The DTM signature for a test of isomorphism between mm-spaces

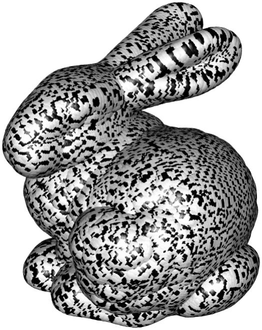
Claire Brécheteau

Université Paris-Sud 11, Laboratoire de Mathématiques d'Orsay and Inria Select Team
and
Inria Saclay, Data Shape Team

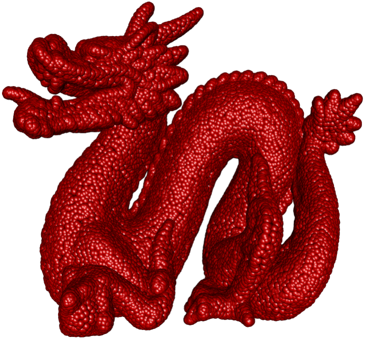
Under the supervision of
Pascal Massart (Université Paris-Sud 11) and Frédéric Chazal (Inria Saclay)

11 octobre 2016

Can a dragon pretend to be a bunny and vice versa ?



Bunny

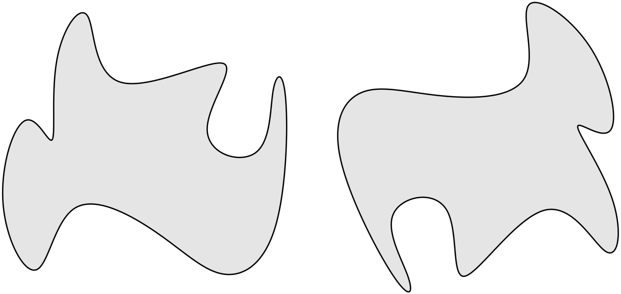


Dragon

Data from the Stanford 3D Scanning Repository

A **metric measure space** (mm-space) is a triple $(\mathscr{X}, \mathrm{d}, \mu)$ s.t. : $(\mathscr{X}, \mathrm{d})$ is a metric space and μ a borel measure supported on $\mathscr{X}$.

A **metric measure space** (mm-space) is a triple $(\mathscr{X},\mathrm{d},\mu)$ s.t. : $(\mathscr{X},\mathrm{d})$ is a metric space and μ a borel measure supported on $\mathscr{X}$.

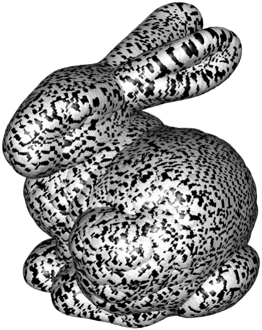


Two mm-spaces $(\mathscr{X},\mathrm{d},\mu)$ and $(\mathscr{Y},\mathrm{d}',\nu)$ are **isomorphic** if :
$\exists \phi : \mathscr{X} \mapsto \mathscr{Y}$ a one-to-one isometry, s.t. for all borel set A,

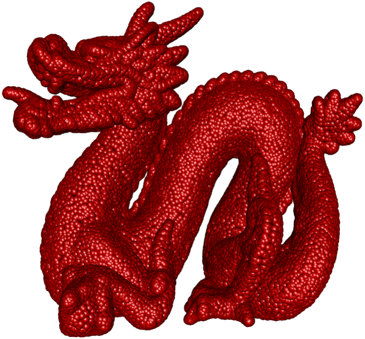
$$\nu(\phi(A)) = \mu(A).$$

How to build a test of level $\alpha > 0$ to test the null

$$H_0 : \text{“}(\mathscr{X}, \mathrm{d}, \mu) \text{ and } (\mathscr{Y}, \mathrm{d}', \nu) \text{ are isomorphic”}\,?$$



$(\mathscr{X}, \mathrm{d}, \mu)$



$(\mathscr{Y}, \mathrm{d}', \nu)$

A first idea : the Gromov-Wasserstein distance

The **Gromov-Wasserstein distance** is defined for two mm-spaces $(\mathscr{X}, d, \mu)$ and $(\mathscr{Y}, d', \nu)$ by :

$$GW(\mathscr{X}, \mathscr{Y}) = \inf_{\pi \in \Pi(\mu,\nu)} \frac{1}{2} \mathbb{E}_{\pi \otimes \pi} \left[\left| d(X_1, X_2) - d'(Y_1, Y_2) \right| \right]$$

with $\Pi(\mu, \nu)$ the set of distributions on $\mathscr{X} \times \mathscr{Y}$ of (X, Y) with $X \sim \mu$ and $Y \sim \nu$.

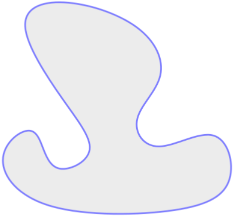
The **Gromov-Wasserstein distance** is defined for two mm-spaces $(\mathscr{X}, d, \mu)$ and $(\mathscr{Y}, d', \nu)$ by :

$$GW(\mathscr{X}, \mathscr{Y}) = \inf_{\pi \in \Pi(\mu,\nu)} \frac{1}{2} \mathbb{E}_{\pi \otimes \pi} \left[\left| d(X_1, X_2) - d'(Y_1, Y_2) \right| \right]$$

with $\Pi(\mu, \nu)$ the set of distributions on $\mathscr{X} \times \mathscr{Y}$ of (X, Y) with $X \sim \mu$ and $Y \sim \nu$.

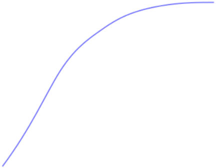
⚠ Too high computational cost.

A second idea : the signatures

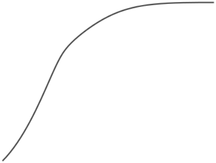


A hat

A cap



Signature of the hat



Signature of the cap

The **shape distribution** :
$\mathscr{L}(d(X_1,X_2))$ for X_1, X_2 i.i.d. from μ.

The **shape distribution** :
$\mathscr{L}(d(X_1,X_2))$ for X_1, X_2 i.i.d. from μ.

The **eccentricity** :
$\mathscr{L}(\mathbb{E}_\mu[d(.,X)](X'))$ for X' from μ.

The **shape distribution** :
$\mathscr{L}(d(X_1, X_2))$ for X_1, X_2 i.i.d. from μ.

The **eccentricity** :
$\mathscr{L}(\mathbb{E}_\mu[d(.,X)](X'))$ for X' from μ.

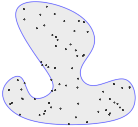
The **local distribution of distances** :
$\mathscr{L}(\mu(\overline{\mathrm{B}}(X,r)))$ for X from μ.

The **shape distribution** :
$\mathscr{L}(d(X_1,X_2))$ for X_1, X_2 i.i.d. from μ.

The **eccentricity** :
$\mathscr{L}(\mathbb{E}_\mu[d(.,X)](X'))$ for X' from μ.

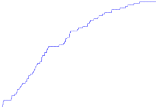
The **local distribution of distances** :
$\mathscr{L}(\mu(\overline{\mathrm{B}}(X,r)))$ for X from μ.

Persistence diagrams, Reeb graphs etc.



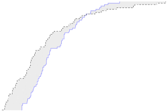
Sample on the hat

Sample on the cap



Empirical signature for the hat

Empirical signature for the cap

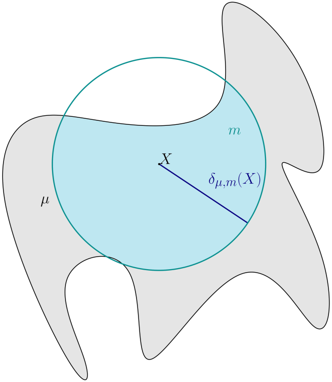


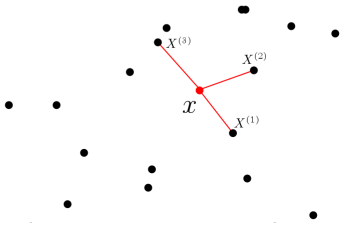
Test statistic

The **distance to a measure** [Chazal, Cohen-Steiner, Mérigot 2009] is defined for all $x \in \mathscr{X}$ and $m \in [0,1]$ by :

$$\mathrm{d}_{\mu,m}(x) = \frac{1}{m}\int_0^m \delta_{\mu,l}(x)\mathrm{d}l$$

with $\delta_{\mu,l}(x) = \inf\{r > 0 \mid \mu(\overline{\mathrm{B}}(x,r)) > m\}$.





Let $\hat{\mu}_N = \sum_{i=1}^{N} \frac{1}{N}\delta_{X_i}$ be some empirical measure in some metric space $(\mathcal{X}, d)$ and $k = Nm$.

Proposition

If $X^{(1)}$, $X^{(2)}$,... $X^{(k)}$ are the k nearest neighbours of x in $\{X_1, X_2, \dots X_N\}$, then :

$$\mathrm{d}_{\hat{\mu}_N, \frac{k}{N}}(x) = \frac{1}{k}\sum_{i=1}^{k} d(X^{(i)}, x).$$

Proposition

The function $\mathrm{d}_{\mu,m}$ *is 1-Lipschitz.*

Proposition

The function $\mathrm{d}_{\mu,m}$ is 1-Lipschitz.

The L_1-**Wasserstein** distance between two measures μ and ν over the same metric space $(\mathcal{X}, d)$ is defined by :

$$W_1(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \mathbb{E}_\pi[d(X,Y)].$$

Proposition

The function $\mathrm{d}_{\mu,m}$ *is 1-Lipschitz.*

The L_1-**Wasserstein** distance between two measures μ and ν over the same metric space $(\mathcal{X}, d)$ is defined by :

$$W_1(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \mathbb{E}_\pi[d(X,Y)].$$

Proposition (Chazal, Cohen-Steiner, Mérigot)

$$\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\nu,m}\|_{\infty,\mathcal{X}} \leq \frac{1}{m}\mathrm{W}_1(\mu,\nu).$$

The DTM signature

Definition

If $X \sim \mu$, the **DTM signature** $\mathrm{d}_{\mu,m}(\mu)$, is the distribution of $\mathrm{d}_{\mu,m}(X)$.

Definition

If $X \sim \mu$, the **DTM signature** $\mathrm{d}_{\mu,m}(\mu)$, is the distribution of $\mathrm{d}_{\mu,m}(X)$.

The L_1-**Wasserstein** distance between two real-valued measures of cdf F_μ and F_ν is equal to :

$$W_1(\mu,\nu) = \int_{\mathbb{R}} \left| F_\mu(t) - F_\nu(t) \right| \, \mathrm{d}t$$

Definition

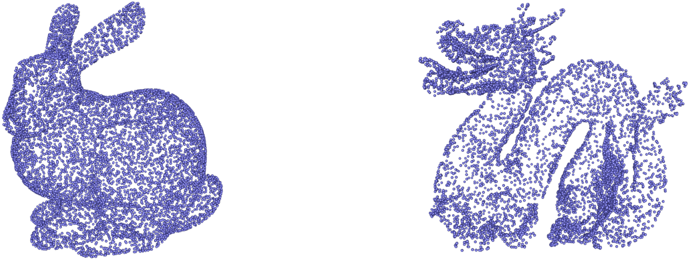
If $X \sim \mu$, the **DTM signature** $\mathrm{d}_{\mu,m}(\mu)$, is the distribution of $\mathrm{d}_{\mu,m}(X)$.

The L_1-**Wasserstein** distance between two real-valued measures of cdf F_μ and F_ν is equal to :

$$W_1(\mu,\nu) = \int_{\mathbb{R}} \left|F_\mu(t) - F_\nu(t)\right| \, \mathrm{d}t$$

Proposition

$$W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)) \leq \frac{1}{m} GW(\mathcal{X}, \mathcal{Y})$$

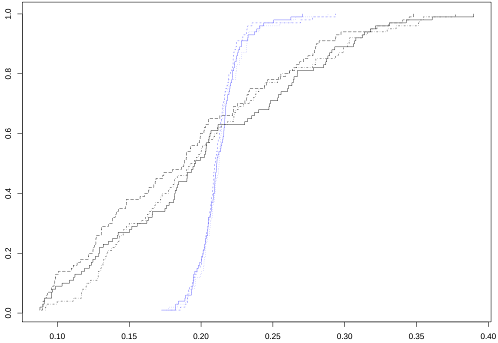


N-samples on the Bunny and the Dragon

Definition

The **empirical DTM signature** is defined by $\mathrm{d}_{\hat{\mu}_{N-n},m}(\hat{\mu}_n)$ from two independent $(N-n)$ and n-samples of law μ.

The test statistic



Empirical DTM signatures for the Bunny and the Dragon
$N = 10000, \; n = 100, \; m = 0.1$

Our **test statistic** : $T = \sqrt{n} W_1(\mathrm{d}_{\hat{\mu}_{N-n},m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_{N-n},m}(\hat{\nu}_n))$.

Bootstrap approximation

Under the null, we approximate
$\mathscr{L}_{N,n,m}(\mu,\nu) = \mathscr{L}(\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_{N-n},m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_{N-n},m}(\hat{\nu}_n)))$

Under the null, we approximate
$\mathscr{L}_{N,n,m}(\mu,\nu) = \mathscr{L}(\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_{N-n},m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_{N-n},m}(\hat{\nu}_n)))$

by the bootstrap law

$$\frac{1}{2}\mathscr{L}^*_{N,n,m}(\hat{\mu}_N) \oplus \frac{1}{2}\mathscr{L}^*_{N,n,m}(\hat{\nu}_N),$$

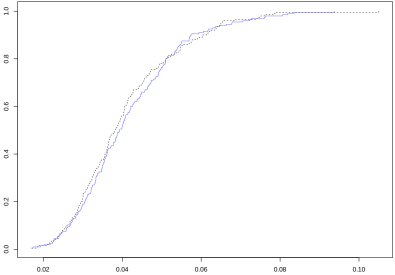
with $\mathscr{L}^*_{N,n,m}(\hat{\mu}_N) = \mathscr{L}^*(\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_N,m}(\mu_n^*), \mathrm{d}_{\hat{\mu}_N,m}(\mu_n^{*\prime}))|\hat{\mu}_N)$.

Under the null, we approximate
$\mathscr{L}_{N,n,m}(\mu,\nu) = \mathscr{L}(\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_{N-n},m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_{N-n},m}(\hat{\nu}_n)))$

by the bootstrap law

$$\frac{1}{2}\mathscr{L}^*_{N,n,m}(\hat{\mu}_N) \oplus \frac{1}{2}\mathscr{L}^*_{N,n,m}(\hat{\nu}_N),$$

with $\mathscr{L}^*_{N,n,m}(\hat{\mu}_N) = \mathscr{L}^*(\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_N,m}(\mu^*_n), \mathrm{d}_{\hat{\mu}_N,m}(\mu^{*\prime}_n))|\hat{\mu}_N)$.



Empirical cdf of $\mathscr{L}_{N,n,m}(\mu,\mu)$ and $\mathscr{L}^*_{N,n,m}(\hat{\mu}_N)$
$N = 10000,\ n = 100,\ m = 0.1,\ N_{MC} = 200$

Assume $N = n + cn^{\rho}$ and $m = c'n^{-\sigma}$ with c, c', ρ, σ positive constants.

Lemma

The quantity $W_1\left(\mathscr{L}_{N,n,m}(\mu,\mu), \mathscr{L}^*_{N,n,m}(\hat{\mu}_N)\right)$ *is upper bounded by :*

$$2\sqrt{n}\left(\mathbb{E}[\|\mathrm{d}_{\hat{\mu}_{N-n},m} - \mathrm{d}_{\mu,m}\|_{\infty,\mathscr{X}}] + W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\mu,m}(\hat{\mu}_N)) + \|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathscr{X}}\right),$$

Assume $N = n + cn^{\rho}$ and $m = c'n^{-\sigma}$ with c, c', ρ, σ positive constants.

Lemma

The quantity $W_1\left(\mathscr{L}_{N,n,m}(\mu,\mu),\mathscr{L}^*_{N,n,m}(\hat{\mu}_N)\right)$ *is upper bounded by :*

$$2\sqrt{n}\left(\mathbb{E}[\|\mathrm{d}_{\hat{\mu}_{N-n},m}-\mathrm{d}_{\mu,m}\|_{\infty,\mathscr{X}}]+W_1(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\mu,m}(\hat{\mu}_N))+\|\mathrm{d}_{\mu,m}-\mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathscr{X}}\right),$$

which is upper bounded by

$$2\frac{\sqrt{n}}{m}\mathbb{E}\left[W_1\left(\hat{\mu}_{N-n},\mu\right)\right]+2\sqrt{n}\left(1+\frac{1}{m}\right)W_1\left(\hat{\mu}_N,\mu\right).$$

Let μ and ν be two borel probability measure supported on a compact subset of euclidean $\mathbb{R}^d$.

Proposition

For some positive constant C, we have,

$$\mathbb{E}[W_1(\mathscr{L}_{N,n,m}(\mu,\mu),\mathscr{L}^*_{N,n,m}(\hat{\mu}_N))] \leq CN^{\frac{1}{2\rho}+\frac{\sigma}{\rho}-\frac{1}{\max\{d,2\}}}[\log(1+N)]^{\mathbb{1}_{d=2}}.$$

Proof : Lemma + Rates of convergence in $\mathbb{R}^d$ for $\mathbb{E}[W_1(\hat{\mu}_N,\mu)]$ from [Fournier, Guillin 2015].

Let μ and ν be two borel probability measure supported on a compact subset of euclidean $\mathbb{R}^d$.

Proposition

For some positive constant C, we have,

$$\mathbb{E}[W_1(\mathscr{L}_{N,n,m}(\mu,\mu),\mathscr{L}^*_{N,n,m}(\hat{\mu}_N))] \leq CN^{\frac{1}{2\rho}+\frac{\sigma}{\rho}-\frac{1}{\max\{d,2\}}}[\log(1+N)]^{\mathbb{1}_{d=2}}.$$

Proof : Lemma + Rates of convergence in $\mathbb{R}^d$ for $\mathbb{E}[W_1(\hat{\mu}_N,\mu)]$ from [Fournier, Guillin 2015].

Proposition

If $\sigma < \frac{\rho}{\max\{d,2\}} - \frac{1}{2}$, then : $W_1(\mathscr{L}_{N,n,m}(\mu,\mu),\mathscr{L}^*_{N,n,m}(\hat{\mu}_N)) \to 0$ a.e. when $N \to \infty$.

Proof : Lemma + Bounds in probability for $W_1(\hat{\mu}_N,\mu)$ from [Fournier, Guillin 2015] + Borel-Cantelli lemma.

Validity for the bootstrap – Ahlfors regular measures on connected compact subsets of $\mathbb{R}^d$

Let $(\mathscr{X}, \delta, \mu)$ be a mm-space with $\mathscr{X}$ a nonempty bounded subset of $\mathbb{R}^d$, $\delta = \|\cdot - \cdot\|_2$ and μ some **Ahlfors b-regular** measure with parameters (a, ∞), that is a Borel probability measure satisfying :

$$\forall x \in \mathscr{X}, \forall r \in \mathbb{R}_+, 1 \wedge ar^b \leq \mu(\mathrm{B}(x,r)) \leq \infty r^b.$$

Validity for the bootstrap – Ahlfors regular measures on connected compact subsets of $\mathbb{R}^d$

Let $(\mathscr{X},\delta,\mu)$ be a mm-space with $\mathscr{X}$ a nonempty bounded subset of $\mathbb{R}^d$, $\delta = \|\cdot - \cdot\|_2$ and μ some **Ahlfors b-regular** measure with parameters (a,∞), that is a Borel probability measure satisfying :

$$\forall x \in \mathscr{X}, \forall r \in \mathbb{R}_+, 1 \wedge ar^b \leq \mu(\mathrm{B}(x,r)) \leq \infty r^b.$$

Let $N = n + cn^\rho$ and $m = c'$ for c, c' and ρ positive constants.

Proposition

Up to a logarithmic term, $\mathbb{E}[W_1(\mathscr{L}_{N,n,m}(\mu,\mu),\mathscr{L}^*_{N,n,m}(\hat{\mu}_N))]$ *is bounded by* $N^{\frac{1}{2\rho}-\frac{1}{2}}$.

Proof : Lemma + Rates of convergence in $\mathbb{R}$ for $\mathbb{E}[W_1(\hat{\mu}_N,\mu)]$ from [Bobkov, Ledoux 2014] + Rates of convergence for $\mathbb{E}[\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_\infty]$ from [Chazal, Massart, Michel 2015].

Let μ be a borel probability measure supported on a compact set of a metric space ; n and N go to ∞ together.

Lemma

If $\sqrt{n}\mathbb{E}[\|\mathrm{d}_{\mu,m}-\mathrm{d}_{\hat{\mu}_{N-n},m}\|_{\infty,\mathscr{X}}]\rightarrow 0$ *or more specifically* $\frac{\sqrt{n}}{m}\mathbb{E}[W_1(\mu,\hat{\mu}_{N-n})]\rightarrow 0$. *Then,*

$$\mathscr{L}_{N,n,m}(\mu,\mu)\rightsquigarrow\mathscr{L}\left(\|\mathbb{G}_{\mu,m}-\mathbb{G}'_{\mu,m}\|_1\right).$$

Moreover, if $\sqrt{n}W_1(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\mu,m}(\hat{\mu}_N))\rightarrow 0$ *a.e. and* $\sqrt{n}\|\mathrm{d}_{\mu,m}-\mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathscr{X}}\rightarrow 0$ *a.e., then for almost every sample* $X_1,X_2,\ldots X_N\ldots$ *:*

$$\mathscr{L}^*_{N,n,m}(\hat{\mu}_N)\rightsquigarrow\mathscr{L}\left(\|\mathbb{G}_{\mu,m}-\mathbb{G}'_{\mu,m}\|_1\right),$$

with $\mathbb{G}_{\mu,m}$ *and* $\mathbb{G}'_{\mu,m}$ *two independent gaussian processes with covariance kernel* $\kappa(s,t)=F_{\mathrm{d}_{\mu,m}(\mu)}(s)\left(1-F_{\mathrm{d}_{\mu,m}(\mu)}(t)\right)$ *for* $s\leq t$.

Proof : $W_1(\mathscr{L}_{N,n,m}(\mu,\mu),\mathscr{L}(W_1(\mathrm{d}_{\mu,m}(\hat{\mu}_n),\mathrm{d}_{\mu,m}(\hat{\mu}'_n))))\rightarrow 0$ and $W_1(\mathscr{L}^*_{N,n,m}(\hat{\mu}_N),\mathscr{L}(W_1(\mathrm{d}_{\mu,m}(\hat{\mu}_n),\mathrm{d}_{\mu,m}(\hat{\mu}'_n))))\rightarrow 0$ + Donsker + [de Acosta, Giné 79]

Here, $N = n + cn^{\rho}$.

Proposition

The two convergence in law occur,

- *in the general case in* $\mathbb{R}^d$, *if* $\rho > \frac{\max\{d,2\}}{2}$,
- *in the Ahlfors regular case, if* $\rho > 1$.

Proof : Lemma + Rates of convergence, in [Chazal, Massart, Michel 2015], [Bobkov, Ledoux 2014] and [Fournier, Guillin 2015].

Here, $N = n + cn^{\rho}$.

Proposition

The two convergence in law occur,

- *in the general case in $\mathbb{R}^d$, if $\rho > \frac{\max\{d,2\}}{2}$,*
- *in the Ahlfors regular case, if $\rho > 1$.*

Proof : Lemma + Rates of convergence, in [Chazal, Massart, Michel 2015], [Bobkov, Ledoux 2014] and [Fournier, Guillin 2015].

$\rightsquigarrow$ If $\mathscr{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1)$ is atomless : Asymptotic level α for the test.

The power of the test

We choose $N = n + n^\rho$ with $\rho > 1$.
Let μ and ν be two Borel measures supported on $\mathscr{X}$ and $\mathscr{Y}$, two compact subsets of $\mathbb{R}^d$.
Assume that $W_1\left(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)\right) > 0$.

Proposition

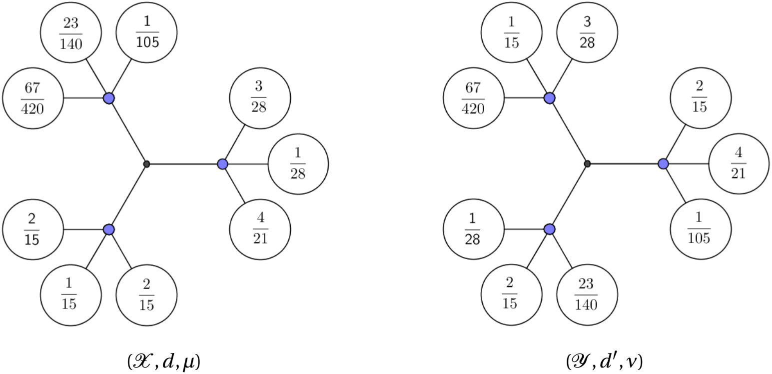
There is $n_0(\mu,\nu)$ such that $\forall n \geq n_0$, the error of second type

$$\mathbb{P}_{(\mu,\nu)}\left(\sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_{N-n},m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_{N-n},m}(\hat{\nu}_n)\right) < \hat{\mathrm{q}}_\alpha\right)$$

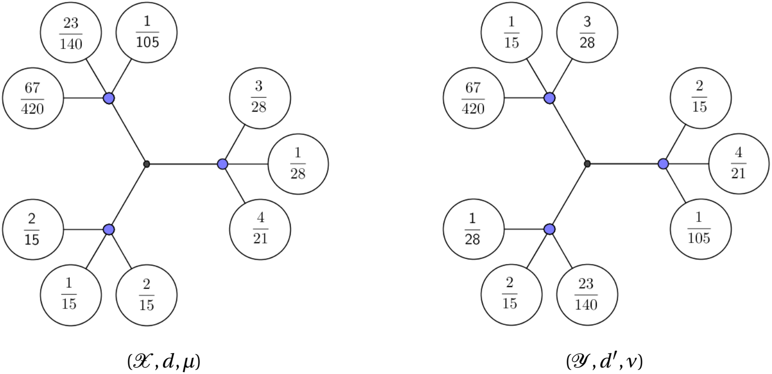
is upper bounded by

$$4\exp\left(-\frac{W_1^2\left(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)\right)}{3\max\left\{\mathrm{Diam}_\mu^2, \mathrm{Diam}_\nu^2\right\}}n\right).$$

Proof : Upper bound for $W_1(\mathscr{L}^*_{N,n,m}(\hat{\mu}_N), \mathscr{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1))$ and then of $\hat{\mathrm{q}}_\alpha$ by a fixed constant with high probability + inequalities with W_1, DTM... + DKW-Massart inequality



Example from Mémoli 2011



Example from Mémoli 2011

⚠ $\mathrm{d}_{\mu,m}(\mu) = \mathrm{d}_{\nu,m}(\nu)$ but $(\mathscr{X}, d, \mu)$ and $(\mathscr{Y}, d', \nu)$ are not isomorphic.

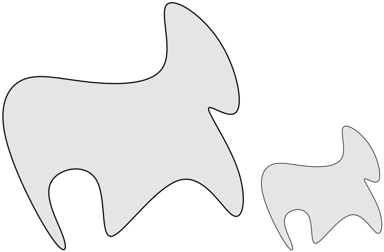
Dilatation of parameter λ, for $0 < \lambda \neq 1$.

Proposition

Let $(\mathscr{X}, \delta, \mu)$ and $(\mathscr{Y}, \gamma, \nu) = (\mathscr{X}, \lambda\delta, \mu)$ be two mm-spaces. We have

$$W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)) = |1 - \lambda| \, \mathbb{E}_\mu[\mathrm{d}_{\mu,m}(X)],$$

for X a random variable of law μ.



Discriminative power of the DTM signature

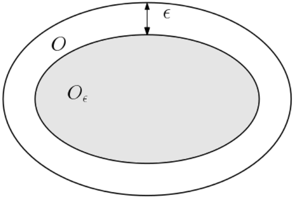
Uniform measures over two nonempty bounded open subsets O and O' of $\mathbb{R}^d$ with different Lebesgue volume.

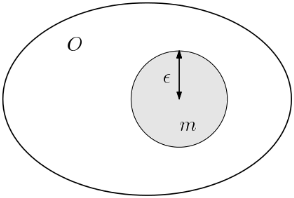
Proposition

If $O = \left(\overline{O}\right)^{\circ}$ *and* $O' = \left(\overline{O'}\right)^{\circ}$. *A lower bound for* $W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu))$ *is given by :*

$$\min\left(\mu\left(O_{\epsilon(m,O)}\right), \mu\left({O'}_{\epsilon(m,O')}\right)\right) \frac{d}{d+1}\left(\frac{m}{\omega_d}\right)^{\frac{1}{d}} \left|\mathrm{Leb}_d(O)^{\frac{1}{d}} - \mathrm{Leb}_d(O')^{\frac{1}{d}}\right|.$$

Here, $O_\epsilon = \left\{x \in O \mid \inf_{y \in \partial O} \|x - y\|_2 \geq \epsilon\right\}$, *and* $\epsilon(m, O) = \left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}$.

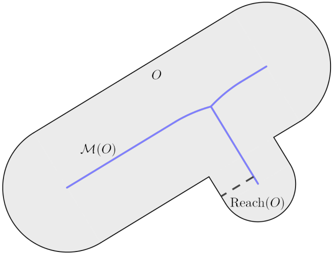




Let $(O, \|\cdot - \cdot\|_2, \mu_O)$ and $(O, \|\cdot - \cdot\|_2, \nu)$ be two mm-spaces with open $\emptyset \neq O \subset \mathbb{R}^d$, $\nu \ll \mu_O$, with $\mu_O = \frac{\mathrm{Leb}_d(. \cap O)}{\mathrm{Leb}_d(O)}$, f the Radon-Nikodym density s.t. :

$$\forall x, y \in O, \ |f(x) - f(y)| \leq L \|x - y\|_2^{\chi}.$$

We assume that $\mathrm{Reach}(O) > 0$.



Medial Axis and Reach of an open set

Proposition

Under the previous assumptions, if one of the following conditions is satisfied, then the quantity $W_1(\mathrm{d}_{\mu_O,m}(\mu_O),\mathrm{d}_{\nu,m}(\nu))$ *is positive :*

- $m < \frac{\omega_d}{\mathrm{Leb}_d(O)}\min\left\{\mathrm{Reach}(O)^d, \left(\frac{\|f\|_{\infty,O}-1}{2L}\right)^{\frac{d}{\chi}}\right\}$;
- $m \in \left[\frac{\omega_d}{\mathrm{Leb}_d(O)}(\mathrm{Reach}(O))^d, \left(\|f\|_{\infty,O} - 2L(\mathrm{Reach}(O))^{\chi}\right)(\mathrm{Reach}(O))^d\frac{\omega_d}{\mathrm{Leb}_d(O)}\right)$;
- $m \in \left[\frac{\omega_d}{\mathrm{Leb}_d(O)}\left(\frac{d}{\chi}\right)^{\frac{d}{\chi}}(2L)^{-\frac{d}{\chi}}, \min\left\{m_0, \frac{\omega_d}{\mathrm{Leb}_d(O)}(\mathrm{Reach}(O))^{d+\chi}\frac{\chi}{d}2L\right\}\right)$,

with $m_0 = \|f\|_{\infty,O}^{\frac{d}{\chi}+1}\frac{\omega_d}{\mathrm{Leb}_d(O)}\left(\frac{d}{\chi}\right)^{\frac{d}{\chi}}(2L)^{-\frac{d}{\chi}}\left(\frac{\chi}{d+\chi}\right)^{\frac{\chi}{d+\chi}}$.

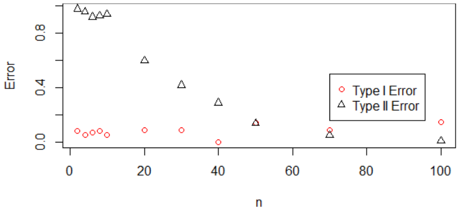
Moreover, a lower bound for $W_1(\mathrm{d}_{\mu_O,m}(\mu_O),\mathrm{d}_{\nu,m}(\nu))$ *is given by :*

$$\frac{1}{1+d}\frac{1}{\mathrm{Leb}_d(O)}\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}\int_{\lambda=1}^{\infty}\frac{1}{\lambda^{\frac{1}{d}}}\max_{\lambda'\geq\lambda}\mathrm{Leb}_d\left(\{f\geq\lambda'\}_{\left(\frac{m}{\omega_d}\frac{\mathrm{Leb}_d(O)}{\lambda'}\right)^{\frac{1}{d}}}\right)\mathrm{d}\lambda.$$

Approximate error of type I, and of type II

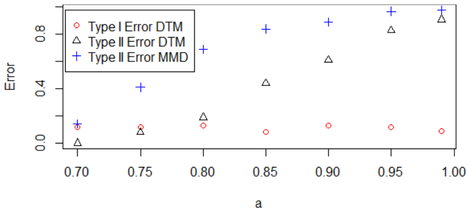
Bunny vs Dragon

$N = 500,\ m = 0.5,\ \alpha = 0.05,\ N_{MC} = 1000$; 100 times



Ball vs Ellipse(1,a)

$N = 200,\ n = 40,\ m = 0.5,\ \alpha = 0.05,\ N_{MC} = 1000$; 100 times



$N = 400$, $N_{MC} = 1000$, $\alpha = 0.05$, experiment replicated 10000 times

m \ n	2	5	10	20	50
0.02	0.0941	0.1029	0.1123	0.1329	0.1730
0.1	0.0557	0.0627	0.0686	0.0801	0.1332
0.2	0.0547	0.0587	0.0632	0.0748	0.1113
0.5	0.0514	0.0586	0.0560	0.0614	0.0896
0.8	0.0534	0.0527	0.0539	0.0671	0.0730

Approximate error of type I, $\hat{\alpha}_0$

$N = 400,\ N_{MC} = 1000,\ \alpha = 0.05,$ experiment replicated 10000 times

m \ n	2	5	10	20	50
0.02	0.0941	0.1029	0.1123	0.1329	0.1730
0.1	0.0557	0.0627	0.0686	0.0801	0.1332
0.2	0.0547	0.0587	0.0632	0.0748	0.1113
0.5	0.0514	0.0586	0.0560	0.0614	0.0896
0.8	0.0534	0.0527	0.0539	0.0671	0.0730

Approximate error of type I, $\hat{\alpha}_0$

m \ n	2	5	10	20	50
0.02	0.1453	0.1755	0.2105	0.2522	0.4058
0.1	0.0652	0.0752	0.0866	0.1103	0.2092
0.2	0.0603	0.0671	0.0687	0.0903	0.2021
0.5	0.0518	0.0586	0.0622	0.0781	0.1334
0.8	0.0516	0.0581	0.0666	0.0730	0.1153

Majoration of the error of type I by permutation, α_0^*

Thank you !