Canad. Math. Bull. 2024, pp. 1–10
http://dx.doi.org/10.4153/S0008439524000377

On the support of measures with fixed marginals with applications in optimal mass transportation

Abbas Moameni

Abstract. Let μ and ν be Borel probability measures on complete separable metric spaces X and Y, respectively. Each Borel probability measure γ on $X \times Y$ with marginals μ and ν can be described through its disintegration $(\gamma_x)_{x\in X}$ with respect to the initial distribution μ. Assume that μ is continuous, i.e., $\mu(\{x\}) = 0$ for all $x \in X$. We shall analyze the structure of the support of the measure γ provided card $(\mathrm{spt}(\gamma_x))$ is finitely countable for μ-a.e. $x \in X$. We shall also provide an application to optimal mass transportation.

1 Introduction

Let X and Y be Polish spaces equipped with Borel probability measures μ on X and ν on Y. Recall that a measure is called continuous if $\mu(\{x\}) = 0$ for all $x \in X$. Let $\Pi(\mu,\nu)$ be the set of Borel probability measures on $X \times Y$ which have X-marginal μ and Y-marginal ν. Let $\gamma \in \Pi(\mu,\nu)$. In what follows, we say that $\gamma \in \Pi(\mu,\nu)$ is concentrated on a set S if the outer measure of its complement is zero, i.e., $\gamma^*(S^c) = 0$. The support of the measure γ is denoted by $\mathrm{spt}(\gamma)$ and is the smallest closed set such that γ is zero on its complement. We now define precisely some notation describing measures concentrated on several graphs.

Definition 1.1 Let X and Y be Polish spaces with Borel probability measures μ on X and ν on Y. Let $k \in \mathbb{N} \cup \{\infty\}$. We say that a measure $\gamma \in \Pi(\mu,\nu)$ is concentrated on the graphs of measurable maps $\{G_i\}_{i=1}^k$ from X to Y, if there exists a sequence of measurable nonnegative functions $\{\alpha_i\}_{i=1}^k$ from X to $\mathbb{R}$ with $\sum_{i=1}^k \alpha_i(x) = 1$ (μ-almost surely) such that for each bounded continuous function $f : X \times Y \to \mathbb{R}$,

$$\int_{X\times Y} f(x,y)\, d\gamma = \sum_{i=1}^k \int_X \alpha_i(x) f(x, G_i x)\, d\mu.$$

In this case, we write $\gamma = \sum_{i=1}^k (Id \times G_i)_\#(\alpha_i \mu)$.

Received by the editors August 18, 2023; accepted May 15, 2024.
Published online on Cambridge Core May 29, 2024.
This work is supported by a grant from the Natural Sciences and Engineering Research Council of Canada.
AMS subject classification: **49Q20**, **49Q22**.
Keywords: Optimal transport, bi-stochastic measures, Choquet theory.

Setting $\Gamma = \mathrm{spt}(\gamma)$, for every $x \in X$, we denote by Γ_x the x-section of Γ, i.e.,

$$\Gamma_x = \{y \in Y;\ (x, y) \in \Gamma\}.$$

Here is our main result in this paper.

Theorem 1.2 *Let μ and ν be Borel probability measures on complete separable metric spaces X and Y, respectively. Assume that at least one of μ or ν is continuous. Let $\gamma \in \Pi(\mu, \nu)$ and $\Gamma = \mathrm{spt}(\gamma)$. The following assertions hold:*

1. *If there exists $m \in \mathbb{N}$ such that $\mathrm{card}(\Gamma_x) \leq m$ for μ-a.e. $x \in X$, then there exists $k \leq m$ and a sequence of Borel measurable maps $\{G_i\}_{i=1}^k$ from X to Y such that the measure γ is concentrated on their graphs.*
2. *If $\mathrm{card}(\Gamma_x) < \infty$ for μ-a.e. $x \in X$, then there exist $k \in \mathbb{N} \cup \{\infty\}$ and a sequence of Borel measurable maps $\{G_i\}_{i=1}^k$ from X to Y such that the measure γ is concentrated on their graphs.*

This theorem has direct applications in the theory of optimal transportation as it provides a precise description of the structure of optimal plans [1, 6, 7, 10–12]. Theorem 1.2 has a straightforward generalization to the multi-marginal case (see Corollary 2.9). We refer to [9] for applications of this result in multi-marginal mass transportation. We also remark that a weaker version of Theorem 1.2 is proved implicitly in [8]. The next section is devoted to the proof of the main theorem.

2 Preliminaries and the proof of Theorem 1.2

We shall need some important preliminaries from the theory of measures before proving Theorem 1.2. Let $(X, \mathcal{B}, \mu)$ be a finite, not necessarily complete measure space, and let (Y, Σ) be a measurable space. The completion of $\mathcal{B}$ with respect to μ is denoted by $\mathcal{B}_\mu$. When necessary, we identify μ with its completion on $\mathcal{B}_\mu$. The push forward of the measure μ by a map $T : (X, \mathcal{B}, \mu) \to (Y, \Sigma)$ is denoted by $T_\# \mu$, i.e.,

$$T_\# \mu(A) = \mu(T^{-1}(A)), \qquad \forall A \in \Sigma.$$

Definition 2.1 Let $T : X \to Y$ be $(\mathcal{B}, \Sigma)$-measurable, and let ν be a positive measure on Σ. We call a map $F : Y \to X$ a $(\Sigma_\nu, \mathcal{B})$-measurable section of T if F is $(\Sigma_\nu, \mathcal{B})$-measurable and $T \circ F = \mathrm{Id}_Y$.

If X is a topological space we denote by $\mathcal{B}(X)$ the set of Borel sets on X. The space of Borel probability measures on a topological space X is denoted by $\mathcal{P}(X)$. The following definition and proposition are essential in the sequel.

Definition 2.2 Let X be a Polish space, let $T : X \to X$ be a surjective Borel measurable map, and let μ be a positive finite measure on $\mathcal{B}(X)$. Denote by $\mathcal{S}(T)$ the set of all measurable sections of T, i.e.,

$$\mathcal{S}(T) = \Big\{F : \big(X, \mathcal{B}(X)_\mu\big) \to \big(X, \mathcal{B}(X)\big);\ T \circ F = Id_X\Big\}.$$

Let $\mathcal{K} \subset \mathcal{S}(T)$. We say that a measurable function $F : \big(X, \mathcal{B}(X)_\mu\big) \to \big(X, \mathcal{B}(X)\big)$ is generated by $\mathcal{K}$ if there exist a sequence $\{F_i\}_{i=1}^\infty \subset \mathcal{K}$ such that

$$X = \cup_{i=1}^\infty \big\{x \in X;\ F(x) = F_i(x)\big\}.$$

We also denote by $\mathcal{G}(\mathcal{K})$ the set of all functions generated by $\mathcal{K}$. It is easily seen that $\mathcal{K} \subseteq \mathcal{G}(\mathcal{K}) \subseteq \mathcal{S}(T)$.

Proposition 2.1 *Let X be a Polish space, let $T : X \to X$ be a surjective Borel measurable map, and let μ be a positive finite measure on $\mathcal{B}(X)$. Let $\mathcal{K}$ be a nonempty subset of $\mathcal{S}(T)$. Then, there exist $k \in \mathbb{N} \cup \{\infty\}$ and a sequence $\{F_i\}_{i=1}^k \subset \mathcal{G}(\mathcal{K})$ such that the following assertions hold:*

1. *For each $i \in \mathbb{N}$ with $i \le k$, we have $\mu(B_i) > 0$, where $\{B_i\}_{i=1}^k$ is defined recursively as follows:*

$$B_1 = X \quad \textit{and} \quad B_{i+1} = \Big\{x \in B_i;\ F_{i+1}(x) \notin \{F_1(x), \dots, F_i(x)\}\Big\} \quad \textit{provided } k > 1.$$

2. *For all $F \in \mathcal{G}(\mathcal{K})$, we have*

$$\mu\Big(\Big\{x \in B_{i+1}^c \setminus B_i^c;\ F(x) \notin \{F_1(x), \dots, F_i(x)\}\Big\}\Big) = 0.$$

3. *If $k \ne \infty$, then for all $F \in \mathcal{G}(\mathcal{K})$,*

$$\mu\Big(\Big\{x \in B_k;\ F(x) \notin \{F_1(x), \dots, F_k(x)\}\Big\}\Big) = 0.$$

Moreover, if either $k \ne \infty$ or, $k = \infty$ and $\mu(\cap_{i=1}^\infty B_i) = 0$, then for every $F \in \mathcal{G}(\mathcal{K})$, the measure $\rho_F = F_\# \mu$ is absolutely continuous with respect to the measure $\sum_{i=1}^k \rho_i$, where $\rho_i = F_{i\#}\mu$.

We refer to Proposition 3.1 in [8] for the proof of Proposition 2.1.

The following result shows that every $(\Sigma_\nu, \mathcal{B}(X))$-measurable map has a $(\Sigma, \mathcal{B}(X))$-measurable representation (see [2, Corollary 6.7.6]). Recall that a Souslin space is the image of a Polish space under a continuous mapping.

Proposition 2.2 *Let ν be a finite measure on a measurable space (Y, Σ), let X be a Souslin space, and let $F : Y \to X$ be a $(\Sigma_\nu, \mathcal{B}(X))$-measurable mapping. Then, there exists a mapping $G : Y \to X$ such that $G = F$ ν-a.e. and $G^{-1}(B) \in \Sigma$ for all $B \in \mathcal{B}(X)$.*

For a measurable map $T : (X, \mathcal{B}(X)) \to (Y, \Sigma, \nu)$ denote by $\mathcal{M}(T, \nu)$ the set of all measures λ on $\mathcal{B}$ so that T pushes λ forward to ν, i.e.,

$$\mathcal{M}(T, \nu) = \{\lambda \in \mathcal{P}(X);\ T_\# \lambda = \nu\}.$$

Evidently, $\mathcal{M}(T, \nu)$ is a convex set. A measure λ is an extreme point of $\mathcal{M}(T, \nu)$ if the identity $\lambda = \theta\lambda_1 + (1 - \theta)\lambda_2$ with $\theta \in (0, 1)$ and $\lambda_1, \lambda_2 \in \mathcal{M}(T, \nu)$ imply that $\lambda_1 = \lambda_2$. The set of extreme points of $\mathcal{M}(T, \nu)$ is denoted by $\operatorname{ext} \mathcal{M}(T, \nu)$.

We recall the following result from [4] in which a characterization of the set $\operatorname{ext} \mathcal{M}(T, \nu)$ is given.

https://doi.org/10.4153/S0008439524000377 Published online by Cambridge University Press

Theorem 2.3 *Let (Y,Σ,ν) be a probability space, let $(X,\mathcal{B}(X))$ be a Hausdorff space with a Radon probability measure λ, and let $T : X \to Y$ be a $(\mathcal{B}(X),\Sigma)$-measurable mapping. Assume that T is surjective and Σ is countably separated. The following conditions are equivalent:*

(i) *λ is an extreme point of $M(T,\nu)$;*
(ii) *there exists a $(\Sigma_\nu,\mathcal{B}(X))$-measurable section $F : Y \to X$ of the mapping T with $\lambda = F_\#\nu$.*

By making use of the Choquet theory in the setting of non-compact sets of measures [13], each $\lambda \in M(T,\nu)$ can be represented as a Choquet-type integral over $\operatorname{ext} M(T,\nu)$. Denote by $\Sigma_{\operatorname{ext} M(T,\nu)}$ the σ-algebra over $\operatorname{ext} M(T,\nu)$ generated by the functions $\rho \to \rho(B)$, $B \in \mathcal{B}(X)$. We have the following result (see [8] for a proof).

Theorem 2.4 *Let X and Y be complete separable metric spaces, and let ν be a probability measure on $\mathcal{B}(Y)$. Let $T : (X,\mathcal{B}(X)) \to (Y,\mathcal{B}(Y))$ be a surjective measurable mapping, and let $\lambda \in M(T,\nu)$. Then, there exists a probability measure ξ on $\sum_{\operatorname{ext} M(T,\nu)}$ such that for each $B \in \mathcal{B}(X)$,*

$$\lambda(B) = \int_{\operatorname{ext} M(T,\nu)} \rho(B)\, d\xi(\rho), \qquad \big(\rho \to \rho(B) \textit{ is measurable}\big).$$

We now recall the notion of isomorphisms for measures.

Definition 2.5 Assume that X and Y are topological spaces with Borel probability measures μ on X and ν on Y. We say that $(X,B(X),\mu)$ is isomorphic to $(Y,B(Y),\nu)$ if there exists a one-to-one map T of X onto Y such that for all $A \in B(X)$, we have $T(A) \in B(Y)$ and $\mu(A) = \nu\big(T(A)\big)$, and for all $B \in B(Y)$, we have $T^{-1}(B) \in B(X)$ and $\mu\big(T^{-1}(B)\big) = \nu(B)$.

Here is the well-known measure isomorphism theorem (see Theorem 17.41 in [5] for a proof).

Theorem 2.6 *Let μ be a Borel probability measure on a Polish space X. If μ is continuous, then $(X,B(X),\mu)$ and $([0,1],\lambda)$, where λ is Lebesgue measure, are isomorphic.*

Lemma 2.7 *Let $\gamma \in \Pi(\mu,\nu)$. If either μ or ν is continuous, then so is γ.*

Proof Assume that μ is continuous. Take $(x,y) \in X \times Y$. It follows that

$$\mu(\{x\}) = \gamma\big(\{x\} \times Y\big) \geq \gamma\big(\{x\} \times \{y\}\big),$$

from which the desired result follows. The proof is similar if ν is continuous. ∎

Proof of Theorem 1.2 We assume that μ is a continuous measure. It follows from Lemma 2.7 that γ is also continuous. It follows from Theorem 2.6 that the Borel measurable spaces $(X,\mathcal{B}(X),\mu)$ and $(X\times Y,\mathcal{B}(X\times Y),\gamma)$ are isomorphic. Thus, there exists an isomorphism $T = (T_1,T_2)$ from $(X,\mathcal{B}(X),\mu)$ onto $(X \times Y,$

https://doi.org/10.4153/S0008439524000377 Published online by Cambridge University Press

$\mathcal{B}(X \times Y), \gamma)$. It can be easily deduced that $T_1 : X \to X$ and $T_2 : X \to Y$ are surjective maps and

$$(T_1)_\# \mu = \mu \quad \& \quad (T_2)_\# \mu = \nu.$$

Consider the convex set

$$\mathcal{M}(T_1, \mu) = \{\lambda \in \mathcal{P}(X);\, (T_1)_\# \lambda = \mu\},$$

and note that $\mu \in \mathcal{M}(T_1, \mu)$. Since $\mu \in \mathcal{M}(T_1, \mu)$, it follows from Theorem 2.4 that there exists a probability measure ξ on $\sum_{\text{ext}\, M(T_1,\mu)}$ such that for each $B \in \mathcal{B}(X)$,

(1) $$\mu(B) = \int_{\text{ext}\, M(T_1,\mu)} \rho(B)\, d\xi(\rho), \qquad \big(\rho \to \rho(B) \text{ is measurable}\big).$$

Since $\Gamma = \text{spt}(\gamma)$, it follows that $T^{-1}(\Gamma)$ is a measurable subset of X with $\mu\big(T^{-1}(\Gamma)\big) = 1$. Let $A_\gamma \in \mathcal{B}(X)$ be the set such that $A_\gamma \subseteq T^{-1}(\Gamma)$ and for all $x \in A_\gamma$, the cardinality of the set Γ_x does not exceed m. It follows from the assumption that $\mu(A_\gamma) = 1$. Since $\mu(X \setminus A_\gamma) = 0$, it follows from (1) that

$$\int_{\text{ext}\, M(T_1,\mu)} \rho(X_1 \setminus A_\gamma)\, d\xi(\rho) = \mu(X \setminus A_\gamma) = 0,$$

and therefore there exists a ξ-full measure subset K_γ of $\text{ext}\, M(T_1, \mu)$ such that $\rho(X \setminus A_\gamma) = 0$ for all $\rho \in K_\gamma$. Let $\mathcal{S}(T_1)$ be the set of all sections of T_1 and define

$$\mathcal{K} := \{F \in \mathcal{S}(T_1);\ \exists \rho \in K_\gamma \text{ with } \mu = F_\# \rho\}.$$

Let $\mathcal{G}(\mathcal{K})$ be the set of all measurable sections of T_1 generated by $\mathcal{K}$ as in Definition 2.2. By Proposition 2.1, there exists a sequence $\{F_i\}_{i=1}^k \subset \mathcal{G}(\mathcal{K})$ with $k \in \mathbb{N} \cup \{\infty\}$ satisfying assertions (1)–(3) in that proposition. Let $B_\gamma := \cap_{i=1}^k F_i^{-1}(A_\gamma)$, and for each $k \in \mathbb{N} \cup \{\infty\}$, define

$$\mathbb{N}_k = \begin{cases} \{1, 2, \ldots, k\}, & k \in \mathbb{N}, \\ \mathbb{N}, & k = \infty. \end{cases}$$

Let $\rho_i := F_{i\#}\mu$ for each $i \in \mathbb{N}_k$. We shall now proceed with the proof in several steps.

Step I: In this step, we show that $\mu\big(B_\gamma\big) = 1$ and

(2) $$\big(x, T_2 \circ F_i(x)\big) \in \Gamma, \qquad \forall x \in B_\gamma,\ \ \forall i \in \mathbb{N}_k.$$

Note first that $\rho_i(X \setminus A_\gamma) = 0$ for each $i \in \mathbb{N}_k$. In fact, for a fixed $i \in \mathbb{N}_k$, since $F_i \in \mathcal{G}(\mathcal{K})$ there exists a sequence $\{F_{\sigma_j}\}_{j=1}^\infty \subset \mathcal{K}$ such that $X = \cup_{j=1}^\infty A_j$, where

$$A_j = \{x \in X;\ F_i(x) = F_{\sigma_j}\}.$$

https://doi.org/10.4153/S0008439524000377 Published online by Cambridge University Press

Let $\sigma_j \in K_\gamma$ be such that the map F_{σ_j} is a push-forward from σ_j to μ. It follows that

$$\begin{aligned}\rho_i(X\setminus A_\gamma) = \mu\big(F_i^{-1}(X\setminus A_\gamma)\big) &= \mu\big((\cup_{j=1}^{\infty} A_j)\cap F_i^{-1}(X\setminus A_\gamma)\big)\\ &\le \sum_{j=1}^{\infty}\mu\big(A_j\cap F_i^{-1}(X\setminus A_\gamma)\big)\\ &= \sum_{j=1}^{\infty}\mu\big(A_j\cap F_{\sigma_j}^{-1}(X\setminus A_\gamma)\big)\\ &\le \sum_{j=1}^{\infty}\mu\big(F_{\sigma_j}^{-1}(X\setminus A_\gamma)\big) = \sum_{j=1}^{\infty}\sigma_j(X\setminus A_\gamma) = 0.\end{aligned}$$

This proves that $\rho_i(X\setminus A_\gamma) = 0$. Since ρ_i is a probability measure, we have that $\rho_i(A_\gamma) = 1$ for every $i \in \mathbb{N}_k$. Therefore, $\mu\big(F_i^{-1}(A_\gamma)\big) = \rho_i(A_\gamma) = 1$. This implies that $\mu(B_\gamma) = \mu\big(\cap_{i=1}^{k} F_i^{-1}(A_\gamma)\big) = 1$. We shall now prove that

$$\big(x, T_2\circ F_i(x)\big)\in\Gamma, \qquad \forall x\in B_\gamma,\ \ \forall i\in\mathbb{N}_k.$$

Since for all $x \in A_\gamma$, we have $T(x) = (T_1x, T_2x) \in \Gamma$, it follows that for each $i \in \mathbb{N}_k$,

$$\big(T_1\circ F_i(x), T_2\circ F_i(x)\big)\in\Gamma, \qquad \forall x\in F_i^{-1}(A_\gamma),$$

from which together with $T_1 \circ F_i = Id_X$ one obtains

$$\big(x, T_2\circ F_i(x)\big)\in\Gamma, \qquad \forall x\in F_i^{-1}(A_\gamma). \tag{3}$$

Thus,

$$\big(x, T_2\circ F_i(x)\big)\in\Gamma, \qquad \forall x\in B_\gamma,\ \ \forall i\in\mathbb{N}_k.$$

This completes the proof of Step I.

Step II: In this step, we assume that assumption (1) of the theorem holds. In this case, we show that $k \le m$.

To do this, let us assume that $k > m$. It follows from Step I that

$$\big(x, T_2\circ F_i(x)\big)\in\Gamma, \qquad \forall x\in B_\gamma,\ \ \forall i\in\{1,\dots,m+1\}. \tag{4}$$

Note that by assertion (1) in Proposition 2.1, we have $\mu(B_{m+1}) > 0$. Since $\mu(B_\gamma) = 1$ and $\mu(B_{m+1}) > 0$, it follows that $B_\gamma \cap B_{m+1} \neq \varnothing$. Take $x \in B_\gamma \cap B_{m+1}$. We have that the cardinality of the set Γ_x is at most m. On the other hand, it follows from (4) that $T_2 \circ F_i(x) \in \Gamma_x$ for all $i \in \{1, 2, \dots, m+1\}$. Thus, there exist $i, j \in \{1, 2, \dots, m+1\}$ with $i < j$ such that $T_2 \circ F_i(x) = T_2 \circ F_j(x)$. Since $T_1 \circ F_i = T_1 \circ F_j = Id_X$ and the map $T = (T_1, T_2)$ is injective, it follows that $F_i(x) = F_j(x)$. On the other hand, $x \in B_{m+1} \subseteq B_j$ from which we have $F_j(x) \notin \{F_1(x), \dots, F_{j-1}(x)\}$. This leads to a contradiction and therefore $k \le m$ in this case.

https://doi.org/10.4153/S0008439524000377 Published online by Cambridge University Press

Step III: In this step, we assume that assumption (2) of the theorem holds. In this case, we prove that if $k = \infty$, then $\mu(\cap_{i=1}^{\infty} B_i) = 0$.

To prove this, let us assume that $k = \infty$ and $\mu(\cap_{i=1}^{\infty} B_i) > 0$. By Step I, we have that $\mu(B_y) = 1$ and

$$(5) \qquad \big(x, T_2 \circ F_i(x)\big) \in \Gamma, \qquad \forall x \in B_y, \ \ \forall i \in \mathbb{N}.$$

Take $x \in \big(\cap_{i=1}^{\infty} B_i\big) \cap B_y$. It follows from (5) that $T_2 \circ F_i x \in \Gamma_x$ for each $i \in \mathbb{N}$. On the other hand, by assumption, we have that $card(\Gamma_x) < \infty$. Thus, there exist i, j with $i < j$ such that $T_2 \circ F_i(x) = T_2 \circ F_j(x)$. As in Step II, since $T_1 \circ F_i = T_1 \circ F_j = Id_X$ and the map $T = (T_1, T_2)$ is injective, it follows that $F_i(x) = F_j(x)$. On the other hand, $x \in \cap_{i=1}^{\infty} B_i \subseteq B_j$ from which we have $F_j(x) \notin \{F_1(x), \ldots, F_{j-1}(x)\}$. This leads to a contradiction and Step III follows.

It now follows from Steps II and III that either $k \neq \infty$ or, if $k = \infty$, then $\mu(\cap_{i=1}^{\infty} B_i) = 0$. On the other hand, Proposition 2.1 yields that if either $k \neq \infty$ or, $k = \infty$ and $\mu(\cap_{i=1}^{\infty} B_i) = 0$, then for every $F \in \mathcal{G}(\mathcal{K})$, the measure $\rho_F = F_{\#}\mu$ is absolutely continuous with respect to the measure $\sum_{i=1}^{k} \rho_i$, where $\rho_i = F_{i\#}\mu$ for $i \in \mathbb{N}_k$. This together with the representation

$$\mu(B) = \int_{\text{ext}\, M(T_1,\mu)} \rho(B)\, d\xi(\rho) = \int_{K_y} \rho(B)\, d\xi(\rho), \qquad \big(\forall B \in \mathcal{B}(X)\big),$$

imply that μ is absolutely continuous with respect to $\sum_{i=1}^{k} \rho_i$. It then follows that there exists a nonnegative measurable function $\alpha : X \to \mathbb{R} \cup \{+\infty\}$ such that

$$\frac{d\mu}{d\big(\sum_{i=1}^{k} \rho_i\big)} = \alpha.$$

Define $\alpha_i = \alpha \circ F_i$ for $i \in \mathbb{N}_k$. We show that $\sum_{i=1}^{k} \alpha_i(x) = 1$ for μ-almost every $x \in X$. In fact, for each $B \in \mathcal{B}(X)$, we have

$$\begin{aligned} \mu(B) &= \mu(T_1^{-1}(B)) \\ &= \sum_{i=1}^{k} \int_{T_1^{-1}(B)} \alpha(x)\, d\rho_i = \sum_{i=1}^{k} \int_{F_i^{-1} \circ T_1^{-1}(B)} \alpha(F_i x)\, d\mu = \sum_{i=1}^{k} \int_B \alpha_i(x)\, d\mu, \end{aligned}$$

from which we obtain $\mu(B) = \sum_{i=1}^{k} \int_B \alpha_i(x)\, d\mu$. Since this holds for all $B \in \mathcal{B}(X)$, we have

$$\sum_{i=1}^{k} \alpha_i(x) = 1, \qquad \mu - a.e.$$

It now follows from Proposition 2.2 that each F_i is μ-a.e. equal to a $(\mathcal{B}(X), \mathcal{B}(X))$-measurable function for which we still denote it by F_i. For each $i \in \mathbb{N}_k$, let

$G_i = T_2 \circ F_i$. We now show that $\gamma = \sum_{i=1}^k (\mathrm{Id} \times G_i)_\#(\alpha_i \mu)$. For each bounded continuous function $f : X \times Y \to \mathbb{R}$, it follows that

$$\int_{X\times Y} f(x,y)\,d\gamma = \int_X f(T_1x, T_2x)\,d\mu = \sum_{i=1}^k \int_X \alpha(x) f(T_1x, T_2x)\,d\rho_i$$
$$= \sum_{i=1}^k \int_X \alpha\big(F_i(x)\big) f\big(T_1 \circ F_i(x), T_2 \circ F_i(x)\big)\,d\mu$$
$$= \sum_{i=1}^k \int_X \alpha_i(x) f\big(x, G_i(x)\big)\,d\mu.$$

Therefore,

$$\gamma = \sum_{i=1}^k (\mathrm{Id} \times G_i)_\#(\alpha_i \mu). \qquad \blacksquare$$

Remark 2.8 It follows from the last part of the proof of Theorem 1.2 that if $G_i(x) = G_j(x)$ for some $x \in X$, then $\alpha_i(x) = \alpha_j(x)$. In fact, let us assume that $G_i(x) = G_j(x)$ for some $x \in X$. It implies that $T_2 \circ F_i(x) = T_2 \circ F_j(x)$. Since $T_1 \circ F_i(x) = T_1 \circ F_j(x) = x$ and $T = (T_1, T_2)$ is injective, we obtain that $F_i(x) = F_j(x)$. This yields that

$$\alpha_i(x) = \alpha \circ F_i(x) = \alpha \circ F_j(x) = \alpha_j(x),$$

as claimed.

It is worth noting that Theorem 1.2 has a straight forward generalization to the multi-marginal case.

Corollary 2.9 *Let $\mu_1, \dots, \mu_n$ be Borel probability measures on complete separable metric spaces $X_1, \dots, X_n$ respectively. Assume that μ_1 is continuous. Let γ be a probability measure on $X_1 \times \cdots \times X_n$ with fixed marginal μ_i on X_i, and let $\Gamma = \mathrm{spt}(\gamma)$. The following assertions hold:*

1. *If there exists $m \in \mathbb{N}$ such that the cardinality of the set*

$$\Gamma_{x_1} := \big\{(x_2, \dots, x_n) \in X_2 \times \cdots \times X_n;\ (x_1, \dots, x_n) \in \Gamma\big\}$$

does not exceed m for μ_1-a.e. $x_1 \in X_1$, then there exists $k \le m$ and a sequence of Borel measurable maps $\{G_i\}_{i=1}^k$ from X_1 to $X_2 \times \cdots \times X_n$ such that the measure γ is concentrated on their graphs.

2. *If $\mathrm{card}\big(\Gamma_{x_1}\big) < \infty$ for μ_1-a.e. $x_1 \in X_1$, then there exist $k \in \mathbb{N} \cup \{\infty\}$ and a sequence of Borel measurable maps $\{G_i\}_{i=1}^k$ from X_1 to $X_2 \times \cdots \times X_n$ such that the measure γ is concentrated on their graphs.*

Proof Let $Y = X_2 \times \cdots \times X_n$ and ν be the projection of γ on Y. It follows that $\gamma \in \Pi(\mu_1, \nu)$. Since μ_1 is continuous the desired result follows from Theorem 1.2. $\blacksquare$

3 Applications in optimal transportation

Here, we shall provide an application of Theorem 1.2. Let $\mathcal{T}$ be a $(2,3)$-torus knot in $\mathbb{R}^3$ (see Figure 1). Our goal is to describe the structure of optimal plans for the cost $c : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ given by

$$c(x,y) = \frac{1}{2}|x-y|^2.$$

Let μ and ν be two probability measures on $\mathcal{T}$. Since the function c is bounded and continuous on $\mathcal{T} \times \mathcal{T}$ it follows that the problem

$$\inf \Big\{ \int_{\mathcal{T}\times\mathcal{T}} c(x,y)\, d\gamma;\ \gamma \in \Pi(\mu,\nu) \Big\}, \tag{6}$$

admits a solution. We have the following result.

Theorem 3.1 *Assume that the nonatomic measure μ is absolutely continuous in each coordinate chart on $\mathcal{T}$. Then any optimal plan of (6) is concentrated on the graphs of at most eight measurable maps.*

Proof By standard results in the theory of optimal transportation, there exist measurable functions $\varphi : \mathcal{T} \to \mathbb{R}$ and $\psi : \mathcal{T} \to \mathbb{R}$ with

$$\psi(y) = \inf_{x\in\mathcal{T}} \{c(x,y) - \varphi(x)\} \qquad \textit{and} \qquad \varphi(x) = \inf_{y\in\mathcal{T}} \{c(x,y) - \psi(y)\}, \tag{7}$$

such that for any optimal plan γ of (6),

$$\text{spt}(\gamma) \subseteq \big\{(x,y) \in \mathcal{T}\times\mathcal{T} \,:\, \varphi(x) + \psi(y) = c(x,y)\big\}.$$

Since $\mathcal{T}$ is bounded, it follows from Lemma C.1 in [3] that φ is locally Lipschitz on $\mathcal{T}$. Let $M = \text{Dom}(D\varphi)$. It follows from Rademacher's theorem together with the absolute continuity of μ that $\mu(M) = 1$. For $x_0 \in M$, if there exist $y_0, y \in \mathcal{T}$ with (x_0, y_0) and $(x_0, y) \in \text{spt}(\gamma)$, then we must have $D_1 c(x_0, y_0) = D_1 c(x_0, y)$. Let $\vec{N}(x_0)$ be the outward normal vector at x_o. If

$$D_1 c(x_0, y_0) = D_1 c(x_0, y),$$

Figure 1: $(2,3)$-*torus knot* $\mathcal{T}$.

https://doi.org/10.4153/S0008439524000377 Published online by Cambridge University Press

then $y - y_0 = \alpha \vec{N}(x_0)$ for some $\alpha \in \mathbb{R}$. This implies that $y = y_0 + \alpha \vec{N}(x_0)$. The latter argument shows that all the points in the set

$$\Big\{ y \in \mathcal{T};\, D_1 c(x_0, y_0) = D_1 c(x_0, y) \Big\},$$

live on a straight line through y_0 and parallel to the normal vector $\vec{N}(x_0)$. On the other hand, one can easily observe that any straight line can intersect the manifold $\mathcal{T}$ in at most eight points. This proves that $\mathrm{card}\big(\Gamma_x\big) \le 8$ is for μ-a.e. $x \in \mathcal{T}$ where $\Gamma_x = \big\{ y \in \mathcal{T};\, (x, y) \in \mathrm{spt}(\gamma) \big\}$. Therefore, by Theorem 1.2, there exist $k \in \{1, 2, \dots, 8\}$ and a sequence of Borel measurable maps $\{G_i\}_{i=1}^k$ from $\mathcal{T}$ to $\mathcal{T}$ such that the measure γ is concentrated on their graphs. ∎

Data availability statement Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Competing interests The author declares no competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

[1] N. Ahmad, H. K. Kim, and R. J. McCann, *Optimal transportation, topology and uniqueness*. Bull. Math. Sci. **1**(2011), 13–32.

[2] V. I. Bogachev, *Measure theory. Vols. I, II*, Springer-Verlag, Berlin, 2007.

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

[4] S. Graf, *Induced σ-homomorphisms and a parametrization of measurable sections via extremal preimage measures*. Math. Ann. **247**(1980), no. 1, 67–80.

[5] A. Kechris, *Classical descriptive set theory*, Graduate Texts in Mathematics, 156, Springer-Verlag, New York, 1995.

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

[7] R. McCann and L. Rifford, *The intrinsic dynamics of optimal transport*. J. Ec. polytech. Math. 3(2016), 67–98.

[8] A. Moameni, *A characterization for solutions of the Monge–Kantorovich mass transport problem*. Math. Ann. **365**(2016), nos. 3–4, 1279–1304.

[9] A. Moameni and B. Pass, *Solutions to multi-marginal optimal transport problems supported on several graphs*. ESAIM Control Optim. Calc. Var. **23**(2017), no. 2, 551–567.

[10] A. Moameni and L. Rifford, *Uniquely minimizing costs for the Kantorovitch problem*. Ann. Fac. Sci. Toulouse Math. (6) **29**(2020), no. 3, 507–563.

[11] S. T. Rachev and L. Rüschendorf, *Mass transportation problems. Vol. I*, Theory. Probability and its Applications (New York), Springer-Verlag, New York, 1998.

[12] C. Villani, *Optimal transport, old and new*, Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 2009.

[13] H. von Weizsäcker and G. Winkler, *Integral representation in the set of solutions of a generalized moment problem*. Math. Ann. **246**(1979/80), no. 1, 23–32.

School of Mathematics and Statistics, Carleton University, Ottawa, ON K1S 5B6, Canada
e-mail: momeni@math.carleton.ca

https://doi.org/10.4153/S0008439524000377 Published online by Cambridge University Press