J Glob Optim (2007) 38:129–141
DOI 10.1007/s10898-006-9087-z

ORIGINAL ARTICLE

Transport and concentration problems with interaction effects

Filippo Santambrogio

Received: 13 January 2006 / Accepted: 19 August 2006 / Published online: 6 October 2006

Abstract After presenting an overview about variational problems on probability measures for functionals involving transport costs and extra terms encouraging or discouraging concentration, we look for optimality conditions, regularity properties and explicit computations in the case where Wasserstein distances and interaction energies are considered.

Keywords Mass transportation · Nonconvex optimization · Monge-Ampère equation · Interaction energy · Regularity

1 Introduction

As a starting point for the paper, we present a short overview of possible variational problems involving transport costs between distributions of mass and their concentrations. The general problem we are interested in is

$$\min_{\mu,\nu\in\mathcal{P}(\Omega)} \mathfrak{F}(\mu,\nu) := T(\mu,\nu) + F(\mu) + G(\nu),$$

where the functional T quantifies in some way the distance between the two probability measures μ and ν according to a transport cost criterion, and F and G are functionals over the space $\mathcal{P}(\Omega)$ (the space of probability measures over a domain Ω) with opposite behaviour: the first favours spread measures and penalizes concentration while the latter, on the other hand, favours concentrated measures. Obviously there are several other sub-problems that may be of interest, for instance the minimizations of the two separate functionals

$$\mathfrak{F}_\nu(\mu) := T(\mu,\nu) + F(\mu) \quad \text{and} \quad \mathfrak{F}^\mu(\nu) = T(\mu,\nu) + G(\nu),$$

F. Santambrogio
Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy
e-mail: f.santambrogio@sns.it

where each time one of the variables is frozen. Also imposing constraints like $F(\mu) \leq H$, $G(\nu) \leq L \ldots$ instead of adding penalizations in the functionals may sometimes be considered (and this is in fact the same as adding penalizations through some $0/+\infty$ functions).

These minimization problems are likely to appear in several phenomena both in nature and in decision science. For instance in [3,5,8] these variational problems have been proposed for urban planning models, where the spread measure μ stands for inhabitants, the concentrated measure ν for services and they have to be close in a transportation distance sense. We may say that the idea of considering the distribution of population and firms in a city and the fact that population prefers to be spread and firms to be next to each other has been brought into the mathematical community, in the optimal transportation framework, through [4] and its authors. On the other hand, a possible choice of the functional $\mathfrak{F}_\nu$ has been proposed recently as a model for the formation of a certain kind of leaves or in general for ramified biological structures: if $\nu = \delta_0$ represents the point where the nutrient arrives to the leaf, then the shape of the leaf is modelled to optimize the quantity of light it receives from the sun taking also into account the effective transport cost for bringing the nutrient all over its shape.

We present here some choices for the functionals T and G. The choice of F is in fact easier since a very good class of concentration-penalizing functionals is given by local convex functionals over measures, for instance

$$F(\mu) = \begin{cases} \int\limits_\Omega f(u)\, d\mathcal{L} & \text{if } \mu = u \cdot \mathcal{L} \\ +\infty & \text{otherwise,} \end{cases} \tag{1.1}$$

for any convex function f with $f(0) = 0$ which is superlinear at infinity. For these functionals we refer to [1]. Here $\mathcal{L}$ is a reference measure that may be chosen as the Lebesgue measure $\mathcal{L}^d$ if $\Omega \subset \mathbb{R}^d$. By Jensen inequality, spread measures with constant density are optimal for these functionals.

Possible choices of T are the following:

- terms involving Monge-Kantorovich optimal transport cost, as in Wasserstein distances $T(\mu, \nu) = W_p^p(\mu, \nu)$ (*Wasserstein*);
- terms taking into account traffic congestion effects, as in [5] where $T(\mu, \nu) = ||\mu - \nu||^2_{X'}$ for a vector space $X \subset H^1(\Omega)$ and this choice is detailed and justified (*congestion*);
- terms reflecting the natural ramified structure of a transportation network (*branching*) as in [10], where a new distance on probability measures is introduced according to this criterion, or in [6], where the same quantity is introduced in a different way, in the case $\mu = \delta_{x_0}$.

This last possibility is the most suitable for model involving natural branching structures like leaves, while the first two seem to be quite natural in urban planning.

For the functional G, before presenting a list of examples, we give a possible definition of the concept of concentration-preferring.

Definition 1 We say that $G\colon \mathcal{P}(\Omega) \to \mathbb{R}$ is a *concentration-preferring* functional if it holds $G(t_\sharp \nu) \leq G(\nu)$ for any measure $\nu \in \mathcal{P}(\Omega)$ and any 1-Lipschitz continuous map $t\colon \Omega \to \Omega$.

It is easy to show that any G with this property is minimized by any measure δ_{x_0}, with $x_0 \in \Omega$. We list here some functionals satisfying this definition:

- $G(\nu) = \sharp\,\mathrm{spt}(\nu)$ (*atomic*), as in location problems, where the corresponding T is usually of Wasserstein type;
- (*subadditive*, see [1])

$$G(\nu) = \begin{cases} \sum\limits_{k\in\mathbb{N}} g(a_k) & \text{if } \nu = \sum\limits_{k\in\mathbb{N}} a_k\delta_{x_k} \\ +\infty & \text{otherwise} \end{cases}$$

  for a subadditive function g with $g(0) = 0$ and $g'(0) = +\infty$ (if $g = 1$ on $(0, +\infty)$ and $g(0) = 0$ we recover the previous case);
- $G(\nu) = \inf\left\{\mathcal{H}^1(\Sigma)\,\middle|\,\mathrm{spt}(\nu) \subset \Sigma,\ \Sigma \text{ compact and connected}\right\}$ as in the irrigation problem, [2], where $T = W_1$ and a constraint on G is considered (*length*);
- $G(\nu) = \int_{\Omega\times\Omega} w(d(x,y))\nu(dx)\nu(dy)$ for an increasing function w (*interaction*).

Actually the two first cases are functionals which decrease under the effect of any map t and not only under 1-Lipschitz ones. The first one has been presented separately and not as a particular case of the second because of its large presence in literature. The last choice for G is a well-known functional on probability measures which represents the interaction energy (or cost) between the particles composing ν. It has been first studied in a transportation framework by McCann in [7], where displacement convexity results are given, with the aim of showing uniqueness results for variational problems.

In [3] and [5] two combinations of these functional have been studied for urban planning purposes: the *Wasserstein + subadditive* and the *congestion + interaction* cases, respectively. The *congestion + subadditive* case has been excluded in [5] since it leads to a trivially infinite functional, and so in this paper we analyze the remaining one, i.e. the *Wasserstein + interaction* case. Many ideas are taken from [5], up to the fact that elliptic regularity is replaced by considerations on Monge–Ampère equation. Moreover, some extra devices are performed and a careful use of Monge–Kantorovich theory is needed.

The main purpose of the paper is giving necessary optimality conditions and trying to identify the global minimizers. Since most optimality conditions are obtained by small perturbations, several statements are valid for local minima as well. Anyway in this paper we exploit an approximation process which does not provide, a priori, information on local minima which are not globally minimizing. At the end of the paper a section is devoted to a specific example where explicit solutions may be obtained.

2 Preliminaries on optimal transportation

In this section we recall some tools and definitions that are well known in the optimal transport community and that we will need in the sequel. Our main reference is [9].

Definition 2 Given two probability measures μ and ν on a space Ω and an l.s.c. cost function $c\colon \Omega\times\Omega \to [0, +\infty]$ we consider the problem

$$(K) \quad \min\left\{\int\limits_{\Omega\times\Omega} c\,d\gamma \,\middle|\, \gamma \in \mathcal{P}(\Omega\times\Omega),\ (\pi_1)_\sharp\gamma = \mu,\ (\pi_2)_\sharp\gamma = \nu\right\}, \tag{2.1}$$

and the minimizers for this problem are called *optimal transport plans* between μ and ν. Should γ be of the form $(id \times t)_\sharp \mu$ for a measurable map $t\colon \Omega \to \Omega$, the map t would be called *optimal transport map* from μ to ν.

An important tool will be duality theory and to introduce it we need in particular the notion of c-transform (a kind of generalization of the well-known Legendre transform).

Definition 3 Given a function $\chi\colon \Omega \to \overline{\mathbb{R}}$ we define its *c-transform* (or c-conjugate function) by

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

Moreover, we say that a function ψ is *c-concave* if there exists χ with $\psi = \chi^c$ and we denote by $\Psi_c(\Omega)$ the set of c-concave functions.

It is well known a duality result stating the following equality (see Theorem 1 together with the following Remark on c-concavity in [9]), also known as Duality Formula:

$$\min(K) = \sup_{\psi \in \Psi_c(\Omega)} \int_\Omega \psi \, d\mu + \int_\Omega \psi^c \, d\nu. \tag{2.2}$$

Definition 4 The functions ψ realizing the maximum in (2.2) are called *Kantorovich potentials* for the transport from μ to ν.

Since we will use $c(x, y) = \frac{1}{2}|x - y|^2$, let us denote by $\Psi_2(\Omega)$ the set of c-concave functions with respect to this quadratic cost. It is not difficult to check that

$$\psi \in \Psi_2(\Omega) \Rightarrow \; x \mapsto \frac{x^2}{2} - \psi \quad \text{is a convex function on } \mathbb{R}^d \text{ restricted to } \Omega.$$

Notice that on a bouded Ω with diameter D any $\psi \in \Psi_2(\Omega)$ is in fact $2D$-Lipschitz continuous. We summarize here some useful results for the quadratic case $c(x, y) = \frac{1}{2}|x - y|^2$, which can be found in Theorem 15 of [9] or throughout its proof.

Theorem 2.1 *Given μ and ν probability measures on a connected $\Omega \subset \mathbb{R}^d$ there exists unique an optimal transport plan γ and it is of the form $(id \times t)_\sharp \mu$, provided μ is absolutely continuous. Moreover, there exists also at least a Kantorovich potential ψ, and the gradient $\nabla\psi$ is uniquely determined μ–a.e. (in particular ψ is unique up to additive constants, provided the density of μ is positive a.e. on Ω). The optimal transport map t and the potential ψ are linked by $t(x) = x - \nabla\psi(x)$ and so t is the gradient of a convex function. Moreover it holds $\psi(x) + \psi^c(t(x)) = c(x, t(x))$ for μ–a.e. x.*

Starting from the values of the problem (K) in (2.1) we can define a set of distances over $\mathcal{P}(\Omega)$. For any $p \geq 1$ set

$$W_p(\mu, \nu) = \left(\min(K) \text{ with } c(x, y) = |x - y|^p\right)^{1/p},$$

and for simplicity we will restrict our analysis to the case $p = 2$. We recall that it holds (see also Theorem 13 in [9]), by the Duality Formula,

$$\frac{1}{2} W_2^2(\mu, \nu) = \sup_{\psi \in \Psi_2(\Omega)} \int_\Omega \psi \, d\nu + \int_\Omega \psi^c \, d\mu. \tag{2.3}$$

The following result on Wasserstein distances can be obtained by putting together Theorem 85 and Theorem 87 of [9].

Theorem 2.2 *If Ω is compact, for any $p \geq 1$ the function W_p is in fact a distance over $\mathcal{P}(\Omega)$ and the convergence with respect to this distance is equivalent to the weak convergence of probability measures. In particular any functional $\mu \mapsto W_p(\mu, \nu)$ is continuous with respect to weak topology.*

The next step of our analysis is concerned with some regularity properties of t and ψ (the optimal transport map and the Kantorovich potential, respectively) and their relations with the densities of μ and ν. It is easy, just by a change-of-variables formula, to transform, in the case of regular maps and densities, the equality $\nu = t_\sharp\mu$ into the PDE $v(t(x)) = u(x)/|Jt|(x)$, where u and v are the densities of μ and ν, respectively, and J denotes the determinant of the Jacobian matrix. Recalling that we may write $t = \nabla\phi$ with ϕ convex, we get the Monge–Ampère equation

$$M\phi = \frac{u}{v(\nabla\phi)}, \tag{2.4}$$

where M denotes the determinant of the Hessian

$$M\phi = \det H\phi = \det\left[\frac{\partial^2\phi}{\partial x_i\,\partial x_j}\right]_{i,j}.$$

This equation up to now is satisfied by $\phi = id - \psi$ in a formal way only.

Definition 5 We say that a function ϕ satisfies (2.4) in the Brenier sense if $(\nabla\phi)_\sharp u\cdot\mathcal{L}^d = v\cdot\mathcal{L}^d$ (and this is actually the sense to be given to this equation); on the other hand we say that ϕ satisfies (2.4) in the classical sense if it is of class C^2 and the equation holds pointwise.

Other notions of solutions (in the Alexandroff or viscosity sense, for instance) are often used but we introduced here only those that we actually need in order to present the following regularity result (which is well summarized in Theorem 50 of [9]):

Theorem 2.3 *If u and v are $C^{0,\alpha}(\Omega)$ and are both bounded from above and from below on the whole Ω by positive constants and Ω is a convex open set, then the unique Brenier solution ϕ of (2.4) belongs to $C^{2,\alpha}(\Omega) \cap C^{1,\alpha}(\overline{\Omega})$ and satisfies the equation in the classical sense.*

3 Optimality conditions for the interaction case

We are here concerned with the minimization problem for the functional $\mathfrak{F}^\mu$, when the transport term is given by $T(\mu,\nu) = \frac{1}{2}W_2^2(\mu,\nu)$ and the concentration one is an interaction term of the form

$$G(\nu) = \int_{\Omega\times\Omega} V(|x-y|^2)\nu(dx)\nu(dy), \tag{3.1}$$

with $V\colon [0,+\infty[\to [0,+\infty[$ a regular increasing function. From now on Ω will be the closure of a convex nonempty open set in $\mathbb{R}^d$.

A priori, a minimizer for this functional may be an arbitrary probability measure on the set Ω, even a singular one. Our goal is to prove, under suitable assumptions and by means of optimality conditions and of an approximation process, that it is in fact an absolutely continuous measure with bounded density.

We provide here an easy optimality condition for the minimization of $\mathfrak{F}^{\mu}$. We do not go into details in the computation because it follows the same scheme as in [3]. The approximation by measures with positive densities that we are going to use works in this case too, while the alternative proof by convex analysis that may be found in [3] does not, simply because there is no convexity in the term G.

Theorem 3.1 *If a probability measure $\nu \in \mathcal{P}(\Omega)$ is a minimizer for $\mathfrak{F}^{\mu}$, then there exists a constant m such that*

$$\psi + T_{\nu} \geq m; \quad \psi + T_{\nu} = m \ \ \nu\text{-a.e.},$$

where ψ is a Kantorovich potential for the transport from ν to μ and we define

$$T_{\nu}(x) = \int_{\Omega} 2V(|x-y|^2)\,\nu(dy).$$

Proof Let us start from the case when μ is absolutely continuous with positive density. In this case we perform convex variations on an optimal measure ν of the form $\nu_t = \nu + t(\nu_1 - \nu)$ for an arbitrary $\nu_1 \in \mathcal{P}(\Omega)$: if we call ψ_t the unique Kantorovich potential from ν_t to μ which vanishes at a certain fixed point $x_0 \in \Omega$, we get (by means of Duality Formula)

$$\int_{\Omega} \psi_t\,d\nu_t + \int_{\Omega} \psi_t^c\,d\mu + \int_{\Omega\times\Omega} V(|x-y|^2)\nu_t(dx)\nu_t(dy)$$
$$\geq \int_{\Omega} \psi_t\,d\nu + \int_{\Omega} \psi_t^c\,d\mu + \int_{\Omega\times\Omega} V(|x-y|^2)\nu(dx)\nu(dy).$$

After erasing the term $\int_{\Omega} \psi_t^c\,d\mu$ and dividing by t we pass to the limit, and we know that ψ_t converges uniformly (by Ascoli-Arzelà) to the unique Kantorovich potential ψ from ν to μ vanishing at x_0. This provides, at the limit,

$$\int_{\Omega} (\psi + T_{\nu})\,d\nu_1 \geq \int_{\Omega} (\psi + T_{\nu})\,d\nu.$$

Being ν_1 arbitrary we get that $\nu-$a.e. the function $\psi + T_{\nu}$ must be equal to its infimum, and this is the thesis.

To generalize the result to an arbitrary measure μ, just proceed by approximation. This can be performed as in [3] and provides the same formula where ψ becomes just one of the possibly many Kantorovich potentials instead of the only one. The main difference between this case and the case of a measure μ with positive density is in fact the lack of uniqueness (even up to additive constants) of the Kantorovich potential. □

The problem in the condition of Theorem 3.1 lies in the fact that the measure ν appears only in a very implicit way (both in ψ and in T_{ν}), and this does not allow to derive any estimate on it. We will consequently need to pass through an approximation process, exactly as in [5]. Fixed a minimizer $\bar{\nu}$ for $\mathfrak{F}^{\mu}$, we will consider a sequence of problems $(P_{\varepsilon})_{\varepsilon}$ given by the minimization of

$$\mathcal{P}(\Omega) \ni \nu \mapsto T(\mu_{\varepsilon}, \nu) + G(\nu) + \delta_{\varepsilon} A(\nu) + \varepsilon W_2^2(\nu, \bar{\nu}_{\varepsilon}),$$

where

- $(\mu_\varepsilon)_\varepsilon$ is a sequence of probability measures approximating μ with Lipschitz continuous strictly positive densities u_ε;
- the functional A is given by

$$A(\nu) = \begin{cases} \int_\Omega a(v)\, d\mathcal{L}^d & \text{if } \nu = v \cdot \mathcal{L}^d, \\ +\infty & \text{otherwise,} \end{cases}$$

  for a convex function $a\colon [0, +\infty[\to [0, +\infty]$ which is superlinear at infinity and blowing up at 0, i.e. $\lim_{t\to 0^+} a(t) = +\infty$, but finite and C^2 with $a'' \geq c > 0$ on $]0, +\infty[$ (for instance $a(t) = t^2 + 1/t$);
- $(\delta_\varepsilon)_\varepsilon$ is a suitably chosen sequence with $\delta_\varepsilon > 0$ and $\delta_\varepsilon \to 0$.
- $(\bar{\nu}_\varepsilon)_\varepsilon$ is a suitably chosen sequence of measures with $\bar{\nu}_\varepsilon \rightharpoonup \bar{\nu}$.

We will prove, exactly as in [5], that this sequence of problems admits an uniform L^∞ bound for their solutions and that we can choose the parameters so that these solutions converge to $\bar{\nu}$, thus obtaining an L^∞ estimate for $\bar{\nu}$. The existence of solutions for P_ε is trivial by the semicontinuity of each term in the sum with respect to the weak convergence of probability measures on the compact set Ω (and moreover any term but A is actually continuous, while A is semicontinuous by convexity).

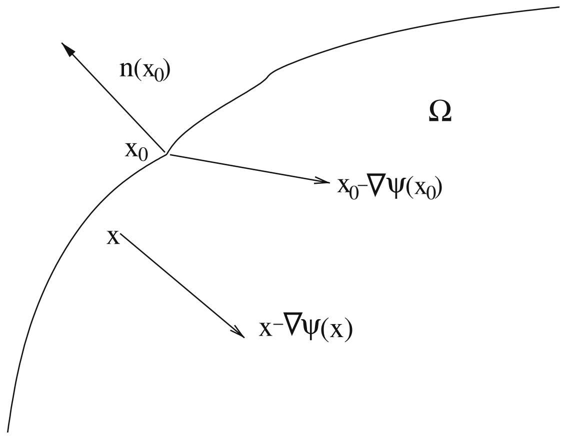
Lemma 3.2 *Suppose that $\mu = u \cdot \mathcal{L}^d$ with $||u||_{L^\infty} \leq M$ and that V is a C^2 function with $V' > 0$. Then any solution ν_ε to the problem P_ε is absolutely continuous and its density is bounded by a constant C depending only on M, d and V.*

Proof First we notice that, thanks to the presence of the term $A(\nu)$ in the minimization problem, ν_ε must be absolutely continuous with strictly positive density almost everywhere. Then we write the optimality conditions for ν_ε with respect to variations of the form $\nu_t = \nu_\varepsilon + t(\nu_1 - \nu_\varepsilon)$. From easy computations we get

$$\psi_\varepsilon + T_{\nu_\varepsilon} + \delta_\varepsilon a'(\nu_\varepsilon) + \varepsilon\chi_\varepsilon = m_\varepsilon \text{ a.e.,}$$

where ψ_ε is the Kantorovich potential for the transport from ν_ε to μ_ε and χ_ε from ν_ε to $\bar{\nu}_\varepsilon$ (they are unique up to additive constants) and m_ε is a suitable constant. We get equality almost everywhere due to the fact that we already know that $\nu_\varepsilon > 0$ (we identify measures and their densities in this context). Since Kantorovich potentials are Lipschitz functions and T_{ν_ε} shares the same regularity of the integrand $(x, y) \mapsto V(|x - y|^2)$, which is C^2 and then Lipschitz, we get that even $a'(\nu_\varepsilon)$ is Lipschitz continuous, and in particular it is bounded. This prevents ν_ε to be close to 0 since it holds $\lim_{t\to 0^+} a'(t) = -\infty$. Thus we get $\nu_\varepsilon \geq c_\varepsilon > 0$. Moreover, $a'(\nu_\varepsilon)$ is Lipschitz continuous and, being a'' bounded from below by a positive constant, also the inverse of a' is Lipschitz. This proves that ν_ε is a Lipschitz continuous function. We can now use regularity theory on Monge–Ampère equation to get $\psi \in C^{2,\alpha}(\Omega) \cap C^{1,\alpha}(\overline{\Omega})$, since both ν_ε and μ_ε are bounded both from above and from below by positive constants (depending on ε, anyway) and are Lipschitz continuous. The same is true for the Kantorovich potential χ_ε by replacing μ_ε by $\bar{\nu}_\varepsilon$. What we can do now is looking for a maximum point x_0 of ν_ε. Notice that such a point will be a minimum point for $\psi_\varepsilon + T_{\nu_\varepsilon} + \varepsilon\chi_\varepsilon$. First we prove that $x_0 \notin \partial\Omega$. To prove this it is sufficient to prove that the gradient of $\psi_\varepsilon + T_{\nu_\varepsilon} + \varepsilon\chi_\varepsilon$ is directed outwards at any point of $\partial\Omega$, i.e. $\nabla(\psi_\varepsilon + T_{\nu_\varepsilon} + \varepsilon\chi_\varepsilon)(x_0) \cdot n(x_0) > 0$ for any $x_0 \in \partial\Omega$, where n is the outward normal vector. From the fact that the optimal transport map t from ν_ε to μ_ε is given by

Fig. 1 Behaviour of $\nabla(\psi)$ near $\partial\Omega$



$t(x) = x - \nabla\psi(x)$ we know that $x - \nabla\psi(x) \in \Omega$ for almost any $x \in \Omega$ (see Fig. 1). In this case, due to continuity up to the boundary of $\nabla\psi$, this holds for any x and also for $x_0 \in \partial\Omega$ and implies $\nabla\psi(x_0) \cdot n(x_0) \geq 0$. Analogously we get $\nabla\chi(x_0) \cdot n(x_0) \geq 0$. For the gradient of T_{ν_ε} it holds

$$\nabla T_{\nu_\varepsilon}(x_0) = \int_\Omega 4V'(|x_0 - y|^2)(x_0 - y)\, \nu_\varepsilon(dy),$$

and so $\nabla T_{\nu_\varepsilon}(x_0) \cdot n(x_0) > 0$ since $V' > 0$ and for almost any $y \in \Omega$ it holds $(x_0 - y) \cdot n(x_0) > 0$. This proves that x_0 lies in the interior of Ω and this allows us to look at the second derivatives. Taking Hessians we have

$$H\psi_\varepsilon(x_0) + HT_{\nu_\varepsilon}(x_0) + \varepsilon H\chi_\varepsilon(x_0) \geq 0,$$

where the letter H denotes Hessians and the inequality is in the sense of positive definite symmetric matrices. Thus we get

$$H\psi_\varepsilon(x_0) \geq -I\left(2||V||_{C^2(\Omega)} + \varepsilon\right),$$

since the second derivatives of T_{ν_ε} may be estimated by those of V and from the fact that $x^2/2 - \chi(x)$ is convex we deduce $H\chi \leq I$. This is a uniform estimate from below for $H\psi_\varepsilon(x_0)$ and for the convex function ϕ given by $\phi(x) = x^2/2 - \psi_\varepsilon(x)$ we have $H\phi(x_0) \leq I\left(1 + \varepsilon + 2||V||_{C^2(\Omega)}\right)$. This implies $M\phi(x_0) \leq (1 + \varepsilon + 2||V||_{C^2(\Omega)})^d$, and, from $\nu_\varepsilon = \mu_\varepsilon(\nabla\phi)M\phi$, we get, for $\varepsilon \leq 1$,

$$\max\ \nu_\varepsilon = \nu_\varepsilon(x_0) \leq 2^d M\left(1 + ||V||_{C^2(\Omega)}\right)^d,$$

which is the desired estimate. □

Remark 1 The proof above of the fact that the gradient is directed outwards (illustrated in Fig. 1 as well) and no maximum point is allowed on the boundary could be used similarly in [5], thus getting rid of the strict convexity assumption in Theorem 6.5 and of the heavy proof of Lemma 6.6. Notice that it could be possible to get the same result even without C^1 regularity for the potentials, just making the proof a bit trickier. It would be sufficient to evaluate the increments of the potential in small balls around x_0 where the gradient is almost everywhere defined and such that $x - \nabla\psi(x),\ x - \nabla\chi(x) \in \Omega$ a.e.

Lemma 3.3 *It is possible to choose the parameters for the problem P_ε, i.e. the numbers δ_ε and the measures $\bar{\nu}_\varepsilon$ and μ_ε so that any sequence of minimizers $(\nu_\varepsilon)_\varepsilon$ converges to $\bar{\nu}$.*

Proof It is sufficient to choose for $\bar{\nu}_\varepsilon$ a sequence of absolutely continuous measures with Lipschitz continuous strictly positive densities such that $\mathfrak{F}^\mu(\bar{\nu}_\varepsilon) \le \mathfrak{F}^\mu(\bar{\nu}) + \varepsilon^2$. Then we have $A(\bar{\nu}_\varepsilon) < +\infty$ and we may choose $\delta_\varepsilon = \varepsilon^2 A(\bar{\nu}_\varepsilon)^{-1}$. For $(\mu_\varepsilon)_\varepsilon$ we can chose any sequence of absolutely continuous measures with Lipschitz continuous strictly positive densities approximating μ in such a way that $W_2(\mu_\varepsilon, \mu) \le \varepsilon^2$. Then we have

$$T(\mu_\varepsilon,\nu_\varepsilon)+G(\nu_\varepsilon)+\delta_\varepsilon A(\nu_\varepsilon)+\varepsilon W_2^2(\nu_\varepsilon,\bar{\nu}_\varepsilon) \le T(\mu_\varepsilon,\bar{\nu}_\varepsilon)+G(\bar{\nu}_\varepsilon)+\delta_\varepsilon A(\bar{\nu}_\varepsilon),$$

which implies

$$\begin{aligned}\mathfrak{F}^\mu(\nu_\varepsilon) + \delta_\varepsilon A(\nu_\varepsilon) + \varepsilon W_2^2(\nu_\varepsilon,\bar{\nu}_\varepsilon) &\le \mathfrak{F}^\mu(\bar{\nu}_\varepsilon) + 4DW_2(\mu_\varepsilon,\mu) + \delta_\varepsilon A(\bar{\nu}_\varepsilon)\\ &\le \mathfrak{F}^\mu(\bar{\nu}) + \varepsilon^2 + 4D\varepsilon^2 + \varepsilon^2\\ &\le \mathfrak{F}^\mu(\nu_\varepsilon) + \varepsilon^2(2+4D).\end{aligned}$$

Finally, this implies $W_2(\nu_\varepsilon,\bar{\nu}_\varepsilon) \le C\sqrt{\varepsilon}$ and, since $\bar{\nu}_\varepsilon \rightharpoonup \bar{\nu}$, we get $\nu_\varepsilon \rightharpoonup \bar{\nu}$. □

Remark 2 This is the point where global optimality of $\bar{\nu}$ plays a crucial role. In fact, should $\bar{\nu}$ be only locally minimizing, we could not use the inequality $\mathfrak{F}^\mu(\bar{\nu}) \le \mathfrak{F}^\mu(\nu_\varepsilon)$, unless we already know that ν_ε is in the domain of minimality of $\bar{\nu}$, i.e. sufficiently close to it.

We can now state our main result and its consequence in the minimization of the whole functional $\mathfrak{F}$.

Theorem 3.4 *Given a compact convex set $\Omega \subset \mathbb{R}^d$ with nonempty interior and a probability measure $\mu \in L^\infty(\Omega)$, if the function V appearing in the definition of the functional G is of class C^2 and $V' > 0$, then the minimization problem for the functional $\mathfrak{F}^\mu$ over the space $\mathcal{P}(\Omega)$ admits at least a solution and any minimizer belongs in fact to the space $L^\infty(\Omega)$.*

Proof As usual the existence is trivial due to continuity and compactness of $\mathcal{P}(\Omega)$ while, for the L^∞ regularity, just apply Lemma 3.2 and Lemma 3.3 □

Corollary 3.5 *Given a compact convex set $\Omega \subset \mathbb{R}^d$ with nonempty interior, a C^1 strictly convex and superlinear function f and a C^2 function V with $V' > 0$, then the minimization problem over the space $\mathcal{P}(\Omega)^2$ for the functional $\mathfrak{F}(\mu,\nu) = \frac{1}{2}W_2^2(\mu,\nu)+F(\mu)+G(\nu)$, where F is defined by (1.1) and G by (3.1), admits a solution and, in any minimizing pair (μ,ν), both μ and ν are in fact absolutely continuous measures $\mu = u\cdot\mathcal{L}^d$, $\nu = v\cdot\mathcal{L}^d$, with $u \in C^0(\Omega)$ and $v \in L^\infty(\Omega)$.*

Proof After the usual proof of existence by the direct method in Calculus of Variations, we refer to [3] for the regularity results on μ. Since such a measure turns out to be absolutely continuous with continuous density (hence bounded), we may apply Theorem 3.4 to get the regularity on ν. □

4 An explicit example

In this section we come back to the whole problem of minimizing the functional $\mathfrak{F}$ in a very particular case, where we can provide almost explicit densities for the solutions. We consider the case

- $T(\mu,\nu) = \frac{1}{2}W_2^2(\mu,\nu)$ and $G(\nu) = \int_{\Omega\times\Omega} V(|x-y|^2)\nu(dx)\nu(dy)$, as in the previous section;
- $V(|x-y|^2) = \frac{\lambda}{2}|x-y|^2$ and so, setting $\mathrm{bar}(\nu) = \int_\Omega y\nu(dy)$, we have $T_\nu(x) = \lambda|x|^2 - 2\lambda x\cdot\mathrm{bar}(\nu) + \lambda\int_\Omega |y|^2\nu(dy)$;
- $F(\mu) = \frac{1}{2}||\mu||^2_{L^2(\Omega)}$, a particular case of what considered in [3].

The framework we obtain is very similar to the one in [5].

Theorem 4.1 *In the specific case described above, any pair of minimizers* (μ,ν) *is shaped as follows:*

- μ *is concentrated on a ball* $B(x_0, r_\lambda)$*(intersected with* Ω*) and has a density u given by*
$$u(x) = \frac{\lambda}{2\lambda+1}\left(r_\lambda^2 - |x-x_0|^2\right);$$
- ν *is concentrated on the ball* $B(x_0, r_\lambda/(2\lambda+1))$ *and it is the image of* μ *under the homothety of ratio* $(2\lambda+1)^{-1}$ *and centre* x_0*, hence it has density v given by*
$$v(x) = \lambda(2\lambda+1)^{d-1}\left(r_\lambda^2 - (2\lambda+1)^2|x-x_0|^2\right);$$
- x_0 *is the barycentre of both* μ *and* ν.

Proof First we write down the optimality conditions given by Theorem 3.1 for the minimization in ν with fixed μ and by [3] for the minimization in μ for fixed ν. We denote by u and v the densities of μ and ν, respectively. We may suppose that the barycentre of ν is the origin, thus obtaining $T_\nu(x) = \lambda|x|^2 + c$. We have

$$\begin{cases} u(x) + \varphi(x) = c_1 & \text{a.e. on } u > 0; \\ \psi(x) + \lambda x^2 = c_2 & \text{a.e. on } v > 0. \end{cases}$$

Here φ and ψ are Kantorovich potentials for the transport from μ to ν and from ν to μ, respectively. From the second condition we can infer $\nabla\psi(x) = -2\lambda x$ a.e. on $v > 0$. Being ν absolutely continuous, this equality is valid $\nu-$a.e.. This means that the optimal transport map t from ν to μ is given by $t(x) = x - \nabla\psi(x) = (2\lambda+1)x$. By uniqueness of the optimal transport plan, which is in this case expressed both as a transport map from ν to μ and viceversa, we know also the optimal transport map s from μ to ν which will be $s(x) = x/(2\lambda+1)$. From duality theory in optimal transportation we know the following equality

$$\varphi(x) + \varphi^c(s(x)) = c(x, s(x)) = \frac{1}{2}|x - s(x)|^2,$$

and thus we get $u(x) = c_1 - \frac{1}{2}|x-s(x)|^2 + \varphi^c(s(x))$. We do already know that u is Lipschitz continuous (by [3]), and this implies that the set $\{u > 0\}$ is an open set. Consequently the same is true for $\{v > 0\}$, which is just an homothety of it. Since φ^c

is a Kantorovich potential from ν to μ, we know that it must agree (up to constants) with ψ on any connected component of the open set $\{v > 0\}$. So, let $\omega \subset \Omega$ be a connected component of $\{u > 0\}$. On $(2\lambda+1)^{-1}\omega$ we have $\varphi^c = \psi + c_3$ and so we get

$$u(x) = c_4 - \frac{1}{2}|x - s(x)|^2 + \psi(s(x)) = c_5 - |x|^2\frac{\lambda}{2\lambda+1}.$$

From this expression it is clear that $\partial\omega \setminus \partial\Omega$ (where u must vanish) is contained in a sphere around 0. This implies that 0 belongs in fact to ω, since no boundary of ω is allowed in the interior of a certain ball around 0. So there is in fact just one connected component for $\{u > 0\}$ and so we get

$$u(x) = \left[c - |x|^2\frac{\lambda}{2\lambda+1}\right]^+. \tag{4.1}$$

From this it is easy to recover the density v of ν since $\nu = s_\sharp\mu$ and we get the thesis. The point x_0 which turns out to be the centre of the balls which are supports for μ and ν is in this notation 0, the barycentre of ν, as in the thesis. It is clear that in this case μ and ν share the same barycentre since they are homothetic. □

Remark 3 In the example of Theorem 4.1 the density v shares the same regularity of u except at the points corresponding to boundary points of Ω where u is positive, i.e. if at $x_0 \in \partial\Omega$ it happens $u(x_0) > 0$ then at $s(x_0)$ we have a jump for v. It is clear from the fact that u is $2\lambda/(2\lambda+1)$-Lipschitz continuous (it follows from the explicit formula) that we have, recalling also $\int_\Omega u\, d\mathcal{L}^d = 1$,

$$1 \leq \left(\inf u + \frac{2\lambda}{2\lambda+1}D\right)|\Omega|,$$

where D is the diameter of Ω. This implies, for small Ω, $\inf u > 0$. In this case u would be positive at any point of $\partial\Omega$ and v discontinuous at any point of $s(\partial\Omega)$. This gives examples when the L^∞ regularity for v cannot be improved up to $v \in C^0(\Omega)$.

Remark 4 In the explicit example above there remain to be determined both the constant c (or the radius r_λ) and the position of the barycentre x_0 in the formula for u. In some simple cases this is possible too. Notice that, once fixed x_0, the constant c may always be recovered by imposing the condition of being probability measures.

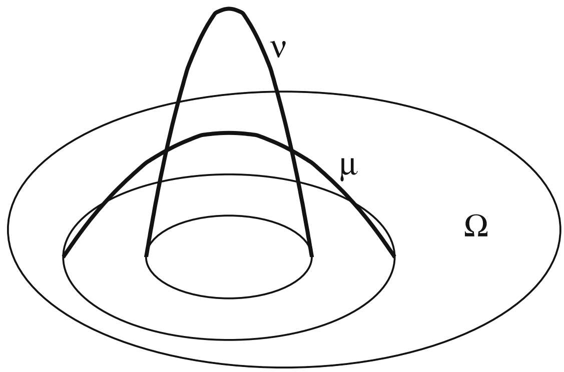


Fig. 2 A solution for a large ball

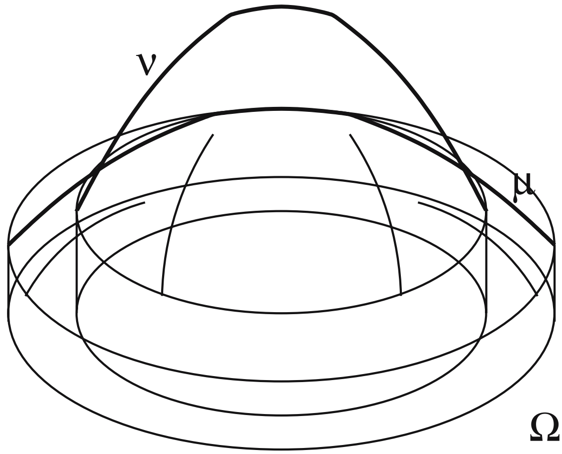


Fig. 3 The solution for a small ball Ω

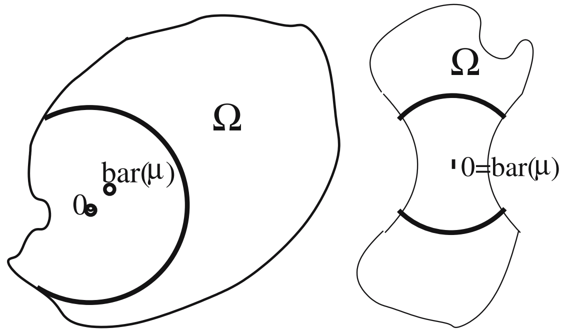


Fig. 4 The position of the centre and of the barycentre

For instance if Ω is a ball, we may see that x_0 may not be the barycentre of a density u shaped as in (4.1) unless the set $B(x_0, 2\lambda^{-1}(2\lambda + 1)) \cap \Omega$ is a ball around x_0. This happens for large Ω whenever the ball $B(x_0, 2\lambda^{-1}(2\lambda + 1))$ does not touch the boundary $\partial\Omega$ or, in general, when x_0 is the centre of the ball Ω. In the first case (Ω a large ball, as in Fig. 2, where the case of a generic Ω is represented) we have several solutions for the problem (non-uniqueness), obtained from each other under translations, and u and v are continuous; in the second (Ω a small ball, Fig. 3) we have uniqueness of the solution, with u a radial continuous function around the centre and v a rescaling of u on a smaller ball.

Remark 5 In general, if Ω is not a ball, the fact that 0 is the barycentre of a distribution of mass which is radial around 0 itself imposes some constraints on the position of 0 with respect to $\partial\Omega$. If the domain Ω cuts a part of the supporting ball from one side, then the centre of the ball could be no longer the barycentre. Figure 4 shows this effect, as well as a situation where the support touches the boundary on two sides and the centre of the ball is actually the barycentre.

Remark 6 It is interesting to see the behaviour of the solution (μ, ν) as $\lambda \to 0, +\infty$. In the first case, as $\lambda \to 0$, from (4.1), we easily get that u tends to a constant density. This comes from the fact that the importance of the functional G decreases, and this allows us to choose ν under no concentration criteria; in particular at the limit we can choose $\nu = \mu$, thus getting $T(\mu, \nu) = 0$: then the only thing to do is choosing μ so that we minimize F, which exactly happens only for constant density measures. On the other hand, as $\lambda \to +\infty$, the role played by G is increasing and in the end we will get a Dirac measure $\nu = \delta_0$, (and Dirac masses are the only minimizers of G). This can be seen from the fact that the homothety ratio between ν and μ tend to 0. The optimal μ can be retrieved from the formula in Theorem 4.1 and we can easily see that it holds $u(x) = (r^2 - |x - x_0|^2)/2$.

Acknowledgements The author acknowledges the support of the project *"Calcolo delle Variazioni"* of the Italian Ministry of Education. Moreover, the ideas of this work, much linked to the previous joint papers, have been discussed together with Prof. Buttazzo and Prof. Carlier, to whom sincere thanks are addressed.

References

1. Bouchitté, G., Buttazzo, G.: New lower semicontinuity results for nonconvex functionals defined on measures. Nonlinear Anal. **15**(7), 679–692 (1990)
2. Buttazzo, G., Oudet, E., Stepanov, E.: Optimal transportation problems with free dirichlet regions. Prog. Nonlinear Diff. Equat. Appl. **51**, 41–65 (2002)
3. Buttazzo, G., Santambrogio, F.: A model for the optimal planning of an urban area. SIAM J. Math. Anal. **37**(2), 514–530 (2005)
4. Carlier, G., Ekeland, I.: The structure of cities. J. Global Optim. **29**(4), 371–376 (2004)
5. Carlier, G., Santambrogio, F.: A variational model for urban planning with traffic congestion. ESAIM Control Optim. Calculus of Variations **11**(4), 595–613 (2005)
6. Maddalena, F., Solimini, S., Morel, J.-M.: A variational model of irrigation patterns. Interf. Free Bound. **5**, 391–415 (2003)
7. McCann, R.J.: A convexity principle for interacting gases. Adv. Math. **128**(1), 153–159 (1997)
8. Santambrogio, F.: Misure ottime per costi di trasporto e funzionali locali (in italian). Laurea Thesis, Università di Pisa, advisor: G. Buttazzo, available at *www.unipi.it/etd* and *cvgmt.sns.it* (2003)
9. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, AMS (2003)
10. Xia, Q.: Optimal paths related to transport problems. Comm. Cont. Math. **5**(2), 251–279 (2003)