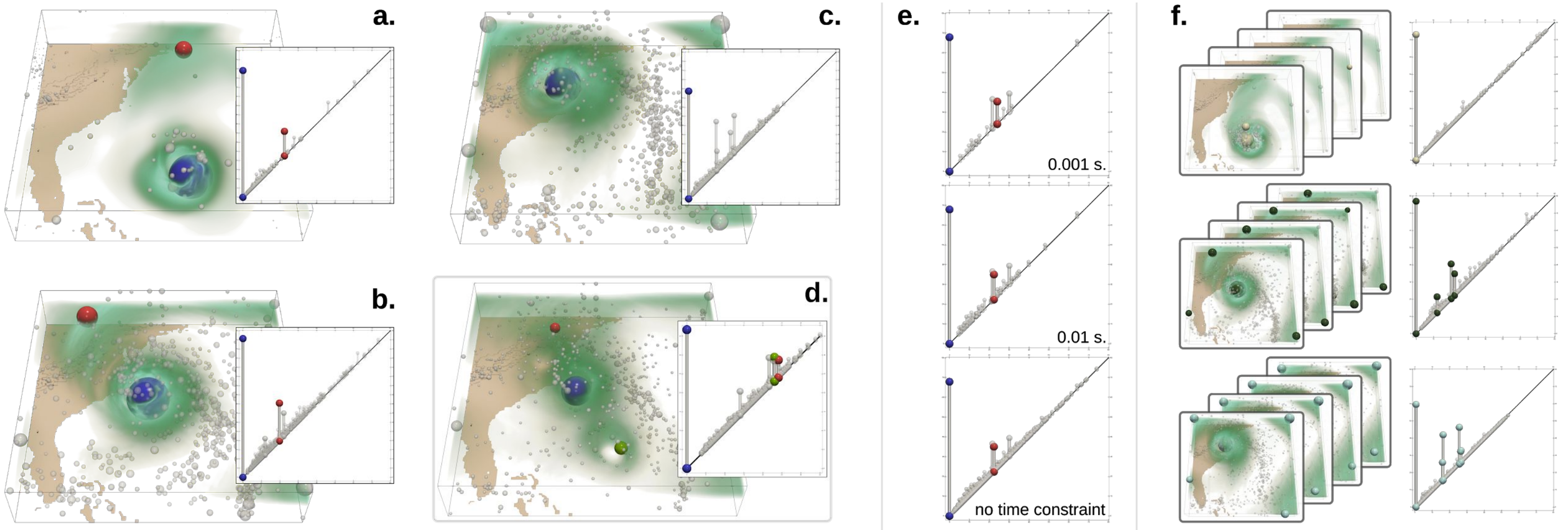
Progressive Wasserstein Barycenters for Persistence Diagrams

Jules Vidal Joseph Budin Julien Tierny



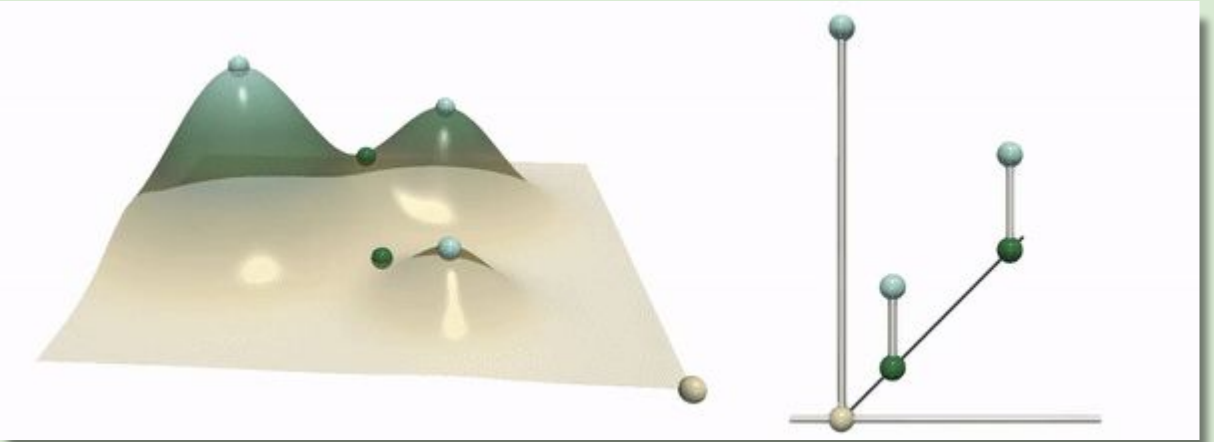
SORBONNE UNIVERSITÉ

Outline

- Introduction
- Preliminaries
  - Wasserstein distances for Persistence Diagrams
  - Solving assignment between diagrams : The Auction algorithm
  - Computing Fréchet means of diagrams : Turner algorithm
- Progressive barycenters
- Application to Ensemble Topological Clustering
- Results

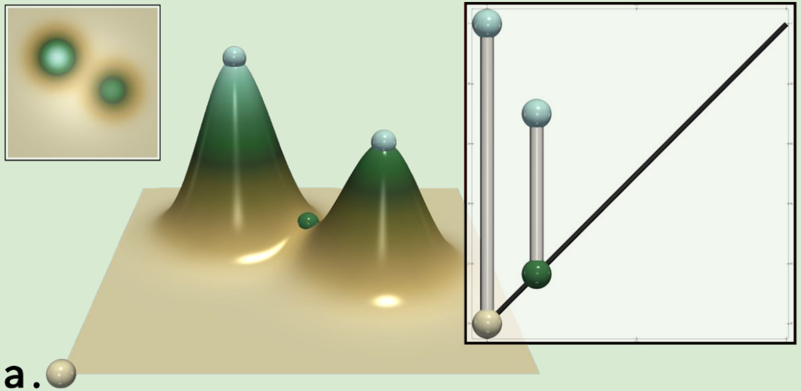
Introduction : The Persistence Diagram (PD)

- Tool from Topological Data Analysis
- Encodes the *Persistence* of topological features
- Highlights important features comparing to noise
- Lightweight topological signature of data

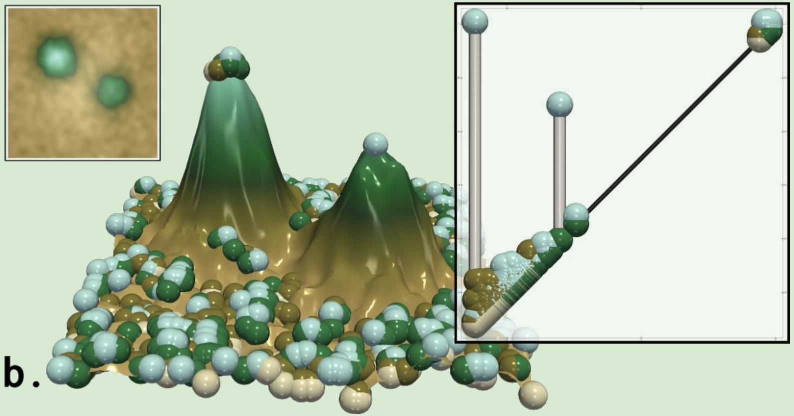


Introduction : The Persistence Diagram (PD)

- Tool from Topological Data Analysis
- Encodes the *Persistence* of topological features
- Highlights important features comparing to noise
- Lightweight topological signature of data



a.



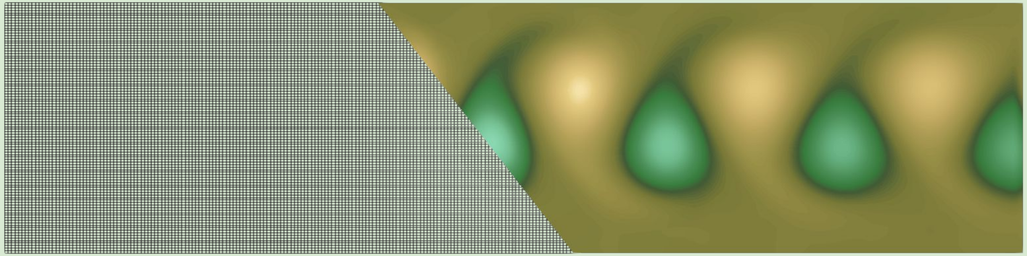
b.

Introduction

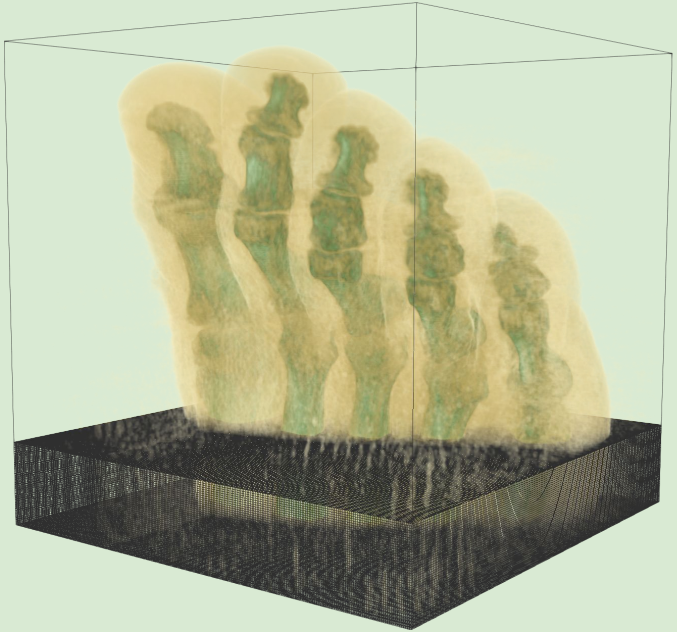
- 2D or 3D data

(physics simulation, medical imaging…)

- High resolution



~ 29,000 cells



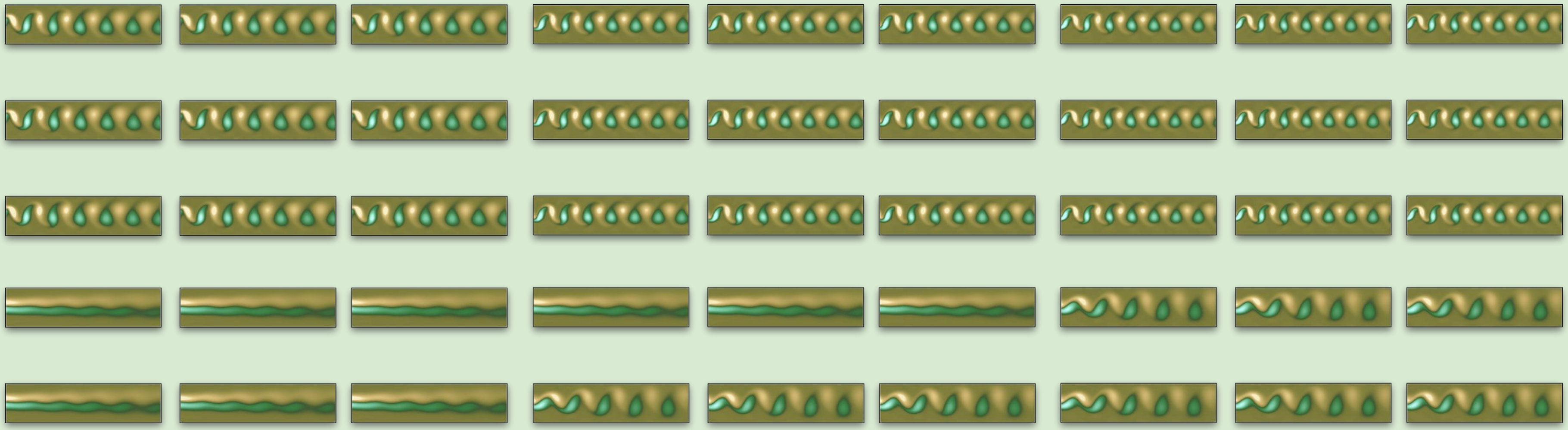
~ 16,500,000 cells

Introduction : Ensemble Analysis

- Ensemble of data-sets
- Characteristics ?

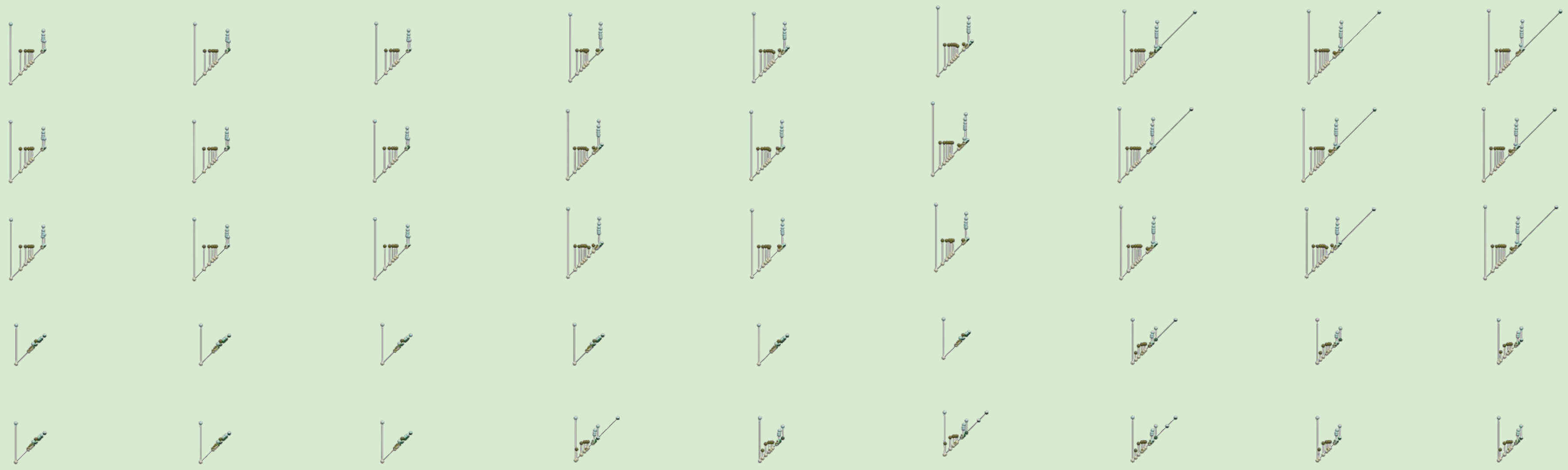
  Commonalities / differences between members

  Average number of topological features, clustering, trend analysis, ...

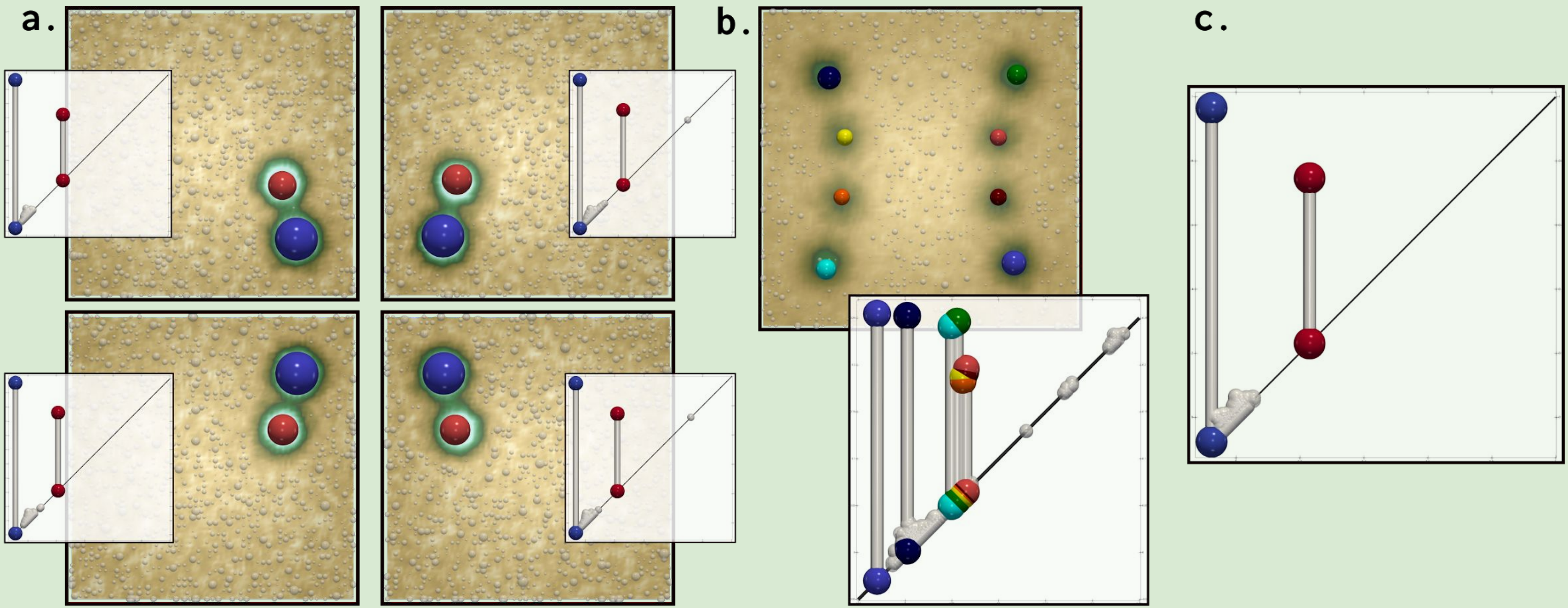


Introduction : Ensemble Analysis

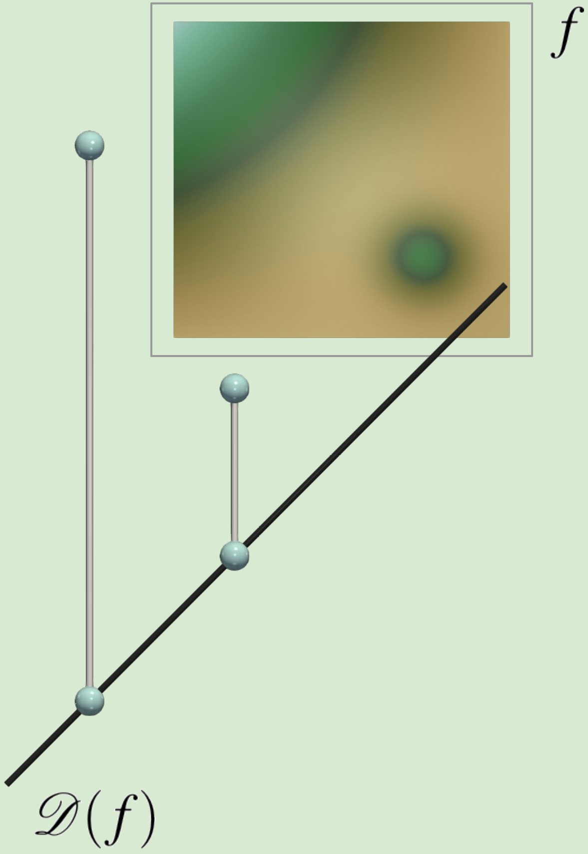
○ Data reduction using TDA

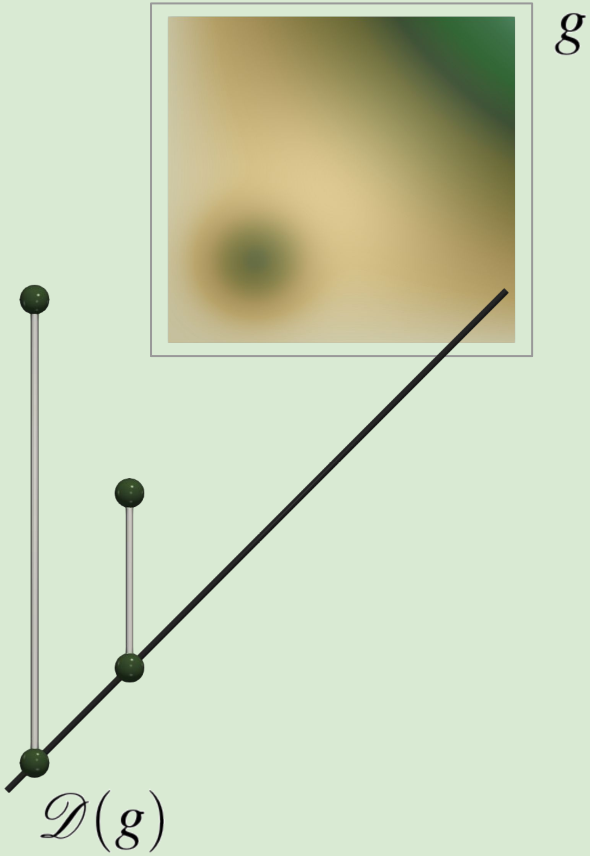


Introduction : Summarization of topological features



Wasserstein distances for PDs





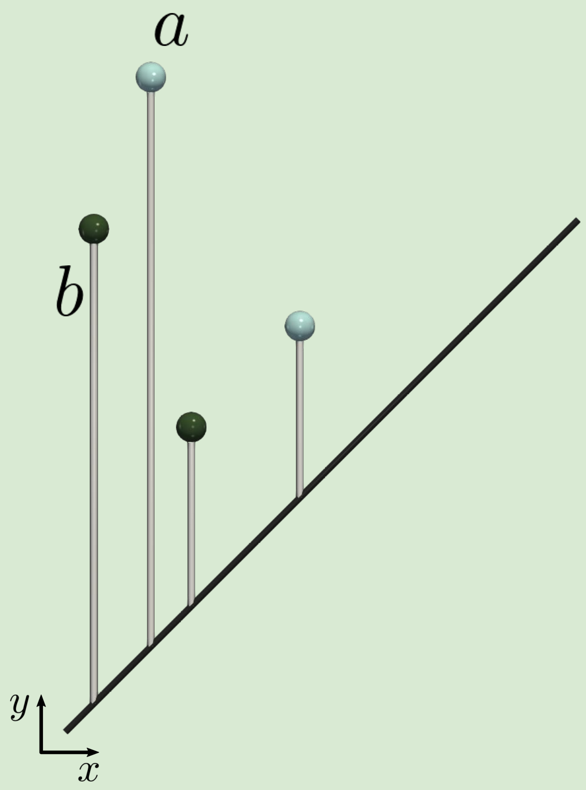
Wasserstein distances for PDs

- Cost of an optimal matching between two diagrams
- Pairwise distance :

$$d_q(a,b) = \left(|x_b - x_a|^q + |y_b - y_a|^q\right)^{1/q}$$

- Wasserstein distance :

$$W_q\big(\mathscr{D}(f), \mathscr{D}(g)\big) = \min_{\phi \in \Phi} \left(\sum_{a \in \mathscr{D}(f)} d_q\big(a, \phi(a)\big)^q \right)^{1/q}$$



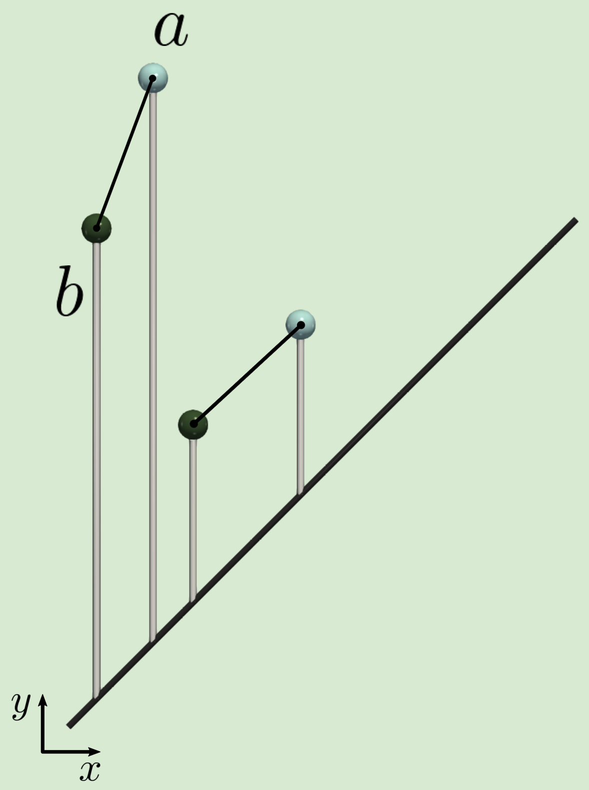
Wasserstein distances for PDs

- Cost of an optimal matching between two diagrams
- Pairwise distance :

$$d_q(a,b) = \left(|x_b - x_a|^q + |y_b - y_a|^q\right)^{1/q}$$

- Wasserstein distance :

$$W_q\big(\mathscr{D}(f), \mathscr{D}(g)\big) = \min_{\phi \in \Phi} \left(\sum_{a \in \mathscr{D}(f)} d_q\big(a, \phi(a)\big)^q \right)^{1/q}$$



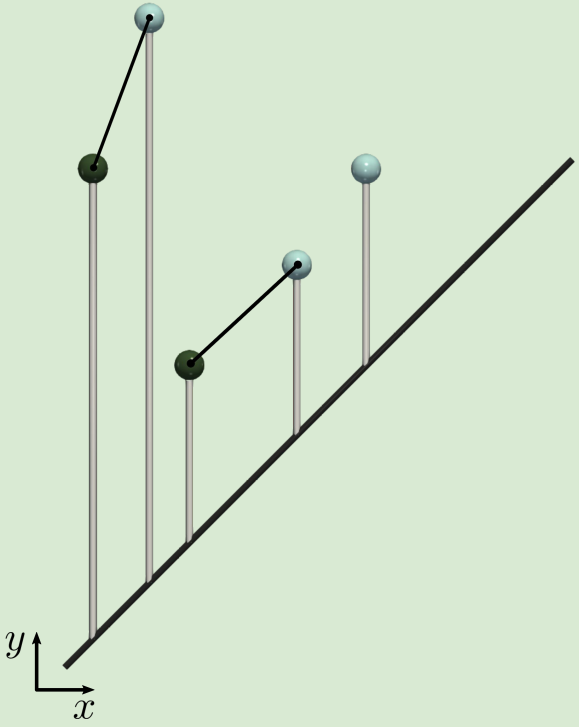
Wasserstein distances for PDs

- Cost of an optimal matching between two diagrams
- Pairwise distance :

$$d_q(a, b) = \left(|x_b - x_a|^q + |y_b - y_a|^q\right)^{1/q}$$

- Wasserstein distance :

$$W_q\big(\mathscr{D}(f), \mathscr{D}(g)\big) = \min_{\phi \in \Phi} \left(\sum_{a \in \mathscr{D}(f)} d_q\big(a, \phi(a)\big)^q \right)^{1/q}$$



Wasserstein distances for PDs

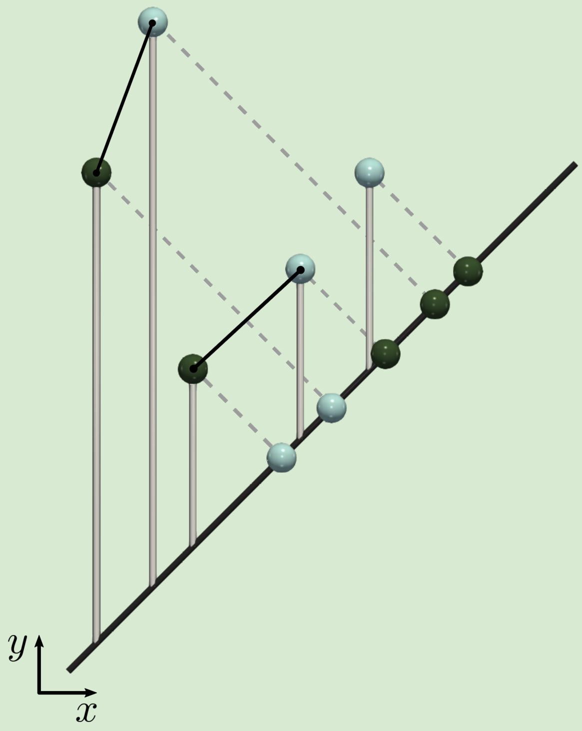
- Cost of an optimal matching between two diagrams
- Pairwise distance :

$$d_q(a,b) = \left(|x_b - x_a|^q + |y_b - y_a|^q\right)^{1/q}$$

- Wasserstein distance :

$$W_q\big(\mathscr{D}(f), \mathscr{D}(g)\big) = \min_{\phi \in \Phi} \left(\sum_{a \in \mathscr{D}(f)} d_q\big(a, \phi(a)\big)^q \right)^{1/q}$$

- Augmented diagrams for a balanced problem



Wasserstein distances for PDs

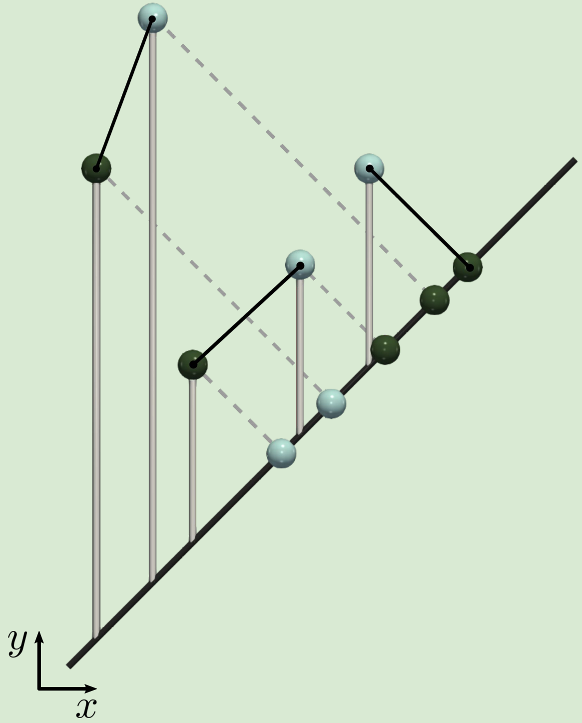
- Cost of an optimal matching between two diagrams
- Pairwise distance :

$$d_q(a,b) = \left(|x_b - x_a|^q + |y_b - y_a|^q\right)^{1/q}$$

- Wasserstein distance :

$$W_q\big(\mathscr{D}(f), \mathscr{D}(g)\big) = \min_{\phi \in \Phi} \left(\sum_{a \in \mathscr{D}(f)} d_q(a, \phi(a))^q \right)^{1/q}$$

- Augmented diagrams for a balanced problem
- Diagonal matchings (≃ removal of a feature)



Auction algorithm

- Approximation of an optimal assignment

- Mimics a real-life auction between *bidders* and *objects.*
- **Auction Round** : several *bids*
  - Each *bidder* ***a*** successively acquires the *object* ***b*** of greatest value $v_{a\to b}$
  - *Objects* prices increase
  - Provides a perfect matching ϕ

$$\widehat{W_2}\left(\mathscr{D}'(f),\mathscr{D}'(g)\right) = \sqrt{\Sigma_{a\in\mathscr{D}'(f)} d_2\left(a,\phi(a)\right)^2}$$

$$v_{a\to b} = -d_2(a,b)^2 - p_b$$

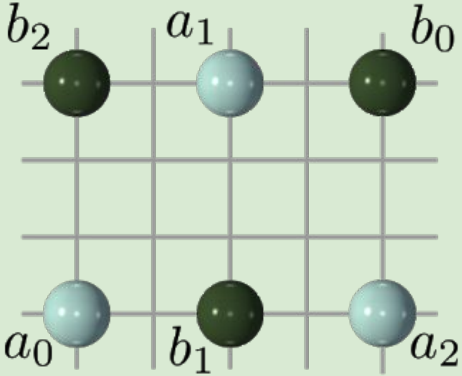
$$p_b \mathrel{+}= \delta_a + \varepsilon$$

- $\widehat{W_2}\left(\mathscr{D}'(f),\mathscr{D}'(g)\right)^2 \leq (1+\gamma)^2\left(\widehat{W_2}\left(\mathscr{D}'(f),\mathscr{D}'(g)\right)^2 - \varepsilon|\mathscr{D}'(f)|\right)$

$$\Longrightarrow \; W_2\left(\mathscr{D}(f),\mathscr{D}(g)\right) \leq \widehat{W_2}\left(\mathscr{D}'(f),\mathscr{D}'(g)\right) \leq (1+\gamma)W_2\left(\mathscr{D}(f),\mathscr{D}(g)\right)$$

- Several Auction Rounds : *ε-scaling*

Auction algorithm : Toy example

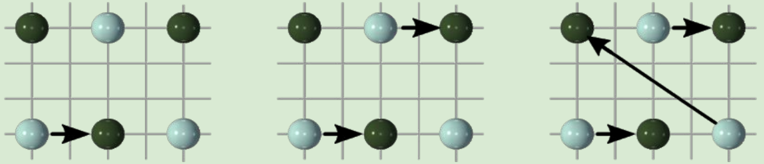


$[v_{a_i \to b_j}]$

	b_0	b_1	b_2
a_0	-25	-4	-9
a_1	-4	-9	-4
a_2	-9	-4	-25

$\varepsilon = 25$ $\widehat{W}_2 = \sqrt{33}$

$a_0 \to b_1$ $a_1 \to b_0$ $a_2 \to b_2$



$p_1 \mathrel{+}= 5 + \varepsilon$

-25	-34	-9
-4	-39	-4
-9	-34	-25

$p_0 \mathrel{+}= \varepsilon$

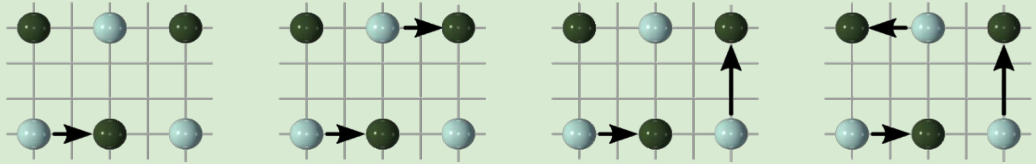
-50	-34	-9
-29	-39	-4
-34	-34	-25

$p_2 \mathrel{+}= 9 + \varepsilon$

-50	-34	-43
-29	-39	-38
-34	-34	-59

$\varepsilon = 5$ $\widehat{W}_2 = \sqrt{17}$

$a_0 \to b_1$ $a_1 \to b_0$ $a_2 \to b_0$ $a_1 \to b_2$



$p_1 \mathrel{+}= 9 + \varepsilon$

-50	-48	-43
-29	-53	-38
-34	-48	-59

$p_0 \mathrel{+}= 9 + \varepsilon$

-64	-48	-43
-43	-53	-38
-48	-48	-59

$p_0 \mathrel{+}= \varepsilon$

-69	-48	-43
-48	-53	-38
-53	-48	-59

$p_2 \mathrel{+}= 10 + \varepsilon$

-69	-48	-58
-48	-53	-53
-53	-48	-74

Turner algorithm

Computation of a Fréchet mean of a set of PDs

$$\mathscr{D}^* = \underset{\mathscr{D} \in \mathbb{D}}{\arg\min} \sum_{\mathscr{D}(f_i) \in \mathscr{F}} W_2\big(\mathscr{D}, \mathscr{D}(f_i)\big)^2$$

- Gradient descent-like approach
- *N* assignment computations for each Relaxation

Algorithm 1 Reference algorithm for Wasserstein Barycenters [94].

Input : Set of diagrams $\mathscr{F} = \{\mathscr{D}(f_1), \mathscr{D}(f_2), \ldots, \mathscr{D}(f_N)\}$
Output : Wasserstein barycenter $\mathscr{D}^*$

$\mathscr{D}^* \leftarrow \mathscr{D}(f_i)$ // with i randomly chosen in $[1, N]$
while $\{\phi_1, \phi_2, \ldots, \phi_N\}$ change **do**
  // Relaxation start
  for $i \in [1, N]$ **do**
    $\phi_i \leftarrow Assignment\big(\mathscr{D}(f_i), \mathscr{D}^*\big)$ // optimizing Eq. 2
  end for
  $\mathscr{D}^* \leftarrow Update(\phi_1, \ldots, \phi_n)$ // arithmetic means in birth/death space
  // Relaxation end
end while
return $\mathscr{D}^*$

Progressive Barycenters

- Naive approach : Auction + Turner

- Key observations :
  - The assignments can be re-used between two *Relaxations*
  - The early Relaxations do not necessarily need great precision in the assignments.
  - Emphase should be put on larger Persistence Pairs
  - Trivial parallelization

Progressive Barycenters

Progressivity in accuracy

- Only one Auction Round at each Relaxation
- Prices memorization
- Global *ε-scaling*

Algorithm 1 Reference algorithm for Wasserstein Barycenters [94].

Input : Set of diagrams $\mathscr{F} = \{\mathscr{D}(f_1), \mathscr{D}(f_2), \dots, \mathscr{D}(f_N)\}$
Output : Wasserstein barycenter $\mathscr{D}^*$

```
D* ← D(f_i)                                  // with i randomly chosen in [1, N]
while {φ_1, φ_2, ..., φ_N} change do
    // Relaxation start
    for i ∈ [1, N] do
        φ_i ← Assignment(D(f_i), D*)         // optimizing Eq. 2
    end for
    D* ← Update(φ_1, ..., φ_n)               // arithmetic means in birth/death space
    // Relaxation end
end while
return D*
```

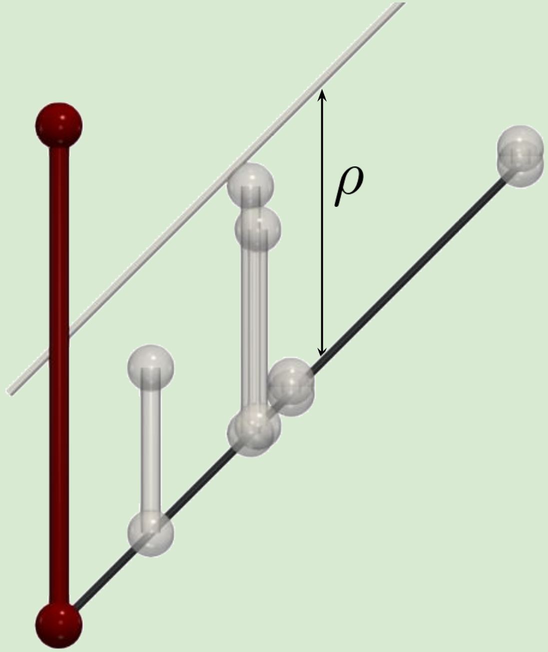
Progressive Barycenters

Persistence-driven Progressivity

- Decreasing persistence threshold ρ
  Pairs are added at each Relaxation

- $\rho = \sqrt{4\varepsilon}$

- Introduction of a time-constraint $t_{\max}$



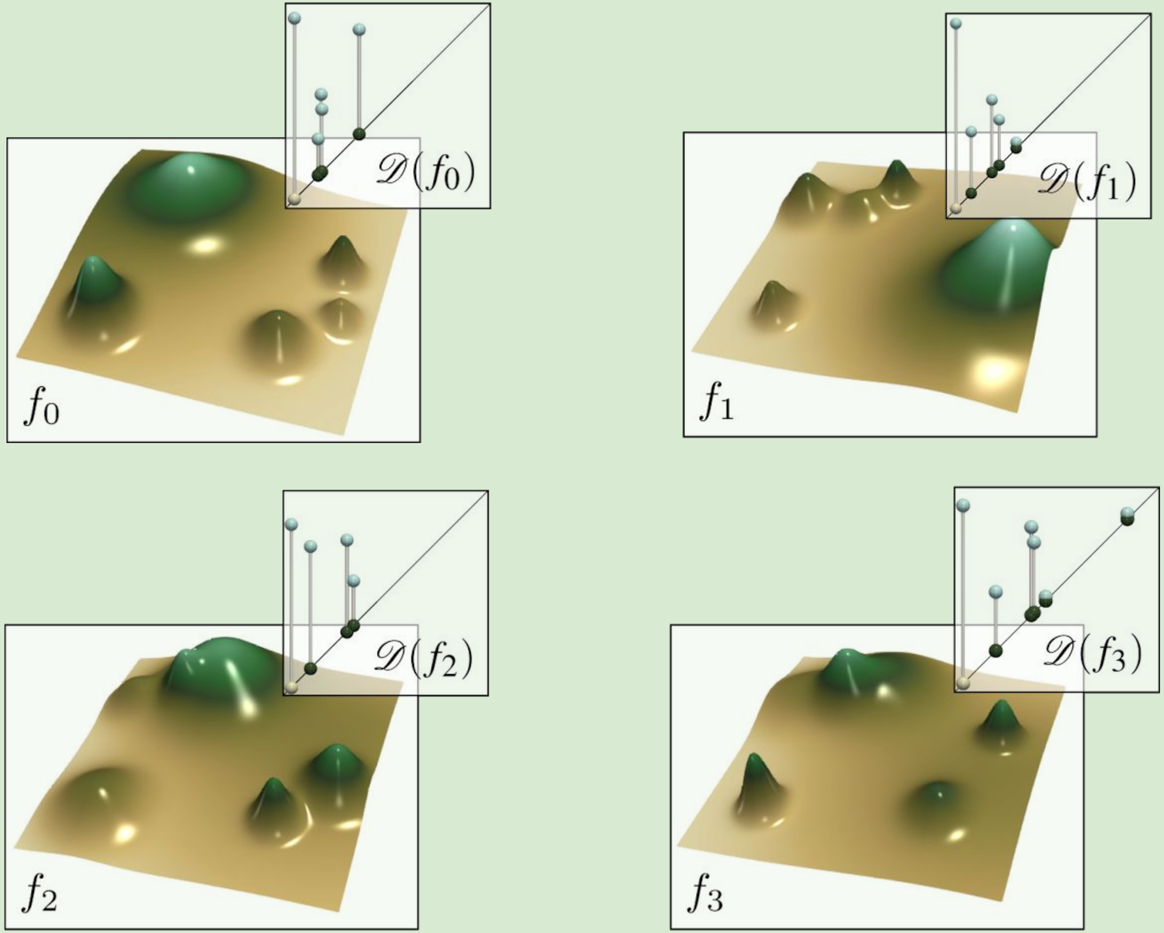
Progressive Barycenters

Algorithm 2 Our overall algorithm for Progressive Wasserstein Barycenters.

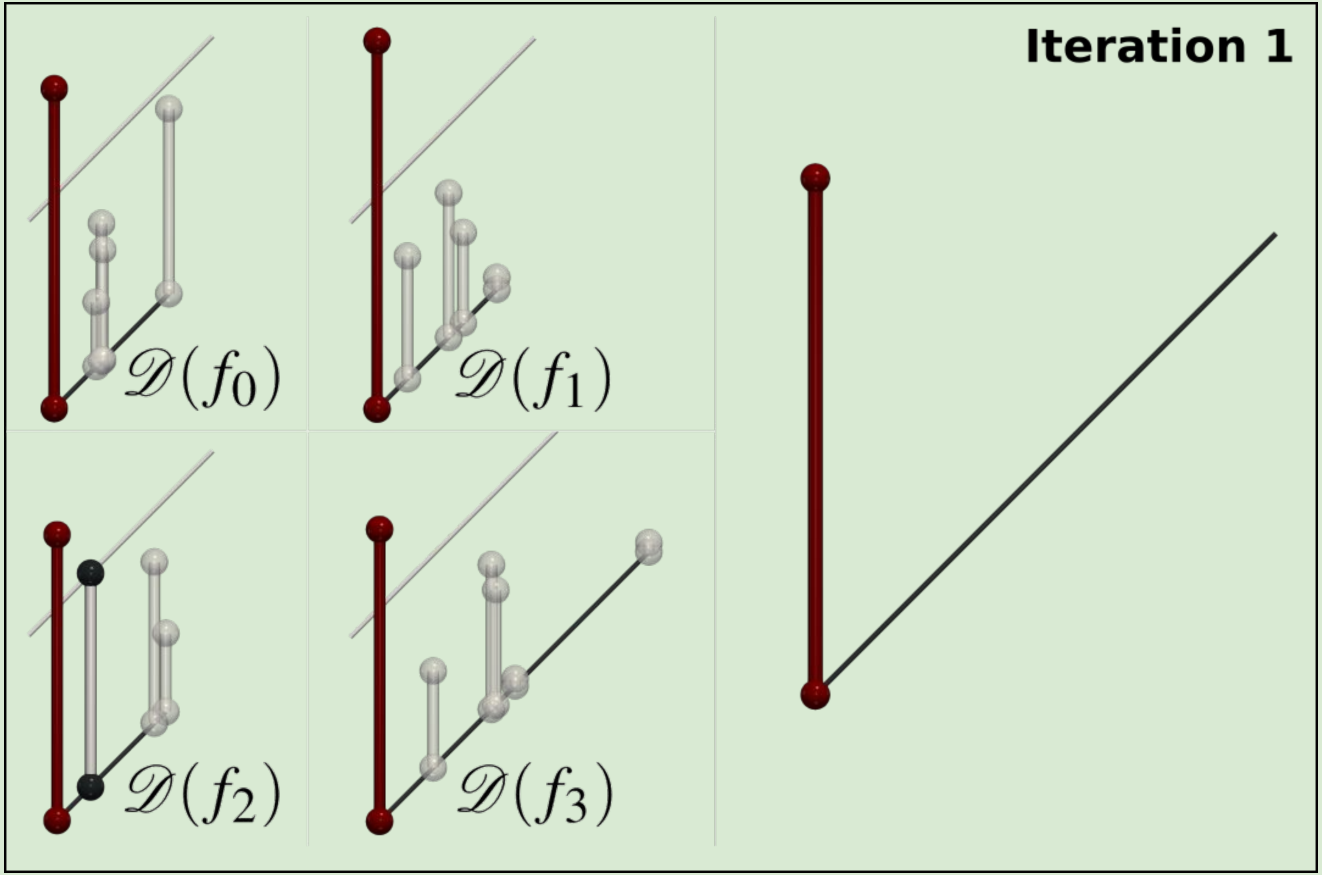
Input : Set of diagrams $\mathscr{F} = \{\mathscr{D}(f_1), \mathscr{D}(f_2), \ldots, \mathscr{D}(f_N)\}$, time constraint t_{max}
Output : Wasserstein barycenter $\mathscr{D}^*_\rho$

$\mathscr{D}^*_\rho \leftarrow \mathscr{D}_\rho(f_i)$ // with i randomly chosen in $[1, N]$
while the Fréchet energy decreases **do**
  // Relaxation start
  for $i \in [1, N]$ **do**
    // In parallel **// Sec. 3.5**
    $\phi_i \leftarrow Assignment\left(\mathscr{D}_\rho(f_i), \mathscr{D}^*_\rho\right)$ **// Sec. 3.2**
  end for
  $\mathscr{D}^*_\rho \leftarrow Update(\phi_1, \ldots, \phi_n)$ // arithmetic means in birth/death space
  $EpsilonScaling()$ **// Sec. 3.3**
  if $t < 0.1 \times t_{max}$ **then** $PersistenceScaling()$ **// Sec. 3.4**
  else if $t >= t_{max}$ **then return** $\mathscr{D}^*_\rho$ **// Sec. 3.6**
  // Relaxation end
end while
return $\mathscr{D}^*_\rho$

Toy Example

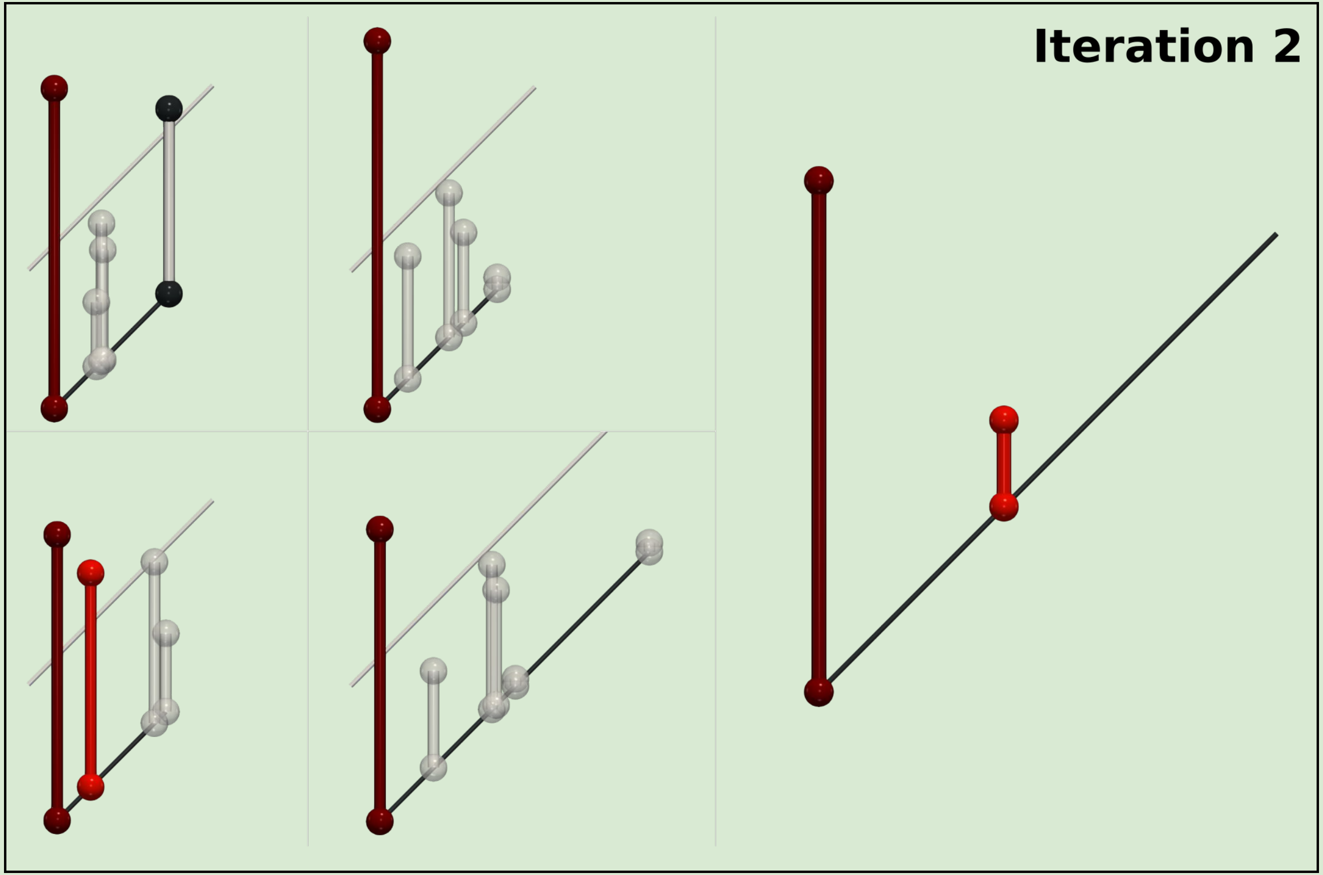


Toy Example

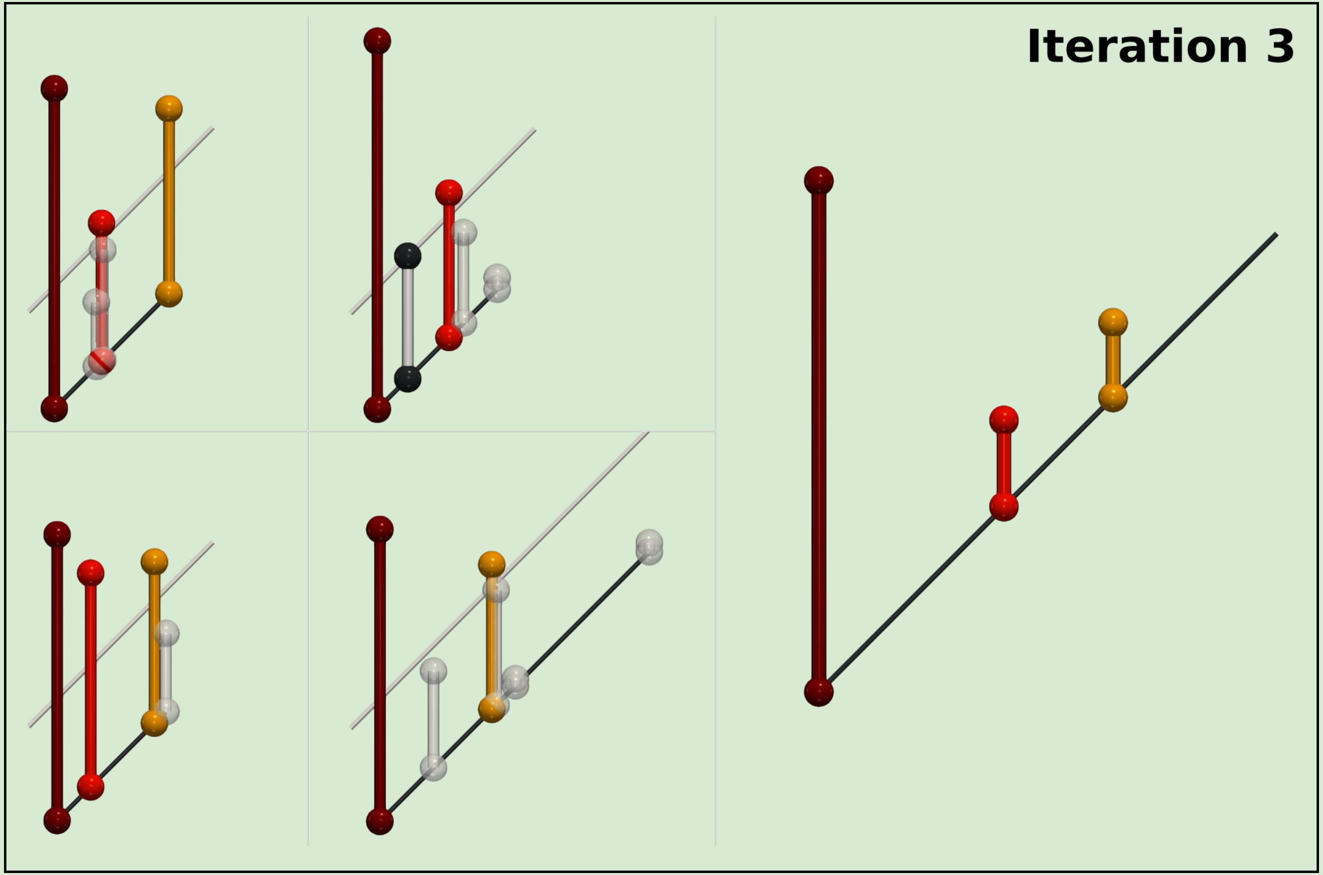


Iteration 1
$\mathscr{D}(f_0)$
$\mathscr{D}(f_1)$
$\mathscr{D}(f_2)$
$\mathscr{D}(f_3)$

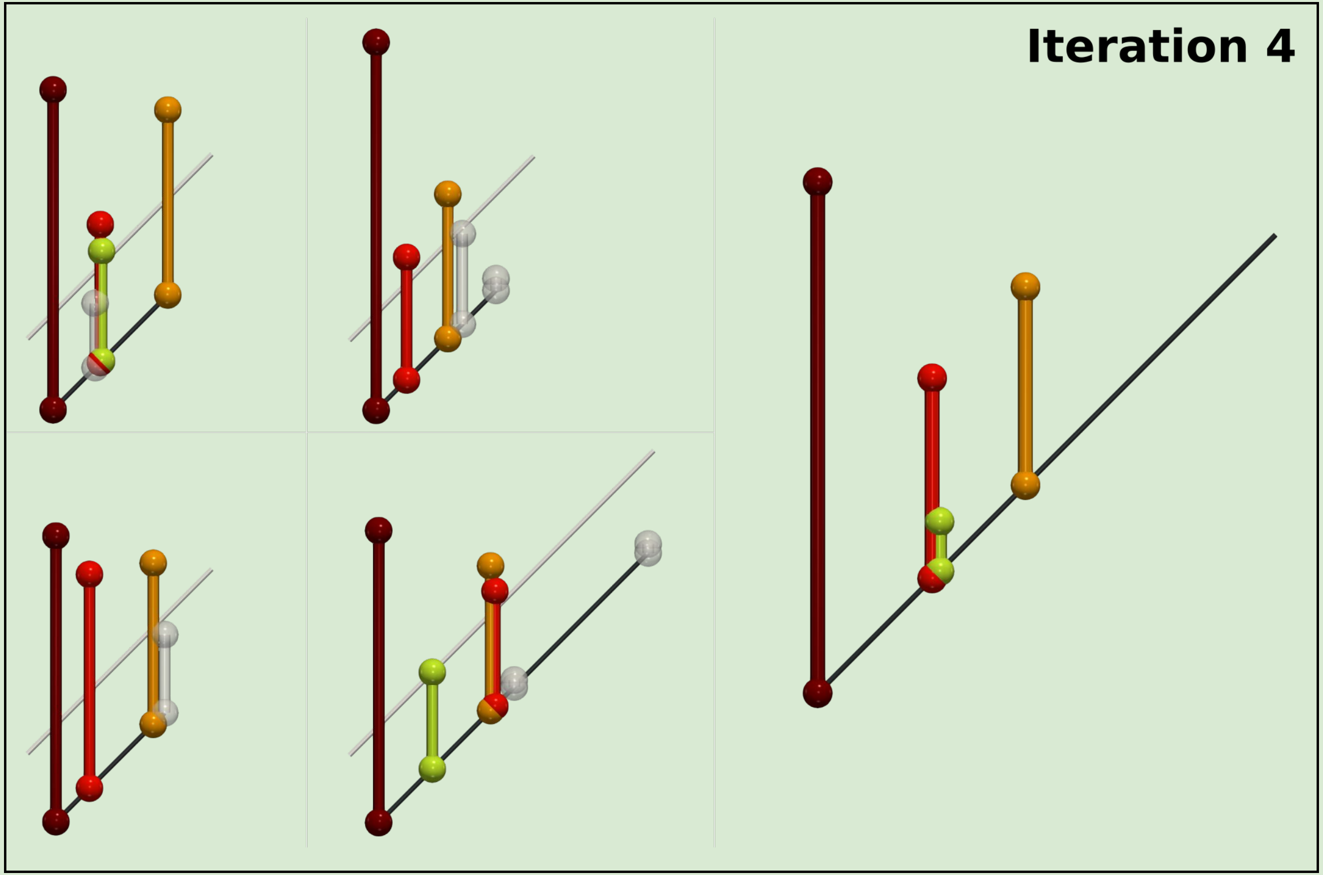
Toy Example



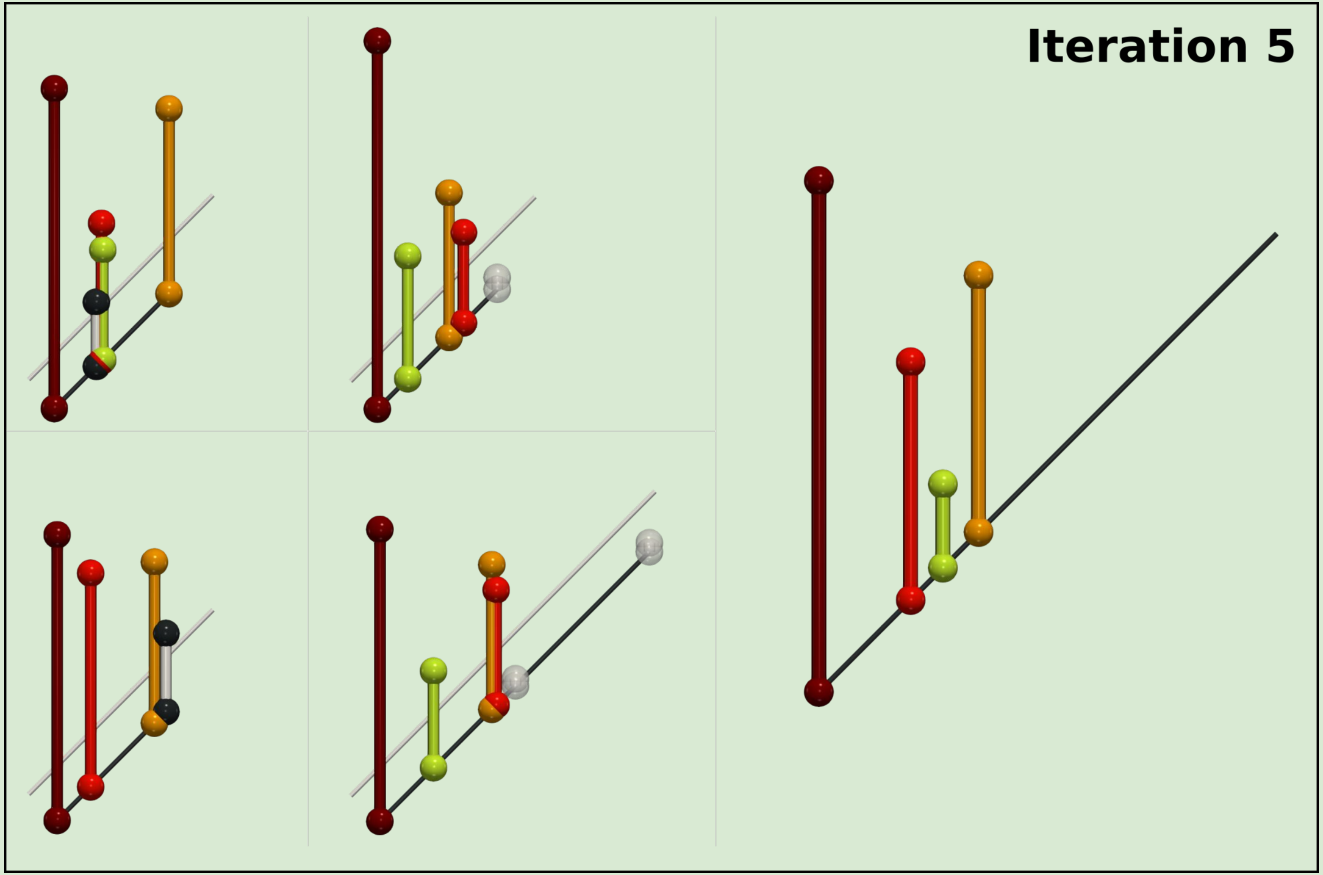
Toy Example



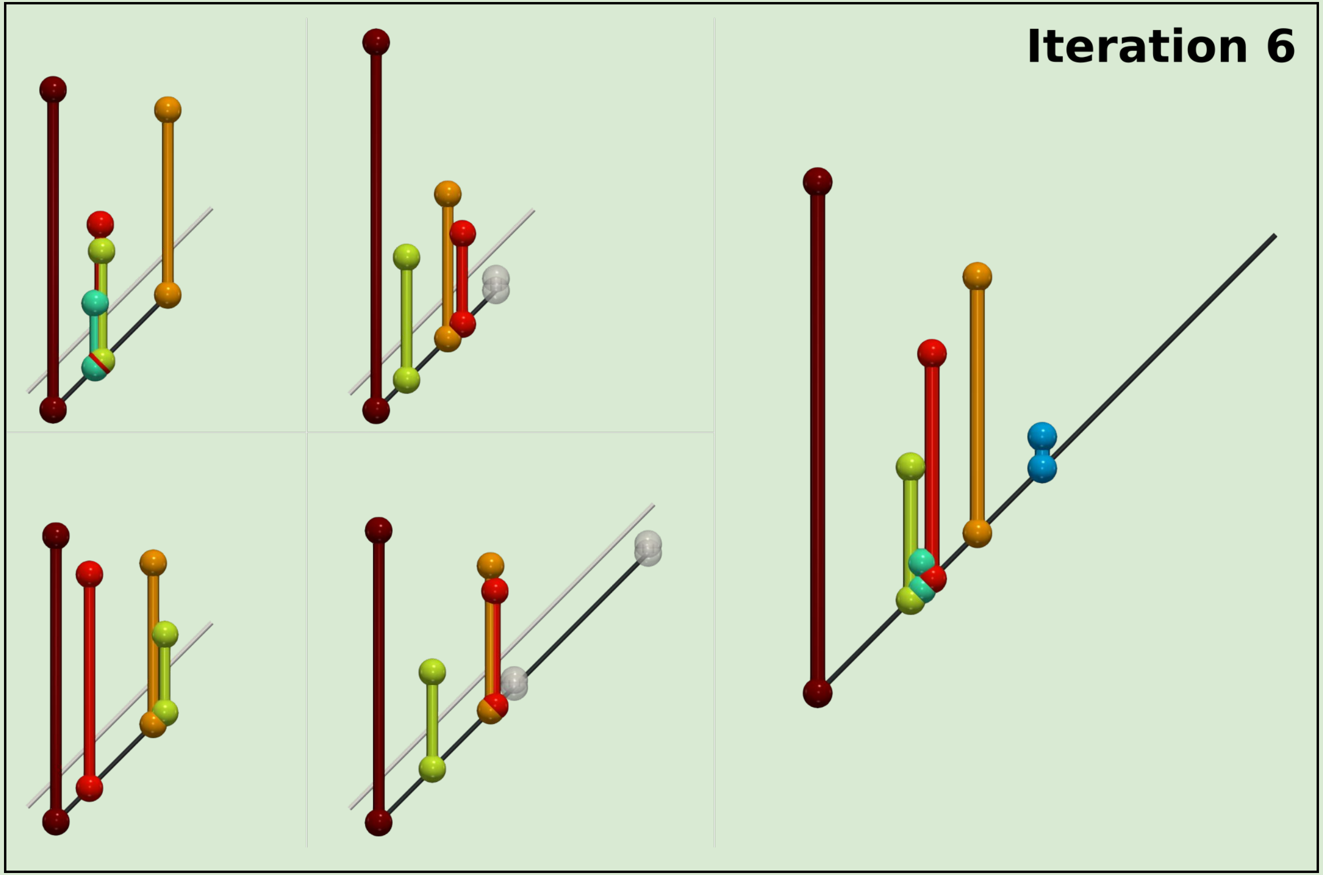
Toy Example



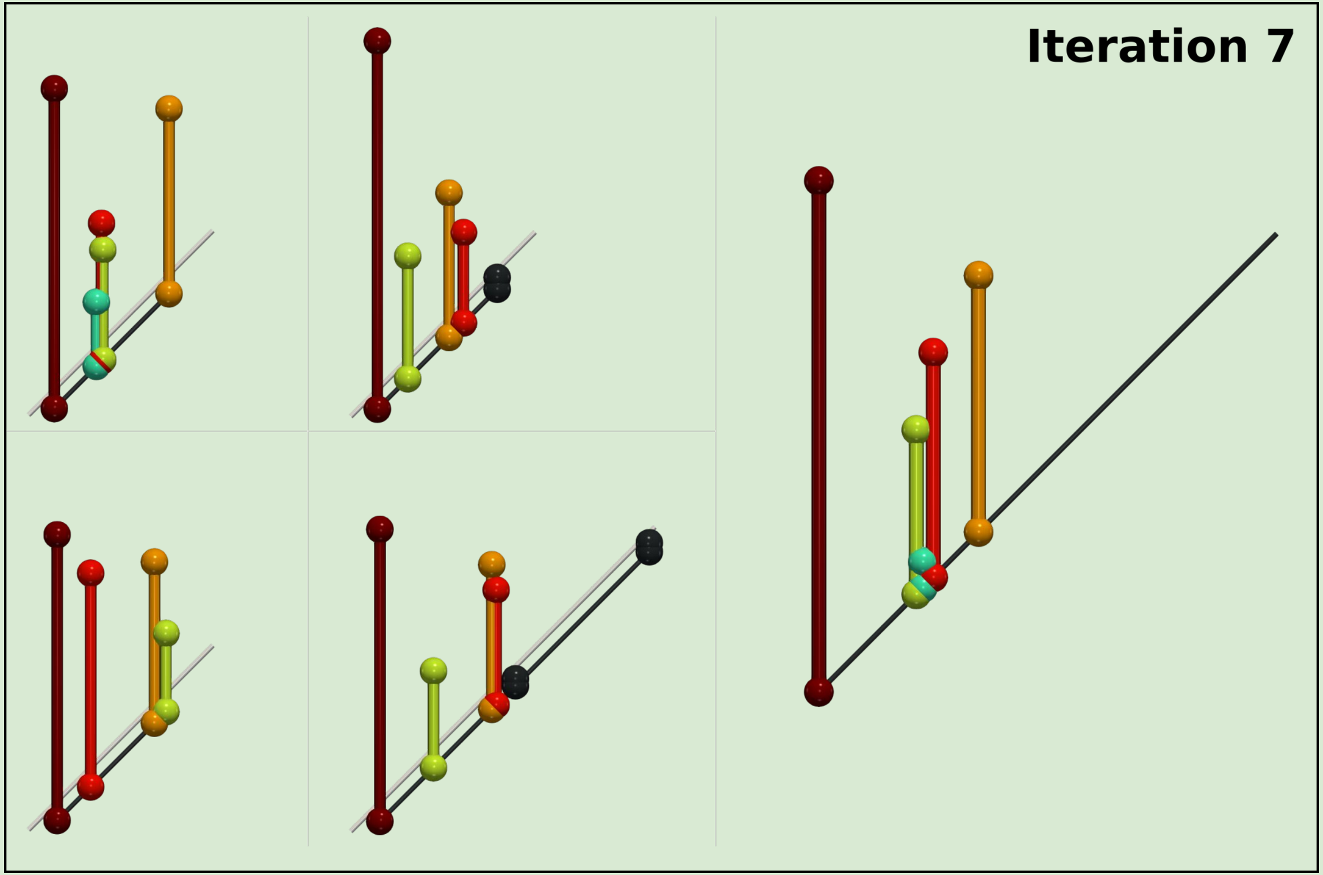
Toy Example



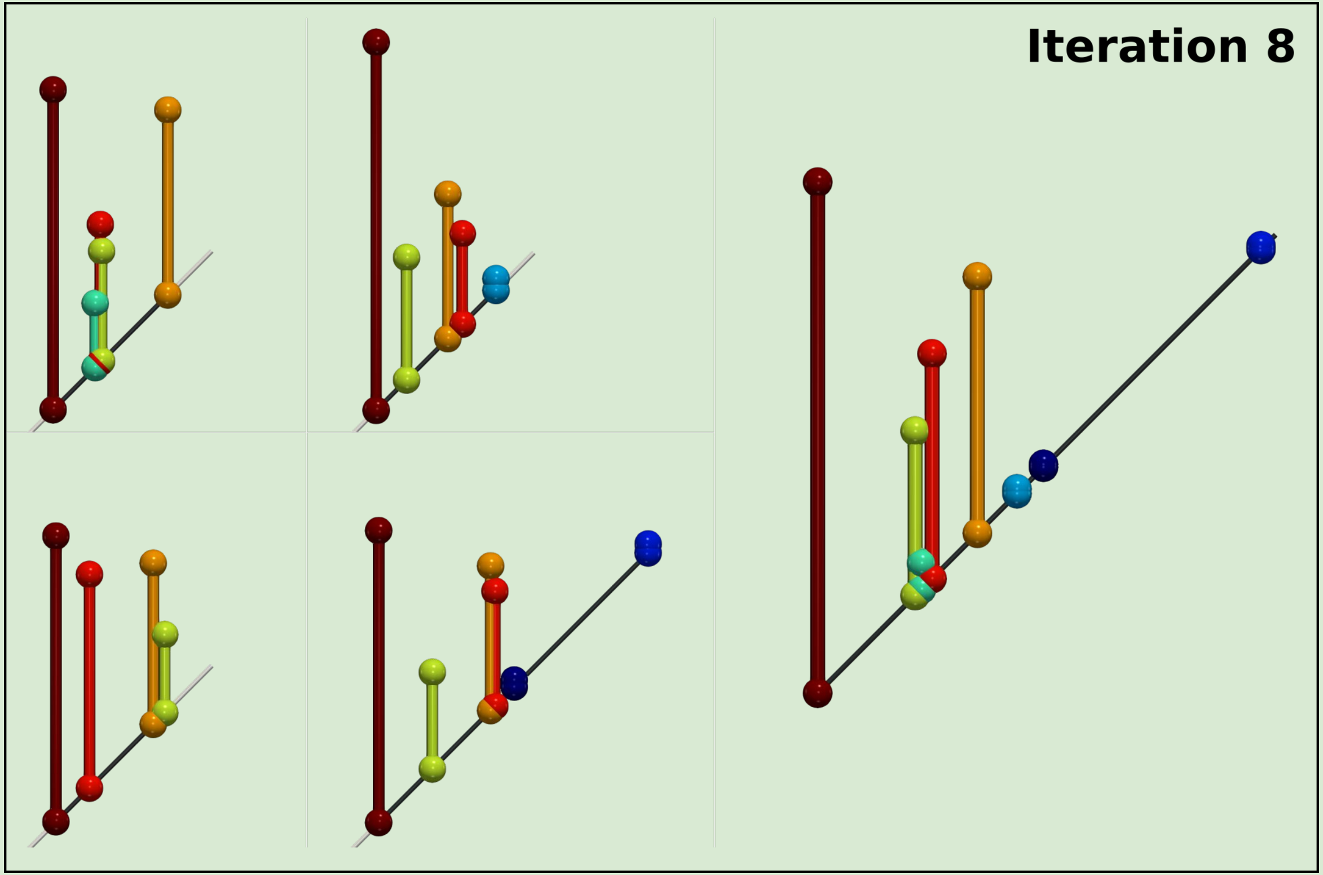
Toy Example



Toy Example



Toy Example



Application to Ensemble Topological Clustering

- Use of the K-Means algorithm
  Two sub-routines : *Assignment* and *Update*

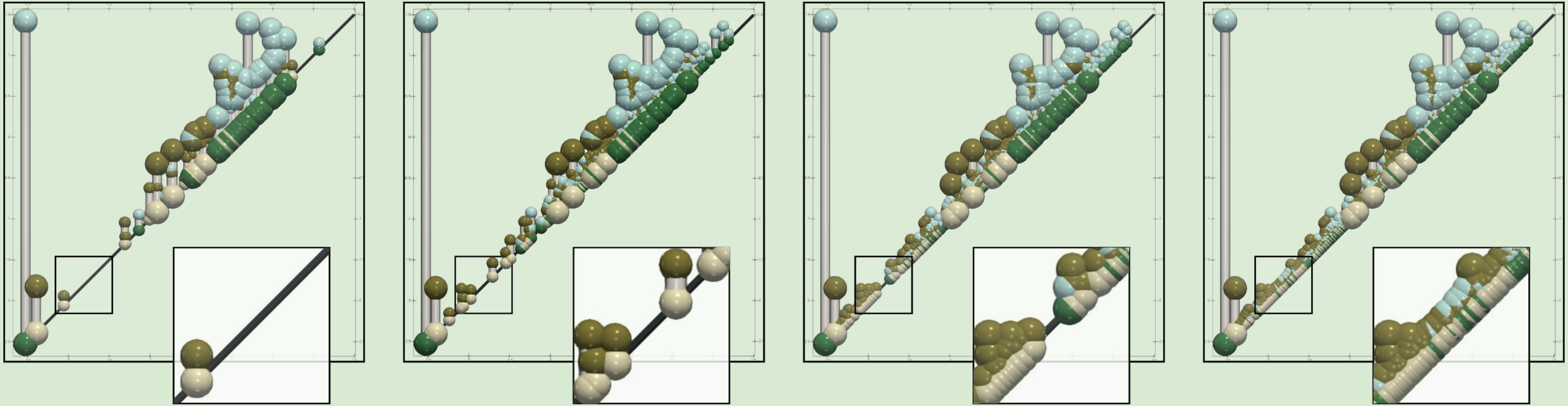
- Progressive accuracy : One relaxation per clustering Update

- Persistence-driven progressivity and time constraint

- Geometrical lifting for the metrics

- K-Means++, Accelerated K-Means

Results



Barycenter computation

Time performance

Data set	N	$\#_{\mathcal{D}(f_i)}$	Sinkhorn [53]	Munkres [94]+[86]	Auction [94]+[51]	**Ours**	Speedup
Gaussians (Fig. 8)	100	2,078	7,499.33	$>$ 24H	8,975.60	785.53	11.4
Vortex Street (Fig. 9)	45	14	54.21	0.14	0.47	0.23	0.6
Starting Vortex (Fig. 10)	12	36	40.98	0.06	0.67	0.28	0.2
Isabel (3D) (Fig. 1)	12	1,337	1,070.57	$>$ 24H	377.42	82.95	4.5
Sea Surface Height (Fig. 11)	48	1,379	4,565.37	$>$ 24H	949.08	75.90	12.5

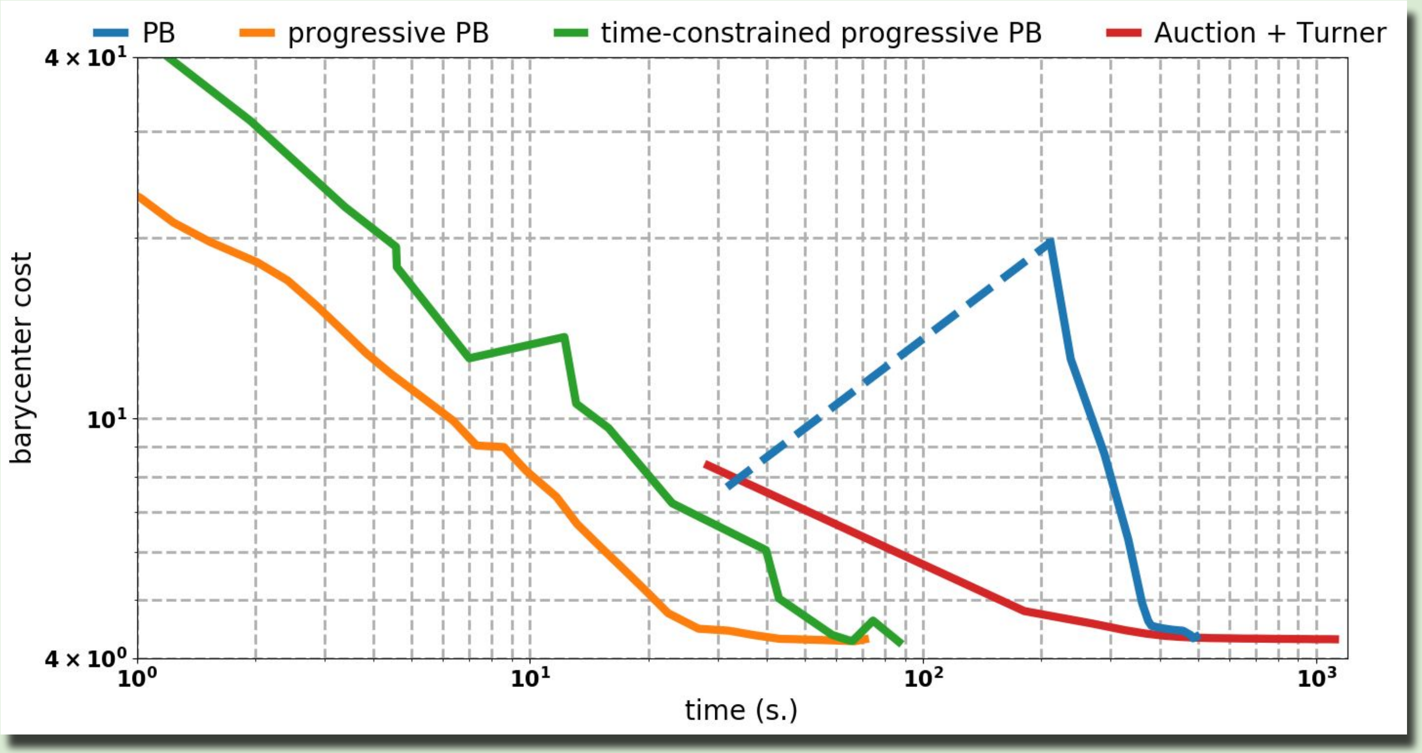
Barycenter computation

Time performance

Data set	N	$\#_{\mathcal{D}(f_i)}$	1 thread	8 threads	Speedup
Gaussians (Fig. 8)	100	2,078	785.53	117.91	6.6
Vortex Street (Fig. 9)	45	14	0.23	0.10	2.3
Starting Vortex (Fig. 10)	12	36	0.28	0.19	1.5
Isabel (3D) (Fig. 1)	12	1,337	82.95	31.75	2.6
Sea Surface Height (Fig. 11)	48	1,379	75.90	19.40	3.9

Barycenter computation

Barycenter quality



Barycenter computation

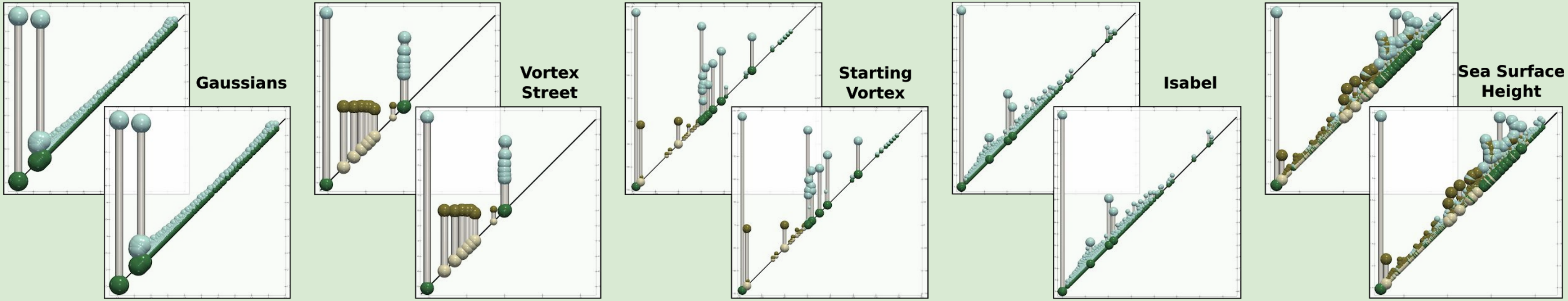
Barycenter quality : Comparison of Fréchet energy

Data set	N	$\#_{\mathcal{D}(f_i)}$	Auction [94]+[51]	Ours	**Ratio**
Gaussians (Fig. 8)	100	2,078	39.4	39.0	0.99
Vortex Street (Fig. 9)	12	36	415.1	412.5	0.99
Starting Vortex (Fig. 10)	45	14	112,787.0	112,642.0	1.00
Isabel (3D) (Fig. 1)	12	1,337	2,395.6	2,337.1	0.98
Sea Surface Height (Fig. 11)	48	1,379	7.2	7.1	0.99

- Disparity of 2% at most

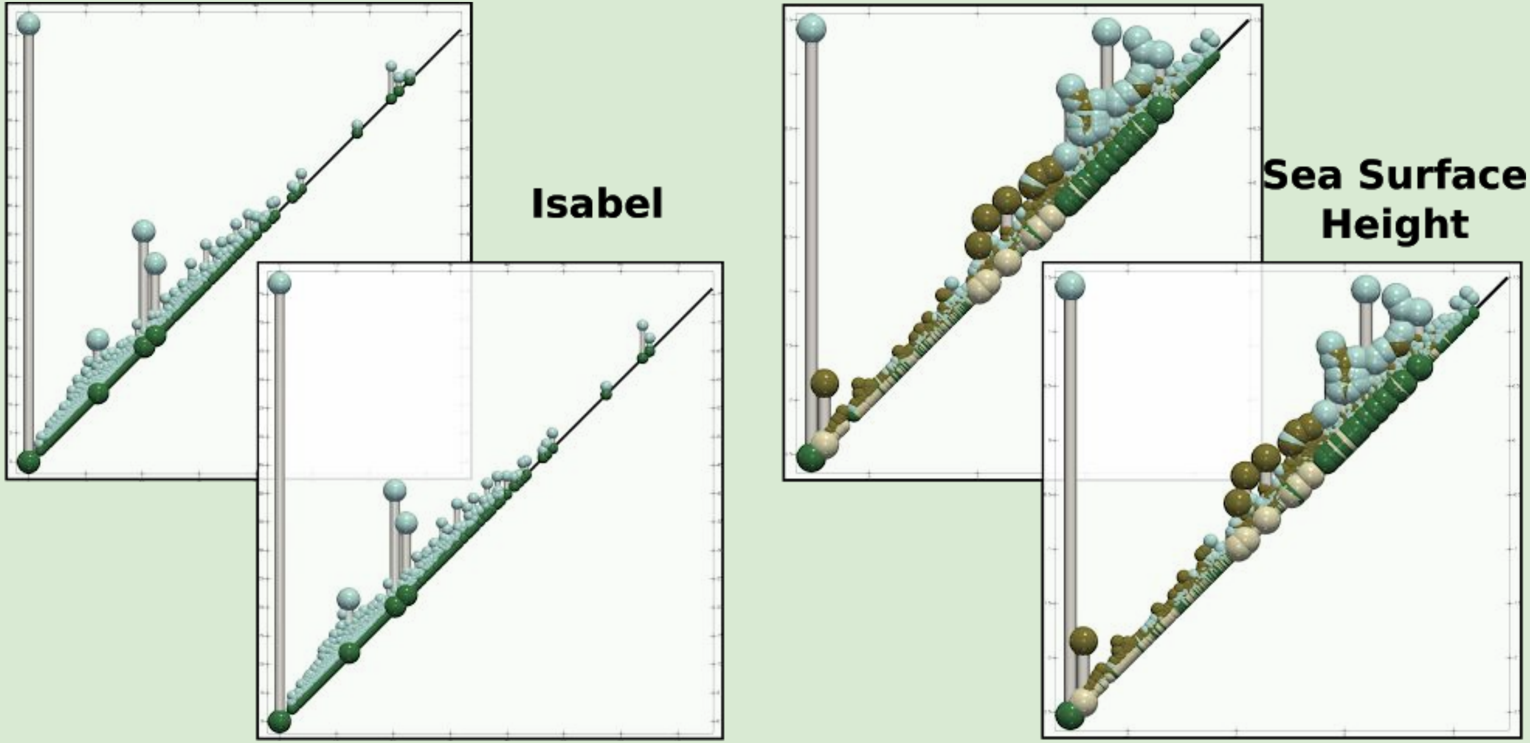
Barycenter computation

Visual quality : Ours (bottom) compared to the Auction approach



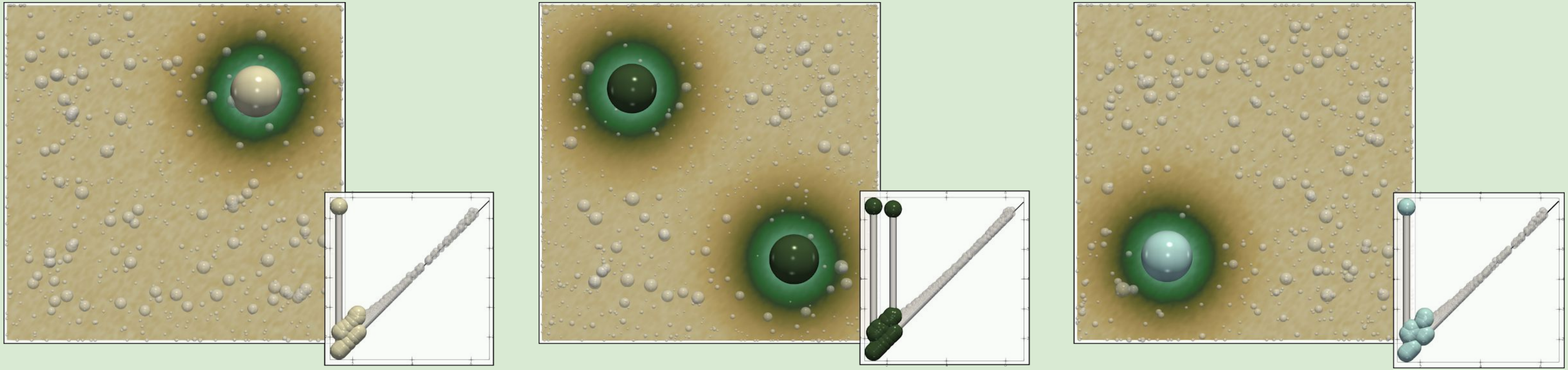
Results : Barycenter computation

Visual quality : Ours (bottom) compared to the Auction approach

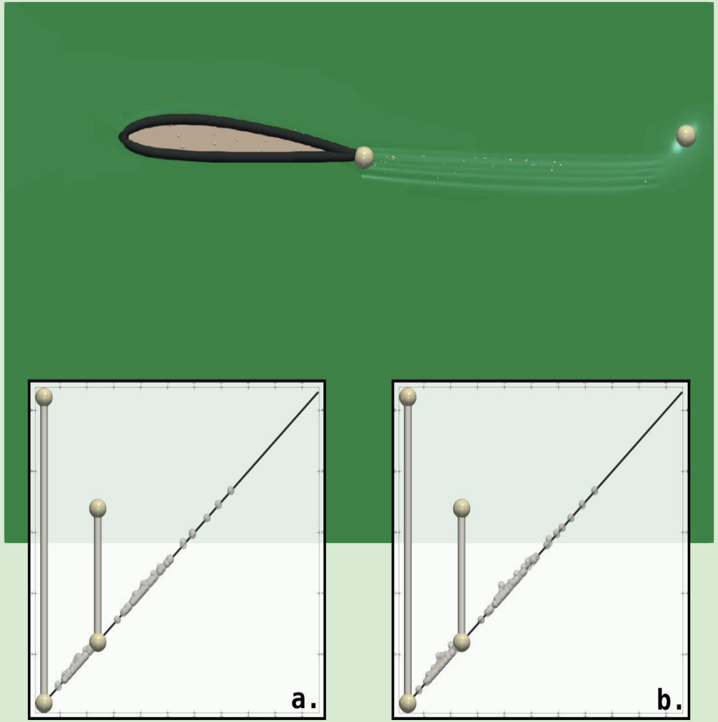


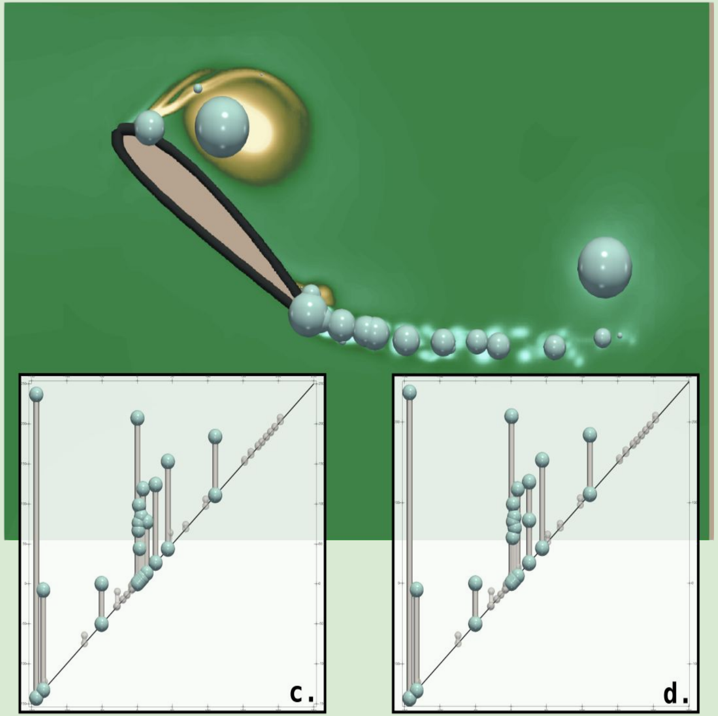
Ensemble Visual Analysis with Topological Clustering

Gaussian Data-set : 100 members, 3 clusters



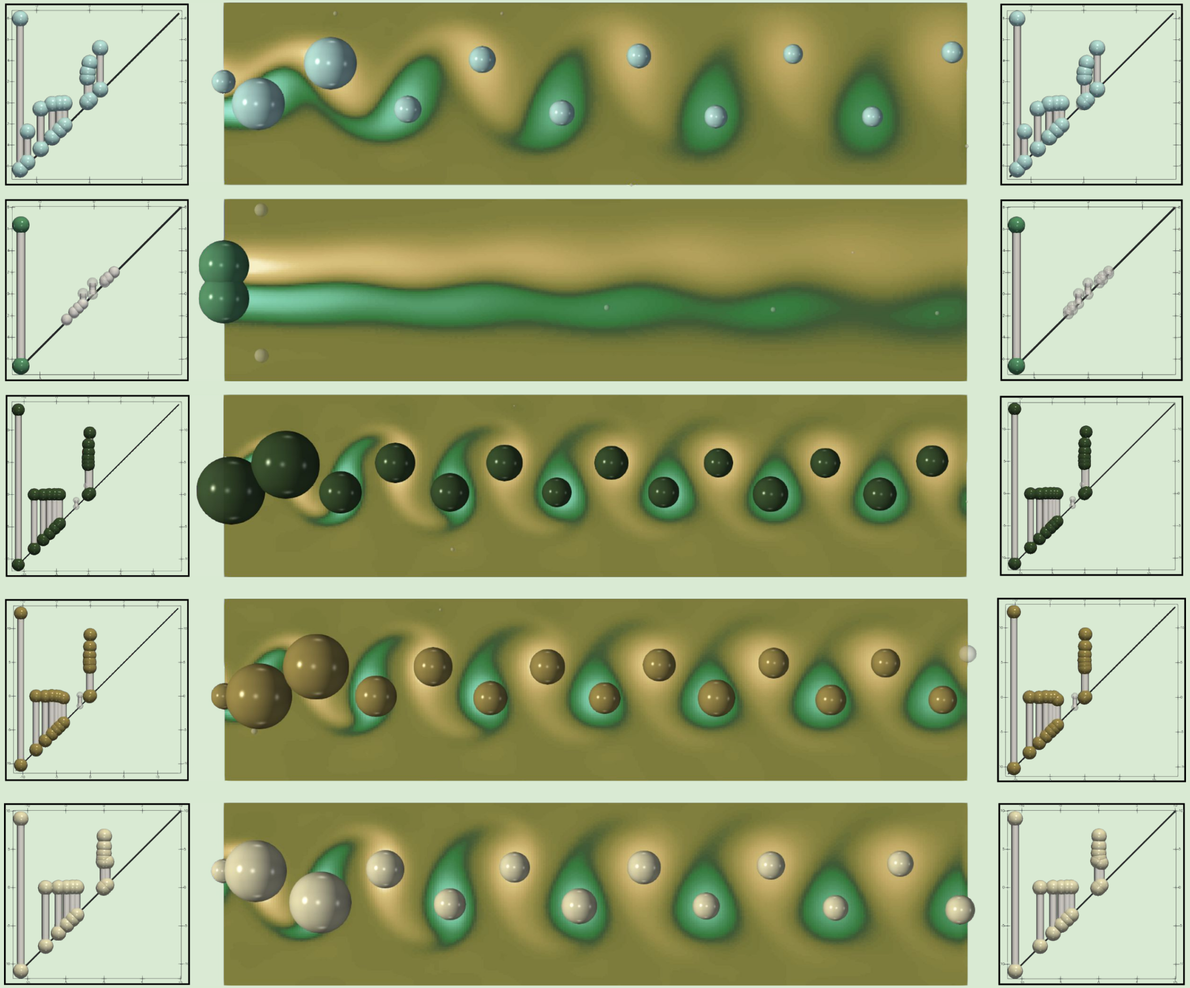
Starting Vortex Data-set : 12 members, 2 clusters



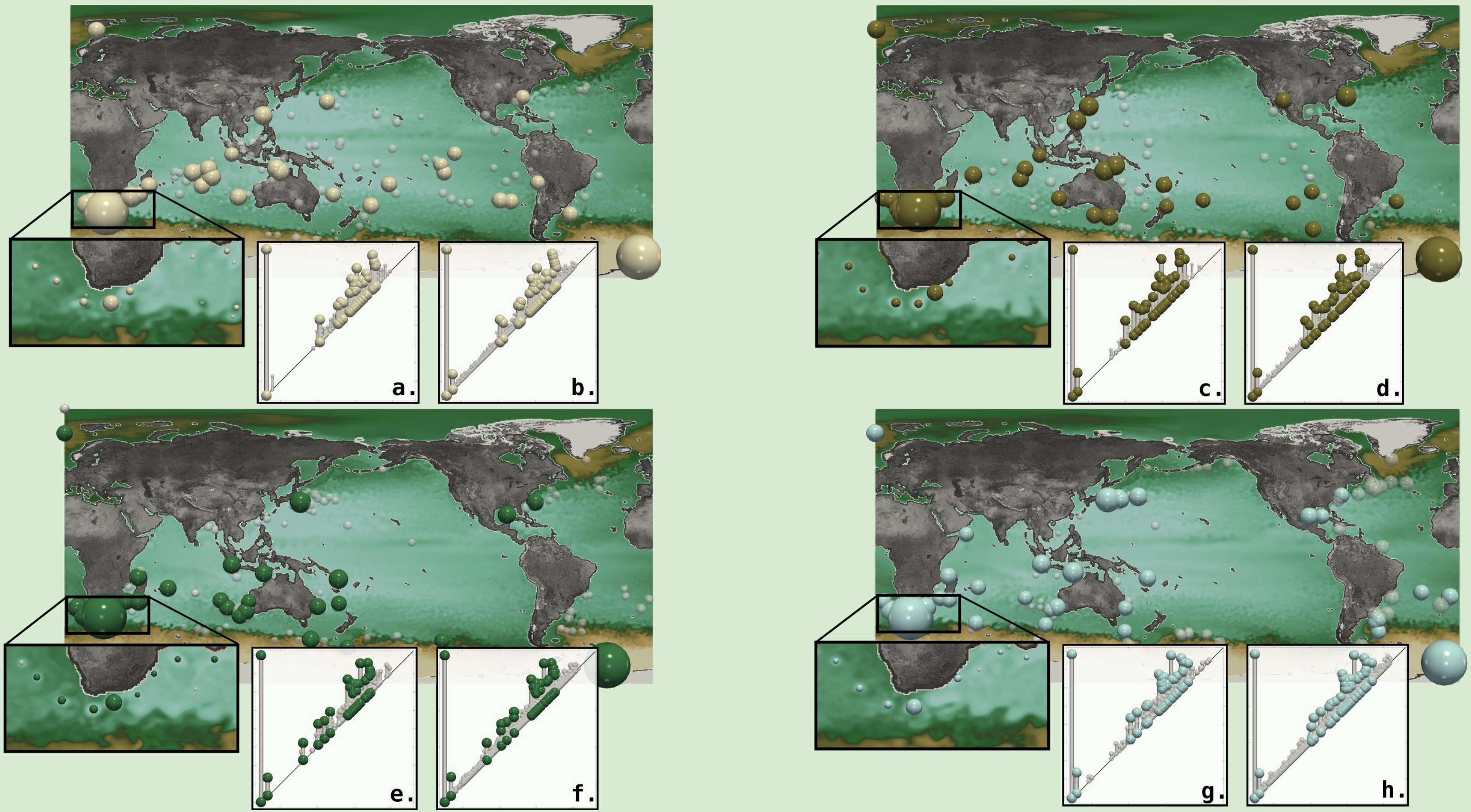


Vortex Street Data-set

45 members, 5 clusters



Sea Surface Height Data-set : 48 members, 4 clusters



Conclusion

- Algorithm for the computation of PD barycenters
- Two layers of progressivity
- Interruptibility
- Interactive analysis of ensembles
- Extension to topological clustering of ensembles
- Open-source implementation in the Topology Tool Kit