Supports of extremal doubly stochastic measures

Abbas Moameni *
School of Mathematics and Statistics
Carleton University
Ottawa, ON, Canada K1S 5B6
momeni@math.carleton.ca

September 22, 2014

Abstract

A doubly stochastic measure on the unit square is a Borel probability measure whose horizental and vertical marginals both coincide with the Lebesgue measure. The set of doubly stochastic measures is convex and compact so its extremal points are of particular interest. The problem #111 of Birkhoff (Lattice Theory 1948) is to provide a necessary and sufficient condition on the support of a doubly stochastic measure to guarantee extremality. It was proved by Beneš and Štěpán that an extremal doubly stochastic measure is concentrated on a set which admits an aperiodic decomposition. Hestir and Williams later found a necessary condition which is nearly sufficient by further refining the aperiodic structure of the support of extremal doubly stochastic measures. Our objective in this work is to provide a more practical necessary and nearly sufficient condition for a set to support an extremal doubly stochastic measure.

Keywords: Doubly stochastic measures, Extremality, Uniqueness

1 Introduction

A doubly $n \times n$ stochastic matrix is a real matrix whose entries are non-negative and whose rows and columns individually sum to one. A classical theorem due to Birkhoff [3] and von Neumann [19] states that the set of doubly stochastic matrices is the convex hull of the set of $n \times n$ permutation matrices. Birkhoff proposed the problem of extending this to an infinite dimensional analog known as Birkhoffs Problem #111 (Lattice Theory, Revised Edition [4]). This project has been taken up at various points since its formulation. A doubly stochastic measure on the square refers to a non-negative Borel probability measure on $[0, 1] \times [0, 1]$ whose horizontal and vertical marginals both coincide with Lebesgue measure m on $[0, 1]$. Let us denote this set of doubly stochastic measures by $\Pi(m, m)$ that is indeed a convex and weak$-*$ compact set. A measure γ in $\Pi(m, m)$ is an extremal point if it cannot be written as a convex combination of measures in $\Pi(m, m)$. Doubly stochastic measures and their extremal points are interesting objects to study for several reasons. For instance, all joint probability distributions can be represented using doubly stochastic measures. In particular there has been an extensive study on the class of extremal doubly stochastic measures whose support is contained in a hairpin set(see e.g. [10, 14, 20, 22, 23]). From the applied probability point of view, doubly stochastic measures are a class of probability measures that is in one-to-one correspondence with the class of copulas (see, e.g., [18]). They are also extremely important in the theory of Monge-Kantorovich optimal mass transportation to prove uniqueness of optimal transference plans (see e.g. [1, 7, 17, 21, 24]).

One can formulate the problem in slightly greater generality, by replacing the two copies of $([0, 1], m)$ with probability spaces (X, μ) and (Y, ν), where X and Y are complete separable metric spaces equipped with Borel probabilty measures μ and ν respectively. Denote by $\Pi(\mu, \nu)$ the set of Borel probability measures

*Supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

on $X \times Y$ having μ and ν for marginals.

Characterizations of extremal doubly stochastic measures originally given by Douglas and Lindenstrauss [9, 15] states that a measure $\gamma \in \Pi(\mu,\nu)$ is extremal if and only if $L^1(X;d\mu) \oplus L^1(Y;d\nu)$ is dense in $L^1(X \times Y; d\mu \otimes \nu)$. This characterization is framed in a functional analytic language which doesn't give a simple test for extremality; nor is it obvious how this criterion could be reduced to a condition on the support of γ in $X \times Y$. Significant further progress was made by Beneš & Štěpán [5]. We shall need a few preliminaries before stating their result. For a map f from a set X to a set Y denote by $Dom(f)$ the domain of f, by $Ran(f)$ the range of f and by $Graph(f)$ the graph of f defined by

$$Graph(f) = \big\{(x, g(x));\ x \in Dom(f)\big\}.$$

For a map g from Y to X, the antigraph of g is denoted by $Antigraph(g)$ and defined by

$$Antigraph(g) = \big\{(g(y), y);\ y \in Dom(g)\big\}.$$

Here we recall the definition of aperiodic representations [5, 8].

Definition 1.1 *Let X and Y be two sets and let $f : X \to Y$ and $g : Y \to X$. Define*

$$T(x) = \begin{cases} g \circ f(x), & x \in Dom(f) \cap f^{-1}\big(Dom(g)\big) = D(T), \\ x, & x \notin D(T). \end{cases}$$

The maps f, g are aperiodic if $x \in D(T)$ implies that $T^n(x) \neq x$ for all $n \geq 1$. If f, g are aperiodic and $Graph(f) \cap Antigraph(g) = \emptyset$ then $S = Graph(f) \cup Antigraph(g)$ is called an aperiodic decomposition of S. Moreover, if $(X, \Sigma(X))$ and $(Y, \Sigma(Y))$ are measure spaces and the maps f and g are measurable we call the maps f, g measure-aperiodic if any T-invariant probability measure defined on $\Sigma(X)$ is supported by $X \backslash D(T)$.

It 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. $\gamma^*(S^c) = 0$. Here is the result of Beneš and Štěpán [5] regarding doubly stochastic measures with aperiodic supports.

Theorem 1.2 (Beneš & Štěpán 1987) *Let $(X, \mathcal{B}(X), \mu)$ and $(Y, \mathcal{B}(Y), \nu)$ be complete separable Borel metric spaces. If γ is an extremal point of $\Pi(\mu,\nu)$ then γ is concentrated on a set which admits an aperiodic decomposition.*
Moreover, let $f : X \to Y$ and $g : Y \to X$ be aperiodic measurable maps and $Graph(f) \cap Antigraph(g) = \emptyset$. Then there exists at most one $\gamma \in \Pi(\mu,\nu)$ that is supported on $S = Graph(f) \cup Antigraph(g)$ provided f and g are measure-aperiodic.

Note that the uniqueness result in the ladder theorem implies extremality as an immediate consequence. Hestir and Williams [13] provided an alternate proof of the latter Theorem while further refining the structure these graphs should take, and rewriting them in terms of limb numbering systems. Here we recall the notion of a numbered limb system proposed by Hestir and Williams [13] to the unit square and adapted by Ahmed-Kim-McCann [1] to $X \times Y$.

Definition 1.3 (Numbered limb system) *Let X and Y be Borel subsets of complete separable metric spaces. A relation $S \subset X \times Y$ is a numbered limb system if there is a sequence of maps $f_{2i} : Dom(f_{2i}) \subset Y \to X$ and $f_{2i-1} : Dom(f_{2i-1}) \subset X \to Y$ such that $S = \cup_{i=1}^{\infty} Graph(f_{2i-1}) \cup Antigraph(f_{2i})$, with*

1. *$Ran(f_i) \subset Dom(f_{i-1})$ for each $i > 1$,*
2. *$Dom(f_i) \cap Dom(f_j) = \emptyset$ for $i - j$ even,*
3. *$Ran(f_1) \cap Dom(h_{2i}) = \emptyset$ for all $i \geq 1$.*

By making use of the axiom of choice, Hestir and Williams deduced from the aperiodically condition of Beneš and Štěpán [5] that each extremal doubly stochastic measure vanishes outside some numbered limb system. Conversely, by assuming that the graphs (and antigraphs) comprising the system are Borel subsets of the square, they proved that vanishing outside a number limb system is sufficient to guarantee extremality of a doubly stochastic measure. Their converse result was extended in the more general setting of subsets $X \times Y$ of complete separable metric spaces, and under a weaker measurability hypothesis on the graphs and antigraphs [1, 7]. The difficulty of applying Theorem 1.2 to prove extremality resides partly in the fact that any geometrical characterization of extremality must be invariant under arbitrary measure-preserving transformations applied independently to the horizontal or vertical variables. In this work we replace the aperiodic and measure-aperiodic hypothesis in Theorem 1.2 with a more practical one.

Definition 1.4 *For functions $f : X \to Y$ and $g : Y \to X$ we say that the graph of f is strongly disjoint from the antigraph of g provided*

1. *$Graph(f) \cap Antigraph(g) = \emptyset$;*
2. *There exists a bounded function $\theta : Y \to R$ such that $\theta\big(f \circ g(y)\big) > \theta(y)$ for every $y \in Dom(f \circ g)$.*

If X and Y are Polish spaces and f, g are Borel measurable, say that the graph of f is strongly disjoint from the antigraph of g in a measurable way if conditions (1) and (2) hold with θ being Borel measurable.

Here we state our main theorem in this paper.

Theorem 1.5 *Let $(X, \mathcal{B}(X), \mu)$ and $(Y, \mathcal{B}(Y), \nu)$ be complete separable Borel metric spaces. Let $f : X \to Y$ and $g : Y \to X$ be two measurable functions such that the graph of f is strongly disjoint from the antigraph of g in a measurable way. Then there exists at most one $\gamma \in \Pi(\mu, \nu)$ that is supported on $S = Graph(f) \cup Antigraph(g)$.*
Moreover, if γ is an extremal point of $\Pi(\mu, \nu)$ then there exist functions $f : X \to Y$ and $g : Y \to X$ such that γ is supported on $Graph(f) \cup Antigraph(g)$ and the graph of f is strongly disjoint from the antigraph of g.

We shall now provide some applications of Theorem 1.5. In the first one we establish a criterion for the uniqueness of measures in $\Pi(\mu, \nu)$ that are supported on the graphs of a countable set of measurable maps.

Theorem 1.6 *Let X and Y be Polish spaces equipped with Borel probability measures μ on X and ν on Y, and let $\{T_i\}_{i=1}^k$ be a (possibly infinite) sequence of measurable maps from X to Y. Assume that the following assertions hold:*

1. *For each $i \geq 2$ the map T_i is injective on the set*
$$D_i := \big\{x \in Dom(T_1) \cap Dom(T_i);\ \ T_1 x \neq T_i x\big\},$$
and $Ran(T_i) \cap Ran(T_j) = \emptyset$ for all $i, j \geq 2$ with $i \neq j$.
2. *There exists a bounded measurable function $\theta : Y \to \mathbb{R}$ with the property that $\theta(T_1 x) > \theta(T_i x)$ on D_i.*

Then there exists at most one $\gamma \in \Pi(\mu, \nu)$ that is supported on $\cup_{i=1}^k Graph(T_i)$.

As an immediate consequence of the latter theorem we recover the following uniqueness result due to Seethoff and Shiflett [22].

Corollary 1.7 *Let $X = Y = [0, 1]$ and $\mu = \nu$ be the Lebesgue measure. If $T_1 \leq T_2$ and one of T_1 or T_2 is injective on $D = \{x; T_1(x) \neq T_2(x)\}$ then there exists at most one $\gamma \in \Pi(\mu, \nu)$ that is supported on the graphs of T_1 and T_2.*

Proof. Suppose T_2 is injective on D. One can define $\theta : D \to \mathbb{R}$ by $\theta(y) = -y$. Since $T_1 \leq T_2$ then $\theta\big(T_1(y)\big) > \theta\big(T_2(y)\big)$ on D. The result then follows from Theorem 1.6. □

As another application of Theorem 1.5, by relaxing the measurability hypotheses required by Hestir and Williams [13], we show that there exists at most one doubly stochastic measure vanishing outside a limb numbering system by imposing some mild measurability assumptions (see Theorem 3.4). Our measurability hypotheses is different from the one established in [1]. We remark that an example in [5] shows that some measurability hypothesis is nevertheless required (see also [16]).

In the next section we shall provide some preliminary and independent results required for the proofs. Section 3 is devoted to the proofs and more applications of Theorem 1.5.

2 Measurable weak sections and extremality

In this section we gather some results from measure theory including Choquet type integral representations and transformations of measures by images and preimages of measurable maps. They are essential for the proof of the main results in the next section.

Let $(X, \mathcal{B}, \mu)$ be a finite, not necessarily complete measure space, and (Y, Σ) 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 $\mathcal{B}_\mu$. A function $T : X \to Y$ is said to be $(\mathcal{B}, \Sigma)$-measurable if and only if $T^{-1}(A) \in \mathcal{B}$ for all $A \in \Sigma$. The push forward of the measure μ by the map T is denoted by $T_\# \mu$, i.e.

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

By the change of variable formula it amounts to saying that $\int_Y f(y)\, d(T_\# \mu) = \int_X f \circ T(x)\, d\mu$, for all bounded measurable functions $f : Y \to \mathbb{R}$. We also have the following definition.

Definition 2.1 *Let $T : X \to Y$ be $(\mathcal{B}, \Sigma)$-measurable and ν 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 $T \circ F = Id_Y$.*

Recall that a Polish space is a separable completely metrizable topological space. A Souslin space (resp. set) is the image of a Polish space (resp. set) under a continuous mapping. Obviously every Polish space is a Souslin space. The following theorem ensures the existence of measurable sections ([6], Theorem 9.1.3). This is indeed a consequence of von Neumann's selection theorem.

Theorem 2.2 *Let X and Y be Souslin spaces and let $T : X \to Y$ be a Borel mapping such that $T(X) = Y$. Then, there exists a mapping $F : Y \to X$ such that $T \circ F(y) = y$ for all $y \in Y$ and F is measurable with respect to every Borel measure on Y.*

If X is a topological space we denote by $\mathcal{B}(X)$ the set of Borel subsets in X. The space of Borel probability measures on a topological space X is denoted by $\mathcal{P}(X)$. 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 $\mathcal{B}(X)$ so that T pushes λ forward to ν, i.e.

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

Evidently $\mathcal{M}(T, \nu)$ is a convex set. A measure λ is an extreme point of $\mathcal{M}(T, \nu)$ if the identity $\lambda = t\lambda_1 + (1 - t)\lambda_2$ with $t \in (0, 1)$ 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 $ext\, \mathcal{M}(T, \nu)$.

We recall the following result from [12] in which a characterization of the set $ext\, \mathcal{M}(T, \nu)$ is given (see also [11] for the case where T is continuous).

Theorem 2.3 *Let (Y, Σ, ν) be a probability space, $(X, \mathcal{B}(X))$ be a Hausdorff space with a Radon probability measure λ, and let $T : X \to Y$ be an $(\mathcal{B}(X), \Sigma)$-measurable mapping. If T is surjective and Σ is countably separated then 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 $\lambda = F_\# \nu$.

Finally, if in addition, Σ is countably generated and for some σ-algebra $\mathcal{S}$ with $\Sigma \subset \mathcal{S} \subset \Sigma_\nu$ and there exists a $(\mathcal{S}, \mathcal{B}(X))$-measurable section of the mapping T, then the indicated conditions are equivalent to the following condition:
(iii) there exists an $(\mathcal{S}, \mathcal{B}(X))$-measurable section F of the mapping T such that $\lambda = F_\# \nu$.

The most interesting for applications is the case where X and Y are Souslin spaces with their Borel σ-algebras and $T : X \to Y$ is a surjective Borel mapping. Then the conditions formulated before assertion (iii) are fulfilled if we take for $\mathcal{S}$ the σ-algebra generated by all Souslin sets. Thus, in this situation, the extreme points of the set $M(T, \nu)$ are exactly the measures of the form $F_\# \nu$, where $F : Y \to X$ is measurable with respect to $(\mathcal{S}, \mathcal{B}(X))$ and $T \circ F(y) = y$ for all $y \in Y$.

We shall now make use of the Choquet theory in the setting of noncompact sets of measures to represent each $\lambda \in M(T,\nu)$ as a Choquet type integral over $ext\, M(T,\nu)$. Let us first recall some notations from von Weizsäcker-Winkler [25]. In the measurable space $(X, \mathcal{B}(X))$, let H be a set of non-negative measures on $\mathcal{B}(X)$. By $\sum_H$ we denote the σ-algebra over H generated by the functions $\varrho \to \varrho(B)$, $B \in \mathcal{B}(X)$. If H is a convex set of measures we denote by $ext\, H$ the set of extreme points of H. The set of tight positive measures on $\mathcal{B}(X)$ is denoted by $\mathcal{M}^+(X)$. For a family $\mathcal{F}$ of real valued Borel measurable functions on X we define

$$\mathcal{M}^+_{\mathcal{F}}(X) = \{\varrho \in \mathcal{M}^+(X);\, \mathcal{F} \subset \mathcal{L}^1(\varrho)\},$$

where $\mathcal{L}^1(\varrho)$ is the set of ϱ-integrable real functions on X. Denote by $\sigma\mathcal{M}^+_{\mathcal{F}}(X)$ the topology on $\mathcal{M}^+_{\mathcal{F}}(X)$ of the functions $\varrho \mapsto \int f\, d\varrho$, $f \in \mathcal{F}$. The weakest topology on $\mathcal{M}^+_{\mathcal{F}}(X)$ that makes the functions $\varrho \mapsto \int f\, d\varrho$ lower semi-continuous for all lower semi-continuous bounded functions f on X is denoted by $v\mathcal{M}^+_{\mathcal{F}}(X)$. Denote by $v\sigma\mathcal{M}^+_{\mathcal{F}}(X)$ the topology generated by $\sigma\mathcal{M}^+_{\mathcal{F}}(X)$ and $v\mathcal{M}^+_{\mathcal{F}}(X)$. Here is the main result of von Weizsäcker-Winkler [25] regarding the Choquet theory in the setting of noncompact sets of measures.

Theorem 2.4 *Let $\mathcal{F}$ be a countable family of real Borel functions on a topological space X. Let H be a convex subset of $\mathcal{M}^+_{\mathcal{F}}(X)$ such that $\sup_{\varrho \in H} \varrho(X) < \infty$. If H is closed with respect to $v\sigma\mathcal{M}^+_{\mathcal{F}}(X)$ then for every $\lambda \in H$ there is a probability measure ξ on $\sum_{ext\, H}$ which represents λ in the following sense*

$$\lambda(B) = \int_{ext\, H} \varrho(B)\, d\xi(\varrho),$$

for every $B \in \mathcal{B}(X)$.

We now use the above theorem to represent each $\lambda \in M(T,\nu)$ as a Choquet type integral over $ext\, M(T,\nu)$.

Theorem 2.5 *Let X and Y be complete separable metric spaces and ν a finite 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 $\sum_{ext\, M(T,\nu)}$ such that for each $B \in \mathcal{B}(X)$,*

$$\lambda(B) = \int_{ext\, M(T,\nu)} \varrho(B)\, d\xi(\varrho).$$

Proof. Note first that any finite Borel measure on a Polish space is tight ([2], Theorem 12.7). It is also known that the Borel σ-algebra of a Souslin space is countably generated and countably separated ([6], Corollary 6.7.5). Let $\mathcal{A}$ be a countable family in $\mathcal{B}(Y)$ which generates $\mathcal{B}(Y)$ as a σ-field. Let

$$\mathcal{F} = \big\{\chi_A \circ T;\, A \in \mathcal{A}\big\},$$

where χ_A is the indicator function of A and $\chi_A \circ T$ is the composition of the map T with the function χ_A. Note that $\mathcal{F}$ is a countable family of real Borel functions on X. It is clear that $M(T,\nu)$ is closed with respect to the topology $v\sigma\mathcal{M}^+_{\mathcal{F}}(X)$. Thus, it follows from Theorem 2.4 that there exists a probability measure ξ on $\sum_{ext\, M(T,\nu)}$ such that for each $B \in \mathcal{B}(X)$ the map $\varrho \to \varrho(B)$ from $ext\, M(T,\nu)$ to $\mathbb{R}$ is measurable and

$$\lambda(B) = \int_{ext\, M(T,\nu)} \varrho(B)\, d\xi(\varrho).$$

$\square$

We conclude this section by another interesting result in [12] that shows that for a map $T : (X, \mathcal{B}(X)) \to (Y, \mathcal{B}(Y), \nu)$ all measurable sections of T can, modulo ν, be parameterized by the pre-image measures of ν and this parametrization can be done in a measurable way.

Theorem 2.6 *Let X and Y be complete separable metric spaces and ν a finite measure on $\mathcal{B}(Y)$. Let $\tilde{\Sigma}$ be the σ-field of universally measurable subsets of Y. Assume that $T : (X, \mathcal{B}(X)) \to (Y, \mathcal{B}(Y))$ is a surjective measurable mapping. Then there exists an $\sum_{ext\, M(T,\nu)} \otimes \tilde{\Sigma} - \mathcal{B}(X)$ measurable map $L : ext\, M(T,\nu) \times Y \to X$ with the following properties:*

i. For fixed $\varrho \in ext\, M(T,\nu)$, the function $L(\varrho,.)$ is an $\tilde{\Sigma} - \mathcal{B}(X)$ measurable section for T.

ii. For every measurable section F for T there exists $\varrho \in ext\, M(T,\nu)$ with $L(\varrho,y) = F(y)$ for ν-a.e. $y \in Y$.

Moreover, if $\lambda \in M(T,\nu)$ then there exists a probability measure ξ on $\sum_{ext\, M(T,\nu)}$ such that

$$\int_X g(x)\, d\lambda = \int_{ext\, M(T,\nu)} \int_X g(x)\, d\varrho(x)\, d\xi(\varrho) = \int_{ext\, M(T,\nu)} \int_Y g\big(L(\varrho,y)\big)\, d\nu(y)\, d\xi(\varrho).$$

for every real Borel measurable function $g \in \mathcal{L}^1(\lambda)$.

Proof. The existence of an $\sum_{ext\, M(T,\nu)} \otimes \tilde{\Sigma} - \mathcal{B}(X)$ measurable map $L : ext\, M(T,\nu) \times Y \to X$ satisfying properties (i) and (ii) was established in Corollary 1 in [12].
For $\lambda \in M(T,\nu)$ it follows from Theorem 2.5 that there exists a Borel probability measure ξ on $\sum_{ext\, M(T,\nu)}$ such that for each $B \in \mathcal{B}(X)$,

$$\lambda(B) = \int_{ext\, M(T,\nu)} \varrho(B)\, d\xi(\varrho).$$

Thus, for every Borel measurable function $g : X \to \mathbb{R}$ in $\mathcal{L}^1(\lambda)$ we have,

$$\int_X g(x)\, d\lambda = \int_{ext\, M(T,\nu)} \int_X g(x)\, d\varrho(x)\, d\xi(\varrho).$$

It now follows from property (i) that for each $\varrho \in ext\, M(T,\nu)$, the function $L(\varrho,.)$ is an $\tilde{\Sigma} - \mathcal{B}(X)$ measurable section for T. This implies that $\sigma = L(\varrho,.)_{\#}\nu$. Therefore,

$$\int_{ext\, M(T,\nu)} \int_X g(x)\, d\varrho(x)\, d\xi(\varrho) = \int_{ext\, M(T,\nu)} \int_Y g\big(L(\varrho,y)\big)\, d\nu(y)\, d\xi(\varrho),$$

as desired. □

3 Proofs and more applications

In this section we shall first prove Theorems 1.5 and 1.6 by using the tools introduced in the previous section. We shall then proceed with more applications of these theorems.

Proof of Theorem 1.5 (The nearly sufficient condition). We will use Theorem 1.2 to prove this part. By assumptions $f : X \to Y$ and $g : Y \to X$ are Borel measurable and the graph of f is strongly disjoint from the antigraph of g in a measurable way. This implies that $Graph(f) \cap Antigraph(g) = \emptyset$ and there exists a Borel measurable bounded function $\theta : Y \to \mathbb{R}$ such that $\theta\big(f \circ g(y)\big) > \theta(y)$ for all $y \in Dom\big(f \circ g\big)$. Define $T : X \to X$ as in Definition 1.1, i.e.,

$$T(x) = \begin{cases} g \circ f(x), & x \in Dom(f) \cap f^{-1}\big(Dom(g)\big) = D(T), \\ x, & x \notin D(T). \end{cases}$$

We shall now proceed with the rest of the proof in two steps. In the first step we show that f and g are aperiodic and in the second step we show that f and g are measure-aperiodic. Then the result follows from Theorem 1.2.

Step 1: Assume that there exist $x \in D(T) = Dom(f) \cap f^{-1}\big(Dom(g)\big)$ and $n \in \mathbb{N}$ such that $(g \circ f)^n(x) = x$. It follows that there exists $y \in g^{-1}(x)$ that

$$f \circ (g \circ f)^{n-1}(x) = y \tag{1}$$

By induction we shall verify that the following inequality holds for every $k \leq n$,

$$\theta\big(f \circ (g \circ f)^{k-1}(x)\big) \geq \theta\big(f(x)\big). \tag{2}$$

It obviously holds for $k=1$. Assuming it holds for $k<n$ we prove that it holds for $k+1$. We have

$$\begin{aligned}
\theta\big(f\circ(g\circ f)^k(x)\big) &= \theta\big(f\circ g\circ(f\circ g)^{k-1}\circ f(x)\big) \\
&\geq \theta\big((f\circ g)^{k-1}\circ f(x)\big) && \big(\text{ since by assumption } \theta\circ f\circ g\geq\theta\big) \\
&= \theta\big(f\circ(g\circ f)^{k-1}(x)\big) \\
&\geq \theta\big(f(x)\big) && \big(\text{ since the inequality holds for } k\big)
\end{aligned}$$

This completes the induction. It now follows from (1) and (2) that

$$\theta(y)=\theta\big(f\circ(g\circ f)^{n-1}(x)\big)\geq\theta\big(f(x)\big).$$

Thus $\theta(y)\geq\theta\big(f(x)\big)$. Taking into account that $y\in g^{-1}(x)$ we have $\theta(y)\geq\theta\big(f\circ g(y)\big)$. This leads to a contradiction as $\theta(y)<\theta\big(f\circ g(y)\big)$.

Step 2: To prove that f and g are measure-aperiodic we need to show that any T-invariant probability measure on $\mathcal{B}(X)$ is supported in $X\setminus D(T)$ where $D(T)=Dom(f)\cap f^{-1}\big(Dom(g)\big)$. Suppose that λ is a probability measure on $\mathcal{B}(X)$ with $T_\#\lambda=\lambda$. Note first that since $T(x)=x$ for each $x\in X\setminus D(T)$ we have that $\lambda\big(T^{-1}(A)\big)=\lambda(A)$ for every measurable subset of $D(T)$. It then implies that $(g\circ f)_\#\lambda=\lambda$ on $D(T)$. Let $f_{|D(T)}$ be the restriction of f on $D(T)$ and let η be the push forward of λ by $f_{|D(T)}$. Since $(g\circ f)_\#\lambda=\lambda$ on $D(T)$, it follows that $g_\#\eta=\lambda$. Let $\mathcal{M}(g,\lambda)$ be the set of positive measures on $\mathcal{B}\big(Dom(g)\big)$ defined by

$$\mathcal{M}(g,\lambda)=\big\{\zeta;\, g_\#\zeta=\lambda\big\}.$$

Note that $\mathcal{M}(g,\lambda)$ is convex and $\eta\in\mathcal{M}(g,\lambda)$. By Theorem 2.3, extreme points of the set $\mathcal{M}(g,\lambda)$ are determined by the preimages of g. Let $\tilde\Sigma$ be the σ-field of universally measurable subsets of $D(T)$. It follows from Theorem 2.6 that there exists an $\sum_{ext\,M(g,\lambda)}\otimes\tilde\Sigma-\mathcal{B}\big(Dom(g)\big)$ measurable map

$$L: ext\,M(g,\lambda)\times D(T)\to Dom(g),$$

such that for fixed $\varrho\in ext\,M(g,\lambda)$, the function $L(\varrho,.)$ is an $\tilde\Sigma-\mathcal{B}\big(Dom(g)\big)$ measurable section for g. Moreover, since $\eta\in M(g,\lambda)$, it follows from Theorem 2.5 that there exists a probability measure ξ on $\sum_{ext\,M(g,\lambda)}$ such that

$$\int_{Dom(g)} g(y)\,d\eta=\int_{ext\,M(g,\lambda)}\int_{Dom(g)} g(y)\,d\varrho(y)\,d\xi(\varrho),$$

for every bounded Borel measurable function g on Y. Since θ is bounded and Borel measurable it follows from Theorem 2.6 that

$$\begin{aligned}
\int_{Dom(g)}\theta(y)\,d\eta &= \int_{ext\,M(g,\lambda)}\int_{Dom(g)}\theta(y)\,d\varrho(y)\,d\xi(\varrho) \\
&= \int_{ext\,M(T,\nu)}\int_{D(T)}\theta\big(L(\varrho,x)\big)\,d\lambda(x)\,d\xi(\varrho). \qquad (3)
\end{aligned}$$

For each $\varrho\in ext\,M(g,\lambda)$, by the properties of the function θ, we have $\theta\big(f\circ g\circ L(\varrho,x)\big)>\theta\big(L(\varrho,x)\big)$ on $D(T)$. Since $g\circ L(\varrho,x)=x$ we obtain

$$\theta\big(f(x)\big)>\theta\big(L(\varrho,x)\big)\qquad\forall x\in D(T). \qquad (4)$$

It now follows (3) and (4) that

$$\begin{aligned}
\int_{D(T)}\theta\big(f(x)\big)\,d\lambda &= \int_{Dom(g)}\theta(y)\,d\eta \\
&= \int_{ext\,M(T,\nu)}\int_{D(T)}\theta\big(L(\varrho,x)\big)\,d\lambda(x)\,d\xi(\varrho) \\
&\leq \int_{ext\,M(T,\nu)}\int_{D(T)}\theta\big(f(x)\big)\,d\lambda(x)\,d\xi(\varrho) \\
&= \int_{D(T)}\theta\big(f(x)\big)\,d\lambda.
\end{aligned}$$

This in fact implies that

$$\int_{ext\,M(T,\nu)} \int_{D(T)} \Big[\theta\big(f(x)\big) - \theta\big(L(\varrho,x)\big)\Big]\, d\lambda(x)\, d\xi(\varrho) = 0.$$

Since the integrand is non-negative there exists $\varrho_0 \in ext\,M(T,\nu)$ such that $\theta\big(L(\varrho_0,x)\big) - \theta\big(f(x)\big) = 0$ for λ almost every $x \in D(T)$. On the other hand by (4) we have that $\theta\big(f(x)\big) > \theta\big(L(\varrho,x)\big)$ for all $x \in D(T)$ and therefore λ must be zero on $D(T)$. This indeed proves that λ must be supported in $X \setminus D(T)$ which completes the proof of Step (2). $\square$

Note that the uniqueness result in Theorem 1.5 implies extremality.

Corollary 3.1 *Let $(X, \mathcal{B}(X), \mu)$ and $(Y, \mathcal{B}(Y), \nu)$ be complete separable Borel metric spaces. Let $f : X \to Y$ and $g : Y \to X$ be two measurable functions such that the graph of f is strongly disjoint from the antigraph of g in a measurable way. If $\gamma \in \Pi(\mu,\nu)$ is supported on $S = Graph(f) \cup Antigraph(g)$ then γ is an extremal point of $\Pi(\mu,\nu)$.*

Proof. Suppose that there exist $\gamma_1, \gamma_2 \in \Pi(\mu,\nu)$ and $0 < t < 1$ such that $\gamma = t\gamma_1 + (1-t)\gamma_2$. It implies that $\gamma \geq \gamma_i$ for $i = 1,2$ and therefore both γ_1 and γ_2 vanish outside S. According to Theorem 1.5 there exists at most one doubly stochastic measure in $\Pi(\mu,\nu)$ supported in S. Hence, $\gamma_1 = \gamma_2$ and the measure γ is an extremal point of $\Pi(\mu,\nu)$. $\square$

Proof of Theorem 1.6. For each $i \geq 2$, since T_i is injective on D_i, we have that $T_i(D_i)$ is a measurable subset of Y ([6], Theorem 6.8.6). Define

$$g : Dom(g) = \cup_{i=2}^{k} T_i(D_i) \subset Y \to X,$$

by $g(y) = T^{-1}_{i|D_i}(y)$ for $y \in T_i(D_i)$ and note that g is measurable. Define

$$f : Dom(f) = Dom(T_1) \subset X \to Y,$$

by $f(x) = T_1(x)$. We shall verify the assumptions of Theorem 1.5 for functions f and g. Note that $Graph(f) \cap Antigraph(g) = \emptyset$. In fact, if $Graph(f) \cap Antigraph(g) \neq \emptyset$ then there exists $x \in Dom(f)$ and $y \in Dom(g)$ with $(x, f(x)) = (g(y), y)$. It then follows that $y = f(x) = T_1(x)$ and $x = T^{-1}_{i|D_i}(y)$ for some $2 \leq i \leq k$. This is a contradiction as $T_1(x) \neq T_i(x)$ on D_i. To conclude we need to verify that $\theta\big(f \circ g(y)\big) > \theta(y)$ for every $y \in Dom(f \circ g)$. Take $y \in Dom(g) \cap g^{-1}\big(Dom(f)\big)$. There exists $i \geq 2$ and $x \in D_i$ such that $y = T_i(x)$. Thus,

$$\theta\big(f \circ g(y)\big) = \theta\big(f \circ g \circ T_i(x)\big) = \theta\big(f(x)\big) = \theta\big(T_1(x)\big) > \theta\big(T_i(x)\big) = \theta(y),$$

from which the result follows. $\square$

By making use of Theorem 1.6 one can easily generalize the result of Seethoff and Shiflett (Corollary 1.7) to higher dimensions. For $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ in $\mathbb{R}^n$ we define $x \preceq y$ if and only if $x_i \leq y_i$ for all $1 \leq i \leq n$.

Corollary 3.2 *Let $X = Y = [0,1]^n$ and $\mu = \nu$ be the n-dimensional Lebesgue measure. Assume that $T_1, T_2 : [0,1]^n \to [0,1]^n$ are such that $T_1 \preceq T_2$ and one of T_1 or T_2 is injective on*

$$D = \big\{x; T_1(x) \neq T_2(x)\big\}.$$

Then there exists at most one $\gamma \in \Pi(\mu,\nu)$ that is supported on the graphs of T_1 and T_2.

Proof Suppose that T_2 is injective on D. One can define $\theta : Y \to \mathbb{R}$ by $\theta(y_1, ..., y_n) = -\sum_{i=1}^{n} y_i$. Since $T_1 \preceq T_2$, it is easily seen that $\theta\big(T_1(x)\big) > \theta\big(T_2(x)\big)$ on D. Thus, all the requirements in Theorem 1.6 are met. $\square$

Here is another application of Theorem 1.6 for maps with disjoint ranges.

Corollary 3.3 *Let X and Y be Polish spaces equipped with Borel probability measures μ on X and ν on Y. Let $\{T_i\}_{i=1}^k$ be a sequence of measurable maps from X to Y such that T_i is injective for each $i \in \{2, ..., k\}$ and $Ran(T_i) \cap Ran(T_j) = \emptyset$ for all $1 \leq i, j \leq k$ with $i \neq j$. If $Ran(T_1)$ is measurable then there exists at most one $\gamma \in \Pi(\mu, \nu)$ that is supported on the graphs of $T_1, T_2, ..., T_k$.*

Proof. Define $\theta(y) = \chi_{Ran(T_1)}(y)$, the indicator function of $Ran(T_1)$. Since $Ran(T_1)$ is measurable we have that θ is a bounded measurable function. For each $i \geq 2$ we have $Ran(T_1) \cap Ran(T_i) = \emptyset$ and therefore for all $x \in Dom(T_1) \cap Dom(T_i)$ we have

$$\theta(T_1 x) = 1 > 0 = \theta(T_i x).$$

Thus the result follows from Theorem 1.6. □

In the following we provide an application of Theorem 1.5 to doubly stochastic measures vanishing outside a limb numbering system.

Theorem 3.4 *Let X and Y be complete separable metric spaces, equipped with Borel probability measures μ on X and ν on Y. Suppose there is a numbered limb system $S = \cup_{i=1}^{\infty} Graph(f_{2i-1}) \cup Antigraph(f_{2i})$ with the property that $\cup_{i=1}^{\infty} Graph(f_{2i-1})$ and $\cup_{i=1}^{\infty} Antigraph(f_{2i})$ are Souslin (e.g. Borel measurable) subsets of $X \times Y$. If $Dom(f_{2i})$ is Borel measurable for each $i \geq 1$ then at most one $\gamma \in \Pi(\mu, \nu)$ vanishes outside of S.*

Proof. Define $g : Dom(g) = \cup_{i=1}^{\infty} Dom(f_{2i}) \subset Y \to X$ by $g(y) = f_{2i}(y)$ when $y \in Dom(f_{2i})$. By disjointness of domains g is a single-valued function and $Antigraph(g) = \cup_{i=1}^{\infty} Antigraph(f_{2i})$. Similarly define the function $f : Dom(f) := \cup_{i=1}^{\infty} Dom(f_{2i-1}) \subset X \to Y$ by $f(x) = f_{2i-1}(x)$ when $x \in Dom(f_{2i-1})$. By assumptions $\cup_{i=1}^{\infty} Graph(f_{2i-1})$ and $\cup_{i=0}^{\infty} Antigraph(f_{2i})$ are Souslin subsets of $X \times Y$. Therefore, $Antigraph(g)$ and $Graph(f)$ are Souslin subsets of the product space from which we obtain that both functions $g : Dom(g) \subset Y \to X$ and $f : Dom(f) \subset X \to Y$ are Borel measurable ([6], Lemma 6.7.1).

We now show that $Graph(f) \cap Antigraph(g) = \emptyset$. If $Graph(f) \cap Antigraph(g) \neq \emptyset$ then there exist $i, j \geq 1$ and $x \in Dom(f_{2i-1})$ and $y \in Dom(f_{2j})$ such that $\big(x, f_{2i-1}(x)\big) = \big(f_{2j}(y), y\big)$. Since $x = f_{2j}(y) \in Ran(f_{2j}) \subset Dom(f_{2j-1})$ we must have $i = j$. Similarly, for $i > 1$, $y = f_{2i-1}(x) \in Rang(f_{2i-1}) \subset Dom(f_{2i-2})$ from which we have $i - 1 = j$ which leads to a contradiction. The case $i = 1$ also leads to a contradiction as $Rang(f_1) \cap Dom(f_{2k}) = \emptyset$ for all $k \geq 1$.

Define $\theta : \cup_{i=}^{\infty} Dom(f_{2i}) \to \mathbb{R}$ by $\theta(y) = 2^{-i}$ for $y \in Dom(f_{2i})$. Since for each $i \geq 1$, $Dom(f_{2i})$ is measurable we have that θ is a bounded Borel measurable function. We show that θ satisfies the assumption of Theorem 1.5. Take $y \in Dom(g) \cap g^{-1}\big(Dom(f)\big)$. Thus, $y \in Dom(f_{2k})$ for some $k > 1$. This implies that $g(y) = f_{2k}(y)$ and since $Ran(f_{2k}) \subset Dom(f_{2k-1})$ we have that

$$f \circ g(y) = f \circ f_{2k}(y) = f_{2k-1} \circ f_{2k}(y).$$

Therefore, $f \circ g(y) \in Ran(f_{2k-1})$. Since $Ran(f_{2k-1}) \subset Dom(f_{2k-2})$ we obtain

$$\theta\big(f \circ g(y)\big) = 2^{-(k-1)} > 2^{-k} = \theta(y).$$

Therefore, $Graph(f)$ is strongly disjoint from $Antigraph(g)$ in a measurable way. It now follows from Theorem 1.5 that at most one $\gamma \in \Pi(\mu, \nu)$ can be supported on

$$Graph(f) \cup Antigraph(g) = \cup_{i=1}^{\infty} Graph(f_{2i-1}) \cup Antigraph(f_{2i}).$$

□

Remark 3.5 *Hestir and Williams [13] proved that vanishing outside a number limb system*

$$S = \cup_{i=1}^{\infty} Graph(f_{2i-1}) \cup Antigraph(f_{2i}),$$

is sufficient to guarantee extremality of a doubly stochastic measure, provided that f_i is Borel measurable for every $i \geq 1$. Their result was later improved by Ahmed-Kim-McCann [1] by showing that if $Graph(f_{2i-1})$ and $Antigraph(f_{2i})$ are γ-measurable subsets of $X \times Y$ for each $i \geq 1$ and for every $\gamma \in \Pi(\mu, \nu)$ vanishing outside of S then at most one γ vanishes outside S.

We conclude this section by completing the proof of Theorem 1.5.

Proof of Theorem 1.5 (The necessary condition). If γ is an extremal point of the convex set $\Pi(\mu,\nu)$ then by the main result of Hestir and Williams [13] there exists a numbered limb system $S=\cup_{i=1}^{\infty}Graph(f_{2i-1})\cup Antigraph(f_{2i})$ such that $\gamma^*(S^c)=0$. Define functions f,g and θ as in the proof of Theorem 3.4. Even though these functions may not be measurable but the graph of f is strongly disjoint from the antigraph of g. $\square$

References

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

[2] C. D. Aliprantis, Kim C. Border, *Infinite dimensional analysis.* Third edition. Springer, Berlin, 2006.

[3] G. Birkhoff. *Tres observaciones sobre el algebra lineal.* Univ. Nac. Tucumán. Revista A 5 (1946) 147-150.

[4] G. Birkhoff. *Lattice Theory (revised edition).* American Mathematical Society, New York, 1948.

[5] V. Beneš, J. Štěpán, *The support of extremal measures with given marginals.* Mathematical Statistics and Probability theory, A, 33-41 (1987).

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

[7] 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) 317-354.

[8] J. L. Denny, *The support of discrete extremal measures with given marginals.* Michigan Math. J. 27 (1980) 59-64.

[9] R. D. Douglas, *On extremal measures and subspace density.* Michigan Math. J. 11 (1964) 243-246.

[10] F. Durante, J. Fernández Sánchez, W. Trutschnig, *Multivariate copulas with hairpin support.* J. Multivariate Anal. 130 (2014), 323-334.

[11] G. A. Edgar, *Measurable weak sections.* Illinois J. Math. 20 (1976), no. 4, 630-646.

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

[13] K. Hestir, S. C. Williams, *Supports of doubly stochastic measures.* Bernoulli 1 (1995) 217-243.

[14] A. Kamiński, P. Mikusiński, H. Sherwood, M. D. Taylor, *Doubly stochastic measures, topology, and latticework hairpins.* J. Math. Anal. Appl. 152 (1) (1990) 252-268.

[15] J. Lindenstrauss. *A remark on doubly stochastic measures.* Amer. Math. Monthly, 72 (1965) 379-382.

[16] V. Losert. *Counterexamples to some conjectures about doubly stochastic measures.* Pacific J. Math. 99 (1982) 387-397.

[17] A. Moameni, *A characterization for solutions of the Monge-Kantorovich mass transport problem.* Submitted (2014), arXiv:1403.3103.

[18] R. B. Nelsen, *An introduction to Copulas.* Springer Series in Statistics, second ed., Springer, New York, 2006.

[19] J. von Neumann. *A certain zero-sum two-person game equivalent to the optimal assignment problem.* Contributions to the Theory of Games, 2:5-12, Princeton, 1953.

[20] J. J. Quesada Molina, J. A. Rodríguez Lallena, *Some remarks on the existence of doubly stochastic measures with latticework hairpin support.* Aequationes Math. 47 (2-3) (1994) 164-174.

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

[22] T. L. Seethoff, R. C. Shiflett, *Doubly stochastic measures with prescribed support.* Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 41 (1977-1978), no. 4, 283-288.

[23] H. Sherwood, M. D. Taylor, *Doubly stochastic measures with hairpin support.* Probab. Theory Related Fields 78 (4) (1988) 617-626.

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

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