Canad. Math. Bull. Vol. 59 (2), 2016 pp. 381–391
http://dx.doi.org/10.4153/CMB-2015-075-3

Supports of Extremal Doubly Stochastic Measures

Abbas Moameni

Abstract. A doubly stochastic measure on the unit square is a Borel probability measure whose horizontal 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 number 111 of Birkhoff 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.

1 Introduction

An $n \times n$ doubly 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 [6] and von Neumann [20] 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 analogue known as Birkhoff's Problem #111 [7]. 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 extensive study on the class of extremal doubly stochastic measures whose support is contained in a hairpin set (see [12, 14, 21, 23, 24]). 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 [19]). They are also extremely important in the theory of Monge-Kantorovich optimal mass transportation to prove uniqueness of optimal transference plans (see [1, 9, 17, 22, 25]).

Received by the editors June 9, 2015; revised November 9, 2015.
Published electronically February 3, 2016.
Supported by a grant from the Natural Sciences and Engineering Research Council of Canada.
AMS subject classification: **49Q15**.
Keywords: optimal mass transport, doubly stochastic measures, extremality, uniqueness.

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 probability measures μ and ν, respectively. Denote by $\Pi(\mu, \nu)$ the set of Borel probability measures on $X \times Y$ having μ and ν as marginals. 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.*, $\gamma^*(S^c) = 0$.

Characterizations of extremal doubly stochastic measures originally given by Douglas and Lindenstrauss [11, 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 $L^1(X \times Y; d\mu \otimes d\nu)$. This characterization is framed in a functional analytic language which does not 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 and Štěpán [3]. We shall need a few preliminaries before stating their result. For a map f from a set X to a set Y denote by $\mathrm{Dom}(f)$ the domain of f, by $\mathrm{Ran}(f)$ the range of f and by $\mathrm{Graph}(f)$ the graph of f defined by $\mathrm{Graph}(f) = \{(x, f(x)); x \in \mathrm{Dom}(f)\}$. For a map g from Y to X, the antigraph of g is denoted by $\mathrm{Antigraph}(g)$ and defined by $\mathrm{Antigraph}(g) = \{(g(y), y); y \in \mathrm{Dom}(g)\}$. If X is a topological space we denote by $\mathcal{B}(X)$ its Borel σ-algebra. Here we recall the definition of aperiodic representations [3, 10].

Definition 1.1 Let X and Y be two sets, and let

$$f\colon \mathrm{Dom}(f) \subseteq X \to Y \quad \text{and} \quad g\colon \mathrm{Dom}(g) \subseteq Y \to X$$

be two maps. Define

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

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 $\mathrm{Graph}(f) \cap \mathrm{Antigraph}(g) = \varnothing$, then $S = \mathrm{Graph}(f) \cup \mathrm{Antigraph}(g)$ is called an aperiodic decomposition of S. Moreover, let $\big(X, \mathcal{B}(X)\big)$ and $\big(Y, \mathcal{B}(Y)\big)$ be Borel measure spaces and let the maps f and g be Borel measurable. Say that the maps f and g are measure-aperiodic if any T-invariant probability measure defined on $\mathcal{B}(X)$ is concentrated on $X \setminus D(T)$.

Here is the result of Beneš and Štěpán [3] regarding doubly stochastic measures with aperiodic supports.

Theorem 1.2 ([3]) *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 $f\colon \mathrm{Dom}(f) \subseteq X \to Y$ and $g\colon \mathrm{Dom}(g) \subseteq Y \to X$ be aperiodic measurable maps with* $\mathrm{Graph}(f) \cap \mathrm{Antigraph}(g) = \varnothing$. *Then there exists at most one $\gamma \in \Pi(\mu, \nu)$ that is concentrated on $S = \mathrm{Graph}(f) \cup \mathrm{Antigraph}(g)$ provided f and g are measure-aperiodic.*

Note that the uniqueness result in the latter theorem implies extremality as an immediate consequence. Hestir and Williams [13] provided an alternate proof of the

https://doi.org/10.4153/CMB-2015-075-3 Published online by Cambridge University Press

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, and 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 $f_{2i}\colon \mathrm{Dom}(f_{2i}) \subset Y \to X$ and $f_{2i-1}\colon \mathrm{Dom}(f_{2i-1}) \subset X \to Y$ such that $S = \bigcup_{i=1}^{\infty}\big(\mathrm{Graph}(f_{2i-1}) \cup \mathrm{Antigraph}(f_{2i})\big)$, with

(i) $\mathrm{Ran}(f_i) \subset \mathrm{Dom}(f_{i-1})$ for each $i > 1$,
(ii) $\mathrm{Dom}(f_i) \cap \mathrm{Dom}(f_j) = \varnothing$ for $i - j$ even,
(iii) $\mathrm{Ran}(f_1) \cap \mathrm{Dom}(f_{2i}) = \varnothing$ for all $i \geq 1$.

By making use of the axiom of choice, Hestir and Williams deduced from the aperiodicity condition of Beneš and Štěpán [3] 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, 9]. By revealing a hidden structure of linear preorder with universally measurable graphs, Bianchini and Caravenna [4, 5] established an interesting result about the extremality of doubly stochastic measures (see also Remark 2.2).

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 $f\colon \mathrm{Dom}(f) \subseteq X \to Y$ and $g\colon \mathrm{Dom}(g) \subseteq Y \to X$ we say that the graph of f is strongly disjoint from the antigraph of g provided

(i) $\mathrm{Graph}(f) \cap \mathrm{Antigraph}(g) = \varnothing$;
(ii) there exists a bounded function $\theta\colon Y \to R$ such that $\theta\big(f \circ g(y)\big) > \theta(y)$ for every $y \in \mathrm{Dom}(f \circ g)$.

If X and Y are Borel 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 (i) and (ii) hold with $\theta\colon Y \to \mathbb{R}$ being Borel measurable.

The existence of a numbered limb system is formally, *i.e.*, without looking at measurability-type properties, equivalent to the existence of a pair f, g whose graph and antigraph are strongly disjoint. Indeed, the construction of f, g from a numbered limb system is done in the proof of Theorem 2.6. Moreover, the construction of a numbered limb system from a "strongly disjoint pair" follows from the aperiodicity proven in the proof of Theorem 1.5 and the theory developed by Hestir and Williams.

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 $f\colon \mathrm{Dom}(f) \subseteq X \to Y$ and $g\colon \mathrm{Dom}(g) \subseteq Y \to X$ 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 concentrated on $S = \mathrm{Graph}(f) \cup \mathrm{Antigraph}(g)$. Moreover, if γ is an extremal point of $\Pi(\mu, \nu)$, then there exist functions $f\colon \mathrm{Dom}(f) \subseteq X \to Y$ and $g\colon \mathrm{Dom}(g) \subseteq Y \to X$ such that γ is concentrated on $\mathrm{Graph}(f) \cup \mathrm{Antigraph}(g)$ and the graph of f is strongly disjoint from the antigraph of g.*

The most important application of the latter theorem is to prove uniqueness results for optimal mass transport problems. To be more precise, let $c\colon X \times Y \to \mathbb{R}$ be a bounded continuous function and consider the following problem (known as the Monge–Kantorovich problem),

$$\text{(MK)} \qquad \inf\Big\{ \int_{X\times Y} c(x, y)\, d\pi; \pi \in \Pi(\mu, \nu)\Big\}.$$

When a measure in $\Pi(\mu, \nu)$ minimizes the cost, it will be called an optimal plan. For each $\gamma \in \Pi(\mu, \nu)$ we define the set-valued function $F_\gamma\colon X \to 2^Y$ by

$$F_\gamma(x) = \big\{ y \in Y; (x, y) \in \mathrm{Supp}(\gamma)\big\},$$

where $\mathrm{Supp}(\gamma)$ stands for the support of the measure γ. The domain of F_γ is defined by $\mathrm{Dom}(F_\gamma) = \{x \in X; F_\gamma(x) \neq \varnothing\}$. We have the following definition.

Definition 1.6 For each $\gamma \in \Pi(\mu, \nu)$ and each continuous function $c\colon X \times Y \to \mathbb{R}$, set $D(\gamma, c) := \big\{ x \in X; \{\mathrm{argmax}\, c(x, y); y \in F_\gamma(x)\}\text{is a singleton}\big\}$, and define the possibly set-valued map $f_{\gamma,c}\colon X \to 2^Y$ by $f_{\gamma,c}(x) = \mathrm{argmax}_{y\in F_\gamma(x)}\, c(x, y)$.

We say that γ is c-extreme if there exists a full μ–measure subset X_0 of X such that

(i) for each $x \in \mathrm{Dom}(F_\gamma)$ the set $\{\mathrm{argmax}\, c(x, y); y \in F_\gamma(x)\}$ is non-empty;

(ii) for all distinct points $x_1, x_2 \in X_0$ the following assertion holds.

$$\big\{ F_\gamma(x_1) \smallsetminus \{y_1\}\big\} \cap \big\{ F_\gamma(x_2) \smallsetminus \{y_2\}\big\} = \varnothing,$$

for all $y_1 \in f_{\gamma,c}(x_1)$ and $y_2 \in f_{\gamma,c}(x_2)$.

The following result was established in [18].

Theorem 1.7 *Let c be a continuous and bounded function. If each optimal plan γ of (MK) is c-extreme, then (MK) has a unique solution.*

The proof of Theorem 1.7 is a combination of the results in Theorem 1.5 and the Kantorovich dual of (MK). Indeed, we can make use of the solutions of the dual problem to act as a measurable function θ in Definition 1.4 to obtain extremality of optimal plans. The interested reader is referred to [18] for the proof of Theorem 1.7 and a comprehensive analysis regarding the uniqueness results for Monge-Kantorovich mass transport problems.

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

Theorem 1.8 *Let X and Y be complete separable Borel metric spaces 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.*

(i) *For each $i \geq 2$ the map T_i is injective on the set*

$$D_i := \left\{ x \in \mathrm{Dom}(T_1) \cap \mathrm{Dom}(T_i); T_1 x \neq T_i x \right\},$$

and $\mathrm{Ran}(T_i) \cap \mathrm{Ran}(T_j) = \varnothing$ for all $i, j \geq 2$ with $i \neq j$.

(ii) *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 concentrated 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 [23].

Corollary 1.9 *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 concentrated on the graphs of T_1 and T_2.*

Proof Suppose T_2 is injective on D. One can define $\theta : Y \to \mathbb{R}$ 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 1.8. ∎

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 2.6). Our measurability hypotheses is different from the one established in [1]. We remark that an example in [3] shows that some measurability hypothesis is nevertheless required (see also [16]).

2 Proofs and More Applications

In this section we shall first prove Theorems 1.5 and 1.8, and then proceed with more applications of these theorems. Recall that a Polish space is a separable completely metrizable topological space. The following result is useful in the sequel.

Lemma 2.1 *Let X and Y be Borel Polish spaces. Suppose that $f: \mathrm{Dom}(f) \subseteq X \to Y$ and $g: \mathrm{Dom}(g) \subseteq Y \to X$ are Borel measurable. The following assertions are equivalent.*

(i) *There exists a bounded measurable function $\theta: Y \to \mathbb{R}$ such that $\theta\big(f \circ g(y)\big) > \theta(y)$ for all $y \in \mathrm{Dom}(f \circ g)$.*

(ii) *There exists a bounded measurable function $\tilde{\theta}: X \to \mathbb{R}$ such that $\widetilde{\theta}\big(g \circ f(x)\big) > \widetilde{\theta}(x)$ for all $x \in \mathrm{Dom}(g \circ f)$.*

https://doi.org/10.4153/CMB-2015-075-3 Published online by Cambridge University Press

Proof (i) → (ii). Since θ is bounded, we have that $\|\theta\|_\infty := \sup_{y\in Y}|\theta(y)| < \infty$. Define $\widetilde{\theta}: X \to \mathbb{R}$ by

$$\widetilde{\theta}(x) = \begin{cases} \theta\big(f(x)\big), & x \in \mathrm{Dom}(f), \\ \|\theta\|_\infty + 1, & x \notin \mathrm{Dom}(f). \end{cases}$$

The boundedness of $\widetilde{\theta}$ follows from the boundedness of θ. It is also easily seen that $\widetilde{\theta}$ is measurable as both θ and f are measurable. We now show that $\widetilde{\theta}\big(g\circ f(x)\big) > \widetilde{\theta}(x)$ for all $x \in \mathrm{Dom}(g \circ f)$. Take $x \in \mathrm{Dom}(g \circ f)$. So it follows that $x \in \mathrm{Dom}(f)$ and $\widetilde{\theta}(x) = \theta\big(f(x)\big)$. We have two cases.

Case I: If $g \circ f(x) \in \mathrm{Dom}(f)$, then $f(x) \in \mathrm{Dom}(f \circ g)$ and

$$\widetilde{\theta}\big(g \circ f(x)\big) = \theta\big(f \circ g \circ f(x)\big) > \theta\big(f(x)\big) = \widetilde{\theta}(x).$$

Case II: If $g \circ f(x) \notin \mathrm{Dom}(f)$, then

$$\widetilde{\theta}\big(g \circ f(x)\big) = \|\theta\|_\infty + 1 > \theta\big(f(x)\big) = \tilde{\theta}(x).$$

This completes the proof of (i) → (ii). The other direction follows from the same argument. ∎

Proof of Theorem 1.5: The nearly sufficient condition We will use Theorem 1.2 to prove this part. By assumptions $f: \mathrm{Dom}(f) \subseteq X \to Y$ and $g: \mathrm{Dom}(g) \subseteq Y \to X$ are Borel measurable and the graph of f is strongly disjoint from the antigraph of g in a measurable way. This implies that $\mathrm{Graph}(f) \cap \mathrm{Antigraph}(g) = \varnothing$ 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 $y \in \mathrm{Dom}\big(f \circ g\big)$. Define $T: X \to X$ as in Definition 1.1, *i.e.*,

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

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

$$f \circ (g \circ f)^{n-1}(x) = y. \tag{2.1}$$

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

$$\theta\big(f \circ (g \circ f)^{k-1}(x)\big) \ge \theta\big(f(x)\big). \tag{2.2}$$

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

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

https://doi.org/10.4153/CMB-2015-075-3 Published online by Cambridge University Press

This completes the induction. It now follows from (2.1) and (2.2) 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 $y \in g^{-1}(x)$, 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 concentrated in $X \setminus D(T)$ where $D(T) = \operatorname{Dom}(f) \cap f^{-1}\big(\operatorname{Dom}(g)\big)$. Suppose that λ is a T−invariant probability measure on $\mathcal{B}(X)$.

By assumption there exists a bounded measurable function $\theta: Y \to \mathbb{R}$ such that $\theta\big(f \circ g(y)\big) > \theta(y)$ for all $y \in \operatorname{Dom}(f \circ g)$. It follows from Lemma 2.1 that there exists a bounded measurable function $\widetilde{\theta}: X \to \mathbb{R}$ such that $\theta\big(g \circ f(x)\big) > \theta(x)$ for all $x \in \operatorname{Dom}(g \circ f)$. Since λ is T−invariant, it follows that

$$\int_X \widetilde{\theta}\big(T(x)\big)\, d\lambda = \int_X \tilde{\theta}(x)\, d\lambda.$$

Since $T(x) = x$ on $X \setminus D(T)$, we obtain that $\int_{D(T)} \widetilde{\theta}\big(T(x)\big)\, d\lambda = \int_{D(T)} \widetilde{\theta}(x)\, d\lambda$. Since $T = g \circ f$ on $D(T)$, it follows from the latter identity that

$$\int_{D(T)} \big[\widetilde{\theta}\big(g \circ f(x)\big) - \widetilde{\theta}(x)\big]\, d\lambda = 0.$$

However, the integrand is non-negative and therefore

$$\widetilde{\theta}\big(g \circ f(x)\big) = \widetilde{\theta}(x), \qquad \lambda\text{-a.e. } x \in D(T).$$

On the other hand, $\widetilde{\theta}\big(g \circ f(x)\big) > \widetilde{\theta}(x)$ for all $x \in D(T)$. This indeed proves that λ must be concentrated in $X \setminus D(T)$ which completes the proof of Step (2). ■

Remark 2.2 The condition that the graph of f and the antigraph of g are strongly disjoint in a measurable way also fulfills the required assumption in [5, Theorem 4.1] from which one obtains a different proof for the nearly sufficient condition in Theorem 1.5.

Note also that the uniqueness result in Theorem 1.5 implies extremality.

Corollary 2.3 *Let $(X, \mathcal{B}(X), \mu)$ and $(Y, \mathcal{B}(Y), \nu)$ be complete separable Borel metric spaces. Let $f: \operatorname{Dom}(f) \subseteq X \to Y$ and $g: \operatorname{Dom}(g) \subseteq Y \to X$ 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 concentrated on $S = \operatorname{Graph}(f) \cup \operatorname{Antigraph}(g)$, 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 γ_1 and γ_2 are absolutely continuous with respect to γ 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 $\Pi(\mu, \nu)$. ■

Proof of Theorem 1.8 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 [8, Theorem 6.8.6]. Define

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

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

$$f: \mathrm{Dom}(f) = \mathrm{Dom}(T_1) \subseteq X \to Y,$$

by $f(x) = T_1(x)$. We shall verify the assumptions of Theorem 1.5 for functions f and g. Note that $\mathrm{Graph}(f) \cap \mathrm{Antigraph}(g) = \varnothing$. In fact, if $\mathrm{Graph}(f) \cap \mathrm{Antigraph}(g) \neq \varnothing$, then there exists $x \in \mathrm{Dom}(f)$ and $y \in \mathrm{Dom}(g)$ with $(x, f(x)) = (g(y), y)$. It then follows that $y = f(x) = T_1(x)$ and $x = T_{i|D_i}^{-1}(y)$ 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 $y \in \mathrm{Dom}(f \circ g)$. Take $y \in \mathrm{Dom}(g) \cap g^{-1}\big(\mathrm{Dom}(f)\big)$. 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. ∎

By making use of Theorem 1.8 one can easily generalize the result of Seethoff and Shiflett, *i.e.*, Corollary 1.9, to higher dimensions. For

$$x = (x_1, \ldots, x_n), \quad y = (y_1, \ldots, y_n)$$

in $\mathbb{R}^n$ we define $x \leq y$ if and only if $x_i \leq y_i$ for all $1 \leq i \leq n$.

Corollary 2.4 *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 $T_1 \leq T_2$ 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 concentrated 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 $\theta(y_1 \ldots, y_n) = -\sum_{i=1}^{n} y_i$. Since $T_1 \leq T_2$, 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 1.8 are met and the result follows accordingly. ∎

Here is another application of Theorem 1.8 for maps with disjoint ranges.

Corollary 2.5 *Let X and Y be complete separable Borel metric spaces 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 $i \in \{2, \ldots, k\}$ and $Ran(T_i) \cap \mathrm{Ran}(T_j) = \varnothing$ for all $1 \leq i, j \leq k$ with $i \neq j$. If $\mathrm{Ran}(T_1)$ is measurable then there exists at most one $\gamma \in \Pi(\mu, \nu)$ that is concentrated on the graphs of $T_1, T_2, \ldots, T_k$.*

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

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

Thus the result follows from Theorem 1.8. ∎

In the following we provide an application of Theorem 1.5 to doubly stochastic measures vanishing outside a limb numbering system (Definition 1.3). Recall that a Souslin set is a continuous image of a Borel set in a Polish space. Obviously every Borel set is a Souslin set.

Theorem 2.6 *Let X and Y be complete separable metric spaces, equipped with Borel probability measures μ on X and ν on Y. Suppose that there is a numbered limb system,*

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

*with the property that $\cup_{i=1}^{\infty} \operatorname{Graph}(f_{2i-1})$ and $\cup_{i=1}^{\infty} \operatorname{Antigraph}(f_{2i})$ are Souslin (*e.g., *Borel measurable) subsets of $X \times Y$. If one of the following assertions holds,*

(i) $\operatorname{Dom}(f_{2i})$ *is Borel measurable for every $i \geq 1$,*

(ii) $\operatorname{Dom}(f_{2i-1})$ *is Borel measurable for every $i \geq 1$,*

then at most one $\gamma \in \Pi(\mu, \nu)$ vanishes outside of S.

Proof Define $g\colon \operatorname{Dom}(g) = \bigcup_{i=1}^{\infty} \operatorname{Dom}(f_{2i}) \subseteq Y \to X$, by $g(y) = f_{2i}(y)$ when $y \in \operatorname{Dom}(f_{2i})$. By disjointness of the domains of f_{2i}, the function g is a single-valued function and $\operatorname{Antigraph}(g) = \bigcup_{i=1}^{\infty} \operatorname{Antigraph}(f_{2i})$. Similarly define the function

$$f\colon \operatorname{Dom}(f) = \bigcup_{i=1}^{\infty} \operatorname{Dom}(f_{2i-1}) \subseteq X \to Y,$$

by $f(x) = f_{2i-1}(x)$ when $x \in \operatorname{Dom}(f_{2i-1})$. By assumptions $\cup_{i=1}^{\infty} \operatorname{Graph}(f_{2i-1})$ and $\cup_{i=0}^{\infty} \operatorname{Antigraph}(f_{2i})$ are Souslin subsets of $X \times Y$. Therefore, $\operatorname{Antigraph}(g)$ and $\operatorname{Graph}(f)$ are Souslin subsets of the product space from which we obtain that both functions $g\colon \operatorname{Dom}(g) \subseteq Y \to X$ and $f\colon \operatorname{Dom}(f) \subseteq X \to Y$ are Borel measurable [8, Lemma 6.7.1].

We now show that $\operatorname{Graph}(f) \cap \operatorname{Antigraph}(g) = \varnothing$. If $\operatorname{Graph}(f) \cap \operatorname{Antigraph}(g) \neq \varnothing$, then there exist $i, j \geq 1$ and $x \in \operatorname{Dom}(f_{2i-1})$ and $y \in \operatorname{Dom}(f_{2j})$ such that

$$\big(x, f_{2i-1}(x)\big) = \big(f_{2j}(y), y\big).$$

Since $x = f_{2j}(y) \in \operatorname{Ran}(f_{2j}) \subset \operatorname{Dom}(f_{2j-1})$, we must have $i = j$. Similarly, if $i > 1$,

$$y = f_{2i-1}(x) \in \operatorname{Rang}(f_{2i-1}) \subset \operatorname{Dom}(f_{2i-2}),$$

from which we have $i - 1 = j$ which leads to a contradiction. The case $i = 1$ also leads to a contradiction as $\operatorname{Rang}(f_1) \cap \operatorname{Dom}(f_{2k}) = \varnothing$ for all $k \geq 1$.

Suppose now that the first assertion in the statement holds and $\operatorname{Dom}(f_{2i})$ is measurable for every $i \geq 1$. The proof for the second assertion is similar. Define $\theta\colon Y \to \mathbb{R}$ by $\theta(y) = 2^{-i}$ if $y \in \operatorname{Dom}(f_{2i})$ for some $i \geq 1$, and $\theta(y) = 1$ if $y \notin \bigcup_{i=}^{\infty} \operatorname{Dom}(f_{2i})$. Since for each $i \geq 1$, $\operatorname{Dom}(f_{2i})$ is measurable we have that θ is a bounded Borel measurable function. We show that θ satisfies the assumption of Theorem 1.5. Take $y \in \operatorname{Dom}(g) \cap g^{-1}\big(\operatorname{Dom}(f)\big)$. Thus, $y \in \operatorname{Dom}(f_{2k})$ for some $k > 1$. This implies that $g(y) = f_{2k}(y)$ and since $\operatorname{Ran}(f_{2k}) \subset \operatorname{Dom}(f_{2k-1})$, we have that

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

https://doi.org/10.4153/CMB-2015-075-3 Published online by Cambridge University Press

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

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

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

$$\mathrm{Graph}(f) \cup \mathrm{Antigraph}(g) = \bigcup_{i=1}^{\infty}\Big(\mathrm{Graph}(f_{2i-1}) \cup \mathrm{Antigraph}(f_{2i})\Big). \qquad \blacksquare$$

Remark 2.7 Hestir and Williams [13] proved that vanishing outside a number limb system $S = \bigcup_{i=1}^{\infty}\big(\mathrm{Graph}(f_{2i-1}) \cup \mathrm{Antigraph}(f_{2i})\big)$, 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, and McCann [1] by showing that if $\mathrm{Graph}(f_{2i-1})$ and $\mathrm{Antigraph}(f_{2i})$ 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] γ is concentrated on a numbered limb system $S = \bigcup_{i=1}^{\infty}\big(\mathrm{Graph}(f_{2i-1}) \cup \mathrm{Antigraph}(f_{2i})\big)$. Define functions f, g and θ as in the proof of Theorem 2.6. Even though these functions may not be measurable but the graph of f is strongly disjoint from the antigraph of g. $\blacksquare$

References

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

[2] C. D. Aliprantis and K. C. Border, *Infinite dimensional analysis.* Third edition. Springer, Berlin, 2006. http://dx.doi.org/10.1007/s13373-011-0002-7

[3] V. Beneš and J. Štěpán, *The support of extremal probability measures with given marginals.* In: Mathematical statistics and probability theory, Vol. A, Reidel, Dordrecht, 1987, pp. 33–41.

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

[5] ———, *On optimality of c-cyclically monotone transference plans.* C. R. Math. Acad. Sci. Paris 348 (2010), no. 11-12, 613–618.

[6] G. Birkhoff, *Tres observaciones sobre el algebra lineal.* Univ. Nac. Tucumán. Revista A (1946) 147-151.

[7] ———, *Lattice theory* Revised edition. American Mathematical Society, New York, 1948.

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

[9] P.-A. Chiappori, R. J. McCann, and L. P. Nesheim, *Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness.* Econom. Theory **42**(2010), no. 2, 317–354. http://dx.doi.org/10.1007/s00199-009-0455-z

[10] J. L. Denny, *The support of discrete extremal measures with given marginals.* Michigan Math. J. **27**(1980), no. 1, 59–64.

[11] R. D. Douglas, *On extremal measures and subspace density.* Michigan Math. J. **11**(1964) 243–246. http://dx.doi.org/10.1307/mmj/1029002309

[12] F. Durante, J. Fernández Sánchez, and W. Trutschnig, *Multivariate copulas with hairpin support.* J. Multivariate Anal. **130**(2014), 323–334. http://dx.doi.org/10.1016/j.jmva.2014.06.009

[13] K. Hestir and S. C. Williams, *Supports of doubly stochastic measures.* Bernoulli **1**(1995), no. 3, 217–243. http://dx.doi.org/10.2307/3318478

[14] A. Kamiński, P. Mikusiński, H. Sherwood, and M. D. Taylor, *Doubly stochastic measures, topology, and latticework hairpins.* J. Math. Anal. Appl. **152**(1990), no. 1, 252–268. http://dx.doi.org/10.1016/0022-247X(90)90102-L

[15] J. Lindenstrauss, *A remark on extreme 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), no. 2, 387–397. http://dx.doi.org/10.2307/2313497

[17] A. Moameni, *A characterization for solutions of the Monge-Kantorovich mass transport problem.* Math. Ann., to appear. http://dx.doi.org/10.1007/s00208-015-1312-y

[18] R. McCann, A. Moameni, and L. Rifford, *Uniqueness results for Monge-Kantorovich mass transport problems.* http://people.math.carleton.ca/~momeni/Publications.html

[19] R. B. Nelsen, *An introduction to copulas.* Second edition. Springer Series in Statistics, Springer, New York, 2006.

[20] J. von Neumann, *A certain zero-sum two-person game equivalent to the optimal assignment problem.* In: Contributions to the theory of games, Vol. 2, Princeton, 1953, pp. 5–12.

[21] J. J. Quesada Molina and J.-A. Rodríguez Lallena, *Some remarks on the existence of doubly stochastic measures with latticework hairpin support.* Aequationes Math. **47**(1994), no. 2-3, 164–174. http://dx.doi.org/10.1007/BF01832957

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

[23] T. L. Seethoff and R. C. Shiflett, *Doubly stochastic measures with prescribed support.* Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **41**(1977/78), no. 4, 283–288. http://dx.doi.org/10.1007/BF00533599

[24] H. Sherwood and M. D. Taylor, *Doubly stochastic measures with hairpin support.* Probab. Theory Related Fields **78**(1988), no. 4, 617–626. http://dx.doi.org/10.1007/BF00353879

[25] C. Villani, *Optimal transport, old and new.* Grundlehren der Mathematischen Wissenschaften, 338. Springer-Verlag, Berlin, 2009.

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