ÉCOLE POLYTECHNIQUE
UNIVERSITÉ PARIS-SACLAY

Journal de l'École polytechnique
Mathématiques

Robert J. McCann & Ludovic Rifford

The intrinsic dynamics of optimal transport

Tome 3 (2016), p. 67-98.

<http://jep.cedram.org/item?id=JEP_2016__3__67_0>

Journal de l'École polytechnique
Mathématiques

Tome 3, 2016, p. 67–98 DOI: 10.5802/jep.29

THE INTRINSIC DYNAMICS OF OPTIMAL TRANSPORT

by Robert J. McCann & Ludovic Rifford

Abstract. — The question of which costs admit unique optimizers in the Monge-Kantorovich problem of optimal transportation between arbitrary probability densities is investigated. For smooth costs and densities on compact manifolds, the only known examples for which the optimal solution is always unique require at least one of the two underlying spaces to be homeomorphic to a sphere. We introduce a (multivalued) dynamics which the transportation cost induces between the target and source space, for which the presence or absence of a sufficiently large set of periodic trajectories plays a role in determining whether or not optimal transport is necessarily unique. This insight allows us to construct smooth costs on a pair of compact manifolds with arbitrary topology, so that the optimal transportation between any pair of probability densities is unique.

Résumé (La dynamique intrinsèque du transport optimal). — Nous nous intéressons aux coûts pour lesquels la solution du problème de transport optimal de Monge-Kantorovitch entre deux mesures de probabilités est unique. À l'heure actuelle, les seuls exemples connus de tels coûts lisses sur des variétés compactes nécessitent que l'une des variétés soit homéomorphe à une sphère. Nous introduisons une dynamique (multivaluée) associée au coût et exhibons des propriétés suffisantes pour l'unicité d'un plan de transport optimal. Cette approche nous permet de construire des coûts lisses sur des variétés compactes quelconques pour lesquels l'unicité d'un plan de transport optimal est assurée.

Contents

Mathematical subject classification (2010). — 49Q20, 28A35.
Keywords. — Optimal transport, Monge-Kantorovitch problem, optimal transport map, optimal transport plan, numbered limb system, sufficient conditions for uniqueness.

The authors thank the University of Nice Sophia-Antipolis, Berkeley's Mathematical Sciences Research Institute (MSRI) and Toronto's Fields' Institute for the Mathematical Sciences for their kind hospitality during various stages of this work. They acknowledge partial support of RJM's research by Natural Sciences and Engineering Research Council of Canada Grant 217006-08, and, during their Fall 2013 residency at MSRI, by the National Science Foundation under Grant No. 0932078 000.

1. Introduction

Let M and N be smooth closed manifolds (meaning compact, without boundary) of dimensions m and $n \geqslant 1$ respectively, and $c : M \times N \to \mathbb{R}$ a continuous cost function. Given two probability measures μ and ν respectively on M and N, the Monge problem consists in minimizing the transportation cost

$$\int_{M\times N} c\big(x, T(x)\big)\, d\mu(x), \tag{1.1}$$

among all transport maps from μ to ν, that is, such that $T_\sharp \mu = \nu$. A classical way to prove existence and uniqueness of optimal transport maps is to relax the Monge problem into the Kantorovitch problem. That problem is a linear optimization problem under convex constraints, it consists in minimizing the transportation cost

$$\int_{M\times N} c(x,y)\, d\gamma(x,y), \tag{1.2}$$

among all transport plans between μ and ν, meaning γ belongs to the set $\Pi(\mu,\nu)$ of non-negative measures having marginals μ and ν. By classical (weak) compactness arguments, minimizers for the Kantorovitch problem always exist. A way to get existence and uniqueness of minimizers for the Monge problems is to show that any minimizer of (1.2) is supported on a graph. Assuming that c is Lipschitz and μ is absolutely continuous with respect to the Lebesgue measure, a condition which guarantees this graph property is the following non-smooth version [6] [22] of the TWIST condition

$$D_x^- c\big(\cdot, y_1\big) \cap D_x^- c\big(\cdot, y_2\big) = \varnothing \qquad \forall\, y_1 \neq y_2 \in N,\, \forall\, x \in M,$$

where $D_x^- c(\cdot, y_i)$ denotes the sub-differential of the function $x \mapsto c(x, y_i)$ at x. In this case, it is well-known how to use linear programming duality to prove that the Kantorovich minimizer is unique, and that Monge's infimum is attained [10] [16].

Examples of Lipschitz costs satisfying the non-smooth TWIST are given by any cost coming from variational problems associated with Tonelli Lagrangians of class $C^{1,1}$ (see [3]), like the square of Riemannian distances (see [19]). Those costs are never C^1 on compact manifolds such as $M \times N$. As a matter of fact, any cost $c : M \times N \to \mathbb{R}$ of class C^1 admits a triple $x \in M, y_1 \in N, y_2 \in N$, (take y_1 with $c(x, y_1) = \min\{c(x, \cdot)\}$ and y_2 with $c(x, y_2) = \max\{c(x, \cdot)\}$) such that

$$\frac{\partial c}{\partial x}\big(x, y_1\big) = \frac{\partial c}{\partial x}\big(x, y_2\big),$$

violating the non-smooth TWIST condition. Indeed, we shall show the following holds.

Theorem 1.1 (Non-genericity of twist). — *Let $c : M \times N \to [0, \infty)$ be a cost function of class C^2. Assume that* $\dim M = \dim N$ *and*

$$\exists\, (\overline{x}, \overline{y}) \in M \times N \quad \textit{such that} \quad \frac{\partial^2 c}{\partial x \partial y}(\overline{x}, \overline{y}) \quad \textit{is invertible.} \tag{1.3}$$

Then there is a pair μ, ν of probability measures respectively on M and N which are both absolutely continuous with respect to the Lebesgue measure for which there is a

unique optimal transport plan for (1.2) and such that this plan is not supported on a graph. The set of costs c satisfying (1.3) is open and dense in $C^2(M \times N; \mathbb{R})$.

The conclusion of Theorem 1.1 implies that solutions for the Monge problem with smooth cost do not generally exist in a compact setting. The purpose of the present paper is to study sufficient conditions on the cost for uniqueness of the Kantorovitch optimizer, and to exhibit smooth costs on arbitrary manifolds for which optimal plans are unique, despite the fact that such plans are not generally concentrated on graphs. Some examples of such costs have been given in [12] [1] (see also [5]). However, if uniqueness is to hold for arbitrary absolutely continuous μ and ν on M and N, all previous examples which we are aware of that involve smooth costs have required at least one of the two compact manifolds be homeomorphic to a sphere. Here we go far beyond this, to construct examples of such costs on compact manifolds whose topology can be arbitrary. Our main idea is to relate the uniqueness of the Kantorovitch optimizer to a multivalued dynamics induced by the cost which does not seem to have been considered previously.

Before stating our results, we need to introduce some definitions.

Denoting the non-negative integers by $\mathbb{N} = \{0, 1, 2, \ldots\}$ and the positive integers by $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$, we begin recalling the well-known notion of c-cyclical monotonicity.

Definition 1.2 (c-cyclical monotonicity). — A set $S \subset M \times N$ is c-cyclically monotone when for all $I \in \mathbb{N}^*$ and $(x_i, y_i) \in S$ for $i = 1, \ldots, I$ with $x_{I+1} = x_1$, we have

$$\sum_{i=1}^{I} [c(x_{i+1}, y_i) - c(x_i, y_i)] \geqslant 0.$$

For given μ, ν and c, it is also well-known [11] that some closed c-cyclically monotone subset $S \subset M \times N$ contains the support of all optimizers to (1.2). Note that of course, any subset of a c-cyclically monotone set is c-cyclically monotone as well. We come now to the concepts which will play a major role.

Definition 1.3 (Alternant chains). — For each $(x, y) \in M \times N$ assume $c(x, \cdot)$ and $c(\cdot, y)$ are differentiable. Fixing $S \subset M \times N$, we call *chain* in S of length $L \geqslant 1$ (or L-*chain* for short) any ordered family of pairs

$$\big((x_1, y_1), \ldots, (x_L, y_L)\big) \in S^L$$

such that the set

$$\big\{(x_1, y_1), \ldots (x_L, y_L)\big\}$$

is c-cyclically monotone and for every $\ell = 1, \ldots, L-1$ there holds, either

$$x_\ell = x_{\ell+1} \quad \text{and} \quad y_\ell \neq y_{\ell+1} = y_{\min\{L, \ell+2\}} \quad \text{and} \quad \frac{\partial c}{\partial x}(x_\ell, y_\ell) = \frac{\partial c}{\partial x}(x_\ell, y_{\ell+1}), \tag{1.4}$$

or

$$y_\ell = y_{\ell+1} \quad \text{and} \quad x_\ell \neq x_{\ell+1} = x_{\min\{L, \ell+2\}} \quad \text{and} \quad \frac{\partial c}{\partial y}(x_\ell, y_\ell) = \frac{\partial c}{\partial y}(x_{\ell+1}, y_\ell). \tag{1.5}$$

The chain is called cyclic if its projections onto M and N each consist of $L/2$ distinct points, in which case L must be even with $y_L = y_1$ and $x_L \neq x_1$ (or $x_L = x_1$ and $y_L \neq y_1$).

Note the existence of any cyclic chain $((x_1, y_1), \dots (x_L, y_L))$ permits the construction of an infinite chain $\{(x_\ell, y_\ell)\}_{\ell \in \mathbb{N}^*}$ by

$$\big(x_{kL+\ell}, y_{kL+\ell}\big) := \big(x_\ell, y_\ell\big) \qquad \forall\, k \geqslant 1, \forall\, \ell \in \{1, \dots, L\}. \tag{1.6}$$

The concept of *alternant chain* is basically a refinement of Hestir and Williams notion of *axial path* [15], to incorporate a choice c of transportation cost. Our first result is the following.

Theorem 1.4 (Optimal transport is unique if long chains are rare)

Fix a cost $c \in C^1(M \times N)$. Choose Borel probability measures μ on M and ν on N, both absolutely continuous with respect to Lebesgue, and let Π_0 denote the set of all optimizers for (1.2) on $\Pi(\mu, \nu)$. Let $E_0 \subset M \times N$ be a σ-compact set which is negligible for all $\gamma \in \Pi_0$, and denote its complement by $\widetilde{\mathcal{S}} := (M \times N) \setminus E_0$. Let E_∞ denote the set of points which occur in k-chains in $\widetilde{\mathcal{S}}$ for arbitrarily large k. Then E_∞ and its projections $\pi^M(E_\infty)$ and $\pi^N(E_\infty)$ are Borel. If $\gamma(E_\infty) = 0$ for every $\gamma \in \Pi_0$, then Π_0 is a singleton.

Remark 1.5 (Extension to singular marginals). — When $c \in C^{1,1}$, we can relax the absolute continuity of μ and ν in the preceding theorem provided neither concentrates positive mass on a $c-c$ hypersurface. Here $c-c$ *hypersurface* refers to one which can be parameterized in local coordinates as the graph of a difference of convex functions [25] [11] [13].

Corollary 1.6 (Sufficient notions of rarity). — *The condition $\gamma(E_\infty) = 0$ in the statement of the theorem, and therefore its conclusions, follow from either $\mu(\pi^M(E_\infty)) = 0$ or $\nu(\pi^N(E_\infty)) = 0$.*

If there is a uniform bound K on the length of all chains in $M \times N$, then our theorem applies a fortiori with $\mathcal{S} = M \times N$ and $E_0 = \varnothing$, since $E_\infty = \varnothing$. We shall see this occurs in many cases of interest, including for the smooth costs that we construct on compact manifolds with arbitrary topology. The bound K will depend on the topology. On the other hand, an obstruction to the uniqueness of optimal plans is the existence of a non-negligible set of periodic orbits. As shown below, such a property is not typical: it fails to occur for costs c in a countable intersection $\mathcal{C}$ of open dense sets. Such a countable intersection is called *residual*.

Theorem 1.7 (Costs admitting cyclic chains are non-generic)

When $\dim M = \dim N$, there is a residual set $\mathcal{C}$ in $C^\infty(M \times N; \mathbb{R})$ such that no cost in $\mathcal{C}$ admits cyclic chains, and for every cost $c \in \mathcal{C}$, there is a nonempty closed set $\Sigma \subset M \times N$ of zero (Lebesgue) volume such that

$$\frac{\partial^2 c}{\partial x \partial y}(x, y) \quad \text{is invertible for any } (x, y) \in M \times N \setminus \Sigma.$$

In the terminology of Hestir and Williams [15], the absence of cyclic chains is sufficient to define (formally) a rooting set whose measurability would be sufficient for uniqueness. We refer the reader to Section 3 for further details on their approach and its aftermath [5] [1] [21]. We do not know if uniqueness of optimal plans between all absolutely continuous measures holds for generic costs. However, elaborating on a celebrated result by Mañé [18] in the framework of Aubry-Mather theory, we are able to prove that uniqueness of optimal transport plans holds for generic costs in C^k if the marginals are fixed. In C^0, such a result was known already to Levin [17].

Theorem 1.8 (Optimal transport between given marginals is generically unique)
Fix Borel probability measures on compact manifolds M and N. For each $k \in \mathbb{N} \cup \{\infty\}$, there exists a residual set $\mathcal{C} \subset C^k(M \times N; \mathbb{R})$ such that for every $c \in \mathcal{C}$, there is a unique optimal plan between μ and ν.

The paper is organized as follows. We provide examples of costs satisfying the above results in Section 2. We develop preliminaries on numbered limb systems and details on Hestir and Williams' rooting sets in Section 3. We give the proofs of Theorem 1.4 in Section 4, of Theorem 1.7 in Section 6, and finally of Theorem 1.8 in the appendix.

2. Examples and applications

2.1. Quadratic cost on a strictly convex set. — Let us begin by recasting an example of Gangbo and McCann [12] into the framework of (alternant) chains.

Fix $N \subset \mathbb{R}^{m+1}$. Let M be the boundary of a strictly convex body $\Omega \subset \mathbb{R}^{m+1}$, that is, a closed set which is the boundary of a bounded open convex set and such that for any $z, z' \in M$,

$$[z, z'] \subset M \quad \Longrightarrow \quad z = z',$$

where $[z, z']$ is the segment joining z to z'. We aim to show that for any measures μ and ν (μ being absolutely continuous w.r.t. the Hausdorff m-dimensional measure $\mathcal{H}^m$ measure on M), we have uniqueness of optimal plans for the cost

$$c(x, y) = \frac{1}{2}|x - y|^2 \qquad \forall\, (x, y) \in M \times N.$$

Let $\mathcal{P}(M \times N)$ denote the Borel probability measure on $M \times N$ and $\pi^M : M \times N \to M$ and $\pi^N : M \times N \to N$ the projections onto the first and second variables. Let μ and ν be probability measures on M and N. We recall that the support $\Gamma \subset M \times N$ of any plan $\overline{\gamma} \in \mathcal{P}(M \times N)$ minimizing

$$\inf\left\{ \int_{M \times N} c(x, y) d\gamma(x, y) \mid \gamma \in \Pi(\mu, \nu) \right\} \tag{2.1}$$

is c-cyclically monotone, which in the case $c(x, y) = |y - x|^2/2$ reads

$$\sum_{i=1}^{I} \langle y_i, x_i - x_{i+1} \rangle \geqslant 0,$$

for all positive integer I, $i=1,\dots,I$, $(x_i,y_i)\in A$, $x_{I+1}=x_1$. The uniqueness of optimal plans will follow easily from the next lemma.

Lemma 2.1 (Interior links are never exposed). — *Fix a hypersurface $M\subset\mathbb{R}^{m+1}$, possibly incomplete. For any submanifold $N\subset\mathbb{R}^{m+1}$ of dimension $n\leqslant m+1$, let $c(x,y)$ denote the restriction of $\frac{1}{2}|x-y|^2$ to $M\times N$. If $((x_0,y),(x_2,y),(x_2,y'),(x_4,y'))$ is a chain in $M\times N$, then no hyperplane strictly separates x_2 from $M\setminus\{x_2\}$.*

Proof. — To derive a contradiction, suppose $((x_0,y),(x_2,y),(x_2,y'),(x_4,y'))$ forms a chain in $M\times\mathbb{R}^{m+1}$, yet x_2 is strictly separated from $M\setminus\{x_2\}$ by a hyperplane with inward normal n_2, i.e.,

$$\langle x-x_2,n_2\rangle>0 \tag{2.2}$$

for all $x\in M\setminus\{x_2\}$. The chain conditions imply $y'-y=\alpha n_2$ for some $\alpha\in\mathbb{R}$.

On the other hand, pairwise monotonicity of the points in the chain imply

$$\langle x_4-x_2,y'-y\rangle=\alpha\langle x_4-x_2,n_2\rangle\geqslant 0$$
$$\langle x_2-x_0,y'-y\rangle=\alpha\langle x_2-x_0,n_2\rangle\geqslant 0.$$

Using (2.2) we deduce $\alpha\geqslant 0$ from the first inequality and $\alpha\leqslant 0$ from the second. But $\alpha=0$ yields $y'=y$, contradicting the definition of a chain. □

As a consequence we have:

Corollary 2.2 (Chain bounds for strictly convex hypersurfaces)

If c is the restriction of $|x-y|^2$ to $M\times N$ as above, where $M\subset\mathbb{R}^{m+1}$ is a strictly convex hypersurface, then $M\times N$ contains no chain of length $L\geqslant 5$. Moreover, the projection of any 4-chain in $M\times N$ onto N consists of three distinct points, while its projection onto M consists of two distinct points. If $N\subset\mathbb{R}^{m+1}$ is also a strictly convex hypersurface, then $M\times N$ contains no chain of length $L\geqslant 4$.

Proof. — If a chain of length $L\geqslant 5$ exists, it begins either with

$$((x_1,y_2),(x_3,y_2),(x_3,y_4),(x_5,y_4)) \tag{2.3}$$

or $((x_2,y_1),(x_2,y_3),(x_4,y_3),(x_4,y_5),(x_6,y_5))$. Since M is strictly convex, each point $x\in M$ is exposed, meaning it can be strictly separated from $M\setminus\{x\}$ by a hyperplane. In the first case Lemma 2.1 would be violated by the chain (2.3) since x_2 is an exposed point of M; in the second it would be violated by the chain $((x_2,y_3),(x_4,y_3),(x_4,y_5),(x_6,y_5))$ since x_4 is an exposed point of M. We are forced to conclude that no chain of length $L\geqslant 5$ can exist. Moreover, any chain of length $L=4$ in $M\times N$ must take the form $((x_2,y_1),(x_2,y_3),(x_4,y_3),(x_4,y_5))$ hence project onto three points $y_i\in N$. The y_i must all be distinct since $y_1\neq y_3\neq y_5$ from the definition of chain, while $y_5=y_1$ would make the chain cyclic, in which case it can be extended to an infinite chain (1.6) contradicting non-existence of a chain of length 5. The projection onto M therefore consists of the two points $x_2\neq x_4$, which are distinct by the definition of chain.

If $N \subset \mathbb{R}^{m+1}$ is also a strictly convex hypersurface then by symmetry, $M \times N$ can contain no chain which projects to more than two points on M and two points on N, hence no chain of length $L \geqslant 4$. □

Example 2.3. — Let us consider the example of the lake that already appeared in [12] and [6]. Let $M = N$ be the unit circle in the plane, that is, the circle centered at the origin of radius 1 equipped with the quadratic cost $c(x, y) = |y - x|^2/2$. Consider a small auxiliary circle centered on the vertical axis, for example the circle centered at $(0, -5/2)$ of radius $1/8$, denote by $\widetilde{\psi}$ the distance function to the disc D enclosed by the small circle (see Figure 2.1). By construction, $\widetilde{\psi}$ is convex and differentiable at every point of M with a gradient of norm 1.

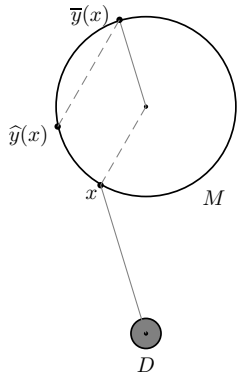

Figure 2.1. The lake

Then we set

$$\psi(x) := \widetilde{\psi}(x) - \frac{1}{2}|x|^2 \qquad \forall\, x \in M$$

and

$$\phi(y) := \min\big\{\psi(x) + c(x, y) \mid x \in M\big\} \qquad \forall\, y \in M.$$

By construction, we check that

$$\psi(x) = \max\big\{\phi(y) - c(x, y) \mid y \in M\big\} \qquad \forall\, x \in M. \tag{2.4}$$

Moreover, for every $x \in M$, the gradient $\overline{y}(x) := \nabla_x \widetilde{\psi} \in \mathbb{R}^2$ belongs to the set $\partial_c \psi(x) \subset M$ of optimizers for (2.4). As a matter of fact, we have by convexity of $\widetilde{\psi}$,

$$\widetilde{\psi}(x') - \widetilde{\psi}(x) \geqslant \langle \overline{y}(x), x' - x\rangle \qquad \forall\, x, x' \in \mathbb{R}^2.$$

Which can be written as

$$\psi(x) \leqslant \psi(x') + \frac{1}{2}\big|x' - \overline{y}(x)\big|^2 - \frac{1}{2}\big|x - \overline{y}(x)\big|^2 \qquad \forall\, x, x' \in \mathbb{R}^2.$$

Taking the minimum over $x' \in M$, we infer that ($\overline{y}(x) \in M$)

$$c\big(x, \overline{y}(x)\big) \leqslant \phi\big(\overline{y}(x)\big) - \psi(x) \qquad \forall\, x \in M$$

which, because $\phi - \psi \leqslant c$ implies that

$$c\big(x, \overline{y}(x)\big) = \phi\big(\overline{y}(x)\big) - \psi(x) \qquad \forall\, x \in M,$$

which means that $\overline{y}(x) = \nabla_x \widetilde{\psi}$ always belongs to $\partial_c \psi(x)$. For every $x \in M$, we set

$$\widehat{y}(x) := \overline{y}(x) + \lambda(x)x \in M,$$

where $\lambda(x) \geqslant 0$ is the largest nonnegative real number λ such that $\overline{y}(x) + \lambda x$ belongs to M (in other terms, $\widehat{y}(x)$ is the intersection of the open semi-line starting from $\overline{y}(x)$ with vector x if the intersection is nonempty and $\widehat{y}(x) = \overline{y}(x)$ otherwise). For every $x \in M$, the point $\widehat{y}(x)$ belongs to $\partial_c \psi(x)$ as well. As a matter of fact, by convexity of M, the fact that the normal to M at x is x itself and the convexity of $\widetilde{\psi}$, we have for every $x, x' \in M$,

$$\langle \widehat{y}(x), x' - x\rangle = \langle \overline{y}(x), x' - x\rangle + \lambda(x)\langle x, x' - x\rangle \leqslant \langle \overline{y}(x), x' - x\rangle \leqslant \widetilde{\psi}(x') - \widetilde{\psi}(x).$$

Proceeding as above we infer that $\widehat{y}(x)$ belongs to $\partial_c \psi(x)$. We can check easily that for every point x close to the south pole $(-1, 0)$ the points $\widehat{y}(x), \overline{y}(x)$ are distinct (see Figure 2.1). Proceeding as in the proof of Theorem 1.1, we can construct an example of optimal transport plan which is not concentrated on a graph.

2.2. Quadratic cost on nested strictly convex sets. — Let

$$\Omega_1 \subset \Omega_2 \subset \cdots \subset \Omega_L. \tag{2.5}$$

be a nested family of strictly convex bodies with differentiable boundaries in $\mathbb{R}^{m+1}$. Set $M = \bigcup_{i=1}^{L} U_i$, where $U_i = \partial\Omega_i \setminus \partial\Omega_{i-1}$ is an embedding of (a portion of) the unit sphere.

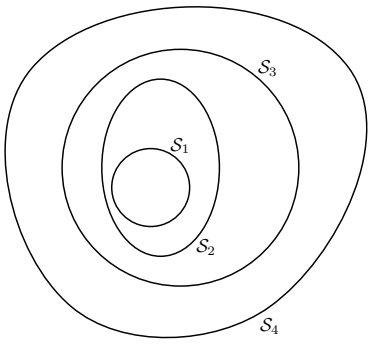

Figure 2.2. Nested convex sets

Lemma 2.4 (Chain length bounds for nested strictly convex boundaries)

If $c(x, y)$ denotes the restriction of $\frac{1}{2}|x-y|^2$ to C^1 manifolds $M_L, N \subset \mathbb{R}^{m+1}$, and $M_L := \partial\Omega_1 \cup \cdots \cup \partial\Omega_L$ is a union of boundaries of a nested sequence (2.5) *of strictly convex bodies $\Omega_i \subset \mathbb{R}^{m+1}$, then $M_L \times N$ contains no chain of length $4L+1$. Moreover, any chain of length $4L$ has projections onto M_L (respectively N) which consist of $2L$ (respectively $2L+1$) distinct points.*

Proof. — We prove the result by induction on L. Corollary 2.2 gives the result for $L = 1$. So assume that the property is proved for $L \geqslant 1$ and prove it for $L + 1$. Note that although M_L may not be a submanifold of $\mathbb{R}^{m+1}$ (if the boundaries of Ω_i and Ω_{i+1} intersect), it may be regarded as C^1 embedding of the disjoint union $\bigcup_{i=1}^{L} U_i$ of potentially incomplete manifolds $U_i = \partial\Omega_i \smallsetminus \partial\Omega_{i-1}$.

Any chain in $M_{L+1} \times N$ of length $4L + 5$ takes one of the forms

$$((x_1, y_2), (x_3, y_2), (x_3, y_4), \ldots, (x_{4L+3}, y_{4L+4}), (x_{4L+5}, y_{4L+4}), (x_{4L+5}, y_{4L+6})) \tag{2.6}$$

or

$$((x_2, y_1), (x_2, y_3), (x_4, y_3), \ldots, (x_{4L+4}, y_{4L+3}), (x_{4L+4}, y_{4L+5}), (x_{4L+6}, y_{4L+5})). \tag{2.7}$$

Strict convexity of $\partial\Omega_{L+1}$ shows any $x \in \partial\Omega_{L+1}$ can be separated from $M_{L+1} \smallsetminus \{x\}$ by a hyperplane. Lemma 2.1 therefore implies $\{x_4, \ldots, x_{4L+3}\} \subset M_L$, so that apart from possibly the first and last pairs of points, the chains (2.6)–(2.7) above are contained in $M_L \times N$. But this contradicts the inductive hypothesis, which asserts that $M_L \times N$ contains no chain of length $4L + 1$.

Similarly, if $M_{L+1} \times N$ contains a chain of length $2L+4$, it must take the form of the first 2L+4 points in (2.7) rather than (2.6); in the latter case $\{x_3, \ldots, x_{4L+3}\} \subset M_L$ whence $M_L \times N$ would contain a chain of length $4L+3$, contradicting the inductive hypothesis. In the former case, $\{x_4, \ldots, x_{4L+2}\} \subset M_L$, whence $M_L \times N$ contains a chain of length $4L$ which the inductive hypothesis asserts is comprised of $2L$ distinct points $X_4^{4L+2} := \{x_4, x_6, \ldots, x_{4L+2}\}$ and $2L+1$ distinct points $Y_3^{4L+3} := \{y_3, y_5, \ldots, y_{4L+3}\}$. Now x_2 and x_{4L+4} both lie outside $X_4^{4L+2} \subset M_L$, since otherwise $M_L \times N$ contains a chain longer than $4L$. Moreover $x_2 \neq x_{4L+4}$, since otherwise we can form a cycle (of length $4L + 2$), hence an infinite chain in $M_{L+1} \times N$, contradicting the length bound already established. Similarly, $y_1 \neq y_{4L+5}$ are disjoint from Y_3^{4L+3}, since otherwise we can extract a cycle and build an infinite chain in $M_{L+1} \times N$. □

2.3. Costs on manifolds

Lemma 2.5 (Diffeomorphism from interior of simplex to punctured sphere)

Fix the standard simplex

$$\Delta = \left\{(t_0, \ldots, t_m) \mid t_i \geqslant 0 \text{ and } \textstyle\sum_{i=0}^{m} t_i = 1)\right\}$$

and unit ball $\Omega = B_1(e_1) \subset \mathbb{R}^{m+1}$ *centered at* $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^{m+1}$. *There is a smooth map* $E : \Delta \to \partial\Omega$ *which acts as a diffeomorphism from* $\Delta \smallsetminus \partial\Delta$ *to* $\partial\Omega \smallsetminus \{0\}$ *such that* E *and all of its derivatives vanish on the boundary* $\partial\Delta$ *of the simplex:* $E(\partial\Delta) = \{0\}$.

Proof. — Let $f : [0, 1] \to [1, 2]$ be a smooth function satisfying the following properties:

(a) f is non-decreasing,
(b) $f(s) = 1$ for every $s \in [0, 1/2]$,
(c) $f(1) = 2$ and all the derivatives of f at $s = 1$ vanish.

Denote by D_m the closed unit disc of dimension m and by $\mathbb{S}^m \subset \mathbb{R}^{m+1}$ the unit sphere. We also denote by $\exp_N : T_N\mathbb{S}^m \to \mathbb{S}^m$ the exponential mapping from the north pole $N = (0, \dots, 0, 1)$ associated with the restriction of the Euclidean metric in $\mathbb{R}^{m+1}$ to $\mathbb{S}^m$. Then we set

$$F(v) = \exp_N \left[\frac{\pi}{2} f(|v|)\, v \right] \qquad \forall\, v \in D_m.$$

By construction, F is smooth on D_m, F is a diffeomorphism from $\mathrm{Int}(D_m)$ to $\mathbb{S}^m \setminus \{S\}$, where S denotes the south pole of $\mathbb{S}^m$, $F(\partial D_m) = \{S\}$ and all the derivatives of F on ∂D_m vanish. Therefore, in order to prove the lemma, it is sufficient to construct a Lipschitz mapping $G : \Delta \to D_m$ which is smooth on $\mathrm{Int}(\Delta)$, is a diffeomorphism from $\mathrm{Int}(\Delta)$ to $\mathrm{Int}(D_m)$, and sends $\partial\Delta$ to ∂D_m.

The simplex Δ is contained in the affine hyperplane

$$H = \big\{ \lambda = (t_0, \dots, t_m) \mid \textstyle\sum_{i=0}^m t_i = 1 \big\}.$$

Let $\bar{t} := (1/(m+1), \dots, 1/(m+1))$ be the center of Δ, we check easily that Δ is contained in the disc centered at $\bar{t}$ with radius $\sqrt{1 - 1/(m+1)}$. For every $t \in \Delta \setminus \{\bar{t}\}$, we set $u_t := (t - \bar{t})/|t - \bar{t}|$ and

$$\rho(t) := \min\big\{ s \geqslant 0 \mid t + s\, u_t \in \partial\Delta \big\} \qquad \forall\, t \in \Delta \setminus \{\bar{t}\}.$$

By construction, the function $\rho : \Delta \setminus \{\bar{t}\} \to [0, +\infty)$ is locally Lipschitz and satisfies for every unit vector $u \in \mathbb{S}^m \cap H_0$ (with $H = \bar{t} + H_0$),

$$\rho\big(\bar{t} + \alpha u\big) = \alpha_u - \alpha \qquad \forall\, \alpha \in (0, \alpha_u],$$

where $\alpha_u > 0$ is the unique $\alpha > 0$ such that $\bar{t} + \alpha u \in \partial\Delta$. We note that since $\Delta \subset \overline{B}\big(\bar{t}, \sqrt{1 - 1/(m+1)}\,\big)$, we have indeed $\alpha_u \in \big(0, \sqrt{1 - 1/(m+1)}\,\big]$ for every $u \in \mathbb{S}^m \cap H_0$. We also observe that the m-dimensional ball $H \cap \overline{B}\,\big(\bar{t}, 1/(2(m+1))\big)$ is contained in the interior of Δ and that $\rho \geqslant 1/(2(m+1))$ on that set. Pick a smooth function $g : [0, +\infty) \to [0, 1]$ satisfying the following properties:

(d) g is non-increasing,
(e) $g(s) = 1 - 3(m+1)s$ for every $s \in [0, 1/(4(m+1))]$,
(f) $g(s) = 0$ for every $s \geqslant 1/(2(m+1))$.

Let D be the m-dimensional unit disc in H centered at $\bar{t}$, define the function $G^0 : \Delta \to D$ by

$$G^0(t) = \bar{t} + [1 - g(\rho(t))]\, \big(t - \bar{t}\big) + g(\rho(t))\, u_t.$$

By construction, G^0 is Lipschitz and smooth on each ray starting from $\bar{t}$. Namely, for each unit vector $u \in \mathbb{S}^m \cap H_0$, we have

$$G^0_u(\alpha) := G^0\big(\bar{t} + \alpha u\big) = \bar{t} + [1 - g(\alpha_u - \alpha)]\, (\alpha\, u) + g(\alpha_u - \alpha)\, u \qquad \forall\, \alpha \in (0, \alpha_u].$$

The derivative of G^0 on each ray $\bar{t} + \mathbb{R}^+ u$ is given by

$$\frac{\partial G^0_u}{\partial \alpha}(\alpha) = [1 - g(\alpha_u - \alpha) + (\alpha - 1)\, g'(\alpha_u - \alpha)]\, u \qquad \forall\, \alpha \in (0, \alpha_u],$$

and there holds

$$1-g\left(\alpha_u-\alpha\right)+(\alpha-1)g'\left(\alpha_u-\alpha\right)$$
$$\geqslant 1-g\left(\alpha_u-\alpha\right)+\Big(\sqrt{1-\frac{1}{m+1}}-1\Big)g'\left(\alpha_u-\alpha\right)$$
$$\geqslant 1-g\left(\alpha_u-\alpha\right)+\Big(\frac{-1}{2(m+1)}\Big)g'\left(\alpha_u-\alpha\right)\geqslant\frac{3}{4},$$

by (d)–(f). Moreover, for every $u\in\mathbb{S}^m\cap H_0$, the ray $\overline{t}+\mathbb{R}^+u$ is invariant by G^0, $G^0_u(\alpha_u)=\overline{t}+u$, and in addition $G^0(t)=t$ whenever $t\in H\cap\overline{B}\left(\overline{t},1/(2(m+1))\right)$. In conclusion, G^0 is Lipschitz and bijective from Δ to D. If we work in polar coordinates $z=(\alpha,u)$ with $\alpha>0$ and $u\in\mathbb{S}^m\cap H_0$, then G^0 reads

$$\widetilde{G}^0(z)=\widetilde{G}^0(\alpha,u)=\left(G_u(\alpha),u\right),$$

for every $z\in\widetilde{\Delta}$, the domain of G^0 in polar coordinates (since G^0 coincides with the identity near $\overline{\lambda}$ we do not care about the singularity at $\alpha=0$). Thus for every z in the interior of $\widetilde{\Delta}$ where $\widetilde{G}^0$ is invertible, the Jacobian matrix of $\widetilde{G}^0$ at z, $J_z\widetilde{G}^0$ is triangular and invertible. Recall that for every z in the interior of $\widetilde{\Delta}$, the generalized Jacobian of $\widetilde{G}^0$ at z is defined as

$$\mathcal{J}_z\widetilde{G}^0:=\mathrm{conv}\Big\{\lim_k J_{z_k}\widetilde{G}^0\mid z_k\underset{k}{\longrightarrow}z,\,G\text{ diff. at }z_k\Big\}.$$

By the above discussion and Rademacher's Theorem, for every z in the interior of $\widetilde{\Delta}$, $\mathcal{J}_z\widetilde{G}^0$ is always a nonempty compact subset of $M_m(\mathbb{R})$ which contains only invertible matrices. In conclusion, for every $t\in\mathrm{Int}(\Delta)$ the generalized Jacobian of G^0 at t satisfies the same properties, it is a nonempty compact subset of $M_m(\mathbb{R})$ which contains only invertible matrices. Thanks to the Clarke Lipschitz Inverse Function Theorem [8], we infer that the Lipschitz mapping $G^0:\Delta\to D$ is locally bi-Lipschitz from $\mathrm{Int}(\Delta)$ to $\mathrm{Int}(D)$. It remains to smooth G^0 in the interior of Δ by fixing G^0 on the boundary $\partial\Delta$.

To this aim, consider a mollifier $\theta:\mathbb{R}^m\to\mathbb{R}$, that is, a smooth function satisfying the following three conditions:

(g) $\theta\geqslant 0$,
(h) $\mathrm{Supp}(\theta)\subset\overline{D}_m$,
(i) $\int_{\mathbb{R}^m}\theta(x)\,dx=1$.

The multivalued mapping $\lambda\in\mathrm{Int}(\Delta)\mapsto\mathcal{J}_\lambda G^0\in M_m(\mathbb{R})$ is upper-semicontinuous (its graph is closed in $\mathrm{Int}(\Delta)\times M_m(\mathbb{R})$) and is valued in the set of compact convex sets of invertible matrices. Hence, there is a continuous function $\varepsilon:\mathrm{Int}(\Delta)\to(0,\infty)$ such that for every $t\in\mathrm{Int}(\Delta)$ and every matrix $A\in M_m(\mathbb{R})$, the following holds

$$d\left(A,\mathrm{conv}\left(\{\mathcal{J}_\beta G\mid\beta\in B(t,\varepsilon(t))\}\right)\right)<\varepsilon(t)\quad\Longrightarrow\quad A\text{ is invertible.}\tag{2.8}$$

Consider also a smooth function $\nu:H\to\mathbb{R}^+$ such that:

(j) $\nu(t)=0$, for every $t\notin\mathrm{Int}(\Delta)$,
(k) $0<\nu(t)<\min\left\{d(t,\partial\Delta),\varepsilon(t)\right\}$, for every $t\in\mathrm{Int}(\Delta)$,
(l) for every $t\in\mathrm{Int}(\Delta)$, $|\nabla_t\nu|\leqslant\varepsilon(t)/K$, where $K>0$ is a Lipschitz constant for G^0.

Then, we define the function $G : \Delta \to D$ by (we identify H_0 with $\mathbb{R}^m$)

$$G(t) = \int_{\mathbb{R}^m} \theta(x)\, G^0\left(t - \nu(t)x\right)\, dx \qquad \forall\, t \in \Delta.$$

By construction, G is Lipschitz on Δ, it coincides with G^0 on $\partial\Delta$, it satisfies $G(\mathrm{Int}(\Delta)) \subset \mathrm{Int}(D)$, and it is smooth on $\mathrm{Int}(\Delta)$. For every $t \in \mathrm{Int}(\Delta)$, its Jacobian matrix at t is given by

$$J_t G = \int_{\mathbb{R}^m} \theta(x)\, J_{t-\nu(t)x} G^0 \cdot \Big(I_n - x \cdot \big(\nabla_t \nu\big)^*\Big)\, dx.$$

Hence, we have for every $t \in \mathrm{Int}(\Delta)$,

$$\left\| J_t G - \int_{\mathbb{R}^m} \theta(x)\, J_{t-\nu(t)x} G\, dx \right\| \leqslant K\ |\nabla_t \nu|$$

and

$$\int_{\mathbb{R}^m} \theta(x)\, J_{t-\nu(t)x} G\, dx \in \mathrm{conv}\big\{ J_\beta G \mid \beta \in B(t, \nu(t)) \big\}.$$

Using (2.8) and (j)–(l), we infer that G is a local diffeomorphism at every point of $\mathrm{Int}(\Delta)$. Moreover, G is surjective. If not, there is $y \in D$ such that y does not belong to the image of G. Since $G = G^0$ on $\partial\Delta$, y does not belong to ∂D. Thus there is $y' \in \partial G(\Delta) \setminus \partial D$. Since G is a local diffeomorphism at any pre-image of y', we get a contradiction. In conclusion, G is a Lipschitz mapping from Δ to D which sends bijectively $\partial\Delta$ to ∂D, which sends $\mathrm{Int}(\Delta)$ to $\mathrm{Int}(D)$, which is surjective, and which is a smooth local diffeomorphism at every point of $\mathrm{Int}(\Delta)$. Moreover, D is simply connected. Hence $G : \Delta \to D$ is a Lipschitz mapping which is a smooth diffeomorphism from $\mathrm{Int}(\Delta)$ to $\mathrm{Int}(D)$. We conclude easily. □

Proposition 2.6 (Smooth costs on arbitrary manifolds leading to unique optimal transport)

Fix smooth closed manifolds M, N. Then there exists a cost $c \in C^\infty(M \times N)$ such that: for any pair of Borel probability measures μ on M and ν on N which charge no $c-c$ hypersurfaces in their respective domains, the minimizer of (2.1) *is unique.*

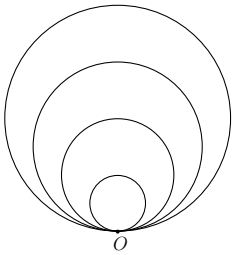

Figure 2.3. A bouquet of nested convex sets

Proof. — Let m and n denote the dimensions of M and N, and assume $m \geqslant n$ without loss of generality. Due to their smoothness, it is a classical result that both manifolds admit smooth triangulations [24] into finitely many (say k_M and k_N) simplices (by compactness).

For each $k \in \{1, 2, \dots, k_M\}$, dilating the map E of Lemma 2.5 by a factor of k induces a smooth map from the k-th simplex of M to the sphere $k\partial B_1(e_1)$ of radius k centered at $(k, 0, \dots, 0) \in \mathbb{R}^{m+1}$. Taken together, these k_M maps define a single smooth map $E_M : M \to \widetilde{M}$ where $\widetilde{M} = \bigcup_{k=1}^{k_M} \partial\Omega_k$ and $\Omega_k = kB_1(e_1) \subset \mathbb{R}^{m+1}$. This map acts as a diffeomorphism from the union of simplex interiors in M to $\widetilde{M} \smallsetminus \{\mathbf{0}_{m+1}\}$ while collapsing their boundaries onto the origin $\mathbf{0}_{m+1}$ in $\mathbb{R}^{m+1}$. Set $M_0 := E_M^{-1}(\mathbf{0}_{m+1})$.

Define the analogous map $E_N : N \to \widetilde{N} \subset \mathbb{R}^{n+1}$ where $\widetilde{N} = \bigcup_{k=1}^{k_N} k\partial B_1(e_1) \subset \mathbb{R}^{n+1}$ and $N_0 = E_N^{-1}(\mathbf{0}_{n+1})$. In case $n < m$, we embed $\mathbb{R}^{n+1}$ into $\mathbb{R}^{m+1}$ by identifying $\mathbb{R}^{n+1}$ with $\{(x_1, \dots, x_{m+1}) \in \mathbb{R}^{m+1} \mid x_{n+2} = \cdots = x_{m+1} = 0\}$.

The cost

$$c(x, y) := |E_M(x) - E_N(y)|^2/2$$

on $M \times N$ then satisfies the conclusions of the proposition. Its smoothness follows from that of E_M and E_N. Lemma 2.4 shows that no chains of length greater than $4k_M$ lie in $(M \smallsetminus M_0) \times (N \smallsetminus N_0)$. On the other hand, the simplex boundaries M_0 lies in a finite union of smooth hypersurfaces, hence are μ-negligible. Similarly, N_0 is ν-negligible. The desired conclusion now follows from Theorem 1.4. □

3. Preliminaries on numbered limb systems

3.1. Classical numbered limb systems. — The concept of numbered limb system was introduced by Hestir and Williams in [15]. Like Benes and Stepan [2], their aim was to find necessary and sufficient conditions on the support of a joint measure to guarantee its extremality in the space of measures which share its marginals.

Definition 3.1 (Numbered limb system). — Let X and Y be subsets of complete separable metric spaces. A relation $S \subset X \times Y$ is a numbered limb system if there are countable disjoint decompositions of $X$and Y,

$$X = \bigcup_{i=0}^{\infty} I_{2i+1} \quad \text{and} \quad Y = \bigcup_{i=0}^{\infty} I_{2i},$$

with a sequence of mappings

$$f_{2i} : \mathrm{Dom}(f_{2i}) \subset Y \longrightarrow X \quad \text{and} \quad f_{2i+1} : \mathrm{Dom}(f_{2i+1}) \subset X \longrightarrow Y$$

such that

$$S = \bigcup_{i=1}^{\infty} \big(\mathrm{Graph}(f_{2i-1}) \cup \mathrm{Antigraph}(f_{2i})\big) \tag{3.1}$$

and

$$\mathrm{Dom}(f_k) \cup \mathrm{Ran}(f_{k+1}) \subset I_k \qquad \forall\, k \geqslant 0. \tag{3.2}$$

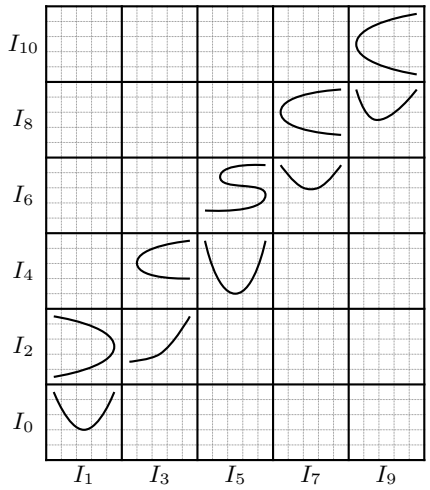

FIGURE 3.1. A numbered limb system with $N = 10$

The following statement from [1] extends and relaxes a celebrated characterization by Hestir and Williams [15]. Here $\pi^X(x,y) = x$ and $\pi^Y(x,y) = y$.

THEOREM 3.2 (Measures on measurable numbered limb systems are simplicial)

Let X and Y be Borel subsets of complete separable metric spaces, equipped with σ-finite Borel measures μ on X and ν on Y. Suppose there is a numbered limb system

$$S = \bigcup_{i=1}^{\infty} \big(\mathrm{Graph}(f_{2i-1}) \cup \mathrm{Antigraph}(f_{2i})\big) \tag{3.3}$$

with the property that $\mathrm{Graph}(f_{2i-1})$ *and* $\mathrm{Antigraph}(f_{2i})$ *are γ-measurable subsets of $X \times Y$ for each $i \geqslant 1$ and for every $\gamma \in \Gamma(\mu,\nu)$ vanishing outside of S. If the system has finitely many limbs or $\mu[X] < \infty$, then at most one $\gamma \in \Gamma(\mu,\nu)$ vanishes outside of S. If such a measure exists, it is given by $\gamma = \sum_{k=1}^{\infty} \gamma_k$ where for every $i \geqslant 1$,*

$$\begin{cases} \gamma_{2i-1} = \big(\mathrm{id}_X \times f_{2i-1}\big)_\sharp \eta_{2i-1}, & \gamma_{2i} = \big(f_{2i} \times \mathrm{id}_Y\big)_\sharp \eta_{2i}, \\ \eta_{2i-1} = \big(\mu - \pi^X_\sharp \gamma_{2i}\big)|_{\mathrm{Dom}\, f_{2i-1}}, & \eta_{2i} = \big(\nu - \pi^Y_\sharp \gamma_{2i+1}\big)|_{\mathrm{Dom}\, f_{2i}}. \end{cases} \tag{3.4}$$

Here η_k is a Borel measure on I_k and f_k is measurable with respect to the η_k completion of the Borel σ-algebra. If the system has $N < \infty$ limbs, $\gamma_k = 0$ for $k > N$, and η_k and γ_k can be computed recursively from the formula above starting from $k = N$.

The statement of Theorem 3.2 from Ahmad, Kim and McCann, like its antecedent in [15], give a sufficient condition for extremality. It is separated from Benes and Stepan [2] and Hestir and Williams' [15] necessary conditions for extremality by the γ-measurability assumed for the graphs and antigraphs (which is satisfied, for example, whenever the graphs and antigraphs are Borel.) For sets S of the form (3.3) whose graphs and antigraphs fail to be measurable, there may exist non-extremal measures vanishing outside of S, as shown by Hestir and Williams using the axiom

of choice [15]. Such issues are further explored by Bianchini and Caravenna [5] and Moameni [21], who arrive at their own criteria for extremality. Moameni's is closest in spirit to the approach developed below based on chain length: he gets his measurability by assuming the existence of a measurable Lyapunov function to distinguish different levels of the dynamics.

4. Proof of Theorem 1.4 and Remark 1.5

Since the source and target spaces are closed manifolds and the cost $c \in C^1$, Gangbo and McCann [11] provide a c-cyclically monotone compact set $\mathcal{S} \subset M \times N$ and Lipschitz potentials $\psi : M \to \mathbb{R}$ and $\phi : M \to \mathbb{R}$ which satisfy

$$\psi(x) = \max\big\{\phi(y) - c(x,y) \mid y \in N\big\} \qquad \forall\, x \in M, \tag{4.1}$$

$$\phi(y) = \min\big\{\psi(x) + c(x,y) \mid x \in M\big\} \qquad \forall\, y \in N, \tag{4.2}$$

$$\mathcal{S} \subset \partial_c\psi := \big\{(x,y) \in M \times N \mid c(x,y) = \phi(y) - \psi(x)\big\}, \tag{4.3}$$

such that any plan $\gamma \in \Pi(\mu,\nu)$ is optimal if and only if $\mathrm{Supp}(\gamma) \subset \mathcal{S}$. Indeed, we henceforth set $\mathcal{S}$ to be the smallest compact set with these properties.

We recall that the c-subdifferential of ψ at $x \in M$ and the c-superdifferential of ϕ at $y \in N$ are defined using (4.3):

$$\partial_c\psi(x) := \big\{y \in N \mid (x,y) \in \partial_c\psi\big\}$$

and

$$\partial^c\phi(y) := \big\{x \in M \mid (x,y) \in \partial_c\psi\big\}.$$

Note that since both ψ and ϕ are Lipschitz and μ and ν are both absolutely continuous with respect to Lebesgue, thanks to Rademacher's theorem, ψ and ϕ are differentiable almost everywhere with respect to μ and ν respectively. Let $\mathrm{Dom}\, d\psi$ denote the subset of M on which ψ is differentiable. Following Clarke [7], for every $x \in M$ (resp. $y \in N$), we denote by $D^*\psi(x)$ and $\partial\psi(x)$ (resp. $D^*\phi(y)$ and $\partial\phi(y)$) the limiting and generalized differentials of ψ at x (resp. ϕ at y) which are defined by (we proceed in the same way with ϕ)

$$D^*\psi(x) = \Big\{\lim_{k\to\infty} p_k \mid p_k = d\psi(x_k), x_k \longrightarrow x, x_k \in \mathrm{Dom}\, d\psi\Big\} \subset T_x^*M,$$

and

$$\partial\psi(x) = \mathrm{conv}\,(D^*\psi(x)) \subset T_x^*M.$$

By Lipschitzness, for every $x \in M$, the sets $D^*\psi(x)$ and $\partial\psi$ are nonempty and compact, and of course $\partial\psi(x)$ is convex. The next three propositions are relatively standard; the lemmas which follow them are new.

Proposition 4.1. — *For $c \in C^1$, the potentials ψ and ϕ of (4.1)–(4.2) satisfy:*

(i) *The mappings $x \in M \mapsto \partial\psi(x)$ and $y \in N \mapsto \partial\phi(y)$ have closed graph.*
(ii) *For every $x \in M$, ψ is differentiable at x if and only if $\partial\psi(x)$ is a singleton.*
(iii) *For every $y \in N$, ϕ is differentiable at y if and only if $\partial\phi(y)$ is a singleton.*
(iv) *The singular sets $M_0 := M \setminus \mathrm{Dom}\, d\psi$ and $N_0 := N \setminus \mathrm{Dom}\, d\phi$ are σ-compact.*

Proof of Proposition 4.1. — Assertion (i) is well-known [7], and follows easily from the definitions of $\partial\psi$ and $\partial\phi$. Let $x \in M$ be such that ψ is differentiable at x. From (4.1)–(4.3) we have

$$-\frac{\partial c}{\partial x}(x,y) = d\psi(x) \qquad \forall\, y \in \partial_c\psi(x). \tag{4.4}$$

Argue by contradiction and assume that $\partial\psi(x)$ is not a singleton. This means that $D^*\psi(x)$ is not a singleton too, let $p \neq q$ be two one-forms in $D^*\psi(x)$. Then there are two sequences $\{x_k\}_k, \{x'_k\}_k$ converging to x such that ψ is differentiable at x_k and x'_k and

$$\lim_{k\to\infty} d\psi(x_k) = p, \qquad \lim_{k\to\infty} d\psi(x'_k) = q.$$

For each k, there are $y_k \in \partial_c\psi(x_k)$, $y'_k \in \partial_c\psi(x'_k)$ such that

$$-\frac{\partial c}{\partial x}\big(x_k, y_k\big) = d\psi(x_k) \qquad \text{and} \qquad -\frac{\partial c}{\partial x}\big(x'_k, y'_k\big) = d\psi(x'_k).$$

By compactness of N, we may assume that the sequences $\{y_k\}_k, \{y'_k\}_k$ converge respectively to some $\overline{y} \in \partial_c\psi(x)$ and $\overline{y}' \in \partial_c\psi(x)$. Passing to the limit, we get

$$-\frac{\partial c}{\partial x}\big(x, \overline{y}\big) = p \qquad \text{and} \qquad -\frac{\partial c}{\partial x}\big(x, \overline{y}'\big) = q,$$

which contradicts (4.4). On the other hand, if $\partial\psi(x)$ is a singleton, then ψ is differentiable at x (indeed, $d\psi : \operatorname{Dom} d\psi \to T^*M$ is continuous at x, so x is a Lebesgue point for $d\psi \in L^\infty_{\text{loc}}$).

(iv) The set of x such that $\partial\psi(x)$ is not a singleton is σ-compact because the multivalued mapping $x \mapsto \partial\psi(x)$ has a closed graph, and the mapping $x \mapsto \operatorname{diam}(\partial\psi(x))$ is upper semicontinuous. For every whole number q, this implies those x with $\operatorname{diam}(\partial\psi(x)) \geqslant 1/q$ form closed subset of the compact manifold M. The singular set $M \smallsetminus \operatorname{Dom} d\psi$ is the union of such subsets over $q = 1, 2, \dots$ σ-compactness of $N \smallsetminus \operatorname{Dom} d\phi = \bigcup_{q=1}^\infty \{y \in N \mid \operatorname{diam}(\partial\phi(y)) \geqslant 1/q\}$ follows by symmetry. □

Proposition 4.2 (Differentiability a.e.) — *The sets*

$$M_0 := M \smallsetminus \operatorname{Dom} d\psi, \quad N_0 := N \smallsetminus \operatorname{Dom} d\phi \quad \textit{and} \quad M_0 \times N_0$$

are σ-compact, and $\mu(M_0) = \nu(N_0) = \gamma(M_0 \times N_0) = 0$ for every plan $\gamma \in \Pi(\mu,\nu)$.

Proof of Proposition 4.2. — Since $\mathcal{S}$ is compact, the σ-compactness of $M_0 \times N_0$ follows from that shown in Proposition 4.1(iv) for M_0 and N_0 (a product of unions being the union of the products). If μ and ν are absolutely continuous with respect to Lebesgue, Rademacher's theorem asserts $\mu[M_0] = 0$ and $\nu[N_0] = 0$. Otherwise $c \in C^{1,1}$, in which case Gangbo and McCann show the potentials ϕ and $-\psi$ are semi-convex [11], meaning their distributional Hessians admit local bounds from below in L^∞. In this case the conclusion of Rademacher's theorem can be sharpened: Zajíček [25] shows M_0 and N_0 to be contained in countably many $c-c$ hypersurfaces, on which μ and ν are assumed to vanish. Finally $\gamma(M_0 \times N_0) \leqslant \gamma(M_0 \times N) = \mu(M_0) = 0$. □

Since our manifolds M and N are compact, any open subset is σ-compact; in particular the complement of $\mathcal{S}$ is σ-compact. In view of this fact and the proposition preceding it, by enlarging E_0 if necessary we may henceforth assume (i) $(M \times N \setminus \mathcal{S}) \subset E_0$ and (ii) $(M_0 \times N) \cup (M \times N_0) \subset E_0$. Then $\widetilde{\mathcal{S}} := M \times N \setminus E_0$ ensures that for all pairs $(x, y) \in \widetilde{\mathcal{S}}$ we have differentiability of ψ at x and of ϕ at y.

Proposition 4.3 (Marginal cost is marginal price). — *For every* $(x, y) \in \widetilde{\mathcal{S}}$, (4.1)–(4.3) *imply*

$$d\psi(x) = -\frac{\partial c}{\partial x}(x, y) \quad \textit{and} \quad d\phi(x) = \frac{\partial c}{\partial y}(x, y). \tag{4.5}$$

Proof of Proposition 4.3. — Let $(x, y) \in \widetilde{\mathcal{S}}$, then we have by (4.1)–(4.2),

$$\begin{aligned} &\phi(y') - c(x, y') \leqslant \psi(x) \quad \forall\, y' \in N \quad \text{and} \quad \phi(y) - c(x, y) = \psi(x) \\ &\psi(x') + c(x', y) \geqslant \phi(y) \quad \forall\, x' \in M \quad \text{and} \quad \psi(x) + c(x, y) = \phi(y). \end{aligned}$$

We conclude easily since both ψ and ϕ are differentiable respectively at x and y. □

We call L-chain in $\widetilde{\mathcal{S}}$ any ordered family of pairs

$$\big((x_1, y_1), \dots (x_L, y_L)\big) \subset \widetilde{\mathcal{S}}^L$$

such that for every $\ell = 1, \dots, L-1$ there holds, either

$$\begin{cases} x_\ell = x_{\ell+1}, \\ y_\ell \neq y_{\ell+1} = y_{\min\{L, \ell+2\}} \end{cases} \quad \text{or} \quad \begin{cases} y_\ell = y_{\ell+1}, \\ x_\ell \neq x_{\ell+1} = x_{\min\{L, \ell+2\}}. \end{cases}$$

Note that by construction, the set of pairs of any L-chain in $\widetilde{\mathcal{S}}$ is c-cyclically monotone as a subset of $\mathcal{S}$, so by (4.5), any L-chain in $\widetilde{\mathcal{S}}$ is indeed an L-chain with respect to c (Definition 1.3). We define the level $\ell(x, y)$ of each $(x, y) \in \widetilde{\mathcal{S}}$ to be the supremum of all natural numbers $L \in \mathbb{N}^*$ such that there is at least one chain $((x_1, y_1), \dots (x_L, y_L))$ in $\widetilde{\mathcal{S}}$ of length L such that $(x, y) = (x_L, y_L)$. Moreover, given a chain $((x_1, y_1), \dots (x_L, y_L))$ with $L \geqslant 2$ in $\widetilde{\mathcal{S}}$, we say that (x_L, y_L) is a horizontal end if $y_L = y_{L-1}$ and a vertical end if $x_L = x_{L-1}$. We set

$$\widetilde{\mathcal{S}}_L := \big\{(x, y) \in \widetilde{\mathcal{S}} \mid \ell(x, y) \geqslant L\big\} \qquad \forall\, L \in \mathbb{N}^*,$$

and denote by $\widetilde{\mathcal{S}}^h_L$ (resp. $\widetilde{\mathcal{S}}^v_L$) the set of pairs $(x, y) \in \widetilde{\mathcal{S}}_L$ such that there exists a L-chain $((x_1, y_1), \dots (x_L, y_L))$ in $\widetilde{\mathcal{S}}$ such that $(x, y) = (x_L, y_L)$ and $y_{L-1} = y_L$ (resp. $x_{L-1} = x_L$). Although projections of Borel sets are not necessarily Borel (see [23]), the following lemma holds.

Lemma 4.4 (Borel measurability). — *The sets* $\widetilde{\mathcal{S}}_1 = \widetilde{\mathcal{S}}$ *and* $\widetilde{\mathcal{S}}^h_2, \widetilde{\mathcal{S}}^v_2, \dots, \widetilde{\mathcal{S}}^h_L, \widetilde{\mathcal{S}}^v_L$ *are Borel: each takes the form* $\bigcup_{p=1}^\infty \bigcap_{q=1}^\infty U_{p,q}$, *where the sets* $U_{1,1}$ *and* $U_{p-1,q} \subset U_{p,q} \subset U_{p,q-1}$ *are open for each* $p, q \geqslant 2$.

Proof of Lemma 4.4. — Given $L \geqslant 3$ odd, we shall show that $\widetilde{\mathcal{S}}^h_L$ has the asserted structure. The other cases are left to the reader. Endow the manifolds M and N with

Riemannian distances d_M and d_N, and let d_L denote the product distance on the product manifold $(M\times N)^L$. For every integer $p\geqslant 1$, denote by S_p the set of L-tuples

$$\big((x_1,y_1),\dots(x_L,y_L)\big)\subset\mathcal{S}^L$$

satisfying for every $i=1,\dots,L-1$,

$$\begin{cases}\text{for } \ell \text{ even: } x_{\ell+1}=x_\ell \text{ and } d_N(y_{\ell+1},y_\ell)\geqslant 1/p,\\ \text{for } \ell \text{ odd: } \ y_{\ell+1}=y_\ell \text{ and } d_M(x_{\ell+1},x_\ell)\geqslant 1/p.\end{cases}$$

Since $\mathcal{S}$ is compact, the set S_p is compact too.

On the other hand, σ-compactness of E_0 yields $(M\times N)\smallsetminus E_0=\bigcap_{q=1}^\infty V_q$ for a monotone sequence of open sets $V_q\subset V_{q-1}$. For every integer $q\geqslant 1$, we denote by S'_q the open set of L-tuples

$$\big((x_1,y_1),\dots(x_L,y_L)\big)\subset \big(V_q\big)^L.$$

A pair $(x,y)\in M\times N$ belongs to $\widetilde{\mathcal{S}}_L^h$ if and only if there is $p\geqslant 1$ such that

$$(x,y)\in \mathrm{Proj}_L\Big(\bigcap_q\big(S_p\cap S'_q\big)\Big),$$

where the projection $\mathrm{Proj}_L:(M\times N)^L\to\mathcal{S}$ is defined by

$$\mathrm{Proj}_L\big((x_1,y_1),\dots(x_L,y_L)\big):=(x_L,y_L).$$

For integers $p,q\geqslant 1$, let S_p^q the set of points which are at distance $d_L<1/q$ from S_p in $(M\times N)^L$. Since S_p is compact, for every p we have

$$\bigcap_q\big(S_p\cap S'_q\big)=\bigcap_q\big(S_p^q\cap S'_q\big)\,.$$

Moreover since for every p, the sequence of sets $\{S_p^q\cap S'_q\}$ is non-increasing with respect to inclusion, we have for every p,

$$\mathrm{Proj}_L\Big(\bigcap_q\big(S_p^q\cap S'_q\big)\Big)=\bigcap_q\mathrm{Proj}_L\left(S_p^q\cap S'_q\right).$$

The open sets $U_{p,q}=\mathrm{Proj}_L\left(S_p^q\cap S'_q\right)$ then have the asserted monotonicities $U_{p-1,q}\subset U_{p,q}\subset U_{p,q-1}$ with respect to p and q, and we find $\widetilde{\mathcal{S}}_L^h=\bigcup_{p=1}^\infty\bigcap_{q=1}^\infty U_{p,q}$ is the desired countable union of G_δ sets. □

Corollary 4.5 (Borel measurability of projections). — *For $i\geqslant 1$, the projections $\pi^M(\widetilde{\mathcal{S}}_i)$ and $\pi^N(\widetilde{\mathcal{S}}_i)$ of $\widetilde{\mathcal{S}}_i$ (and of $\widetilde{\mathcal{S}}_i^h,\widetilde{\mathcal{S}}_i^v$ if $i>1$) take the form $\bigcup_{p=1}^\infty\bigcap_{q=1}^\infty V_{pq}$, where $V_{1,1}$ and $V_{p-1,q}\subset V_{p,q}\subset V_{p,q-1}$ are open for each $p,q\geqslant 2$.*

Proof. — If $\widetilde{\mathcal{S}}_i^{h/v}=\bigcup_p\bigcap_q U_{p,q}^{h/v}$ for $i>1$ with $U_{p-1,q}^{h/v}\subset U_{p,q}^{h/v}\subset U_{p,q-1}^{h/v}$ then setting $V_{p,q}=\pi^M(U_{p,q})$ with $U_{p,q}=U_{p,q}^h\cup U_{p,q}^v$ shows $\pi^M(\widetilde{\mathcal{S}}_i)=\bigcup_p\bigcap_q V_{p,q}$ as desired. The other cases are similar. □

We recall that a set $S\subset M\times N$ is called a *graph* if for every $(x,y)\in S$ there is no $y'\neq y$ such that $(x,y')\in S$. A set $S\subset M\times N$ is called an *antigraph* if for every $(x,y)\in S$ there is no $x'\neq x$ such that $(x',y)\in S$. Any graph is the graph of a function defined on a subset of M and valued in N while any antigraph is the graph

of a function defined on a subset of N and valued in M. We call Borel graph or Borel antigraph any graph or antigraph which is a Borel set in $M \times N$. We are now ready to construct our numbered limb system.

Motivated by the inclusion $\widetilde{\mathcal{S}}_{k+1} \subset \widetilde{\mathcal{S}}_k$, we set $E_1 := \widetilde{\mathcal{S}}_1 \smallsetminus \widetilde{\mathcal{S}}_2$,

$$E_k := \widetilde{\mathcal{S}}_k \smallsetminus \widetilde{\mathcal{S}}_{k+1} \text{ and } \begin{cases} E_k^h := E_k \cap \widetilde{\mathcal{S}}_k^h, \quad E_k^{h-} = E_k \smallsetminus \widetilde{\mathcal{S}}_k^v, \\ E_k^v := E_k \cap \widetilde{\mathcal{S}}_k^v, \quad E_k^{v-} = E_k \smallsetminus \widetilde{\mathcal{S}}_k^h, \qquad \forall\, k \geqslant 2. \\ E_k^{hv} := E_k^h \cap E_k^v, \end{cases} \tag{4.6}$$

Notice that E_k consists precisely of the points in $\widetilde{\mathcal{S}}$ at level k. All these sets are Borel according to Lemma 4.4. Letting $E_\infty := \bigcap_{k=1}^\infty \widetilde{\mathcal{S}}_k$ gives a decomposition

$$\widetilde{\mathcal{S}} = E_\infty \cup E_1 \cup \left(\bigcup_{k=2}^\infty \left(E_k^{h-} \cup E_k^{hv} \cup E_k^{v-} \right) \right) \tag{4.7}$$

of $\mathcal{S}$ into disjoint Borel sets. The next lemma implies the E_k^h are graphs and the E_k^v are antigraphs; E_1 is simultaneously a graph and an antigraph, as are the E_k^{hv}.

Lemma 4.6 (Graph and antigraph properties)

(a) *Let $(x, y_i) \in E_i$ and $(x, y_j) \in E_j$ with $j \geqslant i \geqslant 1$ and $y_i \neq y_j$. Then $i \geqslant 2$. Moreover, if $j > i$ then $(x, y_i) \in E_i^h$ and $(x, y_j) \in E_{i+1}^{v-}$ so $j = i + 1$; otherwise $j = i$ and both $(x, y_i), (x, y_j) \in E_i^{v-}$.*

(b) *Similarly, suppose $(x_i, y) \in E_i$ and $(x_j, y) \in E_j$ with $j \geqslant i \geqslant 1$ and $x_i \neq x_j$. Then $i = 2$ and if $j > i$ then $(x_i, y) \in E_i^v$ and $(x_j, y) \in E_{i+1}^{h-}$ so $j = i + 1$; otherwise $j = i$ and both $(x, y_i), (x, y_j) \in E_i^{h-}$.*

Proof

(a) Let $(x, y_i) \in E_i$ and $(x, y_j) \in E_j$ with $j \geqslant i \geqslant 1$ and $y_i \neq y_j$. Then $(x, y_i), (x, y_j)$ form a 2-chain and both points lie in $\widetilde{\mathcal{S}}_2$, forcing $i \geqslant 2$. If $(x, y_j) \in E_j^h$, there is a j-chain in $\widetilde{\mathcal{S}}$ terminating in the horizontal end (x, y_j). Appending (x, y_i) to this chain produces a chain of length $j + 1$ with vertical end (x, y_i), whence $i = \ell(x, y_i) \geqslant j + 1$. This contradicts our hypothesis $i \leqslant j$. We therefore conclude $(x, y_j) \in E_j^{v-}$. Note that if $(x, y_i) \in E_i^h$, the same argument shows

$$j = \ell(x, y_j) \geqslant i + 1. \tag{4.8}$$

Whether or not this is true, $\mathcal{S}$ contains a j-chain

$$\big((x'_1, y'_1), \ldots, (x'_j, y'_j)\big) \tag{4.9}$$

terminating in the vertical end (x, y_j), so

$$x = x'_j = x'_{j-1}, \quad \text{and} \quad y_j = y'_j \neq y'_{j-1}.$$

Now either (c) $(x, y_i) \in E_i^h$ or (d) $(x, y_i) \in E_i^{v-}$. In case (c), we claim $y'_{j-1} = y_i$. Otherwise the sequence

$$\big((x'_1, y'_1), \ldots, (x'_{j-1}, y'_{j-1}), (x, y_i)\big)$$

would be a j-chain in $\widetilde{\mathcal{S}}$ of length $j \leqslant \ell(x, y_i) = i$, contradicting (4.8). Thus $y'_{j-1} = y_i$ and $i = \ell(x, y_i) \geqslant j - 1$, which implies equality holds in (4.8).

In case (d), $(x, y_i) \in E_i^{v-}$, we replace (x'_j, y'_j) with (x, y_i) in (4.9) to produce a chain of length $j \leqslant \ell(x, y_i) = i$, forcing $i = j$ as desired.

Part (b) of the lemma now follows from part (a) by symmetry. □

We define the graphs and antigraphs of our numbered limb system.

$$G_1 := E_1 \cup E_2^{h-},$$

$$G_{2i} := E_{2i}^{v-} \cup E_{2i}^{hv} \cup E_{2i+1}^{v-} = E_{2i}^{v} \cup E_{2i+1}^{v-}, \tag{4.10}$$

and
$$G_{2i+1} := E_{2i+1}^{h-} \cup E_{2i+1}^{hv} \cup E_{2i+2}^{h-} = E_{2i+1}^{h} \cup E_{2i+2}^{h-}$$

for all integers $i \in \mathbb{N}^*$, and adopt the convention $G_0 = \varnothing$.

Lemma 4.7 (Disjointness of domains and ranges). — *For $k \in \mathbb{N}$ set*

$$I_k = \begin{cases} \pi^M(G_k \cup G_{k+1}) & \text{if } k \text{ odd and} \\ \pi^N(G_k \cup G_{k+1}) & \text{if } k \text{ even.} \end{cases} \tag{4.11}$$

Then the subsets $\{I_{2i+1}\}_{i=0}^{\infty}$ of M are disjoint, as are the subsets $\{I_{2i}\}_{i=1}^{\infty}$ of N.

Proof. — For $i \in \mathbb{N}$, we shall show the sets $I_{2i+1} \subset M$ are disjoint. Disjointness of the subsets $\{I_{2i}\}_{i=1}^{\infty}$ of N is proved similarly, using Lemma 4.6(b).

To derive a contradiction, suppose $x \in I_{2i+1} \cap I_{2j+1}$ with $i < j$. Depending on whether $i = 0$ or $i \geqslant 1$, there exist

$$(x, y) \in E_1 \cup E_2^{h-} \cup E_{2i+1}^{h} \cup E_{2i+2}^{h-} \cup E_{2i+2}^{v} \cup E_{2i+3}^{v-}$$

and
$$(x, y') \in E_{2j+1}^{h} \cup E_{2j+2}^{h-} \cup E_{2j+2}^{v} \cup E_{2j+3}^{v-}.$$

Since $2i + 2 < 2j + 1$ disjointness of the E_k imply $y \neq y'$. Lemma 4.6(a) then asserts $(x, y') \in E_{2j+1}^{v-} \cup E_{2j+2}^{v-}$ — the desired contradiction. Thus the subsets $\{I_{2i+1}\}_{i=0}^{\infty}$ of M are disjoint. □

Lemma 4.8 (Numbered limbs). — *The Borel sets $\{G_{2i+1}\}_{i=0}^{\infty}$ of (4.10) form the graphs and $\{G_{2i}\}_{i=1}^{\infty}$ form the antigraphs of a numbered limb system: $G_{2i+1} = \mathrm{Graph}(f_{2i+1})$ and $G_{2i} = \mathrm{Antigraph}(f_{2i})$, with $\mathrm{Dom}\, f_k \cup \mathrm{Ran}\, f_{k+1} \subset I_k$ from (4.11) for all $k \in \mathbb{N}$.*

Proof. — The sets G_k are Borel by their construction (4.6), (4.10) and Lemma 4.4. If $i > 0$ we claim $G_{2i+1} := E_{2i+2}^{h-} \cup E_{2i+1}^{hv} \cup E_{2i+1}^{h-}$ is a graph: Let $(x, y) \neq (x, y')$ be distinct points in G_{2i+1}. Lemma 4.6(a) asserts that at least one of the two points lies in E_{2i+1}^{v-} or E_{2i+2}^{v-} — a contradiction. The fact that G_{2i} is an antigraph follows by symmetry, and the fact that G_1 is a graph is checked similarly.

We can therefore write $G_{2i+1} = \mathrm{Graph}(f_{2i+1})$ and $G_{2i} = \mathrm{Antigraph}(f_{2i})$ for some sequence of maps $f_k : \mathrm{Dom}\, f_k \to \mathrm{Ran}\, f_k$ with domains $\mathrm{Dom}\, f_k \subset M$ and ranges $\mathrm{Ran}\, f_k \subset N$ if k odd, and $\mathrm{Dom}\, f_k \subset N$ and $\mathrm{Ran}\, f_k \subset M$ if k even. The fact that $\mathrm{Dom}\, f_k \cup \mathrm{Ran}\, f_{k+1} \subset I_k$ follows directly from (4.11), while Lemma 4.7 implies disjointness of the $I_{2i+1} \subset M$ and of the $I_{2i} \subset N$. If $\widetilde{M} = M \setminus \bigcup_{i=0}^{\infty} I_{2i+1}$ or $\widetilde{N} := N \setminus \bigcup_{i=0}^{\infty} I_{2i}$

is non-empty, we replace I_0 by $I_0 \cup \widetilde{N}$ and I_1 by $I_1 \cup \widetilde{M}$ to complete our verification of the properties of a numbered limb system (Definition 3.1). □

Proof of Theorem 1.4 *and Remark* 1.5. — To recapitulate: Gangbo and McCann [11] provide a c-compact set $\mathcal{S}$ containing the support of every optimizer $\gamma \in \Pi_0$, and a pair of Lipschitz potentials (4.1)–(4.3) such that $\mathcal{S} \subset \partial_c \psi$. We take $\mathcal{S}$ to be the minimal such set without loss of generality. Proposition 4.2 shows $M_0 := M \setminus \operatorname{Dom} d\psi$ to be μ-negligible and $N_0 := N \setminus \operatorname{Dom} d\phi$ to be ν-negligible; both are σ-compact by Proposition 4.1. Without loss of generality, we therefore assume $M_0 \times N_0 \subset E_0$ and $M \times N \setminus \mathcal{S} \subset E_0$, the γ-negligible σ-compact set. Lemma 4.8 provides a decomposition (4.7) of $\widetilde{\mathcal{S}} := M \times N \setminus E_0$ into a numbered limb system consisting of Borel graphs and antigraphs — apart from a Borel set $E_\infty = \bigcap \widetilde{\mathcal{S}}_k$. But we have $\gamma(E_\infty) = 0$ for each $\gamma \in \Pi_0$ by hypothesis. Theorem 3.2 therefore asserts that at most one $\gamma \in \Pi_0$ vanishes outside $\widetilde{\mathcal{S}} \setminus E_\infty$. But since all $\gamma \in \Pi_0$ have this property, Π_0 must be a singleton. Finally, since $\widetilde{\mathcal{S}}_{k+1} \subset \widetilde{\mathcal{S}}_k$ we see $\pi^M(E_\infty) = \bigcap_{k=1}^\infty \pi^M(\widetilde{\mathcal{S}}_k)$ and $\pi^N(E_\infty)$ are Borel using Corollary 4.5. □

5. Proof of Theorem 1.1

Noting $\dim M = \dim N$, let $(\overline{x}, \overline{y}) \in M \times N$ be such that $\frac{\partial^2 c}{\partial y \partial x}(\overline{x}, \overline{y})$ is invertible. The mapping

$$F \,:\, y \in N \longmapsto \frac{\partial c}{\partial x}(\overline{x}, y) \in T^*_{\overline{x}} M$$

is C^1 and since its differential at $\overline{y}$ is not singular, its image contains an open set in $T^*_{\overline{x}} M$. By Sard's theorem (see [9, §3.4.3]), the image of critical points of F has Lebesgue measure zero, so we may assume without loss of generality that $F(\overline{y})$ is a regular value of F, meaning there is no y with $\frac{\partial^2 c}{\partial y \partial x}(\overline{x}, y)$ singular such that $F(y) = F(\overline{y})$. The next lemma then follows from topological arguments.

Lemma 5.1 (Generic failure of twist). — *Fix $(\overline{x}, \overline{y}) \in M \times N$ such that $F(\overline{y})$ is a regular value of $F(y) = \frac{\partial c}{\partial x}(\overline{x}, y)$. There is $\widehat{y} \in N$ such that $F(\widehat{y}) = F(\overline{y})$, i.e.,*

$$\frac{\partial c}{\partial x}\big(\overline{x}, \overline{y}\big) = \frac{\partial c}{\partial x}\big(\overline{x}, \widehat{y}\big). \tag{5.1}$$

Proof of Lemma 5.1. — We argue by contradiction and assume that

$$\forall\, y \in N, \quad y \neq \overline{y} \implies F(y) \neq F(\overline{y}).$$

Note that since F is a local diffeomorphism in a neighborhood of $\overline{y}$, the above condition still holds if we replace F by $\widetilde{F}$ a smooth (of class C^∞) regularization of F sufficiently close to F. So without loss of generality we may assume that F is smooth. Define the mapping $G : N \setminus \{\overline{y}\} \to \mathbb{S}^{n-1}$ by

$$G(y) := \frac{F(y) - F(\overline{y})}{|F(y) - F(\overline{y})|} \qquad \forall\, y \in N \setminus \{\overline{y}\}.$$

The mapping G is smooth, so by Sard's Theorem it has a regular value λ. Then the set

$$G^{-1}(\lambda) := \big\{y \in N \setminus \{\overline{y}\} \mid G(y) = \lambda\big\}$$

is a one dimensional submanifold of $N \setminus \{\overline{y}\}$. Moreover, since the differential of F at $\overline{y}$ is invertible, there are a open neighborhood $\mathcal{U}$ of $\overline{y}$ and a C^1 curve $\gamma : [-\varepsilon, \varepsilon] \to N$ with $\gamma(0) = \overline{y}$ and $\dot\gamma(0) \neq 0$ such that

$$G\left(\gamma(\pm t)\right) = \pm\lambda \qquad \forall\, t \in (0, \varepsilon]$$

and

$$G^{-1}(\lambda) \cap \mathcal{U} = \gamma\left((0, \varepsilon)\right).$$

This shows that the closure of $G^{-1}(\lambda)$ is a compact one dimensional submanifold whose boundary is $\overline{y}$. But the boundary of any compact submanifold of dimension one is a finite set with even cardinal (see [20]), a contradiction. □

We need now to construct a c-convex function whose c-subdifferential at each point near $\overline{x}$ takes values near both $\overline{y}$ and $\widehat{y}$. We note that since $F(\overline{y})$ is a regular value of $F(y) = \frac{\partial c}{\partial x}(\overline{x}, y)$ and $F(\widehat{y}) = F(\overline{y})$, both linear mappings $D_yF(\overline{y}), D_yF(\widehat{y})$ are invertible.

Lemma 5.2. — *There is a pair of functions $\psi : M \to \mathbb{R}$, $\phi : N \to \mathbb{R}$ such that*

$$\psi(x) = \max_{y \in N}\big\{\phi(y) - c(x, y)\big\} \qquad \forall\, x \in M \tag{5.2}$$

and

$$\phi(y) = \min\big\{\psi(x) + c(x, y) \mid x \in M\big\} \qquad \forall\, y \in N, \tag{5.3}$$

together with an open neighborhood $\overline{U}$ of $\overline{x}$, two open neighborhoods $\overline{V} \subset N$ of $\overline{y}$ and $\widehat{V} \subset N$ of $\widehat{y}$ with $\overline{V} \cap \widehat{V} = \varnothing$, and two C^1 diffeomorphisms

$$\overline{y} : \overline{U} \longrightarrow \overline{V}, \quad \widehat{y} : \overline{U} \longrightarrow \widehat{V}$$

with

$$\overline{y}(\overline{x}) = \overline{y} \quad \textit{and} \quad \widehat{y} = \widehat{y}(\overline{x}),$$

such that

$$\partial_c\psi(x) \cap \big(\overline{V} \cup \widehat{V}\big) = \big\{\overline{y}(x), \widehat{y}(x)\big\} \qquad \forall\, x \in \overline{U}. \tag{5.4}$$

Proof of Lemma 5.2. — Since we work locally in neighborhoods of $\overline{x}$, $\overline{y}$ and $\widehat{y}$, taking charts, we may assume that we work in $\mathbb{R}^n$. For every symmetric $n \times n$ matrix Q, there is a function $f : M \to \mathbb{R}$ of class C^2 such that

$$d_{\overline{x}} f = -\frac{\partial c}{\partial x}\big(\overline{x}, \overline{y}\big) = -\frac{\partial c}{\partial x}\big(\overline{x}, \widehat{y}\big) \tag{5.5}$$

and

$$\mathrm{Hess}_{\overline{x}}\, f = Q. \tag{5.6}$$

Let Q be fixed such that

$$Q + \frac{\partial^2 c}{\partial x^2}\big(\overline{x}, \overline{y}\big),\, Q + \frac{\partial^2 c}{\partial x^2}\big(\overline{x}, \widehat{y}\big) > 0, \tag{5.7}$$

we claim that there is a c-convex function $\psi : M \to \mathbb{R}$ which coincides with f in an neighborhood of f and which satisfies the required properties. Since both $\frac{\partial^2 c}{\partial x \partial y}(\overline{x}, \overline{y})$ and $\frac{\partial^2 c}{\partial x \partial y}(\overline{x}, \widehat{y})$ are invertible and (5.5) holds with $\overline{y} \neq \widehat{y}$, the Implicit Function Theorem yields a open neighborhood $\overline{U} \subset M$ of $\overline{x}$, two disjoint open neighborhoods $\overline{V}, \widehat{V} \subset N$ of $\overline{y}, \widehat{y}$ respectively, and two functions of class C^1

$$x \in \overline{U} \longmapsto \overline{y}(x) \in \overline{V}, \quad x \in \overline{U} \longmapsto \widehat{y}(x) \in \widehat{V}$$

such that

$$\text{(5.8)} \qquad \begin{cases} \overline{y}(\overline{x}) = \overline{y} \\ \widehat{y}(x) = \widehat{y} \end{cases} \quad \text{and} \quad \begin{cases} \dfrac{\partial c}{\partial x}(x, \overline{y}(x)) = -d_x f \\ \dfrac{\partial c}{\partial x}(x, \widehat{y}(x)) - d_x f \end{cases} \qquad \forall\, x \in \overline{U}.$$

Taking one derivative at $\overline{x}$ in the latter yields

$$\frac{\partial^2 c}{\partial x^2}\big(\overline{x}, \overline{y}\big) + \frac{\partial^2 c}{\partial y \partial x}\big(\overline{x}, \overline{y}\big) \frac{\partial \overline{y}}{\partial x}(\overline{x}) = \frac{\partial^2 c}{\partial x^2}\big(\overline{x}, \widehat{y}\big) + \frac{\partial^2 c}{\partial y \partial x}\big(\overline{x}, \widehat{y}\big) \frac{\partial \widehat{y}}{\partial x}(\overline{x}) = -\operatorname{Hess}_{\overline{x}} f,$$

which can be written as

$$\begin{cases} \dfrac{\partial \overline{y}}{\partial x}(\overline{x}) = -\Big(\dfrac{\partial^2 c}{\partial y \partial x}\big(\overline{x}, \overline{y}\big)\Big)^{-1} \Big[\operatorname{Hess}_{\overline{x}} f + \dfrac{\partial^2 c}{\partial x^2}\big(\overline{x}, \overline{y}\big)\Big] \\ \dfrac{\partial \widehat{y}}{\partial x}(\overline{x}) = -\Big(\dfrac{\partial^2 c}{\partial y \partial x}\big(\overline{x}, \widehat{y}\big)\Big)^{-1} \Big[\operatorname{Hess}_{\overline{x}} f + \dfrac{\partial^2 c}{\partial x^2}\big(\overline{x}, \widehat{y}\big)\Big]. \end{cases}$$

Therefore, by (5.6)–(5.7) we infer that $\frac{\partial \overline{y}}{\partial x}(\overline{x})$ and $\frac{\partial \widehat{y}}{\partial x}(\overline{x})$ are invertible. Then restricting $\overline{U}, \overline{V}, \widehat{V}$ if necessary, we may assume that the mappings

$$x \in \overline{U} \longmapsto \overline{y}(x) \in \overline{V}, \quad x \in \overline{U} \longmapsto \widehat{y}(x) \in \widehat{V}$$

are diffeomorphisms. Moreover, the functions of class C^2 given by

$$\overline{G} : x \in \overline{U} \longmapsto f(x) - f(\overline{x}) + c\,(x, \overline{y}(x)) - c\,(\overline{x}, \overline{y}(x))$$

and

$$\widehat{G} : x \in \overline{U} \longmapsto f(x) - f(\overline{x}) + c\,(x, \widehat{y}(x)) - c\,(\overline{x}, \widehat{y}(x))$$

satisfy (using (5.6)–(5.8))

$$\overline{G}(\overline{x}) = \widehat{G}(\overline{x}) = 0, \quad d_{\overline{x}} \overline{G} = d_{\overline{x}} \widehat{G} = 0, \quad \operatorname{Hess}_{\overline{x}} \overline{G} = \operatorname{Hess}_{\overline{x}} \widehat{G} = -\Big[Q + \frac{\partial^2 c}{\partial x^2}\big(\overline{x}, \overline{y}\big)\Big] < 0,$$

so we may also assume that

$$\text{(5.9)} \qquad \begin{cases} f(x') - f(x) + c\,(x', \overline{y}(x')) - c\,(x, \overline{y}(x')) < 0 \\ f(x') - f(x) + c\,(x', \widehat{y}(x')) - c\,(x, \widehat{y}(x')) < 0 \end{cases} \qquad \forall\, x' \in \overline{U} \smallsetminus \{x\},\ \forall\, x \in \overline{U}.$$

As a matter of fact, freezing x in the first line of (5.9) and setting

$$\overline{G}_x(x') = f(x') - f(x) + c\,(x', \overline{y}(x')) - c\,(x, \overline{y}(x')) \qquad \forall\, x' \in \overline{U},$$

we check that for every $x \in \overline{U}$, we have

$$\overline{G}_x(x) = 0, \quad d_x \overline{G}_x = d_x f + \frac{\partial c}{\partial x}\big(x, \overline{y}(x)\big) = 0$$

and for every $x' \in \overline{U}$

$$d_{x'}\overline{G}_x = d_{x'}f + \frac{\partial c}{\partial x}\big(x', \overline{y}(x')\big) + \Big[\frac{\partial c}{\partial y}\big(x', \overline{y}(x')\big) - \frac{\partial c}{\partial y}\big(x, \overline{y}(x')\big)\Big]\frac{\partial \overline{y}}{\partial x}(x')$$

which implies

$$\mathrm{Hess}_x\, \overline{G}_x = \mathrm{Hess}_x\, f + \frac{\partial^2 c}{\partial x^2}\big(x, \overline{y}(x)\big) + \frac{\partial^2 c}{\partial y \partial x}\big(x, \overline{y}(x)\big)\frac{\partial \overline{y}}{\partial x}(x)$$

Define the functions $\phi_0 : N \to \mathbb{R}$ and $\psi : M \to \mathbb{R}$ by

$$\phi_0(y) := \begin{cases} f\left(\overline{y}^{-1}(y)\right) + c\left(\overline{y}^{-1}(y), y\right) & \text{if } y \in \overline{V} \\ f\left(\widehat{y}^{-1}(y)\right) + c\left(\widehat{y}^{-1}(y), y\right) & \text{if } y \in \widehat{V} \\ -\infty & \text{if } y \notin \overline{V} \cup \widehat{V} \end{cases} \qquad \forall\, y \in N$$

and

$$\psi(x) = \max_{y \in N}\big\{\phi_0(y) - c(x, y)\big\} \qquad \forall\, x \in M.$$

We observe that we have for every $x \in \overline{U}$,

$$\phi_0\big(\overline{y}(x)\big) - c\big(x, \overline{y}(x)\big) = \phi_0\big(\widehat{y}(x)\big) - c\big(x, \widehat{y}(x)\big).$$

Then we have

$$\psi(x) = \max_{x' \in \overline{U}}\big\{\phi_0\big(\overline{y}(x')\big) - c\big(x, \overline{y}(x')\big)\big\} = \max_{x' \in \overline{U}}\big\{\phi_0\big(\widehat{y}(x')\big) - c\big(x, \widehat{y}(x')\big)\big\} \qquad \forall\, x \in M.$$

By the above construction and (5.9), we have for every $x \in \overline{U}$ and any $x' \in \overline{U} \setminus \{x\}$,

$$\begin{aligned} \phi_0\big(\overline{y}(x)\big) - c\big(x, \overline{y}(x)\big) &= f(x) \\ &> f(x') + c\big(x', \overline{y}(x')\big) - c\big(x, \overline{y}(x')\big) = \phi_0\big(\overline{y}(x')\big) - c\big(x, \overline{y}(x')\big). \end{aligned}$$

We infer that

$$\psi(x) = \phi_0\big(\overline{y}(x)\big) - c\big(x, \overline{y}(x)\big) = \phi_0\big(\widehat{y}(x)\big) - c\big(x, \widehat{y}(x)\big) = f(x) \qquad \forall\, x \in \overline{U}.$$

Setting

$$\phi(y) = \min\big\{\psi(x) + c(x, y) \mid x \in M\big\} \qquad \forall\, y \in N,$$

we check that (5.2)–(5.4) are satisfied.

$$\psi(x) = \max_{y \in N}\big\{\phi(y) - c(x, y)\big\} \qquad \forall\, x \in M. \qquad \square$$

Returning to the proof of the second case, let us consider an absolutely continuous probability measure μ on M whose support is contained in $\overline{U}$. Then define the nonnegative measures $\overline{\nu}, \widehat{\nu}$ on N by

$$\overline{\nu} := \frac{1}{2}\overline{y}_\sharp \mu \quad \text{and} \quad \widehat{\nu} := \frac{1}{2}\widehat{y}_\sharp \mu,$$

and set

$$\nu := \overline{\nu} + \widehat{\nu}.$$

Since the functions $\overline{y}$ and $\widehat{y}$ are diffeomorphism, ν is an absolutely continuous probability measure on N whose support in contained in $\overline{V} \cup \widehat{V}$. Moreover, the plan $\overline{\gamma}$ defined by

$$\overline{\gamma} := \frac{1}{2}\,(Id, \overline{y})_\sharp\,\mu + \frac{1}{2}\,(Id, \widehat{y})_\sharp\,\mu$$

has marginals μ and ν and is concentrated on the set of $(x, y) \in M \times N$ with $x \in \overline{U}$ and $y \in \partial_c\psi(x) \cap (\overline{V} \cup \widehat{V})$. By (5.2)–(5.3), any plan γ with marginals μ and ν satisfies

$$\begin{aligned}\int_{M\times N} c(x,y)\,d\gamma(x,y) &\geqslant \int_{M\times N} [\phi(y) - \psi(x)]\;d\gamma(x,y)\\ &= \int_N \phi(y)\,d\nu(y) - \int_M \psi(x)\,d\mu(x)\\ &= \int_{\overline{V}\cup\widehat{V}} \phi(y)\,d\nu(y) - \int_{\overline{U}} \psi(x)\,d\mu(x)\\ &= \int_{M\times N} c(x,y)\,d\overline{\gamma}(x,y),\end{aligned}$$

with equality in the first inequality if and only if γ is concentrated on the set of $(x, y) \in M \times N$ with $x \in \overline{U}$ and $y \in \partial_c\psi(x) \cap (\overline{V} \cup \widehat{V})$. This shows that $\overline{\gamma}$ is the unique optimal plan with marginals μ and ν.

It remains to show that the set of costs satisfying (1.3) is open and dense in $C^2(M \times N; \mathbb{R})$. The openness is obvious. Let us prove the density. Let c be fixed in $C^2(M \times N; \mathbb{R})$ such that (1.3) is not satisfied. Let $\overline{r} \in \{0, \dots, n-1\}$ be the maximum of the rank of $\frac{\partial^2 c}{\partial y \partial x}(x, y)$ for $(x, y) \in M \times N$, pick some $(\overline{x}, \overline{y}) \in M \times N$ such that

$$\operatorname{rank}\Big(\frac{\partial^2 c}{\partial y \partial x}(\overline{x}, \overline{y})\Big) = \overline{r}.$$

Since the rank mapping is lower semicontinuous, there are two open sets $U \subset M$ and $V \subset N$ such that the rank of $\frac{\partial^2 c}{\partial y \partial x}(x, y)$ is equal to $\overline{r}$ for any $(x, y) \in U \times V$. Moreover restricting U and V if necessary and taking local charts, we may assume that we work in $\mathbb{R}^n$. Let $X : V \to \mathbb{R}^n$ be the mapping defined by $X(y) = \frac{\partial c}{\partial x}(\overline{x}, y)$, for any $y \in V$. Doing a change of coordinates in x and y if necessary, we may assume that the $\overline{r} \times \overline{r}$ matrix

$$G = \Big(\frac{\partial X_i}{\partial y_j}(\overline{y})\Big)_{1 \leqslant i,j \leqslant \overline{r}}$$

is invertible. Define the mapping $G : V \to \mathbb{R}^n$ by

$$G(y_1, \dots, y_n) = (X(y)_1, \dots, X(y)_{\overline{r}}, y_{\overline{r}+1}, \dots, y_n) \qquad \forall\, y \in V.$$

The function G is of class C^1 and by construction the differential of G at $\overline{y}$ is invertible. Then G is a local diffeomorphism from a open neighborhood $V' \subset V$ of $\overline{y}$ onto an open neighborhood Z of $\overline{z} := G(\overline{y})$. The function X in z coordinates is given by

$$\widetilde{X}(z) := X_x\left(G^{-1}(z)\right) \qquad \forall\, z \in Z.$$

By construction, we have

$$\widetilde{X}(z)_i = z_i \qquad \forall\, i = 1, \dots, \overline{r},\ \forall\, z \in Z.$$

Therefore, since $\widetilde{X}$ has rank $\overline{r}$, the coordinates $\big(\widetilde{X}_{\overline{r}+1}, \dots, \widetilde{X}_n\big)$ do not depend upon the variables $z_{\overline{r}+1}, \dots, z_n$. Let $\delta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be the smooth function defined by

$$\delta(x,z) = \sum_{i=\overline{r}+1}^{n} x_i z_i \qquad \forall\, x, z \in \mathbb{R}^n$$

and let $\varphi : \mathbb{R}^n \to [0,1]$ be a cut-off function which is equal to 1 in a neighborhood of $G(\overline{y})$ and 0 outside Z. Then for every $\varepsilon > 0$ the function

$$\widetilde{c} \,:\, (x,z) \longmapsto c\big(x, G^{-1}(z)\big) + \varepsilon\varphi(z)\delta(x,z)$$

has a mixed partial derivative which is invertible at $(\overline{x}, \overline{z})$ and tends to c (in (x,z) coordinates) in C^2 topology as $\varepsilon > 0$ goes to zero.

6. Generic costs in smooth topology

The proof of Theorem 1.7 follows by classical transversality arguments. We refer the reader to [14] for further details on the results from Thom transversality theory that we use below.

Recall that $\dim M = \dim N = n$. Denote by $J^2(M \times N; \mathbb{R})$ the smooth manifold of 2-jets from $M \times N$ to $\mathbb{R}$ and denote by V the set consisting of 2-jets $((x,y), \lambda, p, H)$ where H is a symmetric matrix consisting of four $n \times n$ blocks

$$H = \begin{bmatrix} H_1 & H_2 \\ H_3 & H_4 \end{bmatrix},$$

with H_2 of corank $\geqslant 1$. The set V is closed and stratified by the smooth submanifolds

$$V_r := \big\{((x,y), \lambda, p, H) \mid \operatorname{rank}(H_2) = r\big\} \qquad \forall\, r = 0, \dots, n-1,$$

of codimension $\geqslant 1$. By the Thom Transversality Theorem (see [14, Th. 4.9, p. 54]), the set $\mathcal{C}_1$ of costs $c \in C^\infty(M \times N; \mathbb{R})$ such that $j^2 c(M \times N)$ is transverse to V is residual. For these costs the set $\Sigma := (j^2 c)^{-1}(V) \subset M \times N$ is stratified of codimension $\geqslant 1$ and it is nonempty. As a matter of fact, for every $x \in M$, the mapping $\frac{\partial c}{\partial x}(x, \cdot) : N \to T_x^* M$ is smooth and its image $\mathcal{I}$ is a compact subset of $T_x^* M$. Thus for every boundary point $p \in \partial\mathcal{I}$, the function $\frac{\partial c}{\partial x}(x, \cdot)$ cannot be a local diffeomorphism in a neighborhood of any $y \in N$ such that $\frac{\partial c}{\partial x}(x,y) = p$, which shows that for such y the linear mapping $\frac{\partial^2 c}{\partial y \partial x}(x,y)$ cannot be invertible. This shows that Σ is not empty. The fact that Σ is stratified of codimension $\geqslant 1$ (and so of zero measure) comes from the fact that it is the inverse image by $j^2 c : M \times N \to J^2(M \times N; \mathbb{R})$ of V which is transverse to $j^2(M \times N)$ (see [14, Th. 4.4, p. 52]).

Using a similar argument, we next show that the set of costs without periodic chains is residual in $C^\infty(M \times N; \mathbb{R})$.

Lemma 6.1 (Cyclic chains yield optimal alternatives). — *Let*

$$\big((x_1, y_1), \dots (x_L, y_L)\big) \in (M \times N)^L$$

be a chain with $x_2 = x_1, x_L \neq x_1$, *and* $y_L = y_1$. *Then* $L = 2K$ *for some integer* $K \geqslant 2$ *and*

$$\sum_{k=0}^{K-1} c\big(x_{2k+1}, y_{2k+1}\big) = \sum_{k=0}^{K-1} c\big(x_{2k+1}, y_{2k+2}\big).$$

Proof of Lemma 6.1. — The fact that $\ell = 2K$ is obvious from the definition of cycle and the fact that $x_2 = x_1$, $x_L \neq x_1$ and $y_L = y_1$. We have for any $k \in \{0, \dots, K-1\}$,

$$x_{2k+2} = x_{2k+1} \quad \text{and} \quad y_{2k+3} = y_{2k+2}.$$

Then, since the set $\{(x_1, y_1), \dots (x_L, y_L)\}$ is c-cyclically monotone, we have

$$\begin{aligned}\sum_{k=0}^{K-1} c\big(x_{2k+1}, y_{2k+1}\big) &\leqslant \sum_{k=0}^{K-1} c\big(x_{2k+1}, y_{2k+3}\big)\\ &= \sum_{k=0}^{K-1} c\big(x_{2k+1}, y_{2k+2}\big) = \sum_{k=0}^{K-1} c\big(x_{2k+2}, y_{2k+2}\big)\\ &\leqslant c\big(x_2, y_{2K}\big) + \sum_{k=1}^{K-1} c\big(x_{2k+2}, y_{2k}\big) = \sum_{k=0}^{K-1} c\big(x_{2k+1}, y_{2k+1}\big).\end{aligned}$$

We conclude easily. □

We need now to work with 1-multijets of smooth functions from $M \times N$ to $\mathbb{R}$. For every even integer $L = 2K \geqslant 4$, we denote by W_L the set of tuples

$$\Big(\big((x_1, y_1), \dots (x_L, y_L)\big), \big(\lambda_1, \dots, \lambda_L\big), \big((p_1^x, p_1^y), \dots (p_L^x, p_L^y)\big)\Big)$$

satisfying

$$(x_i, y_i) \neq (x_j, y_j) \qquad \forall\, i \neq j \in \{1, \dots, L\},$$

$$\begin{cases} x_{2k+2} = x_{2k+1} \\ y_{2k+3} = y_{2k+2}, \end{cases} \qquad \begin{cases} p^x_{2k+2} = p^x_{2k+1} \\ p^y_{2k+3} = p^y_{2k+2}, \end{cases}$$

for all $k \in \{0, K-1\}$ and

$$\sum_{k=0}^{K-1} \lambda_{2k+1} = \sum_{k=0}^{K-1} \lambda_{2k+2}.$$

The set W_L is a submanifold of $J^1_L(M \times N; \mathbb{R})$ of dimension

$$D = 4Kn + L - 1 = (2n+1)L - 1$$

and $J^1_L(M \times N; \mathbb{R})$ has dimension $(4n+1)L$. Thus W_L has codimension $2nL + 1$.

By the Multijet Transversality Theorem (see [14, Th. 4.13, p. 57]), for every $K = 2, 3, \dots$, the set $\mathcal{C}_K$ of costs c for which $j^1_{2K}c$ is transverse to W_{2K} is residual. The intersection

$$\mathcal{C} = \mathcal{C}_1 \cap \left(\bigcap_{K=2}^{\infty} \mathcal{C}_K\right)$$

satisfies the conclusions of Theorem 1.7.

Appendix. Generic uniqueness of optimal plans for fixed marginals

Elaborating on a celebrated result by Mañé [18] in the framework of Aubry-Mather theory, it is possible to prove that for fixed marginals the set of costs for which uniqueness of optimal transport plans holds is generic. Such a result was first obtained by Levin [17]. We include an argument here for comparison.

Let M and N be smooth closed manifolds (meaning compact, without boundary) of dimension $n \geqslant 1$, $c : M \times N \to \mathbb{R}$ be a cost function in $C^k(M \times N; \mathbb{R})$ with $k \in \mathbb{N} \cup \{\infty\}$, and μ, ν be two Borel probability mesures, we recall that $\Pi(\mu, \nu)$ denotes the set of probability measures in $M \times N$ with first and second marginals μ and ν. By the way, a measure on $M \times N$ is a continuous linear functional on the set of continuous functions $C^0(M \times N; \mathbb{R})$ and the set $E = C^0(M \times N; \mathbb{R})^*$ of such measures is equipped with the topology of weak-$*$ convergence saying that some sequence $(\pi_\ell)_\ell$ in E converges to $\pi \in E$ if and only if

$$\lim_{\ell \to \infty} \int_{M \times N} f \, d\pi_\ell = \int_{M \times N} f \, d\pi,$$

for every $f \in C^0(M \times N; \mathbb{R})$. The following is classical.

Lemma A.1. — *The set $\Pi(\mu, \nu)$ is a nonempty compact convex set in E.*

The following will also be useful. We refer the reader to [14] for the definition of the C^k-topology.

Lemma A.2. — *The mapping*

$$(\pi, c) \in E \times C^k(M \times N; \mathbb{R}) \longmapsto \int_{M \times N} c \, d\pi$$

is continuous with respect to the weak-$*$ topology on E and the C^k-topology on $C^k(M \times N; \mathbb{R})$. Moreover, for every $\pi_1, \pi_2 \in \Pi(\mu, \nu)$ with $\pi_1 \neq \pi_2$, there is $c \in C^k(M \times N; \mathbb{R})$ such that

$$\int_{M \times N} c \, d\pi_1 \neq \int_{M \times N} c \, d\pi_2.$$

For every $c \in C^k(M \times N; \mathbb{R})$ let $\mathcal{M}(c)$ be the set of optimal transport plans between μ and ν, that is,

$$\mathcal{M}(c) := \left\{ \pi \in \Pi(\mu, \nu) \mid \int_{M \times N} c \, d\pi \leqslant \int_{M \times N} c \, d\pi', \, \forall \, \pi' \in \Pi(\mu, \nu) \right\}.$$

By construction, $\mathcal{M}(c)$ is a nonempty compact convex subset of $\Pi(\mu, \nu)$.

Theorem A.3 (Levin). — *There exists a residual set $\mathcal{C} \subset C^k(M \times N; \mathbb{R})$ such that for every $c \in \mathcal{C}$, the set $\mathcal{M}(c)$ is a singleton.*

Here *residual* refers to a countable intersection of open dense sets. Theorem A.3 follows easily from results of Mañé [18] (or from arguments developed subsequently by Bernard and Contreras [4]). For sake of completeness we provide its proof which is based (following the approach of Bernard and Contreras [4]) on the next lemma. It

shows that near any given cost function can be found another for which the minimizing facet of $K := \Pi(\mu,\nu)$ has arbitrarily small diameter.

Lemma A.4. — *The weak-$*$ topology on K can be metrized by a distance $\widetilde{d}$ with the following property. Let $c_0 \in C^k(M\times N;\mathbb{R})$ be fixed. For every neighborhood U of c_0 in $C^k(M\times N;\mathbb{R})$ and every $\varepsilon>0$, there is $c\in U$ such that*

$$\operatorname{diam}\bigl(\mathcal{M}(c)\bigr) < \varepsilon.$$

Proof of Lemma A.4. — Let U and $\varepsilon>0$ be fixed. By compactness of $K := \Pi(\mu,\nu)$ with respect to the weak-$*$ topology, there is a sequence $\{f_\ell\}_{\ell\in\mathbb{N}}$ of continuous functions that defines a metric $\widetilde{d}$ on $\Pi(\mu,\nu)$ by

$$\widetilde{d}(\pi_1,\pi_2) = \sum_{\ell=0}^{\infty} \frac{1}{2^\ell}\left|\int_{M\times N} f_\ell\, d\pi_1 - \int_{M\times N} f_\ell\, d\pi_2\right| \qquad \forall\, \pi_1,\pi_2\in\Pi(\mu,\nu),$$

which is compatible with the weak topology. We claim that there is an integer $\overline{\ell}>0$ and

$$c_1,\dots,c_{\overline{\ell}} \in C^k(M\times N;\mathbb{R})$$

such that the continuous map

$$P_{\overline{\ell}} : \Pi(\mu,\nu) \longrightarrow \mathbb{R}^{\overline{\ell}}$$

defined by

$$P_{\overline{\ell}}(\pi) := \left(\int_{M\times N} c_1\, d\pi, \dots, \int_{M\times N} c_{\overline{\ell}}\, d\pi\right) \qquad \forall\, \pi\in\Pi(\mu,\nu),$$

satisfies

$$\operatorname{diam}\bigl(P_{\overline{\ell}}^{-1}(p)\bigr) < \varepsilon \qquad \forall\, p\in\mathbb{R}^{\overline{\ell}}. \tag{A.1}$$

where the latter refers to the diameter with respect to $\widetilde{d}$ of the set of measures in $\Pi(\mu,\nu)$ sent to p through $P_{\overline{\ell}}$. For every $c\in C^k(M\times N;\mathbb{R})$, let

$$W_c := \left\{(\pi_1,\pi_2) \mid \int_{M\times N} c\, d\pi_1 \neq \int_{M\times N} c\, d\pi_2\right\}.$$

By Lemma A.2, the sets W_c are open and their union covers the complement of the diagonal $D = \{(\pi,\pi) \mid \pi\in K\}$. Since this complement is open in the metrizable set $K\times K$, we can extract a countable subcovering from this covering. So there is a sequence $\{c_\ell\}_{\ell\in\mathbb{N}}$ such that

$$K\times K \smallsetminus D = \bigcup_{\ell\in\mathbb{N}} W_{c_\ell}. \tag{A.2}$$

We need to check that $P_{\overline{\ell}}$ satisfies (A.1) if $\overline{\ell}$ is large enough. If not, there are two sequences $\{\pi_\ell^1\}_\ell, \{\pi_\ell^2\}_\ell$ in K such that

$$P_\ell\bigl(\pi_\ell^1\bigr) = P_\ell\bigl(\pi_\ell^2\bigr) \quad\text{and}\quad \widetilde{d}\bigl(\pi_\ell^1,\pi_\ell^2\bigr) \geqslant \varepsilon \qquad \forall\,\ell.$$

Then up to taking subsequences, $\{\pi^1_\ell\}_\ell$ and $\{\pi^2_\ell\}_\ell$ converge respectively to some $\pi^1, \pi^2 \in K$ with $\widetilde{d}(\pi^1, \pi^2) \geqslant \varepsilon$. But by (A.2), there is m such that

$$\int_{M\times N} c_m \, d\pi^1 \neq \int_{M\times N} c_m \, d\pi^2.$$

But we have

$$P_m\big(\pi^1_\ell\big) = P_m\big(\pi^2_\ell\big) \qquad \forall\, \ell \geqslant m,$$

which passing to the limit gives $P_m\big(\pi^1\big) = P_m\big(\pi^2\big)$, a contradiction.

Let $K' := P_{\overline{\ell}}(K)$ which is a nonempty convex compact set in $\mathbb{R}^{\overline{\ell}}$, denote by $\Psi : \mathbb{R}^{\overline{\ell}} \to \mathbb{R}$ the function defined by

$$\Psi(x) := \begin{cases} \min\Big\{\int_{M\times N} c_0 \, d\pi \mid \pi \in K \text{ s.t. } P_{\overline{\ell}}(\pi) = x\Big\} & \text{if } x \in K' \\ +\infty & \text{if } x \notin K' \end{cases} \qquad \forall\, x \in \mathbb{R}^{\overline{\ell}},$$

and denote by $\Phi : \mathbb{R}^{\overline{\ell}} \to \mathbb{R}$ its conjugate, that is,

$$\begin{aligned} \Phi(y) &:= \sup_{x\in\mathbb{R}^{\overline{\ell}}} \big\{\langle y, x\rangle - \Psi(x)\big\} = \max_{x\in K'} \big\{\langle y, x\rangle - \Psi(x)\big\} \\ &= \max_{\pi\in K}\bigg\{\int_{M\times N} \sum_{\ell=1}^{\overline{\ell}} y_\ell c_\ell \, d\pi - \Psi\big(P_{\overline{\ell}}(\pi)\big)\bigg\}, \end{aligned}$$

for every $y \in \mathbb{R}^{\overline{\ell}}$. By construction, Φ is convex and finite on $\mathbb{R}^{\overline{\ell}}$, moreover for every $\overline{y} \in \mathbb{R}^{\overline{\ell}}$ and every $\overline{x} \in \mathbb{R}^{\overline{\ell}}$ such that $\Phi\big(\overline{y}\big) = \langle \overline{y}, \overline{x}\rangle - \Psi\big(\overline{x}\big)$, we have

$$\Phi\big(\overline{y}\big) + \langle y - \overline{y}, \overline{x}\rangle = \langle y, \overline{x}\rangle - \Psi\big(\overline{x}\big) \leqslant \Phi(y) \qquad \forall\, y \in \mathbb{R}^{\overline{\ell}}.$$

This means that $\overline{x}$ belongs to $\partial\Phi\big(\overline{y}\big)$, the subdifferential of Φ at $\overline{y}$. If in addition $\overline{\pi} \in K$ satisfies $P_{\overline{\ell}}\big(\overline{\pi}\big) = \overline{x}$ and

$$\int_{M\times N}\bigg(c_0 - \sum_{\ell=1}^{\overline{\ell}} \overline{y}_\ell c_\ell\bigg)\, d\overline{\pi} \leqslant \int_{M\times N}\bigg(c_0 - \sum_{\ell=1}^{\overline{\ell}} \overline{y}_\ell c_\ell\bigg)\, d\pi \qquad \forall\, \pi \in K,$$

then by definition of Ψ, we have

$$\int_{M\times N} \sum_{\ell=1}^{\overline{\ell}} \overline{y}_\ell c_\ell \, d\pi - \Psi\big(P_{\overline{\ell}}(\pi)\big) \leqslant \int_{M\times N} \sum_{\ell=1}^{\overline{\ell}} \overline{y}_\ell c_\ell \, d\overline{\pi} - \Psi\big(\overline{x}\big), \qquad \forall\, \pi \in K.$$

This means that

$$\mathcal{M}\bigg(c_0 - \sum_{\ell=1}^{\overline{\ell}} \overline{y}_\ell c_\ell\bigg) \subset P_{\overline{\ell}}^{-1}\left(\partial\Phi\big(\overline{y}\big)\right).$$

By Rademacher's theorem, for almost every $\overline{y} \in \mathbb{R}^{\overline{\ell}}$ the set $\partial\Phi\big(\overline{y}\big)$ is a singleton. We conclude by (A.1). □

Let us now prove Theorem A.3.

Proof of Theorem A.3. — For every integer $\ell > 0$, let us denote by $\mathcal{C}_\ell$ the set of $c \in C^k(M \times N; \mathbb{R})$ such that

$$\operatorname{diam}\big(\mathcal{M}(c)\big) < \frac{1}{\ell}.$$

By the continuity part in Lemma A.2, each set $\mathcal{C}_\ell$ is open and by Lemma A.4, it is dense as well. Then the set

$$\mathcal{C} := \bigcap_{\ell \in \mathbb{N}^*} \mathcal{C}_\ell$$

does the job. □

References

[1] N. Ahmad, H. K. Kim & R. J. McCann – "Optimal transportation, topology and uniqueness", *Bull. Sci. Math.* **1** (2011), no. 1, p. 13–32.

[2] V. Beneš & J. Štěpán – "The support of extremal probability measures with given marginals", in *Mathematical statistics and probability theory, Vol. A (Bad Tatzmannsdorf, 1986)*, Reidel, Dordrecht, 1987, p. 33–41.

[3] P. Bernard & B. Buffoni – "Optimal mass transportation and Mather theory", *J. Eur. Math. Soc. (JEMS)* **9** (2007), no. 1, p. 85–121.

[4] P. Bernard & G. Contreras – "A generic property of families of Lagrangian systems", *Ann. of Math. (2)* **167** (2008), no. 3, p. 1099–1108.

[5] S. Bianchini & L. Caravenna – "On the extremality, uniqueness and optimality of transference plans", *Bull. Inst. Math. Acad. Sinica* **4** (2009), no. 4, p. 353–454.

[6] P.-A. Chiappori, R. J. McCann & L. P. Nesheim – "Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness", *Econom. Theory* **42** (2010), no. 2, p. 317–354.

[7] F. H. Clarke – "Generalized gradients and applications", *Trans. Amer. Math. Soc.* **205** (1975), p. 247–262.

[8] ———, *Optimization and nonsmooth analysis*, Canadian Mathematical Society Series of Monographs and Advanced Texts, J. Wiley & Sons, Inc., New York, 1983.

[9] H. Federer – *Geometric measure theory*, Grundlehren Math. Wiss., vol. 153, Springer-Verlag New York Inc., New York, 1969.

[10] W. Gangbo – "Quelques problèmes d'analyse non convexe", Habilitation thesis, Université de Metz, 1995.

[11] W. Gangbo & R. J. McCann – "The geometry of optimal transportation", *Acta Math.* **177** (1996), no. 2, p. 113–161.

[12] ———, "Shape recognition via Wasserstein distance", *Quart. Appl. Math.* **58** (2000), no. 4, p. 705–737.

[13] N. Gigli – "On the inverse implication of Brenier-McCann theorems and the structure of $(P_2(M), W_2)$", *Methods Appl. Anal.* **18** (2011), no. 2, p. 127–158.

[14] M. Golubitsky & V. Guillemin – *Stable mappings and their singularities*, Graduate Texts in Math., vol. 14, Springer-Verlag, New York-Heidelberg, 1973.

[15] K. Hestir & S. C. Williams – "Supports of doubly stochastic measures", *Bernoulli* **1** (1995), no. 3, p. 217–243.

[16] V. L. Levin – "Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem", *Set-Valued Anal.* **7** (1999), no. 1, p. 7–32.

[17] ———, "On the generic uniqueness of an optimal solution in an infinite-dimensional linear programming problem", *Dokl. Akad. Nauk* **421** (2008), no. 1, p. 21–23.

[18] R. Mañé – "Generic properties and problems of minimizing measures of Lagrangian systems", *Nonlinearity* **9** (1996), no. 2, p. 273–310.

[19] R. J. McCann – "Polar factorization of maps on Riemannian manifolds", *Geom. Funct. Anal.* **11** (2001), no. 3, p. 589–608.

[20] J. W. Milnor – *Topology from the differentiable viewpoint*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, N.J., 1997, Revised reprint of the 1965 original.
[21] A. Moameni – "Supports of extremal doubly stochastic measures", preprint, 2014.
[22] L. Rifford – *Sub-Riemannian geometry and optimal transport*, Springer Briefs in Mathematics, Springer, Cham, 2014.
[23] S. M. Srivastava – *A course on Borel sets*, Graduate Texts in Math., vol. 180, Springer-Verlag, New York, 1998.
[24] H. Whitney – *Geometric integration theory*, Princeton University Press, Princeton, N.J., 1957.
[25] L. Zajíček – "On the differentiability of convex functions in finite and infinite dimensional spaces", *Czechoslovak Math. J.* **29** (1979), no. 3, p. 340–348.

Manuscript received March 26, 2015
accepted November 2, 2015

Robert J. McCann, Department of Mathematics, University of Toronto,
Toronto, Ontario, Canada M5S 2E4
E-mail : mccann@math.toronto.edu
Url : http://www.math.toronto.edu/mccann/

Ludovic Rifford, Université Nice Sophia Antipolis, Laboratoire J.A. Dieudonné, UMR CNRS 7351
Parc Valrose, 06108 Nice Cedex 02, France
& Institut Universitaire de France
E-mail : ludovic.rifford@math.cnrs.fr
Url : http://math.unice.fr/~rifford/