A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions

Alessio Figalli[*], Nicola Gigli[†]

Abstract

In this paper we introduce a new transportation distance between non-negative measures inside a domain Ω. This distance enjoys many nice properties, for instance it makes the space of non-negative measures inside Ω a geodesic space without any convexity assumption on the domain. Moreover we will show that the gradient flow of the entropy functional $\int_\Omega [\rho \log(\rho) - \rho]\, dx$ with respect to this distance coincides with the heat equation, subject to the Dirichlet boundary condition equal to 1.

Résumé

Dans ce papier, nous introduisons une nouvelle distance sur l'espace des mesures positive dans un domaine Ω. Cette distance satisfait plusieurs propriétés intéressantes : par exemple, elle fait de l'espace des mesures positives dans Ω un espace géodésique sans aucune hypothèse de convexité sur le domaine. De plus, on montre que le flot gradient de la fonctionnelle d'entropie $\int_\Omega [\rho \log(\rho) - \rho]\, dx$ par rapport à cette distance donne lieu à l'équation de la chaleur, avec condition de Dirichlet égale à 1 sur le bord.

Keywords: transportation distances, gradient flows, heat equation, Dirichlet boundary conditions.

1 Introduction

Nowadays, it is well-know that transportation distances between probability measures can be successfully used to study evolutionary equations. More precisely, one of the most surprisingly achievement of [8, 11, 12] has been that many evolution equations of the form

$$\frac{d}{dt}\rho(t) = \operatorname{div}\Big(\nabla\rho(t) - \rho(t)\nabla V - \rho(t)(\nabla W * \rho(t))\Big)$$

[*] *Previous address:* Ecole Polytechnique, CMLS. 91128 Palaiseau cedex, France. *Current address:* The University of Texas at Austin. 1 University Station, C1200. Austin TX 78712, USA. e-mail: figalli@math.utexas.edu

[†] Université de Bordeaux. 351 Cours de la Libération, 33405 Talence cedex, France. Partially supported by KAM Faible, ANR-07-BLAN-0361. e-mail: nicolagigli@googlemail.com

can be seen as gradient flows of some entropy functionals on the space of probability measures with respect to the Wasserstein distance

$$W_2(\mu,\nu) = \inf\left\{ \sqrt{\int |x-y|^2 \, d\boldsymbol{\gamma}(x,y)} \; : \; \pi^1_\# \boldsymbol{\gamma} = \mu, \; \pi^2_\# \boldsymbol{\gamma} = \nu \right\}.$$

Besides the fact that this interpretation allows to prove entropy estimates and functional inequalities (see [13, 14] for more details on this area, which is still very active and in continuous evolution), this point of view provides a powerful variational method to prove existence of solutions to the above equations: given a time step $\tau > 0$, construct an approximate solution by iteratively minimizing

$$\rho \quad \mapsto \quad \frac{W_2(\rho,\rho_0)}{2\tau} + \int \Big[\rho \log(\rho) + \rho V + \frac{1}{2}\rho (W * \rho) \Big] \, dx.$$

We refer to [2] for a general description of this approach.

Let us observe that the choice of the distance on the space of probability measures plays a key role, and by changing it one can construct solutions to more general classes of evolution equations, see for instance [1, 5, 7]. However, all the distances considered up to now need the two measures to have the same mass (which up to a scaling can always be assumed equal to 1). In particular, since the mass remains constant along the evolution, if one restricts to measures concentrated on a bounded domain, then the approach described above will always produce solutions to parabolic equations with Neumann boundary conditions.

Motivated by the intent to find an analogous approach to construct solutions of evolution equations subject to Dirichlet boundary condition, in this paper we introduce a new transportation distance Wb_2 between measures. As we will see, the main features of the distance Wb_2 are:

- It metrizes the weak convergence of positive Borel measures in Ω belonging to the space
$$\mathcal{M}_2(\Omega) := \left\{ \mu \; : \; \int d^2(x,\partial\Omega) \, d\mu(x) < \infty \right\}, \tag{1}$$
see Proposition 2.7. Observe that $\mathcal{M}_2(\Omega)$ contains all positive finite measures on Ω and that the claim we are making is perfectly analogous to what happens for the common Wasserstein distances, but without any mass constraint.

- The resulting metric space $(\mathcal{M}_2(\Omega), Wb_2)$ is always geodesic, see Proposition 2.9. This is a particularly interesting property compared to what happens in the classical Wasserstein space: indeed the space $(\mathscr{P}(\Omega), W_2)$ is geodesic if and only if Ω is convex. In our case, the convexity of the open set is not required. (Actually, not even connectedness is needed!)

- The natural approach via minimizing movements to the study of the gradient flow of the entropy leads to weak solution of the heat equation with Dirichlet boundary condition, see Theorem 3.5. Interesting enough, with this approach the regularity of the boundary of Ω does not play any role.

As a drawback, the entropy functional do not have the same nice properties it has in the classical Wasserstein space. In particular:

- It is *not* geodesically convex. Still, it has some sort of convexity properties along geodesics, see Remark 3.4.

- Due to the lack of geodesic convexity, we were not able to prove any kind of contractivity result for the flow.
- Actually, we are not even able to prove uniqueness of the limit of the minimizing movements scheme. (Of course one knows by standard PDE techniques that weak solutions of the heat equation with Dirichlet boundary conditions are unique, therefore a posteriori it is clear that the limit has to be unique - what we are saying here is that we do not know whether such uniqueness may be deduced a priori via techniques similar, e.g., to those appeared in [2].)

The distance Wb_2 is defined in the following way (the '*b*' stands to recall that we have some room to play with the *b*oundary of Ω). Let $\Omega \subset \mathbb{R}^d$ be a bounded open set, and let $\mathcal{M}_2(\Omega)$ be defined by (1). We define the distance Wb_2 on $\mathcal{M}_2(\Omega)$ as a result of the following problem:

Problem 1.1 (A variant of the transportation problem) *Let $\mu, \nu \in \mathcal{M}_2(\Omega)$. The set of admissible couplings* $\textsc{Adm}(\mu,\nu)$ *is defined as the set of positive measures $\boldsymbol{\gamma}$ on $\overline{\Omega}\times\overline{\Omega}$ satisfying*

$$\pi^1_{\#}\boldsymbol{\gamma}_{|\Omega} = \mu, \qquad \pi^2_{\#}\boldsymbol{\gamma}_{|\Omega} = \nu. \tag{2}$$

For any non-negative measure $\boldsymbol{\gamma}$ on $\overline{\Omega}\times\overline{\Omega}$, we define its cost $C(\boldsymbol{\gamma})$ *as*

$$C(\boldsymbol{\gamma}) := \int_{\overline{\Omega}\times\overline{\Omega}} |x-y|^2 d\boldsymbol{\gamma}(x,y).$$

The distance $Wb_2(\mu,\nu)$ is then defined as:

$$Wb_2^2(\mu,\nu) := \inf_{\boldsymbol{\gamma}\in\textsc{Adm}(\mu,\nu)} C(\boldsymbol{\gamma}).$$

The difference between Wb_2 and W_2 relies on the fact that an admissible coupling is a measure on the *closure* of $\Omega\times\Omega$, rather than just on $\Omega\times\Omega$, and that the marginals are required to coincide with the given measures only inside Ω. This means that we can use $\partial\Omega$ as an infinite reserve: we can 'take' as mass as we wish from the boundary, or 'give' it back some of the mass, provided we pay the transportation cost. This is why this distance is well defined for measures which do not have the same mass.

Although this approach could be applied for more general costs than just $|x-y|^2$ and for a wider class of entropy functionals, we preferred to provide a complete result only in the particular case of the heat equation, in order to avoid technicalities and generalizations which would just obscure the main ideas. We refer to Section 4 for some possible generalizations, a comparison between our and the classical L^2-approach, and some open problems.

Acknowledgements: We gratefully thank the referee for his valuable comments, which helped us having a clearer understanding of the properties of Wb_2.

2 General properties of the distance Wb_2

The aim of this section is to describe the main properties of the distance Wb_2.

For any positive Borel measure μ in Ω, define $m_2(\mu)$ as

$$m_2(\mu) := \int d^2(x,\partial\Omega)d\mu(x),$$

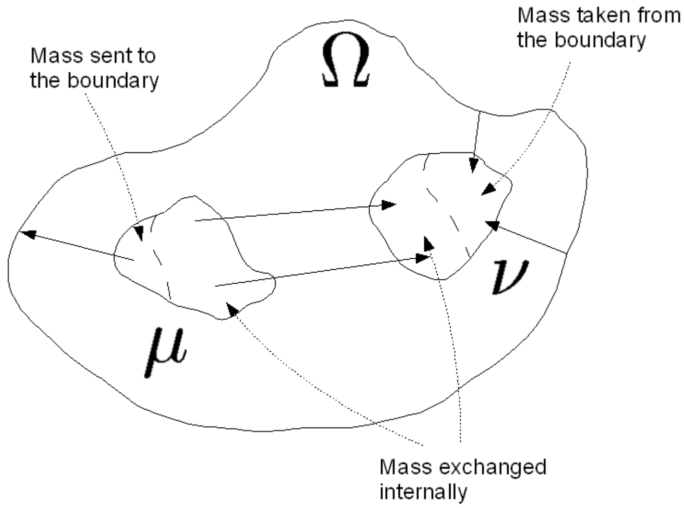

Figure 1: Example of admissible transport plan

so that $\mathcal{M}_2(\Omega)$ is precisely the set of measures μ such that $m_2(\mu) < \infty$. Observe that if $A \subset \Omega$ is any set which is at a positive distance r from $\partial\Omega$ and $\mu \in \mathcal{M}_2(\Omega)$, then the inequality

$$\infty > m_2(\mu) \geq \int_A r^2 d\mu = r^2\mu(A), \tag{3}$$

shows that $\mu(A) < \infty$.

Let $\boldsymbol{\gamma}$ be a non-negative measure on $\overline{\Omega} \times \overline{\Omega}$. We will write $\boldsymbol{\gamma}_A^B$ for the restriction of $\boldsymbol{\gamma}$ to the rectangle $A \times B \subset \overline{\Omega} \times \overline{\Omega}$. Observe that there is a natural splitting of $\boldsymbol{\gamma}$ into 4 parts:

$$\boldsymbol{\gamma} = \boldsymbol{\gamma}_\Omega^\Omega + \boldsymbol{\gamma}_\Omega^{\partial\Omega} + \boldsymbol{\gamma}_{\partial\Omega}^\Omega + \boldsymbol{\gamma}_{\partial\Omega}^{\partial\Omega}.$$

We now remark that, if $\boldsymbol{\gamma} \in \textsc{Adm}(\mu, \nu)$, then

$$\boldsymbol{\gamma} - \boldsymbol{\gamma}_{\partial\Omega}^{\partial\Omega} \in \textsc{Adm}(\mu, \nu) \quad \text{and} \quad C(\boldsymbol{\gamma} - \boldsymbol{\gamma}_{\partial\Omega}^{\partial\Omega}) \leq C(\boldsymbol{\gamma}).$$

Hence, when looking for optimal plans, it is not restrictive to assume that

$$\boldsymbol{\gamma}_{\partial\Omega}^{\partial\Omega} = 0. \tag{4}$$

Observe that from the bound

$$\boldsymbol{\gamma}(A \times \overline{\Omega} \cup \overline{\Omega} \cup A) \leq \boldsymbol{\gamma}(A \times \overline{\Omega}) + \boldsymbol{\gamma}(\overline{\Omega} \times A) = \mu(A) + \nu(A) < \infty,$$

valid for any Borel set $A \subset \Omega$ with positive distance from $\partial\Omega$ and any $\boldsymbol{\gamma} \in \textsc{Adm}(\mu, \nu)$ satisfying (4), it easily follows the weak compactness of the set of admissible plans satisfying (4) (in duality with functions in $C_c(\overline{\Omega} \times \overline{\Omega} \setminus \partial\Omega \times \partial\Omega)$). Thus from the weak lower semicontinuity of

$$\boldsymbol{\gamma} \mapsto C(\boldsymbol{\gamma}),$$

we get the existence of optimal plans

We will denote the set of optimal plans by $\textsc{Opt}(\mu, \nu)$, and we will always assume that an optimal plan satisfies (4).

We now prove that Wb_2 is a distance on $\mathcal{M}_2(\Omega)$. For the triangle inequality we need the following variant of the classical gluing lemma (see [2, Lemma 5.3.2]):

Lemma 2.1 (A variant of the gluing lemma) *Fix $\mu_1, \mu_2, \mu_3 \in \mathcal{M}_2(\Omega)$, and let $\boldsymbol{\gamma}^{12} \in \mathrm{Adm}(\mu_1, \mu_2)$, $\boldsymbol{\gamma}^{23} \in \mathrm{Adm}(\mu_2, \mu_3)$ such that $(\boldsymbol{\gamma}^{12})_{\partial\Omega}^{\partial\Omega} = (\boldsymbol{\gamma}^{23})_{\partial\Omega}^{\partial\Omega} = 0$. Then there exists a positive Borel measure $\boldsymbol{\gamma}^{123}$ on $\overline{\Omega} \times \overline{\Omega} \times \overline{\Omega}$ such that*

$$\pi^{12}_{\#}\boldsymbol{\gamma}^{123} = \boldsymbol{\gamma}^{12} + \boldsymbol{\sigma}^{12},$$
$$\pi^{23}_{\#}\boldsymbol{\gamma}^{123} = \boldsymbol{\gamma}^{23} + \boldsymbol{\sigma}^{23},$$

where $\boldsymbol{\sigma}^{12}$ and $\boldsymbol{\sigma}^{23}$ are both concentrated on the diagonal of $\partial\Omega \times \partial\Omega$, i.e. on the set of pairs of points $\{(x, x) \, : \, x \in \partial\Omega\}$.

Let us point out that, in contrast with the classical result, in our case the second marginal of $\boldsymbol{\gamma}^{12}$ on $\overline{\Omega}$ does not necessarily coincides with the first marginal of $\boldsymbol{\gamma}^{23}$, and so the two measures cannot be 'glued' together in a trivial way.

Proof. In order to clarify the structure of the proof, it is convenient to see μ_1, μ_2, μ_3 as measures on $\mathcal{M}_2(\Omega_1), \mathcal{M}_2(\Omega_2), \mathcal{M}_2(\Omega_3)$ respectively, where $\Omega_1, \Omega_2, \Omega_3$ are three distinct copies of Ω. In this way we have $\boldsymbol{\gamma}^{12} \in \mathcal{M}_+(\overline{\Omega}_1 \times \overline{\Omega}_2)$, $\boldsymbol{\gamma}^{23} \in \mathcal{M}_+(\overline{\Omega}_2 \times \overline{\Omega}_3)$, and $\boldsymbol{\gamma}^{123} \in \mathcal{M}_+(\overline{\Omega}_1 \times \overline{\Omega}_2 \times \overline{\Omega}_3)$. However, since in fact $\Omega_1 = \Omega_2 = \Omega_3$, sometimes we will identify Ω_2 with Ω, Ω_1, or Ω_3. Furthermore, we will use π^2 to denote both the canonical projection from $\overline{\Omega}_1 \times \overline{\Omega}_2$ onto $\overline{\Omega}_2$, and the one from $\overline{\Omega}_2 \times \overline{\Omega}_3$ onto $\overline{\Omega}_2$.

From the hypothesis we know that

$$\pi^2_{\#}(\boldsymbol{\gamma}^{12})^{\Omega}_{\overline{\Omega}} = \mu^2 = \pi^2_{\#}(\boldsymbol{\gamma}^{23})^{\overline{\Omega}}_{\Omega},$$

also, since μ^2 is locally finite an application of (a simple variant of) the classical gluing lemma guarantees the existence of a plan $\tilde{\boldsymbol{\gamma}}^{123} \in \mathcal{M}_+(\overline{\Omega}_1 \times \Omega_2 \times \overline{\Omega}_3)$ such that

$$\pi^{12}_{\#}\tilde{\boldsymbol{\gamma}}^{123} = (\boldsymbol{\gamma}^{12})^{\Omega}_{\overline{\Omega}},$$
$$\pi^{23}_{\#}\tilde{\boldsymbol{\gamma}}^{123} = (\boldsymbol{\gamma}^{23})^{\overline{\Omega}}_{\Omega}.$$

Then define

$$\boldsymbol{\sigma}^{12} := (\pi^2, \pi^2, \pi^3)_{\#}\Big((\boldsymbol{\gamma}^{23})^{\Omega}_{\partial\Omega}\Big) \in \mathcal{M}(\partial\Omega_2 \times \partial\Omega_2 \times \overline{\Omega}_3) = \mathcal{M}(\partial\Omega_1 \times \partial\Omega_2 \times \overline{\Omega}_3),$$
$$\boldsymbol{\sigma}^{23} := (\pi^1, \pi^2, \pi^2)_{\#}\Big((\boldsymbol{\gamma}^{12})^{\partial\Omega}_{\Omega}\Big) \in \mathcal{M}(\overline{\Omega}_1 \times \partial\Omega_2 \times \partial\Omega_2) = \mathcal{M}(\overline{\Omega}_1 \times \partial\Omega_2 \times \partial\Omega_3),$$

and finally define

$$\boldsymbol{\gamma}^{123} := \tilde{\boldsymbol{\gamma}}^{123} + \boldsymbol{\sigma}^{12} + \boldsymbol{\sigma}^{23}.$$

We prove that $\boldsymbol{\gamma}^{123}$ satisfies the thesis. Observe that

$$\begin{aligned}\pi^{12}_{\#}\boldsymbol{\gamma}^{123} &= \pi^{12}_{\#}\boldsymbol{\gamma}^{123} + \pi^{12}_{\#}\boldsymbol{\sigma}^{12} + \pi^{12}_{\#}\boldsymbol{\sigma}^{23} = (\boldsymbol{\gamma}^{12})^{\Omega}_{\overline{\Omega}} + (\pi^2, \pi^2)_{\#}\Big((\boldsymbol{\gamma}^{23})^{\Omega}_{\partial\Omega}\Big) + (\pi^1, \pi^2)_{\#}\Big((\boldsymbol{\gamma}^{12})^{\partial\Omega}_{\Omega}\Big) \\ &= (\boldsymbol{\gamma}^{12})^{\Omega}_{\overline{\Omega}} + (\pi^2, \pi^2)_{\#}\Big((\boldsymbol{\gamma}^{23})^{\Omega}_{\partial\Omega}\Big) + (\boldsymbol{\gamma}^{12})^{\partial\Omega}_{\Omega} = \boldsymbol{\gamma}^{12} + (\pi^2, \pi^2)_{\#}\Big((\boldsymbol{\gamma}^{23})^{\Omega}_{\partial\Omega}\Big),\end{aligned}$$

and that $(\pi^2, \pi^2)_{\#}((\boldsymbol{\gamma}^{23})^{\Omega}_{\partial\Omega})$ is concentrated on the diagonal of $\partial\Omega \times \partial\Omega$. Similar for $\pi^{23}_{\#}\boldsymbol{\gamma}^{123}$. □

Theorem 2.2 (Wb_2 is a distance on $\mathcal{M}_2(\Omega)$) *The function Wb_2 is a distance on the set $\mathcal{M}_2(\Omega)$ which is lower semicontinuous with respect to weak convergence in duality with functions in $C_c(\Omega)$.*

Proof. The facts that $Wb_2(\mu,\nu) = 0$ if and only if $\mu = \nu$ and the symmetry are obvious. For the triangle inequality we need to use the version of gluing lemma we just proved. Fix $\mu_1, \mu_2, \mu_3 \in \mathcal{M}_2(\Omega)$ and let $\boldsymbol{\gamma}^{12}, \boldsymbol{\gamma}^{23}$ be two optimal plans from μ_1 to μ_2 and from μ_2 to μ_3 respectively. Use lemma 2.1 to find a 3-plan $\boldsymbol{\gamma}^{123}$ such that $\pi^{1,2}_\#\boldsymbol{\gamma}^{123} = \boldsymbol{\gamma}^{12} + \boldsymbol{\sigma}^{12}$ and $\pi^{2,3}_\#\boldsymbol{\gamma}^{123} = \boldsymbol{\gamma}^{23} + \boldsymbol{\sigma}^{23}$, with $\boldsymbol{\sigma}^{12}$ and $\boldsymbol{\sigma}^{12}$ concentrated on the diagonals of $\partial\Omega \times \partial\Omega$. Then we have $(\pi^1_\#\boldsymbol{\gamma}^{123})_{|\Omega} = (\pi^1_\#\boldsymbol{\gamma}^{12} + \boldsymbol{\sigma}^{12})_{|\Omega} = \mu_1$. Similarly, we have $(\pi^3_\#\boldsymbol{\gamma}^{123})_{|\Omega} = \mu_3$, therefore $\pi^{1,2}\boldsymbol{\gamma}^{123} \in \textsc{Adm}(\mu_1, \mu_3)$ and it holds

$$
\begin{aligned}
Wb_2(\mu_1,\mu_3) &\leq \sqrt{\int |x_1 - x_3|^2 d\boldsymbol{\gamma}^{123}} \\
&\leq \sqrt{\int |x_1 - x_2|^2 d\boldsymbol{\gamma}^{123}} + \sqrt{\int |x_2 - x_3|^2 d\boldsymbol{\gamma}^{123}} \\
&= \sqrt{\int |x_1 - x_2|^2 d(\boldsymbol{\gamma}^{12} + \boldsymbol{\sigma}^{12})} + \sqrt{\int |x_2 - x_3|^2 d(\boldsymbol{\gamma}^{23} + \boldsymbol{\sigma}^{23})} \\
&= \sqrt{\int |x_1 - x_2|^2 d\boldsymbol{\gamma}^{12}} + \sqrt{\int |x_2 - x_3|^2 d\boldsymbol{\gamma}^{23}} \\
&= Wb_2(\mu_1,\mu_2) + Wb_2(\mu_2,\mu_3),
\end{aligned}
$$

where in the fourth step we used the fact that $\boldsymbol{\sigma}^{12}$ and $\boldsymbol{\sigma}^{23}$ are concentrated on a diagonal.

For the lower semicontinuity, let (μ_n), (ν_n) be two sequences weakly converging to μ, ν respectively, and for any $n \in \mathbb{N}$ choose $\boldsymbol{\gamma}_n \in \textsc{Opt}(\mu_n, \nu_n)$. It is immediate to check that the sequence $(\boldsymbol{\gamma}_n)$ is relatively compact in duality with functions in $C_c(\overline{\Omega} \times \overline{\Omega} \setminus \partial\Omega \times \partial\Omega)$, so that up to passing to a subsequence, not relabeled, we may assume that $(\boldsymbol{\gamma}_n)$ weakly converges to some $\boldsymbol{\gamma}$ in duality with $C_c(\overline{\Omega} \times \overline{\Omega} \setminus \partial\Omega \times \partial\Omega)$. Since obviously $\pi^1_\#\boldsymbol{\gamma}_{|\Omega} = \mu$ and $\pi^2_\#\boldsymbol{\gamma}_{|\Omega} = \nu$ we have

$$
Wb_2^2(\mu,\nu) \leq \int |x-y|^2 d\boldsymbol{\gamma} \leq \varliminf_{n\to\infty} \int |x-y|^2 d\boldsymbol{\gamma}_n(x,y) = \varliminf_{n\to\infty} Wb_2^2(\mu_n,\nu_n).
$$

□

From now on, $P : \Omega \to \partial\Omega$ will be a measurable map such that

$$
|x - P(x)| = d(x, \partial\Omega) \qquad \forall\, x \in \Omega.
$$

It is well-known that such a map is uniquely defined on $\mathcal{L}^d$-a.e. $x \in \Omega$. (Indeed, $P(x)$ is uniquely defined whenever the Lipschitz function $d(\cdot, \partial\Omega)$ is differentiable, and is given by $P(x) = x - \nabla d(x, \partial\Omega)^2/2$.) Here we are just defining it on the whole Ω via a measurable selection argument (we omit the details).

We will use the notation $\mathrm{Id} : \overline{\Omega} \to \overline{\Omega}$ to denote the identity map on $\overline{\Omega}$.

To better understand the properties of optimal plans, let us set $c(x,y) := |x-y|^2$ and

$$
\tilde{c}(x,y) := \min\left\{|x-y|^2, d^2(x,\partial\Omega) + d^2(y,\partial\Omega)\right\}.
$$

Also, define the set $\mathcal{A} \subset \overline{\Omega} \times \overline{\Omega}$ by

$$
\mathcal{A} := \left\{(x,y) \in \overline{\Omega} \times \overline{\Omega} \ : \ |x-y|^2 \leq d^2(x,\partial\Omega) + d^2(y,\partial\Omega)\right\}. \tag{5}
$$

Recall that a function φ on $\overline{\Omega}$ is said c-concave provided

$$\varphi(x) = \inf_{y \in \overline{\Omega}} c(x, y) - \psi(y),$$

for some $\psi : \overline{\Omega} \to \mathbb{R}$. The c-transform of φ is the function φ^c defined by

$$\varphi^c(y) = \inf_{x \in \overline{\Omega}} c(x, y) - \varphi(y),$$

and the c-superdifferential $\partial^c_+ \varphi$ of the c-concave function φ is the set

$$\partial^c_+ \varphi := \Big\{(x, y) \in \overline{\Omega} \times \overline{\Omega} \ : \ \varphi(x) = c(x, y) - \varphi^c(y)\Big\}.$$

Analogously, one can speak about $\tilde{c}$-concavity, $\tilde{c}$-transform, and $\tilde{c}$-superdifferential. Let us remark that since $c, \tilde{c}$ are Lipschitz on $\overline{\Omega} \times \overline{\Omega}$, c-concave and $\tilde{c}$-concave functions are Lipschitz too.

Proposition 2.3 (Characterization of optimal plans) *Let $\boldsymbol{\gamma}$ be a Borel measure on $\overline{\Omega} \times \overline{\Omega} \setminus \partial\Omega \times \partial\Omega$ satisfying $\int \Big[d^2(x, \partial\Omega) + d^2(y, \partial\Omega)\Big]\, d\boldsymbol{\gamma}(x, y) < \infty$. Then the following three things are equivalent:*

- *(i)* *$\boldsymbol{\gamma}$ is optimal for the couple $\pi^1_\# \boldsymbol{\gamma}_{|\Omega}, \pi^2_\# \boldsymbol{\gamma}_{|\Omega}$ for Problem 1.1,.*
- *(ii)* *$\boldsymbol{\gamma}$ is concentrated on $\mathcal{A}$ and the set* $\mathrm{supp}(\boldsymbol{\gamma}) \cup \partial\Omega \times \partial\Omega$ *is $\tilde{c}$-cyclically monotone.*
- *(iii)* *there exists a c-concave function φ such that φ and φ^c are both identically* 0 *on $\partial\Omega$, and* $\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^c_+ \varphi$.

Also, for each optimal plan $\boldsymbol{\gamma}$ it holds

$$|x - y| = d(x, \partial\Omega), \qquad \boldsymbol{\gamma}^{\partial\Omega}_{\Omega}\text{-a.e. } (x, y). \tag{6}$$

Similarly for $\boldsymbol{\gamma}^{\Omega}_{\partial\Omega}$.

Moreover, if $(\boldsymbol{\gamma}^n)$ is a sequence of optimal plans for Problem 1.1 (each one with respect to to its own marginals) which weakly converges to some plan $\boldsymbol{\gamma}$ in duality with functions in $C_c(\overline{\Omega} \times \overline{\Omega} \setminus \partial\Omega \times \partial\Omega)$, then $\boldsymbol{\gamma}$ is optimal as well.

Finally, given $\mu, \nu \in \mathcal{M}_2(\Omega)$ there exists a c-concave function φ such that φ and φ^c are both identically 0 *on $\partial\Omega$, and every optimal plan $\boldsymbol{\gamma}$ between μ and ν is concentrated on $\partial^c_+ \varphi$.*

Proof. Let us first assume that $\boldsymbol{\gamma}$ has finite mass. Define $\overline{\mu} := \pi^1_\# \boldsymbol{\gamma}$, $\overline{\nu} := \pi^2_\# \boldsymbol{\gamma}$, and let μ, ν be the restriction of $\overline{\mu}, \overline{\nu}$ to Ω respectively.

We start proving that $(i) \Rightarrow (ii)$. We show first that $\boldsymbol{\gamma}$ is concentrated on $\mathcal{A}$. Define the plan $\tilde{\boldsymbol{\gamma}}$ by

$$\tilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma}_{|\mathcal{A}} + (\pi^1, P \circ \pi^1)_\# \Big(\boldsymbol{\gamma}_{|\overline{\Omega} \times \overline{\Omega} \setminus \mathcal{A}}\Big) + (P \circ \pi^2, \pi^2)_\# \Big(\boldsymbol{\gamma}_{|\overline{\Omega} \times \overline{\Omega} \setminus \mathcal{A}}\Big), \tag{7}$$

and observe that $\tilde{\boldsymbol{\gamma}} \in \textsc{Adm}(\mu, \nu)$ and

$$\begin{aligned}\int |x - y|^2\, d\tilde{\boldsymbol{\gamma}}(x, y) &= \int_{\mathcal{A}} |x - y|^2\, d\boldsymbol{\gamma} + \int_{\overline{\Omega} \times \overline{\Omega} \setminus \mathcal{A}} \Big[d^2(x, \partial\Omega) + d^2(y, \partial\Omega)\Big]\, d\boldsymbol{\gamma}(x, y) \\ &\leq \int |x - y|^2\, d\boldsymbol{\gamma}(x, y),\end{aligned}$$

with *strict* inequality if $\gamma(\overline{\Omega} \times \overline{\Omega} \setminus \mathcal{A}) > 0$. Thus from the optimality of γ we deduce that γ is concentrated on $\mathcal{A}$. This implies that

$$\int \tilde{c}(x,y)\, d\gamma(x,y) = \int |x-y|^2\, d\gamma(x,y). \tag{8}$$

Now we show that γ is an optimal transport plan (in the classical sense) from $\overline{\mu}$ to $\overline{\nu}$ for the cost $\tilde{c}$. Suppose by contradiction that it is not optimal. Then there exists some plan η such that $\pi^1_{\#}\eta = \overline{\mu}$, $\pi^2_{\#}\eta = \overline{\nu}$ and

$$\int \tilde{c}(x,y)\, d\eta < \int \tilde{c}(x,y)\, d\gamma. \tag{9}$$

Let $\tilde{\eta}$ be the plan constructed via formula (7) replacing γ by η. As before, from $\eta \in \textsc{Adm}(\mu,\nu)$ we derive $\tilde{\eta} \in \textsc{Adm}(\mu,\nu)$, and

$$\int |x-y|^2 d\tilde{\eta}(x,y) = \int \tilde{c}(x,y) d\eta(x,y).$$

Hence from (8) and (9) we contradict the optimality of γ for Problem 1.1.

This shows that any optimal plan γ is an optimal transport plan (in the classical sense) from $\overline{\mu}$ to $\overline{\nu}$ for the cost $\tilde{c}$. Hence, applying the standard transport theory from optimal transport, we deduce that the support of any optimal plan for Problem 1.1 is $\tilde{c}$-cyclically monotone.

Now, take $\overline{x} \in \partial\Omega$ and observe that the plan $\overline{\gamma} := \gamma + \delta_{\overline{x},\overline{x}}$ is still optimal for Problem 1.1. Hence by the above argument the set $\mathrm{supp}(\gamma) \cup \{(\overline{x},\overline{x})\}$ is $\tilde{c}$-cyclically monotone. From the validity of

$$\begin{aligned} \tilde{c}(x,y) = 0, \qquad & \forall x,y \in \partial\Omega, \\ \tilde{c}(x,z) + \tilde{c}(y,z) \geq \tilde{c}(x,y) \qquad & \forall x,y \in \overline{\Omega},\ z \in \partial\Omega, \end{aligned}$$

and it is easy to verify the $\tilde{c}$-cyclically monotonicity of $\mathrm{supp}(\gamma) \cup \{(\overline{x},\overline{x})\}$ implies that the set

$$\mathrm{supp}(\gamma) \cup \partial\Omega \times \partial\Omega$$

is $\tilde{c}$-cyclically monotone as well, as desired.

Now we prove that $(ii) \Rightarrow (iii)$. From the standard theory of transport problems (see e.g. [14, Theorem 5.10]) there exists a $\tilde{c}$-concave function such that $\mathrm{supp}(\gamma) \cup \partial\Omega \times \partial\Omega \subset \partial^{\tilde{c}}_{+}\varphi$. We claim that φ and $\varphi^{\tilde{c}}$ are both constant on $\partial\Omega$. Indeed, since $(x,y) \in \partial^{\tilde{c}}_{+}\varphi$ for any $(x,y) \in \partial\Omega \times \partial\Omega$ we have

$$\varphi(x) + \varphi^{\tilde{c}}(y) = \tilde{c}(x,y) = 0, \qquad \forall\, x,y \in \partial\Omega,$$

which gives the claim. In particular, up to adding a constant, we can assume that φ is identically 0 on $\partial\Omega$, which implies in particular that $\varphi^{\tilde{c}}$ is identically 0 on $\partial\Omega$ too.

The fact that φ is c-concave follows immediately from the fact that for any $y \in \overline{\Omega}$ the function

$$x \quad \mapsto \quad c(x,y) = \min\left\{|x-y|^2, \inf_{z\in\partial\Omega} |x-z|^2 + d^2(y,\partial\Omega)\right\},$$

is c-concave. It remains to prove that γ is concentrated on $\partial^c_{+}\varphi$ and that $\varphi^c = 0$ on $\partial\Omega$. For the first part, we observe that

$$\partial^{\tilde{c}}_{+}\varphi \cap \mathcal{A} \subset \partial^c_{+}\varphi. \tag{10}$$

Indeed, assume that $(x_0, y_0) \in \partial^{\tilde{c}}_+\varphi \cap \mathcal{A}$. Then

$$\begin{aligned}\varphi(x_0) &= \tilde{c}(x_0, y_0) - \varphi^{\tilde{c}}(y_0),\\ \varphi(x) &\le \tilde{c}(x, y_0) - \varphi^{\tilde{c}}(y_0), \qquad \forall x \in \overline{\Omega}.\end{aligned}$$

Moreover, since $(x_0, y_0) \in \mathcal{A}$ we have $\tilde{c}(x_0, y_0) = c(x_0, y_0)$ while in general $\tilde{c}(x, y_0) \le c(x, y_0)$. Hence

$$\begin{aligned}\varphi(x_0) &= c(x_0, y_0) - \varphi^{\tilde{c}}(y_0),\\ \varphi(x) &\le c(x, y_0) - \varphi^{\tilde{c}}(y_0), \qquad \forall x \in \overline{\Omega},\end{aligned}$$

which easily implies that $(x_0, y_0) \in \partial^c_+\varphi$.

For the second part, observe that the diagonal $\{(x, x) : x \in \partial\Omega\}$ is included both in $\partial^{\tilde{c}}_+\varphi$ and in $\mathcal{A}$, thus from (10) it is included in $\partial^c_+\varphi$. This means that

$$\varphi^c(x) = \varphi(x) + \varphi^c(x) = c(x, x) = 0, \qquad \forall x \in \partial\Omega,$$

which gives $\varphi^c(x) \equiv 0$ in $\partial\Omega$ as desired.

We finally show $(iii) \Rightarrow (i)$. Let $\tilde{\boldsymbol{\gamma}}$ be any plan in $\mathrm{Adm}(\mu, \nu)$. Since $\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^c\varphi(x)$, we have $\varphi(x) + \varphi^c(y) = c(x, y) = |x - y|^2$ on the support of $\boldsymbol{\gamma}$, while for general x, y it holds $\varphi(x)+\varphi^c(y) \le c(x, y) = |x-y|^2$. Also, the functions φ, φ^c are Lipschitz (so in particularly bounded), and thus integrable with respect to any measure with finite mass. Furthermore, since φ is identically 0 on $\partial\Omega$ and $\pi^1_\#\boldsymbol{\gamma}_{|\Omega} = \pi^1_\#\tilde{\boldsymbol{\gamma}}_{|\Omega}$, we have $\int \varphi d\pi^1_\#\boldsymbol{\gamma} = \int \varphi d\pi^1_\#\tilde{\boldsymbol{\gamma}}$. The analogous result holds for for φ^c. Thanks to these considerations we obtain

$$\begin{aligned}\int |x - y|^2\, d\boldsymbol{\gamma}(x, y) &= \int \left[\varphi(x) + \varphi^c(y)\right]\, d\boldsymbol{\gamma}(x, y)\\ &= \int \varphi(x)\, d\pi^1_\#\boldsymbol{\gamma}(x) + \int \varphi^c(y)\, d\pi^2_\#\boldsymbol{\gamma}(y)\\ &= \int \varphi(x)\, d\pi^1_\#\tilde{\boldsymbol{\gamma}}(x) + \int \varphi^c(y)\, d\pi^1_\#\tilde{\boldsymbol{\gamma}}(x)\\ &= \int [\varphi(x) + \varphi^c(y)]\, d\tilde{\boldsymbol{\gamma}}(x, y)\\ &\le \int |x - y|^2\, d\tilde{\boldsymbol{\gamma}},\end{aligned}$$

which concludes the proof.

Now, let us consider the case when γ has infinite mass.

The proof of $(i) \Rightarrow (ii)$ works as in the case of finite mass. Indeed, the only argument coming from the classical transport theory that we used is the implication 'support not $\tilde{c}$-cyclically monotone implies plan not optimal for the cost $\tilde{c}$', and it is immediate to check that the classical argument of finding a better competitor from the lack of $\tilde{c}$-cyclical monotonicity of the support works also for infinite mass.

The implication $(ii) \Rightarrow (iii)$ follows as above, as the statements (ii) and (iii) concern only properties of the support of $\boldsymbol{\gamma}$.

To prove $(iii) \Rightarrow (i)$, the only difficulty comes from the fact that a priori φ and φ^c may be not integrable. However, it is easy to see that the c-concavity of φ combined with the fact that φ and φ^c both vanish on $\partial\Omega$ implies

$$\varphi(x) \le d^2(x, \partial\Omega), \quad \varphi^c(x) \le d^2(x, \partial\Omega), \qquad \forall\, x \in \overline{\Omega}.$$

Hence both φ and φ^c are semi-integrable, and this allow us to conclude as above. (See for instance Step 4 in the proof of [2, Theorem 6.1.4].)

To prove (6), let

$$A := \Big\{(x,y) \in \Omega \times \partial\Omega \ : \ |x-y| > d(x,\partial\Omega) = |x - P(x)|\Big\},$$

and assume by contradiction that $\boldsymbol{\gamma}_\Omega^{\partial\Omega}(A) > 0$. Then we define

$$\tilde{\boldsymbol{\gamma}}_\Omega^{\partial\Omega} := (\mathrm{Id}, P)_\# \pi^1_\# \boldsymbol{\gamma}_\Omega^{\partial\Omega},$$

and set

$$\tilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma}_\Omega^\Omega + \tilde{\boldsymbol{\gamma}}_\Omega^{\partial\Omega} + \boldsymbol{\gamma}_{\partial\Omega}^\Omega.$$

Since $\pi^1_\# \tilde{\boldsymbol{\gamma}}_\Omega^{\partial\Omega} = \pi^1_\# \boldsymbol{\gamma}_\Omega^{\partial\Omega}$ we have $\pi^1_\# \tilde{\boldsymbol{\gamma}} = \pi^1_\# \boldsymbol{\gamma}$. Moreover $\pi^2_\# \tilde{\boldsymbol{\gamma}}_{|\Omega} = \pi^2_\# \boldsymbol{\gamma}_{|\Omega}$ by construction, so that $\tilde{\boldsymbol{\gamma}} \in \mathrm{Adm}(\mu_0, \mu_1)$. Since

$$\int_{\Omega\times\partial\Omega} |x-y|^2 \, d\tilde{\boldsymbol{\gamma}}(x,y) < \int_{\Omega\times\partial\Omega} |x-y|^2 \, d\boldsymbol{\gamma}(x,y),$$
$$\int_{\overline{\Omega}\times\overline{\Omega}\setminus\Omega\times\partial\Omega} |x-y|^2 \, d\tilde{\boldsymbol{\gamma}}(x,y) = \int_{\overline{\Omega}\times\overline{\Omega}\setminus\Omega\times\partial\Omega} |x-y|^2 \, d\boldsymbol{\gamma}(x,y),$$

we have $C(\tilde{\boldsymbol{\gamma}}) < C(\boldsymbol{\gamma})$, which gives the desired contradiction.

The stability of optimal plans now follows as in the classical optimal transport problem by exploiting the equivalence between (i) and (ii), see for instance [14, Theorem 5.20]. Finally, the last statement follows from the following observation: let $(\boldsymbol{\gamma}_i)_{i\geq 0} \subset \mathrm{Opt}(\mu,\nu)$ be a countable dense subset, and define

$$\boldsymbol{\gamma}_\infty := \sum_{i\geq 0} \frac{1}{2^i} \gamma_i.$$

Then $\boldsymbol{\gamma}_\infty \in \mathrm{Opt}(\mu,\nu)$ by the convexity of the constraints (2) and the linearity of the cost. Furthermore, since its support contains the supports of all the $\boldsymbol{\gamma}_i$'s, and since they are dense inside $\mathrm{Opt}(\mu,\nu)$, the support of $\boldsymbol{\gamma}_\infty$ contains that of any optimal plan. Hence it suffices to apply $(i) \Rightarrow (iii)$ to γ_∞, to conclude. □

Remark 2.4 The idea on which is based the proof of the last part of the above proposition is well-known for the classical transport problem. Recently, the first author used the same tool to prove a similar result for the optimal partial transport problem (see [6]). Observe also that here the fact that the cost function is the squared distance does not play any crucial role. Therefore many of the statements in this section hold for much more general cost functions (we will not stress this point any further).

The following result is the analogue in our setting of Brenier's theorem on existence and uniqueness of optimal transport maps [3, 4]:

Corollary 2.5 (On uniqueness of optimal plans) *Let $\mu,\nu \in \mathcal{M}_2(\Omega)$, and fix $\boldsymbol{\gamma} \in \mathrm{Opt}(\mu,\nu)$. Then:*

(i) If $\mu \ll \mathcal{L}^d$, then $\boldsymbol{\gamma}_\Omega^{\overline{\Omega}}$ is unique, and it is given by $(\mathrm{Id}, T)_\#\mu$, where $T : \Omega \to \overline{\Omega}$ is the gradient of a convex function. (However, $\boldsymbol{\gamma}$ as a whole may be not uniquely defined as there may be multiple ways of bringing the mass from the boundary to ν if no hypothesis on ν are made).

(ii) If $\mu, \nu \ll \mathcal{L}^d$, then $\boldsymbol{\gamma}$ is unique.

Proof. Thanks to the equivalence $(i) \Leftrightarrow (iii)$ of the previous theorem, the result follows exactly as in the classical transport problem, see for instance [2, Theorem 6.2.4 and Remark 6.2.11]. $\square$

Remark 2.6 Let us point out that given a sequence $(\mu_n) \subset \mathcal{M}_2(\Omega)$ weakly converges to some $\mu \in \mathcal{M}_2(\Omega)$ in duality with functions in $C_c(\Omega)$, the following two things are equivalent:

- $m_2(\mu_n) \to m_2(\mu)$,
- $\displaystyle \lim_{r\to 0} \sup_{n\in\mathbb{N}} \int_{\{d(x,\partial\Omega)\le r\}} d^2(x, \partial\Omega)\, d\mu_n(x) = 0.$

Proposition 2.7 (The space $(\mathcal{M}_2(\Omega), Wb_2)$) *A sequence $(\mu_n) \subset \mathcal{M}_2(\Omega)$ converges to $\mu \in \mathcal{M}_2(\Omega)$ with respect to Wb_2 if and only if it converges weakly in duality with functions in $C_c(\Omega)$ and $m_2(\mu_n) \to m_2(\mu)$.*

Moreover the space $(\mathcal{M}_2(\Omega), Wb_2)$ is Polish, and the subset $\mathcal{M}_{\le M}(\Omega)$ of $\mathcal{M}_2(\Omega)$ consisting of measures with mass less or equal to $M \in \mathbb{R}$ is compact.

Proof. Suppose that $Wb_2(\mu_n, \mu) \to 0$, and let $\mathbf{0}$ denote the vanishing measure. Then, since $m_2(\mu) = Wb_2(\mu, \mathbf{0})$, from the triangle inequality we immediately get $m_2(\mu_n) \to m_2(\mu)$. Now, given $\varphi \in C_c(\Omega)$, fix $\varepsilon > 0$ and find a Lipschitz function ψ such that

$$\begin{gathered} \operatorname{supp}(\psi) \subset \operatorname{supp}(\varphi), \\ \sup_{x\in\Omega} |\varphi(x) - \psi(x)| \le \varepsilon. \end{gathered}$$

Observe that from inequality (3) and the uniform bound on $m_2(\mu_n), m_2(\mu)$, we have that the mass of μ_n, μ on $\operatorname{supp}(\varphi)$ is uniformly bounded by some constant C. Thus, choosing $\boldsymbol{\gamma}_n \in \textsc{Opt}(\mu_n, \mu)$ we have

$$\begin{aligned} \left| \int \varphi\, d\mu_n - \int \varphi\, d\mu \right| &\le 2C\varepsilon + \left| \int \psi\, d\mu_n - \int \psi\, d\mu \right| \\ &= 2C\varepsilon + \left| \int \psi(x)\, d\boldsymbol{\gamma}_n(x,y) - \int \psi(y)\, d\boldsymbol{\gamma}_n(y,y) \right| \\ &\le 2C\varepsilon + \int_{\operatorname{supp}(\psi)\times\operatorname{supp}(\psi)} |\psi(x) - \psi(y)|\, d\boldsymbol{\gamma}_n(x,y) \\ &\le 2C\varepsilon + \operatorname{Lip}(\psi) \int_{\operatorname{supp}(\psi)\times\operatorname{supp}(\psi)} |x-y|\, d\boldsymbol{\gamma}_n(x,y) \\ &\le 2C\varepsilon + C\operatorname{Lip}(\psi) \sqrt{\int_{\operatorname{supp}(\psi)\times\operatorname{supp}(\psi)} |x-y|^2\, d\boldsymbol{\gamma}_n(x,y)} \\ &\le 2C\varepsilon + C\operatorname{Lip}(\psi)\, Wb_2(\mu_n, \mu). \end{aligned}$$

Letting first $n \to \infty$ and then $\varepsilon \to 0$, we obtain the weak convergence.

Conversely, let (μ_n) be a sequence weakly converging to μ and satisfying $m_2(\mu_n) \to m_2(\mu)$, and choose $\boldsymbol{\gamma}_n \in \textsc{Opt}(\mu_n, \mu)$. Up to passing to a subsequence, thanks to Proposition 2.3 we may assume that $(\boldsymbol{\gamma}_n)$ weakly converges to some optimal plan $\boldsymbol{\gamma}$ in duality with functions in $C_c(\overline{\Omega} \times \overline{\Omega} \setminus \partial\Omega \times \partial\Omega)$. Choose $r > 0$, define

$$A_r := \Big\{ (x,y) \in \overline{\Omega} \times \overline{\Omega} \ : \ d(x, \partial\Omega) < r,\ d(y, \partial\Omega) < r \Big\},$$

and recall that $\mathrm{supp}(\boldsymbol{\gamma}) \subset \mathcal{A}$, $\mathcal{A}$ being defined by (5). Hence

$$\varlimsup_{n\to\infty} \int\limits_{\overline{\Omega}\times\overline{\Omega}} |x-y|^2 \, d\boldsymbol{\gamma}_n(x,y) \le \varlimsup_{n\to\infty} \int\limits_{\overline{\Omega}\times\overline{\Omega}\setminus A_r} |x-y|^2 \, d\boldsymbol{\gamma}_n(x,y) + \varlimsup_{n\to\infty} \int\limits_{A_r} |x-y|^2 \, d\boldsymbol{\gamma}_n(x,y)$$

$$\le \int\limits_{\overline{\Omega}\times\overline{\Omega}\setminus A_r} |x-y|^2 \, d\boldsymbol{\gamma}(x,y) + 2 \varlimsup_{n\to\infty} \int\limits_{A_r} d^2(x,\partial\Omega) \, d\boldsymbol{\gamma}_n(x,y) + 2 \varlimsup_{n\to\infty} \int\limits_{A_r} d^2(y,\partial\Omega) \, d\boldsymbol{\gamma}_n(x,y)$$

$$= \int\limits_{\overline{\Omega}\times\overline{\Omega}\setminus A_r} |x-y|^2 \, d\boldsymbol{\gamma}(x,y) + 2 \varlimsup_{n\to\infty} \int\limits_{\{d(x,\partial\Omega)\le r\}} d^2(x,\partial\Omega) \, d\mu_n(x) + 2 \varlimsup_{n\to\infty} \int\limits_{\{d(y,\partial\Omega)\le r\}} d^2(y,\partial\Omega) \, d\mu(y),$$

where in the second step we used the fact that $\overline{\Omega} \times \overline{\Omega} \setminus A_r$ is closed. Letting $r \downarrow 0$, using Remark 2.6, the stability of optimality statement of Proposition 2.3 above and observing that the result does not depend on the subsequence chosen, we get

$$\varlimsup_{n\to\infty} Wb_2^2(\mu_n,\mu) = \int\limits_{\overline{\Omega}\times\overline{\Omega}} |x-y|^2 \, d\boldsymbol{\gamma}(x,y) = Wb_2^2(\mu,\mu) = 0$$

as desired.

The claim on the compactness of $\mathcal{M}_{\le M}(\Omega)$ is easy. It is also immediate to check that $\cup_M \mathcal{M}_{\le M}$ is dense in $\mathcal{M}_2(\Omega)$, so that to prove the separability of $\mathcal{M}_2(\Omega)$ it is enough to prove the separability of each of the $\mathcal{M}_{\le M}(\Omega)$'s, which follows by standard means by considering the set of rational combination of Dirac masses centered at rational points. Thus we only prove completeness. Let (μ_n) be a Cauchy sequence with respect to Wb_2. We observe that $m_2(\mu_n)$ are uniformly bounded. Moreover thanks to inequality (3) the set $\{\mu_n\}$ is weakly relatively compact in duality with functions in $C_c(\Omega)$, which implies the existence of a subsequence (μ_{n_k}) weakly converging to some measure μ. By the lower semicontinuity of Wb_2 with respect to weak convergence we get

$$m_2(\mu) = W_2(\mu,\mathbf{0}) \le \varliminf_{k\to\infty} W_2(\mu_{n_k},\mathbf{0}),$$

so that $\mu \in \mathcal{M}_2(\Omega)$. Again by the lower semicontinuity of Wb_2 we have

$$Wb_2(\mu,\mu_m) \le \varliminf_{k\to\infty} Wb_2(\mu_{n_k},\mu_m),$$

so

$$\varlimsup_{m\to\infty} Wb_2(\mu,\mu_m) \le \varlimsup_{m,k\to\infty} Wb_2(\mu_{n_k},\mu_m) = 0.$$

□

Remark 2.8 Note carefully that in the above proposition we are talking about weak convergence in duality with functions with compact support in Ω, and not, e.g., with continuous and bounded functions in $\overline{\Omega}$. Indeed, the mass can 'disappear' inside the boundary, so that in general we only have

$$\liminf_{n\to\infty} \mu_n(\Omega) \ge \mu(\Omega),$$

for any sequence $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_2(\Omega)$ such that $Wb_2(\mu_n,\mu) \to 0$.

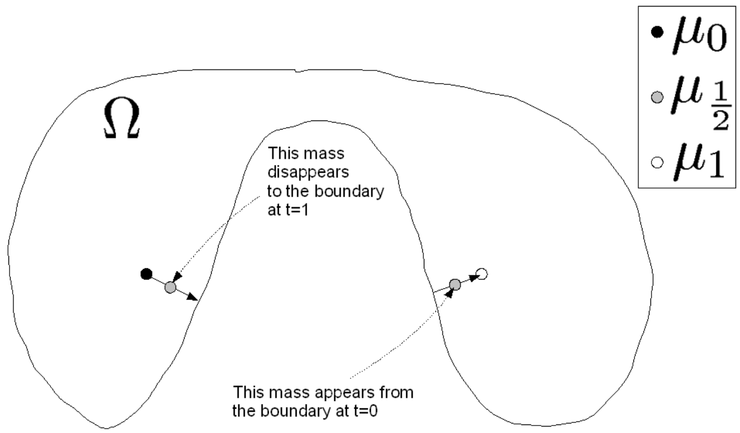

Figure 2: Geodesic interpolation is always possible in the space $(\mathcal{M}_2(\Omega), Wb_2)$. Indeed, the mass can 'appear' only at $t = 0$, can 'vanish' only at $t = 1$, and for $t \in (0, 1)$ it moves along straight segments inside Ω. In particular, in the open interval $(0, 1)$, a geodesic with respect to Wb_2 is also a geodesic with respect to W_2.

Proposition 2.9 (Geodesics) *The space $(\mathcal{M}_2(\Omega), Wb_2)$ is a geodesic space. A curve $[0, 1] \ni t \mapsto \mu_t$ is a minimizing geodesic with constant speed if and only if there exists $\boldsymbol{\gamma} \in \mathrm{Opt}(\mu_0, \mu_1)$ such that*

$$\mu_t = ((1-t)\pi^1 + t\pi^2)_{\#}\boldsymbol{\gamma}, \qquad \forall\, t \in (0, 1). \tag{11}$$

Also, given a geodesic (μ_t), for any $t \in (0, 1)$ and $s \in [0, 1]$ there is a unique optimal plan γ_t^s from μ_t to μ_s, which is given by

$$\boldsymbol{\gamma}_t^s := ((1-t)\pi^1 + t\pi^2, (1-s)\pi^1 + s\pi^2)_{\#}\boldsymbol{\gamma},$$

where $\boldsymbol{\gamma} \in \mathrm{Opt}(\mu_0, \mu_1)$ is the plan which induces the geodesic via Equation (11). Furthermore, the plan $\boldsymbol{\gamma}_t^s$ is the unique optimal transport plan from μ_t to μ_s for the classical transport problem.

In particular, the space $(\mathcal{M}_2(\Omega), Wb_2)$ is non-branching, and the mass of μ_t inside Ω is constant - possibly infinite - for $t \in (0, 1)$.

Observe that Equation (11) does *not* hold for $t = 0, 1$, as the marginals of $\boldsymbol{\gamma}$ generally charge also $\partial\Omega$. We further remark that such a statement would be false for the classical Wasserstein distance W_2. Indeed, if $\boldsymbol{\gamma}$ is an optimal plan for W_2, then the measures μ_t defined by (11) will *not* in general be concentrated in Ω, unless Ω is convex.

Proof. The only new part with respect to the classical case is that, if $\boldsymbol{\gamma}$ is an optimal plan from μ_0 to μ_1, then the measures μ_t defined by (11) are concentrated in Ω (and not just in its convex hull). Once this result is proved, the rest of the proof becomes exactly the same as in the standard case of the Wasserstein distance, see [2, Paragraph 7.2]. Hence, we are going to prove only this new part.

To this aim, recall that thanks to Proposition 2.3 we know that an optimal plan $\boldsymbol{\gamma}$ is concentrated on the set $\mathcal{A}$ defined in (5). Thus to conclude it is enough to show that for every $(x, y) \in \mathcal{A}$ the segment $t \mapsto (1-t)x + ty$ is entirely contained in Ω. Argue by contradiction and assume that for some $\bar{t} \in (0,1)$ it holds $(1-\bar{t})x + \bar{t}y \notin \Omega$, then from

$$d(x, \partial\Omega) \leq |x - (1-\bar{t})x + \bar{t}y| = \bar{t}\,|x-y|,$$
$$d(y, \partial\Omega) \leq |y - (1-\bar{t})x + \bar{t}y| = (1-\bar{t})\,|x-y|,$$

we deduce

$$d^2(x, \partial\Omega) + d^2(y, \partial\Omega) \leq (\bar{t}^2 + (1-\bar{t})^2)\,|x-y|^2 < |x-y|^2,$$

which contradicts $(x,y) \in \mathcal{A}$. □

Remark 2.10 (A comparison between Wb_2 and W_2) Let $\mu, \nu \in \mathcal{M}_2(\Omega)$ and assume that $0 < \mu(\Omega) = \nu(\Omega) < \infty$. Then any plan $\boldsymbol{\gamma}$ which is optimal for the classical transportation cost is admissible for the new one. Therefore we have the inequality:

$$Wb_2(\mu, \nu) \leq W_2(\mu, \nu), \qquad \forall\, \mu, \nu \in \mathcal{M}_2(\Omega) \text{ s.t. } \mu(\Omega) = \nu(\Omega) > 0. \tag{12}$$

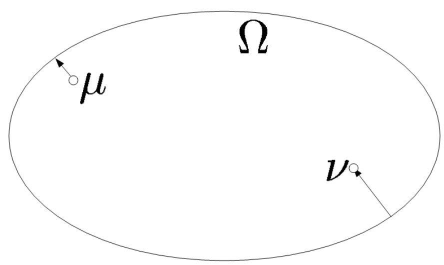

Figure 3: For measures with the same amount of mass, the distance Wb_2 is smaller than the classical W_2: as the picture shows, it may be much better to exchange the mass with the boundary rather than internally.

Proposition 2.11 (An estimate on the directional derivative) *Let $\mu, \nu \in \mathcal{M}_2(\Omega)$ and $\boldsymbol{w} : \Omega \to \mathbb{R}^d$ a bounded vector field with compact support. Also, let $\boldsymbol{\gamma} \in \mathrm{Opt}(\mu, \nu)$, and define $\mu_t := (\mathrm{Id} + t\boldsymbol{w})_\# \mu$. Then*

$$\limsup_{t \to 0} \frac{Wb_2^2(\mu_t, \nu) - Wb_2^2(\mu, \nu)}{t} \leq -2 \int \langle \boldsymbol{w}(x), y - x \rangle \, d\boldsymbol{\gamma}(x, y).$$

Proof. Observe that since $\boldsymbol{w}$ is compactly supported in Ω, for $t > 0$ sufficiently small $\mu_t \in \mathcal{M}_2(\Omega)$, so that the statement makes sense. Now it is simple to check that the plan $\boldsymbol{\gamma}_t$ defined by

$$\boldsymbol{\gamma}_t := ((\mathrm{Id} + t\boldsymbol{w}) \circ \pi^1, \pi^2)_\# \boldsymbol{\gamma},$$

belongs to $\mathrm{Adm}(\mu_t, \nu)$. Hence

$$\begin{aligned} Wb_2^2(\mu_t, \nu) &\leq \int |x-y|^2 \, d\boldsymbol{\gamma}_t(x,y) = \int |x + t\boldsymbol{w}(x) - y|^2 \, d\boldsymbol{\gamma}(x,y) \\ &= Wb_2^2(\mu, \nu) - 2t \int \langle \boldsymbol{w}(x), y - x \rangle \, d\boldsymbol{\gamma}(x,y) + t^2 \int |\boldsymbol{w}(x)|^2 \, d\boldsymbol{\gamma}(x,y), \end{aligned}$$

and the conclusion follows. □

3 The heat equation with Dirichlet boundary condition as a ‘gradient flow’

This section contains an application of our new transportation distance: we are going to show that the gradient flow of the entropy functional $\int_\Omega [\rho \log(\rho) - \rho]\, dx$ coincides with the heat equation, with Dirichlet boundary condition equal to 1. To prove such a result, we will first study some of the properties of the entropy, showing in particular a lower bound on its slope, see Proposition 3.2. Then, following the strategy introduced in [8], we will apply the minimizing movement scheme to prove our result. Finally we will show that the discrete solutions constructed by minimizing movements enjoy a comparison principle: if $(\rho_k^\tau)_{k\in\mathbb{N}}$ and $(\tilde{\rho}_k^\tau)_{k\in\mathbb{N}}$ are two discrete solution for a time step $\tau > 0$, and $\rho_0^\tau \le \tilde{\rho}_0^\tau$, then $\rho_k^\tau \le \tilde{\rho}_k^\tau$ for all $k \in \mathbb{N}$. Letting $\tau \to 0$, this monotonicity result allows to recover the classical maximum principle for the heat equation.

To be clear: we will not state any result concerning existence of the gradient flow of the entropy (we will not identify the slope of the entropy, nor the infinitesimal description of the distance Wb_2). What we will do is a work ‘by hands’: we will show that we have compactness in the minimizing movements scheme and prove that any limit is a weak solution of the heat equation with Dirichlet boundary conditions.

3.1 The entropy

The entropy functional $E : \mathcal{M}_2(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$E(\mu) := \begin{cases} \displaystyle\int_\Omega e(\rho(x))\, dx & \text{if } \mu = \rho\mathcal{L}^d|_\Omega, \\ +\infty & \text{otherwise,} \end{cases}$$

where $e : [0, +\infty) \to [0, +\infty)$ is given by

$$e(z) := z \log(z) - z + 1.$$

From now on, since we will often deal with absolutely continuous measures, by abuse of notation we will sometimes use ρ to denote the measure $\rho\mathcal{L}^d|_\Omega$. In particular, we will write $\textsc{Adm}(\rho, \rho')$ in place of $\textsc{Adm}(\rho\mathcal{L}^d|_\Omega, \rho'\mathcal{L}^d|_\Omega)$.

Proposition 3.1 (Semicontinuity and compactness of sublevels) *The functional $E : \mathcal{M}_2(\Omega) \to \mathbb{R} \cup \{+\infty\}$ takes value in $[0, +\infty]$, it is lower semicontinuous with respect to Wb_2, and its sublevels are compact.*

Proof. If $\mu = \rho\mathcal{L}^d|_\Omega$, thanks to Jensen inequality we have

$$e\left(\frac{\mu(\Omega)}{|\Omega|}\right) = e\left(\frac{1}{|\Omega|}\int_\Omega \rho\, dx\right) \le \frac{1}{|\Omega|}\int_\Omega e(\rho)\, dx = \frac{E(\mu)}{|\Omega|}. \tag{13}$$

This inequality bounds the mass of ρ in terms of the entropy, which gives the relative compactness of the sublevels of E thanks to Proposition 2.7. The bound $E(\mu) \ge 0$ is immediate as $e \ge 0$. Finally, the lower semicontinuity follows from the convexity and superlinearity of e and from fact that convergence with respect to Wb_2 implies weak convergence (see Proposition 2.7). □

We recall that the *slope* of the functional E defined on the metric space $(\mathcal{M}_2(\Omega), Wb_2)$ is defined as:

$$|\nabla E|(\mu) := \limsup_{\nu\to\mu} \frac{(E(\mu) - E(\nu))^+}{Wb_2(\mu,\nu)}.$$

Proposition 3.2 (Bound of the slope in terms of Fisher's information) *The slope of E is bounded from below by the square root of the Fisher information $F : \mathcal{M}_2(\Omega) \to [0, +\infty]$:*

$$F(\mu) := \begin{cases} 4\displaystyle\int_\Omega \left|\nabla\sqrt{\rho}\right|^2 dx & \text{if } \mu = \rho\mathcal{L}^d|_\Omega \text{ and } \sqrt{\rho} \in H^1(\Omega), \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. Take $\mu \in \mathcal{M}_2(\Omega)$, define $m := \mu(\Omega)$, and let $\mathcal{M}_m(\Omega)$ be the set of non-negative measures on Ω with mass m. On $\mathcal{M}_m(\Omega)$, we can consider the Wasserstein distance W_2. Consider the functional $E : (\mathcal{M}_m(\Omega), W_2) \to \mathbb{R} \cup \{+\infty\}$. It is well-known that $|\nabla E|(\mu) = \sqrt{F(\mu)}$ for all $\mu \in \mathcal{M}_1$, see [2, Chapter 10]. Then, it is easily checked by a scaling argument that the formula remains true for arbitrary $m \geq 0$. Hence, taking into account inequality (12), we obtain

$$|\nabla E|(\mu) \geq \limsup_{\mathcal{M}_m(\Omega)\ni\nu\to\mu} \frac{(E(\mu) - E(\nu))^+}{Wb_2(\mu,\nu)} \geq \limsup_{\mathcal{M}_m(\Omega)\ni\nu\to\mu} \frac{(E(\mu) - E(\nu))^+}{W_2(\mu,\nu)} = \sqrt{F(\mu)},$$

as desired. □

Proposition 3.3 (A directional derivative of E) *Let $\mu = \rho\mathcal{L}^d \in \mathcal{M}_2(\Omega)$ be such that $E(\mu) < +\infty$, and let $\boldsymbol{w} : \Omega \to \mathbb{R}^d$ be a C^∞ vector field with compact support. Define $\mu_t := (\mathrm{Id} + t\boldsymbol{w})_\#\mu$. Then*

$$\lim_{t\to 0} \frac{E(\mu_t) - E(\mu)}{t} = \int_\Omega \rho \,\mathrm{div}\boldsymbol{w}\, dx.$$

Proof. Since $\boldsymbol{w}$ is compactly supported, $\mu_t \in \mathcal{M}_2(\Omega)$ for sufficiently small t, and the proof is exactly the same as the one in the Wasserstein case. □

Remark 3.4 [A source of difficulties] It is important to underline that the entropy E is *not* geodesically convex on the space $(\mathcal{M}_2(\Omega), Wb_2)$. Indeed, since for instance the mass can disappear at the boundary for $t = 1$, it is possible that an high concentration of mass near $\partial\Omega$ gives $\lim_{t\uparrow 1} E(\mu_t) = +\infty$, while $E(\mu_1) < +\infty$. (Observe that, once the mass has reached $\partial\Omega$, it does not contribute any more to the energy!) Still, since for $t, s \in (0, 1)$ the optimal transport plan for Wb_2 coincides with the optimal transport plan for W_2 (see Proposition 2.9), $t \mapsto E(\mu_t)$ is convex in the open interval $(0, 1)$ (see for instance [2, Chapter 9]).

3.2 Minimizing movements for the entropy

In this paragraph we apply the minimizing movements to construct a weak solution to the heat equation with Dirichlet boundary condition.

We briefly review the minimizing movement scheme, referring to [2] for a detailed description and general results. Fix $\rho_0 \in \mathcal{M}_2(\Omega)$ such that $E(\rho_0) < +\infty$ (given the

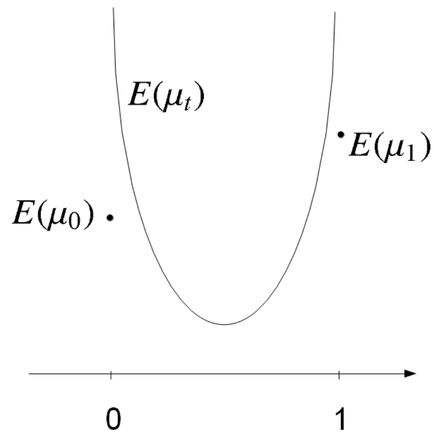

Figure 4: For typical μ_0, μ_1, a geodesic connecting them takes mass from the boundary at $t = 0$ and leaves mass at $t = 1$. In this case the graph of $t \mapsto E(\mu_t)$ looks like in the picture: in the interval $(0, 1)$ the function is convex and converges to $+\infty$ as $t \to 0, 1$. The value of $E(\mu_0)$ and $E(\mu_1)$ has basically no connection with the values in intermediate times.

lack of convexity of E, we need to assume that the entropy at the initial point is finite, thus in particular the measure is absolutely continuous), and fix a time step $\tau > 0$. Set $\rho_0^\tau := \rho_0$, and define recursively ρ_{n+1}^τ as the unique minimizer of

$$\mu \quad \mapsto \quad E(\mu) + \frac{Wb_2^2(\mu, \rho_n^\tau)}{2\tau}$$

(see Proposition 3.6 below). Then, we define the *discrete solution* $t \mapsto \rho^\tau(t) \in \mathcal{M}_2(\Omega)$ by

$$\rho^\tau(t) := \rho_n^\tau \qquad \text{for } t \in [n\tau, (n+1)\tau).$$

We recall that the space $W_0^{1,1}(\Omega)$ is defined as the closure of $C_0^\infty(\Omega)$ with respect to the $W^{1,1}$-norm. (Observe that this definition requires no smoothness assumptions on $\partial\Omega$.) Then we say that $f \in W^{1,1}(\Omega)$ has trace 1 if $f - 1 \in W_0^{1,1}(\Omega)$. (More in general, given a smooth function $\phi : \overline{\Omega} \to \mathbb{R}$, one may say that $f \in W^{1,1}(\Omega)$ has trace ϕ if $f - \phi \in W_0^{1,1}(\Omega)$.)

Our main theorem is the following:

Theorem 3.5 *With the above notation, for any sequence $\tau_k \downarrow 0$ there exists a subsequence, not relabeled, such that, for any $t \geq 0$, $\rho^{\tau_k}(t)$ converges to some limit measure $\rho(t)$ in $(\mathcal{M}_2(\Omega), Wb_2)$ as $k \to \infty$. The map $t \mapsto (\rho(t) - 1)$ belongs to $L^2_{loc}([0, +\infty), W_0^{1,1}(\Omega))$, and $t \mapsto \rho(t)$ is a weak solution of the heat equation*

$$\begin{cases} \dfrac{d}{dt}\rho(t) & = \Delta\rho(t), \\ \rho(0) & = \rho_0. \end{cases} \tag{14}$$

We recall that a weakly continuous curve of measure $t \mapsto \mu_t \in \mathcal{M}_2(\Omega)$ is said to be a weak solution of (14) if

$$\int_\Omega \varphi \, d\mu_s(x) - \int_\Omega \varphi \, d\mu_t(x) = \int_t^s \left(\int_\Omega \Delta\varphi \, d\mu_r(x) \right) dr, \qquad \forall\, 0 \leq t < s, \ \forall\, \varphi \in C_c^\infty(\Omega),$$

In order to prove this theorem, we need the following lemma, which describes the behavior of a single step of the minimizing movements scheme.

Proposition 3.6 (A step of the minimizing movement) *Let $\mu \in \mathcal{M}_2(\Omega)$ and $\tau > 0$. Then there exists a unique minimum $\mu_\tau \in \mathcal{M}_2(\Omega)$ of*

$$\sigma \mapsto E(\sigma) + \frac{Wb_2^2(\mu, \sigma)}{2\tau}. \tag{15}$$

Such a minimum satisfies:

(i) $\mu_\tau = \rho_\tau \mathcal{L}^d|_{\Omega}$, *with* $\rho_\tau - 1 \in W_0^{1,1}(\Omega)$.

(ii) The restriction to $\Omega \times \overline{\Omega}$ of any optimal transport plan from μ_τ to μ is induced by a map T, which satisfies

$$\frac{T(x) - x}{\tau}\rho_\tau(x) = -\nabla\rho_\tau(x), \qquad \mathcal{L}^d - a.e.\ x \tag{16}$$

Proof. The existence of a minimum $\mu_\tau = \rho_\tau \mathcal{L}^d|_{\Omega}$ follows by a standard compactness-semicontinuity argument, while the uniqueness is a direct consequence of the convexity of $Wb_2^2(\cdot, \mu)$ with respect to usual linear interpolation of measures and the strict convexity of $E(\cdot)$.

It is well known that at minimum of (15) the slope is finite (see [2, Lemma 3.1.3]). Hence $\sqrt{\rho_\tau} \in H^1(\Omega)$ by Proposition 3.2, and so $\rho_\tau \in W^{1,1}(\Omega)$ by Hölder inequality. Moreover, thanks to (27) below we have

$$e^{-d(x,\partial\Omega)^2/(2\tau)} \leq \rho_\tau(x) \leq e^{3\mathrm{Diam}(\Omega)d(x,\partial\Omega)/(2\tau)} \qquad \forall\, x \in \Omega, \tag{17}$$

which easily implies that ρ_τ has trace 1 on $\partial\Omega$ (we postpone the proof of (17) to the next section, where we will prove also other useful inequalities on ρ_τ - see Proposition 3.7). This shows (*i*).

To prove (*ii*), we start by observing that Corollary 2.5 and the absolute continuity of μ_τ guarantee the existence of T. Now, choose a C^∞ vector field $\boldsymbol{w}$ with compact support in Ω and define $\rho_\tau^t := (\mathrm{Id} + t\boldsymbol{w})_\# \rho_\tau$. Using the minimality of ρ_τ we get

$$E(\rho_\tau^t) - E(\rho_\tau) + \frac{Wb_2^2(\rho_\tau^t, \mu) - Wb_2^2(\rho_\tau, \mu)}{2\tau} \geq 0.$$

Dividing by t and letting $t \downarrow 0$, thanks to Propositions 3.3 and 2.11 we get

$$\int_\Omega \rho \,\mathrm{div}\boldsymbol{w}\, dx - \int_\Omega \langle \boldsymbol{w}, \frac{T - \mathrm{Id}}{\tau}\rangle \rho\, dx \geq 0.$$

Exchanging $\boldsymbol{w}$ with $-\boldsymbol{w}$ and exploiting the arbitrariness of $\boldsymbol{w}$ the result follows. □

To prove Theorem 3.5 we will use the following a priori bound for the discrete solution, see [2, Lemma 3.2.2 and Equation (3.2.3)]:

$$\frac{1}{2}\sum_{i=n}^{m-1} \frac{Wb_2^2(\rho_i^\tau, \rho_{i+1}^\tau)}{\tau} + \frac{\tau}{2}\sum_{i=n}^{m-1} |\nabla E|^2(\rho_i^\tau) \leq E(\rho_m^\tau) - E(\rho_n^\tau) \qquad \forall\, n \leq m \in \mathbb{N}. \tag{18}$$

Proof of Theorem 3.5. **- Compactness argument.** Let $\{\tau_k\}_{k\in\mathbb{N}}$ be a sequence converging to 0. First of all we observe that, thanks to (13) and the inequality $E(\rho^{\tau_k}(t)) \leq E(\rho_0)$, the mass of the measures $\rho^{\tau_k}(t)$ is uniformly bounded for all $k \in \mathbb{N}$, $t \geq 0$. Then a standard diagonal argument shows that there exists a subsequence, not relabeled, such that $\rho^{\tau_k}(t)$ converges to some $\rho(t)$ in $(\mathcal{M}_2(\Omega), Wb_2)$ for any $t \in \mathbb{Q}_+$. Now, thanks to the uniform bound on the discrete speed

$$\frac{1}{2}\sum_{i=n}^{m-1} \frac{Wb_2^2(\rho_i^{\tau_k}, \rho_{i+1}^{\tau_k})}{\tau} \leq E(\rho_m^\tau) - E(\rho_n^\tau) \leq E(\rho_0),$$

(which is a direct consequence of (18)) we easily get

$$Wb_2(\rho^{\tau_k}(t), \rho^{\tau_k}(s)) \leq \sqrt{2\, E(\rho_0)\, [t - s + \tau_k]} \qquad \forall\, 0 \leq s \leq t, \tag{19}$$

which implies the convergence of $\rho^{\tau_k}(t)$ for every $t \geq 0$.

- Any limit point is a weak solution of the heat equation. Let $\tau_k \downarrow 0$ be a sequence such that $\rho^{\tau_k}(t)$ converges to some $\rho(t)$ in $(\mathcal{M}_2(\Omega), Wb_2)$ for any $t \geq 0$. We want to prove that $t \mapsto \rho(t)$ is a weak solution of the heat equation. For any $\tau > 0$, $n \in \mathbb{N}$, let T_n^τ be the map which induces $(\boldsymbol{\gamma}_n^\tau)_\Omega^{\overline{\Omega}}$, where $\boldsymbol{\gamma}_n^\tau \in \mathrm{O{\scriptsize PT}}(\rho_{n+1}^\tau, \rho_n^\tau)$ (see Corollary 2.5(i)). Fix $\varphi \in C_c^\infty(\Omega)$ and observe that

$$\begin{aligned}
\int_\Omega \varphi \rho_{n+1}^\tau \, dx - \int_\Omega (\varphi \circ T_n^\tau)\rho_{n+1}^\tau \, dx &= \int_\Omega \left(\int_0^1 \langle \nabla\varphi \circ ((1-\lambda)T_n^\tau + \lambda \mathrm{Id}), \mathrm{Id} - T_n^\tau \rangle d\lambda \right) \rho_{n+1}^\tau \, dx \\
&= -\int_\Omega \langle \nabla\varphi, T_n^\tau - \mathrm{Id}\rangle \rho_{n+1}^\tau \, dx + R(\tau, n) \\
&= \tau \int_\Omega \langle \nabla\varphi, \nabla\rho_{n+1}^\tau \rangle \, dx + R(\tau, n) \\
&= -\tau \int_\Omega \Delta\varphi\, \rho_{n+1}^\tau \, dx + R(\tau, n),
\end{aligned} \tag{20}$$

where at the third step we used (16), and the reminder term $R(\tau, n)$ is bounded by

$$|R(\tau, n)| \leq (\mathrm{Lip}\nabla\varphi) \int_\Omega |T_n^\tau - \mathrm{Id}|^2 \rho_{n+1}^\tau \, dx = \mathrm{Lip}(\nabla\varphi) Wb_2^2(\rho_n^\tau, \rho_{n+1}^\tau). \tag{21}$$

Now, since the support of φ is contained in Ω and $\left((T_n^\tau)_\# \rho_{n+1}^\tau\right)|_\Omega = \pi_\#^2 \left((\boldsymbol{\gamma}_n^\tau)_\Omega^\Omega\right)$ we have

$$\int_\Omega \varphi \rho_n^\tau \, dx - \int_\Omega (\varphi \circ T_n^\tau) \rho_{n+1}^\tau \, dx = \int_{\overline{\Omega} \times \Omega} \varphi(y)\, d(\boldsymbol{\gamma}_n^\tau)_{\partial\Omega}^\Omega(x, y).$$

By Proposition 2.3 we have $|x - y| = d(y, \partial\Omega)$ for $(\boldsymbol{\gamma}_n^\tau)_{\partial\Omega}^\Omega$-a.e. (x, y), which implies

$$\begin{aligned}
Wb_2^2(\rho_{n+1}^\tau, \rho_n^\tau) &\geq \int_{\overline{\Omega} \times \mathrm{supp}(\varphi)} |x - y|^2 \, d(\boldsymbol{\gamma}_n^\tau)_{\partial\Omega}^\Omega(x, y) \\
&= \int_{\overline{\Omega} \times \mathrm{supp}(\varphi)} d(y, \partial\Omega)^2 \, d(\boldsymbol{\gamma}_n^\tau)_{\partial\Omega}^\Omega(x, y) \geq c_\varphi \, (\boldsymbol{\gamma}_n^\tau)_{\partial\Omega}^\Omega(\overline{\Omega} \times \mathrm{supp}(\varphi)),
\end{aligned}$$

where $c_\varphi := \min_{y \in \mathrm{supp}(\varphi)} d(y, \partial\Omega)^2 > 0$. Hence

$$\left| \int_\Omega \varphi \rho_n^\tau \, dx - \int_\Omega (\varphi \circ T_n^\tau) \rho_{n+1}^\tau \, dx \right| \leq \frac{\|\varphi\|_\infty}{c_\varphi} Wb_2^2(\rho_{n+1}^\tau, \rho_n^\tau).$$

Combining the above estimate with (20) and (21), we obtain

$$\int_\Omega \varphi\rho_{n+1}^\tau \, dx - \int_\Omega \varphi\rho_n^\tau \, dx \; = -\tau \int_\Omega \Delta\varphi \, \rho_{n+1}^\tau \, dx + \tilde{R}(\tau, n), \tag{22}$$

where

$$\left|\tilde{R}(\tau, n)\right| \leq \left(\mathrm{Lip}(\nabla\varphi) + \frac{\|\varphi\|_\infty}{c_\varphi}\right) Wb_2^2(\rho_n^\tau, \rho_{n+1}^\tau). \tag{23}$$

Now, choose $0 \leq t < s$, let $\tau = \tau_k$, and add up Equation (22) from $n = [t/\tau_k]$ to $m = [s/\tau_k] - 1$ to get

$$\int_\Omega \varphi\rho^{\tau_k}(s) \, dx - \int_\Omega \varphi\rho^{\tau_k}(t) \, dx = \int_{\tau_k[t/\tau_k]}^{\tau_k[s/\tau_k]} \left(\int_\Omega \Delta\varphi \, \rho^{\tau_k}(r) \, dx\right) dr + \sum_{n=[t/\tau_k]}^{[s/\tau_k]-1} \tilde{R}(\tau_k, [r/\tau_k]).$$

We want to take the limit in the above equation as $\tau_k \downarrow 0$. The Wb_2-convergence of $\rho^{\tau_k}(r)$ to $\rho(r)$ combined with Proposition 2.7 gives that the left hand side converges to $\int_\Omega \varphi\rho(s) \, dx - \int_\Omega \varphi\rho(t) \, dx$. For the same reason $\int_\Omega \Delta\varphi \, \rho^{\tau_k}(r) \, dx \to \int_\Omega \Delta\varphi \, \rho(r) \, dx$ for any $r \geq 0$. Thus, since the mass of the measures $\rho^{\tau_k}(t)$ is uniformly bounded we get

$$\int_\Omega |\Delta\varphi \, \rho^{\tau_k}(r)| \, dx \leq \|\Delta\varphi\|_\infty \int_\Omega \rho^{\tau_k}(r) \, dx \leq C_0$$

for some positive constant C_0, so that by the dominated convergence theorem we get

$$\int_{\tau_k[t/\tau_k]}^{\tau_k[s/\tau_k]} \left(\int_\Omega \Delta\varphi \, \rho^{\tau_k}(r) \, dx\right) dr \to \int_t^s \left(\int_\Omega \Delta\varphi \, \rho(r) \, dx\right) dr,$$

as $\tau_k \downarrow 0$. Finally, thanks to (18) and (23) the reminder term is bounded by

$$\begin{aligned}\left|\sum_{n=[t/\tau_k]}^{[s/\tau_k]-1} \tilde{R}(\tau_k, [r/\tau_k])\right| &\leq \left(\mathrm{Lip}(\nabla\varphi) + \frac{\|\varphi\|_\infty}{c_\varphi}\right) \sum_{n=[t/\tau_k]}^{[s/\tau_k]-1} Wb_2^2(\rho_n^{\tau_k}, \rho_{n+1}^{\tau_k}) \\ &\leq 2\tau_k \left(\mathrm{Lip}(\nabla\varphi) + \frac{\|\varphi\|_\infty}{c_\varphi}\right) E(\rho_0),\end{aligned}$$

and thus it goes to 0 as $\tau_k \downarrow 0$. In conclusion we proved that

$$\int_\Omega \varphi\rho(s) \, dx - \int_\Omega \varphi\rho(t) \, dx = \int_t^s \left(\int_\Omega \Delta\varphi \, \rho(r) \, dx\right) dr, \qquad \forall \, 0 \leq t < s, \; \forall \, \varphi \in C_c^\infty(\Omega).$$

Thanks to Equation (19) it is immediate to check that the curve $t \mapsto \rho(t) \in \mathcal{M}_2(\Omega)$ is continuous with respect to Wb_2, and therefore weakly continuous. Finally, since $\rho^\tau(0) = \rho_0$ for any $\tau > 0$, $\rho(0) = \rho_0$ and the initial condition is satisfied.

- The curve $t \mapsto (\rho(t) - 1)$ belongs to $L^2_{loc}([0, +\infty), W_0^{1,1}(\Omega))$. From inequality (18) and Proposition 3.2 we know that

$$\int_0^\infty \left(\int_\Omega \left|\nabla \sqrt{\rho^{\tau_k}(t)}\right|^2 dx\right) dt \leq \frac{1}{4} \int_0^\infty |\nabla E|^2(\rho^{\tau_k}(t)) \, dt \leq \frac{E(\rho_0)}{2},$$

which means that the functions $t \mapsto \sqrt{\rho^{\tau_k}(t)}$ are equibounded in $L^2_{loc}([0, +\infty), H_0^1(\Omega))$. This implies that $t \mapsto \sqrt{\rho(t)}$ belongs to $L^2_{loc}([0, +\infty), H^1(\Omega))$, so that by Hölder inequality $t \mapsto \rho(t) \in L^2_{loc}([0, +\infty), W^{1,1}(\Omega))$. Moreover, thanks to Fatou lemma,

$$\int_0^\infty \liminf_{k \to +\infty} \left(\int_\Omega \left|\nabla \sqrt{\rho^{\tau_k}(t)}\right|^2 dx\right) dt < +\infty,$$

which gives

$$\liminf_{k\to+\infty}\int_\Omega\left|\nabla\sqrt{\rho^{\tau_k}(t)}\right|^2dx<+\infty\qquad\text{for a.e. } t\geq 0,$$

and by Hölder inequality we get

$$\liminf_{k\to+\infty}\int_\Omega\left|\nabla\rho^{\tau_k}(t)\right|dx<+\infty\qquad\text{for a.e. } t\geq 0,$$

Now, for any t such that the above liminf is finite, consider a subsequence k_n (depending on t) such that

$$\sup_{n\in\mathbb{N}}\int_\Omega\left|\nabla\rho^{\tau_{k_n}}(t)\right|dx<+\infty.$$

Then, recalling that $\rho^{\tau_k}(t)\to\rho(t)$ in $(\mathcal{M}_2(\Omega),Wb_2)$, since $\rho^{\tau_{k_n}}(t)$ is uniformly bounded in $W^{1,1}(\Omega)$ and belong to $W_0^{1,1}(\Omega)$ by Proposition 3.6(i), we easily get that $\rho^{\tau_k}(t)\to\rho(t)$ weakly in $W^{1,1}(\Omega)$, and $\rho(t)-1\in W_0^{1,1}(\Omega)$ as desired. □

3.3 A comparison principle

In this section we prove the following monotonicity result for the minimizing movement scheme of E with respect to Wb_2: if we have two measures $\mu,\tilde\mu$ satisfying $\mu\geq\tilde\mu$, then $\mu_\tau\geq\tilde\mu_\tau$ for every $\tau\geq 0$, where $\mu_\tau,\tilde\mu_\tau$ are the unique minimizers of (15) for μ and $\tilde\mu$ respectively. It is interesting to underline that:

- Once monotonicity for the single time step is proven, a maximum principle for weak solutions of heat equation can be proved as a direct consequence, see Corollary 3.9.
- Although our strategy is not new (for instance, it has been used in the context of the classical transportation problem in [10, 1] to prove a maximum principle), the fact of having no mass constraints makes it more efficient, and the properties of minimizers that we are able to deduce are in some sense stronger.
- The argument that we are going to use holds in much more general situations, see Remark 3.10. (This in not the case when one deals with the classical transportation problem, where the fact that the cost function satisfies $c(x,x)\leq c(x,y)$ for all $x,y\in\Omega$ plays an important role, see [1, 7].)

The proof of the monotonicity relies on a set of inequalities valid for each minimizer of (15). In the next proposition we are going to assume that $\mu=\rho\mathcal{L}^d|_\Omega\in\mathcal{M}_2(\Omega)$ is an absolutely continuous measure and that $\tau>0$ is a fixed time step. Also, we will denote by $\mu_\tau=\rho_\tau\mathcal{L}^d|_\Omega$ the unique minimizer of (15) (which is absolutely continuous by Proposition 3.6), by $\boldsymbol{\gamma}$ the unique optimal plan for (ρ,ρ_τ), by T the map which induces $\boldsymbol{\gamma}_\Omega^{\overline{\Omega}}$, and by S the map which induces $\boldsymbol{\gamma}_{\overline{\Omega}}^{\Omega}$ seen from ρ_τ (see Corollary 2.5).

Proposition 3.7 *With the notation above, the following inequalities hold:*

- *Let $y_1,y_2\in\Omega$ be Lebesgue points for ρ_τ, and assume that y_1 is also a Lebesgue point for S. Then*

$$\log(\rho_\tau(y_1))+\frac{|y_1-S(y_1)|^2}{2\tau}\leq\log(\rho_\tau(y_2))+\frac{|y_2-S(y_1)|^2}{2\tau}.\tag{24}$$

- *Let $x \in \Omega$ be a Lebesgue point for both ρ and T, and assume that $T(x) \in \partial\Omega$. Assume further that $y \in \Omega$ is a Lebesgue point for ρ_τ. Then*

$$\frac{|x - T(x)|^2}{2\tau} \leq \log(\rho_\tau(y)) + \frac{|x-y|^2}{2\tau}. \tag{25}$$

- *Let $y_1 \in \Omega$ be a Lebesgue point for ρ_τ. Then, for any $y_2 \in \partial\Omega$, we have*

$$\log(\rho_\tau(y_1)) + \frac{|y_1 - S(y_1)|^2}{2\tau} \leq \frac{|y_2 - S(y_1)|^2}{2\tau}. \tag{26}$$

- *Let $y \in \Omega$ be a Lebesgue point for ρ_τ. Then*

$$-\frac{d^2(y,\partial\Omega)}{2\tau} \leq \log(\rho_\tau(y)) \leq \frac{3d(y,\partial\Omega)\,\mathrm{Diam}(\Omega)}{2\tau}. \tag{27}$$

  (This is the key inequality that shows that minimizers have trace 1 *on $\partial\Omega$, see* (17)*.)*

- *Let $y \in \Omega$ be a Lebesgue point for both ρ_τ and S, and assume that $S(y) \in \partial\Omega$. Then*

$$\log(\rho_\tau(y)) + \frac{d^2(y,\partial\Omega)}{2\tau} = 0. \tag{28}$$

Proof. **- Heuristic argument.** We start with (24). Consider a point $y_1 \in \Omega$ and observe that the mass $\rho_\tau(y_1)$ comes from $S(y_1)$. (It does not matter whether $S(y_1) \in \Omega$ or $S(y_1) \in \partial\Omega$.) We now make a small perturbation of ρ_1 in the following way: we pick a small amount of mass from $S(y_1)$ and, instead than moving it to y_1, we move it to y_2. In terms of entropy, we are earning $\log(\rho_1(S(y_1)))$ because of the less mass in $S(y_1)$ and paying $\log(\rho_1(y_2))$ because of the greater amount of mass at y_2. In terms of the transportation cost, we are earning $\frac{|y_1-S(y_1)|^2}{2\tau}$ and paying $\frac{|y_2-S(y_1)|^2}{2\tau}$. But since ρ_1 is a minimizer of (15), what we are earning must be less or equal to what we are paying, and we get (24).

Inequality (25) is analogous: here we are just considering those points x which are sent to the boundary by T. In this case, if we decide to send some small mass at x onto a point $y \in \Omega$, we are not earning in terms of entropy but just paying $\log(\rho_\tau(y))$, while in terms of cost we are earning $\frac{|x-T(x)|^2}{2\tau}$ and paying $\frac{|x-y|^2}{2\tau}$.

To prove inequality (27) we argue as follows. Consider first a point $y \in \Omega$ and perturb ρ_τ by picking some small mass from one of the nearest point to y on $\partial\Omega$ and putting it onto y. In this way we pay $\log(\rho_\tau(y))$ in terms of entropy and $\frac{d^2(y,\partial\Omega)}{2\tau}$ in terms of cost, so that by minimality we get

$$\frac{d^2(y,\partial\Omega)}{2\tau} \geq -\log(\rho_\tau(y)). \tag{29}$$

The other part of the inequality comes by taking some small mass at y and putting it on one of the nearest point to y on $\partial\Omega$, say $P(y)$: we earn $\log(\rho_\tau(y))$ in terms of entropy and $\frac{|S(y)-y|^2}{2\tau}$ in terms of cost, and we are paying $\frac{|S(y)-P(y)|^2}{2\tau}$ more because of the new cost. This gives

$$\log(\rho_\tau(y)) + \frac{|S(y)-y|^2}{2\tau} \leq \frac{|S(y)-P(y)|^2}{2\tau} \leq \frac{|S(y)-y|^2 + 3|y-P(y)|\,\mathrm{Diam}(\Omega)}{2\tau},$$

from which the claim follows.

The proof of (28) is a sort of converse of (29). Indeed, as $S(y) \in \partial\Omega$ we know that the mass of y is coming from the boundary. Hence we can perturb ρ_τ by taking a bit less of mass from the boundary, so that there is a bit less of mass in y. In this way we obtain the opposite of (27), and equality holds.

- Rigorous proof. We will prove rigorously only (24), the proof of the other inequalities being analogous.

Fix $y_1, y_2 \in \Omega$, and two positive real numbers $r, \varepsilon > 0$, with r small enough so that $B_r(y_1) \cup B_r(y_2) \subset \Omega$. Let $\mathrm{Tr} : \mathbb{R}^d \to \mathbb{R}^d$ be the map defined by $\mathrm{Tr}(y) := y - y_1 + y_2$, and let $\boldsymbol{\gamma} \in \textsc{Opt}(\rho, \rho_\tau)$ be the unique optimal plan. Define the plan $\boldsymbol{\gamma}^{r,\varepsilon}$ as

$$\boldsymbol{\gamma}^{r,\varepsilon} := \boldsymbol{\gamma}_{\overline{\Omega}}^{B_r(y_1)^c} + (1-\varepsilon)\boldsymbol{\gamma}_{\overline{\Omega}}^{B_r(y_1)} + \varepsilon\Big((\pi^1, \mathrm{Tr})_\# \boldsymbol{\gamma}_{\overline{\Omega}}^{B_r(y_1)}\Big),$$

and set

$$\mu_\tau^{r,\varepsilon} := \pi^2_\# \boldsymbol{\gamma}^{r,\varepsilon}.$$

Observe that $\pi^1_\# \boldsymbol{\gamma}^{r,\varepsilon} = \pi^1_\# \boldsymbol{\gamma}$, $\boldsymbol{\gamma}^{r,\varepsilon} \in \textsc{Adm}(\rho, \mu_1^{r,\varepsilon})$, and $\mu_\tau^{r,\varepsilon} = \rho_\tau^{r,\varepsilon} \mathcal{L}^d$, with

$$\rho_\tau^{r,\varepsilon}(y) = \begin{cases} \rho_\tau(y) & \text{if } y \in B_r(y_1)^c \cap B_r(y_2)^c, \\ (1-\varepsilon)\rho_\tau(y) & \text{if } y \in B_r(y_1), \\ \rho_\tau(y) + \varepsilon\rho_\tau(y - y_2 + y_1) & \text{if } y \in B_r(y_2). \end{cases}$$

From the minimality of ρ_τ we get

$$\int_\Omega e(\rho_\tau)\, dx + \frac{1}{2\tau} C(\boldsymbol{\gamma}) \le \int_\Omega e(\rho_\tau^{r,\varepsilon})\, dx + \frac{1}{2\tau} C(\boldsymbol{\gamma}^{r,\varepsilon}).$$

Hence

$$\begin{aligned} \int_{B_r(y_1)\cup B_r(y_2)} & e(\rho_\tau(y))\, dy + \frac{1}{2\tau} \int_{B_r(y_1)\cup B_r(y_2)} |y - S(y)|^2 \rho_\tau(y)\, dy \\ &\le \int_{B_r(y_1)} e((1-\varepsilon)\rho_\tau(y))\, dy + \frac{(1-\varepsilon)}{2\tau} \int_{B_r(y_1)} |y - S(y)|^2 \rho_1(y)\, dy \\ &\quad + \int_{B_r(y_2)} e(\rho_\tau(y) + \varepsilon\rho_\tau(y - y_1 + y_2))\, dy \\ &\quad + \frac{1}{2\tau} \int_{B_r(y_2)} |y - S(y)|^2 (\rho_\tau(y) + \varepsilon\rho_\tau(y - y_1 + y_2))\, dy, \end{aligned}$$

which we write as

$$\begin{aligned} \int_{B_r(y_1)} & \Big(e(\rho_\tau(y)) - e((1-\varepsilon)\rho_\tau(y)) + \frac{\varepsilon}{2\tau} |y - S(y)|^2 \rho_\tau(y) \Big) dy \\ &\le \int_{B_r(y_2)} \Big(e(\rho_\tau(y) + \varepsilon\rho_\tau(y - y_2 + y_1)) - e(\rho_\tau(y)) \\ &\qquad + \frac{\varepsilon}{2\tau} |y - S(y)|^2 \rho_\tau(y - y_2 + y_1) \Big) dy. \end{aligned}$$

Dividing by ε and letting $\varepsilon \downarrow 0$ we obtain

$$\begin{aligned} \int_{B_r(y_1)} & \left(e'(\rho_\tau(y)) + \frac{1}{2\tau} |y - S(y)|^2 \right) \rho_\tau(y)\, dy \\ &\le \int_{B_r(y_2)} \left(e'(\rho_1(y)) + \frac{1}{2\tau} |y - S(y)|^2 \right) \rho_\tau(y - y_2 + y_1)\, dy. \end{aligned}$$

Now, since y_1, y_2 are both Lebesgue points of ρ_τ, and y_1 is also a Lebesgue point of S, dividing both sides by $\mathcal{L}^d(B_r(0))$ and letting $r \downarrow 0$ we obtain (24). □

Proposition 3.8 (Monotonicity) *Let $\mu \geq \tilde{\mu} \in \mathcal{M}_2(\Omega)$, $\tau > 0$, and $\mu_\tau, \tilde{\mu}_\tau$ the minima of the minimizing problem* (15)*. Then $\mu_\tau \geq \tilde{\mu}_\tau$.*

Proof. From the uniqueness part of Proposition 3.6, it follows easily that the map $\mu \mapsto \mu_\tau$ is continuous with respect to the weak topology. Therefore we can assume by approximation that both μ and ν are absolutely continuous, say $\mu = \rho\mathcal{L}^d$ and $\tilde{\mu} = \tilde{\rho}\mathcal{L}^d$. Also, recall that by Proposition 3.6(i) both μ_τ and ν_τ are absolutely continuous, say $\mu_\tau = \rho_\tau\mathcal{L}^d$ and $\tilde{\mu}_\tau = \tilde{\rho}_\tau\mathcal{L}^d$. Let $\boldsymbol{\gamma} \in \mathrm{Opt}(\rho, \rho_\tau)$ and $\tilde{\boldsymbol{\gamma}} \in \mathrm{Opt}(\tilde{\rho}, \tilde{\rho}_\tau)$, and let $T, \tilde{T}$ be the maps which induce $\boldsymbol{\gamma}_\Omega^{\overline{\Omega}}$ and $\tilde{\boldsymbol{\gamma}}_\Omega^{\overline{\Omega}}$ respectively.

Argue by contradiction, and assume that $A := \{\tilde{\rho}_\tau > \rho_\tau\} \subset \Omega$ satisfies $\tilde{\rho}_\tau(A) > 0$. Two cases arise: either $\tilde{\boldsymbol{\gamma}}_{\overline{\Omega}}^A$ is concentrated on $\Omega \times A$ or it is not, i.e. either the mass of $\tilde{\rho}_\tau$ in A comes entirely from Ω or it is partly taken from the boundary.

Case 1: the mass of $\tilde{\rho}_\tau$ in A comes entirely from Ω. Let $B := \tilde{T}^{-1}(A)$, and observe that $\tilde{\mu}(B) = \tilde{\mu}_\tau(A)$. Let $C \subset B$ be the set of points $x \in B$ such that $T(x) \notin A$. We remark that $\mu(C) > 0$, as otherwise we would have

$$\mu_\tau(A) \geq \mu_\tau(T(B)) = \mu(T^{-1}(T(B))) \geq \mu(B) \geq \tilde{\mu}(B) = \tilde{\mu}_\tau(A),$$

which contradicts the definition of A. Define

$$C_1 := \Big\{x \in C \;:\; T(x) \in \Omega\Big\}, \qquad C_2 := \Big\{x \in C \;:\; T(x) \in \partial\Omega\Big\}.$$

Since $C = C_1 \cup C_2$, either $\mu(C_1) > 0$ or $\mu(C_2) > 0$. Suppose we are in the first case. Then, as both $T|_{C_1}$ and $\tilde{T}|_{C_1}$ map subsets of the support of $\tilde{\rho}$ of positive Lebesgue measure into sets of positive Lebesgue measure, we can find $x \in C_1$ a Lebesgue point for both T and $\tilde{T}$ such that $T(x)$ and $\tilde{T}(x)$ are Lebesgue points for both ρ_τ and $\tilde{\rho}_\tau$. With this choice of x we apply (24) with $y_1 = T(x)$ and $y_2 = \tilde{T}(x)$ to get

$$\log\left(\rho_\tau(T(x))\right) + \frac{|x - T(x)|^2}{2\tau} \leq \log\left(\rho_\tau(\tilde{T}(x))\right) + \frac{|x - \tilde{T}(x)|^2}{2\tau}$$

Similarly, using (24) for $\tilde{\rho}_\tau$ with $y_1 = \tilde{T}(x)$ and $y_2 = T(x)$ we obtain

$$\log\left(\tilde{\rho}_\tau(\tilde{T}(x))\right) + \frac{|x - \tilde{T}(x)|^2}{2\tau} \leq \log\left(\tilde{\rho}_\tau(T(x))\right) + \frac{|x - T(x)|^2}{2\tau}$$

Adding up the last two inequalities we get

$$\log\left(\rho_\tau(T(x))\right) + \log\left(\tilde{\rho}_\tau(\tilde{T}(x))\right) \leq \log\left(\rho_\tau(\tilde{T}(x))\right) + \log\left(\tilde{\rho}_\tau(T(x))\right),$$

which contradicts definition of C_1 and the choice of x, as we have

$$\begin{aligned} T(x) \notin A \quad &\Rightarrow \quad \rho_\tau(T(x)) \geq \tilde{\rho}_\tau(T(x)) \quad &&\Rightarrow \quad \log\left(\rho_\tau(T(x))\right) \geq \log\left(\tilde{\rho}_\tau(T(x))\right), \\ \tilde{T}(x) \in A \quad &\Rightarrow \quad \tilde{\rho}_\tau(\tilde{T}(x)) > \rho_\tau(\tilde{T}(x)) \quad &&\Rightarrow \quad \log\left(\tilde{\rho}_\tau(\tilde{T}(x))\right) > \log\left(\rho_\tau(\tilde{T}(x))\right). \end{aligned}$$

It remains to exclude the possibility $\mu(C_2) > 0$. Fix $x \in C_2$ a Lebesgue point for both T and $\tilde{T}$, such that $\tilde{T}(x)$ is a Lebesgue point for both ρ_τ and $\tilde{\rho}_\tau$. We apply (25) with $y = \tilde{T}(x)$ to obtain

$$\frac{|x - T(x)|^2}{2\tau} \leq \log\left(\rho_\tau(\tilde{T}(x))\right) + \frac{|x - \tilde{T}(x)|^2}{2\tau}.$$

Now, we use (26) for $\tilde{\rho}_\tau$ with $y_1 = \tilde{T}(x)$, $S(y_1) = x$, and $y_2 = T(x)$, to get

$$\log\left(\tilde{\rho}_\tau(\tilde{T}(x))\right) + \frac{|x - \tilde{T}(x)|^2}{2\tau} \leq \frac{|x - T(x)|^2}{2\tau}.$$

Since $\tilde{T}(x) \in A$ we have $\rho_\tau(\tilde{T}(x)) < \tilde{\rho}_\tau(\tilde{T}(x))$, which together with the above inequalities implies

$$\begin{aligned}\frac{|x-T(x)|^2}{2\tau} &\leq \log\left(\rho_\tau(\tilde{T}(x))\right) + \frac{|x-\tilde{T}(x)|^2}{2\tau} \\ &< \log\left(\tilde{\rho}_\tau(\tilde{T}(x))\right) + \frac{|x-\tilde{T}(x)|^2}{2\tau} \leq \frac{|x-T(x)|^2}{2\tau},\end{aligned}$$

again a contradiction.

Case 2: the mass of $\tilde{\rho}_\tau$ in A comes partly from $\partial\Omega$. Let $\tilde{S}$ be the map which induces $\tilde{\boldsymbol{\gamma}}_{\overline{\Omega}}^{\Omega}$ seen from $\tilde{\rho}_\tau$, and let $D \subset A$ be the set of points y such that the mass $\tilde{\rho}_\tau(y)$ comes from the boundary, i.e. $D := \{y \in A \,:\, \tilde{S}(y) \in \partial\Omega\}$. Fix $y \in D$ a Lebesgue point for ρ_τ, $\tilde{\rho}_\tau$, and $\tilde{S}$. Thanks to (27) we have

$$\log\left(\rho_\tau(y)\right) + \frac{d^2(y, \partial\Omega)}{2\tau} \geq 0,$$

while applying (28) with $\tilde{\rho}_\tau$ (recall that $\tilde{S}(y) \in \partial\Omega$) we obtain

$$\log\left(\tilde{\rho}_\tau(y)\right) + \frac{d^2(y, \partial\Omega)}{2\tau} = 0.$$

But this is absurd as $y \in D \subset A$. □

Thanks to Proposition 3.8, we immediately obtain the following:

Corollary 3.9 (Comparison principle) *Let $\mu_0, \nu_0 \in \mathcal{M}_2(\Omega)$, assume that $\mu_0 \geq \tilde{\mu}_0$, and let $\tau_k \downarrow 0$ be a sequence of time steps such that the corresponding discrete solutions $\mu^{\tau_k}(t), \tilde{\mu}^{\tau_k}(t)$ associated to $\mu_0, \tilde{\mu}_0$ respectively converge to two solutions $\mu_t, \tilde{\mu}_t$ of the heat equation, as described in Theorem 3.5. Then $\mu_t \geq \tilde{\mu}_t$ for all $t \in [0, +\infty)$.*

Remark 3.10 (Different energies and costs) The proof of the above theorem relies entirely on the set of inequalities proved in Proposition 3.7. Here we want to point out that a corresponding version of such inequalities is true in more general cases.

Indeed, let $c : \overline{\Omega} \times \overline{\Omega} \to \mathbb{R} \cup \{+\infty\}$ be a continuous cost function, and define the *Cost of transport* as the infimum of

$$\int_{\overline{\Omega}\times\overline{\Omega}} c(x,y)\, d\boldsymbol{\gamma}(x,y),$$

among all $\boldsymbol{\gamma} \in \textsc{Adm}(\mu_0, \mu)$. Let $e : [0, +\infty) \to \mathbb{R}$ be a superlinear convex function. Then, a minimizer ρ_1 for

$$\rho \quad \mapsto \quad \int_\Omega e(\rho(x))dx + \text{ Cost of transport } (\rho, \mu_0),$$

always exists, and arguing as in the proof of Proposition 3.7 it is possible to check that for ρ_1-a.e. y_1, y_2, and any x such that (x, y_1) belongs to the support of an optimal plan from μ to ρ_1, we have

$$e'_-(\rho_1(y_1)) + c(x, y_1) \leq e'_+(\rho_1(y_2)) + c(x, y_2),$$

and similarly for the other inequalities. Then the convexity of e implies that

$$e'_-(z_1) \leq e'_+(z_1) \leq e'_-(z_2) \qquad \forall\, 0 \leq z_1 < z_2,$$

and the proof of the monotonicity goes on like in the case we analyzed. In particular it is interesting to observe that the choice $c(x,y) = |x-y|^2$ in this setting does not play any role.

4 Comments and open problems

- All our results could be extended to more general cost function and more general entropies. For instance, by considering $c(x, y) = |x - y|^p$ with $p > 1$, and $e(z) = z\log(z) - \alpha z$ with $\alpha \in \mathbb{R}$, one can construct a weak solution of

$$\begin{cases} \frac{d}{dt}\rho(t) &= \Delta_p \rho(t), \\ \rho(0) &= \rho_0, \end{cases}$$

(where $\Delta_p\rho$ denotes the p-Laplacian of ρ), subject to the Dirichlet boundary condition

$$\rho(t)_{|\partial\Omega} = e^{\alpha-1}, \qquad \text{for a.e. } t \geq 0.$$

- It is interesting to observe that our approach - as in the classical Wasserstein one - allows to introduce a drift term in the diffusion: by considering the entropy $\int_\Omega [\rho\log\rho - \rho + V\rho]\,dx$ for some smooth function $V : \overline{\Omega} \to \mathbb{R}$ we obtain a weak solution of

$$\begin{cases} \frac{d}{dt}\rho(t) &= \Delta\rho(t) + \mathrm{div}(\rho\nabla V) \\ \rho(0) &= \rho_0, \end{cases}$$

subject to the Dirichlet boundary condition

$$\rho(t)_{|\partial\Omega} = e^{-V}, \qquad \text{for a.e. } t \geq 0.$$

- A standard approach for constructing weak solutions to the heat equation with Dirichlet boundary condition equal to a function ϕ consists viewing the equation as the gradient flow of $\int_\Omega |\nabla\rho|^2$ on the set of functions $\rho \in H^1_\phi(\Omega) := \{\rho \in H^1(\Omega) : \mathrm{trace}(\rho) = \phi\}$, with respect to the L^2-norm. However, although this approach allows to treat general boundary conditions, it cannot be used to add a drift term: given $F = F(x, u, p) : \Omega \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$, the gradient flow of a functional of the form $\int_\Omega F(x, \rho, \nabla\rho)\,dx$ is given by

$$\frac{d}{dt}\rho(t) = \mathrm{div}_x\Big(F_p(x, \rho(t), \nabla\rho(t))\Big) - F_u(x, \rho(t), \nabla\rho(t)),$$

and it is easy to check by a direct computation that there is no choice of F which allows to obtain $\Delta\rho(t) - \mathrm{div}(\rho\nabla V)$ as the right-hand side.

- Although it is possible to prove uniqueness of solution by purely PDE methods, it is not clear to us if one can use a transportation approach to prove this result. In particular it is not clear if, as in the classical Wassestein case, $t \mapsto Wb_2(\rho_t, \tilde{\rho}_t)$ is decreasing along gradient flows of the entropy $\int_\Omega \rho\log(\rho)\,dx$.

- In Proposition 2.11 we only proved an upper bound for the derivative of Wb_2. We conjecture that the following formula should be true: let $t \mapsto \mu_t$ an absolutely continuous curve with values in $(\mathcal{M}_+(\Omega), Wb_2)$. Then:

  (a) There exists a velocity field $\boldsymbol{w}_t \in L^1_{loc}([0, +\infty), L^2(\Omega, \mu_t))$ such that

$$\frac{d}{dt}\mu_t + \mathrm{div}(\boldsymbol{w}_t\mu_t) = 0$$

  in $[0, +\infty) \times \Omega$. (Observe that, since by definition the continuity equation can be tested only against smooth functions with support inside $[0, +\infty) \times \Omega$, the mass of μ_t is not necessarily constant.)

(b) Given $\mu \in \mathcal{M}_2(\Omega)$, for a.e. $t \geq 0$ we have

$$\frac{d}{dt} Wb_2^2(\mu_t, \mu) = -2 \int_{\Omega\times\Omega} \langle \boldsymbol{w}_t, y - x\rangle \, d\boldsymbol{\gamma}(x, y),$$

where $\boldsymbol{\gamma}$ is any optimal plan between μ_t and μ.

- For any Borel subset $\Gamma \subset \partial\Omega$, one can define a variant of our distance: given two non-negative measures μ and ν on Ω, we set

$$Wb_2^\Gamma(\mu, \nu) := \inf_{\pi^1_\# \boldsymbol{\gamma}|_{\overline{\Omega}\setminus\Gamma} = \mu, \pi^2_\# \boldsymbol{\gamma}|_{\overline{\Omega}\setminus\Gamma} = \nu} \int_{\overline{\Omega}\times\overline{\Omega}} |x - y|^2 d\boldsymbol{\gamma}(x, y).$$

The difference between this distance and the Wb_2-distance considered in this paper is that now only Γ can be used as an infinite reserve of mass, and not the whole $\partial\Omega$. This distance may be useful to study evolution equations where one what to impose Dirichlet boundary conditions on Γ and Neumann conditions on $\partial\Omega \setminus \Gamma$, at least when Ω is convex. Moreover, it is likely that this distance may be used to study crowds motions, where some people want to escape from Ω and the only available exit in on Γ (see for instance [9], where the authors use a W_2-gradient flow approach to model this kind of phenomena).

References

[1] M. Agueh, *Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory*, Adv. Differential Equations, 10 (2005), no. 3, 309–360.

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in spaces of probability measures*, Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[3] Y. Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs*, (French) C. R. Acad. Sci. Paris Sér. I Math., 305 (1987), no. 19, 805–808.

[4] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math., 44 (1991), no. 4, 375–417.

[5] J. Dolbeault, B. Nazaret, G. Savaré, *A new class of transport distances between measures*, Calc. Var. Partial Differential Equations, 34 (2009), no. 2, 193–231.

[6] A. Figalli, *The optimal partial transport problem*, Arch. Ration. Mech. Anal., to appear.

[7] A. Figalli, W. Gangbo, T. Yolku, *A variational method for class of parabolic PDEs*, preprint, 2009.

[8] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29 (1998), no. 1, 1–17.

[9] B. Maury, A. Roudneff-Chupin, F. Santambrogio, *A macroscopic Crowd Motion Model of the gradient-flow type*, Math. Models Methods Appl. Sci., to appear.

[10] F. Otto, *Dynamics of labyrinthine pattern formation in magnetic fluids: a mean-field theory*, Arch. Rational Mech. Anal., 141 (1998), no. 1, 63–103.

[11] F. Otto, *Evolution of microstructure in unstable porous media flow: a relaxational approach*, Comm. Pure Appl. Math., 52 (1999), no. 7, 873–915.

[12] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation,* Comm. Partial Differential Equations, 26 (2001), no. 1-2, 101–174

[13] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, 58, American Mathematical Society, Providence, RI, 2003.

[14] C. Villani, *Optimal transport, old and new*, Grundlehren des mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 338, Springer-Verlag, Berlin-New York, 2009.