Unbalanced Optimal Transport and OT between metric spaces

Rémi Flamary, **Laetitia Chapel**, Nicolas Courty

June 22, 2022

Ecole de recherche (part. 2) Rennes – Juin 2022
Divergences statistiques et géométriques pour l'apprentissage machine

Extensions of Optimal Transport

Relaxation and extensions

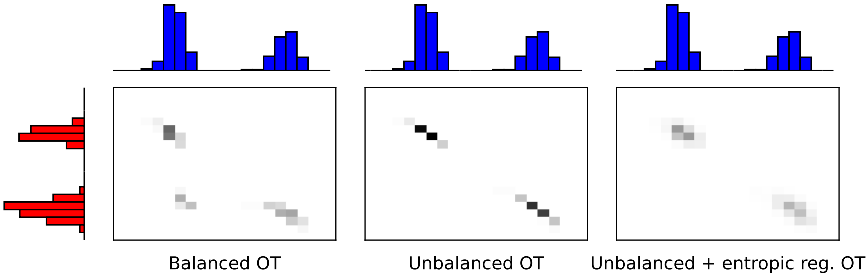
- OT is a powerful formulation for several ML applications.
- But as illustrated by entropic regularization, one can also penalize the optimization problem to get a better/more representative problem.
- Several extensions and variants of OT has been studied by mathematicians and ML practitioners.

Extensions of Optimal Transport

- **Unbalanced OT** and **Partial OT**, can transport between distributions with different total mass and/or outliers
- **Multi-marginal OT**, searches for a transport between more than two distributions.
- **Gromov-Wasserstein OT**, searches for a transport across metric spaces.

Unbalanced and partial Optimal Transport

Limitation of OT: dealing with different masses and/or outliers

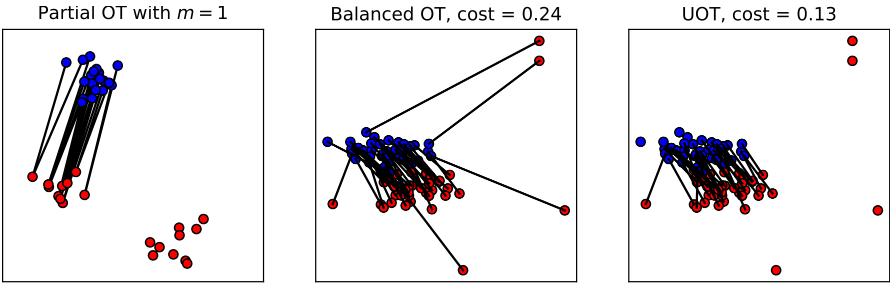


Unbalanced Optimal transport (UOT) [Benamou, 2003]

$$\min_{\mathbf{T}\geq 0} \quad \langle \mathbf{T}, \mathbf{C} \rangle_F + \lambda^u \left(D_\varphi(\mathbf{T}\mathbf{1}_m, \mathbf{a}) + D_\varphi(\mathbf{T}^\top \mathbf{1}_n, \mathbf{b}) \right)$$

- D_φ is a a Bregman divergence penalizing the violation of the marginal constraints.
- $\lambda \to +\infty$ recovers the original OT problem when $\|\mathbf{a}\|_1 = \|\mathbf{b}\|_1$
- Does not transport all the mass and allows distributions with different total mass.

Limitation of OT: dealing with different masses and/or outliers

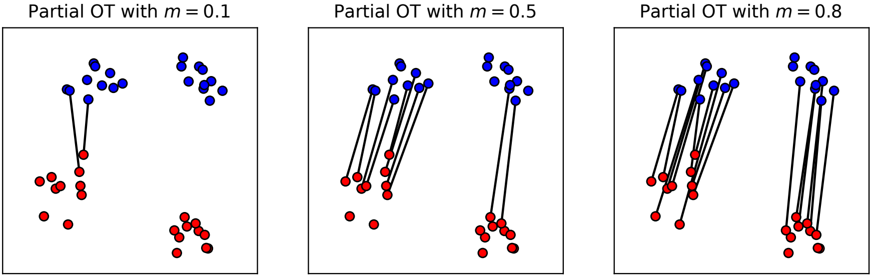


Unbalanced Optimal transport (UOT) [Benamou, 2003]

$$\min_{\mathbf{T}\geq 0} \quad \langle \mathbf{T}, \mathbf{C} \rangle_F + \lambda^u \left(D_\varphi(\mathbf{T}\mathbf{1}_m, \mathbf{a}) + D_\varphi(\mathbf{T}^\top \mathbf{1}_n, \mathbf{b}) \right)$$

- D_φ is a a Bregman divergence penalizing the violation of the marginal constraints.
- $\lambda \to +\infty$ recovers the original OT problem when $\|\mathbf{a}\|_1 = \|\mathbf{b}\|_1$
- Does not transport all the mass and allows distributions with different total mass.

Partial Optimal Transport



Partial OT [Caffarelli and McCann, 2010, Figalli, 2010]

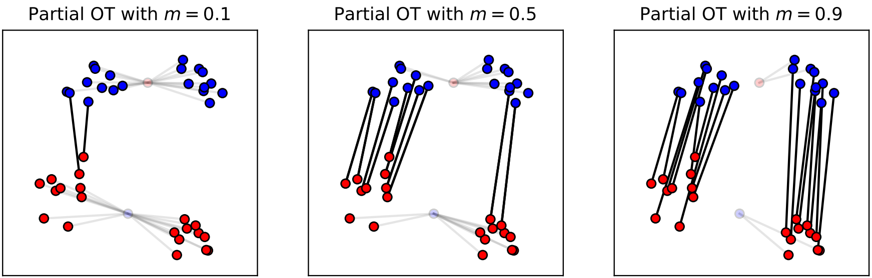
When D_φ is the total variation, comes down to

$$\min_{\mathbf{T}\in\Pi^m(\mu_s,\mu_t)} \quad \left\{ \langle \mathbf{T}, \mathbf{C} \rangle_F = \sum_{i,j} T_{i,j} c_{i,j} \right\}$$

$$\Pi^m(\mu_s, \mu_t) = \left\{ \mathbf{T} \in (\mathbb{R}^+)^{n_s \times n_t} | \, \mathbf{T}\mathbf{1}_{n_t} \leq \mathbf{a}, \mathbf{T}^T \mathbf{1}_{n_s} \leq \mathbf{b}, \mathbf{1}_{n_s}^T \mathbf{T} \mathbf{1}_{n_t} = m \right\}$$

- The equality constraint is on the total transported mass that must be equal to m.
- Allows distributions with different total mass when $m \leq \min(\mathbf{1}_{n_s}^T \mathbf{a}, \mathbf{1}_{n_t}^T \mathbf{b})$

Solving Partial OT



Partial OT solver [Chapel et al., 2020]

- Partial OT can be used solved using standard OT solvers using dummy variables.

$$\min_{\widetilde{\mathbf{T}} \in \widetilde{\Pi}(\color{red}{\mu_s}, \color{blue}{\mu_t})} \quad \left\{ \left\langle \widetilde{\mathbf{T}}, \widetilde{\mathbf{C}} \right\rangle_F = \sum_{i,j} \widetilde{T}_{i,j} \widetilde{c}_{i,j} \right\}$$

where $\widetilde{\Pi}(\color{red}{\mu_s}, \color{blue}{\mu_t}) = \left\{ \widetilde{\mathbf{T}} \in (\mathbb{R}^+)^{n_s+1 \times n_t+1} \middle| \, \widetilde{\mathbf{T}} \mathbf{1}_{n_t+1} = \color{red}{\widetilde{\mathbf{a}}}, \widetilde{\mathbf{T}}^T \mathbf{1}_{n_s+1} = \color{blue}{\widetilde{\mathbf{b}}} \right\}$ and

$$\widetilde{\mathbf{T}} = \begin{bmatrix} \mathbf{T} & \mathbf{q} \\ \mathbf{p}^T & 0 \end{bmatrix}, \quad \widetilde{\mathbf{C}} = \begin{bmatrix} \mathbf{C} & \mathbf{0}_{n_s} \\ \mathbf{0}_{n_t}^T & c_{max} \end{bmatrix}, \quad \color{red}{\widetilde{\mathbf{a}}} = \begin{bmatrix} \color{red}{\mathbf{a}} \\ \color{red}{\mathbf{a}}^T \mathbf{1}_{n_s} - m \end{bmatrix}, \quad \color{blue}{\widetilde{\mathbf{b}}} = \begin{bmatrix} \color{blue}{\mathbf{b}} \\ \color{blue}{\mathbf{b}}^T \mathbf{1}_{n_t} - m \end{bmatrix},$$

where $c_{max} > 0, \ \forall i, j$ and $\mathbf{p}, \mathbf{q}$ contains the mass not transported.

Solving unbalanced OT

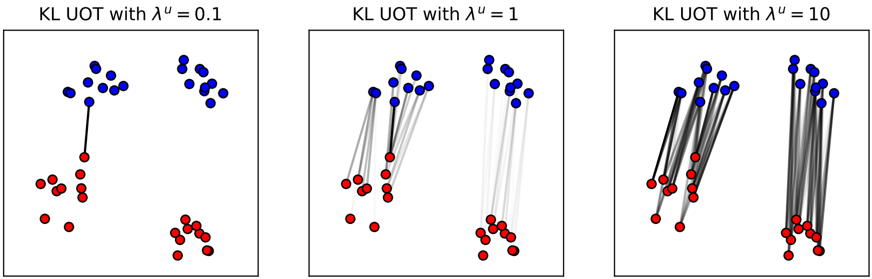
Regularized UOT

$$\min_{\mathbf{T}\geq 0} \quad \langle \mathbf{T}, \mathbf{C} \rangle_F + \lambda^u D_\varphi(\mathbf{T}\mathbf{1}_m, \mathbf{a}) + \lambda^u D_\varphi(\mathbf{T}^\top \mathbf{1}_n, \mathbf{b}) + \lambda \Omega(\mathbf{T}) \tag{1}$$

- Entropic regularization leads to convex problem [Chizat et al., 2018].
- Can be solved in the dual using block coordinate ascent.
- Algorithm similar to sinkhorn (fast, easy to implement).
- Can be debiased to get a proper divergence [Séjourné et al., 2019].

Non regularized UOT

- Smooth convex optimization problem under positivity constraints.
- Classical approach is to use L-BFGS under box constraint [Byrd et al., 1995].
- Problem is actually equivalent to non-negative regression



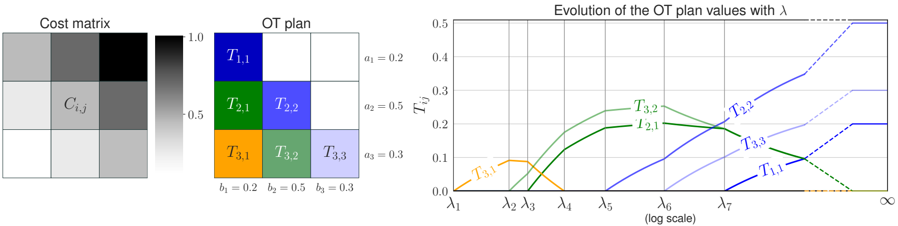
Unbalanced Optimal Transport (UOT) with KL divergence

$$\min_{\mathbf{T}\geq 0} \quad \langle \mathbf{T}, \mathbf{C}\rangle_F + \lambda^u \left(\mathrm{KL}(\mathbf{T}\mathbf{1}_m, \mathbf{a}) + \mathrm{KL}(\mathbf{T}^\top \mathbf{1}_n, \mathbf{b})\right)$$

- Iterative algorithm [Chapel et al., 2021]

$$\boldsymbol{T}^{(k+1)} = \mathrm{diag}\left(\frac{\mathbf{a}}{\boldsymbol{T}^{(k)} 1_{n_t}}\right)^{\frac{1}{2}} \left(\boldsymbol{T}^{(k)} \odot \exp\left(-\frac{\boldsymbol{C}}{2\lambda^u}\right)\right) \mathrm{diag}\left(\frac{\mathbf{b}}{\boldsymbol{T}^{(k)\top} 1_{n_s}}\right)^{\frac{1}{2}}$$

- Amenable to GPU computation



Unbalanced Optimal Transport (UOT) with quadratic divergence

$$\min_{\mathbf{T}\geq 0} \quad \langle \mathbf{T}, \mathbf{C} \rangle_F + \lambda^u \left(\|\mathbf{T}\mathbf{1}_m - \mathbf{a}\|_2 + \left\|\mathbf{T}^\top \mathbf{1}_n, \mathbf{b}\right\|_2 \right)$$

- Solutions are piecewise linear with $1/\lambda^u$
- Resolution algorithm similar to LARS [Chapel et al., 2021]
    1. start with $\lambda^u = 0$
    2. increase λ^u until there is a change on the support of $\mathbf{T}$
    3. update $\mathbf{T}$, solving incrementally a linear equation
    4. repeat until $\lambda^u = \infty$

Multi-marginal Optimal Transport (MMOT)

Multi-marginal Optimal Transport

Optimization problem

- Let $\mu_k = \sum_{i=1}^{n_k} a_i^k \delta_{\mathbf{x}_i^k}$ with $k \in \{1, \ldots, K\}$ be discrete distributions.
- The MMOT problem can be expressed as

$$\min_{\mathbf{T} \in \Pi(\{\mu_k\}_k)} \quad \sum_{k=1}^{K} \sum_{i_k=1}^{n_k} T_{i_1,\ldots,i_K} C_{i_1,\ldots,i_K}$$

- Where $C_{i_1,\ldots,i_K} = c(\mathbf{x}_{i_1}^1, \ldots, \mathbf{x}_{i_K}^K)$ the constant set is defined as:

$$\Pi(\{\mu_k\}_k) = \left\{ \mathbf{T} \in (\mathbb{R}^+)^{n_1 \times \cdots \times n_K} | \sum_{q \neq k} \sum_{i_k=1}^{n_k} T_{i_1,\ldots,i_K} = a_{i_q}, \ \forall q, i_q \in \{1, n_q\} \right\}$$

Properties of MMOT (review in [Pass, 2015])

- Search for a joint distribution (expressed as a tensor for discrete distributions).
- When $K = 2$, $\mathbf{T}$ is a matrix and we recover classical OT problem.
- Can be used to recover the Wasserstein barycenter fo specific cost c [Agueh and Carlier, 2011].

Solving Multi-marginal Optimal Transport

Solving exact MMOT

- Linear program (LP) but with dimensionality exponential in the number of marginal.
- In the primal LP with $\prod_k n_k$ variables and $\sum_k n_k$ constraints.
- Very complex to solve for medium to large scale problems.

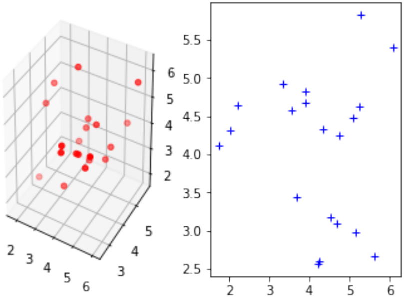
Entropic MMOT

$$\min_{\mathbf{T}\in\Pi(\{\mu_k\}_k)} \quad \sum_{k=1}^{K}\sum_{i_k=1}^{n_k} T_{i_1,\ldots,i_K} C_{i_1,\ldots,i_K} + \lambda\Omega(\mathbf{T})$$

- Problem becomes smooth and strictly convex.
- Can be solved using Bregman projections [Benamou et al., 2015].
- The solution is of the form $\mathbf{T} = \exp(-\mathbf{C}/\lambda) \odot \bigotimes_k \mathbf{u}_k$ where $\mathbf{u}_k$ are positive scaling updated at each projections.

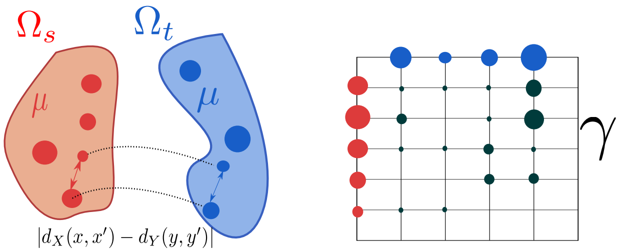
Gromov-Wasserstein and transport across metric spaces

Can you transport between different spaces?



- Ω_s : source space, Ω_t : target space.
- Both domains/spaces do not share the same variables.
- There is no $c(\mathbf{x}, \mathbf{y})$ between the two domains.
- They are related (observe similar objects) but not registered.
- Example: multi-modality with observations on different objects.

Gromov-Wasserstein divergence



Inspired from Gabriel Peyré

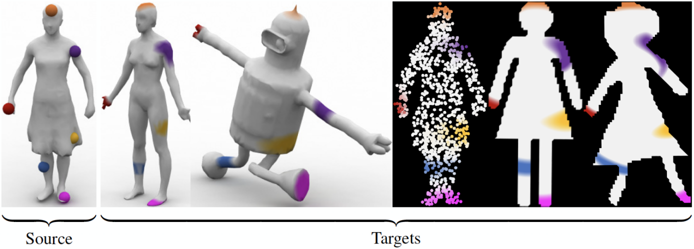
GW for discrete distributions [Memoli, 2011]

$$\mathrm{GW}_p(\mu_s,\mu_t) = \left(\min_{T\in\Pi(\mu_s,\mu_t)} \sum_{i,j,k,l} |D_{i,k} - D'_{j,l}|^p T_{i,j}\, T_{k,l}\right)^{\frac{1}{p}}$$

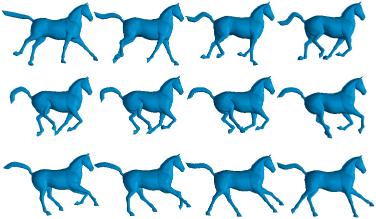
with $\mu_s = \sum_i a_i \delta_{\mathbf{x}_i^s}$ and $\mu_t = \sum_j b_j \delta_{x_j^t}$ and $D_{i,k} = \|\mathbf{x}_i^s - \mathbf{x}_k^s\|, D'_{j,l} = \|\mathbf{x}_j^t - \mathbf{x}_l^t\|$

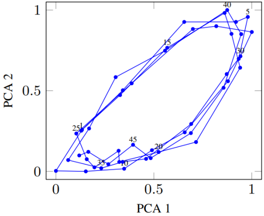
- Distance between metric measured spaces : across different spaces.
- Search for an OT plan that preserve the pairwise relationships between samples.
- Invariant to isometry in either spaces (e.g. rotations and translation).

Shape matching between 3D and 2D surfaces



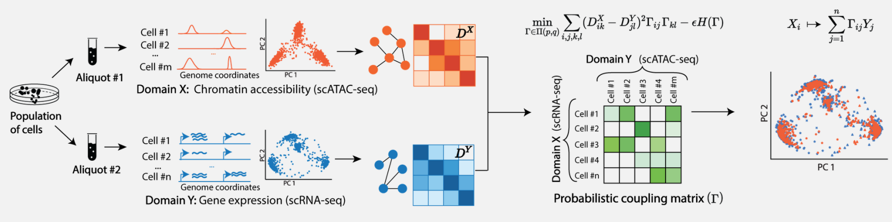
Multidimensional scaling (MDS) of shape collection



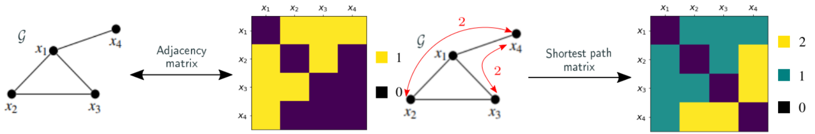


Applications of GW [Demetci et al., 2020]

Unsupervised cell-to-cell alignment of single-cell multi-omic datasets.



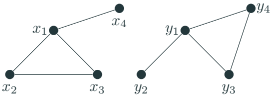
Gromov-Wasserstein between graphs



Modeling the graph structure with a pairwise matrix D

- An undirected graph $\mathcal{G} := (\mathbf{V}, \mathbf{E})$ is defined by $\mathbf{V} = \{\mathbf{x_i}\}_{\mathbf{i} \in [\mathbf{N}]}$ set of the $\mathbf{N}$ nodes and $\mathbf{E} = \{(\mathbf{x_i}, \mathbf{x_j}) | \mathbf{x_i} \leftrightarrow \mathbf{x_j}\}$ set of edges.
- Structure represented as a symmetric matrix $\boldsymbol{D}$ of relations between the nodes.
- Possible choices: **Adjacency matrix**, Laplacian matrix, Shortest path matrix.

Graph as a distribution $(\boldsymbol{D}, \boldsymbol{h})$



- Graph represented as $\mu_X = \sum_i h_i \delta_{x_i}$.
- The positions x_i are implicit and represented as the pairwise matrix $\boldsymbol{D}$.
- h_i are the masses on the nodes of the graphs (uniform by default).

Solving the Gromov Wasserstein optimization problem

$$\mathrm{GW}_p^p(\mu_s, \mu_t) = \min_{\mathbf{T} \in \Pi(\mu_s, \mu_t)} \sum_{i,j,k,l} |D_{i,k} - D'_{j,l}|^p T_{i,j}\, T_{k,l}$$

Optimization problem

- Quadratic Program (Wasserstein is a linear program).
- Nonconvex, NP-hard, related to Quadratic Assignment Problem (QAP).

Optimization algorithm

- Local solution can be obtained with conditional gradient (Frank-Wolfe) [Vayer et al., 2018] (each iteration is an OT problem).
- Iterative algorithm

$$\mathbf{T}^{(t+1)} = \tau^t \operatorname*{argmin}_{\mathbf{T} \in \Pi(\mu_s, \mu_t)} \left\langle \mathbf{G}^{(t)}, \mathbf{T} \right\rangle_F + (1 - \tau^t)\mathbf{T}^{(t)}$$

Where $G_{i,j}^{(t)} = 2 \sum_{k,l} |D_{i,k} - D'_{j,l}|^p T_{k,l}^{(t)}$ is the gradient of the GW loss at previous point $\mathbf{T}^{(t)}$ and τ is a descent step.

Entropic Gromov-Wasserstein

Optimization Problem [Peyré et al., 2016]

$$\mathrm{GW}_{p,\epsilon}^{p}(\mu_s,\mu_t) = \min_{\mathbf{T}\in\Pi(\mu_s,\mu_t)} \sum_{i,j,k,l} |D_{i,k}-D'_{j,l}|^p T_{i,j}\,T_{k,l} + \epsilon \sum_{i,j} T_{i,j}\log T_{i,j} \tag{2}$$

Smoothes the original GW with a convex and smooth entropic term.

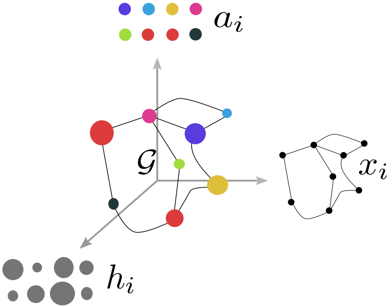
Solving the entropic $\mathcal{GW}$ [Peyré et al., 2016]

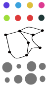
- Problem (3) can be solved using a KL mirror descent.
- This is equivalent to solving at each iteration t

$$\mathbf{T}^{(t+1)} = \min_{\boldsymbol{\gamma}\in\mathcal{P}} \quad \left\langle \mathbf{T}, \mathbf{G}^{(t)} \right\rangle_F + \epsilon \sum_{i,j} T_{i,j}\log T_{i,j}$$

  Where $G_{i,j}^{(t)} = 2\sum_{k,l}|D_{i,k}-D'_{j,l}|^p T_{k,l}^{(t)}$ is the gradient of the GW loss at previous point $\mathbf{T}^{(k)}$.
- Problem above can be solved using a Sinkhorn-Knopp algorithm of entropic OT.
- Very fast approximation exist for low rank distances [Scetbon et al., 2021].

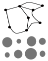
Labeled graphs as distributions





$\Big\} \; \mu = \sum_i h_i \delta_{(x_i, a_i)}$

$\Big\} \; \mu_A = \sum_i h_i \delta_{a_i}$



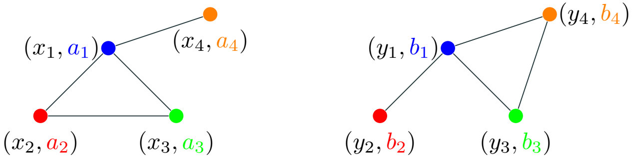
$\Big\} \; \mu_X = \sum_i h_i \delta_{x_i}$

Graph data representation

$$\mu = \sum_{i=1}^{n} h_i \delta_{(x_i a_i)}$$

- Nodes are weighted by their mass h_i.
- But no common metric between the structure points x_i of two different graphs.
- Features values a_i can be compared through the common metric

Fused Gromov-Wasserstein distance



Fused Gromov Wasserstein distance

$\mu_s = \sum_{i=1}^{n} h_i \delta_{x_i, a_i}$ and $\mu_t = \sum_{j=1}^{m} g_j \delta_{y_j, b_j}$

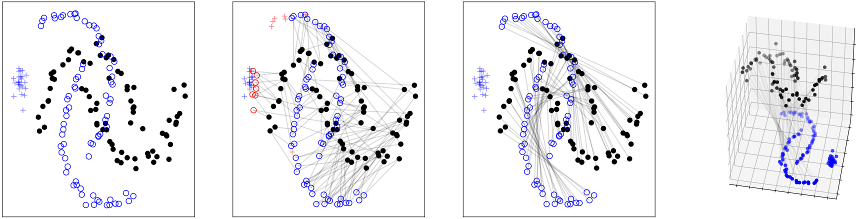
$$\mathrm{FGW}_{p,q,\alpha}(D, D', \mu_s, \mu_t) = \left(\min_{\mathbf{T} \in \Pi(\mu_s, \mu_t)} \sum_{i,j,k,l} \left((1-\alpha) C_{i,j}^{q} + \alpha |D_{i,k} - D'_{j,l}|^{q} \right)^{p} T_{i,j} \, T_{k,l} \right)^{\frac{1}{p}}$$

with $D_{i,k} = \|x_i - x_k\|$ and $D'_{j,l} = \|y_i - y_l\|$ and $C_{i,j} = \|a_i - b_j\|$

- Parameters $q > 1$, $\forall p \geq 1$.
- $\alpha \in [0, 1]$ is a trade off parameter between structure and features.
- Can be solved like a GW problem

Unbalanced and Partial Gromov-Wasserstein

Unbalanced Gromov-Wasserstein



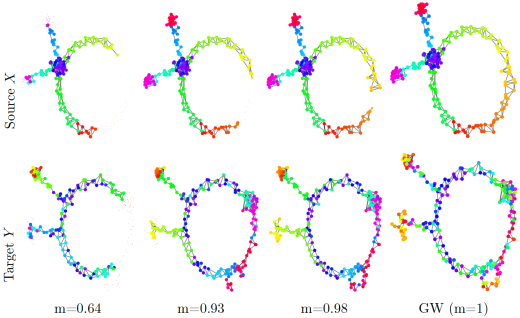
Unbalanced Gromov-Wasserstein

$$\mathsf{UGW}_p^p(\color{red}{\mu_s}, \color{blue}{\mu_t}) = \min_{\mathbf{T}\geq 0} \quad \langle \boldsymbol{L} \odot \mathbf{T}, \mathbf{T} \rangle + \lambda^u \left(D_\varphi(\mathbf{T}\mathbf{1}_m, \color{red}{\mathbf{a}}) + D_\varphi(\mathbf{T}^\top \mathbf{1}_n, \color{blue}{\mathbf{b}}) \right)$$

with $\sum_{i,j,k,l} |\color{red}{D_{i,k}} - \color{blue}{D'_{j,l}}|^q| = L_{ijkl}$ or [Séjourné et al., 2020]

$$\mathsf{UGW}_p^p(\color{red}{\mu_s}, \color{blue}{\mu_t}) = \min_{\mathbf{T}\geq 0} \quad \langle \boldsymbol{L} \odot \mathbf{T}, \mathbf{T} \rangle + \lambda^u \left(D_\varphi(\mathbf{T}\mathbf{1}_m \odot \mathbf{T}\mathbf{1}_m, \color{red}{\mathbf{a} \odot \mathbf{a}}) + D_\varphi(\mathbf{T}^\top \mathbf{1}_n \odot \mathbf{T}^\top \mathbf{1}_n, \color{blue}{\mathbf{b} \odot \mathbf{b}}) \right)$$

Solving Unbalanced Gromov-Wasserstein



From [Séjourné et al., 2020]

Partial Gromov-Wasserstein

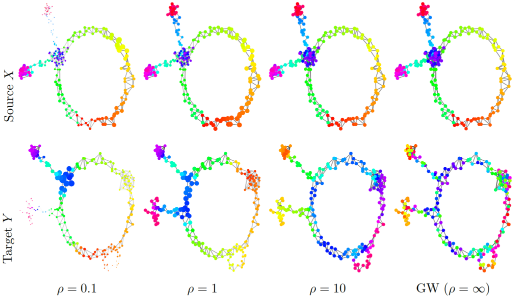
- Where D_φ =TV, it gives rise to the partial Gromov-Wassertein.

$$\min_{\mathbf{T}\in\Pi^m(\mu_s,\mu_t)} \sum_{i,j,k,l} |D_{i,k} - D'_{j,l}|^p T_{i,j}\, T_{k,l}$$

$$\Pi^m(\mu_s,\mu_t) = \left\{\mathbf{T} \in (\mathbb{R}^+)^{n_s\times n_t} \,|\, \mathbf{T}\mathbf{1}_{n_t} \leq \mathbf{a}, \mathbf{T}^T\mathbf{1}_{n_s} \leq \mathbf{b}, \mathbf{1}_{n_s}^T\mathbf{T}\mathbf{1}_{n_t} = m\right\}$$

- FW based-resolution algorithm, with inner loop solving partial-OT.

Solving Unbalanced Gromov-Wasserstein



From [Séjourné et al., 2020]

Unbalanced Gromov-Wasserstein: conic formulation

- With quadratic penalties [Séjourné et al., 2020]

$$\mathrm{UGW}_p^p(\mu_s, \mu_t) = \min_{\mathbf{T} \in \Pi(\mu_s, \mu_t)} \langle \boldsymbol{L} \odot \mathbf{T}, \mathbf{T} \rangle + \lambda^u \left(D_\varphi(\mathbf{T}\mathbf{1}_m \odot \mathbf{T}\mathbf{1}_m, \mathbf{a} \odot \mathbf{a}) + D_\varphi(\mathbf{T}^\top \mathbf{1}_n \odot \mathbf{T}^\top \mathbf{1}_n, \mathbf{b} \odot \mathbf{b}) \right)$$

- Can be approximated using a bi-convex relaxation.

Summary for Part 2

Optimal transport

- Theoretically grounded ways of comparing probability distributions.
- Several variants exists depending on the application.
- Unbalanced/partial OT relaxes the marginal constraints.
- Multi-marginal OT when comparing more than 2 distributions.
- Gromov-Wasserstein for data living in different spaces.

Optimization

- Solving OT is a linear program, GW a quadratic program.
- (Entropic) unbalanced Wasserstein can be solved efficiently.
- Reference for computational OT: [Peyré et al., 2019].

Next step: how to use it in machine learning applications ?

References i

[Agueh and Carlier, 2011] Agueh, M. and Carlier, G. (2011).
Barycenters in the wasserstein space.
SIAM Journal on Mathematical Analysis, 43(2):904–924.

[Benamou, 2003] Benamou, J.-D. (2003).
Numerical resolution of an “unbalanced” mass transport problem.
ESAIM: Mathematical Modelling and Numerical Analysis, 37(5):851–868.

[Benamou et al., 2015] Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., and Peyré, G. (2015).
Iterative Bregman projections for regularized transportation problems.
SISC.

[Byrd et al., 1995] Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995).
A limited memory algorithm for bound constrained optimization.
SIAM Journal on scientific computing, 16(5):1190–1208.

[Caffarelli and McCann, 2010] Caffarelli, L. A. and McCann, R. J. (2010).
Free boundaries in optimal transport and monge-ampere obstacle problems.
Annals of mathematics, pages 673–730.

[Chapel et al., 2020] Chapel, L., Alaya, M. Z., and Gasso, G. (2020).
Partial optimal tranport with applications on positive-unlabeled learning.
Advances in Neural Information Processing Systems, 33:2903–2913.

[Chapel et al., 2021] Chapel, L., Flamary, R., Wu, H., Févotte, C., and Gasso, G. (2021).
Unbalanced optimal transport through non-negative penalized linear regression.
In *Neural Information Processing Systems (NeurIPS)*.

[Chizat et al., 2018] Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X. (2018).
Scaling algorithms for unbalanced optimal transport problems.
Mathematics of Computation, 87(314):2563–2609.

References iii

[Demetci et al., 2020] Demetci, P., Santorella, R., Sandstede, B., Noble, W. S., and Singh, R. (2020).

Gromov-wasserstein optimal transport to align single-cell multi-omics data.

bioRxiv.

[Figalli, 2010] Figalli, A. (2010).

The optimal partial transport problem.

Archive for rational mechanics and analysis, 195(2):533–560.

[Memoli, 2011] Memoli, F. (2011).

Gromov wasserstein distances and the metric approach to object matching.

Foundations of Computational Mathematics, pages 1–71.

[Pass, 2015] Pass, B. (2015).

Multi-marginal optimal transport: theory and applications.

ESAIM: Mathematical Modelling and Numerical Analysis, 49(6):1771–1790.

References iv

[Peyré et al., 2019] Peyré, G., Cuturi, M., et al. (2019).

Computational optimal transport: With applications to data science.

Foundations and Trends® in Machine Learning, 11(5-6):355–607.

[Peyré et al., 2016] Peyré, G., Cuturi, M., and Solomon, J. (2016).

Gromov-wasserstein averaging of kernel and distance matrices.

In *ICML*, pages 2664–2672.

[Scetbon et al., 2021] Scetbon, M., Peyré, G., and Cuturi, M. (2021).

Linear-time gromov wasserstein distances using low rank couplings and costs.

arXiv preprint arXiv:2106.01128.

[Séjourné et al., 2019] Séjourné, T., Feydy, J., Vialard, F.-X., Trouvé, A., and Peyré, G. (2019).

Sinkhorn divergences for unbalanced optimal transport.

arXiv preprint arXiv:1910.12958.

References v

[Séjourné et al., 2020] Séjourné, T., Vialard, F.-X., and Peyré, G. (2020).
The unbalanced gromov wasserstein distance: Conic formulation and relaxation.
arXiv preprint arXiv:2009.04266.

[Solomon et al., 2016] Solomon, J., Peyré, G., Kim, V. G., and Sra, S. (2016).
Entropic metric alignment for correspondence problems.
ACM Transactions on Graphics (TOG), 35(4):72.

[Vayer et al., 2018] Vayer, T., Chapel, L., Flamary, R., Tavenard, R., and Courty, N. (2018).
Fused gromov-wasserstein distance for structured objects: theoretical foundations and mathematical properties.