

ON THE MARTINGALE SCHRÖDINGER BRIDGE BETWEEN TWO DISTRIBUTIONS

MARCEL NUTZ AND JOHANNES WIESEL

ABSTRACT. We study a martingale Schrödinger bridge problem: given two probability distributions, find their martingale coupling with minimal relative entropy. Our main result provides Schrödinger potentials for this coupling. Namely, under certain conditions, the log-density of the optimal coupling is given by a triplet of real functions representing the marginal and martingale constraints. The potentials are also described as the solution of a dual problem.

1. INTRODUCTION

The Schrödinger bridge problem originates in a question of Schrödinger [31] about the most likely evolution of a particle system from a distribution μ into another distribution ν . This dynamic problem can be reduced to the *static* Schrödinger bridge problem

$$(1) \quad \inf_{\pi \in \Pi(\mu, \nu)} H(\pi | R)$$

where $H(\cdot | R)$ denotes relative entropy with respect to the reference measure R and $\Pi(\mu, \nu)$ denotes the set of couplings; i.e., joint distributions with marginals μ and ν . See [12, 22] for surveys. The problem (1) includes the entropically regularized optimal transport (EOT) problem

$$(2) \quad \inf_{\pi \in \Pi(\mu, \nu)} \int c(x, y) \pi(dx, dy) + H(\pi | \mu \otimes \nu)$$

where c is a given cost function, by choosing $dR \propto e^{-c} d(\mu \otimes \nu)$. EOT is a regularized version of Monge–Kantorovich optimal transport. Thanks to its computational and statistical properties, EOT has been the subject of hundreds of works in the last few years (see, e.g., the monograph [28] for extensive references). A central result in EOT states that a coupling π is optimal for (2) if and only if its density is of the form

$$(3) \quad \frac{d\pi}{d(\mu \otimes \nu)}(x, y) = e^{f(x) + g(y) - c(x, y)}$$

for some functions f, g which are called the *Schrödinger potentials* and can also be seen as the solution of the dual problem of (2). Specifically, (f, g) exist in $L^1(\mu) \times L^1(\nu)$ if $c \in L^1(\mu \otimes \nu)$; see, e.g., [24]. Potentials are key objects for most aspects of EOT; we give just three examples: On the computational side, potentials are the output of the Sinkhorn–Knopp algorithm, the Gauss–Seidel iteration commonly used to compute EOT [28]. On the statistical side, sample complexity bounds of EOT are based on growth properties of

Date: January 10, 2024.

2020 Mathematics Subject Classification. 60G42, 90C05, 94A17.

Key words and phrases. Martingale Schrödinger bridge, Potentials, Duality.

MN acknowledges support by NSF Grants DMS-1812661, DMS-2106056.

JW acknowledges support by NSF Grant DMS-2345556.

the potentials [15]. In a mathematical application, regularity properties of the potentials are used to prove a generalization of Caffarelli's contraction theorem [6]. The rigorous construction of the potentials is surprisingly delicate and was achieved for general costs by [13, 29, 30]. Counterexamples in those references highlight some pitfalls in earlier results, related to integrability and convergence issues.

In this paper we study a *martingale Schrödinger bridge* problem, first introduced by [20] in financial mathematics. Financially, the martingale constraint for couplings corresponds to a risk-neutral pricing model that is calibrated to given marginals. The marginals, in turn, represent observed prices of call options. Starting with [3, 14, 21], the optimal transport problem with martingale constraint (*martingale optimal transport*) has been studied in great detail over the last decade; see, e.g., [4, 33] for extensive references. The martingale Schrödinger bridge problem can be seen as entropically regularized martingale optimal transport. Specifically, let μ, ν be distributions on \mathbb{R} with finite first moment. A coupling $\pi \in \Pi(\mu, \nu)$ is a martingale if its disintegration $\pi(dx, dy) = \mu(dx) \otimes \pi_x(dy)$ satisfies $\int (y - x) \pi_x(dy) = 0$ for μ -a.e. $x \in \mathbb{R}$. Let $\mathcal{M}(\mu, \nu)$ be the set of martingale couplings of μ and ν , then we study the minimization

$$(4) \quad \inf_{\pi \in \mathcal{M}(\mu, \nu)} H(\pi | \mu \otimes \nu).$$

This is connected to the EOT problem (2): one may suspect that (4) is equivalent to an EOT problem with a suitably chosen cost function c ; namely, a Lagrange multiplier for the martingale constraint.

Assuming that $\mathcal{M}(\mu, \nu) \neq \emptyset$, standard arguments show that (4) admits a unique optimizer π^* . Our main result (Theorem 2.4) is that, under certain conditions, this optimizer has a density of the form

$$(5) \quad \frac{d\pi^*}{d(\mu \otimes \nu)}(x, y) = e^{f(x) + g(y) - h(x)(y-x)}$$

for some functions $f \in L^1(\mu)$, $g \in L^1(\nu)$, and $h \in L^0(\mu)$ with $h(x)(y - x) \in L^1(\pi)$ for all $\pi \in \mathcal{M}_{\text{fin}}(\mu, \nu) := \{\pi \in \mathcal{M}(\mu, \nu) : H(\pi | \mu \otimes \nu) < \infty\}$. These functions are unique (up to shift by an affine function) and constitute the potentials of our martingale Schrödinger bridge problem. Like in (3), the functions (f, g) can be seen as Lagrange multipliers for the marginal constraints (μ, ν) . The additional term $h(x)(y - x)$ in (5) is a Lagrange multiplier for the martingale constraint. We see that (4) is formally equivalent to the EOT problem (2) with cost $c(x, y) = h(x)(y - x)$. To be precise, the equivalence holds if $h(x)(y - x)$ has good integrability. However we will see that this can fail: in Example 4.1, the positive part of that function is not integrable under $\mu \otimes \nu$, and then the cost $c(x, y) = h(x)(y - x)$ does not fall within the setting generally covered by EOT theory. That may give a hint that the potentials cannot be constructed by simply invoking a standard result from EOT.

Existence of the potentials in the aforementioned sense yields a strong duality theorem (Corollary 2.5). Indeed,

$$\begin{aligned} & \inf_{\pi \in \mathcal{M}(\mu, \nu)} H(\pi | \mu \otimes \nu) \\ &= \sup_{\tilde{f}, \tilde{g}, \tilde{h}} \int \tilde{f}(x) \mu(dx) + \int \tilde{g}(y) \nu(dy) - \log \int e^{\tilde{f}(x) + \tilde{g}(y) - \tilde{h}(x)(y-x)} \mu(dx) \nu(dy) \end{aligned}$$

where the supremum is taken over all $\tilde{f} \in L^1(\mu)$, $\tilde{g} \in L^1(\nu)$, and $\tilde{h} \in L^0(\mu)$ with $\tilde{h}(x)(y-x) \in L^1(\pi)$ for all $\pi \in \mathcal{M}_{\text{fin}}(\mu, \nu)$. Moreover, the dual supremum is attained: the maximizers are precisely the potentials (f, g, h) , leading to the formula $\inf_{\pi \in \mathcal{M}(\mu, \nu)} H(\pi | \mu \otimes \nu) = \int f(x) \mu(dx) + \int g(y) \nu(dy)$.

First statements about potentials for martingale Schrödinger bridges can be found in [20] for a dynamic setting, though those formal statements are not meant to be mathematically rigorous. Compared with the setting of classical Schrödinger bridges and EOT, the additional dual term $h(x)(y-x)$ due to the martingale constraint creates an interaction between the variables x, y which turns out to be a substantial difficulty. This difficulty was first highlighted by a counterexample of [1]. The example exhibits an L^p -convergent sequence $g_n(y) - h_n(x)(y-x)$ with nice functions g_n, h_n where the limit is not of the form $g(y) - h(x)(y-x)$ with an integrable g . The follow-up work [26] showed that this stems from a loss of integrability in the passage to the limit. The results closest to the present paper are obtained in [27] for a martingale Schrödinger bridge problem where only the second marginal ν is given (indeed, a pier rather than a bridge). As there is no constraint on the first marginal, the dual takes the form $g(y) - h(x)(y-x)$ without the function f as in (5). The approach in [27] rests on the flexibility of manipulating the support of the first marginal, a nonstarter in the present setting where μ is prescribed. A different type of result is obtained in [26] by algebraic considerations. It implies the algebraic form of the log-density of π^* , but it yields functions that a priori are not integrable (in general, possibly even not measurable). As far as we could see, the result of [26] does not help in establishing the present results.

Instead, we use a completely new approach. We first study an auxiliary problem where the delicate martingale constraint is replaced by a weaker constraint that is easier to handle. We show that the coupling $\hat{\pi}$ solving that relaxed problem has a density of the desired form (5). Second, we show that $\hat{\pi}$ is actually a martingale, which implies that $\hat{\pi} = \pi^*$ is also the solution of the martingale Schrödinger bridge problem. As a result, the density of π^* has the desired form. A more detailed outline of the approach can be found in Remark 3.1.

Continuing with the broadly related literature, similar potentials also play an important role in [18, 19] tackling the joint S&P 500/VIX calibration problem; here there are additional constraints pertaining to observing prices on VIX derivatives. The existence of integrable potentials is posited, but showing existence from first principles remains an open problem. To the best of our knowledge, the present result is the first to provide integrable potentials for a martingale Schrödinger bridge with more than one marginal. In a different direction, [8] develop a version of Sinkhorn's algorithm for martingale optimal transport using the martingale Schrödinger bridge as a regularized version of the optimal transport problem. See also [17] for a related algorithm using a different relaxation. In another related direction, [9, 10] study an entropic martingale optimal transport problem different from the present one. Instead of minimizing entropy over couplings, these works consider a martingale version of the transport problem proposed in [23]. In that problem, the marginal constraints μ, ν are relaxed into an entropic penalty, whereas the entropy of the coupling does not enter the picture.

The remainder of the paper is organized as follows. Section 2 details the problem formulation and states the main results. The proofs are reported in Section 3, while Section 4 concludes with an example illustrating a technical condition used in the main results.

2. MAIN RESULTS

We write $\mathcal{B}(\mathcal{X})$ and $\mathcal{P}(\mathcal{X})$ for the collections of Borel sets and probability measures on \mathcal{X} , respectively. Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ be probability measures on \mathbb{R} in convex order, denoted by $\mu \preceq_c \nu$ and defined by $\int \varphi(x) \mu(dx) \leq \int \varphi(x) \nu(dx)$ for any convex function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$. This implies that μ and ν have the same mean. Without loss of generality, we can then assume that the measures are centered,

$$(6) \quad \int x \mu(dx) = \int y \nu(dy) = 0.$$

Their set of couplings is defined by

$$\Pi(\mu, \nu) = \{\pi \in \mathcal{P}(\mathbb{R} \times \mathbb{R}) : \pi(A \times \mathbb{R}) = \mu(A), \pi(\mathbb{R} \times A) = \nu(A) \text{ for all } A \in \mathcal{B}(\mathbb{R})\}.$$

For $\pi \in \Pi(\mu, \nu)$, we denote by π_x the conditional probability distribution of π given x , meaning that the Markov kernel $(\pi_x)_{x \in \mathbb{R}}$ satisfies $\pi(A \times B) = \int_A \pi_x(B) \mu(dx)$ for all $A, B \in \mathcal{B}(\mathbb{R})$. The set of martingale couplings is defined by

$$\mathcal{M}(\mu, \nu) = \left\{ \pi \in \Pi(\mu, \nu) : \int (y - x) \pi_x(dy) = 0 \quad \mu\text{-a.s.} \right\}.$$

By Strassen's theorem [32], $\mu \preceq_c \nu$ is equivalent to $\mathcal{M}(\mu, \nu) \neq \emptyset$. If $\mu = \delta_0$, then $\mathcal{M}(\mu, \nu) = \{\mu \otimes \nu\}$ and all our results are trivial. Hence we may assume without loss of generality that $\mu \neq \delta_0$, which in view of $\mu \preceq_c \nu$ also implies that $\nu \neq \delta_0$.

We consider the martingale Schrödinger bridge problem

$$(7) \quad \inf_{\pi \in \mathcal{M}(\mu, \nu)} H(\pi | \mu \otimes \nu),$$

where H denotes the relative entropy

$$H(\pi | \mu \otimes \nu) = \begin{cases} \int \log(d\pi/d(\mu \otimes \nu)) d\pi & \text{if } \pi \ll \mu \otimes \nu, \\ \infty & \text{otherwise.} \end{cases}$$

For (7) to be well-posed, we clearly require some $\pi \in \mathcal{M}(\mu, \nu)$ with $H(\pi | \mu \otimes \nu) < \infty$. A stronger condition is necessary for our main result (8) below: if (8) holds, then the coupling $\pi^* \in \mathcal{M}(\mu, \nu)$ is equivalent to $\mu \otimes \nu$, denoted $\bar{\pi} \sim \mu \otimes \nu$. We state these necessary conditions as a standing assumption.

Assumption 2.1. *There exists $\bar{\pi} \in \mathcal{M}(\mu, \nu)$ such that*

$$\bar{\pi} \sim \mu \otimes \nu \quad \text{and} \quad H(\bar{\pi} | \mu \otimes \nu) < \infty.$$

The existence of a martingale coupling equivalent to $\mu \otimes \nu$ can be characterized in terms of the marginals μ and ν as follows.

Remark 2.2. Define the potential function φ_ρ of $\rho \in \mathcal{P}(\mathbb{R})$ by $\varphi_\rho(x) = \int |x - y| \rho(dy)$. Then $\varphi_\mu \leq \varphi_\nu$ due to $\mu \preceq_c \nu$. Existence of a martingale coupling $\bar{\pi} \sim \mu \otimes \nu$ is equivalent to a *strict* inequality on an open interval of full μ -measure. This condition is known in martingale optimal transport theory as irreducibility: $(\mu, \nu) \in \mathcal{P}(\mathbb{R}) \times \mathcal{P}(\mathbb{R})$ is called irreducible if the set $I := \{\varphi_\mu < \varphi_\nu\}$ is connected and satisfies $\mu(I) = 1$.

To see the equivalence, note that if (μ, ν) is irreducible, then [5, Lemma 3.3 and its proof] show that there exists $\bar{\pi} \in \mathcal{M}(\mu, \nu)$ with $\bar{\pi} \sim \mu \otimes \nu$. Conversely, if (μ, ν) is not irreducible, there exists $x \in \mathbb{R}$ such that $\varphi_\mu(x) = \varphi_\nu(x)$ and either $\mu((-\infty, x]) > 0$ and $\mu((x, \infty)) > 0$, or $\mu((-\infty, x)) > 0$ and $\mu([x, \infty)) > 0$. In the first case, the set $A := (-\infty, x] \times (x, \infty)$

satisfies $\pi(A) = 0$ for all $\pi \in \mathcal{M}(\mu, \nu)$ by [5, Theorem 3.2] but $(\mu \otimes \nu)(A) > 0$, so that $\bar{\pi}$ cannot exist. The second case is analogous with $A := (-\infty, x) \times [x, \infty)$.

Remark 2.2 shows that a strict inequality between the potential functions φ_μ, φ_ν is necessary for our main result to hold. A strict inequality means that the difference of the potential functions is uniformly bounded away from zero on compact subintervals of I . Our proof of the main result uses a slightly stronger condition. We require that either the difference is uniformly bounded away from zero on the whole interval I , or if the potentials touch at an endpoint of I , then at least they should meet at a nonzero angle (which means that ν has an atom at that point). This condition can be phrased succinctly as follows, where we write $\text{supp}(\mu)$ for the support of μ .

Assumption 2.3. *Let $s_- = \inf(\text{supp}(\mu))$ and $s_+ = \sup(\text{supp}(\mu))$ be the left and right endpoints of the support of μ . We assume that $\nu((-\infty, s_-]) > 0$ and $\nu([s_+, \infty)) > 0$.*

If we denote by t_\pm the endpoints of the support of ν , then Assumption 2.3 means that either $t_+ > s_+$, or $t_+ = s_+$ and $\nu(\{s_+\}) > 0$, and similarly for s_- . It follows that μ has bounded support, though this will not be used directly. In fact, Assumption 2.3 will be used only a single time, to show uniform integrability of a certain sequence in Step 2 of the proof of Lemma 3.16. We do not know if Theorem 2.4 holds without Assumption 2.3. Example 4.1 highlights that certain integrability issues arise when the potentials are not properly separated.

We can now state our main result.

Theorem 2.4 (Potentials). *Under Assumptions 2.1 and 2.3, there is a unique optimizer $\pi^* \in \mathcal{M}(\mu, \nu)$ of (7), and its density has the form*

$$(8) \quad \frac{d\pi^*}{d(\mu \otimes \nu)}(x, y) = e^{f(x) + g(y) - h(x)(y-x)}$$

for some functions $f \in L^1(\mu)$, $g \in L^1(\nu)$ and $h \in L^0(\mu)$ with $h(x)(y-x) \in L^1(\pi)$ for all $\pi \in \mathcal{M}_{\text{fin}}(\mu, \nu) := \{\pi \in \mathcal{M}(\mu, \nu) : H(\pi|\mu \otimes \nu) < \infty\}$. We call any such triplet potentials of (7).

The potentials are unique up to an affine shift. More precisely, (f', g', h') are potentials if and only if there are constants $\alpha, \beta \in \mathbb{R}$ such that

$$f'(x) - f(x) = \alpha x + \beta \quad \mu\text{-a.s.}, \quad g(y) - g'(y) = \alpha y + \beta \quad \nu\text{-a.s.}, \quad h - h' = \alpha \quad \mu\text{-a.s.}$$

Theorem 2.4 implies a strong duality result.

Corollary 2.5 (Strong duality). *Define the dual domain*

$$\mathcal{D} = \left\{ (\tilde{f}, \tilde{g}, \tilde{h}) \in L^1(\mu) \times L^1(\nu) \times L^0(\mu) : \tilde{h}(x)(y-x) \in L^1(\pi) \text{ for all } \pi \in \mathcal{M}_{\text{fin}}(\mu, \nu) \right\}.$$

We have

$$\begin{aligned} & \inf_{\pi \in \mathcal{M}(\mu, \nu)} H(\pi|\mu \otimes \nu) \\ &= \sup_{(\tilde{f}, \tilde{g}, \tilde{h}) \in \mathcal{D}} \int \tilde{f}(x) \mu(dx) + \int \tilde{g}(y) \nu(dy) - \log \int e^{\tilde{f}(x) + \tilde{g}(y) - \tilde{h}(x)(y-x)} \mu(dx) \nu(dy) \end{aligned}$$

and the argmax of the right-hand side is precisely the set of potentials (f, g, h) . In particular,

$$\inf_{\pi \in \mathcal{M}(\mu, \nu)} H(\pi|\mu \otimes \nu) = \int f(x) \mu(dx) + \int g(y) \nu(dy).$$

Proof. If $\pi = \pi^*$ and $(\tilde{f}, \tilde{g}, \tilde{h})$ are potentials, the two sides are equal because the last integral has value $\pi^*(\mathbb{R} \times \mathbb{R}) = 1$. The result thus follows from the weak duality inequality, shown in Lemma 3.19. \square

3. PROOFS

The primal existence and uniqueness are straightforward. Indeed, by the lower semi-continuity of $\pi \mapsto H(\pi|\mu \otimes \nu)$ and weak compactness of $\mathcal{M}(\mu, \nu)$ there exists a minimizer $\pi^* \in \mathcal{M}(\mu, \nu)$ for the martingale Schrödinger bridge problem (7),

$$\inf_{\pi \in \mathcal{M}(\mu, \nu)} H(\pi|\mu \otimes \nu) = H(\pi^*|\mu \otimes \nu),$$

and the minimizer is unique by the strict convexity of $\pi \mapsto H(\pi|\mu \otimes \nu)$. Alternately, existence and uniqueness also follow from the general result of [7] on entropy minimization.

The essential part of Theorem 2.4 is to construct the functions f, g, h together with their integrability properties. As the proof is quite long, the following remark sketches its broad idea.

Remark 3.1 (Proof idea). We try to relax the delicate martingale constraint into a weaker constraint that is easier to handle. Specifically, we consider

$$(9) \quad \inf_{\pi \in \widehat{\mathcal{M}}(\mu, \nu)} H(\pi|\mu \otimes \nu),$$

where

$$(10) \quad \widehat{\mathcal{M}}(\mu, \nu) = \left\{ \pi \in \Pi(\mu, \nu) : x \int (y - x) \pi_x(dy) \geq 0 \quad \mu\text{-a.s.} \right\}.$$

Clearly $\widehat{\mathcal{M}}(\mu, \nu) \supseteq \mathcal{M}(\mu, \nu)$. We will establish the desired properties for the optimizer $\hat{\pi}$ of (9), and then show that it is in fact also the optimizer for (7). The intuition for this is as follows. The measure $\pi = \mu \otimes \nu$ satisfies $\pi_x = \nu$ and hence

$$\int (y - x) \pi_x(dy) = \int y \nu(dy) - x = -x, \quad \pi := \mu \otimes \nu.$$

That is, the sign of the barycenter of π_x is opposite to the sign of x , which is the exact opposite of the condition in (10). As the optimizer $\hat{\pi}$ of (9) strives to be “close” to $\mu \otimes \nu$ by its definition, we may speculate that it saturates the constraint in (10), which is to say that $\hat{\pi} \in \mathcal{M}(\mu, \nu)$.

Our proof below starts from the dual of the convex problem (9). We will show that

$$\inf_{\pi \in \widehat{\mathcal{M}}(\mu, \nu)} H(\pi|\mu \otimes \nu) = \sup_{h \in C_b(\mathbb{R}) : xh(x) \leq 0} \inf_{\pi \in \Pi(\mu, \nu)} \int h(x)(y - x) \pi(dx, dy) + H(\pi|\mu \otimes \nu),$$

where $C_b(\mathbb{R})$ denotes the bounded continuous functions on \mathbb{R} . Taking a maximizing sequence (h_m) for the supremum, we define (f_m, g_m) as the corresponding Schrödinger potentials from EOT theory and argue that the sequence of measures $(f_m, g_m, h_m)_*(\mu \otimes \nu)$ is tight (where $*$ denotes pushforward). Denoting the weak limit by $(f, g, h)_*(\mu \otimes \nu)$, we prove that (f, g, h) give rise to the form (8) for $\hat{\pi}$, and that the optimizers $\hat{\pi}$ and π^* coincide. (Throughout this paper, measures are endowed with the weak topology induced by $C_b(\mathbb{R})$.)

Moving on to the actual proof, we recall the centering (6) as well as $\mu \neq \delta_0$ and $\nu \neq \delta_0$. We also define $\text{sign}(x) := \mathbf{1}_{\mathbb{R}^+}(x) - \mathbf{1}_{\mathbb{R}^-}(x)$, where

$$\mathbb{R}^+ := [0, \infty), \quad \mathbb{R}^- := (-\infty, 0).$$

We first introduce a probability measure $P \in \Pi(\mu, \nu)$ which will act as a counterpart of $\mu \otimes \nu$ in the sense that its kernel satisfies inequalities opposite to the ones that ν (the kernel of $\mu \otimes \nu$) satisfies.

Lemma 3.2. *There exists $P \in \Pi(\mu, \nu)$ such that $H(P|\mu \otimes \nu) < \infty$ and μ -a.s.,*

$$\int (y - x) P_x(dy) \begin{cases} > 0 & \text{for } x \geq 0, \\ < 0 & \text{for } x < 0. \end{cases}$$

Proof. By Assumption 2.1, there exists $\bar{\pi} \in \mathcal{M}(\mu, \nu)$ satisfying $\bar{\pi} \sim \mu \otimes \nu$ and $H(\bar{\pi}|\mu \otimes \nu) < \infty$. Fix an arbitrary coupling

$$\gamma \in \Pi\left(\frac{1}{\mu(\mathbb{R}^+)}\mu|_{\mathbb{R}^+}, \frac{1}{\mu(\mathbb{R}^-)}\mu|_{\mathbb{R}^-}\right);$$

it exists due to $\mu \neq \delta_0$. We claim that

$$\varepsilon(x^+, x^-) := \bar{\pi}_{x^+}((-\infty, x^-)) \wedge \bar{\pi}_{x^-}(x^+, \infty) > 0 \quad \gamma(dx^+, dx^-)\text{-a.s.}$$

Indeed, suppose that $\bar{\pi}_{x^+}((-\infty, x^-)) = 0$ on a set of positive γ -measure, then as $\bar{\pi}_x \sim \nu$ μ -a.s. due to $\bar{\pi} \sim \mu \otimes \nu$, we have

$$\bar{\pi}_{x^+}((-\infty, x^-)) = 0 = \nu((-\infty, x^-)) = 0 = \bar{\pi}_{x^-}((-\infty, x^-)).$$

As $\bar{\pi}_{x^-}$ has barycenter x^- , this implies that $\bar{\pi}_{x^-} = \delta_{x^-}$. This contradicts that $\bar{\pi}_x \sim \nu$ and ν is not a Dirac measure.

Next, we define

$$\begin{aligned} P_{x^+, x^-}^+ &:= \mu(\mathbb{R}^+) \left(\bar{\pi}_{x^+} + \varepsilon(x^+, x^-) \mu(\mathbb{R}^-) \left[\frac{\bar{\pi}_{x^-}|_{(x^+, \infty)}}{\bar{\pi}_{x^-}((x^+, \infty))} - \frac{\bar{\pi}_{x^+}|_{(-\infty, x^-)}}{\bar{\pi}_{x^+}((-\infty, x^-))} \right] \right), \\ P_{x^+, x^-}^- &:= \mu(\mathbb{R}^-) \left(\bar{\pi}_{x^-} + \varepsilon(x^+, x^-) \mu(\mathbb{R}^+) \left[\frac{\bar{\pi}_{x^+}|_{(-\infty, x^-)}}{\bar{\pi}_{x^+}((-\infty, x^-))} - \frac{\bar{\pi}_{x^-}|_{(x^+, \infty)}}{\bar{\pi}_{x^-}((x^+, \infty))} \right] \right) \end{aligned}$$

and set

$$(11) \quad P(dx, dy) := \int_{\{x^- \in \mathbb{R}^-\}} P_{x^+, x^-}^+(dy) \gamma(dx, dx^-) + \int_{\{x^+ \in \mathbb{R}^+\}} P_{x^+, x^-}^-(dy) \gamma(dx^+, dx).$$

To see that $P \in \Pi(\mu, \nu)$, we consider an arbitrary Borel set $A \subseteq \mathbb{R}$ and compute

$$\begin{aligned} P(A \times \mathbb{R}) &= \int_{\{(x, x^-, y) \in A \times \mathbb{R}^- \times \mathbb{R}\}} P_{x^+, x^-}^+(dy) \gamma(dx, dx^-) + \int_{\{(x, x^+, y) \in A \times \mathbb{R}^+ \times \mathbb{R}\}} P_{x^+, x^-}^-(dy) \gamma(dx^+, dx) \\ &= \mu(\mathbb{R}^+) \int_{\{(x, x^-) \in A \cap \mathbb{R}^+ \times \mathbb{R}^-\}} \gamma(dx, dx^-) + \mu(\mathbb{R}^-) \int_{\{(x, x^+) \in A \cap \mathbb{R}^- \times \mathbb{R}^+\}} \gamma(dx^+, dx) \\ &= \mu(\mathbb{R}^+) \frac{1}{\mu(\mathbb{R}^+)} \mu(A \cap \mathbb{R}^+) + \mu(\mathbb{R}^-) \frac{1}{\mu(\mathbb{R}^-)} \mu(A \cap \mathbb{R}^-) = \mu(A) \end{aligned}$$

as well as

$$\begin{aligned}
& P(\mathbb{R} \times A) \\
&= \int_{\{(x, x^-, y) \in \mathbb{R}^+ \times \mathbb{R}^- \times A\}} P_{x, x^-}^+(dy) \gamma(dx, dx^-) + \int_{\{(x, x^+, y) \in \mathbb{R}^- \times \mathbb{R}^+ \times A\}} P_{x^+, x}^-(dy) \gamma(dx^+, dx) \\
&= \int_{\mathbb{R}^+} \bar{\pi}_x(A) \mu(dx) + \int_{\mathbb{R}^-} \bar{\pi}_x(A) \mu(dx) \\
&\quad + \int_{\{(x^+, x^-) \in \mathbb{R}^+ \times \mathbb{R}^-\}} \varepsilon(x^+, x^-) \mu(\mathbb{R}^-) \mu(\mathbb{R}^+) \left(\left[\frac{\bar{\pi}_{x^-}(A \cap (x^+, \infty))}{\bar{\pi}_{x^-}((x^+, \infty))} - \frac{\bar{\pi}_{x^+}(A \cap (-\infty, x^-))}{\bar{\pi}_{x^+}((-\infty, x^-))} \right] \right. \\
&\quad \left. + \left[\frac{\bar{\pi}_{x^+}(A \cap (-\infty, x^-))}{\bar{\pi}_{x^+}((-\infty, x^-))} - \frac{\bar{\pi}_{x^-}(A \cap (x^+, \infty))}{\bar{\pi}_{x^-}((x^+, \infty))} \right] \right) \gamma(dx^+, dx^-) = \nu(A).
\end{aligned}$$

Furthermore,

$$\int (y - x) P_x(dy) \begin{cases} > 0 & \text{for } x \geq 0, \\ < 0 & \text{for } x < 0 \end{cases}$$

follows directly from the construction of P .

It remains to show $H(P|\mu \otimes \nu) < \infty$. To that end, write

$$(12) \quad (\mu \otimes \nu)(dx, dy) = \mu(\mathbb{R}^+) \int_{\{x^- \in \mathbb{R}^-\}} \nu(dy) \gamma(dx, dx^-) + \mu(\mathbb{R}^-) \int_{\{x^+ \in \mathbb{R}^+\}} \nu(dy) \gamma(dx^+, dx)$$

and recall that for any $R, Q \in \mathcal{P}(\mathbb{R}^2)$ with first marginals R^1, Q^1 we have the “chain rule”

$$(13) \quad H(R|Q) = H(R^1|Q^1) + \int H(R_x|Q_x) R^1(dx).$$

Comparing (11) and (12) and using joint convexity of $H(\cdot|\cdot)$ twice, we have

$$\begin{aligned}
H(P|\mu \otimes \nu) &\leq \mu(\mathbb{R}^+) \int_{\{x^- \in \mathbb{R}^-\}} H\left(\int \frac{P_{\cdot, x^-}^+}{\mu(\mathbb{R}^+)} \gamma_{x^-}(\cdot) \middle| \nu \otimes \gamma_{x^-}\right) \frac{1}{\mu(\mathbb{R}^-)} \mu(dx^-) \\
&\quad + \mu(\mathbb{R}^-) \int_{\{x^+ \in \mathbb{R}^+\}} H\left(\int \frac{P_{x^+, \cdot}^-}{\mu(\mathbb{R}^-)} \gamma_{x^+}(\cdot) \middle| \nu \otimes \gamma_{x^+}\right) \frac{1}{\mu(\mathbb{R}^+)} \mu(dx^+).
\end{aligned}$$

Using (13), we obtain

$$\begin{aligned}
H(P|\mu \otimes \nu) &\leq \mu(\mathbb{R}^+) \int_{\{(x, x^-) \in \mathbb{R}^+ \times \mathbb{R}^-\}} H(P_{x, x^-}^+ / \mu(\mathbb{R}^+) | \nu) \gamma(dx, dx^-) \\
&\quad + \mu(\mathbb{R}^-) \int_{\{(x^+, x) \in \mathbb{R}^+ \times \mathbb{R}^-\}} H(P_{x^+, x}^- / \mu(\mathbb{R}^-) | \nu) \gamma(dx^+, dx) \\
&\leq \mu(\mathbb{R}^+) \int_{\{(x, x^-) \in \mathbb{R}^+ \times \mathbb{R}^-\}} H(\bar{\pi}_x | \nu) + H(\mu(\mathbb{R}^-) \bar{\pi}_{x^-} | \nu) \gamma(dx, dx^-) \\
&\quad + \mu(\mathbb{R}^-) \int_{\{(x^+, x) \in \mathbb{R}^+ \times \mathbb{R}^-\}} H(\bar{\pi}_x | \nu) + H(\mu(\mathbb{R}^+) \bar{\pi}_{x^+} | \nu) \gamma(dx^+, dx) + C \\
&\leq 2 \int H(\bar{\pi}_x | \nu) \mu(dx) + C < \infty,
\end{aligned}$$

where $C > 0$ is chosen such that $x \log(x) + C \geq 0$. This concludes the proof. \square

Next, we consider the convex set $\widehat{\mathcal{M}}(\mu, \nu) \supseteq \mathcal{M}(\mu, \nu)$ announced in (10),

$$\widehat{\mathcal{M}}(\mu, \nu) := \left\{ \pi \in \Pi(\mu, \nu) : x \int (y - x) \pi_x(dy) \geq 0 \quad \mu\text{-a.s.} \right\}.$$

As $\widehat{\mathcal{M}}(\mu, \nu)$ is weakly compact, there is a unique $\hat{\pi} \in \widehat{\mathcal{M}}(\mu, \nu)$ such that

$$(14) \quad H(\hat{\pi}|\mu \otimes \nu) = \inf_{\pi \in \widehat{\mathcal{M}}(\mu, \nu)} H(\pi|\mu \otimes \nu) \leq \inf_{\pi \in \mathcal{M}(\mu, \nu)} H(\pi|\mu \otimes \nu) < \infty.$$

We start our analysis with a duality result.

Lemma 3.3. *We have*

$$\begin{aligned} & \inf_{\pi \in \widehat{\mathcal{M}}(\mu, \nu)} H(\pi|\mu \otimes \nu) \\ &= \sup_{h \in C_b(\mathbb{R}) : xh(x) \leq 0} \inf_{\pi \in \Pi(\mu, \nu)} \int h(x)(y - x) \pi(dx, dy) + H(\pi|\mu \otimes \nu) \\ &= \sup_{h \in C_b(\mathbb{R}) : xh(x) \leq 0} \sup_{f, g \in C_b(\mathbb{R})} \int f(x) \mu(dx) + \int g(y) \nu(dy) \\ & \quad - \int e^{f(x) + g(y) - h(x)(y - x)} \mu(dx) \nu(dy) + 1. \end{aligned}$$

Proof. We first claim that

$$(15) \quad \inf_{\pi \in \widehat{\mathcal{M}}(\mu, \nu)} H(\pi|\mu \otimes \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \sup_{h \in C_b(\mathbb{R}) : xh(x) \leq 0} \int h(x)(y - x) \pi(dx, dy) + H(\pi|\mu \otimes \nu).$$

Here “ \geq ” holds because $\int h(x)(y - x) \pi(dx, dy) \leq 0$ for all $\pi \in \widehat{\mathcal{M}}(\mu, \nu)$ and $h \in C_b(\mathbb{R})$ with $xh(x) \leq 0$. To see the inequality “ \leq ” we fix $\pi \in \Pi(\mu, \nu) \setminus \widehat{\mathcal{M}}(\mu, \nu)$ and show that there exists $h \in C_b(\mathbb{R})$ with $xh(x) \leq 0$ such that $\int h(x)(y - x) \pi(dx, dy) > 0$. By scaling, this will show that the supremum has value $+\infty$.

Indeed, let $\varphi(x) := \int (y - x) \pi_x(dy)$ and $A := \{x : x\varphi(x) < 0\}$. Then $\varphi \in L^1(\mu)$ and $\mu(A) > 0$. As $0 \notin A$, either $\mu(A \cap (0, \infty)) > 0$ or $\mu(A \cap (-\infty, 0)) > 0$. We treat the first case; the second is analogous. Consider the finite signed measure $d\rho = \varphi d\mu$ on $(0, \infty)$. By our assumption, its negative part $\rho_- \neq 0$. As $C_b(0, \infty)$ is distribution-determining and $0 \neq \rho_- \neq \rho_+$, there exists a nonnegative $s \in C_b(0, \infty)$ with $\int s d\rho_- > \int s d\rho_+$. Moreover, we can choose s such that $s(0+) = 0$, and then s can be extended to a bounded continuous function on \mathbb{R} that vanishes on $(-\infty, 0]$. Taking $h(x) := -s(x)$, we have $xh(x) \leq 0$ and

$$\int h(x)(y - x) \pi(dx, dy) = \int h(x)\varphi(x) \mu(dx) = \int h(x)(\rho_+ - \rho_-)(dx) > 0,$$

completing the proof of (15).

Note that the sets $\{h \in C_b(\mathbb{R}) : xh(x) \leq 0\}$ and $\Pi(\mu, \nu)$ are convex, and $\Pi(\mu, \nu)$ is compact for the weak topology. Furthermore $h \mapsto \int h(x)(y - x) \pi(dx, dy)$ is concave and $\pi \mapsto \int h(x)(y - x) \pi(dx, dy)$ is convex and continuous on $\Pi(\mu, \nu)$; the continuity follows from the fact that $(x, y) \mapsto h(x)(y - x)$ is a continuous function with $|h(x)(y - x)| \leq C(1 + |x| + |y|)$ and the assumption that the marginals μ, ν have finite first moments. Using Sion’s minimax

theorem and (15),

$$\begin{aligned} \inf_{\pi \in \widehat{\mathcal{M}}(\mu, \nu)} H(\pi | \mu \otimes \nu) &= \inf_{\pi \in \Pi(\mu, \nu)} \sup_{h \in C_b(\mathbb{R}): xh(x) \leq 0} \int h(x)(y-x) \pi(dx, dy) + H(\pi | \mu \otimes \nu) \\ &= \sup_{h \in C_b(\mathbb{R}): xh(x) \leq 0} \inf_{\pi \in \Pi(\mu, \nu)} \int h(x)(y-x) \pi(dx, dy) + H(\pi | \mu \otimes \nu). \end{aligned}$$

We focus on the inner minimization for fixed h , which is an EOT problem with cost function $c(x, y) := h(x)(y - x)$. As $|h(x)(y - x)| \leq C(1 + |x| + |y|)$, we have $c \in L^1(\mu \otimes \nu)$ and c is equivalent to a nonnegative cost after adding $C(1 + |x| + |y|)$. EOT duality (specifically [24, Theorem 4.7 and Remark 4.4]) allows us to conclude that the above infimum equals

$$\sup_{f \in L^1(\mu), g \in L^1(\nu)} \int f(x) \mu(dx) + \int g(y) \nu(dy) - \int e^{f(x)+g(y)-h(x)(y-x)} \mu(dx) \nu(dy) + 1.$$

Here we can replace L^1 with L^∞ by monotone approximation, and then further replace L^∞ with $C_b(\mathbb{R})$ by a standard density argument, completing the proof. \square

Corollary 3.4. *Define*

$$L_m(\mathbb{R}) := \{h : \mathbb{R} \rightarrow [-m, m] : h \text{ is } m\text{-Lipschitz, } xh(x) \leq 0\}.$$

Then

$$\begin{aligned} (16) \quad & \inf_{\pi \in \widehat{\mathcal{M}}(\mu, \nu)} H(\pi | \mu \otimes \nu) \\ &= \sup_{m \in \mathbb{N}} \sup_{h \in L_m(\mathbb{R})} \inf_{\pi \in \Pi(\mu, \nu)} \int h(x)(y-x) \pi(dx, dy) + H(\pi | \mu \otimes \nu) \\ &= \sup_{m \in \mathbb{N}} \inf_{\pi \in \Pi(\mu, \nu)} \sup_{h \in L_m(\mathbb{R})} \int h(x)(y-x) \pi(dx, dy) + H(\pi | \mu \otimes \nu). \end{aligned}$$

Proof. By Lemma 3.3 and an approximation argument,

$$\begin{aligned} & \inf_{\pi \in \widehat{\mathcal{M}}(\mu, \nu)} H(\pi | \mu \otimes \nu) \\ &= \sup_{h \in C_b(\mathbb{R}): xh(x) \leq 0} \inf_{\pi \in \Pi(\mu, \nu)} \int h(x)(y-x) \pi(dx, dy) + H(\pi | \mu \otimes \nu) \\ &= \sup_{m \in \mathbb{N}} \sup_{h \in L_m(\mathbb{R})} \inf_{\pi \in \Pi(\mu, \nu)} \int h(x)(y-x) \pi(dx, dy) + H(\pi | \mu \otimes \nu). \end{aligned}$$

Specifically, the approximation argument uses tightness of $\Pi(\mu, \nu)$ and the fact that on any given compact interval, we can approximate a given $h \in C_b(\mathbb{R})$ with $xh(x) \leq 0$ uniformly by Lipschitz functions satisfying the same condition.

The last part of (16) now follows by another application of Sion's minimax theorem similar to the proof of Lemma 3.3. \square

Lemma 3.5. *Fix $m \in \mathbb{N}$. The problem*

$$\sup_{h \in L_m(\mathbb{R})} \inf_{\pi \in \Pi(\mu, \nu)} \int h(x)(y-x) \pi(dx, dy) + H(\pi | \mu \otimes \nu)$$

has a maximizer $h_m \in L_m(\mathbb{R})$ and the problem

$$\inf_{\pi \in \Pi(\mu, \nu)} \sup_{h \in L_m(\mathbb{R})} \int h(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu)$$

has a minimizer $\pi_m \in \Pi(\mu, \nu)$. Moreover, (h_m, π_m) is a saddle point:

$$\begin{aligned} & \sup_{h \in L_m(\mathbb{R})} \inf_{\pi \in \Pi(\mu, \nu)} \int h(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu) \\ &= \inf_{\pi \in \Pi(\mu, \nu)} \int h_m(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu) \\ (17) \quad &= \int h_m(x)(y-x) \pi_m(dx, dy) + H(\pi_m|\mu \otimes \nu) \\ &= \sup_{h \in L_m(\mathbb{R})} \int h(x)(y-x) \pi_m(dx, dy) + H(\pi_m|\mu \otimes \nu) \\ &= \inf_{\pi \in \Pi(\mu, \nu)} \sup_{h \in L_m(\mathbb{R})} \int h(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu). \end{aligned}$$

Finally, we have $\pi_m \in \Pi_m(\mu, \nu)$, where

$$\Pi_m(\mu, \nu) := \left\{ \pi \in \Pi(\mu, \nu) : \sup_{h \in L_m(\mathbb{R})} \int h(x)(y-x) \pi(dx, dy) \leq H(\hat{\pi}|\mu \otimes \nu) \right\}$$

and $\hat{\pi} \in \widehat{\mathcal{M}}(\mu, \nu)$ denotes the unique optimizer of (9).

Proof. By the Arzela–Ascoli theorem, $L_m(\mathbb{R})$ is compact with respect to uniform convergence on compact subsets. The map

$$L_m(\mathbb{R}) \ni h \mapsto \int h(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu)$$

is continuous for that convergence, so that the infimum

$$L_m(\mathbb{R}) \ni h \mapsto \inf_{\pi \in \Pi(\mu, \nu)} \int h(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu)$$

is upper semicontinuous. The existence of a maximizer $h_m \in L_m(\mathbb{R})$ follows. Similarly, weak compactness of $\Pi(\mu, \nu)$ and integrability of μ, ν yield a minimizer $\pi_m \in \Pi(\mu, \nu)$. To verify the saddle point property, note that

$$\begin{aligned} & \sup_{h \in L_m(\mathbb{R})} \inf_{\pi \in \Pi(\mu, \nu)} \int h(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu) \\ &= \inf_{\pi \in \Pi(\mu, \nu)} \int h_m(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu) \\ &\leq \int h_m(x)(y-x) \pi_m(dx, dy) + H(\pi_m|\mu \otimes \nu) \\ &\leq \sup_{h \in L_m(\mathbb{R})} \int h(x)(y-x) \pi_m(dx, dy) + H(\pi_m|\mu \otimes \nu) \\ &= \inf_{\pi \in \Pi(\mu, \nu)} \sup_{h \in L_m(\mathbb{R})} \int h(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu). \end{aligned}$$

Using once more Sion's minimax theorem, the order of infimum and supremum can be exchanged, showing that equality holds throughout (17).

It remains to show $\pi_m \in \Pi_m(\mu, \nu)$. As $xh_m(x) \leq 0$ and $x \int (y-x) \hat{\pi}_x(dy) \geq 0$, we have

$$\int h_m(x)(y-x) \hat{\pi}(dx, dy) = \int h_m(x) \int (y-x) \hat{\pi}_x(dy) \mu(dx) \leq 0.$$

Using the definition of π_m , it follows that

$$\begin{aligned} H(\hat{\pi}|\mu \otimes \nu) &\geq \int h_m(x)(y-x) \hat{\pi}(dx, dy) + H(\hat{\pi}|\mu \otimes \nu) \\ &\geq \int h_m(x)(y-x) \pi_m(dx, dy) + H(\pi_m|\mu \otimes \nu) \\ &\stackrel{(17)}{=} \sup_{h \in L_m(\mathbb{R})} \int h(x)(y-x) \pi_m(dx, dy) + H(\pi_m|\mu \otimes \nu) \\ &\geq \sup_{h \in L_m(\mathbb{R})} \int h(x)(y-x) \pi_m(dx, dy). \end{aligned} \quad \square$$

The following holds for arbitrary couplings in $\Pi_m(\mu, \nu)$, but will be applied to π_m as defined in Lemma 3.5.

Lemma 3.6. *Let $\pi_m \in \Pi_m(\mu, \nu)$, $m \geq 1$. If (π_m) converges weakly, the limit is in $\widehat{\mathcal{M}}(\mu, \nu)$.*

Proof. Let $\pi_m \rightarrow \pi$. We need to prove $x \int (y-x) \pi_x(dy) \geq 0$ μ -a.s. It suffices to show

$$\int h(x)(y-x) \pi(dx, dy) \leq 0 \quad \text{for all } h \in C_b(\mathbb{R}) \text{ satisfying } xh(x) \leq 0.$$

By the definition of $\Pi_m(\mu, \nu)$,

$$\sup_{h \in L_m(\mathbb{R})} \int h(x)(y-x) \pi_m(dx, dy) \leq H(\hat{\pi}|\mu \otimes \nu),$$

or equivalently

$$\sup_{h \in L_1(\mathbb{R})} \int h(x)(y-x) \pi_m(dx, dy) \leq \frac{H(\hat{\pi}|\mu \otimes \nu)}{m}.$$

Letting $m \rightarrow \infty$, we deduce for any $h \in C_b(\mathbb{R})$ satisfying $xh(x) \leq 0$ that

$$\begin{aligned} \int h(x)(y-x) \pi(dx, dy) &\leq \lim_{m \rightarrow \infty} \sup_{h \in L_1(\mathbb{R})} \int h(x)(y-x) \pi_m(dx, dy) \\ &\leq \limsup_{m \rightarrow \infty} \frac{H(\hat{\pi}|\mu \otimes \nu)}{m} = 0. \end{aligned} \quad \square$$

The next lemma will yield weak pre-compactness of $(h_m)_{m \geq 1}$, where $h_m \in L_m(\mathbb{R})$ was defined in Lemma 3.5.

Lemma 3.7. *Recall $P \in \Pi(\mu, \nu)$ as defined in Lemma 3.2. There exists $m_0 \in \mathbb{N}$ such that*

$$(18) \quad \int |h_m(x)| \left| \int (y-x) P_x(dy) \right| \mu(dx) \leq H(P|\mu \otimes \nu) < \infty \quad \text{for all } m \geq m_0.$$

The sequence (h_m) is bounded in $L^1(\tilde{\mu})$, for the finite nonnegative measure $\tilde{\mu} \sim \mu$ defined by

$$\frac{d\tilde{\mu}}{d\mu}(x) := \left| \int (y-x) P_x(dy) \right|.$$

Proof. By (16) and the definition of h_m in Lemma 3.5,

$$\inf_{\pi \in \Pi(\mu, \nu)} \int h_m(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu) \uparrow H(\hat{\pi}|\mu \otimes \nu), \quad m \rightarrow \infty.$$

Therefore, as $H(\hat{\pi}|\mu \otimes \nu) > 0$, there exists $m_0 \in \mathbb{R}$ such that

$$\inf_{\pi \in \Pi(\mu, \nu)} \int h_m(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu) \geq 0, \quad m \geq m_0.$$

Next, let $P \in \Pi(\mu, \nu)$ be as in Lemma 3.2. We conclude that for $m \geq m_0$,

$$\begin{aligned} 0 &\leq \inf_{\pi \in \Pi(\mu, \nu)} \int h_m(x)(y-x) \pi(dx, dy) + H(\pi|\mu \otimes \nu) \\ (19) \quad &\leq \int h_m(x)(y-x) P(dx, dy) + H(P|\mu \otimes \nu). \end{aligned}$$

As $x \int (y-x) P_x(dy) \geq 0$ and $x h_m(x) \leq 0$, we have $h_m(x) \int (y-x) P_x(dy) \leq 0$ and thus (19) becomes

$$0 \leq - \int |h_m(x)| \left| \int (y-x) P_x(dy) \right| \mu(dx) + H(P|\mu \otimes \nu)$$

or equivalently (18).

As $P \in \Pi(\mu, \nu)$ and μ, ν have finite first moments, $\tilde{\mu}$ is a finite measure, and $\tilde{\mu} \sim \mu$ by Lemma 3.2. The bound (18) can then be rephrased as $\|h_m\|_{L^1(\tilde{\mu})} \leq H(P|\mu \otimes \nu) < \infty$. \square

Corollary 3.8. *The sequence $(h_m)_{m \geq 1}$ is bounded in μ -probability.*

Proof. This follows from $\mu \sim \tilde{\mu}$ and boundedness of (h_m) in $L^1(\tilde{\mu})$, by a general fact of measure theory. We recall the argument for the convenience of the reader.

As $\mu, \tilde{\mu}$ are finite equivalent measures, $\varphi := d\mu/d\tilde{\mu} \in L^1(\tilde{\mu})$. Let $\varepsilon > 0$. Then $\varphi \in L^1(\tilde{\mu})$ yields $\delta > 0$ such that $\int_A \varphi d\tilde{\mu} \leq \varepsilon$ whenever A is a Borel set with $\tilde{\mu}(A) \leq \delta$. In particular, choosing $K > 0$ such that $\tilde{\mu}(|h_m| > K) \leq K^{-1} \tilde{\mu}(\mathbb{R}) \|h_m\|_{L^1(\tilde{\mu})} \leq \delta$ for all $m \geq 1$,

$$\mu(|h_m| > K) = \int_{\{|h_m| > K\}} \varphi d\tilde{\mu} \leq \varepsilon, \quad m \geq 1. \quad \square$$

Lemma 3.9. *The coupling $\pi_m \in \Pi(\mu, \nu)$ defined in Lemma 3.5 has a density of the form*

$$(20) \quad \frac{d\pi_m}{d(\mu \otimes \nu)}(x, y) = e^{f_m(x) + g_m(y) - h_m(x)(y-x)}$$

for some functions $(f_m, g_m) \in L^1(\mu) \times L^1(\nu)$ satisfying

$$(21) \quad \int f_m(x) \mu(dx) + \int g_m(y) \nu(dy) = \int h_m(x)(y-x) \pi_m(dx, dy) + H(\pi_m|\mu \otimes \nu)$$

and

$$(22) \quad \begin{aligned} \int e^{f_m(x) + g_m(y) - h_m(x)(y-x)} \mu(dx) &= 1 \quad \forall y \in \mathbb{R}, \\ \int e^{f_m(x) + g_m(y) - h_m(x)(y-x)} \nu(dy) &= 1 \quad \forall x \in \mathbb{R}. \end{aligned}$$

Proof. This follows from the definition of π_m as the solution of the EOT problem with cost $c(x, y) := h_m(x)(y - x)$. Specifically, we have $c \in L^1(\mu \otimes \nu)$ and $c(x, y) \geq -C(1 + |x| + |y|)$. The claims then follow from [24, Remark 4.4, Theorem 4.7, Theorem 4.2, Equation (4.11) and subsequent discussion]. \square

For later use, we emphasize that the identities in (22) hold everywhere (not just almost-surely) for the chosen versions of f_m, g_m .

Next, we establish bounds and equicontinuity properties of (f_m, g_m, h_m) which are inspired by [25] and crucial for the passage to the limit $m \rightarrow \infty$.

Lemma 3.10. *Fix $\delta \in (0, 1)$. There exist compact sets $\mathcal{X}_{\text{cpt}, \delta}, \mathcal{Y}_{\text{cpt}, \delta} \subseteq \mathbb{R}$ and Borel sets*

$$\tilde{A}_{m, \delta} \subseteq A_{m, \delta} \subseteq \mathcal{X}_{\text{cpt}, \delta}, \quad B_{m, \delta} \subseteq \mathcal{Y}_{\text{cpt}, \delta} \quad \text{with} \quad \mu(\tilde{A}_{m, \delta}) \wedge \nu(B_{m, \delta}) \geq 1 - \delta$$

such that $(h_m \mathbf{1}_{A_{m, \delta}})_{m \in \mathbb{N}}$ is uniformly bounded and for all $m \in \mathbb{N}$,

$$(23) \quad |g_m(y) - g_m(\tilde{y})| \leq \sup_{x \in A_{m, \delta}} |h_m(x)(y - \tilde{y})| - \log(1 - \delta) \quad \forall y, \tilde{y} \in B_{m, \delta},$$

$$(24) \quad - \left(\log \int_{B_{m, \delta}} e^{g_m(y) - h_m(x)(y - x)} \nu(dy) - \log(1 - \delta) \right) \leq f_m(x) \\ \leq - \log \int_{B_{m, \delta}} e^{g_m(y) - h_m(x)(y - x)} \nu(dy)$$

for all $x \in \tilde{A}_{m, \delta}$.

Proof. Set $\kappa := \delta^2 < \delta$. Choose compacts $\mathcal{X}_{\text{cpt}, \delta}$ and $\mathcal{Y}_{\text{cpt}, \delta}$ with $\mu(\mathcal{X}_{\text{cpt}, \delta}) \geq 1 - \kappa^2/4$ and $\nu(\mathcal{Y}_{\text{cpt}, \delta}) \geq 1 - \kappa^2/2$. By Corollary 3.8 there exists $K > 0$ such that $\mu(|h_m| > K) \leq \kappa^2/4$ for all $m \in \mathbb{N}$. Define $A_{m, \delta} := \mathcal{X}_{\text{cpt}, \delta} \cap \{|h_m| \leq K\}$. Then $\mu(A_{m, \delta}) \geq 1 - \kappa^2/2$. As $\pi_m \in \Pi(\mu, \nu)$,

$$(25) \quad \pi_m(A_{m, \delta} \times \mathcal{Y}_{\text{cpt}, \delta}) \geq 1 - \kappa^2.$$

Consider the set

$$B_{m, \delta} = \left\{ y \in \mathcal{Y}_{\text{cpt}, \delta} : \int_{A_{m, \delta}} e^{f_m(x) + g_m(y) - h_m(x)(y - x)} \mu(dx) \geq 1 - \kappa \right\}.$$

We claim that its complement $B_{m, \delta}^c$ satisfies

$$(26) \quad p_m := \nu(B_{m, \delta}^c) \leq \kappa.$$

Indeed, (22) yields

$$\int_{A_{m, \delta}} e^{f_m(x) + g_m(y) - h_m(x)(y - x)} \mu(dx) \leq \int e^{f_m(x) + g_m(y) - h_m(x)(y - x)} \mu(dx) = 1$$

and thus

$$\begin{aligned}
1 - \kappa^2 &\stackrel{(25)}{\leq} \pi_m(A_{m,\delta} \times \mathcal{Y}_{\text{cpt},\delta}) = \int_{\mathcal{Y}_{\text{cpt},\delta}} \int_{A_{m,\delta}} e^{f_m(x)+g_m(y)-h_m(x)(y-x)} \mu(dx) \nu(dy) \\
&\leq \int_{B_{m,\delta}^c} \int_{A_{m,\delta}} e^{f_m(x)+g_m(y)-h_m(x)(y-x)} \mu(dx) \nu(dy) \\
&\quad + \int_{B_{m,\delta}} \int_{A_{m,\delta}} e^{f_m(x)+g_m(y)-h_m(x)(y-x)} \mu(dx) \nu(dy) \\
&\leq (1 - \kappa)p_m + (1 - p_m) = 1 - p_m\kappa,
\end{aligned}$$

which implies (26). Next, we observe from the definition of $B_{m,\delta}$ that for $y \in B_{m,\delta}$,

$$\begin{aligned}
(27) \quad & - \left(\log \int_{A_{m,\delta}} e^{f_m(x)-h_m(x)(y-x)} \mu(dx) - \log(1 - \kappa) \right) \leq g_m(y) \\
& \leq - \log \int_{A_{m,\delta}} e^{f_m(x)-h_m(x)(y-x)} \mu(dx).
\end{aligned}$$

Let $y, \tilde{y} \in B_{m,\delta}$ and assume without loss of generality that $g_m(y) \geq g_m(\tilde{y})$. Then

$$\begin{aligned}
|g_m(y) - g_m(\tilde{y})| &\leq \log \int_{A_{m,\delta}} e^{f_m(x)-h_m(x)(\tilde{y}-x)} \mu(dx) - \log(1 - \kappa) \\
&\quad - \log \int_{A_{m,\delta}} e^{f_m(x)-h_m(x)(y-x)} \mu(dx) \\
&= \log \int_{A_{m,\delta}} e^{h_m(x)(y-x)-h_m(x)(\tilde{y}-x)+f_m(x)-h_m(x)(y-x)} \mu(dx) - \log(1 - \kappa) \\
&\quad - \log \int_{A_{m,\delta}} e^{f_m(x)-h_m(x)(y-x)} \mu(dx) \\
&\leq \log \left(e^{\sup_{x \in A_{m,\delta}} |h_m(x)(y-x)-h_m(x)(\tilde{y}-x)|} \int_{A_{m,\delta}} e^{f_m(x)-h_m(x)(y-x)} \mu(dx) \right) \\
&\quad - \log(1 - \kappa) - \log \int_{A_{m,\delta}} e^{f_m(x)-h_m(x)(y-x)} \mu(dx) \\
&= \sup_{x \in A_{m,\delta}} |h_m(x)(y - \tilde{y})| - \log(1 - \kappa).
\end{aligned}$$

This concludes the proof of (23). Turning to (24), note that by (25), (26) and the definition of $B_{m,\delta}$,

$$\begin{aligned}
(28) \quad & \pi_m(A_{m,\delta} \times B_{m,\delta}) \geq \pi_m(A_{m,\delta} \times \mathcal{Y}_{\text{cpt},\delta}) - \pi_m(A_{m,\delta} \times B_{m,\delta}^c) \\
& \geq 1 - \kappa^2 - \int_{B_{m,\delta}^c} \int_{A_{m,\delta}} e^{f_m(x)+g_m(y)-h_m(x)(y-x)} \mu(dx) \nu(dy) \\
& \geq 1 - \kappa^2 - \kappa(1 - \kappa) = 1 - \kappa = 1 - \delta^2,
\end{aligned}$$

where we used our definition $\kappa := \delta^2$ which ensures in particular that $\kappa \in (0, \delta)$. Define

$$\tilde{A}_{m,\delta} = \left\{ x \in A_{m,\delta} : \int_{B_{m,\delta}} e^{f_m(x) + g_m(y) - h_m(x)(y-x)} \nu(dy) \geq 1 - \delta \right\}.$$

Arguing as for (26) and (27), now using (28) instead of (25), we see that $\mu(\tilde{A}_{m,\delta}^c) \leq \delta$ and that (24) holds for $x \in \tilde{A}_{m,\delta}$. \square

Lemma 3.11. *Recall the sets $\tilde{A}_{m,\delta}$ and $B_{m,\delta}$ from Lemma 3.10. Given $A, B \in \mathcal{B}(\mathbb{R})$ with $\mu(A), \nu(B) > \delta$, there exist a subsequence $(m_k)_{k \geq 1} \subseteq \mathbb{N}$ and $x_0, y_0 \in \mathbb{R}$ such that*

$$x_0 \in \tilde{A}_{m_k,\delta} \cap A \quad \text{and} \quad y_0 \in B_{m_k,\delta} \cap B \quad \text{for all } k \geq 1.$$

Proof. We show the claim for x_0 ; the proof for y_0 is similar. Define

$$D_m := \bigcap_{l=m}^{\infty} (\tilde{A}_{l,\delta} \cap A)^c.$$

Then $(D_m)_{m \geq 1}$ are increasing and the claim is equivalent to $\bigcap_m D_m^c \neq \emptyset$. By Lemma 3.10 we have $\mu(\tilde{A}_{m,\delta}) \geq 1 - \delta$ and thus $\inf_m \mu(\tilde{A}_{m,\delta} \cap A) > 0$. As a consequence,

$$0 < \limsup_{m \rightarrow \infty} \mu(\tilde{A}_{m,\delta} \cap A) \leq \limsup_{m \rightarrow \infty} \mu(D_m^c),$$

showing that $\mu(\bigcap_m D_m^c) > 0$. \square

Applying Lemma 3.11 with $B = \mathbb{R}$ and taking a subsequence if necessary, we may assume that there exists a common point $y_0 \in B_{m,\delta}$ for all $m \in \mathbb{N}$. For the remainder of the proof we choose the normalization

$$(29) \quad g_m(y_0) = 0 \quad \text{for all } m \geq 1,$$

which is achieved by replacing (f_m, g_m) with $(f_m + g_m(y_0), g_m - g_m(y_0))$. We remark that a second normalization has been made implicitly: we have $h_m(0) = 0$ since h_m is continuous and $xh_m(x) \leq 0$. For ease of reference, we summarize the consequences of Lemma 3.10 and (29).

Lemma 3.12. *For any $\delta \in (0, 1)$, the sequences*

$$(f_m \mathbf{1}_{\tilde{A}_{m,\delta}})_{m \geq 1}, \quad (g_m \mathbf{1}_{B_{m,\delta}})_{m \geq 1}, \quad (h_m \mathbf{1}_{A_{m,\delta}})_{m \geq 1} \quad \text{are uniformly bounded.}$$

In particular, $(f_m)_{m \geq 1}$ and $(h_m)_{m \geq 1}$ are μ -tight, and $(g_m)_{m \geq 1}$ is ν -tight.

Proof. Recall from Lemma 3.10 that $(h_m \mathbf{1}_{A_{m,\delta}})$ is uniformly bounded and $\mu(A_{m,\delta}) > 1 - \delta$, showing the claims for (h_m) . (In fact, tightness of (h_m) is tautological with Corollary 3.8.) Lemma 3.10 also states that

$$|g_m(y) - g_m(\tilde{y})| \leq \sup_{x \in A_{m,\delta}} |h_m(x)(y - \tilde{y})| - \log(1 - \delta) \quad \forall y, \tilde{y} \in B_{m,\delta}, \quad m \in \mathbb{N},$$

that $B_{m,\delta} \subseteq \mathcal{V}_{\text{cpt},\delta}$ and $\nu(B_{m,\delta}) \geq 1 - \delta$ for all $m \in \mathbb{N}$. Together with (29) and $y_0 \in B_{m,\delta}$ this yields

$$(30) \quad |g_m(y)| \leq \sup_{x \in A_{m,\delta}} |h_m(x)(y - y_0)| - \log(1 - \delta) \quad \forall y \in B_{m,\delta}, \quad m \in \mathbb{N}.$$

As the right-hand side is bounded uniformly in $m \in \mathbb{N}$, the claims for (g_m) follow. For (f_m) we use (24):

$$\begin{aligned} - \left(\log \int_{B_{m,\delta}} e^{g_m(y) - h_m(x)(y-x)} \nu(dy) - \log(1 - \delta) \right) &\leq f_m(x) \\ &\leq - \log \int_{B_{m,\delta}} e^{g_m(y) - h_m(x)(y-x)} \nu(dy) \end{aligned}$$

for all $x \in \tilde{A}_{m,\delta}$ and all $m \in \mathbb{N}$. By (30), $g_m(y)$ is uniformly bounded on $B_{m,\delta} \subseteq \mathcal{Y}_{\text{cpt},\delta}$. Moreover, y is uniformly bounded on that set by compactness, while $h_m(x)$ and x are uniformly bounded on $A_{m,\delta} \subseteq \mathcal{X}_{\text{cpt},\delta}$. The claims for (f_m) now follow from $\tilde{A}_{m,\delta} \subseteq A_{m,\delta}$ and $\mu(\tilde{A}_{m,\delta}) \geq 1 - \delta$. \square

We shall use the following version of Prokhorov's theorem. (Compared, e.g., with [11, Theorem 3.2.2, p. 100], this version constructs the limit F on (\mathbb{R}, μ) without extending the probability space.) We denote by $F_*\mu(A) := \mu(F^{-1}(A))$ the pushforward of the measure μ by the Borel function $F : \mathbb{R} \rightarrow \mathbb{R}^d$.

Lemma 3.13. *Let $F_n : \mathbb{R} \rightarrow \mathbb{R}^d$ be Borel and let $(F_n)_{n \geq 1}$ be μ -tight. After passing to a subsequence, there exists $F : \mathbb{R} \rightarrow \mathbb{R}^d$ such that $F_n \rightarrow F$ weakly under μ ; i.e., $(F_n)_*\mu \rightarrow F_*\mu$.*

Proof. Let S be the set of atoms of μ . We can decompose $\mu = \mu_p + \mu_c$, where $\mu_p = \mu|_S$ is a purely atomic measure supported on S and $\mu_c = \mu|_{\mathbb{R} \setminus S}$ is atomless.

The set S is countable and by tightness, $(F_n(x))_n$ is bounded for each $x \in S$. Using a diagonal argument, $(F_n(x))_n$ converges for all $x \in S$, after taking a subsequence. If $\mu_c = 0$, setting $F(x) = \lim_n F_n(x)$ for $x \in S$ and $F(x) = 0$ for $x \in \mathbb{R} \setminus S$ completes the proof.

Suppose $\mu_c \neq 0$. The sequence $(F_n)_*\mu_c = (F_n)_*(\mu - \mu_p)$ is again tight. Applying Prokhorov's theorem to $(F_n)_*\mu_c$ and taking another subsequence, we obtain $(F_n)_*\mu_c \rightarrow \tilde{\mu}$ for a measure $\tilde{\mu}$ with $\tilde{\mu}(\mathbb{R}^d) = \mu_c(\mathbb{R})$. As (\mathbb{R}, μ_c) is atomless, there exists a random vector with law $\tilde{\mu}$; i.e., a Borel function $G : \mathbb{R} \rightarrow \mathbb{R}^d$ with $G_*\mu_c = \tilde{\mu}$.

We define $F(x) := \lim_n F_n(x)$ for $x \in S$ and $F(x) := G(x)$ for $x \in \mathbb{R} \setminus S$. \square

Lemma 3.14. *There are Borel functions $\hat{f}, \hat{g}, \hat{h} : \mathbb{R} \rightarrow \mathbb{R}$ such that, after taking a subsequence, $(f_m, h_m) \rightarrow (\hat{f}, \hat{h})$ weakly under μ and $g_m \rightarrow \hat{g}$ weakly under ν .*

Moreover, the functions $U_m(x, y) := f_m(x) + g_m(y) - h_m(x)(y - x)$ converge weakly under $(\mu \otimes \nu)$ to $\hat{U}(x, y) := \hat{f}(x) + \hat{g}(y) - \hat{h}(x)(y - x)$.

Proof. The first part follows by applying Lemma 3.13 together with Lemma 3.12. For the second part, note that

$$(f_m(x), h_m(x), x)_*\mu \rightarrow (\hat{f}(x), \hat{h}(x), x)_*\mu.$$

Similarly $(g_m(y), y)_*\nu \rightarrow (\hat{g}(y), y)_*\nu$. Thus by independence,

$$(f_m(x), h_m(x), x, g_m(y), y)_*(\mu \otimes \nu) \rightarrow (\hat{f}(x), \hat{h}(x), x, \hat{g}(y), y)_*(\mu \otimes \nu).$$

The claim follows. \square

Next, we connect the weak convergence of f_m, g_m, h_m to weak convergence of π_m . The limit $\hat{\pi}$ will be identified as $\hat{\pi}$ (and eventually as π^*) in subsequent steps.

Lemma 3.15. *Define the measure $\hat{\pi}$ on \mathbb{R}^2 by*

$$(31) \quad \frac{d\hat{\pi}}{d(\mu \otimes \nu)}(x, y) = e^{\hat{f}(x) + \hat{g}(y) - \hat{h}(x)(y-x)}.$$

Then $\pi_m \rightarrow \hat{\pi}$ weakly for $m \rightarrow \infty$.

Proof. Recall from (20) that π_m has Radon–Nikodym derivative

$$\frac{d\pi_m}{d(\mu \otimes \nu)}(x, y) = e^{f_m(x) + g_m(y) - h_m(x)(y-x)}.$$

To show $\pi_m \rightarrow \hat{\pi}$, we fix an arbitrary Borel set $S \subseteq \mathbb{R} \times \mathbb{R}$ with $(\mu \otimes \nu)(\partial S) = 0$ and prove

$$(32) \quad \lim_{m \rightarrow \infty} |\pi_m(S) - \hat{\pi}(S)| = 0$$

(∂S denotes the boundary of S). Indeed, fix $\delta > 0$ and recall from Lemma 3.12 that there exists $C > 0$ such that

$$(33) \quad g_m \leq C \quad \text{on } B_{m,\delta} \text{ for all } m \in \mathbb{N},$$

$$(34) \quad f_m \leq C \quad \text{on } \tilde{A}_{m,\delta} \text{ for all } m \in \mathbb{N},$$

and that $\mu(\tilde{A}_{m,\delta}) \wedge \nu(B_{m,\delta}) \geq 1 - \delta$. Recalling also that (h_m) is uniformly bounded on $A_{m,\delta}$, we have, after enlarging $C > 0$ if necessary,

$$\sup_{m \in \mathbb{N}} \mu(|h_m| \geq \sqrt{C}) \leq \delta/2, \quad \mu(|x| \geq \sqrt{C}/2) \vee \nu(|y| \geq \sqrt{C}/2) \leq \delta/4,$$

and thus

$$(35) \quad \begin{aligned} \pi_m(|h_m(x)(y-x)| \geq C) &\leq \pi_m(|h_m(x)| \geq \sqrt{C}) + \pi_m(|x-y| \geq \sqrt{C}) \\ &\leq \mu(|h_m(x)| \geq \sqrt{C}) + \mu(|x| \geq \sqrt{C}/2) + \nu(|y| \geq \sqrt{C}/2) \\ &\leq \delta/2 + \delta/4 + \delta/4 = \delta. \end{aligned}$$

Define measures π_m^C and $\hat{\pi}^C$ by

$$\begin{aligned} \frac{d\pi_m^C}{d(\mu \otimes \nu)}(x, y) &:= e^{[f_m(x) + g_m(y) - h_m(x)(y-x)] \wedge 3C}, \\ \frac{d\hat{\pi}^C}{d(\mu \otimes \nu)}(x, y) &:= e^{[\hat{f}(x) + \hat{g}(y) - \hat{h}(x)(y-x)] \wedge 3C}. \end{aligned}$$

We first note that by the monotone convergence theorem,

$$(36) \quad \int_S e^{[\hat{f}(x) + \hat{g}(y) - \hat{h}(x)(y-x)] \wedge 3C} \mu(dx) \nu(dy) \uparrow \int_S e^{\hat{f}(x) + \hat{g}(y) - \hat{h}(x)(y-x)} \mu(dx) \nu(dy)$$

for $C \uparrow \infty$. On the other hand, using (33)–(35),

$$(37) \quad \begin{aligned} \pi_m(S) &\geq \pi_m^C(S) \geq \pi_m^C(S \cap \{f_m < C\} \cap \{g_m < C\} \cap \{-h_m(x)(y-x) < C\}) \\ &= \pi_m(S \cap \{f_m < C\} \cap \{g_m < C\} \cap \{-h_m(x)(y-x) < C\}) \\ &\geq \pi_m(S) - \mu(f_m \geq C) - \nu(g_m \geq C) - \pi_m(-h_m(x)(y-x) \geq C) \\ &\geq \pi_m(S) - 3\delta. \end{aligned}$$

Lastly,

$$(38) \quad \left| \int_S e^{[f_m(x) + g_m(y) - h_m(x)(y-x)] \wedge 3C} - e^{[\hat{f}(x) + \hat{g}(y) - \hat{h}(x)(y-x)] \wedge 3C} \mu(dx) \nu(dy) \right| \rightarrow 0$$

for $m \rightarrow \infty$ by Lemma 3.14 and $(\mu \otimes \nu)(\partial S) = 0$. Combining (36)–(38) yields

$$\begin{aligned} & \lim_{m \rightarrow \infty} |\pi_m(S) - \hat{\pi}(S)| \\ & \leq \limsup_{C \rightarrow \infty} \limsup_{m \rightarrow \infty} (|\pi_m(S) - \pi_m^C(S)| + |\pi_m^C(S) - \hat{\pi}^C(S)| + |\hat{\pi}^C(S) - \hat{\pi}(S)|) \leq 3\delta. \end{aligned}$$

As $\delta > 0$ was arbitrary, (32) follows and the proof is complete. \square

We now come to the key technical step of identifying the limit and showing integrability.

Lemma 3.16. *We have $(\hat{f}, \hat{g}) \in L^1(\mu) \times L^1(\nu)$ and $\hat{h}(x)(y - x) \in L^1(\hat{\pi})$. Moreover,*

$$\begin{aligned} \hat{\pi} &= \hat{\pi}, \\ H(\hat{\pi}|\mu \otimes \nu) &= \int \hat{f}(x) \mu(dx) + \int \hat{g}(y) \nu(dy), \\ \int \hat{h}(x)(y - x) \hat{\pi}(dx, dy) &= 0. \end{aligned}$$

Proof. Step 1. We first show $\hat{\pi} = \hat{\pi}$. As $\pi_m \rightarrow \hat{\pi}$ by Lemma 3.15 and $\pi_m \in \Pi_m(\mu, \nu)$ by Lemma 3.5, we have $\hat{\pi} \in \widehat{\mathcal{M}}(\mu, \nu)$ by Lemma 3.6. Recall from (17) that

$$\begin{aligned} & \int h_m(x)(y - x) \pi_m(dx, dy) + H(\pi_m|\mu \otimes \nu) \\ (39) \quad &= \sup_{h \in L_m(\mathbb{R})} \int h(x)(y - x) \pi_m(dx, dy) + H(\pi_m|\mu \otimes \nu) \\ &\geq H(\pi_m|\mu \otimes \nu). \end{aligned}$$

Using the lower semicontinuity of $\pi \mapsto H(\pi|\mu \otimes \nu)$, we conclude that

$$\begin{aligned} & H(\hat{\pi}|\mu \otimes \nu) \leq \liminf_{m \rightarrow \infty} H(\pi_m|\mu \otimes \nu) \\ (40) \quad &\stackrel{(39)}{\leq} \liminf_{m \rightarrow \infty} \int h_m(x)(y - x) \pi_m(dx, dy) + H(\pi_m|\mu \otimes \nu) \\ &\stackrel{(17), (16)}{=} \inf_{\pi \in \widehat{\mathcal{M}}(\mu, \nu)} H(\pi|\mu \otimes \nu) \end{aligned}$$

and hence $\hat{\pi} = \hat{\pi}$.

Step 2. Our next goal is to show that $(f_m^+, g_m^+)_m$ is $(\mu \otimes \nu)$ -uniformly integrable. Note that the Schrödinger system (22) can be stated as

$$\begin{aligned} & \int e^{f_m(x) - h_m(x)(y - x)} \mu(dx) = e^{-g_m(y)} \quad \forall y \in \mathbb{R}, \\ (41) \quad & \int e^{g_m(y) - h_m(x)y} \nu(dy) = e^{-f_m(x) - h_m(x)x} \quad \forall x \in \mathbb{R}. \end{aligned}$$

As μ is centered and $\mu \neq \delta_0$, we have $\mu((-\infty, 0)) \wedge \mu((0, \infty)) > 0$. After taking subsequences, Lemmas 3.12 and 3.11 yield numbers $C < 0 < D$ such that $(f_m(C), h_m(C), f_m(D), h_m(D))$ is bounded uniformly in $m \in \mathbb{N}$. Recall from Assumption 2.3 that $\nu((-\infty, s_-]) > 0$ and $\nu([s_+, \infty)) > 0$, where s_-, s_+ are the left and right endpoints of the support of μ . Applying again Lemmas 3.12 and 3.11, and taking another subsequence, there exist numbers $E \leq s_- < s_+ \leq F$ such that $(g_m(E), g_m(F))$ is bounded uniformly in $m \in \mathbb{N}$.

Recalling that $h_m(x)x \leq 0$, we have $h_m(C) \geq 0$ and $h_m(D) \leq 0$. Using also that $\text{sign}(E-x)\text{sign}(F-x) \leq 0$ for $x \in [E, F]$, we obtain

$$\begin{aligned} e^{f_m^+(x)} &\leq 1 + e^{f_m(x)-h_m(x)(E-x)} + e^{f_m(x)-h_m(x)(F-x)} & \forall x \in [E, F], \\ e^{g_m^+(y)} &\leq 1 + e^{g_m(y)-h_m(C)y} + e^{g_m(y)-h_m(D)y} & \forall y \in \mathbb{R} \end{aligned}$$

for all $m \in \mathbb{N}$. Combining this with (41),

$$\begin{aligned} \int e^{f_m^+(x)} \mu(dx) &\leq \int [1 + e^{f_m(x)-h_m(x)(E-x)} + e^{f_m(x)-h_m(x)(F-x)}] \mu(dx) \\ &\leq 1 + e^{-g_m(E)} + e^{-g_m(F)} \end{aligned}$$

and

$$\begin{aligned} \int e^{g_m^+(y)} \nu(dy) &\leq \int [1 + e^{g_m(y)-h_m(C)y} + e^{g_m(y)-h_m(D)y}] \nu(dy) \\ &\leq 1 + e^{-f_m(C)-h_m(C)C} + e^{-f_m(D)-h_m(D)D}. \end{aligned}$$

Recalling boundedness of $(f_m(C), h_m(C), f_m(D), h_m(D), g_m(E), g_m(F))$ and the de la Vallée-Poussin theorem, we have shown that $(f_m^+, g_m^+)_m$ is $(\mu \otimes \nu)$ -uniformly integrable.

Step 3. As $(h_m(x), x)_*\mu \rightarrow (\hat{h}(x), x)_*\mu$, we also have $(xh_m(x))_*\mu \rightarrow (x\hat{h}(x))_*\mu$ and thus

$$(42) \quad \mu(\{x : x\hat{h}(x) \leq 0\}) \geq \limsup_{m \rightarrow \infty} \mu(\{x : xh_m(x) \leq 0\}) = 1$$

by the Portmanteau theorem. Recalling that $\hat{\pi} = \hat{\pi}$ is a probability measure,

$$(43) \quad \int e^{\hat{f}(x)+\hat{g}(y)-\hat{h}(x)(y-x)} \nu(dy) \mu(dx) = 1.$$

Combining (42) and (43) with Lemma 3.3 yields

$$\begin{aligned} \int \hat{f}(x) \mu(dx) + \int \hat{g}(y) \nu(dy) &\leq \sup_{h \in C_b(\mathbb{R}), xh(x) \leq 0} \sup_{f, g \in C_b(\mathbb{R})} \int f(x) \mu(dx) + \int g(y) \nu(dy) \\ &\quad - \int e^{f(x)+g(y)-h(x)(y-x)} \nu(dy) \mu(dx) + 1 \\ &= H(\hat{\pi}|\mu \otimes \nu). \end{aligned}$$

On the other hand, by (40) and (21) we have

$$\begin{aligned} H(\hat{\pi}|\mu \otimes \nu) &\leq \liminf_{m \rightarrow \infty} \int h_m(x)(y-x) \pi_m(dx, dy) + H(\pi_m|\mu \otimes \nu) \\ &= \liminf_{m \rightarrow \infty} \int f_m(x) \mu(dx) + \int g_m(y) \nu(dy) \\ &\leq \limsup_{m \rightarrow \infty} \int f_m(x) \mu(dx) + \limsup_{m \rightarrow \infty} \int g_m(y) \nu(dy). \end{aligned}$$

The weak convergence $(f_m)_*\mu \rightarrow \hat{f}_*\mu$ and $(g_m)_*\nu \rightarrow \hat{g}_*\nu$, together with the $(\mu \otimes \nu)$ -uniform integrability of (f_m^+, g_m^+) and the Portmanteau theorem, yields $(\hat{f}^+, \hat{g}^+) \in L^1(\mu) \times L^1(\nu)$ and

$$\limsup_{m \rightarrow \infty} \int f_m(x) \mu(dx) + \limsup_{m \rightarrow \infty} \int g_m(y) \nu(dy) \leq \int \hat{f}(x) \mu(dx) + \int \hat{g}(y) \nu(dy).$$

Putting the last three displays together, we see that equality holds throughout. In particular, this yields

$$(44) \quad H(\hat{\pi}|\mu \otimes \nu) = \int \hat{f}(x) \mu(dx) + \int \hat{g}(y) \nu(dy)$$

and thus $(\hat{f}, \hat{g}) \in L^1(\mu) \times L^1(\nu)$.

In general, if $\pi \in \Pi(\mu, \nu)$ satisfies $\pi \ll \mu \otimes \nu$, then $\log(d\pi/d(\mu \otimes \nu))^- \in L^1(\pi)$ as $x \log x$ is bounded from below. If $H(\pi|\mu \otimes \nu) = \int \log(d\pi/d(\mu \otimes \nu)) d\pi < \infty$, the positive part is also integrable, so that $\log(d\pi/d(\mu \otimes \nu)) \in L^1(\pi)$. Applying this to $\hat{\pi} = \hat{\pi}$ and using (31), we see that $\hat{f}(x) + \hat{g}(y) - \hat{h}(x)(y - x) \in L^1(\hat{\pi})$, and as $(\hat{f}, \hat{g}) \in L^1(\mu) \times L^1(\nu)$, it follows that $\hat{h}(x)(y - x) \in L^1(\hat{\pi})$. In particular, we can write $H(\hat{\pi}|\mu \otimes \nu) = \int \log(d\hat{\pi}/d(\mu \otimes \nu)) d\hat{\pi}$ as

$$H(\hat{\pi}|\mu \otimes \nu) = \int \hat{f}(x) \mu(dx) + \int \hat{g}(y) \nu(dy) - \int \hat{h}(x)(y - x) \hat{\pi}(dx, dy)$$

and conclude by (44) that $\int \hat{h}(x)(y - x) \hat{\pi}(dx, dy) = 0$. \square

Lemma 3.17. *We have $\hat{\pi} \in \mathcal{M}(\mu, \nu)$.*

Proof. As $\hat{\pi} \in \widehat{\mathcal{M}}(\mu, \nu)$, we have $\int (y - x) \hat{\pi}_x(dy) \geq (\leq) 0$ for μ -a.e. $x \geq (\leq) 0$. Suppose for contradiction that $\hat{\pi} \notin \mathcal{M}(\mu, \nu)$; that is, $A = \{x : \int (y - x) \hat{\pi}_x(dy) \neq 0\}$ satisfies $\mu(A) > 0$. As μ, ν are both centered, $A_{\pm} := \{x : \int (y - x) \hat{\pi}_x(dy) > (<) 0\}$ then both have positive μ -mass. Moreover, we must have $A_{\pm} \subseteq \mathbb{R}_{\pm}$ μ -a.s. In particular, there exists a non-negligible set of points $x_2 \leq 0 \leq x_1$ for which

$$(45) \quad \int (y - x_1) \hat{\pi}_{x_1}(dy) > 0 \quad \text{and} \quad \int (y - x_2) \hat{\pi}_{x_2}(dy) < 0.$$

As

$$\int \hat{h}(x) \int (y - x) \hat{\pi}_x(dy) \mu(dx) = \int \hat{h}(x)(y - x) \hat{\pi}(dx, dy) = 0$$

by Lemma 3.16, we obtain that $\hat{h}(x_1) = \hat{h}(x_2) = 0$ μ -a.s., so that (45) can be stated as

$$(46) \quad \int y e^{\hat{f}(x_1) + \hat{g}(y)} \nu(dy) > x_1 \quad \text{and} \quad \int y e^{\hat{f}(x_2) + \hat{g}(y)} \nu(dy) < x_2.$$

Combining (46) and $x_2 \leq 0 \leq x_1$ we obtain

$$(47) \quad \int y e^{\hat{f}(x_2) + \hat{g}(y)} \nu(dy) < x_2 \leq x_1 < \int y e^{\hat{f}(x_1) + \hat{g}(y)} \nu(dy)$$

and in particular $\hat{f}(x_1) \neq \hat{f}(x_2)$. On the other hand,

$$1 = \int e^{\hat{f}(x_i) + \hat{g}(y) - \hat{h}(x_i)(y - x_i)} \nu(dy) = \int e^{\hat{f}(x_i) + \hat{g}(y)} \nu(dy)$$

for $i = 1, 2$ implies $\hat{f}(x_1) = \hat{f}(x_2)$, a contradiction. \square

Proof of Theorem 2.4: Existence. As $\hat{\pi} \in \mathcal{M}(\mu, \nu)$ by Lemma 3.17, it follows from (14) that $\pi^* = \hat{\pi}$. The density of that coupling has the desired form and properties by Lemmas 3.15 and 3.16, except that it remains to verify $\hat{h}(x)(y - x) \in L^1(\pi)$ for all $\pi \in \mathcal{M}(\mu, \nu)$ satisfying $H(\pi|\mu \otimes \nu) < \infty$. Indeed, given such π , [24, Lemma 1.4] yields that $\log(d\pi^*/d(\mu \otimes \nu))^+ \in L^1(\pi)$. As $\log(d\pi^*/d(\mu \otimes \nu)) = \hat{f}(x) + \hat{g}(y) - \hat{h}(x)(y - x)$ with $(\hat{f}, \hat{g}) \in L^1(\mu) \times L^1(\nu)$, it follows that $(\hat{h}(x)(y - x))^- \in L^1(\pi)$. Now the martingale property of π and the Fubini–Tonelli

theorem for kernels imply that $\int \hat{h}(x)(y-x) \pi(dx, dy) = 0$ and in particular $\hat{h}(x)(y-x) \in L^1(\pi)$. \square

It remains to show the uniqueness part of Theorem 2.4. Given potentials (f, g, h) and constants $\alpha, \beta \in \mathbb{R}$, the functions $f'(x) := f(x) + \alpha x + \beta$, $g'(y) := g(y) - \alpha y - \beta$ and $h'(x) := h(x) - \alpha$ satisfy $f'(x) + g'(y) - h'(x)(y-x) = f(x) + g(y) - h(x)(y-x)$, so that (f', g', h') are again potentials. The next result implies that this affine shift is the only source of non-uniqueness, and completes the proof of Theorem 2.4.

Lemma 3.18. *Let $\nu \neq \delta_0$ and $\pi \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$. Suppose that*

$$\frac{d\pi}{d(\mu \otimes \nu)}(x, y) = e^{f(x)+g(y)-h(x)(y-x)} = e^{f'(x)+g'(y)-h'(x)(y-x)} \quad \pi\text{-a.s.}$$

for some functions $f, f', g, g', h, h' : \mathbb{R} \rightarrow \mathbb{R}$. Then there are $\alpha, \beta \in \mathbb{R}$ such that

$$f'(x) - f(x) = \alpha x + \beta \quad \mu\text{-a.s.}, \quad g(y) - g'(y) = \alpha y + \beta \quad \nu\text{-a.s.}, \quad h - h' = \alpha \quad \mu\text{-a.s.}$$

Proof. Note that $\pi \sim \mu \otimes \nu$. Thus we have

$$(48) \quad f(x) + g(y) - h(x)(y-x) = f'(x) + g'(y) - h'(x)(y-x), \quad (x, y) \in A$$

for some $A \in \mathcal{B}(\mathbb{R} \times \mathbb{R})$ with $(\mu \otimes \nu)(A) = 1$. In general, given any $A \in \mathcal{B}(\mathbb{R} \times \mathbb{R})$ with $(\mu \otimes \nu)(A) = 1$, there exist a set $A_0 \subseteq A$ with $(\mu \otimes \nu)(A_0) = 1$ and a point $(x_0, y_0) \in A_0$ such that for all $(x, y) \in A_0$, we have $(x_0, y) \in A_0$ and $(x, y_0) \in A_0$. The construction can be found in the proof of [2, Lemma 4.3] or, more explicitly, in [16, Lemma 4.5]. Fixing $x = x_0$ in (48), it follows for all y with $(x_0, y) \in A_0$ (and hence ν -a.s.) that

$$g(y) - g'(y) = f'(x_0) - f(x_0) + h(x_0)(y - x_0) - h'(x_0)(y - x_0).$$

In particular, $g(y) - g'(y) = \alpha y + \beta$ for some $\alpha, \beta \in \mathbb{R}$. We deduce

$$f'(x) - f(x) + [h(x) - h'(x)](y-x) = \alpha y + \beta \quad (\mu \otimes \nu)\text{-a.s.}$$

As the support of ν includes at least two points, it follows that $h - h' = \alpha$ μ -a.s. Then, we also see that $f' - f - \alpha x = \beta$ μ -a.s. \square

Finally, we show the weak duality that was used in the proof of Corollary 2.5.

Lemma 3.19 (Weak duality). *Let $(\tilde{f}, \tilde{g}, \tilde{h}) \in \mathcal{D}$ and $\pi \in \mathcal{M}(\mu, \nu)$. Then*

$$H(\pi | \mu \otimes \nu) \geq \int \tilde{f}(x) \mu(dx) + \int \tilde{g}(y) \nu(dy) - \log \int e^{\tilde{f}(x)+\tilde{g}(y)-\tilde{h}(x)(y-x)} \mu(dx) \nu(dy).$$

Proof. Let $(\tilde{f}, \tilde{g}, \tilde{h}) \in \mathcal{D}$ and $\pi \in \mathcal{M}(\mu, \nu)$. We may assume that $H(\pi | \mu \otimes \nu) < \infty$, so that $\varphi := \frac{d\pi}{d(\mu \otimes \nu)}$ exists with values in $[0, \infty)$. We can write

$$\begin{aligned} \int e^{\tilde{f}(x)+\tilde{g}(y)-\tilde{h}(x)(y-x)} \mu(dx) \nu(dy) &\geq \int e^{\tilde{f}(x)+\tilde{g}(y)-\tilde{h}(x)(y-x)} \mathbf{1}_{\{\varphi > 0\}} \mu(dx) \nu(dy) \\ &= \int e^{\tilde{f}(x)+\tilde{g}(y)-\tilde{h}(x)(y-x) - \log \varphi(x,y)} \pi(dx, dy). \end{aligned}$$

Hence by Jensen's inequality,

$$\begin{aligned}
& -\log \int e^{\tilde{f}(x)+\tilde{g}(y)-\tilde{h}(x)(y-x)} \mu(dx) \nu(dy) \\
& \leq \int [-\tilde{f}(x) - \tilde{g}(y) + \tilde{h}(x)(y-x) + \log \varphi(x, y)] \pi(dx, dy) \\
& = -\int \tilde{f}(x) \mu(dx) - \int \tilde{g}(y) \nu(dy) + H(\pi|\mu \otimes \nu),
\end{aligned}$$

where the last equality used $\pi \in \mathcal{M}(\mu, \nu)$ and the integrability properties of $(\tilde{f}, \tilde{g}, \tilde{h}) \in \mathcal{D}$ under $\pi \in \mathcal{M}_{\text{fin}}(\mu, \nu)$, as well as the definition of $H(\pi|\mu \otimes \nu)$. \square

4. CLOSING REMARKS

We conclude with an example regarding Remark 2.2 and Assumption 2.3. While the assumption is satisfied in the example, it shows that the integrability of h is delicate when the potentials functions φ_μ and φ_ν are not uniformly separated.

Example 4.1. Let $b \geq a := 1/2$ and $\nu = (\delta_{-b} + \delta_b)/2$. Moreover, let $\mu \preceq_c \nu$ satisfy $\text{supp}(\mu) = [-a, a]$ and $\mu(\{\pm a\}) = 0$. There exists a unique martingale coupling $\pi \in \mathcal{M}(\mu, \nu)$; namely, $\pi = \mu(dx) \otimes \pi_x(dy)$ for

$$\pi_x = \frac{b-x}{2b} \delta_{-b} + \frac{b+x}{2b} \delta_b.$$

In particular, Assumption 2.1 is satisfied. Direct calculation yields that $\frac{d\pi}{d(\mu \otimes \nu)}(x, y) = \exp[f(x) + g(y) - h(x)(y-x)]$ for

$$f(x) = \log\left(\frac{b-x}{2b}\right) - (b+x) \log\left(\frac{b-x}{b+x}\right), \quad g(y) = 0, \quad h(x) = \frac{1}{2b} \log\left(\frac{b-x}{b+x}\right).$$

These functions are unique up to an affine transformation. The functions f and g are always bounded, whereas the function h is bounded on $\text{supp}(\mu)$ when $b > a$ but unbounded for $b = a$.

Focusing on the boundary case $b = a$, then more specifically, h is unbounded from below and from above, while $c(x, y) := h(x)(y-x)$ is bounded from below but unbounded from above. This leads to ce^{-c} being bounded and in particular $c \in L^\infty(\pi)$. Assumption 2.3 is satisfied due to the atoms of ν , hence $c \in L^1(\pi)$ also follows from Theorem 2.4.

Next, we specialize to μ defined by

$$\frac{d\mu}{dx} = \frac{C}{(a-|x|) \log(a-|x|)^2} \mathbf{1}_{[-a, a]}(x),$$

where $C = -\log(a)/2$ is the normalizing constant. Then we have $h \notin L^1(\mu)$ as

$$\begin{aligned}
\int |h(x)| \mu(dx) &= -2 \int_0^a \frac{1}{2a} \log\left(\frac{a-x}{a+x}\right) \frac{C}{(a-x) \log(a-x)^2} dx \\
&\geq -\frac{1}{a} \int_0^a \log\left(\frac{a-x}{a}\right) \frac{C}{(a-x) \log(a-x)^2} dx \\
&= \frac{1}{a} \log(a) \int_0^a \frac{C}{(a-x) \log(a-x)^2} dx - \frac{1}{a} \int_0^a \frac{C}{(a-x) \log(a-x)} dx \\
&= \frac{\log(a)}{2a} + \infty,
\end{aligned}$$

and a similar calculation shows that $c(x, y) := h(x)(x - y) \notin L^1(\mu \otimes \nu)$.

In the case $b = a$, the potential functions φ_μ and φ_ν (cf. Remark 2.2) are not uniformly separated. While the example is covered by Assumption 2.3 due to the presence of atoms, the boundary case illustrates that the integrability of the potentials can be delicate.

REFERENCES

- [1] B. Acciaio, M. Larsson, and W. Schachermayer. The space of outcomes of semi-static trading strategies need not be closed. *Finance Stoch.*, 21(3):741–751, 2017.
- [2] M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer. Optimal and better transport plans. *J. Funct. Anal.*, 256(6):1907–1927, 2009.
- [3] 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.
- [4] M. Beiglböck, B. Jourdain, W. Margheriti, and G. Pammer. Approximation of martingale couplings on the line in the adapted weak topology. *Probab. Theory Related Fields*, 183(1-2):359–413, 2022.
- [5] M. Beiglböck, M. Nutz, and N. Touzi. Complete duality for martingale optimal transport on the line. *Ann. Probab.*, 45(5):3038–3074, 2017.
- [6] S. Chewi and A.-A. Pooladian. An entropic generalization of Caffarelli’s contraction theorem via covariance inequalities. *C. R. Math. Acad. Sci. Paris*, 361:1471–1483, 2023.
- [7] I. Csiszár. I -divergence geometry of probability distributions and minimization problems. *Ann. Probab.*, 3:146–158, 1975.
- [8] H. De March and P. Henry-Labordère. Building arbitrage-free implied volatility: Sinkhorn’s algorithm and variants. *Preprint SSRN:3326486*, 2019.
- [9] A. Doldi and M. Frittelli. Entropy martingale optimal transport and nonlinear pricing-hedging duality. *Finance Stoch.*, 27(2):255–304, 2023.
- [10] A. Doldi, M. Frittelli, and E. Rosazza Gianin. On entropy martingale optimal transport theory. *Preprint*, 2023.
- [11] R. Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fourth edition, 2010.
- [12] H. Föllmer. Random fields and diffusion processes. In *École d’Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, volume 1362 of *Lecture Notes in Math.*, pages 101–203. Springer, Berlin, 1988.
- [13] H. Föllmer and N. Gantert. Entropy minimization and Schrödinger processes in infinite dimensions. *Ann. Probab.*, 25(2):901–926, 1997.
- [14] 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.
- [15] A. Genevay, L. Chizat, F. Bach, M. Cuturi, and G. Peyré. Sample complexity of Sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1574–1583. PMLR, 2019.
- [16] P. Ghosal, M. Nutz, and E. Bernton. Stability of entropic optimal transport and Schrödinger bridges. *J. Funct. Anal.*, 283(9):Paper No. 109622, 2022.
- [17] G. Guo and J. Oblój. Computational methods for martingale optimal transport problems. *Ann. Appl. Probab.*, 29(6):3311–3347, 2019.
- [18] J. Guyon. The joint S&P 500/VIX smile calibration puzzle solved. *Risk*, 2020.
- [19] J. Guyon. Dispersion-constrained martingale Schrödinger problems and the exact joint S&P 500/VIX smile calibration puzzle. *Preprint SSRN:3853237*, 2021.
- [20] P. Henry-Labordère. From (martingale) Schrödinger bridges to a new class of stochastic volatility model. *Preprint SSRN:3353270*, 2019.
- [21] D. Hobson. Robust hedging of the lookback option. *Finance Stoch.*, 2(4):329–347, 1998.
- [22] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.
- [23] M. Liero, A. Mielke, and G. Savaré. Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures. *Invent. Math.*, 211(3):969–1117, 2018.
- [24] M. Nutz. *Introduction to Entropic Optimal Transport*. Lecture notes, Columbia University, 2021. https://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf.

- [25] M. Nutz and J. Wiesel. Entropic optimal transport: convergence of potentials. *Probab. Theory Related Fields*, 184(1-2):401–424, 2022.
- [26] M. Nutz, J. Wiesel, and L. Zhao. Limits of semistatic trading strategies. *Math. Finance*, 33(1):185–205, 2023.
- [27] M. Nutz, J. Wiesel, and L. Zhao. Martingale Schrödinger bridges and optimal semistatic portfolios. *Finance Stoch.*, 27(1):233–254, 2023.
- [28] G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
- [29] L. Rüschendorf and W. Thomsen. Note on the Schrödinger equation and I -projections. *Statist. Probab. Lett.*, 17(5):369–375, 1993.
- [30] L. Rüschendorf and W. Thomsen. Closedness of sum spaces and the generalized Schrödinger problem. *Teor. Veroyatnost. i Primenen.*, 42(3):576–590, 1997.
- [31] E. Schrödinger. Über die Umkehrung der Naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Akad. Wiss., Berlin. Phys. Math.*, 144:144–153, 1931.
- [32] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
- [33] J. Wiesel. Continuity of the martingale optimal transport problem on the real line. *Ann. Appl. Probab.*, 33(6A):4645–4692, 2023.

(MN) DEPARTMENTS OF STATISTICS AND MATHEMATICS, COLUMBIA UNIVERSITY, 1255 AMSTERDAM AVENUE, NEW YORK, NY 10027, USA

Email address: `mnutz@columbia.edu`

(JW) DEPARTMENT OF MATHEMATICS, CARNEGIE MELLON UNIVERSITY, WEAN HALL, PITTSBURGH, PA 15213, USA

Email address: `wiesel@cmu.edu`