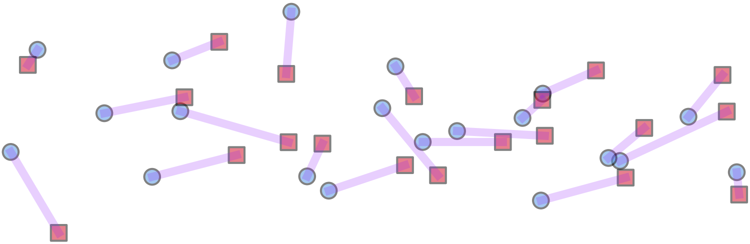
Theory and Computation of Optimal Transport Variants

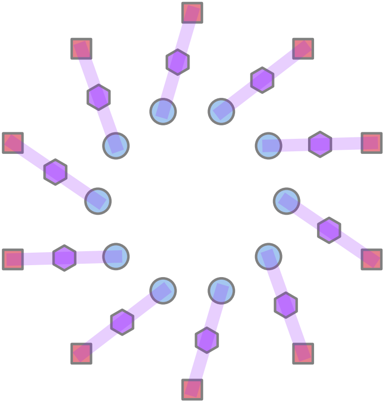
PhD Defence - supervision: Julie Delon and Agnès Desolneux

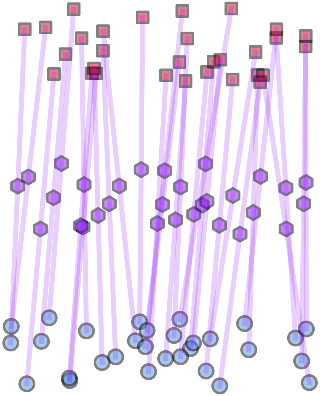


LABORATOIRE MAP5

Université Paris Cité

Discrete Monge Problem [Mon81]

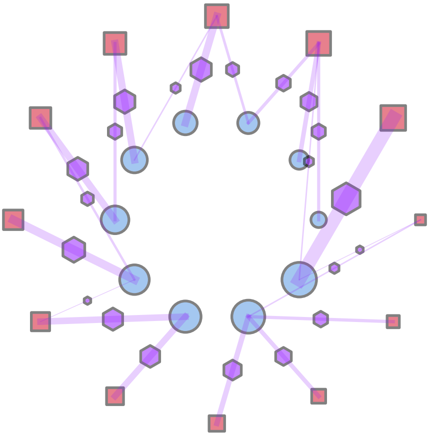


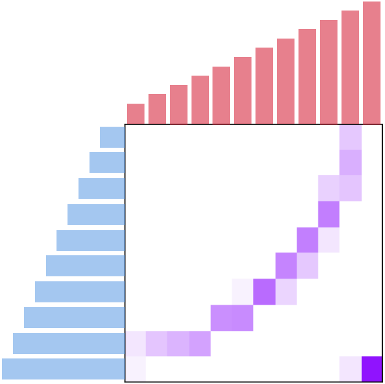


Discrete Monge Problem:

$$\min_{\sigma \in \mathfrak{S}_n} \frac{1}{n} \sum_{i=1}^{n} c(x_i, y_{\sigma(i)}).$$

Discrete Kantorovich Problem [Kan42]





Discrete Kantorovich Problem:

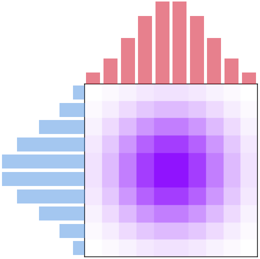
$$\min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P\mathbf{1}=a \\ P^\top \mathbf{1}=b}} \sum_{i=1}^{n} \sum_{j=1}^{m} c(x_i, y_j) P_{i,j}.$$

Comparing Measures with Optimal Transport

Discrete OT Cost

$\mu := \sum_{i=1}^n a_i\delta_{x_i}, \ \nu := \sum_{j=1}^m b_j\delta_{y_j}$

$$\mathcal{T}_c(\mu,\nu) := \min_{\substack{P\in\mathbb{R}_+^{n\times m}\\ P\mathbf{1}=a\\ P^\top\mathbf{1}=b}} \sum_{i=1}^n\sum_{j=1}^m c(x_i,y_j)P_{i,j}.$$

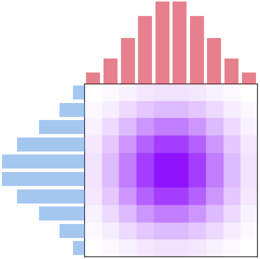


Comparing Measures with Optimal Transport

Discrete OT Cost

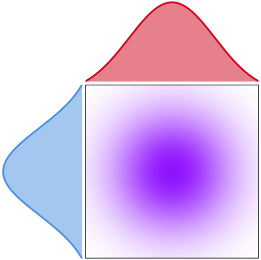
$\mu := \sum_{i=1}^n a_i \delta_{x_i}, \ \nu := \sum_{j=1}^m b_j \delta_{y_j}$

$$\mathcal{T}_c(\mu, \nu) := \min_{\substack{P \in \mathbb{R}_+^{n\times m} \\ P\mathbf{1} = a \\ P^\top \mathbf{1} = b}} \sum_{i=1}^n \sum_{j=1}^m c(x_i, y_j) P_{i,j}.$$



OT Cost $\mu \in \mathcal{P}(\mathcal{X}),\ \nu \in \mathcal{P}(\mathcal{Y})$

$$\mathcal{T}_c(\mu, \nu) := \inf_{\substack{\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}) \\ \pi_{\mathcal{X}} = \mu \\ \pi_{\mathcal{Y}} = \nu}} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y) \mathrm{d}\pi(x, y).$$

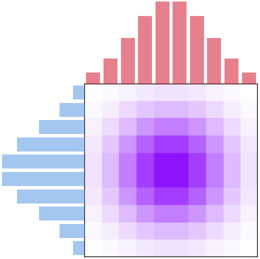


Comparing Measures with Optimal Transport

Discrete OT Cost

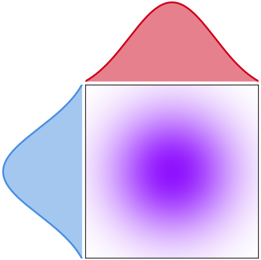
$\mu := \sum_{i=1}^n a_i \delta_{x_i}, \ \nu := \sum_{j=1}^m b_j \delta_{y_j}$

$$\boxed{\mathcal{T}_c(\mu, \nu) := \min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P\mathbf{1} = a \\ P^\top \mathbf{1} = b}} \sum_{i=1}^n \sum_{j=1}^m c(x_i, y_j) P_{i,j}.}$$



OT Cost $\mu \in \mathcal{P}(\mathcal{X}), \ \nu \in \mathcal{P}(\mathcal{Y})$

$$\boxed{\mathcal{T}_c(\mu, \nu) := \inf_{\substack{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \\ \pi_\mathcal{X} = \mu \\ \pi_\mathcal{Y} = \nu}} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \mathrm{d}\pi(x, y).}$$



p-Wasserstein Distance: $\mathcal{X} = \mathcal{Y}$ and $c(x, y) = d_\mathcal{X}(x, y)^p$.

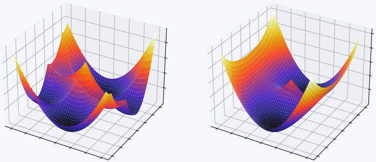
Random Variable Formulation: $\mathcal{T}_c(\mu, \nu) = \inf_{(X,Y):\ X \sim \mu, Y \sim \nu} \mathbb{E}\left[c(X, Y)\right].$

Part A: Optimal Transport Discrepancies as Losses

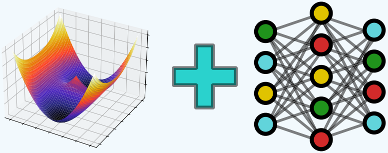
Chapter A.I: Reconstructing Discrete Measures from Projections [TFD24b].

$$\left\{\gamma \in \mathcal{P}(\mathbb{R}^d) : \forall i \in [\![p]\!],\ P_i\#\gamma = P_i\#\gamma_Z\right\}.$$

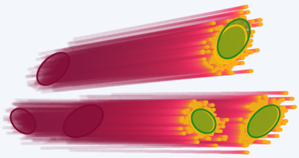
Chapter A.II: Properties of Discrete Sliced Wasserstein Losses [TFD24a].



Chapter A.III: Convergence of SGD for Training Neural Networks with Sliced Wasserstein Losses [Tan23].



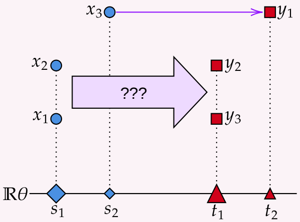
Chapter A.IV: Differentiable Expectation-Maximisation and Applications to Gaussian Mixture Model Optimal Transport [Boï+25].



Part B: Variants of Optimal Transport Maps and Plans

Chapter B.I: Constrained Approximate Optimal Transport Maps [TDD25] $\underset{g \in G}{\operatorname{argmin}} \mathcal{T}_c(g\#\mu, \nu)$.

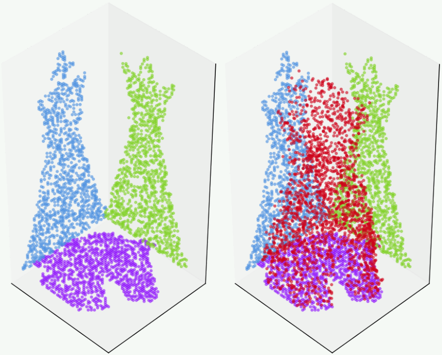
Chapter B.II: Sliced OT Plans [TCD25].



Chapter B.III: Sliced Gromov Wasserstein.

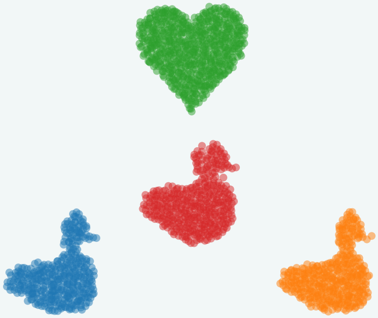
Part C: Optimal Transport Barycentres

Chapter C.I: (Blind) Generalised Wasserstein Barycentres.



Extends [DGS21].

Chapter C.II: Computing OT Barycentres [TDG24; TDG25].



Extends [CD14; Álv+16].

Part D: Some Contributions to Kernel Methods

Chapter D.I: A Gentle Introduction to RKHS.

Chapter D.II: On Gradients of Convex Functions in RKHS.

Chapter D.III: Explicit Universal and Approximate Universal Kernels on Compact Metric Spaces [Tan25].

List of Papers

Published works

[Tan23] **ET**. "Convergence of SGD for Training Neural Networks with Sliced Wasserstein Losses". *Transactions on Machine Learning Research* (Oct. 2023).

[TFD24a] **ET**, Rémi Flamary and Julie Delon. "Properties of Discrete Sliced Wasserstein Losses". *Mathematics of Computation* (Jun. 2024).

[TFD24b] **ET**, Rémi Flamary and Julie Delon. "Reconstructing discrete measures from projections. Consequences on the empirical Sliced Wasserstein Distance". *Comptes Rendus. Mathématique* 362 (Jun. 2024), pp. 1121-1129.

[TDD25] **ET**, Agnès Desolneux, and Julie Delon. "Constrained Approximate Optimal Transport Maps". *ESAIM: Control, Optimisation and Calculus of Variations*. (Aug. 2025)

Pre-prints

[TDG24] **ET**, Julie Delon and Nathaël Gozlan. "Computing Barycentres of Measures for Generic Transport Costs". *arxiv preprint* 2501.04016 (Dec. 2024).

[Tan25] **ET**. "Explicit Universal and Approximate-Universal Kernels on Compact Metric Spaces". *arxiv preprint* 2506.03661 (Jun. 2025).

[TCD25] **ET**, Laetitia Chapel and Julie Delon. "Sliced Optimal Transport Plans". *arxiv preprint* 2508.01243 (Aug. 2025).

[Boï+25] Samuel Boïté*, **ET***, Julie Delon, Agnès Desolneux and Rémi Flamary. "Differentiable Expectation-Maximisation and Applications to Gaussian Mixture Model Optimal Transport". *arxiv preprint* 2509.02109 (Sept. 2025). (*: equal contribution)

[Sis+25] Keanu Sisouk, **ET**, Julie Delon and Julien Tierny. "Robust Barycenters of Persistence Diagrams". *arxiv preprint* 2509.14904 (Sept. 2025).

Open-Source Code Contributions

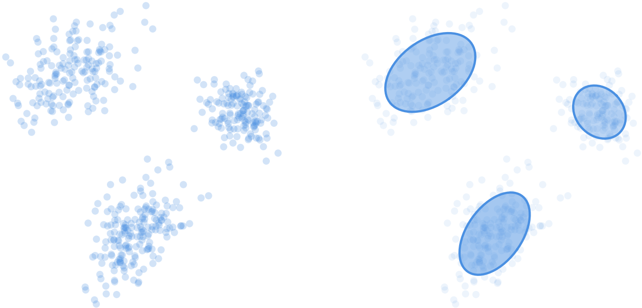
Github repositories:

- (Blind) generalised Wasserstein barycentres [DGS21]: eloitanguy/bgwb.
- OT barycentres [TDG24]: eloitanguy/ot_bar.
- GMM-OT toolbox [DD20; Boï+25] (WIP).

Contributions to the `POT` library [Fla+21]:

- Generalised Wasserstein barycentres [DGS21]: PR 372.
- Nearest Brenier potentials [PdC20; TDD25]: PR 526.
- GMM OT [DD20; Boï+25]: PR 649.
- Fixed-point solvers for OT barycentres [TDG24]: PR 715.
- Sliced OT plans [TCD25]: PR 767.

Differentiable EM and Applications to GMM OT [Boï+25]

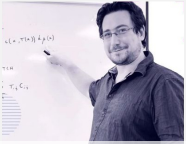


$$\text{Data } X \in (\mathbb{R}^d)^n \xmapsto{\ \nabla\ ?\ } \mathrm{EM}(X) : \text{GMM}$$

Samuel Boïté

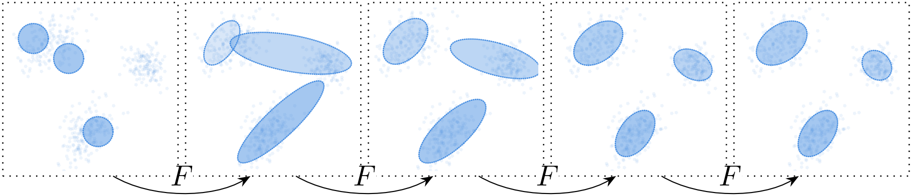
Julie Delon

Agnès Desolneux



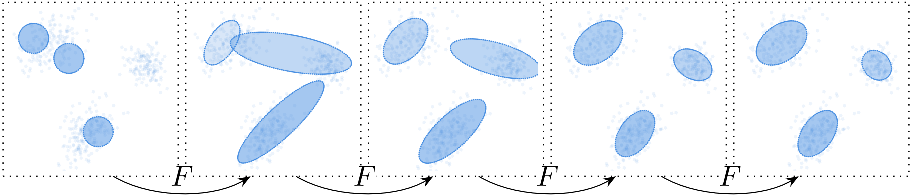
Rémi Flamary

Expectation-Maximisation as a Fixed-Point Algorithm



$$\boxed{\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).}$$

Expectation-Maximisation as a Fixed-Point Algorithm

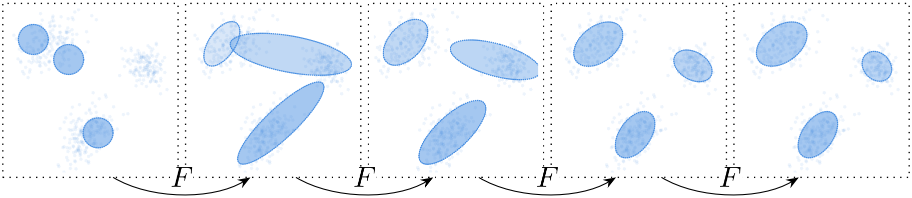


$$\boxed{\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).}$$

Differentiation methods:

- Automatic: $\dfrac{\partial \theta_T}{\partial X} = \dfrac{\partial}{\partial X}\left[F_X^T(\theta_0)\right].$

Expectation-Maximisation as a Fixed-Point Algorithm



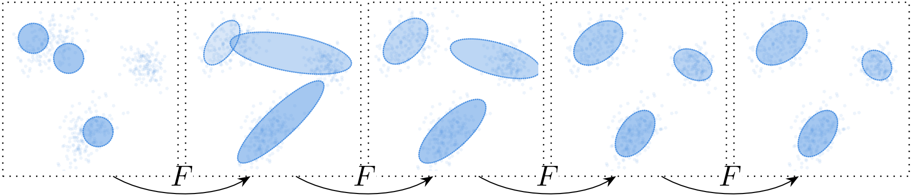
$$\boxed{\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).}$$

Differentiation methods:

- Automatic: $\dfrac{\partial \theta_T}{\partial X} = \dfrac{\partial}{\partial X}\left[F_X^T(\theta_0)\right].$

$$F(\theta_\infty, X) = \theta_\infty.$$

Expectation-Maximisation as a Fixed-Point Algorithm



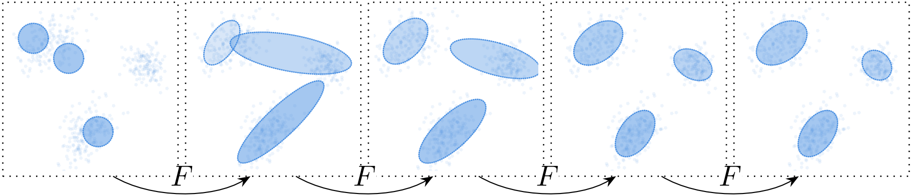
$$\boxed{\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).}$$

Differentiation methods:

- Automatic: $\dfrac{\partial \theta_T}{\partial X} = \dfrac{\partial}{\partial X}\left[F_X^T(\theta_0)\right].$

$$\frac{\partial}{\partial X}\Big[F(\theta_\infty, X)\Big] = \frac{\partial \theta_\infty}{\partial X}.$$

Expectation-Maximisation as a Fixed-Point Algorithm



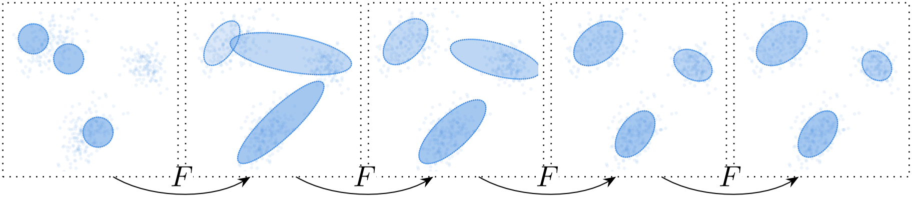
$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

- Automatic: $\dfrac{\partial \theta_T}{\partial X} = \dfrac{\partial}{\partial X}\left[F_X^T(\theta_0)\right].$

$$\frac{\partial F}{\partial \theta}(\theta_\infty, X)\frac{\partial \theta_\infty}{\partial X} + \frac{\partial F}{\partial X}(\theta_\infty, X) = \frac{\partial \theta_\infty}{\partial X}.$$

Expectation-Maximisation as a Fixed-Point Algorithm



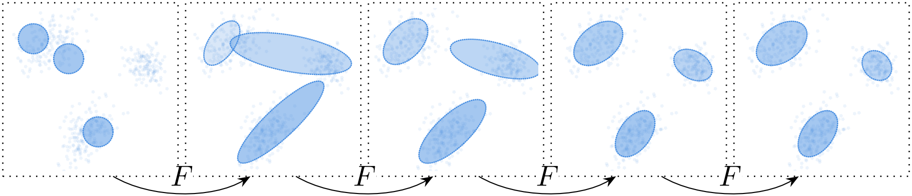
$$\boxed{\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).}$$

Differentiation methods:

- Automatic: $\dfrac{\partial \theta_T}{\partial X} = \dfrac{\partial}{\partial X}\left[F_X^T(\theta_0)\right].$

$$\frac{\partial \theta_\infty}{\partial X} = \left(I - \frac{\partial F}{\partial \theta}(\theta_\infty, X)\right)^{-1} \frac{\partial F}{\partial X}(\theta_\infty, X).$$

Expectation-Maximisation as a Fixed-Point Algorithm

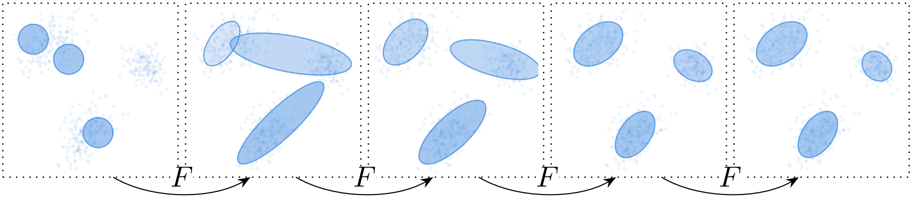


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

Differentiation methods:

- Automatic: $\dfrac{\partial \theta_T}{\partial X} = \dfrac{\partial}{\partial X}\left[F_X^T(\theta_0)\right].$
- Implicit: $\dfrac{\partial \theta_T}{\partial X} \approx \left(I - \dfrac{\partial F}{\partial \theta}(\theta_T, X)\right)^{-1} \dfrac{\partial F}{\partial X}(\theta_T, X).$

Expectation-Maximisation as a Fixed-Point Algorithm

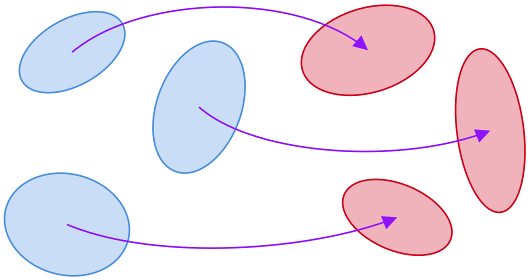


$$\theta_t = F(\theta_{t-1}, X) = F_X^t(\theta_0).$$

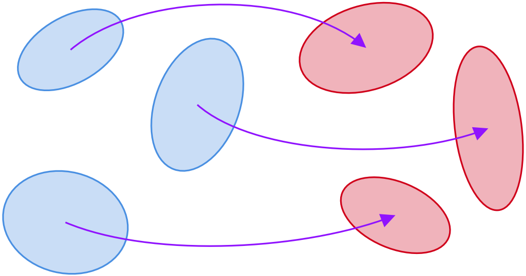
Differentiation methods:

- Automatic: $\dfrac{\partial \theta_T}{\partial X} = \dfrac{\partial}{\partial X}\left[F_X^T(\theta_0)\right].$
- Implicit: $\dfrac{\partial \theta_T}{\partial X} \approx \left(I - \dfrac{\partial F}{\partial \theta}(\theta_T, X)\right)^{-1} \dfrac{\partial F}{\partial X}(\theta_T, X).$
- One-Step: $\dfrac{\partial \theta_T}{\partial X} \approx \dfrac{\partial F}{\partial X}(\theta_T, X).$

Crash Course on GMMOT [DD20]



Crash Course on GMMOT [DD20]

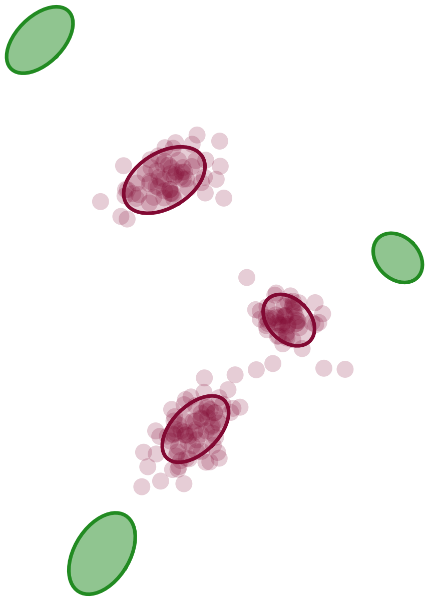


Mixture-Wasserstein Distance [DD20]

$$\mathrm{MW}_2^2(\mu_0,\mu_1) = \min_{P\in\Pi(w_0,w_1)} \sum_{k,\ell} P_{k\ell}\mathrm{W}_2^2(\mu_{0,k},\mu_{1,\ell})$$
$$= \min_{\pi\in\Pi(\mu_0,\mu_1)\cap\mathrm{GMM}} \int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 \mathrm{d}\pi(x_1,x_2).$$

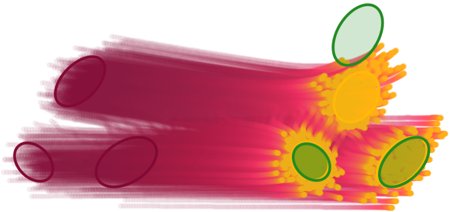
EM-MW2 as a Loss

Gradient Descent on $X \longmapsto \mathrm{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.

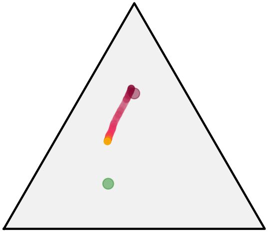


Fixing GMM Weights

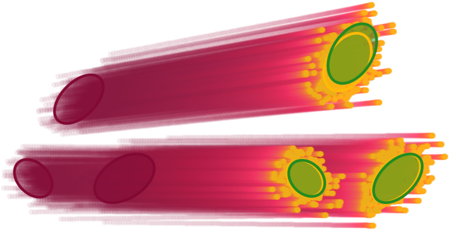
Gradient Descent on $X \longmapsto \mathrm{MW}_2^2(\mu(F_X^T(\theta_0)), \nu)$.



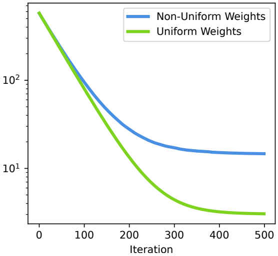
Particle flow (non-uniform).



Weight evolution (non-uniform).

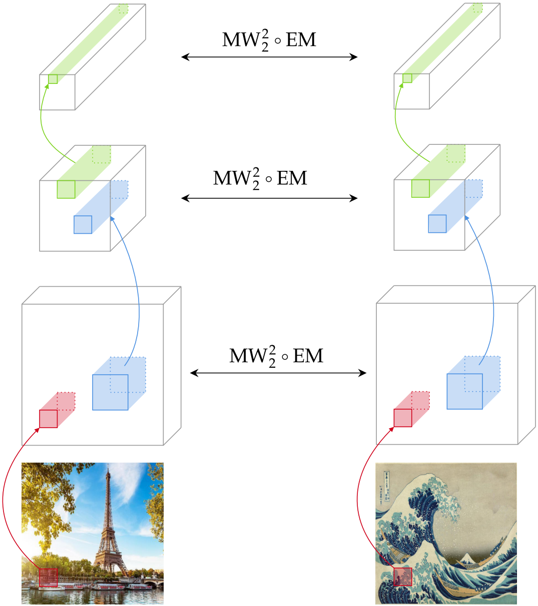


Particle flow (uniform).



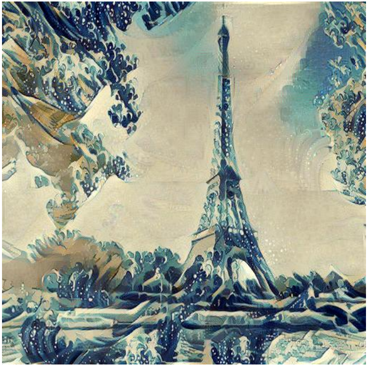
Energy evolutions.

Application: Style Transfer



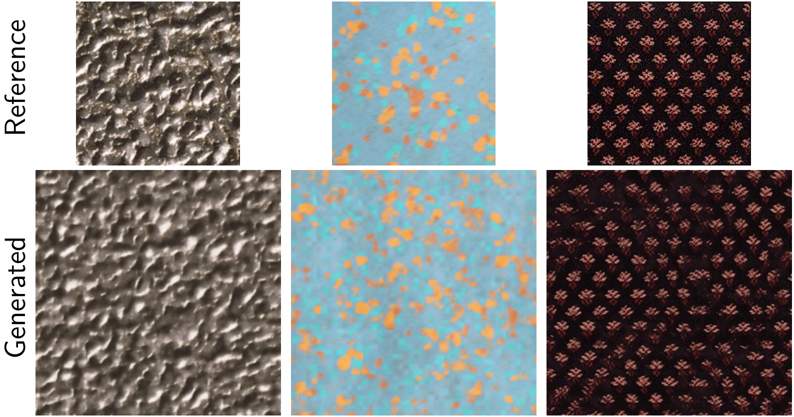
Optimised Image

Style Target

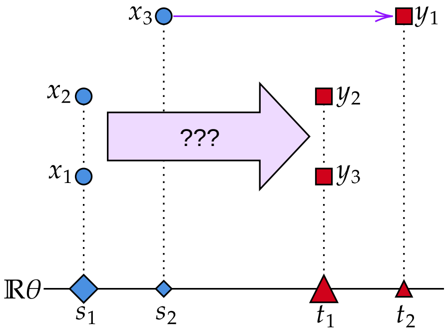


Application: Texture Synthesis

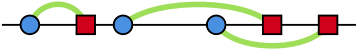
$$\min_X \sum_{\ell=1}^{L} \lambda_\ell \mathrm{MW}_2^2\Big(F^T \circ \mathrm{Patches}_{p\times p} \circ \mathrm{Downscale}_{s_\ell}(X), F^T \circ \mathrm{Patches}_{p\times p} \circ \mathrm{Downscale}_{s_\ell}(Y)\Big).$$



Sliced OT Plans [TCD25]

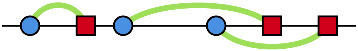


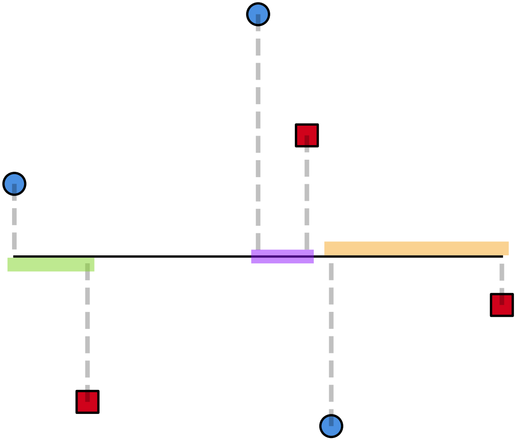
Sliced Wasserstein Distance [Rab+12]



Sliced Wasserstein Distance [Rab+12]

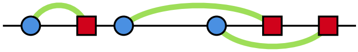
$$\mathrm{SW}_2^2(\mu_1, \mu_2) := \int_{\mathbb{S}^{d-1}} \mathrm{W}_2^2(P_\theta \# \mu_1, P_\theta \# \mu_2) \mathrm{d}\sigma(\theta).$$

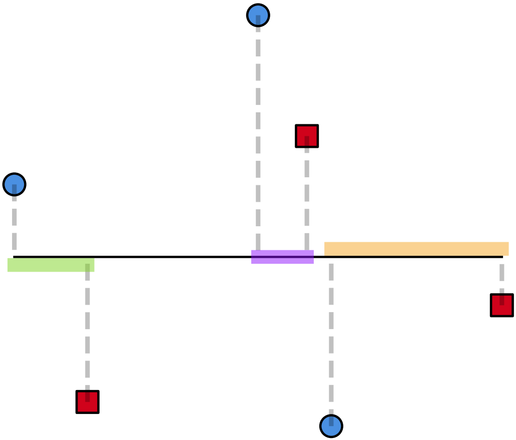




Sliced Wasserstein Distance [Rab+12]

$$\mathrm{SW}_2^2(\mu_1, \mu_2) := \int_{\mathbb{S}^{d-1}} \mathrm{W}_2^2(P_\theta \# \mu_1, P_\theta \# \mu_2) \mathrm{d}\sigma(\theta).$$

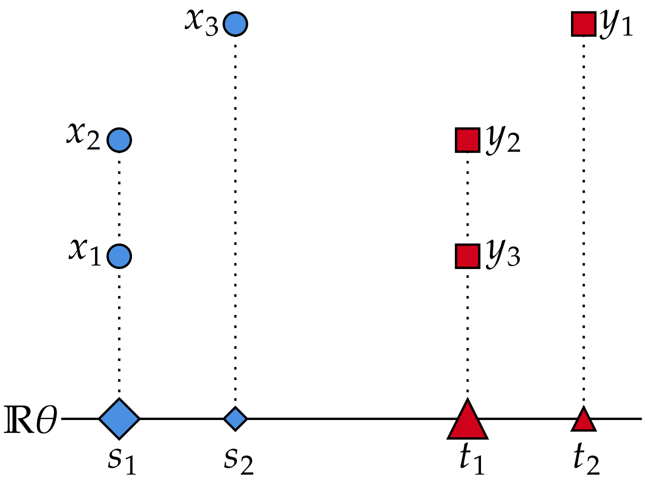




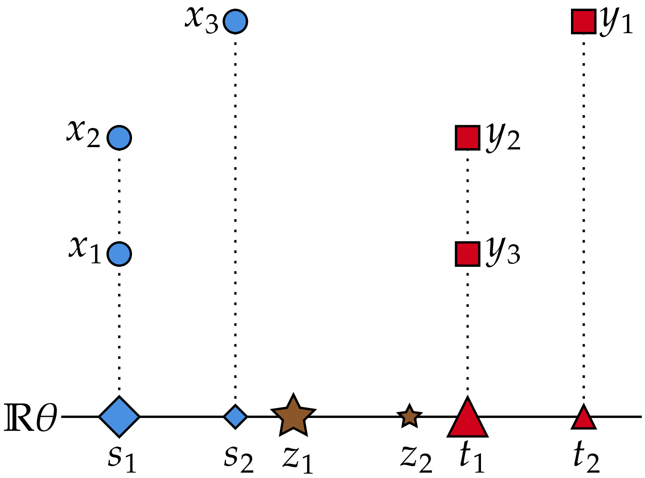
Sliced plan propositions:

- SWGG [Mah+23]
  ↳ Pivot-Sliced
- Expected SW [Liu+24]
  ↳ more theory

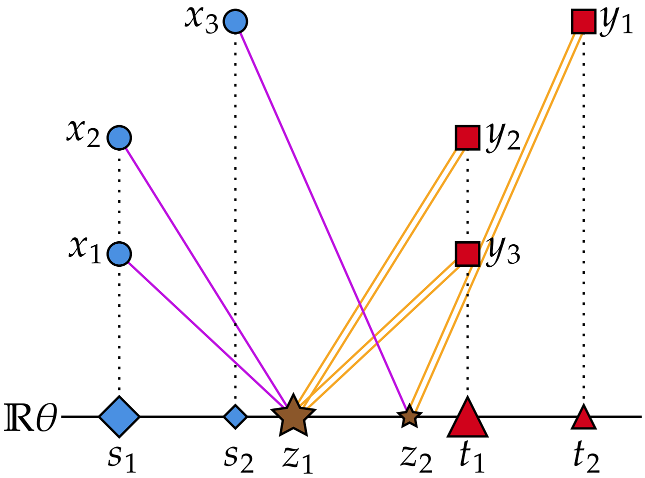
Pivot-Sliced: Definition and Reformulation



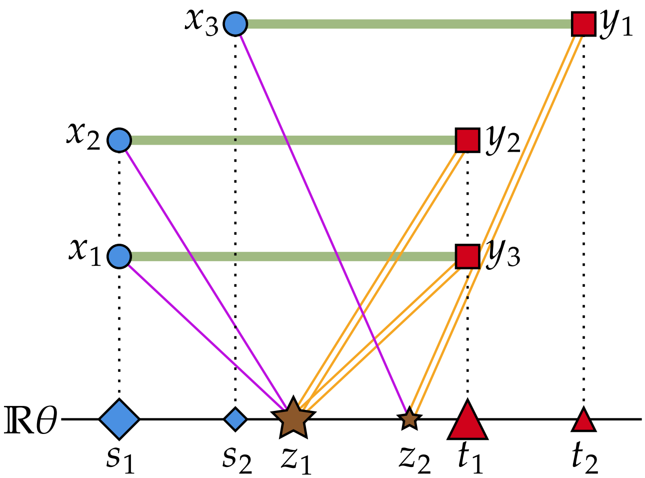
Pivot-Sliced: Definition and Reformulation



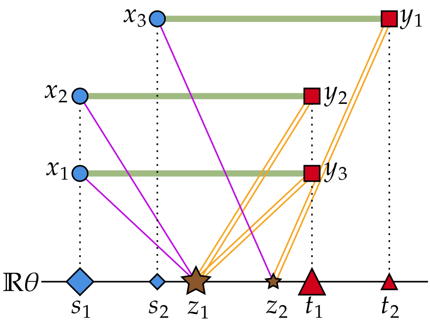
Pivot-Sliced: Definition and Reformulation



Pivot-Sliced: Definition and Reformulation



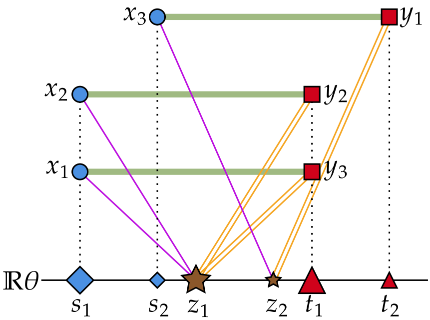
Pivot-Sliced: Definition and Reformulation



$$\Gamma(\nu, \mu_1, \mu_2) := \left\{\rho \in \mathcal{P}_2(\mathbb{R}^{3d}) : \rho_{0,1} \in \Pi^*(\nu, \mu_1) \text{ and } \rho_{0,2} \in \Pi^*(\nu, \mu_2)\right\}.$$

$$\mathrm{PS}_\theta^2(\mu_1, \mu_2) := \min_{\rho \in \Gamma(\mu_\theta, \mu_1, \mu_2)} \int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho(y, x_1, x_2).$$

Pivot-Sliced: Definition and Reformulation



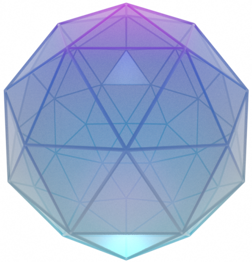
$$\Gamma(\nu, \mu_1, \mu_2) := \left\{\rho \in \mathcal{P}_2(\mathbb{R}^{3d}) : \rho_{0,1} \in \Pi^*(\nu, \mu_1) \text{ and } \rho_{0,2} \in \Pi^*(\nu, \mu_2)\right\}.$$

$$\mathrm{PS}_\theta^2(\mu_1, \mu_2) := \min_{\rho \in \Gamma(\mu_\theta, \mu_1, \mu_2)} \int_{\mathbb{R}^{3d}} \|x_1 - x_2\|_2^2 \mathrm{d}\rho(y, x_1, x_2).$$

Equivalent Formulation

$$\mathrm{PS}_\theta^2(\mu_1, \mu_2) = \min_{\substack{\omega \in \Pi(\mu_1, \mu_2) \\ (P_\theta, P_\theta)\#\omega \in \Pi^*(P_\theta\#\mu_1, P_\theta\#\mu_2)}} \int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 \mathrm{d}\omega(x_1, x_2).$$

Pivot-Sliced: Point Cloud Monge Formulation

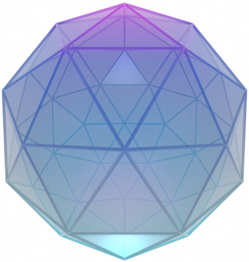


Extreme Optimality [Thm 2.7 in BT97]

Let $\mathcal{P}$ be a convex and compact polytope.

$$\min_{x \in \mathcal{P}} c \cdot x = \min_{x \in \operatorname{Extr} \mathcal{P}} c \cdot x.$$

Pivot-Sliced: Point Cloud Monge Formulation



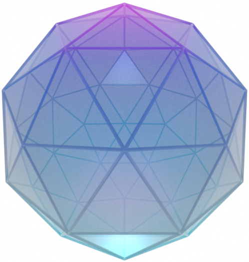
Extreme Optimality [Thm 2.7 in BT97]

Let $\mathcal{P}$ be a convex and compact polytope.

$$\min_{x \in \mathcal{P}} c \cdot x = \min_{x \in \operatorname{Extr} \mathcal{P}} c \cdot x.$$

$$\mathrm{PS}_\theta^2 \left(\frac{1}{n} \sum_{i=1}^n \delta_{x_i}, \frac{1}{n} \sum_{j=1}^n \delta_{y_j} \right)$$

Pivot-Sliced: Point Cloud Monge Formulation



Extreme Optimality [Thm 2.7 in BT97]

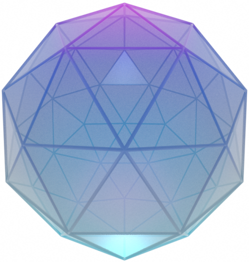
Let $\mathcal{P}$ be a convex and compact polytope.

$$\min_{x \in \mathcal{P}} c \cdot x = \min_{x \in \operatorname{Extr} \mathcal{P}} c \cdot x.$$

$$\mathrm{PS}_\theta^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \frac{1}{n}\sum_{j=1}^n \delta_{y_j}\right) = \min_{P \in \mathcal{P}} \sum_{i=1}^n \sum_{j=1}^n \|x_i - y_j\|_2^2 P_{i,j},$$

LP formulation

Pivot-Sliced: Point Cloud Monge Formulation



Extreme Optimality [Thm 2.7 in BT97]

Let $\mathcal{P}$ be a convex and compact polytope.

$$\min_{x \in \mathcal{P}} c \cdot x = \min_{x \in \operatorname{Extr} \mathcal{P}} c \cdot x.$$

$$\mathrm{PS}_\theta^2 \left(\frac{1}{n} \sum_{i=1}^n \delta_{x_i}, \frac{1}{n} \sum_{j=1}^n \delta_{y_j} \right)$$

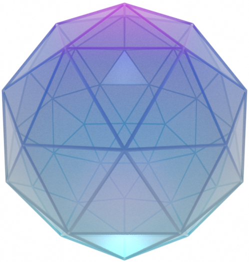
LP formulation

$$= \min_{P \in \mathcal{P}} \sum_{i=1}^n \sum_{j=1}^n \|x_i - y_j\|_2^2 P_{i,j},$$

E.P. Thm

$$= \min_{P \in \operatorname{Extr} \mathcal{P}} \sum_{i=1}^n \sum_{j=1}^n \|x_i - y_j\|_2^2 P_{i,j},$$

Pivot-Sliced: Point Cloud Monge Formulation



Extreme Optimality [Thm 2.7 in BT97]

Let $\mathcal{P}$ be a convex and compact polytope.

$$\min_{x \in \mathcal{P}} c \cdot x = \min_{x \in \operatorname{Extr} \mathcal{P}} c \cdot x.$$

$$\mathrm{PS}_\theta^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \frac{1}{n}\sum_{j=1}^n \delta_{y_j}\right)$$

LP formulation

$$= \min_{P \in \mathcal{P}} \sum_{i=1}^n \sum_{j=1}^n \|x_i - y_j\|_2^2 P_{i,j},$$

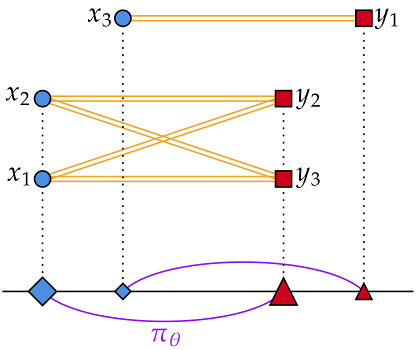
E.P. Thm

$$= \min_{P \in \operatorname{Extr} \mathcal{P}} \sum_{i=1}^n \sum_{j=1}^n \|x_i - y_j\|_2^2 P_{i,j},$$

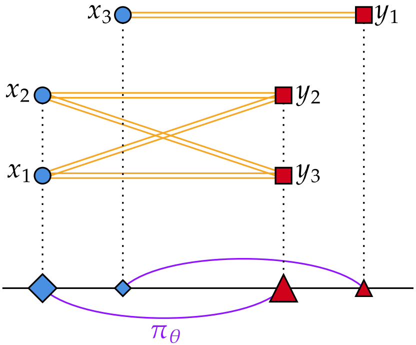
$\operatorname{Extr} \mathcal{P} = \cdots$

$$= \min_{(\sigma,\tau) \in \mathfrak{S}_\theta(X,Y)} \frac{1}{n} \sum_{i=1}^n \|x_{\sigma(i)} - y_{\tau(i)}\|_2^2.$$

Expected Sliced: Definition



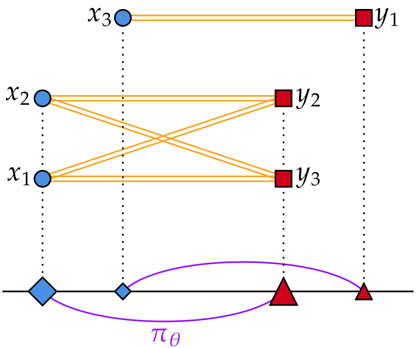
Expected Sliced: Definition



Definition of $\gamma_\theta[\mu_1, \mu_2]$

$$\int_{\mathbb{R}^{2d}} \phi(x_1, x_2) \mathrm{d}\gamma_\theta(x_1, x_2) = \int_{\mathbb{R}^2} \left(\int_{P_\theta^{-1}(s) \times P_\theta^{-1}(t)} \phi(x_1, x_2) \mathrm{d}\mu_1^s(x_1) \mathrm{d}\mu_2^t(x_2) \right) \mathrm{d}\pi_\theta(s, t).$$

Expected Sliced: Definition

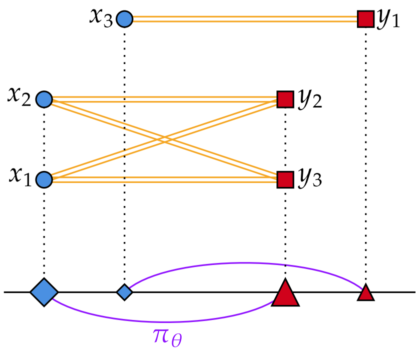


Definition of $\gamma_\theta[\mu_1, \mu_2]$

$$\int_{\mathbb{R}^{2d}} \phi(x_1, x_2) \mathrm{d}\gamma_\theta(x_1, x_2) = \int_{\mathbb{R}^2} \left(\int_{P_\theta^{-1}(s) \times P_\theta^{-1}(t)} \phi(x_1, x_2) \mathrm{d}\mu_1^s(x_1) \mathrm{d}\mu_2^t(x_2) \right) \mathrm{d}\pi_\theta(s, t).$$

$$\mathrm{LS}_\theta^2(\mu_1, \mu_2) := \int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 \mathrm{d}\gamma_\theta[\mu_1, \mu_2](x_1, x_2).$$

Expected Sliced: Definition



Definition of $\gamma_\theta[\mu_1, \mu_2]$

$$\int_{\mathbb{R}^{2d}} \phi(x_1, x_2) \mathrm{d}\gamma_\theta(x_1, x_2) = \int_{\mathbb{R}^2} \left(\int_{P_\theta^{-1}(s) \times P_\theta^{-1}(t)} \phi(x_1, x_2) \mathrm{d}\mu_1^s(x_1) \mathrm{d}\mu_2^t(x_2) \right) \mathrm{d}\pi_\theta(s, t).$$

$$\mathrm{LS}_\theta^2(\mu_1, \mu_2) := \int_{\mathbb{R}^{2d}} \|x_1 - x_2\|_2^2 \mathrm{d}\gamma_\theta[\mu_1, \mu_2](x_1, x_2).$$

$$\overline{\gamma}[\mu_1, \mu_2, \sigma] := \int_{\mathbb{S}^{d-1}} \gamma_\theta[\mu_1, \mu_2] \mathrm{d}\sigma(\theta); \quad \mathrm{ES}_\sigma^2(\mu_1, \mu_2) := \int_{\mathbb{R}^{2d}} \| \cdot - \cdot \|_2^2 \mathrm{d}\overline{\gamma}[\mu_1, \mu_2, \sigma].$$

Are (min)-Pivot-Sliced and Expected Sliced Distances?

$$\min \mathrm{PS}^2(\mu_1, \mu_2) := \min_{\theta \in \mathbb{S}^{d-1}} \mathrm{PS}_\theta^2(\mu_1, \mu_2);$$

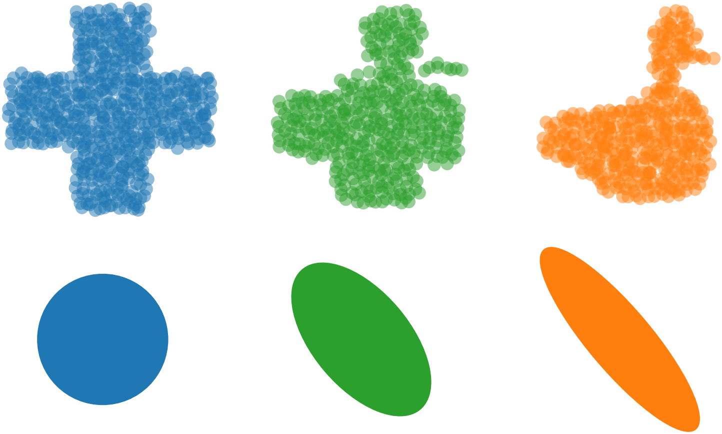
$$\mathrm{ES}_\varpi^2(\mu_1, \mu_2) := \int_{\mathbb{R}^{2d}} \| \cdot - \cdot \|_2^2 \mathrm{d}\overline{\gamma}[\mu_1, \mu_2, \varpi].$$

	PS_θ	$\min \mathrm{PS}$	LS_θ	ES_ϖ
$D(\mu_1, \mu_2) = 0 \implies \mu_1 = \mu_2$	✔	✔	✔	✔
$D(\mu, \mu) = 0$	✔	✔	✘ discrete only	✘
Triangle Inequality	✘ AC only	✘	✔	✔

Optimal Transport Barycentres [TDG24]

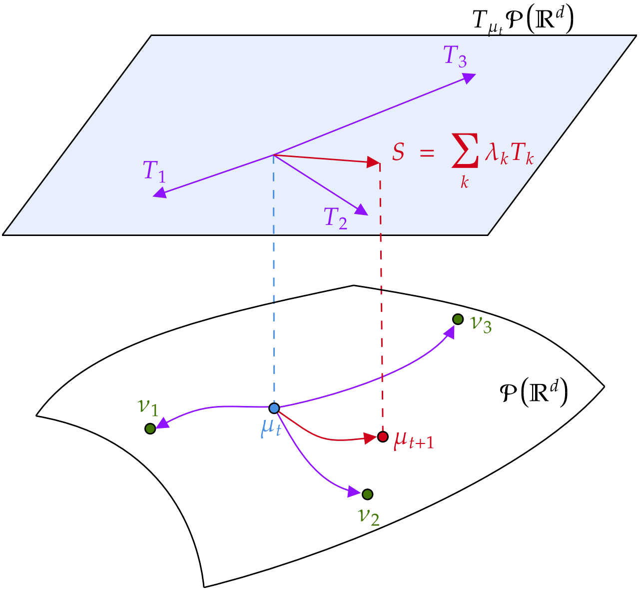
2-Wasserstein Barycentres [AC11]

$$\underset{\mu \in \mathcal{P}(\mathbb{R}^d)}{\operatorname{argmin}} \sum_{k=1}^{K} \lambda_k \mathrm{W}_2^2(\mu, \nu_k).$$



Fixed-Point Method [Álv+16]

Assumptions: $c(x,y) = \|x-y\|_2^2$, AC measures on $\mathbb{R}^d$.



Generalising Wasserstein Barycentres

Setting:

- $(\mathcal{X}, d_{\mathcal{X}})$ compact metric space for barycentres μ.
- $(\mathcal{Y}_k, d_{\mathcal{Y}_k})$ compact metric spaces for measures ν_k.
- $c_k : \mathcal{X} \times \mathcal{Y}_k \longrightarrow \mathbb{R}_+$ continuous cost functions.

Generalising Wasserstein Barycentres

Setting:

- $(\mathcal{X}, d_{\mathcal{X}})$ compact metric space for barycentres μ.
- $(\mathcal{Y}_k, d_{\mathcal{Y}_k})$ compact metric spaces for measures ν_k.
- $c_k : \mathcal{X} \times \mathcal{Y}_k \longrightarrow \mathbb{R}_+$ continuous cost functions.

$$\underset{\mu \in \mathcal{P}(\mathcal{X})}{\operatorname{argmin}} V(\mu), \quad V(\mu) := \sum_{k=1}^{K} \mathcal{T}_{c_k}(\mu, \nu_k).$$

Generalising Wasserstein Barycentres

Setting:

- $(\mathcal{X}, d_{\mathcal{X}})$ compact metric space for barycentres μ.
- $(\mathcal{Y}_k, d_{\mathcal{Y}_k})$ compact metric spaces for measures ν_k.
- $c_k : \mathcal{X} \times \mathcal{Y}_k \longrightarrow \mathbb{R}_+$ continuous cost functions.

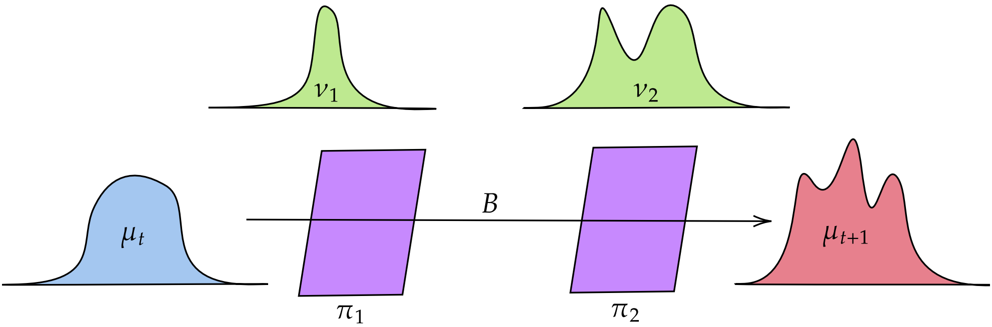
$$\underset{\mu \in \mathcal{P}(\mathcal{X})}{\text{argmin}}\ V(\mu), \quad V(\mu) := \sum_{k=1}^{K} \mathcal{T}_{c_k}(\mu, \nu_k).$$

Assumption: The ground barycenter function

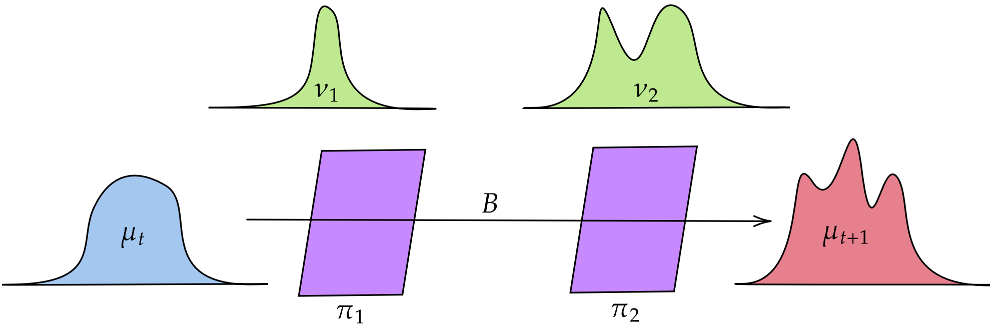
$$B(y_1, \cdots, y_k) := \underset{x \in \mathcal{X}}{\text{argmin}} \sum_{k=1}^{K} c_k(x, y_k)$$

is well-defined.

Fixed-point Algorithm



Fixed-point Algorithm



$$\Gamma(\mu) := \left\{(X, Y_1, \cdots, Y_K) \ : \ (X, Y_k) \in \Pi^*_{c_k}(\mu, \nu_k)\right\},$$

$$G := \begin{cases} \mathcal{P}(\mathcal{X}) & \rightrightarrows \quad \mathcal{P}(\mathcal{X}) \\ \mu & \mapsto \quad \{\mathrm{Law}\left[B(Y_1, \cdots, Y_K)\right] \ : (X, Y_1, \cdots, Y_K) \in \Gamma(\mu)\} \end{cases}.$$

$$\boxed{\mu_{t+1} \in G(\mu_t).}$$

Algorithm Convergence

Ground Barycentre Inequality [Lemma 3.8 in TDG24]

$$\sum_k c_k(x, y_k) \geq \sum_k c_k(\overline{x}, y_k) + \delta(x, \overline{x}), \quad \overline{x} := B(y_1, \cdots, y_K).$$

Case $\|x - y\|_2^2$: simply $\sum_k \lambda_k \|x - y_k\|_2^2 = \sum_k \|\overline{x} - y_k\|_2^2 + \|x - \overline{x}\|_2^2$.

Algorithm Convergence

Ground Barycentre Inequality [Lemma 3.8 in TDG24]

$$\sum_k c_k(x, y_k) \geq \sum_k c_k(\overline{x}, y_k) + \delta(x, \overline{x}), \quad \overline{x} := B(y_1, \cdots, y_K).$$

Case $\|x - y\|_2^2$: simply $\sum_k \lambda_k \|x - y_k\|_2^2 = \sum_k \|\overline{x} - y_k\|_2^2 + \|x - \overline{x}\|_2^2$.

Decrease Property [Proposition 3.9 in TDG24]

$$\forall \overline{\mu} \in G(\mu),\ V(\mu) \geq V(\overline{\mu}) + \mathcal{T}_\delta(\mu, \overline{\mu}).$$

If μ^* is a barycentre then $G(\mu^*) = \{\mu^*\}$.

Algorithm Convergence

Ground Barycentre Inequality [Lemma 3.8 in TDG24]

$$\sum_k c_k(x, y_k) \geq \sum_k c_k(\overline{x}, y_k) + \delta(x, \overline{x}), \quad \overline{x} := B(y_1, \cdots, y_K).$$

Case $\|x-y\|_2^2$: simply $\sum_k \lambda_k \|x - y_k\|_2^2 = \sum_k \|\overline{x} - y_k\|_2^2 + \|x - \overline{x}\|_2^2$.

Decrease Property [Proposition 3.9 in TDG24]

$$\forall \overline{\mu} \in G(\mu),\ V(\mu) \geq V(\overline{\mu}) + \mathcal{T}_\delta(\mu, \overline{\mu}).$$

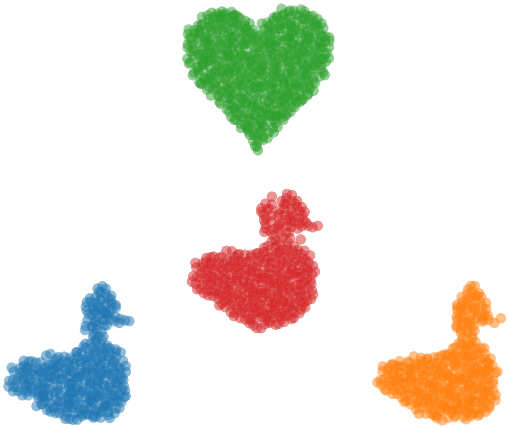
If μ^* is a barycentre then $G(\mu^*) = \{\mu^*\}$.

Convergence [Theorem 3.10 in TDG24]

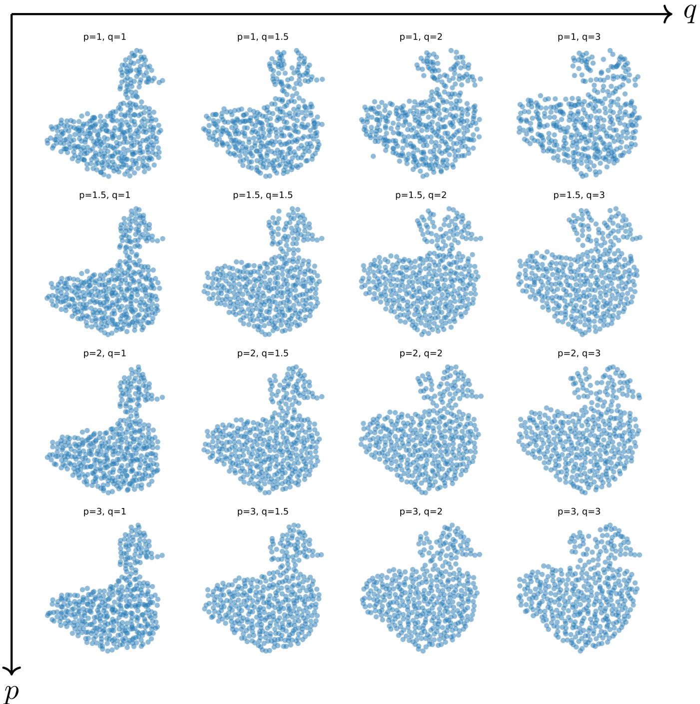
If μ is a subsequential limit of (μ_t) then $\mu \in G(\mu)$.

Application: Barycentres for p-norms with power q

Cost: $c(x, y) = \|x - y\|_p^q$.



Application: Barycentres for p-norms with power q



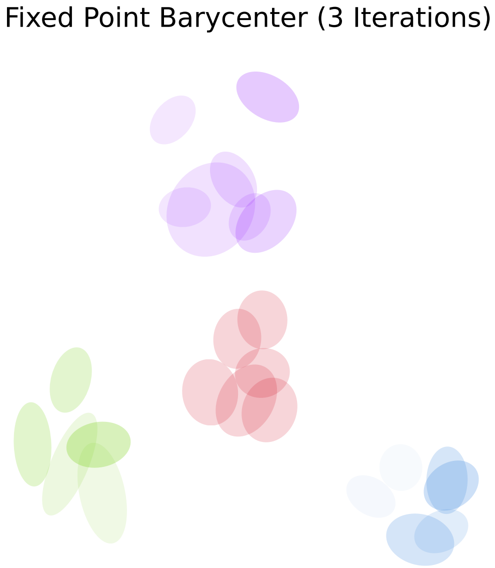
Application: GMM Barycentres

`ot.gmm.gmm_barycenter_fixed_point`

$$\mu = \sum_{i=1}^{n} a_i \delta_{\mathcal{N}(m_i, S_i)},$$

$$\nu_k = \sum_{j=1}^{n_k} b_k \delta_{\mathcal{N}(m_{k,j}, S_{k,j})},$$

$$V(\mu) = \sum_{k=1}^{K} \lambda_k \mathcal{T}_{\mathrm{W}_2^2}(\mu, \nu_k).$$



https://pythonot.github.io/auto_examples/barycenters/plot_gmm_barycenter.html

Conclusion and Outlook

Future Directions

- Flows in the GMM Space and Mixture-Wasserstein Geometry
  ↳ Infinite mixtures?
  ↳ Continuous-time and discrete-time dynamics?

Conclusion and Outlook

Future Directions

- Flows in the GMM Space and Mixture-Wasserstein Geometry
  ↳ Infinite mixtures?
  ↳ Continuous-time and discrete-time dynamics?

- RKHS representations of OT Maps
  ↳ RKHS representation of gradients of convex functions?
  ↳ A constrained map problem in RKHS?

Conclusion and Outlook

Future Directions

- Flows in the GMM Space and Mixture-Wasserstein Geometry
  ↳ Infinite mixtures?
  ↳ Continuous-time and discrete-time dynamics?

- RKHS representations of OT Maps
  ↳ RKHS representation of gradients of convex functions?
  ↳ A constrained map problem in RKHS?

- A (formally) Riemannian framework for OT barycentres
  ↳ See the FP algo as a “gradient descent”?
  ↳ Convergence rates?
  ↳ Guarantees for heuristic variants?

Conclusion and Outlook

Future Directions

- Flows in the GMM Space and Mixture-Wasserstein Geometry
  ↳ Infinite mixtures?
  ↳ Continuous-time and discrete-time dynamics?

- RKHS representations of OT Maps
  ↳ RKHS representation of gradients of convex functions?
  ↳ A constrained map problem in RKHS?

- A (formally) Riemannian framework for OT barycentres
  ↳ See the FP algo as a "gradient descent"?
  ↳ Convergence rates?
  ↳ Guarantees for heuristic variants?


Thanks!

Bibliography

[AC11] Martial Agueh and Guillaume Carlier. "Barycenters in the Wasserstein Space". In: *SIAM Journal on Mathematical Analysis* 43.2 (2011), pp. 904–924. eprint: https://doi.org/10.1137/100805741 (cit. on p. 49).

[Álv+16] Pedro C Álvarez-Esteban, E Del Barrio, JA Cuesta-Albertos, and C Matrán. "A fixed-point approach to barycenters in Wasserstein space". In: *Journal of Mathematical Analysis and Applications* 441.2 (2016), pp. 744–762 (cit. on pp. 9, 50).

[BT97] Dimitris Bertsimas and John N Tsitsiklis. *Introduction to linear optimization*. Vol. 6. Athena scientific Belmont, MA, 1997 (cit. on pp. 38–42).

[Boï+25] Samuel Boïté, Eloi Tanguy, Julie Delon, Agnès Desolneux, and Rémi Flamary. *Differentiable Expectation-Maximisation and Applications to Gaussian Mixture Model Optimal Transport*. 2025. arXiv: 2509.02109 [cs.LG] (cit. on pp. 7, 11–13).

[CD14] Marco Cuturi and Arnaud Doucet. "Fast Computation of Wasserstein Barycenters". In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by Eric P. Xing and Tony Jebara. Vol. 32. Proceedings of Machine Learning Research. Beijing, China: PMLR, June 2014, pp. 685–693 (cit. on p. 9).

[DD20] Julie Delon and Agnes Desolneux. "A Wasserstein-type distance in the space of Gaussian mixture models". In: *SIAM Journal on Imaging Sciences* 13.2 (2020), pp. 936–970 (cit. on pp. 12, 22, 23).

[DGS21] Julie Delon, Nathaël Gozlan, and Alexandre Saint-Dizier. *Generalized Wasserstein barycenters between probability measures living on different subspaces*. 2021 (cit. on pp. 9, 12).

[Fla+21] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. "POT: Python Optimal Transport". In: *Journal of Machine Learning Research* 22.78 (2021), pp. 1–8 (cit. on p. 12).

[Kan42] Leonid V Kantorovich. "On the translocation of masses". In: *Dokl. Akad. Nauk. USSR (NS)*. Vol. 37. 1942, pp. 199–201 (cit. on p. 3).

[Liu+24] Xinran Liu, Rocio Diaz Martin, Yikun Bai, Ashkan Shahbazi, Matthew Thorpe, Akram Aldroubi, and Soheil Kolouri. "Expected Sliced Transport Plans". In: *The Thirteenth International Conference on Learning Representations*. 2024 (cit. on pp. 29–31).

[Mah+23] Guillaume Mahey, Laetitia Chapel, Gilles Gasso, Clément Bonet, and Nicolas Courty. "Fast Optimal Transport through Sliced Generalized Wasserstein Geodesics". In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine. Vol. 36. Curran Associates, Inc., 2023, pp. 35350–35385 (cit. on pp. 29–31).

[Mon81] Gaspard Monge. "Mémoire sur la théorie des déblais et des remblais". In: *Mem. Math. Phys. Acad. Royale Sci.* (1781), pp. 666–704 (cit. on p. 2).

[PdC20] François-Pierre Paty, Alexandre d'Aspremont, and Marco Cuturi. "Regularity as regularization: Smooth and strongly convex brenier potentials in optimal transport". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 1222–1232 (cit. on p. 12).

[Rab+12] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. "Wasserstein barycenter and its application to texture mixing". In: *Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3*. Springer. 2012, pp. 435–446 (cit. on pp. 29–31).

[Sis+25] Keanu Sisouk, Eloi Tanguy, Julie Delon, and Julien Tierny. *Robust Barycenters of Persistence Diagrams*. 2025. arXiv: 2509.14904 [cs.LG] (cit. on p. 11).

[Tan23] Eloi Tanguy. "Convergence of SGD for Training Neural Networks with Sliced Wasserstein Losses". In: *Transactions on Machine Learning Research* (2023) (cit. on pp. 7, 11).

[Tan25] Eloi Tanguy. *Explicit Universal and Approximate-Universal Kernels on Compact Metric Spaces*. 2025. arXiv: 2506.03661 [math.FA] (cit. on pp. 10, 11).

[TCD25] Eloi Tanguy, Laetitia Chapel, and Julie Delon. *Sliced Optimal Transport Plans*. 2025. arXiv: 2508.01243 [math.OC] (cit. on pp. 8, 11, 12, 28).

[TDG24] Eloi Tanguy, Julie Delon, and Nathaël Gozlan. *Computing Barycentres of Measures for Generic Transport Costs*. 2024. arXiv: 2501.04016 [math.NA] (cit. on pp. 9, 11, 12, 48, 56–58).

[TDG25] Eloi Tanguy, Julie Delon, and Nathaël Gozlan. "Un algorithme de point fixe pour calculer des barycentres robustes entre mesures". In: *GRETSI*. Strasbourg, France, Aug. 2025 (cit. on p. 9).

[TFD24a] Eloi Tanguy, Rémi Flamary, and Julie Delon. "Properties of Discrete Sliced Wasserstein Losses". In: *Mathematics of Computation* (June 2024) (cit. on pp. 7, 11).

[TFD24b] Eloi Tanguy, Rémi Flamary, and Julie Delon. "Reconstructing discrete measures from projections. Consequences on the empirical Sliced Wasserstein Distance". In: *Comptes Rendus. Mathématique* 362 (2024), pp. 1121–1129 (cit. on pp. 7, 11).

[TDD25] Tanguy, Eloi, Desolneux, Agnès, and Delon, Julie. "Constrained Approximate Optimal Transport Maps". In: *ESAIM: COCV* 31 (2025), p. 70 (cit. on pp. 8, 11, 12).

EM Formulas

$$F(\theta, X) = (w', m', \Sigma') :$$

$$\gamma_{ik}(\theta) = \frac{w_k g_{m_k,\Sigma_k}(x_i)}{\sum_{\ell=1}^{K} w_\ell g_{m_\ell,\Sigma_\ell}(x_i)},$$

$$w'_k = \frac{1}{n}\sum_{i=1}^{n} \gamma_{ik}(\theta) ,$$

$$m'_k = \frac{\sum_{i=1}^{n} \gamma_{ik}(\theta) x_i}{\sum_{j=1}^{n} \gamma_{jk}(\theta)} ,$$

$$\Sigma'_k = \frac{\sum_{i=1}^{n} \gamma_{ik}(\theta)(x_i - m'_k)(x_i - m'_k)^\top}{\sum_{j=1}^{n} \gamma_{jk}(\theta)} .$$

Application: Colour Transfer

Gradient Descent on $X \in (\mathbb{R}^3)^{h \times w} \longmapsto \mathrm{MW}_2^2\left(\mu\left(F_X^T(\theta)\right), \mu\left(F_Y^T(\theta_0)\right)\right)$.

Source

Result

Target

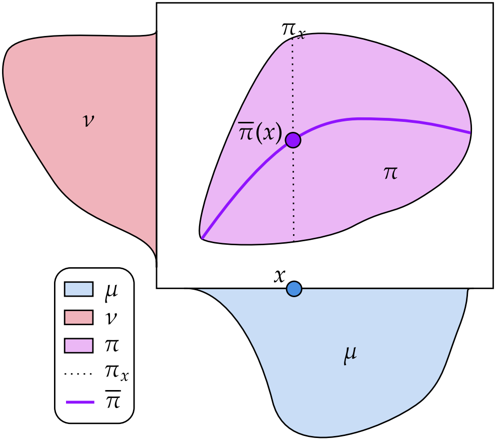
Balanced result

Unbalanced result

Corrupted target

Barycentric Projections

Replace a coupling π with a map $\overline{\pi}$.



$$\overline{\pi}(x) = \int y \mathrm{d}\pi_x(y).$$

$$\overline{\pi}(x) = \mathbb{E}_{(X,Y)\sim\pi}[Y|X=x].$$

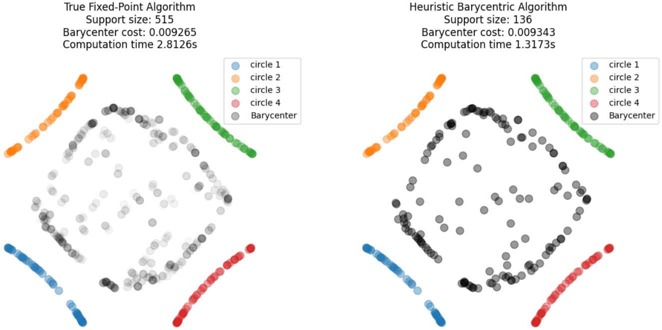
$$\overline{\pi} = \underset{f\in L^2(\mu)}{\mathrm{argmin}} \int \|f(x)-y\|_2^2 \mathrm{d}\pi(x,y).$$

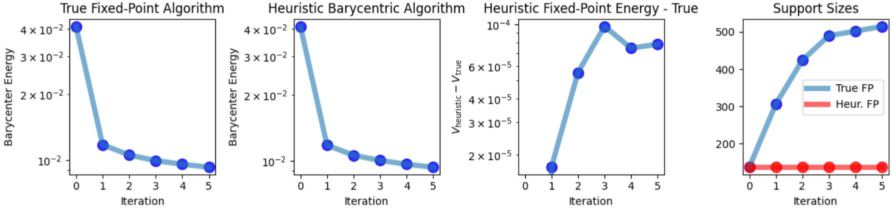
$$H(\mu) = \left\{ B(\overline{\pi_1}, \cdots, \overline{\pi_K})\#\mu,\ \pi_k \in \Pi^*_{c_k}(\mu, \nu_k) \right\}.$$

No guarantees... yet?

Illustration: $c_k(x, y) = \|P_{\text{circle}\#k}(x) - y\|_2^2$

`ot.lp.generalized_free_support_barycenter`





https://pythonot.github.io/auto_examples/barycenters/plot_free_support_barycenter_generic_cost.html