Author(s): Leskelä, Lasse; Vihola, Matti

Title: Conditional convex orders and measurable martingale couplings

Year: 2017

Version:

Please cite the original version:
Leskelä, L., & Vihola, M. (2017). Conditional convex orders and measurable martingale couplings. Bernoulli, 23(4A), 2784-2807. https://doi.org/10.3150/16-BEJ827

Bernoulli **23**(4A), 2017, 2784–2807
DOI: 10.3150/16-BEJ827

Conditional convex orders and measurable martingale couplings

LASSE LESKELÄ[1] and MATTI VIHOLA[2]

[1]*Department of Mathematics and Systems Analysis, School of Science, Aalto University, PO Box 11100, FI-00076 Aalto, Finland. E-mail: lasse.leskela@aalto.fi*
[2]*Department of Mathematics and Statistics, University of Jyväskylä, PO Box 35, FI-40014 Jyväskylä, Finland. E-mail: matti.vihola@iki.fi*

Strassen's classical martingale coupling theorem states that two random vectors are ordered in the convex (resp. increasing convex) stochastic order if and only if they admit a martingale (resp. submartingale) coupling. By analysing topological properties of spaces of probability measures equipped with a Wasserstein metric and applying a measurable selection theorem, we prove a conditional version of this result for random vectors conditioned on a random element taking values in a general measurable space. We provide an analogue of the conditional martingale coupling theorem in the language of probability kernels, and discuss how it can be applied in the analysis of pseudo-marginal Markov chain Monte Carlo algorithms. We also illustrate how our results imply the existence of a measurable minimiser in the context of martingale optimal transport.

Keywords: conditional coupling; convex stochastic order; increasing convex stochastic order; martingale coupling; pointwise coupling; probability kernel

1. Introduction and main results

1.1. Convex stochastic orders

Stochastic orders and relations provide powerful tools to compare distributions of random variables and processes, and they have been used in various applications [22,25,28,33]. We focus here on two closely related stochastic orders which are characterised by expectations of convex functionals, the convex order and the increasing convex order. The convex order is a common measure of 'variability' or 'dispersion' of random variables and vectors, and it arises naturally for example in majorisation [24]. The increasing convex order allows to compare also random vectors with different means.

Let μ and ν be probability measures on $\mathbb{R}^d$. We say that μ is less than ν in the *convex order*, denoted $\mu \le_{\mathrm{cx}} \nu$, if

$$\int \phi \, \mathrm{d}\mu \le \int \phi \, \mathrm{d}\nu \tag{1.1}$$

for all convex $\phi : \mathbb{R}^d \to \mathbb{R}_+$. We say that μ is less than ν in the *increasing convex order*, denoted $\mu \le_{\mathrm{icx}} \nu$, if (1.1) holds for all convex $\phi : \mathbb{R}^d \to \mathbb{R}_+$ which are increasing with respect to the usual coordinate-wise partial order $x \le y$.

The following type of characterisation of convex orders in terms of martingale couplings will be of our main interest. We denote by $\mathcal{M}_n(\mathbb{R}^d)$ (resp. $\mathcal{M}_n^*(\mathbb{R}^d)$) the set of probability measures λ on $(\mathbb{R}^d)^n$ such that λ is the joint distribution of some $\mathbb{R}^d$-valued martingale (resp. submartingale) (X_t) parametrised by $t \in \{1, \dots, n\}$. Recall that a *coupling* of probability measures $\mu_1, \dots, \mu_n$ on $\mathbb{R}^d$ is a probability measure on $(\mathbb{R}^d)^n$ having $\mu_1, \dots, \mu_n$ as its marginal distributions.

Theorem 1.1 (Strassen [31]). *For any probability measures μ and ν on $\mathbb{R}^d$ with finite first moments*:

(i) *$\mu \le_{\mathrm{cx}} \nu$ if and only if μ and ν admit a coupling $\lambda \in \mathcal{M}_2(\mathbb{R}^d)$*,
(ii) *$\mu \le_{\mathrm{icx}} \nu$ if and only if μ and ν admit a coupling $\lambda \in \mathcal{M}_2^*(\mathbb{R}^d)$.*

Stochastic orders are often expressed in the notation of random variables instead of probability measures. Let X and Y be random vectors on $\mathbb{R}^d$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then we denote $X \le_{\mathrm{cx}} Y$ (resp. $X \le_{\mathrm{icx}} Y$) if the corresponding probability distributions $\mathbb{P} \circ X^{-1}$ and $\mathbb{P} \circ Y^{-1}$ are ordered according to $\le_{\mathrm{cx}}$ (resp. $\le_{\mathrm{icx}}$), that is,

$$\mathbb{E}\phi(X) \le \mathbb{E}\phi(Y) \tag{1.2}$$

for all convex (resp. increasing convex) functions $\phi : \mathbb{R}^d \to \mathbb{R}_+$. Recall that a coupling of random vectors $X_1, \dots, X_n$ on $\mathbb{R}^d$ is a random vector $(\hat{X}_1, \dots, \hat{X}_d)$ defined on some probability space and taking values in $(\mathbb{R}^d)^n$ such that $\hat{X}_i \stackrel{\mathrm{d}}{=} X_i$ for all i, where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution. In this notation, Theorem 1.1 can be reformulated as follows.

Theorem 1.2. *For any real-valued random vectors X and Y with finite first moments*:

(i) *$X \le_{\mathrm{cx}} Y$ if and only if X and Y admit a coupling $(\hat{X}, \hat{Y})$ which satisfies $\hat{X} = \mathbb{E}(\hat{Y}|\hat{X})$ almost surely.*
(ii) *$X \le_{\mathrm{icx}} Y$ if and only if X and Y admit a coupling $(\hat{X}, \hat{Y})$ which satisfies $\hat{X} \le \mathbb{E}(\hat{Y}|\hat{X})$ almost surely.*

1.2. Main results

The main contribution of the present paper is the following theorem which extends the martingale characterisation in Theorem 1.1 to pairs of probability measures indexed by a parameter θ with values in some measurable space S. Recall that a *probability kernel* from S to $\mathbb{R}^d$ is a map $P : (\theta, B) \mapsto P_\theta(B)$ such that

- P_θ is a probability measure on $\mathbb{R}^d$ for every $\theta \in S$, and
- $\theta \mapsto P_\theta(B)$ is measurable for every Borel set $B \subset \mathbb{R}^d$.

We say that P has *finite first moments* if $\int |x| P_\theta(\mathrm{d}x) < \infty$ for all θ. We extend the notion of coupling to probability kernels as follows. Let P and Q be probability kernels from S to $\mathbb{R}^d$, and assume that R is a probability kernel from S to $\mathbb{R}^d \times \mathbb{R}^d$. We say that R is a *pointwise coupling* of P and Q if R_θ is a coupling of P_θ and Q_θ for every θ.

Theorem 1.3. *For any probability kernels P and Q from a measurable space S to $\mathbb{R}^d$ with finite first moments*:

(i) $P_\theta \le_{\mathrm{cx}} Q_\theta$ *for all* θ *if and only if* P *and* Q *admit a pointwise coupling* R *such that* $R_\theta \in \mathcal{M}_2(\mathbb{R}^d)$ *for all* θ,

(ii) $P_\theta \le_{\mathrm{icx}} Q_\theta$ *for all* θ *if and only if* P *and* Q *admit a pointwise coupling* R *such that* $R_\theta \in \mathcal{M}_2^*(\mathbb{R}^d)$ *for all* θ.

Conditional versions of integral stochastic orders may be defined by considering conditional analogues of (1.2). Let Z be a random element with values in a measurable space S, defined on the same probability space as random vectors X and Y on $\mathbb{R}^d$. Then we denote $X \mid Z \le_{\mathrm{cx}} Y \mid Z$ (resp. $X \mid Z \le_{\mathrm{icx}} Y \mid Z$) if

$$\mathbb{E}\big(\phi(X) \mid Z\big) \le \mathbb{E}\big(\phi(Y) \mid Z\big) \qquad \text{almost surely}$$

for all convex (resp. increasing convex) functions $\phi : \mathbb{R}^d \to \mathbb{R}$ such that $\phi(X)$ and $\phi(Y)$ are integrable. As a corollary of Theorem 1.3, we will prove the following conditional analogue of Theorem 1.2. Here a *Z-conditional coupling of X and Y* is a random element $(\hat{X}, \hat{Y}, \hat{Z})$ such that $(\hat{X}, \hat{Z}) \overset{\mathrm{d}}{=} (X, Z)$ and $(\hat{Y}, \hat{Z}) \overset{\mathrm{d}}{=} (Y, Z)$.

Theorem 1.4. *For any real-valued random vectors X and Y with finite first moments and any random element Z in a measurable space S*:

(i) $X \mid Z \le_{\mathrm{cx}} Y \mid Z$ *if and only if* X *and* Y *admit a* Z*-conditional coupling* $(\hat{X}, \hat{Y}, \hat{Z})$ *such that* $\hat{X} = \mathbb{E}(\hat{Y} \mid \hat{X}, \hat{Z})$ *almost surely.*

(ii) $X \mid Z \le_{\mathrm{icx}} Y \mid Z$ *if and only if* X *and* Y *admit a* Z*-conditional coupling* $(\hat{X}, \hat{Y}, \hat{Z})$ *such that* $\hat{X} \le \mathbb{E}(\hat{Y} \mid \hat{X}, \hat{Z})$ *almost surely.*

1.3. Related work

Theorem 1.1 extends by induction to the case where one has countably many distributions $(\mu_n)_{n\in\mathbb{N}}$ with $\mu_n \le_{\mathrm{cx}} \mu_{n+1}$ or $\mu_n \le_{\mathrm{icx}} \mu_{n+1}$. Kellerer [18] extended this to the uncountable setting, by showing that a collection of probability distributions parametrised by $t \in \mathbb{R}_+$ satisfies $\mu_s \le_{\mathrm{cx}} \mu_t$ (resp. $\mu_s \le_{\mathrm{icx}} \mu_t$) for all $s \le t$ if and only if there exists a martingale (resp. submartingale) (X_t) with X_t distributed according to μ_t for all $t \in \mathbb{R}_+$. This relation is further explored in the recent monograph [13]; see also [23]. The 'if' part of Theorem 1.1 can be proved by a simple application of Jensen's inequality, whereas the 'only if' part is more subtle. Strassen's proof [31], Theorems 8 and 9, uses the Hahn–Banach theorem. Müller and Stoyan [25], Theorem 1.5.20 and Corollary 1.5.21, provide a more constructive proof, still relying on a limiting argument. In fact, Strassen's work [31] addresses more general integral stochastic orders, defined by requiring (1.1) for a general class of functions ϕ. This allows to define orderings of random variables with values in general measurable spaces, as further investigated by Shortt [29] and Hirshberg and Shortt [14]; see also Kertz and Rösler [19]. Another direction of extending the theory of stochastic orders is to consider nontransitive relations, see Leskelä [22]. Conditional stochastic orders have been considered earlier more generally by Rüschendorf [27], following the work due to Whitt [36,37].

The main result of this article (Theorem 1.3) extends Theorem 1.1 to parametrised collections of ordered pairs of probability distributions, in contrast with ordered sequences as in [13,18]. The proof of Theorem 1.3 is based on measurability properties of related set-valued mappings and an application of a measurable selection theorem of Kuratowski and Ryll-Nardzewski [20]. We are unaware of earlier results which would be directly applicable in this context. However, similar results related to martingale couplings have appeared recently in the context of optimal transport. Beiglböck and Juillet [6] consider the problem of finding an optimal transport plan under the constraint that the transport plan is a martingale. The work of Fontbona, Guérin and Méléard [10] has the most similarities with our developments. With the notation above, they consider finding a measurable optimal transport plan between P_θ and Q_θ. The work of Hobson [16], brought to our attention by a referee, provides an explicit Skorokhod embedding of two univariate convex ordered distributions. This embedding could be used to prove our result in the scalar case.

1.4. Outline of the rest of the paper

Section 2 discusses the definitions and basic properties related to conditional convex stochastic orders. The proofs of Theorems 1.3 and 1.4 are given in Section 3 after analysing the measurability of related set-valued mappings.

Our problem was initially motivated by applied work on so-called pseudo-marginal Markov chain Monte Carlo algorithms [2]. In Section 4, we summarise the application and discuss why a martingale coupling is crucial in this context. We discuss in Section 5 some extensions of our results and their applicability in the context of martingale optimal transport.

2. Conditional convex orders

2.1. Definitions and basic properties

We denote the d-dimensional Euclidean space by $\mathbb{R}^d$, the real line by $\mathbb{R}^1 = \mathbb{R}$ and the set of positive real numbers by $\mathbb{R}_+$. We follow the convention that a number x is *positive* if $x \geq 0$ and a function f is *increasing* if $f(x) \leq f(y)$ for all $x \leq y$, with the usual coordinate-wise partial order, which holds if all the coordinates are ordered by $x_i \leq y_i$ for $1 \leq i \leq d$. Unless otherwise mentioned, all measures on a topological space will considered as measures defined on the corresponding Borel sigma-algebra. A random vector X is called *integrable* if $\mathbb{E}|X| < \infty$. When X and Y are integrable, it is not hard to verify that $X \leq_{\mathrm{cx}} Y$ (resp. $X \leq_{\mathrm{icx}} Y$) if and only if (1.2) holds for all convex (resp. increasing convex) $\phi : \mathbb{R}^d \to \mathbb{R}$ such that $\phi(X)$ and $\phi(Y)$ are integrable.

The following definition extends the Z-conditional order in Section 1 to an order conditioned on a sigma-algebra. Let X and Y be integrable random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and let $\mathcal{F} \subset \mathcal{A}$ be a sigma-algebra. We denote $X \mid \mathcal{F} \leq_{\mathrm{cx}} Y \mid \mathcal{F}$ (resp. $X \mid \mathcal{F} \leq_{\mathrm{icx}} Y \mid \mathcal{F}$) if

$$\mathbb{E}\big(\phi(X) \mid \mathcal{F}\big) \leq \mathbb{E}\big(\phi(Y) \mid \mathcal{F}\big) \qquad \text{almost surely}$$

for all convex (resp. increasing convex) functions $\phi : \mathbb{R}^d \to \mathbb{R}$ such that $\phi(X)$ and $\phi(Y)$ are integrable. When this is the case we say that X is less than Y in the *conditional convex* (resp. *increasing convex*) *order given* $\mathcal{F}$. In the special case when $\mathcal{F} = \sigma(Z)$ is generated by a random element Z with values in some measurable space, we write $X \mid Z \le_{\mathrm{cx}} Y \mid Z$ and $X \mid Z \le_{\mathrm{icx}} Y \mid Z$.

We state next a proposition which suggests that conditional convex orders can be seen as interpolations between (unconditional) convex orders and the corresponding strong stochastic orders.

Proposition 2.1. *Let X and Y be integrable random vectors defined on $(\Omega, \mathcal{A}, \mathbb{P})$ and let $\mathcal{F} \subset \mathcal{G}$ be subsigma-algebras of $\mathcal{A}$.*

(i) $X \mid \mathcal{G} \le_{\mathrm{icx}} Y \mid \mathcal{G} \Rightarrow X \mid \mathcal{F} \le_{\mathrm{icx}} Y \mid \mathcal{F} \Rightarrow X \le_{\mathrm{icx}} Y$.
(ii) $X \mid \mathcal{G} \le_{\mathrm{cx}} Y \mid \mathcal{G} \Rightarrow X \mid \mathcal{F} \le_{\mathrm{cx}} Y \mid \mathcal{F} \Rightarrow X \le_{\mathrm{cx}} Y$.
(iii) $X \mid \mathcal{A} \le_{\mathrm{icx}} Y \mid \mathcal{A} \iff X \le Y$ *almost surely.*
(iv) $X \mid \mathcal{A} \le_{\mathrm{cx}} Y \mid \mathcal{A} \iff X = Y$ *almost surely.*

Proof. For (i) assume that $X \mid \mathcal{G} \le_{\mathrm{icx}} Y \mid \mathcal{G}$, and let ϕ be an increasing convex function such that $\phi(X)$ and $\phi(Y)$ are integrable. Then by the tower property of conditional expectations,

$$\mathbb{E}\big(\phi(X) \mid \mathcal{F}\big) = \mathbb{E}\big[\mathbb{E}\big(\phi(X) \mid \mathcal{G}\big) \mid \mathcal{F}\big] \le \mathbb{E}\big[\mathbb{E}\big(\phi(Y) \mid \mathcal{G}\big) \mid \mathcal{F}\big] = \mathbb{E}\big(\phi(Y) \mid \mathcal{F}\big)$$

almost surely. Therefore $X \mid \mathcal{F} \le_{\mathrm{icx}} Y \mid \mathcal{F}$. The second implication in (i) follows by writing the above inequality for $\mathcal{F} = \{\varnothing, \Omega\}$. Part (ii) follows similarly.

Part (iii) is direct, and for (iv), notice that $X \mid \mathcal{A} \le_{\mathrm{cx}} Y \mid \mathcal{A}$ implies $X \mid \mathcal{A} \le_{\mathrm{icx}} Y \mid \mathcal{A}$ and $-X \mid \mathcal{A} \le_{\mathrm{icx}} -Y \mid \mathcal{A}$. By (iii), we conclude that $X = Y$ almost surely. The reverse implication is trivial. □

2.2. Countable characterisations

Instead of testing the expectations of all (increasing) convex functions, the following lemma states that it is enough to restrict to a countable family of such functions.

Lemma 2.2. *There exist countable sets of convex functions $\mathcal{C}$ and increasing convex functions $\mathcal{C}_+$ such that*

$$X \le_{\mathrm{cx}} Y \iff \mathbb{E}\phi(X) \le \mathbb{E}\phi(Y) \quad \text{for all } \phi \in \mathcal{C},$$
$$X \le_{\mathrm{icx}} Y \iff \mathbb{E}\phi(X) \le \mathbb{E}\phi(Y) \quad \text{for all } \phi \in \mathcal{C}_+.$$

The proof of Lemma 2.2 is given in Appendix A.

In the univariate case, Lemma 2.2 follows from the following well-known characterisations [28], Theorems 3.A.2 and 4.A.2. Here $(x)_+ := \max\{0, x\}$ denotes the positive part of a number x.

Proposition 2.3. *Let X and Y be integrable random variables. Then*

$$X \le_{\mathrm{cx}} Y \iff \mathbb{E}|X-t| \le \mathbb{E}|Y-t| \quad \text{for all } t \in \mathbb{R},$$
$$X \le_{\mathrm{icx}} Y \iff \mathbb{E}(X-t)_+ \le \mathbb{E}(Y-t)_+ \quad \text{for all } t \in \mathbb{R}.$$

Remark 2.4. It is easy to see that we may restrict to $t \in \mathbb{Q}$ in Proposition 2.3, implying that in the univariate case, we may take $\mathcal{C} = \{x \mapsto |x-t| : t \in \mathbb{Q}\}$ and $\mathcal{C}_+ = \{x \mapsto (x-t)_+ : t \in \mathbb{Q}\}$ in Lemma 2.2.

The characterisations in Proposition 2.3 are often easier to check in practice. In the insurance context, the quantity $\mathbb{E}(X-t)_+$ has an interpretation as a stop-loss [7]. Unfortunately, such simple parametrisations are not available in the multivariate case; see the discussion in [25], page 98. Both Lemma 2.2 and proposition extend naturally to the conditional case; see Lemma 2.7 and Proposition 2.8.

2.3. Characterisations using regular conditional distributions

If X is a real-valued random vector defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{F} \subset \mathcal{A}$ is a sigma-algebra, recall that a *regular conditional distribution of X given $\mathcal{F}$* is a map $(\omega, B) \mapsto P^{\mathcal{F}}_{\omega}(B)$ such that $P^{\mathcal{F}}_{\omega}$ is a probability measure on $\mathbb{R}^d$ for every ω, and $\omega \mapsto P^{\mathcal{F}}_{\omega}(B)$ is a version of $\mathbb{E}(\mathbf{1}(X \in B) \mid \mathcal{F})$ for every Borel set $B \subset \mathbb{R}^d$. Hence, $P^{\mathcal{F}}$ is a random probability measure, and the probability that $P^{\mathcal{F}}$ assigns to a Borel set B is an $\mathcal{F}$-measurable random variable with expectation $\mathbb{P}(X \in B)$. If $P^{\mathcal{F}}$ is a regular conditional distribution of a X given $\mathcal{F}$, then

$$\mathbb{E}\big(\phi(X) \mid \mathcal{F}\big) = \int \phi(x) P^{\mathcal{F}}(\mathrm{d}x) \tag{2.1}$$

almost surely for any ϕ such that $\phi(X)$ is integrable [17], Theorem 6.4.

The next result shows that conditional convex orders can be expressed equivalently by the corresponding orders of the related conditional distributions.

Proposition 2.5. *Assume that X and Y are integrable random vectors defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and let $\mathcal{F} \subset \mathcal{A}$ be a sigma-algebra. Let $P^{\mathcal{F}}$ and $Q^{\mathcal{F}}$ stand for regular conditional distributions of X and Y given $\mathcal{F}$, respectively. Then,*

$$X \mid \mathcal{F} \le_{\mathrm{cx}} Y \mid \mathcal{F} \iff P^{\mathcal{F}} \le_{\mathrm{cx}} Q^{\mathcal{F}} \quad \text{almost surely,} \tag{i}$$
$$X \mid \mathcal{F} \le_{\mathrm{icx}} Y \mid \mathcal{F} \iff P^{\mathcal{F}} \le_{\mathrm{icx}} Q^{\mathcal{F}} \quad \text{almost surely.} \tag{ii}$$

Proof. Assume first that $P^{\mathcal{F}} \le_{\mathrm{cx}} Q^{\mathcal{F}}$ almost surely. Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a convex function such that $\phi(X)$ and $\phi(Y)$ are integrable. Then by (2.1),

$$\mathbb{E}\big(\phi(Y) \mid \mathcal{F}\big) - \mathbb{E}\big(\phi(X) \mid \mathcal{F}\big) = \int \phi(y) Q^{\mathcal{F}}(\mathrm{d}y) - \int \phi(x) P^{\mathcal{F}}(\mathrm{d}x) \ge 0$$

almost surely. As a consequence, $X \mid \mathcal{F} \le_{\mathrm{cx}} Y \mid \mathcal{F}$.

To prove the converse in (i), assume that $X \mid \mathcal{F} \le_{\mathrm{cx}} Y \mid \mathcal{F}$. Let Ω_0 be the event that $P^{\mathcal{F}}$ and $Q^{\mathcal{F}}$ have finite first moments. Then $\mathbb{P}(\Omega_0) = 1$. Recall Lemma 2.2, fix a function $f \in \mathcal{C}$ and define

$$Z_f(\omega) = \int f(x) Q^{\mathcal{F}}_{\omega}(\mathrm{d}x) - \int f(x) P^{\mathcal{F}}_{\omega}(\mathrm{d}x)$$

for $\omega \in \Omega_0$, and let $Z_f(\omega) = 0$ otherwise. Then by (2.1),

$$Z_f = \mathbb{E}\big(f(Y) \mid \mathcal{F}\big) - \mathbb{E}\big(f(X) \mid \mathcal{F}\big) \ge 0$$

almost surely. This further implies that $\inf_{f \in \mathcal{C}} Z_f \ge 0$ almost surely. We conclude from Lemma 2.2 that $P^{\mathcal{F}} \le_{\mathrm{cx}} Q^{\mathcal{F}}$ almost surely.

The proof if (ii) is identical, except with functions $f \in \mathcal{C}_+$. □

Let us now consider the case where the sigma-algebra $\mathcal{F} = \sigma(Z)$ is generated by a random element Z taking values in a general measurable space S. Then for any random vector X defined on the same probability space as Z there exists [17], Thm 6.3, a probability kernel P from S to $\mathbb{R}$ such that $\omega \mapsto P_{Z(\omega)}(B)$ is a version of $\mathbb{E}(\mathbf{1}(X \in B)|Z)$ for every Borel set $B \subset \mathbb{R}^d$. Such P is called a *regular conditional distribution of X given Z*, and we note that $(\omega, B) \mapsto P_{Z(\omega)}(B)$ is a regular conditional distribution of X given $\sigma(Z)$ in the sense defined in the beginning of the section. In this case, the conditional convex and increasing convex orders can be characterised as follows.

Proposition 2.6. *Let X and Y be integrable random vectors and Z a random element in a measurable space S, all defined on a common probability space. If P and Q are regular conditional distributions of X and Y given Z, then*

$$X \mid Z \le_{\mathrm{cx}} Y \mid Z \quad \Longleftrightarrow \quad P_\theta \le_{\mathrm{cx}} Q_\theta \quad \textit{for } \mu\textit{-almost every } \theta \in S, \tag{i}$$

$$X \mid Z \le_{\mathrm{icx}} Y \mid Z \quad \Longleftrightarrow \quad P_\theta \le_{\mathrm{icx}} Q_\theta \quad \textit{for } \mu\textit{-almost every } \theta \in S, \tag{ii}$$

where μ stands for the distribution of Z.

Proof. Let $\mathcal{F} = \sigma(Z)$ and denote $P^{\mathcal{F}}_{\omega}(B) = P_{Z(\omega)}(B)$ and $Q^{\mathcal{F}}_{\omega}(B) = Q_{Z(\omega)}(B)$ for $\omega \in \Omega$ and Borel sets $B \subset \mathbb{R}$. Then $P^{\mathcal{F}}$ and $Q^{\mathcal{F}}$ are regular conditional distributions of X and Y given $\mathcal{F}$, respectively. Let $S_0 = \{\theta \in S : P_\theta \le_{\mathrm{cx}} Q_\theta\}$. The argument used in the proof of Proposition 2.5 shows that

$$S_0 = \bigcap_{f \in \mathcal{C}} \left\{\theta \in S : \int f(x) P_\theta(\mathrm{d}x) \le \int f(y) Q_\theta(\mathrm{d}y)\right\},$$

from which we conclude that S_0 is a measurable subset of S. Proposition 2.5 now tells us that $X \mid Z \le_{\mathrm{cx}} Y \mid Z$ if and only if $P_{Z(\omega)} \le_{\mathrm{cx}} Q_{Z(\omega)}$ for $\mathbb{P}$-almost every ω. The latter condition is equivalent to requiring that $\mu(S_0) = \mathbb{P}(Z \in S_0) = 1$. Hence, we have proved claim (i). The proof of claim (ii) is analogous. □

As another corollary of Proposition 2.5, we obtain the following conditional version of Lemma 2.2.

Lemma 2.7. *Let X and Y be integrable random vectors defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and let $\mathcal{F} \subset \mathcal{A}$ be a sigma-algebra. Then, there exist countable sets of convex functions $\mathcal{C}$ and $\mathcal{C}_+$ such that*

$$X \mid \mathcal{F} \leq_{\mathrm{cx}} Y \mid \mathcal{F} \quad \Longleftrightarrow \quad \mathbb{E}\big[f(X)|\mathcal{F}\big] \leq \mathbb{E}\big[f(Y)|\mathcal{F}\big], \qquad f \in \mathcal{C}, \tag{i}$$

$$X \mid \mathcal{F} \leq_{\mathrm{icx}} Y \mid \mathcal{F} \quad \Longleftrightarrow \quad \mathbb{E}\big[f(X)|\mathcal{F}\big] \leq \mathbb{E}\big[f(Y)|\mathcal{F}\big], \qquad f \in \mathcal{C}_+, \tag{ii}$$

where the inequalities on the right hold almost surely for any $f \in \mathcal{C}$ or $f \in \mathcal{C}_+$.

Proof. The forward directions of both claims follow trivially, as $f \in \mathcal{C}$ are convex and $f \in \mathcal{C}_+$ are increasing convex functions.

For the opposite direction, assume that the inequality on the right of (i) holds for all $f \in \mathcal{C}$ almost surely. Let $P^{\mathcal{F}}$ and $Q^{\mathcal{F}}$ be regular conditional distributions of X and Y given $\mathcal{F}$, respectively. Then

$$\int f(x) P^{\mathcal{F}}(\mathrm{d}x) \leq \int f(y) Q^{\mathcal{F}}(\mathrm{d}y) \tag{2.2}$$

almost surely for all $f \in \mathcal{C}$. Let Ω_0 be the event that (2.2) holds for all $f \in \mathcal{C}$, then $\mathbb{P}(\Omega_0) = 1$. Lemma 2.2 hence implies that $P^{\mathcal{F}}_\omega \leq_{\mathrm{cx}} Q^{\mathcal{F}}_\omega$ for all $\omega \in \Omega_0$, and Proposition 2.5 shows that $X \mid \mathcal{F} \leq_{\mathrm{cx}} Y \mid \mathcal{F}$. The opposite direction of claim (ii) is proved in a similar way. □

We also state the conditional version of Proposition 2.3, which follows from Lemma 2.7 as suggested in Remark 2.4.

Proposition 2.8. *Let X and Y be integrable random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and let $\mathcal{F} \subset \mathcal{A}$ be a sigma-algebra. Then,*

$$X \mid \mathcal{F} \leq_{\mathrm{cx}} Y \mid \mathcal{F} \quad \Longleftrightarrow \quad \mathbb{E}\big[|X-t||\mathcal{F}\big] \leq \mathbb{E}\big[|Y-t||\mathcal{F}\big],$$

$$X \mid \mathcal{F} \leq_{\mathrm{icx}} Y \mid \mathcal{F} \quad \Longleftrightarrow \quad \mathbb{E}\big[(X-t)_+|\mathcal{F}\big] \leq \mathbb{E}\big[(Y-t)_+|\mathcal{F}\big],$$

where the inequalities on the right hold almost surely for any $t \in \mathbb{R}$.

3. Proofs of the main results

This section is devoted to proving Theorems 1.3 and 1.4. Our proof of Theorem 1.3 is based on a measurable selection theorem of Kuratowski and Ryll-Nardzewski [20]. To apply it, we first need to analyse the regularity of coupling constructions and probability kernels with respect to suitable measurable structures on spaces of probability measures. Because convex orders are essentially restricted to probability measures with finite first moments, our natural choice is to consider Borel sigma-algebras generated by the Wasserstein metric which will be discussed in

Section 3.1. A similar measurability analysis for the topology corresponding to convergence in distribution has been carried out in [22]. The space of martingale distributions with respect to the Wasserstein metric is analysed in Section 3.2, whereas Section 3.3 establishes crucial measurability properties of probability kernels and marginalising maps. Section 3.4 concludes the proof of Theorem 1.3 and Section 3.5 concludes the proof of Theorem 1.4.

3.1. Wasserstein metric

For a probability measure μ on S and a measurable function $f : S \to S'$, we denote by $f_\#\mu = \mu \circ f^{-1}$ the pushforward measure of μ by f. When $S = S_1 \times \cdots \times S_d$, we denote the ith coordinate projection by $\pi^i(x_1, \dots, x_d) := x_i$. Then $\pi^i_\#\mu$ equals the ith marginal distribution of μ. The set of couplings of $\mu \in \mathcal{P}(S_1)$ and $\nu \in \mathcal{P}(S_2)$ will be denoted by

$$\Gamma(\mu, \nu) := \big\{\lambda \in \mathcal{P}(S_1 \times S_2) : \pi^1_\#\lambda = \mu, \pi^2_\#\lambda = \nu\big\}.$$

Let us recall the definition of the Wasserstein (a.k.a. Kantorovich–Rubinstein) metric between two probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$:

$$d_W(\mu, \nu) := \min_{\lambda \in \Gamma(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y| \lambda(\mathrm{d}x \times \mathrm{d}y).$$

The minimum is attained by lower semicontinuity properties and the relative compactness of $\Gamma(\mu, \nu)$, and the map d_W is a metric on $\mathcal{P}_1(\mathbb{R}^d)$ [1], Section 7.1.

The space $\mathcal{P}_1(\mathbb{R}^d)$ equipped with the Wasserstein metric is a complete separable metric space [1], Proposition 7.1.5. The same proposition also shows that $d_W(\mu_n, \mu) \to 0$ if and only if $\mu_n \to \mu$ in distribution and (μ_n) is uniformly integrable in the sense that

$$\sup_n \int_{\mathbb{R}^d} |x| \mathbf{1}\big(|x| > t\big) \mu_n(\mathrm{d}x) \to 0 \qquad \text{as } t \to \infty.$$

Hereafter, we equip $\mathcal{P}_1(\mathbb{R}^d)$ by the topology induced by d_W.

The following results are probably well-known in transport theory, but we were unable to find them in the literature. We provide proofs in Appendix A for the reader's convenience.

Lemma 3.1. *The ith marginal map $\pi^i_\# : \mathcal{P}_1((\mathbb{R}^d)^n) \to \mathcal{P}_1(\mathbb{R}^d)$ is continuous for all i.*

Lemma 3.2. *For any $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$, the set of couplings $\Gamma(\mu, \nu)$ is compact in $\mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$.*

3.2. Two-parameter martingales and submartingales

Recall that $\mathcal{M}_2(\mathbb{R}^d)$ (resp. $\mathcal{M}_2^*(\mathbb{R}^d)$) denotes the collection of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ which are joint distributions of a two-parameter martingale (resp. submartingale). The following elementary lemmas stated without a proof give convenient ways to characterise these collections.

Lemma 3.3. *The following are equivalent for any* $\lambda \in \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$:

(i) $\lambda \in \mathcal{M}_2(\mathbb{R}^d)$.
(ii) $\mathbb{E}[Y \mid X] = X$ *a.s. for any random vector* (X, Y) *with distribution* λ.
(iii) $\int y\mathbf{1}(x \in A)\lambda(\mathrm{d}x \times \mathrm{d}y) = \int x\mathbf{1}(x \in A)\lambda(\mathrm{d}x \times \mathrm{d}y)$ *for all Borel sets* $A \subset \mathbb{R}^d$.
(iv) $\int y\phi(x)\lambda(\mathrm{d}x \times \mathrm{d}y) = \int x\phi(x)\lambda(\mathrm{d}x \times \mathrm{d}y)$ *for all continuous bounded* $\phi : \mathbb{R}^d \to \mathbb{R}_+$.

Lemma 3.4. *The following are equivalent for any* $\lambda \in \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$:

(i) $\lambda \in \mathcal{M}_2^*(\mathbb{R}^d)$.
(ii) $\mathbb{E}[Y \mid X] \geq X$ *a.s. for any random vector* (X, Y) *with distribution* λ.
(iii) $\int y\mathbf{1}(x \in A)\lambda(\mathrm{d}x \times \mathrm{d}y) \geq \int x\mathbf{1}(x \in A)\lambda(\mathrm{d}x \times \mathrm{d}y)$ *for all Borel sets* $A \subset \mathbb{R}^d$.
(iv) $\int y\phi(x)\lambda(\mathrm{d}x \times \mathrm{d}y) \geq \int x\phi(x)\lambda(\mathrm{d}x \times \mathrm{d}y)$ *for all continuous bounded* $\phi : \mathbb{R}^d \to \mathbb{R}_+$.

The following lemma shows that martingale and submartingale measures form closed sets with respect to the Wasserstein metric.

Lemma 3.5. *The sets* $\mathcal{M}_2(\mathbb{R}^d)$ *and* $\mathcal{M}_2^*(\mathbb{R}^d)$ *are closed in* $\mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$.

Proof. Assume that $\mu_n \in \mathcal{M}_2^*(\mathbb{R}^d)$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$ such that $d_W(\mu_n, \mu) \to 0$. Then $\mu_n \to \mu$ in distribution and (μ_n) is uniformly integrable. Let $\phi : \mathbb{R}^d \to \mathbb{R}_+$ be continuous and bounded. By Lemma 3.4, it is sufficient to verify that

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (x_2 - x_1)\phi(x_1)\mu(\mathrm{d}x) \geq 0. \tag{3.1}$$

To do this, let $g(x) := (x_2 - x_1)\phi(x_1)$, fix $t > 0$ and choose a continuous function $k_t : \mathbb{R}^d \times \mathbb{R}^d \to [0, 1]$ such that $k_t(x) = 1$ for $|x| \leq t$ and $k_t(x) = 0$ for $|x| > t + 1$. Let us write $g = g_t^0 + g_t^1$ where $g_t^0(x) := g(x)k_t(x)$ and $g_t^1(x) := g(x)(1 - k_t(x))$. Then

$$\mu_n(g) - \mu(g) = \bigl(\mu_n\bigl(g_t^0\bigr) - \mu\bigl(g_t^0\bigr)\bigr) + \bigl(\mu_n\bigl(g_t^1\bigr) - \mu\bigl(g_t^1\bigr)\bigr).$$

Now g_t^0 is continuous and bounded, so that $\mu_n(g_t^0) \to \mu(g_t^0)$ by convergence in distribution. Moreover, $|g_t^1(x)| \leq 2|x|\|\phi\|_\infty \mathbf{1}(|x| > t)$. This bound together with uniform integrability shows that $\sup_n(\mu_n(g_t^1) - \mu(g_t^1)) \to 0$ as $t \to \infty$. We can make the last two terms on the right-hand side above arbitrarily close to zero by choosing t large enough, uniformly in n. Then by letting $n \to \infty$, we may conclude that $\mu_n(g) \to \mu(g)$ as $n \to \infty$. The submartingale property implies by Lemma 3.4 that $\mu_n(g) \geq 0$ for all n, so we conclude that $\mu(g) \geq 0$ and therefore (3.1) is valid.

The proof that $\mathcal{M}_2(\mathbb{R}^d)$ is closed is identical, with equality in (3.1). □

3.3. Measurability of the coupling map

In what follows, we consider set-valued mappings (a.k.a. multifunctions [30]) from a measurable space $(S, \mathcal{S})$ to the topological space $\mathcal{P}_1((\mathbb{R}^d)^n)$ equipped with the Wasserstein metric. A set-valued mapping G maps a point $\theta \in S$ to a set $G(\theta) \subset \mathcal{P}_1((\mathbb{R}^d)^n)$. The *set-valued inverse* of such a mapping G is defined by

$$G^-(A) := \{\theta \in S : G(\theta) \cap A \neq \varnothing\}, \qquad A \subset \mathcal{P}_1\big((\mathbb{R}^d)^n\big).$$

The set-valued map G is called *measurable* if $G^-(A) \in \mathcal{S}$ for all closed $A \subset \mathcal{P}_1((\mathbb{R}^d)^n)$. By expressing an open set $U \subset \mathcal{P}_1((\mathbb{R}^d)^n)$ as a countable union of closed balls, we see that the measurability of G implies that $G^-(U) \in \mathcal{S}$ also for open sets U.

Proposition 3.6. *Let P and Q be probability kernels from S to $\mathbb{R}^d$ with finite first moments. Then*

$$F(\theta) := \Gamma(P_\theta, Q_\theta)$$

is measurable as a set-valued mapping from S to $\mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$.

The proof of Proposition 3.6 is based on the three auxiliary lemmas which will be stated and proved next.

Lemma 3.7. *Let P be probability kernel from S to $\mathbb{R}^d$ with finite first moments. Then $\theta \mapsto P_\theta$ is a measurable map from S to $\mathcal{P}_1(\mathbb{R}^d)$.*

Proof. Let us first verify that $\theta \mapsto P_\theta f$ is measurable for every Borel function $f : \mathbb{R}^d \to \mathbb{R}$ such that $\int |f(y)| P_\theta(dy) < \infty$ for all $\theta \in S$. Choose a sequence of simple Borel functions such that $f_n \to f$ and $|f_n| \le |f|$ pointwise. By linearity, $\theta \mapsto P_\theta f_n$ is measurable for any n. By dominated convergence,

$$P_\theta f = \lim_{n\to\infty} P_\theta f_n$$

by which $\theta \mapsto P_\theta f$ is measurable as a pointwise limit of measurable functions.

Let then $B_\varepsilon(\mu)$ denote the closed d_W-ball with radius $\varepsilon > 0$ and centre $\mu \in \mathcal{P}_1(\mathbb{R}^d)$. We will next show that the preimages $A_{\varepsilon,\mu} := \{\theta : P_\theta \in B_\varepsilon(\mu)\}$ of closed balls are measurable. By Lemma B.1 in Appendix B, there exists a countable set $\mathcal{T}_d$ of 1-Lipschitz functions on $\mathbb{R}^d$ such that

$$A_{\varepsilon,\mu} = \Big\{\theta : \sup_{g\in\mathcal{T}_d} \big[P_\theta(g) - \mu(g)\big] \le \varepsilon\Big\} = \bigcap_{g\in\mathcal{T}_d} \big\{\theta : \big[P_\theta(g) - \mu(g)\big] \le \varepsilon\big\}.$$

Therefore, $A_{\varepsilon,\mu}$ is measurable as a countable intersection of measurable sets.

Let then U be an open set in $\mathcal{P}_1(\mathbb{R}^d)$. Because $\mathcal{P}_1(\mathbb{R}^d)$ is a separable metric space, U may be expressed as a countable union of d_W-balls $B_1, B_2, \dots,$ and therefore

$$\{\theta : P_\theta \in U\} = \bigcup_{i=1}^\infty \{\theta : P_\theta \in B_i\}$$

is measurable. This implies the claim. □

We next consider the marginaliser map $\Pi : \mathcal{P}((\mathbb{R}^d)^n) \to \mathcal{P}(\mathbb{R}^d)^n$ defined by

$$\Pi(\mu) = \left(\pi^1_{\#}\mu, \dots, \pi^n_{\#}\mu\right).$$

It takes a probability measure on $(\mathbb{R}^d)^n$ as its input and returns its marginal distributions on $\mathbb{R}^d$. If the input of Π has a finite first moment, then so do its marginal distributions. Therefore, we may also consider Π as a mapping from $\mathcal{P}_1((\mathbb{R}^d)^n)$ onto $\mathcal{P}_1(\mathbb{R}^d)^n$.

Lemma 3.8. *Let S and S' be Polish spaces and $f : S \to S'$ a Borel map such that $f^{-1}(y)$ is compact for all $y \in S'$. Then f maps closed sets into Borel sets.*

Proof. By [30], Propositions 3.1.21 and 3.1.23, the graph of f

$$\operatorname{graph}(f) := \left\{\left(x, f(x)\right) : x \in S\right\}$$

is a Borel set in $S \times S'$. For any closed set $A \subset S$, the image $f(A)$ can be represented as a projection of the set

$$B := \operatorname{graph}(f) \cap \left(A \times S'\right).$$

Observe next that for any $y \in S'$ the section

$$\left\{x : (x, y) \in B\right\} = f^{-1}(y) \cap A$$

is compact. Therefore, Novikov's theorem [30], Theorem 4.7.11, implies that $f(A)$ is Borel. □

Lemma 3.9. *The marginaliser map $\Pi : \mathcal{P}_1((\mathbb{R}^d)^n) \to \mathcal{P}_1(\mathbb{R}^d)^n$ defined by*

$$\Pi(\mu) = \left(\pi^1_{\#}\mu, \dots, \pi^n_{\#}\mu\right)$$

maps closed sets into Borel sets.

Proof. Π is continuous by Lemma 3.1, and hence also Borel. The spaces $\mathcal{P}_1((\mathbb{R}^d)^n)$ and $\mathcal{P}_1(\mathbb{R}^d)^n$ are Polish. The preimage of Π for any singleton is compact by Lemma 3.2, because $\Pi^{-1}(\{\nu_1, \dots, \nu_n\}) = \Gamma(\nu_1, \dots, \nu_n)$. The rest follows from Lemma 3.8. □

Proof of Proposition 3.6. We write the set of couplings of P_θ and Q_θ again as a preimage of the marginaliser,

$$F(\theta) = \Pi^{-1}\left(\left\{(P_\theta, Q_\theta)\right\}\right).$$

Note that $F(\theta) \cap A \neq \varnothing$ if and only if $\mu \in F(\theta)$ for some $\mu \in A$, that is, $\Pi(\mu) = (P_\theta, Q_\theta)$ for some $\mu \in A$. Therefore, the set-valued inverse of F may be written as

$$F^-(A) = \left\{\theta \in S : F(\theta) \cap A \neq \varnothing\right\} = \left\{\theta \in S : (P_\theta, Q_\theta) \in \Pi(A)\right\}.$$

By Lemma 3.9, $\Pi(A)$ is a Borel set in $\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$ whenever $A \subset \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$ is closed. By Lemma 3.7, the maps $\theta \mapsto P_\theta$ and $\theta \mapsto Q_\theta$ are measurable from S to $\mathcal{P}_1(\mathbb{R}^d)$. Thus also the map $\theta \mapsto (P_\theta, Q_\theta)$ is measurable from S to $\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$. We may hence conclude that $F^-(A)$ is a measurable subset of S for any closed $A \subset \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$. □

3.4. Proof of Theorem 1.3

Assume that $P_\theta \le_{\mathrm{cx}} Q_\theta$ for all $\theta \in S$. Consider the set-valued mapping $G(\theta) := F(\theta) \cap \mathcal{M}$, where $F(\theta) = \Gamma(P_\theta, Q_\theta)$ is the set of couplings of P_θ and Q_θ, and $\mathcal{M} := \mathcal{M}_2(\mathbb{R}^d)$ is the collection of joint distributions of two-parameter martingales. Proposition 3.6 shows that F is a measurable set-valued mapping from S to the subsets of $\mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$. For any $A \subset \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$, the set-valued inverse of G can be written as

$$G^-(A) = F^-(\mathcal{M} \cap A).$$

Because $\mathcal{M}$ is closed by Lemma 3.5, we see that G is a measurable set-valued mapping from S to the subsets of $\mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$. Furthermore, because $F(\theta)$ is compact for all θ by Lemma 3.2, also $G(\theta)$ is compact for all θ. Hence, G is a measurable compact-valued mapping from S to the subsets of $\mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$. Strassen's coupling characterisation (Theorem 1.1) implies that $G(\theta)$ is nonempty for all θ. A measurable selection theorem of Kuratowski and Ryll-Nardzewski [20] (see alternatively [30], Theorem 5.2.1) now implies that there exists a measurable selection for G, that is, a measurable function $g : S \to \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$ such that $g(\theta) \in G(\theta)$ for all θ. Let us now define a map $(\theta, B) \mapsto R_\theta(B)$ by setting

$$R_\theta(B) := \mathrm{ev}_B\big(g(\theta)\big)$$

for $\theta \in S$ and Borel sets $B \subset \mathbb{R}^d \times \mathbb{R}^d$, where $\mathrm{ev}_B(\mu) = \mu(B)$. Then $R_\theta \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$ is a coupling of P_θ and Q_θ for every $\theta \in S$. We are left with showing that $\theta \mapsto R_\theta(B)$ is measurable for any Borel set $B \subset \mathbb{R}^d \times \mathbb{R}^d$. This follows because the map $\mathrm{ev}_B : \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d) \to \mathbb{R}$ is measurable by Lemma C.1 in Appendix C. Hence, R is a pointwise coupling of the probability kernels P and Q.

If $P_\theta \le_{\mathrm{icx}} Q_\theta$ for all $\theta \in S$, then by repeating the above construction with $\mathcal{M}$ replaced by $\mathcal{M}^* := \mathcal{M}_2^*(\mathbb{R}^d)$ we obtain a probability kernel R which is a pointwise coupling of P and Q such that $R_\theta \in \mathcal{M}^*$ for all $\theta \in S$.

Finally, we note that if R is pointwise coupling of P and Q such that $R_\theta \in \mathcal{M}_2(\mathbb{R}^d)$ (resp. $\mathcal{M}_2^*(\mathbb{R}^d)$) for all θ, then Theorem 1.1 immediately implies that $P_\theta \le_{\mathrm{cx}} Q_\theta$ (resp. $P_\theta \le_{\mathrm{icx}} Q_\theta$) for all θ. □

3.5. Proof of Theorem 1.4

Let us first prove the forward implication in (ii). Suppose that $X \mid Z \le_{\mathrm{icx}} Y \mid Z$. Let P and Q be regular conditional distributions of X and Y given Z, respectively, and denote the distribution of Z by μ. Then by Proposition 2.6, $P_\theta \le_{\mathrm{icx}} Q_\theta$ for all $\theta \in S$ outside a set of μ-measure zero. By redefining P_θ and Q_θ as equal on this set of μ-measure zero, we may assume that $P_\theta \le_{\mathrm{icx}} Q_\theta$ for all $\theta \in S$. By Theorem 1.3, there exists a probability kernel R from S to $\mathbb{R}^d \times \mathbb{R}^d$ which is a pointwise coupling of P and Q and satisfies $R_\theta \in \mathcal{M}_2^*(\mathbb{R}^d)$ for all θ.

Let $(\hat{X}, \hat{Y}, \hat{Z})$ be a random element in $\mathbb{R}^d \times \mathbb{R}^d \times S$ with distribution

$$\lambda(\mathrm{d}x \times \mathrm{d}y \times \mathrm{d}\theta) := \mu(\mathrm{d}\theta) R_\theta(\mathrm{d}x \times \mathrm{d}y).$$

Because $\lambda(\mathrm{d}x \times \mathbb{R}^d \times \mathrm{d}\theta) = \mu(\mathrm{d}\theta) P_\theta(\mathrm{d}x)$ and $\lambda(\mathbb{R}^d \times \mathrm{d}y \times \mathrm{d}\theta) = \mu(\mathrm{d}\theta) Q_\theta(\mathrm{d}y)$, it follows that $(\hat{X}, \hat{Z}) \stackrel{\mathrm{d}}{=} (X, Z)$ and $(\hat{Y}, \hat{Z}) \stackrel{\mathrm{d}}{=} (Y, Z)$. Hence, $(\hat{X}, \hat{Y}, \hat{Z})$ is a Z-conditional coupling of X and Y. We still need to verify that

$$\hat{X} \le \mathbb{E}(\hat{Y} \mid \hat{X}, \hat{Z}) \qquad \text{almost surely.} \tag{3.2}$$

For any measurable $A \subset \mathbb{R}^d \times S$, by denoting $A_\theta := \{x \in \mathbb{R}^d : (x, \theta) \in A\}$, we see with the help of Lemma 3.4 that

$$\begin{aligned} \mathbb{E}\big[\mathbb{E}[\hat{Y} \mid \hat{X}, \hat{Z}]\mathbf{1}\big((\hat{X}, \hat{Z}) \in A\big)\big] &= \mathbb{E}\big[\hat{Y}\mathbf{1}\big((\hat{X}, \hat{Z}) \in A\big)\big] \\ &= \int \mu(\mathrm{d}\theta) \int y \mathbf{1}(x \in A_\theta) R_\theta(\mathrm{d}x \times \mathrm{d}y) \\ &\ge \int \mu(\mathrm{d}\theta) \int x \mathbf{1}(x \in A_\theta) R_\theta(\mathrm{d}x \times \mathrm{d}y) \\ &= \mathbb{E}\big[\hat{X}\mathbf{1}\big((\hat{X}, \hat{Z}) \in A\big)\big], \end{aligned} \tag{3.3}$$

because $R_\theta \in \mathcal{M}_2^*(\mathbb{R}^d)$ for all θ. This implies (3.2).

To prove the other direction in (ii), assume next that $(\hat{X}, \hat{Y}, \hat{Z})$ is a Z-conditional coupling of X and Y satisfying (3.2). Recall Lemma 2.7 and let $f \in \mathcal{C}_+$, then conditional Jensen's inequality implies that

$$f(\hat{X}) \le f\big(\mathbb{E}(\hat{Y} \mid \hat{X}, \hat{Z})\big) \le \mathbb{E}\big(f(\hat{Y}) | \hat{X}, \hat{Z}\big)$$

almost surely. By taking $\hat{Z}$-conditional expectations on both sides above, it follows that

$$\mathbb{E}\big(f(\hat{X}) | \hat{Z}\big) \le \mathbb{E}\big(f(\hat{Y}) | \hat{Z}\big).$$

Because $(\hat{X}, \hat{Z}) \stackrel{\mathrm{d}}{=} (X, Z)$ and $(\hat{Y}, \hat{Z}) \stackrel{\mathrm{d}}{=} (Y, Z)$, we may remove the hats above to conclude that

$$\mathbb{E}\big(f(X) | Z\big) \le \mathbb{E}\big(f(Y) | Z\big) \tag{3.4}$$

almost surely. By Lemma 2.7, this implies $X \mid Z \le_{\mathrm{icx}} Y \mid Z$.

The proof of the forward implication of claim (i) is obtained by imitating the proof of (ii); by replacing the inequality in (3.2) and (3.3) by equality, and applying Lemma 3.3 in place of Lemma 3.4. Similarly, the reverse implication of claim (i) is obtained by using $f \in \mathcal{C}$ in place of $f \in \mathcal{C}_+$ in (3.4). □

4. Application to pseudo-marginal Markov chain Monte Carlo

We discuss here briefly the application which initially motivated the present work. The application focuses on Markov chain Monte Carlo (MCMC) algorithms targeting a probability distribution π on a general state space X. In particular, the interest lies in the so-called pseudo-marginal

MCMC with transition probability

$$K(x, w; \mathrm{d}y \times \mathrm{d}u) := q(x, \mathrm{d}y) Q_y(\mathrm{d}u) \min\left\{1, r(x, y)\frac{u}{w}\right\} + \mathbf{1}_{\mathrm{d}y \times \mathrm{d}u}(x, t)\rho(x, w),$$

parametrised by a proposal kernel $(x, B) \mapsto q(x, B)$ on X and an auxiliary kernel $(x, B) \mapsto Q_x(B)$ from X to $\mathbb{R}_+$, satisfying $\int Q_x(\mathrm{d}w)w = 1$ for every $x \in \mathsf{X}$. The function $r(x, y)$ is the Radon–Nikodym derivative $\frac{\pi(\mathrm{d}y)}{\pi(\mathrm{d}x)} \frac{q(y,\mathrm{d}x)}{q(x,\mathrm{d}y)}$ whenever well-defined, and zero otherwise cf. [34], and the 'probability of rejection' $\rho(x, w) \in [0, 1]$ is such that K defines a transition probability. We advise an interested reader to consult [4] for details and [2,3] and references therein for more thorough introduction to the method.

It is not difficult to check that K is reversible with respect to the distribution

$$\tilde{\pi}(\mathrm{d}x \times \mathrm{d}w) = \pi(\mathrm{d}x) Q_x(\mathrm{d}w) w,$$

and it is evident that $\tilde{\pi}$ admits π as its first marginal. This means that, if the Markov chain $(X_k, W_k)_{k \geq 1}$ with transition probability K is irreducible, the ergodic averages approximate the integral of any π-integrable function $f : \mathsf{X} \to \mathbb{R}$:

$$\frac{1}{n} \sum_{k=1}^{n} f(X_k) \stackrel{n \to \infty}{\longrightarrow} \pi(f) := \int_{\mathsf{X}} f(x) \pi(\mathrm{d}x).$$

The so-called *asymptotic variance* is a common MCMC efficiency criterion, which is informative about the asymptotic rate of convergence above. It is defined in the present setting for any $f \in L^2(\pi) := \{f : \mathsf{X} \to \mathbb{R} : \pi(f^2) < \infty\}$ through

$$\sigma^2(K, f) := \lim_{n \to \infty} \mathbb{E}\left[\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left[f\left(X'_k\right) - \pi(f)\right]\right]^2, \tag{4.1}$$

where $(X'_k, W'_k)_{k \geq 1}$ is a stationary version of the MCMC chain – that is, $(X'_1, W'_1) \sim \tilde{\pi}$ and $(X'_k, W'_k)_{k \geq 1}$ follows the transition probability K. The limit in (4.1) always exists, but can be infinite [34]. In the pseudo-marginal context, we are interested in how the choice of the laws $\{Q_x\}_{x \in \mathsf{X}}$ affects the asymptotic variance.

The usual method to compare asymptotic variances of reversible Markov chains is Peskun's theorem [26] and its generalisations [8,21,34]. It states that if two Markov transition probabilities K and K' are reversible with respect to the same probability distribution μ, then

$$\sigma^2(K, f) \leq \sigma^2\left(K', f\right) \qquad \text{for all } f \in L^2(\mu),$$

if and only if

$$\langle g, Kg \rangle_\mu \leq \left\langle g, K'g \right\rangle_\mu \qquad \text{for all } g \in L^2(\mu),$$

where $\langle f, g \rangle_\mu := \int f(x) g(x) \mu(\mathrm{d}x)$. This is inapplicable in the present application, as the two Markov chains K and K' with $\{Q_x\}_{x \in \mathsf{X}}$ and $\{Q'_x\}_{x \in \mathsf{X}}$ are reversible with respect to different invariant distributions $\tilde{\pi}$ and $\tilde{\pi}'$, respectively.

Because the interest lies only in functions which are constant in the second coordinate, it is still possible to pursue such an ordering. Indeed, given a pointwise martingale coupling R_x of Q_x and Q'_x, which exists by Theorem 1.3 if $Q_x \le_{cx} Q'_x$ for all $x \in \mathsf{X}$, it turns out to be possible to deduce a 'Peskun-type' order of the asymptotic variances [4], Theorem 10,

$$\sigma^2(K, f) \le \sigma^2\big(K', f\big) \qquad \text{for all } f(x, w) = f(x) \in L^2(\pi). \tag{4.2}$$

We will next briefly summarise why a strong martingale coupling as in Theorem 1.3 is fundamental to prove this result.

The key of the proof of (4.2) relies in 'embedding' the two Markov kernels K and K' on a common Hilbert space. The martingale coupling allows to construct the following Markov kernels $\breve{K}$ and $\breve{K}'$ and a distribution $\breve{\pi}$:

$$\begin{aligned}
\breve{K}(x, w, v; \mathrm{d}y \times \mathrm{d}u \times \mathrm{d}t) &= q(x, \mathrm{d}y) R_y(\mathrm{d}u \times \mathrm{d}t) \frac{t}{u} \min\left\{1, r(x, y)\frac{u}{w}\right\} \\
&\quad + \mathbf{1}_{\mathrm{d}y \times \mathrm{d}u \times \mathrm{d}t}(x, w, v)\rho(x, w), \\
\breve{K}'(x, w, v; \mathrm{d}y \times \mathrm{d}u \times \mathrm{d}t) &= q(x, \mathrm{d}y) R_y(\mathrm{d}u \times \mathrm{d}t) \min\left\{1, r(x, y)\frac{t}{v}\right\} \\
&\quad + \mathbf{1}_{\mathrm{d}y \times \mathrm{d}u \times \mathrm{d}t}(x, w, v)\rho'(x, v), \\
\breve{\pi}(\mathrm{d}x \times \mathrm{d}t \times \mathrm{d}u) &= \pi(\mathrm{d}x) R_x(\mathrm{d}w \times \mathrm{d}v) v.
\end{aligned}$$

It is not difficult to check that both $\breve{K}$ and $\breve{K}'$ are reversible with respect to $\breve{\pi}$, and $\breve{\pi}$ coincides marginally with $\tilde{\pi}$ and $\tilde{\pi}'$ so that $\tilde{\pi}(\mathrm{d}x \times \mathrm{d}w) = \breve{\pi}(\mathrm{d}x \times \mathrm{d}w \times \mathbb{R}_+)$ and $\tilde{\pi}'(\mathrm{d}x \times \mathrm{d}v) = \breve{\pi}(\mathrm{d}x \times \mathbb{R}_+ \times \mathrm{d}v)$. Similarly, the kernels $\breve{K}$ and $\breve{K}'$ coincide marginally with K and K'; see [4], Lemma 20. This construction enables the Hilbert space techniques, on $L^2(\breve{\pi})$, to be used. The martingale coupling allows to show that ([4], Theorem 22(b)),

$$\langle g, \breve{K} g\rangle_{\breve{\pi}} \le \big\langle g, \breve{K}' g\big\rangle_{\breve{\pi}} \qquad \text{for all } g(x, w, v) = g(x, w) \text{ with } g \in L^2(\tilde{\pi}),$$

which ultimately leads to the order $\sigma^2(K, f) \le \sigma^2(K', f)$ for all $f(x, w) = f(x)$ with $f \in L^2(\pi)$.

5. Extensions and implications

We discuss next some extensions and implications of our results. In Strassen's original paper, Theorem 1.1 is formulated for countably many distributions instead of a pair. Extension of Theorems 1.3 and 1.4 into a context with countably many kernels is straightforward. For instance, we may formulate the following result.

Proposition 5.1. *Suppose that for each $i \in \mathbb{N}$, $(\theta, B) \mapsto P_\theta^{(i)}(B)$ is a probability kernel from S to $\mathbb{R}^d$.*

(i) $P_\theta^{(i)} \le_{\mathrm{cx}} P_\theta^{(i+1)}$ *for all* $i \in \mathbb{N}$ *if and only if there exists a pointwise coupling* R *of* $\{P_\theta^{(i)}\}_{i\in\mathbb{N}}$ *such that* $R_\theta \in \mathcal{M}_{\mathbb{N}}(\mathbb{R}^d)$.

(ii) $P_\theta^{(i)} \le_{\mathrm{icx}} P_\theta^{(i+1)}$ *for all* $i \in \mathbb{N}$ *if and only if there exists a pointwise coupling* R *of* $\{P_\theta^{(i)}\}_{i\in\mathbb{N}}$ *such that* $R_\theta \in \mathcal{M}^*_{\mathbb{N}}(\mathbb{R}^d)$.

More precisely, R *above is a kernel from* S *to* $(\mathbb{R}^d)^{\mathbb{N}}$ *such that* $R_\theta(\cdot)$ *is the law of the* $\mathbb{R}^d$*-valued (sub-)martingale* $(X_\theta^{(i)})_{i\ge 1}$ *such that* $\mathbb{P}(X_\theta^{(i)} \in A) = P_\theta^{(i)}(A)$.

Proof. For, (i) assume that for each $i \in \mathbb{N}$ $P_\theta^{(i)} \le_{\mathrm{cx}} P_\theta^{(i+1)}$ and let $R_\theta^{(i)}$ stand for their pointwise coupling. There exist kernels $T^{(i)}$ from $S \times \mathbb{R}^d$ to $\mathbb{R}^d$ (regular conditional probabilities) such that

$$R_\theta^{(i)}(\mathrm{d}x_{i-1} \times \mathrm{d}x_i) = P_\theta^{(i-1)}(\mathrm{d}x_{i-1}) T^{(i)}_{\theta, x_{i-1}}(\mathrm{d}x_i).$$

We may define R_θ inductively through its finite-dimensional distributions by letting $R_\theta(\mathrm{d}x_1 \times \mathrm{d}x_2 \times \mathbb{R}^{\mathbb{N}}) = R_\theta^{(2)}(\mathrm{d}x_1 \times \mathrm{d}x_2)$ and for $i \ge 3$

$$R_\theta\big(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_i \times \mathbb{R}^{\mathbb{N}}\big) = R_\theta\big(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_{i-1} \times \mathbb{R}^{\mathbb{N}}\big) T^{(i)}_{\theta, x_{i-1}}(\mathrm{d}x_i).$$

The other direction follows from Jensen's inequality. The proof of (ii) follows similar lines. □

The following characterisation of increasing convex orders in terms of convex stochastic order and strong stochastic order [25], Theorem 3.4.3, is sometimes convenient.

Theorem 5.2. *If* $X \le_{\mathrm{icx}} Y$, *then there exist a probability space with random variables* $\hat{X}, \hat{W}, \hat{Y}$ *such that* $\hat{X} \overset{\mathrm{d}}{=} X$, $\hat{Y} \overset{\mathrm{d}}{=} Y$, $\hat{X} \le \hat{W}$ *almost surely and* $\hat{W} \le_{\mathrm{cx}} \hat{Y}$.

We record the following result, which is a conditional version Theorem 5.2.

Proposition 5.3. *If* $X \mid Z \le_{\mathrm{icx}} Y \mid Z$, *then there exist a probability space with random variables* $\hat{X}$, $\hat{Y}$, $\hat{Z}$ *and* $\hat{W}$ *such that* $(\hat{X}, \hat{Z}) \overset{\mathrm{d}}{=} (X, Z)$, $(\hat{Y}, \hat{Z}) \overset{\mathrm{d}}{=} (Y, Z)$, $\hat{X} \le \hat{W}$ *almost surely and* $\hat{W} \mid \hat{Z} \le_{\mathrm{cx}} \hat{Y} \mid \hat{Z}$.

Proof. We may take the triple $(\hat{X}, \hat{Y}, \hat{Z})$ from Theorem 1.4 and set $\hat{W} := \mathbb{E}(\hat{Y} \mid \hat{X}, \hat{Z})$. Then $\hat{X} \le \hat{W}$, and for any convex $\phi : \mathbb{R}^d \to \mathbb{R}$, Jensen's inequality yields $\mathbb{E}(\phi(\hat{W})|\hat{Z}) \le \mathbb{E}(\phi(\hat{Y})|\hat{Z})$. □

We next turn into so-called martingale optimal transport problem [6] which is linked to applications in mathematical finance, e.g., [5,9,11,15]. Optimal transport problems, in general, mean finding a coupling of two probability measures μ such that the 'cost' $\mu(c) := \iint c(x, y)\mu(\mathrm{d}x \times \mathrm{d}y)$ is minimised. Usually the minimisation is over all couplings, but in the martingale optimal transport the minimisation is constrained to martingale couplings.

We illustrate that when a parametric version of such a problem is considered, our results allow to ensure that minimisers can be chosen in a measurable manner in this context.

Proposition 5.4. *Suppose that P and Q are probability kernels from S to $\mathbb{R}^d$ and $c : \mathbb{R}^d \times \mathbb{R}^d \to (-\infty, \infty]$ is lower semi-continuous and satisfies the lower bound*

$$c(x, y) \geq -C\big(1 + |x| + |y|\big) \qquad \text{for all } x, y \in \mathbb{R}^d$$

for some finite constant C.

(i) *If $P_\theta \leq_{\mathrm{cx}} Q_\theta$ for all θ, then there exists a measurable optimal martingale transport plan γ, that is, a kernel $(\theta, B) \mapsto \gamma_\theta(B)$ from S to $\mathbb{R}^d \times \mathbb{R}^d$ such that for every θ, γ_θ is a martingale coupling of P_θ and Q_θ which minimises $\mu(c)$ over all martingale couplings μ of P_θ and Q_θ.*

(ii) *If $P_\theta, Q_\theta \in \mathcal{P}_1(\mathbb{R}^d)$, then there exists a measurable optimal transport plan γ^*, that is, a kernel $(\theta, B) \mapsto \gamma^*_\theta(B)$ from S to $\mathbb{R}^d \times \mathbb{R}^d$ such that for every θ, γ^*_θ is a coupling of P_θ and Q_θ which minimises $\mu(c)$ over all couplings μ of P_θ and Q_θ.*

Proof. Consider first (i), and denote for brevity $\Gamma(\theta) := \Gamma(P_\theta, Q_\theta) \cap \mathcal{M}_2(\mathbb{R}^d)$. Recall that $\theta \to \Gamma(\theta)$ is compact-valued and measurable; see the proof of Theorem 1.4. Denote $v_\theta := \inf_{\mu \in \Gamma(\theta)} \mu(c)$, and let us check that $v_\theta > -\infty$. Let $\mu_1, \mu_2, \ldots \in \Gamma(\theta)$, then because $\Gamma(\theta)$ is compact, there exists a convergent subsequence $\mu'_n \to \mu$. By assumption,

$$c_-(x, y) := -\min\big\{c(x, y), 0\big\} \leq C\big(1 + |x| + |y|\big),$$

implying that c_- is uniformly integrable with respect to (μ'_n). Lemma 5.17 of [1] states that then $\liminf_{n\to\infty} \mu'_n(c) \geq \mu(c) > -\infty$.

Define next for each $q \in \mathbb{Q}$ the level sets $L_q := \{\mu \in \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d) : \mu(c) \leq q\}$, which are closed following the argument above. The set-valued mapping

$$\tilde{L}_q(\theta) := \begin{cases} L_q \cap \Gamma(\theta), & \text{if } L_q \cap \Gamma(\theta) \neq \varnothing, \\ \Gamma(\theta), & \text{otherwise,} \end{cases}$$

is compact-valued. Let us turn next into showing that $\theta \mapsto \tilde{L}_q(\theta)$ is a measurable as a set-valued mapping, by considering the set-valued inverse of a closed F

$$\begin{aligned} \tilde{L}_q^-(F) &= \big\{\theta \in S : \tilde{L}_q(\theta) \cap F \neq \varnothing\big\} \\ &= \big\{\theta \in S : \Gamma(\theta) \cap L_q \cap F \neq \varnothing\big\} \cup \big\{\theta : \Gamma(\theta) \cap F \neq \varnothing, \Gamma(\theta) \cap L_q = \varnothing\big\} \\ &= \Gamma^-(F \cap L_q) \cup \big(\Gamma^-(F) \setminus \Gamma^-(L_q)\big), \end{aligned}$$

which is measurable due to the measurability of $\theta \mapsto \Gamma(\theta)$.

It is straightforward to check that

$$\Gamma_{\mathrm{opt}}(\theta) := \big\{\mu \in \Gamma(\theta) : \mu(c) = v_\theta\big\} = \bigcap_{q \in \mathbb{Q}} L_q(\theta).$$

Because $\Gamma_{\text{opt}}(\theta)$ is a countable intersection of compact-valued mappings $\theta \to \tilde{L}_q(\theta)$, and because $\mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$ is a complete separable metric space, it follows ([12], Theorems 3.1 and 4.1), that it is measurable as a set-valued mapping. Because Γ_{opt} is non-empty, compact-valued and measurable, we may apply the measurable selection theorem [20] and the evaluation map ev_B as in the proof of Theorem 1.4 to conclude the existence of the desired γ_θ.

The proof of (ii) is similar, because also $\theta \to \Gamma^*(\theta) := \Gamma(P_\theta, Q_\theta)$ is compact-valued and measurable by Lemma 3.2 and Proposition 3.6. □

We record the following remarks about Proposition 5.4:

(i) The results may be extended into countably many $P_\theta^{(i)}$ in similar lines as Proposition 5.1.
(ii) The assumptions on the cost function c coincide with those of Beiglböck, Henry-Labordère and Penkner [5], who consider the martingale optimal transport problem in the scalar case.
(iii) Proposition 5.4(ii) is probably well known, but we included it for completeness. Indeed, Corollary 5.22 of Villani [35] is similar, without the integrability assumption on P_θ and Q_θ, but with constant lower bound and continuity assumption on c.

Appendix A: Proofs of Lemmas 2.2, 3.1 and 3.2

Proof of Lemma 2.2. We can take $\mathcal{C}$ as the countable family of max-affine convex functions $f : \mathbb{R}^d \to \mathbb{R}$ taking the form

$$f(x) = \max\{\alpha_1^T x + \beta_1, \dots, \alpha_n^T x + \beta_n\},$$

where $n \in \mathbb{N}$ and $\alpha_i, \beta_i \in \mathbb{Q}^d$.

To confirm this, take first a non-negative convex $\phi : \mathbb{R}^d \to \mathbb{R}_+$ with $\mathbb{E}\phi(X) < \infty$. It is not difficult to see that for any $\varepsilon > 0$, we may find a piecewise linear function g defined as an infinite maximum of affine functions with $\alpha_i, \beta_i \in \mathbb{Q}^d$

$$g(x) = \max\{\alpha_i^T x + \beta_i : i \in \mathbb{N}\},$$

such that $|g(x) - \phi(x)| \le \varepsilon/2$ for all $x \in \mathbb{R}^d$. Consequently, $|\mathbb{E}g(x) - \mathbb{E}\phi(x)| \le \varepsilon/2$. Taking

$$g_n(x) := \max\{\alpha_i^T x + \beta_i : i = 1, \dots, n\},$$

then $g_n \in \mathcal{C}$ and $g_n(x) \uparrow g(x)$ pointwise. We conclude by monotone convergence that there exists $g_m \in \mathcal{C}$ such that $|\mathbb{E}\phi(x) - \mathbb{E}g_m(x)| \le \varepsilon$.

For general $\phi : \mathbb{R}^d \to \mathbb{R}_+$, with $\mathbb{E}\phi(X)$ finite, it is sufficient to observe that

$$\lim_{n\to\infty} \mathbb{E} \max\{\phi(x), -n\} = \mathbb{E}\phi(x),$$

and then one can take any $\phi_m = (\phi(x) - m)_+$ and apply the result above to conclude the existence of $f \in \mathcal{C}$ such that $|\mathbb{E}\phi(X) - \mathbb{E}f(X)|$ is arbitrarily small.

Similarly one can take $\mathcal{C}_+$ as the set of increasing $f \in \mathcal{C}$. □

Proof of Lemma 3.1. Assume that $\mu_n \to \mu \in \mathcal{P}_1((\mathbb{R}^d)^n)$. Then $\mu_n \to \mu$ in distribution and (μ_n) is uniformly integrable. If $f : \mathbb{R}^d \to \mathbb{R}$ is continuous and bounded, then so is $f \circ \pi_i : (\mathbb{R}^d)^n \to \mathbb{R}$. Therefore, $(\pi^i_\# \mu_n)(f) = \mu_n(f \circ \pi_i) \to \mu(f \circ \pi_i) = (\pi^i_\# \mu)(f)$. Thus, $\pi^i_\# \mu^n \to \pi^i_\# \mu$ in distribution. It is also easy to see that $(\pi^i_\# \mu_n)$ is uniformly integrable because

$$\begin{aligned}\int_{\mathbb{R}^d} |x_i| \mathbf{1}\big(|x_i| > t\big) \pi^i_\# \mu_n(\mathrm{d}x_i) &= \int_{(\mathbb{R}^d)^n} |x_i| \mathbf{1}\big(|x_i| > t\big) \mu_n(\mathrm{d}x) \\ &\le \int_{(\mathbb{R}^d)^n} |x| \mathbf{1}\big(|x| > t\big) \mu_n(\mathrm{d}x).\end{aligned}$$ □

Proof of Lemma 3.2. Let $\lambda \in \Gamma(\mu, \nu)$. Note that $|(x, y)|/2 \le \max\{|x|, |y|\} =: |x| \vee |y|$ for all $x, y \in \mathbb{R}^d$. Therefore, for any $t > 0$

$$\begin{aligned}&\frac{1}{2} \int \big|(x, y)\big| \mathbf{1}\big(\big|(x, y)\big| > t\big) \lambda(\mathrm{d}x \times \mathrm{d}y) \\ &\quad \le \int \big(|x| \vee |y|\big) \mathbf{1}\big(2\big(|x| \vee |y|\big) > t\big) \lambda(\mathrm{d}x \times \mathrm{d}y) \\ &\quad \le \int |x| \mathbf{1}\left(|x| > \frac{t}{2}\right) \mu(\mathrm{d}x) + \int |y| \mathbf{1}\left(|y| > \frac{t}{2}\right) \nu(\mathrm{d}y).\end{aligned}$$

Because the measures μ and ν have finite first moments, the right-hand side above tends to zero as $t \to \infty$, uniformly with respect to $\lambda \in \Gamma(\mu, \nu)$ We conclude that $\Gamma(\mu, \nu)$ is uniformly integrable and hence also tight. By [1], Proposition 7.1.5, it follows that $\Gamma(\mu, \nu)$ is relatively compact in $\mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d)$.

To verify that $\Gamma(\mu, \nu)$ is closed, it suffices to observe that it can be written as a preimage $\Gamma(\mu, \nu) = \Pi^{-1}(\{(\mu, \nu)\})$ of the map $\Pi : \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d) \to \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$ defined by $\Pi(\lambda) = (\pi^1_\# \lambda, \pi^2_\# \lambda)$ which is continuous by Lemma 3.1. □

Appendix B: Wasserstein distance as a countable supremum

Let $\mathrm{Lip}_1(\mathbb{R}^d) := \{f : \mathbb{R}^d \to \mathbb{R} : |f(x) - f(y)| \le |x - y| \text{ for all } x, y \in \mathbb{R}^d\}$ stand for the set of 1-Lipschitz functions on $\mathbb{R}^d$.

Lemma B.1. *There exists a countable subset* $\mathcal{T}_d \subset \mathrm{Lip}_1(\mathbb{R}^d)$ *such that*

$$d_W(\mu, \nu) = \sup_{f \in \mathrm{Lip}_1(\mathbb{R})} \big[\mu(f) - \nu(f)\big] = \sup_{g \in \mathcal{T}_d} \big[\mu(g) - \nu(g)\big] \tag{B.1}$$

for all $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$.

Proof. The first equality in (B.1) is the Kantorovich–Rubinstein duality [35], Remark 6.5. Inspired by [32], Theorem 3.1.5, we may take $\mathcal{T}_d$ as all functions of the form

$$\begin{aligned} g(x) &= \min\bigl\{g_1(x), g_2(x), \dots, g_n(x)\bigr\}, \\ g_k(x) &= q_k + |x - y_k|, \end{aligned} \tag{B.2}$$

where $n \in \mathbb{N}$, $q_k \in \mathbb{Q}$ and $y_k \in \mathbb{Q}^d$.

Clearly $\mathcal{T}_d \subset \operatorname{Lip}_1(\mathbb{R}^d)$, and for any function $f \in \operatorname{Lip}_1(\mathbb{R}^d)$, any compact set $K \subset \mathbb{R}^d$ and $\varepsilon > 0$ there exists $g \in \mathcal{T}_d$ such that

$$\sup_{x \in K} \bigl|f(x) - g(x)\bigr| \le \varepsilon.$$

Namely, take $y_1, \dots, y_n$ such that for all $x \in K$ there exists $m(x)$ such that $|x - y_{m(x)}| \le \varepsilon/3$, and choose q_k such that $0 \le q_k - f(y_k) \le \varepsilon/3$. Then, $g(x) \ge f(x)$ for all $x \in \mathbb{R}^d$, and for any $x \in K$ we have

$$\begin{aligned} \bigl|g(x) - f(x)\bigr| &= \bigl(g(x) - g(y_{m(x)})\bigr) + \bigl(g(y_{m(x)}) - f(y_{m(x)})\bigr) + \bigl(f(y_{m(x)}) - f(x)\bigr) \\ &\le \frac{\varepsilon}{3} + \bigl(q_{m(x)} - f(y_{m(x)})\bigr) + \frac{\varepsilon}{3}. \end{aligned}$$

Fix then $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $f \in \operatorname{Lip}_1(\mathbb{R}^d)$, which we may assume without loss of generality to satisfy $f(0) = 0$. For any $\varepsilon > 0$, we may find $M < \infty$ such that

$$\int_{|x|>M} |x| \bigl(\mu(\mathrm{d}x) + \nu(\mathrm{d}x)\bigr) \le \frac{\varepsilon}{8},$$

because μ, ν are integrable. Let $g \in \mathcal{T}_d$ such that $|f(x) - g(x)| \le \varepsilon/8$ for all $|x| \le M$. Then,

$$\mu(g) - \nu(g) \ge \mu(f) - \nu(f) - \mu\bigl(|f - g|\bigr) - \nu\bigl(|f - g|\bigr) \ge \mu(f) - \nu(f) - \varepsilon,$$

because

$$\begin{aligned} \mu\bigl(|f - g|\bigr) &\le \frac{\varepsilon}{8} + \int_{|x|>M} \bigl(\bigl|f(x)\bigr| + \bigl|g(x)\bigr|\bigr)\mu(\mathrm{d}x) \\ &\le \frac{\varepsilon}{8} + 2\int_{|x|>M} |x|\mu(\mathrm{d}x) + \bigl|g(0)\bigr|, \end{aligned}$$

so $\mu(|f - g|) \le \varepsilon/2$ and similarly $\nu(|f - g|) \le \varepsilon/2$. □

Appendix C: From measure-valued mappings to kernels

Lemma C.1. *For any Borel set $B \subset \mathbb{R}^d$, the evaluation map $\operatorname{ev}_B : \mu \mapsto \mu(B)$ from $\mathcal{P}_1(\mathbb{R}^d)$ to $\mathbb{R}$ is measurable with respect to the Borel sigma-algebra generated by the Wasserstein metric on $\mathcal{P}_1(\mathbb{R}^d)$.*

Proof. Assume first that B is open. Let f_n be bounded positive continuous functions such that $f_n \uparrow \mathbf{1}_B$ pointwise; such functions exist by Urysohn's lemma.

Note that for each n, the map $\Phi_n : \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$ defined by $\Phi_n(\mu) = \mu(f_n)$ is continuous and thus measurable. Furthermore, the monotone convergence theorem implies that $\Phi_n(\mu) \uparrow \mathrm{ev}_B(\mu)$ for every μ in $\mathcal{P}_1(\mathbb{R}^d)$. Thus the map ev_B is measurable, being a pointwise limit of measurable maps.

We next show that the claim holds for any Borel set. Denote by $\mathcal{E}$ the collection of Borel sets $B \subset \mathbb{R}^d$ such that ev_B is measurable. If $A, B \in \mathcal{E}$ and $A \subset B$, then $\mathrm{ev}_{B\setminus A}(\mu) = \mathrm{ev}_B(\mu) - \mathrm{ev}_A(\mu)$, so $B \setminus A \in \mathcal{E}$. Similarly, one can show that $\mathcal{E}$ is closed under monotone unions, and clearly $\mathbb{R}^d \in \mathcal{E}$. We conclude that $\mathcal{E}$ is a Dynkin's λ-system which contains the open sets of $\mathbb{R}^d$. Because the collection of open sets is closed under finite intersections, an application of a monotone class theorem ([17], Theorem 1.1), shows that $\mathcal{E}$ contains all Borel sets of $\mathbb{R}^d$. □

Acknowledgements

We thank Eero Saksman, Julio Backhoff and a referee for useful remarks. L. Leskelä was working at the University of Jyväskylä while the major part of this article was written. M. Vihola was supported by the Academy of Finland projects 250575 and 274740.

References

[1] Ambrosio, L., Gigli, N. and Savaré, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Basel: Birkhäuser. MR2401600

[2] Andrieu, C. and Roberts, G.O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648

[3] Andrieu, C. and Vihola, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Probab.* **25** 1030–1077. MR3313762

[4] Andrieu, C. and Vihola, M. (2016). Establishing some order amongst exact approximations of MCMCs. *Ann. Appl. Probab.* To appear.

[5] Beiglböck, M., Henry-Labordère, P. and Penkner, F. (2013). Model-independent bounds for option prices – A mass transport approach. *Finance Stoch.* **17** 477–501. MR3066985

[6] Beiglböck, M. and Juillet, N. (2016). On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44** 42–106. MR3456332

[7] Borglin, A. and Keiding, H. (2002). Stochastic dominance and conditional expectation – An insurance theoretical approach. *Geneva Pap. Risk Insur., Theory* **27** 31–48.

[8] Caracciolo, S., Pelissetto, A. and Sokal, A.D. (1990). Nonlocal Monte Carlo algorithm for self-avoiding walks with fixed endpoints. *J. Stat. Phys.* **60** 1–53. MR1063212

[9] Dolinsky, Y. and Soner, H.M. (2014). Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields* **160** 391–427. MR3256817

[10] Fontbona, J., Guérin, H. and Méléard, S. (2010). Measurability of optimal transportation and strong coupling of martingale measures. *Electron. Commun. Probab.* **15** 124–133. MR2643592

[11] Galichon, A., Henry-Labordère, P. and Touzi, N. (2014). A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.* **24** 312–336. MR3161649

[12] Himmelberg, C.J. (1975). Measurable relations. *Fund. Math.* **87** 53–72. MR0367142
[13] Hirsch, F., Profeta, C., Roynette, B. and Yor, M. (2011). *Peacocks and Associated Martingales, with Explicit Constructions*. *Bocconi & Springer Series* **3**. Milan: Springer. MR2808243
[14] Hirshberg, A. and Shortt, R.M. (1998). A version of Strassen's theorem for vector-valued measures. *Proc. Amer. Math. Soc.* **126** 1669–1671. MR1443832
[15] Hobson, D. and Neuberger, A. (2012). Robust bounds for forward start options. *Math. Finance* **22** 31–56. MR2881879
[16] Hobson, D.G. (1998). The maximum maximum of a martingale. In *Séminaire de Probabilités XXXII*. *Lecture Notes in Math.* **1686** 250–263. Berlin: Springer. MR1655298
[17] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications*. New York: Springer. MR1876169
[18] Kellerer, H.G. (1972). Markov-Komposition und eine Anwendung auf Martingale. *Math. Ann.* **198** 99–122. MR0356250
[19] Kertz, R.P. and Rösler, U. (2000). Complete lattices of probability measures with applications to martingale theory. In *Game Theory, Optimal Stopping, Probability and Statistics*. *Institute of Mathematical Statistics Lecture Notes – Monograph Series* **35** 153–177. Beachwood, OH: IMS. MR1833858
[20] Kuratowski, K. and Ryll-Nardzewski, C. (1965). A general theorem on selectors. *Bull. Acad. Polon. Sci., Sér. Sci. Math. Astron. Phys.* **13** 397–403. MR0188994
[21] Leisen, F. and Mira, A. (2008). An extension of Peskun and Tierney orderings to continuous time Markov chains. *Statist. Sinica* **18** 1641–1651. MR2469328
[22] Leskelä, L. (2010). Stochastic relations of random variables and processes. *J. Theoret. Probab.* **23** 523–546. MR2644874
[23] Lowther, G. (2008). Fitting martingales to given marginals. Preprint. Available at arXiv:0808.2319v1.
[24] Marshall, A.W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. *Mathematics in Science and Engineering* **143**. New York: Academic Press. MR0552278
[25] Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR1889865
[26] Peskun, P.H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60** 607–612. MR0362823
[27] Rüschendorf, L. (1991). On conditional stochastic ordering of distributions. *Adv. in Appl. Probab.* **23** 46–63. MR1091092
[28] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. New York: Springer. MR2265633
[29] Shortt, R.M. (1983). Strassen's marginal problem in two or more dimensions. *Z. Wahrsch. Verw. Gebiete* **64** 313–325. MR0716489
[30] Srivastava, S.M. (1998). *A Course on Borel Sets*. *Graduate Texts in Mathematics* **180**. New York: Springer. MR1619545
[31] Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Stat.* **36** 423–439. MR0177430
[32] Stroock, D.W. (1993). *Probability Theory, an Analytic View*. Cambridge: Cambridge Univ. Press. MR1267569
[33] Szekli, R. (1995). *Stochastic Ordering and Dependence in Applied Probability*. *Lecture Notes in Statistics* **97**. New York: Springer. MR1329324
[34] Tierney, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8** 1–9. MR1620401
[35] Villani, C. (2009). *Optimal Transport: Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Berlin: Springer. MR2459454
[36] Whitt, W. (1980). Uniform conditional stochastic order. *J. Appl. Probab.* **17** 112–123. MR0557440

[37] Whitt, W. (1985). Uniform conditional variability ordering of probability distributions. *J. Appl. Probab.* **22** 619–633. MR0799285

Received June 2014 and revised October 2015