UNBALANCED OPTIMAL TRANSPORT
THEORY AND APPLICATIONS

Lénaïc Chizat[1]

[1]DMA, École Normale Supérieure
Advisors : G. Peyré and F-X. Vialard

SUMMARY

The need to extend optimal transport (OT) theory to measures of arbitrary mass has often come forward in applications but there was lacking a general *unbalanced* optimal transport theory until 2015, when several papers appeared on the subject. In this overview, we present (1) a selection of theoretical results, (2) a numerical algorithm and (3) illustrations and applications.

CLASSICAL OPTIMAL TRANSPORT

Given probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ on measurable spaces and a *cost* function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$, the OT problem writes

$$C(\mu,\nu) \overset{\text{def}}{=} \min_{\gamma\in\Pi(\mu,\nu)} \left\{ \int_{X\times Y} c(x,y)\,\mathrm{d}\gamma(x,y) \right\}$$

where $\Pi(\mu,\nu)$ is the set of couplings between μ and ν i.e. measures $\gamma \in \mathcal{P}(X\times Y)$ such that $P^X_\# \gamma = \mu$ and $P^Y_\# \gamma = \nu$.

Key results:

- characterization of minimizers through convex duality theory;
- if $c(x,y) = dist(x,y)^2$ then $C(\mu,\nu)^{\frac{1}{2}}$ is the *Wasserstein* metric on $\mathcal{P}(X)$;
- fluid dynamic formulation if $X = Y \subset \mathbb{R}^d$ and $c(x,y) = |y-x|^2$:

$$C(\mu,\nu) = \inf_{\substack{\rho_0=\mu\\ \rho_1=\nu}} \left\{ \int_0^1 \left(\int_X |v_t(x)|_2^2\, \mathrm{d}\rho_t(x) \right) \mathrm{d}t \,;\, \partial_t \rho + \mathrm{div}(\rho v) = 0 \right\} .$$

Some applications (among many others):

- PDEs : gives a metric structure on $\mathcal{P}(X)$ for gradient flows;
- nonparametric comparison of data distributions in machine learning;
- image registration, shape interpolation, color transfer...

In many applications, the mass constraint on the marginals is not natural.

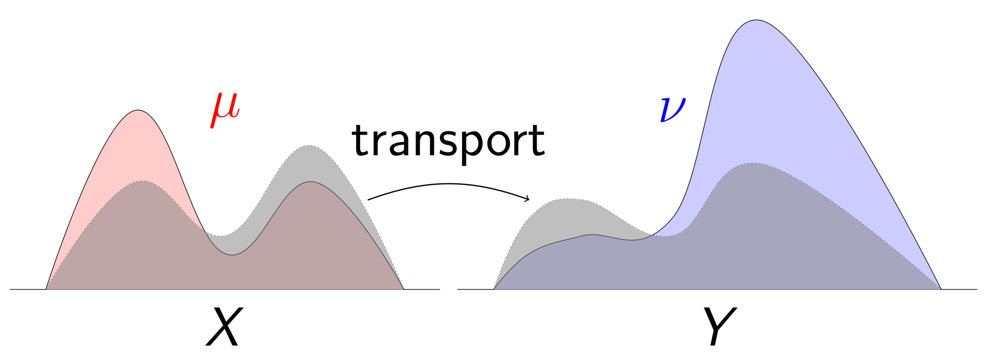
UNBALANCED OPTIMAL TRANSPORT

Optimal entropy-transport problems formulation [6]

Given nonnegative measures $\mu \in \mathcal{M}_+(X)$, $\nu \in \mathcal{M}_+(Y)$ and a cost c, solve

$$C(\mu,\nu) \overset{\text{def}}{=} \min_{\gamma\in\mathcal{M}_+(X\times Y)} \left\{ \int_{X\times Y} c(x,y)\,\mathrm{d}\gamma(x,y) + D_X(P^X_\#\gamma|\mu) + D_Y(P^Y_\#\gamma|\nu) \right\}$$

where D_X and D_Y are convex *φ-divergence* functionals.



	Name of $C(\mu,\nu)$	Constraint/divergence $D(\lambda\|\mu)$	Cost $c(x,y)$
(i)	W_2^2	$\lambda = \mu$	$d(x,y)^2$
(ii)	Partial OT	$\|\lambda-\mu\|_{TV}$	-
(iii)	Range OT	$\lambda \in [\alpha\mu, \beta\mu]$	-
(iv)	GHK^2	$\mathrm{KL}(\lambda\|\mu)$	$d(x,y)^2$
(v)	WF^2	$\mathrm{KL}(\lambda\|\mu)$	$-\log\cos^2_+(d(x,y))$

Table 1: Examples of entropy-transport problems. Here d metric on $X = Y$ and KL is the Kullback-Leibler divergence defined as $\int_X (\sigma\log(\sigma) - \sigma + 1)\,\mathrm{d}\mu$ if $\lambda = \sigma\mu$, $+\infty$ otherwise.

Classical results in optimal transport theory have their "unbalanced" counterpart. In particular, the natural extension of W_2 is WF in Table 1 (v):

The Wasserstein-Fisher-Rao metric [1, 5, 6]

The quantity WF defined in Table 1(v):

- defines a complete metric on $\mathcal{M}_+(X)$ (geodesic if X is geodesic);
- its square is equal to the dynamic formulation

$$\inf_{\substack{\rho_0=\mu\\ \rho_1=\nu}} \left\{ \int_0^1 \left(\int_X (|v_t(x)|_2^2 + \tfrac{1}{4} g_t(x)^2)\, \mathrm{d}\rho_t(x) \right) \mathrm{d}t \,;\, \partial_t \rho + \mathrm{div}(\rho v) = \rho g \right\}$$

- endows $\mathcal{M}_+(X)$ with a "Riemannian-like" structure which tensor is an inf-convolution between the Wasserstein and the Fisher-Rao tensor.

Geodesic for WF on $\mathcal{M}_+(\mathbb{R}^2)$

NUMERICAL RESOLUTION

If $X = (x_i)_I$ and $Y = (y_j)_J$ are discretized, solving unbalanced optimal transport problems (and WF barycenters, gradient flows...) requires to solve problems of the form (with F_1, F_2 convex, l.s.c.)

$$\min_{\gamma\in\mathbb{R}^{I\times J}} \sum_{i,j} c_{i,j}\gamma_{i,j} + F_1\Big(\sum_i \gamma_{i,j}\Big) + F_2\Big(\sum_j \gamma_{i,j}\Big) + \iota_{\mathbb{R}_+^{I\times J}}(\gamma)$$

Following a now standard technique in classical OT, we propose to:

- replace positivity constraint by the entropy barrier $\epsilon\sum\gamma_{i,j}(\log(\gamma_{i,j})-1)$;
- solve this smoothed problem with the *iterative scaling* algorithm (below);
- optionally decrease ϵ, solve again with a better initialization, and so on.

Iterative scaling algorithm

Defining $K = (\exp(-c_{i,j}/\epsilon))_{I\times J}$, $\mathrm{prox}^{\mathrm{KL}}_F(x) = \mathrm{argmin}_y\, F(y) + \mathrm{KL}(y|x)$ and $b^{(0)} = \mathbb{1}_I$, compute iteratively:

$$b^{(\ell+1)} \leftarrow \frac{\mathrm{prox}^{\mathrm{KL}}_{F_1/\epsilon}(Ka^\ell)}{Ka^\ell} \quad \text{and} \quad a^{(\ell+1)} \leftarrow \frac{\mathrm{prox}^{\mathrm{KL}}_{F_2/\epsilon}(K^T b^{\ell+1})}{K^T b^{\ell+1}} .$$

Proposition: $(a_i^\ell K_{i,j} b_j^\ell)_{i,j}$ converges to a minimizer of the smoothed problem.

ILLUSTRATIONS AND APPLICATIONS

Unbalanced optimal transport in 1d

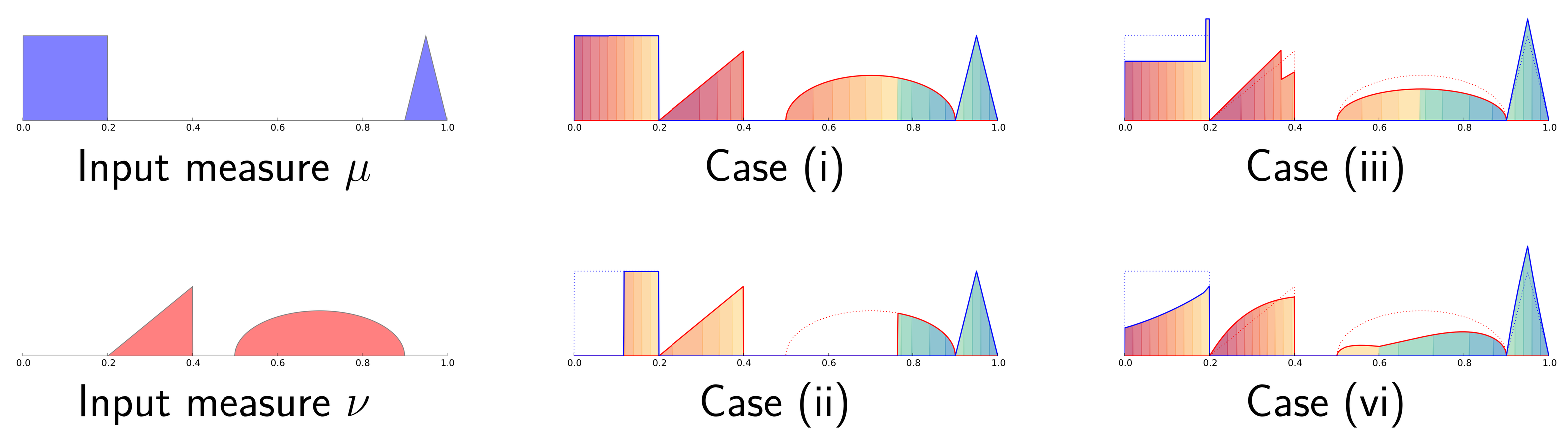
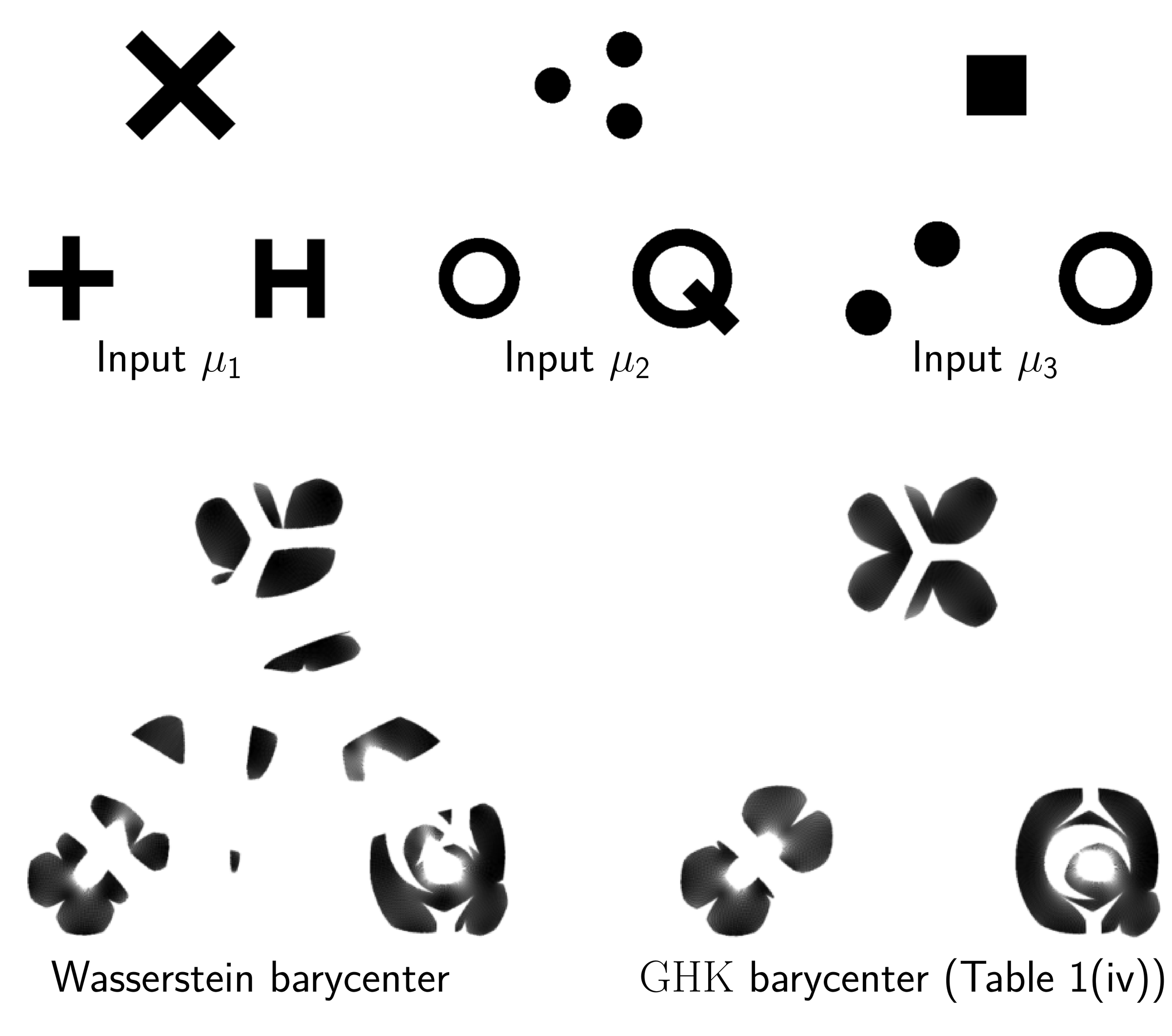
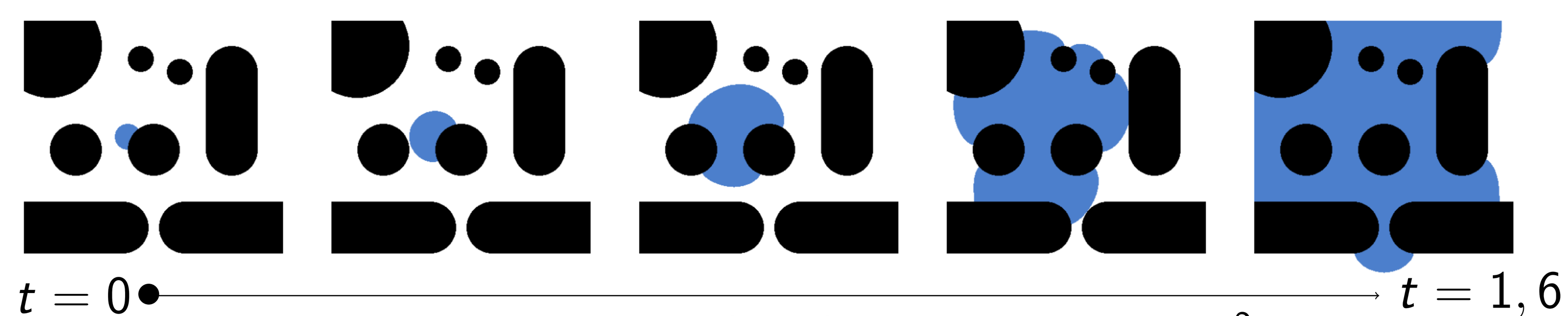


Figure 1: Marginals of γ^* for problems in Table 1 with $c(x,y) = |y-x|^2$ and $X = [0,1]$.

Barycenters in 2d



Gradient flow of the functional $G(\mu) \overset{\text{def}}{=} -\mu(X)$ on the space of positive densities smaller than 1, endowed with the metric WF: one recovers mechanical tumor growth models of Hele-Shaw type [4].



Evolution of a tumor (blue) on a domain in $\mathbb{R}^2$.

REFERENCES

[1] L. Chizat, G. Peyré, B. Schmitzer, F.-X. Vialard, An interpolating distance between optimal transport and Fisher-Rao metrics, FoCM, 2015.

[2] —, Unbalanced optimal transport: geometry and Kantorovich formulation, 2015.

[3] —, Scaling algorithms for unbalanced transport problems, 2016.

[4] S. Di Marino, L. Chizat, A tumor growth Hele-Shaw problem as a gradient flow, in preparation.

[5] S. Kondratyev, L. Monsaingeon, D. Vorotnikov, A new optimal transport distance on the space of finite radon measures, 2015.

[6] M. Liero, A. Mielke, G. Savaré, Optimal entropy-transport problems and the Hellinger-Kantorovich distance , 2015