

**DFG** Deutsche  
Forschungsgemeinschaft  
Priority Programme 1962

*Orlicz Space Regularization of Continuous  
Optimal Transport Problems*

Dirk Lorenz, Hinrich Mahler



Preprint Number SPP1962-161

received on March 25, 2021

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/>

# Orlicz space regularization of continuous optimal transport problems

Dirk Lorenz\*

Hinrich Mahler\*

March 5, 2021

In this work we analyze regularized optimal transport problems in the so-called Kantorovich form, i.e. given two Radon measures on two compact sets, the aim is to find a transport plan, which is another Radon measure on the product of the sets, that has these two measures as marginals and minimizes the sum of a certain linear cost function and a regularization term. We focus on regularization terms where a Young's function applied to the (density of the) transport plan is integrated against a product measure. This forces the transport plan to belong to a certain Orlicz space. The predual problem is derived and proofs for strong duality and existence of primal solutions of the regularized problem are presented. Existence of (pre-)dual solutions is shown for the special case of  $L^p$ -regularization for  $p \geq 2$ . Moreover, two results regarding  $\Gamma$ -convergence are stated: The first is concerned with marginals that do not lie in the appropriate Orlicz space and guarantees  $\Gamma$ -convergence to the original Kantorovich problem, when smoothing the marginals. The second result gives convergence of a regularized and discretized problem to the unregularized, continuous problem.

## 1. Introduction

In this paper we consider the optimal transport problem in the Kantorovich form in the following setting: For compact sets  $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ , probability measures  $\mu_1, \mu_2$  on  $\Omega_1, \Omega_2$ , respectively, and a real-valued continuous cost function  $c : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$  we want to solve

$$\inf_{\pi} \int_{\Omega_1 \times \Omega_2} c \, d\pi \quad (\text{OT})$$

where the infimum is taken over all probability measures on  $\Omega_1 \times \Omega_2$  which have  $\mu_1$  and  $\mu_2$  as their first and second marginals, respectively. This problem has been well studied, and an overview is given in the recent books [31, 33, 27]. For example, it is known that the problem has a solution  $\bar{\pi}$  and that the support of  $\bar{\pi}$  is contained in the so-called  $c$ -superdifferential of a  $c$ -concave function on  $\Omega_1$ , see [1, Theorem 1.13]. 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

---

\*Institute of Analysis and Algebra, TU Braunschweig, 38092 Braunschweig, Germany, (d.lorenz@tu-braunschweig.de, h.mahler@tu-braunschweig.de)

respect to the Lebesgue measure. This motivates the use of regularization of the continuous problem to obtain approximate solutions that are absolutely continuous w.r.t. given measures. That in turn allows to apply classical discretization techniques to solve the regularized problem approximately.

A regularization method that received much attention recently is regularization with the negative entropy of  $\pi$ , i.e. adding a term  $\int_{\Omega_1 \times \Omega_2} \Phi(\pi) d\lambda$  with  $\Phi(t) = t \log(t)$  and some measure  $\lambda$  on  $\Omega_1 \times \Omega_2$  [8, 11, 12, 10, 2]. Since  $\pi$  is a measure, one has to interpret  $\Phi(\pi)$  appropriately: One should think of  $\pi$  as the Radon-Nikodym derivative of  $\pi$  with respect to the regularization measure  $\lambda$ , and we will make this distinction explicit in the following. In [10] entropic regularization with respect to the Lebesgue measure is considered and it is shown that the analysis of entropically regularized optimal transport problems naturally takes place in the function space  $L \log L$  (also called Zygmund space [3]) and that optimal plans for entropic regularization are always in  $L \log L(\Omega_1 \times \Omega_2)$  and exist if and only if the marginals are in the spaces  $L \log L(\Omega_i)$ . These spaces are an example of so-called Orlicz spaces [28]. This motivates the analysis of regularization in arbitrary Orlicz spaces in this paper. Another motivation to study a more general regularization comes from the fact that regularization with the  $L^2$ -norm have been shown to be beneficial in some applications, see [29, 4, 13, 20]. Using the product of the marginals  $\lambda = \mu_1 \times \mu_2$  for regularization has been considered in the case of entropic regularization [17, 27, 32]. In this case one can show existence of the dual problem with different techniques. These observations motivate us to consider regularization with Young's function with respect to general measures. Notable regularizations that our approach covers are  $L^p$  regularization with  $p > 1$  arbitrary and the Tsallis entropy [25].

The notion of Orlicz spaces in the context of convex integral functionals has previously been used in [18], where the author considers a more general setting than the one presented here. More precisely, the spaces used in [18] are a generalization of the Orlicz spaces used here, which are also known as Musielak-Orlicz spaces [24]. Existence of both primal and dual optimizers are covered. By choosing  $\gamma^*(z, t) = \varepsilon \cdot \Phi(t) + c(z)t + A(z)$  with  $A(z) := \min_t \varepsilon \cdot \Phi(t) + c(z)t$  and regularization parameter  $\varepsilon$ , a problem similar to the one considered here is recovered. 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 [18] is to weaken the necessary assumptions as much as possible, the overall setting is more abstract and the proofs rely heavily on the author's previous work [19]. Here we aim for a self-contained, more elementary treatment of the problem.

During the review process of the present work, we became aware of the preprint [22], which also considers regularization of the Kantorovich problem with Young's functions. While the underlying domains  $\Omega_i$  are chosen to be general complete separable metric spaces, only regularization with respect to the marginal measures is considered and this allows the authors to derive existence of dual solutions independent of the form of the regularization.

Moreover, in [26] regularization with general convex functionals  $F : \mathfrak{M}(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}$  is considered. A duality result similar to our [Theorem 3.1](#) is derived and existence of primal solutions is covered. However, [26, Theorem 2] implies the existence of *continuous* dual optimizers. As [Example 3.7](#) below demonstrates, this can not be the case in general.

## 1.1. Notation and problem statement

Let us first fix some notation before we formulate our problem. The spaces of Radon and probability measures on  $\Omega \subset \mathbb{R}^n$  will be denoted by  $\mathfrak{M}(\Omega)$  and  $\mathcal{P}(\Omega)$ , respectively. The cone of non-negative

Radon measures will be denoted by  $\mathfrak{M}_+(\Omega)$ . With  $C(\Omega)$  and  $C_b(\Omega)$  we denote the spaces of continuous functions and bounded, continuous functions, respectively. The Lebesgue measure will be denoted by  $\mathcal{L}$  and integrals w.r.t. the Lebesgue measure are simply denoted by  $dx$  with the appropriate integration variable  $x$ .

In the following we will consider compact domains  $\Omega_1, \Omega_2$  equipped with finite measures  $\lambda_1 \in \mathfrak{M}_+(\Omega_1)$  and  $\lambda_2 \in \mathfrak{M}_+(\Omega_2)$ , respectively. The measures  $\lambda_1$  and  $\lambda_2$  will be assumed to have full support, i.e.  $\text{spt } \lambda_i = \Omega_i$ , for  $i = 1, 2$ . We will denote  $\Omega := \Omega_1 \times \Omega_2$  and  $\lambda := \lambda_1 \otimes \lambda_2$ . For the space of  $p$ -integrable functions on  $\Omega$  with respect to the measure  $\nu$ , the symbol  $L^p(\Omega, d\nu)$  will be used. When a measure  $\nu$  is absolutely continuous with respect to another measure  $\mu$ , written as  $\nu \ll \mu$ , the Radon-Nikodym derivative of  $\nu$  w.r.t. to  $\mu$ , i.e. the density of  $\nu$  w.r.t  $\mu$ , will be denoted by  $\frac{d\nu}{d\mu}$ .

The characteristic function of a set  $A$  will be denoted by  $\mathbb{1}_A$ . For two functions  $f : \Omega_1 \rightarrow \mathbb{R}$  and  $g : \Omega_2 \rightarrow \mathbb{R}$ , denote by  $f \oplus g : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ ,  $(x_1, x_2) \mapsto f(x_1) + g(x_2)$  the outer sum of  $f$  and  $g$ . This notation generalizes to measures  $\nu_1, \nu_2$  on  $\Omega_1, \Omega_2$ , respectively, by  $\nu_1 \oplus \nu_2 := \nu_1 \otimes \mathcal{L} + \mathcal{L} \otimes \nu_2$ . For  $\nu \in \mathfrak{M}(\Omega_1)$  and  $f : \Omega_1 \rightarrow \Omega_2$ , the pushforward of  $\nu$  by  $f$  will be denoted as  $f_{\#}\nu$ , i.e. the measure on  $\Omega_2$  defined by  $f_{\#}\nu(A) := \nu(f^{-1}(A))$ . Most importantly, the pushforward of the coordinate projections  $P_i : \Omega_1 \times \Omega_2 \rightarrow \Omega_i$ ,  $P_i(x_1, x_2) = x_i$  will be used. Note, that  $(P_i)_{\#}\pi$  is the  $i$ -th marginal of  $\pi \in \mathfrak{M}(\Omega_1 \times \Omega_2)$ . For real-valued functions  $f$  we denote by  $f_+ := \max(f, 0)$  the positive part and by  $f_- := -\min(f, 0)$  the negative part. Finally, for a function  $g : [0, \infty) \rightarrow \mathbb{R} \cup \{\infty\}$  denote by  $g_{\infty}$  its extension to the real line by infinity, i.e.

$$g_{\infty}(x) := \begin{cases} g(x), & x \geq 0, \\ \infty, & \text{else.} \end{cases}$$

The Orlicz space regularized Kantorovich problem of optimal transport considered in this work now reads as

$$\inf_{\substack{\pi \in \mathcal{P}(\Omega), \pi \ll \lambda \\ (P_i)_{\#}\pi = \mu_i, i=1,2}} \int_{\Omega} c \, d\pi + \gamma \int_{\Omega} \Phi_{\infty}\left(\frac{d\pi}{d\lambda}\right) \, d\lambda, \quad (\text{P})$$

where  $\Phi$  is a so-called *Young's function*. Note that the regularization of  $\pi$  is employed w.r.t. some product measure  $\lambda$ . Important cases are  $\lambda = \mathcal{L}$  and  $\lambda = \mu_1 \otimes \mu_2$ . Note also that  $\pi$  is required to be absolutely continuous w.r.t.  $\lambda$ . This is due to the fact that even for Young's functions  $\Phi$  satisfying modest conditions like  $\frac{\Phi(t)}{t} \not\rightarrow \infty$ , as  $t \rightarrow \infty$ , by e.g. [16, Theorem 5.19] the optimal  $\pi$  may have a singular part w.r.t.  $\lambda$ . That however, would make the process of regularizing futile. Therefore we will require  $\lim_{t \rightarrow \infty} \Phi(t)/t = \infty$  throughout the paper.

Now consider the marginal constraints. As we will see later in [Lemma 2.9](#) and [Theorem 3.4](#), it is necessary that the marginals  $\mu_i$  are absolutely continuous with respect to  $\lambda_i$  and that  $\Phi(\frac{d\mu_i}{d\lambda_i})$  is integrable with respect to  $\lambda_i$ . To formulate the marginal constraints in terms of the densities, we recall that  $(P_1)_{\#}\pi = \mu_1$  if for all  $\lambda_1$ -measurable sets  $A$  it holds that  $\pi(A \times \Omega_2) = \mu_1(A)$ . In terms of densities and integrals, this reads as

$$\int_{\Omega_2} \int_A \frac{d\pi}{d\lambda} \, d\lambda_1 \, d\lambda_2 = \int_A d\mu_1 = \int_A \frac{d\mu_1}{d\lambda_1} \, d\lambda_1.$$

Using Fubini's theorem we get

$$\int_A \left( \int_{\Omega} \frac{d\pi}{d\lambda} \, d\lambda_2 - \frac{d\mu_1}{d\lambda_1} \right) \, d\lambda_1 = 0,$$

and hence, the marginal constraints read as

$$\int_{\Omega_2} \frac{d\pi}{d\lambda} d\lambda_2 = \frac{d\mu_1}{d\lambda_1} \quad \lambda_1\text{-a.e.} \quad \text{and} \quad \int_{\Omega_1} \frac{d\pi}{d\lambda} d\lambda_1 = \frac{d\mu_2}{d\lambda_2} \quad \lambda_2\text{-a.e.}$$

Note that for the integral  $\int_{\Omega} c d\pi = \int_{\Omega} c \frac{d\pi}{d\lambda} d\lambda$  to exist, the cost function  $c$  does not need to be continuous and the problem may be formulated for more general cost functions. However, some of the results in this work require  $c$  to be continuous and for simplicity this shall be assumed throughout the paper.

Let us summarize our assumptions:

**Assumption 1.1.** For  $i = 1, 2$  let  $\Omega_i \subset \mathbb{R}^n$  be compact domains equipped with finite measures  $\lambda_i \in \mathfrak{M}_+(\Omega_i)$  with  $\text{spt } \lambda_i = \Omega_i$ . Let  $c : \Omega \rightarrow \mathbb{R}$  be a continuous cost function. Let  $\Phi$  be such that  $\lim_{t \rightarrow \infty} \Phi(t)/t = \infty$ . Finally, for  $i = 1, 2$  let  $\mu_i \in \mathcal{P}(\Omega_i)$  such that  $\mu_i \ll \lambda_i$  and  $\Phi(\frac{d\mu_i}{d\lambda_i})$  is integrable with respect to  $\lambda_i$ .

## 1.2. Contribution and Organization

The notions of Young's functions and Orlicz spaces are introduced in [Section 2](#) alongside some auxiliary results that will be used in the later sections. [Section 3](#) deals with the question of existence of solutions in the framework of Fenchel duality. The first contribution ([Theorem 3.4](#)) guarantees existence of solutions of problem (P), which generalizes the corresponding results of [[10](#), [20](#)]. Afterwards, the predual problem is analyzed for the special case  $\Phi(t) = t^p/p$  for  $p > 1$ . As second contribution [Theorem 3.13](#) gives existence of dual optimizers in  $L^q$ , where  $1/p + 1/q = 1$  and  $p \geq 2$ . This generalizes the corresponding result of [[20](#)]. In [Section 4](#)  $\Gamma$ -convergence of different related problems is considered. First, a continuous, regularized problem with arbitrary marginals is considered. [Theorem 4.2](#) extends [[10](#), [Theorem 5.1](#)] and guarantees  $\Gamma$ -convergence to the unregularized problem (OT) when smoothing the marginals. Note that the case of  $\Gamma$ -convergence for fixed marginals in  $L \log L(\Omega)$  has been treated in [[8](#), [Theorem 2.7](#)] for  $\Omega = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ . While [Theorem 4.2](#) is stated only for compact  $\Omega$ , it allows for marginals not in  $L \log L(\Omega)$  and a coupled reduction of the regularization and the smoothing parameter. The final contribution ([Theorem 4.9](#)) is the proof of  $\Gamma$ -convergence of a discretized and regularized optimal transport problem to the unregularized continuous problem (OT). The result covers both entropic and quadratic regularization as special cases. Some of the results in this paper are a direct generalization of results of previous papers and their proofs also follow the general proof strategy, but we still include these results and sketches of the proofs for the sake of completeness.

## 2. Young's functions and Orlicz spaces

In this section, some notions about Young's functions and Orlicz spaces are introduced. For a more detailed introduction, see [[3](#), [28](#)].

**Definition 2.1** (Young's function [[3](#), [Definitions IV.8.1, IV.8.11](#)]). *i) Let  $\varphi : [0, \infty) \rightarrow [0, \infty]$  be increasing and lower semi-continuous, with  $\varphi(0) = 0$ . Suppose that  $\varphi$  is neither identically zero nor identically infinite on  $(0, \infty)$ . Then the function  $\Phi$  defined by  $\Phi(t) := \int_0^t \varphi(s) ds$  is said to be a Young's function.*

ii) Let  $\psi(s) := \inf \{t \mid \varphi(t) \geq s\}$ . Then, the function  $\Psi$  defined by  $\Psi(t) := \int_0^t \psi(s) ds$  is said to be the complementary Young's function of  $\Phi$ .

By definition, Young's functions are convex and for a Young's function  $\Phi$  it holds that the complementary Young's function  $\Psi$  is also a Young's function and actually equal to the convex conjugate  $\Phi^*$ .

The negative entropy regularization uses the function  $\Phi(t) = t \log(t)$  which is not a Young's function, but the function  $t \mapsto (t \log(t))_+$  is. Hence, we introduce a slight generalization of the notion of Young's function to be able to treat this case as well.

**Definition 2.2** (Quasi-Young's functions). *We say that  $\check{\Phi}$  is a quasi-Young's function if it is convex, lower semi-continuous and  $\Phi(t) := \max(0, \check{\Phi}(t))$  is a Young's function. In this case we say that  $\check{\Phi}$  is associated with  $\Phi$ .*

Note that convexity of  $\Phi$  shows that  $\check{\Phi}$  is also bounded from below.

**Example 2.3.** *The function  $\check{\Phi}(t) = t \log(t)$  is a quasi-Young's function associated with the Young's function  $\Phi(t) = (t \log(t))_+$ .*

**Definition 2.4** (Luxemburg and Orlicz spaces [3, Definition IV.8.10]). *Let  $\Phi$  be a Young's function,  $\Omega \subset \mathbb{R}^n$  and  $\nu \in \mathfrak{M}_+(\Omega)$ . Define the Luxemburg norm of a measurable function  $f : \Omega \rightarrow \mathbb{R}$  w.r.t.  $\nu$  as*

$$\|f\|_{L^\Phi(\Omega, d\nu)} := \inf \left\{ \gamma \geq 0 \mid \int_{\Omega} \Phi \left( \frac{|f|}{\gamma} \right) d\nu \leq 1 \right\}.$$

Then the space

$$L^\Phi(\Omega, d\nu) := \left\{ f : \Omega \rightarrow \mathbb{R} \text{ measurable} \mid \|f\|_{L^\Phi(\Omega, d\nu)} < \infty \right\}$$

of measurable functions on  $\Omega$  with finite Luxemburg norm is called the Orlicz space of  $\Phi$  w.r.t.  $\nu$ .

**Remark 2.5** ([9, Remark 1]). *The bound 1 in the definition of the Luxemburg norm can be replaced by any  $a \in (0, \infty)$ . That is, all norms defined by*

$$\|f\|_{L^\Phi(\Omega, d\nu), a} := \inf \left\{ \gamma \geq 0 \mid \int_{\Omega} \Phi \left( \frac{|f|}{\gamma} \right) d\nu \leq a \right\}$$

are equivalent. This can be seen by combining the inequalities  $\|\cdot\|_{L^\Phi(\Omega, d\nu), b} \leq \|\cdot\|_{L^\Phi(\Omega, d\nu), a}$  and  $a\|\cdot\|_{L^\Phi(\Omega, d\nu), a} \leq b\|\cdot\|_{L^\Phi(\Omega, d\nu), b}$  for  $0 < a < b$ .

The definition of Orlicz spaces does not immediately allow for the concept of quasi-Young's functions to be incorporated. However, the following results establish the desired connection.

**Lemma 2.6.** *Let  $\Phi$  be a Young's function and  $\Omega \subset \mathbb{R}^n$ ,  $\nu \in \mathfrak{M}_+(\Omega)$  with  $\nu(\Omega) < \infty$  and let  $t_0 \geq 0$ . Then  $\check{\Phi}$  defined by  $\check{\Phi}(t) := \max\{0, \Phi(t) - \Phi(t_0)\}$  is a Young's function and  $L^\Phi(\Omega, d\nu) = L^{\check{\Phi}}(\Omega, d\nu)$ .*

A proof of [Lemma 2.6](#) can be found in [Appendix A](#).

**Corollary 2.7.** Let  $\Phi$  be a Young's function,  $\check{\Phi}$  a quasi-Young's function associated with  $\Phi$ ,  $\nu$  a measure and  $\Omega \subset \mathbb{R}^n$  with  $\nu(\Omega) < \infty$ . Then,  $f \in L^\Phi(\Omega, d\nu)$  if and only if

$$\int_{\Omega} \check{\Phi} \left( \frac{|f|}{\gamma} \right) d\nu \leq 1$$

for some  $\gamma < \infty$ .

*Proof.* Let  $t_0 = \inf\{t \mid \check{\Phi}(t) \geq 0 \forall \tau \geq t\}$  and  $\check{\Phi}(t) := \max\{0, \Phi(t) - \Phi(t_0)\}$ . Let  $\gamma$  be such that

$$\int_{\Omega} \check{\Phi} \left( \frac{|f(x)|}{\gamma} \right) d\nu(x) \leq 1.$$

Then

$$1 \geq \int_{\Omega} \check{\Phi} \left( \frac{|f(x)|}{\gamma} \right) d\nu(x) \geq \int_{\Omega} \check{\Phi} \left( \frac{|f(x)|}{\gamma} \right) d\nu(x)$$

together with [Lemma 2.6](#), shows one implication.

For the other implication, by [Remark 2.5](#) it suffices to show that  $\int_{\Omega} \check{\Phi} (|f(x)|/\gamma) d\nu(x) < \infty$  whenever  $\int_{\Omega} \check{\Phi} (|f(x)|/\gamma) d\nu(x) < \infty$ . However, this holds trivially, since  $\check{\Phi}$  is bounded from below and  $\Omega$  has finite measure w.r.t.  $\nu$ .  $\square$

[Lemma 2.6](#) and [Corollary 2.7](#) state that the definitions of  $\|\cdot\|_{L^\Phi(\Omega, d\nu)}$  and  $L^\Phi(\Omega, d\nu)$  are essentially independent of whether  $\Phi$  is a Young's function or just a quasi-Young's function. To simplify notation,  $\|\cdot\|_{L^{\check{\Phi}}(\Omega, d\nu)}$  and  $L^{\check{\Phi}}(\Omega, d\nu)$  will therefore be used for quasi-Young's functions  $\check{\Phi}$  as well. Note that for a quasi-Young's function  $\check{\Phi}$  associated with a Young's function  $\Phi$ ,  $L^{\check{\Phi}}(\Omega, d\nu) = L^\Phi(\Omega, d\nu)$ , while in general  $\|\cdot\|_{L^{\check{\Phi}}(\Omega, d\nu)}$  and  $\|\cdot\|_{L^\Phi(\Omega, d\nu)}$  are equivalent but not equal. Moreover for complementary Young's functions  $\Phi$  and  $\Psi = \Phi^*$  which are proper, locally integrable and satisfy a certain growth condition, called the  $\Delta_2$ -property near infinity, it holds that  $(L^\Phi(\Omega, d\nu))^*$  is canonically isometrically isomorphic to  $(L^\Psi(\Omega, d\nu), \|\cdot\|_{L^\Psi(\Omega, d\nu)})$  (see, e.g. [14]).

**Example 2.8** ( $L \log L$  and  $L_{\exp}$ ). Let  $\Phi(t) = t \log t$  and  $\check{\Phi}(t) = (t \log(t))_+$ . The space of measurable functions  $f$  with  $\int_{\Omega} \check{\Phi}(|f|) d\nu < \infty$  is called  $L \log L(\Omega, d\nu)$ . By the above corollary, the space of measurable functions  $g$  with  $\int_{\Omega} \Phi(|g|) d\nu < \infty$  is equal to  $L \log L(\Omega, d\nu)$ . The complementary Young's function  $\check{\Psi}$  of  $\check{\Phi}$  is given by

$$\check{\Psi}(t) = \begin{cases} t, & t \leq 1 \\ e^{t-1}, & \text{else.} \end{cases}$$

As  $\Phi$  satisfies the  $\Delta_2$ -property near infinity, the dual space of  $L \log L(\Omega, d\nu)$  is thus given by the space of measurable functions  $h$  that satisfy  $\int_{\Omega} \check{\Psi}(|h|) d\nu < \infty$ , which is called  $L_{\exp}(\Omega, d\nu)$ .

The following result states that the marginals of a transport plan with density in  $L^\Phi$  also have density in the respective  $L^\Phi$  space.

**Lemma 2.9.** Let  $\nu_i \in \mathfrak{M}_+(\Omega)$  such that  $\nu_i(\Omega_i) < \infty$ , for  $i = 1, 2$  and set  $\nu := \nu_1 \otimes \nu_2$ . Let  $\pi \in \mathfrak{M}_+(\Omega)$ .

If  $\frac{d\pi}{d\nu} \in L^\Phi(\Omega, d\nu)$  for a quasi-Young's function  $\Phi$ , then  $\frac{d(P_i)_\# \pi}{d\nu_i} \in L^\Phi(\Omega_i, d\nu_i)$  for  $i = 1, 2$  with

$$\left\| \frac{d(P_i)_\# \pi}{d\nu_i} \right\|_{L^\Phi(\Omega_i, d\nu_i)} \leq \max(1, \nu_{3-i}(\Omega_{3-i})) \left\| \frac{d\pi}{d\nu} \right\|_{L^\Phi(\Omega, d\nu)}.$$



*Proof.* Setting  $\frac{d\pi}{d(v_1 \otimes v_2)}(x_1, x_2) = f(x_1, x_2)$ , one observes, that  $\frac{d(P_2)_{\#}\pi}{dv_2} = \int_{\Omega_1} f(x_1, x_2) dv_1$ , and similarly for  $\frac{d(P_1)_{\#}\pi}{dv_1}$ . The proof given for [10, Lemma 2.11] then holds with minor modifications.  $\square$

**Remark 2.10.** *The above result immediately yields that no optimal solution  $\bar{\pi}$  of problem (P) can exist for any  $\gamma > 0$ , if  $\mu_i \not\ll \lambda_i$ .*

Next, a few facts are derived, which will be useful for the analysis of both the primal and dual regularized optimal transport problems and  $\Gamma$ -convergence.

**Lemma 2.11.** *Let  $\Phi$  be a quasi-Young's function and  $f$  such that  $\|f\|_{L^\Phi(\Omega, dv)} > 1$ . Then,  $\int_{\Omega} \Phi(|f|) dv \geq \|f\|_{L^\Phi(\Omega, dv)}$ .*

*Proof.* Noting that  $\Phi(0) \leq 0$ , the proof given for [10, Lemma 2.6] holds with minor adjustments.  $\square$

For complementary Young's functions  $\Phi$  and  $\Psi$ , we have the following result on conjugating scaled Young's functions.

**Lemma 2.12.** *Let  $\Phi$  be a Young's function,  $\Psi = \Phi^*$  and  $\gamma \neq 0$ .*

1. *Then for  $\tilde{\Psi}(t) := \gamma\Psi(t/\gamma)$  it holds that  $\tilde{\Psi}^* = \gamma\Phi$ .*
2. *Let  $\check{\Phi}$  be a quasi-Young's function associated with  $\Phi$ . Then for  $\check{\Phi}^*(t) := \gamma\check{\Phi}^*(t/\gamma)$  it holds that  $(\check{\Phi}^*)^* = \gamma\check{\Phi}$ .*

*Proof.* The assertion follows directly from the definition of the convex conjugate and the Fenchel-Moreau theorem.  $\square$

The following Lemma 2.13 gives some useful insights into the behavior of the objective function of problem (P) under perturbation of  $\pi$ . Recall that for Young's functions  $\Phi$  we required  $\lim_{t \rightarrow \infty} \Phi(t)/t = \infty$ .

**Lemma 2.13.** *Let  $v \in \mathfrak{M}_+(\Omega)$ ,  $(\pi_k) \subset \mathcal{P}(\Omega)$  and  $\pi \in \mathcal{P}(\Omega)$  such that  $\pi_k \xrightarrow{*} \pi$  and  $\pi_k \ll v$ . Let  $g := \Phi_\infty$  for some quasi-Young's function  $\Phi$ . Then the following statements hold.*

1. *Let  $\pi \not\ll v$ . Then*

$$\liminf_{k \rightarrow \infty} \int_{\Omega} g \left( \frac{d\mu_k}{dv} \right) dv = \infty.$$

2. *Let  $\pi \ll v$ . Then*

$$\liminf_{k \rightarrow \infty} \int_{\Omega} g \left( \frac{d\mu_k}{dv} \right) dv \geq \int_{\Omega} g \left( \frac{d\mu}{dv} \right) dv.$$

*Proof.* Since  $g$  grows superlinearly at  $\infty$ , the recession function  $g^\infty(t) = \lim_{h \rightarrow \infty} \frac{g(s+ht) - g(s)}{h}$  (which is independent of  $s$ ) is infinite for all  $t > 0$ . By [16, Theorem 5.19], it holds for every sequence  $\pi_k$  which weakly\* converges to  $\pi$  that

$$\liminf_{k \rightarrow \infty} \int_{\Omega} g\left(\frac{d\mu_k}{d\nu}\right) d\nu \geq \int_{\Omega} g\left(\frac{d\pi}{d\nu}\right) d\nu + \int_{\Omega} g^\infty\left(\frac{d\pi}{d|\pi_s|}\right) d|\pi_s|,$$

where  $\pi_s$  denotes the unique measure singular to  $\nu$  in the Lebesgue decomposition  $\pi = \pi_s + \pi_{ac}$  (e.g. [16, Theorem 1.115]) and  $\pi_{ac}$  denotes the corresponding measure with  $\pi_{ac} \ll \nu$ .

1. It suffices to show  $\frac{d\pi}{d|\pi_s|}(x) > 0$  for every  $x \in \text{spt}|\pi_s|$ . First note that since  $\pi \in \mathcal{P}(\Omega)$ ,  $\pi_s$  is non-negative and hence  $|\pi_s| = \pi_s$ . Let now  $C$  be a bounded, convex closed set containing the origin in its interior. Then by [16, Definition 1.156], for every  $x \in \text{spt} \pi_s$  it holds that

$$\begin{aligned} \frac{d\pi}{d\pi_s}(x) &= \lim_{r \searrow 0} \frac{\pi((x+rC) \cap \Omega)}{\pi_s((x+rC) \cap \Omega)} \\ &= \lim_{r \searrow 0} \frac{\pi_{ac}((x+rC) \cap \Omega) + \pi_s((x+rC) \cap \Omega)}{\pi_s((x+rC) \cap \Omega)} \\ &= \lim_{r \searrow 0} \frac{\pi_{ac}((x+rC) \cap \Omega)}{\pi_s((x+rC) \cap \Omega)} + 1 \\ &\geq 1, \end{aligned}$$

since  $\pi_s((x+rC) \cap \Omega) > 0$  for all  $r$  because of  $x \in \text{spt} \pi_s$ . In fact,

$$\lim_{r \searrow 0} \pi_{ac}((x+rC) \cap \Omega) = 0$$

and  $\frac{d\pi}{d|\pi_s|}(x) = 1$  for every  $x \in \text{spt}|\pi_s|$ .

2. The second statement follows directly, since for  $\pi \ll \nu$  it holds that  $\pi_s = 0$ . □

### 3. Existence of Solutions

In this section, we show strong duality for the regularized mass transport (P) using Fenchel duality in the spaces  $\mathcal{P}(\Omega)$  and  $C(\Omega)$ . The result will then be used to study the question of existence of solutions for both the primal and the dual problem.

**Theorem 3.1** (Strong duality). *Let  $\Phi$  be a quasi-Young's function and Assumption 1.1 hold. If  $(\Phi_\infty)^*(-c/\gamma)$  is integrable w.r.t.  $\lambda$ , then the predual problem to (P) is*

$$\sup_{\substack{\alpha_i \in C(\Omega_i), \\ i=1,2}} \int_{\Omega_1} \alpha_1 d\mu_1 + \int_{\Omega_2} \alpha_2 d\mu_2 - \gamma \int_{\Omega} (\Phi_\infty)^* \left( \frac{\alpha_1 \oplus \alpha_2 - c}{\gamma} \right) d\lambda \quad (\text{P}^*)$$

and strong duality holds. Furthermore, if the supremum is finite, (P) possess a minimizer.

*Proof.* Strong duality holds by standard arguments (see e.g. [6, Theorem 4.4.3]) and (assuming a finiteness of the supremum) the primal problem (P) possesses a minimizer.

To derive the dual problem, we start from the primal problem, express the equality conditions  $\int_{\Omega_2} \frac{d\pi}{d\lambda} d\lambda_2 = \frac{d\mu_1}{d\lambda_1}$  and  $\int_{\Omega_1} \frac{d\pi}{d\lambda} d\lambda_1 = \frac{d\mu_2}{d\lambda_2}$  as suprema over continuous functions and get

$$\begin{aligned} & \inf_{\substack{\frac{d\pi}{d\lambda} \in L^\Phi(\Omega, d\lambda), \\ \int_{\Omega_2} \frac{d\pi}{d\lambda} d\lambda_2 = \frac{d\mu_1}{d\lambda_1}, \\ \int_{\Omega_1} \frac{d\pi}{d\lambda} d\lambda_1 = \frac{d\mu_2}{d\lambda_2}}} \int_{\Omega} c \frac{d\pi}{d\lambda} + \gamma \tilde{\Phi}\left(\frac{d\pi}{d\lambda}\right) d\lambda \\ &= \inf_{\frac{d\pi}{d\lambda} \in L^\Phi(\Omega, d\lambda)} \sup_{\substack{\alpha_i \in C(\Omega_i), \\ i=1,2}} \int_{\Omega} c \frac{d\pi}{d\lambda} + \gamma \tilde{\Phi}\left(\frac{d\pi}{d\lambda}\right) d\lambda + \int_{\Omega_1} \left(\frac{d\mu_1}{d\lambda_1} - \int_{\Omega_2} \frac{d\pi}{d\lambda} d\lambda_2\right) \alpha_1 d\lambda_1 + \int_{\Omega_2} \left(\frac{d\mu_2}{d\lambda_2} - \int_{\Omega_1} \frac{d\pi}{d\lambda} d\lambda_1\right) \alpha_2 d\lambda_2 \\ &= \sup_{\substack{\alpha_i \in C(\Omega_i), \\ i=1,2}} \inf_{\frac{d\pi}{d\lambda} \in L^\Phi(\Omega, d\lambda)} \left( \int_{\Omega} (c - \alpha_1 \oplus \alpha_2) \frac{d\pi}{d\lambda} + \gamma \tilde{\Phi}\left(\frac{d\pi}{d\lambda}\right) d\lambda \right) + \int_{\Omega_1} \alpha_1 d\mu_1 + \int_{\Omega_2} \alpha_2 d\mu_2 \end{aligned}$$

The integrand of the first integral in (P\*) is normal, so that it can be conjugated pointwise [30, Theorem 2]. Carrying out the conjugation with the help of Lemma 2.12, one obtains the claim.  $\square$

**Example 3.2.** 1. Using  $\Phi(t) = t \log t$  and  $\lambda_i = \mathcal{L}_i$ , one obtains the result for  $L \log L$  as stated in [10, Theorem 3.1]. In this case it holds that  $(\Phi_\infty)^*(r) = \exp(r) = \Phi^*(r)$ .

2. Using  $\Phi(t) = \frac{1}{2}t^2$  and  $\lambda_i = \mathcal{L}_i$ , one obtains the result for  $L^2$  as stated in [20]. In this case it holds that  $(\Phi_\infty)^*(r) = \max(0, r)^2$ .

One can show that in general  $(\Phi_\infty)^*(r) = \Phi^*(r)$  for  $r \geq \inf_{\tau > 0} \partial\Phi(\tau)$  and equal to  $-\Phi(0)$  otherwise.

**Remark 3.3.** Theorem 3.1 does not claim that the supremum is attained, i.e. that the predual problem (P\*) admits a solution. Moreover, the solutions of (P\*) cannot be unique since one can add and subtract constants to  $\alpha_1$  and  $\alpha_2$ , respectively, without changing the functional value. If however  $(\Phi_\infty)^*$  is strictly convex, then the functional in (P\*) is strictly concave up to such constant and therefore any solution is uniquely determined by this constant. This is the case, e.g., for functions  $\Phi$  with superlinear growth at  $\infty$ .

### 3.1. Existence result for the primal problem

The duality result can now be used to address the question of existence of a solution to (P).

**Theorem 3.4** (Existence of solutions of (P)). *Let Assumption 1.1 hold. If  $(\Phi_\infty)^*(-c/\gamma)$  is integrable w.r.t.  $\lambda$  and*

$$\frac{d\mu_i}{d\lambda_i} \in L^\Phi(\Omega_i, d\lambda_i), \quad i = 1, 2 \implies \frac{d(\mu_1 \otimes \mu_2)}{d\lambda} \in L^\Phi(\Omega, d\lambda) \quad (3.1)$$

*we have that problem (P) admits a minimizer  $\bar{\pi}$  if and only if  $\frac{d\mu_i}{d\lambda_i} \in L^\Phi(\Omega_i, d\lambda_i)$  for  $i = 1, 2$ . In this case,  $\bar{\pi} \in L^\Phi(\Omega, d\lambda)$  and the minimizer is unique, if  $\Phi$  is strictly convex.*

*Proof.* The proof given in [10, Theorem 3.3] for  $\Phi(t) = t \log t$  holds for arbitrary  $\Phi$  and arbitrary product measures  $\lambda = \lambda_1 \otimes \lambda_2$ , since it only relies on Lemma 2.9 and condition (3.1).  $\square$

**Example 3.5.** Since  $\frac{d(\mu_1 \otimes \mu_2)}{d\lambda} = \frac{d\mu_1}{d\lambda_1} \otimes \frac{d\mu_2}{d\lambda_2}$ , condition (3.1) is satisfied e.g. when  $\Phi$  satisfies either  $\Phi(xy) \leq C\Phi(x)\Phi(y)$  for some  $C > 0$  or  $\Phi(xy) \leq C_1x\Phi(y) + C_2\Phi(x)y$  for some  $C_1, C_2 \geq 0$ . For  $\Phi(t) = t^p/p$ ,  $p > 1$ , both conditions hold trivially. For  $\Phi(t) = t \log t$  the second condition holds, since  $\log(xy) = \log(x) + \log(y)$ .

### 3.2. Existence result for the predual problem with $L^p(\Omega, d\lambda)$ regularization

The question of existence of solutions to the predual problem (P\*) proves to be difficult for general Young's functions. There are results that show existence for the predual problem in the entropic case [10] and in the quadratic case [20] (both considering the penalty w.r.t. the Lebesgue measure), but their proofs are quite different in nature. In the case where  $\lambda = \mu_1 \otimes \mu_2$  (i.e. the product of the marginals) and  $\Phi(t) = t \log(t)$ , one can show dual existence in a different way, see [17] for a proof based on the convergence of the Sinkhorn method and [32] for a sketch of a proof that uses regularity of the cost function. Our methods do not allow to show existence of solutions for the dual problem in the general case considered up to now. Hence, we consider a special case in the following: We specialize to the case of  $L^p(\Omega, d\lambda)$ -regularization for a measure  $\lambda = \lambda_1 \otimes \lambda_2$  with  $\mu_i \ll \lambda_i$ ,  $i = 1, 2$ . That is, we use the Young's functions  $\Phi(t) = t^p/p$  for  $p > 1$ , and thus,  $\Phi^*(t) = t^q/q$ , with  $1/p + 1/q = 1$ , and the predual is actually also the dual. Moreover,  $(\Phi_\infty)^*(s) = \frac{1}{q}(s_+)^q$ . To keep notation clean,  $(t_+)^p$  will be abbreviated as  $t_+^p$  with slight abuse of notation and similarly for  $(t_-)^p$ .

**Assumption 3.6.** Let Assumption 1.1 hold. In addition, let the cost function  $c$  be continuous and fulfill  $c \geq c^\dagger$  for some constant  $c^\dagger > -\infty$ . Furthermore, let the marginals  $\mu_i$  with  $\frac{d\mu_i}{d\lambda_i} \in L^p(\Omega_i, d\lambda_i)$  satisfy  $\frac{d\mu_i}{d\lambda_i} \geq \delta > 0$   $\lambda_i$ -a.e. for  $i = 1, 2$ .

Note that the latter condition, i.e. that the densities should be bounded away from zero, can be guaranteed by proper choice of  $\lambda_i$ , e.g.  $\lambda_i = \mu_i$  gives  $\frac{d\mu_i}{d\lambda_i} = 1$ . It can not be expected for problem (P\*) to have continuous optimizers  $\alpha_1, \alpha_2$ , as the following example demonstrates.

**Example 3.7.** For  $i = 1, 2$  let  $\Omega_i = [-1, 1]$ ,  $\lambda_i = \mathcal{L}|_{\Omega_i} + \delta_0$ , where  $\delta_0$  denotes the Dirac measure at zero, and  $\mu_i = \delta_0$ . Moreover, let  $c \equiv 0$ . Clearly the minimum in problem (P) is  $\frac{1}{p}$ , which is attained for  $\pi = \delta_0 \times \delta_0$ . By Theorem 3.1, the optimal value of problem (P\*) is  $\frac{1}{p}$ , as well. Going on, it holds

$$\begin{aligned} & \sup_{\alpha_i \in C(\Omega)} \int_{\Omega_1} \alpha_1 d\mu_1 + \int_{\Omega_2} \alpha_2 d\mu_2 - \frac{1}{q} \int_{\Omega} (\alpha_1 \oplus \alpha_2)_+^q d\lambda \\ &= \sup_{\alpha_i \in C(\Omega)} \alpha_1(0) + \alpha_2(0) - \frac{1}{q} (\alpha_1(0) + \alpha_2(0))_+^q - \frac{1}{q} \int_{\Omega} (\alpha_1 \oplus \alpha_2)_+^q d\mathcal{L} \end{aligned}$$

If  $\alpha_1 \oplus \alpha_2 \leq 0$ , the supremum ( $\frac{1}{p}$ ) can clearly not be attained. Hence, assume  $(\alpha_1 \oplus \alpha_2)(x) > 0$  for some  $x \in \Omega$ , which implies  $\int_{\Omega} (\alpha_1 \oplus \alpha_2)_+^q d\mathcal{L} > 0$ , as  $\alpha_i \in C(\Omega)$ . Thus,

$$\begin{aligned} & \alpha_1(0) + \alpha_2(0) - \frac{1}{q} (\alpha_1(0) + \alpha_2(0))_+^q - \frac{1}{q} \int_{\Omega} (\alpha_1 \oplus \alpha_2)_+^q d\mathcal{L} \\ & < (\alpha_1 \oplus \alpha_2)(0) - \frac{1}{q} ((\alpha_1 \oplus \alpha_2)(0))_+^q \leq 1 - \frac{1}{q} = \frac{1}{p}, \end{aligned}$$

where equality holds for  $(\alpha_1 \oplus \alpha_2)(0) = 1$ . Due to the strict inequality, the supremum in problem (P\*) can not be attained for  $\alpha_i \in C(\Omega)$ ,  $i = 1, 2$ .

However, with the help of the following result, we can define a variant of the predual problem, for which existence of minimizers can be shown.

**Lemma 3.8.** *Let  $\alpha_i : \Omega_i \rightarrow \mathbb{R} \cup \{\pm\infty\}$ ,  $i = 1, 2$  such that  $(\alpha_1 \oplus \alpha_2 - c)_+ \in L^q(\Omega, d\lambda)$ . Then  $(\alpha_i)_+ \in L^q(\Omega_i, d\lambda_i)$ ,  $i = 1, 2$ .*

*Proof.* First note, that

$$\begin{aligned} (\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c + c)_+^q &\leq ((\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+ + c_+)^q = 2^q \left( \frac{(\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+}{2} + \frac{c_+}{2} \right)^q \\ &\leq 2^{q-1} ((\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+^q + c_+^q), \end{aligned}$$

so that  $(\alpha_1 \oplus \alpha_2)_+ \in L^q(\Omega, d\lambda)$ . This implies  $(\alpha_i)_+ < \infty$   $\lambda_i$ -a.e. for  $i = 1, 2$ .

We now consider  $(\alpha_i)_+$ ,  $i = 1, 2$  separately and start with  $i = 1$ . Let  $M \subset \Omega_2$  be such that  $\alpha_2 \geq 0$   $\lambda_2$ -a.e. on  $M$ . If  $\lambda_2(M) = 0$  the assertion holds trivially, so we assume  $\lambda_2(M) > 0$ . For  $g \in L^p(\Omega_2, d\lambda_2)$  denote  $C_g := \int_M g d\lambda_2$ . Then,

$$\begin{aligned} \|(\alpha_1)_+\|_q &= \sup \left\{ \int_{\Omega_1} (\alpha_1)_+ g_1 d\lambda_1 \mid g_1 \in L^p(\Omega_1, d\lambda_1), \|g_1\|_p = 1 \right\} \\ &= \sup \left\{ \int_M \int_{\Omega_1} (\alpha_1)_+ (g_1 \otimes g_2) d\lambda_1 d\lambda_2 \mid g_1 \in L^p(\Omega_1, d\lambda_1), \|g_1\|_p = 1, g_2 = \mathbf{1}_M \lambda_2(M)^{-\frac{1}{p}} \right\} \\ &\leq \sup \left\{ \int_M \int_{\Omega_1} (\alpha_1 \oplus \alpha_2)_+ (g_1 \otimes g_2) d\lambda_1 d\lambda_2 \mid g_i \in L^p(\Omega_i, d\lambda_i), \|g_i\|_p = 1, i = 1, 2 \right\} \\ &\leq \sup \left\{ \int_{\Omega} (\alpha_1 \oplus \alpha_2)_+ f d\lambda \mid f \in L^p(\Omega, d\lambda), \|f\|_p = 1 \right\} \\ &= \|(\alpha_1 \oplus \alpha_2)_+\|_q < \infty, \end{aligned}$$

where the first inequality is justified because thanks to  $(\alpha_1)_+ \geq 0$  we can take the supremum over  $g_1 \geq 0$ . With the same argumentation, we can prove  $(\alpha_2)_+ \in L^q(\Omega_2, d\lambda_2)$ .  $\square$

Hence, the objective function of problem  $(P^*)$  is also well defined for functions  $\alpha_i \in L^1(\Omega_i, d\lambda_i)$ ,  $i = 1, 2$ , with  $(\alpha_1 \oplus \alpha_2 - c)_+ \in L^q(\Omega, d\lambda)$ . We now consider the following variant of the predual problem:

$$\begin{aligned} \min \left\{ \Lambda(\alpha_1, \alpha_2) := \frac{1}{q} \|(\alpha_1 \oplus \alpha_2 - c)_+\|_{L^q(\Omega, d\lambda)}^q - \gamma^{q-1} \int_{\Omega_1} \alpha_1 d\mu_1 - \gamma^{q-1} \int_{\Omega_2} \alpha_2 d\mu_2 \right. \\ \left. \mid \alpha_i \in L^1(\Omega_i, d\lambda_i), i = 1, 2, \frac{1}{\gamma} (\alpha_1 \oplus \alpha_2 - c)_+ \in L^q(\Omega, d\lambda) \right\}. \end{aligned} \quad (P^\dagger)$$

The strategy in this section is as follows.

1. First, show that problem  $(P^\dagger)$  admits a solution  $(\bar{\alpha}_1, \bar{\alpha}_2) \in L^1(\Omega_1, d\lambda_1) \times L^1(\Omega_2, d\lambda_2)$ .
2. Then, prove that  $\bar{\alpha}_1$  and  $\bar{\alpha}_2$  possess higher regularity, namely that they are functions in  $L^q(\Omega_i, d\lambda_i)$ .

The objective function needs to be extended to allow to deal with weak-\* converging sequences. To that end, define

$$G : L^q(\Omega, d\lambda) \ni w \mapsto \int_{\Omega} \left( \frac{1}{q} w_+^q - w\mu \right) d\lambda \in \mathbb{R},$$

where  $\mu := \gamma^{q-1} \frac{d(\mu_1 \otimes \mu_2)}{d\lambda}$ . Note that in the case  $\lambda_i = \mu_i$ , the variable  $\mu$  is given by  $1 \cdot \gamma^{q-1}$ . Then, thanks to the normalization of  $\mu_1$  and  $\mu_2$ ,

$$\Lambda(\alpha_1, \alpha_2) = G(\alpha_1 \oplus \alpha_2 - c) - \int_{\Omega} c\mu d\lambda \quad \forall \alpha_1, \alpha_2 \in L^q.$$

Of course,  $G$  is also well defined as a functional on the feasible set of problem (P<sup>†</sup>) and this functional will be denoted by the same symbol to ease notation. In order to extend  $G$  to the space of Radon measures, consider for a given measure  $w \in \mathfrak{M}_+(\Omega)$  the Hahn-Jordan decomposition  $w = w_+ - w_-$  and assume  $\frac{dw_+}{d\lambda} \in L^q(\Omega, d\lambda)$ . Then,  $\int_{\Omega} \frac{d(w_+\mu)}{d\lambda} d\lambda$  is finite for  $\mu$  as in [Assumption 3.6](#). Regarding the negative part, we define  $\int_{\Omega} \mu dw_- := \infty$ , whenever this expression is not properly defined, as  $w_-$  and  $\mu$  are both positive. Combining this, we always have  $-\int_{\Omega} \mu dw \in \mathbb{R} \cup \{\infty\}$  and define

$$G(w) := \int_{\Omega} \frac{1}{q} \frac{dw_+^q}{d\lambda} d\lambda - \int_{\Omega} \mu dw.$$

**Remark 3.9.** *If  $w \ll \lambda$ , then  $\frac{dw_+}{d\lambda} \in L^1(\Omega, d\lambda)$  and  $\frac{dw_+}{d\lambda}(x) = \max\{0, \frac{dw}{d\lambda}(x)\}$   $\lambda$ -a.e. in  $\Omega$ . Hence, both functionals denoted by  $G$  coincide on  $L^q(\Omega, d\lambda)$ , which justifies this notation.*

The following auxiliary results are generalizations of the corresponding results in [\[20\]](#). The first lemma covers the coercivity of  $G$  in  $L^1(\Omega, d\lambda)$ . To keep notation simple, from now on we will abbreviate  $\|\cdot\|_{L^p(\Omega, d\lambda)}$  by  $\|\cdot\|_p$  and similarly for  $\|\cdot\|_{L^p(\Omega_i, d\lambda_i)}$ ,  $i = 1, 2$ , where the underlying space will be clear from the context.

**Lemma 3.10.** *Let [Assumption 3.6](#) hold and suppose that a sequence  $(w_n) \subset L^q(\Omega, d\lambda)$  fulfills*

$$G(w_n) \leq C < \infty \quad \forall n \in \mathbb{N}$$

*for some  $C > 0$ . Then, the sequences  $(w_n)_+$  and  $(w_n)_-$  are bounded in  $L^q(\Omega, d\lambda)$  and  $L^1(\Omega, d\lambda)$ , respectively.*

*Proof.* This proof follows the outline of the proof in [\[20, Lemma 2.5\]](#).

We rewrite  $G$  as  $G(w) = \int_{\Omega} \frac{1}{q} w_+^q - w_+\mu d\lambda + \int_{\Omega} w_-\mu d\lambda$ . The positivity of  $\mu$  implies

$$\frac{1}{q} \|(w_n)_+\|_q^q = G(w_n) + \int_{\Omega} (w_n)_+\mu d\lambda - \int_{\Omega} (w_n)_-\mu d\lambda \leq C + \|\mu\|_p \|(w_n)_+\|_q,$$

which gives the first assertion. The second one can be seen by making use of  $\mu \geq \gamma^{q-1} \delta^2$  with  $\delta$  from [Assumption 3.6](#), which yields the estimate

$$\begin{aligned} C \geq G(w_n) &= \frac{1}{q} \int_{\Omega} (w_n)_+^q d\lambda - \int_{\Omega} (w_n)_+\mu d\lambda + \int_{\Omega} (w_n)_-\mu d\lambda \\ &\geq \frac{1}{q} \|(w_n)_+\|_q^q - \|\mu\|_p \|(w_n)_+\|_q + \gamma^{q-1} \delta^2 \|(w_n)_-\|_1 \\ &\geq -\|\mu\|_p \|(w_n)_+\|_q + \gamma^{q-1} \delta^2 \|(w_n)_-\|_1. \end{aligned}$$

Since  $\|(w_n)_+\|_q$  is already known to be bounded, the second assertion holds. □

The next lemma provides a lower semi-continuity result for  $G$  w.r.t. weak-\* convergence in  $\mathfrak{M}(\Omega)$ . Note that the extension of  $G$  as introduced above is needed, here.

**Lemma 3.11.** *Let Assumption 3.6 hold and a sequence  $(w_n) \subset L^q(\Omega, d\lambda)$  be given such that  $w_n \xrightarrow{*} \bar{w}$  in  $\mathfrak{M}(\Omega)$  and  $G(w_n) \leq C < \infty$  for all  $n \in \mathbb{N}$ . Then it holds that  $\frac{d\bar{w}_+}{d\lambda} \in L^q(\Omega, d\lambda)$  and*

$$G(\bar{w}) \leq \liminf_{n \rightarrow \infty} G(w_n).$$

*Proof.* The proof given in [20, Lemma 2.6] only uses Lemma 3.10 and fundamental properties of  $L^2(\Omega, d\mathcal{L})$ , which also hold for  $L^q(\Omega, d\lambda)$ ,  $q > 1$ , and can thus be readily extended.  $\square$

Now all prerequisites for proving the existence result for problem  $(P^\dagger)$  are gathered.

**Proposition 3.12.** *Let Assumption 3.6 hold. Then, problem  $(P^\dagger)$  admits a solution  $(\bar{\alpha}_1, \bar{\alpha}_2) \in L^1(\Omega_1, d\lambda_1) \times L^1(\Omega_2, d\lambda_2)$ .*

*Proof.* In [20, Proposition 2.9] the statement is proven for  $p = 2$  via the classical direct method of the calculus of variations using only [20, Lemmas 2.7 & 2.8] and Lemmas 3.10 and 3.11, where [20, Lemmas 2.7 & 2.8] are rather technical results holding independently of the choice of  $p > 1$  and  $\lambda_1, \lambda_2$ . Hence, the proof also holds for  $p > 1$ .  $\square$

Next, it is shown that  $\alpha_i, i = 1, 2$  are indeed functions in  $L^q(\Omega_i, d\lambda_i)$ .

**Theorem 3.13.** *Let Assumption 3.6 hold and let  $p \geq 2$ . Then every optimal solution  $(\bar{\alpha}_1, \bar{\alpha}_2)$  from Proposition 3.12 satisfies  $\bar{\alpha}_i \in L^q(\Omega_i, d\lambda_i)$ ,  $i = 1, 2$ . Moreover, the negative parts of  $\bar{\alpha}_i$  are bounded and the function  $\frac{1}{\gamma}(\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+$  has the marginals  $\frac{d\mu_1}{d\lambda_1}$  and  $\frac{d\mu_2}{d\lambda_2}$ .*

*Proof.* We only consider the the negative parts, as for the positive parts we already have  $(\alpha_i)_+ \in L^q(\Omega_i, d\lambda_i)$  by Lemma 3.8.

First note, that for functions  $f_i : \Omega_i \rightarrow \mathbb{R}$ ,  $i = 1, 2$  and  $g : \Omega_1 \rightarrow \mathbb{R}$  it holds

$$((f_1 + g) \oplus f_2)_+ = \max\left(0, (f_1 \oplus f_2) + g \oplus 0\right) \leq (f_1 \oplus f_2)_+ + g_+ \oplus 0,$$

where 0 is to be understood as the constant mapping  $x_2 \mapsto 0$ . Let now  $\varphi \in C_c^\infty(\Omega_1)$  and fix some  $0 < t \leq 1$ . Then, thanks to

$$0 \leq ((\bar{\alpha}_1 + t\varphi) \oplus \bar{\alpha}_2 - c)_+ \leq (\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+ + t\varphi_+ \oplus 0, \quad (3.2)$$

Proposition 3.12 implies that  $((\bar{\alpha}_1 + t\varphi) \oplus \bar{\alpha}_2 - c)_+ \in L^q(\Omega, d\lambda)$ , so that  $(\bar{\alpha}_1 + t\varphi, \bar{\alpha}_2)$  is feasible for problem  $(P^\dagger)$ . Therefore, the optimality of  $(\bar{\alpha}_1, \bar{\alpha}_2)$  for problem  $(P^\dagger)$  yields

$$\frac{1}{q} \int_{\Omega} \frac{1}{t} \left( ((\bar{\alpha}_1 + t\varphi) \oplus \bar{\alpha}_2 - c)_+^q - (\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+^q \right) d\lambda - \gamma^{q-1} \int_{\Omega_1} \frac{d\mu_1}{d\lambda_1} \varphi d\lambda_1 \geq 0$$

for all  $0 < t \leq 1$ . Owing to the continuous differentiability of  $\mathbb{R} \ni r \mapsto r_+^q \in \mathbb{R}$ , the first integrand converges to  $q(\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+^{q-1} \varphi$   $\lambda$ -a.e. in  $\Omega$ .

Moreover, for  $x \geq 0$ , the mapping  $[0, 1] \ni t \mapsto (x + t\varphi_+)^q \in \mathbb{R}$  is Lipschitz continuous with Lipschitz constant  $q\varphi_+ (x + \varphi_+)^{q-1}$ . Together with (3.2), this gives

$$\begin{aligned} \frac{1}{t} \left( ((\bar{\alpha}_1 + t\varphi) \oplus \bar{\alpha}_2 - c)_+^q - (\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+^q \right) &\leq \frac{1}{t} \left( ((\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+ + t\varphi_+ \oplus 0)^q - (\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+^q \right) \\ &\leq q\varphi_+ ((\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+ + \varphi_+ \oplus 0)^{q-1} \\ &\leq q\varphi_+ (2 \max \{ (\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+, \varphi_+ \oplus 0 \})^{q-1} \\ &\leq q\varphi_+ 2^{q-1} \left( (\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+^{q-1} + (\varphi_+ \oplus 0)^{q-1} \right), \end{aligned} \quad (3.3)$$

since  $(x + y)^r \leq (2 \max\{x, y\})^r$  for all  $r > 0$ ,  $x, y \geq 0$ . As  $g \in L^q(\Omega, d\lambda)$  with  $\lambda(\Omega) < \infty$  implies  $\int_{\Omega} g^{q-1} d\lambda < \infty$  for all  $q > 1$  (see e.g. [15, Proposition 6.12]), the right-hand side is integrable.

Hence, due to Lebesgue's dominated convergence theorem, passing to the limit  $t \searrow 0$  is allowed and yields

$$\int_{\Omega_1} \left( \int_{\Omega_2} (\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+^{q-1} d\lambda_2 - \gamma^{q-1} \frac{d\mu_1}{d\lambda_1} \right) \varphi d\lambda_1 \geq 0.$$

Since  $\varphi \in C_c^\infty(\Omega_1)$  was arbitrary, the fundamental lemma of the calculus of variations gives

$$\int_{\Omega_2} (\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c)_+^{q-1} d\lambda_2 = \gamma^{q-1} \frac{d\mu_1}{d\lambda_1} \quad (3.4)$$

$\lambda_1$ -a.e. in  $\Omega_1$ . Next, define a sequence of functions  $(f_n)$  by

$$f_n(x_2) := (-n + \bar{\alpha}_2(x_2) - c^\dagger)_+ \quad \forall n \in \mathbb{N},$$

where  $c^\dagger$  is the lower bound for  $c$  from Assumption 3.6. It holds that  $f_n \in L^{q-1}(\Omega_2, d\lambda_2)$ , which can be seen as follows: Since  $\bar{\alpha}_2 \in L^1(\Omega_2, d\lambda_2)$ , clearly  $(\bar{\alpha}_2)^{q-1} \in L^{1/q-1}(\Omega_2, d\lambda_2)$ . Furthermore, since  $\Omega_2$  is compact and  $q \leq 2$ , it holds  $L^{1/q-1}(\Omega_2, d\lambda_2) \hookrightarrow L^1(\Omega_2, d\lambda_2)$ . Consequently,  $(\bar{\alpha}_2)^{q-1} \in L^1(\Omega_2, d\lambda_2)$  and  $f_n^{q-1}$  is also integrable for every  $n$ .

The functions  $f_n$  satisfy  $f_n \geq 0$  and  $f_n \searrow 0$   $\lambda_2$ -a.e. in  $\Omega_2$  for  $n \rightarrow \infty$ , so that the monotone convergence theorem gives

$$\int_{\Omega_2} f_n^{q-1} d\lambda_2 \xrightarrow{n \rightarrow \infty} 0.$$

Thus, there exists  $N \in \mathbb{N}$  such that

$$\int_{\Omega_2} (-N + \bar{\alpha}_2 - c^\dagger)_+^{q-1} d\lambda_2 < \gamma^{q-1} \delta,$$

with the threshold  $\delta > 0$  from Assumption 3.6. Now assume that  $\bar{\alpha}_1 \leq -N$   $\lambda_1$ -a.e. on a set  $E \subset \Omega_1$  with  $\lambda_1(E) > 0$ . Then

$$\int_{\Omega_2} (\bar{\alpha}_1 \oplus \bar{\alpha}_2 - c^\dagger)_+^{q-1} d\lambda_2 \leq \int_{\Omega_2} (-N + \bar{\alpha}_2 - c^\dagger)_+^{q-1} d\lambda_2 < \gamma^{q-1} \delta \leq \gamma^{q-1} \frac{d\mu_1}{d\lambda_1}$$

$\lambda_1$ -a.e. in  $E$ , which is a contradiction to (3.4). Therefore  $\bar{\alpha}_1 > -N$   $\lambda_1$ -a.e. in  $\Omega_1$ , which even implies that  $(\bar{\alpha}_1)_- \in L^\infty(\Omega_1, d\lambda_1)$ . Concerning  $(\bar{\alpha}_2)_-$ , one may argue exactly the same way to conclude that  $(\bar{\alpha}_2)_- \in L^\infty(\Omega_1, d\lambda_1)$ , too.  $\square$



**Remark 3.14.** While it seems clear that proving a generalization of [Theorem 3.13](#) to general Young's functions or even quasi-Young's functions  $\Phi$  is likely to be complicated or even impossible without making strict assumptions on  $\Phi$ , not even the existence result for optimizers in  $L^1$  can be generalized directly. The problem occurs in [Lemma 3.10](#), which could not be extended to the case of Young's functions or quasi-Young's functions  $\Phi$  in this work. That additional assumptions on  $\Phi$  might be necessary for [Lemma 3.10](#) to hold can be seen as follows.

In the general case, the function  $G$  would be defined as

$$G : L^q(\Omega, d\lambda) \ni w \mapsto \int_{\Omega} \left( \gamma \tilde{\Phi}^* \left( \frac{w}{\gamma} \right) - w\mu \right) d\lambda \in \mathbb{R},$$

where  $\mu := \frac{d(\mu_1 \otimes \mu_2)}{d\lambda}$  and in the proof of [Lemma 3.10](#), an inequality of the form

$$C \|w_n\|_{L^{\Phi^*}(\Omega, d\lambda)} \leq \gamma \int_{\Omega} \tilde{\Phi}^* \left( \frac{w_n}{\gamma} \right) d\lambda \quad \forall n \in \mathbb{N}$$

would be necessary. For this to hold, it would suffice to know

$$\Phi(\|f\|_{L^{\Phi}(\Omega, d\lambda)}) \leq C \int_{\Omega} \Phi(|f|) d\lambda$$

for (quasi-)Young's functions  $\Phi$ , but this is not true in general as the Young's function  $\Phi(t) = \max(t^2, t^3)$  and  $\Omega = (0, 1)$  shows. Indeed for  $f = a\mathbb{1}_{(0,b)}$  for some  $a, b \in (0, 1)$  one readily computes that  $\|f\|_{L^{\Phi}(\Omega, d\mathcal{L})} = ab^{1/3}$  and the above mentioned inequality would be

$$a^2 b^{2/3} \leq C a^2 b,$$

which is not possible for any constant  $C$  independent of  $b$ .<sup>1</sup> This counterexample indicates that both the growth of  $\Phi$  at infinity and at zero are important properties for this problem.

## 4. $\Gamma$ -Convergence

We return to the general setting, i.e. we only assume that [Assumption 1.1](#) holds, and consider results on  $\Gamma$ -convergence of problems related to problem (P). Recall from, e.g., [7], that a sequence  $(F_n)$  of functionals  $F_n : X \rightarrow \mathbb{R} \cup \{\infty\}$  on a metric space  $X$  is said to  $\Gamma$ -converge to a functional  $F : X \rightarrow \mathbb{R} \cup \{\infty\}$ , 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.

<sup>1</sup>We thank the user harfe from mathoverflow who provided this counterexample to our question <https://mathoverflow.net/q/333925>.

#### 4.1. Continuous case

When considering arbitrary measures as marginals, their densities w.r.t.  $\lambda_i$  may not be in  $L^\Phi(\Omega, d\lambda_i)$  and by [Theorem 3.4](#), problem (P) will not admit a solution in that case. One may therefore consider smoothed marginals  $\mu_i^\delta$  with  $\frac{d\mu_i^\delta}{d\lambda_i} \in L^\Phi(\Omega_i, d\lambda_i)$ ,  $i = 1, 2$ , converging to  $\mu_1$  and  $\mu_2$ , respectively and show that the regularized problem with these marginals  $\Gamma$ -converges to the unregularized problem with the original marginals. Even in the case where  $\frac{d\mu_i}{d\lambda_i} \in L^\Phi(\Omega_i, d\lambda_i)$ , smoothing of the marginals allows for a fairly short proof of  $\Gamma$ -convergence.

Let  $\varphi$  be a smooth, compactly supported, non-negative kernel on  $\mathbb{R}^n$  with unit integral (w.r.t. the Lebesgue measure) and set

$$\varphi_r(x) = \frac{1}{r^n} \varphi\left(\frac{x}{r}\right), \quad G_r := \varphi_r \otimes \varphi_r.$$

Since the marginals and transport plans will be smoothed by convolutions, the domains  $\Omega_1$  and  $\Omega_2$  will be extended slightly to take care of boundary effects. Hence, let  $\tilde{\Omega}_1, \tilde{\Omega}_2$  be compact supersets of  $\Omega_1, \Omega_2$ , respectively, such that

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

which is large enough to contain the supports of the smoothed marginals  $\mu_i^r$ ,  $i = 1, 2$  for  $r \leq 1$  (and the width of the convolution kernels will be assumed to be small enough for this in the following). For a function (or measure)  $f$  on  $\Omega_1$  denote by  $\tilde{f}$  the extension of  $f$  onto  $\tilde{\Omega}_1$  by zero (and analogously for functions and measures on  $\Omega_2$  and  $\Omega_1 \times \Omega_2$ ). Let  $\hat{\lambda}_i$  be the extension of  $\lambda_i$  onto  $\tilde{\Omega}_i$  by the Lebesgue measure and  $\hat{\lambda} = \hat{\lambda}_1 \otimes \hat{\lambda}_2$ . Let  $\hat{c}$  be a continuous extension of  $c$  onto  $\tilde{\Omega}_1 \times \tilde{\Omega}_2$  and let

$$F_Y[\pi] = \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \hat{c} d\pi + \gamma \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \Phi_\infty\left(\frac{d\pi}{d\hat{\lambda}}\right) d\hat{\lambda},$$

where we set the second integral to  $+\infty$ , if  $\pi \not\ll \hat{\lambda}$ , and

$$E_Y^{v_1, v_2}[\pi] = \begin{cases} F_Y[\pi] & \text{if } 0 \leq \pi \in \mathcal{P}(\tilde{\Omega}_1 \times \tilde{\Omega}_2), (P_i)_\# \pi = v_i, i = 1, 2, \\ \infty & \text{else,} \end{cases}$$

for  $v_i \in \mathfrak{M}_+(\Omega)$ ,  $i = 1, 2$ .

First, we state an auxiliary result ensuring that the marginal constraints are preserved by convolution. For simplicity, we state it for measures on  $\mathbb{R}^n$  (but we could restrict everything to the respective domains  $\Omega_i, \tilde{\Omega}_i$ ). A proof is given in [Appendix B](#).

**Lemma 4.1.** *Let  $v_1, v_2 \in \mathfrak{M}(\mathbb{R}^n)$ , let  $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^n)$ , with  $\mu_i \ll v_i$  and let  $v := v_1 \otimes v_2$ . Let  $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$  with  $(P_i)_\# \pi = \mu_i$ ,  $i = 1, 2$  such that  $\pi \ll v$ . Let*

$$\mu_i^\delta := \varphi_\delta * \mu_i, \quad i = 1, 2 \quad \text{and} \quad \pi_\delta := G_\delta * \pi.$$

*Then  $(P_i)_\# \pi_\delta = \mu_i^\delta$  for  $i = 1, 2$ .*

**Theorem 4.2** ( $\Gamma$ -convergence for smoothed marginals). *Let [Assumption 1.1](#) hold and let  $(\gamma, \delta) \xrightarrow{\Phi} 0$  denote*

$$\gamma \rightarrow 0, \quad \delta \rightarrow 0, \quad \gamma \Phi_+ \left( \frac{1}{\delta^{2n}} \right) \rightarrow 0.$$

Define the smoothed marginals as  $\mu_i^\delta = \varphi_\delta * \tilde{\mu}_i$  for  $i = 1, 2$ . Then it holds that

$$\Gamma\text{-}\lim_{(\gamma, \delta) \xrightarrow{\Phi} 0} E_Y^{\mu_1^\delta, \mu_2^\delta} = E_0^{\mu_1, \mu_2}$$

with respect to weak-\* convergence in  $\mathfrak{M}(\Omega_1 \times \Omega_2)$ . Moreover, if  $\gamma, \delta \rightarrow 0$  are chosen such that

$$\gamma \left\| \frac{d\mu_1^\delta}{d\lambda_1} \right\|_{L^\Phi(\Omega_1, d\lambda_1)} \rightarrow \infty \quad \text{or} \quad \gamma \left\| \frac{d\mu_2^\delta}{d\lambda_2} \right\|_{L^\Phi(\Omega_2, d\lambda_2)} \rightarrow \infty,$$

then  $E_Y^{\mu_1^\delta, \mu_2^\delta}$  does not have a finite  $\Gamma$ -limit. More precisely, even for feasible  $\pi_\delta$  (i.e. with marginals  $\mu_i^\delta$ ) it holds that

$$\lim_{(\gamma, \delta) \xrightarrow{\Phi} 0} E_Y^{\mu_1^\delta, \mu_2^\delta}[\pi_\delta] = \infty.$$

*Proof.* This proof follows the outline of the proof in [10, Theorem 5.1].

- i) *lim inf-condition:* Let  $\pi_\delta \xrightarrow{*} \pi$ , then  $\lim_{\delta \rightarrow 0} F_0[\pi_\delta] = F_0[\tilde{\pi}]$  due to  $\hat{c}$  being continuous and bounded. Furthermore,

$$\int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \Phi_\infty\left(\frac{d\tilde{\pi}}{d\hat{\lambda}}\right) d\hat{\lambda} \geq -C \cdot \hat{\lambda}(\tilde{\Omega}_1 \times \tilde{\Omega}_2)$$

for some  $C > 0$  only dependent on  $\Phi$  for any  $\tilde{\pi} \geq 0$ . Thus,

$$F_0[\tilde{\pi}] = \lim_{(\gamma, \delta) \xrightarrow{\Phi} 0} F_0[\pi_\delta] - \gamma C \cdot \hat{\lambda}(\tilde{\Omega}_1 \times \tilde{\Omega}_2) \leq \liminf_{(\gamma, \delta) \xrightarrow{\Phi} 0} F_\gamma[\pi_\delta].$$

Finally, the pushforward operator is weak-\* continuous, which implies that the marginal constraints are preserved under weak-\* convergence of  $\pi_\delta$ ,  $\mu_1^\delta$ , and  $\mu_2^\delta$  (note that  $\mu_i^\delta \xrightarrow{*} \mu_i$ ).

- ii) *lim sup-condition:* It suffices to consider a recovery sequence for  $\pi \in \mathcal{P}(\Omega_1 \times \Omega_2)$ , because for  $\pi \in \mathcal{P}(\tilde{\Omega}_1 \times \tilde{\Omega}_2) \setminus \mathcal{P}(\Omega_1 \times \Omega_2)$  the marginal conditions for  $\mu_1$  and  $\mu_2$  can never be satisfied.

If  $E_0^{\tilde{\mu}_1, \tilde{\mu}_2}[\tilde{\pi}] = \infty$ , then  $E_Y^{\tilde{\mu}_1, \tilde{\mu}_2}[\tilde{\pi}] = \infty$  for every  $\gamma$  and the lim sup condition holds trivially. Let therefore  $E_0^{\tilde{\mu}_1, \tilde{\mu}_2}[\tilde{\pi}]$  be finite. First, note that  $\pi$  thus satisfies the marginal constraints. Considering two Borel sets  $A_i \subset \Omega_i$ ,  $i = 1, 2$  such that  $(\mu_1 \otimes \mu_2)(A_1 \times A_2) = 0$ , it is straightforward to see, that also  $\pi(A_1 \times A_2) = 0$ . Hence,  $\pi \ll \mu_1 \otimes \mu_2$  and by the assumption  $\mu_i \ll \lambda_i$ ,  $i = 1, 2$ , we also have  $\pi \ll \lambda$ . Set now  $\pi_\delta = G_\delta * \tilde{\pi}$ . Then  $0 \leq \pi_\delta$  and  $\pi_\delta \xrightarrow{*} \tilde{\pi}$  as well as  $(P_i)_\# \pi_\delta = \mu_i^\delta$  for  $i = 1, 2$ , by Lemma 4.1. Finally, by Young's convolution inequality,

$$\frac{d\pi_\delta}{d\hat{\lambda}} \leq \|G_\delta\|_{L^\infty(\Omega, d\hat{\lambda})} \left\| \frac{d\tilde{\pi}}{d\hat{\lambda}} \right\|_{L^1(\Omega, d\hat{\lambda})} \leq \frac{C}{\delta^{2n}}$$

for some constant  $C > 0$ . Thus,

$$\begin{aligned} \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \Phi_\infty\left(\frac{d\pi_\delta}{d\hat{\lambda}}\right) d\hat{\lambda} &= \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \Phi\left(\frac{d\pi_\delta}{d\hat{\lambda}}\right) d\hat{\lambda} \leq \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \Phi_+\left(\frac{d\pi_\delta}{d\hat{\lambda}}\right) d\hat{\lambda} \\ &\leq \hat{\lambda}(\tilde{\Omega}_1 \times \tilde{\Omega}_2) \gamma \Phi_+\left(\frac{C}{\delta^{2n}}\right), \end{aligned} \tag{4.1}$$

and the right hand side vanishes for  $(\gamma, \delta) \xrightarrow{\Phi} 0$  by the assumption on the (coupled) convergences of  $\gamma$  and  $\delta$ . Therefore,

$$E_0^{\tilde{\mu}_1, \tilde{\mu}_2}[\tilde{\pi}] = \lim_{(\gamma, \delta) \xrightarrow{\Phi} 0} \left( F_0[\pi_\delta] + \hat{\lambda} \left( \tilde{\Omega}_1 \times \tilde{\Omega}_2 \right) \gamma \Phi_+ \left( \frac{C}{\delta^{2n}} \right) \right) \geq \lim_{(\gamma, \delta) \xrightarrow{\Phi} 0} F_\gamma[\pi_\delta].$$

iii) For the second statement, recall from [Lemma 2.9](#) that

$$\begin{aligned} \gamma \left\| \frac{d\mu_1^\delta}{d\hat{\lambda}_1} \right\|_{L^\Phi(\Omega_1, d\hat{\lambda}_1)} &\leq \gamma \max \left( 1, \hat{\lambda}_2(\Omega_1) \right) \left\| \frac{d\pi_\delta}{d\hat{\lambda}} \right\|_{L^\Phi(\Omega, d\hat{\lambda})}, \\ \gamma \left\| \frac{d\mu_2^\delta}{d\hat{\lambda}_2} \right\|_{L^\Phi(\Omega_2, d\hat{\lambda}_2)} &\leq \gamma \max \left( 1, \hat{\lambda}_1(\Omega_2) \right) \left\| \frac{d\pi_\delta}{d\hat{\lambda}} \right\|_{L^\Phi(\Omega, d\hat{\lambda})}. \end{aligned}$$

By [Lemma 2.11](#), this immediately yields  $F_\gamma[\pi_\delta] \rightarrow \infty$ , and the assertion follows.  $\square$

The assumption  $\gamma \Phi_+ \left( \frac{1}{\delta^{2n}} \right) \rightarrow 0$  is much stronger than necessary for some Young's functions. For example consider  $\Phi(t) = t^p/p$  with  $p > 1$  or  $\Phi(t) = t \log t$ . In those cases the condition

$$\gamma \delta^{2n} \Phi_+ \left( \frac{1}{\delta^{2n}} \right) \rightarrow 0$$

suffices, as the following result states. For  $\Phi(t) = t \log t$  this gives exactly the result in [[10](#), Theorem 5.1].

**Corollary 4.3.** *Let  $\Phi$  be a quasi-Young's function such that  $t^{-1}\Phi(t)$  is monotone. Then it suffices to assume*

$$\gamma \delta^{2n} \Phi_+ \left( \frac{1}{\delta^{2n}} \right) \rightarrow 0$$

in [Theorem 4.2](#).

*Proof.* An alternate estimate to [\(4.1\)](#) can be given. Using the monotonicity of  $t^{-1}\Phi(t)$  we obtain

$$\begin{aligned} \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \Phi \left( \frac{d\pi_\delta}{d\hat{\lambda}} \right) d\hat{\lambda} &\leq \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \Phi_+ \left( \frac{d\pi_\delta}{d\hat{\lambda}} \right) d\hat{\lambda} \\ &= \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \frac{d\pi_\delta}{d\hat{\lambda}} \left( \frac{d\pi_\delta}{d\hat{\lambda}} \right)^{-1} \Phi_+ \left( \frac{d\pi_\delta}{d\hat{\lambda}} \right) d\hat{\lambda} \\ &\leq \frac{\delta^{2n}}{C} \Phi_+ \left( \frac{C}{\delta^{2n}} \right) \int_{\tilde{\Omega}_1 \times \tilde{\Omega}_2} \frac{d\pi_\delta}{d\hat{\lambda}} d\hat{\lambda} \\ &= \frac{\delta^{2n}}{C} \Phi_+ \left( \frac{C}{\delta^{2n}} \right) \end{aligned}$$

where again the right hand side vanishes for  $(\gamma, \delta) \xrightarrow{\Phi} 0$ . The assertion follows as in [Theorem 4.2](#).  $\square$

## 4.2. Discretized problems

Here we describe a discretization of problem (P) and show two approximation results:

- i)  $\Gamma$ -convergence of the discretized and regularized problem towards the continuous, *regularized* problem (P)
- ii)  $\Gamma$ -convergence of the discretized and regularized problem towards the continuous, *unregularized* problem (OT)

We recall the problem data: Marginals  $\mu_i \in \mathcal{P}(\Omega_i)$  and finite measures  $\lambda_i \in \mathfrak{M}_+(\Omega_i)$  with  $\mu_i \ll \lambda_i$  for  $i = 1, 2$  and a continuous and positive cost function  $c$  on  $\Omega$ . We have  $\lambda = \lambda_1 \otimes \lambda_2$  and aim to discretize the problem

$$\min_{\pi \in \mathcal{P}(\Omega)} \int_{\Omega} c \frac{d\pi}{d\lambda} + \gamma \Phi_{\infty} \left( \frac{d\pi}{d\lambda} \right) d\lambda, \quad \text{s.t.} \quad \int_{\Omega_2} \frac{d\pi}{d\lambda} d\lambda_2 = \frac{d\mu_1}{d\lambda_1} \quad \lambda_1\text{-a.e.} \\ \int_{\Omega_1} \frac{d\pi}{d\lambda} d\lambda_1 = \frac{d\mu_2}{d\lambda_2} \quad \lambda_2\text{-a.e.} \quad (4.2)$$

We do a Galerkin discretization with piecewise constant functions. For  $k \in \mathbb{N}$  let  $(Q_j^{i,k})$  be a sequence of finite partitions of  $\Omega_i$  such that for every  $j$  there is an  $l$  with  $Q_j^{i,k+1} \subset Q_l^{i,k}$  and such that  $(Q_i^{1,k})$ ,  $(Q_j^{2,k})$ , and  $I_{ij}^k := Q_i^{1,k} \times Q_j^{2,k}$  satisfy the following assumption:

**Assumption 4.4.** *Let  $A \subset \Omega$  be a Borel set and  $\varepsilon > 0$ . Then there exists some  $K \in \mathbb{N}$  such that for all  $K \leq k \in \mathbb{N}$  the sets  $A_+^k, A_-^k$  defined by*

$$A_-^k := \bigcup_{\{(i,j) \mid I_{i,j}^k \subseteq A\}} I_{i,j}^k, \quad A_+^k := \bigcup_{\{(i,j) \mid I_{i,j}^k \cap A \neq \emptyset\}} I_{i,j}^k$$

satisfy

$$v(A_+^k) - v(A_-^k) < \varepsilon, \quad (4.3)$$

for all  $v \in \mathfrak{M}_+(\Omega)$ .

**Remark 4.5.** *If Assumption 4.4 is fulfilled for  $\lambda = \lambda_1 \otimes \lambda_2$ , condition (4.3) holds analogously for  $\lambda_1$  and  $\lambda_2$ , which can be seen as follows: For  $A^i \subset \Omega_i$ ,  $i = 1, 2$  let  $A_+^{i,k}$  and  $A_-^{i,k}$  be defined analogously based on  $Q_j^{i,k}$ . It holds*

$$A_-^{1,k} \times \Omega_2 = (A^1 \times \Omega_2)_-^k \subset \Omega \\ A_+^{1,k} \times \Omega_2 = (A^1 \times \Omega_2)_+^k \subset \Omega.$$

Thus,

$$\lambda_1(A_+^{1,k}) - \lambda_1(A_-^{1,k}) = (P_1)_{\#} \lambda(A_+^{1,k}) - (P_1)_{\#} \lambda(A_-^{1,k}) = \lambda(A_+^{1,k} \times \Omega_2) - \lambda(A_-^{1,k} \times \Omega_2) \\ = \lambda((A^1 \times \Omega_2)_+^k) - \lambda((A^1 \times \Omega_2)_-^k) < \varepsilon$$

by (4.3) and the argument holds analogously for  $\lambda_2$ .

[Assumption 4.4](#) yields the following auxiliary result about piecewise constant approximation of measures.

**Lemma 4.6.** *Let [Assumption 4.4](#) hold. Let  $v \in \mathfrak{M}_+(\Omega)$ ,  $v^i \in \mathfrak{M}_+(\Omega_i)$ ,  $i = 1, 2$  and define*

$$v_k := \sum_{i,j} \frac{v(I_{i,j}^k)}{\lambda(I_{i,j}^k)} \mathbb{1}_{I_{i,j}^k}, \quad v_k^i := \sum_j \frac{v^i(Q_j^{i,k})}{\lambda_i(Q_j^{i,k})} \mathbb{1}_{Q_j^{i,k}}, \quad i = 1, 2.$$

Then  $v_k \lambda \xrightarrow{*} v$  and  $v_k^i \lambda_i \xrightarrow{*} v^i$ .

A proof of [Lemma 4.6](#) can be found in [Appendix C](#). We now set

$$\begin{aligned} \mu_{1,i}^k &:= \mu_1(Q_i^{1,k}), & \mu_{2,j}^k &:= \mu_2(Q_j^{2,k}) \\ \lambda_{1,i}^k &:= \lambda_1(Q_i^{1,k}), & \lambda_{2,j}^k &:= \lambda_2(Q_j^{2,k}). \end{aligned}$$

Then by [Lemma 4.6](#)

$$\sum_i \frac{\mu_{1,i}^k}{\lambda_{1,i}^k} \mathbb{1}_{Q_i^{1,k}} \lambda_1 \xrightarrow{*} \mu_1 \qquad \sum_j \frac{\mu_{2,j}^k}{\lambda_{2,j}^k} \mathbb{1}_{Q_j^{2,k}} \lambda_2 \xrightarrow{*} \mu_2.$$

Note that division by zero is not a problem here, since  $\lambda_1$  and  $\lambda_2$  were assumed to have full support and hence  $\lambda_{1,i}^k, \lambda_{2,j}^k \neq 0$  for all  $i, j, k$ . We define the finite-dimensional spaces

$$\mathcal{V}^{1,k} := \langle \{\mathbb{1}_{Q_i^{1,k}} \mid i\} \rangle, \quad \mathcal{V}^{2,k} := \langle \{\mathbb{1}_{Q_j^{2,k}} \mid j\} \rangle, \quad \mathcal{V}^k = \mathcal{V}^{1,k} \otimes \mathcal{V}^{2,k} = \langle \{\mathbb{1}_{I_{ij}^k} \mid i, j\} \rangle.$$

The discrete problem is then one in  $\mathcal{V}^k$ , namely

$$\inf_{\pi \in \mathcal{V}^k} \int_{\Omega} c \pi \, d\lambda + \gamma \int_{\Omega} \Phi_{\infty}(\pi) \, d\lambda, \quad \text{s.t.} \quad \begin{aligned} \int_{\Omega_2} \pi \, d\lambda_2 &= \sum_i \frac{\mu_{1,i}^k}{\lambda_{1,i}^k} \mathbb{1}_{Q_i^{1,k}}, \\ \int_{\Omega_1} \pi \, d\lambda_1 &= \sum_j \frac{\mu_{2,j}^k}{\lambda_{2,j}^k} \mathbb{1}_{Q_j^{2,k}} \end{aligned} \tag{PD}$$

If we discretize the sought-after density  $\frac{d\pi}{d\lambda}$  by

$$\frac{d\pi}{d\lambda} = \sum_{ij} p_{ij} \mathbb{1}_{I_{ij}^k} \in \mathcal{V}^k,$$

we can derive the optimization problem for the unknown coefficients  $p_{ij}$  as follows: The objective function is

$$\int_{\Omega} c \frac{d\pi}{d\lambda} + \gamma \Phi_{\infty}\left(\frac{d\pi}{d\lambda}\right) \, d\lambda = \sum_{ij} \int_{I_{ij}^k} c \, d\lambda \, p_{ij} + \gamma \Phi_{\infty}(p_{ij}) \lambda_{1,i}^k \lambda_{2,j}^k.$$

The first marginal constraint is given by

$$\int_{\Omega_2} \sum_{ij} p_{ij} \mathbb{1}_{Q_i^{1,k}}(x_1) \mathbb{1}_{Q_j^{2,k}}(x_2) \, d\lambda_2(x_2) = \sum_i \frac{\mu_{1,i}^k}{\lambda_{1,i}^k} \mathbb{1}_{Q_i^{1,k}}(x_1)$$

and this leads to the equation

$$\sum_j p_{ij} \lambda_{2,j}^k = \frac{\mu_{1,i}^k}{\lambda_{1,i}^k}.$$

A similar equation can be derived for the other marginal constraint. With

$$c_{ij}^k := \frac{1}{\lambda_{1,i}^k \lambda_{2,j}^k} \int_{I_{ij}^k} c \, d\lambda,$$

we arrive at the fully discretized problem

$$\begin{aligned} \min_p \sum_{ij} \left( c_{ij}^k p_{ij} + \gamma \Phi_\infty(p_{ij}) \right) \lambda_{1,i}^k \lambda_{2,j}^k, \quad & \sum_j p_{ij} \lambda_{2,j}^k = \frac{\mu_{1,i}^k}{\lambda_{1,i}^k} \\ & \sum_i p_{ij} \lambda_{1,i}^k = \frac{\mu_{2,j}^k}{\lambda_{2,j}^k}. \end{aligned}$$

This is a finite-dimensional convex minimization problem with linear constraints. Several general methods could be used to solve this problem numerically, see, e.g. [4, 27, 20, 21]. The following theorem now guarantees that the discretized and regularized problem converges to the continuous regularized problem (P). The proof makes use of the so-called Jensen's inequality (see e.g. [5, Theorem 2.12.19]), which we recall for the convenience of the reader: Let  $\Omega \subset \mathbb{R}^d$ ,  $\nu \in \mathfrak{M}_+(\mathbb{R}^d)$  with  $\nu(\Omega) < \infty$ . Let  $\varphi : \mathbb{R}^d \supset K \rightarrow \mathbb{R}^d$  be convex and let  $f : \Omega \rightarrow M \subset \mathbb{R}^d$  be a  $\nu$ -integrable function with  $M \subset K$  such that  $\varphi(f)$  is  $\nu$ -integrable. Then,

$$\nu(\Omega) \cdot \varphi\left(\frac{1}{\nu(\Omega)} \int_\Omega f \, d\nu\right) \leq \int_\Omega \varphi(f) \, d\nu.$$

**Theorem 4.7.** *Let Assumptions 1.1 and 4.4 hold. Let  $c \in C(\Omega, [0, \infty))$ ,  $\Phi$  be a quasi-Young's function, and let  $\gamma > 0$ . Then the minimization problem (PD)  $\Gamma$ -converges to problem (4.2) as  $k \rightarrow \infty$  w.r.t. weak-\* convergence in  $\mathfrak{M}(\Omega)$ .*

*Proof.* We define,  $F, F_k : \mathfrak{M}(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}$  via

$$F_k(\pi) = \begin{cases} \int_\Omega c \, d\pi + \gamma \int_\Omega \Phi_\infty\left(\frac{d\pi}{d\lambda}\right) \, d\lambda, & \frac{d(P_1)_\# \pi}{d\lambda_1} = \sum_i \frac{\mu_{1,i}^k}{\lambda_{1,i}^k} \mathbb{1}_{Q_i^{1,k}}, \frac{d(P_2)_\# \pi}{d\lambda_2} = \sum_j \frac{\mu_{2,j}^k}{\lambda_{2,j}^k} \mathbb{1}_{Q_j^{2,k}}, \\ 0 \leq \pi \ll \lambda, \frac{d\pi}{d\lambda} \in \mathcal{V}_k & \\ \infty, & \text{else} \end{cases}$$

and

$$F(\pi) = \begin{cases} \int_\Omega c \, d\pi + \gamma \int_\Omega \Phi_\infty\left(\frac{d\pi}{d\lambda}\right) \, d\lambda, & 0 \leq \pi \ll \lambda, (P_i)_\# \pi = \mu_i, i = 1, 2 \\ \infty, & \text{else.} \end{cases}$$

Given an arbitrary  $\pi \in \mathfrak{M}(\Omega)$ , we now check the two conditions for  $\Gamma$ -convergence.

i) *lim inf-condition:* Let  $(\pi_k) \subset \mathfrak{M}(\Omega)$  such that  $\pi_k \xrightarrow{*} \pi$ .

If  $F(\pi) < \infty$ , pass to a subsequence (denoted by the same symbol) with finite values  $F_k(\pi_k)$ .

Because  $\pi_k \xrightarrow{*} \pi$ , we have

$$\int_\Omega c \, d\pi_k \rightarrow \int_\Omega c \, d\pi,$$

and moreover,  $\liminf_{k \rightarrow \infty} \int_{\Omega} \Phi_{\infty} \left( \frac{d\pi_k}{d\lambda} \right) d\lambda \geq \int_{\Omega} \Phi_{\infty} \left( \frac{d\pi}{d\lambda} \right) d\lambda$  by [Lemma 2.13](#).

If  $F(\pi) = \infty$ , assume for a contradiction that  $\liminf_{k \rightarrow \infty} F_k(\pi_k) < \infty$ . Pass to a subsequence (not renamed)  $(\pi_k)$  with  $\lim_{k \rightarrow \infty} F_k(\pi_k) = \liminf_{k \rightarrow \infty} F_k(\pi_k)$ .

Since,  $\pi_k \xrightarrow{*} \pi$  and, by [Lemma 4.6](#),  $\sum_i \frac{\mu_{1,i}^k}{\lambda_{1,i}^k} \mathbb{1}_{Q_i^{1,k}} \xrightarrow{*} \frac{d\mu_1}{d\lambda_1}$ , and  $\sum_j \frac{\mu_{2,j}^k}{\lambda_{2,j}^k} \mathbb{1}_{Q_j^{2,k}} \xrightarrow{*} \frac{d\mu_2}{d\lambda_2}$ , we see that  $\frac{d\pi_k}{d\lambda}$  always satisfies the marginal constraints and positivity. Hence,  $F(\pi) = \infty$  can not occur due to violation of these constraints.

So, if  $\pi$  satisfies the marginals but  $\pi \not\ll \lambda$  or  $\pi \ll \lambda$  with  $\int_{\Omega} \Phi_{\infty} \left( \frac{d\pi}{d\lambda} \right) d\lambda = \infty$ , then  $\liminf_{k \rightarrow \infty} F_k(\pi_k) = \infty$  by [Lemma 2.13](#), which again is a contradiction.

Thus,  $\liminf F_k(\pi_k) \geq F(\pi)$  in every case.

- ii) *lim sup-condition*: If  $F(\pi) = \infty$ , the condition  $\limsup_{k \rightarrow \infty} F_k(\pi_k)$  is trivially fulfilled for  $\pi_k := \pi$ . Hence, consider  $F(\pi) < \infty$  and define  $(\pi_k)$  by

$$\frac{d\pi_k}{d\lambda} := \sum_{i,j} \frac{\pi(I_{ij}^k)}{\lambda(I_{ij}^k)} \mathbb{1}_{I_{ij}^k}.$$

Note that by assumption  $\lambda(I_{ij}^k) > 0$  and therefore  $\pi_k \geq 0$  by definition. One easily sees that

$$\int_{\Omega_1} \frac{d\pi_k}{d\lambda} d\lambda_1 = \sum_j \frac{\mu_{2,j}^k}{\lambda_{2,j}^k} \mathbb{1}_{Q_j^{2,k}}, \quad \text{and} \quad \int_{\Omega_2} \frac{d\pi_k}{d\lambda} d\lambda_2 = \sum_i \frac{\mu_{1,i}^k}{\lambda_{1,i}^k} \mathbb{1}_{Q_i^{1,k}},$$

i.e.,  $F_k(\pi_k) < \infty$ . In particular,  $\pi_k \xrightarrow{*} \pi$  by [Lemma 4.6](#).

It remains to show  $\limsup_{k \rightarrow \infty} F_k(\pi_k) \leq F(\pi)$ . As before, we have

$$\int_{\Omega} c d\pi_k \rightarrow \int_{\Omega} c d\pi.$$

Moreover, we get from Jensen's inequality that

$$\Phi \left( \frac{\pi(I_{ij}^k)}{\lambda(I_{ij}^k)} \right) = \Phi \left( \frac{1}{\lambda(I_{ij}^k)} \int_{I_{ij}^k} \frac{d\pi}{d\lambda} d\lambda \right) \leq \frac{1}{\lambda(I_{ij}^k)} \int_{I_{ij}^k} \Phi \left( \frac{d\pi}{d\lambda} \right) d\lambda.$$

With this we obtain

$$\begin{aligned} \int_{\Omega} \Phi \left( \frac{d\pi_k}{d\lambda} \right) d\lambda &= \int_{\Omega} \Phi \left( \sum_{i,j} \frac{\pi(I_{ij}^k)}{\lambda(I_{ij}^k)} \mathbb{1}_{I_{ij}^k} \right) d\lambda = \sum_{i,j} \int_{I_{ij}^k} d\lambda \Phi \left( \frac{\pi(I_{ij}^k)}{\lambda(I_{ij}^k)} \right) = \sum_{i,j} \lambda(I_{ij}^k) \Phi \left( \frac{\pi(I_{ij}^k)}{\lambda(I_{ij}^k)} \right) \\ &\leq \sum_{i,j} \int_{I_{ij}^k} \Phi \left( \frac{d\pi}{d\lambda} \right) d\lambda = \int_{\Omega} \Phi \left( \frac{d\pi}{d\lambda} \right) d\lambda \end{aligned}$$

which shows the lim sup-condition.  $\square$

$\Gamma$ -convergence for *simultaneously* decreasing the regularization parameter *and* refining the discretization proves to be a harder problem. The following example shows that the convergence rate of  $\gamma_k$  must be linked to the convergence rate of the discretization.



**Example 4.8.** Let  $\Omega = [0, 1]^2$ ,  $\mu_1 = \mu_2 = \delta_0$  the Dirac measure at zero and  $\lambda_i = \mathcal{L}_i$  the Lebesgue measure on  $\Omega_i$ . Then clearly  $\pi = \delta_0$  is the only feasible and thus the optimal transport plan. Let  $0 < h_k$  be a sequence, which is to be chosen.

We consider  $\Phi(t) = t \log t$  and have the discretized optimal plan

$$\pi_k = \frac{\pi([0, h_k]^2)}{\lambda([0, h_k]^2)} \mathbb{1}_{[0, h_k]^2} = \frac{\mathbb{1}_{[0, h_k]^2}}{h_k^2}.$$

However, it holds that

$$\gamma_k \int_{\Omega} \Phi(\pi_k) d\lambda = \gamma_k \log(h_k^{-2}),$$

and hence, the lim sup condition can not be fulfilled if  $h_k$  is such that  $\gamma_k \log(h_k^{-2}) \not\rightarrow 0$  (which holds, e.g., for  $h_k = \exp(-1/\gamma_k)$ ).

Let now  $\tilde{\lambda}_i = \delta_0$  be the Dirac measure at zero. The discretized optimal plan now is

$$\tilde{\pi}_k = \frac{\pi([0, h_k]^2)}{\tilde{\lambda}([0, h_k]^2)} \mathbb{1}_{[0, h_k]^2} = \mathbb{1}_{[0, h_k]^2}, \quad \text{with } \tilde{\lambda} := \tilde{\lambda}_1 \otimes \tilde{\lambda}_2$$

and hence,

$$\gamma_k \int_{\Omega} \Phi(\tilde{\pi}_k) d\tilde{\lambda} = \gamma_k \Phi(\tilde{\pi}_k(0)) = \gamma_k \Phi(1) = 0.$$

Thus,  $h_k$  may be chosen arbitrary in this case.

Using the insights of [Example 4.8](#) the desired result can be formulated.

**Theorem 4.9.** Under the assumptions of [Theorem 4.7](#), let  $(h_k)$  be a positive sequence with  $h_k \leq \lambda(I_{ij}^k)$  for all  $i, j$  and let  $(\gamma_k)$  be a sequence converging to zero such that

$$\gamma_k \Phi_+(h_k^{-1}) \rightarrow 0. \quad (4.4)$$

Then the minimization problem

$$\inf \left\{ \int_{\Omega} c d\pi + \gamma_k \int_{\Omega} \Phi_{\infty} \left( \frac{d\pi}{d\lambda} \right) d\lambda \mid \pi \ll \lambda, 0 \leq \frac{d\pi}{d\lambda} \in \mathcal{V}^k, \right. \\ \left. \frac{d(P_1)_{\#} \pi}{d\lambda_1} = \sum_i \frac{\mu_{1,i}^k}{\lambda_{1,i}^k} \mathbb{1}_{Q_i^{1,k}}, \frac{d(P_1)_{\#} \pi}{d\lambda_2} = \sum_j \frac{\mu_{2,j}^k}{\lambda_{2,j}^k} \mathbb{1}_{Q_j^{2,k}} \right\}$$

$\Gamma$ -converges to (OT) as  $k \rightarrow \infty$  w.r.t. weak-\* convergence in  $\mathfrak{M}(\Omega)$ .

*Proof.* Let  $F_k$  be defined as in the proof of [Theorem 4.7](#) and let  $F : \mathfrak{M}(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}$  be defined via

$$F(\pi) = \begin{cases} \int_{\Omega} c d\pi, & 0 \leq \pi \in \mathcal{P}(\Omega), (P_i)_{\#} \pi = \mu_i, i = 1, 2, \\ \infty, & \text{else.} \end{cases}$$

Given an arbitrary  $\pi \in \mathfrak{M}(\Omega)$ , check now the two conditions for  $\Gamma$ -convergence.

i) *lim inf-condition*: Let  $\pi$  be any measure and  $(\pi_k)$  be such that  $\pi_k \xrightarrow{*} \pi$ .

In the case  $F(\pi) = \infty$  we can see that  $F_k(\pi_k) < \infty$  has to be violated infinitely often (otherwise  $\pi$  would have to be feasible as in the proof of [Theorem 4.7](#)). For  $F(\pi) < \infty$  we get

$$\int_{\Omega} c \, d\pi_k \rightarrow \int_{\Omega} c \, d\pi.$$

Moreover, since  $\Phi$  is bounded from below, we can extract a subsequence such that  $\int_{\Omega} \Phi\left(\frac{d\pi_k}{d\lambda}\right) d\lambda$  is bounded from above and below and obtain

$$\lim_{k \rightarrow \infty} \gamma_k \int_{\Omega} \Phi\left(\frac{d\pi_k}{d\lambda}\right) d\lambda = 0$$

which proves  $\liminf F_k(\pi_k) \geq F(\pi)$ .

ii) *lim sup-condition*: For  $F(\pi) = \infty$  we have nothing to prove.

If  $F(\pi) < \infty$ , define  $\pi_k$  as in the proof of [Theorem 4.7](#). Then, the marginal constraints are satisfied and  $\pi_k \xrightarrow{*} \pi$ . Hence  $\int_{\Omega} c \, d\pi_k \rightarrow \int_{\Omega} c \, d\pi$  and it remains to show that

$$\limsup_{k \rightarrow \infty} \gamma_k \int_{\Omega} \Phi\left(\frac{d\pi_k}{d\lambda}\right) d\lambda \leq 0.$$

Using  $0 \leq \pi(I_{ij}^k) \leq \pi(\Omega) = 1$  and monotonicity of  $\Phi_+$  we get

$$\begin{aligned} \int_{\Omega} \Phi_{\infty}\left(\frac{d\pi_k}{d\lambda}\right) d\lambda &= \sum_{ij} \Phi\left(\frac{\pi(I_{ij}^k)}{\lambda(I_{ij}^k)}\right) \lambda(I_{ij}^k) \leq \sum_{ij} \Phi_+\left(\frac{\pi(I_{ij}^k)}{\lambda(I_{ij}^k)}\right) \lambda(I_{ij}^k) \\ &\leq \sum_{ij} \Phi_+\left(\frac{1}{\lambda(I_{ij}^k)}\right) \lambda(I_{ij}^k) \leq \Phi_+(h_k^{-1}) \sum_{ij} \lambda(I_{ij}^k) \\ &= \Phi_+(h_k^{-1}) \lambda(\Omega) = \Phi_+(h_k^{-1}). \end{aligned} \tag{4.5}$$

Hence,

$$\gamma_k \int_{\Omega} \Phi\left(\frac{d\pi_k}{d\lambda}\right) d\lambda \leq \gamma_k \Phi_+(h_k^{-1}) \rightarrow 0$$

as desired.  $\square$

**Corollary 4.10.** *Theorems 4.7 and 4.9 remain true if instead of a continuous cost function  $c$ , a sequence of step functions  $(c_k)$ , constant on the sets  $I_{ij}^k$  and converging to  $c$  uniformly, is used.*

*Proof.* Rewrite the integral  $\int_{\Omega} c_k \, d\pi_k$  as the dual pairing  $\langle c_k, \pi_k \rangle$ , where  $(c_k) \subset C_b(\Omega)$  converges strongly and  $\pi_k \subset \mathfrak{M}(\Omega)$  converges weakly-\*. Because the variation norm  $\|\pi_k\|$  is bounded ([\[23, Corollary 2.6.10\]](#)),  $\langle c_k, \pi_k \rangle \rightarrow \langle c, \pi \rangle$  can be seen by standard arguments, which yields the assertion.  $\square$

Similarly to [Corollary 4.3](#), for  $t^{-1}\Phi(t)$  monotone the assumption on  $(\gamma_k)$  can be weakened.

**Corollary 4.11.** *Let  $\Phi$  be a quasi-Young's function such that  $t^{-1}\Phi(t)$  is monotone. Then it suffices to assume*

$$\gamma_k h_k \Phi_+(h_k^{-1}) \rightarrow 0$$

instead of condition (4.4) in Theorem 4.9.

*Proof.* Using the monotonicity of  $t^{-1}\Phi(t)$ , (4.5) can be refined to

$$\begin{aligned} \int_{\Omega} \Phi_{\infty}\left(\frac{d\pi_k}{d\lambda}\right) d\lambda &= \sum_{ij} \Phi\left(\frac{\pi(I_{ij}^k)}{\lambda(I_{ij}^k)}\right) \frac{\lambda(I_{ij}^k)}{\pi(I_{ij}^k)} \frac{\pi(I_{ij}^k)}{\lambda(I_{ij}^k)} \lambda(I_{ij}^k) \leq \sum_{ij} \Phi_+\left(\frac{\pi(I_{ij}^k)}{\lambda(I_{ij}^k)}\right) \frac{\lambda(I_{ij}^k)}{\pi(I_{ij}^k)} \pi(I_{ij}^k) \\ &\leq \sum_{ij} \Phi_+\left(\left(\lambda(I_{ij}^k)\right)^{-1}\right) \lambda(I_{ij}^k) \pi(I_{ij}^k) \leq h_k \Phi_+(h_k^{-1}) \sum_{ij} \pi(I_{ij}^k) \\ &= h_k \Phi_+(h_k^{-1}) \pi(\Omega) = h_k \Phi_+(h_k^{-1}) \end{aligned}$$

and the assertion follows as in Theorem 4.9.  $\square$

## 5. Conclusion

Employing regularization in Orlicz spaces to the optimal transport problem allows to generalize the existence results of [10, 20] for both the primal and the predual problem and under mild assumptions, the results hold for regularization w.r.t. product measures  $\lambda = \lambda_1 \otimes \lambda_2$ . More precisely, primal solutions exist if and only if the marginals are functions in the appropriate Orlicz spaces and existence of optimizers in  $L^q$  for the predual problem has been shown for the special case  $\Phi(t) = t^p/p$ ,  $p \geq 2$ .

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 gives a generalization of the corresponding result [10, Theorem 5.1] for  $\Phi(t) = t \log t$ . Similarly, a combined regularization and discretization approach leads to another family of approximations. It could be proven that, again,  $\Gamma$ -convergence is guaranteed if the regularization parameter and the discretization fineness are coupled in an appropriate way.

Existence of solutions of the dual problem for general (quasi-)Young's functions has been considered in a different framework in [18]. Still, future work might investigate, if the result can also be achieved by the approach considered here. Moreover, numerical methods for solving the regularized problem (P) have not been discussed here.

## Acknowledgements

The authors would like to thank two anonymous reviewers for a number of valuable comments and suggestions regarding the presentation.

## Appendix

### A. Proof of Lemma 2.6

*Proof.* Let  $\Phi(t) := \int_0^t \varphi(s) ds$  as in Definition 2.1. Then with  $\tilde{\varphi} := \varphi \cdot \mathbb{1}_{(t_0, \infty)}$  it holds that  $\tilde{\Phi}(t) := \int_0^t \tilde{\varphi}(s) ds$  and hence,  $\tilde{\Phi}$  is a Young's function.

To show  $L^{\tilde{\Phi}}(\Omega, d\nu) = L^{\Phi}(\Omega, d\nu)$  it suffices to show that for any  $f$

$$\|f\|_{L^{\Phi}(\Omega, d\nu)} < \infty \Leftrightarrow \|f\|_{L^{\tilde{\Phi}}(\Omega, d\nu)} < \infty.$$

If  $\|f\|_{L^{\Phi}(\Omega, d\nu)} < \infty$ , then  $\|f\|_{L^{\tilde{\Phi}}(\Omega, d\nu)} < \infty$  trivially. Let therefore  $\|f\|_{L^{\tilde{\Phi}}(\Omega, d\nu)} < \infty$ , i.e. there is a  $\gamma$  such that

$$\int_{\Omega} \tilde{\Phi}\left(\frac{|f(x)|}{\gamma}\right) dx = \int_{\Omega_{\gamma}(f)} \Phi\left(\frac{|f(x)|}{\gamma}\right) dx - \nu(\Omega_{\gamma}(f))\Phi(t_0) \leq 1,$$

where  $\Omega_{\gamma}(f) := \left\{x \mid \frac{|f(x)|}{\gamma} \geq t_0\right\}$ . Let  $r \in (0, 1]$  and write

$$\int_{\Omega} \Phi\left(r \frac{|f(x)|}{\gamma}\right) dx = \left( \int_{\left\{x \mid \frac{|f(x)|}{\gamma} \leq t_0\right\}} + \int_{\left\{x \mid t_0 \leq \frac{|f(x)|}{\gamma} \leq \frac{t_0}{r}\right\}} + \int_{\left\{x \mid \frac{t_0}{r} \leq \frac{|f(x)|}{\gamma}\right\}} \right) \Phi\left(r \frac{|f(x)|}{\gamma}\right) dx.$$

Since  $\Phi(0) = 0$  and  $\Phi$  is convex, for every  $s \in [0, 1]$  it holds that  $\Phi(sx) \leq s\Phi(x)$  and we obtain an upper bound for the first integral by

$$\int_{\left\{x \mid \frac{|f(x)|}{\gamma} \leq t_0\right\}} \Phi\left(r \frac{|f(x)|}{\gamma}\right) dx \leq r \int_{\left\{x \mid \frac{|f(x)|}{\gamma} \leq t_0\right\}} \Phi\left(\frac{|f(x)|}{\gamma}\right) dx \leq rC$$

for some constant  $C$ . With the same argument,

$$\begin{aligned} \int_{\left\{x \mid \frac{t_0}{r} \leq \frac{|f(x)|}{\gamma}\right\}} \Phi\left(r \frac{|f(x)|}{\gamma}\right) dx &\leq r \int_{\left\{x \mid \frac{t_0}{r} \leq \frac{|f(x)|}{\gamma}\right\}} \Phi\left(\frac{|f(x)|}{\gamma}\right) dx \leq r \int_{\left\{x \mid t_0 \leq \frac{|f(x)|}{\gamma}\right\}} \Phi\left(\frac{|f(x)|}{\gamma}\right) dx \\ &= r \left( \int_{\Omega} \tilde{\Phi}\left(\frac{|f(x)|}{\gamma}\right) dx + \nu(\Omega_{\gamma}(f))\Phi(t_0) \right) \\ &\leq r (1 + \nu(\Omega_{\gamma}(f))\Phi(t_0)). \end{aligned}$$

Finally, since  $\left\{x \mid t_0 \leq \frac{|f(x)|}{\gamma} \leq \frac{t_0}{r}\right\} \subseteq \left\{x \mid t_0 \leq \frac{|f(x)|}{\gamma}\right\}$ , the same estimate holds for the second integral. Combining the estimates for the three integrals one obtains

$$\int_{\Omega} \Phi\left(r \frac{|f(x)|}{\gamma}\right) dx \leq r (C + 2 \cdot (1 + \nu(\Omega_{\gamma}(f))\Phi(t_0))),$$

which is less or equal than 1 for  $r$  small enough. □

## B. Proof of Lemma 4.1

*Proof.* We only treat the case  $i = 1$  (the case  $i = 2$  being analogous). First note that

$$\begin{aligned} \frac{d(P_1)_\# \pi}{dv_1}(x_1) &= \int_{\mathbb{R}^n} \frac{d\pi}{dv}(x_1, x_2) dv_2(x_2) = \int_{\mathbb{R}^n} \frac{d\pi}{dv}(x_1, x_2) dv_2(x_2) \int_{\mathbb{R}^n} \varphi_\delta(y_2) dy_2 \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \varphi_\delta(y_2 - x_2) \frac{d\pi}{dv}(x_1, x_2) dy_2 dv_2(x_2), \end{aligned}$$

where we used Fubini's theorem. Using this (remembering  $G_\delta = \varphi_\delta \otimes \varphi_\delta$ ) and Fubini again we get

$$\begin{aligned} (\varphi_\delta * (P_1)_\# \pi)(y_1) &= \int_{\mathbb{R}^n} \varphi_\delta(y_1 - x_1) d((P_1)_\# \pi)(x_1) = \int_{\mathbb{R}^n} \varphi_\delta(y_1 - x_1) \frac{d(P_1)_\# \pi}{dv_1}(x_1) dv_1(x_1) \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n \times \mathbb{R}^n} G_\delta(y_1 - x_1, y_2 - x_2) \frac{d\pi}{dv}(x_1, x_2) dv(x_1, x_2) dy_2 \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n \times \mathbb{R}^n} G_\delta(y_1 - x_1, y_2 - x_2) d\pi(x_1, x_2) dy_2 \\ &= \int_{\mathbb{R}^n} (G_\delta * \pi)(y_1, y_2) dy_2 = (P_1)_\# \pi_\delta(y_1). \end{aligned}$$

Then, for every measurable set  $A \subset \mathbb{R}^n$ ,

$$\begin{aligned} \mu_1^\delta(A) &= \int_A (\varphi_\delta * \mu_1)(x) dx = \int_A (\varphi_\delta * (P_1)_\# \pi)(x) dx \\ &= \int_A ((P_1)_\# \pi_\delta)(x) dx = ((P_1)_\# \pi_\delta)(A), \end{aligned}$$

which concludes the proof.  $\square$

## C. Proof of Lemma 4.6

*Proof.* Let  $A \subset \Omega$  be a Borel set,  $\varepsilon > 0$  and  $A_-^k, A_+^k$  as in Assumption 4.4. Then for all  $i, j$  it holds

$$\lambda(I_{i,j}^k \cap A_+^k) = \begin{cases} \lambda(I_{i,j}^k), & I_{i,j}^k \cap A \neq \emptyset, \\ 0, & \text{else} \end{cases}$$

and similarly for  $\lambda(I_{i,j}^k \cap A_-^k)$ . In combination with (4.3) this yields

$$\begin{aligned} \int_A v_k d\lambda - v(A) &\leq \int_{A_+^k} v_k d\lambda - v(A_+^k) = \sum_{i,j} \frac{v(I_{i,j}^k)}{\lambda(I_{i,j}^k)} \lambda(I_{i,j}^k \cap A_+^k) - v(A_+^k) \\ &= \sum_{\{(i,j) \mid I_{i,j}^k \cap A \neq \emptyset\}} v(I_{i,j}^k) - v(A_+^k) = v(A_+^k) - v(A_+^k) < \varepsilon, \end{aligned}$$

for  $k$  large enough. Using an analogous argument for a lower bound, we get

$$\left| \int_A v_k d\lambda - v(A) \right| < \varepsilon,$$

which yields the first assertion. Analogous argumentation proves the result for  $v^i \in \mathfrak{M}_+(\Omega_i)$ .  $\square$

## References

- [1] L. Ambrosio and N. Gigli. A user’s guide to optimal transport. In *Modelling and Optimisation of Flows on Networks*, pages 1–155. Springer, 2013. doi:[10.1007/978-3-642-32160-3\\_1](https://doi.org/10.1007/978-3-642-32160-3_1).
- [2] 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(2):A1111–A1138, 2015. doi:[10.1137/141000439](https://doi.org/10.1137/141000439).
- [3] 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).
- [4] M. Blondel, V. Seguy, and A. Rolet. Smooth and sparse optimal transport, 2017. arXiv:[1710.06276](https://arxiv.org/abs/1710.06276).
- [5] V. I. Bogachev. *Measure theory*. Springer, Berlin New York, 2007.
- [6] J. Borwein and Q. J. Zhu. *Techniques of variational analysis*. Springer, New York, 2005.
- [7] 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).
- [8] 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, 12 2015. doi:[10.1137/15M1050264](https://doi.org/10.1137/15M1050264).
- [9] A. Caruso. Two properties of norms in Orlicz spaces. *Le Matematiche*, 56(1):183–194, 2001.
- [10] C. Clason, D. A. Lorenz, H. Mahler, and B. Wirth. Entropic regularization of continuous optimal transport problems. *Journal of Mathematical Analysis and Applications*, 494(1):124432, Feb. 2021. doi:[10.1016/j.jmaa.2020.124432](https://doi.org/10.1016/j.jmaa.2020.124432).
- [11] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transportation distances. *Advances in Neural Information Processing Systems*, 26, 06 2013.
- [12] M. Cuturi, G. Peyré, and A. Rolet. A smoothed dual approach for variational Wasserstein problems. *SIAM Journal on Imaging Sciences*, 9, 03 2015. doi:[10.1137/15M1032600](https://doi.org/10.1137/15M1032600).
- [13] A. Dessein, N. Papadakis, and J.-L. Rouas. Regularized optimal transport and the rot mover’s distance. *The Journal of Machine Learning Research*, 19(1):590–642, 2018.
- [14] L. Diening, P. Harjulehto, P. Hästö, and M. Růžička. *Lebesgue and Sobolev Spaces with Variable Exponents*, volume 2017 of *Lecture Notes in Mathematics*. Springer, 4 2011. doi:[10.1007/978-3-642-18363-8](https://doi.org/10.1007/978-3-642-18363-8).
- [15] G. B. Folland. *Real analysis : modern techniques and their applications*. Wiley, New York, 1999.
- [16] I. Fonseca and G. Leoni. *Modern Methods in the Calculus of Variations:  $L^p$  Spaces*. Springer Monographs in Mathematics. Springer, 2007. doi:[10.1007/978-0-387-69006-3](https://doi.org/10.1007/978-0-387-69006-3).
- [17] A. Genevay. *Entropy-regularized optimal transport for machine learning*. PhD thesis, Université Paris Dauphine, 2019.
- [18] C. Léonard. Minimization of entropy functionals. *Journal of Mathematical Analysis and Applications*, 346(1):183 – 204, 2008. doi:[10.1016/j.jmaa.2008.04.048](https://doi.org/10.1016/j.jmaa.2008.04.048).

- [19] C. Léonard. Convex minimization problems with weak constraint qualifications. *Journal of Convex Analysis*, 17(1):321–348, 2010.
- [20] D. Lorenz, P. Manns, and C. Meyer. Quadratically regularized optimal transport. *Applied Mathematics & Optimization*, 09 2019. doi:[10.1007/s00245-019-09614-w](https://doi.org/10.1007/s00245-019-09614-w).
- [21] D. A. Lorenz and H. Mahler. Orlicz-space regularization for optimal transport and algorithms for quadratic regularization, 2019. arXiv:[1909.06082](https://arxiv.org/abs/1909.06082).
- [22] S. D. Marino and A. Gerolin. Optimal transport losses and Sinkhorn algorithm with general convex regularization, 2020. arXiv:[2007.00976](https://arxiv.org/abs/2007.00976).
- [23] R. E. Megginson. *An Introduction to Banach Space Theory (Graduate Texts in Mathematics)*. Springer, 1998. doi:[10.1007/978-1-4612-0603-3](https://doi.org/10.1007/978-1-4612-0603-3).
- [24] J. Musielak. *Orlicz Spaces and Modular Spaces*. Springer Berlin Heidelberg, 1983. doi:[10.1007/bfb0072210](https://doi.org/10.1007/bfb0072210).
- [25] B. Muzellec, R. Nock, G. Patrini, and F. Nielsen. Tsallis regularized optimal transport and ecological inference. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), Feb. 2017. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/10854>.
- [26] F.-P. Paty and M. Cuturi. Regularized optimal transport is ground cost adversarial. In H. Daumé III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7532–7542. PMLR, 13–18 Jul 2020. URL: <http://proceedings.mlr.press/v119/paty20a.html>.
- [27] G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. doi:[10.1561/22000000073](https://doi.org/10.1561/22000000073).
- [28] M. M. Rao and Z. D. Ren. *Theory of Orlicz Spaces*. Pure and Applied Mathematics. Dekker, 1991.
- [29] L. Roberts, L. Razoumov, L. Su, and Y. Wang. Gini-regularized optimal transport with an application to spatio-temporal forecasting, 2017. arXiv:[1712.02512](https://arxiv.org/abs/1712.02512).
- [30] R. T. Rockafellar. Integrals which are convex functionals. *Pacific J. Math.*, 24:525–539, 1968. doi:[10.2140/pjm.1968.24.525](https://doi.org/10.2140/pjm.1968.24.525).
- [31] F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling (Progress in Nonlinear Differential Equations and Their Applications)*. Birkhäuser, 2015. doi:[10.1007/978-3-319-20828-2](https://doi.org/10.1007/978-3-319-20828-2).
- [32] F.-X. Vialard. An elementary introduction to entropic regularization and proximal methods for numerical optimal transport. Lecture, May 2019. URL: <https://hal.archives-ouvertes.fr/hal-02303456>.
- [33] 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](https://doi.org/10.1007/978-3-540-71050-9).