The Hidden Topology of a Noisy Point Cloud

(Part III)

A critical reading of
"Geometric Inference for Probability Measures"
by Chazal, Steiner & Merigot, 2011.

A. Pedrini, M. Piastra

Measures

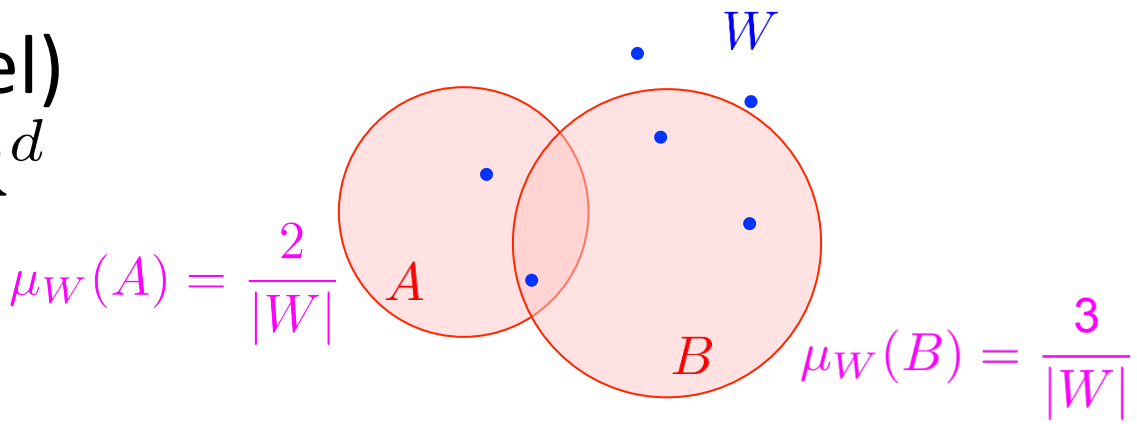
- A measure μ is a function that maps every (Borel) subset $A \subseteq \mathbb{R}^d$ to a non-negative number $\mu(A)$ and which is *countably additive*
- It is *finite* if $\mu(\mathbb{R}^d) < \infty$
  A *probability* measure is such that $\mu(\mathbb{R}^d) = 1$
  (we will consider *probability* measures only, from now on)
- The *support* of μ is the smallest closed subset K such that $\mu(\mathbb{R}^d \setminus K) = 0$

A measure from a point cloud

- Let W be a finite point cloud
- The *uniform measure* on W is defined as

$$\mu_W(A) := \frac{|A \cap W|}{|W|}$$

for every (Borel) subset $A \subseteq \mathbb{R}^d$



The uniform measure is a special case: it could be different

Distance to a compact

- Distance to a compact subset $C \subset \mathbb{R}^d$

$$d_C(x) := \inf_{y \in C} |x - y|$$

- Alternative formulation

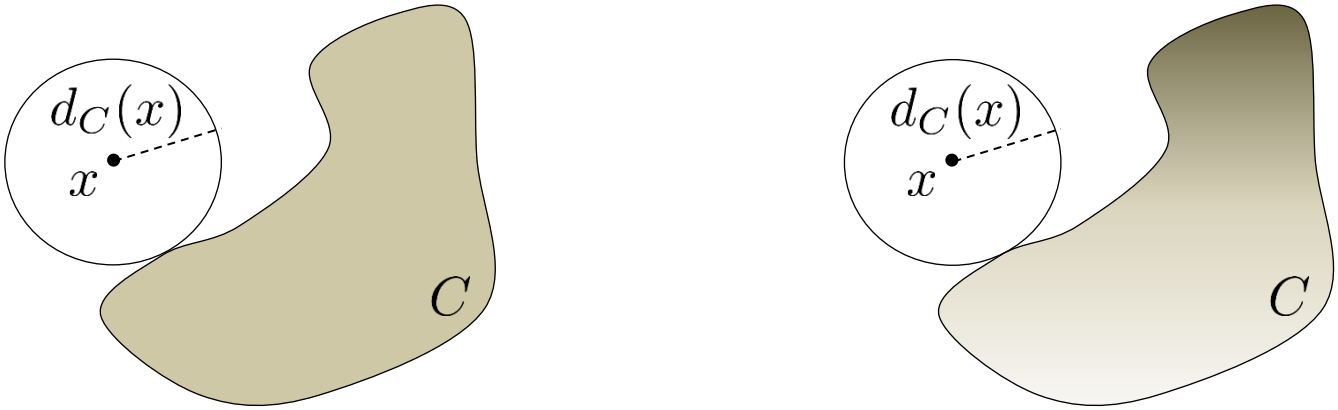
$$d_C(x) := \inf\{r \mid \overline{B}_r(x) \cap C \neq \emptyset\}$$

where $\overline{B}_r(x)$ is a closed ball of radius r centered in x

From a compact to a measure

- Let μ_C be a measure having the compact C as its support
- The distance to C is

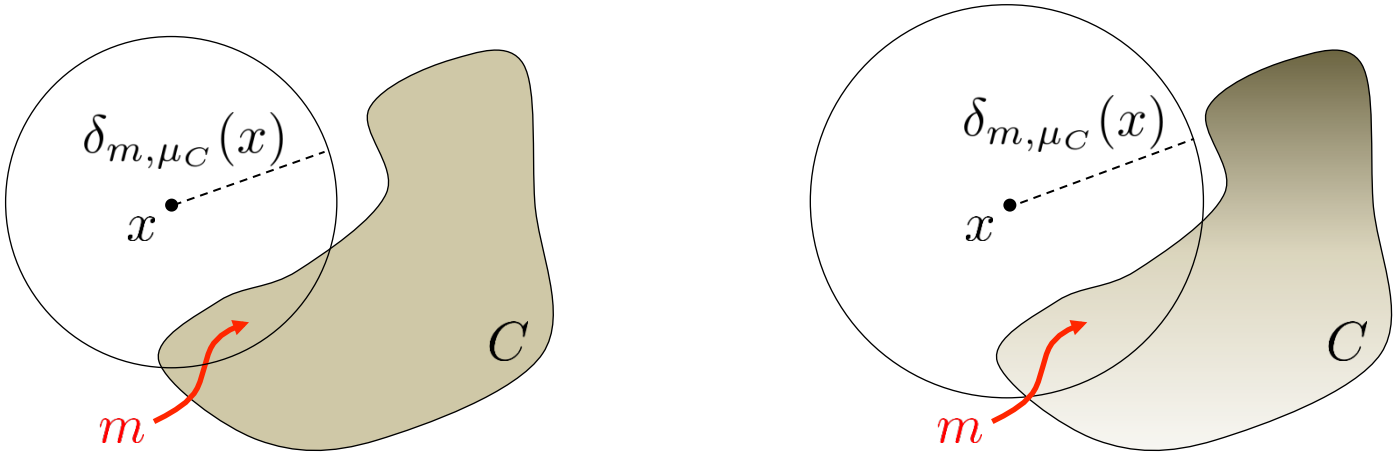
$$d_C(x) := \inf\{r \mid \mu_C(\overline{B}_r(x)) > 0\}$$



A pseudo-distance

- Let μ_C be a measure having the compact C as its support
- As a generalization, define (for any $m \in [0, 1)$)

$$\delta_{m,\mu_C}(x) := \inf\{r \mid \mu_C(\overline{B}_r(x)) > m\}$$



A pseudo-distance

- For a point cloud W, with uniform μ_W:

$$\delta_{m,\mu_W}(x) := \inf\{r \mid \mu_W(\overline{B}_r(x)) > m\}$$

$$= \inf\{r \mid \frac{|\overline{B}_r(x) \cap W|}{|W|} > m\}$$

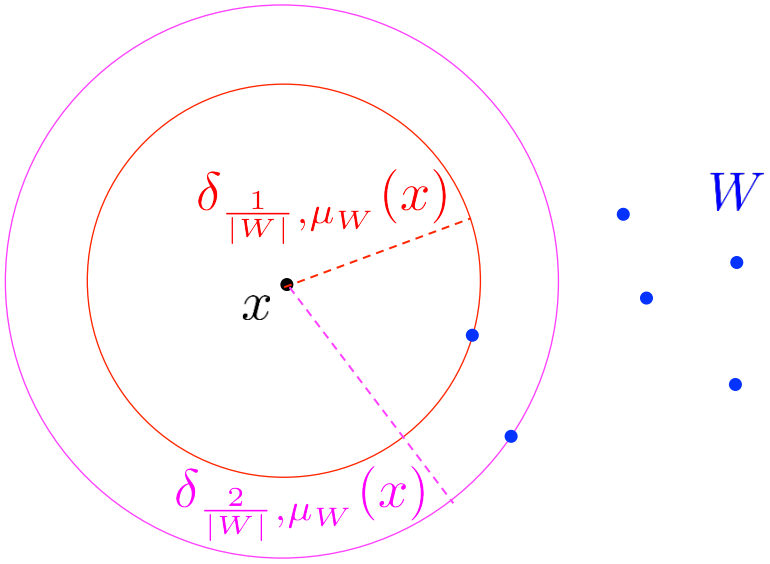
$$= d(x, w_k), \ \frac{k-1}{|W|} \leq m < \frac{k}{|W|}$$

where w_k is the k-nearest point to x in W

A pseudo-distance

$$\delta_{m,\mu_W}(x) := d(x, w_k), \quad \frac{k-1}{|W|} \leq m < \frac{k}{|W|}$$

where w_k is the k-nearest point from x in W



A pseudo-distance

- However, δ_{m,μ_C} is ***not*** distance-like

  This entails that neither the *isotopy lemma* nor the *reconstruction theorem* apply to δ_{m,μ_C}

- Furthermore, the map $\mu_C \mapsto \delta_{m,\mu_C}$ is ***not*** continuous in any reasonable sense

  A slight perturbation that turns μ_C into μ'_C can induce substantial differences between δ_{m,μ_C} and δ_{m,μ'_C}

A pseudo-distance

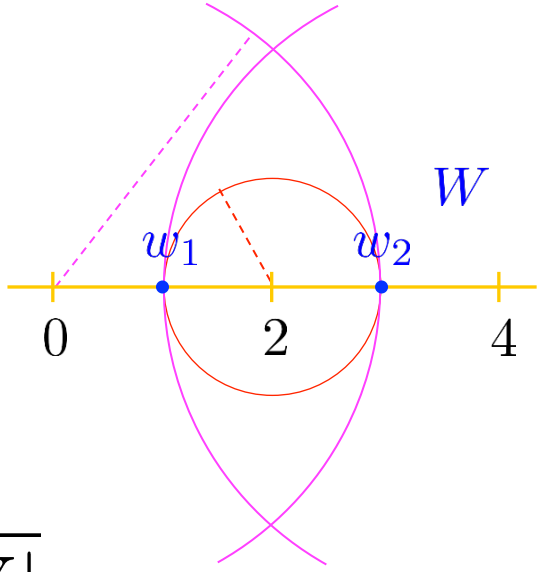
- δ_{m,μ_W} is *not* distance like

  Counter-example:

$$\mu_W(w_1) = \mu_W(w_2) = \frac{1}{2}$$

$$m := \frac{1}{2} \Rightarrow \frac{k-1}{|W|} \leq \frac{1}{2} < \frac{k}{|W|}$$

$$\Rightarrow k = 2$$



  i.e. $\delta_{\frac{1}{2},\mu_W}$ is the distance to the second-nearest point in W

A pseudo-distance

- δ_{m,μ_W} is *not* distance like

  Counter-example:

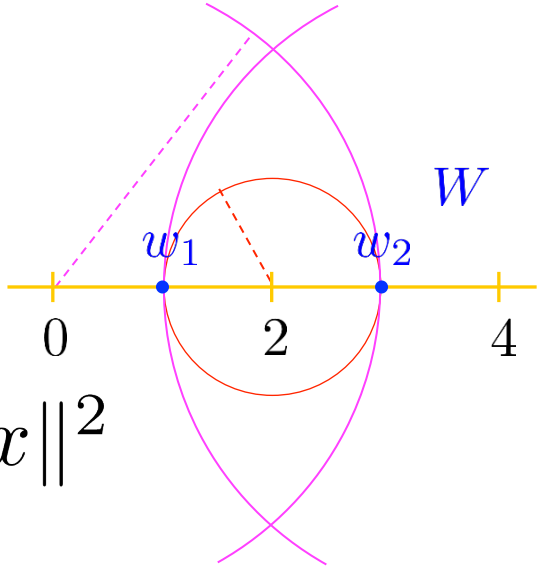
  for $\delta^2_{\frac{1}{2},\mu_W}$ to be *semi-concave*

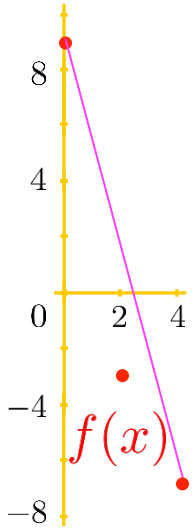
$$x \mapsto f(x) := \delta_{\frac{1}{2},\mu_W}(x)^2 - \|x\|^2$$

must be *concave.*

However:

$$f(0) = |w_2 - 0|^2 - 0^2 = 3^2 = 9$$
$$f(2) = |w_1 - 2|^2 - 2^2 = 1^2 - 2^2 = -3$$
$$f(4) = |w_1 - 4|^2 - 4^2 = 3^2 - 4^2 = -7$$

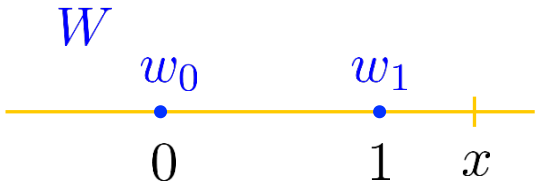




$f(x)$ *is* *not* concave

A pseudo-distance

- The map $\mu_W \mapsto \delta_{m,\mu_W}$ is *not* continuous in any reasonable sense

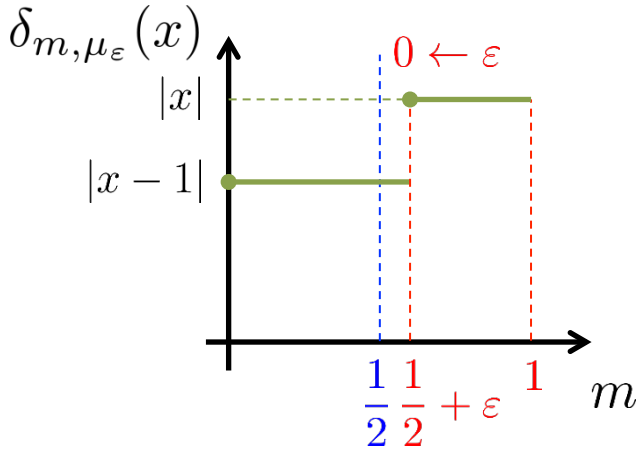


$$\mu_\varepsilon(w_0) := \frac{1}{2} - \varepsilon \qquad \mu_\varepsilon(w_0) := \frac{1}{2} + \varepsilon$$

For $x > \dfrac{1}{2}$ *:*

$$\varepsilon > 0 \;\Rightarrow\; \delta_{1/2,\mu_\varepsilon}(x) = |x-1|$$

$$\varepsilon = 0 \;\Rightarrow\; \delta_{1/2,\mu_\varepsilon}(x) = |x|$$



i.e. the function $\varepsilon \mapsto \delta_{1/2,\mu_\varepsilon}$ *is not continuous*

Distance to a measure

- Let μ_C be a measure having the compact C as its support
- Define, for any $m \in [0, 1)$

$$d_{m,\mu_C}(x) := \left(\frac{1}{m} \int_0^m \delta^2_{u,\mu_C}(x)\, \mathrm{d}u \right)^{\frac{1}{2}}$$

- This *is* distance-like (see article)

Distance to a measure

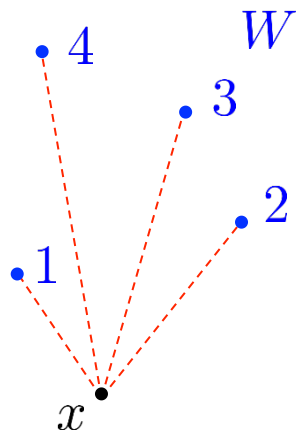
- In the case of a point cloud W assuming $m_0 = k_0/|W|$ the distance becomes

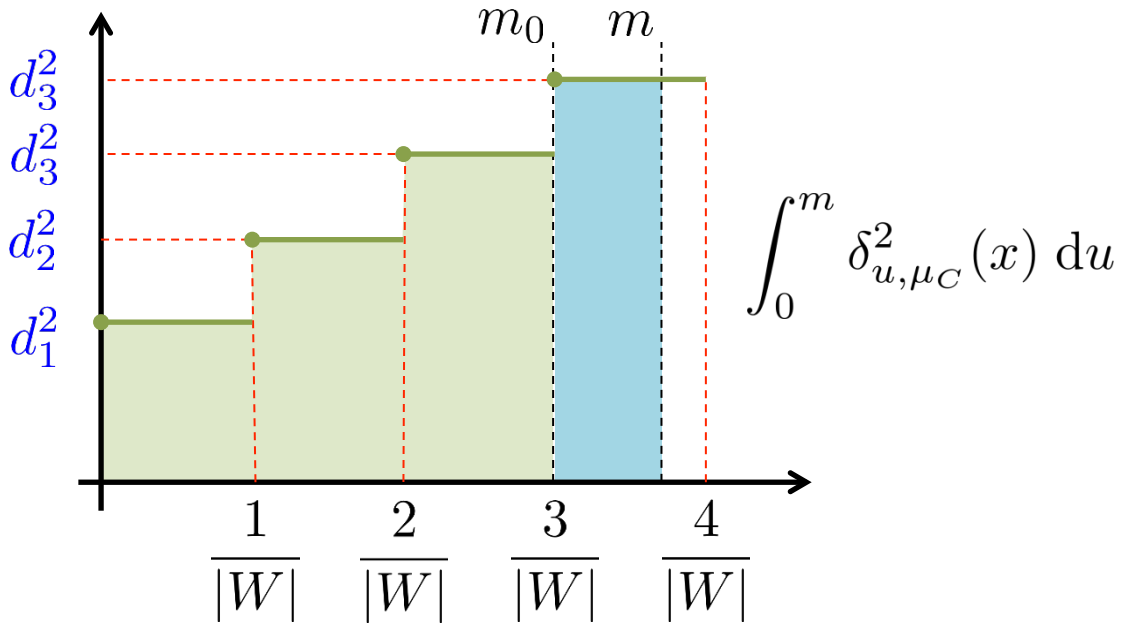
$$
\begin{aligned}
d_{m_0,\mu_W}(x) &:= \left(\frac{1}{m_0} \int_0^{m_0} \delta_{u,\mu_W}^2(x)\, \mathrm{d}u \right)^{\frac{1}{2}} \\
&= \left(\frac{1}{m_0} \sum_{k=1}^{k_0} \frac{1}{|W|} d^2(x, w_k) \right)^{\frac{1}{2}} \\
&= \left(\frac{1}{k_0} \sum_{k=1}^{k_0} d^2(x, w_k) \right)^{\frac{1}{2}}
\end{aligned}
$$

Distance to a measure

$$d_{m_0,\mu_W}(x) := \left(\frac{1}{k_0} \sum_{k=1}^{k_0} d^2(x, w_k) \right)^{\frac{1}{2}}$$

$$m_0 = k_0/|W|$$





Distance to a measure

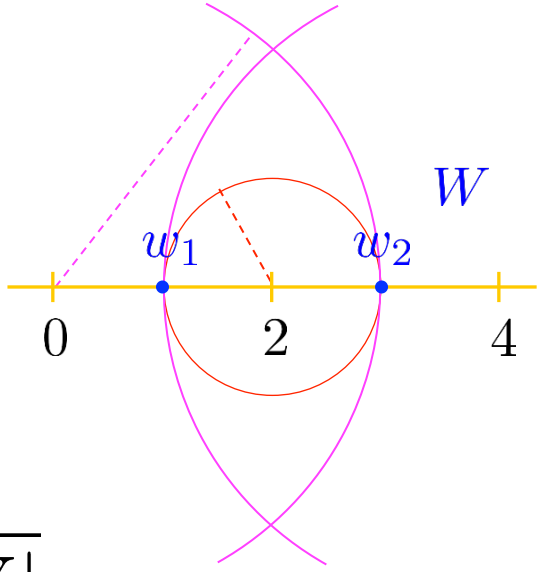
- d_{m_0,μ_W} *is* distance-like

  Not a proof, just an example:

$$\mu_W(w_1) = \mu_W(w_2) = \frac{1}{2}$$

$$m := \frac{1}{2} \Rightarrow \frac{k-1}{|W|} \leq \frac{1}{2} < \frac{k}{|W|}$$

$$\Rightarrow k = 2$$



Distance to a measure

- d_{m_0,μ_W} *is* distance-like

  Not a proof, just an example:

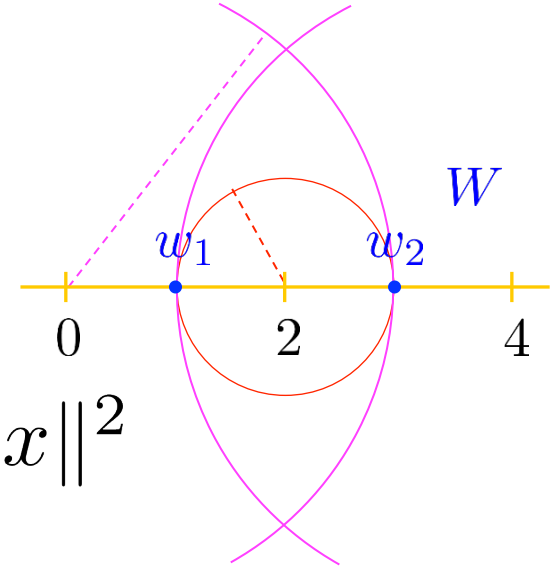
  for $d^2_{\frac{1}{2},\mu_W}$ to be *semi-concave*

  $$x \mapsto f(x) := d_{\frac{1}{2},\mu_W}(x)^2 - \|x\|^2$$

  must be *concave.*

  We have:

  $$f(0) = \left(\tfrac{1}{2}(|1-0|^2 + |3-0|^2)\right)^{\frac{1}{2}} - 0^2 = \sqrt{5}$$
  $$f(2) = \left(\tfrac{1}{2}(|1-2|^2 + |3-2|^2)\right)^{\frac{1}{2}} - 2^2 = -3$$
  $$f(4) = \left(\tfrac{1}{2}(|1-4|^2 + |3-4|^2)\right)^{\frac{1}{2}} - 4^2 = \sqrt{5} - 16$$



Distance to a measure

- d_{m_0,μ_W} *is* distance-like

  Not a proof, just an example:

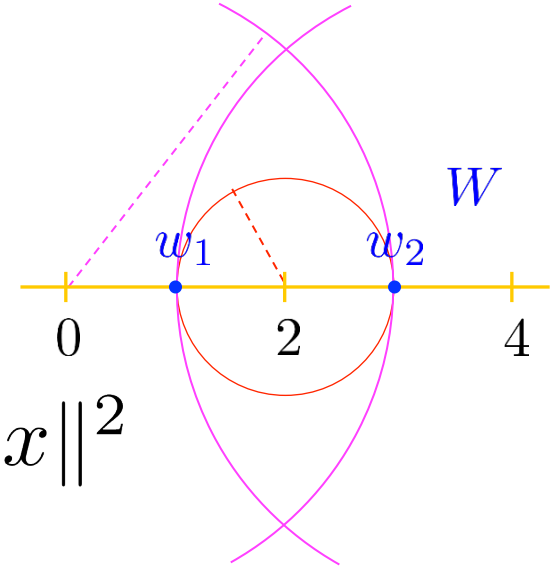
  for $d^2_{\frac{1}{2},\mu_W}$ to be *semi-concave*

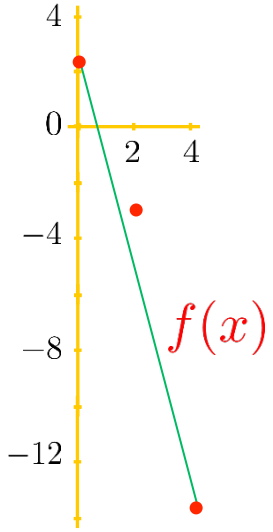
$$x \mapsto f(x) := d_{\frac{1}{2},\mu_W}(x)^2 - \|x\|^2$$

must be *concave.*

We have:

$$f(0) = \sqrt{5}$$
$$f(2) = -3$$
$$f(4) = \sqrt{5} - 16$$





$f(x)$ ***is*** concave

Being distance-like

- The Isotopy Lemma holds for $d_{m,\mu}$
  (see article)

- Given two numbers $0 < r_1 < r_2$ such that there are no 0-critical points in $d_{m,\mu}^{-1}([r_1, r_2])$, all the sublevel sets

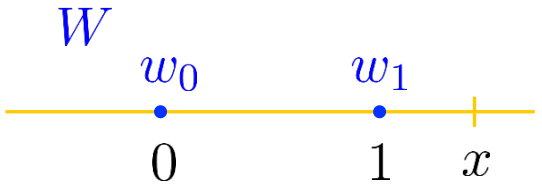
$$d_{m,\mu}^{-1}([0, r]), \ \forall r \in [r_1, r_2]$$

  are *isotopic*

Distance to a measure

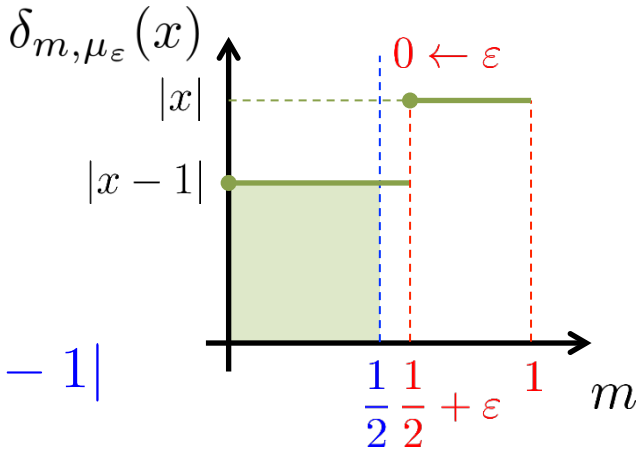
- The map $\mu_W \mapsto d_{m_0,\mu_W}$ is continuous

Not a proof, just an example



$$\mu_\varepsilon(w_0) := \frac{1}{2} - \varepsilon \qquad \mu_\varepsilon(w_0) := \frac{1}{2} + \varepsilon$$

For $x > \frac{1}{2}$ *:*

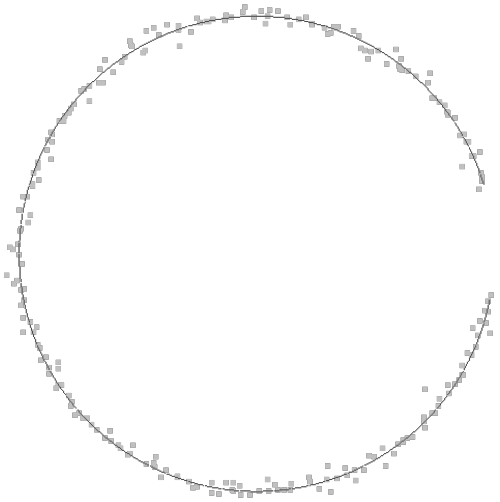


$$d_{\frac{1}{2},\mu_\varepsilon}(x) = \left(2\int_0^{\frac{1}{2}} |x-1|^2 du\right)^{\frac{1}{2}} = |x-1|$$

(*A rigorous definition requires Wasserstein distance*)

Distances and level sets

- Topology of the *sublevel sets* of d_{m_0,μ_W}

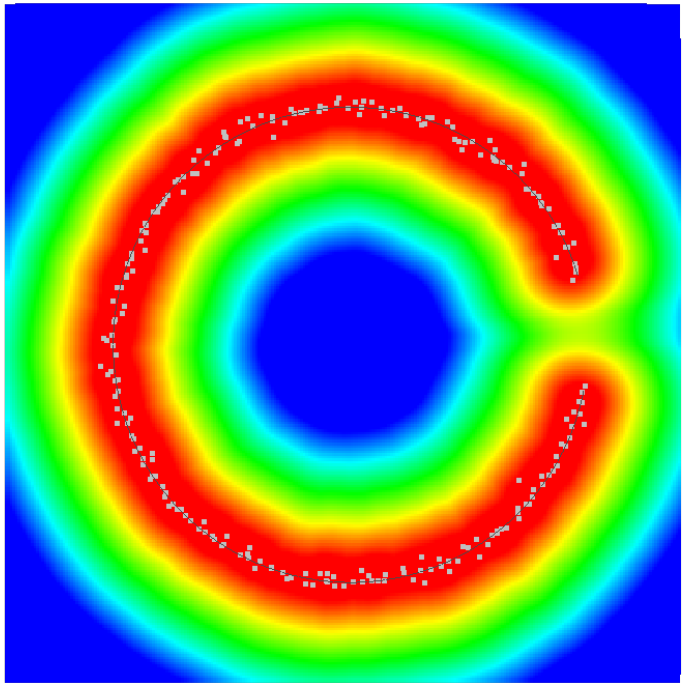


In black: C_{shape}
Grey dots: W
238 point samples from
$\mathcal{U}(C_{shape}) * \mathcal{N}(0, 0.03)$

Which values of m_0 guarantee the correct *homotopy type*?
A rigorous answer requires Wasserstein distance

Distances and level sets

- Topology of the *sublevel sets* of d_{m_0,μ_W}



In black: C_{shape}
Grey dots: W
238 point samples from
$\mathcal{U}(C_{shape}) * \mathcal{N}(0, 0.03)$

In color:
$d_{m_0,\mu_W}(x),\ m_0 = 0.029$
$(i.e.\ k_0 = 7)$

Which values of m_0 guarantee the correct *homotopy type*?
A rigorous answer requires Wasserstein distance

Distances and level sets

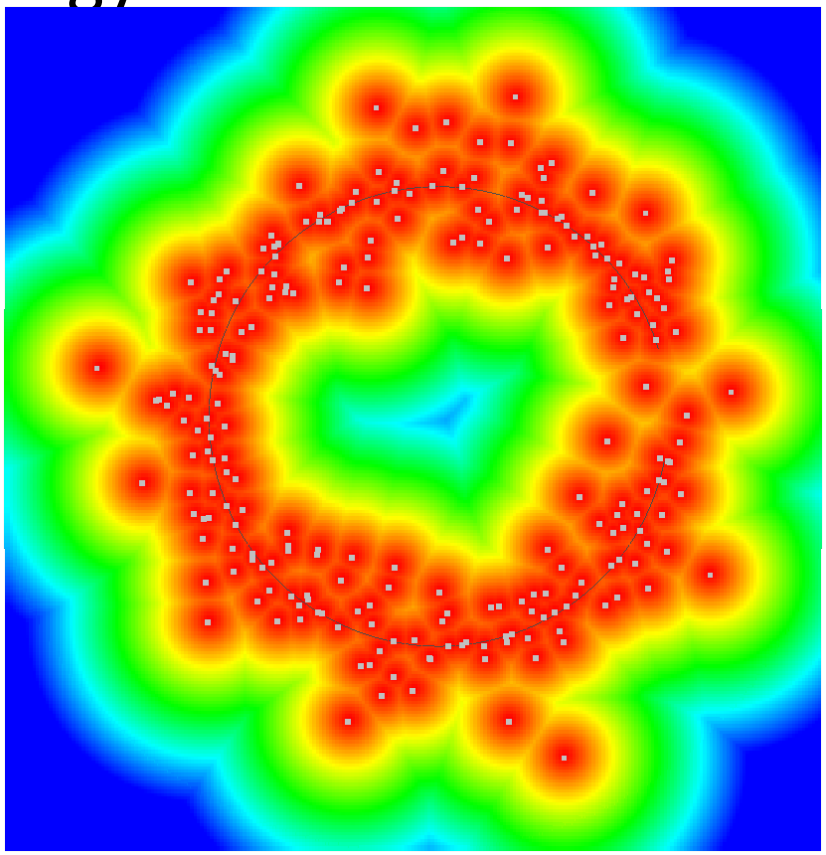
- Topology of the *sublevel sets* of d_{m_0,μ_W}

In black: C_{shape}
Grey dots: W
238 point samples from
$\mathcal{U}(C_{shape}) * \mathcal{N}(0, 0.20)$

Which values of m_0 guarantee the correct *homotopy type*?
A rigorous answer requires Wasserstein distance

Distances and level sets

- Topology of the *sublevel sets* of d_{m_0,μ_W}



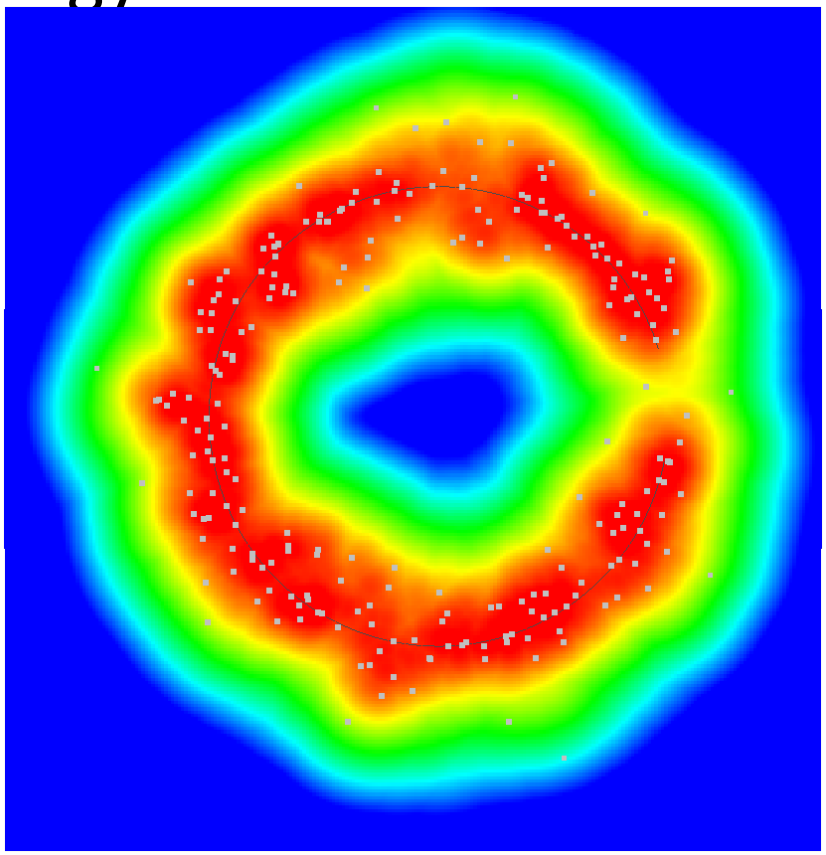
In black: C_{shape}
Grey dots: W
238 point samples from
$\mathcal{U}(C_{shape}) * \mathcal{N}(0, 0.20)$

In color:
$d_{m_0,\mu_W}(x),\ m_0 = 0.004$
$(i.e.\ k_0 = 1)$

Which values of m_0 guarantee the correct *homotopy type*?
A rigorous answer requires Wasserstein distance

Distances and level sets

- Topology of the *sublevel sets* of d_{m_0,μ_W}



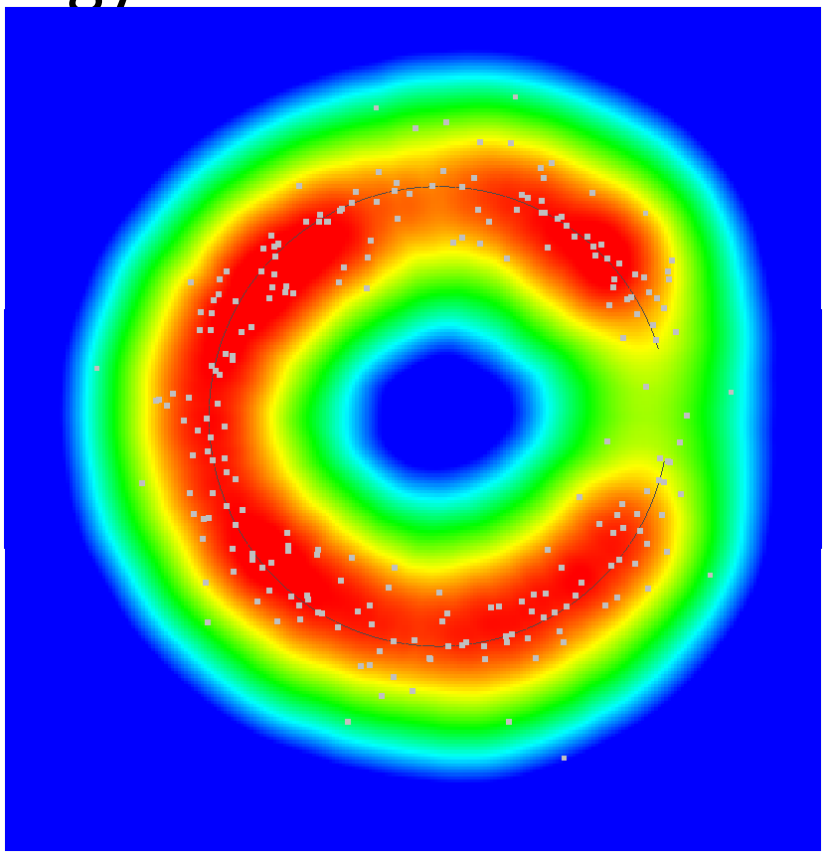
In black: C_{shape}
Grey dots: W
238 point samples from
$\mathcal{U}(C_{shape}) * \mathcal{N}(0, 0.20)$

In color:
$d_{m_0,\mu_W}(x),\ m_0 = 0.029$
$(i.e.\ k_0 = 7)$

Which values of m_0 guarantee the correct *homotopy type*?
A rigorous answer requires Wasserstein distance

Distances and level sets

- Topology of the *sublevel sets* of d_{m_0,μ_W}



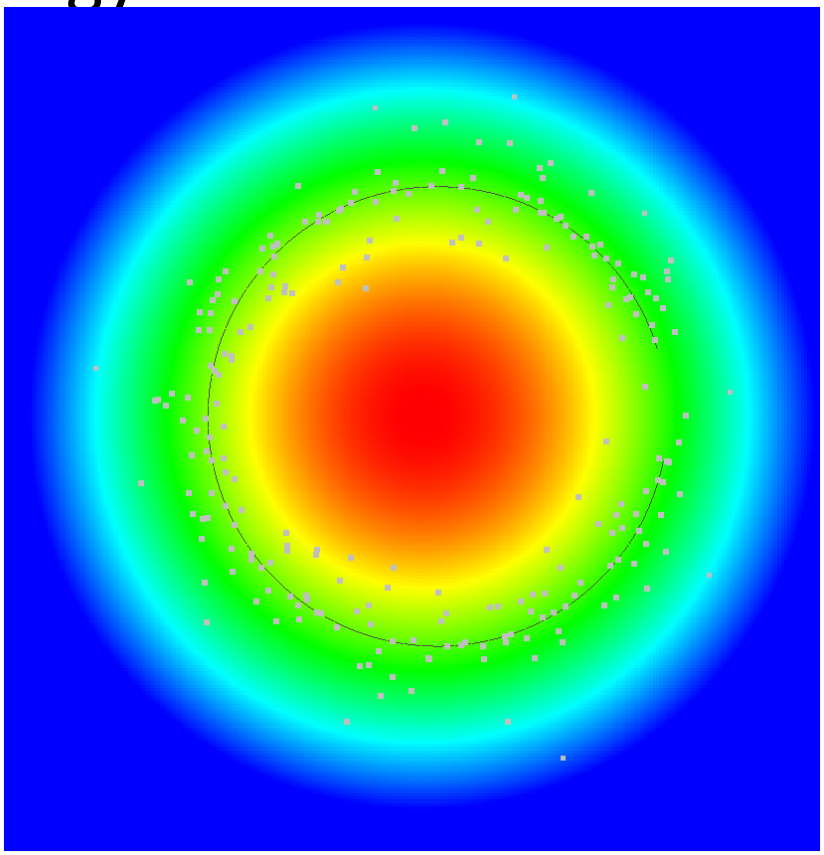
In black: C_{shape}
Grey dots: W
238 point samples from
$\mathcal{U}(C_{shape}) * \mathcal{N}(0, 0.20)$

In color:
$d_{m_0,\mu_W}(x),\ m_0 = 0.159$
$(i.e.\ k_0 = 38)$

Which values of m_0 guarantee the correct *homotopy type*?
A rigorous answer requires Wasserstein distance

Distances and level sets

- Topology of the *sublevel sets* of d_{m_0,μ_W}



In black: C_{shape}
Grey dots: W
238 point samples from
$\mathcal{U}(C_{shape}) * \mathcal{N}(0, 0.20)$

In color:
$d_{m_0,\mu_W}(x),\ m_0 = 1.000$
$(i.e.\ k_0 = 238)$

Which values of m_0 guarantee the correct *homotopy type*?
A rigorous answer requires Wasserstein distance

Distances and level sets

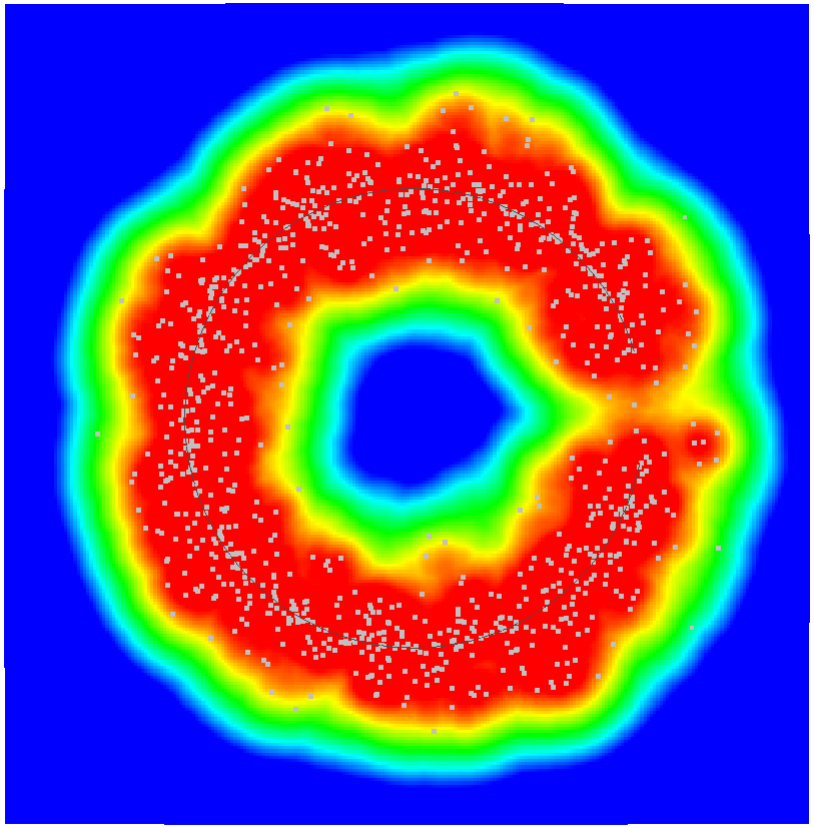
- Topology of the *sublevel sets* of d_{m_0,μ_W}

In black: C_{shape}
Grey dots: W
952 point samples from
$\mathcal{U}(C_{shape}) * \mathcal{N}(0, 0.20)$

Which values of m_0 guarantee the correct *homotopy type*?
A rigorous answer requires Wasserstein distance

Distances and level sets

- Topology of the *sublevel sets* of d_{m_0,μ_W}



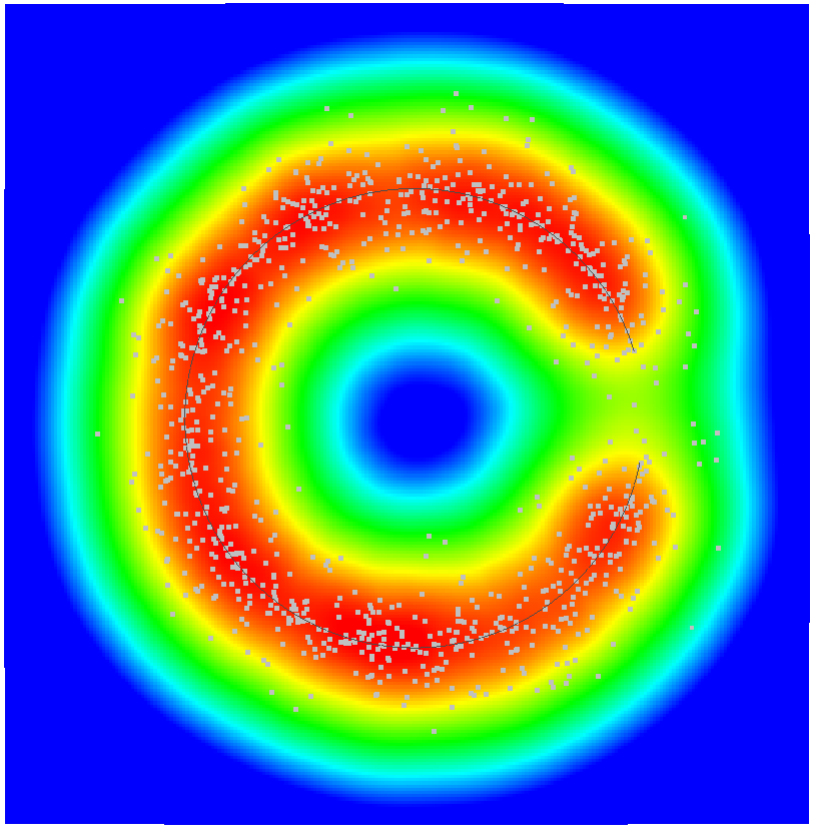
In black: C_{shape}
Grey dots: W
952 point samples from
$\mathcal{U}(C_{shape}) * \mathcal{N}(0, 0.20)$

In color:
$d_{m_0,\mu_W}(x),\ m_0 = 0.007$
$(i.e.\ k_0 = 7)$

Which values of m_0 guarantee the correct *homotopy type*?
A rigorous answer requires Wasserstein distance

Distances and level sets

- Topology of the *sublevel sets* of d_{m_0,μ_W}



In black: C_{shape}
Grey dots: W
952 point samples from
$\mathcal{U}(C_{shape}) * \mathcal{N}(0, 0.20)$

In color:
$d_{m_0,\mu_W}(x),\ m_0 = 0.100$
$(i.e.\ k_0 = 96)$

Which values of m_0 guarantee the correct *homotopy type*?
A rigorous answer requires Wasserstein distance

Wasserstein distance

Distance btw. *two* measures

The intuitive idea: transport

- An example

  There are two *producers* $x_1, x_2 \in \mathbb{R}$
  producing quantities $\alpha_1, \alpha_2 > 0$

  And two *consumers* $y_1, y_2 \in \mathbb{R}$
  consuming quantities $\beta_1, \beta_2 > 0$,
  with $\alpha_1 + \alpha_2 = \beta_1 + \beta_2$

  What is the most effective way to transport quantities from producers to consumers?

Distance btw. *two* measures

The intuitive idea: transport

- An example

  There are two *producers* $x_1, x_2 \in \mathbb{R}$
  producing quantities $\alpha_1, \alpha_2 > 0$

  And two *consumers* $y_1, y_2 \in \mathbb{R}$
  consuming quantities $\beta_1, \beta_2 > 0$,
  with $\alpha_1 + \alpha_2 = \beta_1 + \beta_2 = 1$ ← i.e. two measures

  The total *cost* of transporting quantities around is:

$$c = \sum_{i\ j} |x_i - y_j| \cdot m_{ij} = \sum_{i\ j} d(x_i, y_j) \cdot m_{ij}$$

Transported quantity

Distance btw. *two* measures

The intuitive idea: transport

- An example

  The *most effective* transport plan is:

$$w = \min \left(\sum_{i\ j} d(x_i, y_j) \cdot m_{ij} \right)$$

  under the constraints:

$$\sum_j m_{ij} = \alpha_i, \ \sum_i m_{ij} = \beta_j, \ m_{ij} \geq 0$$

This generalizes to any two measures defined on two finite point clouds

Wasserstein distance

A.k.a. *Earth Mover's Distance*, EMD

- Given *two* measures μ, ν
  a *transport plan* is a joint measure $\pi_{\mu,\nu}$
  such that for every (Borel) subset $A \subseteq \mathbb{R}^d$
  $$\pi_{\mu,\nu}(A, \mathbb{R}^d) = \mu(A), \ \pi_{\mu,\nu}(\mathbb{R}^d, A) = \nu(A)$$

- The *cost* of a transport plan is
  $$\mathcal{C}(\pi_{\mu,\nu}) := \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \mathrm{d}\pi_{\mu,\nu}(x, y) \right)^{\frac{1}{2}}$$

Wasserstein distance

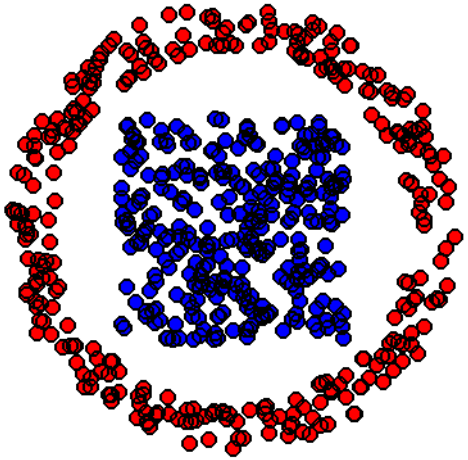
- The *Wasserstein distance* of two measures μ, ν is

$$\mathcal{W}_2(\mu, \nu) := \inf_{\pi_{\mu,\nu}} \mathcal{C}(\pi_{\mu,\nu})$$

(Fear not: we won't compute it)

Wasserstein distance

- A remarkable example: two point clouds of the same size and uniformly distributed mass



A *transport plan* (in this case) is a bijection between the blue dots and the red dots:

$$\pi_{B,R} : B \to R$$

The cost of a transport plan is:

$$\mathcal{C}(\pi_{B,R}) := \left(\sum_{i=1}^{N} \|b_i - \pi_{B,R}(b_i)\|^2 \right)^{\frac{1}{2}}$$

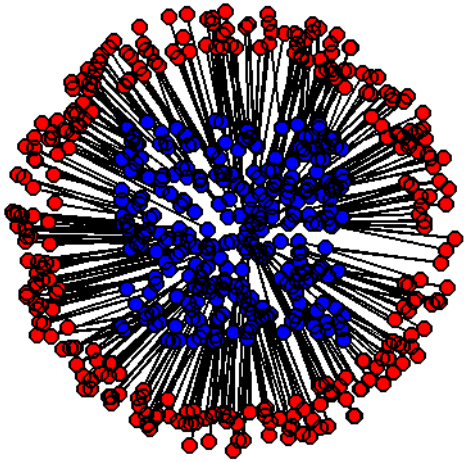
In this case $\mathcal{W}_2$ is:

$$\mathcal{W}_2(B,R) := \min_{\pi_{B,R}} \mathcal{C}(\pi_{B,R})$$

This image comes from:
http://www.numerical-tours.com/matlab/optimaltransp_4_matching_sliced/

Wasserstein distance

- A remarkable example: two point clouds of the same size and uniform mass



A *transport plan* (in this case) is a bijection between the blue dots and the red dots:

$$\pi_{B,R} : B \to R$$

The cost of a transport plan is:

$$\mathcal{C}(\pi_{B,R}) := \left(\sum_{i=1}^{N} \|b_i - \pi_{B,R}(b_i)\|^2 \right)^{\frac{1}{2}}$$

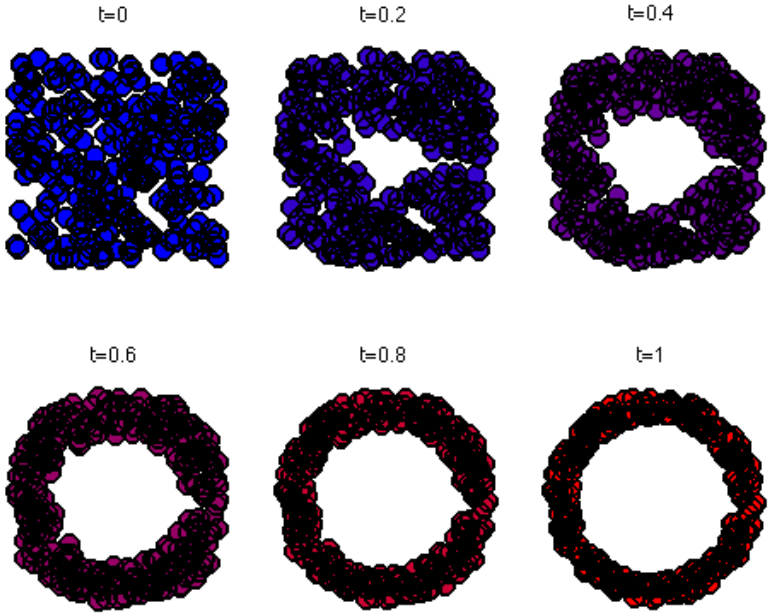
In this case $\mathcal{W}_2$ is:

$$\mathcal{W}_2(B,R) := \min_{\pi_{B,R}} \mathcal{C}(\pi_{B,R})$$

This image comes from:
http://www.numerical-tours.com/matlab/optimaltransp_4_matching_sliced/

Wasserstein distance

- A remarkable example: two point clouds of the same size and uniform mass



An iterative MATLAB algorithm finds $\mathcal{W}_2$

This image comes from:
http://www.numerical-tours.com/matlab/optimaltransp_4_matching_sliced/

Wasserstein distance

- A remarkable example: two gaussians

$$\mathcal{W}_2(\mathcal{N}_{m_1,\Sigma_1},\mathcal{N}_{m_2,\Sigma_2})^2 =$$

$$\|m_1 - m_2\|^2 + \mathrm{Tr}\left(\Sigma_1 + \Sigma_2 - 2\left(\Sigma_1^{\frac{1}{2}}\Sigma_2\Sigma_1^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)$$

Special case:

$$\mathcal{W}_2(\mathcal{N}_{m,0},\mathcal{N}_{m,\Sigma})^2 = \mathrm{Tr}\,(\Sigma)$$

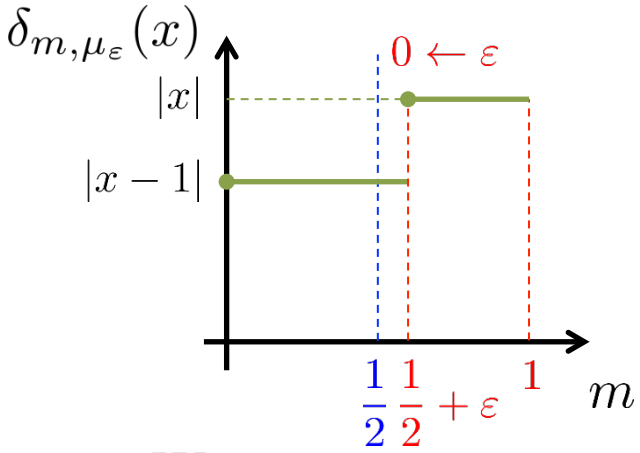
See also:
http://djalil.chafai.net/blog/2010/04/30/wasserstein-distance-between-two-gaussians/

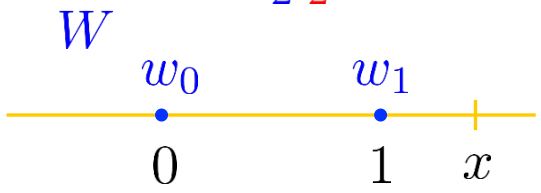
Distance function stability

- The map $\mu_W \mapsto \delta_{m,\mu_W}$ is not *continuous* with respect to $\mathcal{W}_2$

Same example as before, considering that:

$$\mathcal{W}_2(\mu_\varepsilon, \mu_0) = \varepsilon$$





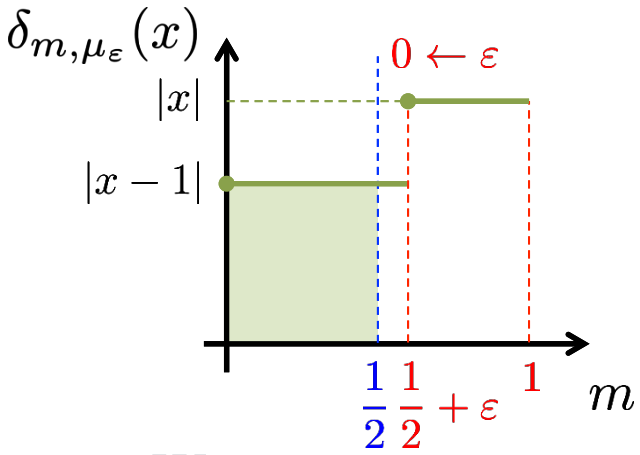
$$\mu_\varepsilon(w_0) := \frac{1}{2} - \varepsilon \qquad \mu_\varepsilon(w_0) := \frac{1}{2} + \varepsilon$$

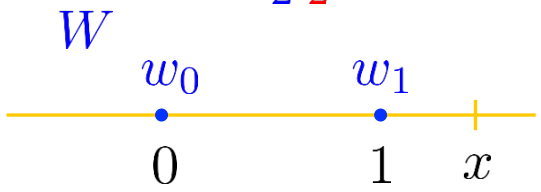
Distance function stability

- The map $\mu_W \mapsto \delta_{m,\mu_W}$ is not *continuous* with respect to $\mathcal{W}_2$
- The map $\mu_W \mapsto d_{m,\mu_W}$ <u>is</u> *continuous* with respect to $\mathcal{W}_2$

Same example as before, considering that:

$$\mathcal{W}_2(\mu_\varepsilon, \mu_0) = \varepsilon$$





$$\mu_\varepsilon(w_0) := \frac{1}{2} - \varepsilon \qquad \mu_\varepsilon(w_0) := \frac{1}{2} + \varepsilon$$

Distance Function Stability

- If μ and μ' are two probability measures on $\mathbb{R}^d$ and $m > 0$ then

$$\|d_{m,\mu} - d_{m,\mu'}\|_\infty \leq \frac{1}{\sqrt{m}} \mathcal{W}_2(\mu, \mu')$$

(Theorem 3.5)

Relevance: the reconstruction theorem requires the two distance-like functions to be close enough
i.e. $\|\varphi - \psi\|_\infty \leq \varepsilon$.

This theorem says that this can be guaranteed for $d_{m,\mu}, d_{m,\mu'}$ when $\mathcal{W}_2(\mu, \mu')$ is small

Reconstruction revisited (I)

- Let μ, μ' be probability measures. If it exists a $\eta > 0$ such that

$$\varepsilon := \frac{W_2(\mu, \mu')}{\sqrt{m}} \leq \frac{r_\eta(d_{m,\mu})}{5 + 4\eta^2}$$

then, for all $\lambda \in (0, r_\eta(d_{m,\mu}))$
and all $r \in [4\varepsilon/\eta^2, r_\eta(d_{m,\mu}) - 3\varepsilon]$

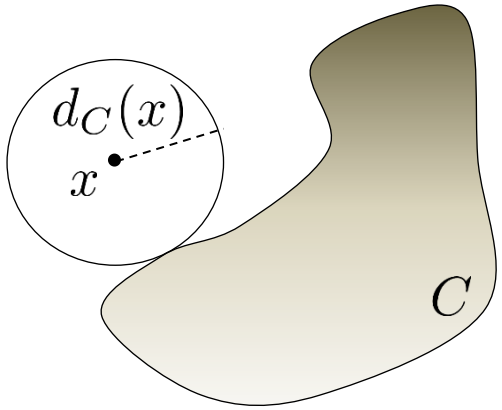
the sublevel sets $d_{m,\mu}^{-1}([0,\lambda])$ $d_{m,\mu'}^{-1}([0,r])$
have the same homotopy type

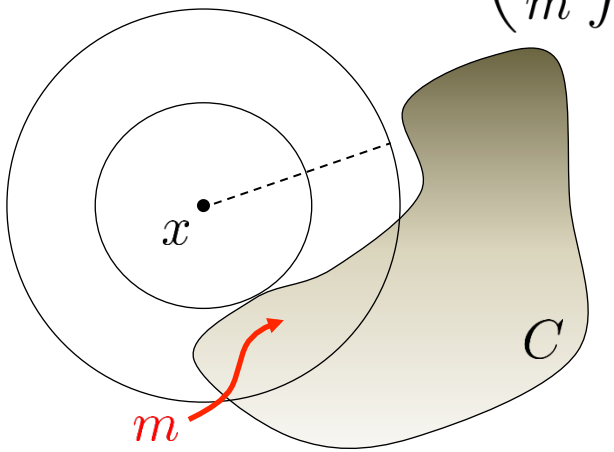
Distance to support

Distance and distance-like

- Consider a compact C and the distance d_C
- Consider also a probability measure μ_C having C as its support
- How does d_C relate to d_{m,μ_C}?

$$d_{m,\mu_C}(x) = \left(\frac{1}{m} \int_0^m \delta^2_{u,\mu_C}(x)\, \mathrm{d}u \right)^{\frac{1}{2}}$$

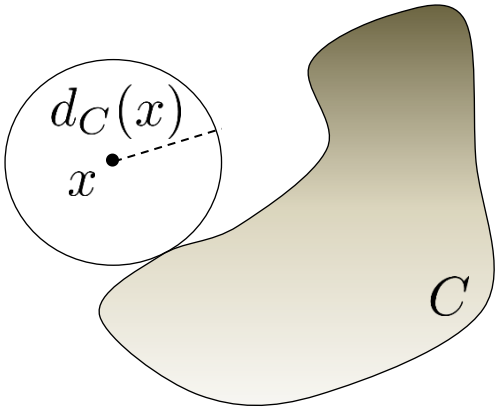


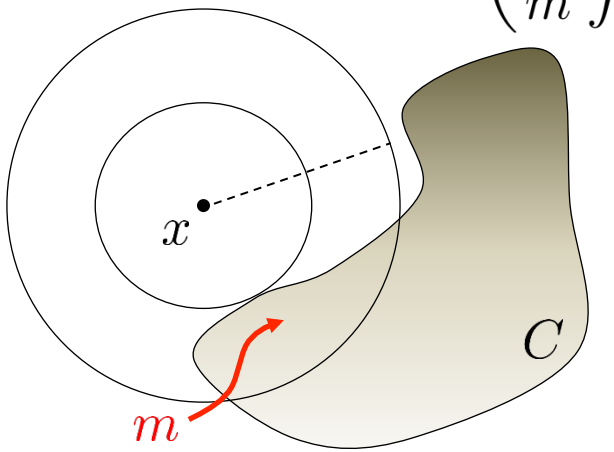


Distance and distance-like

- When the support of a probability measure μ is a compact C, then d_C is the uniform limit of $d_{m,\mu}$ as m converges to 0

$$d_{m,\mu_C}(x) = \left(\frac{1}{m} \int_0^m \delta^2_{u,\mu_C}(x)\, \mathrm{d}u \right)^{\frac{1}{2}}$$





Dimension at most k

- A probability measure μ_C with support C *has dimension at most* k iff k is the smallest integer for which $\exists c_\mu > 0$ such that $\forall p \in C$ there exists an $r' \in (0, 1)$ such that

$$\mu_C(\overline{B}_r(p)) \geq c_\mu r^k, \ r \in (0, r')$$

Ok, what does it mean?

Dimension at most k

- A probability measure μ_C with support C *has dimension at most* k iff k is the smallest integer for which $\exists c_\mu > 0$ such that $\forall p \in C$ there exists an $r' \in (0,1)$ such that

$$\mu_C(\overline{B}_r(p)) \geq c_\mu r^k, \ r \in (0, r')$$

Why the *smallest* k ?

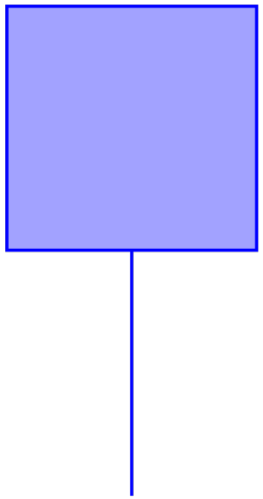
Assume $\mu_C(\overline{B}_r(p)) = c_1 r^2$

then $c_1 r^2 \geq c_1 r^3$ since $1 \geq r$

but $c_1 r^2 \not\geq c_2 r$ since $r \not\geq \frac{c_2}{c_1}$

Dimension at most k

$$\mu_C(\overline{B}_r(p)) \geq c_\mu r^k, \ r \in (0, r')$$



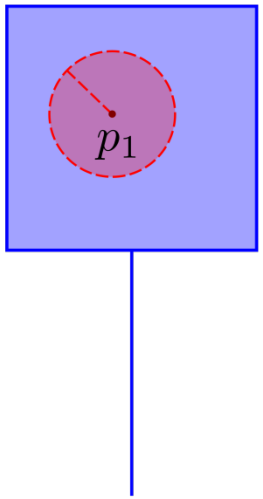
$A_{square} = 1$ $L_{segment} = 1$

$a := \mu_C(square)$ $b := \mu_C(segment)$

$a + b = 1$

Dimension at most k

$$\mu_C(\overline{B}_r(p)) \geq c_\mu r^k, \ r \in (0, r')$$



$A_{square} = 1$ $L_{segment} = 1$

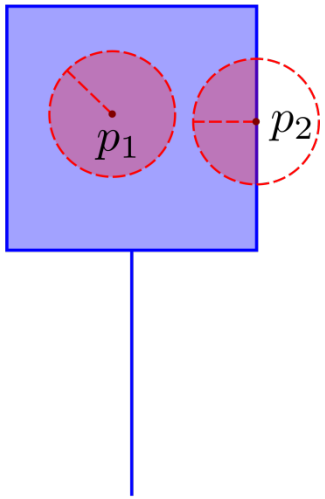
$a := \mu_C(square)$ $b := \mu_C(segment)$

$a + b = 1$

$\mu_C(\overline{B}_r(p_1)) = \pi a r^2$

Dimension at most k

$$\mu_C(\overline{B}_r(p)) \geq c_\mu r^k, \ r \in (0, r')$$



$A_{square} = 1$ $L_{segment} = 1$

$a := \mu_C(square)$ $b := \mu_C(segment)$

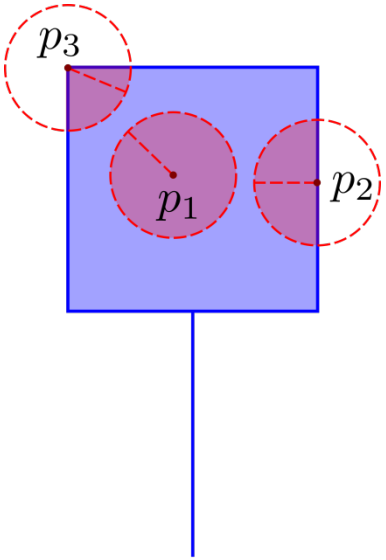
$a + b = 1$

$\mu_C(\overline{B}_r(p_1)) = \pi a r^2$

$\mu_C(\overline{B}_r(p_2)) = \frac{\pi}{2} a r^2$

Dimension at most k

$$\mu_C(\overline{B}_r(p)) \geq c_\mu r^k, \ r \in (0, r')$$



$A_{square} = 1$ $L_{segment} = 1$

$a := \mu_C(square)$ $b := \mu_C(segment)$

$a + b = 1$

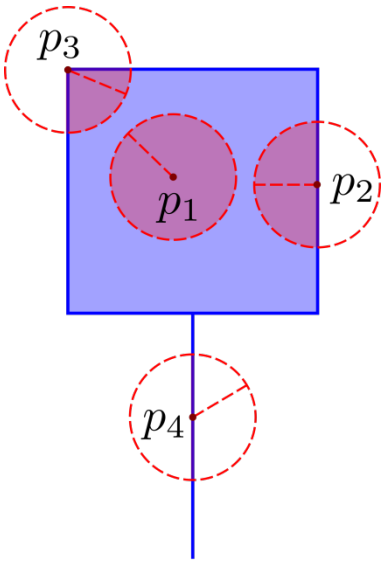
$$\mu_C(\overline{B}_r(p_1)) = \pi a r^2$$

$$\mu_C(\overline{B}_r(p_2)) = \frac{\pi}{2} a r^2$$

$$\mu_C(\overline{B}_r(p_3)) = \frac{\pi}{4} a r^2$$

Dimension at most k

$$\mu_C(\overline{B}_r(p)) \geq c_\mu r^k, \ r \in (0, r')$$



$A_{square} = 1$ $L_{segment} = 1$

$a := \mu_C(square)$ $b := \mu_C(segment)$

$a + b = 1$

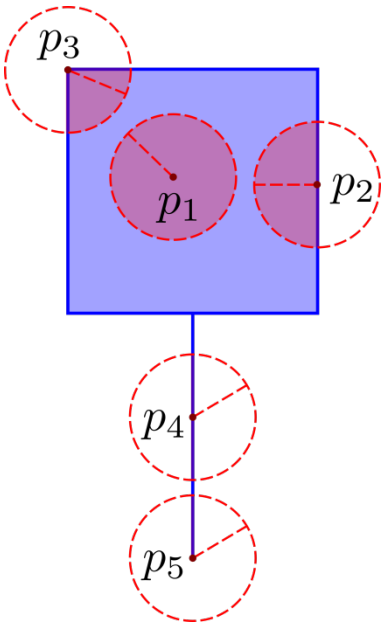
$\mu_C(\overline{B}_r(p_1)) = \pi a r^2$ $\mu_C(\overline{B}_r(p_4)) = 2br$

$\mu_C(\overline{B}_r(p_2)) = \frac{\pi}{2} a r^2$

$\mu_C(\overline{B}_r(p_3)) = \frac{\pi}{4} a r^2$

Dimension at most k

$$\mu_C(\overline{B}_r(p)) \geq c_\mu r^k, \ r \in (0, r')$$



$A_{square} = 1$ $L_{segment} = 1$

$a := \mu_C(square)$ $b := \mu_C(segment)$

$a + b = 1$

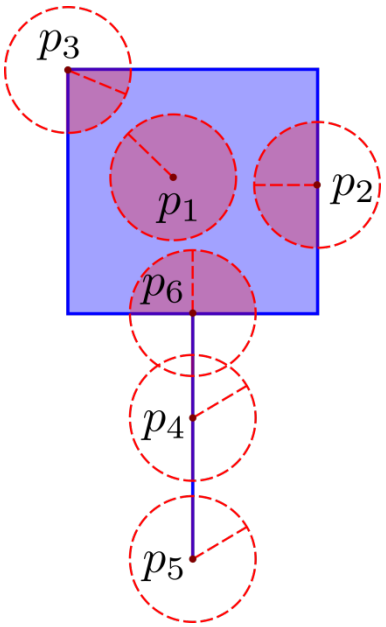
$\mu_C(\overline{B}_r(p_1)) = \pi a r^2$ $\mu_C(\overline{B}_r(p_4)) = 2br$

$\mu_C(\overline{B}_r(p_2)) = \frac{\pi}{2} a r^2$ $\mu_C(\overline{B}_r(p_5)) = br$

$\mu_C(\overline{B}_r(p_3)) = \frac{\pi}{4} a r^2$

Dimension at most k

$$\mu_C(\overline{B}_r(p)) \geq c_\mu r^k, \ r \in (0, r')$$



$A_{square} = 1$ $L_{segment} = 1$

$a := \mu_C(square)$ $b := \mu_C(segment)$

$a + b = 1$

$\mu_C(\overline{B}_r(p_1)) = \pi a r^2$ $\mu_C(\overline{B}_r(p_4)) = 2br$

$\mu_C(\overline{B}_r(p_2)) = \frac{\pi}{2} a r^2$ $\mu_C(\overline{B}_r(p_5)) = br$

$\mu_C(\overline{B}_r(p_3)) = \frac{\pi}{4} a r^2$ $\mu_C(\overline{B}_r(p_6)) = br + \frac{\pi}{2} a r^2$

The *squared lollipop* has dimension at most 2, with $c_\mu = \frac{\pi}{4} a$

Distance to measure's support

- Let μ_C be a probability measure with support C having dimension at most k , with constant c_μ
  Then

$$\|d_C - d_{m,\mu_C}\|_\infty \leq \left(\frac{m}{c_\mu}\right)^{\frac{1}{k}}$$

Putting everything together

Compacts & Measures

Assume we have:

- A compact set C to be reconstructed
- A measure μ_C having C as its support
- A *noisy* measure μ_C^N
  (e.g. $\mu_C^N := \mu_C * \mu^N$, where μ^N models noise)
- An *empirical*, *noisy* measure μ_C^{NW} obtained from a *finite* point sample W of μ_C^N

Compacts & Measures

Assume we have (examples):

- A circle to be reconstructed
- Source of coherent light from one side (i.e μ_C *not* uniform on C)
- A *noisy* measure μ_C^N (e.g. $\mu_C^N := \mu_C * \mu^N$, where μ^N models noise)
- An *empirical*, *noisy* measure μ_C^{NW} obtained from a *finite* point sample W of μ_C^N

Reconstruction (final)

Then:

$$\|d_C - d_{m,\mu_C^N W}\|_\infty \leq$$

$$\|d_C - d_{m,\mu_C}\|_\infty +$$

$$\|d_{m,\mu_C} - d_{m,\mu_C^N}\|_\infty + \|d_{m,\mu_C^N} - d_{m,\mu_C^N W}\|_\infty$$

Reconstruction (final)

Then:

$$\|d_C - d_{m,\mu_C^{NW}}\|_\infty \leq$$

$$\left(\frac{m}{c_\mu}\right)^{\frac{1}{k}} +$$

$$\frac{W_2(\mu_C, \mu_C^N) + W_2(\mu_C^N, \mu_C^{NW})}{\sqrt{m}}$$

Reconstruction (final)

- Let μ have dimension at most k and support C such that for some $\eta > 0$, $r_\eta(C) > 0$.
  Let μ' be another measure such that
  $\|d_C - d_{m,\mu'}\|_\infty \leq$

$$\varepsilon := \left(\frac{m}{c_\mu}\right)^{\frac{1}{k}} + \frac{W_2(\mu,\mu')}{\sqrt{m}} \leq \frac{r_\eta(C)}{5+4\eta^2}$$

- Then C^r, $0 < r < r_\eta(C)$
  and $d_{m,\mu'}^{-1}([0,\lambda])$, $\lambda \in [4\varepsilon/\eta^2, r_\eta(C) - 3\varepsilon]$
  have the same homotopy type

Reconstruction (final)

- Given that, in a specific problem, the values $c_\mu, W_2(\mu, \mu'), \eta, r_\eta(C)$ are given
- The problem is then finding an m such that:

$$\left(\frac{m}{c_\mu}\right)^{\frac{1}{k}} + \frac{W_2(\mu, \mu')}{\sqrt{m}} \leq \frac{r_\eta(C)}{5 + 4\eta^2}$$

(Unfortunately, there is no known closed-form solution for m to this inequality)