

**DFG** Deutsche  
Forschungsgemeinschaft  
Priority Programme 1962

# *Entropic Regularization of Continuous Optimal Transport Problems*

Christian Clason, Dirk A. Lorenz, Hinrich Mahler, Benedikt Wirth



Preprint Number SPP1962-121

received on September 23, 2019

Edited by  
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)  
Leibniz Institute in the Forschungsverbund Berlin e.V.  
Mohrenstraße 39, 10117 Berlin, Germany  
E-Mail: [spp1962@wias-berlin.de](mailto:spp1962@wias-berlin.de)

World Wide Web: <http://spp1962.wias-berlin.de/>

# ENTROPIC REGULARIZATION OF CONTINUOUS OPTIMAL TRANSPORT PROBLEMS

Christian Clason\*    Dirk A. Lorenz†    Hinrich Mahler‡    Benedikt Wirth§

**Abstract** We analyze continuous optimal transport problems in the so-called Kantorovich form, where we seek a transport plan between two marginals that are probability measures on compact subsets of Euclidean space. We consider the case of regularization with the negative entropy, which has attracted attention because it can be solved in the discrete case using the very simple Sinkhorn algorithm. We first analyze the problem in the context of classical Fenchel duality and derive a strong duality result for a predual problem in the space of continuous functions. However, this problem may not admit a minimizer, which prevents obtaining primal-dual optimality conditions that can be used to justify the Sinkhorn algorithm on the continuous level. We then show that the primal problem is naturally analyzed in the Orlicz space of functions with finite entropy and derive a dual problem in the corresponding dual space, for which existence can be shown and primal-dual optimality conditions can be derived. For marginals that do not have finite entropy, we finally show Gamma-convergence of the regularized problem with smoothed marginals to the original Kantorovich problem.

## 1 INTRODUCTION

The Kantorovich formulation of optimal transport is the problem of finding a transport plan that describes how to move some measure onto another measure of the same mass such that a certain cost functional is minimal [12]. Specifically, let  $\Omega_1$  and  $\Omega_2$  be two compact subset of  $\mathbb{R}^{n_1}$  and  $\mathbb{R}^{n_2}$ , respectively. For given probability measures  $\mu$  on  $\Omega_1$  and  $\nu$  on  $\Omega_2$  and a continuous cost function  $c : \Omega_1 \times \Omega_2 \rightarrow [0, \infty)$ , the goal is to find a measure  $\pi$  on  $\Omega_1 \times \Omega_2$  such that the cost  $\int_{\Omega_1 \times \Omega_2} c \, d\pi$  is minimal among all  $\pi$  that have  $\mu$  and  $\nu$  as marginals. This problem has been well studied, and we refer to the recent books [21, 19] for an overview. For example it is known that the problem has a solution  $\pi$  and that the support of  $\pi$  is contained in the so-called  $c$ -superdifferential of a  $c$ -concave function on  $\Omega_1$ , see [1, Thm. 1.13]. (This is sometimes called the *Fundamental Theorem of Optimal Transport*.) In the case where  $\Omega_1$  and  $\Omega_2$  are both subsets of  $\mathbb{R}^n$  and where  $c(x_1, x_2) = |x_1 - x_2|^2$  is the squared Euclidean distance, this implies that the support of an optimal plan  $\pi$  is singular with respect

---

B.W.'s research has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Cells-in-Motion Cluster of Excellence (EXC 1003 – CiM), University of Münster, Germany, as well as Germany's Excellence Strategy – EXC 2044 –, Mathematics Münster: Dynamics – Geometry – Structure. B.W. was additionally supported by the Alfried Krupp Prize for Young University Teachers awarded by the Alfried Krupp von Bohlen und Halbach-Stiftung.

\*Faculty of Mathematics, University of Duisburg-Essen, 45117 Essen, Germany ([christian.clason@uni-due.de](mailto:christian.clason@uni-due.de), ORCID: [0002-9948-8426](https://orcid.org/0002-9948-8426))

†Institute of Analysis and Algebra, TU Braunschweig, 38092 Braunschweig, Germany ([d.lorenz@tu-braunschweig.de](mailto:d.lorenz@tu-braunschweig.de), ORCID: [0000-0002-7419-769X](https://orcid.org/0000-0002-7419-769X))

‡Institute of Analysis and Algebra, TU Braunschweig, 38092 Braunschweig, Germany ([h.mahler@tu-braunschweig.de](mailto:h.mahler@tu-braunschweig.de), ORCID: [0000-0001-9108-549X](https://orcid.org/0000-0001-9108-549X))

§Applied Mathematics Münster, University of Münster, Einsteinstraße 62, 48149 Münster, Germany ([benedikt.wirth@uni-muenster.de](mailto:benedikt.wirth@uni-muenster.de), ORCID: [0000-0003-0393-1938](https://orcid.org/0000-0003-0393-1938))

to the Lebesgue measure. This motivates the use of regularization of the continuous problem to obtain approximate solutions that are functions instead of measures and of discretization techniques to solve the regularized problem.

In this work we focus on *entropic regularization* by adding the negative entropy of  $\pi$  to the objective function. This forces the optimal plan to be a measure which has a density with respect to the Lebesgue measure. Furthermore, in the discrete setting, this allows the numerical solution by the very simple Sinkhorn algorithm [13, 8, 3].

To fully state the regularized optimal transport problem, we introduce some notation. With  $\mathcal{M}(\Omega)$  and  $\mathcal{P}(\Omega)$  we denote the set of Radon and probability measures on  $\Omega \subset \mathbb{R}^n$ , respectively. The Lebesgue measure will be denoted by  $\mathcal{L}$  (and the set on which it is defined will always be clear from the context), and integrals with respect to the Lebesgue measure are just denoted by  $dx$  with the appropriate integration variable  $x$ . In the case where a measure  $\pi$  has a density with respect to the Lebesgue measure, we will also use  $\pi$  for that density. For a measure  $\mu \in \mathcal{M}(\Omega_1)$  and  $g : \Omega_1 \rightarrow \Omega_2$  we denote by  $g\#\mu$  the pushforward of  $\mu$  by  $g$ , i.e., a measure on  $\Omega_2$  defined by  $g\#\mu(B) = \mu(g^{-1}(B))$  for all measurable sets  $B \subset \Omega_2$ . In particular, we will use the coordinate projections  $P_i : \Omega_1 \times \Omega_2 \rightarrow \Omega_i$ ,  $P_i(x_1, x_2) = x_i$ , and the fact that  $P_{i\#}\pi$  is the  $i$ th marginal of  $\pi \in \mathcal{M}(\Omega_1 \times \Omega_2)$ . The entropically regularized Kantorovich problem of optimal mass transport between  $\mu \in \mathcal{P}(\Omega_1)$ ,  $\nu \in \mathcal{P}(\Omega_2)$  is then given by

$$(P) \quad \inf_{\substack{\pi \in \mathcal{P}(\Omega_1 \times \Omega_2), \\ (P_1)\#\pi = \mu, (P_2)\#\pi = \nu}} \int_{\Omega_1 \times \Omega_2} c \, d\pi + \gamma \int_{\Omega_1 \times \Omega_2} \pi(\log \pi - 1) \, d(x_1, x_2).$$

A purely formal application of convex duality then yields the predual problem

$$(D) \quad \sup_{\substack{\alpha \in C(\Omega_1) \\ \beta \in C(\Omega_2)}} \int_{\Omega_1} \alpha(x_1) \, d\mu(x_1) + \int_{\Omega_2} \beta(x_2) \, d\nu(x_2) - \gamma \int_{\Omega_1 \times \Omega_2} \exp\left(\frac{-c(x_1, x_2) + \alpha(x_2) + \beta(x_1)}{\gamma}\right) \, d(x_1, x_2).$$

Having a primal and a dual problem, it is now possible to write down the system of Fenchel–Rockafellar extremality conditions and derive and analyze algorithms to solve this system; in fact, this is one of the possible ways of deriving the Sinkhorn algorithm in the discrete case. However, the existence of solutions to (D) – which is necessary to rigorously obtain extremality conditions – is not obvious in the continuous case, and we could not find any results in this direction. This motivated the study described in this work. As it turns out, neither (P) nor (D) may admit a solution in the considered spaces. For the primal problem, it is necessary and sufficient for the marginals to be in the Banach space  $L \log L$  of functions of finite entropy; correspondingly, a reformulation of the predual problem in the *dual* space  $L_{\exp}$  allows showing existence of a maximizer and deriving the continuous analog of the primal-dual optimality conditions that can be used as the starting point for the classical Sinkhorn algorithm.

We briefly comment on related literature. The continuous optimal transport problem has been analyzed in the survey paper [14] where the relation to the so-called dynamic Schrödinger problem is made. There, the author explicitly states that existence for the dual problem is neither obvious nor proven. Another even more recent survey [10] presents an existence proof for a reparameterized optimality system based on the convergence analysis for a continuous variant of Sinkhorn’s algorithm (and attributes the proof and the algorithm to Fortet [11]). In [6], primal existence and  $\Gamma$ -convergence (for  $\gamma \rightarrow 0$  and fixed marginals with densities with finite entropy) has been shown in the subset of the space of measures which have a density with respect to the Lebesgue measure that has finite entropy. Furthermore, [7] analyzes the problem (for unbalanced transport, i.e., for marginals with different mass) in  $L^1$  and derives a dual formulation in  $L^\infty$ . In both cases, the choice of spaces does not seem to reflect the nature of the functional well; correspondingly, the question of existence of a solution of the respective dual problem is not answered. Finally, [15] analyzes regularization with the  $L^2$  norm of  $\pi$

and derives existence of solutions of the dual problem; to the best of our knowledge, this seems to be the first paper where existence of solutions of the dual of regularized optimal transport problems has been shown.

The rest of the paper is organized as follows. [Section 2](#) recalls statements about functions of finite entropy and the duality of the respective Orlicz space  $L \log L$ . In [Section 3](#), we analyze the regularized optimal transport problem (P), show existence and uniqueness in  $L \log L$  for the solution of the primal problem (P), and derive the dual problem and show existence of solutions for the dual problem. We finally state a result on  $\Gamma$ -convergence for the combined regularization and smoothing of marginals that do not have finite entropy in [Section 5](#).

## 2 REVIEW OF FUNCTIONS OF FINITE ENTROPY AND THE SPACE $L \log L$

Entropic regularization deals with positive integrable functions of finite entropy. These functions are closely connected to the space  $L \log L$ , a special case of (Birnbaum–)Orlicz spaces, and hence we collect some facts about this space which are mainly taken from [\[17, 4\]](#); see also [\[20\]](#). We consider a compact domain  $\Omega \subset \mathbb{R}^n$ , which we assume for simplicity to have unit Lebesgue measure (although the following only requires the measure to be finite). We denote the *neg-entropy* of a measurable function  $f : \Omega \rightarrow \mathbb{R}$  by

$$E(f) = \int_{\Omega} |f(x)| \log(|f(x)|) dx,$$

where we set  $0 \log 0 = 0$  as usual. Note that since  $x \log x \geq -1/e$ , the neg-entropy always lies in the interval  $[-1/e, \infty]$ . We say that  $f$  has finite entropy if  $E(f) < \infty$ . Following [\[4\]](#), the space  $L \log L(\Omega)$  consists of all measurable functions  $f$  for which

$$\int_{\Omega} |f(x)| \log^+(|f(x)|) dx < \infty,$$

where  $\log^+(x) = \max(\log(x), 0)$ .

**Proposition 2.1** (see [\[16, Thm. 1.2\]](#)). *A nonnegative measurable function  $f$  on a set with finite measure has finite entropy if and only if  $f \in L \log L(\Omega)$ .*

It turns out that  $L \log L(\Omega)$  can be normed such that it becomes a Banach space and that its dual has a natural characterization. In the following, we sketch the central constructions and main results based on the *decreasing rearrangement* (which is one of several alternative approaches to introduce a norm on  $L \log L(\Omega)$ ).<sup>1</sup> Here and throughout the rest of the paper, we use the usual shorthand  $\{f > \lambda\}$  for the set  $\{x \in \Omega : f(x) > \lambda\}$  as well as similar such sets.

**Definition 2.2.** For a measurable function  $f$  on  $\Omega \subset \mathbb{R}^n$ , we define its *decreasing rearrangement*  $f^* : [0, \infty) \rightarrow [0, \infty)$  by

$$f^*(t) = \sup \{ \lambda > 0 : \mathcal{L}(\{|f| > \lambda\}) > t \}.$$

Moreover, we define  $f^{**} : [0, \infty) \rightarrow [0, \infty)$  by

$$f^{**}(t) = \frac{1}{t} \int_0^t f^*(s) ds.$$

We collect some properties of  $f^*$  and  $f^{**}$  that are either obvious or can be found in [\[4\]](#).

**Lemma 2.3.**

<sup>1</sup>The following notations should not be confused with that of the Fenchel conjugate and biconjugate. For this reason, we avoid introducing the latter notation in this work.

- (i)  $f^*$  is nonnegative and decreasing.
- (ii)  $f^{**}$  is nonnegative, decreasing, and continuous.
- (iii)  $f^* \leq f^{**}$ .
- (iv) For  $t_1, t_2 \geq 0$  and two measurable functions  $f$  and  $g$ , it holds that

$$(f + g)^*(t_1 + t_2) \leq f^*(t_1) + g^*(t_2).$$

- (v) For  $a \in \mathbb{R}$  it holds that

$$(af)^{**} = |a| f^{**}.$$

- (vi) If  $\Omega$  is resonant (i.e., the measure on  $\Omega$  is either nonatomic or purely atomic with all atoms of equal measure), then we have for any two measurable functions  $f$  and  $g$  that

$$(f + g)^{**} \leq f^{**} + g^{**}.$$

The following proposition gives a characterization of functions in  $L \log L(\Omega)$ .

**Proposition 2.4** (see [4, Lem. IV.6.2]). *For a measurable function  $f$  on  $\Omega$ , it holds that  $f \in L \log L(\Omega)$  if and only if*

$$\int_0^1 f^*(t) \log\left(\frac{1}{t}\right) dt < \infty.$$

Integration by parts shows that

$$\int_0^1 f^*(t) \log\left(\frac{1}{t}\right) dt = \left[ \int_0^t f^*(s) ds \log\left(\frac{1}{t}\right) \right]_{t=0}^1 - \int_0^1 \int_0^t f^*(s) ds \left(-\frac{1}{t}\right) dt = \int_0^1 f^{**}(t) dt.$$

Hence, **Proposition 2.4** together with **Lemma 2.3** (v) and (vi) shows that we can define a norm on  $L \log L(\Omega)$  by

$$\|f\|_{L \log L} := \int_0^1 f^{**}(t) dt.$$

**Lemma 2.5.** *The space  $L \log L(\Omega)$  is a separable Banach space.*

*Proof.* This is contained in [4, Thm. IV.6.4 and Thm. IV.6.5]: The Lebesgue measure on  $\Omega$  is separable (in the sense of [4, Def. I.5.4]). By [4, Thm. I.5.5], it hence suffices to show that  $L \log L(\Omega)$  has absolutely continuous norm (noting that it is an order ideal by definition, see [4, Def. I.3.7]). But this is shown in [4, proof of Thm. IV.6.5].  $\square$

The dual space of  $L \log L(\Omega)$  is described in [4, Thm. IV.6.5].

**Proposition 2.6.** *If  $\Omega$  has unit measure,*

$$L \log L(\Omega)^* = L_{\exp}(\Omega) := \left\{ f : \Omega \rightarrow \mathbb{R} \text{ measurable} : \int_{\Omega} \exp(\lambda |f|) dx < \infty \text{ for some } \lambda > 0 \right\}.$$

*The space  $L_{\exp}(\Omega)$  with the norm*

$$\|f\|_{L_{\exp}} = \sup_{0 < t < 1} \frac{f^{**}(t)}{1 + \log\left(\frac{1}{t}\right)}$$

*is a Banach space. Moreover, for all  $1 < p < \infty$ , the following embeddings hold*

$$L^{\infty}(\Omega) \hookrightarrow L_{\exp}(\Omega) \hookrightarrow L^p(\Omega) \hookrightarrow L \log L(\Omega) \hookrightarrow L^1(\Omega).$$

**Lemma 2.7.**  $L \log L(\Omega)$  and  $L_{\exp}(\Omega)$  are not reflexive.

*Proof.* By [4, Cor. I.4.4] it is sufficient to show that  $L_{\exp}(\Omega)$  contains functions that do not have absolutely continuous norm. By [4, Prop. I.3.5] it is thus sufficient to provide an  $f \in L_{\exp}(\Omega)$  and a sequence  $f_n \geq 0$  with  $f \geq f_n \rightarrow 0$  almost everywhere but  $\|f_n\|_{L_{\exp}} > \varepsilon$  for some fixed  $\varepsilon > 0$ . Let  $f \in L_{\exp}(\Omega)$  be arbitrary with  $f^*(s) = -\log s$  and take  $f_n(x) = f(x)$  if  $f(x) \geq n$  and  $f_n(x) = 0$  else. Then it is straightforward to show  $\|f\|_{L_{\exp}} = \|f_n\|_{L_{\exp}} = 1$  while  $f_n \rightarrow 0$  pointwise.  $\square$

**Example 2.8.** Consider  $\mu(x) = \exp(-1/\sqrt{x})$ . It holds that  $\mu \in L \log L([0, 1])$  (since  $\mu$  is bounded) and hence that  $E(\mu) < \infty$ , but note that  $E'(\mu) = \log(\mu) + 1$  is not in  $L_{\exp}$ . To see that, define  $\alpha(x) = \log(\mu(x)) = -1/\sqrt{x}$ . Its decreasing rearrangement is  $\alpha^*(x) = 1/\sqrt{x}$  and hence

$$\alpha^{**}(t) = \frac{1}{t} \int_0^t \frac{1}{\sqrt{s}} ds = \frac{2}{\sqrt{t}}.$$

But

$$\frac{\alpha^{**}(t)}{1 + \log\left(\frac{1}{t}\right)} = \frac{2}{\sqrt{t}(1 + \log\left(\frac{1}{t}\right))}$$

is unbounded near zero, and hence the  $L_{\exp}$ -norm of  $\alpha$  is not finite.

For the next lemma, we use the elementary fact that for all  $a, b > 0$  we have  $\log^+(ab) \leq \log^+(a) + \log^+(b)$ .

**Lemma 2.9.** If  $\mu \in L \log L(\Omega_1)$ ,  $\nu \in L \log L(\Omega_2)$ , and  $\pi = \mu \otimes \nu$  (i.e.,  $\pi(x_1, x_2) := \mu(x_1)\nu(x_2)$ ), then  $\pi \in L \log L(\Omega_1 \times \Omega_2)$ .

*Proof.* We simply estimate

$$\begin{aligned} & \int_{\Omega_1 \times \Omega_2} |\mu(x_1)\nu(x_2)| \log^+ |\mu(x_1)\nu(x_2)| d(x_1, x_2) \\ & \leq \int_{\Omega_1} |\mu(x_1)| \log^+ |\mu(x_1)| dx_1 \int_{\Omega_2} |\nu(x_2)| dx_2 + \int_{\Omega_1} |\mu(x_1)| dx_1 \int_{\Omega_2} |\nu(x_2)| \log^+ |\nu(x_2)| dx_2 \end{aligned}$$

and use that all terms on the right-hand side are finite since  $L \log L(\Omega) \subset L^1(\Omega)$ .  $\square$

Alternatively,  $L \log L(\Omega)$  can be introduced as the Orlicz space with Young function  $\Phi_{\log}(s) = s \log^+ s$  and *Luxemburg norm*

$$(2.1) \quad \|f\|_{\Phi_{\log}} = \inf \left\{ \gamma > 0 : \int_{\Omega} \Phi_{\log} \left( \frac{|f|}{\gamma} \right) dx \leq 1 \right\},$$

which is an equivalent norm to  $\|f\|_{L \log L}$ ; see [4, Def. IV.8.10]. Likewise,  $L_{\exp}(\Omega)$  can be introduced as the Orlicz space with Young function

$$\Phi_{\exp}(s) = \begin{cases} s & \text{if } 0 \leq s \leq 1, \\ e^{s-1} & \text{if } s > 1, \end{cases}$$

and *Luxemburg norm*

$$(2.2) \quad \|f\|_{\Phi_{\exp}} = \inf \left\{ \gamma > 0 : \int_{\Omega} \Phi_{\exp} \left( \frac{|f|}{\gamma} \right) dx \leq 1 \right\},$$

which is equivalent to  $\|f\|_{L_{\exp}}$ .

**Lemma 2.10.** *Let  $L^\Phi(\Omega)$  denote the Orlicz space with convex Young function  $\Phi$ . If  $\|u\|_{L^\Phi} \geq 1$ , then  $\int_\Omega \Phi(|u|) dx \geq \|u\|_{L^\Phi}$ .*

*Proof.* For any  $\gamma < \|u\|_{L^\Phi}$ , it holds that  $\int_\Omega \Phi\left(\frac{|u|}{\gamma}\right) dx > 1$ . It follows from the convexity of  $\Phi$  and  $\Phi(0) = 0$  that

$$\begin{aligned} \frac{1}{\gamma} \int_\Omega \Phi(|u|) dx &= \int_\Omega \frac{1}{\gamma} \Phi(|u|) + \left(1 - \frac{1}{\gamma}\right) \Phi(0) dx \\ &\geq \int_\Omega \Phi\left(\frac{1}{\gamma}|u| + \left(1 - \frac{1}{\gamma}\right) 0\right) dx = \int_\Omega \Phi\left(\frac{|u|}{\gamma}\right) dx > 1. \end{aligned}$$

Letting  $\gamma \rightarrow \|u\|_{L^\Phi}$ , the claim follows.  $\square$

We next derive a few facts that will be useful for the analysis of the primal and dual regularized optimal transport problems. First, we consider a function  $\pi \in L \log L(\Omega_1 \times \Omega_2)$  and its pushforwards under the coordinate projections

$$(P_1)_\# \pi(x_1) = \int_{\Omega_2} \pi(x_1, x_2) dx_2, \quad (P_2)_\# \pi(x_2) = \int_{\Omega_1} \pi(x_1, x_2) dx_1.$$

The following result states that these marginals are also in  $L \log L$ .

**Lemma 2.11.** *If  $\pi \in L \log L(\Omega_1 \times \Omega_2)$ , then  $(P_i)_\# \pi \in L \log L(\Omega_i)$  for  $i \in \{1, 2\}$  with*

$$\|(P_i)_\# \pi\|_{\Phi_{\log}} \leq \max(1, \mathcal{L}(\Omega_{3-i})) \|\pi\|_{\Phi_{\log}}.$$

*Proof.* Using the convexity of  $\Phi(s) = s \log^+(s)$  and Jensen's inequality, we obtain

$$\int_{\Omega_1 \times \Omega_2} \Phi\left(\frac{\pi}{\gamma}\right) d(x_1, x_2) \geq \mathcal{L}(\Omega_1) \int_{\Omega_2} \Phi\left(\frac{1}{\mathcal{L}(\Omega_1)} \int_{\Omega_1} \frac{\pi}{\gamma} dx_1\right) dx_2 \geq \int_{\Omega_2} \Phi\left(\int_{\Omega_1} \frac{\pi}{\gamma \max(1, \mathcal{L}(\Omega_1))} dx_1\right) dx_2$$

where we used  $\ell \Phi(s/\ell) \geq \Phi(s)$  for  $\ell \leq 1$  and  $\ell \Phi(s/\ell) \geq \Phi(s/\ell)$  otherwise. Thus we obtain

$$\begin{aligned} \|\pi\|_{\Phi_{\log}} &= \min \left\{ \gamma \geq 0 : \int_{\Omega_1 \times \Omega_2} \Phi\left(\frac{\pi}{\gamma}\right) dx_1 dx_2 \leq 1 \right\} \\ &\geq \min \left\{ \gamma \geq 0 : \int_{\Omega_2} \Phi\left(\int_{\Omega_1} \frac{\pi}{\gamma \max(1, \mathcal{L}(\Omega_1))} dx_1\right) dx_2 \leq 1 \right\} \\ &= \min \left\{ \gamma \geq 0 : \int_{\Omega_2} \Phi\left(\frac{(P_2)_\# \pi}{\gamma \max(1, \mathcal{L}(\Omega_1))}\right) dx_2 \leq 1 \right\} \\ &= \frac{\|(P_2)_\# \pi\|_{\Phi_{\log}}}{\max(1, \mathcal{L}(\Omega_1))}. \end{aligned}$$

The claim for  $(P_1)_\#$  follows similarly.  $\square$

As a corollary, we obtain a characterization of  $L_{\exp}(\Omega)$  on tensor product spaces. We define

$$(\alpha \oplus \beta)(x_1, x_2) := \alpha(x_1) + \beta(x_2).$$

**Corollary 2.12.** *It holds that  $\alpha \in L_{\exp}(\Omega_1)$  and  $\beta \in L_{\exp}(\Omega_2)$  if and only if  $\alpha \oplus \beta \in L_{\exp}(\Omega_1 \times \Omega_2)$ .*

*Proof.* The mapping  $(\alpha, \beta) \mapsto \alpha \oplus \beta$  is the adjoint of  $\pi \mapsto ((P_1)_\# \pi, (P_2)_\# \pi)$ , and hence one implication follows from the fact that  $L \log L(\Omega)^* = L_{\exp}(\Omega)$ .



For the other implication, we use the Luxemburg norm and Jensen's inequality with  $\Phi \equiv \Phi_{\text{exp}}$  to observe that

$$\begin{aligned} \|\alpha \oplus \beta\|_{\Phi_{\text{exp}}} &= \min \left\{ \gamma \geq 0 : \int_{\Omega_1 \times \Omega_2} \Phi \left( \frac{\alpha(x_1) + \beta(x_2)}{\gamma} \right) dx_1 dx_2 \leq 1 \right\} \\ &\geq \min \left\{ \gamma \geq 0 : \mathcal{L}(\Omega_1) \int_{\Omega_2} \Phi \left( \frac{\frac{1}{\mathcal{L}(\Omega_1)} \int_{\Omega_1} \alpha(x_1) dx_1 + \beta(x_2)}{\gamma} \right) dx_2 \leq 1 \right\} \\ &\geq \min \left\{ \gamma \geq 0 : \int_{\Omega_2} \Phi \left( \min(1, \mathcal{L}(\Omega_1)) \frac{\frac{1}{\mathcal{L}(\Omega_1)} \int_{\Omega_1} \alpha(x_1) dx_1 + \beta(x_2)}{\gamma} \right) dx_2 \leq 1 \right\} \\ &= \min(1, \mathcal{L}(\Omega_1)) \left\| \beta + \frac{1}{\mathcal{L}(\Omega_1)} \int_{\Omega_1} \alpha dx_1 \right\|_{\Phi_{\text{exp}}}. \end{aligned}$$

This shows that  $\beta$  plus a constant is in  $L_{\text{exp}}(\Omega)$  and hence that  $\beta$  itself is in  $L_{\text{exp}}(\Omega)$ . Arguing similarly for  $\alpha$ , we obtain the claim.  $\square$

### 3 FENCHEL DUALITY IN $\mathcal{M}$ AND $C$

In this section, we study the primal and dual problems for entropically regularized mass transport, i.e.,

$$(P) \quad \inf_{\substack{\pi \in \mathcal{P}(\Omega_1 \times \Omega_2), \\ (P_1)_\# \pi = \mu, (P_2)_\# \pi = \nu}} \int_{\Omega_1 \times \Omega_2} c d\pi + \gamma \int_{\Omega_1 \times \Omega_2} \pi (\log \pi - 1) d(x_1, x_2),$$

and

$$(D) \quad \sup_{\substack{\alpha \in C(\Omega_1) \\ \beta \in C(\Omega_2)}} \int_{\Omega_1} \alpha(x_1) d\mu(x_1) + \int_{\Omega_2} \beta(x_2) d\nu(x_2) - \gamma \int_{\Omega_1 \times \Omega_2} \exp \left( \frac{-c(x_1, x_2) + \alpha(x_2) + \beta(x_1)}{\gamma} \right) d(x_1, x_2),$$

using Fenchel duality in the canonical spaces  $\mathcal{M}(\Omega_1 \times \Omega_2)$  and  $C(\Omega_1) \times C(\Omega_2)$ . We use the general framework as outlined in, e.g., [9, Sec. III.4] or [2, Chap. 9]. All throughout the article, we assume that  $\mu \in \mathcal{P}(\Omega_1)$ ,  $\nu \in \mathcal{P}(\Omega_2)$ ,  $c \in C(\Omega_1 \times \Omega_2)$ , and  $\gamma > 0$ . We further assume for the sake of simplicity that  $\Omega_1$  and  $\Omega_2$  are compact.

We begin with a strong duality result for (P) and (D). A similar result in  $L^1(\Omega)$  instead of  $\mathcal{M}(\Omega)$  is [7, Thm. 3.2], but we state the theorem and its proof because we use a slightly different setting as well as for the sake of completeness.

**Proposition 3.1 (strong duality).** *The predual problem to (P) is (D), and strong duality holds. Furthermore, if the supremum in (D) is finite, (P) admits a minimizer.*

*Proof.* First, by the Riesz–Markov representation theorem,  $\mathcal{M}(\Omega)$  is the dual space of  $C(\Omega)$  for compact  $\Omega$ . Furthermore, Slater's condition is fulfilled with  $\alpha, \beta = 0$  so that strong duality holds and – assuming a finite supremum – the primal problem (P) possesses a minimizer. In addition, the integrand of the last integral in (D) is normal so that it can be conjugated pointwise [18]. Carrying out the conjugation,

we obtain

$$\begin{aligned}
& \sup_{\substack{\alpha \in C(\Omega_1) \\ \beta \in C(\Omega_2)}} \int_{\Omega_1} \alpha \, d\mu + \int_{\Omega_2} \beta \, d\nu - \gamma \int_{\Omega_1 \times \Omega_2} \exp\left(\frac{-c(x_1, x_2) + \alpha(x_2) + \beta(x_1)}{\gamma}\right) \, d(x_1, x_2) \\
&= \sup_{\substack{\alpha \in C(\Omega_1) \\ \beta \in C(\Omega_2)}} \int_{\Omega_1} \alpha \, d\mu + \int_{\Omega_2} \beta \, d\nu + \int_{\Omega_1 \times \Omega_2} \min_{\pi \geq 0} (c(x_1, x_2) - \alpha(x_1) - \beta(x_2)) \cdot \pi(x_1, x_2) \\
&\quad + \gamma \pi(\log \pi - 1) \, d(x_1, x_2) \\
&= \sup_{\substack{\alpha \in C(\Omega_1) \\ \beta \in C(\Omega_2)}} \min_{\substack{\pi \in \mathcal{M}(\Omega_1 \times \Omega_2) \\ \pi \geq 0}} \int_{\Omega_1 \times \Omega_2} c \pi + \gamma \pi(\log \pi - 1) \, d(x_1, x_2) + \int_{\Omega_1} \alpha \, d(\mu - (P_1)_\# \pi) + \int_{\Omega_2} \beta \, d(\nu - (P_2)_\# \pi) \\
&= \min_{\substack{\pi \in \mathcal{P}(\Omega_1 \times \Omega_2) \\ (P_1)_\# \pi = \mu, (P_2)_\# \pi = \nu}} \int_{\Omega_1 \times \Omega_2} c \, d\pi + \gamma \int_{\Omega_1 \times \Omega_2} \pi(\log \pi - 1) \, d(x_1, x_2),
\end{aligned}$$

which is (P). □

**Remark 3.2.** Note that [Proposition 3.1](#) does *not* claim that the supremum is attained, i.e., that the predual problem (D) admits a solution. The proposition should also be compared to [7, Thm. 3.2], which similarly characterizes solutions under the condition that the dual problem attains a maximizer.

In addition, solutions to (D) cannot be unique since we can add and subtract constants to  $\alpha$  and  $\beta$ , respectively, without changing the functional value. On the other hand, up to such a constant, the functional in (D) is strictly concave, and therefore any solution is uniquely determined by this constant.

We can use this duality argument in combination with the results of [Section 2](#) to address the question of existence of a solution to (P).

**Theorem 3.3.** *Problem (P) admits a minimizer  $\bar{\pi}$  if and only if  $\mu \in L \log L(\Omega_1)$  and  $\nu \in L \log L(\Omega_2)$ . In this case, the minimizer is unique and lies in  $L \log L(\Omega_1 \times \Omega_2)$ .*

*Proof.* By [Proposition 2.1](#), the energy is bounded if and only if  $\bar{\pi} \in L \log L(\Omega_1 \times \Omega_2)$ . However, by [Lemma 2.11](#), this is the case only if  $\mu = (P_1)_\# \bar{\pi} \in L \log L(\Omega_1)$  and similarly for  $\nu$ . This shows that the conditions are necessary for a finite energy. For sufficiency, we first note that for  $\mu \in L \log L(\Omega_1)$  and  $\nu \in L \log L(\Omega_2)$ , the tensor product  $\pi = \mu \otimes \nu$  is a feasible candidate with finite energy by [Lemma 2.9](#). Thus, the infimum in (P) is finite, and weak duality – which always holds due to the properties of supremum and infimum – shows that the supremum in (D) is finite as well. Existence of a solution for (P) now follows from [Proposition 3.1](#).

Uniqueness and regularity of the minimizer then are a direct consequence of the strict convexity of the entropy and [Proposition 2.1](#). □

In case a minimizer exists, we can characterize its support as follows.

**Proposition 3.4.** *A minimizer  $\bar{\pi} \in L \log L(\Omega_1 \times \Omega_2)$  of (P) satisfies  $\text{supp } \bar{\pi} = \text{supp } \mu \times \text{supp } \nu$ , i.e.,*

$$\mathcal{L}(\{\bar{\pi} > 0\} \Delta (\{\mu > 0\} \times \{\nu > 0\})) = 0.$$

*Proof.* The fact that  $\text{supp } \bar{\pi} \subset \text{supp } \mu \times \text{supp } \nu$  follows from the marginal constraints and the nonnegativity of  $\bar{\pi}$ . It remains to show that  $\text{supp } \bar{\pi} \supset \text{supp } \mu \times \text{supp } \nu$ . For a contradiction, assume  $\bar{\pi} = 0$  on a set  $A \subset \text{supp } \mu \times \text{supp } \nu$  of positive measure. Since the Lebesgue measure is regular, we thus have  $\bar{\pi} = 0$  on a set  $\omega_1 \times \omega_2$  with  $\omega_1 \subset \text{supp } \mu$  and  $\omega_2 \subset \text{supp } \nu$  open. We may choose  $\omega_1, \omega_2$  small enough and  $\varepsilon > 0$  small enough such that there are  $\tilde{\omega}_1 \subset \Omega_1 \setminus \omega_1$  and  $\tilde{\omega}_2 \subset \Omega_2 \setminus \omega_2$  with  $\bar{\pi} > \varepsilon$  on  $(\tilde{\omega}_1 \times \omega_2) \cup (\omega_1 \times \tilde{\omega}_2)$ .

Let  $\kappa_i := \frac{\mathcal{L}(\omega_i)}{\mathcal{L}(\tilde{\omega}_i)}$  for  $i = 1, 2$ ,  $\kappa := \mathcal{L}(\omega_1) \cdot \mathcal{L}(\omega_2)$ , and

$$\tilde{\pi} = \bar{\pi} + t \left[ \chi_{\omega_1 \times \omega_2}(x_1, x_2) + \kappa_1 \kappa_2 \chi_{\tilde{\omega}_1 \times \tilde{\omega}_2}(x_1, x_2) - \kappa_1 \chi_{\tilde{\omega}_1 \times \omega_2}(x_1, x_2) - \kappa_2 \chi_{\omega_1 \times \tilde{\omega}_2}(x_1, x_2) \right]$$

for  $0 < t < \varepsilon / \min\{\kappa_1, \kappa_2\}$ , where  $\chi_A$  is the characteristic function of the set  $A$ . Then  $\tilde{\pi}$  is feasible. We will now argue that for small enough  $t$  we have

$$\int_M c \tilde{\pi} d(x_1, x_2) + \gamma \int_M \tilde{\pi} \log \tilde{\pi} d(x_1, x_2) \leq \int_M c \bar{\pi} d(x_1, x_2) + \gamma \int_M \bar{\pi} \log \bar{\pi} d(x_1, x_2),$$

where  $M := (\omega_1 \cup \tilde{\omega}_1) \times (\omega_2 \cup \tilde{\omega}_2)$ . Note that  $\bar{\pi} = \tilde{\pi}$  on  $\Omega_1 \times \Omega_2 \setminus M$ .

First, consider  $\int_M c \tilde{\pi} d(x_1, x_2)$ . Since  $c$  is continuous and finite,  $\int_M c \tilde{\pi} d(x_1, x_2) - \int_M c \bar{\pi} d(x_1, x_2)$  is finite and hence

$$\int_M c \tilde{\pi} d(x_1, x_2) = \int_M c \bar{\pi} d(x_1, x_2) + t C_0$$

for some constant  $C_0$ . Now, consider the entropy of  $\tilde{\pi}$ . Since  $\bar{\pi} = 0$  on  $\omega_1 \times \omega_2$ , we have

$$\int_{\omega_1 \times \omega_2} \tilde{\pi} \log \tilde{\pi} d(x_1, x_2) = \int_{\omega_1 \times \omega_2} t \log t d(x_1, x_2) = \kappa t \log t.$$

Using the inequality  $f(y) \geq f(x) + f'(x)(y - x)$  for convex  $f$ , we can estimate

$$\begin{aligned} \int_{\omega_1 \times \tilde{\omega}_2} \tilde{\pi} \log \tilde{\pi} d(x_1, x_2) &= \int_{\omega_1 \times \tilde{\omega}_2} (\bar{\pi} - \kappa_2 t) \log(\bar{\pi} - \kappa_2 t) d(x_1, x_2) \\ &\leq \int_{\omega_1 \times \tilde{\omega}_2} \bar{\pi} \log \bar{\pi} d(x_1, x_2) - \kappa_2 t \int_{\omega_1 \times \tilde{\omega}_2} \log(\bar{\pi} - \kappa_2 t) d(x_1, x_2) - \kappa t \end{aligned}$$

and similarly for  $\int_{\tilde{\omega}_1 \times \omega_2} \tilde{\pi} \log \tilde{\pi} d(x_1, x_2)$ . Again using the above inequality we have

$$\begin{aligned} \int_{\tilde{\omega}_1 \times \tilde{\omega}_2} \tilde{\pi} \log \tilde{\pi} d(x_1, x_2) &\leq \int_{\tilde{\omega}_1 \times \tilde{\omega}_2} \bar{\pi} \log \bar{\pi} d(x_1, x_2) + \kappa_1 \kappa_2 t \int_{\tilde{\omega}_1 \times \tilde{\omega}_2} \log(\bar{\pi} + \kappa_1 \kappa_2 t) d(x_1, x_2) + \kappa t \\ &\leq \int_{\tilde{\omega}_1 \times \tilde{\omega}_2} \bar{\pi} \log \bar{\pi} d(x_1, x_2) + \kappa_1 \kappa_2 t \int_{\tilde{\omega}_1 \times \tilde{\omega}_2} \bar{\pi} + t \kappa_1 \kappa_2 d(x_1, x_2). \end{aligned}$$

We obtain

$$\begin{aligned} \int_M \tilde{\pi} \log \tilde{\pi} d(x_1, x_2) - \int_M \bar{\pi} \log \bar{\pi} d(x_1, x_2) &\leq \kappa t \log t - \kappa_2 t \int_{\omega_1 \times \tilde{\omega}_2} \log(\bar{\pi} - \kappa_2 t) d(x_1, x_2) - \kappa t \\ &\quad - \kappa_1 t \int_{\tilde{\omega}_1 \times \omega_2} \log(\bar{\pi} - \kappa_1 t) d(x_1, x_2) - \kappa t + \kappa_1 \kappa_2 t \int_{\tilde{\omega}_1 \times \tilde{\omega}_2} \bar{\pi} + t \kappa_1 \kappa_2 d(x_1, x_2). \end{aligned}$$

The right-hand side is of the form  $g(t) = \kappa t \log t + h(t)$  with  $h$  differentiable at 0. We can therefore estimate

$$g(t) \leq \kappa t \log t + C_1 t = t(\kappa \log t + C_1)$$

for some  $C_1 > 0$  big enough and small  $t$ .

Putting the estimates for cost and entropy together yields

$$\begin{aligned} \int_M c \tilde{\pi} d(x_1, x_2) + \gamma \int_M \tilde{\pi} \log \tilde{\pi} d(x_1, x_2) \\ \leq \int_M c \bar{\pi} d(x_1, x_2) + \int_M \bar{\pi} \log \bar{\pi} d(x_1, x_2) + t(\gamma \kappa \log t + C_0 + \gamma C_1) \end{aligned}$$

for  $t$  small enough. However, the last term will be negative for  $t$  small enough, which shows that  $\bar{\pi}$  is not optimal in contradiction to the assumption.  $\square$

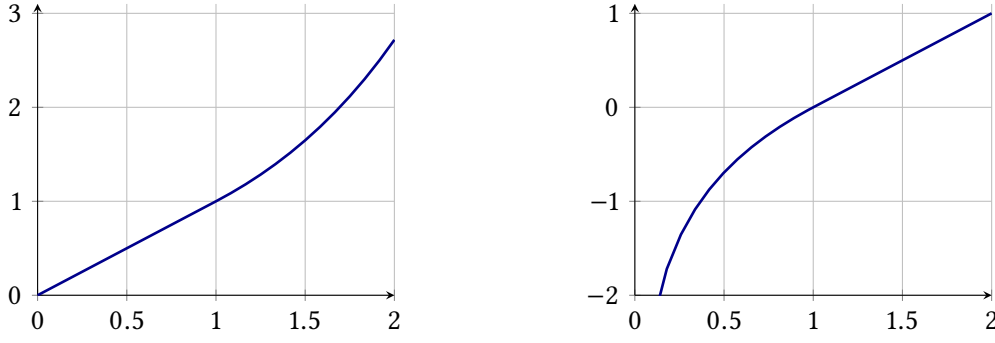


Figure 1: The functions  $\Phi$  (left) and  $\Psi = \log \circ \Phi$  (right).

**Theorem 3.3** shows that the correct setting for (P) is in fact  $L \log L(\Omega)$  rather than  $\mathcal{M}(\Omega)$ . In the next section, we will prove existence of solutions for a suitable modified dual problem of (P) and justify a pointwise almost everywhere optimality system that in turn justifies the use of the Sinkhorn algorithm.

#### 4 DUALITY IN $L \log L$ AND $L_{\text{exp}}$

In this section, we consider (P) in the space  $L \log L(\Omega_1 \times \Omega_2)$ . To derive a dual problem in  $L_{\text{exp}}(\Omega_1) \times L_{\text{exp}}(\Omega_2)$ , we shall perform the variable substitution

$$\Phi(s) = \begin{cases} \infty & \text{if } s < 0, \\ s & \text{if } s \in [0, 1], \\ e^{s-1} & \text{else,} \end{cases} \quad \text{and} \quad \Psi(s) = \log \Phi(s) = \begin{cases} -\infty & \text{if } s \leq 0, \\ \log s & \text{if } s \in (0, 1), \\ s - 1 & \text{else.} \end{cases}$$

Note that  $\Phi$  is convex and  $\Psi$  concave, see also Figure 1, and that the function  $\Phi = \Phi_{\text{exp}}$  is the Young function of  $L_{\text{exp}}$  from (2.2).

We now substitute  $e^{\alpha/\gamma} = \Phi(u_1)$  and  $e^{\beta/\gamma} = \Phi(u_2)$ , i.e.,

$$(4.1) \quad u_1 = \begin{cases} e^{\alpha/\gamma} & \text{if } \alpha \leq 0, \\ \frac{\alpha}{\gamma} + 1 & \text{else,} \end{cases} \quad u_2 = \begin{cases} e^{\beta/\gamma} & \text{if } \beta \leq 0, \\ \frac{\beta}{\gamma} + 1 & \text{else,} \end{cases}$$

which conversely implies that  $\alpha = \gamma \log(\Phi(u_1)) = \gamma \Psi(u_1)$  and  $\beta = \gamma \log(\Phi(u_2)) = \gamma \Psi(u_2)$ . Using this substitution, we obtain that

$$\begin{aligned} & \int_{\Omega_1} \alpha(x_1) d\mu(x_1) + \int_{\Omega_2} \beta(x_2) d\nu(x_2) - \gamma \int_{\Omega_1 \times \Omega_2} \exp\left(\frac{-c(x_1, x_2) + \alpha(x_1) + \beta(x_2)}{\gamma}\right) d(x_1, x_2) \\ &= -\gamma \int_{\Omega_1 \times \Omega_2} \Phi(u_1(x_1))\Phi(u_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}} d(x_1, x_2) + \gamma \int_{\Omega_1} \Psi(u_1)\mu dx_1 + \gamma \int_{\Omega_2} \Psi(u_2)\nu dx_2. \end{aligned}$$

Instead of the predual problem (D), we thus consider the reformulated problem

$$(D_{\text{exp}}) \quad \sup_{\substack{u_1 \in L_{\text{exp}}(\Omega_1) \\ u_2 \in L_{\text{exp}}(\Omega_2) \\ u_1, u_2 \geq 0}} -\gamma \int_{\Omega_1 \times \Omega_2} \Phi(u_1(x_1))\Phi(u_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}} d(x_1, x_2) + \gamma \int_{\Omega_1} \Psi(u_1)\mu dx_1 + \gamma \int_{\Omega_2} \Psi(u_2)\nu dx_2.$$

This substitution renders the problem nonconvex but, as we will see, allows to prove existence of solutions.

In the following, we assume that  $\mu, \nu \in L \log L(\Omega)$  as required for existence for the primal problem and that  $c \in C(\Omega_1 \times \Omega_2)$ . We also recall that the Luxemburg norms  $\|\cdot\|_{\Phi_{\exp}}$  and  $\|\cdot\|_{\Phi_{\log}}$  are equivalent norms on  $L_{\exp}(\Omega)$  and  $L \log L(\Omega)$ , respectively. Our aim is to apply Tonelli's direct method to  $(D_{\exp})$  by showing that the functional

$$B(u_1, u_2) := \int_{\Omega_1 \times \Omega_2} \Phi(u_1(x_1))\Phi(u_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}} d(x_1, x_2) - \int_{\Omega_1} \Psi(u_1)\mu dx_1 - \int_{\Omega_2} \Psi(u_2)\nu dx_2$$

is radially unbounded and lower semi-continuous in the right topology. We first need the following lemma.

**Lemma 4.1.** *If  $\|v\chi_{\{v>0\}}\|_{\Phi_{\exp}} > 1$ , then  $\|v\chi_{\{v>0\}}\|_{\Phi_{\exp}} \leq \log\left(\frac{1}{\mathcal{L}(\Omega)} \int_{\Omega} \Phi((v+1)\chi_{\{v>0\}}) dx\right) / \log\frac{e}{\mathcal{L}(\Omega)}$ .*

*Proof.* Set  $\gamma_\varepsilon = \|v\chi_{\{v>0\}}\|_{\Phi_{\exp}} - \varepsilon$  for an arbitrary small enough  $\varepsilon > 0$ , then by Jensen's inequality we have

$$\begin{aligned} \mathcal{L}(\Omega) \left( \frac{1}{\mathcal{L}(\Omega)} \int_{\Omega} \Phi((v+1)\chi_{\{v>0\}}) dx \right)^{\frac{1}{\gamma_\varepsilon}} &= \mathcal{L}(\Omega) \left( \frac{1}{\mathcal{L}(\Omega)} \int_{\Omega} \chi_{\{v>0\}} e^v dx \right)^{\frac{1}{\gamma_\varepsilon}} \\ &\geq \int_{\Omega} \chi_{\{v>0\}} (e^v)^{\frac{1}{\gamma_\varepsilon}} dx \\ &\geq e \int_{\Omega} \Phi\left(\frac{v\chi_{\{v>0\}}}{\gamma_\varepsilon}\right) dx > e, \end{aligned}$$

therefore  $\gamma_\varepsilon < \log\left(\frac{1}{\mathcal{L}(\Omega)} \int_{\Omega} \Phi((v+1)\chi_{\{v>0\}}) dx\right) / \log\frac{e}{\mathcal{L}(\Omega)}$ . Letting  $\varepsilon \rightarrow 0$ , the claim follows.  $\square$

We next capture the invariance inherited from (D) as described in Remark 3.2.

**Lemma 4.2.** *Let  $u_i \in L_{\exp}(\Omega_i)$ ,  $i = 1, 2$ , with  $B(u_1, u_2) < \infty$ . If for an arbitrary  $K \in \mathbb{R}$  we set  $\tilde{u}_1 = \Psi^{-1}(\Psi(u_1) - K)$  and  $\tilde{u}_2 = \Psi^{-1}(\Psi(u_1) + K)$ , then  $B(\tilde{u}_1, \tilde{u}_2) = B(u_1, u_2)$ . In particular, by choosing  $K$  appropriately, we can always achieve  $\int_{\Omega_1} \Phi(\tilde{u}_1) dx_1 = \|\tilde{u}_1\|_{\Phi_{\exp}} = 1$ .*

*Proof.* This is a direct consequence of the invariance of the cost functional in (D) under the mapping  $(\alpha, \beta) \mapsto (\alpha - K, \beta + K)$ .  $\square$

Modulo this invariance we now obtain coercivity.

**Lemma 4.3.** *Let  $u_1^n$ ,  $n = 1, 2, \dots$ , be a sequence of unit vectors in  $L_{\exp}(\Omega_1)$ , then  $\|u_2^n\|_{\Phi_{\exp}} \rightarrow \infty$  for  $n \rightarrow \infty$  implies  $B(u_1^n, u_2^n) \rightarrow \infty$  as  $n \rightarrow \infty$ .*

*Proof.* Due to  $\|u_1^n\|_{\Phi_{\exp}} = 1$  we have  $\int_{\Omega_1} \Phi(u_1^n) dx_1 = 1$  as well as

$$\int_{\Omega_1} \Psi(u_1^n) \mu dx_1 \leq \int_{\Omega_1} (u_1^n - 1)\mu dx_1 \leq \|u_1^n\|_{\Phi_{\exp}} \|\mu\|_{\Phi_{\log}} - 1 = \|\mu\|_{\Phi_{\log}} - 1.$$

Analogously we obtain

$$\int_{\Omega_2} \Psi(u_2^n) \nu dx_2 \leq \int_{\Omega_2} \max(u_2^n - 1, 0)\nu dx_2 \leq \|\max(u_2^n - 1, 0)\|_{\Phi_{\exp}} \|\nu\|_{\Phi_{\log}}.$$

Hence for  $C = \exp(\max_{\Omega_1 \times \Omega_2} c/\gamma)$  we have

$$\begin{aligned} B(u_1^n, u_2^n) &\geq C \int_{\Omega_1} \Phi(u_1^n) dx_1 \int_{\Omega_2} \Phi(u_2^n) dx_2 - \int_{\Omega_1} \Psi(u_1^n) \mu dx_1 - \int_{\Omega_2} \Psi(u_2^n) \nu dx_2 \\ &\geq C \int_{\Omega_2} \Phi(u_2^n) dx_2 - \|\mu\|_{\Phi_{\log}} + 1 - \|\max(u_2^n - 1, 0)\|_{\Phi_{\exp}} \|\nu\|_{\Phi_{\log}}. \end{aligned}$$

Since  $\|u_2^n\|_{\Phi_{\text{exp}}} \rightarrow \infty$  we also have  $\|\max(u_2^n - 1, 0)\|_{\Phi_{\text{exp}}} \rightarrow \infty$  as  $n \rightarrow \infty$  so that by Lemma 4.1 we have

$$\|\max(u_2^n - 1, 0)\|_{\Phi_{\text{exp}}} \leq \log \left( \frac{1}{\mathcal{L}(\Omega_1)} \int_{\Omega_1} \Phi(u_2^n) dx_2 \right) / \log \frac{e}{\mathcal{L}(\Omega_2)}$$

and thus

$$B(u_1^n, u_2^n) \geq C \mathcal{L}(\Omega_2) \left( \frac{e}{\mathcal{L}(\Omega_2)} \right)^{\|\max(u_2^n - 1, 0)\|_{\Phi_{\text{exp}}}} - \|\mu\|_{\Phi_{\log}} + 1 - \|\max(u_2^n - 1, 0)\|_{\Phi_{\text{exp}}} \|\mu\|_{\Phi_{\log}} \rightarrow \infty,$$

the desired contradiction.  $\square$

**Lemma 4.4.** *B is sequentially weakly-\* lower semi-continuous on  $L_{\text{exp}}(\Omega_1) \times L_{\text{exp}}(\Omega_2)$ .*

*Proof.* Let  $(u_1^n, u_2^n) \rightharpoonup^* (u_1, u_2)$  in  $L_{\text{exp}}(\Omega_1) \times L_{\text{exp}}(\Omega_2)$ . Then we have in particular  $(u_1^n, u_2^n) \rightharpoonup (u_1, u_2)$  in  $L^p(\Omega_1) \times L^p(\Omega_2)$  for any  $1 \leq p < \infty$ . Thus,  $\int_{\Omega_1} -\Psi(u_1^n) \mu dx_1$  and  $\int_{\Omega_2} -\Psi(u_2^n) \nu dx_2$  are sequentially lower semi-continuous due to the convexity of the integrands.

It remains to show lower semi-continuity of  $\int_{\Omega_1 \times \Omega_2} \Phi(u_1(x_1)) \Phi(u_2(x_2)) e^{-\frac{c(x_1, x_2)}{\gamma}} d(x_1, x_2)$ . For fixed  $N > 0$ , decompose  $\Omega_1$  and  $\Omega_2$  into a finite number of subsets  $\Omega_i^k$  of size at most  $\frac{1}{N}$  and define  $c_{kl} = \min_{(x_1, x_2) \in \Omega_1^k \times \Omega_2^l} e^{-c(x_1, x_2)/\gamma}$ . We then have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int_{\Omega_1 \times \Omega_2} \Phi(u_1^n(x_1)) \Phi(u_2^n(x_2)) e^{-\frac{c(x_1, x_2)}{\gamma}} d(x_1, x_2) &\geq \liminf_{n \rightarrow \infty} \int_{\Omega_1 \times \Omega_2} \Phi(u_1^n(x_1)) \Phi(u_2^n(x_2)) \sum_{k, l} c_{kl} \chi_{\Omega_1^k \times \Omega_2^l} d(x_1, x_2) \\ &= \liminf_{n \rightarrow \infty} \sum_{k, l} c_{kl} \int_{\Omega_1 \times \Omega_2} \Phi(u_1^n(x_1)) \Phi(u_2^n(x_2)) \chi_{\Omega_1^k}(x_1) \chi_{\Omega_2^l}(x_2) d(x_1, x_2) \\ &\geq \sum_{k, l} c_{kl} \liminf_{n \rightarrow \infty} \int_{\Omega_1} \Phi(u_1^n(x_1)) \chi_{\Omega_1^k}(x_1) dx_1 \liminf_{n \rightarrow \infty} \int_{\Omega_2} \Phi(u_2^n(x_2)) \chi_{\Omega_2^l}(x_2) dx_2 \\ &\geq \sum_{k, l} c_{kl} \int_{\Omega_1} \Phi(u_1(x_1)) \chi_{\Omega_1^k}(x_1) dx_1 \int_{\Omega_2} \Phi(u_2(x_2)) \chi_{\Omega_2^l}(x_2) dx_2 \\ &= \int_{\Omega_1 \times \Omega_2} \Phi(u_1(x_1)) \Phi(u_2(x_2)) \sum_{k, l} c_{kl} \chi_{\Omega_1^k \times \Omega_2^l}(x_1, x_2) d(x_1, x_2) \\ &\rightarrow \int_{\Omega_1 \times \Omega_2} \Phi(u_1(x_1)) \Phi(u_2(x_2)) e^{-\frac{c(x_1, x_2)}{\gamma}} d(x_1, x_2) \quad \text{as } N \rightarrow \infty \end{aligned}$$

by the monotone convergence theorem, since  $\sum_{k, l} c_{kl} \chi_{\Omega_1^k \times \Omega_2^l} \nearrow e^{-c/\gamma}$  monotonically (assuming that the decompositions  $(\Omega_1^k)_k$  and  $(\Omega_2^l)_l$  for  $N + 1$  are obtained from the decompositions for  $N$  by refinement).  $\square$

**Theorem 4.5 (dual existence).** *Problem  $(D_{\text{exp}})$  possesses a maximizer  $(\bar{u}_1, \bar{u}_2) \in L_{\text{exp}}(\Omega_1) \times L_{\text{exp}}(\Omega_2)$ .*

*Proof.* We need to show that  $B$  possesses a minimizer. The energy  $B$  is finite at, e.g.,  $u_1 \equiv 1 \equiv u_2$ . Thus we may consider a minimizing sequence  $(u_1^n, u_2^n)$  in  $L_{\text{exp}}(\Omega_1) \times L_{\text{exp}}(\Omega_2)$ , where by Lemma 4.2 we may assume  $\|u_1^n\|_{\Phi_{\text{exp}}} = 1$  without loss of generality. Lemma 4.3 now implies boundedness of  $\|u_2^n\|_{\Phi_{\text{exp}}}$  so that by the Banach–Alaoglu theorem we may extract a weakly-\* convergent subsequence from  $(u_1^n, u_2^n)$  (recall that  $L \log L(\Omega_1 \times \Omega_2)$  is separable by Lemma 2.5). The claim now follows from the lower semi-continuity of  $B$  along that subsequence by Lemma 4.4.  $\square$

From optimizers  $\bar{u}_1$  and  $\bar{u}_2$  we obtain by backsubstitution  $\bar{\alpha} := \gamma\Psi(\bar{u}_1)$  and  $\bar{\beta} := \gamma\Psi(\bar{u}_2)$  as a candidate for a solution of the original predual problem (D). However, these are in general not admissible since  $\bar{u}_1 \in L_{\text{exp}}(\Omega_1)$  and  $\bar{u}_2 \in L_{\text{exp}}(\Omega_2)$  does not imply the needed regularity of  $\bar{\alpha} \in C(\Omega_1)$  and  $\bar{\beta} \in C(\Omega_2)$ : The positive parts of  $\bar{\alpha}$  and  $\bar{\beta}$  (which equal the positive parts of  $\bar{u}_1 + 1$  and  $\bar{u}_2 + 1$ , respectively) are in  $L_{\text{exp}}$ , but the negative parts need not even be functions as they could be  $-\infty$  everywhere.

Nevertheless, from (D<sub>exp</sub>) one sees that  $\bar{u}_1 \geq 0$   $\mu$ -almost everywhere and  $\bar{u}_2 \geq 0$   $\nu$ -almost everywhere, and hence  $\bar{u}_1$  and  $\bar{u}_2$  are at least  $\mu$ - and  $\nu$ -measurable, respectively. We will derive more information on  $\bar{\alpha}$  and  $\bar{\beta}$  from the necessary optimality conditions.

First, we have again a strong duality result relating (D<sub>exp</sub>) to (P).

**Proposition 4.6 (strong duality).** *Let  $\mu \in L \log L(\Omega_1)$ ,  $\nu \in L \log L(\Omega_2)$ , and  $c \in C(\Omega_1 \times \Omega_2)$ . Then, both (P) and (D<sub>exp</sub>) admit a solution, and their optimal values coincide.*

*Proof.* Existence for both problems follows from Theorems 3.3 and 4.5. To show their equality, by Proposition 3.1 it suffices to show that the value of (D) equals that of (D<sub>exp</sub>). First, let  $\alpha \in C(\Omega_1)$  and  $\beta \in C(\Omega_2)$  be arbitrary and set  $u_1 := \Psi^{-1}(\alpha/\gamma)$  and  $u_2 := \Psi^{-1}(\beta/\gamma)$ . By substitution, we see that

$$\int_{\Omega_1} \alpha \mu \, d(x_1) + \int_{\Omega_2} \beta \nu \, d(x_2) - \gamma \int_{\Omega_1 \times \Omega_2} \exp\left(\frac{-c(x_1, x_2) + \alpha(x_2) + \beta(x_1)}{\gamma}\right) d(x_1, x_2) \leq \max_{u_1, u_2 \geq 0} -\gamma B(u_1, u_2),$$

and taking the supremum over all  $\alpha, \beta$  yields that the value of (D) is at most that of (D<sub>exp</sub>).

It thus remains to show that the value of (D<sub>exp</sub>) can be achieved by (D). Let  $\bar{u}_1, \bar{u}_2$  be optimal. By the monotone convergence theorem,  $B(\bar{u}_1, \bar{u}_2) = \lim_{n \rightarrow \infty} B(\max\{\bar{u}_1, \frac{1}{n}\}, \max\{\bar{u}_2, \frac{1}{n}\})$  and also

$$B(\max\{\bar{u}_1, \frac{1}{n}\}, \max\{\bar{u}_2, \frac{1}{n}\}) = \lim_{N \rightarrow \infty} B(\min\{\max\{\bar{u}_1, \frac{1}{n}\}, N\}, \min\{\max\{\bar{u}_2, \frac{1}{n}\}, N\}).$$

Hence  $B(\bar{u}_1, \bar{u}_2)$  can be arbitrarily well approximated by  $B(u_1, u_2)$  with  $\Psi(u_1) \in L^\infty(\Omega_1)$  and  $\Psi(u_2) \in L^\infty(\Omega_2)$ . Now let  $\alpha_n \in C(\Omega_1)$  and  $\beta_n \in C(\Omega_2)$  with  $\alpha_n \rightarrow \gamma\Psi(u_1)$  in  $L^2(\Omega_1)$  and  $\beta_n \rightarrow \gamma\Psi(u_2)$  in  $L^2(\Omega_2)$ . Here we may assume  $\alpha_n, \beta_n$  to be uniformly bounded so that (upon restricting to a subsequence) we additionally have  $\alpha_n \rightharpoonup^* \gamma\Psi(u_1)$  and  $\beta_n \rightharpoonup^* \gamma\Psi(u_2)$  in  $L^\infty(\Omega)$ . Now

$$\int_{\Omega_1} \alpha_n \mu \, dx_1 + \int_{\Omega_2} \beta_n \nu \, dx_2 \rightarrow \gamma \left[ \int_{\Omega_1} \Psi(u_1) \mu \, dx_1 + \int_{\Omega_2} \Psi(u_2) \nu \, dx_2 \right]$$

due to the weak-\* convergence, and

$$\int_{\Omega_1 \times \Omega_2} \exp\left(\frac{\alpha_n(x_1) + \beta_n(x_2) - c(x_1, x_2)}{\gamma}\right) d(x_1, x_2) \rightarrow \int_{\Omega_1 \times \Omega_2} \Phi(u_1(x_1))\Phi(u_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}} d(x_1, x_2)$$

due to the strong convergence of  $e^{\frac{\alpha_n(x_1) + \beta_n(x_2)}{\gamma}}$ . □

Having established primal and dual existence, we can now show how the solution of the dual problem can be used to solve the primal problem.

**Theorem 4.7 (optimality conditions).** *Let  $\mu \in L \log L(\Omega_1)$ ,  $\nu \in L \log L(\Omega_2)$ , and  $c \in C(\Omega_1 \times \Omega_2)$ . Then solutions  $(\bar{u}_1, \bar{u}_2) \in L_{\text{exp}}(\Omega_1) \times L_{\text{exp}}(\Omega_2)$  of (D<sub>exp</sub>) satisfy*

$$(4.2a) \quad \int_{\Omega_2} \Phi(\bar{u}_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}} dx_2 \Phi(\bar{u}_1(x_1)) = \mu(x_1),$$

$$(4.2b) \quad \int_{\Omega_1} \Phi(\bar{u}_1(x_1))e^{-\frac{c(x_1, x_2)}{\gamma}} dx_1 \Phi(\bar{u}_2(x_2)) = \nu(x_2),$$

for  $\mu$ -almost every  $x_1 \in \Omega_1$  and  $\nu$ -almost every  $x_2 \in \Omega_2$ . Furthermore,  $\bar{\pi}$  defined by

$$(4.3) \quad \bar{\pi}(x_1, x_2) = \Phi(\bar{u}_1(x_1))\Phi(\bar{u}_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}},$$

is the solution of (P).

*Proof.* Let  $\bar{u}_1, \bar{u}_2$  be solutions of the dual problem. First, note that  $\{\bar{u}_1 > 0\} \supset \{\mu > 0\}$  and  $\{\bar{u}_2 > 0\} \supset \{\nu > 0\}$  (up to a Lebesgue-negligible set) since otherwise  $\int_{\Omega_1} \Psi(\bar{u}_1)\mu \, dx_1 + \int_{\Omega_2} \Psi(\bar{u}_2)\nu \, dx_2 = -\infty$ . For any  $\varepsilon > 0$ , consider an arbitrary  $\varphi \in L_{\text{exp}}(\Omega_1) \cap L^\infty(\Omega_1)$  with  $\varphi = 0$  on  $\{\bar{u}_1 < \varepsilon\}$ . The optimality of  $\bar{u}_1$  then implies that

$$\begin{aligned} 0 &= -\frac{d}{dt}B(\bar{u}_1 + t\varphi, \bar{u}_2)|_{t=0} = \int_{\{\bar{u}_1 \geq \varepsilon\}} \Psi'(\bar{u}_1)\varphi\mu \, dx_1 \\ &\quad - \int_{\{\bar{u}_1 \geq \varepsilon\} \times \Omega_2} \Phi'(\bar{u}_1(x_1))\varphi(x_1)\Phi(\bar{u}_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}} \, d(x_1, x_2) \\ &= \int_{\{\bar{u}_1 \geq \varepsilon\}} \varphi(x_1)\Phi'(\bar{u}_1(x_1)) \left[ \frac{\mu(x_1)}{\Phi(\bar{u}_1(x_1))} - \int_{\Omega_2} \Phi(\bar{u}_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}} \, dx_2 \right] \, dx_1. \end{aligned}$$

(Note that for  $\bar{u}_1 > 0$ , both  $\Phi$  and  $\Psi$  are differentiable.) Since  $\varphi$  was arbitrary and  $\Phi'(\bar{u}_1) \neq 0$  for  $\bar{u}_1 > 0$ , the fundamental lemma of the calculus of variations then implies that

$$0 = \mu(x_1) - \Phi(\bar{u}_1(x_1)) \int_{\Omega_2} \Phi(\bar{u}_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}} \, dx_2 \quad \text{for } \mu\text{-almost all } x_1 \in \Omega_1 \text{ with } \bar{u}_1(x_1) \geq \varepsilon.$$

Since  $\varepsilon > 0$  was arbitrary and  $\mu(x_1) = 0$  whenever  $\bar{u}_1(x_1) = 0$ , the above equation even holds for  $\mu$ -almost all  $x_1 \in \Omega_1$ , which yields (4.2a). Equation (4.2b) is derived analogously.

Now we show that  $\bar{\pi}$  defined by (4.3) is a solution of the primal problem. Note that we have  $\{\bar{\pi} > 0\} = \{\bar{u}_1 > 0\} \times \{\bar{u}_2 > 0\}$ , and thus from (4.2a) and (4.2b) we obtain  $\{\bar{\pi} > 0\} = \{\mu > 0\} \times \{\nu > 0\}$ . Let now  $\tilde{\pi} \in L \log L(\Omega_1 \times \Omega_2)$  be the solution to (P), which exists by Theorem 3.3 under the assumptions on  $\mu$  and  $\nu$ . From Proposition 3.4, we know that  $\{\tilde{\pi} > 0\} = \{\mu > 0\} \times \{\nu > 0\}$  as well and hence that  $\{\tilde{\pi} = 0\} = \{\bar{\pi} = 0\}$  up to a set of Lebesgue measure zero. Using the convexity of the integrands, the fact that  $\log \pi$  is Gâteaux differentiable at  $\bar{\pi} > 0$ , and (4.3) after taking the logarithm, we can estimate

$$\begin{aligned} &\left[ \int_{\Omega_1 \times \Omega_2} c\tilde{\pi} \, d(x_1, x_2) + \gamma \int_{\Omega_1 \times \Omega_2} \tilde{\pi}(\log \tilde{\pi} - 1) \, d(x_1, x_2) \right] \\ &\quad - \left[ \int_{\Omega_1 \times \Omega_2} c\bar{\pi} \, d(x_1, x_2) + \gamma \int_{\Omega_1 \times \Omega_2} \bar{\pi}(\log \bar{\pi} - 1) \, d(x_1, x_2) \right] \\ &\quad \geq \int_{\Omega_1 \times \Omega_2} c(\tilde{\pi} - \bar{\pi}) \, d(x_1, x_2) + \gamma \int_{\{\tilde{\pi} > 0\}} \partial[\bar{\pi}(\log \bar{\pi} - 1)](\tilde{\pi} - \bar{\pi}) \, d(x_1, x_2) \\ &\quad = \int_{\{\tilde{\pi} > 0\}} (\tilde{\pi} - \bar{\pi}) [c + \gamma \log \bar{\pi}] \, d(x_1, x_2) \\ &\quad = \gamma \int_{\Omega_1 \times \Omega_2} (\tilde{\pi} - \bar{\pi}) [\Psi(\bar{u}_1(x_1)) + \Psi(\bar{u}_2(x_2))] \, d(x_1, x_2) \\ &\quad = \gamma \int_{\Omega_1} \int_{\Omega_2} \tilde{\pi} - \bar{\pi} \, dx_2 \Psi(\bar{u}_1(x_1)) \, dx_1 + \gamma \int_{\Omega_2} \int_{\Omega_1} \tilde{\pi} - \bar{\pi} \, dx_1 \Psi(\bar{u}_2(x_2)) \, dx_2. \end{aligned}$$

Both  $\tilde{\pi}$  and  $\bar{\pi}$  have the marginals  $\mu$  and  $\nu$  (by feasibility and (4.2), respectively), and hence  $\int_{\Omega_2} \tilde{\pi} - \bar{\pi} \, dx_2 = 0 = \int_{\Omega_1} \tilde{\pi} - \bar{\pi} \, dx_1$ . The last expression thus equals zero, showing that  $\bar{\pi}$  is also optimal. Since  $\tilde{\pi}$  is the unique solution to (P), we have that  $\bar{\pi} = \tilde{\pi}$  as claimed.  $\square$



**Remark 4.8.** The optimality system (4.2) is the basis of the Sinkhorn algorithm. First, note that one only needs to find  $\bar{u}_1$  and  $\bar{u}_2$  that solve (4.2a) and (4.2b); an optimal plan  $\bar{\pi}$  is then obtained from (4.3). Furthermore, (4.2a) and (4.2b) can be rewritten using the feasibility of  $\bar{\pi}$  as

$$(4.4a) \quad \int_{\Omega_2} \Phi(\bar{u}_2(x_2)) e^{-\frac{c(x_1, x_2)}{\gamma}} dx_2 \Phi(\bar{u}_1(x_1)) = \mu(x_1),$$

$$(4.4b) \quad \int_{\Omega_1} \Phi(\bar{u}_1(x_1)) e^{-\frac{c(x_1, x_2)}{\gamma}} dx_1 \Phi(\bar{u}_2(x_2)) = \nu(x_2)$$

for almost all  $(x_1, x_2) \in \Omega_1 \times \Omega_2$ . The Sinkhorn method now solves the nonlinear system (4.4) by alternatingly solving the equations: Given  $u_2^n$ , compute  $u_1^{n+1}$  by solving (4.4a), i.e., setting

$$u_1^{n+1}(x_1) = \Phi^{-1} \left( \frac{\mu(x_1)}{\int_{\Omega_2} \Phi(u_2^n(x_2)) \exp\left(-\frac{c(x_1, x_2)}{\gamma}\right) dx_2} \right),$$

and then solve (4.4b) with  $u_2^{n+1}$  to obtain

$$u_2^{n+1}(x_2) = \Phi^{-1} \left( \frac{\nu(x_2)}{\int_{\Omega_1} \Phi(u_1^{n+1}(x_1)) \exp\left(-\frac{c(x_1, x_2)}{\gamma}\right) dx_1} \right).$$

Note that the original Sinkhorn method is usually formulated directly in  $\Phi(u_1)$  and  $\Phi(u_2)$ , cf. [10, Sec. 5.3.1].

Finally, the optimality conditions (4.2a) and (4.2b) allow us to conclude which problem is solved by  $(\bar{\alpha}, \bar{\beta})$ .

**Corollary 4.9.** Let  $(\bar{u}_1, \bar{u}_2) \in L_{\text{exp}}(\Omega_1 \times \Omega_2)$  be a solution of  $(D_{\text{exp}})$ . Then  $\bar{\alpha} := \gamma\Psi(\bar{u}_1) \in L^1(\Omega_1, \mu)$  and  $\bar{\beta} := \gamma\Psi(\bar{u}_2) \in L^1(\Omega_2, \nu)$  are solutions of

$$(D_{L^1}) \quad \sup_{\substack{\alpha \in L^1(\Omega_1, \mu) \\ \beta \in L^1(\Omega_2, \nu)}} \int_{\Omega_1} \alpha d\mu + \int_{\Omega_2} \beta d\nu - \gamma \int_{\Omega_1 \times \Omega_2} \exp\left(\frac{-c(x_1, x_2) + \alpha(x_2) + \beta(x_1)}{\gamma}\right) d(x_1, x_2).$$

*Proof.* First, note that the mapping  $x_1 \mapsto \int_{\Omega_2} \Phi(\bar{u}_2(x_2)) e^{-\frac{c(x_1, x_2)}{\gamma}} dx_2$  is continuous and thus attains a minimum  $\underline{c} > 0$  and a maximum  $\bar{c} > 0$  on the (assumed to be) compact set  $\Omega_1$ . From the optimality condition (4.2a), we thus obtain that

$$\underline{c}e^{\bar{\alpha}/\gamma} = \underline{c}\Phi(\bar{u}_1) \leq \mu \leq \bar{c}\Phi(\bar{u}_1) = \bar{c}e^{\bar{\alpha}/\gamma}.$$

This implies that  $\log \mu - K \leq \bar{\alpha}/\gamma \leq \log \mu + K$  for some  $K > 0$ . We thus have

$$\frac{1}{\gamma} \int_{\Omega_1} |\bar{\alpha}| \mu dx_1 \leq K \int_{\Omega_1} \mu dx_1 + \int_{\Omega_1} |\log \mu| \mu dx_1.$$

Since  $\mu \in L \log L(\Omega_1)$ , we deduce that the right-hand side is finite and hence that  $\bar{\alpha}$  is integrable with respect to  $\mu$ , i.e.,  $\bar{\alpha} \in L^1(\Omega_1, \mu)$ . The result for  $\bar{\beta}$  follows analogously.  $\square$

**Remark 4.10.** As for (D) and as formalized in Lemma 4.2, solutions to  $(D_{\text{exp}})$  are not unique.

## 5 $\Gamma$ -LIMIT

We now turn to  $\Gamma$ -convergence of the regularized problem. Recall from, e.g., [5], that a sequence  $\{F_n\}$  of functionals  $F_n : X \rightarrow \overline{\mathbb{R}}$  on a metric space  $X$  is said to  $\Gamma$ -converge to a functional  $F : X \rightarrow \overline{\mathbb{R}}$ , written  $F = \Gamma\text{-}\lim_{n \rightarrow \infty} F_n$ , if

(i) for every sequence  $\{x_n\} \subset X$  with  $x_n \rightarrow x$ ,

$$F(x) \leq \liminf_{n \rightarrow \infty} F_n(x_n),$$

(ii) for every  $x \in X$ , there is a sequence  $\{x_n\} \subset X$  with  $x_n \rightarrow x$  and

$$F(x) \geq \limsup_{n \rightarrow \infty} F_n(x_n).$$

It is a straightforward consequence of this definition that if  $F_n$   $\Gamma$ -converges to  $F$  and  $x_n$  is a minimizer of  $F_n$  for every  $n \in \mathbb{N}$ , then every cluster point of the sequence  $\{x_n\}$  is a minimizer to  $F$ . Furthermore,  $\Gamma$ -convergence is stable under perturbations by continuous functionals.

Here we aim to approximate optimal transport plans  $\tilde{\pi}$  of the unregularized problem for marginals  $\mu$  and  $\nu$  which are *not* required to be in  $L \log L(\Omega)$ , i.e., we allow actual measures as marginals. In this case we cannot use these marginals for the regularized problems as well, since these will admit no solutions by [Theorem 3.3](#). We therefore consider smoothed marginals  $\mu_\gamma$  and  $\nu_\gamma$  in  $L \log L(\Omega)$  converging to  $\mu$  and  $\nu$ , respectively, and show that the regularized problem with these marginals  $\Gamma$ -converges to the unregularized problem with the original marginals. (The case of  $\Gamma$ -convergence for *fixed* marginals in  $L \log L(\Omega)$  has been treated in [6, Thm. 2.7].)

Let  $B$  be a smooth, compactly supported, nonnegative kernel with unit integral, and for  $\delta > 0$  and  $n \in \mathbb{N}$  set

$$B_\delta^n(x) := \frac{1}{\delta^n} B\left(\frac{x}{\delta}\right), \quad G_\delta(x_1, x_2) := B_\delta^{n_1}(x_1) B_\delta^{n_2}(x_2).$$

Since we will smooth the marginals and the transport plans by convolutions, we will need to slightly extend the domains  $\Omega_1$  and  $\Omega_2$  to avoid boundary effects. Hence, let  $\tilde{\Omega}_1$  and  $\tilde{\Omega}_2$  be compact supersets of  $\Omega_1$  and  $\Omega_2$ , respectively, such that

$$\Omega_i + \text{supp } B \subseteq \tilde{\Omega}_i, \quad i = 1, 2,$$

which are large enough to contain the supports of  $\mu_\delta := \mu * B_\delta^{n_1}$  and  $\nu_\delta := \nu * B_\delta^{n_2}$  for  $\delta \leq 1$ . (Here and in the following, we assume that the width of the convolution kernels will be small enough.) For a function or measure  $f$  on  $\Omega_1$ , we denote by  $\tilde{f}$  the extension of  $f$  to  $\tilde{\Omega}_1$  by zero (and analogously for functions and measures on  $\Omega_2$  and  $\Omega_1 \times \Omega_2$ ). Let  $\hat{c}$  be a continuous extension of  $c$  onto  $\tilde{\Omega}_1 \times \tilde{\Omega}_2$  and set

$$F_\gamma[\pi] := \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \hat{c} \, d\pi + \gamma \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \pi(\log \pi - 1) \, d(x_1, x_2),$$

$$E_\gamma^{\mu, \nu}[\pi] := \begin{cases} F_\gamma[\pi] & \text{if } \pi \in \mathcal{P}(\tilde{\Omega}_1 \times \tilde{\Omega}_2), (P_1)_\# \pi = \mu, (P_2)_\# \pi = \nu, \\ \infty & \text{else.} \end{cases}$$

Using smoothed marginals  $\mu_\delta, \nu_\delta$  and coupling  $\gamma$  and  $\delta$  in an appropriate way, we can then show  $\Gamma$ -convergence of  $E_\gamma^{\mu_\delta, \nu_\delta}$  to  $E_0^{\mu, \nu}$  as  $\gamma, \delta \rightarrow 0$ .

**Theorem 5.1.** *Let  $\mu \in \mathcal{P}(\Omega_1)$ ,  $\nu \in \mathcal{P}(\Omega_2)$ , and  $\gamma, \delta > 0$  be such that*

$$\gamma \rightarrow 0, \quad \delta \rightarrow 0, \quad \gamma \log(\delta) \rightarrow 0,$$

which is denoted in the following by  $(\gamma, \delta) \rightarrow 0$ . Define  $\mu_\delta = B_\delta^{n_1} * \tilde{\mu}$  and  $\nu_\delta = B_\delta^{n_2} * \tilde{\nu}$ . Then it holds that

$$\Gamma\text{-}\lim_{(\gamma, \delta) \rightarrow 0} E_Y^{\mu_\delta, \nu_\delta} = E_0^{\mu, \nu}$$

with respect to weak-\* convergence in  $\mathcal{M}(\tilde{\Omega}_1 \times \tilde{\Omega}_2)$ .

On the other hand, if  $\gamma, \delta \rightarrow 0$  are chosen such that  $\gamma \|\mu_\delta\|_{\Phi_{\log}} \rightarrow \infty$  or  $\gamma \|\nu_\delta\|_{\Phi_{\log}} \rightarrow \infty$ , then  $E_Y^{\mu_\delta, \nu_\delta}$  does not have a finite  $\Gamma$ -limit. More precisely, even for a family of feasible  $\pi_\delta$  (i.e., with marginals  $\mu_\delta$  and  $\nu_\delta$ ) it holds that

$$\lim_{\gamma, \delta \rightarrow 0} E_Y^{\mu_\delta, \nu_\delta}[\pi_\delta] = \infty.$$

*Proof.* For the first statement we verify the two conditions in the definition of  $\Gamma$ -convergence.

*Ad (i):* Let  $\pi_\delta \rightharpoonup^* \tilde{\pi}$ , then  $\lim_{\delta \rightarrow 0} F_0[\pi_\delta] = F_0[\tilde{\pi}]$  since  $\hat{c}$  is continuous and bounded. Since  $t(\log t - 1) \geq -1$ , we also have that

$$\int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \pi_\delta(\log \pi_\delta - 1) d(x_1, x_2) \geq -|\tilde{\Omega}_1 \times \tilde{\Omega}_2|$$

and thus,

$$F_0[\tilde{\pi}] = \lim_{\delta \rightarrow 0} F_0[\pi_\delta] - \lim_{\gamma \rightarrow 0} \gamma |\tilde{\Omega}_1 \times \tilde{\Omega}_2| \leq \liminf_{(\gamma, \delta) \rightarrow 0} F_Y[\pi_\delta].$$

Finally, the condition on the marginals is continuous with respect to weak-\* convergence of  $\pi_\delta, \mu_\delta$ , and  $\nu_\delta$  (note that  $\mu_\delta, \nu_\delta \rightharpoonup^* \tilde{\mu}, \tilde{\nu}$ ).

*Ad (ii):* It suffices to consider a recovery sequence for  $\pi \in \mathcal{P}(\Omega_1 \times \Omega_2)$ , because the marginal conditions for  $\mu$  and  $\nu$  can never be satisfied for  $\pi \in \mathcal{P}(\tilde{\Omega}_1 \times \tilde{\Omega}_2) \setminus \mathcal{P}(\Omega_1 \times \Omega_2)$ . If  $E_0^{\mu, \nu}[\tilde{\pi}] = \infty$ , then the lim sup condition holds trivially. Let therefore  $E_0^{\mu, \nu}[\tilde{\pi}]$  be finite. We set  $\pi_\delta := G_\delta * \tilde{\pi}$ . Then  $\pi_\delta \rightharpoonup^* \tilde{\pi}$  as well as  $(P_1)_\# \pi_\delta = \mu_\delta$ ,  $(P_2)_\# \pi_\delta = \nu_\delta$ . Since by Young's convolution inequality  $\pi_\delta \leq \|G_\delta\|_{L^\infty} \|\tilde{\pi}\|_{L^1} \leq \frac{C}{\delta^N}$  for some constant  $C > 0$  and  $N := n_1 + n_2$  and we have

$$\int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \pi_\delta d(x_1, x_2) = \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \pi * G_\delta d(x_1, x_2) = \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \pi d(x_1, x_2) \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} G_\delta d(x_1, x_2) = 1,$$

we conclude that

$$\gamma \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \pi_\delta(\log \pi_\delta - 1) d(x_1, x_2) \leq \gamma \left( \log \left( \frac{C}{\delta^N} \right) - 1 \right) \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \pi_\delta d(x_1, x_2) = -\gamma(1 + N \log \delta - \log C).$$

The right-hand side vanishes for  $(\gamma, \delta) \rightarrow 0$  by the assumption on the (coupled) convergences of  $\gamma$  and  $\delta$ . Hence,

$$E_0^{\mu, \nu}[\tilde{\pi}] = \lim_{(\gamma, \delta) \rightarrow 0} [F_0[\pi_\delta] - \gamma(1 + N \log \delta - \log C)] \geq \limsup_{(\gamma, \delta) \rightarrow 0} F_Y[\pi_\delta].$$

For the second statement, recall from [Lemma 2.11](#) that

$$\gamma \|\mu_\delta\|_{\Phi_{\log}} \leq \max(1, \mathcal{L}(\Omega_1)) \gamma \|\pi_\delta\|_{\Phi_{\log}}, \quad \gamma \|\nu_\delta\|_{\Phi_{\log}} \leq \max(1, \mathcal{L}(\Omega_2)) \gamma \|\pi_\delta\|_{\Phi_{\log}}.$$

By [Lemma 2.10](#), this immediately yields  $F_Y[\pi_\delta] \rightarrow \infty$ , and the assertion follows.  $\square$

The conditions on  $\gamma$  and  $\delta$  are in particular satisfied for  $\delta = c\gamma$  for some  $c > 0$ .

## 6 CONCLUSION

In contrast to the original Kantorovich formulation of optimal transport problems, their entropic regularization is well-posed only for marginals with finite entropy. Restricting the regularized problem to such functions and applying Fenchel duality in the space  $L \log L(\Omega)$  allows deriving primal-dual optimality conditions that can be interpreted pointwise almost everywhere and used to derive a continuous version of the popular Sinkhorn algorithm. For marginals that do not have finite entropy, a combined regularization and smoothing approach leads to a family of well-posed approximations that  $\Gamma$ -converge to the original Kantorovich formulation if the regularization and smoothing parameters are coupled in an appropriate way.

This work can be extended in several directions. For example, we have considered the usual setting where the entropic penalty is taken with respect to Lebesgue density, but other choices (such as the product measure of the marginals) are possible and may lead to well-posedness and duality for a larger class of marginals. Naturally, a challenging but worthwhile issue would be a convergence analysis of the Sinkhorn algorithm in the considered Orlicz spaces  $L \log L(\Omega)$  and  $L_{\exp}(\Omega)$ .

## REFERENCES

- [1] L. Ambrosio and N. Gigli, A user's guide to optimal transport, in *Modelling and Optimisation of Flows on Networks*, Springer, 2013, 1–155, [doi:10.1007/978-3-642-32160-3\\_1](https://doi.org/10.1007/978-3-642-32160-3_1).
- [2] H. Attouch, G. Buttazzo, and G. Michaille, *Variational Analysis in Sobolev and BV Spaces*, volume 6 of MPS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, PA, 2006, [doi:10.1137/1.9781611973488](https://doi.org/10.1137/1.9781611973488).
- [3] J. D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, Iterative Bregman Projections for Regularized Transportation Problems, *SIAM Journal on Scientific Computing* 37 (2015), A1111–A1138, [doi:10.1137/141000439](https://doi.org/10.1137/141000439).
- [4] C. Bennett and R. Sharpley, *Interpolation of Operators*, volume 129 of Pure and Applied Mathematics, Academic Press, Inc., Boston, MA, 1988, [doi:10.1016/s0079-8169\(13\)62909-8](https://doi.org/10.1016/s0079-8169(13)62909-8).
- [5] A. Braides,  *$\Gamma$ -convergence for Beginners*, volume 22 of Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, Oxford, 2002, [doi:10.1093/acprof:oso/9780198507840.001.0001](https://doi.org/10.1093/acprof:oso/9780198507840.001.0001).
- [6] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer, Convergence of entropic schemes for optimal transport and gradient flows, *SIAM Journal on Mathematical Analysis* 49 (2017), 1385–1418, [doi:10.1137/15m1050264](https://doi.org/10.1137/15m1050264).
- [7] L. Chizat, G. Peyré, B. Schmitzer, and F. X. Vialard, Scaling algorithms for unbalanced optimal transport problems, *Math. Comp.* 87 (2018), 2563–2609, [doi:10.1090/mcom/3303](https://doi.org/10.1090/mcom/3303).
- [8] M. Cuturi, Sinkhorn distances: Lightspeed computation of optimal transport, in *Advances in Neural Information Processing Systems (NIPS)*, volume 26, Curran Associates Inc., USA, 2013, 2292–2300, [arXiv:1306.0895](https://arxiv.org/abs/1306.0895).
- [9] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, volume 28 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999, [doi:10.1137/1.9781611971088](https://doi.org/10.1137/1.9781611971088).

- [10] M. Essid and M. Pavon, Traversing the Schrödinger Bridge Strait: Robert Fortet’s Marvelous Proof Redux, *Journal of Optimization Theory and Applications* 181 (2019), 23–60, [doi:10.1007/s10957-018-1436-9](https://doi.org/10.1007/s10957-018-1436-9).
- [11] R. Fortet, Résolution d’un système d’équations de M. Schrödinger, *J. Math. Pure Appl.* IX 1 (1940), 83–105.
- [12] L. Kantorovitch, On the translocation of masses, *C. R. (Doklady) Acad. Sci. URSS (N.S.)* 37 (1942), 199–201, [doi:10.1287/mnsc.5.1.1](https://doi.org/10.1287/mnsc.5.1.1).
- [13] P. A. Knight, The Sinkhorn–Knopp Algorithm: Convergence and Applications, *SIAM Journal on Matrix Analysis and Applications* 30 (2008), 261–275, [doi:10.1137/060659624](https://doi.org/10.1137/060659624).
- [14] C. Léonard, A survey of the Schrödinger problem and some of its connections with optimal transport, *Discrete Contin. Dyn. Syst.* 34 (2014), 1533–1574, [doi:10.3934/dcds.2014.34.1533](https://doi.org/10.3934/dcds.2014.34.1533).
- [15] D. A. Lorenz, P. Manns, and C. Meyer, Quadratically regularized optimal transport, *arXiv* (2019), [arXiv:1903.01112](https://arxiv.org/abs/1903.01112).
- [16] I. Navrotskaya and P. J. Rabier,  $L \log L$  and finite entropy, *Adv. Nonlinear Anal.* 2 (2013), 379–387, [doi:10.1515/anona-2013-0018](https://doi.org/10.1515/anona-2013-0018).
- [17] M. M. Rao and Z. D. Ren, *Theory of Orlicz Spaces*, Pure and Applied Mathematics, Dekker, 1991.
- [18] R. T. Rockafellar, Integrals which are convex functionals, *Pacific J. Math.* 24 (1968), 525–539, [doi:10.2140/pjm.1968.24.525](https://doi.org/10.2140/pjm.1968.24.525).
- [19] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, volume 87 of Progress in Nonlinear Differential Equations and their Applications, Birkhäuser/Springer, Cham, 2015, [doi:10.1007/978-3-319-20828-2](https://doi.org/10.1007/978-3-319-20828-2).
- [20] B. Simon, *Convexity: An Analytic Viewpoint*, volume 187, Cambridge University Press, 2011, [doi:10.1017/cb09780511910135](https://doi.org/10.1017/cb09780511910135).
- [21] C. Villani, *Optimal Transport. Old and New*, volume 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 2009, [doi:10.1007/978-3-540-71050-9](https://doi.org/10.1007/978-3-540-71050-9).