# DISTANCES

Pb : Stabilité/Continuité des constructions présentées par Clément et Noël :
Données initiales "proches" $\Longrightarrow$ Résultats finaux "proches" ?

X, Y esp. métriques (nuages de points) $\rightsquigarrow$ Filtrations de complexes simpliciaux (Čech ou VR) $\rightsquigarrow$ Diagrammes de Persistance

dgm(X), dgm(Y)

Théorème : $\underbrace{D_b}_{\text{distance Bottleneck}}(dgm(X), dgm(Y)) \leq 2 . \underbrace{D_{GH}}_{\text{distance de Gromov-Hausdorff}}(X, Y)$

## 1 _ Distances de Hausdorff & Gromov-Hausdorff (BBI, chap. 7)

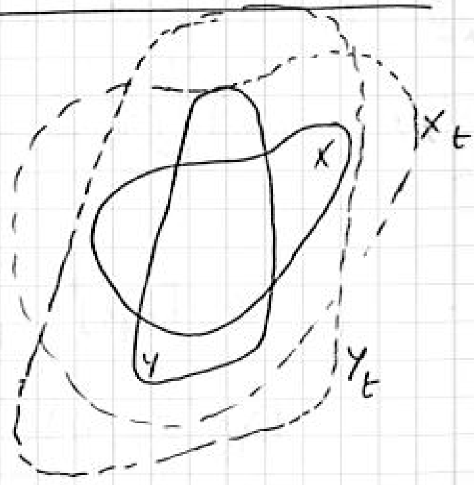
Définition : $(M,d)$ esp. métrique, $\mathcal{K}(M) = \{\text{compacts de } M\}$, $X, Y \in \mathcal{K}(M)$.

$$D_H(X,Y) = \operatorname{Inf}\{t > 0 \mid X \subset Y_t \text{ et } Y \subset X_t\}$$

$$= \operatorname{Max}\left( \max_{x \in X}(d(Y,x)), \max_{y \in Y}(d(X,y)) \right)$$

où $X_t = \{d(X,\cdot) < t\}$

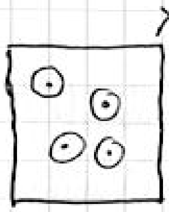
$= \bigcup_{x \in X} B(x,t)$ t-voisinage de X.




Thm : $D_H$ est une distance sur $\mathcal{K}(M)$

Exemples : $(M,d) = \mathbb{R}^2$, euclidien.

- $X = [0,1]^2$ , $x_1, \dots, x_N \in \mathring{X}$

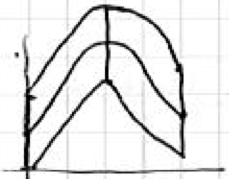



$$X_k = X \setminus \bigcup_{i=1}^{N} B(x_i, \tfrac{1}{k})$$

$D_H(X, X_k) = \frac{1}{k}$ donc $X_k \xrightarrow{D_H} X$

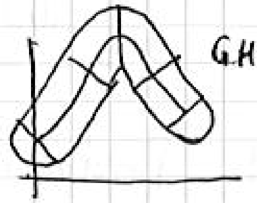
Pas de continuité de la topologie.

- « La convergence Hausdorff est une c.v. $C^0$ »

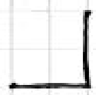



$f, f_k : [0,1] \longrightarrow \mathbb{R}$ continues

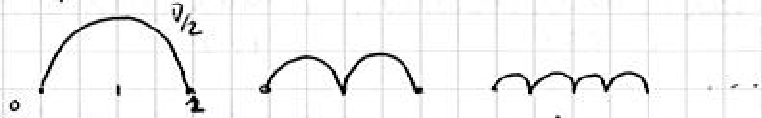
$G, G_k$ graphes de $f, f_k$




$G_k \xrightarrow{D_H} G$ si et ssi $f_k \xrightarrow{C^0} f$

(On a $D_H(G_f, G_g) \le \|f-g\|_\infty$, mais pas d'inégalité dans l'autre sens $f_k(t) = t^k$, $G_{f_k} \longrightarrow$  )

- Rq : Pas de continuité de la mesure de Hausdorff 1-dim.




$\|f_k\|_\infty \longrightarrow 0$ donc $G_k \xrightarrow{D_H} [0,1] \times \{0\}$

$\mathcal{H}^1(G_k) = \frac{\pi}{2}$ et $\mathcal{H}^1(G) = 1$.

- $X = [0,1]^2$ , $X_k = \frac{1}{k}\mathbb{Z}^2 \cap [0,1]^2 = \left\{ \left(\frac{i}{k}, \frac{j}{k}\right) \;\middle|\; i,j \in [\![0,k]\!] \right\}$

$D_H(X_k, X) = \frac{1}{k\sqrt{2}}$ , donc $X_k \xrightarrow{D_H} X$

et $\mathcal{H}^2(X_k) = 0$ , $\mathcal{H}^2(X) = 1$.

Pas de continuité du volume.

Propriétés :

- $\begin{cases} \mathcal{K}(M) \longrightarrow \mathbb{R} \\ X \longmapsto \text{diam} \end{cases}$ est 2-lipschitzienne.
- Si $(M,d)$ est compact (resp. complet) alors $(\mathcal{K}(M), D_H)$ est compact (resp. complet)
- Dans $\mathbb{R}^n$ euclidien, $X \longmapsto \mathcal{H}^n(X)$ est s.c.s.
  $\limsup(\mathcal{H}^n(X_k)) \leq \mathcal{H}^n(X)$ si $X_k \xrightarrow{D_H} X$

La bonne nouvelle : le dernier exemple montre qu'un réseau de points "assez serrés" est $D_H$-proche de $X$

Distance de Gromov-Hausdorff :

Dans $\mathbb{R}^2$, $\tau_u$ translation de vecteur $u$.

$X$ et $\tau_u(X)$ "sont les mêmes" mais $D_H(X, \tau_u(X)) = |u|$

Définition : $(X_1, d_1)$ et $(X_2, d_2)$ espaces métriques compacts

$$D_{GH}(X_1, X_2) = \operatorname{Inf}\left\{ t > 0 \;\middle|\; \begin{array}{l} \exists (M,d) \text{ esp. métrique}, \ \phi_i : X_i \longrightarrow M \text{ injections isométriques} \\ \text{t.q. } D_H^M(\phi_1(X_1), \phi_2(X_2)) < t \end{array} \right.$$

$X_1 = X$, $X_2 = \tau_u(X)$, $(Y, \delta) = (\mathbb{R}^2, \|.\|_2)$, $\phi_1 = +\mathrm{Id}$, $\phi_2 = \tau_{-u}$

$D_H(\phi_1(X_1), \phi_2(X_2)) = D_H(X, \tau_{-u}(\tau_u(X))) = D_H(X, X) = 0$.

Donc $D_{GH}(X_1, X_2) = 0$.

Théorème : $D_{GH}$ est une "distance" :

- Positive, symétrique, inégalité triangulaire.
- $D_{GH}(X_1, X_2) = 0$ ssi $X_1$ et $X_2$ isométriques.

Rq : Si $X_1, X_2 \subset (M,d)$, $D_{GH}(X_1, X_2) \leq D_H^M(X_1, X_2)$

les exemples précédents sont aussi des exemples de C.V. G-H.

## L'importance des nuages de points

- $S \subset (X,d)$ est un $\varepsilon$-réseau si $X = \bigcup_{x \in S} B(x,\varepsilon)$ ou $\forall x \in X \quad d(S,x) < \varepsilon$.

  En particulier, $D_{GH}(S,X) \leq \varepsilon$.

- $(X,d_X)$ et $(Y,d_Y)$ sont des $\varepsilon$-approx. l'un de l'autre s'il existe des $\varepsilon$-réseaux $S = \{x_1,\dots,x_N\} \subset X$ et $T = \{y_1,\dots,y_N\} \subset Y$ t.q. $\forall\, i,j \quad |d_X(x_i,x_j) - d_Y(y_i,y_j)| < \varepsilon$

**Proposition :**

1. Si $X$ et $Y$ sont $\varepsilon$-approx l'un de l'autre, alors $D_{GH}(X,Y) \leq 3.\varepsilon$.
2. Si $D_{GH}(X,Y) \leq \varepsilon$ alors $X$ et $Y$ sont des $5.\varepsilon$-approx l'un de l'autre.

| | | | |
|---|---|---|---|
| $(Y,d_Y)$ | $T_\varepsilon \subset Y$ | | $dgm(T_\varepsilon)$ |
| $(X,d_X)$ $\rightsquigarrow$ | $S_\varepsilon \subset X$ $\rightsquigarrow$ | Complexes de Cech ou V.R. $\rightsquigarrow$ | $dgm(S_\varepsilon)$ |
| esp. métrique compact | $\varepsilon$-réseau, $\#(S) < +\infty$ | | |

$$D_b(dgm(T_\varepsilon), dgm(S_\varepsilon)) \leq 2.D_{GH}(T_\varepsilon, S_\varepsilon) \leq 2.D_{GH}(X,Y) + 4\varepsilon$$

**Critère de compacité :** $(X_n,d_n)_{n\in\mathbb{N}}$ esp. métriques compacts.

Si
1. $\exists D \ \forall n \quad diam(X_n) \leq D$
2. $\forall \varepsilon > 0 \ \exists N \ \forall n \quad X_n$ admet un $\varepsilon$-réseau ~~approx~~ $S_n \subset X_n$ t.q. $\#(S_n) \leq N$.

Alors $(X_n, d_n)_{n\in\mathbb{N}}$ admet une s.ste qui c.v. G.H.

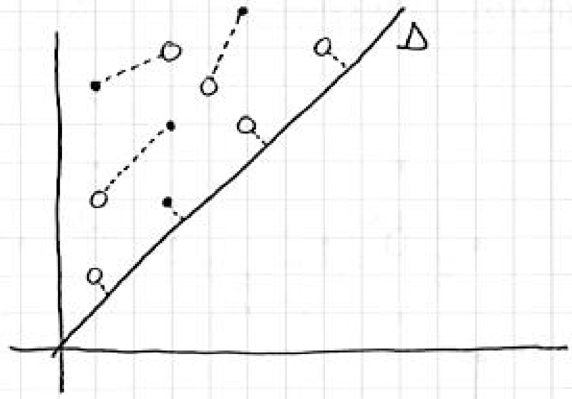
## 2 - Distances entre diagrammes de persistance.

Les diagrammes de persistance sont des familles de points de $\mathbb{R}^2$ qui encodent les changement de l'homologie de $\check{C}ech_t(X_\bullet)$ ou $Rips_t(X_\bullet)$ quand $t$ évolue. On les note comme des $n$-uplets pour tenir compte des multiplicités.

$$dgm(X) = (x_1, \dots, x_n) \subset \{(\underbrace{s}_{\text{apparition}}, t) \in \mathbb{R}^2 \mid 0 \leq s < t\} =: \mathcal{D}$$

disparition.

Il y a un diagramme par dimension d'Homologie, et chacun des $x_i$ correspond à une barre du code-barre




Pour définir une distance entre diagrammes, on "marie" leurs points.

$dgm_1 = (x_1, \dots, x_n)$

$dgm_2 = (y_1, \dots, y_p)$

Un "mariage partiel" de $n$ et $p$ éléments est $M \subset [\![1,n]\!] \times [\![1,p]\!]$ t.q.

- $\forall i \in [\![1,n]\!] \quad \#(M \cap \{i\} \times [\![1,p]\!]) \leq 1$
- $\forall j \in [\![1,p]\!] \quad \#(M \cap [\![1,n]\!] \times \{j\}) \leq 1$

"Pas de polygamie"

Si $M$ est un mariage partiel, on note

$C_1(M) = \{ i \in [\![1,n]\!] \mid M \cap \{i\} \times [\![1,p]\!] = \emptyset \}$

$C_2(M) = \{ j \in [\![1,p]\!] \mid M \cap [\![1,n]\!] \times \{j\} = \emptyset \}$

"les célibataires".

Le coût du mariage est

$$J_\infty(M) = \operatorname{Max}\Big[ \{ \|x_i - y_j\| \mid (i,j) \in M \} \cup \{ \|x_i - \pi_\Delta(x_i)\| \mid i \in C_1(M) \} \cup \{ \|y_j - \pi_\Delta(y_j)\| \mid j \in C_2(M) \} \Big]$$

où $\pi_\Delta$ est la proj. $\perp$ sur $\Delta$.

**Définition :** $D_b(dgm_1, dgm_2) = \text{Min}\{J_\infty(M) \mid M \text{ mariage partiel}\}$

De façon analogue on définit pour $r \geqslant 1$

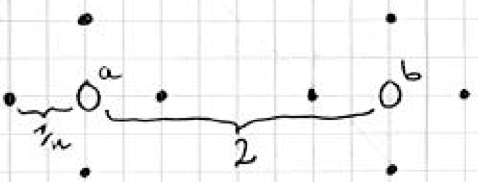
$$J_r(M) = \sum_{(i,j)\in M} \|x_i - y_j\|^r + \sum_{i\in C_1(M)} \|x_i - \pi_\Delta(x_i)\|^r + \sum_{i \in C_2(M)} \|y_j - \pi_\Delta(y_j)\|^r$$

et $W_r(dgm_1, dgm_2) = \left[\underset{M}{\text{Min}}\,(J_r(M))\right]^{1/r}$

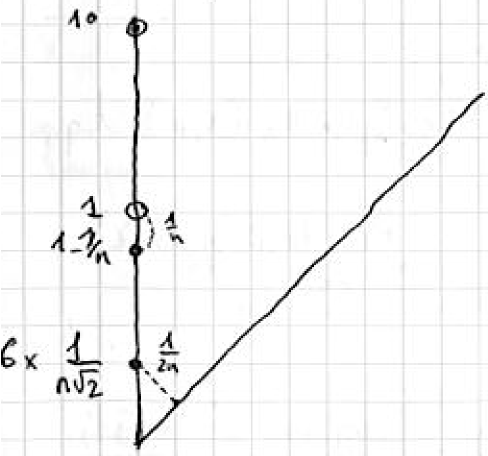
"distance de Wasserstein"  [Des résultats de stabilité existent aussi pour $W_r$, mais il faut des hypothèses.] ??

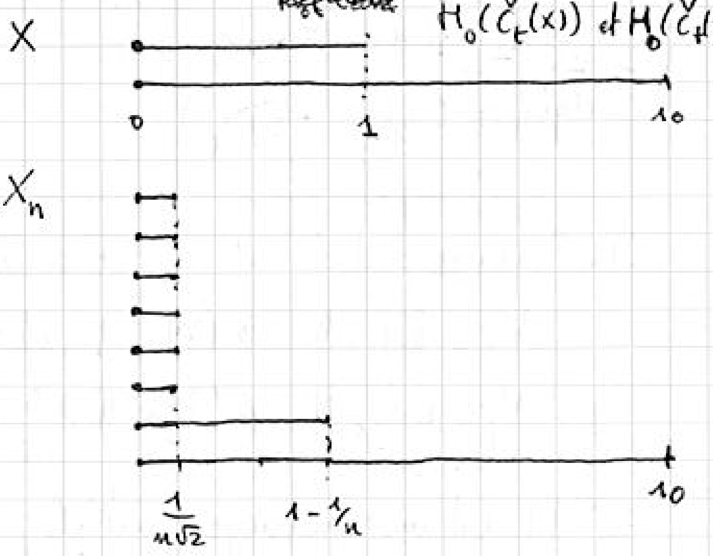
Un exemple : $X = \{a, b\} \subset \mathbb{R}^2$ $\|a-b\| = 2$

$X_n$ = 8 points à dist. $\frac{1}{n}$ de a ou b. comme sur le dessin.




Codes-barre pour $H_0(\check{C}_t(X))$ et $H_0(\check{C}_t(X_n))$




$$D_b(dgm(X), dgm(X_n)) = \frac{1}{n}$$

$$W_1(dgm(X), dgm(X_n)) = \frac{1}{n} + 6 \cdot \frac{1}{2n}$$

$$D_{GH}(X, X_n) = \frac{1}{n}$$

Le nombre de point du nuage $X_n$ apparaît dans le calcul de $W_1$. Possibilité dans l'esp. de Hilbert $\ell^2$ de construire $X_n$, $X = \{a, b\}$ tels que

$X_n \xrightarrow{G.H} X$ et $\lim_{n\to\infty} W_1(dgm(X), dgm(X_n)) = c > 0$.

# Une autre façon d'écrire $D_b$ et $W_r$.

$\overline{\mathcal{D}} = \mathcal{D} \cup \{\Delta\}$ (où $\mathcal{D} = \{(s,t) \mid 0 \leq s < t\}$)

$d : \overline{\mathcal{D}} \longrightarrow \mathbb{R}$ . $d(x,y) = \|x-y\|$ si $x,y \in \mathcal{D}$

. $d(x,\Delta) = d(\Delta,x) = \|x - \pi_\Delta(x)\|$ si $x \in \mathcal{D}$

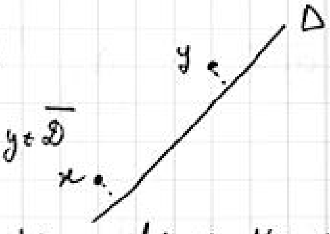
. $d(\Delta,\Delta) = 0$.

WARNING : $d$ n'est pas une distance sur $\overline{\mathcal{D}}$ :

$d(x,z) \leq d(x,y) + d(y,z)$ si $x,y,z \in \mathcal{D}$

$d(x,\Delta) \leq d(x,y) + d(y,\Delta)$ si $x \in \mathcal{D}, y \in \overline{\mathcal{D}}$

mais $d(x,z)$ ?? $d(x,\Delta) + d(\Delta,z)$ si $x,z \in \mathcal{D}$




$d(x,y) > d(x,\Delta) + d(\Delta,y)$

$dgm_1 = (x_1, \dots, x_n) \in \mathcal{D}^n$, $dgm_2 = (y_1, \dots, y_p) \in \mathcal{D}^p$

$\overline{dgm_1} = (x_1, \dots, x_n, \overbrace{\underset{x_{n+1}}{\Delta}, \Delta, \dots, \underset{x_{n+p}}{\Delta}}^{p}) \in (\overline{\mathcal{D}})^{n+p}$

$\overline{dgm_2} = (y_1, \dots\dots, y_p, \underbrace{\overset{y_{p+1}}{\Delta}, \dots \overset{y_{p+n}}{\Delta}}_{n})$

Si $\sigma \in \mathcal{S}_{n+p}$ $\overline{J}_\infty(\sigma) = \underset{i \in [\![1,n+p]\!]}{\mathrm{Max}} \left(d(x_i, y_{\sigma(i)})\right)$, $\overline{J}_r(\sigma) = \sum_{i=1}^{n+p} d(x_i, y_{\sigma(i)})^r$

$\sigma \in \mathcal{S}_{n+p} \overset{\Phi}{\longmapsto}$ Mariage partiel $\overset{\Phi(\sigma)}{M}$ de $dgm_1$ et $dgm_2$ (on oublie les $\Delta \longleftrightarrow \Delta$)

et $\overline{J}_\infty(\sigma) = J_\infty(M)$

$\overline{J}_r(\sigma) = J_r(M)$

Si $M$ est un mariage partiel, $\underbrace{\ }_{\Phi^{-1}(M) \neq \emptyset \text{ et}}$ $\forall \sigma \in \Phi^{-1}(M)$ $\overline{J}_\infty(\sigma) = J_\infty(M)$ et

$\overline{J}_r(\sigma) = J_r(M)$.

$$D_b(dgm_1, dgm_2) = \underset{\sigma \in \mathcal{S}_{n+p}}{\mathrm{Min}} \left(\overline{J}_\infty(\sigma)\right) = \underset{\sigma \in \mathcal{S}_{n+1}}{\mathrm{Min}} \left[\underset{i \in [\![1,n+p]\!]}{\mathrm{Max}} \left(d(x_i, y_{\sigma(i)})\right)\right]$$

$$W_r(dgm_1, dgm_2) = \left[\underset{\sigma \in \mathcal{S}_{n+p}}{\mathrm{Min}} \left(\sum_{i=1}^{n+p} d(x_i, y_{\sigma(i)})^r\right)\right]^{1/r}$$

Ecrites comme ça, ces distances ressemblent bcp à des distances de Wasserstein sur des espaces de mesures $\rightsquigarrow$ utile pour les calculs.

## Preuve de l'inégalité triangulaire.

$dgm_1 = (x_1, \dots, x_n)$ , $dgm_2 = (y_1, \dots, y_p)$ , $dgm_3 = (z_1, \dots, z_q)$

$$
\begin{aligned}
dgm_1 &= \overbrace{\Delta, \dots\dots, \Delta}^{q}, x_1, \dots\dots, x_n, \overbrace{\Delta, \dots\dots, \Delta}^{p} \\
& \qquad\qquad\qquad\qquad\qquad \Big\downarrow \sigma \\
dgm_2 &= \underbrace{\Delta, \dots\dots, \Delta}_{q}, y_1, \dots\dots, y_p, \overbrace{\Delta, \dots, \Delta}^{n} \\
& \qquad\qquad\qquad\qquad\qquad \Big\downarrow \tau \\
dgm_3 &= \underbrace{\Delta, \dots\dots, \Delta}_{p}, z_1, \dots\dots, z_q, \underbrace{\Delta, \dots, \Delta}_{n}
\end{aligned}
$$

On choisit $\sigma$ ~~et~~ $\tau \in \mathcal{S}_{n+p+q}$ t.q. $D_b(dgm_1, dgm_2) = \overline{J}_\infty(\sigma)$

et $D_b(dgm_2, dgm_3) = \overline{J}_\infty(\tau)$

$$D_b(dgm_1, dgm_3) \leq \overline{J}_\infty(\tau \circ \sigma) = \underset{i \in [\![1, n+p+q]\!]}{\mathrm{Max}} \left( d(x_i, z_{\tau\circ\sigma(i)}) \right)$$

<u>Rque</u>: $y_{\sigma(i)} = \Delta \iff \left( x_i = \Delta \text{ ou } z_{\tau\circ\sigma(i)} = \Delta \right)$

on évite le cas problématique d' l'inég. triangulaire.

D'où $\forall i \in [\![1, n+p+q]\!]$ $d(x_i, z_{\tau\circ\sigma(i)}) \leq d(x_i, y_{\sigma(i)}) + d(y_{\sigma(i)}, z_{\tau\circ\sigma(i)})$

$$
\begin{aligned}
D_b(dgm_1, dgm_3) &\leq \mathrm{Max}_i \left( d(x_i, y_{\sigma(i)}) \right) + \mathrm{Max}_i \left( d(y_{\sigma(i)}, z_{\tau(\sigma(i))}) \right) \\
&\leq \overline{J}_\infty(\sigma) + \overline{J}_\infty(\tau) = D_b(dgm_1, dgm_2) + D_b(dgm_2, dgm_3).
\end{aligned}
$$

## Lien avec la distance de Wasserstein usuelle:

$(E, d_E)$ esp. métrique , $c(x,y) = d_E(x,y)^r$ coût de transport d'une unité de masse de $x$ à $y$.

$\mu, \nu \in \mathcal{P}(E) = \{$mesures de proba sur $E\}$

Un "transport" $\gamma$ de $\mu$ à $\nu$ a un coût global $\overline{J}_r(\gamma)$ qui est une "somme des coût ponctuels".

$\widetilde{W}_r(\mu, \nu) = \left[ \underset{\gamma}{\mathrm{Inf}} (\overline{J}_r(\gamma)) \right]^{1/r}$ est une distance de Wasserstein sur $\mathcal{P}(E)$

(Il faut une bonne notion de "transport" pour que l'inf soit fini.)

Cas particulier : $\mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}$ et $\nu = \frac{1}{N} \sum_{j=1}^{N} \delta_{y_j}$

On montre dans ce cas que

$$\widetilde{W}_r(\mu,\nu) = \frac{1}{N^{1/r}} \left[ \min_{\sigma \in \mathcal{S}_N} \left( \sum_{i=1}^{N} d_E(x_i, y_{\sigma(i)})^r \right) \right]^{1/r}$$

Retour aux diagrammes :

$$dgm_1 = (x_1, \dots, x_n) \rightsquigarrow \mu_1 = \frac{1}{n+p} \left( \sum_{i=1}^{n} \delta_{x_i} + p\,\delta_\Delta \right)$$

$$dgm_2 = (y_1, \dots, y_p) \rightsquigarrow \mu_2 = \frac{1}{n+p} \left( \sum_{j=1}^{p} \delta_{y_j} + n\,\delta_\Delta \right)$$

$$\mu_1, \mu_2 \in \mathcal{P}(\overline{\mathcal{D}})$$

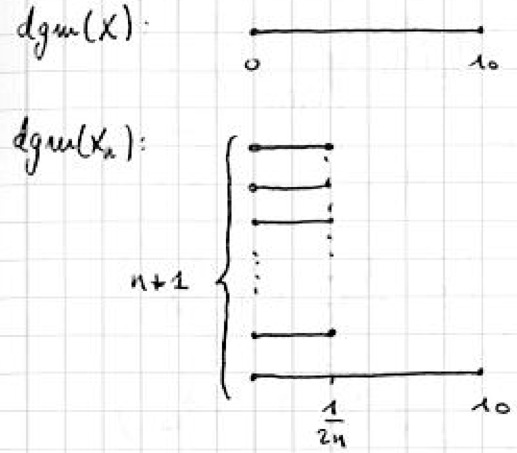
$$W_r(dgm_1, dgm_2) = (n+p)^{1/r}\, \widetilde{W}_r(\mu_1, \mu_2)$$

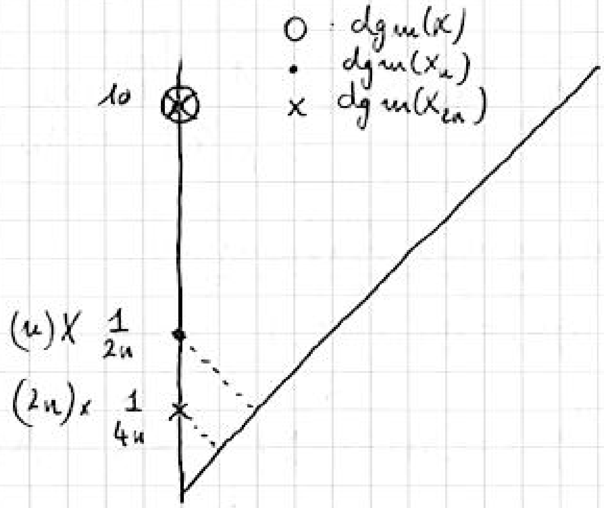
Deux remarques :

- $d$ n'est pas une distance sur $\overline{\mathcal{D}}$, donc le $\widetilde{W}_r$ construit à partir de $d$ n'est pas une distance sur $\mathcal{P}(\overline{\mathcal{D}})$.
- La mesure $\mu_1$ dépend de $dgm_1$ et $dgm_2$ (à travers $p$). De même pour $\mu_2$. En particulier, on ne plonge pas l'espace des diagrammes dans un espace de mesures.

Un autre exemple : $H_0(\check{C}_t(X))$ et $H_0(\check{C}_t(X_n))$ où

$$X = [0,1] \quad , \quad X_n = \left\{ \frac{k}{n} \mid k \in [\![0,1]\!] \right\}$$




. $D_b(dgm(X), dgm(X_n)) = \frac{1}{n.2\sqrt{2}}$ et $W_1(dgm(X), dgm(X_n)) = n \times \frac{1}{n2\sqrt{2}} = \frac{1}{2\sqrt{2}}$

$D_{GH}(X, X_n) = \frac{1}{2n}$

. Pour comparer deux nuages de points : le mariage optimal pour $D_b$ et $W_1$ consiste à envoyer les $n$ points $(0, \frac{1}{2n})$ de dgm($X_n$) sur $n$ points $(0, \frac{1}{4n})$ de dgm($X_{2n}$), ce qui donne

$D_b(dgm(X_n), dgm(X_{2n})) = \frac{1}{4n}$ et $W_1(dgm(X_n), dgm(X_{2n})) = \frac{1}{4} + \frac{1}{4\sqrt{2}}$

mais $\lim_{n \to \infty} D_{GH}(X_n, X_{2n}) = 0$