

Efficient Computation of approximate Kantorovich-Wasserstein distance for large spatial maps

Stefano Gualandi^(a), Fabio Ricciato^(b)

(a) Università di Pavia, Mathematics Department

(b) European Commission, DG Eurostat, Unit A5, Methodology,
Innovation in Official Statistics

email: stefano.gualandi@unipv.it
mastodon: [@famo2spaghi@sigmoid.social](https://mastodon.social/@famo2spaghi)

Package SKWD: <https://github.com/eurostat/Spatial-KWD>
R wrapper@CRAN: <https://cran.r-project.org/web/packages/SpatialKWD>
Python Wrapper: <https://pypi.org/project/Spatial-KWD/>

Eurostat 2022, 14-12-2022

Outline

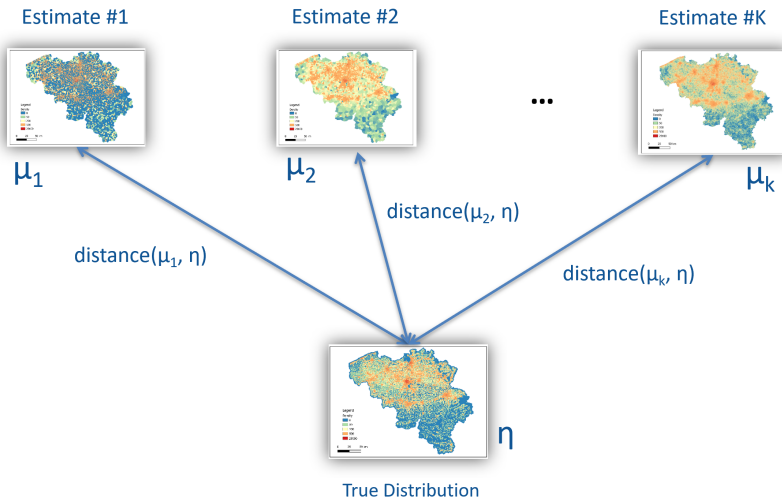
1 Spatial Maps and Optimal Transport

2 The Spatial-KWD Package

3 StatBel use case

4 Wasserstein Barycenters

Motivating example



Computing similarity (distances) between kD -Histograms

In general, we have two classes of similarity Measures for histograms:

- ① **Bin-By-Bin Dissimilarity Measures:** Minkowski-Form, Histogram Intersection, Kullback-Leibler Divergence, Jeffrey Divergence, χ^2 -Statistics.
- ② **Cross-Bin Dissimilarity Measures:** Quadratic-form distance, Match distance, **Earth Mover Distance (EMD)**.

SEE: Rubner, Tomasi, Guibas, L.J., 2000. The earth mover's distance as a metric for image retrieval. International journal of computer vision, 40(2), pp.99-121.

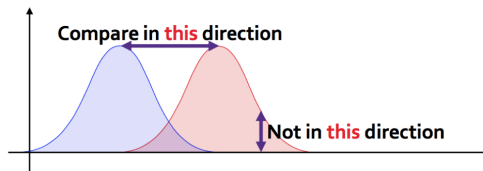
Computing similarity (distances) between kD -Histograms

In general, we have two classes of similarity Measures for histograms:

- ① **Bin-By-Bin Dissimilarity Measures:** Minkowski-Form, Histogram Intersection, Kullback-Leibler Divergence, Jeffrey Divergence, χ^2 -Statistics.
- ② **Cross-Bin Dissimilarity Measures:** Quadratic-form distance, Match distance, **Earth Mover Distance (EMD)**.

SEE: Rubner, Tomasi, Guibas, L.J., 2000. The earth mover's distance as a metric for image retrieval. International journal of computer vision, 40(2), pp.99-121.

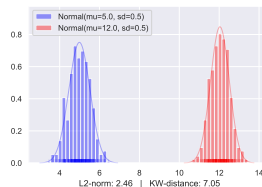
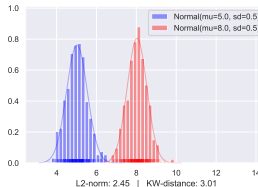
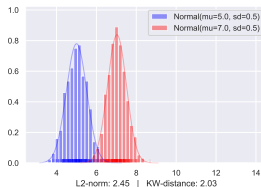
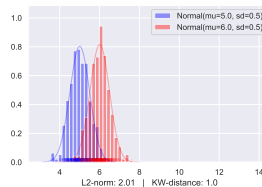
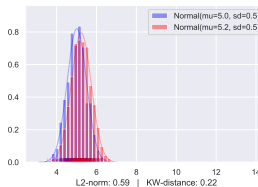
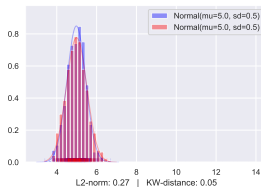
The Earth Mover Distance captures an “horizontal” notion of distance:



Credits: <https://jeremykun.com/2018/03/05/earthmover-distance/>

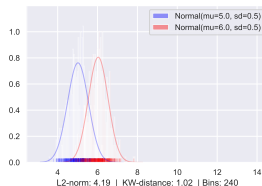
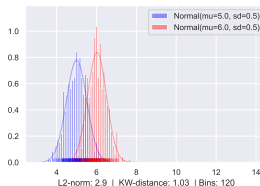
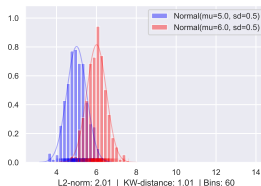
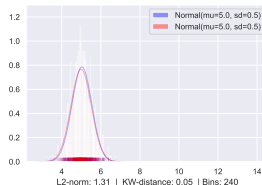
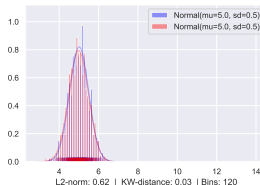
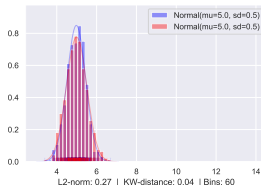
Computing similarity (distances) between 1D-Histograms

Let us take n samples from two Normal distributions: **BLU** and **RED**
Initially, they have the same mean and std. Then, we move the **RED** away.
How can we measure the similarity of the two distributions?



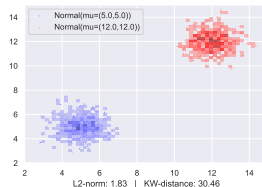
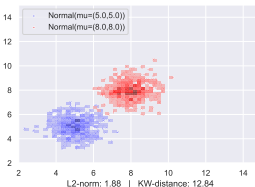
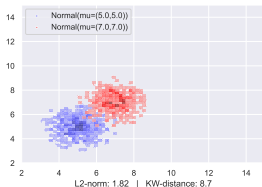
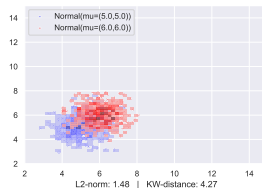
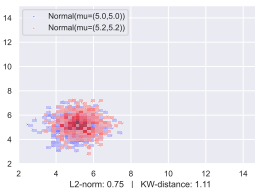
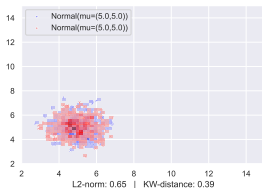
Computing similarity (distances) between 1D-Histograms

Let us take n samples from two Normal distributions: **BLU** and **RED**
 We move the **RED** away and we change the **number of bins**.
How can we measure the similarity of the two distributions?



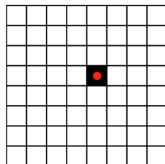
Earth Mover Distance: 2D-Histograms and Spatial Maps

Let us take n samples from **two 2D Normal distributions: BLUE and RED**. Initially, they have the same mean and std. Then, we move the **RED** away. **How can we measure the similarity of the two distributions?**

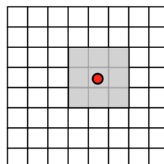


Issues with f -divergence functions for spatial maps

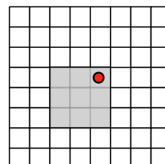
O



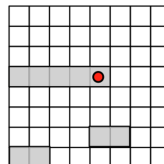
A



B

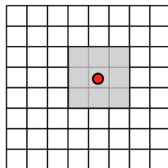


C

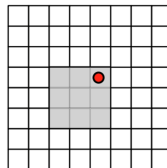


Issues with f -divergence functions for spatial maps

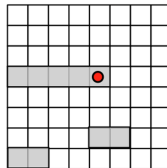
A



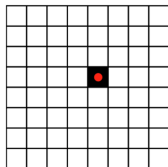
B



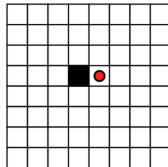
C



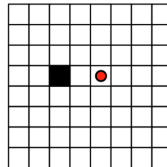
O



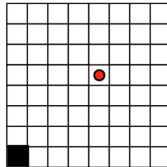
E



F



G



This talk

KEY POINT: We need a similarity measure or, that is a **metric function** to compare two **SPATIAL MAPS**.

This talk

KEY POINT: We need a similarity measure or, that is a **metric function** to compare two **SPATIAL MAPS**.

OUR PROPOSAL: to use the **Kantorovich Wasserstein Distance (KWD)** to measure the similarity between **2D-Histograms** and **SPATIAL MAPS**.

This talk

KEY POINT: We need a similarity measure or, that is a **metric function** to compare two **SPATIAL MAPS**.

OUR PROPOSAL: to use the **Kantorovich Wasserstein Distance (KWD)** to measure the similarity between **2D-Histograms** and **SPATIAL MAPS**.

OUR SELLING: We have an efficient implementation with two user-friendly wrappers in R and Python to a C++ core implementation.

Bassetti, Gualandi, Veneroni. On the Computation of Kantorovich–Wasserstein Distances Between Two-Dimensional Histograms by Uncapacitated Minimum Cost Flows.
SIAM Journal on Optimization 30 (3), 2441-2469

Optimal Transport: From Monge to ...

Optimal Transport originates from the work of Monge on finding the **minimum cost plan** for transporting a **distribution of *molécules*** from its origins to a given **target distribution** (1781).

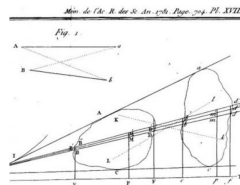


M É M O I R E S U R L A T H É O R I E D E S D É B L A I S E T D E S R E M B L A I S.

Par M. M O N G E.

Lorsqu'on doit transporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit transporter, & le nom de *Remblai* à l'espace qu'elles doivent occuper après le transport.

Le prix du transport d'une molécule étant, toutes choses d'ailleurs égales, proportionnel à son poids & à l'espace qu'on lui fait parcourir, & par conséquent le prix du transport total devant être proportionnel à la somme des produits des molécules multipliées chacune par l'espace parcouru, il s'en suit que le déblai & le remblai étant donnés de figure & de position, il n'est pas indifférent que telle molécule du déblai soit transportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine distribution à faire des molécules du premier dans le second, d'après laquelle la somme de ces produits fera la moindre possible, & le prix du transport total fera un minimum.



Optimal Transport: From Monge to ...

Optimal Transport originates from the work of Monge on finding the **minimum cost plan** for transporting a **distribution of *molécules*** from its origins to a given **target distribution** (1781).

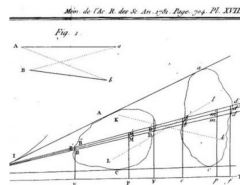


M É M O I R E
S U R L A
T H É O R I E D E S D É B L A I S
E T D E S R E M B L A I S.

Par M. M O N G E.

Lorsqu'on doit transporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit transporter, & le nom de *Remblai* à l'espace qu'elles doivent occuper après le transport.

Le prix du transport d'une molécule étant, toutes choses d'ailleurs égales, proportionnel à son poids & à l'espace qu'on lui fait parcourir, & par conséquent le prix du transport total devant être proportionnel à la somme des produits des molécules multipliées chacune par l'espace parcouru, il s'en suit que le déblai & le remblai étant donnés de figure & de position, il n'est pas indifférent que telle molécule du déblai soit transportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine distribution à faire des molécules du premier dans le second, d'après laquelle la somme de ces produits sera la moindre possible, & le prix du transport total fera un *minimum*.



In Combinatorial Optimization terms, the original Monge's problem is an **Assignment Problem**, which can be solved in strongly polynomial time with flow algorithms: $O(nm + n^2 \log(n))$ on general graphs and in $O(n^3)$ on bipartite complete graphs.

Ahuja, Magnanti, Orlin. Network flows: Theory, Algorithms, and Applications, 1988

Optimal Transport: From Monge to Kantorovich

An important contribution to Optimal Transport is due to Kantorovich, *Mathematical methods in the organization and planning of production*, (1939).

Apart from the Duality theory of Linear Programming, Kantorovich is the father of the **Transport Metric**, a.k.a. the **Wasserstein Metric**

Long History of the Monge-Kantorovich Transportation Problem

(Marking the centennial of L.V. Kantorovich's birth!)

A. M. Vershik



Figure 1. L.V. Kantorovich in his youth.

Leonid Vital'evich Kantorovich (1912–1986) was one of the great mathematicians and economists of the twentieth century.

In 2012, the centenary of his birth was marked in St. Petersburg. Short histories were presented describing some of the main parts of his legacy, which continue in importance today: *duality in linear programming*, the so-called “*Monge-Kantorovich transportation problem*,” and the “*Kantorovich metric*.” It is true that 2012 was also the 70th anniversary of the publication of his historic paper on *the transport metric*. The present article offers a somewhat expanded version of my talk on that occasion.

L. V. Kantorovich the Person

We remember Leonid Vital'evich Kantorovich for his massive contributions to foundations of mathematics, computational mathematics, and other areas, and in particular as one of the founders of mathematical economics. He began as a child prodigy, entering Leningrad University *at the age of 14*. His first paper on descriptive set theory, which

reception was quite different. The ideas of Kantorovich, and Kantorovich himself, were not appreciated in Soviet Russia. For a long time—until the end of the 1990s—his ideas on mathematical economics were considered in official circles as anti-Marxist, consequently it was prohibited and even dangerous to study and develop them. Such a story, carried much further, is well-known in Soviet biology (“*perestroika*”). So, between 1947 and the late 1950s Kantorovich never mentioned his ideas about mathematical economics and linear programming. My colleagues and I heard his lectures on functional analysis (later published as a book with G. P. Akhiezer) but knew nothing of his economically related work. He lectured openly on the subject only after the beginning of Khrushchev’s “*otvet*” (“*thaw*”), the liberalization of 1955–1956.

L. V. Kantorovich (LV hereinafter) has left a rich record of the defence of scientific truth. A majority of the Soviet economists of the generation of the 1940s and 1970s were pupils of LV, and many mathematicians (including the author) considered themselves his pupils.



Optimal Transport: From Monge to Kantorovich

An important contribution to Optimal Transport is due to Kantorovich, *Mathematical methods in the organization and planning of production*, (1939). Apart from the Duality theory of Linear Programming, Kantorovich is the father of the **Transport Metric**, a.k.a. the **Wasserstein Metric**

Long History of the Monge-Kantorovich Transportation Problem

(Marking the centennial of L.V. Kantorovich's birth!)

A. M. Vershik



Figure 1. L.V. Kantorovich in his youth.

Leonid Vital'evich Kantorovich (1912–1986) was one of the great mathematicians and economists of the twentieth century.

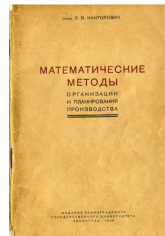
In 2012, the centenary of his birth was marked in St. Petersburg. Short histories were presented describing some of the main parts of his legacy, which continue in importance today: *duality in linear programming*, the so-called “*Monge-Kantorovich transportation problem*,” and the “*Kantorovich metric*.” It is that 2012 was also the 70th anniversary of the publication of his historic paper on *the transport metric*. The present article offers a somewhat expanded version of my talk on that occasion.

L. V. Kantorovich the Person

We remember Leonid Vital'evich Kantorovich for his massive contributions to foundations of mathematics, computational mathematics, and other areas, and in particular as one of the founders of mathematical economics. He began as a child prodigy, entering Leningrad University *at the age of 14*. His first paper on descriptive set theory, which

reception was quite different. The ideas of Kantorovich, and Kantorovich himself, were not appreciated in Soviet Russia. For a long time—until the end of the 1990s—his ideas on mathematical economics were considered in official circles as anti-Marxist, consequently it was prohibited and even dangerous to study and develop them. Such a story, carried much further, is well-known in Soviet biology (“*perestroika*”). So, between 1947 and the late 1950s Kantorovich never mentioned his ideas about mathematical economics and linear programming. My colleagues and I heard his lectures on functional analysis (later published as a book with G. P. Akilov) but knew nothing of his economically related work. He lectured openly on the subject only after the beginning of Khrushchev’s “*otvet*” (“*thaw*”), the liberalization of 1953–1956.

L. V. Kantorovich (LV hereafter) has left a rich record of the defence of scientific truth. A majority of the Soviet economists of the generation of the 1940s and 1950s were pupils of LV, and many mathematicians (including the author) considered themselves his pupils.

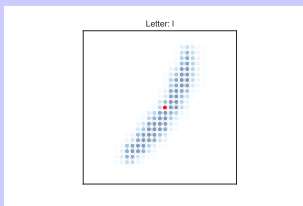
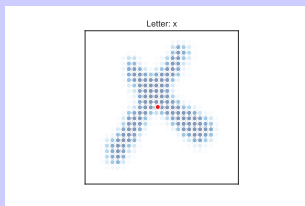


The Kantorovich reformulation of Monge’s problem gives a **Transportation Problem**: the complexity raises to $O(n \log(n)(m + n \log(n)))$ on general graphs, and to $O(n^3 \log(n))$ for bipartite complete graphs.

Orlin. A faster strongly polynomial minimum cost flow algorithm.
Operations research 41(2) 338-350 (1993)

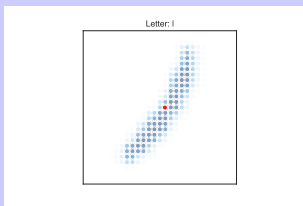
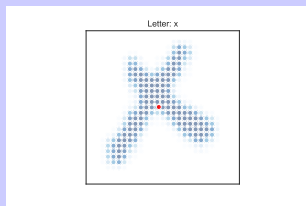
Discrete Probability measures

Example: **Spatial maps** are discrete measures: $x_i = (x_{i1}, x_{i2})$ is the tile/pixel position, and μ_i is the tile/pixel intensity at x_i . Two toy examples:



Discrete Probability measures

Example: **Spatial maps** are discrete measures: $x_i = (x_{i1}, x_{i2})$ is the tile/pixel position, and μ_i is the tile/pixel intensity at x_i . Two toy examples:

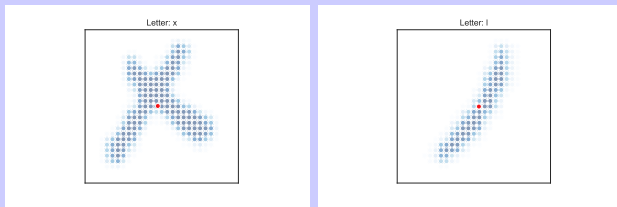


We define the Dirac δ -measure centered at $x \in X \subseteq \mathbb{R}^q$ via

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \subseteq X \\ 0 & \text{if } x \notin A \subseteq X. \end{cases}$$

Discrete Probability measures

Example: **Spatial maps** are discrete measures: $x_i = (x_{i1}, x_{i2})$ is the tile/pixel position, and μ_i is the tile/pixel intensity at x_i . Two toy examples:



We define the Dirac δ -measure centered at $x \in X \subseteq \mathbb{R}^q$ via

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \subseteq X \\ 0 & \text{if } x \notin A \subseteq X. \end{cases}$$

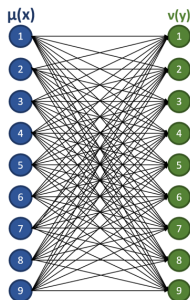
Let μ and ν be two discrete probability measures defined on $X, Y \subseteq \mathbb{R}^q$

$$\mu(A) = \sum_{i=1}^m \mu_i \delta_{x_i}(A) \quad \text{and} \quad \nu(A) = \sum_{j=1}^n \nu_j \delta_{y_j}(A)$$

where $\sum_{i=1}^m \mu_i = 1$ and $\sum_{j=1}^n \nu_j = 1$.

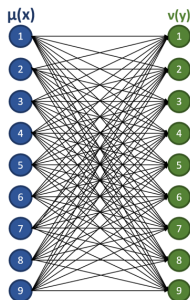
Computing the KWD between two spatial distributions

Given two empirical distributions μ and ν defined over a space X , we can define a complete bipartite graph:



Computing the KWD between two spatial distributions

Given two empirical distributions μ and ν defined over a space X , we can define a complete bipartite graph:



Using the bipartite graph

$B = (V \cup W, E)$, we can write the LP *bipartite* model:

$$\min \sum_{ij \in E} c_{ij} \pi_{ij} \quad (1)$$

$$\text{s.t.} \quad \sum_{ij \in E} \pi_{ij} = \mu_i, \quad \forall i \in V, \quad (2)$$

$$\sum_{ij \in E} \pi_{ij} = \nu_j, \quad \forall j \in W, \quad (3)$$

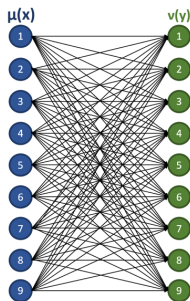
$$\pi_{ij} \geq 0, \quad \forall \{i, j\} \in E. \quad (4)$$

where the cost are defined as:

$$c(i, j) := d(x_i, x_j) = \|x_i - x_j\|_2. \quad (5)$$

Computing the KWD between two spatial distributions

Given two empirical distributions μ and ν defined over a space X , we can define a complete bipartite graph:



Using the bipartite graph

$B = (V \cup W, E)$, we can write the LP *bipartite* model:

$$\min \sum_{ij \in E} c_{ij} \pi_{ij} \quad (1)$$

$$\text{s.t.} \quad \sum_{ij \in E} \pi_{ij} = \mu_i, \quad \forall i \in V, \quad (2)$$

$$\sum_{ij \in E} \pi_{ij} = \nu_j, \quad \forall j \in W, \quad (3)$$

$$\pi_{ij} \geq 0, \quad \forall \{i, j\} \in E. \quad (4)$$

where the cost are defined as:

$$c(i, j) := d(x_i, x_j) = \|x_i - x_j\|_2. \quad (5)$$

REMARK

We need to solve an optimization problem to compute the KW-distance.

Optimal Transport: State-of-the-art (?)

- We compute Wasserstein Distance using our **Network Simplex algorithm**.

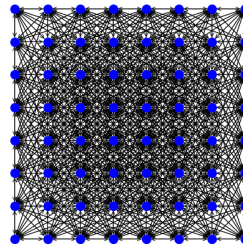
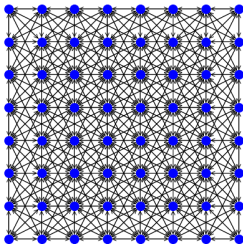
Optimal Transport: State-of-the-art (?)

- We compute Wasserstein Distance using our **Network Simplex algorithm**.
- Modern Computational Optimal Transport for Machine Learning has focused on **entropic regularized costs** ($c_{ij}x_{ij} + \epsilon x_{ij} \log(x_{ij})$), which add a regularization parameter depending on a parameter ϵ . For $\epsilon \rightarrow 0$, we recover the standard Optimal Transport, (in theory).
- The **Sinkhorn algorithm** is highly parallelizable,

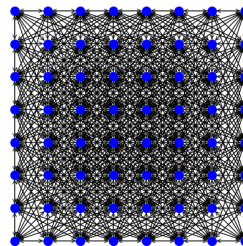
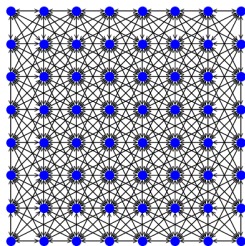
Optimal Transport: State-of-the-art (?)

- We compute Wasserstein Distance using our **Network Simplex algorithm**.
- Modern Computational Optimal Transport for Machine Learning has focused on **entropic regularized costs** ($c_{ij}x_{ij} + \epsilon x_{ij} \log(x_{ij})$), which add a regularization parameter depending on a parameter ϵ . For $\epsilon \rightarrow 0$, we recover the standard Optimal Transport, (in theory).
- The **Sinkhorn algorithm** is highly parallelizable, but suffers from stability numerical issues.

Flow Networks with coprime index $L = 2$ (left) and $L = 3$ (right).



Flow Networks with **coprime index** $L = 2$ (left) and $L = 3$ (right).



Using the flow network $G = (N, A, \mathbf{c}, \mathbf{b})$, we can formulate the following Linear Program, which is equivalent to P_B :

$$P_A : \quad \min \quad \sum_{ij \in A} c_{ij} f_{ij} \quad (6)$$

$$\text{s.t.} \quad \sum_{ij \in \delta^+(i)} f_{ij} - \sum_{ji \in \delta^-(i)} f_{ji} = b_i, \quad \forall i \in N, \quad (7)$$

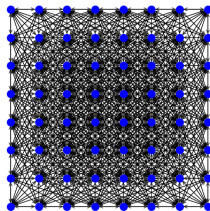
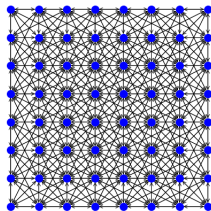
$$f_{ij} \geq 0, \quad \forall \{i, j\} \in A. \quad (8)$$

The impact of the coprime index L

REMARK: The impact of the parameter $L=3$

Maximum percentage error from the optimal value of P_{A_L} with respect to the optimal value obtained with the bipartite graphs:

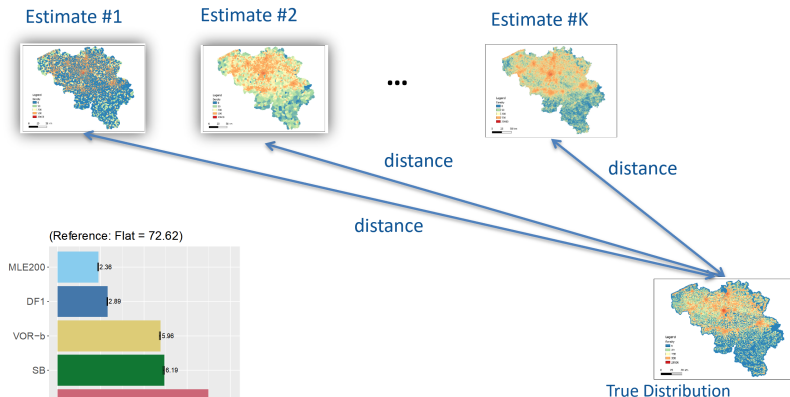
L	1	2	3	5	10	15
max error	7.61%	2.68%	1.29%	0.486%	0.124%	0.055%
$ A_L $	$O(8n)$	$O(16n)$	$O(32n)$	$O(80n)$	$O(256n)$	$O(576n)$



Outline

- 1 Spatial Maps and Optimal Transport
- 2 The Spatial-KWD Package
- 3 StatBel use case
- 4 Wasserstein Barycenters

Desiderata: Comparing *easily* different estimates



```
compareOneToMany(Coordinates, Weights, L = 3, recode = TRUE,
  method = "approx", algorithm = "colgen",
  model = "mincostflow", verbosity = "silent",
  timelimit = 14400, opt_tolerance = 1e-06,
  unbalanced = FALSE, unbal_cost = 1e+09, convex = TRUE)
```

Main functions of the Spatial-KWD

REMARK: The four main functions implemented in Spatial-KWD are:

- ① `compareOneToOne(Coordinates, Weights, L=3)`
- ② `compareOneToMany(Coordinates, Weights, L = 3)`
- ③ `compareAll(Coordinates, Weights, L=3)`
- ④ `focusArea(Coordinates, Weights, x=15, y=15, radius=5)`

```

1 # Import SpatialKWD package (available on CRAN)
2 library(SpatialKWD)
3
4 # Random coordinates
5 N = 90
6 Xs <- as.integer(runif(N, 0, 31))
7 Ys <- as.integer(runif(N, 0, 31))
8 Coordinates <- matrix(c(Xs, Ys), ncol=2, nrow=N)
9
10 # Random weights
11 m <- 4
12 Weights <- matrix(runif(m*N, 0, 1), ncol=m)

```

Example: compareAll(Coordinates, Weights)

```

1  # Import SpatilKWD package (available on CRAN)
2  library(SpatialKWD)
3
4  # Random coordinates
5  N = 90
6  Xs <- as.integer(runif(N, 0, 31))
7  Ys <- as.integer(runif(N, 0, 31))
8  Coordinates <- matrix(c(Xs, Ys), ncol=2, nrow=N)
9
10 # Random weights
11 m <- 3
12 Weights <- matrix(runif(m*N, 0, 1), ncol=m)
13
14 # Compute distance
15 print("Compare all pairwise distances with an approximate
      algorithm:")
16 d <- compareAll(Coordinates, Weights, L=3)
17
18 cat("L: 3, runtime:", d$runtime, " distances:", "\n")
19 print(matrix(d$distance, ncol=3, nrow=3))

```

Main functions of the Spatial-KWD

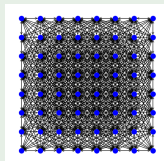
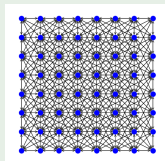
```

1 compareAll( Coordinates , Weights , L = 3, recode = TRUE,
2               method = "approx", algorithm = "colgen",
3               model="mincostflow", verbosity = "silent",
4               timelimit = 14400, opt_tolerance = 1e-06,
5               unbalanced = FALSE, unbal_cost = 1e+09,
6               convex = TRUE)

```

REMARK: Understanding the impact of the parameter L=3

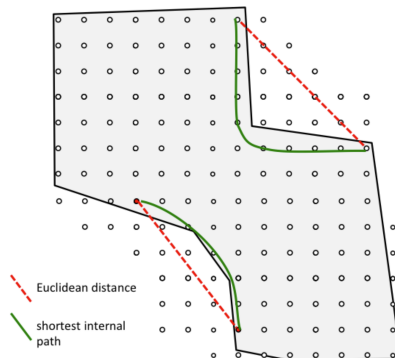
L	1	2	3	5	10	15
max error	7.61%	2.68%	1.29%	0.486%	0.124%	0.055%
$ A_L $	$O(8n)$	$O(16n)$	$O(32n)$	$O(80n)$	$O(256n)$	$O(576n)$



List of parameters of the **Spatial-KWD** library

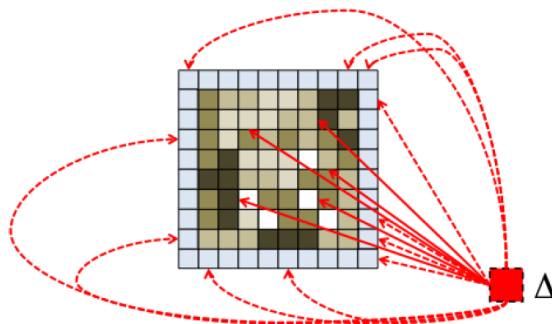
Name	Type	Possible Values	Default Value
KWD_PAR_METHOD	string	KWD_VAL_EXACT KWD_VAL_APPROX	KWD_VAL_APPROX
KWD_PAR_MODEL	string	KWD_VAL_BIPARTITE KWD_VAL_MINCOSTFLOW	KWD_VAL_MINCOSTFLOW
KWD_PAR_ALGORITHM	string	KWD_VAL_FULLMODEL KWD_VAL_COLGEN	KWD_VAL_COLGEN
KWD_PAR_RECODE	string	KWD_VAL_TRUE KWD_VAL_FALSE	KWD_VAL_TRUE
KWD_PAR_CONVEXHULL	string	KWD_VAL_TRUE KWD_VAL_FALSE	KWD_VAL_TRUE
KWD_PAR_UNBALANCED	string	KWD_VAL_FALSE KWD_VAL_ALLBIN KWD_VAL_BORDERBIN KWD_VAL_CUSTOMBIN	KWD_VAL_FALSE
KWD_PAR_AREA	string	KWD_VAL_CIRCULAR KWD_VAL_SQUARED	KWD_VAL_CIRCULAR
KWD_PAR_VERBOSITY	string	KWD_VAL_INFO KWD_VAL_DEBUG KWD_VAL_SILENT	KWD_VAL_INFO
KWD_PAR_UNBALANCED_COST	double	> 0	max distance
KWD_PAR_TIMELIMIT	double	> 0	max double
KWD_PAR_OPTTOLERANCE	double	$[10^{-9}, 10^{-1}]$	10^{-6}
KWD_PAR_COPRIMEINDEX	integer	≥ 1	3

Convex vs. Nonconvex regions



We support Euclidean distance versus shortest internal path, in case of nonconvex spatial maps, via a package parameter.

Focus Area Support via Unbalanced Optimal Transport



We have implemented different options for dealing with the artificial node v_0 : all tiles connected with the virtual node v_0 , only the tiles on the border, and all the tiles specified by the user of the software library.

Computational Evaluation 1/2

To assess the efficiency and the scalability of our solver we have run a number of experiments using synthetic data derived from the official data by **StatBel** (<https://statbel.fgov.be/>) at

[https://data.gov.be/en/dataset/
ea430cb8-2805-4c67-8392-9e2a47c1ef55](https://data.gov.be/en/dataset/ea430cb8-2805-4c67-8392-9e2a47c1ef55)

Computational Evaluation 1/2

To assess the efficiency and the scalability of our solver we have run a number of experiments using synthetic data derived from the official data by **StatBel** (<https://statbel.fgov.be/>) at

<https://data.gov.be/en/dataset/ea430cb8-2805-4c67-8392-9e2a47c1ef55>

resolution	nodes (tiles)	arcs	iterations	runtime in sec.
1000 × 1000	39 303	64 349	89 227	0.52
500 × 500	156 240	243 231	162 366	1.32
250 × 250	621 868	1 092 105	671 988	5.06
125 × 125	2 482 118	4 425 571	5 962 135	572.81

Comparing the runtime in seconds (on our desktop computer) of the capacitated network simplex algorithm for computing the distance within a focus area, using the coprime index $L = 3$.

Computational Evaluation 2/2

resolution	supports	nodes	L	arcs	normalized values		
					UB	runtime	iterations
1000×1000	32 140	39 302	2	54 728	1.000267	0.78	0.65
			3	60 169	1.000015	0.73	0.70
			5	64 706	1.000001	0.81	0.79
			10	68 213	1.000000	0.90	0.90
			15	69 621	1.000000	0.95	0.95
			20	70 304	1.000000	1.00	1.00

The UB and runtime ratios are computed using the results obtained $L = 20$.

Each row gives the average over 9 instances.

Computational Evaluation 2/2

resolution	supports	nodes	L	arcs	normalized values		
					UB	runtime	iterations
1000×1000	32 140	39 302	2	54 728	1.000267	0.78	0.65
			3	60 169	1.000015	0.73	0.70
			5	64 706	1.000001	0.81	0.79
			10	68 213	1.000000	0.90	0.90
			15	69 621	1.000000	0.95	0.95
			20	70 304	1.000000	1.00	1.00

The UB and runtime ratios are computed using the results obtained $L = 20$.

Each row gives the average over 9 instances.

platform	L	125 × 125	250 × 250	500 × 500	1000 × 1000
cluster	2	11 638.5	455.5	28.8	3.4
	3	12 211.2	533.5	29.7	3.1
	5	14 195.7	587.3	30.8	3.5
	10	-	610.9	33.1	3.9
desktop	2	13 602.5	411.8	19.1	2.9
	3	13 925.6	474.7	19.5	2.7
	5	19 296.9	525.5	20.2	2.4
	10	-	534.0	23.2	1.9

Runtime in seconds using the coprime index L versus tile resolutions.

Spatial Density Estimation

Our package is used to quantify and compare the **goodness of fit** for different estimated **density distributions**, obtained with different estimation methods, against the known ground truth density (that is known since the study is based on synthetically generated scenarios)

- The paper pre-print is available at

<https://arxiv.org/abs/2009.05410>

- The open-source notebook developed for this study by [Marco Ramljak](#) (who is presenting later in the afternoon session) is available at

https://r-ramljak.github.io/MNO_mobdensity/

Questions so far?

Package SKWD <https://github.com/eurostat/Spatial-KWD>

R wrapper@CRAN <https://cran.r-project.org/web/packages/SpatialKWD>

Python Wrapper <https://pypi.org/project/Spatial-KWD/>

Questions so far?

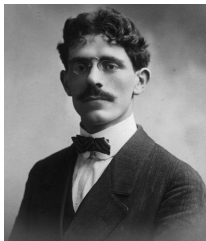
Package SKWD <https://github.com/eurostat/Spatial-KWD>

R wrapper@CRAN <https://cran.r-project.org/web/packages/SpatialKWD>

Python Wrapper <https://pypi.org/project/Spatial-KWD/>

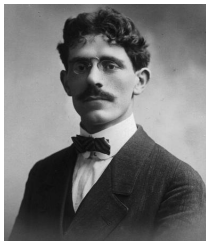


Statistical Inference: *Éléments typiques*



Éléments typiques. *La statistique a constamment besoin de donner une idée de ce que sont, en gros, les éléments d'une collection complexe, difficile à saisir dans son ensemble, en les assimilant à un seul élément choisi aussi ressemblant que possible à tous les éléments de la collection et qui, pour cette raison, soit appelé élément typique de cette collection.* *Maurice Fréchet, 1945.*

Statistical Inference: *Éléments typiques*



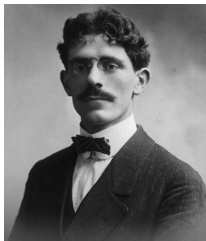
Éléments typiques. *La statistique a constamment besoin de donner une idée de ce que sont, en gros, les éléments d'une collection complexe, difficile à saisir dans son ensemble, en les assimilant à un seul élément choisi aussi ressemblant que possible à tous les éléments de la collection et qui, pour cette raison, soit appelé élément typique de cette collection.* *Maurice Fréchet, 1945.*

Example 1: Numbers in $\mathbb{R}$

Consider the collection of numbers $C = \{2, 3, 4, 6, 7, 12\}$.

QUESTION: Which is the *Élément typique* of C ?

Statistical Inference: *Éléments typiques*



Éléments typiques. *La statistique a constamment besoin de donner une idée de ce que sont, en gros, les éléments d'une collection complexe, difficile à saisir dans son ensemble, en les assimilant à un seul élément choisi aussi ressemblant que possible à tous les éléments de la collection et qui, pour cette raison, soit appelé élément typique de cette collection.* *Maurice Fréchet, 1945.*

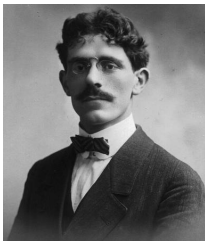
Example 1: Numbers in $\mathbb{R}$

Consider the collection of numbers $C = \{2, 3, 4, 6, 7, 12\}$.

QUESTION: Which is the *Élément typique* of C ?

Mean:

Statistical Inference: *Éléments typiques*



Éléments typiques. *La statistique a constamment besoin de donner une idée de ce que sont, en gros, les éléments d'une collection complexe, difficile à saisir dans son ensemble, en les assimilant à un seul élément choisi aussi ressemblant que possible à tous les éléments de la collection et qui, pour cette raison, soit appelé élément typique de cette collection.* *Maurice Fréchet, 1945.*

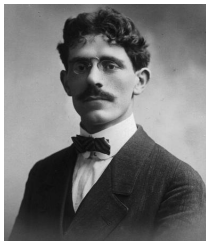
Example 1: Numbers in $\mathbb{R}$

Consider the collection of numbers $C = \{2, 3, 4, 6, 7, 12\}$.

QUESTION: Which is the *Élément typique* of C ?

Mean: $\rho^* = \arg \min_{\rho \in \mathbb{R}} \frac{1}{2} \sum_{c \in C} \|\rho - c\|_2^2$

Statistical Inference: *Éléments typiques*



Éléments typiques. *La statistique a constamment besoin de donner une idée de ce que sont, en gros, les éléments d'une collection complexe, difficile à saisir dans son ensemble, en les assimilant à un seul élément choisi aussi ressemblant que possible à tous les éléments de la collection et qui, pour cette raison, soit appelé élément typique de cette collection.* *Maurice Fréchet, 1945.*

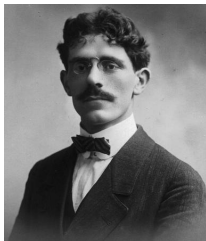
Example 1: Numbers in $\mathbb{R}$

Consider the collection of numbers $C = \{2, 3, 4, 6, 7, 12\}$.

QUESTION: Which is the *Élément typique* of C ?

Mean: $\rho^* = \arg \min_{\rho \in \mathbb{R}} \frac{1}{2} \sum_{c \in C} \|\rho - c\|_2^2 \rightarrow$

Statistical Inference: *Éléments typiques*



Éléments typiques. *La statistique a constamment besoin de donner une idée de ce que sont, en gros, les éléments d'une collection complexe, difficile à saisir dans son ensemble, en les assimilant à un seul élément choisi aussi ressemblant que possible à tous les éléments de la collection et qui, pour cette raison, soit appelé élément typique de cette collection.* *Maurice Fréchet, 1945.*

Example 1: Numbers in $\mathbb{R}$

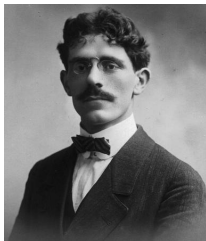
Consider the collection of numbers $C = \{2, 3, 4, 6, 7, 12\}$.

QUESTION: Which is the *Élément typique* of C ?

Mean: $\rho^* = \arg \min_{\rho \in \mathbb{R}} \frac{1}{2} \sum_{c \in C} \|\rho - c\|_2^2 \rightarrow \rho^* = \frac{\sum_{c \in C} c}{|C|} \approx 5.66$

Median:

Statistical Inference: *Éléments typiques*



Éléments typiques. *La statistique a constamment besoin de donner une idée de ce que sont, en gros, les éléments d'une collection complexe, difficile à saisir dans son ensemble, en les assimilant à un seul élément choisi aussi ressemblant que possible à tous les éléments de la collection et qui, pour cette raison, soit appelé élément typique de cette collection.* *Maurice Fréchet, 1945.*

Example 1: Numbers in $\mathbb{R}$

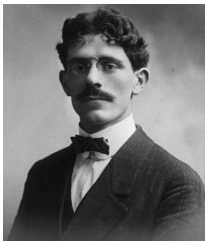
Consider the collection of numbers $C = \{2, 3, 4, 6, 7, 12\}$.

QUESTION: Which is the *Élément typique* of C ?

Mean: $\rho^* = \arg \min_{\rho \in \mathbb{R}} \frac{1}{2} \sum_{c \in C} \|\rho - c\|_2^2 \rightarrow \rho^* = \frac{\sum_{c \in C} c}{|C|} \approx 5.66$

Median: $\rho^* = \arg \min_{\rho \in \mathbb{R}} \frac{1}{2} \sum_{c \in C} \|\rho - c\|_2$

Statistical Inference: *Éléments typiques*



Éléments typiques. *La statistique a constamment besoin de donner une idée de ce que sont, en gros, les éléments d'une collection complexe, difficile à saisir dans son ensemble, en les assimilant à un seul élément choisi aussi ressemblant que possible à tous les éléments de la collection et qui, pour cette raison, soit appelé élément typique de cette collection.* *Maurice Fréchet, 1945.*

Example 1: Numbers in $\mathbb{R}$

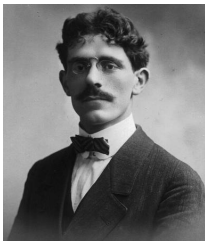
Consider the collection of numbers $C = \{2, 3, 4, 6, 7, 12\}$.

QUESTION: Which is the *Élément typique* of C ?

Mean: $\rho^* = \arg \min_{\rho \in \mathbb{R}} \frac{1}{2} \sum_{c \in C} \|\rho - c\|_2^2 \rightarrow \rho^* = \frac{\sum_{c \in C} c}{|C|} \approx 5.66$

Median: $\rho^* = \arg \min_{\rho \in \mathbb{R}} \frac{1}{2} \sum_{c \in C} \|\rho - c\|_2 \rightarrow$

Statistical Inference: *Éléments typiques*



Éléments typiques. *La statistique a constamment besoin de donner une idée de ce que sont, en gros, les éléments d'une collection complexe, difficile à saisir dans son ensemble, en les assimilant à un seul élément choisi aussi ressemblant que possible à tous les éléments de la collection et qui, pour cette raison, soit appelé élément typique de cette collection.* *Maurice Fréchet, 1945.*

Example 1: Numbers in $\mathbb{R}$

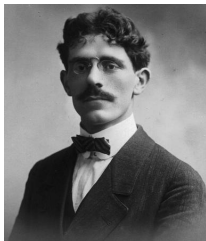
Consider the collection of numbers $C = \{2, 3, 4, 6, 7, 12\}$.

QUESTION: Which is the *Élément typique* of C ?

Mean: $\rho^* = \arg \min_{\rho \in \mathbb{R}} \frac{1}{2} \sum_{c \in C} \|\rho - c\|_2^2 \rightarrow \rho^* = \frac{\sum_{c \in C} c}{|C|} \approx 5.66$

Median: $\rho^* = \arg \min_{\rho \in \mathbb{R}} \frac{1}{2} \sum_{c \in C} \|\rho - c\|_2 \rightarrow \rho^* \in [4 \dots 6] \quad (\dots \text{many solutions!})$

Statistical Inference: *Éléments typiques*



Éléments typiques. *La statistique a constamment besoin de donner une idée de ce que sont, en gros, les éléments d'une collection complexe, difficile à saisir dans son ensemble, en les assimilant à un seul élément choisi aussi ressemblant que possible à tous les éléments de la collection et qui, pour cette raison, soit appelé élément typique de cette collection.* *Maurice Fréchet, 1945.*

Example 1: Numbers in $\mathbb{R}$

Consider the collection of numbers $C = \{2, 3, 4, 6, 7, 12\}$.

QUESTION: Which is the *Élément typique* of C ?

Mean: $\rho^* = \arg \min_{\rho \in \mathbb{R}} \frac{1}{2} \sum_{c \in C} \|\rho - c\|_2^2 \rightarrow \rho^* = \frac{\sum_{c \in C} c}{|C|} \approx 5.66$

Median: $\rho^* = \arg \min_{\rho \in \mathbb{R}} \frac{1}{2} \sum_{c \in C} \|\rho - c\|_2 \rightarrow \rho^* \in [4 \dots 6] \quad (\text{...many solutions!})$

QUESTION: How generalize *mean* and *median* to distributions in $\mathbb{R}^q$?

Mean and Medians of distributions

Example 2: Images (Double ellipses and MNIST digits)



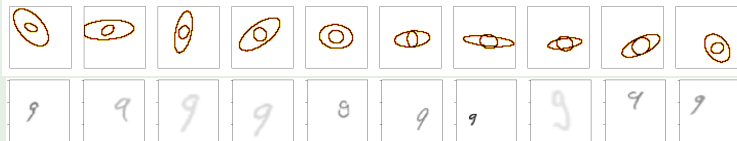
Mean and Medians of distributions

Example 2: Images (Double ellipses and MNIST digits)



Mean and Medians of distributions

Example 2: Images (Double ellipses and MNIST digits)



Definition 1 (Fréchet mean)

Let $\mu_1, \dots, \mu_k \in \mathcal{X}$ with $(\mathcal{X}, \text{distance})$ a metric space, $k \geq 2$, and $\lambda_1, \dots, \lambda_k \geq 0$, such that $\sum_{i=1}^k \lambda_i = 1$. For any $\alpha \geq 1$, a Fréchet mean of order α is any optimal solution to the following problem:

$$\rho^* = \arg \min \left\{ \sum_{i=1}^k \lambda_i \text{distance}^\alpha(\rho, \mu_i) \mid \rho \in \mathcal{X} \right\}$$

When $\alpha = 2$, we get a **(weighted) Fréchet barycenter**.

When $\alpha = 1$, we get a **(weighted) Fréchet median**.

Mean and Medians of distributions

Definition 2 (Wasserstein Means)

Let $\mathcal{P}_2(\mathbb{R}^q)$ be the space of probability measures on $\mathbb{R}^q$ with finite second moments equipped with a Wasserstein distance of order α .

Mean and Medians of distributions

Definition 2 (Wasserstein Means)

Let $\mathcal{P}_2(\mathbb{R}^q)$ be the space of probability measures on $\mathbb{R}^q$ with finite second moments equipped with a Wasserstein distance of order α . Given $k \geq 2$ probability measures $\mu_1, \dots, \mu_k \in \mathcal{P}_2(\mathbb{R}^q)$, a Wasserstein mean of $\mu_1, \dots, \mu_k$ is any optimal solution to

$$\rho^* = \arg \min \left\{ \sum_{i=1, \dots, k} \lambda_i W^\alpha(\rho, \mu_i) \mid \rho \in \mathcal{P}_2(\mathbb{R}^q) \right\}$$

where $\lambda_1, \dots, \lambda_k \geq 0$, such that $\sum_{i=1}^k \lambda_i = 1$.

When $\alpha = 2$, we get a **(weighted) Wasserstein barycenter**.

When $\alpha = 1$, we get a **(weighted) Wasserstein median**.

Mean and Medians of distributions

Definition 2 (Wasserstein Means)

Let $\mathcal{P}_2(\mathbb{R}^q)$ be the space of probability measures on $\mathbb{R}^q$ with finite second moments equipped with a Wasserstein distance of order α . Given $k \geq 2$ probability measures $\mu_1, \dots, \mu_k \in \mathcal{P}_2(\mathbb{R}^q)$, a Wasserstein mean of $\mu_1, \dots, \mu_k$ is any optimal solution to

$$\rho^* = \arg \min \left\{ \sum_{i=1, \dots, k} \lambda_i W^\alpha(\rho, \mu_i) \mid \rho \in \mathcal{P}_2(\mathbb{R}^q) \right\}$$

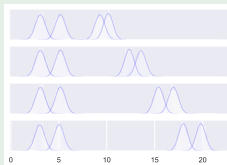
where $\lambda_1, \dots, \lambda_k \geq 0$, such that $\sum_{i=1}^k \lambda_i = 1$.

When $\alpha = 2$, we get a **(weighted) Wasserstein barycenter**.

When $\alpha = 1$, we get a **(weighted) Wasserstein median**.

Example 3: Distributions in $\mathcal{P}_2(\mathbb{R})$

<https://github.com/stegua/ot1d>



Mean and Medians of distributions

Definition 2 (Wasserstein Means)

Let $\mathcal{P}_2(\mathbb{R}^q)$ be the space of probability measures on $\mathbb{R}^q$ with finite second moments equipped with a Wasserstein distance of order α . Given $k \geq 2$ probability measures $\mu_1, \dots, \mu_k \in \mathcal{P}_2(\mathbb{R}^q)$, a Wasserstein mean of $\mu_1, \dots, \mu_k$ is any optimal solution to

$$\rho^* = \arg \min \left\{ \sum_{i=1, \dots, k} \lambda_i W^\alpha(\rho, \mu_i) \mid \rho \in \mathcal{P}_2(\mathbb{R}^q) \right\}$$

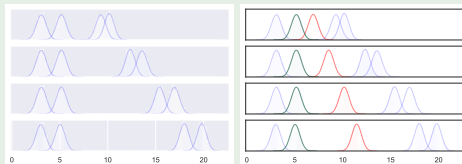
where $\lambda_1, \dots, \lambda_k \geq 0$, such that $\sum_{i=1}^k \lambda_i = 1$.

When $\alpha = 2$, we get a **(weighted) Wasserstein barycenter**.

When $\alpha = 1$, we get a **(weighted) Wasserstein median**.

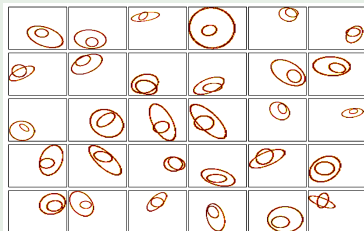
Example 3: Distributions in $\mathcal{P}_2(\mathbb{R})$

<https://github.com/stegua/ot1d>



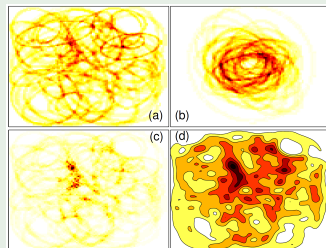
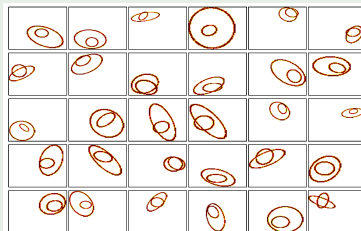
Barycenters of images [Cuturi and Doucet, 2014]

Example 4: Distributions in $\mathcal{P}_2(\mathbb{R}^2)$



Barycenters of images [Cuturi and Doucet, 2014]

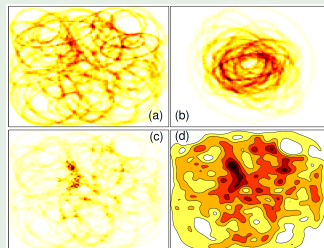
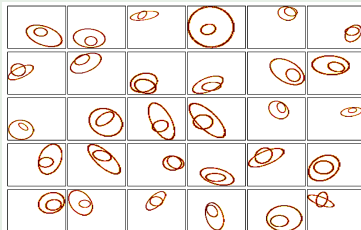
Example 4: Distributions in $\mathcal{P}_2(\mathbb{R}^2)$



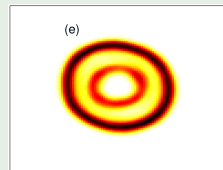
- (a) Euclidean Barycenter
- (b) Reentered Euclidean Barycenter
- (c) Jeffrey Centroid
- (d) Gaussian kernel-based
- (e) **Wasserstein Barycenter**

Barycenters of images [Cuturi and Doucet, 2014]

Example 4: Distributions in $\mathcal{P}_2(\mathbb{R}^2)$



- (a) Euclidean Barycenter
- (b) Reentered Euclidean Barycenter
- (c) Jeffrey Centroid
- (d) Gaussian kernel-based
- (e) **Wasserstein Barycenter**



Thanks!

