Probability Theory and Related Fields (2022) 183:359–413
https://doi.org/10.1007/s00440-021-01103-y

Approximation of martingale couplings on the line in the adapted weak topology

M. Beiglböck[1] · B. Jourdain[2] · W. Margheriti[2] · G. Pammer[3]

Received: 6 February 2021 / Revised: 23 November 2021 / Accepted: 9 December 2021 /
Published online: 2 February 2022

Abstract

Our main result is to establish stability of martingale couplings: suppose that π is a martingale coupling with marginals μ, ν. Then, given approximating marginal measures $\tilde{\mu} \approx \mu, \tilde{\nu} \approx \nu$ in convex order, we show that there exists an approximating martingale coupling $\tilde{\pi} \approx \pi$ with marginals $\tilde{\mu}, \tilde{\nu}$. In mathematical finance, prices of European call/put option yield information on the marginal measures of the arbitrage free pricing measures. The above result asserts that small variations of call/put prices lead only to small variations on the level of arbitrage free pricing measures. While these facts have been anticipated for some time, the actual proof requires somewhat intricate stability results for the adapted Wasserstein distance. Notably the result has consequences for several related problems. Specifically, it is relevant for numerical approximations, it leads to a new proof of the monotonicity principle of martingale optimal transport and it implies stability of weak martingale optimal transport as well as optimal Skorokhod embedding. On the mathematical finance side this yields continuity of the robust pricing problem for exotic options and VIX options with respect to market data. These applications will be detailed in two companion papers.

W. Margheriti acknowledges support from the 'Chaire Risques Financiers', Fondation du Risque. G. Pammer acknowledges support from the Austrian Science Fund (FWF) through Grant Number W1245.

✉ G. Pammer
gudmund.pammer@math.ethz.ch

M. Beiglböck
mathias.beiglboeck@univie.ac.at

B. Jourdain
benjamin.jourdain@enpc.fr

W. Margheriti
william.margheriti@enpc.fr

1 University of Vienna, Vienna, Austria

2 CERMICS, Ecole des Ponts, INRIA, Marne-la-Vallée, Paris, France

3 ETH Zürich, Zürich, Switzerland

Keywords Martingale optimal transport · Adapted Wasserstein distance · Stability

1 Introduction

Before carefully explaining all required notation and describing relevant literature, let us give a first description of our main result and its relevance for the martingale transport theory.

While classical transport theory is concerned with the set $\Pi(\mu, \nu)$ of *couplings* or *transport plans* of probability measures μ, ν, the martingale variant restricts the problem to the set $\Pi_M(\mu, \nu)$ of *martingale couplings*, that is, transport plans which preserve the barycenter of each particle. Even though the main interest lies in the case where μ, ν are probabilities on the real line, many of the basic arguments and results appear significantly more involved in the martingale context. A basic explanation lies in the rigidity of the martingale condition that makes classically simple approximation results quite intricate. Specifically, the martingale theory has been missing a counterpart to the following straightforward fact of the classical transport theory:

Fact 1.1 (Stability of couplings) Let $\pi \in \Pi(\mu, \nu)$ and assume that $\mu^k, \nu^k, k \in \mathbb{N}$, are probabilities that converge weakly to μ and ν. Then there exist couplings $\pi^k \in \Pi(\mu^k, \nu^k), k \in \mathbb{N}$ converging weakly to π.

This result is so basic and straightforward that its implicit use is easily overlooked. Note however that it plays a crucial role in a number of occasions, e.g. for stability of optimal transport, providing numerical approximations, or in the characterisation of optimality through cyclical monotonicity.

The main result of this article is to establish Fact 1.1 for martingale transports on the real line, see Theorem 2.6 below. This closes a gap in the theory of martingale transport and yields basic fundamental results in a unified fashion that is much closer to the classical theory. It allows to address questions in martingale optimal transport, optimal Skorokhod embedding and robust finance that have previously remained open. These applications are considered systematically in two accompanying articles, see [12] for the first of the two. Among other results, we establish therein the stability of the superreplication bound for VIX futures as well as the stability of the stretched Brownian motion. Moreover, we derive sufficiency of a monotonicity principle, in the spirit of cyclical monotonicity of classical optimal transport, for the weak martingale optimal transport problem and are able to generalize the results concerning the corresponding notion of monotonicity in martingale optimal transport.

We note that while virtually all (to the best of our knowledge) applications of martingale optimal transport are concerned with the case where μ, ν are supported on $\mathbb{R}$, it is a highly intriguing challenge to extend the martingale transport theory to the case where μ, ν are supported on $\mathbb{R}^d, d > 1$. In a remarkable contrast to our main result, stability of martingale optimal transport breaks down in higher dimensions as has been recently established by Brückerhoff and Juillet [17].

1.1 The martingale optimal transport problem

Let (X, d_X), (Y, d_Y) be Polish spaces and $C : X \times Y \to \mathbb{R}_+$ be a nonnegative measurable function. Denote by $\mathcal{P}(X)$ the set of probability measures on X. For $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, the classical Optimal Transport problem consists in minimising

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{X\times Y} C(x, y)\, \pi(dx, dy), \tag{OT}$$

where $\Pi(\mu, \nu)$ denotes the set of probability measures in $\mathcal{P}(X \times Y)$ with the first marginal μ and the second marginal ν. When $X = Y$ and $C = d_X^r$ for some $r \geq 1$, (OT) corresponds to the well-known Wasserstein distance with index r to the power r, denoted $\mathcal{W}_r^r(\mu, \nu)$, see [4,51,53,54] for a study in depth.

The theory of OT goes back to Monge [44] in its original formulation and Kantorovich [39] in its modern formulation. It was rediscovered many times under various forms and has an impressive scope of applications. A variant of OT that is motivated by applications in mathematical finance, in particular in model-independent pricing, was introduced in [11] in a discrete time setting and in [26] in a continuous time setting. Compared to the usual OT, the difference is that one requires an additional martingale constraint to (OT) which reflects the condition for a financial market to be free of arbitrage.

In detail, the Martingale Optimal Transport (MOT) problem is formulated as follows: given $\pi \in \mathcal{P}(\mathbb{R}\times\mathbb{R})$, we denote by $(\pi_x)_{x\in X}$ a regular conditional disintegration with respect to its first marginal μ. We then 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. Let $C : \mathbb{R}\times\mathbb{R} \to \mathbb{R}_+$ be a nonnegative measurable function and μ, ν be two probability distributions on the real line with finite first moment. Then the MOT problem consists in minimising

$$\inf_{\pi \in \Pi_M(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} C(x, y)\, \pi(dx, dy), \tag{MOT}$$

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 [52], the existence of a martingale coupling between two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with finite first moment is equivalent to $\mu \leq_c \nu$, where $\leq_c$ denotes the convex order. We recall that two finite positive measures μ, ν on $\mathbb{R}$ with finite first moment and are said to be in the convex order if and only if we have

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

for every convex function $f : \mathbb{R} \to (-\infty, \infty]$. Note that there holds equality for all affine functions, from which we deduce that μ and ν have equal masses and satisfy $\int_{\mathbb{R}} x\, \mu(dx) = \int_{\mathbb{R}} y\, \nu(dy)$.

For adaptations of classical optimal transport theory to the MOT problem, we refer to [33,34,36]. Concerning duality results, we refer to [15,18,21,23]. We also refer to [20,22,27,46] for the multi-dimensional case and to [10,14] for connections to Skorokhod embedding problem.

Concerning the numerical resolution of the MOT problem, we refer to the articles [2,3,19,30,32]. When μ and ν are finitely supported, then the MOT problem amounts to linear programming. In the general case, once the MOT problem is discretised by approximating μ and ν by probability measures with finite support and in the convex order, Alfonsi, Corbetta and Jourdain [3] raised the question of the convergence of optimal costs of the discretised problem towards the costs of the original problem. A first partial result was obtained by Juillet [38] who established stability of left-curtain coupling. Guo and Obłój [30] establish the result under moment conditions. More recently, [9,55] independently gave a definite positive answer.

1.2 The adapted Wasserstein distance

The stability result shown in [9] involves Wasserstein convergence. More precisely, let $\mu^k, \nu^k \in \mathcal{P}(\mathbb{R})$, $k \in \mathbb{N}$ be in the convex order and respectively converge to μ and ν in $\mathcal{W}_r$. Under mild assumption, for all $k \in \mathbb{N}$ there exists $\pi^k \in \Pi_M(\mu^k, \nu^k)$, optimal for (MOT), and any accumulation point of $(\pi^k)_{k\in\mathbb{N}}$ with respect to the $\mathcal{W}_r$-convergence is a martingale coupling between μ and ν optimal for (MOT).

However, it turns out that the usual weak topology / Wasserstein distance is not well suited in setups where accumulation of information plays a distinct role, e.g. in mathematical finance. Indeed, the symmetry of this distance does not take into account the temporal structure of stochastic processes. It is easy to convince oneself that two stochastic processes very close in Wasserstein distance can yield radically unalike information, as illustrated in [5, Figure 1]. Therefore, one needs to strengthen, the usual topology of weak convergence accordingly. Over time numerous researchers have independently introduced refinements of the weak topology, we mention Hellwig's information topology [31], Aldous's extended weak topology [1], the nested distance / adapted Wasserstein distance of Plug-Pichler [47] and the optimal stopping topology [6]. Strikingly, all those seemingly different definitions lead to same topology in the present discrete time [6, Theorem 1.1] framework. We refer to this topology as the *adapted weak topology*. A natural compatible metric is given by the *adapted Wasserstein distance*, see [16,43,47–50] among others.

Fix $x_0 \in X$ and $r \geq 1$. We denote the set of all probability measures on X with finite r-th moment by $\mathcal{P}_r(X)$, i.e.

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

Let $\mathcal{M}(X)$ (resp. $\mathcal{M}_r(X)$) denote the set of all finite positive measures (resp. with finite r-th moment). The sets $\mathcal{M}(X)$ and $\mathcal{M}_r(X)$, resp. are equipped with the weak topology induced by the set $C_b(X)$ of all real-valued absolutely bounded continuous functions on X and, resp., the set $\Phi_r(X)$ of all real-valued continuous functions on X, $C(X)$, which satisfy the growth constraint

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

A sequence $(\mu^k)_{k\in\mathbb{N}}$ converges in $\mathcal{M}_r(X)$ to μ if and only if

$$\forall f \in \Phi_r(X), \quad \mu^k(f) \underset{k\to+\infty}{\longrightarrow} \mu(f). \tag{1.1}$$

If moreover μ and μ^k, $k \in \mathbb{N}$, have equal masses, then the convergence (1.1) can be equivalently formulated (see for instance [54, Theorem 6.9]) in terms of the Wasserstein distance with index r:

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

Given $m_0 > 0$, we can then equip the set of finite positive measures in $\mathcal{M}_r(X\times Y)$ with mass m_0 with the Wasserstein topology. However, we can also equip it with a stronger topology, namely the adapted Wasserstein topology. It is induced by the metric $\mathcal{AW}_r$ defined for all $\pi, \pi' \in \mathcal{M}_r(X \times Y)$ such that $\pi(X \times Y) = \pi'(X \times Y) = m_0$ by

$$\mathcal{AW}_r(\pi, \pi') = \inf_{\chi\in\Pi(\mu,\mu')} \left(\int_{X\times X} \left(d_X^r(x, x') + \mathcal{W}_r^r(\pi_x, \pi'_{x'})\right) \chi(dx, dx') \right)^{\frac{1}{r}}, \tag{1.2}$$

where μ, resp. μ', is the first marginal of π, resp. π'. It is easy to check that $\mathcal{W}_r \le \mathcal{AW}_r$, and therefore $\mathcal{AW}_r$ indeed induces a stronger topology than $\mathcal{W}_r$. Another useful point of view is the following: let $J : \mathcal{M}(X \times Y) \to \mathcal{M}(X \times \mathcal{P}(Y))$ be the inclusion map defined for all $\pi = \mu \times \pi_x \in \mathcal{M}(X \times Y)$ by

$$J(\pi)(dx, dp) = \mu(dx)\, \delta_{\pi_x}(dp).$$

For all $\pi, \pi' \in \mathcal{M}_r(X \times Y)$ with equal masses, their adapted Wasserstein distance coincides with

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

It follows that the topology induced by $\mathcal{AW}_r$ coincides with the weak topology induced by J.

Finally, let us mention the interpretation of the adapted Wasserstein distance in terms of bicausal couplings (cf. [7]). Let $\pi, \pi' \in \mathcal{P}_r(X \times Y)$. Let Z_1, Z_2, Z'_1, Z'_2 be random variables such that the distribution of (Z_1, Z_2, Z'_1, Z'_2) is a $\mathcal{W}_r$-optimal coupling between π and π'. In many cases, there exists a Monge transport map $T : X \times Y \to X \times Y$ such that $(Z_1', Z_2') = T(Z_1, Z_2)$. As mentioned in [5], the temporal structure of stochastic processes is then not taken into account since the present value Z_1' is determined from the future value Z_2. Therefore, it is more suitable to restrict to couplings (Z_1, Z_2, Z_1', Z_2') between π and π' such that the conditional distribution of Z_1' (resp. Z_1) given (Z_1, Z_2) (resp. (Z_1', Z_2')) is equal to the conditional distribution of Z_1' (resp. Z_1) given Z_1 (resp. Z_1').

Let μ and μ' denote the respective first marginal distributions of π and π' and let $\eta \in \Pi(\pi, \pi')$ be a coupling between π and π'. Let $\chi(dx, dx') = \int_{(y,y')\in Y\times Y} \eta(dx, dy, dx', dy') \in \Pi(\mu, \mu')$. We write $\chi(dx, dx') = \mu(dx)\, \chi_x(dx') = \mu'(dx')\, \overleftarrow{\chi}_{x'}(dx)$. Then η is called bicausal if and only if

$$\int_{y'\in Y} \eta(dx, dy, dx', dy') = \pi(dx, dy)\, \chi_x(dx')$$
$$\text{and} \quad \int_{y\in Y} \eta(dx, dy, dx', dy') = \pi'(dx', dy')\, \overleftarrow{\chi}_{x'}(dx).$$

We denote by $\Pi_{bc}(\pi, \pi')$ the set of bicausal couplings between π and π'. Let $(\gamma_{(x,x')}(dy, dy'))_{(x,x')\in X\times X}$ be a probability kernel such that $\eta(dx, dy, dx', dy') = \chi(dx, dx')\, \gamma_{(x,x')}(dy, dy')$. Another useful characterisation is that η is bicausal if and only if $\chi(dx, dx')$-almost everywhere, $\gamma_{(x,x')}(dy, dy') \in \Pi(\pi_x, \pi_{x'})$. Then the adapted Wasserstein distance coincides with

$$\mathcal{AW}_r(\pi, \pi') = \inf_{\eta\in\Pi_{bc}(\pi,\pi')} \left(\int_{X\times Y} \left(d_X^r(x, x') + d_Y^r(y, y') \right) \eta(dx, dy, dx', dy') \right)^{\frac{1}{r}}.$$

One of the objectives of the present paper is to prove that well-known stability results for the $\mathcal{W}_r$-convergence also hold for the $\mathcal{AW}_r$-convergence. More details are given in Sect. 2.

1.3 Outline

Section 2 presents the main result of this article, Theorem 2.6. We also provide a discussion of the result and give a sketch of its proof in order to help seeing through the technical details provided later on.

In Sect. 3 we provide certain technical lemmas which allow us to deal with difficulties specific to the adapted Wasserstein distance with more ease. They mainly explore properties of approximations and when addition (in a sense explained below) is continuous.

Section 4 focuses on the convex order. It deals with potential functions which are a convenient tool to address the convex order in dimension one.

Section 5 is devoted to the proof of the main theorem. Before entering into actual argument, we establish that is is enough to prove $\mathcal{AW}_1$-convergence for irreducible pairs of marginals.

2 Main result

Our main result is Theorem 2.6 below. Before stating it, we give a proposition which enlightens us why the conclusion of the theorem should be at least hoped for. We also state a generalisation of this proposition to Polish spaces. Then, we state a proposition which is a key result to argue that the theorem needs only to be proved when the limit pair is irreducible. Next, we state the theorem together with a sketch of its proof. It is understood that (X, d_X) and (Y, d_Y) denote arbitrary Polish spaces and that (x_0, y_0) is a fixed element of $X \times Y$.

As already mentioned above, it is well-known (and easy to show) that when one considers convergent sequences of marginals $(\mu^k)_{k\in\mathbb{N}}$, $(\nu^k)_{k\in\mathbb{N}}$ (with equal masses) to $\mu, \nu \in \mathcal{M}_r(X)$, then, informally speaking, we have[1]

$$\Pi(\mu^k, \nu^k) \underset{k\to+\infty}{\longrightarrow} \Pi(\mu, \nu) \quad \text{in } \mathcal{W}_r, \tag{2.1}$$

i.e., any sequence with convergent marginals has accumulation points in $\Pi(\mu, \nu)$, and for any $\pi \in \Pi(\mu, \nu)$ it holds

$$\inf_{\pi^k\in\Pi(\mu^k,\nu^k)} \mathcal{W}_r^r(\pi, \pi^k) \le \mathcal{W}_r^r(\mu, \mu^k) + \mathcal{W}_r^r(\nu, \nu^k) \underset{k\to+\infty}{\longrightarrow} 0. \tag{2.2}$$

Indeed, if $\eta^k \in \Pi(\mu^k, \mu)$, resp. $\tau^k \in \Pi(\nu, \nu^k)$ is optimal for $\mathcal{W}_r(\mu^k, \mu)$, resp. $\mathcal{W}_r(\nu, \nu^k)$, then the measure $\eta^k(dx^k, dx)\,\pi_x(dy)\,\tau_y^k(dy^k)$ is a coupling between $\pi(dx, dy)$ and $\int_{(x,y)\in X\times Y} \eta^k(dx^k, dx)\,\pi_x(dy)\,\tau_y^k(dy^k)$ which belongs to $\Pi(\mu^k, \nu^k)$ and

$$\begin{aligned}
&\inf_{\pi^k\in\Pi(\mu^k,\nu^k)} \mathcal{W}_r^r(\pi, \pi^k)\\
&\quad\le \int_{X\times X\times Y\times Y} \left(d_X^r(x^k, x) + d_Y^r(y, y^k)\right) \eta^k(dx^k, dx)\,\pi_x(dy)\,\tau_y^k(dy^k)\\
&\quad= \mathcal{W}_r^r(\mu, \mu^k) + \mathcal{W}_r^r(\nu, \nu^k).
\end{aligned}$$

The next two propositions establish (2.1) with respect to $\mathcal{AW}_r$ for finite positive measures with common mass. The first one is formulated for $X = Y = \mathbb{R}$ and provides under mild assumptions an estimate of $\inf_{\pi^k\in\Pi(\mu^k,\nu^k)} \mathcal{AW}_r^r(\pi, \pi^k)$ with respect to the marginals as in (2.2). Its proof relies on unidimensional tools, which we recall here. For η a probability distribution on $\mathbb{R}$, we denote by $F_\eta : x \mapsto \eta((-\infty, x])$ its cumulative distribution function, and by $F_\eta^{-1} : (0, 1) \to \mathbb{R}$ its quantile function defined for all $u \in (0, 1)$ by

$$F_\eta^{-1}(u) = \inf\{x \in \mathbb{R} \mid F_\eta(x) \ge u\}.$$

The following properties are standard results (see for instance [37, Section 6] for proofs):

[1] Note that this can be made precise in terms of hemicontinuity. We also refer to [45].

(a) F_η is càdlàg i.e. right-continuous with left-hand limits, F_η^{-1} is càglàd i.e. left-continuous with right-hand limits;
(b) For all $(x, u) \in \mathbb{R} \times (0, 1)$,

$$F_\eta^{-1}(u) \leq x \iff u \leq F_\eta(x), \tag{2.3}$$

which implies, using the notation $F_\eta(y-)$ for the left-hand limit of F_η at $y \in \mathbb{R}$,

$$F_\eta(x-) < u \leq F_\eta(x) \implies x = F_\eta^{-1}(u), \tag{2.4}$$

$$\text{and} \quad F_\eta(F_\eta^{-1}(u)-) \leq u \leq F_\eta(F_\eta^{-1}(u)); \tag{2.5}$$

(c) For $\eta(dx)$-almost every $x \in \mathbb{R}$,

$$0 < F_\eta(x), \quad F_\eta(x-) < 1 \quad \text{and} \quad F_\eta^{-1}(F_\eta(x)) = x; \tag{2.6}$$

(d) The image of the Lebesgue measure on $(0, 1)$ by F_η^{-1} is η.

The property (d) is referred to as the inverse transform sampling.

Proposition 2.1 *Let $\mu, \mu^k, \nu, \nu^k \in \mathcal{M}_r(\mathbb{R})$, $k \in \mathbb{N}$, be measures of equal masses such that μ^k (resp. ν^k) converges to μ (resp. ν) in $\mathcal{W}_r$. Let $\pi \in \Pi(\mu, \nu)$. Then:*

(a) *There exists a sequence $\pi^k \in \Pi(\mu^k, \nu^k)$, $k \in \mathbb{N}$, converging to π in $\mathcal{AW}_r$;*
(b) *If for all $x \in \mathbb{R}$ and $k \in \mathbb{N}$ with $\mu^k(\{x\}) > 0$, there exists $x' \in \mathbb{R}$ such that*

$$\mu((-\infty, x')) \leq \mu^k((-\infty, x)) < \mu^k((-\infty, x]) \leq \mu((-\infty, x'])$$

(which is for instance always satisfied for μ^k non-atomic) then

$$\mathcal{AW}_r^r(\pi, \pi^k) \leq \mathcal{W}_r^r(\mu, \mu^k) + \mathcal{W}_r^r(\nu, \nu^k). \tag{2.7}$$

Remark 2.2 If π is a martingale coupling, i.e. $\int_\mathbb{R} y'\, \pi_{x'}(dy') = x'$, $\mu(dx')$-almost everywhere, then for $\chi^k \in \Pi(\mu^k, \mu)$ an optimal coupling for $\mathcal{AW}_r(\pi^k, \pi)$, we have

$$\begin{aligned}
&\int_\mathbb{R} \left| x - \int_\mathbb{R} y\, \pi_x^k(dy) \right|^r \mu^k(dx) \\
&= \int_{\mathbb{R}\times\mathbb{R}} \left| x - \int_\mathbb{R} y\, \pi_x^k(dy) \right|^r \chi^k(dx, dx') \\
&\leq 2^{r-1} \int_{\mathbb{R}\times\mathbb{R}} \left(|x - x'|^r + \left| x' - \int_\mathbb{R} y\, \pi_x^k(dy) \right|^r \right) \chi^k(dx, dx') \\
&= 2^{r-1} \int_{\mathbb{R}\times\mathbb{R}} \left(|x - x'|^r + \left| \int_\mathbb{R} y'\, \pi_{x'}(dy') - \int_\mathbb{R} y\, \pi_x^k(dy) \right|^r \right) \chi^k(dx, dx') \\
&\leq 2^{r-1} \int_{\mathbb{R}\times\mathbb{R}} \left(|x - x'|^r + \mathcal{W}_1^r(\pi_x^k, \pi_{x'}) \right) \chi^k(dx, dx') \\
&\leq 2^{r-1} \mathcal{AW}_r^r(\pi, \pi^k) \underset{k\to+\infty}{\longrightarrow} 0.
\end{aligned}$$

In that sense, π^k, $k \in \mathbb{N}$ is almost a sequence of martingale couplings.

In the setting of Proposition 2.1 and Remark 2.2, if μ^k and ν^k are also in the convex order and π is a martingale coupling, then in view of Remark 2.2 one would naturally expect that π^k can be slightly modified into a martingale coupling and still converge to π in $\mathcal{AW}_r$. This actually requires a considerable amount of work and is the main message of Theorem 2.6 below. We mention that the previous proposition generalises to arbitrary Polish spaces X and Y, as the next proposition states, but unfortunately without providing an estimate.

Proposition 2.3 *Let $\mu, \mu^k \in \mathcal{M}_r(X)$, $\nu, \nu^k \in \mathcal{M}_r(Y)$, $k \in \mathbb{N}$, all with equal masses and such that μ^k (resp. ν^k) converges to μ (resp. ν) in $\mathcal{W}_r$. Let $\pi \in \Pi(\mu, \nu)$. Then there exists a sequence $\pi^k \in \Pi(\mu^k, \nu^k)$, $k \in \mathbb{N}$, converging to π in $\mathcal{AW}_r$.*

The next proposition is a key ingredient which allows us to reduce the proof of Theorem 2.6 below to the case of irreducible pairs of marginals. For $\mu \in \mathcal{M}_1(\mathbb{R})$, we denote by u_μ its potential function, that is the map defined for all $y \in \mathbb{R}$ by $u_\mu(y) = \int_{\mathbb{R}} |y - x| \, \mu(dx)$ (see Sect. 4 for more details). We recall that a pair (μ, ν) of finite positive measures in convex order is called irreducible if $I = \{u_\mu < u_\nu\}$ is an interval and then, $\mu(I) = \mu(\mathbb{R})$ and $\nu(\overline{I}) = \nu(\mathbb{R})$.

Remark 2.4 If (μ, ν) is an irreducible pair of non-zero measures in the convex order and $a \in \mathbb{R}$ is such that $\nu([a, +\infty)) = 0$, then the convex order implies $\mu([a, +\infty)) = 0$, hence

$$u_\mu(a) = a - \int_{\mathbb{R}} x \, \mu(dx) = a - \int_{\mathbb{R}} y \, \nu(dy) = u_\nu(a),$$

so $a \notin I$. Similarly, $\nu((-\infty, a]) = 0 \implies a \notin I$. We deduce that ν must assign positive mass to any neighbourhood of each of the boundaries of I.

According to [13, Theorem A.4], for any pair (μ, ν) of probability measures in convex order, there exist $N \subset \mathbb{N}$ and a sequence $(\mu_n, \nu_n)_{n \in N}$ of irreducible pairs of sub-probability measures in convex order such that

$$\mu = \eta + \sum_{n \in N} \mu_n, \quad \nu = \eta + \sum_{n \in \mathbb{N}} \nu_n \quad \text{and} \quad \{u_\mu < u_\nu\} = \bigcup_{n \in N} \{u_{\mu_n} < u_{\nu_n}\},$$

where the union is disjoint and $\eta = \mu|_{\{u_\mu = u_\nu\}}$. The sequence $(\mu_n, \nu_n)_{n \in N}$ is unique up to rearrangement of the pairs and is called the decomposition of (μ, ν) into irreducible components. Moreover, for any martingale coupling $\pi \in \Pi_M(\mu, \nu)$, there exists a unique sequence of martingale couplings $\pi_n \in \Pi_M(\mu_n, \nu_n)$, $n \in N$ such that

$$\pi = \chi + \sum_{n \in N} \pi_n,$$

where $\chi = (\mathrm{id}, \mathrm{id})_* \eta$ and $*$ denotes the pushforward operation. This sequence satisfies

$$\forall n \in N, \quad \pi_n(dx, dy) = \mu_n(dx) \, \pi_x(dy). \tag{2.8}$$

Proposition 2.5 *Let $(\mu^k,\nu^k)_{k\in\mathbb{N}}$ be a sequence of pairs of probability measures on the real line in convex order which converge to (μ,ν) in $\mathcal{W}_1$. Let $(\mu_n,\nu_n)_{n\in N}$ be the decomposition of (μ,ν) into irreducible components and $\eta=\mu|_{\{u_\mu=u_\nu\}}$. Then there exists for any $k\in\mathbb{N}$ a decomposition of (μ^k,ν^k) into pairs of sub-probability measures $(\mu_n^k,\nu_n^k)_{n\in N}$, (η^k,υ^k) which are in convex order such that*

$$\eta^k+\sum_{n\in N}\mu_n^k=\mu^k,\quad \upsilon^k+\sum_{n\in N}\nu_n^k=\nu^k,\quad k\in\mathbb{N},\tag{2.9}$$

$$\lim_{k\to+\infty}\eta^k=\eta,\quad \lim_{k\to+\infty}\mu_n^k=\mu_n,\quad \lim_{k\to+\infty}\nu_n^k=\nu_n,\quad \lim_{k\to+\infty}\upsilon^k=\eta\quad \textit{in } \mathcal{W}_1.\tag{2.10}$$

We can now state our main result, namely Theorem 2.6 below. Any martingale coupling whose marginals are approximated by probability measures in convex order can be approximated by martingale couplings with respect to the adapted Wasserstein distance.

Theorem 2.6 *Let $\mu^k,\nu^k\in\mathcal{P}_r(\mathbb{R})$, $k\in\mathbb{N}$, be in convex order and respectively converge to μ and ν in $\mathcal{W}_r$. 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 $\mathcal{AW}_r$.*

Proof (Sketch of the proof) We will first argue that it is enough to consider the case $r=1$. Thanks to Proposition 2.5, we can also reduce the proof to the case of irreducible pairs of marginals (μ,ν), whose single irreducible component is denoted $(\ell,\rho)=I$.

Step 1. Fix a martingale coupling $\pi\in\Pi_M(\mu,\nu)$. When directly approximating π we would face technical obstacles. First, for K a compact subset of I, $\mu|_K\times\pi_x$ is not necessarily compactly supported. Moreover, ν may put mass on the boundary of I. To overcome successively these two difficulties, the kernel π_x is first compactified to a compact set $[-R,R]$, where $R>0$ (when $|\ell|\vee|\rho|<\infty$, one may choose R equal to this maximum), and then pushed forward by the map $y\mapsto\alpha(y-x)+x$, where $\alpha\in(0,1)$. This yields a martingale coupling $\pi^{R,\alpha}$ close to π and easier to approximate, between μ and a probability measure $\nu^{R,\alpha}$ dominated by ν in the convex order. We find compact sets $K,L\subset I$ such that the restriction $\pi^{R,\alpha}|_{K\times\mathbb{R}}$ is compactly supported on $K\times L$ and concentrated on $K\times\mathring{L}$, where $\mathring{L}$ denotes the interior of L. Since, by irreducibility, ν puts mass onto any neighbourhood of the boundary of I, $\nu^{R,\alpha}$ assigns positive mass to two open sets L_-, L_+ on both sides of K with positive distance to K. This is summarised in Fig. 1, where J denotes a compact subset of I that is large enough.

Step 2. It is possible to find an approximating sequence $(\hat\pi^k=\hat\mu^k\times\hat\pi_x^k)_{k\in\mathbb{N}}$ for the sub-probability martingale coupling $\pi^{R,\alpha}|_{K\times\mathbb{R}}$ from step 1. Unfortunately $\hat\pi^k$ is not necessarily a martingale coupling. Therefore, we free up some mass, and use the one available on the left and right of K in L_- and L_+ to adjust the barycenters of the kernels π_x^k. Hence we find a sequence $(\tilde\pi^k=\hat\mu^k\times\tilde\pi_x^k)_{k\in\mathbb{N}}$ of sub-probability martingale couplings approximating $\pi^{R,\alpha}|_{K\times\mathbb{R}}$.

Step 3. By construction, up to multiplication by a factor smaller than and close to 1, the first marginal of $\tilde\pi^k$ satisfies $\hat\mu^k\le\mu^k$. Moreover, its second marginal denoted $\tilde\nu^k$ is

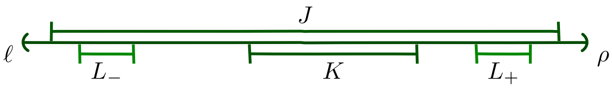

Fig. 1 Intervals involved in the proof. The boundaries of the closed intervals are vertical bars and those of the open intervals are parentheses

such that there exists a probability measure $\nu^{R,\alpha,k}$ which satisfies $\tilde{\nu}^k \leq \nu^{R,\alpha,k} \leq_c \nu^k$. Then by using the uniform convergence of potential functions, we show that for k sufficiently large there exist sub-probability martingale couplings $\eta^k \in \Pi_M(\mu^k - \hat{\mu}^k, \nu^{R,\alpha,k} - \tilde{\nu}^k)$ so that the sum $\eta^k + \tilde{\pi}^k$ is a martingale coupling in $\Pi_M(\mu^k, \nu^{R,\alpha,k})$, where the second marginal is dominated by ν^k in the convex order.

Step 4. In the last step, we use the inverse-transform martingale coupling between $\nu^{R,\alpha,k}$ and ν^k, see [37], to change $\eta^k + \tilde{\pi}^k$ to a martingale coupling $\pi^k \in \Pi(\mu^k, \nu^k)$. Finally, we estimate the $\mathcal{AW}_1$-distance of π to π^k. □

3 On the adapted weak topology

We begin this section with a lemma on uniform integrability which will prove very handy throughout the paper. We formulate it for finite positive measures on X, but it is understood that (X, x_0) is replaced with (Y, y_0) for measures on Y.

Lemma 3.1 *Let $r \geq 1$ and $\mu \in \mathcal{M}_r(X)$. For $\epsilon > 0$, let*

$$I_\epsilon^r(\mu) := \sup_{\substack{\tau \in \mathcal{M}(X) \\ \tau \leq \mu,\ \tau(X) \leq \epsilon}} \int_X d_X^r(x, x_0)\, \tau(dx). \tag{3.1}$$

(a) *I_ϵ^r is monotone in μ, i.e., $\mu \leq \mu' \in \mathcal{M}_r(X)$ implies that $I_\epsilon^r(\mu) \leq I_\epsilon^r(\mu')$.*
(b) *The value of $I_\epsilon^r(\mu)$ vanishes as $\epsilon \to 0$.*
(c) *For any $\mu' \in \mathcal{M}_r(X)$ such that $\mu(X) = \mu'(X)$ we have*

$$I_\epsilon^r(\mu) \leq 2^{r-1}\left(I_\epsilon^r(\mu') + \mathcal{W}_r^r(\mu, \mu')\right). \tag{3.2}$$

(d) *Let $\mu, \mu^k \in \mathcal{M}_r(X)$, $k \in \mathbb{N}$ be with equal masses such that μ^k converges weakly to μ. Then*

$$\mathcal{W}_r(\mu^k, \mu) \underset{k \to +\infty}{\longrightarrow} 0 \iff \sup_{k \in \mathbb{N}} I_\varepsilon^r(\mu^k) \underset{\varepsilon \to 0}{\longrightarrow} 0 \quad \textit{and} \quad \sup_{k \in \mathbb{N}} \int_X d_X^r(x, x_0)\, \mu^k(dx) < +\infty.$$

(e) *Finally, if $X = \mathbb{R}^d$ and $\mu \leq_c \nu$ with $\nu \in \mathcal{M}_1(\mathbb{R}^d)$, then $I_\epsilon^1(\mu) \leq I_\epsilon^1(\nu)$.*

Remark 3.2 If $\mu(X) \leq \epsilon$, then $I_\epsilon^r(\mu)$ is simply the r-th moment of μ.

Proof The first point (a) is an easy consequence of the definition of I_ϵ^r.

Next we check (b). Let $\mu \in \mathcal{M}_r(X)$ be such that $\mu(X) > 0$. Since

$$I^r_\varepsilon(\mu) = \mu(X) I^r_{\frac{\varepsilon}{\mu(X)}}\left(\frac{\mu}{\mu(X)}\right), \tag{3.3}$$

to check convergence of $I^r_\varepsilon(\mu)$ to 0 as $\varepsilon \to 0$, we may suppose that $\mu \in \mathcal{P}_r(X)$. Let $\varepsilon \in (0, 1)$. For $\eta \in \mathcal{M}_r(X)$, we denote by $\overline{\eta}$ the image of η under the map $x \mapsto d^r_X(x, x_0)$. Let $\tau \in \mathcal{M}(X)$ be such that $\tau \leq \mu$ and $0 < \tau(X) \leq \varepsilon$. Since $\tau \leq \mu$, we have $\overline{\tau} \leq \overline{\mu}$. Using (2.5) for the last inequality, we get for all $u \in (0, 1)$

$$\begin{aligned} 1 - F_{\overline{\tau}/\tau(X)}\left(F^{-1}_{\overline{\mu}}(1 - \tau(X)u)\right) &= \frac{\overline{\tau}((F^{-1}_{\overline{\mu}}(1 - \tau(X)u), +\infty))}{\tau(X)} \\ &\leq \frac{\overline{\mu}((F^{-1}_{\overline{\mu}}(1 - \tau(X)u), +\infty))}{\tau(X)} \leq u, \end{aligned}$$

hence $F_{\overline{\tau}/\tau(X)}(F^{-1}_{\overline{\mu}}(1 - \tau(X)u)) \geq 1 - u$ and by (2.3), $F^{-1}_{\overline{\tau}/\tau(X)}(1 - u) \leq F^{-1}_{\overline{\mu}}(1 - \tau(X)u)$. Using the inverse transform sampling, we deduce

$$\begin{aligned} &\int_X d^r_X(x, x_0)\, \tau(dx) \\ &= \tau(X) \int_0^1 F^{-1}_{\overline{\tau}/\tau(X)}(1 - u)\, du \\ &\leq \tau(X) \int_0^1 F^{-1}_{\overline{\mu}}(1 - \tau(X)u)\, du = \int_{1-\tau(X)}^1 F^{-1}_{\overline{\mu}}(u)\, du \leq \int_{1-\varepsilon}^1 F^{-1}_{\overline{\mu}}(u)\, du. \end{aligned} \tag{3.4}$$

Hence $I^r_\varepsilon(\mu) \leq \int_{1-\varepsilon}^1 F^{-1}_{\overline{\mu}}(u)\, du$ where the right-hand side vanishes as $\varepsilon \to 0$ since, as $\mu \in \mathcal{P}_r(X)$, $\int_0^1 F^{-1}_{\overline{\mu}}(u)\, du = \int_X d_X(x, x_0)^r\, \mu(dx) < +\infty$. Let us check the equality

$$I^r_\varepsilon(\mu) = \int_{1-\varepsilon}^1 F^{-1}_{\overline{\mu}}(u)\, du, \tag{3.5}$$

that will come in handy for the proof of claim (e) by setting

$$\tau^*(dx) = \left(\mathbb{1}_{A\varepsilon}(x) + \frac{F_{\overline{\mu}}(y_\varepsilon) - (1 - \varepsilon)}{\mu(B_\varepsilon)} \mathbb{1}_{B\varepsilon}(x)\right) \mu(dx), \tag{3.6}$$

where

$$y_\varepsilon = F^{-1}_{\overline{\mu}}(1 - \varepsilon), \quad A_\varepsilon = \{x \in \mathbb{R} \mid d^r_X(x, x_0) > y_\varepsilon\} \quad \text{and} \quad B_\varepsilon = \{x \in \mathbb{R} \mid d^r_X(x, x_0) = y_\varepsilon\},$$

and the second summand of the right-hand side in (3.6) is taken to be zero if $\mu(B_\varepsilon) = 0$. Since $A_\varepsilon \cap B_\varepsilon = \emptyset$ and, by (2.5),

$$\mu(B_\varepsilon) = \overline{\mu}(\{y_\varepsilon\}) = F_{\overline{\mu}}(y_\varepsilon) - F_{\overline{\mu}}(F_{\overline{\mu}}^{-1}(1-\varepsilon)-) \geq F_{\overline{\mu}}(y_\varepsilon) - (1-\varepsilon),$$

hence $\tau^* \leq \mu$. Moreover, $\overline{\tau^*}$ is the measure dominated by $\overline{\mu}$ with mass equal to ε which is the largest in stochastic order. Indeed, one easily checks that

$$\overline{\tau^*}(dy) = \mathbb{1}_{y>y_\varepsilon}\overline{\mu}(dy) + \left(F_{\overline{\mu}}(y_\varepsilon) - (1-\varepsilon)\right)\delta_{y_\varepsilon}(dy) \text{ so that } \overline{\tau^*}(\mathbb{R}) = \varepsilon,$$
$$\forall y \in \mathbb{R},\ F_{\overline{\tau^*}/\varepsilon}(y) = \mathbb{1}_{y\geq y_\varepsilon}\frac{F_{\overline{\mu}}(y) - (1-\varepsilon)}{\varepsilon} \text{ and } \forall u \in (0,1),\ F_{\overline{\tau^*}/\varepsilon}^{-1}(1-u) = F_{\overline{\mu}}^{-1}(1-\varepsilon u).$$

With the inverse transform sampling, the latter equality implies that

$$\int_X d_X^r(x, x_0)\,\tau^*(dx) = \varepsilon\int_0^1 F_{\overline{\tau^*}/\varepsilon}^{-1}(u)du = \varepsilon\int_0^1 F_{\overline{\mu}}^{-1}(1-\varepsilon u)du = \int_{1-\varepsilon}^1 F_{\overline{\mu}}^{-1}(u)\,du$$

so that (3.5) holds.

To see (c), fix $\mu' \in \mathcal{M}(X)$ with $\mu(X) = \mu'(X)$. We denote by $\pi(dx, dx') = \mu(dx)\,\pi_x(dx') \in \Pi(\mu, \mu')$ a $\mathcal{W}_r$-optimal coupling. Let $\tau \in \mathcal{M}(X)$ be such that $\tau \leq \mu$ and $\tau(X) \leq \epsilon$. Let $\tau' \in \mathcal{M}(X)$ be defined by

$$\tau'(dx') = \int_{x\in X} \pi_x(dx')\,\tau(dx).$$

Since π is element of $\Pi(\mu, \mu')$, we find $\tau' \leq \mu'$ and $\tau(X) = \tau'(X)$. Then

$$\begin{aligned}\int_X d_X^r(x, x_0)\,\tau(dx) &\leq 2^{r-1}\int_{X\times X}\left(d_X^r(x', x_0) + d_X^r(x, x')\right)\pi_x(dx')\,\tau(dx)\\ &\leq 2^{r-1}\left(I_\epsilon^r(\mu') + \int_{X\times X} d_X^r(x, x')\,\pi(dx, dx')\right),\end{aligned}$$

which shows by optimality of π the assertion.

We now show (d). Let $\mu, \mu^k \in \mathcal{M}_r(X)$ be with equal masses such that μ^k converges weakly to μ. According to (3.3), we may suppose that $\mu, \mu^k \in \mathcal{P}_r(X)$.

Suppose that $\mathcal{W}_r(\mu^k, \mu)$ vanishes as k goes to $+\infty$. Then the sequence of the r-th moments of μ^k, $k \in \mathbb{N}$ is bounded since it converges to the r-th moment of μ. Let $\eta > 0$. Let $k_0 \in \mathbb{N}$ be such that for all $k > k_0$, $\mathcal{W}_r^r(\mu^k, \mu) < \eta$. Then (c) yields for $\varepsilon > 0$

$$\sup_{k\in\mathbb{N}} I_\varepsilon^r(\mu^k) \leq \sum_{k\leq k_0} I_\varepsilon^r(\mu^k) + \sup_{k>k_0} I_\varepsilon^r(\mu^k) \leq \sum_{k\leq k_0} I_\varepsilon^r(\mu^k) + 2^{r-1}(I_\varepsilon^r(\mu) + \eta).$$

According to (b) we then get

$$\limsup_{\varepsilon\to 0}\sup_{k\in\mathbb{N}} I_\varepsilon^r(\mu^k) \leq 2^{r-1}\eta.$$

Since $\eta > 0$ is arbitrary, we deduce that $\sup_{k\in\mathbb{N}} I_\varepsilon^r(\mu^k)$ vanishes with ε.

Conversely, suppose that $\sup_{k\in\mathbb{N}} I_\varepsilon^r(\mu^k)$ vanishes with ε and the sequence of the r-th moments of μ^k, $k \in \mathbb{N}$ is bounded. By Skorokhod's representation theorem, there exist random variables X and X^k, $k \in \mathbb{N}$, defined on a common probability space such that X, resp. X^k is distributed according to μ, resp. μ^k and X^k converges almost surely to X. Then for all $M > 0$,

$$\begin{aligned}\mathcal{W}_r^r(\mu^k,\mu) &\le \mathbb{E}[d_X^r(X^k,X)]\\ &= \mathbb{E}[d_X^r(X^k,X)\mathbb{1}_{\{d_X^r(X^k,X)<M\}}] + \mathbb{E}[d_X^r(X^k,X)\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}].\end{aligned}$$

By the dominated convergence theorem, we deduce

$$\limsup_{k\to+\infty} \mathcal{W}_r^r(\mu^k,\mu) \le \limsup_{k\to+\infty} \mathbb{E}[d_X^r(X^k,X)\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}].$$

Let us then prove that the right-hand side vanishes as M goes to $+\infty$. Let $\eta > 0$. Let $\varepsilon > 0$ be such that $I_\varepsilon^r(\mu) + \sup_{k\in\mathbb{N}} I_\varepsilon^r(\mu^k) < \eta$. By Markov's inequality, we have

$$\sup_{k\in\mathbb{N}} \mathbb{E}[\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}] \le \sup_{k\in\mathbb{N}} \frac{\mathbb{E}[d_X^r(X^k,X)]}{M} \le \frac{2^{r-1}}{M} \sup_{k\in\mathbb{N}} \int_X d_X^r(x,x_0)\,(\mu^k+\mu)(dx),$$

where the right-hand side vanishes as M goes to $+\infty$. Therefore, there exists $M_0 > 0$ such that for all $k \in \mathbb{N}$ and $M > M_0$,

$$\begin{aligned}&\mathbb{E}[d_X^r(X^k,X)\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}]\\ &\quad\le 2^{r-1}\left(\mathbb{E}[d_X^r(X^k,x_0)\,\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}] + \mathbb{E}[d_X^r(x_0,X)\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}]\right)\\ &\quad\le 2^{r-1}\left(I_\varepsilon^r(\mu^k) + I_\varepsilon^r(\mu)\right) < 2^{r-1}\eta.\end{aligned}$$

Therefore, for all $M > M_0$,

$$\limsup_{k\to+\infty} \mathbb{E}[d_X^r(X^k,X)\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}] \le 2^{r-1}\eta.$$

Since η is arbitrary, this proves the assertion.

Finally, we want to show (e). Let $X = \mathbb{R}^d$ and $\mu \le_c \nu$ with $\nu \in \mathcal{M}_1(\mathbb{R}^d)$. According to (3.3), we may suppose that $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^d)$. Again, we write $\overline{\mu}$ and $\overline{\nu}$ for the pushforward measures of μ and ν under the map $(x \mapsto |x-x_0|^r)$. First, we note that $\overline{\mu}$ is dominated by $\overline{\nu}$ in the increasing convex order. Indeed, let $f \in C(X)$ be convex and nondecreasing, then $x \mapsto f(|x-x_0|^r)$ constitutes a convex, continuous function. Thus,

$$\int_{\mathbb{R}} f(y)\,\overline{\mu}(dy) = \int_{\mathbb{R}^d} f(|x-x_0|^r)\,\mu(dx) \le \int_{\mathbb{R}^d} f(|x-x_0|^r)\,\nu(dx) = \int_{\mathbb{R}} f(y)\,\overline{\nu}(dy).$$

The convex increasing order is characterised by the following family of inequalities (see for instance [2, Theorem 2.4]): for all $0 \le \varepsilon \le 1$,

$$\int_{1-\varepsilon}^{1} F_{\overline{\mu}}^{-1}(y)\,dy \le \int_{1-\varepsilon}^{1} F_{\overline{\nu}}^{-1}(y)\,dy.$$

The identity (3.5) concludes the proof. □

We now prove Proposition 2.1. A handy tool in the construction of the approximative sequence $(\pi^k)_{k\in\mathbb{N}}$ are copulas. Recall that a two-dimensional copula is an element C of $\Pi(\lambda,\lambda)$ where λ is the uniform distribution on $(0,1)$. A coupling π is an element of $\Pi(\mu,\nu)$ if and only if it can be written as the push-forward of a copula C under the quantile map $(F_\mu^{-1}, F_\nu^{-1})\colon (0,1)\times(0,1)\to\mathbb{R}\times\mathbb{R}$. Clearly, if C is a copula then $\pi = (F_\mu^{-1}, F_\nu^{-1})_* C$ is contained in $\Pi(\mu,\nu)$. On the other hand, if $\pi\in\Pi(\mu,\nu)$ is given, we can construct a copula C by

$$C(du,dv) = \mathbb{1}_{(0,1)}(u)\,du\,C_u(dv),$$

where C_u is given by

$$C_u = ((y,w)\mapsto F_\nu(y-) + w\nu(\{y\}))_*(\pi_{F_\mu^{-1}(u)}\times\lambda). \tag{3.7}$$

In particular, we have that $u\mapsto C_u$ is constant on the jumps on F_μ. The fact that the second marginal distribution of C is indeed uniformly distributed on $(0,1)$ is a direct consequence of the inverse transform sampling and the well-known result (see for instance [37, Lemma 6.6] for a proof) that for any $\eta\in\mathcal{P}(\mathbb{R})$,

$$((z,w)\mapsto F_\eta(z-) + w\eta(\{z\}))_*(\eta\times\lambda) = \lambda. \tag{3.8}$$

Finally, we check the identity $\pi = (F_\mu^{-1}, F_\nu^{-1})_* C$. Let $w\in(0,1]$ and continue by distinguishing two cases: On the one hand, if $\nu(\{y\})>0$ then we have by (2.4)

$$F_\nu^{-1}(F_\nu(y-) + w\nu(\{y\})) = y. \tag{3.9}$$

On the other hand, we derive from (2.6) that (3.9) holds for ν-almost every $y\in\{z\in\mathbb{R}\colon \nu(\{z\})=0\}$. Hence, we obtain for λ-almost every $u\in(0,1)$

$$\pi_{F_\mu^{-1}(u)} = (F_\nu^{-1})_* C_u \tag{3.10}$$

and conclude with $\pi = (F_\mu^{-1}, F_\nu^{-1})_* C$.

Proof (Proof of Proposition 2.1) Because of homogeneity of the $\mathcal{AW}_r$- and $\mathcal{W}_r$-distances, we can suppose w.l.o.g. that μ,μ^k,ν,ν^k and π are probability measures. Let C be the copula defined by $C(du,dv) = \mathbb{1}_{(0,1)}(u)\,du\,C_u(dv)$, where C_u is given by (3.7).

In order to define π^k, we construct associated copulas C^k where $u \mapsto C^k_u$ is constant on the jumps of F_{μ^k}. Let

$$\begin{aligned}
&\theta_k\colon \mathbb{R}\times(0,1)\to(0,1),\ (x,w)\mapsto F_{\mu^k}(x-)+w\mu^k(\{x\}),\\
&C^k_u(dv)=\int_{w=0}^1 C_{\theta_k\left(F^{-1}_{\mu^k}(u),w\right)}(dv)\,dw,\\
&\pi^k=(F^{-1}_{\mu^k},F^{-1}_{\nu^k})_*C^k=(F^{-1}_{\mu^k},F^{-1}_{\nu^k})_*(\mathbb{1}_{(0,1)}(u)\,du\,C^k_u(dv)).
\end{aligned}$$

The fact that C^k is a copula, and therefore $\pi^k\in\Pi(\mu^k,\nu^k)$, is a direct consequence of (3.8) and the inverse transform sampling. Since $u\mapsto C_u$ and $u\mapsto C^k_u$ are constant on the jumps of F_μ and F_{μ^k} respectively, reasoning like in the derivation of (3.10), we have for du-almost every u in $(0,1)$

$$\pi_{F^{-1}_\mu(u)}=(F^{-1}_\nu)_*C_u,\quad \pi^k_{F^{-1}_{\mu^k}(u)}=(F^{-1}_{\nu^k})_*C^k_u.$$

Moreover, since $(u\mapsto(F^{-1}_\mu(u),F^{-1}_{\mu^k}(u)))_*\lambda$ is a coupling between μ and μ^k, namely the comonotonous coupling, we have using the definition of $\mathcal{AW}_r(\pi,\pi^k)$ as an infimum over $\Pi(\mu,\mu^k)$, cf. (1.2),

$$\begin{aligned}
\mathcal{AW}^r_r(\pi,\pi^k)&\le\int_0^1\left(|F^{-1}_\mu(u)-F^{-1}_{\mu^k}(u)|^r+\mathcal{W}^r_r(\pi_{F^{-1}_\mu(u)},\pi^k_{F^{-1}_{\mu^k}(u)})\right)du\\
&=\mathcal{W}^r_r(\mu,\mu^k)+\int_0^1\mathcal{W}^r_r\left((F^{-1}_\nu)_*C_u,(F^{-1}_{\nu^k})_*C^k_u\right)du.
\end{aligned}\tag{3.11}$$

By Minkowski's inequality we have

$$\begin{aligned}
\left(\int_0^1\mathcal{W}^r_r\left((F^{-1}_\nu)_*C_u,(F^{-1}_{\nu^k})_*C^k_u\right)du\right)^{\frac1r}&\le\left(\int_0^1\mathcal{W}^r_r\left((F^{-1}_\nu)_*C_u,(F^{-1}_\nu)_*C^k_u\right)du\right)^{\frac1r}\\
&\quad+\left(\int_0^1\mathcal{W}^r_r\left((F^{-1}_\nu)_*C^k_u,(F^{-1}_{\nu^k})_*C^k_u\right)du\right)^{\frac1r}.
\end{aligned}\tag{3.12}$$

Since for any $\eta\in\mathcal{P}(\mathbb{R})$ the map $F^{-1}_\eta\circ F^{-1}_{C^k_u}$ is non-decreasing, we have (see for instance [3, Lemma A.3]) that for dw-almost every $w\in(0,1)$,

$$F^{-1}_\eta(F^{-1}_{C^k_u}(w))=F^{-1}_{(F^{-1}_\eta)_*C^k_u}(w).$$

Hence, we deduce

$$
\begin{aligned}
&\int_{(0,1)} \mathcal{W}_r^r\left((F_\nu^{-1})_* C_u^k, (F_{\nu^k}^{-1})_* C_u^k\right) du \\
&\quad = \int_{(0,1)} \int_{(0,1)} |F_\nu^{-1}(F_{C_u^k}^{-1}(w)) - F_{\nu^k}^{-1}(F_{C_u^k}^{-1}(w))|^r \, dw \, du \\
&\quad = \int_{(0,1)} \int_{(0,1)} |F_\nu^{-1}(v) - F_{\nu^k}^{-1}(v)|^r \, C_u^k(dv) \, du \\
&\quad = \int_{(0,1)} |F_\nu^{-1}(v) - F_{\nu^k}^{-1}(v)|^r \, dv = \mathcal{W}_r^r(\nu, \nu^k) \to 0,
\end{aligned}
\tag{3.13}
$$

where we used inverse transform sampling in the second equality. At this stage, we can already show (b) of Proposition 2.1. Indeed, the assumption made in (b) ensures that any jump of F_{μ_k} is included in a jump of F_μ. We already noted that $u \mapsto C_u$ is constant on the jumps of F_μ and therefore also constant on the jumps of F_{μ^k}. This yields for all $u, w \in (0,1)$ that $C_{\theta_k(F_{\mu^k}^{-1}(u), w)} = C_u$ and particularly $C_u^k = C_u$, which causes the first term on the right-hand side of (3.12) to vanish. Then the estimate (2.7) follows immediately from (3.11), (3.12) and (3.13).

To obtain (a) and in view of (3.11), (3.12) and (3.13), it is sufficient to show

$$
\int_0^1 \mathcal{W}_r^r\left((F_\nu^{-1})_* C_u, (F_\nu^{-1})_* C_u^k\right) du \to 0.
$$

This is achieved in two steps: First, we show for du-almost every $u \in (0,1)$ that

$$
\mathcal{W}_r((F_\nu^{-1})_* C_u, (F_\nu^{-1})_* C_u^k) \to 0. \tag{3.14}
$$

Second, we prove that

$$
u \mapsto \mathcal{W}_r^r((F_\nu^{-1})_* C_u, (F_\nu^{-1})_* C_u^k) \quad k \in \mathbb{N}, \tag{3.15}
$$

is uniformly integrable on $(0,1)$ with respect to λ.

To show (3.14), note that $\mathcal{W}_r$-convergence is already determined by a countable family $\mathcal{C} \subset \Phi_r(\mathbb{R})$ (see [25, Theorem 4.5.(b)]). For this reason, it is sufficient to show that for all $f \in \mathcal{C}$, for du-almost every $u \in (0,1)$,

$$
\int_{(0,1)} f(F_\nu^{-1}(v)) \, C_u^k(dv) \to g(u) := \int_{(0,1)} f(F_\nu^{-1}(v)) \, C_u(dv), \quad k \to +\infty,
\tag{3.16}
$$

where the integrals are du-almost everywhere well defined because of the inverse transform sampling, the fact that $f \in \Phi_r(\mathbb{R})$ and $\nu \in \mathcal{P}_r(\mathbb{R})$. For $u \in (0,1)$, let $x_u = F_\mu^{-1}(u)$ and $x_u^k = F_{\mu^k}^{-1}(u)$. Let $\mathcal{U} \subset (0,1)$ be the set of continuity points of F_μ^{-1}

and define

$$\mathcal{U}_c = \{u \in \mathcal{U} \mid F_\mu \text{ is continuous at } x_u\} \quad \text{and}$$
$$\mathcal{U}_d = \{u \in \mathcal{U}\backslash\mathcal{U}_c \mid u \in (F_\mu(x_u-), F_\mu(x_u))\}.$$

By monotonicity of F_μ^{-1}, the complement of $\mathcal{U}$ in $(0, 1)$ is at most countable, and since μ has countably many atoms, the complement of $\mathcal{U}_d$ in $\mathcal{U}\backslash\mathcal{U}_c$ is also at most countable. We deduce that it is sufficient to show (3.16) for du-almost all $u \in \mathcal{U}_c \cup \mathcal{U}_d$. Let then $u \in \mathcal{U}$. If $\mu^k(\{x_u^k\}) = 0$, then $C_u^k = C_u$ and

$$\int_{(0,1)} f(F_\nu^{-1}(v))\, C_u^k(dv) = g(u).$$

From now on and until (3.16) is proved, we suppose w.l.o.g. that $\mu^k(\{x_u^k\}) > 0$ for all $k \in \mathbb{N}$. Then

$$\int_{(0,1)} f(F_\nu^{-1}(v))\, C_u^k(dv) = \frac{1}{\mu^k(\{x_u^k\})} \int_{F_{\mu^k}(x_u^k-)}^{F_{\mu^k}(x_u^k)} g(w)\, dw. \tag{3.17}$$

Define $l_k = \inf_{n\geq k} x_u^n$ and $r_k = \sup_{n\geq k} x_u^n$. Since $u \in \mathcal{U}$ we find $l_k \nearrow x_u$ and $r_k \searrow x_u$ when k goes to $+\infty$. Due to right continuity of F_μ and left continuity of $x \mapsto F_\mu(x-)$ we have

$$F_\mu(x_u-) = \lim_p F_\mu(l_p-) \quad \text{and} \quad \lim_p F_\mu(r_p) = F_\mu(x_u).$$

By Portmanteau's theorem and monotonicity of cumulative distribution functions we have

$$\begin{aligned} F_\mu(l_p-) &\leq \liminf_k F_{\mu^k}(l_p-) \leq \liminf_k F_{\mu^k}(x_u^k-) \leq \limsup_k F_{\mu^k}(x_u^k) \\ &\leq \limsup_k F_{\mu^k}(r_p) \leq F_\mu(r_p). \end{aligned}$$

By taking the limit $p \to +\infty$, we find

$$F_\mu(x_u-) \leq \liminf_k F_{\mu^k}(x_u^k-) \leq \limsup_k F_{\mu^k}(x_u^k) \leq F_\mu(x_u). \tag{3.18}$$

By (2.5), the interval $[F_{\mu^k}(x_u^k-), F_{\mu^k}(x_u^k)]$ contains u, and if $u \in \mathcal{U}_c$, then (3.18) implies that its length $\mu^k(\{x_u^k\})$ vanishes when k goes to $+\infty$. Consequently, (3.17) and the Lebesgue differentiation theorem yield that for du-almost every $u \in \mathcal{U}_c$,

$$\int_{(0,1)} f(F_\nu^{-1}(v))\, C_u^k(dv) \to g(u).$$

Suppose now $u \in \mathcal{U}_d$ and define

$$a_k = F_{\mu^k}(x_u^k-) \vee F_\mu(x_u-), \quad b_k = F_{\mu^k}(x_u^k) \wedge F_\mu(x_u).$$

Note that on the interval (a_k, b_k) the function g is constant equal to $g(u)$, so (3.17) amounts to

$$\int_{(0,1)} f(F_\nu^{-1}(v))\, C_u^k(dv)$$
$$= \frac{1}{\mu^k(\{x_u^k\})} \left(\int_{b_k}^{F_{\mu^k}(x_u^k)} g(w)\, dw + \int_{a_k}^{b_k} g(u)\, dw + \int_{F_{\mu^k}(x_u^k-)}^{a_k} g(w)\, dw \right).$$

According to (3.18),

$$a_k - F_{\mu^k}(x_u^k-) \to 0 \quad \text{and} \quad F_{\mu^k}(x_u^k) - b_k \to 0, \quad k \to +\infty. \tag{3.19}$$

Moreover, having (2.5) in mind it is clear that

$$\begin{aligned} F_{\mu^k}(x_u^k-) < a_k &\implies \mu^k(\{x_u^k\}) \geq u - F_\mu(x_u-), \\ \text{and} \quad b_k < F_{\mu^k}(x_u^k) &\implies \mu^k(\{x_u^k\}) \geq F_\mu(x_u) - u. \end{aligned} \tag{3.20}$$

Using the latter fact and the equality

$$b_k - a_k = \mu_k(\{x_u^k\}) - (F_{\mu_k}(x_u^k) - b_k) - (a_k - F_{\mu_k}(x_u^k-)),$$

we get

$$1 - \frac{F_{\mu_k}(x_u^k) - b_k}{F_\mu(x_u) - u} - \frac{a_k - F_{\mu_k}(x_u^k-)}{u - F_\mu(x_u-)} \leq \frac{b_k - a_k}{\mu_k(\{x_u^k\})} \leq 1.$$

Hence by (3.19) we have $\frac{b_k - a_k}{\mu_k(\{x_u^k\})} \to 1$ as k goes to $+\infty$, which implies that $\frac{1}{\mu^k(\{x_u^k\})} \int_{a_k}^{b_k} g(u)\, dw \to g(u)$ as $k \to +\infty$. Therefore, we just have to show that

$$\frac{1}{\mu^k(\{x_u^k\})} \left(\int_{b_k}^{F_{\mu^k}(x_u^k)} g(w)\, dw + \int_{F_{\mu^k}(x_u^k-)}^{a_k} g(w)\, dw \right) \to 0, \quad k \to +\infty. \tag{3.21}$$

Note that we can assume w.l.o.g. that for all $k \in \mathbb{N}$ either $F_{\mu^k}(x_u^k-) < a_k$ or $b_k < F_{\mu^k}(x_u^k)$. Let $d = (u - F_\mu(x_u-)) \wedge (F_\mu(x_u) - u)$, which is positive since $u \in \mathcal{U}_d$.

Then we have by (3.20)

$$\frac{1}{\mu^k(\{x_u^k\})}\left|\int_{b_k}^{F_{\mu^k}(x_u^k)} g(w)\,dw + \int_{F_{\mu^k}(x_u^k-)}^{a_k} g(w)\,dw\right| \\ \leq \frac{1}{d}\left|\int_{b_k}^{F_{\mu^k}(x_u^k)} g(w)\,dw + \int_{F_{\mu^k}(x_u^k-)}^{a_k} g(w)\,dw\right|. \tag{3.22}$$

By the inverse transform sampling and the facts that $f \in \Phi_r(\mathbb{R})$ and $\nu \in \mathcal{P}_r(\mathbb{R})$, we have $\int_0^1 |g(w)|\,dw = \int_{\mathbb{R}} |f(y)|\,\nu(dy) < +\infty$. Then (3.21) is a direct consequence of (3.22), (3.19) and the dominated convergence theorem. Hence (3.14) is proved for du-almost every $u \in (0,1)$.

Next, we show uniform integrability of (3.15). We can estimate

$$\mathcal{W}_r^r((F_\nu^{-1})_* C_u, (F_\nu^{-1})_* C_u^k) \leq 2^{r-1}\left(\int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u(dv) + \int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u^k(dv)\right).$$

Since by the inverse transform sampling we have

$$\int_{(0,1)}\int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u(dv)\,du = \int_{\mathbb{R}} |y|^r\,\nu(dy) < \infty,$$

it is enough to show uniform integrability of $u \mapsto \int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u^k(dv)$, $k \in \mathbb{N}$.

On the one hand, using the inverse transform sampling and $\nu \in \mathcal{P}_r(\mathbb{R})$, we have

$$\forall k \in \mathbb{N},\quad \int_{(0,1)}\int_{(0,1)} |F_\nu^{-1}(v)|^r C_u^k(dv)\,du = \int_{\mathbb{R}} |y|^r\,\nu(dy) < +\infty.$$

On the other hand, let $\epsilon > 0$ and A be a measurable subset of $(0,1)$ such that $\lambda(A) < \epsilon$. We have

$$\int_A\int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u^k(dv)\,du = \int_{\mathbb{R}} |y|^r\,(F_\nu^{-1})_*\tau^k(dy),$$

where $\tau^k(dv) = \int_{u=0}^1 \mathbb{1}_A(du)\,C_u^k(dv)\,du$. Note that $\tau^k \leq \lambda$, $(F_\nu^{-1})_*\tau^k \leq \nu$ and $(F_\nu^{-1})_*\tau^k(\mathbb{R}) = \tau^k((0,1)) = \lambda(A)$. Therefore,

$$\sup_{\substack{A\in\mathcal{B}((0,1)),\\ \lambda(A)\leq\epsilon}}\ \sup_k \int_A\int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u^k(dv)\,du \leq I_\epsilon^r(\nu),$$

where $I_\varepsilon^r(\nu)$ is defined by (3.1). By Lemma 3.1, the right-hand side converges to 0 with $\epsilon \to 0$. This yields uniform integrability of (3.15), which completes the proof. □

As mentioned in Sect. 2, Proposition 2.1 generalises to Polish spaces. Unsurprisingly, the proof of Proposition 2.3 requires radically different tools from its unidimensional equivalent. In particular, we need to recall the so-called Weak Optimal Transport (WOT) problem introduced by Gozlan, Roberto, Samson and Tetali [29] and studied in [28]. Let $C : X \times \mathcal{P}_r(Y) \to \mathbb{R}_+$ be nonnegative, continuous, strictly convex in the second argument and such that there exists a constant $K > 0$ which satisfies

$$\forall (x, p) \in X \times \mathcal{P}_r(Y), \quad C(x, p) \leq K \left(1 + d_X^r(x, x_0) + \int_Y d_Y^r(y, y_0)\, p(dy) \right). \tag{3.23}$$

Then the WOT problem consists in minimising

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

In view of the definition (1.3) of the adapted Wasserstein distance which involves measures on the extended space $X \times \mathcal{P}(Y)$, it is natural to consider an extension of (WOT) which also involves this space. Hence we also consider the extended problem

$$V'_C(\mu, \nu) := \inf_{P \in \Lambda(\mu,\nu)} \int_{X \times \mathcal{P}(Y)} C(x, p)\, P(dx, dp), \tag{WOT'}$$

where $\Lambda(\mu, \nu)$ is the set of couplings between μ and an arbitrary measure on $\mathcal{P}(Y)$ with mean ν, that is

$$\Lambda(\mu, \nu) = \left\{ P \in \mathcal{P}(X \times \mathcal{P}(Y)) \mid \int_{(x',p) \in X \times \mathcal{P}(Y)} \delta_{x'}(dx)\, p(dy)\, P(dx', dp) \in \Pi(\mu, \nu) \right\}. \tag{3.24}$$

Remark 3.3 We gather here useful results on weak transport problems which hold under the standing assumptions on C:

(a) according to [8, Theorem 1.2] and the paragraph following this theorem, (WOT) admits a unique minimiser π^*;
(b) As a consequence of the necessary optimality condition [9, Theorem 2.2], $J(\pi^*)$ is the only minimiser of (WOT'). Indeed, if we assume the opposite then there is a minimizer $P^* \in \Lambda(\mu, \nu)$ of (WOT') which does not lie in the image of $\Pi(\mu, \nu)$ under J. Hence, any measurable set $\mathcal{A} \subset X \times \mathcal{P}_r(Y)$ with $P^*(\mathcal{A}) = 1$ contains $(x, p), (x, q) \in \mathcal{A}$ with $p \neq q$. Due to strict convexity of C in its second argument, we find

$$C\left(x, \frac{p+q}{2}\right) < \frac{1}{2}\left(C(x, p) + C(x, q)\right).$$

Since $\mathcal{A}$ was an arbitrary set supporting P^*, the strict inequality above contradicts the necessary optimality condition in [9, Theorem 2.2];

(c) $V(\mu,\nu)=V'(\mu,\nu)$ [8, Lemma 2.1];

(d) Stability of (WOT) and (WOT'): Let $\mu^k\in\mathcal{P}_r(X)$, $\nu^k\in\mathcal{P}_r(Y)$, $k\in\mathbb{N}$ converge respectively to $\mu\in\mathcal{P}_r(X)$ and $\nu\in\mathcal{P}_r(Y)$ in $\mathcal{W}_r$. For $k\in\mathbb{N}$, let $\pi^k\in\Pi(\mu^k,\nu^k)$ be optimal for $V(\mu^k,\nu^k)$. Then π^k, resp. $J(\pi^k)$, converges to the unique minimiser π^*, resp. $J(\pi^*)$, in $\mathcal{W}_r$ [9, Theorem 1.3 and Corollary 2.9]. In particular, this shows that π^k converges to π^* even in $\mathcal{AW}_r$.

Proof of Proposition 2.3 Since $\nu\in\mathcal{P}_r(Y)$, we have that

$$\int_X\int_Y d_Y^r(y,y_0)\,\pi_x(dy)\,\mu(dx)=\int_Y d_Y^r(y,y_0)\,\nu(dy)<+\infty, \tag{3.25}$$

hence up to a modification on a μ-null set, we can suppose w.l.o.g. that for all $x\in X$, $\pi_x\in\mathcal{P}_r(Y)$. Let $\varepsilon>0$ and $y_0\in Y$. Define for $R>0$ the $\mathcal{W}_r$-open ball B_R of radius $R^{1/r}$ and centre δ_{y_0} and the set

$$A_R=\{x\in X\mid \pi_x\in B_R\}=\left\{x\in X\;\middle|\;\int_Y d_Y^r(y,y_0)\,\pi_x(dy)<R\right\}.$$

By (3.25) again, μ is concentrated on $\bigcup_{R>0}A_R$ and we can choose R large enough such that

$$\mu(X\backslash A_R)<\epsilon.$$

Since μ is a probability measure on the Polish space X, it is a Radon measure. Moreover, $\mathcal{P}_r(Y)$ endowed with $\mathcal{W}_r$ is a separable metric space, hence it is second-countable. Therefore we can apply Lusin's theorem to the map $X\ni x\mapsto\pi_x\in\mathcal{P}_r(Y)$ in order to deduce the existence of a closed set $F\subset A_R$ such that

$$\mu(X\backslash F)<\epsilon\quad\text{and}\quad x\mapsto\pi_x\text{ restricted to }F\text{ is continuous.}$$

Let $\widetilde{\mathcal{M}}_r(Y)$ be the linear space of all finite signed measures on Y, the positive and negative parts of which are contained in $\mathcal{M}_r(Y)$, equipped with the weak topology induced by $\Phi_r(Y)$. Since weak topologies are locally convex, an extension of Tietze's theorem [24, Theorem 4.1] yields the existence of a continuous map $x\mapsto\overline{\pi}_x$ defined on X with values in $\widetilde{\mathcal{M}}_r(Y)$ such that $\overline{\pi}_x=\pi_x$ for all $x\in F$ and

$$\{\overline{\pi}_x\mid x\in X\}\subset\mathrm{co}\{\pi_x\mid x\in F\}\subset B_R,$$

where co denotes the convex hull.

Next, we define a nonnegative, continuous, strictly convex in the second argument function which satisfies a condition of the form (3.23) in order to use the results on weak transport problems detailed in Remark 3.3. Let $\{g_k\mid k\in\mathbb{N}\}\subset\Phi_1(Y)$ be a family of 1-Lipschitz continuous functions and absolutely bounded by 1, which separates $\mathcal{P}(Y)$

(see [25, Theorem 4.5.(a)]). We have for any pair $p, p' \in \mathcal{P}(Y)$, $p \neq p'$ that there is $l \in \mathbb{N}$ such that

$$\int_Y g_l(y)\, p(dy) \neq \int_Y g_l(y)\, p'(dy). \tag{3.26}$$

Define $C : X \times \mathcal{P}_r(Y) \to \mathbb{R}_+$ for all $(x, p) \in X \times \mathcal{P}_r(Y)$ by

$$C(x, p) := \rho(\overline{\pi}_x, p) + \sum_{k\in\mathbb{N}} \frac{1}{2^k} \left| \int_Y g_k(y)\, \overline{\pi}_x(dy) - \int_Y g_k(y)\, p(dy) \right|^2,$$

where $\rho : \mathcal{P}(Y) \times \mathcal{P}(Y) \to [0, 1]$ is defined for all $p, p' \in \mathcal{P}(Y)$ by

$$\rho(p, p') = \inf_{\chi\in\Pi(p,p')} \int_{Y\times Y} (d_Y(y, y') \wedge 1)\, \chi(dy, dy').$$

Since ρ can be interpreted as a Wasserstein distance with respect to a bounded distance, it is immediate that it is a metric on $\mathcal{P}(Y)$ which induces the weak convergence topology. On the one hand, the map $(x, p) \mapsto \rho(\overline{\pi}_x, p)$ is continuous by continuity of $x \mapsto \overline{\pi}_x$. On the other hand, by Kantorovich and Rubinstein's duality theorem and Jensen's inequality, we have for all $(x, p), (x', p') \in X \times \mathcal{P}_r(Y)$

$$\begin{aligned}
&\sum_{k\in\mathbb{N}} \frac{1}{2^k} \left| \left| \int_Y g_k(y)\, \overline{\pi}_x(dy) - \int_Y g_k(y)\, p(dy) \right|^2 - \left| \int_Y g_k(y)\, \overline{\pi}_{x'}(dy) - \int_Y g_k(y)\, p'(dy) \right|^2 \right| \\
&\quad = \sum_{k\in\mathbb{N}} \frac{1}{2^k} \left| \int_Y g_k(y)\, \overline{\pi}_x(dy) - \int_Y g_k(y)\, p(dy) + \int_Y g_k(y)\, \overline{\pi}_{x'}(dy) - \int_Y g_k(y)\, p'(dy) \right| \\
&\qquad \times \left| \int_Y g_k(y)\, \overline{\pi}_x(dy) - \int_Y g_k(y)\, p(dy) - \int_Y g_k(y)\, \overline{\pi}_{x'}(dy) + \int_Y g_k(y)\, p'(dy) \right| \\
&\quad \le \sum_{k\in\mathbb{N}} \frac{4}{2^k} \left(\left| \int_Y g_k(y)\, \overline{\pi}_x(dy) - \int_Y g_k(y)\, \overline{\pi}_{x'}(dy) \right| + \left| \int_Y g_k(y)\, p(dy) - \int_Y g_k(y)\, p'(dy) \right| \right) \\
&\quad \le 8 \left(\mathcal{W}_1(\overline{\pi}_x, \overline{\pi}_{x'}) + \mathcal{W}_1(p, p') \right) \le 8 \left(\mathcal{W}_r(\overline{\pi}_x, \overline{\pi}_{x'}) + \mathcal{W}_r(p, p') \right),
\end{aligned}$$

where the right-hand side vanishes when (x', p') converges to (x, p) by continuity of $x \mapsto \overline{\pi}_x$. We deduce that C is continuous.

Note that ρ is convex in the second argument. Therefore, to obtain strict convexity of $C(x, \cdot)$ in the second argument, it is sufficient to verify that

$$F(p) = \sum_{k\in\mathbb{N}} \frac{1}{2^k} \left| \int_Y g_k(y)\, p(dy) \right|^2$$

is strictly convex. Let $p, p' \in \mathcal{P}(Y)$, $p \neq p'$ and $l \in \mathbb{N}$ such that (3.26) holds. Hence, strict convexity of the square proves

$$\left|\alpha \int_Y g_l(y)\, p(dy) + (1-\alpha) \int_Y g_l(y)\, p'(dy)\right|^2 < \alpha \left|\int_Y g_l(y)\, p(dy)\right|^2 + (1-\alpha)\left|\int_Y g_l(y)\, p'(dy)\right|^2,$$

which yields strict convexity of F on $\mathcal{P}(Y)$.

Moreover, we have for all $(x, p) \in X \times \mathcal{P}_r(Y)$, $C(x, p) \leq 1 + 8 = 9$, hence C satisfies (3.23). Remember the definitions of V_C and V'_C given in (WOT) and (WOT'). Since for all $x \in F$, $C(x, \pi_x) = C(x, \overline{\pi}_x) = 0$, we have

$$V_C(\mu, \nu) \leq \int_{X \setminus F} C(x, \pi_x)\, \mu(dx) < 9\varepsilon.$$

Let $\pi^{*,\varepsilon} \in \Pi(\mu, \nu)$ be optimal for $V_C(\mu, \nu)$. For $P, P' \in \mathcal{P}(X \times \mathcal{P}(Y))$, let

$$\tilde{\rho}(P, P') = \inf_{\chi \in \Pi(P, P')} \int_{X \times \mathcal{P}(Y) \times X \times \mathcal{P}(Y)} \big(\big(d_X(x, x') + \rho(p, p')\big) \wedge 1\big)\, \chi(dx, dp, dx', dp').$$

Since $\mu(dx)\, \delta_{\pi_x}(dp)\, \delta_x(dx')\, \delta_{\pi^{*,\varepsilon}_{x'}}(dp')$ is a coupling between $J(\pi)$ and $J(\pi^{*,\varepsilon})$, we can estimate

$$\begin{aligned}
&\tilde{\rho}(J(\pi), J(\pi^{*,\varepsilon})) \\
&\quad \leq \int_X \rho(\pi_x, \pi^{*,\varepsilon}_x)\, \mu(dx) \\
&\quad \leq \int_F \rho(\pi_x, \pi^{*,\varepsilon}_x)\, \mu(dx) \\
&\qquad + \int_{X \setminus F} \int_Y (d_Y(y, y_0) \wedge 1)\, (\pi_x + \pi^{*,\varepsilon}_x)(dy)\, \mu(dx) \\
&\quad \leq V_C(\mu, \nu) + 2\varepsilon < 11\varepsilon.
\end{aligned}$$

For $k \in \mathbb{N}$, let $\pi^{k,\varepsilon} \in \Pi(\mu^k, \nu^k)$ be optimal for $V_C(\mu^k, \nu^k)$. Then $J(\pi^{k,\varepsilon})$ is optimal for $V'_C(\mu^k, \nu^k)$ by Remark 3.3 (b), and converges to $J(\pi^{*,\varepsilon})$ in $\mathcal{W}_r$ and therefore weakly by Remark 3.3 (d). Then we get

$$\begin{aligned}
&\limsup_{k \to +\infty} \tilde{\rho}(J(\pi^{k,\varepsilon}), J(\pi)) \\
&\quad \leq \limsup_{k \to +\infty} \Big(\tilde{\rho}(J(\pi^{k,\varepsilon}), J(\pi^{*,\varepsilon})) + \tilde{\rho}(J(\pi^{*,\varepsilon}), J(\pi))\Big) \leq 11\varepsilon. \qquad (3.27)
\end{aligned}$$

So far $\epsilon > 0$ was arbitrary. Therefore, there exists a strictly increasing sequence $(k_N)_{N\in\mathbb{N}^*}$ of positive integers such that

$$\forall N \in \mathbb{N}^*, \quad \forall k \geq k_N, \quad \tilde{\rho}(J(\pi^{k,1/N}), J(\pi)) \leq \frac{12}{N}.$$

For $k \in \mathbb{N}$, let $N_k = \max\{N \in \mathbb{N}^* \mid k \geq k_N\}$, where the maximum of the empty set is defined as 1. Since $(k_N)_{N\in\mathbb{N}^*}$ is strictly increasing, we find that $N_k \to +\infty$ as $k \to +\infty$. Then the sequence of couplings

$$\pi^k = \pi^{k,1/N_k} \in \Pi(\mu^k, \nu^k), \ k \in \mathbb{N}$$

is such that $\tilde{\rho}(J(\pi^k), J(\pi))$ vanishes as k goes to $+\infty$, and therefore $J(\pi^k)$ converges weakly to $J(\pi)$. Moreover, since $\mathcal{W}_r$-convergence is equivalent to weak convergence coupled with convergence of the r-moments, we have that the r-moments of μ^k and ν^k respectively converge to the r-moments of μ and ν, which implies

$$\begin{aligned}
&\int_{X\times\mathcal{P}(Y)} \left(d_X^r(x, x_0) + \mathcal{W}_r^r(p, \delta_{y_0})\right) J(\pi^k)(dx, dp) \\
&= \int_X \left(d_X^r(x, x_0) + \mathcal{W}_r^r(\pi_x^k, \delta_{y_0})\right) \mu^k(dx) \\
&= \int_X d_X^r(x, x_0)\, \mu^k(dx) + \int_Y d_Y^r(y, y_0)\nu^k(dy) \\
&\underset{k\to+\infty}{\longrightarrow} \int_X d_X^r(x, x_0)\, \mu(dx) + \int_Y d_Y^r(y, y_0)\, \nu(dy) \\
&= \int_{X\times\mathcal{P}(Y)} \left(d_X^r(x, x_0) + \mathcal{W}_r^r(p, \delta_{y_0})\right) J(\pi)(dx, dp).
\end{aligned}$$

We deduce that $J(\pi^k)$ converges to $J(\pi)$ in $\mathcal{W}_r$ as $k \to +\infty$. According to (1.3), $\pi^{k,\varepsilon}$ converges to $\pi^{*,\varepsilon}$ in $\mathcal{AW}_r$, which concludes the proof. □

In the proof of Theorem 2.6 we need to be able to confine approximative sequences of couplings to certain sets. The next result provides all necessary tools for this.

Lemma 3.4 *Let $\mu, \mu^k \in \mathcal{M}_r(X)$, $\nu, \nu^k \in \mathcal{M}_r(Y)$, $k \in \mathbb{N}$ all with equal masses and $\pi^k \in \Pi(\mu^k, \nu^k)$, $k \in \mathbb{N}$, converge to $\pi \in \Pi(\mu, \nu)$ in $\mathcal{AW}_r$. Let also $A \subset X$ be measurable and $B \supset A$ be open.*

(i) *There are $\tilde{\mu}^k \leq \mu^k|_B$ and $\epsilon_k \geq 0$, $k \in \mathbb{N}$ such that $\tilde{\mu}^k(B) = (1 - \epsilon_k)\mu(A)$ and $\tilde{\pi}^k := \tilde{\mu}^k \times \pi_x^k$ satisfies*

$$\mathcal{AW}_r(\tilde{\pi}^k, (1 - \epsilon_k)\pi|_{A\times Y}) + \epsilon_k \underset{k\to+\infty}{\longrightarrow} 0.$$

(ii) *Let $C \subset Y$ be an open set on which ν is concentrated. There are $\hat{\mu}^k \leq \tilde{\mu}^k$, $\hat{\nu}^k \leq \nu^k$, $\hat{\pi}^k = \hat{\mu}^k \times \hat{\pi}^k_x \in \Pi(\hat{\mu}^k, \hat{\nu}^k)$ concentrated on $B \times C$ and $\epsilon'_k \geq 0$, $k \in \mathbb{N}$ such that*

$$\mathcal{AW}_r^r(\hat{\pi}^k, (1-\epsilon'_k)\pi|_{A\times Y}) + \int_X \mathcal{W}_r^r(\hat{\pi}^k_x, \pi^k_x)\, \hat{\mu}^k(dx) + \varepsilon'_k \underset{k\to+\infty}{\longrightarrow} 0.$$

Proof To give the reader some guidance we first give an informal description of the strategy of the proof: In order to find $(\tilde{\pi}^k)_{k\in\mathbb{N}}$ and $(\hat{\pi}^k)_{k\in\mathbb{N}}$, we first pick, for $k \in \mathbb{N}$, optimizers $\chi^k \in \Pi(\mu^k, \mu)$ for $\mathcal{AW}_r(\pi^k, \pi)$. Denote by $\tilde{\pi}^k$ the composition of the first marginal of $\chi^k|_{B\times A}$ with the kernel $(\pi^k_x)_{x\in X}$. By approximation arguments we will then deduce that $\tilde{\pi}^k$ has the desired properties. In the last step, we adequately modify $\tilde{\pi}^k$ to a coupling $\hat{\pi}^k$ with second marginal concentrated on C.

Both assertions are trivial if $\mu(A) = 0$ (and also when $A = X$). So assume that $\mu(A) > 0$.

(i) Let $\chi^k \in \Pi(\mu^k, \mu)$ be optimal for $\mathcal{AW}_r(\pi^k, \pi)$ and $\tilde{\mu}^k$ be the first marginal of $\chi^k|_{B\times A}$, $k \in \mathbb{N}$. We set $\tilde{\pi}^k = \tilde{\mu}^k \times \pi^k_x$ and

$$\varepsilon_k = 1 - \frac{\chi^k(B\times A)}{\chi^k(X\times A)} = 1 - \frac{\tilde{\mu}^k(X)}{\mu(A)}. \tag{3.28}$$

Let us prove that ε_k goes to 0 as $k \to \infty$ before checking that the same holds for $\mathcal{AW}_r(\tilde{\pi}^k, (1-\epsilon_k)\pi|_{A\times Y})$.

Let $\chi = (\mathrm{id}, \mathrm{id})_*\mu$. Since $\chi^k(dx_1, dx_2)\, \delta_{(x_2,x_2)}(dx_3, dx_4)$ defines a coupling in $\Pi(\chi^k, \chi)$, we find

$$\begin{aligned}\mathcal{W}_r^r(\chi^k, \chi) &\leq \int_{X^4} (d_X(x_1, x_3)^r + d_X(x_2, x_4)^r)\, \chi^k(dx_1, dx_2)\, \delta_{(x_2,x_2)}(dx_3, dx_4)\\ &= \int_{X\times X} d_X(x_1, x_2)^r\, \chi^k(dx_1, dx_2) \leq \mathcal{AW}_r^r(\pi^k, \pi) \to 0,\ k \to +\infty.\end{aligned}$$

Further, let $P : \mathcal{P}_r(X\times X) \to \mathcal{P}(X\times X)$ be the homeomorphism given by

$$P(\eta)(dx_1, dx_2) = \frac{(1 + d_X(x_1, x_0)^r + d_X(x_2, x_0)^r)\, \eta(dx_1, dx_2)}{\int_{X\times X}(1 + d_X(x'_1, x_0)^r + d_X(x'_2, x_0)^r)\, \eta(dx'_1, dx'_2)},$$

for $\eta \in \mathcal{P}_r(X \times X)$. Recall (1.1), then it is easy to deduce that $P(\eta') \to P(\eta)$ weakly if and only if $\eta' \to \eta$ in $\mathcal{W}_r$. In particular, we find that $P(\chi^k) \to P(\chi)$ weakly as k goes to $+\infty$. Let $f \in \Phi_r(X \times X)$ and

$$\varphi\colon X\times X\colon (x_1, x_2) \mapsto \frac{\mathbb{1}_{X\times A}(x_1, x_2) f(x_1, x_2)}{1 + d_X(x_1, x_0)^r + d_X(x_2, x_0)^r}.$$

Then φ is a bounded measurable map which is continuous w.r.t. the first coordinate. As a consequence of [42, Lemma 2.1], we find

$$\int_{X\times X} \varphi(x_1, x_2)\, P(\chi^k)(dx_1, dx_2) \to \int_{X\times X} \varphi(x_1, x_2)\, P(\chi)(dx_1, dx_2), \quad k \to +\infty,$$

which amounts to

$$\int_{X\times X} f(x_1,x_2)\,\chi^k|_{X\times A}(dx_1,dx_2) \to \int_{X\times X} f(x_1,x_2)\,\chi|_{X\times A}(dx_1,dx_2), \quad k\to+\infty.$$

Therefore (1.1) yields $\mathcal{W}_r$-convergence of $\chi^k|_{X\times A}$ to $\chi|_{X\times A}$. By Portmanteau's theorem, we have

$$\begin{aligned}\chi^k(B\times A)\le\chi^k(X\times A)=&\mu(A)=\chi|_{X\times A}(B\times B)\\ &\le\liminf_{k\to+\infty}\chi^k|_{X\times A}(B\times B)=\liminf_{k\to+\infty}\chi^k(B\times A),\end{aligned}$$

By the first equality in (3.28), we deduce that ϵ_k, $k\in\mathbb{N}$ is a null sequence of nonnegative real numbers. We now want to show that

$$\mathcal{AW}_r(\tilde\mu^k\times\pi_x^k,(1-\epsilon_k)\mu|_A\times\pi_x)\to 0. \tag{3.29}$$

On the one hand, denoting by $\bar\mu^k$ the second marginal of $\chi^k|_{B\times A}$, we have that

$$\begin{aligned}\mathcal{AW}_r^r(\tilde\mu^k\times\pi_x^k,\bar\mu^k\times\pi_x)&\le\int_{X\times X}\left(d_X^r(x,x')+\mathcal{W}_r^r(\pi_x^k,\pi_{x'})\right)\chi^k|_{B\times A}(dx,dx')\\ &\le\int_{X\times X}\left(d_X^r(x,x')+\mathcal{W}_r^r(\pi_x^k,\pi_{x'})\right)\chi^k(dx,dx')\\ &=\mathcal{AW}_r^r(\pi^k,\pi)\to 0,\quad k\to+\infty.\end{aligned}\tag{3.30}$$

On the other hand, let

$$\check\mu^k=(1-\epsilon_k)\mu|_A,\,\zeta^k=\check\mu^k\wedge\bar\mu^k\quad\text{and}\alpha_k=\bar\mu^k(X)-\zeta^k(X)=\check\mu^k(X)-\zeta^k(X).$$

Let $\overline{\chi}^k\in\Pi(\bar\mu^k-\zeta^k,\check\mu^k-\zeta^k)$ be optimal for $\mathcal{AW}_r^r((\bar\mu^k-\zeta^k)\times\pi_x,(\check\mu^k-\zeta^k)\times\pi_x)$. Since $((\mathrm{id},\mathrm{id})_*\zeta^k+\overline{\chi}^k)$ is a coupling between $\bar\mu^k$ and $\check\mu^k$, we find

$$\begin{aligned}\mathcal{AW}_r(\bar\mu^k\times\pi_x,\check\mu^k\times\pi_x)&\le\int_X\left(d_X^r(x,x')+\mathcal{W}_r^r(\pi_x,\pi_{x'})\right)\overline{\chi}^k(dx,dx')\\ &=\mathcal{AW}_r^r((\bar\mu^k-\zeta^k)\times\pi_x,(\check\mu^k-\zeta^k)\times\pi_x)\\ &\le\mathcal{AW}_r((\bar\mu^k-\zeta^k)\times\pi_x,\alpha_k\delta_{(x_0,y_0)})+\mathcal{AW}_r((\check\mu^k-\zeta^k)\times\pi_x,\alpha_k\delta_{(x_0,y_0)}).\end{aligned}$$

In the next estimates we use (3.1). Note that the first marginal of $(\bar{\mu}^k - \zeta^k) \times \pi_x$ is dominated by μ whereas its second marginal is dominated by ν. Thus, denoting $\tau^k(dy) = \int_X \pi_x(dy)\,(\bar{\mu}^k - \zeta^k)(dx)$, we find

$$
\begin{aligned}
&\mathcal{AW}_r^r((\bar{\mu}^k - \zeta^k) \times \pi_x, \alpha_k \delta_{(x_0,y_0)}) = \int_X \left(d_X^r(x,x_0) + \mathcal{W}_r^r(\pi_x, \delta_{y_0})\right)(\bar{\mu}^k - \zeta^k)(dx) \\
&= \int_X d_X^r(x,x_0)\,(\bar{\mu}^k - \zeta^k)(dx) + \int_Y d_Y^r(y,y_0)\,\tau^k(dy) \\
&\le I_{\alpha_k}^r(\mu) + I_{\alpha_k}^r(\nu).
\end{aligned}
$$

Similarly, we find

$$
\mathcal{AW}_r^r((\check{\mu}^k - \zeta^k) \times \pi_x, \alpha_k \delta_{(x_0,y_0)}) \le I_{\alpha_k}^r(\mu) + I_{\alpha_k}^r(\nu).
$$

If we can show that α_k vanishes for $k \to +\infty$, then we find by Lemma 3.1 (b) that

$$
\mathcal{AW}_r(\bar{\mu}^k \times \pi_x, \check{\mu}^k \times \pi_x) \underset{k\to+\infty}{\longrightarrow} 0, \tag{3.31}
$$

and the triangle inequality together with (3.30) and (3.31) yield the assertion, (3.29).

Since $\check{\mu}^k, \bar{\mu}^k \le \mu|_A$, the densities of $\check{\mu}^k$ and $\bar{\mu}^k$ with respect to $\mu|_A$ satisfy $\frac{d\check{\mu}^k}{d\mu|_A}, \frac{d\bar{\mu}^k}{d\mu|_A} \le 1$. Then we conclude by

$$
\begin{aligned}
\alpha_k &= \bar{\mu}^k(X) - \zeta^k(X) \\
&= \int_A \left(\frac{d\bar{\mu}^k}{d\mu|_A}(x) - \frac{d\check{\mu}^k}{d\mu|_A}(x)\right)^+ \mu(dx) \le \int_A \left(1 - \frac{d\check{\mu}^k}{d\mu|_A}(x)\right)\mu(dx) \\
&= \mu(A) - \check{\mu}^k(A) = \varepsilon_k \mu(A) \underset{k\to+\infty}{\longrightarrow} 0.
\end{aligned}
$$

(ii) Let $\tilde{\nu}^k$ and $\tilde{\nu}$ denote the second marginals of $\tilde{\mu}^k \times \pi_x^k$ and $\mu|_A \times \pi_x$ respectively. Since $\mu|_A \times \pi_x \le \mu \times \pi_x$ with the second marginal ν of the right-hand side concentrated on C, $\tilde{\nu}$ is concentrated on C and $\tilde{\nu}(C) = \mu(A)$. In a similar way, since $\tilde{\mu}^k \times \pi_x^k \le \mu^k \times \pi_x^k$, we have $\tilde{\nu}^k \le \nu^k$. In order to modify $\tilde{\mu}^k \times \pi_x^k$ into a coupling with second marginal concentrated on C, we consider $\tilde{\mu}^k \times \mathring{\pi}_x^k$ with $\mathring{\pi}_x^k(dz) = \int_Y \mathring{\chi}_y^k(dz)\,\pi_x^k(dy)$ where the coupling $\mathring{\chi}^k \in \Pi(\tilde{\nu}^k, (1-\epsilon_k)\tilde{\nu})$ is $\mathcal{W}_r$-optimal. To enable comparison of the second marginal with ν^k as in the statement, we take advantage of the inequality $\tilde{\nu}^k \le \nu^k$ and introduce $\tilde{\mu}^k \times \hat{\pi}_x^k$ with $\hat{\pi}_x^k(dt) = \int_Y \hat{\chi}_z^k(dt)\,\mathring{\pi}_x^k(dz)$ where the coupling $\hat{\chi}^k \in \Pi((1-\epsilon_k)\tilde{\nu}, (1-\epsilon_k)\frac{\tilde{\nu}(C)}{\tilde{\nu}^k(C)}\tilde{\nu}^k|_C)$ is $\mathcal{W}_r$-optimal. The second marginal of $\hat{\pi}^k = \tilde{\mu}^k \times \hat{\pi}_x^k$ is $(1-\epsilon_k)\frac{\tilde{\nu}(C)}{\tilde{\nu}^k(C)}\tilde{\nu}^k|_C$. By the equality $\tilde{\nu}(C) = \mu(A)$ and (3.28) for the equality then the definition of $\tilde{\nu}^k$ for the inequality, one has

$$
(1-\epsilon_k)\frac{\tilde{\nu}(C)}{\tilde{\nu}^k(C)} = \frac{\tilde{\mu}^k(X)}{\tilde{\nu}^k(C)} \ge 1.
$$

Setting $\hat{\mu}^k = \frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)}\tilde{\mu}^k \le \tilde{\mu}^k$ then ensures that the second marginal $\hat{\nu}^k = \tilde{\nu}^k|_C$ of $\hat{\mu}^k \times \hat{\pi}_x^k$ is both concentrated on C and not greater than ν^k. Moreover, $\hat{\nu}^k(C) = \hat{\nu}^k(Y) = \hat{\mu}^k(X) \le \tilde{\mu}^k(X)$ with the right-hand side not greater than $\mu(A)$ by (3.28). Hence

$$\epsilon_k' := 1 - \frac{\tilde{\nu}^k(C)}{\mu(A)} \in [0, 1]\cdot \tag{3.32}$$

Then it remains to show that

$$\mathcal{AW}_r(\hat{\pi}^k, (1-\epsilon_k')\pi|_{A\times Y}) + \int_X \mathcal{W}_r^r(\hat{\pi}_x^k, \pi_x^k)\,\hat{\mu}^k(dx) + \epsilon_k' \underset{k\to+\infty}{\longrightarrow} 0. \tag{3.33}$$

Since we have

$$\begin{aligned} \pi_x^k(dy)\,\mathring{\chi}_y^k(dz) \in \Pi(\pi_x^k, \mathring{\pi}_x^k), \quad \mathring{\pi}_x^k(dz)\,\hat{\chi}_z^k(dt) \in \Pi(\mathring{\pi}_x^k, \hat{\pi}_x^k),\\ \int_{x\in X} \hat{\mu}^k(dx)\,\pi_x^k(dy)\,\mathring{\chi}_y^k(dz) = \frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)}\mathring{\chi}^k(dy,dz),\\ \int_{x\in X} \hat{\mu}^k(dx)\,\mathring{\pi}_x^k(dz)\,\hat{\chi}_z^k(dt) = \frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)}\hat{\chi}^k(dz,dt), \end{aligned}$$

we find plugging the expressions (3.28) and (3.32) that

$$\begin{aligned} &\mathcal{AW}_r^r(\hat{\mu}^k \times \pi_x^k, \hat{\mu}^k \times \hat{\pi}_x^k)\\ &\le \int_X \mathcal{W}_r^r(\pi_x^k, \hat{\pi}_x^k)\,\hat{\mu}^k(dx)\\ &\le 2^{r-1}\int_X \left(\mathcal{W}_r^r(\pi_x^k, \mathring{\pi}_x^k) + \mathcal{W}_r^r(\mathring{\pi}_x^k, \hat{\pi}_x^k)\right)\hat{\mu}^k(dx)\\ &\le 2^{r-1}\int_X \left(\int_{Y\times Y} d_Y^r(y,z)\,\pi_x^k(dy)\,\mathring{\chi}_y^k(dz) + \int_{Y\times Y} d_Y^r(z,t)\,\mathring{\pi}_x^k(dz)\,\hat{\chi}_z^k(dt)\right)\hat{\mu}^k(dx)\\ &= 2^{r-1}\left(\frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)}\int_{Y\times Y} d_Y^r(y,z)\,\mathring{\chi}^k(dy,dz) + \frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)}\int_{Y\times Y} d_Y^r(z,t)\,\hat{\chi}^k(dz,dt)\right)\\ &= 2^{r-1}\left(\frac{1-\varepsilon_k'}{1-\varepsilon_k}\mathcal{W}_r^r(\tilde{\nu}^k, (1-\varepsilon_k)\tilde{\nu}) + \frac{1}{\mu(A)}\mathcal{W}_r^r(\tilde{\nu}^k(C)\tilde{\nu}, \tilde{\nu}(C)\tilde{\nu}^k|_C)\right). \end{aligned}$$

To see convergence to 0, note that since $\mathcal{AW}_r$ dominates $\mathcal{W}_r$, we find by continuity of the projection on the second marginal that (3.29) implies

$$\mathcal{W}_r(\tilde{\nu}^k, (1-\epsilon_k)\tilde{\nu}) \to 0, \quad k \to +\infty.$$

Using Portmanteau's theorem and the fact that $(1-\varepsilon_k) \to 1$ as k goes to $+\infty$, we have for all nonnegative function $f \in \Phi_r(Y)$

$$\limsup_{k\to+\infty} \tilde{\nu}^k(\mathbb{1}_C f) \le \limsup_{k\to+\infty} \tilde{\nu}^k(f) = \tilde{\nu}(f) = \tilde{\nu}(\mathbb{1}_C f) \le \liminf_{k\to+\infty} \tilde{\nu}^k(\mathbb{1}_C f),$$

hence

$$\tilde{\nu}^k|_C(f) \to \tilde{\nu}(f), \quad k \to +\infty. \tag{3.34}$$

Moreover, (3.34) applied with $f = 1$ yields $\tilde{\nu}^k(C) \to \tilde{\nu}(C) = \mu(A)$ as k goes to $+\infty$, hence ε'_k vanishes as k goes to $+\infty$ and

$$\mathcal{W}_r(\tilde{\nu}^k(C)\tilde{\nu}, \tilde{\nu}(C)\tilde{\nu}^k|_C) \to 0, \quad k \to +\infty.$$

We deduce that

$$\mathcal{AW}_r^r(\hat{\mu}^k \times \pi_x^k, \hat{\mu}^k \times \hat{\pi}_x^k) \le \int_X \mathcal{W}_r^r(\pi_x^k, \hat{\pi}_x^k)\, \hat{\mu}^k(dx) \to 0, \quad k \to +\infty.$$

On the other hand, by the definition of $\hat{\mu}^k$ as $\frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)}\tilde{\mu}^k$, (3.28) and (3.32) we have $\hat{\mu}^k = \frac{1-\varepsilon'_k}{1-\varepsilon_k}\tilde{\mu}^k$, hence

$$\mathcal{AW}_r(\hat{\mu}^k \times \pi_x^k, (1-\varepsilon'_k)\mu|_A \times \pi_x) = \frac{1-\varepsilon'_k}{1-\varepsilon_k}\mathcal{AW}_r(\tilde{\mu}^k \times \pi_x^k, (1-\varepsilon_k)\mu|_A \times \pi_x),$$

where the right-hand side vanishes as k goes to $+\infty$ by the first part. Then (3.33) follows by triangle inequality and the latter convergences, which completes the proof. □

The addition of measures is continuous with respect to the weak and Wasserstein topology. More precisely, we have the estimate

$$\mathcal{W}_r^r(\mu + \mu', \nu + \nu') \le \mathcal{W}_r^r(\mu, \nu) + \mathcal{W}_r^r(\mu', \nu')$$

for all measures $\mu, \mu', \nu, \nu' \in \mathcal{P}_r(X)$ such that μ and ν, resp. μ' and ν' have equal masses.

When considering the adapted weak topology, the next example disproves a comparable statement.

Example 3.5 Let $X = Y = \mathbb{R}$, and $\pi^k = \delta_{\left(\frac{1}{k},1\right)}$, $\chi^k = \delta_{\left(-\frac{1}{k},-1\right)}$, $k \in \mathbb{N}$. Then both sequences are convergent in $\mathcal{AW}_1$, but

$$\mathcal{AW}_1(\pi^k + \chi^k, \delta_{(0,1)} + \delta_{(0,-1)}) = \frac{2}{k} + 2$$

does not vanish.

However, we show in the next lemma that the addition of measures with respect to the adapted weak topology can still be considered to be continuous in a certain sense if one of the limits has mass significantly smaller than the other.

Lemma 3.6 *Let* $\hat{\mu}, \hat{\mu}^k, \hat{\nu}, \hat{\nu}^k \in \mathcal{M}_r(Y)$, $k \in \mathbb{N}$ *be with equal masses and* $\tilde{\mu}, \tilde{\mu}^k, \tilde{\nu}, \tilde{\nu}^k \in \mathcal{M}_r(Y)$, $k \in \mathbb{N}$ *be with equal masses smaller than* ε. *Let* $\hat{\pi}^k \in \Pi(\hat{\mu}^k, \hat{\nu}^k)$, $\tilde{\pi}^k \in \Pi(\tilde{\mu}^k, \tilde{\nu}^k)$, $k \in \mathbb{N}$, $\hat{\pi} \in \Pi(\hat{\mu}, \hat{\nu})$ *and* $\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})$. *Let* $\mu = \hat{\mu} + \tilde{\mu}$ *and* $\nu = \hat{\nu} + \tilde{\nu}$. *Then*

(a) *We have for all* $k \in \mathbb{N}$

$$\begin{aligned} &\mathcal{AW}_r^r(\hat{\pi}^k + \tilde{\pi}^k, \hat{\pi} + \tilde{\pi}) \\ &\quad \le \mathcal{AW}_r^r(\hat{\pi}^k, \hat{\pi}) + 2^{r-1}\left(I_\varepsilon^r(\tilde{\mu}) + I_\varepsilon^r(\tilde{\mu}^k) + I_\varepsilon^r(\tilde{\nu}) + I_\varepsilon^r(\tilde{\nu}^k) + 2I_\varepsilon^r(\hat{\nu}) + 2I_\varepsilon^r(\hat{\nu}^k)\right) \\ &\quad \le \mathcal{AW}_r^r(\hat{\pi}^k, \hat{\pi}) + (2^{r-1})^2\left(\mathcal{W}_r^r(\tilde{\mu}^k, \tilde{\mu}) + \mathcal{W}_r^r(\tilde{\nu}^k, \tilde{\nu}) + 2\mathcal{W}_r^r(\hat{\nu}^k, \hat{\nu})\right) \\ &\qquad + 2^{r-1}(1 + 2^{r-1})I_\varepsilon^r(\mu) + 3 \cdot 2^{r-1}(1 + 2^{r-1})I_\varepsilon^r(\nu), \end{aligned} \tag{3.35}$$

where $I_\varepsilon^r(\cdot)$ *is defined by* (3.1).

(b) *If* $(\hat{\pi}^k)_{k\in\mathbb{N}}$ *converges to* $\hat{\pi}$ *in* $\mathcal{AW}_r$ *and* $(\mu^k = \hat{\mu}^k + \tilde{\mu}^k)_{k\in\mathbb{N}}$, *resp.* $(\nu^k = \hat{\nu}^k + \tilde{\nu}^k)_{k\in\mathbb{N}}$, *converges to* μ, *resp.* ν, *in* $\mathcal{W}_r$, *then*

$$\limsup_{k\to+\infty} \mathcal{AW}_r^r(\hat{\pi}^k + \tilde{\pi}^k, \hat{\pi} + \tilde{\pi}) \le C(I_\varepsilon^r(\mu) + I_\varepsilon^r(\nu)), \tag{3.36}$$

where $C > 0$ *depends only on* r.

Proof The second inequality of (3.35) is easily deduced from the first one, (3.2) and the fact that $I_\varepsilon^r(\tilde{\mu}) \le I_\varepsilon^r(\mu)$, $I_\varepsilon^r(\tilde{\nu}) \le I_\varepsilon^r(\nu)$ and $I_\varepsilon^r(\hat{\nu}) \le I_\varepsilon^r(\nu)$.

To see (b), assume for a moment that the first inequality of (3.35) holds true and suppose

$$\hat{\pi}^k \to \hat{\pi} \text{ in } \mathcal{AW}_r, \quad \mu^k = \hat{\mu}^k + \tilde{\mu}^k \to \mu \quad \text{and} \quad \nu^k = \hat{\nu}^k + \tilde{\nu}^k \to \nu \text{ in } \mathcal{W}_r$$

as $k \to +\infty$. Using Lemma 3.1 (a) and then (c), we obtain

$$\begin{aligned} &\limsup_{k\to+\infty} \mathcal{AW}_r^r(\hat{\pi}^k + \tilde{\pi}^k, \hat{\pi} + \tilde{\pi}) \\ &\quad \le C' \limsup_{k\to+\infty} \left(I_\varepsilon^r(\mu^k) + I_\varepsilon^r(\nu^k) + I_\varepsilon^r(\mu) + I_\varepsilon^r(\nu)\right) \\ &\quad \le C \limsup_{k\to+\infty} \left(\mathcal{W}_r^r(\mu^k, \mu) + \mathcal{W}_r^r(\nu^k, \nu) + I_\varepsilon^r(\mu) + I_\varepsilon^r(\nu)\right) \\ &\quad = C(I_\varepsilon^r(\mu) + I_\varepsilon^r(\nu)), \end{aligned}$$

where $C, C' > 0$ depend only on r. Hence (b) is proved.

To conclude the proof, it remains to show the first inequality in (3.35). Let $\hat{\rho}^k \in \Pi(\hat{\mu}^k, \hat{\mu})$ be optimal for $\mathcal{AW}_r(\hat{\pi}^k, \hat{\pi})$ and $\tilde{\rho}^k \in \Pi(\tilde{\mu}^k, \tilde{\mu})$ be arbitrary. We write

$\rho^k = \hat{\rho}^k + \tilde{\rho}^k$. Then

$$\begin{aligned}&\mathcal{AW}_r^r(\hat{\pi}^k + \tilde{\pi}^k, \hat{\pi} + \tilde{\pi}) \\ &\quad \leq \int_{X\times X} \left(d_X^r(x, x') + \mathcal{W}_r^r((\hat{\pi}^k + \tilde{\pi}^k)_x, (\hat{\pi} + \tilde{\pi})_{x'}) \right) \rho^k(dx, dx'). \end{aligned} \tag{3.37}$$

Let $\hat{p} = \frac{d\hat{\mu}}{d\mu}$ and $\hat{p}^k = \frac{d\hat{\mu}^k}{d\mu^k}$. Notice that $\hat{p}$ and $\hat{p}^k$ take values in [0, 1]. The identities

$$\begin{aligned}(\hat{\pi} + \tilde{\pi})(dx, dx') &= \mu(dx) \left(\hat{p}(x)\, \hat{\pi}_x(dx') + (1 - \hat{p}(x))\, \tilde{\pi}_x(dx') \right), \\ (\hat{\pi}^k + \tilde{\pi}^k)(dx, dx') &= \mu^k(dx) \left(\hat{p}^k(x)\, \hat{\pi}_x^k(dx') + (1 - \hat{p}^k(x))\, \tilde{\pi}_x^k(dx') \right),\end{aligned}$$

provide representations for the disintegrations of $(\hat{\pi} + \tilde{\pi})$ and $(\hat{\pi}^k + \tilde{\pi}^k)$ respectively for $\mu(dx)$- and $\mu^k(dx)$-almost every x:

$$(\hat{\pi} + \tilde{\pi})_x = \hat{p}(x)\, \hat{\pi}_x + (1 - \hat{p}(x))\tilde{\pi}_x, \quad (\hat{\pi}^k + \tilde{\pi}^k)_x = \hat{p}^k(x)\, \hat{\pi}_x^k + (1 - \hat{p}^k(x))\, \tilde{\pi}_x^k.$$

Thus, we have when letting $\alpha_+^k(x, x') = (\hat{p}^k(x) - \hat{p}(x'))^+$, $\alpha_-^k(x, x') = (\hat{p}^k(x) - \hat{p}(x'))^-$ and $\beta^k(x, x') = \hat{p}^k(x) \wedge \hat{p}(x')$ that

$$\begin{aligned}&\mathcal{W}_r^r((\hat{\pi}^k + \tilde{\pi}^k)_x, (\hat{\pi} + \tilde{\pi})_{x'}) \\ &\quad \leq \mathcal{W}_r^r(\beta^k(x, x')\, \hat{\pi}_x^k, \beta^k(x, x')\, \hat{\pi}_{x'}) \\ &\qquad + \mathcal{W}_r^r\left(\alpha_+^k(x, x')\, \hat{\pi}_x^k + (1 - \hat{p}^k(x))\, \tilde{\pi}_x^k, \alpha_-^k(x, x')\, \hat{\pi}_{x'} + (1 - \hat{p}(x'))\, \tilde{\pi}_{x'} \right) \\ &\quad \leq \beta^k(x, x')\mathcal{W}_r^r(\hat{\pi}_x^k, \hat{\pi}_{x'}) + 2^{r-1} \Big(\alpha_+^k(x, x')\mathcal{W}_r^r(\hat{\pi}_x^k, \delta_{y_0}) + (1 - \hat{p}^k(x))\mathcal{W}_r^r(\tilde{\pi}_x^k, \delta_{y_0}) \\ &\qquad + \alpha_-^k(x, x')\mathcal{W}_r^r(\hat{\pi}_{x'}, \delta_{y_0}) + (1 - \hat{p}(x'))\mathcal{W}_r^r(\tilde{\pi}_{x'}, \delta_{y_0}) \Big).\end{aligned} \tag{3.38}$$

Since $\beta^k(x, x') = \hat{p}^k(x) \wedge \hat{p}(x') \leq 1$, we deduce from (3.37), (3.38) and $\mathcal{AW}_r$-optimality of $\hat{\rho}^k$

$$
\begin{aligned}
\mathcal{AW}_r^r(\hat\pi^k+\tilde\pi^k,\hat\pi+\tilde\pi) \le \mathcal{AW}_r^r(\hat\pi^k,\hat\pi) &+ \int_{X\times X} d_X(x,x')^r\,\tilde\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \hat p^k(x)\mathcal{W}_r^r(\hat\pi_x^k,\delta_{y_0})\,\tilde\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \hat p(x')\mathcal{W}_r^r(\hat\pi_{x'},\delta_{y_0})\,\tilde\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \alpha_+^k(x,x')\mathcal{W}_r^r(\hat\pi_x^k,\delta_{y_0})\,\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} (1-\hat p^k(x))\mathcal{W}_r^r(\tilde\pi_x^k,\delta_{y_0})\,\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \alpha_-^k(x,x')\mathcal{W}_r^r(\hat\pi_{x'},\delta_{y_0})\,\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} (1-\hat p(x'))\mathcal{W}_r^r(\tilde\pi_{x'},\delta_{y_0})\,\rho^k(dx,dx').
\end{aligned}
\tag{3.39}
$$

Recall that $\tilde\rho^k$ has marginals $\tilde\mu^k$ and $\tilde\mu$ with total mass smaller than ϵ. By (3.1) we find

$$
\int_{X\times X} d_X(x,x')^r\,\tilde\rho^k(dx,dx') \le 2^{r-1}\left(I_\varepsilon^r(\tilde\mu^k)+I_\varepsilon^r(\tilde\mu)\right). \tag{3.40}
$$

Concerning the marginals of $\hat p^k(x)\,\rho(dx,dx')$ and $\hat p(x')\,\rho(dx,dx')$, we find the relations

$$
\hat p^k(x)\,\tilde\mu^k(dx) = (1-\hat p^k(x))\,\hat\mu^k(dx), \quad \hat p(x')\,\tilde\mu(dx') = (1-\hat p(x'))\,\hat\mu(dx').
$$

Again by (3.1), we find since $\tilde\rho^k \in \Pi(\tilde\mu^k,\tilde\mu)$, $\hat\pi^k \in \Pi(\hat\mu^k,\hat\nu^k)$ and $\hat\pi \in \Pi(\hat\mu,\hat\nu)$ that

$$
\begin{aligned}
&\int_{X\times X} \hat p^k(x)\mathcal{W}_r^r(\hat\pi_x^k,\delta_{y_0})\,\tilde\rho^k(dx,dx') \\
&\quad = \int_{X\times X} (1-\hat p^k(x))\mathcal{W}_r^r(\hat\pi_x^k,\delta_{y_0})\,\hat\mu^k(dx) \le I_\epsilon^r(\hat\nu^k),
\end{aligned}
\tag{3.41}
$$

$$
\begin{aligned}
&\int_{X\times X} \hat p(x')\mathcal{W}_r^r(\hat\pi_{x'},\delta_{y_0})\,\tilde\rho^k(dx,dx') \\
&= \int_{X\times X} (1-\hat p(x'))\mathcal{W}_r^r(\hat\pi_{x'},\delta_{y_0})\,\hat\mu(dx') \le I_\epsilon^r(\hat\nu).
\end{aligned}
\tag{3.42}
$$

We deduce from (3.39) and (3.40)-(3.42) that it is sufficient to show

$$\int_{X\times X} \alpha_+^k(x,x')\mathcal{W}_r^r(\hat{\pi}_x^k,\delta_{y_0})\,\rho^k(dx,dx') \le I_\epsilon^r(\hat{\nu}^k), \tag{3.43}$$

$$\int_{X\times X} (1-\hat{p}^k(x))\mathcal{W}_r^r(\tilde{\pi}_x^k,\delta_{y_0})\,\rho^k(dx,dx') \le I_\varepsilon^r(\tilde{\nu}^k), \tag{3.44}$$

$$\int_{X\times X} \alpha_-^k(x,x')\mathcal{W}_r^r(\hat{\pi}_{x'},\delta_{y_0})\,\rho^k(dx,dx') \le I_\varepsilon^r(\hat{\nu}), \tag{3.45}$$

$$\int_{X\times X} (1-\hat{p}(x'))\mathcal{W}_r^r(\tilde{\pi}_{x'},\delta_{y_0})\,\rho^k(dx,dx') \le I_\varepsilon^r(\tilde{\nu}). \tag{3.46}$$

To see (3.44) and (3.46), note that

$$(1-\hat{p}^k(x))\,\mu^k(dx) = \tilde{\mu}^k(dx) \quad \text{and} \quad (1-\hat{p}(x'))\,\mu(dx') = \tilde{\mu}(dx'). \tag{3.47}$$

As a consequence, the first marginal of $(1-\hat{p}^k(x))\,\rho^k(dx,dx')$ is $\tilde{\mu}^k$, whereas the second marginal of $(1-\hat{p}(x'))\,\rho^k(dx,dx')$ coincides with $\tilde{\mu}$. Hence, as the mass of $\tilde{\mu}^k$ and $\tilde{\mu}$ does not exceed ϵ, we have

$$\begin{aligned}
&\int_{X\times X} (1-\hat{p}^k(x))\mathcal{W}_r^r(\tilde{\pi}_x^k,\delta_{y_0})\,\rho^k(dx,dx') \\
&\quad = \int_X \mathcal{W}_r^r(\tilde{\pi}_x^k,\delta_{y_0})\,\tilde{\mu}^k(dx) = \mathcal{W}_r^r(\tilde{\nu}^k,\delta_{y_0}) = I_\epsilon^r(\tilde{\nu}^k), \\
&\int_{X\times X} (1-\hat{p}(x'))\mathcal{W}_r^r(\tilde{\pi}_{x'},\delta_{y_0})\,\rho^k(dx,dx') \\
&\quad = \int_X \mathcal{W}_r^r(\tilde{\pi}_{x'},\delta_{y_0})\,\tilde{\mu}(dx') = \mathcal{W}_r^r(\tilde{\nu},\delta_{y_0}) = I_\epsilon^r(\tilde{\nu}).
\end{aligned}$$

Next, we show (3.43) and (3.45). To this end, denoting $\rho^k(dx,dx') = \mu^k(dx)\,\rho_x^k(dx') = \mu(dx')\,\overleftarrow{\rho}_{x'}^k(dx)$, we have

$$\begin{aligned}
&\alpha_+^k(x,x')\,\rho^k(dx,dx') \le \hat{p}^k(x)\,\rho^k(dx,dx') \\
&\quad = \frac{d\hat{\mu}^k}{d\mu^k}(x)\,\mu^k(dx)\,\rho_x^k(dx') = \hat{\mu}^k(dx)\,\rho_x^k(dx'), \\
&\alpha_-^k(x,x')\,\rho^k(dx,dx') \le \hat{p}(x')\,\rho^k(dx,dx') \\
&\quad = \frac{d\hat{\mu}}{d\mu}(x')\,\mu(dx')\,\overleftarrow{\rho}_{x'}^k(dx) = \hat{\mu}(dx')\,\overleftarrow{\rho}_{x'}^k(dx).
\end{aligned}$$

In particular, the first marginal of $\alpha_+^k(x,x')\,\rho^k(dx,dx')$, denoted here by τ^k, is dominated by $\hat{\mu}^k$, whereas the second marginal of $\alpha_-^k(x,x')\,\rho^k(dx,dx')$, denoted here by $\tau^{k'}$, is dominated by $\hat{\mu}$. Concerning the masses of τ^k and $\tau^{k'}$, remember (3.47),

$\alpha_+^k(x,x') \le 1 - \hat{p}(x')$ and $\alpha_-^k(x,x') \le 1 - \hat{p}^k(x)$, thus,

$$\tau^k(X) = \int_{X\times X} \alpha_+^k(x,x')\,\rho^k(dx,dx') \le \int_X (1-\hat{p}(x'))\,\mu(dx') = \tilde{\mu}(X) \le \epsilon,$$
$$\tau^{k\prime}(X) = \int_{X\times X} \alpha_-^k(x,x')\,\rho^k(dx,dx') \le \int_X (1-\hat{p}^k(x))\,\mu^k(dx) = \tilde{\mu}^k(X) \le \epsilon.$$

Using (3.1), we conclude with

$$\int_{X\times X} \alpha_+^k(x,x')\mathcal{W}_r^r(\hat{\pi}_x^k,\delta_{y_0})\,\rho^k(dx,dx') = \int_X \mathcal{W}_r^r(\hat{\pi}_x^k,\delta_{y_0})\,\tau(dx) \le I_\epsilon^r(\hat{\nu}^k),$$
$$\int_{X\times X} \alpha_-^k(x,x')\mathcal{W}_r^r(\hat{\pi}_{x'},\delta_{y_0})\,\rho^k(dx,dx') = \int_X \mathcal{W}_r^r(\hat{\pi}_{x'},\delta_{y_0})\,\tau'(dx') \le I_\epsilon^r(\hat{\nu}).$$

□

The addition on $\mathcal{M}_r(X\times Y)$ is continuous with respect to the adapted weak topology provided the limits have singular first marginal distributions. We recall that two positive measures μ,ν are called singular if and only if there exists a measurable set $A\subset X$ such that $\mu(A^{\complement}) = 0 = \nu(A)$.

Lemma 3.7 *Let $\pi,\chi\in\mathcal{M}_r(X\times Y)$ be such that their respective first marginals are singular. Let $\pi^k,\chi^k\in\mathcal{M}_r(X\times Y)$, $k\in\mathbb{N}$ converge to π and χ respectively in $\mathcal{AW}_r$. Then*

$$\pi^k+\chi^k \underset{k\to+\infty}{\longrightarrow} \pi+\chi \quad \text{in } \mathcal{AW}_r.$$

Proof Let μ_1,μ_2,μ_1^k and μ_2^k denote the respective first marginals of π,χ,π^k and χ^k. Due to singularity, there is a measurable set $A\subset X$ such that $\mu_1(A^{\complement}) = 0 = \mu_2(A)$.

Suppose first that for all $k\in\mathbb{N}$, $\mu_1^k(A^{\complement}) = 0 = \mu_2^k(A)$. Let $\rho_1^k\in\Pi(\mu_1^k,\mu_1)$, resp. $\rho_2^k\in\Pi(\mu_2^k,\mu_2)$, be an optimal coupling for $\mathcal{AW}_r(\pi^k,\pi)$, resp. $\mathcal{AW}_r(\chi^k,\chi)$. Since almost surely

$$(\pi^k+\chi^k)_x = \mathbb{1}_A(x)\,\pi_x^k + \mathbb{1}_{A^{\complement}}(x)\chi_x^k \quad\text{and}\quad (\pi+\chi)_x = \mathbb{1}_A(x)\,\pi_x + \mathbb{1}_{A^{\complement}}(x)\chi_x,$$

we have

$$\begin{aligned}
&\mathcal{AW}_r^r(\pi^k+\chi^k,\pi+\chi)\\
&\quad\le \int_{X\times X}\left(d_X^r(x,x') + \mathcal{W}_r^r((\pi^k+\chi^k)_x,(\pi+\chi)_{x'})\right)\,(\rho_1^k+\rho_2^k)(dx,dx')\\
&\quad= \int_{X\times X}\left(d_X^r(x,x') + \mathcal{W}_r^r(\pi_x^k,\pi_{x'})\right)\,\rho_1^k(dx,dx')\\
&\qquad+ \int_{X\times X}\left(d_X^r(x,x') + \mathcal{W}_r^r(\chi_x^k,\chi_{x'})\right)\,\rho_2^k(dx,dx')\\
&\quad= \mathcal{AW}_r^r(\pi^k,\pi) + \mathcal{AW}_r^r(\chi^k,\chi) \to 0, \quad k\to+\infty.
\end{aligned}$$

Let us now go back to the general case. Let $\varepsilon > 0$. Since X is a Polish space, μ_1 and μ_2 are inner regular, so there exist two compact sets $K_1 \subset A$ and $K_2 \subset A^{\complement}$ such that

$$\mu_1(K_1^{\complement}) < \varepsilon \quad \text{and} \quad \mu_2(K_2^{\complement}) < \varepsilon.$$

Since X is metrizable, it is normal, hence we can separate the closed, disjoint sets K_1 and K_2 by open, disjoint sets $\tilde{K}_1$ and $\tilde{K}_2$ where $K_1 \subset \tilde{K}_1$ and $K_2 \subset \tilde{K}_2$. Then Lemma 3.4 (i) provides sequences $(\tilde{\mu}_1^k \times \pi_x^k)_{k\in\mathbb{N}}$ and $(\tilde{\mu}_2^k \times \chi_x^k)_{k\in\mathbb{N}}$ with values in $\mathcal{M}(X \times Y)$ and null sequences $(\varepsilon_k)_{k\in\mathbb{N}}$ and $(\eta_k)_{k\in\mathbb{N}}$ with values in $[0, 1]$, such that $\tilde{\mu}_1^k \le \mu_1^k|_{\tilde{K}_1}$, $\tilde{\mu}_2^k \le \mu_2^k|_{\tilde{K}_2}$ and, for $k \to +\infty$,

$$\begin{aligned}&\mathcal{AW}_r^r(\tilde{\mu}_1^k \times \pi_x^k, (1-\varepsilon_k)\pi|_{K_1\times Y})\\ &\quad + \mathcal{AW}_r^r(\tilde{\mu}_2^k \times \chi_x^k, (1-\eta_k)\chi|_{K_2\times Y}) \to 0.\end{aligned}$$

To apply Lemma 3.6 (b), let $0 < \epsilon' \le \epsilon$ be such that $\epsilon'(\mu_1(K_1) + \mu_2(K_2)) < \epsilon$. Let k be sufficiently large such that $\epsilon^k \wedge \eta^k < \epsilon'$. We consider the sequences

$$\begin{aligned}&\hat{\pi}^k = \frac{1-\epsilon'}{1-\epsilon^k}\tilde{\mu}_1^k \times \pi_x^k + \frac{1-\epsilon'}{1-\eta^k}\tilde{\mu}_2^k \times \chi_x^k, \quad \hat{\pi} = (1-\epsilon')\left(\pi|_{K_1\times Y} + \chi|_{K_2\times Y}\right),\\ &\tilde{\pi}^k = \pi^k + \chi^k - \hat{\pi}^k, \quad \tilde{\pi} = \pi + \chi - \hat{\pi},\end{aligned}$$

where $\tilde{\pi}^k$ is well-defined in $\mathcal{M}_r(X \times Y)$ since $\epsilon^k < \epsilon'$ and $\eta^k < \epsilon'$. Note that as $k \to +\infty$,

$$\begin{aligned}&\mathcal{AW}_r^r\left(\frac{1-\epsilon'}{1-\epsilon^k}\tilde{\mu}_1^k \times \pi_x^k, (1-\epsilon')\pi|_{K_1\times Y}\right)\\ &\qquad\qquad = \frac{1-\epsilon'}{1-\epsilon^k}\mathcal{AW}_r^r\left(\tilde{\mu}_1^k \times \pi_x^k, (1-\epsilon^k)\pi|_{K_1\times Y}\right) \to 0,\\ &\mathcal{AW}_r^r\left(\frac{1-\epsilon'}{1-\eta^k}\tilde{\mu}_2^k \times \chi_x^k, (1-\epsilon')\chi|_{K_2\times Y}\right)\\ &\qquad\qquad = \frac{1-\epsilon'}{1-\eta^k}\mathcal{AW}_r^r\left(\tilde{\mu}_2^k \times \chi_x^k, (1-\eta^k)\chi|_{K_2\times Y}\right) \to 0.\end{aligned}$$

Since the first marginal distributions of $\tilde{\mu}_1^k \times \pi_x^k$ and $(1-\varepsilon_k)\pi|_{K_1\times Y}$, resp. $\tilde{\mu}_2^k \times \chi_x^k$ and $(1-\eta_k)\chi|_{K_2\times Y}$, are concentrated on $\tilde{K}_1$, resp. $\tilde{K}_2$, and since $\tilde{K}_1$ and $\tilde{K}_2$ are disjoint, we have according to the preceding part that

$$\mathcal{AW}_r^r(\hat{\pi}^k, \hat{\pi}) \to 0, \quad k \to +\infty.$$

Due to $\mathcal{AW}_r$-convergence of $(\pi^k)_{k\in\mathbb{N}}$ and $(\chi^k)_{k\in\mathbb{N}}$, we obtain $\mathcal{W}_r$-convergence of the marginals of $\pi^k + \chi^k$ to the marginals of $\pi + \chi$. Furthermore, we have

$$\tilde{\pi}^k(X \times Y) = \tilde{\pi}(X \times Y) \le \mu_1(K_1^c) + \mu_2(K_2^c) + \epsilon'(\mu_1(K_1) + \mu_2(K_2)) < 3\epsilon.$$

Then (3.36) yields

$$\limsup_{k\to+\infty} \mathcal{AW}_r^r(\pi^k+\chi^k,\pi+\chi) = \limsup_{k\to+\infty} \mathcal{AW}_r^r(\hat\pi^k+\tilde\pi^k,\hat\pi+\tilde\pi)$$
$$\le C\left(I_{3\varepsilon}^r(\mu_1+\mu_2)+I_{3\varepsilon}^r(\nu_1+\nu_2)\right),$$

where ν_1 and ν_2 denote the respective second marginals of π and χ, and the constant C only depends on r. Therefore, the right-hand side vanishes as $\epsilon \to 0$ according to Lemma 3.1 (b), which concludes the proof. □

4 Auxiliary results on the convex order in dimension one

We recall that the convex order on $\mathcal{M}_1(\mathbb{R})$ is defined by

$$\mu \le_c \nu \iff \quad \forall f : \mathbb{R} \to \mathbb{R} \text{ convex}, \quad \mu(f) \le \nu(f).$$

The following assertions can be found for instance be found in [35, Section 2]: for all $(m_0, m_1) \in \mathbb{R}_+^* \times \mathbb{R}$, there is a one-to-one correspondence between finite positive measures $\mu \in \mathcal{M}_1(\mathbb{R})$ with mass m_0 such that $\int_\mathbb{R} y\,\mu(dy) = m_1$ and the set of functions $u : \mathbb{R} \to \mathbb{R}^+$ which satisfy

(i) u is convex;
(ii) $u(y) - m_0|y - m_1|$ goes to 0 as $|y|$ tends to $+\infty$.

Any function which satisfies (i) and (ii) is then called a potential function. As noted above, the potential function of μ is denoted by

$$u_\mu(y) = \int_\mathbb{R} |y - x|\,\mu(dx).$$

Potential functions can of course also be considered in greater generality than on the real line, but this is not relevant for our purposes.

A sequence $(\mu^k)_{k\in\mathbb{N}}$ of finite positive measures with equal masses on the line converges in $\mathcal{W}_1$ to μ if and only if the sequence of potential functions $(u_{\mu^k})_{k\in\mathbb{N}}$ converges pointwise to u_μ. In that case, since for all $y \in \mathbb{R}$ the map $x \mapsto |y - x|$ is Lipschitz continuous with constant 1, we have by Kantorovich and Rubinstein's duality theorem that

$$\sup_{y\in\mathbb{R}} |u_{\mu^k}(y) - u_\mu(y)| \le \mathcal{W}_1(\mu^k,\mu) \to 0, \quad k \to +\infty,$$

hence we even have uniform convergence on $\mathbb{R}$ of potential functions.

For all $m_1 \in \mathbb{R}$, the set of all finite positive measures on the real line with mean m_1 is a lattice [40, Proposition 1.6], and even a complete lattice [41] for the convex order. Then all $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$ with mean m_1 have a supremum, denoted $\mu \vee_c \nu$, and an infimum, denoted $\mu \wedge_c \nu$, with respect to the convex order. In that context it is convenient to work with potential functions since they provide simple characterisations of those bounds:

$$\mu \vee_c \nu \text{ is defined as the measure with potential function } u_\mu \vee u_\nu,$$
$$\mu \wedge_c \nu \text{ is defined as the measure with potential function } \operatorname{co}(u_\mu \wedge u_\nu),$$

where co is the convex hull.

Lemma 4.1 *Let $(\mu^k)_{k\in\mathbb{N}}$, $(\nu^k)_{k\in\mathbb{N}}$ be two sequences of $\mathcal{M}_1(\mathbb{R})$ converging respectively to μ and ν in $\mathcal{W}_1$. Suppose that there exists $(m_0, m_1) \in \mathbb{R}_+^* \times \mathbb{R}$ such that $\mu^k(\mathbb{R}) = \nu^k(\mathbb{R}) = m_0$ and $\int_\mathbb{R} x\, \mu^k(dx) = \int_\mathbb{R} y\, \nu^k(dy) = m_1$ for all $k \in \mathbb{N}$. Then*

$$\lim_{k\to+\infty} \mathcal{W}_1(\mu^k \vee_c \nu^k, \mu \vee_c \nu) = 0 \quad \text{and} \quad \lim_{k\to+\infty} \mathcal{W}_1(\mu^k \wedge_c \nu^k, \mu \wedge_c \nu) = 0.$$

Proof Convergence in $\mathcal{W}_1$ is equivalent to pointwise convergence of the potential functions. Thus, the convergence of $\mu^k \vee_c \nu^k$ to $\mu \vee_c \nu$ in $\mathcal{W}_1$ is a consequence of the pointwise convergence of $u_{\mu^k \vee_c \nu^k} = u_{\mu^k} \vee u_{\nu^k}$ to $u_\mu \vee u_\nu = u_{\mu \vee_c \nu}$.

To show convergence of $\mu^k \wedge_c \nu^k$ to $\mu \wedge_c \nu$ in $\mathcal{W}_1$, it is sufficient to show for all $x \in \mathbb{R}$

$$u_{\mu^k \wedge_c \nu^k}(x) = \operatorname{co}(u_{\mu^k} \wedge u_{\nu^k})(x) \to \operatorname{co}(u_\mu \wedge u_\nu)(x) = u_{\mu \wedge_c \nu}(x), \quad k \to +\infty. \tag{4.1}$$

Since u_{μ^k} and u_{ν^k} converge uniformly on $\mathbb{R}$ to u_μ and u_ν respectively, we have uniform convergence of $u_{\mu^k} \wedge u_{\nu^k}$ to $u_\mu \wedge u_\nu$. Let $\varepsilon > 0$ and $k_0 \in \mathbb{N}$ be such that for all $k \geq k_0$,

$$\sup_{x\in\mathbb{R}} |(u_{\mu^k} \wedge u_{\nu^k})(x) - (u_\mu \wedge u_\nu)(x)| \leq \varepsilon.$$

For all $k \geq k_0$, we find

$$\operatorname{co}(u_\mu \wedge u_\nu) - \varepsilon \leq (u_\mu \wedge u_\nu) - \epsilon \leq u_{\mu^k} \wedge u_{\nu^k},$$
$$\operatorname{co}(u_{\mu^k} \wedge u_{\nu^k}) - \varepsilon \leq (u_{\mu^k} \wedge u_{\nu^k}) - \epsilon \leq u_\mu \wedge u_\nu.$$

Thus, as the convex hull is the supremum over all dominated, convex functions, this yields

$$\operatorname{co}(u_\mu \wedge u_\nu) - \varepsilon \leq \operatorname{co}(u_{\mu^k} \wedge u_{\nu^k}) \leq \operatorname{co}(u_\mu \wedge u_\nu) + \varepsilon,$$

which establishes (4.1) and completes the proof. □

We now provide the proof of Proposition 2.5 which is the key argument to see that it is enough to prove our main result, namely Theorem 2.6, for irreducible pairs of marginals.

Proof of Proposition 2.5 To construct the desired decomposition, pick for all $k \in \mathbb{N}$ a coupling $\pi^k \in \Pi_M(\mu^k, \nu^k)$. Let l_n and r_n denote the left and right boundary of the open interval $\{u_{\mu_n} < u_{\nu_n}\}$ on which μ_n is concentrated, and set

$$\mu_n^k(dx) = \int_{u=F_\mu(l_n)}^{F_\mu(r_n-)} \delta_{F_{\mu^k}^{-1}(u)}(dx)\, du, \quad \nu_n^k(dy) = \int_{u=F_\mu(l_n)}^{F_\mu(r_n-)} \pi^k_{F_{\mu^k}^{-1}(u)}(dy)\, du.$$

These are the respective marginals of $\tilde{\pi}^{k,n}$ on $\mathbb{R}^2$ given by

$$\tilde{\pi}^{k,n}(dx, dy) = \int_{u=F_\mu(l_n)}^{F_\mu(r_n-)} \delta_{F_{\mu^k}^{-1}(u)}(dx)\, \pi^k_{F_{\mu^k}^{-1}(u)}(dy)\, du. \tag{4.2}$$

Since π^k is a martingale coupling, we have $\mu_n^k \leq_c \nu_n^k$. Finally define

$$J = [0, 1]\setminus \bigcup_{n\in N} (F_\mu(l_n), F_\mu(r_n-)),$$

and set

$$\eta^k(dx) = \int_{u\in J} \delta_{F_{\mu^k}^{-1}(u)}(dx)\, du, \quad \upsilon^k(dy) = \int_{u\in J} \pi^k_{F_{\mu^k}^{-1}(u)}(dy)\, du.$$

These are the respective marginals of $\tilde{\pi}^k$ defined by

$$\tilde{\pi}^k(dx, dy) = \int_{u\in J} \delta_{F_{\mu^k}^{-1}(u)}(dx)\, \pi^k_{F_{\mu^k}^{-1}(u)}(dy)\, du,$$

which is again a martingale coupling with marginals (η^k, υ^k), thus, $\eta^k \leq_c \upsilon^k$.

Using inverse transform sampling for the second equality, we find

$$\begin{aligned}\left(\tilde{\pi}^k + \sum_{n\in N} \tilde{\pi}^{k,n}\right)(dx, dy) &= \int_{u=0}^{1} \delta_{F_{\mu^k}^{-1}(u)}(dx)\, \pi^k_{F_{\mu^k}^{-1}(u)}(dy)\, du \\ &= \int_{x^k\in\mathbb{R}} \delta_{x^k}(dx)\, \pi^k_{x^k}(dy)\, \mu^k(dx^k) \\ &= \mu^k(dx)\, \pi^k_x(dy) = \pi^k(dx, dy).\end{aligned}$$

Concerning the marginals, we deduce

$$\eta^k + \sum_{n\in N} \mu_n^k = \mu^k \quad \text{and} \quad \upsilon^k + \sum_{n\in N} \nu_n^k = \nu^k.$$

For all $(\tau, u, l, r) \in \mathcal{P}_1(\mathbb{R}) \times (0, 1) \times \mathbb{R} \times \mathbb{R}$, we have by (2.3):

$$F_\tau(l) < u < F_\tau(r-) \implies l < F_\tau^{-1}(u) < r \implies F_\tau(l) < u \leq F_\tau(r-). \tag{4.3}$$

Since $\mu_n(dx) = \mathbb{1}_{(l_n,r_n)}(x)\,\mu(dx)$, using (4.3) for the second equality we find

$$\mu_n(dx) = \int_{x' \in (l_n,r_n)} \delta_{x'}(dx)\,\mu(dx) = \int_{u=F_\mu(l_n)}^{F_\mu(r_n-)} \delta_{F_\mu^{-1}(u)}(dx)\,du.$$

We deduce that

$$\begin{aligned}\eta(dx) &= \left(\mu - \sum_{n\in N}\mu_n\right)(dx)\\ &= \int_{u=0}^{1} \delta_{F_\mu^{-1}(u)}(dx)\,du - \sum_{n\in N}\int_{u=F_\mu(l_n)}^{F_\mu(r_n-)} \delta_{F_\mu^{-1}(u)}(dx)\,du\\ &= \int_{u\in J} \delta_{F_\mu^{-1}(u)}(dx)\,du.\end{aligned}$$

Since the monotone rearrangement yields an optimal coupling, we have

$$\mathcal{W}_1(\eta^k,\eta) + \sum_{n\in N}\mathcal{W}_1(\mu_n^k,\mu_n) = \int_0^1 |F_{\mu^k}^{-1}(u) - F_\mu^{-1}(u)|\,du = \mathcal{W}_1(\mu^k,\mu),$$

hence

$$\lim_{k\to+\infty}\mathcal{W}_1(\eta^k,\eta) = 0 = \lim_{k\to+\infty}\mathcal{W}_1(\mu_n^k,\mu_n), \quad \forall n \in N.$$

Since the marginals of π^k converge weakly, the sequences $(\mu^k)_{k\in\mathbb{N}}$ and $(\nu^k)_{k\in\mathbb{N}}$ are tight, and so is $(\pi^k)_{k\in\mathbb{N}}$. For $n \in N$, $\tilde{\pi}^{k,n}$ is dominated by π^k, hence $(\tilde{\pi}^{k,n})_{k\in\mathbb{N}}$ is tight and therefore relatively compact. Moreover, by $\mathcal{W}_1$-convergence of $(\mu^k)_{k\in\mathbb{N}}$ and $(\nu^k)_{k\in\mathbb{N}}$, the sequences $\left(\int_{\mathbb{R}}|x|\,\mu^k(dx)\right)_{k\in\mathbb{N}}$ and $\left(\int_{\mathbb{R}}|y|\,\nu^k(dy)\right)_{k\in\mathbb{N}}$ converge and are in particular bounded. Hence the sequences $\left(\int_{\mathbb{R}}|x|\,\mu_n^k(dx)\right)_{k\in\mathbb{N}}$ and $\left(\int_{\mathbb{R}}|y|\,\nu_n^k(dy)\right)_{k\in\mathbb{N}}$ are bounded as well and admit convergent subsequences. Since the $\mathcal{W}_1$-convergence is equivalent to the weak convergence plus convergence of the first moments, we deduce that the sequence $(\tilde{\pi}^{k,n})_{k\in\mathbb{N}}$ is relatively compact in $\mathcal{W}_1$. Since $(\pi^k)_{k\in\mathbb{N}}$ is tight, from any subsequence we can extract a further subsequence denoted by $(\pi^{k_j})_{j\in\mathbb{N}}$ which converges weakly to some $\pi \in \Pi_M(\mu,\nu)$. There are subsequences $(\tilde{\pi}^{k_j,n})_{j\in\mathbb{N}}$ converging in $\mathcal{W}_1$ to a measure $\tilde{\pi}_n$. Moreover $\tilde{\pi}_n \le \pi$ with $\pi \in \Pi_M(\mu,\nu)$ denoting the weak limit of a subsequence of the tight sequence $(\pi^{k_j})_{j\in\mathbb{N}}$. The first marginal of $\tilde{\pi}_n$ coincides with μ_n due to the continuity of the projection, thus,

$$\tilde{\pi}_n \le \pi|_{(l_n,r_n)\times\mathbb{R}} =: \pi_n.$$

As $\tilde{\pi}_n(\mathbb{R}\times\mathbb{R}) = \mu_n((l_n,r_n)) = \pi_n(\mathbb{R}\times\mathbb{R})$, there must hold equality, i.e., $\tilde{\pi}_n = \pi_n$ and $\int_{x\in\mathbb{R}}\tilde{\pi}_n(dx,dy) = \nu_n(dy)$. By continuity of the projection, we deduce that $\lim_{j\to\infty}\mathcal{W}_1(\nu_n^{k_j},\nu_n) = 0$ and, since the limit does not depend on the subsequence, $(\nu_n^k)_{k\in\mathbb{N}}$ converges in $\mathcal{W}_1$ to ν_n. Analogously, we find that $(\upsilon^k)_{k\in\mathbb{N}}$ converges to η. □

The next two lemmas explore the influence of certain scaling and restrictions of measure on condition that the transformed measures are in convex order.

Lemma 4.2 *Let $r \geq 1$ and $\mu \in \mathcal{M}_r(\mathbb{R}^d)$ be a finite positive measure. Let $m_1 = \int_{\mathbb{R}} x\,\mu(dx)$ and μ^α, $\alpha \in \mathbb{R}_+$ be the image of μ by $y \mapsto \alpha(y - m_1) + m_1$. Then for all $\alpha, \beta \in \mathbb{R}_+$,*

$$\mathcal{W}_r(\mu^\alpha, \mu^\beta) = |\beta - \alpha| \left(\int_{\mathbb{R}^d} |x - m_1|^r \, \mu(dx) \right)^{\frac{1}{r}} = |\beta - \alpha| \mathcal{W}_r(\mu^0, \mu^1). \quad (4.4)$$

Moreover, $(\mu^\alpha)_{\alpha \in \mathbb{R}^+}$ constitutes a peacock, i.e., $\alpha \leq \beta \in \mathbb{R}_+$ implies $\mu^\alpha \leq_c \mu^\beta$.

Proof Let $\alpha \leq \beta \in \mathbb{R}_+$. By the triangle inequality we obtain

$$\begin{aligned}\left(\int_{\mathbb{R}^d} |x - m_1|^r \, \mu^\beta(dx) \right)^{\frac{1}{r}} &= \mathcal{W}_r(\delta_{m_1}, \mu^\beta) \leq \mathcal{W}_r(\delta_{m_1}, \mu^\alpha) + \mathcal{W}_r(\mu^\alpha, \mu^\beta) \\ &= \left(\int_{\mathbb{R}^d} |x - m_1|^r \, \mu^\alpha(dx) \right)^{\frac{1}{r}} + \mathcal{W}_r(\mu^\alpha, \mu^\beta).\end{aligned}$$

Thus,

$$\begin{aligned}\mathcal{W}_r(\mu^\alpha, \mu^\beta) &\geq \left(\int_{\mathbb{R}^d} |x - m_1|^r \, \mu^\beta(dx) \right)^{\frac{1}{r}} - \left(\int_{\mathbb{R}^d} |x - m_1|^r \, \mu^\alpha(dx) \right)^{\frac{1}{r}} \\ &= (\beta - \alpha) \left(\int_{\mathbb{R}^d} |x - m_1|^r \, \mu(dx) \right)^{\frac{1}{r}}.\end{aligned}$$

Since the image of μ under $x \mapsto (\alpha(x - m_1) + m_1, \beta(y - m_1) + m_1)$ is a coupling between μ^α and μ^β, we also have the reverse inequality

$$\mathcal{W}_r(\mu^\alpha, \mu^\beta) \leq (\beta - \alpha) \left(\int_{\mathbb{R}^d} |x - m_1|^r \, \mu(dx) \right)^{\frac{1}{r}},$$

which proves (4.4).

To see that $(\mu^\alpha)_{\alpha \in \mathbb{R}^+}$ is a peacock, we fix again $\alpha \leq \beta \in \mathbb{R}_+$ and a convex function f on $\mathbb{R}^d$. By convexity, we have

$$\begin{aligned}\mu^\alpha(f) &= \int_{\mathbb{R}^d} f(\alpha(x - m_1) + m_1) \, \mu(dx) \\ &\leq \int_{\mathbb{R}^d} \left(\frac{\alpha}{\beta} f(\beta(x - m_1) + m_1) + \left(1 - \frac{\alpha}{\beta} \right) f(m_1) \right) \mu(dx) \leq \mu^\beta(f).\end{aligned}$$

□

Lemma 4.3 *For all $p \in \mathcal{P}_1(\mathbb{R})$ with barycentre $m_1 \in \mathbb{R}$ and $R \geq 0$, let p^R be defined by*

$$p^R = p \wedge_c \left(\frac{R - m_1}{2R}\,\delta_{-R} + \frac{R + m_1}{2R}\,\delta_R \right) \quad \text{if} \quad R \geq |m_1|,$$

and $p^R = \delta_{m_1}$ otherwise. Then

(a) *For all $R > 0$, $p^R \leq_c p$, and if $R \geq |m_1|$, then p^R is concentrated on $[-R, R]$.*
(b) *We have*

$$\mathcal{W}_1(p^R, p) \underset{R \to +\infty}{\longrightarrow} 0.$$

Proof Let $p \in \mathcal{P}_1(\mathbb{R})$ be with barycentre $m_1 \in \mathbb{R}$. For all $R \geq |m_1|$, let $\eta^R = \frac{R-m_1}{2R}\,\delta_{-R} + \frac{R+m_1}{2R}\,\delta_R$, so that $p^R = p \wedge_c \eta^R$. If $R < |m_1|$ then $p^R = \delta_{m_1}$ so we clearly have $p^R \leq_c p$. Else, $p^R \leq_c p$ still holds by definition of the convex infimum. Moreover, since η^R is concentrated on $[-R, R]$, so is p^R by domination in the convex order, hence (a) is proved.

To show (b), it suffices to verify pointwise convergence of the corresponding potential functions, i.e., for all $y \in \mathbb{R}$,

$$u_{p \wedge \eta^R}(y) = \mathrm{co}(u_p \wedge u_{\eta^R})(y) \to u_p(y), \quad R \to +\infty. \tag{4.5}$$

Let $\varepsilon > 0$. Since $u_p(y) - |y - m_1|$ vanishes as $|y| \to +\infty$, there exists $M > 0$ such that

$$\forall y \in \mathbb{R}, \quad |y| > M \implies u_p(y) \leq |y - m_1| + \varepsilon.$$

Let $R_0 = |m_1| + \sup_{x \in [-M,M]} u_p(x)$ and $R \geq R_0$. The map u_{η^R} is a piecewise affine function which changes slope at $-R$ and R and such that $u_{\eta^R}(y) \to +\infty$ as $|y| \to +\infty$. It therefore attains its minimum either at $-R$ where it is equal to $R + m_1$ or at R where it is equal to $R - m_1$, and this minimum is equal to $R - |m_1|$. We deduce that for all $y \in \mathbb{R}$, $u_{\eta^R}(y) \geq R - |m_1|$. Moreover, $\delta_{m_1} \leq_c \eta^R$, hence we also have $u_{\eta^R}(y) \geq |y - m_1|$ for all $y \in \mathbb{R}$. Let $y \in \mathbb{R}$. If $|y| \leq M$, then

$$u_p(y) \leq \sup_{x \in [-M,M]} u_p(x) = R_0 - |m_1| \leq R - |m_1| \leq u_{\eta^R}(y).$$

If, on the other hand, $|y| > M$, then

$$u_p(y) \leq |y - m_1| + \varepsilon \leq u_{\eta^R}(y) + \varepsilon.$$

We deduce that for all $y \in \mathbb{R}$ and $R \geq R_0$, $u_p(y) - \varepsilon \leq (u_p \wedge u_{\eta^R})(y)$. Thus, as the convex hull is the supremum over all dominated, convex functions, this yields

$$u_p - \varepsilon \leq \mathrm{co}(u_p \wedge u_{\eta^R}) \leq u_p,$$

which proves (4.5) and completes the proof. □

5 Proof of the main theorem

We consider the setting of Theorem 2.6. Before entering its technical proof, we argue that it is sufficient to consider the case $r = 1$ and that we can assume w.l.o.g. that (μ, ν) is irreducible.

When considering a sequence of couplings $(\pi^k)_{k\in\mathbb{N}}$ which converges in $\mathcal{AW}_1$ to $\pi \in \Pi(\mu, \nu)$, whose sequence of marginal distributions $(\mu^k, \nu^k)_{k\in\mathbb{N}}$ is converging in $\mathcal{W}_r$, one can deduce $\mathcal{AW}_r$-convergence for the sequence of couplings. This is due to (1.3), and $\mathcal{W}_r$-convergence being equivalent to weak convergence plus convergence of the r-moments. To see the latter, we find, when equipping $X \times \mathcal{P}_r(Y)$ with the product metric $((x, p), (x', p')) \mapsto (d_X^r(x, x') + \mathcal{W}_r^r(p, p'))^{1/r}$,

$$\int_{X\times\mathcal{P}_r(Y)} \left(d_X^r(x, x_0) + \mathcal{W}_r^r(p, \delta_{y_0})\right) J(\pi^k)(dx, dp) = \mathcal{W}_r^r(\mu^k, \delta_{x_0}) + \mathcal{W}_r^r(\nu^k, \delta_{y_0})$$
$$\underset{k\to+\infty}{\longrightarrow} \mathcal{W}_r^r(\mu, \delta_{x_0}) + \mathcal{W}_r^r(\nu, \delta_{y_0}) = \int_{X\times\mathcal{P}_r(Y)} \left(d_X^r(x, x_0) + \mathcal{W}_r^r(p, \delta_{y_0})\right) J(\pi)(dx, dp). \tag{5.1}$$

A direct consequence is the following lemma, according to which proving Theorem 2.6 for $r = 1$ is sufficient.

Lemma 5.1 *In the setting of Theorem 2.6, assume that there exists a sequence of martingale couplings $\pi^k \in \Pi_M(\mu^k, \nu^k)$, $k \in \mathbb{N}$ converging to π in $\mathcal{AW}_1$. Then this sequence also converges to π in $\mathcal{AW}_r$.*

Next, Proposition 2.5 is the key ingredient to show that it is enough to prove Theorem 2.6 when (μ, ν) is irreducible.

Lemma 5.2 *If the conclusion of Theorem 2.6 holds for $r = 1$ and for any irreducible pair of marginals (μ, ν), then it holds for $r = 1$ and for any pair (μ, ν) in the convex order.*

Proof In the setting of Theorem 2.6, fix $\pi \in \Pi_M(\mu, \nu)$. Denote by $(\mu_n, \nu_n)_{n\in N}$ the decomposition of (μ, ν) into irreducible components with

$$\mu = \eta + \sum_{n\in N} \mu_n, \quad \nu = \eta + \sum_{n\in N} \nu_n.$$

By Proposition 2.5, we can find sub-probability measures $(\eta^k, \upsilon^k)_{k\in\mathbb{N}}$, $(\mu_n^k)_{(k,n)\in\mathbb{N}\times N}$, $(\nu_n^k)_{(k,n)\in\mathbb{N}\times N}$ such that

$$\eta^k \leq_c \upsilon^k, \quad \mu_n^k \leq_c \nu_n^k \quad \forall (k, n) \in \mathbb{N} \times N,$$
$$\eta^k \to \eta, \quad \upsilon^k \to \eta, \quad \mu_n^k \to \mu_n, \quad \nu_n^k \to \nu_n \quad \text{in } \mathcal{W}_1, \quad k \to +\infty.$$

For $k \in \mathbb{N}$, let $\chi^k \in \Pi_M(\eta^k, \upsilon^k)$ be a martingale coupling between η^k and υ^k. Since the marginals both converge to η in $\mathcal{W}_1$, $(\chi^k)_{k\in\mathbb{N}}$ is tight and any accumulation point with respect to the weak topology belongs to $\Pi_M(\eta, \eta)$. Since $\chi := (\mathrm{id}, \mathrm{id})_*\eta$ is the only martingale coupling between η and itself, $(\chi^k)_{k\in\mathbb{N}}$ converges weakly to χ as k goes to $+\infty$ and even in $\mathcal{W}_1$ according to (5.1). We can show that this convergence also holds in $\mathcal{AW}_1$.

Indeed, according to Proposition 2.1, there exists a sequence $\tilde{\chi}^k \in \Pi(\eta^k, \upsilon^k)$, $k \in \mathbb{N}$, converging to χ in $\mathcal{AW}_1$. Then

$$\begin{aligned}\mathcal{AW}_1(\chi^k, \tilde{\chi}^k) &\le \int_{\mathbb{R}} \mathcal{W}_1(\chi^k_x, \tilde{\chi}^k_x)\, \eta^k(dx) \le \int_{\mathbb{R}} \left(\mathcal{W}_1(\chi^k_x, \delta_x) + \mathcal{W}_1(\delta_x, \tilde{\chi}^k_x)\right)\, \eta^k(dx)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} |x'-x|\, (\chi^k_x + \tilde{\chi}^k_x)(dx')\, \eta^k(dx) = \int_{\mathbb{R}} |x'-x|\, (\chi^k + \tilde{\chi}^k)(dx, dx').\end{aligned}$$

Since $(x, x') \mapsto |x'-x| \in \Phi_1(\mathbb{R}^2)$ and χ^k and $\tilde{\chi}^k$ converge to χ in $\mathcal{W}_1$, we deduce, using (1.1), that

$$\int_{\mathbb{R}} |x'-x|\, (\chi^k + \tilde{\chi}^k)(dx, dx') \to 2\int_{\mathbb{R}} |x'-x|\, \chi(dx, dx') = 0, \quad k \to +\infty,$$

hence,

$$\mathcal{AW}_1(\chi^k, \chi) \le \mathcal{AW}_1(\chi^k, \tilde{\chi}^k) + \mathcal{AW}_1(\tilde{\chi}^k, \chi) \to 0, \quad k \to +\infty.$$

By assumption, we can find for any $n \in N$ a sequence $(\pi^{k,n})_{k\in\mathbb{N}}$ of martingale couplings between μ^k_n and ν^k_n, $k \in \mathbb{N}$, which converges in $\mathcal{AW}_1$ to π_n as k goes to $+\infty$, where π_n denotes π restricted to the n-th irreducible component given by (2.8). By Lemma 3.7, we have for all $p \in N$ that

$$\chi^k + \sum_{n\in N,\ n\le p} \pi^{k,n} \to \chi + \sum_{n\in N,\ n\le p} \pi_n \quad \text{in } \mathcal{AW}_1, \quad k \to +\infty.$$

Moreover, the respective marginals of $\chi^k + \sum_{n\in N} \pi^{k,n}$, namely μ^k and ν^k, converge in $\mathcal{W}_1$ to the respective marginals of $\chi + \sum_{n\in N} \pi_n$, namely μ and ν. Therefore, according to Lemma 3.6 (b), there exists a constant $C > 0$ such that

$$\limsup_k \mathcal{AW}_1\left(\chi^k + \sum_{n\in N} \pi^{k,n}, \chi + \sum_{n\in N} \pi_n\right) \le C\left(I^1_{\varepsilon_p}(\mu) + I^1_{\varepsilon_p}(\nu)\right),$$

where $\varepsilon_p = \sum_{n\in N, n>p} \mu_n(\mathbb{R})$ where by convention the sum over an empty set is 0. Clearly, $(\epsilon_p)_{p\in N}$ tends to 0, thus Lemma 3.1 (b) reveals that the right-hand side vanishes as p goes to $\sup N$. This proves that $\pi^k = \chi^k + \sum_{n\in N} \pi^{k,n} \in \Pi_M(\mu^k, \nu^k)$ converges in $\mathcal{AW}_1$ to $\pi = \chi + \sum_{n\in N} \pi^k \in \Pi_M(\mu, \nu)$. □

Proof of Theorem 2.6 *Step 1.* Due to Lemma 5.1 and Lemma 5.2, we may suppose w.l.o.g. that $r = 1$ and (μ, ν) is irreducible with component $I = (\ell, \rho)$, $\ell \in \mathbb{R} \cup \{-\infty\}$, $\rho \in \mathbb{R} \cup \{+\infty\}$. Next, we define auxiliary martingale couplings close to π which will be easier to approximate in the limit. We define them with the same first marginal distribution whereby the second marginal distribution is smaller with respect to the convex order. These auxiliary couplings will satisfy two key properties: first, their second marginal distribution must be concentrated on a compact subset of I when the first marginal distribution is itself concentrated on a certain compact subset K of I. Second, it is essential that their second marginal distribution has positive mass on some two compact subsets of I on both sides of K.

Fix $\epsilon \in (0, \frac{1}{2})$. Choose a compact subset $K = [a, b]$ of I with

$$\mu(K^{\complement}) < \varepsilon. \tag{5.2}$$

Instead of directly approximating π, we initially consider the martingale coupling $\pi^{R,\alpha}$ whose definition is given below. For any $R > 0$, let $(\pi_x^R)_{x \in \mathbb{R}}$ be the probability kernel obtained by virtue of Lemma 4.3. By Lemma 4.3 (b) we have for all $x \in \mathbb{R}$ that $\pi_x^R \leq_c \pi_x$. Therefore,

$$\mathcal{W}_1(\pi_x^R, \pi_x) \leq 2 \int_{\mathbb{R}} |y| \, \pi_x(dy),$$

where the right-hand side is a μ-integrable function of x. By Lemma 4.3 (b) we find $\pi_x^R \to \pi_x$ in $\mathcal{W}_1$ as $R \to +\infty$. Let $\pi^R := \mu \times \pi_x^R$, then dominated convergence yields

$$\mathcal{AW}_1(\pi^R, \pi) \leq \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x) \, \mu(dx) \to 0, \quad R \to +\infty.$$

Denote by ν^R the second marginal of π^R. Consequently, ν^R converges to ν for the $\mathcal{W}_1$-distance and $\nu^R \leq_c \nu$ for all $R > 0$. Let $\tilde{a}$ and $\tilde{b}$ be real numbers such that $\tilde{a} \in (\ell, a)$ and $\tilde{b} \in (b, \rho)$, for instance

$$\tilde{a} = \frac{\ell + a}{2} \vee (a - 1) \text{ and } \tilde{b} = (b + 1) \wedge \frac{b + \rho}{2}.$$

Since (μ, ν) is irreducible on I, according to Remark 2.4, ν assigns positive mass to any neighbourhood in $\overline{I}$ of the endpoints ℓ and ρ of I. From now on, we use the notational convention that for all $c \in \mathbb{R} \cup \{\pm\infty\}$,

$$[-\infty, c) = \{x \in \mathbb{R} \mid x < c\}, \quad (c, +\infty] = \{x \in \mathbb{R} \mid c < x\} \quad \text{and} \quad [-\infty, +\infty] = \mathbb{R}.$$

In particular, $\overline{I} = [\ell, \rho] \subset \mathbb{R}$.

Then $[\ell, \tilde{a})$ and $(\tilde{b}, \rho]$ are relatively open on $\overline{I}$ with $\nu^R(\overline{I}) = 1 = \nu(\overline{I})$, so Portmanteau's theorem yields

$$\liminf_{R\to+\infty} \nu^R([\ell, \tilde{a})) \geq \nu([\ell, \tilde{a})) > 0, \quad \liminf_{R\to+\infty} \nu^R((\tilde{b}, \rho]) \geq \nu((\tilde{b}, \rho]) > 0.$$

Thus, we deduce that we can choose $R > 0$ such that

$$R \geq |a| \vee |b|, \quad \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x)\, \mu(dx) < \varepsilon, \quad \nu^R([\ell, \tilde{a})) > 0, \quad \text{and} \\ \nu^R((\tilde{b}, \rho]) > 0. \tag{5.3}$$

Let $\pi_x^{R,\alpha}$ be the image of π_x^R by $y \mapsto \alpha(y - x) + x$ when $\alpha \in (0, 1)$. Then $\pi^{R,\alpha} := \mu \times \pi_x^{R,\alpha}$ satisfies by Lemma 4.2

$$\begin{aligned}
&\mathcal{AW}_1(\varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}, \pi) \\
&\quad\leq \int_{\mathbb{R}} \mathcal{W}_1(\varepsilon\pi_x + (1-\varepsilon)\pi_x^{R,\alpha}, \pi_x)\, \mu(dx) \\
&\quad\leq (1-\varepsilon) \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^{R,\alpha}, \pi_x)\, \mu(dx) \\
&\quad\leq \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^{R,\alpha}, \pi_x^R)\, \mu(dx) + \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x)\, \mu(dx) \\
&\quad= (1-\alpha) \int_{\mathbb{R}} \int_{\mathbb{R}} |x - y|\, \pi_x^R(dy)\, \mu(dx) + \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x)\, \mu(dx) \\
&\quad\leq (1-\alpha) \left(\int_{\mathbb{R}} |x|\, \mu(dx) + \int_{\mathbb{R}} |y|\, \nu^R(dy) \right) + \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x)\, \mu(dx),
\end{aligned}$$

where the right-hand side converges to $\int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x)\, \mu(dx) < \varepsilon$ for $\alpha \to 1$. Note that $\frac{2R-a-\tilde{a}}{2R-2\tilde{a}}, \frac{b+\tilde{b}+2R}{2\tilde{b}+2R} \in (0, 1)$, so we can choose $\alpha \in (0, 1)$ such that

$$\mathcal{AW}_1(\varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}, \pi) < \epsilon \quad \text{and} \quad \alpha \geq \frac{2R - a - \tilde{a}}{2R - 2\tilde{a}} \vee \frac{b + \tilde{b} + 2R}{2\tilde{b} + 2R}. \tag{5.4}$$

Let L be a compact subset of I such that the interior $\mathring{L}$ of L satisfies

$$[(-R) \vee (\alpha\ell + (1-\alpha)a), R \wedge (\alpha\rho + (1-\alpha)b)] \subset \mathring{L}.$$

Because $R \geq (-a) \vee b$ and thereby $[a, b] = K \subset [-R, R]$, we have that $\mu|_K \times \pi_x^R$ is concentrated on $K \times ([-R, R] \cap \overline{I})$. Furthermore, for any $(x, y) \in K \times ([-R, R] \cap \overline{I})$, we find $\alpha y + (1-\alpha)x \in \mathring{L}$. Hence, $\mu|_K \times \pi_x^{R,\alpha}$ is concentrated on $K \times \mathring{L}$.

Denote the second marginal of $\pi^{R,\alpha}$ by $\nu^{R,\alpha}$. Since

$$(x, y) \in (\ell, R) \times [\ell, \tilde{a}) \implies \ell < (1-\alpha)x + \alpha y < R - \alpha(R - \tilde{a}) \leq \frac{a + \tilde{a}}{2},$$

we have that

$$
\begin{aligned}
&\nu^{R,\alpha}\left(\left(\ell, \frac{a+\tilde{a}}{2}\right)\right) \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\left(\ell, \frac{a+\tilde{a}}{2}\right)}(y)\, \pi^{R,\alpha}(dx, dy) = \int_{\mathbb{R}^2} \mathbb{1}_{\left(\ell, \frac{a+\tilde{a}}{2}\right)}(\alpha y + (1-\alpha)x)\, \pi^R(dx, dy) \\
&\geq \int_{\mathbb{R}^2} \mathbb{1}_{(\ell,R)\times[\ell,\tilde{a})}(x, y)\, \pi^R(dx, dy) = \int_{(\ell,R)} \pi_x^R((-\infty, \tilde{a}))\, \mu(dx).
\end{aligned}
$$

If $x \in [R, +\infty)$, then $\pi_x^R = \delta_x$ and since $R \geq \tilde{a}$, $\pi_x^R((-\infty, \tilde{a})) = 0$. Added to the fact that μ is concentrated on I, we obtain

$$
\begin{aligned}
\int_{(\ell,R)} \pi_x^R((-\infty, \tilde{a}))\, \mu(dx) &= \int_{\mathbb{R}} \pi_x^R((-\infty, \tilde{a}))\, \mu(dx) \\
&= \nu^R((-\infty, \tilde{a})) = \nu^R([\ell, \tilde{a})) > 0.
\end{aligned}
$$

We deduce that

$$
\nu^{R,\alpha}\left(\left(\ell, \frac{a+\tilde{a}}{2}\right)\right) > 0, \text{ and similarly, } \nu^{R,\alpha}\left(\left(\frac{b+\tilde{b}}{2}, \rho\right)\right) > 0. \tag{5.5}
$$

To summarise, we have constructed a martingale coupling $\pi^{R,\alpha} \in \Pi_M(\mu, \nu^{R,\alpha})$ close to π with respect to the $\mathcal{AW}_1$-distance in view of (5.4), whose restriction $\pi^{R,\alpha}|_{K\times\mathbb{R}}$ is compactly supported on $K \times L$ and concentrated on $K \times \mathring{L}$. Moreover, the second marginal distribution $\nu^{R,\alpha}$ is dominated by ν in the convex order and assigns positive mass on both sides of K according to (5.5).

Step 2. In the next step we construct a sequence of sub-probability martingale couplings supported on a compact cube $J \times J$ $(K \subset J \subset I)$ converging to $\pi^{R,\alpha}|_{K\times\mathbb{R}}$.

Our first goal is to find a sequence $\nu^{R,\alpha,k}$, $k \in \mathbb{N}$, such that $\mu^k \leq_c \nu^{R,\alpha,k} \leq_c \nu^k$ and

$$
\mathcal{W}_1(\nu^{R,\alpha,k}, \nu^{R,\alpha}) \to 0, \quad k \to \infty. \tag{5.6}
$$

Defining $\nu^{R,\alpha,k}$ by

$$
\nu^{R,\alpha,k} = \nu^k \wedge_c (\mu^k \vee_c T_k(\nu^{R,\alpha})),
$$

where T_k denotes the translation by the difference between the common barycentre of μ^k and ν^k and the common barycentre of ν and $\nu^{R,\alpha}$, i.e., $\int_{\mathbb{R}} y\, \nu^k(dy) - \int_{\mathbb{R}} y\, \nu^{R,\alpha}(dy)$, fulfils these requirements. Indeed

$$
\mathcal{W}_1(T_k(\nu^{R,\alpha}), \nu^{R,\alpha}) = \left| \int_{\mathbb{R}} y\, \nu^k(dy) - \int_{\mathbb{R}} y\, \nu(dy) \right| \leq \mathcal{W}_1(\nu^k, \nu) \to 0,
$$

as k goes to $+\infty$. Then Lemma 4.1 provides $\nu^{R,\alpha,k} \to \nu \wedge_c (\mu \vee_c \nu^{R,\alpha}) = \nu^{R,\alpha}$ in $\mathcal{W}_1$ as k goes to $+\infty$. By Proposition 2.1 we can approximate $\pi^{R,\alpha}$ with couplings $\pi^{R,\alpha,k} \in \Pi(\mu^k, \nu^{R,\alpha,k})$ in $\mathcal{AW}_1$. Unfortunately the sequence $\pi^{R,\alpha,k}$, $k \in \mathbb{N}$ does not have to consist of solely martingale couplings. Therefore, we have to adjust the barycentres of its disintegrations, $(\pi_x^{R,\alpha,k})$ to obtain martingale kernels and thereby martingale couplings. Due to (5.5) and inner regularity of $\nu^{R,\alpha}$, we find compact sets

$$L_- \subset \left(\ell, \frac{a+\tilde{a}}{2}\right), \quad L_+ \subset \left(\frac{b+\tilde{b}}{2}, \rho\right)$$

with $\nu^{R,\alpha}$-positive measure. Let $\tilde{\ell}, \tilde{\rho} \in I$, be such that $\tilde{\ell} < \inf(L \cup L-)$ and $\sup(L \cup L_+) < \tilde{\rho}$. Then define

$$\tilde{L}_- = \left(\tilde{\ell}, \frac{a+\tilde{a}}{2}\right), \quad \tilde{L}_+ = \left(\frac{b+\tilde{b}}{2}, \tilde{\rho}\right) \quad \text{and} \quad \tilde{K} = \left(\frac{3a+\tilde{a}}{4}, \frac{3b+\tilde{b}}{4}\right), \tag{5.7}$$

so that $\tilde{L}_-$, $\tilde{L}_+$ and $\tilde{K}$ are bounded and open intervals covering respectively L_-, L_+ and K and such that the distance e between $\tilde{L}_- \cup \tilde{L}_+$ and $\tilde{K}$ is positive:

$$e = \inf\left\{|x-y| \mid (x,y) \in (\tilde{L}_- \cup \tilde{L}_+) \times \tilde{K}\right\} \geq \frac{a-\tilde{a}}{4} \wedge \frac{\tilde{b}-b}{4} > 0.$$

Denoting $J = [\tilde{\ell}, \tilde{\rho}]$, Fig. 2 summarises the construction.

The respective restrictions of $\nu^{R,\alpha,k}$ to $\tilde{L}_-$ and $\tilde{L}_+$ are denoted by ν_-^k and ν_+^k, respectively. Since $\tilde{L}_-$ and $\tilde{L}_+$ are open, Portmanteau's theorem ensures that eventually (for k sufficiently large) ν_-^k and ν_+^k each have more total mass than some constant $\delta > 0$.

By Lemma 3.4 (ii) there are $\hat{\mu}^k \leq \mu^k$, $\hat{\nu}^k \leq \nu^{R,\alpha,k}$, $\hat{\pi}^k = \hat{\mu}^k \times \hat{\pi}_x^k \in \Pi(\hat{\mu}^k, \hat{\nu}^k)$ concentrated on $\tilde{K} \times \mathring{L}$, and $\epsilon_k \geq 0$ such that

$$\mathcal{AW}_1(\hat{\pi}^k, (1-\epsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}) + \epsilon_k \to 0, \quad k \to +\infty. \tag{5.8}$$

The following procedure shows that there are for $\hat{\mu}^k(dx)$-almost every x unique constants $c_-^k(x), c_+^k(x) \in [0, +\infty)$ and $d^k(x) \in [1, +\infty)$ such that

$$\tilde{\pi}_x^k := \frac{\hat{\pi}_x^k + c_+^k(x)\nu_+^k + c_-^k(x)\nu_-^k}{d^k(x)} \in \mathcal{P}(\mathbb{R}), \quad \int_{\mathbb{R}} y\, \tilde{\pi}_x^k(dy) = x, \quad c_-^k(x) \wedge c_+^k(x) = 0.$$

Note that the constraint $c_-^k(x) \wedge c_+^k(x) = 0$ provides

$$\int_{\mathbb{R}} y\, \hat{\pi}_x^k(dy) \leq x \implies c_-^k(x) = 0, \quad \int_{\mathbb{R}} y\, \hat{\pi}_x^k(dy) \geq x \implies c_+^k(x) = 0. \tag{5.9}$$

We require $\tilde{\pi}^k_x$ to be a probability measure with mean x, thus,

$$1 + c^k_+(x)\nu^k_+(\mathbb{R}) + c^k_-(x)\nu^k_-(\mathbb{R}) = d^k(x), \tag{5.10}$$

$$\int_{\mathbb{R}} y\,\hat{\pi}^k_x(dy) + c^k_+(x)\int_{\mathbb{R}} y\,\nu^k_+(dy) + c^k_-(x)\int_{\mathbb{R}} y\,\nu^k_-(dy) = xd^k(x). \tag{5.11}$$

Combining (5.9) with (5.10) and (5.11) yields

$$c^k_-(x) = \frac{\int_{\mathbb{R}} y\,\hat{\pi}^k_x(dy) - x}{\int_{\mathbb{R}}(x-y)\,\nu^k_-(dy)} \vee 0 \in \left[0, \frac{\left|x - \int_{\mathbb{R}} y\,\hat{\pi}^k_x(dy)\right|}{e\nu^k_-(\mathbb{R})}\right],$$

$$c^k_+(x) = \frac{x - \int_{\mathbb{R}} y\,\hat{\pi}^k_x(dy)}{\int_{\mathbb{R}}(y-x)\,\nu^k_+(dy)} \vee 0 \in \left[0, \frac{\left|x - \int_{\mathbb{R}} y\,\hat{\pi}^k_x(dy)\right|}{e\nu^k_+(\mathbb{R})}\right],$$

$$d^k(x) = 1 + c^k_-(x)\nu^k_-(\mathbb{R}) + c^k_+(x)\nu^k_+(\mathbb{R}) \in \left[1, 1 + \frac{\left|x - \int_{\mathbb{R}} y\,\hat{\pi}^k_x(dy)\right|}{e}\right].$$

Remember from (5.7) that $L \cup \tilde{L}_- \cup \tilde{L}_+ \subset [\tilde{\ell}, \tilde{\rho}] \subset I$. Then we obtain for $\hat{\mu}^k(dx)$-almost every x the estimate

$$\begin{aligned}\mathcal{W}_1(\tilde{\pi}^k_x, \hat{\pi}^k_x) &\le \mathcal{W}_1\left(\frac{c^k_+(x)\nu^k_+ + c^k_-(x)\nu^k_-}{d^k(x)}, \frac{d^k(x)-1}{d^k(x)}\hat{\pi}^k_x\right) \le \frac{d^k(x)-1}{d^k(x)}|\tilde{\rho} - \tilde{\ell}| \\ &\le \frac{\left|x - \int_{\mathbb{R}} y\,\hat{\pi}^k_x(dy)\right|}{e}|\tilde{\rho} - \tilde{\ell}|.\end{aligned}$$

Hence, the adapted Wasserstein distance between $\hat{\pi}^k$ and $\tilde{\pi}^k = \hat{\mu}^k \times \tilde{\pi}^k_x$ satisfies

$$\begin{aligned}\mathcal{AW}_1(\tilde{\pi}^k, \hat{\pi}^k) &\le \int_{\mathbb{R}} \mathcal{W}_1(\tilde{\pi}^k_x, \hat{\pi}^k_x)\,\hat{\mu}^k(dx) \le \frac{|\tilde{\rho} - \tilde{\ell}|}{e}\int_{\mathbb{R}}\left|x - \int_{\mathbb{R}} y\,\hat{\pi}^k_x(dy)\right|\hat{\mu}^k(dx) \\ &\le \frac{|\tilde{\rho} - \tilde{\ell}|}{e}\mathcal{AW}_1(\hat{\pi}^k, (1-\epsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}),\end{aligned}$$

where we used Remark 2.2 with exponent $r = 1 = 2^{r-1}$ in the last inequality. The triangle inequality and (5.8) then yield

$$\lim_k \mathcal{AW}_1(\tilde{\pi}^k, (1-\epsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}) \to 0, \quad k \to \infty. \tag{5.12}$$

Next we bound the total mass which we require to fix the barycentres. We find that

$$\begin{aligned}\int_{\mathbb{R}} \frac{c^k_-(x) + c^k_+(x)}{d^k(x)}\,\hat{\mu}^k(dx) &\le \frac{1}{e(\nu^k_-(\mathbb{R}) \wedge \nu^k_+(\mathbb{R}))}\int_{\mathbb{R}}\left|x - \int_{\mathbb{R}} y\,\hat{\pi}^k_x(dy)\right|\hat{\mu}^k(dx) \\ &\le \frac{\mathcal{AW}_1(\hat{\pi}^k, (1-\epsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}})}{e(\nu^k_-(\mathbb{R}) \wedge \nu^k_+(\mathbb{R}))} \to 0, \quad k \to +\infty,\end{aligned}$$

where we used Remark 2.2 again for the last inequality and the fact that $\nu_-^k(\mathbb{R}) \wedge \nu_+^k(\mathbb{R}) \geq \delta$ for k large enough for the limit. Consequently, when $\tilde{\nu}^k$ denotes the second marginal of $\tilde{\pi}^k$, we have for k sufficiently large that

$$\begin{aligned}(1-2\epsilon)\tilde{\nu}^k &\leq (1-2\epsilon)\hat{\nu}^k + (1-2\epsilon)(\nu_-^k + \nu_+^k)\int_{\mathbb{R}} \frac{c_-^k(x) + c_+^k(x)}{d^k(x)}\,\hat{\mu}^k(dx)\\ &\leq (1-\epsilon)\nu^{R,\alpha,k}.\end{aligned}$$

Step 3. In this step, we complement the martingale coupling $(1-2\epsilon)\tilde{\pi}^k$ to a martingale coupling with marginals μ^k and $\epsilon\nu^k + (1-\epsilon)\nu^{R,\alpha,k}$ for k sufficiently large. Recall that $\tilde{\pi}^k \in \Pi_M(\hat{\mu}^k, \tilde{\nu}^k)$ and $\pi^{R,\alpha}|_{K\times\mathbb{R}} \in \Pi_M(\mu|_K, \check{\nu}^{R,\alpha})$, where $\check{\nu}^{R,\alpha}$ is the second marginal distribution of $\pi^{R,\alpha}|_{K\times\mathbb{R}}$, are concentrated on the compact cube $J \times J$ and

$$\mathcal{AW}_1(\tilde{\pi}^k, (1-\epsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}) \to 0, \quad k \to +\infty.$$

Furthermore, since $(1-\epsilon)\pi^{R,\alpha} - (1-2\epsilon)\pi^{R,\alpha}|_{K\times\mathbb{R}}$ is a martingale coupling with marginals

$$(1-\epsilon)\mu - (1-2\epsilon)\mu|_K \quad \text{and} \quad (1-\epsilon)\nu^{R,\alpha} - (1-2\epsilon)\check{\nu}^{R,\alpha},$$

we deduce by irreducibility of the pair (μ, ν) on I irreducibility of the pair of sub-probability measures

$$\varepsilon\mu + (1-\varepsilon)\mu - (1-2\varepsilon)\mu|_K \quad \text{and} \quad \varepsilon\nu + (1-\varepsilon)\nu^{R,\alpha} - (1-2\varepsilon)\check{\nu}^{R,\alpha},$$

whose potential functions satisfy

$$0 \leq u_\mu - u_{(1-2\varepsilon)\mu|_K} < u_{\varepsilon\nu+(1-\varepsilon)\nu^{R,\alpha}} - u_{(1-2\varepsilon)\check{\nu}^{R,\alpha}} \quad \text{on } I.$$

Since those potential functions are continuous, there exists $\tau > 0$ such that they have distance greater τ on J. By uniform convergence of potential functions, for $k \in \mathbb{N}$ sufficiently large we have

$$0 \leq u_{\mu^k} - u_{(1-2\varepsilon)\hat{\mu}^k} + \frac{\tau}{2} \leq u_{\varepsilon\nu^k+(1-\varepsilon)\nu^{R,\alpha,k}} - u_{(1-2\epsilon)\tilde{\nu}^k} \quad \text{on } J.$$

On the complement of J we have $u_{(1-2\varepsilon)\hat{\mu}^k} = u_{(1-2\varepsilon)\tilde{\nu}^k}$ since both measures are concentrated on J and satisfy $(1-2\varepsilon)\int_{\mathbb{R}} x\,\hat{\mu}^k(dx) = (1-2\varepsilon)\int_{\mathbb{R}} y\,\tilde{\nu}^k(dy)$. Therefore,

$$0 \leq u_{\mu^k} - u_{(1-2\varepsilon)\hat{\mu}^k} \leq u_{\varepsilon\nu^k+(1-\varepsilon)\nu^{R,\alpha,k}} - u_{(1-2\varepsilon)\tilde{\nu}^k} \quad \text{on } J^c.$$

By Strassen's theorem [52], there exists $\eta^k \in \Pi_M(\mu^k - (1-2\varepsilon)\hat{\mu}^k, \varepsilon\nu^k + (1-\varepsilon)\nu^{R,\alpha,k} - (1-2\varepsilon)\tilde{\nu}^k)$. Finally, we write

$$\overline{\pi}^k = \eta^k + (1-2\varepsilon)\tilde{\pi}^k \in \Pi_M(\mu^k, \varepsilon\nu^k + (1-\varepsilon)\nu^{R,\alpha,k}).$$

Step 4. In the last step, we show that the sequence constructed in this way is eventually close to the original martingale coupling π in adapted Wasserstein distance.

The marginals of $\overline{\pi}^k$ are converging in $\mathcal{W}_1$ to $(\mu, \epsilon\nu + (1-\epsilon)\nu^{R,\alpha})$ as k goes to $+\infty$. We have according to (5.12) that

$$\mathcal{AW}_1\left((1-2\varepsilon)\frac{1-\varepsilon}{1-\varepsilon_k}\tilde{\pi}^k, (1-2\varepsilon)(1-\varepsilon)\pi^{R,\alpha}|_{K\times\mathbb{R}}\right) \to 0, \quad k\to\infty.$$

For k large enough so that $\varepsilon_k \leq \varepsilon$,

$$\begin{aligned}
&\bar{\pi}^k(\mathbb{R}^2) - (1-2\varepsilon)\frac{1-\varepsilon}{1-\varepsilon_k}\tilde{\pi}^k(\mathbb{R}^2) \\
&\quad = \eta^k(\mathbb{R}^2) + (1-2\varepsilon)\frac{\varepsilon-\varepsilon_k}{1-\varepsilon_k}\tilde{\pi}^k(\mathbb{R}^2) \\
&\quad = \left(\varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha} - (1-2\varepsilon)(1-\varepsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}\right)(\mathbb{R}^2) \\
&\qquad + (1-2\varepsilon)(\varepsilon-\varepsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}(\mathbb{R}^2) \\
&\quad = 1-(1-2\varepsilon)(1-\varepsilon)\mu(K) \leq 4\varepsilon,
\end{aligned}$$

where we used $\mu(K) \geq 1-\varepsilon$ for the last inequality. Hence applying Lemma 3.6 (b), with $(\hat{\pi}^k, \hat{\pi}, \tilde{\pi}^k, \tilde{\pi}, \varepsilon)$ replaced by

$$\begin{aligned}
&\Bigg((1-2\varepsilon)\frac{1-\varepsilon}{1-\varepsilon_k}\tilde{\pi}^k, (1-2\varepsilon)(1-\varepsilon)\pi^{R,\alpha}|_{K\times\mathbb{R}}, \\
&\qquad \eta^k + (1-2\varepsilon)\frac{\varepsilon-\varepsilon_k}{1-\varepsilon_k}\tilde{\pi}^k, \varepsilon\pi + (1-\varepsilon)\left(\pi^{R,\alpha} - (1-2\varepsilon)\pi^{R,\alpha}|_{K\times\mathbb{R}}\right), 4\varepsilon\Bigg),
\end{aligned}$$

we obtain

$$\limsup_k \mathcal{AW}_1(\overline{\pi}^k, \varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}) \leq C(I_{4\varepsilon}(\mu) + I_{4\varepsilon}(\varepsilon\nu + (1-\varepsilon)\nu^{R,\alpha})),$$

with C given by Lemma 3.6 (b) and depending only on the exponent $r = 1$. Since $\nu^{R,\alpha} \leq_c \nu$, then $\varepsilon\nu + (1-\varepsilon)\nu^{R,\alpha} \leq_c \nu$, so using Lemma 3.1 (e), the triangle inequality and (5.4), we get

$$\begin{aligned}
&\limsup_k \mathcal{AW}_1(\overline{\pi}^k, \pi) \\
&\quad \leq \limsup_k \left(\mathcal{AW}_1(\overline{\pi}^k, \varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}) + \mathcal{AW}_1(\varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}, \pi)\right) \\
&\quad \leq C(I_{4\varepsilon}(\mu) + I_{4\varepsilon}(\nu)) + \varepsilon.
\end{aligned}$$

Since the right-hand side only depends on ϵ and vanishes as ϵ goes to 0, we can reason like in the proof of Proposition 2.3 (from (3.27)) to find a null sequence $(\tilde{\epsilon}_k)_{k\in\mathbb{N}}$, two

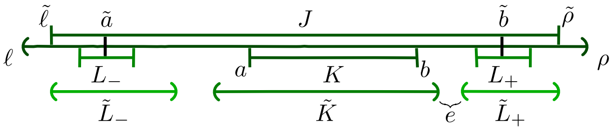

Fig. 2 Points and intervals involved in the proof. The boundaries of the closed intervals are vertical bars and those of the open intervals are parentheses

sequences $(R_k)_{k\in\mathbb{N}}$, $(\alpha_k)_{k\in\mathbb{N}}$ with values respectively in $\mathbb{R}_+^*$ and $(0,1)$, and martingale couplings

$$\mathring{\pi}^k \in \Pi_M(\mu^k, \tilde{\epsilon}_k \nu^k + (1-\tilde{\epsilon}_k)\nu^{R_k,\alpha_k,k}), \quad k \in \mathbb{N}$$

such that

$$\mathcal{AW}_1(\mathring{\pi}^k, \pi) \to 0, \quad k \to +\infty. \tag{5.13}$$

In particular, the $\mathcal{W}_1$-distance of their second marginal distributions vanishes as k goes to $+\infty$, hence the triangle inequality yields, for $k \to +\infty$,

$$\begin{aligned}&\mathcal{W}_1(\tilde{\epsilon}_k\nu^k + (1-\tilde{\epsilon}_k)\nu^{R_k,\alpha_k,k}, \nu^k)\\ &\quad\le \mathcal{W}_1(\tilde{\epsilon}_k\nu^k + (1-\tilde{\epsilon}_k)\nu^{R_k,\alpha_k,k}, \nu) + \mathcal{W}_1(\nu,\nu^k) \to 0.\end{aligned}$$

Remember that $\nu^{R_k,\alpha_k,k} \le_c \nu^k$, hence $\tilde{\epsilon}_k\nu^k + (1-\tilde{\epsilon}_k)\nu^{R_k,\alpha_k,k} \le_c \nu^k$. Then by [37, Theorem 2.12], there exist martingale couplings $M^k \in \Pi_M(\tilde{\epsilon}_k\nu^k + (1-\tilde{\epsilon}_k)\nu^{R_k,\alpha_k,k}, \nu^k)$, $k \in \mathbb{N}$ such that, for $k \to +\infty$,

$$\int_{\mathbb{R}\times\mathbb{R}} |x-y|\, M^k(dx,dy) \le 2\mathcal{W}_1(\tilde{\epsilon}_k\nu^k + (1-\tilde{\epsilon}_k)\nu^{R_k,\alpha_k,k}, \nu^k) \to 0. \tag{5.14}$$

Let then

$$\pi^k(dx,dy) = \mu^k(dx)\int_{z\in\mathbb{R}} M_z^k(dy)\,\mathring{\pi}_x^k(dz) \in \Pi_M(\mu^k,\nu^k).$$

Using the fact that for $\mu^k(dx)$-almost every x, $\mathring{\pi}_x^k(dz)\, M_z^k(dy) \in \Pi(\mathring{\pi}_x^k, \pi_x^k)$, we get

$$\begin{aligned}\mathcal{AW}_1(\pi^k,\mathring{\pi}^k) &\le \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^k, \mathring{\pi}_x^k)\,\mu^k(dx) \le \int_{\mathbb{R}\times\mathbb{R}\times\mathbb{R}} |z-y|\,\mu^k(dx)\,M_z^k(dy)\,\mathring{\pi}_x^k(dz)\\ &= \int_{\mathbb{R}\times\mathbb{R}} |z-y|\,M^k(dy,dz),\end{aligned}$$

where the right-hand side vanishes by (5.14) as k goes to $+\infty$. Then (5.13) and the triangle inequality yield

$$\mathcal{AW}_1(\pi^k, \pi) \leq \mathcal{AW}_1(\pi^k, \mathring{\pi}^k) + \mathcal{AW}_1(\mathring{\pi}^k, \pi) \to 0, \quad k \to +\infty,$$

which concludes the proof. □

Funding Open access funding provided by Swiss Federal Institute of Technology Zurich.

References

1. Aldous, D.J.: Weak convergence and general theory of processes. Unpublished incomplete draft of monograph; Department of Statistics, University of California, Berkeley (1981)
2. Alfonsi, A., Corbetta, J., Jourdain, B.: Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds. Int. J. Theor. and Appl. Finance **22**(3) (2019)
3. Alfonsi, A., Corbetta, J., Jourdain, B.: Sampling of probability measures in the convex order by Wasserstein projection. Ann. l'Inst. Henri Poincaré B Probab. Stat. **56**(3), 1706–1729 (2020)
4. Ambrosio, L., Gigli, N.: A user's guide to optimal transport. In: Modelling and Optimisation of Flows on Networks. Lecture Notes in Math., vol. 2062, pp. 1–155. Springer, Heidelberg (2013)
5. Backhoff-Veraguas, J., Bartl, D., Beiglböck, M., Eder, M.: Adapted Wasserstein distances and stability in mathematical finance. Finance Stochast. **24**(3), 601–632 (2020)
6. Backhoff-Veraguas, J., Bartl, D., Beiglböck, M., Eder, M.: All adapted topologies are equal. Probab. Theory Related Fields 1–48 (2020)
7. Backhoff-Veraguas, J., Bartl, D., Beiglböck, M., Wiesel, J.: Estimating processes in adapted Wasserstein distance. Ann. Appl. Prob. (2019) (**to appear**)
8. Backhoff-Veraguas, J., Beiglböck, M., Pammer, G.: Existence, duality, and cyclical monotonicity for weak transport costs. Calc. Var. Partial. Differ. Equ. **58**(6), 203 (2019)
9. Backhoff-Veraguas, J., Pammer, G.: Stability of martingale optimal transport and weak optimal transport. Ann. Appl. Prob. (2022) (**to appear**)
10. Beiglböck, M., Cox, A.M., Huesmann, M.: Optimal transport and Skorokhod embedding. Invent. Math. **208**(2), 327–400 (2017)
11. Beiglböck, M., Henry-Labordère, P., Penkner, F.: Model-independent bounds for option prices: a mass transport approach. Finance Stochast. **17**(3), 477–501 (2013)
12. Beiglböck, M., Jourdain, B., Margheriti, W., Pammer, G.: Stability of the weak martingale optimal transport problem. arXiv arXiv:2109.06322 (2021)
13. Beiglböck, M., Juillet, N.: On a problem of optimal transport under marginal martingale constraints. Ann. Probab. **44**(1), 42–106 (2016)
14. Beiglböck, M., Nutz, M., Stebegg, F.: Fine Properties of the optimal Skorokhod embedding problem. JEMS (2020). https://doi.org/104171/JEMS/1122
15. Beiglböck, M., Nutz, M., Touzi, N.: Complete duality for martingale optimal transport on the line. Ann. Probab. **45**(5), 3038–3074 (2017)
16. Bion-Nadal, J., Talay, D.: On a Wasserstein-type distance between solutions to stochastic differential equations. Ann. Appl. Probab. **29**(3), 1609–1639 (2019)
17. Brückerhoff, M., Juillet, N.: Instability of martingale optimal transport in dimension d $\geq$ 2. arXiv arXiv:2101.06964 (2021)
18. Cheridito, P., Kiiski, M., Prömel, D.J., Soner, H.M.: Martingale optimal transport duality. Math. Ann. **379**, 1685–1712 (2021)

19. De March, H.: Entropic approximation for multi-dimensional martingale optimal transport. arXiv arXiv:1812.11104 (2018)
20. De March, H.: Local structure of multi-dimensional martingale optimal transport. arXiv arXiv:1805.09469 (2018)
21. De March, H.: Quasi-sure duality for multi-dimensional martingale optimal transport. arXiv arXiv:1805.01757 (2018)
22. De March, H., Touzi, N.: Irreducible convex paving for decomposition of multi-dimensional martingale transport plans. Ann. Probab. **47**(3), 1726–1774 (2019)
23. Dolinsky, Y., Soner, H.M.: Martingale optimal transport and robust hedging in continuous time. Probab. Theory Relat. Fields **160**(1–2), 391–427 (2014)
24. Dugundji, J.: An extension of Tietze's theorem. Pac. J. Math. **1**(3), 353–367 (1951)
25. Ethier, S.N., Kurtz, T.G.: Markov Processes: Characterization and Convergence, 2nd revised Wiley, Hoboken (2005)
26. Galichon, A., Henry-Labordère, P., Touzi, N.: A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. Ann. Appl. Probab. **24**(1), 312–336 (2014)
27. Ghoussoub, N., Kim, Y.-H., Lim, T.: Structure of optimal martingale transport plans in general dimensions. Ann. Probab. **47**(1), 109–164 (2019)
28. Gozlan, N., Roberto, C., Samson, P.-M., Shu, Y., Tetali, P.: Characterization of a class of weak transport-entropy inequalities on the line. Ann. Inst. Henri Poincaré Probab. Stat. **54**(3), 1667–1693 (2018)
29. Gozlan, N., Roberto, C., Samson, P.-M., Tetali, P.: Kantorovich duality for general transport costs and applications. J. Funct. Anal. **273**(11), 3327–3405 (2017)
30. Guo, G., Obłój, J.: Computational methods for martingale optimal transport problems. Ann. Appl. Probab. **29**(6), 3311–3347 (2019)
31. Hellwig, M.F.: Sequential decisions under uncertainty and the maximum theorem. J. Math. Econ. **25**(4), 443–464 (1996)
32. Henry-Labordère, P.: (Martingale) optimal transport and anomaly detection with neural networks: a primal-dual algorithm. arXiv arXiv:1904.04546 (2019)
33. Henry-Labordère, P., Tan, X., Touzi, N.: An explicit version of the one-dimensional Brenier's theorem with full marginals constraint. Stoch. Process. Appl. **126**(9), 2800–2834 (2016)
34. Henry-Labordère, P., Touzi, N.: An explicit martingale version of the one-dimensional Brenier theorem. Finance Stochast. **20**(3), 635–668 (2016)
35. Hirsch, F., Roynette, B.: A new proof of Kellerer's theorem. ESAIM Probab. Stat. **16**, 48–60 (2012)
36. Hobson, D., Neuberger, A.: Robust bounds for forward start options. Math. Finance **22**(1), 31–56 (2012)
37. Jourdain, B., Margheriti, W.: A new family of one dimensional martingale couplings. Electron. J. Probab. **25** (2020)
38. Juillet, N.: Stability of the shadow projection and the left-curtain coupling. Ann. Inst. Henri Poincaré Probab. Stat. **52**(4), 1823–1843 (2016)
39. Kantorovich, L.V.: On the translocation of masses. Dokl. Akad. Nauk SSSR **37**(7–8), 227–229 (1942)
40. Kertz, R.P., Rösler, U.: Stochastic and convex orders and lattices of probability measures, with a martingale interpretation. Israel J. Math. **77**(1–2), 129–164 (1992)
41. Kertz, R.P., Rösler, U.: Complete lattices of probability measures with applications to martingale theory. Lect. Notes Monogr. Ser. **35**, 153–177 (2000)
42. Lacker, D.: Dense sets of joint distributions appearing in filtration enlargements, stochastic control, and causal optimal transport. arXiv arXiv:1805.03185 (2018)
43. Lassalle, R.: Causal transference plans and their Monge–Kantorovich problems. Stoch. Anal. Appl. **36**(3), 452–484 (2018)
44. Monge, G.: Mémoire sur la théorie des déblais et des remblais. Histoire de l'académie Royale des Sciences de Paris (1781)
45. Neufeld, A., Sester, J.: On the stability of the martingale optimal transport problem: a set-valued map approach (2021)
46. Obłój, J., Siorpaes, P.: Structure of martingale transports in finite dimensions. arXiv arXiv:1702.08433 (2017)
47. Pflug, G.C., Pichler, A.: A distance for multistage stochastic optimization models. SIAM J. Optim. **22**(1), 1–23 (2012)
48. Pflug, G.C., Pichler, A.: Multistage Stochastic Optimization. Springer Series in Operations Research and Financial Engineering, Springer, Cham (2014)

49. Pflug, G.C., Pichler, A.: Dynamic generation of scenario trees. Comput. Optim. Appl. **62**(3), 641–668 (2015)
50. Pflug, G.C., Pichler, A.: From empirical observations to tree models for stochastic optimization: convergence properties. SIAM J. Optim. **26**(3), 1715–1740 (2016)
51. Santambrogio, F.: Optimal transport for applied mathematicians. In: Progress in Nonlinear Differential Equations and their Applications. Calculus of Variations, PDEs, and Modeling, vol 87. Springer, Cham (2015)
52. Strassen, V.: The existence of probability measures with given marginals. Ann. Math. Stat. **36**(2), 423–439 (1965)
53. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence (2003)
54. Villani, C.: Optimal Transport, Old and New. Grundlehren der mathematischen Wissenschaften, vol. 338. Springer, Berlin (2009)
55. Wiesel, J.: Continuity of the martingale optimal transport problem on the real line. arXiv arXiv:1905.04574 (2020)