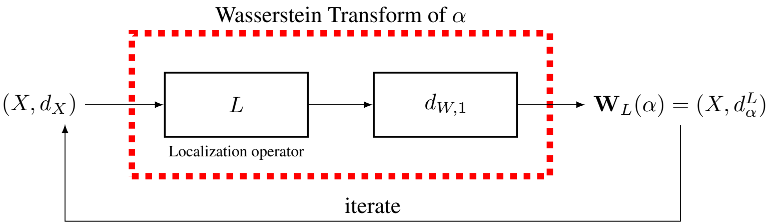
The Wasserstein Transform

Facundo Mémoli*, Zane Smith†, Zhengchao Wan*

*The Ohio State University
†The University of Minnesota
https://research.math.osu.edu/networks/



[1]Supported by NSF AF 1526513, NSF DMS 1723003, NSF DMS 1547357, and NSF CCF 1740761

The Wasserstein Transform

Fix a metric space (X, d_X) and a parameter $\varepsilon > 0$. Given a dataset α, i.e., a probability measure on X with full support, for each point $x \in X$, let

$$m_\alpha^{(\varepsilon)}(x) := \frac{\alpha|_{B_\varepsilon(x)}}{\alpha(B_\varepsilon(x))}.$$

The Wasserstein Transform

Fix a metric space (X, d_X) and a parameter $\varepsilon > 0$. Given a dataset α, i.e., a probability measure on X with full support, for each point $x \in X$, let

$$m_\alpha^{(\varepsilon)}(x) := \frac{\alpha|_{B_\varepsilon(x)}}{\alpha(B_\varepsilon(x))}.$$

The *Wasserstein transform* $\mathbf{W}_\varepsilon$ applied to α gives the distance function $d_\alpha^{(\varepsilon)}$ on X defined by

$$d_\alpha^{(\varepsilon)}(x, x') := d_{W,1}\left(m_\alpha^{(\varepsilon)}(x), m_\alpha^{(\varepsilon)}(x')\right), \forall x, x' \in X.$$

The Wasserstein Transform

Fix a metric space (X, d_X) and a parameter $\varepsilon > 0$. Given a dataset α, i.e., a probability measure on X with full support, for each point $x \in X$, let

$$m_\alpha^{(\varepsilon)}(x) := \frac{\alpha|_{B_\varepsilon(x)}}{\alpha(B_\varepsilon(x))}.$$

The *Wasserstein transform* $\mathbf{W}_\varepsilon$ applied to α gives the distance function $d_\alpha^{(\varepsilon)}$ on X defined by

$$d_\alpha^{(\varepsilon)}(x, x') := d_{W,1}\left(m_\alpha^{(\varepsilon)}(x), m_\alpha^{(\varepsilon)}(x')\right), \forall x, x' \in X.$$

Note:

- as $\varepsilon \to 0$ one has $d_\alpha^{(\varepsilon)}(x, x') \to d_X(x, x')$ and
- as $\varepsilon \to \infty$ one has $d_\alpha^{(\varepsilon)}(x, x') \to 0$

The Wasserstein Transform

Fix a metric space (X, d_X) and a parameter $\varepsilon > 0$. Given a dataset α, i.e., a probability measure on X with full support, for each point $x \in X$, let

$$m_\alpha^{(\varepsilon)}(x) := \frac{\alpha|_{B_\varepsilon(x)}}{\alpha(B_\varepsilon(x))}.$$

The *Wasserstein transform* $\mathbf{W}_\varepsilon$ applied to α gives the distance function $d_\alpha^{(\varepsilon)}$ on X defined by

$$d_\alpha^{(\varepsilon)}(x, x') := d_{W,1}\left(m_\alpha^{(\varepsilon)}(x), m_\alpha^{(\varepsilon)}(x')\right), \forall x, x' \in X.$$

Note:

- as $\varepsilon \to 0$ one has $d_\alpha^{(\varepsilon)}(x, x') \to d_X(x, x')$ and
- as $\varepsilon \to \infty$ one has $d_\alpha^{(\varepsilon)}(x, x') \to 0$

Theorem

$\mathbf{W}_\varepsilon$ *is stable when restricted to certain subsets of* $\mathcal{P}_f(X)$.

Wasserstein Transform: Generalizing Mean Shift

The Wasserstein Transform can be viewed as a generalization of Mean Shift.

Wasserstein Transform: Generalizing Mean Shift

The Wasserstein Transform can be viewed as a generalization of Mean Shift.

Indeed, because when given two probability measures α and β on Euclidean space, $d_{W,1}(\alpha, \beta)$ is bounded below by the distance between their means, we get:

Wasserstein Transform: Generalizing Mean Shift

The Wasserstein Transform can be viewed as a generalization of Mean Shift.

Indeed, because when given two probability measures α and β on Euclidean space, $d_{W,1}(\alpha, \beta)$ is bounded below by the distance between their means, we get:

$$d_{\alpha}^{L^{ms}}(x, x') := ||mean(m_{\alpha}^{(\varepsilon)}(x)) - mean(m_{\alpha}^{(\varepsilon)}(x'))|| \leq d_{\alpha}^{(\varepsilon)}(x, x')$$

Wasserstein Transform: Generalizing Mean Shift

The Wasserstein Transform can be viewed as a generalization of Mean Shift.

Indeed, because when given two probability measures α and β on Euclidean space, $d_{W,1}(\alpha, \beta)$ is bounded below by the distance between their means, we get:

$$d_\alpha^{L^{ms}}(x, x') := ||mean(m_\alpha^{(\varepsilon)}(x)) - mean(m_\alpha^{(\varepsilon)}(x'))|| \leq d_\alpha^{(\varepsilon)}(x, x')$$

This means our stability result for $\mathbf{W}_\varepsilon$ gives stability for Mean Shift, which remained unproven.

Wasserstein Transform: Generalizing Mean Shift

The Wasserstein Transform can be viewed as a generalization of Mean Shift.

Indeed, because when given two probability measures α and β on Euclidean space, $d_{W,1}(\alpha, \beta)$ is bounded below by the distance between their means, we get:

$$d_\alpha^{L^{ms}}(x, x') := ||mean(m_\alpha^{(\varepsilon)}(x)) - mean(m_\alpha^{(\varepsilon)}(x'))|| \leq d_\alpha^{(\varepsilon)}(x, x')$$

This means our stability result for $\mathbf{W}_\varepsilon$ gives stability for Mean Shift, which remained unproven.

Theorem

Mean Shift *is stable when restricted to certain subsets of* $\mathcal{P}_f(X)$*.*

Ameliorating Chaining Effect

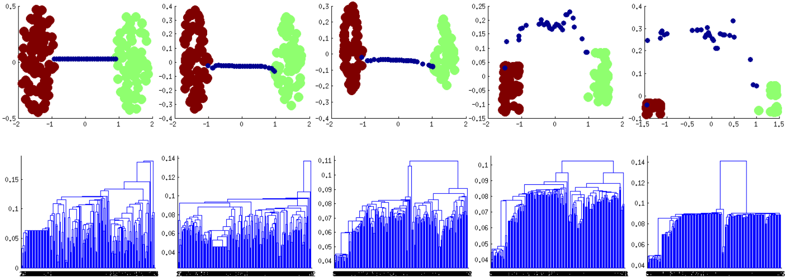


Figure: **WT ameliorates chaining effect.** A dumbbell shape consisting of two disk shaped blobs each with 100 points and separated by a thin chain of 30 points in the plane with Euclidean distance. The diameter of the initial shape was approximately 4. From left to right: 0,1, 2, 3, and 4, iterations of $\mathbf{W}_\varepsilon$ for $\varepsilon = 0.3$.

Denoising of a Circle: WT v.s. Meanshift

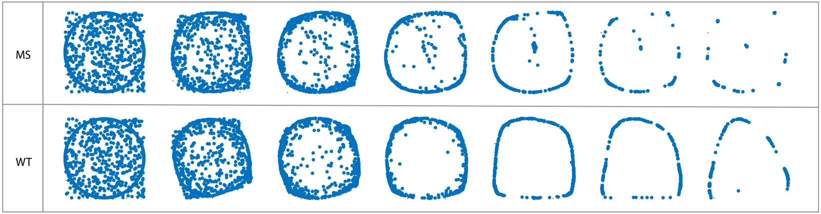


Figure: **Denoising of a circle: several iterations of mean shift vs. $\mathbf{W}_\varepsilon$.** In each case ε was chosen to be 0.3 relative to the diameter at each iteration. This is useful as preprocessing step in **TDA**.

Discussion

- See the paper for performance of WT in classification tasks on MNIST and Grassmannian manifold data.
- Implementation: exploit Sinkhorn/entropic regularization of OT.
- Future work:
  (1) investigate theoretical behaviour of the iterated WT: its connection with Ricci/gradient flows.
  (2) study the experimental performance of versions of the WT based on l_p-Wasserstein distances for $p > 1$ and/or other localization operators.