Unbalanced optimal transport: formulation and efficient computational solutions

Laetitia Chapel
laetitia.chapel@irisa.fr

IRISA, Rennes, France

Institut Agro Rennes-Angers

GDR IASIS, Feb. 2025

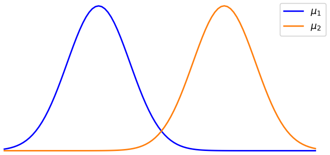
Table of Contents

Optimal transport
Balanced Optimal transport: Monge formulation

- **Balanced** optimal transport

$$\mathcal{OT}(\mu_1, \mu_2) \triangleq \inf_{\boldsymbol{t}} \int c(x, \boldsymbol{t}(x)) \, d\mu_1(x)$$

where $\boldsymbol{t}$ is a **transport map** and $\boldsymbol{t}_\# \mu_1 = \mu_2$

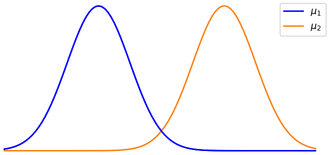


Optimal transport
Balanced Optimal transport: Monge formulation

- **Balanced** optimal transport

$$\mathcal{OT}(\mu_1, \mu_2) \triangleq \inf_{\boldsymbol{t}} \int c(x, \boldsymbol{t}(x))\, d\mu_1(x)$$

where $\boldsymbol{t}$ is a **transport map** and $\boldsymbol{t}_{\#}\mu_1 = \mu_2$

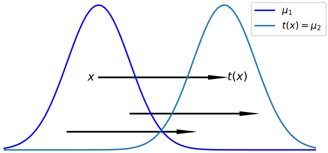


Optimal transport
Balanced Optimal transport: Monge formulation

- **Balanced** optimal transport

$$\mathcal{OT}(\mu_1, \mu_2) \triangleq \inf_{\boldsymbol{t}} \int c(x, \boldsymbol{t}(x))\, d\mu_1(x)$$

where $\boldsymbol{t}$ is a **transport map** and $\boldsymbol{t}_{\#}\mu_1 = \mu_2$



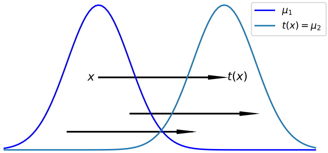
Defines for each particle located at x what is its destination $\boldsymbol{t}(x)$

Optimal transport
Balanced Optimal transport: Monge formulation

- **Balanced** optimal transport

$$\mathcal{OT}(\mu_1, \mu_2) \triangleq \inf_{\boldsymbol{t}} \int c(x, \boldsymbol{t}(x)) \, d\mu_1(x)$$

where $\boldsymbol{t}$ is a **transport map** and $\boldsymbol{t}_{\#}\mu_1 = \mu_2$



Defines for each particle located at x what is its destination $\boldsymbol{t}(x)$

- Among other conditions, implies that μ_1 and μ_2 have the same masses (no mass creation nor destruction)

Optimal transport
Balanced Optimal transport: Kantorovich formulation

- **Balanced** optimal transport

Linear loss

$$\mathcal{OT}(\mu_1, \mu_2) \triangleq \inf_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{X \times Y} c(x, y) \, d\gamma(x, y)$$

where $\Gamma(\mu_1, \mu_2) \stackrel{\text{def}}{=} \{\gamma \in \mathcal{M}_+(X \times Y) \text{ s.t. } (\pi_x)_\# \gamma = \mu_1 \text{ and } (\pi_y)_\# \gamma = \mu_2\}$ with $\pi_x : X \times Y \rightarrow X$.

Marginal constraints

Optimal transport
Balanced Optimal transport: Kantorovich formulation

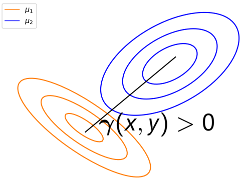
- **Balanced** optimal transport

Linear loss

$$\mathcal{OT}(\mu_1,\mu_2) \triangleq \inf_{\gamma \in \Gamma(\mu_1,\mu_2)} \int_{X\times Y} c(x,y)\, d\gamma(x,y)$$

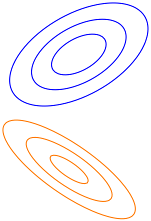
where $\Gamma(\mu_1,\mu_2) \stackrel{\text{def}}{=} \{\gamma \in \mathcal{M}_+(X\times Y) \text{ s.t. } (\pi_x)_\#\gamma = \mu_1 \text{ and } (\pi_y)_\#\gamma = \mu_2\}$ with $\pi_x : X\times Y \to X$.

Marginal constraints



with $(\pi_x)_\#\gamma = \mu_1$

and $(\pi_y)_\#\gamma = \mu_2$



The **transport plan** $\gamma(x,y)$ specifies for each pair (x,y) how many particles go from x to y

- Still implies that μ_1 and μ_2 have the same masses

Optimal transport

Balanced Optimal transport: Kantorovich formulation

- **Balanced** optimal transport

$$\mathcal{OT}(\mu_1, \mu_2) \triangleq \inf_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{X \times Y} c(x, y) \, d\gamma(x, y)$$

where $\Gamma(\mu_1, \mu_2) \stackrel{\text{def}}{=} \{\gamma \in \mathcal{M}_+(X \times Y) \text{ s.t. } (\pi_x)_\# \gamma = \mu_1 \text{ and } (\pi_y)_\# \gamma = \mu_2\}$ with $\pi_x : X \times Y \to X$.

- Can be rewritten with a penalty term

$$\mathcal{OT}(\mu_1, \mu_2) = \inf_{\gamma \geq 0} \int_{X \times Y} c(x, y) \, d\gamma(x, y) + \iota_{\{=\}} ((\pi_x)_\# \gamma | \mu_1) + \iota_{\{=\}} ((\pi_y)_\# \gamma | \mu_2)$$

with $\iota_{\{=\}}(\nu|\mu)$ is 0 if $\nu = \mu$ and ∞ otherwise.

Optimal transport

Balanced Optimal transport: Kantorovich formulation

- **Balanced** optimal transport

$$\mathcal{OT}(\mu_1, \mu_2) \triangleq \inf_{\gamma \in \Gamma(\mu_1,\mu_2)} \int_{X \times Y} c(x,y) \, d\gamma(x,y)$$

where $\Gamma(\mu_1, \mu_2) \stackrel{\text{def}}{=} \{\gamma \in \mathcal{M}_+(X \times Y) \text{ s.t. } (\pi_x)_\# \gamma = \mu_1 \text{ and } (\pi_y)_\# \gamma = \mu_2\}$ with $\pi_x : X \times Y \to X$.

- Can be rewritten with a penalty term

$$\mathcal{OT}(\mu_1, \mu_2) = \inf_{\gamma \geq 0} \int_{X \times Y} c(x,y) \, d\gamma(x,y) + l_{\{=\}}((\pi_x)_\# \gamma | \mu_1) + l_{\{=\}}((\pi_y)_\# \gamma | \mu_2)$$

with $l_{\{=\}}(\nu|\mu)$ is 0 if $\nu = \mu$ and ∞ otherwise.

- When the distributions are **discrete** $\mu_1 = \sum_{i=1}^n h_i \delta_{x_i}$ and $\mu_2 = \sum_{j=1}^m g_j \delta_{y_j}$, it is written

$$\mathcal{OT}(\mu_1, \mu_2) = \min_{\gamma \in \Gamma(\mu_1,\mu_2)} \sum_{i,j} C_{i,j} \gamma_{i,j}$$

It is the same as the problem between their associated probability weight vectors **h** and **g**, with the cost matrix **C** depending on the support of μ_1 and μ_2:

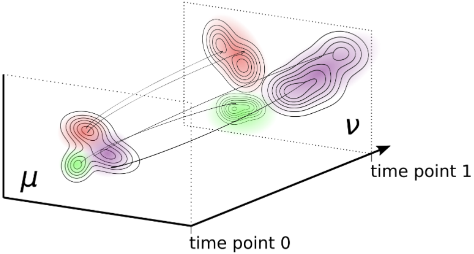
$$\mathcal{OT}_{\mathbf{C}}(\mathbf{h}, \mathbf{g}) = \mathcal{OT}(\mu_1, \mu_2)$$

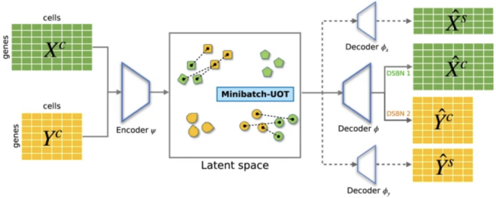
with $C_{i,j} = C(x_i, y_j)$ and $\gamma \in \mathbb{R}^{n \times m}$

Optimal transport

Balanced Optimal transport in action

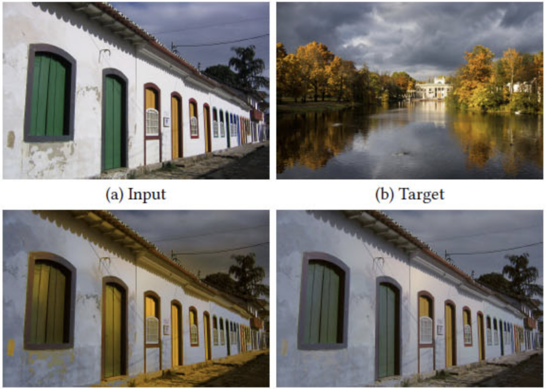
- But, in many applications, we **cannot/do not want to have the same masses**, we may want to **discard some outliers or limit the impact of the noise** or we would like to **reweight** the distributions
  - In biology, there are different cell proliferation or death in different sub-populations [14] or we may want to identify common genes [4].





Optimal transport
Balanced Optimal transport in action

- But, in many applications, we **cannot/do not want to have the same masses**, we may want to **discard some outliers or limit the impact of the noise** or we would like to **reweight** the distributions
  - In biology, there are different cell proliferation or death in different sub-populations [14] or we may want to identify common genes [4].
  - In color transfer, to account for different proportions of colors [2]



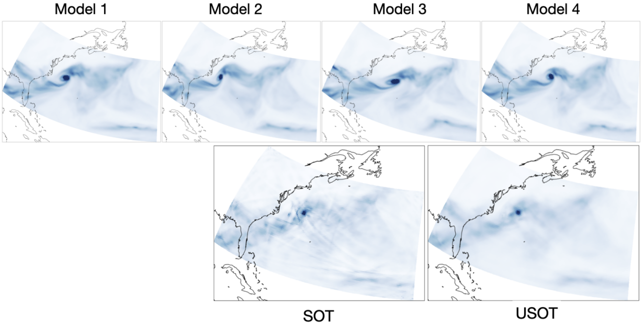
(a) Input (b) Target

(c) Full histogram matching (d) Partial histogram matching

Optimal transport

Balanced Optimal transport in action

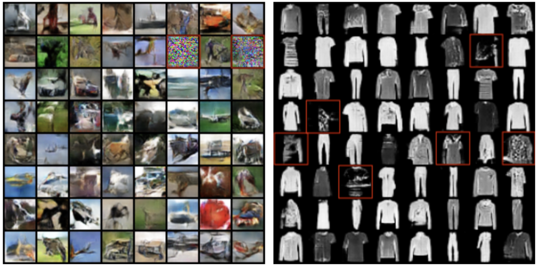
- But, in many applications, we **cannot/do not want to have the same masses**, we may want to **discard some outliers or limit the impact of the noise** or we would like to **reweight** the distributions
  - In biology, there are different cell proliferation or death in different sub-populations [14] or we may want to identify common genes [4].
  - In color transfer, to account for different proportions of colors [2]
  - In geophysics, when averaging different models [12]



Optimal transport

Balanced Optimal transport in action

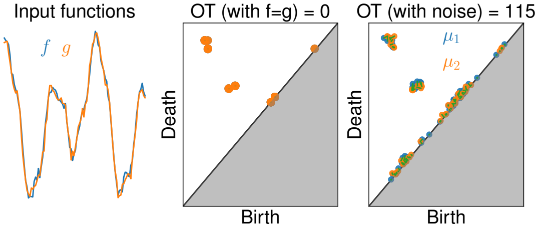
- But, in many applications, we **cannot/do not want to have the same masses**, we may want to **discard some outliers or limit the impact of the noise** or we would like to **reweight** the distributions
  - In biology, there are different cell proliferation or death in different sub-populations [14] or we may want to identify common genes [4].
  - In color transfer, to account for different proportions of colors [2]
  - In geophysics, when averaging different models [12]
  - In machine learning, when some of the points are out of the distribution, for instance with WGAN [13]



Optimal transport

Balanced Optimal transport in action

- But, in many applications, we **cannot/do not want to have the same masses**, we may want to **discard some outliers or limit the impact of the noise** or we would like to **reweight** the distributions
  - In biology, there are different cell proliferation or death in different sub-populations [14] or we may want to identify common genes [4].
  - In color transfer, to account for different proportions of colors [2]
  - In geophysics, when averaging different models [12]
  - In machine learning, when some of the points are out of the distribution, for instance with WGAN [13]
  - In topological analysis, to extract (topological) features such as gaps, connected component



Optimal transport
Balanced Optimal transport in action

- But, in many applications, we **cannot/do not want to have the same masses**, we may want to **discard some outliers or limit the impact of the noise** or we would like to **reweight** the distributions
  - In biology, there are different cell proliferation or death in different sub-populations [14] or we may want to identify common genes [4].
  - In color transfer, to account for different proportions of colors [2]
  - In geophysics, when averaging different models [12]
  - In machine learning, when some of the points are out of the distribution, for instance with WGAN [13]
  - In topological analysis, to extract (topological) features such as gaps, connected component
- How to define outlier and noise-robust OT?
  - define robust variants of OT (e.g. medians of means OT, low rank constraints on the OT plan)
  - *pick a dedicated ground cost* to avoid too much influence of samples that are too far away from the distributions
  - **allow for some mass variation**
    - *destroy mass*, in order to discard some outliers
    - *rebalance the weights*, in order to account for noise
  - **Unbalanced Optimal Transport** is often used in this context

Table of Contents

Unbalanced Optimal Transport
Definition

- **key idea**: relax the mass conservation constraint

NUMERICAL RESOLUTION OF AN "UNBALANCED" MASS TRANSPORT PROBLEM

JEAN-DAVID BENAMOU[1]

Abstract. We introduce a modification of the Monge–Kantorovitch problem of exponent 2 which accommodates non balanced initial and final densities. The augmented Lagrangian numerical method introduced in [6] is adapted to this "unbalanced" problem. We illustrate the usability of this method on an idealized error estimation problem in meteorology.

Mathematics Subject Classification. 35J60, 65K10, 78A05, 90B99.

Received: April 1st, 2003.

reg. parameter

2.4. The mixed distance

In this paper we propose to work on unbalanced data by considering the mixed Wasserstein/L^2-distance in the following sense: given two possibly unbalanced densities ρ_0 and ρ_1, find $\tilde{\rho}_1$ – the closest density to ρ_1 in the L^2-sense – which minimizes the Wasserstein distance $d_{\text{wass}}(\rho_0, \tilde{\rho}_1)$. It can be formulated as

$$\inf_{\tilde{\rho}_1} \left\{ d_{\text{wass}}(\rho_0, \tilde{\rho}_1)^2 + \frac{\gamma}{2} d_{L^2}(\tilde{\rho}_1, \rho_1)^2 \right\} \quad (16)$$

surrogate target distrib.

$\int \rho_0(x)dx = \int \tilde{\rho}_1(y)dy$

$\tilde{\rho}_1$ should be close to ρ_1

Unbalanced Optimal Transport
Definition

- Regularizing the **balanced** optimal transport, by replacing the hard constraints with some divergence D

Linear loss

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \inf_{\gamma \geq 0} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \, d\gamma(x, y)$$

reg

$$+ \lambda \left(D((\pi^1)_\# \gamma | \mu_1) + D((\pi^2)_\# \gamma | \mu_2) \right)$$

Marginal constraints

with $\lambda \geq 0$: relaxing the constraints.
When $\lambda \to \infty$ we recover the balanced OT problem.

Unbalanced Optimal Transport
Definition

- Regularizing the **balanced** optimal transport, by replacing the hard constraints with some divergence D

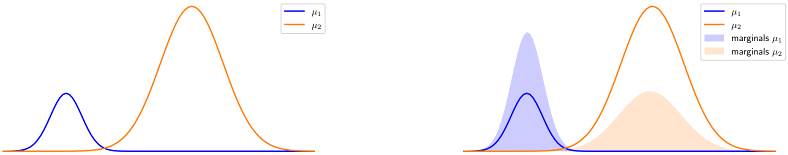
Linear loss

$$\mathcal{UOT}(\mu_1,\mu_2) \triangleq \inf_{\gamma \geq 0} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\, d\gamma(x,y) + \lambda \left(D((\pi^1)_\# \gamma | \mu_1) + D((\pi^2)_\# \gamma | \mu_2) \right)$$

reg

Marginal constraints

with $\lambda \geq 0$: relaxing the constraints.
When $\lambda \to \infty$ we recover the balanced OT problem.
When the overall masses are different



Unbalanced Optimal Transport
Definition

- Regularizing the **balanced** optimal transport, by replacing the hard constraints with some divergence D

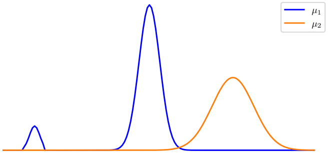
Linear loss

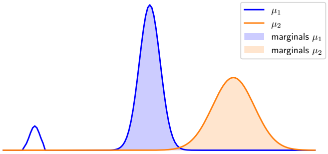
$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \inf_{\gamma \geq 0} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\, d\gamma(x,y) + \lambda \left(D((\pi^1)_{\#}\gamma | \mu_1) + D((\pi^2)_{\#}\gamma | \mu_2) \right)$$

reg

Marginal constraints

with $\lambda \geq 0$: relaxing the constraints.
When $\lambda \to \infty$ we recover the balanced OT problem.
When there are some outliers





Unbalanced Optimal Transport
Definition

- Regularizing the **balanced** optimal transport, by replacing the hard constraints with some divergence D

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \inf_{\gamma \geq 0} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\, d\gamma(x,y) + \lambda \left(D((\pi^1)_\# \gamma | \mu_1) + D((\pi^2)_\# \gamma | \mu_2) \right)$$

Linear loss: $c(x,y)$; reg: λ; Marginal constraints: $D((\pi^1)_\# \gamma | \mu_1) + D((\pi^2)_\# \gamma | \mu_2)$

with $\lambda \geq 0$: relaxing the constraints.
When $\lambda \to \infty$ we recover the balanced OT problem.

- Depending on D, has often similar properties as OT (is a distance, weak convergence etc.)

Unbalanced Optimal Transport
Definition

- Regularizing the **balanced** optimal transport, by replacing the hard constraints with some divergence D

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \inf_{\gamma \geq 0} \int_{\mathbb{R}^d \times \mathbb{R}^d} \underbrace{c(x,y)}_{\text{Linear loss}} d\gamma(x,y) + \underbrace{\lambda}_{\text{reg}} \Big(\underbrace{D((\pi^1)_\# \gamma | \mu_1) + D((\pi^2)_\# \gamma | \mu_2)}_{\text{Marginal constraints}} \Big)$$

  with $\lambda \geq 0$: relaxing the constraints.
  When $\lambda \to \infty$ we recover the balanced OT problem.
- Depending on D, has often similar properties as OT (is a distance, weak convergence etc.)
- Questions:
  - How to write the problem for discrete distributions?
  - Which D?
  - how to solve the problem?

Unbalanced Optimal Transport
Discrete UOT

- We denote $\hat{\mu}_1 = (\pi^1)_\#\gamma$ and $\hat{\mu}_2 = (\pi^2)_\#\gamma$ the marginals of γ
- When the distributions are discrete $\mu_1 = \sum_{i=1}^n h_i\delta_{x_i}$ and $\mu_2 = \sum_{j=1}^m g_j\delta_{y_j}$, it is written

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(D((\pi^1)_\#\gamma|\mu_1) + D((\pi^2)_\#\gamma|\mu_2) \right)$$

or [8]

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\hat{\mu}_1, \hat{\mu}_2 \geq 0} \mathcal{OT}(\hat{\mu}_1, \hat{\mu}_2) + \lambda \left(D(\hat{\mu}_1|\mu_1) + D(\hat{\mu}_2|\mu_2) \right)$$

$\Rightarrow$ **OT between surrogate distributions** $\hat{\mu}_1$ and $\hat{\mu}_2$ + deviation penalty

Unbalanced Optimal Transport
Discrete UOT

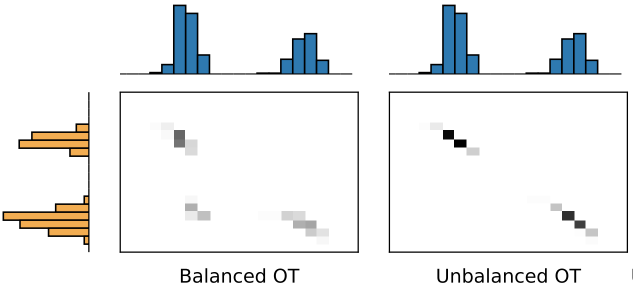
- We denote $\hat{\mu}_1 = (\pi^1)_\# \gamma$ and $\hat{\mu}_2 = (\pi^2)_\# \gamma$ the marginals of γ
- When the distributions are discrete $\mu_1 = \sum_{i=1}^n h_i \delta_{x_i}$ and $\mu_2 = \sum_{j=1}^m g_j \delta_{y_j}$, it is written

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j} \gamma_{i,j} + \lambda \left(D((\pi^1)_\# \gamma | \mu_1) + D((\pi^2)_\# \gamma | \mu_2) \right)$$

or [8]

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\hat{\mu}_1, \hat{\mu}_2 \geq 0} \mathcal{OT}(\hat{\mu}_1, \hat{\mu}_2) + \lambda \left(D(\hat{\mu}_1 | \mu_1) + D(\hat{\mu}_2 | \mu_2) \right)$$

$\Rightarrow$ **OT between surrogate distributions** $\hat{\mu}_1$ and $\hat{\mu}_2$ + deviation penalty



Unbalanced Optimal Transport
Discrete UOT

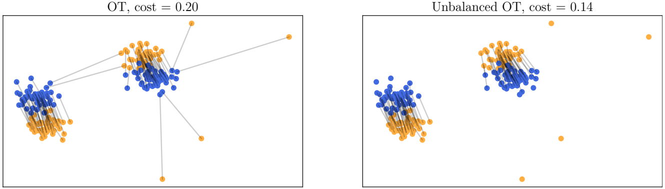
- We denote $\hat{\mu}_1 = (\pi^1)_\#\gamma$ and $\hat{\mu}_2 = (\pi^2)_\#\gamma$ the marginals of γ
- When the distributions are discrete $\mu_1 = \sum_{i=1}^n h_i \delta_{x_i}$ and $\mu_2 = \sum_{j=1}^m g_j \delta_{y_j}$, it is written

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(D((\pi^1)_\#\gamma|\mu_1) + D((\pi^2)_\#\gamma|\mu_2) \right)$$

or [8]

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\hat{\mu}_1, \hat{\mu}_2 \geq 0} \mathcal{OT}(\hat{\mu}_1, \hat{\mu}_2) + \lambda \left(D(\hat{\mu}_1|\mu_1) + D(\hat{\mu}_2|\mu_2) \right)$$

$\Rightarrow$ **OT between surrogate distributions** $\hat{\mu}_1$ and $\hat{\mu}_2$ + deviation penalty



Unbalanced Optimal Transport
Discrete UOT

- We denote $\hat{\mu}_1 = (\pi^1)_\# \gamma$ and $\hat{\mu}_2 = (\pi^2)_\# \gamma$ the marginals of γ
- When the distributions are discrete $\mu_1 = \sum_{i=1}^n h_i \delta_{x_i}$ and $\mu_2 = \sum_{j=1}^m g_j \delta_{y_j}$, it is written

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(D((\pi^1)_\#\gamma | \mu_1) + D((\pi^2)_\#\gamma | \mu_2) \right)$$

or [8]

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\hat{\mu}_1, \hat{\mu}_2 \geq 0} \mathcal{OT}(\hat{\mu}_1, \hat{\mu}_2) + \lambda \left(D(\hat{\mu}_1 | \mu_1) + D(\hat{\mu}_2 | \mu_2) \right)$$

$\Rightarrow$ **OT between surrogate distributions** $\hat{\mu}_1$ and $\hat{\mu}_2$ + deviation penalty

- It is very often restated as

$$\mathcal{UOT}_{\mathbf{C}}(\mathbf{h}, \mathbf{g}) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(D(\gamma \mathbb{1}_m | \mathbf{h}) + D(\gamma^\top \mathbb{1}_n | \mathbf{g}) \right)$$

in which the divergence does not depend on the support of μ_1 and μ_2 $\Rightarrow$ **allows creating/destroying mass**

Unbalanced Optimal Transport
Discrete UOT

- We denote $\hat{\mu}_1 = (\pi^1)_\# \gamma$ and $\hat{\mu}_2 = (\pi^2)_\# \gamma$ the marginals of γ
- When the distributions are discrete $\mu_1 = \sum_{i=1}^n h_i \delta_{x_i}$ and $\mu_2 = \sum_{j=1}^m g_j \delta_{y_j}$, it is written

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(D((\pi^1)_\#\gamma|\mu_1) + D((\pi^2)_\#\gamma|\mu_2) \right)$$

or [8]

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\hat{\mu}_1, \hat{\mu}_2 \geq 0} \mathcal{OT}(\hat{\mu}_1, \hat{\mu}_2) + \lambda \left(D(\hat{\mu}_1|\mu_1) + D(\hat{\mu}_2|\mu_2) \right)$$

$\Rightarrow$ **OT between surrogate distributions** $\hat{\mu}_1$ and $\hat{\mu}_2$ + deviation penalty ②

- It is very often restated as

$$\mathcal{UOT}_{\mathbf{C}}(\mathbf{h}, \mathbf{g}) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(D(\gamma \mathbb{1}_m|\mathbf{h}) + D(\gamma^\top \mathbb{1}_n|\mathbf{g}) \right)$$

in which the divergence does not depend on the support of μ_1 and μ_2 $\Rightarrow$ **allows creating/destroying mass** ①

Unbalanced Optimal Transport
Motivation and questions

- Selecting the right notion of discrepancy is the key
- Diverse spectrum of formulations (e.g. Sliced Unbalanced OT [1, 6])
- Does not need to assume common ground cost (Unbalanced Gromov-Wasserstein [5, 12])
- Additional regularization (entropic) can also be considered
- Discuss several discrepancies for the regularization
  - give their main features
  - give associated computational methods

Table of Contents

① UOT for creating and destroying mass
Partial Optimal Transport

$$\mathcal{UOT}_{\mathbf{C}}(\mathbf{h},\mathbf{g}) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(\|\gamma \mathbb{1}_m - \mathbf{h}\|_1 + \|\gamma^\top \mathbb{1}_n - \mathbf{g}\|_1 \right)$$

D is L_1 penalty

① UOT for creating and destroying mass
Partial Optimal Transport

$$\mathcal{UOT}_{\mathbf{c}}(\mathbf{h},\mathbf{g}) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(\|\gamma \mathbb{1}_m - \mathbf{h}\|_1 + \|\gamma^\top \mathbb{1}_n - \mathbf{g}\|_1 \right)$$

D is L_1 penalty

is equivalent to writing

$$\mathcal{UOT}_{\mathbf{c}}(\mathbf{h},\mathbf{g}) = \inf_{\gamma \in \Gamma_{\leq}(\mathbf{h},\mathbf{g})} \sum_{i,j} C_{i,j}\gamma_{i,j}$$

where $\Gamma_{\leq(\mathbf{h},\mathbf{g})} = \{\gamma \geq 0,\ \gamma \mathbb{1}_m \leq \mathbf{h} \text{ and } \gamma^\top \mathbb{1}_n \leq \mathbf{g} \text{ and } \mathbb{1}_n^\top \gamma \mathbb{1}_m = s\}$

amount of mass to be transported

① UOT for creating and destroying mass

Partial Optimal Transport

$$\mathcal{UOT}_{\mathbf{c}}(\mathbf{h},\mathbf{g}) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(\|\gamma \mathbb{1}_m - \mathbf{h}\|_1 + \|\gamma^\top \mathbb{1}_n - \mathbf{g}\|_1 \right)$$

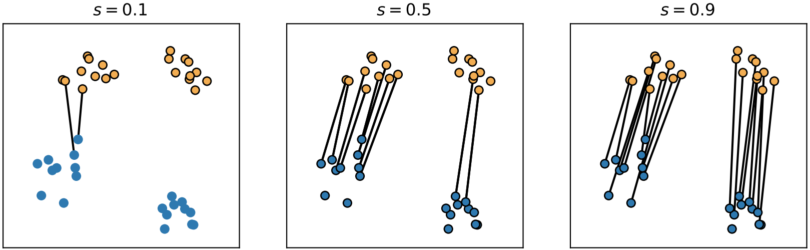
D is L_1 penalty

is equivalent to writing

$$\mathcal{UOT}_{\mathbf{c}}(\mathbf{h},\mathbf{g}) = \inf_{\gamma \in \Gamma_{\leq}(\mathbf{h},\mathbf{g})} \sum_{i,j} C_{i,j}\gamma_{i,j}$$

where $\Gamma_{\leq(\mathbf{h},\mathbf{g})} = \{\gamma \geq 0,\ \gamma\mathbb{1}_m \leq \mathbf{h} \text{ and } \gamma^\top\mathbb{1}_n \leq \mathbf{g} \text{ and } \mathbb{1}_n^\top\gamma\mathbb{1}_m = s\}$

amount of mass to be transported

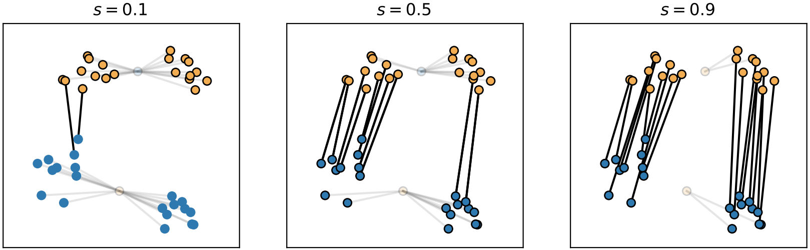


① UOT for creating and destroying mass
Partial Optimal Transport

$$\mathcal{UOT}_{\mathbf{c}}(\mathbf{h}, \mathbf{g}) \triangleq \inf_{\gamma \in \Gamma_{\leq}(\mathbf{h},\mathbf{g})} \sum_{i,j} C_{i,j}\gamma_{i,j}$$

where $\Gamma_{\leq(\mathbf{h},\mathbf{g})} = \{\gamma \geq 0, \ \gamma \mathbb{1}_m \leq \mathbf{h} \text{ and } \gamma^\top \mathbb{1}_n \leq \mathbf{g} \ \text{ and } \ \mathbb{1}_n^\top \gamma \mathbb{1}_m = s \ \}$

- Can be solved easily by adding *dummy* points $h_{n+1} = \|\boldsymbol{g}\|_1 - s$ and $g_{m+1} = \|\boldsymbol{h}\|_1 - s$ with null cost and solve the extended OT problem [5, 3]



- Any OT solver can be used!

① UOT for creating and destroying mass
Unbalanced Optimal Transport with L_2 penalty

$$\mathcal{UOT}_{\mathbf{c}}(\mathbf{h},\mathbf{g}) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(\|\boldsymbol{\gamma}\mathbb{1}_m - \mathbf{h}\|_2^2 + \|\boldsymbol{\gamma}^\top \mathbb{1}_n - \mathbf{g}\|_2^2 \right)$$

D is squared L_2 penalty

is equivalent to writing, in a vectorial form:

$$\mathcal{UOT}_{\mathbf{c}}(\mathbf{h},\mathbf{g}) \triangleq \min_{\gamma \geq 0} \|\boldsymbol{H}\gamma_v - \boldsymbol{y}\|_2^2 + \frac{1}{\lambda} \boldsymbol{c}^\top \|\gamma_v\|_1$$

linear regression pb

weighted L1 (Lasso) regul.

where $\boldsymbol{c} = \text{vec}(\boldsymbol{C})$, $\gamma_v = \text{vec}(\gamma)$, $\boldsymbol{y}^\top = [\boldsymbol{h}^\top, \boldsymbol{g}^\top]$ and $\boldsymbol{H}$ is a design matrix.

① UOT for creating and destroying mass
Unbalanced Optimal Transport with L_2 penalty

$$\mathcal{UOT}_{\mathbf{C}}(\mathbf{h},\mathbf{g}) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(\|\gamma \mathbb{1}_m - \mathbf{h}\|_2^2 + \|\gamma^\top \mathbb{1}_n - \mathbf{g}\|_2^2 \right)$$

D is squared L_2 penalty

is equivalent to writing, in a vectorial form:

$$\mathcal{UOT}_{\mathbf{C}}(\mathbf{h},\mathbf{g}) \triangleq \min_{\gamma \geq 0} \|\boldsymbol{H}\gamma_v - \boldsymbol{y}\|_2^2 + \frac{1}{\lambda} \boldsymbol{c}^\top \|\gamma_v\|_1$$

linear regression pb

weighted L1 (Lasso) regul.

where $\boldsymbol{c} = \text{vec}(\boldsymbol{C})$, $\gamma_v = \text{vec}(\gamma)$, $\boldsymbol{y}^\top = [\boldsymbol{h}^\top, \boldsymbol{g}^\top]$ and $\boldsymbol{H}$ is a design matrix.
We can borrow the tools from a large literature on solving those problems!

① UOT for creating and destroying mass

Unbalanced Optimal Transport with L_2 penalty

Regularization path of UOT: a LARS-like algorithm [7]

- Solutions are piecewise linear with $\frac{1}{\lambda}$
- We can find the set of all solutions for all λ values
  1. start with $\lambda = 0$
  2. loop
  3. increase λ until there is a change on the support of γ_V
  4. update γ_V (incremental resolution of linear equations)
  5. repeat until $\lambda = \infty$

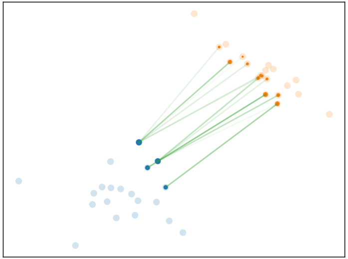


Table of Contents

② UOT with surrogate distributions
Taking into account the support with surrogate distributions, formulation

- For now, we have considered the following formulation

$$\mathcal{UOT}_{\mathbf{c}}(\mathbf{h}, \mathbf{g}) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(D(\gamma \mathbb{1}_m | \mathbf{h}) + D(\gamma^\top \mathbb{1}_n | \mathbf{g}) \right)$$

in which the divergence does not depend on the support of μ_1 and μ_2 $\Rightarrow$ **allow some mass variation**

- What if we also take into account the support of the samples?

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\hat{\mu}_1, \hat{\mu}_2 \geq 0} \mathcal{OT}(\hat{\mu}_1, \hat{\mu}_2) + \lambda \left(D(\hat{\mu}_1 | \mu_1) + D(\hat{\mu}_2 | \mu_2) \right)$$

② UOT with surrogate distributions

Taking into account the support with surrogate distributions, formulation

- For now, we have considered the following formulation

$$\mathcal{UOT}_{\mathbf{C}}(\mathbf{h},\mathbf{g}) \triangleq \min_{\gamma \geq 0} \sum_{i,j} C_{i,j}\gamma_{i,j} + \lambda \left(D(\gamma \mathbb{1}_m|\mathbf{h}) + D(\gamma^\top \mathbb{1}_n|\mathbf{g}) \right)$$

in which the divergence does not depend on the support of μ_1 and μ_2 $\Rightarrow$ **allow some mass variation**

- What if we also take into account the support of the samples?

$$\mathcal{UOT}(\mu_1,\mu_2) \triangleq \min_{\hat{\mu}_1,\hat{\mu}_2 \geq 0} \mathcal{OT}(\hat{\mu}_1,\hat{\mu}_2) + \lambda \left(D(\hat{\mu}_1|\mu_1) + D(\hat{\mu}_2|\mu_2) \right)$$

- Which D?
  - MMD [11]
  - OT [9]
- The price is that it involves optimization over all possible joint measures

② UOT with surrogate distributions

UOT with OT penalty

- **Unbalanced OT with an OT penalty: rebalancing the weigths RebOT [10]**

$$\mathcal{UOT}(\mu_1,\mu_2) \triangleq \min_{\hat{\mu}_1,\hat{\mu}_2 \geq 0} \mathcal{OT}(\hat{\mu}_1,\hat{\mu}_2) + \lambda \left(\mathcal{OT}(\hat{\mu}_1,\mu_1) + \mathcal{OT}(\hat{\mu}_2,\mu_2) \right)$$

$\Rightarrow$ **do not allow some mass variation, rather *rebalance* the mass** as the mass of $\hat{\mu}_i$ should be equal to μ_i
$\hat{\mu}_1$ and $\hat{\mu}_2$ provide *compressed* representation of μ_1 and μ_2

② UOT with surrogate distributions

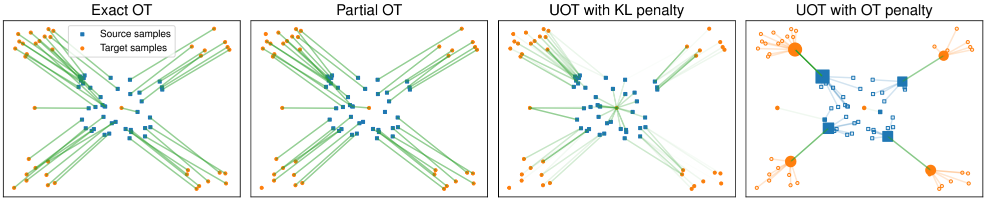
UOT with OT penalty

- **Unbalanced OT with an OT penalty: rebalancing the weigths RebOT [10]**

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\hat{\mu}_1, \hat{\mu}_2 \geq 0} \mathcal{OT}(\hat{\mu}_1, \hat{\mu}_2) + \lambda \left(\mathcal{OT}(\hat{\mu}_1, \mu_1) + \mathcal{OT}(\hat{\mu}_2, \mu_2) \right)$$

$\Rightarrow$ **do not allow some mass variation, rather *rebalance* the mass** as the mass of $\hat{\mu}_i$ should be equal to μ_i
$\hat{\mu}_1$ and $\hat{\mu}_2$ provide *compressed* representation of μ_1 and μ_2

- Can be solved with any convex solver (e.g. CVXPY), is a distance



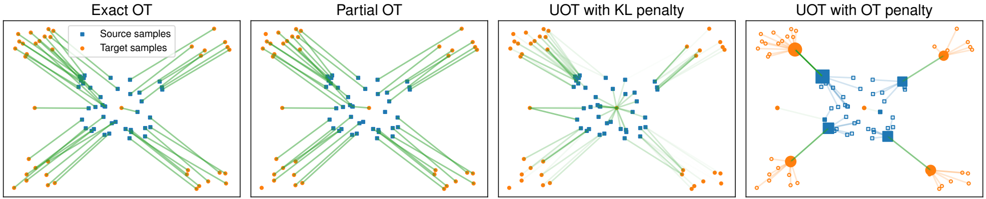
② UOT with surrogate distributions
UOT with OT penalty

- **Unbalanced OT with an OT penalty: rebalancing the weigths RebOT [10]**

$$\mathcal{UOT}(\mu_1, \mu_2) \triangleq \min_{\hat{\mu}_1, \hat{\mu}_2 \geq 0} \mathcal{OT}(\hat{\mu}_1, \hat{\mu}_2) + \lambda \left(\mathcal{OT}(\hat{\mu}_1, \mu_1) + \mathcal{OT}(\hat{\mu}_2, \mu_2) \right)$$

$\Rightarrow$ **do not allow some mass variation, rather *rebalance* the mass** as the mass of $\hat{\mu}_i$ should be equal to μ_i
$\hat{\mu}_1$ and $\hat{\mu}_2$ provide *compressed* representation of μ_1 and μ_2

- Can be solved with any convex solver (e.g. CVXPY), is a distance

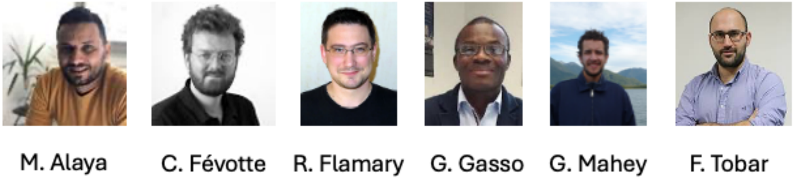


- Outliers: points with small mass on the rebalanced distribution $\hat{\mu}_1$ and $\hat{\mu}_2$

Table of Contents

Unbalanced Optimal Transport
Conclusion and pen challenges

- Conclusion
  - UOT is mandatory for many applications
  - (many) efficient solvers exist
  - implementation in POT python toolbox [1]
- Some open challenges
  - outlier removal?
  - which statistical guarantees?



[1] many figures have been generated with POT https://pythonot.github.io/

Unbalanced optimal transport: formulation and efficient computational solutions

Laetitia Chapel
laetitia.chapel@irisa.fr

IRISA, Rennes, France

Institut Agro Rennes-Angers

GDR IASIS, Feb. 2025

Table of Contents

Bibliography I

[1] Clément Bonet et al. "Slicing Unbalanced Optimal Transport". In: *Transactions on Machine Learning Research* (2024).

[2] Nicolas Bonneel and David Coeurjolly. "Spot: sliced partial optimal transport". In: *ACM Transactions on Graphics (TOG)* (2019).

[3] Luis A Caffarelli and Robert J McCann. "Free boundaries in optimal transport and Monge-Ampere obstacle problems". In: *Annals of mathematics* (2010).

[4] Kai Cao et al. "A unified computational framework for single-cell data integration with optimal transport". In: *Nature Communications* (2022).

[5] Laetitia Chapel, Mokhtar Z Alaya, and Gilles Gasso. "Partial optimal tranport with applications on positive-unlabeled learning". In: *NeurIPS* (2020).

[6] Laetitia Chapel and Romain Tavenard. "One for all and all for one: Efficient Computation of Partial Wasserstein Distances on the Line". In: *International Conference on Learning Representations*. 2025.

[7] Laetitia Chapel et al. "Unbalanced optimal transport through non-negative penalized linear regression". In: *NeurIPS* (2021).

[8] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. "Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures". In: *Inventiones mathematicae* 211.3 (2018), pp. 969–1117.

Bibliography II

[9] Chi-Heng Lin, Mehdi Azabou, and Eva Dyer. “Making transport more robust and interpretable by moving data through a small number of anchor points”. In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, 18–24 Jul 2021, pp. 6631–6641.

[10] Guillaume Mahey et al. “Rebalanced optimal transportation: A Wasserstein penalty for unbalanced OT”. In: *preprint* (2024).

[11] Piyushi Manupriya, J Saketha Nath, and Pratik Jawanpuria. “MMD-Regularized Unbalanced Optimal Transport”. In: *arXiv preprint arXiv:2011.05001* (2020).

[12] Thibault Séjourné et al. “Unbalanced Optimal Transport meets Sliced-Wasserstein”. In: *arXiv preprint arXiv:2306.07176* (2023).

[13] G. Staerman et al. “When OT meets MoM: Robust estimation of Wasserstein Distance”. In: *AISTATS*. 2021.

[14] Karren D Yang and Caroline Uhler. “Scalable Unbalanced Optimal Transport using Generative Adversarial Networks”. In: *International Conference on Learning Representations*. 2018.