An Homogeneous Unbalanced Regularized Optimal Transport Model with Applications to Optimal Transport with Boundary

Théo Lacombe
Laboratoire d'Informatique Gaspard Monge,
Univ. Gustave Eiffel, CNRS, LIGM, F-77454 Marne-la-Vallée, France

Abstract

This work studies how the introduction of the entropic regularization term in unbalanced Optimal Transport (OT) models may alter their homogeneity with respect to the input measures. We observe that in common settings (including balanced OT and unbalanced OT with Kullback-Leibler divergence to the marginals), although the optimal transport cost itself is not homogeneous, optimal transport plans and the so-called Sinkhorn divergences are indeed homogeneous. However, homogeneity does not hold in more general Unbalanced Regularized Optimal Transport (UROT) models, for instance those using the Total Variation as divergence to the marginals. We propose to modify the entropic regularization term to retrieve an UROT model that is homogeneous while preserving most properties of the standard UROT model. We showcase the importance of using our Homogeneous UROT (HUROT) model when it comes to regularize Optimal Transport with Boundary, a transportation model involving a spatially varying divergence to the marginals for which the standard (inhomogeneous) UROT model would yield inappropriate behavior.

1 INTRODUCTION

Optimal Transport (OT) literature can be traced back to the seminal work of Monge (1784), where Monge proposes a way to interpolate between two distributions of mass, represented by two probabilities measures α, β supported on some space Ω, while minimizing a cost representing the total effort spent to move each element of mass in α to a corresponding one in β. In its modern formulation due to Kantorovich (1942), the OT problem is introduced as a linear program $\mathrm{OT}(\alpha,\beta) := \min_\pi \iint c(x,y)\mathrm{d}\pi(x,y)$ over transport plans $\pi \in \Pi(\alpha,\beta)$ that correspond to measures supported on $\Omega \times \Omega$ whose marginals are exactly α and β. Here, $c(x,y)$ denotes the cost of transporting some mass located at x to y. When $\Omega \subset \mathbb{R}^d$ is convex and $c(x,y) = \|x-y\|^p$, the infimum value reached (to the power $1/p$) defines a metric between probability measures supported on Ω called the Wasserstein distance. In addition, any optimal $\pi \in \Pi(\alpha,\beta)$ induces an interpolation between α and β by setting $\mu_t := A_t\#\pi : X \mapsto \pi(A_t^{-1}(X))$ for measurable $X \subset \Omega$, where $A_t(x,y) := (1-t)x + ty$. This curve turns out to be a geodesic between α and β for the Wasserstein distance and can also be understood as a minimal solution of the so-called continuity equation (see for instance (Villani, 2008, Thm. 7.21) and (Santambrogio, 2015, §5.4)). More generally, gradient flows induced by transportation problems are closely related to evolutionary equations (Ambrosio et al., 2005).

Naturally, this physical interpretation suggests that optimal transport models should be homogeneous with respect to the input measures α and β: loosely speaking, encoding the mass of α and β in grams or in kilograms should not change the structure of the solutions we obtain to describe the behavior of a physical system. Formally, it means that if π is an optimal transport plan between α and β, we expect $\lambda\pi$ (or, at least, some scaled version of π) to be an optimal transport plan between $\lambda\alpha$ and $\lambda\beta$, for any scaling factor $\lambda > 0$. Fortunately, this clearly holds in the standard formulation of OT (the objective function and the constraints are linear). While this formulation is restricted to measures with the same total masses (and, by homogeneity, boils down to probability measures), models of Unbalanced OT (UOT) have been proposed to handle measures with possibly different total masses by relaxing the marginal constraints (see

Proceedings of the 26[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206.

(Chizat et al., 2015; Liero et al., 2015) and Section 2). Of interest in this work and developed in Section 5 is the framework of *Optimal Transport with Boundary* (OTB) proposed by Figalli and Gigli (2010) to model heat diffusion process with Dirichlet boundary conditions. Their model enables the comparison of measures with different total masses by allowing the transportation of any amount of mass to, and from, the boundary $\partial\Omega$ of the domain Ω provided we pay the corresponding cost $c(\cdot, \partial\Omega)$. Here as well, all these models of UOT are homogeneous.

A parallel line of development—mainly popularized by the work of Cuturi (2013)—proposes to regularize the standard OT model between probability measures by adding an entropic regularization term $+\varepsilon \mathrm{KL}(\pi|\alpha \otimes \beta)$ where $\varepsilon > 0$ is a regularization parameter, and $\mathrm{KL}(\mu|\nu) = \int \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\mathrm{d}\mu$ denotes the Kullback-Leibler divergence (here, between probability measures). This approach was initially motivated by computational aspects: the resulting problem becomes strictly convex and can be solved efficiently using the Sinkhorn algorithm, a fixed-point algorithm that only involves matrix manipulations (hence usable efficiently on modern hardware as GPUs). Nonetheless, this model appears to be supported by strong theoretical properties, in particular through the introduction of an "unbiased" version called the Sinkhorn divergences (Ramdas et al., 2017; Genevay et al., 2018; Feydy et al., 2019), presented in Section 2. Unbalanced and Regularized OT have been mixed together in the works (Chizat et al., 2018; Séjourné et al., 2021) in a setting that covers most UOT models (though not directly the OTB one). However, the resulting Unbalanced Regularized OT model (UROT) may fail to be homogeneous, mostly because of the introduction of the (non-linear) term $\alpha \otimes \beta$. In particular, naive adaptations of (Séjourné et al., 2021) to introduce an entropic regularization in the OTB model will suffer with heavy inhomogeneity, hindering its use in practice and calling for the development of an entropic regularization term that would preserve homogeneity.

Outline and Contributions. This paper is organized in the following way:

- Section 2 presents the background on OT theory on which this work relies, including its regularized and unbalanced variants.
- Section 3 studies the (in)homogeneity properties of Unbalanced Regularized OT in its standard formulation. We prove in particular that in the natural settings of balanced OT and KL-penalized marginals, although the transport cost itself is not homogeneous, the corresponding Sinkhorn divergence appears to be homogeneous thanks to the addition of a "mass bias" proposed by Séjourné et al. It gives a new perspective in favor of the use of this "unbiased" formulation of entropic OT in these contexts. We show that, however, in a more general setting homogeneity does not hold in the standard UROT model.
- Section 4 introduces a model of *Homogeneous* Unbalanced Regularized OT (HUROT). This model enjoys most of the properties of the standard one (UROT): it is solved by applying the Sinkhorn algorithm to renormalized measures—hence can be implemented faithfully based on existing code, is continuous with respect to the weak convergence of measures, and the corresponding Sinkhorn divergence is positive without the need to introduce a mass bias term.
- Eventually, Section 5 introduces a model of Regularized OT with Boundary (ROTB). We showcase the importance of enforcing homogeneity in this model using the approach developed in Section 4. We prove that the resulting ROTB model, in addition to the properties it shares with the HUROT model (continuity, positivity of the Sinkhorn divergence, etc.), implies the same notion of convergence as its unregularized counterpart, which legitimates our approach as a consistent way to regularize this spatially varying UOT model.

Our implementation is available at `https://github.com/tlacombe/homogeneousUROT`. All proofs and some complementary remarks have been deferred to the appendix.

2 BACKGROUND

Preliminary Definitions and Notation. In this work, Ω denotes a compact subset of $\mathbb{R}^d$, $c : \Omega \times \Omega \to \mathbb{R}_+$ is a cost function that is assumed to satisfy $c(x,x) = 0$, to be symmetric, and Lipschitz continuous on Ω, typically $c(x,y) = \|x-y\|^2$. The set $\mathcal{M}(\Omega)$ denotes the space of (non-negative) Radon measures supported on Ω, and $\mathcal{P}(\Omega) = \{\alpha \in \mathcal{M}(\Omega),\ m(\alpha) = 1\}$ denotes the subset of probability measures, that is measures of total mass $m(\alpha) := \alpha(\Omega) = 1$. With the exception of Section 5, we also assume that the total masses of the measures are finite.

Given a measure $\alpha \in \mathcal{M}(\Omega)$ and a function $f \in \mathcal{C}(\Omega)$, we use the notation $\langle \cdot, \cdot \rangle$ to denote the duality product, that is $\langle f, \alpha \rangle := \int_\Omega f(x)\mathrm{d}\alpha(x)$. We say that a sequence of measures $(\alpha_n)_n \in \mathcal{M}(\Omega)^{\mathbb{N}}$ converges *weakly* toward some $\alpha_\infty \in \mathcal{M}(\Omega)$, denoted by $\alpha_n \xrightarrow{w} \alpha_\infty$, if for any continuous (bounded) map f one has $\langle f, \alpha_n \rangle \to \langle f, \alpha_\infty \rangle$. Note that this implies $m(\alpha_n) \to m(\alpha_\infty)$.

A function $\varphi : [0, +\infty) \to [0, +\infty]$ is said to be an *entropy function* if it is convex, lower-semi-continuous and satisfies $\varphi(1) = 0$. We also set the convention $\varphi(p) = +\infty$ whenever $p < 0$ and, in this work, we will only consider entropy functions that satisfy $\varphi(0) < \infty$. Of interest is its *Legendre transform*, defined by $\varphi^* : q \mapsto \sup_{p \geq 0} pq - \varphi(p)$. An entropy function φ induces a φ-*divergence*:

$$D_\varphi(\alpha|\beta) := \langle \varphi \circ \frac{\mathrm{d}\alpha}{\mathrm{d}\beta}, \beta \rangle = \int_\Omega \varphi \left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}(x) \right) \mathrm{d}\beta(x). \quad (1)$$

Among the notorious choices to define a φ-divergence, one has $\varphi(p) = p \log(p) - p + 1$, whose Legendre transform is $\varphi^*(q) = e^q - 1$, and which defines the so-called *Kullback-Leibler* divergence $D_\varphi = \mathrm{KL}$. As another example that will play an important role in this work, the *Total Variation* between measures can also be retrieved as a φ-divergence by taking $\varphi(p) = |1 - p|$, yielding $D_\varphi(\alpha|\beta) = \int_\Omega |\mathrm{d}\alpha(x) - \mathrm{d}\beta(x)| =: \mathrm{TV}(\alpha - \beta)$. Finally, the *convex indicator function* is defined by $\imath_c(p) = 0$ if $p = 1$, and $+\infty$ otherwise, so that $D_{\imath_c}(\alpha|\beta) = 0$ if $\alpha = \beta$, and $+\infty$ otherwise. Note that $\imath_c^* = \mathrm{id}$, the identity map.

Finally, a function $F : \mathcal{X} \to \mathcal{Y}$ (for some Banach spaces $\mathcal{X}, \mathcal{Y}$) is said to be h-homogeneous if there exists a constant $h > 0$ such that for any $(\lambda, x) \in \mathbb{R} \times \mathcal{X}$ we have $F(\lambda x) = \lambda^h F(x)$. When $h = 1$, we will simply say that F is homogeneous.

Balanced Regularized Optimal Transport. Let $\alpha, \beta \in \mathcal{P}(\Omega)$ denote two probability measures. We denote by $\Pi(\alpha, \beta) := \{\pi \in \mathcal{M}(\Omega \times \Omega),\ \pi(\cdot, \Omega) = \alpha, \pi(\Omega, \cdot) = \beta\}$ the corresponding *set of transport plans between α and β*, that is the measures π supported on $\Omega \times \Omega$ whose marginals π_1, π_2 are equal to α, β, respectively. The optimal transport cost between α and β is defined as

$$\mathrm{OT}(\alpha, \beta) := \inf_{\pi \in \Pi(\alpha, \beta)} \langle \pi, c \rangle, \quad (2)$$

and any minimizer of this problem is said to be an optimal transport plan between the two measures.

In 2013, Cuturi significantly contributed to popularize the practical use of OT (in particular in the machine learning community) by observing that its entropic regularized version can be solved efficiently on modern hardware (Cuturi, 2013), see (Peyré et al., 2019) for an extensive overview of the computational aspects of OT. In its modern form, this regularized problem reads, for a parameter $\varepsilon > 0$,

$$\mathrm{OT}_\varepsilon(\alpha, \beta) := \inf_{\pi \in \Pi(\alpha, \beta)} \langle \pi, c \rangle + \varepsilon \mathrm{KL}(\pi | \alpha \otimes \beta), \quad (3)$$

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

where (3) is referred to as the primal problem and (4) as its dual. It is worth noting that despite its appealing computational properties, OT_ε does not define a proper divergence between probability measures. In particular, $\mathrm{OT}_\varepsilon(\alpha, \alpha) \neq 0$ in general, and probably worse, the map $\alpha \mapsto \mathrm{OT}_\varepsilon(\alpha, \beta)$ is **not** minimized for $\alpha = \beta$. This *entropic bias* (Janati et al., 2020), can be corrected by introducing the associated *Sinkhorn divergence* (Ramdas et al., 2017; Genevay et al., 2018), defined by

$$\mathrm{Sk}_\varepsilon(\alpha, \beta) := \mathrm{OT}_\varepsilon(\alpha, \beta) - \frac{1}{2}\mathrm{OT}_\varepsilon(\alpha, \alpha) - \frac{1}{2}\mathrm{OT}_\varepsilon(\beta, \beta). \quad (5)$$

Deeply studied in (Feydy et al., 2019), it can be proved that $\mathrm{Sk}_\varepsilon(\alpha, \beta) \geq 0$, with equality if, and only if, $\alpha = \beta$.

Unbalanced Sinkhorn Divergences. The above formulations are restricted to measures α, β with the same total masses $m(\alpha) = m(\beta)$. This setting is referred to as *balanced* OT. One celebrated way to extend (2) to measures of different total masses is to relax the marginal constraints using a φ-divergence. The *unbalanced* OT (UOT) problem reads, for a given entropy function φ:

$$\mathrm{OT}_\varphi(\alpha, \beta) = \inf_{\pi \in \mathcal{M}(\Omega \times \Omega)} \langle c, \pi \rangle + D_\varphi(\pi_1 | \alpha) + D_\varphi(\pi_2 | \beta). \quad (6)$$

Following (Chizat et al., 2018), unbalanced and regularized OT can be mixed together yielding the following problems, dual of each other:

$$\begin{aligned} \mathrm{OT}_{\varepsilon, \varphi}(\alpha, \beta) := \inf_{\pi \in \mathcal{M}(\Omega \times \Omega)} & \langle \pi, c \rangle + D_\varphi(\pi_1 | \alpha) + D_\varphi(\pi_2 | \beta) \\ & + \varepsilon \mathrm{KL}(\pi | \alpha \otimes \beta) & (7) \\ = \sup_{f, g \in \mathcal{C}(\Omega)} & \langle -\varphi^*(-f), \alpha \rangle + \langle -\varphi^*(-g), \beta \rangle \\ & - \varepsilon \langle e^{\frac{f \oplus g - c}{\varepsilon}} - 1, \alpha \otimes \beta \rangle & (8) \end{aligned}$$

In the following, we will refer to this formulation as the *standard Unbalanced Regularized OT (UROT) model*. Note that setting $\varphi = \imath_c$ retrieves (3) (balanced regularized OT) and setting $\varepsilon = 0$ retrieves (6) (unbalanced OT). This model has been deeply studied in (Séjourné et al., 2021). In particular, authors prove that the dual problem (8) can be solved by iterating an adapted version of the Sinkhorn algorithm that reads (Séjourné et al., 2021, Def. 3) which consists of building a sequence $(f_t, g_t)_t$ defined by

$$\begin{aligned} f_{t+1}(x) &= -\mathrm{aprox}_{\varepsilon, \varphi^*} \left(\varepsilon \log \langle e^{\frac{g_t - c(x, \cdot)}{\varepsilon}}, \beta \rangle \right), \\ g_{t+1}(y) &= -\mathrm{aprox}_{\varepsilon, \varphi^*} \left(\varepsilon \log \langle e^{\frac{f_{t+1} - c(\cdot, y)}{\varepsilon}}, \alpha \rangle \right), \end{aligned} \quad (9)$$

where $\mathrm{aprox}_{\varepsilon, \varphi^*}$ is the *anisotropic proximity operator* (Séjourné et al., 2021, Def. 2) associated to (the

Legendre transform of) the divergence φ defined by $\mathrm{aprox}_{\varepsilon,\varphi^*}(p) := \arg\min_{q\in\mathbb{R}} \varepsilon e^{\frac{p-q}{\varepsilon}} + \varphi^*(q)$. Crucially, a couple $(f, g) \in \mathcal{C}(\Omega)$ is optimal for (8) if, and only if, it is a fixed point of the map $(f_t, g_t) \mapsto (f_{t+1}, g_{t+1})$ (Séjourné et al., 2021, Prop. 8). Furthermore, any sequence $(f_t, g_t)_t$ built following (9) is guaranteed to converge towards such a fixed point (that is, an optimal pair of potentials) under mild assumptions (Séjourné et al., 2021, Thm. 1) that are satisfied in this work (namely, c must be Lipschitz continuous on Ω, and one must be able to restrict (8) to a compact subset of $\mathcal{C}(\Omega)$, which is possible in our settings of interest: $D_\varphi = D_{\imath_c}$, TV or KL, see (Séjourné et al., 2021, Lemmas 8, 9)). Eventually, if (f, g) is optimal for the dual problem (8), then

$$\pi := \exp\left(\frac{f \oplus g - c}{\varepsilon}\right) \alpha \otimes \beta \tag{10}$$

is optimal for the primal problem (7). Finally, the authors introduce the unbalanced Sinkhorn divergence between α and β:

$$\begin{aligned} \mathrm{Sk}_{\varepsilon,\varphi}(\alpha, \beta) :=& \mathrm{OT}_{\varepsilon,\varphi}(\alpha, \beta) + \frac{\varepsilon}{2}(m(\alpha) - m(\beta))^2 \\ &- \frac{1}{2}\mathrm{OT}_{\varepsilon,\varphi}(\alpha, \alpha) - \frac{1}{2}\mathrm{OT}_{\varepsilon,\varphi}(\beta, \beta). \end{aligned} \tag{11}$$

They prove that this formulation enjoys most of the properties of its balanced counterpart (5): it is continuous with respect to the weak convergence, non-negative, satisfies $\mathrm{Sk}_{\varepsilon,\varphi}(\alpha, \beta) = 0 \Leftrightarrow \alpha = \beta$, is convex with respect to each of its entries, and induces the same topology as weak convergence on the set $\mathcal{M}_{\leq m}(\Omega)$ of Radon measures with total mass uniformly bounded by $m > 0$, that is $\mathrm{Sk}_{\varepsilon,\varphi}(\alpha_n, \alpha) \to 0 \Leftrightarrow \alpha_n \xrightarrow{w} \alpha$. Note the presence of the term $+\frac{\varepsilon}{2}(m(\alpha) - m(\beta))^2$, called the *mass bias*, that is required to make the above assertions on $\mathrm{Sk}_{\varepsilon,\varphi}$ correct.

3 (IN)HOMOGENEITY IN THE STANDARD MODEL

In this section, we study the homogeneity properties of the standard model (7) with respect to the couple of input measures (α, β). First, let us stress that non-regularized OT, should it be balanced (2) or not (6), is homogeneous in (α, β), that is $\mathrm{OT}_{\varepsilon=0,\varphi}(\lambda\alpha, \lambda\beta) = \lambda \cdot \mathrm{OT}_{\varepsilon=0,\varphi}(\alpha, \beta)$ for any $\lambda \geq 0$. Furthermore, if π is an optimal transport plan between α and β, then $\lambda\pi$ is an optimal transport plan between $\lambda\alpha$ and $\lambda\beta$. As mentioned in the introduction, this behavior is desirable as an optimal transport plan may be used as a way to interpolate between α and β, and it would be surprising that a change of scale in the masses of the measures induces a structural change in the interpolation between the two measures. However, the addition of the entropic regularization term which, in the dual (8), reads $-\varepsilon \langle e^{\frac{f\oplus g-c}{\varepsilon}} - 1, \alpha \otimes \beta\rangle$ induces a seemingly peculiar behavior in terms of homogeneity. Namely, if we let

$$\begin{aligned} J_{(\alpha,\beta)}(f, g) :=& \langle -\varphi^*(-f), \alpha\rangle + \langle -\varphi^*(-g), \beta\rangle \\ &- \varepsilon \langle e^{\frac{f\oplus g-c}{\varepsilon}} - 1, \alpha \otimes \beta\rangle, \end{aligned} \tag{12}$$

one has

$$\begin{aligned} J_{(\lambda\alpha,\lambda\beta)}(f, g) =& \lambda \langle -\varphi^*(-f), \alpha\rangle + \lambda \langle -\varphi^*(-g), \beta\rangle \\ &- \lambda^2\varepsilon \langle e^{\frac{f\oplus g-c}{\varepsilon}} - 1, \alpha \otimes \beta\rangle, \end{aligned}$$

inducing a quadratic term in λ that may hinder homogeneity.

The Balanced Case. We first consider the case of regularized balanced optimal transport (3); where $\varphi = \imath_c$. The following lemma describes the effect of a scaling of the measures on the sequence of potentials produced by the Sinkhorn algorithm (9).

Lemma 3.1. *Let $\alpha, \beta \in \mathcal{M}(\Omega)$ be two measures of total mass $m(\alpha) = m(\beta) = m$. Fix $(f_0, g_0) \in \mathcal{C}(\Omega)$ and let $(f_t, g_t)_{t\geq 1}$ denote the sequence of dual potentials produced iterating (9) starting from (f_0, g_0) for the couple (α, β). Let $(f_t^{(\lambda)}, g_t^{(\lambda)})_t$ denote the sequence produced starting from (f_0, g_0) for the couple $(\lambda\alpha, \lambda\beta)$. Then, for all $t \geq 1$, $(f_t^{(\lambda)}, g_t^{(\lambda)}) = (f_t - \varepsilon \log(\lambda), g_t)$.*

Hence, scaling the measures by a factor λ reflects as a shift of $-\varepsilon \log(\lambda)$ in the first potential of the sequence produced by the Sinkhorn algorithm, yielding a series of results summarized in the following corollary.

Corollary 3.2. *Let $\alpha, \beta \in \mathcal{M}(\Omega)$ be two measures of total mass $m(\alpha) = m(\beta) = m$.*

1. *If (f, g) is a couple of optimal potentials for the dual problem for the couple (α, β), then $(f - \varepsilon \log(\lambda), g)$ is optimal for $(\lambda\alpha, \lambda\beta)$.*
2. *If π is an optimal transport plan for (α, β), then $\lambda\pi$ is optimal for $(\lambda\alpha, \lambda\beta)$.*
3. *We have $\mathrm{OT}_\varepsilon(\lambda\alpha, \lambda\beta) = \lambda \cdot \mathrm{OT}_\varepsilon(\alpha, \beta) + \varepsilon\lambda(\lambda - 1)m^2 - \varepsilon \log(\lambda)\lambda m$, that is, the optimal transport cost is not homogeneous.*
4. *We have $\mathrm{Sk}_\varepsilon(\lambda\alpha, \lambda\beta) = \lambda \cdot \mathrm{Sk}_\varepsilon(\alpha, \beta)$.*

Overall, the quantities of interest behave in a reasonable way, in particular the solutions of the primal problem are homogeneous. Interestingly, the Sinkhorn divergence cancels the inhomogeneous behavior appearing in OT_ε, giving an additional argument in favor of using this debiased (and homogenized!) quantity to compare probability measures using regularized OT.

Remark 3.3. *We warn the reader interested in computational OT that the inhomogeneity appearing in* OT_ε *may lead to ill-behavior in numerical applications. Indeed, in practice, the Sinkhorn algorithm (9) does not exactly reach a fixed point and is instead run until some stopping criterion is reached. For instance, one may stop the iterations when the relative change in the objective value* $v_t := J_{(\alpha,\beta)}(f_t, g_t)$ *is lesser than some* $\tau > 0$*, that is when* $\left|\frac{v_{t+1}-v_t}{v_t}\right| < \tau$*. However, the inhomogeneous behavior in* v_t *implies that for a given* τ*, the number of iterations needed to reach the criterion when comparing* α *and* β *may differ from the one needed when comparing* $\lambda\alpha$ *and* $\lambda\beta$*. Thus, even though in theory the (optimal) transportation plans of both couples should be the same (up to the scaling factor* λ*), the numerical outputs (transport plan, Sinkhorn divergence, etc.) provided by the Sinkhorn algorithm may not satisfy this property. An illustration of this phenomenon using* `PythonOptimalTransport` *is provided in the appendix.*

The KL Case. We now propose to derive the same study using $\varphi(p) = p\log(p) - p + 1$, that is $D_\varphi = \mathrm{KL}$, a common choice in unbalanced optimal transport to penalize the marginal errors. In this context, $\varphi^*(q) = e^q - 1$, and $\mathrm{aprox}_{\varepsilon,\varphi^*}(p) = \frac{1}{1+\varepsilon}p$. The following proposition summarizes the important properties of this model as far as homogeneity is concerned.

Proposition 3.4. *Let* $\alpha, \beta \in \mathcal{M}(\Omega)$*. Then,*

1. *If* (f, g) *is a pair of optimal potentials for the couple* (α, β)*, then* $\left(f - \frac{\varepsilon^2}{(1+\varepsilon)^2-1}\log(\lambda), g - \frac{\varepsilon^2}{(1+\varepsilon)^2-1}\log(\lambda)\right)$ *is optimal for the couple* $(\lambda\alpha, \lambda\beta)$*.*
2. *If* π *is optimal for* (α, β)*, then* $\lambda^h\pi$ *is optimal for* $(\lambda\alpha, \lambda\beta)$*, where* $h = 2 - \frac{2}{2+\varepsilon}$*.*
3. *The Sinkhorn divergence* $\mathrm{Sk}_{\varepsilon,\varphi}$ *is* h*-homogeneous (while* $\mathrm{OT}_{\varepsilon,\varphi}$ *is not* h*-homogeneous).*

As in the balanced case, the conclusions here are mostly positive: though the optimization problem (8) itself is not (h-)homogeneous, the optimal transport plans are h-homogeneous, and so is the Sinkhorn divergence (thanks to the addition of the mass bias term!).

Remark 3.5. *Proposition 3.4 can be slightly generalized: whenever the anisotropic proximal operator is linear—*$\mathrm{aprox}_{\varepsilon,\varphi^*}(p) = \kappa p$ *for some* $\kappa \in (0, 1]$*—, the optimal transport plans and the Sinkhorn divergence are* $h = \frac{2}{1+\kappa}$*-homogeneous. Note that this leads to* $\varphi^*(q) = \frac{\varepsilon}{\left(\frac{1}{\kappa}-1\right)}\left(e^{\frac{q}{\varepsilon}\left(\frac{1}{\kappa}-1\right)} - 1\right)$*, that is equivalent to use* $\rho\mathrm{KL}$ *as the marginal penalty with* $\rho = \frac{\varepsilon}{\left(\frac{1}{\kappa}-1\right)}$*. We believe that this condition may be necessary as well: a non-linearity in the* aprox *operator prevents* h*-homogeneity to occur and the family of divergences* $(\rho\mathrm{KL})_{\rho\in[0,+\infty]}$ *is the only one that makes the UROT problem homogeneous. Elements of proof in the case* $h = 1$ *are provided in the appendix (i.e. we conclude that* $\mathrm{aprox}_{\varepsilon,\varphi^*} = \mathrm{id}$*). It means that* balanced *regularized OT is the only instance that provides a (1-)homogeneous formulation when using* $\varepsilon\mathrm{KL}(\pi|\alpha\otimes\beta)$ *as the entropic regularization term. In Section 4, we show how to correct the entropic regularization term in unbalanced cases to retrieve a 1-homogeneous formulation, without sacrificing the good properties of* $\mathrm{Sk}_{\varepsilon,\varphi}$*.*

Inhomogeneity in General: the TV Case. The two previous case studies, which fall in the setting "aprox is linear", may suggest that the apparent inhomogeneity in the formulation of the unbalanced regularized OT problem does not have much aftermaths: the structure of optimal transport plans is preserved and the Sinkhorn divergence is h-homogeneous. As this encompasses both balanced regularized OT (3) and UOT with KL-relaxation of the marginal constraints—arguably covering most applications of UROT in practice—this may explain why behaviors related to (in)homogeneity did not receive much attention in the OT community so far.

We now give an example for which inhomogeneity (in particular, of the optimal transport plan) occurs: the case of Total Variation (TV). This choice of marginal divergence encodes the *Optimal Partial Transport* model proposed by Figalli (2010) and, roughly, induces a distance threshold above which one prefers to destroy/create mass rather than transporting it and has found different applications in machine learning (Chapel et al., 2020; Mukherjee et al., 2021; Fatras et al., 2021). Formally, this setting corresponds to taking $\varphi(p) = |1-p|$, yielding $D_\varphi(\pi_1|\alpha) = \mathrm{TV}(\pi_1 - \alpha)$, $\varphi^*(q) = \max(-1, q)$ and $\mathrm{aprox}_{\varepsilon,\varphi^*}(p) = \max(-1, \min(p, 1))$.

The Sinkhorn updates used to produce a sequence $(f_t^{(\lambda)}, g_t^{(\lambda)})_t$ for the couple $(\lambda\alpha, \lambda\beta)$ read

$$f_{t+1}^{(\lambda)} = \min\left(\max\left(-1, -\varepsilon\log\langle e^{\frac{g_t^{(\lambda)}-c}{\varepsilon}}, \beta\rangle - \varepsilon\log(\lambda)\right), 1\right),$$
$$g_{t+1}^{(\lambda)} = \min\left(\max\left(-1, -\varepsilon\log\langle e^{\frac{f_{t+1}^{(\lambda)}-c}{\varepsilon}}, \alpha\rangle - \varepsilon\log(\lambda)\right), 1\right). \tag{13}$$

Here, $\mathrm{aprox}_{\varepsilon,\varphi^*}$ exhibits sharp changes of behavior when its argument get higher than 1 (or lower than -1). This is the source of an inhomogeneous behavior: when the scaling factor $\lambda \to \infty$ (or $\to 0$), this affects the Sinkhorn updates hence the returned (optimal) potentials, transport plan, and Sinkhorn divergence.

Numerical Illustration. To empirically illustrate

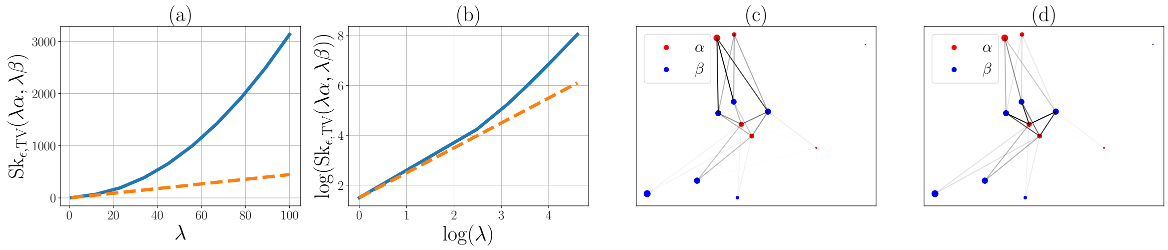

Figure 1: **Inhomogeneity when $D_\varphi = \mathrm{TV}$.** *(a)* The Sinkhorn divergence between $\lambda\alpha$ and $\lambda\beta$ for $\lambda \in [1, 100]$ for the standard UROT model. Dashed line correspond to the homogeneous behavior $\lambda \cdot \mathrm{Sk}_{\varepsilon,\mathrm{TV}}(\alpha, \beta)$. *(b)* The same curve in log-log scale. *(c,d)* The optimal transport plans for $\lambda = 1$ and $\lambda = 100$, respectively. Width of the lines linking x in α to y in β are proportional to $\mathrm{d}\pi(x, y)$. The transport plans are not proportional to each other, showcasing the structural change in the interpolation when rescaling the measures.

the possible inhomogeneous behavior of $\mathrm{Sk}_{\varepsilon,\mathrm{TV}}$, we propose the following experiment. We randomly sample two measures α, β with $n = 5$ and $m = 7$ points, respectively and random (non-negative) weights on their support distributed uniformly between 0 and 1. We then compute the Sinkhorn divergence $\mathrm{Sk}_{\varepsilon,\mathrm{TV}}(\lambda\alpha, \lambda\beta)$ for $\lambda \in [1, 100]$ from the optimal dual potentials obtained by iterating (13) and the corresponding transport plans through the relation (10). Figure 1 showcases the dependence of the result on λ. The plot (a) shows that $\mathrm{Sk}_{\varepsilon,\mathrm{TV}}$ cannot be 1-homogeneous. If $\mathrm{Sk}_{\varepsilon,\mathrm{TV}}$ was h-homogeneous for some h, one would expect that $\log(\mathrm{Sk}_{\varepsilon,\mathrm{TV}}(\lambda\alpha, \lambda\beta)) = h\log(\lambda) + \log(\mathrm{Sk}_{\varepsilon,\mathrm{TV}}(\alpha, \beta))$, that would yield a line of slope h in log-log scale. Plot (b) in Figure 1 shows that this does not hold overall: a slope break occurs around $\log(\lambda) \sim 2.5$, as a consequence of the non-linearity in $\mathrm{aprox}_{\varepsilon,\mathrm{TV}}$. This reflects in structural changes in the resulting transport plans as illustrated in the subplots (c,d). Computations are run with $\varepsilon = 1$.

4 HOMOGENEOUS UNBALANCED REGULARIZED OPTIMAL TRANSPORT (HUROT)

In this section, by slightly changing the entropic regularization term appearing in (7), we introduce a model of unbalanced regularized OT that presents the advantage of being homogeneous in a broad setting. Let fix $\alpha, \beta \in \mathcal{M}(\Omega)$ and assume for now that they have positive total masses: $m(\alpha) > 0$, $m(\beta) > 0$, that is belong to $\mathcal{M}(\Omega)\backslash\{0\}$. Let also $m_a(\alpha, \beta) := \frac{1}{2}(m(\alpha) + m(\beta))$, $m_g(\alpha, \beta) := \sqrt{m(\alpha)m(\beta)}$ and $m_h(\alpha, \beta) := 2\left(\frac{1}{m(\alpha)} + \frac{1}{m(\beta)}\right)^{-1}$ denote the arithmetic, geometric and harmonic mean of $m(\alpha)$ and $m(\beta)$, respectively. When it is clear from the context, we will simply write m_a, m_g and m_h instead.

Definition 4.1. *For $\pi \in \mathcal{M}(\Omega \times \Omega)$ and $\alpha, \beta \in \mathcal{M}(\Omega)\backslash\{0\}$, introduce*

$$R(\pi|\alpha, \beta) := \frac{1}{2}\left(\mathrm{KL}\left(\pi|\frac{\alpha}{m(\alpha)} \otimes \beta\right) + \mathrm{KL}\left(\pi|\alpha \otimes \frac{\beta}{m(\beta)}\right)\right). \tag{14}$$

The homogeneous unbalanced regularized optimal transport (HUROT) problem is defined as

$$\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha, \beta) := \inf_{\pi} \langle c, \pi\rangle + D_\varphi(\pi_1|\alpha) + D_\varphi(\pi_2|\beta) + \varepsilon R(\pi|\alpha, \beta). \tag{15}$$

The regularization term (14) can be seen as the average of two entropic regularization terms whose reference measure have β as second marginal and α as first marginal, respectively.

Proposition 4.2 (Dual formulation)**.** *One has:*

$$\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha, \beta) = \sup_{f,g \in \mathcal{C}(\Omega)} \langle -\varphi^*(-f), \alpha\rangle + \langle -\varphi^*(-g), \beta\rangle - \varepsilon\left\langle \frac{e^{\frac{f \oplus g - c}{\varepsilon}}}{m_g} - \frac{1}{m_h}, \alpha \otimes \beta\right\rangle. \tag{16}$$

Furthermore, if f, g is optimal for (16), then

$$\pi := \exp\left(\frac{f \oplus g - c}{\varepsilon}\right)\frac{\alpha \otimes \beta}{m_g(\alpha, \beta)} \tag{17}$$

is an optimal transport plan for the problem (15).

Using first order conditions on the dual, we can derive the Homogeneous Sinkhorn algorithm:

$$\begin{aligned} f_{t+1} &= -\mathrm{aprox}_{\varepsilon,\varphi^*}\left(\varepsilon \log\langle e^{\frac{g_t - c}{\varepsilon}}, \frac{\alpha}{m_g}\rangle\right), \\ g_{t+1} &= -\mathrm{aprox}_{\varepsilon,\varphi^*}\left(\varepsilon \log\langle e^{\frac{f_{t+1} - c}{\varepsilon}}, \frac{\beta}{m_g}\rangle\right). \end{aligned} \tag{18}$$

This iterative algorithm can be seen as the standard Sinkhorn algorithm (9) applied to the renormalized measures $\left(\frac{\alpha}{m_g(\alpha,\beta)}, \frac{\beta}{m_g(\alpha,\beta)}, \right)$ and benefits from all the properties proved in (Séjourné et al., 2021). In particular, it converges toward a fixed point (f, g) that is an optimal couple of potentials for the HUROT model. Numerically, optimal potentials can thus be directly obtained using dedicated software such as POT (Flamary et al., 2021) without requiring further development, and can be re-injected in the objective function $J^{[H]}_{(\alpha,\beta)}$ to get the corresponding homogeneous transport cost $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha,\beta)$. Thus, computing HUROT is not harder than solving the usual UROT problem.

We now state the homogeneity of the HUROT model.

Proposition 4.3. $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}$ *is* 1*-homogeneous. Furthermore, if* (f,g) *is a pair of optimal dual potentials for the couple* (α,β)*, then it is also optimal for the couple* $(\lambda\alpha,\lambda\beta)$.

Corollary 4.4. *If* π *is an optimal transport plan for the HUROT model* (15) *for the couple of measures* (α,β)*, then* $\lambda\pi$ *is optimal for the couple* $(\lambda\alpha,\lambda\beta)$.

Proposition 4.5. *Let* $(f_0,g_0)\in\mathcal{C}(\Omega)$*,* α,β *be two non-zero measures, and* $\lambda>0$*. The sequence* $(f^{(\lambda)}_t,g^{(\lambda)}_t)_t$ *produced by* (18) *for* $(\lambda\alpha,\lambda\beta)$ *initialized at* (f_0,g_0) *is independent of* λ.

As the standard UROT model, HUROT is also continuous with respect to the weak convergence.

Proposition 4.6 (Continuity of the HUROT model)**.** *Let* $\alpha,\beta\in\mathcal{M}(\Omega)\backslash\{0\}$*. Consider two sequences* $(\alpha_n)_n,(\beta_n)_n$ *in* $\mathcal{M}(\Omega)\backslash\{0\}$ *that weakly converge toward* α *and* β*, respectively. Then* $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha_n,\beta_n)\to\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha,\beta)$.

We can now introduce the corresponding notion of (homogeneous) Sinkhorn divergence.

Definition 4.7. *Let* $\alpha,\beta\in\mathcal{M}(\Omega)$ *with* $m(\alpha),m(\beta)>0$*. The homogeneous Sinkhorn divergence between* α *and* β *is defined as*

$$\begin{aligned}\mathrm{Sk}^{[H]}_{\varepsilon,\varphi}(\alpha,\beta):=&\mathrm{OT}^{[H]}_{\varepsilon,\varphi^*}(\alpha,\beta)\\&-\frac{1}{2}\mathrm{OT}^{[H]}_{\varepsilon,\varphi^*}(\alpha,\alpha)-\frac{1}{2}\mathrm{OT}^{[H]}_{\varepsilon,\varphi^*}(\beta,\beta).\end{aligned}\tag{19}$$

By construction, $\mathrm{Sk}^{[H]}_{\varepsilon,\varphi^*}$ is homogeneous. Interestingly, it is also non-negative under standard assumptions, without needing a "mass bias" term.

Proposition 4.8. *Let* $K_\varepsilon(x,y)=e^{-\frac{c(x,y)}{\varepsilon}}$*, and assume that* K_ε *is a positive definite kernel. Then,* $\mathrm{Sk}^{[H]}_{\varepsilon,\varphi^*}(\alpha,\beta)\geq 0$*, with equality if, and only if,* $\alpha=\beta$.

Continuity around the Null Measure. Previously in this section, we only considered the HUROT model whenever $\alpha,\beta\neq 0$. As in the standard case (Séjourné et al., 2021, §4.6), assessing continuity of our model around the null measure requires specific care. Recall that we assume $\varphi(0)<\infty$.

Proposition 4.9. *Let* $\beta\in\mathcal{M}(\Omega)\backslash\{0\}$*. Define* $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(0,\beta):=\left(\varphi(0)+\frac{\varepsilon}{2}\right)m(\beta)$*. Let* $(\alpha_n)_n$ *be a sequence of non-null measures that weakly converges toward the null measure:* $\alpha_n\xrightarrow{w}0$*. Then* $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha_n,\beta)\to\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(0,\beta)$*. Furthermore, if we set* $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(0,0):=0$*, then for any sequences* $(\alpha_n)_n,(\beta_n)_n$ *that both weakly converge toward the null measure, one has* $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha_n,\beta_n)\to 0$.

5 APPLICATION TO OPTIMAL TRANSPORT WITH BOUNDARY

Definition and Motivation. Optimal Transport with Boundary (OTB) was introduced by Figalli and Gigli (2010) as a way to model heat diffusion equations with specific boundary conditions. We first give a brief introduction to this model as introduced by the authors in their seminal paper.

We consider an open bounded domain $\Omega\subset\mathbb{R}^d$. Let $\overline{\Omega}$ be its closure and $\partial\Omega$ denote its boundary. For the sake of simplicity, we assume that the cost function $c:\overline{\Omega}\times\overline{\Omega}\to\mathbb{R}_+$ is given by $c(x,y)=\|x-y\|^2$, though most of the approach developed in the following would adapt to more general symmetric Lipschitz continuous cost functions. To alleviate notations, we introduce $c_{\partial\Omega}(x):=c(x,\partial\Omega)=c(\partial\Omega,x)=\inf_{y\in\partial\Omega}c(x,y)$. We also assume that $\partial\Omega$ is regular enough so that there exist a measurable map $P:\Omega\to\partial\Omega$ such that $c(x,P(x))=c_{\partial\Omega}(x)$. Now, let α,β be two locally finite Radon measures supported on Ω. Formally, we introduce the set of *admissible plans*

$$\begin{aligned}\mathrm{Adm}(\alpha,\beta):=\big\{&\pi\in\mathcal{M}(\overline{\Omega}\times\overline{\Omega}),\\&\forall A,B\subset\Omega,\pi(A\times\overline{\Omega})=\alpha(A),\\&\pi(\overline{\Omega}\times B)=\beta(B)\big\}.\end{aligned}\tag{20}$$

Now, consider the following optimization problem:

$$\mathrm{FG}(\alpha,\beta)=\inf_{\pi\in\mathrm{Adm}(\alpha,\beta)}\iint_{\overline{\Omega}\times\overline{\Omega}}c(x,y)\mathrm{d}\pi(x,y).\tag{21}$$

To guarantee that $\mathrm{FG}(\alpha,\beta)<+\infty$, we restrict to measures α,β that have finite *total persistence*, where the total persistence of a measure $\mu\in\mathcal{M}(\Omega)$ is defined as $\mathrm{Pers}(\mu):=\mathrm{FG}(\mu,0)=\int_\Omega c_{\partial\Omega}(x)\mathrm{d}\mu(x)$. We will note by $\mathcal{M}^c(\Omega)$ the set of such measures.

The first step to propose a relevant entropic regularization is to rephrase it in a formalism much closer to the standard UROT model. For $\mu\in\mathcal{M}^c(\Omega)$, define the renormalized measure $\hat{\mu}$ by $\forall A\subset\Omega$ Borel, $\hat{\mu}(A):=\int_A c_{\partial\Omega}(x)\mathrm{d}\mu(x)$. Note in particular the relation $m(\hat{\mu})=\mathrm{Pers}(\mu)<\infty$.

Proposition 5.1. *Let $\alpha, \beta \in \mathcal{M}^c(\Omega)$. Then,*

$$\mathrm{FG}(\alpha,\beta) = \inf_{\pi\in\mathcal{M}(\Omega\times\Omega)} \langle c,\pi\rangle + \int_\Omega \varphi\left(x, \frac{\mathrm{d}\pi_1}{\mathrm{d}\hat\alpha}\right)\mathrm{d}\hat\alpha + \int_\Omega \varphi\left(x, \frac{\mathrm{d}\pi_2}{\mathrm{d}\hat\beta}\right)\mathrm{d}\hat\beta, \tag{22}$$

where

$$\varphi(x,z) = \begin{cases} |1 - c_{\partial\Omega}(x)\cdot z| & \text{if } z \in \left[0, \frac{1}{c_{\partial\Omega}(x)}\right] \\ +\infty & \text{otherwise.}\end{cases} \tag{23}$$

This proposition allows us to express $\mathrm{FG}(\alpha,\beta)$ in a formalism much closer to standard (non-regularized) unbalanced OT (6): it only involves measures with finite total masses and turns the cost of transporting mass to the boundary $\partial\Omega$ into a penalty between the marginals of π and $(\hat\alpha,\hat\beta)$. This enables the development of a regularized model for OTB (ROTB).

Definition 5.2. *Let $\alpha,\beta \in \mathcal{M}^c(\Omega)\backslash\{0\}$ and $\varepsilon > 0$ be a regularization parameter. The corresponding Homogeneous Regularized Optimal Transport with Boundary (ROTB) problem is given by*

$$\mathrm{FG}_\varepsilon(\alpha,\beta) := \inf_{\pi\in\mathcal{M}(\Omega\times\Omega)} \langle c,\pi\rangle + \varepsilon R(\pi|\hat\alpha,\hat\beta) + \int_\Omega \varphi\left(x, \frac{\mathrm{d}\pi_1}{\mathrm{d}\hat\alpha}\right)\mathrm{d}\hat\alpha + \int_\Omega \varphi\left(x, \frac{\mathrm{d}\pi_2}{\mathrm{d}\hat\beta}\right)\mathrm{d}\hat\beta, \tag{24}$$

where φ is the divergence defined in (23) and R is defined in (14).

Despite this primal problem involves a spatially varying divergence, its dual essentially boils down to a standard problem applied to the renormalized measures $\hat\alpha$ and $\hat\beta$ in this particular setting, allowing us to adapt the results of Section 4 seamlessly.

Proposition 5.3. *Let $\alpha,\beta \in \mathcal{M}^c(\Omega)\backslash\{0\}$. One has, with $m_g = m_g(\hat\alpha,\hat\beta)$ and $m_h = m_h(\hat\alpha,\hat\beta)$,*

$$\mathrm{FG}_\varepsilon(\alpha,\beta) = \sup_{f,g\in\mathcal{C}(\Omega)} \langle \min(1, f/c_{\partial\Omega}), \hat\alpha\rangle + \langle \min(1, g/c_{\partial\Omega}), \hat\beta\rangle - \varepsilon\,\langle \frac{e^{\frac{f\oplus g - c}{\varepsilon}}}{m_g} - \frac{1}{m_h}, \hat\alpha\otimes\hat\beta\rangle\,. \tag{25}$$

Furthermore, if (f,g) is optimal for (25), then $\pi := \exp\left(\frac{f\oplus g - c}{\varepsilon}\right)\frac{\hat\alpha\otimes\hat\beta}{m_g(\hat\alpha,\hat\beta)}$ is optimal for (24).

From this dual formulation, we can derive the corresponding Sinkhorn algorithm with $m_g = m_g(\hat\alpha,\hat\beta)$:

$$\begin{aligned} f_{t+1}(x) &:= \min\left(c_{\partial\Omega}(x), -\varepsilon\log\left(c_{\partial\Omega}(x)\,\langle e^{\frac{g-c(x,\cdot)}{\varepsilon}}, \frac{\hat\beta}{m_g}\rangle\right)\right), \\ g_{t+1}(y) &:= \min\left(c_{\partial\Omega}(y), -\varepsilon\log\left(c_{\partial\Omega}(y)\,\langle e^{\frac{f-c(\cdot,y)}{\varepsilon}}, \frac{\hat\beta}{m_g}\rangle\right)\right). \end{aligned} \tag{26}$$

Finally, we introduce the corresponding notion of Sinkhorn divergence:

$$\mathrm{SkFG}_\varepsilon(\alpha,\beta) := \mathrm{FG}_\varepsilon(\alpha,\beta) - \frac{1}{2}\mathrm{FG}_\varepsilon(\alpha,\alpha) - \frac{1}{2}\mathrm{FG}_\varepsilon(\beta,\beta). \tag{27}$$

This problem enjoys the same properties as the HUROT model introduced in Section 4, as summarized in the following proposition.

Proposition 5.4 (Properties of ROTB)**.**

1. *FG_ε and $\mathrm{SkFG}_\varepsilon$ are 1-homogeneous. The sequence of potentials produced by (26) for the couple of measures $(\lambda\alpha,\lambda\beta)$ is independent of λ, so are the optimal potentials. If π is an optimal plan for (α,β), $\lambda\pi$ is optimal for the couple $(\lambda\alpha,\lambda\beta)$.*
2. *FG_ε is continuous with respect to the weak convergence of the renormalized measures: $\widehat{\alpha_n} \xrightarrow{w} \hat\alpha \Rightarrow \mathrm{FG}_\varepsilon(\alpha_n,\beta) \to \mathrm{FG}_\varepsilon(\alpha,\beta)$. This holds in particular around the null measure by setting $\mathrm{FG}_\varepsilon(0,\beta) := \left(1+\frac{\varepsilon}{2}\right)\mathrm{Pers}(\beta)$ and $\mathrm{FG}_\varepsilon(0,0) = 0$.*
3. *Under the same assumptions as in Proposition 4.8, $\mathrm{SkFG}_\varepsilon(\alpha,\beta) \geq 0$, with equality if and only if $\alpha=\beta$.*

Finally, we state that FG_ε induces the same topology as FG, supporting the use of this model as a proper entropic regularization of the OTB model.

Proposition 5.5. *FG_ε induces the same notion of convergence as FG, that is, for any sequence $(\alpha_n)_n \in \mathcal{M}^c(\Omega)^{\mathbb{N}}$ and any $\alpha \in \mathcal{M}^c(\Omega)$, $\mathrm{SkFG}_\varepsilon(\alpha_n,\alpha) \to 0 \Leftrightarrow \widehat{\alpha_n} \xrightarrow{w} \hat\alpha \Leftrightarrow \mathrm{SkFG}(\alpha_n,\alpha) \to 0$.*

Numerical illustration. To showcase the importance of using an homogeneous model in the context of OTB, we propose the following experiment. Inspired by the context of Topological Data Analysis (see the supplementary materials), we consider the half-plane $\Omega = \{(t_1,t_2),\ t_1 < t_2\} \subset \mathbb{R}^2$ hence $\partial\Omega = \{(t,t),\ t\in\mathbb{R}\}$.We sample two measures α,β with $n = 5$ and $m = 10$ points respectively, and with weight 1 on each point. We then compute the OTB Sinkhorn divergence $\mathrm{SkFG}_\varepsilon(\lambda\alpha,\lambda\beta)$ for $\lambda \in [0.01, 100]$ using our homogeneous model and the Sinkhorn divergence obtained using the standard UROT model (9). Figure 3 showcases the dependence of the result on λ. As expected, our model exhibits 1-homogeneity.

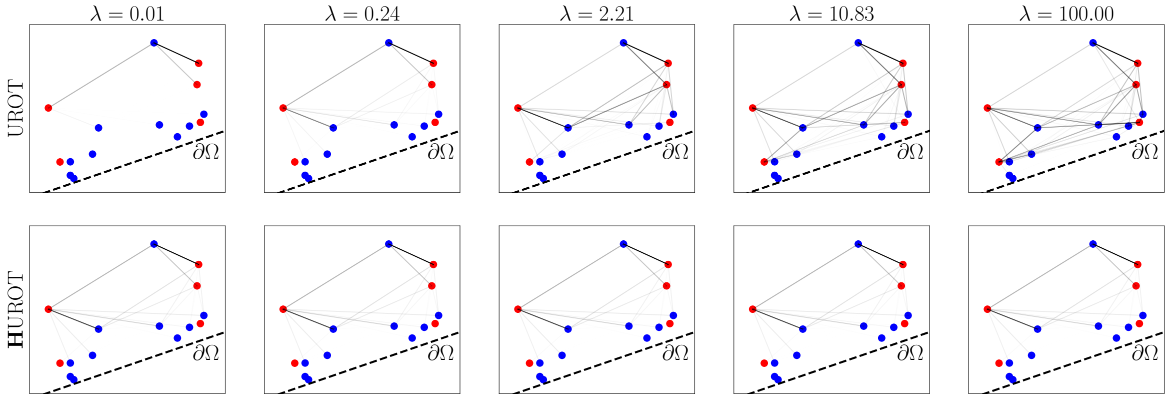

Figure 2: **Impact of inhomogeneity on the transport plan for the OTB model.** *(Top)* The transport plans obtained for the couple $(\lambda\alpha, \lambda\beta)$ for varying λ using the standard UROT model. Inhomogeneity reflects in structural changes in the resulting transport plan: increasing λ tends to weight transportation near the boundary $\partial\Omega$. *(Bottom)* The transportation plans using the HUROT model. As expected, varying λ only rescales the transport plan.

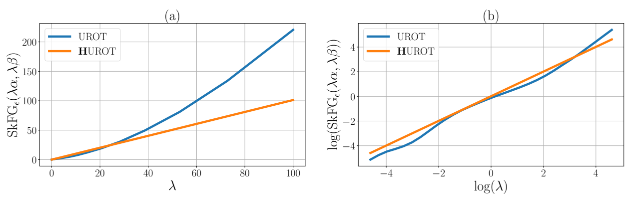

Figure 3: **Importance of homogeneity for the OTB model.** *(a)* The evolution of $\mathrm{FG}_\varepsilon(\lambda\alpha, \lambda\beta)$ for $\lambda \in [0.01, 100]$ using either our homogeneous regularization term (14) (**H**UROT) or the standard one $+\varepsilon\mathrm{KL}(\pi|\hat{\alpha}\otimes\hat{\beta})$ (UROT). As expected, the HUROT model yields a straight line of slope 1. *(b)* Same curve in log-log scale. The various slope breaks illustrate a highly non-homogeneous behavior.

In contrast, the standard model yields a highly inhomogeneous behavior which reflects in many structural changes in the resulting transport plans as showcased in Figure 2. Computations are run with $\varepsilon = 1$.

6 DISCUSSION

We believe that the homogeneous UROT model we propose can provide a good alternative to the standard model of Unbalanced Regularized Optimal Transport proposed by Séjourné et al., especially when *(i)* the marginal divergence induces a "cut-off" as do the Total Variation or spatially varying divergences involved in OT with boundary and when *(ii)* the masses of the measures considered may be ill-defined (e.g. depend on the choice of a unit of measurement) or may largely vary on the considered sample.

Note that enforcing homogeneity in the regularization term comes with some price. In particular, in contrast with the standard UROT model, at fixed β, the map $\alpha \mapsto \mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha, \beta)$ is *a priori* not convex with respect to linear interpolation of measures $(1-t)\alpha + t\alpha'$. Since the resulting homogeneous Sinkhorn divergence still shares key properties with the standard one, wondering whether there exist a convex reparametrization of $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}$ is an important question. Other type of convexity properties, for instance along the interpolation curves described by the optimal transport plans, may also be investigated.

Acknowledgments The author wants to thank the anonymous reviewers along with V. Divol, T. Séjourné and F.-X. Vialard for fruitful discussions that contributed to the development of this work.

References

Ambrosio, L., Gigli, N., and Savaré, G. (2005). *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media.

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.

Chazal, F. and Michel, B. (2021). An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence*, 4.

Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X. (2015). Unbalanced optimal transport: geometry and kantorovich formulation. *arXiv preprint arXiv:1508.05216*.

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.

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300.

Divol, V. and Lacombe, T. (2021). Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport. *Journal of Applied and Computational Topology*, 5(1):1–53.

Edelsbrunner, H. and Harer, J. (2010). *Computational topology: an introduction*. American Mathematical Soc.

Fatras, K., Séjourné, T., Flamary, R., and Courty, N. (2021). Unbalanced minibatch optimal transport; applications to domain adaptation. In *International Conference on Machine Learning*, pages 3186–3197. PMLR.

Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S.-i., Trouvé, A., and Peyré, G. (2019). Interpolating between optimal transport and mmd using sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2681–2690. PMLR.

Figalli, A. (2010). The optimal partial transport problem. *Archive for rational mechanics and analysis*, 195(2):533–560.

Figalli, A. and Gigli, N. (2010). A new transportation distance between non-negative measures, with applications to gradients flows with dirichlet boundary conditions. *Journal de mathématiques pures et appliquées*, 94(2):107–130.

Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., et al. (2021). Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8.

Genevay, A., Peyré, G., and Cuturi, M. (2018). Learning generative models with sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, pages 1608–1617. PMLR.

Janati, H., Cuturi, M., and Gramfort, A. (2020). Debiased sinkhorn barycenters. In *International Conference on Machine Learning*, pages 4692–4701. PMLR.

Kantorovich, L. V. (1942). On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201.

Lacombe, T., Cuturi, M., and Oudot, S. (2018). Large scale computation of means and clusters for persistence diagrams using optimal transport. In *Advances in Neural Information Processing Systems*.

Liero, M., Mielke, A., and Savaré, G. (2015). Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Inventiones mathematicae*, pages 1–149.

Monge, G. (1784). *Mémoire sur la théorie des déblais et des remblais*. Histoire de l'Académie Royale des Sciences de Paris.

Mukherjee, D., Guha, A., Solomon, J. M., Sun, Y., and Yurochkin, M. (2021). Outlier-robust optimal transport. In *International Conference on Machine Learning*, pages 7850–7860. PMLR.

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.

Ramdas, A., Trillos, N. G., and Cuturi, M. (2017). On wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, 19(2):47.

Santambrogio, F. (2015). Optimal transport for applied mathematicians. *Birkäuser, NY*.

Séjourné, T., Feydy, J., Vialard, F.-X., Trouvé, A., and Peyré, G. (2021). Sinkhorn divergences for unbalanced optimal transport. *arXiv preprint arXiv:1910.12958v2*.

Villani, C. (2008). *Optimal transport: old and new*, volume 338. Springer Science & Business Media.

A Delayed proofs

Complementary Notations: The following notions are used in the some proofs. Given a function $K : \Omega \times \Omega \to \mathbb{R}_+$ and $\mu, \nu \in \mathcal{M}(\Omega)$, we introduce

$$\langle \mu, \nu \rangle_K := \iint K(x,y)\mathrm{d}\mu(x)\mathrm{d}\nu(y),$$
$$\|\mu - \nu\|_K^2 := \langle \mu - \nu, \mu - \nu \rangle_K .$$

We say that K defines a positive definite kernel when $\|\mu - \nu\|_K \geq 0$, with equality if and only if $\mu = \nu$. In the following, we assume that $K_\varepsilon : (x,y) \mapsto e^{-\frac{c(x,y)}{\varepsilon}}$ defines a positive definite kernel for any $\varepsilon > 0$ (which holds if, for instance, $c(x,y) = \|x-y\|$ or $\|x-y\|^2$).

A.1 Delayed proofs from Section 3

Proof of Lemma 3.1. In this context, $\varphi^* = \mathrm{id}$ and subsequently, $\mathrm{aprox}_{\varepsilon,\varphi^*} = \mathrm{id}$, hence the Sinkhorn iterations (9) simply read

$$f_{t+1} = -\varepsilon \log \langle e^{\frac{g_t - c}{\varepsilon}}, \beta \rangle ,$$
$$g_{t+1} = -\varepsilon \log \langle e^{\frac{f_{t+1} - c}{\varepsilon}}, \alpha \rangle .$$

We observe that $f_1^{(\lambda)} = -\varepsilon \log \langle e^{\frac{g_0 - c}{\varepsilon}}, \lambda\beta \rangle = f_1 - \varepsilon \log(\lambda)$. Therefore, $g_1^{(\lambda)} = -\varepsilon \log \langle e^{\frac{f_1 - \varepsilon \log(\lambda) - c}{\varepsilon}}, \lambda\alpha \rangle = g_1$, and thus $f_2^{(\lambda)} = -\varepsilon \log \langle e^{\frac{g_1 - c}{\varepsilon}}, \lambda\alpha \rangle = -\varepsilon \log \langle e^{\frac{g_1 - c}{\varepsilon}}, \alpha \rangle - \varepsilon \log(\lambda) = f_2 - \varepsilon \log(\lambda)$. A simple induction gives the conclusion. □

Proof of Corollary 3.2.

1. Since the sequence of potentials $(f_t, g_t)_t$ converges to (f, g) which are optimal for (α, β), it follows from Lemma 3.1 that $(f_t^{(\lambda)}, g_t^{(\lambda)}) \to (f - \varepsilon \log(\lambda), g)$ which must also be a fixed point of the Sinkhorn loop, hence a pair of optimal potentials for the couple $(\lambda\alpha, \lambda\beta)$.

2. Using the primal-dual relation (10), we know that $\pi = e^{\frac{f \oplus g - c}{\varepsilon}} \mathrm{d}\alpha \otimes \beta$ is optimal for the couple (α, β). Therefore, from the previous point,

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

   is optimal for the couple $(\lambda\alpha, \lambda\beta)$.

3. Using $(f - \varepsilon \log(\lambda), g)$ in the dual relation (8), we have

$$\begin{aligned}\mathrm{OT}_\varepsilon(\lambda\alpha, \lambda\beta) =& \lambda \langle f, \alpha \rangle - \varepsilon \log(\lambda)\lambda m(\alpha) + \lambda \langle g, \beta \rangle - \lambda\varepsilon \langle e^{\frac{f \oplus g - c}{\varepsilon}} - \lambda, \alpha \otimes \beta \rangle ,\\ =& \lambda\big(\langle f, \alpha \rangle + \langle g, \beta \rangle - \varepsilon \langle e^{\frac{f \oplus g - c}{\varepsilon}} - 1 + (1 - \lambda), \alpha \otimes \beta \rangle \big) - \varepsilon \log(\lambda)\lambda m(\alpha)\\ =& \lambda\mathrm{OT}_\varepsilon(\alpha, \beta) + \varepsilon\lambda(\lambda - 1)m(\alpha)m(\beta) - \varepsilon \log(\lambda)\lambda m(\alpha)\\ =& \lambda\mathrm{OT}_\varepsilon(\alpha, \beta) + \varepsilon\lambda(\lambda - 1)m^2 - \varepsilon \log(\lambda)\lambda m.\end{aligned}$$

4. The homogeneity of Sk_ε follows from the fact that the "inhomogeneous terms" $+\varepsilon\lambda(\lambda - 1)m^2 - \varepsilon \log(\lambda)\lambda m$ cancel in the definition of the balanced Sinkhorn divergence.

□

Proof of Proposition 3.4. As in the balanced case, we first investigate the behavior of the Sinkhorn algorithm under rescaling of the measures. Let $(f_0, g_0) \in \mathcal{C}(\Omega)$, let $(f_t, g_t)_t$ denote the sequence obtained when iterating the Sinkhorn loop for the couple (α, β) initialized at (f_0, g_0), and let $(f_t^{(\lambda)}, g_t^{(\lambda)})_t$ be the one obtained for the couple $(\lambda\alpha, \lambda\beta)$ with the same initialization. We prove the following by induction:

$$\begin{aligned} f_t^{(\lambda)} &= f_t - \varepsilon u_t \log(\lambda) \\ g_t^{(\lambda)} &= g_t - \varepsilon v_t \log(\lambda), \end{aligned}$$

where $(u_t, v_t) \in \mathbb{R} \times \mathbb{R}$ are real sequences following the relations $u_{t+1} = T(v_t)$ and $v_{t+1} = T(u_{t+1})$ with $T(x) = \frac{1-x}{1+\varepsilon}$, with $u_0, v_0 = 0$. Indeed,

$$\begin{aligned} f_{t+1}^{(\lambda)} &= -\frac{\varepsilon}{1+\varepsilon} \log \int e^{\frac{g_t^{(\lambda)} - c}{\varepsilon}} \mathrm{d}\lambda\beta \\ &= -\frac{\varepsilon}{1+\varepsilon} \log \int e^{\frac{g_t - c}{\varepsilon}} \lambda^{1-v_t} \mathrm{d}\beta \\ &= f_t - \varepsilon \frac{1 - v_t}{1+\varepsilon} \log(\lambda). \end{aligned}$$

A similar computation holds for the second potentials $(g_t)_t$.

The sequences $(u_t)_t$ and $(v_t)_t$ converge to the fixed point of $T \circ T$, given by

$$\begin{aligned} x = T \circ T(x) \Leftrightarrow x &= \frac{x + \varepsilon}{(1+\varepsilon)^2} \\ \Leftrightarrow x &= \frac{\varepsilon}{(1+\varepsilon)^2 - 1} = \frac{1}{2+\varepsilon}, \end{aligned}$$

proving the result linking $(f_t^{(\lambda)}, g_t^{(\lambda)})$ and (f_t, g_t).

From this, simple computations prove the claims:

1. Follows from the fact that $(f_t, g_t)_t$ converges to a couple of optimal dual potentials for (α, β).
2. Follows from the fact that

$$\begin{aligned} &\exp\left(\frac{f \oplus g - 2\frac{\varepsilon}{2+\varepsilon}\log(\lambda) - c}{\varepsilon}\right) \lambda^2 \mathrm{d}\alpha \otimes \beta \\ =&\exp\left(\frac{f \oplus g - c}{\varepsilon}\right) \lambda^{-\frac{2}{2+\varepsilon}} \lambda^2 \mathrm{d}\alpha \otimes \beta = \lambda^h \pi \end{aligned}$$

   is an optimal transport plan for the couple $(\lambda\alpha, \lambda\beta)$.

3. The shift in the potentials induces a change in the objective value $J_{(\lambda\alpha,\lambda\beta)}$ reading

$$\begin{aligned} &\langle 1 - e^{-f + \frac{\varepsilon}{2+\varepsilon}\log(\lambda)}, \lambda\alpha \rangle + \langle 1 - e^{-g + \frac{\varepsilon}{2+\varepsilon}\log(\lambda)}, \lambda\beta \rangle - \varepsilon \langle e^{\frac{f \oplus g - c - 2\frac{\varepsilon}{2+\varepsilon}\log(\lambda)}{\varepsilon}} - 1, \lambda^2 \alpha \otimes \beta \rangle \\ =&\lambda^h \langle 1 - e^{-f}, \alpha \rangle + (\lambda - \lambda^h) m(\alpha) + \lambda^h \langle 1 - e^{-g}, \beta \rangle + (\lambda - \lambda^h) m(\beta) \\ &- \varepsilon \lambda^h \langle e^{\frac{f \oplus g - c}{\varepsilon}} - 1, \alpha \otimes \beta \rangle - \varepsilon(\lambda^h - \lambda^2) m(\alpha) m(\beta) \\ =&\lambda^h \mathrm{OT}_\varepsilon(\alpha, \beta) + (\lambda - \lambda^h)(m(\alpha) + m(\beta)) - \varepsilon(\lambda^h - \lambda^2) m(\alpha) m(\beta). \end{aligned}$$

   Here as well, the non-homogeneous part cancels when considering the Sinkhorn divergence. Note that the linear term involving $(m(\alpha) + m(\beta))$ disappears when adding $-\frac{1}{2}\mathrm{OT}_{\varepsilon,\varphi}(\alpha, \alpha) - \frac{1}{2}\mathrm{OT}_{\varepsilon,\varphi}(\beta, \beta)$, but adding the mass bias term $\frac{\varepsilon}{2}\lambda^2 (m(\alpha) - m(\beta))^2$ is required to cancel the product term that involves $m(\alpha)m(\beta)$.

□

Element of proof for Remark 3.5. We provide elements of proof of the following statement: fix f_0, g_0 (initial points of the Sinkhorn algorithm for UROT) and suppose that for any $\alpha, \beta, \lambda > 0$, $\pi^{(\lambda)} = \lambda\pi$, where π is the optimal transport plan obtained when running the algorithm for the couple (α, β) and $\pi^{(\lambda)}$ the one when running the algorithm with $(\lambda\alpha, \lambda\beta)$, both starting from f_0, g_0 (to avoid ill-definition of the potentials up to constant terms, see (Séjourné et al., 2021, Lemma 4,5)). Let A denote the $\mathrm{aprox}_\varphi^\varepsilon$ operator for the sake of concision. We make in addition the following assumptions to simplify the proof: (i) A is differentiable, (ii) for any α, β, there exist x_0 in $\mathrm{spt}(\alpha)$ such that the map $\lambda \mapsto f^{(\lambda)}(x_0)$ is differentiable at $\lambda = 1$, (iii) the quantity $G := \log \int e^{\frac{g-c}{\varepsilon}} \mathrm{d}\beta$ covers $\mathbb{R}$ as β varies (i.e. we can make G arbitrary), (iv) $A(0) = 0$. Then, necessarily $A(p) = p$ for all $p \in \mathbb{R}$

We believe that these assumptions hold quite generally but do not investigate this in detail in this work. Assuming that the transport plan is 1-homogeneous means that, for any $\lambda > 0$ and any α, β, we have

$$\lambda^2 e^{\frac{f^{(\lambda)} \oplus g^{(\lambda)} - c}{\varepsilon}} \mathrm{d}\alpha \otimes \beta = \lambda e^{\frac{f \oplus g - c}{\varepsilon}} \mathrm{d}\alpha \otimes \beta,$$

that is, $\alpha \otimes \beta$-a.e.,

$$f^{(\lambda)} \oplus g^{(\lambda)} = f \oplus g - \varepsilon \log(\lambda).$$

In particular, $f^{(\lambda)} - f = \varepsilon \log(\tau)$ where τ is a constant (of x) that depends on ε and λ. Note in particular the relation $\partial_\lambda \log(\tau) = \partial_\lambda f^{(\lambda)}$. Re-injecting this in the fixed-point equations satisfied by $f, g, f^{(\lambda)}$ and $g^{(\lambda)}$, we obtain the relation

$$A(\varepsilon G - \varepsilon \log(\tau)) = A(\varepsilon G) - \varepsilon \log(\tau).$$

Deriving this quantity in λ and then evaluating at $\lambda = 1$ (hence $\tau = 1$), yields

$$A'(\varepsilon G) = 1,$$

with arbitrary G. Since $A(0) = 0$, this implies that $A(p) = p$ for all $p \in \mathbb{R}$.

□

A.2 Delayed proofs from Sections 4 and 5

Proof of Proposition 4.3. The proof simply follows from introducing

$$J^{[H]}_{(\alpha,\beta)}(f, g) := \langle -\varphi^*(-f), \alpha \rangle + \langle -\varphi^*(-g), \beta \rangle - \varepsilon \left\langle \frac{e^{\frac{f \oplus g - c}{\varepsilon}}}{m_g(\alpha, \beta)} - \frac{1}{m_h(\alpha, \beta)}, \alpha \otimes \beta \right\rangle$$

and observing that for any $\lambda > 0$, since $m_g(\lambda\alpha, \lambda\beta) = \lambda m_g(\alpha, \beta)$ and $m_h(\lambda\alpha, \lambda\beta) = \lambda m_h(\alpha, \beta)$, we have

$$J^{[H]}_{(\lambda\alpha,\lambda\beta)}(f, g) = \lambda \cdot J^{[H]}_{(\alpha,\beta)}(f, g),$$

yielding the conclusion. □

Proof of Proposition 4.2. The only computations that differ from the proof of duality appearing in (Séjourné et al., 2021) are those corresponding to our slightly modified entropic regularization term.

Introduce $\xi := \frac{\mathrm{d}\pi}{\mathrm{d}\alpha \otimes \beta}$ to alleviate notation.

$$\begin{aligned}
&\frac{\varepsilon}{2} \left(\mathrm{KL}(\pi | \frac{\alpha}{m(\alpha)} \otimes \beta) + \mathrm{KL}(\pi | \alpha \otimes \frac{\beta}{m(\beta)}) \right) \\
=& \frac{\varepsilon}{2} (\langle \xi \log(\xi) - \xi + \log(m(\alpha))\xi + \frac{1}{m(\alpha)}, \alpha \otimes \beta \rangle + \langle \xi \log(\xi) - \xi + \log(m(\beta))\xi + \frac{1}{m(\beta)}, \alpha \otimes \beta \rangle) \\
=& \varepsilon \langle \xi \log(\xi) - \xi + \log(\sqrt{m(\alpha) m(\beta)})\xi + \frac{1}{2} \left(\frac{1}{m(\alpha)} + \frac{1}{m(\beta)} \right), \alpha \otimes \beta \rangle \\
=& \varepsilon \left\langle \xi \log(\xi) - \xi + \log(m_g)\xi + \frac{1}{m_h}, \alpha \otimes \beta \right\rangle .
\end{aligned}$$

In order to obtain the primal-dual relationship, we write

$$
\begin{aligned}
&-\sup_{\pi} \langle f \oplus g, \pi\rangle - \langle c, \pi\rangle - \frac{\varepsilon}{2}(\mathrm{KL}(\pi|\frac{\alpha}{m(\alpha)} \otimes \beta) + \mathrm{KL}(\pi|\alpha \otimes \frac{\beta}{m(\beta)}))\\
=&\inf_{\xi} \langle -(f \oplus g - c)\xi, \alpha \otimes \beta\rangle + \varepsilon \langle \xi \log(\xi) - \xi + \log(m_g)\xi + \frac{1}{m_h}, \alpha \otimes \beta\rangle\\
=&\inf_{\xi} \langle -(f \oplus g - c)\xi + \varepsilon(\xi \log(\xi) - \xi + \log(m_g)\xi + \frac{1}{m_h}), \alpha \otimes \beta\rangle.
\end{aligned}
$$

This optimization problem in ξ yields the primal-dual relation (17).

$$
\xi = \frac{1}{m_g} e^{\frac{f \oplus g - c}{\varepsilon}}, \tag{28}
$$

so that the term

$$
\frac{\varepsilon}{2}\left(\mathrm{KL}\left(\pi|\frac{\alpha}{m(\alpha)} \otimes \beta\right) + \mathrm{KL}\left(\pi|\alpha \otimes \frac{\beta}{m(\beta)}\right)\right)
$$

is equal to

$$
\begin{aligned}
&-(f \oplus g - c)\frac{e^{\frac{f \oplus g - c}{\varepsilon}}}{m_g} + \frac{e^{\frac{f \oplus g - c}{\varepsilon}}}{m_g}(f \oplus g - c) - \varepsilon \log(m_g)\frac{e^{\frac{f \oplus g - c}{\varepsilon}}}{m_g} - \varepsilon\frac{e^{\frac{f \oplus g - c}{\varepsilon}}}{m_g} + \varepsilon \log(m_g)\frac{e^{\frac{f \oplus g - c}{\varepsilon}}}{m_g} + \varepsilon\frac{1}{m_h}\\
=&-\varepsilon\left(\frac{e^{\frac{f \oplus g - c}{\varepsilon}}}{m_g} - \frac{1}{m_h}\right).
\end{aligned}
$$

Eventually

$$
\mathrm{OT}_{\varepsilon}(\alpha, \beta) = \sup_{f,g} \langle -\varphi^*(-f), \alpha\rangle + \langle -\varphi^*(-g), \beta\rangle - \varepsilon \langle \frac{e^{\frac{f \oplus g - c}{\varepsilon}}}{m_g} - \frac{1}{m_h}, \alpha \otimes \beta\rangle. \tag{29}
$$

□

Proof of Proposition 4.6. We know that (f_n, g_n) is optimal for the HUROT model for the couple (α_n, β_n) if and only if it is optimal for the standard model for the couple $\left(\frac{\alpha_n}{m_g(\alpha_n,\beta_n)}, \frac{\beta_n}{m_g(\alpha_n,\beta_n)}\right)$ which converges (as $\alpha_n, \beta_n, \alpha, \beta \neq 0$) to $\left(\frac{\alpha}{m_g(\alpha,\beta)}, \frac{\beta}{m_g(\alpha,\beta)}\right)$.

Using (Séjourné et al., 2021, Prop. 10 and Thm. 2), it implies in the settings considered in this work ($\varphi = \imath_c$, KL or TV) that $(f_n, g_n)_n$ converges (uniformly) toward a pair (f, g) that is optimal (in the HUROT model) for the couple (α, β) and, by continuity of the objective functional in (α, β, f, g) it follows that $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha_n, \beta_n) \to \mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha, \beta)$. □

Proof of Proposition 4.8. The proof of this proposition rely on the following result, adapted from (Séjourné et al., 2021, Prop. 14) (its proof can be found below).

Lemma A.1. *One has*

$$
\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha, \alpha) = \sup_{f \in \mathcal{C}(\Omega)} 2\langle -\varphi^*(-f), \alpha\rangle - \varepsilon \langle e^{\frac{f \oplus f - c}{\varepsilon}} - 1, \frac{\alpha \otimes \alpha}{m(\alpha)}\rangle \tag{30}
$$

Let f_α and g_β be the minimizers of $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha, \alpha)$ and $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\beta, \beta)$, respectively. Note the relation

$$
\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha, \alpha) = 2\langle -\varphi^*(-f_\alpha), \alpha\rangle - \varepsilon \left\| e^{\frac{f_\alpha}{\varepsilon}} \frac{\alpha}{\sqrt{m(\alpha)}} \right\|^2_{K_\varepsilon} + \varepsilon m(\alpha)
$$

and symmetrically in β.

As f_α and g_β are sub-optimal for the dual problem corresponding to $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha,\beta)$, we have:

$$\begin{aligned}
\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha,\beta) \geq & \langle -\varphi^*(-f_\alpha),\alpha\rangle + \langle -\varphi^*(-g_\beta),\beta\rangle - \varepsilon \langle e^{\frac{f_\alpha \oplus g_\beta - c}{\varepsilon}}, \frac{\alpha\otimes\beta}{\sqrt{m(\alpha)m(\beta)}}\rangle \frac{\varepsilon}{2}(m(\alpha)+m(\beta)) \\
\geq & \langle -\varphi^*(-f_\alpha),\alpha\rangle + \langle -\varphi^*(-g_\beta),\beta\rangle v - \varepsilon \langle e^{\frac{f_\alpha}{\varepsilon}} \frac{\alpha}{\sqrt{m(\alpha)}}, e^{\frac{g_\beta}{\varepsilon}} \frac{\beta}{\sqrt{m(\beta)}}\rangle_{K_\varepsilon} + \frac{\varepsilon}{2}(m(\alpha)+m(\beta)) \\
\geq & \frac{1}{2}\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha,\alpha) + \frac{1}{2}\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\beta,\beta) + \frac{\varepsilon}{2}\left\| e^{\frac{f_\alpha}{\varepsilon}} \frac{\alpha}{\sqrt{m(\alpha)}}\right\|^2_{K_\varepsilon} + \frac{\varepsilon}{2}\left\| e^{\frac{g_\beta}{\varepsilon}} \frac{\beta}{\sqrt{m(\beta)}}\right\|^2_{K_\varepsilon} \\
& - \varepsilon \langle e^{\frac{f_\alpha}{\varepsilon}} \frac{\alpha}{\sqrt{m(\alpha)}}, e^{\frac{g_\beta}{\varepsilon}} \frac{\beta}{\sqrt{m(\beta)}}\rangle_{K_\varepsilon}
\end{aligned}$$

so that

$$\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha,\beta) - \frac{1}{2}\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha,\alpha) - \frac{1}{2}\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\beta,\beta) \geq \left\| e^{\frac{f_\alpha}{\varepsilon}} \frac{\alpha}{\sqrt{m(\alpha)}} - e^{\frac{g_\beta}{\varepsilon}} \frac{\beta}{\sqrt{m(\beta)}}\right\|_{K_\varepsilon} \geq 0 \tag{31}$$

which proves the non-negativity.

Furthermore, the equality case reads $e^{\frac{f_\alpha}{\varepsilon}} \frac{\alpha}{\sqrt{m(\alpha)}} = e^{\frac{g_\beta}{\varepsilon}} \frac{\beta}{\sqrt{m(\beta)}}$. By the characterization of f_α and g_β as fixed point of their respective Sinkhorn algorithms, we have

$$\begin{aligned}
f_\alpha &= -\mathrm{aprox}_{\varepsilon,\varphi^*}\left(\varepsilon \log \langle e^{\frac{f_\alpha - c}{\varepsilon}}, \frac{\alpha}{m(\alpha)}\rangle\right), \\
g_\beta &= -\mathrm{aprox}_{\varepsilon,\varphi^*}\left(\varepsilon \log \langle e^{\frac{g_\beta - c}{\varepsilon}}, \frac{\beta}{m(\beta)}\rangle\right).
\end{aligned}$$

Using the equality case aforementioned, we have

$$\begin{aligned}
f_\alpha &= -\mathrm{aprox}_{\varepsilon,\varphi^*}\left(\varepsilon \log \langle e^{\frac{g_\beta - c}{\varepsilon}}, \frac{\beta}{m_g(\alpha,\beta)}\rangle\right), \\
g_\beta &= -\mathrm{aprox}_{\varepsilon,\varphi^*}\left(\varepsilon \log \langle e^{\frac{f_\alpha - c}{\varepsilon}}, \frac{\alpha}{m_g(\alpha,\beta)}\rangle\right).
\end{aligned}$$

Therefore, (f_α, g_β) is actually an optimal couple for the HUROT problem between α and β, as a fixed point of the corresponding Sinkhorn map.

From this, we can write the optimal transport plans $\pi_{\alpha\beta}, \pi_{\alpha\alpha}, \pi_{\beta\beta}$ between the corresponding couple of measures as

$$\begin{aligned}
\pi_{\alpha\beta} &= e^{\frac{f_\alpha \oplus g_\beta - c}{\varepsilon}} \frac{\mathrm{d}\alpha\otimes\beta}{m_g(\alpha,\beta)}, \\
\pi_{\alpha\alpha} &= e^{\frac{f_\alpha \oplus f_\alpha - c}{\varepsilon}} \frac{\mathrm{d}\alpha\otimes\alpha}{m(\alpha)}, \\
\pi_{\beta\beta} &= e^{\frac{g_\beta \oplus g_\beta - c}{\varepsilon}} \frac{\mathrm{d}\beta\otimes\beta}{m(\beta)},
\end{aligned}$$

which actually reads

$$\pi_{\alpha\beta} = \pi_{\alpha\alpha} = \pi_{\beta\beta}.$$

Let π denote this common transportation plan. Since $\mathrm{Sk}^{[H]}_{\varepsilon,\varphi}(\alpha,\beta) = 0$, and observing that the terms $\langle c,\pi\rangle, D_\varphi(\pi_1|\alpha)$ and $D_\varphi(\pi_2|\beta)$ in the primal problems cancel each other, and using the relations

$$\begin{aligned}
&2\mathrm{KL}(\pi|\alpha\otimes\beta) - \mathrm{KL}(\pi|\alpha\otimes\alpha) - \mathrm{KL}(\pi|\beta\otimes\beta) = 0, \\
&\frac{1}{2}\left(\mathrm{KL}\left(\pi|\frac{\alpha\otimes\beta}{m(\alpha)}\right) + \mathrm{KL}\left(\pi|\frac{\alpha\otimes\beta}{m(\beta)}\right)\right) = \mathrm{KL}(\pi|\alpha\otimes\beta) + m(\pi)\log(m_g(\alpha,\beta)) + m_a(\alpha,\beta) - m(\alpha)m(\beta),
\end{aligned}$$

we can write

$$
\begin{aligned}
0 =& \frac{1}{2}\left(\mathrm{KL}\left(\pi|\frac{\alpha\otimes\beta}{m(\alpha)}\right)+\mathrm{KL}\left(\pi|\frac{\alpha\otimes\beta}{m(\beta)}\right)\right)-\frac{1}{2}\mathrm{KL}\left(\pi|\frac{\alpha\otimes\alpha}{m(\alpha)}\right)-\frac{1}{2}\mathrm{KL}\left(\pi|\frac{\beta\otimes\beta}{m(\beta)}\right),\\
=& m(\pi)\log(m_g)+m_a-m(\alpha)m(\beta)-\frac{1}{2}m(\pi)\log(m(\alpha))-\frac{1}{2}m(\alpha)+\frac{1}{2}m(\alpha)^2\\
&-\frac{1}{2}m(\pi)\log(m(\beta))-\frac{1}{2}m(\beta)+\frac{1}{2}m(\beta)^2\\
=& \frac{1}{2}(m(\alpha)-m(\beta))^2
\end{aligned}
$$

which implies that $m(\alpha)=m(\beta)=: m$. From this, it follows that

$$
\begin{aligned}
f_\alpha &= -\mathrm{aprox}_{\varepsilon,\varphi^*}\left(\varepsilon\log\langle e^{\frac{f_\alpha-c}{\varepsilon}},\frac{\alpha}{m}\rangle\right)\\
&= -\mathrm{aprox}_{\varepsilon,\varphi^*}\left(\varepsilon\log\langle e^{\frac{g_\beta-c}{\varepsilon}},\frac{\beta}{m}\rangle\right)\\
&= g_\beta,
\end{aligned}
$$

hence $\alpha=\beta$. ☐

Proof of Lemma A.1. Let $f\in\mathcal{C}(\Omega)$ be optimal in (30). Using the couple (f,f) in (16), we get

$$
\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha,\alpha)\geq\sup_{f\in\mathcal{C}(\Omega)}2\langle-\varphi^*(-f),\alpha\rangle-\varepsilon\langle e^{\frac{f\oplus f-c}{\varepsilon}}-1,\frac{\alpha\otimes\alpha}{m(\alpha)}\rangle.
$$

Now, let $\pi=\exp\left(\frac{f\oplus f-c}{\varepsilon}\right)\frac{\mathrm{d}\alpha\otimes\alpha}{m(\alpha)}$. By the symmetry of c, its marginals are given by $\pi_1=\pi_2=\langle e^{\frac{f-c}{\varepsilon}},\frac{\alpha}{m(\alpha)}\rangle e^{f/\varepsilon}\alpha$. As π is suboptimal in (15), we get

$$
\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha,\alpha)\leq\langle\pi,c\rangle+2D_\varphi(\pi_1|\alpha)+\varepsilon\mathrm{KL}\left(\pi|\frac{\alpha\otimes\alpha}{m(\alpha)}\right).
$$

Now, observe that for $i\in\{1,2\}$,

$$
\frac{\mathrm{d}\pi_i}{\mathrm{d}\alpha}=\langle e^{\frac{f-c}{\varepsilon}},\frac{\alpha}{m(\alpha)}\rangle e^{f/\varepsilon}\in\partial\varphi^*(-f),
$$

and since $\varphi^*(q)=\sup_p pq-\varphi(p)$, we have that $\forall x\in\Omega,\ \varphi^*(-f(x))=-f(x)\frac{\mathrm{d}\pi_1}{\mathrm{d}\alpha}-\varphi\left(\frac{\mathrm{d}\pi_1}{\mathrm{d}\alpha}\right)$. Therefore,

$$
\begin{aligned}
D_\varphi(\pi_1|\alpha)&=\langle\varphi\left(\frac{\mathrm{d}\pi_1}{\mathrm{d}\alpha}\right),\alpha\rangle\\
&=\langle-f\frac{\mathrm{d}\pi_1}{\mathrm{d}\alpha}-\varphi^*(-f),\alpha\rangle\\
&=-\langle f,\pi_1\rangle+\langle-\varphi^*(-f),\pi_1\rangle.
\end{aligned}
$$

On the other hand, denoting $\zeta=\exp\left(\frac{f\oplus f-c}{\varepsilon}\right)$, we have

$$
\begin{aligned}
\varepsilon\mathrm{KL}\left(\pi|\frac{\alpha\otimes\alpha}{m(\alpha)}\right)=&\varepsilon\langle\log(\zeta)\zeta-\zeta+1,\frac{\alpha\otimes\alpha}{m(\alpha)}\rangle\\
=&\langle f\oplus f-c,\pi\rangle-\varepsilon\langle e^{\frac{f\oplus f-c}{\varepsilon}}-1,\frac{\alpha\otimes\alpha}{m(\alpha)}\rangle\\
=&2\langle f,\pi_1\rangle-\langle c,\pi\rangle-\varepsilon\langle e^{\frac{f\oplus f-c}{\varepsilon}}-1,\frac{\alpha\otimes\alpha}{m(\alpha)}\rangle.
\end{aligned}
$$

Summing the terms together yields the result. ☐

Proof of Proposition 4.9. The proof where only $\alpha_n \to 0$ follows the spirit of the one of (Séjourné et al., 2021, Prop. 18), though requiring specific adaptation related to our regularization term. When both measures go to 0, we can leverage the homogeneity of our model to prove the claim easily.

• Using that $\alpha_n \otimes \beta$ is a suboptimal transport plan for (15), we have

$$\begin{aligned}
&\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha_n, \beta) \\
\leq &\langle c, \alpha_n \otimes \beta \rangle + D_\varphi(m(\beta)\alpha_n|\alpha_n) + D_\varphi(m(\alpha_n)\beta|\beta) + \varepsilon R(\alpha_n \otimes \beta|\alpha_n, \beta) \\
\leq &\langle c, \alpha_n \otimes \beta \rangle + m(\alpha_n)\varphi(m(\beta)) + m(\beta)\varphi(m(\alpha_n)) + \varepsilon(m(\alpha_n)m(\beta)\log(m_g(\alpha_n, \beta)) \\
&+ \frac{1}{2}(m(\alpha_n) + m(\beta)) - m(\alpha_n)m(\beta)) \\
\to &\varphi(0)m(\beta) + \frac{\varepsilon}{2}m(\beta).
\end{aligned}$$

On the other hand, Jensen inequality applied to D_φ allows us to write

$$\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha_n, \beta) \geq \inf_\pi \langle c, \pi \rangle + m(\alpha_n)\varphi(m(\pi)) + m(\beta)\varphi(m(\pi)) + \varepsilon R(\pi|\alpha_n, \beta) =: F_n(\pi).$$

We observe that

$$\lim_{n\to\infty} F_n(\pi) \begin{cases} \geq \langle c, \pi \rangle + m(\beta)\varphi(m(\pi)) + \frac{\varepsilon}{2}\mathrm{KL}(\pi|0) = +\infty & \text{if } \pi \neq 0, \\ = m(\beta)\varphi(0) + \frac{\varepsilon}{2}m(\beta) & \text{if } \pi = 0, \end{cases}$$

where the second equality follows from the relation

$$R(\pi|\alpha_n, \beta) = \mathrm{KL}(\pi|\alpha_n \otimes \beta) - m(\pi)\log(m_g(\alpha_n, \beta)) + m_a(\alpha_n, \beta) - m(\alpha_n)m(\beta)$$

which evaluates to $\frac{1}{2}m(\beta)$ for $\pi = 0$ and $\alpha_n \to 0$.

As F_n is lower-semi-continuous, it follows that $\lim_n \mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha_n, \beta) \geq \varphi(0)m(\beta) + \frac{\varepsilon}{2}m(\beta)$, and finally

$$\lim_{n\to\infty} \mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha_n, \beta) = \left(\varphi(0) + \frac{\varepsilon}{2}\right)m(\beta),$$

proving the continuity of $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}$ around couple of the form $(0, \beta)$ when $\beta \neq 0$.

• We now consider two sequences $\alpha_n, \beta_n \xrightarrow{w} 0$. Define $M_n = \max(m(\alpha_n), m(\beta_n))$. Using the homogeneity of our model, we can write

$$\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha_n, \beta_n) = M_n \cdot \mathrm{OT}^{[H]}_{\varepsilon,\varphi}\left(\frac{\alpha_n}{M_n}, \frac{\beta_n}{M_n}\right).$$

Using $\frac{\alpha_n}{M_n} \otimes \frac{\beta_n}{M_n}$ as a suboptimal transport plan, we have (note that the two measures have total masses ≤ 1)

$$\mathrm{OT}^{[H]}_{\varepsilon,\varphi}\left(\frac{\alpha_n}{M_n}, \frac{\beta_n}{M_n}\right) \leq \|c\|_\infty + \varphi\left(\frac{\alpha_n}{M_n}\right) + \varphi\left(\frac{\beta_n}{M_n}\right) + 1.$$

As φ is bounded over $[0,1]$, it follows that $\left(\mathrm{OT}^{[H]}_{\varepsilon,\varphi}\left(\frac{\alpha_n}{M_n}, \frac{\beta_n}{M_n}\right)\right)_n$ is bounded as well, hence since $M_n \to 0$,

$$\lim_{n\to\infty} \mathrm{OT}^{[H]}_{\varepsilon,\varphi}(\alpha_n, \beta_n) = 0,$$

proving the continuity in this case as well. □

Proof of Proposition 5.1. Let $\alpha, \beta \in \mathcal{M}^c(\Omega)$ and $\pi \in \mathrm{Adm}(\alpha, \beta)$. Without loss of generality, we can assume that $\pi(\partial\Omega \times \partial\Omega) = 0$ (Figalli and Gigli, 2010, Eq. (4)) and that $\forall A \subset \Omega,\ \pi(A \times \partial\Omega) = \pi(A \times P(A))$. Let also $\pi_1 = \pi(\cdot \times \Omega)$ and $\pi_2 = \pi(\Omega \times \cdot)$, that are the marginals of the restricted plan $\pi_{|\Omega\times\Omega}$. Note the constraints

$\pi_1 \leq \alpha,\ \pi_2 \leq \beta$. It allows us to write

$$\begin{aligned}
&\iint_{\overline{\Omega}\times\overline{\Omega}} c(x,y)\mathrm{d}\pi(x,y) \\
&= \iint_{\Omega\times\Omega} c(x,y)\mathrm{d}\pi + \int_{\Omega\times\partial\Omega} c_{\partial\Omega}(x)\mathrm{d}\pi + \int_{\partial\Omega\times\Omega} c_{\partial\Omega}(y)\mathrm{d}\pi \\
&= \iint_{\Omega\times\Omega} c(x,y)\mathrm{d}\pi + \int_{\Omega\times\partial\Omega} c_{\partial\Omega}(x)\mathrm{d}(\alpha-\pi_1) + \int_{\partial\Omega\times\Omega} c_{\partial\Omega}(y)\mathrm{d}(\beta-\pi_2) \\
&= \iint_{\Omega\times\Omega} c(x,y)\mathrm{d}\pi + \int_{\Omega\times\partial\Omega} \mathrm{d}(\hat{\alpha} - c_{\partial\Omega}(x)\pi_1) + \int_{\partial\Omega\times\Omega} \mathrm{d}(\hat{\beta} - c_{\partial\Omega}(y)\pi_2) \\
&= \iint_{\Omega\times\Omega} c(x,y)\mathrm{d}\pi + \int_{\Omega\times\partial\Omega} \left(1 - c_{\partial\Omega}(x)\frac{\mathrm{d}\pi_1}{\mathrm{d}\hat{\alpha}}\right)\mathrm{d}\hat{\alpha} + \int_{\Omega\times\partial\Omega} \left(1 - c_{\partial\Omega}(y)\frac{\mathrm{d}\pi_2}{\mathrm{d}\hat{\beta}}\right)\mathrm{d}\hat{\beta} \\
&= \iint_{\Omega\times\Omega} c(x,y)\mathrm{d}\pi + \int_{\Omega} \varphi\left(x, \frac{\mathrm{d}\pi_1}{\mathrm{d}\hat{\alpha}}\right)\mathrm{d}\hat{\alpha} + \int_{\Omega} \varphi\left(x, \frac{\mathrm{d}\pi_2}{\mathrm{d}\hat{\beta}}\right)\mathrm{d}\hat{\beta}.
\end{aligned}$$

From this, we observe that $\pi \in \mathrm{Adm}(\alpha,\beta)$ induces a plan $\pi' = \pi_{|\Omega\times\Omega} \in \mathcal{M}(\Omega\times\Omega)$ which implies that

$$\mathrm{FG}(\alpha,\beta) \geq \inf_{\pi' \in \mathcal{M}(\Omega\times\Omega)} \langle c, \pi'\rangle + \int_{\Omega} \varphi\left(x, \frac{\mathrm{d}\pi_1}{\mathrm{d}\hat{\alpha}}\right)\mathrm{d}\hat{\alpha} + \int_{\Omega} \varphi\left(x, \frac{\mathrm{d}\pi_2}{\mathrm{d}\hat{\beta}}\right)\mathrm{d}\hat{\beta} =: F(\pi').$$

Conversely, consider $\pi' \in \mathcal{M}(\Omega\times\Omega)$. Let π'_1, π'_2 denote its marginals. Observe that if $\pi'_1 \not\leq \alpha$ or $\pi'_2 \not\leq \beta$, the choice of φ implies that $F(\pi') = +\infty$, so we can restrict to such plans. They naturally induce an element $\pi \in \mathrm{Adm}(\alpha,\beta)$ defined by $\pi = \pi'$ on $\Omega\times\Omega$, and $\forall A \subset \Omega, B \subset \partial\Omega,\ \pi(A\times B) = (\alpha - \pi'_1)(P^{-1}(B)\cap A)$ (and symmetrically in β, π'_2), and $\iint_{\overline{\Omega}\times\overline{\Omega}} c\mathrm{d}\pi = F(\pi')$, proving the claim by taking the infimum. ☐

Proof of Proposition 5.5. The fact that $\widehat{\alpha_n} \xrightarrow{w} \hat{\alpha} \Leftrightarrow \mathrm{FG}(\alpha_n,\alpha) \to 0$ is already known (Divol and Lacombe, 2021, Cor. 3.2). Therefore, it remains to show that $\mathrm{SkFG}_\varepsilon(\alpha_n,\alpha) \to 0 \Leftrightarrow \widehat{\alpha_n} \xrightarrow{w} \hat{\alpha}$.

The converse implication is given by the continuity of $\mathrm{SkFG}_\varepsilon$ with respect to the weak convergence of the normalized measures (Proposition 5.4). Now, assume that $\mathrm{SkFG}_\varepsilon(\alpha_n,\alpha) \to 0$. If the sequence $(\widehat{\alpha_n})_n$ has uniformly bounded mass (i.e. $(\alpha_n)_n$ has uniformly bounded total persistence), we know that it must be compact with respect to the weak convergence (as Ω is bounded). If so, extracting a converging subsequence converging to some limit $\widehat{\alpha_\infty}$ yields by continuity $\mathrm{SkFG}_\varepsilon(\alpha_\infty,\alpha) = 0$ and thus $\alpha_\infty = \alpha$. This makes $(\widehat{\alpha_n})_n$ a compact sequence with $\hat{\alpha}$ as unique limit, implying $\widehat{\alpha_n} \xrightarrow{w} \hat{\alpha}$.

Therefore, it remains to show that $\mathrm{SkFG}_\varepsilon(\alpha_n,\alpha) \to 0 \Rightarrow \sup_n \mathrm{Pers}(\alpha_n) = \sup_n m(\widehat{\alpha_n}) < +\infty$. Let f_n denotes the optimal symmetric potential for the dual problem (25) corresponding to the couple (α_n,α_n), and $f_{\hat{\alpha}}$ be the one corresponding to the couple (α,α). The optimality condition on f_n gives $f_n \geq -c_{\partial\Omega} \geq -L$, where $L = \mathrm{diam}(\Omega)$. Assume first that $\alpha_n, \alpha \neq 0$. One has

$$\mathrm{SkFG}_\varepsilon(\alpha_n,\alpha) \geq \frac{\varepsilon}{2}\| e^{\frac{f_n}{\varepsilon}} \frac{\widehat{\alpha_n}}{\sqrt{m(\widehat{\alpha_n})}} - e^{\frac{f_{\hat{\alpha}}}{\varepsilon}} \frac{\widehat{\alpha}}{\sqrt{m(\widehat{\alpha})}} \|_{K_\varepsilon}.$$

Since $\mathrm{SkFG}_\varepsilon(\alpha_n,\alpha) \to 0$, one has $\sup_n \left\| e^{\frac{f_n}{\varepsilon}} \frac{\widehat{\alpha_n}}{\sqrt{m(\widehat{\alpha_n})}} \right\|_{K_\varepsilon} < \infty$, and since $(f_n)_n$ is (uniformly) lower bounded, necessarily, $(m(\widehat{\alpha_n}))_n$ is bounded, proving the claim. If $\alpha = 0$, the same reasoning yields $e^{\frac{f_n}{\varepsilon}} \frac{\widehat{\alpha_n}}{\sqrt{m(\widehat{\alpha_n})}} \xrightarrow{w} 0$, thus $\sup_n m(\widehat{\alpha_n}) < \infty$ and $\alpha_n \xrightarrow{w} 0$. ☐

B Complementary remarks

Remark B.1. *The presence of the term $+\frac{\varepsilon}{2}(m(\alpha)-m(\beta))^2$ in (11), called the* mass bias*, is required to make the unbalanced Sinkhorn divergence non-negative (and convex). Intuitively, this term arises from the constant term $-\varepsilon\langle -1, \alpha\otimes\beta\rangle = \varepsilon m(\alpha)m(\beta)$ in (8): while in the balanced case ($m(\alpha) = m(\beta)$), these terms cancel each other when computing the Sinkhorn divergence (5), in the unbalanced case, they yield a constant term*

$\varepsilon(m(\alpha)m(\beta) - \frac{1}{2}m(\alpha)^2 - \frac{1}{2}m(\beta)^2) = -\frac{\varepsilon}{2}(m(\alpha) - m(\beta))^2$ *that must be compensated by the mass bias term to ensure the good behavior of the model, in particular its non-negativity.*

Remark B.2. *It may be appealing to replace the entropic regularization term* (14) *by* $\varepsilon\mathrm{KL}\left(\pi|\frac{\alpha\otimes\beta}{m_g(\alpha,\beta)}\right)$. *This indeed leads to an homogeneous problem that shares most of the properties of the proposed* $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}$. *Actually, the dual formulation would read*

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

so that the two quantities only differ from a constant term and are substantially equivalent. Note also that $\pi \mapsto \frac{\varepsilon}{2}\left(\mathrm{KL}\left(\pi|\frac{\alpha}{m(\alpha)}\otimes\beta\right) + \mathrm{KL}\left(\pi|\alpha\otimes\frac{\beta}{m(\beta)}\right)\right)$ *is minimized for* $\pi = \frac{\alpha\otimes\beta}{m_g(\alpha,\beta)}$, *so both entropic terms play morally the same role.*

The one we propose presents the advantage of leading to a Sinkhorn divergence that does not need the introduction of a mass bias term: using $\varepsilon\mathrm{KL}\left(\pi|\frac{\alpha\otimes\beta}{m_g(\alpha,\beta)}\right)$ *would require to add* $+\varepsilon(\sqrt{m(\alpha)} - \sqrt{m(\beta)})^2$ *to the corresponding Sinkhorn divergence to make it positive. Interestingly, this mass bias correspond to a sort of Hellinger distance between the masses of the two measures.*

Remark B.3. *The key (and essentially sole) difference between* (22) *and* (6) *is the dependence of the divergence* φ *on the location* x, *a situation referred to as "spatially varying divergence" in (Séjourné et al., 2021,* Remark *3). This formalism is substantially equivalent to the standard one and most computations adapt seamlessly with the choice of* φ *used in this section. The HUROT model could have been presented directly in the more general context of spatially varying divergences in Section 4, but this would have required several additional assumptions on* φ *and would have hinder the use of many results of (Séjourné et al., 2021) directly. For the sake of simplicity, we prefer to deal with spatially varying divergences only in this section and for the particular choice* (23) *of* φ *that allows us to retrieve (when* $\varepsilon = 0$*) the model of Figalli and Gigli* (21).

Remark B.4. *The formulation* (22) *shows that OT with boundary can be recast as a (spatially varying) UOT problem involving the couple of renormalized measures* $(\hat{\alpha}, \hat{\beta})$, *justifying to use this couple as reference measure in the entropic reference measure in* (24). *Intuitively, it makes the entropic regularization term sensitive to the geometry of the problem, down weighting the points close to the boundary* $\partial\Omega$. *Formally, the choice of* $(\hat{\alpha}, \hat{\beta})$ *as reference is theoretically supported by the fact that the Sinkhorn divergence corresponding to* FG_ε *induces the same convergence as the non-regularized problem* (21), *as detailed in Proposition 5.5.*

Remark B.5 (Links between OTB and Topological Data Analysis.)**.** *The OTB transportation model has not been widely used in OT literature to the best of our knowledge*[1]. *However, it has been recently shown in (Divol and Lacombe, 2021) that the metric* FG *does exactly coincide with the metrics used by the Topological Data Analysis (TDA) community to compare* Persistence diagrams *(PDs), a type of descriptor routinely used to compare objects with respect to their topological properties, see (Edelsbrunner and Harer, 2010; Chazal and Michel, 2021) for an overview. This connection appeared to be fruitful and enabled the adaptation of various tools—both theoretical and computational ones—existing in the OT literature to the context of TDA. In a related work (Lacombe et al., 2018), still in the context of TDA, authors proposed a regularized version of* (21) *by (substantially) adding a term* $+\varepsilon\mathrm{KL}(\pi|\mathcal{L})$, *where* $\mathcal{L}$ *denotes the Lebesgue measure on* $\Omega\times\Omega$. *However, using the Lebesgue measure (or even* $\alpha\otimes\beta$*) as reference measure (aside from non-homogeneity) has several drawbacks. It is only properly defined for measures with finite total masses, indicating possible problems when the masses of the measures get large in practice—even though the exact distances could be mostly unchanged if the additional mass is close to the boundary* $\partial\Omega$. *In the same vein, it does not follow the spirit of OT with boundary, which tells that points near* $\partial\Omega$ *have a lesser importance.*

Remark B.6. *Contrary to the standard UROT model,* $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}(0,\beta)$ *depends on* ε *(the result is simply* $\varphi(0)m(\beta)$ *in the standard model). This can be seen as an artifact of the fact that our model directly encompasses the "mass bias" in the functional* $\mathrm{OT}^{[H]}_{\varepsilon,\varphi}$.

Remark B.7. *The development of different numerical tools in the context of OT with boundary, in particular regularized Fréchet means, is a natural follow-up of this work. Note that in the context of topological data analysis (which is related to OT with boundary, see Remark B.5), regularized barycenters for persistence diagrams have*

[1] In comparison, for instance, to the UROT model.

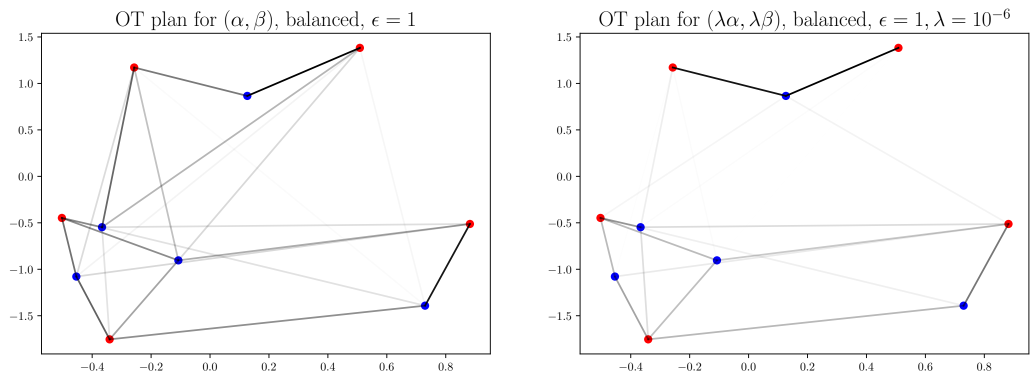

Figure 4: Optimal transport plans returned by `pot.bregman.empirical_sinkhorn` for $(\lambda\alpha, \lambda\beta)$ with $\lambda = 1$ (left) and $\lambda = 10^{-6}$ (right) for two measures α, β randomly sampled with $n = 5$ points; $m(\alpha) = m(\beta) = 1$ (mass $1/n$ on each point). Entropic regularization parameter $\varepsilon = 1$ (`reg=1.` in `pot`).

been developed (Lacombe et al., 2018). However, the proposed approach uses the Lebesgue measure as reference measure in their entropic regularization term. This yields points near the boundary of the space, which tend to outnumber farther points in applications, to outweigh them as well. Using instead the reweighted measures $\hat{\alpha}, \hat{\beta}$ *and our homogeneous formulation is likely to improve the quality of the numerical results that can be obtained.*

We eventually provide a numerical illustration of the warning described in Remark 3.3. Even if the entropic OT plan is *theoretically* 1-homogeneous in the balanced case, the *stopping criterion* for Sinkhorn may suffer from the inhomogeneity of the dual potentials. This happens for the method `pot.bregman.empirical_sinkhorn` for instance, where current implementation (as of version `0.8.2`) implies a different numerical behavior for measures when they have a (same) low mass (which may simply reflect a change of units), as illustrated in Figure 4: the returned OT plans for rescaled measures exhibit different structure as they do not reach the stopping criterion at the same time. We stress nonetheless that the stopping criterion used in `POT 0.8.2`, namely checking the marginal error, appears to be much more robust to mass rescaling than a more naive approach such as checking the variation of the objective function (which may be a natural idea though). This issue cannot occur with HUROT because the sequence of dual potentials produced by the (adapted) Sinkhorn algorithm are insensitive to scaling.