Electron. Commun. Probab. **29** (2024), article no. 17, 1–12.
https://doi.org/10.1214/24-ECP586
ISSN: 1083-589X

**ELECTRONIC COMMUNICATIONS in PROBABILITY**

# Some duality results for equivalence couplings and total variation

Luca Pratelli* Pietro Rigo†

**Abstract**

Let $(\Omega, \mathcal{F})$ be a measurable space and $E \subset \Omega \times \Omega$. Suppose that $E \in \mathcal{F} \otimes \mathcal{F}$ and the relation on $\Omega$ defined as $x \sim y \Leftrightarrow (x, y) \in E$ is reflexive, symmetric and transitive. Following [7], say that $E$ is strongly dualizable if there is a sub-$\sigma$-field $\mathcal{G} \subset \mathcal{F}$ such that

$$\min_{P \in \Gamma(\mu,\nu)} (1 - P(E)) = \max_{A \in \mathcal{G}} |\mu(A) - \nu(A)|$$

for all probabilities $\mu$ and $\nu$ on $\mathcal{F}$. This paper investigates strong duality. Essentially, it is shown that $E$ is strongly dualizable provided some mild modifications are admitted. Let $\mathcal{G}_0$ be the $E$-invariant sub-$\sigma$-field of $\mathcal{F}$. One result is that, for all probabilities $\mu$ and $\nu$ on $\mathcal{F}$, there is a probability $\nu_0$ on $\mathcal{F}$ such that

$$\nu_0 = \nu \text{ on } \mathcal{G}_0 \quad \text{and} \quad \min_{P \in \Gamma(\mu,\nu_0)} (1 - P(E)) = \max_{A \in \mathcal{G}_0} |\mu(A) - \nu(A)|.$$

In the other results, $(\Omega, \mathcal{F})$ is a standard Borel space and the min over $\Gamma(\mu, \nu)$ is replaced by the inf over $\Gamma(\mu, \nu)$ in the definition of strong duality. Then, $E$ is strongly dualizable provided $\mathcal{G}$ is allowed to depend on $(\mu, \nu)$ or it is taken to be the universally measurable version of the $E$-invariant $\sigma$-field.

**Keywords:** duality; equivalence relation; finitely additive probability measure; optimal transport; total variation.
**MSC2020 subject classifications:** 60A10; 60E05; 28A35; 49N15; 49Q22.
Submitted to ECP on November 14, 2023, final version accepted on March 10, 2024.

## 1 Introduction

Throughout, $(\Omega, \mathcal{F})$ is a measurable space, $\mathcal{P}(\mathcal{F})$ the collection of all probability measures on $\mathcal{F}$, and $E \subset \Omega \times \Omega$ a measurable equivalence relation. This means that $E \in \mathcal{F} \otimes \mathcal{F}$ and the relation on $\Omega$ defined as

$$x \sim y \quad \Leftrightarrow \quad (x, y) \in E$$

is reflexive, symmetric and transitive.

The following notion of duality has been recently introduced by Jaffe [7]. Given a sub-$\sigma$-field $\mathcal{G} \subset \mathcal{F}$, the pair $(E, \mathcal{G})$ is said to satisfy *strong duality* if

$$\min_{P \in \Gamma(\mu,\nu)} (1 - P(E)) = \|\mu - \nu\|_{\mathcal{G}} \qquad \text{for all } \mu, \nu \in \mathcal{P}(\mathcal{F}).$$

*Accademia Navale di Livorno, Italy. E-mail: Luca_pratelli@marina.difesa.it
†Università di Bologna, Italy. E-mail: pietro.rigo@unibo.it

Here, as usual, $\Gamma(\mu,\nu)$ is the set of all probability measures on $\mathcal{F}\otimes\mathcal{F}$ with marginals $\mu$ and $\nu$ and the notation "$\min$" asserts that the infimum is actually achieved. Moreover,

$$\|\mu-\nu\|_{\mathcal{G}}=\sup_{A\in\mathcal{G}}|\mu(A)-\nu(A)|$$

is the total variation between $\mu$ and $\nu$ on $\mathcal{G}$.

Obviously, strong duality is strictly connected to mass transportation and Kantorovich duality; see Section 2. In addition, strong duality is intriguing from the foundational point of view and plays a role in some probabilistic frameworks, including stochastic calculus, point processes and random sequence simulation; see Section 2 of [7].

Say that $E$ is *strongly dualizable* if $(E,\mathcal{G})$ satisfies strong duality for *some* sub-$\sigma$-field $\mathcal{G}\subset\mathcal{F}$. Various conditions for $E$ to be strongly dualizable are given in [7] (see e.g. Theorems 3.13 and 3.14) but no measurable equivalence relation which fails to be strongly dualizable is known to date. This suggests the *conjecture* that, under mild conditions on $(\Omega,\mathcal{F})$ (say $(\Omega,\mathcal{F})$ is a standard Borel space), every measurable equivalence relation is strongly dualizable.

This paper focus on strong duality and includes three results. In a sense, these results state that $E$ is strongly dualizable as soon as a few mild modifications are admitted. Let

$$\mathcal{G}_0=\big\{A\in\mathcal{F}:\,1_A(x)=1_A(y)\text{ for all }(x,y)\in E\big\}$$

be the $E$-invariant sub-$\sigma$-field of $\mathcal{F}$. It is quite intuitive that $\mathcal{G}_0$ plays a role as regards strong duality. In fact, $E$ is strongly dualizable if and only if $(E,\mathcal{G}_0)$ satisfies strong duality; see [7, Proposition 3.15]. Our first result is that, for all $\mu,\,\nu\in\mathcal{P}(\mathcal{F})$, there is $\nu_0\in\mathcal{P}(\mathcal{F})$ satisfying

$$\nu_0=\nu\text{ on }\mathcal{G}_0\quad\text{and}\quad\min_{P\in\Gamma(\mu,\nu_0)}(1-P(E))=\|\mu-\nu\|_{\mathcal{G}_0}.$$

Roughly speaking, the above condition means that strong duality is always true up to changing one between $\mu$ and $\nu$ out of $\mathcal{G}_0$. This is quite reasonable, after all, for $\|\mu-\nu\|_{\mathcal{G}_0}$ only involves the restrictions of $\mu$ and $\nu$ on $\mathcal{G}_0$.

Next, suppose $(\Omega,\mathcal{F})$ is a standard Borel space and denote by $\widehat{\mathcal{F}}$ the collection of those subsets of $\Omega$ which are universally measurable with respect to $\mathcal{F}$; see Section 2. Define

$$\mathcal{G}_1=\big\{A\in\widehat{\mathcal{F}}:\,1_A(x)=1_A(y)\text{ for all }(x,y)\in E\big\}.$$

This time, $\mathcal{G}_1$ is not a sub-$\sigma$-field of $\mathcal{F}$. However, by our second result, one obtains

$$\inf_{P\in\Gamma(\mu,\nu)}(1-P(E))=\|\mu-\nu\|_{\mathcal{G}_1}\qquad\text{for all }\mu,\,\nu\in\mathcal{P}(\mathcal{F}).\tag{1.1}$$

In addition, the $\inf$ is achieved if $P$ is allowed to be finitely additive. Precisely,

$$\min_{P\in M(\mu,\nu)}(1-P(E))=\|\mu-\nu\|_{\mathcal{G}_1}\qquad\text{for all }\mu,\,\nu\in\mathcal{P}(\mathcal{F})$$

where $M(\mu,\nu)$ is the collection of finitely additive probabilities on $\mathcal{F}\otimes\mathcal{F}$ with marginals $\mu$ and $\nu$.

To state the third result, for each $B\subset\Omega$, define

$$\mathcal{G}_B=\big\{A\in\mathcal{F}:1_A(x)=1_A(y)\text{ for all }(x,y)\in E\cap(B\times B)\big\}.$$

Then, for all $\mu,\,\nu\in\mathcal{P}(\mathcal{F})$, there is a set $B\in\mathcal{F}$ such that

$$\mu(B)=\nu(B)=1\quad\text{and}\quad\inf_{P\in\Gamma(\mu,\nu)}(1-P(E))=\|\mu-\nu\|_{\mathcal{G}_B}.$$

If compared with (1.1), the latter result has the advantage that $\mathcal{G}_B$ is a sub-$\sigma$-field of $\mathcal{F}$ but the disadvantage that $\mathcal{G}_B$ is not universal, for it depends on the pair $(\mu,\nu)$. Note also that $\mu(B)=\nu(B)=1$ and $E\cap(B\times B)$ is a measurable equivalence relation on $B$. Therefore, for fixed $(\mu,\nu)$, one can replace $\Omega$ with $B$ and $E$ with $E\cap(B\times B)$. After doing this, everything works as regards the total variation side of strong duality.

A last remark is in order. For fixed $\mu,\,\nu\in\mathcal{P}(\mathcal{F})$, let us call *equivalence coupling problem* the minimization of $1-P(E)$ over $P\in\Gamma(\mu,\nu)$ and *total variation problem* the maximization of $|\mu(A)-\nu(A)|$ over $A\in\mathcal{G}$. In this paper, since $E$ is given, the equivalence coupling problem is regarded as *primal* while the total variation problem is viewed as *dual*. But of course this perspective can be reverted. Indeed, [7] contains results in which the total variation problem is primal and the equivalence coupling problem is dual.

## 2 Preliminaries

In this section, we introduce some further notation and recall a few known facts.

Let $(S,\mathcal{E})$ be a measurable space. Then, $\mathcal{P}(\mathcal{E})$ denotes the set of probability measures on $\mathcal{E}$ and $b\mathcal{E}$ the set of bounded $\mathcal{E}$-measurable functions $f:S\rightarrow\mathbb{R}$. For each $\mu\in\mathcal{P}(\mathcal{E})$, we write

$$\mu(f)=\int f\,d\mu\qquad\text{whenever } f\in b\mathcal{E},$$

and we denote by $\mu^*$ and $\mu_*$ the outer and inner measures corresponding to $\mu$. Precisely, $\mu^*$ and $\mu_*$ are defined as

$$\mu^*(A)=\inf\bigl\{\mu(B):B\in\mathcal{E},\;B\supset A\bigr\}\quad\text{and}\quad\mu_*(A)=\sup\bigl\{\mu(B):B\in\mathcal{E},\;B\subset A\bigr\}$$

for all $A\subset S$. Moreover, we let

$$\widehat{\mathcal{E}}=\bigcap_{\mu\in\mathcal{P}(\mathcal{E})}\overline{\mathcal{E}}^{\mu}$$

where $\overline{\mathcal{E}}^{\mu}$ is the completion of $\mathcal{E}$ with respect to $\mu$. The elements of $\widehat{\mathcal{E}}$ are usually called *universally measurable* with respect to $\mathcal{E}$. With a slight abuse of notation, for each $\mu\in\mathcal{P}(\mathcal{E})$, the unique extension of $\mu$ to $\widehat{\mathcal{E}}$ is still denoted by $\mu$.

If $T$ is any topological space, $\mathcal{B}(T)$ denotes the Borel $\sigma$-field. We say that $T$ is *Polish* if its topology is induced by a distance $d$ such that $(T,d)$ is a complete separable metric space. If $T$ is Polish, each analytic subset $A\subset T$ is universally measurable with respect to $\mathcal{B}(T)$, that is, $A\in\widehat{\mathcal{B}(T)}$.

The measurable space $(S,\mathcal{E})$ is a *standard Borel space* if $\mathcal{E}=\mathcal{B}(S)$ for some Polish topology on $S$.

A probability $\mu\in\mathcal{P}(\mathcal{E})$ is *perfect* if, for any $\mathcal{E}$-measurable function $f:S\rightarrow\mathbb{R}$, there is a Borel set $B\in\mathcal{B}(\mathbb{R})$ such that $B\subset f(S)$ and $\mu(f\in B)=1$. In a sense, perfectness is a non-topological version of the notion of tightness. In fact, if $S$ is separable metric and $\mathcal{E}=\mathcal{B}(S)$, then $\mu$ is perfect if and only if it is tight. In particular, each element of $\mathcal{P}(\mathcal{E})$ is perfect whenever $(S,\mathcal{E})$ is a standard Borel space. We refer to [13] for more on perfect probability measures.

As regards duality theory in mass transportation, we just mention a result by Ramachandran and Rüschendorf [14, Theorem 4]. For more information, the interested reader is referred to [2], [3], [9], [12], [16], [19] and references therein. Given $\mu,\,\nu\in\mathcal{P}(\mathcal{E})$, let $\Gamma(\mu,\nu)$ be the collection of probability measures $P$ on $\mathcal{E}\otimes\mathcal{E}$ with marginals $\mu$ and $\nu$, i.e.

$$P(A\times S)=\mu(A)\quad\text{and}\quad P(S\times A)=\nu(A)\quad\text{for all } A\in\mathcal{E}.$$

Moreover, let $c:S\times S\rightarrow\mathbb{R}$ be a bounded measurable cost function. (Boundedness of $c$ is generally superfluous and has been assumed for the sake of simplicity only). A *primal minimizer*, or an *optimal coupling*, is a probability measure $P \in \Gamma(\mu,\nu)$ such that $P(c) \le Q(c)$ for each $Q \in \Gamma(\mu,\nu)$. For a primal minimizer to exist, it suffices that $S$ is separable metric, $\mathcal{E} = \mathcal{B}(S)$, $\mu$ and $\nu$ are perfect, and the cost $c$ is lower semi-continuous. To state the duality result, we denote by $L$ the set of pairs $(f,g)$ satisfying

$$f,\, g \in b\mathcal{E} \quad \text{and} \quad f(x) + g(y) \le c(x,y) \text{ for all } (x,y) \in S \times S.$$

Then, in view of [14, Theorem 4], one obtains

$$\inf_{P \in \Gamma(\mu,\nu)} P(c) = \sup_{(f,g)\in L} \big\{\mu(f) + \nu(g)\big\}$$

provided at least one between $\mu$ and $\nu$ is perfect.

We finally turn to total variation distance. Let $\mathcal{D} \subset \mathcal{E}$ be a sub-$\sigma$-field and $\mu,\, \nu \in \mathcal{P}(\mathcal{E})$. The total variation between $\mu$ and $\nu$ on $\mathcal{D}$ is

$$\|\mu - \nu\|_{\mathcal{D}} = \sup_{A\in\mathcal{D}} |\mu(A) - \nu(A)| = \sup_{\substack{f \in b\mathcal{D} \\ 0 \le f \le 1}} |\mu(f) - \nu(f)|.$$

It is well known that $\|\cdot\|_{\mathcal{D}}$ can be written as

$$\|\mu - \nu\|_{\mathcal{D}} = \mu(A) - \nu(A) \qquad \text{for a suitable } A \in \mathcal{D}. \tag{2.1}$$

A last remark is in order. If $(\Phi, \mathcal{C}, \mathbb{Q})$ is any probability space and $H \subset \Phi$ is an arbitrary subset, there is a probability measure $\mathbb{P}$ on the $\sigma$-field $\sigma\big(\mathcal{C} \cup \{H\}\big)$ such that $\mathbb{P} = \mathbb{Q}$ on $\mathcal{C}$ and $\mathbb{P}(H) = \mathbb{Q}_*(H)$; see e.g. Theorem 1.12.14, p. 58, of [5]. As a consequence,

$$\|\mu - \nu\|_{\mathcal{D}} \le \mathbb{Q}_*\big(X \ne Y\big)$$

whenever $X,\, Y : (\Phi, \mathcal{C}) \to (S, \mathcal{D})$ are measurable maps such that $\mathbb{Q}(X \in A) = \mu(A)$ and $\mathbb{Q}(Y \in A) = \nu(A)$ for all $A \in \mathcal{D}$. Define in fact $H = \{X \ne Y\}$. Then, for every $A \in \mathcal{D}$,

$$\begin{aligned} &|\mu(A) - \nu(A)| = |\mathbb{Q}(X \in A) - \mathbb{Q}(Y \in A)| = |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \\ &= |\mathbb{P}\big(X \in A,\, X \ne Y\big) - \mathbb{P}\big(Y \in A,\, X \ne Y\big)| \le \mathbb{P}\big(X \ne Y\big) = \mathbb{Q}_*\big(X \ne Y\big). \end{aligned}$$

## 3 Two weak results

The results of this section have been termed "weak" as they concern the $\inf$ and not the $\min$ over $\Gamma(\mu,\nu)$.

It is quite intuitive that, when investigating strong duality, the partition of $\Omega$ in the equivalence classes of $E$ plays a role. Let $\Pi$ denote such a partition, i.e.

$$\Pi = \big\{[x] : x \in \Omega\big\} \qquad \text{where } [x] = \big\{y \in \Omega : (x,y) \in E\big\}.$$

The $\sigma$-fields $\mathcal{G}_0$ and $\mathcal{G}_1$, introduced in Section 1, can be written as

$$\mathcal{G}_0 = \big\{A \in \mathcal{F} : A \text{ is a union of elements of } \Pi\big\},$$
$$\mathcal{G}_1 = \big\{A \in \widehat{\mathcal{F}} : A \text{ is a union of elements of } \Pi\big\},$$

where $\widehat{\mathcal{F}}$ denotes the universally measurable $\sigma$-field with respect to $\mathcal{F}$. By "a union of elements of $\Pi$", we mean "an arbitrary union of elements of $\Pi$"; in particular, the union is not necessarily countable. In descriptive set theory and ergodic theory, the sets which are union of elements of $\Pi$ are usually called $E$-invariant sets. Another useful fact, often used in the sequel, is

$$1_A(x) - 1_A(y) \le 1 - 1_E(x,y) \qquad \text{for all } (x,y) \in \Omega \times \Omega \tag{3.1}$$

provided the set $A \subset \Omega$ is a union of elements of $\Pi$.

Our starting point is the following.

**Theorem 3.1.** *If* $(\Omega, \mathcal{F})$ *is a standard Borel space, then*

$$\inf_{P\in\Gamma(\mu,\nu)}(1-P(E)) = \|\mu-\nu\|_{\mathcal{G}_1} \qquad \text{for all } \mu,\, \nu \in \mathcal{P}(\mathcal{F}).$$

*Proof.* Let $\mu,\, \nu \in \mathcal{P}(\mathcal{F})$. In the notation of Section 2, let $c = 1 - 1_E$ and

$$L = \big\{(f,g):\ f,\, g \in b\mathcal{F} \text{ and } f(x)+g(y) \le c(x,y) \text{ for all } (x,y)\in\Omega\times\Omega\big\}.$$

Since $(\Omega,\mathcal{F})$ is standard Borel, $\mu$ and $\nu$ are perfect. Hence, by the duality result mentioned in Section 2, it follows that

$$\inf_{P\in\Gamma(\mu,\nu)}(1-P(E)) = \inf_{P\in\Gamma(\mu,\nu)} P(c) = \sup_{(f,g)\in L}\big\{\mu(f)+\nu(g)\big\}.$$

Given $(f,g)\in L$, define

$$\phi = \big(f - \sup f + 1\big)^+ \quad \text{and} \quad \psi = g + \sup f - 1.$$

On noting that

$$\sup f + \sup g = \sup_{(x,y)}\{f(x)+g(y)\} \le \sup c \le 1,$$

one obtains $(\phi,\psi)\in L$. Moreover, $0\le\phi\le 1$ and $\mu(\phi)+\nu(\psi)\ge\mu(f)+\nu(g)$. Hence,

$$\inf_{P\in\Gamma(\mu,\nu)}(1-P(E)) = \sup_{(f,g)\in L}\big\{\mu(f)+\nu(g)\big\} = \sup_{\substack{(f,g)\in L\\ 0\le f\le 1}}\{\mu(f)+\nu(g)\}.$$

Next, fix $\epsilon>0$ and take $(f,g)\in L$ such that $0\le f\le 1$ and

$$\mu(f)+\nu(g)+\epsilon > \inf_{P\in\Gamma(\mu,\nu)}(1-P(E)).$$

Define

$$h(x) = \sup_{y\in[x]} f(y)$$

and note that $h(x)+g(y)\le c(x,y)$ for all $(x,y)$. Letting $y=x$, one obtains

$$g(x) \le c(x,x) - h(x) = -h(x) \qquad \text{for all } x\in\Omega.$$

Since $h(x)=h(y)$ whenever $(x,y)\in E$, for each $a\in\mathbb{R}$ the set $\big\{h>a\big\}$ is a union of elements of $\Pi$. Moreover,

$$\big\{h>a\big\} = \big\{x\in\Omega : (x,y)\in E \text{ and } f(y)>a \text{ for some } y\in\Omega\big\}$$

is the projection on the first coordinate of the set

$$E\cap\big(\Omega\times\{f>a\}\big)\in\mathcal{F}\otimes\mathcal{F}.$$

Since $(\Omega,\mathcal{F})$ is standard Borel, the projection theorem yields $\big\{h>a\big\}\in\widehat{\mathcal{F}}$; see e.g. Theorem A1.4, page 562, of [8]. Hence, $\big\{h>a\big\}\in\mathcal{G}_1$. To sum up,

$$h\in b\mathcal{G}_1,\quad 0\le h\le 1,\quad h\ge f,\quad -h\ge g.$$

Therefore,

$$\|\mu-\nu\|_{\mathcal{G}_1} = \sup_{\substack{f\in b\mathcal{G}_1\\ 0\le f\le 1}} |\mu(f)-\nu(f)| \ge \mu(h)-\nu(h)$$
$$\ge \mu(f)+\nu(g) > \inf_{P\in\Gamma(\mu,\nu)}(1-P(E)) - \epsilon.$$

Finally, fix $P \in \Gamma(\mu,\nu)$ and $A \in \mathcal{G}_1$. Since $A$ is a union of elements of $\Pi$, inequality (3.1) yields

$$1 - P(E) \geq \int \big\{1_A(x) - 1_A(y)\big\}\, P(dx, dy) = \mu(A) - \nu(A).$$

Hence,

$$\inf_{P \in \Gamma(\mu,\nu)} (1 - P(E)) \geq \|\mu - \nu\|_{\mathcal{G}_1}$$

and this concludes the proof. □

The function $h$ involved in the proof of Theorem 3.1 is also called the least $E$-invariant majorant of $f$. Letting $c = 1 - 1_E$ and recalling that $0 \leq f \leq 1$, one obtains

$$-h(x) = \inf_{y \in [x]} \big\{-f(y)\big\} = \inf_{y \in \Omega} \big\{1 - 1_E(x,y) - f(y)\big\} = \inf_{y \in \Omega} \big\{c(x,y) - f(y)\big\}.$$

Hence, in the mass transportation terminology, $-h$ is the $c$-transform of $f$. In general, $h$ is $\widehat{\mathcal{F}}$-measurable, as proved above, but not necessarily $\mathcal{F}$-measurable; see Remarks 5.5 and 5.11 of [19]. For this reason, $\mathcal{G}_1$ (and not $\mathcal{G}_0$) comes into play. An open question is to assign conditions on $E$ under which $h$ is $\mathcal{F}$-measurable. Under such conditions, $\|\mu - \nu\|_{\mathcal{G}_1}$ could be replaced by $\|\mu - \nu\|_{\mathcal{G}_0}$ in Theorem 3.1. We also note that, for a lower semi-continuous cost function $c$, measurability of the $c$-transform is discussed in [19, p. 69]. This discussion, however, does not fit to $c = 1 - 1_E$.

If regarded as a tool to get strong duality, Theorem 3.1 has two gaps:

- $\mathcal{G}_1$ is not a sub-$\sigma$-field of $\mathcal{F}$;
- Theorem 3.1 is a weak result, for it involves the $\inf$ and not the $\min$ over $\Gamma(\mu,\nu)$.

The second gap is concerned in the next section. Here, we focus on the first, that is, we replace $\mathcal{G}_1$ with a suitable sub-$\sigma$-field of $\mathcal{F}$.

**Theorem 3.2.** *If $(\Omega,\mathcal{F})$ is a standard Borel space, then, for all $\mu$, $\nu \in \mathcal{P}(\mathcal{F})$, there is a set $B \in \mathcal{F}$ such that*

$$\mu(B) = \nu(B) = 1 \quad \textit{and} \quad \inf_{P \in \Gamma(\mu,\nu)} (1 - P(E)) = \|\mu - \nu\|_{\mathcal{G}_B}$$

*where $\mathcal{G}_B = \big\{A \in \mathcal{F} : 1_A(x) = 1_A(y)$ for all $(x,y) \in E \cap (B \times B)\big\}$.*

*Proof.* Let $\mu$, $\nu \in \mathcal{P}(\mathcal{F})$. By (2.1) and Theorem 3.1, there is $D \in \mathcal{G}_1$ such that

$$\inf_{P \in \Gamma(\mu,\nu)} (1 - P(E)) = \|\mu - \nu\|_{\mathcal{G}_1} = \mu(D) - \nu(D).$$

Since $D$ is universally measurable with respect to $\mathcal{F}$, there is $A \in \mathcal{F}$ such that

$$\frac{\mu + \nu}{2}\big(A \Delta D\big) = 0,$$

or equivalently $\mu(A\Delta D) = \nu(A\Delta D) = 0$. Let

$$T = \big\{(x,y) \in E : 1_A(x) \neq 1_A(y)\big\}.$$

Since $D$ is a union of elements of $\Pi$, then $1_D(x) = 1_D(y)$ for all $(x,y) \in E$. Hence,

$$P(T) = P\big\{(x,y) \in E : 1_D(x) \neq 1_D(y)\big\} = P(\emptyset) = 0 \quad \text{for each } P \in \Gamma(\mu,\nu)$$

where the first equality is because $\mu(A\Delta D) = \nu(A\Delta D) = 0$. By a (deep) result of Arveson, Haydon and Shulman, since $(\Omega,\mathcal{F})$ is standard Borel and $P(T) = 0$ for all $P \in \Gamma(\mu,\nu)$, there is $B \in \mathcal{F}$ such that $\mu(B) = \nu(B) = 1$ and

$$T \subset (B^c \times \Omega) \cup (\Omega \times B^c);$$

see [1, Theorems 1.4.2 and 1.4.3], [6, Corollary, p. 500] and [16, p. 2345]. Therefore $A \in \mathcal{G}_B$, which in turn implies

$$\inf_{P \in \Gamma(\mu,\nu)} (1 - P(E)) = \mu(D) - \nu(D) = \mu(A) - \nu(A) \leq \|\mu - \nu\|_{\mathcal{G}_B}.$$

To prove the reverse inequality, fix any $C \in \mathcal{G}_B$ and $P \in \Gamma(\mu, \nu)$. Then,

$$P(B \times B) = 1 \quad \text{and} \quad 1_C(x) - 1_C(y) \leq 1 - 1_E(x, y) \quad \text{for all } (x, y) \in B \times B.$$

Hence,

$$\mu(C) - \nu(C) = \int \big(1_C(x) - 1_C(y)\big)\, P(dx, dy) \leq 1 - P(E),$$

which in turn implies $\|\mu - \nu\|_{\mathcal{G}_B} \leq \inf_{P \in \Gamma(\mu,\nu)} (1 - P(E))$. □

The advantage of Theorem 3.2 with respect to Theorem 3.1 is that $\mathcal{G}_B$ is a sub-$\sigma$-field of $\mathcal{F}$ while $\mathcal{G}_1$ is not. The disadvantage is that $\mathcal{G}_B$ is not universal, for it depends on the pair $(\mu, \nu)$. However, for fixed $(\mu, \nu)$, since $\mu(B) = \nu(B) = 1$ and $E \cap (B \times B)$ is a measurable equivalence relation on $B$, it is reasonable to replace $\Omega$ with $B$ and $E$ with $E \cap (B \times B)$. In other terms, for fixed $(\mu, \nu)$, it makes sense to involve $\mathcal{G}_B$ in the notion of strong duality.

## 4 Existence of primal minimizers

Quite surprisingly, in mass transportation theory, existence of primal minimizers seems to have received only a little attention to date; see e.g. [2] and [7, p. 4]. To our knowledge, when the cost $c$ is not lower semi-continuous, the only available results are in [12, Theorem 2.3.10] and require $c$ to be suitably approximable by regular costs. However, such results do not apply to our case where $c = 1 - 1_E$.

Let $(\Omega, \mathcal{F})$ be a standard Borel space and $c = 1 - 1_E$. Then, $c$ is lower semi-continuous if and only if $E$ is closed, and in this case $E$ is strongly dualizable. Similarly, $E$ is strongly dualizable if its equivalence classes are the atoms of a countably generated sub-$\sigma$-field of $\mathcal{F}$, or if $E$ is the union of an increasing sequence of strongly dualizable equivalence relations; see [7, Theorems 3.13 and 3.14] and [11, Theorem 1]. As noted above, however, we are not aware of any general condition for a primal minimizer to exist. In the sequel, we discuss two strategies for circumventing this problem.

The first strategy is possibly expected and lies in using finitely additive probabilities. Let

$$M(\mu, \nu) = \big\{\text{finitely additive probabilities on } \mathcal{F} \otimes \mathcal{F} \text{ with marginals } \mu \text{ and } \nu\big\}.$$

**Theorem 4.1.** *Let $(\Omega, \mathcal{F})$ be a standard Borel space. Then,*

$$\min_{P \in M(\mu,\nu)} (1 - P(E)) = \|\mu - \nu\|_{\mathcal{G}_1} \qquad \textit{for all } \mu, \nu \in \mathcal{P}(\mathcal{F}).$$

*Moreover, for all $\mu$, $\nu \in \mathcal{P}(\mathcal{F})$ there is $B \in \mathcal{F}$ such that*

$$\mu(B) = \nu(B) = 1 \quad \textit{and} \quad \min_{P \in M(\mu,\nu)} (1 - P(E)) = \|\mu - \nu\|_{\mathcal{G}_B}.$$

*Proof.* Just apply Theorems 3.1 and 3.2 and note that, by Theorem 2 of [17],

$$\min_{P \in M(\mu,\nu)} (1 - P(E)) = \inf_{P \in \Gamma(\mu,\nu)} (1 - P(E)).$$ □

A remark on Theorem 4.1 is in order. Let $P$ be a finitely additive primal minimizer, in the sense that $P \in M(\mu,\nu)$ and $1 - P(E) = \inf_{Q\in\Gamma(\mu,\nu)}(1 - Q(E))$. Moreover, let $\mathcal{R}$ be the field generated by the measurable rectangles $A \times B$ with $A,\, B \in \mathcal{F}$. Since $\mu$ and $\nu$ are perfect (due to $(\Omega,\mathcal{F})$ is standard Borel), the restriction $P|\mathcal{R}$ is $\sigma$-additive; see e.g. [15, Theorem 6]. Hence, it is tempting to define $P'$ as the only $\sigma$-additive extension of $P|\mathcal{R}$ to $\sigma(\mathcal{R}) = \mathcal{F} \otimes \mathcal{F}$. Then, $P' \in \Gamma(\mu,\nu)$ but it is *not* necessarily true that $P'(E) = P(E)$. Hence, $P'$ needs not be a primal minimizer.

The second strategy for dealing with primal minimizers is summarized by the next result.

**Theorem 4.2.** *For all $\mu,\, \nu \in \mathcal{P}(\mathcal{F})$, there is $\nu_0 \in \mathcal{P}(\mathcal{F})$ such that*

$$\nu_0 = \nu \text{ on } \mathcal{G}_0 \quad \text{and} \quad \min_{P\in\Gamma(\mu,\nu_0)}(1 - P(E)) = \|\mu - \nu\|_{\mathcal{G}_0}.$$

Before proving Theorem 4.2, we provide a lemma (which is possibly of some independent interest).

**Lemma 4.3.** *Let $(S,\mathcal{E})$ be a measurable space, $\mathcal{D} \subset \mathcal{E}$ a sub-$\sigma$-field and $\mu,\, \nu \in \mathcal{P}(\mathcal{E})$. Then, there are a probability space $(\Phi,\mathcal{A},\mathbb{P})$ and two measurable maps $X,\, Y : (\Phi,\mathcal{A}) \to (S,\mathcal{E})$ such that*

$$\mathbb{P}(X \in A) = \mu(A) \text{ for all } A \in \mathcal{E}, \quad \mathbb{P}(Y \in A) = \nu(A) \text{ for all } A \in \mathcal{D},$$
$$\big\{X \neq Y\big\} \in \mathcal{A} \quad \text{and} \quad \mathbb{P}(X \neq Y) = \|\mu - \nu\|_{\mathcal{D}}.$$

*Proof.* For any measure $\gamma$ on $\mathcal{E}$, we write $\gamma|\mathcal{D}$ to denote the restriction of $\gamma$ on $\mathcal{D}$.

Suppose first $\mu|\mathcal{D} = \nu|\mathcal{D}$. Let $\Phi = S \times S$, $\mathcal{C} = \mathcal{E} \otimes \mathcal{E}$ and $X(a,b) = a$ and $Y(a,b) = b$ for all $(a,b) \in S \times S$. Define also

$$\mathbb{Q}(C) = \mu\big\{x \in S : (x,x) \in C\big\} \qquad \text{for all } C \in \mathcal{C}.$$

Then, $\mathbb{Q}(X \in A) = \mathbb{Q}(Y \in A) = \mu(A)$ for all $A \in \mathcal{E}$. In particular, since $\mu = \nu$ on $\mathcal{D}$, then $\mathbb{Q}(Y \in A) = \nu(A)$ for all $A \in \mathcal{D}$. Moreover, since $\mathbb{Q}(C) = 0$ whenever $C \in \mathcal{C}$ and $C \subset \big\{X \neq Y\big\}$, one obtains $\mathbb{Q}_*\big(X \neq Y\big) = 0$ (where $\mathbb{Q}_*$ is the inner measure corresponding to $\mathbb{Q}$). Hence, by the extension theorem mentioned in Section 2, $\mathbb{Q}$ can be extended to a probability measure $\mathbb{P}$ on

$$\mathcal{A} = \sigma\big(\mathcal{C} \cup \big\{X \neq Y\big\}\big)$$

such that

$$\mathbb{P}(X \neq Y) = \mathbb{Q}_*\big(X \neq Y\big) = 0 = \|\mu - \nu\|_{\mathcal{D}}.$$

Suppose now that $\mu|\mathcal{D} \neq \nu|\mathcal{D}$. Define

$$\lambda = \mu + \nu, \quad f = \frac{d\,(\mu|\mathcal{D})}{d\,(\lambda|\mathcal{D})}, \quad g = \frac{d\,(\nu|\mathcal{D})}{d\,(\lambda|\mathcal{D})}, \quad \text{and}$$
$$\gamma(A) = \frac{1}{\|\mu - \nu\|_{\mathcal{D}}} \int_A (g - f)^+ \, d\lambda \qquad \text{for all } A \in \mathcal{E}.$$

Since $\int (g - f)^+ \, d\lambda = \|\mu - \nu\|_{\mathcal{D}}$, such a $\gamma$ is a probability measure on $\mathcal{E}$. Let $(\Phi,\mathcal{C},\mathbb{Q})$ be any probability space which supports three independent random variables $U,\, X,\, Z$ with $U$ uniformly distributed on $(0,1)$ and

$$\mathbb{Q}(X \in A) = \mu(A) \quad \text{and} \quad \mathbb{Q}(Z \in A) = \gamma(A) \quad \text{for all } A \in \mathcal{E}.$$

Define

$$G = \big\{f(X)\,U > g(X)\big\}, \quad Y = Z \text{ on } G \quad \text{and} \quad Y = X \text{ on } G^c.$$

Then,

$$\mathbb{Q}(G) = \mathbb{Q}\left[f(X) > g(X),\, U > \frac{g(X)}{f(X)}\right] = \int_{\{f>g\}} \left(1 - \frac{g}{f}\right) f\, d\lambda$$
$$= \int_{\{f>g\}} (f-g)\, d\lambda = \|\mu - \nu\|_{\mathcal{D}}.$$

Moreover, for each $A \in \mathcal{E}$,

$$\mathbb{Q}(Y \in A) = \mathbb{Q}\big(G \cap \{Z \in A\}\big) + \mathbb{Q}\big(G^c \cap \{X \in A\}\big)$$
$$= \mathbb{Q}(G)\, \mathbb{Q}(Z \in A) + \mathbb{Q}\left[f(X)\, U \le g(X),\, X \in A\right]$$
$$= \int_A (g-f)^+ d\lambda + \int_{A\cap\{f>g\}} \frac{g}{f}\, d\mu + \mu\big(A \cap \{f \le g\}\big).$$

If $A \in \mathcal{D}$, since $f = \frac{d\,(\mu|\mathcal{D})}{d\,(\lambda|\mathcal{D})}$, one obtains

$$\int_{A\cap\{f>g\}} \frac{g}{f}\, d\mu + \mu\big(A \cap \{f \le g\}\big) = \int_{A\cap\{f>g\}} \frac{g}{f}\, f\, d\lambda + \int_{A\cap\{f\le g\}} f\, d\lambda = \int_A (f \wedge g)\, d\lambda.$$

Therefore,

$$\mathbb{Q}(Y \in A) = \int_A (g-f)^+ d\lambda + \int_A (f \wedge g)\, d\lambda = \int_A g\, d\lambda = \nu(A) \qquad \text{for each } A \in \mathcal{D}.$$

It follows that

$$\|\mu - \nu\|_{\mathcal{D}} \le \mathbb{Q}_*(X \ne Y) \le \mathbb{Q}^*(X \ne Y) \le \mathbb{Q}(G) = \|\mu - \nu\|_{\mathcal{D}}$$

where the first inequality has been discussed in Section 2. Hence, to conclude the proof, it suffices to take $(\Phi, \mathcal{A}, \mathbb{P})$ as the completion of $(\Phi, \mathcal{C}, \mathbb{Q})$. □

Lemma 4.3 slightly improves some known results; see [4, Proposition 3.1] and [18, Lemma 2.1]. We also recall that the diagonal $\Delta = \{(x,x) : x \in S\}$ does not necessarily belong to $\mathcal{E} \otimes \mathcal{E}$; see e.g. Exercise 3.10.44 of [5]. A characterization of the measurable spaces $(S, \mathcal{E})$ such that $\Delta \in \mathcal{E} \otimes \mathcal{E}$ is in [5, Theorem 6.5.7].

***Proof of Theorem 4.2***. By Lemma 4.3, applied with $S = \Omega$, $\mathcal{E} = \mathcal{F}$ and $\mathcal{D} = \mathcal{G}_0$, there are a probability space $(\Phi, \mathcal{A}, \mathbb{P})$ and two measurable maps $X, Y : (\Phi, \mathcal{A}) \to (\Omega, \mathcal{F})$ such that

$$P(X \in A) = \mu(A) \text{ for all } A \in \mathcal{F}, \quad P(Y \in A) = \nu(A) \text{ for all } A \in \mathcal{G}_0,$$
$$\big\{X \ne Y\big\} \in \mathcal{A} \quad \text{and} \quad \mathbb{P}(X \ne Y) = \|\mu - \nu\|_{\mathcal{G}_0}.$$

Up to replacing $(\Phi, \mathcal{A}, \mathbb{P})$ with its completion, it can be assumed that $(\Phi, \mathcal{A}, \mathbb{P})$ is complete. Because of (3.1),

$$1_{\{X\in A\}} - 1_{\{Y\in A\}} \le 1_{\{(X,Y)\notin E\}} \le 1_{\{X\ne Y\}} \qquad \text{for each } A \in \mathcal{G}_0.$$

Therefore,

$$\mu(A) - \nu(A) = \int \big(1_{\{X\in A\}} - 1_{\{Y\in A\}}\big)\, d\mathbb{P} \le \mathbb{P}_*\big((X,Y) \notin E\big)$$
$$\le \mathbb{P}^*\big((X,Y) \notin E\big) \le \mathbb{P}(X \ne Y) = \|\mu - \nu\|_{\mathcal{G}_0} \qquad \text{for each } A \in \mathcal{G}_0,$$

which in turn implies

$$\|\mu - \nu\|_{\mathcal{G}_0} \le \mathbb{P}_*\big((X,Y) \notin E\big) \le \mathbb{P}^*\big((X,Y) \notin E\big) \le \|\mu - \nu\|_{\mathcal{G}_0}.$$

Since $(\Phi, \mathcal{A}, \mathbb{P})$ is complete, one obtains

$$\big\{(X,Y)\notin E\big\}\in\mathcal{A}\quad\text{and}\quad \mathbb{P}\big((X,Y)\notin E\big)=\|\mu-\nu\|_{\mathcal{G}_0}.$$

To conclude the proof, note that $\big\{(X,Y)\in H\big\}\in\mathcal{A}$ for each $H\in\mathcal{F}\otimes\mathcal{F}$ and define

$$\nu_0(A)=\mathbb{P}(Y\in A)\quad\text{and}\quad P(H)=\mathbb{P}\big((X,Y)\in H\big)\qquad\text{for all } A\in\mathcal{F} \text{ and } H\in\mathcal{F}\otimes\mathcal{F}.$$

Then, $\nu_0=\nu$ on $\mathcal{G}_0$, $P\in\Gamma(\mu,\nu_0)$ and

$$1-P(E)=\|\mu-\nu\|_{\mathcal{G}_0}\leq 1-Q(E)\qquad\text{for each } Q\in\Gamma(\mu,\nu_0).$$ □

It is worth noting that, in Theorem 4.2, $(\Omega,\mathcal{F})$ is not required to be a standard Borel space. In addition, Theorem 4.2 has the following useful consequence.

**Corollary 4.4.** *Let $(\Omega,\mathcal{F})$ be a standard Borel space.*

*(a) If $E\in\mathcal{F}\otimes\mathcal{G}_0$, then*

$$\min_{P\in\Gamma(\mu,\nu)}(1-P(E))=\|\mu-\nu\|_{\mathcal{G}_0}\qquad\textit{for all } \mu,\,\nu\in\mathcal{P}(\mathcal{F}).$$

*(b) If $E\in\widehat{\mathcal{F}}\otimes\mathcal{G}_1$ (and even if $E\notin\mathcal{F}\otimes\mathcal{F}$), then*

$$\min_{P\in\Gamma(\mu,\nu)}(1-P(E))=\|\mu-\nu\|_{\mathcal{G}_1}\qquad\textit{for all } \mu,\,\nu\in\mathcal{P}(\mathcal{F}).$$

*Proof.* Recall that, for each $\gamma\in\mathcal{P}(\mathcal{F})$, the only extension of $\gamma$ to $\widehat{\mathcal{F}}$ is still denoted by $\gamma$. Moreover, since $(\Omega,\mathcal{F})$ is standard Borel, every probability measure on $\widehat{\mathcal{F}}$ is perfect.

**Part (a).** Suppose $E\in\mathcal{F}\otimes\mathcal{G}_0$. Given $\mu,\,\nu\in\mathcal{P}(\mathcal{F})$, by Theorem 4.2, there are $\nu_0\in\mathcal{P}(\mathcal{F})$ and $P_0\in\Gamma(\mu,\nu_0)$ such that $\nu_0=\nu$ on $\mathcal{G}_0$ and $1-P_0(E)=\|\mu-\nu\|_{\mathcal{G}_0}$. In addition, since $\mu$ and $\nu$ are perfect, by Theorem 9 of [15], there is $P\in\Gamma(\mu,\nu)$ such that $P=P_0$ on $\mathcal{F}\otimes\mathcal{G}_0$. Since $E\in\mathcal{F}\otimes\mathcal{G}_0$, one obtains $P(E)=P_0(E)$. Therefore,

$$P\in\Gamma(\mu,\nu)\quad\text{and}\quad 1-P(E)=\|\mu-\nu\|_{\mathcal{G}_0}\leq 1-Q(E)\qquad\text{for each } Q\in\Gamma(\mu,\nu)$$

where the inequality is by (3.1).

**Part (b).** Suppose $E\in\widehat{\mathcal{F}}\otimes\mathcal{G}_1$. Let $\widehat{\Gamma}(\mu,\nu)$ be the collection of probability measures $\widehat{P}$ on $\widehat{\mathcal{F}}\otimes\widehat{\mathcal{F}}$ such that

$$\widehat{P}(A\times\Omega)=\mu(A)\quad\text{and}\quad\widehat{P}(\Omega\times A)=\nu(A)\quad\text{for all } A\in\widehat{\mathcal{F}}.$$

Since $E\in\widehat{\mathcal{F}}\otimes\mathcal{G}_1$ and $\mu$ and $\nu$ are perfect (where $\mu$ and $\nu$ are now regarded as probability measures on $\widehat{\mathcal{F}}$) the proof of part (a) can be repeated with $(\Omega,\widehat{\mathcal{F}})$ and $\mathcal{G}_1$ in the place of $(\Omega,\mathcal{F})$ and $\mathcal{G}_0$. Hence,

$$1-\widehat{P}(E)=\|\mu-\nu\|_{\mathcal{G}_1}\quad\text{for some}\quad\widehat{P}\in\widehat{\Gamma}(\mu,\nu).$$

Finally, denoting by $P$ the restriction of $\widehat{P}$ on $\mathcal{F}\otimes\mathcal{F}$, one obtains

$$P\in\Gamma(\mu,\nu)\quad\text{and}\quad 1-P(E)=\|\mu-\nu\|_{\mathcal{G}_1}\leq 1-Q(E)\qquad\text{for each } Q\in\Gamma(\mu,\nu).$$ □

Corollary 4.4-(a) slightly improves [7, Theorem 3.13]. The former requires in fact $E\in\mathcal{F}\otimes\mathcal{G}_0$ while the latter $E\in\mathcal{G}_0\otimes\mathcal{G}_0$. However, we do not know of any example where $E\in\mathcal{F}\otimes\mathcal{G}_0$ but $E\notin\mathcal{G}_0\otimes\mathcal{G}_0$. Instead, to our knowledge, Corollary 4.4-(b) is new. Among other things, since $E$ is not forced to belong to $\mathcal{F}\otimes\mathcal{F}$, it allows to handle situations where $E$ is analytic but not Borel.

**Example 4.5.** Let $\Omega$ be a Polish space and $\mathcal{F} = \mathcal{B}(\Omega)$. A subset of $\Omega$ is a $G_\delta$ if it is a countable intersection of open sets. In particular, open and closed subsets of $\Omega$ are both $G_\delta$. The next result is a consequence of Corollary 4.4-(b).

*If $E$ is analytic and the equivalence classes of $E$ are $G_\delta$, then*

$$\min_{P \in \Gamma(\mu,\nu)} (1 - P(E)) = \|\mu - \nu\|_{\mathcal{G}_1} \qquad \text{for all } \mu,\, \nu \in \mathcal{P}(\mathcal{F}).$$

To prove this claim, it suffices to show that $E \in \widehat{\mathcal{F}} \otimes \mathcal{G}_1$. For $A \in \mathcal{F}$, define

$$A^* = \{x \in \Omega : \exists\, y \in A \text{ such that } (x, y) \in E\}.$$

Then, $A^*$ is analytic, as it is the projection on the first coordinate of the analytic set $E \cap (\Omega \times A)$. Hence, $A^* \in \widehat{\mathcal{F}}$. Since $A^*$ is a union of equivalence classes of $E$, one also obtains $A^* \in \mathcal{G}_1$. Having noted this fact, fix a countable basis $\mathcal{U}$ for the topology of $\Omega$ and define

$$\mathcal{V} = \sigma(U^* : U \in \mathcal{U}).$$

Then, $\mathcal{V}$ is countably generated and $\mathcal{V} \subset \mathcal{G}_1$. If $A$ and $B$ are any disjoint $G_\delta$ sets, there is $U \in \mathcal{U}$ such that

$$A \cap U \neq \emptyset \text{ and } B \cap U = \emptyset \quad \text{or} \quad A \cap U = \emptyset \text{ and } B \cap U \neq \emptyset;$$

see the proof of Lemma 2 in [10]. Hence, if $A$ and $B$ are two disjoint equivalence classes of $E$, then

$$A \subset U^* \text{ and } B \cap U^* = \emptyset \quad \text{or} \quad A \cap U^* = \emptyset \text{ and } B \subset U^*$$

for some $U \in \mathcal{U}$. This implies that the equivalence classes of $E$ are precisely the atoms of $\mathcal{V}$. Finally, since $\mathcal{V}$ is countably generated, there is a function $f : \Omega \to \mathbb{R}$ such that $\mathcal{V} = \sigma(f)$. Therefore,

$$E = \big\{(x, y) : f(x) = f(y)\big\} \in \mathcal{V} \otimes \mathcal{V} \subset \widehat{\mathcal{F}} \otimes \mathcal{G}_1.$$

We close this paper with a last result. While not practically useful, it still provides some information on primal minimizers.

**Theorem 4.6.** *Let $(\Omega, \mathcal{F})$ be a standard Borel space and $P \in \Gamma(\mu, \nu)$ for some $\mu,\, \nu \in \mathcal{P}(\mathcal{F})$. Then, $P$ is a primal minimizer (with respect to $c = 1 - 1_E$) if and only if*

$$P(E) = 1 - P(A \times A^c) \qquad \text{for some } A \in \mathcal{G}_1. \tag{4.1}$$

*Proof.* By (2.1), there is $A \in \mathcal{G}_1$ such that $\|\mu - \nu\|_{\mathcal{G}_1} = \mu(A) - \nu(A)$. Hence, if $P$ is a primal minimizer, Theorem 3.1 implies

$$\begin{aligned} 1 - P(E) = \|\mu - \nu\|_{\mathcal{G}_1} = \mu(A) - \nu(A) &= \int \big(1_A(x) - 1_A(y)\big)\, P(dx, dy) \\ &\le P(A \times A^c) \le P(E^c) = 1 - P(E). \end{aligned}$$

Thus, condition (4.1) holds. Conversely, if (4.1) holds for some $A \in \mathcal{G}_1$, then

$$\begin{aligned} P\big\{(x, y) : 1_A(x) - 1_A(y) = 1 - 1_E(x, y)\big\} &= P(E) + P\big\{(x, y) \in E^c : 1_A(x) - 1_A(y) = 1\big\} \\ &= P(E) + P\big(E^c \cap (A \times A^c)\big) = P(E) + P(A \times A^c) = 1. \end{aligned}$$

Therefore, for each $Q \in \Gamma(\mu, \nu)$,

$$\begin{aligned} 1 - P(E) &= \int \big(1_A(x) - 1_A(y)\big)\, P(dx, dy) = \mu(A) - \nu(A) \\ &= \int \big(1_A(x) - 1_A(y)\big)\, Q(dx, dy) \le 1 - Q(E) \end{aligned}$$

where the last inequality is by (3.1). Hence, $P$ is a primal minimizer. □

Theorem 4.6 is very similar to Proposition 3.12 of [7]. Both provide characterizations of primal minimizers. The only difference is that, in Proposition 3.12, it is required a priori that $E$ is strongly dualizable.

## References


[1] Arveson, W.: Operator algebras and invariant subspaces, *Ann. Math.* **100**, (1974), 433-533. MR0365167

[2] Beiglböck, M., Léonard, C. and Schachermayer, W.: On the duality theory for the Monge-Kantorovich transport problem. In: *Optimal transportation, London Math. Soc. Lecture Note Ser.* **413**, (2014), 216-265, Cambridge Univ. Press. MR3328997

[3] Beiglböck, M., Nutz, M. and Touzi, N.: Complete duality for martingale optimal transport on the line. *Ann. Probab.* **45**, (2017), 3038-3074. MR3706738

[4] Berti, P., Pratelli, L. and Rigo, P.: Skorohod representation theorem via disintegrations. *Sankhya* **72**, (2010), 208-220. MR2658171

[5] Bogachev, V.I.: *Measure theory, Vol. I and II.* Springer, Berlin, 2007. MR2267655

[6] Haydon, R. and Shulman, V.: On a measure-theoretic problem of Arveson. *Proc. Amer. Math. Soc.* **124**, (1996), 497–503. MR1301501

[7] Jaffe, A.Q.: A strong duality principle for equivalence couplings and total variation. *Electronic J. Probab.* **28**, (2023), 1-33. MR4664456

[8] Kallenberg, O.: *Foundations of modern probability, Second edition.* Springer, New York, 2002. MR1876169

[9] Kellerer, H.G.: Duality theorems for marginal problems. *Z.Wahrscheinlichkeitstheorie Verw. Geb.* **67**, (1984), 399-432. MR0761565

[10] Miller, D.E.: A selector for equivalence relations with $G_\delta$ orbits, *Proc. Amer. Math. Soc.* **72**, (1978), 365-369. MR0515142

[11] Pratelli, L. and Rigo, P.: A strong version of the Skorohod representation theorem. *J. Theor. Probab.* **36**, (2023), 372-389. MR4585120

[12] Rachev, S.T. and Rüschendorf, L.: *Mass transportation problems, Volume I: Theory.* Springer, New York, 1998. MR1619171

[13] Ramachandran, D.: Perfect measures I and II. *ISI-Lecture Notes Series* **5** and **7**, (1979), New Delhi, Macmillan. MR0553600

[14] Ramachandran, D. and Rüschendorf, L.: A general duality theorem for marginal problems. *Probab. Theory Relat. Fields* **101**, (1995), 311-319. MR1324088

[15] Ramachandran, D.: The marginal problem in arbitrary product spaces. In: *Distributions with fixed marginals and related topics* (Rüschendorf, Schweizer and Taylor eds.) IMS Lect. Notes Monog. Series **28**, (1996), 260-272. MR1485537

[16] Rigo, P.: A note on duality theorems in mass transportation. *J. Theor. Probab.* **33**, (2020), 2337-2350. MR4166202

[17] Rigo, P.: Finitely additive mass transportation. *Bernoulli* **30**, (2024).

[18] Sethuraman, J.: Some extensions of the Skorohod representation theorem. *Sankhya* **64**, (2002), 884-893. MR1981517

[19] Villani, C.: *Optimal transport. Old and new.* Vol. 338 of Grundlehren der mathematischen Wissenschaften, Springer, Berlin, 2009. MR2459454


**Acknowledgments.** We are grateful to an anonymous referee for many comments and remarks which improved this paper.