Linearised Optimal Transport Distances

Kantorovich Initiative

Matthew Thorpe

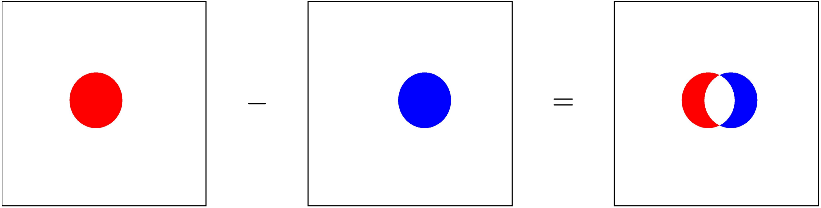
Joint Work with Tianji Cai (University of California Santa Barbara), Junyi Cheng (University of California Santa Barbara), Oliver Crook (University of Oxford), Mihai Cucuringu (University of Oxford), Soheil Kolouri (University of Vanderbilt), Serim Park (Twitter), Gustavo Rohde (University of Virginia), Bernhard Schmitzer (University of Götingen), Caola Schönlieb (University of Cambridge), Dejan Slepčev (Carnegie Mellon University), and Kostas Zygalakis (University of Edinburgh)

Department of Mathematics
University of Manchester

4th May 2023

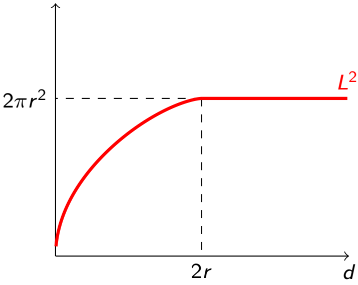
Euclidean Distances

A Euclidean distance considers a pixel-by-pixel difference.



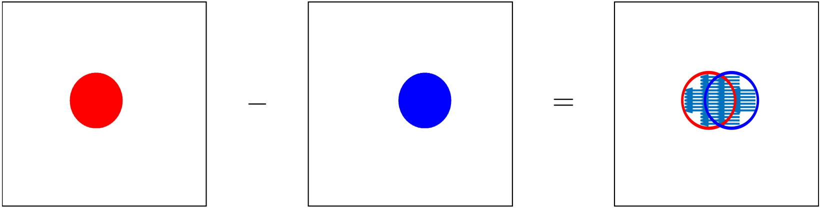
For example, the L^2 distance:

$$d_{L^2}(f,g) = \sqrt{\int_\Omega |f(x) - g(x)|^2 \, \mathrm{d}x}.$$



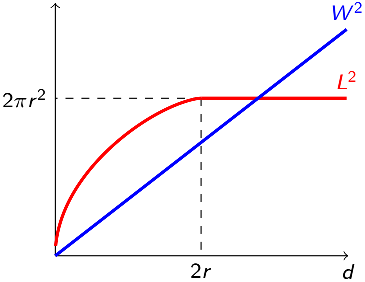
Lagrangian Distances

A Lagrangian distance considers the distance moved.



For example, the Wasserstein distance:

$$d_{W^2}(f, g) \sim \text{size of translation.}$$



Motivation

The Wasserstein distance is *great* as a distance between signals/images, because...

1. Lagrangian modelling,
2. simple to understand compared to other Lagrangian methods such as large deformation diffeomorphic metric mapping,
3. metric properties (in particular symmetry).
4. geodesics and Riemannian structure,
5. theoretical and characterising properties such as existence of optimal transport maps and optimal transport plans (under appropriate conditions).

Motivation

The Wasserstein distance is *great* as a distance between signals/images, because...

1. Lagrangian modelling,
2. simple to understand compared to other Lagrangian methods such as large deformation diffeomorphic metric mapping,
3. metric properties (in particular symmetry).
4. geodesics and Riemannian structure,
5. theoretical and characterising properties such as existence of optimal transport maps and optimal transport plans (under appropriate conditions).

But,...

1. it places restrictive conditions on the input, in particular signals have to be probability measures,
2. computationally expensive (despite recent advances),
3. there is a lack of off-the-shelf data analysis tools.

Motivation

The Wasserstein distance is *great* as a distance between signals/images, because...

1. Lagrangian modelling,
2. simple to understand compared to other Lagrangian methods such as large deformation diffeomorphic metric mapping,
3. metric properties (in particular symmetry).
4. geodesics and Riemannian structure,
5. theoretical and characterising properties such as existence of optimal transport maps and optimal transport plans (under appropriate conditions).

But,...

1. it places restrictive conditions on the input, in particular signals have to be probability measures,
2. computationally expensive (despite recent advances),
3. there is a lack of off-the-shelf data analysis tools.

Solution: **linearise an unbalanced/functional optimal transport distances!**

Balanced Optimal Transport

Let $\mu, \nu \in \mathcal{P}(\Omega)$. The Wasserstein distance can be defined in one of three ways.

Balanced Optimal Transport

Let $\mu, \nu \in \mathcal{P}(\Omega)$. The Wasserstein distance can be defined in one of three ways.

1. Monge formulation:

$$\mathrm{d}^2_{\mathrm{W}^2}(\mu, \nu) := \inf_{T \,:\, T_{\#}\mu = \nu} \int_\Omega |x - T(x)|^2 \, \mathrm{d}\mu(x);$$

Balanced Optimal Transport

Let $\mu, \nu \in \mathcal{P}(\Omega)$. The Wasserstein distance can be defined in one of three ways.

1. Monge formulation:
$$\mathrm{d}^2_{\mathrm{W}^2}(\mu, \nu) := \inf_{T\,:\,T_\#\mu=\nu} \int_\Omega |x - T(x)|^2 \, \mathrm{d}\mu(x);$$

2. Kantorovich formulation:
$$\mathrm{d}^2_{\mathrm{W}^2}(\mu, \nu) := \min_{\pi\in\Pi(\mu,\nu)} \int_{\Omega\times\Omega} |x - y|^2 \, \mathrm{d}\pi(x, y);$$

Balanced Optimal Transport

Let $\mu, \nu \in \mathcal{P}(\Omega)$. The Wasserstein distance can be defined in one of three ways.

1. Monge formulation:
$$\mathrm{d}^2_{\mathrm{W}^2}(\mu,\nu) := \inf_{T\,:\,T_{\#}\mu=\nu} \int_\Omega |x - T(x)|^2 \, \mathrm{d}\mu(x);$$
2. Kantorovich formulation:
$$\mathrm{d}^2_{\mathrm{W}^2}(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)} \int_{\Omega\times\Omega} |x - y|^2 \, \mathrm{d}\pi(x,y);$$
3. Benamou–Brenier formulation:

$$\mathrm{d}^2_{\mathrm{W}^2}(\mu,\nu) := \inf \left\{ \int_0^1 \int_\Omega \left\| \frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t}(x) \right\|^2 \mathrm{d}\rho_t(x)\,\mathrm{d}t \;:\; (\rho,\omega) \in \mathcal{CE}(\mu,\nu) \right\}$$

where
$$(\rho,\omega) \in \mathcal{CE}(\mu,\nu) \Leftrightarrow \frac{\partial\rho}{\partial t} + \nabla_x\omega = 0, \rho_0 = \mu, \rho_1 = \nu.$$

Balanced Optimal Transport

Let $\mu, \nu \in \mathcal{P}(\Omega)$. The Wasserstein distance can be defined in one of three ways.

1. Monge formulation:

$$\mathrm{d}^2_{\mathrm{W}^2}(\mu,\nu) := \inf_{T\,:\,T_{\#}\mu=\nu} \int_\Omega |x - T(x)|^2 \,\mathrm{d}\mu(x);$$

2. Kantorovich formulation:

$$\mathrm{d}^2_{\mathrm{W}^2}(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)} \int_{\Omega\times\Omega} |x - y|^2 \,\mathrm{d}\pi(x,y);$$

3. Benamou–Brenier formulation:

$$\mathrm{d}^2_{\mathrm{W}^2}(\mu,\nu) := \inf\left\{ \int_0^1 \int_\Omega \left\| \frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t}(x) \right\|^2 \mathrm{d}\rho_t(x)\,\mathrm{d}t \;:\; (\rho,\omega) \in \mathcal{CE}(\mu,\nu) \right\}$$

where

$$(\rho,\omega) \in \mathcal{CE}(\mu,\nu) \Leftrightarrow \frac{\partial\rho}{\partial t} + \nabla_x\omega = 0, \rho_0 = \mu, \rho_1 = \nu.$$

Under appropriate conditions all three are equivalent.

1. Let $v_t = \frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t}$, then

$$\mathrm{d}^2_{\mathrm{W}^2}(\mu, \nu) = \int_0^1 \int_\Omega \|v_t(x)\|^2 \, \mathrm{d}\rho_t(x) \, \mathrm{d}t.$$

1. Let $v_t = \frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t}$, then

$$\mathrm{d}^2_{\mathrm{W}^2}(\mu, \nu) = \int_0^1 \int_\Omega \|v_t(x)\|^2 \,\mathrm{d}\rho_t(x)\,\mathrm{d}t.$$

2. If $T_t^* = tT^* + (1-t)\mathrm{Id}$ is the optimal map then $\mu_t = [T_t^*]_\# \mu$ is the geodesic between μ and ν.

The Riemannian Structure of Wasserstein Spaces

1. Let $v_t = \frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t}$, then
$$\mathrm{d}^2_{\mathrm{W}^2}(\mu,\nu) = \int_0^1 \int_\Omega \|v_t(x)\|^2 \,\mathrm{d}\rho_t(x)\,\mathrm{d}t.$$
2. If $T_t^* = tT^* + (1-t)\mathrm{Id}$ is the optimal map then $\mu_t = [T_t^*]_\# \mu$ is the geodesic between μ and ν.
3. Moreover $v_t \circ T_t^* = T^* - \mathrm{Id}$ and
$$\int_\Omega \|v_t(x)\|^2 \,\mathrm{d}\rho_t(x) = \int_\Omega \|v_0\|^2 \,\mathrm{d}\mu(x)$$
for all $t \in [0,1]$.

The Riemannian Structure of Wasserstein Spaces

1. Let $v_t = \frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t}$, then

$$\mathrm{d}^2_{\mathrm{W}^2}(\mu, \nu) = \int_0^1 \int_\Omega \|v_t(x)\|^2 \,\mathrm{d}\rho_t(x)\,\mathrm{d}t.$$

2. If $T_t^* = tT^* + (1-t)\mathrm{Id}$ is the optimal map then $\mu_t = [T_t^*]_\# \mu$ is the geodesic between μ and ν.
3. Moreover $v_t \circ T_t^* = T^* - \mathrm{Id}$ and

$$\int_\Omega \|v_t(x)\|^2 \,\mathrm{d}\rho_t(x) = \int_\Omega \|v_0\|^2 \,\mathrm{d}\mu(x)$$

   for all $t \in [0, 1]$.
4. Hence $\mathrm{d}^2_{\mathrm{W}^2}(\mu, \nu) = \int_\Omega \|v_0\|^2 \,\mathrm{d}\mu(x)$.

The Riemannian Structure of Wasserstein Spaces

1. Let $v_t = \frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t}$, then
$$\mathrm{d}^2_{\mathrm{W}^2}(\mu, \nu) = \int_0^1 \int_\Omega \|v_t(x)\|^2 \,\mathrm{d}\rho_t(x)\,\mathrm{d}t.$$
2. If $T_t^* = tT^* + (1-t)\mathrm{Id}$ is the optimal map then $\mu_t = [T_t^*]_{\#}\mu$ is the geodesic between μ and ν.
3. Moreover $v_t \circ T_t^* = T^* - \mathrm{Id}$ and
$$\int_\Omega \|v_t(x)\|^2 \,\mathrm{d}\rho_t(x) = \int_\Omega \|v_0\|^2 \,\mathrm{d}\mu(x)$$
for all $t \in [0, 1]$.
4. Hence $\mathrm{d}^2_{\mathrm{W}^2}(\mu, \nu) = \int_\Omega \|v_0\|^2 \,\mathrm{d}\mu(x)$.
5. Let $g_{\mathrm{W}^2}(\mu; u, v) = \int_\Omega u \cdot v \,\mathrm{d}\mu$, then
$$\mathrm{d}^2_{\mathrm{W}^2}(\mu, \nu) = g_{\mathrm{W}^2}(\mu; v_0, v_0).$$

The Linear Wasserstein Distance

1. Let $\mathrm{Log}_{\mathrm{W}^2}(\mu;\nu) = v_0$, so

$$\mathrm{d}_{\mathrm{W}^2}(\mu,\nu) = \|\mathrm{Log}_{\mathrm{W}^2}(\mu;\nu)\|_{\mathrm{L}^2(\mu)}.$$

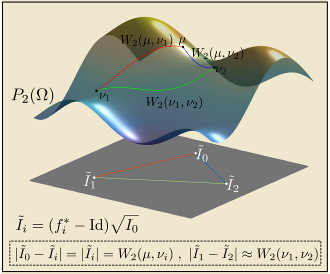


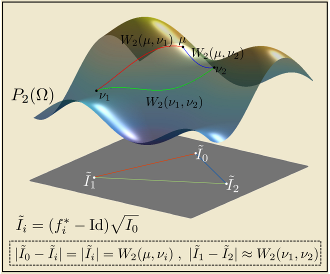
Figure credit: Soheil Kolouri.

The Linear Wasserstein Distance

1. Let $\mathrm{Log}_{\mathrm{W}^2}(\mu;\nu) = v_0$, so

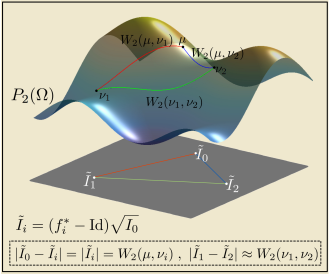
$$\mathrm{d}_{\mathrm{W}^2}(\mu,\nu) = \|\mathrm{Log}_{\mathrm{W}^2}(\mu;\nu)\|_{\mathrm{L}^2(\mu)}.$$

2. Now (following Wang, Slepčev, Basu, Ozolek and Rohde (2013)) we define



$$\mathrm{d}_{\mathrm{W}^2,\mu,\mathrm{lin}}(\mu_1,\mu_2) = \|\mathrm{Log}_{\mathrm{W}^2}(\mu;\mu_1) - \mathrm{Log}_{\mathrm{W}^2}(\mu;\mu_2)\|_{\mathrm{L}^2(\mu)}.$$

Figure credit: Soheil Kolouri.



1. Let $\mathrm{Log}_{\mathrm{W}^2}(\mu;\nu) = v_0$, so

$$\mathrm{d}_{\mathrm{W}^2}(\mu,\nu) = \|\mathrm{Log}_{\mathrm{W}^2}(\mu;\nu)\|_{\mathrm{L}^2(\mu)}.$$

2. Now (following Wang, Slepčev, Basu, Ozolek and Rohde (2013)) we define

$$\mathrm{d}_{\mathrm{W}^2,\mu,\mathrm{lin}}(\mu_1,\mu_2) = \|\mathrm{Log}_{\mathrm{W}^2}(\mu;\mu_1) - \mathrm{Log}_{\mathrm{W}^2}(\mu;\mu_2)\|_{\mathrm{L}^2(\mu)}.$$

3. Linear embedding map:

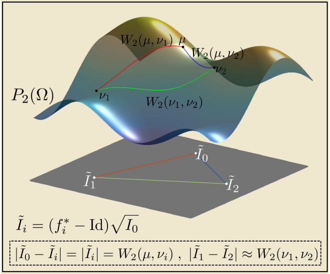
$$P_{\mathrm{W}^2,\mu,\mathrm{lin}}(\mu_i) = \mathrm{Log}_{\mathrm{W}^2}(\mu;\mu_i).$$

Figure credit: Soheil Kolouri.

The Linear Wasserstein Distance

1. Let $\mathrm{Log}_{\mathrm{W}^2}(\mu;\nu) = v_0$, so

$$\mathrm{d}_{\mathrm{W}^2}(\mu,\nu) = \|\mathrm{Log}_{\mathrm{W}^2}(\mu;\nu)\|_{\mathrm{L}^2(\mu)}.$$



2. Now (following Wang, Slepčev, Basu, Ozolek and Rohde (2013)) we define

$$\mathrm{d}_{\mathrm{W}^2,\mu,\mathrm{lin}}(\mu_1,\mu_2) = \|\mathrm{Log}_{\mathrm{W}^2}(\mu;\mu_1) - \mathrm{Log}_{\mathrm{W}^2}(\mu;\mu_2)\|_{\mathrm{L}^2(\mu)}.$$

3. Linear embedding map:

$$P_{\mathrm{W}^2,\mu,\mathrm{lin}}(\mu_i) = \mathrm{Log}_{\mathrm{W}^2}(\mu;\mu_i).$$

4. **Linear Optimal Transport Assumption:**

$$\mathrm{d}_{\mathrm{W}^2}(\mu_1,\mu_2) \approx \mathrm{d}_{\mathrm{W}^2,\mu,\mathrm{lin}}(\mu_1,\mu_2) = \|P_{\mathrm{W}^2,\mu,\mathrm{lin}}(\mu_1) - P_{\mathrm{W}^2,\mu,\mathrm{lin}}(\mu_2)\|_{\mathrm{L}^2(\mu)}.$$

Figure credit: Soheil Kolouri.

Approximate Numerical Method

1. Solve the Kantorovich formulation to find π^* (e.g. Sinkhorns algorithm)

$$\mathrm{d}^2_{\mathrm{W}^2}(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)} \int_{\Omega\times\Omega} |x-y|^2 \,\mathrm{d}\pi(x,y).$$

2. Extract T^* the optimal Monge map from $\pi^* = (\mathrm{Id}\times T^*)_{\#}\mu$

$$\mathrm{d}^2_{\mathrm{W}^2}(\mu,\nu) := \inf_{T\,:\,T_{\#}\mu=\nu} \int_{\Omega} |x-T(x)|^2 \,\mathrm{d}\mu(x).$$

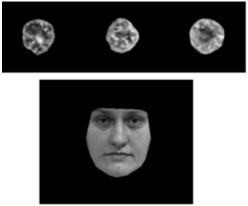
3. Compute the velocity map at time $t=0$, i.e. $v_0 = T^* - \mathrm{Id}$

$$\mathrm{d}^2_{\mathrm{W}^2}(\mu,\nu) = \int_{\Omega} \|v_0\|^2 \,\mathrm{d}\mu(x).$$

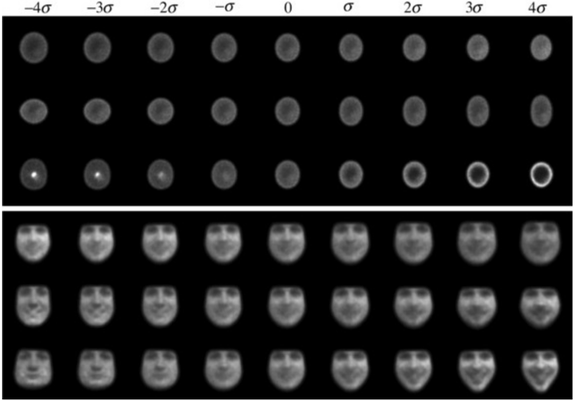
Road map:

$$\nu \quad\mapsto\quad \pi^* \quad\mapsto\quad T^* \quad\mapsto\quad v_0.$$

Example Data:

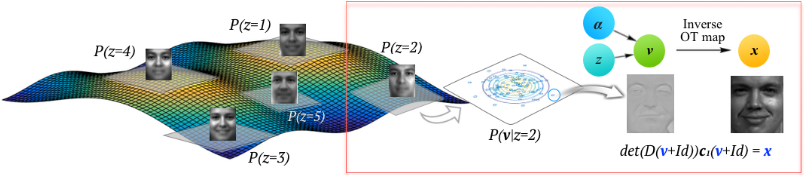


Principle Component Analysis on Linear Embedding:



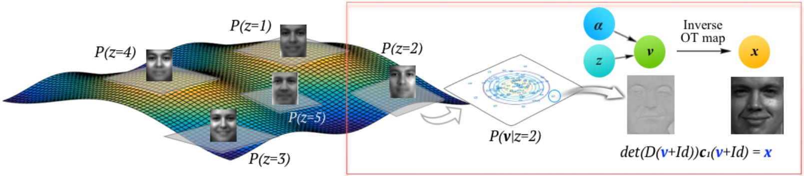
Source: Wang, Slepčev, Basu, Ozolek and Rohde, *A Linear Optimal Transportation Framework for Quantifying and Visualizing Variations in Sets of Images*, International Journal of Computer Vision 101(2):254–269, 2013.

1. **Aim:** Generate new data points from the Wasserstein manifold of images.



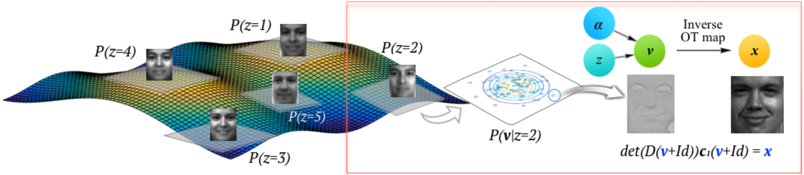
Data Generating

1. **Aim:** Generate new data points from the Wasserstein manifold of images.
2. **Idea:** Approximate the manifold at K-points.



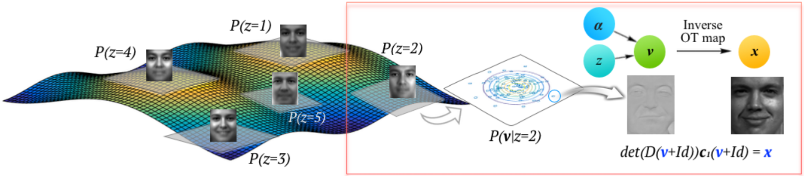
Data Generating

1. **Aim:** Generate new data points from the Wasserstein manifold of images.
2. **Idea:** Approximate the manifold at K-points.
3. **Strategy:**



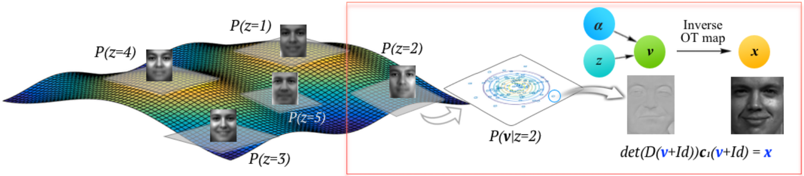
Data Generating

1. **Aim:** Generate new data points from the Wasserstein manifold of images.
2. **Idea:** Approximate the manifold at K-points.
3. **Strategy:**
    1. Cluster the data $\{\mu_i\}_{i=1}^n$ into K groups.



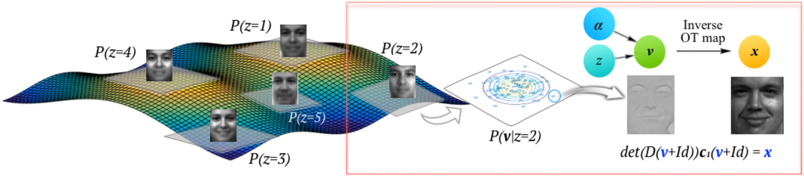
Data Generating

1. **Aim:** Generate new data points from the Wasserstein manifold of images.
2. **Idea:** Approximate the manifold at K-points.
3. **Strategy:**
    1. Cluster the data $\{\mu_i\}_{i=1}^n$ into K groups.
    2. For each cluster find the centre ν_k which will define the K points we approximate the manifold by.



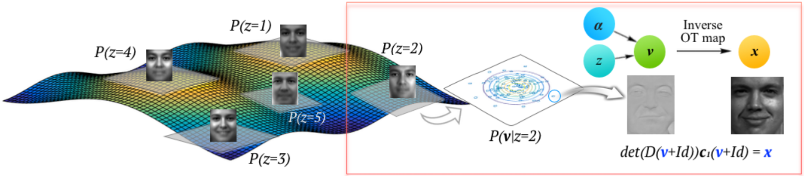
Data Generating

1. **Aim:** Generate new data points from the Wasserstein manifold of images.
2. **Idea:** Approximate the manifold at K-points.
3. **Strategy:**
   1. Cluster the data $\{\mu_i\}_{i=1}^n$ into K groups.
   2. For each cluster find the centre ν_k which will define the K points we approximate the manifold by.
   3. At each of the K centres model the tangent space by a Gaussian with mean m_k and covariance W_k.



Data Generating

1. **Aim:** Generate new data points from the Wasserstein manifold of images.
2. **Idea:** Approximate the manifold at K-points.
3. **Strategy:**
    1. Cluster the data $\{\mu_i\}_{i=1}^n$ into K groups.
    2. For each cluster find the centre ν_k which will define the K points we approximate the manifold by.
    3. At each of the K centres model the tangent space by a Gaussian with mean m_k and covariance W_k.
    4. To generate a new data point (i) sample a cluster centre $k \in \{1, \ldots, K\}$, then (ii) sample a tangent vector $v \sim N(m_k, W_k)$, finally (iii) create a new image by pushing forward the cluster centre ν_k by the transport map $T = v + \mathrm{Id}$.



Are we Learning New Images?

1. Top row, all 19 original images.
2. Second and third rows, generated images.

Source: Park and T., *Representing and Learning High Dimensional Data with the Optimal Transport Map from a Probabilistic Viewpoint*, CVPR, 2018.

Unbalanced Optimal Transport via Benamou–Brenier

1. Recall the continuity equation:

$$(\rho, \omega) \in \mathcal{CE}(\mu, \nu) \Leftrightarrow \frac{\partial \rho}{\partial t} + \nabla_x \omega = 0, \rho_0 = \mu, \rho_1 = \nu.$$

Unbalanced Optimal Transport via Benamou–Brenier

1. Recall the continuity equation:

$$(\rho, \omega) \in \mathcal{CE}(\mu, \nu) \Leftrightarrow \frac{\partial \rho}{\partial t} + \nabla_x \omega = 0, \rho_0 = \mu, \rho_1 = \nu.$$

2. We now consider the continuity equation with source:

$$(\rho, \omega, \zeta) \in \mathcal{CES}(\mu, \nu) \Leftrightarrow \frac{\partial \rho}{\partial t} + \nabla_x \omega = \zeta, \rho_0 = \mu, \rho_1 = \nu.$$

Unbalanced Optimal Transport via Benamou–Brenier

1. Recall the continuity equation:

$$(\rho, \omega) \in \mathcal{CE}(\mu, \nu) \Leftrightarrow \frac{\partial \rho}{\partial t} + \nabla_x \omega = 0, \rho_0 = \mu, \rho_1 = \nu.$$

2. We now consider the continuity equation with source:

$$(\rho, \omega, \zeta) \in \mathcal{CES}(\mu, \nu) \Leftrightarrow \frac{\partial \rho}{\partial t} + \nabla_x \omega = \zeta, \rho_0 = \mu, \rho_1 = \nu.$$

3. The Kondratyev, Monsaingeon and Vorotnikov (2016), Chizat, Peyré, Schmitzer and Vialard (2018, 2018a), and Liero, Mielke and Savaré (2018) model:

$$\int_0^1 \int_\Omega \left(\frac{\mathrm{d}\zeta_t}{\mathrm{d}\rho_t}(x) \right)^2 \mathrm{d}\rho_t(x) \, \mathrm{d}t.$$

Unbalanced Optimal Transport via Benamou–Brenier

1. Recall the continuity equation:

$$(\rho,\omega)\in\mathcal{CE}(\mu,\nu)\Leftrightarrow\frac{\partial\rho}{\partial t}+\nabla_x\omega=0,\rho_0=\mu,\rho_1=\nu.$$

2. We now consider the continuity equation with source:

$$(\rho,\omega,\zeta)\in\mathcal{CES}(\mu,\nu)\Leftrightarrow\frac{\partial\rho}{\partial t}+\nabla_x\omega=\zeta,\rho_0=\mu,\rho_1=\nu.$$

3. The Kondratyev, Monsaingeon and Vorotnikov (2016), Chizat, Peyré, Schmitzer and Vialard (2018, 2018a), and Liero, Mielke and Savaré (2018) model:

$$\int_0^1\int_\Omega\left(\frac{\mathrm{d}\zeta_t}{\mathrm{d}\rho_t}(x)\right)^2\,\mathrm{d}\rho_t(x)\,\mathrm{d}t.$$

4. The Hellinger–Kantorovich distance:

$$\mathrm{d}^2_{\mathrm{HK}}(\mu,\nu):=\inf_{(\rho,\omega,\zeta)\in\mathcal{CES}(\mu,\nu)}\int_0^1\int_\Omega\left(\left\|\frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t}\right\|^2+\frac{1}{4}\left(\frac{\mathrm{d}\zeta_t}{\mathrm{d}\rho_t}\right)^2\right)\mathrm{d}\rho_t\,\mathrm{d}t.$$

Soft Marginal Kantorovich Form

1. Let KL be the Kullback–Leibler divergence

$$\mathrm{KL}(\mu|\nu) = \int \varphi\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\,\mathrm{d}\nu$$

if $\mu \ll \nu$ and where $\varphi(s) = s\log(s) - s + 1$.

Soft Marginal Kantorovich Form

1. Let KL be the Kullback–Leibler divergence

$$\mathrm{KL}(\mu|\nu) = \int \varphi\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\,\mathrm{d}\nu$$

   if $\mu \ll \nu$ and where $\varphi(s) = s\log(s) - s + 1$.

2. Let

$$c(x,y) = \begin{cases} -2\log(\cos\|x-y\|) & \text{if } \|x-y\| < \frac{\pi}{2} \\ +\infty & \text{else.} \end{cases}$$

Soft Marginal Kantorovich Form

1. Let KL be the Kullback–Leibler divergence

$$\mathrm{KL}(\mu|\nu) = \int \varphi\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right) \mathrm{d}\nu$$

if $\mu \ll \nu$ and where $\varphi(s) = s\log(s) - s + 1$.

2. Let

$$c(x,y) = \begin{cases} -2\log(\cos\|x-y\|) & \text{if } \|x-y\| < \frac{\pi}{2} \\ +\infty & \text{else.} \end{cases}$$

3. Then, (Liero, Mielke and Saveré (2018))

$$\mathrm{d}^2_{\mathrm{HK}}(\mu,\nu) = \inf_{\pi\in\mathcal{M}_+(\Omega^2)} \left\{ \int_{\Omega^2} c\,\mathrm{d}\pi + \mathrm{KL}(P_{1\#}\pi|\mu) + \mathrm{KL}(P_{2\#}\pi|\nu) \right\}.$$

1. Let KL be the Kullback–Leibler divergence

$$\mathrm{KL}(\mu|\nu) = \int \varphi\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\,\mathrm{d}\nu$$

   if $\mu \ll \nu$ and where $\varphi(s) = s\log(s) - s + 1$.

2. Let

$$c(x,y) = \begin{cases} -2\log(\cos\|x-y\|) & \text{if } \|x-y\| < \frac{\pi}{2} \\ +\infty & \text{else.} \end{cases}$$

3. Then, (Liero, Mielke and Saveré (2018))

$$\mathrm{d}^2_{\mathrm{HK}}(\mu,\nu) = \inf_{\pi\in\mathcal{M}_+(\Omega^2)} \left\{ \int_{\Omega^2} c\,\mathrm{d}\pi + \mathrm{KL}(P_{1\#}\pi|\mu) + \mathrm{KL}(P_{2\#}\pi|\nu) \right\}.$$

4. Furthermore, there exists π^*, T^* and $\tilde{\mu}$ such that $\pi^* = (\mathrm{Id}\times T^*)_{\#}\tilde{\mu}$ is optimal.

PARENTAL
ADVISORY
EXPLICIT CONTENT

Warning: Long (and uninformative) equations are present on the next slide.

Hellinger–Kantorovich Geodesics via Optimal Plans

Let $\mu, \nu \in \mathcal{M}_+(\Omega)$, π^* optimal and T^* be the Monge map $\pi^* = (\mathrm{Id} \times T^*)_\# \tilde{\mu}$. Let $\tilde{\mu} = P_{1\#}\pi^*$, $\tilde{\nu} = P_{2\#}\pi^*$ and write

$$\mu = u\tilde{\mu} + \mu^\perp \qquad \nu = w\tilde{\nu} + \nu^\perp.$$

Then a geodesic is given by

$$\begin{aligned}
\tilde{\rho}_t &= X\left(t; \cdot, u(\cdot), T^*(\cdot), w \circ T^*(\cdot)\right)_\# \left[M\left(t; \cdot, u(\cdot), T^*(\cdot), w \circ T^*(\cdot)\right) \tilde{\mu}\right] \\
\rho_t &= \tilde{\rho}_t + (1-t)^2\mu^\perp + t^2\nu^\perp \\
\omega_t &= X\left(t; \cdot, u(\cdot), T^*(\cdot), w \circ T^*(\cdot)\right)_\# \left[M\left(t; \cdot, u(\cdot), T^*(\cdot), w \circ T^*(\cdot)\right) \frac{\partial X}{\partial t}\left(t; \cdot, u(\cdot), T^*(\cdot), w \circ T^*(\cdot)\right) \tilde{\mu}\right] \\
\tilde{\zeta}_t &= X\left(t; \cdot, u(\cdot), T^*(\cdot), w \circ T^*(\cdot)\right)_\# \left[\frac{\partial M}{\partial t}\left(t; \cdot, u(\cdot), T^*(\cdot), w \circ T^*(\cdot)\right) \tilde{\mu}\right] \\
\zeta_t &= \tilde{\zeta}_t - 2(1-t)\mu^\perp + 2t\nu^\perp.
\end{aligned}$$

where

$$\begin{aligned}
M(t) &= (1-t)^2 m_0 + t^2 m_1 + 2t(1-t)\sqrt{m_0 m_1}\cos\|x_0 - x_1\| \\
\varphi(t) &= \cos^{-1}\left(\frac{(1-t)\sqrt{m_0} + t\sqrt{m_1}\cos(\|x_0 - x_1\|)}{\sqrt{M(t)}}\right) \\
X(t) &= x_0 + \frac{x_1 - x_0}{\|x_0 - x_1\|}\varphi(t).
\end{aligned}$$

Time Independent Benamou–Brenier Form

Thm: Let $\mu, \nu \in \mathcal{M}_+(\Omega)$ and $\pi^* = (\mathrm{Id} \times T^*)_\# \tilde{\mu}$ be optimal. Let (ρ, ω, ζ) be the geodesics constructed on the previous slide. Set for $t \in [0, 1)$:

$$v_t = \frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t} \qquad\qquad \alpha_t = \frac{\mathrm{d}\tilde{\zeta}_t}{\mathrm{d}\rho_t} - 2(1-t)\frac{\mathrm{d}\mu^\perp}{\mathrm{d}\rho_t}.$$

Time Independent Benamou–Brenier Form

Thm: Let $\mu, \nu \in \mathcal{M}_+(\Omega)$ and $\pi^* = (\mathrm{Id} \times T^*)_\# \tilde{\mu}$ be optimal. Let (ρ, ω, ζ) be the geodesics constructed on the previous slide. Set for $t \in [0, 1)$:

$$v_t = \frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t} \qquad \alpha_t = \frac{\mathrm{d}\tilde{\zeta}_t}{\mathrm{d}\rho_t} - 2(1-t)\frac{\mathrm{d}\mu^\perp}{\mathrm{d}\rho_t}.$$

Then

$$v_0(x) = \begin{cases} \frac{T^*(x)-x}{\|T^*(x)-x\|}\sqrt{\frac{w(T^*(x))}{u(x)}}\sin(\|T^*(x)-x\|) & \tilde{\mu}\text{-a.e.}, \\ 0 & \mu^\perp\text{-a.e.}, \end{cases}$$

$$\alpha_0(x) = \begin{cases} 2\left(\sqrt{\frac{w(T^*(x))}{u(x)}}\cos(\|T^*(x)-x\|) - 1\right) & \tilde{\mu}\text{-a.e.}, \\ -2 & \mu^\perp\text{-a.e.} \end{cases}$$

Time Independent Benamou–Brenier Form

Thm: Let $\mu, \nu \in \mathcal{M}_+(\Omega)$ and $\pi^* = (\mathrm{Id} \times T^*)_\# \tilde{\mu}$ be optimal. Let (ρ, ω, ζ) be the geodesics constructed on the previous slide. Set for $t \in [0, 1)$:

$$v_t = \frac{\mathrm{d}\omega_t}{\mathrm{d}\rho_t} \qquad \alpha_t = \frac{\mathrm{d}\tilde{\zeta}_t}{\mathrm{d}\rho_t} - 2(1-t)\frac{\mathrm{d}\mu^\perp}{\mathrm{d}\rho_t}.$$

Then

$$v_0(x) = \begin{cases} \frac{T^*(x)-x}{\|T^*(x)-x\|}\sqrt{\frac{w(T^*(x))}{u(x)}}\sin(\|T^*(x)-x\|) & \tilde{\mu}\text{-a.e.,} \\ 0 & \mu^\perp\text{-a.e.,} \end{cases}$$

$$\alpha_0(x) = \begin{cases} 2\left(\sqrt{\frac{w(T^*(x))}{u(x)}}\cos(\|T^*(x)-x\|) - 1\right) & \tilde{\mu}\text{-a.e.,} \\ -2 & \mu^\perp\text{-a.e.} \end{cases}$$

and

$$\mathrm{d}^2_{\mathrm{HK}}(\mu, \nu) = \int_\Omega \left(\|v_0\|^2 + \frac{1}{4}(\alpha_0)^2\right) \mathrm{d}\mu + \|\nu^\perp\|.$$

Linear Hellinger–Kantorovich Distance

1. One can show that $\tilde{\mu}, \mu^\perp \perp \nu^\perp$, so $\mu \perp \nu^\perp$.

Linear Hellinger–Kantorovich Distance

1. One can show that $\tilde{\mu}, \mu^\perp \perp \nu^\perp$, so $\mu \perp \nu^\perp$.
2. In particular, if $\mathrm{spt}(\mu) = \Omega$ then $\nu^\perp = 0$, and

$$\mathrm{d}^2_{\mathrm{HK}}(\mu, \nu) = \int_\Omega \left(\|v_0\|^2 + \frac{1}{4}(\alpha_0)^2 \right) \, \mathrm{d}\mu.$$

Linear Hellinger–Kantorovich Distance

1. One can show that $\tilde{\mu}, \mu^\perp \perp \nu^\perp$, so $\mu \perp \nu^\perp$.
2. In particular, if $\mathrm{spt}(\mu) = \Omega$ then $\nu^\perp = 0$, and

$$\mathrm{d}^2_{\mathrm{HK}}(\mu,\nu) = \int_\Omega \left(\|v_0\|^2 + \frac{1}{4}(\alpha_0)^2 \right) \mathrm{d}\mu.$$

3. Let $\mathrm{Log}_{\mathrm{HK}}(\mu;\nu) = (v_0, \alpha_0)$, so

$$\mathrm{d}_{\mathrm{HK}}(\mu,\nu) = \|\mathrm{Log}_{\mathrm{HK}}(\mu;\nu)\|_{\mathrm{L}^2(\mu)}.$$

Linear Hellinger–Kantorovich Distance

1. One can show that $\tilde{\mu}, \mu^\perp \perp \nu^\perp$, so $\mu \perp \nu^\perp$.
2. In particular, if $\mathrm{spt}(\mu) = \Omega$ then $\nu^\perp = 0$, and

$$\mathrm{d}^2_{\mathrm{HK}}(\mu, \nu) = \int_\Omega \left(\|v_0\|^2 + \frac{1}{4}(\alpha_0)^2 \right) \, \mathrm{d}\mu.$$

3. Let $\mathrm{Log}_{\mathrm{HK}}(\mu; \nu) = (v_0, \alpha_0)$, so

$$\mathrm{d}_{\mathrm{HK}}(\mu, \nu) = \|\mathrm{Log}_{\mathrm{HK}}(\mu; \nu)\|_{\mathrm{L}^2(\mu)}.$$

4. Now we define

$$\mathrm{d}_{\mathrm{HK},\mu,\mathrm{lin}}(\mu_1, \mu_2) = \|\mathrm{Log}_{\mathrm{HK}}(\mu; \mu_1) - \mathrm{Log}_{\mathrm{HK}}(\mu; \mu_2)\|_{\mathrm{L}^2(\mu)}.$$

Linear Hellinger–Kantorovich Distance

1. One can show that $\tilde{\mu}, \mu^\perp \perp \nu^\perp$, so $\mu \perp \nu^\perp$.
2. In particular, if $\mathrm{spt}(\mu) = \Omega$ then $\nu^\perp = 0$, and

$$\mathrm{d}^2_{\mathrm{HK}}(\mu, \nu) = \int_\Omega \left(\|v_0\|^2 + \frac{1}{4}(\alpha_0)^2 \right) \mathrm{d}\mu.$$

3. Let $\mathrm{Log}_{\mathrm{HK}}(\mu; \nu) = (v_0, \alpha_0)$, so

$$\mathrm{d}_{\mathrm{HK}}(\mu, \nu) = \|\mathrm{Log}_{\mathrm{HK}}(\mu; \nu)\|_{\mathrm{L}^2(\mu)}.$$

4. Now we define

$$\mathrm{d}_{\mathrm{HK},\mu,\mathrm{lin}}(\mu_1, \mu_2) = \|\mathrm{Log}_{\mathrm{HK}}(\mu; \mu_1) - \mathrm{Log}_{\mathrm{HK}}(\mu; \mu_2)\|_{\mathrm{L}^2(\mu)}.$$

5. Linear embedding map:

$$P_{\mathrm{HK},\mu,\mathrm{lin}}(\mu_i) = \mathrm{Log}_{\mathrm{HK}}(\mu; \mu_i).$$

Linear Hellinger–Kantorovich Distance

1. One can show that $\tilde{\mu}, \mu^{\perp} \perp \nu^{\perp}$, so $\mu \perp \nu^{\perp}$.
2. In particular, if $\mathrm{spt}(\mu) = \Omega$ then $\nu^{\perp} = 0$, and

$$\mathrm{d}_{\mathrm{HK}}^2(\mu, \nu) = \int_{\Omega} \left(\|v_0\|^2 + \frac{1}{4}(\alpha_0)^2 \right) \, \mathrm{d}\mu.$$

3. Let $\mathrm{Log}_{\mathrm{HK}}(\mu; \nu) = (v_0, \alpha_0)$, so

$$\mathrm{d}_{\mathrm{HK}}(\mu, \nu) = \|\mathrm{Log}_{\mathrm{HK}}(\mu; \nu)\|_{\mathrm{L}^2(\mu)}.$$

4. Now we define

$$\mathrm{d}_{\mathrm{HK},\mu,\mathrm{lin}}(\mu_1, \mu_2) = \|\mathrm{Log}_{\mathrm{HK}}(\mu; \mu_1) - \mathrm{Log}_{\mathrm{HK}}(\mu; \mu_2)\|_{\mathrm{L}^2(\mu)}.$$

5. Linear embedding map:

$$P_{\mathrm{HK},\mu,\mathrm{lin}}(\mu_i) = \mathrm{Log}_{\mathrm{HK}}(\mu; \mu_i).$$

6. **Linear Hellinger–Kantorovich Assumption:**

$$\mathrm{d}_{\mathrm{HK}}(\mu_1, \mu_2) \approx \mathrm{d}_{\mathrm{HK},\mu,\mathrm{lin}}(\mu_1, \mu_2) = \|P_{\mathrm{HK},\mu,\mathrm{lin}}(\mu_1) - P_{\mathrm{HK},\mu,\mathrm{lin}}(\mu_2)\|_{\mathrm{L}^2(\mu)}.$$

Approximate Numerical Method

1. Solve the Kantorovich formulation to find π^* (e.g. Sinkhorns algorithm)

$$\mathrm{d}^2_{\mathrm{HK}}(\mu,\nu) = \inf_{\pi \in \mathcal{M}_+(\Omega^2)} \left\{ \int_{\Omega^2} c \,\mathrm{d}\pi + \mathrm{KL}(P_{1\#}\pi|\mu) + \mathrm{KL}(P_{2\#}\pi|\nu) \right\}.$$

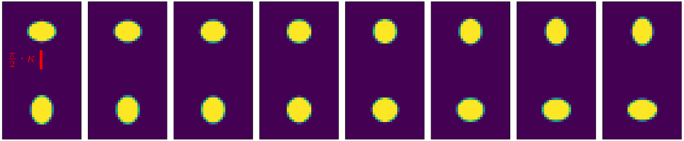
2. Extract T^* the optimal Monge map from $\pi^* = (\mathrm{Id} \times T^*)_{\#}\tilde{\mu}$ and the densities u, w.
3. Compute the velocity and growth maps at time $t = 0$, i.e. v_0, α_0 using the previous theorem

$$\mathrm{d}^2_{\mathrm{HK}}(\mu,\nu) = \int_{\Omega} \left(\|v_0\|^2 + \frac{1}{4}(\alpha_0)^2 \right) \mathrm{d}\mu.$$

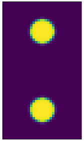
Road map:

$$\nu \quad \mapsto \quad \pi^* \quad \mapsto \quad (T^*, u, w) \quad \mapsto \quad (v_0, \alpha_0).$$

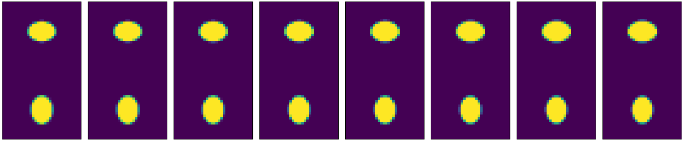
A Toy Example: Data and Barycentres



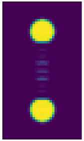
(a) samples for different elongations p_1 (sizes p_2 fixed)



(c) HK barycenter

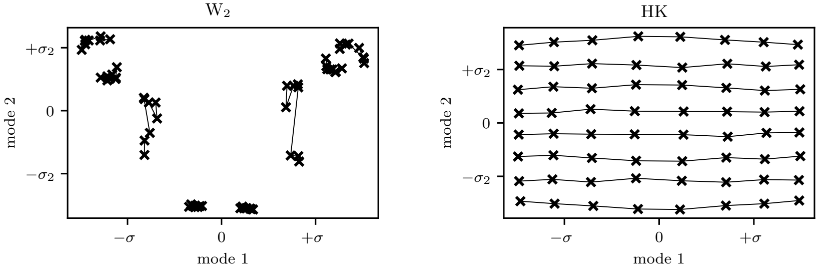


(b) samples for different sizes p_2 (elongations p_1 fixed)

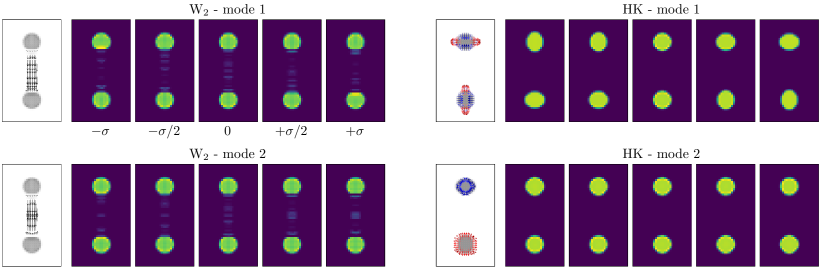


(d) W_2 barycenter

A Toy Example: 2D PCA Projection



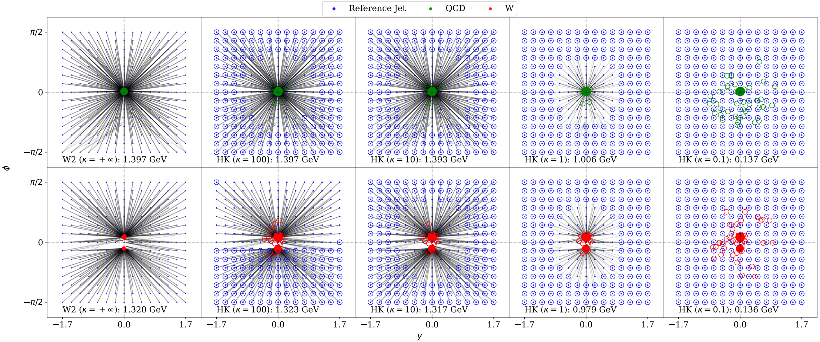
A Toy Example: Dominant Eigenmodes



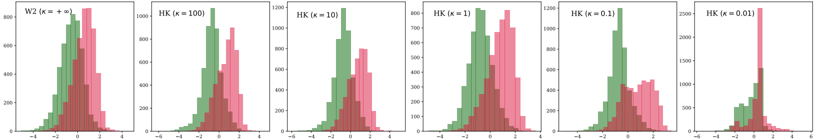
For each mode, the quiver plot on the left shows the initial velocity field v_0, for HK the color of the arrows encodes α_0 (blue means decrease, red increase of mass). The five images on the right visualize the exponential map evaluated between $-\sigma$ and σ where σ denotes the standard deviation along the considered mode.

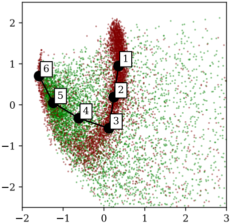
Collider Events: Data

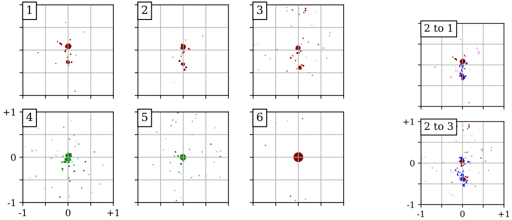
Aim: Jet tagging. In particular, can we label W boson jets and QCD (quark or gluon) jets from a simulated dataset of particle collider events observed in the rapidity-azimuth plan (i.e. $\Omega \subset \mathbb{R}^2$).



Collider Events: LDA and PCA







Collider Events: Labelling

Table: Results for the W vs. QCD jet tagging task using LDA, kNN and SVM on the (unbalanced) linearized OT embeddings for various length scale parameters κ ($\kappa = +\infty$ denotes balanced the Wasserstein distance).

	length scale κ	$+\infty$	100	10	5	1	0.7	0.5	0.3	0.1	0.05	0.01
LDA	**AUC**	0.694	0.733	0.746	0.747	0.752	0.751	0.748	0.760	**0.765**	0.763	0.642
	TPR	0.684	0.684	0.703	0.721	0.724	0.740	0.736	0.692	0.704	0.731	**0.770**
	FPR	0.296	0.218	0.211	0.226	0.220	0.239	0.239	**0.171**	0.174	0.205	0.486
	run time	several seconds										
kNN	**AUC**	0.821	0.818	0.819	0.818	0.829	0.841	**0.849**	0.847	0.821	0.772	0.671
	TPR	0.771	0.763	0.768	0.763	0.760	0.791	0.798	0.809	**0.821**	0.783	0.733
	FPR	0.128	0.127	0.130	0.126	0.102	0.110	**0.100**	0.114	0.181	0.238	0.390
	hyperpar. k	30	20	30	20	10	20	10	20	10	10	30
	run time	1.5 hours										
SVM	**AUC**	0.842	0.842	0.842	0.841	0.849	0.851	**0.856**	0.853	0.845	0.806	0.694
	TPR	0.817	0.819	0.817	0.819	0.823	0.829	**0.832**	0.829	0.788	0.741	0.787
	FPR	0.133	0.134	0.134	0.137	0.126	0.127	0.120	0.124	**0.099**	0.128	0.401
	hyperpar. C	1	1	1	1	1	1	1	1	1	10	10
	hyperpar. γ	100	100	100	100	100	100	100	100	1000	1000	100000
	run time	5 hours										

TL^p Definition

1. Aim: define a *Lagrangian* distance for functions.

TL^p Definition

1. Aim: define a *Lagrangian* distance for functions.
2. The idea is to treat signals as a pair (f, μ) where $f \in \mathrm{L}^p(\mu)$.

TL^p Definition

1. Aim: define a *Lagrangian* distance for functions.
2. The idea is to treat signals as a pair (f, μ) where $f \in \mathrm{L}^p(\mu)$.
3. Mostly we consider when μ is the uniform measure (either continuous or discrete), but one could also trivially adapt in order to weight features of the signal, for example.

TL^p Definition

1. Aim: define a *Lagrangian* distance for functions.
2. The idea is to treat signals as a pair (f, μ) where $f \in \mathrm{L}^p(\mu)$.
3. Mostly we consider when μ is the uniform measure (either continuous or discrete), but one could also trivially adapt in order to weight features of the signal, for example.
4. Note that we can compare signals on different domains.

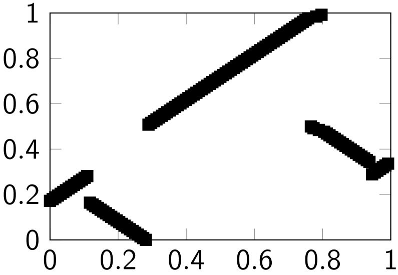
TL^p Definition

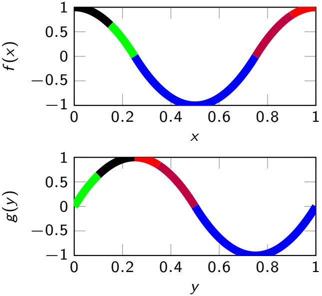
1. Aim: define a *Lagrangian* distance for functions.
2. The idea is to treat signals as a pair (f, μ) where $f \in \mathrm{L}^p(\mu)$.
3. Mostly we consider when μ is the uniform measure (either continuous or discrete), but one could also trivially adapt in order to weight features of the signal, for example.
4. Note that we can compare signals on different domains.
5. TL^p definition (Monge formulation):

$$\mathrm{d}^p_{\mathrm{TL}^p}((f,\mu),(g,\nu)) = \inf_{T:T_{\#}\mu=\nu} \int_X |x - T(x)|^p + |f(x) - g(T(x))|^p \,\mathrm{d}\mu(x).$$

A Simple Example

For example consider the functions $f(x) = \cos(2\pi x)$ and $g(y) = \sin(2\pi y)$ defined on $[0, 1]$ with the uniform measure. The optimal plan using the TL^2 distance is given below.





Relationship Between TL^p and OT: via the Cost Function

1. Optimal transport problems:

Relationship Between TL^p and OT: via the Cost Function

1. Optimal transport problems:

$$\mathcal{M}(\mu, \nu) = \inf_{T : T_{\#}\mu = \nu} \int_X c(x, T(x)) \, \mathrm{d}\mu(x),$$

Relationship Between TL^p and OT: via the Cost Function

1. Optimal transport problems:

$$\mathcal{M}(\mu,\nu) = \inf_{T:T_{\#}\mu=\nu} \int_X c(x, T(x))\,\mathrm{d}\mu(x),$$

$$\mathrm{d}^p_{\mathrm{W}^p}(\mu,\nu) = \inf_{T:T_{\#}\mu=\nu} \int_X |x - T(x)|^p\,\mathrm{d}\mu(x),$$

Relationship Between TL^p and OT: via the Cost Function

1. Optimal transport problems:

$$\mathcal{M}(\mu,\nu) = \inf_{T:T_{\#}\mu=\nu} \int_X c(x, T(x)) \,\mathrm{d}\mu(x),$$

$$\mathrm{d}^p_{\mathrm{W}^p}(\mu,\nu) = \inf_{T:T_{\#}\mu=\nu} \int_X |x - T(x)|^p \,\mathrm{d}\mu(x),$$

$$\mathrm{d}^p_{\mathrm{TL}^p}((f,\mu),(g,\nu)) = \inf_{T:T_{\#}\mu=\nu} \int_X |x - T(x)|^p + |f(x) - g(T(x))|^p \,\mathrm{d}\mu(x).$$

Relationship Between TL^p and OT: via the Cost Function

1. Optimal transport problems:

$$\mathcal{M}(\mu,\nu) = \inf_{T:T_{\#}\mu=\nu} \int_X c(x, T(x))\,\mathrm{d}\mu(x),$$

$$\mathrm{d}^p_{\mathrm{W}^p}(\mu,\nu) = \inf_{T:T_{\#}\mu=\nu} \int_X |x - T(x)|^p\,\mathrm{d}\mu(x),$$

$$\mathrm{d}^p_{\mathrm{TL}^p}((f,\mu),(g,\nu)) = \inf_{T:T_{\#}\mu=\nu} \int_X |x - T(x)|^p + |f(x) - g(T(x))|^p\,\mathrm{d}\mu(x).$$

2. So TL^p is a special case of OT with cost function $c(x,y;f,g) = |x-y|^p + |f(x) - f(y)|^p$.

Relationship Between TL^p and OT: via the Cost Function

1. Optimal transport problems:

$$\mathcal{M}(\mu,\nu) = \inf_{T:T_\#\mu=\nu} \int_X c(x, T(x))\,\mathrm{d}\mu(x),$$

$$\mathrm{d}^p_{\mathrm{W}^p}(\mu,\nu) = \inf_{T:T_\#\mu=\nu} \int_X |x - T(x)|^p\,\mathrm{d}\mu(x),$$

$$\mathrm{d}^p_{\mathrm{TL}^p}((f,\mu),(g,\nu)) = \inf_{T:T_\#\mu=\nu} \int_X |x - T(x)|^p + |f(x) - g(T(x))|^p\,\mathrm{d}\mu(x).$$

2. So TL^p is a special case of OT with cost function $c(x,y;f,g) = |x-y|^p + |f(x) - f(y)|^p$.
3. *This is useful for numerics: Any numerical method for OT that can deal with arbitrary cost function can be used to compute* TL^p*.*

Relationship Between TL^p and OT: via the Cost Function

1. Optimal transport problems:

$$\mathcal{M}(\mu,\nu) = \inf_{T:T_{\#}\mu=\nu} \int_X c(x, T(x))\,\mathrm{d}\mu(x),$$

$$\mathrm{d}^p_{\mathrm{W}^p}(\mu,\nu) = \inf_{T:T_{\#}\mu=\nu} \int_X |x - T(x)|^p\,\mathrm{d}\mu(x),$$

$$\mathrm{d}^p_{\mathrm{TL}^p}((f,\mu),(g,\nu)) = \inf_{T:T_{\#}\mu=\nu} \int_X |x - T(x)|^p + |f(x) - g(T(x))|^p\,\mathrm{d}\mu(x).$$

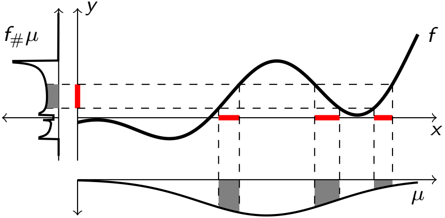
2. So TL^p is a special case of OT with cost function $c(x, y; f, g) = |x - y|^p + |f(x) - f(y)|^p$.
3. *This is useful for numerics: Any numerical method for OT that can deal with arbitrary cost function can be used to compute* TL^p*.*
4. This includes Cuturi's entropy regularised approach (Sinkhorn algorithm).

Relationship Between TL^p and OT: via Graph Projections

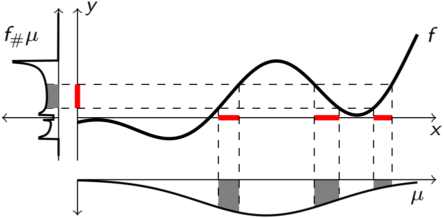
1. The cost function $c(x, y; f, g)$ is not necessarily continuous, therefore the previous relationship with OT is not useful for transferring theoretical properties.

Relationship Between TL^p and OT: via Graph Projections

1. The cost function $c(x, y; f, g)$ is not necessarily continuous, therefore the previous relationship with OT is not useful for transferring theoretical properties.
2. Define $\tilde{\mu} = (\mathrm{Id} \times f)_{\#}\mu$, $\tilde{\nu} = (\mathrm{Id} \times g)_{\#}\nu$ as the measures μ and ν raised onto the graphs of f and g.



1. The cost function $c(x, y; f, g)$ is not necessarily continuous, therefore the previous relationship with OT is not useful for transferring theoretical properties.
2. Define $\tilde{\mu} = (\mathrm{Id} \times f)_{\#}\mu$, $\tilde{\nu} = (\mathrm{Id} \times g)_{\#}\nu$ as the measures μ and ν raised onto the graphs of f and g.

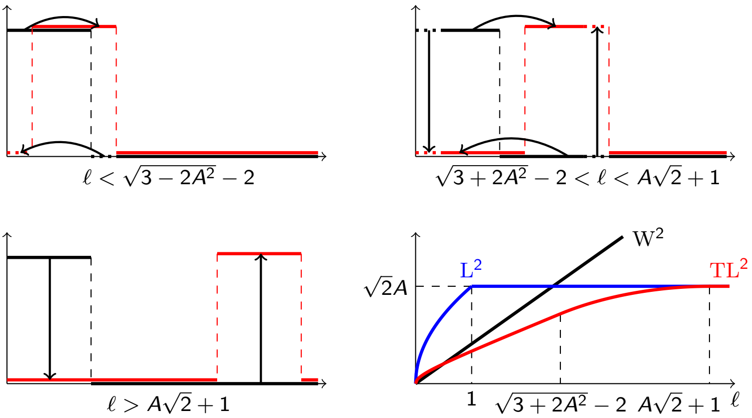


3. In which case we have

$$\mathrm{d}^p_{\mathrm{TL}^p}((f,\mu),(g,\nu)) = \min_{T:T_{\#}\tilde{\mu}=\tilde{\nu}} \int_{X\times\mathbb{R}} |\tilde{x} - T(\tilde{x})|^p \,\mathrm{d}\tilde{\mu}(\tilde{x}).$$

TL^p Translations

TL^2 transport between $f(x) = A\chi_{[0,1]}$ and $g(x) = f(x-\ell)$ with the uniform measure.



TL^p Properties: Summary

1. Signals can be negative and not all of the same size (i.e. not integrate to the same value).
2. Can discriminate between fast oscillating signals (true for L^p, false for W^p).
3. Can track translations for further than L^p (but not as far as W^p).
4. Existing numerical methods for OT are available.
5. The distance defines a metric.
6. We have the existence of plans.
7. Maps exist in the discrete case when $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ and $\nu = \frac{1}{n}\sum_{j=1}^n \delta_{y_j}$.
8. Disadvantages: no geodesics, not complete.

Linear TL^2 Distance

- Fix a reference point $(f, \mu) \in \mathrm{TL}^2$ and let T^* is the TL^2-optimal transport map between (f, μ) and (g, ν). I.e. $T^*_{\#}\mu = \nu$ and

$$\mathrm{d}^2_{\mathrm{TL}^2}((f, \mu), (g, \nu)) = \int_X |x - T^*(x)|^2 + |f(x) - g(T^*(x))|^2 \,\mathrm{d}\mu(x).$$

Linear TL^2 Distance

- Fix a reference point $(f, \mu) \in \mathrm{TL}^2$ and let T^* is the TL^2-optimal transport map between (f, μ) and (g, ν). I.e. $T^*_\# \mu = \nu$ and

$$\mathrm{d}^2_{\mathrm{TL}^2}((f, \mu), (g, \nu)) = \int_X |x - T^*(x)|^2 + |f(x) - g(T^*(x))|^2 \, \mathrm{d}\mu(x).$$

- Assume $\mu = \frac{1}{N} \sum_{k=1}^N \delta_{z_k}$ then we define

$$P_{\mathrm{TL}^2,(f,\mu),\mathrm{lin}}(g, \nu) = (P_1(g, \nu), P_2(g, \nu)) \in \mathbb{R}^{2N}$$
$$[P_1(g, \nu)]_k = T^*(z_k) - z_k$$
$$[P_2(g, \nu)]_k = g(T^*(z_k)) - f(z_k).$$

Linear TL^2 Distance

- Fix a reference point $(f, \mu) \in \mathrm{TL}^2$ and let T^* is the TL^2-optimal transport map between (f, μ) and (g, ν). I.e. $T^*_{\#}\mu = \nu$ and

$$\mathrm{d}^2_{\mathrm{TL}^2}((f,\mu),(g,\nu)) = \int_X |x - T^*(x)|^2 + |f(x) - g(T^*(x))|^2 \,\mathrm{d}\mu(x).$$

- Assume $\mu = \frac{1}{N}\sum_{k=1}^N \delta_{z_k}$ then we define

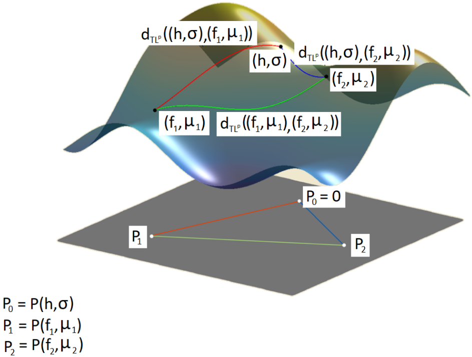
$$\begin{aligned} P_{\mathrm{TL}^2,(f,\mu),\mathrm{lin}}(g,\nu) &= (P_1(g,\nu), P_2(g,\nu)) \in \mathbb{R}^{2N} \\ [P_1(g,\nu)]_k &= T^*(z_k) - z_k \\ [P_2(g,\nu)]_k &= g(T^*(z_k)) - f(z_k). \end{aligned}$$

- The linear TL^2 distance

$$\mathrm{d}_{\mathrm{TL}^2,(f,\mu),\mathrm{lin}}((g,\nu),(h,\omega)) = \|P_{\mathrm{TL}^2,(f,\mu),\mathrm{lin}}(g,\nu) - P_{\mathrm{TL}^2,(f,\mu),\mathrm{lin}}(h,\omega)\|_{\ell^2}.$$

Properties of Linear TL^2

- If $\nu = \frac{1}{N}\sum_{k=1}^{N} \delta_{x_k}$ then $P_{\mathrm{TL}^2,(f,\mu),\mathrm{lin}}(g,\nu) \in \ell^2$.
- $P_{\mathrm{TL}^2,(f,\mu),\mathrm{lin}}(f,\mu) = \underline{0}$.
- $\mathrm{d}_{\mathrm{TL}^2,(f,\mu),\mathrm{lin}}((f,\mu),(g,\nu)) = \mathrm{d}_{\mathrm{TL}^2}((f,\mu),(g,\nu))$.



Spatially Correlated Histogram Specification

Histogram Specification: The problem of matching one histogram $\varphi(y) := f_{\#}\mu(y) = \frac{1}{N}\{x : f(x) = y\}$ with another ψ, i.e. find a map $T : X \to Y$ such that $\psi = T_{\#}\varphi$.

Spatially Correlated Histogram Specification

Histogram Specification: The problem of matching one histogram $\varphi(y) := f_{\#}\mu(y) = \frac{1}{N}\{x : f(x) = y\}$ with another ψ, i.e. find a map $T : X \to Y$ such that $\psi = T_{\#}\varphi$.

Colour Transfer: Colour one image with the palette of an exemplar image.

Spatially Correlated Histogram Specification

Histogram Specification: The problem of matching one histogram $\varphi(y) := f_{\#}\mu(y) = \frac{1}{N}\{x : f(x) = y\}$ with another ψ, i.e. find a map $T : X \to Y$ such that $\psi = T_{\#}\varphi$.

Colour Transfer: Colour one image with the palette of an exemplar image.

W^2 Solution: (For greyscale images) define histograms φ, ψ from the images and let T be the optimal Monge map between them. The recoloured image is $\hat{f} = g \circ T$.

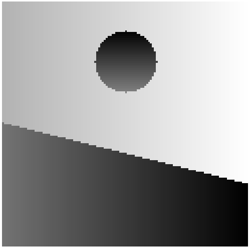
Histogram Specification: The problem of matching one histogram $\varphi(y) := f_{\#}\mu(y) = \frac{1}{N}\{x : f(x) = y\}$ with another ψ, i.e. find a map $T : X \to Y$ such that $\psi = T_{\#}\varphi$.

Colour Transfer: Colour one image with the palette of an exemplar image.

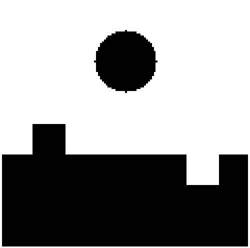
W^2 Solution: (For greyscale images) define histograms φ, ψ from the images and let T be the optimal Monge map between them. The recoloured image is $\hat{f} = g \circ T$.

TL^2 Solution: Let T be the TL^2 optimal map between (f, μ) and (g, ν) (f, g may be RGB images).

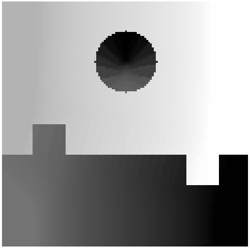
Histogram Specification: Synthetic



(a) Exemplar image.



(b) Original image to be shaded.



(c) The TL^2 solution.

(a) Exemplar image.

(b) Original image to be coloured.

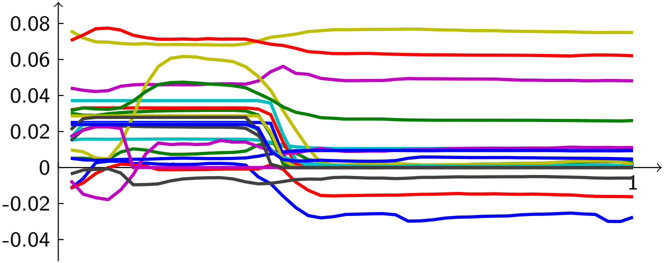
(c) TL^2 solution.

(d) W^2 solution.

(e) Reinhard, Ashikhmin, Gooch and Shirley's method.

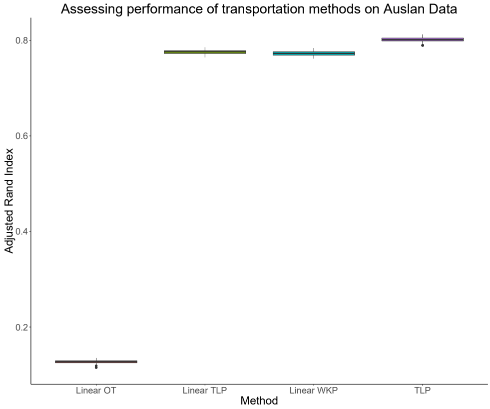
(f) Pitié and Kokaram's method.

1. The AUSLAN data set is a set of 95 words ‘spoken’ by a native AUSLAN (Australian sign language) using 22 sensors on a cyberglove.
2. 27 signals in each class, so a total of 2565 signals.



AUSLAN Results

Accuracy:



Computation time:

Method	Linear W^2	Linear TL^2	Linear $\mathrm{TW}^{k,p}$	TL^p
CPU times (seconds)	12.1	13.0	13.5	91200

References I

1. **Cai, Cheng, Schmtzer and T., *The Linearized Hellinger–Kantorovich Distance*, SIAM Journal on Imaging Sciences, 15(1):45–83, 2022.**
2. Chizat, Peyré, Schmitzer and Vialard, *Unbalanced Optimal Transport: Dynamic and Kantorovich Formulations*, Journal of Functional Analysis, 274(11):3090-3123, 2018.
3. Chizat, Peyré, Schmitzer and Vialard, *An Interpolating Distance Between Optimal Transport and Fisher–Rao Metrics*, Foundations of Computational Mathematics, 18(1):1-44, 2018a.
4. **Crook, Cucuringu, Hurst, Schönlieb, T. and Zygalakis, *A Linear Transportation Lp Distance for Pattern Recognition*, preprint arXiv:2009.11262, 2020.**
5. Gangbo, Li, Osher and Puthawala, *Unnormalized Optimal Transport*, preprint arXiv:1902.03367, 2019.
6. **Kolouri, Park, T., Slepčev and Rohde, *Optimal Mass Transport: Signal Processing and Machine-Learning Applications*, IEEE Signal Processing Magazine, 34(4):43-59, 2017.**
7. Kondratyev, Monsaingeon and Vorotnikov, *A new optimal transport distance on the space of finite Radon measures*, Advances in Differential Equations, 21:1117–1164, 2016.
8. Lee, Lai, Li and Osher, *Generalized Unnormalised Optimal Transport and its Fast Algorithms*, preprint arXiv:2001.11530, 2020.

9. Liero, Mielke and Savaré, *Optimal Entropy-Transport Problems and a New Hellinger–Kantorovich Distance Between Positive Measures*, Inventiones Mathematicae, 211(3):969-1117, 2018.
10. **Park and T., *Representing and Learning High Dimensional Data with the Optimal Transport Map from a Probabilistic Viewpoint*, CVPR, 2018.**
11. **T., Park, Kolouri, Rohde and Slepčev, *A Transportation Lp Distance for Signal Analysis*, Journal of Mathematical Imaging and Vision, 59(2):187-210, 2017.**
12. Wang, Slepčev, Basu, Ozolek and Rohde, *A Linear Optimal Transportation Framework for Quantifying and Visualizing Variations in Sets of Images*, International Journal of Computer Vision 101(2):254–269, 2013.

Thank you for listening!

People worry that computers will get too smart and take over the world, but the real problem is that they're too stupid and they've already taken over the world.

— Pedro Domingos