DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

Optimal transport and mathematical finance through the lens of the adapted weak topology

verfasst von / submitted by

Dipl.-Ing. Gudmund Pammer, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2021 / Vienna, 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 796 605 405
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Mathematik
Betreut von / Supervisor:	Ass.-Prof. Dr. Julio Backhoff-Veraguas
	Univ.-Prof. Dr. Mathias Beiglböck

He used often to say there
was only one Road; that it
was like a great river: its
springs were at every
doorstep, and every path was
its tributary. “It’s a dangerous
business, Frodo, going out of
your door,” he used to say.
“You step into the Road, and
if you don’t keep your feet,
there is no knowing where
you might be swept off to.”

Frodo on Bilbo in
The Fellowship of the Ring

To my family, and Mathias and Manu,
Anne, Julio, Stefan, Daniel and Dani,
and all the other beautiful minds
I had the privilege to get to know.

I am glad you are here with me.
Here at the end of all things.

Frodo in
The Return of the King

Abstract

In recent years multiple novel variants of optimal transport (OT) were introduced such as weak optimal transport (WOT), martingale optimal transport (MOT), and entropic optimal transport to name but a few. Even though all of these variations are tightly connected to the original spirit of OT, the motivation behind them could not be more diverse: Gozlan et al. [8] initiate the field of weak optimal transport motivated by geometric inequalities; martingale optimal transport can be used to obtain model-independent prices for options in mathematical finance [2]; the entropic regularization of OT gained traction most notably due to Cuturi [4] who made OT computationally tractable.

In the classical interpretation of OT, there is an initial mass distribution μ and a terminal mass distribution ν. If transporting mass from one point x to another point y has a given fixed cost $c(x, y)$, then the aim is to move all mass μ to ν under minimal cost. OT is an essentially symmetric problem in the sense that mass transport can be interpreted both ways. From this perspective, MOT and WOT exhibit a substantial difference to OT as they incorporate asymmetries by treating coordinates disparately. This asymmetry leads to the necessity of introducing fitting, new tools for studying MOT and WOT. For this reason, a change of perspective can be fruitful: A different view on admissible transport plans π is to interpret them as the law of 1-timestep stochastic processes (X, Y) with marginal distributions (μ, ν). Stochastic processes are intrinsically asymmetric due to the role of time and information encoded in their filtrations.

The weak topology is the go-to topology in the OT setup but fails for acknowledging the introduced asymmetries of WOT and MOT. A promising candidate (particularly in the light of the stochastic process point of view) is the so-called adapted weak topology, which embraces the most important features of the weak topology in addition to a temporal structure.

My PhD project aims to investigate questions arising from the theory of optimal transport and mathematical finance through the lens of the adapted weak topology. This enables us to recover the basic fundaments of OT for WOT like the existence of optimizers, duality, and a geometric description of optimizers. It provides adequate tools to tackle the hitherto open question of stability in MOT which is also of great relevance from a mathematical finance perspective. This understanding also bears fruit in model-independent finance, where it is used to obtain robust sub- and superhedging prices in fixed income markets, and to identify extremal models.

Zusammenfassung

In den letzten Jahren wurden mehrere neue Varianten des optimalen Transports (OT) eingeführt, wie z.B. optimaler schwacher Transport (WOT), optimaler Martingaltransport (MOT) und entropischer optimaler Transport, um nur einige zu nennen. Auch wenn all diese Variationen eng mit dem ursprünglichen Geist von OT verbunden sind, könnten die dahinterliegenden Motivationen nicht vielfältiger sein: Gozlan et al. [8] initiieren das durch geometrische Ungleichungen motivierte Feld des optimalen schwachen Transports; Martingaltransport kann verwendet werden, um modellunabhängige Preise für Optionen in der Finanzmathematik zu erhalten [2]; die entropische Regularisierung von OT gewann vor allem durch Cuturi [4] an Zugkraft, der OT rechnerisch handhabbar machte.

In der klassischen Interpretation des OT gibt es eine initiale Massenverteilung μ und eine terminale Massenverteilung ν. Wenn der Massentransport von einem Punkt x zu einem anderen Punkt y mit gegebenen Kosten $c(x, y)$ verbunden ist, dann besteht das Ziel darin, die gesamte Masse μ mit minimalen Kosten nach ν zu befördern. Im Wesentlichen ist OT ein symmetrisches Problem in dem Sinne, dass der Massetransport in beide Richtungen gelesen werden kann. Aus dieser Perspektive weisen MOT und WOT einen wesentlichen Unterschied zu OT auf, da sie Asymmetrien enthalten, indem sie Koordinaten unterschiedlich behandeln. Aus dieser Asymmetrie ergibt sich die Notwendigkeit der Einführung passender, neuer Instrumente zur Untersuchung von MOT und WOT. Aus diesem Grund kann ein Perspektivenwechsel fruchtbar sein: Ein anderer Blickwinkel auf zulässige Transportpläne π ist jene, sie als die Verteilung von stochastischen 1-Zeitschritt-Prozessen (X, Y) mit Randverteilungen (μ, ν) zu sehen. Stochastische Prozesse sind aufgrund der Rolle von Zeit und Information, welche in ihrer Filtration kodiert ist, inhärent asymmetrisch.

Die schwache Topologie ist die Go-To-Topologie im OT-Setup, welche aber versäumt, die eingeführten Asymmetrien von WOT und MOT zu berücksichtigen. Ein vielversprechender Kandidat (insbesondere unter dem Gesichtspunkt von stochastischen Prozessen) ist die so genannte adaptierte schwache Topologie, die zusätzlich zu einer temporalen Struktur die wichtigsten Merkmale der schwachen Topologie aufweist.

Das Dissertationsprojekt zielt darauf ab, Fragen, die sich aus der Theorie des optimalen Transports und der Finanzmathematik ergeben, durch die Linse der adaptierten schwachen Topologie zu untersuchen. Dies ermöglicht es uns, die elementaren Grundlagen von OT auch für WOT wiederzuerlangen, wie die Existenz von Optimierern, Dualität und eine geometrische Beschreibung von Optimierern. Außerdem stellt es geeignete Instrumente zur Verfügung, um die bisher offene Frage der Stabilität von MOT zu lösen, die auch aus finanzmathematischer Sicht von großer Relevanz ist. Dieses Sichtweise trägt auch Früchte in modelindependent finance, wo es Anwendung findet, um robuste Sub- und Super-Hedging-Preise in einem Markt für festverzinsliche Wertpapiere herzuleiten und um extremale Modelle zu identifizieren.

Contents

Introduction and summary of results

Optimal transport

The optimal transport problem was first stated by Gaspard Monge, a French mathematician of the 18th century. In his formulation, he considered mass distributions μ and ν with equal mass and a function c describing the cost of moving mass from one point to another. He was interested in finding optimal ways of moving one pile of mass μ to another pile ν, i.e., minimizing

$$\int c\,(x, T(x))\;\mu(dx),$$

where T is a *transport map* describing where the mass lying at x is carried to. T has to satisfy the constraint that after moving the mass accordingly we end up distributed according to ν, i.e., $T(X) \sim \nu$ when $X \sim \mu$ or equivalently $T_{\#}\mu = \nu$.

Unfortunately, in this formulation basic fundamentals like the existence of optimizers and duality are non-trivial questions. Even for very 'well-behaved' cost functions c there do not necessarily exist optimizers or even an admissible candidate map T.

Almost 200 years later Leonid Kantorovich, a Russian mathematician, proposed a relaxation to Monge's original formulation which notably constitutes the starting point of modern OT: instead of looking for transport maps $T\colon X \to Y$, he admits that mass is (infinitesimally) split and transported to different locations. We phrase this in mathematical language. For a probability measure μ on the Polish space X and a probability measure ν on the Polish space Y, he considers all transport plans π which are probability measures on the product $X \times Y$ with marginals (μ, ν). Transport plans π are also called couplings with marginals μ, ν and the set of all such couplings is denoted by $\Pi(\mu, \nu)$. We want to minimize

$$\int c(x, y)\,\pi(dx, dy)$$

over all $\pi \in \Pi(\mu, \nu)$. As a result of this subtle modification, questions that previously seemed impossible to solve became fairly easy to answer with soft arguments. As an illustrative example consider the question of the existence of minimizers. By Portmanteau's theorem, the map $\pi \mapsto \pi(c)$ is lower semicontinuous w.r.t. the weak topology when c is lower semicontinuous, and lower bounded. Further, as an easy consequence of Prokhorov's theorem, we find that $\Pi(\mu, \nu)$ is weakly compact.

Thus, any minimizing sequence admits optimal accumulation points. We proceed to elaborate on the specific projects in more detail.

Weak optimal Transport

In Paper 1 we consider the WOT problem of Gozlan et al. [8], which permits cost functions of form $C\colon X \times \mathcal{P}(Y) \to \mathbb{R}$. We want to minimize

$$\int C(x, \pi_x)\, \mu(dx) \tag{WOT}$$

over all $\pi \in \Pi(\mu, \nu)$ where $(\pi_x)_{x \in X}$ denotes a regular disintegration of π w.r.t. its first marginal. It is clear that WOT is a generalization of OT as setting $C(x, p) := \int c(x, y)\, p(dy)$ for an OT cost c yields the same minimization problem. Even though duality was already accomplished in [8] under additional assumptions, the basic questions of existence and duality remained unanswered in the generality we are used to from OT. The trick here is – in analogy to Kantorovich who passed from transport maps to transport plans – to consider probability measures in $\mathcal{P}(X \times Y)$ as elements of a larger space, namely $\mathcal{P}(X \times \mathcal{P}(Y))$. A crucial property of this embedding is that it preserves relative compactness. When C is lower semicontinuous, lower-bounded and convex in its second argument, we can thereby show that (WOT) is lower semicontinuous as a function of π. By the same soft line of argument as in OT there exist optimizers.

Probably the most studied problem in OT is the Wasserstein distance $\mathcal{W}_2$ on $\mathbb{R}^d$ where $c(x, y) = |x - y|^2$ and $|\cdot|$ the euclidean distance on $\mathbb{R}^d$, i.e.,

$$\mathcal{W}_2(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \left(\int |x - y|^2\, \pi(dx, dy) \right)^{\frac{1}{2}}.$$

At the core of our understanding of the Wasserstein distance lies the Brenier theorem which closely knits together optimality with (gradients of) convex functions. An appealing variation is contained in the WOT framework under the name *Wasserstein projection* in convex order:

$$\mathcal{V}_2(\mu, \nu) := \inf_{\eta \leq_c \nu} \mathcal{W}_2(\mu, \eta) = \inf_{\pi \in \Pi(\mu, \nu)} \left(\int \left| \int y - x\, \pi_x(dy) \right|^2 \mu(dx) \right)^{\frac{1}{2}}, \tag{1}$$

where $\leq_c$ denotes the convex order, i.e., $\eta \leq_c \nu$ if and only if $\eta(f) \leq \nu(f)$ for every convex $f\colon \mathbb{R}^d \to \mathbb{R}$. The analogon of the Brenier theorem here is called the Brenier-Strassen theorem. This theorem was introduced by Gozlan and Juillet in [7] and it states that the minimizer η^* of $\mathcal{V}_2$ is the image measure of μ under the 1-Lipschitz gradient of a convex function. This theorem is named in conjunction with Strassen since he discovered that there is a martingale with given marginals if and only if the marginals are in convex order, see [11]. The particular WOT problem displayed in (1) was studied in numerous works during the past couple of years due to its interesting properties. Notably, it was used in a novel proof of Caffarelli's contraction theorem [5] and provides a convenient tool in the study of MOT, see Paper 7. Our treatment of WOT allowed us to prove the Brenier-Strassen theorem in Paper 1 and Paper 3 without requiring that μ, ν are compactly supported, which was assumed in an earlier version of [7].

In Paper 2 we extensively investigate the 1-dimensional Wasserstein projection. When μ does not weight single points, the coupling π^* given by the *monotone rearrangement* $F_\nu^{-1} \circ F_\mu$, the concatenation of cumulative distribution function F_μ and quantile map F_ν^{-1}, minimizes

$$\mathcal{W}_\theta(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int \theta(x-y)\, \pi(dx,dy),$$

for all convex $\theta\colon \mathbb{R} \to \mathbb{R}^+$. If θ is strictly convex and the value of $\mathcal{W}_\theta(\mu,\nu)$ is finite, π^* is even the unique minimizer. We call a monotone, 1-Lipschitz map $T\colon \mathbb{R} \to \mathbb{R}$ with $T(\mu) \leq_c \nu$ admissible. Then the *weak monotone rearrangement* is given by the (unique) admissible map which is maximal in the sense that $S(\mu) \leq_c T(\mu)$ for every other admissible S. Let $\mathcal{V}_\theta$ be given by (1) where we substitute the squared modulus with θ. For any admissible map T the coupling $\pi^T = (x, T(x))_\# \mu$ is optimal for $\mathcal{V}_\theta(\mu, T(\mu))$ and $\mathcal{W}_\theta(\mu, T(\mu))$, in particular, the values coincide. Let T be the weak monotone rearrangement between μ and ν. Then gluing π^T with any martingale coupling $\pi^M \in \Pi_M(T(\mu),\nu)$ yields an optimizer of $\mathcal{V}_\theta(\mu,\nu)$, where the existence of such martingale measures is guaranteed by Strassen's theorem. If θ is strictly convex, any optimizer of $\mathcal{V}_\theta(\mu,\nu)$ can be decomposed into π^T and a martingale coupling $\pi^M \in \Pi_M(T(\mu),\nu)$. Consider a process (X,Y,Z) where $(X,Y) = (X,T(X)) \sim \pi^T$ and $(Y,Z) \sim \pi^M$. The process first starts in $X \sim \mu$ and tries to move optimally in convex order to ν as martingale steps are free by the barycentric structure of the cost. From this perspective, we describe T as the unique admissible map with the property that on those intervals where T is strictly contracting, the process (Y,Z) remains constant. Remarkably this offers a natural way of joining two probability measures by a deterministic process and a martingale.

In Paper 3 we revisit several fields like the Schrödinger problem, optimal mechanism design, and semimartingale transport and recast it in the framework of WOT. Our viewpoint unifies the approaches and sometimes allows us to strengthen the original results.

Paper 4 is roughly split into two parts. The first part deals with a topic from mathematical finance: model-independence in a fixed-income market. The main realization here is that given market data in the form of marginal distributions of a process, the model-independent upper and lower bounds can be treated in the WOT set-up. Concretely we encounter a reformulation of the bounds closely related to the Wasserstein projection, which allows us to prove pricing-hedging dualities and to find explicit representations of optimizers. In the second part we lay the theoretical groundwork by investigating the derived class of problems from a WOT point of view. There we obtain duality formulas for cost functions $C\colon X \times \mathcal{P}(Y) \to \mathbb{R}$ where the value of $C(x,p)$ for fixed x depends solely on the barycenter of p. A standard assumption in WOT is the convexity of $C(x,\cdot)$ for all $x \in X$, whose sole purpose seems to be of technical nature. Let $\mu \in \mathcal{P}(X)$ be continuous. We reason that for bounded, continuous cost $C\colon X \times \mathcal{P}(Y) \to \mathbb{R}$ the minimum of (WOT) over $\Pi(\mu,\nu)$ is unchanged when exchanging C with $(x,p) \mapsto (C(x,\cdot))^{**}(p)$, that is for fixed x the convex hull of $C(x,\cdot)$.

Martingale Optimal Transport

As a research area MOT was kicked off in [2] by Beiglböck, Henry-Larbordère, and Penkner, and in [6] by Galichon, Henri-Larbordère, and Touzi. For notational convenience we focus on the 1-step instance. They propose to consider only transport plans π whose law is given by a 1-time step real-valued martingale, i.e., $(X, Y) \sim \pi$ and $\mathbb{E}_\pi[Y|X] = X$ almost surely. The set of all transport plans satisfying this additional constraint is denoted by $\Pi_M(\mu, \nu)$, which leads to the following constrained version of the OT problem:

$$\inf_{\pi \in \Pi_M(\mu,\nu)} \int c(x, y)\, \pi(dx, dy). \tag{MOT}$$

The main motivation of studying problems of this kind can be derived from robust mathematical finance. Martingale measures and mathematical finance are deeply connected by the *Fundamental Theorem of Asset Pricing* which states that a market is arbitrage-free if and only if there exists an equivalent martingale measure. Given the marginal distributions $(\mu, \nu), \mu \leq_c \nu$ for a single real-valued asset $(S_t)_{t=1}^2$ there can exist a multitude of martingale measures π describing the dynamics of a market. Therefore, in this setting model-independent pricing of an option $c(S_1, S_2)$ leads to the optimization problem (MOT).

Investigating the marginal dependence of the optimal value in (MOT) is of significant importance when dealing with real-world data or to obtain a better understanding of its numerical resulotion. The continuous dependence of value and optimizers on marginals is called stability. The question of stability was open up to recently and its answer required a combination of novel approaches.

Even though WOT encompasses MOT simply by setting

$$C(x, p) = \begin{cases} \int c(x, y)\, p(dy) & \int y\, p(dy) = x, \\ +\infty & \text{else}, \end{cases}$$

the hereby defined WOT cost C is usually not regular enough to handle MOT as a special case of WOT. To achieve comparable statements in the case of MOT we have to refine and specify the arguments with novel ideas.

In Paper 5 we show the stability of MOT and WOT by lending the notion of cyclical monotonicity and amending it for the specific structure of MOT and WOT. Cyclical monotonicity is a fundamental concept of OT going back at least to Rockafellar and Rüschendorf which provides a geometric and finitistic description of optimality of transport plans. The core idea here is to find a suitable notion that is preserved under limits and at the same time sufficient for optimality. This property is baptized C-monotonicity (for WOT) and martingale C-monotonicity (for MOT) respectively.

In Paper 6 we answer the question of approximating martingale couplings by sequences of martingale couplings. In mathematical language, this means the following: prescribe $(\mu^k, \nu^k)_{k\in\mathbb{N}}$ of pairs in convex order converging in the 1-Wasserstein distance to (μ, ν) and choose any $\pi \in \Pi_M(\mu, \nu)$. Is there a sequence of martingale couplings $\pi^k \in \Pi_M(\mu^k, \nu^k)$, $k \in \mathbb{N}$ converging to π at least in the 1-Wasserstein metric? In this paper, we show that this convergence can be achieved not only in $\mathcal{W}_1$ but even in the adapted weak topology. This is realized by exploiting the concept of potential functions – a one-dimensional tool characterizing the convex order – and combining it with the fruitful methods set forth in Paper 1.

Natural applications of the results are discussed in the follow-up paper Paper 7: We show the stability of the WOT-adaptation of (MOT)

$$\inf_{\pi \in \Pi_M(\mu,\nu)} \int C(x, \pi_x)\, \mu(dx).$$

As a special case, we find that the law of the 1-dimensional stretched Brownian motion [1] joining a pair (μ, ν) in the convex order continuously depends on the marginal distribution, (μ, ν). Moreover, we show continuous dependence of the super-hedging price of VIX-options (see e.g. [10]) on the marginals. We show that martingale C-monotonicity is a necessary and sufficient optimality criterion, which allows us to extend existing results on finite optimality being sufficient for optimality as in [3, 9].

Bibliographic information on papers contained in this thesis

Paper 1.

Existence, duality, and cyclical monotonicity for weak transport costs
Julio Backhoff-Veraguas, Mathias Beiglböck, Gudmund Pammer.
Published in: Calculus of Variations and Partial Differential Equations (2019), 58: 203.
Contribution of the author of this dissertation:
The paper was worked out in close collaboration.
Online versions available at: https://arxiv.org/abs/1809.05893
https://doi.org/10.1007/s00526-019-1624-y

Paper 2.

Weak monotone rearrangement on the line
Julio Backhoff-Veraguas, Mathias Beiglböck, Gudmund Pammer.
Published in: Electronic Communications in Probability (2020), Vol. 25, paper no. 18, 1-16.
Contribution of the author of this dissertation:
The paper was worked out in close collaboration.
Online versions available at: https://arxiv.org/abs/1902.05763
https://doi.org/10.1214/20-ECP292

Paper 5.

Stability of martingale optimal transport and weak optimal transport
Julio Backhoff-Veraguas, Gudmund Pammer.
Not published elsewhere yet.
Contribution of the author of this dissertation:
The paper was worked out in close collaboration.
Online versions available at: https://arxiv.org/abs/1904.04171

Paper 3.

Applications of weak transport theory
Julio Backhoff-Veraguas, Gudmund Pammer.
Not published elsewhere yet.
Contribution of the author of this dissertation:
The paper was worked out in close collaboration.
Online versions available at: https://arxiv.org/abs/2003.05338

Paper 4.

Model-independence in a fixed-income market and weak transport for non-convex costs
Beatrice Acciaio, Mathias Beiglböck, Gudmund Pammer.
Not published elsewhere yet.
Contribution of the author of this dissertation:
The paper was worked out in close collaboration.

Paper 6.

Approximation of martingale couplings in the adapted weak topology: part I

Mathias Beiglböck, Benjamin Jourdain, William Margheriti, Gudmund Pammer.

Not published elsewhere yet.

Contribution of the author of this dissertation:

The paper was worked out in close collaboration.

Paper 7.

Approximation of martingale couplings in the adapted weak topology: part II

Mathias Beiglböck, Benjamin Jourdain, William Margheriti, Gudmund Pammer.

Not published elsewhere yet.

Contribution of the author of this dissertation:

The paper was worked out in close collaboration.

Bibliography for the introduction

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

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

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

[4] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013.

[5] M. Fathi, N. Gozlan, and M. Prodhomme. A proof of the Caffarelli contraction theorem via entropic regularization. *Calculus of Variations and Partial Differential Equations*, 59(96), 2020.

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

[7] N. Gozlan and N. Juillet. On a mixture of Brenier and Strassen theorems. *Proceedings of the London Mathematical Society*, 120(3):434–463, 2020.

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

[9] C. Griessler. An extended footnote on finitely minimal martingale measures. *ArXiv e-prints*, June 2016.

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

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

EXISTENCE, DUALITY, AND CYCLICAL MONOTONICITY FOR WEAK TRANSPORT COSTS

J. BACKHOFF-VERAGUAS, M. BEIGLBÖCK, AND G. PAMMER

Abstract. The optimal weak transport problem has recently been introduced by Gozlan et. al. [25]. We provide general existence and duality results for these problems on arbitrary Polish spaces, as well as a necessary and sufficient optimality criterion in the spirit of cyclical monotonicity. As an application we extend the Brenier-Strassen Theorem of Gozlan-Juillet [23] to general probability measures on $\mathbb{R}^d$ under minimal assumptions.

A driving idea behind our proofs is to consider the set of transport plans with a new ('adapted') topology which seems better suited for the weak transport problem and allows to carry out arguments which are close to the proofs in the classical setup.

Keywords: Optimal Transport, cyclical monotonicity, Brenier's Theorem, Strassen's Theorem, weak transport costs, weak adapted topology, duality.
Mathematics Subject Classification (2010): 60G42, , 90C46, 58E30.

1. Introduction

1.1. **Notation.** This article is concerned with the optimal transport problem for weak costs, as initiated by Gozlan, Roberto, Samson and Tetali in [25]. To state it (see (1.1) below) we introduce some basic notation. We write $\mathcal{P}(Z)$ for the set of probability measures on a Polish space Z is and equip $\mathcal{P}(Z)$ with the usual weak topology. Throughout X and Y are Polish spaces, $\mu \in \mathcal{P}(X)$, and $\nu \in \mathcal{P}(Y)$. We write $\Pi(\mu, \nu)$ for the set of all couplings on $X \times Y$ with marginals μ and ν. Given a coupling π on $X \times Y$ we denote a regular disintegration with respect to the first marginal by $(\pi_x)_{x \in X}$. We consider cost functionals of the form

$$C : X \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\};$$

usually it is assumed that C is lower bounded and lower semicontinuous in an appropriate sense, and that $C(x, \cdot)$ is convex. The weak transport problem is then defined as

$$V(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x, \pi_x) \mu(dx). \tag{1.1}$$

1.2. **Literature.** The initial works of Gozlan et al. [25, 24] are mainly motivated by applications to geometric inequalities. Indeed, particular costs of the form (1.1) were already considered by Marton [30, 29] and Talagrand [40, 41]. Further papers directly related to [25] include [37, 36, 38, 21, 23]. Notably the weak transport problem (1.1) also yields a natural framework to investigate a number of related problems: it appears in the recursive formulation of the causal transport problem [7], in [1, 2, 15, 6] it is used to investigate martingale optimal transport problems, in [3] it is applied to prove stability of pricing and hedging in mathematical finance, it appears in the characterization of optimal mechanism for the multiple good monopolist [19] and motivates the investigation of linear transfers in [17]. A more classical example is given by entropy-regularized optimal transport (i.e. the Schrödinger problem); see [28] and the references therein.

University of Vienna. Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria. MB gratefully acknowledges support by FWF-grant Y00782. GP acknowledges support from the Austrian Science Fund (FWF) through grant number W 1245. All authors thank the anonymous referee for insightful comments that lead to a significant improvement of the article.

1.3. **Main results.** We will establish analogues of three fundamental facts in optimal transport theory: existence of optimizers, duality, and characterization of optimizers through c-cyclical monotonicity. We make the important comment, that these concepts (in particular existence and duality) have been previously studied for the weak transport problem. However, the results available so far may be too restrictive for certain applications.

Our goal is to establish these results at a level of generality that mimics the framework usually considered in the optimal transport literature (i.e. lower bounded, lower semicontinuous cost function). We emphasize that this extension is in fact required to treat specific examples of interest, cf. Section 1.3.4 below.

We briefly hint at the novel viewpoint which makes this extension possible: In a nutshell, the technicalities of the weak transport problem appear intricate and tedious since kernels $(\pi_x)_x$ are notoriously ill behaved with respect to weak convergence of measures on $\mathcal{P}(X \times Y)$. In the present paper we circumvent this difficulty by embedding $\mathcal{P}(X \times Y)$ into the bigger space $\mathcal{P}(X \times \mathcal{P}(Y))$. This idea is borrowed from the investigation of process distances (cf. [33, 5, 4]) and will allow us to carry out proofs that closely resemble familiar arguments from classical optimal transport.

1.3.1. *Primal Existence.* As a first contribution we will establish in Section 2 the following basic existence results.

Theorem 1.1 (Existence I)**.** *Assume that $C\colon X \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ is jointly lower semicontinuous, bounded from below, and convex in the second argument. Then, the problem*

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x, \pi_x)\mu(dx),$$

admits a minimizer.

Notably, Gozlan et.al. prove existence of minimizers under the assumption that $\pi \mapsto \int C(x, \pi_x)\, d\mu(x)$ is continuous on the set of all transport plans with first marginal μ, whereas our aim is to establish existence based on properties of the function C. We also note that Theorem 1.1 was first established by Alibert, Bouchitté, and Champion in [2] in the case where X, Y are compact spaces.

In fact the assumptions of Theorem 1.1 may be more restrictive than they initially appear. Indeed, as the cost function defined in (1.5) below is not lower semicontinuous with respect to weak convergence, we will need to employ a refined version of Theorem 1.1 to carry out our application in Theorem 1.4 below.

Given a compatible metric d_Y on the Polish space Y and $t \in [1, \infty)$, we write $\mathcal{P}^t_{d_Y}(Y)$ for the set of probability measures $\nu \in \mathcal{P}(Y)$ such that $\int d_Y(y, y_0)^t\, \nu(dy) < \infty$ for some (and then any) $y_0 \in Y$ and denote the t-Wasserstein metric on $\mathcal{P}^t_{d_Y}(Y)$ by $\mathcal{W}_t$ (see e.g. [42, Chapter 7]). In the sequel we make the convention that, whenever we refer to $\mathcal{P}^t_{d_Y}(Y)$, it is assumed that this set is equipped with the topology generated by $\mathcal{W}_t$. On the other hand, regarding the Polish space X, we fix from now on a compatible bounded metric d_X.

Theorem 1.2 (Existence II)**.** *Assume that $\nu \in \mathcal{P}^t_{d_Y}(Y)$. Let $C\colon X \times \mathcal{P}^t_{d_Y}(Y) \to \mathbb{R} \cup \{+\infty\}$ be jointly lower semicontinuous with respect to the product topology on $X \times \mathcal{P}^t_{d_Y}(Y)$, bounded from below, and convex in the second argument. Then, the problem*

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x, \pi_x)\mu(dx),$$

admits a minimizer.

We emphasize that Theorem 1.1 is a special case of Theorem 1.2. To see this, just take d_Y to be a compatible bounded metric. We also note that if C is strictly convex in the second argument and $V(\mu, \nu) < \infty$, then the minimizer $\pi^* \in \Pi(\mu, \nu)$ is unique. We report our proofs in Section 2.

1.3.2. *Duality.* We fix a compatible metric d_Y on Y and introduce the space

$$\Phi_{b,t} := \{\psi : Y \to \mathbb{R} \text{ continuous s.t. } \exists a, b, \ell \in \mathbb{R}, y_0 \in Y, \ell \leq \psi(\cdot) \leq a + b d_Y(y_0, \cdot)^t\}. \quad (1.2)$$

To each $\psi \in \Phi_{b,t}$ we associate the function

$$R_C\psi(x) := \inf_{p \in \mathcal{P}^t_{d_Y}(Y)} p(\psi) + C(x, p). \quad (1.3)$$

We remark that $R_C\psi(\cdot)$ is universally measurable if C is measurable ([16, Proposition 7.47]) and so the integral $\mu(R_C\psi)$ is well defined for all $\mu \in \mathcal{P}(Y)$ if C is lower-bounded.

Theorem 1.3. *Let $C\colon X \times \mathcal{P}^t_{d_Y}(Y) \to \mathbb{R} \cup \{\infty\}$ be jointly lower semicontinuous with respect to the product topology on $X \times \mathcal{P}^t_{d_Y}(Y)$, bounded from below, and convex in the second argument. Then we have for all $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}^t_{d_Y}(Y)$*

$$V(\mu, \nu) = \sup_{\psi \in \Phi_{b,t}} \mu(R_C\psi) - \nu(\psi). \quad (1.4)$$

The proof of Theorem 1.3 is provided in Section 3. We also refer to this section for a comparison of earlier duality results of Gozlan, Roberto, Samson, and Tetali [25, Theorem 9.6] and Alibert, Bouchitté, and Champion [2, Theorem 4.2].

1.3.3. *C-monotonicity.* Besides primal existence and duality, another fundamental result in classical optimal transport is the characterization of optimality through the notion of *cyclical monotonicity*; see [35, 22] as well as the monographs [34, 42, 43]. More recently, variants of this 'monotonicity priniciple' have been applied in transport problems for finitely or infinitely many marginals [32, 18, 26, 11, 44], the martingale version of the optimal transport problem [12, 31, 13], the Skorokhod embedding problem [9] and the distribution constrained optimal stopping problem [14].

We provide in Definition 5.1 below, a concept analogous to cyclical monotonicity (which we call C-monotonicity) for weak transport costs C. We show that every optimal transport plan is C-monotone in a very general setup. Conversely, we have that every C-monotone transport plan is optimal under certain regularity assumptions. See Theorems 5.3 and 5.6 respectively.

We note that related concepts already appeared in [6, Proposition 4.1] (where necessity of a 2-step optimality condition is established) and in [23] (necessity in the case of compactly supported measures and a quadratic cost criterion). To the best of our knowledge, our sufficient criterion is the first of its kind for weak transport costs.

We remark that the 2-step monotonicity principle for weak transport costs has already proved vital in [6] for the construction of a martingale counterpart to the Brenier theorem and the Benamou-Brenier formula. On the other hand, we conjecture that this monotonicity principle could be used in order to generalize [23] to non-quadratic costs.

1.3.4. *A general Brenier-Strassen theorem.* As an application of our abstract results we extend the Brenier-Strassen theorem [23, Theorem 1.2] of Gozlan and Juillet to the case of general probabilities on $X = Y = \mathbb{R}^d$ under the assumption that μ has finite second moment and ν has finite first moment. We thus drop the condition in [23] that the marginals have compact support. For this part we set

$$C(x, \rho) := \left| x - \int y \rho(dy) \right|^2, \quad (1.5)$$

and write $\leq_c$ for the convex order of probability measures.

Theorem 1.4. *Let $\mu \in \mathcal{P}^2(\mathbb{R}^d)$ and $\nu \in \mathcal{P}^1(\mathbb{R}^d)$. There exists a unique $\mu^* \leq_c \nu$ such that*

$$\mathcal{W}_2(\mu^*, \mu)^2 = \inf_{\eta \leq_c \nu} \mathcal{W}_2(\eta, \mu)^2 = V(\mu, \nu). \quad (1.6)$$

Moreover, there exists a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ of class C^1 with $\nabla\varphi$ being 1-Lipschitz, such that $\nabla\varphi(\mu) = \mu^*$. Finally, an optimal coupling $\pi^* \in \Pi(\mu, \nu)$ for $V(\mu, \nu)$ exists, and a coupling $\pi \in \Pi(\mu, \nu)$ is optimal for $V(\mu, \nu)$ if and only if $\int y \pi_x(dy) = \nabla\varphi(x)$ μ-a.s.

Existence of μ^* and the expression (1.6) were first proved by Gozlan et al [24] for $d = 1$ and by Alfonsi, Corbetta, Jourdain [1] for arbitrary $d \in \mathbb{N}$. Indeed a general version of (1.6), appealing to $\mathcal{W}_p$ and probabilities $\mu, \nu \in \mathcal{P}^p(\mathbb{R}^d)$ is provided in [1]. All other statements in the above theorem were originally established by Gozlan and Juillet [23] under the assumption of compactly supported measures μ, ν. The proof of Theorem 1.4 is given in Section 6.

Note added in revision. In an updated version of [23], Gozlan and Juillet have also removed the compactness assumption in Theorem 1.4. Their proof is based on duality arguments and in particular differs from the one given here.

2. Existence of minimizers

A principal idea behind the proofs of this paper is to endow the set of transport plans $\mathcal{P}(X\times Y)$ with a topology that is finer than the usual weak topology and which appropriately accounts for the asymmetric role of X and Y in the context of weak transport. This can be formalized by embedding $\mathcal{P}(X \times Y)$ into the bigger space $\mathcal{P}(X \times \mathcal{P}(Y))$. I.e., given a transport plan π, we will consider its disintegration $(\pi_x)_{x\in X}$ (w.r.t. its first marginal) and view it as a *Monge-type* coupling in the larger space $\mathcal{P}(X \times \mathcal{P}(Y))$. It turns out that on this 'extended' space the minimization problems Theorem 1.2 and Theorem 1.1 can be handled more efficiently.

We need to introduce additional notation: for a probability measure $\pi \in \mathcal{P}(X \times Y)$ with not further specified marginals, we write $\pi(dx\times Y)$ and $\pi(X\times dy)$ for its X-marginal and Y-marginal respectively. At several instances we use the projection from a product space onto one of its components. This map is usually denoted by $\mathrm{proj}_\bullet$ where the subscript describes the component, e.g. $\mathrm{proj}_X\colon X \times Y \to X$ stands for the projection onto the X-component. Denoting by $(\pi_x)_{x\in X}$ a regular disintegration of π with respect to $\pi(dx\times Y)$, we consider the measurable map

$$\begin{aligned} \kappa_\pi\colon X &\to X\times\mathcal{P}(Y) \\ x &\mapsto (x, \pi_x). \end{aligned}$$

We define the embedding $J\colon \mathcal{P}(X \times Y) \to \mathcal{P}(X \times \mathcal{P}(Y))$ by setting for $\pi \in \mathcal{P}(X \times Y)$ with X-marginal $\mu(dx) = \pi(dx \times Y)$

$$J(\pi) := (\kappa_\pi)_\#(\mu). \tag{2.1}$$

The map J is well-defined since κ_π is $\pi(dx \times Y)$-almost surely unique. Note that elements in $\mathcal{P}(X\times Y)$ precisely correspond to those elements of $\mathcal{P}(X\times\mathcal{P}(Y))$ which are concentrated on a graph of a measurable function from X to $\mathcal{P}(Y)$.

The intensity $I(P) \in \mathcal{P}(Y)$ of $P \in \mathcal{P}(\mathcal{P}(Y))$ is uniquely determined by

$$I(P)(f) = \int_{\mathcal{P}(Y)} p(f)P(dp) \quad \forall f \in C_b(Y). \tag{2.2}$$

The set of all probability measures $P \in \mathcal{P}(X\times\mathcal{P}(Y))$ with X-marginal μ and '$\mathcal{P}(Y)$-marginal intensity' ν is denoted by

$$\Lambda(\mu,\nu) := \left\{P \in \mathcal{P}(X \times \mathcal{P}(Y)) \mid \mathrm{proj}_X P = \mu,\ I(\mathrm{proj}_{\mathcal{P}(Y)}(P)) = \nu\right\}. \tag{2.3}$$

Similar to (2.2), we define the intensity of $P \in \mathcal{P}(X \times \mathcal{P}(Y))$ as the unique measure $\hat{I}(P) \in \mathcal{P}(X \times Y)$ such that

$$\int_{X\times Y} f(x,y)\hat{I}(P)(dx,dy) = \int_{X\times\mathcal{P}(Y)} \int_Y f(x,y)p(dy)P(dx,dp) \quad \forall f \in C_b(X \times Y). \tag{2.4}$$

Note that while J is in general not continuous (cf. Example 2.2), the mappings I and $\hat{I}$ are continuous.

Using (2.1) and (2.4) we find that

$$\Lambda(\mu,\nu) = \hat{I}^{-1}(\Pi(\mu,\nu)) \text{ and } J(\Pi(\mu,\nu)) \subseteq \Lambda(\mu,\nu).$$

Also note that $\hat{I}$ is the left-inverse of J, i.e., $\hat{I} \circ J(\pi) = \pi$ for $\pi \in \mathcal{P}(X \times Y)$. We now describe the relation between minimization problems on $\Pi(\mu, \nu)$ and $\Lambda(\mu, \nu)$:

Lemma 2.1. *Let $C\colon X \times \mathcal{P}(Y) \to \mathbb{R} \cup \{-\infty, +\infty\}$ be measurable, lower-bounded, and convex in the second argument. Then*

$$V(\mu, \nu) = \hat{V}(\mu, \nu), \tag{2.5}$$

where V was defined in (1.1) *and*

$$\hat{V}(\mu, \nu) := \inf_{P \in \Lambda(\mu,\nu)} \int_{X \times \mathcal{P}(Y)} C(x, p) P(dx, dp). \tag{2.6}$$

Proof. For any $\pi \in \Pi(\mu, \nu)$ we have $J(\pi) \in \Lambda(\mu, \nu)$ and

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

Thus,

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x, \pi_x) \mu(dx) \geq \inf_{P \in \Lambda(\mu,\nu)} \int_{X \times \mathcal{P}(Y)} C(x, p) P(dx, dp).$$

Now, letting $P \in \Lambda(\mu, \nu)$, we easily derive from (2.4) that $\hat{I}(P) \in \Pi(\mu, \nu)$ and $\hat{I}(P)_x = \int_{\mathcal{P}(Y)} p\, P_x(dp)$ for μ-a.e x. Using convexity we conclude

$$\begin{aligned}
\int_{X \times \mathcal{P}(Y)} C(x, p) P(dx, dp) &= \int_X \int_{\mathcal{P}(Y)} C(x, p) P_x(dp) \mu(dx) \\
&\geq \int_X C(x, \hat{I}(P)_x) \mu(dx) \\
&\geq \inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x, \pi_x) \mu(dx).
\end{aligned}$$

□

2.1. **Existence of minimizers.** The purpose of this subsection is to establish Theorem 1.2, or more precisely, a strengthened version of it; see Theorem 2.9 below. To this end we need a number of auxiliary results.

We start by stressing that, in general, the embedding J is not continuous. In fact:

Example 2.2. The map J is continuous if and only if X is discrete or $|Y| = 1$. Indeed, given X discrete and a sequence $(\pi^k)_{k \in \mathbb{N}} \in \mathcal{P}(X \times Y)^{\mathbb{N}}$ which weakly converges to π, we have that $\pi^k(x \times Y) \to \pi(x \times Y)$ from which $\pi^k_x(dy) = \frac{\pi^k(x,dy)}{\pi^k(x \times Y)}$ converges weakly to $\pi_x(dy) = \frac{\pi(x,dy)}{\pi(x \times Y)}$ if $\pi(x \times Y) > 0$. Consequently if $f \in C_b(X \times \mathcal{P}(Y))$, then

$$\begin{aligned}
&\lim_k |J(\pi^k)(f) - J(\pi)(f)| \\
&\quad \leq \limsup_k \sum_x |f(x, \pi^k_x)(\pi^k(x \times Y) - \pi(x \times Y))| + \sum_x |f(x, \pi^k_x) - f(x, \pi_x)| \pi(x \times Y) \\
&\quad = 0.
\end{aligned}$$

Therefore $(J(\pi^k))_{k \in \mathbb{N}}$ converges weakly to $J(\pi)$. On the other hand, suppose there is a sequence $(x_k)_{k \in \mathbb{N}} \in X^{\mathbb{N}}$ of distinct points converging to some $x \in X$, as well as $p, q \in \mathcal{P}(Y)$ with $p \neq q$. For $k \in \mathbb{N}$ define a probability measure on $\mathcal{P}(X \times Y)$ by

$$\pi^k(dx, dy) := \frac{1}{2}(\delta_{x_{k+1}}(dx) p(dy) + \delta_{x_k}(dx) q(dy)).$$

A short computation yields

$$\begin{aligned}
\lim_k J(\pi^k) &= \lim_k \frac{1}{2} \left(\delta_{(x_{k+1}, p)} + \delta_{(x_k, q)} \right) = \frac{1}{2} \left(\delta_{(x,p)} + \delta_{(x,q)} \right), \\
J(\lim_k \pi^k) &= J\Big(\frac{1}{2} \delta_x (p + q)\Big) = \delta_{(x, \frac{1}{2}(p+q))},
\end{aligned}$$

which shows that J is discontinuous.

On the bright side, J possesses a crucial feature: it maps relatively compact sets to relatively compact sets. We prove this in Lemma 2.6 below. But first we need to digress into the characterization of tightness on $\mathcal{P}(\mathcal{P}(Y))$ and subspaces thereof. The following can be found in [39, p. 178, Ch. II].

Lemma 2.3. *A set $\mathcal{A} \subseteq \mathcal{P}(\mathcal{P}(Y))$ is tight if and only if the set of its intensities $I(\mathcal{A})$ is tight in $\mathcal{P}(Y)$.*

We need to refine Lemma 2.3 for our purposes, since we equip $\mathcal{P}^t_{d_Y}(Y)$ with the $\mathcal{W}_t$-topology instead of the weak topology.

Lemma 2.4. *A set $\mathcal{A} \subseteq \mathcal{P}^t_{\mathcal{W}_t}(\mathcal{P}^t_{d_Y}(Y))$ is relatively compact if and only if the set of its intensities $I(\mathcal{A})$ is relatively compact in $\mathcal{P}^t_{d_Y}(Y)$.*

The proof of Lemma 2.4 heavily relies on the following lemma, for which we include a proof for sake of completeness.

Lemma 2.5. *A set $\mathcal{A} \subseteq \mathcal{P}^t_{d_Y}(Y)$ is relatively compact if and only if it is tight and*

$$\exists y' \in Y\, \forall \varepsilon > 0\, \exists R > 0 \colon \sup_{\mu \in \mathcal{A}} \int_{B_R(y')^c} d_Y(y, y')^t \mu(dy) < \varepsilon. \tag{2.7}$$

Note that if (2.7) holds for some $y' \in Y$ it automatically holds for any $y' \in Y$.

Proof of Lemma 2.4. Since continuous maps preserve relative compactness in Hausdorff spaces, the first implication follows by continuity of I. To show the reverse implication, let $I(\mathcal{A})$ be relatively compact in $\mathcal{P}^t_{d_Y}(Y)$. First we show for fixed $y' \in Y$ that

$$\forall \varepsilon > 0\, \exists R_\varepsilon > 0 \colon \sup_{P \in \mathcal{A}} \int_{\{p \colon \mathcal{W}_t(p, \delta_{y'})^t \geq R_\varepsilon\}} \mathcal{W}_t(p, \delta_{y'})^t P(dp) < \varepsilon. \tag{2.8}$$

Fix $\varepsilon > 0$. There exist $K > 0$ and $r > 0$ such that for all $P \in \mathcal{A}$

$$\int_{\mathcal{P}^t_{d_Y}(Y)} \mathcal{W}_t(p, \delta_{y'})^t P(dp) = \int_Y d_Y(y, y')^t I(P)(dy) \leq K$$

$$\int_{\mathcal{P}^t_{d_Y}(Y)} \int_{B_r(y')^c} d_Y(y, y')^t p(dy) P(dp) = \int_{B_r(y')^c} d_Y(y, y')^t I(P)(dy) < \frac{\varepsilon}{2}, \tag{2.9}$$

where $B_r(y') = \{y \in Y \colon d_Y(y, y') < r\}$. Set $R_\varepsilon = \frac{2r^t K}{\varepsilon}$ and $A_{R_\varepsilon} = \left\{p \in \mathcal{P}^t_{d_Y}(Y) \colon \mathcal{W}_t(p, \delta_{y'})^t \geq R_\varepsilon\right\}$, then

$$\sup_{P \in \mathcal{A}} P(A_{R_\varepsilon}) \leq \sup_{P \in \mathcal{A}} \frac{1}{R_\varepsilon} \int_{A_{R_\varepsilon}} \mathcal{W}_t(p, \delta_{y'})^t P(dp) \leq \frac{K}{R_\varepsilon}$$

and

$$\sup_{P \in \mathcal{A}} \int_{A_{R_\varepsilon}} \int_{B_r(y')} d_Y(y, y')^t p(dy) P(dp) \leq \sup_{P \in \mathcal{A}} P(A_{R_\varepsilon}) r^t \leq \frac{\varepsilon}{2}. \tag{2.10}$$

Putting (2.9) and (2.10) together shows (2.8).

It remains to show that $\mathcal{A}$ is tight in $\mathcal{P}(\mathcal{P}^t_{d_Y}(Y))$. By Lemma 2.3 we have that $\mathcal{A}$ is tight in $\mathcal{P}(\mathcal{P}(Y))$, i.e., given $\varepsilon > 0$ there is a compact set $K_\varepsilon \subseteq \mathcal{P}(Y)$ such that for all $P \in \mathcal{A}$ we have $P(K_\varepsilon) \geq 1 - \varepsilon$. We will construct a set $\tilde{K}_\varepsilon \subseteq K_\varepsilon$ which is compact in $\mathcal{P}^t_{d_Y}(Y)$ and satisfies $P(\tilde{K}_\varepsilon) \geq 1 - 2\varepsilon$ in $P \in \mathcal{A}$. To this end, take a sequence of radii $(R_n)_{n \in \mathbb{N}}$ such that

$$\sup_{P \in \mathcal{A}} P\left(\left\{p \colon \int_{\{y \colon d_Y(y, y')^t > R_n\}} d_Y(y, y')^t p(dy) \geq \frac{1}{n}\right\}\right) < \frac{\varepsilon}{2^n},$$

which is possible since

$$P\left(\left\{p \colon \int_{\{y \colon d_Y(y, y')^t > R_n\}} d_Y(y, y')^t p(dy) \geq \frac{1}{n}\right\}\right) \leq n \int_{\{y \colon d_Y(y, y')^t > R_n\}} d_Y(y, y')^t I(P)(dy),$$

can be chosen sufficiently small, uniformly for $P \in \mathcal{A}$. The set

$$\tilde{K}_\varepsilon := \left\{ p \in K_\varepsilon \colon \int_{\{y\colon d_Y(y,y')^t > R_n\}} d_Y(y,y')^t p(dy) \le \frac{1}{n}, \quad n \in \mathbb{N} \right\}$$

is compact in $\mathcal{P}^t_{d_Y}(Y)$ (cf. Lemma 2.5). Finally, given $P \in \mathcal{A}$ we obtain

$$P(\tilde{K}_\varepsilon) \ge P(K_\varepsilon) - \sum_n P\left(\left\{ p\colon \int_{\{y\colon d_Y(y,y')^t > R_n\}} d_Y(y,y') p(dy) \ge \frac{1}{n} \right\} \right) \ge 1 - 2\varepsilon$$

as desired □

Proof of Lemma 2.5.

'⇒': Since the topology induced by $\mathcal{W}_t$ on $\mathcal{P}^t_{d_Y}(Y)$ is finer than the weak topology on $\mathcal{P}^t_{d_Y}(Y)$, relative compactness in $\mathcal{W}_t$ implies relative compactness with respect to the weak topology. Therefore, Prokhorov's theorem yields tightness. Suppose for contradiction that (2.7) fails, i.e. there exist $y' \in Y$ and $\varepsilon > 0$ such that for all $N \in \mathbb{N}$ there is $\mu_N \in \mathcal{A}$ s.t.

$$\int_{B_N(y')^c} d_Y(y,y')^t \mu_N(dy) \ge \varepsilon.$$

In particular,

$$\lim_{R\to\infty} \liminf_N \int_{B_R(y')^c} d_Y(y',y)^t \mu_N(dy) \ge \varepsilon. \tag{2.11}$$

Due to relative compactness we find for any sequence in $\mathcal{A}$ an accumulation point. Then, from the definition of $\mathcal{W}_t$-convergence, see [43, Definition 6.8 (*iii*)], we deduce

$$\lim_{R\to\infty} \liminf_N \int_{B_R(y')^c} d_Y(y',y)^t \mu_N(dy) = 0,$$

which contradicts (2.11). Hence, (2.7) is satisfied.

'⇐': Let $\mathcal{A}$ be tight such that (2.7) holds. Then, any sequence $(\mu_k)_{k\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ has an accumulation point $\mu \in \mathcal{P}(Y)$ with respect to the weak topology. Without loss of generality assume that $\mu_k \to \mu$ for $k \to \infty$. By monotone convergence

$$\begin{aligned}\int d_Y(y,y')^t \mu(dy) &= \lim_{R\to\infty} \int R \wedge d_Y(y,y')^t \mu(dy) \\ &\le \lim_{R\to\infty} \liminf_{n\to\infty} \int R \wedge d_Y(y,y')^t \mu_n(dy) \le \sup_n \int d_Y(y,y')^t \mu_n(dy).\end{aligned}$$

Hence, by (2.7) we can choose (for $\varepsilon = 1$, say) $R > 0$ such that

$$\int_Y d_Y(y,y')^t \mu(dy) \le \sup_n \int_{B_R(y')} d_Y(y,y')^t \mu_n(dy) + 1 < \infty,$$

which shows that $\mu \in \mathcal{P}^t_{d_Y}(Y)$.

Next, fix $\varepsilon > 0$. Pick $R > 0$ such that

$$\int_Y d_Y(y,y')^t - R^t \wedge d_Y(y,y')^t \mu(dy) < \varepsilon,$$

$$\sup_n \int_{B_R(y')^c} d_Y(y,y')^t \mu_n(dy) < \varepsilon.$$

By weak convergence we know that

$$\lim_k \int_Y R^t \wedge d_Y(y,y')^t \mu_k(dy) \to \int_Y R^t \wedge d_Y(y,y')^t \mu(dy).$$

Hence we may pick k_0 such that for all $k \ge k_0$

$$\left| \int_Y R^t \wedge d_Y(y,y')^t (\mu_k - \mu)(dy) \right| < \varepsilon.$$

Thus we have for $k \geq k_0$

$$\Big| \int_Y d_Y(y,y')^t(\mu_k - \mu)(dy)\Big| < 3\varepsilon.$$

Since ε was arbitrary, we obtain that the t-moments are converging, which implies convergence in $\mathcal{W}_t$. □

We recall that on Y we are usually given a compatible complete metric d_Y, whereas on X we fix a compatible bounded metric d_X. We thus endow the product spaces $X \times Y$ and $X \times \mathcal{P}^t_{d_Y}(Y)$ with natural (product) metrices d and $\hat{d}$ defined respectively by

$$d((x,y),(x_0,y_0)) = d_X(x,x_0) + d_Y(y,y_0), \tag{2.12}$$

$$\hat{d}((x,p),(x_0,p_0)) = d_X(x,x_0) + \mathcal{W}_t(p,p_0). \tag{2.13}$$

We can now state and prove the crucial property of J:

Lemma 2.6. *If $\Pi \subseteq \mathcal{P}^t_d(X \times Y)$ is relatively compact then $J(\Pi) \subseteq \mathcal{P}^t_{\hat{d}}(X \times \mathcal{P}^t_{d_Y}(Y))$ is relatively compact. Conversely, if $\Lambda \in \mathcal{P}^t_{\hat{d}}(X \times \mathcal{P}^t_{d_Y}(Y))$ is relatively compact then $\hat{I}(\Lambda) \subseteq \mathcal{P}^t_d(X \times Y)$ is relatively compact.*

Proof. Since continuous maps preserve relative compactness in Hausdorff spaces, we immediately deduce relative compactness of $\hat{I}(\Lambda)$, and the sets $\Pi^X \subseteq \mathcal{P}(X)$ and $\Pi^Y \subseteq \mathcal{P}^t_{d_Y}(Y)$ consisting respectively of the X- and Y-marginals of the elements in Π.

Denote now respectively by $\Pi^X_J \subseteq \mathcal{P}(X)$ and $\Pi^Y_J \subseteq \mathcal{P}^t_{\mathcal{W}_t}(\mathcal{P}^t_{d_Y}(Y))$ the set of X- and $\mathcal{P}(Y)$-marginals of the elements in $J(\Pi)$. Clearly $\Pi^X_J = \Pi^X$. By Lemma 2.4, the set Π^Y_J is relatively compact in $\mathcal{P}^t_{\mathcal{W}_t}(\mathcal{P}^t_{d_Y}(Y))$ if and only if the set $I(\Pi^Y_J)$ is relatively compact in $\mathcal{P}^t_{d_Y}(Y)$. However, if m is equal to the $\mathcal{P}(Y)$-marginal of $J(\pi)$, then $I(m)$ is equal to the Y-marginal of π. It follows that $I(\Pi^Y_J) \subseteq \Pi^Y$ is relatively compact and so is Π^Y_J. Since the marginals of $J(\Pi)$ are relatively compact, we conclude that $J(\Pi)$ itself is relatively compact. □

It is convenient to introduce the following assumptions, which we will often require:

Definition 2.7 (A)**.** *Given Polish spaces X, Y, we say that a function*

$$C\colon X \times \mathcal{P}^t_{d_Y}(Y) \to \mathbb{R} \cup \{+\infty\}$$

satisfies Condition (A) *if the following hold:*

- *C is lower semicontinuous with respect to the product topology of*

$$(X, d_X) \times (\mathcal{P}^t_{d_Y}(Y), \mathcal{W}_t),$$

- *C is bounded from below.*

If in addition for all $x \in X$ the map $p \mapsto C(x,p)$ is convex, i.e.

$$p, q \in \mathcal{P}^t_{d_Y}(Y), \alpha \in [0,1] \Rightarrow \; C(x, \alpha p + (1-\alpha)q) \leq \alpha C(x,p) + (1-\alpha)C(x,q), \tag{2.14}$$

then we say that C satisfies Condition (A+).

We now show that under Condition (A+) the cost functional defining the weak transport problem is lower semicontinuous:

Proposition 2.8. *Let $C\colon X \times \mathcal{P}^t_{d_Y}(Y) \to \mathbb{R} \cup \{+\infty\}$ satisfy condition* (A)*. Then the map*

$$\mathcal{P}^t_{\hat{d}}(X \times \mathcal{P}^t_{d_Y}(Y)) \ni P \mapsto \int_{X \times \mathcal{P}^t_{d_Y}(Y)} C(x,p) P(dx, dp) \tag{2.15}$$

is lower semicontinuous. If C satisfies condition (A+) *then the map*

$$\mathcal{P}^t_d(X \times Y) \ni \pi \mapsto \int_X C(x, \pi_x) \pi(dx \times Y) \tag{2.16}$$

is lower semicontinuous.

Proof. Let $P^k \to P$ in $\mathcal{P}^t_{\hat{d}}(X \times \mathcal{P}^t_{d_Y}(Y))$. Similar to [20, Theorem A.3.12], we can approximate C from below by $\hat{d}$-Lipschitz functions and obtain lower semicontinuity of (2.15), i.e.,

$$\liminf_k \int_{X\times\mathcal{P}(Y)} C(x,p)\, P^k(dx,dp) \geq \int_{X\times\mathcal{P}(Y)} C(x,p)\, P(dx,dp).$$

To show lower semicontinuity of (2.16), let $\pi^k \to \pi$ in $\mathcal{P}^t_d(X\times Y)$ and denote $P^k = J(\pi^k)$. We may assume that $\liminf_k \int_X C(x,\pi^k_x)\pi^k(dx\times Y) = \lim_k \int_X C(x,\pi^k_x)\pi^k(dx\times Y)$ by selecting a subsequence. By Lemma 2.6 we know that $\{P^k\}_k$ is relatively compact in $\mathcal{P}^t_{\hat{d}}(X\times\mathcal{P}^t_{d_Y}(Y))$. Denote by P an accumulation point of $\{P^k\}_k$. From now on we work along a subsequence converging to P. Observe that

$$\int_X C(x,\pi^k_x)\pi^k(dx\times Y) = \int_{X\times\mathcal{P}(Y)} C(x,p)\, P^k(dx,dp).$$

Hence, we find by the first part that

$$\liminf_k \int_{X\times\mathcal{P}(Y)} C(x,p)\, P^k(dx,dp) \geq \int_{X\times\mathcal{P}(Y)} C(x,p)\, P(dx,dp).$$

Observe that the X-marginal of P equals the X-marginal of π, so by convexity of $C(x,\cdot)$ we then have

$$\begin{aligned}\liminf_k \int_X C(x,\pi^k_x)\pi^k(dx\times Y) &\geq \int_{X\times\mathcal{P}(Y)} C(x,p)\, P_x(dp)\pi(dx\times Y)\\ &\geq \int_X C\Big(x, \int_{\mathcal{P}(Y)} p(dy)P_x(dp)\Big)\pi(dx\times Y).\end{aligned}$$

Now, if f is continuous bounded on $X\times Y$, we have

$$\int_{X\times Y} f(x,y)\pi^k(dx,dy) \to \int_{X\times Y} f(x,y)\pi(dx,dy).$$

But the function $F(x,p) := \int_Y f(x,y)p(dy)$ is easily seen to be continuous and bounded in $X\times\mathcal{P}(Y)$. Hence $\int F dP^k \to \int F dP$ and by the structure of F we deduce

$$\int_{X\times Y} f(x,y)\pi(dx,dy) = \int F dP = \int_{X\times\mathcal{P}(Y)} \int_Y f(x,y)p(dy)P(dx,dp).$$

This shows for the disintegration $(\pi_x)_{x\in X}$ of π that $\pi_x(dy) = \int_{\mathcal{P}(Y)} p(dy)\, P_x(dp)$ for $\pi(dx\times Y)$-almost every x. So we conclude

$$\liminf_k \int_X C(x,\pi^k_x)\pi^k(dx\times Y) \geq \int_X C(x,\pi_x)\pi(dx\times Y).$$

□

We are finally ready to provide our main existence result:

Theorem 2.9. *Let $C\colon X\times\mathcal{P}^t_{d_Y}(Y) \to \mathbb{R}\cup\{+\infty\}$ satisfy Condition* (A). *If $\Lambda \subseteq \mathcal{P}^t_{\hat{d}}(X\times\mathcal{P}^t_{d_Y}(Y))$ is compact, then there exists a minimizer $P^* \in \Lambda$ of*

$$\inf_{P\in\Lambda} \int_{X\times\mathcal{P}(Y)} C(x,p)P(dx,dp).$$

In particular $\mathcal{P}(X)\times\mathcal{P}^t_{d_Y}(Y) \ni (\mu,\nu) \mapsto \hat{V}(\mu,\nu)$ is lower semicontinuous and $\hat{V}(\mu,\nu)$ is attained (recall (2.6)*). Assume now that C fulfils Condition* (A+) *and $\Pi \subseteq \mathcal{P}^t_d(X\times Y)$ is compact. Then there exists a minimizer $\pi^* \in \Pi$ of*

$$\inf_{\pi\in\Pi} \int_X C(x,\pi_x)\pi(dx\times Y).$$

In particular $\mathcal{P}(X)\times\mathcal{P}^t_{d_Y}(Y) \ni (\mu,\nu) \mapsto V(\mu,\nu)$ is lower semicontinuous and $V(\mu,\nu)$ is attained (recall (1.1)*).*

Proof. The existence of minimizers in Λ and Π are direct consequences of their compactness and the lower semicontinuity of the objective functionals (Proposition 2.8).

We move to the study of $\hat{V}$. Let $(\mu_k, \nu_k) \to (\mu, \nu)$ in $\mathcal{P}(X) \times (\mathcal{P}^t_{d_Y}, \mathcal{W}_t)$. For any $k \in \mathbb{N}$ we find an optimizer P^*_k of $\hat{V}(\mu_k, \nu_k)$. Note that the set $\{P^*_k : k \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}^t_{\hat{d}}(X \times \mathcal{P}^t_{d_Y}(Y))$. Therefore, we can find again a converging subsequence with limit point in $\Pi(\mu, \nu)$. Without loss of generality we assume

$$\liminf_k \hat{V}(\mu_k, \nu_k) = \lim_k \hat{V}(\mu_k, \nu_k).$$

Using lower semicontinuity of the objective functional shows the assertion for $\hat{V}$. By Lemma 2.1 the lower semicontinuity of V is immediate. □

Of course Theorems 1.1 and 1.2 are particular cases of the second half of Theorem 2.9. More generally: if A is compact in $\mathcal{P}(X)$ and B is compact in $(\mathcal{P}^t_{d_Y}(Y), \mathcal{W}_t)$, then $\Pi := \bigcup_{\mu \in A, \nu \in B} \Pi(\mu, \nu)$ is compact in $\mathcal{P}^t_d(X \times Y)$ and Theorem 2.9 applies.

3. Duality

We denote by Φ_t the set of continuous functions on Y which satisfy the growth constraint

$$\exists y_0 \in Y, \exists a, b \in \mathbb{R}_+, \forall y \in Y : |\psi(y)| \leq a + b d_Y(y, y_0)^t,$$

and by $\Phi_{b,t}$ the subset of functions in Φ_t which are bounded from below. Further, we recall the notion of C-conjugate : The C-conjugate of a measurable function $\psi : Y \to \mathbb{R}$, denoted $R_C\psi$, is given by

$$R_C\psi(x) := \inf_{p \in \mathcal{P}^t_{d_Y}(Y)} p(\psi) + C(x, p). \tag{3.1}$$

We obtain Theorem 1.3 as a particular case of the following:

Theorem 3.1. *Let $C : X \times \mathcal{P}^t_{d_Y}(Y) \to \mathbb{R} \cup \{+\infty\}$ satisfy Condition* (A)*. Then*

$$\inf_{P \in \Lambda(\mu,\nu)} \int_{X \times \mathcal{P}(Y)} C(x, p) P(dx, dp) = \sup_{\psi \in \Phi_{b,t}} -\nu(\psi) + \int_X R_C\psi(x)\mu(dx). \tag{3.2}$$

If moreover C satisfies Condition (A+)*, then*

$$V(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x, \pi_x)\mu(dx) = \sup_{\psi \in \Phi_{b,t}} -\nu(\psi) + \int_X R_C\psi(x)\mu(dx). \tag{3.3}$$

Remark 3.2. A proof of Theorem 1.3 can be obtained by means of [25, Theorem 9.6], since we may verify the hypotheses therein thanks to our Proposition 2.8. We prefer to obtain the slightly stronger Theorem 3.1 via self-contained arguments. The primal-dual equality (3.3) was obtained in [2, Theorem 4.2] in the case when X, Y are compact spaces.

Proof of Theorem 3.1. Fix $y_0 \in Y$. Define the auxiliary cost function $\widetilde{C} : X \times \mathcal{P}^t_{d_Y}(Y)$ by

$$\widetilde{C}(x, p) := C(x, p) + \mathcal{W}_t(p, \delta_{y_0})^t$$

and $F : \mathcal{P}^t_{d_Y}(Y) \to \mathbb{R} \cup \{+\infty\}$ by

$$\begin{aligned} F(m) &:= \inf_{P \in \Lambda(\mu,m)} \int_{X \times \mathcal{P}(Y)} \widetilde{C}(x, p) P(dx, dp) \\ &= \inf_{P \in \Lambda(\mu,m)} \int_{X \times \mathcal{P}(Y)} C(x, p) P(dx, dp) + \int_Y d_Y(y, y_0)^t m(dy). \end{aligned} \tag{3.4}$$

Since the integrand $\widetilde{C}$ is bounded from below and lower semicontinuous we can apply Proposition 2.8 and find that F is lower semicontinuous on $\mathcal{P}^t_{d_Y}(Y)$. Note that for any $\alpha \in [0, 1]$ and $m_1, m_2 \in \mathcal{P}^t_{d_Y}(Y)$ we have

$$P_i \in \Lambda(\mu, m_i),\ i = 1, 2 \implies \alpha P_1 + (1 - \alpha) P_2 \in \Lambda(\mu, \alpha m_1 + (1 - \alpha) m_2),$$

and, particularly, it follows that F is convex. We can extend F to the set $\mathcal{M}^t_{d_Y}(Y)$ of bounded signed measures with finited t-moment (i.e. $m \in \mathcal{M}^t_{d_Y}(Y)$ implies $\int_Y d_Y(y, y_0)^t |m|(dy) < \infty$

for some y_0) by setting $F(m) = +\infty$ if $m \notin \mathcal{P}^t_{d_Y}(Y)$. We equip the space $\mathcal{M}^t_{d_Y}(Y)$ with the topology induced by Φ_t. It follows that the extension of F is still convex and lower semicontinuous. Now, the spaces Φ_t and $\mathcal{M}^t_{d_Y}(Y)$ are in separating duality. Define the convex conjugate $F^* \colon \Phi_t \to \mathbb{R} \cup \{+\infty\}$ of F by

$$F^*(\psi) = \sup_{m \in \mathcal{P}^t_{d_Y}(Y)} m(\psi) - F(m). \tag{3.5}$$

Observe that $F^*(\psi) = \lim_{k\to+\infty} F^*(\psi \wedge k)$, by monotone convergence. We may apply the Fenchel duality theorem [45, Theorem 2.3.3], and then replace Φ_t by $\Phi_{b,t}$, obtaining:

$$\begin{aligned}
F(m) &= \sup_{\psi \in \Phi_t} m(\psi) - F^*(\psi) \\
&= \sup_{-\psi \in \Phi_{b,t}} m(\psi) - F^*(\psi) \\
&= \sup_{\psi \in \Phi_{b,t}} m(-\psi) - F^*(-\psi).
\end{aligned}$$

Now we show that

$$F^*(-\psi) = -\int_X R_{\widetilde{C}}\psi(x)\mu(dx). \tag{3.6}$$

Rewriting (3.5) yields

$$\begin{aligned}
F^*(-\psi) &= \sup_{m \in \mathcal{P}^t_{d_Y}(Y)} m(-\psi) - \inf_{P \in \Lambda(\mu,m)} \int_{X\times\mathcal{P}(Y)} \widetilde{C}(x,p)P(dx,dp) \\
&= \sup_{\substack{m \in \mathcal{P}^t_{d_Y}(Y) \\ P \in \Lambda(\mu,m)}} -\int_X \left(\int_{\mathcal{P}(Y)} p(\psi) + \widetilde{C}(x,p)P_x(dp) \right) \mu(dx) \\
&= -\inf_{\substack{m \in \mathcal{P}^t_{d_Y}(Y) \\ P \in \Lambda(\mu,m)}} \int_X \left(\int_{\mathcal{P}(Y)} p(\psi) + \widetilde{C}(x,p)P_x(dp) \right) \mu(dx) \\
&\le -\int_X R_{\widetilde{C}}\psi(x)\mu(dx).
\end{aligned}$$

To show the converse inequality, we assume without loss of generality that

$$\int_X \widetilde{R_C}\psi(x)\mu(dx) < +\infty.$$

For all $x \in X$ the value of $R_{\widetilde{C}}\psi(x)$ is finite, because ψ is bounded from below. Fix $\varepsilon > 0$. The map $R_{\widetilde{C}}\psi(\cdot)$ is lower semianalytic by [16, Proposition 7.47] and by [16, Proposition 7.50] there exists an analytically measurable probability kernel $(\tilde{p}_x)_{x\in X} \in (\mathcal{P}^t_{d_Y}(Y))^X$ such that for all $x \in X$

$$p_x(\psi) + \widetilde{C}(x, p_x) \le R_{\widetilde{C}}\psi(x) + \varepsilon.$$

Then, we immediately obtain

$$\int_X p_x(\psi) + \widetilde{C}(x,p_x)\mu(dx) \le \int_X R_{\widetilde{C}}\psi(x)\mu(dx) + \varepsilon.$$

The term $\delta_{p_x}(dp)\mu(dx)$ uniquely defines a probability measure $\tilde{P}$ on $X \times \mathcal{P}(Y)$. Since $\widetilde{C}$ and ψ are bounded from below, we infer that

$$\mathcal{W}_t(\mathrm{proj}_Y \hat{I}(\tilde{P}), \delta_{y_0})^t = \int_{X\times\mathcal{P}(Y)} \mathcal{W}_t(p, \delta_{y_0})^t \tilde{P}(dx,dp) < +\infty,$$

and in particular $\mathrm{proj}_Y \hat{I}(\tilde{P}) \in \mathcal{P}^t_{d_Y}(Y)$. Clearly $\tilde{P} \in \Lambda(\mu, \mathrm{proj}_Y \hat{I}(\tilde{P}))$, so

$$\begin{aligned}
-\int_X \widetilde{R_C}\psi(x)\mu(dx) &\le \mathrm{proj}_Y(\hat{I}(\tilde{P}))(-\psi) - \int_{X\times\mathcal{P}(Y)} C(x,p) + \mathcal{W}_t(p,\delta_{y_0})^t \tilde{P}(dx,dp) + \varepsilon \\
&\le \mathrm{proj}_Y(\hat{I}(\tilde{P}))(-\psi) - F\left(\mathrm{proj}_Y \hat{I}(\tilde{P})\right) + \varepsilon \\
&\le F^*(-\psi) + \varepsilon,
\end{aligned}$$

and since ε was arbitrary, we have shown (3.6).

So far, we know that

$$F(m) = \sup_{\psi \in \Phi_{b,t}} -m(\psi) + \int_X R_{\widetilde{C}}\psi(x)\mu(dx).$$

Define $f(y) := d_Y(y, y_0)^t$ and note that $R_C(\psi + f)(x) = R_{\widetilde{C}}\psi(x)$ for all $x \in X$, as well as $\psi + f \in \Phi_{b,t}$ for $\psi \in \Phi_{b,t}$. From (3.4) we get

$$\begin{aligned}\inf_{P \in \Lambda(\mu,m)} P(C) &= F(m) - \mathcal{W}_t(m, \delta_{y_0})^t \\ &= \sup_{\psi \in \Phi_{b,t}} -m(\psi + f) + \int_X R_{\widetilde{C}}\psi(x)\mu(dx) \\ &= \sup_{\psi \in \Phi_{b,t}} -m(\psi) + \int_X R_C\psi(x)\mu(dx),\end{aligned}$$

which shows (3.2).

If for all $x \in X$ the map $C(x, \cdot)$ is convex, then (3.3) follows by Lemma 2.1 and (3.2). □

4. On the restriction property

The restriction property of optimal transport roughly states that if a coupling is optimal, then the conditioning of the coupling to a subset is also optimal given its marginals. This property fails for weak optimal transport, as we illustrate with a simple example:

Example 4.1. Let $X = Y = \mathbb{R}$, $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, $\nu = \frac{1}{4}\delta_{-2} + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_2$ and $C(x,\rho) = \left(x - \int y\rho(dy)\right)^2$. We consider the weak transport problem with these ingredients, and observe that an optimal coupling is given by

$$\pi = \frac{1}{4}[\delta_{(1,2)} + \delta_{(1,0)} + \delta_{(-1,0)} + \delta_{(-1,-2)}],$$

since it produces a cost equal to zero. Consider the set $K = \{(x,y) : y \neq 0\}$ and $\tilde{\pi}(dx,dy) = \pi(dx,dy|K)$ the conditioning of π to the set K, i.e. $\tilde{\pi}(S) := \frac{\pi(S\cap K)}{\pi(K)}$. It follows that

$$\tilde{\pi} = \frac{1}{2}[\delta_{(1,2)} + \delta_{(1,-2)}],$$

and denoting by $\tilde{\mu}$ and $\tilde{\nu}$ the first and second marginals of $\tilde{\pi}$, we have $\tilde{\mu} = \mu$ and $\tilde{\nu} = \frac{1}{2}\delta_2 + \frac{1}{2}\delta_{-2}$. With $\tilde{\mu}$ and $\tilde{\nu}$ and again the cost C as ingredients, an optimizer for the weak transport problem is given by

$$\hat{\pi} = \frac{3}{8}\delta_{(1,2)} + \frac{1}{8}\delta_{(1,-2)} + \frac{1}{8}\delta_{(-1,2)} + \frac{3}{8}\delta_{(-1,-2)},$$

since this time this coupling produces a cost equal to zero. On the other hand the cost of $\tilde{\pi}$ is equal to 1, and so $\tilde{\pi}$ is not optimal between is marginals.

However, we can state the following positive result.[1]

Proposition 4.2. *Suppose that π is optimal between the marginals μ and ν, $V(\mu,\nu) < \infty$, and that $C(x,\cdot)$ is convex. Let $0 \leq \tilde{\mu} \leq \mu$ be a non-negative measure such that $0 \not\equiv \tilde{\mu}$ and define $\hat{\mu} = \tilde{\mu}/\tilde{\mu}(X)$. Then $\hat{\pi}(dx,dy) := \hat{\mu}(dx)\pi_x(dy)$ is optimal between its marginals.*

Proof. By contradiction, suppose there exists a coupling χ with the same marginals as $\hat{\pi}$ such that

$$\int C(x,\chi_x)\hat{\mu}(dx) < \int C(x,\hat{\pi}_x)\hat{\mu}(dx).$$

[1] In a preliminary version of this article the restriction property Proposition 4.2 was used to derive Theorem 1.4 from the compact version given by Gozlan and Juillet [23]. Following the insightful suggestion of the anonymous referee, we now give a more self contained argument that does not require Proposition 4.2 / [23]. We have decided to keep Proposition 4.2 since it might be of some independent interest.

Now define $\pi^* := \pi + \tilde{\mu}(X)[\chi - \hat{\pi}] = \pi - \tilde{\mu}.\pi_x + \tilde{\mu}(X)\chi$. Observe that π^* has marginals μ, ν, and $\pi^*(X \times Y) = 1$. We also have $\pi^* \geq 0$ since $\tilde{\mu} \leq \mu$, so π^* is a probability measure. Of course $0 \leq \frac{d\tilde{\mu}}{d\mu} \leq 1$ and clearly $\pi^*_x = \left(1 - \frac{d\tilde{\mu}}{d\mu}(x)\right)\pi_x + \frac{d\tilde{\mu}}{d\mu}(x)\chi_x$. Therefore

$$\begin{aligned}\int C(x, \pi^*_x)\mu(dx) &= \int C\left(x, \left(1 - \frac{d\tilde{\mu}}{d\mu}(x)\right)\pi_x + \frac{d\tilde{\mu}}{d\mu}(x)\chi_x\right)\mu(dx) \\ &\leq \int C(x, \pi_x)\mu(dx) + \int [C(x, \chi_x) - C(x, \pi_x)]\tilde{\mu}(dx) \\ &< \int C(x, \pi_x)\mu(dx),\end{aligned}$$

where we used convexity in the first inequality and that $V(\mu, \nu) < \infty$ in the second one. □

5. C-Monotonicity for weak transport costs

Cyclical monotonicity plays a crucial role in classical optimal transport [35, 22]. This has inspired similar development for weak transport costs in [6, 23]:

Definition 5.1 (C-monotonicity)**.** *We say that a coupling $\pi \in \Pi(\mu, \nu)$ is C-monotone if there exists a measurable set $\Gamma \subseteq X$ with $\mu(\Gamma) = 1$, such that for any finite number of points $x_1, \ldots, x_N$ in Γ and measures $m_1, \ldots, m_N$ in $\mathcal{P}(Y)$ with $\sum_{i=1}^N m_i = \sum_{i=1}^N \pi_{x_i}$, the following inequality holds:*

$$\sum_{i=1}^N C(x_i, \pi_{x_i}) \leq \sum_{i=1}^N C(x_i, m_i).$$

We first show that C-monotonicity is necessary for optimality under minimal assumptions. We then provide strengthened assumptions under which C-monotonicity is sufficient.

5.1. ***C*-monotonicity: necessity.** We denote by S_N the set of permutations of the set $\{1, \ldots, N\}$. If $\vec{z} := (z_1 \ldots, z_n)$ is any N-vector, and $\sigma \in S_N$, we naturally overload the notation by defining

$$\sigma(\vec{z}) := (z_{\sigma(1)}, \ldots, z_{\sigma(N)}).$$

Recall the notation (1.1) for the weak transport problem, and the following lemma, which is employed prominently in the proof of Theorem 5.3.

Lemma 5.2 ([10, Proposition 2.1])**.** *Let $X_1, \ldots, X_n$, $n \geq 2$, be Polish spaces equipped with probability measures $\mu_i \in \mathcal{P}(X_i)$, $i = 1, \ldots, n$. Then for any analytic set $B \subseteq X_1 \times \cdots \times X_n$ one of the following holds:*

(a) For every $i = 1, \ldots, n$ there is a μ_i-null set $A_i \subseteq X_i$ s.t.

$$B \subseteq \bigcup_{i=1}^n \mathrm{proj}_{X_i}^{-1}(A_i).$$

(b) There exists a coupling $\pi \in \Pi(\mu_1, \ldots, \mu_n)$ with $\pi(B) > 0$.

The previous lemma is originally stated only for Borel sets, but the same proof technique also works for analytic sets.

Our main result, concerning the necessity of C-monotonicity is the following:

Theorem 5.3. *Let C be jointly measurable and $C(x, \cdot)$ be convex and lower semicontinuous for all x. Assume that π^* is optimal for $V(\mu, \nu)$ and $|V(\mu, \nu)| < \infty$. Then π^* is C-monotone.*

Proof. Let $N \in \mathbb{N}$. Then

$$\mathcal{D}_N := \Big\{((x_1, \ldots, x_N), (m_1, \ldots, m_N)) \in X^N \times \mathcal{P}(Y)^N : \\ \sum_{i=1}^N \pi^*_{x_i} = \sum_{i=1}^N m_i \text{ and } \sum_{i=1}^N C(x_i, \pi^*_{x_i}) > \sum_{i=1}^N C(x_i, m_i)\Big\},$$

is an analytic set. Write

$$D_N := \mathrm{proj}_{X^N}(\mathcal{D}_N).$$

By Jankov-von Neumann uniformization [27, Theorem 18.1] there is an analytically measurable function $f_N\colon D_N \to \mathcal{P}(Y)^N$ such that graph$(f_N) \subseteq \mathcal{D}_N$. We can extend f_N to X^N by defining it on $X^N \setminus D_N$ as the Borel-measurable map $\vec{x} \mapsto (\pi^*_{x_1}, \dots, \pi^*_{x_N})$. Observe that for all $\sigma \in S_N$, we have $(\sigma, \sigma)(\mathcal{D}_N) = \mathcal{D}_N$. Thanks to this, and Lemma 5.4 below, we can assume without loss of generality that f_N satisfies

$$f_N \circ \sigma = \sigma \circ f_N \quad \forall \sigma \in S_N.$$

We write $f^i_N(\vec{x})$ for the i-th element of the vector $f_N(\vec{x}) \in \mathcal{P}(Y)^N$.

Assume that there exists a coupling $Q \in \Pi(\mu^N) = \Pi(\mu, \dots, \mu)$ such that $Q(D_N) > 0$. We now show that this is in conflict with optimality of π^*. We clearly may assume that Q is symmetric, i.e. such that for all $\sigma \in S_N$ we have $Q(B) = Q(\sigma(B))$ for all $B \in \mathcal{B}(X^N)$ (in other words $\sigma(Q) = Q$). We define the possible contender $\tilde{\pi}$ of π^* by

$$\tilde{\pi}(dx_1, dy) := \mu(dx_1) \int_{X^{N-1}} Q_{x_1}(dx_2, \dots, dx_n) f^1_N(x_1, \dots, x_N)(dy), \tag{5.1}$$

which is legitimate owing to all measurability precautions we have taken. We will prove

(1) $\tilde{\pi} \in \Pi(\mu, \nu)$,
(2) $\int \mu(dx) C(x, \pi^*_x) > \int \mu(dx) C(x, \tilde{\pi}_x)$.

Ad (1): Evidently the first marginal of $\tilde{\pi}$ is μ. Write $\sigma_i \in S_N$ for the permutation that merely interchanges the first and i-th component of a vector. By the symmetric properties of Q and f_N we find

$$\begin{aligned}
\int_X \mu(dx_1) \tilde{\pi}_{x_1}(dy) &= \int_{X^N} Q(dx_1, \dots, dx_N) f^1_N(\vec{x})(dy) \\
&= \frac{1}{N} \sum_{i=1}^N \int_{X^N} \sigma_i(Q)(dx_1, \dots, dx_N) f^i_N(\vec{x})(dy) \\
&= \frac{1}{N} \sum_{i=1}^N \int_{X^N} Q(dx_1, \dots, dx_N) \pi^*_{x_i}(dy) \\
&= \nu(dy).
\end{aligned}$$

Ad (2): On D_N holds by construction the strict inequality

$$\sum_{i=1}^N C(x_i, f^i_N(\vec{x})) < \sum_{i=1}^N C(x_i, \pi_{x_i}).$$

Using convexity of $C(x, \cdot)$ and the symmetry properties of Q and f_N, we find

$$\begin{aligned}
\int_X C(x, \tilde{\pi}_x) \mu(dx) &= \int_X \mu(dx_1) C\left(x_1, \int_{X^{N-1}} Q_{x_1}(dx_2, \dots, dx_N) f^1_N(\vec{x})\right) \\
&\leq \int_{X^N} Q(d\vec{x}) C(x_1, f^1_N(\vec{x})) \\
&= \frac{1}{N} \sum_{i=1}^N \int_{X^N} Q(d\vec{x}) C(x_i, f^i_N(\vec{x})) \\
&< \frac{1}{N} \sum_{i=1}^N \int_{X^N} Q(d\vec{x}) C(x_i, \pi_{x_i}) = \int_X C(x, \pi_x) \mu(dx),
\end{aligned}$$

yielding a contradiction to the optimality of π^*.

We conclude that no measure Q with the stated properties exists. By Lemma 5.2, we obtain that D_N is contained in a set of the form $\bigcup_{k=1}^N \mathrm{proj}_k^{-1}(M_N)$ where $\mu(M_N) = 0$ and proj_k denotes the projection from X^N to its k-th component. Since $N \in \mathbb{N}$ was arbitrary, we can define the set $\Gamma := (\bigcup_{N \in \mathbb{N}} M_N)^C$ with $\mu(\Gamma) = 1$, which has the desired property. □

The missing bit in the above proof is Lemma 5.4. By [27, Theorem 7.9] there exists for every Polish space X a closed subset F of the Baire space $\mathcal{N} := \mathbb{N}^{\mathbb{N}}$ and a continuous bijection $h_X\colon F \to X$. On the Baire space the lexicographic order naturally provides a total order. Hence, X inherits the total order of $F \subseteq \mathcal{N}$ by virtue of h_X and its Borel-measurable inverse $h_X^{-1} := g_X$, namely:

$$x, y \in X\colon\ x \le y \Leftrightarrow h_X^{-1}(x) = g_X(x) \le h_X^{-1}(y) = g_X(y).$$

Lemma 5.4. *The set*

$$A = \left\{\vec{x} \in X^N\colon\ x_1 \le x_2 \le \ldots \le x_N\right\},$$

is Borel-measurable. Given $f\colon A \subseteq X^N \to Y^N$ an analytically measurable function, there exists an analytically measurable extension $\hat{f}\colon X^N \to Y^N$ such that for any $\sigma \in S_N$

$$\hat{f} \circ \sigma = \sigma \circ \hat{f}.$$

Proof of Lemma 5.4. Let $\hat{A} = \left\{\vec{a} \in \mathcal{N}^N\colon\ a_1 \le a_2 \le \ldots \le a_N\right\}$, and define $g\colon \mathcal{N}^N \to S_N$ by $g(\vec{a}) = \sigma$ where $\sigma \in S_N$ satisfies

- $\sigma(\vec{a}) \in \hat{A}$
- for each i, j such that $0 \le i < j \le N$ it holds

$$a_i = a_j \implies \sigma(i) < \sigma(j).$$

With these precautions $g(\vec{a}) = \sigma$ is indeed well defined. For each $\sigma \in S_N$ we define also $B_\sigma \subseteq \mathcal{N}^N$ by

$$B_\sigma := \left\{\vec{a} \in \mathcal{N}^N\colon\ g(\vec{a}) = \sigma\right\} = \left\{\vec{a} \in \mathcal{N}^N\colon\ a_{\sigma(1)} \le_\sigma^1 a_{\sigma(2)} \le_\sigma^2 \ldots \le_\sigma^{N-1} a_{\sigma(N)}\right\},$$

where the order $\le_\sigma^i$ is defined depending on σ by

$$\le_\sigma^i := \begin{cases} \le & \sigma(i) \le \sigma(i+1), \\ < & \text{else.} \end{cases}$$

It follows from this representation that B_σ is Borel-measurable. We introduce

$$X^N \ni \vec{x} \mapsto g_X^N(\vec{x}) := (g_X(x_1), g_X(x_2), \ldots, g_X(x_N)) \in F^N \subseteq \mathcal{N}^N.$$

Then the set

$$A_\sigma := \{\vec{x} \in X^N\colon\ g \circ g_X^N(\vec{x}) = \sigma\} = (g_X^N)^{-1}(B_\sigma),$$

is Borel-measurable. In particular, $A_{id} = A$ is Borel-measurable. Note that $\cup_\sigma A_\sigma = X^N$ and $A_{\sigma_1} \cap A_{\sigma_2} = \emptyset$ if $\sigma_1 \not\equiv \sigma_2$. We can apply Lemma 5.5, proving the continuity[2] of

$$\mathcal{N}^N \ni \vec{a} \mapsto G(a) := g(\vec{a})(\vec{a}) \in \mathcal{N}^N.$$

We define the candidate for the desired extension of f by

$$\begin{aligned} &\hat{f}\colon X^N \to Y^N, \\ &\quad \vec{x}\ \mapsto (g \circ g_X^N(\vec{x}))^{-1}\left(f \circ (g_X^N)^{-1} \circ G \circ g_X^N(\vec{x})\right), \end{aligned}$$

which is well defined since $G \circ g_X^N(\vec{x}) \in \hat{A}$, so that $(g_X^N)^{-1} \circ G \circ g_X^N(\vec{x}) \in A$. As a composition of analytically measurable function, $\hat{f}$ inherits this property. It is also clear that $\hat{f}(\vec{x}) = f(\vec{x})$ if $\vec{x} \in A$. Finally, for any $\sigma \in S_N$ and $\vec{x} \in X^N$, we easily find

$$\sigma^{-1}(\hat{f} \circ \sigma(\vec{x})) = \hat{f}(\vec{x}).$$

□

[2]In fact one obtains $\max_{i\in\{1,\ldots,N\}} d_{\mathcal{N}}(g(\vec{a})(\vec{a})_i, g(\vec{b})(\vec{b})_i) \le \max_{i\in\{1,\ldots,N\}} d_{\mathcal{N}}(a_i, b_i)$, for $d_{\mathcal{N}}$ the metric on $\mathcal{N}$ that we recall in Lemma 5.5.

Lemma 5.5. *Let each of $a, b \in \mathcal{N}^N$ be increasing vectors.*[3] *Then for any permutation $\sigma \in S_N$ we have*

$$\max_{i\in\{1,\dots,N\}} d_{\mathcal{N}}(a_i, b_i) \le \max_{i\in\{1,\dots,N\}} d_{\mathcal{N}}(a_i, b_{\sigma(i)}), \tag{5.2}$$

where the metric $d_{\mathcal{N}}$ on $\mathcal{N}$ is given by

$$d_{\mathcal{N}}(a,b) = \begin{cases} 0 & a = b \\ \frac{1}{\min\{n\in\mathbb{N}\colon a(n)\neq b(n)\}} & else. \end{cases}$$

Proof. We show the assertion by induction. For $N = 1$ (5.2) holds trivially. Now assume that (5.2) holds for $N = k$. Given $\sigma \in S_{k+1}$ and $a, b \in \mathcal{N}^{k+1}$ increasing, we know that any $\tilde\sigma \in S_k$ yields

$$\max_{i\in\{1,\dots,k\}} d_{\mathcal{N}}(a_i, b_i) \le \max_{i\in\{1,\dots,k\}} d_{\mathcal{N}}(a_i, b_{\tilde\sigma(i)}).$$

If $\sigma(k+1) = k+1$ the assertion follows by the inductive hypothesis. So let $\sigma(k+1) \neq k+1$ and write $k_1 = \sigma(k+1)$ and $k_2 = \sigma^{-1}(k+1)$. Define a permutation $\hat\sigma \in S_k$ by

$$\hat\sigma(i) = \begin{cases} \sigma(i) & i \neq k_1 \\ k_2 & i = k_1 \end{cases}$$

Since that $a_{k_2} \le a_{k+1}$ and $b_{k_1} \le b_{k+1}$, then

$$\begin{aligned} a_{k_2} \le b_{k_1} &\implies a_{k_2} \le b_{k_1} \le b_{k+1} \implies d_{\mathcal{N}}(a_{k_2}, b_{k_1}) \le d_{\mathcal{N}}(a_{k_2}, b_{k+1}), \\ a_{k_2} \ge b_{k_1} &\implies a_{k+1} \ge a_{k_2} \ge b_{k_1} \implies d_{\mathcal{N}}(a_{k_2}, b_{k_1}) \le d_{\mathcal{N}}(a_{k+1}, b_{k_1}), \end{aligned}$$

and particularly

$$\max_{i\in\{1,\dots,k\}} d_{\mathcal{N}}(a_i, b_{\hat\sigma(i)}) \le \max_{i\in\{1,\dots,k+1\}} d_{\mathcal{N}}(a_i, b_{\sigma(i)}). \tag{5.3}$$

On the other hand, clearly

$$\begin{aligned} a_{k+1} \ge b_{k+1} &\implies d_{\mathcal{N}}(a_{k+1}, b_{k+1}) \le d_{\mathcal{N}}(a_{k+1}, b_{k_1}), \\ a_{k+1} \le b_{k+1} &\implies d_{\mathcal{N}}(a_{k+1}, b_{k+1}) \le d_{\mathcal{N}}(a_{k_2}, b_{k+1}). \end{aligned}$$

This and (5.3) yield $\max_{i\in\{1,\dots,k+1\}} d_{\mathcal{N}}(a_i, b_i) \le \max_{i\in\{1,\dots,k+1\}} d_{\mathcal{N}}(a_i, b_{\sigma(i)})$, so concluding the inductive step. □

5.2. **C-monotonicity: sufficiency.** The conditions under which Theorem 5.3 holds are rather mild. If we assume further continuity properties of C, the next theorem establishes that C-monotonicity is also a sufficient criterion for optimality, resembling the classical case. For weak transport costs, we don't know of any comparable result in the literature.

We recall that, for the given compatible complete metric d_Y on Y, we denote by $\mathcal{W}_1$ the 1-Wasserstein distance [42, Chapter 7].

Theorem 5.6. *Let $\nu \in \mathcal{P}^1_{d_Y}(Y)$. Assume that $C\colon X \times \mathcal{P}^1_{d_Y}(Y) \to \mathbb{R}$ satisfies Condition* (A+) *and is $\mathcal{W}_1$-Lipschitz in the second argument is the sense that for some $L \ge 0$:*

$$|C(x,p) - C(x,q)| \le L\mathcal{W}_1(p,q), \ \ \forall x \in X, \forall p, q \in \mathcal{P}^1_{d_Y}(Y). \tag{5.4}$$

If π is C-monotone then π is an optimizer of $V(\mu,\nu)$.

In the proof we will use the following auxiliary result, which we will establish subsequently:

Lemma 5.7. *Let $\nu \in \mathcal{P}^1_{d_Y}(Y)$. Assume that $C\colon X \times \mathcal{P}^1_{d_Y}(Y) \to \mathbb{R}$ satisfies Condition* (A+) *and is $\mathcal{W}_1$-Lipschitz in the sense of* (5.4). *Then*

$$\inf_{\pi\in\Pi(\mu,\nu)} \int C(x,\pi_x)\,\mu(dx) = \sup_{\substack{\varphi\in\Phi_{b,1} \\ \|\varphi\|_{Lip}\le L}} \mu(R_C\varphi) - \nu(\varphi), \tag{5.5}$$

where $R_C\varphi$ is defined as in (3.1).

[3] A vector $v = (v_i)_{i=1}^N \in \mathcal{N}^N$ is increasing if for any $1 \le i < j \le N$ we have $v_i \le v_j$, where inequality here is meant in the lexicographic order on $\mathcal{N}$.

Proof of Theorem 5.6. Let π be C-monotone. There is an increasing sequence $(K_n)_{n\in\mathbb{N}}$ of compact sets on Y such that $\nu(K_n) \nearrow 1$. From this we can refine the μ-full measurable set Γ in the definition of C-monotonicity, see Definition 5.1, so that for each $x \in \Gamma$ we have $\lim_n \pi_x(K_n) = 1$ and $\pi_x \in \mathcal{P}^1_{d_Y}(Y)$. Our goal is to construct a dual optimizer $\varphi \in \Phi_1$ to π such that

$$\pi_x(\varphi) + C(x,\pi_x) - R_C\varphi(x) = 0 \quad \forall x \in \Gamma.$$

When this is achieved, Theorem 1.3 and the following arguments show that π is optimal as desired:

$$\begin{aligned}\int_X C(x,\pi_x)\mu(dx) &= \int_\Gamma C(x,\pi_x)\mu(dx) = \int_\Gamma [R_C(\varphi)(x) - \pi_x(\varphi)]\mu(dx)\\ &\le \liminf_{k\to-\infty} \int_X [R_C(\varphi\vee k)(x) - \pi_x(\varphi\vee k)]\mu(dx)\\ &\le \sup_{\varphi\in\Phi_{b,1}} \mu(R_C\varphi) - \nu(\varphi)\\ &\le \inf_{\tilde\pi\in\Pi(\mu,\nu)} \int_X C(x,\tilde\pi_x)\mu(dx),\end{aligned}$$

where we used that

$$\liminf_{k\to-\infty} R_C(\varphi\vee k)(x) = \inf_{k\le 0} R_C(\varphi\vee k)(x) = R_C\varphi(x) \quad \forall x \in X.$$

Let us prove the existence of a dual optimizer in Φ_1. Let $G \subseteq \Gamma$ be a finite subset. By definition of C-monotonicity, we conclude that the coupling $\frac{1}{|G|}\sum_{x_i\in G}\delta_{x_i}(dx)\pi_{x_i}(dy)$ is optimal for the weak transport problem determined by the cost C and its first and second marginals. We can apply Lemma 5.7 in this context and obtain

$$\inf_{\|\varphi\|_{Lip}\le L} \sum_{x\in G} \pi_x(\varphi) + C(x,\pi_x) - R_C\varphi(x) = 0. \tag{5.6}$$

We fix $y_0 \in K_1$ and, without loss of generality, find a maximizing sequence $(\varphi_k)_{k\in\mathbb{N}}$ of (5.6) such that for all $k \in \mathbb{N}$ the function φ_k is L-Lipschitz and $\varphi_k(y_0) = 0$. Note that for all $x \in G$

$$\pi_x(\varphi_k) + C(x,\pi_x) - R_C\varphi_k(x) \to 0,$$

since by definition $\pi_x(\varphi_k) + C(x,\pi_x) - R_C\varphi_k(x) \ge 0$. By the Arzelà-Ascoli theorem we find for any $n \in \mathbb{N}$ a subsequence of $(\varphi_k)_{k\in\mathbb{N}}$ and a L-Lipschitz continuous function ψ_n on K_n such that

$$\lim_j \varphi_{k_j}(y) = \psi_n(y) \quad \forall y \in K_n.$$

Thus by a diagonalization argument we can assume without loss of generality that the maximizing sequence converges uniformly for every K_n to a given L-Lipschitz function $\tilde\psi$ defined on

$$A := \bigcup_n K_n.$$

We can extend $\tilde\psi$ from A to all of Y, obtaining an everywhere L-Lipschitz function, via

$$\psi(y) = \inf_{z\in A} \tilde\psi(z) + L d_Y(z,y). \tag{5.7}$$

From (5.7) we find $R_C\psi(x) = \inf_{p\in\mathcal{P}^1_{d_Y}(A)} p(\psi) + C(x,p)$. Indeed, by [16, Proposition 7.50] there is for any $\varepsilon > 0$ an analytically measurable function $T_\varepsilon\colon Y \to A$ with

$$\tilde\psi(T_\varepsilon(y)) + L d_Y(T_\varepsilon(y),y) \le \psi(y) + \varepsilon,$$

from which after integrating with respect to p and using the definition of the Wasserstein distance we deduce

$$p(\psi) - T_\varepsilon(p)(\psi) + C(x,p) - C(x,T_\varepsilon(p)) \ge -\varepsilon + L\mathcal{W}_1(p,T_\varepsilon(p)) + C(x,p) - C(x,T_\varepsilon(p)) \ge -\varepsilon,$$

where we used (5.4) in the last inequality. Therefore, it is actually possible to restrict infimum in $R_C\psi(x)$ to $\mathcal{P}^1_{d_Y}(A)$, and we conclude

$$\limsup_k R_C\varphi_k(x) \le \inf_{p\in\mathcal{P}^1_{d_Y}(A)} p(\psi) + C(x,p) = R_C\psi(x). \tag{5.8}$$

By dominated convergence, and the fact that $\pi_x(A) = 1$, we have

$$\lim_k \pi_x(\varphi_k) = \pi_x(\psi), \tag{5.9}$$

which yields

$$0 = \liminf_k \pi_x(\varphi_k) + C(x,\pi_x) - R_C\varphi_k(x) \geq \pi_x(\psi) + C(x,\pi_x) - R_C\psi(x) \geq 0, \tag{5.10}$$

by definition of $R_C\psi(x)$.

For $G \subseteq Y$ define Ψ_G as the set of all L-Lipschitz continuous functions on A, vanishing at the point y_0, and satisfying

$$\pi_x(\psi) + C(x,\pi_x) - R_C\psi(x) = 0 \quad \forall x \in G.$$

The previous arguments show that, for each finite $G \subseteq \Gamma$, the set Ψ_G is nonempty. We now check that Ψ_G is closed in the topology of pointwise convergence: Let $(\psi_\alpha)_{\alpha\in\mathcal{I}}$ be a net in Ψ_G which converges pointwise to a function φ on A. Since A is the countable union of compact sets, it is possible to extract a sequence $(\psi_{\alpha_k})_{k\in\mathbb{N}}$ of the net such that

$$\psi_{\alpha_k} \to \varphi \quad \text{pointwise on } A \text{ and uniformly on each } K_n,$$

from which φ is L-Lipschitz on A and can be extended to an L-Lipschitz continuous function ψ on Y, see (5.7). By repeating previous arguments (see (5.8), (5.9) and (5.10)) we obtain that $\varphi \in \Psi_G$.

Note that Ψ_G is a closed subset of $\prod_{y\in A}[-Ld(y,y_0), Ld(y,y_0)]$ which is compact in the topology of pointwise convergence by Tychonoff's theorem. Further, the collection $\{\Psi_G : G \subseteq \Gamma, |G| < \infty\}$ satisfies the finite intersection property, since if $G_1,\dots,G_n$ are finite then

$$\bigcap_{i\leq n} \Psi_{G_i} \supseteq \Psi_{\cup_{i\leq n} G_i} \neq \emptyset.$$

Therefore it is possible to find $\varphi \in \bigcap_{G\subseteq\Gamma,\, |G|<\infty} \Psi_G$. Again extend φ, from A to Y, by a L-Lipschitz function as usual. Thus, we have found the desired dual optimizer. □

Proof of Lemma 5.7. By Theorem 1.3 we have

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_X C(x,\pi_x)\mu(dx) = \sup_{\varphi\in\Phi_{b,1}} \mu(R_C\varphi) - \nu(\varphi). \tag{5.11}$$

By Theorem 1.2 we find a minimizer $\pi^* \in \Pi(\mu,\nu)$ of $V(\mu,\nu)$. Now we proceed by taking a maximizing sequence $(\varphi_k)_{k\in\mathbb{N}}$ for the right-hand side of (5.11). Note that we can choose each φ_k, in addition to being below-bounded and continuous, in a way such that it attains its infimum, i.e., there exists $y_k \in Y$ such that

$$-\infty < b_k := \inf_{y\in Y} \varphi_k(y) = \varphi_k(y_k). \tag{5.12}$$

Indeed, this can be done by using e.g. $\varphi_k \vee (b_k + \frac{1}{k})$ instead. Then

$$\lim_k \nu\left(\varphi_k - \varphi_k \vee \left(b_k + \frac{1}{k}\right)\right) = 0, \quad R_C\varphi_k \leq R_C\left(\varphi_k \vee \left(b_k + \frac{1}{k}\right)\right),$$

and the following computation shows that $(\varphi_k \vee (b_k + \frac{1}{k}))_{k\in\mathbb{N}}$ is another maximizing sequence:

$$\begin{aligned} 0 &= \lim_k \int_X [\pi^*_x(\varphi_k) + C(x,\pi^*_x) - R_C\varphi_k(x)]\mu(dx) \\ &\geq \lim_k \int_X \left[\pi^*_x\left(\varphi_k \vee \left(b_k + \frac{1}{k}\right)\right) + C(x,\pi^*_x) - R_C\left(\varphi_k \vee \left(b_k + \frac{1}{k}\right)\right)(x)\right]\mu(dx) \geq 0. \end{aligned}$$

So let φ_k attain its infimum as in (5.12). We want to show that we can choose the sequence to be Lipschitz with constant L. For this purpose we infer additional properties of potential minimizers of $R_C\varphi_k$. Define for each function φ_k the Borel-measurable sets

$$A_k := \left\{ y \in Y \colon \sup_{y\neq z\in Y} \frac{\varphi_k(y) - \varphi_k(z)}{d_Y(y,z)} \leq L \right\} \neq \emptyset,$$

$$\mathcal{Y}_k := \{(y,z) \in Y \times A_k : \varphi_k(y) - \varphi_k(z) > L d_Y(y,z)\}.$$

That $A_k \neq \emptyset$ follows since the minimizers of φ_k form a subset. We also stress that

$$\mathrm{proj}_1(\mathcal{Y}_k) = A_k^c.$$

Indeed, it is apparent that $\mathrm{proj}_1(\mathcal{Y}_k) \subseteq A_k^c$. To see the converse, assume $y \in A_k^c \cap \mathrm{proj}_1(\mathcal{Y}_k)^c$. Define $Z(z') := \{z \in Y : \varphi_k(z') - \varphi_k(z) > L d_Y(z,z')\}$. If there exists $\tilde{z} \in Z(y) \cap A_k$, we obtain a contradiction to $y \in \mathrm{proj}_1(\mathcal{Y}_k)^c$. Let $z_0 := y$ and inductively set $z_l \in Z(z_{l-1})$ such that

$$\inf_{z \in Z(z_{l-1})} \varphi_k(z) + \frac{1}{2^l} \geq \varphi_k(z_l). \tag{5.13}$$

We have for any natural numbers $0 \leq i < n$

$$\varphi_k(z_i) - \varphi_k(z_n) = \sum_{l=i}^{n} \varphi_k(z_{l-1}) - \varphi_k(z_l) > L \sum_{l=i}^{n} d_Y(z_{l-1}, z_l). \tag{5.14}$$

The r.h.s. is bounded from below by $L d_Y(z_i, z_n)$ and so as before we see that $z_n \in A_k$ provides a contradiction. We therefore assume for all l that $z_l \notin A_k$. The above inequality yields by lower-boundedness of φ_k that $(z_l)_{l\in\mathbb{N}}$ is a Cauchy sequence in Y. Writing $\bar{z}$ for its limit point, we conclude from (5.14) that $\varphi_k(z_i) - \varphi_k(\bar{z}) > L d_Y(z_i, \bar{z})$ and consequentely $Z(\bar{z}) \subseteq Z(z_i)$. Since then $\inf\{\varphi_k(z) : z \in Z(z_i)\} \leq \inf\{\varphi_k(z) : z \in Z(\bar{z})\}$ and from (5.13), we deduce $\inf\{\varphi_k(z) : z \in Z(\bar{z})\} \geq \varphi_k(\bar{z})$. Thus $Z(\bar{z}) = \emptyset$, implying $\bar{z} \in A_k$ and yielding a contradiction to $y \in \mathrm{proj}_1(\mathcal{Y}_k)^c$. All in all, we have proven that $A_k^c = \mathrm{proj}_1(\mathcal{Y}_k)$.

By Jankov-von Neumann uniformization [27, Theorem 18.1] there is an analytically measurable selection $T_k : \mathrm{proj}_1(\mathcal{Y}_k) \to A_k$. We set T_k on $A_k = \mathrm{proj}_1(\mathcal{Y}_k)^c$ as the identity. Then T_k maps from Y to A_k and for any $p \in \mathcal{P}^t_{d_Y}(Y)$ we have

$$\begin{aligned}
C(x, T_k(p)) &\leq C(x,p) + L\mathcal{W}_1(p, T_k(p)) \\
&\leq C(x,p) + L \int_Y d_Y(y, T_k(y)) p(dy) \\
&\leq C(x,p) + \int_Y [\varphi_k(y) - \varphi_k(T_k(y))] p(dy) \\
&= C(x,p) + p(\varphi_k) - T_k(p)(\varphi_k).
\end{aligned}$$

Therefore, we can assume that potential minimizers of $R_C\varphi_k$ are concentrated on A_k:

$$R_C\varphi_k(x) = \inf_{p \in \mathcal{P}^1_{d_Y}(Y)} p(\varphi_k) + C(x,p) = \inf_{p \in \mathcal{P}^1_{d_Y}(A_k)} p(\varphi_k) + C(x,p). \tag{5.15}$$

We introduce a family of L-Lipschitz continuous functions by

$$\psi_k(y) := \inf_{z \in A_k} \varphi_k(z) + L d_Y(y,z) = \inf_{z \in Y} \varphi_k(z) + L d_Y(y,z) \quad \forall y \in Y,$$

where equality holds thanks to $\mathrm{proj}_1(\mathcal{Y}_k) = A_k^c$, since for $z \in A_k^c$ we find $(z, \hat{z}) \in \mathcal{Y}_k$, and so

$$\varphi_k(z) + L d_Y(y,z) > \varphi_k(\hat{z}) + L(d_Y(y,z) + d_Y(z,\hat{z})) \geq \varphi_k(\hat{z}) + L d_Y(y,\hat{z}).$$

Then $\varphi_k \geq \psi_k$ where equality holds precisely on A_k. Similarly to before, we find a measurable selection $\hat{T}_k : Y \to A_k$ such that $\psi_k(\hat{T}_k(y)) + L d_Y(y, \hat{T}_k(y)) \leq \psi_k(y) + \varepsilon$. For any $p \in \mathcal{P}^t_{d_Y}(Y)$ we have

$$C(x, \hat{T}_k(p)) \leq C(x,p) + L \int_Y d_Y(y, \hat{T}_k(y)) p(dy) \leq C(x,p) + p(\psi_k) - \hat{T}_k(p)(\psi_k) + \varepsilon.$$

Since ε is arbitrary, by the same argument as in (5.15), we can restrict $\mathcal{P}^1_{d_Y}(Y)$ to $\mathcal{P}^1_{d_Y}(A_k)$ in the definition of $R_C\psi_k$. Hence, $R_C\varphi_k(x) = R_C\psi_k(x)$ and

$$\begin{aligned}\int_X C(x,\pi_x^*)\mu(dx) &= \lim_k \int_X \left[-\pi_x^*(\varphi_k) + R_C\varphi_k(x)\right]\mu(dx)\\ &\le \lim_k \int_X \left[-\pi_x^*(\psi_k) + R_C\psi_k(x)\right]\mu(dx)\\ &\le \lim_k \int_X \left[-\pi_x^*(\psi_k) + \pi_x^*(\psi_k) + C(x,\pi_x^*)\right]\mu(dx)\\ &= \int_X C(x,\pi_x^*)\mu(dx).\end{aligned}$$

□

6. On the Brenier-Strassen Theorem of Gozlan and Juillet

In this part we take $X = Y = \mathbb{R}^d$, equipped with the Euclidean metric, and

$$C_\theta(x,\rho) := \theta\left(x - \int y\rho(dy)\right),$$

where $\theta : \mathbb{R}^d \to \mathbb{R}_+$ is convex. As usual we denote by $V(\cdot,\cdot)$ the value of the weak transport problem with this cost functional (see (1.1)). We have

Lemma 6.1. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\nu \in \mathcal{P}^1(\mathbb{R}^d)$. Then*

$$\inf_{\eta\le_c\nu}\ \inf_{\pi\in\Pi(\mu,\eta)} \int \theta(x-z)\pi(dx,dz) = V(\mu,\nu). \tag{6.1}$$

Proof. Given π feasible for $V(\mu,\nu)$, we define $T(x) := \int y\pi^x(dy)$ and notice that $T(\mu) \le_c \nu$ by Jensen's inequality. From this we deduce that the l.h.s. of (6.1) is smaller than the r.h.s. For the reverse inequality, let $\varepsilon > 0$ and say $\bar\eta \le_c \nu$ is such that

$$\inf_{\eta\le_c\nu}\ \inf_{\pi\in\Pi(\mu,\eta)} \int \theta(x-z)\pi(dx,dz) + \varepsilon \ge \inf_{\pi\in\Pi(\mu,\bar\eta)} \int \theta(x-z)\pi(dx,dz) \ge \int \theta(x-z)\bar\pi(dx,dz) - \varepsilon,$$

for some $\bar\pi \in \Pi(\mu,\bar\eta)$. By Strassen theorem there is a martingale measure $m(dz,dy)$ with first marginal $\bar\eta$ and second marginal ν. Define $\pi(dx,dy) := \int_z \bar\pi^z(dx)m^z(dy)\bar\eta(dz)$, so then π has x-marginal μ and y-marginal ν, and furthermore $\int y\pi^x(dx) = \int z\bar\pi^x(dx)$ (μ-a.s.), by the martingale property of m. Thus, by Jensen's inequality:

$$\int \theta(x-z)\bar\pi_x(dz)\mu(dx) \ge \int \theta\left(x - \int z\bar\pi_x(dz)\right)\mu(dx) = \int \theta\left(x - \int y\pi_x(dy)\right)\mu(dx) \ge V(\mu,\nu).$$

Taking $\varepsilon \to 0$ we conclude. □

We now provide the proof of Theorem 1.4, in which case $\theta(\cdot) = |\cdot|^2$:

Proof of Theorem 1.4. We have $V(\mu,\nu) < \infty$, since the product coupling yields a finite cost. Lemma 6.1 established the rightmost equality in (1.6). The existence of an optimizer π to $V(\mu,\nu)$ follows from Theorem 1.2. By the necessary monotonicity principle (Theorem 5.3) there exists a measurable set $\Gamma \subseteq X$ with $\mu(\Gamma) = 1$ such that for any finite number of points $x_1,\dots,x_N$ in Γ and measures $m^1,\dots,m^N$ in $\mathcal{P}(\mathbb{R}^d)$ with $\sum_{i=1}^N m^i = \sum_{i=1}^N \pi^{x_i}$ the following inequality holds:

$$\sum_{i=1}^N \left|x^i - \int y\pi^{x^i}(dy)\right|^2 \le \sum_{i=1}^N \left|x^i - \int y m^i(dy)\right|^2. \tag{6.2}$$

In particular, if we let

$$T(x) := \int y\pi_x(dy),$$

and σ is any permutation, then

$$\sum_i \left|x^i - T(x^i)\right|^2 \leq \sum_{i=1}^N \left|x^i - T(x^{\sigma(i)})\right|^2. \tag{6.3}$$

Let us introduce $p(dx, dz) := \mu(dx)\delta_{T(x)}(dz)$ and observe that its z-marginal is $T(\mu)$. By Rockafellar's theorem ([42, Theorem 2.27]) the support of p is contained in the graph of the subdifferential of a closed convex function. Then by the Knott-Smith optimality criterion ([42, Theorem 2.12]) the coupling p attains $\mathcal{W}_2(\mu, T(\mu))$. Since by Jensen clearly $T(\mu) \leq_c \nu$, this establishes the remaining equality in (1.6) and shows further that $V(\mu, \nu) = \mathcal{W}_2(\mu, T(\mu))^2$ and $\mu^* := T(\mu)$. The uniqueness of μ^* follows the same argument as in the proof of [23, Proposition 1.1].

We can use (6.2) and argue verbatim as in [23, Remark 3.1] showing that T is actually 1-Lipschitz on Γ. We will now prove that T is (μ-a.s. equal to) the gradient of a continuously differentiable convex function. The key remark is that the coupling p is also optimal for $V(\mu, T(\mu))$. Indeed, we have

$$V(\mu, \nu) \leq \inf_{\eta \leq_c T(\mu)} \mathcal{W}_2(\mu, \nu)^2 = V(\mu, T(\mu)) \leq \int |x - T(x)|^2 \mu(dx) = V(\mu, \nu).$$

Take any $\mathcal{W}_2$-approximative sequence $(\mu^k)_{k\in\mathbb{N}}$ of μ such that for all $k \in \mathbb{N}$

$$\mu^k \ll \lambda \ll \mu^k,$$

where λ denotes the d-dimensional Lebesgue measure. This can be easily achieved by scaled convolution with a non-degenerate Gaussian kernel. By stability of the considered weak transport problem [8, Theorem 1.5], and using the previously shown, we obtain for each μ^k a 1-Lipschitz map T^k defined this time everywhere in $\mathbb{R}^d$ with

$$\mathcal{W}_2(\mu^k, T^k(\mu^k))^2 = V(\mu^k, \nu),$$

and $T^k(\mu^k) \to T(\mu)$ in $\mathcal{W}_1$. By Brenier's theorem [42, Theorem 2.12 (*ii*)] we find for each $k \in \mathbb{N}$ some convex function $\varphi^k \colon \mathbb{R}^d \to \mathbb{R}$, $\varphi(0) = 0$, and $\nabla\varphi^k(x) = T^k(x)$ λ-a.e. x. By continuity of T^k we have $\nabla\varphi^k(x) = T^k(x)$ for all $x \in \mathbb{R}^d$.

We want to show that $(\varphi^k)_{k\in\mathbb{N}}$ is suitably relatively compact. By tightness of μ^k and $T^k(\mu^k)$ we find compact sets $K_1, K_2 \subseteq \mathbb{R}^d$ with

$$\inf_k \mu^k(K_1) > \frac{1}{2}, \quad \inf_k T^k(\mu^k)(K_2) > \frac{1}{2}.$$

In particular, the sets $(T^k(K_1) \cap K_2)_{k\in\mathbb{N}}$ are all non-empty. The compactness of K_1 and K_2, and the 1-Lipschitz property of each T^k, imply then the existence of $x \in K_1$ such that $\sup_k |T^k(x)| < \infty$. Hence, $(T^k)_{k\in\mathbb{N}}$ is pointwise bounded and uniformly 1-Lipschitz. Thanks to Arzelà-Ascoli's theorem and a diagonalization argument, we can select a subsequence $(T^{k_j})_{j\in\mathbb{N}}$ of $(T^k)_{k\in\mathbb{N}}$ which converges locally uniformly to some 1-Lipschitz function $\tilde{T} \colon \mathbb{R}^d \to \mathbb{R}^d$. Since being a gradient field is preserved under locally uniform limits, we have that $\tilde{T}$ is a gradient field, and φ^{k_j} converges pointwise to some φ with $\varphi(0) = 0$ and $\nabla\varphi = \tilde{T}$. In particular φ is convex and of class $C^1(\mathbb{R}^d)$.

Finally, for any $f \in C_b(\mathbb{R}^d)$ and $\varepsilon > 0$, we find an index $j_0 \in \mathbb{N}$ such that for all $j \geq j_0$:

$$|T^{k_j}(\mu^{k_j})(f) - \tilde{T}(\mu)(f)| \leq |T^{k_j}(\mu^{k_j})(f) - \tilde{T}(\mu^{k_j})(f)| + |\tilde{T}(\mu^{k_j})(f) - \tilde{T}(\mu)(f)| < \varepsilon,$$

where the first summand can be chosen sufficiently small for large j by locally uniform convergence of T^{k_k} to $\tilde{T}$ and the second one by weak convergence of μ^{k_j} to μ. All in all, we deduce that $T^{k_j}(\mu^{k_j})$ converges weakly to $\tilde{T}(\mu)$, which must therefore match $T(\mu)$. Furthermore, $\mu(dx)\delta_{\tilde{T}(x)}(dy)$ defines an optimizer for the weak transport problem (1.1) between μ and ν with cost (1.5). By uniqueness of the optimizers we conclude $T = \tilde{T}$ μ-almost surely. In particular, T is μ-almost everywhere the gradient of the convex function $\varphi \in C^1(\mathbb{R}^d)$. □

References

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

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

[3] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. Adapted Wasserstein Distances and Stability in Mathematical Finance. *arXiv e-prints*, page arXiv:1901.07450, Jan 2019.

[4] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. All Adapted Topologies are Equal. *arXiv e-prints*, page arXiv:1905.00368, May 2019.

[5] J. Backhoff-Veraguas, M. Beiglböck, M. Eder, and A. Pichler. Fundamental properties of process distances. *ArXiv e-prints*, 2017.

[6] J. Backhoff-Veraguas, M. Beiglböck, M. Huesmann, and S. Källblad. Martingale Benamou–Brenier: a probabilistic perspective. *ArXiv e-prints*, Aug. 2018.

[7] J. Backhoff-Veraguas, M. Beiglböck, Y. Lin, and A. Zalashko. Causal transport in discrete time and applications. *SIAM Journal on Optimization*, 27(4):2528–2562, 2017.

[8] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Weak monotone rearrangement on the line. *arXiv e-prints*, page arXiv:1902.05763, Feb 2019.

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

[10] M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer. Optimal and better transport plans. *J. Funct. Anal.*, 256(6):1907–1927, 2009.

[11] M. Beiglböck and C. Griessler. A land of monotone plenty. *Annali della SNS, to appear*, Apr. 2016.

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

[13] M. Beiglböck, M. Nutz, and N. Touzi. Complete Duality for Martingale Optimal Transport on the Line. *Ann. Probab., to appear*, 2016.

[14] M. Beiglboeck, M. Eder, C. Elgert, and U. Schmock. Geometry of Distribution-Constrained Optimal Stopping Problems. *Probab. Theory Relat. Fields, to appear*, 2018.

[15] M. Beiglboeck and N. Juillet. Shadow couplings. *ArXiv e-prints*, Sept. 2016.

[16] D. P. Bertsekas and S. E. Shreve. *Stochastic optimal control*, volume 139 of *Mathematics in Science and Engineering*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1978. The discrete time case.

[17] M. Bowles and N. Ghoussoub. A Theory of Transfers: Duality and convolution. *arXiv e-prints*, page arXiv:1804.08563, Apr 2018.

[18] M. Colombo, L. De Pascale, and S. Di Marino. Multimarginal optimal transport maps for one-dimensional repulsive costs. *Canad. J. Math.*, 67(2):350–368, 2015.

[19] C. Daskalakis, A. Deckelbaum, and C. Tzamos. Strong Duality for a Multiple-Good Monopolist. *Econometrica*, 85(3):735–767, 2017.

[20] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations*, volume 902. John Wiley & Sons, 2011.

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

[22] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[23] N. Gozlan and N. Juillet. On a mixture of brenier and strassen theorems. *arXiv preprint arXiv:1808.02681*, 2018.

[24] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali. Characterization of a class of weak transport-entropy inequalities on the line. *Ann. Inst. Henri Poincaré Probab. Stat.*, 54(3):1667–1693, 2018.

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

[26] C. Griessler. c-cyclical monotonicity as a sufficient criterion for optimality in the multi-marginal Monge-Kantorovich problem. *ArXiv e-prints*, Jan. 2016.

[27] A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[28] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.

[29] K. Marton. A measure concentration inequality for contracting markov chains. *Geometric & Functional Analysis GAFA*, 6(3):556–571, 1996.

[30] K. Marton et al. Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration. *The Annals of Probability*, 24(2):857–866, 1996.

[31] M. Nutz and F. Stebegg. Canonical Supermartingale Couplings. *Ann. Probab., to appear*, Sept. 2018.

[32] B. Pass. On the local structure of optimal measures in the multi-marginal optimal transportation problem. *Calc. Var. Partial Differential Equations*, 43(3-4):529–536, 2012.

[33] G. C. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM J. Optim.*, 22(1):1–23, 2012.

[34] S. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I*. Probability and its Applications (New York). Springer-Verlag, New York, 1998.
[35] S. T. Rachev and L. Rüschendorf. A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.*, 32(1):48–54, 1990.
[36] P.-M. Samson. Transport-entropy inequalities on locally acting groups of permutations. *Electron. J. Probab.*, 22:Paper No. 62, 33, 2017.
[37] Y. Shu. From hopf-lax formula to optimal weak transfer plan. *arXiv preprint arXiv:1609.03405*, 2016.
[38] Y. Shu. Hamilton-Jacobi equations on graph and applications. *Potential Anal.*, 48(2):125–157, 2018.
[39] A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer, 1991.
[40] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 81(1):73–205, 1995.
[41] M. Talagrand. New concentration inequalities in product spaces. *Inventiones mathematicae*, 126(3):505–563, 1996.
[42] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
[43] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.
[44] D. Zaev. On the Monge-Kantorovich problem with additional linear constraints. *ArXiv e-prints*, 2014.
[45] C. Zalinescu. *Convex analysis in general vector spaces*. World scientific, 2002.

WEAK MONOTONE REARRANGEMENT ON THE LINE

J. BACKHOFF-VERAGUAS, M. BEIGLBÖCK, AND G. PAMMER

Abstract. Weak optimal transport has been recently introduced by Gozlan et al. The original motivation stems from the theory of geometric inequalities; further applications concern numerics of martingale optimal transport and stability in mathematical finance.

In this note we provide a complete geometric characterization of the *weak* version of the classical monotone rearrangement between measures on the real line, complementing earlier results of Alfonsi, Corbetta, and Jourdain.

Keywords: Convex order, mass transport, martingale transport, monotone rearrangement.
Mathematics Subject Classification (2010): 60E15, 60G42, 60J05.

1. Introduction

Recently, there has been a growing interest in weak transport problems as introduced by Gozlan, Roberto, Samson and Tetali [14]. While the original motivation mainly stems from applications to geometric inequalities (cf. Marton [16, 15] and Talagrand [21, 22]), weak transport problems appear also in a number of further topics, including martingale transport [2, 3, 9, 5], the causal transport problem [6, 1], and stability in mathematical finance [4].

1.1. **Framework and main results.** Write $\Pi(\mu,\nu)$ for the set of couplings between $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$. Starting with the seminal article of Gangbo-McCann [11] problems of the form

$$W_\theta(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} \theta(x-y)\pi(dx,dy), \tag{1.1}$$

where $\theta\colon \mathbb{R}^d \to \mathbb{R}$ denotes a convex function have received particular attention in optimal transport. A related problem that is considered in the weak transport literature is

$$V_\theta(\mu,\nu) := \inf_{\mu^*\leq_c\nu} W_\theta(\mu,\mu^*). \tag{1.2}$$

Here $\leq_c$ denotes convex order, i.e. $\mu \leq_c \nu$ iff $\int \varphi\, d\mu \leq \int \varphi\, d\nu$ for all convex $\varphi : \mathbb{R}^d \to \mathbb{R}$.

The problem (1.2), and in particular its one dimensional version, is investigated in [2, 14, 13, 19, 17, 20, 10, 12, 7, 4]. The main purpose of this note is to give a complete geometric characterization of the optimizer μ^* in one dimension.

We write $\mathcal{P}_\rho(\mathbb{R}^d)$ for the set of probability measures on $\mathbb{R}^d$ which have finite ρ-moment.

Definition 1.1. *Fix $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$. We call a function $S\colon \mathbb{R}\to\mathbb{R}$ admissible if it satisfies*

(i) S is increasing,
(ii) S is 1-Lipschitz,
(iii) $S(\mu) \leq_c \nu$.

MB gratefully acknowledges support by FWF-Grant Y00782. GP acknowledges support from the AustrianScience Fund (FWF) through Grant No. W 1245. All authors thank an anonymous referee for insightful comments that lead to a significant improvement of the article.

Theorem 1.2. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. There exists an admissible T (μ-a.s. unique) which is maximal in the sense that $S(\mu) \leq_c T(\mu)$ for every other admissible S.*

If $\theta\colon \mathbb{R} \to \mathbb{R}$ is convex then $\mu^* := T(\mu)$ is an optimizer of (1.2). If $\theta\colon \mathbb{R} \to \mathbb{R}$ is strictly convex and $V_\theta(\mu, \nu)$ is finite, $T(\mu)$ is the unique optimizer of (1.2).

We call the (μ-a.s. unique) map T in Theorem 1.2 the *weak monotone rearrangement.*

A consequence of Theorem 1.2 is that the optimizer of (1.2) does not depend on the choice of the convex function θ. We find this fact non-trivial as well as remarkable and highlight that it is not new: different independent proofs were given by Gozlan, Roberto, Samson and Tetali [14], Alfonsi, Corbetta, Jourdain [2] and Shu [20]. For $d > 1$ the optimizer of (1.2) may depend on θ, see [2, Example 2.4].

The map T can be explicitly characterized in geometric terms using the notion of irreducibility introduced in [8]: Measures $\eta, \nu \in P_1(\mathbb{R})$ are in convex order iff their potential functions satisfy

$$u_\eta(y) := \textstyle\int_{\mathbb{R}} |x-y| \eta(dx) \leq \int_{\mathbb{R}} |x-y| \nu(dx) =: u_\nu(y). \tag{1.3}$$

By continuity, the set U where this inequality is strict is open. Hence $U = \bigcup_n I_n$, where (I_n) is an at most countable family of disjoint open intervals; these intervals I_n are called *irreducible* with respect to (η, ν).

Theorem 1.3. *The weak monotone rearrangement T of $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ is the unique admissible map which has slope 1 on each interval $T^{-1}(I)$, where I is irreducible w.r.t. $(T(\mu), \nu)$.*

We emphasize that [2, Proposition 4.3] yields that every optimiser μ^* of (1.2) is of the form $T(\mu)$ for an admissible map T and thus implies a part of Theorem 1.2. Likewise, it can be seen from [2, Proposition 4.3] that this mapping T has slope 1 on each interval $T^{-1}(I)$, I irreducible w.r.t. $(T(\mu), \nu)$. The novelty of Theorem 1.3 is that there exists only one admissible mapping with this irreducibility property.

1.2. **Connection with martingale transport plans.** Intuitively, the irreducible intervals of (η, ν) are the components where we need to 'expand' η in order to transform it into ν. In this sense Theorem 1.3 asserts that the mass of μ can *either* concentrate between μ and $T(\mu)$, *or* it can expanded between $T(\mu)$ and ν (see Figure 1).

To make this precise, we recall from [14] that (1.2) can be reformulated as

$$V_\theta(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \textstyle\int_{\mathbb{R}^d} \theta\Big(x - \int_{\mathbb{R}^d} y \pi_x(dy)\Big) \mu(dx), \tag{1.4}$$

where $(\pi_x)_{x \in \mathbb{R}^d}$ denotes a regular disintegration of coupling π w.r.t. its first marginal μ.

The set of optimizers of (1.4) is also straightforward to express in terms of T: Write $\Pi_M(\eta, \nu)$ for the set of *martingale couplings* (or martingale transport plans), i.e. $\pi \in \Pi(\eta, \nu)$ which satisfy barycenter$(\pi_x) = x$, η-a.s. By Strassen's theorem $\Pi_M(\eta, \nu)$ is nonempty iff $\eta \leq_c \nu$. Using this notation, $\pi \in \Pi(\mu, \nu)$ is optimal iff there exists a martingale coupling $\pi^M \in \Pi_M(T(\mu), \nu)$ such that π is the concatenation of the transports described by T and π^M:

$$\pi(A \times B) = \textstyle\int_A \mu(dx) \pi^M_{T(x)}(B). \tag{1.5}$$

Any $\pi^M \in \Pi_M(\eta, \nu)$ can be decomposed based on the family of irreducible intervals $(I_n)_n$: denoting $F := (\cup_n I_n)^c$ by [8, Appendix A] we have

$$\pi^M = \textstyle\sum_n \pi^M_{|I_n \times \bar{I}_n} + (\mathrm{Id}, \mathrm{Id})(\mu_{|F}). \tag{1.6}$$

Plainly, (1.6) asserts that any martingale transport plan can move mass only within the individual irreducible intervals, whereas particles $x \in F$ have to stay put.

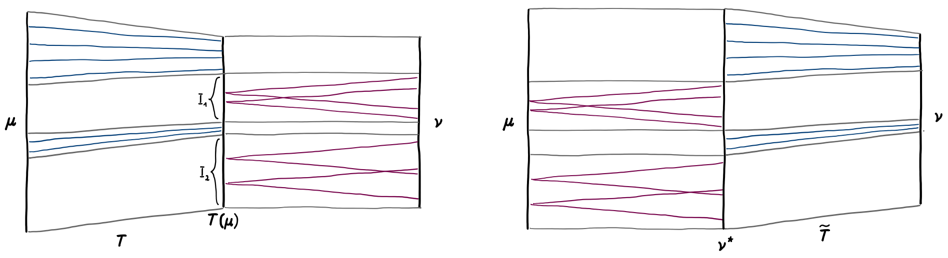

FIGURE 1. The solutions to (1.2)/(1.4) and (1.7), respectively. Blue lines depict contractive parts of the map, purple lines depict areas with (non trivial) martingale transport.

1.3. **A reverse problem.** Alfonsi, Corbetta, and Jourdain [2] proved that the same value is obtained when reversing the order of transport and convex order relaxation in (1.2), i.e.

$$V_\theta(\mu,\nu) = \inf_{\mu \le_c \nu^*} W_\theta(\nu^*,\nu); \tag{1.7}$$

moreover they find ([2, Proposition 4.3]) a monotone mapping which is optimal between the optimizer of (1.7) and ν as well as between μ and the optimizer of (1.2).

As a counterpart to Theorems 1.2 and 1.3 we establish the following result, strengthening the connection between (1.2) and (1.7).

Theorem 1.4. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$. Then there exists a unique $\le_c$-smallest measure ν^*, $\mu \le_c \nu^*$, which can be pushed onto ν by an increasing 1-Lipschitz mapping. Moreover we have*

(1) If $\theta\colon \mathbb{R} \to \mathbb{R}$ is convex then ν^* is an optimizer of (1.7). If $\theta\colon \mathbb{R}\to\mathbb{R}$ is strictly convex and $V_\theta(\mu,\nu)$ is finite, ν^* is the unique optimizer of (1.7).
(2) There exists a (ν^*-unique) increasing 1-Lipschitz mapping $\tilde{T}$ which pushes ν^* onto ν; $\tilde{T}$ has slope 1 on each interval I, where I is irreducible w.r.t. (μ,ν^*).
(3) $\tilde{T}$ is the weak monotone rearrangement between μ and ν.

As in the previous section, (1.7) could be interpreted as a concatenation of a martingale transport with the weak monotone rearrangement.

1.4. **An auxiliary result.** We close this introductory section with an auxiliary result that will be important in the proofs of our results. Since it might be of independent interest we provide the d-dimensional version. We denote the topology induced by the ρ-Wasserstein distance on the space of probability measures on $\mathbb{R}^d$ by $\mathcal{W}_\rho$.

Theorem 1.5 (Stability)**.** *Let $1 \le \rho < \infty$, $(\mu^k)_{k\in\mathbb{N}} \in \mathcal{P}_\rho(\mathbb{R}^d)^{\mathbb{N}}$, $(\nu^k)_{k\in\mathbb{N}} \in \mathcal{P}_1(\mathbb{R}^d)^{\mathbb{N}}$ and $\theta\colon \mathbb{R}^d \to \mathbb{R}$ convex and such that for some constant $c > 0$ it holds*

$$\theta(x) \le c(1+|x|^\rho) \quad \forall x \in \mathbb{R}^d. \tag{1.8}$$

If $\mu^k \to \mu$ in $\mathcal{W}_\rho$ and $\nu^k \to \nu$ in $\mathcal{W}_1$, then $\lim_k V_\theta(\mu^k,\nu^k) = V_\theta(\mu,\nu)$. If additionally θ is strictly convex, we have that

(1) $\arg\min_{\eta\le_c \nu^k} W_\theta(\mu^k,\nu^k) \to \arg\min_{\eta \le_c \nu} W_\theta(\mu,\nu)$ in $\mathcal{W}_1$,
(2) the sequence of maps T^k, where $T^k(\mu) \le_c \nu$ and $V_\theta(\mu,\nu^k) = W_\theta(\mu,T^k(\mu))$, converges in μ-probability to T, where $T(\mu) \le_c \nu$ and $V_\theta(\mu,\nu) = W_\theta(\mu,T(\mu))$.

2. C-Monotonicity implies geometric characterization

In this part we prove the following

Theorem 2.1. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ and $\theta \colon \mathbb{R} \to \mathbb{R}$ strictly convex. If $V_\theta(\mu,\nu)$ yields a finite value, for any optimizer $\pi \in \Pi(\mu,\nu)$ of (1.4), the map*

$$T(x) := \int_{\mathbb{R}} y \pi_x(dy),$$

is independent of the specific optimizer π. This map is further admissible in the sense of Definition 1.1 and it has slope 1 on $T^{-1}(I)$ if I is an irreducible interval for $(T(\mu),\nu)$.

To prove Theorem 2.1, we need some further properties connected to irreducibility:

Lemma 2.2. *Suppose $\{u_\mu < u_\nu\} =: (a,b)$. Then for any $\pi \in \Pi_M(\mu,\nu)$, any regular disintegration $(\pi_x)_{x\in\mathbb{R}}$ w.r.t. μ and any $c \in (a,b)$ such that $\mu((a,c)) > 0$, $\mu((c,b)) > 0$, there are $x \in (a,c]$, $y \in [c,b)$, $x \neq y$, such that the supports of π_x and π_y overlap, i.e.*

$$\operatorname{int}(\operatorname{co}(\operatorname{supp}(\pi_x))) \cap \operatorname{co}(\operatorname{supp}(\pi_y)) \cup \operatorname{int}(\operatorname{co}(\operatorname{supp}(\pi_y))) \cap \operatorname{co}(\operatorname{supp}(\pi_x)) \neq \emptyset. \quad (2.1)$$

Proof. To show this assertion, we assume the opposite. So there exist $c \in \{u_\mu < u_\nu\}$, $\pi \in \Pi_M(\mu,\nu)$ with fixed disintegration $(\pi_x)_{x\in\mathbb{R}}$ w.r.t. μ and

$$\mu((a,c)) > 0, \quad \mu((c,b)) > 0,$$

so that for all x, y with $a < x \leq c \leq y < b$, $x \neq y$, we have

$$\operatorname{int}(\operatorname{co}(\operatorname{supp}(\pi_x))) \cap \operatorname{co}(\operatorname{supp}(\pi_y)) \cup \operatorname{int}(\operatorname{co}(\operatorname{supp}(\pi_y))) \cap \operatorname{co}(\operatorname{supp}(\pi_x)) = \emptyset. \quad (2.2)$$

Since π is a martingale coupling, and by (2.2), there exists $d \in (a,b)$ with

$$\begin{aligned} \operatorname{supp}(\pi_x) &\subseteq (-\infty, d] \quad \text{for } \mu\text{-a.e. } x < c, \\ \operatorname{supp}(\pi_y) &\subseteq [d, \infty) \quad \text{for } \mu\text{-a.e. } y > c. \end{aligned} \quad (2.3)$$

Write d_+ for the largest and d_- for the smallest d such that (2.3) holds. Note then that $d_-, d_+ \in (a,b)$. We have either $\operatorname{supp}(\pi_c) \subseteq [d_-, d_+]$ or $\mu(\{c\}) = 0$, which in any case implies $\mu([c \wedge d_-, c \vee d_+] \setminus \{c\}) = 0$. Thus, we infer

$$\begin{aligned} 1 &= \pi((-\infty,c) \times (-\infty,d_-) \cup \{c\} \times [d_-,d_+] \cup (c,\infty) \times (d_+,\infty)) \\ &= \pi((-\infty,d_-) \times (-\infty,d_-] \cup \{c\} \times [d_-,d_+] \cup (d_+,\infty) \times [d_+,\infty)), \end{aligned}$$

and we conclude by contradicting $u_\mu(d_-) < u_\nu(d_-)$ since

$$\begin{aligned} \int_{\mathbb{R}} |x - d_-| \mu(dx) &= \int_{(a,d_-)} |x - d_-| \mu(dx) + \int_{[d_-,b)} |x - d_-| \mu(dx) \\ &= \int_{(a,d_-)} |y - d_-| \pi_x(dy) \mu(dx) + \int_{[d_-,b)} |y - d_-| \pi_x(dy) \mu(dx) \\ &= \int_{\mathbb{R}} |x - d_-| \nu(dx). \end{aligned}$$

□

Lemma 2.3. *Let $p, q \in \mathcal{P}_1(\mathbb{R})$ have overlapping supports (cf. (2.1)). Then there exists a continuous map $[0,1] \ni \alpha \mapsto (p_\alpha, q_\alpha) \in \mathcal{P}(\mathbb{R}) \times \mathcal{P}(\mathbb{R})$ such that*

$$p_\alpha + q_\alpha = p + q \quad \forall \alpha \in [0,1], \quad (2.4)$$

and such that for some $\beta \in (0,1)$ the functions

$$[\beta, 1] \ni \alpha \mapsto \int_{\mathbb{R}} z p_\alpha(dz), \quad [\beta, 1] \ni \alpha \mapsto \int_{\mathbb{R}} z q_\alpha(dz) \quad (2.5)$$

are strictly increasing and decreasing, respectively.

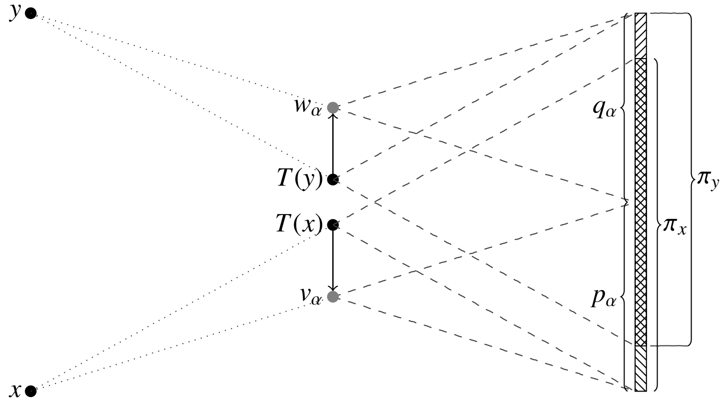

FIGURE 2. Sketch of usage of Lemma 2.2 and Lemma 2.3 to find contradiction to C-monotonicity of an C-optimal coupling π.

Proof. Let $\alpha \in [0,1]$ and define the inverse distribution functions by

$$s_\alpha := \inf\{x \in \mathbb{R} \colon F_p(x) \geq \alpha\}, \quad t_\alpha := \inf\{x \in \mathbb{R} \colon F_q(x) \geq \alpha\},$$

where F_p and F_q denote the cumulative distribution functions of p and q, respectively. Define two auxiliary measures

$$\tilde{p}_\alpha := p|_{(-\infty,s_\alpha)} + (\alpha - F_p(s_\alpha-))\delta_{\{s_\alpha\}}, \quad \tilde{q}_\alpha := q|_{(-\infty,t_\alpha)} + (\alpha - F_q(t_\alpha-))\delta_{\{t_\alpha\}}.$$

We define the probability measure p_α by taking the leftmost α mass of p and the leftmost $1-\alpha$ mass of q, i.e., let p_α be given by $p_\alpha := \tilde{p}_\alpha + \tilde{q}_{1-\alpha}$. To satisfy (2.4), we set $q_\alpha := p + q - p_\alpha$, which also yields continuity of $\alpha \mapsto (p_\alpha, q_\alpha)$. Since the supports of p and q are overlapping, see (2.1), we find constants $c_1, c_2 \in \mathbb{R}$ with $c_1 > c_2$ such that

$$\alpha_1 := p([c_1, +\infty)) > 0 \text{ and } \alpha_2 := q((-\infty, c_2]) > 0.$$

Let $0 < \alpha_3 < \alpha_1 \wedge \alpha_2$, then for any $1 - \alpha_3 \leq \alpha < \alpha' \leq 1$ we have

$$\begin{aligned} \textstyle\int_{\mathbb{R}} z p_{\alpha'}(dz) - \int_{\mathbb{R}} z p_\alpha(dz) &= \textstyle\int_{\mathbb{R}} z\ (\tilde{p}_{\alpha'} - \tilde{p}_\alpha)(dz) + \int_{\mathbb{R}} z\ (\tilde{q}_{1-\alpha'} - \tilde{q}_{1-\alpha})(dz) \\ &\geq \textstyle\int_{\mathbb{R}} c_1(\tilde{p}_{\alpha'} - \tilde{p}_\alpha)(dz) - \int_{\mathbb{R}} c_2(\tilde{q}_{1-\alpha} - \tilde{q}_{1-\alpha'})(dz) \\ &= (\alpha' - \alpha)(c_1 - c_2) > 0, \end{aligned}$$

and conclude that for $\beta := 1 - \alpha_3$ the maps defined in (2.5) are strictly monotone. □

An important tool in the proof of Theorem 2.1 is C-monotonicity, a concept which was introduced for the weak optimal transport problem in [5, 12, 7]. In the current setting it takes the following form:

Definition 2.4 (C-monotonicity)**.** *A coupling $\pi \in \Pi(\mu,\nu)$ is C-monotone (with cost $C(x,p) = \theta\big(x - \int yp(dy)\big)$) if there exists a measurable set $\Gamma \subseteq \mathbb{R}$ with $\mu(\Gamma) = 1$, such that for any finite number of points $x_1, \ldots, x_N$ in Γ and probability measures $m_1, \ldots, m_N \in \mathcal{P}_1(\mathbb{R})$ with $\sum_{i=1}^N m_i = \sum_{i=1}^N \pi_{x_i}$, we have*

$$\textstyle\sum_{i=1}^N \theta\left(x_i - \int y\, \pi_{x_i}(dy)\right) \leq \sum_{i=1}^N \theta\left(x_i - \int y\, m_i(dy)\right).$$

Proof of Theorem 2.1. Let π^* and $\tilde{\pi}^*$ be optimal for the weak optimal transport problem (1.4). Clearly, $\frac{\pi^* + \tilde{\pi}^*}{2}$ constitutes another optimizer of (1.4). Strict convexity of θ yields that

$$\textstyle\int_{\mathbb{R}} y \pi^*_x(dy) = \int_{\mathbb{R}} y \frac{\pi^* + \tilde{\pi}^*}{2}(dy) = \int_{\mathbb{R}} y \tilde{\pi}^*_x(dy) \quad \mu\text{-a.e. } x,$$

showing that T is independent of the choice of the optimizer.

By the monotonicity principle [7, Theorem 5.2] π^* is C-monotone, therefore, there exists a set $\Gamma \subseteq \mathbb{R}$ with $\mu(\Gamma) = 1$ and such that for all $x, y \in \Gamma$, $p_1, p_2 \in \mathcal{P}_1(\mathbb{R})$ satisfying $\pi^*_x + \pi^*_y = p_1 + p_2$, we have

$$\theta\Big(x - \textstyle\int_{\mathbb{R}} z\pi^*_x(dy)\Big) + \theta\Big(y - \int_{\mathbb{R}} z\pi^*_y(dz)\Big) \le \theta\Big(x - \int_{\mathbb{R}} zp_1(dz)\Big) + \theta\Big(y - \int_{\mathbb{R}} zp_2(dz)\Big). \tag{2.6}$$

Let $T(x) = \int_{\mathbb{R}} y\pi^*_x(dy)$. If $x < y$ and $T(y) < T(x)$ we would have by strict convexity that

$$\frac{\theta(x - T(y)) - \theta(x - T(x))}{x - T(y) - (x - T(x))} < \frac{\theta(y - T(y)) - \theta(y - T(x))}{y - T(y) - (y - T(x))},$$

which then would contradict (2.6) for the choice $p_1 = \pi^*_y$ and $p_2 = \pi^*_x$. Hence T is μ-almost surely increasing. Letting $x, y \in \Gamma$, for any $\alpha \in (0, 1]$ we define the convex combination of π^*_x and π^*_y by

$$p_1^\alpha = (1 - \alpha)\pi^*_x + \alpha\pi^*_y, \quad p_2^\alpha = \alpha\pi^*_x + (1 - \alpha)\pi^*_y.$$

Due to T being a monotone map, we can assume that $x \le y$ and $T(x) \le T(y)$. Plugging p_1^α and p_2^α into (2.6), dividing this inequality by α, and sending $\alpha \searrow 0$, yields

$$\Big(\partial_+\theta(x - T(x)) - \partial_-\theta(y - T(y))\Big)(T(x) - T(y)) \ge 0, \tag{2.7}$$

where $\partial_+\theta$ denotes the right-hand side derivative and $\partial_-\theta$ the left-hand side derivative, respectively. By convexity of θ, (2.7) is equivalent to

$$\Big(x - T(x) - y + T(y)\Big)(T(x) - T(y)) \ge 0.$$

Hence, $|x - y| \ge |T(x) - T(y)|$ and T is μ-a.e. 1-Lipschitz.

Further, we have that $T(\mu) \le_c \nu$ and $\bar\pi := (T, id)(\pi^*) \in \Pi_M(T(\mu), \nu)$. Since the cost only depends on x and $T(x)$, but not on the martingale part (i.e., the part which transports mass from $T(x)$ according to $\pi^*_x(dy)$), it could happen that T is constant on some set I, i.e., $\forall x, y \in I$ we have

$$\textstyle\int_{\mathbb{R}} z\pi^*_x(dz) = \int_{\mathbb{R}} z\pi^*_y(dz),$$

but π^*_x and π^*_y do not have to match necessarily. Hence, by choosing some other optimizer (the one which averages on I over all π^*_z appropriately), we may assume that π^*_x is constant on I. Therefore, we assume without loss of generality $\bar\pi_{T(x)} = \pi^*_x$, μ-a.e. Let $(I_k)_{k\in\mathbb{N}}$ be the intervals given by the decomposition of $(T(\mu), \nu)$ into irreducible intervals. Assume that there is an interval I_k so that on $T^{-1}(I_k)$ the map T does not have μ-a.s. slope 1. Then Lemma 2.2 provides two points $\tilde x, \tilde y$ in $T(\Gamma \cap I_k)$ and two corresponding points $x, y \in \Gamma \cap I_k$ such that

$$x < y, \quad T(x) = \tilde x < \tilde y = T(y), \quad x - T(x) > y - T(y),$$

and the overlapping condition (2.1) holds for $\bar\pi_{T(x)} = \pi^*_x$, $\bar\pi_{T(y)} = \pi^*_y$. Lemma 2.3 allows us to define measures p_α and q_α on $\mathbb{R}$ such that

$$\pi^*_x + \pi^*_y = \bar\pi_{T(x)} + \bar\pi_{T(y)} = p_\alpha + q_\alpha.$$

For a graphical depiction compare with Figure 2. Hence,

$$T(x) > v_\alpha := \textstyle\int_{\mathbb{R}} zp_\alpha(dz), \quad T(y) < w_\alpha := \int_{\mathbb{R}} zq_\alpha(dz),$$

are strictly monotone, continuous maps on $[\beta, 1]$. Therefore, we find $\alpha \in (\beta, 1)$ with

$$x - v_1 = x - T(x) < x - v_\alpha \le y - w_\alpha < y - T(y) = y - w_1.$$

By strict convexity of θ we find

$$\frac{\theta(y-w_\alpha)-\theta(x-T(x))}{y-w_\alpha-x+T(x)} < \frac{\theta(y-T(y))-\theta(x-\nu_\alpha)}{y-T(y)-x+\nu_\alpha},$$

which then yields a contradiction to C-monotonicity:

$$\begin{aligned}\theta\Big(x-\textstyle\int_{\mathbb{R}} z\pi_x^*(dz)\Big)&+\theta\Big(y-\textstyle\int_{\mathbb{R}} z\pi_y^*(dz)\Big)\\ &=\theta(x-T(x))+\theta(y-T(y))>\theta(x-v_\alpha)+\theta(y-w_\alpha)\\ &=\theta\Big(x-\textstyle\int_{\mathbb{R}} zp_\alpha(dz)\Big)+\theta\Big(y-\textstyle\int_{\mathbb{R}} zq_\alpha(dz)\Big).\end{aligned}$$

□

3. Sufficiency of the geometric characterization

Naturally the question arises whether any map T satisfying the properties in Theorem 2.1 must be optimal. The aim of this section is to establish this:

Theorem 3.1. *Let $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$. Then any coupling $\pi\in\Pi(\mu,\nu)$ for which $T(x):=\int_{\mathbb{R}} y\pi_x(dy)$ is admissible (in the sense of Definition 1.1) with slope 1 on each interval $T^{-1}(I)$, where I is irreducible w.r.t. $(T(\mu),\nu)$, is optimal for (1.4), i.e.,*

$$V_\theta(\mu,\nu)=\int_{\mathbb{R}}\theta(x-T(x))\mu(dx).$$

The proof is based on dual optimizers and their explicit representation. As long as T is strictly increasing, [19, Theorem 2.1] provides dual optimizers to $V_\theta(\mu,T(\mu))$. Investigating dual optimizers further, we are able to show here $V_\theta(\mu,T(\mu))=V_\theta(\mu,\nu)$. First, Lemma 3.2 helps us to carefully approximate the increasing map T with strictly increasing maps T_ε.

Lemma 3.2. *Let $T\colon\mathbb{R}\to\mathbb{R}$ be an increasing map, with $T-id$ decreasing. Then, for any $\varepsilon>0$ there is a strictly increasing map $T_\varepsilon\colon\mathbb{R}\to\mathbb{R}$, with $T_\varepsilon-id$ decreasing, such that*

$$|T-T_\varepsilon|_\infty\le\varepsilon,$$

and T is affine with slope 1 on an interval I if and only if T_ε is affine with slope 1 on I.

Proof. Since T is increasing we know that the pre-image of any point under T corresponds to an interval. Therefore, we can find at most countable many, disjoint intervals $(I_k)_{k\in\mathbb{N}}$ of finite length, where $T(I_k)$ is a singleton and T is strictly increasing on the complement, i.e., on $\bigcap_k I_k^c$. For any $\varepsilon>0$, we define $g_\varepsilon\colon\mathbb{R}\to\mathbb{R}$

$$g_\varepsilon(x):=\begin{cases}\frac{\varepsilon}{\lambda(I_k)2^k} & \exists k\in\mathbb{N}\colon x\in I_k,\\ 0 & \text{else.}\end{cases}\tag{3.1}$$

Then the map $T_\varepsilon(x):=T(x)+\int_{-\infty}^x[g_\varepsilon(y)\wedge 1]\,dy$ satisfies the desired properties. □

Let $T\colon\mathbb{R}\to\mathbb{R}$ be an increasing 1-Lipschitz function. Then T induces a unique decomposition of $\mathbb{R}$ into at most countably many maximal, closed, disjoint intervals $(I_k)_k$ and a (G_δ-set) G such that for all $k\in\mathbb{N}$ the map $T|_{I_k}$ is affine with slope 1 and $T|_G$ is properly contractive, i.e., for any two points $x,y\in G$ we have $|T(x)-T(y)|<|x-y|$. Below we call the intervals $(I_k)_k$ *irreducible* w.r.t. T.

Proof of Theorem 3.1. The convex function θ can be approximated by an increasing sequence of convex Lipschitz functions $(\theta_n)_{n\in\mathbb{N}}$, e.g.

$$\begin{aligned}\theta_n(x):=\theta(x)\,1_{[-n,n]}(x)+[\theta(n)+&(x-n)\partial_+\theta(n)]\,1_{(n,+\infty)}\\ &+[\theta(-n)+(x+n)\partial_-\theta(-n)]\,1_{(-\infty,-n)}.\end{aligned}$$

By monotone convergence we find[1] $\sup_n V_{\theta_n}(\mu,\nu) = V_\theta(\mu,\nu)$. Indeed, if π^n optimizes $V_{\theta_n}(\mu,\nu)$ and assuming w.l.o.g. that $\pi_n \to \pi$, then

$$\begin{aligned}\lim_n \int \theta_n \left(x - \int y\pi_x^n(dy)\right)\mu(dx) &\geq \lim_n \int \theta_m \left(x - \int y\pi_x^n(dy)\right)\mu(dx)\\ &\geq \int \theta_m \left(x - \int y\pi_x(dy)\right)\mu(dx),\end{aligned}$$

by [7, Proposition 2.8]. Thus the claim follows by taking the supremum in m.

From this, it suffices to consider the case when θ is Lipschitz continuous. By Lemma 3.2 we find for any $\varepsilon > 0$ a strictly increasing map T_ε, such that $T_\varepsilon - id$ is decreasing, the decompositions of T and T_ε match (i.e., T is affine with slope 1 on an interval I iff T_ε is affine with slope 1 on I), and $|T_\varepsilon - T|_\infty \leq \varepsilon$. Then [19, Theorem 2.1] provides a convex, Lipschitz continuous function $f_\varepsilon \colon \mathbb{R} \to \mathbb{R}$ such that for all $x \in \mathbb{R}$

$$Rf_\varepsilon(x) := \inf_{y\in\mathbb{R}} f_\varepsilon(y) + \theta(x-y) = f_\varepsilon(T_\varepsilon(x)) + \theta(x - T_\varepsilon(x)),$$

which is even affine on the parts where T_ε is affine. Write $S_\varepsilon(x) = \int_{-\infty}^x [g_\varepsilon(z) \wedge 1]dz$, where g_ε is defined as in (3.1) so $T_\varepsilon = T + S_\varepsilon$. In the following we will show that f_ε is a dual optimizer (cf. the dual problem as in [14, Theorem 2.11]) to the weak transport problem V_θ between the marginals of the coupling π^ε, where π^ε is defined as the push-forward measure of π by the function

$$(x, y) \mapsto (x, y + S_\varepsilon(x)).$$

First, we compute the barycenters of π_x^ε:

$$\int_{\mathbb{R}} y\pi_x^\varepsilon(dy) = \int_{\mathbb{R}} y + S_\varepsilon(x)\pi_x(dy) = T_\varepsilon(x),$$

and $T_\varepsilon(\mu) \leq_c \mathrm{proj}_2(\pi^\varepsilon) =: \nu_\varepsilon$. Given the sets $(I_k)_{k\in\mathbb{N}}$ and F from the decomposition of $(T(\mu),\nu)$ into irreducible intervals, we find the sets $(I_k^\varepsilon)_{k\in\mathbb{N}}$ and F_ε from the decomposition of $(T_\varepsilon(\mu),\nu_\varepsilon)$ into irreducible intervals by setting

$$I_k^\varepsilon := I_k + S_\varepsilon(x) \quad \text{for any } y \in I_k \text{ and } x \in T^{-1}(y), \quad F_\varepsilon := \bigcap_k (I_k^\varepsilon)^c.$$

In view of the structure of martingale couplings, see [8, Theorem A.4], we find that for μ-a.e. $x \in T_\varepsilon^{-1}(I_k^\varepsilon)$ we have $\mathrm{supp}(\pi_x^\varepsilon) \subseteq \bar{I}_k^\varepsilon$ and $\pi_x^\varepsilon = \delta_{T(x)+S_\varepsilon(x)}$ μ-a.e. on F_ε. Since T has slope 1 on each interval I which is irreducible w.r.t. $(T(\mu),\nu)$, we infer the same for the map T_ε and the irreducible intervals of $(T_\varepsilon(\mu),\nu_\varepsilon)$. The next computation establishes duality of the pair $(\pi^\varepsilon, f_\varepsilon)$, where we use affinity of f_ε on the irreducible components of $(T_\varepsilon(\mu),\nu_\varepsilon)$:

$$\begin{aligned}\int_{\mathbb{R}} \theta\left(x - \int_{\mathbb{R}} y\pi_x^\varepsilon(dy)\right)\mu(dx) &= \int_{\mathbb{R}} \theta\left(x - T_\varepsilon(x)\right)\mu(dx) = \int_{\mathbb{R}} Rf_\varepsilon(x) - f_\varepsilon(T_\varepsilon(x))\mu(dx)\\ &= \int_{\mathbb{R}} Rf_\varepsilon(x)\mu(dx) - \int_{\mathbb{R}} f_\varepsilon\left(\int_{\mathbb{R}} y\pi_x^\varepsilon(dy)\right)\mu(dx)\\ &= \int_{\mathbb{R}} Rf_\varepsilon(x)\mu(dx) - \int_{\mathbb{R}} f_\varepsilon(y)\nu_\varepsilon(dy).\end{aligned}$$

This easily proves that π^ε is optimal for the optimal weak transport problem $V_\theta(\mu,\nu_\varepsilon)$. Drawing the limit for $\varepsilon \searrow 0$, we observe

$$V_\theta(\mu,\nu_\varepsilon) = \int_{\mathbb{R}} \theta(x - T_\varepsilon(x))\mu(dx) \to \int_{\mathbb{R}} \theta(x - T(x))\mu(dx).$$

As θ is Lipschitz, we can apply stability Theorem 1.5 and obtain optimality of π. □

[1]Recall that by [13] the optimizer of the weak transport problem does not depend on the convex cost.

4. Geometry of the weak monotone rearrangement

We can summarize Theorems 2.1 and 3.1 as follows: There exists an admissible map T with slope 1 on $T^{-1}(I)$ whenever I is an irreducible interval w.r.t. $(T(\mu),\nu)$, such that π is optimal for (1.4) iff $T(x)=\int y\pi_x(dy)$ (μ-a.s.). We now show that this map is the weak monotone rearrangement and is therefore the maximum in convex order of the set[2]

$$M(\mu,\nu):=\{S\colon \mathbb{R}\to\mathbb{R}\colon S \text{ is increasing and 1-Lipschitz}, S(\mu)\le_c \nu\}\,.$$

Heuristically speaking, if the maximum in convex order of the set $M(\mu,\nu)$ is again given by an increasing, 1-Lipschitz map, then this map is as close as possible to a shifted identity. In turn, this is favourable when trying to find the minimum in convex order of

$$\{(id-S)(\mu)\colon S\in M(\mu,\nu)\}\,,$$

which gives reason to why there should exist a single optimizer to (1.4) for all convex θ.

As preparation to establishing Theorem 1.3, we prove Lemma 4.1 and Lemma 4.2.

Lemma 4.1. *Let $\mu\in\mathcal{P}_1(\mathbb{R})$, $T,S\colon\mathbb{R}\to\mathbb{R}$ be increasing maps with*

$$\textstyle\int_{\mathbb{R}} T(x)\mu(dx)=\int_{\mathbb{R}} S(x)\mu(dx),$$

then the maximum w.r.t. the convex order[3] of $T(\mu)$ and $S(\mu)$ is given by $T^*(\mu)$ for T^* an increasing map. If in addition T,S are L-Lipschitz with $L>0$, then T^* is also L-Lipschitz.

Proof. The maximum of $T(\mu)$ and $S(\mu)$ w.r.t. the convex order is uniquely determined by the maximum of its potential functions, i.e. $u_{S(\mu)}\vee u_{T(\mu)}=u_{S(\mu)\vee T(\mu)}$. The right-hand side derivative of the potential function can be expressed by the cumulative distribution function, namely $\partial_+u_\mu(x)=2F_\mu(x)-1$. By continuity of the potential functions, we find a minimal partition of $\mathbb{R}$ into at most countably many disjoint intervals $((l_k,r_k])_{k\in\mathbb{N}}$, where $u_{T(\mu)}=u_{S(\mu)}$ on $\{l_k,r_k\}$, and restricted onto $(l_k,r_k]$ one of the following holds true:

(a) $u_{S(\mu)}|_{(l_k,r_k]}\le u_{T(\mu)}|_{(l_k,r_k]}$,
(b) $u_{S(\mu)}|_{(l_k,r_k]}\ge u_{T(\mu)}|_{(l_k,r_k]}$.

Suppose w.l.o.g. (a) holds, then

$$F_{S(\mu)\vee T(\mu)}(x)=F_{T(\mu)}(x)\quad x\in(l_k,r_k]. \tag{4.1}$$

By monotonicity of the maps S and T, we can define T^* on $[\tilde l_k,\tilde r_k)$ where[4]

$$\tilde l_k=\sup T^{-1}(\{l_k\})\vee\sup S^{-1}(\{l_k\}),\quad \tilde r_k=\sup T^{-1}(\{r_k\})\vee\sup S^{-1}(\{r_k\}),$$

by

$$T^*(x)=\begin{cases}T(x) & x\in T^{-1}((l_k,r_k])\cap(\tilde l_k,\tilde r_k],\\ r_k & x\in(\tilde l_k,\tilde r_k]\setminus T^{-1}((l_k,r_k]).\end{cases} \tag{4.2}$$

Hence, T^* is an increasing map and

$$F_{T^*(\mu)}(x)=F_{T(\mu)}(x)=F_{T(\mu)\vee S(\mu)}(x)\quad x\in(l_k,r_k),$$
$$F_{T^*(\mu)}(r_k)=F_{S(\mu)}(r_k)=F_{T(\mu)\vee S(\mu)}(r_k).$$

---

[2] We abuse terminology here, meaning maximum of $\{S(\mu):S\in M(\mu,\nu)\}$.

[3] The maximum in convex order of $\nu,\eta\in\mathcal{P}_1(\mathbb{R})$, denoted $\nu\vee\eta$, satisfies $\nu,\eta\le_c\nu\vee\eta$ as well as $\nu\vee\eta\le_c\xi$ whenever $\nu,\eta\le_c\xi$. One can easily characterize it by the identity $u_{\nu\vee\eta}(\cdot)=u_\nu(\cdot)\vee u_\eta(\cdot)$. Similarly, the minimum of ν and η, denoted $\nu\wedge\eta$, is characterized by $u_{\nu\wedge\eta}$ being equal to the greatest convex minorant of $u_\nu\wedge u_\eta$.

[4] We use the convention that the maximum/supremum of the empty set equals $-\infty$ and the preimage under T and S of $\pm\infty$ is $\pm\infty$.

Therefore, $F_{T^*(\mu)} = F_{S(\mu)\vee T(\mu)}$ which is equivalent to $S(\mu) \vee T(\mu)$ being given by the map T^*. If T and S are in addition L-Lipschitz, it follows by construction that T^* is L-Lipschitz. □

Lemma 4.2. *Let $\eta_1 \leq_c \eta_2$ and $T_2(\eta_2) \leq_c T_1(\eta_1)$, where T_1, T_2 are increasing and 1-Lipschitz. Then $(id - T_1)(\eta_1) \leq_c (id - T_2)(\eta_2)$.*

In particular, if T and S are increasing 1-Lipschitz maps s.t. $\int T(x)\mu(dx) = \int S(x)\mu(dx)$, and we denote by R the increasing 1-Lipschitz map with $R(\mu) = S(\mu) \vee T(\mu)$, which exists by Lemma 4.1, then $(id - R)(\mu) \leq_c (id - S)(\mu) \wedge (id - T)(\mu)$.

Proof. By [18, Theorem 3.A.5] we have that for any two probability measures μ and ν on $\mathbb{R}$ the convex order, i.e., $\mu \leq_c \nu$, is characterized by

$$\int_0^\alpha F_\mu^{-1}(u)du \geq \int_0^\alpha F_\nu^{-1}(u)du \quad \forall \alpha \in [0,1], \tag{4.3}$$

with equality when $\alpha = 1$. For any increasing, continuous map $U\colon \mathbb{R} \to \mathbb{R}$, any measure $\mu \in \mathcal{P}(\mathbb{R})$ and $u \in (0,1)$ we find

$$\begin{aligned} F_{U(\mu)}^{-1}(u) &= \inf\{x \in \mathbb{R}\colon F_{U(\mu)}(x) \geq u\} \\ &= \inf\{x \in \mathbb{R}\colon F_\mu(\sup U^{-1}(\{x\})) \geq u\} \\ &= \inf\{U(y)\colon y \in \mathbb{R} \text{ and } F_\mu(y) \geq u\} \\ &= U(F_\mu^{-1}(u)). \end{aligned} \tag{4.4}$$

Now all tools are assembled to show the assertion. Note that $(id - T_i)$, $i = 1, 2$, are increasing maps and using the equations (4.3) and (4.4), we compute for $\alpha \in [0,1]$

$$\begin{aligned} \int_0^\alpha F_{(id-T_1)(\eta_1)}^{-1}(u)du &= \int_0^\alpha (id - T_1)(F_{\eta_1}^{-1}(u))du = \int_0^\alpha F_{\eta_1}(u) - F_{T_1(\eta_1)}^{-1}(u)du \\ &\geq \int_0^\alpha F_{\eta_2}(u) - F_{T_2(\eta_2)}^{-1}(u)du = \int_0^\alpha F_{(id-T_2)(\eta_2)}^{-1}(u)du. \end{aligned}$$ □

Proof of Theorem 1.3. Existence of an admissible map T which has slope 1 on each interval $T^{-1}(I)$, where I is irreducible w.r.t. $(T(\mu), \nu)$, was already shown in Theorem 2.1. Therefore, it remains to show that the map is maximal. Denote by T the map given by Theorem 2.1 associated with an optimizer to (1.4) and some strictly convex $\theta\colon \mathbb{R} \to \mathbb{R}$. Let S be an arbitrary map in $M(\mu, \nu)$. Then Lemma 4.2 states that

$$(id - R)(\mu) \leq_c (id - T)(\mu),$$

where R is defined as the increasing, 1-Lipschitz map such that $R(\mu) = S(\mu) \vee T(\mu)$. Additionally to existence, strict convexity of θ ensures μ-almost sure uniqueness of T in the sense that for any optimal coupling π we have $\int_{\mathbb{R}} y\pi_x(dy) = T(x)$ μ-a.s. Thus, $R(\mu) = T(\mu)$ and $T = R$ μ-almost surely. □

Proof of Theorem 1.2. This is a direct consequence of Theorem 2.1, which provides existence of a map with the desired geometric properties, and Theorem 1.3, which provides the equivalence between the geometric properties and maximality. □

5. On the reverse problem of Alfonsi, Corbetta, and Jourdain

We aim to prove Theorem 1.4 pertaining the *reverse* problem (1.7).

Lemma 5.1. *Let $\eta_1, \eta_2 \in \mathcal{P}_1(\mathbb{R})$ with $\int_{\mathbb{R}} y\eta_1(dy) = \int_{\mathbb{R}} y\eta_2(dy)$, and $T_1, T_2\colon \mathbb{R} \to \mathbb{R}$ be increasing maps with*

$$T_1(\eta_1) = T_2(\eta_2) =: \nu.$$

Denote the minimum in convex order of η_1 and η_2 by η. Then there exists an increasing map T^* such that $T^*(\eta) = \nu$. If in addition, the maps are L-Lipschitz with $L > 0$, then the same holds true for T^*.

Proof. Suppose there exist increasing maps T_i and measures η_i, $i = 1, 2$, such that $T_1(\eta_1) = \nu = T(\eta_2)$. Then the potential function of the minimum η of η_1 and η_2 w.r.t. the convex order is given by the convex hull of u_{η_1} and u_{η_2}. The potential function u_η completely specifies the cumulative distribution function through $\partial_+ u_\eta = 2F_\eta - 1$. Thus, we can find a partition of $\mathbb{R}$ into countably many, disjoint intervals $I_k = [a_k, b_k) \cap \mathbb{R}$. For $k \in \mathbb{N}$ we find $i \neq j \in \{1, 2\}$ with $u_\eta(a_k) = u_{\eta_i}(a_k)$, $u_\eta(b_k) = u_{\eta_j}(b_k)$ such that one of the following holds

(a) $u_\eta(x) = u_{\eta_i}(x)$ on I_k,

(b) $F_\eta(a_k) = F_\eta(b_k-) = F_{\eta_j}(b_k-)$ and $u_\eta(x) < u_{\eta_1}(x) \wedge u_{\eta_2}(x)$ on (a_k, b_k).

According to this decomposition, we can define an increasing map T^* via

$$T^*(x) = \begin{cases} T_i(x) & x \in I_k, \text{ (a) holds,} \\ T_i(a_k) & x \in I_k, \text{ (b) holds.} \end{cases}$$

Note that T^* is L-Lipschitz if T_i, $i = 1, 2$, are L-Lipschitz. Let $y \in \mathbb{R}$, due to the continuity of the maps T_1 and T_2, we can find points $x_1, x_2 \in \mathbb{R}$ with $p = F_\nu(y)$, $x_1 = F_{\eta_1}^{-1}(p)$ $x_2 = F_{\eta_2}^{-1}(p)$. Assume that $i = 1, j = 2$ with $x_1 \in I_k$. If (a) holds, then we have $F_\eta(x) = F_{\eta_1}(x)$ for all $x \in I_k$. Now presume that (b) holds, then $F_{\eta_2}(b_k-) = F_\eta(a_k) \leq F_{\eta_1}(x_1)$. Then

$$\begin{aligned} F_{\eta_2}(b_k) = p &\implies x_2 = b_k \text{ and } F_\eta(b_k) = p, \\ F_{\eta_2}(b_k) < p &\implies \exists l \in \mathbb{N}\colon x_2 \in I_l \text{ s.t. } i = 2 \text{ and (a) holds.} \end{aligned}$$

Hence, by monotonicity of the map T^* we conclude $F_{T^*(\eta)} = F_\nu$. □

Proof of Theorem 1.4. Define ν^* via the weak monotone rearrangement T between μ and ν, as μ on the contraction parts of T and accordingly shifted ν on the affine (irreducible) intervals such that $\tilde{T}(\nu^*) = \nu$. Let $\eta \geq_c \mu$. Then for any strictly convex $\theta\colon \mathbb{R} \to \mathbb{R}$ and coupling $\pi^2 \in \Pi(\eta, \nu)$ we have $\int_{\mathbb{R}\times\mathbb{R}} \theta(y - z)\pi^2(dy, dz) \geq \int_{\mathbb{R}} \theta(y - \int_{\mathbb{R}} z\pi^2_y(dz))\eta(dy)$, with equality iff π^2 is actually given by a map. Hence, if the optimizer π^2 of $W_\theta(\eta, \nu)$ is not given by a map and $\pi^1 \in \Pi_M(\mu, \eta)$, we have

$$W_\theta(\eta, \nu) > \int_{\mathbb{R}} \theta(x - \int_{\mathbb{R}} \int_{\mathbb{R}} z\pi^2_y(dz)\pi^1_x(dy))\mu(dx) \geq V_\theta(\mu, \nu).$$

Thus, by the structure of the weak monotone rearrangement, we deduce optimality of ν^* for Problem (1.7). To show uniqueness of (1.7), assume that η attains the minimum of (1.7) and the optimizer of $W_\theta(\eta, \nu)$ is given by the map R. For any martingale coupling $\pi^1 \in \Pi_M(\mu, \eta)$ we define a map by

$$L(x) := \int_{\mathbb{R}} R(y)\pi^1_x(dy).$$

Then by optimality $L(\mu) = T(\mu)$ and, in particular,

$$\int_{\mathbb{R}} \theta(y - R(y))\eta(dy) = \int_{\mathbb{R}} \theta(x - L(x))\mu(dx) = \int_{\mathbb{R}} \theta(x - T(x))\mu(dx),$$

which shows $L = T$ μ-almost surely. By strict convexity, we have

$$y - R(y) = x - L(x) = x - T(x) \quad \pi^1\text{-a.s.}$$

Since π^1 was arbitrary in $\Pi_M(\mu, \eta)$ we get that R is affine with slope 1 on I, whenever I is an irreducible interval w.r.t. (μ, η). Therefore η and ν^* restricted to I coincide. Hence, $\eta = \nu^*$.

We finally show that ν^* is minimal in the convex order as stated. By Lemma 5.1, we can assume $\mu \leq_c \eta \leq_c \nu^*$ and that η can be pushed forward onto ν via an increasing 1-Lipschitz map S. It follows by Lemma 4.2 that $(id - S)(\eta) \leq_c (id - \tilde{T})(\nu^*)$, so

$$V_\theta(\mu, \nu) \leq \int \theta(x - S(x))\eta(dx) \leq \int \theta(x - \tilde{T}(x))\nu^*(dx) = V_\theta(\mu, \nu),$$

and by the uniqueness obtained above we deduce $\eta = \nu^*$. □

6. Stability of barycentric weak transport problems in multiple dimensions

The final part of the article is concerned with stability of the weak optimal transport problem under barycentric costs, see (1.4). Unlike in the rest of the article we work here on $\mathbb{R}^d$. The final aim is to prove Theorem 1.5.

One surprising aspect of this result is that we only require $\nu^k \to \nu$ in $\mathcal{W}_1$ and not necessarily in $\mathcal{W}_\rho$. This relates to the conditional expectation in (1.4) being 'inside of θ.' We first prove an illuminating intermediate result:

Proposition 6.1. *Let $1 \le \rho < \infty$, $(\mu^k)_{k\in\mathbb{N}} \in \mathcal{P}_\rho(\mathbb{R}^d)^{\mathbb{N}}$, $(\nu^k)_{k\in\mathbb{N}} \in \mathcal{P}_\rho(\mathbb{R}^d)^{\mathbb{N}}$ and $\theta\colon \mathbb{R}^d \to \mathbb{R}$ convex and satisfying the growth condition (1.8). Suppose that $\mu^k \to \mu$ and $\nu^k \to \nu$ in $\mathcal{W}_\rho$, and that $\eta \le_c \nu$. Then there exist $\eta^k \le_c \nu^k$ such that*

(i) $\eta^k \to \eta$ in $\mathcal{W}_\rho$,
(ii) $\lim_k W_\theta(\mu^k, \eta^k) = W_\theta(\mu, \eta)$.

Proof. It is well-known that (i) together with the stated convergence of the μ^k's implies (ii), so we proceed to prove the former. Let π^k be an optimal coupling attaining $W_\rho(\nu, \nu^k)$. Let M be any martingale coupling with first marginal η and second marginal ν, the existence of which is guaranteed by the assumption $\eta \le_c \nu$ together with Strassen's theorem. We convene on the notation $M(dx, dy)$ and $\pi(dy, dz)$, and define the measure

$$P(dx, dy, dz) = M_x(dx)\pi^k_y(dz)\nu(dy).$$

This measure has η, ν and ν^k as first, second and third marginals. We next define $R^k(x)$ as the conditional expectation under P of the third variable given the first one, namely

$$R^k(x) = \int\int z\, \pi^k_y(dz) M_x(dy).$$

Next we introduce $\eta^k := R^k(\eta)$ so by definition $\eta^k \le_c \nu^k$. Finally

$$\begin{aligned} W_\rho(\eta, \eta^k)^\rho &\le \int |x - R^k(x)|^\rho \eta(dx) \\ &= \int \left| \int y\, M_x(dy) - \int \left(\int z\, \pi^k_y(dz) \right) M_x(dy) \right|^\rho \eta(dx) \\ &\le \int \left| y - \int z\, \pi^k_y(dz) \right|^\rho \nu(dy) \\ &\le \int\int |y - z|^\rho \pi^k(dy, dz) \;=\; W_\rho(\nu, \nu^k)^\rho, \end{aligned} \tag{6.1}$$

by the martingale property and two applications of Jensen's inequality. The desired conclusion follows. □

Remark 6.2. In the context of the previous proposition, if η is supported in finitely many atoms, then the condition that $\nu^k \to \nu$ in $\mathcal{W}_\rho$ can be relaxed to convergence in $\mathcal{W}_1$. To wit, if $\eta = \sum_{i=1}^{\ell} \alpha_i \delta_{x^i}$, one can take $\rho = 1$ in (6.1) and prove

$$\forall i \le \ell : \; |x^i - R^k(x^i)| \le (\min\{\alpha_j\})^{-1} W_1(\nu, \nu^k),$$

so taking ρ-power and integrating w.r.t. η we get $W_\rho(\eta, \eta^k)^\rho \le K\, W_1(\nu, \nu^k)^\rho \to 0$.

The previous remark shows that we need to reduce to the finite-support setting. We carry to this in the next two lemmas:

Lemma 6.3. *Let $\eta \in \mathcal{P}_1(\mathbb{R}^d)$. Then for any $\varepsilon > 0$ there is a compactly supported, sub-probability measure $\tilde\eta$ with*

$$\tilde\eta \le \eta, \quad \tilde\eta(\mathbb{R}^d) \ge 1 - \varepsilon, \quad \int_{\mathbb{R}^d} z\eta(dz) = \tfrac{1}{\tilde\eta(\mathbb{R}^d)} \int_{\mathbb{R}^d} z\tilde\eta(dz). \tag{6.2}$$

Proof. We first partition $\mathbb{R}^d$ into countably many, disjoint d-dimensional cubes $(Q_k^\delta)_{k\in\mathbb{N}}$ of length $\delta > 0$. Define an approximation η^δ of η by

$$\eta^\delta := \textstyle\sum_{k\in\mathbb{N}} \delta_{z_k^\delta}\, \eta(Q_k^\delta), \quad z_k^\delta := \begin{cases} \frac{1}{\eta(Q_k^\delta)} \int_{Q_k^\delta} z\eta(dz) & \eta(Q_k^\delta) > 0, \\ 0 & \text{else.} \end{cases}$$

Note that $\eta^\delta \leq_c \eta$ and $\eta^\delta \to \eta$ in $\mathcal{W}_1$ when $\delta \searrow 0$. If there exists an approximation η^δ such that the assertion holds, then it is straightforward to construct the corresponding measure for η, which in turn satisfies the assertion with respect to η. W.l.o.g., we may assume that

$$\Big\{\textstyle\sum_{i=1}^{2d} \alpha_i v_i \colon (\alpha_i)_{i=1}^{2d} \in \mathbb{R}_+^{2d}, v_1, \dots, v_{2d} \in \Big\{x - \bar z \in \mathbb{R}^d \colon x \in \operatorname{supp}(\eta)\Big\}\Big\} = \mathbb{R}^d,$$

where $\bar z$ denotes the barycenter of η. Then we can find $\delta > 0$ such that

$$\Big\{\textstyle\sum_{i=1}^{2d} \alpha_i v_i \colon (\alpha_i)_{i=1}^{2d} \in \mathbb{R}_+^{2d}, v_1, \dots, v_{2d} \in \Big\{x - \bar z \in \mathbb{R}^d \colon x \in \operatorname{supp}(\eta^\delta)\Big\}\Big\} = \mathbb{R}^d.$$

Let $z_{n_1}^\delta, \dots, z_{n_{2d}}^\delta$ span $\mathbb{R}^d$ in the sense above and

$$\eta^\delta(z_{n_j}^\delta) = \eta(Q_{n_j}^\delta) > 0 \quad j = 1, \dots, 2d.$$

For any $\varepsilon > 0$ there is a $\tilde\varepsilon \in (0, \varepsilon)$ such that

$$\Big\{\textstyle\sum_{i=1}^{2d} \alpha_i z_{n_i} \colon (\alpha_i)_{i=1}^{2d} \in \mathbb{R}_+^{2d}, \sum_{i=1}^{2d} \alpha_i < \varepsilon\Big\} \supseteq B_{\tilde\varepsilon}(0).$$

Besides, there exists a compact set $K \subseteq \mathbb{R}^d$ such that

$$\eta^\delta(K^c) < \tilde\varepsilon, \quad \big|\bar z - \textstyle\int_K z\eta^\delta(dz)\big| < \tilde\varepsilon$$

and $z_{n_1}^\delta, \dots, z_{n_{2d}}^\delta \in K$. Therefore, we find $(\tilde\alpha_i)_{i=1}^{2d} \in \mathbb{R}_+^{2d}$ with

$$\bar z - \textstyle\int_K z\eta^\delta(dz) = \sum_{i=1}^{2d} \tilde\alpha_i z_{n_i}, \quad \sum_{i=1}^{2d} \tilde\alpha_i < \varepsilon.$$

If ε is chosen smaller than $\eta^\delta(z_{n_i}^\delta)$ for all $i = 1, \dots, 2d$, we can define $\tilde\eta^\delta$ via

$$\tilde\eta^\delta := \eta^\delta\restriction_K - \textstyle\sum_{i=1}^{2d} \tilde\alpha_i \delta_{z_{n_i}}. \qquad \square$$

Lemma 6.4. *Let $\mu, \eta \in \mathcal{P}_\rho(\mathbb{R}^d)$ and $\theta \colon \mathbb{R}^d \to \mathbb{R}$ convex satisfying the growth condition* (1.8)*. Then there exists a sequence $(\eta^k)_{k\in\mathbb{N}}$ of finitely supported measures with $\eta^k \leq_c \eta$, $\eta^k \to \eta$ in $\mathcal{W}_\rho$ and $W_\theta(\mu, \eta^k) \to W_\theta(\mu, \eta)$.*

Proof. For any $\varepsilon > 0$ we find a compact set $K_\varepsilon \subseteq \operatorname{supp}(\eta)$ such that $\int_{K_\varepsilon^c} |y|^\rho \eta(dy) < \varepsilon$. For any $\delta > 0$, the set K_ε can be covered by finitely many, disjoint sets $(A_i^{\varepsilon,\delta})_{i=1}^{N_{\varepsilon,\delta}}$ with diameter smaller than δ and $K_\varepsilon = \bigcup_i A_i^{\varepsilon,\delta}$. Define the measure $\eta^{\varepsilon,\delta}$ by

$$\eta^{\varepsilon,\delta} = \delta_{z_\varepsilon} \eta(K_\varepsilon^c) + \textstyle\sum_{i=1}^{N_\delta} \delta_{z_i^{\varepsilon,\delta}} \eta(A_i^{\varepsilon,\delta}),$$

where the points z_ε and $(z_i^{\varepsilon,\delta})_{i=1}^{N_{\varepsilon,\delta}}$ are given by

$$z_\varepsilon := \tfrac{1}{\eta(K_\varepsilon^c)} \textstyle\int_{K_\varepsilon^c} y\eta(dy), \quad z_i^{\varepsilon,\delta} := \tfrac{1}{\eta(A_i^{\varepsilon,\delta})} \int_{A_i^{\varepsilon,\delta}} y\eta(dy).$$

By construction, there exists a martingale coupling between $\eta^{\varepsilon,\delta}$ and η, thus, $\eta^{\varepsilon,\delta} \leq_c \eta$. Note that θ restricted to K_ε is Lipschitz continuous. Then drawing the limit $\delta \to 0$ yields

$$\textstyle\sum_{i=1}^{N_{\varepsilon,\delta}} \delta_{z_i^{\varepsilon,\delta}} \eta(A_i^{\varepsilon,\delta}) \to \eta|_{K_\varepsilon} \text{ in } \mathcal{W}_\rho,$$

and by convexity, $|z_\varepsilon|^\rho \eta(K_\varepsilon^c) \leq \int_{K_\varepsilon^c} |y|^\rho \eta(dy)$. Choosing $\delta(\varepsilon)$ sufficiently small, we have

$$\mathcal{W}_\rho(\eta, \eta^{\varepsilon,\delta(\varepsilon)}) \leq 2\varepsilon \text{ and } \eta^{\varepsilon,\delta(\varepsilon)} \to \eta \text{ in } \mathcal{W}_\rho.$$

By the growth condition (1.8) we obtain $W_\theta(\mu, \eta^{\varepsilon,\delta(\varepsilon)}) \to W_\theta(\mu, \eta)$. $\square$

We can now prove a version of Proposition 6.1 under weaker assumptions:

Lemma 6.5. *Let $(\nu^k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{P}_1(\mathbb{R}^d)$ and let $(\mu^k)_{k\in\mathbb{N}}$ be a sequence $\mathcal{P}_\rho(\mathbb{R}^d)$ with $\nu^k \to \nu$ in $\mathcal{W}_1$, $\mu^k \to \mu$ in $\mathcal{W}_\rho$, where $\rho \geq 1$, and let $\theta\colon \mathbb{R}^d \to \mathbb{R}$ be a convex functions satisfying the growth constraint* (1.8)*. Then for any $\eta \leq_c \nu$ we find a sequence of $\eta^k \leq_c \nu^k$ such that $W_\theta(\mu^k, \eta^k) \to W_\theta(\mu, \eta)$ and $\eta^k \to \eta$ in $\mathcal{W}_1$.*

Proof. W.l.o.g. assume that θ is positive. By Lemma 6.3, we can find for any $\varepsilon > 0$ a compactly supported $\hat\eta = \tilde\eta + (1 - \tilde\eta(\mathbb{R}^d))\delta_{\bar z} \in \mathcal{P}_\rho(\mathbb{R}^d)$ with $\hat\eta \leq_c \eta$, $\mathcal{W}_1(\hat\eta, \eta) < \varepsilon$ and

$$W_\theta(\mu, \hat\eta) \leq W_\theta(\mu, \eta) + c \textstyle\int_{\mathbb{R}^d} 1 + |x - \bar z|^\rho \mu(dx) < \infty,$$

where $\bar z := \int_{\mathbb{R}^d} y\eta(dy)$. Using stability of classical optimal transport, see [23, Theorem 5.20], we may assume that $|W_\theta(\mu, \hat\eta) - W_\theta(\mu, \eta)| < \varepsilon$. By Lemma 6.4 we may reduce to the case of finitely supported $\hat\eta$. We conclude the proof with Remark 6.2. □

Finally we can give the pending proof of Theorem 1.5:

Proof of Theorem 1.5. Lower-semicontinuity of the map $(\mu, \nu) \mapsto V_\theta(\mu, \nu)$ follows from [7, Theorem 1.3]. By [7, Lemma 6.1] we have

$$V_\theta(\mu, \nu) = \inf\nolimits_{\eta \leq_c \nu} W_\theta(\mu, \eta), \tag{6.3}$$

where the infimum is attained for a measure $\eta \leq_c \nu$. By Lemma 6.5 we find a sequence $\eta^k \leq_c \nu^k$, so that again using [7, Lemma 6.1] we have

$$V_\theta(\mu, \nu) = W_\theta(\mu, \eta) = \lim\nolimits_k W_\theta(\mu^k, \eta^k) \geq \limsup\nolimits_k V_\theta(\mu^k, \nu^k).$$

If θ is strictly convex, the infimum in (6.3) is attained by a unique probability measure $\eta^k \leq_c \nu^k$,[5] which in turn is the push-forward of μ^k under a μ^k-uniquely defined map T^k. Moreover, the W_θ-optimal transport plan $\pi^k \in \Pi(\mu^k, \nu^k)$ is uniquely determined by $\mu^k(dx)\delta_{T^k(x)}(dy)$: Suppose the contrary and let $T'(x) := \int_{\mathbb{R}^d} y\pi^k_x(dy)$, then $T'(\mu^k) \leq_c \nu^k$ and we find the contradiction

$$V_\theta(\mu^k, \nu^k) \leq \textstyle\int_{\mathbb{R}^d} \theta(x - T'(x))\mu^k(dx) < \int_{\mathbb{R}^d \times \mathbb{R}^d} \theta(x - y)\pi^k(dx, dy) = V_\theta(\mu^k, \nu^k).$$

Hence, by convergence of the values of V_θ and tightness of $(\eta_k)_{k\in\mathbb{N}}$, we deduce the convergence of the η^k to the optimal $\eta \leq_c \nu$ in $\mathcal{W}_1$. Suppose that $\mu^k = \mu$ for all $k \in \mathbb{N}$, then due to the uniqueness of the optimal transport maps T between T and $T(\mu)$, we can apply Theorem [23, Corollary 5.23] and obtain convergence of the transport maps T^k to T. □

References

[1] B. Acciaio, J. Backhoff-Veraguas, and A. Zalashko. Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization. *Forthcoming at Stoch. Processes and their Applications*, 2019.

[2] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of probability measures in the convex order by Wasserstein projection. *Annales de l'Institut Henri Poincaré B, Probability and Statistics*, 56(3):1706–1729, 2020.

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

[4] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. Adapted Wasserstein Distances and Stability in Mathematical Finance. *Finance and Stochastics*, to appear.

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

[6] J. Backhoff-Veraguas, M. Beiglböck, Y. Lin, and A. Zalashko. Causal transport in discrete time and applications. *SIAM Journal on Optimization*, 27(4):2528–2562, 2017.

[5] The uniqueness of η^k, T^k and π^k was already shown in [2, Theorem 2.1] for $|\cdot|^\rho$, $\rho > 1$

[7] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Forthcoming at Calc. Variations and Partial Diff. Eq.*, 2019.
[8] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Annals of Probability*, 44(1):42–106, 2016.
[9] M. Beiglböck and N. Juillet. Shadow couplings. *ArXiv e-prints*, Sept. 2016.
[10] M. Fathi and Y. Shu. Curvature and transport inequalities for Markov chains in discrete spaces. *Bernoulli*, 24(1):672–698, 2018.
[11] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.
[12] N. Gozlan and N. Juillet. On a mixture of Brenier and Strassen theorems. *arXiv preprint arXiv:1808.02681*, 2018.
[13] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali. Characterization of a class of weak transport-entropy inequalities on the line. *Ann. Inst. Henri Poincaré Probab. Stat.*, 54(3):1667–1693, 2018.
[14] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, 273(11):3327–3405, 2017.
[15] K. Marton. A measure concentration inequality for contracting Markov chains. *Geometric & Functional Analysis GAFA*, 6(3):556–571, 1996.
[16] K. Marton et al. Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration. *The Annals of Probability*, 24(2):857–866, 1996.
[17] P.-M. Samson. Transport-entropy inequalities on locally acting groups of permutations. *Electron. J. Probab.*, 22:Paper No. 62, 33, 2017.
[18] M. Shaked and J. G. Shanthikumar. *Stochastic orders.* Springer Science & Business Media, 2007.
[19] Y. Shu. From Hopf-Lax formula to optimal weak transfer plan. *arXiv preprint arXiv:1609.03405*, 2016.
[20] Y. Shu. Hamilton-Jacobi equations on graph and applications. *Potential Anal.*, 48(2):125–157, 2018.
[21] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 81(1):73–205, 1995.
[22] M. Talagrand. New concentration inequalities in product spaces. *Inventiones mathematicae*, 126(3):505–563, 1996.
[23] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften.* Springer, 2009.

APPLICATIONS OF WEAK TRANSPORT THEORY

J. BACKHOFF-VERAGUAS AND G. PAMMER

Abstract. Motivated by applications to geometric inequalities, Gozlan, Roberto, Samson, and Tetali [34] introduced a transport problem for 'weak' cost functionals. Basic results of optimal transport theory can be extended to this setup in remarkable generality.

In this article we collect several problems from different areas that can be recast in the framework of weak transport theory, namely: the Schrödinger problem, the Brenier–Strassen theorem, optimal mechanism design, linear transfers, semimartingale transport. Our viewpoint yields a unified approach and often allows to strengthen the original results.

keywords: Schrödinger problem, Brenier–Strassen theorem, linear transfers, semimartingale transport, optimal mechanism design, weak transport problem, duality, cyclical monotonicity.

1. Overview

The optimal transport problem for weak costs was first introduced by Gozlan, Roberto, Samson and Tetali [34] and has immediately generated interest in several groups of researchers, see [45, 46, 29, 32, 1, 2] among others. To present the basic problem we introduce some notation. Throughout X and Y denote Polish spaces. Given probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ we write $\Pi(\mu,\nu)$ for the set of all couplings on $X \times Y$ with marginals μ and ν. Given a coupling π on $X \times Y$ we denote a regular disintegration with respect to the first marginal by $(\pi_x)_{x\in X}$.

We consider cost functionals of the form

$$C : X \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\},$$

where C is lower bounded and lower semicontinuous, and $C(x,\cdot)$ is assumed to be convex on $\mathcal{P}(Y)$ for every $x \in X$. The weak transport problem is then to determine

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

The classical transport problem is included via $C(x,p) = \int c(x,y)\,dp(y)$ for a given cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$.

Fundamental results in classical optimal transport theory include *existence*, *duality*, and *cyclical monotonicity* of optimizers. Through a series of contributions (see [34, 2, 32, 6]) it has been understood that these results extend in full generality to the weak transport setup. We will recall these results in some detail in Section 2 below.

The purpose of the present article is to advertise the power and flexibility of optimal weak transport theory through the investigation of a broad variety of applications thereof. We will consider the following problems:

- The *Schrödinger problem* has recently received particular attention since it provides a regularized version of the transport problem that is numerically much more tractable than the classical counterpart. On a technical level, the difference is that quadratic costs are replaced by the entropy wrt a reference measure. Based on monotonicity for weak transport costs we give a short proof of the fundamental characterization of optimizers in the Schrödinger problem. (Section 4.)
- The *Brenier–Strassen Theorem* of Gozlan and Juillet [32] yields that 1-Lipschitz maps that are gradients to convex functions are the optimizers of transport problems with barycentric costs. This result plays a role in the probabilistic proof of the Caffarelli contraction theorem [28]. We provide a short new derivation which is based on monotonicity for weak transport costs and emphasizes the similarity with the classical Brenier Theorem. (Section 5.)
- A significant problem in the economics literature is to *optimize the revenue for a multiple good monopolist*. The influential article [25] of Daskalakis, Deckelbaum, and Tzamos suggests a systematic framework to study such problems and provides a dual characterization. We link the multiple good monopolist problem to optimal weak transport and use the weak transport duality theorem to recover and strengthen the results of [25]. (Section 6.)
- Bowles and Ghoussoub [18] have recently introduced the class of *linear transfers* which includes many specific couplings between probability measures. A main theorem of [18] provides the representation of linear transfers through weak transport under the assumption of compactness of the underlying spaces. We provide a short new derivation based on weak transport duality that allows to drop the compactness condition. In particular, this implies that the representation result of Bowles and Ghoussoub is valid for the important case of Euclidean space. (Section 7.)
- The *semimartingale transportation problem* was introduced by Tan and Touzi [50] and extends classical optimal transport to the case where mass is transported along the trajectories of a semimartingale. As such, it also contains both martingale optimal transport [30, 13, 14], and the drift control framework of Mikami and Thieullen [41], as particular cases. We show that weak transport theory can be used to strengthen the main duality result of [50]. (Section 8.)

The paper is organized as follows. In Section 2 we give a brief overview of previous works connected to optimal weak transport. In Section 3 we review the basic results of optimal weak transport theory together with the necessary notation. Then we describe the various applications announced above in Sections 4 - 8.

2. Literature connected to the weak transport problem

The weak transport problem was introduced by Gozlan, Roberto, Samson and Tetali [34], and shortly afterwards by Aliberti, Bouchitte and Champion [2]. The problem has also been designated "general transport problem" and "non-linear transport problem" respectively.

The initial works of Gozlan et al. [34, 33] are mainly motivated by applications to geometric inequalities. Indeed, particular costs of the form (OWT) were already considered by Marton [40, 39] and Talagrand [48, 49]. The theory for problem (OWT) has been further developed in [1, 34, 33, 45, 44, 46, 29, 32, 6, 9, 7]: Basic results of existence and duality are established in the articles [34, 2, 6]. The notion of *C-monotonicity* was developed in [4, 32, 6] as an analogue of classical c-cyclical montonicity in order to provide a characterization of optimizers to the weak transport problem. A weak transport analogue to the case of quadratic costs in classical optimal transport, is the case of barycentric costs. This case has received particular attention, and we refer to [33, 45, 44, 46, 29, 32, 9, 7, 1].

The weak transport viewpoint is useful for a number of problems loosely related to stochastic optimization: it appears in the recursive formulation of the causal transport problem [5], in [1, 2, 15, 4, 36, 22] it is used to investigate martingale optimal transport problems, in [3] it is applied to prove stability of pricing and hedging in mathematical finance and, as mentioned above, it appears in the recent probabilistic proof to the Caffarelli contraction theorem [28].

3. Fundamental results of weak transport theory

For $t \geq 1$, $\mathcal{P}_t(X)$ denotes the set of Borel probability measures with finite t-th moment for some fixed metric d_X (compatible with the topology on X), i.e., a Borel probability measure μ is in $\mathcal{P}_t(X)$ iff for some $x_0 \in X$ we have

$$\int_X d_X(x, x_0)^t \mu(dx) < \infty.$$

The set of continuous functions on X which are dominated by a multiple of $1 + d_X(x, x_0)^t$, is denoted by $\Phi_t(X)$. We equip the set of probability measures $\mathcal{P}_t(X)$ with the t-th Wasserstein topology. Specifically, a sequence $(\mu_k)_{k\in\mathbb{N}}$ converges to $\mu \in \mathcal{P}_t(X)$ if $\mu_k(f) := \int f d\mu_k$ converges to $\mu(f)$ for all $f \in \Phi_t(X)$. The space $\mathcal{P}(X)$ itself is equipped with the usual weak topology. The same conventions apply to Y instead of X.

We have already mentioned the basic assumptions on the function C in the introductory section above. We recall it and make it more precise in the following definition.

Definition 3.1 (A)**.** *We say $C\colon X \times \mathcal{P}_t(Y) \to \mathbb{R} \cup \{+\infty\}$ satisfies property* (A) *iff*

i) C is lower-semi continuous wrt the product topology on $X \times \mathcal{P}_t(Y)$,
ii) C is bounded from below,

iii) the map $p \mapsto C(x,p)$ is convex, i.e., for all $x \in X$ and $p,q \in \mathcal{P}_t(Y)$ we have

$$C(x, \lambda p + (1-\lambda)q) \leq \lambda C(x,p) + (1-\lambda)C(x,q) \quad \lambda \in [0,1].$$

From now on until the end of this section, the cost function C is assumed to satisfy property (A).

We will need the following existence and continuity result from [6]:

Theorem 3.2 (Existence and semicontinuity)**.** *The infimum in* (OWT) *is attained and the value $V_C(\mu,\nu)$ depends in a lower semicontinuous way on the marginals $(\mu,\nu) \in \mathcal{P}(X) \times \mathcal{P}_t(Y)$.*

In optimal transport, the renowned Kantorovich duality states that for lower semi-continuous and lower bounded cost functions $c\colon X \times Y \to \mathbb{R} \cup \{+\infty\}$ we have

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c(x,y)\, \pi(dx,dy) = \sup_{\substack{f \in L_1(\mu),\ g \in L_1(\nu), \\ f+g \leq c}} \mu(f) + \nu(g). \tag{3.1}$$

Here duality takes the following form, which resembles (and generalizes) (3.1), cf. [6, Theorem 3.1]:

Theorem 3.3 (Kantorovich duality for weak transport)**.** *The weak transport problem* (OWT) *admits the dual representation*

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x,\pi_x)\, \mu(dx) = \sup \mu(f) + \nu(g), \tag{3.2}$$

where the supremum is taken over functions $f \in L_1(\mu)$, $g \in \Phi_t(Y)$, satisfying

$$f(x) + p(g) \leq C(x,p)$$

for $x \in X$, $p \in \mathcal{P}_t(Y)$.

We will often use Theorem 3.3 in the following (equivalent) form:

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x,\pi_x)\mu(dx) = \sup_{g \in \Phi_t(Y)} -\nu(g) + \int_X R_C g(x)\mu(dx), \tag{3.3}$$

where $R_C g(x) = \inf_{p \in \mathcal{P}_t(Y)} p(g) + C(x,p)$. Moreover, the right-hand supremum in (3.3) can be restricted to functions in $g \in \Phi_t(Y)$ which are bounded from below.

We also recall a further consequence of the proof of Theorem 3.3 in [6] that provides further insight into the dual problem. The convex conjugate of $\nu \mapsto V_C(\mu,\nu)$ admits a rather concrete representation: For any $g \in \Phi_t(Y)$ we have

$$\sup_{\nu \in \mathcal{P}_t(Y)} \nu(g) - V_C(\mu,\nu) = -\int_X R_C g(x)\mu(dx). \tag{3.4}$$

The notion of c-cyclical monotonicity constitutes a necessary (and often also sufficient) optimality criterion for transport plans in classical optimal transport, i.e., for any measurable cost $c\colon X \times Y \to \mathbb{R} \cup \{-\infty, +\infty\}$ if $\pi^* \in \Pi(\mu, \nu)$ is an optimizer of

$$V_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{X\times Y} c(x, y)\pi(dx, dy),$$

where $|V_c(\mu, \nu)| < \infty$, then there exists $\Gamma \subseteq X \times Y$ with $\pi^*(\Gamma) = 1$ such that for all $N \in \mathbb{N}$ and $(x_k, y_k)_{k=1}^N$ in Γ we have for any permutation σ of $\{1, \dots, N\}$ that

$$\sum_{k=1}^N c(x_k, y_k) \le \sum_{k=1}^N c(x_k, y_{\sigma(k)}).$$

The importance of c-cyclical monotonicity has been understood at least since the publication of the seminal article [31]. See [52, 17, 10] for minimal conditions that guarantee equivalence of optimality and c-cyclical monotonicity. More recently, variants of this 'monotonicity priniciple' have been applied in transport problems for finitely or infinitely many marginals, the martingale version of the transport problem, the Skorokhod embedding problem, and the distribution constrained optimal stopping problem, see [43, 21, 35, 12, 14, 42, 53, 11] among others.

In the context of optimal weak transport the corresponding concept is C-monotonicity. Early versions can be found in [4, 6, 32], while the following definition as well as the subsequent result are taken from [9, Section 2].

Definition 3.4. *A set $\Gamma \subseteq X \times \mathcal{P}_t(Y)$ is called C-monotone iff for any finite subset of points $(x_k, p_k)_{k=1}^N$ of Γ we have*

$$\sum_{k=1}^N C(x, p_k) \le \sum_{k=1}^N C(x, q_k) \quad \sum_{k=1}^N q_k = \sum_{k=1}^N p_k.$$

A coupling π with first marginal μ is called C-monotone iff there is a C-monotone set Γ such that $(id_X, \delta_{\pi_x})_\# \mu$ is concentrated on Γ.

Theorem 3.5 (C-monotonicity)**.** *If* (OWT) *is finitely valued, then any optimizer is C-monotone.*

The reverse implication holds also true if C is sufficiently regular, see [9, Theorem 2.2].

Theorem 3.6. *Assume that $\mu \in \mathcal{P}_t(X)$, $\nu \in \mathcal{P}_t(Y)$. If C is continuous and $|C(x, p)| \le R\,(d_X(x, x_0)^t + \int_Y d_Y(y, y_0)^t p(dy))$ for some R, then any C-monotone coupling $\pi \in \Pi(\mu, \nu)$ is optimal for* (OWT).

4. Structure of optimizers in the Schrödinger problem

We refer the reader to [38] for a survey on the Schrödinger problem / entropic transport problem (cf. (4.4) below). Starting with the articles [24, 16, 27], the Schrödinger problem has received significant attention as a regularized, numerically tractable version of the classical transport problem.

The main goal of this section is to recover, in Corollary 4.3 below, the characterization of entropic cost optimal transport plans through the product structure of their density. Given a Polish space Z, the relative entropy of $\mu \in \mathcal{P}(Z)$ wrt a 'reference measure' $\nu \in \mathcal{P}(Z)$ is defined as

$$H(\mu|\nu) = \begin{cases} \int_X \log\left(\frac{d\mu}{d\nu}\right)\mu(dx) & \mu \ll \nu, \\ +\infty & \text{else.} \end{cases}$$

Theorem 4.1. *Let γ be a probability measure equivalent to the product measure $\mu \otimes \nu$ for $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. If the coupling $\pi^* \in \Pi(\mu, \nu)$ is optimal for the problem*

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_X H(\pi_x|\gamma_x)\mu(dx) < +\infty, \tag{4.1}$$

then there is a measurable set $\Gamma \subseteq X \times \mathcal{P}(Y)$ with $\mu(\{x \in X : (x, \pi^*_x) \in \Gamma\}) = 1$ such that for all $(x, \pi^*_x), (z, \pi^*_z) \in \Gamma$ there is a constant $\alpha > 0$ so that

$$\frac{d\pi^*_x}{d\gamma_x} = \alpha \frac{d\pi^*_z}{d\gamma_z} \quad \nu\text{-a.e.} \tag{4.2}$$

Proof. The minimization problem (4.1) constitutes a weak transport problem. Therefore, there is by Theorem 3.5 a measurable set $\Gamma \subseteq X \times \mathcal{P}(Y)$ such that π^* is C-monotone on Γ. Moreover, we may assume w.l.o.g. that $\gamma_x \sim \nu$ for all $(x, \pi^*_x) \in \Gamma$ and $H(\pi^*_x|\gamma_x) < \infty$. Fix $(x, \pi^*_x), (z, \pi^*_z) \in \Gamma$.

We want to show that $\pi^*_x \sim \pi^*_z$. By C-monotonicity, cf. Definition 3.4, we have for all $q_x, q_z \in \mathcal{P}(Y)$ with $q_x + q_z = \pi^*_x + \pi^*_z$ that

$$H(\pi^*_x|\gamma_x) + H(\pi^*_z|\gamma_z) \leq H(q_x|\gamma_x) + H(q_z|\gamma_z).$$

Then the corresponding first order optimality condition reads as

$$\pi^*_x\left(\log\left(\frac{d\pi^*_x}{d\gamma_x}\right)\right) + \pi^*_z\left(\log\left(\frac{d\pi^*_z}{d\gamma_z}\right)\right) \leq q_x\left(\log\left(\frac{d\pi^*_x}{d\gamma_x}\right)\right) + q_z\left(\log\left(\frac{d\pi^*_z}{d\gamma_z}\right)\right). \tag{4.3}$$

This inequality can be easily deduced by differentiation along the segment joining π^*_x with q_x and π^*_z with q_z, i.e., by computing

$$\left.\frac{d}{dt}\right|_{t=0} H(q^t_x|\gamma_x) + H(q^t_z|\gamma_z),$$

where $q^t_\cdot = (1-t)\pi^*_\cdot + tq_\cdot$. Assume that there exists a Borel measurable set $A \subseteq Y$ with $\pi^*_x(A) > 0$ but $\pi^*_z(A) = 0$. Then on A we have for ν-a.e. y that $\frac{d\pi^*_z}{d\gamma_z}(y) = 0$. It is straightforward that one can find q_x, q_z such that $q_x + q_z = \pi^*_x + \pi^*_z$, $q_z(A) > 0$ and $q_x\left(\log\left(\frac{d\pi^*_x}{d\gamma_x}\right)\right) < +\infty$, so we omit the technical details. As a consequence we find

$$q_x\left(\log\left(\frac{d\pi^*_x}{d\gamma_x}\right)\right) + q_z\left(\log\left(\frac{d\pi^*_z}{d\gamma_z}\right)\right) = -\infty,$$

which contradicts the first order optimality criterion since the left-hand side of (4.3) is non-negative. Hence by symmetry π^*_x is equivalent to π^*_z and as a consequence we have $\pi^*_x \sim \nu$ for all $(x, \pi^*_x) \in \Gamma$.

Applying Lemma 4.2 below to the pair π^*_x, π^*_z yields a contradiction to C-monotonicity. $\square$

Lemma 4.2. *Let p_i, γ_i, $i \in \{1,2\}$ and ν be equivalent probability measures on Y. If $H(p_i|\gamma_i)$, $i = 1,2$, is finite and there exists no constant $\alpha \in \mathbb{R}$ such that*

$$\frac{dp_1}{d\gamma_1} = \alpha \frac{dp_2}{d\gamma_2} \quad \nu\text{-a.e.},$$

then there are two probability measures $q_1, q_2 \in \mathcal{P}(Y)$ with $q_1 + q_2 = p_1 + p_2$ and

$$H(p_1|\gamma_1) + H(p_2|\gamma_2) > H(q_1|\gamma_1) + H(q_2|\gamma_2).$$

Proof. Since any Polish space is Borel-isomorphic to a measurable subset of $[0,1]$ we may assume that $Y = [0,1]$. By the inverse transform sampling we may assume that ν is the uniform distribution on $[0,1]$. For $i \in \{1,2\}$, define the densities

$$g_i\colon [0,1] \to (0,+\infty), \quad g_i = \frac{dp_i}{d\gamma_i}; \quad f_i\colon [0,1] \to (0,+\infty), \quad f_i = \frac{d\gamma_i}{d\nu}.$$

By Lusin's theorem we find for any $\varepsilon > 0$ a compact set $K_\varepsilon \subseteq [0,1]$ with mass $\nu(K_\varepsilon) \geq 1 - \varepsilon$ such that $h := \frac{g_1}{g_2}\colon [0,1] \to (0,+\infty)$ is continuous on K_ε. By the assumption that g_1 is not a multiple of g_2 we find two disjoint intervals $[a_1,b_1],[a_2,b_2] \subseteq (0,+\infty)$, $b_1 < a_2$ with

$$\nu(h^{-1}([a_1,b_1])) > 0 \text{ and } \nu(h^{-1}([a_2,b_2])) > 0.$$

Choosing $\varepsilon > 0$ sufficiently small the closed, disjoint sets

$$A_1 = h^{-1}([a_1,b_1]) \cap K_\varepsilon \text{ and } A_2 = h^{-1}([a_2,b_2]) \cap K_\varepsilon$$

have positive mass under ν. W.l.o.g. we can assume that $\nu(A_1) = \nu(A_2) > 0$. The map

$$\begin{aligned} T\colon A_1 &\to A_2 \\ x &\mapsto F^{-1}_{\nu|_{A_2}} \circ F_{\nu|_{A_1}}(x) \end{aligned}$$

provides a measure preserving bijection. Let $S : [0,1] \to [0,+\infty)$ be defined as

$$S(y) = \begin{cases} \min\left\{g_1(y)f_1(y),\ g_2(y)f_2(y),\ \hat{g}_1(y)\hat{f}_1(y),\ \hat{g}_2(y)\hat{f}_2(y)\right\} & y \in A_1, \\ -\min\left\{g_1(z)f_1(z),\ g_2(z)f_2(z),\ g_1(y)f_1(y),\ g_2(y)f_2(y)\right\} & y \in A_2, \\ 0 & \text{else,} \end{cases}$$

where $\hat{g}_i = g_i \circ T$, $\hat{f}_i = f_i \circ T$ and $z = T^{-1}(y)$. Then we define for $t \in [-1,1]$ the probability measures p_1^t and p_2^t on $[0,1]$ by

$$p_1^t(dy) = p_1(dy) + tS(y)\nu(dy), \quad p_2^t(dy) = p_2(dy) - tS(y)\nu(dy).$$

The respective densities are then given by

$$\frac{dp_1^t}{d\gamma_1} = g_1 + t\frac{S}{f_1}, \quad \frac{dp_2^t}{d\gamma_2} = g_2 - t\frac{S}{f_2}.$$

Differentiation of

$$\begin{aligned}&H(p_1^t|\gamma_1) + H(p_2^t|\gamma_2)\\&= \int_Y \log\left(g_1(y) + t\frac{S}{f_1}(y)\right) p_1^t(dy) + \int_Y \log\left(g_2(y) + t\frac{S}{f_2}(y)\right) p_2^t(dy)\end{aligned}$$

with respect to t and evaluation at 0 yields

$$\int_Y S(y)\log\left(\frac{g_1}{g_2}(y)\right)\nu(dy) = \int_{A_1} S(y)\log\left(\frac{g_1(y)g_2(T(y))}{g_2(y)g_1(T(y))}\right)\nu(dy) < 0.$$

By strict convexity of $t \mapsto H(p_1^t|\gamma_1)+H(p_2^t|\gamma_2)$ we conclude that there exists a $t_0 \in [-1,1]$ with

$$H(p_1|\gamma_1) + H(p_2|\gamma_2) > H(p_1^{t_0}|\gamma_1) + H(p_2^{t_0}|\gamma_2).$$

□

Finally we obtain the main result of this section (see [23, Corollary 3.2]).

Corollary 4.3. *Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ and $\gamma \in \mathcal{P}(X \times Y)$ a probability measure equivalent to $\mu \otimes \nu$. Assume that the value of the corresponding entropic optimal transport problem is finite, i.e.,*

$$\inf_{\pi\in\Pi(\mu,\nu)} H(\pi|\gamma) < \infty. \tag{4.4}$$

A coupling $\pi^* \in \Pi(\mu,\nu)$ minimizes (4.4) *if and only $H(\pi^*|\gamma) < \infty$ and there exist measurable functions f and g such that*

$$\frac{d\pi^*}{d\gamma}(x,y) = f(x)g(y) \quad \gamma\text{-a.e.}$$

Proof. We use the notation of Theorem 4.1 and rewrite

$$H(\pi|\gamma) = \int_X H(\pi_x|\gamma_x)\mu(dx) + H(\mu|\gamma_0), \tag{4.5}$$

where $\gamma_0 \in \mathcal{P}(X)$ is the X-marginal of γ. Lower semicontinuity and strict convexity of the relative entropy yield the existence of a minimizer $\pi^* \in \Pi(\mu,\nu)$ of (4.4). The identity (4.5) shows that π^* also minimizes (4.1). By Theorem 4.1, calling Γ the set with μ-full X-projection therein, we fix $(z,\pi_z^*) \in \Gamma$ and define $g(y) := \frac{d\pi_z^*}{d\gamma_z}(y)$ as well as $h(x) = \alpha$ for α as in (4.2). Thus we have

$$\frac{d\pi_x^*}{d\gamma_x}(y) = h(x)g(y) \quad (x,\pi_x^*) \in \Gamma,\ \gamma_x\text{-a.e. } y.$$

Hence,

$$\frac{d\pi^*}{d\gamma}(x,y) = \frac{d\mu}{d\gamma_0}(x)h(x)g(y) = f(x)g(y) \quad \gamma\text{-a.e.},$$

where f is defined appropriately. Conversely, assume that we have two couplings π, π' with finite entropy, with marginals μ and ν, and such that their densities w.r.t. γ are of product form. Let

$$\frac{d\pi}{d\gamma}(x,y) = f(x)g(y) \text{ and } \frac{d\pi'}{d\gamma}(x,y) = f'(x)g'(y)$$

and $\log(fg)fg, \log(f'g')f'g' \in L^1(\gamma)$. Since the marginals of π and π' coincide, we have for any $h \in L^1(\mu) \oplus L^1(\nu)$ that

$$\int h(x,y)(fg - f'g')\gamma(dx,dy) = 0. \tag{4.6}$$

We can approximate $\log(fg)$ by elements in $L^1(\mu) \oplus L^1(\nu)$ such that on $[\log(fg) \geq 0]$ we have $h_n \geq 0$ and $h_n \nearrow \log(fg)$ and similar applies on $[\log(fg) \leq 0]$. Hence

$$\int \log(fg) d\pi = \lim_n \int h_n fg d\gamma = \lim_n \int h_n f'g' d\gamma = \int \log(fg) d\pi'.$$

Therefore, $\log(fg)f'g' \in L^1(\gamma)$ and $\log(f'g')fg \in L^1(\gamma)$ and

$$H(\pi|\gamma) - H(\pi|\pi') = \int \log(f'g') d\pi = \int \log(f'g') d\pi' = H(\pi'|\gamma).$$

Especially, we have shown $H(\pi|\gamma) = H(\pi'|\gamma)$ since $H(\pi|\pi')$ is non-negative, which implies $H(\pi|\pi')$ and $H(\pi'|\pi)$ vanish, so $\pi = \pi'$. □

5. Brenier-Strassen Theorem

A fundamental result in the theory of optimal transport is Brenier's theorem [19, 20] which asserts that the optimizer of the Wasserstein-2 distance on $\mathbb{R}^d$ between μ, ν is given by the gradient of a convex function φ. Specifically, the optimal plan $\pi^* \in \Pi(\mu,\nu)$ is of the form $(id, \nabla\varphi)(\mu)$ and $\nabla\varphi(\mu) = \nu$. We refer to, e.g. [51, Theorem 2.12] for more details and bibliographical remarks.

Strassen's theorem [47] asserts that given marginals $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ there exists a martingale $(Z_i)_{i=1,2}$ with $Z_1 \sim \mu, Z_2 \sim \nu$ provided that $\mu \leq_c \nu$. Here $\leq_c$ denotes the usual convex order, i.e. $\mu \leq_c \nu$ means that $\int \varphi \, d\mu \leq_c \int \psi \, d\nu$ for all convex functions φ. Note that the condition $\mu \leq_c \nu$ is not only sufficient for the existence of a martingale with these marginals but in fact also necessary by Jensen's inequality.

An intermediate version between the classical $\mathcal{W}_2$-problem and Strassen's Theorem on the existence of martingales [47] was recently investigated in various forms [32, 34, 33, 1, 46, 6, 7]. The following represents the weak transport analogue of the classical $\mathcal{W}_2$-distance:

$$V_2(\mu,\nu)^2 := \inf_{\eta \leq_c \nu} \mathcal{W}_2(\mu,\eta)^2 \tag{5.1}$$

$$= \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d} \left| x - \int_{\mathbb{R}^d} y \pi_x(dy) \right|^2 \mu(dx). \tag{5.2}$$

Note that the equality of (5.1) and (5.2) is a straightforward consequence of Strassen's Theorem.

The Brenier-Strassen theorem of Gozlan and Juillet [32] (see also Shu [45] for the one-dimensional case) asserts that $\eta^* \leq_c \nu$ is optimal for (5.1) if and only if there exists a convex function φ with 1-Lipschitz gradient such that $\nabla\varphi(\mu) = \eta^*$. The proof of the *if*-part of the statement can be done by showing dual attainment or more directly using C-monotonicity whereas the *only if*-part was thus far only shown via duality.

The goal of this section is to recover the *only if*-clause by means of sufficiency of C-monotonicity. We believe that this new proof is appealing in that it mimics the proof of Brenier's theorem via classical cyclical monotonicity, underlining the similarity of the two results.

To provide an intuition for the proof, we recall the main idea in the classical case: Let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ be a convex and differentiable function. For any finite number of points $x_1, \dots, x_n \in \mathbb{R}^d$ it is immediate that

$$\sum_{k=1}^{n} \langle x_{i+1} - x_i, \nabla\varphi(x_i)\rangle \leq 0. \tag{5.3}$$

Completing the square yields

$$\sum_{k=1}^{n} |x_i - \nabla\varphi(x_i)|^2 \leq \sum_{k=1}^{n} |x_{i+1} - \nabla\varphi(x_i)|^2, \tag{5.4}$$

which shows optimality for any measure supported on finitely many points in the graph of $\nabla\varphi$. Therefore, since cyclical monotonicity is sufficient for optimality, the coupling $(id, \nabla\varphi)(\mu)$ turns out to be optimal for the quadratic distance transport problem.

In the subsequent proof, we will use as a black box a characterization of convex functions with 1-Lipschitz gradient in the spirit of cyclical monotonicity, see [55]. This permits to draw the desired conclusion for the problem (5.1) in analogy to step from (5.3) to (5.4).

Theorem 5.1. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ where $\mathbb{R}^d$ is equipped with the standard euclidean norm $|\cdot|$. Let η^* be in convex order dominated by ν. Then the following are equivalent:*

(1) The values of $\mathcal{W}_2(\mu, \eta^*)$ and $V_2(\mu, \nu)$ coincide, i.e. η^* is a solution of the optimization problem (5.1)
(2) There is a convex function $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ such that $\nabla\varphi\colon \mathbb{R}^d \to \mathbb{R}^d$ is 1-Lipschitz, $\eta^* = \nabla\varphi(\mu)$ and $(id, \nabla\varphi)_\# \mu$ is the unique optimizer of (5.1).

Theorem 5.1 can be found in [32, Theorem 2.1]. A proof of the first implication using different arguments is given in [6, Theorem 1.4].

Proof of Theorem 5.1. Here we will only show "2 $\implies$ 1". To this end, we want to verify that

$$\Gamma := \left\{(x, \delta_{\nabla\varphi(x)}) \in \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \colon x \in \mathbb{R}^d\right\}$$

is C-monotone, where $C(x, p) := \left|x - \int y p(dy)\right|^2$.

Let $N \in \mathbb{N}$, $x_1, \dots, x_N \in \mathbb{R}^d$ with $y_i = \nabla\varphi(x_i)$. We have to show that if $\sum_{i=1}^N m_i = \sum_{i=1}^N \delta_{y_i}$ then $\sum_{i=1}^N |x_i - y_i|^2 \leq \sum_{i=1}^N |x_i - \int y m_i(dy)|^2$. Clearly we must have $m_i = \sum_{j=1}^N \alpha_{i,j}\delta_{y_j}$ for all i, where $(\alpha_{i,j}) \in \mathbb{R}_+^{N\times N}$ is a bistochastic matrix. We can rewrite $\alpha = (\alpha_{i,j})$ as a convex combination of permutation matrices $(P_\sigma)_{\sigma\in\Sigma}$, where Σ denotes the set of permutations on $\{1, \dots, N\}$:

$$\alpha = \sum_{\sigma\in\Sigma} \beta_\sigma P_\sigma, \quad \sum_{\sigma\in\Sigma} \beta_\sigma = 1, \quad \beta_\sigma \geq 0.$$

Note that the map $F\colon \mathbb{R}^{N!} \to \mathbb{R}$

$$F((\beta_\sigma)_{\sigma\in\Sigma}) = \frac{1}{2}\sum_{i=1}^N \Big| x_i - \sum_{\sigma\in\Sigma} \beta_\sigma y_{\sigma(i)} \Big|^2$$

is convex. We seek to show that

$$\hat\beta_\sigma = \begin{cases} 1 & \sigma = id, \\ 0 & \text{else}, \end{cases}$$

is a minimum on the simplex on $\mathbb{R}^{N!}$. The gradient of F at $\hat\beta$ has the following form

$$\nabla F(\hat\beta) = \left(\sum_{i=1}^N (x_i - y_i)\cdot(y_i - y_{\sigma(i)}) \right)_{\sigma\in\Sigma}.$$

For any permutation $\sigma \in \Sigma$ we compute

$$\begin{aligned} \sum_{i=1}^N (x_i - y_i)\cdot(y_i - y_{\sigma(i)}) &= \sum_{i=1}^N x_i \cdot (y_i - y_{\sigma(i)}) - \frac{1}{2}|y_i - y_{\sigma(i)}|^2 \\ &= \sum_{i=1}^N (x_i - x_{\sigma^{-1}(i)})\cdot y_i - \frac{1}{2}|y_i - y_{\sigma^{-1}(i)}|^2 \\ &\geq 0. \end{aligned}$$

The last inequality above follows from [55, Lemma 4], which states that φ being convex and 1-Lipschitz is equivalent to

$$\varphi(z) - \varphi(x) \geq \nabla\varphi(x)\cdot(z-x) + \frac{1}{2}|\nabla\varphi(z) - \nabla\varphi(x)|^2 \quad \forall x, z \in \mathbb{R}^d,$$

so by summing this last inequality we get

$$\begin{aligned} 0 &= \sum_{i=1}^N \varphi(x_{\sigma(i)}) - \varphi(x_i) \\ &\leq \sum_{i=1}^N (x_i - x_{\sigma(i)})\cdot y_i - \frac{1}{2}|y_i - y_{\sigma(i)}|^2. \end{aligned}$$

We conclude that $\nabla F(\hat\beta)$ is pointwise non-negative proving optimality of $\hat\beta$ on the simplex by convexity of F.

All in all, the coupling $(id, \nabla\varphi)_\#\mu$ is concentrated on the C-monotone set Γ. Since both marginals are in $\mathcal{P}_2(\mathbb{R}^d)$ we can apply Theorem 3.6 which yields its optimality for the weak transport problem $V_2(\mu,\nu)$. □

An immediate consequence of the Brenier-Strassen Theorem 5.1, coupled with the proof therein, is the following modification of the classical Rockafellar theorem. We denote the subdifferential of a convex function φ at $x \in \mathbb{R}^d$ by $\partial\varphi(x)$ and by $\partial\varphi = \{(x,y)\colon x \in \mathbb{R}^d, y \in \partial\varphi(x)\}$ its graph.

Corollary 5.2 (Rockafellar-Strassen)**.** *Let $L \in \mathbb{R}^+$ and $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$. The following are equivalent:*

(1) Γ *satisfies for all*[1] $n \in \mathbb{N}$, $((x_1, y_1), \dots, (x_n, y_n)) \in \Gamma$,

$$\sum_{i=1}^{n} (x_{i+1} - x_i) \cdot y_i + \frac{1}{2L} |y_{i+1} - y_i|^2 \leq 0. \tag{5.5}$$

(2) *there exists a convex* $\varphi \in C^1(\mathbb{R}^d)$ *with L-Lipschitz gradient such that* $\Gamma \subseteq \partial\varphi$.

Proof. The implication '2 $\Longrightarrow$ 1' can be easily deduced from [55, Lemma 4]. For '1 $\Longrightarrow$ 2': Since we can always consider $\tilde{\Gamma} = \{(x, y) \colon (x, Ly) \in \Gamma\}$, which sastisfies (5.5) for $L = 1$, we can assume w.l.o.g. that $L = 1$. Let μ be a probability measure on $\mathbb{R}^d$ such that

$$\operatorname{supp}(\mu) = \operatorname{cl}(\{x \colon \exists y, (x, y) \in \Gamma\}).$$

By (5.5), if $(x, y), (x, y') \in \Gamma$ then $y = y'$. This defines a map $T \colon \operatorname{proj}_1 \Gamma \to \operatorname{proj}_2 \Gamma$. Further, for all $(x, y), (x', y') \in \Gamma$, we have

$$-|y - y'|^2 \geq (x - x') \cdot (y - y') \geq -|x - x'||y - y'|,$$

which shows that T is 1-Lipschitz. The coupling $\pi = (id, T)_\# \mu$ is optimal for (5.2) between its marginals μ and η^*, by the reasoning in the proof of Theorem 5.1 . Hence, by Theorem 5.1 there is a convex function φ with 1-Lipschitz gradient such that $(id, \nabla\varphi)_\# \mu$ is the unique optimizer. Thus, $\nabla\varphi = T$ (μ-a.s.). Due to continuity we conclude $\nabla\varphi = T$ on $\operatorname{proj}_1 \Gamma$ and $\Gamma \subseteq \partial\varphi$. □

6. Multiple-Good monopoly problem

The goal of this section is to recover the main result of Daskalakis, Deckelbaum, and Tzamos [25] in Corollary 6.2 below. We will not discuss the economic interpretation and just mention that it can be interpreted as a Kantorovich-Rubinstein Theorem for specific weak transport costs.

We will obtain Corollary 6.2 as a consequence of the more general result Theorem 6.1. We first introduce some notation. Fix $X := \mathbb{R}^d$ and equip it with the coordinate-wise partial order $\leq$. Let $\Phi_t^{icx}(\mathbb{R}^d)$ consist of all $\leq$-increasing, convex functions in $\Phi_t(\mathbb{R}^d)$. For $\mu, \nu \in \mathcal{P}_1(X)$ we write $\mu \leq_{icx} \nu$ iff $\int f \, d\mu \leq \int f \, d\nu$ for all $f \in \Phi_1^{icx}(X)$. Once again, it follows from the results of Strassen [47] that $\mu \leq_{icx} \nu$ is tantamount to the existence of a stochastic process $(Z_i)_{i=1,2}$ satisfying

$$Z_1 \sim \mu, Z_2 \sim \nu, Z_1 \leq \mathbb{E}[Z_2 | Z_1].$$

Theorem 6.1. *Let* $\theta \in \Phi_{b,t}(\mathbb{R}^d)$ *be convex and denote*

$C_{\theta,icx}(x, p) := \inf_{q \leq_{icx} p} \theta\left(x - \int y q(dy)\right)$ *and* $R_\theta\varphi(x) := \inf_{y \leq z, z \in \mathbb{R}^d} \varphi(z) + \theta(x - y)$.

Then all of the following optimization problems yield the same value:

i) $\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d} C_{\theta,icx}(x, \pi_x) \mu(dx)$,

ii) $\inf_{\tilde{\nu} \leq_{icx} \nu} \inf_{\pi \in \Pi(\mu,\tilde{\nu})} \int \theta(x - y) \pi(dx, dy)$,

iii) $\inf_{\mu \leq_{icx} \tilde{\mu}, \tilde{\nu} \leq_{icx} \nu} \inf_{\pi \in \Pi(\tilde{\mu},\tilde{\nu})} \int \theta(x - y) \pi(dx, dy)$,

[1] Here $(x_{n+1}, y_{n+1}) := (x_1, y_1)$.

iv) $\sup_{\varphi\in\Phi_t^{icx}(\mathbb{R}^d)} -\nu(\varphi) + \int R_\theta\varphi(x)\mu(dx)$.

Proof. "$i) = ii)$": For $(i) \geq (ii)$ we first take π to be an (almost) optimizer of (i) and by a measurable selection argument take q_x to be an (almost) optimizer of $C_{\theta,icx}(x,\pi_x)$. Defining $T(x) := \int y q_x(dy)$ we remark that $T(\mu) \leq_{icx} \nu$, and so $(i) \geq \int \theta(x - T(x))\mu(dx) \geq (ii)$. For the converse, take $\tilde\nu \leq_{icx} \nu$ and $P \in \Pi(\mu,\tilde\nu)$ which are (almost) optimizers of (ii). By Strassen's result [47, Theorem 9] there exists a submartingale coupling $\tilde\pi \in \Pi(\mu,\nu)$, i.e. a coupling π satisfying $\int y\, d\pi_x \geq x$.

Next we define $\pi(dx,dy) = \int \tilde\pi_{\tilde x}(dy)P(dx,d\tilde x)$ which belongs to $\Pi(\mu,\nu)$. Since by definition $\theta(x-\tilde x) \geq C_{\theta,icx}(x,\tilde\pi_{\tilde x})$, we have by Jensen's inequality

$$
\begin{aligned}
(ii) = \int \theta(x-\tilde x)P(dx,d\tilde x) &\geq \int C_{\theta,icx}(x,\tilde\pi_{\tilde x})P(dx,d\tilde x)\\
&\geq \int C_{\theta,icx}\Big(x, \int \tilde\pi_{\tilde x}P_x(d\tilde x)\Big)\mu(dx)\\
&\geq \int C_{\theta,icx}(x,\pi_x)\mu(dx) \geq (i),
\end{aligned}
$$

where we used that $C_{\theta,icx}(x,\cdot)$ is convex if θ is convex.

"$ii) = iii)$": Fix $\mu \leq_{icx} \tilde\mu$ and $\tilde\nu \leq_{icx} \nu$. Let $X_1 \leq_{icx} X_2$ and $Y_1 \leq_{icx} Y_2$ be $\mathbb{R}^d$-valued random variables on some probability space whose laws satisfy

$$X_1 \sim \mu, \quad X_2 \sim \tilde\mu, \quad Y_1 \sim \tilde\nu, \quad Y_2 \sim \nu.$$

The order $\leq_{icx}$ between random variables has to be understood in the following sense, (where the existence of such random variables is again provided by [47, Theorem 9])

$$X \leq_{icx} Y \iff \mathbb{E}[Y|X] \geq X \quad \text{a.s.}$$

Define the random variable $Z = X_1 + \mathbb{E}[Y_1 - X_2|X_1]$. Then Z is in increasing convex order to Y_2, since

$$Z \leq_{icx} Z + \mathbb{E}[X_2 - X_1|X_1] = \mathbb{E}[Y_1|X_1] \leq_{icx} Y_1 \leq_{icx} Y_2.$$

Applying Jensen's inequality yields

$$\mathbb{E}[\theta(X_2 - Y_1)] \geq \mathbb{E}[\theta(\mathbb{E}[X_2 - Y_1|X_1])] = \mathbb{E}[\theta(X_1 - Z)] \geq ii).$$

"$i) = iv)$": By duality, see Theorem 3.3, we have

$$i) = \sup_{\varphi\in\Phi_{b,t}(\mathbb{R}^d)} -\nu(\varphi) + \int R_{C_{\theta,icx}}\varphi(x)\mu(dx). \tag{6.1}$$

It remains to show that we can additionally restrict the infimum to convex and increasing functions which will be accomplished using the double c-convexification trick. To this end, we note that

$$C_{\theta,icx}(x,p) = \inf\nolimits_{x-\int yp(dy)\leq z} \theta(z) =: \hat\theta\Big(x - \int yp(dy)\Big),$$

where $\hat\theta$ is a function on $\mathbb{R}^d$ which is bounded from below and convex. Further note

$$x \mapsto \hat\theta(x-y) \text{ is increasing}, \quad y \mapsto \hat\theta(x-y) \text{ is decreasing}. \tag{6.2}$$

Analogously to Gozlan et al. [33, Proof of Theorem 2.11 (2)] we find that

$$\inf_{y\in\mathbb{R}^d} \varphi(y) + \hat\theta(x-y) = R_{C_{\theta,icx}}\varphi(x) = R_{C_{\theta,icx}}\hat\varphi(x) = \inf_{y\in\mathbb{R}^d} \hat\varphi(y) + \hat\theta(x-y),$$

where $\hat\varphi$ denotes the convex envelope of φ. The inf-convolution $\psi := R_{C_{\theta,icx}}\hat\varphi$ is therefore bounded from below, convex and increasing. Note that for any $p \in \mathcal{P}(\mathbb{R}^d)$ with barycenter $\int yp(dy) = z$ we have

$$\psi(x) - p(\hat\varphi) \le \psi(x) - \hat\varphi(z) \le \hat\theta(x-z),$$

which allows us to $\hat\theta$-convexify $\hat\varphi$, i.e.,

$$\bar\varphi(y) = \sup_x \psi(x) - \hat\theta(x-y),$$

which is in particular an increasing function by (6.2) with

$$\hat\varphi \ge \bar\varphi \ge \psi(y) - \hat\theta(0) \ge \min_y \psi(y) - \hat\theta(0).$$

Again, we find that the convex envelope of $\bar\varphi$

$$\tilde\varphi(z) = \inf_{p\in\mathcal{P}_t(Y):\ z=\int yp(dy)} p(\bar\varphi) \tag{6.3}$$

is increasing, convex and dominated by $\bar\varphi$, thus,

$$\tilde\psi(x) := R_{C_{\theta,icx}}\tilde\varphi(x) = R_{C_{\theta,icx}}\bar\varphi(x) \ge \psi(x) = R_{C_{\theta,icx}}\varphi(x).$$

Finally, we obtain (since $\tilde\varphi \le \varphi$ and $\psi \le \tilde\psi$) that

$$\mu(\psi) - \nu(\varphi) \le \mu(\tilde\psi) - \nu(\tilde\varphi),$$

which shows that we can replace $\Phi_{b,t}(\mathbb{R}^d)$ with $\Phi^{icx}_{b,t}(\mathbb{R}^d)$ in (6.1). □

The proof of Corollary 6.2 resembles the proof of Kantorovich-Rubinstein duality when one already knows that classical Kantorovich duality holds.

Corollary 6.2. *Setting $\theta(x) = |x|$ where $|\cdot|$ is a norm on $\mathbb{R}^d$, we have*

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d} \inf_{z\le\int y\pi_x(dy)} |x-z|\mu(dx) = \sup_{\varphi\in\Phi^{icx}_{b,1}(\mathbb{R}^d)\text{ and 1-Lipschitz}} \mu(\varphi) - \nu(\varphi).$$

Proof. First we show that given $\varphi \in \Phi^{icx}_{b,1}(\mathbb{R}^d)$ the inf-convolution $\psi := R_{C_{|\cdot|,icx}}\varphi$ is 1-Lipschitz: Let $x, x' \in \mathbb{R}^d$ then

$$R_{C_{|\cdot|,icx}}\varphi(x) - R_{C_{|\cdot|,icx}}\varphi(x') \le \sup_{z\le y,\ y\in\mathbb{R}^d} |x-y| - |x'-y| \le |x-x'|.$$

By the proof of Theorem 6.1 ψ is additionally bounded from below, convex and increasing. Using the notation in the proof of Theorem 6.1 the mapping $\bar\varphi$ is increasing, bounded from below and also 1-Lipschitz. Hence, for $x, x' \in \mathbb{R}^d$ the increasing, convex and lower bounded function $\tilde\varphi$, see (6.3), is 1-Lipschitz

$$\tilde\varphi(x) - \tilde\varphi(x') \le \sup_{p\in\mathcal{P}_1(Y)} \mathcal{W}_1(p, T_{x'-x}(p)) = |x-x'|,$$

where $T_{x'-x}$ is the translation by $x' - x$. By the conclusion of the proof of Theorem 6.1 we can assume w.l.o.g. that $\varphi \in \Phi^{icx}_{b,1}(\mathbb{R}^d)$ is 1-Lipschitz.

It remains to show that $R_{C_{|\cdot|,icx}}\varphi$ and φ coincide. By definition of the inf-convolution we have $\varphi \geq R_{C_{|\cdot|,icx}}$. We find by 1-Lipschitz continuity of φ

$$-\varphi(x) + \varphi(y) + \inf_{z\leq y} |x - z| \geq \inf_{z\leq y} -\varphi(x) + \varphi(z) + |x - z| \geq 0,$$

which shows the reverse inequality. □

7. Backward transfers

Bowles and Ghoussoub [18] suggest a notion of linear transfer between probability measures which is more encompassing than mass transportation but still admits important traits of the dual theory of mass transport. In particular, they identify many examples that illustrate the scope of their approach.

A main result of Bowles and Ghoussoub yields a representation of linear transfers through weak transport problems, see [18, Theorem 3.1]. In this section, we recover [18, Theorem 3.1] as an application of the weak transport duality theorem. Notably, this approach extends the result of Bowles and Ghoussoub from compact to general Polish spaces without additional effort.[2]

To present the notion of linear transfer, we introduce some notation. The basic object of interest are functionals $\mathcal{T}\colon \mathcal{P}(X) \times \mathcal{P}_t(Y) \to \mathbb{R} \cup \{+\infty\}$. We will use the Legendre transform $\mathcal{T}_\mu^*$ of $\mathcal{T}_\mu = \mathcal{T}(\mu, \cdot)$, which is given by

$$\mathcal{T}_\mu^*(g) = \sup_{\nu\in\mathcal{P}_t(Y)} \nu(g) - \mathcal{T}(\mu, \nu).$$

Since the set $\Phi_t(Y)$ is in separating duality with $\mathcal{M}_t(Y)$, the Fenchel duality theorem [54, Theorem 2.3.3] states

$$\mathcal{T}_\mu(\nu) = \mathcal{T}_\mu^{**}(\nu) = \sup_{g\in\Phi_t(Y)} \nu(g) - \mathcal{T}_\mu^*(g), \tag{7.1}$$

if $\mathcal{T}_\mu$ is proper convex, bounded from below and lower semicontinuous.

Definition 7.1. *A proper, convex, bounded from below and lower semicontinuous functional $\mathcal{T}\colon \mathcal{P}(X) \times \mathcal{P}_t(Y) \to \mathbb{R} \cup \{+\infty\}$ is called backward linear transfer if there exists a map T from $\Phi_t(Y)$ to the set of universally measurable functions on X bounded from below by an element in $\Phi_t(X)$, with the following property: for each $\mu \in \mathcal{P}(X)$ with $\inf_{\nu\in\mathcal{P}_t(Y)} \mathcal{T}(\mu, \nu) < \infty$ the Legendre transform $\mathcal{T}_\mu^*$ can be represented as*

$$\mathcal{T}_\mu^*(g) = \mu(T(g)) \quad \forall g \in \Phi_t(Y). \tag{7.2}$$

Theorem 7.2. *Let $\mathcal{T}\colon \mathcal{P}(X) \times \mathcal{P}_t(Y) \to \mathbb{R} \cup \{+\infty\}$ be such that*

$$\forall x \in X,\ \exists p \in \mathcal{P}_t(Y)\colon \quad \mathcal{T}(\delta_x, p) < \infty. \tag{7.3}$$

Then the following are equivalent

(i) $\mathcal{T}$ is a backward linear transfer,

[2] As noted on [18, page 3], the right setting for most applications of linear transfers should be '... complete metric spaces, Riemannian manifolds or at least $\mathbb{R}^n$.'. In this respect the extension of [18, Theorem 3.1] beyond the compact setup seems relevant.

(ii) there is a lower semicontinuous cost function $C\colon X \times \mathcal{P}_t(Y) \to \mathbb{R} \cup \{+\infty\}$ which is bounded from below and convex in the second argument such that for all $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}_t(Y)$

$$\mathcal{T}(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x, \pi_x)\mu(dx).$$

Proof. Evidently, if $\mathcal{T}$ is given as a backward linear transfer, the cost function C has to satisfy $C(x,p) = \mathcal{T}(\delta_x, p)$ and T satisfies on $\Phi_t(Y)$

$$T(g)(x) = \mathcal{T}^*_{\delta_x} g = \sup_{p \in \mathcal{P}_t(Y)} p(g) - C(x,p) = -R_C(-g)(x).$$

Hence by (7.2) and optimal weak transport duality (Theorem 3.3) we have

$$\begin{aligned}
\mathcal{T}(\mu,\nu) = \mathcal{T}_\mu(\nu) &= \sup_{g \in \Phi_t(Y)} \nu(g) - \mathcal{T}^*_\mu(g) \\
&= \sup_{g \in \Phi_t(Y)} \nu(g) - \int_X R_C(-g)(x)\mu(dx) \\
&= \sup_{g \in \Phi_t} -\nu(g) + \int_X R_C g(x)\mu(dx) \\
&= \inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x,\pi_x)\mu(dx).
\end{aligned}$$

Conversely, if (ii) holds, then Theorem 3.3 reveals

$$\mathcal{T}_\mu(\nu) = \sup_{g \in \Phi_t(Y)} \nu(g) - \mu(T(g)),$$

for $T(g)(x) = R_C(-g)(x)$, and by (3.4) we have $\mu(T(g)) = \mathcal{T}^*_\mu(g)$. □

8. Semimartingale Transport Duality

In this part we need to set up some terminology before stating the actual problem. Let

$$C = C([0,1]; \mathbb{R}^d)$$

denote the continuous path space equipped with the supremum norm $\|\cdot\|_\infty$ and its Borel σ-field. With

$$W = (W(t))_{t \in [0,1]}$$

we denote the canonical (coordinate) process on C, defined by $W(t)(\omega) = \omega(t)$, so that W is a standard d-dimensional Brownian motion under the Wiener measure $\mathbb{W}$. Let $\mathcal{F} = (\mathcal{F}_t)_{t \in [0,1]}$ denote the $\mathbb{W}$-complete filtration generated by W. As usual, we denote by $L^0(\mathbb{W})$ the space of (real-valued) random variables quotiented with the $\mathbb{W}$-a.s. identification, and by $L^\infty(\mathbb{W})$ the essentially bounded elements of $L^0(\mathbb{W})$. We will likewise identify processes that are $dt \times d\mathbb{W}$-almost surely equal. Finally, we denote by $\mathcal{S}^d_+$ the set of symmetric positive semi-definite matrices of size $d \times d$. We fix from now on a matrix norm on $\mathbb{R}^{d \times d}$.

We consider

$$g : [0,1] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{S}^d_+ \to \mathbb{R} \cup \{\infty\},$$

and assume

Assumption 8.1.

(1) *g is jointly measurable and lower-bounded.*

(2) *For each $t \in [0,1]$ the function*

$$\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{S}_+^d \ni (q,a) \mapsto g(t,x,q,a),$$

is jointly lower semicontinuous. Furthermore

$$(q,a) \mapsto g(t,x,q,a)$$

is convex for each fixed (t,x).

(3) *Either g is finite and coercive in the sense that*

$$\lim_{|q|\vee|a|\to\infty} \inf_{t,x} \frac{g(t,x,q,a)}{|q|+|a|} = +\infty, \tag{8.1}$$

or

$$dom(g(t,x,\cdot,\cdot)) := \{(q,a) \in \mathbb{R}^d \times \mathcal{S}_+^d \ : \ g(t,x,q,a) < \infty\}$$

is a compact convex set which does not depend of (t,x).

For $Q \in \mathcal{P}(C)$ we denote by

$$\mathsf{M}_0^{ac}(Q)$$

the space of continuous $\mathbb{R}^d$-valued Q-martingales, which are started at zero, whose quadratic variation matrix is absolutely continuous and integrable: Namely $M \in \mathsf{M}_0^{ac}(Q)$ iff it is a Q-martingale started at zero, $\frac{d\langle M\rangle_t}{dt}$ exists Q-a.s. and

$$\mathbb{E}^Q[|\langle M\rangle(1)|] < \infty.$$

This last condition is equivalent to asking

$$\mathbb{E}^Q\left[\int_0^1 \left|\frac{d\langle M\rangle(t)}{dt}\right| dt\right] < \infty.$$

On the other hand we write $\mathcal{L}^1(Q)$ for the set of progressively measurable $\mathbb{R}^d$-valued processes which are integrable with respect to $dt \times dQ$.

We can now introduce the set of semimartingale laws relevant to our work:

$$S := \left\{ Q \in \mathcal{P}(C) : \begin{array}{l} W(\cdot) = \int_0^\cdot q^Q(s)ds + M^Q(\cdot) \text{ under } Q, \\ \text{for some } M^Q \in \mathsf{M}_0^{ac}(Q),\ q^Q \in \mathcal{L}^1(Q) \end{array} \right\}$$

We remark that for $Q \in S$ the process q^Q above is uniquely determined. Likewise, the $\mathcal{S}_+^d$-valued process

$$a^Q := \frac{d\langle M^Q\rangle_t}{dt}$$

is uniquely determined.

Let

$$\alpha^g : S \to \mathbb{R} \cup \{+\infty\},$$

be given by

$$\alpha^g(Q) := \mathbb{E}^Q\left[\int_0^1 g(t, W(t), q^Q(t), a^Q(t))\,dt\right]. \tag{8.2}$$

Note that $\alpha^g(Q)$ is well-defined and takes values in $\mathbb{R}\cup\{+\infty\}$, as g is bounded from below. As a final bit of notation, we introduce $\Pi(\mu,\nu)$ for the set of those $Q \in S$ with initial and final marginals equal to μ and ν respectively. We further write $\Pi(\mu,\cdot)$ when only the inital marginal is prescribed.

For $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$ we define[3]

$$V(\mu,\nu) := \inf_{Q\in\Pi(\mu,\nu)} \alpha^g(Q).$$

This is a stochastic mass transport problem (or optimal transport of semimartingales) as introduced by Tan and Touzi [50]. Specification of α^g allow to cover classical optimal transport, martingale transport, and some instances of the Schrödinger problem in this framework. We now prove a duality result, originally obtained by the aforementioned authors, by means of optimal weak transport:

Theorem 8.2. *Under the standing assumptions, we have*

$$V(\mu,\nu) = \sup_{\psi\in C_b(\mathbb{R}^d)} \left\{\mu(\tilde\Psi) - \nu(\psi)\right\},$$

where $\tilde\psi(x) := \inf_{Q\in\Pi(\delta_x,\cdot)} \mathbb{E}^Q\left[\int_0^1 g(t, W(t), q^Q(t), a^Q(t))\,dt + \Psi(X_1)\right]$.

Proof. Define $C(x,p) := V(\delta_x, p)$.

Remark that if $Q_1 \in \Pi(\delta_x, p_1)$ and $Q_2 \in \Pi(\delta_x, p_2)$, then for $\alpha \in [0,1]$ we have $\alpha Q_1 + (1-\alpha)Q_2 \in \Pi(\delta_x, \alpha p_1 + (1-\alpha)p_2)$. As we will see in Theorem 8.3, the function $\alpha^g(\cdot)$ is convex. From these facts it follows that $C(x,\cdot)$ is convex for each x fixed. Further, Theorem 8.3 also shows that $\alpha^g(\cdot)$ is coercive and lower semicontinuous, which then implies that $C(x,\cdot)$ is lower semicontinuous for each x fixed.

A standard measurable selection argument shows that

$$V(\mu,\nu) = \inf_{\pi\in Cpl(\mu,\nu)} \int V(\delta_x, \pi^x)\mu(dx) = \inf_{\pi\in Cpl(\mu,\nu)} \int C(x,\pi^x)\mu(dx),$$

where we wrote $Cpl(\mu,\nu)$ for the set of measures in $\mathbb{R}^d\times\mathbb{R}^d$ with the given marginals. Applying the duality (3.3) we deduce

$$V(\mu,\nu) = \sup_{\psi}\left\{\int R_C\psi(x)\mu(dx) - \nu(\psi)\right\},$$

[3]Equivalently, one may minimize the functional $\mathbb{E}[\int_0^1 g(t, X(t), q(t), \sigma(t)\sigma'(t))dt]$ over all semimartingales $dX(t) = q(t)dt + \sigma(t)dB(t)$ on some stochastic basis, such that $X(0) \sim \mu, X(1) \sim \nu$.

where

$$
\begin{aligned}
R_C\psi(x) &:= \inf_p \{p(\psi) + C(x,p)\} \\
&= \inf_{Q\in\Pi(\delta_x,\cdot)} \mathbb{E}^Q\left[\int_0^1 g(t,W(t),q^Q(t),a^Q(t))\,dt + \Psi(X_1)\right] \\
&= \tilde{\psi}(x).
\end{aligned}
$$

□

For the previous result we employed:

Theorem 8.3. *The functional α^g is convex, lower semicontinuous with respect to weak convergence, and coercive in the sense that $\{Q : \alpha^g(Q) \le c\}$ is weakly compact for each $c \in \mathbb{R}$.*

In order to prove this we need the following auxiliary result first:

Lemma 8.4. *Suppose $(q_n)_n$ is a sequence of $L^1([0,1],dt;\mathbb{R}^d)$-valued random variables possibly defined in different probability spaces, and call $A_n(t) := \int_0^t q_n(s)ds$. In the same space where q_n is defined we are given a further C-valued random variable M_n such that $M_n(0) = 0$, M_n is a martingale wrt. its completed filtration, and such that $a_n(t) := \frac{\langle M_n\rangle(t)}{dt}$ exists a.s. Finally assume the existence of $c > 0$ such that, for all n,*

$$
\mathbb{E}\left[\int_0^t g(t, M_n(t) + A_n(t), q_n(t), a_n(t))dt\right] \le c. \tag{8.3}
$$

Then there exist an $L^1([0,1],dt;\mathbb{R}^d)$-valued random variable A, a C-valued random variable M, and subsequences A_{n_k} and M_{n_k} such that

1. *A_{n_k} converges in law in C to A,*
2. *M_{n_k} converges in law in C to M,*
3. *$A(t) = \int_0^t q(s)ds$, some $L^1([0,1],dt;\mathbb{R}^d)$-valued random variable q,*
4. *$\langle M\rangle(t) = \int_0^t a dt$, some $\mathcal{S}^d_+$-valued process a with $\mathbb{E}\left[\int_0^1 |a(t)|dt\right] < \infty$*
5. *the following inequality holds:*

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^t g(t, M(t) + A(t), q(t), a(t))dt\right] \\
&\le \liminf_{k\to\infty} \mathbb{E}\left[\int_0^t g(t, M_{n_k}(t) + A_{n_k}(t), q_{n_k}(t), a_{n_k}(t))dt\right]
\end{aligned} \tag{8.4}
$$

In particular, the laws of (A_n) and $(M_n)_n$ form tight sequences.

The following proofs follow very closely the arguments in [8]. As a small technical improvement over [50], we observe that the coercivity condition (3.1) assumed here, is weaker than the analogue in the cited paper.

Proof of Lemma 8.4.
We first check tightness. If the final part of Assumption 8.1(3) holds, this is trivial. Otherwise, by (8.1), for each $r > 0$ we may find $N > 0$ such that $g(t,x,q,a) \ge r|q| + r|a|$ whenever $|q| \vee |a| \ge N$. Moreover, there

exists $b \geq 0$ such that $g(t, x, q, a) \geq -b$. In particular, for all (t, q) we have $|q| + |a| \leq 2N + \frac{1}{r}(g(t, x, q, a) + b)$. Hence, for $0 \leq s < t \leq 1$,

$$\begin{aligned} |A_n(t) - A_n(s)| &\leq \int_s^t |q_n(u)| du \\ &\leq \frac{1}{r} \int_s^t (g(u, W(u), q_n(u), a_n(u)) + b)\, du + 2N(t - s) \\ &\leq \frac{1}{r} \int_0^1 g(u, W(u), q_n(u), a_n(u))\, du + \frac{b}{r} + N(t - s). \end{aligned}$$

Hence, for any $\delta_n \downarrow 0$, (8.3) yields

$$\limsup_{n\to\infty} \sup_\tau \mathbb{E}|A_n(\tau + \delta_n) - A_n(\tau)| \leq \limsup_{n\to\infty} \left(\frac{c + b}{r} + N\delta_n \right) = \frac{c + b}{r},$$

where the $\sup_\tau$ is over all stopping times with values in $[0, 1 - \delta_n]$. As $r > 0$ was arbitrary, this shows that

$$\limsup_{n\to\infty} \sup_\tau \mathbb{E}|A_n(\tau + \delta_n) - A_n(\tau)| = 0,$$

and from Aldous' criterion for tightness [37, Theorem 16.11] we conclude that (A_n) is tight. The Cauchy-Schwartz inequality and similar calculations allow to conclude that

$$\mathbb{E}[|M_n(\tau + \delta) - M_n(\tau)|] \leq \sqrt{N\delta + \frac{c + b}{r}},$$

so as before (M_n) is a tight sequence. Furthermore, (M_n) is in fact precompact in the 1-Wasserstein space $\mathcal{W}_1(C)$ of measures on C which integrate the supremum of the norm of the path of the canonical process. This follows by showing that

$$\lim_{K\to\infty} \sup_n \mathbb{E}\left[\sup_{t\leq 1} |M_n(t)| 1_{\sup_{t\leq 1} |M_n(t)|\geq K} \right] = 0,$$

which is a consequence of Cauchy-Schwartz, Doob's inequality, and Assumption 8.1(3).

Passing to a subsequence and applying Skorokhod's representation, let us now assume that there exists continuous process A and M such that $A_n \to A$ and $M_n \to M$ almost surely (in C), with all processes defined on some common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The process M is in fact a martingale thanks to $\mathcal{W}_1(C)$-precompactness. From Assumption 8.1(3) and a standard argument as in the de la Vallée Poisson Theorem, we conclude that $\{q_n : n \in \mathbb{N}\} \subseteq L^1 := L^1([0, 1] \times \Omega,\ dt \otimes d\mathbb{P})$ is uniformly integrable and thus weakly precompact. Similarly, using further that $\{a_n : n \in \mathbb{N}\}$ is weakly precompact in the Bochner space $L^1_d := L^1([0, 1] \times \Omega,\ dt \otimes d\mathbb{P}; \mathbb{R}^{d\times d})$ of matrix-valued integrable processes if and only if $\{|a_n| : n \in \mathbb{N}\}$ is uniformly integrable (cf. [26]), we deduce that $\{a_n : n \in \mathbb{N}\}$ is weakly precompact. By passing to a further subsequence, we may now assume that $q_n \to q$ weakly in L^1, that $a_n \to a$ weakly in L^1_d, and that a is almost surely $\mathcal{S}^d_+$-valued. Because g is bounded from below and lower semicontinuous in its last three variables, the map $(\bar{X}, \bar{q}, \bar{a}) \mapsto \mathbb{E} \int_0^1 g(t, \bar{X}(t), \bar{q}(t), \bar{a}(t)) dt$ is lower semicontinuous in the norm topology of $C \times L^1 \times L^1_d$, by Fatou's lemma. Because it is also convex in the last two, this map is therefore weakly lower semicontinuous when $L^1 \times L^1_d$ is given the weak topology. This yields (8.4). By dominated convergence, it holds for each bounded random variable Z that

$$\mathbb{E}[ZA(t)] = \lim_{n\to\infty} \mathbb{E}[ZA_n(t)] = \lim_{n\to\infty} \mathbb{E}\left[Z \int_0^t q_n(s)ds\right] = \mathbb{E}\left[Z \int_0^t q(s)ds\right].$$

Hence $A(t) = \int_0^t q(s)ds$ a.s. for each t, and by continuity we have $A(\cdot) = \int_0^\cdot q(s)ds$ a.s. A similar argument shows that for each $1 \le i, j \le d$ and $0 \le s \le t \le 1$ and Z measurable up to time s, we have

$$\begin{aligned} 0 = &\lim_{n\to\infty} \mathbb{E}\left[Z\left(M^i_n M^j_n(t) - M^i_n M^j_n(s) - \int_s^t a^{i,j}_n(r)dr\right)\right] \\ =&\mathbb{E}\left[Z\left(M^i M^j(t) - M^i M^j(s) - \int_s^t a^{i,j}(r)dr\right)\right], \end{aligned}$$

from which $\langle M\rangle(\cdot) = \int_0^\cdot a(r)dr$. □

Proof of Theorem 8.3.

Convexity: Let $\lambda \in [0, 1]$, and fix $Q_0, Q_1 \in S$. We work on the extended probability space $C \times \{0, 1\}$, and we write (W, X) to denote the identity map on this space. We define a measure M on $C \times \{0, 1\}$ by requiring that the second marginal of M be $\lambda\delta_0 + (1 - \lambda)\delta_1$, and the conditional law of W given X be Q_X. In particular, the first marginal of M is precisely $Q := \lambda Q_0 + (1 - \lambda)Q_1$. Abbreviate $q_i := q^{Q_i}$ and $a_i := a^{Q_i}$. It easily follows that the process

$$W(t) - \int_0^t q_X(s)ds$$

defines an M-martingale with respect to the filtration $\overline{\mathcal{F}} = (\overline{\mathcal{F}}_t)_{t\in[0,1]}$ defined by $\overline{\mathcal{F}}_t = \mathcal{F}_t \otimes \sigma(X)$ on the product space. Furthermore, the quadratic variation of $W(\cdot) - \int_0^\cdot q_X(s)ds$ has a density explicitly given by $t \mapsto a_X(t)$. Now define the processes $q = (q(t))_{t\in[0,1]}$ and $a = (a(t))_{t\in[0,1]}$ respectively as the optional projections of the processes $(q_X(t))_{t\in[0,1]}$ and $(a_X(t))_{t\in[0,1]}$ on the filtration $\mathcal{F}$ generated by W. In particular,

$$\begin{aligned} q(t) &= \mathbb{E}^M[q_X(t)\,|\,(W_s)_{s\le t}] = \mathbb{E}^M[\mathbf{1}_{X=0}q_0(t) + \mathbf{1}_{X=1}q_1(t)\,|\,(W_s)_{s\le t}], \\ a(t) &= \mathbb{E}^M[a_X(t)\,|\,(W_s)_{s\le t}] = \mathbb{E}^M[\mathbf{1}_{X=0}a_0(t) + \mathbf{1}_{X=1}a_1(t)\,|\,(W_s)_{s\le t}], \end{aligned}$$

A few computations reveal that $W(\cdot) - \int_0^\cdot q(t)dt$ is still an M-martingale. On the other hand, since $W(\cdot) - \int_0^\cdot q_X(s)ds$ has a_X as the density of its quadratic variation under M, so does W itself. But then for all $i \le i, j \le d$, if R is bounded and $\mathcal{F}_t$-measurable and $0 \leq h \leq 1-t$, we have

$$
\begin{aligned}
0 &= \mathbb{E}^M\left[\left(W^iW^j(t+h) - W^iW^j(t) - \int_t^{t+h} a_X^{i,j}(s)ds\right)R\right] \\
&= \mathbb{E}^M\left[\left(W^iW^j(t+h) - W^iW^j(t) - \int_t^{t+h} a^{i,j}(s)ds\right)R\right] \\
&= \mathbb{E}^Q\left[\left(W^iW^j(t+h) - W^iW^j(t) - \int_t^{t+h} a^{i,j}(s)ds\right)R\right],
\end{aligned}
$$

recalling that Q is the first marginal of M. Since the M martingale $W(\cdot) - \int_0^{\cdot} q(t)dt$ is $\mathcal{F}$-adapted, it follows that it is a Q-martingale (when seen as living in the filtered probability space $(C, \mathcal{F}, Q)$) and the above display shows that the density of its quadratic variation is precisely a. In summary, we conclude that $Q \in S$, and that $q = q^Q$ as well as $a = a^Q$. Finally, using Jensen's inequality, we compute

$$
\begin{aligned}
&\lambda\alpha^g(Q_0) + (1-\lambda)\alpha^g(Q_1) \\
=&\lambda\mathbb{E}^{Q_0}\left[\int_0^1 g(t, W(t), q_0(t), a_0(t))dt\right] + (1-\lambda)\mathbb{E}^{Q_1}\left[\int_0^1 g(t, W(t), q_1(t), a_1(t))dt\right] \\
=&\mathbb{E}^M\left[\int_0^1 g(t, W(t), q_X(t), a_X(t))dt\right] \\
\geq&\mathbb{E}^M\left[\int_0^1 g(t, W(t), q(t), a(t))dt\right] \\
=&\mathbb{E}^Q\left[\int_0^1 g(t, W(t), q(t), a(t))dt\right] \\
=&\alpha^g(Q).
\end{aligned}
$$

Inf-compactness: Let $c \in \mathbb{R}$ and $\Lambda_c := \{Q : \alpha^g(Q) \leq c\}$. It is convenient in this step and the next to define

$$
W^Q(t) := W(t) - \int_0^t q^Q(s)ds, \quad t \in [0,1],
$$

for $Q \in S$, noting that W^Q is a Q-martingale with volatility a^Q. Letting $A^Q(t) := \int_0^t q^Q(s)ds$, it follows from Lemma 8.4 that $\{Q \circ (A^Q)^{-1} : Q \in \Lambda_c\} \subseteq \mathbb{P}(C)$ is tight. On the other hand, $\{Q \circ (W^Q)^{-1} : Q \in \Lambda_c\}$ is tight as well by the same argument. Since each marginal is tight, we deduce that $\{Q \circ (W^Q, A^Q)^{-1} : Q \in \Lambda_c\} \subseteq \mathbb{P}(C \times C)$ is tight. Finally, by continuous mapping, the set $\{Q \circ (W^Q + A^Q)^{-1} : Q \in \Lambda_c\} = \Lambda_c$ is tight.

Lower semicontinuity: Suppose $\{Q_n : n \in \mathbb{N}\} \subseteq \Lambda_c$ with $Q_n \to Q$ weakly for some $Q \in \mathbb{P}(C)$. We must show that Q belongs to Λ_c. Define the continuous process

$$
A_n(t) := \int_0^t q^{Q_n}(s)ds = W(t) - W^{Q_n}(t),
$$

for each n. As in the previous point, $\{Q_n \circ (W, W^{Q_n}, A^n)^{-1} : n \in \mathbb{N}\}$ is tight. Relabelling a subsequence, suppose that $Q_n \circ (W, W^{Q_n}, A^n)^{-1}$ converges weakly to the law of some C^3-valued random variable (X, B, A). Using Lemma 8.4, we may assume also that $A(t) = \int_0^t q(s)ds$ and $a(t) := \frac{\langle B\rangle(t)}{dt}$ satisfying

$$\mathbb{E}\int_0^1 g(t, X(t), q(t), a(t))dt \leq \liminf \mathbb{E}^{Q_n}\int_0^1 g(t, W(t), q^{Q_n}(t), a^{Q_n}(t))dt \leq c.$$

Clearly W^{Q_n} is a martingale in the filtration of (W, W^{Q_n}, A^n), and hence B is a martingale in the filtration of (X, B, A). Finally, notice that

$$X(t) = B(t) + A(t) = B(t) + \int_0^t q(s)ds,$$

as the same relation holds in the pre-limit. A standard argument shows that $X - \int_0^\cdot \widehat{q}(s)ds$ is a martingale with respect to the filtration of X, where $\widehat{q}$ is the optional projection of q onto such filtration. Writing

$$X(t) = B(t) + \int_0^t [q(s) - \widehat{q}(s)]ds + \int_0^t \widehat{q}(s)ds =: \tilde{B}(t) + \int_0^t \widehat{q}(s)ds,$$

we deduce that $\tilde{B}$ is an X-adapted martingale with density of quadratic variation a. By convexity of $g(t, x, \cdot, \cdot)$, we have

$$\mathbb{E}\int_0^1 g(t, X(t), \widehat{q}(t), a(t))dt \leq \mathbb{E}\int_0^1 g(t, X(t), q(t), a(t))dt \leq c.$$

Recalling that Q denoted the law of X, we conclude that $Q \in \Lambda_c$.

□

References

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

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

[3] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. Adapted Wasserstein Distances and Stability in Mathematical Finance. *arXiv e-prints*, 2019.

[4] J. Backhoff-Veraguas, M. Beiglböck, M. Huesmann, and S. Källblad. Martingale Benamou–Brenier: a probabilistic perspective. *Ann. Probab., to appear*, 2020.

[5] J. Backhoff-Veraguas, M. Beiglböck, Y. Lin, and A. Zalashko. Causal transport in discrete time and applications. *SIAM Journal on Optimization*, 27(4):2528–2562, 2017.

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

[7] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Weak monotone rearrangement on the line. *Electronic Communications in Probability*, 25, 2020.

[8] J. Backhoff-Veraguas, D. Lacker, and L. Tangpi. Non-exponential Sanov and Schilder theorems on Wiener space: BSDEs, Schrödinger problems and Control. *Forthcoming at Annals of Applied Probability*, 2018.

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

[10] M. Beiglböck. Cyclical monotonicity and the ergodic theorem. *Ergodic Theory Dynam. Systems*, 35(3):710–713, 2015.
[11] M. Beiglböck, A. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. *Invent. Math.*, 208(2):327–400, 2017.
[12] M. Beiglböck and C. Griessler. A land of monotone plenty. *Annali della SNS, to appear*, Apr. 2016.
[13] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices—a mass transport approach. *Finance Stoch.*, 17(3):477–501, 2013.
[14] M. Beiglböck, N. Juillet, et al. On a problem of optimal transport under marginal martingale constraints. *The Annals of Probability*, 44(1):42–106, 2016.
[15] M. Beiglboeck and N. Juillet. Shadow couplings. *ArXiv e-prints*, Sept. 2016.
[16] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.
[17] S. Bianchini and L. Caravenna. On optimality of c-cyclically monotone transference plans. *C. R. Math. Acad. Sci. Paris*, 348(11-12):613–618, 2010.
[18] M. Bowles and N. Ghoussoub. Mather measures and ergodic properties of kantorovich operators associated to general mass transfers. 2019.
[19] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(19):805–808, 1987.
[20] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.
[21] M. Colombo, L. De Pascale, and S. Di Marino. Multimarginal optimal transport maps for one-dimensional repulsive costs. *Canad. J. Math.*, 67(2):350–368, 2015.
[22] A. M. Cox and M. Vidmar. The structure of non-linear martingale optimal transport problems. *arXiv preprint arXiv:1903.06606*, 2019.
[23] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, pages 146–158, 1975.
[24] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013.
[25] C. Daskalakis, A. Deckelbaum, and C. Tzamos. Strong Duality for a Multiple-Good Monopolist. *Econometrica*, 85(3):735–767, 2017.
[26] J. Diestel. Uniform integrability: an introduction. 1991.
[27] S. Eckstein and M. Kupper. Computation of optimal transport and related hedging problems via penalization and neural networks. *Applied Mathematics & Optimization*, pages 1–29, 2019.
[28] M. Fathi, N. Gozlan, and M. Prodhomme. A proof of the caffarelli contraction theorem via entropic regularization. *arXiv preprint arXiv:1904.06053*, 2019.
[29] M. Fathi and Y. Shu. Curvature and transport inequalities for Markov chains in discrete spaces. *Bernoulli*, 24(1):672–698, 2018.
[30] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.*, 24(1):312–336, 2014.
[31] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.
[32] N. Gozlan and N. Juillet. On a mixture of brenier and strassen theorems. *arXiv preprint arXiv:1808.02681*, 2018.
[33] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali. Characterization of a class of weak transport-entropy inequalities on the line. *Ann. Inst. Henri Poincaré Probab. Stat.*, 54(3):1667–1693, 2018.
[34] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, 273(11):3327–3405, 2017.
[35] C. Griessler. c-cyclical monotonicity as a sufficient criterion for optimality in the multi-marginal Monge-Kantorovich problem. *ArXiv e-prints*, Jan. 2016.

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

[37] O. Kallenberg. *Foundations of modern probability*. Springer Science & Business Media, 2006.

[38] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.

[39] K. Marton. A measure concentration inequality for contracting markov chains. *Geometric & Functional Analysis GAFA*, 6(3):556–571, 1996.

[40] K. Marton et al. Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration. *The Annals of Probability*, 24(2):857–866, 1996.

[41] T. Mikami and M. Thieullen. Duality theorem for the stochastic optimal control problem. *Stochastic Process. Appl.*, 116(12):1815–1835, 2006.

[42] M. Nutz and F. Stebegg. Canonical Supermartingale Couplings. *Ann. Probab., to appear*, Sept. 2018.

[43] B. Pass. On the local structure of optimal measures in the multi-marginal optimal transportation problem. *Calc. Var. Partial Differential Equations*, 43(3-4):529–536, 2012.

[44] P.-M. Samson. Transport-entropy inequalities on locally acting groups of permutations. *Electron. J. Probab.*, 22:Paper No. 62, 33, 2017.

[45] Y. Shu. From hopf-lax formula to optimal weak transfer plan. *arXiv preprint arXiv:1609.03405*, 2016.

[46] Y. Shu. Hamilton-Jacobi equations on graph and applications. *Potential Anal.*, 48(2):125–157, 2018.

[47] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.

[48] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 81(1):73–205, 1995.

[49] M. Talagrand. New concentration inequalities in product spaces. *Inventiones mathematicae*, 126(3):505–563, 1996.

[50] X. Tan and N. Touzi. Optimal transportation under controlled stochastic dynamics. *Ann. Probab.*, 41(5):3201–3240, 2013.

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

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

[53] D. Zaev. On the Monge-Kantorovich problem with additional linear constraints. *Mathematical Notes*, 98(5-6):725–741, 2015.

[54] C. Zalinescu. *Convex analysis in general vector spaces*. World scientific, 2002.

[55] X. Zhou. On the fenchel duality between strong convexity and lipschitz continuous gradient. *arXiv preprint arXiv:1803.06573*, 2018.

WEAK TRANSPORT FOR NON-CONVEX COSTS AND MODEL-INDEPENDENCE IN A FIXED-INCOME MARKET

B. ACCIAIO, M. BEIGLBÖCK, AND G. PAMMER

Abstract. We consider a model-independent pricing problem in a fixed-income market and show that it leads to a weak optimal transport problem as introduced by Gozlan et al. We use this to characterize the extremal models for the pricing of caplets on the spot rate and to establish a first robust super-replication result that is applicable to fixed-income markets.

Notably, the weak transport problem exhibits a cost function which is non-convex and thus not covered by the standard assumptions of the theory. In an independent section, we establish that weak transport problems for general costs can be reduced to equivalent problems that do satisfy the convexity assumption, extending the scope of weak transport theory. This part could be of its own interest independent of our financial application, and is accessible to readers who are not familiar with mathematical finance notations.

keywords: fixed-income markets, robust pricing and hedging, weak transport problem

1. Introduction

Prompted by Hobson's seminal paper (Hobson, 1998), the field of model-independent finance has seen a steep development. Typically the framework consists in a market where some assets are dynamically traded, while some derivatives are statically traded at time zero. Then, without imposing any model or underlying probability measure, one seeks for robust pricing bounds for other derivatives. Usually the payoffs are already expressed in discounted terms, and the pricing problem corresponds to looking for market-compatible martingale measures, that is, probability measures on the path space such that the discounted asset prices - taken to be the canonical processes - are martingales, and reproduce the observed marked prices.

A widely used assumptions is the availability, for static trading, of call options with maturity T written on an asset S, for all strikes K, which determines the distribution of S_T under the pricing measure. However, this reasoning crucially relies on the existence of a deterministic numéraire for discounting. This point is crucial for the use of tools from the Optimal Transport and Skorokhod embedding, see Hobson (2011); Obłój (2004); Beiglböck et al. (2013); Galichon et al. (2014); Beiglböck and Juillet (2016); Dolinsky and Soner (2014); Campi et al. (2017); Beiglböck et al. (2017); Cox et al. (2019); Cheridito et al. (2019), among many others.

In the present paper, we do not assume the existence of a deterministic bank account, or of a bank account at all. Instead, we consider a fixed-income market, where bonds are dynamically traded, and some options on them are statically traded. A similar setting is considered by Acciaio and Grbac (2020), who assume finitely many call options with possibly different maturities being traded on a bond, and investigate market consistency with absence of arbitrage. This analysis follows the spirit of Davis and Hobson (2007), and displays in a simple market the different characteristics of a robust framework when stochastic discounting is allowed. On the other hand, the focus of the present paper is on robust pricing. We will use bonds as numéraires, thus the pricing measures will be the so-called forward measures, a notion that we recall in the next paragraph.

For $T > 0$, we refer to a (zero coupon) bond with maturity T as a T-bond, and denote its price at time $t \le T$ by $p(t, T)$. A T-forward measure $\mathbb{Q}_T$ is a probability measure on $\mathcal{F}_T$ such that every traded asset expressed in units of the T-bond is a martingale. The pricing formula, at time s, for a claim Φ with maturity t, for $s \le t \le T$, is then given by

$$\frac{V_s(\Phi)}{p(s,T)} = \mathbb{E}^{\mathbb{Q}_T}\left[\frac{\Phi}{p(t,T)}\middle|\mathcal{F}_s\right]. \tag{1.1}$$

In particular, having the prices at time 0 of all call options with maturity T on a given asset, identifies the distribution of that asset at time T under the T-forward measure.

1.1. **Robust pricing setting.**

Setup 1.1. *We consider maturities* $0 < T_1 < T_2 < T_3$, *and let* T_2*- and* T_3*-bonds be (dynamically) traded in the market. We also let call options with maturity* T_1 *on the* T_2*-bond, and call options with maturity* T_2 *on the* T_3*-bond be (statically) traded at time zero, for every strike* K*, and denote the respective prices by* $C(T_1, T_2, K)$ *and* $C(T_2, T_3, K)$*.*

A main contribution of this paper is to provide a robust superreplication theorem that applies to general derivatives written on $p(T_1, T_2)$, $p(T_1, T_3)$, i.e. we will consider payoffs of the form

$$\Phi(p(T_1, T_2), p(T_1, T_3)). \tag{1.2}$$

Since puts with maturity T_i on T_j-bonds correspond to caplets on the spot rate $F(T_i, T_j)$, our setting in particular captures the case where $T_{i+1} - T_i = 6$ months, as 6 months caplets are much more liquid than the 12 months ones. Indeed, in this specific case we also identify the extremal models, leading to lower and upper pricing bounds, as well as the corresponding sub- and superreplication strategies. (see Section 2.2).

1.2. **Robust superreplication theorem.** In our setup it is most convenient to take the T_2-bond as a numéraire. In particular, we consider the *discounted* processes

$$X_i := \frac{p(T_i,T_1)}{p(T_i,T_2)}, i = 0, 1, \quad Y_i := \frac{p(T_i,T_3)}{p(T_i,T_2)}, i = 0, 1, 2. \tag{1.3}$$

For the convenience of the reader, we detail in Section 2.1 below how to switch between undiscounted quantities and quantities expressed in T_2-bond units. In the present introductory section we will consistently refer to quantities expressed in T_2-bonds. The assumption in Setup 1.1 then asserts that options on X_1 and Y_2 are liquidly traded. In the spirit of Breeden and Litzenberger (1978), this amounts to the identification of probabilities μ, ν on $\mathbb{R}_+$ with finite first moments, such that for derivatives φ, ψ

$$\text{price}(\varphi(X_1)) = \int \varphi \, d\mu, \quad \text{price}(\psi(Y_2)) = \int \psi \, d\nu. \tag{1.4}$$

In view of the classical (in the sense of *non-robust*) theory of fixed-income markets, we expect pricing functionals to arise from T_2-forward measures. A generic T_2-forward measure is compatible with the information given by the market if (compare (1.1))

$$\text{law}_{\mathbb{Q}_{T_2}}(X_1) \sim \mu, \quad \text{law}_{\mathbb{Q}_{T_2}}(Y_2) \sim \nu. \tag{1.5}$$

We write $Q(\mu, \nu)$ for the class of all setups $(\Omega, \mathcal{F}, (\mathcal{F}_{T_i})_{i=0}^2, \mathbb{Q}_{T_2})$, $p(T_i, T_j)$, $i \le j \le 3$, $i \le 2$, satisfying the *marginal constraints* (1.5) and the *martingale constraints*

$$(X_i)_{i=0,1}, \quad (Y_i)_{i=0,1,2} \quad \text{are } (\mathcal{F}_{T_i})\text{-martingales}. \tag{1.6}$$

We will slightly abuse notation in that we write $\mathbb{Q}_{T_2} \in Q(\mu, \nu)$ when we really mean that we run over all such setups. Writing $\hat{\Phi}(X_1, Y_2)$ for the payoff (1.2) in terms of T_2-bonds (see Section 2.1), we arrive at the primal optimization problem

$$\hat{P}_F := \sup_{\mathbb{Q}_{T_2} \in Q(\mu,\nu)} \mathbb{E}^{\mathbb{Q}_{T_2}}[\hat{\Phi}(X_1, Y_1)]. \tag{1.7}$$

All our results are still valid if T_1-bonds are also dynamically traded.

We stress that problem (1.7) differs significantly from a usual martingale optimal transport problem in that our objective functional is written on Y_1, while the marginal constraint concerns Y_2. We also note that, in contrast to most problems considered in robust finance, it is delicate to pass from (1.7) to a problem on a canonical setup, cf. Proposition 1.4.

Switching to the dual problem of (1.7), one is allowed to trade statically in vanilla payoffs φ, ψ written on X_1, Y_2 and to trade dynamically in $(Y_i)_{i=1,2}$. I.e. we consider superhedges of the form

$$\hat{\Phi}(X_1, Y_1) \leq \varphi(X_1) + \psi(Y_2) + \Delta(Y_1)(Y_2 - Y_1). \tag{1.8}$$

Since our goal is to obtain a robust hedge, we shall require that (1.8) holds for arbitrary $x_1, y_1, y_2 \in \mathbb{R}_+$. According to (1.4), the costs for this strategy (in discounted terms) amount to $\int \varphi \, d\mu + \int \psi \, d\nu$. Our main superreplication theorem is then:

Theorem 1.2. *Assume that μ, ν are probabilities on $\mathbb{R}_+$ with finite r-th moments, for some $r \geq 1$. Assume that $\hat{\Phi}(x, y)$ is upper semicontinuous and bounded from above by a multiple of $|x|^r + |y|^r$. Then we have*

$$\hat{P}_F = \inf \left\{ \int \varphi \, d\mu + \int \psi \, d\nu : \hat{\Phi}(x_1, y_1) \leq \varphi(x_1) + \psi(y_2) + \Delta(y_1)(y_2 - y_1) \text{ for all } x_1, y_1, y_2 \geq 0 \right\},$$

where the infimum is taken over continuous functions φ and ψ that are bounded in absolute terms by a multiple of $x \mapsto (1 + |x|)^r$ and Δ is an arbitrary function .

Remark 1.3. It is worth noticing that the T_3-bond plays no special role for our results, in the sense that, by replacing it with a generic asset S, we would get the same superreplication result. That is, if in Setup 1.1, instead of the T_3-bond, we assume that a generic asset S is dynamically traded and T_2-calls on it statically traded, then Theorem 1.2 holds true for any derivative of the form $\Phi(p(T_1, T_2), S_{T_1})$. In this case, from market prices we deduce the distributions under $\mathbb{Q}_{T_2}$ of $X_1 = 1/p(T_1, T_2)$ and of S_{T_2}.

We note that robust versions of the superhedging duality / FTAP have seen particular focus in robust finance, see e.g. Hobson and Neuberger (2012); Acciaio et al. (2016); Bouchard and Nutz (2015); Burzoni et al. (2017); Cheridito et al. (2017), among many others.

1.3. **Weak transport formulation.** Our proof of Theorem 1.2 is based on new results in the theory of weak optimal transport (WOT) (see Section 3). Specifically, we will show that WOT problems for general costs can be reduced to the much better understood case of WOT problems that satisfy convexity constraints. This reduction enables us to rewrite the pricing problem (1.7) as a weak transport problem for barycentric costs. To formulate this, we write $\Pi(\mu, \nu)$ for the set of all couplings of the probabilities μ, ν, and for $\pi \in \Pi(\mu, \nu)$ we denote its disintegration w.r.t. μ by $(\pi_x)_x$. For a measure p on $\mathbb{R}$ that has a finite first moment we write $b(p)$ for its barycenter. Using these notions, it is possible to rewrite (1.7) as an optimization problem over measures on $\mathbb{R}^2$.

Proposition 1.4. *Assume that μ, ν are probabilities on $\mathbb{R}_+$ with finite r-th moment, for some $r \geq 1$, and that μ is continuous. Assume that $\hat{\Phi}(x, y)$ is continuous and has at most growth of order $|x|^r + |y|^r$. Then*

$$\hat{P}_F = \sup_{\pi \in \Pi(\mu, \nu)} \int \hat{\Phi}(x, b(\pi_x)) \, d\mu(x). \tag{1.9}$$

The reformulation of the robust pricing problem given in (1.9) plays a key role in our explicit solution of the weak transport problem in the case where Φ is a put or call option (see Section 2.2). (1.9) can also be used to formulate our superreplication theorem 1.2 in the language of optimal

In principle we could allow for the hedge Δ to depend also on the value of X_1, but as part of our results we will obtain that this does not change the actual value of the superhedging problem.

In fact it suffices to consider functions Δ which are increasing.

transport. We state here the corresponding theorem to highlight the connection with the classical Monge-Kantorovich duality:

Theorem 1.5. *Assume that μ is a continuous probability on a Polish space $\mathcal{X}$ and ν a probability on $\mathbb{R}^d$, $d \in \mathbb{N}$, with finite r-th moments, for some $r \geq 1$. Assume that $c : \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}$ is continuous and has at most growth of order $d_{\mathcal{X}}(x, x_0)^r + |y|^r$ (where x_0 is an arbitrary fixed point in $\mathcal{X}$ and $d_{\mathcal{X}}$ is a compatible, complete metric on $\mathcal{X}$). Then we have*

$$\sup_{\pi \in \Pi(\mu,\nu)} \int c(x, b(\pi_x))\, d\mu(x) \tag{1.10}$$

$$= \inf \left\{ \int \varphi\, d\mu + \int \psi\, d\nu : c(x, y_1) \leq \varphi(x) + \psi(y_2) + \Delta(y_1)(y_2 - y_1) \text{ for all } x, y_1, y_2 \right\} \tag{1.11}$$

$$= \inf \left\{ \int \varphi\, d\mu + \int \psi\, d\nu : c(x, y) \leq \varphi(x) + \psi(y) \text{ for all } x, y,\ \psi \text{ convex} \right\}, \tag{1.12}$$

where the infimum is taken over continuous functions φ, ψ that are bounded in absolute terms by a multiple of $x \mapsto (1 + d_{\mathcal{X}}(x, x_0))^r$, $y \mapsto (1 + |y|)^r$, respectively, and Δ is measurable.

1.4. **Organisation of the paper.** In Section 2 we reformulate the robust pricing problem in discounted terms (with all quantities expressed in units of T_2-bonds) and illustrate in the case of caplets our results on superreplication duality and characterization of primal and dual optimizers. In Section 3 we formulate and prove our new results for the weak optimal transport problem, and from these we derive the results stated in the introduction.

2. Robust pricing problem

2.1. **On discounting.** Throughout this section, we work in the framework described in Sections 1.1-1.2, thus in particular under Setup 1.1. Recall that we discount our financial instruments by expressing them in terms of T_2-bonds, see (1.3). In particular we consider

$$X_1 = \tfrac{1}{p(T_1,T_2)}, \quad Y_1 = \tfrac{p(T_1,T_3)}{p(T_1,T_2)}, \quad Y_2 = p(T_2, T_3).$$

Then a generic T_2-forward measure is compatible with the information given by the market if for all $K \geq 0$ (cf. (1.1))

$$\begin{aligned} \frac{C(T_1, T_2, K)}{p(0, T_2)} &= \mathbb{E}^{\mathbb{Q}_{T_2}} \left[\frac{(p(T_1, T_2) - K)^+}{p(T_1, T_2)} \right] = \mathbb{E}^{\mathbb{Q}_{T_2}} \left[(1 - KX_1)^+ \right], \\ \frac{C(T_2, T_3, K)}{p(0, T_2)} &= \mathbb{E}^{\mathbb{Q}_{T_2}} \left[\frac{(p(T_2, T_3) - K)^+}{p(T_2, T_2)} \right] = \mathbb{E}^{\mathbb{Q}_{T_2}} \left[(Y_2 - K)^+ \right]. \end{aligned} \tag{2.1}$$

Therefore, by Breeden and Litzenberger (1978), from the observed call prices we can deduce the distribution under $\mathbb{Q}_{T_2}$ of X_1 and Y_2, which we denoted by μ and ν, respectively.

We are interested in a derivative of the form $\Phi(p(T_1, T_2), p(T_1, T_3))$, with maturity T_1. In terms of T_2-bonds, its payoff equals

$$\frac{\Phi(p(T_1, T_2), p(T_1, T_3))}{p(T_1, T_2)} = \Phi\left(\frac{1}{X_1}, \frac{Y_1}{X_1}\right) X_1.$$

This justifies our notation in Section 1.2, where we denoted the discounted payoff as a function of X_1, Y_1:

$$\hat{\Phi}(X_1, Y_1), \tag{2.2}$$

i.e. formally we define

$$\hat{\Phi}(x_1, y_1) := \Phi\left(\frac{1}{x_1}, \frac{y_1}{x_1}\right) x_1. \tag{2.3}$$

According to a T_2-forward measure $\mathbb{Q}_{T_2}$, the discounted price for such a derivative is then

$$\mathbb{E}^{\mathbb{Q}_{T_2}}\left[\frac{\Phi(p(T_1,T_2),p(T_1,T_3))}{p(T_1,T_2)}\right]=\mathbb{E}^{\mathbb{Q}_{T_2}}\left[\hat{\Phi}(X_1,Y_1)\right].$$

Therefore, the robust price bounds are given by optimization problems (cf. (1.7))

$$\inf_{\mathbb{Q}_{T_2}\in\mathcal{Q}(\mu,\nu)}\mathbb{E}^{\mathbb{Q}_{T_2}}[\hat{\Phi}(X_1,Y_1)],\qquad \sup_{\mathbb{Q}_{T_2}\in\mathcal{Q}(\mu,\nu)}\mathbb{E}^{\mathbb{Q}_{T_2}}[\hat{\Phi}(X_1,Y_1)]. \tag{2.4}$$

On the other hand, the products available for trading in the market suggest the following semi-static hedging strategies, again expressed in discounted terms:

$$\varphi(X_1)+\psi(Y_2)+\Delta(X_1,Y_1)(Y_2-Y_1). \tag{2.5}$$

Here the static parts $\varphi(X_1)$, $\psi(Y_2)$ can be approximated by call options written on $p(T_1,T_2)$ and $p(T_2,T_3)$, respectively, whereas the dynamic part $\Delta(X_1,Y_1)(Y_2-Y_1)$ corresponds to rebalancing the portfolio in T_1 in a self-financing way, between T_2- and T_3-bonds, according to Δ. The quantity in (2.5) then amounts to the value of such a portfolio at time T_2. In fact, we find that it suffices to rebalance at time T_1 only depending on the value of Y_1, that is, it suffices to consider strategies of the form $\Delta(Y_1)$ rather than $\Delta(X_1,Y_1)$, see Theorem 1.2.

2.2. **Robust pricing of caplets: primal and dual optimizers.** In this section we consider vanilla derivatives of the type $\Phi=(p(T_1,T_3)-K)^+$, and comment on their primal and dual optimizers. W.l.o.g. we can consider $K=1$. In discounted terms, this leads to considering the following function in (2.2):

$$\hat{\Phi}(x,y)=(y-x)^+.$$

By Proposition 1.4, the robust pricing bounds in (2.4) can be computed via

$$\inf_{\pi\in\Pi(\mu,\nu)}\int(b(\pi_x)-x)^+\,d\mu(x), \tag{2.6}$$

$$\sup_{\pi\in\Pi(\mu,\nu)}\int(b(\pi_x)-x)^+\,d\mu(x), \tag{2.7}$$

respectively. In Section 3.4 we discuss in detail the WOT for cost functions of the form $C(x,p)=\theta(b(p)-x)$, where θ is convex or concave. In particular we will determine the primal as well as dual optimizers and thus obtain the solutions to the problems (2.6) and (2.7):

- *Lower bound.* The optimizer for the lower bound (2.6) is determined by the so called *weak monotone rearrangement* of μ and ν (see Alfonsi et al. (2020) and Backhoff-Veraguas et al. (2020b) for details): Briefly, by using $*$ to indicate a push-forward measure, and $\le_c$ to denote convex order between measures , the family

  $$\{T:\mathbb{R}\to\mathbb{R}, T\text{ is monotone, 1-Lip}, T_*(\mu)\le_c\nu\}$$

  has a unique element $\hat{T}$ for which the Wasserstein-1 distance $\mathcal{W}_1(\mu,\hat{T}_*(\mu))$ is minimized. (2.6) is then minimized if we put $Y_1=\hat{T}(X_1)$ and couple Y_1 and Y_2 by an arbitrary martingale coupling of $T_*(\mu)$ and ν. In particular, in this market model, $X_1=\frac{1}{p(T_1,T_2)}$ and $Y_1=\frac{p(T_1,T_3)}{p(T_1,T_2)}$ are comonotone.

  The corresponding sub-replication strategy is determined in Example 3.13, (3.35). It consists in the static position of being short a call on X_1 (that is, short puts on $p(T_1,T_2)$) and holding a call on Y_2 (that is, holding a call on $p(T_2,T_3)$), and on rearranging the portfolio at time T_1 (between the T_2 and the T_3 bonds) in order to hold one unit of the T_3-bond if Y_1 is above a certain threshold and else investing all in the T_2-bond.

Two measures $\nu',\nu\in\mathcal{P}_1(\mathbb{R}^d)$ are in convex order, denoted by $\nu'\le_c\nu$, if $\int\theta d\nu'\le\int\theta d\nu$ for all convex $\theta\colon\mathbb{R}^d\to\mathbb{R}$.

- *Upper bound.* On the other hand, the upper bound (2.6) is obtained in a market where $Y_1 = Y_2$ a.s. and where X_1, Y_1 are coupled through the *anticomonotone coupling*. In particular,
$$p(T_1, T_3) = p(T_1, T_2)p(T_2, T_3) \quad \text{a.s.}$$
The corresponding superreplication strategy (Example 3.13, (3.36)) consists in a static position of holding a put on X_1 (that is, long calls on $p(T_1, T_2)$) and being short a put on Y_2 (that is, short a put on $p(T_2, T_3)$), and on rearranging the portfolio at time T_1 (between the T_2 and the T_3 bonds) in order to be one unit short in the T_3-bond if Y_1 is below a certain threshold and else investing all in the T_2-bond.

3. Weak Optimal Transport problem

For the convenience of the reader who is potentially unfamiliar with the mathematical finance aspect of this work, this section can be read independently of the previous sections. The optimal transport problem for weak transport costs (WOT) was initiated by Gozlan et al. (2017) motivated by applications in geometric inequalities. Their contribution kicked off a vivid research activity by various groups: Gozlan et al. (2018); Alfonsi et al. (2019, 2020); Gozlan and Juillet (2020); Alibert et al. (2019); Shu (2020); Backhoff-Veraguas et al. (2019, 2020b); Backhoff-Veraguas and Pammer (2019, 2020).

In this section we use techniques established in WOT to prove our main results (stated in the Introduction). On the way, we prove some auxiliary results that might be of independent interest.

Specifically, we will show in Theorem 3.7 that, for μ, C continuous, the WOT transport problem can be reduced to the case where the cost function is convex in the second argument. We prove that
$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}} C(x, \pi_x)\, \mu(dx) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}} C^{**}(x, \pi_x)\, \mu(dx), \tag{3.1}$$
where for each x, the function $C^{**}(x, \cdot)$ denotes the convex hull of the function $p \mapsto C(x, p)$. This reduction is relevant since virtually all results obtained in WOT so far assume convexity in the second argument.

In Subsection 3.3 we consider a particular class of WOT problems, where the cost function is barycentric. We draw a connection of barycentric costs to a specific class of OT problems where the second marginal is not fixed, but has to satisfy a convex order constraint. Moreover we derive duality results for semi-continuous costs of barycentric type that closely resemble classical Monge-Kantorovich duality.

3.1. **Notation and setting.** Fix $r \in [1, +\infty)$. We denote by $\mathcal{X}$ and $\mathcal{Y}$ Polish spaces with compatible metrics $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$. We denote by $\mathcal{P}_r(\mathcal{X})$ the set of probability measures on $\mathcal{X}$ which finitely integrate $x \mapsto d_{\mathcal{X}}(x, x_0)^r$ (for some $x_0 \in \mathcal{X}$ and therefore also for any), and we equip it with the Wasserstein distance $\mathcal{W}_r$. For any pair of measures $(\mu, \nu) \in \mathcal{P}_r(\mathcal{X}) \times \mathcal{P}_r(\mathcal{X})$ their distance is given by
$$\mathcal{W}_r(\mu, \nu)^r = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d_{\mathcal{X}}(x, x')^r\, \pi(dx, dx').$$
The set of continuous functions which are absolutely bounded by a multiple of $1 + d_{\mathcal{X}}(x, x_0)^r$ is denoted by $\Phi_r(\mathcal{X})$. For products $\mathcal{X} \times \mathcal{Y}$ of Polish spaces, we will consider the metric $d_r((x, x'), (y, y'))^r = d_{\mathcal{X}}(x, x')^r + d_{\mathcal{Y}}(y, y')^r$ and the corresponding Wasserstein space $\mathcal{P}_r(\mathcal{X} \times \mathcal{Y})$. We will often use the notation $\mu(f)$ to abbreviate $\int_{\mathcal{X}} f(x)\, \mu(dx)$.

3.1.1. *The weak optimal transport problem.* We define the weak transport problem for a measurable cost function $C\colon \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to \overline{\mathbb{R}}$, where we write $\overline{\mathbb{R}}$ for the extended real line $\mathbb{R} \cup \{\pm\infty\}$, as

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}} C(x,\pi_x)\,\mu(dx). \tag{WOT}$$

As the disintegration of a measure is measurable but by no means (weakly) continuous, we observe an inherent asymmetry in (WOT). To be able to obtain results like existence of minimizers and duality, additional assumptions were made about the cost function C in the literature. A minimal assumptions to have both existence and duality is, besides lower semicontinuity and lower boundedness of C, that

$$\forall x \in \mathcal{X}\colon\ p \mapsto C(x,p) \text{ is convex,} \tag{3.2}$$

see Backhoff-Veraguas et al. (2019). The key observation therein is that the spaces $\mathcal{X}$ and $\mathcal{Y}$ play different roles: $\mathcal{X}$ is like the state space of a stochastic process at time 1 and $\mathcal{Y}$ like the state space at time 2.

3.1.2. *The adapted weak topology.* When we regard $\mathcal{X} \times \mathcal{Y}$ as the path space of discrete-time stochastic processes (with two times), it is apparent that the weak topology on $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ is not fit to account for closeness of stochastic processes as it is ignorant to the temporal information structure. For this reason it is of interest to equip $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with a stronger topology, the so-called adapted weak topology. A more detailed discussion of this topology can be found in Backhoff-Veraguas et al. (2020a) and Eder (2019). In our setting, where the interest lies on a single time step, this topology can be introduced in the style of Backhoff-Veraguas et al. (2019) by virtue of the following embedding:

$$J\colon \mathcal{P}(\mathcal{X}\times\mathcal{Y}) \to \mathcal{P}(\mathcal{X}\times\mathcal{P}(\mathcal{Y}))\colon \pi \mapsto (x,\pi_x)_*(\mathrm{proj}^1_*\,\pi), \tag{3.3}$$

where proj^i stands for the projection onto the i-th component. The adapted weak topology on $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ is then defined as the initial topology induced by J. Note that J restricted to $\mathcal{P}_r(\mathcal{X}\times\mathcal{Y})$ maps to $\mathcal{P}_r(\mathcal{X}\times\mathcal{P}_r(\mathcal{Y}))$, where (one possible candidate for) the metric on the latter space is given by $d_{\mathcal{X}\times\mathcal{P}_r(\mathcal{Y})}((x,p),(x',p')) := \sqrt[r]{d_{\mathcal{X}}(x,x')^r + \mathcal{W}_r(p,p')^r}$. We call the initial topology of J restricted to $\mathcal{P}_r(\mathcal{X}\times\mathcal{Y})$ the r-adapted Wasserstein topology which is induced by the metric

$$\mathcal{AW}_r(\pi,\pi') := \mathcal{W}_r(J(\pi),J(\pi')), \tag{3.4}$$

where $\pi,\pi' \in \mathcal{P}_r(\mathcal{X}\times\mathcal{Y})$, and $\mathcal{W}_r$ here is the Wasserstein distance on $\mathcal{P}_r(\mathcal{X}\times\mathcal{P}_r(\mathcal{Y}))$ w.r.t. $d_{\mathcal{X}\times\mathcal{P}_r(\mathcal{Y})}$.

3.1.3. *The relaxation of WOT.* In light of the embedding J it seems natural to introduce a relaxed version of (WOT). As we are interested in processes with prescribed marginals (μ,ν), we start by making sense of what it means for a measure in $\mathcal{P}(\mathcal{X}\times\mathcal{P}(\mathcal{Y}))$ to have ν as its “second marginal". To this end, consider the intensity maps $\hat{I}\colon \mathcal{P}(\mathcal{X}\times\mathcal{P}(\mathcal{Y})) \to \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ and $I\colon \mathcal{P}(\mathcal{P}(\mathcal{Y})) \to \mathcal{P}(\mathcal{Y})$, where for $P \in \mathcal{P}(\mathcal{X}\times\mathcal{P}(\mathcal{Y}))$ and $Q \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ their intensities $\hat{I}(P)$ and $I(Q)$ are given by the unique measures satisfying

$$\hat{I}(P)(f) = \int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})}\int_{\mathcal{Y}} f(x,y)\,p(dy)\,P(dx,dp) \quad \forall f \in C_b(\mathcal{X}\times\mathcal{Y}), \tag{3.5}$$

$$I(Q)(f) = \int_{\mathcal{P}(\mathcal{Y})}\int_{\mathcal{Y}} f(y)\,p(dy)\,Q(dp) \quad \forall f \in C_b(\mathcal{Y}), \tag{3.6}$$

respectively. Note that $\hat{I}$ is the left inverse of J, i.e. $\hat{I}(J(\pi)) = \pi$ for any $\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$. We denote by $\mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{Y})$ the set of probability measures on $\mathcal{X}\times\mathcal{Y}$ concentrated on the graph of a measurable function. Then J maps in fact into $\mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{P}(\mathcal{Y}))$, and $\hat{I}$ restricted to $\mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{P}(\mathcal{Y}))$ is its inverse.

We define $\Lambda(\mu,\nu) \subseteq \mathcal{P}(\mathcal{X}\times\mathcal{P}(\mathcal{Y}))$ by

$$\Lambda(\mu,\nu) := \left\{P \in \mathcal{P}(\mathcal{X}\times\mathcal{P}(\mathcal{Y}))\colon\ \hat{I}(P) \in \Pi(\mu,\nu)\right\}. \tag{3.7}$$

Since $\mathrm{proj}^1_*(\hat{I}(P)) = \mathrm{proj}^1_* P$ and $\mathrm{proj}^2_*(\hat{I}(P)) = I(\mathrm{proj}^2_* P)$, we obtain

$$\Lambda(\mu,\nu) = \left\{P \in \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y}))\colon\ \mathrm{proj}^1_* P = \mu, I(\mathrm{proj}^2_* P) = \nu\right\} = \bigcup_{\substack{\nu' \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))\\ I(\nu')=\nu}} \Pi(\mu,\nu'). \tag{3.8}$$

Clearly, we have that $J(\Pi(\mu,\nu)) \subseteq \Lambda(\mu,\nu)$, but there is even more to it. As long as μ is atomless, we have by the next lemma that $J(\Pi(\mu,\nu))$ is dense in $\Lambda(\mu,\nu)$.

Lemma 3.1. *Let $\mu \in \mathcal{P}(\mathcal{X})$ be continuous and $\nu \in \mathcal{P}(\mathcal{Y})$. Then the set $J(\Pi(\mu,\nu)) = \Lambda(\mu,\nu) \cap \mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{P}(\mathcal{Y}))$ is dense in $\Lambda(\mu,\nu)$.*

Proof. It is well-known that, for a continuous probability measure μ on some Polish space $\mathcal{X}$ and an arbitrary probability measure ν' on some Polish space $\mathcal{Z}$, the set $\Pi(\mu,\nu') \cap \mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{Z})$ is dense in $\Pi(\mu,\nu')$. This means here that, for each $\nu' \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))$, we have denseness of $\Pi(\mu,\nu') \cap \mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{P}(\mathcal{Y}))$ in $\Pi(\mu,\nu')$. Hence, (3.8) yields denseness of

$$\Lambda(\mu,\nu) \cap \mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{P}(\mathcal{Y})) = \bigcup_{\substack{\nu' \in \mathcal{P}(\mathcal{P}(\mathcal{Y}))\\ I(\nu')=\nu}} \Pi(\mu,\nu') \cap \mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{P}(\mathcal{Y}))$$

in $\Lambda(\mu,\nu)$. Finally, since $\hat{I}$ restricted to $\mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{P}(\mathcal{Y}))$ is the inverse of J, we have that $J\colon \Pi(\mu,\nu) \to \mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{P}(\mathcal{Y})) \cap \Lambda(\mu,\nu)$ is a bijection and $J(\Pi(\mu,\nu)) = \mathcal{P}(\mathcal{X} \rightsquigarrow \mathcal{P}(\mathcal{Y})) \cap \Lambda(\mu,\nu)$. □

3.1.4. *On the process interpretation of the relaxation.* A common interpretation of probability measures $\pi \in \Pi(\mu,\nu)$ in optimal transport is to view π as the law of a stochastic process (X,Y) where $X \sim \mu$ and $Y \sim \nu$. Due to the inherent asymmetry of (WOT), this point of view seems even more natural in the framework of weak optimal transport. Then (WOT) is equivalent to

$$\inf_{\substack{(X,Y)\\ X\sim\mu, Y\sim\nu}} \mathbb{E}[C(X, \mathscr{L}(Y|X))].$$

We can relax this problem by considering triplets $(X,Y,\mathcal{F})$ where $X \sim \mu$, $Y \sim \nu$ and $\mathcal{F} \supseteq \sigma(X)$. The corresponding minimization problem reads as

$$\inf_{\substack{(X,Y,\mathcal{F})\\ X\sim\mu, Y\sim\nu, \mathcal{F}\supseteq\sigma(X)}} \mathbb{E}[C(X, \mathscr{L}(Y|\mathcal{F}))]. \tag{3.9}$$

Any such triplet $(X,Y,\mathcal{F})$ (on some abstract probability space) induces an element of $\mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y}))$ via $\mathscr{L}(X, \mathscr{L}(Y|\mathcal{F}))$. In the same way, elements of $\mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y}))$ allow to be interpreted as 1-time step stochastic processes with non-trivial filtration, which is fleshed out in the next lemma.

Lemma 3.2. *Let $(\mu,\nu) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$. Let $\mathbf{X}$ be a 1-time step stochastic process $(X_1 = X, X_2 = Y)$ with filtration $(\mathcal{F}_t)_{t=1}^2$ on an abstract probability space such that $X \sim \mu$ and $Y \sim \nu$. Then there exists $P^{\mathbf{X}} \in \Lambda(\mu,\nu)$ with*

$$\mathbb{E}\left[F(X, \mathscr{L}(Y|\mathcal{F}_1))\right] = \int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} F(x,p)\, P^{\mathbf{X}}(dx,dp) \tag{3.10}$$

for all measurable $F\colon \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to \overline{\mathbb{R}}$ for which the left-hand side in (3.10) *is well-defined.*

On the other hand, any $P \in \Lambda(\mu,\nu)$ induces a process $\mathbf{X}$ with $X \sim \mu$ and $Y \sim \nu$ such that (3.10) *holds for any F where the right-hand side of* (3.10) *is well-defined with $P^{\mathbf{X}} = P$.*

Proof. Define $P^{\mathbf{X}}$ as the law of $(X, \mathscr{L}(Y|\mathcal{F}_1))$, then (3.10) is satisfied. To see that $P^{\mathbf{X}}$ belongs to $\Lambda(\mu,\nu)$, we note that $X \sim \mu = \mathrm{proj}^1_* P^{\mathbf{X}}$. Moreover, for any $f \in C_b(Y)$ we find

$$\int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} \int_{\mathcal{Y}} f(y)\, p(dy)\, P^{\mathbf{X}}(dx,dp) = \mathbb{E}\left[\int_{\mathcal{Y}} f(y)\, \mathscr{L}(Y|\mathcal{F}_1)(dy)\right] = \mathbb{E}\left[f(Y)\right],$$

which proves that the intensity of the second marginal of $P^{\mathbf{X}}$ is ν, hence, $P^{\mathbf{X}} \in \Lambda(\mu,\nu)$ by (3.8).

Now, fix $P \in \Lambda(\mu,\nu)$. Define a filtered probability space by $\mathcal{X} \times \mathcal{P}(\mathcal{Y}) \times \mathcal{Y}$ with σ-algebra $\mathcal{F} = \mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{P}(\mathcal{Y})) \otimes \mathcal{B}(\mathcal{Y})$, probability measure $\mathbb{P}(dx,dp,dy) = P(dx,dp)p(dy)$, and filtration $\mathcal{F}_1 = \sigma((x,p,y) \mapsto (x,p))$, $\mathcal{F}_2 = \mathcal{F}$. We consider the process $(X(x,p,y) = x, Y(x,p,y) = y)$. Clearly, we have $\mathscr{L}(X) = \mu$. To see that $\mathscr{L}(Y) = \nu$, it suffices to show $\mathscr{L}(Y|\mathcal{F}_1)(x,p,y) = p$ $\mathbb{P}$-almost surely, since then $\mathscr{L}(Y) = \mathbb{E}[\mathscr{L}(Y|\mathcal{F}_1)] = \int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} p\, P(dx,dp) = I(\mathrm{proj}^2_* P) = \nu$. By $\mathbb{P}(dx,dp,dy) = P(dx,dp)p(dy)$ we have

$$\mathscr{L}(Y|\mathcal{F}_1)(x,p,y) = \mathbb{P}_{x,p} = p \quad \mathbb{P}\text{-almost surely.}$$

As a consequence we get $Y \sim \nu$ and

$$\begin{aligned}\int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} F(x,p)\, P(dx,dp) &= \int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})\times\mathcal{Y}} F(x,p)\, \mathbb{P}(dx,dp,dy) \\ &= \mathbb{E}\left[F(X,\mathscr{L}(Y|\mathcal{F}_1))\right] = \int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} F(x,p)\, P^{\mathbf{X}}(dx,dp),\end{aligned}$$

for any $F\colon \mathcal{X}\times\mathcal{P}(\mathcal{Y}) \to \overline{\mathbb{R}}$ such that the left-hand side is well-defined. This equation determines the distribution of $P^{\mathbf{X}}$, thus we conclude with $P = P^{\mathbf{X}}$. □

In Lemma 3.2 we found that for each element of $\Lambda(\mu,\nu)$ there is a stochastic process $\mathbf{X}$ such that (3.10) is satisfied and vice versa. In this sense, $\Lambda(\mu,\nu)$ can be seen as the set of stochastic processes $\mathbf{X}$ with not necessarily canonical filtration, which are at time 1 distributed according to μ and at time 2 according to ν. The following result is an immediate consequence.

Corollary 3.3. *The values of* (3.9) *and*

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

coincide.

3.2. **Weak transport duality for non-convex cost.** The main result of this section is Theorem 3.7, that can be seen as a justification for the convexity assumption on the cost C, in the sense of (3.2). First we show, in Proposition 3.5, that the relaxed problem (WOT') coincides with a weak transport problem (WOT) when the cost C is convexified. Then in Theorem 3.7 we obtain that the values of (WOT) and (WOT') coincide (even without convexification of C) for continuous cost when the first marginal μ is atomless. Hence, when one is solely interested in the value of (WOT) for continuous cost, it is of no harm to convexify C and thus work in the framework of classical WOT theory.

For the sake of completeness we include the following (slightly more general) version of the first assertion of (Backhoff-Veraguas et al., 2019, Theorem 3.1), where we weaken the assumption that C is bounded from below to (3.11). So far the weak transport duality was usually stated in the form of (3.12) using an inf-convolution instead of pairs of dual functions, c.f. (Gozlan et al., 2017, Theorem 9.6), (Alibert et al., 2019, Theorem 4.2) and (Backhoff-Veraguas et al., 2019, Theorem 3.1). Another minor contribution of the next theorem is the harmonisation of the weak transport duality to the Kantorovich duality (known from classical optimal transport) in terms of dual pairs instead of utilizing an inf-convolution, see (3.12) and (3.13).

Theorem 3.4 (Duality)**.** *Let* $(\mu,\nu) \in \mathcal{P}_r(\mathcal{X}) \times \mathcal{P}_r(\mathcal{Y})$ *and* $C\colon \mathcal{X}\times\mathcal{P}_r(\mathcal{Y}) \to \mathbb{R}\cup\{+\infty\}$ *be lower semicontinuous such that there is a constant* $K > 0$ *and* $(x_0,y_0) \in \mathcal{X}\times\mathcal{Y}$ *with*

$$C(x,p) \geq -K\left(1 + d_{\mathcal{X}}(x,x_0)^r + \mathcal{W}_r(p,\delta_{y_0})^r\right). \tag{3.11}$$

Then

$$\inf_{P\in\Lambda(\mu,\nu)} \int C(x,p)\, P(dx,dp) = \sup_{\psi\in\Phi_r(\mathcal{Y})} \nu(\psi) + \int_{\mathcal{X}} R_C\psi(x)\,\mu(dx) \tag{3.12}$$

$$= \sup_{\substack{\varphi\in\Phi_r(\mathcal{X}),\,\psi\in\Phi_r(\mathcal{Y})\\ \varphi(x)+p(\psi)\le C(x,p)}} \nu(\psi) + \mu(\psi), \tag{3.13}$$

where $R_C\psi(x) := \inf\{-p(\psi) + C(x,p) \colon p \in \mathcal{P}_r(\mathcal{Y})\}$ *and* $\int_{\mathcal{X}} R_C\psi(x)\,\mu(dx) := -\infty$ *if the integral of* $R_C\psi$ *w.r.t.* μ *is not well-defined.*

Proof. Define the auxiliary cost function $\tilde{C}\colon \mathcal{X}\times\mathcal{P}(\mathcal{Y}) \to [0,+\infty)$ by

$$\tilde{C}(x,p) := C(x,p) + K\left(1 + d_{\mathcal{X}}(x,x_0)^r + \mathcal{W}_r(p,\delta_{y_0})^r\right),$$

where $(x,p)\in\mathcal{X}\times\mathcal{P}(\mathcal{Y})$. For any $P\in\Lambda(\mu,\nu)$ we have

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} 1 + d_{\mathcal{X}}(x,x_0)^r + \mathcal{W}_r(p,\delta_{y_0})^r\, P(dx,dp)\\
&\qquad = 1 + \int_{\mathcal{X}} d_{\mathcal{X}}(x,x_0)^r\,\mu(dx) + \int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})}\int_{\mathcal{Y}} d_{\mathcal{Y}}(y,y_0)^r\, p(dy)\, P(dx,dp)\\
&\qquad = 1 + \int_{\mathcal{X}} d_{\mathcal{X}}(x,x_0)^r\,\mu(dx) + \int_{\mathcal{Y}} d_{\mathcal{Y}}(y,y_0)^r\,\nu(dy),
\end{aligned}$$

and therefore

$$\begin{aligned}
\int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} \tilde{C}(x,p)\,P(dx,dp) &= \int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} C(x,p)\,P(dx,dp)\\
&\qquad + K\left(1 + \int_{\mathcal{X}} d_{\mathcal{X}}(x,x_0)^r\,\mu(dx) + \int_{\mathcal{Y}} d_{\mathcal{Y}}(y,y_0)^r\,\nu(dy)\right).
\end{aligned}$$

For any $\psi\in\Phi_r(\mathcal{Y})$ we write

$$\tilde{\psi}(y) := \psi(y) + K\left(1 + d_{\mathcal{Y}}(y,y_0)^r\right) \in \Phi_r(\mathcal{Y})$$

and we obtain the following relation between the inf-convolutions of C and $\tilde{C}$:

$$\begin{aligned}
R_{\tilde{C}}\psi(x) &= \inf_{p\in\mathcal{P}_r(\mathcal{Y})} -p(\psi) + \tilde{C}(x,p)\\
&= \inf_{p\in\mathcal{P}_r(\mathcal{Y})} -p(\psi) + C(x,p) + K\left(1 + d_{\mathcal{X}}(x,x_0)^r + \int_{\mathcal{Y}} d_{\mathcal{Y}}(y,y_0)^r\,p(dy)\right)\\
&= \inf_{p\in\mathcal{P}_r(\mathcal{Y})} -p(\tilde{\psi}) + C(x,p) + Kd_{\mathcal{X}}(x,x_0)^r\\
&= R_C\tilde{\psi}(x) + Kd_{\mathcal{X}}(x,x_0)^r.
\end{aligned}$$

In particular, the μ-integral of $R_{\tilde{C}}\psi$ is well-defined if and only if the μ-integral of $R_C\tilde{\psi}$ is. Since $\tilde{C}$ is bounded from below and lower semicontinuous, (Backhoff-Veraguas et al., 2019, Theorem 3.1) yields

$$\begin{aligned}
\inf_{P\in\Lambda(\mu,\nu)}\int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})}\tilde{C}(x,p)\,P(dx,dp) &= \sup_{\psi\in\Phi_r(\mathcal{Y})}\nu(\psi)+\int_{\mathcal{X}}R_{\tilde{C}}\psi(x)\,\mu(dx)\\
&= \sup_{\psi\in\Phi_r(\mathcal{Y})}\nu(\psi)+\int_{\mathcal{X}}R_C\tilde{\psi}(x)+Kd_{\mathcal{X}}(x,x_0)^r\,\mu(dx)\\
&= K\left(1+\int_{\mathcal{X}}d_{\mathcal{X}}(x,x_0)^r\,\mu(dx)+\int_{\mathcal{Y}}d_{\mathcal{Y}}(y,y_0)^r\,\nu(dy)\right)\\
&\quad+\sup_{\psi\in\Phi_r(\mathcal{Y})}\nu(\tilde{\psi})+\int_{\mathcal{X}}R_C\tilde{\psi}(x)\,\mu(dx).
\end{aligned}$$

Rearranging the terms and relabeling $\tilde{\psi}$ as ψ yields (3.12).

The next goal is to show (3.13), which resembles more closely the classical Kantorovich duality. By the first part of the proof, we may assume w.l.o.g. that C is nonnegative. It is well-known that for any lower semicontinuous, nonnegative function f (on a metric space), there exists a sequence of nonnegative functions $(f_k)_{k\in\mathbb{N}}$ such that f_k is k-Lipschitz continuous and absolutely bounded by k, and $f_k \nearrow f$.

Let $(C_k)_{k\in\mathbb{N}}$ be such a sequence for C. Monotone convergence and existence of minimizers, see (Backhoff-Veraguas et al., 2019, Theorem 2.9), yield

$$\sup_k\inf_{P\in\Lambda(\mu,\nu)}\int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})}C_k(x,p)\,P(dx,dp)=\inf_{P\in\Lambda(\mu,\nu)}\int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})}C(x,p)\,P(dx,dp).$$

Assume for a moment that (3.13) holds for all $k\in\mathbb{N}$, then

$$\begin{aligned}
\inf_{P\in\Lambda(\mu,\nu)}\int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})}C(x,p)\,P(dx,dp) &= \sup_k \sup_{\substack{\varphi\in\Phi_r(\mathcal{X}),\,\psi\in\Phi_r(\mathcal{Y})\\ \varphi(x)+p(\psi)\le C_k(x,p)}}\nu(\psi)+\mu(\varphi)\\
&\le \sup_{\substack{\varphi\in\Phi_r(\mathcal{X}),\,\psi\in\Phi_r(\mathcal{Y})\\ \varphi(x)+p(\psi)\le C(x,p)}}\nu(\psi)+\mu(\varphi)\\
&\le \inf_{P\in\Lambda(\mu,\nu)}\int C(x,p)\,P(dx,dp).
\end{aligned}$$

Therefore, the inequalities are in fact equalities and (3.13) holds even for C as in the statement of the theorem.

It is sufficient to show (3.13) for k-Lipschitz weak transport costs C, $k\in\mathbb{N}$. When C is k-Lipschitz, its inf-convolution $R_C\psi$ is also k-Lipschitz as the next computation shows:

$$|R_C\psi(x)-R_C\psi(x')|\le\sup_{p\in\mathcal{P}_r(\mathcal{Y})}|C(x,p)-C(x',p)|\le kd_{\mathcal{X}}(x,x').$$

As a consequence we have that $R_C\psi\in\Phi_r(\mathcal{X})$, and

$$\sup_{\psi\in\Phi_r(\mathcal{Y})}\nu(\psi)+\int_{\mathcal{X}}R_C\psi(x)\,\mu(dx)\ge\sup_{\substack{\varphi\in\Phi_r(\mathcal{X}),\,\psi\in\Phi_r(\mathcal{Y})\\ \varphi(x)+p(\psi)\le C(x,p)}}\nu(\psi)+\mu(\varphi).$$

On the other hand, the reverse inequality is easy to show: Let $\varphi\in\Phi_r(\mathcal{X})$ and $\psi\in\Phi_r(\mathcal{Y})$ such that $\varphi(x)+p(\psi)\le C(x,p)$. Then $\varphi(x)\le -p(\psi)+C(x,p)$ for all $(x,p)\in\mathcal{X}\times\mathcal{P}_r(\mathcal{Y})$, hence,

$$\varphi(x)\le\inf_{p\in\mathcal{P}_r(\mathcal{Y})}-p(\psi)+C(x,p)=R_C\psi(x)\quad\forall x\in\mathcal{X}.$$

The supremum in the duality of (Backhoff-Veraguas et al., 2019, Theorem 3.1) is taken over all functions $\psi\in\Phi_r(\mathcal{Y})$ which are bounded from above. Therefore, $R_{\tilde{C}}\psi$ is bounded from below and its μ-integral is well-defined. We circumvent this by defining the μ-integral of $R_C\psi$ appropriately, that is $-\infty$, whenever it is not well-defined.

Therefore we have

$$\nu(\psi) + \mu(\varphi) \le \nu(\psi) + \int_{\mathcal{X}} R_C\psi(x)\,\mu(dx),$$

which readily shows the reverse inequality and completes the proof. □

Proposition 3.5. *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}_r(\mathcal{Y})$. Assume that $C\colon \mathcal{X} \times \mathcal{P}_r(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ is measurable and bounded from below as in* (3.11)*. For each $x \in \mathcal{X}$, we denote by C_x^{**} the lower semicontinuous convex envelope of $C(x,\cdot)$, and define $\tilde{C}(x,p) := \inf_{Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y})), I(Q)=p} \int C(x,q)\,Q(dq)$. Then*

$$\inf_{P \in \Lambda(\mu,\nu)} \int C(x,p)\,P(dx,dp) = \inf_{P \in \Lambda(\mu,\nu)} \int \tilde{C}(x,p)\,P(dx,dp) \tag{3.14}$$

$$= \inf_{\pi \in \Pi(\mu,\nu)} \int \tilde{C}(x,\pi_x)\,\mu(dx) \tag{3.15}$$

$$\ge \inf_{P \in \Lambda(\mu,\nu)} \int C_x^{**}(p)\,P(dx,dp) \tag{3.16}$$

$$= \inf_{\pi \in \Pi(\mu,\nu)} \int C_x^{**}(\pi_x)\,\mu(dx). \tag{3.17}$$

Moreover, if $C(x,\cdot)$ is lower semicontinuous for all $x \in \mathcal{X}$, there is equality.

Proof. Let $P \in \Lambda(\mu,\nu)$. For disintegrations $(P_x)_{x\in\mathcal{X}}$ and $(\pi_x)_{x\in\mathcal{X}}$ of P and π w.r.t. μ, respectively, we have by definition of the intensity maps, c.f. (3.5) and (3.6), the relation

$$I(P_x) = \pi_x \quad \mu\text{-almost surely.}$$

Due to (3.18), we have

$$\int_{\mathcal{X}\times\mathcal{P}_r(\mathcal{Y})} C(x,p)\,P(dx,dp) \ge \int_{\mathcal{X}\times\mathcal{P}_r(\mathcal{Y})} \tilde{C}(x,p)\,P(dx,dp) = \int_{\mathcal{X}}\int_{\mathcal{P}_r(\mathcal{Y})} \tilde{C}(x,p)\,P_x(dp)\,\mu(dx)$$

$$\ge \int_{\mathcal{X}} \tilde{C}(x,\pi_x)\,\mu(dx) \ge \int_{\mathcal{X}} C_x^{**}(x,\pi_x)\,\mu(dx).$$

Evidently, (3.16) is dominated by (3.17), and at the same time we have because of Jensen's inequality

$$\int_{\mathcal{X}\times\mathcal{P}_r(\mathcal{Y})} C_x^{**}(p)\,P(dx,dp) = \int_{\mathcal{X}}\int_{\mathcal{P}_r(\mathcal{Y})} C_x^{**}(p)\,P_x(dp)\,\mu(dx) \ge \int_{\mathcal{X}} C_x^{**}(I(P_x))\,\mu(dx),$$

which proves that (3.16) and (3.17) coincide.

It remains to show that the left-hand side of (3.14) is dominated by (3.15). Pick $\pi \in \Pi(\mu,\nu)$, and note that the map $(x,p) \mapsto \tilde{C}(x,p)$ is lower semianalytic by (Bertsekas and Shreve, 1978, Proposition 7.47). Moreover, $D := \{(x,p,Q) \in \mathcal{X}\times\mathcal{P}_r(\mathcal{Y})\times\mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))\colon I(Q) = p\}$ is closed. Therefore, by (Bertsekas and Shreve, 1978, Proposition 7.50) there is, for each $\varepsilon > 0$, an analytically measurable function $(x,p) \mapsto Q^{x,p}$ with

$$\int_{\mathcal{P}_r(\mathcal{Y})} C(x,q)\,Q^{x,p}(dq) \le \tilde{C}(x,q) + \varepsilon.$$

The intensity of $P(dx,dq) := \mu(dx)\,Q^{x,\pi_x}(dq)$ is given by

$$\hat{I}(P)(dx,dy) = \mu(dx)\,I(Q^{x,\pi_x})(dy) = \mu(dx)\,\pi_x(dy) = \pi(dx,dy).$$

Thus $P \in \Lambda(\mu,\nu)$ and

$$\int_{\mathcal{X}\times\mathcal{P}_r(\mathcal{Y})} C(x,p)\,P(dx,dp) = \int_{\mathcal{X}}\int_{\mathcal{P}_r(\mathcal{Y})} C(x,p)\,Q^{x,\pi_x}(dp)\,\mu(dx) \le \int_{\mathcal{X}} \tilde{C}(x,\pi_x)\,\mu(dx) + \varepsilon.$$

We see that (3.15) dominates the left-hand side of (3.14) by recalling that $\varepsilon > 0$ was arbitrary. □

Lemma 3.6. *Under the assumption of Proposition 3.5, we have that, for all* $(x,p) \in \mathcal{X} \times \mathcal{P}_r(\mathcal{Y})$ *and* $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$ *with* $I(Q) = p$,

$$C_x^{**}(p) \leq \tilde{C}(x,p) \leq \int_{\mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))} \tilde{C}(x,q)\, Q(dq). \tag{3.18}$$

Moreover, if $C(x,\cdot)$ *is lower semicontinuous there is equality, that means* $C_x^{**}(p) = \tilde{C}(x,p)$ *for all* $p \in \mathcal{P}_r(\mathcal{Y})$.

Proof. By Fenchel's duality theorem, for all $p \in \mathcal{P}_r(\mathcal{Y})$ we have that

$$C_x^{**}(p) = \sup_{\psi \in \Phi_r(\mathcal{Y})} p(\psi) - \sup_{q \in \mathcal{P}_r(\mathcal{Y})} q(\psi) - C(x,q) = \sup_{\psi \in \Phi_r(\mathcal{Y})} p(\psi) + R_C\psi(x),$$

and $C_x^{**}(p) \leq C(x,p)$. Therefore, we have for any $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$ with $I(Q) = p$ that

$$C_x^{**}(p) \leq \int_{\mathcal{P}_r(\mathcal{Y})} C_x^{**}(q)\, Q(dq) \leq \int_{\mathcal{P}_r(\mathcal{Y})} C(x,q)\, Q(dq).$$

Taking the infimum over all $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$ with $I(Q) = p$ shows the first inequality in (3.18).

To see the second inequality in (3.18), we observe that the map $q \mapsto \tilde{C}(x,q)$ is by (Bertsekas and Shreve, 1978, Proposition 7.47) lower semianalytic, and that $D = \{(p,Q) \in \mathcal{P}_r(\mathcal{Y}) \times \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))\colon I(Q) = p\}$ is closed. Therefore, by (Bertsekas and Shreve, 1978, Proposition 7.50), for each $\varepsilon > 0$ there is an analytically measurable function $q \mapsto Q^q$ with

$$\int_{\mathcal{P}_r(\mathcal{Y})} C(x,q)\, Q^q(dq) \leq \tilde{C}(x,q) + \varepsilon \quad \forall q \in \mathcal{P}_r(\mathcal{Y}).$$

We now compute the intensity of $\hat{Q}(dq) := \int_{\mathcal{P}_r(\mathcal{Y})} Q^{q'}(dq)\, Q(dq')$:

$$I(\hat{Q}) = \int_{\mathcal{P}_r(\mathcal{Y})} I(Q^{q'})\, Q(dq') = \int_{\mathcal{P}_r(\mathcal{Y})} q'\, Q(dq') = I(Q) = p.$$

Hence,

$$\tilde{C}(x,p) \leq \int_{\mathcal{P}_r(\mathcal{Y})} C(x,q)\, \hat{Q}(dq) \leq \int_{\mathcal{P}_r(\mathcal{Y})} \tilde{C}(x,q)\, Q(dq) + \varepsilon,$$

which proves the second inequality in (3.18) as ε is arbitrary.

If $C(x,\cdot)$ is lower semicontinuous, we have by Theorem 3.4

$$\sup_{\psi \in \Phi_r(\mathcal{Y})} p(\psi) + R_C\psi(x) = \inf_{P \in \Lambda(\delta_x, p)} C(x',p)\, P(dx',dp) = \tilde{C}(x,p),$$

whence, $C_x^{**}(p) = \tilde{C}(x,p)$ for all $p \in \mathcal{P}_r(\mathcal{Y})$. □

Theorem 3.7. *Assume that* $\mu \in \mathcal{P}(\mathcal{X})$ *is continuous,* $\nu \in \mathcal{P}_r(\mathcal{Y})$, *and* $C \in \Phi_r(\mathcal{X} \times \mathcal{P}_r(\mathcal{Y}))$. *Then*

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}} C(x,\pi_x)\, \mu(dx) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}} C_x^{**}(\pi_x)\, \mu(dx). \tag{3.19}$$

In particular, the minimization problems (3.14)-(3.17) *yield the same value.*

Proof. By Proposition 3.5 it remains to show that

$$\inf_{P \in \Lambda(\mu,\nu)} \int C(x,p)\, P(dx,dp) \geq \inf_{\pi \in \Pi(\mu,\nu)} \int C(x,\pi_x)\, \mu(dx).$$

As μ is continuous, Lemma 3.1 provides for any $P \in \Lambda(\mu,\nu)$ a sequence $P^k \in \Lambda(\mu,\nu) \cap \mathcal{P}_r(\mathcal{X} \rightsquigarrow \mathcal{P}_r(\mathcal{Y}))$ such that P^k converges to P weakly. Since $(\mu,\nu) \in \mathcal{P}_r(\mathcal{X}) \times \mathcal{P}_r(\mathcal{Y})$ is fixed, this convergence holds even in $\mathcal{W}_r$ (on $\mathcal{P}_r(\mathcal{X} \times \mathcal{P}_r(\mathcal{Y}))$). Let $\pi^k \in \Pi(\mu, \nu)$ with $J(\pi^k) = P^k$, $k \in \mathbb{N}$. Then

$$\begin{aligned}\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}} C(x, \pi_x)\, \mu(dx) &\leq \liminf_{k \to +\infty} \int_{\mathcal{X}} C(x, \pi^k_x)\, \mu(dx) \\ &= \lim_{k \to +\infty} \int_{\mathcal{X} \times \mathcal{P}_r(\mathcal{Y})} C(x, p)\, P^k(dx, dp) = \int_{\mathcal{X} \times \mathcal{P}_r(\mathcal{Y})} C(x, p)\, P(dx, dp).\end{aligned}$$

As P was an arbitrary element of $\Lambda(\mu, \nu)$, this concludes the proof. □

Proof of Proposition 1.4. First note that clearly the robust optimization problem in (1.7) can be reformulated as in (3.9). Then (1.9) follows from Corollary 3.3, Proposition 3.5 and Theorem 3.7. □

Remark 3.8. Recall the definition of $\tilde{C}(x, \cdot)$ from Proposition 3.5. Assume that $C(x, \cdot)$ is concave in the following sense:

$$\int_{\mathcal{P}_r(\mathcal{Y})} C(x, q)\, Q(dq) \leq C(x, I(Q)) \quad \forall Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y})), \tag{3.20}$$

which is satisfied as soon as $C(x, \cdot)$ is concave, lower/upper semicontinuous and sufficiently bounded from above, by Lemma 3.9. It turns out that then $\tilde{C}(x, \cdot)$ is in fact linear. Indeed, let $p \in \mathcal{P}_r(\mathcal{Y})$. The pushforward measure $Q^* := (y \mapsto \delta_y)_* p$ clearly constitutes an element of $\mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$ where $I(Q^*) = p$. Consequentially we deduce

$$\tilde{C}(x, p) = \inf_{\substack{Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y})) \\ I(Q) = p}} \int_{\mathcal{P}_r(\mathcal{Y})} C(x, q)\, Q(dq) \leq \int_{\mathcal{P}_r(\mathcal{Y})} C(x, q)\, Q^*(dq) = \int_{\mathcal{Y}} C(x, \delta_y)\, p(dy). \tag{3.21}$$

On the other hand, due to (3.20) we have

$$\int_{\mathcal{Y}} C(x, \delta_y)\, p(dy) = \int_{\mathcal{P}_r(\mathcal{Y})} \int_{\mathcal{Y}} C(x, \delta_y)\, q(dy)\, Q(dq) \leq \int_{\mathcal{P}_r(\mathcal{Y})} C(x, q)\, Q(dq), \tag{3.22}$$

for all $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$ with $I(Q) = p$. Thus the left-hand side of (3.22) is dominated by $\tilde{C}(x, p)$. By combining this with (3.21), we derive

$$\tilde{C}(x, p) = \int_{\mathcal{Y}} C(x, \delta_y)\, p(dy).$$

Therefore, (3.14) reduces to a classical optimal transport problem with cost $c(x, y) := C(x, \delta_y)$:

$$\inf_{P \in \Lambda(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{P}_r(\mathcal{Y})} C(x, p)\, P(dx, dp) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y)\, \pi(dx, dy).$$

Lemma 3.9. *Let $F \colon \mathcal{P}_r(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ be convex. Assume that one of the following holds:*

(a) F is lower semicontinuous;
(b) F is upper semicontinuous and upper bounded by a multiple of $p \mapsto \mathcal{W}_r(p, p_0)^r$, where $p_0 \in \mathcal{P}_r(\mathcal{Y})$ is fixed.

Then we have

$$F(I(Q)) \leq \int_{\mathcal{P}_r(\mathcal{Y})} F(q)\, Q(dq) \quad \forall Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y})).$$

Proof. Fix $Q \in \mathcal{P}_r(\mathcal{P}_r(\mathcal{Y}))$. If (a) is satisfied, the inequality is a simple consequence of the representation of F via the Fenchel-Moreau theorem.

Now assume that (b) holds. Since $\mathcal{P}_r(\mathcal{Y})$ is Polish, there exists by (Kallenberg, 2002, Lemma 3.22) a map $h\colon (0,1) \to \mathcal{P}_r(\mathcal{Y})$ such that $Q = h_*\lambda$, where λ denotes the Lebesgue measure on $(0,1)$. We define, for $1 \le k \le n \in \mathbb{N}$,

$$Q_k^n(dq) := n \int_{\frac{k-1}{n}}^{\frac{k}{n}} \delta_{h(t)}\, dt.$$

It is immediate that $\frac{1}{n}\sum_{k=1}^n Q_k^n = Q$, and

$$\begin{aligned}
\mathcal{W}_r\left(\frac{1}{n}\sum_{k=1}^n \delta_{I(Q_k^n)}, Q\right)^r &= \mathcal{W}_r\left(\frac{1}{n}\sum_{k=1}^n \delta_{n\int_{\frac{k-1}{n}}^{\frac{k}{n}} h(t)\,dt}, \int_0^1 \delta_{h(t)}\,dt\right)^r \\
&\le \frac{1}{n}\sum_{k=1}^n \mathcal{W}_r\left(\delta_{n\int_{\frac{k-1}{n}}^{\frac{k}{n}} h(t)\,dt}, n\int_{\frac{k-1}{n}}^{\frac{k}{n}} \delta_{h(t)}\,dt\right)^r \\
&\le \sum_{k=1}^n \int_{\frac{k-1}{n}}^{\frac{k}{n}} \mathcal{W}_r\left(n\int_{\frac{k-1}{n}}^{\frac{k}{n}} h(t)\,dt, h(s)\right)^r ds \\
&\le n\sum_{k=1}^n \int_{\frac{k-1}{n}}^{\frac{k}{n}}\int_{\frac{k-1}{n}}^{\frac{k}{n}} \mathcal{W}_r(h(t),h(s))^r\,dt\,ds \\
&\le \sup\left\{\int_{(0,1)^2} \mathcal{W}_r(h(t),h(s))^r \chi(dt,ds)\colon \chi \in \Pi(\lambda,\lambda)\right. \\
&\qquad\qquad \left.\int_{(0,1)^2} |s-t|^r\,\chi(dt,ds) \le \frac{1}{n^r}\right\},
\end{aligned}$$

where we applied multiple times Jensen's inequality (which is possible in this settings thanks to the first part of this lemma). The right-hand side vanishes for $n \to +\infty$ by (Eder, 2019, Lemma 2.7). Due to convexity of F, we have

$$F(I(Q)) = F\left(\frac{1}{n}\sum_{k=1}^n I(Q_k^n)\right) \le \frac{1}{n}\sum_{k=1}^n F(I(Q_k^n)).$$

We conclude by taking the limit superior for $n \to +\infty$, which yields by upper semicontinuity

$$F(I(Q)) \le \limsup_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n F(I(Q_k^n)) \le \int_{\mathcal{P}_r(\mathcal{Y})} F(q)\,Q(dq).$$

□

3.3. **Barycentric cost functions.** A cost function $C\colon \mathcal{X}\times\mathcal{P}_1(\mathbb{R}^d) \to \overline{\mathbb{R}}$ is called barycentric if, for any $x \in \mathcal{X}$, $p, q \in \mathcal{P}_r(\mathbb{R}^d)$, we have the implication

$$b(p) = b(q) \implies C(x,p) = C(x,q), \tag{3.23}$$

where we recall that $b(p)$ denotes the barycenter of p. An equivalent way of introducing a barycentric cost function C is to select any $c\colon \mathcal{X}\times\mathbb{R}^d \to \overline{\mathbb{R}}$ and let

$$C(x,p) = c\,(x, b(p)), \quad (x,p) \in \mathcal{X}\times\mathcal{P}_1(\mathbb{R}^d).$$

Weak optimal transport problems where the cost function C is of barycentric-type have been recently investigated by various authors in different contexts, see Gozlan et al. (2018, 2017); Gozlan and Juillet (2020); Shu (2020); Alfonsi et al. (2020); Daskalakis et al. (2017); Backhoff-Veraguas et al. (2019, 2020b); Backhoff-Veraguas and Pammer (2020). Typically in these papers, $\mathcal{X}$ was also given as $\mathbb{R}^d$ and $C(x,p) = \theta(b(p) - x)$ for some convex $\theta\colon \mathbb{R}^d \to \mathbb{R}$. Motivated by functional inequalities, this particular problem was explored by Gozlan et al. (2017, 2018); Shu (2020); Gozlan and Juillet (2020). On the other hand, in Fathi et al. (2020) it was used to give a new proof to Cafarelli's contraction theorem. In Alfonsi et al. (2020) the authors are mainly motivated by applications in robust mathematical finance. They use the barycentric WOT problem to construct a sampling technique preserving the convex order, which can then be used to approximate martingale optimal transport problems. Finally, in Daskalakis et al. (2017) the barycentric WOT problem appears in the context of revenue maximization and mechanism design in economy. For a clearer connection of this topic to WOT we refer to Backhoff-Veraguas and Pammer (2020).

The next proposition fleshes out the intrinsic connection of (WOT') with barycentric C and the convex order (and thereby convex functions). Recall that we denote by $\leq_c$ the convex order between measures. For the special case when the cost $C(x,p)$ is given by $\theta(b(p)-x)$ the results of the next proposition can be (partially) found in Gozlan et al. (2017) and Backhoff-Veraguas et al. (2019).

Proposition 3.10. *Let $\mu \in \mathcal{P}_r(\mathcal{X})$, $\nu \in \mathcal{P}_r(\mathbb{R}^d)$, and C be measurable and barycentric such $\int C(x,p)\,P(dx,dp)$ is well-defined for any $P \in \Lambda(\mu,\nu)$. Then*

(a) $$\inf_{P\in\Lambda(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{P}_r(\mathbb{R}^d)} C(x,p)\,P(dx,dp) = \inf_{\nu'\leq_c\nu}\inf_{\pi\in\Pi(\mu,\nu')} \int_{\mathcal{X}\times\mathbb{R}^d} C(x,\delta_y)\,\pi(dx,dy).$$

(b) *In addition, if C is lower semicontinuous and satisfies* (3.11)*, then*

$$\begin{aligned}
&\inf_{P\in\Lambda(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{P}_r(\mathbb{R}^d)} C(x,p)\,P(dx,dp)\\
&= \sup\Big\{\nu(\psi)+\mu(R_C\psi)\colon \psi\in\Phi_r(\mathbb{R}^d) \text{ and concave}\Big\}\\
&= \sup\Big\{\mu(\varphi)+\nu(\psi)\colon \varphi\in\Phi_r(\mathcal{X}), \text{concave } \psi\in\Phi_r(\mathbb{R}^d), \varphi(x)+\psi(y)\leq C(x,\delta_y)\ \forall(x,y)\in\mathcal{X}\times\mathbb{R}^d\Big\}\\
&= \sup\Big\{\mu(\varphi)+\nu(\psi)\colon \varphi\in\Phi_r(\mathcal{X}), \psi\in\Phi_r(\mathbb{R}^d), \Delta\colon\mathbb{R}^d\to\mathbb{R}^d,\\
&\qquad\qquad \varphi(x)+\psi(z)+\Delta(y)(z-y)\leq C(x,\delta_y)\ \forall(x,y,z)\in\mathcal{X}\times\mathbb{R}^d\times\mathbb{R}^d\Big\},
\end{aligned}$$

where $R_C\psi$ is as in Theorem 3.4.

Remark 3.11. Similar results hold for the corresponding WOT problem, c.f. (WOT), when additional convexity/concavity is assumed. When $C(x,\cdot)$ permits for μ-almost every x Jensen's inequality in the following form

$$C\left(x,\int_{\mathcal{P}_r(\mathbb{R}^d)} q\,Q(dq)\right)\leq\int_{\mathcal{P}_r(\mathbb{R}^d)} C(x,q)\,Q(dq) \tag{3.24}$$

for all $Q\in\mathcal{P}_r(\mathcal{P}_r(\mathbb{R}^d))$, then $C(x,\cdot)$ coincides μ-almost surely with the function $\tilde{C}(x,\cdot)$ that was defined in Proposition 3.5. So, if either (3.24) is satisfied or μ is continuous, we have, additionally to (a) of Proposition 3.10, that

$$\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathcal{X}} C(x,\pi_x)\,\mu(dx) = \inf_{\nu'\leq_c\nu}\inf_{\pi\in\Pi(\mu,\nu')}\int_{\mathcal{X}} C(x,\delta_y)\,\pi(dx,dy),$$

by (3.15) and (3.19), respectively. For the same reason, we have, additionally to (b) of Proposition 3.10, that

$$\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathcal{X}} C(x,\pi_x)\,\mu(dx) = \sup_{\psi\in\Phi_r(\mathbb{R}^d),\ concave} \nu(\psi)+\mu(R_C\psi).$$

Proof of Proposition 3.10. We first show (a). Let $\pi \in \Pi(\mu, \nu')$ with $\nu' \leq_c \nu$. By Strassen's theorem there is a martingale coupling $\pi^M \in \Pi_M(\nu', \nu)$, where $\Pi_M(\nu', \nu) := \{\pi' \in \Pi(\nu', \nu) \colon b(\pi'_x) = x \ \mu\text{-a.s.}\}$. Define

$$P(dx, dp) := \mu(dx) \int_{\mathbb{R}^d} \delta_{\pi^M_y}(dp)\, \pi_x(dy).$$

Evidently, the first marginal of P is μ. Then the next line of computations establishes $P \in \Lambda(\mu, \nu)$ thanks to (3.8):

$$\mathrm{proj}_2\left(\hat{I}(P)\right) = \int_{\mathcal{X}\times\mathbb{R}^d} \pi^M_y\, \pi(dx, dy) = \int_{\mathbb{R}^d} \pi^M_y\, \nu'(dy) = \mathrm{proj}_2(\pi^M) = \nu.$$

Since C is barycentric, we find

$$\int_{\mathcal{X}\times\mathcal{P}_1(\mathbb{R}^d)} C(x,p)\, P(dx,dp) = \int_{\mathcal{X}\times\mathcal{P}_1(\mathbb{R}^d)} C(x, \pi^M_y)\, \pi(dx,dy) = \int_{\mathcal{X}\times\mathbb{R}^d} C(x,\delta_y)\, \pi(dx,dy). \quad (3.25)$$

On the other hand, any $P \in \Lambda(\mu,\nu)$ induces a measure $\pi \in \mathcal{P}(\mathcal{X}\times\mathbb{R}^d)$ with second marginal $\nu' \leq_c \nu$, given by $\pi := ((x,p) \mapsto (x, b(p)))_* P$, so that

$$\int_{\mathcal{X}\times\mathcal{P}_1(\mathbb{R}^d)} C(x,p)\, P(dx,dp) = \int_{\mathcal{X}\times\mathcal{P}_1(\mathbb{R}^d)} C\left(x, \delta_{b(p)}\right) P(dx,dp)$$
$$= \int_{\mathcal{X}\times\mathbb{R}^d} C(x,\delta_y)\, \pi(dx,dy). \quad (3.26)$$

To see that $\mathrm{proj}_2\, \pi = \nu'$ is in convex order with ν, we pick any convex $\theta \colon \mathbb{R}^d \to \mathbb{R}$ and find that

$$\nu'(\theta) = \int_{\mathcal{X}\times\mathcal{P}_1(\mathbb{R}^d)} \theta(b(p))\, P(dx,dp) \leq \int_{\mathcal{X}\times\mathcal{P}_1(\mathbb{R}^d)} p(\theta)\, P(dx,dp) = \nu(\theta),$$

by Jensen's inequality, which means that $\nu' \leq_c \nu$. Hence, item (a) follows from (3.25) and (3.26).

Now, we prove (b). Let $\psi \in \Phi_r(\mathbb{R}^d)$ and ψ^{**} be the convex envelope of ψ. For any $y \in \mathbb{R}^d$, we have by Jensen's inequality

$$\psi^{**}(y) = \inf_{p\in\mathcal{P}_r(\mathbb{R}^d),\ b(p)=y} p(\psi^{**}) \leq \inf_{p\in\mathcal{P}_r(\mathbb{R}^d),\ b(p)=y} p(\psi).$$

On the other hand, the right-hand side is convex and dominated by ψ, whence the inequality is actually an equality. Therefore,

$$R_C\psi(x) = \inf_{p\in\mathcal{P}_r(\mathbb{R}^d)} -p(\psi) + C(x,p) = \inf_{y\in\mathbb{R}^d} C(x,\delta_y) + \inf_{p\in\mathcal{P}_r(\mathbb{R}^d),\ b(p)=y} -p(\psi) \quad (3.27)$$
$$= \inf_{y\in\mathbb{R}^d} C(x,\delta_y) + (-\psi)^{**}(y) = R_C(-(-\psi)^{**})(x). \quad (3.28)$$

Note that $R_C\psi$ is μ-integrable with $\mu(R_C\psi) \in [-\infty, +\infty)$. As by definition $(-\psi)^{**} \leq -\psi$, we have

$$\nu(\psi) + \mu(R_C\psi) \leq \nu(-(-\psi)^{**}) + \mu(R_C(-(-\psi)^{**})),$$

whenever $\mu(R_C\psi) > -\infty$. Since $-(-\psi)^{**}$ is concave and the case $\mu(R_C\psi) = -\infty$ is here of no interest, we have shown the first equality in item (b).

The second equality can be shown with the same arguments as in the proof of Theorem 3.4. For the third equality, consider any pair $(\varphi,\psi) \in \Phi_r(\mathcal{X}) \times \Phi_r(\mathbb{R}^d)$. On the one hand, if ψ is concave, then there exists a measurable selection $\Delta \colon \mathbb{R}^d \to \mathbb{R}^d$ of the subgradient of $-\psi$. Therefore,

$$\varphi(x) + \psi(z) + \Delta(y)(z-y) \leq \varphi(x) + \psi(y),$$

whence (φ,ψ,Δ) is admissible in the last supremum whenever (φ,ψ) is admissible in the second supremum.

On the other hand, if there is a measurable $\Delta\colon \mathbb{R}^d \to \mathbb{R}^d$ with

$$\varphi(x) + \psi(z) + \Delta(y)(z-y) \le C(x, \delta_y),$$

then we can define a concave function $\tilde{\psi}(z) := \inf\{C(x,\delta_y) - \varphi(x) - \Delta(y)(z-y)\colon (x,y) \in \mathcal{X} \times \mathbb{R}^d\}$. Since $\tilde{\psi} \ge \psi$, we find that (φ, ψ) is admissible for the second supremum and $\mu(\varphi)+\nu(\psi) \le \mu(\varphi)+\nu(\tilde{\psi})$, which concludes the proof. □

Proof of Theorem 1.2 and Theorem 1.5. As already noticed, the robust optimization problem in (1.7) can be reformulated as in (3.9). Then Theorem 1.5 follows directly from Theorem 3.7 and Proposition 3.10, whereas Theorem 1.2 is a consequence of Corollary 3.3 and again Proposition 3.10. □

3.4. **Primal and dual optimizers for convex barycentric costs.** In this section we consider costs of the form

$$C(x,p) = \theta(b(p) - x), \tag{3.29}$$

with $\theta\colon \mathbb{R} \to \mathbb{R}$ convex, and investigate primal and dual optimizers for the problems

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}} \theta(b(\pi_x) - x)\,\mu(dx), \qquad \sup_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}} \theta(b(\pi_x) - x)\,\mu(dx). \tag{3.30}$$

We first consider the maximization problem in (3.30). From Remark 3.8, we have that

$$\sup_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}} \theta(b(\pi_x) - x)\,\mu(dx) = \sup_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}} \theta(y-x)\,\pi(dx,dy),$$

that is, the problem reduces to a classical transport problem. A minimizer here is the anticomonotone coupling between μ and ν.

Let us now consider the minimization problem in (3.30). If $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$, then Proposition 3.10 gives

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}} \theta(b(\pi_x) - x)\,\mu(dx) = \inf_{\eta\le_c\nu}\inf_{\pi\in\Pi(\mu,\eta)} \int_{\mathbb{R}} \theta(y-x)\,\pi(dx,dy). \tag{3.31}$$

A minimal measure on $\mathbb{R}$ for the right-hand side, denoted by $\nu^* \le_c \nu$, is given by the image of μ under the weak monotone rearrangement T^* of μ and ν, see Backhoff-Veraguas et al. (2020b). Note that $T^* = F_{\nu^*}^{-1} \circ F_\mu$ μ-almost surely.

In what follows we provide a (semi-)explicit representation of the dual optimisers.

Corollary 3.12. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$, and $\theta\colon \mathbb{R}\to\mathbb{R}$ be convex. Then we have*

$$\inf_{\pi\in\Pi(\mu,\nu)} \int \theta\,(b(\pi_x) - x)\,\mu(dx) = \sup_{\psi\in\Phi_1(\mathbb{R}),\,concave} \nu(\psi) + \mu(R_C\psi), \tag{3.32}$$

$$\sup_{\pi\in\Pi(\mu,\nu)} \int \theta(y-x)\,\pi(dx,dy) = \inf_{\psi\in\Phi_1(\mathbb{R}),\,convex} \nu(\psi) + \mu(\overline{R_c}\psi). \tag{3.33}$$

If there exist $f \in L^1(\mu)$, $g \in L^1(\nu)$ such that $\theta(y-x) \le f(x) + g(y)$, then (3.32) *is attained by $\underline{\psi}$, see* (3.39)*, where $\underline{T}(x)$ is the weak monotone rearrangement of μ and ν.*

If there exist $\tilde{f} \in L^1(\mu)$, $\tilde{g} \in L^1(\nu)$ such that $\theta(y-x) \ge \tilde{f}(x) + \tilde{g}(y)$, then (3.33) *is attained by $\overline{\psi}$, see* (3.42)*, where $\overline{T}(x) := F_\nu^{-1}(1 - F_\mu(x))$.*

Proof. The first part of the corollary follows from Remark 3.8, Remark 3.11, and Proposition 3.10. The primal optimizer of the upper bound is given by the anticomonotone coupling $\pi^a \in \Pi(\mu,\nu)$. It satisfies

$$\pi_x^a\left([\overline{T}(x+), \overline{T}(x)]\right) = 1 \quad \mu\text{-almost surely},$$

whence $\overline{\psi}(y) + \overline{R}_C\psi(y) = \theta(y-x)$ π^a-almost surely. Moreover, $\overline{\psi}$ is convex on its domain, since $z \mapsto z - \overline{S}(z)$ and $\partial_-\theta(z)$ are nondecreasing. If $\theta(y-x) \geq \tilde{f}(x) + \tilde{g}(y)$ for $\tilde{f} \in L^1(\mu)$ and $\tilde{g} \in L^1(\nu)$, then the integral of $\overline{\psi}$ w.r.t. ν and the integral of $\overline{R}_C\psi$ w.r.t. μ are well-defined in $(-\infty, +\infty]$. Therefore, $\overline{\psi}$ is a dual optimizer, since

$$\int_{\mathbb{R}\times\mathbb{R}} \theta(y-x)\,\pi^a(dx,dy) = \int_{\mathbb{R}\times\mathbb{R}} \overline{\psi}(y) + \overline{R}_C\psi(x)\,\pi^a(dx,dy) = \nu(\overline{\psi}) + \mu(\overline{R}_C\psi).$$

By (Backhoff-Veraguas et al., 2020b, Theorem 3.1), we know that $\pi^* \in \Pi(\mu,\nu)$ given by

$$\pi^*(dx,dy) := \mu(dx)\,\pi^M_{\underline{T}(x)}(dy)$$

minimizes (3.32), where π^M is any martingale coupling with marginals (μ,ν^*), and $\underline{T}$ is nondecreasing and 1-Lipschitz. Due to (Backhoff-Veraguas et al., 2020b, Theorem 1.3) and the accompanying concept of irreducible components, see (Beiglböck and Juillet, 2016, Appendix A), we have that

$$\pi^M_{\underline{T}(x)}\left(\{y \in \mathbb{R} \colon \underline{T}(x) - x = \underline{T}(y) - y\}\right) = 1 \quad \mu\text{-almost surely.} \tag{3.34}$$

Since $\underline{T}$ is nondecreasing and 1-Lipschitz, the map $x \mapsto \underline{T}(x) - x$ is nonincreasing, thus, $\{y \in \mathbb{R} \colon \underline{T}(x) - x = \underline{T}(y) - y\}$ is a closed interval for every $x \in \mathbb{R}$. For every z, z' in the interior of $\underline{T}(\mathbb{R})$ there are minimal $x, x' \in \mathbb{R}$ with $\underline{T}(x) = z$ and $\underline{T}(x') = z'$. Therefore, $\underline{S}(z) = z$, $\underline{S}(z') = x'$, and

$$z < z' \implies z - \underline{S}(z) = \underline{T}(x) - x \geq \underline{T}(x') - x' = z' - \underline{S}(z'),$$

i.e. $z \mapsto z - \underline{S}(z)$ is decreasing, from which we deduce concavity of $\underline{\psi}$. Furthermore, $\underline{\psi}$ restricted to $\{y \in \mathbb{R} \colon \underline{T}(y) - y = \underline{T}(x) - x\}$ is linear, whereby we find by (3.34) that μ-almost surely

$$\int_{\mathbb{R}} \underline{\psi}(y)\,\pi^M_{\underline{T}(x)}(dy) = \underline{\psi}\left(b(\pi^M_{\underline{T}(x)})\right) = \underline{\psi}(\underline{T}(x)),$$

and therefore

$$\underline{\psi}(\underline{T}(x)) + R_C\underline{\psi}(x) = \theta(\underline{T}(x) - x).$$

If $\theta(y-x) \leq f(x) + g(y)$ for $f \in L^1(\mu)$ and $g \in L^1(\nu)$, then again the μ-integral of $R_C\underline{\psi}$ and the ν-integral of $\underline{\psi}$ are well-defined in $[-\infty, +\infty)$. Due to concavity of $\underline{\psi}$, we observe that its ν^*-integral is well-defined in $[-\infty, +\infty)$, and $R_C\underline{\psi}$ is convex (as the infimum over a jointly convex function) yielding $R_C\underline{\psi} \in L^1(\mu)$. Therefore, $\underline{\psi}$ is a dual optimizer, since

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}} \theta(y-x)\,\pi^*(dx,dy) &= \int_{\mathbb{R}} \theta(\underline{T}(x) - x)\,\mu(dx) = \int_{\mathbb{R}} \underline{\psi}(\underline{T}(x)) + R_C\underline{\psi}(x)\,\mu(dx)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \underline{\psi}(y)\,\pi^M_{\underline{T}(x)}(dy) + R_C\underline{\psi}(x)\,\mu(dx) = \nu(\underline{\psi}) + \mu(R_C\underline{\psi}).\end{aligned}$$

□

Example 3.13. In the setting of Corollary 3.12, let θ be given by $y \mapsto y^+$, and assume that μ and ν are equivalent to the Lebesgue measure restricted to some interval I. Then the left-hand derivative of θ satisfies

$$\partial_-\theta(y-x) = \begin{cases} 0 & y \leq x, \\ 1 & \text{else.} \end{cases}$$

We have already seen in the proof of Corollary 3.12 that $y \mapsto y - \underline{S}(y)$ and $y \mapsto y - \overline{S}(y)$ are nonincreasing and nondecreasing, respectively. Therefore, there are uniquely determined points $\underline{a}, \overline{a} \in \overline{\mathbb{R}}$ such that

$$y - \underline{S}(y) > 0 \quad \forall y < \underline{a} \qquad \text{and} \qquad y - \overline{S}(y) > 0 \quad \forall y > \overline{a}.$$

If $\underline{a} \in \mathbb{R}$, then

$$\underline{\psi}(y) = \int_{y_0}^{y} 1_{(-\infty,\underline{a})}(z)\, dz = (\underline{a} - y_0)^+ - (\underline{a} - y)^+,$$
$$R_{\underline{\psi}}(x) = \inf_{y \in \mathbb{R}} (\underline{a} - y)^+ + (y - x)^+ - (\underline{a} - y_0)^+ = (\underline{a} - x)^+ - (\underline{a} - y_0)^+,$$

whereas if $\overline{a} \in \mathbb{R}$, then

$$\overline{\psi}(y) = \int_{y_0}^{y} 1_{(\overline{a},+\infty)}(z)\, dz = (y - \overline{a})^+ - (y_0 - \overline{a})^+,$$
$$\overline{R}_{\overline{\psi}}(x) = \sup_{y \in \mathbb{R}} -(y - \overline{a})^+ + (y_0 - \overline{a})^+ + (y - x)^+ = -(x - \overline{a})^+ + (y_0 - \overline{a})^+.$$

Note that the constants are canceling out. Therefore, removing the constants and differentiating yields the triplets

$$\Big(\underline{\varphi}(x_1), \underline{\psi}(y_2), \underline{\Delta}(y_1)\Big) := \Big((\underline{a} - x_1)^+, -(\underline{a} - y_2)^+, -\partial_-\theta(\underline{a} - y_1)\Big), \tag{3.35}$$
$$\Big(\overline{\varphi}(x_1), \overline{\psi}(y_2), \overline{\Delta}(y_1)\Big) := \Big(- (x_1 - \overline{a})^+, (y_2 - \overline{a})^+, \partial_-\theta(y_1 - \overline{a})\Big), \tag{3.36}$$

for $(x_1, y_1, y_2) \in \mathbb{R}^3$, which are optimizers of

$$\sup \{\mu(\varphi) + \nu(\psi) \colon \varphi(x_1) + \psi(y_2) + \Delta(y_1)(y_2 - y_1) \le (y_1 - x_1)^+ \text{ for all } x_1, y_1, y_2\},$$
$$\inf \{\mu(\varphi) + \nu(\psi) \colon \varphi(x_1) + \psi(y_2) + \Delta(y_1)(y_2 - y_1) \ge (y_1 - x_1)^+ \text{ for all } x_1, y_1, y_2\},$$

respectively.

We remark that the construction in Lemma 3.14 is by no means novel, but it is still included for the sake of completeness as the authors are unaware of a fitting reference. Shu constructed in an earlier version of Shu (2020) a dual optimizer to the corresponding weak transport problem under the assumption that T is strictly increasing, and remarks that the assumption is simply to avoid technicalities.

Lemma 3.14. *Let $\theta\colon \mathbb{R} \to \mathbb{R}$ be convex, $T\colon I \to \mathbb{R}$ a map where $I \subseteq \mathbb{R}$, $[a, b] = \overline{co(T(I))}$, and $y_0 \in (a, b)$.*

(a) Assume that T is nondecreasing and define

$$\underline{S}(y) := \inf\{x \colon T(x) \ge y\} \quad y \in (a, b), \tag{3.37}$$
$$\underline{\psi}(y) := \int_{y_0}^{y} \partial_-\theta\Big(z - \underline{S}(z)\Big)\, dz \quad y \in (a, b). \tag{3.38}$$

Then $\underline{\psi}$ is continuous on (a, b) and satisfies for all $y' \in [T(x-), T(x)] \cap [a, b]$ that

$$-\underline{\psi}(y') + \theta(y' - x) = \inf_{y \in [a,b]} -\underline{\psi}(y) + \theta(y - x). \tag{3.39}$$

(b) Assume that T is nonincreasing and define

$$\overline{S}(y) := \sup\{x \colon T(x) \ge y\} \quad y \in (a, b), \tag{3.40}$$
$$\overline{\psi}(y) := \int_{y_0}^{y} \partial_-\theta\Big(z - \overline{S}(z)\Big)\, dz \quad y \in (a, b). \tag{3.41}$$

Then $\overline{\psi}$ is continuous on (a, b) and satisfies for all $y' \in [T(x+), T(x)] \cap [a, b]$ that

$$-\overline{\psi}(y') + \theta(y' - x) = \sup_{y \in [a,b]} \overline{\psi}(y) + \theta(y - x). \tag{3.42}$$

Proof. Since θ is convex, it is locally absolutely continuous, thus

$$\theta(y') - \theta(y) = \int_y^{y'} \partial_- \theta(z)\, dz,$$

where we write $\partial_-\theta$ for the left-hand derivative of θ, which is nondecreasing. For fixed $x \in I$ we introduce

$$\underline{f}(z) := -\underline{\psi}(z) + \theta(z - x) \text{ and } \overline{f}(z) := -\overline{\psi}(z) + \theta(z - x),$$

where $z \in (a, b)$. We have dz-almost surely on (a, b) the derivatives

$$\frac{d}{dz}\underline{f}(z) = -\partial_-(z - \underline{S}(z)) + \partial_-\theta(z - x) \text{ and } \frac{d}{dz}\overline{f}(z) = -\partial_-(z - \overline{S}(z)) + \partial_-(z - x).$$

The maps $\underline{S}$ and $\overline{S}$ are nonincreasing and nondecreasing, respectively. If T is monotone, then for any $z \in (a, b)$ there are $\underline{x}, \overline{x} \in I$ such that

$$\begin{aligned} T \text{ nondecreasing}: \ & x \in (-\infty, \underline{x}] \cap I \implies T(x) < z; \quad \text{and } x \in [\overline{x}, +\infty) \cap I \implies T(x) > z; \\ T \text{ nonincreasing}: \ & x \in (-\infty, \underline{x}] \cap I \implies T(x) > z; \quad \text{and } x \in [\overline{x}, +\infty) \cap I \implies T(x) < z. \end{aligned}$$

Hence, $\underline{S}$ and $\overline{S}$ are real-valued, which also shows that $\underline{\psi}$ and $\overline{\psi}$ are well-defined and continuous on (a, b). From now on, we implicitly assume that T is nondecreasing when talking about $\underline{f}$ and $\underline{S}$, whereas we assume that T is nonincreasing when talking about $\overline{f}$ and $\overline{S}$. We have

$$\begin{aligned} z \in (T(x-), T(x)) &\implies x \le \underline{S}(z) \le \underline{S}(T(x)) \le x \iff x = \underline{S}(z), \\ z \in (T(x+), T(x)) &\implies x \ge \overline{S}(z) \ge \overline{S}(T(x)) \ge x \iff x = \overline{S}(z). \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} \frac{d}{dz}\underline{f}|_{(T(x-),T(x))} = 0 &\implies \underline{f}|_{[T(x-),T(x)]} = \underline{f}(T(x)), \\ \frac{d}{dz}\overline{f}|_{(T(x+),T(x))} = 0 &\implies \overline{f}|_{[T(x+),T(x)]} = \overline{f}(T(x)). \end{aligned}$$

It is now sufficient to show that

$$\frac{d}{dz}\underline{f}|_{(a,T(x)]} \le 0 \text{ and } \frac{d}{dz}\underline{f}|_{[(T(x),b)} \ge 0; \quad \frac{d}{dz}\overline{f}|_{(a,T(x)]} \ge 0 \text{ and } \frac{d}{dz}\overline{f}|_{[T(x),b)} \le 0. \tag{3.43}$$

Due to monotonicity and by definition of $\underline{S}$ and $\overline{S}$, we find:

$$\begin{aligned} z \in (a, T(x)) &\implies \begin{cases} \underline{S}(z) \le \underline{S}(T(x)) \le x \implies & z - \underline{S}(z) \ge z - x, \\ x \le \overline{S}(z) \implies & z - \overline{S}(z) \le z - x, \end{cases} \\ z \in (T(x), b) &\implies \begin{cases} x \le \underline{S}(z) \implies & z - x \ge z - \underline{S}(z), \\ \overline{S}(z) \le \overline{S}(T(x)) \le x \implies & z - x \le z - \overline{S}(z). \end{cases} \end{aligned} \tag{3.44}$$

As the left-hand derivative of a convex function is nondecreasing, (3.43) follows from (3.44). □

References

Acciaio, B., Beiglböck, M., Penkner, F., and Schachermayer, W. (2016). A model-free version of the fundamental theorem of asset pricing and the super-replication theorem. *Mathematical Finance*, 26(2):233–251.

Acciaio, B. and Grbac, Z. (2020). Consistency of traded option prices and absence of arbitrage in the presence of stochastic interest rates. *preprint*.

Alfonsi, A., Corbetta, J., and Jourdain, B. (2019). Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds. *International Journal of Theoretical and Applied Finance*, 22(3).

Alfonsi, A., Corbetta, J., and Jourdain, B. (2020). Sampling of probability measures in the convex order by Wasserstein projection. *Annales de l'Institut Henri Poincaré B, Probability and Statistics*, 56(3):1706–1729.

Alibert, J.-J., Bouchitté, G., and Champion, T. (2019). A new class of costs for optimal transport planning. *European Journal of Applied Mathematics*, 30(6):1229–1263.

Backhoff-Veraguas, J., Bartl, D., Beiglböck, M., and Eder, M. (2020a). Adapted Wasserstein distances and stability in mathematical finance. *Finance and Stochastics*, 24(3):601–632.

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

Backhoff-Veraguas, J., Beiglböck, M., Pammer, G., et al. (2020b). Weak monotone rearrangement on the line. *Electronic Communications in Probability*, 25.

Backhoff-Veraguas, J. D. and Pammer, G. (2019). Stability of martingale optimal transport and weak optimal transport. *arXiv e-prints*, page arXiv:1904.04171.

Backhoff-Veraguas, J. D. and Pammer, G. (2020). Applications of weak transport theory. *arXiv preprint arXiv:2003.05338*.

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

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

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

Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic optimal control*, volume 139 of *Mathematics in Science and Engineering*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London. The discrete time case.

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

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

Burzoni, M., Frittelli, M., and Maggis, M. (2017). Model-free superhedging duality. *The Annals of Applied Probability*, 27(3):1452–1477.

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

Cheridito, P., Kiiski, M., Promel, D. J., and Soner, H. M. (2019). Martingale optimal transport duality.

Cheridito, P., Kupper, M., and Tangpi, L. (2017). Duality formulas for robust pricing and hedging in discrete time. *SIAM J. Financial Math.*, 8(1):738–765.

Cox, A. M. G., Obł ój, J., and Touzi, N. (2019). The Root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach. *Probab. Theory Related Fields*, 173(1-2):211–259.

Daskalakis, C., Deckelbaum, A., and Tzamos, C. (2017). Strong Duality for a Multiple-Good Monopolist. *Econometrica*, 85(3):735–767.

Davis, M. H. and Hobson, D. G. (2007). The range of traded option prices. *Mathematical Finance*, 17(1):1–14.

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

Eder, M. (2019). Compactness in Adapted Weak Topologies. *arXiv e-prints*, page arXiv:1905.00856.

Fathi, M., Gozlan, N., and Prodhomme, M. (2020). A proof of the Caffarelli contraction theorem via entropic regularization. *Calculus of Variations and Partial Differential Equations*, 59(3):96.

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

Gozlan, N. and Juillet, N. (2020). On a mixture of Brenier and Strassen theorems. *Proceedings of the London Mathematical Society*, 120(3):434–463.

Gozlan, N., Roberto, C., Samson, P.-M., Shu, Y., and Tetali, P. (2018). Characterization of a class of weak transport-entropy inequalities on the line. *Ann. Inst. Henri Poincaré Probab. Stat.*, 54(3):1667–1693.

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

Hobson, D. (1998). Robust hedging of the lookback option. *Finance and Stochastics*, 2:329–347.

Hobson, D. (2011). The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, pages 267–318. Springer, Berlin.

Hobson, D. and Neuberger, A. (2012). Robust bounds for forward start options. *Math. Finance*, 22(1):31–56.

Kallenberg, O. (2002). *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition.

Obłój, J. (2004). The Skorokhod embedding problem and its offspring. *Probab. Surv.*, 1:321–390.

Shu, Y. (2020). From Hopf–Lax formula to optimal weak transfer plan. *SIAM Journal on Mathematical Analysis*, 52(3):3052–3072.

STABILITY OF MARTINGALE OPTIMAL TRANSPORT AND WEAK OPTIMAL TRANSPORT

J. BACKHOFF-VERAGUAS AND G. PAMMER

Abstract. Under mild regularity assumptions, the transport problem is stable in the following sense: if a sequence of optimal transport plans $\pi^1, \pi^2, \dots$ converges weakly to a transport plan π, then π is also optimal (between its marginals).

Alfonsi, Corbetta and Jourdain [3] asked whether the same property is true for the martingale transport problem. This question seems particularly pressing since martingale transport is motivated by robust finance where data is naturally noisy. On a technical level, stability in the martingale case appears more intricate than for classical transport since martingale optimal transport plans are not characterized by a 'monotonicity'-property of their supports.

In this paper we give a positive answer and establish stability of the martingale transport problem. As a particular case, this recovers the stability of the left curtain coupling established by Juillet [35]. An important auxiliary tool is an unconventional topology which takes the temporal structure of martingales into account. Our techniques also apply to the the weak transport problem introduced by Gozlan, Roberto, Samson and Tetali.

Keywords: stability, martingale transport, weak transport, causal transport, weak adapted topology, robust finance.

1. Introduction and main results

Let X and Y be Polish spaces and consider a continuous function $c\colon X \times Y \to [0, \infty)$. Given probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, the classical transport problem is

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

where $\Pi(\mu, \nu)$ denotes the set of couplings with X-marginal μ and Y-marginal ν. A classical result in optimal transport asserts that $\pi \in \Pi(\mu, \nu)$ is optimal for (OT) iff its support $\operatorname{supp} \pi$ is c-cyclically monotone [42, 43]. One useful consequence of this characterization of optimality is the stability of (OT) with respect to the marginals μ, ν as well as the cost function c. Indeed, the link between monotonicity and stability becomes apparent once one realizes that the notion of monotonicity is itself stable.

In this article we consider the martingale optimal transport problem from the point of view of monotonicity and stability. In fact, since this problem is an instance of a weak optimal transport problem (where the cost is singular, i.e., also takes the value $+\infty$), we will likewise study the latter class of problems for regular cost functions from this viewpoint.

1.1. Stability of martingale optimal transport.

The martingale optimal transport problem is a variant of (OT) stemming from robust mathematical finance (cf. [32, 13, 41, 23, 20, 17, 11, 7, 34, 35, 38, 19, 33, 24, 30] among many others). In order to define this problem, we take $X = Y = \mathbb{R}$, suppose that μ, ν have finite first moments, and introduce the set $\Pi_M(\mu, \nu)$ of martingale couplings with marginals μ, ν. To be precise, a transport plan π is a martingale coupling iff

$$\int_{\mathbb{R}} y\, \pi_x(dy) = x \quad \mu\text{-a.s.},$$

where $\{\pi_x\}_{x \in \mathbb{R}}$ denotes a regular disintegration of the second coordinate given the first one. By a famous result of Strassen, the set $\Pi_M(\mu, \nu)$ is non-empty iff μ is smaller than ν in convex order. The martingale optimal transport problem[1] is given by

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

under the convention that the infimum over the empty set is $+\infty$.

The main result of the article is the stability of (MOT). This gives a positive answer to the question posed by Alfonsi, Corbetta and Jourdain in [3, Section 5.3] in the case $d = 1$.

Let $r \geq 1$. We denote by $\mathcal{P}_r(\mathbb{R})$ the set of probability measures with finite rth-moments and by $\mathcal{W}_r$ the topology of r-Wasserstein convergence on $\mathcal{P}_r(\mathbb{R})$, cf. [42].

Theorem 1.1 (MOT Stability). *Let $c, c_k\colon \mathbb{R}\times\mathbb{R} \to [0,\infty)$, $k \in \mathbb{N}$, be continuous cost functions such that c_k converges uniformly to c. Let $\{\mu_k\}_{k\in\mathbb{N}}, \{\nu_k\}_{k\in\mathbb{N}}$ be sequences in $\mathcal{P}_1(\mathbb{R})$ converging in $\mathcal{W}_1$ to μ and ν, respectively. For each $k \in \mathbb{N}$ let $\pi^k \in \Pi_M(\mu_k,\nu_k)$ be an optimizer of* (MOT) *with cost c_k between the marginals μ_k and ν_k. If $c(x,y) \leq a(x) + b(y)$ with $a \in L^1(\mu)$, $b \in L^1(\nu)$, and*

$$\limsup_{k\to\infty}\int_{\mathbb{R}\times\mathbb{R}} c_k(x,y)\,\pi^k(dx,dy) < \infty,$$

then any weak accumulation point of $\{\pi^k\}_{k\in\mathbb{N}}$ is an optimizer of (MOT) *for the cost function c. In particular if the latter has a unique optimizer π, then $\pi^k \to \pi$ weakly.*

Corollary 1.2. *Let $c, c_k\colon \mathbb{R}\times\mathbb{R} \to [0,\infty)$, $k \in \mathbb{N}$, be continuous cost functions such that c_k converges uniformly to c. Let $\{\mu_k\}_{k\in\mathbb{N}}, \{\nu_k\}_{k\in\mathbb{N}}$ be sequences in $\mathcal{P}_r(\mathbb{R})$ converging in $\mathcal{W}_r$ to μ and ν, respectively, and μ_k is smaller in convex order than ν_k, $k \in \mathbb{N}$. Suppose that*

$$c(x,y) \leq K\,(1 + |x|^r + |y|^r), \text{ for some } K > 0.$$

Then we have

$$\lim_{k\to\infty}\inf_{\pi\in\Pi_M(\mu_k,\nu_k)}\int_{\mathbb{R}\times\mathbb{R}} c_k(x,y)\,\pi(dx,dy) = \inf_{\pi\in\Pi_M(\mu,\nu)}\int_{\mathbb{R}\times\mathbb{R}} c(x,y)\,\pi(dx,dy).$$

We remark that Juillet has obtained in [35] the stability of the left-curtain coupling and hence stability for martingale transport for specific costs. These results are recovered as particular cases of our main result.

Guo and Obłój in [28] introduce and study the convergence of a computational method for martingale transport where the marginals are discretely approximated and the martingale constraint is allowed to fail with a vanishing error.[2] In independent work Wiesel [44] proved stability of the value function of the martingale optimal transport problem in 1 dimension by estimating the distance between an arbitrary coupling to its projection w.r.t. the adapted Wasserstein distance. For further details on the adapted Wasserstein distance we refer to [5].

1.2. **Stability of optimal weak transport.** Gozlan, Roberto, Samson and Tetali [26] proposed the following non-linear generalization of (OT): Given a cost function $C : X \times \mathcal{P}(Y) \to \mathbb{R}$ the optimal weak transport problem is

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

[1]The multidimensional version of the martingale transport problem is defined analogously, although the mathematical finance application is less clear.

[2]Note that the updated version [29] (listed on arxiv.org on April 8th 2019) shows in Proposition 4.7 that the optimal value of MOT is continuous w.r.t. $(\mu,\nu) \in \mathcal{P}_2(\mathbb{R})\times\mathcal{P}_2(\mathbb{R})$ provided that $\mathcal{P}_2(\mathbb{R})\times\mathcal{P}_2(\mathbb{R})$ is equipped with $\mathcal{W}_2$-convergence and c is assumed to be Lipschitz continuous.

The authors of the present article decided to post this article on arxiv.org concurrently to emphasize the independence of our work. We also note that the main focus of [28]/[29] lies on the numerics of martingale transport. In contrast to Theorem 1.1 the proof of [29, Proposition 4.7] is based on the dual problem rather than stability of the optimal couplings.

Observe that one may consider cost functions of the form

$$C_M(x,p) := \begin{cases} \int_{\mathbb{R}^d} c(x,y)\, p(dy) & \int_{\mathbb{R}^d} y\, p(dy) = x, \\ +\infty & \text{else,} \end{cases} \tag{1.1}$$

and in this way (MOT) is a special case of (WOT).

While the original motivation for (WOT) mainly stems from applications to geometric inequalities (cf. Marton [37, 36] and Talagrand [39, 40]), weak transport problems appear also in a number of further topics, including martingale transport [2, 4, 15, 7, 9], the causal transport problem [8, 1], and stability in mathematical finance [5]. In fact, recently some works have considered non-linear martingale transport problems for cost functionals $C\colon \mathbb{R}\times\mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$, i.e. where the cost can also depend on the kernel as in (WOT), cf. [7, 31, 18]. We also refer to [25, 21] for some unexpected applications of weak transport problems in analysis and probability.

The second main contribution of the article is the stability of optimal weak transport. Throughout the article we fix a compatible metric on Y, and for a real number $r \geq 1$ we denote by $\mathcal{P}_r(Y)$ the set of probability measures which integrate the function $d_Y(y_0,\cdot)^r$ for some $y_0 \in Y$, where d_Y denotes a metric on Y. We endow $\mathcal{P}_r(Y)$ with the the r-Wasserstein topology denoted by $\mathcal{W}_r$, and denote the space of continuous functions on X by $C(X)$.

Theorem 1.3 (WOT Stability)**.** *Let $C, C_k : X\times\mathcal{P}_r(Y) \to [0,\infty)$, $k\in\mathbb{N}$, be continuous cost functions such that C_k converges uniformly to C, and either one of the following holds*

(a) $\{C(x,\cdot)\colon x\in X\} \subseteq C(P_r(Y))$ is an equicontinuous family of convex functions,
(b) $\mu\in\mathcal{P}_r(X)$, and there is a constant $K>0$, $x_0\in X$, $y_0\in Y$ such that

$$C(x,p) \leq K\left(1 + d_X(x,x_0)^r + \int_Y d_Y(y,y_0)^r\, p(dy)\right).$$

Let $\{\mu_k\}_{k\in\mathbb{N}}$, $\{\nu_k\}_{k\in\mathbb{N}}$ be sequences in $\mathcal{P}(X)$ and $\mathcal{P}_r(Y)$, which converge weakly to μ and in $\mathcal{W}_r$ to ν, respectively. For each $k\in\mathbb{N}$ let $\pi^k \in \Pi(\mu_k,\nu_k)$ be an optimizer of (WOT) *with cost function C_k between the marginals μ_k and ν_k. If*

$$\limsup_{k\to\infty} \int_X C_k(x,\pi^k_x)\, \mu_k(dx) < \infty,$$

then any weak accumulation point of $\{\pi^k\}_{k\in\mathbb{N}}$ is an optimizer of (WOT) *for the cost function C. In particular if the latter has a unique optimizer π, then $\pi^k \to \pi$ weakly.*

We now describe the main idea used in the proofs of Theorems 1.1 and 1.3.

1.3. **Monotonicity and the correct topology on the set of couplings.** The article [9] investigates the optimal weak transport problem by essentially enlarging the original state space $X\times Y$ to $X\times\mathcal{P}(Y)$. We briefly review this idea since it points to the right notion of monotonicity which will be useful in proving the above stability results.

First we introduce the embedding map

$$\begin{aligned} J\colon \mathcal{P}(X\times Y) &\to \mathcal{P}(X\times\mathcal{P}(Y)), \\ \pi &\mapsto \mathrm{proj}_X(\pi)(dx)\, \delta_{\pi_x}(dp), \end{aligned} \tag{1.2}$$

in words, if $(X,Y)\sim\pi$ then $(X,\pi_X)\sim J(\pi)$, and its left-inverse, the $X\times Y$-intensity map $\hat{I}$,

$$\begin{aligned} \hat{I}\colon \mathcal{P}(X\times\mathcal{P}(Y)) &\to \mathcal{P}(X\times Y), \\ P &\mapsto \int_{p\in\mathcal{P}(Y)} p(dy)\, P(dx,dp). \end{aligned} \tag{1.3}$$

Despite the fact that J is seldom continuous, it does enjoy a key property: it preserves the relative compactness of a set. As a consequence, one can easily obtain optimizers for the following suitable extension of (WOT) as soon as C is lower semicontinuous:

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

where $\Lambda(\mu, \nu)$ is the set of couplings $P \in \mathcal{P}(X \times \mathcal{P}(Y))$ with $\hat{I}(P) \in \Pi(\mu, \nu)$. When $C(x, \cdot)$ is furthermore convex, the extended problem (WOT') is equivalent to the original one, and in addition, (WOT) can be shown to admit an optimizer by means of the natural projection operator $\hat{I}$ from $P(X \times \mathcal{P}(Y))$ onto $\mathcal{P}(X \times Y)$.

This idea of using an embedding (which preserves relative compactness) into a larger space can be appreciated in the following terms: On the original space $\mathcal{P}(X \times Y)$ we consider the initial topology of J when the target space is given the weak topology. This initial topology has been studied in [6, 5] and given the name *adapted weak topology*. One immediate observation is that the cost functional of (WOT) is lower semicontinuous in this topology if C is likewise lower semicontinuous.

Since under the stated conditions it makes no difference to work with (WOT) or (WOT'), it helps to consider the latter simpler linear problem in order to inform our intuition. By analogy with optimal transport we define

Definition 1.4 (C-monotonicity). *A coupling $\pi \in \Pi(\mu, \nu)$ is C-monotone iff there exists a μ-full set $\Gamma \subseteq X$ such that for any finite number of points $x_1, \ldots, x_N \in \Gamma$ and $q_1, \ldots, q_N \in \mathcal{P}(Y)$ with $\sum_{i=1}^N \pi_{x_i} = \sum_{i=1}^N q_i$, we have*

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

It was shown under mild assumptions in [9] that optimality of π for (WOT) implies C-monotonicity in the sense of Definition 1.4 above. The reverse implication(sufficiency) was shown to be true under the additional assumption that the cost function C is uniformly $\mathcal{W}_1$-Lipschitz in the second argument. In the present article we will generalize this result (and largely simplify the arguments) in Theorem 2.2. Once we are equipped with this necessary and sufficient criterion for optimality, the stability result Theorem 1.3 becomes a consequence of the fact that the notion of C-monotonicity is itself stable.

Although martingale optimal transport is a particular case of optimal weak transport, in this work we treat the two problems separately. The reason for this lies in the fact that our approach to optimal weak transport requires regularity of the cost, which the 'embedded cost' C_M, see (1.1), does not provide any longer. To deal with the singular cost C_M, we refine the notion of C-monotonicity when it comes to martingale couplings and additionally employ techniques which currently only work in dimension one. We define

Definition 1.5 (Martingale C-monotonicity). *A coupling $\Pi_M(\mu, \nu)$ is martingale C-monotone iff there exists a μ-full set $\Gamma \subseteq \mathbb{R}^d$ such that for any finite number of points $x_1, \ldots, x_N \in \Gamma$ and $q_1, \ldots, q_N \in \mathcal{P}_1(\mathbb{R}^d)$ with $\sum_{i=1}^N \pi_{x_i} = \sum_{i=1}^N q_i$ and $\int_{\mathbb{R}^d} y\, q_i(dy) = x_i$, we have*

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

Then the key to proving Theorem 1.1 boils down to two arguments: that martingale C-monotonicity is sufficient for optimality, and that this notion of monitonicity is itself stable.

The idea that a monotonicity viewpoint for martingale optimal transport has to be based on the disintegrations of a coupling, or put otherwise, that optimality cannot be read-off from the support of a martingale coupling, goes to Juillet [35]. This author in fact showed that for the left-curtain coupling (c.f. [14]) the support of an optimizer may contain sub-optimal paths; see Figure 1 for an illustration. This is in stark contrast with classical optimal transport, where for say continuous bounded cost functions c it is known that c-cyclical monotonicity of the support is equivalent to optimality. This failure ultimately underlies the need for our Definition 1.5.

1.4. **Outline.** The article is written so that the sections concerning optimal weak transport and martingale optimal transport are largely independent. Section 2 deals with the stability of optimal weak transport, whereas Section 3 explores the stability of martingale optimal

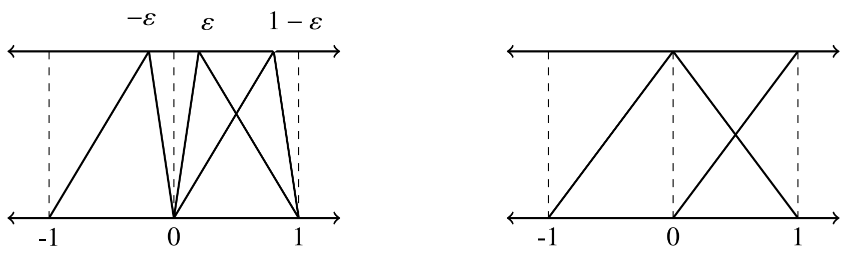

FIGURE 1. Left: Six possibe paths of the left-curtain coupling between a uniform distribution on the interval $[-1, 1]$ and the measure $1/4(\delta_{-1} + \delta_1 + 2\delta_0)$, c.f. [35, Example 2.11]. Right: Three paths contained in the support of this coupling (as can be seen from the left picture after sending $\varepsilon \to 0$). These violate the "left-curtain" condition.

transport. Section 4 studies the relationship between classical optimal transport and optimal weak transport. Finally Section 5 wraps up the article with a number of remarks, followed by an appendix containing auxiliary results.

2. On the weak transport problem

Let us complement Definition 1.4 with a more complete list of monotonicity properties: see Remark 2.4 (a) for a comparision of these definitions.

Definition 2.1.

(1) We call $\Gamma \subseteq X \times \mathcal{P}(Y)$ C-monotone iff for any finite number of points

$$(x_1, p_1), \ldots, (x_N, p_N) \in \Gamma \text{ and } q_1, \ldots, q_N \in \mathcal{P}(Y) \text{ with } \sum_{i=1}^N p_i = \sum_{i=1}^N q_i,$$

we have

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

(2) A probability measure $P \in \mathcal{P}(X \times \mathcal{P}(Y))$, which is concentrated on a C-monotone set, is called C-monotone.

The next theorem greatly extends the sufficiency of C-monotonicity first obtained in [9, Theorem 5.5]. We recall that $\mathcal{P}_r(Y)$ denotes the space of probability measures on Y with finite r-th moment, $r \ge 1$, i.e., $p \in \mathcal{P}_r(Y)$ iff $p \in \mathcal{P}(Y)$ and

$$\int_Y d_Y(y, y_0)^r \, p(dy) < \infty$$

for some $y_0 \in Y$, and thus all $y_0 \in Y$. It is well-known that $\mathcal{P}_r(Y)$ equipped with the topology of the Wasserstein metric $\mathcal{W}_r$ becomes a Polish space, where $\mathcal{W}_r(p, q)$ for $p, q \in \mathcal{P}_r(Y)$ is defined via

$$\mathcal{W}_r(p, q)^r = \inf_{\pi \in \Pi(p,q)} \int_{Y \times Y} d_Y(y_1, y_2)^r \, \pi(dy_1, dy_2).$$

Furthermore, we fix on $X \times \mathcal{P}_r(Y)$ the metric $d((x, p), (x', p'))^r := d_X(x, x')^r + \mathcal{W}_r(p, p')^r$ for $(x, p), (x', p') \in X \times \mathcal{P}_r(Y)$, and denote by $\mathcal{P}_r(X \times \mathcal{P}_r(Y))$ the space of probability measures on $X \times \mathcal{P}_r(Y)$, which finitely integrate $(x, p) \mapsto d((x, p), (x', p'))^r$ for given $(x', p') \in X \times \mathcal{P}_r(Y)$, equiped with the $\mathcal{W}_r$-metric associated with the metric d. Similarly, we fix on $\mathcal{P}_r(Y)$ the Wasserstein metric $\mathcal{W}_r$, and denote by $\mathcal{P}_r(\mathcal{P}_r(Y))$, the space of probability measures on $\mathcal{P}_r(Y)$, which finitely integrate $p \mapsto \mathcal{W}_r(p, p')^r$ for given $p' \in \mathcal{P}_r(Y)$, equiped with the r-th Wasserstein topology. Note that these definitions are independent of the choice of $(x', p') \in X \times \mathcal{P}_r(Y)$ and $p' \in \mathcal{P}_r(Y)$, respectively.

Recall the definition of the set $\Lambda(\mu,\nu)$ given after (WOT'), i.e.,

$$\Lambda(\mu,\nu) := \left\{P \in \mathcal{P}(X \times \mathcal{P}(Y))\colon \hat{I}(P) \in \Pi(\mu,\nu)\right\}. \tag{2.1}$$

Theorem 2.2. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}_r(Y)$, $C\colon X \times \mathcal{P}_r(Y) \to \mathbb{R}$ measurable. Assume either of the following conditions:*

(a) $\{C(x,\cdot)\colon x \in X\} \subseteq C(\mathcal{P}_r(Y))$ *is equicontinuous continuous, i.e.*

$$\theta(\delta) := \sup\left\{|C(x,p) - C(x,q)|\colon x \in X, (p,q) \in \mathcal{P}_r(Y)^2 \text{ s.t.} \mathcal{W}_r(p,q)^r \le \delta\right\},$$

vanishes for $\delta \searrow 0$.

(b) $\mu \in \mathcal{P}_r(X)$*, C is jointly continuous and there is $K \in \mathbb{R}$, $x_0 \in X$, $y_0 \in Y$ s.t. for all $x \in X$:*

$$C(x,p) \le K\left(1 + d_X(x,x_0)^r + \int d_Y(y,y_0)^r\, p(dy)\right).$$

Then $P \in \Lambda(\mu,\nu)$ is optimal for (WOT') *iff P is C-monotone. Similarly, if $p \mapsto C(x,p)$ is convex, then $\pi \in \Pi(\mu,\nu)$ is optimal for* (WOT) *iff π is C-monotone.*

Remark 2.3. Note that under the assumptions of Theorem 2.2 C-monotonicity of $\pi \in \Pi(\mu,\nu)$ yields C-monotonicity of $J(\pi) \in \Lambda(\mu,\nu)$, thus, $J(\pi)$ is optimal for (WOT') and particularly π is optimal for (WOT). We will complete this discussion in Remark 2.4.

At the same time, we see in Section 5 Example 5.1 that convexity of $p \mapsto C(x,p)$ is necessary for optimal couplings being C-monotone. Example 5.2 shows that additional regularity (to lower semicontinuity and boundedness) of the cost C is required for C-monotonicity being a sufficient optimality criterion.

Proof. Let P be C-monotone concentrated on the C-monotone set Γ. Fix $P' \in \Lambda(\mu,\nu)$. We argue as in [10] for classical (linear) optimal transport. Take any iid sequences $(X_n)_{n\in\mathbb{N}}$ of X-valued random variables, and any iid sequences $(Y_n)_{n\in\mathbb{N}}$, $(Z_n)_{n\in\mathbb{N}}$ of $\mathcal{P}(Y)$-valued random variables, on some probability space $(\Omega,\mathbb{P})$, with

$$(X_n,Y_n) \sim P, \quad (X_n,Z_n) \sim P'.$$

In particular by the law of large numbers, we find $\mathbb{P}$-almost surely

$$\int_{X\times\mathcal{P}_r(Y)} C(x,p)P'(dx,dp) - \int_{X\times\mathcal{P}_r(Y)} C(x,p)P(dx,dp)$$
$$= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N C(X_n,Z_n) - C(X_n,Y_n). \tag{2.2}$$

Note that for any function $g \in C(Y)$, which is majorized by $y \mapsto 1 + d_Y(y,y_0)^r$, we have

$$\mathbb{E}[Y_n(g)] = \nu(g),$$

where $p(g)$ for $p \in \mathcal{P}(Y)$ denotes the integral $\int_Y g(y)\,p(dy)$. Then the law of large numbers implies almost surely

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N Y_n(g) = \nu(g) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N Z_n(g).$$

By standard separability arguments, we find $\mathbb{P}$-almost surely

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N Y_n = \nu = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N Z_n \quad \mathbb{P}\text{-a.s.} \tag{2.3}$$

where convergence holds in $\mathcal{W}_r$. Let $\omega \in \Omega$ be in a $\mathbb{P}$-full set s.t.

$$\lim_{N\to\infty}\mathcal{W}_r\left(\frac{1}{N}\sum_{n=1}^N Y_n(\omega), \frac{1}{N}\sum_{n=1}^N Z_n(\omega)\right) = 0,$$

and $(X_n, Y_n)(\omega) \in \Gamma$ for all $n \in \mathbb{N}$. From now on we omit the ω argument. For each $N \in \mathbb{N}$, we denote a $\mathcal{W}_r$-optimal coupling in $\Pi(\frac{1}{N}\sum_{n=1}^N Z_n, \frac{1}{N}\sum_{n=1}^N Y_n)$ by χ^N, i.e.

$$\mathcal{W}_r\left(\frac{1}{N}\sum_{n=1}^N Z_n, \frac{1}{N}\sum_{n=1}^N Y_n\right)^r = \int_{Y\times Y} d_Y(z,y)^r \chi^N(dz,dy).$$

We denote by $\{\chi_z^N\}_{z\in Y}$ a regular disintegration of χ^N given its projection in the first coordinate (marginal). Defining $\chi^N Z_n(dy) := \int_{z\in Y} Z_n(dz)\chi_z^N(dy)$, we find

$$\frac{1}{N}\sum_{n=1}^N \mathcal{W}_r(Z_n, \chi^N Z_n)^r \le \frac{1}{N}\sum_{n=1}^N \int_{Y^2} d_Y(z,y)^r \chi_z^N(dy) Z_n(dz) = \mathcal{W}_r\left(\frac{1}{N}\sum_{n=1}^N Z_n, \frac{1}{N}\sum_{n=1}^N Y_n\right)^r. \tag{2.4}$$

Moreover, $\sum_{n=1}^N \chi^N Z_n = \sum_{n=1}^N Y_n$, so by C-monotonicity

$$\frac{1}{N}\sum_{n=1}^N C(X_n, \chi^N Z_n) - C(X_n, Y_n) \ge 0. \tag{2.5}$$

In the case of (a), we apply Lemma 6.2 and get $\tilde{\theta}\colon \mathbb{R}^+ \to \mathbb{R}^+$, which is continuous, concave, and vanishing at 0, and which we rename θ for simplicity. Hence, by Jensen's inequality,

$$\frac{1}{N}\sum_{n=1}^N \theta\left(\mathcal{W}_r(Z_n,\chi^N Z_n)^r\right) \le \theta\left(\frac{1}{N}\sum_{n=1}^N \mathcal{W}_r(Z_n,\chi^N Z_n)^r\right) \le \theta\left(\mathcal{W}_r\left(\frac{1}{N}\sum_{n=1}^N Z_n, \frac{1}{n}\sum_{n=1}^N Y_n\right)^r\right),$$

which vanishes as $N \to +\infty$. Using C-monotonicity of P and uniform continuity, we obtain $\mathbb{P}$-almost surely

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^N C(X_n, Z_n) - C(X_n, Y_n)\\
&\qquad = \frac{1}{N}\sum_{n=1}^N C(X_n, Z_n) - C(X_n,\chi^N Z_n) + \frac{1}{N}\sum_{n=1}^N C(X_n,\chi^N Z_n) - C(X_n, Y_n)\\
&\qquad\qquad \ge -\frac{1}{N}\sum_{n=1}^N \theta(\mathcal{W}_r(Z_n,\chi^N Z_n))\\
&\qquad\qquad\qquad \ge -\theta\left(\left(\mathcal{W}_r\left(\frac{1}{N}\sum_{n=1}^N Y_n(\omega), \frac{1}{N}\sum_{n=1}^N Z_n(\omega)\right)\right)\right) \to 0.
\end{aligned} \tag{2.6}$$

In the case of (b), we define the random variable P'_N taking values in $X \times \mathcal{P}(Y)$ by

$$P'_N := \frac{1}{N}\sum_{n=1}^N \delta_{X_n,\chi^N Z_n}.$$

By (2.4) we find $\mathbb{P}$-almost surely

$$\mathcal{W}_r(P'_N, P')^r \le \frac{1}{N}\sum_{n=1}^N \mathcal{W}_r(Z_n,\chi^N Z_n) \le \mathcal{W}_r\left(\frac{1}{N}\sum_{i=1}^N Z_n, \frac{1}{N}\sum_{n=1}^N Y_n\right)^r,$$

the right-hand side of which converges a.s. to zero as we have already seen in (2.3).[3] Then, continuity and growth of C yields

$$\int_{X\times\mathcal{P}_r(Y)} C(x,p)\, P'(dx,dp) - \int_{X\times\mathcal{P}_r(Y)} C(x,p)\, P'_N(dx,dp) \to 0. \tag{2.7}$$

[3]Here, the Wasserstein distance on the left-hand side of the equation is taken on the space $X \times \mathcal{P}_r(Y)$ w.r.t. the metric $d((x_1,p_1),(x_2,p_2))^r = d_X(x_1,x_2)^r + \mathcal{W}_r(p_1,p_2)^r$.

Hence, in both cases we have $\mathbb{P}$-almost surely

$$\int_{X\times\mathcal{P}_r(Y)} C(x,p)\,P'(dx,dp) - \int_{X\times\mathcal{P}_r(Y)} C(x,p)\,P(dx,dp)$$
$$\geq \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} C(X_n,Z_n) - C(X_n,Y_n) \geq 0,$$

where we used (2.2),(2.5), (2.6), and (2.7). □

2.1. **Stability of C-Monotonicity.** Recall the embedding of (1.2)

$$J\colon \mathcal{P}(X\times Y) \to \mathcal{P}(X\times\mathcal{P}(Y)),$$
$$\pi \mapsto \mathrm{proj}_X(\pi)(dx)\,\delta_{\pi_x}(dp).$$

The intensity $I(Q)$ of some measure $Q \in \mathcal{P}(\mathcal{P}(Y))$ is uniquely defined as the probability measure $I(Q) \in \mathcal{P}(Y)$ with

$$I(Q)(f) = \int_{\mathcal{P}(Y)}\int_Y f(y)p(dy)\,Q(dp) \quad \forall f\in C_b(Y),$$

that is $I(Q)(dy) = \int_{p\in\mathcal{P}(Y)} p(dy)\,Q(dp)$.

Remark 2.4. In the light of this embedding it appears to be natural to consider C-monotonicity on the enhanced space $X\times\mathcal{P}(Y)$.

(a) $\pi \in \Pi(\mu,\nu)$ is C-monotone iff $J(\pi)$ is C-monotone: On the one hand, if π is C-monotone it is possible to find a measurable set $\tilde\Gamma \subseteq X$ such that $\mu(\tilde\Gamma) = 1$ which defines via
$$\Gamma = \left\{(x,p)\in\tilde\Gamma\times\mathcal{P}(Y)\colon\, p = \pi_x\right\},$$
a C-monotone set. Therefore, equivalently to Definition 2.1, we can demand that there exists a C-monotone set $\Gamma \subseteq X\times\mathcal{P}(Y)$ such that $(x,\pi_x)\in\Gamma$ for μ-almost every $x\in X$.

On the other hand, if $J(\pi)$ is C-montone, then there exists a C-montone set Γ with
$$1 = J(\pi)(\Gamma) = \mu\left(\{x\in X\colon\,(x,\pi_x)\in\Gamma\}\right).$$
Consider the μ-full, analytically measurable set $\{x\in X\colon\,(x,\pi_x)\in\Gamma\}$. As analyitcally measurable sets are universally measurable, it admits a Borel measurable subset $\tilde\Gamma$ with $\mu(\tilde\Gamma) = 1$. Thus, π is C-monotone (on $\tilde\Gamma$) in the sense of Definition 1.4.

(b) If $\Gamma \subseteq X\times\mathcal{P}(Y)$ is C-monotone, and $C\colon X\times\mathcal{P}(Y)\to\mathbb{R}$ is convex in the second argument, i.e., for all $x\in X$, $(p,q)\in\mathcal{P}(Y)^2$, and $\alpha\in[0,1]$
$$C(x,\alpha p + (1-\alpha)q) \leq \alpha C(x,p) + (1-\alpha)C(x,q),$$
then the enlarged set
$$\tilde\Gamma := \left\{\left(x, \frac{1}{k}\sum_{i=1}^{k} p_i\right)\colon\, x\in X,\ (x,p_i)\in\Gamma,\ i=1,\dots,k\in\mathbb{N}\right\}$$
is also C-monotone. Likewise, if $C(x,\cdot)$ is further continuous for all $x\in X$, then
$$\hat\Gamma := \left\{(x,p)\in X\times\mathcal{P}(Y)\colon\, x\in X, p\in\overline{\mathrm{co}(\Gamma_x)}\right\},$$
is C-monotone, where $\overline{\mathrm{co}}$ stands for the closed convex hull and Γ_x denotes the x-fibre of Γ, that is $\{p\in\mathcal{P}(Y)\colon\,(x,p)\in\Gamma\}$.

(c) We observe that the set $\Lambda(\mu,\nu)$ (see (2.1)) can be characterized by a family of continuous functions $\mathcal{F}\subseteq C(X\times\mathcal{P}_r(Y))$: $P\in\Lambda(\mu,\nu)$ if and only if
$$\int_{X\times\mathcal{P}(Y)} f(x)\,P(dx,dp) = \int_X f(x)\,\mu(dx) \quad \forall f\in C_b(X),$$
$$\int_{X\times\mathcal{P}(Y)}\int_Y g(y)\,p(dy)\,P(dx,dp) = \int_Y g(y)\,\nu(dy) \quad \forall g\in C_b(Y).$$

As a further observation we have the equivalence of C-monotonicity as in Definition 2.1 and C-finite optimality under the linear constraints $\mathcal{F}$

$$\mathcal{F} = \Big\{ f \in C_b(X \times \mathcal{P}(Y)) \colon \exists g \in C_b(X),\ h \in C_b(Y) \\ \text{s.t. } f(x,p) \equiv g(x) \text{ or } f(x,p) = \int_Y h(y)\, p(dy) \Big\},$$

which was introduced in [12, Definition 1.2].

Theorem 2.5. *Let $C\colon X \times \mathcal{P}_r(Y) \to [0,\infty]$ be measurable and $P^* \in \mathcal{P}_r(X \times \mathcal{P}_r(Y))$ optimal for* (WOT') *with finite value. Then P^* is C-monotone. In particular, if C satisfies for all $x \in X$ and $Q \in \mathcal{P}(\mathcal{P}(Y))$*

$$C(x, I(Q)) \leq \int_{\mathcal{P}(Y)} C(x,p)\, Q(dp), \tag{2.8}$$

then any optimizer π^* of (WOT) *with finite value is C-monotone.*

Proof. The first assertion is a consequence of Remark 2.4.(c) and [12, Theorem 1.4]. To show the second assertion, let $P \in \Lambda(\mu,\nu)$. Then $I(P_x)\,\mu(dx) \in \Pi(\mu,\nu)$ and by (2.8)

$$\int_{X\times\mathcal{P}(Y)} C(x,p)\, P(dx,dp) \geq \int_X C(x, I(P_x))\, \mu(dx).$$

Hence, $J(\pi^*)$ is optimal for (WOT'). We deduce from the first part of the proof combined with Remark 2.4 (a) C-monotonicity of π. □

The assumption that C is lower bounded is as a matter of fact not necessary to deduce C-monotonicity in the classical optimal weak transport setting, cf. [9, Theorem 5.2]. Note that (2.8) holds when $C(x,\cdot)$ is lower semicontinuous and convex. Indeed, by similar approximation arguments as in Theorem 2.2 we find for any $Q \in \mathcal{P}(\mathcal{P}(Y))$ a sequence of measures $\{p_k\}_{k\in\mathbb{N}} \subseteq \mathcal{P}(Y)$ such that

$$\frac{1}{N}\sum_{i=1}^N p_i \to I(Q) \quad \text{in } \mathcal{W}_r, \qquad \frac{1}{N}\sum_{i=1}^N C(x,p_i) \to \int_{X\times\mathcal{P}(Y)} C(x,p)\, Q(dp).$$

Thus,

$$\begin{aligned}\int_{X\times\mathcal{P}(Y)} C(x,p) Q(dp) &= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N C(x,p_i) \\ &\geq \liminf_{N\to\infty} C\left(x, \frac{1}{N}\sum_{i=1}^N p_i\right) \geq C\left(x, I(Q)\right).\end{aligned}$$

Lemma 2.6. *Let $p_i, m_i \in \mathcal{P}_r(Y)$, $i = 1,\dots,N$, with $\sum_{i=1}^N p_i = \sum_{i=1}^N m_i$, and $\{p_1^k,\dots,p_N^k\}$, $k \in \mathbb{N}$, be sequences in $\mathcal{P}_r(Y)$ with $p_i^k \to p_i$ in $\mathcal{W}_r$. Then there exist approximative sequences $\{m_1^k,\dots,m_N^k\}_{k\in\mathbb{N}}$ of competitors, i.e.,*

$$\sum_{i=1}^N p_i^k = \sum_{i=1}^N m_i^k \text{ and } m_i^k \to m_i \text{ in } \mathcal{W}_r.$$

Proof. Since $\sum_{i=1}^N m_i = \sum_{i=1}^N p_i$, we find sub-probability measures $m_{i,j}$, with the property that $m_{i,j}(A) \leq \min\{p_i(A), m_j(A)\}$ for all A measurable, and

$$m_j = \sum_{i=1}^N m_{i,j}, \qquad p_i = \sum_{j=1}^N m_{i,j}.$$

Denote by $(\chi_z^{k,i})_{z\in Y}$ a regular disintegration of a $\mathcal{W}_r$-optimal transport plan $\chi^{k,i} \in \Pi(p_i, p_i^k)$ w.r.t. to its first marginal p_i. Let $i, j \in \{1,\dots,N\}$ and define

$$m_{i,j}^k(dy) := \int_{z\in Y} \chi_z^{k,i}(dy)\, m_{i,j}(dz), \qquad m_i^k := \sum_{j=1}^N m_{j,i}^k.$$

Since

$$\sum_{j=1}^N \mathcal{W}_r(m_{i,j}^k, m_{i,j})^r \le \sum_{j=1}^N \int_{Y\times Y} d_Y(z,y)^r \,\chi_z^{k,i}(dy)\, m_{i,j}(dz)$$
$$\le \int_{Y\times Y} d_Y(z,y)^r \,\chi_z^{k,i}(dy)\, p_i(dz) = \mathcal{W}_r(p_i^k, p_i)^r,$$

we deduce the convergence of $m_{i,j}^k$ to $m_{i,j}$, and in consequence, the convergence of m_i^k to m_i in $\mathcal{W}_r$. Finally observe that

$$\sum_{i=1}^N m_i^k = \sum_{i=1}^N \sum_{j=1}^N m_{j,i}^k = \sum_{j=1}^N \int_{z\in Y} \chi_z^{k,j}(dy)\, p_j(dz) = \sum_{j=1}^N p_j^k$$

so indeed $\{m_1^k, \dots, m_N^k\}$ are feasible competitors of $\{p_1^k, \dots, p_N^k\}$ such that for $i = 1, \dots, N$, $m_i^k \to m_i$ in $\mathcal{W}_r$. □

Lemma 2.7. *Let $C \in C(X\times \mathcal{P}_r(Y))$, $\varepsilon \ge 0$, and $N \in \mathbb{N}$. Then the set*

$$\Gamma_N^\varepsilon := \Big\{ (x_i, p_i)_{i=1}^N \in (X\times\mathcal{P}_r(Y))^N \Big| \forall m_1,\dots,m_N \in \mathcal{P}_r(Y) \text{ s.t. } \sum_{i=1}^N p_i = \sum_{i=1}^N m_i, \text{ we have } \sum_{i=1}^N C(x_i,p_i) \le \sum_{i=1}^N C(x_i,m_i) + \varepsilon \Big\} \tag{2.9}$$

is a closed subset of $(X\times\mathcal{P}_r(Y))^N$.

Proof. Take any convergent sequence $(x_i^k, p_i^k)_{i=1}^N \in \Gamma_N^\varepsilon$, $k\in\mathbb{N}$, such that

$$x_i^k \to x_i \text{ in } X, \quad p_i^k \to p_i \text{ in } \mathcal{W}_r.$$

Assume that $(x_i, m_i)_{i=1}^N$ is a competitor, i.e., $\sum_{i=1}^N p_i = \sum_{i=1}^N m_i$. Lemma 2.6 provides an approximative sequence of competitors, and by continuity of C we conclude. □

The key ingredient towards stability of (WOT) is the following result concerning stability of the notion of C-monotonicity.

Theorem 2.8. *Let $C, C_k \in C(X\times\mathcal{P}_r(Y))$, $k\in\mathbb{N}$, and C_k converges uniformly to C. If $P, P^k \in \mathcal{P}_r(X\times\mathcal{P}_r(Y))$, $k\in\mathbb{N}$, such that*

(a) for all $k\in\mathbb{N}$ the measure P^k is C_k-monotone,
(b) the sequence $(P^k)_{k\in\mathbb{N}}$ converges to P,

then P is C-monotone. Moreover, if $\pi, \pi^k \in \mathcal{P}_r(X\times Y)$ and C_k is convex in the second argument, $k\in\mathbb{N}$, such that

(a') for all $k\in\mathbb{N}$ the measure π^k is C_k-monotone,
(b') the sequence $(\pi^k)_{k\in\mathbb{N}}$ converges to π,

then π is C-monotone.

Proof. The aim is to construct a C-monotone set Γ on which P is concentrated. So, we write $P^{k,\otimes N}$ and $P^{\otimes N}$ for the N-fold product measure of P^k and P where $N\in\mathbb{N}$. By C_k-monotonicity and uniform convergence we find for any $\varepsilon > 0$ a natural number k_0 such that $P^{k,\otimes N}$, $k \ge k_0$, is concentrated on Γ_N^ε, see (2.9). Lemma 2.7 combined with the Portmanteau theorem yield that $P^{\otimes N}$ is concentrated on Γ_N^ε:

$$1 = \limsup_k P^{k,\otimes N}(\Gamma_N^\varepsilon) \le P^{\otimes N}(\Gamma_N^\varepsilon) = 1.$$

As a consequence, we find that $P^{\otimes N}$ gives full measure to the closed set $\Gamma_N := \Gamma_N^0$. Sets of the form $\bigotimes_{i=1}^N O_i$ where O_i is open in $X\times\mathcal{P}_r(Y)$ form a basis of the product topology on

$(X \times \mathcal{P}_r(Y))^N$. Hence, we can cover the open set Γ_N^c by countably many sets of the form $\bigotimes_{i=1}^N O_i$ where O_i is open in $X \times \mathcal{P}_r(Y)$,

$$\Gamma_N^c = \bigcup_{k\in\mathbb{N}} \bigotimes_{i=1}^N O_{i,k}.$$

In particular, we deduce for any $k \in \mathbb{N}$

$$0 = P^{\otimes N}\left(\bigotimes_{i=1}^N O_{i,k}\right) = \prod_{i=1}^N P(O_{i,k}).$$

We find open sets A_N such that

$$A_N := \bigcup_{\substack{k\in\mathbb{N}, i\in\{1,\dots,N\}\\ P(O_{i,k})=0}} O_{i,k}, \quad P(A_N) = 0,$$

$$\Gamma_N^c \subseteq \bigcup_{i=1}^N (X \times \mathcal{P}_r(Y))^{i-1} \times A_N \times (X \times \mathcal{P}_r(Y))^{N-i}.$$

Since $N \in \mathbb{N}$ was arbitrary we define the closed and C-monotone set

$$\Gamma := \left(\bigcup_{N\in\mathbb{N}} A_N\right)^c, \quad P(\Gamma) = 1, \quad \Gamma^N \subseteq \Gamma_N.$$

With the taken precautions it poses no challenge to verify that P is C-monotone on Γ.

To show the second assertion, we embed $\pi^k \in \mathcal{P}(X \times Y)$ into $\mathcal{P}(X \times \mathcal{P}(Y))$ owing to the map J. Note that $\Lambda(\mu, \nu)$ is a closed subset of $\mathcal{P}_r(X \times \mathcal{P}_r(Y))$. By the same line of argument as in [9, Lemma 2.6] we find relative compactness, thus compactness of $\Lambda(\mu, \nu)$. Therefore, we find an accumulation point $P \in \mathcal{P}(X \times \mathcal{P}(Y))$ of $(J(\pi^k))_{k\in\mathbb{N}}$. Note that

$$\mu(dx)\, I(P_x) =: \pi \in \Pi(\mu, \nu),$$

determines a coupling, which is likewise an accumulation point of $\{\pi^k\}_{k\in\mathbb{N}}$. Since P is concentrated on the C-monotone set Γ, we find for any $x \in X$ such that $P_x(\Gamma_x) = 1$ a sequence of measures $p_i^x \in \Gamma_x \subseteq \mathcal{P}(Y)$, $i \in \mathbb{N}$, with

$$q_n^x := \frac{1}{n}\sum_{i=1}^n p_i^x \to I(P_x) = \pi_x, \quad n \to \infty, \quad \text{in } \mathcal{W}_r.$$

By Remark 2.4 (b) we know that (x, q_n^x) is contained in the C-monotone set Γ. By closure of Γ we conclude $(x, \pi_x) \in \Gamma$ for μ-a.e. x, and C-monotonicity of π. □

From Theorems 2.2 and 2.8 we easily deduce the following corollary, which has Theorem 1.3 in the introduction as a particular case:

Corollary 2.9. *Let $C, C_k \in C(X \times \mathcal{P}_r(Y))$, $k \in \mathbb{N}$, be non-negative cost functions such that C_k converges uniformly to C and one of the following holds:*

(a) $\{C(x, \cdot) : x \in X\}$ is equicontinuous,

(b) $\mu \in \mathcal{P}_r(X)$ and there is a constant $K > 0$, $x_0 \in X$, $y_0 \in Y$ such that for all $(x, p) \in X \times \mathcal{P}(X)$

$$C(x, p) \le K\left(1 + d_X(x, x_0)^r + \int_Y d_Y(y, y_0)^r\, p(dy)\right).$$

Let $\{\mu_k\}_{k\in\mathbb{N}}$ and $\{\nu_k\}_{k\in\mathbb{N}}$ be two sequences of probability measures on $\mathcal{P}(X)$ and $\mathcal{P}_r(Y)$, respectively, where μ_k converges weakly to μ, and ν_k converges in $\mathcal{W}_r$ to ν. Let $P^k \in \Lambda(\mu_k, \nu_k)$ be optimizers of (WOT') *with cost function C_k between the marginals μ_k and ν_k. If*

$$\limsup_k P^k(C_k) < \infty,$$

then any weak accumulation point of $\{P^k\}_{k\in\mathbb{N}}$ is an optimizer of (WOT') *for the cost C.*

If moreover $C_k(x, \cdot)$ and $C(x, \cdot)$ are convex, then an analogous statement holds in the case of (WOT).

3. Stability of martingale optimal transport

In this section we consider the martingale optimal transport problem (MOT), and $X = Y = \mathbb{R}^d$. A generalization of c-cyclical monotonicity under additional linear constraints were suggested by [12, 45], which also encompass (MOT). For (MOT), the set of linear constraints $\mathcal{F}_M \subseteq C(\mathbb{R}^d \times \mathbb{R}^d)$ takes the shape

$$\mathcal{F}_M := \left\{ f \in C(\mathbb{R}^d \times \mathbb{R}^d) \colon f(x,y) = g(x)(y-x) \text{ and } g \in C_b(\mathbb{R}^d) \right\}.$$

Definition 3.1.

(1) *A measures $\alpha' \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is called a $\mathcal{F}_M$-competitor of $\alpha \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ iff their marginals coincide and $\alpha(f) = \alpha'(f)$ for all $f \in \mathcal{F}_M$.*

(2) *We call $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$ $(c, \mathcal{F}_M)$-monotone iff for any probability measure α, finitely supported on Γ, and any competitor α', we have $\alpha(c) \le \alpha'(c)$.*

(3) *A martingale coupling $\pi \in \Pi_M$, which is supported on a $(c, \mathcal{F}_M)$-monotone set, is called $(c, \mathcal{F}_M)$-monotone.*

Remark 3.2. Definition 3.1, Point (2), implies for any $(c, \mathcal{F}_M)$-monotone set Γ that for all sequences $(x_1, p_1), \ldots, (x_N, p_N) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$ with p_i finitely supported on the fibre Γ_{x_i} that

$$\sum_{i=1}^N \int_{\mathbb{R}^d} c(x_i, y)\, p_i(dy) \le \sum_{i=1}^N \int_{\mathbb{R}^d} c(x_i, y)\, q_i(dy), \tag{3.1}$$

where $q_1, \ldots, q_N \in \mathcal{P}_1(\mathbb{R}^d)$, $\sum_{i=1}^N q_i = \sum_{i=1}^N p_i$ and $\int_{\mathbb{R}^d} y\, p_i(dy) = \int_{\mathbb{R}^d} y\, q_i(dy)$.

On the other hand, suppose Γ is a set such that (3.1) holds for all $N \in \mathbb{N}$, and collections $(x_1, p_1), \ldots, (x_N, p_N) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$ with p_i finitely supported on Γ_{x_i}. Given any measure $\alpha \in \mathcal{P}(\Gamma)$ supported on the finite set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, we have optimality of $\{(x_1, \alpha_{x_1}), \ldots, (x_n, \alpha_{x_n})\}$ where $\alpha_{x_i}(dy) = \frac{\alpha(x_i, dy)}{\alpha(x_i \times \mathbb{R}^d)}$ over all competing sequences as in (3.1). Therefore, $\tilde{\alpha}(dx, dy) := \sum_{i=1}^n \delta_{x_i}(dx)\, \alpha_{x_i}(dy)$ defines an optimal coupling between its marginals under all competitors, i.e., for all $\gamma \in \Pi(\mathrm{proj}_1(\tilde{\alpha}), \mathrm{proj}_2(\tilde{\alpha}))$ with $\int_{\mathbb{R}^d} y \gamma_x(dy) = \int_{\mathbb{R}^d} y\, \alpha_x(dy)$ we have

$$\frac{1}{n} \sum_{i=1}^n c(x_i, y_i) = \tilde{\alpha}(c) \le \gamma(c). \tag{3.2}$$

By the duality theorem of linear programming, we find dual optimizers of the linear program given by (3.2), i.e., maximizers of

$$\max_{f(x)+g(y)+\Delta(x)\cdot(x-y) \le c(x,y)} \sum_{i=1}^N f(x_i) + \sum_{i=1}^N \alpha_{x_i}(g) + \sum_{i=1}^N \Delta(x_i) \cdot \int_{\mathbb{R}^d} (x_i - y) \alpha_{x_i}(dy).$$

The complementary slackness condition, which reads here as

$$(f(x_i) + g(y_j) + \Delta(x_i) \cdot (x_i - y_j)) \tilde{\alpha}(x_i, y_j) = 0 \quad \forall 1 \le i, j \le n,$$

yields that optimality of the dual optimizer is independent of the definitive choice of the measure α – as long as $\operatorname{supp} \alpha \subseteq \operatorname{supp} \tilde{\alpha}$. Hence, we deduce $(c, \mathcal{F}_M)$-monotonicity of α, and Γ is $(c, \mathcal{F}_M)$-monotone. In Definition 3.3 we introduce a notion of martingale C-monotonicity for weak transport costs which by above reasoning naturally extends $(c, \mathcal{F}_M)$-monotonicity, see Definition 3.1, to weak transport costs.

We will see in Lemma 3.7 that under given conditions $(c, \mathcal{F}_M)$-monotonicity of a coupling is equivalent to martingale C-monotonicity (cf. Definition 1.5).

By [12], optimizers of (MOT) are concentrated on $(c, \mathcal{F}_M)$-monotone sets. If c is continuous, then the reverse implication was shown in one dimension by Beiglböck and Juillet [14] and Griessler [27], but for arbitrary dimensions $d \in \mathbb{N}$ it remains unanswered.

Let us regard two natural generalizations of (MOT):

$$\inf_{\pi\in\Pi_M(\mu,\nu)}\int_{\mathbb{R}} C(x,\pi_x)\,\mu(dx), \tag{MWOT}$$

$$\inf_{P\in\Lambda_M(\mu,\nu)}\int_{\mathbb{R}\times\mathcal{P}(\mathbb{R})} C(x,p)\,P(dx,dp), \tag{MWOT'}$$

where $\Lambda_M(\mu,\nu)$ is the set of all $P\in\Lambda(\mu,\nu)$ giving full measure to

$$\left\{(x,p)\in\mathbb{R}^d\times\mathcal{P}_1(\mathbb{R}^d)\colon x=\int_{\mathbb{R}^d} y\,p(dy)\right\},$$

i.e., $P\in\Lambda_M(\mu,\nu)$ iff $P\in\Lambda(\mu,\nu)$ and

$$\int_{\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)} f(x,p)\,P(dx,dp)=0 \quad \forall f\in\tilde{\mathcal{F}}_M,$$

where the set of martingale constraints $\tilde{\mathcal{F}}_M$ is given by

$$\tilde{\mathcal{F}}_M := \Big\{ f\in C_b(\mathbb{R}^d\times\mathcal{P}_1(\mathbb{R}^d))\colon \exists g\in C_b(\mathcal{P}(\mathbb{R}^d)), h\in C_b(\mathbb{R}^d)$$
$$\text{s.t. } f(x,p)=g(p)h(x)\int_{\mathbb{R}^d}(x-y)\,p(dy)\Big\}.$$

Definition 3.3.

(1) We call $\Gamma\subseteq\mathbb{R}^d\times\mathcal{P}_1(\mathbb{R}^d)$ martingale C-monotone iff for any $N\in\mathbb{N}$, any collection $(x_1,p_1),\dots,(x_N,p_N)\in\Gamma$, and $q_1,\dots,q_N\in\mathcal{P}_1(\mathbb{R}^d)$ such that $\sum_{i=1}^N p_i=\sum_{i=1}^N q_i$ and $\int_{\mathbb{R}^d} y\,p_i(dy)=\int_{\mathbb{R}^d} y\,q_i(dy)$, we have

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

(2) A probability measure $P\in\mathcal{P}(\mathbb{R}^d\times\mathcal{P}_1(\mathbb{R}^d))$, which is supported on a martingale C-monotone set, is then called martingale C-monotone.

(3) A probability measure $\pi\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$ is called martingale C-monotone if $J(\pi)\in\mathcal{P}(\mathbb{R}^d\times\mathcal{P}_1(\mathbb{R}^d))$ is martingale C-monotone. (This is equivalent to Definition 1.5.)

Again, by [12, Theorem 1.4] we show in the following theorem that martingale C-monotonicity is a necessary optimality criterion.

Theorem 3.4. *Let $C\colon X\times\mathcal{P}_r(\mathbb{R}^d)\to[0,\infty]$ be measurable and $P^*\in\Lambda_M(\mu,\nu)$ optimal for* (MWOT') *with finite value. Then P^* is martingale C-monotone. Moreover, if C additionally satisfies for all $x\in X$ and $Q\in\mathcal{P}_r(\mathcal{P}_r(\mathbb{R}^d))$*

$$C(x,I(Q))\le\int_{\mathcal{P}(\mathbb{R}^d)} C(x,p)\,Q(dp), \tag{3.3}$$

then any optimizer π^* of (MWOT) *with finite value is martingale C-monotone.*

As before (3.3) holds when $C(x,\cdot)$ is lower semicontinuous and convex, and in particular for $C(x,p)=\int_{\mathbb{R}^d} c(x,y)\,p(dy)$ when $c\colon\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ is lower semicontinuous and lower bounded.

Proof. Since (MWOT') is an optimal transport problem under additional linear constraints, the first statement is a consequence of [12, Theorem 1.4]. To show the second assertion, we note that any martingale coupling $\pi\in\Pi_M(\mu,\nu)$ naturally induces an element in $\Lambda_M(\mu,\nu)$ by the embedding J, c.f. Section 1.3. Let $P\in\Lambda_M(\mu,\nu)$, then $I(P_x)\,\mu(dx)\in\Pi_M(\mu,\nu)$ and by (3.3)

$$\int_{\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)} C(x,p)\,P(dx,dp)\ge\int_{\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)} C(x,I(P_x))\,\mu(dx).$$

Hence, $J(\pi^*)$ is optimal for (MWOT'), and we deduce from the first part martingale C-monotonicity of $J(\pi^*)$. Due to similar reasoning as in Remark 2.4 (a) we conclude that π^* is also martingale C-montone. □

From here on we assume that

$$d = 1,$$

but we hope that in the future a similar approach can be developed for higher dimensions.

Lemma 3.5. *Let $N \in \mathbb{N}$ and $p_i \in \mathcal{P}_r(\mathbb{R})$ with competitor $q_i \in \mathcal{P}_r(\mathbb{R})$, $i = 1, \ldots, N$, i.e.,*

$$\sum_{i=1}^N p_i = \sum_{i=1}^N q_i, \qquad \int_{\mathbb{R}} y\, p_i(dy) = \int_{\mathbb{R}} y\, q_i(dy).$$

Suppose there are sequences $\{p_1^k, \ldots, p_N^k\}$, $k \in \mathbb{N}$, of measures in $\mathcal{P}_r(\mathbb{R})$ with $p_i^k \to p_i$ in $\mathcal{W}_r$. Then there exist approximative sequences $\{q_1^k, \ldots, q_N^k\}_{k \in \mathbb{N}}$ of competitors, i.e.,

$$\sum_{i=1}^N p_i^k = \sum_{i=1}^N q_i^k, \ \int_{\mathbb{R}} y\, p_i^k(dy) = \int_{\mathbb{R}} y\, q_i^k(dy), \ \text{and } q_i^k \to q_i \text{ in } \mathcal{W}_r.$$

Since the proof of Lemma 3.5 is slightly demanding, we first give for convenience of the reader a more concrete version of the argument in the simpler setting of $N = 2$:

Proof of Lemma 3.5 for $N = 2$. W.l.o.g. $q_1 \neq p_1$. Applying Lemma 2.6 we find a sequence $\{q_i^k\}_{k \in \mathbb{N}}$ which converges to q_i in $\mathcal{W}_r$ and such that $\sum_{i=1}^N q_i^k = \sum_{i=1}^N p_i^k$. We may further assume $\int_{\mathbb{R}} y\, q_1^k(dy) < \int_{\mathbb{R}} y\, p_1^k(dy)$. We can decompose the measures q_1, q_2, p_1, p_2 into sub-probability measures $m_{i,j}$, $i, j \in \{1, 2\}$ such that

$$p_i = m_{i,1} + m_{i,2}, \quad q_j = m_{1,j} + m_{2,j}.$$

By equality of the mean values of q_1 and p_1, we find that

$$\int_{\mathbb{R}} y\, m_{1,2}(dy) = \int_{\mathbb{R}} y\, m_{2,1}(dy).$$

Thus, we find disjoint, open intervals I_1, I_2 with $\min(m_{1,2}(I_1), m_{2,1}(I_2)) > 0$ and $\sup(I_2) < \inf(I_1)$. Similarly, for each $k \in \mathbb{N}$ we can decompose $q_1^k, q_2^k, p_1^k, p_2^k$ in the same manner and obtain by the construction in Lemma 2.6 that $m_{i,j}^k$ converges to $m_{i,j}$ in $\mathcal{W}_r$. When well-defined denote by $\alpha_k > 0$ the constant such that

$$\int_{\mathbb{R}} y\, q_1^k(dy) + \alpha_k \left(\frac{1}{m_{1,2}^k(I_1)} \int_{I_1} y\, m_{1,2}^k(dy) - \frac{1}{m_{2,1}^k(I_2)} \int_{I_2} y\, m_{2,1}^k(dy) \right) = \int_{\mathbb{R}} y\, p_1^k(dy).$$

By $\mathcal{W}_r$-convergence, we have on the one hand $\liminf_k m_{1,2}^k(I_1) > 0$ and $\liminf_k m_{2,1}^k(I_2) > 0$, and on the other,

$$\lim_{k \to \infty} \int_{\mathbb{R}} y q_1^k(dy) - \int_{\mathbb{R}} y p_1^k(dy) = 0,$$

implying that α_k is well-defined for k sufficiently large, and $\alpha_k \to 0$. Therefore, there is an index $k_0 \in \mathbb{N}$ such that

$$\tilde{q}_1^k = q_1^k + \alpha_k \left(\frac{m_{1,2}^k|_{I_1}}{m_{1,2}^k(I_1)} - \frac{m_{2,1}^k|_{I_2}}{m_{2,1}^k(I_2)} \right), \quad \tilde{q}_2^k = q_2^k - \alpha_k \left(\frac{m_{1,2}^k|_{I_1}}{m_{1,2}^k(I_1)} - \frac{m_{2,1}^k|_{I_2}}{m_{2,1}^k(I_2)} \right),$$

are both probability measures for $k \geq k_0$, and $\{\tilde{q}_1^k, \tilde{q}_2^k\}_{k \geq k_0}$ has the desired properties. □

Proof of Lemma 3.5 General Case. Let $s \in \mathbb{R}$, $F_p \colon \mathbb{R} \to [0, 1]$ the cumulative distribution function of $p \in \mathcal{P}(\mathbb{R})$, and define

$$I_s^1 := \{i \in \{1, \ldots, N\} \colon F_{p_i}(s) = 1\}, \quad I_s^2 := \{i \in \{1, \ldots, N\} \colon F_{q_i}(s) = 1\}.$$

As a preparatory step, we show that

$$j \in \{1, \ldots, N\} \setminus I_s^1 \implies p_j((-\infty, s)) = 0, \tag{3.4}$$

already implies that $I_s^2 = I_s^1$ and

$$j \in \{1,\dots,N\} \setminus I_2^1 \implies q_j((-\infty, s)) = 0. \tag{3.5}$$

This is achieved by observing the barycenters: assume (3.4) and note that

$$\begin{aligned} 0 &= \textstyle\sum_{i\in I_s^1} \int_{\mathbb{R}} y\, p_i(dy) - \int_{\mathbb{R}} y\, q_i(dy) \\ &= \textstyle\int_{\mathbb{R}} y \left(\sum_{i\in I_s^1} p_i - q_i\right)^+ (dy) - \int_{\mathbb{R}} y \left(\sum_{i\in I_s^1} q_i - p_i\right)^+ (dy), \end{aligned} \tag{3.6}$$

where $(\cdot)^+$ of a signed measure denotes its positive part. Moreover, as $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i$ we obtain

$$\sum_{i\in I_s^1} q_i|_{(-\infty,s)} \le \sum_{i=1}^N p_i|_{(-\infty,s)} \le \sum_{i\in I_s^1} p_i,$$

and consequently $(\sum_{i\in I_s^1} p_i - q_i)^+$ is concentrated on $(-\infty, s]$, whereas $(\sum_{i\in I_s^1} q_i - p_i)^+$ is concentrated on $[s, +\infty)$. From (3.6) follows that $\left(\sum_{i\in I_s^1} p_i - q_i\right)^+ = \left(\sum_{i\in I_s^1} q_i - p_i\right)^+ = 0$, thus $\sum_{i\in I_s^1} p_i = \sum_{i\in I_s^1} q_i$ and $I_s^1 \subseteq I_s^2$. Assume that there is $j \in I_s^2 \setminus I_s^1$. Then

$$\sum_{i=1}^N q_i|_{(-\infty,s)} = \sum_{i=1}^N p_i|_{(-\infty,s)} = \sum_{i\in I_s^1} p_i|_{(-\infty,s)} = \sum_{i\in I_s^1} q_i|_{(-\infty,s)}$$

shows that $q_j((-\infty, s)) = 0$. Since $j \in I_s^2$ we have that $q_j((s, +\infty)) = 0$, and consequently $q_j = \delta_s$. The barycenter of p_j coincides with s, but from $j \notin I_s^1$ we deduce $p_j((s, +\infty)) > 0$ and $p_j((-\infty, s)) > 0$, which violates (3.4). Therefore $I_s^1 = I_s^2$, and (3.5) is satisfied.

Clearly, if there exists $s \in \mathbb{R}$ such that either (3.4) or (3.5) holds, then by the preparatory step we get $I_s^1 = I_s^2 =: I_s$ and $\sum_{i\in I_s} p_i = \sum_{i\in I_s} q_i$, and

$$j \in \{1,\dots,N\} \setminus I_s \implies p_j((-\infty, s)) = 0, \quad j \in \{1\dots,N\} \setminus I_s \implies q_j((-\infty, s)) = 0.$$

In this case we can split the problem into two parts: Finding sequences of competitors for the index sets I_s and $\{1,\dots,N\} \setminus I_s$. It is sufficient to show the existence of such a sequence for the sub problem I_s, where we also assume that s is minimal.

Thus, assume without loss of generality that s is minimal and $I_s^1 = I_s^2 = \{1,\dots,N\}$, $N > 1$. Denote the convex hull of the support of q_i by

$$S_i := \mathrm{co}(\mathrm{supp}\, q_i) \quad i = 1,\dots,N.$$

We can also assume without loss of generality that (p_i, q_j) are pairwise different for $i, j \in \{1,\dots,N\}$.

Let $i_{\min}, i_{\max} \in \{1,\dots,N\}$ be such that $\inf S_{i_{\min}}$ is minimal and $\sup S_{i_{\max}}$ is maximal. Under these assumptions, we get that $p_{i_{\min}}, q_{i_{\min}}, p_{i_{\max}}$, and $q_{i_{\max}}$ cannot be concentrated on a single point, thus,

$$\lambda(S_{i_{\min}}) > 0 \text{ and } \lambda(S_{i_{\max}}) > 0, \tag{3.7}$$

where λ is the Lebesgue measure.

Applying now Lemma 2.6 we find for every $i \in \{1,\dots,N\}$ a sequence $\{q_i^k\}_{k\in\mathbb{N}}$ with $\sum_{i=1}^N q_i^k = \sum_{i=1}^N p_i^k$ and which converges to q_i in $\mathcal{W}_r$, in particular,

$$\lim_{k\to\infty} \left| \int_{\mathbb{R}} y p_i^k(dy) - \int_{\mathbb{R}} y q_i^k(dy) \right| = 0. \tag{3.8}$$

Consider the following case. Let $i, j \in \{1,\dots,N\}$ such that one of the following is true: either

$$\lambda(S_j \cap S_i) > 0 \quad \text{or} \quad S_j \subseteq \mathrm{int}(S_i), \tag{3.9}$$

where $\mathrm{int}(S_i)$ denotes the interior of S_i. Then there are open intervals $O_{i,j}^+, O_{i,j}^-, O_{j,i}^+, O_{j,i}^-$ such that

$$\begin{aligned} q_i(O_{i,j}^t) > 0, \quad q_j(O_{j,i}^t) > 0, \quad t \in \{-,+\}, \\ \sup O_{i,j}^+ < \inf O_{j,i}^-, \quad \sup O_{j,i}^+ < \inf O_{i,j}^-. \end{aligned}$$

By weak convergence of q_i^k to q_i, the Portmanteau theorem implies

$$\begin{aligned}\liminf_{k\to\infty} q_i^k(O_{i,j}^+) > 0, \quad \liminf_{k\to\infty} q_i^k(O_{i,j}^-) > 0,\\ \liminf_{k\to\infty} q_j^k(O_{j,i}^+) > 0, \quad \liminf_{k\to\infty} q_j^k(O_{j,i}^-) > 0.\end{aligned} \tag{3.10}$$

In particular, when k is sufficiently large we can by (3.8) and (3.10) replace q_i^k and q_j^k with the probability measures

$$\tilde q_j^k := q_j^k + \alpha_+^k \left(\frac{q_i^k|_{O_{i,j}^-}}{q_i^k(O_{i,j}^-)} - \frac{q_j^k|_{O_{j,i}^+}}{q_j^k(O_{j,i}^+)} \right) + \alpha_-^k \left(\frac{q_i^k|_{O_{i,j}^+}}{q_i^k(O_{i,j}^+)} - \frac{q_j^k|_{O_{j,i}^-}}{q_j^k(O_{j,i}^-)} \right),$$

and $\tilde q_i^k := q_i^k - \tilde q_j^k + q_j^k$, where $\{\alpha_-^k\}_{k\in\mathbb{N}}$ and $\{\alpha_+^k\}_{k\in\mathbb{N}}$ are non-negative sequences converging to zero, and $\tilde q_i^k$ and p_i^k have the same barycenter (i.e. if the barycenter of q_j^k is smaller than that of p_j^k we set $\alpha_+^k > 0$ and $\alpha_-^k = 0$, etc.). Thus, $\{\tilde q_i^k\}_{k\in\mathbb{N}}$ and $\{\tilde q_j^k\}_{k\in\mathbb{N}}$ converge in $\mathcal{W}_r$ to q_i and q_j, respectively.

We will call $i \in \{1, \dots, N\}$ a *pivot*, if for all $\hat\imath \in \{1, \dots, N\}$ with $\sup S_{\hat\imath} \le \inf S_i$, the barycenters of $q_{\hat\imath}^k$ are correct (when k is sufficiently large), that means, there is $k_0 \in \mathbb{N}$ such that

$$k \ge k_0 \implies \int_{\mathbb{R}} y\, q_{\hat\imath}^k(dy) = \int_{\mathbb{R}} y\, p_{\hat\imath}^k(dy). \tag{3.11}$$

For such a pivot i, consider all $j \neq i$ with (3.9). By the previous step in this proof, there exists $k_1 \in \mathbb{N}$ such that we can change q_i^k and q_j^k for $k \ge k_1 \in \mathbb{N}$ (and denote the probability measures $\tilde q_i^k$ and $\tilde q_j^k$ for notational convenience again by q_i^k and q_j^k respectively) such that

$$k \ge k_1 \implies \int_{\mathbb{R}} y\, q_j^k(dy) = \int_{\mathbb{R}} y\, p_j^k(dy). \tag{3.12}$$

There are two possible cases:

(1) (3.9) held true for all $j \neq i$ with $\inf S_i < \sup S_j$, which would imply that (given k is sufficiently large), we have not only corrected the barycenter of q_j^k but also the one of q_i^k, since

$$\int_{\mathbb{R}} y q_i^k(dy) = \sum_{l=1}^N \int_{\mathbb{R}} y p_l^k(dy) - \sum_{j\neq i} \int_{\mathbb{R}} y q_j^k(dy) = \int_{\mathbb{R}} y p_i^k(dy).$$

Hence, we have found the desired sequences.

(2) There are indices $k \in \{1, \dots, N\}$ with $\sup S_i \le \inf S_k \le \sup S_k$. Due to minimality of s, there is an index $l \in \{1, \dots, N\}$ with $\inf S_l < \sup S_i < \sup S_l$. Hence (3.9) holds and we can use $\{q_l^k\}_{k\in\mathbb{N}}$ to fix the barycenters of $\{q_i^k\}_{k\in\mathbb{N}}$ for k sufficiently large. Moreover, l is a pivot: Let $\hat l \in \{1, \dots, N\}$ with $\sup S_{\hat l} \le \inf S_l$, then either $\sup S_{\hat l} \le \inf S_i$ whereby (3.11) holds for $\hat l$, or $\inf S_i < \sup S_{\hat l} \le \inf S_l < \sup S_i$. In the latter case, we have that $\hat l$ satiesfies (3.9) and therefore (3.12). Notice that

$$\sup S_l > \sup S_i. \tag{3.13}$$

Taking $i = i_{\min}$ as our initial pivot, we can iterate the above reasoning. Since there are only finitely many elements, the procedure terminates as ensured by (3.13). □

The key ingredient of this part is the following stability result concerning the notion of martingale C-monotonicity:

Theorem 3.6. *Let $C, C_k \in C(\mathbb{R} \times \mathcal{P}_r(\mathbb{R}))$, $k \in \mathbb{N}$, and C_k converges uniformly to C. If $P, P^k \in \mathcal{P}_r(\mathbb{R} \times \mathcal{P}_r(\mathbb{R}))$, $k \in \mathbb{N}$, are such that*

(a) for all $k \in \mathbb{N}$ the measure P^k is martingale C_k-monotone,

(b) the sequence $\{P^k\}_{k\in\mathbb{N}}$ converges to P,

then P is martingale C-monotone. Moreover, if $\pi, \pi^k \in \mathcal{P}_r(\mathbb{R} \times \mathbb{R})$ and C_k is convex in the second argument and such that

(a') for all $k \in \mathbb{N}$ the measure π^k is martingale C_k-monotone,

(b') the sequence $\{\pi^k\}_{k\in\mathbb{N}}$ converges to π,

then π is martingale C-monotone.

Proof. The proof runs parallel to the one of Theorem 2.8: Using Lemma 3.5 we can alter Lemma 2.7 such that for all $\varepsilon \geq 0$ and $N \in \mathbb{N}$ the set

$$\tilde{\Gamma}_N^\varepsilon := \left\{(x_i, p_i)_{i=1}^N \in (\mathbb{R} \times \mathcal{P}_r(\mathbb{R}))^N \middle| \forall m_1, \ldots, m_N \in P_r(\mathbb{R}) \text{ s.t. } \sum_{i=1}^N p_i = \sum_{i=1}^N m_i \text{ and} \right.$$
$$\left. \int_{\mathbb{R}} y\, m_i(dy) = \int_{\mathbb{R}} y\, p_i(dy) \text{ for } i = 1, \ldots, N, \text{ we have } \sum_{i=1}^N C(x_i, p_i) \leq \sum_{i=1}^N C(x_i, m_i) + \varepsilon \right\}$$

is closed. The aim is to construct a martingale C-monotone set Γ on which P is concentrated. So, we write $P^{k,\otimes N}$ and $P^{\otimes N}$ for the N-fold product measure of P^k resp. P where $N \in \mathbb{N}$. By martingale C_k-monotonicity and uniform convergence we find for any $\varepsilon > 0$ a natural number k_0 such that $P^{k,\otimes N}$, $k \geq k_0$, is concentrated on $\tilde{\Gamma}_N^\varepsilon$. As $\tilde{\Gamma}_N^\varepsilon$ is closed, the Portmanteau theorem yields that $P^{\otimes N}$ is concentrated on $\tilde{\Gamma}_N^\varepsilon$:

$$1 = \limsup_k P^{k,\otimes N}(\tilde{\Gamma}_N^\varepsilon) \leq P^{\otimes N}(\tilde{\Gamma}_N^\varepsilon) = 1.$$

As a consequence, we find that $P^{\otimes N}$ gives full measure to the closed set $\tilde{\Gamma}_N := \tilde{\Gamma}_N^0$. With the same line of argument as in the proof of Theorem 2.8, we can find from here a closed, martingale C-monotone set $\tilde{\Gamma}$ with $P(\tilde{\Gamma}) = 1$.

To show the second assertion, we embed $\pi^k \in \mathcal{P}_r(\mathbb{R} \times \mathbb{R})$ into $\mathcal{P}_r(\mathbb{R} \times \mathcal{P}_r(\mathbb{R}))$ owing to the map J. Then, by compactness of $\Lambda_M(\mu, \nu)$, we find an accumulation point $P \in \mathcal{P}_r(\mathbb{R} \times \mathcal{P}_r(\mathbb{R}))$ of $\{J(\pi^k)\}_{k\in\mathbb{N}}$. Note that P gives full measure to $\{(x, p) \in \mathbb{R}\times\mathcal{P}_1(\mathbb{R})\colon x = \int_{\mathbb{R}} y\, p(dy)\}$, and

$$\mu(dx)\, I(P_x) =: \pi \in \Pi_M(\mu, \nu)$$

determines a martingale coupling, which is likewise a $\mathcal{W}_r$-accumulation point of $\{\pi^k\}_{k\in\mathbb{N}}$. Since $P \in \mathcal{P}_r(\mathbb{R}\times\mathcal{P}_r(\mathbb{R}))$ we have μ-almost surely that $I(P_x) \in \mathcal{P}_r(\mathbb{R})$. As P is concentrated on the martingale C-monotone set $\tilde{\Gamma}$, we find for any $x \in X$ such that $P_x(\tilde{\Gamma}_x) = 1$ and $I(P_x) \in \mathcal{P}_r(\mathbb{R})$ a sequence of measures $p_i^x \in \tilde{\Gamma}_x \subseteq \mathcal{P}_r(\mathbb{R})$, $i \in \mathbb{N}$, with

$$q_n^x := \frac{1}{n} \sum_{i=1}^n p_i^x \to I(P_x) = \pi_x, \quad n \to \infty, \quad \text{in } \mathcal{W}_r.$$

Since C is convex in the second argument, we can reason as in Remark 2.4 (b), and find that for μ-a.e. x (x, q_n^x) is already contained in the martingale C-monotone set $\tilde{\Gamma}$. Then by closure of $\tilde{\Gamma}$ we conclude $(x, \pi_x) \in \tilde{\Gamma}$ for μ-a.e. x, and by the same reasoning as in Remark 2.4 (a) martingale C-monotonicity of π. □

Lemma 3.7. *Let $\pi \in \Pi_M(\mu, \nu)$, $b \in L^1(\nu)$, $c\colon X \times Y \to \mathbb{R}$ be jointly measurable, and $c(x, \cdot)$ be upper semicontinuous and upper bounded by a postive multiple of b for all $x \in \mathbb{R}$. Then π is $(c, \mathcal{F}_M)$-monotone if and only if π is martingale C-monotone (with $C(x, p) := \int_{\mathbb{R}} c(x, y)p(dy)$).*

Proof. Let $\hat{\Gamma} \subseteq \mathbb{R} \times \mathbb{R}$ be $(c, \mathcal{F}_M)$-monotone with $\pi(\hat{\Gamma}) = 1$. Consider the Borel measurable set

$$\Gamma := \left\{(x, p) \in X \times \mathcal{P}_1(Y)\colon p(\hat{\Gamma}_x) = 1, \int_{\mathbb{R}} y\, p(dy) = x, \int_{\mathbb{R}} |b(y)|\, p(dy) < \infty\right\},$$

where $(x, \pi_x) \in \Gamma$ for μ-almost every $x \in X$. Take any sequence $(x_1, p_1), \ldots, (x_N, p_N) \in \Gamma$ with competitors $q_1, \ldots q_N \in \mathcal{P}_1(\mathbb{R})$, i.e.,

$$\sum_{i=1}^N p_i = \sum_{i=1}^N q_i, \quad \int_{\mathbb{R}} y\, p_i(dy) = \int_{\mathbb{R}} y\, q_i(dy).$$

We find for any $(x_i, p_i) \in \Gamma$, a sequence of finitely supported measures $\{p_i^k\}_{k\in\mathbb{N}}$ where for all $k \in \mathbb{N}$ we have that $p_i^k(\hat{\Gamma}_{x_i}) = 1$, $\int_{\mathbb{R}} y\, p_i^k(dy) = x_i$, and

$$\lim_{k\to\infty} \int_{\mathbb{R}} c(x_i, y)\, p_i^k(dy) = \int_{\mathbb{R}} c(x_i, y)\, p_i(dy), \tag{3.14}$$

$$\lim_{k\to\infty} \int_{\mathbb{R}} |b(y)|\, p_i^k(dy) = \int_{\mathbb{R}} |b(y)|\, p(dy). \tag{3.15}$$

Thus, Lemma 3.5 provides $\{q_1^k, \dots, q_N^k\}$, sequences of feasible and finitely supported competitors with corresponding limit points $\{q_1, \dots, q_N\}$. Then $(c, \mathcal{F}_M)$-monotonicity yields

$$\begin{aligned}\sum_{i=1}^N \int_{\mathbb{R}} c(x_i, y)\, p_i(dy) &= \lim_{k\to\infty} \sum_{i=1}^N \int_{\mathbb{R}} c(x_i, y)\, p_i^k(dy) \\ &\le \limsup_{k\to\infty} \sum_{i=1}^N \int_{\mathbb{R}} c(x_i, y)\, q_i^k(dy) \le \sum_{i=1}^N \int_{\mathbb{R}} c(x_i, y)\, q_i(dy),\end{aligned}$$

where we used (3.14) for the first equality, and upper semicontinuity of $c(x, \cdot)$, upper boundedness of $c(x, \cdot) - a(x)b(\cdot)$ for some $a(x) > 0$, and (3.15) in the last inequality. We have shown that Γ is martingale C-monotone. Since $(x, \pi_x) \in \Gamma$ for μ-a.e. $x \in \mathbb{R}$, we conclude that π is martingale C-monotone.

Now, let π be martingale C-monotone on a Borel measurable set $\Gamma \subseteq \mathbb{R} \times \mathcal{P}(\mathbb{R})$, that is, $J(\pi)(\Gamma) = 1$. Due to the variation of Lusin's theorem, Lemma 6.3, there is an analytically measurable set $\hat{\Gamma} \subseteq X \times Y$ satisfying:

(i) For any $(x, p) \in \Gamma$ we have that p is concentrated on the fibre $\hat{\Gamma}_x = \{y \in Y : (x, y) \in \hat{\Gamma}\}$, i.e., $p(\hat{\Gamma}_x) = 1$.

(ii) For any $(x, y) \in \hat{\Gamma}$ we find $(x, p) \in \Gamma$ and a Borel measurable set $K \subseteq \hat{\Gamma}_x$ such that
   (1) c restricted to the fibre $\{x\} \times K$ is continuous,
   (2) $y \in \operatorname{supp}(p|_K) \cap K$ and

$$\int_{B_\delta(y)\cap K} \frac{c(x, z)}{p(B_\delta(y) \cap K)}\, p(dz) \to c(x, y) \quad \text{for } \delta \searrow 0. \tag{3.16}$$

In particular, we have that π is concentrated on $\hat{\Gamma}$ by property (i). To see that $\hat{\Gamma}$ is $(c, \mathcal{F}_M)$-monotone, take a finite subset of $\hat{\Gamma}$, i.e., $G := \{(x_1, y_1), \dots, (x_N, y_N)\} \subseteq \hat{\Gamma}$. Let α be supported on G and β be a competitor, i.e.,

$$\operatorname{proj}_1(\alpha) = \operatorname{proj}_1(\beta), \quad \operatorname{proj}_2(\alpha) = \operatorname{proj}_2(\beta),$$
$$\int_{\mathbb{R}} y\, \alpha_{x_i}(dy) = \int_{\mathbb{R}} y \beta_{x_i}(dy), \quad i = 1, \dots, N.$$

By property (ii) of $\hat{\Gamma}$ we find for each (x_i, y_i), $1 \le i \le N$, $(x_i, p_i) \in \Gamma$ and sets K_i such that c is continuous on $\{x_i\} \times K_i$, $y_i \in K_i$, and (3.16) holds. Let

$$\alpha^k(dx, dy) := \sum_{i=1}^N \delta_{x_i}(dx) \frac{\alpha^k(x_i, y_i)}{p_i(K_i \cap B_{\frac{1}{k}(y_i)})} p_i|_{K_i \cap B_{\frac{1}{k}}(y_i)}(dy), \quad k \in \mathbb{N}.$$

Then α^k converges weakly to α, and by Lemma 3.5 we find a sequence $\{\beta^k\}_{k\in\mathbb{N}}$, where β^k is a competitor of α^k, which converges weakly to β. As $c(x, \cdot)$ is upper semicontinuous and finitely valued, there exists $k_0 \in \mathbb{N}$ with

$$\sup\left\{c(x, y) : y \in \bigcup_{i=1}^N B_{\frac{1}{k_0}}(y_i)\right\} < \infty. \tag{3.17}$$

Thus, we have

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}} c(x, y)\, \alpha(dx, dy) &= \lim_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} c(x, y)\, \alpha^k(dx, dy) \\ &\le \liminf_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} c(x, y) \beta^k(dx, dy) \le \int_{\mathbb{R}\times\mathbb{R}} c(x, y) \beta(dx, dy),\end{aligned}$$

where we obtain the first equality due to (3.16), the first inequality due to martingale C-monotonicity, and the final inequality due to upper semicontinuity and (3.17). □

Proof of Theorem 1.1. Since π^k is optimizer of (MOT) for cost c_k with marginals μ_k and ν_k, π^k is $(c_k, \mathcal{F}_M)$-monotone by [12, Theorem 1.4]. By Lemma 3.7, we find that π^k is martingale C_k-monotone. Then Theorem 3.6 shows that martingale monotonicity is preserved for $k \to \infty$. Hence, π is martingale C-monotone and thus $(c, \mathcal{F}_M)$-monotone by Lemma 3.7. Finally, π is optimal for (MOT) with cost c by [27, Theorem 1.3]. □

Proof of Corollary 1.2. Let π^k be optimal for (MOT) for the cost function c_k and marginal measures μ_k, ν_k. We may apply Theorem 1.1 showing that every accumulation point (with respect to weak convergence) of $\{\pi^k\}_{k\in\mathbb{N}}$ is an optimizer for (MOT) for the cost function c and marginal measures μ, ν. On $\mathbb{R}^2$ we may choose the metric $D((x,y),(\bar{x},\bar{y})) = \sqrt[r]{|x-\bar{x}|^r + |y-\bar{y}|^r}$ in order to define the r-Wasserstein metric on $\mathcal{P}_r(\mathbb{R}^2)$. Then

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}} D((0,0),(x,y))^r\,\pi^k(dx,dy) &= \int_{\mathbb{R}} |x|^r\,\mu_k(dx) + \int_{\mathbb{R}} |y|^r\,\nu_k(dy)\\ &\to \int_{\mathbb{R}} |x|^r\,\mu(dx) + \int_{\mathbb{R}} |y|^r\,\nu(dy) = \int_{\mathbb{R}\times\mathbb{R}} D((0,0),(x,y))\,\pi(dx,dy),\end{aligned}$$

as k tends to ∞ for any coupling π with marginals μ, ν. It follows by [43, Definition 6.8 (i)] that the accumulation points of $\{\pi^k\}_{k\in\mathbb{N}}$ under the weak topology or under the $\mathcal{W}_r$-topology coincide. The following inequality is immediate

$$\liminf_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} c_k(x,y)\,\pi^k(dx,dy) \geq \inf_{\pi\in\Pi_M(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\,\pi(dx,dy).$$

In order to prove

$$\limsup_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} c_k(x,y)\,\pi^k(dx,dy) \leq \inf_{\pi\in\Pi_M(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\,\pi(dx,dy),$$

it suffices to observe that if (for some subsequence which we do not track) $\pi^k \to \pi$ in $\mathcal{W}_r$, then $\int_{\mathbb{R}\times\mathbb{R}} c_k(x,y)\pi^k(dx,dy) \to \int_{R\times\mathbb{R}} c(x,y)\,\pi(dx,dy)$, since π must be optimal for the r.h.s. But this is clear since $c_k \to c$ uniformly and c is dominated by a positive multiple of $(x,y) \mapsto 1 + D((0,0),(x,y))^r$. □

4. The relation of OT and WOT

In this part we explore the relationship between (classical) c-cyclical monotonicity in optimal transport, and C-monotonicity in optimal weak transport, in the case when $C(x,p) = \int_Y c(x,y)p(dy)$. We recall the notion of c-cyclical monotonicity, cf. [43, Definition 5.1]:

Definition 4.1 (c-cyclical monotonicity)**.** *Let $c\colon X\times Y \to \mathbb{R}$.*

(1) A set $\Gamma \subseteq X\times Y$ is called c-cyclically monotone iff for any $N\in\mathbb{N}$, and any collection $(x_1,y_1),\dots,(x_N,y_N)\in\Gamma$, we have

$$\sum_{i=1}^N c(x_i,y_i) \leq \sum_{i=1}^N c(x_i,y_{i+1}),$$

with the convention $y_{N+1} = y_1$.

(2) A probability measure $\pi \in \mathcal{P}(X\times Y)$, which is concentrated on a c-cyclically monotone set, is then called c-cyclically monotone.

Our main result of this section is:

Theorem 4.2. *Let $\pi\in\Pi(\mu,\nu)$, $b\in L^1(\nu)$, $c\colon X\times Y\to\mathbb{R}$ be jointly measurable, and $c(x,\cdot)$ be upper semicontinuous and upper bounded by a positive multiple of b for all $x\in X$. Then π is c-cyclically monotone if and only if π is C-monotone (with $C(x,p) := \int_Y c(x,y)\,p(dy)$).*

Lemma 4.3. *Let $c\colon X \times Y \to \mathbb{R}$ be jointly measurable. Let $\pi \in \Pi(\mu, \nu)$ be C-monotone with $C(x, p) := \int_Y c(x, y)\, p(dy)$ and $c(x, \cdot) \in L^1(\pi_x)$ for μ-almost every $x \in X$. Then for all $A \in \mathcal{B}(X \times Y)$ the restriction of π to A, i.e., $\pi|_A$, is also C-monotone.*

Proof. Let $x_1, \dots, x_N \in \Gamma \cap \{x \in X \colon \pi_x(A_x) > 0\}$, $N \in \mathbb{N}$, where Γ is a C-monotone subset and $A_x := \{y \in Y \colon (x, y) \in A\}$. To shorten notation we write

$$p_i := \pi_{x_i}|_{A_{x_i}}, \quad \bar{p}_i := \frac{1}{p_i(A_{x_i})} p_i, \quad i = 1, \dots, N.$$

Without loss of generality we can assume that restricting and disintegrating commutes, $\bar{p}_i = (\pi|_A)_{x_i}$ for all $i = 1, \dots, N$. Let $n > 1$ be a natural number with reciprocal value smaller than $\min_i p_i(A_{x_i})$. Define the probability measures

$$r_i = \frac{n\pi_{x_i} - \bar{p}_i}{n - 1} \in \mathcal{P}(Y), \quad i = 1, \dots, N,$$

which by linearity of $p \mapsto C(x, p)$ satisfy

$$n \sum_{i=1}^N C(x_i, \pi_{x_i}) = \sum_{i=1}^N C(x_i, \bar{p}_i) + (n - 1) C(x_i, r_i). \tag{4.1}$$

To show C-monotonicity of $\pi|_A$, let $\bar{q}_1, \dots, \bar{q}_N \in \mathcal{P}(Y)$ with $\sum_{i=1}^N \bar{q}_i = \sum_{i=1}^N \bar{p}_i$. Define

$$q_{i,j} := \begin{cases} \bar{q}_i & j = 1, \\ r_i & 2 \le j \le n, \end{cases} \quad i = 1, \dots, N,$$

whence, $q_i := \frac{1}{n} \sum_{j=1}^n q_{i,j} \in \mathcal{P}(Y)$, $i = 1, \dots, N$, define competitors of $(\pi_{x_i})_i$, since

$$\sum_{i=1}^N q_i = \frac{1}{n} \sum_{i=1}^N \sum_{j=1}^n q_{i,j} = \sum_{i=1}^N \pi_{x_i}.$$

From C-monotonicity of π we derive that

$$n \sum_{i=1}^N C(x_i, \pi_{x_i}) \le n \sum_{i=1}^N C(x_i, q_i) = \sum_{i=1}^N \sum_{i=1}^n C(x_i, q_{i,j}),$$

which is by (4.1) equivalent to

$$\sum_{i=1}^N C(x_i, \bar{p}_i) \le \sum_{i=1}^N C(x_i, \bar{q}_i).$$

□

The next proposition has Theorem 4.2 as an immediate corollary:

Proposition 4.4. *Let $c\colon X \times Y \to \mathbb{R}$ be jointly measurable, $b\colon Y \to \mathbb{R}$ be measurable, and $c(x, \cdot)$ be upper semicontinuous and dominated by a positive multiple of b for each $x \in X$. If $\Gamma \subseteq X \times \mathcal{P}(Y)$ is a C-monotone analytic set (where $C(x, p) := \int_Y c(x, y) p(dy)$), then the set $\hat{\Gamma} \subseteq X \times Y$ from Lemma 6.3 is c-cyclically monotone. Conversely, if a c-cyclical monotone set $\hat{\Gamma} \subseteq X \times Y$ is given, then*

$$\Gamma := \left\{(x, p) \in X \times \mathcal{P}(Y) \colon p(\hat{\Gamma}_x) = 1, b \in L^1(p)\right\}$$

is C-monotone.

Proof. Suppose that Γ is C-monotone. Let a finite number of points $(x_1, y_1), \dots, (x_N, y_N)$ in $\hat{\Gamma}$ be given. We find $(x_i, p_i) \in \Gamma$ with $p_i(\hat{\Gamma}_{x_i}) = 1$ as in Lemma 6.3. This means that for each y_i there is a Borel measurable set $K_i \subseteq Y$ with $y_i \in \operatorname{supp}(p_i|_{K_i})$, $c|_{\{x_i\} \times K_i}$ is continuous, and

$$p_i^\varepsilon := \frac{1}{p_i(B_\varepsilon(y_i) \cap K_i)} p_i|_{B_\varepsilon(y_i) \cap K_i}, \quad \int_{K_i \cap B_\varepsilon(y_i)} \frac{c(x_i, z)}{p(K_i \cap B_\varepsilon(y_i))}\, p(dz) \to c(x_i, y_i).$$

The measure $\frac{1}{N}\sum_{i=1}^N \delta_{x_i} p_i$ is by construction C-monotone. By the restriction property of Lemma 4.3, the measure $\frac{1}{N}\sum_{i=1}^N \delta_{x_i} p_i^\varepsilon$ is likewise C-monotone. Then by upper semicontinuity of $c(x,\cdot)$ and (3.17) (which is here applicable), we conclude (with the convention that $N+1=1$):

$$\sum_{i=1}^N c(x_i,y_i) = \lim_{\varepsilon\searrow 0}\sum_{i=1}^N C(x_i,p_i^\varepsilon) \le \limsup_{\varepsilon\searrow 0}\sum_{i=1}^N C(x_i,p_{i+1}^\varepsilon) \le \sum_{i=1}^N c(x_i,y_{i+1}).$$

Now let $\hat\Gamma$ be a c-cyclical monotone set. Recall that $c(x,\cdot)$ is dominated by b and $b\in L^1(p)$ for all $(x,p)\in\Gamma$. Therefore, by the law of large numbers, we have for an iid sequence $(Y_i)_{i\in\mathbb{N}}$ distributed according to p that almost surely

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n c(x,Y_i) = \int_Y c(x,y)\,p(dy),\quad \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n |b(Y_i)| = \int_Y |b(y)|\,p(dy).$$

Thus, we can approximate p by discrete measures concentrated on $\hat\Gamma$ and thereby obtain C-monotonicity of Γ: Let $(x_1,p_1),\dots,(x_N,p_N)\in\Gamma$ and $q_1,\dots,q_N\in\mathcal{P}(Y)$ with $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i$. We find by discretely approximating $p_1,\dots,p_N$ on $\hat\Gamma$, and then using c-cyclical monotonicity, the sequence of competitors constructed in Lemma 2.6, and upper semicontinuity of $c(x,\cdot)$ and upper boundedness of $c(x,\cdot)-a(x)b(\cdot)$ for some $a(x)>0$ that

$$\sum_{i=1}^N C(x_i,p_i) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^N\sum_{j=1}^n c(x_i,y_j^i) \le \limsup_n \sum_{i=1}^N\int_Y c(x_i,y)\,q_i^n(dy) \le \sum_{i=1}^N C(x_i,q_i).$$

□

5. Epilogue

Convexity is a natural assumption in the setting of weak transport. It is known that convexity of $C(x,\cdot)$ is necessary to obtain general existence of minimizers in the space of couplings, see [4, Example 3.2]. Similarly, convexity is required for C-monotonicity to be a necessary optimality criterion:

Example 5.1. Let $X=\{0\}$, $Y=\{0,1\}$, $\mu=\delta_0$, $\nu=\frac{1}{2}(\delta_0+\delta_1)$, and

$$C(x,p)=\min(p(\{0\}),p(\{1\})).$$

Then C is continuous and concave on $X\times\mathcal{P}(Y)$, but the only (and therefore optimal) coupling $\mu\otimes\nu\in\Pi(\mu,\nu)$ is not C-monotone. Indeed,

$$2C(0,\nu)=1>0=C(0,\delta_0)+C(0,\delta_1).$$

In the classical optimal transport setting c-cyclical monotonicity implies optimality already when the cost function c is bounded from below and real valued, see [10]. A similar conclusion cannot be drawn in optimal weak transport as C-monotonicity does not even imply optimality when C is lower continuous:

Example 5.2. Let $X=[0,1]$, $Y=[0,1]$ and $C(x,p):=p_{s,\delta_x}([0,1])=p([0,1]\setminus\{x\})$, which is a measurable cost function (c.f. Proposition 6.5) and lower semicontinuous for fixed $x\in[0,1]$: Given a weakly convergent sequence $p_k\rightharpoonup p\in\mathcal{P}(Y)$, the Portmanteau theorem yields

$$\liminf_k C(x,p_k)=\liminf_k p_k([0,1]\setminus\{x\})\ge p([0,1]\setminus\{x\})=C(x,p).$$

The product coupling $\pi:=\lambda\otimes\lambda\in\Pi(\lambda,\lambda)$ where λ denotes the uniform distribution on $[0,1]$ is in fact C-monotone: Since $\pi_x=\lambda$ (λ-almost surely), we have for any $x_1,\dots,x_N\in[0,1]$ and $q_1,\dots,q_N\in\mathcal{P}(Y)$ with

$$\sum_{i=1}^N \pi_{x_i} = N\lambda = \sum_{i=1}^N q_i$$

that q_i is absolutely continuous to λ, hence,

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

But the unique optimizer of (WOT) with cost C is given by $\pi^*(dx, dy) := \lambda(dx)\delta_x(dy)$ and

$$\int_{[0,1]} C(x, \pi^*_x)\, \lambda(dx) = 0 < 1 = \int_{[0,1]} C(x, \pi_x)\, \lambda(dx).$$

Even when C is given as the integral with respect to some $c\colon X \times Y \to \mathbb{R}$, we cannot hope that C-monotonicity implies optimality and/or c-cyclical monotonicity, as the next example shows.

Example 5.3. Let $X = [0, 1]$, $Y = [0, 1]$ and $C(x, p) = \int_{[0,1]} c(x, y)p(dy)$ with $c(x, y) := \mathbb{1}_{\{x\}}(y)$. As in the previous example, the product coupling $\pi = \lambda \otimes \lambda$ is C-monotone, but not optimal, whereas $\pi^*(dx, dy) = \lambda(dx)\delta_x(dy)$ is optimal and in particular c-cyclical monotone.

The failure of C-monotonicity to provide optimality in these simple settings (the cost function C is even bounded and lower semicontinuous) is caused by the manner it varies over '$X \times Y$': The variation over X is pointwise (similar to c-cyclical monotonicity), whereas over Y variations are taken in a weak sense, i.e., we require that the Y-intensities of the two competing sequences to agree. Here, C-monotonicity is unable to detect the jump from 1 to 0 at x, and we could argue that C-monotonicity yields optimality of π under all couplings χ in $\Pi(\lambda, \lambda)$ such that $\chi_x \ll \lambda$ for λ-almost all $x \in [0, 1]$. To be able to compare with all competing couplings, more regularity of C in 'Y-direction' is necessary, e.g., upper semicontinuity as in Theorem 4.2 or uniform equicontinuity as in Theorem 2.2, but the question remains open precisely how much regularity is required. The authors conjecture that C being upper Carathéodory, i.e., $C\colon X \times \mathcal{P}(Y) \to \mathbb{R}$ is jointly measurable and for fixed $x \in X$ the map $p \mapsto C(x, p)$ is upper semicontinuous, is required for sufficiency of C-monotonicity as optimality criterion.

Notably, Example 5.2, when taking $\tilde{C} := -C$ as cost, provides an upper semicontinuous cost function which is convex (in fact linear) in the following sense: Let $Q \in \mathcal{P}(\mathcal{P}([0, 1]))$, then

$$\int_{\mathcal{P}([0,1])} \tilde{C}(x, p)Q(dp) = \tilde{C}\left(x, \int_{\mathcal{P}([0,1])} pQ(dp)\right). \tag{5.1}$$

Therefore, C being lower semicontinuous and convex is a strictly stronger assumption than the convexity property stated in (5.1) together with measurability, as demanded in Theorem 2.5.

6. Appendix

The existence of a concave modulus of continuity was employed in the proof of Theorem 2.2. We split the argument into a general part, and a part optimized for the setting of the paper.

Lemma 6.1. *Let (Y, d) be a metric space, $r \geq 1$, and $f\colon X \times Y \to \mathbb{R}$ be d-uniformly continuous uniformly in $x \in X$, i.e. such that*

$$\theta(\delta) := \sup\{|f(x, a) - f(x, b)|\colon x \in X,\ a, b \in Y \text{ s.t. } d(a, b)^r \leq \delta\},$$

vanishes for $\delta \searrow 0$. Assume additionally that

$$\forall c > 0\colon \sup_{\substack{a,b\in Y\colon d(a,b)\geq c,\\ x\in X}} \frac{|f(x, a) - f(x, b)|}{d(a, b)^r} < \infty.$$

Then there is $\tilde{\theta} : [0, \infty) \to [0, \infty]$ concave such that $\lim_{\delta\searrow 0} \tilde{\theta}(\delta) = 0$ and

$$|f(x, a) - f(x, b)| \leq \tilde{\theta}\,(d(a, b)^r)\,.$$

Proof. Fixing $\varepsilon > 0$, there is δ_ε such that $|f(x,a) - f(x,b)| \le \varepsilon$ if $d(a,b) \le \delta_\varepsilon$. If on the other hand $d(a,b) > \delta_\varepsilon$ then

$$|f(x,a) - f(x,b)| \le d(a,b)^r \sup_{\substack{\bar a,\bar b\in Y:\, d(\bar a,\bar b)\ge\delta_\varepsilon,\\ \bar x\in X}} \frac{|f(\bar x,\bar a) - f(\bar x,\bar b)|}{d(\bar a,\bar b)^r} =: d(a,b)^r\, K_\varepsilon.$$

Hence overall $|f(x,a) - f(x,b)| \le \varepsilon + d(a,b)^r K_\varepsilon$ and in particular

$$\theta(\delta) \le \varepsilon + \delta K_\varepsilon,$$

and $K_\varepsilon < \infty$. Denoting by $\tilde\theta$ the concave envelope of θ, i.e. the infimum over all affine functions majorizing θ, we conclude $\tilde\theta(\delta) \le \varepsilon + \delta K_\varepsilon$, and so $\lim_{\delta\to 0}\tilde\theta(\delta) = 0$, as ε was arbitrary. Finally remark that $|f(x,a) - f(x,b)| \le \theta(d(a,b)^r) \le \tilde\theta(d(a,b)^r)$.

□

Lemma 6.2. *Let $C\colon X\times\mathcal{P}_r(Y)\to\mathbb{R}$ and suppose that $C(x,\cdot)$ is $\mathcal{W}_r$-uniformly continuous uniformly in $x\in X$, i.e. such that*

$$\theta(\delta) := \sup\left\{|C(x,p) - C(x,q)|\colon x\in X, (p,q)\in\mathcal{P}_r(Y)^2 \text{ s.t.} \mathcal{W}_r(p,q)^r \le \delta\right\},$$

vanishes for $\delta\searrow 0$. Then there is $\tilde\theta : [0,\infty)\to[0,\infty]$ concave such that $\lim_{\delta\searrow 0}\tilde\theta(\delta) = 0$ and

$$|C(x,p) - C(x,q)| \le \tilde\theta\left(\mathcal{W}_r(p,q)^r\right).$$

Proof. We apply Lemma 6.1. It suffices to show that for any $c > 0$

$$\sup_{p,q\in\mathcal{P}_r(Y)\colon \mathcal{W}_r(p,q)^r\ge c} \frac{|C(x,p) - C(x,q)|}{\mathcal{W}_r(p,q)^r} < \infty. \tag{6.1}$$

To this end choose $0 < \delta \le \frac{c}{2}$ such that $\theta(\delta) < \infty$. For any $p,q\in\mathcal{P}_r(Y)$ with $\mathcal{W}_r(p,q)^r \ge c$ there is $N\ge 2$ such that

$$(N-1)\delta \le \mathcal{W}_r(p,q)^r \le N\delta.$$

Denoting $[p,q]_\alpha = (1-\alpha)p + \alpha q$, by convexity of optimal transport we have[4] for all $\alpha,\beta\in[0,1]$

$$\mathcal{W}_r([p,q]_\alpha,[p,q]_\beta)^r \le |\alpha-\beta|\mathcal{W}_r(p,q)^r.$$

Hence,

$$\begin{aligned}|C(x,p) - C(x,q)| &\le \sum_{k=1}^N \left|C(x,[p,q]_{(k-1)/N}) - C(x,[p,q]_{k/N})\right|\\ &\le \sum_{k=1}^N \theta\left(\frac{\mathcal{W}_r(p,q)^r}{N}\right) \le N\theta(\delta) \le \frac{\theta(\delta)\mathcal{W}_r(p,q)^r N}{\delta(N-1)} \le 2\mathcal{W}_r(p,q)^r\frac{\theta(\delta)}{\delta}.\end{aligned}$$

and we conclude that the left-hand side of (6.1) is bounded by $\frac{2\theta(\delta)}{\delta} < \infty$. □

The following lemma can be viewed as a measurable version of Lusin's theorem. It is applied in Section 3 and 4 to connect martingale C-monontone respectively C-monotone sets Γ with $(c,\mathcal{F}_M)$-monotone respectively c-cyclically monotone sets $\hat\Gamma$. Namely this is done in Lemma 3.7 and Proposition 4.4.

Lemma 6.3. *Let $\Gamma\subseteq X\times\mathcal{P}(Y)$ be analytic, and $c\colon X\times Y\to\mathbb{R}\cup\{\pm\infty\}$ be Borel measurable. Then there exists an analytic set $\hat\Gamma\subseteq X\times Y$ with the following properties:*

(i) For any $(x,p)\in\Gamma$ we have that p is concentrated on the fibre $\hat\Gamma_x = \{y\in Y\colon (x,y)\in\hat\Gamma\}$, i.e., $p(\hat\Gamma_x) = 1$.
(ii) For any $(x,y)\in\hat\Gamma$ we find $(x,p)\in\Gamma$ and a Borel measurable set $K\subseteq\hat\Gamma_x$ such that
   (a) c restricted to the fibre $\{x\}\times K$ is continuous,

[4]Indeed, if $\beta > \alpha$ we write $[p,q]_\beta := \frac{1-\beta}{1-\alpha}[p,q]_\alpha + \frac{\beta-\alpha}{1-\alpha}q$ and first deduce $\mathcal{W}_r([p,q]_\alpha,[p,q]_\beta)^r \le \frac{\beta-\alpha}{1-\alpha}\mathcal{W}_r([p,q]_\alpha,q)^r$, by convexity of optimal transport with respect to marginals. Iterating the argument we find $\mathcal{W}_r([p,q]_\alpha,[p,q]_\beta)^r \le (\beta-\alpha)\mathcal{W}_r(p,q)^r$.

(b) $y \in \operatorname{supp}(p) \cap K$ *and*

$$\int_{B_\delta(y)\cap K} \frac{c(x,z)}{p(B_\delta(y)\cap K)} p(dz) \to c(x,y) \quad \textit{for } \delta \searrow 0.$$

Proof. Without loss of generality, we can assume that c is bounded. We follow similar ideas as [22].

To illustrate the idea, let for a moment $K \subseteq Y$ be any measurable set, and write $c_x(y) := c(x,y)$. Continuity of $c_x|_{\{x\}\times K}$ means that for any open set $O \subseteq \mathbb{R}$ the preimage $c_x^{-1}(O) \cap K$ is open in the trace topology on $Y \cap K$. For any radius $\alpha > 0$ and open set $O \subseteq \mathbb{R}$, we can consider an α-doughnut around $c_x^{-1}(O)$, i.e.,

$$\tilde{A}(\alpha) := \left\{ y \in Y \colon \exists z \in c_x^{-1}(O) \text{ s.t. } d_Y(y,z) < \alpha \right\} \setminus c_x^{-1}(O).$$

Clearly, $\tilde{A}(\alpha) \cup c_x^{-1}(O)$ is open and $c_x^{-1}(O) = (\tilde{A}(\alpha) \cup c_x^{-1}(O)) \cap \tilde{A}^c(\alpha)$, thus, $c_x^{-1}(O)$ is a relatively open subset of $\tilde{A}^c(\alpha)$. Note that for any $p \in \mathcal{P}(Y)$, we have by outer regularity that $p(\tilde{A}(\alpha))$ vanishes with α.

Next, we generalize the above reasoning adequately to the product space $X \times Y$ and countably many open sets. Denote by $\{U_n\}_{n\in\mathbb{N}}$ a countable basis of the topology on $\mathbb{R}$. In analogy to $\tilde{A}$, we define for any sequence of radii in Y-direction $\alpha = (\alpha_n)_{n\in\mathbb{N}} \in \mathbb{R}_+^{\mathbb{N}}$, $\alpha_n > 0$ for all $n \in \mathbb{N}$, the $\mathcal{B}(X \times Y)$-measurable set

$$A(\alpha) := \bigcup_{n\in\mathbb{N}} \left\{ (x,y) \in X \times Y \colon \exists (x,z) \in c^{-1}(U_n) \text{ s.t. } d_Y(y,z) < \alpha_n \right\} \setminus c^{-1}(U_n).$$

We see likewise that $c^{-1}(U_n) \cap A(\alpha)^c \cap \{x\} \times Y$ is relatively open in $A(\alpha)^c \cap \{x\} \times Y$. As $\{U_n\}_{n\in\mathbb{N}}$ forms a basis of the topology on $\mathbb{R}$, we find for any open set $O \subseteq \mathbb{R}$ a subset $N \subseteq \mathbb{N}$ with $\bigcup_{n\in N} U_n = O$, thus,

$$c^{-1}(O) \cap A(\alpha)^c \cap \{x\} \times Y = \bigcup_{n\in N} c^{-1}(U_n) \cap A(\alpha)^c \cap \{x\} \times Y,$$

is also relatively open in $A(\alpha)^c \cap \{x\} \times Y$, and consequently $c|_{A(\alpha)^c\cap\{x\}\times Y}$ is continuous for all $x \in X$.

Let $(x,p,\alpha) \in X \times \mathcal{P}(Y) \times \mathbb{R}_+^{\mathbb{N}}$ and $\alpha_n > 0$ for all $n \in \mathbb{N}$, $K := \operatorname{proj}_Y(A(\alpha)^c \cap \{x\} \times Y)$. We have shown for any $y \in \operatorname{supp}(p|_K)$ that $\delta \searrow 0$ implies

$$\frac{1}{p(K \cap B_\delta(y))} p|_{K\cap B_\delta(y)} \to \delta_y \quad \text{in } \mathcal{P}(Y).$$

Since c is bounded and continuous on $\{x\} \times K$, we have for $\delta \searrow 0$

$$\int_{K\cap B_\delta(y)} \frac{c(x,z)}{p(K \cap B_\delta(y))} p(dz) \to c(x,y). \tag{6.2}$$

For any $(x,p) \in X \times \mathcal{P}(Y)$ we have by outer regularity of p that for any $n \in \mathbb{N}$ and $\delta \searrow 0$

$$p\left(\left\{ y \in Y \colon \exists z \in c_x^{-1}(U_n) \text{ s.t. } d_Y(y,z) < \delta \right\}\right) \to p\left(c_x^{-1}(U_n)\right).$$

Therefore, for any $\varepsilon > 0$ there is $\alpha \in \mathbb{R}_+^{\mathbb{N}}$ with $\alpha_n > 0, n \in \mathbb{N}$ such that

$$p\left(\left\{ y \in Y \colon \exists z \in c_x^{-1}(U_n) \text{ s.t. } d_Y(y,z) < \alpha_n \right\} \setminus c_x^{-1}(U_n)\right) < \frac{\varepsilon}{2^n},$$

whence

$$\delta_x \otimes p\,(A(\alpha)) \le \sum_{n\in\mathbb{N}} p\left(\left\{ y \in Y \colon \exists z \in c_x^{-1}(U_n) \text{ s.t. } d_Y(y,z) < \alpha_n \right\}\right) < \varepsilon.$$

The set

$$M^\varepsilon := \left\{ (x,p,\alpha) \in X \times \mathcal{P}(Y) \times \mathbb{R}_+^{\mathbb{N}} \colon \delta_x \otimes p(A(\alpha)) < \varepsilon \right\}$$

satisfies that $\operatorname{proj}_{X\times\mathcal{P}(Y)} M^\varepsilon = X \times \mathcal{P}(Y)$.

We claim that $(x,p,\alpha) \mapsto \delta_x \otimes p|_{A(\alpha)^c}$ and M^ε are both Borel measurable: Evidently, $\alpha \mapsto A(\alpha)$ meets the requirements of Lemma 6.4 below, and consequently the function

$(x, y, \alpha) \mapsto f_A(x, y, \alpha) := \mathbb{1}_{A(\alpha)}(x, y)$ is Borel measurable. From here, we deduce Borel measurability of

$$(x, p, \alpha) \mapsto \delta_x \otimes p(\{x\} \times B \setminus A(\alpha)) = \int_B |1 - f_A(x, y, \alpha)|\ p(dy),$$

for every $B \in \mathcal{B}(Y)$, which yields our claim.

Finally, we are in a position to define $\hat{\Gamma}$: Since Γ is analytic and M^ε is Borel measurable, the set $M^\varepsilon_\Gamma := M^\varepsilon \cap \Gamma \times \mathbb{R}^{\mathbb{N}}_+$ is yet again analytic. Consider the analytic set

$$\begin{aligned}\Theta^\varepsilon :=& \left\{(x, p, \alpha, y) \in X \times \mathcal{P}(Y) \times \mathbb{R}^N_+ \times Y \colon \exists (x, p, \alpha) \in M^\varepsilon_\Gamma \text{ s.t. } y \in \operatorname{supp}(p), \mathbb{1}_{A(\alpha)}(x, y) = 0\right\}\\ =& M^\varepsilon_\Gamma \cap \left\{(x, p, \alpha, y) \in X \times \mathcal{P}(Y) \times \mathbb{R}^{\mathbb{N}}_+ \times Y \colon y \in \operatorname{supp}(p), f_A(x, y, \alpha) = 0\right\},\end{aligned}$$

where we used that Y admits a countable basis $(O_k)_{k\in\mathbb{N}}$ of its topology, whereby

$$y \in \operatorname{supp}(p) \iff \forall k \in \mathbb{N} \colon \text{either } p(O_k) > 0 \text{ or } y \notin O_k.$$

Since Θ^ε is analytic, its $X \times Y$-projection is analytic too. Thus,

$$\hat{\Gamma} := \bigcup_{k\in\mathbb{N}} \operatorname{proj}_{X\times Y} \Theta^{\frac{1}{k}}$$

is analytic. Recall that for any $(x, p) \in \Gamma$ and $k \in \mathbb{N}$ there is $\alpha^k \in (0, \infty)^{\mathbb{N}}$ with $\delta_x \otimes p(A(\alpha^k)) < \frac{1}{k}$. As for every $k \in \mathbb{N}$ we have

$$\begin{aligned}p(\hat{\Gamma}_x) &= \delta_x \otimes p(\hat{\Gamma} \cap X \times \operatorname{supp}(p)) \geq \delta_x \otimes p(\{x\} \times \{y \in Y \colon f_A(x, y, \alpha^k) = 0\})\\ &= 1 - \delta_x \otimes p(A(\alpha^k)) > 1 - \frac{1}{k},\end{aligned}$$

we derive that $\hat{\Gamma}$ satisfies item (i).

On the other hand, if $(x, y) \in \hat{\Gamma}$, then there is $k \in \mathbb{N}$ and $(x, p, \alpha) \in M^\varepsilon_\Gamma$ with $(x, p, \alpha, y) \in \Theta^{\frac{1}{k}}$. Set $K := \operatorname{proj}_Y(A(\alpha)^c \cap \{x\} \times Y) \cap \operatorname{supp}(p)$ and note that $\{x\} \times \{p\} \times \{\alpha\} \times K \subseteq \Theta^\varepsilon$, that means $K \subseteq \hat{\Gamma}_x$. Recall (6.2). By the paragraph preceding (6.2), we have continuity of $c|_{\{x\}\times K}$, and conclude that item (ii) is satisfied. □

Lemma 6.4. *Let $A \colon \mathbb{R}^{\mathbb{N}}_+ \to \mathcal{B}(X \times Y)$ be given s.t.*

(a) for any[5] $\alpha, \beta \in \mathbb{R}^{\mathbb{N}}_+$, $\alpha \leq \beta$ we have $A(\alpha) \subseteq A(\beta)$,

(b) for any $\{\alpha^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^{\mathbb{N}}_+$ with $\alpha^k_n \nearrow \alpha_n, n \in \mathbb{N}$ and $\alpha = (\alpha_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}_+$, we have $A(\alpha^k) \nearrow A(\alpha)$.

Then the function

$$f_A \colon X \times Y \times \mathbb{R}^{\mathbb{N}}_+ \to \mathbb{R} \colon (x, y, \alpha) \mapsto \mathbb{1}_{A(\alpha)}(x, y)$$

is Borel measurable.

Proof. The function f_A is the indicator function of the set

$$\mathcal{A} := \left\{(x, y, \alpha) \in X \times Y \times \mathbb{R}^{\mathbb{N}}_+ \colon (x, y) \in A(\alpha)\right\} = \bigcup_{\alpha\in\mathbb{R}^{\mathbb{N}}} A(\alpha) \times \{\alpha\}.$$

We show that $\mathcal{A}$ coincides with the Borel measurable set $\mathcal{A}'$,

$$\mathcal{A}' := \bigcup_{q\in Q} A(q) \times \bigotimes_{n\in\mathbb{N}} [q_n, \infty),$$

where $Q := \{\alpha \in \mathbb{Q}^{\mathbb{N}}_+ \colon \exists n \in \mathbb{N} \text{ s.t. } \alpha_k = 0\ \forall k \geq n\}$. From property (a) of A we deduce $\mathcal{A} \supseteq \mathcal{A}'$. Now let $(x, y, \alpha) \in \mathcal{A}$, then we find a sequence $\{q^k\}_{k\in\mathbb{N}} \subseteq Q$ with $q^k \nearrow \alpha$, and by (b), there exists an index $k_0 \in \mathbb{N}$ such that $(x, y) \in A(q^k)$ for all $k \geq k_0$. Thus,

$$(x, y, \alpha) \in A(q^{k_0}) \times \bigotimes_{n\in\mathbb{N}} [q^{k_0}_n, \infty) \subseteq \mathcal{A}',$$

and $\mathcal{A} \subseteq \mathcal{A}'$. Therefore $\mathcal{A}$ is a Borel measurable set and f_A a Borel measurable function. □

[5]In this part we write $\alpha \leq \beta \iff \alpha_i \leq \beta_i$ for all i.

We conclude this section by showing the following measurable variant of the Lebesgue decomposition theorem, quoted in Example 5.2.

Proposition 6.5. *Let $\mathcal{M}_+(X)$ the space of all finite measures on X be equipped with the topology of weak convergence of measures. Then the map*

$$T\colon \mathcal{M}_+(X)\times\mathcal{M}_+(X)\to\mathcal{M}_+(X)\times\mathcal{M}_+(X),$$
$$(p,q)\mapsto(q_{ac,p},q_{s,p}),$$

where $T(p,q)$ is the unique Lebesgue decomposition of q w.r.t. p, i.e.,

$$q_{ac,p}+q_{s,p}=q,\quad q_{ac,p}\ll p,\quad q_{s,p}\perp p,$$

is measurable.

Proof. Essentially we are defining a family of measurable functions: for any $\delta>0$ and $A\in\mathcal{B}(X)$ let

$$F_{\delta,A}\colon \mathcal{M}_+(X)\times\mathcal{M}_+(X)\times\mathcal{B}(X)\to\mathbb{R},$$
$$(p,q,B)\mapsto\begin{cases}q(A\setminus B) & p(B)<\delta,\\ q(A) & \text{else.}\end{cases}\tag{6.3}$$

In turn, this allows us to define $T(p,q)(A)$ as a the limit of countable infima of measurable functions. Since X is Polish there exists a countable family $\mathcal{O}$ of open sets on X such that $\sigma(\mathcal{O})=\mathcal{B}(X)$. Denote by $\mathcal{R}(\mathcal{O})$ the set-theoretic ring generated by $\mathcal{O}$, which is again countable. Then, for any $\varepsilon>0$, $A\in\mathcal{B}(X)$ and $p\in\mathcal{P}(X)$ there is a set $A_\varepsilon\in\mathcal{R}(\mathcal{O})$ such that

$$p(A\setminus A_\varepsilon\cup A_\varepsilon\setminus A)<\varepsilon.\tag{6.4}$$

Using the classical Lebesgue decomposition theorem, for any pair $(p,q)\in\mathcal{M}_+(X)\times\mathcal{M}_+(X)$ there is a set $\tilde B\subseteq X$ with $p(\tilde B)=0$ and

$$q_{ac,p}(A)=q(A\setminus\tilde B)\quad\forall A\in\mathcal{B}(X).$$

By [16, Proposition 4.5.3] we have

$$\lim_{\delta\searrow 0}\sup_{B\in\mathcal{B}(X)\colon\, p(B)<\delta}q_{ac,p}(B)=0,$$

whence,

$$\begin{aligned}q_{ac,p}(A)&=\lim_{\delta\searrow 0}\inf_{B\in\mathcal{B}(X)\colon\, p(B)<\delta}q_{ac,p}(A\setminus B)\\&=\lim_{\delta\searrow 0}\inf_{B\in\mathcal{B}(X)\colon\, p(B)<\delta}q_{ac,p}(A\setminus(\tilde B\cup B))+q_{s,p}(A\setminus(\tilde B\cup B))\\&=\lim_{\delta\searrow 0}\inf_{B\in\mathcal{B}(X)\colon\, p(B)<\delta}q(A\setminus B)\\&=\lim_{\delta\searrow 0}\inf_{B\in\mathcal{R}(\mathcal{O})\colon\, p(B)<\delta}q(A\setminus B),\end{aligned}$$

where we used for the last equality the approximation property (6.4).

Thus $F_A\colon\mathcal{M}_+(X)\times\mathcal{M}_+(X)\to\mathbb{R}$ defined by

$$F_A(p,q):=q_{ac,p}(A)=\inf_{\delta\searrow 0}\inf_{\{B\in\mathcal{B}(X)\colon\, p(B)<\delta\}}q(A\setminus B)=\lim_{k\to\infty}\inf_{B\in\mathcal{R}(\mathcal{O})}F_{\frac1k,A}(p,q,B)\tag{6.5}$$

is Borel measurable, and as $A\in\mathcal{B}(X)$ was arbitrary we conclude that T is measurable. □

References

[1] B. Acciaio, J. Backhoff, and A. Zalashko. Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization. *Stochastic Processes and their Applications*, 130(5):2918–2953, 2020.

[2] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems. *Available at SSRN 3072356*, 2017.

[3] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of probability measures in the convex order by Wasserstein projection. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 1706–1729. Institut Henri Poincaré, 2020.

[4] J.-J. Alibert, G. Bouchitté, and T. Champion. A new class of costs for optimal transport planning. *European Journal of Applied Mathematics*, 30(6):1229–1263, 2019.

[5] J. Backhoff, D. Bartl, M. Beiglböck, and E. Manu. Adapted Wasserstein distances and stability in mathematical finance. *Finance and Stochastics*, 24(3):601–632, 2020.
[6] J. Backhoff, M. Beiglböck, M. Eder, and A. Pichler. Fundamental properties of process distances. *Stochastic Processes and their Applications*, 2020.
[7] J. Backhoff, M. Beiglböck, M. Huesmann, S. Källblad, et al. Martingale Benamou–Brenier: a probabilistic perspective. *Annals of Probability*, 48(5):2258–2289, 2020.
[8] J. Backhoff, M. Beiglbock, Y. Lin, and A. Zalashko. Causal transport in discrete time and applications. *SIAM Journal on Optimization*, 27(4):2528–2562, 2017.
[9] J. Backhoff, M. Beiglböck, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, 58(6):203, 2019.
[10] M. Beiglböck. Cyclical monotonicity and the ergodic theorem. *Ergodic Theory and Dynamical Systems*, 35(3):710–713, 2015.
[11] M. Beiglböck, A. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. *Inventiones Mathematicae*, 208(2):327–400, 2017.
[12] M. Beiglböck and C. Griessler. A land of monotone plenty. *Annali della SNS, to appear*, Apr. 2016.
[13] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: A mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.
[14] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Annals of Probability*, 44(1):42–106, 2016.
[15] M. Beiglboeck and N. Juillet. Shadow couplings. *ArXiv e-prints*, Sept. 2016.
[16] V. I. Bogachev. *Measure theory*, volume 1. Springer Science and Business Media, 2007.
[17] L. Campi, I. Laachir, and C. Martini. Change of numeraire in the two-marginals martingale transport problem. *Finance and Stochastics*, 21(2):471–486, June 2017.
[18] A. Cox and M. Vidmar. The structure of non-linear martingale optimal transport problems. *ArXiv e-prints*, 2019.
[19] H. De March, N. Touzi, et al. Irreducible convex paving for decomposition of multidimensional martingale transport plans. *The Annals of Probability*, 47(3):1726–1774, 2019.
[20] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probability Theory and Related Fields*, 160(1-2):391–427, 2014.
[21] M. Fathi, N. Gozlan, and M. Prodhomme. A proof of the Caffarelli contraction theorem via entropic regularization. *Calculus of Variations and Partial Differential Equations*, 59(96), 2020.
[22] M. B. Feldman. A proof of Lusin's theorem. *The American Mathematical Monthly*, 88(3):191–192, 1981.
[23] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Annals of Applied Probability*, 24(1):312–336, 2014.
[24] N. Ghoussoub, Y.-H. Kim, and T. Lim. Structure of optimal martingale transport plans in general dimensions. *The Annals of Probability*, 47(1):109–164, 2019.
[25] N. Gozlan and N. Juillet. On a mixture of Brenier and Strassen theorems. *Proceedings of the London Mathematical Society*, 120(3):434–463, 2020.
[26] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, 273(11):3327–3405, 2017.
[27] C. Griessler. An extended footnote on finitely minimal martingale measures. *ArXiv e-prints*, June 2016.
[28] G. Guo and J. Obloj. Computational Methods for Martingale Optimal Transport problems. *ArXiv e-prints*, page arXiv:1710.07911v1, Oct 2017.
[29] G. Guo and J. Obloj. Computational Methods for Martingale Optimal Transport problems. *ArXiv e-prints*, page arXiv:1710.07911v2, April 2019.
[30] I. Guo, G. Loeper, and S. Wang. Local volatility calibration by optimal transport. In *2017 MATRIX Annals*, pages 51–64. Springer, 2019.
[31] J. Guyon, R. Menegaux, and M. Nutz. Bounds for VIX futures given S&P 500 smiles. *Finance and Stochastics*, 21(3):593–630, 2017.
[32] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Mathematical Finance*, 22(1):31–56, 2012.
[33] M. Huesmann and D. Trevisan. A Benamou–Brenier formulation of martingale optimal transport. *Bernoulli*, 25(4A):2729–2757, 2019.
[34] B. Jourdain and W. Margheriti. A new family of one dimensional martingale couplings. *Electronic Journal of Probability*, 25, 2020.
[35] N. Juillet. Stability of the shadow projection and the left-curtain coupling. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 52, pages 1823–1843. Institut Henri Poincaré, 2016.
[36] K. Marton. A measure concentration inequality for contracting Markov chains. *Geometric & Functional Analysis GAFA*, 6(3):556–571, 1996.
[37] K. Marton et al. Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration. *The Annals of Probability*, 24(2):857–866, 1996.
[38] J. Obłój and P. Siorpaes. Structure of martingale transports in finite dimensions. arXiv:1702.08433, 2017.
[39] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 81(1):73–205, 1995.

[40] M. Talagrand. New concentration inequalities in product spaces. *Inventiones mathematicae*, 126(3):505–563, 1996.
[41] X. Tan and N. Touzi. Optimal transportation under controlled stochastic dynamics. *Annals of Probability*, 41(5):3201–3240, 2013.
[42] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
[43] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.
[44] J. Wiesel. Continuity of the martingale optimal transport problem on the real line. *ArXiv e-prints*, 2019.
[45] D. A. Zaev. On the Monge–Kantorovich problem with additional linear constraints. *Mathematical Notes*, 98(5-6):725–741, 2015.

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

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

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 a 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.

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

1 Introduction

Subsequently we will carefully explain all required notation and describe relevant literature in some detail. Before going into this, we provide here a first description of our main result and its relevance for martingale transport theory.

In brief, while classical transport theory is concerned with couplings $\Pi(\mu, \nu)$ of probability measures μ, ν, the martingale variant restricts the problem to the set of couplings which also satisfy the martingale constraint, in symbols $\Pi_M(\mu, \nu)$. Already in the simple (and most relevant) case where μ, ν are probabilities on the real line, many arguments and result appear significantly more involved in the martingale context. A basic explanation lies 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 classical transport theory:

Fact 1.1 (Stability of Couplings)**.** *Let $\pi \in \Pi(\mu, \nu)$ and assume that μ^k, ν^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 is so basic and straightforward that its implicit use is easily overlooked. Note however that it plays a crucial role in a number of instances, 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 the martingale counterpart of Fact 1.1, see Theorem 2.5 below. This closes a gap in the theory of martingale transport. It yields basic fundamental results in a unified fashion that it is much closer to the classical theory and it allows to address questions in Martingale optimal transport, optimal Skorokhod embedding and robust finance that have previously remained open. We will consider these applications systematically in two accompanying articles.

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

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

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

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

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 first marginal μ and 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, 42, 44, 45] for a study in depth.

The theory of OT goes back to Gaspard Monge [37] and, resp. Kantorovich [32] 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 suitable for applications in mathematical finance, in particular model-independent pricing was introduced in [11] in a discrete time setting and in [22] in a continuous time setting. Compared to 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 $\pi \in \mathcal{P}(X \times Y)$, we denote by $(\pi_x)_{x\in X}$ its 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 [43], the existence of a martingale coupling between two probability measures $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$ 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 $\mathbb{R}^d$ with finite first moment and are said to be in the convex order iff we have

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

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

For adaptations of classical optimal transport theory to the MOT problem, we refer to Henry-Labordère, Tan and Touzi [28] and Henry-Labordère and Touzi [29]. Concerning duality results, we refer to Beiglböck, Nutz and Touzi [14], Beiglböck, Lim and Obłój [13] and De March [18]. We also refer to De March [17] and De March and Touzi [19] for the multi-dimensional case.

Concerning the numerical resolution of the MOT problem, we refer to the articles of Alfonsi, Corbetta and Jourdain [2, 3], De March [16], Guo and Obłój [25] and Henry-Labordère [27]. 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 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 [31] who established stability of left-curtain coupling. Guo and Obłój [25] establish the result under moment conditions. More recently, Backhoff-Veraguas and Pammer [10] and Wiesel [46] independently gave a definite positive answer.

1.2 The Adapted Wasserstein distance

The stability result shown by Backhoff-Veraguas and Pammer 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 wellsuited 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 haven independently introduced refinements of the weak topology, we mention Hellwig's information topology [26], Aldous's extended weak topology [1], the nested distance / adapted Wasserstein distance of Plug-Pichler [38] and the optimal stopping topology [6]. Strikingly, all those seemingly different definitions lead to same topology in 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 [38, 39, 40, 41, 36, 15] 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.

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

Similarly, we denote by $\mathcal{M}(X)$ (resp. $\mathcal{M}_r(X)$) 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

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

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

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

If moreover μ and μ^k, $k \in \mathbb{N}$, have equal mass, then the convergence (1.1) can be equivalently formulated 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 $\mathcal{W}_r \leq \mathcal{AW}_r$, 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 mass, their adapted Wasserstein distance coincides with

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

Therefore, the topology induced by $\mathcal{AW}_r$ coincides with the initial topology w.r.t. J.

Finally, let us mention the interpretation of adapted Wasserstein distance in terms of bicausal couplings as done in [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 $\chi(dx, dx') = \int_{(y,y')\in Y\times Y} \eta(dx, dy, dx', dy') \in \Pi(\mu, \mu')$. With a slight abuse of notation we write $\chi(dx, dx') = \mu(dx)\, \chi_x(dx') = \mu'(dx')\, \chi_{x'}(dx)$. Then η is called bicausal iff

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

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 η is bicausal iff $\chi(dx, dx')$-almost everywhere, $\gamma_{(x,x')}(dy, dy') \in \Pi(\pi_x, \pi_{x'})$. Then the adapted Wasserstein distance coincides with

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

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

1.3 Outline

Section 2 presents the main result of this article, namely Theorem 2.5 below. We give an explanation of how the conclusion of this theorem was expectable. Moreover, we give a sketch of its proof in order to help seeing through the technical details.

Section 3 deals with technical lemmas which allow us to deal with difficulties specific to the adapted Wasserstein distance with more ease. They mainly explore properties about 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. Moreover, it contains the proof of a key result which allows us to see that the main theorem needs only to be proved for irreducible pairs of marginals.

Section 5 is as the name suggests devoted to the proof of the main theorem. Before entering its technical proof, we prove that is is enough to prove $\mathcal{AW}_1$-convergence for irreducible pairs of marginals.

2 Main result

Our main result is Theorem 2.5 below. Before, we state a proposition which enlightens us about why the conclusion of the theorem should be expected (or at least hoped for). We also state the 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 with a sketch of its proof. It is understood that (X, d_X) and (Y, d_Y) denote any Polish spaces, and (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 $(\mu^k)_{k\in\mathbb{N}}$, $(\nu^k)_{k\in\mathbb{N}}$ all with equal mass respectively to $\mu, \nu \in \mathcal{M}_r(X)$, then, informally speaking, there holds

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

i.e., any sequence with convergent marginals has accumulation points in $\Pi(\mu, \nu)$, and for any $\pi \in \Pi(\mu, \nu)$ there 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, let $\eta^k \in \Pi(\mu^k, \mu)$, resp. $\tau^k \in \Pi(\nu, \nu^k)$ be optimal for $\mathcal{W}_r(\mu^k, \mu)$, resp. $\mathcal{W}_r(\nu, \nu^k)$ and set

$$\pi^k(dx^k, dy^k) = \int_{(x,y)\in X\times Y} \eta^k(dx^k, dx)\, \pi_x(dy)\, \tau^k_y(dy^k).$$

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

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

The following properties are standard results (see for instance [30, Section 6] for proofs):

(a) F_η is càdlàg, F_η^{-1} is càglàd ;

(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

$$F_\eta(x-) < u \leq F_\eta(x) \implies x = F_\eta^{-1}(u), \tag{2.4}$$
$$\text{and} \quad F_\eta(F_\eta^{-1}(u)-) \leq u \leq F_\eta(F_\eta^{-1}(u)); \tag{2.5}$$

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

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

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

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

Proposition 2.1. *Let $\mu, \mu^k, \nu, \nu^k \in \mathcal{M}_r(\mathbb{R})$, $k \in \mathbb{N}$, be with equal mass 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 when μ^k is non-atomic, then

$$\mathcal{AW}_r^r(\pi, \pi^k) \leq \mathcal{W}_r^r(\mu, \mu^k) + \mathcal{W}_r^r(\nu, \nu^k). \tag{2.6}$$

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, $\pi^k, k \in \mathbb{N}$ is almost a sequence of martingale couplings.

In the setting of Proposition 2.1, 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 2.5 below. We mention that the previous proposition generalises to any 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 mass 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 2.5 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 Section 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, $\mu(I) = \mu(\mathbb{R})$ and $\nu(\overline{I}) = \nu(\mathbb{R})$. If $a \in \mathbb{R}$ 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 [12, 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

$$\forall n \in N, \quad \pi_n(dx, dy) = \mu_n(dx) \, \pi_x(dy). \tag{2.7}$$

Proposition 2.4. *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*

$$\eta^k + \sum_{n \in N} \mu_n^k = \mu^k, \quad \upsilon^k + \sum_{n \in N} \nu_n^k = \nu^k, \quad k \in \mathbb{N}, \tag{2.8}$$

$$\lim_{k \to +\infty} \eta^k = \eta, \quad \lim_{k \to +\infty} \mu_n^k = \mu_n, \quad \lim_{k \to +\infty} \nu_n^k = \nu_n, \quad \lim_{k \to +\infty} \upsilon^k = \eta \quad \textit{in } \mathcal{W}_1. \tag{2.9}$$

We can now state our main result, namely Theorem 2.5 below. Any martingale coupling whose marginals are approximated by probability measures in convex order can be approximated by martingale couplings with respect to the adapated Wasserstein distance.

Theorem 2.5. *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 $\mathcal{AW}_r$.*

Sketch of the proof. We will first argue that it is enough to consider the case $r = 1$. Thanks to Proposition 2.4, we can also reduce the proof to the case of irreducible pairs of marginals (μ, ν), whose single irreducible component is denoted $(l, r) = I$.

Step 1. Fix any 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 this, the kernel π_x is first compactified to a compact set $[-R, R]$, where $R > 0$, 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 to 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 Figure 1, where J denotes a compact subset of I large enough.

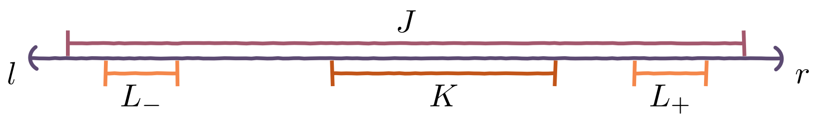

Figure 1: Intervals involved in the proof. The boundaries of the closed intervals are vertical bars and those of the open intervals are parenthesis.

Step 2. It is possible to find an approximative sequence $(\hat{\pi}^k = \hat{\mu}^k \times \hat{\pi}^k_x)_{k\in\mathbb{N}}$ of 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 barycentres of the kernels π^k_x. Hence we find a sequence $(\tilde{\pi}^k = \hat{\mu}^k \times \tilde{\pi}^k_x)_{k\in\mathbb{N}}$ 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 $\hat{\mu}^k \leq \mu^k$. Moreover, its second marginal denoted $\tilde{\nu}^k$ is 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 [30], 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 weak adapted 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 $\varepsilon > 0$, let*

$$I^r_\varepsilon(\mu) := \sup_{\substack{\tau\in\mathcal{M}(X) \\ \tau\leq\mu,\ \tau(X)\leq\varepsilon}} \int_X d^r_X(x, x_0)\, \tau(dx). \tag{3.1}$$

(a) I^r_ε is monotone in μ, i.e., $\mu \leq \mu' \in \mathcal{M}_r(X)$ implies that $I^r_\varepsilon(\mu) \leq I^r_\varepsilon(\mu')$.

(b) The value of $I^r_\varepsilon(\mu)$ vanishes as $\varepsilon \to 0$.

(c) For any $\mu' \in \mathcal{M}_r(X)$ such that $\mu(X) = \mu'(X)$ we have

$$I^r_\varepsilon(\mu) \leq 2^{r-1}\left(I^r_\varepsilon(\mu') + \mathcal{W}^r_r(\mu, \mu')\right). \tag{3.2}$$

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

$$\mathcal{W}_r(\mu^k, \mu) \underset{k\to+\infty}{\longrightarrow} 0 \iff \sup_{k\in\mathbb{N}} I^r_\varepsilon(\mu^k) \underset{\varepsilon\to 0}{\longrightarrow} 0 \quad \text{and} \quad \sup_{k\in\mathbb{N}} \int_X d^r_X(x,x_0)\, \mu^k(dx) < +\infty.$$

(e) *Finally, if $X = \mathbb{R}^d$ and $\mu \le_c \nu$ with $\nu \in \mathcal{M}_1(\mathbb{R}^d)$, then $I^1_\varepsilon(\mu) \le I^1_\varepsilon(\nu)$.*

Remark 3.2. If $\mu(X) \le \varepsilon$, then $I^r_\varepsilon(\mu)$ is simply the r-th moment of μ.

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

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 $x \mapsto d^r_X(x,x_0)$. It is then enough to show that

$$I^r_\varepsilon(\mu) = \int_{1-\varepsilon}^1 F^{-1}_{\overline{\mu}}(u)\, du. \tag{3.4}$$

Indeed, since $\mu \in \mathcal{P}_r(X)$ we have $\int_X d_X(x,x_0)^r\, \mu(dx) = \int_0^1 F^{-1}_{\overline{\mu}}(u)\, du < +\infty$, so we conclude by dominated convergence that $I^r_\varepsilon(\mu)$ vanishes as $\varepsilon \to 0$. Of course an appropriate upper bound would be sufficient, but (3.4) will come in handy for the proof of the last part. Let $\tau^* \in \mathcal{M}(X)$ be such that $\overline{\tau^*}$ is the right-most measure dominated by $\overline{\mu}$ with mass equal to ε, namely

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

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 is zero if $\mu(B_\varepsilon) = 0$. By (2.5) and definition of y_ε, we have

$$F_{\overline{\mu}}(y_\varepsilon-) \le 1-\varepsilon, \tag{3.5}$$

hence $\mu(B_\varepsilon) = \overline{\mu}(\{y_\varepsilon\}) \ge F_{\overline{\mu}}(y_\varepsilon) - (1-\varepsilon)$ and therefore $\tau^* \le \mu$. Moreover,

$$\tau^*(\mathbb{R}) = \mu(A_\varepsilon) + F_{\overline{\mu}}(y_\varepsilon) - (1-\varepsilon) = \overline{\mu}((y_\varepsilon, +\infty)) - (1 - F_{\overline{\mu}}(y_\varepsilon)) + \varepsilon = \varepsilon.$$

Let us show that

$$\int_X d^r_X(x,x_0)\, \tau^*(dx) = \int_{1-\varepsilon}^1 F^{-1}_{\overline{\mu}}(u)\, du. \tag{3.6}$$

According to (3.5), $(1-\varepsilon, F_{\overline{\mu}}(y_\varepsilon)] \subset (F_{\overline{\mu}}(y_\varepsilon-), F_{\overline{\mu}}(y_\varepsilon)]$, so by (2.4), for all $u \in (1-\varepsilon, F_{\overline{\mu}}(y_\varepsilon)]$, $F^{-1}_{\overline{\mu}}(u) = y_\varepsilon$. Then, using the inverse transform sampling and (2.3) for the second equality, we get

$$\begin{aligned}
\int_X d^r_X(x,x_0)\, \tau^*(dx) &= \int_{\mathbb{R}} y \mathbb{1}_{(y_\varepsilon,+\infty)}(y)\, \overline{\mu}(dy) + (F_{\overline{\mu}}(y_\varepsilon) - (1-\varepsilon)) y_\varepsilon \\
&= \int_{F_{\overline{\mu}}(y_\varepsilon)}^1 F^{-1}_{\overline{\mu}}(u)\, du + \int_{1-\varepsilon}^{F_{\overline{\mu}}(y_\varepsilon)} F^{-1}_{\overline{\mu}}(u)\, du \\
&= \int_{1-\varepsilon}^1 F^{-1}_{\overline{\mu}}(u)\, du,
\end{aligned}$$

hence (3.6) holds.

Let $\tau \in \mathcal{M}(X)$ be such that $\tau \le \mu$ and $0 < \tau(X) \le \varepsilon$. In view of (3.6), it is enough in order to prove (3.4) to show that

$$\int_X d^r_X(x,x_0)\, \tau(dx) \le \int_{1-\varepsilon}^1 F^{-1}_{\overline{\mu}}(u)\, du. \tag{3.7}$$

Since $\tau \le \mu$, we have $\overline{\tau} \le \overline{\mu}$. Using (2.5) for the last inequality, we get for all $u \in (0,1)$

$$1 - F_{\overline{\tau}/\tau(X)}\left(F_{\overline{\mu}}^{-1}(1-\tau(X)u)\right) = \frac{\overline{\tau}((F_{\overline{\mu}}^{-1}(1-\tau(X)u), +\infty))}{\tau(X)} \le \frac{\overline{\mu}((F_{\overline{\mu}}^{-1}(1-\tau(X)u), +\infty))}{\tau(X)} \le u,$$

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

$$\begin{aligned}\int_X d_X^r(x,x_0)\,\tau(dx) &= \tau(X)\int_0^1 F_{\overline{\tau}/\tau(X)}^{-1}(u)\,du = \tau(X)\int_0^1 F_{\overline{\tau}/\tau(X)}^{-1}(1-u)\,du \\ &\le \tau(X)\int_0^1 F_{\overline{\mu}}^{-1}(1-\tau(X)u)\,du = \int_{1-\tau(X)}^1 F_{\overline{\mu}}^{-1}(u)\,du \le \int_{1-\varepsilon}^1 F_{\overline{\mu}}^{-1}(u)\,du,\end{aligned}$$

hence (3.7) 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 $\tau \le \mu$ and $\tau(X) \le \varepsilon$. 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 $\tau' \le \mu'$ and $\tau(X) = \tau'(X)$. Then

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

which shows by optimality of π the assertion.

We now show (d). Let $\mu, \mu^k \in \mathcal{M}_r(X)$ be with equal mass 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 $\varepsilon > 0$

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

According to (b) we then get

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

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 $M > 0$,

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

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 vanished 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 $M > M_0$,

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

Therefore, for all $M > M_0$,

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

Since η is arbitrary, this proves the assertion.

Finally, we want to show (e). Let $X = \mathbb{R}^d$ and $\mu \leq_c \nu$ 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,

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

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

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

The identity (3.4) 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 $\Pi(\mu, \nu)$. Contrary, when $\pi \in \Pi(\mu, \nu)$ 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

$$C_u = ((y, w) \mapsto F_\nu(y-) + w\nu(\{y\}))_*(\pi_{F_\mu^{-1}(u)} \times \lambda). \tag{3.8}$$

In particular, we have that $u \mapsto C_u$ is constant on the jumps on F_μ. By the inverse transform sampling, showing that $\pi = (F_\mu^{-1}, F_\nu^{-1})_* C$ amounts to showing that the second marginal distribution of C is indeed uniformly distributed on $(0, 1)$, which is a direct consequence of the inverse transform sampling and the well-known result (see for instance [30, Lemma 6.6] for a proof) that for any $\eta \in \mathcal{P}(\mathbb{R})$,

$$((z, w) \mapsto F_\eta(z-) + w\eta(\{z\}))_*(\eta \times \lambda) = \lambda. \tag{3.9}$$

Moreover, we have by (2.4) that $F_\nu^{-1}(F_\nu(y-) + w\nu(\{y\})) = y$ for all $(y, w) \in \mathbb{R} \times (0, 1]$, hence for all $u \in (0, 1)$,

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

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 (3.8).

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

$$\theta_k\colon \mathbb{R} \times (0, 1) \to (0, 1),\ (x, w) \mapsto F_{\mu^k}(x-) + w\mu^k(\{x\}),$$

$$C_u^k(dv) = \int_{w=0}^1 C_{\theta_k\left(F_{\mu^k}^{-1}(u), w\right)}(dv)\,dw,$$

$$\pi^k = (F^{-1}_{\mu^k}, F^{-1}_{\nu^k})_* C^k = (F^{-1}_{\mu^k}, F^{-1}_{\nu^k})_* (\mathbb{1}_{(0,1)}(u)\, du\, C^k_u(dv)).$$

The fact that C^k is a copula, and therefore $\pi^k \in \Pi(\mu^k, \nu^k)$, is a direct consequence of (3.9) 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

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

By Minkowski's inequality we have

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

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

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

where we used inverse transform sampling in the second equality. At this stage, we can already show (b). 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 $C_{\theta_k(F^{-1}_{\mu^k}(u), w)} = C_u$ and particularly $C^k_u = C_u$, which lets vanish the first term in the right-hand side of (3.12). Then the estimate (2.6) 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

$$\int_0^1 \mathcal{W}^r_r\left((F^{-1}_\nu)_* C_u, (F^{-1}_\nu)_* C^k_u \right) du \to 0.$$

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

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

Second, we prove that

$$u \mapsto \mathcal{W}^r_r((F^{-1}_\nu)_* C_u, (F^{-1}_\nu)_* C^k_u) \quad k \in \mathbb{N}, \tag{3.15}$$

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 [21, 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

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

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

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

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) writes

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

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, it is clear with (2.5) in mind 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)

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

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 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

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

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

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

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

As mentioned in Section 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 [24] and studied in [23], formulated according to our needs. 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 any 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 assumption we made on C:

(a) (WOT) admits a unique minimiser π^* [9, Theorem 1.2];

(b) As a consequence of the necessary optimality criterion [10, Theorem 2.2] and [10, Remark 2.3.(a)], $J(\pi^*)$ is the only minimiser of (WOT')

(c) $V(\mu,\nu) = V'(\mu,\nu)$ [9, 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 $\mathcal{W}_r$ [10, Theorem 1.3 and Corollary 2.8]. In particular, this shows that π^k converges to π^* even in $\mathcal{AW}_r$.

Proof of Proposition 2.3. 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

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

Since $\nu \in \mathcal{P}_r(\mathbb{R})$, $\int_X \int_Y d_Y^r(y,y_0)\, \pi_x(dy)\, \mu(dx) = \int_Y d_Y^r(y,y_0)\, \nu(dy) < +\infty$, hence μ is concentrated on $\bigcup_{R>0} A_R$ and we can choose R large enough such that

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

By Lusin's theorem, there exists a closed set $F \subset A_R$ such that

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

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 [20, 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 [21, 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

$$\int_Y g_l(y)\, p(dy) \neq \int_Y g_l(y)\, p'(dy). \tag{3.25}$$

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 $(x,p), (x',p') \in X \times \mathcal{P}_r(Y)$

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

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 (3.25) 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 $(x,p) \in X \times \mathcal{P}_r(Y)$, $C(x,p) \le 1+8 = 9$, 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

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

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

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

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

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

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

$$\limsup_{k\to+\infty} \tilde{\rho}(J(\pi^{k,\varepsilon}), J(\pi)) \le \limsup_{k\to+\infty} \left(\tilde{\rho}(J(\pi^{k,\varepsilon}), J(\pi^{*,\varepsilon})) + \tilde{\rho}(J(\pi^{*,\varepsilon}), J(\pi))\right) \le 11\varepsilon. \tag{3.26}$$

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

$$\forall N \in \mathbb{N}^*, \quad \forall k \ge k_N, \quad \tilde{\rho}(J(\pi^{k,2^{-rN}}), J(\pi)) \le 12 \cdot 2^{-rN}.$$

For $k \in \mathbb{N}$, let $N_k = \max\{N \in \mathbb{N} \mid k \ge k_N\}$, where the maximum of the empty set is defined as 0. Since k_N is strictly increasing, we find that $N_k \to +\infty$ as $k \to +\infty$. Then the sequence of couplings

$$\pi^k = \pi^{k,2^{-rN_k}} \in \Pi(\mu^k,\nu^k), \;\; k \in \mathbb{N}$$

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

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

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 2.5 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 mass and $\pi^k \in \Pi(\mu^k,\nu^k)$, $k \in \mathbb{N}$, converge to $\pi \in \Pi(\mu,\nu)$ in $\mathcal{AW}_r$.*

(i) Let $A \subset X$ be measurable and $B \supset A$ be open. There are $\tilde{\mu}^k \le \mu^k|_B$ and $\varepsilon_k \ge 0$, $k \in \mathbb{N}$ such that $\tilde{\pi}^k := \tilde{\mu}^k \times \pi^k_x$ satisfies

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

(ii) Let $C \subset X$ 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 $\varepsilon'_k \geq 0$, $k \in \mathbb{N}$ such that

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

Proof. Both assertions are trivial if $\mu(A) = 0$. 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 $\chi = (\mathrm{id}, \mathrm{id})_*\mu$.
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

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

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 easily deductible that $P(\eta') \to P(\eta)$ weakly if and only if $\eta' \to \eta$ in $\mathcal{W}_r$. In particular, we find $P(\chi^k) \to P(\chi)$ 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 component. As a consequence of [35, 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

$$\chi^k(B \times A) \leq \mu(A) = \chi|_{X\times A}(B \times B) \leq \liminf_{k\to+\infty} \chi^k|_{X\times A}(B \times B) = \liminf_{k\to+\infty} \chi^k(B \times A),$$

hence $\varepsilon_k := 1 - \frac{\chi^k(B\times A)}{\mu(A)}$, $k \in \mathbb{N}$ is a null sequence of nonnegative real numbers. Denote by $\tilde{\mu}^k$, resp. $\bar{\mu}^k$ the first resp. second, marginal of $\chi^k|_{B\times A}$, $k \in \mathbb{N}$. Note that

$$1 - \varepsilon_k = \frac{\tilde{\mu}^k(X)}{\mu(A)}. \tag{3.27}$$

We want to show that

$$\mathcal{AW}_r(\tilde{\mu}^k \times \pi^k_x, (1 - \varepsilon_k)\mu|_A \times \pi_x) \to 0. \tag{3.28}$$

On the one hand, note that

$$\begin{aligned}\mathcal{AW}_r^r(\tilde{\mu}^k \times \pi^k_x, \bar{\mu}^k \times \pi_x) &\leq \int_{X\times X} \left(d_X^r(x, x') + \mathcal{W}_r^r(\pi^k_x, \pi_{x'})\right) \chi^k|_{B\times A}(dx, dx') \\ &\leq \int_{X\times X} \left(d_X^r(x, x') + \mathcal{W}_r^r(\pi^k_x, \pi_{x'})\right) \chi^k(dx, dx') \\ &= \mathcal{AW}_r^r(\pi^k, \pi) \to 0, \quad k \to +\infty.\end{aligned} \tag{3.29}$$

On the other hand, let

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

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

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

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

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

Similarly, we find

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

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

$$\mathcal{AW}_r(\bar{\mu}^k \times \pi_x, \check{\mu}^k \times \pi_x) \underset{k \to +\infty}{\longrightarrow} 0, \tag{3.30}$$

and the triangle inequality with (3.29) and (3.30) yield the assertion, (3.28).

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

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

which is conclusive.

(ii) To show the last assertion, we denote by $\tilde{\nu}^k$ and $\tilde{\nu}$ the second marginal of $\tilde{\mu}^k \times \pi_x^k$ and $\mu|_A \times \pi_x$ respectively. Since we have to remove mass outside of C from $\tilde{\nu}^k$ and the corresponding kernels, we consider the $\mathcal{W}_r$-optimal couplings

$$\mathring{\chi}^k \in \Pi(\tilde{\nu}^k, (1-\varepsilon_k)\tilde{\nu}) \quad \text{and} \quad \hat{\chi}^k \in \Pi(\tilde{\nu}^k(C)\tilde{\nu}, \tilde{\nu}(C)\tilde{\nu}^k|_C).$$

From now on, we denote by $\hat{\mu}^k = \tilde{\mu}^k \frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)} \leq \tilde{\mu}^k$. Using the couplings $\mathring{\chi}^k$ and $\hat{\chi}^k$ we define the kernels

$$\mathring{\pi}_x^k(dz) = \int_Y \mathring{\chi}_y^k(dz)\,\pi_x^k(dy) \quad \text{and} \quad \hat{\pi}_x^k(dt) = \int_Y \hat{\chi}_z^k(dt)\,\mathring{\pi}_x^k(dz).$$

Let then $\hat{\pi}^k := \hat{\mu}^k \times \hat{\pi}_x^k$, whose second marginal denoted $\hat{\nu}^k$ satisfies

$$\hat{\nu}^k(dt) = \int_{x \in X} \int_{y \in Y} \int_{z \in Z} \hat{\mu}^k(dx)\,\pi_x^k(dy)\,\mathring{\chi}_y^k(dz)\,\hat{\chi}_z^k(dt)$$

$$= \frac{\tilde{\nu}(C)(1-\varepsilon_k)}{\tilde{\mu}^k(X)}\tilde{\nu}^k|_C(dt) = \frac{\tilde{\nu}(C)}{\mu(A)}\tilde{\nu}^k|_C(dt).$$

So $\hat{\nu}^k$ is concentrated on C. Moreover, we have $\tilde{\nu}^k|_C \leq \nu^k$, and since ν is concentrated on C, $\tilde{\nu}(C) = \tilde{\nu}(Y) = \mu|_A(X) = \mu(A)$, hence $\hat{\nu}^k = \tilde{\nu}^k|_C \leq \nu^k$. For $k \in \mathbb{N}$, let $\varepsilon'_k \in \mathbb{R}$ be such that

$$1 - \varepsilon'_k = \frac{\tilde{\nu}^k(C)}{\mu(A)}. \tag{3.31}$$

For $k \in \mathbb{N}$, using (3.27), we have $\tilde{\nu}^k(C) \leq \tilde{\mu}^k(X) = (1-\varepsilon_k)\mu(A) \leq \mu(A)$, hence ε'_k is nonnegative. Then it remains to show that

$$\mathcal{AW}_r(\hat{\pi}^k, (1-\varepsilon'_k)\pi|_{A\times Y}) + \int_X \mathcal{W}_r^r(\hat{\pi}^k_x, \pi^k_x)\,\hat{\mu}^k(dx) + \varepsilon'_k \underset{k\to+\infty}{\longrightarrow} 0. \tag{3.32}$$

Since we have

$$\pi^k_x(dy)\,\mathring{\chi}^k_y(dz) \in \Pi(\pi^k_x, \mathring{\pi}^k_x), \quad \mathring{\pi}^k_x(dz)\,\hat{\chi}^k_z(dt) \in \Pi(\mathring{\pi}^k_x, \hat{\pi}^k_x),$$
$$\int_{x\in X} \hat{\mu}^k(dx)\,\pi^k_x(dy)\,\mathring{\chi}^k_y(dz) = \frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)}\mathring{\chi}^k(dy,dz),$$
$$\int_{x\in X} \hat{\mu}^k(dx)\,\mathring{\pi}^k_x(dz)\,\hat{\chi}^k_z(dt) = \frac{(1-\varepsilon_k)}{\tilde{\mu}^k(X)}\hat{\chi}^k(dz,dt),$$

we find plugging the expressions (3.27) and (3.31) that

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

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 (3.28) implies

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

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

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

hence

$$\tilde{\nu}^k|_C(f) \to \tilde{\nu}(f), \quad k \to +\infty. \tag{3.33}$$

Moreover, (3.33) 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

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

On the other hand, by (3.27) and (3.31) we have $\hat{\mu}^k = \frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)}\tilde{\mu}^k = \frac{1-\varepsilon'_k}{1-\varepsilon_k}\tilde{\mu}^k$, hence

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

where the right-hand side vanishes as k goes to $+\infty$ by the first part. Then (3.32) 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 mass.

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

Example 3.5. Let $X = Y = \mathbb{R}$, and $\pi^k = \delta_{(\frac{1}{k},1)}$, $\chi^k = \delta_{(-\frac{1}{k},-1)}$, $k \in \mathbb{N}$. 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 weak adapted topology could still be somehow considered continuous 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 mass and $\tilde{\mu}, \tilde{\mu}^k, \tilde{\nu}, \tilde{\nu}^k \in \mathcal{M}_r(Y)$, $k \in \mathbb{N}$ be with equal mass 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 $k \in \mathbb{N}$

$$\begin{aligned}
&\mathcal{AW}^r_r(\hat{\pi}^k + \tilde{\pi}^k, \hat{\pi} + \tilde{\pi}) \\
&\quad\le \mathcal{AW}^r_r(\hat{\pi}^k, \hat{\pi}) + 2^{r-1}\left(I^r_\varepsilon(\tilde{\mu}) + I^r_\varepsilon(\tilde{\mu}^k) + I^r_\varepsilon(\tilde{\nu}) + I^r_\varepsilon(\tilde{\nu}^k) + 2I^r_\varepsilon(\hat{\nu}) + 2I^r_\varepsilon(\hat{\nu}^k)\right) \\
&\quad\le \mathcal{AW}^r_r(\hat{\pi}^k, \hat{\pi}) + (2^{r-1})^2\left(\mathcal{W}^r_r(\tilde{\mu}^k, \tilde{\mu}) + \mathcal{W}^r_r(\tilde{\nu}^k, \tilde{\nu}) + 2\mathcal{W}^r_r(\hat{\nu}^k, \hat{\nu})\right) \\
&\qquad + 2^{r-1}(1+2^{r-1})I^r_\varepsilon(\mu) + 3 \cdot 2^{r-1}(1+2^{r-1})I^r_\varepsilon(\nu),
\end{aligned} \tag{3.34}$$

where $I^r_\varepsilon(\cdot)$ 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

$$\limsup_{k\to+\infty} \mathcal{AW}^r_r(\hat{\pi}^k + \tilde{\pi}^k, \hat{\pi} + \tilde{\pi}) \le C(I^r_\varepsilon(\mu) + I^r_\varepsilon(\nu)), \tag{3.35}$$

where $C > 0$ depends only on r.

Proof. The second inequality of (3.34) is easily deduced from the first one, (3.2) and the fact that $I^r_\varepsilon(\tilde{\mu}) \le I^r_\varepsilon(\mu)$, $I^r_\varepsilon(\tilde{\nu}) \le I^r_\varepsilon(\nu)$ and $I^r_\varepsilon(\hat{\nu}) \le I^r_\varepsilon(\nu)$.

To see (b), assume for a moment that the first inequality of (3.34) 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 (3.2), we get

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

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

To conclude the proof, it remains to show the first inequality in (3.34). 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

$$\mathcal{AW}_r^r(\hat\pi^k + \tilde\pi^k, \hat\pi + \tilde\pi) \le \int_{X\times X} \left(d_X^r(x,x') + \mathcal{W}_r^r((\hat\pi^k+\tilde\pi^k)_x, (\hat\pi+\tilde\pi)_{x'})\right)\, \rho^k(dx,dx'). \tag{3.36}$$

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

$$(\hat\pi+\tilde\pi)(dx,dx') = \mu(dx)\left(\hat p(x)\,\hat\pi_x(dx') + (1-\hat p(x))\,\tilde\pi_x(dx')\right),$$
$$(\hat\pi^k+\tilde\pi^k)(dx,dx') = \mu^k(dx)\left(\hat p^k(x)\,\hat\pi^k_x(dx') + (1-\hat p^k(x))\,\tilde\pi^k_x(dx')\right),$$

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:

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

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

$$\begin{aligned}
&\mathcal{W}_r^r((\hat\pi^k+\tilde\pi^k)_x, (\hat\pi+\tilde\pi)_{x'}) \\
&\quad\le \mathcal{W}_r^r(\beta^k(x,x')\,\hat\pi^k_x, \beta^k(x,x')\,\hat\pi_{x'}) \\
&\qquad + \mathcal{W}_r^r\left(\alpha^k_+(x,x')\,\hat\pi^k_x + (1-\hat p^k(x))\,\tilde\pi^k_x, \alpha^k_-(x,x')\,\hat\pi_{x'} + (1-\hat p(x'))\,\tilde\pi_{x'}\right) \\
&\quad\le \beta^k(x,x')\mathcal{W}_r^r(\hat\pi^k_x,\hat\pi_{x'}) + 2^{r-1}\left(\alpha^k_+(x,x')\mathcal{W}_r^r(\hat\pi^k_x,\delta_{y_0}) + (1-\hat p^k(x))\mathcal{W}_r^r(\tilde\pi^k_x,\delta_{y_0})\right. \\
&\qquad \left. + \alpha^k_-(x,x')\mathcal{W}_r^r(\hat\pi_{x'},\delta_{y_0}) + (1-\hat p(x'))\mathcal{W}_r^r(\tilde\pi_{x'},\delta_{y_0})\right).
\end{aligned} \tag{3.37}$$

Since $\beta^k(x,x') = \hat p^k(x)\wedge\hat p(x') \le 1$, we deduce from (3.36), (3.37) and $\mathcal{AW}_r$-optimality of $\hat\rho^k$

$$\begin{aligned}
\mathcal{AW}_r^r(\hat\pi^k+\tilde\pi^k, \hat\pi+\tilde\pi) \le \mathcal{AW}_r^r(\hat\pi^k,\hat\pi) &+ \int_{X\times X} d_X(x,x')^r\,\tilde\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \hat p^k(x)\mathcal{W}_r^r(\hat\pi^k_x,\delta_{y_0})\,\tilde\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \hat p(x')\mathcal{W}_r^r(\hat\pi_{x'},\delta_{y_0})\,\tilde\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \alpha^k_+(x,x')\mathcal{W}_r^r(\hat\pi^k_x,\delta_{y_0})\,\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} (1-\hat p^k(x))\mathcal{W}_r^r(\tilde\pi^k_x,\delta_{y_0})\,\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \alpha^k_-(x,x')\mathcal{W}_r^r(\hat\pi_{x'},\delta_{y_0})\,\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} (1-\hat p(x'))\mathcal{W}_r^r(\tilde\pi_{x'},\delta_{y_0})\,\rho^k(dx,dx').
\end{aligned} \tag{3.38}$$

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

$$\int_{X\times X} d_X(x,x')^r\,\tilde\rho^k(dx,dx') \le 2^{r-1}\left(I^r_\varepsilon(\tilde\mu^k) + I^r_\varepsilon(\tilde\mu)\right). \tag{3.39}$$

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

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

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

$$\int_{X\times X} \hat p^k(x)\mathcal{W}_r^r(\hat\pi^k_x,\delta_{y_0})\,\tilde\rho^k(dx,dx') = \int_{X\times X} (1-\hat p^k(x))\mathcal{W}_r^r(\hat\pi^k_x,\delta_{y_0})\,\hat\mu^k(dx) \le I^r_\varepsilon(\hat\nu^k), \tag{3.40}$$

$$\int_{X\times X} \hat{p}(x')\mathcal{W}_r^r(\hat{\pi}_{x'},\delta_{y_0})\,\tilde{\rho}^k(dx,dx') = \int_{X\times X} (1-\hat{p}(x'))\mathcal{W}_r^r(\hat{\pi}_{x'},\delta_{y_0})\,\hat{\mu}(dx') \le I_\varepsilon^r(\hat{\nu}). \tag{3.41}$$

We deduce from (3.38) and (3.39)-(3.41) that it is sufficient to show

$$\int_{X\times X} \alpha_+^k(x,x')\mathcal{W}_r^r(\hat{\pi}_x^k,\delta_{y_0})\,\rho^k(dx,dx') \le I_\varepsilon^r(\hat{\nu}^k), \tag{3.42}$$

$$\int_{X\times X} (1-\hat{p}^k(x))\mathcal{W}_r^r(\tilde{\pi}_x^k,\delta_{y_0})\,\rho^k(dx,dx') \le I_\varepsilon^r(\tilde{\nu}^k), \tag{3.43}$$

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

$$\int_{X\times X} (1-\hat{p}(x'))\mathcal{W}_r^r(\tilde{\pi}_{x'},\delta_{y_0})\,\rho^k(dx,dx') \le I_\varepsilon^r(\tilde{\nu}). \tag{3.45}$$

To see (3.43) and (3.45), note that

$$(1-\hat{p}^k(x))\,\mu^k(dx) = \tilde{\mu}^k(dx) \quad \text{and} \quad (1-\hat{p}(x'))\,\mu(dx') = \tilde{\mu}(dx'). \tag{3.46}$$

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

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

Next, we show (3.42) and (3.44). To this end, denoting with a slight abuse of notation $\rho^k(dx,dx') = \mu^k(dx)\,\rho_x^k(dx') = \mu(dx')\,\rho_{x'}^k(dx)$, we have

$$\alpha_+^k(x,x')\,\rho^k(dx,dx') \le \hat{p}^k(x)\,\rho^k(dx,dx') = \frac{d\hat{\mu}^k}{d\mu^k}(x)\,\mu^k(dx)\,\rho_x^k(dx') = \hat{\mu}^k(dx)\,\rho_x^k(dx'),$$
$$\alpha_-^k(x,x')\,\rho^k(dx,dx') \le \hat{p}(x')\,\rho^k(dx,dx') = \frac{d\hat{\mu}}{d\mu}(x')\,\mu(dx')\,\rho_{x'}^k(dx) = \hat{\mu}(dx')\,\rho_{x'}^k(dx).$$

In particular, the first marginal of $\alpha_+^k(x,x')\,\rho^k(dx,dx')$, denoted here by τ, is dominated by $\hat{\mu}^k$, whereas the second marginal of $\alpha_-^k(x,x')\,\rho^k(dx,dx')$, denoted here by τ', is dominated by $\hat{\mu}$. Concerning the mass of τ and τ', remember (3.46), $\alpha_+^k(x,x') \le 1-\hat{p}(x')$ and $\alpha_-^k(x,x') \le 1-\hat{p}^k(x)$, thus,

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

Using (3.1), we conclude with

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

□

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 iff 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

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

Let us now go back to the general case. 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 Polish, 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

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

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

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

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

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

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

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

Then (3.35) yields

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

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 $\varepsilon \to 0$ according to Lemma 3.1 (b), which concludes the proof. $\square$

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

$$\mu \leq_c \nu \iff \forall f\colon \mathbb{R} \to \mathbb{R} \text{ convex}, \quad \mu(f) \leq \nu(f).$$

The following assertions can be found in [12, Section 4]: 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 $u\colon \mathbb{R} \to \mathbb{R}^+$ 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 suffices (i) and (ii) is then called a potential function. The potential function of μ is denoted by

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

A sequence $(\mu^k)_{k\in\mathbb{N}}$ of finite positive measures with equal mass on the line converges in $\mathcal{W}_1$ to μ iff 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

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

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

In dimension one, for all $m_1 \in \mathbb{R}$, the set of all finite positive measures with mean m_1 is a lattice [33, Proposition 1.6], and even a complete lattice [34]. 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*

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

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

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

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

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

which establishes (4.1) and completes the proof. □

We now provide the key argument to see that it is enough to prove Theorem 2.5 for irreducible pairs of marginals.

Proof of Proposition 2.4. 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 π_n^k on $\mathbb{R}^2$ given by

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

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

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

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

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

$$= \int_{u\in J} \delta_{F_\mu^{-1}(u)}(dx)\, du.$$

Since the monotone rearrangement gives 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 any $n \in N$, π_n^k is dominated by π^k, hence $(\pi_n^k)_{k\in\mathbb{N}}$ 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 $\left(\int_\mathbb{R} |y|\, \nu^k(dy)\right)_{k\in\mathbb{N}}$ are convergent and therefore 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 and the convergence of the first moments, we deduce that the sequence $(\pi_n^k)_{k\in\mathbb{N}}$ is relatively compact in $\mathcal{W}_1$. There are subsequences $(\pi_n^{k_j})_{j\in\mathbb{N}}$ converging in $\mathcal{W}_1$ to a measure $\tilde{\pi}_n$, where $\tilde{\pi}_n \le \pi$. The first marginal of $\tilde{\pi}_n$ coincides with μ_n due to continuity of the projection, thus,

$$\tilde{\pi}_n \le \pi|_{I_n\times\mathbb{R}} = \pi_n.$$

As $\tilde{\pi}_n(\mathbb{R}\times\mathbb{R}) = \mu_n(I_n) = \pi_n(\mathbb{R}\times\mathbb{R})$, there must hold equality, i.e., $\tilde{\pi}_n = \pi_n$. In particular, $(\pi_n^k)_{k\in\mathbb{N}}$ has only a single accumulation point and is therefore $\mathcal{W}_1$-convergent. Due to continuity of the projection, $(\nu_n^k)_{k\in\mathbb{N}}$ converges in $\mathcal{W}_1$ to ν_n for each $n \in N$. Analogously, we find that $(\upsilon^k)_{k\in\mathbb{N}}$ converges to η. □

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

Lemma 4.2. *Let $r \ge 1$ 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 $\alpha, \beta \in \mathbb{R}_+$,*

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

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

Remark 4.3. Let $\eta_0, \eta_1 \in \mathcal{M}_r(\mathbb{R}^d)$ be with equal mass and γ be an optimal coupling for $\mathcal{W}_r(\eta_0,\eta_1)$. For all $t \in [0,1]$, let η_t be the image of γ by $(y,z) \mapsto (1-t)y + tz$. It is well known that the curve $[0,1] \ni t \mapsto \eta_t$ is a constant speed geodesic for W_r connecting η_0 to η_1. Moreover, for all $0 \le s \le t \le 1$, the image of γ by $((1-s)y+sz, (1-t)y+tz)$ is an optimal coupling for $\mathcal{W}_r(\eta_s,\eta_t)$.

In particular for $\alpha < \beta \in \mathbb{R}_+$, $\eta_0 = \delta_{m_1}$ and $\eta_1 = \mu^\beta$, we have $\gamma(dy,dz) = \delta_x(dy)\, \mu^\beta(dz)$ and for $t = \frac{\alpha}{\beta}$, $\eta_t = \mu^\alpha$. Therefore, the image of γ by $(y,z) \mapsto ((1-t)y+tz, z)$, that is the image of μ by $y \mapsto (\alpha(y-x)+x, \beta(y-x)+x)$, is an optimal coupling for $\mathcal{W}_r(\mu^\alpha,\mu^\beta)$, hence (4.4) is proved.

Since the proof is brief, we add it here for the sake of completeness.

Proof of Lemma 4.2. Let $\alpha \le \beta \in \mathbb{R}_+$. We have by the triangle inequality for $\mathcal{W}_r$ that

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

Thus,

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

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

Since the image of μ by $x \mapsto (\alpha(x - m_1) + m_1, \beta(y - m_1) + m_1)$ 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 any convex function f on $\mathbb{R}^d$. Due to 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 4.4. *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*

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

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), we show that 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 $R_0 = |m_1| + \sup_{y \in [-M, M]} u_p(y)$ 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, $\delta_{m_1} \leq_c u_{\eta^R}$, hence we also have $u_p(y) \geq |y - m_1|$ for all $y \in \mathbb{R}$. Let $y \in \mathbb{R}$. If $|y| \leq M$, then

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

Else if $|y| > M$ and $u_{\eta^R}(y) < u_p(y)$, 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 2.5. Before entering its technical proof, we argue that it is sufficient to consider the case $r = 1$ and that we can consider without loss of generality 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 coupled with convergence of the r-moments. To see the latter, we find, when equipping $X \times \mathcal{P}_r(Y)$ with the product metric $((x,p),(x',p')) \mapsto (d_X^r(x,x') + \mathcal{W}_r^r(p,p'))^{1/r}$,

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

A direct consequence is the following lemma, according to which proving Theorem 2.5 for $r = 1$ is sufficient.

Lemma 5.1. *In the setting of Theorem 2.5, 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 2.4 is the key ingredient to show that it is enough to prove Theorem 2.5 when (μ, ν) is irreducible.

Lemma 5.2. *If the conclusion of Theorem 2.5 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 2.5, 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 2.4, 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

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

For $k \in \mathbb{N}$, let $\chi^k \in \Pi_M(\eta^k, \upsilon^k)$ be any 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

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

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 $(\pi_n^k)_{k\in\mathbb{N}}$ of martingale couplings between μ_n^k and ν_n^k, $k \in \mathbb{N}$, 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 (2.7). By Lemma 3.7, we have for all $p \in N$ that

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

Moreover, the respective marginals of $\chi^k + \sum_{n\in N} \pi_n^k$, 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

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

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, $(\varepsilon_p)_{p\in N}$ is a null sequence, thus Lemma 3.1 (b) reveals that the right-hand side vanishes as p goes to $\sup N$. This proves that $\pi^k = \chi^k + \sum_{n\in N} \pi_n^k \in \Pi_M(\mu^k, \nu^k)$ converges in $\mathcal{AW}_1$ to $\pi = \chi + \sum_{n\in N} \pi^k \in \Pi_M(\mu, \nu)$. □

Proof of Theorem 2.5. Step 1. Due to Lemma 5.1 and Lemma 5.2, we may suppose without loss of generality that $r = 1$ and (μ, ν) is irreducible with component $I = (l, r)$, $l \in \mathbb{R} \cup \{-\infty\}$, $r \in \mathbb{R} \cup \{+\infty\}$. 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 two compact subsets of I on both sides of K.

Fix $\varepsilon \in (0, \frac{1}{2})$. 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 4.4. By Lemma 4.4 (a) we have for all $x \in \mathbb{R}$ that $\pi_x^R \le_c \pi_x$. Therefore,

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

where the right-hand side is a μ-integrable function of x. By Lemma 4.4 (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

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

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

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

Since (μ, ν) is irreducible on I, ν assigns positive mass to any neighbourhood of the endpoints l and r 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, $\overline{I} = [l, r] \subset \mathbb{R}$. Then $[l, \tilde{a})$ and $(\tilde{b}, r]$ are relatively open on the common support $\overline{I}$ of the probability measures ν^R and ν, so Portmanteau's theorem yields

$$\liminf_{R\to+\infty} \nu^R([l, \tilde{a})) \ge \nu([l, \tilde{a})) > 0, \quad \liminf_{R\to+\infty} \nu^R((\tilde{b}, r]) \ge \nu((\tilde{b}, r]) > 0.$$

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

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

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

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

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 $\frac{2R-a-\tilde{a}}{2R-2\tilde{a}}, \frac{b+\tilde{b}+2R}{\tilde{2}b+2R} \in (0,1)$, so we can choose $\alpha \in (0,1)$ such that

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

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

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

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

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

we have that

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

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

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

We deduce that

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

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, 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.

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

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

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

$$\tilde L_- = \left(\tilde l, \frac{a+\tilde a}{2}\right), \quad \tilde L_+ = \left(\frac{b+\tilde b}{2}, \tilde r\right) \quad \text{and} \quad \tilde K = \left(\frac{3a+\tilde a}{4}, \frac{3b+\tilde b}{4}\right), \tag{5.7}$$

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 $J = [\tilde l, \tilde r]$, Figure 2 summarises the construction.

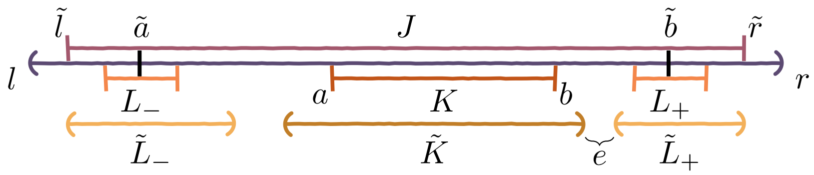

Figure 2: Points and intervals involved in the proof. The boundaries of the closed intervals are vertical bars and those of the open intervals are parenthesis.

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 $\varepsilon_k \geq 0$ such that

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

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

$$\int_{\mathbb{R}} y\, \hat{\pi}^k_x(dy) \le x \implies c^k_-(x) = 0, \quad \int_{\mathbb{R}} y\, \hat{\pi}^k_x(dy) \ge x \implies c^k_+(x) = 0. \tag{5.9}$$

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

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

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

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

$$\begin{aligned}
c^k_-(x) &= \frac{\int_{\mathbb{R}} y\, \hat{\pi}^k_x(dy) - x}{\int_{\mathbb{R}} (x-y)\, \nu^k_-(dy)} \vee 0 \in \left[0, \frac{\left|x - \int_{\mathbb{R}} y\, \hat{\pi}^k_x(dy)\right|}{e\nu^k_-(\mathbb{R})}\right], \\
c^k_+(x) &= \frac{x - \int_{\mathbb{R}} y\, \hat{\pi}^k_x(dy)}{\int_{\mathbb{R}} (y-x)\, \nu^k_+(dy)} \vee 0 \in \left[0, \frac{\left|x - \int_{\mathbb{R}} y\, \hat{\pi}^k_x(dy)\right|}{e\nu^k_+(\mathbb{R})}\right], \\
d^k(x) = 1 + c^k_-(x)\nu^k_-(\mathbb{R}) &+ c^k_+(x)\nu^k_+(\mathbb{R}) \in \left[1, 1 + \frac{\left|x - \int_{\mathbb{R}} y\, \hat{\pi}^k_x(dy)\right|}{e}\right].
\end{aligned}$$

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

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

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

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

where we used Remark 2.2 in the last inequality. The triangle inequality and (5.8) then yield

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

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

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

where we used Remark 2.2 again for the last inequality and the fact that $\nu^k_-(\mathbb{R}) \wedge \nu^k_+(\mathbb{R}) \ge \delta$ 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

$$(1-2\varepsilon)\tilde{\nu}^k \le (1-2\varepsilon)\hat{\nu}^k + (1-2\varepsilon)(\nu^k_- + \nu^k_+) \int_{\mathbb{R}} \frac{c^k_-(x) + c^k_+(x)}{d^k(x)}\, \hat{\mu}^k(dx) \le (1-\varepsilon)\nu^{R,\alpha,k}.$$

Step 3. In this step, we complement the martingale coupling $(1-2\varepsilon)\tilde{\pi}^k$ to a martingale coupling with marginals μ^k and $\varepsilon\nu^k + (1-\varepsilon)\nu^{R,\alpha,k}$ 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

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

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

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

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

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

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

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

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,

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

By Strassen's theorem [43], 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 $(\mu, \varepsilon\nu+(1-\varepsilon)\nu^{R,\alpha})$ as k goes to $+\infty$. We have according to (5.12) that

$$(1-2\varepsilon)\mathcal{AW}_1(\tilde{\pi}^k, (1-\varepsilon_k)\pi^{R,\alpha}) \to 0, \quad k \to \infty.$$

Since $\mu(K) \ge 1-\varepsilon$, $(1-2\varepsilon)(1-\varepsilon) > 1-3\varepsilon$, and ε_k vanishes as k goes to $+\infty$, we have for $k \in \mathbb{N}$ large enough

$$\begin{aligned}
\eta^k(\mathbb{R}^2) &= \left(\varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha} - (1-2\varepsilon)(1-\varepsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}\right)(\mathbb{R}^2) \\
&= 1 - (1-2\varepsilon)(1-\varepsilon_k)\mu(K) \\
&\le 3\varepsilon,
\end{aligned}$$

hence (3.35) writes

$$\limsup_k \mathcal{AW}_1(\overline{\pi}^k, \varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}) \le C(I_{3\varepsilon}(\mu) + I_{3\varepsilon}(\varepsilon\nu + (1-\varepsilon)\nu^{R,\alpha})),$$

where C depends only on the exponent $r = 1$. Since $\nu^{R,\alpha} \le_c \nu$, then $\varepsilon\nu + (1-\varepsilon)\nu^{R,\alpha} \le_c \nu$, so using Lemma 3.1 (e), the triangle inequality and (5.4), we get

$$\begin{aligned}
\limsup_k \mathcal{AW}_1(\overline{\pi}^k, \pi) &\le \limsup_k \left(\mathcal{AW}_1(\overline{\pi}^k, \varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}) + \mathcal{AW}_1(\varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}, \pi)\right) \\
&\le C(I_{3\varepsilon}(\mu) + I_{3\varepsilon}(\nu)) + \varepsilon.
\end{aligned}$$

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 (3.26)) to find a null sequence $(\tilde{\varepsilon}_k)_{k\in\mathbb{N}}$, two 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

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

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

$$\mathcal{W}_1(\tilde{\varepsilon}_k \nu^k + (1-\tilde{\varepsilon}_k)\nu^{R_k,\alpha_k,k}, \nu^k) \leq \mathcal{W}_1(\tilde{\varepsilon}_k \nu^k + (1-\tilde{\varepsilon}_k)\nu^{R_k,\alpha_k,k}, \nu) + \mathcal{W}_1(\nu, \nu^k) \to 0, \quad k \to +\infty.$$

Remember that $\nu^{R_k,\alpha_k,k} \leq_c \nu^k$, hence $\tilde{\varepsilon}_k \nu^k + (1-\tilde{\varepsilon}_k)\nu^{R_k,\alpha_k,k} \leq_c \nu^k$. Then by [30, Theorem 2.11], there exist martingale couplings $M^k \in \Pi_M(\tilde{\varepsilon}_k \nu^k + (1-\tilde{\varepsilon}_k)\nu^{R_k,\alpha_k,k}, \nu^k)$, $k \in \mathbb{N}$ such that

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

Let then π^k be the joint distribution of the first and third marginals of $\mathring{\pi}^k \dot{\otimes} M^k$, that is

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

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

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

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. $\square$

References

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

[2] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds. *International Journal of Theoretical and Applied Finance*, 22(3), 2019.

[3] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of probability measures in the convex order by Wasserstein projection. *Annales de l'Institut Henri Poincaré B, Probability and Statistics*, 56(3):1706–1729, 2020.

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

[5] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. Adapted wasserstein distances and stability in mathematical finance. *Finance and Stochastics*, 24(3):601–632, 2020.

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

[7] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and J. Wiesel. Estimating processes in adapted wasserstein distance. *arXiv e-prints:2002.07261*, February 2020.

[8] J. Backhoff Veraguas, M. Beiglböck, and G. Pammer. Existence, Duality, and Cyclical monotonicity for weak transport costs. *arXiv e-prints:1809.05893*, September 2018.

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

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

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

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

[13] M. Beiglböck, T. Lim, and J. Obłój. Dual attainment for the martingale transport problem. *Bernoulli*, 25(3):1640–1658, 2019.

[14] M. Beiglböck, M. Nutz, and N. Touzi. Complete Duality for Martingale Optimal Transport on the Line. *Annals of Probability*, 45(5):3038–3074, 2017.

[15] J. Bion-Nadal and D. Talay. On a wasserstein-type distance between solutions to stochastic differential equations. *Annals of Applied Probability*, 29(3):1609–1639, 2019.

[16] H. De March. Entropic approximation for multi-dimensional martingale optimal transport. *arXiv e-prints:1812.11104*, December 2018.

[17] H. De March. Local structure of multi-dimensional martingale optimal transport. *arXiv e-prints:1805.09469*, November 2018.

[18] H. De March. Quasi-sure duality for multi-dimensional martingale optimal transport. *arXiv e-prints:1805.01757*, May 2018.

[19] H. De March and N. Touzi. Irreducible convex paving for decomposition of multi-dimensional martingale transport plans. *Annals of Probability*, 47(3):1726–1774, 2019.

[20] J. Dugundji. An extension of tietze's theorem. *Pacific Journal of Mathematics*, 1(3):353–367, 1951.

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

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

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

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

[25] G. Guo and J. Obłój. Computational methods for martingale optimal transport problems. *Annals of Applied Probability*, 29(6):3311–3347, 2019.

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

[27] P. Henry-Labordère. (Martingale) optimal transport and anomaly detection with neural networks: A primal-dual algorithm. *arXiv preprint arXiv:1904.04546*, 2019.

[28] P. Henry-Labordère, X. Tan, and N. Touzi. An explicit version of the one-dimensional brenier's theorem with full marginals constraint. *Stochastic Processes and their Applications*, 126(9):2800–2834, 2016.

[29] P. Henry-Labordère and N. Touzi. An explicit martingale version of the one-dimensional Brenier theorem. *Finance and Stochastics*, 20(3):635–668, 2016.

[30] B. Jourdain and W. Margheriti. A new family of one dimensional martingale couplings. *arXiv e-prints:1808.01390*, March 2019.

[31] N. Juillet. Stability of the shadow projection and the left-curtain coupling. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 52(4):1823–1843, 2016.

[32] L. V. Kantorovich. On the translocation of masses. *Doklady Akademii Nauk SSSR*, 37(7-8):199–201, 1942.

[33] R. P. Kertz and U. Rösler. Stochastic and convex orders and lattices of probability measures, with a martingale interpretation. *Israel Journal of Mathematics*, 77(1-2):129–164, 1992.

[34] R. P. Kertz and U. Rösler. Complete lattices of probability measures with applications to martingale theory. *Lecture Notes-Monograph Series*, 35:153–177, 2000.

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

[36] R. Lassalle. Causal transference plans and their Monge-Kantorovich problems. *Stochastic Analysis and Applications*, 36(3):452–484, 2018.

[37] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'académie Royale des Sciences de Paris*, 1781.

[38] G. C. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM Journal on Optimization*, 22(1):1–23, 2012.

[39] G. C. Pflug and A. Pichler. *Multistage stochastic optimization.* Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2014.

[40] G. C. Pflug and A. Pichler. Dynamic generation of scenario trees. *Computational Optimization and Applications*, 62(3):641–668, 2015.

[41] G. C. Pflug and A. Pichler. From empirical observations to tree models for stochastic optimization: convergence properties. *SIAM Journal on Optimization*, 26(3):1715–1740, 2016.

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

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

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

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

[46] J. Wiesel. Continuity of the martingale optimal transport problem on the real line. *arXiv e-prints:1905.04574*, January 2020.

Monotonicity and stability of the Weak Martingale Optimal Transport problem

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

October 24, 2020

Abstract

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

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

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

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

1 Introduction and motivations

1.1 The Weak Optimal Transport problem

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

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

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

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

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

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

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

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

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

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

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

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

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

It is well known that the latter is equivalent to

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

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

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

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

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

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

It will prove important to observe that

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

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

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

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

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

1.2 The Martingale Weak Optimal Transport problem

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

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

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

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

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

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

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

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

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

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

1.3 Outline

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

2 Main results

2.1 An extension of the weak and adapted topologies

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.2 Stability

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

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

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

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

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

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

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

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

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

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

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

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

(A) C is continuous.

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

Then there holds

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

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

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

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

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

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

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

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

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

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

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

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

(A) C is continuous.

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

Then there holds

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

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

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

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

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

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

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

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

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

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

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

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

(b) π^* is the joint probability distribution of (M^*_0, M^*_1), and conversely

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

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

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

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

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

$$\mathcal{AW}^r_r\left(\mathcal{L}(M^k_0, (M^k_t)_{t\in[0,1]}), \mathcal{L}(M^*_0, (M^*_t)_{t\in[0,1]})\right) \le \left(\frac{r}{r-1}\right)^r \mathcal{AW}^r_r\left(\mathcal{L}(M^k_0, M^k_1), \mathcal{L}(M^*_0, M^*_1)\right),$$

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

2.3 Stability of the superreplication bound for VIX futures

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.4 Martingale monotonicity

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

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

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

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

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

for every competitor α' of α.

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

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

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

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

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

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

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

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

hence π is optimal.

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

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

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

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

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

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

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

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

3 Stability

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

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

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

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

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

We now show that the convergence

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The proof of Corollary 2.8 relies on the following Lemma.

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

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

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

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

□

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

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

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

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

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

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

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

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

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

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

We deduce that

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

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

4 Martingale monotonicity

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

proving optimality of π. $\square$

5 Appendix

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and their topologies are equal.

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

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

and their topologies are equal.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Conversely, the following are equivalent:

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

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

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

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

(a) The map

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

is lower semicontinuous.

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

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

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

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

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

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

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

and thus (5.13). □

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

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

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

are continuous.

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

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

which shows by (5.6) continuity of I.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5.2 A Pormanteau-like theorem for Carathéodory maps

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\mathcal{W}_2^2(\pi_x^*, \gamma) + \mathcal{W}_2^2(\pi_{x'}^*, \gamma) \le \mathcal{W}_2^2(p, \gamma) + \mathcal{W}_2^2(p', \gamma). \tag{5.25}$$

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

$$\begin{aligned}
\mathcal{W}_2^2(\pi_x^*, \gamma) + \mathcal{W}_2^2(\pi_{x'}^*, \gamma) &\le \mathcal{W}_2^2\left(\frac{1}{2}(\pi_x^* + p), \gamma\right) + \mathcal{W}_2^2\left(\frac{1}{2}(\pi_{x'}^* + p'), \gamma\right) \\
&< \frac{1}{2}\mathcal{W}_2^2(\pi_x^*, \gamma) + \frac{1}{2}\mathcal{W}_2^2(p, \gamma) + \frac{1}{2}\mathcal{W}_2^2(\pi_{x'}^*, \gamma) + \frac{1}{2}\mathcal{W}_2^2(p', \gamma),
\end{aligned}$$

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

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

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

Suppose that

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

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

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

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

$$0 < F_{\pi_{x'}^*}(x-) \le F_{\pi_{x'}^*}(x) < 1. \tag{5.29}$$

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

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

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

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

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

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

hence

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

Combining these inequalities yields

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

which proves (5.27) and completes the proof.

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

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

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

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

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

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

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

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

and, in particular,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Thus,

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

Moreover there holds

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

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

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

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

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

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

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

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

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

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

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

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

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

To conclude, it remains to show that

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

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

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gudmund Pammer

Curriculum Vitae

✉ *gudmund.pammer@univie.ac.at*

Scientific Career

2017–current **Project assistant**.
In the group of Prof. Mathias Beiglböck and Prof. Walter Schachermayer
2019–current: University of Vienna, Faculty of Mathematics
2017–2019: Technical University of Vienna, Faculty of Mathematics
Research topics:
- Mathematical finance
- Optimal transport, martingale optimal transport, weak optimal transport

2015 – 2017 **Undergraduate research assistant**.
In the group of Prof. Clemens Heitzinger
Technical University of Vienna, Faculty of Mathematics
Project title: START project no. Y660PDE Models for Nanotechnology
Research topics:
- Stochastic Galerkin methods, stochastic collocation methods
- High-dimensional integration of multisymmetric functions
- Bayesian estimation

Education

2017–current **PhD**, *Mathematics*, PhD program: Dissipation and Dispersion in Nonlinear Partial Differential equations, University of Vienna.
Thesis: *Optimal Transport and Mathematical Finance through the lens of the Adapted Weak Topology*
Supervisors: Prof. Mathias Beiglböck and Prof. Julio Backhoff-Veraguas

2015–2017 **Master of Science**, *Mathematics (with honors)*, Technical University of Vienna.
Thesis: *Distribution constrained optimal stopping problems in discrete time*
Supervisor: Prof. Mathias Beiglböck

2012–2015 **Bachelor of Science**, *Mathematics (with honors)*, Technical University of Vienna.
Thesis: *Existence, uniqueness, and a comparison of two non-intrusive methods for the stochastic nonlinear Poisson–Boltzmann equation*
Supervisor: Prof. Clemens Heitzinger

Honors & Awards

2017 **Leistungsstipendium**, *Vienna University of Technology*.
2016 **Stiftungsstipendium**, *Vienna University of Technology*.
2011 **Young Scientist Award**, *Johannes Kepler University Linz*, 2. Price.

Publications

[10] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. The weak monotone rearrangement on the line. *Electronic Communications in Probability*, 2020.

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

[8] C. Heitzinger, G. Pammer, and S. Rigger. Cubature formulas for multisymmteric functions and applications to stochastic partial differential equations. *SIAM/ASA Journal on Uncertainty Quantification*, 2019.

[7] C. Heitzinger, G. Pammer, and S. Rigger. Existence, Uniqueness, and a Comparison of Nonintrusive Methods for the Stochastic Nonlinear Poisson–Boltzmann Equation. *SIAM/ASA Journal on Uncertainty Quantification*, 2018.

Preprints

[6] J. Backhoff-Veraguas and G. Pammer. Stability of martingale optimal transport and weak optimal transport. *Annals of Applied Probability*, under revision, 2020.

[5] J. Backhoff-Veraguas and G. Pammer. Applications of weak transport theory. *arXiv e-prints*, arXiv:2003.05338, 2019.

[4] B. Acciaio, M. Beiglböck, and G. Pammer. Model-independence in a fixed-income market and weak transport for non-convex costs. *https://homepage.univie.ac.at/gudmund.pammer/publications.html*, 2020.

[3] K. Johnson, M. Beiglböck, M. Eder, A. Grass, J. Hermisson, G. Pammer, J. Polechova, D. Toneian, B. Wölfl. Estimating the reproduction number in the presence of superspreading. *https://homepage.univie.ac.at/gudmund.pammer/publications.html*, 2020.

[2] M. Beiglböck, B. Jourdain, W. Margheriti and G. Pammer. Approximation of martingale couplings in the adapted weak topology. *https://homepage.univie.ac.at/gudmund.pammer/publications.html*, 2020.

[1] M. Beiglböck, B. Jourdain, W. Margheriti and G. Pammer. Monotonicity and stability of the weak martingale optimal transport problem. *https://homepage.univie.ac.at/gudmund.pammer/publications.html*, 2020.

In preparation

W. Margheriti and G. Pammer. On the stability of extremal models in mathematical finance.

D. Bartl and G. Pammer. The Wasserstein space of filtered processes.

G. Pammer. Stability of multiperiod martingale optimal transport on the line.

M. Beiglböck, G. Pammer, and A. Posch. The Knothe-Rosenblatt distance metrizes the adapted weak topology.

M. Beiglböck, G. Pammer, and S. Schrott. Stochastic process definition of the adapted Wasserstein distance.

Conferences & Talks

2020 PDE Afternoon, March 4, Vienna (Austria).

2019 Vienna Zurich Symposium for young researchers in Financial Mathematics and related fields, November 27 – 29, Vienna (Austria).

2019 PDE Afternoon, October 9, Vienna (Austria).

2019 Eight or Nine Talks on Contemporary Optimal Transport Problems, July 4 – 5, Strasbourg (France).

2019 ESI: Optimal Transport summer school, May 6 – 10, Vienna (Austria).

2019 PDE Afternoon, March 6, Vienna (Austria).

2018 PDE Afternoon, October 3, Vienna (Austria).

2018 Robust Techniques in Quantitative Finance, September 3 – 7, Oxford (UK).

2018 SIAM Conference on Uncertainty Quantification, April 16 – 19, California (USA).

2018 PDE Afternoon, March 6, Vienna (Austria).

2017 Austrian Numerical Analysis Day, May 4 – 5, Salzburg (Austria).

2016 SIAM Conference on Uncertainty Quantification, April 5 – 8, Lausanne (Switzerland).

Languages

Fluent: German, English

Intermediate: French