On the Wasserstein distance between mutually singular measures

Giuseppe Buttazzo, Guillaume Carlier, Maxime Laborde

July 1, 2017

Abstract

We study the Wasserstein distance between two measures μ, ν which are mutually singular. In particular, we are interested in minimization problems of the form

$$W(\mu, \mathcal{A}) = \inf \big\{ W(\mu, \nu) \ : \ \nu \in \mathcal{A} \big\}$$

where μ is a given probability and $\mathcal{A}$ is contained in the class $\mu^{\perp}$ of probabilities that are singular with respect to μ. Several cases for $\mathcal{A}$ are considered; in particular, when $\mathcal{A}$ consists of L^1 densities bounded by a constant, the optimal solution is given by the characteristic function of a domain. Some regularity properties of these optimal domains are also studied. Some numerical simulations are included, as well as the double minimization problem

$$\min \big\{ P(B) + kW(A, B) \ : \ |A \cap B| = 0, \ |A| = |B| = 1 \big\},$$

where $k > 0$ is a fixed constant, $P(A)$ is the perimeter of A, and both sets A, B may vary.

Keywords: Wasserstein distance, singular measures, perimeter penalization.
2010 Mathematics Subject Classification: 49J45, 49M29, 49Q20, 49A99, 49B99.

1 Introduction

In this paper, we consider, for a given probability measure μ on $\mathbb{R}^d$, the optimization problem

$$W(\mu, \mathcal{A}) = \inf \big\{ W(\mu, \nu) \ : \ \nu \in \mathcal{A} \big\}, \tag{1.1}$$

where W denotes the p-Wasserstein distance ($p \geq 1$ is fixed) and $\mathcal{A}$ is a class of probabilities that are singular with respect to μ. For all the background about mass transportation and Wasserstein distances, we refer to the books [14] and [11].

Problems of this kind arise in some models of bilayer membranes, for which we refer for instance to [9] and references therein. Here we consider only the mathematical issues, that appear to be very rich.

When $\mathcal{A}$ coincides with the class $\mu^{\perp}$ of all the probabilities that are singular with respect to μ, the optimization problem (1.1) becomes trivial, in the sense that $W(\mu, \mathcal{A}) = 0$, as shown in Proposition 3.1. The same happens when $\mathcal{A} = \mu^{\perp} \cap L^1$ and μ is singular with respect to the Lebesgue measure $\mathcal{L}^d$ (see Proposition 3.2).

On the contrary, when μ has a nonzero absolutely continuous part with respect to $\mathcal{L}^d$ and $\mathcal{A} = \mu^{\perp} \cap L^1$, the optimization problem (1.1) has a nontrivial *generalized* solution ν providing a nonzero minimal value $W(\mu, \mathcal{A})$ (see Proposition 3.6). If $\mu \in L^1$ (or in

slightly more general cases, see Remark 3.7) this probability ν can be expressed through the distance function $d(x)$ of x to the boundary $\partial S(\mu)$ of a concentration set $S(\mu)$ of μ:

$$\nu = \big(x - d(x)\nabla d(x)\big)^{\#}\mu$$

being $\#$ the push-forward operator. In some more regular cases $S(\mu)$ reduces to spt μ and some explicit examples are provided in Example 3.8 and Example 3.9.

The most interesting situation occurs when in the optimization problem (1.1) we impose an upper bound on the competing probabilities ν; more precisely, we take $\mathcal{A} = \mathcal{A}_\phi$, where

$$\mathcal{A}_\phi = \big\{\rho \in \mathcal{P}_c \cap L^1 \ : \ \rho \perp \mu, \ \rho \le \phi\big\},$$

being ϕ a fixed nonnegative integrable function, with $\int \phi\,dx > 1$. In this case the minimum value $W(\mu, \mathcal{A}_\phi)$ is reached by a characteristic function; more precisely, in Theorem 3.10 we show that

$$W(\mu, \mathcal{A}_\phi) = W(\mu, \phi 1_A)$$

for a suitable set A. Under additional assumptions on μ, we show in Theorem 3.13 that this set A has a finite perimeter.

This allows us to consider in Section 4 the joint minimization problem with a perimeter penalization

$$\min\big\{P(B) + kW(A, B) \ : \ |A \cap B| = 0, \ |A| = |B| = 1\big\},$$

where $k > 0$ is a fixed parameter and both A and B may vary. Here we denoted shortly by $W(A, B)$ the Wasserstein distance between the characteristic functions 1_A, 1_B. We show in Theorem 4.1 that an optimal solution A^*, B^* exists and prove some regularity results, namely that A^* has finite perimeter and that B^* is a quasi-minimizer of the perimeter.

Finally, in Section 5 we present some numerical simulations in the case $p = 2$.

2 Notation and preliminaries

In the following, our ambient space is $\mathbb{R}^d$; we denote by $\mathcal{P}_c$ the class of all probabilities on $\mathbb{R}^d$ with compact support. Analogously, we denote by L^p the space of p-integrable functions on $\mathbb{R}^d$ and, for a given nonnegative $\phi \in L^p$, by L^p_ϕ the class of nonnegative functions $u \in L^p$ with $u \le \phi$. We recall the following definitions for measures on $\mathbb{R}^d$.

- μ is said *concentrated* on a Borel set A if

$$\mu(E) = \mu(E \cap A) \qquad \text{for every Borel set } E$$

  or equivalently

$$\mu(E \setminus A) = 0 \qquad \text{for every Borel set } E.$$

- μ is absolutely continuous with respect to ν if

$$\mu(E) = 0 \qquad \text{whenever } \nu(E) = 0.$$

  In this case, we use the notation $\mu \ll \nu$. By the Radon-Nikodym derivation theorem, when $\mu \ll \nu$ there exists a unique (up to ν a.e.) nonnegative function $h \in L^1_\nu$ such that

$$\mu(E) = \int_E h\,d\nu \qquad \text{for every Borel set } E.$$

In this case, we use the notation $\mu = h\nu$. The function h above can be obtained (ν a.e.) as

$$h(x) = \lim_{r \to 0} \frac{\mu\big(B_r(x)\big)}{\nu\big(B_r(x)\big)}$$

where $B_r(x)$ is the ball of radius r centered at x.

- μ and ν are *mutually singular* if there exists a Borel set A such that μ is concentrated on A and ν is concentrated on $\mathbb{R}^d \setminus A$, that is

$$\mu(E \setminus A) = \nu(E \cap A) = 0 \qquad \text{for every Borel set } E.$$

In this case we use the notation $\mu \perp \nu$. For a fixed $\mu \in \mathcal{P}_c$ we denote by $\mu^\perp$ the class

$$\mu^\perp = \big\{\nu \in \mathcal{P}_c \ : \ \nu \perp \mu\big\}.$$

- The *Lebesgue decomposition* of μ with respect to ν is the unique way of writing $\mu = \mu_1 + \mu_2$ with $\mu_1 \ll \nu$ and $\mu_2 \perp \nu$. The measures μ_1 and μ_2 are called the *absolutely continuous part* and the *singular part* of μ with respect to ν. When the measure ν is fixed in the context, we write $\mu = \mu^a + \mu^s$ and, by the Radon-Nikodym derivation theorem we then have

$$\mu = h\nu + \mu^s \qquad \text{with } h \in L^1_\nu \text{ and } \mu^s \perp \mu.$$

Often, when no ambiguity is possible, we simply write h instead of $h\nu$ identifying an absolutely continuous measure (with respect to ν) with its L^1_ν density.

For a fixed $p \geq 1$ and $\mu, \nu \in \mathcal{P}_c$ we denote by $W(\mu, \nu)$ the Wasserstein distance

$$W(\mu, \nu) = \left(\min \Big\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \, d\gamma \ : \ \gamma \in \Pi(\mu, \nu) \Big\} \right)^{1/p},$$

where $\Pi(\mu, \nu)$ is the class of probabilities γ on $\mathbb{R}^d \times \mathbb{R}^d$ having μ and ν as *marginals*, that is

$$\pi_1^\# \gamma = \mu \qquad \text{and} \qquad \pi_2^\# \gamma = \nu,$$

being $\#$ the push-forward operation defined, for a map $f : X \to Y$ between two measurable spaces, as

$$(f^\# \mu)(E) = \mu\big(f^{-1}(E)\big) \qquad \text{for all measurable } E \text{ in } Y.$$

In the sequel, we denote by $\mathcal{L}^d$ the Lebesgue measure on $\mathbb{R}^d$ and by δ_x the Dirac measure at the point x.

3 Some optimization problems

In this section, we fix a probability $\mu \in \mathcal{P}_c$ and we consider the Wasserstein distance ($p \geq 1$ is fixed) from μ to some subclasses $\mathcal{A} \subset \mu^\perp$. In other words, we consider the optimization problem

$$W(\mu, \mathcal{A}) = \inf \big\{ W(\mu, \nu) \ : \ \nu \in \mathcal{A} \big\}.$$

The first case we consider is $\mathcal{A} = \mu^\perp$.

Proposition 3.1. *For every $\mu \in \mathcal{P}_c$ we have $W(\mu, \mu^\perp) = 0$.*

Proof. Let μ^s be the singular part of μ with respect to the Lebesgue measure $\mathcal{L}^d$; the measure μ^s is concentrated on a set N which is negligible with respect to $\mathcal{L}^d$. Then we can find a sequence (x_n) in $\mathbb{R}^d \setminus N$ which is dense in $\mathbb{R}^d$, and a sequence (μ_n) in $\mathcal{P}_c$ of the form

$$\mu_n = \sum_{k \in \mathbb{N}} a_{n,k} \delta_{x_k}$$

such that $\mu_n \to \mu$ in the weak* convergence. By the choice of the sequence (x_n) the measures μ_n are singular with respect to μ and, since $\mu_n \to \mu$ weakly*, we have $W(\mu, \mu_n) \to 0$, as required. $\qquad\square$

In the next step, we consider the class

$$\mathcal{A} = \mu^\perp \cap L^1 = \left\{ \nu \in \mathcal{P}_c \ : \ \nu \perp \mu, \ \nu \ll \mathcal{L}^d \right\}.$$

Proposition 3.2. *For every $\mu \perp \mathcal{L}^d$ we have $W(\mu, \mu^\perp \cap L^1) = 0$.*

Proof. Let N be a negligible set with respect to $\mathcal{L}^d$ where μ is concentrated. We can find a sequence (x_n) in $\mathbb{R}^d \setminus N$ dense in $\mathbb{R}^d$ and a sequence (μ_n) in $\mathcal{P}_c$ of the form

$$\mu_n = \sum_{k \in \mathbb{N}} a_{n,k} \delta_{x_k}$$

such that $\mu_n \to \mu$ in the weak* convergence. Since L^1 is weakly* dense in $\mathcal{P}_c$ we can find L^1 functions ρ_n with compact support such that $W(\mu_n, \rho_n) \leq 1/n$. Since $\mu \perp \mathcal{L}^d$ we have that $\rho_n \in \mu^\perp \cap L^1$ and so

$$W(\mu, \mu^\perp \cap L^1) \leq \lim_n W(\mu, \rho_n) \leq \lim_n W(\mu, \mu_n) + W(\mu_n, \rho_n) = 0$$

as required. $\qquad\square$

The situation becomes more interesting when μ is not singular with respect to $\mathcal{L}^d$.

Proposition 3.3. *Let $\mu \in \mathcal{P}_c$ and assume there exists $\delta > 0$ such that the set $\{\mu^a \geq \delta\}$ contains an open set. Then $W(\mu, \mu^\perp \cap L^1) > 0$.*

Proof. Assume by contradiction that there exists a sequence (ρ_n) in $\mu^\perp \cap L^1$ with $\rho_n \to \mu$ in the weak* convergence. Then we have

$$\int \rho_n \mu^a \, dx = 0 \qquad \text{for every } n \in \mathbb{N}.$$

Let A be an open set contained in $\{\mu^a \geq \delta\}$; then

$$\int_A \rho_n \, dx \leq \frac{1}{\delta} \int_{\{\mu^a \geq \delta\}} \rho_n \mu^a \, dx \leq \frac{1}{\delta} \int \rho_n \mu^a \, dx = 0 \qquad \text{for every } n \in \mathbb{N}.$$

Since A is open we have

$$\mu(A) \leq \liminf_n \int_A \rho_n \, dx = 0,$$

which is impossible, since $\mu^a \geq \delta$ on A. $\qquad\square$

Corollary 3.4. *If $\mu \in \mathcal{P}_c$ is such that μ^a is (or coincides a.e. with) a lower semicontinuous function and $\mu^a \neq 0$, then $W(\mu, \mu^\perp \cap L^1) > 0$.*

4

Example 3.5. The regularity assumption of Proposition 3.3 cannot be removed. Take indeed in dimension one an open dense set A in $(-1, 1)$ with unit measure and take $\mu = 1_{A^c}$. Since A is dense, by finite sums of Dirac masses at points of A, we can weakly* approximate every measure supported in $[-1, 1]$; hence approximating a Dirac mass by a smooth function we can construct a sequence ρ_n of smooth functions, compactly supported in A, that converge to μ in the weak* convergence of measures. We have then

$$\mu^a \neq 0 \qquad \text{and} \qquad \int \rho_n \, d\mu = 0 \quad \forall n \in \mathbb{N}.$$

In the next proposition, we characterize the quantity $W(\mu, \mu^\perp \cap L^1)$ (which is positive under the assumption of Proposition 3.3).

Proposition 3.6. *Let $\mu \in \mathcal{P}_c$. Then there exists $\nu \in \mathcal{P}_c$ such that*

$$W(\mu, \mu^\perp \cap L^1) = W(\mu, \nu). \tag{3.1}$$

The measure ν is concentrated on ∂S, where S is a concentration set for μ. Moreover, if $\mu \in L^1$ we have that ν is unique and given by

$$\nu = \left(x - d(x) \nabla d(x) \right)^{\#} \mu \tag{3.2}$$

where $d(x)$ denotes the distance of the point x to ∂S.

Proof. Let $\rho_n \in \mu^\perp \cap L^1$ be such that

$$W(\mu, \mu^\perp \cap L^1) = \lim_n W(\mu, \rho_n). \tag{3.3}$$

Then, up to a subsequence, we may assume that $\rho_n \to \nu$ weakly*, for a suitable $\nu \in \mathcal{P}_c$. Hence

$$\lim_n W(\mu, \rho_n) = W(\mu, \nu).$$

In order to see that ν is concentrated on ∂S, where S is a concentration set for μ, notice that, heuristically, to achieve the minimal Wasserstein distance, every point on S has to be transported out of the S in the shortest way. Hence ν has to be concentrated on ∂S. To prove this fact in a precise way, since $\rho_n \perp \mu$ there is a set R_n such that ρ_n is concentrated on R_n and μ is concentrated on R_n^c. If we take $S = \cap_n R_n^c$ we have

$$\nu(\text{int } S) \leq \liminf_n \int_{\text{int } S} \rho_n \, dx \leq \liminf_n \int_{R_n^c} \rho_n \, dx = 0,$$

which proves that ν is concentrated on $(\text{int } S)^c$. We show now that ν is concentrated on the compact set $\overline{S}$. Otherwise, we could find a compact K set disjoint from $\overline{S}$ such that $\nu(K) > 0$, since both K and $\overline{S}$ are compact there is an open neighbourhood U of K whose closure is still disjoint from K. Hence there is open set U with compact closure such that $\text{dist}(\overline{U}, \overline{S}) = 4\delta > 0$ and $\nu(U) > 0$. Since U is open, taking a subsequence if necessary, we can also assume that for some $\alpha > 0$, we have $\rho_n(U) \geq \alpha$ for every n. We are going to show that this implies the existence of a sequence $\hat{\rho}_n \in \mu^\perp \cap L^1$ and of a positive constant $C > 0$ such that for every n

$$W(\mu, \hat{\rho}_n) \leq W(\mu, \rho_n) - C , \tag{3.4}$$

which contradicts (3.3). We first choose an almost optimal map T_n such that $T_n^{\#} \rho_n = \mu$ and

$$\int |T_n(y) - y|^p \, d\rho_n(y) \leq W^p(\mu, \rho_n) + 2^{-n}.$$

For $y \in U$, let $F_n(y)$ be a point on the segment $[y, T_n(y)]$ lying at distance 2δ from $\overline{S}$ (measurable selection arguments imply that this map can be chosen measurable). By construction, for ρ_n a.e. $y \in U$ we then have that

$$|T_n(y) - y| \geq |T_n(y) - F_n(y)| + 2\delta. \tag{3.5}$$

Let then η be a smooth compactly supported probability density having its support in the unit ball B_δ of radius δ and centered at the origin and define the plan $\hat{\gamma}_n$ by

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x, y) \, d\hat{\gamma}_n(x, y) := \int_{U^c} \varphi(T_n(y), y) \, d\rho_n(y) + \int_{U \times \mathbb{R}^d} \varphi(T_n(y), F_n(y) - z) \, d\rho_n(y) \, d\eta(z)$$

for every $\varphi \in C(\mathbb{R}^d \times \mathbb{R}^d)$. Obviously the first marginal of $\hat{\gamma}_n$ is μ and we denote by $\hat{\rho}_n$ its second marginal, i.e. $\hat{\rho}_n$ is the sum of the restriction of ρ_n to U^c and the convolution of η with the image of ρ_n restricted to U by F_n, it is therefore in L^1. Also $\hat{\rho}_n \in \mu^\perp$ since $F_n(y) - z$ remains at distance δ from $\overline{S}$ for ρ_n-a.e. $y \in U$ and every z in B_δ. Since $\hat{\gamma}_n \in \Pi(\mu, \hat{\rho}_n)$, we have

$$W^p(\mu, \hat{\rho}_n) \leq \int_{U^c} |y - T_n(y)|^p \, d\rho_n(y) + \int_U \left(|F_n(y) - T_n(y)| + \delta \right)^p \, d\rho_n(y).$$

On the other hand, recalling (3.5)

$$W^p(\mu, \rho_n) + 2^{-n} \geq \int_{U^c} |y - T_n(y)|^p \, d\rho_n(y) + \int_U \left(|F_n(y) - T_n(y)| + 2\delta \right)^p \, d\rho_n(y)$$

which, together with $\rho_n(U) \geq \alpha > 0$, proves (3.4).

Assume now that $\mu \in L^1$, so that we may take S as the set of points x such that

$$\lim_{r \to 0} \frac{\mu(B_r(x))}{|B_r(x)|} > 0.$$

Then the optimal transport map is given by

$$T(x) = x - d(x)\nabla d(x) \tag{3.6}$$

where $d(x)$ denotes the distance of the point x to ∂S, and so the measure ν is given by (3.2). The uniqueness of ν follows from the fact that $\mu \in L^1$ and that $\nabla d(x)$ is well defined for a.e. point $x \in S$. $\qquad \square$

Remark 3.7. When μ has a singular part with respect to $\mathcal{L}^d$ formula (3.2) still provides a measure ν which verifies (3.1). In this case, writing $\mu = \mu^a + \mu^s$ it is easy to see that $\nu = \mu$ on $\mathbb{R}^d \setminus S(\mu^a)$ where $S(\mu^a)$ denotes a concentration set for μ^a, hence in this region the Wasserstein cost vanishes. On the other hand, if μ^s does not vanish on $S(\mu^a)$, it is transported on $\partial S(\mu^a)$ by the transport map in (3.6); the only case to be made precise is when on $\left(S(\mu^s) \right) \cap \left(S(\mu^a) \right)$ the function d is not differentiable, and so ∇d is not defined. In this case, these singular points x have more than one projection on $\partial S(\mu^a)$, and every choice of $T(x)$ as one of these projections (or also as a transport plan, sending x in any subset of its projections) gives a measure ν verifying (3.1).

Example 3.8. Let $\mu = 1_Q$ be the characteristic function of a rectangle Q with sides a and b, with $a \leq b$; we assume for simplicity that the center of the rectangle is at the origin, that is $Q = [-b/2, b/2] \times [-a/2, a/2]$. Using Proposition 3.6 we obtain that the measure ν in (3.1) is concentrated on the boundary of Q and its boundary density ρ is given by:

$$\rho = \begin{cases} a/2 - |y| & \text{on the vertical sides} \\ (b/2 - |x|) \wedge a/2 & \text{on the horizontal sides.} \end{cases}$$

In Figure 1, we represent the boundary density on ∂Q as a boundary thickness.
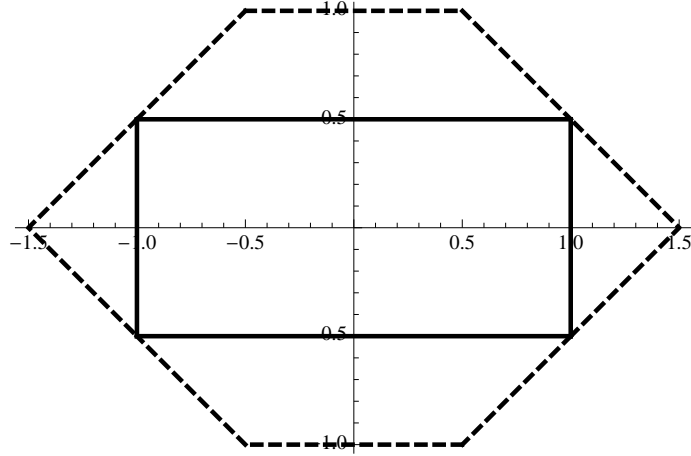
Figure 1: The optimal boundary density (dashed line) on a rectangle.

Example 3.9. Let $\mu = 1_{B^+}$ the characteristic function of the half circle B^+ of radius 1; we assume for simplicity that the center is at the origin, that is $B^+ = \{x^2 + y^2 \leq 1, \ y \geq 0\}$. Using Proposition 3.6 we obtain that the measure ν in (3.1) is concentrated on the boundary of B^+ and the optimal transport map is

$$T(x,y) = \begin{cases} (x,0) & \text{if } 2y < 1 - x^2 \\ \dfrac{(x,y)}{\sqrt{x^2 + y^2}} & \text{if } 2y > 1 - x^2. \end{cases}$$

The boundary density ρ can be then obtained by elementary calculations as

$$\rho = \begin{cases} (1 - x^2)/2 & \text{on the bottom diameter, } -1 \leq x \leq 1 \\ \dfrac{1}{2}\left(1 - \dfrac{1}{(1 + \sin\theta)^2}\right) & \text{on the upper boundary, } 0 \leq \theta \leq \pi. \end{cases}$$

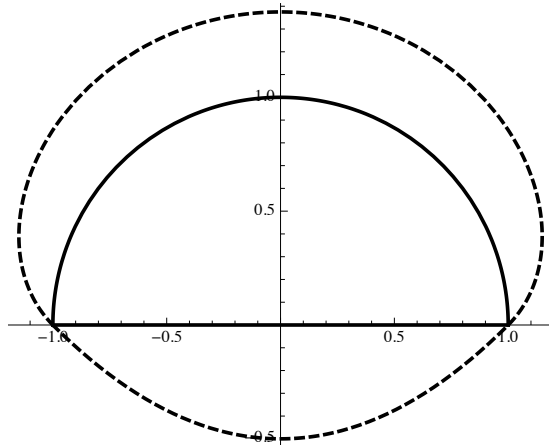In Figure 2, we represent the boundary density on ∂B^+ as a boundary thickness.

Figure 2: The optimal boundary density (dashed line) on a half circle.

We consider now, as a class $\mathcal{A} \subset \mu^\perp$ the class

$$\mathcal{A}_\phi = \mu^\perp \cap L^1_\phi = \left\{ \rho \in \mathcal{P}_c \cap L^1 \ : \ \rho \perp \mu, \ \rho \leq \phi \right\},$$

where ϕ is a given nonnegative integrable function with $\int \phi \, dx > 1$.

Theorem 3.10. *For every $\mu \in \mathcal{P}_c$ there exists a set A with $\mu(A) = 0$ and such that*

$$W(\mu, \mathcal{A}_\phi) = W(\mu, \phi 1_A).$$

Proof. We have

$$W(\mu, \mathcal{A}_\phi) = \inf \left\{ W(\mu, \theta\phi) \ : \ 0 \leq \theta \leq 1, \ \int \theta\phi \, dx = 1, \ \int \theta\phi\mu^a \, dx = 0 \right\}. \qquad (3.7)$$

Moreover, since μ is compactly supported, we may reduce ourselves to consider in (3.7) only functions θ supported in a ball B_R with R large enough. Then, by the weak* L^∞ compactness of bounded L^∞ classes, in (3.7) the infimum is actually a minimum; we denote by θ a minimizer. We can assume that $\phi > 0$ since on the set where ϕ vanishes one may take $\theta = 0$ and there is nothing to prove. We want to show that the set $\{0 < \theta(x) < 1\}$ is Lebesgue negligible. There are at most countably many connected components of the support of $\theta\phi$ which are of positive Lebesgue measure; we denote them by A_n and we are left to show that for each n, and each $\delta \in (0, 1)$, the set $E_\delta \cap A_n$ with $E_\delta = \{\delta \leq \theta(x) \leq 1 - \delta\}$ is Lebesgue negligible. Assume by contradiction that it has a positive measure. Then, for every $h \in L^\infty$ such that

$$h = 0 \text{ outside } E_\delta \cap A_n, \qquad \int h\phi \, dx = 0, \qquad (3.8)$$

for ε small enough $\theta + \varepsilon h$ is admissible for the above minimization problem. Using the dual Kantorovich formulation and denoting by u_ε a Kantorovich potential between μ and $\nu_\varepsilon := (\theta + \varepsilon h)\phi$, we have

$$\int \phi h u_\varepsilon \, dx \leq W(\mu, \theta\phi) - W(\mu, \nu_\varepsilon) \leq 0.$$

Being Kantorovich potentials uniformly continuous, it may be assumed that, up to extraction of some subsequence, u_ε converges to some Kantorovich potential u for $W(\mu, \theta\phi)$. A priori any such cluster point may depend on h, but it is also known (see for instance [11]) that Kantorovich potentials [1] are defined uniquely up to an additive constant on each A_n. In other words, u does not depend on h up to possibly an additive constant, but with (3.8) we have $\int \phi h u \, dx \leq 0$, hence by changing h into $-h$ we in fact have $\int \phi h u \, dx = 0$. Since h is arbitrary we deduce that u coincides a.e. with a constant on $E_\delta \cap A_n$ and we then also have $\nabla u = 0$ a.e. on $E_\delta \cap A_n$. Therefore the optimal transport T from $\theta\phi$ to μ is the identity map on $E_\delta \cap A_n$ and this implies that $\mu \geq \delta\phi 1_{E_\delta \cap A_n}$ which clearly contradicts the fact that $\int \theta\phi\mu^a \, dx = 0$. $\qquad \square$

Remark 3.11. It is well-known (see [11]) that when $\mu \in L^1 \cap \mathcal{P}_c$ and $p > 1$, $\nu \mapsto W^p(\mu, \nu)$ is strictly convex (note that the value of an optimal transport is always convex with respect to the marginals), in this case the optimal θ in (3.7) is unique and then so is the set A.

---

[1]For a differentiable transport cost, this is well-known but for the distance case one has to use the fact that in addition the measures μ and $\theta\phi$ are orthogonal, hence if γ is an optimal plan it gives no mass to the diagonal which is the set where the cost is not differentiable.

In general, we should not expect that the set A of Theorem 3.10 has a finite perimeter, as the example below shows.

Example 3.12. Take $\phi = 1$, $\mu = \sum_{n \in \mathbb{N}} c_n \delta_{x_n}$ where x_n is the center of a ball of radius $r_n = (c_n/\omega_d)^{1/d}$ (being ω_d the Lebesgue measure of the unit ball in $\mathbb{R}^d$). We may choose the balls $B(x_n, r_n)$ all disjoint, so that the set A of Theorem 3.10 coincides with $\cup_{n \in \mathbb{N}} B(x_n, r_n)$. We then have

$$P(A) = d\omega_d^{1/d} \sum_{n \in \mathbb{N}} c_n^{(d-1)/d}.$$

To have $\mathcal{L}^d(A) = 1$ and $P(A) = +\infty$ it is now enough to choose c_n such that

$$\sum_{n \in \mathbb{N}} c_n = 1, \qquad \sum_{n \in \mathbb{N}} c_n^{(d-1)/d} = +\infty.$$

A possible array of the balls $B(x_n, r_n)$ is shown in Figure 3.
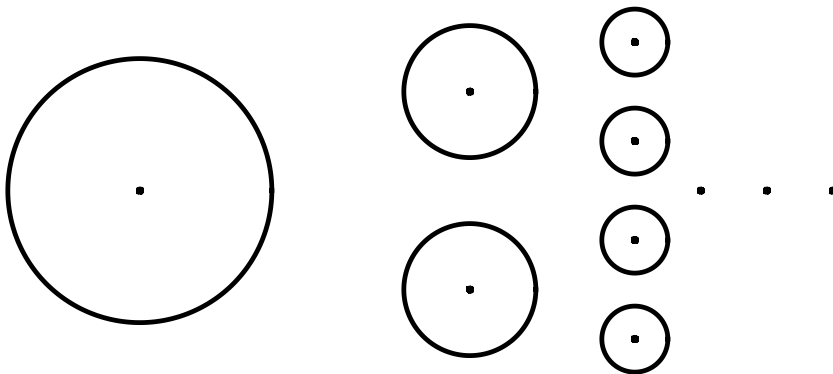
Figure 3: An optimal domain A without finite perimeter.

By a suitable approximation of Dirac masses by smooth functions, we may construct a counterexample similar to Example 3.12, of a $\mu \in L^\infty$ for which the set A of Theorem 3.10 does not have a finite perimeter. Therefore, some extra assumptions on μ are needed in order to have $P(A) < +\infty$.

Theorem 3.13. *Let $p = 2$ and $\phi = 1$; let $\mu \in \mathcal{P}_c \cap BV$ be such that the set $S_\mu = \{\mu(x) > 0\}$ has a finite perimeter. Then the set A of Theorem 3.10 has a finite perimeter. More precisely, we have*

$$P(A) \le \int |\nabla \mu| + 2P(S_\mu).$$

Proof. It is enough to apply Theorem 1.2 of [6] with $\Omega = \mathbb{R}^d$, $g = \mu$ and $f = 1_{\{\mu=0\}}$. $\square$

Remark 3.14. Both Theorem 1.2 in [6] and Theorem 3.13 are stated in the case $p = 2$. Similar results hold also in the case $p \ge 1$, as it was communicated to us by S. Di Marino [7].

4   Perimeter penalization

In this section, we consider the minimum problem

$$(\mathcal{P}) \qquad \min \left\{ P(B) + kW(B, A) \ : \ |A \cap B| = 0, \ |A| = |B| = 1 \right\}, \qquad (4.1)$$

where $k > 0$ is a fixed constant and $P(B)$ denotes the perimeter of the set B. Note that the minimization above if performed with respect to both sets A, B. We also consider, for a fixed bounded Lipschitz domain D of $\mathbb{R}^d$ with $|D| > 1$, the problem

$$(\mathcal{P}_D) \qquad \min\big\{P(B) + kW(B, A) \ : \ B \subset D, \ |A \cap B| = 0, \ |A| = |B| = 1\big\}. \quad (4.2)$$

In both problems $(\mathcal{P})$ and $(\mathcal{P}_D)$, we have slightly abused notations, denoting $W(B, A)$ the Wasserstein distance between the uniform measures on B and A respectively.

Theorem 4.1. *For every bounded domain D of $\mathbb{R}^d$ and for every $k > 0$ the minimization problem (4.2) admits a solution.*

Proof. Let (B_n, A_n) be a minimizing sequence for the minimization problem (4.2). Since $P(B_n)$ are bounded and $B_n \subset D$, possibly passing to subsequences we may assume that $B_n \to B^*$ strongly in L^1. Analogously, since 1_{A_n} are bounded by 1 and compactly supported, we may assume that $1_{A_n} \rightharpoonup \bar{\theta}$ weakly* in L^∞ for a suitable $\bar{\theta}$ with $0 \le \bar{\theta} \le 1$. We have

$$\int_{B^*} \bar{\theta} \, dx = \lim_n \int_{B_n} 1_{A_n} \, dx = 0.$$

By Theorem 3.10, the minimization problem

$$\min\Big\{W(B^*, \theta) \ : \ \int \theta = 1, \ 0 \le \theta \le 1, \ \int_{B^*} \theta = 0\Big\}$$

admits a solution which is the characteristic function of a set A^*. By the minimality of A^* we have

$$W(B^*, A^*) \le W(B^*, \bar{\theta}) = \lim_n W(B_n, A_n)$$

and, by the lower semicontinuity of the perimeter with respect to the strong L^1-convergence, we deduce that

$$P(B^*) + kW(B^*, A^*) \le \liminf_n \Big(P(B_n) + kW(B_n, A_n)\Big),$$

which concludes the proof $\qquad\qquad\square$

In the two-dimensional case, $d = 2$, we can take $D = \mathbb{R}^2$ in Theorem 4.1.

Theorem 4.2. *In the case $d = 2$ for every $k > 0$ the minimization problem (4.1) admits a solution.*

Proof. We can repeat the proof of Theorem 4.1 as soon as we can show that for a suitable minimizing sequence (B_n, A_n) the sets B_n remain uniformly bounded. Let then (B_n, A_n) be a minimizing sequence for problem (4.1), let $B_{n,k}$ be the connected components of B_n and let $A_{n,k}$ be the part of A_n that is transported on $B_{n,k}$. Then we have

$$\begin{aligned} P(B_n) &= \sum_k P(B_{n,k}) \\ W^p(B_n, A_n) &= \sum_k W^p(B_{n,k}, A_{n,k}) \ge \sum_k W^p(B_{n,k}, \bar{A}_{n,k}) \end{aligned} \qquad (4.3)$$

where $\bar{A}_{n,k}$ denotes a solution of the minimum problem

$$\min\big\{W(B_{n,k}, A) \ : \ |A| = |B_{n,k}|, \ |A \cap B_{n,k}| = 0\big\}.$$

We can now construct a new minimizing sequence $(\tilde{B}_n, \tilde{A}_n)$ by translating the sets $B_{n,k}$ and $\bar{A}_{n,k}$ in such a way that $\tilde{B}_{n,k}$ is contained in a square of side $P(B_{n,k})$ and $\tilde{A}_{n,k}$ in a concentric square of side $2P(B_{n,k})$. Arranging these squares in an array we obtain that

$$\mathrm{diam}(\tilde{B}_n) \leq \sum_k 4\,\mathrm{diam}(B_{n,k}) \leq \sum_k 2P(B_{n,k}) = 2P(B_n)$$

which shows that the sets $\tilde{B}_n$ are uniformly bounded. The fact that $(\tilde{B}_n, \tilde{A}_n)$ is still a minimizing sequence for problem (4.1) follows from the fact that

$$P(\tilde{B}_n) = P(B_n)$$
$$W^p(B_n, A_n) \geq \sum_k W^p(B_{n,k}, \bar{A}_{n,k}) = \sum_k W^p(\tilde{B}_{n,k}, \tilde{A}_{n,k}) = W^p(\tilde{B}_n, \tilde{A}_n)$$

that are consequences of (4.3). □

Remark 4.3. Even if we expect that a result similar to the one in Theorem 4.2 holds for every dimension, the proof we provided uses the fact that for a connected set its diameter is bounded by its perimeter, which only holds in dimension two. It would be interesting to find an alternative proof of Theorem 4.2 valid for every dimension d.

Example 4.4. In the one-dimensional case $d = 1$, if W is the p-Wasserstein distance ($p \geq 1$), an easy calculation gives that, taking B as the union of n disjoint equal intervals (of length $1/n$ each) and A which surrounds them symmetrically,

$$P(B) + kW(B, A) = 2n + \frac{k}{2n}.$$

Then the optimal solution (B^*, A^*) of problem (4.1) is obtained by taking B^* given by

$$\begin{array}{ll} 1 \text{ interval} & \text{if } 0 \leq k \leq 8 \\ 2 \text{ intervals} & \text{if } 8 \leq k \leq 24 \\ \qquad \cdots & \\ n \text{ intervals} & \text{if } 4(n-1)n \leq k \leq 4n(n+1). \end{array}$$

We now address the regularity of optimal B's. Indeed, let D be a bounded domain, $p \geq 1$, $k > 0$, $\alpha > 0$ and consider the (slightly more general) shape optimization problem

$$\inf_{B \subset D,\ |B|=1} P(B) + F^\alpha(B) \tag{4.4}$$

where

$$F(B) := \inf_{A \subset D,\ |A \cap B|=0,\ |A|=1} W_p^p(1_A, 1_B).$$

If B solves (4.4) it satisfies

$$P(B) \leq P(B') + k(F^\alpha(B') - F^\alpha(B)) \leq P(B') + C(F(B') - F(B)) \tag{4.5}$$

for every $B' \subset D$ with $|B'| = 1$ (where we have used the fact that $F(B')$ is bounded and bounded away from 0). We now have

Lemma 4.5. *There is a constant M such that for any B and B', subsets of D with same unit measure,*

$$F(B') - F(B) \leq M|B \Delta B'|.$$

Proof. Write $B' = (B \setminus B^-) \cup B^+$ with $|B^- \setminus B| = 0$, $|B^+ \cap B| = 0$, $\sigma = |B^-| = |B^+|$, $|B \Delta B'| = 2\sigma$. Let A such that $A \subset D$, $|A \cap B| = 0$, $|A| = 1$. Set $A^- := B^+ \cap A$ and choose $A^+ \subset B^-$ of same measure (less than σ) as A^- and then set $A' := (A \setminus A^-) \cup A^+$. By construction, we have $A' \subset D$, $|A' \cap B'| = 0$, $|A'| = 1$. It follows from Lemma 6.2 in [1] that for any $\gamma \in \Pi(1_A, 1_B)$ there exists a $\gamma' \in \Pi(1_{A'}, 1_{B'})$ such that $W_1(\gamma, \gamma') \le W_1(1_A, 1_{A'}) + W_1(1_B, 1_{B'}) \le C\sigma$, and then since $(x, y) \mapsto |x - y|^p$ is Lipschitz on $D \times D$, we also have

$$F(B') \le \int_{A' \times B'} |x - y|^p \, d\gamma'(x, y) \le \int_{A \times B} |x - y|^p \, d\gamma(x, y) + M\sigma$$

minimizing the right-hand side in A and $\gamma \in \Pi(1_A, 1_B)$ gives the desired result. $\qquad \square$

Thanks to (4.5), Lemma 4.5 and the theory of quasi-minimizers of the perimeter (see the seminal work of Tamanini [12] and Xia [13] for the case of a volume constraint as in the present context), we have:

Theorem 4.6. *If B solves (4.4), its reduced boundary $\partial^* B$ is a $C^{1,1/2}$ hypersurface and the Hausdorff dimension of $(\partial B \setminus \partial^* B) \cap D$ is at most $d - 8$.*

5 Numerical simulations

In this section, we give some numerical simulations of problem (3.7) where $p = 2$,

$$\bar{\theta} = \operatorname{argmin}\left\{W(\mu, \theta) \ : \ 0 \le \theta \le \phi, \ \int \theta \, dx = 1, \ \int \theta \mu^a \, dx = 0\right\}.$$

Theorem 3.10 and Remark 3.11 say that there exists a unique minimizer and $\bar{\theta}$ is the characteristic function of a domain multiplied by ϕ. This problem can be rewritten as

$$\bar{\gamma} = \operatorname{argmin}\left\{\iint |x - y|^2 \, d\gamma(x, y) + G_1(\gamma) + G_2(\gamma) + G_3(\gamma)\right\}, \tag{5.1}$$

where, $\bar{\theta} = \pi_1^\# \bar{\gamma}$ and

$$G_1(\gamma) = \chi_{\{\pi_1^\# \gamma = \mu\}} := \begin{cases} 0 & \text{if } \pi_1^\# \gamma = \mu \\ +\infty & \text{otherwise} \end{cases}, \quad G_2(\gamma) = \chi_{\{\pi_2^\# \gamma \le \phi\}}, \quad G_3(\gamma) = \chi_{\{\pi_2^\# \gamma \perp \mu\}}.$$

In the discrete setting, μ and θ are replaced by

$$\mu = \sum_{i=1}^N \mu_i \delta_{x_i}, \qquad \theta = \sum_{j=1}^N \theta_i \delta_{x_j},$$

where $\{x_i\}_{i=1}^N$ is a discretization of D, a compact subset of $\mathbb{R}^2$ such that $\operatorname{spt} \mu, \operatorname{spt} \bar{\theta} \subset D$. Then problem (5.1) becomes

$$\bar{\gamma} = \operatorname{argmin}_{\gamma \in \mathbb{R}^{N \times N}}\left\{\sum_{i,j=1}^N c_{i,j} \gamma_{i,j} + G_1(\gamma) + G_2(\gamma) + G_3(\gamma)\right\},$$

where $c_{i,j} = |x_i - x_j|^2$ and the marginal maps are defined by

$$(\pi_1^\# \gamma)_i = \sum_{j=1}^N \gamma_{i,j}, \qquad (\pi_2^\# \gamma)_j = \sum_{i=1}^N \gamma_{i,j}.$$

This problem can be easily solved using the well-known entropic regularization method, [3, 8, 2]. It consists in regularizing W_2 by the entropy of the transport plan. Given a regularization parameter $\varepsilon > 0$, we solve

$$\min_{\gamma \in \mathbb{R}_+^{N \times N}} \left\{ \sum_{i,j=1}^{N} c_{i,j} \gamma_{i,j} + \varepsilon \sum_{i,j=1}^{N} \gamma_{i,j} \big(\ln(\gamma_{i,j}) - 1 \big) + G_1(\gamma) + G_2(\gamma) + G_3(\gamma) \right\},$$

or,

$$\min_{\gamma \in \mathbb{R}_+^{N \times N}} \left\{ KL(\gamma | \eta_\varepsilon) + \frac{1}{\varepsilon} \big(G_1(\gamma) + G_2(\gamma) + G_3(\gamma) \big) \right\}, \tag{5.2}$$

where $(\eta_\varepsilon)_{i,j} = e^{-c_{i,j}/\varepsilon}$ and KL is the Kullback-Leibler divergence

$$KL(\gamma | \eta) := \sum_{i,j=1}^{N} \gamma_{i,j} \left(\ln \left(\frac{\gamma_{i,j}}{\eta_{i,j}} \right) - 1 \right).$$

Solving problem (5.2) is equivalent to solve the proximal problem

$$\mathrm{prox}_G^{KL}(\eta_\varepsilon) = \mathrm{argmin}_{\gamma \in \mathbb{R}_+^{N \times N}} \left\{ KL(\gamma | \eta_\varepsilon) + G(\gamma) \right\},$$

where $G := G_1 + G_2 + G_3$. This problem can be solved using the proximal splitting algorithm introduced by Peyré in the setting of entropic regularization of Wasserstein gradient flows, [10]. This scheme was extended in [5] to unbalanced transport problem and used in [4] to compute Cournot-Nash equilibria. It is well known that the solution of (5.2) is of the form $\overline{\gamma}_{i,j} = a_i(\eta_\varepsilon)_{i,j} b_j$, where $a, b \in \mathbb{R}^N$ and $\eta_\varepsilon \in \mathbb{R}^{N \times N}$. The splitting proximal algorithm corresponds to alternate proximal problems with G_l, $l \in \{1, 2, 3\}$, instead of solving directly (5.2). In our special case, we initialize the algorithm by

$$a_i^0 = c_j^0 = d_j^0 = 1 \quad \text{and} \quad b_j^0 = c_j^0 d_j^0 \qquad \text{for all } i, j \in \{1, \ldots, N\},$$

and $\gamma_{i,j}^0 = a_i^0 (\eta_\varepsilon)_{i,j} b_j^0$, then iteratively, for $k \geq 1$,

$$a_i^k = \frac{\mathrm{prox}_{G_1}^{KL} \left(\big(\sum_{j=1}^{N} (\eta_\varepsilon)_{i,j} b_j^{k-1} \big)_i \right)_i}{\sum_{j=1}^{N} (\eta_\varepsilon)_{i,j} b_j^{k-1}},$$

$$c_j^k = \frac{\mathrm{prox}_{G_2}^{KL} \left(\big(\sum_{i=1}^{N} a_i^k (\eta_\varepsilon)_{i,j} d_j^{k-1} \big)_j \right)_j}{\sum_{i=1}^{N} a_i^k (\eta_\varepsilon)_{i,j} d_j^{k-1}},$$

$$d_j^k = \frac{\mathrm{prox}_{G_3}^{KL} \left(\big(\sum_{i=1}^{N} a_i^k (\eta_\varepsilon)_{i,j} c_j^k \big)_j \right)_j}{\sum_{i=1}^{N} a_i^k (\eta_\varepsilon)_{i,j} c_j^k}.$$

Then $b_j^k = c_j^k d_j^k$ and $\gamma_{i,j}^k = a_i^k (\eta_\varepsilon)_{i,j} b_j^k$. We refer to [10, 5] for the convergence of this algorithm to a solution of (5.2). The advantage of this method is that computing $\mathrm{prox}_{G_l}^{KL}$ can be done easily. Indeed, for all $\theta \in \mathbb{R}^N$,

$$\mathrm{prox}_{G_1}^{KL}(\theta)_i = \mu_i ,$$
$$\mathrm{prox}_{G_2}^{KL}(\theta)_j = \min(\phi, \theta_j) ,$$
$$\mathrm{prox}_{G_3}^{KL}(\theta)_j = \theta_j 1_{\{\mu_j = 0\}} .$$

We now present some numerical results obtained using this algorithm. In the sequel, all computations are done with Matlab, using $N = 500 \times 500$ in the discretization of

13

$D = [-4, 4]^2$, except for the triangle in Figure 5 where $D = [-2, 2]^2$. In Figures 4 and 5, we represent the regularized solution of problem (3.7), $\overline{\theta}_\varepsilon = \pi_2^\# \gamma^k$, with k large enough and $\varepsilon = 0.01$. In Figure 4, μ is the white rectangle or the white half circle, $\phi = 1$, and we remark that $\overline{\theta}_\varepsilon$ is the characteristic function of the black set, as expected. In Figure 5, the initial densities are given by the characteristic functions of a triangle or a non-convex Pacman shape or a disconnected domain given by the union of rectangles and ellipses. The first column represents the regularized solution of problem (3.7) with $\phi = 1$ and in the second column, $\phi(x, y) = (x+1)^2 + 1$. We remark that the solution $\overline{\theta}_\varepsilon$, with initial density the characteristic function of a non-convex Pacman shape, fills the hole in μ, when $\phi = 1$, but not entirely in the case $\phi(x, y) = (x + 1)^2 + 1 \geq 1$. In both cases, the support of the solution is not connected. As expected, in the second column, the regularized solutions of problem (3.7), are given by ϕ multiplied by the characteristic function of a set.
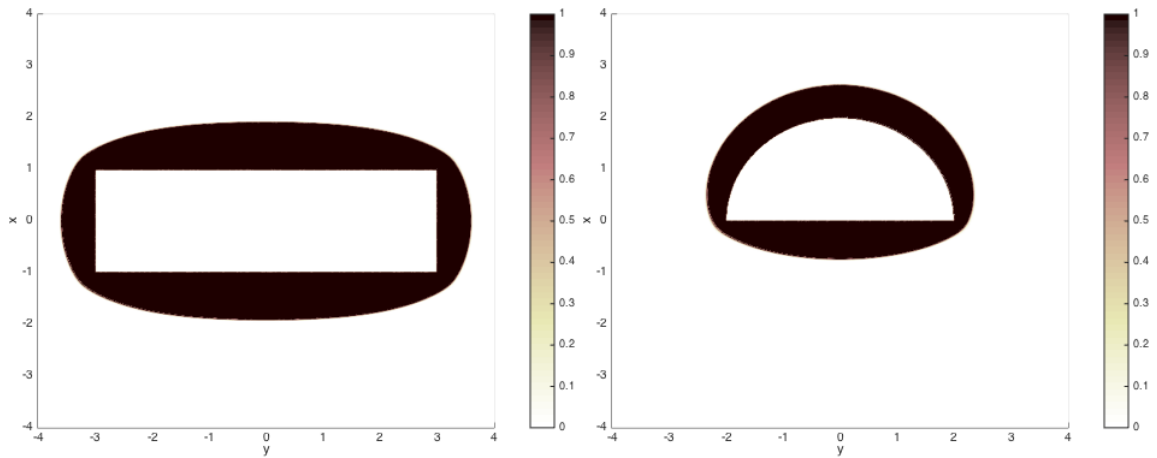

Figure 4: The optimal density for a rectangle and for a half circle.

-

References

[1] J.-B. Baillon, G. Carlier: *From discrete to continuous Wardrop equilibria.* Networks and Hetereogeneous Media, **7** (2) (2012), 219–241.

[2] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, G. Peyré: *Iterative Bregman projections for regularized transportation problems.* SIAM J. Sci. Compt., 37(2):A1111-A1138, 2015..

[3] M. Cuturi: *Sinkhorn distances: Lightspeed computation of optimal transport.* In Advancesin Neural InformationProcessing systems (Nips), 26, pages 2292-2300,2013.
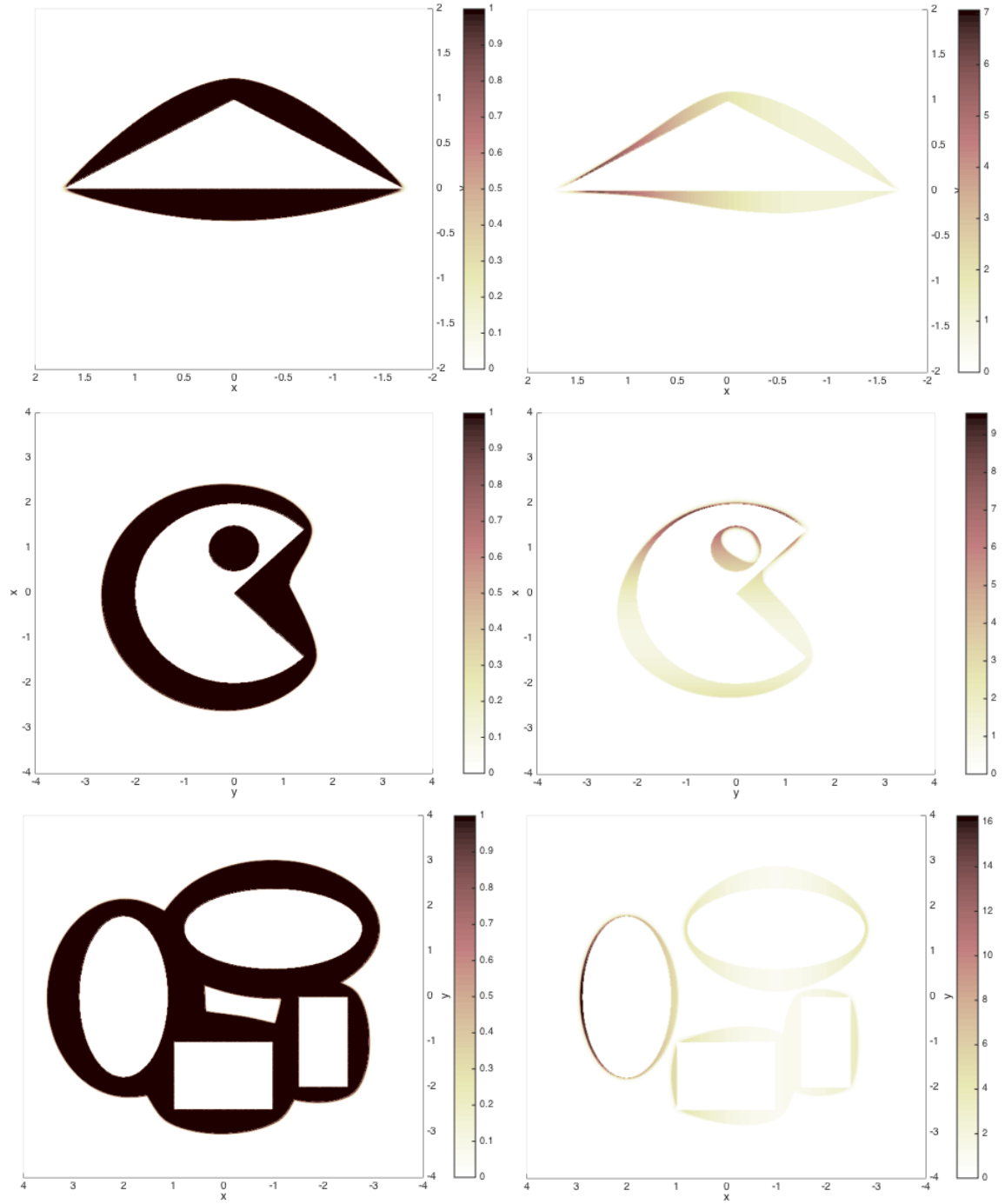
Figure 5: The optimal density for a triangle, for a pacman like domain and for a disconnected domain. In the first column, $\phi = 1$ and in the second column, the congestion function ϕ is given by $\phi(x,y) = (x+1)^2 + 1$.

[4] A. BLANCHET, G. CARLIER, L. NENNA: *Computation of Cournot-Nash equilibiria by entropic regularization.* preprint, 2016.

[5] L. CHIZAT, G. PEYRÉ, B. SCHMITZER, F.-X. VIALARD: *Scaling algorithms for unbalanced transport problems.* Technical report,

http://arxiv.org/abs/1607.05816,2016.

[6] G. De Philippis, A.R. Mészáros, F. Santambrogio, B. Velichkov: *BV Estimates in Optimal Transportation and Applications.* Arch. Rational Mech. Anal., **219** (2016), 829–860.

[7] S. Di Marino: *The five gradients inequality for $p \geq 1$.* Private communication.

[8] A. Galichon, B. Salanié: *Matching with trade-offs: Revealed preferences over competing characters.* Technical report, Preprint SSRN-1487307, 2009.

[9] L. Lussardi, M.A. Peletier, M. Röger: *Variational analysis of a mesoscale model for bilayer membranes.* J. Fixed Point Theory Appl., **15** (1) (2014), 217–240.

[10] G. Peyré: *Entropic approximation of Wasserstein gradient flows.* SIAM J. Imaging Sci., 8(4):2323-2351, 2015..

[11] F. Santambrogio: *Optimal Transport for Applied Mathematicians.* Progress in Nonlinear Differential Equations and Their Applications **87**, Birkhäuser, Basel (2015).

[12] I. Tamanini Boundaries of Caccioppoli sets with Hölder-continuous normal vector J. Reine Angew. Math. **334** (1982), 27–39

[13] Q. Xia Regularity of minimizers of quasi perimeters with a volume constraint Interfaces and Free Boundaries **7** (3) 2005, 339–352

[14] C. Villani: *Optimal Transport. Old and New.* Grundlehren der mathematischen Wissenschaften **338**, Springer, Berlin (2009).

Giuseppe Buttazzo: Dipartimento di Matematica, Università di Pisa
Largo B. Pontecorvo 5, 56127 Pisa - ITALY
buttazzo@dm.unipi.it
http://www.dm.unipi.it/pages/buttazzo/

Guillaume Carlier: CEREMADE UMR CNRS 7534, Université de Paris Dauphine
Place de Lattre de Tassigny, 75775 Paris Cedex 16 - FRANCE
carlier@ceremade.dauphine.fr
https://www.ceremade.dauphine.fr/∼carlier/

Maxime Laborde: Department of Mathematics and Statistics, McGill University, Montréal
805 Rue Sherbrooke Ouest, Montréal-CANADA
maxime.laborde@mcgill.ca