (non-convex) optimization problem.
- But before discussing them let's think of possible approaches based on $\mathbf{dist}_1$.

Throughout this section X and Y are finite metric spaces.

3.1 Brute force (?)

- Fix $n \in \mathbb{N}$. Assume you have $N_{X,n}^{(R,s)}$ and $N_{Y,n}^{(R',s')}$.
- We can then estimate the GH distance from above:

$$d_{\mathcal{GH}}(X,Y) \leqslant R+R'+d_{\mathcal{I}}(N_X,N_Y).$$

- This could work in principle. But there may be obstructions!! Burago and Kleiner, and also McCullen (see references in [**MS05**]) constructed two nets in R^2 such that when you force a bijective mapping between them, then $d_{\mathcal{I}}$ is infinity!.

- What is OK, is fixing one of them, say N_X, and then searching for $\mathbb{Y}$ in Y that minimizes

$$d_{\mathcal{I}}(\mathbb{Y}, N_X).$$

  Indeed, Corollary 1 guarantees this. We see more about this later.

3.2 A connection with the QAP

Recall the $\mathbf{dist}_1$ expression for the GH distance:

$$d_{\mathcal{GH}}(X,Y) = \frac{1}{2}\min_R \max_{(x,y),(x',y')\in R} |d_X(x,x') - d_Y(y,y')| \tag{1}$$

Remark 1. *We want to argue that expression (1) is reminiscent of the QAP (Quadratic Assignment Problem). This will let us loosely infer something about the inherent complexity of computing the Gromov-Hausdorff distance. Let's restrict ourselves to the case of finite metric spaces,* $\mathbb{X} = \{x_1,\ldots,x_n\}$ *and* $\mathbb{Y} = \{y_1,\ldots,y_m\}$*. For* $R \in \mathcal{R}(\mathbb{X},\mathbb{Y})$ *let* δ_{ij}^R *equal* 1 *if* $(x_i,y_j)\in R$ *and* 0 *otherwise. Then we have:*

$$d_{\mathcal{GH}}(X,Y) = \frac{1}{2}\min_R \max_{i,k,j,l} \Gamma_{ijkl}\ \delta_{ij}^R\delta_{kl}^R$$

where $\Gamma_{ijkl} := |d_X(x_i,x_k) - d_Y(y_j,y_l)|$*. Now, one can obtain a family of related problems by relaxing the* max *to a sum as follows: Fix* $p \geqslant 1$*, then one can also consider the problem:*

$$(\mathcal{P}_p) \qquad \min_R \sum_{ij}\sum_{kl} (\Gamma_{ikjl})^p\ \delta_{ij}^R\delta_{kl}^R.$$

Note that one can recast the above problem as follows. Let $\Delta \subset \mathbb{R}^{n\times m}$ *denote the set of matrices defined by the constraints below:*

1. $\delta_{ij} \in \{0,1\}$ *for all* i,j
2. $\sum_i \delta_{ij} \geqslant 1$ *for all* j
3. $\sum_j \delta_{ij} \geqslant 1$ *for all* i

and let $K_p(\delta) := \sum_{ij}\sum_{kl}(\Gamma_{ijkl})^p\delta_{ij}\delta_{kl}$*. then* $(\mathcal{P}_p)$ *is equivalent to*

$$\min_{\delta\in\Delta} K_p(\delta)$$

*which can be regarded as a generalized version of the QAP. In the standard QAP ([**?**])* $n = m$ *and the inequalities 2. and 3. defining* Δ *above are actually equalities, what forces each* δ *to be a permutation matrix.*

Actually, we argue next that, when $n = m$*,* $(\mathcal{P}_p)$ *reduces to a QAP. It is known that the QAP is an NP-hard problem, see references in [**M07**].*

Indeed, it is clear that for any $\delta \in \Delta$ *there exist* $\pi \in \Pi_n$ *(*$n\times n$ *permutations matrices) such that* $\delta_{ij} \geqslant \pi_{ij}$ *for all* $1 \leqslant i,j \leqslant n$*. Then, since* $(\Gamma_{ijkl})^p$ *is non negative for all* $1 \leqslant i,j,k,l \leqslant n$*, it follows that* $K_p(\delta) \geqslant K_p(\pi)$*. Therefore the minimal value of* $K_p(\delta)$ *is attained at some* $\delta \in \Pi_n$*.*

3.3 A naive relaxation

Recall problem $(\mathcal{P})$ in Remark 1. We now relax condition 1 in the definition of Δ to read

$$\delta_{ij} \in [0,1].$$

We let Δ' be the new set of matrices we obtain. Consider the problem given by

$$(\mathcal{P}') \qquad \min_{\delta \in \Delta'} K_p(\delta).$$

A few comments are in order:

- Problem $(\mathcal{P}')$ is a QOP with linear and bound constraints. It is however not necessarily convex.
- One can directly attempt to solve this problem using off the shelf tools for continuous variable optimization, e.g. gradient descent. One may, however, only converge to local minima.
- It is obvious that $\min_{\delta \in \Delta} K_p(\delta) \geqslant \min_{\delta \in \Delta'} K_p(\delta)$. However, it is unclear whether a bound can be obtained in the reverse direction, that is, can one relate $d_{\mathcal{GH}}$ with $\min_{\delta \in \Delta'} K_p(\delta)$?
- Assume that $\mathbb{X}$ and $\mathbb{Y}$ are sample sets from certain underlying "nice" metric spaces X and Y, respectively. It is unclear what is the behaviour of $\min_{\delta \in \Delta'} K_p(\delta)$ when both $\mathbb{X} \to X$ and $\mathbb{Y} \to Y$ (sampling consistency). It is clear that by imposing that δ be a finite *measure* on $\mathbb{X} \times \mathbb{Y}$ one could hope for a notion of limit problem.[2] This is basically the feature of the approach we'll see later on teh course (Gromov-Wasserstein or L^p GH distances)

3.4 The approach of [MS05]

- This approach did not try to compute the GH directly.
- The main observation was that $(*)$ can be simplified if one can drop the cross-term $C(\phi, \psi)$. Indeed, if one does that, then, what one obtaines in a lower bound for the GH distance, and more imprtantly, the two problems (optimize over ϕ and ψ) *become decoupled*:

$$d_{\mathcal{F}}(X,Y) := \max(\inf_{\phi} \operatorname{dis}(\phi), \inf_{\psi} \operatorname{dis}(\psi)).$$

- The challenge was, roughly, to build an upper bound for $d_{\mathcal{GH}}(X,Y)$ using $d_{\mathcal{F}}$ (and some more stuff).

4 Assignment for next class

1. Finish [**MS05**], get a good idea fo the computation of GH distance.
2. Read [**BBK**], get a good idea of their proposed computation of the GH distance.
3. Next class: Implementation details of [**MS05**] and [**BBK06**]. Critiques for the GH distance. Shape Contexts, Shape Distributions and other approaches. The Gromov-Wasserstein (or L^p GH distances).

[2] Namely, that "sums converge to integrals."