Generalizing and Scaling Optimal Transport

Généralisation et Passage à L'Échelle du Transport Optimal

2025-05-11

Rémi Emonet

Célébration des 10 ans de la FIL

Université Jean Monnet Saint-Étienne

Laboratoire Hubert Curien UMR • CNRS • 5516 • Saint-Étienne

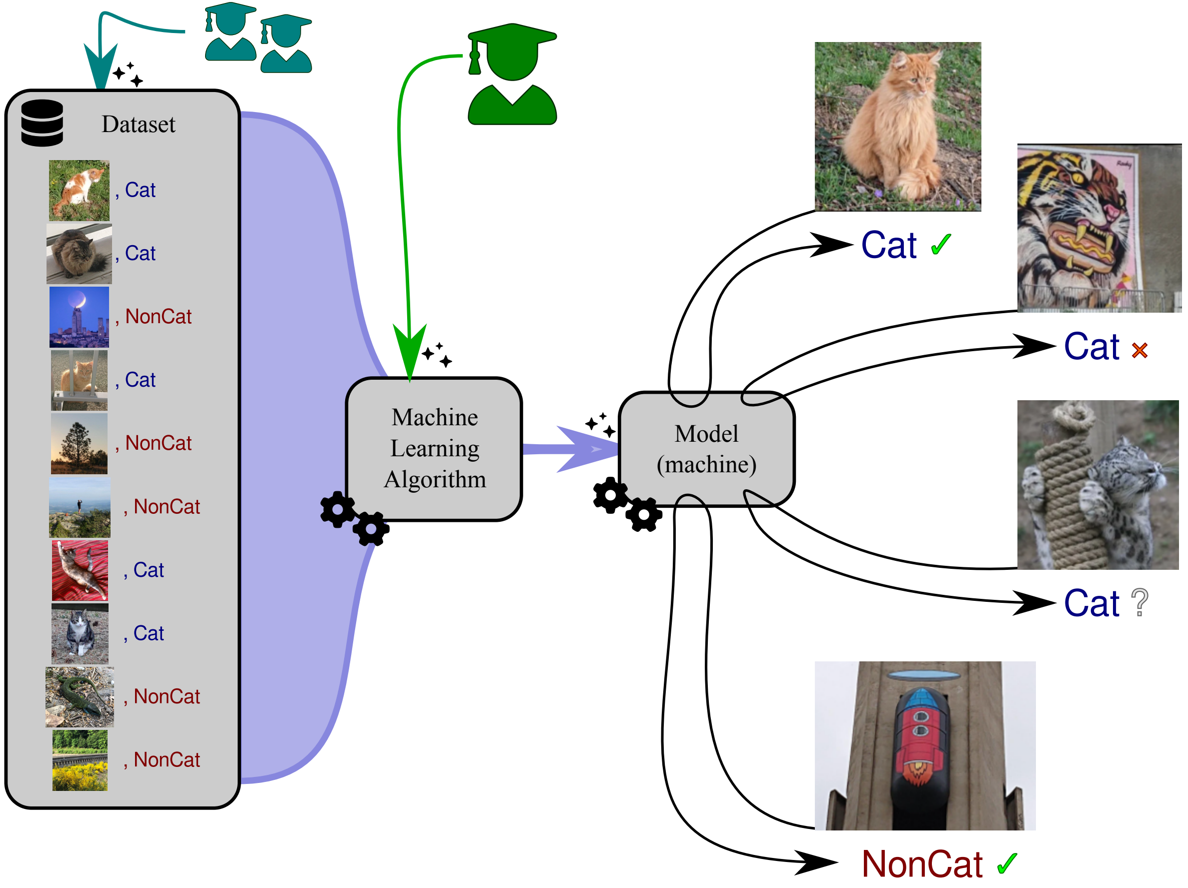
cnrs

Inria

institut universitaire de France

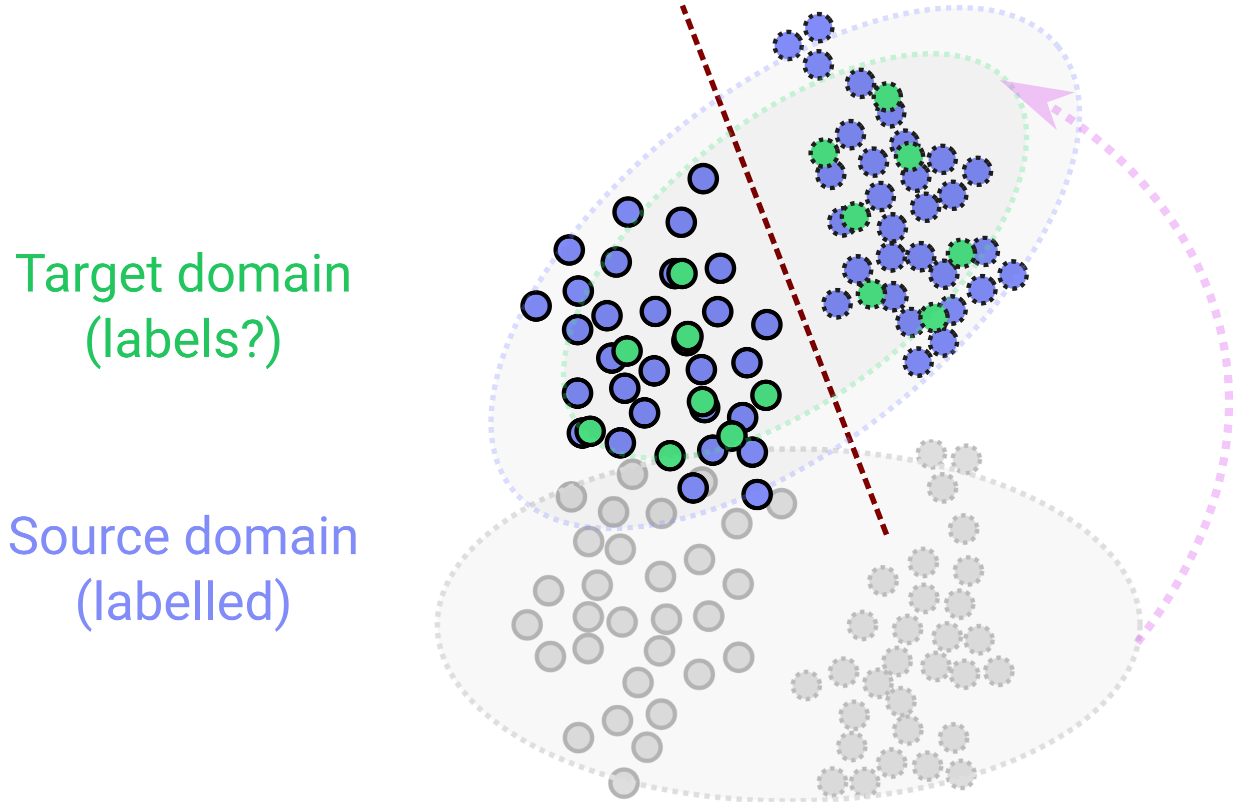
Introduction

What is Machine Learning?



- theoretical foundations
- algorithms
- certificates
- many facets
  - imbalance
  - fairness
  - privacy
  - explainability
- ...

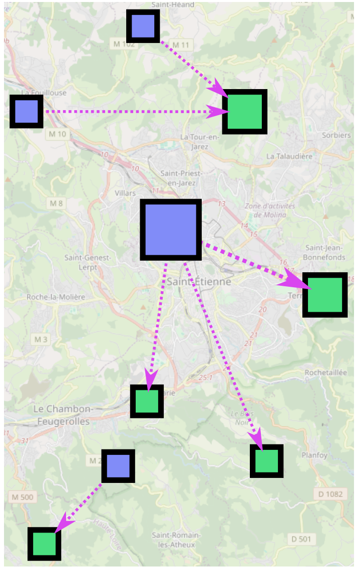
Transfer Learning and Domain Adaptation

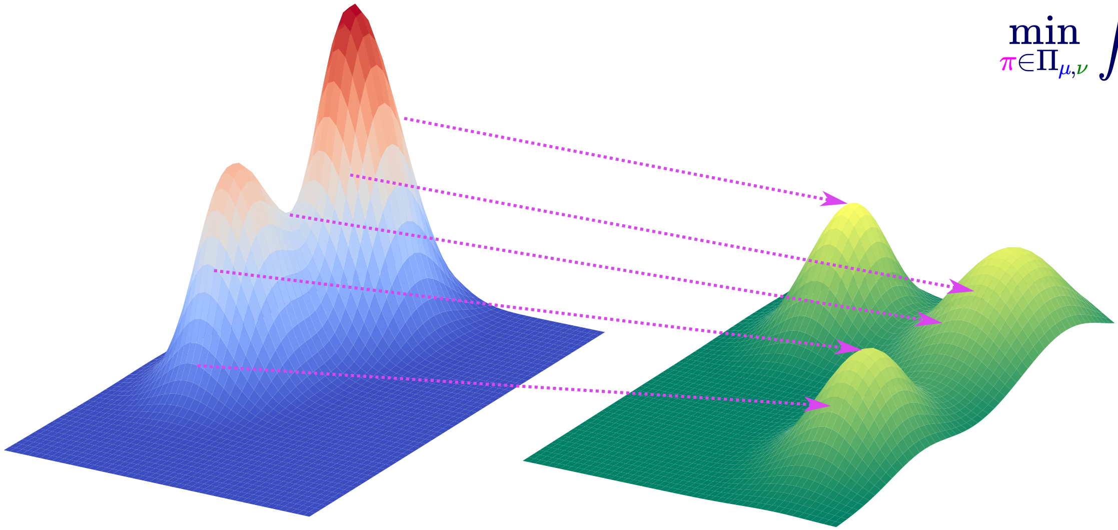


Distribution/dataset alignment

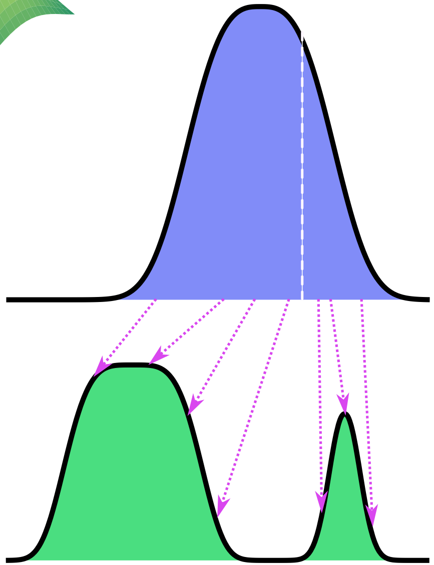
- key for domain adaptation
- need a notion of distance

Optimal Transport: distance between distributions (discrete, continuous)





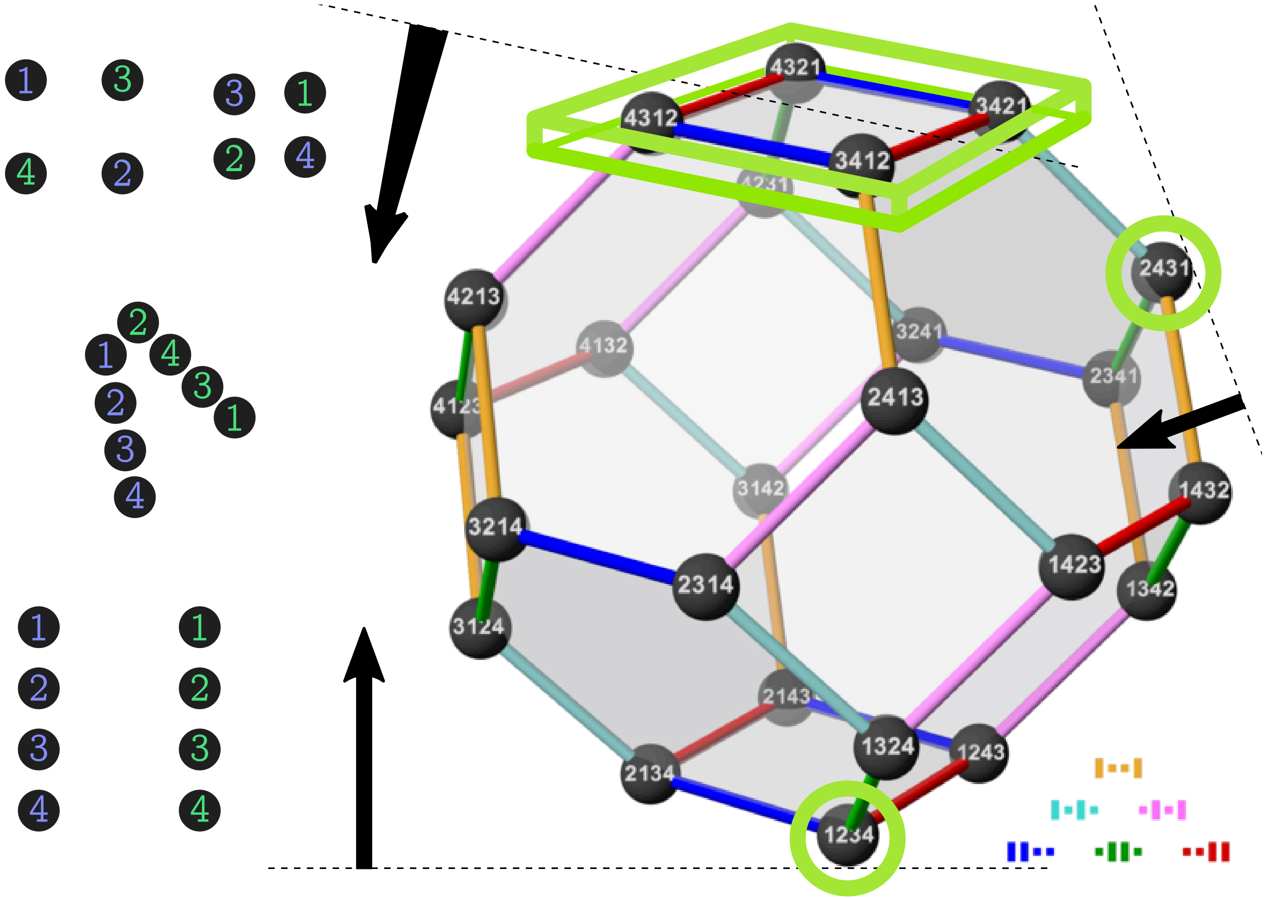
$$\min_{\pi \in \Pi_{\mu,\nu}} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y)$$



$$\min_{T \in \Pi_{a,b}} \sum_{i,j} T_{ij} C_{ij}$$

$$\min_{T \in \Pi_{a,b}} < T, C >$$

Birkhoff Polytope, Permutohedron and Computation Complexity



- OT Problem
  - Linear problem
  - Linear constraints
  - Birkhoff polytope
- Complexity
  - $O(n^3)$ (general case)
  - $O(n\log(n))$ in 1D

Overview

- Introduction
- Generalized Optimal Transport as a Modelling Tool
- Scaling Optimal Transport
- Generative Models and Optimal Transport
- Deriving Generalization Guarantees

Overview:

- Introduction
- Generalized Optimal Transport as a Modelling Tool
- Scaling Optimal Transport
- Generative Models and Optimal Transport
- Deriving Generalization Guarantees

Generalized Optimal Transport as a Modelling Tool

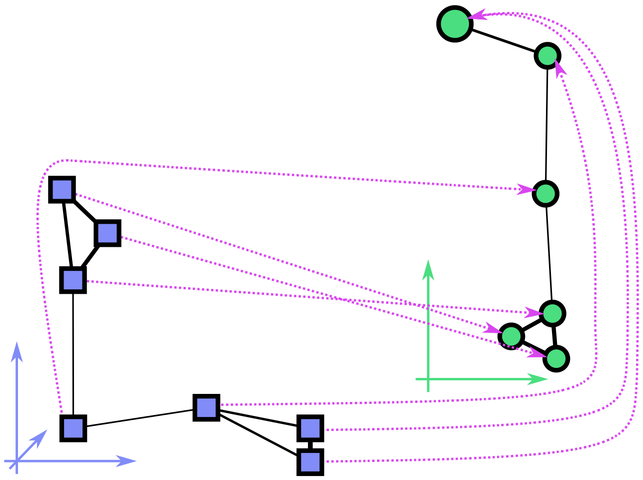
Transport Between Heterogeneous Domains: Gromov-Wasserstein

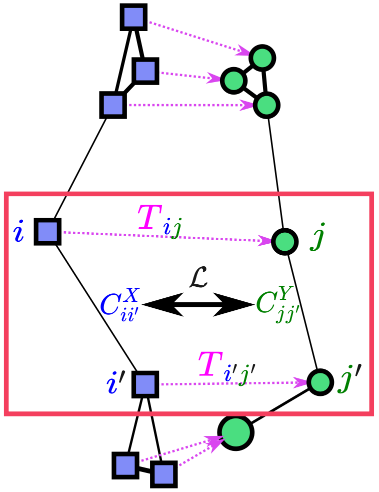
Reminder, optimal transport problem $OT(C) = \min_{T \in \Pi_{a,b}} \sum_{i,j} T_{ij} C_{ij} = \min_{T \in \Pi_{a,b}} < T, C >$

⇒ **limited to data in the same space** (to define C)

Gromov-Wasserstein problem $GW(C^X, C^Y) = \min_{T \in \Pi_{a,b}} \sum_{i,j} \sum_{i',j'} T_{ij} T_{i'j'} \mathcal{L}(C^X_{ii'}, C^Y_{jj'})$

- matches source points to target points
- minimize the difference of distances
  - between a pair of source points
  - and a pair of target points
  - that are mapped onto each other
- ⇒ graph matching





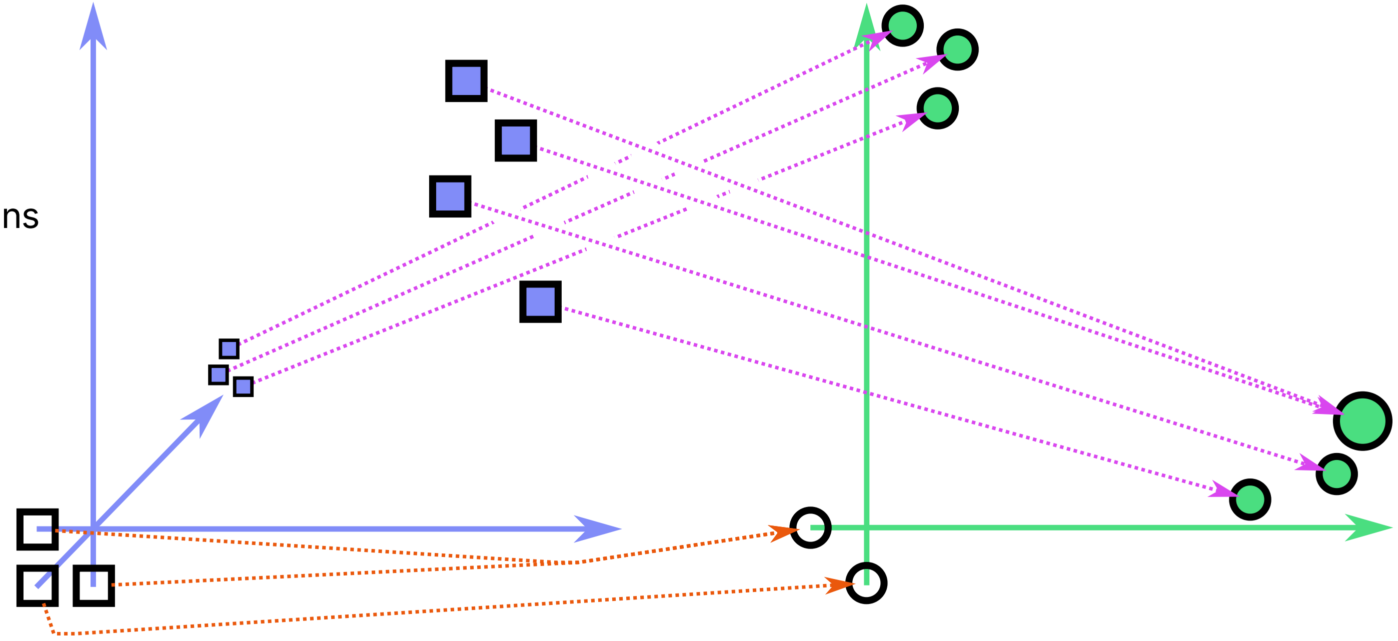
Fused Gromov Wassertein « $FGW(C, C^X, C^Y) = (1 - \beta) \cdot OT(C) + \beta \cdot GW(C^X, C^Y)$ »

Co-Optimal Transport problem

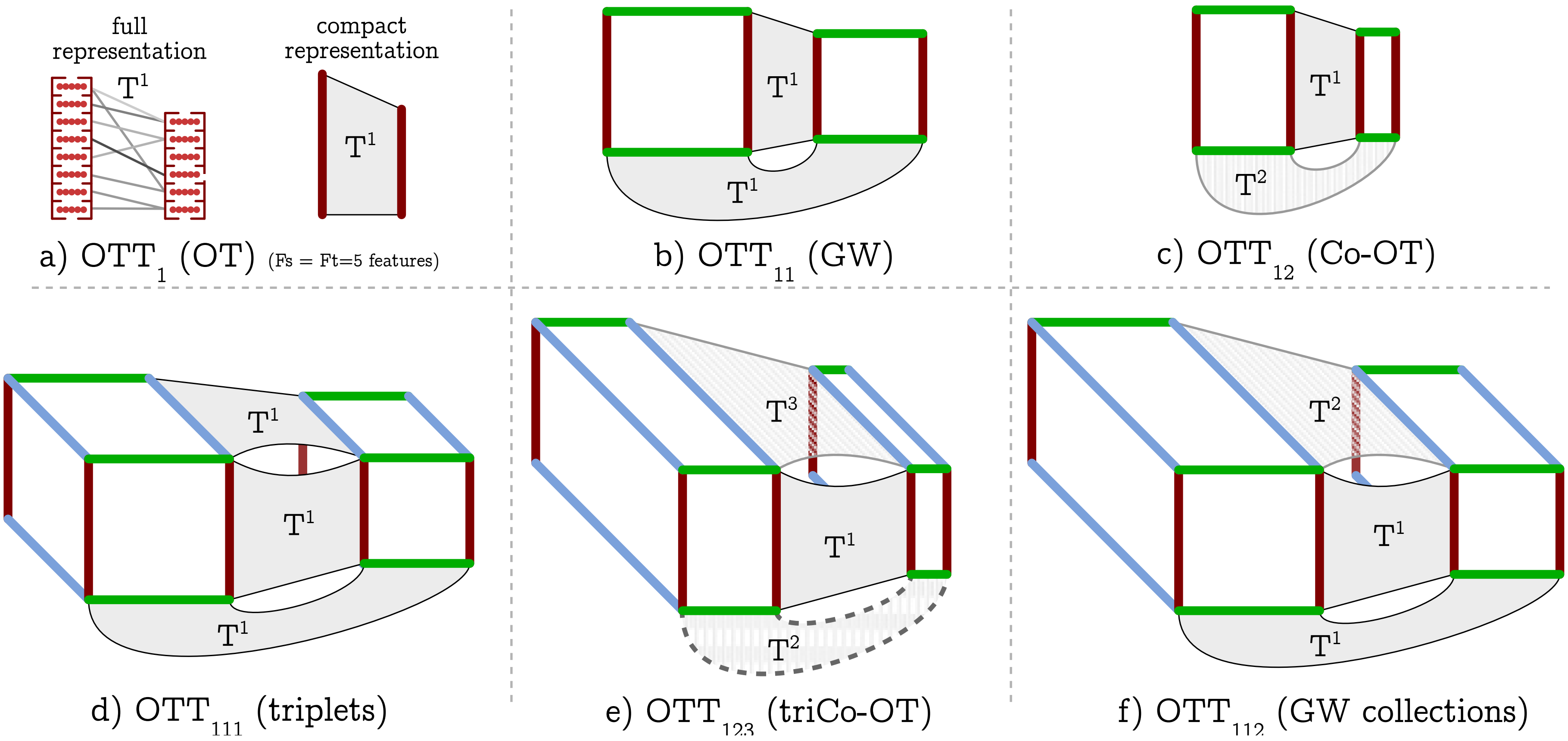
$$Co-OT(X,Y) = \min_{T^1 \in \Pi_{a^1,b^1}} \min_{T^2 \in \Pi_{a^2,b^2}} \sum_{i^1,j^1} \sum_{i^2,j^2} {T^1}_{i^1 j^1} {T^2}_{i^2 j^2} \mathcal{L}(X_{i^1 i^2}, Y_{j^1 j^2})$$

Matches

- source/target points
- sources/target dimensions
- with 2 transport plans



Generalization: OTT$_*$ (*Optimal Tensor Transport*)



More axes of generalization

- Mixed problems ("fused")
- Partial transport, relaxed constraints
- Multi-marginal
- Class information handling (for domain adaptation)
- Representation/metric learning
- Time-series constraints

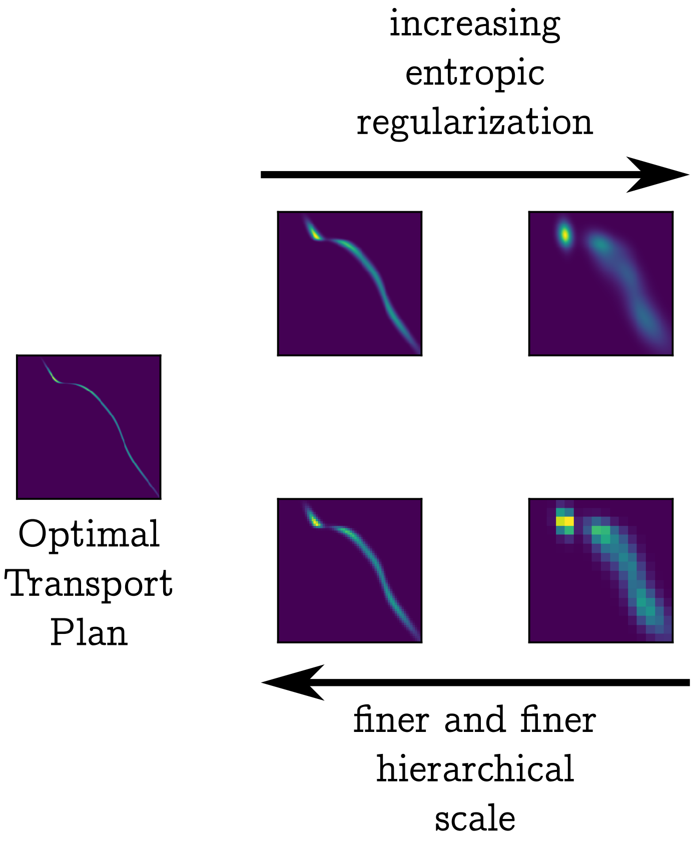
Overview:

- Introduction
- Generalized Optimal Transport as a Modelling Tool
- Scaling Optimal Transport
- Generative Models and Optimal Transport
- Deriving Generalization Guarantees

Scaling Optimal Transport

Main Means for Scaling Optimal Transport (and Generalized OT)

- Entropic regularization
- Hierarchical (for spatial or groupeable data)
- Sliced approaches / random projections
- Stochastic approximations

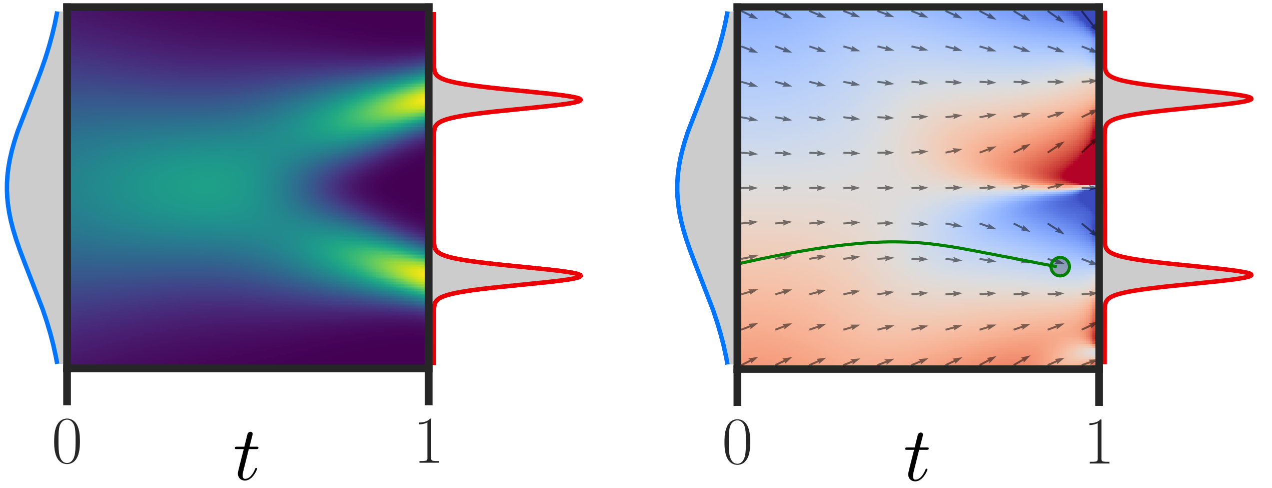


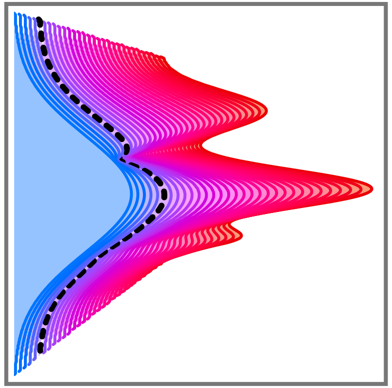
Overview:

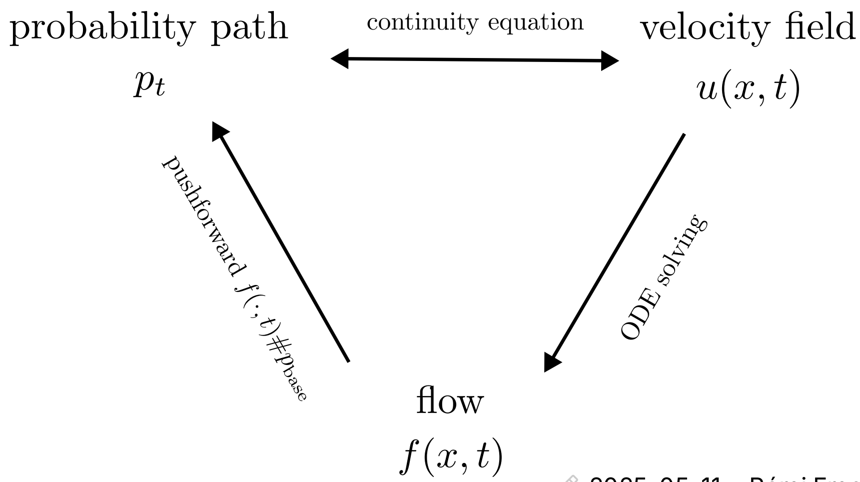
- Introduction
- Generalized Optimal Transport as a Modelling Tool
- Scaling Optimal Transport
- Generative Models and Optimal Transport
- Deriving Generalization Guarantees

Generative Models and Optimal Transport

Conditional Flow Matching / Diffusion / Optimal Transport







- OT for Conditional Flow Matching (OT-CFM)
- (Transport) Optimality of CFM couplings?
- CFM to divide-and-conquer OT

Overview:

- Introduction
- Generalized Optimal Transport as a Modelling Tool
- Scaling Optimal Transport
- Generative Models and Optimal Transport
- Deriving Generalization Guarantees

Deriving Generalization Guarantees

- Probabilistic view of (generalized) optimal transport
- **what**: e.g., sample complexity, convergence, generalization bounds
- **how**: e.g., the PAC-Bayesian framework

Overview:

- Introduction
- Generalized Optimal Transport as a Modelling Tool
- Scaling Optimal Transport
- Generative Models and Optimal Transport
- Deriving Generalization Guarantees

Conclusion

Generalizing and Scaling Optimal Transport

Généralisation et Passage à L'Échelle du Transport Optimal

2025-05-11

Rémi Emonet

Célébration des 10 ans de la FIL

Thank You! Questions?

Université Jean Monnet Saint-Étienne

Laboratoire Hubert Curien UMR • CNRS • 5516 • Saint-Étienne

cnrs

Inria

IUF institut universitaire de France

Overview:

Supplementary slides

SaGroW : Algorithme d'optimisation pour GW et OTT

Rappel, problème de Gromov-Wasserstein (ré-écrit)

$$GW(C^X, C^Y) = \min_{T \in \Pi_{a,b}} \sum_i \sum_j T_{ij} \sum_{i',j'} T_{i'j'} \mathcal{L}(C^X_{ii'}, C^Y_{jj'})$$

Algorithm SaGroW

Require: $\boldsymbol{a}$, $\boldsymbol{b}$ (probability vectors of μ and ν), $\mathcal{C}^{\mathcal{X}}$, $\mathcal{C}^{\mathcal{Y}}$ (cost matrices), $\mathcal{L}$ (loss function), M (number of samples), ϵ (entropy regularization), α (partial update weight)

$\boldsymbol{T}_0 = \boldsymbol{a}\boldsymbol{b}^\top$
for s= 0 **to** S-1 **do**
  $(j_m, l_m) \sim \text{Sample}(\boldsymbol{T}_s)\ \forall m \in [\![1, M]\!]$
  $\widehat{\boldsymbol{\Lambda}}_{ik} = \frac{1}{M}\sum_{m=1}^{M} \mathcal{L}(\boldsymbol{C}^{\mathcal{X}}_{i,j_m}, \boldsymbol{C}^{\mathcal{Y}}_{k,l_m})\ \forall i, k \in [\![1, N]\!]$
  $\boldsymbol{T}'_s$ = solve the regularized OT problem $(\boldsymbol{a}, \boldsymbol{b}, \widehat{\boldsymbol{\Lambda}}, \epsilon)$
  $\boldsymbol{T}_{s+1} = (1-\alpha)\boldsymbol{T}_s + \alpha \boldsymbol{T}'_s$
end for
return $\boldsymbol{T}_{S-1}$

- termes vues comme des espérances mathématiques
- algorithme stochastique par échantillonnage
- complexité contrôlée et possibilité de transport 1d