SHADOW COUPLINGS

MATHIAS BEIGLBÖCK AND NICOLAS JUILLET

Abstract. A classical result of Strassen asserts that given probabilities μ, ν on the real line which are in convex order, there exists a *martingale coupling* with these marginals, i.e. a random vector (X_1, X_2) such that $X_1 \sim \mu, X_2 \sim \nu$ and $\mathbb{E}[X_2|X_1] = X_1$. Remarkably, it is a non trivial problem to construct particular solutions to this problem. Based on the concept of *shadow* for measures in convex order, we introduce a family of such martingale couplings, each of which admits several characterizations in terms of optimality properties / geometry of the support set / representation through a Skorokhod embedding. As a particular element of this family we recover the (left-)curtain martingale transport, which has recently been studied [11, 22, 15, 9] and which can be viewed as a martingale analogue of the classical monotone rearrangement. As another canonical element of this family we identify a martingale coupling that resembles the usual *product coupling* and appears as an optimizer in the general transport problem recently introduced by Gozlan et al. In addition, this coupling provides an explicit example of a Lipschitz kernel, shedding new light on Kellerer's proof of the existence of Markov martingales with specified marginals.

Keywords: Strassen's theorem, Kellerer's theorem, peacocks, (martingale) optimal transport, general transport costs, Skorokhod embedding

1. Introduction

1.1. **Outline.** Given Polish spaces X, Y, a measure π on $X \times Y$ with marginals μ and ν is called *transport plan*[1] from μ to ν or a *coupling* of μ and ν. Let $\Pi(\mu, \nu)$ be the space of transport plans of marginals μ and ν. We will usually consider probability measures μ, ν on the real line having first moments. Our primary interest lies in the set of *martingale transport plans* which is defined as

$$\begin{aligned}\Pi_M(\mu, \nu) &= \{\pi = \mathrm{Law}(X, Y) \in \Pi(\mu, \nu),\ \mathbb{E}(Y|X) = X\} \\ &= \{\pi \in \Pi(\mu, \nu) : \textstyle\int y \,\mathrm{d}\pi_{x,\cdot} = x \text{ for } \mu\text{-a.e. } x\}.\end{aligned}$$

Here the constraint $\mathbb{E}(Y|X) = X$ means that $\mathbb{E}(Y|X = x) = x$ for μ-almost every $x \in \mathbb{R}$, while $(\pi_{x,\cdot})_{x\in\mathbb{R}}$ denotes the disintegration of π with respect to μ, i.e. the family of conditional laws. By Jensen's inequality, the existence of a martingale transport plan $\pi \in \Pi_M(\mu, \nu)$ implies that μ, ν are in the convex order $\mu \leq_C \nu$, i.e. $\int \phi \,\mathrm{d}\mu \leq \int \phi \,\mathrm{d}\nu$ for every convex $\phi : \mathbb{R} \to \mathbb{R}$. Conversely Strassen [47], established that $\Pi_M(\mu, \nu)$ is nonempty whenever μ, ν are in convex order. While Strassen derived the existence of martingale transport plans as a rather direct consequence of the Hahn-Banach theorem, it seems harder to provide elementary / natural constructions of such martingale transport plans.

2010 *Mathematics Subject Classification.* 60G42.

Key words and phrases. couplings, martingales, peacocks, convex order, optimal transport.

The first author acknowledges support through FWF grant Y782. The second author was partially supported by the "Programme ANR JCJC GMT" (ANR 2011 JS01 011 01).

[1] Throughout, we will introduce new notations in the place they are needed. For the convenience of the reader, we also collect them in the Appendix.

With the aim of defining a systematic martingale coupling we [11] introduced the (left-)curtain coupling π_{lc} which can be seen as a martingale analogue of the monotone rearrangement coupling (alias the *quantile coupling*). An explicit description of the curtain coupling is provided when μ is finitely supported in [11, Section 2]. Another construction using differential equations is given by Henry-Labordère and Touzi [22] for sufficiently regular distributions. Moreover they establish that the curtain coupling is relevant for the pricing of variance swaps. According to [33] the operation of coupling μ and ν through the curtain coupling is continuous so that left-curtain couplings for general measures μ and ν can be approximated using either of the two mentioned constructions, see [33, Remark 2.18]. Hobson and Norgilas [29, 30] deepen the connection to mathematical finance, in particular they use the curtain coupling to obtain robust bounds for the arbitrage free prices of American put options. In [9], a link to the field of Skorokhod embedding is established. Nutz and Stebegg provide extensions to a supermartingale setup [40] and, together with Tan, to a multi-period setup [41]. The continuous time setting was explored in [34] and [21] as a limit of the multi-period setting but with a slightly different approach (with respect to [41]).

In this article we reconsider the particular role of the left-curtain coupling as a distinguished coupling in the set $\Pi_M(\mu,\nu)$. We will define an infinite family of martingale couplings. Roughly speaking, the (left-)curtain coupling will then be recovered as one extreme element of the this family, while at the other end of the spectrum we will obtain a new and rather different type of systematic coupling that we shall call sunset coupling π_{sun}. Whereas the curtain coupling shares a number of properties with the monotone rearrangement coupling in (classical optimal transport), the sunset coupling can be seen as the martingale analogue of the product coupling $\mu\times\nu$. In view of this it is natural that π_{sun} does not appear as an optimizer of the martingale version of the transport problem. However, we shall see in Theorem 1.1 and Section 5 below that it enjoys some optimality properties of a different type. A further particular property is that π_{sun} yields a concrete example of a martingale transport plan which has the *Lipschitz(-Markov) property* (see §1.3 in this introduction and Section 4).

Before diving into the technical details, we try to convey an intuitive description of the subsequent constructions.

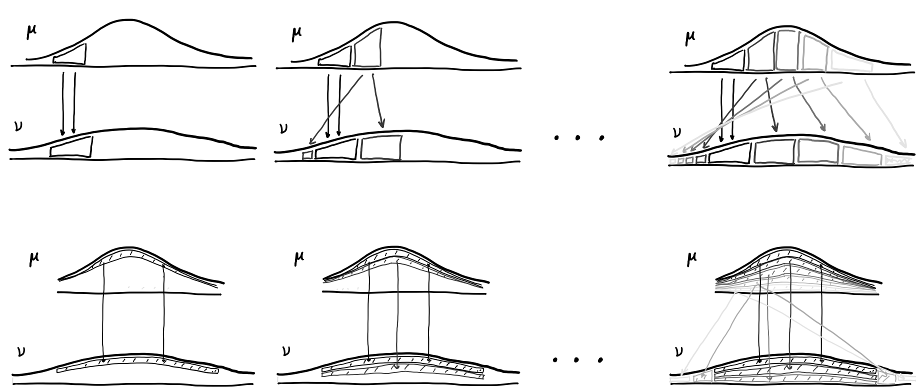

FIGURE 1. Curtain closing from left to right; sun setting from top to bottom.

We start by describing the curtain coupling. All particles of μ are transported to ν, spreading their mass due to the martingale property. Starting with the left most particle and moving to the right, each particle greedily tries to spread its mass as little as possible, given the portion of ν that still needs to be filled, see the upper part of Figure 1. Rather than 'closing the curtain from left to right' as in Figure 1, we could obtain variants of the curtain coupling by 'closing the curtain from right to left', 'from the middle to boundary' (see Subsections 3.1.1, 3.1.3, respectively), etc.

Another (and more interesting) variation is to prioritise the mass in μ not from left to right, but rather from top to bottom. This leads to the sunset coupling depicted in the lower part of Figure 1.

Our first aim will be to rigorously describe the class of martingale couplings arising from constructions as hinted here. Subsequently we will describe applications and links to other themes of research.

1.2. **Main Theorem.** We write λ for the Lebesgue measure on the unit interval and assume that $\mu \leq_C \nu$. Throughout *lift* or *source*[2] of μ will refer to a probability $\hat{\mu} \in \Pi(\lambda, \mu)$ that will serve as a parameter in the construction of a general version of the left-curtain coupling. The set of *lifted martingale transport plans* is

$$\hat{\Pi}_M(\hat{\mu}, \nu) := \left\{ \hat{\pi} \in \Pi(\hat{\mu}, \nu) : \int y \, \mathrm{d}\hat{\pi}_{u,x,\cdot} = x \text{ for } \hat{\mu}\text{-a.e. } (u, x) \right\},$$

where $(\hat{\pi}_{u,x,\cdot})_{(x,u)\in\mathbb{R}\times[0,1]}$ denotes the disintegration of $\hat{\pi}$ with respect to $\hat{\mu}$.

$$\begin{tikzcd}
 & \boxed{\text{“lifted” } \hat{\theta} \in \mathcal{P}([0,1]\times\mathbb{R}^d)} \arrow[dl, shift left, "\text{disintegration}"] \arrow[dr, leftrightarrow, "\hat{\theta}_{[0,u],\cdot}(A) = \hat{\theta}([0,u]\times A)"] & \\
\boxed{\hat{\theta}_{u,\cdot} \in \mathcal{P}(\mathbb{R}^d), u \in [0,1]} \arrow[ur, shift left, "\text{integration}"] \arrow[rr, shift left, "\text{primitive}"] & & \boxed{\hat{\theta}_{[0,u],\cdot} \in \mathcal{M}(\mathbb{R}^d), u \in [0,1]} \arrow[ll, shift left, "\text{derivative}"]
\end{tikzcd}$$

We shall use two further ways to denote the objects $\hat{\mu} \in \Pi(\lambda, \mu) \subseteq \mathcal{P}([0,1]\times\mathbb{R})$ and $\hat{\pi} \in \hat{\Pi}_M(\hat{\mu}, \nu) \subseteq \mathcal{P}([0,1]\times\mathbb{R}^2)$, respectively: given a measure $\hat{\theta}$ on $[0,1]\times\mathbb{R}^d$ (where $d = 1, 2$ so that θ stands for μ or π), with $\mathrm{proj}_{[0,1]}(\hat{\theta}) = \lambda$, we write

(1) $(\hat{\theta}_{u,\cdot})_{u\in[0,1]}$ for the (λ-a.s. unique) disintegration[3] of $\hat{\theta}$ with respect to λ.
(2) $(\hat{\theta}_{[0,u],\cdot})_{u\in[0,1]}$ for the family of measures defined for every $u \in [0,1]$ by

$$\hat{\theta}_{[0,u],\cdot}(A) = \hat{\theta}([0,u]\times A) = \int_0^u \hat{\theta}_{s,\cdot}(A)\, ds \tag{1}$$

where $A \subseteq \mathbb{R}^d$.

Our main result is the following.

Theorem 1.1. *Let μ, ν be real probability measures in convex order and $\hat{\mu} \in \Pi(\lambda, \mu)$. There exists a unique $\hat{\pi} \in \hat{\Pi}_M(\hat{\mu}, \nu)$ satisfying any, and then all of the following properties:*

---

[2]In the recent preprint [14] which considers a continuum time version of the present construction, also the term *parametrization* is used.

[3]We write $\hat{\theta}_{u,\cdot}$ instead of the more commonly used $\hat{\theta}_u$ to emphasize that the disintegration is understood with respect to the first coordinate. In fact, we will also consider disintegrations with respect to to other coordinates subsequently.

(1) $\hat\pi$ *minimizes*

$$\hat\gamma \mapsto \int (1-u)\sqrt{1+y^2}\,\mathrm{d}\hat\gamma(u,x,y) \tag{2}$$

on the set $\hat\Pi_M(\hat\mu,\nu)$*.*

(2) $\hat\pi = \mathrm{Law}(U, B_0, B_\tau)$*, where* U *has uniform law on* $[0,1]$*,* (B_t) *is one dimensional Brownian motion,* $\mathrm{Law}(U,B_0) = \hat\mu$ *and* τ *is the hitting time of the process* $t \mapsto (U, B_t)$ *into a* left barrier *(i.e. a Borel set* $R \subseteq [0,1]\times\mathbb{R}$ *such that* $(u,x)\in R, v\le u$ *implies* $(v,x)\in R$*).*

(3) $\hat\pi(\hat\Gamma) = 1$ *for a Borel set* $\hat\Gamma \subseteq [0,1]\times\mathbb{R}\times\mathbb{R}$ *which is* monotone *in the sense that for all* s,t,x,x',y^-,y^+,y'

$$s<t, (s,x,y^-), (s,x,y^+), (t,x',y') \in \Gamma \Rightarrow y' \notin]y^-,y^+[.$$

(4) For all $u\in[0,1]$*, the projection of* $\hat\pi_{[0,u],\cdot,\cdot}$ *onto the second coordinate is the* shadow *of* $\hat\mu_{[0,u],\cdot}$ *onto the measure* ν*.*

We add some comments to this result:

- In (1) the cost $c : (u,x,y) \mapsto (1-u)\sqrt{1+y^2}$ can be replaced by any positive $c(u,x,y) = \phi(u)\psi(y)$ where $\phi \ge 0$ is strictly decreasing and $\psi \ge 0$ is strictly convex and the minimum over $\hat\Pi_M(\hat\mu,\nu)$ is finite. More generally the same conclusions holds for a nonnegative c with $\partial^{(3)}_{uyy}c < 0$ in a weak sense. Alternative assumptions to $c\ge 0$ are that $\int|\phi|\,\mathrm{d}\lambda, \int|\psi(y)|\,\mathrm{d}\nu < \infty$ or that $c(x,y) \ge A + Bx + Cy$.
- In the setting of (2), $\hat\pi \in \hat\Pi_M(\hat\mu,\nu)$ implies that the martingale $(B_{t\wedge\tau})_{t\ge0}$ is uniformly integrable, see Proposition 5.3 below.
- To make sense of the last point, note that if $\mu'(A) \le \mu(A)$ for every Borel set[4] and $\mu \le_C \nu$, then the set $\{\nu' : \mu' \le_C \nu' \text{ and } \nu' \le \nu\}$ is nonempty and has a smallest element $S^\nu(\mu')$ with respect to $\le_C$, the *shadow* of μ' onto the measure ν (cf. [11, Lemma 4.6] / Definition 2.1 below). Intuitively speaking, among all measures $\nu' \le \nu$ which are larger than μ' in convex order, $S^\nu(\mu')$ is the most concentrated one.

We call the unique element of $\hat\Pi_M(\hat\mu,\nu)$ characterized in Theorem 1.1 the *lifted shadow coupling* with *lift* (or *source*) $\hat\mu$. Its projection onto the two last coordinates is an element π of $\Pi_M(\mu,\nu)$ that we call *shadow coupling* of μ and ν associated to the source $\hat\mu$. Note that π is $\hat\pi_{[0,1],\cdot,\cdot}$ in the terminology introduced in Equation (1).

Figure 2 illustrates Theorem 1.1. Left part: the measure $\hat\mu$ has first marginal λ (as any lifted measure $\hat\mu$), and second marginal μ as depicted on the left side of the figure. On the vertical line with abscissa u starts a 1-dimensional Brownian motion (represented on the figure with double arrows for a possible starting position (u,x) in the support of $\hat\mu$) which will hit the set R at a position (u,y) where it is stopped. The measure π is the joint law in $\mathbb{R}^2$ of the starting position x and the end position y. The measure $\hat\nu$ is the law of the end position (u,y), where the starting position is (u,x) distributed according to $\hat\mu(du,dx)$.

Middle part: the measure $\hat\nu_{u,\cdot}(dy)$ is the law of the end position for a Brownian motion starting on the vertical line according to $\hat\mu_{u,\cdot}(dx)$.

Right part: when considering starting positions (u',x) with $u' \le u$ for a reference $u\in[0,1]$, the measure $\hat\pi$ restricted to $A = [0,u]\times\mathbb{R}\times\mathbb{R}$ has projections $(\mathrm{proj}_{u'})_\#\hat\pi|_A = \lambda|_{[0,u]}$, $(\mathrm{proj}_x)_\#\hat\pi|_A = \hat\mu_{[0,u],\cdot}$ and $(\mathrm{proj}_y)_\#\hat\pi|_A = \hat\nu_{[0,u],\cdot}$. Moreover $(\mathrm{proj}_{x,y})_\#\hat\pi|_A = \hat\pi_{[0,u],\cdot,\cdot}$. To (roughly) read $\hat\pi_{[0,u],\cdot,\cdot}$ on the picture, we start from the horizontal line of coordinate x,

[4] This relation is denoted by $\mu' \le_+ \mu$ later.

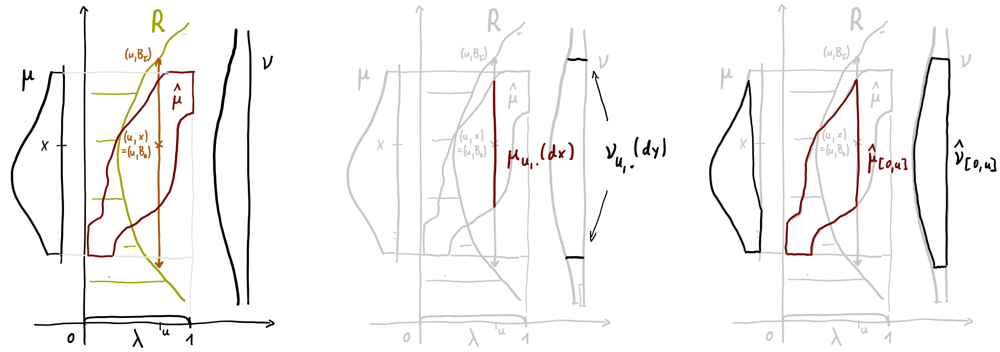

FIGURE 2.

consider the mass of $\hat{\mu}$ when disintegrated until value u (the measure $\hat{\mu}_{\cdot,x}|_{[0,u]}$), start the usual vertical Brownian motions until they hit the barrier and project them back to the vertical axis (that, here, can properly be called y-axis). We have described the transition kernel implied by $\hat{\pi}_{[0,u],\cdot,\cdot}$, from x distributed according to $\hat{\mu}_{[0,u],\cdot}$ to the target (probability) measure that we may denote by $\hat{\pi}_{[0,u],x,\cdot}$.

If the source $\hat{\mu}$ is concentrated on the graph of a 1-1 function $T : [0,1] \to \mathbb{R}$ there is an obvious correspondence between elements of $\Pi_M(\mu,\nu)$ and $\hat{\Pi}_M(\hat{\mu},\nu)$. In particular the optimality property stated in Theorem 1.1 (1) then translates to optimality properties for the martingale version of the transport problem; early papers to investigate such problems include [28, 8, 18, 16, 13, 27, 15, 5, 12]. For general sources, the shadow coupling does not exhibit particular optimality properties for the martingale transport problem. However, it is characterized by a general optimality problem in the sense of Gozlan et al. [20]. We shall discuss this in Section 5 below.

Natural choices of sources lead to natural martingale couplings of shadow type. We shall be particularly interested in the cases where $\hat{\mu}$ is either the quantile or the product coupling of λ and μ:

- The quantile coupling (or monotone rearrangement coupling, see §2.1 and §3.1) of λ and μ is the unique coupling $\hat{\mu}$ whose support is the graph of an increasing function. Considering the corresponding lifted shadow coupling in $\hat{\Pi}_M(\hat{\mu},\nu)$, we recover the *left-curtain coupling* introduced in [11]. We shall henceforth denote this coupling by π_{lc}.
  Notably most of the results established for π_{lc} in [11] are a particular consequence of Theorem 1.1 (cf. Remark 5.6).
- The *sunset coupling* π_{sun} is based on the product source $\hat{\mu} = \lambda \times \mu$, i.e. the independent coupling of λ and μ.

Note that the sources of the above martingale couplings are the two most natural coupling methods for elements in the space without constraint $\Pi(\mu,\nu)$. Looking at the measure μ as the hypograph of its unit density function, we note that the curves $(\hat{\mu}_{[0,u],\cdot})_{u\in[0,1]}$ and $(\hat{\mu}_{u,\cdot})_{u\in[0,1]}$ reminds the reader of a curtain closed from the left in the first case and from the bottom in the second case. This further motivates the names *curtain coupling* and *sunset coupling*. The lifted versions are naturally denoted by $\hat{\pi}_{\mathrm{lc}}$, $\hat{\pi}_{\mathrm{sun}}$ and called *lifted curtain coupling*, *lifted sunset coupling* respectively.

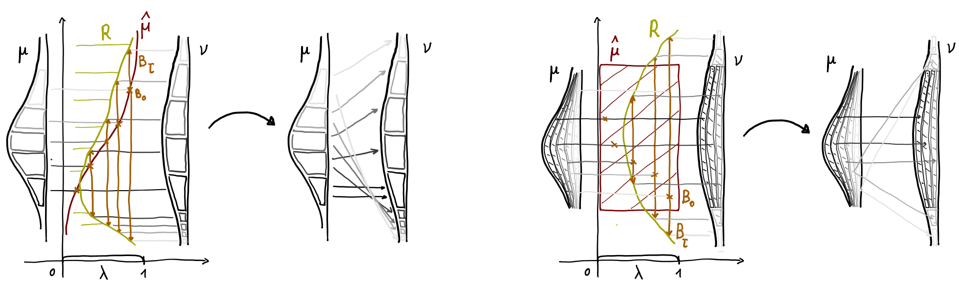

FIGURE 3. From quantile to curtain; from product to sunset.

1.3. **Other systematic methods to define a martingale transport.** A main theme of this paper is to explore systematic methods $(\mu,\nu) \mapsto \pi \in \Pi_M(\mu,\nu)$ to select a special martingale transport plan for the pairs (μ,ν) with $\mu \preceq_C \nu$. By "special" we mean here an element π that can be uniquely characterized with respect to one or several mathematical features / theories. This problem has a rich history and is still being written. Basically, the type of solutions may be into two categories. The most classical is the one of Skorokhod embedding, that may be summarized as follows: starting from μ stop a continuous martingale (mostly the Brownian motion distributed like μ at time 0) so that the distribution at the stopping time is ν. This problem admits a variety of distinct solutions, e.g. the one obtained by Rost [45] that is particularly natural from the perspective of potential theory, the one of Root [44], which is canonical in the sense that it has minimal variance among all stopping times that solve the Skorokhod problem. Another celebrated solution based on recursion theory was given by Azema-Yor [3]. We note that the original construction of Azema-Yor is remarkably simple and explicit but restricted to the case where Brownian motion is started at a constant. In contrast the extension to non-trivial starting laws (due to Hobson [25]) is no longer explicit. We refer the reader to the comprehensive surveys of Obłłój [42] and Hobson [26] on the Skorokhod problem.
The constructions that are the focus of the present paper could be considered more elementary in that their definitions do not rely on an (auxilliary) Brownian motion. Even though computational aspects are not the topic of this paper, we believe that shadow constructions are more adapted to concrete computations (in Section 3.1 we give some explicit formulas when μ and ν are uniform measures). In the context of martingale optimal transport we also mention the recent construction of Jourdain and Marghereti [31] that is particularly useful in the context of inequalities for the usual Wasserstein distance.

1.4. **Kellerer's theorem and the sunset coupling.** Kellerer's Theorem [35] states that if a family of measures $(\mu_t)_{t\in\mathbb{R}_+}$ satisfies $s \le t \Rightarrow \mu_s \preceq_C \mu_t$, there exists a martingale $(X_t)_{t\in\mathbb{R}_+}$ with $\mathrm{Law}(X_t) = \mu_t$ for every t. The martingale can moreover be supposed to be *Markov* and this is, as far as we are concerned, the most spectacular achievement of this theorem. In contemporary terms (see [23]), $(\mu_t)_{t\in\mathbb{R}_+}$ is called a *peacock* and $(X_t)_{t\in\mathbb{R}_+}$ is a Markov martingale *associated* to this peacock.

To the best of our knowledge, all proofs of Kellerer's theorem are based on approximation arguments using sequences of Markov processes. Here the obstacle is that the Markov-property is *not* preserved when passing to the limit. The key insight of Kellerer was to consider Lipschitz-Markov processes, that is Markov-processes whose transition kernels have the following Lipschitz property: a kernel $P : x \mapsto P(x) \in \mathcal{P}$ is called *Lipschitz* (or more precisely 1-*Lipschitz*) if $W(P(x), P(x')) \le |x - x'|$ for all x, x' (see (4) for the definition of the Wasserstein-1 distance W). It is then not hard to see that the property of being a Lipschitz-Markov process *is* preserved when passing to the limit in the sense of finite-dimensional distributions.

Martingale transport plans can be seen as one step martingales and it is possible to compose several of them to define a discrete Markov martingale. The main technical step in Kellerer's proof is therefore to show that given measures μ, ν in convex order there exists a martingale transport plan π whose transition kernel P has the Lipschitz property, see Theorem 1.3 just bellow. Accordingly, to the concept of *Lipschitz kernel P* we add the one of *Lipschitz martingale transport plan π*. The following remark provides several reformulations for it.

Remark 1.2. In Kellerer's terminology [35, 36], martingale transport plans π appear as pairs consisting of an initial measure μ and a so-called dilation P, i.e. a transition kernel satisfying $\int y \, \mathrm{d}P(y) = x$ and $\mu P = \nu$. Note that $P(x) = \pi_{x,\cdot}$, holds μ-almost surely. Thus π is a martingale transport plan if there exists a version $(\tilde{\pi}_{x,\cdot})_{x\in\mathbb{R}}$ of $(\pi_{x,\cdot})_{x\in\mathbb{R}}$ that satisfies $W(\tilde{\pi}_{x,\cdot}, \tilde{\pi}_{x',\cdot}) \le |x-x'|$ for every $x, x' \in \mathbb{R}$ (Recall that $x \to \pi_{x,\cdot}$ is apriori only μ-almost surely defined). As explained in [32] after Definition 5, this is also equivalent to the existence of $A \subseteq \mathbb{R}$ of full measure (for μ) such that $W(\pi_{x,\cdot}, \pi_{x',\cdot}) \le |x-x'|$ is satisfied for every $x, x' \in A$. Slightly abusing notation, we will occasionally identify martingale transport plans with their kernels.

Theorem 1.3 (Kellerer's key result to Kellerer's theorem on Markov martingales)**.** *Let $\mu, \nu \in \mathcal{P}$ be real probability measures in convex order. Then there exists a Lipschitz martingale transport plan π in $\Pi_M(\mu, \nu)$, i.e. $\pi \in \Pi_M(\mu, \nu)$ such that*

$$W(\pi_{x,\cdot}, \pi_{x',\cdot}) \le |x - x'| \tag{3}$$

holds for any x, x' in a set of μ-full measure.

Let us add that in higher dimensions $d \ge 2$ Lipschitz martingale transport plans may not exist among the elements of $\Pi_M(\mu, \nu)$ for some pairs $\mu \le \nu$ (see [32, Proposition 1]). As Lipschitz kernels and their variants are the only known methods for proving the Kellerer theorem, still to our knowledge, it is an open problem whether this theorem holds in dimensions greater than or equal to two.

While the various extremal martingale couplings constructed in [28, 11, 27, 15, 46] do not have the Lipschitz property, we shall see that the sunset coupling has the Lipschitz property. Moreover, as we will see in Section 4 it connects to Kellerer's original proof (in [35]) of the existence of Lipschitz kernels.

2. Preparations and construction

2.1. **Concepts related to the martingale transport problem.** We consider the space $\mathcal{M}$ of positive measures on $\mathbb{R}$ with finite first moments. The subspace of probability measures with finite expectations is denoted by $\mathcal{P}$. In higher dimensions we denote the corresponding spaces by $\mathcal{M}(\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{R}^d)$. For μ, $\nu \in \mathcal{M}$, the Wasserstein-1 distance is defined by

$$W(\mu, \nu) = \sup_{f \in \mathrm{Lip}(1)} \left| \int f \, \mathrm{d}\mu - \int f \, \mathrm{d}\nu \right| \tag{4}$$

and endows $(\mathcal{P}, W)$ with $\mathcal{T}_1$, the usual topology for probability measures with finite first moments[5]. In (4), the supremum is taken over all 1-Lipschitz functions $f : \mathbb{R} \to \mathbb{R}$. We also consider W (with the same definition) on the subspace $m\mathcal{P} = \{\mu \in \mathcal{M} | \mu(\mathbb{R}) = m\} \subseteq \mathcal{M}$ of measures of mass m.

According to the Kantorovich duality theorem, an alternative definition in the case $\mu, \nu \in \mathcal{P}$ is

$$W(\mu, \nu) = \inf_{(\Omega, X, Y)} \mathbb{E}(|Y - X|) \tag{5}$$

where X, $Y : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ are random variables of laws μ and ν. The infimum is taken over all joint laws (X, Y), the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ being part of the minimization problem. Without loss of generality $(\Omega, \mathcal{F}, \mathbb{P})$ can be assumed to be $([0, 1], \mathcal{B}, \lambda)$ where λ is the Lebesgue measure and $\mathcal{B}$ the σ-algebra of Borel sets on $[0, 1]$. On this probability space the quantile functions G_μ and G_ν in fact realize a minimizing coupling, called *quantile coupling*. In particular one has $W(\mu, \nu) = \int_0^1 |G_\nu - G_\mu|$. Recall that the *quantile function* of θ is defined by $G_\theta(u) = \inf\{x \in [-\infty, \infty] : \theta((-\infty, x]) \geq u\}$ as the generalized inverse of the *cumulative distribution function* $F_\theta : x \in \mathbb{R} \mapsto \theta((-\infty, x])$. Finally, an application of Fubini's theorem yields $W(\mu, \nu) = \int_{\mathbb{R}} |F_\nu - F_\mu|$.

A special choice of a 1-Lipschitz function in (4) is the function $f_t : x \in \mathbb{R} \to |x - t| \in \mathbb{R}$. Therefore if $\mu_n \to \mu$ in $\mathcal{M}$, the sequence of functions $u_{\mu_n} : t \mapsto \int f_t(x)\, \mathrm{d}\mu_n(x)$ converges to u_μ pointwise. The converse statement also holds if all the measures have the same mass and barycenter (see [24, Proposition 2.3] or [11, Proposition 4.2]). For $\mu \in \mathcal{M}$, the function u_μ is usually called the *potential function* of μ. Note that u_μ is a convex function with (weak) second derivative is 2μ.

2.2. **Bijection between curves, primitive curves, and lifted measures.** We elaborate on the equivalent expressions of the lifted measures introduced in Subsection 1.2. In short, we are representing the same mathematical object in three ways: we consider the measure $\hat{\theta}$ that may be $\hat{\mu} \in \mathcal{P}([0, 1] \times \mathbb{R}^2)$ or $\hat{\pi} \in \mathcal{P}([0, 1] \times \mathbb{R}^2)$, the almost surely defined disintegration $(\hat{\theta}_{u,\cdot})_{u \in [0,1]}$, and the primitive curve $(\hat{\theta}_{[0,u],\cdot})_{u \in [0,1]}$. We first recall the integrability conditions. The lifted measure $\hat{\theta}$ is a probability measure on $[0, 1] \times \mathbb{R}^d$ (where $d = 1$ or $d = 2$) such that $\hat{\theta}(\hat{\rho}) < +\infty$ is finite where $\rho : x \mapsto \|x\|_{\mathbb{R}^d}$ and $\hat{\rho}(u, x) = \rho(x)$. This integrability condition corresponds to $\hat{\theta}_{[0,1],\cdot}(\rho) < +\infty$ for the primitive curve $(\hat{\theta}_{[0,u],\cdot})_{u \in [0,1]}$ and $\int_0^1 \hat{\theta}_{u,\cdot}(\rho)\, \mathrm{d}u < +\infty$ for $(\hat{\theta}_{u,\cdot})_{u \in [0,1]}$. The marginal condition asserts that $\hat{\theta} \in \Pi(\lambda, \theta)$ for some $\theta \in \mathcal{P}(\mathbb{R}^d)$. In terms of the primitive curve this can be expressed by asserting that $\hat{\theta}_{[0,1],\cdot} = \theta$ and $\hat{\theta}_{[0,u],\cdot}(\mathbb{R}) = u$. The equivalent condition on $(\hat{\theta}_{u,\cdot})_{u \in [0,1]}$ is that λ-almost surely $\hat{\theta}_{u,\cdot} \in \mathcal{P}$ and $\theta = \int_0^1 \hat{\theta}_{u,\cdot}\, \mathrm{d}u$. Note that from a probabilistic point of view, if (U, X) is a random vector of law $\hat{\theta}$ with $U \sim \lambda$, the other representations are given by $(u \times \mathrm{Law}(X | U \leq u))_{u \in [0,1]}$ and $(\mathrm{Law}(X | U = u))_{u \in [0,1]}$. Finally the object that we will ultimately be most interested in is not the source $\hat{\theta}$ but $\theta := \hat{\theta}_{[0,1],\cdot} = \mathrm{Law}(X)$ (in particular for $\theta = \pi$ where the measure in on $\mathbb{R}^2$).

In what follows we explain that the derivative of the primitive curve $(\hat{\theta}_{[0,u],\cdot})_{u \in [0,1]}$ can be considered with respect to $\mathcal{T}_1$, the weak topology with respect to continuous functions which have at most linear growth. Let us start with $\hat{\theta} \in \Pi(\lambda, \theta)$. We disintegrate the measure with respect to the first marginal and obtain an a.s. uniquely determined family $(\hat{\theta}_{u,\cdot})_{u \in [0,1]}$ such that for almost every u, $\hat{\theta}_{u,\cdot} \in \mathcal{P}(\mathbb{R}^d)$. (We can assume that the measure is

[5] A sequence a signed measure $(\mu_n)_n$ converges to μ if $\int \phi d\mu_n \to \int \phi d\mu$ for every continuous function ϕ with growth at most linear.

zero for the other parameters.) Define $\hat\theta_{[0,u],\cdot}$ for $u \in [0,1]$ by

$$\hat\theta_{[0,u],\cdot}(A) = \hat\theta([0,u]\times A) = \int_0^u \hat\theta_{s,\cdot}(A)\,\mathrm{d}s$$

for $A \subseteq \mathbb{R}^d$ Borel. Given a function $f : x \in \mathbb{R}^d \to \mathbb{R}$ with $f(x)/(1+\|x\|)$ bounded, the function $s \mapsto \hat\theta_{s,\cdot}(f)$ is measurable and in $L^1([0,1])$. Hence at almost every time $u \in [0,1]$ the function $t \mapsto \hat\theta_{[0,u],\cdot}(f) = \int_0^u \hat\theta_{t,\cdot}(f)\,\mathrm{d}t$ is differentiable with derivative $\hat\theta_{u,\cdot}(f)$. It is important that the set $L \subseteq [0,1]$ of times at which the derivative exists for all f is a Borel set of full measure, as we will verify in the next paragraph. Before establishing this claim, note that this permits us to define a canonical disintegration $(\tilde{\hat\theta}_{u,\cdot})_{u\in[0,1]}$: we define the measure $\tilde{\hat\theta}_{u,\cdot}$ as the derivative if $u \in L$, and zero otherwise.

We turn now to the claim: let $\mathcal{X}$ be a countable set of functions which is dense in the space $C_c(\mathbb{R}^d)$ of continuous functions with compact support and let $\mathcal{X}_+$ be $\mathcal{X} \cup \{\rho\}$ where $\rho(x) = \|x\|_{\mathbb{R}^d}$. Let $L \subseteq [0,1]$ be the set such that at any time $u \in L$, $\hat\theta_{u,\cdot}$ is a probability measure and $u \mapsto \hat\theta_{[0,u],\cdot}(f)$ has derivative $\hat\theta_{u,\cdot}(f)$ for any $f \in \mathcal{X}_+$ and note that L has full mass. Then, as an increment h goes to zero the measure $h^{-1}(\hat\theta_{[0,u+h],\cdot} - \hat\theta_{[0,u],\cdot})$ weakly converges to $\hat\theta_{u,\cdot}$ and as ρ is a continuous function with linear growth, convergences holds also in $\mathcal{T}_1$, cf. [48, Theorem 7.12]. Thus L is a set of differentiation for any function with finite first moment.

2.3. **General description of the construction.** In what follows we shortly explain the systematic scheme to define $\hat\pi \in \hat\Pi_M(\hat\mu,\nu)$ when the marginals $\hat\mu \in \Pi(\lambda,\mu)$ and $\nu \in \mathcal{P}$ are given. Recall that the resulting coupling $\pi = \hat\pi_{[0,1],\cdot,\cdot} = (\mathrm{proj}_{x,y})_{\#}\hat\pi$ whose marginals are $\mu = \hat\mu_{[0,1],\cdot}$ and ν fits more naturally to the theory of optimal transportation than the lifted coupling $\hat\pi$.

Represent $\hat\mu$ in the form $(\hat\mu_{[0,u],\cdot})_{u\in[0,1]}$. The first canonical operation, called *shadow projection* on ν, consists in building the curve $(\hat\nu_{[0,u],\cdot})_{u\in[0,1]}$ from it (see Definition 2.1). Hence the construction is complete if on a set $L \subseteq [0,1]$ of differentiation (of full measure) of $(\hat\mu_{[0,u],\cdot})_u$ and $(\hat\nu_{[0,u],\cdot})_u$ we know for every $u \in L$ how to canonically choose a joint law $\hat\pi_{u,\cdot,\cdot}$ of the derivatives $\hat\mu_{u,\cdot}$ and $\hat\nu_{u,\cdot}$. In our situation, the martingale constraint on the one side and the fact that we use the convex shadow projection of Definition 2.1 on the other side will make this choice uniquely determined. As we will see $\hat\pi_{u,\cdot,\cdot}$ is related to Kellerer's *hitting projection* (see Definition 2.6). We will thus obtain $(\hat\pi_{u,\cdot})_{u\in[0,1]}$ and equivalently $(\hat\pi_{[0,u],\cdot})_{u\in[0,1]}$ and $\hat\pi$. This construction will be carried out in detail in the proof of Theorem 2.9.

2.4. **Order relations, convex shadow and alternative shadows.** On $\mathcal{M}$ we write $\mu \preceq_{C,+} \nu$ if there exists $\eta \in \mathcal{M}$ with $\mu \preceq_C \eta$ and $\eta \preceq_+ \nu$. Here $\preceq_+$ means $\eta(A) \le \nu(A)$ for every Borel set A. The order $\preceq_{C,+}$ can also be characterized by asserting $\mu(f) \le \nu(f)$ for every convex positive function f. We also introduce the stochastic order $\mu \preceq_{\mathrm{sto}} \nu$ that holds if $\mu(f) \le \nu(f)$ for every integrable increasing function. This is equivalent to $G_\mu \le G_\nu$ or $F_\mu \ge F_\nu$. See [33] for more details on these order relations in the context of martingale optimal transport.

Definition-Proposition 2.1 (Definition of the convex shadow)**.** If μ and ν are positive real measures and $\mu \preceq_{C,+} \nu$ there exists a unique measure $\eta \in \mathcal{M}$ such that

- $\mu \preceq_C \eta$
- $\eta \preceq_+ \nu$
- If η' satisfies the two first conditions (i.e. $\mu \preceq_C \eta' \preceq_+ \nu$), one has $\eta \preceq_C \eta'$.

This measure η is called the *convex shadow* or simply *shadow* of μ in ν and we denote it by $S^\nu(\mu)$.

In the two next examples we illustrate the general description of Subsection 2.3 with two alternative types of shadow projections that may at the same time be more intuitive and be useful for future comparisons in the paper.

Example 2.2. In this example we illustrate how the quantile coupling of μ and ν can be defined going along the construction lines given in §2.3. Let us first recall that the shortest formula for this quantile coupling is $(G_\mu, G_\nu)_\# \lambda$. Similarly note that, as $G_\lambda = \mathrm{Id}$, the quantile coupling $\hat{\mu} = (\mathrm{Id}, G_\mu)_\# \lambda$ is actually the source for the left-curtain coupling. As we will see $\hat{\mu}$ is also the source for our alternative definition of $(G_\mu, G_\nu)_\# \lambda$ as a shadow coupling where a "fake" shadow replaces the (convex) shadow of Definition 2.1).

We first write μ as the superposition of Dirac measures in the following way:

$$\mu = \int_0^1 \hat{\mu}_{u,\cdot}\, \mathrm{d}u$$

with $\hat{\mu}_{u,\cdot} = \delta_{G_\mu(u)}$ for every $u \in [0,1]$. This corresponds to $\hat{\mu}_{[0,u],\cdot} = (G_\mu)_\# \lambda|_{[0,u]}$ and, as mentioned above, $\hat{\mu} = (\mathrm{Id}, G_\mu)_\# \mu$. We perform the same decomposition for ν:

$$\nu = \int_0^1 \hat{\nu}_{u,\cdot}\, \mathrm{d}u,$$

where $\hat{\nu}_{u,\cdot} = \delta_{G_\nu(u)}$ for every $u \in [0,1]$. The quantile transport plan is now obtained in the form

$$\pi = \int_0^1 \hat{\pi}_{u,\cdot,\cdot}\, \mathrm{d}u$$

since $\hat{\pi}_{u,\cdot,\cdot} = \delta_{(G_\mu, G_\nu)(u)}$ is the unique transport plan in $\Pi(\hat{\mu}_{u,\cdot}, \hat{\nu}_{u,\cdot})$.

To see this presentation as an example of shadow coupling we have firstly to see $\hat{\nu}_{[0,u],\cdot}$ as the shadow of $\hat{\mu}_{[0,u],\cdot}$, with a "fake shadow" projection adapted to the present example. Secondly the curve $(\hat{\nu}_{u,\cdot})_u$ is derived from $(\hat{\nu}_{[0,u],\cdot})_u$. To define the fake shadow in a similar fashion as Definition 2.1 we simply say that $\hat{\nu}_{[0,u],\cdot}$ is the measure $\eta \preceq_+ \nu$ with the same mass as $\hat{\mu}_{[0,u],\cdot}$ such that any other $\eta' \preceq_+ \nu$ of mass u satisfies $\eta \preceq_{\mathrm{sto}} \eta'$. Of course $\hat{\nu}_{[0,u],\cdot}$ is simply $(G_\nu)_\# \lambda|_{[0,u]}$ the restriction of ν to the quantiles of level smaller that u up to the fact that we may have to 'break' atoms at $G_\nu(u)$.

Example 2.3. Let us go further than Example 2.2: in preparation to §3.1.3 and §3.2, let us notice that if $\mu \preceq_{\mathrm{sto}} \nu$, the "fake shadow" $(G_\nu)_\# \lambda|_{[0,u]}$ is also greater than $(G_\mu)_\# \lambda|_{[0,u]}$ in stochastic order so that under the condition $\mu \preceq_{\mathrm{sto}} \nu$, the fake shadow is also a "stochastic shadow" in the sense we are going to provide here (compare with Definition 2.1 of the (convex) "true" shadows):

Let μ and ν be positive real measures and let E_μ^ν be the set of measures $\eta \in \mathcal{M}$ such that

- $\mu \preceq_{\mathrm{sto}} \eta$
- $\eta \preceq_+ \nu$

Assume $E_\mu^\nu \neq \emptyset$. Then there exists $\eta \in E_\mu^\nu$ with $\eta \preceq_{\mathrm{sto}} \eta'$ for every $\eta' \in E_\mu^\nu$. We call this measure the *stochastic shadow* of μ in ν.

Based on this definition under the condition $\mu \preceq_{\mathrm{sto}} \nu$ the coupling of Example 2.2 is truly a *stochastic* shadow coupling adapted to the stochastic order, exactly as shadow couplings of the present paper are in fact *convex* shadow couplings adapted to the convex order.

Convex shadows are sometimes difficult to determine. An important fact is that they have the smallest variance among the admissible measures η'. Indeed, $\eta \preceq_C \eta'$ implies $\int x\,\mathrm{d}\eta = \int x\,\mathrm{d}\eta'$ and $\int x^2\mathrm{d}\eta \le \int x^2\mathrm{d}\eta'$ with equality if and only if $\eta = \eta'$ or $\int x^2\mathrm{d}\eta = +\infty$.

Example 2.4 (Shadow of an atom, Example 4.7 in [11]). Let δ be an atom of mass α at a point x. Assume that $\delta \preceq_{C,+} \nu$. Then $S^\nu(\delta)$ is the restriction of ν between two quantiles, more precisely it is $\nu' = (G_\nu)_\# \lambda_{]s;s'[}$ where $s' - s = \alpha$ and the barycenter of ν' is x.

The following result is one of the most important on the structure of shadows (Theorem 4.8 of [11]).

Proposition 2.5 (Structure of shadows)**.** *Let γ_1, γ_2 and ν be elements of $\mathcal{M}$ and assume $\gamma_1 + \gamma_2 \preceq_{C,+} \nu$. Then, we have $\gamma_2 \preceq_{C,+} \nu - S^\nu(\gamma_1)$ and*

$$S^\nu(\gamma_1 + \gamma_2) = S^\nu(\gamma_1) + S^{\nu - S^\nu(\gamma_1)}(\gamma_2).$$

An important consequence is that if $(\hat{\mu}_{[0,u],\cdot})_{u\in[0,1]}$ is a primitive curve and $\hat{\mu}_{[0,1],\cdot} \preceq_C \nu$, then the curve $(\hat{\nu}_{[0,u],\cdot})_{u\in[0,1]}$ satisfies $\hat{\nu}_{[0,u],\cdot}(\mathbb{R}) = u$ and using $\gamma_1 = \hat{\mu}_{[0,u],\cdot}$ and $\gamma_1 + \gamma_2 = \hat{\mu}_{[0,u'],\cdot}$ we obtain $\hat{\nu}_{[0,u],\cdot} \preceq_+ \hat{\nu}_{[0,u'],\cdot}$ for every $u \le u'$. Hence $(\hat{\nu}_{[0,u],\cdot})_u$ is a primitive curve. In the next subsection we consider the derivatives of such curves $(\hat{\nu}_{[0,u],\cdot})_u$ and introduce for this the proper infinitesimal version of the shadow projection.

2.5. **Hitting projection.** We denote by $\mathcal{F}(\mathbb{R})$ the space of closed subsets of $\mathbb{R}$, and $\mathcal{I}$ the subspace of those elements $T \in \mathcal{F}(\mathbb{R})$ such that $\sup T = -\inf T = +\infty$. The space $\mathcal{F}(\mathbb{R})$ is endowed with a natural topology presented in [36, §2.1], i.e. the coarsest topology such that $F \in \mathcal{F}(\mathbb{R}) \mapsto d(x, F)$ is continuous for every $x \in \mathbb{R}$.

Definition 2.6 (Hitting projection of measure in/to a set)**.** Let T be an element of $\mathcal{I}$. For every $x \in \mathbb{R}$, let $x_T^- = \sup(T \cap (-\infty, x])$ and $x_T^+ = \inf(T \cap [x, +\infty))$. The Kellerer dilation ([36, Definition 16]) is given by

$$P_T(x, \cdot) = \begin{cases} \delta_x & \text{if } x \in T; \\ (x_T^+ - x_T^-)^{-1}[(x_T^+ - x)\delta_{x_T^-} + (x - x_T^-)\delta_{x_T^+}] & \text{otherwise.} \end{cases}$$

Hence if $\mu \in \mathcal{P}$, the hitting projection of μ in T is $\nu = \mu P_T$ and the hitting coupling of μ and ν is given by $\pi(A \times B) = \int_A P_T(x, B)\,\mathrm{d}\mu(x)$; we shall abbreviate this by $\pi = \mu(\mathrm{id} \times P_T)$.

Note that if T is not an element of $\mathcal{I}$ but $\operatorname{supp}(\mu) \subseteq [\inf T, \sup T]$, the kernel P_T still makes sense μ-almost surely.

Proposition 2.7. *For $\hat{\mu} \in \Pi(\lambda, \mu)$ and $\mu \preceq_C \nu$, let $u \mapsto \hat{\mu}_{[0,u],\cdot}$ have right derivative $\hat{\mu}_{u_0,\cdot}$ at u_0 and let $\hat{\nu}_{[0,u],\cdot}$ be $S^\nu(\hat{\mu}_{[0,u],\cdot})$. Then $(\hat{\nu}_{[0,u],\cdot})$ has a right derivative at u_0. This derivative is given by $\hat{\mu}_{u_0,\cdot} P_T$ and $\operatorname{supp}(\hat{\mu}_{u_0,\cdot}) \subseteq [\inf T, \sup T]$ where T is the support of $\hat{\nu}_{]u_0,1],\cdot} := \nu - \hat{\nu}_{[0,u_0],\cdot}$.*

Proof. Consider $h^{-1}(\hat{\nu}_{[0,u_0+h],\cdot} - \hat{\nu}_{[0,u_0],\cdot}) = h^{-1}(S^\nu(\hat{\mu}_{[0,u_0+h],\cdot}) - S^\nu(\hat{\mu}_{[0,u_0],\cdot})) =: \sigma_h$. According to Proposition 2.5, σ_h equals

$$h^{-1} S^{\hat{\nu}_{]u_0,1],\cdot}}(\hat{\mu}_{]u_0,u_0+h],\cdot}),$$

where we set $\hat{\mu}_{]u,v],\cdot} = \hat{\mu}_{[0,v],\cdot} - \hat{\mu}_{[0,u],\cdot}$. However, we know that $h^{-1}(\hat{\mu}_{[0,u_0+h],\cdot} - \hat{\mu}_{[0,u_0],\cdot}) = h^{-1}\hat{\mu}_{]u_0,u_0+h],\cdot}$ tends to $\hat{\mu}_{u_0,\cdot}$ as $h \downarrow 0$. An easy scaling analysis shows that σ_h can in fact be written as

$$\sigma_h = S^{h^{-1}\hat{\nu}_{]u_0,1],\cdot}}(h^{-1}\hat{\mu}_{]u_0,u_0+h],\cdot}).$$

Formally, in the limit, we are considering the shadow projection of $\hat{\mu}_{u_0,\cdot}$ into the infinite measure $\infty \cdot \hat{\nu}_{]u_0,1],\cdot}$. Here it is easy to believe that the support of $\hat{\nu}_{]u_0,1],\cdot}$, denoted by T in our statement plays the leading role. We are indeed left with the proof that σ_h converges to $\hat{\mu}_{u_0,\cdot}P_T$ as $h \downarrow 0$ (and of $\operatorname{supp}(\hat{\mu}_{u_0,\cdot}) \subseteq [\inf T, \sup T]$). Since $\mathcal{T}_1$ is metric, it is enough to replace h by a sequence $(h_n)_n$ and apply the postponed Lemma 2.8 to the sequence $(H_n)_n$ given through $H_n := h_n^{-1}$, the sequence of probability measures $h_n^{-1}\hat{\mu}_{]u_0,u_0+h_n],\cdot}$ for η_n, and the measure $\hat{\nu}_{]u_0,1],\cdot}$ for υ. Note that this lemma delivers at the same time the convergence and the claimed inclusion. □

Lemma 2.8. *Let $(H_n)_n$ be a sequence of positive numbers tending to infinity, $(\eta_n)_n$ a sequence of probability measures (with finite first moment) converging to η in $\mathcal{P}$ and υ a positive measure. Assume $\eta_n \preceq_{C,+} H_n\upsilon$ for every $n \geq 1$. Then, denoting by T the support* $\operatorname{supp}(\upsilon)$ *of υ it holds* $\operatorname{supp}(\eta) \subseteq [\inf T, \sup T]$ *(replacing the bounds by $\pm\infty$ if necessary) and $S^{H_n\upsilon}(\eta_n) \to \eta P_T$ in $\mathcal{P}$.*

Proof. Note first that due to the convex order relation $\eta_n \preceq_{C,+} H_n\upsilon$ we have $\operatorname{supp}(\eta_n) \subseteq \operatorname{supp}(\upsilon) \subseteq [\inf T, \sup T]$ so that $\eta_n([\inf T, \sup T]) = 1$, for every n. Letting n tend to infinity we find $\operatorname{supp}(\eta) \subseteq [\inf T, \sup T]$ as well.

1. Let us prove the result if $\eta_n = \delta_x$ for every $n \in \mathbb{N}$. We prove in fact a somewhat stronger statement: if γ_n has mass less than or equal to one and $\gamma_n \preceq_+ H_n\upsilon$, the sequence $S^{H_n\upsilon-\gamma_n}(\delta_x)$ converges to ηP_T. Moreover $x \in T^\circ$, $x \in T^c$ or $x \in \partial T$. In either of these cases the result easily follows from Example 2.4.

2. We assume now that for every $n \in \mathbb{N}$, we have $\eta_n = \eta = \sum_{k=1}^n a_k\delta_{x_k}$. We proceed by induction. The initial step $n = 1$ has just been established. We assume the statement for $n-1 \geq 1$ and prove it for n by using the decomposition $\eta = \eta' + a_n\delta_n$ where $\eta' = \sum_{k=1}^{n-1} a_k\delta_{x_k}$. By Proposition 2.5 we have

$$S^{H_n\upsilon}(\eta' + a_n\delta_{x_n}) = S^{H_n\upsilon}(\eta') + S^{\beta_n}(a_n\delta_{x_n})$$

where $\beta_n = H_n\upsilon - S^{H_n\upsilon}(\eta')$. Each of the two terms converges to the Kellerer projection of η' respectively $a_n\delta_n$ onto T. Note that for the second projection we used the full strength of the statement proved in 1.

3. A general measure η can be approximated using a convex combination of Dirac masses η_k with $\eta_k \preceq_C \eta$ and such that $\eta_k \to \eta$ [33, Point 3. in the proof of Proposition 2.34]. We have

$$W(S^{H_n\upsilon}(\eta_k), S^{H_n\upsilon}(\eta)) \leq W(\eta_k, \eta).$$

This tends to zero uniformly in n as k goes to infinity. But $S^{H_n\upsilon}(\eta_k) \to \eta_k P_T$ as n tends to infinity and the composition with P_T is continuous (cf. [36, Section 2.2], this can be understood easily from the action of P_T on the potential functions). Hence we obtain the result for any constant sequence $\eta_n = \eta$.

4. If η_n is a non-constant sequence

$$W(S^{H_n\upsilon}(\eta_n), \eta P_T) \leq W(S^{H_n\upsilon}(\eta_n), S^{H_n\upsilon}(\eta)) + W(S^{H_n\upsilon}(\eta), \eta P_T),$$

which tends to zero as required ([33, Proposition 2.34]). □

Based on the preparations above we can now rigorously introduce the *lifted shadow couplings*.

Theorem 2.9 (Existence, construction, and uniqueness of the lifted shadow coupling)**.** *Let μ and ν be elements of $\mathcal{P}$ with $\mu \preceq_C \nu$. Let $\hat{\mu}$ be an element of $\Pi(\lambda, \mu)$. Then there exists a unique element $\hat{\pi} \in \hat{\Pi}_M(\hat{\mu}, \nu)$, the lifted shadow coupling of $\hat{\mu}$ and ν, such that for every $u \in [0,1]$, the marginals of $\hat{\pi}_{[0,u],\cdot,\cdot}$ are $\hat{\mu}_{[0,u],\cdot}$ and its shadow projection $S^\nu(\hat{\mu}_{[0,u],\cdot})$. Denoting this second marginal by $\hat{\nu}_{[0,u],\cdot}$ and the derivative of $(\hat{\nu}_{[0,u],\cdot})_{u\in[0,1]}$ by $\hat{\nu}_{u,\cdot}$, we have moreover $\hat{\nu}_{u,\cdot} = \hat{\mu}_{u,\cdot}P_{T(u)}$ where $T(u)$ is the support of $\hat{\nu}_{]u,1],,\cdot} := \hat{\nu}_{[0,1],\cdot} - \hat{\nu}_{[0,u],\cdot}$.*

Note that Theorem 2.9 implies in particular that there exists a unique $\hat{\pi}$ satisfying Theorem 1.1 (4).

Proof of Theorem 2.9. Let $\hat{\nu}_{[0,u],\cdot} = S^{\nu}(\hat{\mu}_{[0,u],\cdot})$ be as in the statement and let $(\hat{\mu}_{u,\cdot})_{u\in[0,1]}$ and $(\hat{\nu}_{u,\cdot})_{u\in[0,1]}$ be the derivative curves. According to Proposition 2.7, $\hat{\nu}_{u,\cdot}$ can for almost every $u \in [0,1]$ be identified with the Kellerer projection of $\hat{\mu}_{u,\cdot}$ in $T(u) := \mathrm{supp}(\hat{\nu}_{]u,1],\cdot})$ where $\hat{\nu}_{]u,1],\cdot} = \nu - \hat{\nu}_{]0,u],\cdot}$. Hence we can define $\hat{\pi}_{u,\cdot} = \hat{\mu}_{u,\cdot}(\mathrm{id}\times P_{T(u)}) \in \Pi_M(\hat{\mu}_{u,\cdot}, \hat{\nu}_{u,\cdot})$ for almost every u, and then the corresponding $\hat{\pi}$ and $\hat{\pi}_{[0,u],\cdot} = \int_0^u \hat{\pi}_{t,\cdot}\,\mathrm{d}t$.

Conversely, let $\hat{\pi}$ and $(\hat{\pi}_{[0,u],\cdot})_{u\in[0,1]}$ be as in the statement. The curve has marginals $\hat{\mu}_{[0,u],\cdot}$ and $\hat{\nu}_{[0,u],\cdot} = S^{\nu}(\hat{\mu}_{[0,u],\cdot})$ so that one can apply Proposition 2.7. At points u where the derivatives $\hat{\pi}_{u,\cdot}$, $\hat{\mu}_{u,\cdot}$ and $\hat{\nu}_{u,\cdot}$ exist, we have $\hat{\pi}_{u,\cdot,\cdot} \in \Pi_M(\hat{\mu}_{u,\cdot}, \hat{\nu}_{u,\cdot})$ and $\hat{\nu}_{u,\cdot} = \hat{\mu}_{u,\cdot}P_{T(u)}$ as in the last paragraph, proving the uniqueness part of the statement. □

3. Generating martingale couplings from particular lifts $\hat{\mu}$

3.1. **Examples of shadow couplings.** We now further discuss three special lifts $\hat{\mu} \in \Pi(\lambda,\mu)$ of μ (with their associated primitive families $(\hat{\mu}_{[0,u],\cdot})_{u\in[0,1]}$) to give rise to particular shadow couplings. The first one is the monotone coupling of λ and μ and the second the independent coupling $\lambda\times\mu$.

- $\hat{\mu}_{[0,u],\cdot} = (G_\mu)_\#\lambda|_{[0,u]}$ (corresponding to the left-curtain coupling ; # is the push-forward operator);
- $\hat{\mu}_{[0,u],\cdot} = u\cdot\mu$ (corresponding to the sunset coupling);
- $\hat{\mu}_{[0,u],\cdot} = S^{\mu}(u.\delta_m)$ where $m = \int x\,\mathrm{d}\mu(x)$ (corresponding to the middle curtain coupling). Recall Example 2.4 for the shadow of an atom.

Figure 4 illustrate the corresponding Skorokhod embeddings when μ and ν are two uniform laws $\mu = \mathbb{1}_{[0,1]}dx$ and $\nu = \frac{1}{3}\mathbb{1}_{[-1,2]}dx$. Based on Remark 3.1 below he corresponding equations for the shadows are summerized in Table 1, while Table 2 presents the equations of the sources of the particular shadow couplings.

	$\hat{\mu}_{[0,u],\cdot}$	$\hat{\mu}_{[0,u],\cdot}$	$\hat{\nu}_{[0,u],\cdot}$
π_{lc}	$\mu\vert_{(-\infty,G_\mu(u))} + (u - F_\mu\circ G_\mu(u))\delta_{G_\mu(u)}$	$\mathbb{1}_{[0,u]}dx$	$\frac{1}{3}\mathbb{1}_{[-u,2u]}dy$
π_{sun}	$u\cdot\mu$	$u\cdot\mathbb{1}_{[0,1]}dx$	$u\cdot\mathbb{1}_{[0,1]}dy$ or $\frac{1}{3}\mathbb{1}_{[1-2u,1+2u]}dy$
π_{mid}	$\mu\vert_{(f(u),g(u))} + c(u)\delta_{f(u)} + d(u)\delta_{g(u)}$	$\mathbb{1}_{[\frac{1-u}{2},\frac{1+u}{2}]}dx$	$\frac{1}{3}\mathbb{1}_{[1-2u,1+2u]}dy$

Table 1. General expression of $\hat{\mu}_{[0,u],\cdot}$ and expressions of $\hat{\mu}_{[0,u],\cdot}$ and $\hat{\nu}_{[0,u],\cdot}$ for uniform measures on $[0,1]$ and $[-1,2]$.

	$\hat{\mu}$	$\hat{\mu}_{u,\cdot}$	$\hat{\mu}_{[0,u],\cdot}$
π_{lc}	$(\mathrm{Id}\times G_\mu)_\#\lambda$	$\delta_{G_\mu(u)}$	$\mu\vert_{(-\infty,G_\mu(u))} + (u - F_\mu\circ G_\mu(u))\delta_{G_\mu(u)}$
π_{sun}	$\lambda\times\mu$	μ	$u\cdot\mu$
π_{mid}	$a\cdot(\mathrm{Id}\times f)_\#\lambda + b\cdot(\mathrm{Id}\times g)_\#\lambda$	$a(u)\delta_{f(u)} + b(u)\delta_{g(u)}$	$\mu\vert_{(f(u),g(u))} + c(u)\delta_{f(u)} + d(u)\delta_{g(u)}$

Table 2. General expressions of $\hat{\mu}$, $\hat{\mu}_{u,\cdot}$ and $\hat{\mu}_{[0,u],\cdot}$ for the the main lift methods.

Remark 3.1 (Shadow of a uniform measure in a uniform measure). In the following we illustrate our three examples with the fixed pair $\mu = \mathbb{1}_{[0,1]}dx$ and $\nu = \frac{1}{3}\mathbb{1}_{[-1,2]}dx$. Therefore it is useful to clearly state that two measures $\eta = m\mathbb{1}_{[a,b]}$ and $\eta' = m'\mathbb{1}_{[a',b']}$ satisfiy $\eta \preceq_{C,+} \eta'$ if and only if $\frac{a+b}{2} \pm \frac{m}{m'}\frac{b-a}{2} \in [a',b']$. In this case the shadow is $m'\mathbb{1}_{[\frac{a+b}{2}-\frac{m}{m'}\frac{b-a}{2},\frac{a+b}{2}+\frac{m}{m'}\frac{b-a}{2}]}$, that is the measure with the same mass and barycenter as η, i.e. the restriction of η' to an interval.

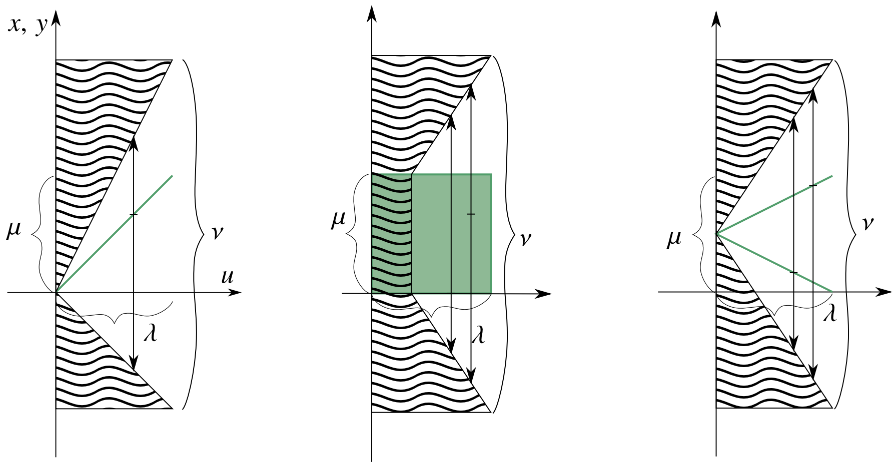

FIGURE 4. i. Left-curtain coupling of uniform measures. The map $G_\mu : [0,1] \to \mathbb{R}$ is identity and $\hat{\mu}$ is uniform on this graph, i.e. on the segment of ends $(0,0)$ and $(1,1)$. ii. Sunset coupling of uniform measures. $\hat{\mu}$ is uniform on the square $[0,1]^2$. iii. Middle-curtain coupling of uniform measures. $\hat{\mu}$ is uniform on the segment with ends $(0,1/2)$ and $(1,1)$ and $(1,0)$. These segments are the graphs of f and g. The measures μ and ν satisfy $\mu \preceq_{DC} \nu$ (an order relation defined in §3.1.3).

3.1.1. *The left- and right-curtain couplings.* This case corresponds to the construction given in [11], even though the construction described there appears slightly different. In fact for $u = F_\mu(x)$ the three marginals of $\hat{\pi}|_{[0,1]\times]-\infty,x]\times\mathbb{R}}$ are $\lambda|_{[0,u]}$, $\mu|_{]-\infty,x]}$ and $S^\nu(\mu_{]-\infty,x]})$ so that for every $x \in \mathbb{R}$, $\pi|_{]-\infty,x]\times\mathbb{R}}$ has marginals $\mu|_{]-\infty,x]}$ and $S^\nu(\mu_{]-\infty,x]})$. In [11] this was used to define the left-curtain coupling π_{lc}.

In an entirely symmetric fashion we can define the right-curtain coupling through $\hat{\mu}_{[0,u],\cdot} = (G_\mu)_\# \lambda|_{[1-u,1]}$ for $u \in [0,1]$.

3.1.2. *The sunset coupling.* In this case we have $\hat{\mu}_{u,\cdot} = \mu$ for almost every $u \in [0,1]$ and $\hat{\nu}_{u,\cdot} = \mu P_{T(u)}$ where $T(u) = \operatorname{supp}(\nu - S^\nu(u\cdot\mu))$. Hence

$$\nu = \int_0^1 \mu P_{T(u)}\,\mathrm{d}u$$

Of course $T(u)$ can be replaced by $T^*(u) = T(u) \cup (]-\infty, \inf \nu] \cup [\sup \nu, +\infty[) \in \mathcal{I}$. We refer the reader to Section 4 for further explanations.

3.1.3. *The middle-curtain coupling.* For diatomic probability measures $\mu = a\delta_f + b\delta_g$ and $\nu = a'\delta_{f'} + b'\delta_{g'}$ the relation $\mu \preceq_C \nu$ holds if and only if $af + bg = a'f' + b'g'$ (same mean) and $[a,b] \subseteq [a',b']$. In this case, as one can easily check, there exists a unique martingale transport plan in $\Pi_M(\mu,\nu)$. It is

$$\pi_{\mathrm{mid}}(\mu,\nu) = \frac{1}{g'-f'}\Big([(g'-f)\delta_{f,f'} + (f-f')\delta_{f,g'}] + [(g'-g)\delta_{g,f'} + (g-f')\delta_{g,g'}]\Big). \tag{6}$$

This basic fact permits us to construct simple martingale couplings for a special class of ordered pairs $\mu \preceq_C \nu$ in a parallel way to the way the quantile coupling is defined in Example 2.3, under $\mu \preceq_{\mathrm{sto}} \nu$. This martingale coupling is defined in [32] without assigning a particular name. Here we shall call it *middle-curtain coupling*. In our case, diatomic probability measures with mean m replace the atoms of Example 2.2. There is a almost surely unique family of diatomic measures $(\hat\mu_{u,\cdot})_{u\in[0,1]}$ such that $\int_0^1(\hat\mu_{u,\cdot})_{u\in[0,1]}\mathrm{d}u = \mu$ and $\hat\mu_{u,\cdot} \preceq_C \hat\mu_{u',\cdot}$ for every $u < u'$. One (maybe too) elaborate way to see this is to define $\hat\mu_{[0,u],\cdot} = S^\mu(u\cdot\delta_m)$. We recall (see Example 2.4) that this measure is the restriction of μ to an interval (with more or less mass at the ends of the interval) of mass u and mean m. Consequently the derivative $\hat\mu_{u,\cdot}$ is the diatomic probability measure announced above. Note that this corresponds to $\hat\mu_{u,\cdot} = a(u)\delta_{f(u)} + b(u)\delta_{g(u)}$ where

- $(a+b)(u) = 1$;
- $a(u)f(u) + b(u)g(u) = m$;
- f is decreasing and g is increasing (we can moreover assume that they are right continuous, as quantile functions are).

In the family $(\hat\mu_{[0,u],\cdot})_{u\in[0,1]}$, the mass appears from the middle which explains the name 'middle-curtain coupling'.

It may happen that

$$\hat\mu_{u,\cdot} \preceq_C \hat\nu_{u,\cdot} \quad \text{for every } u\in[0,1] \tag{7}$$

where we denote by $(\hat\nu_{u,\cdot})_{u\in[0,1]}$ and $\hat\nu_{[0,u],\cdot}$ the measures defined in the same way as $\hat\mu_{u,\cdot}$ and $\hat\mu_{[0,u],\cdot}$. Integrating the martingales couplings of (6) gives a martingale coupling

$$\pi_{\mathrm{mid}}(\mu,\nu) = \int_0^1 \pi_{\mathrm{mid}}(\hat\mu_{u,\cdot},\hat\nu_{u,\cdot})\,\mathrm{d}u \tag{8}$$

of μ and ν, so that, in particular, $\mu\preceq_C\nu$. However, contrary to what happens in Example 2.3 for the stochastic order and the quantile coupling, $\mu\preceq_C\nu$ is not equivalent to (7). We call the latter necessary (but not sufficient) condition the *diatomic convex order* and denote it by $\mu\preceq_{DC}\nu$. Due to the non-uniqueness of the derivative curve, a condition that better define this order is probably $S^\mu(u\cdot\delta_m)\preceq_C S^\nu(u\cdot\delta_m)$ for every $u\leq 1$. It can be easily proved to be equivalent.

As the notation π_{mid} suggests, the coupling in (6) and (8) are the middle-curtain couplings. The second formula generalises the first one. Let us see that the middle-curtain coupling is a shadow coupling. We already expressed $\hat\mu_{[0,u],\cdot} = S^\mu(u\cdot\delta_m)$. As we assumed $\mu\preceq_{DC}\nu$, we have

$$u\cdot\delta_m \preceq_C \underbrace{S^\mu(u\cdot\delta_m)}_{\hat\mu_{[0,u],\cdot}} \preceq_C \underbrace{S^\nu(u\cdot\delta_m)}_{\hat\nu_{[0,u],\cdot}} \leq \nu$$

so that $S^\nu(\hat\mu_{[0,u],\cdot}) = \hat\nu_{[0,u],\cdot}$ is not difficult to derive from Definition 2.1. Therefore, under $\mu\preceq_{DC}\nu$ the coupling (8) is the shadow coupling of source $\hat\mu$. If $\mu\preceq_C\nu$ but not $\mu\preceq_{DC}\nu$, equation (8) has no meaning because $\hat\mu_{u,\cdot}\preceq_C\hat\nu_{u,\cdot}$ is not satisfied for every $u\in[0,1]$, so that $\pi_{\mathrm{mid}}(\hat\mu_{u,\cdot},\hat\nu_{u,\cdot})$ does not exist. However, the shadow coupling is still defined, which permits to naturally extend the idea of a middle-curtain coupling originally intoduced in [32] to any ordered pair $\mu\preceq_C\nu$. Note that in this case, $S^\nu(\hat\mu_{[0,u],\cdot})\neq S^\nu(u\cdot\delta_m)$ at least for one $u\in[0,1]$.

Remark 3.2. Finally, note that for every family of probability measures $(\mu_t)_{t\in T}$ indexed by any partial order T so that $s\leq t$ implies $\mu_s\preceq_{DC}\mu_t$, there exists a martingale $(X_t)_{t\in T}$ such that $\mathrm{Law}(X_s,X_t)$ is the middle-curtain coupling of μ_s and μ_t for every $s<t\in T$, [32, Theorem 4].

3.2. **Comparison with the stochastic order; quantile and independent couplings.** We have seen in Examples 2.2 and 2.3 that with the quantile coupling as source $\hat\mu$, and modifying the definition of shadow (fake shadow or stochastic shadow respectively), the related shadow couplings are in both examples the quantile coupling of μ and ν. In this sense the left-curtain coupling can be seen as the quantile coupling with respect to the convex order. The left-curtain and the quantile coupling are also analogous on the level of optimality properties, see [11, Sections 1.2, 1.3].

While the left-curtain coupling can be viewed as the quantile coupling of the convex order world, we will explain next in which sense the sunset coupling corresponds to the independent (aka product) coupling. This comparison is based on the fake shadow of Example 2.2 but is unfortunately not coherent with Example 2.3. Thus, $\hat{\nu}_{[0,u],\cdot} = (G_\nu)_\# \lambda|_{[0,u]}$. We take the same source $\hat{\mu} = \lambda \times \mu$ as for the sunset coupling, so that $\hat{\mu}_{u,\cdot} = \mu$. It is then easy to identify the derivative in the target space as $\hat{\nu}_{u,\cdot} = \delta_{G_\nu(u)}$. But $\mu \times \delta_{G_\nu(u)}$ is the unique element of $\Pi(\mu, \delta_{G_\nu(u)})$. We thus obtain

$$\hat{\pi}_{[0,1],\cdot,\cdot} = \pi = \int_0^1 \mu \times \delta_{G_\nu(u)}\, \mathrm{d}u.$$

This coupling is nothing but the product $\mu \times \nu$. Recall that for the same source $\lambda \times \mu$ and the convex shadow (of Definition 2.1) we obtain the sunset coupling.

4. On Theorem 1.3 by Kellerer and the sunset coupling

In this section we give a new proof of Theorem 1.3 which is the key step in Kellerer's theorem on the existence of Markov martingales with given marginals. Recall that Theorem 1.3 is a result on the existence of a *special element* of $\Pi_M(\mu, \nu)$ where μ and ν are in convex ordering, extending Strassen theorem. Concretely we are looking for a *Lipschitz* martingale transport, i.e. some $\pi \in \Pi_M(\mu, \nu)$ such that $x \in \mathbb{R} \to \pi_{x,\cdot} \in (\mathcal{P}(\mathbb{R}), W)$ is Lipschitz on a set of full μ measure. Recall Remark 1.2 for equivalent formulations. As we will see here, the sunset coupling $\pi_{\mathrm{sun}} \in \Pi_M(\mu, \nu)$ is such an element. In Theorem 4.5 we will moreover characterize it as the unique Lipschitz martingale coupling satisfying some additional conditions.

Both the sunset coupling and the coupling (abstractly) defined by Kellerer for his proof of Theorem 1.3 come together with an integral representation of elements that are extreme in the sense of Choquet's theory. In the next paragraph we provide a reminder on this theory and describe Kellerer's proof of Theorem 1.3. Paragraph 4.2 shows how Theorem 1.1 entails such a representation. In §4.3 we comment on the uniqueness of this representation, see Theorem 4.5.

4.1. Kellerer's proof ([35, 36]) of Theorem 1.3.

Let us present the Choquet representation yielding a Lipschitz element in $\Pi_M(\mu, \nu)$. It is based on the integral representation of the elements of the convex set $E(\mu) = \{\eta \in \mathcal{P} : \mu \preceq_C \eta\}$. Kellerer establishes in [36, Theorem 1] that for a given $\mu \in \mathcal{P}$ the extreme points of $E(\mu)$ exactly are the measures μP_T where P_T is the Kellerer dilation for a set $T \in \mathcal{I}$ defined in Definition 2.6.

Recall that if T is not an element of $\mathcal{I}$ but $\mathrm{supp}(\mu) \subseteq [\inf T, \sup T]$, the kernel P_T still makes sense μ-almost surely. Note that if $\mu \preceq_C \nu$, this remark applies to the set $T = \mathrm{supp}(\nu)$ since $\mathrm{supp}(\mu) \subseteq [\inf T, \sup T]$. The Kellerer dilation possesses important properties.

Proposition 4.1. *Let μ be an element of $\mathcal{P}$ and $T \in \mathcal{F}(\mathbb{R})$ satisfy* $\mathrm{supp}(\mu) \subseteq [\inf T, \sup T]$. *Then $W(P_T(x,\cdot), P_T(x',\cdot)) \leq |x - x'|$ for every x, x' in* $\mathrm{supp}(\mu)$. *Moreover the hitting coupling $\mu(\mathrm{id} \times P_T)$ is the unique element of $\Pi_M(\mu, \mu P_T)$.*

Proof. Part 1. is on the Lipschitz property and 2. on the uniqueness.

1. This is for instance explained in [6, §3.1]. Briefly, recall that for two probability measures κ and κ' of $\mathcal{P}$, we have $W(\kappa, \kappa') = \int_{[0,1]} |G_{\kappa'} - G_\kappa|$. However, as the law of G_κ on $([0,1], \lambda)$ is κ, the integral $\int_{[0,1]} G_\kappa$ is the expectation $\int x \, d\kappa(x)$. The same holds

for κ'. Therefore the condition $G_\kappa \leq G_{\kappa'}$ (equivalent to $\kappa \leq_{\mathrm{sto}} \kappa'$) yields $W(\kappa,\kappa') = \left|\int_{[0,1]} G_{\kappa'} - \int_{[0,1]} G_\kappa\right|$ where we usually only have an inequality. This can be applied to $\kappa = P_T(x,\cdot)$ and $\kappa' = P_T(x',\cdot)$ as soon as $x \leq x'$ because these measures clearly are in stochastic order.

2. This is [36, Satz 25]. Alternatively, one may consider the canonical decomposition of (μ,ν) into irreducible components described in [11, Theorem 8.4] that we recall now. The set $\{u_\mu < u_\nu\}$ is open and hence consists of a (finite or countable) union of open intervals $(I_n)_{n\geq 1}$. We denote by I_0 the closed set $\{u_\mu = u_\nu\}$. Write $\mu_n = \mu|_{I_n}$ and $\mu = \mu_0 + \sum_{n\geq 1}\mu_n$. By [11, Theorem 8.4] we can write $\nu = \nu_0 + \sum_{n\geq 1}\nu_n$ where $\nu_0 = \mu_0$, for $n \geq 1$, the measures ν_n are concentrated on $\bar{I}_n$ and any $\pi \in \Pi_M(\mu,\nu)$ can be decomposed as $\pi = (\mathrm{Id}\times\mathrm{Id})_\#(\mu_0) + \sum \pi_n$ with $\pi_n \in \Pi_M(\mu_n,\nu_n)$, $n \geq 0$.

In the present situation, from the definitions of u_μ, u_ν and P_T, the potential functions are the same on T, i.e. $T \subseteq I_0$. If I is an open connected interval of $\mathbb{R}\setminus T$, we have moreover $\nu(I) = (\mu P_T)(I) = 0$ so that u_ν is affine on each interval $\bar{I}$ (recall that its second derivative is 2ν). As u_μ is convex on I we have $u_\mu < u_\nu$ if $\mu(I) > 0$ or $u_\mu = u_\nu$ if $\mu(I) = \nu(I) = 0$. Hence, the connected intervals $(I_n)_{n\geq 1}$ of $\{u_\mu < u_\nu\}$ are among the connected intervals I of $\mathbb{R}\setminus T$. We obtain that $\nu_n(I_n) = 0$ and thus ν_n is the atomic measure concentrated on ∂I_n with $\mu_n \leq_C \nu_n$. This implies that necessarily $\pi_n = \mu_n(\mathrm{id}\times P_T)$ and hence $\pi = \mu(\mathrm{id}\times P_T)$. □

We can now state the integral representation established by Kellerer [36, Theorem 1] together with the Lipschitz property. Recall that Theorem 4.2 entails Theorem 1.3.

Theorem 4.2 (A Choquet representation established by Kellerer)**.** *Let μ and $\nu \in \mathcal{P}$ satisfy $\mu \leq_C \nu$. Then there exists a probability measure χ on $\mathcal{I}$ such that $\int_{\mathcal{I}} d(0,T)\,\mathrm{d}\chi(T) < \infty$ and*

$$\nu = \int_{\mathcal{I}} \mu P_T \,\mathrm{d}\chi(T) =: \mu P_\chi =: \mu_\chi. \tag{9}$$

Moreover P_χ is a Lipschitz kernel (and $\mu(\mathrm{Id}\times P_\chi) \in \Pi_M(\mu,\nu)$ is a Lipschitz martingale transport plan).

Scheme of the proof by Kellerer. 1. Kellerer establishes that for a given $\mu \in \mathcal{P}$ the extreme points of $E(\mu) = \{\eta \in \mathcal{P} : \mu \leq_C \eta\}$ exactly are the measures μP_T (see [36, §3.1, 3.2]). The latter set is not compact but going first through the spaces $\{\nu \in \mathcal{P},\ \mu \leq_C \eta \leq_C \mu P_S\}$ that are compact and convex for every set $S \in \mathcal{I}$, Kellerer is able to derive a Choquet representation in [36, Theorem 4] so that any $\nu \in E(\mu)$ can be represented in the form μ_χ.

2. The Lipschitz estimate is based on the estimate in Proposition 4.1 and the inequality

$$W\left(\int_{\mathcal{I}} P_T(x) d\chi(T), \int_{\mathcal{I}} P_T(x') d\chi(T)\right) \leq \int_{\mathcal{I}} W(P_T(x), P_T(x')) d\chi(T)$$

that can be obtained as a simple consequence of the formula $W(\kappa,\kappa') = \int_{\mathbb{R}} |F_{\kappa'} - F_\kappa|$. In Kellerer's paper [35] this estimate is established in Satz 20 (for the similar case of submartingale couplings). □

In [36], uniqueness of a measure on the extreme elements of $E(\mu)$ is not claimed and can easily be disproved. Set for instance $\mu = \delta_0$ and $\nu = \frac{\delta_{-2}+\delta_{-1}+\delta_1+\delta_2}{4}$. Taking the probability measure $\chi = 2^{-1}(\delta_{T_1} + \delta_{T_2})$ on $T_1 = \mathbb{R}\setminus]-1,1[$ and $T_2 = \mathbb{R}\setminus]-2,2[$ on the one hand and $\chi' = 8^{-1}(3\delta_{T_1'} + 3\delta_{T_2'} + 2\delta_{T_3'})$ on $T_1' = \mathbb{R}\setminus]-1,2[$, $T_2' = \mathbb{R}\setminus]-2,1[$ and $T_3' = \mathbb{R}\setminus]-2,2[$ on the other hand we obtain two different representations of $\nu = \mu_\chi = \mu_{\chi'}$. Another, more trivial, type of non-uniqueness can be observed: in the previous example T_1 can also be replaced for instance by $\mathbb{R}\setminus(]-1,1[\cup]-20,-15[)$ providing the same measure on $\mathcal{P}$ but another measure on $\mathcal{I}$.

4.2. **Our proof of Theorem 1.3 and Theorem 4.2 using π_{sun}.** Let us now discuss how $\pi_{\text{sun}} \in \Pi_M(\mu, \nu)$ compares to Kellerer's proof of the existence of Lipschitz kernels. In fact according to the barrier characterization in Theorem 1.1, we have

$$\pi_{\text{sun}} = \int_0^1 \hat{\pi}_{u,\cdot,\cdot}\, \mathrm{d}u.$$

Here $\hat{\pi}_{u,\cdot,\cdot} = \mathrm{Law}((B_0, B_\tau) \mid U = u)$ where U is uniform on $[0,1]$ and (B_t) is a Brownian motion with starting distribution $\mu_{u,\cdot}$, that is identically $\mu_{u,\cdot} = \mu$ for the sunset coupling, and τ (conditioned on $U = u$) is the hitting time of the barrier's vertical section $R_u := \{y \in \mathbb{R} : (u, y) \in R\}$, see Figure 2. Therefore, considering the second marginal only we obtain

$$\nu = \int_0^1 \mu P_{R_u}\, \mathrm{d}u. \tag{10}$$

This corresponds to (9) in Theorem 4.2 for which we have therefore obtained a new proof (and a fortiori for Theorem 1.3). Note that π_{sun} is systematically constructed from μ and ν, which was not the case for $\mu(\mathrm{Id} \times P_\chi)$ in the proof by Kellerer. In the next subsection we refine this remark with a uniqueness statement, see Theorem 4.5.

Remark 4.3. According to Theorem 1.1 (4), for every $u \in [0,1]$ the transport plan $\hat{\pi}_{u,\cdot,\cdot}$ transfers $\hat{\mu}_{u,\cdot} = \mu$ onto the right derivative of $u \mapsto \hat{\nu}_{[0,u],\cdot} = S^\nu(\mu_{[0,u],\cdot})$. The existence of such a construction is already clear from Theorem 2.9 and Proposition 2.7 where see that that this transport is given through the *hitting projection* of μ onto $\mathrm{supp}(\nu - \hat{\nu}_{[0,u],\cdot})$, the support of $\nu - \hat{\nu}_{[0,u],\cdot}$. Note however that the corresponding uniqueness result of Theorem 4.5 will really rely on Theorem 1.1, not on Theorem 2.9 only.

Remark 4.4. After Remark 4.3, a systematic choice for R_u is the support of $\nu - \hat{\nu}_{[0,u],\cdot}$. The fact that R_u may not be an element of $\mathcal{I}$ as it should be according to Theorem 4.2 can easily repaired by replacing it by $(-\infty, \min R_u] \cup R_u \cup [\max R_u, \infty)$. Comparing (10) with Theorem 4.2 again It is not evident that $\int_0^1 d(0, R_u)\, du < \infty$. However, according to Lemma 15 in [36], this holds if and only if ν has finite first moments, i.e. is an element of $\mathcal{P}$.

4.3. **Uniqueness statement beyond Theorems 1.3 and 4.2.** We stress that the uniqueness in Theorem 1.1 permits us to guarantee that there exists a *unique* Choquet representation through a family $(R_u)_{u \in [0,1]}$ that is *ordered* in the sense that $u \leq v$ implies that $R_u \supseteq R_v$. Indeed, the vertical sections of a barrier defined as in Theorem 1.1 (2) are ordered in this sense (cf. Figure 3).

If some Choquet representation of ν is given by measures obtained using the hitting projection on sets R_u that are ordered in the above sense, then these sets constitute a barrier and based on the uniqueness assertion in Theorem 1.1, the family $(\hat{\nu}_{u,\cdot})_{u \in [0,1]}$ with $\nu = \mu P_{R_u}$ is the one associated to the sunset coupling.

While until §4.2 we only used results contained in Theorem 2.9, for the first time in this paper we now make use of the full strength our main theorem (whose proof is given in the next section). In particular we not only assume that the sunset coupling exists but also that it is uniquely determined by the properties of barriers.

Theorem 4.5 (Sunset coupling and Lipschitz kernel)**.** *Let $\mu, \nu \in \mathcal{P}$ satisfy $\mu \preceq_C \nu$. Then there exists a probability measure χ on $\mathcal{I}$ such that $\int_{\mathcal{I}} d(0,T)\, \mathrm{d}\chi(T) < \infty$ and*

$$\nu(A) = \int_{\mathcal{I}} \mu P_T(A)\, \mathrm{d}\chi(T) =: \mu P_\chi(A) =: \mu_\chi(A). \tag{11}$$

Moreover P_χ is a Lipschitz kernel (and $\mu(\mathrm{Id} \times P_\chi) \in \Pi_M(\mu, \nu)$ is a Lipschitz martingale transport plan).

A possible choice of χ is the uniform measure on $(R_u)_{u\in[0,1]}$ where R_u is a decreasing family for $\subseteq$. Moreover if χ' is another measure associated to a decreasing family $(R'_u)_{u\in[0,1]}$ with $\nu = \mu P_{\chi'}$ and $\mu(\mathrm{Id}\times P_{\chi'}) \in \Pi_M(\mu,\nu)$, then $\mu(\mathrm{Id}\times P_\chi) = \mu(\mathrm{Id}\times P_{\chi'})$. (This measure is the sunset coupling of μ and ν, that is the shadow coupling of μ with source $\hat\mu = \lambda\times\mu$ and target ν.)

Hence we have obtained an advanced version of Theorem 1.3 both related to the original proof by Kellerer and the Skorokhod embedding problem. Since other proofs that $\Pi_M(\mu,\nu)$ contains at least one Lipschitz kernel are given by solutions of the Skorokhod embedding problem –Root's embedding ([44]) is considered in [10] and Hobson's embedding ([25]) in [37, Lemma 3.3]– with Theorem 1.1 we finally spans a bridge between the two methods of using Choquet's theorem and applying Skorokhod embedding techniques, respectively.

5. Proof of Theorem 1.1

Throughout this section we assume that μ,ν are in convex order and that $\hat\mu \in \Pi(\lambda,\mu)$. Given a measurable $c : [0,1]\times\mathbb{R}\times\mathbb{R} \to \mathbb{R}_+$ we consider the optimization problem

$$P := P_c := \inf\left\{\int c\,\mathrm{d}\hat\pi : \hat\pi \in \hat\Pi_M(\hat\mu,\nu)\right\}. \tag{12}$$

Proposition 5.1. *Assume that $\hat\pi$ is the lifted shadow coupling corresponding to a curve $(\hat\pi_{[0,u],\cdot})_{u\in[0,1]}$ as in Theorem 1.1 (4). Then for all $p\in[0,1], q\in\mathbb{R}$, $\hat\pi$ is an optimizer of* (12) *for $c_{p,q}(u,x,y) = \mathbb{1}_{u\le p}|y-q|$.*

Conversely if $\hat\gamma \in \hat\Pi_M(\hat\mu,\nu)$ is a minimizer for every $c_{p,q}$ where $(p,q)\in[0,1]\times\mathbb{R}$, then $\hat\gamma$ is the lifted shadow coupling from μ to ν with source $\hat\mu$.

Proof. Let $\hat\pi \in \hat\Pi_M(\hat\mu,\nu)$ be a shadow coupling and (p,q) as in the statement. Then $\int c_{p,q}\,\mathrm{d}\hat\pi = \int |y-q|\,\mathrm{d}\hat\nu_{[0,p],\cdot}(y)$ where we recall $\hat\nu_{[0,p],\cdot} = S^\nu(\hat\mu_{[0,p],\cdot})$. More generally if $\hat\gamma$ is an element of $\hat\Pi_M(\hat\mu,\nu)$ we have $\int c_{p,q}\mathrm{d}\hat\gamma = \int |y-q|\,\mathrm{d}\beta^p(y)$ where $\hat\mu_{[0,p],\cdot} \preceq_C \beta^p$ and $\beta^p \preceq_+ \nu$ (in fact $\beta^p := (\mathrm{proj}_y)_\# \gamma|_{[0,p]\times\mathbb{R}\times\mathbb{R}}$)). Therefore $\hat\nu_{[0,q],\cdot} \preceq_C \beta^p$ and as $y\mapsto|y-q|$ is convex, we have proved that $\hat\pi$ is a minimizer.

If $\hat\gamma \in \hat\Pi_M(\hat\mu,\nu)$ is a minimizer for $c_{p,q}$ for any $p\in[0,1]$, then the measures β^p and $\hat\nu_{[0,p],\cdot}$ have the same potential function. Thus, they are equal and the curve $p \mapsto (\mathrm{proj}_y)_\#\hat\gamma|_{[0,p]\times\mathbb{R}\times\mathbb{R}})$ is completely determined. As also noticed in Theorem 1.1, since $\hat\mu$ is known such couplings are uniquely determined, so that $\hat\gamma = \hat\pi$. ◻

Recall from Theorem 1.1 that a set $\hat\Gamma \subseteq [0,1]\times\mathbb{R}\times\mathbb{R}$ is called *monotone* if for all u,v,x,x',y^-,y^+,y' such that $u<v, (u,x,y^-),(u,x,y^+),(v,x',y')\in\Gamma$ it holds $y'\notin]y^-,y^+[$.

Proposition 5.2. *Assume that $\hat\pi \in \Pi_M(\mu,\nu)$ satisfies one of the following assumptions:*

(1) For all $p\in[0,1], q\in\mathbb{R}$, $\hat\pi$ is an optimizer of (12) *for $c_{p,q}(u,x,y) = \mathbb{1}_{u\le p}|y-q|$.*

(2) $\hat\pi$ is an optimizer of (12) *for $c(u,x,y) = (1-u)\sqrt{1+y^2}$.*

Then there is a monotone set $\hat\Gamma$ such that $\hat\pi(\hat\Gamma) = 1$.

Proof. We will establish the assertion under the first assumption, the argument based on the second assumption is very similar.

Using the notation and the monotonicity principle from [7], we pick for each $(p,q) \in ([0,1]\times\mathbb{R})\cap\mathbb{Q}^2$ a monotoncity set $\Gamma_{(p,q)}$ for the cost function and set

$$\Gamma := \bigcap_{(p,q)\in[0,1]\times\mathbb{R}\cap\mathbb{Q}^2} \Gamma_{(p,q)}.$$

Assume for contradiction that there exist $s, t, x, x', y^-, y^+, y'$ such that

$$s < t, \text{ and } a := (s, x, y^-),\ b := (s, x, y^+),\ c := (t, x', y') \in \Gamma, \text{ and } y' \in]y^-, y^+[.$$

Pick λ such that $y' = (1-\lambda)y^- + \lambda y^+$, $p \in]s, t[$, and q very close to y' (in comparison to y^-, y^+). Set

$$a' := (t, x', y^-),\ b' := (t, x', y^+),\ c' := (s, x, y').$$

Then

$$\alpha := (1-\lambda)\delta_a + \lambda\delta_b + \delta_c \text{ and } \alpha' := (1-\lambda)\delta_{a'} + \lambda\delta_{b'} + \delta_{c'}$$

are competitors with $\operatorname{supp}\alpha \subseteq \hat\Gamma$ and $\int c_{p,q}\,\mathrm{d}\alpha > \int c_{p,q}\,\mathrm{d}\alpha'$, contradiction. □

In the next result we establish that any $\hat\pi$ which is monotone admits a barrier representation as in Theorem 1.1 (2).

Proposition 5.3. *Let $\hat\pi \in \Pi_M(\hat\mu, \nu)$ be a transport plan concentrated on a monotone set $\hat\Gamma$. Define barriers*

$$R_o := \{(s, y) \in [0,1]\times\mathbb{R} : \exists t > s, (t, y) \in \hat\Gamma\} \tag{13}$$

$$R_c := \{(s, y) \in [0,1]\times\mathbb{R} : \exists t \ge s, (t, y) \in \hat\Gamma\}. \tag{14}$$

Consider a process $(Z_t)_{t\ge0} = (Z^1_t, Z^2_t)_{t\ge0} = (Z^1_0, Z^2_t)_{t\ge0}$ on some probability space which takes values in $[0,1]\times\mathbb{R}$ and is specified through

(1) $Z_0 \sim \hat\mu$,
(2) $Z_t = Z_0 + (0, B_t)$, where $(B_t)_t$ is (one dimensional) Brownian motion.

and write τ_o, τ_c for the first time Z hits R_o respectively R_c.

Then

- *$\tau_o = \tau_c$ a.s.*
- *It holds*

$$(Z_0, Z_{\tau_o}) \sim (Z_0, Z_{\tau_c}) \sim \hat\pi. \tag{15}$$

- *The martingales $t \mapsto Z_{t\wedge\tau}$ and $t \mapsto B_{t\wedge\tau}$ are uniformly integrable.*
- *There exist Borel maps $T_{up}, T_{down} : [0,1]\times\mathbb{R} \to \mathbb{R}, T_{down}(x) \le x \le T_{up}(x)$ such that*

$$\hat\pi\{(u, x, T_i(x)) : i \in \{up, down\}, (u, x) \in [0,1]\times\mathbb{R}\} = 1.$$

Proof. Fix a disintegration $(\pi_{u,x})$ of $\hat\pi$ with respect to $\hat\mu$ and write $\Gamma_{u,x}$ for the section of $\hat\Gamma$ in (u, x). Then $\hat\mu(\Gamma_0) = 1$, where

$$\Gamma_0 = \left\{(u, x) : \pi_{u,x}(\Gamma_{u,x}) = 1, \int |y|\,\mathrm{d}\pi_{u,x} < \infty, \int y\,\mathrm{d}\pi_{u,x} = x\right\}.$$

Define for each $(u, x) \in [0,1]\times\mathbb{R}$, $\tau_{u,x}$ to be (say) the Azema-Yor solution of the Skorokhod embedding problem such that $B_{\tau_{u,x}} \sim \pi_{u,x}$. Then define a stopping time τ such that conditionally on $Z_0 = (u, x)$ we have $\tau = \tau_{u,x}$. It follows that $(Z_0, Z^2_\tau) \sim \hat\pi$ and that for all elements ω of a full measure set Ω_0 we have $(Z^1_0(\omega), Z^2_0(\omega), Z^2_\tau(\omega)) \in \hat\Gamma$. Next we claim that there exists a full measure subset Ω_1 of Ω_0 such that for all $\omega \in \Omega_1$ and every $t < \tau(\omega)$, the following assertion holds true:

Continuation Assertion on (ω, t). There are $\omega_i \in \Omega_0, i = 1, 2$ satisfying

(1) $t < \tau(\omega_i)$, $(Z_s(\omega))_{s\le t} = (Z_s(\omega_i))_{s\le t}$ for $i = 1, 2$,
(2) $Z^2_\tau(\omega_1) < Z^2_t(\omega) < Z^2_\tau(\omega_2)$.

Assume for contradiction that the set

$$\{(\omega,t) : t < \tau(\omega) \text{ and Continuation Assertion fails}\} =: D$$

is not evanescent, i.e. that $\mathrm{proj}_\Omega(D)$ does not have $\mathbb{P}$-measure 0. Set

$$D^- := \{(\omega,t) \in [\![0,\tau[\![: \omega_1 \in \Omega_0, t < \tau(\omega_1), (Z_s(\omega_1))_{s\le t} = (Z_s(\omega))_{s\le t} \Rightarrow Z_\tau(\omega_1) \le Z_t(\omega)\}$$
$$D^+ := \{(\omega,t) \in [\![0,\tau[\![: \omega_2 \in \Omega_0, r < \tau(\omega_2), (Z_s(\omega_2))_{s\le t} = (Z_s(\omega_2))_{s\le t} \Rightarrow Z_\tau(\omega_2) \ge Z_t(\omega)\}$$

such that $D = D^- \cup D^+$. If D is not evanescent, then by the optional section theorem there exists a stopping time σ such that $\mathbb{P}(\sigma < \infty) > 0$ and

$$\{(\omega,\sigma(\omega)) : \sigma(\omega) < \infty\} \subseteq D^- \text{ or } \{(\omega,\sigma(\omega)) : \sigma(\omega) < \infty\} \subseteq D^+.$$

Combined with the strong Markov property this leads to a contradiction with the optional stopping theorem.

We claim that on Ω_1

$$\tau_c \le \tau \le \tau_o. \tag{16}$$

Note that the first inequality is satisfied by definition of τ_c. To establish the second inequality we assume for contradiction that there exists $\omega \in \Omega_1$ such that $\tau_o(\omega) < \tau(\omega)$.

Then $t^* := \min\{t \ge 0 : Z_t(\omega) \in R_o\} < \tau(\omega)$. Set $y' := Z^2_{t^*}(\omega)$ and $(u,x) = (Z^1_0(\omega), Z^2_0(\omega))$. By definition of R_o, there exist $v > u$ and x' such that $(v,x',y') \in \hat\Gamma$. Pick $\omega_i, i = 1,2$ according to the Continuation Assertion. Setting $y_i = Z^2_\tau(\omega_i), i = 1,2$, we have $(u,x,y_i) \in \hat\Gamma$, contradiction.

By Lemma 5.4 $\tau_c = \tau_o$ almost surely hence (15) holds.

To see that $(Z_{t\wedge\tau})$ (respectively $(B_{t\wedge\tau})$) is uniformly integrable we recall a result of Monroe [39] which asserts that a solution τ of the Skorokhod problem is minimal (i.e. there is no strictly smaller solution) if and only if Brownian motion up to time τ is uniformly integrable. In the present context it is straightforward to verify that τ provides a minimal embedding of ν with respect to Z^2 (we refer to [9, Proposition 4.1] for complete details), hence $(Z^2_{t\wedge\tau})$ is uniformly integrable.

The rest is immediate. □

In the proof we used the following lemma from [6] (we include the proof for the convenience of the reader).

Lemma 5.4. *Let $\hat\mu$ be a probability measure on $\mathbb{R}^2$ such that the projection onto the horizontal axis $\mathrm{proj}_x \hat\mu$ is continuous (in the sense of not having atoms) and let $\phi : \mathbb{R} \to \mathbb{R}$ be a Borel function. Set*

$$R_o := \{(x,y) : x > \phi(y)\}, \quad R_c := \{(x,y) : x \ge \phi(y)\}.$$

Start a vertically moving Brownian motion in μ and define

$$\tau_o := \inf\{t : (x, y + B_t) \in R_o\}, \quad \tau_c := \inf\{t : (x, y + B_t) \in R_c\}.$$

Then $\tau_c = \tau_o$ almost surely.

Proof. Obviously $\tau_c \le \tau_o$.

We say that y is a local minimum of ϕ if $\phi(y') \ge \phi(y)$ for all y' in a neighborhood of y. Set

$$I := \{\phi(y) : y \text{ is a local minimum of } \phi\}.$$

It is then not difficult to prove (and certainly well known) that I is at most countable: assume by contradiction that there exist an uncountably family $A \subseteq \mathbb{R}$ and corresponding neighborhoods $]a-\varepsilon_{sun}, a+\varepsilon_a[, a \in A$ such that $\phi(x) \ge \phi(a)$ for $x \in]a-\varepsilon_a, a+\varepsilon_a[$ and $a \ne a'$

implies $f(a) \neq f(a')$. Passing to an uncountable subset of A, we can assume that there is some $\eta > 0$ such that $\varepsilon_a > \eta$ for all $a \in A$. For $a \neq a'$ we cannot have $|a - a'| < \eta$ for then $a \in]a' - \varepsilon_{a'}, a' + \varepsilon_{a'}[$ as well as $a' \in]a - \varepsilon_a, a + \varepsilon_a[$ which would imply that $f(a) = f(a')$. Hence $|a - a'| \geq \eta$ which implies that A is countable, giving a contradiction.

On the complement of $I \times \mathbb{R}$ we have almost surely

$$\tau_o = 0 \quad \Longleftrightarrow \quad \tau_c = 0 \tag{17}$$

as a consequence of the strong Markov property. □

We have thus obtained an interpretation of monotone transport plans in terms of a barrier-type solution to the Skorokhod problem. This interpretation is useful for us since it allows us to use a short argument of Loynes [38] (which in turn builds on Root [43]) to show that there is only one monotone transference plan.

Lemma 5.5 (cf. Loynes [38])**.** *Let $\hat{\pi}_1$, $\hat{\pi}_2$ be monotone transport plans in $\hat{\Pi}_M(\hat{\mu}, \nu)$, with corresponding maps $T^{\mathbb{P}} = (T^i_{up}, T^i_{down})$ and denote by $R^{\hat{\pi}_i}, i = 1, 2$ the corresponding 'closed' barriers as in Proposition 5.3. Then $\tau_{R^{\hat{\pi}_1}} = \tau_{R^{\hat{\pi}_2}}$, a.s.*

Proof. For a set $A \subseteq \mathbb{R}$, we abbreviate $R_i(A) := R^{\hat{\pi}_i} \cap (\mathbb{R} \times A)$ and $\tau_i = \tau_{R^{\hat{\pi}_i}}$ for $i = 1, 2$. Denote

$$K := \{y : m_1(y) > m_2(y)\} \quad \text{where} \quad m_i(y) := \sup\{m : (m, y) \in R^{\hat{\pi}_i}\},\ i = 1, 2. \tag{18}$$

Fix a trajectory $(Z_t)_t = (Z_t(\omega))_t$ such that $Z^2_{\tau_2} \in K$. Then $(Z_t)_t$ hits $R_2(K)$ before it enters $R_2(K^C)$. But then $(Z)_t$ also hits $R_1(K)$ before it enters $R_1(K^C)$. Hence

$$B_{\tau_2} \in K \quad \Longrightarrow \quad B_{\tau_1} \in K.$$

As both stopping times embed the same measure, this implication is an equivalence almost surely, and we may set $\Omega_K := \{B_{\tau_1} \in K\} = \{B_{\tau_2} \in K\}$. On Ω_K we have $\tau_1 \leq \tau_2$ while $\tau_1 \geq \tau_2$ on Ω^C_K. Then, for all Borel subset $A \subseteq \mathbb{R}$:

$$\mathbb{P}[B_{\tau_1 \wedge \tau_2} \in A] = \mathbb{P}[B_{\tau_1 \wedge \tau_2} \in A, \Omega_K] + \mathbb{P}[B_{\tau_1 \wedge \tau_2} \in A, \Omega^c_K] \tag{19}$$

$$= \mathbb{P}[B_{\tau_1} \in A, \Omega_K] + \mathbb{P}[B_{\tau_2} \in A, \Omega^c_K] \tag{20}$$

$$= \mathbb{P}[B_{\tau_1} \in A \cap K] + \mathbb{P}[B_{\tau_2} \in A \cap K^c] \tag{21}$$

$$= \mathbb{P}[B_{\tau_2} \in A \cap K] + \mathbb{P}[B_{\tau_2} \in A \cap K^c] \tag{22}$$

$$= \mathbb{P}[B_{\tau_2} \in A] \tag{23}$$

since $B_{\tau_i} \sim \nu$. Hence $\tau_1 \wedge \tau_2$ embeds ν. Similarly, we see that $\tau_1 \vee \tau_2$ also embeds ν. Since τ_1 and τ_2 are both minimal embeddings, we deduce that $\tau_1 \wedge \tau_2 = \tau_1$ as well as $\tau_1 \wedge \tau_2 = \tau_2$. □

Taking the results of this section we can now establish our main theorem.

Proof of Theorem 1.1. We have already seen in Theorem 2.9 that there exists $\hat{\pi} \in \hat{\Pi}_M(\hat{\mu}, \nu)$ satisfying Theorem 1.1 (4). By virtue of Propositions 5.1, 5.2 we have that $\hat{\pi}$ is monotone as required in 1.1 (3) and by Proposition 5.3 $\hat{\pi}$ admits a barrier type representation as in 1.1 (2). Moreover, by Lemma 5.4 we find that there exists a unique such $\hat{\pi}$.

Finally, by the standard compactness-continuity argument there exists $\hat{\pi}$ which solves the optimization problem in Theorem 1.1 (1), by Proposition 5.2 it is monotone and hence uniquely determined as before. □

Note that in the proof of Theorem 1.1, we did not use the uniqueness part in the statement of Theorem 2.9; rather we have obtained a second derivation of this uniqueness property based on Lemma 5.5.

We close this section with a remark on the implication of the above results for the curtain coupling.

Remark 5.6. We consider the curtain coupling π_{lc} corresponding to the case where the source $\hat{\mu}$ is given by the monotone rearrangement between Lebesgue measure and μ. Assume for simplicity that μ has no atoms such that the source $\hat{\mu}$ is concentrated on the graph of a 1-1 function elements of $\Pi_M(\mu,\nu)$ correspond in a 1-1 manner to elements of $\hat{\Pi}_M(\hat{\mu},\nu)$. It then follows from the respective optimality property of $\hat{\pi}$ that π_{lc} minimizes

$$\gamma \mapsto \int \phi(x)\psi(y)\,\mathrm{d}\gamma(x,y) \tag{24}$$

on the set $\Pi_M(\mu,\nu)$, where $\phi \geq 0$ is strictly decreasing and $\psi \geq 0$ is strictly convex and the minimum over $\Pi_M(\mu,\nu)$ is finite. Moreover, there exist a Borel set $S \subseteq \mathbb{R}$ and two measurable functions $T_1, T_2 : S \to \mathbb{R}$ such that

(1) π_{lc} is concentrated on the graphs of T_1 and T_2.
(2) For all $x \in \mathbb{R}$, $T_1(x) \leq x \leq T_2(x)$.
(3) For all $x < x' \in \mathbb{R}$, $T_2(x) < T_2(x')$ and $T_1(x') \notin]T_1(x), T_2(x)[$.

This recovers [11, Corollary 1.6].

6. The sunset coupling as a non-optimizer and shadow couplings as optimizers to general transport problems.

An important message of [11, 22] is that the left-curtain couplings are characterized as the optimizers to martingale optimal transport problems for a large class of cost functions. This goes together with the fact that the support of the left-curtain coupling is typically a very 'small' set – if μ is continuous, it is contained in the graphs of two functions.

In contrast, the sunset coupling typically has a 'large' support. Hence we do not expect it to solve a martingale transport problem except in trivial instances. This is underlined by the following simple example.

Example 6.1. Let μ, ν be measures in convex order such that

(1) $\mathrm{conv}(\mathrm{supp}(\mu)) \cap \mathrm{supp}(\nu) = \emptyset$
(2) μ, ν consist of finitely many atoms.

Assume c is such that the sunset coupling is optimal for the martingale transport problem. Then *all* elements of $\Pi_M(\mu,\nu)$ are optimal for the martingale transport problem.

Proof. We first note that $\mathrm{supp}(\pi_{\mathrm{sun}}) = \mathrm{supp}(\mu) \times \mathrm{supp}(\nu)$ under our assumptions on μ, ν.

In the present atomic case, the martingale transport problem can be formulated as a linear programming problem and it admits a natural dual problem for which strong duality holds. It follows from this that every martingale transport plan $\pi \in \Pi_M(\mu,\nu)$ satisfying $\mathrm{supp}(\pi) \subseteq \mathrm{supp}(\pi_{\mathrm{sun}})$ is optimal. □

For simplicity, we have stated Example 6.1 for discrete marginals but (with some work) it is not difficult to see that the same phenomenon carries over to more general cases.

However we find it interesting to note that shadow couplings possess optimality properties in a different sense:

The *general* optimal transport problem was introduced by Gozlan, Roberto, Samson and Tetali [20] having applications to geometric inequalities in mind and has immediately generated interest in several groups of researchers, see [17, 19, 1, 2]. As in the classical case, one optimizes over the set of transport plans $\pi \in \Pi(\mu, \nu)$, where μ, ν are probabilities on Polish spaces X, Y. In contrast to the classical case, more general cost functionals are considered. Writing $\mathcal{P}(Y)$ for the set of all probability measures on Y, a general cost is a function $C : X \times \mathcal{P}(Y) \to [0, \infty]$ and its associated transport costs are

$$T_C(\nu|\mu) := \inf_{\pi \in \Pi(\mu,\nu)} \int C(x, \pi_{x,\cdot})\, \mathrm{d}\mu(x), \tag{25}$$

where we use $(\pi_{x,\cdot})_{x\in\mathbb{R}}$ to denote disintegration with respect to μ.

The shadow couplings appear as optimizers to such general transport problems. E.g. we will see below that the sunset coupling is the unique optimizer for the general transport cost function

$$C(x, \bar\pi) := \inf_{\alpha \in \hat\Pi_M(\lambda \times \delta_x, \bar\pi)} \int (1-u)\sqrt{1+y^2}\, \mathrm{d}\alpha(u, x', y). \tag{26}$$

Here the function $(u, y) \mapsto (1-u)\sqrt{1+y^2}$ could be replaced by any function of the form $(u, y) \mapsto \phi(u)\psi(y)$, where ϕ is strictly increasing and ψ strictly convex and sufficiently integrable with respect to the given marginals. The cost function defined in (26) exhibits a relatively intuitive behavior: if $\bar\pi$ does not have center x the costs equal $+\infty$. If $\bar\pi$ is centered around x, the more $\bar\pi$ is spread out, the higher are the costs.

We note that the existence of optimizers for cost functions of the above type is guaranteed by abstract results for the theory of general optimal transport problems [4, 19].

Proposition 6.2. *Let μ and ν be in convex order and fix a disintegration $(\hat\mu_{\cdot,x})_x$ of $\hat\mu$ with respect to μ and set*

$$C^{\hat\mu}(x, \bar\pi) := \inf_{\alpha \in \hat\Pi_M(\hat\mu_{\cdot,x} \times \delta_x, \bar\pi)} \int (1-u)\sqrt{1+y^2}\, \mathrm{d}\alpha(u, x', y). \tag{27}$$

Then the shadow coupling associated to $\hat\mu$ is the unique optimizer of the general transport problem associated to $C^{\hat\mu}$.

Proof. Given $\hat\pi \in \hat\Pi_M(\hat\mu, \nu)$, write π for the corresponding martingale transport $\pi \in \Pi_M(\mu, \nu)$ and $(\pi_{x,\cdot})_{x\in\mathbb{R}}$ for its disintegration with respect to μ. We note that $\hat\pi$ can be μ-a.s. uniquely represented in the form

$$\hat\pi(A \times B \times C) = \int \mathrm{d}\mu(x) \int \mathrm{d}\alpha_x(u, x', z)\, \mathbb{1}_{A\times B\times C}(u, x, y), \tag{28}$$

where (α_x) is the (measurable) family with $\alpha_x \in \hat\Pi_M(\hat\mu_{\cdot,x} \times \delta_x, \bar\pi_x)$ and $(\mathrm{proj}_{u,y})_\#\alpha_x = \hat\pi_{\cdot,x,\cdot}$. We then find

(29)

$$\begin{aligned} \inf_{\hat\pi \in \hat\Pi_M(\hat\mu,\nu)} \int (1-u)\sqrt{1+y^2}\, \mathrm{d}\hat\pi &= \inf_{\pi\in\Pi_M(\mu,\nu)} \int \inf_{\alpha \in \hat\Pi_M(\hat\mu_{\cdot,x}\times\delta_x, \pi_{x,\cdot})} \left(\int (1-u)\sqrt{1+y^2}\, \mathrm{d}\alpha(u, x', y) \right) \mathrm{d}\mu(x) \\ &= \inf_{\pi\in\Pi(\mu,\nu)} \int C^{\hat\mu}(x, \pi_{x,\cdot})\, \mathrm{d}\mu(x). \end{aligned}$$

□

We note that the solution to the optimization problems (26) / (27) is straightforward to characterize in the non-trivial case where x is the barycenter of $\bar\pi$: The optimizer α is the unique element of $\hat\Pi_M(\hat\mu_{\cdot,x} \times \delta_x, \bar\pi)$ which is concentrated on the graphs of two functions $T^x_{\mathrm{up}} : [0, 1] \to [x, \infty)$, $T^x_{\mathrm{down}} : [0, 1] \to (-\infty, x]$, where T^x_{up} is increasing and T^x_{down} is decreasing.

In the particular case, where the source is the quantile coupling, we obtain that the left curtain coupling is concentrated on the graph of two functions $(x,u)\mapsto T^x_{\mathrm{up}}(x,u)$, $(x,u)\mapsto T^x_{\mathrm{down}}(x,u)$. We can thus recover the main result of Hobson and Norgilas [30].

7. Appendix: Notation

We gather here our notation and terminology so that the reader can refer to it when needed. This also appear at the appropriate place where it is introduced first.

7.1. **Probability and positive finite measures.** The letters $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ stand for probability measures on $\mathbb{R}$ and $\boldsymbol{\pi}$ for measures on $\mathbb{R}^2$ with marginals μ and ν. We denote the Lebesgue measure on $[0,1]$ by $\boldsymbol{\lambda}$. The lifted measures $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\nu}}$ and $\hat{\boldsymbol{\pi}}$ stand for probability measures with first marginal λ and second marginal μ, ν or π respectively. Note here that we call here the probability measure π on $\mathbb{R}^2$ the second marginal of $\hat{\pi}$, so that is in particular a probability measure on $\mathbb{R}^3$. Finally $\boldsymbol{\delta_x}$ stands for the Dirac measure in x.

We respect a special convention for conditional measures / disintegration / conditional laws linked together with a consistent habit concerning the use of the variables $\boldsymbol{u\in[0,1]}$, $\boldsymbol{x}\in\mathbb{R}$ and $\boldsymbol{y}\in\mathbb{R}$ for λ, μ and ν respectively. As an example $\boldsymbol{(\pi_{x,\cdot})_{x\in\mathbb{R}}}$ is a family of probability measures on $\mathbb{R}$ that is a disintegration of π with respect to μ. Similarly with $\boldsymbol{(\hat{\mu}_{u,\cdot})_{u\in[0,1]}}$, $\boldsymbol{(\hat{\mu}_{\cdot,x})_{x\in\mathbb{R}}}$ and $\boldsymbol{(\hat{\nu}_{\cdot,y})_{y\in\mathbb{R}}}$. The measures $\boldsymbol{(\hat{\pi}_{u,\cdot})_{u\in[0,1]}}$ are two dimensional so that we also use the notation $\boldsymbol{(\hat{\pi}_{u,\cdot,\cdot})_{u\in[0,1]}}$ depending on the context. We will also make use of $\boldsymbol{(\hat{\pi}_{u,x,\cdot})_{u,x\in[0,1]\times\mathbb{R}}}$ for the disintegration in conditional laws on $\mathbb{R}$ of $\hat{\pi}$ with respect to $\hat{\mu}$.

We will use integrated versions of the conditional measures (alias partial marginals) as for instance $\boldsymbol{\hat{\pi}_{[0,u],\cdot,\cdot}}$ or $\boldsymbol{\hat{\mu}_{[0,u],\cdot}}$, that are measures of mass u. For $u=1$ we will recover the full marginals $\pi=\hat{\pi}_{[0,1],\cdot}$ and $\mu=\hat{\mu}_{[0,1],\cdot}$.

7.2. **Spaces.** We denote by $\boldsymbol{\mathcal{P}(E)}$ and $\boldsymbol{\mathcal{M}(E)}$ the space of probability measures, or, respectively, of positive measures on a Polish space E. In the particular case $E=\mathbb{R}$ we simply write $\mathcal{P}$ and $\mathcal{M}$. We denote by $\boldsymbol{\Pi(\mu,\nu)}$ the space of measures with first marginal μ and second marginal ν. The set $\boldsymbol{\Pi_M(\mu,\nu)}$ is the subset of martingale transport plans as defined in §1.1. Those are two dimensional measures. We denote by $\boldsymbol{\hat{\Pi}_M(\hat{\mu},\nu)}$ the space of lifted transport plans as in §1.2. It is a subspace of $\mathcal{P}(\mathbb{R}^3)$.

7.3. **Orders on finite measures.** The usual order on positive measures is $\boldsymbol{\preceq_+}$. It is defined by $\mu\preceq_+\nu$ if and only if $\mu(A)\leq\nu(A)$ for every measurable set A, or equivalently if $\int f\mathrm{d}\mu\leq\int f\mathrm{d}\mu$ for every positive $f:\mathbb{R}\to\mathbb{R}^+$. The convex order $\boldsymbol{\preceq_C}$ is defined for positive measure of the same mass and finite first moment. We have $\mu\preceq_C\nu$ if and only if $\int f\mathrm{d}\mu\leq\int f\mathrm{d}\mu$ for every convex function $f:\mathbb{R}\to\mathbb{R}$. The stochastic order $\boldsymbol{\preceq_{\mathrm{sto}}}$ is defined for positive measures of the same mass. We have $\mu\preceq_{\mathrm{sto}}\nu$ if and only if $\int f\mathrm{d}\mu\leq\int f\mathrm{d}\mu$ for every increasing bounded function $f:\mathbb{R}\to\mathbb{R}$. We introduce also $\boldsymbol{\preceq_{C,\mathrm{sto}}}$ for positive measure of the same mass and finite first moment. We have $\mu\preceq_{\mathrm{sto}}\nu$ if and only if $\int f\mathrm{d}\mu\leq\int f\mathrm{d}\mu$ for every increasing convex function $f:\mathbb{R}\to\mathbb{R}$. Finally, see §3.1.3 for $\preceq_{DC}$.

7.4. **Operations on the measures.** We denote by $\boldsymbol{F_\theta}$ and $\boldsymbol{G_\theta}$ the cumulative distribution function and the quantile function of a one-dimensional probability measure θ, respectively. More precisely $F_\theta(x)=\theta((-\infty,x])$ and G_θ is the unique left continuous and increasing function such that $(G_\theta)_\#\lambda=\theta$. Recall that $G_\theta(u)=\inf_{x\in\mathbb{R}}\{F_\theta(x)\geq t\}$. We denote by $\alpha\times\beta$ the product measure of two measures. The restriction $\boldsymbol{\theta|_A}$ of θ to a set A is defined by $\theta|_A(\cdot)=\theta(A\cap\cdot)$ on the same measured space as θ. If f is mapping, the notation $\boldsymbol{f_\#\theta}$ stands for the pushed-forward measure $\theta(f^{-1}(\cdot))$. For instance $(\mathrm{proj}_{x,y})\#\hat{\pi}=\pi$. We denote by $\mathbf{supp}(\boldsymbol{\alpha})$ the support of the measure α, that is the smallest closed set F with $\alpha(\mathbb{R}\setminus F)=0$. In Subsection 2.4 we define the central notion of shadow of $\mu\in\mathcal{M}$ in

$\nu \in \mathcal{M}$, that we denote by $\boldsymbol{S^\nu(\mu)}$ (Note that, here, the measures must not be of mass one. However $\mu(\mathbb{R}) \leq \nu(\mathbb{R})$ because $\mu \preceq_{C,+} \nu$.)

7.5. **Random elements.** The letter $\boldsymbol{U}$ stands for a uniform random variable on $[0, 1]$. It has law λ and fits well with our notation $u \in [0, 1]$. We denote the standard Brownian motion by $\boldsymbol{(B_t)_{t\geq 0}}$. For random times τ_0 and τ_1 we denote by $\boldsymbol{\tau_0 \wedge \tau_1}$ for the minimum of the two. The random time interval $\boldsymbol{[\![0, \tau[\![}$, where τ is a positive random variable on a probability space Ω, is the set $\{(\omega, t) \in \Omega \times \mathbb{R}_+ : \ t < \tau(\omega))\}$.

7.6. **Others.** We denote the indicator functions by $\mathbb{1}$. For instance

$$\mathbb{1}_{u\leq p}(u, x, y) = \begin{cases} 1 & \text{if } u \leq p \\ 0 & \text{otherwise.} \end{cases}$$

References

[1] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems. *ArXiv e-prints*, Sept. 2017.

[2] J.-J. Alibert, G. Bouchitte, and T. Champion. A new class of cost for optimal transport planning. *hal-preprint*, 2018.

[3] J. Azéma and M. Yor. Une solution simple au problème de Skorokhod. In *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)*, volume 721 of *Lecture Notes in Math.*, pages 90–115. Springer, Berlin, 1979.

[4] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calc. Var. Partial Differential Equations*, 58(6):Paper No. 203, 28, 2019.

[5] M. Beiglböck, A. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. *Invent. Math.*, 208(2):327–400, 2017.

[6] M. Beiglböck, A. M. G. Cox, and M. Huesmann. The geometry of multi-marginal Skorokhod Embedding. *Probab. Theory Related Fields*, 176(3-4):1045–1096, 2020.

[7] M. Beiglböck and C. Griessler. A land of monotone plenty. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 19(1):109–127, 2019.

[8] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: A mass transport approach. *Finance Stoch.*, 17(3):477–501, 2013.

[9] M. Beiglböck, P. Henry-Labordère, and N. Touzi. Monotone martingale transport plans and Skorokhod embedding. *Stochastic Process. Appl.*, 127(9):3005–3013, 2017.

[10] M. Beiglböck, M. Huesmann, and F. Stebegg. Root to Kellerer. *Séminaire de Probabilités XLVIII*, pages 1–12, 2016.

[11] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, 44(1):42–106, 2016.

[12] M. Beiglböck, M. Nutz, and N. Touzi. Complete duality for martingale optimal transport on the line. *Ann. Probab.*, 45(5):3038–3074, 2017.

[13] B. Bouchard and M. Nutz. Arbitrage and duality in nondominated discrete-time models. *The Annals of Applied Probability*, 25(2):823–859, 2015.

[14] M. Brückerhoff, M. Huesmann, and N. Juillet. Shadow martingales – a stochastic mass transport approach to the peacock problem, 2020.

[15] L. Campi, I. Laachir, and C. Martini. Change of numeraire in the two-marginals martingale transport problem. *Finance Stoch.*, 21(2):471–486, June 2017.

[16] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Relat. Fields*, 160(1-2):391–427, 2014.

[17] M. Fathi and Y. Shu. Curvature and transport inequalities for Markov chains in discrete spaces. *Bernoulli*, 24(1):672–698, 2018.

[18] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.*, 24(1):312–336, 2014.

[19] N. Gozlan and N. Juillet. On a mixture of Brenier and Strassen theorems. *Proc. Lond. Math. Soc. (3)*, 120(3):434–463, 2020.

[20] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, 273(11):3327–3405, 2017.

[21] P. Henry-Labordère, X. Tan, and N. Touzi. An explicit martingale version of the one-dimensional Brenier's theorem with full marginals constraint. *Stochastic Process. Appl.*, 126(9):2800–2834, 2016.
[22] P. Henry-Labordère and N. Touzi. An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.*, 20(3):635–668, 2016.
[23] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. *Peacocks and associated martingales, with explicit constructions*, volume 3 of *Bocconi & Springer Series*. Springer, Milan, 2011.
[24] F. Hirsch and B. Roynette. A new proof of Kellerer's theorem. *ESAIM Probab. Stat.*, 16:48–60, 2012.
[25] D. Hobson. The maximum maximum of a martingale. In *Séminaire de Probabilités, XXXII*, volume 1686 of *Lecture Notes in Math.*, pages 250–263. Springer, Berlin, 1998.
[26] D. Hobson. The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, pages 267–318. Springer, Berlin, 2011.
[27] D. Hobson and M. Klimmek. Robust price bounds for the forward starting straddle. *Finance Stoch.*, 9(1):189–214, Apr. 2015.
[28] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Math. Finance*, 22(1):31–56, 2012.
[29] D. Hobson and D. Norgilas. Robust bounds for the American put. *Finance Stoch.*, 23(2):359–395, 2019.
[30] D. G. Hobson and D. Norgilas. The left-curtain martingale coupling in the presence of atoms. *Ann. Appl. Probab., to appear*, Feb. 2019.
[31] B. Jourdain and W. Margheriti. A new family of one dimensional martingale couplings. *arXiv:1808.01390*, 2019.
[32] N. Juillet. Peacocks parametrised by a partially ordered set. In *Séminaire de Probabilités XLVIII*, volume 2168 of *Lecture Notes in Math.*, pages 13–32. Springer, Cham, 2016.
[33] N. Juillet. Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1823–1843, 2016.
[34] N. Juillet. Martingales associated to peacocks using the curtain coupling. *Electron. J. Probab.*, 23:Paper No. 8, 29, 2018.
[35] H. Kellerer. Markov-Komposition und eine Anwendung auf Martingale. *Math. Ann.*, 198:99–122, 1972.
[36] H. G. Kellerer. Integraldarstellung von Dilationen. In *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Tech. Univ., Prague, 1971; dedicated to the memory of Antonín Špaček)*, pages 341–374. Academia, Prague, 1973.
[37] G. Lowther. Fitting Martingales To Given Marginals. *ArXiv e-prints*, Aug. 2008.
[38] R. M. Loynes. Stopping times on Brownian motion: Some properties of Root's construction. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 16:211–218, 1970.
[39] I. Monroe. On embedding right continuous martingales in Brownian motion. *Ann. Math. Statist.*, 43:1293–1311, 1972.
[40] M. Nutz and F. Stebegg. Canonical supermartingale couplings. *Ann. Probab.*, 46(6):3351–3398, 2018.
[41] M. Nutz, F. Stebegg, and X. Tan. Multiperiod martingale transport. *Stochastic Process. Appl.*, 130(3):1568–1615, 2020.
[42] J. Obłój. The Skorokhod embedding problem and its offspring. *Probab. Surv.*, 1:321–390, 2004.
[43] R. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific J. Math.*, 17:497–510, 1966.
[44] D. H. Root. The existence of certain stopping times on Brownian motion. *Ann. Math. Statist.*, 40:715–718, 1969.
[45] H. Rost. The stopping distributions of a Markov Process. *Invent. Math.*, 14:1–16, 1971.
[46] F. Stebegg. Model-Independent Pricing of Asian Options via Optimal Martingale Transport. *ArXiv e-prints*, Dec. 2014.
[47] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
[48] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, 2009.