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 with respect to the Lebesgue measure, which has attracted attention because it can be solved by the very simple Sinkhorn algorithm. We first analyze the regularized 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. We then show that the primal problem is naturally analyzed in the Orlicz space of functions with finite entropy in the sense that the entropically regularized problem admits a minimizer if and only if the marginals have finite entropy. We then derive a dual problem in the corresponding dual space, for which existence can be shown by purely variational arguments 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 [22]. Specifically, let Ω_1 and Ω_2 be two compact subset of $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, respectively. For given probability measures μ on Ω_1 and ν on Ω_2 and a continuous cost function $c : \Omega_1 \times \Omega_2 \to [0, \infty)$, the goal is to find a measure π on $\Omega_1 \times \Omega_2$ such that the cost $\int_{\Omega_1 \times \Omega_2} c \, d\pi$ is minimal among all π that have μ and ν as marginals. This problem has been well studied, and we refer to the recent books [34, 32] for an overview. For example, it is known that the problem has a solution π and that the support of π is contained in the so-called c-superdifferential of a c-concave function on Ω_1, see [2, Thm. 1.13]. (This is sometimes called the *fundamental theorem of optimal transport.*) In the case where Ω_1 and Ω_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 optimal plans π are singular with respect to the Lebesgue measure. Hence, the optimal plan is not a measurable function, and so standard approximation techniques from numerical analysis (e.g. by piecewise constant or piecewise linear functions) are not applicable. This motivates the use of regularization of the continuous problem to obtain approximate solutions that are

[*]Faculty of Mathematics, University of Duisburg-Essen, 45117 Essen, Germany (christian.clason@uni-due.de, ORCID: 0000-0002-9948-8426)

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

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

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

functions instead of measures, which in turn can be treated by classical discretization techniques in order to solve the regularized problem.

In this work we focus on *entropic regularization* by adding a multiple of the negative entropy of π (with respect to the Lebesgue measure) to the objective function. This forces the optimal plan to be a measure that has a density with respect to the Lebesgue measure. Furthermore, in the discrete setting, this allows to solve the problem numerically by the very simple Sinkhorn algorithm [23, 17, 4].

Notation and problem statement.   To fully state the regularized optimal transport problem, we introduce some notation. By $\mathcal{M}(\Omega)$ 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 $\mathcal{L}$ (the set on which it is defined being clear from the context), and integrals with respect to the Lebesgue measure are simply denoted by dx with the appropriate integration variable x. We write $L^p(\Omega, d\mu)$ for the space of p-integrable functions with respect to the measure μ but omit the set Ω if it is clear from the context. If no measure is given, L^p always refers to the space with respect to the Lebesgue measure. In the case where the measure π has a density with respect to the Lebesgue measure, we will also use π for that density. For $\mu \in \mathcal{M}(\Omega_1)$ and $g : \Omega_1 \to \Omega_2$, we denote by $g_\# \pi$ the pushforward of μ by g, i.e., the measure on Ω_2 defined by $g_\# \pi(B) = \pi(g^{-1}(B))$ for all measurable sets $B \subset \Omega_2$. In particular, we will use the coordinate projections $P_i : \Omega_1 \times \Omega_2 \to \Omega_i$, $P_i(x_1, x_2) = x_i$, and the fact that $P_{i\#}\pi$ is the ith marginal of $\pi \in \mathcal{M}(\Omega_1 \times \Omega_2)$. The entropically regularized Kantorovich problem of optimal mass transport between $\mu \in \mathcal{P}(\Omega_1)$ and $\nu \in \mathcal{P}(\Omega_2)$ is then given by

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

(Note that we used the negative entropy of π with respect to the Lebesgue measure for regularization. One could also consider regularization by adding $\gamma \int_{\Omega_1 \times \Omega_2} \pi(\log \pi - 1) \, d\theta$ for some other measure θ, e.g., the product measure $\mu \otimes \nu$ [18], but we will not pursue this further.) A purely formal application of convex duality then yields the predual problem

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

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. As it turns out, neither (P) nor (D) may admit a solution in the considered spaces. As we will show, it is necessary and sufficient to obtain existence of a primal solution 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 by purely variational methods. For marginals that are not in $L \log L$, we show Γ-convergence of minimizers of regularized problems with suitably smoothed marginals.

Related work.   The continuous optimal transport problem has been analyzed in the survey paper [24] where the relation to the so-called dynamic Schrödinger problem is made. Another survey [20] 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 [21]). A detailed overview of the connections between optimal transport, the Schrödinger problem, and the Sinkhorn algorithm from a stochastic control viewpoint is given in the even more recent survey

[14]. In [11], primal existence has been shown in the subset of the space of measures which have a density of finite entropy with respect to the Lebesgue measure. Furthermore, [15] analyzes the problem (for unbalanced transport, i.e., for marginals with different mass) in L^1 and derives a dual formulation in L^∞. However, the question of existence of a solution of the respective dual problem is not answered. In [9], this gap was closed through a contraction argument using the Hilbert metric. More precisely, [9, Thm. 3.1] guarantees the existence of dual solutions in L^∞ provided that the feasible set of the dual problem is not empty. Moreover, if a certain constraint qualification holds, the dual optimizers x and y can be shown to satisfy $x(s) + y(t) = \log u_0(s,t)$ a.e., where u_0 denotes the optimal primal solution. Here u_0, x, and y correspond to $\bar\pi$, α, β of (P) and (D) via $u_0 = e^{\frac{c}{\gamma}}\bar\pi$, $x = \alpha/\gamma$, and $y = \beta/\gamma$. A similar result is also stated in the more recent work [13, Thm. 6], which shows the existence of dual optimizers if the marginals are absolutely continuous probability measures. (The relation of q and v in [13] to our notation is $q = e^{\frac{-c}{\gamma}}$ and $v = \bar\pi e^{\frac{c}{\gamma}}$.) Another approach to prove the existence of unique solutions (even in the multi-marginal case) is presented in [12, Thm 4.3]. The authors show that a certain map is a bijection, which yields existence of dual solutions α and β in L^∞ if the marginals are functions in L^∞ as well. Moreover, in [6] a compactness argument is used to show the existence of a fixed point of the Sinkhorn iteration; in contrast to our work, the entropy penalization there is considered with respect to the product measure of the marginals.

Previous works [6, 9, 12, 13] tackle the problem of existence of dual solutions under various conditions in standard Lebesgue spaces. For marginals of finite entropy, [16, Cor. 3.2] already states that dual solutions exist and satisfy $\bar\alpha \in L^1(\Omega_1,\mu)$ and $\bar\beta \in L^1(\Omega_2,v)$ (in our notation; the notation there uses $Q = \bar\pi e^{\frac{-c}{\gamma}}$, $P_1 = \mu$, $a = e^{\frac{-\alpha}{\gamma}}$, $R = e^{\frac{-c}{\gamma}}\mathcal{L}$, $P_2 = v$, and $b = e^{\frac{-\beta}{\gamma}}$). Note that while the primal solution Q is in $L\log L(\Omega,\mathcal{L})$, the analysis takes place in $L\log L(\Omega, e^{\frac{-c}{\gamma}}\mathcal{L})$. Moreover, as the authors of [9] note, [16] fails to elaborate a crucial step of the argumentation. This gap was closed only later in [8]. None of the mentioned works considered necessary conditions for existence. Finally, [25] analyzes regularization with the L^2 norm of π and derives existence of solutions of the dual problem.

The notion of Orlicz spaces in the context of convex integral functionals has previously been used in [26], where existence of both primal and dual optimizers are covered in a more general setting. More precisely, the spaces used in [26], which are also known as Musielak–Orlicz spaces [28], are a generalization of the Orlicz spaces used here. The setting considered here can be recovered in two different ways: In section 7.3 (a), the above referenced results of [16] are recovered as a special case (where again our case corresponds to choosing $R = e^{\frac{-c}{\gamma}}\mathcal{L}$). Moreover, choosing $m(z) = e^{\frac{-c(z)}{\gamma}}$ in the second example (titled *a variant of the Boltzmann entropy*) in section 7.1 gives a problem very similar to the one considered here. The difference lies in the fact that the cost function c is part of the definition of the relevant Musielak–Orlicz spaces in this case, and hence the analysis takes place in different spaces. As the aim of [26] is to weaken the necessary assumptions as much as possible, the overall setting is more abstract, and the proofs rely heavily on the authors previous work [27]. Here we aim for a self-contained, more elementary, treatment of (P).

Regarding Γ-convergence, the limit for $\gamma \to 0$ and fixed marginals with densities with finite entropy was considered recently in [11].

Organization.   The next Section 2 recalls statements about functions of finite entropy and the duality of the respective Orlicz space $L\log L$. In Section 3, we collect and prove (for the sake of completeness) results on the regularized optimal transport problem (P) in the context of duality of continuous functions and measures. In particular, Theorem 3.3 shows that primal solutions exist if and only if the marginals are in the space $L\log L$. Hence, we analyze the problem in Section 4 in the context of $L\log L$ and $L_{\exp}$. We show existence and uniqueness of the primal problem in $L\log L$, derive the dual problem and show existence of solutions for the dual problem in $L_{\exp}$. We finally show a result on Γ-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 [30, 5, 1]; see also [33]. We consider a compact domain $\Omega \subset \mathbb{R}^n$ and denote the *neg-entropy* of a measurable function $f : \Omega \to \mathbb{R}$ by

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

where we set $0 \log 0 = 0$ as usual. Note that since $s \log s \geq -1/e$ for every $s \geq 0$, the neg-entropy always lies in the interval $[-\mathcal{L}(\Omega)/e, \infty]$. We say that f has finite entropy if $E(f) < \infty$. Following [5], we define

$$L \log L(\Omega) := \left\{ f : \Omega \to \mathbb{R} \text{ measurable} : \int_\Omega |f(x)| \log^+(|f(x)|) \, dx < \infty \right\},$$

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

Proposition 2.1 ([29, 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 recall the central constructions and main results based on so-called Young functions.

Definition 2.2 (Young functions). Let $\varphi : [0, \infty) \to [0, \infty]$ be increasing and lower semi-continuous with $\varphi(0) = 0$. Suppose that φ is neither identically zero nor identically infinite on $(0, \infty)$. Then the function Φ, defined by

$$\Phi(t) := \int_0^t \varphi(s) \, ds,$$

is said to be a *Young function*. Moreover, the function Ψ defined by

$$\Psi(s) := \max_{t \geq 0} \{ st - \Phi(t) \}$$

is called the *complementary Young function* of Φ.

Any Young function is continuous and convex on its domain, and the complementary Young function Ψ is again a Young function. The notion of Young functions gives rise to a generalization of L^p spaces through the definition of the so-called Luxemburg norm.

Definition 2.3 (Luxemburg norm and Orlicz spaces). Let Φ be a Young function. The Luxemburg norm of a measurable function $f : \Omega \to \mathbb{R}$ is defined as

$$(2.1) \qquad \|f\|_\Phi = \inf \left\{ \gamma > 0 : \int_\Omega \Phi\left(\frac{|f|}{\gamma} \right) dx \leq 1 \right\}.$$

The space of all measurable functions with finite Luxemburg norm is called *Orlicz space* and denoted by $L^\Phi(\Omega)$.

Remark 2.4. General Orlicz norms do not scale in a simple way with the size of the set Ω. Writing $\mathbb{1}_A$ for the characteristic function of the set $A \subset \Omega$, i.e., $\mathbb{1}_A(x) = 1$ if $x \in A$ and 0 else, the p-norm (corresponding to the Young function $\Phi(t) = t^p$) of $\mathbb{1}_\Omega$ equals $\|\mathbb{1}_\Omega\|_p = \mathcal{L}(\Omega)^{1/p}$. For a strictly increasing Young function Φ, we obtain the more complicated result $\|\mathbb{1}_\Omega\|_\Phi = (\Phi^{-1}(\mathcal{L}(\Omega)^{-1}))^{-1}$. As a consequence, some results in the following depend on the size of the domain. One could get rid of this dependence by adapting the definition of the norm to, e.g.,

$$\|f\|_\Phi = \inf \left\{ \gamma > 0 : \frac{1}{\mathcal{L}(\Omega)} \int_\Omega \Phi\left(\frac{|f|}{\gamma} \right) dx \leq 1 \right\}.$$

However, since this definition would be nonstandard, we refrain from doing so.

Moreover, note that

$$\int_\Omega \Phi\left(\frac{|f|}{\|f\|_\Phi}\right) \mathrm{d}x \le 1$$

is always true, but equality may fail to hold. For a counterexample, see, e.g., [33, Example 2.8].

Theorem 2.5 ([1, Thm. 8.10]). *$L^\Phi(\Omega)$ is a Banach space with respect to the Luxemburg norm.*

We will also need the following estimate.

Lemma 2.6. *Let $L^\Phi(\Omega)$ denote the Orlicz space with convex Young function Φ and $u \in L^\Phi(\Omega)$ with $\|u\|_\Phi > 1$. Then $\int_\Omega \Phi(|u|) \, \mathrm{d}x \ge \|u\|_\Phi$.*

Proof. For any $1 \le \gamma < \|u\|_\Phi$, it holds that $\int_\Omega \Phi\big(\frac{|u|}{\gamma}\big) \, \mathrm{d}x > 1$. It then follows from the convexity of Φ and $\Phi(0) = 0$ that

$$\begin{aligned}
\frac{1}{\gamma}\int_\Omega \Phi(|u|) \, \mathrm{d}x &= \int_\Omega \tfrac{1}{\gamma}\Phi(|u|) + \left(1 - \tfrac{1}{\gamma}\right)\Phi(0) \, \mathrm{d}x \\
&\ge \int_\Omega \Phi\left(\tfrac{1}{\gamma}|u| + \left(1 - \tfrac{1}{\gamma}\right)0\right) \mathrm{d}x = \int_\Omega \Phi\left(\frac{|u|}{\gamma}\right) \mathrm{d}x > 1.
\end{aligned}$$

Letting $\gamma \to \|u\|_\Phi$, the claim follows. $\qquad\square$

Note that by Remark 2.4, Lemma 2.6 does *not* hold for $\|u\|_\Phi = 1$.

Using $\Phi_{\log}(s) = s \log^+ s$ as Young function now immediately yields $L \log L(\Omega) = L^{\Phi_{\log}}(\Omega)$. The complementary Young function

$$(2.2) \qquad\qquad \Phi_{\exp}(s) = \begin{cases} s & \text{if } 0 < s \le 1, \\ e^{s-1} & \text{if } s > 1 \end{cases}$$

of $\Phi_{\log}$ now provides a natural way to define the Orlicz space $L^{\Phi_{\exp}}(\Omega) =: L_{\exp}(\Omega)$. In fact, $L_{\exp}(\Omega)$ is the dual space of $L \log L(\Omega)$.

Proposition 2.7 ([5, Thm. IV.6.5]). *If Ω has finite Lebesgue measure, then $L \log L(\Omega)^* = L_{\exp}(\Omega)$ (up to equivalence of norms). Moreover, for all $1 < p < \infty$, the following embeddings hold*

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

The Luxemburg norms (2.1) on $L^\Phi(\Omega)$ are equivalent to the norms defined in [5, Def. IV.6.3] (in [5, Def. IV.6.3], the norms for $L \log L(\Omega)$ and $L_{\exp}(\Omega)$ are dual to each other). The constants in this norm equivalence will in the following generically be denoted by c_Φ. Note that [5, Thm. IV.6.5] is stated for domains with unit Lebesgue measure, but the case of general finite measure follows by a simple rescaling.

We also have the following properties, which follow from Theorem 8.21 b and Theorem 8.19 in [1], respectively, by observing that $\Phi_{\log}$ is so-called Δ-regular (c.f. [1, Def. 8.7]) but $\Phi_{\exp}$ is not.

Lemma 2.8.

  (i) *The space $L \log L(\Omega)$ is separable.*

  (ii) *The spaces $L \log L(\Omega)$ and $L_{\exp}(\Omega)$ are not reflexive.*

The following example shows that the desired optimality conditions cannot be derived by simply setting the Gâteaux derivative to zero.

Example 2.9. E is not Gâteaux-differentiable on $L \log L((0,1))$. Indeed, consider $f(x) = \exp(-1/\sqrt{x})$. Then it holds that $f \in L \log L((0,1))$ (since f is bounded) and hence that $E(f) < \infty$, but note that the formal Gâteaux derivative $E'(f) = \log(f) + 1$ is not in $L_{\exp}((0,1))$. To see this, note that $\log(f) + 1 = 1 - \frac{1}{\sqrt{x}}$ is not in $L^2((0,1))$ and thus by Proposition 2.7 is not in $L_{\exp}((0,1))$.

We next derive a few facts that will be useful for the analysis of the primal and dual regularized optimal transport problems. For the first lemma, we use the elementary fact that for all $a, b > 0$ we have $\log^+(ab) \le \log^+(a) + \log^+(b)$.

Lemma 2.10. *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

$$\int_{\Omega_1 \times \Omega_2} |\mu(x_1)\nu(x_2)| \log^+ |\mu(x_1)\nu(x_2)| \, \mathrm{d}(x_1, x_2)$$

$$\le \int_{\Omega_1} |\mu(x_1)| \log^+ |\mu(x_1)| \, \mathrm{d}x_1 \int_{\Omega_2} |\nu(x_2)| \, \mathrm{d}x_2 + \int_{\Omega_1} |\mu(x_1)| \, \mathrm{d}x_1 \int_{\Omega_2} |\nu(x_2)| \log^+ |\nu(x_2)| \, \mathrm{d}x_2$$

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

Next, we consider a function $\pi \in L \log L(\Omega_1 \times \Omega_2)$ and its pushforwards under the coordinate projections

$$(P_1)_\# \pi(x_1) = \int_{\Omega_2} \pi(x_1, x_2) \, \mathrm{d}x_2, \qquad (P_2)_\# \pi(x_2) = \int_{\Omega_1} \pi(x_1, x_2) \, \mathrm{d}x_1.$$

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*

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

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

$$\int_{\Omega_1 \times \Omega_2} \Phi\Big(\frac{|\pi|}{\gamma}\Big) \, \mathrm{d}(x_1, x_2) \ge \mathcal{L}(\Omega_1) \int_{\Omega_2} \Phi\Big(\frac{1}{\mathcal{L}(\Omega_1)} \int_{\Omega_1} \frac{|\pi|}{\gamma} \, \mathrm{d}x_1\Big) \, \mathrm{d}x_2 \ge \int_{\Omega_2} \Phi\Big(\int_{\Omega_1} \frac{|\pi|}{\gamma \max(1, \mathcal{L}(\Omega_1))} \, \mathrm{d}x_1\Big) \, \mathrm{d}x_2$$

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

$$\|\pi\|_{\Phi_{\log}} = \min\left\{\gamma \ge 0 : \int_{\Omega_1 \times \Omega_2} \Phi\Big(\frac{|\pi|}{\gamma}\Big) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \le 1\right\}$$

$$\ge \min\left\{\gamma \ge 0 : \int_{\Omega_2} \Phi\Big(\int_{\Omega_1} \frac{|\pi|}{\gamma \max(1, \mathcal{L}(\Omega_1))} \, \mathrm{d}x_1\Big) \, \mathrm{d}x_2 \le 1\right\}$$

$$= \min\left\{\gamma \ge 0 : \int_{\Omega_2} \Phi\Big(\frac{(P_2)_\# |\pi|}{\gamma \max(1, \mathcal{L}(\Omega_1))}\Big) \, \mathrm{d}x_2 \le 1\right\}$$

$$\ge \min\left\{\gamma \ge 0 : \int_{\Omega_2} \Phi\Big(\frac{(P_2)_\# \pi}{\gamma \max(1, \mathcal{L}(\Omega_1))}\Big) \, \mathrm{d}x_2 \le 1\right\}$$

$$= \frac{\|(P_2)_\# \pi\|_{\Phi_{\log}}}{\max(1, \mathcal{L}(\Omega_1))}.$$

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

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

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

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

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 $\Phi \equiv \Phi_{\exp}$ to observe that

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

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

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.,

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

and

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

using Fenchel duality in the canonical spaces $\mathcal{M}(\Omega_1 \times \Omega_2)$ and $C(\Omega_1) \times C(\Omega_2)$. Most of the results in this section are classical [16, 9], but we include the results with proofs for the sake of completeness.

We use the general framework as outlined in, e.g., [19, Sec. III.4] or [3, Chap. 9]. All throughout the following, we assume that $\mu \in \mathcal{P}(\Omega_1)$, $\nu \in \mathcal{P}(\Omega_2)$, $c \in C(\Omega_1 \times \Omega_2)$, $\gamma > 0$, and that Ω_1 and Ω_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 [15, Thm. 3.2], but we state the theorem and its proof because we use a slightly different setting.

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 Ω. 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 [31]. Carrying out the conjugation,

we obtain

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

$$= \sup_{\substack{\alpha \in C(\Omega_1) \\ \beta \in C(\Omega_2)}} \int_{\Omega_1} \alpha \, d\mu + \int_{\Omega_2} \beta \, d\nu + \int_{\Omega_1 \times \Omega_2} \min_{\pi \geq 0}(c(x_1,x_2) - \alpha(x_1) - \beta(x_2))\pi(x_1,x_2) + \gamma\pi(\log\pi - 1) \, d(x_1,x_2)$$

$$= \sup_{\substack{\alpha \in C(\Omega_1) \\ \beta \in C(\Omega_2)}} \min_{\substack{\pi \in \mathcal{M}(\Omega_1 \times \Omega_2) \\ \pi \geq 0}} \int_{\Omega_1 \times \Omega_2} c\pi + \gamma\pi(\log\pi - 1) \, d(x_1,x_2) + \int_{\Omega_1} \alpha \, d(\mu-(P_1)_\#\pi) + \int_{\Omega_2} \beta \, d(\nu-(P_2)_\#\pi)$$

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

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 [15, 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 α and β, 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). (Naturally, existence under the stated condition can also be shown using Tonelli's direct method; here we give a proof based on the already shown convex duality for the sake of conciseness.)

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 ν. This shows that the conditions are necessary to have 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 2.10. 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. Here and throughout the rest of the paper, we use the usual shorthand $\{f > \lambda\}$ for the set $\{x \in \Omega : f(x) > \lambda\}$. We also recall from Remark 2.4 that $\mathbb{1}_A$ refers to the characteristic function of the set A. The following result can also be found in [9, Thm. 2.7], but the proof there needs a constraint qualification for the primal problem which we do not need in this formulation. We present a full proof for the sake of completeness.

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

Proof. The fact that $\operatorname{supp} \bar{\pi} \subset \operatorname{supp} \mu \times \operatorname{supp} \nu$ follows from the marginal constraints and the nonnegativity of $\bar{\pi}$. It remains to show that $\operatorname{supp} \bar{\pi} \supset \operatorname{supp} \mu \times \operatorname{supp} \nu$. For a contradiction, assume there is some $\hat{x} \in (\operatorname{supp} \mu \times \operatorname{supp} \nu) \setminus \operatorname{supp} \bar{\pi}$, then there exists a radius $r > 0$ such that $\bar{\pi} = 0$ on each ball $B_s(\hat{x})$ with $s < r$, but $\mu(P_1(B_s(\hat{x}))) > 0$ and $\nu(P_2(B_s(\hat{x}))) > 0$. In particular, there exist $\omega_1 \subset P_1(B_{r/2}(\hat{x}))$ and

$\omega_2 \subset P_2(B_{r/2}(\hat{x}))$ such that $\mu(\omega_i) > 0$ and $\mathcal{L}(\omega_i) > 0$ for $i = 1, 2$, but $\bar{\pi}(\omega_1 \times \omega_2) = 0$. We may choose ω_1, ω_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 nonzero Lebesgue measure and with $\bar{\pi} > \varepsilon$ on $(\tilde{\omega}_1 \times \omega_2) \cup (\omega_1 \times \tilde{\omega}_2)$.

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

$$\tilde{\pi} = \bar{\pi} + t \left[\mathbb{1}_{\omega_1 \times \omega_2}(x_1, x_2) + \kappa_1 \kappa_2 \mathbb{1}_{\tilde{\omega}_1 \times \tilde{\omega}_2}(x_1, x_2) - \kappa_1 \mathbb{1}_{\tilde{\omega}_1 \times \omega_2}(x_1, x_2) - \kappa_2 \mathbb{1}_{\omega_1 \times \tilde{\omega}_2}(x_1, x_2) \right]$$

for $0 < t < \varepsilon/\min\{\kappa_1, \kappa_2\}$. Then $\tilde{\pi}$ is feasible. We will now argue that for small enough t we have

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

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

$$\int_M c\tilde{\pi} \, \mathrm{d}(x_1, x_2) = \int_M c\bar{\pi} \, \mathrm{d}(x_1, x_2) + tC_0$$

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

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

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

$$\int_{\omega_1 \times \tilde{\omega}_2} \tilde{\pi} \log \tilde{\pi} \, \mathrm{d}(x_1, x_2) = \int_{\omega_1 \times \tilde{\omega}_2} (\bar{\pi} - \kappa_2 t) \log(\bar{\pi} - \kappa_2 t) \, \mathrm{d}(x_1, x_2)$$

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

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

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

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

We obtain

$$\int_M \tilde{\pi} \log \tilde{\pi} \, \mathrm{d}(x_1, x_2) - \int_M \bar{\pi} \log \bar{\pi} \, \mathrm{d}(x_1, x_2) \leq \kappa t \log t - \kappa_2 t \int_{\omega_1 \times \tilde{\omega}_2} \log(\bar{\pi} - \kappa_2 t) \, \mathrm{d}(x_1, x_2) - \kappa t$$

$$- \kappa_1 t \int_{\tilde{\omega}_1 \times \omega_2} \log(\bar{\pi} - \kappa_1 t) \, \mathrm{d}(x_1, x_2) - \kappa t + \kappa_1 \kappa_2 t \int_{\tilde{\omega}_1 \times \tilde{\omega}_2} \bar{\pi} + t\kappa_1 \kappa_2 \, \mathrm{d}(x_1, x_2).$$

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.

Combining the estimates for cost and entropy yields

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

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

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. $\qquad\square$

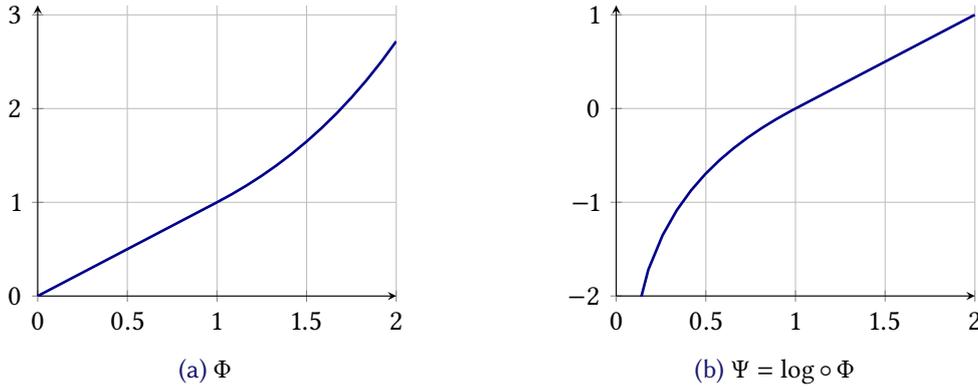(a) Φ　　　　　　　　　　　　　　　　　(b) $\Psi = \log \circ \Phi$

Figure 1: Young functions for $L \log L$ and $L_{\exp}$

Theorem 3.3 shows that the natural setting for the entropically regularized problem (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 can be used as a basis for deriving the Sinkhorn algorithm.

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

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

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

see Figure 1. Note that Φ is convex and Ψ concave and that the function Φ coincides with the Young function $\Phi_{\exp}$ from (2.2), which is associated with $L_{\exp}$.

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

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

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

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

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

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

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

$$
+ \gamma \int_{\Omega_1} \Psi(u_1)\mu \, dx_1 + \gamma \int_{\Omega_2} \Psi(u_2)\nu \, dx_2 \,.
$$

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

$$(4.2) \quad B(u_1, u_2) := \int_{\Omega_1 \times \Omega_2} \Phi(u_1(x_1)) \Phi(u_2(x_2)) e^{-\frac{c(x_1, x_2)}{\gamma}} \, d(x_1, x_2) - \int_{\Omega_1} \Psi(u_1) \mu \, dx_1 - \int_{\Omega_2} \Psi(u_2) \nu \, dx_2$$

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

Lemma 4.1. *If* $\|v \mathbb{1}_{\{v > 0\}}\|_{\Phi_{\exp}} > \max(1, \mathcal{L}(\Omega))$, *then*

$$\|v \mathbb{1}_{\{v > 0\}}\|_{\Phi_{\exp}} \leq \log \left(\frac{1}{\mathcal{L}(\Omega)} \int_\Omega \Phi((v + 1) \mathbb{1}_{\{v > 0\}}) \, dx \right) / \log \frac{e}{\min(1, \mathcal{L}(\Omega))} \,.$$

Proof. Set $\gamma_\varepsilon = \|v \mathbb{1}_{\{v > 0\}}\|_{\Phi_{\exp}} - \varepsilon$ for some $\varepsilon > 0$ such that still $\gamma_\epsilon > \max(1, \mathcal{L}(\Omega))$. Then it holds that $\int_\Omega \Phi\left(\frac{v \mathbb{1}_{\{v > 0\}}}{\gamma_\epsilon}\right) dx > \|v \mathbb{1}_{\{v > 0\}}\|_{\Phi_{\exp}} > \max(1, \mathcal{L}(\Omega))$.

By Jensen's inequality we have

$$
\begin{aligned}
\frac{\min(1, \mathcal{L}(\Omega))}{e} \left(\frac{1}{\mathcal{L}(\Omega)} \int_\Omega \Phi((v + 1) \mathbb{1}_{\{v > 0\}}) \, dx \right)^{\frac{1}{\gamma_\epsilon}} &= \frac{\min(1, \mathcal{L}(\Omega))}{e} \left(\frac{1}{\mathcal{L}(\Omega)} \int_\Omega \mathbb{1}_{\{v > 0\}} e^v \, dx \right)^{\frac{1}{\gamma_\epsilon}} \\
&\geq \frac{\min(1, \frac{1}{\mathcal{L}(\Omega)})}{e} \int_\Omega \mathbb{1}_{\{v > 0\}} (e^v)^{\frac{1}{\gamma_\epsilon}} \, dx \\
&= \min(1, \tfrac{1}{\mathcal{L}(\Omega)}) \int_\Omega \mathbb{1}_{\{v > 0\}} e^{\frac{v}{\gamma_\epsilon} - 1} \, dx \\
&\geq \min(1, \tfrac{1}{\mathcal{L}(\Omega)}) \int_\Omega \Phi\left(\frac{v \mathbb{1}_{\{v > 0\}}}{\gamma_\varepsilon} \right) dx \\
&> \min(1, \tfrac{1}{\mathcal{L}(\Omega)}) \max(1, \mathcal{L}(\Omega)) = 1.
\end{aligned}
$$

Taking logarithms, we deduce that $\gamma_\varepsilon < \log \left(\frac{1}{\mathcal{L}(\Omega)} \int_\Omega \Phi((v + 1) \mathbb{1}_{\{v > 0\}}) \, dx \right) / \log \frac{e}{\min(1, \mathcal{L}(\Omega))}$, and letting $\varepsilon \to 0$ yields the claim. $\qquad \square$

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* $\int_{\Omega_1} \Phi(\tilde{u}_1) \, dx_1 = 1$.

Proof. Note that $u_1 > 0$ μ-a.e. and $u_2 > 0$ ν-a.e. as $B(u_1, u_2) < \infty$. By construction, the same holds for $\tilde{u}_1$ and $\tilde{u}_2$. The first statement is now a direct consequence of the invariance of the cost functional in (D) under the mapping $(\alpha, \beta) \mapsto (\alpha - K, \beta + K)$. For the second statement, first note that $\int_{\Omega_1} \Phi(\tilde{u}_1) \, dx_1$ is continuous in K. Moreover,

$$\int_{\Omega_1} \Phi(\tilde{u}_1) \, dx_1 \to 0, \quad \text{for } K \to \infty \quad \text{and}$$

$$\int_{\Omega_1} \Phi(\tilde{u}_1) \, dx_1 \to \infty, \quad \text{for } K \to -\infty$$

so that the assertion follows by the intermediate value theorem. $\qquad \square$

Remark 4.3. While $\|\tilde{u}_1\|_{\Phi_{\exp}} = 1$ implies $\int_{\Omega_1} \Phi(\tilde{u}_1)\, dx_1 \leq 1$ and $\int_{\Omega_1} \Phi(\tilde{u}_1)\, dx_1 = 1$ implies $\|\tilde{u}_1\|_{\Phi_{\exp}} \leq 1$, in general we cannot achieve both equalities simultaneously due to Remark 2.4.

Modulo this invariance we now obtain coercivity.

Lemma 4.4. *Let u_1^n, $n = 1, 2, \ldots,$ be a sequence in $L_{\exp}(\Omega_1)$ such that $\int_{\Omega_1} \Phi(u_1^n) = 1$ for all n. Then $\|u_2^n\|_{\Phi_{\exp}} \to \infty$ for $n \to \infty$ implies $B(u_1^n, u_2^n) \to \infty$ as $n \to \infty$.*

Proof. Without loss of generality we may assume the u_2^n to be nonnegative, since replacing u_2^n with its absolute value decreases $B(u_1^n, u_2^n)$ without changing $\|u_2^n\|_{\Phi_{\exp}}$. Due to $\int_{\Omega_1} \Phi(u_1^n)\, dx_1 = 1$ we have $\|u_1^n\|_{\Phi_{\exp}} \leq 1$ and thus

$$\int_{\Omega_1} \Psi\left(u_1^n\right) \mu\, dx_1 \leq \int_{\Omega_1} (u_1^n - 1)\mu\, dx_1 \leq c_\Phi \|u_1^n\|_{\Phi_{\exp}} \|\mu\|_{\Phi_{\log}} - 1 \leq c_\Phi \|\mu\|_{\Phi_{\log}} - 1,$$

where c_Φ denotes the generic equivalence constant for the duality from Proposition 2.7. Analogously we obtain

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

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

$$B(u_1^n, u_2^n) \geq C \int_{\Omega_1} \Phi(u_1^n)\, dx_1 \int_{\Omega_2} \Phi(u_2^n)\, dx_2 - \int_{\Omega_1} \Psi\left(u_1^n\right) \mu\, dx_1 - \int_{\Omega_2} \Psi\left(u_2^n\right) \nu\, dx_2$$

$$\geq C \int_{\Omega_2} \Phi(u_2^n)\, dx_2 - c_\Phi \|\mu\|_{\Phi_{\log}} + 1 - c_\Phi \|\max(u_2^n - 1, 0)\|_{\Phi_{\exp}} \|\nu\|_{\Phi_{\log}}.$$

Since $\|u_2^n\|_{\Phi_{\exp}} \to \infty$ and u_2^n is nonnegative, we also have $\|\max(u_2^n - 1, 0)\|_{\Phi_{\exp}} \to \infty$ as $n \to \infty$ and therefore

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

by Lemma 4.1. Now Lemma 2.6 implies that $\int_{\Omega_2} \Phi(u_2^n)\, dx \to \infty$ for $n \to \infty$ and therefore that

$$B(u_1^n, u_2^n) \geq C \int_{\Omega_2} \Phi(u_2^n)\, dx - c_\Phi \|\mu\|_{\Phi_{\log}} + 1 - c_\Phi \frac{\log\left(\frac{1}{\mathcal{L}(\Omega_2)} \int_{\Omega_2} \Phi(u_2^n)\, dx_2\right)}{\log\left(\frac{e}{\min(1, \mathcal{L}(\Omega_2))}\right)} \|\nu\|_{\Phi_{\log}} \to \infty,$$

which yields the desired contradiction. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 4.5. *B is sequentially weakly-$*$ lower semi-continuous on $L_{\exp}(\Omega_1) \times L_{\exp}(\Omega_2)$.*

Proof. Let $(u_1^n, u_2^n) \rightharpoonup^* (u_1, u_2)$ in $L_{\exp}(\Omega_1) \times L_{\exp}(\Omega_2)$. 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 $1 \leq p < \infty$. Since $-\Psi$ is a lower semi-continuous and convex integrand, it thus follows, e.g., by [3, Thm. 13.1.1] that

$$\liminf_{n \to \infty} \int_{\Omega_1} -\Psi(u_1^n)\mu\, dx_1 \geq \int_{\Omega_1} -\Psi(u_1)\mu\, dx_1, \qquad \liminf_{n \to \infty} \int_{\Omega_2} -\Psi(u_2^n)\mu\, dx_1 \geq \int_{\Omega_2} -\Psi(u_2)\mu\, dx_2,$$

and hence that these functionals are weak-$*$ sequentially lower semicontinuous on $L_{\exp}(\Omega_1) \times L_{\exp}(\Omega_2)$.

It remains to show weak-$*$ 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 Ω_1 and Ω_2 into a finite number of subsets Ω_i^k with $\mathcal{L}(\Omega_i^k) \leq \frac{1}{N}$. We further

assume that the decompositions $(\Omega_1^k)_k$ and $(\Omega_2^k)_k$ for $N+1$ are obtained from the decompositions for N by refinement. Defining $c_{kl} = \min_{(x_1,x_2) \in \Omega_1^k \times \Omega_2^l} e^{-c(x_1,x_2)/\gamma}$, we then have

$$\liminf_{n\to\infty} \int_{\Omega_1 \times \Omega_2} \Phi(u_1^n(x_1))\Phi(u_2^n(x_2))e^{-\frac{c(x_1,x_2)}{\gamma}}\, \mathrm{d}(x_1,x_2)$$

$$\geq \liminf_{n\to\infty} \int_{\Omega_1 \times \Omega_2} \Phi(u_1^n(x_1))\Phi(u_2^n(x_2)) \sum_{k,l} c_{kl} \mathbb{1}_{\Omega_1^k \times \Omega_2^l}\, \mathrm{d}(x_1,x_2)$$

$$= \liminf_{n\to\infty} \sum_{k,l} c_{kl} \int_{\Omega_1 \times \Omega_2} \Phi(u_1^n(x_1))\Phi(u_2^n(x_2)) \mathbb{1}_{\Omega_1^k}(x_1) \mathbb{1}_{\Omega_2^l}(x_2)\, \mathrm{d}(x_1,x_2)$$

$$\geq \sum_{k,l} c_{kl} \liminf_{n\to\infty} \int_{\Omega_1} \Phi(u_1^n(x_1)) \mathbb{1}_{\Omega_1^k}(x_1)\, \mathrm{d}x_1 \, \liminf_{n\to\infty} \int_{\Omega_2} \Phi(u_2^n(x_2)) \mathbb{1}_{\Omega_2^l}(x_2)\, \mathrm{d}x_2 \, .$$

Similarly as above, it follows from the lower semi-continuity and convexity of Φ that $u \mapsto \int_{\Omega_1^k} \Phi(u)\, \mathrm{d}x_1$ and $v \mapsto \int_{\Omega_2^l} \Phi(v)\, \mathrm{d}x_2$ are sequentially weakly-$*$ lower semi-continuous on $L_{\exp}(\Omega_1^k)$ and $L_{\exp}(\Omega_2^l)$, respectively. Hence

$$\sum_{k,l} c_{kl} \liminf_{n\to\infty} \int_{\Omega_1} \Phi(u_1^n(x_1)) \mathbb{1}_{\Omega_1^k}(x_1)\, \mathrm{d}x_1 \, \liminf_{n\to\infty} \int_{\Omega_2} \Phi(u_2^n(x_2)) \mathbb{1}_{\Omega_2^l}(x_2)\, \mathrm{d}x_2$$

$$\geq \sum_{k,l} c_{kl} \int_{\Omega_1} \Phi(u_1(x_1)) \mathbb{1}_{\Omega_1^k}(x_1)\, \mathrm{d}x_1 \int_{\Omega_2} \Phi(u_2(x_2)) \mathbb{1}_{\Omega_2^l}(x_2)\, \mathrm{d}x_2$$

$$= \int_{\Omega_1 \times \Omega_2} \Phi(u_1(x_1))\Phi(u_2(x_2)) \sum_{k,l} c_{kl} \mathbb{1}_{\Omega_1^k \times \Omega_2^l}(x_1,x_2)\, \mathrm{d}(x_1,x_2)$$

$$\to \int_{\Omega_1 \times \Omega_2} \Phi(u_1(x_1))\Phi(u_2(x_2))e^{-\frac{c(x_1,x_2)}{\gamma}}\, \mathrm{d}(x_1,x_2) \quad \text{as } N \to \infty$$

by the monotone convergence theorem, since $\sum_{k,l} c_{kl} \mathbb{1}_{\Omega_1^k \times \Omega_2^l} \nearrow e^{-c/\gamma}$ monotonically. $\qquad\square$

Theorem 4.6 (dual existence). *Problem* $(\mathrm{D}_{\exp})$ *possesses a maximizer* $(\bar{u}_1, \bar{u}_2) \in L_{\exp}(\Omega_1) \times L_{\exp}(\Omega_2)$.

Proof. We show that B possesses a minimizer. The energy B is finite at, e.g., $u_1 \equiv 1 \equiv u_2$. We thus may consider a minimizing sequence (u_1^n, u_2^n) in $L_{\exp}(\Omega_1) \times L_{\exp}(\Omega_2)$, where by Lemma 4.2 we may assume $\int_{\Omega_1} \Phi_{\exp}(u_1^n)\, \mathrm{d}x_1 = 1$ without loss of generality. Lemma 4.4 now implies boundedness of $\|u_2^n\|_{\Phi_{\exp}}$ so that by the Banach–Alaoglu theorem we may extract a weakly-$*$ convergent subsequence from (u_1^n, u_2^n) (recalling that $L \log L(\Omega_1 \times \Omega_2)$ is separable by Lemma 2.8). The claim now follows from the lower semi-continuity of B along that subsequence by Lemma 4.5. $\qquad\square$

From dual solutions $\bar{u}_1$ 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 $\bar{u}_1 \in L_{\exp}(\Omega_1)$ and $\bar{u}_2 \in L_{\exp}(\Omega_2)$ 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 $L_{\exp}$, but the negative parts need not even be functions as they could be $-\infty$ everywhere.

Nevertheless, from $(\mathrm{D}_{\exp})$ one sees that $\bar{u}_1 \geq 0$ μ-almost everywhere and $\bar{u}_2 \geq 0$ ν-almost everywhere, and hence $\bar{u}_1$ and $\bar{u}_2$ are at least μ- and ν-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 $(\mathrm{D}_{\exp})$ to (P).

Proposition 4.7 (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* ($D_{\exp}$) *admit a solution, and their optimal values coincide.*

Proof. Existence for both problems follows from Theorems 3.3 and 4.6. To show their equality, by Proposition 3.1 it suffices to show that the value of (D) equals that of ($D_{\exp}$). 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

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

and taking the supremum over all α, β yields that the value of (D) is at most that of ($D_{\exp}$).

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

$$B\left(\max\left\{\bar{u}_1, \tfrac{1}{n}\right\}, \max\left\{\bar{u}_1, \tfrac{1}{n}\right\}\right) = \lim_{N \to \infty} B\left(\min\left\{\max\left\{\bar{u}_1, \tfrac{1}{n}\right\}, N\right\}, \min\left\{\max\left\{\bar{u}_1, \tfrac{1}{n}\right\}, N\right\}\right).$$

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 $\alpha_n \to \gamma\Psi(u_1)$ in $L^2(\Omega_1)$ and $\beta_n \to \gamma\Psi(u_2)$ in $L^2(\Omega_2)$. Here we may assume α_n, β_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

$$\int_{\Omega_1} \alpha_n\mu \, \mathrm{d}x_1 + \int_{\Omega_2} \beta_n\nu \, \mathrm{d}x_2 \to \gamma \left[\int_{\Omega_1} \Psi(u_1)\mu \, \mathrm{d}x_1 + \int_{\Omega_2} \Psi(u_2)\nu \, \mathrm{d}x_2\right]$$

due to the weak-$*$ convergence. Finally, as α_n, β_n converge in L^2, $e^{\frac{\alpha_n(x_1) + \beta_n(x_2)}{\gamma}}$ converges a.e. (after passing to a subsequence). Using uniform boundedness of α_n, β_n, the dominated convergence theorem yields

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

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 4.8 (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 $(\bar{u}_1, \bar{u}_2) \in L_{\exp}(\Omega_1) \times L_{\exp}(\Omega_2)$ of* ($D_{\exp}$) *satisfy*

$$(4.3a) \qquad \int_{\Omega_2} \Phi(\bar{u}_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}} \, \mathrm{d}x_2 \, \Phi(\bar{u}_1(x_1)) = \mu(x_1),$$

$$(4.3b) \qquad \int_{\Omega_1} \Phi(\bar{u}_1(x_1))e^{-\frac{c(x_1, x_2)}{\gamma}} \, \mathrm{d}x_1 \, \Phi(\bar{u}_2(x_2)) = \nu(x_2),$$

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

$$(4.4) \qquad \bar{\pi}(x_1, x_2) = \Phi(\bar{u}_1(x_1))\Phi(\bar{u}_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}}$$

is the solution of (P).

Proof. Let $\bar{u}_1, \bar{u}_2$ be solutions of the dual problem. We start with deriving the necessary conditions (4.3). 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 $\int_{\Omega_1} \Psi(\bar{u}_1)\mu \, dx_1 + \int_{\Omega_2} \Psi(\bar{u}_2)\nu \, dx_2 = -\infty$. Let now $\varepsilon > 0$ be arbitrary and consider any $\varphi \in L_{\exp}(\Omega_1) \cap L^\infty(\Omega_1)$ with $\varphi = 0$ on $\{\bar{u}_1 < \varepsilon\}$. We next argue that the dual functional B given in (4.2) is directionally differentiable in $(\bar{u}_1, \bar{u}_2)$ with respect to its first argument in direction φ. Since both $s \mapsto \Phi(s)$ and $s \mapsto \Psi(s)$ are differentiable at $s > 0$, so are the integrands pointwise almost everywhere on $\{\bar{u}_1 \geq \varepsilon\} \times \Omega_2$. It therefore suffices to show that the pointwise directional derivatives are integrable in order to differentiate under the integral. For the first term in B, we have almost everywhere on $\{\bar{u}_1 \geq \varepsilon\} \times \Omega_2$ that

$$
\Phi'(\bar{u}_1; \varphi)\Phi(\bar{u}_2)e^{-\frac{c}{\gamma}} = \begin{cases} \Phi(\bar{u}_1)\varphi\Phi(\bar{u}_2)e^{-\frac{c}{\gamma}} & \bar{u}_1 > 1, \\ \varphi\Phi(\bar{u}_2)e^{-\frac{c}{\gamma}} & \text{else,} \end{cases}
$$
$$
\leq \frac{1}{\varepsilon}\|\varphi\|_\infty \Phi(\bar{u}_1)\Phi(\bar{u}_2)e^{-\frac{c}{\gamma}},
$$

which is integrable on $\{\bar{u}_1 \geq \varepsilon\} \times \Omega_2$ since $\bar{u}_1$ and $\bar{u}_2$ are feasible for ($D_{\exp}$). An integrable lower bound is obtained similarly using $\varphi \in L^\infty(\Omega_1)$.

For the second term in B, the chain rule and differentiability of Φ yields almost everywhere on $\{\bar{u}_1 \geq \varepsilon\} \times \Omega_2$ that

$$
\Psi'(\bar{u}_1; \varphi) = \frac{1}{\Phi(\bar{u}_1)}\Phi'(\bar{u}_1)\varphi = (\bar{u}_1^{-1}\mathbb{1}_{\{\bar{u}_1 < 1\}} + \mathbb{1}_{\{\bar{u}_1 \geq 1\}})\varphi \leq \|\varphi\|_\infty(\varepsilon^{-1}\mathbb{1}_{\{\bar{u}_1 < 1\}} + \mathbb{1}_{\{\bar{u}_1 \geq 1\}}),
$$

where the right-hand side is integrable with respect to $d\mu$.

From the dominated convergence theorem, it thus follows that the partial directional derivative of B in the first direction is given by

$$
\begin{aligned}
B_1'(\bar{u}_1, \bar{u}_2; \varphi) = & -\int_{\{\bar{u}_1 \geq \varepsilon\}} \Psi'(\bar{u}_1; \varphi)\mu \, dx_1 \\
& + \int_{\{\bar{u}_1 \geq \varepsilon\} \times \Omega_2} \Phi'(\bar{u}_1(x_1); \varphi(x_1))\Phi(\bar{u}_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}} \, d(x_1, x_2) \\
= & -\int_{\{\bar{u}_1 \geq \varepsilon\}} \max\{\Phi(\bar{u}_1(x_1)), 1\}\varphi(x_1)\left[\frac{\mu(x_1)}{\Phi(\bar{u}_1(x_1))} - \int_{\Omega_2} \Phi(\bar{u}_2(x_2))e^{-\frac{c(x_1, x_2)}{\gamma}} \, dx_2\right] dx_1,
\end{aligned}
$$

where we have again used the integrability of the integrand to apply Fubini's Theorem in order to iterate the double integrals and used $\Phi'(s) = \max\{\Phi(s), 1\}$ and $\Psi'(s) = \Phi(s)^{-1}\Phi'(s)$.

By the specific choice of φ, we have $\bar{u}_1 \pm t\varphi \geq 0$ for all $t > 0$ sufficiently small. The optimality of $(\bar{u}_1, \bar{u}_2)$ thus implies that

$$
0 = B_1'(\bar{u}_1, \bar{u}_2; \varphi),
$$

and since φ was arbitrary on $\{\bar{u}_1 \geq \varepsilon\}$ and $\max\{\Phi(\bar{u}_1), 1\} > 0$, we must therefore have that

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

Furthermore, since $\varepsilon > 0$ was arbitrary and $\mu(x_1) = 0$ whenever $\bar{u}_1(x_1) = 0$, this equation even holds for μ-almost all $x_1 \in \Omega_1$, which yields (4.3a). Equation (4.3b) is derived analogously.

Now we show that $\bar{\pi}$ defined by (4.4) is a solution of the primal problem. First note that by construction, $\bar{\pi}$ is feasible (i.e., is non-negative and has the correct marginals). Since strong duality holds by Proposition 4.7, it thus suffices to show that the primal objective functional evaluated in $\bar{\pi}$ is equal to

the dual optimal objective value (D_{exp}). To that end, we insert (4.4) into the objective functional in (P) and obtain (using again the convention that $0 \log 0 = 0$)

$$
\begin{aligned}
\int_{\Omega_1 \times \Omega_2} c\bar{\pi} + \gamma\bar{\pi}(\log\bar{\pi} - 1)\, \mathrm{d}(x_1, x_2) &= \int_{\Omega_1 \times \Omega_2} c\Phi(\bar{u}_1)\Phi(\bar{u}_2)e^{-\frac{c}{\gamma}}\, \mathrm{d}(x_1, x_2) \\
&\quad + \gamma\Phi(\bar{u}_1)\Phi(\bar{u}_2)e^{-\frac{c}{\gamma}}\left(\Psi(\bar{u}_1) + \Psi(\bar{u}_2) - \frac{c}{\gamma} - 1\right)\mathrm{d}(x_1, x_2) \\
&= -\gamma \int_{\Omega_1 \times \Omega_2} \Phi(\bar{u}_1)\Phi(\bar{u}_2)e^{-\frac{c}{\gamma}}\, \mathrm{d}(x_1, x_2) \\
&\quad + \gamma \int_{\Omega_1 \times \Omega_2} \Psi(\bar{u}_1)\Phi(\bar{u}_1)\Phi(\bar{u}_2)e^{-\frac{c}{\gamma}}\, \mathrm{d}(x_1, x_2) \\
&\quad + \gamma \int_{\Omega_1 \times \Omega_2} \Psi(\bar{u}_2)\Phi(\bar{u}_1)\Phi(\bar{u}_2)e^{-\frac{c}{\gamma}}\, \mathrm{d}(x_1, x_2)\,.
\end{aligned}
$$

Since $\bar{u}_i \geq 0$ and hence $\Phi(\bar{u}_i) \geq 0$, we have that $\Psi(\bar{u}_i)\Phi(\bar{u}_i) = \log(\Phi(\bar{u}_i))\Phi(\bar{u}_i) \geq -\frac{1}{e}$. Furthermore, we have assumed $\mathcal{L}(\Omega_i) < \infty$ and can thus shift the integrand to allow applying Tonelli's Theorem in the second and third integral. Inserting (4.3), the right-hand side now coincides with $B(\bar{u}_1, \bar{u}_2)$. Hence strong duality holds for $(\bar{u}_1, \bar{u}_2)$ and $\bar{\pi}$, and thus the latter is a solution to (P). $\qquad \square$

Remark 4.9. The optimality system (4.3) can be used to derive the Sinkhorn algorithm. First, note that one only needs to find $\bar{u}_1$ and $\bar{u}_2$ that solve (4.3a) and (4.3b); an optimal plan $\bar{\pi}$ is then obtained from (4.4). The Sinkhorn method now solves the nonlinear system (4.3) by alternatingly solving the equations: Given u_2^n, compute u_1^{n+1} by solving (4.3a), i.e., setting

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

and then solve (4.3b) with u_2^{n+1} to obtain

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

Formulating this iteration directly in $\Phi(u_1)$ and $\Phi(u_2)$, we obtain the original Sinkhorn algorithm, cf. [20, Sec. 5.3.1].

Remark 4.10. The optimality system (4.3) also corresponds to the so-called Schrödinger system [14, Eq. (4.12)–(4.13) or (4.14)], i.e., the system of equations which characterizes the solution to the so-called Schrödinger bridge problem (essentially, the most likely transition path of a hot gas between the initial and final gas distribution μ and ν). Existence of solutions to that system was typically shown based on iterative approximation schemes (analogous to but predating the Sinkhorn algorithm; see the discussion in [14]). There are also alternative proofs exploiting the variational nature of the problem; however, these are not as straightforward as identifying (4.3) as the optimality conditions to an optimization problem which has a solution. In [7], for example, a minimizing sequence for the dual problem (D) is used to construct a sequence of measures of the type (4.4) that is then shown to converge to a solution of the Schrödinger bridge problem.

Finally, the optimality conditions (4.3a) and (4.3b) allow us to conclude which problem is solved by $(\bar{\alpha}, \bar{\beta})$.

Corollary 4.11. Let $\mu \in L \log L(\Omega_1)$, $\nu \in L \log L(\Omega_2)$, and $c \in C(\Omega_1 \times \Omega_2)$. Let $(\bar{u}_1, \bar{u}_2) \in L_{\exp}(\Omega_1) \times L_{\exp}(\Omega_2)$ be a solution of $(D_{\exp})$. 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

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

and the values of (D_{L^1}) and $(D_{\exp})$ coincide.

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 Ω_1. From the optimality condition (4.3a), 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 μ, i.e., $\bar{\alpha} \in L^1(\Omega_1, \mu)$. The result for $\bar{\beta}$ follows analogously. Finally, it follows from a density argument that (D_{L^1}) cannot exceed $(D_{\exp})$. Indeed, assume there are $\alpha \in L^1(\Omega_1, \mu)$ and $\beta \in L^1(\Omega_2, \nu)$ with an objective functional value C strictly larger than $(D_{\exp})$. By invoking the monotone convergence theorem as in the proof of Proposition 4.7, we may assume without loss of generality that α and β are bounded. Defining now $u_1 = \Psi^{-1}(\frac{\alpha}{\gamma}) \in L^\infty(\Omega_1) \subset L_{\exp}(\Omega_1)$ and $u_2 = \Psi^{-1}(\frac{\beta}{\gamma}) \in L^\infty(\Omega_2) \subset L_{\exp}(\Omega_2)$ shows that $(D_{\exp})$ is no smaller than C, the desired contradiction. $\qquad \square$

Remark 4.12. As for (D) and as formalized in Lemma 4.2, solutions to $(D_{\exp})$ are not unique.

5   Γ-LIMIT

We now turn to Γ-convergence of the regularized problem. Recall from, e.g., [10], that a sequence $\{F_n\}$ of functionals $F_n : X \to \overline{\mathbb{R}}$ on a metric space X is said to Γ-converge to a functional $F : X \to \overline{\mathbb{R}}$, written $F = \Gamma\text{-}\lim_{n \to \infty} F_n$, if

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

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

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

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

It is a straightforward consequence of this definition that if F_n Γ-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, Γ-convergence is stable under perturbations by continuous functionals.

Here we aim to approximate optimal transport plans $\bar{\pi}$ of the unregularized problem for marginals μ and ν which are *not* required to be in $L \log L(\Omega)$, i.e., we allow arbitrary 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 μ_γ and ν_γ in $L \log L(\Omega)$ converging to μ and ν, respectively, and show that the regularized problem with these marginals Γ-converges to the unregularized problem with the original marginals. The conceptually different case of Γ-convergence for *fixed, non-mollified* marginals (which then, however, need to be of finite entropy) has been treated in [11, Thm. 2.7]. Our setting with smoothed marginals allows simpler constructions in the lim sup-inequality since a given transport plan is merely approximated via mollification. A further difference to [11, Thm. 2.7] is that we work on a compact set Ω instead of $\mathbb{R}^n$ and need to couple the smoothing parameter to the regularization parameter to obtain Γ-convergence.

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

$$B_\delta^n(x) := \tfrac{1}{\delta^n} B(\tfrac{x}{\delta}), \qquad G_\delta(x_1, x_2) := B_\delta^{n_1}(x_1) B_\delta^{n_2}(x_2).$$

Since we will smooth the marginals and the transport plans by convolutions, we will need to slightly extend the domains Ω_1 and Ω_2 to avoid boundary effects. Hence, let $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ be compact supersets of Ω_1 and Ω_2, respectively, such that

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

and 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 Ω_1, we denote by $\tilde{f}$ the extension of f to $\tilde{\Omega}_1$ by zero (and analogously for functions and measures on Ω_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

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

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

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

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

$$\gamma \to 0, \qquad \delta \to 0, \qquad \gamma \log(\delta) \to 0,$$

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

$$\underset{(\gamma, \delta) \to 0}{\Gamma\text{-}\lim} \ E_\gamma^{\mu_\delta, \nu_\delta} = E_0^{\mu, \nu}$$

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

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

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

Proof. For the first statement, we verify the two conditions in the definition of Γ-convergence.

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

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

and thus that

$$F_0[\tilde{\pi}] = \lim_{\delta \to 0} F_0[\pi_\delta] - \lim_{\gamma \to 0} \gamma \left| \tilde{\Omega}_1 \times \tilde{\Omega}_2 \right| \le \liminf_{(\gamma, \delta) \to 0} F_\gamma[\pi_\delta] .$$

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

(ii): It suffices to consider a recovery sequence for $\pi \in \mathcal{P}(\Omega_1 \times \Omega_2)$, because the marginal conditions for μ and ν 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 $\pi_\delta \le \|G_\delta\|_{L^\infty} \|\tilde{\pi}\|_{L^1} \le \frac{C}{\delta^N}$ for some constant $C > 0$ and $N := n_1 + n_2$ and we have

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

we conclude that

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

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

$$E_0^{\mu,\nu}[\tilde{\pi}] = \lim_{(\gamma,\delta) \to 0} \left[F_0[\pi_\delta] - \gamma(1 + N \log \delta - \log C) \right] \ge \limsup_{(\gamma,\delta) \to 0} F_\gamma[\pi_\delta] .$$

For the second statement, recall from Lemma 2.11 that

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

so that $\gamma \|\pi_\delta\|_{\Phi_{\log}} \to \infty$. By Lemma 2.6, this immediately yields $\gamma \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \pi_\delta \log^+ \pi_\delta \, \mathrm{d}(x_1, x_2) \to \infty$, which implies

$$\gamma \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \pi_\delta (\log \pi_\delta - 1) \, \mathrm{d}(x_1, x_2) = \gamma \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \pi_\delta \log \pi_\delta \, \mathrm{d}(x_1, x_2) - \gamma \to \infty$$

and thus $F_\gamma[\pi_\delta] \to \infty$ so that the assertion follows. $\qquad \square$

The conditions on γ and δ 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 Γ-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. More general penalties have been considered in a different framework in [26], and other choices (such as the product measure of the marginals) are possible in the approach considered here as well 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 $L_{\exp}(\Omega)$.

ACKNOWLEDGMENTS

REFERENCES

[1] R. A. Adams, *Sobolev Spaces*, volume 140 of Pure and Applied Mathematics, Academic Press, Inc., Boston, MA, second edition, 2003.

[2] 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.

[3] 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), 2006, doi:10.1137/1.9781611973488.

[4] 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.

[5] 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.

[6] R. J. Berman, The Sinkhorn algorithm, parabolic optimal transport and geometric Monge–Ampère equations, 2017.

[7] A. Beurling, An automorphism of product measures, *Annals of Mathematics* 72 (1960), 189–200, doi:10.2307/1970151.

[8] J. M. Borwein and A. S. Lewis, Decomposition of multivariate functions, *Canadian Journal of Mathematics* 44 (1992), 463–482, doi:10.4153/cjm-1992-030-9.

[9] J. M. Borwein, A. S. Lewis, and R. D. Nussbaum, Entropy minimization, DAD problems, and doubly stochastic kernels, *Journal of Functional Analysis* 123 (1994), 264–307, doi:10.1006/jfan.1994.1089.

[10] A. Braides, Γ-*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.

[11] 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.

[12] G. Carlier and M. Laborde, A differential approach to the multi-marginal Schrödinger system, *SIAM Journal on Mathematical Analysis* 52 (2020), 709–717, doi:10.1137/19m1253800.

[13] Y. Chen, T. T. Georgiou, and M. Pavon, Entropic and displacement interpolation: a computational approach using the Hilbert metric, *SIAM Journal on Applied Mathematics* 76 (2015), 2375–2396, doi:10.1137/16m1061382.

[14] Y. Chen, T. T. Georgiou, and M. Pavon, Stochastic control liasons: Richard Sinkhorn meets Gaspard Monge on a Schroedinger bridge, 2020.

[15] 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.

[16] I. Csiszar, *I*-divergence geometry of probability distributions and minimization problems, *Ann. Probab.* 3 (1975), 146–158, doi:10.1214/aop/1176996454.

[17] 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.

[18] M. Cuturi and G. Peyré, Semidual regularized optimal transport, *SIAM Review* 60 (2018), 941–965, doi:10.1137/18m1208654.

[19] 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.

[20] 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.

[21] R. Fortet, Résolution d'un systeme d'equations de M. Schrödinger, *J. Math. Pure Appl. IX* 1 (1940), 83–105.

[22] 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.

[23] 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.

[24] 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.

[25] D. A. Lorenz, P. Manns, and C. Meyer, Quadratically regularized optimal transport, *Applied Mathematics & Optimization* (2019), doi:10.1007/s00245-019-09614-w.

[26] C. Léonard, Minimization of entropy functionals, *Journal of Mathematical Analysis and Applications* 346 (2008), 183 – 204, doi:10.1016/j.jmaa.2008.04.048.

[27] C. Léonard, Convex minimization problems with weak constraint qualifications, *Journal of Convex Analysis* 17 (2010), 321–348.

[28] J. Musielak, *Orlicz Spaces and Modular Spaces*, Springer Berlin Heidelberg, 1983, doi:10.1007/bfb0072210.

[29] 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.

[30] M. M. Rao and Z. D. Ren, *Theory of Orlicz Spaces*, Pure and Applied Mathematics, Dekker, 1991.

[31] R. T. Rockafellar, Integrals which are convex functionals, *Pacific J. Math.* 24 (1968), 525–539, doi:10.2140/pjm.1968.24.525.

[32] 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.

[33] B. Simon, *Convexity: An Analytic Viewpoint*, volume 187, Cambridge University Press, 2011, doi:10.1017/cbo9780511910135.

[34] C. Villani, *Optimal Transport. Old and New*, volume 338 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 2009, doi:10.1007/978-3-540-71050-9.