Monotonicity and stability of the Weak Martingale Optimal Transport problem

M. Beiglböck* B. Jourdain† W. Margheriti†, ‡ G. Pammer*, §

October 24, 2020

Abstract

Motivated by the study of the concentration of measure phenomenon and their connection with transport-entropy inequalities, Gozlan, Roberto, Samson and Tetali [20] introduced a general notion of transport cost problem called the Optimal Weak Transport (WOT) problem. Backhoff-Veraguas, Beiglböck and Pammer [6] and Backhoff-Veraguas and Pammer [7] proved that under some regularity assumption on the cost function, existence, uniqueness and stability hold for the WOT problem. Thanks to a result proved in the companion paper [10], we recover those result in a different way.

Because the martingale constraint reflects the condition for a financial market to be arbitrage free, it is natural in the context of mathematical finance to consider the martingale counterpart of the WOT problem, namely the Martingale Optimal Weak Transport (WMOT) problem. Thanks to the main theorem of the companion paper, we prove the existence, the uniqueness and most importantly the stability of the WMOT problem under mild regularity assumption of the cost function.

We also prove that martingale C-monotonicity is sufficient for optimality of the WMOT problem, that the so called Wasserstein projections are Lipschitz continuous in dimension 1 and finally we establish the convergence in a space of extended martingale couplings. We discuss a consequence on the superreplication bound for VIX futures.

Keywords: Martingale Optimal Transport, Adapted Wasserstein distance, Robust finance, Weak transport, Stability, Convex order, Martingale couplings.

1 Introduction and motivations

1.1 The Weak Optimal Transport problem

Motivated by the study of the concentration of measure phenomenon and their connection with transport-entropy inequalities, whose extensive study can be found in [25, 18, 14], Gozlan, Roberto, Samson and Tetali [20] introduced a general notion of transport cost problem called the Optimal Weak Transport (WOT) problem, which they studied with Shu in [19]. In order to define it we introduce some notation. Let X and Y be two Polish spaces respectively endowed with the compatible and complete metrics d_X and d_Y. Let μ be in the set $\mathcal{P}(X)$ of probability measures on X, $\nu \in \mathcal{P}(Y)$ and $C : X \times \mathcal{P}(Y) \to \mathbb{R}_+$ be a nonnegative measurable function. We denote by $\Pi(\mu, \nu)$ the set of couplings between μ and ν, that is $\pi \in \Pi(\mu, \nu)$ iff $\pi \in \mathcal{P}(X \times Y)$ is such that for any measurable subsets $A \subset X$ and $B \subset Y$, $\pi(A \times Y) = \mu(A)$ and $\pi(X \times B) = \nu(B)$. Then the WOT problem consists in the minimisation

$$V_C(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x, \pi_x)\, \mu(dx), \tag{WOT}$$

*University of Vienna, Austria. Emails: mathias.beiglboeck@univie.ac.at, gudmund.pammer@univie.ac.at

†Université Paris-Est, Cermics (ENPC), INRIA F-77455 Marne-la-Vallée, France. E-mails: benjamin.jourdain@enpc.fr, william.margheriti@enpc.fr

‡acknowledges support from the "Chaire Risques Financiers", Fondation du Risque.

§acknowledges support from the Austrian Science Fund (FWF) through grant number W1245.

where for all $\pi \in \Pi(\mu,\nu)$, $(\pi_x)_{x\in\mathbb{R}}$ denotes a disintegration of π with respect to its first marginal, which we write $\pi(dx,dy) = \mu(dx)\,\pi_x(dy)$, or with a slight abuse of notation, $\pi = \mu \times \pi_x$ if the context is not ambiguous. Note that for a measurable map $c : X \times Y \to \mathbb{R}_+$, the WOT problem with the cost function $C : (x,p) \mapsto \int_Y c(x,y)\,p(dy)$ amounts to the classical Optimal Transport (OT) problem already discussed in the companion paper [10]. In particular it was mentioned that the OT theory covers an impressive range of applications. This particularity seems to be shared with the WOT problem, which benefits of high flexibility. One could for instance see the recent work of Backhoff-Veraguas and Pammer [8] and the references inside for an investigation of a connection of the WOT problem with the Schrödinger problem, the Brenier-Strassen Theorem, optimal mechanism design, linear transfers and semimartingale transport.

In order to gain some insight on the WOT problem, we recall some results of paramount importance, namely existence, uniqueness and stability. To formulate those results we need to introduce a more specific setting. From now on, we fix $r \geq 1$ and x_0, y_0 two arbitrary elements of X and Y respectively, their specific value having no impact on our study. Let $\mathcal{P}^r(X)$ denote the set of all probability measures on X with finite r-th moment, i.e.

$$\mathcal{P}^r(X) = \left\{ p \in \mathcal{P}(X) \mid \int_X d_X^r(x,x_0)\,p(dx) < +\infty \right\}. \tag{1.1}$$

Let $\mathcal{C}(X)$ denote the set of all real-valued continuous functions on X. The set $\mathcal{P}^r(X)$ is equipped with the weak topology induced by

$$\Phi^r(X) = \left\{ f \in C(X) \mid \exists \alpha > 0,\ \forall x \in X,\ |f(x)| \leq \alpha(1 + d_X^r(x,x_0)) \right\}.$$

Then a sequence $(p^k)_{k\in\mathbb{N}}$ converges in $\mathcal{P}^r(X)$ to p iff

$$\forall g \in \Phi^r(X), \quad p^k(g) := \int_X g(x)\,p^k(dx) \underset{k\to+\infty}{\longrightarrow} p(g) := \int_X g(x)\,p(dx).$$

It is well known that the latter is equivalent to

$$\mathcal{W}_r(p^k,p) := \inf_{\pi\in\Pi(p^k,p)} \left(\int_{X\times X} d_X^r(x,y)\,\pi(dx,dy) \right)^{\frac{1}{r}} \underset{k\to+\infty}{\longrightarrow} 0,$$

where $\mathcal{W}_r$ is the Wasserstein distance with index r, which is a metric on $\mathcal{P}^r(X)$ compatible with its topology, turning $\mathcal{P}^r(X)$ into a complete separable metric space, see [3, 26, 28, 29] for much more details.

Back to the WOT problem, consider a cost function $C : X \times \mathcal{P}^r(Y) \to \mathbb{R}_+$ being lower semicontinuous and convex in its second argument. Backhoff-Veraguas, Beiglböck and Pammer [6] prove that for all $(\mu,\nu) \in \mathcal{P}(X) \times \mathcal{P}^r(Y)$, the infimum $V_C(\mu,\nu)$ is attained. Moreover, the map $(\mu,\nu) \mapsto V_C(\mu,\nu)$ is lower semicontinuous. Backhoff-Veraguas and Pammer [7] also prove the stability of the WOT problem: suppose that $C \in \Phi^r(X \times \mathcal{P}^r(Y))$ and let $\mu^k \in \mathcal{P}(X), \nu^k \in \mathcal{P}^r(Y)$, $k \in \mathbb{N}$ converge respectively weakly to $\mu \in \mathcal{P}(X)$ and in $\mathcal{W}_r$ to $\nu \in \mathcal{P}^r(Y)$ as k goes to $+\infty$. For $k \in \mathbb{N}$, let $\pi^k \in \Pi(\mu^k,\nu^k)$ be optimal for $V(\mu^k,\nu^k)$. Then any accumulation point of $(\pi^k)_{k\in\mathbb{N}}$ for the weak convergence topology is a minimiser of $V_C(\mu,\nu)$. If the latter has a unique minimiser π^*, which happens for instance if C is strictly convex in its second argument, then π^k converges weakly to π^* as k goes to $+\infty$. In the latter case, when μ^k, μ belong to $\mathcal{P}^r(X)$ and the convergence of μ^k to μ holds in $\mathcal{W}_r$, then one can easily show that

$$\pi^k \underset{k\to+\infty}{\longrightarrow} \pi^* \text{ in } \mathcal{W}_r. \tag{1.2}$$

However, the topology induced by the Wasserstein distance in not always well suited for any setting, especially in mathematical finance, since its symmetry does not take into account the temporal structure of martingales. As explained in the companion paper [10], it is sometimes necessary to strengthen the usual topology and therefore consider the adapted Wasserstein distance $\mathcal{AW}_r$ of index r defined for all couplings $\pi = \mu \times \pi_x, \pi' = \mu' \times \pi'_x \in \mathcal{P}(X \times Y)$ by

$$\mathcal{AW}_r(\pi,\pi') = \inf_{\chi\in\Pi(\mu,\mu')} \left(\int_{X\times X} \left(d_X^r(x,x') + \mathcal{W}_r^r(\pi_x,\pi'_{x'}) \right) \chi(dx,dx') \right)^{\frac{1}{r}}.$$

It will prove important to observe that

$$\mathcal{AW}_r(\pi,\pi') = \mathcal{W}_r(J(\pi), J(\pi')), \tag{1.3}$$

where J is the trivial embedding map from $\mathcal{P}(X \times Y)$ to $\mathcal{P}(X \times \mathcal{P}(Y))$, namely

$$J : \mathcal{P}(X \times Y) \ni \pi = \mu \times \pi_x \mapsto \mu(dx)\,\delta_{\pi_x}(dp) \in \mathcal{P}(X \times \mathcal{P}(Y)). \tag{1.4}$$

There exist other ways to adapt the usual weak topoly: Hellwig's information topology [23], Aldous's extended weak topology [1] or the optimal stopping topology [4]. But we do not lose generality by using the topology induced by the adapted Wasserstein distance since strikingly, all those apparently independent topologies are actually equal, at least in discrete time [4, Theorem 1.1]. By the connection Backhoff-Veraguas and Pammer establish between the WOT problem and an extended version of it, in the setting of (1.2) when C is strictly convex in its second argument, we can derive the convergence of $J(\pi^k)$ to $J(\pi^*)$ in $\mathcal{W}_r$ as k goes to $+\infty$, which by (1.3) is equivalent to the convergence of the minimiser π^k of $V_C(\mu^k,\nu^k)$ to the only minimiser π^* of $V_C(\mu,\nu)$ in $\mathcal{AW}_r$.

In the companion paper we prove that any coupling whose marginals are approximated by probability measures can be approximated by couplings with respect to the adapted Wasserstein distance (see Proposition 2.3 below). We show in the present paper that this result allows us to recover the existence, uniqueness and most importantly the stability of the WOT problem under less regularity assumption on the cost function.

1.2 The Martingale Weak Optimal Transport problem

The classical OT theory is not sufficient to solve some major problems raised by the field of mathematical finance, such as robust model-independent pricing. Indeed, Beiglböck, Henry-Labordère and Penkner [9] showed in a discrete time setting and Galichon, Henry- Labordère and Touzi [17] in a continuous time setting that one would need an additional martingale constraint to the OT problem in order to get model-free bounds of an option price. This martingale constraint reflects the condition for a financial market to be arbitrage free. This leads to the formulation of the Martingale Optimal Transport (MOT) problem recalled in the companion paper [10]. With regard to this, it is natural to study a martingale counterpart of the WOT problem.

Let $d \in \mathbb{N}^*$, $C : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}_+$ and $\mu,\nu \in \mathcal{P}^1(\mathbb{R}^d)$. Then the Martingale Optimal Weak Transport (WMOT) problem consists in the minimisation

$$V_C^M(\mu,\nu) := \inf_{\pi \in \Pi_M(\mu,\nu)} \int_{\mathbb{R}^d} C(x,\pi_x)\,\mu(dx), \tag{WMOT}$$

where $\Pi_M(\mu,\nu)$ denotes the set of martingale couplings between μ and ν, that is

$$\Pi_M(\mu,\nu) = \left\{ \pi = \mu \times \pi_x \in \Pi(\mu,\nu) \mid \mu(dx)\text{-almost everywhere, } \int_{\mathbb{R}} y\,\pi_x(dy) = x \right\}.$$

According to Strassen's theorem [27], the existence of a martingale coupling between two probability measures $\mu,\nu \in \mathcal{P}^1(\mathbb{R}^d)$ is equivalent to $\mu \le_c \nu$, where $\le_c$ denotes the convex order. We recall that two finite positive measures μ,ν on $\mathbb{R}^d$ with finite first moment are said to be in the convex order iff we have

$$\int_{\mathbb{R}^d} f(x)\,\mu(dx) \le \int_{\mathbb{R}^d} f(y)\,\nu(dy),$$

for every convex function $f : \mathbb{R}^d \to \mathbb{R}$. Note that by evaluating this inequality for the constant function equal to 1, the identity function and their opposites, we have that μ and ν have equal mass and satisfy $\int_{\mathbb{R}^d} x\,\mu(dx) = \int_{\mathbb{R}^d} y\,\nu(dy)$. For a measurable map $c : X \times Y \to \mathbb{R}_+$, the WMOT problem with the cost function $C : (x,p) \mapsto \int_Y c(x,y)\,p(dy)$ amounts to the MOT problem already discussed in the companion paper [10].

The main result of the companion paper is that any martingale coupling whose marginals are approximated by probability measures in the convex order can be approximated by martingale couplings with respect to the adapted Wasserstein distance, see Theorem 2.4 below. Similarly as for the WOT problem, we prove in the present paper thanks to the latter theorem the existence, the uniqueness and most importantly the stability of the WMOT problem under reasonable regularity assumptions on the cost function.

In particular we recover the stability of the MOT problem proved by Backhoff-Veraguas and Pammer [7]. To do so, they used the tool of martingale C-monotonicity by proving that it was a stable necessary optimality criterion. However the question remained open whether any martingale coupling satisfying this condition is optimal. We show here that it is indeed the case under mild regularity assumptions on the cost function.

1.3 Outline

We state in Section 2 the main results of the present paper, namely the stability of the WOT and the WMOT problems, the sufficient optimality criterion of martingale C-monotonicity for the WMOT problem, the Lipschitz continuity of the so called Wasserstein projections and the convergence of an extended space of martingale couplings. We also connect this work with an application on the superreplication bound for VIX futures. Those results have their own devoted section. Section 3 consists of the unified proof of the stability of the WOT and the WMOT problems. Section 4 consists in showing that martingale C-monotonicity is sufficient for optimality for the WMOT problem. Finally Section 5 is an appendix which gathers the proofs of useful lemmas.

2 Main results

2.1 An extension of the weak and adapted topologies

For $r \geq 1$, the Wasserstein distance $\mathcal{W}_r$ is widely used to measure the distance between two probability measures with finite r-th moment. In order to measure the distance between two couplings, one could also use the stronger adapted Wasserstein distance for reasons discussed above. Despite being very handy, those distances sometimes lack topological convenience. For example, the $\mathcal{W}_r$-balls $\{p \in \mathcal{P}^r(X) \mid \mathcal{W}_r(p, \delta_{x_0}) \leq R\}$, $R > 0$, are not compact for the $\mathcal{W}_r$-distance topology. This observation is not without consequences since it stood in the way of our proof that martingale C-monotonicity is a sufficient optimality criterion for the WMOT problem (see Section 2.4 below).

In order to overcome that hurdle, we choose in the present paper to work in a finer topology which benefits of more convenient and flexible properties. We give the definition here as well as some insight to understand its basic properties. All proofs and technical details are deferred to Section 5.1 below.

Definition 2.1. Let $f\colon X \to [1, +\infty)$ be continuous. We consider the space

$$\mathcal{P}_f(X) = \{p \in \mathcal{P}(X) \mid p(f) < +\infty\}.$$

We equip $\mathcal{P}_f(X)$ with the topology induced by the following convergence: a sequence $(p_k)_{k\in\mathbb{N}} \in \mathcal{P}_f(X)^{\mathbb{N}}$ converges in $\mathcal{P}_f(X)$ to p iff one of the two following assertions is satisfied:

(i) $p_k \underset{k\to+\infty}{\longrightarrow} p$ in $\mathcal{P}(X)$ and $p_k(f) \underset{k\to+\infty}{\to} p(f)$.

(ii) $p_k(h) \underset{k\to+\infty}{\longrightarrow} p(h)$ for all $h \in \Phi_f(X) := \{h \in \mathcal{C}(X) \mid \exists \alpha > 0,\ \forall x \in X,\ |h(x)| \leq \alpha f(x)\}$.

Unless explicitly stated otherwise, $\mathcal{P}(X)$ is endowed with the weak convergence topology; for $r \geq 1$, $\mathcal{P}^r(X)$ is endowed with the $\mathcal{W}_r$-distance topology; for $f : X \to [1, +\infty)$ continuous, $\mathcal{P}_f(X)$ is endowed with the topology induced by the convergence (2.1). When f is the map $x \mapsto 1 + d_X^r(x, x_0)$, then $\mathcal{P}_f(X) = \mathcal{P}^r(X)$ and the two topologies match. Hence the reader who is not willing to consider this extension may completely disregard it and consistently view $\mathcal{P}_f(X)$ as the usual Wasserstein space $\mathcal{P}^r(X)$.

We will mainly address convergences of probability measures in terms of topology. However it will sometimes prove useful to consider the metric $\overline{\mathcal{W}}_f$ defined on $\mathcal{P}_f(X)$ by

$$\forall p,q\in\mathcal{P}_f(X),\quad \overline{\mathcal{W}}_f(p,q):=\sup_{\substack{h\colon X\to[-1,1],\\ h\text{ is 1-Lipschitz}}}(p(fh)-q(fh)),\tag{2.1}$$

which is a complete metric compatible with the topology on $\mathcal{P}_f(X)$.

A continuous function $g:Y\to[1,+\infty)$ can naturally be lifted to a continuous function $\hat g:\mathcal{P}_g(Y)\to[1,+\infty)$ by setting

$$\forall p\in\mathcal{P}_f(Y),\quad \hat g(p)=p(g).\tag{2.2}$$

We will often consider probability measures on $\mathcal{P}(Y)$. A very convenient aspect of the extended topology is that the spaces $\mathcal{P}_{\hat g}(\mathcal{P}(Y))$ and $\mathcal{P}_{\hat g}(\mathcal{P}_g(Y))$ and their respective topologies, a priori different, are actually equal. If moreover $\mathcal{P}_g(Y)$ is endowed with the metric $\overline{\mathcal{W}}_g$, then those topological spaces are also equal to $\mathcal{P}^1(\mathcal{P}_g(Y))$, whose definition is given by (1.1) with $(1,\mathcal{P}_g(Y),\overline{\mathcal{W}}_g)$ replacing (r,X,d_X). Therefore one can freely switch between the topological spaces $\mathcal{P}_{\hat g}(\mathcal{P}(Y))$, $\mathcal{P}_{\hat g}(\mathcal{P}_g(Y))$ and $\mathcal{P}^1(\mathcal{P}_g(Y))$.

It is also possible to extend the adapted weak topology, in the spirit of (1.3). Recall the map J defined by (1.4) which embeds $\mathcal{P}(X\times Y)$ into $\mathcal{P}(X\times\mathcal{P}(Y))$. For two real-valued functions f and g respectively defined on X and Y, we denote by $f\oplus g$ the map $X\times Y\ni(x,y)\mapsto f(x)+g(y)$.

Definition 2.2. Let $f:X\to[1,+\infty)$ and $g:Y\to[1,+\infty)$ be continuous. For $k\in\mathbb{N}$, let $\mu^k,\mu\in\mathcal{P}_f(X)$, $\nu^k,\nu\in\mathcal{P}_g(Y)$, $\pi^k\in\Pi(\mu^k,\nu^k)$ and $\pi\in\Pi(\mu,\nu)$. We say that $(\pi^k)_{k\in\mathbb{N}}$ converges in $\mathcal{AW}_{f\oplus\hat g}$ to π if one of the two following equivalent assertions is satisfied:

(i) $J(\pi^k)\underset{k\to+\infty}{\longrightarrow}J(\pi)$ in $\mathcal{P}_{f\oplus\hat g}(X\times\mathcal{P}(Y))$.

(ii) $J(\pi^k)\underset{k\to+\infty}{\longrightarrow}J(\pi)$ in $\mathcal{P}(X\times\mathcal{P}(Y))$, $\mu^k(f)\underset{k\to+\infty}{\longrightarrow}\mu(f)$ and $\nu^k(g)\underset{k\to+\infty}{\longrightarrow}\nu(g)$.

There also holds the convenient fact that $\mathcal{P}_{f\oplus\hat g}(X\times\mathcal{P}(Y))$ and $\mathcal{P}_{f\oplus\hat g}(X\times\mathcal{P}_g(Y))$ and their respective topologies are equal, hence we can rephrase (i) as $J(\pi^k)\underset{k\to+\infty}{\longrightarrow}J(\pi)$ in $\mathcal{P}_{f\oplus\hat g}(X\times\mathcal{P}_g(Y))$. When f and g are respectively the maps $x\mapsto 1+d_X^r(x,x_0)$ and $y\mapsto 1+d_Y^r(y,y_0)$, then $(\pi^k)_{k\in\mathbb{N}}$ converges in $\mathcal{AW}_{f\oplus\hat g}$ to π iff it converges in $\mathcal{AW}_r$. Once again, the reader may skip this extension and consider as he wishes that convergences in $\mathcal{AW}_{f\oplus\hat g}$ mean convergences in $\mathcal{AW}_r$.

2.2 Stability

The stability of the WOT and WMOT problems, or more generally of any optimal transport problem, with respect to their marginals, are of paramount importance. Indeed, those problems are often computationally solvable when the marginals are finitely supported. It is therefore natural to discretise the marginals and solve the discretised problem, but this approach works only if we know that the discretised cost converges to the original one, which is assured when the stability holds. Another issue is that the marginals often derive from noisy data. In that context, if the stability of the cost function with respect to the marginals does not hold, then solving it is meaningless.

The proof of the stability of the WOT problem relies on the following extension to the finer topology of the approximation of couplings on the line in the weak adapted topology proved in the companion paper [10, Proposition 2.3]. This extension is an easy consequence of the equivalence stated in Definition 2.2.

Proposition 2.3. *Let $f\colon X\to[1,+\infty)$ and $g\colon Y\to[1,+\infty)$ be continuous. Let $\mu^k\in\mathcal{P}_f(\mathbb{R})$, $\nu^k\in\mathcal{P}_g(\mathbb{R})$, $k\in\mathbb{N}$, respectively converge to μ and ν in $\mathcal{P}_f(\mathbb{R})$ and $\mathcal{P}_g(\mathbb{R})$ respectively. Then there is for any $\pi\in\Pi(\mu,\nu)$ a sequence of couplings $\pi^k\in\Pi(\mu^k,\nu^k)$, $k\in\mathbb{N}$ converging to π in $\mathcal{AW}_{f\oplus\hat g}$.*

In the martingale setting, we recall the main theorem of the companion paper, namely that any martingale couplings whose marginals are approximated by probability measures in the convex order can be approximated by martingale couplings with respect to the adapted Wasserstein distance. We state it in the setting of our extended topology, which is also a direct consequence of the equivalence stated in Definition 2.2. For $r \geq 1$, we denote by $\mathcal{F}^r(X)$ the set of continuous functions $f : X \to [1, +\infty)$ which dominate $x \mapsto 1 + d_X^r(x, x_0)$, that is

$$\forall x \in X, \quad f(x) \geq 1 + d_X^r(x, x_0). \tag{2.3}$$

Theorem 2.4. *Let $f \in \mathcal{F}^1(\mathbb{R})$ and $g \in \mathcal{F}^1(\mathbb{R})$. Let $\mu^k \in \mathcal{P}_f(\mathbb{R})$ and $\nu^k \in \mathcal{P}_g(\mathbb{R})$, $k \in \mathbb{N}$, be in convex order and respectively converge to μ and ν in $\mathcal{P}_f(\mathbb{R})$ and $\mathcal{P}_g(\mathbb{R})$. Let $\pi \in \Pi_M(\mu, \nu)$. Then there exists a sequence of martingale couplings $\pi^k \in \Pi_M(\mu^k, \nu^k)$, $k \in \mathbb{N}$ converging to π in $\mathcal{AW}_{f \oplus \hat{g}}$.*

We recall that a sequence $(\mu^k)_{k \in \mathbb{N}}$ of probability measures on X is said to converge strongly to some $\mu \in \mathcal{P}(X)$ iff for any measurable subset $A \subset X$, $\mu^k(A)$ converges to $\mu(A)$ as k goes to $+\infty$.

Theorem 2.5. *Let $f : X \to [1, +\infty)$ and $g : Y \to [1, +\infty)$ be continuous. Let X and Y be Polish spaces, $C : X \times \mathcal{P}_g(Y) \to \mathbb{R}$ be convex in the second argument, lower semicontinuous and such that there exists a constant $K > 0$ which satisfies for all $(x, p) \in X \times \mathcal{P}_g(Y)$*

$$|C(x, p)| \leq K \left(f(x) + \int_Y g(y)\, p(dy) \right). \tag{2.4}$$

For $k \in \mathbb{N}$, let $\mu^k \in \mathcal{P}_f(X)$ and $\nu^k \in \mathcal{P}_g(Y)$ converge in $\mathcal{P}_f(X)$ and $\mathcal{P}_g(Y)$ as $k \to +\infty$ to μ and ν respectively. Then

(a) Existence: there exists $\pi^* \in \Pi(\mu, \nu)$ which minimises $V_C(\mu, \nu)$.

(b) Stability of the cost function: suppose that one of the following holds true:

(A) C is continuous.

(B) C is continuous in the second argument and μ^k converges strongly to μ as $k \to +\infty$.

Then there holds

$$V_C(\mu^k, \nu^k) \underset{k \to +\infty}{\longrightarrow} V_C(\mu, \nu). \tag{2.5}$$

(c) Stability of the minimisers: suppose that (2.5) holds. For $k \in \mathbb{N}$ let $\pi^{k,*} \in \Pi(\mu^k, \nu^k)$ be a minimiser of $V_C(\mu^k, \nu^k)$. Then any accumulation point of $(\pi^{k,*})_{k \in \mathbb{N}}$ for the weak convergence topology is a minimiser of $V_C(\mu, \nu)$. If the latter has a unique minimiser π^*, then

$$\pi^{k,*} \underset{k \to +\infty}{\longrightarrow} \pi^* \quad \text{in } \mathcal{P}_{f \oplus g}(X \times Y). \tag{2.6}$$

(d) Uniqueness: if C is strictly convex in the second argument, then the minimisers are unique and the convergence (2.6) holds in $\mathcal{AW}_{f \oplus \hat{g}}$.

Note that the WMOT problem is in fact a particular case of the WOT problem. Indeed, the resolution of the WMOT problem between $\mu, \nu \in \mathcal{P}^1(\mathbb{R}^d)$ for a cost function $C : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}_+$ amounts to the resolution of the WOT problem between the same marginals and the cost function

$$\tilde{C} : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \ni (x, p) \mapsto \left\{ \begin{array}{cl} C(x, p) & \text{if } \int_Y y\, p(dy) = x \\ +\infty & \text{else} \end{array} \right. \in \mathbb{R}_+ \cup \{+\infty\}.$$

Since infinite-valued cost functions are not admissible in the setting of Theorem 2.5, the case of the stability of the WMOT problem requires its own statement.

Theorem 2.6. *Let $f \in \mathcal{F}^1(\mathbb{R}^d)$ and $g \in \mathcal{F}^1(\mathbb{R}^d)$. Let $C : \mathbb{R}^d \times \mathcal{P}_g(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ be convex in the second argument, lower semicontinuous and such that there exists a constant $K > 0$ which satisfies for all $(x, p) \in \mathbb{R}^d \times \mathcal{P}_g(\mathbb{R}^d)$*

$$|C(x,p)| \leq K \left(f(x) + \int_Y g(y)\, p(dy) \right). \tag{2.7}$$

For $k \in \mathbb{N}$, let $\mu^k \in \mathcal{P}_f(\mathbb{R}^d)$ and $\nu^k \in \mathcal{P}_g(\mathbb{R}^d)$ be such that $\mu^k \leq_c \nu^k$ and μ^k, resp. ν^k, converges to μ in $\mathcal{P}_f(\mathbb{R}^d)$, resp. ν in $\mathcal{P}_g(\mathbb{R}^d)$ as $k \to +\infty$. Then

(a') Existence: there exists $\pi^* \in \Pi_M(\mu, \nu)$ which minimises $V_C^M(\mu, \nu)$.

(b') Stability of the cost function in dimension 1: suppose that $d = 1$ and one of the following holds true:

(A) C is continuous.

(B) C is continuous in the second argument and μ^k converges strongly to μ as $k \to +\infty$.

Then there holds

$$V_C^M(\mu^k, \nu^k) \underset{k \to +\infty}{\longrightarrow} V_C^M(\mu, \nu). \tag{2.8}$$

(c') Stability of the minimisers: suppose that (2.8) holds. For $k \in \mathbb{N}$ let $\pi^{k,*} \in \Pi_M(\mu^k, \nu^k)$ be a minimiser of $V_C^M(\mu^k, \nu^k)$. Then any accumulation point of $(\pi^{k,*})_{k \in \mathbb{N}}$ for the weak convergence topology is a minimiser of $V_C^M(\mu, \nu)$. If the latter has a unique minimiser π^*, then

$$\pi^{k,*} \underset{k \to +\infty}{\longrightarrow} \pi^* \quad \text{in } \mathcal{P}_{f \oplus g}(\mathbb{R}^d \times \mathbb{R}^d). \tag{2.9}$$

(d') Uniqueness: if C is strictly convex in the second argument, then the minimisers are unique and the convergence (2.9) holds in $\mathcal{AW}_{f \oplus \hat{g}}$.

Remark 2.7. We actually show that (a'), (b'), (c') and (d') are still valid when there exist two Polish subspaces X and Y of $\mathbb{R}^d$ such that f is defined on X, g on Y, (2.7) holds for all $(x, p) \in X \times \mathcal{P}_g(Y)$ such that $\int_{\mathbb{R}^d} y\, p(dy) = x$, $\mu^k \in \mathcal{P}_f(X)$ converges in $\mathcal{P}_f(X)$ to $\mu \in \mathcal{P}_f(X)$ and $\nu^k \in \mathcal{P}_f(Y)$ converges in $\mathcal{P}_f(Y)$ to $\nu \in \mathcal{P}_f(Y)$ as $k \to +\infty$.

We can exhibit a strong connection betwen the (WMOT) problem and a Benamou-Brenier type formulation of the MOT problem suggested by [5]. In dimension one, this problem consists for two probability measures μ, ν in the convex order in maximising

$$MT(\mu, \nu) := \sup \mathbb{E}\left[\int_0^1 \sigma_t\, dt \right] \tag{MBB}$$

over all filtered probability spaces $(\Omega, (\mathcal{F}_t)_{t \in [0,1]}, \mathbb{P})$, real-valued $(\mathcal{F}_t)_{t \in [0,1]}$-progressive process $(\sigma_t)_{t \in [0,1]}$ and real-valued $(\mathcal{F}_t)_{t \in [0,1]}$-Brownian motions $(B_t)_{t \in [0,1]}$ such that the process

$$(M_t)_{t \in [0,1]} = \left(M_0 + \int_0^t \sigma_s\, dB_s \right)_{t \in [0,1]}$$

is a continuous martingale which satisfies $M_0 \sim \mu$ and $M_1 \sim \nu$. When the second moment of ν is finite, then (MBB) has a unique maximiser $(M_t^*)_{t \in [0,1]}$ [5, Theorem 1.5] called the stretched Brownian motion from μ to ν, since it is the martingale subject to the constraints $M_0^* \sim \mu$ and $M_1^* \sim \nu$ which correlates the most with the Brownian motion.

Let $C_2 : \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$ be defined for all $(x, p) \in \mathbb{R} \times \mathcal{P}_2(\mathbb{R})$ by $C_2(x, p) = \mathcal{W}_2^2(p, \mathcal{N}(0, 1))$, where $\mathcal{N}(0, 1)$ denotes the unidimensional standard normal distribution. Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$ be in the convex order and $V_{C_2}^M(\mu, \nu)$ be the value function given by (WMOT) for the cost function C_2. Let $\pi^* \in \Pi_M(\mu, \nu)$ be optimal for $V_{C_2}^M(\mu, \nu)$ and M^* be the stretched Brownian motion from μ to ν. Then Remark 2.1, Theorem 2.2 and Remark 2.3 from [5] imply that

(a) $MT(\mu,\nu) = \frac{1}{2}\left(1 + \int_{\mathbb{R}} |y|^2\,\nu(dy) - V_{C_2}^M(\mu,\nu)\right)$;

(b) π^* is the joint probability distribution of (M_0^*, M_1^*), and conversely

$$\forall t \in [0,1], \quad M_t^* = \mathbb{E}\left[F^{-1}_{\pi^*_X}(F_{\mathcal{N}(0,1)}(B_1))|X, (B_s)_{0\le s\le t}\right], \tag{2.10}$$

where $X \sim \mu$ is a random variable independent of the Brownian motion $(B_t)_{t\in[0,1]}$, and F_η, resp. F_η^{-1} denotes the cumulative distribution function, resp. the quantile function of a probability distribution $\eta \in \mathcal{P}(\mathbb{R})$.

As a consequence of Theorem 2.6, for $r \ge 2$, the stretched Brownian motion between converging marginals in $\mathcal{W}_r$ converges in $\mathcal{AW}_r$ to the stretched Brownian motion between the limits. Because of the martingale structure, we even have an estimate of the $\mathcal{AW}_r$-distance between the joint probability distributions of the initial position and the whole trajectory.

Corollary 2.8 (Stability of the unidimensional stretched Brownian motion)**.** *Let $r \ge 2$ and $\mu^k, \nu^k, \mu, \nu \in \mathcal{P}^r(\mathbb{R})$, $k \in \mathbb{N}$ be such that for all $k \in \mathbb{N}$, $\mu^k \le_c \nu^k$ and μ^k, resp. ν^k, converges to μ, resp. ν, in $\mathcal{W}_r$.*

For $k \in \mathbb{N}$, let M^k be the stretched Brownian motion from μ^k to ν^k and M^* be the stretched Brownian motion from μ to ν. Equipping $C([0,1])$ with the supremum distance and denoting by $\mathcal{L}(Z)$ the law of any random variable Z, we have the estimate

$$\mathcal{AW}_r^r\left(\mathcal{L}(M_0^k, (M_t^k)_{t\in[0,1]}), \mathcal{L}(M_0^*, (M_t^*)_{t\in[0,1]})\right) \le \left(\frac{r}{r-1}\right)^r \mathcal{AW}_r^r\left(\mathcal{L}(M_0^k, M_1^k), \mathcal{L}(M_0^*, M_1^*)\right),$$

and the right-hand side vanishes as k goes to $+\infty$.

2.3 Stability of the superreplication bound for VIX futures

The Volatility Index (VIX), often referred to as the Fear Index, is a popular measure to determine market sentiment. When investors expect the market to move vigorously, they typically tend to purchase more options, which has an impact on implied volatility levels. The VIX is by definition the implied volatility calculated on a 30 days horizon on the S&P 500. The more the VIX increases, the more demand is expressed for options, which become more expensive. In that case the market is described as volatile. Conversely, a decreasing VIX often means less demand and therefore decreasing option prices, hence the market is perceived as calm.

We consider a financial market composed of two financial assets: the risk-free asset and the S&P 500 $(S_t)_{t\in\{T_1,T_2\}}$, tradable at dates T_1 and $T_2 = T_1 + 30$ days. We suppose known the market price of call options for any strike $K \ge 0$, so that by the Breeden-Litzenberger formula [15] we get the respective probability distributions μ and ν of S_{T_1} and S_{T_2}. We allow trading at time 0 in vanilla options with maturities T_1 and T_2, and trading at time T_1 in the S&P 500 and the forward-starting log-contract, that is the option with payoff $\frac{-2}{T_2-T_1}\ln\frac{S_{T_2}}{S_{T_1}}$ at T_2. In this setting, Guyon, Menegaux and Nutz [22] derive the model-independent arbitrage-free upper bound for the VIX future expiring at T_1, given by the smallest superreplication price at time 0

$$P_{\text{super}}(\mu,\nu) = \inf\left(\int_{(0,+\infty)} u_1(x)\,\mu(dx) + \int_{(0,+\infty)} u_2(y)\,\nu(dy)\right), \tag{2.11}$$

where the infimum is taken over all $(u_1,u_2) \in L^1(\mu)\times L^1(\nu)$ and measurable maps Δ^S, Δ^L such that for all $(x,y,v) \in (0,+\infty)^2 \times [0,+\infty)$,

$$u_1(x) + u_2(y) + \Delta^S(x,v)(y-x) + \Delta^L(x,v)\left(-\frac{2}{T_2-T_1}\ln\frac{y}{x} - v^2\right) - v \tag{2.12}$$

is nonnegative. Similarly, the model-independent arbritrage-free lower bound for the VIX future expiring at T_1 is given by the largest subreplication price at time 0

$$P_{\text{sub}}(\mu,\nu) = \sup\left(\int_{(0,+\infty)} u_1(x)\,\mu(dx) + \int_{(0,+\infty)} u_2(y)\,\nu(dy)\right),$$

where the supremum is taken over all $(u_1,u_2) \in L^1(\mu)\times L^1(\nu)$ and measurable maps Δ^S,Δ^L such that for all $(x,y,v)\in(0,+\infty)^2\times[0,+\infty)$, (2.12) is nonpositive.

Note that the primal problem $P_{\text{super}}(\mu,\nu)$ involves in (2.12) three variables x,y,s, which stand respectively for the S&P 500 at time T_1, the S&P 500 at time T_2, and the VIX at time T_1. We would then naturally expect the dual formulation to involve three marginals. Strikingly, the dual side of the superreplication of the VIX takes the form of a WMOT problem with 2 marginals only thanks to concavity of the square root, see [22, Proposition 4.10].

Proposition 2.9 (Guyon, Menegaux, Nutz, 2017)**.** *Let $0<T_1<T_2$ and $f\colon[1,+\infty)\to\mathbb{R}_+$ be given by $f(x)=|\ln(x)|+|x|$. Let $\mu,\nu\in\mathcal{P}_f((0,+\infty))$ be in the convex order, then the dual problem D_{super} consists of*

$$D_{super}(\mu,\nu) = \sup_{\pi\in\Pi_M(\mu,\nu)} \int_{(0,+\infty)} C_{VIX}(x,\pi_x)\,\mu(dx), \tag{2.13}$$

when $C_{VIX}\colon(0,+\infty)\times\mathcal{P}_f((0,+\infty)),(x,p)\mapsto\sqrt{-\frac{2}{T_2-T_1}\int_{(0,+\infty)}\ln\left(\frac{y}{x}\right)\,p(dy)}$. The values of $P_{super}(\mu,\nu)$ and $D_{super}(\mu,\nu)$ coincide.

The fact that μ and ν are defined on $(0,+\infty)$ motivated Remark 2.7, that is the consideration of the stability of the WMOT problem in the setting of Polish subspaces of $\mathbb{R}^d$. Note that C_{VIX} is indeed an admissible weak transport cost: on $\{(x,p)\in(0,+\infty)\times\mathcal{P}_f((0,+\infty))\mid\int_{(0,+\infty)}y\,p(dy)=x\}$ it is well-defined and continuous, and the map $p\mapsto C_{VIX}(x,p)$ is concave on $\{p\in\mathcal{P}_f((0,+\infty))\mid\int_{\mathbb{R}}y\,p(dy)=x\}$ for fixed $x\in(0,+\infty)$. Hence, we can apply Theorem 2.6 and find that the robust superreplication bound for VIX futures depends continuously on the marginals:

Corollary 2.10. *In the setting of Proposition 2.9, let the pairs $\mu^k,\nu^k\in\mathcal{P}_f((0,+\infty))$, $k\in\mathbb{N}$, be in convex order and converge in $\mathcal{P}_f(\mathbb{R})$ to μ and ν respectively. Then there exist maximisers $\pi^{k,*}\in\Pi_M(\mu^k,\nu^k)$,*

$$\lim_{k\to+\infty} D_{super}(\mu^k,\nu^k) = D_{super}(\mu,\nu),$$

and any weak accumulation point of $(\pi^{k,*})_{k\in\mathbb{N}}$ maximises $D_{super}(\mu,\nu)$.

2.4 Martingale monotonicity

A remarkable tool in the theory of optimal transport is cyclical monotonicity. It allows to determine optimality of a coupling only by knowing its support. In its spirit the notion of finite optimality was developed in context of martingale optimal transport in [11] and [21].

Definition 2.11 (Competitor)**.** Let $\alpha=\mu\times\alpha_x\in\mathcal{P}^1(\mathbb{R}\times\mathbb{R})$. We call $\alpha'=\mu'\times\alpha'_x\in\mathcal{P}^1(\mathbb{R}\times\mathbb{R})$ a competitor of α, if

$$\mu=\mu' \quad\text{and}\quad \int_{\mathbb{R}} y\,\alpha'_x(dy)=\int_{\mathbb{R}} y\,\alpha_x(dy),\quad \mu(dx)\text{-almost everywhere.}$$

Definition 2.12 (Finite optimality)**.** Let $c\colon\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ be a cost function. We say that a Borel set $\Gamma\subset\mathbb{R}\times\mathbb{R}$ is *finitely optimal* for c if for every probability measure $\alpha\in\mathcal{P}(\mathbb{R}\times\mathbb{R})$ finitely supported on Γ, we have

$$\int_{\mathbb{R}\times\mathbb{R}} c(x,y)\,\alpha(dx,dy) \le \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\,\alpha'(dx,dy),$$

for every competitor α' of α.

Under the assumption that c is continuous and sufficiently integrable, there was shown in [11, Lemma A.2] and [21, Theorem 1.3] that a martingale coupling is optimal if it is concentrated on a finitely optimal set.

Recently the notion of martingale C-monotonicity, c.f. [7], was introduced for martingale optimal weak transport (WMOT), which was therein used to show stability of the martingale optimal transport problem.

Definition 2.13 (Martingale C-monotonicity)**.** We say that a Borel set $\Gamma \subset \mathbb{R} \times \mathcal{P}^1(\mathbb{R})$ is *martingale C-monotone* iff for any $N \in \mathbb{N}$, any collection $(x_1, p_1), \dots, (x_N, p_N) \in \Gamma$ and $q_1, \dots, q_N \in \mathcal{P}^1(\mathbb{R})$ such that $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i$ and $\int_{\mathbb{R}} y\, p_i(dy) = \int_{\mathbb{R}} y\, q_i(dy)$, we have

$$\sum_{i=1}^N C(x_i, p_i) \le \sum_{i=1}^N C(x_i, q_i).$$

So far, it was known that martingale C-monotonicity is a necessary optimality criterion in the following sense, c.f. [7, Theorem 3.4]: let $\pi^* \in \Pi_M(\mu, \nu)$ be a martingale coupling which minimises (WMOT), then $J(\pi^*)$ is concentrated on a martingale C-monotone set. This means explicitly that there is a martingale C-monotone set Γ with

$$(x, \pi_x) \in \Gamma \quad \text{for } \mu(dx)\text{-almost every } x. \tag{2.14}$$

Remark 2.14. Conversely, if $\pi \in \Pi_M(\mu, \nu)$ is a finitely supported coupling of the form $\frac{1}{N}\sum_{i=1}^N \delta_{x_i}(dx)\, p_i(dy)$ for $x_1 < \cdots < x_n \in \mathbb{R}$ and $p_1, \cdots, p_N \in \mathcal{P}^1(\mathbb{R})$ and satisfies (2.14), then it is optimal. Indeed, in that case $(x_1, p_1), \cdots, (x_N, p_N) \in \Gamma$ and any martingale coupling $\pi' \in \Pi_M(\mu, \nu)$ is of the form $\pi'(dx, dy) = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}(dx)\, q_i(dy)$, where $q_1, \cdots, q_N \in \mathcal{P}^1(\mathbb{R})$ are such that $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i$ and for all $i \in \{1, \cdots, N\}$, $\int_{\mathbb{R}} y\, p_i(dy) = x_i = \int_{\mathbb{R}} y\, q_i(dy)$. By definition of martingale C-monotonicity, we get

$$\int_{\mathbb{R}\times\mathbb{R}} C(x, \pi_x)\, \mu(dx) = \frac{1}{N}\sum_{i=1}^N C(x_i, p_i) \le \frac{1}{N}\sum_{i=1}^N C(x_i, q_i) = \int_{\mathbb{R}^2} C(x, \pi'_x)\, \mu(dx),$$

hence π is optimal.

However, the question remained open if any martingale coupling satisfying (2.14) is optimal. The stability result, Theorem 2.6, allows us to confirm that this is indeed the case.

Theorem 2.15 (Sufficiency)**.** *Let $f\colon \mathbb{R} \to [1, +\infty)$ and $g\colon \mathbb{R} \to [1, +\infty)$ be continuous. Let $\mu \in \mathcal{P}_f(\mathbb{R})$ and $\nu \in \mathcal{P}_g(\mathbb{R})$ be in convex order, and $C\colon \mathbb{R} \times \mathcal{P}_g(\mathbb{R}) \to \mathbb{R}$ be a measurable cost function, continuous in the second argument and such that there exists a constant $K > 0$ which satisfies*

$$\forall (x, p) \in \mathbb{R} \times \mathcal{P}_g(\mathbb{R}), \quad C(x, p) \le K\left(f(x) + \int_{\mathbb{R}} g(y)\, p(dy)\right),$$

Let Γ be martingale C-monotone and $\pi \in \Pi_M(\mu, \nu)$ be such that we have (2.14)*. Then π is optimal for* (WMOT)*.*

In turn Theorem 2.15 allows us to strengthen [11, Lemma A.2] and [21, Theorem 1.3] by assuming less continuity of the cost function.

Corollary 2.16. *Let $f\colon \mathbb{R} \to [1, +\infty)$ and $g\colon \mathbb{R} \to [1, +\infty)$ be continuous. Let $\mu \in \mathcal{P}_f(\mathbb{R})$ and $\nu \in \mathcal{P}_g(\mathbb{R})$ be in convex order, $c\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be measurable and $y \mapsto c(x, y)$ be continuous for all $x \in \mathbb{R}$. Furthermore, let $K > 0$ be a constant such that*

$$c(x, y) \le K(f(x) + g(y)), \quad \forall (x, y) \in \mathbb{R} \times \mathbb{R}.$$

Then $\pi \in \Pi_M(\mu, \nu)$ is finitely optimal if and only if π is optimal for the MOT problem.

3 Stability

This section is devoted to the proof of Theorem 2.5 and Theorem 2.6 about the stability of (WOT) and (WMOT), and the corollary on the stability of the stretched Brownian motion in dimension one.

Proof of Theorem 2.5 and Theorem 2.6. First, we prove (a) and (a'). Let $\hat{\Pi}(\mu,\nu) = \Pi(\mu,\nu)$ and $\hat{V}_C(\mu,\nu) = V_C(\mu,\nu)$ in the setting of Theorem 2.5, and $\hat{\Pi}(\mu,\nu) = \Pi_M(\mu,\nu)$ and $\hat{V}_C(\mu,\nu) = V_C^M(\mu,\nu)$ in the setting of Theorem 2.6.

Let $(\pi^n)_{n\in\mathbb{N}} \in \hat{\Pi}(\mu,\nu)^{\mathbb{N}}$ be such that $\int_X C(x,\pi^n_x)\,\mu(dx)$ converges to $\hat{V}_C(\mu,\nu)$ as $n \to +\infty$. By tightness of μ and ν we deduce the existence of a subsequence $(\pi^{n_l})_{l\in\mathbb{N}}$ of $(\pi^n)_{n\in\mathbb{N}}$ which converges to some $\pi^* \in \hat{\Pi}(\mu,\nu)$ with respect to the weak convergence topology and therefore the topology of $\mathcal{P}_{f\oplus g}(X\times Y)$ since $\pi^{n_l}(f\oplus g) = \mu(f) + \nu(f) = \pi^*(f\oplus g)$ for all $l \in \mathbb{N}$. By Proposition 5.8 (b) below we then have

$$\hat{V}_C(\mu,\nu) \le \int_X C(x,\pi^*_x)\,\mu(dx) \le \liminf_{l\to+\infty} \int_X C(x,\pi^{n_l}_x)\,\mu(dx) = \hat{V}_C(\mu,\nu),$$

which shows that π^* is a minimiser for $\hat{V}_C(\mu,\nu)$ and proves (a) and (a').

We now show that the convergence

$$\hat{V}_C(\mu^k,\nu^k) \underset{k\to+\infty}{\longrightarrow} \hat{V}_C(\mu,\nu) \tag{3.1}$$

holds under either Assumption (A) or Assumption (B) in the setting of Theorem 2.5, and in the setting of Theorem 2.6 as soon as $d = 1$, which will prove (b) and (b'). Let π^* be a minimiser of $\hat{V}_C(\mu,\nu)$. By Proposition 2.3 in the setting of Theorem 2.5 and Theorem 2.4 in the setting of Theorem 2.6 if $d = 1$, there exists a sequence $\pi^k \in \hat{\Pi}(\mu^k,\nu^k)$, $k \in \mathbb{N}$, which converges to π^* in $\mathcal{AW}_{f\oplus\hat{g}}$, which is equivalent to $J(\pi^k)$ converging to $J(\pi^*)$ in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$.

Under Assumption (A), we then have by Lemma 5.12 (b) that

$$\int_{X\times\mathcal{P}_g(Y)} C(x,p)\,J(\pi^k)(dx,dp) \underset{k\to+\infty}{\longrightarrow} \int_{X\times\mathcal{P}_g(Y)} C(x,p)\,J(\pi^*)(dx,dp). \tag{3.2}$$

Under Assumption (B), the strong convergence of $(\mu^k)_{k\in\mathbb{N}}$ to μ and the weak convergence of $(J(\pi^k))_{k\in\mathbb{N}}$ to $J(\pi^*)$ imply by Lemma 5.11 (b) that $(J(\pi^k))_{k\in\mathbb{N}}$ converges stably to $J(\pi^*)$, hence (3.2) still holds by Lemma 5.12 (d).

Using (3.2) for the second equality, we then have

$$\begin{aligned}
\limsup_{k\to+\infty} \hat{V}_C(\mu^k,\nu^k) &\le \limsup_{k\to+\infty} \int_X C(x,\pi^k_x)\,\mu^k(dx) \\
&= \limsup_{k\to+\infty} \int_{X\times\mathcal{P}_g(Y)} C(x,p)\,J(\pi^k)(dx,dp) \\
&= \int_{X\times\mathcal{P}_g(Y)} C(x,p)\,J(\pi^*)(dx,dp) \\
&= \hat{V}_C(\mu,\nu).
\end{aligned} \tag{3.3}$$

Let $(\hat{V}_C(\mu^{k_l},\nu^{k_l}))_{l\in\mathbb{N}}$ be a subsequence of $(\hat{V}_C(\mu^k,\nu^k))_{k\in\mathbb{N}}$ converging to $\liminf_{k\to+\infty}\hat{V}_C(\mu^k,\nu^k)$. Let $\tilde{\pi} \in \hat{\Pi}(\mu,\nu)$ be an accumulation point of $(\pi^{k_l,*})_{l\in\mathbb{N}}$ with respect to the weak convergence topology, which exists by tightness of the marginals. Note that in the martingale case, the fact that $\tilde{\pi}$ is a martingale coupling is guaranteed by the $\mathcal{W}_1$-convergence of the marginals. Then by Proposition 5.8 (b) below, we find that

$$\liminf_{k\to+\infty} \hat{V}_C(\mu^k,\nu^k) = \lim_{l\to+\infty} \int_X C(x,\pi^{k_l,*}_x)\,\mu^{k_l}(dx) \ge \int_X C(x,\tilde{\pi}_x)\,\mu(dx) \ge \hat{V}_C(\mu,\nu).$$

With (3.3), we conclude that $\lim_{k\to+\infty}\hat{V}_C(\mu^k,\nu^k)=\hat{V}_C(\mu,\nu)$, which proves (3.1).

Let us now prove (c) and (c'), assuming that (3.1) holds. For $k\in\mathbb{N}$, let $\pi^{k,*}\in\hat{\Pi}(\mu^k,\nu^k)$ be a minimiser of $\hat{V}_C(\mu^k,\nu^k)$. For any subsequence $(\pi^{k_j,*})_{j\in\mathbb{N}}$ of $(\pi^{k,*})_{k\in\mathbb{N}}$ converging weakly to some $\tilde{\pi}$, Proposition 5.8 (b) below ensures that

$$\hat{V}_C(\mu,\nu)=\lim_{j\to+\infty}\hat{V}_C(\mu^{k_j},\nu^{k_j})=\lim_{j\to+\infty}\int_X C(x,\pi_x^{k_j,*})\,\mu(dx)\geq\int_X C(x,\tilde{\pi}_x)\,\mu(dx)\geq\hat{V}_C(\mu,\nu),$$

so $\tilde{\pi}$ is a minimiser of $\hat{V}_C(\mu,\nu)$. In particular if $\hat{V}_C(\mu,\nu)$ has a unique minimiser π^*, it is the unique accumulation point with respect to the weak convergence topology of the tight sequence $(\pi^{k,*})_{k\in\mathbb{N}}$, which therefore converges to π^* weakly and even in $\mathcal{P}_{f\oplus g}(X\times Y)$ since its marginals converge in $\mathcal{P}_f(X)$ and $\mathcal{P}_g(Y)$ respectively. Hence (c) and (c') are proved.

Finally, let us show (d) and (d'). The strict convexity of $C(x,\cdot)$ for all $x\in X$ yields uniqueness of the minimisers. Indeed when $\pi,\tilde{\pi}\in\hat{\Pi}(\mu,\nu)$ then $\frac{1}{2}(\pi+\tilde{\pi})\in\hat{\Pi}(\mu,\nu)$. When, moreover, $\pi\neq\tilde{\pi}$, then $\mu(\{x\in X\mid\pi_x\neq\tilde{\pi}_x\})>0$ and since $C(x,\frac{1}{2}(\pi_x+\tilde{\pi}_x))\leq\frac{1}{2}(C(x,\pi_x)+C(x,\tilde{\pi}_x))$ with strict inequality when $\pi_x\neq\tilde{\pi}_x$,

$$\int_X C\left(x,\frac{\pi_x+\tilde{\pi}_x}{2}\right)\mu(dx)<\frac{1}{2}\left(\int_X C(x,\pi_x)\,\mu(dx)+\int_X C(x,\tilde{\pi}_x)\,\mu(dx)\right).\tag{3.4}$$

Let then π^* be the only minimiser of $\hat{V}_C(\mu,\nu)$. To conclude the proof, it is enough to show that $J(\pi^{k,*})$ converges to $J(\pi^*)$ in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$ as k goes to $+\infty$. Let $P^*\in\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$ be an accumulation point of $(J(\pi^{k,*}))_{k\in\mathbb{N}}$, which exists by Lemma 5.7.

To conclude, it suffices to show that $P^*=J(\pi^*)$, which is achieved in three steps. Let $\hat{\Lambda}(\mu,\nu)=\Lambda(\mu,\nu)$ (see the definition (5.3) below) in the setting of Theorem 2.5 and $\hat{\Lambda}(\mu,\nu)=\Lambda_M(\mu,\nu)$ (see the definition (5.5) below) in the setting of Theorem 2.6. First we show that

$$P^*\in\hat{\Lambda}(\mu,\nu).\tag{3.5}$$

Next, we show that $J(\pi^*)$ and P^* both minimise

$$\tilde{V}_C(\mu,\nu):=\inf_{P\in\hat{\Lambda}(\mu,\nu)}\int_{X\times\mathcal{P}_g(Y)}C(x,p)\,P(dx,dp).$$

Finally, we show the uniqueness of minimisers of $\tilde{V}_C(\mu,\nu)$.

Let $(J(\pi^{k_l,*}))_{l\in\mathbb{N}}$ be a subsequence converging to P^* in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$. By Lemma 5.9 below we have

$$\int_{(x,p)\in X\times\mathcal{P}_g(Y)}p(dy)\,J(\pi^{k_l,*})(dx,dp)\underset{l\to+\infty}{\longrightarrow}\int_{(x,p)\in X\times\mathcal{P}_g(Y)}p(dy)\,P^*(dx,dp),$$

where the convergence holds in $\mathcal{P}_{f\oplus g}(X\times Y)$ as l goes to $+\infty$. Since the left-hand side is ν^{k_l}, which converges to ν in $\mathcal{W}_g$ and therefore in the weak topology, we deduce by uniqueness of the limit that the right-hand side is ν, hence $P^*\in\Lambda(\mu,\nu)$. In the setting of Theorem 2.6, since, as $f,g\in\mathcal{F}^1(\mathbb{R}^d)$, $X\times\mathcal{P}_g(Y)\ni(x,p)\mapsto\left|x-\int_Y y\,p(dy)\right|\in\Phi_{f\oplus\hat{g}}(X\times\mathcal{P}_g(Y))$, we have that

$$0=\int_{X\times\mathcal{P}_g(Y)}\left|x-\int_Y y\,p(dy)\right|\,J(\pi^{k_l,*})(dx,dp)\underset{l\to+\infty}{\longrightarrow}\int_{X\times\mathcal{P}_g(Y)}\left|x-\int_Y y\,p(dy)\right|\,P^*(dx,dp),$$

hence $P^*\in\Lambda_M(\mu,\nu)$.

Let us show that $J(\pi^*)$ and P^* both minimise $\tilde{V}_C(\mu,\nu)$. Note that since $P^*\in\hat{\Lambda}(\mu,\nu)$, we have $P^*(X\times\mathcal{P}_g(Y))=1$. Since $(J(\pi^{k_l,*}))_{l\in\mathbb{N}}$ converges to P^* in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$, we find with Lemma 5.12 like in the derivation of (3.2) that

$$\int_{X\times\mathcal{P}_g(Y)}C(x,p)\,J(\pi^{k_l,*})(dx,dp)\underset{l\to+\infty}{\longrightarrow}\int_{X\times\mathcal{P}_g(Y)}C(x,p)\,P^*(dx,dp).\tag{3.6}$$

Then (3.6), the definition of $\pi^{k_l,*}$ and last (a), resp. (a'), yield

$$\begin{aligned}\int_{X\times\mathcal{P}_g(Y)} C(x,p)\,P^*(dx,dp) &= \lim_{l\to+\infty}\int_{X\times\mathcal{P}_g(Y)} C(x,p)\,J(\pi^{k_l,*})(dx,dp)\\ &= \lim_{l\to+\infty}\hat{V}_C(\mu^{k_l},\nu^{k_l})\\ &= \hat{V}_C(\mu,\nu)\\ &= \int_{X\times\mathcal{P}_g(Y)} C(x,p)\,J(\pi^*)(dx,dp).\end{aligned}$$

Let $P(dx,dp)=\mu(dx)\,P_x(dp)\in\hat{\Lambda}(\mu,\nu)$. Then $\mu(dx)\int_{p\in\mathcal{P}_f(Y)}p(dy)\,P_x(dp)\in\hat{\Pi}(\mu,\nu)$, so by Proposition 5.10 below for the last inequality,

$$\begin{aligned}\int_{X\times\mathcal{P}_f(Y)} C(x,p)\,J(\pi^*)(dx,dp) &= \int_X C(x,\pi_x^*)\,\mu(dx)\\ &= \hat{V}_C(\mu,\nu)\\ &\leq \int_X C\left(x,\int_{p\in\mathcal{P}_f(Y)}p(dy)\,P_x(dp)\right)\mu(dx)\\ &\leq \int_X\int_{\mathcal{P}_f(Y)} C(x,p)\,P_x(dp)\,\mu(dx),\end{aligned}\tag{3.7}$$

which proves that $J(\pi^*)$ minimises $\tilde{V}_C(\mu,\nu)$, and so does P^*.

We now prove that $J(\pi^*)$ is the only minimiser of $\tilde{V}_C(\mu,\nu)$. To do so, we first prove that any minimiser of $\tilde{V}_C(\mu,\nu)$ is in the image of J. Let then $\tilde{P}$ be such a minimiser. For $x\in X$, let $\tilde{\pi}_x(dy)=\int_{p\in\mathcal{P}_f(Y)}p(dy)\,\tilde{P}_x(dp)$ and $\tilde{\pi}(dx,dy)=\mu(dx)\,\tilde{\pi}_x(dy)$. Then $J(\tilde{\pi})\in\hat{\Lambda}(\mu,\nu)$ and Proposition 5.10 below yields

$$\int_{X\times\mathcal{P}_f(Y)} C(x,p)\,J(\tilde{\pi})(dx,dp)=\int_X C(x,\tilde{\pi}_x)\,\mu(dx)\leq\int_X\int_{\mathcal{P}_f(Y)} C(x,p)\,\tilde{P}_x(dp)\,\mu(dx).$$

By optimality of $\tilde{P}$, this inequality is an equality, hence for $\mu(dx)$-almost every $x\in X$ we have

$$C(x,\tilde{\pi}_x)=\int_{\mathcal{P}_f(Y)} C(x,p)\,\tilde{P}_x(dp),$$

and therefore $\tilde{P}_x=\delta_{\tilde{\pi}_x}$ by the equality case of Proposition 5.10 below, or equivalently $\tilde{P}=J(\tilde{\pi})$. Therefore any minimiser of $\tilde{V}_C(\mu,\nu)$ is contained in $J(\hat{\Pi}(\mu,\nu))$.

Recall that for all $\pi\in\hat{\Pi}(\mu,\nu)$ we have

$$\int_X C(x,\pi_x)\,\mu(dx)=\int_{X\times\mathcal{P}_g(Y)} C(x,p)\,J(\pi)(dx,dp).$$

With (3.7), we deduce that $P\in\hat{\Lambda}(\mu,\nu)$ is a minimiser of $\tilde{V}_C(\mu,\nu)$ iff P is the image of a minimiser of $\hat{V}_C(\mu,\nu)$ by J. By (3.4) the minimiser of $\hat{V}_C(\mu,\nu)$ is unique. This shows the uniqueness of minimisers of $\tilde{V}_C(\mu,\nu)$, and therefore the uniqueness of accumulation points of $(J(\pi^{k,*}))_{k\in\mathbb{N}}$, which is conclusive. □

The proof of Corollary 2.8 relies on the following Lemma.

Lemma 3.1. *Let $\rho>1$, and $C_\rho:\mathbb{R}\times\mathcal{P}^\rho(\mathbb{R})\to\mathbb{R}$ be defined for all $(x,p)\in\mathbb{R}\times\mathcal{P}^\rho(\mathbb{R})$ by $C_\rho(x,p)=\mathcal{W}_\rho^\rho(p,\gamma)$, where $\gamma\in\mathcal{P}^\rho(\mathbb{R})$ does not weight points. Let $V^M_{C_\rho}$ be the value function given by* (WMOT) *for the cost function C_ρ.*

Let $r\geq\rho$ and $\mu^k,\nu^k\in\mathcal{P}^r(\mathbb{R})$, $k\in\mathbb{N}$ be in convex order and converge respectively to μ and ν in $\mathcal{W}_r$. Then $\lim_{k\to+\infty}V^M_{C^\rho}(\mu^k,\nu^k)=V^M_{C^\rho}(\mu,\nu)$ and the optimisers are converging in $\mathcal{AW}_r$.

Proof. By Theorem 2.6 it is sufficient to show that $p \mapsto \mathcal{W}_\rho^\rho(\gamma, p)$ is strictly convex. Since γ does not weight points, the unique $\mathcal{W}_\rho$-optimal coupling between γ and $p \in \mathcal{P}_\rho(\mathbb{R})$ is the comonotonous coupling χ^p given by the map $x \mapsto F_p^{-1}(F_\gamma(x))$ i.e. the image of γ by $x \mapsto (x, F_p^{-1}(F_\gamma(x)))$. For $q \in \mathcal{P}_\rho(\mathbb{R})$ and $\lambda \in (0,1)$ the coupling $\chi = (1-\lambda)\chi^p + \lambda \chi^q$ between γ and $(1-\lambda)p + \lambda q$ is not given by a map unless $F_q^{-1}(u) = F_p^{-1}(u)$ for all $u \in (0,1)$ i.e. $p = q$. Therefore, when $p \neq q$,

$$(1-\lambda)\mathcal{W}_\rho^\rho(\gamma, p) + \lambda \mathcal{W}_\rho^\rho(\gamma, q) = \int |x-y|^\rho \, \chi(dx, dy) > \mathcal{W}_\rho^\rho(\gamma, (1-\lambda)p + \lambda q).$$

□

We can now prove the stability of the unidimensional stretched Brownian motion.

Proof of Corollary 2.8. Let $\gamma = \mathcal{N}(0,1)$ be the unidimensional standard normal distribution and $C_2 : \mathbb{R} \times \mathcal{P}^2(\mathbb{R}) \to \mathbb{R}$ be defined for all $(x,p) \in \mathbb{R} \times \mathcal{P}^2(\mathbb{R})$ by $C_2(x,p) = \mathcal{W}_2^2(p, \gamma)$. Let $V_{C_2}^M$ be the value function given by (WMOT) for the cost function C_2.

In the setting of Corollary 2.8, let $\pi^* \in \Pi_M(\mu, \nu)$, resp. $\pi^k \in \Pi_M(\mu^k, \nu^k)$ be optimal for $V_{C_2}^M(\mu, \nu)$, resp. $V_{C_2}^M(\mu^k, \nu^k)$. For $(x, b) \in \mathbb{R} \times \mathbb{R}^{[0,1]}$, let $B = (B_t)_{t \in [0,1]}$ be a Brownian motion and

$$G^k(x,b) = \left(\mathbb{E}\left[F_{\pi_x^k}^{-1}(F_\gamma(B_1 - B_t + b_t))\right]\right)_{t\in[0,1]} \text{ and } G^*(x,b) = \left(\mathbb{E}\left[F_{\pi_x^*}^{-1}(F_\gamma(B_1 - B_t + b_t))\right]\right)_{t\in[0,1]}.$$

According to (2.10), $(M_0^k, (M_t^k)_{t\in[0,1]})$ and $(M_0^*, (M_t^*)_{t\in[0,1]})$ are respectively distributed according to

$$\eta^k(dx, df) := \mu^k(dx)\,(G^k(x,\cdot)_* W)(df) \quad \text{and} \quad \eta^*(dx, df) := \mu(dx)\,(G^*(x,\cdot)_* W)(df),$$

where W denotes the Wiener measure on $C([0,1])$. Let $\chi^k \in \Pi(\mu^k, \mu)$ be optimal for $\mathcal{AW}_r(\pi^k, \pi)$. Then

$$\mathcal{AW}_r^r(\eta^k, \eta^*) \le \int_{\mathbb{R}\times\mathbb{R}} \left(|x-x'|^r + \mathcal{W}_r^r(G^k(x,\cdot)_* W, G(x',\cdot)_* W)\right) \chi^k(dx, dx').$$

According to (2.10), for $\mu^k(dx)$-almost every $x \in \mathbb{R}$, $G^k(x, B)$ is the stretched Brownian motion from δ_x to π_x^k, hence it is a continuous $(\mathcal{F}_t)_{t\in[0,1]}$-martingale, where $(\mathcal{F}_t)_{t\in[0,1]}$ is the natural filtration associated to B. Similarly, for $\mu(dx')$-almost every $x \in \mathbb{R}$, $G^*(x', B)$ is a continuous $(\mathcal{F}_t)_{t\in[0,1]}$-martingale. Therefore, for $\chi^k(dx, dx')$-almost every $(x, x') \in \mathbb{R}\times\mathbb{R}$, $G^k(x,B) - G^*(x',B)$ is a continuous $(\mathcal{F}_t)_{t\in[0,1]}$-martingale. Using Doob's martingale inequality for the second inequality, the fact that $F_\gamma(B_1)$ is uniformly distributed on $(0,1)$ for the first equality and the fact that the comonotonous coupling between π_x^k and $\pi_{x'}^*$ is optimal for $\mathcal{W}_r(\pi_x^k, \pi_{x'}^*)$ for the second equality, we get for $\chi^k(dx, dx')$-almost every $(x, x') \in \mathbb{R}\times\mathbb{R}$

$$\begin{aligned}
\mathcal{W}_r^r(G^k(x,\cdot)_* W, G(x',\cdot)_* W) &\le \mathbb{E}\left[\sup_{t\in[0,1]} \left|G^k(x,B)_t - G^*(x',B)_t\right|^r\right] \\
&\le \left(\frac{r}{r-1}\right)^r \mathbb{E}[|G^k(x,B)_1 - G^*(x',B)_1|^r] \\
&= \left(\frac{r}{r-1}\right)^r \mathbb{E}[|F_{\pi_x^k}^{-1}(F_\gamma(B_1)) - F_{\pi_{x'}^*}^{-1}(F_\gamma(B_1)|^r] \\
&= \left(\frac{r}{r-1}\right)^r \mathcal{W}_r^r(\pi_x^k, \pi_{x'}^*).
\end{aligned}$$

We deduce that

$$\mathcal{AW}_r^r(\eta^k, \eta^*) \le \left(\frac{r}{r-1}\right)^r \int_{\mathbb{R}\times\mathbb{R}} \left(|x-x'|^r + \mathcal{W}_r^r(\pi_x^k, \pi_{x'}^*)\right) \chi^k(dx, dx') = \left(\frac{r}{r-1}\right)^r \mathcal{AW}_r^r(\pi^k, \pi^*),$$

where the right-hand side vanishes as k goes to $+\infty$ in virtue of Lemma 3.1. □

4 Martingale monotonicity

In this section we prove the claim that martingale C-monotonicity is sufficient for optimality for (WMOT).

For $g : Y \to [1, +\infty)$ continuous, we denote

$$\mathcal{F}_g(Y) := \{f : Y \to [1, +\infty) \text{ continuous } \mid \forall y \in Y,\ f(y) \geq g(y)\}. \tag{4.1}$$

When $Y = \mathbb{R}^d$ for some $d \in \mathbb{N}^*$, we denote

$$\mathcal{F}_g^+(\mathbb{R}^d) := \left\{ f \in \mathcal{F}_g(\mathbb{R}^d) \mid \exists h : \mathbb{R}_+ \to [1, +\infty),\ \frac{h(t)}{t} \underset{t \to +\infty}{\longrightarrow} +\infty \text{ and } f = h \circ g \right\}. \tag{4.2}$$

Proof of Theorem 2.15. Let $h \in \mathcal{F}_g(\mathbb{R})$ be such that $\nu(h) < +\infty$, whose purpose will be revealed later in the proof. To demonstrate the main idea without further technical details, we assume for now that μ is concentrated on a Polish subset $\tilde{K} \subset \mathbb{R}$ and the restriction $C|_{\tilde{K} \times \mathcal{P}_h(\mathbb{R})}$ is continuous. Let $X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$ be independent random variables identically distributed according to μ and $\mathcal{G} \subset \Phi_{f \oplus \hat{g}}(\mathbb{R} \times \mathcal{P}(\mathbb{R}))$ be a countable family which determines the convergence in $\mathcal{P}_{f \oplus \hat{g}}(\mathbb{R} \times \mathcal{P}(\mathbb{R}))$ (see [16, Theorem 4.5.(b)]). By the law of large numbers, almost surely, for all $\psi \in \mathcal{G}$,

$$\begin{aligned}
\frac{1}{n} \sum_{k=1}^n \int_{\mathbb{R} \times \mathcal{P}(\mathbb{R})} \psi(x, p)\, \delta_{(X_k, \pi_{X_k})}(dx, dp) &= \frac{1}{n} \sum_{k=1}^n \psi(X_k, \pi_{X_k}) \\
&\underset{n \to +\infty}{\longrightarrow} \mathbb{E}[\psi(X_1, \pi_{X_1})] \\
&= \int_{\mathbb{R}} \psi(x, \pi_x)\, \mu(dx) \\
&= \int_{\mathbb{R} \times \mathcal{P}(\mathbb{R})} \psi(x, p)\, J(\pi)(dx, dp),
\end{aligned} \tag{4.3}$$

Moreover, almost surely, for all $n \in \mathbb{N}$,

$$(X_n, \pi_{X_n}) \in \Gamma \cap (\tilde{K} \times \mathcal{P}_g(\mathbb{R})), \tag{4.4}$$

and by the law of large numbers again, we have almost surely

$$\frac{1}{n} \sum_{k=1}^n \pi_{X_k}(h) \underset{n \to +\infty}{\longrightarrow} \mathbb{E}[\pi_{X_1}(h)] = \int_{\mathbb{R}} \pi_x(h)\, \mu(dx) = \nu(h). \tag{4.5}$$

Let then $\omega \in \Omega$ be such that (4.3), (4.4) and (4.5) hold when evaluated at ω and set $x_n = X_n(\omega)$ and $\pi^n(dx, dy) = \frac{1}{n} \sum_{k=1}^n \delta_{x_k}(dx)\, \pi_{x_k}(dy)$ for $n \in \mathbb{N}$. Then π^n has first marginal $\mu^n = \frac{1}{n} \sum_{k=1}^n \delta_{x_k}$ and second marginal $\nu^n = \int_{x \in \mathbb{R}} \pi_x(dy)\, \mu^n(dx)$. We deduce that π^n is a martingale C-monotone coupling between μ^n and ν^n which satisfies

$$J(\pi^n) = \frac{1}{n} \sum_{k=1}^n \delta_{(x_k, \pi_{x_k})} \underset{n \to +\infty}{\longrightarrow} J(\pi) \quad \text{in } \mathcal{P}_{f \oplus \hat{g}}(\tilde{K} \times \mathcal{P}(\mathbb{R})),$$

Thus π^n converges to π in $\mathcal{AW}_{f \oplus \hat{g}}$ as n goes to $+\infty$. In particular we have convergence of the marginals in $\mathcal{P}_f(\mathbb{R})$ and $\mathcal{P}_g(\mathbb{R})$ respectively, and even convergence in $\mathcal{P}_h(\mathbb{R})$ of the second marginals since $\nu^n(h)$ converges to $\nu(h)$ as $n \to +\infty$, consequence of (4.5) evaluated at ω. Note that due to martingale C-monotonicity of π^n, we have according to Remark 2.14 that

$$V_C^M(\mu^n, \nu^n) = \int_{\mathbb{R}} C(x, \pi_x^n)\, \mu^n(dx),$$

where we recall that the value function V_C^M is defined in (WMOT). Since μ^n, resp. ν^n converges to μ, resp. ν in $\mathcal{P}_f(\mathbb{R})$, resp. $\mathcal{P}_h(\mathbb{R})$, by Theorem 2.6 we have convergence of the optimal values $V_C^M(\mu^n,\nu^n) \to V_C^M(\mu,\nu)$ as n goes to $+\infty$. By convergence of $(\nu^n(h))_{n\in\mathbb{N}}$ to $\nu(h)$, the convergence $J(\pi^n) \to J(\pi)$ is not only in $\mathcal{P}_{f\oplus\hat{g}}(\tilde{K}\times\mathcal{P}(\mathbb{R}))$, but even in $\mathcal{P}_{f\oplus\hat{h}}(\tilde{K}\times\mathcal{P}(\mathbb{R}))$ (see Definition 2.2) and therefore $\mathcal{P}_{f\oplus\hat{h}}(\tilde{K}\times\mathcal{P}_h(\mathbb{R}))$ by Lemma 5.2 (b) below. In that context, $C|_{\tilde{K}\times\mathcal{P}_h(\mathbb{R})} \in \Phi_{f\oplus\hat{h}}(\tilde{K}\times\mathcal{P}_h(\mathbb{R}))$, so

$$\begin{aligned}\int_{\mathbb{R}} C(x,\pi_x)\,\mu(dx) &= \int_{\tilde{K}\times\mathcal{P}_h(\mathbb{R})} C(x,p)\,J(\pi)(dx,dp)\\ &= \lim_{n\to+\infty}\int_{\tilde{K}\times\mathcal{P}_h(\mathbb{R})} C(x,p)\,J(\pi^n)(dx,dp)\\ &= \lim_{n\to+\infty}\int_{\mathbb{R}} C(x,\pi^n_x)\,\mu^n(dx)\\ &= \lim_{n\to+\infty} V_C^M(\mu^n,\nu^n)\\ &= V_C^M(\mu,\nu),\end{aligned}$$

hence π is optimal for $V_C^M(\mu,\nu)$.

Next, we drop the additional joint-continuity assumption on C. Since $\nu(g) < +\infty$, there exists by the de La Vallée Poussin theorem $h \in \mathcal{F}_g^+(\mathbb{R})$ such that $\nu(h) < +\infty$. For $N \in \mathbb{N}^*$, let $B_N = \{p \in \mathcal{P}_g(\mathbb{R}) \mid p(h) \le N\}$, which is a compact subset of $\mathcal{P}_g(\mathbb{R})$ by Lemma 5.6 below, and $\mathcal{C}(B_N)$ be the set of continuous functions from B_N to $\mathbb{R}$, endowed with the topology of uniform convergence. The map $\phi^N\colon \mathbb{R}\to\mathcal{C}(B_N)$ given by $\phi^N(x) = C(x,\cdot)|_{B_N}$ is Borel measurable due to [2, Theorem 4.55]. Let $\varepsilon \in (0,1)$. By Lusin's theorem there is for every $N\in\mathbb{N}^*$ a compact set $K^N \subset \mathbb{R}$ such that the restriction $\phi^N|_{K^N}$ is continuous and $\mu(K^N) \ge 1 - \frac{\varepsilon}{2^N}$. We have

$$\mu\left(\bigcap_{N\in\mathbb{N}^*} K^N\right) \ge 1 - \sum_{N\in\mathbb{N}^*}\mu\left((K^N)^c\right) \ge 1 - \sum_{N\in\mathbb{N}^*}\frac{\varepsilon}{2^N} = 1-\varepsilon.$$

Let $K^\varepsilon = \bigcap_{N\in\mathbb{N}^*} K^N$, then for all $N\in\mathbb{N}^*$ the restriction $\phi^N|_{K^\varepsilon}$ is continuous. We claim that $C|_{K^\varepsilon\times\mathcal{P}_h(\mathbb{R})}$ is continuous w.r.t. the product topology of $\mathbb{R}\times\mathcal{P}_h(\mathbb{R})$. To this end, take any sequence $(x_k,p_k)_{k\in\mathbb{N}} \in (K^\varepsilon\times\mathcal{P}_h(\mathbb{R}))^{\mathbb{N}}$ with limit point $(x,p)\in K^\varepsilon\times\mathcal{P}_h(\mathbb{R})$. Since $p_k \to p$ in $\mathcal{P}_h(\mathbb{R})$ as k goes to $+\infty$, the sequence $(p_k(h))_{k\in\mathbb{N}}$ is convergent and therefore bounded so there exists $N\in\mathbb{N}$ such that $p, p_k \in B_N$ for all $k\in\mathbb{N}$. As $\phi^N(x_k)$ converges uniformly to $\phi^N(x)$, we have

$$C(x_k,p_k) = \phi^N(x_k)(p_k) \underset{k\to+\infty}{\longrightarrow} \phi^N(x,p) = C(x,p).$$

Therefore, $C|_{K^\varepsilon\times\mathcal{P}_h(\mathbb{R})}$ is continuous.

Let $\mu^\varepsilon = \frac{1}{\mu(K^\varepsilon)}\mu|_{K^\varepsilon}$, $\pi^\varepsilon = \mu^\varepsilon\times\pi_x = \frac{1}{\mu(K^\varepsilon)}\pi|_{K^\varepsilon\times\mathbb{R}}$ and ν^ε be the second marginal of π^ε. Obviously μ^ε is concentrated on K^ε. Since $\mu(K^\varepsilon)\mu^\varepsilon \le \mu$ and $\pi^\varepsilon_x = \pi_x$, π^ε is martingale C-monotone and satisfies $(x,\pi^\varepsilon_x)\in\Gamma$ for $\mu^\varepsilon(dx)$-almost every x. Finally, $\mu(K^\varepsilon)\nu^\varepsilon(h) = \int_{K^\varepsilon}\pi_x(h)\,\mu(dx) \le \nu(h) < +\infty$, hence $\nu^\varepsilon \in \mathcal{P}_h(\mathbb{R})$. Therefore the reasoning of the first part applied with $(K^\varepsilon,\mu^\varepsilon,\nu^\varepsilon,\pi^\varepsilon)$ replacing $(\tilde{K},\mu,\nu,\pi)$ proves that π^ε is optimal for $V_C^M(\mu^\varepsilon,\nu^\varepsilon)$.

Next, we convince ourselves that $J(\pi^\varepsilon)$ converges to $J(\pi)$ stably in $\mathcal{P}_{f\oplus\hat{h}}(\mathbb{R}\times\mathcal{P}(\mathbb{R}))$: let $\psi\colon\mathbb{R}\times\mathcal{P}(\mathbb{R})$ be measurable and absolutely dominated by a positive multiple of $f\oplus\hat{h}$, then

$$J(\pi^\varepsilon)(\psi) = \int_{\mathbb{R}}\psi(x,\pi_x)\,\mu^\varepsilon(dx) = \frac{1}{\mu(K^\varepsilon)}\int_{K^\varepsilon}\psi(x,\pi_x)\,\mu(dx) \underset{\varepsilon\to0}{\longrightarrow} \int_{\mathbb{R}}\psi(x,\pi_x)\,\mu(dx) = J(\pi)(\psi),$$

where we employed dominated convergence and that $1-\varepsilon \le \mu(K^\varepsilon) \le 1$. In particular, the marginals $(\mu^\varepsilon)_{\varepsilon>0}$ converge to μ in $\mathcal{P}_f(\mathbb{R})$ and strongly, whereas the marginals $(\nu^\varepsilon)_{\varepsilon>0}$ converge to ν in $\mathcal{P}_h(\mathbb{R})$ for $\varepsilon\searrow0$. Using item (d) of Lemma 5.12 yields

$$\lim_{n\to+\infty} V_C^M(\mu^{1/n},\nu^{1/n}) = \lim_{n\to+\infty}\int_{\mathbb{R}\times\mathcal{P}_h(\mathbb{R})} C(x,p)\,J(\pi^{1/n})(dx,dp)$$

$$= \int_{\mathbb{R}\times\mathcal{P}_h(\mathbb{R})} C(x,p)\, J(\pi)(dx,dp) = \int_{\mathbb{R}} C(x,\pi_x)\,\mu(dx).$$

The marginal sequences $(\mu^{1/n})_{n\in\mathbb{N}^*}$ and $(\nu^{1/n})_{n\in\mathbb{N}^*}$ satisfy the assumptions of Theorem 2.6. Hence, by item (b') with (B) of the very same theorem we have that

$$V_C^M(\mu,\nu) = \lim_{n\to+\infty} V_C^M(\mu^{1/n},\nu^{1/n}) = \int_{\mathbb{R}} C(x,\pi_x)\,\mu(dx),$$

proving optimality of π. □

5 Appendix

The adapted weak topology is defined as the initial topology under the trivial embedding map J from $\mathcal{P}(X\times Y)$ to $\mathcal{P}(X\times\mathcal{P}(Y))$, namely

$$J : \mathcal{P}(X\times Y) \ni \pi = \mu\times\pi_x \mapsto \mu(dx)\,\delta_{\pi_x}(dp) \in \mathcal{P}(X\times\mathcal{P}(Y)). \tag{5.1}$$

Conversely, it is widely known that we can associate to a probability measure $P\in\mathcal{P}(\mathcal{P}(Y))$ its intensity $I(P)(dy) = \int_{p\in\mathcal{P}(Y)} p(dy)\,P(dp) \in \mathcal{P}(Y)$. For the extended space $\mathcal{P}(X\times\mathcal{P}(Y))$ we naturally define the extended intensity $\hat{I}$ by

$$\hat{I} : \mathcal{P}(X\times\mathcal{P}(Y)) \ni P \mapsto \int_{p\in\mathcal{P}(Y)} p(dy)\,P(dx,dp) \in \mathcal{P}(X\times Y), \tag{5.2}$$

which associates to each $P\in\mathcal{P}(X\times\mathcal{P}(Y))$ a coupling $\hat{I}(P)\in\mathcal{P}(X\times Y)$. We note that $\hat{I}$ is the left-inverse of J.

For $(\mu,\nu)\in\mathcal{P}(X)\times\mathcal{P}(Y)$, we define the set of extended couplings $\Lambda(\mu,\nu)$ between μ and ν as the set of probability measures on $\mathcal{P}(X\times\mathcal{P}(Y))$ whose extended intensity is a coupling between μ and ν, that is

$$\Lambda(\mu,\nu) = \left\{ P = \mu\times P_x \in \mathcal{P}(X\times\mathcal{P}(Y)) \mid \int_{(x,p)\in X\times\mathcal{P}(Y)} p(dy)\,P(dx,dp) = \nu(dy) \right\}. \tag{5.3}$$

If $f\colon X\to\mathbb{R}^+$ and $g\colon Y\to\mathbb{R}^+$ are measurable functions, then any $P\in\Lambda(\mu,\nu)$ satisfies

$$\begin{aligned} &\int_{X\times\mathcal{P}(Y)} f(x)\,P(dx,dp) = \int_X f(x)\,\mu(dx), \\ \text{and} \quad &\int_{X\times\mathcal{P}(Y)} \int_Y g(y)\,p(dy)\,P(dx,dp) = \int_Y g(y)\,\nu(dy). \end{aligned} \tag{5.4}$$

For $\mu,\nu\in\mathcal{P}^1(\mathbb{R}^d)$, we define the martingale counterpart $\Lambda_M(\mu,\nu)$ of $\Lambda(\mu,\nu)$ as the set of probability measures on $\mathcal{P}^1(\mathbb{R}^d\times\mathcal{P}^1(\mathbb{R}^d))$, whose extended intensity is a martingale coupling between μ and ν, that is

$$\Lambda_M(\mu,\nu) = \left\{ P\in\Lambda(\mu,\nu) \mid \int_{\mathbb{R}^d} y\,p(dy) = x,\ P(dx,dp)\text{-almost everywhere} \right\}. \tag{5.5}$$

5.1 Extension from $\mathcal{P}^r$ to $\mathcal{P}_f$.

We recall that unless explicitly stated otherwise, $\mathcal{P}(Y)$ is endowed with the weak convergence topology, and for any continuous map $f : Y\to[1,+\infty)$ we endow the space $\mathcal{P}_f(Y) = \{p\in\mathcal{P}(Y) \mid p(f) < +\infty\}$ with the topology induced by the following convergence: a sequence $(p_k)_{k\in\mathbb{N}}\in\mathcal{P}_f(Y)^{\mathbb{N}}$ converges in $\mathcal{P}_f(Y)$ to p iff p_k converges weakly to p and $p_k(f)$ converges to $p(f)$ as $k\to+\infty$.

As mentioned in Section 2, this extension emerged from the need to overcome the inconvenience of the non-compacity of the $\mathcal{W}_r$-balls $\{p\in\mathcal{P}^r(Y) \mid \mathcal{W}_r(p,\delta_{y_0}) \le R\}$, $R>0$ for the $\mathcal{W}_r$-distance topology. All the following lemmas together show that this extension enjoys nearly the same flexibility as the usual Wasserstein distance topology and most importantly benefits of a helpful compacity result, see Lemma 5.6 below.

Remark 5.1. We continue with some remarks on the structure of $\mathcal{P}_f(Y)$:

(1) Convergence in $\mathcal{P}_f(Y)$ can be described differently: let $(p_k)_{k\in\mathbb{N}}$ converge to p in $\mathcal{P}_f(Y)$, and let $g \in \mathcal{C}(Y)$ be such that $0 \le g \le f$. By Portmanteau's theorem we have $p(g) \le \liminf_{k\to+\infty} p_k(g)$ and $p(f)-p(g) = p(f-g) \le \liminf_{k\to+\infty} p_k(f-g) = p(f) - \limsup_{k\to+\infty} p_k(g)$, hence $\limsup_{k\to+\infty} p_k(g) \le p(g)$. We deduce that

$$p_k \underset{k\to+\infty}{\longrightarrow} p \text{ in } \mathcal{P}_f(Y) \iff p_k(g) \underset{k\to+\infty}{\longrightarrow} p(g), \quad \forall g \in \Phi_f(Y), \tag{5.6}$$

when $\Phi_f(Y) := \{g \in \mathcal{C}(Y) \mid g \text{ is absolutely dominated by a positive multiple of } f\}$.

It is immediate that for $r \ge 1$, this topology is finer than the one induced by $\mathcal{W}_r$ on $\mathcal{P}_f(Y)$ if f belongs to the set $\mathcal{F}^r(Y)$ of real-valued continuous functions defined on Y and bounded from below by $y \mapsto 1 + d_Y^r(y, y_0)$.

(2) The set $\mathcal{P}_f(Y)$ is naturally embedded into the set $\mathcal{M}_+(Y)$ of all bounded positive Borel measures on Y, endowed with the weak topology, via the following continuous injection

$$\iota\colon \mathcal{P}_f(Y) \to \mathcal{M}_+(Y), \quad \iota(p)(dy) = f(y)\, p(dy).$$

Clearly, the topology on $\mathcal{P}_f(Y)$ coincides with the initial topology under ι. Even more, the set $\iota(\mathcal{P}_f(Y)) = \{m \in \mathcal{M}_+(Y) \colon m(\frac{1}{f}) = 1\}$ is a closed subset of $\mathcal{M}_+(Y)$ since $\frac{1}{f}$ is continuous and bounded. As such, we deduce that $\mathcal{P}_f(Y)$ is a Polish space.

(3) By [12, Theorem 8.3.2 and the preceding discussion], we have that the weak topology on $\mathcal{M}_+(Y)$ is induced by the norm

$$\|m_1 - m_2\|_0 := \sup_{\substack{g\colon Y\to[-1,1] \\ g \text{ is 1-Lipschitz}}} (m_1(g) - m_2(g)).$$

This permits us to define a metric on $\mathcal{P}_f(Y)$ via

$$\overline{\mathcal{W}}_f(p,q) := \sup_{\substack{g\colon Y\to[-1,1], \\ g \text{ is 1-Lipschitz}}} (p(fg) - q(fg)) = \|\iota(p) - \iota(q)\|_0. \tag{5.7}$$

Thus, $\overline{\mathcal{W}}_f$ is a complete metric compatible with the topology on $\mathcal{P}_f(Y)$.

From now on, we equip $\mathcal{P}_f(Y)$ with $\overline{\mathcal{W}}_f$. A continuous function $f\colon Y \to [1,+\infty)$ can naturally be lifted to a continuous function $\hat{f}\colon \mathcal{P}_f(Y) \to [1,+\infty)$ by setting

$$\hat{f}(p) := p(f). \tag{5.8}$$

Let us recall some notation. For any probability $P \in \mathcal{P}(\mathcal{P}(Y))$ we denote its intensity $I(P) \in \mathcal{P}(Y)$, defined by $I(P)(dy) = \int_{\mathcal{P}(Y)} p(dy)\, P(dp)$. Then we have $P(\hat{f}) = I(P)(f)$. For two maps $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$, we denote $f \oplus g : X \times Y \ni (x,y) \mapsto f(x) + g(y)$.

As we are solely interested in topological properties, the next lemma shows that we can freely switch between the spaces $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$, $\mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$, and $\mathcal{P}^1(\mathcal{P}_f(Y))$, the latter's definition being given by (1.1) with $(1, \mathcal{P}_f(Y), \overline{\mathcal{W}}_f)$ replacing (r, X, d_X).

Lemma 5.2. *(a) Let $f : Y \to [1,+\infty)$ be continuous. Then*

$$\mathcal{P}_{\hat{f}}(\mathcal{P}(Y)) = \mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y)), \tag{5.9}$$

and their topologies are equal. If moreover one endows $\mathcal{P}_f(Y)$ with the metric $\overline{\mathcal{W}}_f$ defined by (2.1), then

$$\mathcal{P}_{\hat{f}}(\mathcal{P}(Y)) = \mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y)) = \mathcal{P}^1(\mathcal{P}_f(Y)), \tag{5.10}$$

and their topologies are equal.

(b) Let $f : X \to [1,+\infty)$ and $g : Y \to [1,+\infty)$ be continuous. Then

$$\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}(Y)) = \mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}_g(Y)), \tag{5.11}$$

and their topologies are equal.

Remark 5.3. The equalities (5.9), (5.10) and (5.11) are to be understood up to an identification, namely we consider that for two measurable sets $Z' \subset Z$, a probability measure $p \in \mathcal{P}(Z)$ belongs to $\mathcal{P}(Z')$ if $p(Z') = 1$, the underlying identification being of course between $p \in \mathcal{P}(Z)$ and the probability measure $p' \in \mathcal{P}(Z')$ defined for any measurable subset $A \subset Z'$ by $p'(A) = p(A \cap Z')$.

Proof. Let us prove (a). The inclusion $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y)) \supset \mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$ is straightforward. Conversely, let $P \in \mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$. Then by definition,

$$P(\hat{f}) = \int_{\mathcal{P}(Y)} p(f)\, P(dp) < +\infty,$$

which can only hold if $p(f)$ is $P(dp)$-almost everywhere finite, or equivalently $P(\mathcal{P}_f(Y)) = 1$, hence $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y)) \subset \mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$ and therefore we have equality. To see that the two topologies match, let us show that

$$P^k \underset{k\to+\infty}{\longrightarrow} P \text{ in } \mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y)) \iff P^k \underset{k\to+\infty}{\longrightarrow} P \text{ in } \mathcal{P}_{\hat{f}}(\mathcal{P}(Y)).$$

Since the topology on $\mathcal{P}_f(Y)$ is finer than the weak topology on $\mathcal{P}(Y)$, we have $\mathcal{C}(\mathcal{P}(Y)) \subset \mathcal{C}(\mathcal{P}_f(Y))$, so the direct implication is trivial. Conversely, suppose that P^k converges in $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$ to P as k goes to $+\infty$. Let $h \in \mathcal{C}(Y)$ be bounded. Then $\hat{h} \in \mathcal{C}(\mathcal{P}(Y))$ is bounded, and $I(P^k)(h) = P^k(\hat{h})$ converges to $P(\hat{h}) = I(P)(h)$ as k goes to $+\infty$. Moreover $I(P^k)(f) = P^k(\hat{f})$ converges to $P(\hat{f}) = I(P)(f)$. This shows that $(I(P^k))_{k\in\mathbb{N}}$ converges in $\mathcal{P}_f(Y)$ to $I(P)$. Therefore $\{I(P^k) \mid k \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}_f(Y)$. We deduce by Lemma 5.4 below that $\{P^k \mid k \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$. Let Q be an accumulation point of $(P^k)_{k\in\mathbb{N}}$ in $\mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$. In particular Q is by the direct implication shown above an accumulation point of $(P^k)_{k\in\mathbb{N}}$ in $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$, hence $Q = P$ by uniqueness of the limit since the topology is metrisable and therefore Hausdorff.

Let us now prove the second part of (a). We endow $\mathcal{P}_f(Y)$ with the metric $\overline{\mathcal{W}}_f$. To see that the sets $\mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$ and $\mathcal{P}^1(\mathcal{P}_f(Y))$ are the same, we find

$$P(\hat{f}) < +\infty \iff \int_{\mathcal{P}(Y)} p(f)\, P(dp) < +\infty \iff \int_{\mathcal{P}(Y)} \overline{\mathcal{W}}_f(p, \delta_{y_0})\, P(dp) < +\infty,$$

which is an easy consequence, as well as the equality of the topologies, of

$$\forall p \in \mathcal{P}_f(Y), \quad p(f) - f(y_0) \le \overline{\mathcal{W}}_f(p, \delta_{y_0}) \le p(f) + f(y_0).$$

Let us now prove (b). We derive the equality $\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}(Y)) = \mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}_g(Y))$ as in (a) since

$$P(f \oplus \hat{g}) = \int_{X\times\mathcal{P}(Y)} (f(x) + p(g))\, P(dx, dp) < +\infty,$$

which can only hold if the second marginal of P is concentrated on $\mathcal{P}_g(Y)$. To see that the topologies are equal, the only nontrivial part is, as in (a), to show that if $(P^k)_{k\in\mathbb{N}}$ converges in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$, then $\{P^k \mid k \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}_g(Y))$. Let then $(P^k)_{k\in\mathbb{N}}$ converge in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$ to some P. Recall moreover the definition of the extended intensity $\hat{I}$ given by (5.2). Let $h_1 : X \to \mathbb{R}$ and $h_2 : Y \to \mathbb{R}$ be two continuous and bounded maps. Then the map $H : X \times \mathcal{P}(Y) \ni (x,p) \mapsto \int_Y h_1(x)h_2(y)\, p(dy)$ is continuous and bounded. Denoting $h : (x,y) \mapsto h_1(x)h_2(y)$, we deduce that $\hat{I}(P^k)(h) = P^k(H)$ converges to $P(H) = \hat{I}(P)(h)$ as k goes to $+\infty$. Hence $(\hat{I}(P^k))_{k\in\mathbb{N}}$ converges weakly to $\hat{I}(P)$. Then by continuity of the projections the first marginal μ^k, resp. the second marginal ν^k of $\hat{I}(P^k)$ converges weakly to the first marginal μ, resp. the second marginal ν of $\hat{I}(P)$. Since the maps $f\oplus\hat{0} : (x,p) \mapsto f(x)$ and $0\oplus\hat{g} : (x,p) \mapsto \hat{g}(p)$ belong to $\mathcal{C}(X \times \mathcal{P}(Y))$ and are dominated by $f \oplus \hat{g}$, we also have that

$$\mu^k(f) = P^k(f \oplus \hat{0}) \underset{k\to+\infty}{\longrightarrow} P(f \oplus \hat{0}) = \mu(f) \quad \text{and} \quad \nu^k(g) = P^k(0 \oplus g) \underset{k\to+\infty}{\longrightarrow} P(0 \oplus g) = \nu(g),$$

which shows that $(\mu^k,\nu^k)_{k\in\mathbb{N}}$ converges in $\mathcal{P}_f(X)\times\mathcal{P}_g(Y)$ to (μ,ν). Therefore $(\hat{I}(P^k))_{k\in\mathbb{N}}$ is tight in $\mathcal{P}(X\times Y)$ and both projections $\{\mu^k \mid k \in \mathbb{N}\}$ and $\{\nu^k \mid k \in \mathbb{N}\}$ are relatively compact respectively in $\mathcal{P}_f(X)$ and $\mathcal{P}_f(Y)$, so by Lemma 5.7 below $\{P^k \mid k \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}_g(Y))$, which proves the claim. □

Lemma 5.4. *A set $\mathcal{A} \subset \mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$ is relatively compact if and only if the set of its intensities $I(\mathcal{A}) \subset \mathcal{P}_f(Y)$ is relatively compact.*

Proof. The first implication follows as in [6, Lemma 2.4] by continuity of I, c.f. Lemma 5.9 below. The reverse implication can be shown by pursuing the same idea as in [6, Lemma 2.4] with slight modification: instead of considering the map $y \mapsto d_Y(y,y')^t$ we use $y \mapsto f(y)$. □

Lemma 5.5. *A set $\mathcal{A} \subset \mathcal{P}_f(Y)$ is relatively compact if and only if it is tight and*

$$\forall\varepsilon > 0, \quad \exists R > 0, \quad \sup_{\mu\in\mathcal{A}} \int_{\{y\in Y \mid f(y)\geq R\}} f(y)\,\mu(dy) < \varepsilon.$$

Proof. The proof of this lemma runs along the lines of [6, Lemma 2.5] when replacing $y \mapsto d_Y(y,y')^t$ by $y \mapsto f(y)$. □

For $g : \mathbb{R}^d \to [1,+\infty)$, recall the definition (4.2) of the set $\mathcal{F}_g^+(\mathbb{R}^d)$.

Lemma 5.6. *Let $d \in \mathbb{N}^*$ and $\mathbb{R}^d$ be endowed with a norm $|\cdot|$, and let $g\colon \mathbb{R}^d \to [1,+\infty)$ be continuous. Then for all $f \in \mathcal{F}_g^+(\mathbb{R}^d)$, the set $B_R := \{p \in \mathcal{P}(\mathbb{R}) \mid p(f) \leq R\}$ is a compact subset of $\mathcal{P}_g(\mathbb{R}^d)$.*

Proof. Let $R \geq 0$, $(p_n)_{n\in\mathbb{N}}$ be a sequence in $B_R^{\mathbb{N}}$ and $\varepsilon > 0$. There exists $r > 0$ such that for all $x \in \mathbb{R}^d$, $|x| \geq r$ implies $f(x) \geq \frac{R}{\varepsilon}$. Let $K = \{x \in \mathbb{R}^d \mid |x| \leq r\}$. For all $n \in \mathbb{N}$, we have $R \geq p_n(f) \geq p_n(\mathbb{R}^d\backslash K)\frac{R}{\varepsilon}$, hence $p_n(\mathbb{R}^d\backslash K) \leq \varepsilon$. So $(p_n)_{n\in\mathbb{N}}$ is tight, and by Prokhorov's theorem there exists a subsequence, still denoted $(p_n)_{n\in\mathbb{N}}$ for notational simplicity, which converges weakly to $p \in \mathcal{P}(\mathbb{R})$. Since f is continuous and nonnegative, we have by Portmanteau's theorem

$$p(f) \leq \liminf_{n\to+\infty} p_n(f) \leq R,$$

so $p(f) \in B_R$. It remains to show that this convergence also holds in $\mathcal{W}_g$. By Skorokhod's representation theorem, there exists for all $n \in \mathbb{N}$ a random variable $Z_n \sim p_n$, such that $(Z_n)_{n\in\mathbb{N}}$ converges almost surely to a random variable $Z \sim p$. For all $n \in \mathbb{N}$ we have

$$p_n(g) = \mathbb{E}[g(Z_n)] \leq \mathbb{E}[f(Z_n)] = p_n(f) \leq R,$$

so by the de La Vallée Poussin theorem, $(g(Z_n))_{n\in\mathbb{N}}$ is uniformly integrable. We deduce by

$$\lim_{n\to+\infty} p_n(g) = p(g)$$

and (5.6) that $(p_n)_{n\in\mathbb{N}}$ converges in $\mathcal{P}_g(\mathbb{R}^d)$ to p, so B_R is compact. □

For a probability measure $\pi \in \mathcal{P}(X \times Y)$, we denote by $\mathrm{proj}_X(\pi)$ and $\mathrm{proj}_Y(\pi)$ its X-marginal and Y-marginal, respectively. Recall moreover the definition of the extended intensity $\hat{I}$ given by (5.2).

Lemma 5.7. *Let $f\colon X \to [1,+\infty)$ and $g\colon Y \to [1,+\infty)$ be continuous. The following are equivalent:*

(a) A set $\Pi \subset \mathcal{P}(X \times Y)$ is tight and both projections, $\mathrm{proj}_X(\Pi) \subset \mathcal{P}_f(X)$ *and* $\mathrm{proj}_Y(\Pi) \subset \mathcal{P}_g(Y)$, *are relatively compact.*

(b) $J(\Pi)$ as a subset of $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}_g(Y))$ is relatively compact.

Conversely, the following are equivalent:

(a') $\Lambda\subset\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}_g(Y))$ is relatively compact.

(b') $\hat{I}(\Lambda)\subset\mathcal{P}(X\times Y)$ is tight, and both projections, $\mathrm{proj}_X(\hat{I}(\Lambda))\subset\mathcal{P}_f(X)$ and $\mathrm{proj}_Y(\hat{I}(\Lambda))\subset\mathcal{P}_g(Y)$, are relatively compact.

Proof. For this lemma works the same proof as in [6, Lemma 2.6] when using Lemma 5.4, the characterisation of relative compactness given in Lemma 5.5 and continuity of $\hat{I}$, see Lemma 5.9. ☐

Proposition 5.8. *Let $f\colon X\to[1,+\infty)$ and $g\colon Y\to[1,+\infty)$ be continuous functions, and $C\colon X\times\mathcal{P}_g(Y)\to\mathbb{R}\cup\{+\infty\}$ be lower semicontinuous and bounded from below by a negative multiple of $f\oplus\hat{g}$. Then*

(a) The map

$$\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}_g(Y))\ni P\mapsto\int_{X\times\mathcal{P}_g(Y)}C(x,p)\,P(dx,dp)\tag{5.12}$$

is lower semicontinuous.

(b) Suppose in addition that for all $x\in X$, the map $p\mapsto C(x,p)$ is convex. Then

$$\mathcal{P}_{f\oplus g}(X\times Y)\ni\pi\mapsto\int_X C(x,\pi_x)\,\mu(dx),\tag{5.13}$$

where μ denotes the X-marginal of π, is lower semicontinuous.

Proof. Lower semicontinuity of (5.12) is obtained by standard arguments. To see (5.13), let $(\pi_k)_{k\in\mathbb{N}}\in\mathcal{P}_{f\oplus g}(X\times Y)^{\mathbb{N}}$ converge in $\mathcal{P}_{f\oplus g}(X\times Y)$ to some π. We find by the first part of Lemma 5.7 an accumulation point $P\in\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$ of $(J(\pi^k))_{k\in\mathbb{N}}$. By possibly passing to a subsequence we can assume that $P^k:=J(\pi^k)$ converges to P in $\mathcal{P}_{f\oplus\hat{g}}(X\times P(Y))$ as k goes to $+\infty$. Write μ^k, $k\in\mathbb{N}$ and μ for the X-marginal of π^k and π, respectively. Due to (5.12), we obtain

$$\begin{aligned}\liminf_{k\to+\infty}\int_X C(x,\pi^k_x)\,\mu^k(dx)&=\liminf_{k\to+\infty}\int_{X\times\mathcal{P}_f(Y)}C(x,p)\,P^k(dx,dp)\\&\geq\int_{X\times\mathcal{P}_f(Y)}C(x,p)\,P(dx,dp)\\&\geq\int_X C\left(x,I(P_x)\right)\,\mu(dx)\\&=\int_X C\left(x,\hat{I}(P)_x\right)\,\mu(dx),\end{aligned}$$

where we used Proposition 5.10 below for the last inequality. Since $\hat{I}$ is continuous by Lemma 5.9 below, we find that $\pi^k=\hat{I}(P^k)\to\hat{I}(P)$ and $\hat{I}(P^k)=\pi^k\to\pi$ as $k\to+\infty$. But the weak topology is Hausdorff and therefore $\pi=\hat{I}(P)$ yielding

$$\liminf_{k\to+\infty}\int_X C(x,\pi^k_x)\,\mu^k(dx)\geq\int_X C(x,\pi_x)\,\mu(dx),$$

and thus (5.13). ☐

Lemma 5.9. *Let $f\colon X\to[1,+\infty)$ and $g\colon Y\to[1,+\infty)$ be continuous. The maps*

$$I\colon\mathcal{P}_{\hat{g}}(\mathcal{P}(Y))\to\mathcal{P}_g(Y),\quad I(P)(dy):=\int_{\mathcal{P}(Y)}p(dy)\,P(dp),\tag{5.14}$$

$$\hat{I}\colon\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))\to P_{f\oplus g}(X\times Y),\quad\hat{I}(P)(dx,dy):=\int_{X\times\mathcal{P}(Y)}p(dy)\,P(dx,dp),\tag{5.15}$$

are continuous.

Proof. Let $(P^k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{P}_{\hat{g}}(\mathcal{P}(Y))$ with limit point P. Let $h \in \mathcal{C}_b(Y)$, then $\hat{h} \in \mathcal{C}_b(P(Y))$. Thus,

$$\lim_{k\to+\infty} I(P^k)(h) = \lim_{k\to+\infty} P^k(\hat{h}) = P(\hat{h}) = I(P)(h),$$
$$\lim_{k\to+\infty} I(P^k)(g) = \lim_{k\to+\infty} P^k(\hat{g}) = P(\hat{g}) = I(P)(g),$$

which shows by (5.6) continuity of I.

Next, let $(P^k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}(Y))$ converging to P. Let $h \in \mathcal{C}_b(X \times Y)$, then $\check{h}(x,p) := \int_Y h(x,y)\, p(dy)$ is contained in $\mathcal{C}_b(X \times \mathcal{P}(Y))$. Again, we find

$$\lim_{k\to+\infty} \hat{I}(P^k)(h) = \lim_{k\to+\infty} P^k(\check{h}) = P(\check{h}) = \hat{I}(P)(h),$$
$$\lim_{k\to+\infty} \hat{I}(P^k)(f\oplus g) = \lim_{k\to+\infty} P^k(f\oplus\hat{g}) = P(f\oplus\hat{g}) = \hat{I}(P)(f\oplus g),$$

whereby we derive continuity of $\hat{I}$ by virtue of (5.6). □

Proposition 5.10. *Let $f\colon X \to [1,+\infty)$ be continuous, $C : \mathcal{P}_f(Y) \to \mathbb{R}$ be convex, lower semicontinuous and lower bounded by a negative multiple of $\hat{f}$. Then for all $Q \in \mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$ holds*

$$C\left(I(Q)\right) \le \int_{\mathcal{P}_f(Y)} C(p)\, Q(dp). \tag{5.16}$$

If moreover C is strictly convex, then (5.16) is an equality iff $Q = \delta_{I(Q)}$.

Proof. Let $Q \in \mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$, $P_n : \Omega \to \mathcal{P}(Y)$, $n \in \mathbb{N}^*$ be independent random variables identically distributed according to Q and $\mathcal{G} \subset \Phi_{\hat{f}}(\mathcal{P}(Y))$ be a countable family which determines the convergence in $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$ (see [16, Theorem 4.5.(b)]). By the law of large numbers, almost surely, for all $\psi \in \mathcal{G}$,

$$\frac{1}{n}\sum_{k=1}^n \psi(P_k) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[\psi(P_1)] = Q(\psi) \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^n C(P_k) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[C(P_1)] = Q(C). \tag{5.17}$$

Let $\omega \in \Omega$ be such that (5.17) holds when evaluated at ω and set $p_n = P_n(\omega)$ for $n \in \mathbb{N}^*$. Then $\left(\frac{1}{n}\sum_{k=1}^n \delta_{p_k}\right)_{n\in\mathbb{N}}$ converges in $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$ to Q. By Lemma 5.9, $\frac{1}{n}\sum_{k=1}^n p_k$ converges to $I(Q)$ as $n \to +\infty$. By lower semicontinuity of C for the first inequality, convexity of C for the second one and (5.17) evaluated at ω for the equality, we get

$$C(I(Q)) \le \liminf_{n\to+\infty} C\left(\frac{1}{n}\sum_{k=1}^n p_k\right) \le \liminf_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n C(p_k) = Q(C). \tag{5.18}$$

If $Q = \delta_{I(Q)}$ we have trivially equality in (5.16). So, assume that Q is not concentrated on a single point, and that C is strictly convex. There are $h \in \Phi_f(Y)$ and $b \in \mathbb{R}$ such that $A = \{p \in \mathcal{P}_f(Y) \mid p(h) \le b\}$ satisfies

$$Q(A) > 0 \text{ and } Q(A^c) > 0. \tag{5.19}$$

Indeed, pick any points $p_1, p_2 \in \mathcal{P}_f(Y)$, $p_1 \neq p_2$ in the support of Q, then the Hahn-Banach separation theorem provides $h \in \Phi_f(Y)$ and $b \in \mathbb{R}$ such that $p_1(h) < b < p_2(h)$. As both points lie in the support of Q, and $\{p \in \mathcal{P}_f(Y) \mid p(f) < b\}$ and $\{p \in \mathcal{P}_f(Y) \mid p(f) > b\}$ are open subsets containing p_1 and p_2, respectively, we obtain (5.19). Write $Q_1(dp) := \mathbb{1}_A \frac{Q(dp)}{Q(A)}$ and $Q_2(dp) := \mathbb{1}_{A^c} \frac{Q(dp)}{Q(A^c)}$. By the definition of A, we have that $I(Q_1)(h) < b < I(Q_2)(h)$ and especially $I(Q_1) \neq I(Q_2)$. By (5.16) we find

$$\int_{\mathcal{P}_f(y)} C(p)\, Q_1(dp) \ge C(I(Q_1)) \text{ and } \int_{\mathcal{P}_f(Y)} C(p)\, Q_2(dp) \ge C(I(Q_2)).$$

Hence, as $Q = Q(A)Q_1 + (1 - Q(A))Q_2$ we get

$$\begin{aligned}\int_{\mathcal{P}_f(Y)} C(p)\, Q(dp) &= \int_{\mathcal{P}_f(Y)} C(p) Q(A)\, Q_1(dp) + \int_{\mathcal{P}_f(Y)} C(p) Q(A^c)\, Q_2(dp) \\ &\geq Q(A) C(I(Q_1)) + (1 - Q(A)) C(I(Q_2)) \\ &> C(Q(A) I(Q_1) + (1 - Q(A)) I(Q_2)) = C(I(Q)),\end{aligned}$$

where we used $I(Q_1) \neq I(Q_2)$ and strict convexity for the last inequality. ☐

5.2 A Pormanteau-like theorem for Carathéodory maps

Let $(\pi^k)_{k\in\mathbb{N}}$ be a sequence of probability measures defined on $X \times Y$ converging in $\mathcal{P}_{f\oplus g}(X \times Y)$ to π, and $c : X \times Y \to \mathbb{R}$ be a (lower) Carathéodory map, that is a measurable function which is (lower semi-)continuous in its second argument. The goal of the present section is to determine in which situation we can connect the asymptotic behaviour of $\int_{X\times Y} c(x, y)\, \pi^k(dx, dy)$ and $\int_{X\times Y} c(x, y)\, \pi(dx, dy)$. We recall that $(\pi^k)_{k\in\mathbb{N}}$ is said to converge stably to π iff for every bounded measurable map $g : X \to \mathbb{R}$ and bounded continuous map $h : Y \to \mathbb{R}$

$$\int_{X\times Y} g(x)h(y)\, \pi^k(dx, dy) \underset{k\to+\infty}{\longrightarrow} \int_{X\times Y} g(x)h(y)\, \pi(dx, dy). \tag{5.20}$$

We say that a sequence $(\mu^k)_{k\in\mathbb{N}}$ of probability measures on $\mathcal{P}(X)$ K-converges in total variation to μ iff for every subsequence $(\mu^{k_i})_{i\in\mathbb{N}}$ we have

$$\frac{1}{n}\sum_{i=1}^{n} \mu^{k_i} \underset{n\to+\infty}{\longrightarrow} \mu \quad \text{in total variation.}$$

Lemma 5.11. *Let $\pi, \pi^k \in \mathcal{P}(X \times Y)$, $k \in \mathbb{N}$ be with respective first marginal μ, μ^k. All of the following statements are equivalent:*

(a) $(\pi^k)_{k\in\mathbb{N}}$ converges to π stably.

(b) $(\pi^k)_{k\in\mathbb{N}}$ converges to π weakly and $(\mu^k)_{k\in\mathbb{N}}$ converges strongly to μ.

(c) $(\pi^k)_{k\in\mathbb{N}}$ converges to π weakly and every subsequence of $(\mu^k)_{k\in\mathbb{N}}$ has an in total variation K-convergent sub-subsequence with limit μ.

Proof. We prove "(a) $\implies$ (b)". The definition of stable convergence given by (5.20) is in the Polish set-up by [13, Theorem 8.10.65 (ii)] equivalent to

$$\int_{X\times Y} c(x, y)\, \pi^k(dx, dy) \underset{k\to+\infty}{\longrightarrow} \int_{X\times Y} c(x, y)\, \pi(dx, dy)$$

for all $c\colon X \times Y \to \mathbb{R}$ which are bounded and Carathéodory. Thus, stable convergence is stronger than weak convergence. For all measurable subsets $A \subset X$, we find by setting $g = \mathbb{1}_A$ and $h = 1$ in (5.20) that

$$\mu^k(A) \underset{k\to+\infty}{\longrightarrow} \mu(A).$$

Next we show "(b) $\implies$ (c)".

Let $\mu^k(dx) = \rho^k(x)\, \mu(dx) + \eta^k(dx)$ be the Lebesgue decomposition of μ^k w.r.t. μ. Since η^k is singular to μ there is $N^k \in \mathcal{B}(X)$ such that $\eta^k(N^k) = \eta^k(X)$ and $\mu(N^k) = 0$. Define $N = \bigcup_{k\in\mathbb{N}} N^k \in \mathcal{B}(X)$, then $\eta^k(N) = \eta^k(X)$ for all $k \in \mathbb{N}$ and $\mu(N)$ vanishes as a countable union of null sets. Thus, $\eta^k(X) = \mu^k(N) \to \mu(N) = 0$ as $k \to +\infty$. Since $(\rho^k)_{k\in\mathbb{N}}$ is bounded in $L^1(\mu)$ there is by Komlós theorem a K-convergent subsequence to some limiting function $\rho \in L^1(\mu)$. We have

$$\frac{1}{n}\sum_{l=1}^{n} \rho^{k_l} \underset{n\to+\infty}{\longrightarrow} \rho, \quad \mu\text{-a.s.}$$

By [12, Corollary 4.5.7] the above convergence even holds in $L^1(\mu)$. We find for any measurable subset $A \subset X$

$$\int_A \frac{1}{n}\sum_{l=1}^{n} \rho^{k_l}(x)\,\mu(dx) \underset{n\to+\infty}{\longrightarrow} \int_A \rho(x)\,\mu(dx) = \mu(A).$$

Hence, $\rho(x) = 1$, $\mu(dx)$-almost surely and

$$\mathrm{TV}\left(\frac{1}{n}\sum_{l=1}^{n}\mu^{k_l}, \mu\right) = \eta^k(X) + \int_X \left|\frac{1}{n}\sum_{l=1}^{n}\rho^{k_l}(x) - 1\right| \mu(dx) \underset{n\to+\infty}{\longrightarrow} 0.$$

Finally we show "(c) $\implies$ (a)". If $(\pi^k)_{k\in\mathbb{N}}$ does not converge stably to π, then there is a bounded Carathéodory function $c\colon X\times Y \to \mathbb{R}$, such that

$$\limsup_{k\to+\infty}\left|\int_{X\times Y} c(x,y)\,\pi^k(dx,dy) - \int_{X\times Y} c(x,y)\,\pi(dx,dy)\right| > 0.$$

Hence, w.l.o.g. there is a subsequence $(\pi^{k_j})_{j\in\mathbb{N}}$ such that $\pi^{k_j}(c) \geq \pi(c) + \delta$ for some $\delta > 0$. Especially, we have for any sub-subsequence $(\pi^{k_{j_i}})_{i\in\mathbb{N}}$ of $(\pi^{k_j})_{j\in\mathbb{N}}$ that

$$\frac{1}{n}\sum_{i=1}^{n}\pi^{k_{j_i}}(c) \geq \pi(c) + \delta, \tag{5.21}$$

whereby the Cesàro-means of the sub-subsequence are not stably convergent. By assumption there exists a subsequence $(\mu^{k_{j_i}})_{i\in\mathbb{N}}$ of $(\mu^{k_j})_{j\in\mathbb{N}}$ which K-converges in total variation to μ. For $n \in \mathbb{N}^*$ define

$$\hat\mu^n = \frac{1}{n}\sum_{i=1}^{n}\mu^{k_{j_i}} \quad \text{and} \quad \hat\pi^n = \frac{1}{n}\sum_{i=1}^{n}\pi^{k_{j_i}}.$$

We will show that $(\hat\pi^n)_{n\in\mathbb{N}^*}$ converges stably to π, which will contradict (5.21) and end the proof. Let $\hat\mu^n(dx) = \hat\pi^n(x)\,\mu(dx) + \hat\eta^n(dx)$ be the Lebesgue decomposition of $\hat\mu^n$ w.r.t. μ. Define the auxiliary sequence

$$\tilde\pi^n(dx,dy) = \big((1\wedge\hat\rho^n(x))\,\hat\pi^n_x(dy) + (1-\hat\rho^n(x))^+\,\pi_x(dy)\big)\,\mu(dx).$$

Let $c\colon X\times Y\to\mathbb{R}$ be Carathéodory and absolutely bounded by K, then

$$\begin{aligned}
&\left|\int_{X\times Y} c(x,y)\;\tilde\pi^n(dx,dy) - \int_{X\times Y} c(x,y)\,\hat\pi^n(dx,dy)\right| \\
&\quad\leq K\left(\int_X |\hat\rho^n(x) - 1\wedge\hat\rho^n(x)|\;\mu(dx) + \int_X (1-\hat\rho^n(x))^+\,\mu(dx) + \hat\eta^n(X)\right) \\
&\quad\leq K\left(\int_X |\hat\rho^n(x) - 1|\;\mu(dx) + 2\hat\eta^n(X)\right) \\
&\quad\leq 2K\,\mathrm{TV}(\hat\mu^n,\mu) \underset{n\to+\infty}{\longrightarrow} 0.
\end{aligned} \tag{5.22}$$

In particular, we have found that $(\tilde\pi^n)_{n\in\mathbb{N}^*}$ converges to π weakly. Note that the first marginal $\tilde\pi^n$ is μ, and therefore [24, Lemma 2.1] yields stable convergence of $\tilde\pi^n$ to π as $n\to+\infty$. By (5.22), we find that $(\hat\pi^n)_{n\in\mathbb{N}^*}$ also stably converges to π. □

Lemma 5.12. *Let $f\colon X\to[1,+\infty)$ and $g\colon Y\to[1,+\infty)$ be continuous, and let $(\pi^k)_{k\in\mathbb{N}}$ converge to π in $\mathcal{P}_{f\oplus g}(X\times Y)$.*

(a) If $c\colon X\times Y\to\mathbb{R}\cup\{+\infty\}$ is lower semicontinuous and bounded from below by a negative multiple of $g\oplus h$, then

$$\liminf_{k\to+\infty}\int_{X\times Y} c(x,y)\,\pi^k(dx,dy) \geq \int_{X\times Y} c(x,y)\,\pi(dx,dy).$$

(b) If $c\colon X\times Y\to\mathbb{R}$ is continuous and absolutely bounded by positive multiple of $g\oplus h$, then

$$\lim_{k\to+\infty}\int_{X\times Y} c(x,y)\,\pi^k(dx,dy)=\int_{X\times Y} c(x,y)\,\pi(dx,dy).$$

(c) If $c\colon X\times Y\to\mathbb{R}\cup\{+\infty\}$ is lower Carathéodory and bounded from below by a negative multiple of $g\oplus h$, and π^k converges to π stably, then

$$\liminf_{k\to+\infty}\int_{X\times Y} c(x,y)\,\pi^k(dx,dy)\geq\int_{X\times Y} c(x,y)\,\pi(dx,dy).$$

(d) If $c\colon X\times Y\to\mathbb{R}$ is Carathéodory and absolutely bounded by a positive multiple of $g\oplus h$, and π^k converges to π stably, then

$$\lim_{k\to+\infty}\int_{X\times Y} c(x,y)\,\pi^k(dx,dy)=\int_{X\times Y} c(x,y)\,\pi(dx,dy).$$

Proof. These results are well-known. Note that by [12, Theorem 8.10.65] we have for every bounded lower Carathéodory map c that $\pi\mapsto\pi(c)$ is lower semicontinuous w.r.t. the topology of stable convergence. $\square$

5.3 On the continuity of the marginal distributions of the stretched Brownian motion

The following Lemma shows that the stretched Brownian motion provides a convenient tool to approximate two probability measures in the convex order with atomless ones still in the convex order.

Lemma 5.13. *Let $\mu,\nu\in\mathcal{P}^2(\mathbb{R})$ be such that $\mu\leq_c\nu$ and (μ,ν) consists of a single irreducible component $I=(l,r)$. Let $(M^*_t)_{t\in[0,1]}$ be the unidimensional stretched Brownian motion from μ to ν. Then*

(a) For each $t\in(0,1)$ the distribution ν_t of M^*_t is atomless.

(b) For all $s,t\in[0,1]$ such that $s<t$, (ν_s,ν_t) consists of the single irreducible component I.

Proof. Let us first prove (a). Let $t\in(0,1)$ and $y\in\mathbb{R}$. Let $\gamma=\mathcal{N}(0,1)$ be the unidimensional standard normal distribution and $C_2:\mathbb{R}\times\mathcal{P}^2(\mathbb{R})\to\mathbb{R}$ be defined for all $(x,p)\in\mathbb{R}\times\mathcal{P}^2(\mathbb{R})$ by $C_2(x,p)=\mathcal{W}_2^2(p,\gamma)$. Let $V^M_{C_2}$ be the value function given by (WMOT) for the cost function C_2 and $\pi^*\in\Pi_M(\mu,\nu)$ be optimal for $V^M_{C_2}(\mu,\nu)$. According to (2.10),

$$M^*_t=\varphi_t(X,B_t),$$

where $X\sim\mu$ is a random variable independent of the Brownian motion $(B_s)_{s\in[0,1]}$ and $\varphi_t:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is defined for all $(x,b)\in\mathbb{R}^2$ by

$$\varphi_t(x,b)=\int_{\mathbb{R}} F^{-1}_{\pi^*_x}\left(F_\gamma\left(\sqrt{1-t}y+b\right)\right)\,\gamma(dy). \tag{5.23}$$

In order to prove that $\mathbb{P}(\{M^*_t=y\})=0$, it clearly suffices to show that for $\mu(dx)$-almost every $x\in\mathbb{R}$,

$$\mathbb{P}(\{\varphi_t(x,B_t)=y\})=0. \tag{5.24}$$

The map $T=F^{-1}_{\pi^*_x}\circ F_\gamma$ is nondecreasing. Let $b>b'$ and assume that $T(\sqrt{1-t}y+b)=T(\sqrt{1-t}y+b')$ for all $y\in\mathbb{R}$. By monotonicity of T we deduce that T has to be constant on all intervals of $\mathbb{R}$ of length $b-b'$ and therefore on $\mathbb{R}$. So assume that T is not constant. Then there exists $y\in\mathbb{R}$ such that $T(\sqrt{1-t}y+b)>T(\sqrt{1-t}y+b')$. As F_γ is continuous and $F^{-1}_{\pi^*_x}$ is left-continuous, we find an $\varepsilon>0$ such that

$$\forall y'\in(y-\varepsilon,y],\quad T(\sqrt{1-t}y'+b)>T(\sqrt{1-t}y'+b'),$$

hence $\varphi_t(x,b) > \varphi_t(x,b')$. We deduce that $b \mapsto \varphi_t(x,b)$ is increasing and therefore one-to-one, hence the equation $\varphi_t(x,b) = y$ has at most one solution b^*. Denoting by b^* any real number if the latter equation has no solution, we then have

$$\mathbb{P}(\{\varphi_t(x,B_t) = y\}) \leq \mathbb{P}(\{B_t = b^*\}) = 0.$$

In order to prove (5.24) and conclude the proof, it remains to show that for $\mu(dx)$-almost every $x \in \mathbb{R}$, the map $F^{-1}_{\pi^*_x} \circ F_\gamma$ is not constant. Since γ is the unidimensional standard normal distribution and π^*_x is a martingale kernel, it is equivalent to show that for $\mu(dx)$-almost every $x \in \mathbb{R}$, $\pi^*_x \neq \delta_x$. This is done using the WMOT monotonicity principle. By (2.14) there exists a martingale C_2-monotone set $\Gamma \subset \mathbb{R} \times \mathcal{P}^1(\mathbb{R})$ such that $(x, \pi^*_x) \in \Gamma$ for all x in a μ-full set $A \subset \mathbb{R}$. This implies that for all $x, x' \in A$ and $p, p' \in \mathcal{P}^1(\mathbb{R})$ such that $\pi^*_x + \pi^*_{x'} = p + p'$, $\int_\mathbb{R} y\, p(dy) = x$ and $\int_\mathbb{R} y\, p'(dy) = x'$, we have

$$\mathcal{W}_2^2(\pi^*_x, \gamma) + \mathcal{W}_2^2(\pi^*_{x'}, \gamma) \leq \mathcal{W}_2^2(p, \gamma) + \mathcal{W}_2^2(p', \gamma). \tag{5.25}$$

Let $x \in A$. To conclude, it suffices to show that $\pi^*_x \neq \delta_x$. Note that if (p, p') is admissible for (5.25), so is $(\frac{1}{2}(\pi^*_x + p), \frac{1}{2}(\pi^*_{x'} + p'))$. In the proof of Lemma 3.1 we show that $q \mapsto \mathcal{W}_2^2(q, \gamma)$ is strictly convex. Therefore, if $p \neq \pi^*_x$ or $p' \neq \pi^*_{x'}$, then

$$\begin{aligned}
\mathcal{W}_2^2(\pi^*_x, \gamma) + \mathcal{W}_2^2(\pi^*_{x'}, \gamma) &\leq \mathcal{W}_2^2\left(\frac{1}{2}(\pi^*_x + p), \gamma\right) + \mathcal{W}_2^2\left(\frac{1}{2}(\pi^*_{x'} + p'), \gamma\right) \\
&< \frac{1}{2}\mathcal{W}_2^2(\pi^*_x, \gamma) + \frac{1}{2}\mathcal{W}_2^2(p, \gamma) + \frac{1}{2}\mathcal{W}_2^2(\pi^*_{x'}, \gamma) + \frac{1}{2}\mathcal{W}_2^2(p', \gamma),
\end{aligned}$$

and the inequality (5.25) is strict. To show that $\pi^*_x \neq \delta_x$ and thereby end the proof, we deduce that it suffices to find $x' \in A$ and two measures $p, p' \in \mathcal{P}^1(\mathbb{R})$ such that

$$\delta_x + \pi^*_{x'} = p + p', \quad \int_\mathbb{R} y\, p(dy) = x, \quad \int_\mathbb{R} y\, p'(dy) = x', \quad p \neq \delta_x, \tag{5.26}$$

$$\text{and} \quad \mathcal{W}_2^2(p, \gamma) + \mathcal{W}_2^2(p', \gamma) \leq \mathcal{W}_2^2(\pi^*_{x'}, \gamma) + \mathcal{W}_2^2(\delta_x, \gamma). \tag{5.27}$$

Suppose that

$$\mu(\{x' \in (l, x] \mid \pi^*_{x'}((x, r)) > 0\}) + \mu(\{x' \in (x, r) \mid \pi^*_{x'}((l, x)) > 0\}) = 0. \tag{5.28}$$

Then for $\mu(dx')$-almost every $x' \in (l, r)$, the sign of $y - x$ is $\pi^*_{x'}(dy)$-almost everywhere constant equal to the sign of $x' - x$, so using the martingale property of $\pi^*_{x'}$ in the third equality, we get that

$$\begin{aligned}
u_\nu(x) = \int_\mathbb{R} |y - x|\, \nu(dy) &= \int_\mathbb{R} \left(\int_\mathbb{R} |y - x|\, \pi^*_{x'}(dy) \right) \mu(dx') \\
&= \int_{(l,x]} \left| \int_\mathbb{R} y\, \pi^*_{x'}(dy) - x \right| \mu(dx') + \int_{(x,r)} \left| \int_\mathbb{R} y\, \pi^*_{x'}(dy) - x \right| \mu(dx') \\
&= \int_\mathbb{R} |x' - x|\, \mu(dx') = u_\mu(x),
\end{aligned}$$

which contradicts the irreducibility of (μ, ν). We deduce that (5.28) does not hold, hence there exists $x' \in A$ such that $x' \leq x$ and $\pi^*_{x'}((x, r)) > 0$, or $x' > x$ and $\pi^*_{x'}((l, x)) > 0$. Since $\pi^*_{x'}$ has mean x', we can find in both cases $\tilde{x} < x < \tilde{y}$ such that $\pi^*_{x'}((\tilde{x}, x)) > 0$ and $\pi^*_{x'}((x, \tilde{y})) > 0$, which implies

$$0 < F_{\pi^*_{x'}}(x-) \leq F_{\pi^*_{x'}}(x) < 1. \tag{5.29}$$

Define for $\alpha, \beta \in [0, 1]$

$$p_\alpha = \int_0^\alpha \delta_{F^{-1}_{\pi^*_{x'}}(u)}\, du, \quad q_\beta = \int_{1-\beta}^1 \delta_{F^{-1}_{\pi^*_{x'}}(u)}\, du.$$

Let $c = \left(\int_0^{F_{\pi^*_{x'}}(x-)}(x - F^{-1}_{\pi^*_{x'}}(u))\,du\right) \wedge \left(\int_{F_{\pi^*_{x'}}(x)}^1 (F^{-1}_{\pi^*_{x'}(x)}(u) - x)\,du\right)$. Since for all $u \in (0,1)$, $u > F_{\pi^*_{x'}}(x) \iff F^{-1}_{\pi^*_{x'}}(u) > x$ and $u < F_{\pi^*_{x'}}(x-) \implies F^{-1}_{\pi^*_{x'}}(u) < x \implies u \le F_{\pi^*_{x'}}(x-)$, the maps $\alpha \mapsto \int_0^\alpha (x - F^{-1}_{\pi^*_{x'}}(u))\,du$ and $\beta \mapsto \int_{1-\beta}^1 (F^{-1}_{\pi^*_{x'}}(u) - x)\,du$ are nondecreasing respectively on $[0, F_{\pi^*_{x'}}(x-)]$ and $[0, 1 - F_{\pi^*_{x'}}(x)]$. Moreover, those two maps are continuous, so we deduce the existence of $\alpha' \in (0, F_{\pi^*_{x'}}(x-)]$ and $\beta' \in (0, 1 - F_{\pi^*_{x'}}(x)]$ such that they both equal c respectively at $\alpha = \alpha'$ and $\beta = \beta'$, hence

$$\int_0^{\alpha'} (F^{-1}_{\pi^*_{x'}}(u) - x)\,du + \int_{1-\beta'}^1 (F^{-1}_{\pi^*_{x'}}(u) - x)\,du = \int_{\mathbb{R}} y\, p_{\alpha'}(dy) + \int_{\mathbb{R}} y\, q_{\beta'}(dy) - (\alpha' + \beta')x = 0.$$

Note that (5.29) implies that $\alpha' + \beta' \in (0,1]$. Then the measures $p = (1 - \alpha' - \beta')\delta_x + p_{\alpha'} + q_{\beta'}$ and $p' = (\alpha' + \beta')\delta_x + \pi^*_{x'} - p_{\alpha'} - q_{\beta'}$ satisfy (5.26). Let $\chi \in \Pi(\pi^*_{x'}, \gamma)$ be the $\mathcal{W}_2$-optimal coupling and denote by $\hat{p} = \pi^*_{x'} - p_{\alpha'} - q_{\beta'}$ and $\tilde{p} = p_{\alpha'} + q_{\beta'}$. Then

$$\begin{aligned}
&\left(\tilde{p}(dy)\,\chi_y(dz) + \delta_x(dy) \int_{t\in\mathbb{R}} \chi_t(dz)\,\hat{p}(dt)\right) \in \Pi(p, \gamma),\\
&\left(\hat{p}(dy)\,\chi_y(dz) + \delta_x(dy) \int_{t\in\mathbb{R}} \chi_t(dz)\,\tilde{p}(dt)\right) \in \Pi(p', \gamma),
\end{aligned}$$

hence

$$\begin{aligned}
\mathcal{W}_2^2(p,\gamma) &\le \int_{\mathbb{R}\times\mathbb{R}} |y - z|^2\, \tilde{p}(dy)\,\chi_y(dz) + \int_{\mathbb{R}\times\mathbb{R}} |x - z|^2\, \hat{p}(dt)\,\chi_t(dz),\\
\mathcal{W}_2^2(p',\gamma) &\le \int_{\mathbb{R}\times\mathbb{R}} |y - z|^2\, \hat{p}(dy)\,\chi_y(dz) + \int_{\mathbb{R}\times\mathbb{R}} |x - z|^2\, \tilde{p}(dt)\,\chi_t(dz).
\end{aligned}$$

Combining these inequalities yields

$$\begin{aligned}
\mathcal{W}_2^2(p,\gamma) + \mathcal{W}_2^2(p',\gamma) &\le \int_{\mathbb{R}\times\mathbb{R}} |y - z|^2\, (\tilde{p} + \hat{p})(dy)\,\chi_y(dz) + \int_{\mathbb{R}\times\mathbb{R}} |x - z|^2\, (\hat{p} + \tilde{p})(dt)\,\chi_t(dz)\\
&= \int_{\mathbb{R}\times\mathbb{R}} |y - z|^2\, \chi(dy, dz) + \int_{\mathbb{R}} |x - z|^2\, \gamma(dz)\\
&= \mathcal{W}_2^2(\pi^*_{x'}, \gamma) + \mathcal{W}_2^2(\delta_x, \gamma),
\end{aligned}$$

which proves (5.27) and completes the proof.

Let us now prove (b). Let $s, t \in [0,1]$ be such that $s < t$. Since $\mu \le_c \nu_s \le_c \nu_t \le_c \nu$, we have $u_\mu \le u_{\nu_s} \le u_{\nu_t} \le u_\nu$, hence $u_{\nu_s} = u_{\nu_t}$ on $I^\complement$. Let $z \in I$. Then

$$\begin{aligned}
u_{\nu_s}(z) &= \mathbb{E}[|M^*_s - z|] = \mathbb{E}[|\mathbb{E}[M^*_t - z | X, (B_u)_{u\in[0,s]}]|]\\
&\le \mathbb{E}[\mathbb{E}[|M^*_t - z| | X, (B_u)_{u\in[0,s]}] = \mathbb{E}[|M^*_t - z|] = u_{\nu_t}(z).
\end{aligned} \tag{5.30}$$

Let us show that the inequality above is strict. This is equivalent to show that given X and $(B_u)_{u\in[0,s]}$, the sign of $M^*_t - z$ is not almost surely constant. Suppose that $\mathbb{P}(M^*_s \le z) > 0$, the case $\mathbb{P}(M^*_s \ge z) > 0$ being treated symmetrically. Then it suffices to find a Borel subset $A \subset \mathbb{R}$ such that

$$\mathbb{P}(M^*_t > z, X \in A, M^*_s \le z) > 0, \tag{5.31}$$

since the martingale property would then imply that $\mathbb{P}(M^*_t < z, X \in A, M^*_s \le z)$ is positive as well. The pair (μ, ν) being irreducible, we have that

$$\mu(A := \{x \in (-\infty, z] \mid \pi^*_x((z, +\infty)) > 0\}) > 0.$$

For fixed $x, y \in \mathbb{R}$, the map $b \mapsto T^t_{x,y}(b) = F^{-1}_{\pi^*_x}(F_\gamma(\sqrt{1-t}y+b))$ is non-decreasing where $\lim_{b\to+\infty} T^t_{x,y}(b) = \lim_{u\nearrow 1} F^{-1}_{\pi^*_x}(u)$. Recall that $y \mapsto T^t_{x,y}(b)$ and $y \mapsto T^s_{x,y}$ are γ-integrable, therefore we have due to monotone convergence

$$\lim_{b\to+\infty} \varphi_t(x,b) = \lim_{b\to+\infty} \int_{\mathbb{R}} T^t_{x,y}(b)\,\gamma(dy) = \lim_{u\nearrow 1} F^{-1}_{\pi^*_x}(u),$$
$$\lim_{b\to-\infty} \varphi_s(x,b) = \lim_{b\to-\infty} \int_{\mathbb{R}} T^s_{x,y}(b)\,\gamma(dy) = \lim_{u\searrow 0} F^{-1}_{\pi^*_x}(u),$$

and, in particular,

$$\forall x \in A, \quad \lim_{b\to+\infty} \varphi_t(x,b) > z \quad \text{and} \quad \lim_{b\to-\infty} \varphi_s(x,b) < z.$$

Again, recall that $x \mapsto \varphi_t(x,b)$ and $x \mapsto \varphi_s(x,b)$ are μ-integrable, therefore we find due to monotone convergence

$$\lim_{b\to+\infty} \int_A \varphi_t(x,b)\,\mu(dx) = \int_A \lim_{b\to+\infty} \varphi_t(x,b)\,\mu(dx) > z\mu(A),$$
$$\lim_{b\to-\infty} \int_A \varphi_s(x,b)\,\mu(dx) = \int_A \lim_{b\to-\infty} \varphi_s(x,b)\,\mu(dx) < z\mu(A).$$

Hence, there are $b_0, b_1 \in \mathbb{R}$ and $A' \subset A$, $\mu(A') > 0$, such that

$$\varphi_t(x,b) > z \text{ and } \varphi_s(x,b') < z,$$

for every $x \in A'$, $b \geq b_0$ and $b' \leq b_1$. Then

$$\begin{aligned}\mathbb{P}(M^*_t > z, X \in A,\ M^*_s \leq z) &\geq \mathbb{P}(B_t \geq b_0, X \in A', M^*_s \leq z)\\ &= \mathbb{P}(B_t \geq b_0, X \in A', B_s \leq b_1) > 0,\end{aligned}$$

which proves (5.31). Hence the inequality in (5.30) is strict and $u_{\nu_s} < u_{\nu_t}$ on I. □

Corollary 5.14. *Let $\mu, \nu \in \mathcal{P}^1(\mathbb{R})$ be such that $\mu \leq_c \nu$ and (μ,ν) is irreducible with component I. Let $\varepsilon > 0$, then there is an atomless $\tilde\nu \in \mathcal{P}^1(\mathbb{R})$ such that*

$$\mathcal{W}_1(\tilde\nu, \nu) < \varepsilon, \quad \mu \leq_c \tilde\nu \leq_c \nu \quad \textit{and} \quad (\mu, \tilde\nu) \textit{ is irreducible with component } I.$$

Proof. It is clear that whenever two measures (μ,ν) have finite second moment, the stretched Brownian motion provides by Corollary 2.8 and Lemma 5.13 (a) a continuous interpolation $(\mu_t)_{t\in[0,1]}$, where $\mu_0 = \mu$ and $\mu_1 = \nu$, such that μ_t is atomless for $t \in (0,1)$. We are going to extend such an interpolation to a case where only first moments are finite. To work around this issue, assume for a moment that we can introduce an intermediary measure $\bar\nu$ with $\mu \leq_c \bar\nu \leq_c \nu$, where the decomposition into irreducible components $(I_n)_{n\in N}$ of $(\bar\nu,\nu)$ consists only of bounded intervals, and $\bar\nu(J) = 0$ for $J = \mathbb{R} \setminus \bigcup_{n\in N} I_n$. For all $n \in N$, let $(\bar\nu|_{I_n}, \nu_n)$ be the irreducible pair associated with I_n in the decomposition of $(\bar\nu,\nu)$. Since I_n is bounded, $\nu_n \in \mathcal{P}^2(\mathbb{R})$ so we can consider the stretched Brownian motion $(M^n_t)_{t\in[0,1]}$ from $\frac{1}{\bar\nu(I_n)}\bar\nu|_{I_n}$ to $\frac{1}{\bar\nu(I_n)}\nu_n$. Since $t \mapsto M^n_t$ is almost surely continuous on $[0,1]$ and I_n is bounded, we find by dominated convergence that the law of M^n_t converges in $\mathcal{W}_1$ to $\frac{1}{\bar\nu(I_n)}\nu_n$ as t tends to 1. Therefore we find for each I_n a time $t_n \in (0,1)$ such that for each $n \in N$ the distribution $\frac{1}{\bar\nu(I_n)}\bar\nu_n$ of $M^n_{t_n}$ satisfies

$$\mathcal{W}_1(\bar\nu_n, \nu_n) < \frac{\varepsilon}{2^{n+1}}.$$

In particular, $\bar\nu_n$ is atomless by Lemma 5.13 (a). We set

$$\tilde\nu := \sum_{n\in N} \bar\nu_n.$$

Thus,

$$\mathcal{W}_1(\tilde{\nu},\nu) < \sum_{n\in N} \frac{\varepsilon}{2^{n+1}} \leq \varepsilon.$$

Moreover there holds

$$u_\mu \leq u_{\bar{\nu}} = \sum_{n\in N} u_{\bar{\nu}|_{I_n}} \leq \sum_{n\in N} u_{\bar{\nu}_n} = u_{\tilde{\nu}} \leq \sum_{n\in N} u_{\nu_n} = u_\nu, \tag{5.32}$$

which implies $\mu \leq_c \tilde{\nu} \leq_c \nu$. For all $n \in N$, $(\bar{\nu}|_{I_n}, \bar{\nu}_n)$ is irreducible by Lemma 5.13 (b), so the second inequality in (5.32) is strict on $\bigcup_{n\in N} I_n$. Since $u_\mu < u_\nu = u_{\tilde{\nu}}$ on $I\backslash\bigcup_{n\in N} I_n$, we deduce that $u_\mu < u_{\tilde{\nu}}$ on I. On $I^{\complement}$, we have $u_\mu = u_\nu$, which implies $u_\mu = u_{\tilde{\nu}}$. Therefore $\{u_\mu < u_{\tilde{\nu}}\} = I$ and $(\mu,\tilde{\nu})$ is irreducible with component I. It remains to show that there is a measure $\hat{\nu}$ with the above mentioned properties.

If $\nu \notin \mathcal{P}^2(\mathbb{R})$, then I has to be unbounded. For simplicity, we assume that $I = \mathbb{R}$, since if $I = (-\infty, b)$ or $I = (a, +\infty)$ with $a, b \in \mathbb{R}$ the construction below also works in these cases with the obvious modifications. To this end, we define iteratively $u_1^1 = \frac{1}{2} = u_2^1$, and for $n \in \mathbb{N}^*$, we choose $u_1^{n+1} \in (0, \frac{1}{2^{n+1}} \wedge u_1^n)$ and $u_2^{n+1} \in ((1 - \frac{1}{2^{n+1}}) \vee u_2^n, 1)$ such that

$$|\{x \in \mathbb{R} \mid F_\nu(x) \in ((u_1^{n+1}, u_1^n))\}| > 1 \quad \text{and} \quad |\{x \in \mathbb{R} \mid F_\nu(x) \in ((u_2^n, u_2^{n+1}))\}| > 1, \tag{5.33}$$

which is possible as $\nu((-\infty, R)) \wedge \nu((R, +\infty)) > 0$ for all $R \in \mathbb{R}$. We have that $\lim_{n\to+\infty} u_1^n = 0$ and $\lim_{n\to+\infty} u_2^n = 1$, and set

$$\hat{\nu} := \sum_{n\in\mathbb{N}^*} \left((u_1^n - u_1^{n+1}) \delta_{\int_{u_1^{n+1}}^{u_1^n} F_\nu^{-1}(u)\, \frac{du}{u_1^n - u_1^{n+1}}} + (u_2^{n+1} - u_2^n) \delta_{\int_{u_2^n}^{u_2^{n+1}} F_\nu^{-1}(u)\, \frac{du}{u_2^{n+1} - u_2^n}} \right),$$

which entails us to define $\bar{\nu} := \hat{\nu} \vee_c \mu$. For all $x \in \mathbb{R}$ we have by inverse transform sampling for the last equality

$$\begin{aligned}
u_{\hat{\nu}}(x) &= \sum_{n\in\mathbb{N}^*} \left(\left| \int_{u_1^{n+1}}^{u_1^n} (F_\nu^{-1}(u) - x)\, du \right| + \left| \int_{u_2^n}^{u_2^{n+1}} (F_\nu^{-1}(u) - x)\, du \right| \right) \\
&\leq \sum_{n\in\mathbb{N}^*} \left(\int_{u_1^{n+1}}^{u_1^n} |F_\nu^{-1}(u) - x|\, du + \int_{u_2^n}^{u_2^{n+1}} |F_\nu^{-1}(u) - x|\, du \right) \\
&= \int_0^1 |F_\nu^{-1}(u) - x|\, du = u_\nu(x),
\end{aligned}$$

where the inequality is strict iff there exists $n \in \mathbb{N}^*$ such that $F_\nu^{-1} - x$ is not constant on (u_1^{n+1}, u_1^n) or (u_2^n, u_2^{n+1}). By monotonicity of F_ν^{-1} the strict inequality is equivalent to $x \in (F_\nu^{-1}(u_1^{n+1}), F_\nu^{-1}(u_1^n))$ or $(F_\nu^{-1}(u_2^n), F_\nu^{-1}(u_2^{n+1}))$ for some $n \in \mathbb{N}^*$. We deduce that $\hat{\nu} \leq_c \nu$ and therefore $\mu \leq_c \bar{\nu} \leq_c \nu$, and the irreducible components of $(\hat{\nu}, \nu)$ are given by the intervals

$$I_n^1 = (F_\nu^{-1}(u_1^{n+1}), F_\nu^{-1}(u_1^n)) \quad \text{and} \quad I_n^2 = (F_\nu^{-1}(u_2^n), F_\nu^{-1}(u_2^{n+1})), \quad n \in \mathbb{N}^*, \tag{5.34}$$

which are indeed nonempty by (5.33) and bounded. In particular for all $n \in \mathbb{N}^*$ we have

$$u_{\hat{\nu}}(F_\nu^{-1}(u_1^n)) = u_\nu(F_\nu^{-1}(u_1^n)) \quad \text{and} \quad u_{\hat{\nu}}(F_\nu^{-1}(u_2^n)) = u_\nu(F_\nu^{-1}(u_2^n)). \tag{5.35}$$

Since $u_{\hat{\nu}} \leq u_\mu \vee u_{\hat{\nu}} = u_{\bar{\nu}} \leq u_\nu$ and (μ,ν) is irreducible, (5.34) is also the decomposition into irreducible components of $(\hat{\nu}, \nu)$, which consists solely of bounded intervals.

To conclude, it remains to show that

$$\bar{\nu}\left(\mathbb{R}\backslash \bigcup_{n\in\mathbb{N}^*} (I_n^1 \cup I_n^2) \right) = \bar{\nu}\left(\left\{ F_\nu^{-1}(u_1^n) \mid n \in \mathbb{N}^* \right\} \cup \left\{ F_\nu^{-1}(u_2^n) \mid n \in \mathbb{N}^* \right\} \right) = 0. \tag{5.36}$$

For each $n \in \mathbb{N}^*$ and bounded neighbourhood of $F_\nu^{-1}(u_1^n)$, there is by irreducibility of (μ, ν) and continuity of potential functions a $\delta > 0$ such that $u_\mu + \delta < u_\nu$ on this neighbourhood. Thus, for y close enough to $F_\nu^{-1}(u_1^n)$, we have $u_\mu(y) < u_{\hat{\nu}}(y)$ due to (5.35), hence

$$u_{\bar{\nu}}(y) = u_{\hat{\nu}}(y) \vee u_\mu(y) = u_{\hat{\nu}}(y), \quad \text{for } y \text{ close enough to } F_\nu^{-1}(u_1^n). \tag{5.37}$$

For each $n \in \mathbb{N}^*$, it is clear from the definition of $\hat{\nu}$ that its restriction to the closure of I_n^1 is concentrated on a single point in I_n^1 and therefore does not charges the boundaries of I_n^1. We recall the easy fact that the potential function of a probability measure is linear on an open intervall iff this measure does not charge this interval. We deduce, with use of (5.37), that we can find an open neighbourhood of $F_\nu^{-1}(u_n^1)$ such that $u_{\hat{\nu}}$ and therefore $u_{\bar{\nu}}$ is linear, which implies that $\bar{\nu}$ does not put mass on $\{F_\nu^{-1}(u_1^n) \mid n \in \mathbb{N}^*\}$. Analogously, we find that $\tilde{\nu}$ does not charge $\{F_\nu^{-1}(u_2^n) \mid n \in \mathbb{N}^*\}$, which proves (5.36). $\square$

References

[1] D. J. Aldous. Weak convergence and general theory of processes. Unpublished incomplete draft of monograph; Department of Statistics, University of California, Berkeley, CA 94720, July 1981.

[2] C. D. Aliprantis and K. C. Border. *Infinite dimensional analysis: A hitchhiker's guide*. Springer, 3rd edition, 2006.

[3] L. Ambrosio and N. Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, volume 2062 of *Lecture Notes in Math.*, pages 1–155. Springer, Heidelberg, 2013.

[4] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. All adapted topologies are equal. *Probability Theory and Related Fields*, pages 1–48, 2020.

[5] J. Backhoff-Veraguas, M. Beiglböck, M. Huesmann, and S. Källblad. Martingale Benamou–Brenier: a probabilistic perspective. *Annals of Probability*, 48(5):2258–2289, 2020.

[6] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, 58(6):203, 2019.

[7] J. Backhoff-Veraguas and G. Pammer. Stability of martingale optimal transport and weak optimal transport. *arXiv e-prints:1904.04171*, April 2019.

[8] J. Backhoff-Veraguas and G. Pammer. Applications of weak transport theory. *arXiv preprint arXiv:2003.05338*, 2020.

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

[10] M. Beiglböck, B. Jourdain, W. Margheriti, and G. Pammer. Approximation of martingale couplings on the line in the weak adapted topology. 2020. [Online; posted October 2020, https://homepage.univie.ac.at/gudmund.pammer/files/papers/ApproxMOT.pdf].

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

[12] V. I. Bogachev. *Measure theory*, volume 1. Springer Science and Business Media, 2007.

[13] V. I. Bogachev. *Measure theory*, volume 2. Springer Science and Business Media, 2007.

[14] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[15] D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651, 1978.

[16] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley-Blackwell, Hoboken, 2nd revised edition, 2005.

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

[18] N. Gozlan and C. Léonard. Transport inequalities. a survey. *Markov Processes And Related Fields*, 16(4):635–736, 2010.

[19] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali. Characterization of a class of weak transport-entropy inequalities on the line. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 54(3):1667–1693, 2018.

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

[21] C. Griessler. An extended footnote on finitely minimal martingale measures. *arXiv preprint arXiv:1606.03106*, 2016.

[22] J. Guyon, R. Menegaux, and M. Nutz. Bounds for vix futures given S&P 500 smiles. *Finance and Stochastics*, 21(3):593–630, 2017.

[23] M. F. Hellwig. Sequential decisions under uncertainty and the maximum theorem. *Journal of Mathematical Economics*, 25(4):443–464, 1996.

[24] D. Lacker. Dense sets of joint distributions appearing in filtration enlargements, stochastic control, and causal optimal transport. *arXiv e-prints:1805.03185*, May 2018.

[25] M. Ledoux. *The concentration of measure phenomenon*. Number 89. American Mathematical Soc., 2001.

[26] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[27] V. Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36(2):423–439, 1965.

[28] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[29] C. Villani. *Optimal Transport, Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.