DSAIDIS Workshop

November 15th

Optimal alignments in machine learning:

The case of spatio-temporal data

Hicham Janati

Assistant professor, Télécom Paris

hicham.janati@telecom-paris.fr

Introduction

Any machine learning pipeline includes some form of:

1. Comparison of distributions
2. Matching of distributions
3. Averaging of distributions

Introduction

Any machine learning pipeline includes some form of:

1. Comparison of distributions
2. Matching of distributions
3. Averaging of distributions

1. Comparison of distributions

"Geometry"

Computational complexity

# of operations to compute $\mathcal{L}(\alpha_n, \beta_n)$

Sample complexity

$\mathbb{E}[\mathcal{L}(\alpha_n, \beta_n) - \mathcal{L}(\alpha, \beta)]$:

1. Comparing distributions

Kullback-Leibler (relative entropy)

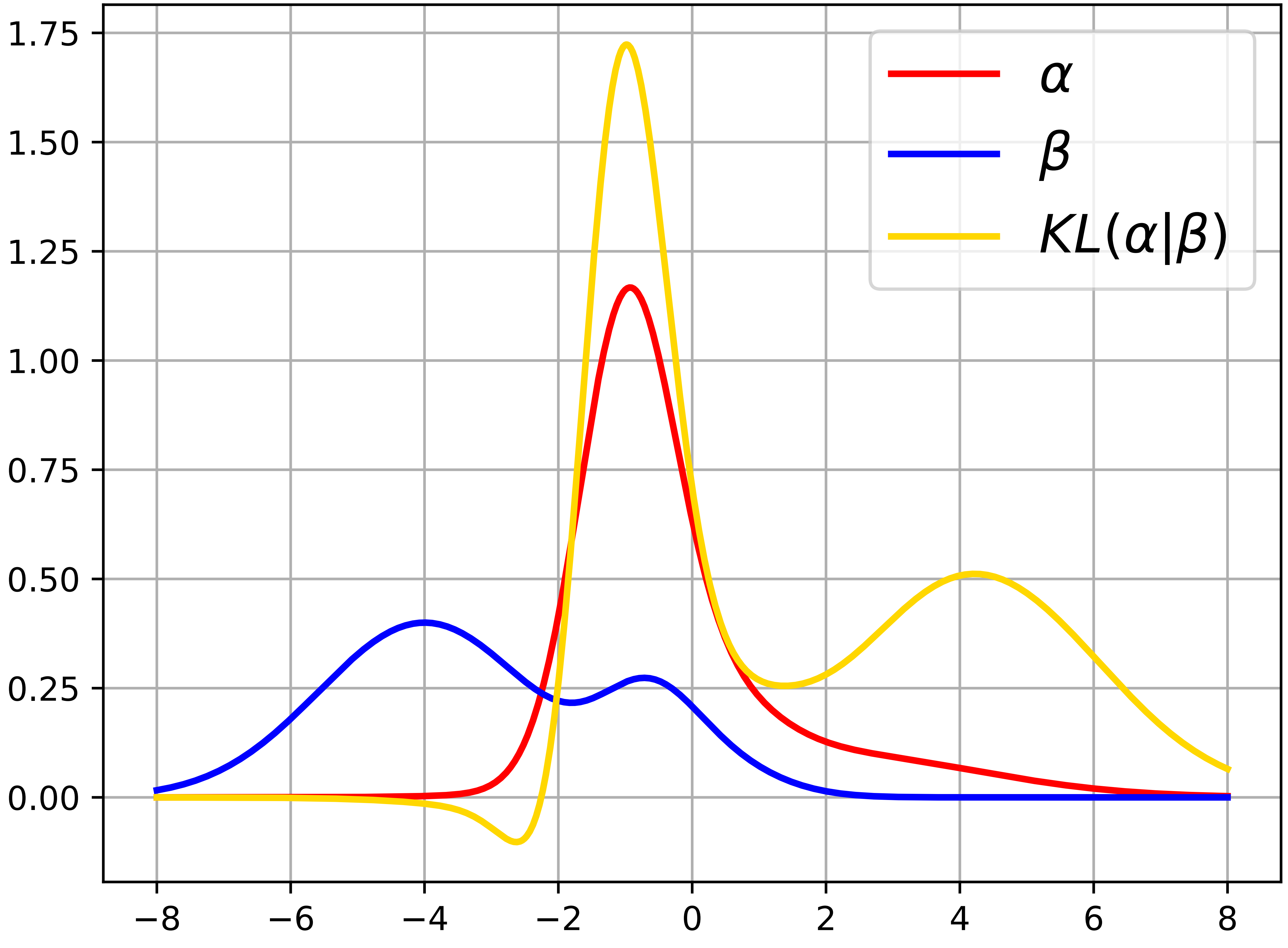
$$\mathrm{KL}(\alpha, \beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$$

"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n, \beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n, \beta_n) - \mathcal{L}(\alpha, \beta)]$:
KL —		

1. Comparing distributions

Kullback-Leibler (relative entropy)

$$\mathrm{KL}(\alpha, \beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$$



“Geometry’

mple complexity

$(\alpha_n, \beta_n) - \mathcal{L}(\alpha, \beta)]$

KL	—

1. Comparing distributions

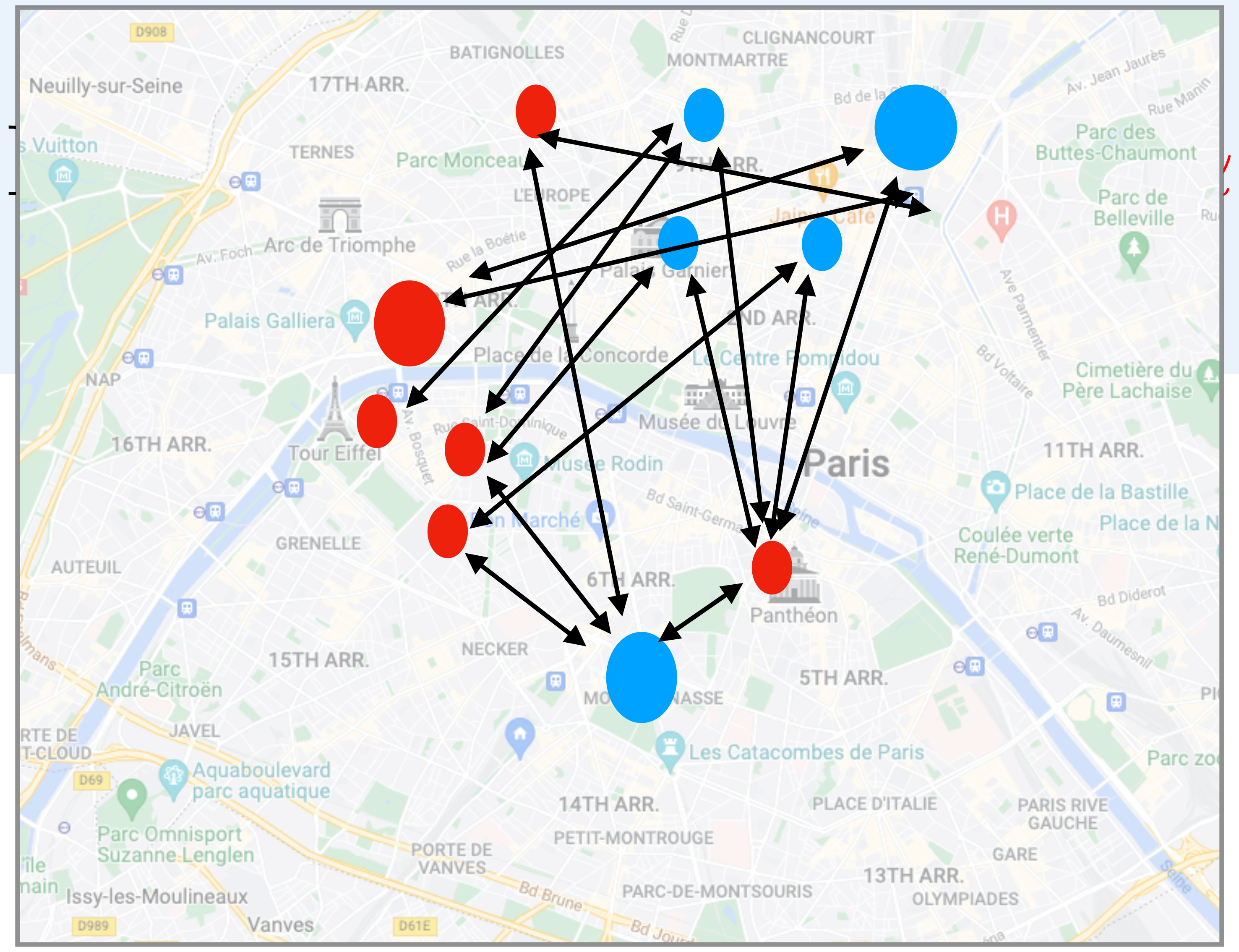
Kullback-Leibler (relative entropy)

$$\mathrm{KL}(\alpha, \beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$$

	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n, \beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n, \beta_n) - \mathcal{L}(\alpha, \beta)]$:
KL	—	$O(n)$	Ill-defined (can be $+\infty$)

1. Comparing distributions

Kullback-Leibler (relative entropy)



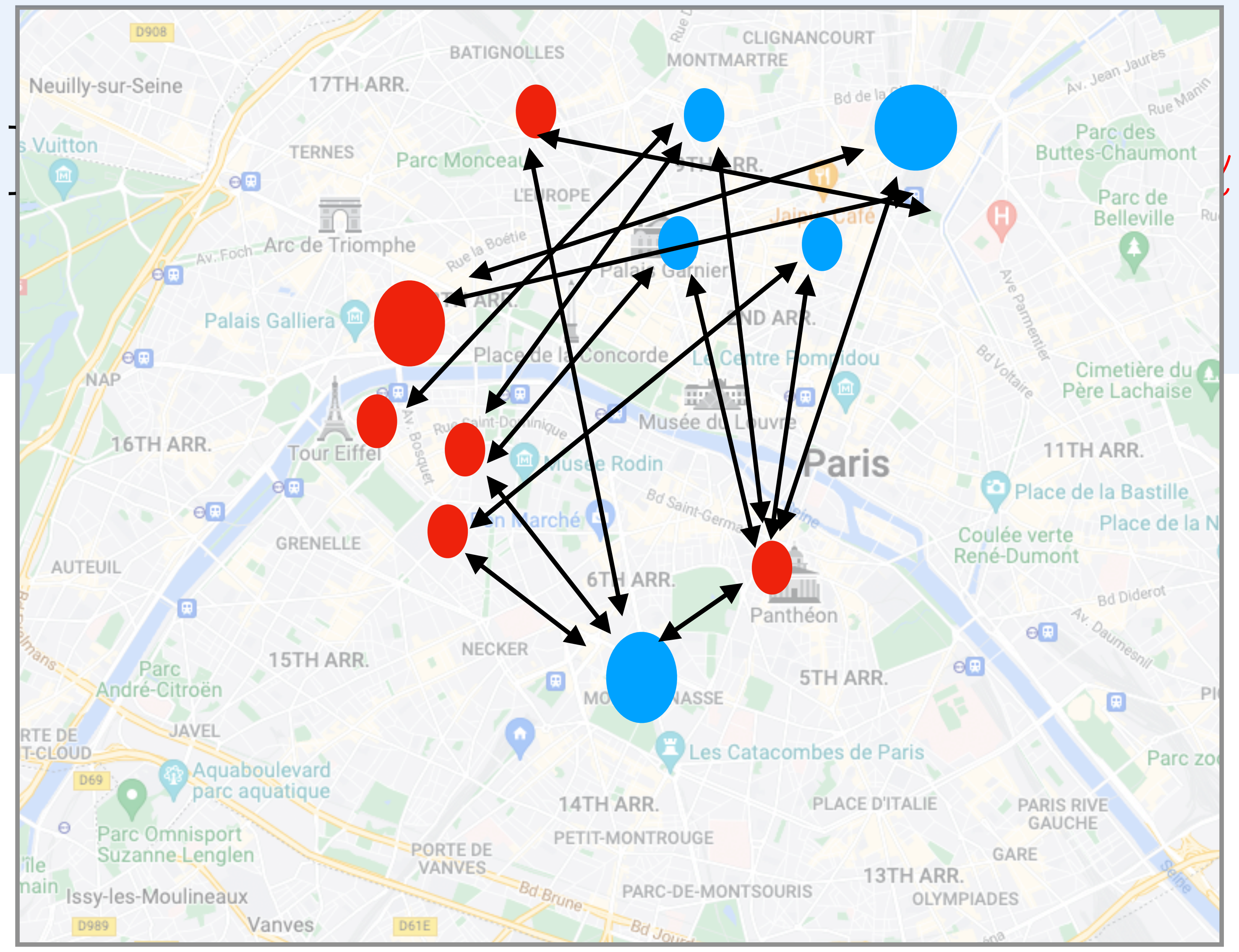
MMD norms

$$\|\alpha - \beta\|_K^2 = \iint K(x,y) d^2(\alpha - \beta)$$

"Geometry"		Computational complexity # of operations to compute $\mathcal{L}(\alpha_n, \beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n, \beta_n) - \mathcal{L}(\alpha, \beta)]$
KL	—	$O(n)$	Ill-defined (can be $+\infty$)

1. Comparing distributions

Kullback-Leibler (relative entropy)



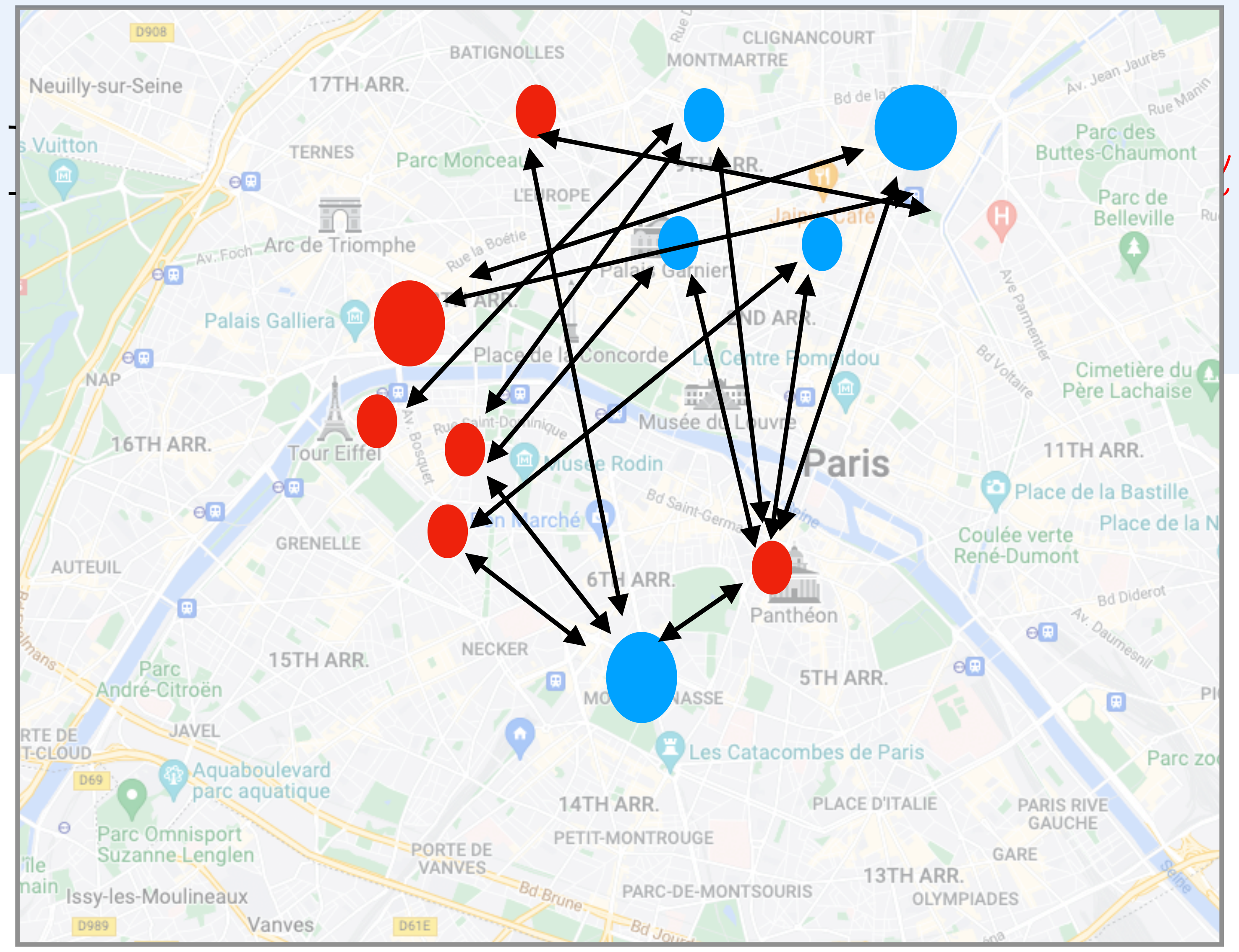
MMD norms

$$\|\alpha-\beta\|_K^2 = \iint K(x,y)\,d^2(\alpha-\beta)$$

	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n, \beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n, \beta_n) - \mathcal{L}(\alpha, \beta)]$
KL	—	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$

1. Comparing distributions

Kullback-Leibler (relative entropy)



MMD norms

$$\|\alpha - \beta\|_K^2 = \iint K(x, y) d^2(\alpha - \beta)$$

Idea: Optimize over interactions to find the “best” pairs to compare ?

“Geometry”

Computational complexity

# of operations to compute $\mathcal{L}(\alpha_n, \beta_n)$

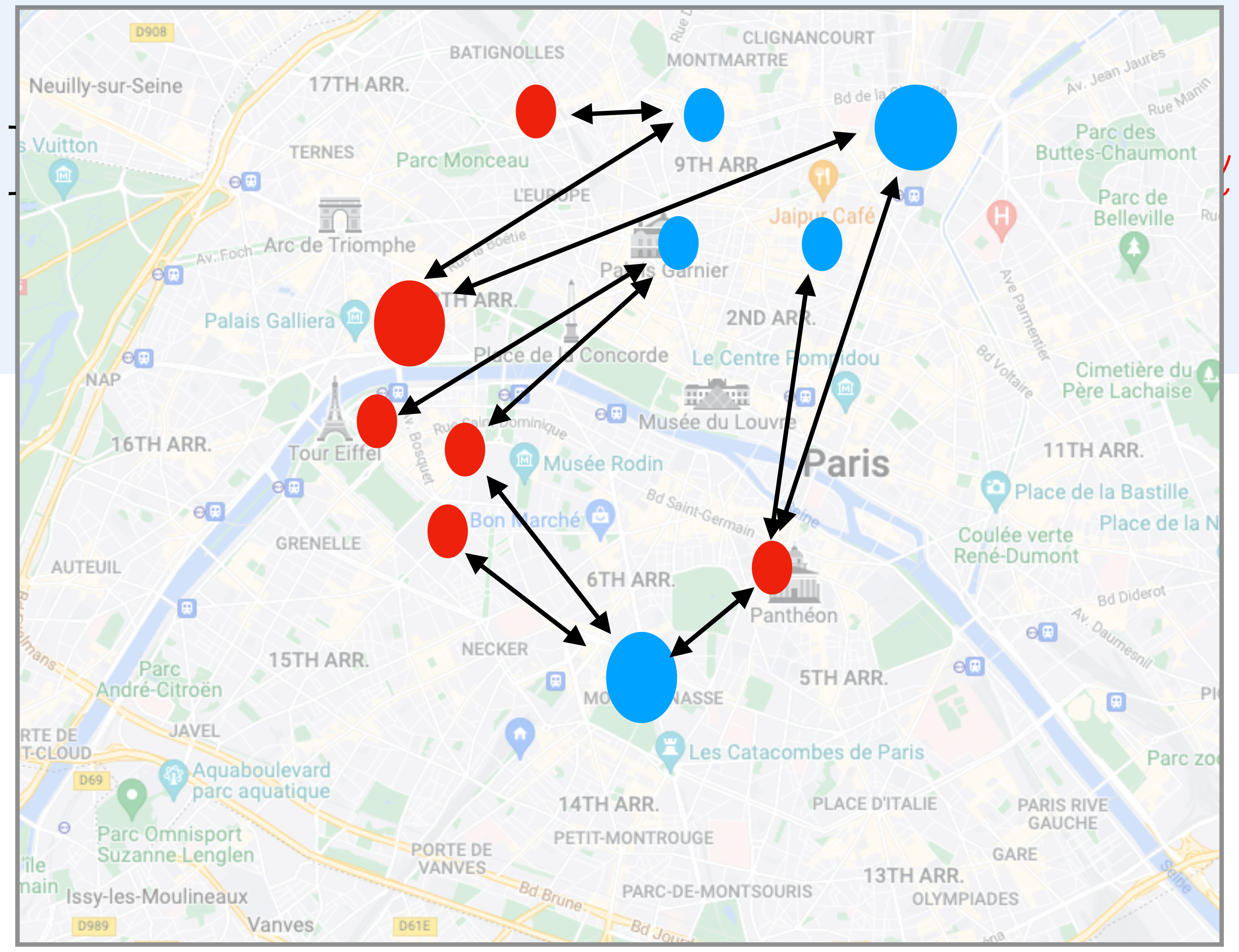
Sample complexity

$\mathbb{E}[\mathcal{L}(\alpha_n, \beta_n) - \mathcal{L}(\alpha, \beta)]$

KL	—	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$

1. Comparing distributions

Kullback-Leibler (relative entropy)



MMD norms

$$\|\alpha - \beta\|_K^2 = \iint K(x,y)d^2(\alpha-\beta)$$

Idea: Optimize over interactions to find the "best" pairs to compare ?

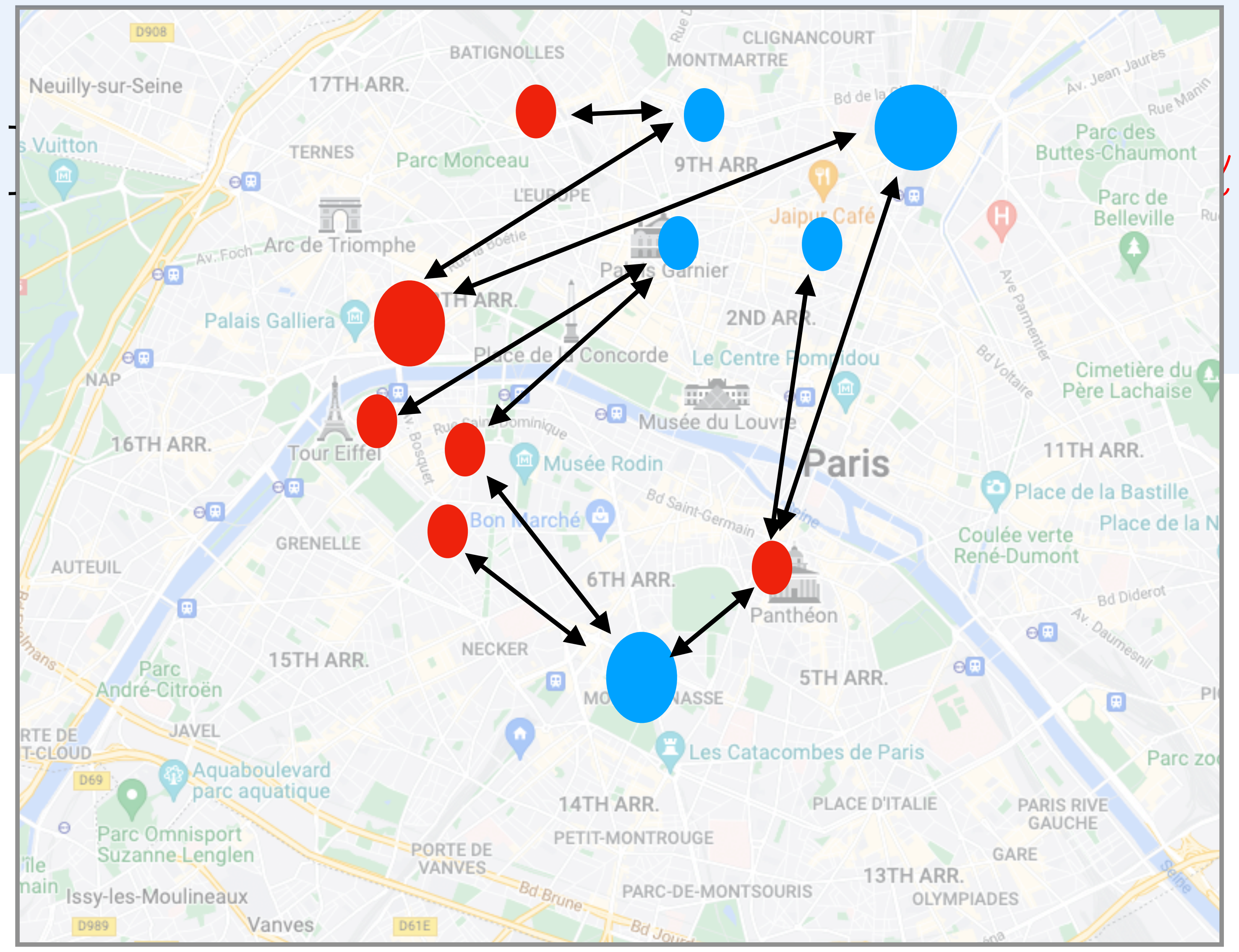
	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n, \beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n, \beta_n) - \mathcal{L}(\alpha, \beta)]$
KL	—	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$

1. Comparing distributions

Optimal Transport

$$\mathrm{OT}(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha,\pi_2=\beta}} \iint C d\pi$$

Kullback-Leibler (relative entropy)



MMD norms

$$\|\alpha-\beta\|_K^2 = \iint K(x,y) d^2(\alpha-\beta)$$

Idea: Optimize over interactions to find the "best" pairs to compare ?

"Geometry"		Computational complexity # of operations to compute $\mathcal{L}(\alpha_n,\beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n,\beta_n)-\mathcal{L}(\alpha,\beta)]$
KL	—	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$

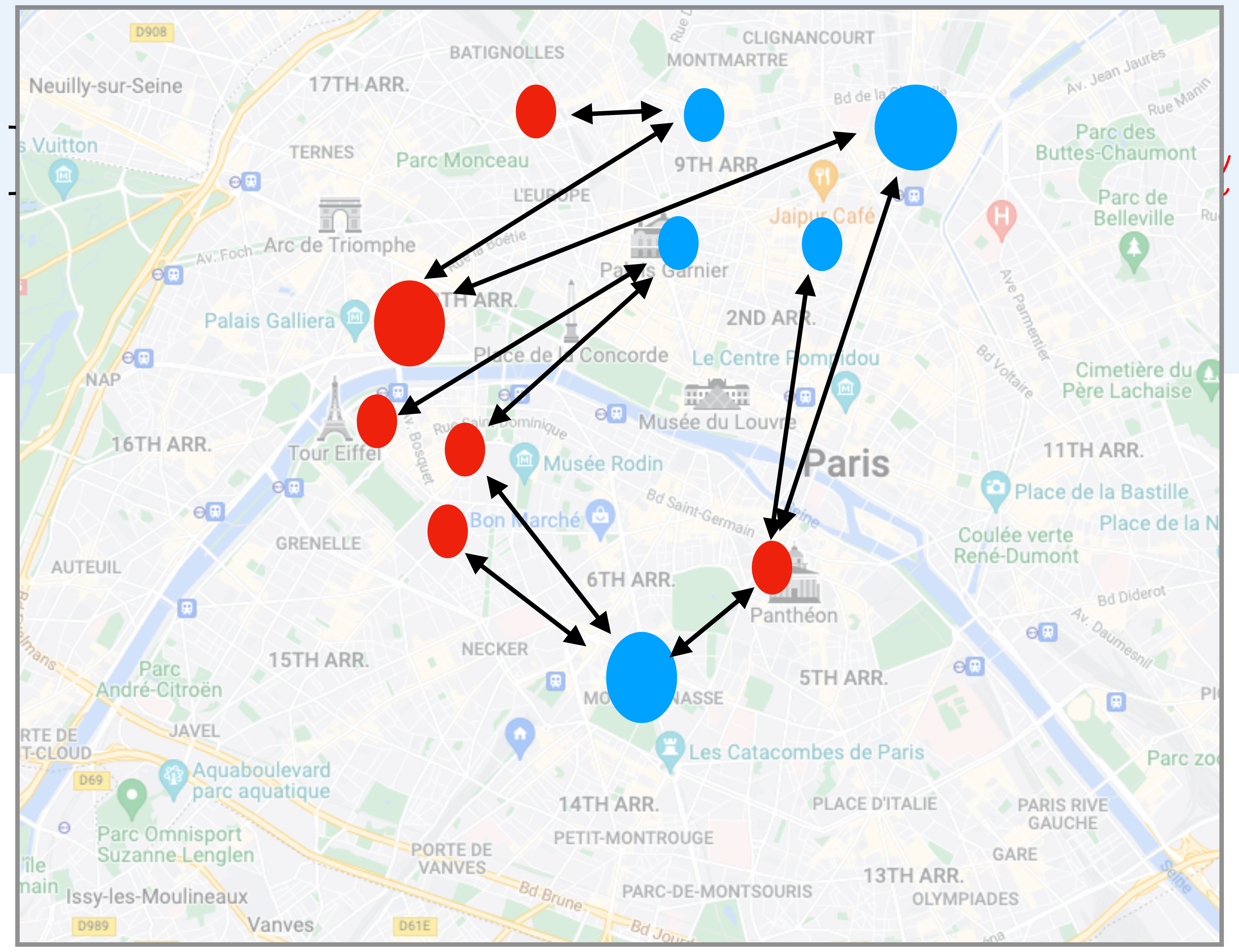
1. Comparing distributions

Optimal Transport

$$\mathrm{OT}(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha,\pi_2=\beta}} \iint C d\pi$$

Discrete case:
$$\min_{P\geq 0} \sum_{i=1}^{6} \sum_{j=A}^{B} C_{ij} P_{ij}$$
$$P\mathbb{1} = \alpha$$
$$P^\top \mathbb{1} = \beta$$

Kullback-Leibler (relative entropy)



MMD norms

$$\|\alpha-\beta\|_K^2 = \iint K(x,y)d^2(\alpha-\beta)$$

Idea: Optimize over interactions to find the "best" pairs to compare ?

"Geometry"

Computational complexity

\# of operations to compute $\mathcal{L}(\alpha_n,\beta_n)$

Sample complexity

$\mathbb{E}[\mathcal{L}(\alpha_n,\beta_n) - \mathcal{L}(\alpha,\beta)]$

KL	—	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$

$$\min_{\substack{P \geq 0 \\ P\mathbb{1} = \alpha \\ P^\top \mathbb{1} = \beta}} \sum_{i=1}^{6} \sum_{j=A}^{B} C_{ij} P_{ij}$$

Goal: Estimate a *2D mass distribution*

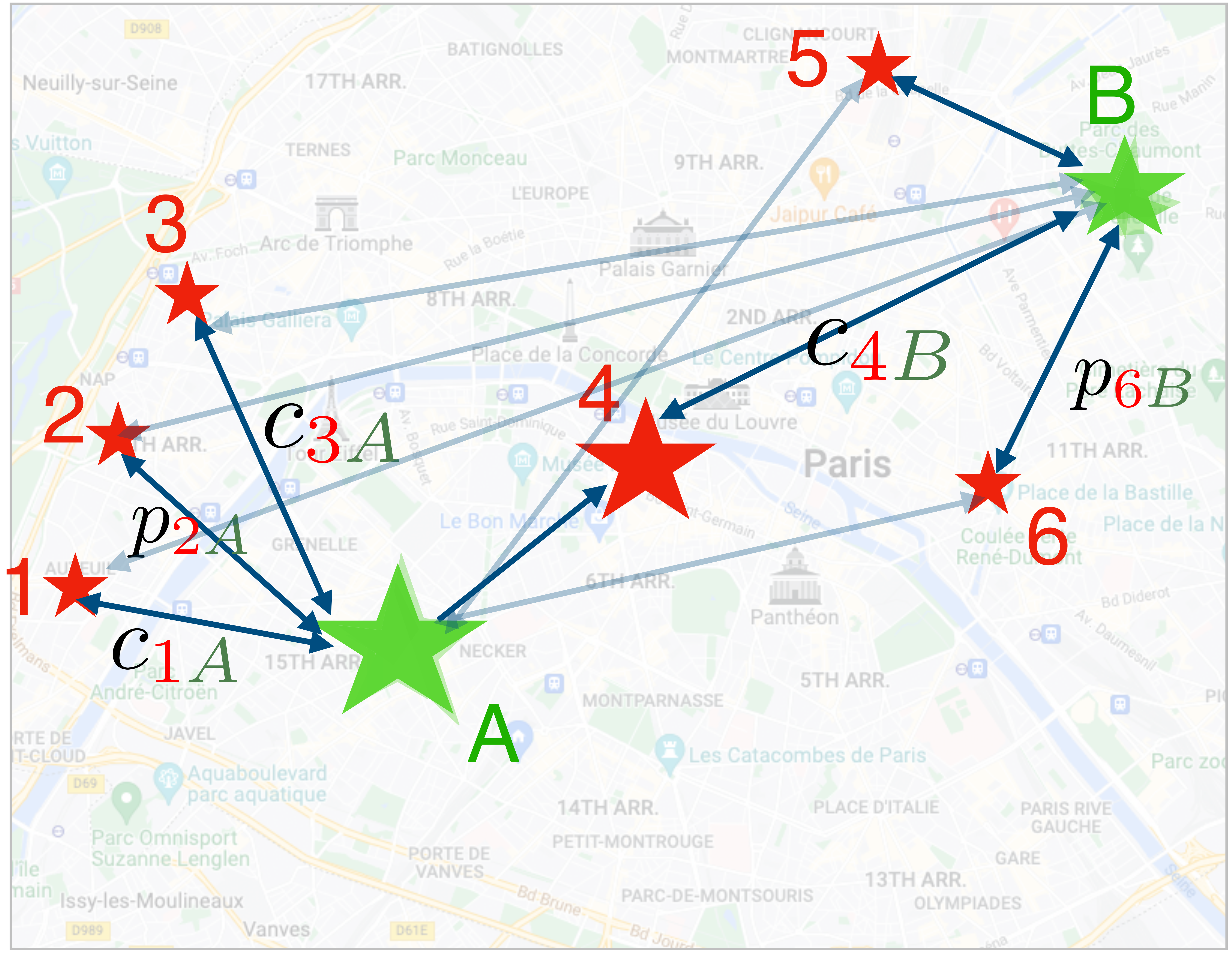
" how much goes where"

Costs / distances matrix

$$\begin{pmatrix} c_{1A} & c_{1B} \\ c_{2A} & c_{2B} \\ \vdots & \vdots \\ c_{6A} & c_{6B} \end{pmatrix}$$

Unknown transport plan

$$\begin{pmatrix} p_{1A} & p_{1B} \\ p_{2A} & p_{2B} \\ \vdots & \vdots \\ p_{6A} & p_{6B} \end{pmatrix}$$



$$\min_{\substack{P \geq 0 \\ P\mathbb{1} = \alpha \\ P^\top \mathbb{1} = \beta}} \sum_{i=1}^{6} \sum_{j=A}^{B} C_{ij} P_{ij}$$

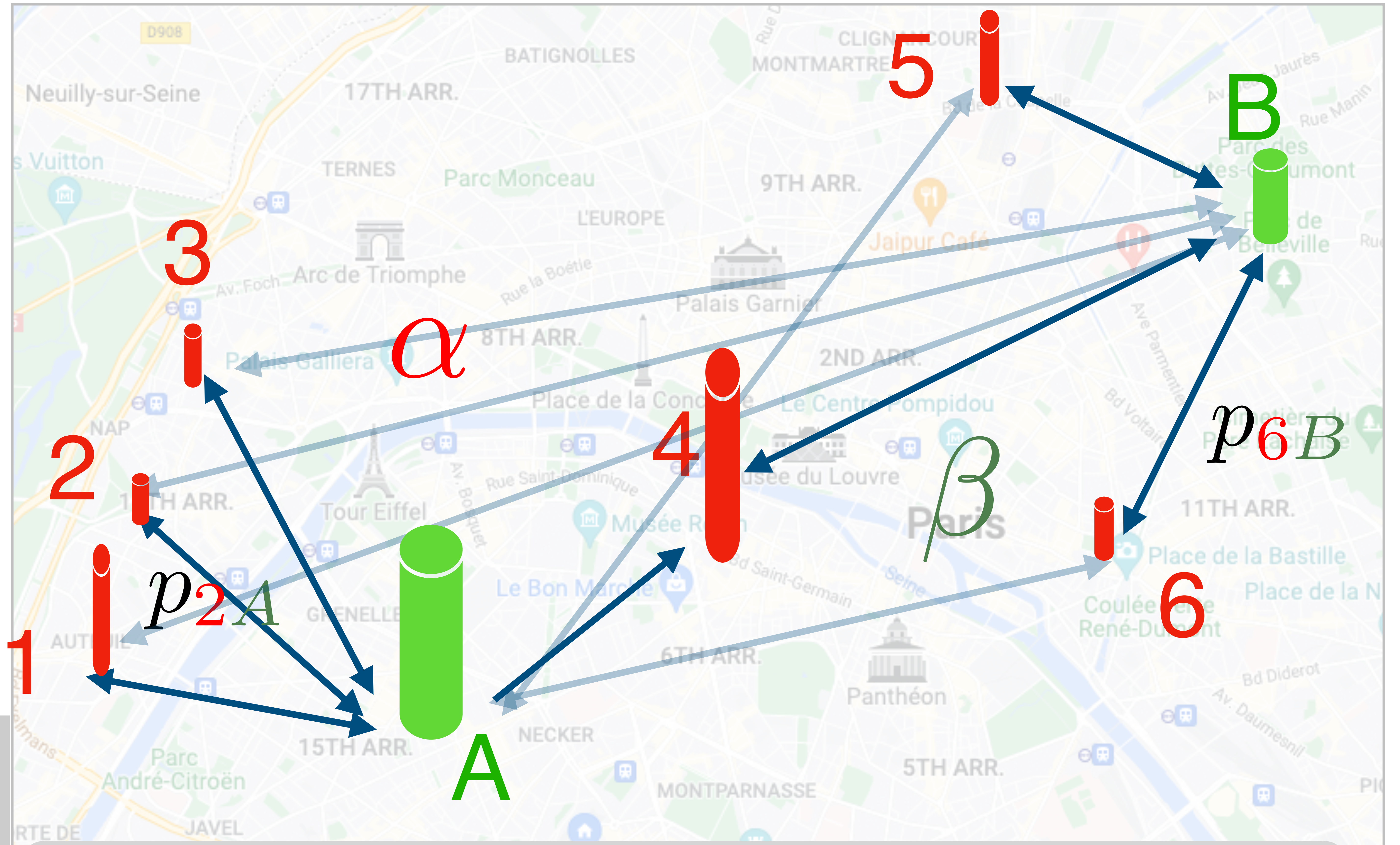
mass conservation constraints

Costs / distances matrix

$$\begin{pmatrix} c_{1A} & c_{1B} \\ c_{2A} & c_{2B} \\ \vdots & \vdots \\ c_{6A} & c_{6B} \end{pmatrix}$$

Unknown transport plan

$$\begin{pmatrix} p_{1A} & p_{1B} \\ p_{2A} & p_{2B} \\ \vdots & \vdots \\ p_{6A} & p_{6B} \end{pmatrix}$$



P is a coupling between α and β

i.e: 2D distribution with marginals α and β

Introduction

Optimal Transport	Kullback-Leibler (relative entropy)	MMD norms
$\mathrm{OT}(\alpha, \beta) = \min_{\substack{\pi \\ \pi_1=\alpha, \pi_2=\beta}} \iint C d\pi$	$\mathrm{KL}(\alpha, \beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$	$\|\alpha - \beta\|_K^2 = \iint K(x, y) d^2(\alpha - \beta)$

	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n, \beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n, \beta_n) - \mathcal{L}(\alpha, \beta)]$:
KL	$-$	$O(n)$	Ill-defined (can be $+\infty$)
MMD	$+$	$O(n^2)$	$O(n^{-1})$

Introduction

Optimal Transport	Kullback-Leibler (relative entropy)	MMD norms
$\mathrm{OT}(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha, \pi_2=\beta}} \iint C d\pi$	$\mathrm{KL}(\alpha,\beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$	$\|\alpha-\beta\|_K^2 = \iint K(x,y) d^2(\alpha-\beta)$

	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n,\beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n,\beta_n) - \mathcal{L}(\alpha,\beta)]$:
KL	$-$	$O(n)$	Ill-defined (can be $+\infty$)
MMD	$+$	$O(n^2)$	$O(n^{-1})$
OT	$+++$	$O(n^3)$	$O(n^{-\frac{2}{d}})$

Introduction

Optimal Transport

$$\mathrm{OT}(\alpha, \beta) = \min_{\substack{\pi \\ \pi_1=\alpha, \pi_2=\beta}} \iint C d\pi$$

Compares the distributions via finding an underlying alignment

Kullback-Leibler (relative entropy)

$$\mathrm{KL}(\alpha, \beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$$

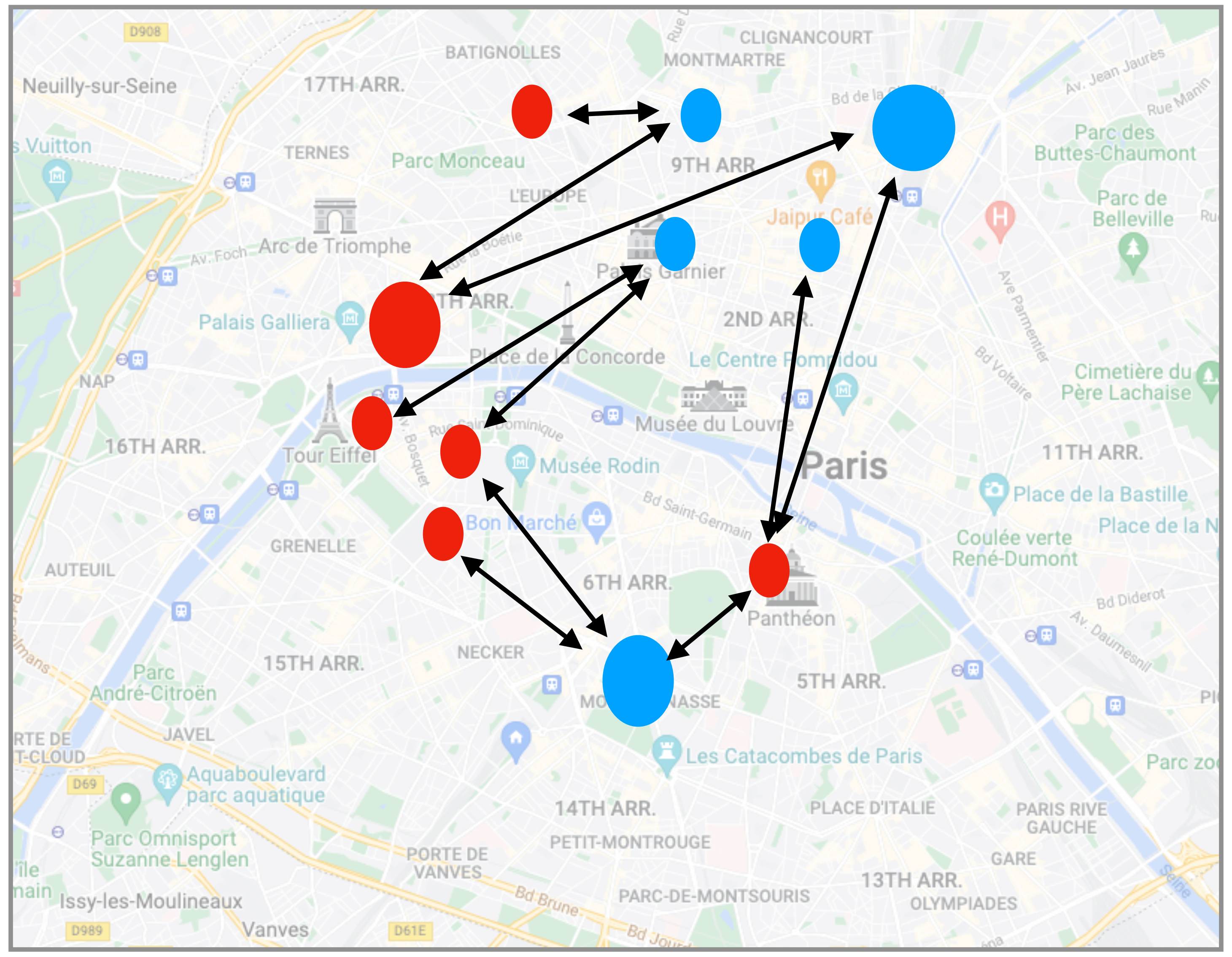
MMD norms

$$\|\alpha - \beta\|_K^2 = \iint K(x, y) d^2(\alpha - \beta)$$

	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n, \beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n, \beta_n) - \mathcal{L}(\alpha, \beta)]$
KL	–	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$
OT	+++	$O(n^3)$	$O(n^{-\frac{2}{d}})$

2. Matching of distributions

Optimal transport finds
a smooth / probabilistic alignment
using some underlying geometry



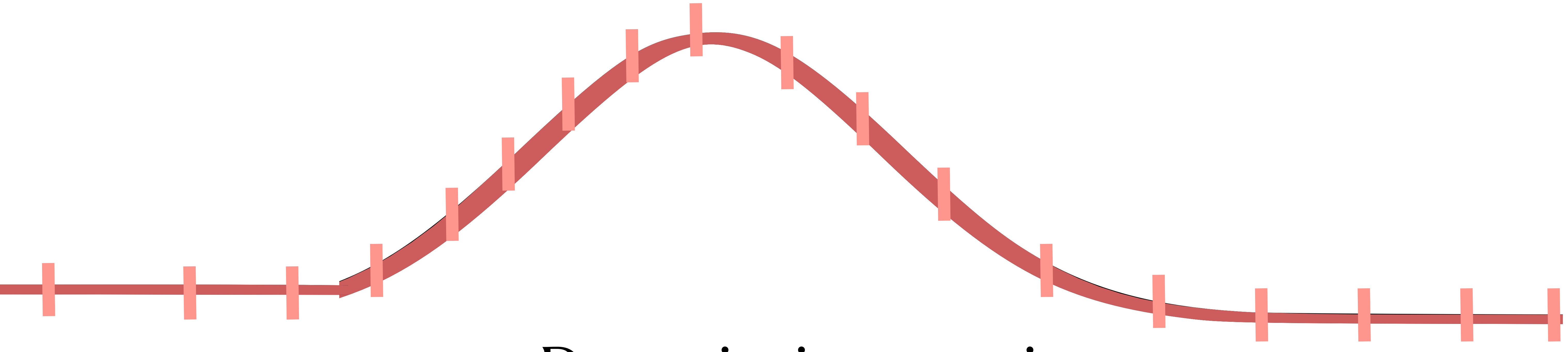
Can this matching be used for time series data ?

2. Matching distributions

Time series alignment: DTW

2. Matching distributions

Time series alignment: DTW



Dynamic time warping

[Sakoe & Chiba, 1972]

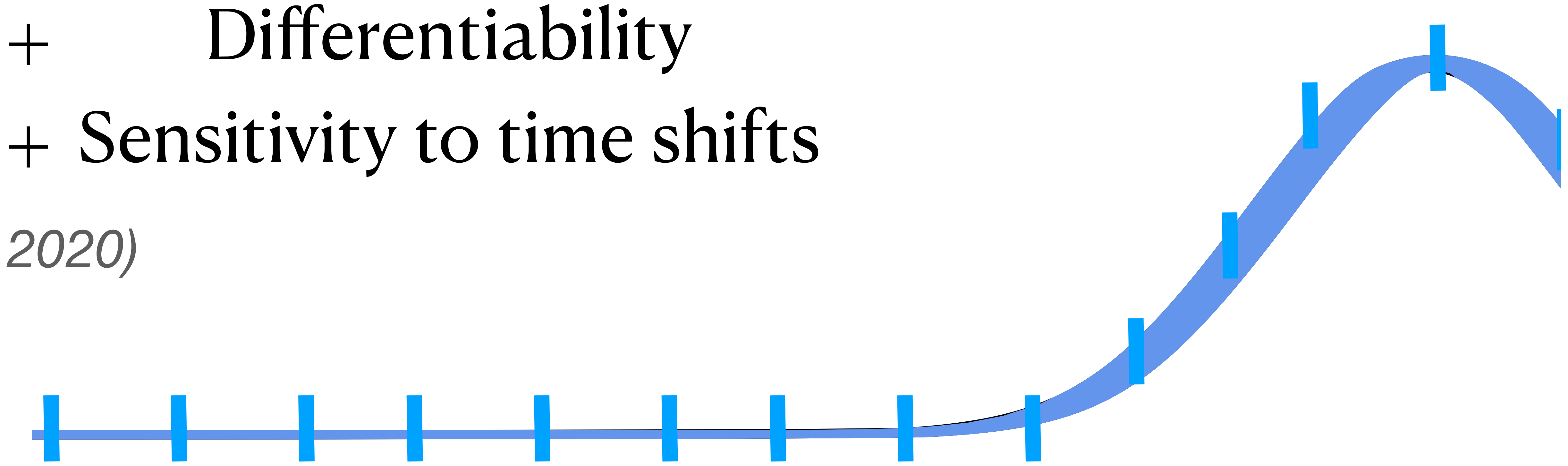
$$\mathrm{dtw}_{\Delta}(x, y) = \min_{A \in \mathcal{A}} \sum_{i,j} A_{ij}\, \Delta(x_i, y_j)$$

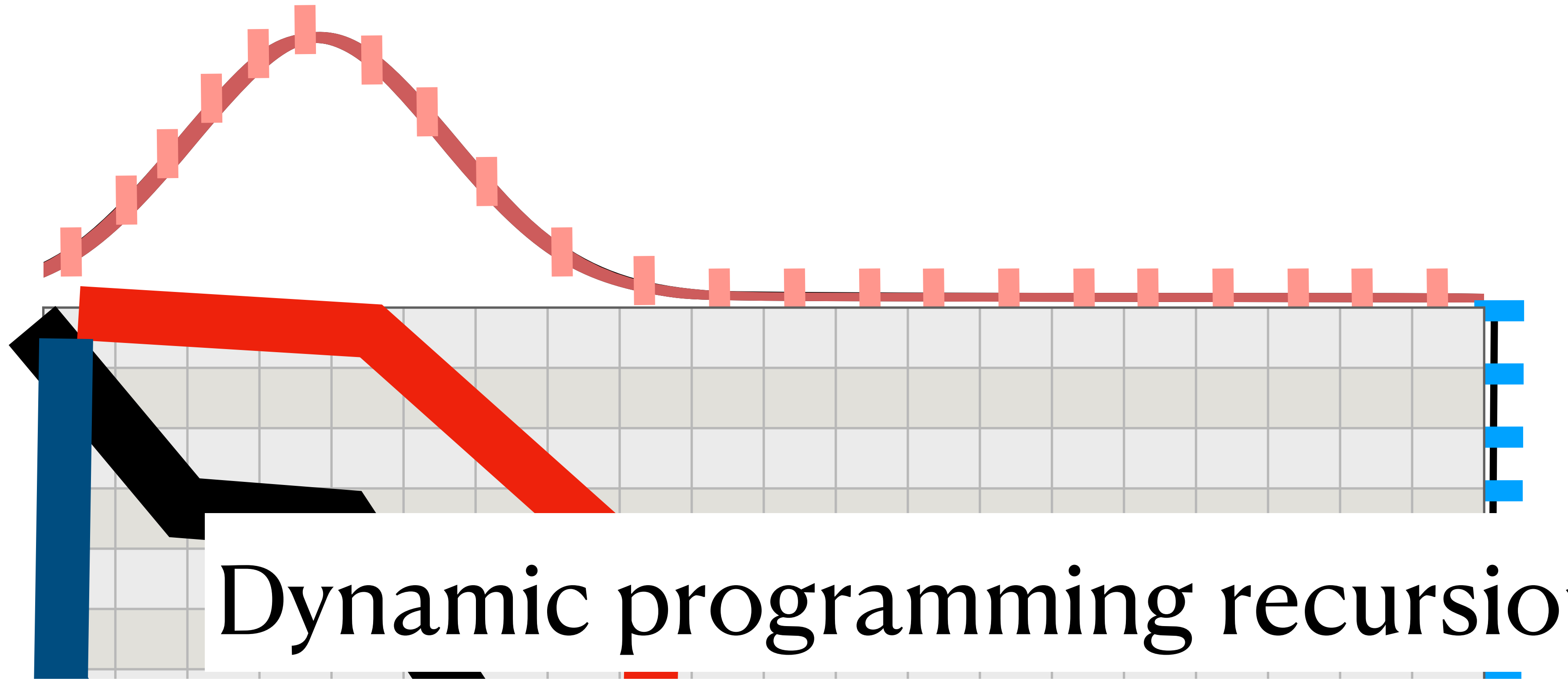
(Cuturi & Blondel, 2017) Softmin

\+ Differentiability

\+ Sensitivity to time shifts

(Janati et al., 2020)





Dynamic programming recursion

Algorithm 1 Forward recursion to compute $\mathbf{dtw}_{\gamma}(\mathbf{x}, \mathbf{y})$ and intermediate alignment costs

Inputs: $\mathbf{x}, \mathbf{y}$, smoothing $\gamma \geq 0$, distance function δ
$r_{0,0} = 0; r_{i,0} = r_{0,j} = \infty; i \in [\![n]\!], j \in [\![m]\!]$
for $j = 1, \ldots, m$ **do**
  for $i = 1, \ldots, n$ **do**
    $r_{i,j} = \delta(x_i, y_j) + \min^{\gamma}\{r_{i-1,j-1}, r_{i-1,j}, r_{i,j-1}\}$
  end for
end for
Output: $(r_{n,m}, R)$

Small recap

1. Comparing distributions

Optimal Transport

$$\mathrm{OT}(\alpha, \beta) = \min_{\substack{\pi \\ \pi_1 = \alpha, \pi_2 = \beta}} \iint C d\pi$$

Kullback-Leibler (relative entropy)

$$\mathrm{KL}(\alpha, \beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$$

MMD norms

$$\|\alpha - \beta\|_K^2 = \iint K(x, y) d^2(\alpha - \beta)$$

2. Matching distributions

Optimal Transport

$$\mathrm{OT}(\alpha, \beta) = \min_{\substack{\pi \\ \pi_1 = \alpha, \pi_2 = \beta}} \iint C d\pi$$

Matching time series data (dtw)

$$\mathrm{dtw}_\Delta(x, y) = \min_{A \in \mathcal{A}} \sum_{i,j} A_{ij}\, \Delta(x_i, y_j)$$

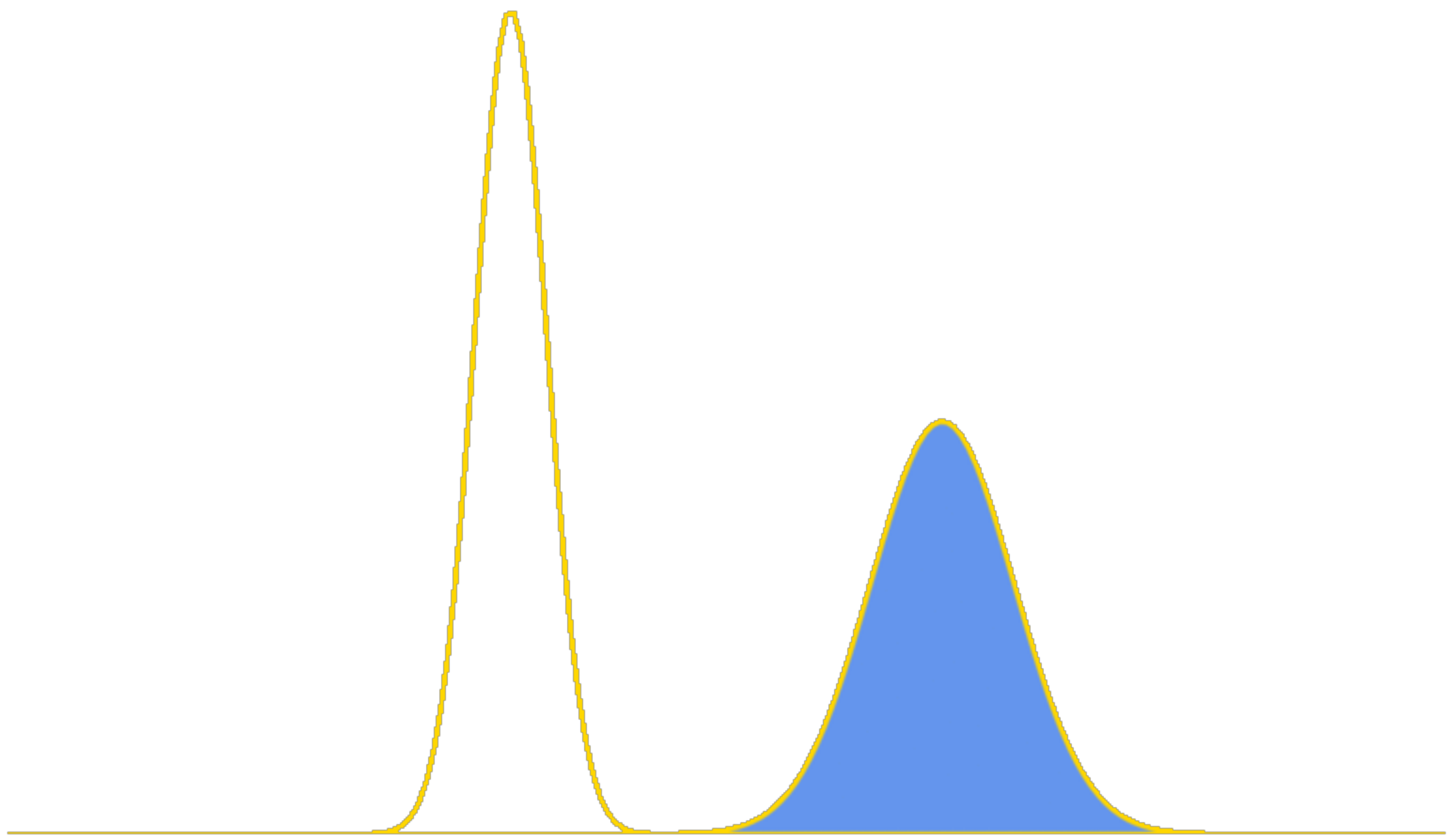
3. Averaging distributions

For any distance above, define the Fréchet mean:

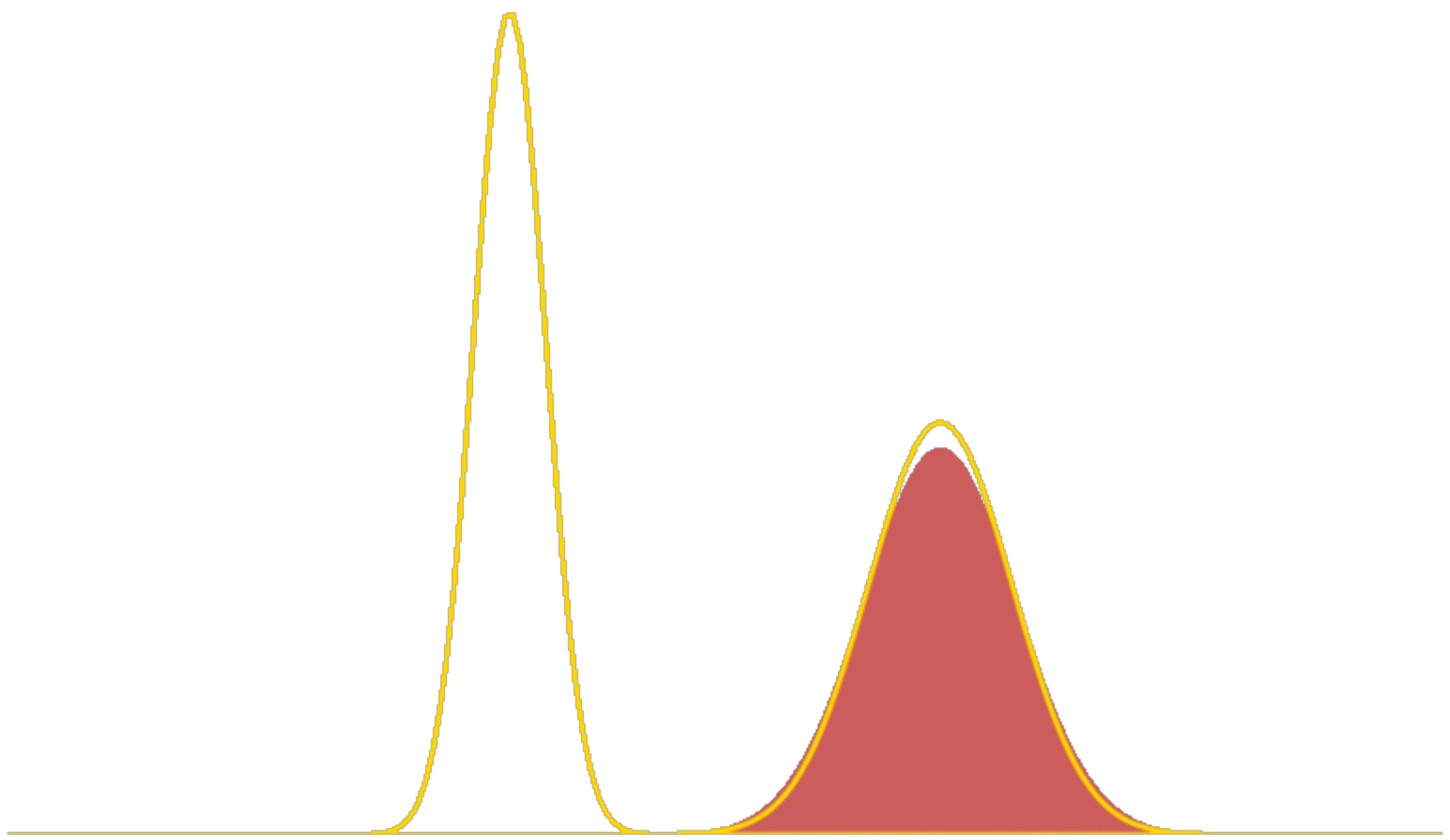
$$\arg\min_{\beta} \sum_{k=1}^{K} w_k \mathrm{Dist}(\alpha_k, \beta)$$

$$\arg\min_{\beta} \sum_{k=1}^{K} w_k \mathrm{Dist}(\alpha_k, \beta)$$

No geometry
(Euclidean / KL)



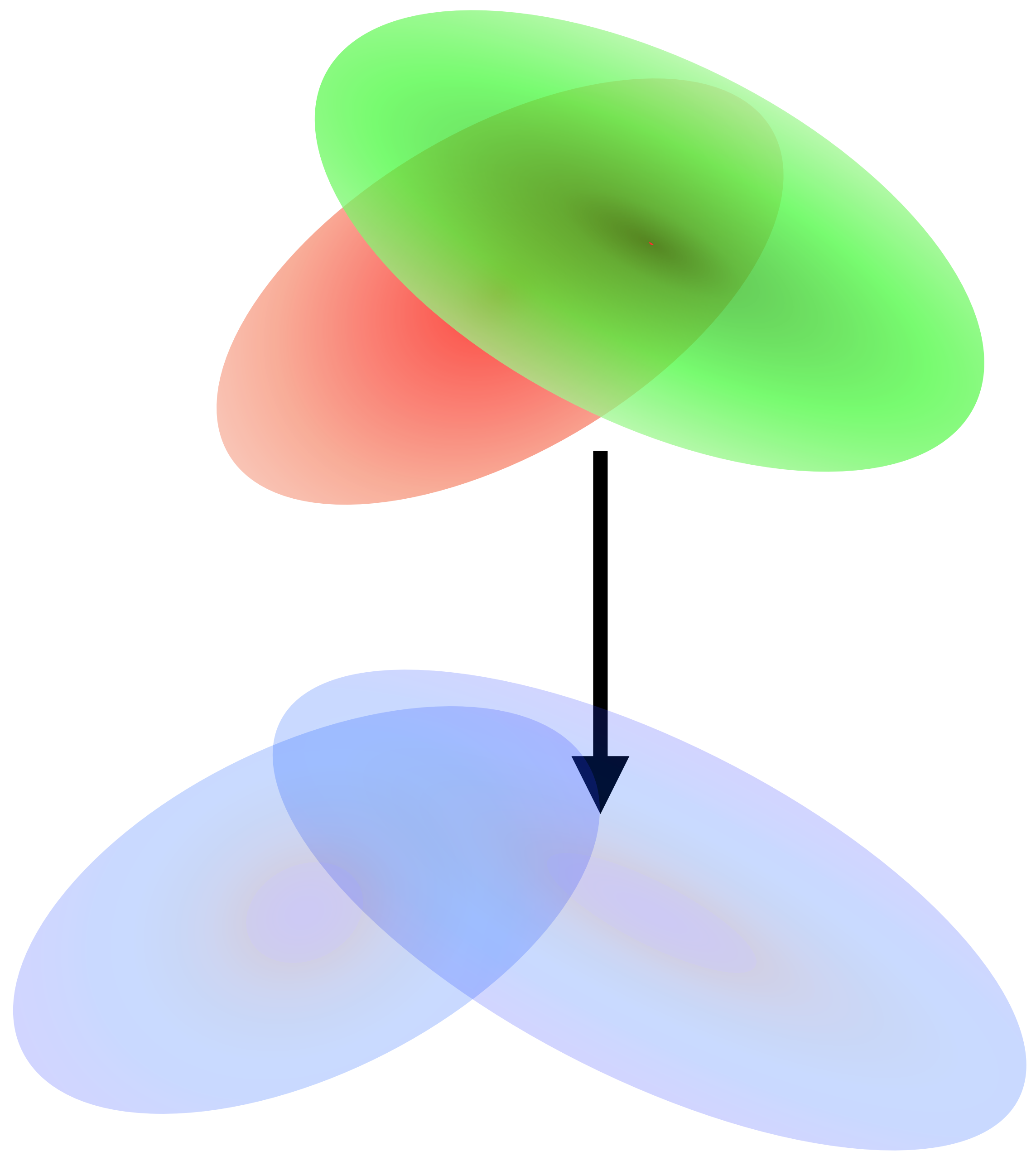
Geometry
(OT)



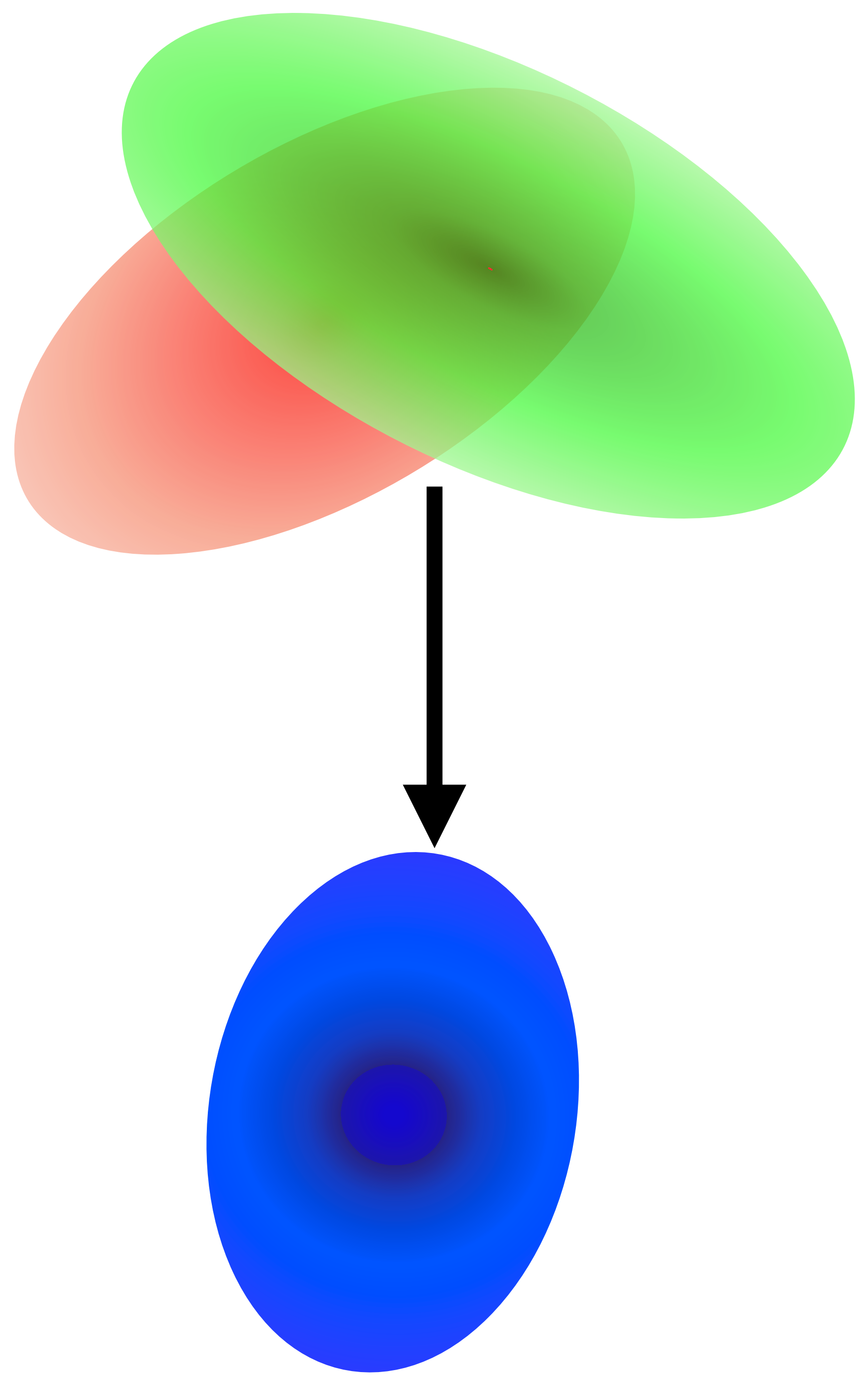
3. Averaging distributions

$$\underset{\beta}{\arg\min} \sum_{k=1}^{K} w_k \mathrm{Dist}(\alpha_k, \beta)$$

No geometry
(Euclidean / KL)

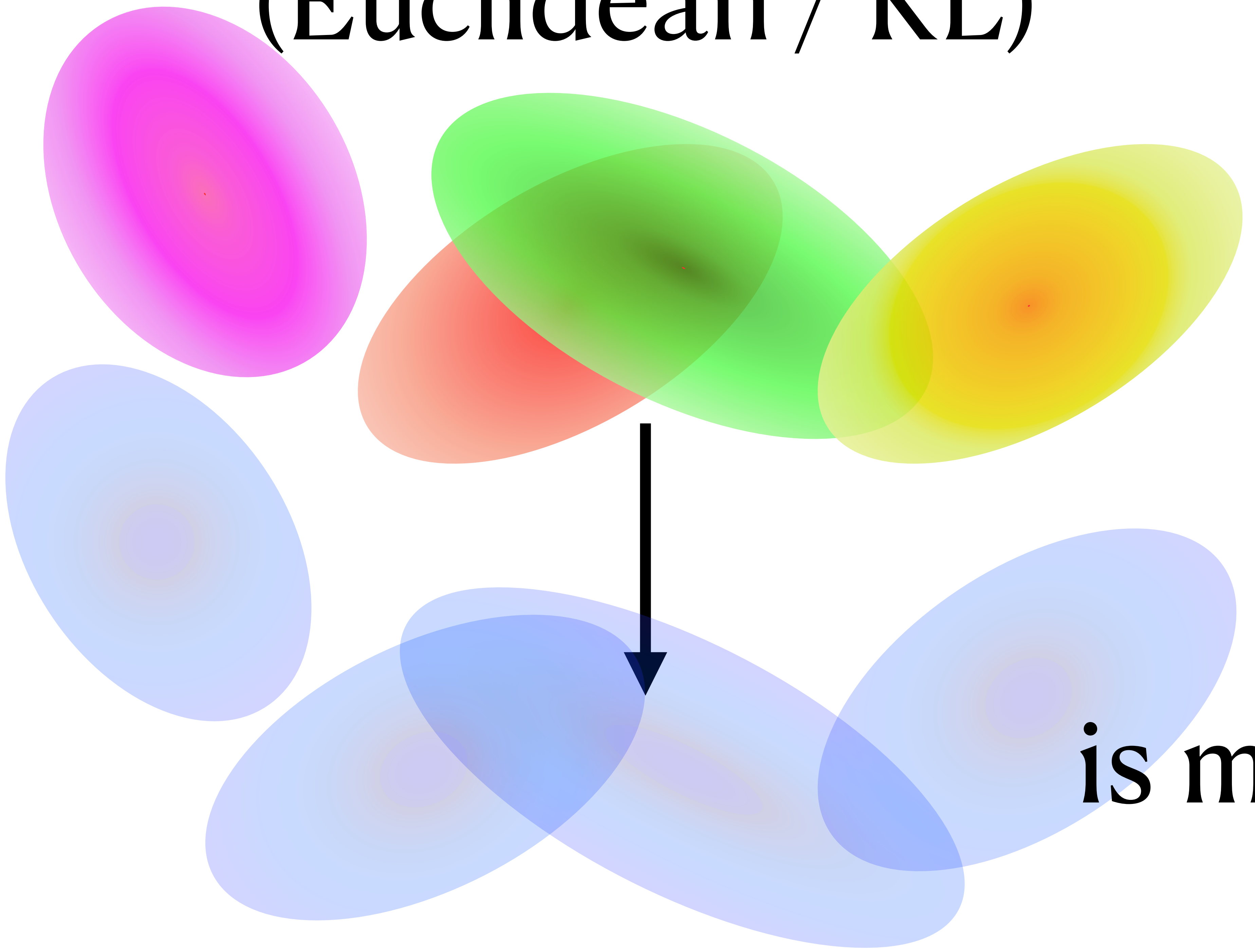


Geometry
(OT)

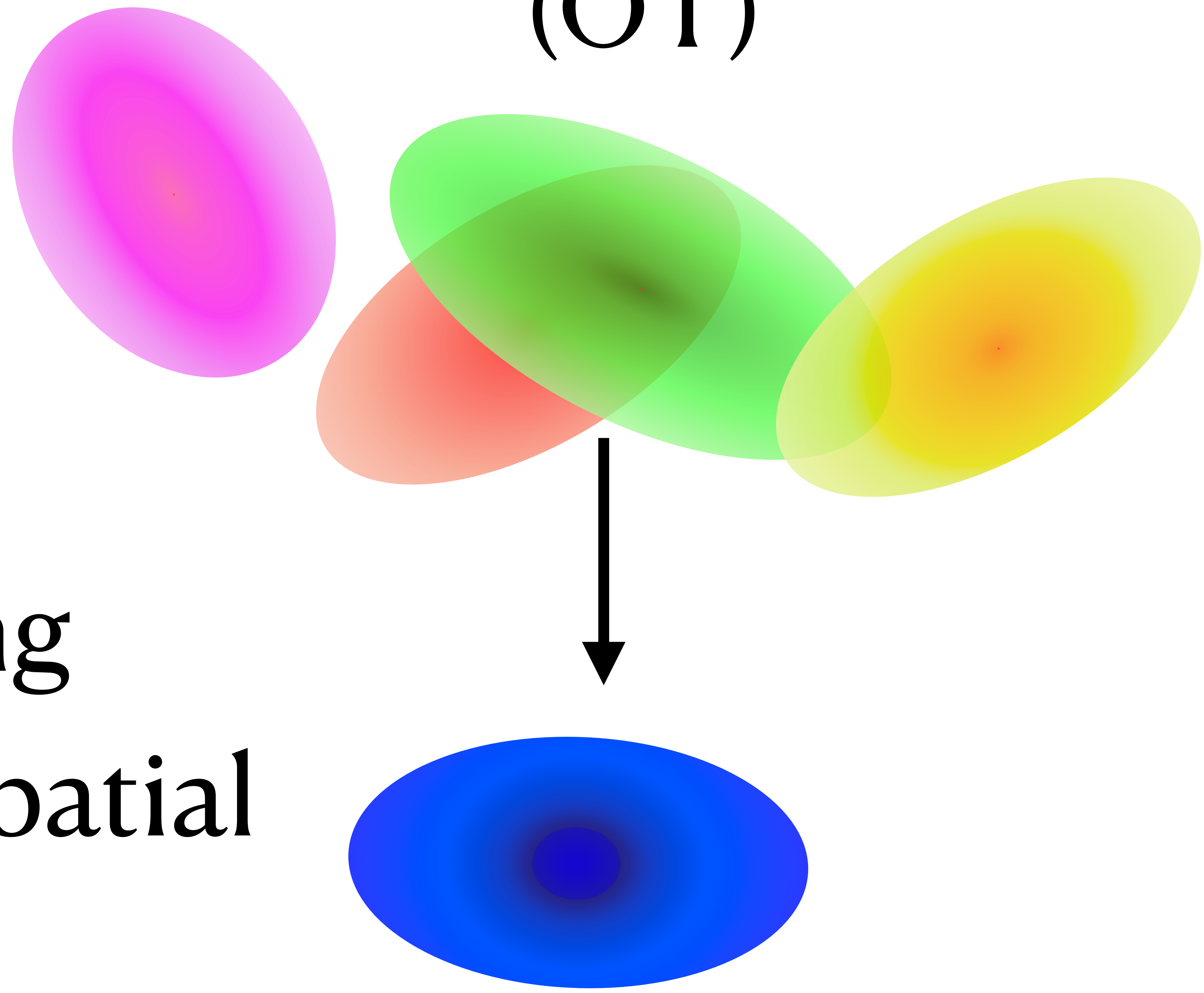


$$\arg\min_{\beta} \sum_{k=1}^{K} w_k \mathrm{Dist}(\alpha_k, \beta)$$

No geometry
(Euclidean / KL)



Geometry
(OT)



OT averaging
is minimizing "spatial
variance"

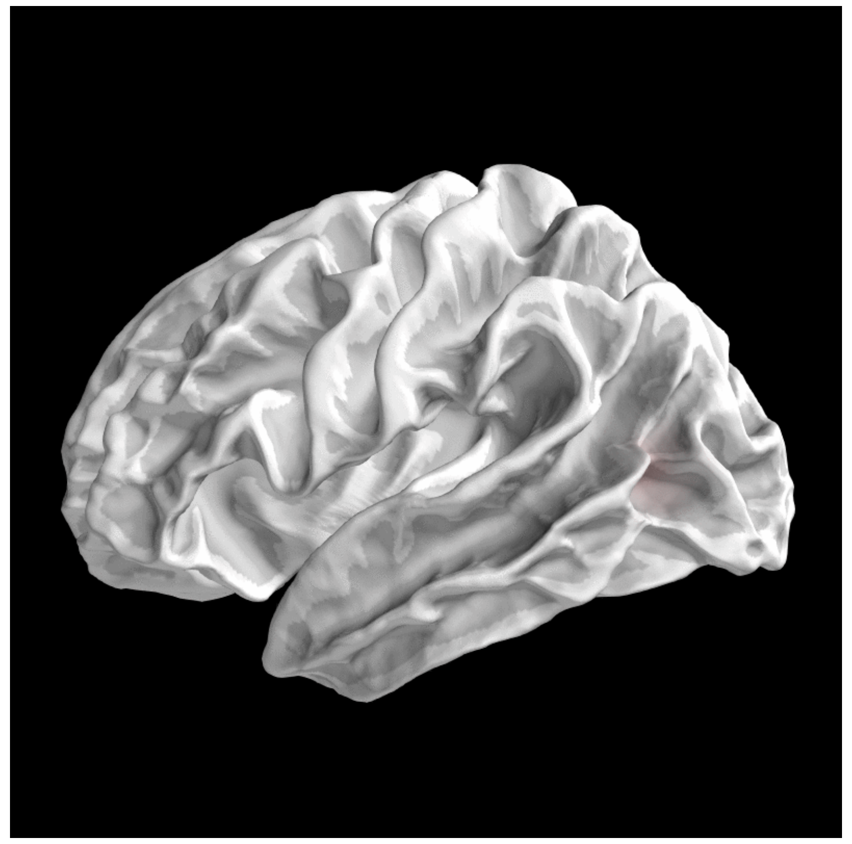
Similarly, averaging time series we can use:

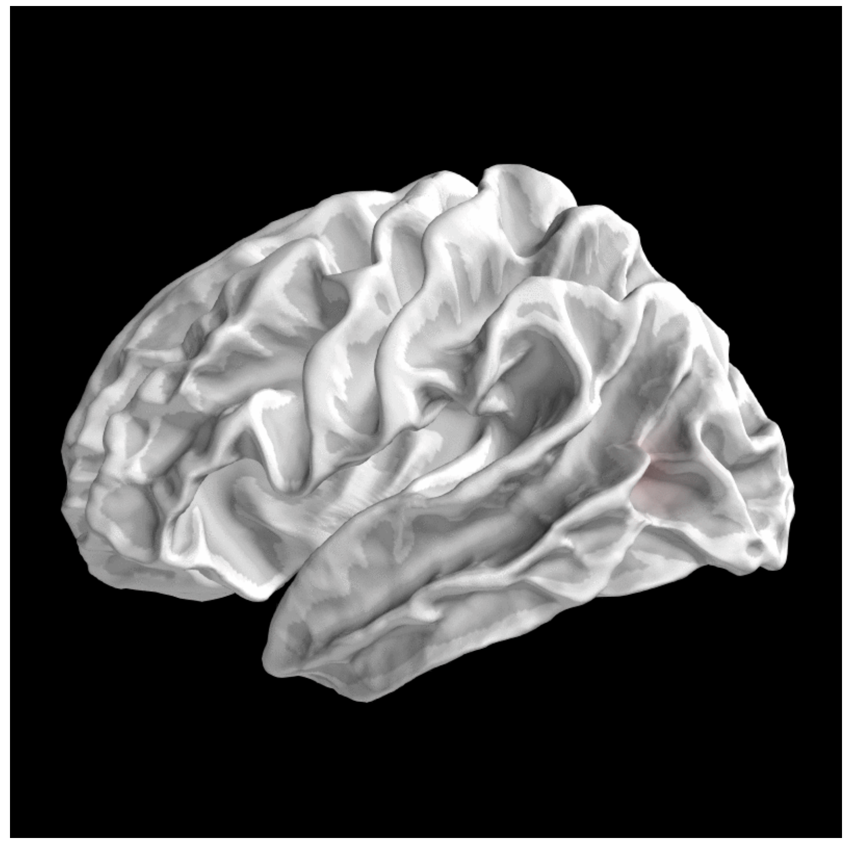
$$\arg\min_{\beta} \sum_{k=1}^{K} w_k \mathrm{Dist}(\alpha_k, \beta)$$

With Dist = dynamic time warping: $\mathrm{dtw}_\Delta(x, y) = \min_{A \in \mathcal{A}} \sum_{i,j} A_{ij}\, \Delta(x_i, y_j)$

What about averaging spatio-temporal data ?

Use DTW with a cost $\Delta(x_i, y_j)$ defined through optimal transport





[Janati et al, 2022]

Closing the loop

Optimal Transport	Kullback-Leibler (relative entropy)	MMD norms
$\mathrm{OT}(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha,\pi_2=\beta}} \iint C d\pi$	$\mathrm{KL}(\alpha,\beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$	$\|\alpha-\beta\|_K^2 = \iint K(x,y) d^2(\alpha-\beta)$

	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n,\beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n,\beta_n)-\mathcal{L}(\alpha,\beta)]$:
KL	−	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$
OT	+++	$O(n^3)$	$O(n^{-\frac{2}{d}})$

Closing the loop

Optimal Transport	Kullback-Leibler (relative entropy)	MMD norms
$\mathrm{OT}(\alpha,\beta) = \min_{\pi_1=\alpha,\pi_2=\beta} \iint C d\pi$	$\mathrm{KL}(\alpha,\beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$	$\|\alpha-\beta\|_K^2 = \iint K(x,y) d^2(\alpha-\beta)$

Entropy-regularized OT $\mathrm{OT}_\varepsilon(\alpha,\beta) = \min_{\pi_1=\alpha,\pi_2=\beta} \iint C d\pi + \varepsilon \mathrm{KL}(\pi, \alpha\otimes\beta)$

[Cuturi et al, 2013]

	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n,\beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n,\beta_n) - \mathcal{L}(\alpha,\beta)]$
KL	—	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$
OT	+++	$O(n^3)$	$O(n^{-\frac{2}{d}})$

Closing the loop

Optimal Transport	Kullback-Leibler (relative entropy)	MMD norms
$\mathrm{OT}(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha, \pi_2=\beta}} \iint C d\pi$	$\mathrm{KL}(\alpha,\beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$	$\|\alpha-\beta\|_K^2 = \iint K(x,y) d^2(\alpha-\beta)$

Entropy-regularized OT $\mathrm{OT}_\varepsilon(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha, \pi_2=\beta}} \iint C d\pi + \varepsilon \mathrm{KL}(\pi, \alpha\otimes\beta)$ Not a distance !

	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n,\beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n,\beta_n) - \mathcal{L}(\alpha,\beta)]$
KL	−	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$
OT	+++	$O(n^3)$	$O(n^{-\frac{2}{d}})$

Introduction

Optimal Transport

$$\mathrm{OT}(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha,\pi_2=\beta}} \iint C d\pi$$

Kullback-Leibler (relative entropy)

$$\mathrm{KL}(\alpha,\beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$$

MMD norms

$$\|\alpha-\beta\|_K^2 = \iint K(x,y)d^2(\alpha-\beta)$$

Entropy-regularized OT

$$\mathrm{OT}_\varepsilon(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha,\pi_2=\beta}} \iint C d\pi + \varepsilon \mathrm{KL}(\pi, \alpha\otimes\beta)$$

Not a distance !

$$S_\varepsilon(\alpha,\beta) = \mathrm{OT}_\varepsilon(\alpha,\beta) - \frac{1}{2}(\mathrm{OT}_\varepsilon(\alpha,\alpha) + \mathrm{OT}_\varepsilon(\beta,\beta))$$

	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n,\beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n,\beta_n) - \mathcal{L}(\alpha,\beta)]$
KL	−	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$
OT	+++	$O(n^3)$	$O(n^{-\frac{2}{d}})$

Introduction

Optimal Transport

$$\mathrm{OT}(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha,\pi_2=\beta}} \iint C d\pi$$

Kullback-Leibler (relative entropy)

$$\mathrm{KL}(\alpha,\beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$$

MMD norms

$$\|\alpha-\beta\|_K^2 = \iint K(x,y) d^2(\alpha-\beta)$$

Entropy-regularized OT

$$\mathrm{OT}_\varepsilon(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha,\pi_2=\beta}} \iint C d\pi + \varepsilon \mathrm{KL}(\pi, \alpha\otimes\beta)$$

$$S_\varepsilon(\alpha,\beta) = \mathrm{OT}_\varepsilon(\alpha,\beta) - \frac{1}{2}(\mathrm{OT}_\varepsilon(\alpha,\alpha) + \mathrm{OT}_\varepsilon(\beta,\beta))$$

	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n,\beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n,\beta_n) - \mathcal{L}(\alpha,\beta)]$
KL	$-$	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$
OT	+++	$O(n^3)$	$O(n^{-\frac{2}{d}})$
S	+++	$O(n^2)$	$O(n^{-1}\varepsilon^{-1/d})$

Introduction

Optimal Transport

$$\mathrm{OT}(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha,\pi_2=\beta}} \iint C d\pi$$

Kullback-Leibler (relative entropy)

$$\mathrm{KL}(\alpha,\beta) = \int \log\left(\frac{d\alpha}{d\beta}\right) d\alpha$$

MMD norms

$$\|\alpha-\beta\|_K^2 = \iint K(x,y) d^2(\alpha-\beta)$$

Entropy-regularized OT

$$\mathrm{OT}_\varepsilon(\alpha,\beta) = \min_{\substack{\pi \\ \pi_1=\alpha,\pi_2=\beta}} \iint C d\pi + \varepsilon \mathrm{KL}(\pi, \alpha\otimes\beta)$$

[Cuturi et al, 2013]

[Feydy et al, 2019]

$$\mathrm{OT}(\alpha,\beta) \longleftarrow S_\varepsilon(\alpha,\beta) = \mathrm{OT}_\varepsilon(\alpha,\beta) - \frac{1}{2}(\mathrm{OT}_\varepsilon(\alpha,\alpha) + \mathrm{OT}_\varepsilon(\beta,\beta)) \longrightarrow \|\alpha-\beta\|_{-C}^2$$

$0 \longleftarrow \varepsilon$ $\qquad$ $\varepsilon \longrightarrow +\infty$

	"Geometry"	Computational complexity # of operations to compute $\mathcal{L}(\alpha_n,\beta_n)$	Sample complexity $\mathbb{E}[\mathcal{L}(\alpha_n,\beta_n) - \mathcal{L}(\alpha,\beta)]$
KL	$-$	$O(n)$	Ill-defined (can be $+\infty$)
MMD	+	$O(n^2)$	$O(n^{-1})$
OT	+++	$O(n^3)$	$O(n^{-\frac{2}{d}})$
S	+++	$O(n^2)$	$O(n^{-1/2}\varepsilon^{-1/d})$

POT: Python Optimal Transport

Contents

POT Python Optimal Transport

0.8.2

Search docs

POT: Python Optimal Transport

Quick start guide

API and modules

Examples gallery

Releases

Contributing to POT

Contributors

Code of Conduct

» POT: Python Optimal Transport

View page source

POT: Python Optimal Transport

Contents

Number of ML papers with OT