An homogeneous formulation for Unbalanced Regularized Optimal Transport

Journée thématique OT-ML
Institut Henri Poincaré
18 novembre 2021

theo.lacombe@univ-eiffel.fr

Standard formulation for regularized OT

Let α, β be two (non-negative, finite) Radon measures supported on a compact domain $\Omega \subset \mathbb{R}^d$. Let $c : \Omega \times \Omega \to \mathbb{R}_+$, $\epsilon > 0$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a divergence (and $\varphi^*(q) := \sup_p \{pq - \varphi(p)\}$ be its Legendre transform). The unbalanced, regularized Optimal Transport problem is defined as

$$\mathrm{OT}_\epsilon(\alpha, \beta) := \inf_{\pi \in \mathcal{M}_+(\Omega \times \Omega)} \underbrace{\langle \pi, c \rangle}_{\text{transport cost}} + \underbrace{D_\varphi(\pi_1|\alpha) + D_\varphi(\pi_2|\beta)}_{\text{marginal errors}} + \underbrace{\epsilon \mathrm{KL}(\pi|\alpha \otimes \beta)}_{\text{entropic regularization}} \quad (1)$$

Standard formulation for regularized OT

Let α, β be two (non-negative, finite) Radon measures supported on a compact domain $\Omega \subset \mathbb{R}^d$. Let $c : \Omega \times \Omega \to \mathbb{R}_+$, $\epsilon > 0$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a divergence (and $\varphi^*(q) := \sup_p \{pq - \varphi(p)\}$ be its Legendre transform). The unbalanced, regularized Optimal Transport problem is defined as

$$\mathrm{OT}_\epsilon(\alpha, \beta) := \inf_{\pi \in \mathcal{M}_+(\Omega \times \Omega)} \underbrace{\langle \pi, c \rangle}_{\text{transport cost}} + \underbrace{D_\varphi(\pi_1 | \alpha) + D_\varphi(\pi_2 | \beta)}_{\text{marginal errors}} + \underbrace{\epsilon \mathrm{KL}(\pi | \alpha \otimes \beta)}_{\text{entropic regularization}} \tag{1}$$

$$= \sup_{f, g \in \mathcal{C}(\Omega)} \langle -\varphi^*(-f), \alpha \rangle + \langle -\varphi^*(-g), \beta \rangle - \epsilon \langle e^{\frac{f \oplus g - c}{\epsilon}} - 1, \alpha \otimes \beta \rangle \tag{2}$$

Standard formulation for regularized OT

Let α, β be two (non-negative, finite) Radon measures supported on a compact domain $\Omega \subset \mathbb{R}^d$. Let $c : \Omega \times \Omega \to \mathbb{R}_+$, $\epsilon > 0$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a divergence (and $\varphi^*(q) := \sup_p \{pq - \varphi(p)\}$ be its Legendre transform). The unbalanced, regularized Optimal Transport problem is defined as

$$\mathrm{OT}_\epsilon(\alpha, \beta) := \inf_{\pi \in \mathcal{M}_+(\Omega \times \Omega)} \underbrace{\langle \pi, c \rangle}_{\text{transport cost}} + \underbrace{D_\varphi(\pi_1|\alpha) + D_\varphi(\pi_2|\beta)}_{\text{marginal errors}} + \underbrace{\epsilon \mathrm{KL}(\pi|\alpha \otimes \beta)}_{\text{entropic regularization}} \tag{1}$$

$$= \sup_{f, g \in \mathcal{C}(\Omega)} \langle -\varphi^*(-f), \alpha \rangle + \langle -\varphi^*(-g), \beta \rangle - \epsilon \langle e^{\frac{f \oplus g - c}{\epsilon}} - 1, \alpha \otimes \beta \rangle \tag{2}$$

Proposition: [Séjourné et al.,2021] Define $\mathrm{aprox} : p \mapsto -\arg\min_q \left\{ \epsilon \exp\left(\frac{p-q}{\epsilon}\right) + \varphi^*(q) \right\}$. Let $f_0, g_0 \in \mathcal{C}(\Omega)$, and let

$$\boxed{\begin{aligned} f_{t+1} &= -\mathrm{aprox}\left(\epsilon \log \int e^{\frac{g_t(y) - c(\cdot, y)}{\epsilon}} \mathrm{d}\beta(y)\right) \\ g_{t+1} &= -\mathrm{aprox}\left(\epsilon \log \int e^{\frac{f_{t+1}(x) - c(x, \cdot)}{\epsilon}} \mathrm{d}\alpha(x)\right). \end{aligned}}$$

Under mild assumptions, $(f_t, g_t)_t \to (f^*, g^*)$ optimal for (2).
Note: using $m(\alpha) = m(\beta)$ and $\varphi^* = \mathrm{id}$ ($\Rightarrow \mathrm{aprox} = \mathrm{id}$), we retrieve the Balanced OT model.

Standard formulation for regularized OT

Let α, β be two (non-negative, finite) Radon measures supported on a compact domain $\Omega \subset \mathbb{R}^d$. Let $c : \Omega \times \Omega \to \mathbb{R}_+$, $\epsilon > 0$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a divergence (and $\varphi^*(q) := \sup_p \{pq - \varphi(p)\}$ be its Legendre transform). The unbalanced, regularized Optimal Transport problem is defined as

$$\mathrm{OT}_\epsilon(\alpha, \beta) := \inf_{\pi \in \mathcal{M}_+(\Omega \times \Omega)} \underbrace{\langle \pi, c \rangle}_{\text{transport cost}} + \underbrace{D_\varphi(\pi_1|\alpha) + D_\varphi(\pi_2|\beta)}_{\text{marginal errors}} + \underbrace{\epsilon \mathrm{KL}(\pi|\alpha \otimes \beta)}_{\text{entropic regularization}} \tag{1}$$

$$= \sup_{f, g \in \mathcal{C}(\Omega)} \langle -\varphi^*(-f), \alpha \rangle + \langle -\varphi^*(-g), \beta \rangle - \epsilon \langle e^{\frac{f \oplus g - c}{\epsilon}} - 1, \alpha \otimes \beta \rangle \tag{2}$$

Proposition: [Séjourné et al.,2021] If (f, g) are optimal for (2), then

$$\pi = \exp\left(\frac{f \oplus g - c}{\epsilon}\right) \mathrm{d}\alpha \mathrm{d}\beta$$

is optimal for (1).

Homogeneity in standard regularized OT

Observation: If we replace α, β by $\lambda\alpha, \lambda\beta$ for some $\lambda > 0$, the dual objective function becomes

$$\lambda \langle -\varphi^*(-f), \alpha\rangle + \lambda \langle -\varphi^*(-g), \beta\rangle - \lambda^2 \epsilon \langle e^{\frac{f \oplus g - c}{\epsilon}} - 1, \alpha \otimes \beta\rangle ,$$

Quadratic term, feels like $\epsilon \to \lambda\epsilon$

and then

$$\mathrm{OT}_\epsilon(\lambda\alpha, \lambda\beta) \neq \lambda \cdot \mathrm{OT}_\epsilon(\alpha, \beta)$$

when $\epsilon > 0$, while this holds when $\epsilon = 0$.

Homogeneity in standard regularized OT

How bad is it?

- In the balanced case ($m(\alpha) = m(\beta) = m$, $\varphi^* = \mathrm{id}$, and $\mathrm{aprox} = \mathrm{id}$):

Homogeneity in standard regularized OT

How bad is it?

- In the balanced case ($m(\alpha) = m(\beta) = m$, $\varphi^* = \mathrm{id}$, and $\mathrm{aprox} = \mathrm{id}$):

  Proposition: Fix (f_0, g_0) as the initial state for Sinkhorn iterations. Let $(f_t, g_t)_t$ be the sequence generated for the couple (α, β).

Homogeneity in standard regularized OT

How bad is it?

- In the balanced case ($m(\alpha) = m(\beta) = m$, $\varphi^* = \mathrm{id}$, and $\mathrm{aprox} = \mathrm{id}$):

  Proposition: Fix (f_0, g_0) as the initial state for Sinkhorn iterations. Let $(f_t, g_t)_t$ be the sequence generated for the couple (α, β). Then, the sequence generated for the couple $(\lambda\alpha, \lambda\beta)$ is $(f_t - \epsilon \log(\lambda), g_t)_t$.

Homogeneity in standard regularized OT

How bad is it?

- In the balanced case ($m(\alpha) = m(\beta) = m$, $\varphi^* = \mathrm{id}$, and $\mathrm{aprox} = \mathrm{id}$):

Proposition: Fix (f_0, g_0) as the initial state for Sinkhorn iterations. Let $(f_t, g_t)_t$ be the sequence generated for the couple (α, β). Then, the sequence generated for the couple $(\lambda\alpha, \lambda\beta)$ is $(f_t - \epsilon\log(\lambda), g_t)_t$.

Corollary: If (f, g) is optimal for (α, β), then $(f - \epsilon\log(\lambda), g)$ is optimal for $(\lambda\alpha, \lambda\beta)$.

$\to$ There is a shift in $f \oplus g$ which diverges as $\lambda \to \infty$ (or $\lambda \to 0$).

Homogeneity in standard regularized OT

How bad is it?

- In the balanced case ($m(\alpha) = m(\beta) = m$, $\varphi^* = \mathrm{id}$, and $\mathrm{aprox} = \mathrm{id}$):

  Proposition: Fix (f_0, g_0) as the initial state for Sinkhorn iterations. Let $(f_t, g_t)_t$ be the sequence generated for the couple (α, β). Then, the sequence generated for the couple $(\lambda\alpha, \lambda\beta)$ is $(f_t - \epsilon \log(\lambda), g_t)_t$.

  Corollary: If (f, g) is optimal for (α, β), then $(f - \epsilon \log(\lambda), g)$ is optimal for $(\lambda\alpha, \lambda\beta)$.

  $\to$ There is a shift in $f \oplus g$ which diverges as $\lambda \to \infty$ (or $\lambda \to 0$).

  Corollary: If π is an optimal transport plan between α and β, then $\lambda\pi$ is optimal for $(\lambda\alpha, \lambda\beta)$.

  $\to$ Good news: the optimal transport plan **is** homogeneous is this setting!

Homogeneity in standard regularized OT

How bad is it?

- In the balanced case ($m(\alpha) = m(\beta) = m$, $\varphi^* = \mathrm{id}$, and $\mathrm{aprox} = \mathrm{id}$):

Proposition: Fix (f_0, g_0) as the initial state for Sinkhorn iterations. Let $(f_t, g_t)_t$ be the sequence generated for the couple (α, β). Then, the sequence generated for the couple $(\lambda\alpha, \lambda\beta)$ is $(f_t - \epsilon \log(\lambda), g_t)_t$.

Corollary: If (f, g) is optimal for (α, β), then $(f - \epsilon \log(\lambda), g)$ is optimal for $(\lambda\alpha, \lambda\beta)$.

$\to$ There is a shift in $f \oplus g$ which diverges as $\lambda \to \infty$ (or $\lambda \to 0$).

Corollary: If π is an optimal transport plan between α and β, then $\lambda\pi$ is optimal for $(\lambda\alpha, \lambda\beta)$.

$\to$ Good news: the optimal transport plan **is** homogeneous is this setting!

Corollary: $\mathrm{OT}_\epsilon(\lambda\alpha, \lambda\beta) = \lambda \mathrm{OT}_\epsilon(\alpha, \beta) + \epsilon\lambda(\lambda - 1)m^2 - \epsilon\lambda \log(\lambda) m$

$\to$ Optimal value is not homogeneous and tends to be over-estimated.

Homogeneity in standard regularized OT

How bad is it?

- In the balanced case ($m(\alpha) = m(\beta) = m$, $\varphi^* = \mathrm{id}$, and $\mathrm{aprox} = \mathrm{id}$):

Proposition: Fix (f_0, g_0) as the initial state for Sinkhorn iterations. Let $(f_t, g_t)_t$ be the sequence generated for the couple (α, β). Then, the sequence generated for the couple $(\lambda\alpha, \lambda\beta)$ is $(f_t - \epsilon \log(\lambda), g_t)_t$.

Corollary: If (f, g) is optimal for (α, β), then $(f - \epsilon \log(\lambda), g)$ is optimal for $(\lambda\alpha, \lambda\beta)$.

$\rightarrow$ There is a shift in $f \oplus g$ which diverges as $\lambda \rightarrow \infty$ (or $\lambda \rightarrow 0$).

Corollary: If π is an optimal transport plan between α and β, then $\lambda\pi$ is optimal for $(\lambda\alpha, \lambda\beta)$.

$\rightarrow$ Good news: the optimal transport plan **is** homogeneous is this setting!

Corollary: $\mathrm{OT}_\epsilon(\lambda\alpha, \lambda\beta) = \lambda \mathrm{OT}_\epsilon(\alpha, \beta) + \epsilon\lambda(\lambda - 1)m^2 - \epsilon\lambda \log(\lambda) m$

$\rightarrow$ Optimal value is not homogeneous and tends to be over-estimated.

Corollary: The (balanced) Sinkhorn divergence $\mathrm{Sk}_\epsilon(\alpha, \beta) := \mathrm{OT}_\epsilon(\alpha, \beta) - \frac{1}{2}\mathrm{OT}_\epsilon(\alpha, \alpha) - \frac{1}{2}\mathrm{OT}_\epsilon(\beta, \beta)$ **is** homogeneous!

Homogeneity in standard regularized OT

How bad is it?

- In the balanced case ($m(\alpha) = m(\beta) = m$, $\varphi^* = \mathrm{id}$, and $\mathrm{aprox} = \mathrm{id}$):

  $\Rightarrow$ Things are going well.

  One (potential) numerical issue:

  It is natural to fix a stopping criterion in the Sinkhorn iterations in terms of relative variations of the objective value:

```
IF (new_ot_estim - ot_estim)/ot_estim < crit:
    STOP.
```

  Non-homogeneity hinders this approach.

Homogeneity in standard regularized OT

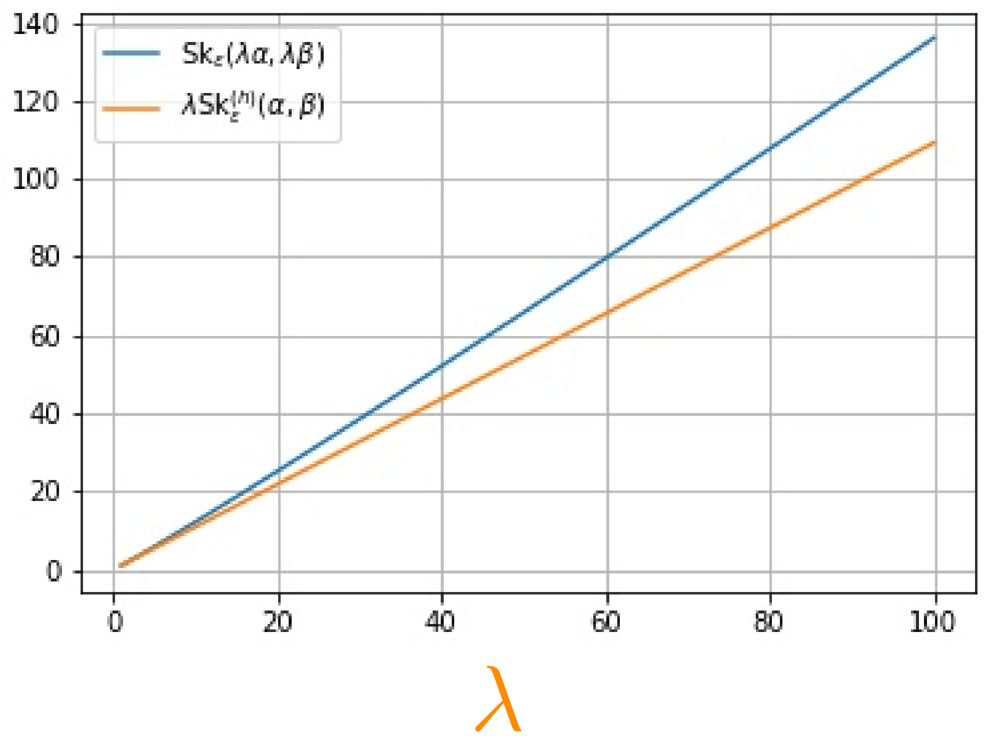
How bad is it?

- And in the general case, where $\varphi^* \neq \mathrm{id}$ (hence $\mathrm{aprox} \neq \mathrm{id}$)?

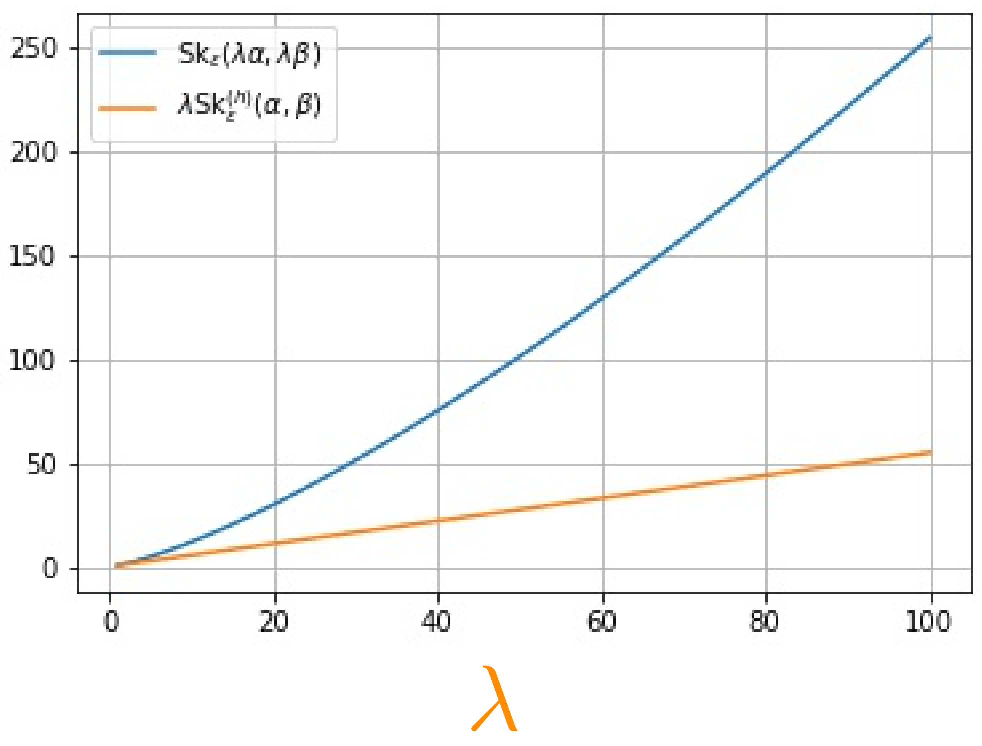
> Things get worse: Sk_ϵ, π, etc. won't be homogeneous for general φ^*.

Depends on the behavior of $\mathrm{aprox}(p + \epsilon \log(\lambda))$.
For instance, if $\varphi(p) = z\log(z) - z + 1$ (KL), $\varphi^*(q) = e^q - 1$ and $\mathrm{aprox}(p) = \frac{1}{1+\epsilon}p$; in this setting, Sk_ϵ is **not** 1-homogeneous anymore.

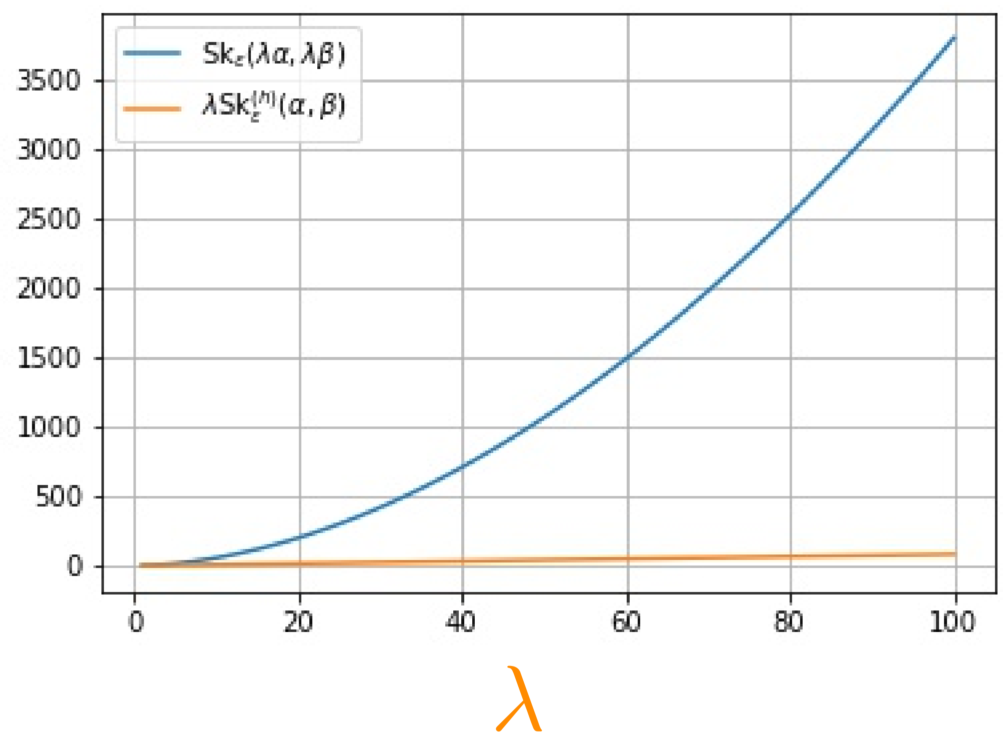
$\epsilon = 0.1$



$\epsilon = 1$



$\epsilon = 10$



Note: numerically, it seems nonetheless that when $D_\varphi = \mathrm{KL}$, then Sk_ϵ is $\underbrace{1 + \epsilon/(2+\epsilon)}_{h}$-homogeneous, i.e. $\mathrm{Sk}_\epsilon(\lambda\alpha, \lambda\beta) = \lambda^h \mathrm{Sk}_\epsilon(\alpha, \beta)$. But if for instance $D_\varphi = \mathrm{TV}$ the total variation, h-homogeneity does not hold.

A candidate homogeneous unbalanced regularized OT model

Let α, β be two (non-negative, finite) Radon measures supported on a compact domain $\Omega \subset \mathbb{R}^d$. Let $c : \Omega \times \Omega \to \mathbb{R}_+$, $\epsilon > 0$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a divergence (and $\varphi^*(q) := \sup_p \{pq - \varphi(p)\}$ be its Legendre transform). We propose to introduce the homogeneous, unbalanced, regularized Optimal Transport problem as:

$$\mathrm{OT}_\epsilon^{(h)}(\alpha, \beta) := \inf_{\pi \in \mathcal{M}_+(\Omega \times \Omega)} \langle \pi, c \rangle + D_\varphi\left(\pi_1 | \alpha \sqrt{\frac{m(\beta)}{m(\alpha)}}\right) + D_\varphi\left(\pi_2 | \beta \sqrt{\frac{m(\alpha)}{m(\beta)}}\right) + \epsilon \mathrm{KL}\left(\pi | \frac{\alpha \otimes \beta}{\sqrt{m(\alpha) m(\beta)}}\right) \tag{3}$$

$$= \sup_{f, g \in \mathcal{C}(\Omega)} \langle -\varphi^*(-f), \alpha \sqrt{\frac{m(\alpha)}{m(\beta)}} \rangle + \langle -\varphi^*(-g), \beta \sqrt{\frac{m(\alpha)}{m(\beta)}} \rangle - \epsilon \langle e^{\frac{f \oplus g - c}{\epsilon}} - 1, \frac{\alpha \otimes \beta}{\sqrt{m(\alpha) m(\beta)}} \rangle \tag{4}$$

A candidate homogeneous unbalanced regularized OT model

Let α, β be two (non-negative, finite) Radon measures supported on a compact domain $\Omega \subset \mathbb{R}^d$. Let $c : \Omega \times \Omega \to \mathbb{R}_+$, $\epsilon > 0$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a divergence (and $\varphi^*(q) := \sup_p \{pq - \varphi(p)\}$ be its Legendre transform). We propose to introduce the homogeneous, unbalanced, regularized Optimal Transport problem as:

$$\mathrm{OT}_\epsilon^{(h)}(\alpha, \beta) := \inf_{\pi \in \mathcal{M}_+(\Omega \times \Omega)} \langle \pi, c \rangle + D_\varphi\left(\pi_1 | \alpha \sqrt{\frac{m(\beta)}{m(\alpha)}}\right) + D_\varphi\left(\pi_2 | \beta \sqrt{\frac{m(\alpha)}{m(\beta)}}\right) + \epsilon \mathrm{KL}\left(\pi | \frac{\alpha \otimes \beta}{\sqrt{m(\alpha) m(\beta)}}\right) \tag{3}$$

$$= \sup_{f, g \in \mathcal{C}(\Omega)} \langle -\varphi^*(-f), \alpha \sqrt{\frac{m(\alpha)}{m(\beta)}} \rangle + \langle -\varphi^*(-g), \beta \sqrt{\frac{m(\alpha)}{m(\beta)}} \rangle - \epsilon \langle e^{\frac{f \oplus g - c}{\epsilon}} - 1, \frac{\alpha \otimes \beta}{\sqrt{m(\alpha) m(\beta)}} \rangle \tag{4}$$

Main observation: This formulation is homogeneous by construction, no matter φ or ϵ.

A candidate homogeneous unbalanced regularized OT model

Let α, β be two (non-negative, finite) Radon measures supported on a compact domain $\Omega \subset \mathbb{R}^d$. Let $c : \Omega \times \Omega \to \mathbb{R}_+$, $\epsilon > 0$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a divergence (and $\varphi^*(q) := \sup_p \{pq - \varphi(p)\}$ be its Legendre transform). We propose to introduce the homogeneous, unbalanced, regularized Optimal Transport problem as:

$$\mathrm{OT}_\epsilon^{(h)}(\alpha, \beta) := \inf_{\pi \in \mathcal{M}_+(\Omega \times \Omega)} \langle \pi, c \rangle + D_\varphi\left(\pi_1 | \alpha \sqrt{\frac{m(\beta)}{m(\alpha)}}\right) + D_\varphi\left(\pi_2 | \beta \sqrt{\frac{m(\alpha)}{m(\beta)}}\right) + \epsilon \mathrm{KL}\left(\pi | \frac{\alpha \otimes \beta}{\sqrt{m(\alpha) m(\beta)}}\right) \tag{3}$$

$$= \sup_{f, g \in \mathcal{C}(\Omega)} \langle -\varphi^*(-f), \alpha \sqrt{\frac{m(\alpha)}{m(\beta)}} \rangle + \langle -\varphi^*(-g), \beta \sqrt{\frac{m(\alpha)}{m(\beta)}} \rangle - \epsilon \langle e^{\frac{f \oplus g - c}{\epsilon}} - 1, \frac{\alpha \otimes \beta}{\sqrt{m(\alpha) m(\beta)}} \rangle \tag{4}$$

Main observation: This formulation is homogeneous by construction, no matter φ or ϵ.

Intuition: Entropic reg. : made homogeneous + reconcile D_φ and KL terms.

push π_1 toward $\alpha \sqrt{m(\beta)/m(\alpha)}$
and π_2 toward $\beta \sqrt{m(\alpha)/m(\beta)}$

push π toward $\frac{\alpha \otimes \beta}{\sqrt{m(\alpha) m(\beta)}}$ $\longrightarrow$ Homogeneous

Compatible

A candidate homogeneous unbalanced regularized OT model

Let α, β be two (non-negative, finite) Radon measures supported on a compact domain $\Omega \subset \mathbb{R}^d$. Let $c : \Omega \times \Omega \to \mathbb{R}_+$, $\epsilon > 0$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a divergence (and $\varphi^*(q) := \sup_p \{pq - \varphi(p)\}$ be its Legendre transform). We propose to introduce the homogeneous, unbalanced, regularized Optimal Transport problem as:

$$\mathrm{OT}_\epsilon^{(h)}(\alpha, \beta) := \inf_{\pi \in \mathcal{M}_+(\Omega \times \Omega)} \langle \pi, c \rangle + D_\varphi\left(\pi_1 | \alpha \sqrt{\frac{m(\beta)}{m(\alpha)}}\right) + D_\varphi\left(\pi_2 | \beta \sqrt{\frac{m(\alpha)}{m(\beta)}}\right) + \epsilon \mathrm{KL}\left(\pi | \frac{\alpha \otimes \beta}{\sqrt{m(\alpha) m(\beta)}}\right) \quad (3)$$

$$= \sup_{f, g \in \mathcal{C}(\Omega)} \langle -\varphi^*(-f), \alpha \sqrt{\frac{m(\alpha)}{m(\beta)}} \rangle + \langle -\varphi^*(-g), \beta \sqrt{\frac{m(\alpha)}{m(\beta)}} \rangle - \epsilon \langle e^{\frac{f \oplus g - c}{\epsilon}} - 1, \frac{\alpha \otimes \beta}{\sqrt{m(\alpha) m(\beta)}} \rangle \quad (4)$$

Adapted Sinkhorn algorithm: For any φ, the sequence $(f_t, g_t)_t$ produced by the fixed-point alg. used between $\lambda\alpha$ and $\lambda\beta$ is independent of $\lambda > 0$.

A candidate homogeneous unbalanced regularized OT model

Let α, β be two (non-negative, finite) Radon measures supported on a compact domain $\Omega \subset \mathbb{R}^d$. Let $c : \Omega \times \Omega \to \mathbb{R}_+$, $\epsilon > 0$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a divergence (and $\varphi^*(q) := \sup_p \{pq - \varphi(p)\}$ be its Legendre transform). We propose to introduce the homogeneous, unbalanced, regularized Optimal Transport problem as:

$$\mathrm{OT}_\epsilon^{(h)}(\alpha, \beta) := \inf_{\pi \in \mathcal{M}_+(\Omega \times \Omega)} \langle \pi, c \rangle + D_\varphi\left(\pi_1 | \alpha \sqrt{\frac{m(\beta)}{m(\alpha)}}\right) + D_\varphi\left(\pi_2 | \beta \sqrt{\frac{m(\alpha)}{m(\beta)}}\right) + \epsilon \mathrm{KL}\left(\pi | \frac{\alpha \otimes \beta}{\sqrt{m(\alpha) m(\beta)}}\right) \tag{3}$$

$$= \sup_{f, g \in \mathcal{C}(\Omega)} \langle -\varphi^*(-f), \alpha \sqrt{\frac{m(\alpha)}{m(\beta)}} \rangle + \langle -\varphi^*(-g), \beta \sqrt{\frac{m(\alpha)}{m(\beta)}} \rangle - \epsilon \langle e^{\frac{f \oplus g - c}{\epsilon}} - 1, \frac{\alpha \otimes \beta}{\sqrt{m(\alpha) m(\beta)}} \rangle \tag{4}$$

Adapted Sinkhorn algorithm: For any φ, the sequence $(f_t, g_t)_t$ produced by the fixed-point alg. used between $\lambda\alpha$ and $\lambda\beta$ is independent of $\lambda > 0$.

Convexity and reparametrization. Unfortunately, $\mathrm{OT}_\epsilon^{(h)}$ is **not** convex in α or β a priori. However, using the reparametrization $\mu := \alpha / \sqrt{m(\alpha)}$ (resp. $\nu := \beta / \sqrt{m(\beta)}$), then it is convex in μ and ν (a priori not jointly). This reparam. is somewhat natural: when $\epsilon \to \infty$, we can expect $\pi_\epsilon \to \mu \otimes \nu$ and $\epsilon \mathrm{KL}(\pi_\epsilon | \dots) \to 0$, yielding $\mathrm{OT}_\epsilon^{(h)}(\alpha, \beta) - \frac{1}{2}\mathrm{OT}_\epsilon^{(h)}(\alpha, \alpha) - \frac{1}{2}\mathrm{OT}_\epsilon^{(h)}(\beta, \beta) \to \mathrm{MMD}(\mu, \nu)$.

Conclusion

Take home messages:

- Balanced, Regularized OT yields homogeneous transport plan and Sinkhorn divergences.

- Unbalanced regularized OT may not be homogeneous.

- Can introduce an homogeneous unbalanced regularized OT model, which seemingly invites us to replace α, β by $\frac{\alpha}{\sqrt{m(\alpha)}}, \frac{\beta}{\sqrt{m(\beta)}}$. Seems to be $\geqslant 0$ (to be proved).

Thank you for your attention!