GEOMETRY OF VECTORIAL MARTINGALE OPTIMAL TRANSPORTATIONS AND DUALITY

TONGSEOK LIM

ABSTRACT. The theory of *Optimal Transport* (OT) and *Martingale Optimal Transport* (MOT) were inspired by problems in economics and finance and have flourished over the past decades, making significant advances in theory and practice. MOT considers the problem of pricing and hedging of a financial instrument, referred to as an option, assuming its payoff depends on a single asset price. In this paper we introduce *Vectorial Martingale Optimal Transport* (VMOT) problem, which considers the more general and realistic situation in which the option payoff depends on multiple asset prices. We address this problem of pricing and hedging given market information – described by vectorial marginal distributions of underlying asset prices – which is an intimately relevant setup in the robust financial framework.

We establish that the VMOT problem, as an infinite-dimensional linear programming, admits an optimizer for its dual program. Such existence result of dual optimizers is significant for several reasons: the dual optimizers describe how a person who is liable for an option payoff can formulate optimal hedging portfolios, and more importantly, they can provide crucial information on the geometry of primal optimizers, i.e. the VMOTs. As an illustration, we show that multiple martingales given marginals must exhibit an extremal conditional correlation structure whenever they jointly optimize the expectation of distance-type cost functions.

Keywords: Optimal transport, Martingale, Duality, Dual attainment, Extremal correlation structure, Infinite-dimensional linear programming

MSC2010 Classification: 90Bxx, 90Cxx, 49Jxx, 49Kxx, 60Dxx, 60Gxx

---

1. Introduction

The theory of *Optimal Transport* (OT) and its probabilistic version *Martingale Optimal Transport* (MOT) were inspired by problems in economics and finance and have flourished over the past decades, making significant advances in theory and practice.

In this paper we introduce the *vectorial Martingale Optimal Transport* (VMOT) problem, which is an intimately relevant setup in view of the robust financial framework dealing with many asset prices.

To further elaborate this point, let us consider a stochastic process $(X_{t,i})_{t \geq 0}$ of e.g. company i's stock price, where $i \in [d] := \{1, 2, ..., d\}$. Under the risk neutral probability we may assume $X_t := \{(X_{t,i})\}_{i \in [d]} \in \mathbb{R}^d$ is a vector-valued martingale. Now in the robust financial framework we shall not assume that the joint probability law of $(X_t)_{t \geq 0}$ is known, since we are unable to determine this joint law from the market information. However, by a standard argument by Breeden and Litzenberger [9], we shall assume that the distribution of each price at each fixed maturity $t_0 > 0$, i.e. $\mathrm{Law}(X_{t_0,i})$, can be observed from the market.

From now on let us focus on two fixed maturity times $0 < t_1 < t_2$, and denote $X_i := X_{t_1,i}$, $Y_i := X_{t_2,i}$, $X = (X_1, ..., X_d)$, $Y = (Y_1, ..., Y_d)$, and assume that (X, Y) is a (one-step, $\mathbb{R}^d$-valued) martingale: $\mathbb{E}[Y|X] = X$. According to the above consideration we do not assume $\mathrm{Law}(X, Y) \in \mathcal{P}(\mathbb{R}^{2d})$ is known, but only the $2d$-number of marginal distributions $\mu_i := \mathrm{Law}(X_i), \nu_i := \mathrm{Law}(Y_i)$ are known and fixed. This naturally leads us to consider the following VMOT problem: Let $\mu_i, \nu_i \in \mathcal{P}_1(\mathbb{R})$, where

$$\mathcal{P}_j(\mathbb{R}^k) := \{\mu \mid \mu \text{ is a probability measure on } \mathbb{R}^k \text{ with } \int |x|^j d\mu(x) < \infty\}.$$

Let us write $\vec{\mu} = (\mu_1, ..., \mu_d)$, $\vec{\nu} = (\nu_1, ..., \nu_d)$. We consider the following space of *Vectorial Martingale Transportations* from $\vec{\mu}$ to $\vec{\nu}$

$$(1.1) \quad \mathrm{VMT}(\vec{\mu}, \vec{\nu}) := \{\pi \in \mathcal{P}(\mathbb{R}^{2d}) \mid \pi = \mathrm{Law}(X, Y), \, \mathbb{E}[Y|X] = X,$$
$$X_i \sim \mu_i, \, Y_i \sim \nu_i \quad \forall \, i \in [d]\}.$$

Let $c : \mathbb{R}^{2d} \to \mathbb{R}$ be a (cost) function. We define the VMOT problem as

$$(1.2) \quad \max / \text{minimize} \ \ \mathbb{E}_\pi[c(X, Y)] \ \ \text{over} \ \ \pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu}).$$

The function c has a natural interpretation in finance as an exotic option whose payoff is determined by the d-number of asset prices (X, Y) at the terminal maturity. $\mathbb{E}_\pi[c(X, Y)]$ shall be considered as a fair price of the option c in the risk-neutral world governed by the law of asset prices π, but since π would not be observed from the financial market, we need to consider all possible laws $\mathrm{VMT}(\vec{\mu}, \vec{\nu})$ given the marginal

information $\vec{\mu}, \vec{\nu}$. Under this information, the max / min value in (1.2) can be interpreted as the upper / lower price bound for c respectively.

Note that without the martingale constraint $\mathbb{E}[Y|X] = X$, (1.2) would be the usual multi-marginal optimal transport problem. But (1.2) still belongs to the class of *infinite-dimensional linear programming*, since the martingale constraint is linear. Now unlike the optimal transport case, the problem (1.2) will be feasible (i.e. $\mathrm{VMT}(\vec{\mu}, \vec{\nu}) \neq \emptyset$) if only if every pair of marginals μ_i, ν_i is in convex order, defined by

$$\mu_i \leq_c \nu_i \ \text{ iff } \ \mu_i, \nu_i \in \mathcal{P}_1 \ \text{ and } \ \int f d\mu_i \leq \int f d\nu_i \ \text{ for every convex } f,$$

as shown by Strassen [54]. Thus we will always assume $\mu_i \leq_c \nu_i$ for all $i \in [d]$ in VMOT problem. And, because we will need to calculate means and potential functions of the measures, we will make the following assumption throughout the paper (unless otherwise specified):

Assumption. All measures appearing in this paper are assumed to have finite first moments.

Our first main result investigates the extremal correlation behavior of the terminal prices Y conditional on X, when the martingales (X, Y) jointly max/minimize a distance-type cost function given vectorial marginals. More precisely, we establish the following theorem.

Theorem 1.1. *Assume all μ_i are absolutely continuous with respect to Lebesgue measure, denoted as $\mu_i \ll \mathcal{L}$. If π solves the maximization problem in (1.2) with the Euclidean distance cost $c(x, y) = |x - y|$, then given X, the conditional distribution of Y lies in the set of extreme points of a convex set $C_X \subseteq \mathbb{R}^d$ depending on X, π–almost surely.*

Here we say that a random variable Y lies in A if $Y \in A$ a.s., and a distribution ξ lies in A if ξ has its full mass in A. We shall also prove an analogous result for the minimization problem in (1.2), and we also note that Theorem 1.1 still holds for more general class of strictly convex costs $c(x, y) = ||x - y||$. See Theorem 2.1 for more details. We further note that the statement of extremal correlation was firstly given and studied by [32] in the context of MOT and two n-dimensional marginals.

Recall that the VMOT problem belongs to the class of infinite-dimensional linear programming, which implies that the problem has its *dual programming* problem. For the (primal) minimization problem in (1.2), its dual problem is given by

$$(1.3) \qquad \sup_{(\varphi_i, \psi_i, h_i) \in \Psi} \sum_{i=1}^{d} \Big(\int \varphi_i d\mu_i + \int \psi_i d\nu_i \Big)$$

where Ψ consists of triplets $\varphi_i, \psi_i : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ and $h_i : \mathbb{R}^d \to \mathbb{R}$ such that $\varphi_i \in L^1(\mu_i)$, $\psi_i \in L^1(\nu_i)$, h_i is bounded for every $i \in [d]$, and

$$(1.4) \quad \sum_{i=1}^{d} \left(\varphi_i(x_i) + \psi_i(y_i) + h_i(x)(y_i - x_i) \right) \leq c(x, y) \quad \forall (x, y) \in \mathbb{R}^{2d}$$

where $x = (x_1, ., , , x_d)$, $y = (y_1, ..., y_d)$. Analogous dual problem for the maximization in (1.2) is easily obtained by changing sign of c to $-c$.

The dual problem also has an important interpretation in finance. Suppose a financial firm is liable for an option c it has sold so that the firm has to pay $c(X, Y)$ at the terminal maturity t_2. To hedge its risk, the company may consider buying European options φ_i, ψ_j where each payoff relies only on a single asset at a fixed time. In addition, the firm may consider holding h_i number of shares of the ith asset held between the t_1 and t_2 maturity, such that its payoff at t_2 is $h_i(X) \cdot (Y_i - X_i)$. Notice h_i is a function of the past prices of all assets $\{X_i\}_{i=1}^{d}$. Thus the left hand side of (1.4) yields the overall payoff of the hedging portfolio $(\varphi_i, \psi_i, h_i)_{i=1}^{d}$, and the inequality (1.4) imposes that the position must subhedge the liability for all scenarios of (X, Y). Now notice the dual of the maximization in (1.2) becomes an optimal superhedging problem.

Now the celebrated *duality* result asserts, under a mild assumption on c, that the primal and dual optimal values coincide (see e.g. [27, 58])

$$(1.5) \qquad P(c) := \inf_{\pi \in \text{VMT}(\mu, \nu)} \mathbb{E}_\pi [c(X, Y)]$$

$$= \sup_{(\varphi_i, \psi_i, h_i) \in \Psi} \sum_{i=1}^{d} \left(\int \varphi_i d\mu_i + \int \psi_i d\nu_i \right) =: D(c).$$

Still a natural question is whether a dual optimizer exists, that is, a "tightly" sub/superhedging portfolio can be constructed. In the past decades, researchers working in OT already realized that the existence of an optimizer to the dual problem – often called *dual attainment* – is the essential key for the investigation of the optimal transport plans. But due to the nature of infinite dimensionality, establishing the dual attainment turns out to be quite subtle, as can be seen e.g. by the work of Brenier [10]. Thus it is no surprise that our second main result on the dual attainment is crucial for establishing Theorem 1.1. For a measure ξ on $\mathbb{R}$ define its *potential function* $u_\xi(x) := \int |x - y| d\xi(y)$. We say $\mu \leq_c \nu$ is *irreducible* if $I := \{u_\mu < u_\nu\}$ is connected and $\mu(I) = \mu(\mathbb{R})$. For more information on the notion of irreducibility, see Section 3 or [7].

Theorem 1.2. *Let $(\mu_i, \nu_i)_{i \in [d]}$ be irreducible pairs of marginals on $\mathbb{R}$. Let $c(x_1, ..., x_d, y_1, ..., y_d)$ be a lower-semicontinuous cost such that*

$|c(x,y)| \leq \sum_{i=1}^{d} \left(v_i(x_i) + w_i(y_i) \right)$ *for some continuous* $v_i \in L^1(\mu_i)$, $w_i \in L^1(\nu_i)$. *Then there exists a* pointwise dual maximizer (PDM), *a triplet of functions* $(\varphi_i, \psi_i, h_i)_{i=1}^{d}$ *that satisfies* (1.4) *tightly in the following pointwise manner (but does not necessarily belong to* Ψ*):*

$$(1.6) \qquad \sum_{i=1}^{d} \left(\varphi_i(x_i) + \psi_i(y_i) + h_i(x)(y_i - x_i) \right) = c(x,y) \quad \pi - a.s.$$

for every VMOT π *which solves the minimization problem in* (1.2).

We analogously define a *pointwise (or pathwise) dual minimizer* (PDm) for which the inequality in (1.4) should be reversed and π in (1.6) now solves the maximization problem in (1.2). This simply corresponds to the change of c to $-c$ in the theorem. We shall sometimes call Theorem 1.2 as *pointwise/pathwise dual attainment* (PDA) for VMOT problems.

The term *pointwise* indicates the fact that (1.4), (1.6) hold in a pointwise manner (i.e., the equality (1.6) is satisfied for π - almost every "point" or "path" (x,y)), and that we do *not* impose the integrability condition, namely $\varphi_i \in L^1(\mu_i)$, $\psi_i \in L^1(\nu_i)$, or h_i is bounded, on the PDM. Thus, a PDM needs not be in Ψ, and moreover, φ_i, ψ_i may assume the value $-\infty$. However, it is indeed true that $\varphi_i \in \mathbb{R}$ μ_i-a.s. and $\psi_i \in \mathbb{R}$ ν_i-a.s., implied by the equation (1.6). See also Remark 3.2.

A couple of pioneering works on the dual attainment for MOT problems on the line ($d = 1$) was established by Beiglböck-Juillet [5] and Beiglböck-Nutz-Touzi [7], where they already observed that the PDA could fail if one insists the integrability condition on PDM/PDm, or if one does not impose the irreducibility on the marginals. Beiglböck-Lim-Obłój [6] established PDA without the irreducibility assumption but imposing further regularity condition on c. Theorem 1.2 is a continuation of these efforts for the case $d \geq 2$, i.e., for the VMOT setup.

Further background. Let us briefly introduce the *optimal transport* (OT) problem, which is the prototype of MOT and VMOT problems. To begin with, let us introduce the notion of *push-forward* of measures. Given a measurable map $F : \mathcal{X} \rightarrow \mathcal{Y}$ and a measure μ on $\mathcal{X}$, the push-forward of μ by F, denoted by $F_{\#}\mu$, is a measure on $\mathcal{Y}$ satisfying

$$F_{\#}\mu(A) = \mu(F^{-1}(A))$$

for every measurable $A \subseteq \mathcal{Y}$.

Let $p^x, p^y : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the projection maps given by $p^x(x,y) = x$, $p^y(x,y) = y$. For a given cost function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ and two

Borel probability measures μ, ν on $\mathbb{R}^d$, the OT problem is given by

$$(1.7) \quad \text{Minimize} \quad \text{cost}[\pi] = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \, d\pi(x, y) \quad \text{over} \quad \pi \in \Pi(\mu, \nu)$$

where $\Pi(\mu, \nu) \subseteq \mathcal{P}(\mathbb{R}^{2d})$ is the set of *transport plans*, or *couplings*, if all $\pi \in \Pi(\mu, \nu)$ have given marginals μ and ν, meaning $p^x_{\#}\pi = \mu$, $p^y_{\#}\pi = \nu$.

Kantorovich [40,41] proposed the general formulation of the problem (1.7) in the 1940s after Gaspard Monge's first formulation in the 1780s [46]. Since then, many important contributions have been made on the subject, as can be seen by part through the works [1, 2, 8, 10, 11, 19, 20, 28, 31, 44, 45, 53, 55–57]. In particular, one of the central questions is when the OT plans are given by transport maps – often called *Monge problem*, especially when $c(x, y) = |x - y|$ – that is, when there exists a map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that a minimizer π in (1.7) is given by $\pi = (\text{Id}, T)_{\#}\mu$, where $\text{Id}(x) = x$ is the identity map on $\mathbb{R}^d$.

To answer, we need to understand the geometry of the OT plans. Such a characteristic structure of OT plan is encoded in its support, which exhibits the so called *c-cyclical monotonicity*. The monotonicity is already useful, but when combined with the celebrated theorem of Rockafellar [51], we obtain the *dual attainment*, a cornerstone of OT: there exist two functions $\varphi, \psi : \mathbb{R}^d \to \mathbb{R}$, called *Kantorovich potentials*, such that for every minimizer π of the problem (1.7), we have

$$(1.8) \qquad \varphi(x) + \psi(y) \leq c(x, y) \quad \forall x \in \mathbb{R}^d, \, \forall y \in \mathbb{R}^d,$$

$$(1.9) \qquad \varphi(x) + \psi(y) = c(x, y) \quad \pi - a.e. \, (x, y).$$

It turns out the dual attainment is essential for understanding OT [31].

Let us view the OT maps in a slightly different angle, and for this we recall the notion of *disintegration* of measures. For a general measure $\pi \in \mathcal{P}(\mathbb{R}^{2d})$, a disintegration of π via its first marginal $\pi^1 := p^x_{\#}\pi$ is

$$(1.10) \qquad\qquad\qquad \pi = \pi_x \otimes \pi^1$$

which means for any Borel sets $A, B \subseteq \mathbb{R}^d$, $\pi(A \times B) = \int_A \pi_x(B) d\pi^1(x)$. The family of probabilities $\pi_x := \{\pi_x\}_x$ is called a disintegration (or *kernel*) of π with respect to π^1. It is known this family uniquely exists π^1- a.e. x. Perhaps probabilistic language helps to understand better. Suppose π is the joint distribution of the $\mathbb{R}^d$-valued random variables X, Y, denoted as $\pi = \text{Law}(X, Y)$. Then $\pi^1 = \text{Law}(X)$, and π_x is the conditional law of Y given $X = x$, that is $\pi_x(B) = \mathbb{P}(Y \in B \,|\, X = x)$.

Now "$\pi \in \Pi(\mu, \nu)$ is given by a transport map T." is equivalent to

$$(1.11) \qquad \pi = (\text{Id}, T)_{\#}\mu \iff \pi = \pi_x \otimes \mu \text{ with } \pi_x = \delta_{T(x)}$$

where δ_y is the *Dirac mass* at y, the probability measure concentrated at y. In this case one could say the solution π exhibits an extremal geometry in the sense that a disintegration π_x, which describe the solution, are as extreme as Dirac masses, the most condensed measures. Finally, we would like to also mention some important applications of OT in economics [18, 22, 29, 50], and in statistical theory [12, 16, 17, 52].

Recently a variant of OT, referred to as *martingale optimal transport* (MOT), was introduced. In MOT problem, we consider the following:

$$(1.12) \quad \text{Minimize Cost}[\pi] = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \, d\pi(x, y) \ \text{ over } \ \pi \in \text{MT}(\mu, \nu)$$

where $\text{MT}(\mu, \nu)$ (Martingale Transport plans) is a subset of $\Pi(\mu, \nu)$ such that each $\pi = \pi_x \otimes \mu \in \text{MT}(\mu, \nu)$ satisfies $\delta_x \leq_c \pi_x$, that is π_x has its *barycenter* at x. Notice the kernel $(\pi_x)_x$ cannot be Dirac masses unless $\pi_x = \delta_x$. Nonetheless, Theorem 1.1 shows the kernel of each MOT for l_1 norm-type costs still exhibits an interesting extremal distribution.

Probabilistic description of the MOT problem is as follows: consider

$$(1.13) \qquad\qquad \text{Minimize} \quad \mathbb{E}_{\mathbb{P}}\left[c(X, Y)\right]$$

over all couples (X, Y) of $\mathbb{R}^d$-valued random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\text{Law}(X) = \mu$, $\text{Law}(Y) = \nu$ and $\mathbb{E}[Y|X] = X$, i.e. (X, Y) is a one-step martingale. Strassen [54] showed $\text{MT}(\mu, \nu)$ is nonempty if and only if μ, ν are in convex order $\mu \leq_c \nu$.

While some pioneering works to investigate the MOT problem have been made including [4, 26, 30, 38, 39], there is a related problem, called the *Skorokhod embedding problem* (SEP) which has a long history in probability theory. Since Hobson [36, 37] recognised the important connection between *model independent finance and asset pricing theory* and SEP (see Obłój [47] for a nice overview of the SEP and Beiglböck-Cox-Huesmann [3] for a link with OT theory), much related research has been done in this context including [13, 21, 34, 35, 49] for instance.

In the early stages of development of MOT theory, most of the research focused on single asset cases. More recently there have been efforts to generalize the theory into higher dimension $d \geq 2$, especially around the theory of duality and its attainment [24, 25, 27, 32, 33, 43, 48]. Nonetheless, to the best of the author's knowledge a complete understanding of the possibility for the dual attainment has not been achieved: the added "trading strategy" term $h(x) \cdot (y - x)$ in the dual formulation seems to drastically change the picture from the OT case, and the potential non-existence of dual optimizers makes the study of

solutions to MOT much more complex in the multi-dimensional setting.

While the MOT formulation (1.12) is a natural analogue of OT, not only it has difficulty in establishing PDA, but also its assumption on the marginal information is rather unrealistic from the financial point of view. This is because MOT assumes the *joint distribution* of the assets $\{X_i\}_i$ and $\{Y_i\}_i$ are known – the μ, ν. However, in practice such information is hardly observable from the market, and this is part of the motivation for us to introduce the current VMOT formulation. Let us recall that the main difference between MOT and VMOT lies in the marginal constraint, since in VMOT only the individual laws of $\{X_i\}_i$ and $\{Y_i\}_i$ – the $\vec{\mu}, \vec{\nu}$ – are assumed to be known. Thus, there are now at least three unknown couplings in the VMOT problem whose structures are of major interest: the VMOT plan $\pi = \mathrm{Law}(X, Y)$, and moreover the induced couplings $\pi^1 := p^x_\# \pi = \mathrm{Law}(X)$ and $\pi^2 := p^y_\# \pi = \mathrm{Law}(Y)$.

This paper is organized as follows. In section 2 we state and prove a more detailed version of Theorem 1.1 by applying PDA. In section 3 we prove the PDA, Theorem 1.2, using a compactness result whose proof is given in an appendix. Section 4 presents further results on the structure of VMOT and optimality of induced couplings.

2. Extremal correlation of martingales – Theorem 1.1

In this section, we state and prove Theorem 2.1 which is a more precise and extended version of Theorem 1.1. We will prove it here by assuming Theorem 1.2. Let us introduce some terminology. A norm $||\cdot||$ on $\mathbb{R}^d$ is called *strictly convex* if its unit ball $B = \{x : ||x|| \leq 1\}$ is strictly convex, i.e. every boundary point of B is an extreme point of B. Euclidean norm $|\cdot|$ clearly belongs to this class. For a set A, $\mathrm{conv}(A)$ is the convex hull of A, and for a convex set A, $\mathrm{Ext}(A)$ is the set of extreme points of A. For a measure ξ, the *support* of ξ, denoted by $\mathrm{supp}\,\xi$, is the smallest closed set on which ξ has its full mass. For measures μ, ν, there exists a unique largest measure $\mu \wedge \nu$ (called the *common mass* of μ, ν) which is dominated by μ and ν, i.e. $\mu \wedge \nu \leq \mu$, $\mu \wedge \nu \leq \nu$, where $\mu \leq \nu$ means $\mu(A) \leq \nu(A)$ for every measurable A. When μ, ν are given by densities f, g respectively, $\mu \wedge \nu$ is given by the density $f \wedge g := \min(f, g)$. $\mu \ll \mathcal{L}^d$ means μ is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^d$, that is μ has a density function.

Theorem 2.1. *Let $(\mu_i, \nu_i)_{i \in [d]}$ be pairs of probability measures on $\mathbb{R}$ in convex order, and $\mu_i \ll \mathcal{L}^1$ for all i. Let $c(x, y) = \pm ||x - y||$ where the*

norm $||\cdot||$ *is strictly convex and* $x \mapsto ||x||$ *is differentiable on* $\mathbb{R}^d \setminus \{0\}$.
Let $\pi = \pi_x \otimes \pi^1$ be any minimizer *for* (1.2) *where* $\pi^1 = p^x_\# \pi$, $\pi^2 = p^y_\# \pi$.

    (1) *If* $c(x,y) = -||x-y||$, *then the support of* π_x *coincides with the extreme points of the convex hull of itself:*

$$\operatorname{supp} \pi_x = \operatorname{Ext}\big(\operatorname{conv}(\operatorname{supp} \pi_x)\big), \quad \pi^1 - a.e.\, x.$$

    (2) *If* $c(x,y) = ||x-y||$, *we have* $D_\#(\pi^1 \wedge \pi^2) \le \pi$ *where* $D(x) = (x,x)$, *meaning that the common marginal* $\pi^1 \wedge \pi^2$ *stays put under* π, *or the minimizer* π *does not move the mass of* $\pi^1 \wedge \pi^2$. *Let* $\tilde{\pi} := \pi - D_\#(\pi^1 \wedge \pi^2)$ *and disintegrate as* $\tilde{\pi} = \tilde{\pi}_x \otimes \tilde{\pi}^1$. *Then*

$$\operatorname{supp} \tilde{\pi}_x = \operatorname{Ext}\big(\operatorname{conv}(\operatorname{supp} \tilde{\pi}_x)\big), \quad \tilde{\pi}^1 - a.e.\, x.$$

    Observe $\tilde{\pi}$ is a martingale transport from $\tilde{\pi}^1 := p^x_\# \tilde{\pi} = \pi^1 - \pi^1 \wedge \pi^2$ to $\tilde{\pi}^2 := p^y_\# \tilde{\pi} = \pi^2 - \pi^1 \wedge \pi^2$, and $\tilde{\pi}^1 \wedge \tilde{\pi}^2 = 0$. Since $D_\#(\pi^1 \wedge \pi^2)$ is an identity transport, we conclude that by minimizing $\mathbb{E}||x-y||$, we get

$$\operatorname{supp} \pi_x = \operatorname{Ext}\big(\operatorname{conv}(\operatorname{supp} \pi_x)\big) \text{ or } \operatorname{supp} \pi_x = \operatorname{Ext}\big(\operatorname{conv}(\operatorname{supp} \pi_x)\big) \cup \{x\}$$

and moreover the latter will be the case for $\pi^1 \wedge \pi^2 -$ a.e. x.

    We remark that [38], [39] studied the structure of MOT w.r.t. $c(x,y) = \pm|x-y|$ in the single-martingale setup, and later [5] rediscovered part of their results from a different approach. Theorem 2.1 may be viewed as a generalization of e.g. [5, Theorem 7.3, Theorem 7.4] to the multiple martingales setup, showing that the martingales given marginals are correlated in an extremal way to achieve the optimum.

Proof. **Step 1.** Let us tentatively assume that (μ_i, ν_i) are irreducible for every i. This assumption will be removed in the last step. Then by Theorem 1.2 we have a PDM $(f_i, -g_i, h_i)_{i \in [d]}$, that is, by denoting $f^\oplus(x) = \sum_{i=1}^d f_i(x_i)$, $g^\oplus(y) = \sum_{i=1}^d g_i(y_i)$ and $h = (h_1, ..., h_d)$, we have

$$(2.1) \quad f^\oplus(x) + h(x) \cdot (y-x) \le c(x,y) + g^\oplus(y) \quad \forall (x,y) \in \mathbb{R}^{2d}, \text{ and}$$

$$(2.2) \quad f^\oplus(x) + h(x) \cdot (y-x) = c(x,y) + g^\oplus(y) \quad \pi - a.s..$$

Recall that in Theorem 1.2, (2.2) is attained as real-valued, yielding f_i is finite μ_i-a.s. and g_i finite ν_i-a.s.. In this step we shall obtain a differential identity (2.13).

    To this end, recall the martingale Legendre transform of $g^\oplus(y)$ is a pair of functions $(\alpha, \tilde{\gamma})$ defined as in [32, Definition 3.1]:

$$(2.3) \qquad \alpha(x) := \sup\{a \in \mathbb{R} \,|\, \text{there exists } b \in \mathbb{R}^d \text{ such that}$$

$$a + b \cdot (y-x) \le c(x,y) + g^\oplus(y) \text{ for all } y\},$$

$$(2.4) \quad \tilde{\gamma}(x) := \{b \in \mathbb{R}^d \,|\, \alpha(x) + b \cdot (y-x) \le c(x,y) + g^\oplus(y) \text{ for all } y\}.$$

In general $\tilde{\gamma}$ is a convex set-valued (possibly empty, but see (2.5) below) function and we may choose a $\mathbb{R}^d$-valued function $\gamma \in \tilde{\gamma}$. Observe that if we define $H_x(y)$ to be the convex envelope of $y \mapsto c(x,y) + g^{\oplus}(y)$, then $y \mapsto \alpha(x) + \gamma(x) \cdot (y-x)$ is an affine tangent function to $H_x(y)$ at x. Now recall that $\mu_i \leq_c \nu_i$ is called *irreducible* if $I_i := \{u_{\mu_i} < u_{\nu_i}\}$ is connected open and $\mu_i(I_i) = \mu_i(\mathbb{R})$. Set $I := \otimes_i I_i \subseteq \mathbb{R}^d$ to be an open rectangle. Now (2.3) gives $f^{\oplus}(x) \leq \alpha(x)$, thus

$$f^{\oplus}(x) + \gamma(x) \cdot (y-x) - c(x,y) \leq \alpha(x) + \gamma(x) \cdot (y-x) - c(x,y)$$
$$\leq g^{\oplus}(y)$$

and equality holds throughout (as real-valued) π - a.s.. The fact g_i finite ν_i-a.s. then yields $I \subseteq \mathrm{conv}\{g^{\oplus}$ is finite$\}$. With this and the fact that c is Lipschitz, [32, Theorem 3.2] yields the following local regularity:

(2.5) α is Lipschitz and γ is bounded on every compact subset of I.

We wish f_i to attain Lipschitz property, so we will take the Legendre transform of f_i with respect to α to obtain such property. But as α is in general only locally Lipschitz, let us take the transform locally as follows: write $I_i =]a_i, b_i[$ and for $n \in \mathbb{N}$, let $I_i^n =]a_i + \frac{1}{n}, b_i - \frac{1}{n}[\cap] - n, n[$ and $I^n = \otimes_i I_i^n$. Now define φ_i successively for $i = 1, 2, ..., d$ by

$$(2.6) \qquad \varphi_i(x_i) := \inf_{x_j \in I_j^n, j \neq i} \left(\alpha(x) - \sum_{j<i} \varphi_j(x_j) - \sum_{j>i} f_j(x_j) \right).$$

Since α is (globally) Lipschitz in I^n (because α is locally Lipschitz in I), (2.6) implies that φ_i is Lipschitz in I_i^n, and

$$f_i \leq \varphi_i \text{ on } I_i^n \quad \text{and} \quad \sum_i \varphi_i(x_i) \leq \alpha(x) \text{ on } I^n.$$

Let $\pi_n := \pi\big|_{I^n \times \mathbb{R}^d}$ be the restriction of π on $I^n \times \mathbb{R}^d$, let $\pi_n^1 = p_{\#}^x \pi_n$, and let $(\mu_i^n)_{i \in [d]}$ be the one-dimensional marginals of π_n^1. As $\mu_i^n \leq \mu_i$ yields $\mu_i^n \ll \mathcal{L}$, each φ_i is differentiable μ_i^n-a.e.. Hence

$$(2.7) \qquad \varphi^{\oplus}(x) := \sum_i \varphi_i(x_i) \text{ is differentiable} \quad \pi_n^1 - a.e.\, x.$$

Now because $f^{\oplus} \leq \varphi^{\oplus} \leq \alpha$, we have

$(2.8) \quad \varphi^{\oplus}(x) + \gamma(x) \cdot (y-x) - c(x,y) \leq g^{\oplus}(y) \quad \forall(x,y) \in I^n \times \mathbb{R}^d,$

$(2.9) \quad \varphi^{\oplus}(x) + \gamma(x) \cdot (y-x) - c(x,y) = g^{\oplus}(y) \quad \pi_n - a.e.$

We may rewrite (2.9) as

(2.10)
For $\pi_n^1 - a.e.\, x$, $\varphi^{\oplus}(x) + \gamma(x) \cdot (y-x) - c(x,y) = g^{\oplus}(y) \quad \pi_x - a.e.\, y$.

Fix x_0 at which (2.10) holds and $\varphi^\oplus$ is differentiable. Let V_0 be the subspace of $\mathbb{R}^d$ spanned by the set $\operatorname{supp} \pi_{x_0} - x_0$ (translation of $\operatorname{supp} \pi_{x_0}$ by $-x_0$). If $\dim V_0 = 0$ then it simply means $\operatorname{supp} \pi_{x_0} = \{x_0\}$ and there is nothing to prove. Thus let us assume that $\dim V_0 \geq 1$. Now [32, Lemma 4.1] showed that, if (2.8), (2.10) hold, then

(2.11)
directional derivative of $\operatorname{proj}_{V_0} \gamma$ exists at x_0 in every direction $u \in V_0$,

where $\operatorname{proj}_{V_0} \gamma$ is the orthogonal projection of the $\mathbb{R}^d$-valued function γ on V_0. While a proof can be found in [32], here let us give some intuition: in (2.8) the function $y \mapsto \varphi^\oplus(x_0) + \gamma(x_0) \cdot (y - x_0) - c(x_0, y)$ is bounded above by $g^\oplus(y)$, but (2.10) tells us that the function is in fact tightly bounded by $g^\oplus(y)$ for π_{x_0} - a.e. y. The facts π_{x_0} has barycenter at x_0 and the functions $\varphi^\oplus(x)$ and $x \mapsto c(x, y)$ are differentiable at $x = x_0$ implies that $\operatorname{proj}_{V_0} \gamma$ has not much room to vary wildly near x_0 in V_0, yielding it is forced to be differentiable at x_0 in V_0.

Next, note that (2.8), (2.10) imply for $\pi_{x_0} - a.e.\, y$

(2.12)
$$\varphi^\oplus(x_0) + \gamma(x_0) \cdot (y - x_0) - c(x_0, y)$$
$$\geq \varphi^\oplus(x) + \gamma(x) \cdot (y - x) - c(x, y) \quad \forall x \in I^n,$$

and notice by continuity of c, (2.12) holds for every $y \in \operatorname{supp} \pi_{x_0}$. Then from (2.7), (2.11), we deduce that for any nonzero vector u in V_0, by taking u-directional derivative in x at x_0, it holds that for any $y \in \operatorname{supp} \pi_{x_0} \setminus \{x_0\}$,

(2.13) $\nabla_u \varphi^\oplus(x_0) + \nabla_u \gamma(x_0) \cdot (y - x_0) - \gamma(x_0) \cdot u - \nabla_u c(x_0, y) = 0.$

From this identity [32] proved the theorem when the cost is given by the Euclidean norm. We will follow a similar line but as we deal with more general strictly convex norms, we shall need a more involved argument.

Step 2. We shall prove the "non-staying" property of the common mass $\pi^1 \wedge \pi^2$ in the case $c(x, y) = -||x - y||$ ($\pi^1 \neq \pi^2$ by irreducibility) and the "staying" property in the case $c(x, y) = +||x - y||$ for the minimization problem in (1.2).

For $c(x, y) = -||x - y||$, (2.12) immediately implies $x_0 \notin \operatorname{supp} \pi_{x_0}$ for every x_0 at which $\varphi^\oplus$ is differentiable, thus π^1-a.s., (which we call the non-staying property of the minimizer π) by the following reason: if $x_0 \in \operatorname{supp} \pi_{x_0}$, then the function $x \mapsto \varphi^\oplus(x) + \gamma(x) \cdot (x_0 - x) + ||x - x_0||$ must attain its maximum at $x = x_0$ by (2.12). But notice that due to the increase of $x \mapsto ||x - x_0||$ the function will strictly increase as x moves away from x_0 along any direction $u \in V_0$ satisfying $\nabla_u(\varphi^\oplus(x) + \gamma(x) \cdot (x_0 - x)) \geq 0$ at $x = x_0$, a contradiction. Notice we have shown

that if $\operatorname{supp} \pi_{x_0} \neq \{x_0\}$ and a PDM exists, then for $c(x, y) = -||x - y||$, $x_0 \notin \operatorname{supp} \pi_{x_0}$ whenever $\nabla \varphi^\oplus$ exists at x_0.

Next, the staying property for the case $c(x, y) = ||x - y||$ refers to the statement $D_\#(\pi^1 \wedge \pi^2) \leq \pi$ as stated in the theorem. This property was proved in e.g. [5, Theorem 7.4] for one dimensional case and then generalized in [43, Theorem 2.3] for general dimension in the MOT setup, but note that a solution π of MMOT is also a solution of MOT between its own marginals π^1 and π^2. Both [5], [43] assumed $c(x, y) = |x - y|$, but we note the same proof works for any strictly convex norm cost. From the staying property, the study of the geometry of π is now reduced to the study of $\tilde\pi := \pi - D_\#(\pi^1 \wedge \pi^2)$. Note that $\tilde\pi$ has no mass on the diagonal $\{(x, x) \,|\, x \in \mathbb{R}^d\}$ and $\tilde\pi$ must solve the VMOT problem with respect to its own one-dimensional marginals.

Step 3. By Step 2 we can assume π_{x_0} has no mass at x_0. Under this assumption we will show $\operatorname{supp} \pi_{x_0}$ is contained in the set of extreme points of $\operatorname{conv}(\operatorname{supp} \pi_{x_0})$. To this end, first we will show that $\operatorname{supp} \pi_{x_0}$ is contained in the boundary of $\operatorname{conv}(\operatorname{supp} \pi_{x_0})$, where the boundary refers to the topology of V_0 and not of $\mathbb{R}^d$. Suppose on the contrary $\operatorname{supp} \pi_{x_0} \not\subseteq \operatorname{bd}\big(\operatorname{conv}(\operatorname{supp} \pi_{x_0})\big)$. Then we can find a point $y \in \operatorname{int}\big(\operatorname{conv}(\operatorname{supp} \pi_{x_0})\big) \cap (\operatorname{supp} \pi_{x_0} \setminus \{x_0\})$ and a subset $\{y_0, y_1, ..., y_m\} \subseteq \operatorname{supp} \pi_{x_0} \setminus \{x_0, y\}$ such that y is a convex combination of these, i.e.,

$$(2.14) \quad \text{There exists } p_i > 0 \text{ for all } i, \sum_{i=0}^m p_i = 1 \text{ such that } y = \sum_{i=0}^m p_i y_i.$$

Then due to the linearity of $y \mapsto \nabla_u \gamma(x) \cdot (y - x)$, from (2.13) we deduce

$$(2.15) \qquad \nabla_u c(x_0, y) = \sum_{i=0}^m p_i \nabla_u c(x_0, y_i) \quad \forall u \in V_0.$$

Using this, we will show that

(2.16) the points $\{y, y_0, y_1, ..., y_m\}$ lie on a ray emanating from x_0.

To see this, take $u = y - x_0$, let $c(x, y) = ||x - y||$, and let $\overline{x_0 y}$ be the infinite line containing x_0, y. Thanks to the strict convexity of $|| \cdot ||$, we will firstly show that $\nabla_u c(x_0, y) < \nabla_u c(x_0, z)$ for any $z \notin \overline{x_0 y}$, where ∇_u refers to the u-directional derivative w.r.t. the x variable. To see this, we consider the following function on $]-1, 1[$

$$\sigma(t) := \big(||z - (x_0 + tu)|| - ||z - x_0||\big) - \big(||y - (x_0 + tu) - ||y - x_0||\big).$$

The claim reads $\sigma'(0) > 0$. Note $||y - x_0|| - ||y - (x_0 + tu)|| = t||u||$, so

$$\sigma(t) = ||z - (x_0 + tu)|| - ||z - x_0|| + t||u||, \quad -1 < t < 1.$$

Notice $\sigma(0) = 0$, $\sigma(t) > 0$ for $t > 0$ and $\sigma(t) < 0$ for $t < 0$ since the norm $||\cdot||$ is strictly convex and $z \notin \overline{x_0 y}$. Also notice σ is convex, hence $\sigma'(0) > 0$, showing the claim in this case. Now if $z \in \overline{x_0 y}$ with z and y not on the same side of x_0, then $\sigma(t) = 2t||u||$ for small t, so that again $\sigma'(0) > 0$ and $\nabla_u c(x_0, y) < \nabla_u c(x_0, z)$. Finally if $z \in \overline{x_0 y}$ with z and y on the same side of x_0, then $\sigma'(0) = 0$ and $\nabla_u c(x_0, y) = \nabla_u c(x_0, z)$. With (2.15) this clearly yields (2.16). Finally, a similar argument yields (2.16) for the cost $c(x, y) = -||x - y||$.

Now if $\dim V_0 \geq 2$, we can choose $\{y_0, y_1, ..., y_m\}$ such that they are not aligned meanwhile (2.14) holds. But (2.16) then forces them to be aligned, a contradiction. If $\dim V_0 = 1$ then since π_{x_0} is centered at x_0, we can choose y_0, y_1 in the opposite direction with respect to x_0, i.e. $(y_0 - x_0) \cdot (y_1 - x_0) < 0$. Also $(y_0 - y) \cdot (y_1 - y) < 0$ by (2.14). Then again (2.16) cannot hold, a contradiction. This yields $\operatorname{supp} \pi_{x_0} \subseteq \operatorname{bd}\big(\operatorname{conv}(\operatorname{supp} \pi_{x_0})\big)$. Now if $\operatorname{supp} \pi_{x_0} \not\subseteq \operatorname{Ext}(\operatorname{conv}(\operatorname{supp} \pi_{x_0}))$, then again we can find $\{y, y_0, y_1, ..., y_m\} \subseteq \operatorname{supp} \pi_{x_0}$ such that (2.14) holds. Then by the above argument we deduce (2.16), a contradiction to the fact that $\{y, y_0, y_1, ..., y_m\} \subseteq \operatorname{bd}\big(\operatorname{conv}(\operatorname{supp} \pi_{x_0})\big)$, since a ray emanating from an interior point of a convex set can intersect with its boundary at one point only. Hence we get $\operatorname{supp} \pi_{x_0} \subseteq \operatorname{Ext}\big(\operatorname{conv}(\operatorname{supp} \pi_{x_0})\big)$, and as the reverse inclusion is clear, we conclude $\operatorname{supp} \pi_{x_0} = \operatorname{Ext}\big(\operatorname{conv}(\operatorname{supp} \pi_{x_0})\big)$. Notice this proves the theorem for $\pi_n = \pi\big|_{I^n \times \mathbb{R}^d}$, the restriction of π on $I^n \times \mathbb{R}^d$ in Step 1. Finally, letting $n \to \infty$ for the domain I^n completes the proof of Theorem 2.1 for any irreducible pair of marginals $(\mu_i, \nu_i)_i$.

Step 4. We will prove the theorem for each pair of marginals (μ_i, ν_i) only in convex order and not necessarily irreducible. It is well known that any convex-ordered pair (μ_i, ν_i) can be decomposed as at most countably many irreducible pairs, and the decomposition is uniquely determined by the potential functions u_{μ_i}, u_{ν_i}. While more details can be found in [5], [7], we provide the statement for reader's convenience.

[7, Proposition 2.3] For each $i \in [d]$, let $(I_{i,k})_{1 \leq k \leq N}$ be the open components of the open set $\{u_{\mu_i} < u_{\nu_i}\}$ in $\mathbb{R}$, where $N \in \mathbb{N} \cup \{\infty\}$. Let $I_{i,0} = \mathbb{R} \setminus \cup_{k \geq 1} I_{i,k}$ and $\mu_{i,k} = \mu\big|_{I_{i,k}}$ for $k \geq 0$, so that $\mu_i = \sum_{k \geq 0} \mu_{i,k}$. Then there exists a unique decomposition $\nu_i = \sum_{k \geq 0} \nu_{i,k}$ such that

$\mu_{i,0} = \nu_{i,0}$, and $(\mu_{i,k}, \nu_{i,k})$ is irreducible for $k \geq 1$ with $\mu_{i,k}(I_{i,k}) = \mu_{i,k}(\mathbb{R})$.

Moreover, any $\pi_i \in \operatorname{MT}(\mu_i, \nu_i)$ admits a unique decomposition $\pi_i = \sum_{k \geq 0} \pi_{i,k}$ such that $\pi_{i,k} \in \operatorname{MT}(\mu_{i,k}, \nu_{i,k})$ for all $k \geq 0$.

Note that $\pi_{i,0}$ must be the identity transport (i.e. $\pi_{i,0}$ is concentrated on the diagonal $\{(x,x) \,|\, x \in I_{i,0}\}$) since it is a martingale and $\mu_{i,0} = \nu_{i,0}$. There is no randomness in $\pi_{i,0}$. Now for the rest of the proof we will assume $d = 2$ to avoid notational difficulty, but one may observe the same argument works for any $d \geq 3$ just with some notational challenge.

Let π be a minimizer in (1.2) and let $\mu_1 = \sum_{k \geq 0} \mu_{1,k}$, $\mu_2 = \sum_{k \geq 0} \mu_{2,k}$ be the unique decomposition of μ_1, μ_2 respectively. Then our domain $\mathbb{R}^2$ is decomposed as $\mathbb{R}^2 = \cup_{m \geq 0, n \geq 0} I_{1,m} \times I_{2,n}$ accordingly. Now the strategy is to study the geometry of π on each domain $I_{m,n} := I_{1,m} \times I_{2,n}$.

Let $\pi_{m,n} := \pi\big|_{I_{m,n} \times \mathbb{R}^2}$ be the restriction of π on $I_{m,n} \times \mathbb{R}^2$. If $m \geq 1$ and $n \geq 1$, then $(\mu_{1,m}, \nu_{1,m})$ and $(\mu_{2,n}, \nu_{2,n})$ are both irreducible and in this case we already established the theorem. On the other extreme, that is if $m = 0$ and $n = 0$, then as mentioned above there is no randomness in $\pi_{0,0}$, that is $\pi_{0,0} \in \mathrm{MT}(\pi_{0,0}^1, \pi_{0,0}^1)$ where $\pi_{0,0}^1$ is a coupling of $\mu_{1,0}$ and $\mu_{2,0}$. Hence $\pi_{0,0}$ represents an identity transport, and thus the theorem obviously holds in this case.

The remaining case is $m = 0$ and $n \geq 1$. Write $\pi = \pi_{0,n}$ for simplicity and recall that $\pi = \mathrm{Law}(X, Y)$ where $X = (X_1, X_2)$, $Y = (Y_1, Y_2)$, and (X_1, Y_1), (X_2, Y_2) are jointly martingales. Let $\gamma_1 = \mathrm{Law}(X_1, Y_1)$ and $\gamma_2 = \mathrm{Law}(X_2, Y_2)$, so that π is a coupling of martingales γ_1 and γ_2. But since $\mu_{1,0} = \nu_{1,0}$ we have $X_1 = Y_1$, thus the cost is simplified as follows:

$$c(x,y) = ||(x_1, x_2) - (y_1, y_2)|| = ||(0, x_2 - y_2)|| =: ||x_2 - y_2||_*$$

where $||\cdot||_*$ is the restriction of the norm $||\cdot||$ on the second coordinate axis in $\mathbb{R}^2$. Hence we have $\mathbb{E}_\pi ||X - Y|| = \mathbb{E}_{\gamma_2} ||X_2 - Y_2||_*$ and this implies γ_2 is an optimizer in $\mathrm{VMT}(\mu_2, \nu_2)$ with respect to $||\cdot||_*$. Now by the irreducibility of (μ_2, ν_2) Step 3 already established the theorem for γ_2, hence the theorem also holds for π as γ_1 is merely an identity transport. This completes the proof of the theorem. $\square$

We end this section with some examples that illustrate Theorem 2.1.

Example 2.2. *We give a maximizer for* (1.2) *with* $c(x,y) = |x - y|$. *As [32] observed, the inequality* $\frac{1}{2}(|x - y| - 1)^2 \geq 0$ *may be rewritten as*

$$(2.17) \quad \sum_{i=1}^{d} \left(\frac{1}{2}\left(x_i^2 - \frac{1}{2d}\right) - \frac{1}{2}\left(y_i^2 + \frac{1}{2d}\right) + x_i(y_i - x_i) \right) \leq -|x - y|.$$

Fix any probability measure $\pi^1 \in \mathcal{P}_2(\mathbb{R}^d)$. *Choose a family of probabilities* $(\pi_x)_x$ *on* $\mathbb{R}^d$ *such that* $\delta_x \leq_c \pi_x$ *and* $\pi_x(S_{x,1}) = 1$ *where* $S_{x,1}$ *is the unit sphere in* $\mathbb{R}^d$ *centered at* x. *Now define a martingale measure* π *by* $\pi := \pi_x \otimes \pi^1$. *Then* π *is a maximizer in* $\mathrm{VMT}(\vec{\mu}, \vec{\nu})$, *where* $\vec{\mu} = (\mu_1, ..., \mu_d)$ *and* $\vec{\nu} = (\nu_1, ..., \nu_d)$ *are the one-dimensional marginals*

induced by π. The optimality of π follows from the fact that while the integration of the left hand side of (2.17) by any $\gamma \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$ is a constant, the inequality (2.17) becomes an equality precisely on the set $G := \{(x, y) : |x - y| = 1\}$, and moreover $\pi(G) = 1$.

Example 2.3. *We shall construct a minimizer for (1.2) with $c(x, y) = |x - y|$. Denote by $\mathrm{Leb}\big|_I$ the Lebesgue measure restricted on a set I. Let $d = 2$, $\mu_1 = \mu_2 = \mathrm{Leb}\big|_{[-1/2,1/2]}$, $\nu_1 = \nu_2 = \frac{1}{2}\mathrm{Leb}\big|_{[-1,1]}$. Recall that the projection of any element in $\mathrm{VMT}(\vec{\mu}, \vec{\nu})$ on the "first martingale space" $\mathbb{R}^2$ belongs to $\mathrm{MT}(\mu_1, \nu_1)$, that is, if $\mathbb{E}[Y|X] = X$ then $\mathbb{E}[Y_1|X_1] = X_1$. This implies $P(c) \geq P_1(c)$, where $P_1(c)$ is the primal value of the one-dimensional MOT problem*

$$(2.18) \qquad P_1(c) := \inf_{\pi_1 \in \mathrm{MT}(\mu_1, \nu_1)} \int_{\mathbb{R} \times \mathbb{R}} |x_1 - y_1| \, d\pi_1(x_1, y_1)$$

simply because $|x - y| \geq |x_1 - y_1|$. It is well known (see e.g. [5, Theorem 7.4]) that the solution to (2.18), say γ^1, is unique and there exist functions $T^- : [-\frac{1}{2}, \frac{1}{2}] \to [-1, -\frac{1}{2}]$, $T^+ : [-\frac{1}{2}, \frac{1}{2}] \to [\frac{1}{2}, 1]$ such that

$$\gamma^1_{x_1} = \frac{1}{2}\left(\delta_{x_1} + \lambda^-(x_1)\delta_{T^-(x_1)} + \lambda^+(x_1)\delta_{T^+(x_1)}\right) \quad \text{where}$$

$$\gamma^1 = \gamma^1_{x_1} \otimes \mu_1 \quad \text{and} \quad \lambda^\pm(x_1) = \left|\frac{T^\mp(x_1) - x_1}{T^+(x_1) - T^-(x_1)}\right|.$$

This means that the martingale γ^1 splits the mass at each $x_1 \in]-1, 1[$ in μ_1 onto three points $\{T^-(x_1), x_1, T^+(x_1)\}$. Since the identity transport $x_1 \mapsto x_1$ has no contribution to the cost (2.18), we see that

(2.19)
$$P_1(c) = \frac{1}{2}\int \left(\lambda^-(x_1)|T^-(x_1) - x_1| + \lambda^+(x_1)|T^+(x_1) - x_1|\right)d\mu_1(x_1).$$

Now let us construct a minimizer $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$. For each $x = (x_1, x_2) \in]-\frac{1}{2}, \frac{1}{2}[^2$ define a kernel $\pi_x \in \mathcal{P}(\mathbb{R}^2)$ by

$$\pi_x = \frac{1}{2}\left(\lambda^-(x_1)\delta_{\left(T^-(x_1), x_2\right)} + \lambda^+(x_1)\delta_{\left(T^+(x_1), x_2\right)}\right.$$

$$\left. + \lambda^-(x_2)\delta_{\left(x_1, T^-(x_2)\right)} + \lambda^+(x_2)\delta_{\left(x_1, T^+(x_2)\right)}\right).$$

Note that π_x splits mass at x toward four directions – west, east, south and north. Now let $\pi^1 \in \Pi(\mu_1, \mu_2)$ be any coupling of μ_1, μ_2, and define $\pi := \pi_x \otimes \pi^1$. Then notice that $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$, and moreover by (2.19), we have $\int |x - y| d\pi = P_1(c)$. Therefore, π is a minimizer.

We note that this example gives a counterexample to the second conjecture given in [32] as π_x is supported at four points rather than three. But the existence question of such polytope-type MOTs still remains open. For related results, we refer to [23].

The next example shows the strict convexity of the norm assumption cannot be omitted in Theorem 2.1.

Example 2.4. *Strict convexity of the norm is necessary for the extremal structure of* VMOT. *For an example, let* $d = 2$, $c(x, y) = \max(|x_1 - y_1|, |x_2 - y_2|)$, $\mu_1 = \mu_2 = \text{Leb}\big|_{[-1/2, 1/2]}$, $\nu_1 = \frac{1}{2}(\delta_{-10} + \delta_{10})$, $\nu_2 = \frac{1}{2}\text{Leb}\big|_{[-1,1]}$. *Observe every* $\gamma \in \text{VMT}(\vec{\mu}, \vec{\nu})$ *leads to the same cost* $\int |x_1 - y_1| \, d\gamma^1$ *where* $\gamma^1 \in \text{MT}(\mu_1, \nu_1)$ *(note* $\text{MT}(\mu_1, \nu_1)$ *is a singleton). Now clearly there are elements in* $\text{VMT}(\vec{\mu}, \vec{\nu})$ *which do not satisfy Theorem 2.1, for example one may take* $\gamma = \gamma^1 \otimes \gamma^2$ *where* $\gamma^2 \in \text{MT}(\mu_2, \nu_2)$ *is independent of* γ^1 *and its kernel* $\{\gamma_x^2\}_x$ *are dispersed onto* $[-1, 1]$.

3. Existence of dual optimizers for VMOT – Theorem 1.2

One of the most important results for the proof of Theorem 1.2 is Proposition 3.1, a compactness property of a class of convex functions.

Recall two probabilities in convex order $\mu \leq_c \nu$ is irreducible if $I = \{x \mid u_\mu(x) < u_\nu(x)\}$ is connected and $\mu(I) = \mu(\mathbb{R})$; see [7, Section 2]. In this case, (I, J) is called the *domain* of (μ, ν) where J is the smallest interval satisfying $\nu(J) = \nu(\mathbb{R})$, that is, J is the union of I and any endpoints of I that are atoms of ν, and moreover, $I = \text{int}(J) = \text{int}(\text{conv}(\text{supp}(\nu)))$. Note that $\mu \leq_c \nu$ is irreducible if and only if for every $\pi = \pi_x \otimes \mu \in \text{MT}(\mu, \nu)$ and $z \in I := \text{int}(\text{conv}(\text{supp}\,\nu))$, we have

$$(3.1) \quad \mu(I_z^\pi) > 0 \quad \text{where} \quad I_z^\pi := \{x \in I \mid z \in \text{int}(\text{conv}(\text{supp}(\pi_x)))\}.$$

In other words, (μ, ν) is irreducible if and only if for any $z \in \text{int}(\text{conv}(\text{supp}\,\nu))$, every $\pi \in \text{MT}(\mu, \nu)$ must "cut across" the point z. This follows directly from the definition of potential functions and Jensen's inequality.

For irreducible pairs $(\mu_i, \nu_i)_{i \in [d]}$ with domains (I_i, J_i), let us define $I = I_1 \times \dots \times I_d$, $J = J_1 \times \dots \times J_d$ and $\mu^\otimes = \mu_1 \otimes \dots \otimes \mu_d$, $\nu^\otimes = \nu_1 \otimes \dots \otimes \nu_d$.

Proposition 3.1. *Let* $(\mu_i, \nu_i)_{i \in [d]}$ *be irreducible pairs of probabilities with domains* $(I_i, J_i)_i$. *Let* $a \in I$, $C \in \mathbb{R}$. *Consider the following class of functions* $\Lambda = \Lambda(a, C, \vec{\mu}, \vec{\nu})$ *where every* $\chi \in \Lambda$ *satisfies the following:*

(1) χ *is a real-valued convex function on* J,
(2) $\chi \geq 0$ *and* $\chi(a) = 0$,
(3) $\int \chi \, d(\nu^\otimes - \mu^\otimes) \leq C$.

Then Λ is locally bounded in the following sense: for each compact subset K of J, there exists $M = M(K)$ such that $\chi \leq M$ on K for every $\chi \in \Lambda$. Moreover, for any sequence $\{\chi_n\}_n$ in Λ there exists a subsequence $\{\chi_{n_j}\}_j$ of $\{\chi_n\}_n$ and a real-valued convex function χ on J such that $\lim_{j\to\infty} \chi_{n_j}(x) = \chi(x)$ for every $x \in J$, and the convergence is uniform on every compact subset of I.

We note that the proposition for the case $d = 1$ was proved in [7]. In this section we prove Theorem 1.2 while assuming Proposition 3.1, whose proof will then be given in an appendix. During the proof the bounding constant C may vary, but observe it does not depend on n.

Proof. **Step 1.** The assumption $|c(x, y)| \leq \sum_{i=1}^{d} \big(v_i(x_i) + w_i(y_i)\big)$ for continuous $v_i \in L^1(\mu_i)$, $w_i \in L^1(\nu_i)$ ensures $P(c) = D(c)$ as in (1.5) (for a proof, see e.g. [58]). Moreover, it is obvious that a dual optimizer exists for $c(x, y)$ iff so does for $\tilde{c}(x, y) := c(x, y) - \sum_{i=1}^{d} \big(v_i(x_i) + w_i(y_i)\big)$. Hence by replacing c with $\tilde{c}$, from now on we will assume that $c \leq 0$.

As $P(c) = D(c) \in \mathbb{R}$, we can find an "approximating dual maximizer" $(f_{i,n}, g_{i,n}, h_{i,n})_{n\in\mathbb{N}}$ which consist of real-valued continuous functions $f_{i,n} \in L^1(\mu_i)$, $g_{i,n} \in L^1(\nu_i)$, and continuous bounded $h_{i,n}$ for every $i \in [d]$ and $n \in \mathbb{N}$, such that the following weak duality holds:

$$(3.2) \qquad \sum_{i=1}^{d} \big(f_{i,n}(x_i) - g_{i,n}(y_i) + h_{i,n}(x)(y_i - x_i)\big) \leq c(x, y),$$

$$(3.3) \qquad \sum_{i=1}^{d} \Big(\int f_{i,n} d\mu_i - \int g_{i,n} d\nu_i\Big) \nearrow P(c) \quad \text{as} \quad n \to \infty.$$

Denote $f_n^{\oplus}(x) = \sum_{i=1}^{d} f_{i,n}(x_i)$, $g_n^{\oplus}(y) = \sum_{i=1}^{d} g_{i,n}(y_i)$, $h_n(x) = \big(h_{1,n}(x), ..., h_{d,n}(x)\big)$. By taking supremum over $x \in \mathbb{R}^d$ in (3.2), we get

$$(3.4) \qquad \chi_n(y) := \sup_{x\in\mathbb{R}^d} \big(f_n^{\oplus}(x) + h_n(x) \cdot (y - x)\big) \leq g_n^{\oplus}(y).$$

Notice χ_n is a convex function on $\mathbb{R}^d$, and by definition of χ_n, we see

$$(3.5) \qquad\qquad\qquad f_n^{\oplus} \leq \chi_n \leq g_n^{\oplus} \quad \forall n \in \mathbb{N}.$$

In this step we shall obtain local uniform boundedness of $\{\chi_n\}_n$ (3.8). Let $a \in I$ be fixed. By subtracting an appropriate linear function $L_n(y) = \nabla L_n(x) \cdot (y - x) + L_n(x)$ from (3.2), that is by replacing $f_n^{\oplus}(x)$ with $f_n^{\oplus}(x) - L_n(x)$, $g_n^{\oplus}(y)$ with $g_n^{\oplus}(y) - L_n(y)$ and $h_n(x)$ with $h_n(x) - \nabla L_n(x)$, we can assume

$$(3.6) \qquad\qquad\qquad \chi_n \geq 0 \quad \text{and} \quad \chi_n(a) = 0 \quad \forall n.$$

Next, observe (3.5) gives

$$\sum_i \int g_{i,n} d\nu_i \geq \int \chi_n d\nu^\oplus \geq \int \chi_n d\mu^\oplus \geq \sum_i \int f_{i,n} d\mu_i$$

where the second inequality is due to the convexity of χ_n and $\mu_i \leq_c \nu_i$. Then by (3.3), there exists a constant $C > 0$ such that

$$(3.7) \qquad \int \chi_n \, d(\nu^\otimes - \mu^\otimes) \leq C \quad \forall n.$$

Let $\{\epsilon_k\}_k$ be a positive decreasing sequence tending to zero as $k \to \infty$, and write $I_i =]a_i, b_i[$ where $-\infty \leq a_i < b_i \leq +\infty$. Then we define the compact interval $J_{i,k} := [c_{i,k}, d_{i,k}]$ for $i \in [d]$ and $k \in \mathbb{N}$ as

If $a_i > -\infty$ then: $\nu_i(a_i) = 0 \Rightarrow c_{i,k} := a_i + \epsilon_k$, $\nu_i(a_i) > 0 \Rightarrow c_{i,k} := a_i$,

If $b_i < +\infty$ then: $\nu_i(b_i) = 0 \Rightarrow d_{i,k} := b_i - \epsilon_k$, $\nu_i(b_i) > 0 \Rightarrow d_{i,k} := b_i$,

If $a_i = -\infty \Rightarrow c_{i,k} := -1/\epsilon_k$. If $b_i = +\infty \Rightarrow d_{i,k} := +1/\epsilon_k$.

Thus for example, if $\nu(a_i) = 0$ and $\nu(b_i) > 0$, then $J_{i,k} = [a_i + \epsilon_k, b_i]$. Let ϵ_1 be small so that $\mu_i(J_{i,1}) > 0$, $\nu_i(J_{i,1}) > 0$ for every $i \in [d]$. Notice $J_{i,k} \nearrow J_i$. Let $J_k := J_{1,k} \times J_{2,k} \times ... \times J_{d,k}$. Then by Proposition 3.1, there exists $\{M_k\}_k$ such that

$$(3.8) \qquad 0 \leq \sup_n \chi_n \leq M_k \quad \text{on} \ \ J_k.$$

Step 2. Let $(f_{i,n}, g_{i,n}, h_{i,n})_{n \in \mathbb{N}}$ be an approximating dual maximizer. We want to show pointwise convergence of $f_{i,n}$ and $g_{i,n}$, i.e. $f_{i,n}(x_i) \to f_i(x_i) \in \mathbb{R}$ for μ_i-a.e. x_i and $g_{i,n}(y_i) \to g_i(y_i) \in \mathbb{R}$ for ν_i-a.e. y_i as $n \to \infty$. However, in establishing the convergence we see that there is an immediate obstacle when $d \geq 2$, that is, in the duality formulation (3.2) one can always replace $(f_{i,n})_{i \in [d]}$ by $(f_{i,n} + C_{i,n})_{i \in [d]}$ for any constants $(C_{i,n})_{i \in [d]}$ satisfying $\sum_i C_{i,n} = 0$ and similarly replace $(g_{i,n})_{i \in [d]}$ by $(g_{i,n} + D_{i,n})_{i \in [d]}$. This means the convergence cannot be shown for any approximating dual maximizer. In this step we will show there exists an approximating dual maximizer which satisfies the convergence.

Take an approximating dual maximizer $(f_{i,n}, g_{i,n}, h_{i,n})_{n \in \mathbb{N}}$. Observe

$$C \geq \int g_n^\oplus d\nu^\otimes - \int f_n^\oplus d\mu^\otimes$$

$$\geq \int \chi_n d\nu^\otimes - \int f_n^\oplus d\mu^\otimes$$

$$\geq \int \chi_n d\mu^\otimes - \int f_n^\oplus d\mu^\otimes$$

$$= \| \chi_n - f_n^\oplus \|_{L^1(\mu^\otimes)} \quad \forall n$$

where the third inequality is by convexity of χ_n and $\mu^\otimes \leq_c \nu^\otimes$. For each $k \in \mathbb{N}$, let $\mu_{i,k}$ be the restriction of μ_i on $J_{i,k}$ then normalized to be a probability measure. Let $\mu_k^\otimes = \otimes_i \mu_{i,k}$, so that $\mu_k^\otimes(J_k) = 1$. Define

$$v_{i,k,n} = \int f_{i,n}\, d\mu_{i,k}, \quad i \in [d], k \in \mathbb{N}, n \in \mathbb{N}.$$

Let us fix $k \in \mathbb{N}$. We claim that

$$\sup_n \| f_{i,n} - v_{i,k,n} \|_{L^1(\mu_{i,k})} \text{ is bounded for each } i \in [d].$$

To see this, recall $\sup_n \| \chi_n - f_n^\oplus \|_{L^1(\mu_k^\otimes)}$ is bounded and so by (3.8),

$$\| M_k - f_n^\oplus \|_{L^1(\mu_k^\otimes)} = \| M_k - \chi_n \|_{L^1(\mu_k^\otimes)} + \| \chi_n - f_n^\oplus \|_{L^1(\mu_k^\otimes)} \leq M_k + C =: C.$$

The constant C may depend on k but not on n. From this and Jensen,

$$(3.9) \qquad |v_{1,k,n} + v_{2,k,n} + ... + v_{d,k,n}| \leq \| f_n^\oplus \|_{L^1(\mu_k^\otimes)} \leq C \text{ for all } n.$$

Next, note that as $f_n^\oplus \leq M_k$ on J_k, by taking supremum we see that

$$\sum_{i=1}^d \sup_{x_i \in J_{i,k}} f_{i,n}(x_i) \leq M_k \quad \forall n,$$

and note that obviously $v_{i,k,n} \leq \sup_{x_i \in J_{i,k}} f_{i,n}(x_i)$, so in particular

$$\sup_{x_1 \in J_{1,k}} f_{1,n}(x_1) + \sum_{i=2}^d v_{i,k,n} \leq M_k.$$

Define $\hat{v}_{1,k,n} = -\sum_{i=2}^d v_{i,k,n}$ and observe

$$C \geq \| M_k - f_n^\oplus \|_{L^1(\mu_k^\otimes)} = M_k - \int (f_{1,n} + \sum_{i=2}^d v_{i,k,n})\, d\mu_{1,k}.$$

This implies that $\sup_n \| f_{1,n} - \hat{v}_{1,k,n} \|_{L^1(\mu_{1,k})}$ is bounded, and then by (3.9), $\sup_n \| f_{1,n} - v_{1,k,n} \|_{L^1(\mu_{1,k})}$ is bounded. The claim is proved.

We are now ready to apply the Komlós lemma, which states that every L^1-bounded sequence of real functions contains a subsequence such that the arithmetic means of all its subsequences converge pointwise almost everywhere. Define $\tilde{v}_{1,k,n} = \frac{1}{n}\sum_{m=1}^n \hat{v}_{1,k,m}$, and $\tilde{v}_{i,k,n} = \frac{1}{n}\sum_{m=1}^n v_{i,k,m}$ for $2 \leq i \leq d$. Observe that by repeated use of Komlós lemma, for every $i \in [d]$ and $k \in \mathbb{N}$, we can find a subsequence $\{f_{i,k,n}\}_n$ of $\{f_{i,n}\}_n$ such that

    (1) $\{f_{i,k+1,n}\}_n$ is a further susequence of $\{f_{i,k,n}\}_n$, and
    (2) $\tilde{f}_{i,k,n}(x_i) - \tilde{v}_{i,k,n}$ converges for $\mu_{i,k}$ - a.e. x_i as $n \to \infty$,

where $\tilde{f}_{i,k,n}(x) = \frac{1}{n}\sum_{m=1}^{n} f_{i,k,m}(x)$. Select the diagonal sequence $F_{i,n} := f_{i,n,n}$ and again define $w_{i,k,n} = \int F_{i,n} d\mu_{i,k}$ for $2 \leq i \leq d$, $\hat{w}_{1,k,n} = -\sum_{i=2}^{d} w_{i,k,n}$, and $\tilde{F}_{i,n}(x_i) = \frac{1}{n}\sum_{m=1}^{n} F_{i,m}(x_i)$, $\tilde{w}_{1,k,n} = \frac{1}{n}\sum_{m=1}^{n} \hat{w}_{1,k,m}$, and $\tilde{w}_{i,k,n} = \frac{1}{n}\sum_{m=1}^{n} w_{i,k,m}$ for $2 \leq i \leq d$. We finally claim that

$$(3.10) \quad \tilde{F}_{i,n}(x_i) - \tilde{w}_{i,1,n} \text{ converges for } \mu_i - a.e.\, x_i \text{ for every } i \in [d].$$

The point is that the dependence on k is now removed. To see this, as $\{F_{i,n}\}_n$ is a subsequence of $\{f_{i,k,n}\}_n$ for every $k \in \mathbb{N}$, by Komlós lemma

$$(3.11) \quad \tilde{F}_{i,n}(x_i) - \tilde{w}_{i,k,n} \text{ converges for } \mu_{i,k} - a.e.\, x_i \text{ for every } i \in [d].$$

In particular, both $\{\tilde{F}_{i,n}(x_i) - \tilde{w}_{i,1,n}\}_n$ and $\{\tilde{F}_{i,n}(x_i) - \tilde{w}_{i,k,n}\}_n$ converge for $\mu_{i,1}$ - a.e. x_i as $n \to \infty$, hence their difference $\{\tilde{w}_{i,1,n} - \tilde{w}_{i,k,n}\}_n$ must converge for any fixed k. With (3.11) this implies (3.10).

Now having an approximating dual maximizer $(F_{i,n}, G_{i,n}, H_{i,n})_n$ where G, H followed the same subsequence as F did (and noting that (3.2) and (3.3) are conserved while taking convex combinations), we can repeat the same procedure for $G_{i,n}$ by starting from the uniform bound $\| G_n^\oplus - \chi_n \|_{L^1(\nu^\otimes)} \leq C$ so that we get a further subsequence for which similar statement as in (3.10) holds for $G_{i,n}$ as well. Then a final application of Komlós lemma (recall $\sum_i \tilde{w}_{i,1,n} = 0$) yields an approximating dual maximizer $(\tilde{F}_{i,n}, \tilde{G}_{i,n}, \tilde{H}_{i,n})_n$ which satisfies the claimed convergence property.

Step 3. We've shown there exists a sequence of functions $(f_{i,n}, g_{i,n}, h_n)_n$ satisfying (3.2), (3.3) for each n, and admitting the limit functions $(f_i, g_i)_{i\in[d]}$ such that as $n \to \infty$, $f_{i,n}(x_i) \to f_i(x_i)$ for μ_i - a.e. x_i and $g_{i,n}(y_i) \to g_i(y_i)$ for ν_i - a.e. y_i. For convenience, define A_i and B_i to be the set of convergence points, that is $\mu_i(A_i) = 1$, $\nu_i(B_i) = 1$, and

$$\lim_{n\to\infty} f_{i,n}(x_i) = f_i(x_i) \in \mathbb{R}, \ \lim_{n\to\infty} g_{i,n}(y_i) = g_i(y_i) \in \mathbb{R} \quad \forall x_i \in A_i,\, y_i \in B_i.$$

Let $A = A_1 \times ... \times A_d$, $B = B_1 \times ... \times B_d$. Note that $A \subseteq I$, $B \subseteq J$, and moreover the interior of the convex hull of B is I, i.e. $\mathrm{int}(\mathrm{conv}(B)) = I$.

In this step we will show the convergence of $\{\chi_n\}_n$ defined in (3.4). Fix $a \in A$ so that $\lim_{n\to\infty} f_n^\oplus(a) = f^\oplus(a) =: \sum_i f_i(a)$. As $\mathrm{int}(\mathrm{conv}(B)) = I$, we can find points in B, say $\{y_1, ..., y_m\}$, such that $\lim_{n\to\infty} g_n^\oplus(y_j) = g^\oplus(y_j)$ for every $1 \leq j \leq m$ and $a \in \mathrm{int}(\mathrm{conv}(\{y_1, ..., y_m\}))$. Then in view of (3.5) we see that both $\{\chi_n(a)\}_n$ and $\{\nabla\chi_n(a)\}_n$ are uniformly bounded in n, where $\nabla\chi_n(a) \in \partial\chi_n(a)$ is a subgradient of the convex function χ_n at a. Hence by taking a subsequence, we can assume that $\{\chi_n(a)\}_n$ and $\{\nabla\chi_n(a)\}_n$ both converge. Now define the affine function $L_n(y) = \chi_n(a) + \nabla\chi_n(a) \cdot (y - a)$ and replace $f_n^\oplus(x)$ with $f_n^\oplus(x) - L_n(x)$,

$g_n^\oplus(y)$ with $g_n^\oplus(y) - L_n(y)$ and $h_n(x)$ with $h_n(x) - \nabla\chi_n(a)$. Then we get $\chi_n(a) = \nabla\chi_n(a) = 0$ while the convergence property of $f_{i,n}, g_{i,n}$ is retained. And (3.7) and Proposition 3.1 yield $\lim_{n\to\infty}\chi_n = \chi$ on J.

Step 4. We will show there exists a function $h : I \to \mathbb{R}^d$ such that

$$(3.12) \quad \sum_{i\in[d]} \big(f_i(x_i) - g_i(y_i)\big) + h(x)\cdot(y-x) \le c(x,y) \quad \forall x,y \in \mathbb{R}^d.$$

We may define $f_i = -\infty$ on $\mathbb{R}\setminus A_i$ and $g_i = +\infty$ on $\mathbb{R}\setminus B_i$ so that they are defined everywhere on $\mathbb{R}$. For a function φ defined on a subset of $\mathbb{R}^d$ and bounded below by an affine function, let $\mathrm{conv}[\varphi] : \mathbb{R}^d \to \mathbb{R}\cup\{\infty\}$ denote the convex envelope of φ. Define

$$(3.13) \qquad H(x,y) := \mathrm{conv}[c(x,\cdot) + g^\oplus(\cdot)](y).$$

Then any measurable choice h satisfying $h(x) \in \partial H(x,\cdot)(x)$ will satisfy (3.12). To see this, we may argue similarly as [7]. Recall

$$f_n^\oplus(x) + h_n(x)\cdot(y-x) \le c(x,y) + g_n^\oplus(y).$$

If we define $H_n(x,y) := \mathrm{conv}[c(x,\cdot) + g_n^\oplus(\cdot)](y)$, then we have

$$f_n^\oplus(x) + h_n(x)\cdot(y-x) \le H_n(x,y) \le c(x,y) + g_n^\oplus(y).$$

In particular by taking $y = x$, we get $f_n^\oplus(x) \le H_n(x,x)$. Next, observe that since the $\limsup$ of convex functions is convex, we have

$$\limsup_{n\to\infty} H_n(x,y) \le \mathrm{conv}[\limsup_{n\to\infty}\big(c(x,\cdot) + g_n^\oplus(\cdot)\big)](y)$$
$$= \mathrm{conv}[c(x,\cdot) + g^\oplus(\cdot)](y) = H(x,y).$$

Then by the convergence $f_n^\oplus \to f$ and the definition of $H(x,y)$, we get

$$f(x) \le H(x,x), \text{ and } H(x,y) \le c(x,y) + g^\oplus(y).$$

Since $\mathrm{int}(\mathrm{conv}(B)) = I$ and $g^\oplus$ is real-valued on B, for each $x \in I$ the convex function $y \mapsto H(x,y)$ is real-valued thus continuous in I. The subdifferential $\partial H(x,\cdot)(y)$ is then nonempty, convex and compact for every $y \in I$. Hence we can choose a measurable function $h : I \to \mathbb{R}^d$ satisfying $h(x) \in \partial H(x,\cdot)(x)$. Any such choice yields (3.12) as follows:

$$f^\oplus(x) + h(x)\cdot(y-x) \le H(x,x) + h(x)\cdot(y-x) \le H(x,y) \le c(x,y) + g^\oplus(y).$$

Step 5. We will show that for any function $h : I \to \mathbb{R}^d$ satisfying

$$(3.14) \qquad f^\oplus(x) - g^\oplus(y) + h(x)\cdot(y-x) \le c(x,y)$$

(we know that such a function exists by Step 4), and for any minimizer π^* for the problem (1.2), in fact we have

$$(3.15) \qquad f^\oplus(x) - g^\oplus(y) + h(x) \cdot (y - x) = c(x, y), \quad \pi^* - a.s..$$

In other words, every minimizer π^* is concentrated on the contact set

$$\Gamma := \{(x, y) \mid f^\oplus(x) - g^\oplus(y) + h(x) \cdot (y - x) = c(x, y)\}$$

whenever h is chosen to satisfy (3.14). This will complete the proof.

To begin, recall $f_{i,n} \to f_i$ μ_i-a.s., $g_{i,n} \to g_i$ ν_i-a.s., and $\chi_n \to \chi$ on J where χ_n is defined as (3.4) with χ being its limit. For any $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$ (not necessarily an optimizer), notice that $c(x, y) \in L^1(\pi)$ by the assumption of Theorem 1.2. Now we claim:

$$\limsup_{n \to \infty} \int \left(f_n^\oplus(x) - g_n^\oplus(y) + h_n(x) \cdot (y - x) \right) d\pi$$
$$\leq \int \left(f^\oplus(x) - g^\oplus(y) + h(x) \cdot (y - x) \right) d\pi.$$

First, let us observe how the claim implies (3.15). Let π^* be any minimizer for (1.2) (note that this exists by the assumption on the cost). Then $c(x, y) \in L^1(\pi^*)$ and $P(c) = \int c(x, y) \, d\pi^*$. Now we deduce

$$P(c) = \lim_{n \to \infty} \int \left(f_n^\oplus(x) - g_n^\oplus(y) + h_n(x) \cdot (y - x) \right) d\pi^*$$
$$\leq \int \left(f^\oplus(x) - g^\oplus(y) + h(x) \cdot (y - x) \right) d\pi^*$$
$$\leq \int c(x, y) \, d\pi^*$$
$$= P(c)$$

hence equality holds throughout, and this implies (3.15).

If we have pointwise convergence $h_n(x) \to h(x)$ then the claim would have been a simple consequence of Fatou's lemma, but we do not know such a convergence a priori. But [7] suggested a clever idea to handle this case and let us carry out a similar scheme in the current $\mathbb{R}^d$ setting.

Fix any $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$ and let $\pi^1 := p_\#^x \pi$, $\pi^2 := p_\#^y \pi$. Then as in Step 2 (but $\pi^1 \leq_c \pi^2$ instead of $\mu^\otimes \leq_c \nu^\otimes$), by (3.3) and (3.5) we get

$$\sup_n ||\chi_n - f_n^\oplus||_{L^1(\pi^1)} < \infty, \quad \sup_n ||g_n^\oplus - \chi_n||_{L^1(\pi^2)} < \infty.$$

From this, as $f_n^\oplus \to f^\oplus$, $g_n^\oplus \to g^\oplus$, $\chi_n \to \chi$, by Fatou's lemma we get

$$\chi - f^\oplus \in L^1(\pi^1), \quad g^\oplus - \chi \in L^1(\pi^2),$$

$$\limsup_{n\to\infty} \int (f_n^\oplus - \chi_n)\, d\pi^1 \leq \int (f^\oplus - \chi)\, d\pi^1,$$

$$\limsup_{n\to\infty} \int (\chi_n - g_n^\oplus)\, d\pi^2 \leq \int (\chi - g^\oplus)\, d\pi^2.$$

This allows us to proceed

$$\limsup_{n\to\infty} \int \left(f_n^\oplus(x) - g_n^\oplus(y) + h_n(x) \cdot (y - x) \right) d\pi$$

$$= \limsup_{n\to\infty} \int \left(f_n^\oplus(x) - \chi_n(x) + \chi_n(y) - g_n^\oplus(y) \right.$$

$$\left. + \chi_n(x) - \chi_n(y) + h_n(x) \cdot (y - x) \right) d\pi$$

$$\leq \int (f^\oplus - \chi)\, d\pi^1 + \int (\chi - g^\oplus)\, d\pi^2$$

$$+ \limsup_{n\to\infty} \int \left(\chi_n(x) - \chi_n(y) + h_n(x) \cdot (y - x) \right) d\pi.$$

To handle the last term, let $\pi = \pi_x \otimes \pi^1$, and let $\xi_n : I \to \mathbb{R}^d$ be a sequence of functions satisfying $\xi_n(x) \in \partial \chi_n(x)$. Then we compute

$$\int \left(\chi_n(x) - \chi_n(y) + h_n(x) \cdot (y - x) \right) d\pi$$

$$= \iint \left(\chi_n(x) - \chi_n(y) + h_n(x) \cdot (y - x) \right) d\pi_x(y)\, d\pi^1(x)$$

$$= \iint \left(\chi_n(x) - \chi_n(y) + \xi_n(x) \cdot (y - x) \right) d\pi_x(y)\, d\pi^1(x),$$

since $\int h_n(x) \cdot (y - x)\, d\pi_x(y) = \int \xi_n(x) \cdot (y - x)\, d\pi_x(y) = 0$. Notice that the last integrand is nonpositive. Thus by Fatou's lemma,

$$\limsup_{n\to\infty} \int \left(\chi_n(x) - \chi_n(y) + h_n(x) \cdot (y - x) \right) d\pi$$

$$\leq \int \limsup_{n\to\infty} \left(\int \left(\chi_n(x) - \chi_n(y) + \xi_n(x) \cdot (y - x) \right) d\pi_x(y) \right) d\pi^1(x)$$

$$\leq \int \left(\int \left(\chi(x) - \chi(y) + \xi(x) \cdot (y - x) \right) d\pi_x(y) \right) d\pi^1(x)$$

for some $\xi(x) \in \partial\chi(x)$ which is a limit point of the bounded sequence $\{\xi_n(x)\}_n$. Finally, in the last line, the inner integral equals

$$\int \big(\chi(x) + h(x) \cdot (y - x) - \chi(y)\big) \, d\pi_x(y).$$

This proves the claim, hence the theorem. $\square$

Remark 3.2. *In Step 5 we deduced* $\chi - f^\oplus \in L^1(\pi^1)$, $g^\oplus - \chi \in L^1(\pi^2)$ *for every* $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$. *Is* $f_i \in L^1(\mu_i)$ *or* $g_i \in L^1(\nu_i)$? *Even in one-dimension and in a fairly mild situation such as* $c(x, y)$ *is* 1-*Lipschitz and* μ, ν *are irreducible and compactly supported, it can happen that neither* $f \in L^1(\mu)$ *nor* $g \in L^1(\nu)$, *as shown by examples in [7]. We may extend the notion of generalized integral of the pair* (f, g) *introduced in [7] to the* **VMOT** *setup via (compare with [7, Definition 4.7])*

$$(3.16) \quad \pi^1(f^\oplus) - \pi^2(g^\oplus) := \pi^1(f^\oplus - \chi) - \pi^2(g^\oplus - \chi) + (\pi^1 - \pi^2)(\chi).$$

For this definition to be meaningful, it is desired that the right hand side would not depend on the choice of π, *since this will then allow us to write the left hand side as e.g.* $\vec{\mu}(f^\oplus) - \vec{\nu}(g^\oplus)$. *Although this looks fairly plausible, we shall not pursue this point further in this paper.*

4. Further results on the structure of VMOT

In this section, we present two more structural results for **VMOT**. Recall the proof of Theorem 2.1 was based on the differential identity

$$\nabla_x \bigg(\sum_{i \in [d]} \big(f_i(x_i) - g_i(y_i) + h_i(x)(y_i - x_i)\big) - c(x, y) \bigg) = 0, \quad \pi^* - a.e. \, (x, y)$$

where π^* is a **VMOT** and $(f_i, -g_i, h_i)_i$ is a **PDM**. On the other hand, the next result shall be obtained via the differential identity in y variable:

$$\nabla_y \bigg(\sum_{i \in [d]} \big(f_i(x_i) - g_i(y_i) + h_i(x)(y_i - x_i)\big) - c(x, y) \bigg) = 0, \quad \pi^* - a.e. \, (x, y).$$

Theorem 4.1. *Assume the same as in Theorem 1.2. Let* S *be a proper subset of* $[d]$ *and assume that* ν_i *is absolutely continuous with respect to Lebesgue measure for every* $i \in S$. *Suppose* $c(x, y)$ *is semiconcave in* y *in the following sense: there exists* $u_i : J_i \to \mathbb{R}$ *for all* $i \in [d]$ *such that*

$$y \mapsto c(x, y) + \sum_i u_i(y_i) \text{ is concave on } J \text{ for every } x \in I.$$

Assume the following twist condition: for each $x \in I$ and $y_i \in J_i$ where $i \in S$, the mapping

$$(4.1) \qquad (y_j)_{j \in [d] \setminus S} \mapsto \left(\frac{\partial c(x,y)}{\partial y_i} \right)_{i \in S} \quad \text{is one-to-one on} \quad \bigotimes_{j \in [d] \setminus S} J_j$$

whenever the derivatives exist. Then there exists a family of functions $F_x : \mathbb{R}^{|S|} \to \mathbb{R}^{d-|S|}$ for each $x \in \mathbb{R}^d$ such that for any minimizer $\pi = \pi_x \otimes \pi^1$ of (1.2), π_x is concentrated on the graph of F_x for π^1-a.e. x.

We note that the definition of semiconcavity here is so broad that there is no regularity imposed on u_i other than that it be real-valued and measurable.

Proof. The same assumption as in Theorem 1.2 ensures there is a PDM. By the semiconcavity on the cost, we can assume that $y \mapsto c(x,y)$ is concave for every $x \in I$ for which $(f_i, -g_i, h_i)_i$ is a PDM, that is

$$f^\oplus(x) + h(x) \cdot (y - x) - c(x,y) \leq g^\oplus(y) \quad \forall x, y \in \mathbb{R}^d,$$

and for any minimizer $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$ of the primal problem (1.2),

$$f^\oplus(x) + h(x) \cdot (y - x) - c(x,y) = g^\oplus(y) \quad \pi - a.e. \, (x,y).$$

Consider the "inverse martingale Legendre transform" (see [32])

$$(4.2) \qquad \beta(y) := \sup_{x \in I} \{ f^\oplus(x) + h(x) \cdot (y - x) - c(x,y) \}.$$

Notice β is convex and $\beta \leq \sum_i g_i$. By replacing g_i's with its Legendre transform ψ_i's with respect to the "cost" β, that is, defining ψ_i successively for $i = 1, 2, ..., d$ by

$$\psi_i(y_i) := \sup_{y_j, j \neq i} \left(\beta(y) - \sum_{j < i} \psi_j(y_j) - \sum_{j > i} g_j(y_j) \right),$$

we find that ψ_i are convex, and

$$\beta(y) \leq \sum_{i=1}^d \psi_i(y_i) \quad \forall y \in \mathbb{R}^d, \quad \text{and} \quad \beta(y) = \sum_{i=1}^d \psi_i(y_i) \quad \pi^2 - a.e. \, y,$$

where $\pi^2 = p^y_\# \pi$ as usual. Let us define the contact set

$$G = \{ (x,y) \in \mathbb{R}^{2d} \, | \, f^\oplus(x) + h(x) \cdot (y - x) - c(x,y) = \beta(y) = \psi^\oplus(y) \}.$$

Let us slightly refine G as follows, and define its projections:

$$H = \{ (x,y) \in G \, | \, \psi_i \text{ is differentiable at } y_i \text{ for every } i \in S \}$$

$$X_H = \{ x \in \mathbb{R}^d \, | \, \exists y \text{ s.t. } (x,y) \in H \}, \quad Y_H = \{ y \in \mathbb{R}^d \, | \, \exists x \text{ s.t. } (x,y) \in H \}.$$

As $\nu_i \ll \mathcal{L}$ for all $i \in S$ and ψ_i is convex, $\pi(H) = 1$ for any VMOT π.

Now we claim that for any $(x, y) \in H$ and $i \in S$,

(4.3) $\qquad \dfrac{\partial \psi_i}{\partial y_i}(y_i), \ \dfrac{\partial \beta}{\partial y_i}(y)$ and $\dfrac{\partial c}{\partial y_i}(x, y)$ exist, and

(4.4) $\qquad \dfrac{\partial \psi_i}{\partial y_i}(y_i) = \dfrac{\partial \beta}{\partial y_i}(y) = h_i(x) - \dfrac{\partial c}{\partial y_i}(x, y).$

Of course $\frac{\partial \psi_i}{\partial y_i}(y_i)$ exists by definition of H. Also from the definition

$$\psi_i(y_i) = \sup_{y_j, j \neq i} \Big(\beta(y) - \sum_{j \neq i} \psi_j(y_j) \Big)$$

we see that the first equality in (4.4) holds since the supremum is attained at $y = (y_i)_i$ and both ψ_i and β are convex. Then in turn since the supremum in (4.2) is attained at (x, y) and both β and $y \mapsto h(x) \cdot (y - x) - c(x, y)$ are convex, the second equality holds as well.

For each $x \in X_H$ define the slice set $H_x := \{ y \, | \, (x, y) \in H \}$, and for $y = (y_1, ..., y_d) \in \mathbb{R}^d$ define the projection $y_S \in \mathbb{R}^{|S|}$ of y to be the collection of those y_i's with $i \in S$. Now by (4.4) and the twist condition (4.1), we see that if $y^1, y^2 \in H_x$ and $y_S^1 = y_S^2$, then we must have $y^1 = y^2$. This implies that we can define a function $F_x : \mathbb{R}^{|S|} \to \mathbb{R}^{d-|S|}$ such that the set H_x is contained in the graph of F_x for every $x \in X_H$. Finally as $\pi(H) = 1$, we have $\pi_x(H_x) = 1$ for π^1-a.e. x, completing the proof. $\qquad \square$

Remark 4.2. *In Theorem 4.1 if one could obtain a dual maximizer $(f_i, -g_i, h_i)_i$ where all g_i's and $y \mapsto c(x, y)$ are differentiable, then the family of maps $\{F_x\}_x$ could be directly obtained by the above differential identity in y; see the next example. But instead the semiconcavity was assumed in the theorem in order to deal with more general costs.*

Example 4.3. *Set $d = 2$ and $c(x, y) = -y_1 y_2$, so that we consider the covariance maximization problem of Y_1, Y_2. Let $(f_i, -g_i, h_i)$ be a PDM. In this case we can assume that g_1, g_2 are convex, since we can replace g_1 by the following Legendre transform $\tilde{g}_1$ (and similarly for g_2)*

$$\tilde{g}_1(y_1) := \sup_{x_1, x_2, y_2} \Big(\sum_{i=1}^{2} \big(f_i(x_i) + h_i(x_1, x_2) \cdot (y_i - x_i) \big) - c(x, y) - g_2(y_2) \Big).$$

So g_1, g_2 are differentiable a.e., and the differential identity in y_1 reads

$$y_2 = g_1'(y_1) - h_1(x_1, x_2)$$

which represents the function $y_2 = F_x(y_1)$ in Theorem 4.1. Interestingly, even though $c(x, y)$ has no dependence on x, F_x still depends on x via the "trading strategy" h_1. This is due to the randomness of X_1, X_2 and the martingale constraint $\mathbb{E}[Y|X] = X$. Note that if X_1, X_2 are nonrandom, then one can take $h_1 \equiv 0$ and the function becomes the

monotone map. Nonetheless, observe that even if X_1, X_2 are random, the dependence of the functions $y_2 = F_x(y_1)$ on x is mild; for any given marginal $(\vec{\mu}, \vec{\nu})$, their graphs are merely translations in y_2 direction.

Theorem 2.1 and Theorem 4.1 discussed the optimal structure of $(\pi_x)_x$ for VMOT $\pi = \pi_x \otimes \pi^1$. Would $\pi^1 = p^x_\# \pi$ and $\pi^2 = p^y_\# \pi$ also have some optimality property? To discuss, recall (2.3), (3.13) where we defined $y \mapsto H(x, y)$ to be the convex envelope of $y \mapsto c(x, y) + g^\oplus(y)$ and set $\alpha(x) = H(x, x)$. Also recall β is defined by (4.2). The following theorem shows the optimality of π^1, π^2 with respect to the costs α, β respectively from a classical duality theorem of Kellerer [42].

Theorem 4.4. *Let $\pi \in \mathrm{VMT}(\vec{\mu}, \vec{\nu})$ be a minimizer for (1.2) and let $(f_i, -g_i, h_i)_{i \in [d]}$ be a dual maximizer. Then π^1, π^2 satisfy the following:*

$$(4.5) \quad \sum_i f_i(x_i) \leq \alpha(x) \quad \forall x \in \mathbb{R}^d, \text{ and } \sum_i f_i(x_i) = \alpha(x) \quad \pi^1 - a.e.\, x,$$

$$(4.6) \quad \sum_i g_i(y_i) \geq \beta(y) \quad \forall y \in \mathbb{R}^d, \text{ and } \sum_i g_i(y_i) = \beta(y) \quad \pi^2 - a.e.\, y.$$

Proof. Recall $f^\oplus(x) + h(x) \cdot (y - x) \leq c(x, y) + g^\oplus(y)$. Then by convexity,

$$(4.7) \qquad f^\oplus(x) + h(x) \cdot (y - x) \leq H(x, y) \leq c(x, y) + g^\oplus(y).$$

By setting $y = x$ in the first inequality, we get $f^\oplus(x) \leq H(x, x) =: \alpha(x)$, and notice from the first inequality that if $f^\oplus(x) = \alpha(x)$ then $h(x)$ must belong to the subdifferential of the convex function $y \mapsto H(x, y)$ at x. Now since (4.7) must hold as equalities for π-a.e. (x, y) and $\delta_x \leq_c \pi_x$, we deduce that $f^\oplus(x) = \alpha(x)$ must hold π^1-a.e. x. This yields (4.5).

The proof of (4.6) is similar. By definition of β, we have

$$(4.8) \qquad f^\oplus(x) + h(x) \cdot (y - x) - c(x, y) \leq \beta(y) \leq g^\oplus(y)$$

and moreover (4.8) must hold as equalities π-a.s.. This yields (4.6). $\square$

Example 4.5. *In Example 2.2 and 2.3, the optimality of π did not imply any constrained structure on π^1, since π^1 could be any coupling of $(\mu_1, ..., \mu_d)$ there. In this example we shall see that for some data (i.e. cost and marginals), π^1 may have to be constrained. As in the previous example, let $d = 2$ and $c(x, y) = -y_1 y_2$.*

It is well known that the monotone coupling, which is the coupling concentrated on an increasing graph, minimizes the cost (i.e. maximizes $\mathbb{E}[Y_1 Y_2]$) among all couplings of ν_1, ν_2. But in our VMOT setting, the monotone coupling may not be feasible to be π^2 since we have to satisfy $\pi^1 \leq_c \pi^2$ but there may not exist such a π^1. Still, π^2 may try to be as close as possible to the monotone coupling, and this in turn may affect the structure of π^1. We explore this phenomenon with explicit examples.

Firstly, as a toy model let the marginals be those in Example 2.3:
$\mu_1 = \mu_2 = \text{Leb}\big|_{[-1/2,1/2]}$, $\nu_1 = \nu_2 = \frac{1}{2}\text{Leb}\big|_{[-1,1]}$ *and let π be a VMOT.*
As maximizing $\mathbb{E}[Y_1 Y_2]$ is equivalent to minimizing $\mathbb{E}\frac{1}{2}|Y_1 - Y_2|^2$, it is
better for π^2 to stay near the diagonal $\Delta := \{(y_1, y_2) \mid y_1 = y_2\}$. With
this simple data $(\vec{\mu}, \vec{\nu})$ it is feasible for π^2 to be supported on Δ, and
since $\pi^1 \leq_c \pi^2$, this implies π^1 must also be supported on Δ. This
uniquely determines π^1, π^2, and any martingale measure π connecting
π^1, π^2 is optimal. Thus a VMOT π is nonunique but π^1, π^2 are unique.

For another example, let μ_1, μ_2 be arbitrary but fixed. The main goal
is to build a VMOT π^* such that its first marginal π^{*1} is the monotone
coupling of arbitrarily given marginals μ_1, μ_2. To this end, set $g_1(y_1) =$
$\frac{1}{2}|y_1|^2$, $g_2(y_2) = \frac{1}{2}|y_2|^2$, so that $c(x,y) + g_1(y_1) + g_2(y_2) = \frac{1}{2}|y_1 - y_2|^2$. Now
since $\frac{1}{2}|y_1 - y_2|^2$ is already convex in y, the first derived cost α appearing
in Theorem 4.4 must be the same, i.e. $\alpha(x) = \frac{1}{2}|x_1 - x_2|^2$. This implies
$h(x) = \nabla\alpha(x) = (x_1 - x_2, x_2 - x_1)$, thus $y \mapsto \alpha(x) + h(x) \cdot (y - x)$ is
the tangent plane to $y \mapsto c(x,y) + g_1(y_1) + g_2(y_2)$ at x. In this simple
case, it is easy to see what the contact set is given by:

$$\begin{aligned} G :=& \{(x,y) \mid \alpha(x) + h(x) \cdot (y - x) = c(x,y) + g_1(y_1) + g_2(y_2)\} \\ =& \{(x,y) \mid y - x = \lambda(1,1) \text{ for some } \lambda \in \mathbb{R}\}. \end{aligned}$$

Now choose a martingale measure π^* which satisfies that $\pi^*(G) = 1$, for
which its first induced coupling $\pi^{*1} = p^x_\# \pi^*$ is the monotone coupling of
μ_1, μ_2. This implies there exist Kantorovich potentials f_1, f_2 such that

(4.9) $f_1(x_1) + f_2(x_2) \leq \alpha(x) \quad \forall x = (x_1, x_2)$, *and*

(4.10) $f_1(x_1) + f_2(x_2) = \alpha(x) \quad \pi^{*1} - a.e.\, x.$

Let ν_1^*, ν_2^* be the one-dimensional marginals of $\pi^{*2} = p^y_\# \pi^*$. Then the
fact $\pi^*(G) = 1$ and (4.9), (4.10) imply that π^* is optimal in the class
$\text{VMT}(\mu_1, \mu_2, \nu_1^*, \nu_2^*)$. Moreover, any optimal $\pi \in \text{VMT}(\mu_1, \mu_2, \nu_1^*, \nu_2^*)$
must satisfy $\pi^1 = \pi^{*1}$ since (4.9), (4.10) must hold for π^1 as well.

In this example the choices $g_i(y_i) = \frac{1}{2}|y_i|^2$ were made with no specific
reason, and due to this the π^* constructed there may not fit to a pre-
scribed marginal data $(\vec{\mu}, \vec{\nu})$. Theorem 1.2 asserts that for any cost and
irreducible marginal data one can obtain a dual optimizer such that
the contact set (1.6) can accommodate all VMOTs. Moreover, we have
illustrated that some careful analysis of the contact set can provide
information on the structure of VMOTs, as shown in Theorem 2.1, 4.1.

Lastly, motivated by this example, we address the following question:
what conditions on the data $(\vec{\mu}, \vec{\nu})$ and cost $c(x,y)$ shall impose on the

induced optimal couplings π^1, π^2 to have some specific structures, e.g. π^1 or π^2 has to lie on "small" sets? We leave this to future research.

APPENDIX A. COMPACTNESS OF THE CONVEX POTENTIALS

We prove Proposition 3.1 which we restate for reader's convenience.

Proposition A.1. *Let $(\mu_i, \nu_i)_{i \in [d]}$ be irreducible pairs of probabilities with domains $(I_i, J_i)_i$. Let $a \in I$, $C \in \mathbb{R}$. Consider the following class of functions $\Lambda = \Lambda(a, C, \vec{\mu}, \vec{\nu})$ where every $\chi \in \Lambda$ satisfies the following:*

(1) χ is a real-valued convex function on J,
(2) $\chi \geq 0$ and $\chi(a) = 0$,
(3) $\int \chi \, d(\nu^{\otimes} - \mu^{\otimes}) \leq C$.

Then Λ is locally bounded in the following sense: for each compact subset K of J, there exists $M = M(K)$ such that $\chi \leq M$ on K for every $\chi \in \Lambda$. Moreover, for any sequence $\{\chi_n\}_n$ in Λ there exists a subsequence $\{\chi_{n_j}\}_j$ of $\{\chi_n\}_n$ and a real-valued convex function χ on J such that $\lim_{j \to \infty} \chi_{n_j}(x) = \chi(x)$ for every $x \in J$, and the convergence is uniform on every compact subset of I.

The following class of two-way martingales will also be used often.

Definition A.2. *(1) We define $\zeta = \zeta_{x \to (y^-, y^+)}$ to be a simple martingale measure in $\mathbb{R} \times \mathbb{R}$ such that $p^x_{\#} \zeta$ is supported at a point (say x), and $p^y_{\#} \zeta$ is supported at two distinct points (say y^-, y^+). Then necessarily $y^- < x < y^+$, and ζ can be written as*

$$\zeta_{x \to (y^-, y^+)} = \frac{y^+ - x}{y^+ - y^-} \delta_{(x, y^-)} + \frac{x - y^-}{y^+ - y^-} \delta_{(x, y^+)}.$$

Then we define $\mathfrak{m}$ to be the set of all such simple martingale measures. (2) Let $\mathfrak{M}$ be the set of martingale measures in $\mathbb{R} \times \mathbb{R}$ such that any $\pi = \pi_x \otimes \pi^1 \in \mathfrak{M}$ satisfies $\pi_x \otimes \delta_x \in \mathfrak{m}$ for π^1–a.e. x, where $\pi^1 = p^x_{\#} \pi$. (3) We define $\mathfrak{M}(\mu, \nu) := \mathfrak{M} \cap \mathrm{MT}(\mu, \nu)$.

Proof of Proposition A.1. **Step 1.** In this step we will prove for the simplest marginal case where $d = 1$ and there exists $\zeta = \zeta_{x \to (y^-, y^+)} \in \mathfrak{m}$ such that $\mu = p^x_{\#} \zeta$ and $\nu = p^y_{\#} \zeta$, that is $\mu = \delta_x$ and $\nu = \frac{y^+ - x}{y^+ - y^-} \delta_{y^-} + \frac{x - y^-}{y^+ - y^-} \delta_{y^+}$. For $x, y \in \mathbb{R}$, $x \neq y$, define the open / closed transport rays

$$]\!]x, y[\![= \{(1-t)x + ty \mid 0 < t < 1\}, \quad [\![x, y]\!] = \{(1-t)x + ty \mid 0 \leq t \leq 1\}.$$

Suppose there is $a \in I =]\!]y^-, y^+[\![$ and $C > 0$ such that $\chi(a) = 0$, $\chi \geq 0$ and $\int \chi \, d(\nu - \mu) \leq C$ for every $\chi \in \Lambda$. First, consider the case $a = x$. Then it is trivial that there is a constant $M > 0$ such that $\chi(y^-) \leq M$ and $\chi(y^+) \leq M$ for all χ, since $\int \chi \, d(\nu - \mu) = \int \chi \, d\nu =$

$\frac{y^+-x}{y^+-y^-}\chi(y^-) + \frac{x-y^-}{y^+-y^-}\chi(y^+) \leq C$. By convexity of χ, we find $\chi \leq M$ on $[\![y^-, y^+]\!]$ for every $\chi \in \Lambda$, proving the theorem in this most simple case.

Next, consider the case that a is away from x, e.g. $y^- < a < x$. The idea is to find probability measures $\tilde{\mu}, \tilde{\nu}$ where $\mu \leq_c \tilde{\mu} \leq_c \tilde{\nu} \leq_c \nu$, and to find a small but positive constant θ (which does not depend on $\chi \in \Lambda$) such that the martingale measure $\eta := \theta\, \zeta_{a \to (y^-, y^+)}$ satisfies $\eta \leq \tilde{\pi}$ for some $\tilde{\pi} \in \mathrm{MT}(\tilde{\mu}, \tilde{\nu})$, i.e., for some martingale measure $\tilde{\pi}$ having marginals $\tilde{\mu}, \tilde{\nu}$. If we can do so, then since $\tilde{\pi} - \eta$ is still a martingale, we infer from $\int \chi\, d(\tilde{\nu} - \tilde{\mu}) \leq C$ that we have $\int \chi\, d(p^y_\# \eta - p^x_\# \eta) \leq C$. Then as $\eta := \theta\, \zeta_{a \to (y^-, y^+)}$, we are reduced to the previous case and the proposition follows. Now for this, we can simply take $\tilde{\nu} = \nu$ and $\tilde{\mu} = p^y_\# \zeta_{x \to (a, y^+)}$. Since $y^- < a < x$, clearly $\tilde{\mu} \leq_c \nu$. Then notice we can take $\theta = \tilde{\mu}(a) = \frac{y^+-x}{y^+-a}$, implying $\tilde{\pi} = \eta + \tilde{\mu}(y^+)\delta_{(y^+, y^+)}$; see figure (1a).

We can describe this procedure in terms of Brownian motion and stopping time, namely, let $(B^x_t)_t$ denote a Brownian motion starting at x, and define the stopping time $\tau = \inf\{t \geq 0 \,|\, B^x_t \notin]y^-, y^+[\}$. Then $\mathrm{Law}(B^x_\tau) = \nu$. Next, define $\tau^1 = \inf\{t \geq 0 \,|\, B^x_t \notin]a, y^+[\}$ and $\tilde{\mu} = \mathrm{Law}(B^x_{\tau^1})$. Then as $\tau^1 \leq \tau$ we have $\tilde{\mu} \leq_c \nu$. Finally, define τ^2 as the following: if $B^x_{\tau^1} = y^+$ then let $\tau^2 = \tau^1$, and if $B^x_{\tau^1} = a$ then let $\tau^2 = \tau^1 + \inf\{t \geq 0 \,|\, B^a_t \notin]y^-, y^+[\}$. We may define $\tilde{\nu} = \mathrm{Law}(B^x_{\tau^2})$, but in this case $\tilde{\nu} = \nu$ simply because $\tau^2 = \tau$.

The case $x < a < y^+$ can be treated similarly and Step 1 is done.

Step 2. In this step we will prove for the case where (μ, ν) are the marginals of a martingale measure of the form $\pi := \sum_{i=1}^N p_i \zeta_{x_i \to (y^-_i, y^+_i)}$. Here $N \geq 2$, $p_i > 0, \sum p_i = 1$, and the family $\zeta_{x_i \to (y^-_i, y^+_i)} \in \mathfrak{m}$, $i = 1, 2, ..., N$ is assumed to satisfy the following *chain condition*: if we let $l_k = \min_{i \leq k} y^-_i$, $r_k = \max_{i \leq k} y^+_i$, then

$$]\!] l_k, r_k [\![\, \cap \,]\!] y^-_{k+1}, y^+_{k+1} [\![\, \neq \, \emptyset, \quad \forall k = 1, ..., N-1.$$

The chain condition immediately implies that (μ, ν) is irreducible. As it is enough to prove this step with $N = 2$ let us assume $N = 2$. Without loss of generality assume $a \in]\!] y^-_1, y^+_1 [\![$ such that $\chi(a) = 0$, $\chi \geq 0$ and $\int \chi\, d(\nu - \mu) \leq C$ for every $\chi \in \Lambda$. Then since $p_1 \zeta_{x_1 \to (y^-_1, y^+_1)} \leq \pi$, by Step 1 we see that there exist $M_1, M'_1 > 0$ such that for every $\chi \in \Lambda$

    (1) $\chi \leq M_1$ on $[\![y^-_1, y^+_1]\!]$, and
    (2) $\nabla\chi \leq M'_1$ on $[\![y^-_1 + \epsilon, y^+_1 - \epsilon]\!]$, for any fixed $\epsilon > 0$.

M'_1 may depend on ϵ but notice it does not depend on $\chi \in \Lambda$. Note that (1) follows from Step 1 and (2) follows from (1) and the fact that χ is convex. Now take ϵ small enough such that

$$]\!] y^-_1 + \epsilon, y^+_1 - \epsilon [\![\, \cap \,]\!] y^-_2 + \epsilon, y^+_2 - \epsilon [\![\, \neq \, \emptyset.$$

(A) Step 1   (B) Step 3
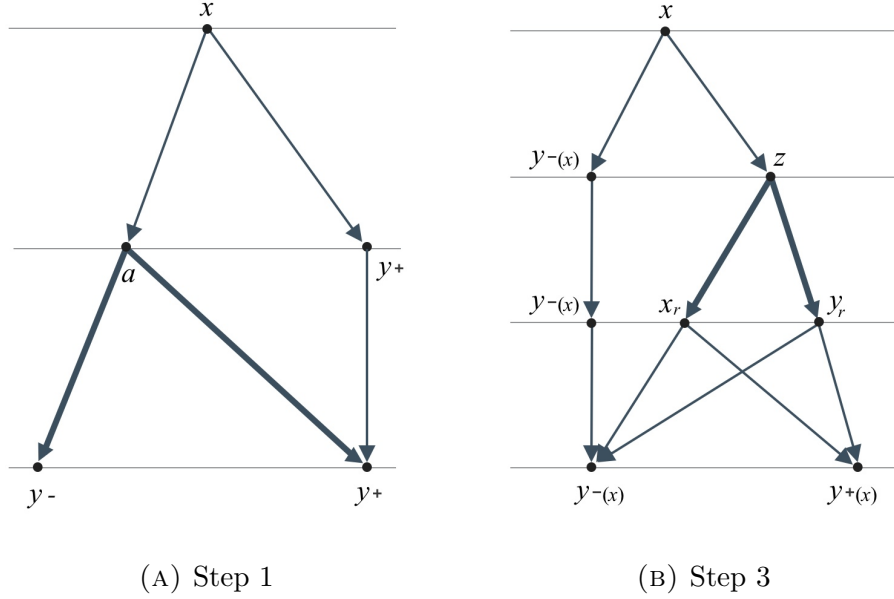
FIGURE 1. Martingale constructions in Step 1 and 3

Fix a point $b \in \,]\!]y_1^- + \epsilon, y_1^+ - \epsilon[\![\, \cap \,]\!]y_2^- + \epsilon, y_2^+ - \epsilon[\![$ and take any $\chi \in \Lambda$. Let L_χ be an affine function which supports the convex function χ at b, that is $L_\chi(b) = \chi(b), \nabla L_\chi(b) \in \partial \chi(b)$. Note that, by (1) and (2), $L_\chi(b)$ and $\nabla L_\chi(b)$ are bounded independently of χ. Let $\tilde{\chi} = \chi - L_\chi$. Then $\tilde{\chi}(b) = 0$, $\tilde{\chi} \geq 0$ and $\int \tilde{\chi} \, d(\nu - \mu) \leq C$, hence by the observation (1), (2) and the Step 1 we deduce there exists $M_2 > 0$ such that for every $\chi \in \Lambda$ we have $\chi \leq M_2$ on $[\![y_2^-, y_2^+]\!]$. Generalization to arbitrary $N \in \mathbb{N}$ is immediate through the chain condition, and this completes Step 2.

Step 3. In this step we will prove for a general irreducible pair (μ, ν) with domain (I, J) but still in dimension $d = 1$. In this case, it is well-known that there exists a probability measure $\tilde{\mu}$ such that $\tilde{\mu}$ is absolutely continuous with respect to Lebesgue measure, $\mu \leq_c \tilde{\mu} \leq_c \nu$, and $(\tilde{\mu}, \nu)$ is irreducible with the same domain (I, J). One way to see this is the following: consider u_μ, u_ν, the potential functions of μ, ν. They coincide outside of I while $u_\mu < u_\nu$ in I by irreducibility. Select a convex function u such that $u_\mu = u = u_\nu$ outside of I while $u_\mu < u < u_\nu$ in I. Then $\tilde{\mu}$ is taken to be the second derivative measure of u, and by selecting sufficiently smooth u one gets the desired absolute continuity.

Hence from now on we will assume without loss of generality that $\mu \ll \mathcal{L}^1$. Also note that we can assume $\mu \wedge \nu = 0$ since the pair

$(\mu - \mu \wedge \nu, \nu - \mu \wedge \nu)$ is irreducible with domain (I, J) and $\int \chi \, d\big((\nu - \mu \wedge \nu) - (\mu - \mu \wedge \nu)\big) \leq C$. Thus, we will assume $\mu \ll \mathcal{L}^1$ and $\mu \wedge \nu = 0$.

Our strategy is to reduce this situation to Step 1 and Step 2. For this, we begin by choosing a martingale measure $\pi \in \mathfrak{M}(\mu, \nu)$. We can obtain such a π by solving the MOT problem w.r.t. the cost $c(x, y) = \pm |x - y|$ [38], [39], [5]. Let $x \mapsto y^-(x)$, $x \mapsto y^+(x)$ be two functions supporting $\pi = \pi_x \otimes \mu$, that is $\pi_x = \frac{y^+(x) - x}{y^+(x) - y^-(x)} \delta_{y^-(x)} + \frac{x - y^-(x)}{y^+(x) - y^-(x)} \delta_{y^+(x)}$ μ - a.e. x.

Now set $\Gamma := \operatorname{supp} \pi$ such that $\pi(\Gamma) = 1$ and for any $(x, y) \in \Gamma$ and $r > 0$, the restricted measure $\pi\big|_{B(x,r) \times B(y,r)}$ has strictly positive mass by the property of support. Pick $(x_0, y_0) \in \Gamma$ and suppose $x_0 < y_0$. Fix $z \in]x_0, y_0[$ and let $r > 0$ be so small that $x_r := x_0 + r < z < y_r := y_0 - r$. Now we claim the following: there exist positive measures $\tilde{\mu}, \tilde{\nu}, \tilde{\mu}_A, \tilde{\nu}_A$ satisfying $\tilde{\mu}_A \leq_c \tilde{\mu} \leq_c \tilde{\nu} \leq_c \tilde{\nu}_A$, martingale measures $\tilde{\pi} \in \operatorname{MT}(\tilde{\mu}, \tilde{\nu})$, $\pi_A \in \operatorname{MT}(\tilde{\mu}_A, \tilde{\nu}_A)$ such that $\pi_A \leq \pi$, and a positive constant $\tilde{\theta}$ such that the measure $\eta := \tilde{\theta} \, \zeta_{z \to (x_r, y_r)}$ satisfies $\eta \leq \tilde{\pi}$. Notice this yields $\int \chi \, d(\nu - \mu) \geq \int \chi \, d(p^y_\# \eta - p^x_\# \eta)$ as desired. See figure (1b).

To see this, note that the set $A := \{x \in B(x_0, r) \mid y^+(x) \in B(y_0, r)\}$ has μ – positive measure since $\pi\big(B(x_0, r) \times B(y_0, r)\big) > 0$. For each $x \in A$, consider the martingale measure $\zeta_{x \to (y^-(x), y^+(x))}$. Recall $y^-(x) < x < x_r < z < y_r < y^+(x)$. Now we apply the same idea as in Step 1: Let $\tau_x = \inf\{t \geq 0 \mid B^x_t \notin]y^-(x), y^+(x)[\}$ and $\tau^1_x = \inf\{t \geq 0 \mid B^x_t \notin]y^-(x), z[\}$. Then define τ^2_x as follows: if $B^x_{\tau^1_x} = y^-(x)$ then $\tau^2_x = \tau^1_x$, and if $B^x_{\tau^1_x} = z$ then $\tau^2_x = \tau^1_x + \inf\{t \geq 0 \mid B^z_t \notin]x_r, y_r[\}$. Then clearly

$$(\text{A.1}) \qquad \delta_x \leq_c \operatorname{Law}(B^x_{\tau^1_x}) \leq_c \operatorname{Law}(B^x_{\tau^2_x}) \leq_c \operatorname{Law}(B^x_{\tau_x}).$$

With $\theta = \theta(x, z) := \operatorname{Law}(B^x_{\tau^1_x})(\{z\}) = \frac{x - y^-(x)}{z - y^-(x)}$, $\theta \, \zeta_{z \to (x_r, y_r)}$ satisfies

$$(\text{A.2}) \qquad \theta \, \zeta_{z \to (x_r, y_r)} \leq \operatorname{Law}(B^x_{\tau^1_x}, B^x_{\tau^2_x}),$$

where $\operatorname{Law}(B^x_{\tau^1_x}, B^x_{\tau^2_x})$ is the joint law of $B^x_{\tau^1_x}$ and $B^x_{\tau^2_x}$ hence is a martingale measure with marginals $\operatorname{Law}(B^x_{\tau^1_x})$, $\operatorname{Law}(B^x_{\tau^2_x})$. Finally, set

$$\tilde{\theta}(z) = \int_A \theta(x, z) \, d\mu(x).$$

Notice $\tilde{\theta}(z) > 0$. We claim that the measure $\tilde{\theta} \, \zeta_{z \to (x_r, y_r)}$ satisfies the desired property. To see this, consider the restricted measure $\pi_A := \pi\big|_{A \times \mathbb{R}} \in \operatorname{MT}(\tilde{\mu}_A, \tilde{\nu}_A)$ where $\tilde{\mu}_A := p^x_\# \pi_A$, $\tilde{\nu}_A := p^y_\# \pi_A$. Define $\tilde{\mu} = \int_A \operatorname{Law}(B^x_{\tau^1_x}) \, d\mu(x)$, $\tilde{\nu} = \int_A \operatorname{Law}(B^x_{\tau^2_x}) \, d\mu(x)$. Then in view of (A.1) we have $\tilde{\mu}_A \leq_c \tilde{\mu} \leq_c \tilde{\nu} \leq_c \tilde{\nu}_A$, and by (A.2) and $\eta := \tilde{\theta} \, \zeta_{z \to (x_r, y_r)}$ we have $\eta \leq \tilde{\pi}$ for some $\tilde{\pi} \in \operatorname{MT}(\tilde{\mu}, \tilde{\nu})$, where $\tilde{\pi}$ is driven by the Brownian motion B^x between times τ^1_x and τ^2_x for $\mu\big|_A$ - a.e. x.

In summary, we have shown that for any $(x, y) \in \Gamma$ and $z \in]x, y[$ and any sufficiently small $r > 0$, we can find $\tilde{\mu}, \tilde{\nu}, \tilde{\mu}_A, \tilde{\nu}_A$ where $\tilde{\mu}_A \leq_c \tilde{\mu} \leq_c \tilde{\nu} \leq_c \tilde{\nu}_A$ and martingale measures $\tilde{\pi} \in \mathrm{MT}(\tilde{\mu}, \tilde{\nu})$, $\pi_A \in \mathrm{MT}(\tilde{\mu}_A, \tilde{\nu}_A)$ such that $\pi_A \leq \pi$, and a positive constant $\tilde{\theta}$ such that the measure $\eta := \tilde{\theta} \, \zeta_{z \to (x_r, y_r)}$ satisfies $\eta \leq \tilde{\pi}$. Now we make the following observation. Since $\pi(\Gamma) = 1$, for any $z \in I$ there exists $(x_z, y_z) \in \Gamma$ such that $z \in]x_z, y_z[$, since otherwise we would have $u_\mu(z) = u_\nu(z)$, contradicting to the irreducibility of (μ, ν). We choose such $(x_z, y_z) \in \Gamma$ for each $z \in I$ so that $\{ \,]x_z, y_z[\, \}_{z \in I}$ is an open cover of I. Hence we can find a sequence $(z_n)_{n \in \mathbb{N}}$ such that $\{ \,]x_{z_n}, y_{z_n}[\, \}_{n \in \mathbb{N}}$ is an open cover of I, and if $l_k := \min_{n \leq k} x_{z_n}$, $r_k := \max_{n \leq k} y_{z_n}$, then $]l_k, r_k[\cap \,]x_{z_{k+1}}, y_{z_{k+1}}[\neq \emptyset$ for all $k \in \mathbb{N}$, namely, the chain condition.

Now Step 2 implies that for each compact interval $K \subseteq I$, there exists $M_K > 0$ such that $\chi \leq M_K$ on K for every $\chi \in \Lambda$, which is the content of the proposition when ν has no mass on the boundary of I. To see this, observe that for a compact K there is $N \in \mathbb{N}$ such that $\{ \,]x_{z_n}, y_{z_n}[\, \}_{n \leq N}$ is an open cover of K, so for small $r > 0$ the shrunken covering $\{ \,]x_{z_n} + r, y_{z_n} - r[\, \}_{n \leq N}$ still covers K and satisfies the chain condition, hence Step 2 applies. This completes Step 3 for the case $\nu(I) = 1$. Now the remaining case when ν assigns positive mass on the boundary of I is simple. Write $I =]a, b[$ and suppose $b \in \mathbb{R}$ and $\nu(b) > 0$. Recall that $\pi \in \mathfrak{M}(\mu, \nu)$ and $x \mapsto (y^-(x), y^+(x))$ is the graph of π. Let $A := \{x \mid y^+(x) = b\}$. $\nu(b) > 0$ implies $\mu(A) > 0$. Take any $x \in A$ satisfying $(x, b) \in \Gamma$. Then the previous argument shows that for any $r > 0$ and $z \in]x + r, b[$, we can find measures $\tilde{\mu}, \tilde{\nu}, \tilde{\mu}_A, \tilde{\nu}_A$ where $\tilde{\mu}_A \leq_c \tilde{\mu} \leq_c \tilde{\nu} \leq_c \tilde{\nu}_A$, martingale measures $\tilde{\pi} \in \mathrm{MT}(\tilde{\mu}, \tilde{\nu})$, $\pi_A \in \mathrm{MT}(\tilde{\mu}_A, \tilde{\nu}_A)$ such that $\pi_A \leq \pi$, and $\tilde{\theta} > 0$ such that $\eta := \tilde{\theta} \, \zeta_{z \to (x+r, b)}$ satisfies $\eta \leq \tilde{\pi}$. Step 2 then yields for any $t \in I$, there exists $M > 0$ such that $\chi \leq M$ on $[t, b]$ for every χ.

The case $\nu(a) > 0$ can be treated similarly, and finally note that if both $\nu(a) > 0$ and $\nu(b) > 0$ then we can conclude that there exists $M > 0$ such that $\chi \leq M$ on $[a, b]$ for every $\chi \in \Lambda$, because $[a, b]$ can be covered with finitely many appropriate intervals that we have discussed so far. The proof is therefore complete for the one-dimension case.

Step 4. In this step we will prove the local uniform bound $\chi \leq M$ for general $d \in \mathbb{N}$. We will see how the ideas from the previous steps can be applied in the same way, but with more notational difficulty; this is why we take a different approach than [7] in establishing the local uniform bound of χ for $d = 1$ in the previous steps.

Fix $d \in \mathbb{N}$. As in Step 1, we begin with the most simple case where for each $i \in [d]$ there is $\zeta_i = \zeta_{x_i \to (y_i^-, y_i^+)} \in \mathfrak{m}$ such that $\mu_i = p_\#^x \zeta_i$, $\nu_i = p_\#^y \zeta_i$, that is, $\mu_i = \delta_{x_i}$ and $\nu_i = \frac{y_i^+ - x_i}{y_i^+ - y_i^-} \delta_{y_i^-} + \frac{x_i - y_i^-}{y_i^+ - y_i^-} \delta_{y_i^+}$. Let $\mu := \otimes_i \mu_i$, $\nu := \otimes_i \nu_i$ and $\zeta := \otimes_i \zeta_i$. Note that the martingale measure $\zeta \in \mathcal{P}(\mathbb{R}^{2d})$ has μ, ν as its d-dimensional marginals. Now we want to prove that there exists $M > 0$ such that $\chi \le M$ on J for all $\chi \in \Lambda$, where $J = \mathrm{conv}(\mathrm{supp}\,\nu)$ is the convex hull of the support of ν. Notice that J is a closed rectangle in $\mathbb{R}^d$ and $\mathrm{supp}\,\nu$ is the set of its vertices, consisting of 2^d points.

Suppose there is $a \in I$ (where $I = \mathrm{int}\,J$) and $C > 0$ such that $\chi(a) = 0$, $\chi \ge 0$ and $\int \chi \, d(\nu - \mu) \le C$ for every $\chi \in \Lambda$. First, consider the case $a = x$, where $x = (x_1, ..., x_d)$. Then just as in Step 1, it is then trivial that there is a constant $M > 0$ such that $\chi \le M$ on $\mathrm{supp}\,\nu$. Then by convexity of χ we get $\chi \le M$ on J for every $\chi \in \Lambda$, proving the proposition in this simple case. Next if $x \ne a = (a_1, ..., a_d)$, recall that in Step 1 we found $\tilde{\mu}_i, \tilde{\nu}_i \in \mathcal{P}(\mathbb{R})$ such that $\mu_i \le_c \tilde{\mu}_i \le_c \tilde{\nu}_i \le_c \nu_i$ and a constant $\theta_i > 0$ such that the measure $\eta_i := \theta_i \, \zeta_{a_i \to (y_i^-, y_i^+)}$ satisfies $\eta_i \le \tilde{\pi}_i$ for some $\tilde{\pi}_i \in \mathrm{MT}(\tilde{\mu}_i, \tilde{\nu}_i)$. Set $\tilde{\mu}^\otimes := \otimes_i \tilde{\mu}_i$, $\tilde{\nu}^\otimes := \otimes_i \tilde{\nu}_i$ and $\eta^\otimes := \otimes_i \eta_i$. Then observe that $\int \chi \, d(\tilde{\nu}^\otimes - \tilde{\mu}^\otimes) \le C$ implies $\int \chi \, d(p_\#^y \eta^\otimes - p_\#^x \eta^\otimes) \le C$. Since $\mathrm{supp}(p_\#^x \eta^\otimes) = \{a\}$ and $\mathrm{supp}(p_\#^y \eta^\otimes) = \mathrm{supp}\,\nu$, we are reduced to the previous case and the local uniform bound follows.

Now notice we can carry out Step 2 and 3 in this higher dimension case exactly the same way, that is, we can find a countable rectangular martingale measures (say $(\zeta_n)_{n \in \mathbb{N}}$) which covers I and satisfies the appropriate chain condition. We saw in Step 2 that the boundedness property of χ on each interval $\mathrm{conv}(\mathrm{supp}\,\zeta_n)$ can propagate along such a chain, and the argument works the same way in higher dimension, simply observing that the intersecting intervals $]\!]y_1^- + \epsilon, y_1^+ - \epsilon[\![\cap]\!]y_2^- + \epsilon, y_2^+ - \epsilon[\![\ne \emptyset$ in Step 2 are now replaced by intersecting chain of rectangles. Moreover if some marginal measures ν_i assign positive mass on the boundary of I_i, then observe that the rectangles can cover up to such boundaries so that we get the desired boundedness of all $\chi \in \Lambda$ up to the boundary. This completes the proof of the local boundedness.

Step 5. It remains to prove the convergence property, and it is a direct consequence of the local bound and an application of Arzelà-Ascoli theorem, as follows: we have shown that every $\chi \in \Lambda$ is uniformly bounded on any compact subset of J, so with convexity of χ and $\chi(a) = 0$, $\chi \ge 0$ we deduce that the derivative of every $\chi \in \Lambda$ is also uniformly bounded on any compact subset of I. Hence by Arzelà-Ascoli theorem we deduce that for any sequence $\{\chi_n\}$ in Λ and any compact set K

in I, there exists a subsequence of $\{\chi_n\}$ that converges uniformly on K. Now by increasing K to I and using diagonal argument, we can find a further subsequence of $\{\chi_n\}$ that converges pointwise in I, and moreover, the convergence is uniform on every compact subset of I.

We can deal with the convergence on $J \setminus I$ the same way and let us explain in dimension 2 to avoid notational difficulty. Suppose, for example, $J_1 = [a_1, b_1[$, $J_2 = [a_2, b_2[$, and $J = J_1 \times J_2$. First of all, at the corner point (a_1, a_2) the sequence $\{\chi_n\}$ is bounded, so we can choose a subsequence which converges at (a_1, a_2). Next, on any compact interval in the open line $\{(x, a_2) \,|\, a_1 < x < b_1\}$ the sequence $\{\chi_n\}$ is also uniformly bounded, so by Arzelà-Ascoli theorem with diagonal selection again we can find a further subsequence which converges everywhere on $\{(x, a_2) \,|\, a_1 < x < b_1\}$. Finally, we can find a further subsequence which converges on $\{(a_1, y) \,|\, a_2 < y < b_2\}$, and hence the sequence converges everywhere on J. This completes the proof of Proposition A.1. $\qquad\square$

References

[1] L. Ambrosio, B. Kirchheim, and A. Pratelli. Existence of optimal transport maps for crystalline norms. *Duke Mathematical Journal.* Volume 125, Number 2 (2004) 207–241.

[2] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. *Optimal transportation and applications. Springer Berlin Heidelberg.*, Volume 1813 of the series Lecture Notes in Mathematics (2003) 123–160.

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

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

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

[6] M. Beiglböck, T. Lim and J. Obłój. Dual attainment for the martingale transport problem. *Bernoulli.* 25 (2019), no. 3, 1640–1658.

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

[8] S. Bianchini and F. Cavalletti. The Monge Problem for Distance Cost in Geodesic Spaces. *Comm. Math. Phys.*, Volume 318, Issue 3 (2013) 615–673.

[9] D.T. Breeden and R.H. Litzenberger. Prices of state-contingent claims implicit in option prices. *J. Business*, 51(4):621–651 (1978).

[10] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, Volume 44, Issue 4 (1991) 375–417.

[11] L.A. Caffarelli, M. Feldman, and R. J. McCann. Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *J. Amer. Math. Soc.*, 15 (2002) 1–26.

[12] G. Carlier, V. Chernozhukov and A. Galichon. Vector quantile regression: an optimal transport approach. *The Annals of Statistics*, 2016, Vol. 44, No. 3, 1165–1192.

[13] L. Carraro, N.E. Karoui and J. Obłój. On Azéma-Yor processes, their optimal properties and the Bachelier-drawdown equation. *Ann. Probab.*, Volume 40, Number 1 (2012) 372–400.

[14] T. Champion and L. De Pascale. The Monge problem for strictly convex norms in $\mathbb{R}^d$. *Journal of the European Mathematical Society.*, 12(6) (2010) 1355–1369.

[15] T. Champion and L. De Pascale. The Monge problem in $\mathbb{R}^d$. *Duke Mathematical Journal.*, Volume 157, Number 3 (2011) 551–572.

[16] V. Chernozhukov, I. Fernandez-Val and A. Galichon. Quantile and Probability Curves without Crossing. *Econometrica* 78(3), pp. 1093–1125 (2010).

[17] V. Chernozhukov, A. Galichon, M. Hallin, and M. Henry. Monge-Kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 2017, Vol. 45, No. 1, 223–256.

[18] V. Chernozhukov, A. Galichon, M. Henry, and B. Pass. Identification of hedonic equilibrium and nonseperable simultaneous equations. *J. Political Econ.*, Volume 129, Number 3 (2021).

[19] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger. A Riemannian interpolation inequality á la Borell, Brascamp and Lieb. *Invent. Math.*, 146, (2), (2001), 219–257.

[20] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger. Prékopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport. *Ann. Fac. Sci. Toulouse Math.*, 6, (15), (2006), 613–635.

[21] A.M.G. Cox, J. Obłój and N. Touzi. The Root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach. *Probab. Theory Related Fields.* 173 (2019), no. 1-2, 211–259.

[22] C. Decker, E.H. Lieb, R.J. McCann and B.K. Stephens. Unique equilibria and substitution effects in a stochastic model of the marriage market. *J. Econom. Theory* 148 (2013) 778–792.

[23] H. De March. Local structure of multi-dimensional martingale optimal transport. arXiv preprint. https://arxiv.org/abs/1805.09469

[24] H. De March. Quasi-sure duality for multi-dimensional martingale optimal transport. arXiv preprint. https://arxiv.org/abs/1805.01757

[25] H. De March and N. Touzi. Irreducible convex paving for decomposition of multidimensional martingale transport plans. *Ann. Probab.* 47 (2019), No. 3, 1726–1774.

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

[27] S. Eckstein, G. Guo, T. Lim, and J. Obłój. Robust Pricing and Hedging of Options on Multiple Assets and Its Numerics. *SIAM J. Financial Math.*, 12 (2021), no. 1, 158–188.

[28] L.C. Evans. Partial differential equations and Monge-Kantorovich mass transfer. *Current Developments in Mathematics.*, (1997) International Press.

[29] A. Figalli,Y-H Kim and R.J. McCann. When is multidimensional screening a convex program? *J. Econom. Theory* 146 (2011) 454–478.

[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. *Annals of Applied Probability.*, Volume 24, Number 1 (2014) 312–336.

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

[32] N. Ghoussoub, Y-H. Kim, and T. Lim. Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.* 47(1): 109–164 (2019).

[33] N. Ghoussoub, Y-H. Kim, and T. Lim. Optimal Brownian stopping when the source and target are radially symmetric distributions. *SIAM J. Control Optim.* 58 (2020), no. 5, 2765–2789.

[34] G. Guo, X. Tan and N. Touzi. On the monotonicity principle of optimal Skorokhod embedding problem. *SIAM J. Control Optim.*, 54-5 (2016) 2478–2489.

[35] G. Guo, X. Tan and N. Touzi. Optimal Skorokhod embedding under finitely-many marginal constraints. *SIAM J. Control Optim.*, 54-4 (2016) 2174–2201.

[36] D. Hobson. Robust hedging of the lookback option. *Finance and Stochastics.*, 2 (1998) 329–347.

[37] D. Hobson. The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, Springer, Berlin (2011) 267–318.

[38] D. Hobson and M. Klimmek. Robust price bounds for the forward starting straddle. *Finance and Stochastics.*, Volume 19, Issue 1 (2014) 189–214.

[39] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Mathematical Finance.*, Volume 22, Issue 1 (2012) 31–56.

[40] L. V. Kantorovich. On the transfer of masses. *Dokl. Akad. Nauk.*, SSSR 37 (1942) 227–229 (Russian).

[41] L. V. Kantorovich. On a problem of Monge. *Uspekhi Mat. Nauk.*, 3 (1948) 225–226.

[42] H. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete.*, 67(4) (1984) 399–432.

[43] T. Lim. Optimal martingale transport between radially symmetric marginals in general dimensions. *Stochastic Processes and their Applications.* Volume 130, Issue 4 (2020) 1897–1912.

[44] X. N. Ma, N. S. Trudinger, and X. J. Wang. Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.*, 177 (2005) no. 2, 151–183.

[45] R. J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128, 153–179 (1997).

[46] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'académie Royale des Sciences de Paris* (1781).

[47] J. Obłój. The Skorokhod embedding problem and its offspring. *Probability Surveys*, 1 (2004) 321–392.

[48] J. Obłój and P. Siorpaes. Structure of martingale transports in finite dimensions. arXiv preprint. https://arxiv.org/abs/1702.08433

[49] J. Obłój and P. Spoida. An iterated Azéma-Yor type embedding for finitely many marginals. *Ann. Probab.*, 45(4): 2210–2247 (2017).

[50] B. Pass. Convexity and multi-dimensional screening for spaces with different dimensions. *J. Econom. Theory.* 147 (2012) 2399–2418.

[51] R.T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific J. Math*, 17 (1966) 497–510.

[52] L. Rüschendorf. Construction of multivariate distributions with marginals Given. *Annals of the Institute of Statistical Mathematics.*, 37 (1), (1985) 225–233.

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

[54] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36 (1965) 423-439.

[55] N. S. Trudinger and X. J. Wang. On strict convexity and continuous differentiability of potential functions in optimal transportation. *Arch. Ration. Mech. Anal.*, 192 (2009) no. 3, 403–418.

[56] C. Villani. *Topics in optimal transportation*, Volume 58, *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI (2003).

[57] C. Villani. *Optimal Transport. Old and New*, Vol. 338, *Grundlehren der mathematischen Wissenschaften.* Springer (2009).

[58] D. Zaev. On the Monge-Kantorovich problem with additional linear constraints. *Mathematical Notes.*, 98(5-6) (2015) 725–741.

Tongseok Lim: Mitchell E. Daniels, Jr. School of Business
Purdue University, West Lafayette, Indiana 47907, USA
    Email address: lim336@purdue.edu