HAL
open science

A strategy for non-strictly convex transport costs and the example of ||x-y||p in R2

Guillaume Carlier, Luigi de Pascale, Filippo Santambrogio

To cite this version:

Guillaume Carlier, Luigi de Pascale, Filippo Santambrogio. A strategy for non-strictly convex transport costs and the example of ||x-y||p in R2. Communications in Mathematical Sciences, 2010, à paraître. hal-00417303

HAL Id: hal-00417303

https://hal.science/hal-00417303v1

Submitted on 15 Sep 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A STRATEGY FOR NON-STRICTLY CONVEX TRANSPORT COSTS AND THE EXAMPLE OF $||x-y||^p$ IN $\mathbb{R}^2$

GUILLAUME CARLIER, LUIGI DE PASCALE, AND FILIPPO SANTAMBROGIO

Abstract. This paper deals with the existence of optimal transport maps for some optimal transport problems with a convex but non strictly convex cost. We give a decomposition strategy to address this issue. As part of our strategy, we have to treat some transport problems, of independent interest, with a convex constraint on the displacement. As an illustration of our strategy, we prove existence of optimal transport maps in the case where the source measure is absolutely continuous with respect to the Lebesgue measure and the transportation cost is of the form $h(\|x-y\|)$ with h strictly convex increasing and $\|.\|$ an arbitrary norm in $\mathbb{R}^2$.

1. Introduction

Given two probability measures μ and ν on $\mathbb{R}^d$ and a transport cost $c:\mathbb{R}^d\to\mathbb{R}$ the corresponding Monge problem consists in minimizing the average transport cost $\int_{\mathbb{R}^d} c(x-T(x))\mu(dx)$ among all *transport maps* T i.e. maps pushing forward μ to ν (which as usual is denoted by $T_\#\mu=\nu$). It is a highly nonconvex problem (whose admissible set may even be empty if μ has atoms for instance) and it is therefore relaxed to the Monge-Kantorovich problem that consists in minimizing $\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x-y)\gamma(dx,dy)$ over $\Pi(\mu,\nu)$, the set of transport plans i.e. of probability measures on $\mathbb{R}^d\times\mathbb{R}^d$ having μ and ν as marginals. To prove existence of an optimal transport map one thus aims to prove that there is an optimal plan γ (existence of such plans holds under very mild assumptions since the Monge-Kantorovich problem is linear) that is in fact induced by a transport map i.e. of the form $\gamma=(\mathrm{id},T)_\#\mu$. To achieve this goal, one usually uses strongly the dual problem that consists in maximizing $\int_{\mathbb{R}^d}\phi d\mu+\int_{\mathbb{R}^d}\psi d\nu$ subject to the constraint that $\phi(x)+\psi(y)\le c(x-y)$. An optimal pair (ϕ,ψ) for the dual is called a pair of Kantorovich potentials. It is very well-known (under reasonable assumptions on the measures) that when c is strictly convex then this strategy gives an optimal transport (see [12] or section 2 below where the arguments is briefly recalled; we also refer to the book [14] for a general overview and recent developments of optimal transport theory).

It is also well-known that lack of strict convexity makes the existence of an optimal transport much more delicate. Even the important case (originally considered by Monge) where c is a norm was well understood only in recent years ([11], [4], [3], [2], [8], [6], [9], [7]). In the case of the Euclidean norm, for example, the direction of the displacement is determined by a Kantorovich potential and transport only takes place on a set of segments called *transport rays*. On the one hand, the lack of strict convexity in the radial direction gives rise to an indeterminacy of the displacement length, but on the other hand, the problem on transport rays is one dimensional and then much simpler. These observations lead to a strategy proof originally due to Sudakov ([13]) consisting in reducing to a one-dimensional problem on each transport ray (monotone transport

Date: August 28 2009.
2000 *Mathematics Subject Classification.* 49Q20, 49K30, 49J45.
Key words and phrases. Monge-Kantorovich problem, optimal transport problem.

for instance) and then glue the solutions together. The most involved part of the full proof consisted in proving that the transport rays have enough regularity so that the disintegrated measures on such rays are non-atomic (see [4], [6] for such a proof).

In the present paper we consider a convex but not strictly convex c. We propose a decomposition strategy that takes advantage of the fact that whenever the displacement is not fully determined it means that it lies in some *face* of c but on such a face the cost is affine and is therefore unchanged if we replace the transport plan by another one which has the same marginals and satisfy the further constraint that the displacement belongs to the face. In the spirit, our strategy can be compared to Sudakov's one but it is different since there is no real analogue of transport rays here. Instead, our strategy, detailed in section 2, involves "face restricted problems" that are optimal transport problems with convex constraints on the displacement. Such constrained problems have, we believe, their own interest and motivation (e.g. due to connections with L^∞ transportation problems as studied in [10]) and we will address some of them in section 3. We will avoid here subtle disintegration arguments to glue together the face restricted problems in general but will instead illustrate in section 4 how our strategy easily produces an optimal transport map in the case of $c(z) = h(\|z\|)$ with h strictly convex increasing and $\|\cdot\|$ an arbitrary norm in $\mathbb{R}^2$. The contributions of this paper are then :

- a general decomposition strategy to deal with convex but non strictly convex costs,
- a contribution to constrained transport problems,
- the proof of existence of optimal transport maps for a class of costs in $\mathbb{R}^2$.

2. Strategy of decomposition

In this section we outline a general decomposition strategy to study existence of an optimal transport for the Monge-Kantorovich problem

$$\min\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x-y)\,\gamma(dx,dy)\ :\ \gamma\in\Pi(\mu,\nu)\right\},$$

where c is convex but not strictly convex. Not all of the steps listed below can be always carried on in full generality. In the following sections we will detail some case where this is possible and illustrate some applications.

The decomposition is based on the following steps:

- Consider an optimal plan γ and look at the optimality conditions by means of a solution (ϕ,ψ) of the dual problem. From the fact that

  $$\phi(x)+\psi(y)=c(x-y)\ \text{ on }\ \operatorname{spt}\gamma \quad\text{and}\quad \phi(x)+\psi(y)\le c(x-y)$$

  one deduces that if x is a differentiability point for ϕ (which is denoted $x\in\operatorname{Diff}(\phi)$),

  $$\nabla\phi(x)\in\partial c(x-y),$$

  which is equivalent to

  $$x-y\in\partial c^*(\nabla\phi(x)). \tag{2.1}$$

  Let us define

  $$\mathcal{F}_c := \{\partial c^*(p)\ :\ p\in\mathbb{R}^d\},$$

  which is the set of all values of the subdifferential multi-map of c^*. These values are those convex sets where the function c is affine, and they will be called *faces* of c.

  Thanks to (2.1), for every fixed x, all the points y such that (x,y) belongs to the support of an optimal transport plan are such that the difference $x-y$ belong to a same face of c. Classically, when these faces are singleton (i.e. when c^* is differentiable, which

is the same as c being strictly convex), this is the way to obtain a transport map, since only one y is admitted for every x.

Equation (2.1) also enables one to classify the points x as follows. For every $K \in \mathcal{F}_c$ we define the set

$$X_K := \{x \in \mathrm{Diff}(\phi) \ : \ \partial c^*(\nabla\phi(x)) = K\}.$$

Hence γ may be decomposed into several subplans γ_K according to the criterion $x \in X_K$, which, as we said, is equivalent to $x - y \in K$.

If the set $\mathcal{F}_c$ is finite or countable, we can define

$$\gamma_K := \gamma_{|X_K \times \mathbb{R}^d},$$

which is the simpler case. Actually, in this case, the marginal $\mu_K := (\pi_1)_\# \gamma_K$ (where $\pi_1(x, y) = x$) is a submeasure of μ, and in particular it inherits the absolute continuity from μ. This is often useful for proving existence of transport maps.

If $\mathcal{F}_c$ is uncountable, in some cases one can still rely on a countable decompositions by considering the set $\mathcal{F}_c^{multi} := \{K \in \mathcal{F}_c \ : \ K \text{ is not a singleton }\}$. If $\mathcal{F}_c^{multi}$ is countable, then one can separate those x such that $\partial c^*(\nabla(\phi(x))$ is a singleton (where a transport already exists) and look at a decomposition for $K \in \mathcal{F}_c^{multi}$ only.

In some other cases, it could be useful to bundle together different possible K's so that the decomposition is countable, even if coarser. We will give an example of this last type in Section 4.

- This reduces the transport problem to a superposition of transport problems of the type

$$\min\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x-y)\,\gamma(dx,dy) \ : \gamma \in \Pi(\mu_K, \nu_K),\ spt\gamma \subset \{x - y \in K\}\right\}.$$

  The advantage is that the cost c restricted to K is easier to study. For instance, if K is a face of c, then c is affine on K and in this case the transport cost does not depend any more on the transport plan.
- If K is a face of c the problem is reduced to find a transport map from μ_K to ν_K satisfying the constraint $x - T(x) \in K$, knowing a priori that a transport plan satisfying the same constraint exists.

  In some cases (for example if K is a convex compact set with non-empty interior) this problem may be reduced to an L^∞ transport problem. In fact if one denotes by $||\cdot||_K$ the (gauge-like) "norm" such that $K = \{x \ : \ ||x||_K \le 1\}$, one has

$$\min\left\{\max\{||x-y||_K,\ (x,y) \in \mathrm{spt}(\gamma)\}, \gamma \in \Pi(\mu,\nu)\right\} \le 1 \tag{2.2}$$

  and the question is whether the same result would be true if one restricted the admissible set to transport maps only (passing from Kantorovich to Monge, say). The answer would be positive if a solution of (2.2) was induced by a transport map T (which is true if $\mu_K \ll \mathcal{L}^d$ and $K = \overline{B(0,1)}$, see [10], but is not known in general). Moreover, the answer is also positive in $\mathbb{R}$ where the monotone transport solves all the L^p problems, and hence the L^∞ as well.
- A positive answer may be also given in case (and it is actually almost equivalent) one is able to select, for instance by a secondary minimization, a particular transport plan satisfying $\mathrm{spt}(\gamma) \subset \{x - y \in K\}$ which is induced by a map. This leads to the very natural question of solving

$$\min\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} [\frac{1}{2}|x-y|^2 + \chi_K(x-y)]\,d\gamma \ : \ \gamma \in \Pi(\mu,\nu)\right\},$$

or, more generally, transport problems where the cost function involves convex constraints on $x - y$. These problems are studied in Section 3. For instance, in the quadratic case above, we can say that the optimal transport T exists and is given by $T(x) = x - \mathrm{proj}_K(\nabla\phi(x))$ (proj_K denoting projection on K), provided two facts hold:
  - μ is absolutely continuous, so that we can assure the existence of a (possibly approximate) gradient of ϕ $\mu-$a.e. under mild regularity assumptions on ϕ;
  - an optimal potential ϕ does actually exist, in a class of functions (Lipschitz, BV) which are differentiable almost everywhere, at least in a weak sense.
- In order to apply the study of convex-constrained problems to the original problem with $c(x - y)$ the first issue (i.e. absolute continuity) does not pose any problem if the decomposition is finite or countable, while it is non trivial, and it presents the same kind of difficulties as in Sudakov's solution of the Monge's problem, in case of disintegration. For interesting papers related to this kind of problems, see for instance [5, 6].
- As far as the second issue is concerned, this is much more delicate, since in general there are no existence results for potentials with non-finite costs. In particular, a counterexample has been provided by Caravenna when $c(x, y) = |x - y| + \chi_K(x - y)$ where $K = \mathbb{R}^+ \times \mathbb{R}^+ \subset \mathbb{R}^2$ and it is easily adapted to the case of quadratic costs with convex constraints. On the other hand, it is easy to think that the correct space to set the dual problem in Kantorovitch theory for this kind of costs would be BV since the the constraints on $x - y$ enable one to control the increments of the potentials ϕ and ψ on some directions, thus giving some sort of monotonicity. Yet, this is not sufficient to find a bound in BV if an L^∞ estimate is not available as well and the counterexample that we mentioned - which gives infinite values for both ϕ and ψ, exactly proves that this kind of estimates are hard to prove.

Remark 2.1. An interesting example that could be approached by this strategy is that of crystalline norms (a problem that has been already solved by a different method in [2]). In this case the faces of the cost c are polyhedral cones but, if the support of the two measures are bounded, we can suppose that they are compact convex polyhedra. This means, thanks to the considerations we did above, that it is possible to perform a finite decomposition and to reduce the problem to some L^∞ minimizations for norms whose unit balls are polyhedra. In particular solving the L^∞ problem for crystalline norms is enough to solve the usual L^1 optimal transport problem for the crystalline norms.

Remark 2.2. If, on the other hand, one wants a simple example that can be completely solved through this strategy and that works in any dimension, one can look at the cost $c(z) = (|z|-1)^2_+$, which vanishes for displacements smaller than one. In this case it is easy to see that $\mathcal{F}_c^{multi}$ has only one element, which is given by the Euclidean ball $\overline{B(0, 1)}$. The associated L^∞ problem has been solved in [10] and this is enough to get the existence of an optimal transport.

3. Constrained transport problems

In this section we see two useful examples of transport problem under the constraint $x-y \in K$. Since the cost we use, due to this constraint, is lower semicontinuous but not finitely valued, it is well-known that duality holds (the minimum of the Kantorovich problem coincides with the supremum of the dual one), but existence of optimizers in the dual problem is not guaranteed. As we underlined in Section 2, this is a key point and what we present here will always assume (artificially) that this existence holds true. In section 4 we will show a relevant example in which this is actually the case.

3.1. **Strictly convex costs with convex constraints.** We start from the easiest transport problem with convex constraints:

Theorem 3.1. *Let μ, ν be two probability measures on $\mathbb{R}^d$, with $\mu \ll \mathcal{L}^d$, K a closed and convex subset of $\mathbb{R}^d$ and $h : \mathbb{R}^d \to \mathbb{R}$ a strictly convex function. Let $c(z) = h(z) + \chi_K(z)$: then the transport problem*

$$\min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x-y)\, \gamma(dx, dy),\ \gamma \in \Pi(\mu, \nu) \right\}$$

admits a unique solution, which is induced by a transport map T, provided that a transport plan with finite cost exists and the dual problem admits a solution (ϕ, ψ) where ϕ is at least approximately differentiable a.e.

Proof. As usual, consider an optimal transport plan γ and the potentials ϕ and ψ optimality conditions on optimal transport plans and optimal potentials for this problem read as

$$\phi(x) + \psi(y) = h(x-y) \ \text{ on } \operatorname{spt}\gamma \subset \{x - y \in K\}$$

and

$$\phi(x) + \psi(y) \le h(x-y) \ \text{ for all } (x, y) \,:\, x - y \in K$$

and, if ϕ is differentiable at x, they lead to

$$\nabla\phi(x) \in \partial c(x-y) = \partial h(x-y) + \mathcal{N}_K(x-y), \tag{3.1}$$

where $N_K(z)$ is the normal cone to K at z. We used the fact that the sub-differential of the sum h and χ_K is the sum of their sub-differentials since h is real-valued and hence continuous. Equation (3.1) is verified by the true gradient of ϕ if it exists but it stays true for the approximate gradient if ϕ is only approximately differentiable.

Yet, when a vector l and a point $\bar{z} \in K$ satisfy

$$l \in \partial h(\bar{z}) + N_K(\bar{z}),$$

thanks to the convexity of h and K, this gives that $\bar{z}$ minimizes $K \ni z \mapsto h(z) - l \cdot z$. Since h is strictly convex this gives the uniqueness of $\bar{z}$, which will depend on l. We will denote it by $\bar{z}(l)$.

In this case, we get $x - y = \bar{z}(\nabla\phi(x))$, which is enough to identify $y = T(x) := x - \bar{z}(\nabla\phi(x))$ and proving existence of a transport map which is necessarily unique. □

Remark 3.2. Notice that, in the case $h(z) = \frac{1}{2}|z|^2$, the point $\bar{z}(l)$ will be the projection of l on K, which gives the nice formula

$$T(x) = x - \operatorname{proj}_K(\nabla\phi(x)),$$

i.e. a generalization of the usual formula for the optimal transport in the quadratic case.

3.2. **Strictly convex costs of one variable and convex constraints.** Let $\mu \ll \mathcal{L}^2$ and ν be two probability measures in $\mathbb{R}^2$, let K be a convex subset of $\mathbb{R}^2$ such that $\overset{\circ}{K} \neq \emptyset$ and consider the cost

$$c(x-y) = h(x_1 - y_1) + \chi_K(x-y),$$

for a function $h : \mathbb{R} \to \mathbb{R}$ increasing and strictly convex.

Theorem 3.3. *Assume that there exist $\gamma \in \Pi(\mu, \nu)$ and ϕ, ψ Lipschitz such that*

$$\phi(x) + \psi(y) \le c(x-y) \quad \forall x, y \tag{3.2}$$

$$\phi(x) + \psi(y) = c(x-y) \quad \gamma - a.e. \tag{3.3}$$

Then there exists an optimal transport map for the cost c between μ and ν

Remark 3.4. The assumption above implies that γ is an optimal plan and the couple (ϕ, ψ) is optimal for the dual problem. In order to have the existence of an optimal γ it is enough to assume that there exists at least one $\sigma \in \Pi(\mu, \nu)$ such that

$$\int_{\mathbb{R}^2 \times \mathbb{R}^2} c(x-y)\sigma(dx, dy) < \infty.$$

Again we recall that this assumption is not enough to have the existence of an optimal couple for the dual problem.

Proof. As $\mathcal{L}^2(\mathbb{R}^2 \setminus \mathrm{Diff}(\phi)) = 0$ also $\mu(\mathbb{R}^2 \setminus \mathrm{Diff}(\phi)) = 0$ then we can restrict our attention to the points of differentiability of ϕ. For each $(x, y) \in spt(\gamma)$ such that $x \in \mathrm{Diff}(\phi)$ by (3.2) and (3.3) we obtain

$$\nabla\phi(x) \in h'(x_1 - y_1)e_1 + \mathcal{N}_K(x-y). \tag{3.4}$$

By the convexity of the functions involved (3.4) is equivalent to

$$x - y \in arg\min_z \{h(z_1) - \nabla\phi(x) \cdot z + \chi_K(z)\}.$$

Let S be the set of those x such that the set on the right-hand side of (3.4) is a singleton

$$arg\min_z \{h(z_1) - \nabla\phi(x) \cdot z + \chi_K(z)\} = \{p\},$$

then y is uniquely determined by

$$y = x - p.$$

Let us consider a decomposition of γ in two parts: $\gamma = \gamma_{|S \times \mathbb{R}^2} + \gamma_{|S^c \times \mathbb{R}^2}$ The first part of the decomposition is already supported on the graph of a Borel map. As far as the second part is concerned, we will prove the existence of a transport map which gives the same cost and the same marginals. This will be done by a sort of one-dimensional decomposition according to the following observations.

Whenever the set on the right-hand side of (3.4) is not a singleton then by the convexity of the function $h(z_1) - \nabla\phi(x) \cdot z$ it is a convex subset of K. Even more, by the strict convexity of h there exist a number $m(\nabla\phi(x))$ such that

$$arg\min_z \{h(z) - \nabla\phi(x) \cdot z + \chi_K(z)\} \subset \{z \cdot e_1 = m(\nabla\phi(x))\} \bigcap K.$$

We claim that if $arg\min_z \{h(z) - \nabla\phi(x) \cdot z + \chi_K(z)\}$ has more than one element then x is a local maximum of ϕ on the line $x + te_2$. In fact assume that there exist two points $y = (m(\nabla\phi(x)), y_2)$ and $\tilde{y} = (m(\nabla\phi(x)), \tilde{y}_2)$ such that $x - y \in arg\min_z \{h(z) - \nabla\phi(x) \cdot z + \chi_K(z)\}$ and $x - \tilde{y} \in arg\min_z \{h(z) - \nabla\phi(x) \cdot z + \chi_K(z)\}$ and assume without loss of generality that $\tilde{y}_2 < y_2$. As $x - y \in K$ also $x + te_2 - y \in K$ for small and positive t (because it belongs to the segment $[x - y, x - \tilde{y}]$), then

$$\phi(x + te_2) + \psi(y) \leq h(x_1 - m(\nabla\phi(x)))$$
$$\phi(x) + \psi(y) = h(x_1 - m(\nabla\phi(x)))$$

and subtracting the second equation from the first

$$\phi(x + te_2) - \phi(x) \leq 0.$$

The same inequality is obtained for small, negative t using $\tilde{y}$. Introduce now the set $M_n = \{x \mid \phi(x + te_2) \leq \phi(x)\ \forall t \in [-\frac{1}{n}, \frac{1}{n}]\}$. The set M_n is closed. Let $M_n^i := M_n \cap \{x : x \cdot e_2 \in [\frac{i}{n}, \frac{i+1}{n}]\}$, as a consequence of local maximality, ϕ is vertically constant on M_n^i. There exists a function $\tilde{\phi}$ depending only on the variable x_1 such that $\tilde{\phi}$ coincide with ϕ on M_n^i. Then on the set $\mathcal{L}(M_n^i) \cap \mathrm{Diff}(\phi)$, $\nabla\phi = app\nabla\phi = app\nabla\tilde{\phi}$ and the latter approximate gradient depends only on

x_2 which implies that $\nabla\phi$ is vertically constant on a subset of full measure of M_n^i. The disjoint union $S^c = \cup_{n,i}[S^c \cap (M_n^i \setminus M_{n-1})]$ induces the decomposition

$$\gamma_{|S^c\times\mathbb{R}^2} = \sum_{n,i} \gamma_{n,i},$$

where $\gamma_{n,i} := \gamma_{|[S^c\cap(M_n^i\setminus M_{n-1})]\times\mathbb{R}^2}$.

Denote by $\mu_{n,i}$ and $\nu_{n,i}$ the marginals of $\gamma_{n,i}$. Clearly $\mu_{n,i}$ is absolutely continuous with respect to $\mathcal{L}^2$. Consider the disintegration according to x_1

$$\mu_{n,i} = a_{n,i}(x_1)\cdot\mathcal{L}^1 \otimes \mu_{n,i}^{x_1}$$

with $\mu_{n,i}^{x_1} \ll \mathcal{L}^1$ for almost every x_1, where $a_{n,i}(x_1)\cdot\mathcal{L}^1$ is the projection of $\mu_{n,i}$ on the first variable (which is absolutely continuous as well). Analogously, we define $\nu_{n,i}^{y_1}$ as the disintegration of $\nu_{n,i}$ with respect to y_1. From what we said, it follows that there exist a map $T_{n,i} : \mathbb{R} \to \mathbb{R}$ such that $spt\gamma_{n,i} \cap \{x \ : \ x\cdot e_1 = x_1\} \subset \{(x,y) \ : \ y\cdot e_1 = T_{n,i}(x_1)\}$. Notice that $T_{n,i}(x_1)$ coincides with $m(\nabla\phi(x))$ for every $x \in M_n^i$ whose first component is x_1.

Also disintegrate

$$\gamma_{n,i} = a_{n,i}(x_1)\cdot\mathcal{L}^1 \otimes \gamma_{n,i}^{x_1}$$

with $\gamma_{n,i}^{x_1} \in \Pi(\mu_{n,i}^{x_1}, \delta_{T_{n,i}(x_1)} \otimes \nu_{n,i}^{T_{n,i}(x_1)})$. It is then enough to replace $\gamma_{n,i}$ with a new transport plan of the form $(\mathrm{id}\times S_{n,i})_\#\mu_{n,i}$ with the following properties

$$S_{n,i}(x_1,x_2) = (T_{n,i}(x_1), S_{n,i}^{x_1}(x_2)),$$
$$(S_{n,i}^{x_1})_\#\mu_{n,i}^{x_1} = \nu_{n,i}^{T_{n,i}(x_1)}, \quad x_2 - S_{n,i}^{x_1}(x_2) \in K_{x_1 - T_{n,i}(x_1)},$$

where we denote, for every t, the section of K at level t by $K_t := \{s \ : \ (t,s)\in K\}$.

Since the h-part of the cost only depends on $x_1 - y_1$ and $\gamma_{n,i}$ and the transport plan issued by the map $S_{n,i}$ realize the same displacements as far as the first coordinates are concerned, this part of the cost does increase. Moreover, the constraint $(x_1 - y_1, x_2 - y_2) \in K$, which amounts to the requirement that $x_2 - y_2$ belongs to a segment $K_{x_1 - y_1}$, is preserved, thanks to the last condition above. Hence, the two transport plans have exactly the same cost.

We are only left to find, for every x_1, the map $S_{n,i}^{x_1}$. For this, it is enough to choose the monotone transport map from $\mu_{n,i}^{x_1}$ to $\nu_{n,i}^{x_1}$. This map is well defined because the first measure is absolutely continuous (in particular there are no atoms). The constraint is preserved because $\gamma_{n,i}^{x_1}$ satisfied it, and the monotone transport map is optimal for any convex transport cost of $x_2 - y_2$.

Notice that the whole map $S_{n,i}$ we are using is measurable since in every ambiguous case we chose the monotone transport map. □

4. Application: $c(x-y) = h(\|x-y\|)$ in $\mathbb{R}^2$

Theorem 4.1. *Let μ, ν be probability measures compactly supported in $\mathbb{R}^2$, with $\mu \ll \mathcal{L}^2$, let $||\cdot||$ be an arbitrary norm of $\mathbb{R}^2$ and $h : \mathbb{R}^+ \to \mathbb{R}^+$ a strictly convex and increasing function. Denote by c the function $c(z) = h(||z||)$, which is still a convex function. Then the transport problem*

$$\min\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x-y)\,\gamma(dx,dy),\ \gamma\in\Pi(\mu,\nu)\right\}$$

admits at least a solution which is induced by a transport map T.

Proof. The strategy to prove this theorem is almost the one described in Section 2, with the additional trick of bundling together some faces of the convex function $h(||x-y||)$. The boundary of the unit ball of the norm $||\cdot||$ has a countable number of flat parts (they are countable since

each one has a positive length, otherwise it would not be a flat part, and we cannot have more than a countable quantity of disjoint positive length segments in the boundary). We call F_i, with $i = 1, 2, \dots$ the closure of these flat parts and we then associate to each face F_i the cone $K_i = \{tz,\ t \geq 0,\ z \in F_i\}$.

Consider an optimal transport plan γ and a pair of Kantorovich potentials (ϕ, ψ) (these objects exist since the cost is continuous). From the general theory of optimal transport and what we underlined in Section 2, it follows that if $(x, y) \in \mathrm{spt}(\gamma)$ and x is a differentiability point of ϕ, then $\nabla\phi(x) \in \partial c(x - y)$. This may be re-written as

$$x - y \in \partial c^*(\nabla\phi(x)).$$

We classify the points x according to $\partial c^*(\nabla\phi(x))$. This subdifferential may be either a singleton or a segment (it cannot be a set of dimension two because otherwise c would be affine on a set with non-empty interior, which is in contradiction with the strict convexity of h). We call X_0 the set of those x such that $\partial c^*(\nabla\phi(x))$ is a singleton. If instead it is a segment, it is a segment of points which share a single vector in the subdifferential of c. This means, thanks to the shape of c, that it is necessarily an homothety of one face F_i:

$$x - y \in \partial c^*(\nabla\phi(x)) = t(x)F_{i(x)} \subset K_{i(x)}.$$

We denote by X_i the set of points x where $c^*(\nabla\phi(x))$ is an homothety of F_i.

The disjoint decomposition given by the sets X_i induces a corresponding decomposition $\gamma = \sum_{i\geq 0} \gamma_i$ where $\gamma_i = \gamma_{|X_i \times \mathbb{R}^2}$. Each sub-transport plan γ_i is an optimal transport plan between its marginals μ_i and ν_i, which are submeasures of μ and ν, respectively. In particular, each measure μ_i is absolutely continuous.

We now build a solution to the optimal transport problem for the cost c that is induced by a transport map by modifying γ in the following way: γ_0 is already induced by a transport map since for every x we only have one y; every γ_i for $i \geq 1$ will be replaced by a new transport plan with the same marginals which does not increase the cost, thanks to Theorem 3.3. Since γ_i sends μ_i to ν_i, is optimal for the cost c and satisfies the additional condition $\mathrm{spt}(\gamma_i) \subset \{(x, y) : x - y \in K_i\}$, it is also optimal for the cost $c(x - y) + \chi_{K_i}(x - y)$. The advantage of this new cost is that c depends on one variable only, when restricted to K_i since there exists a new basis (e_1, e_2) such that $||z|| = z_1$ for every $z \in K_i$ which is written as (z_1, z_2) in this new basis (the direction of the vector e_1 being the normal direction to F_i). Theorem 3.3 precisely states that the optimal transport problem for a cost which is given by a strictly convex function of one variable plus a constraint imposing that $x - y$ belongs to a given convex set admits a solution induced by a transport map, provided the first measure is absolutely continuous and some γ, ϕ and ψ satisfying equations 3.2 and 3.3 exist. This last assumption is exactly satisfied by taking γ_i and the original potentials ϕ and ψ for the problem with cost c from μ to ν, since they satisfied

$$\phi(x) + \psi(y) \leq h(||x - y||) \leq h(||x - y||) + \chi_{K_i}(x - y),$$
$$\phi(x) + \psi(y) = h(||x - y||) = h(||x - y||) + \chi_{K_i}(x - y) \text{ on } \mathrm{spt}(\gamma_{K_i}). \qquad \square$$

Acknowledgments

The work of the authors has been partially supported by the University Paris Dauphine (CEREMADE) and the University of Pisa through the project "Optimal Transportation and related topics". G.C. and F.S. gratefully acknowledge the support of the Agence Nationale de la Recherche through the project "ANR-07-BLAN-0235 OTARIE".

The authors wish to thank the Centro de Ciencias de Benasque Pedro Pascual (CCBPP) where this work was completed during the program "Partial differential equations, optimal design and numerics", August 23 - September 4 2009.

References

[1] Ambrosio, L., Lecture Notes on Optimal Transportation Problems, Mathematical aspects of evolving interfaces (Funchal, 2000), Lecture Notes in Math., **1812**, Springer, Berlin, 2003, 1–52.

[2] Ambrosio, L., Kirchheim, B., Pratelli, A., *Existence of optimal transport maps for crystalline norms*, Duke Math. J., **125** (2004), no. 2, 207–241.

[3] Ambrosio, L., Pratelli, A.,, Existence and stability results in the L^1 theory of optimal transportation, Optimal transportation and applications (Martina Franca, 2001), Lecture Notes in Math., **1813**, Springer, Berlin, 2003, 123–160.

[4] Caffarelli, L.A., Feldman, M., McCann, R.J., *Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs*, J. Amer. Math. Soc. **15** (2002), no. 1, 1–26.

[5] L. Caravenna , S. Daneri. *The disintegration of the Lebesgue measure on the faces of a convex function,* preprint, 2009, available at `http://cvgmt.sns.it`

[6] L. Caravenna *A proof of Sudakov theorem with strictly convex norms*, preprint, 2008, available at `http://cvgmt.sns.it`.

[7] L. Caravenna *An existence result for the Monge problem in $\mathbb{R}^n$ with norm cost functions*, preprint, 2008, available at `http://cvgmt.sns.it`.

[8] T. Champion and L. De Pascale, *The Monge problem for strictly convex norms in $\mathbb{R}^d$*, to appear in J.E.M.S., preprint available at `http://cvgmt.sns.it`.

[9] T. Champion and L. De Pascale, *The Monge problem in $\mathbb{R}^d$*, preprint, 2009, available at `http://cvgmt.sns.it`.

[10] Champion, T. De Pascale, L., Juutinen, P. *The ∞-Wasserstein distance: local solutions and existence of optimal transport maps*, SIAM J. of Mathematical Analysis **40** (2008), no. 1, 1-20.

[11] Evans, L. C., Gangbo, W., *Differential Equations Methods for the Monge–Kantorovich Mass Transfer Problem*, Mem. Amer. Math. Soc., Vol. **137** (1999).

[12] Gangbo, W., McCann, R. J., *The geometry of optimal transportation*, Acta Math., **177** (1996), 113–161.

[13] Sudakov, V. N., Geometric problems in the theory of infinite-dimensional probability distributions. Cover to cover translation of Trudy Mat. Inst. Steklov 141 (1976). Proc. Steklov Inst. Math. 1979, no. 2, i–v, 1–178.

[14] Villani, C., Topics in optimal transportation. Graduate Studies in Mathematics, 58, American Mathematical Society (2003)

G.C. CEREMADE, Université Paris Dauphine, Pl. de Lattre de Tassigny, 75775, Paris Cedex 16, FRANCE, `carlier@ceremade.dauphine.fr,`

L.D.P. Dipartimento di Matematica Applicata, Universitá di Pisa, Via Buonarroti 1/c, 56127 Pisa, ITALY, `depascal@dm.unipi.it,`

F.S. CEREMADE, Université Paris Dauphine, Pl. de Lattre de Tassigny, 75775, Paris Cedex 16, FRANCE, `filippo@ceremade.dauphine.fr.`