Fréchet Means for Distributions of Persistence Diagrams
(Turner, Mileyko, Mukherkee, Harer 2014)

Idea: compute means and variance wrt Wasserstein distances

$$d_{L^2}(X,Y) = \sqrt{\inf_{\phi: X \to Y} \sum_{x \in X} \| x - \phi(x) \|^2}$$

Let $D_{L^2} = \{ X \mid d_{L^2}(X, \emptyset) < \infty \}$. The Fréchet function of a prob. distribution ρ on D_{L^2} is

$$F_\rho : D_{L^2} \to \mathbb{R} \qquad Y \longmapsto \int_{D_{L^2}} d(X,Y)\, d\rho(X)$$

The Fréchet Variance is $\inf_{Y \in D_{L^2}} F_\rho(Y)$ and any Y which achieves this is a Fréchet Mean.

For a finite sample of diagrams $X_1, \ldots, X_m$ we can consider the distribution $\rho = \frac{1}{m}\sum_{i=1}^{m} \delta_{X_i}$

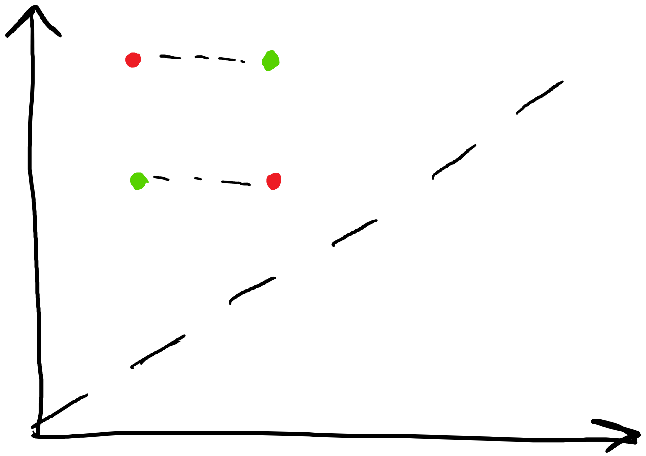
Then $F_\rho(Y) = \frac{1}{m}\sum_{i=1}^{m} d(X_i, Y)$

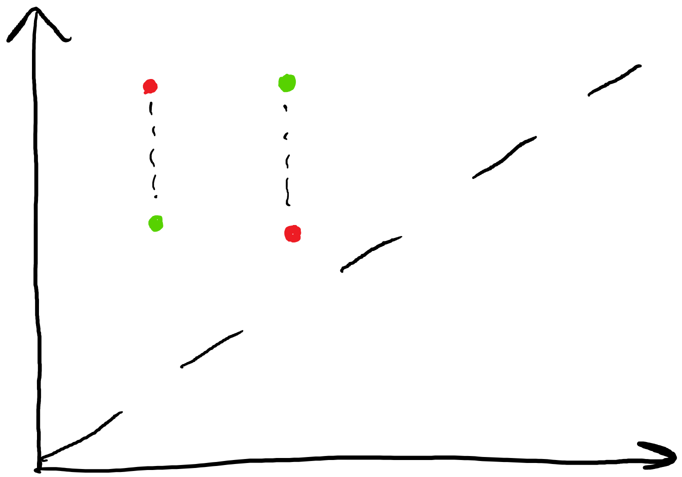
Computation of Fréchet means is well-studied for CAT(k) spaces (~ spaces with curvature bdd above), but ...

Proposition 2.4. *D_{L^2} is not in CAT(k) for any $k > 0$.*

proof: $X, Y \in D_{L^2} \in CAT(k)$ with $d(X,Y) < \frac{\pi}{\sqrt{k}}$

$\Rightarrow \exists$ unique geodesic from X to Y. But





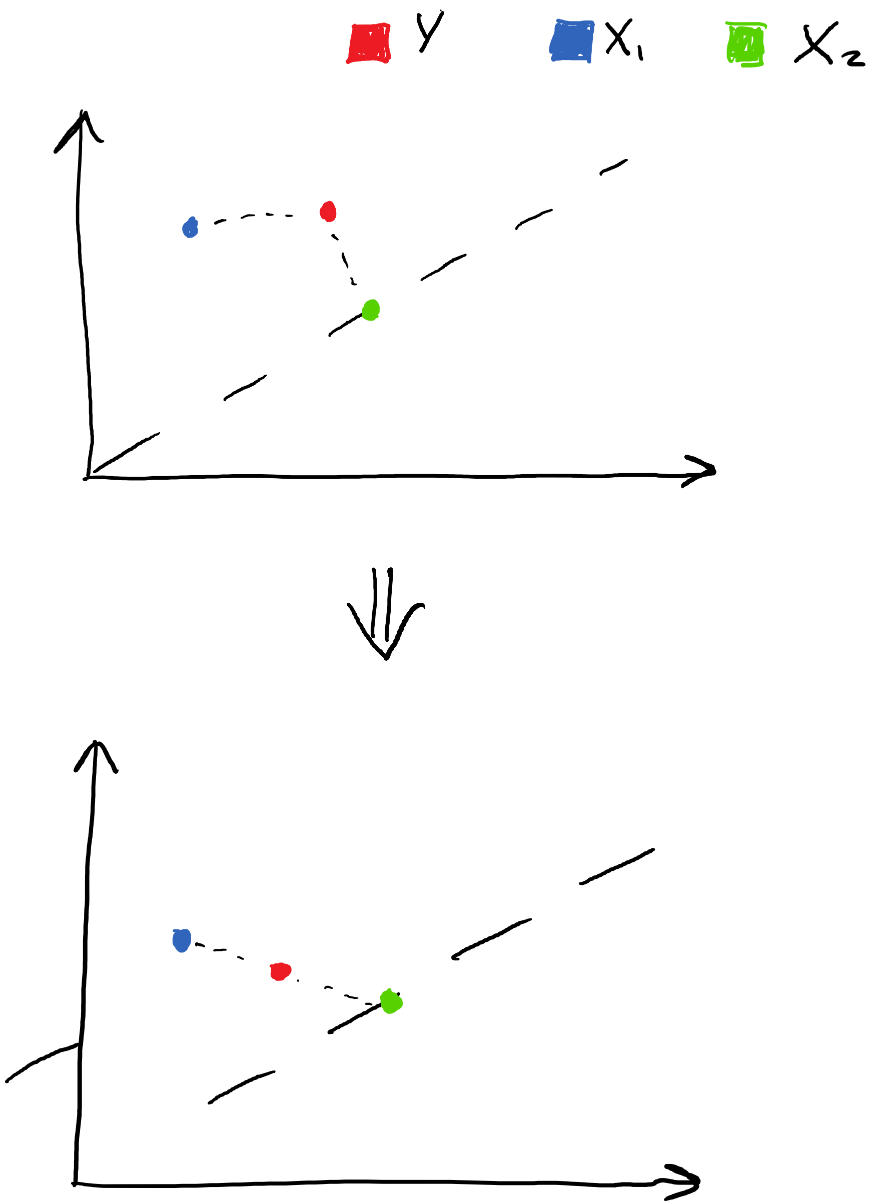
Algorithm 1: Algorithm for computing the Fréchet mean Y from persistence diagrams $X_1, \dots, X_m$.

input : persistence diagrams $\{X_1, \dots, X_m\}$

return: Fréchet mean $\{Y\}$

Draw $i \sim \text{Uniform}(1, ..., n)$; `/* randomly draw a diagram */`
Initialize $Y \leftarrow X_i$; `/* initialize Y */`

stop $\leftarrow$ false ;
repeat
  $K = |Y|$; `/* the number of non-diagonal points in Y */`
  for i=1,. . . , m **do**
    $(y^j, x_i^j) \leftarrow \text{Hungarian}(Y, X_i)$; `/* compute optimal pairings between each X_i and Y using the Hungarian algorithm */`
  for j=1,. . . K **do**
    $y^j \leftarrow \text{mean}_{i=1,..,m}(x_i^j)$ `/* set each non-diagonal point in Y to the arithmetic mean of its pairings */`
  if $\mathit{Hungarian}(Y, X_i) = (y_j, x_i^j)$ **then** stop $\leftarrow$ true `/* The points in the updated Y are optimal pairings w.r.t. each X_i */`
until *stop=true*;
return: Y

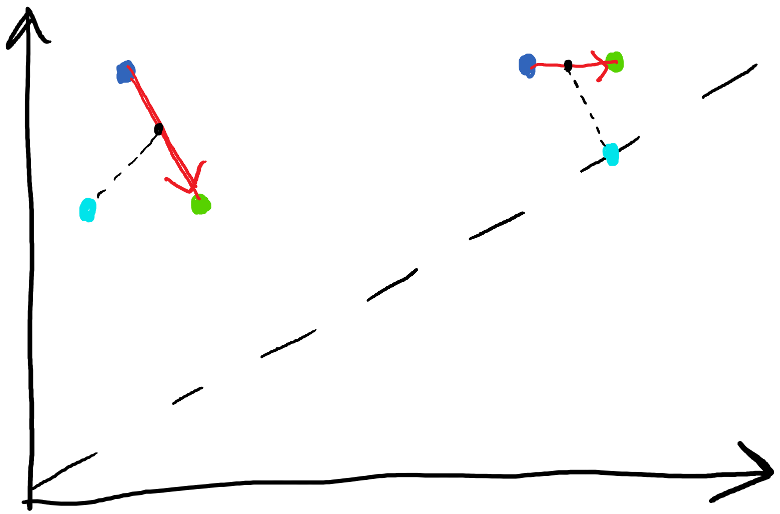


Theorem 2.5. *The space of persistence diagrams $\mathcal{D}_{L^2}$ with metric d given in (1) is a non-negatively curved Alexandrov space.*

iff

$$d'(Z,\gamma(t))^2 \geq t d'(Z,Y)^2 + (1-t)d'(Z,X)^2 - t(1-t)d'(X,Y)^2. \tag{4}$$

$\forall \gamma \, (X \rightsquigarrow Y)$, Z



proof idea: Let $\phi : X \to Y$ induce γ.

Construct pairings

(i) $Z \to X$ via $Z \xrightarrow{\text{opt}} \gamma(t) \longrightarrow X$

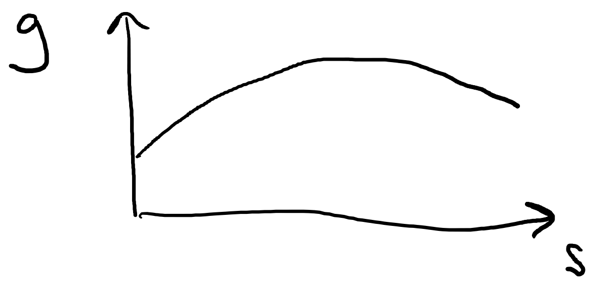
(ii) $Z \to Y$ via $Z \xrightarrow{\text{opt}} X \xrightarrow{\phi} Y$

Then for each pt in Z (4) holds, but the pairings might not be optimal, so its an inequality.

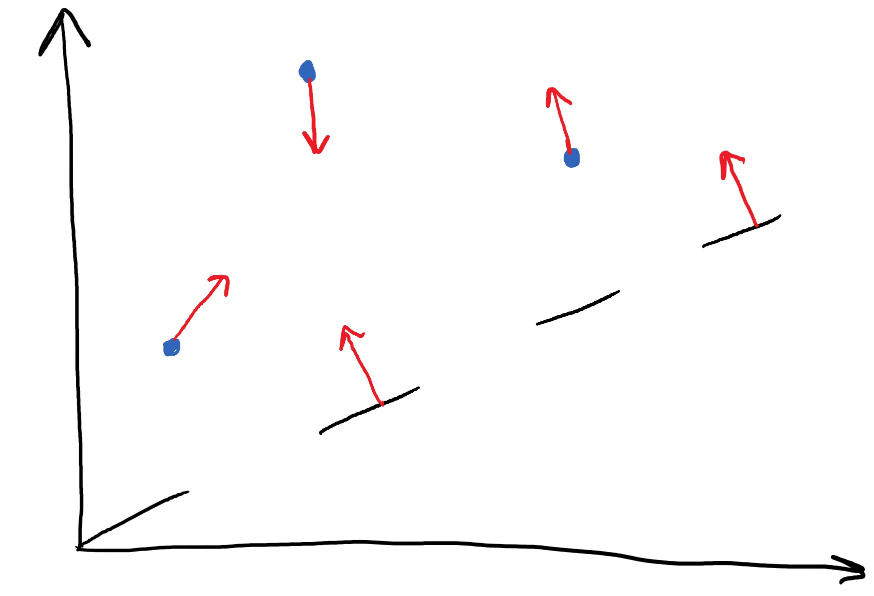
Proposition 2.6. *If the support of ρ is bounded (as in has bounded diameter) then the corresponding Fréchet function is semiconcave.*

In fact it's 2-concave — i.e. $\forall \gamma$ unit speed

$$g = F_\rho \circ \gamma(s) - s^2 \text{ is concave}$$



proof: follows from ineq. (4)



$$\hat{\Sigma}_y = \left\{ \gamma \;\middle|\; \begin{array}{l} \text{unit speed,} \\ \text{starts at } y \end{array} \right\}$$

$$\angle_y(\gamma, \eta) = \cos^{-1}\left(\lim_{s,t \downarrow 0} \frac{s^2 + t^2 - d(\gamma(s), \eta(t))^2}{2st} \right)$$

Σ = completion of $\hat{\Sigma}_y$ wrt $\angle_y$.

Tangent cone $T_y = \dfrac{\Sigma \times [0, \infty)}{\Sigma \times \{0\}}$.

$$\langle (\gamma, s), (\eta, t) \rangle = st \cos \angle_y(\gamma, \eta)$$

Definition 2.7 (Gradients and supporting vectors). Given an open set $\Omega \subset \mathcal{A}$ and a function $f : \Omega \to \mathbb{R}$ we denote by $\nabla_p f$ the *gradient* of a function f at a point $p \in \Omega$. $\nabla_p f$ is the vector $v \in T_p$ such that

(i) $d_p f(x) \leq \langle v, x \rangle$ for all $x \in T_p$
(ii) $d_p f(v) = \langle v, v \rangle$.

For a semiconcave f the gradient exists and is unique (Theorem 1.7 in [15]). We say $s \in T_p$ is a *supporting vector* of f at p if $d_p f(x) \leq -\langle s, x \rangle$ for all $x \in T_p$. Note that $-\nabla_p f$ is a supporting vector if it exists in the tangent cone at p.

Lemma 2.8. (i) *If s is a supporting vector then $\|s\| \geq \|\nabla_p f\|$.*
(ii) *If p is local minimum of f and s is a supporting vector of f at p then $s = 0$.*

(i) $0 \leq \langle \nabla f + s, \nabla f + s \rangle = \langle \nabla f, \nabla f \rangle + 2 \langle \nabla f, s \rangle + \langle s, s \rangle$

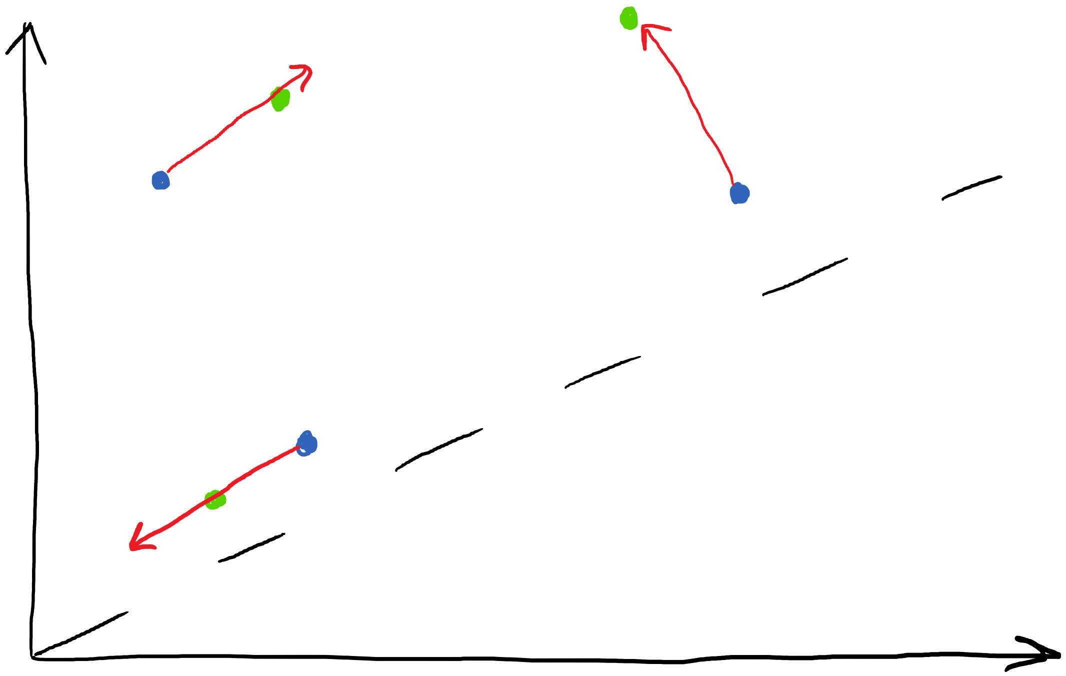
$\langle \nabla f, \nabla f \rangle = df(\nabla f) \leq -\langle s, \nabla f \rangle$

(ii) $d_p f(s) \geq 0$, so $-\langle s, s \rangle \geq d_p f(s) \geq 0$

$\Rightarrow \langle s, s \rangle = 0 \Rightarrow s = 0$.

Proposition 2.9. *Let $Y \in \mathcal{D}_{L^2}$. For each $X \in \mathcal{D}_{L^2}$ let $F_X : Z \mapsto d(X,Z)^2$.*

(i) *If γ is a distance achieving geodesic from Y to X, then the tangent vector to γ at Y of length $2d(X,Y)$ is a supporting vector at Y for F_X.*
(ii) *If s_X is a supporting vector at Y for the function F_X for each $X \in supp(\rho)$ then $s = \int s_X d\rho(X)$ is a supporting vector at Y of the Fréchet function F corresponding to the distribution ρ.*



in particular, if

$\rho = \frac{1}{m}\sum_i \delta_{x_i}$, then

s is just an average.

Lemma 3.1. *If $W = \{w_i\}$ is a local minimum of the Fréchet function $F = \frac{1}{m}\sum_{j=1}^{m} F_j$ F then there is a unique optimal pairing from W to each of the X_j which we denote as ϕ_j and each w_i is the arithmetic mean of the points $\{\phi_j(w_i)\}_{j=1,2\ldots m}$. Furthermore if w_k and w_l are off-diagonal points such that $\|w_k - w_l\| = 0$ then $\|\phi_j(w_k) - \phi_j(w_l)\| = 0$ for each j.*

Let s_j be the vectors in TX_j tangent to the pairings ϕ_j of length $d(X_j, Y)$. Then $\frac{2}{m}\sum_j s_j$ is a support vector for F.

So we know $\frac{2}{m}\sum_j s_j = 0$. But now this implies each pt of W is the mean of the points its paired with.

Proposition 3.2. *Let X and Y be diagrams, each with only finitely many off diagonal points, such that there is a unique optimal pairing ϕ_X^Y between them and no off diagonal point in X matches the diagonal in Y. We further stipulate that if y_k and y_l are off-diagonal points with $\|y_k - y_l\| = 0$ then $\|(\phi_X^Y)^{-1}(y_k) - (\phi_X^Y)^{-1}(y_l)\| = 0$. There is some $r > 0$ such that for every $Z \in B(Y, r)$ there is a unique optimal pairing between X and Z and this optimal pairing is induced from the one from X to Y. By this we mean there is a unique optimal pairing ϕ_Y^Z from Y to Z and that the unique optimal pairing from X to Z is $\phi_Y^Z \circ \phi_X^Y$.*

Furthermore, if $X_1, X_2, \ldots, X_m$ and Y are diagrams with finitely many off-diagonal points such that there is a unique optimal pairing $\phi_{X_i}^Y$ between X_i and Y for each i with the same conditions as above, then there is some $r > 0$ such that for every $Z \in B(Y, r)$ there is a unique optimal pairing between each X_i and Z and this optimal pairing is induced by the one from X_i to Y.

Now note that the mean of $\{x_1, \ldots, x_n\}$ minimises $\sum_i \|x_i - y\|^2$.

So as long as the pairings stay the same nearby, the mean is a local minima