The Annals of Probability
2019, Vol. 47, No. 1, 109–164
https://doi.org/10.1214/18-AOP1258

STRUCTURE OF OPTIMAL MARTINGALE TRANSPORT PLANS IN GENERAL DIMENSIONS

BY NASSIF GHOUSSOUB[*,1], YOUNG-HEON KIM[*,†,1,2] AND TONGSEOK LIM[‡,3]

University of British Columbia[*], *Korea Institute for Advanced Study*[†] *and ShanghaiTech University*[‡]

Given two probability measures μ and ν in "convex order" on $\mathbb{R}^d$, we study the profile of one-step martingale plans π on $\mathbb{R}^d \times \mathbb{R}^d$ that optimize the expected value of the modulus of their increment among all martingales having μ and ν as marginals. While there is a great deal of results for the real line (i.e., when $d = 1$), much less is known in the richer and more delicate higher-dimensional case that we tackle in this paper. We show that many structural results can be obtained, provided the initial measure μ is absolutely continuous with respect to the Lebesgue measure. One such a property is that μ-almost every x in $\mathbb{R}^d$ is transported by the optimal martingale plan into a probability measure π_x concentrated on the extreme points of the closed convex hull of its support. This will be established for the distance cost $c(x, y) = |x - y|$ in the two-dimensional case, and also for any $d \geq 3$ as long as the marginals are in "subharmonic order." In some cases, π_x is supported on the vertices of a $k(x)$-dimensional polytope, such as when the target measure is discrete. Duality plays a crucial role in our approach, even though, in contrast to standard optimal transports, the dual extremal problem may not be attained in general. We show however that "martingale supporting" Borel subsets of $\mathbb{R}^d \times \mathbb{R}^d$ can be decomposed into a collection of mutually disjoint components by means of a "convex paving" of the source space, in such a way that when the martingale is optimal for a general cost function, each of the components then supports a restricted optimal martingale transport whose dual problem is attained. This decomposition is used to obtain structural results in cases where global duality is not attained. On the other hand, it shows that certain "optimal martingale supporting" Borel sets can be viewed as higher-dimensional versions of Nikodym-type sets. The paper focuses on the distance cost, but much of the results hold for general Lipschitz cost functions.

Received April 2017; revised February 2018.

[1]Supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC).

[2]Supported by an Alfred P. Sloan Research Fellowship.

[3]Supported in part by a doctoral graduate fellowship from the University of British Columbia; also supported by the Austrian Science Foundation (FWF) through grant Y782 and from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 335421. Also supported by ShanghaiTech University.

MSC2010 subject classifications. Primary 49N15, 60G42; secondary 52A40.

Key words and phrases. Optimal transport, martingale, Choquet boundary, duality, convex paving.

CONTENTS

1. Introduction. We study the profile of one-step martingales π on $\mathbb{R}^d \times \mathbb{R}^d$ that optimize the expected value of the modulus of their increment, among all martingales with two given marginals μ and ν in convex order. More precisely, we investigate the structure of conditional probabilities $(\pi_x)_{x \in \operatorname{supp} \mu}$ on $\mathbb{R}^d$ which describe how a given particle at x is propagated under such transport plans. These questions originate in mathematical finance and are variations on the original Monge–Kantorovich problem, where one considers all couplings of the given marginals and not only those of martingale type [7, 15, 22, 26, 28, 29]. However, unlike solutions of the Monge–Kantorovich problem, which are often supported on graphs (such as the well-known Brenier solution [7] for the cost given by the squared distance), the additional martingale constraint forces the transport to split the elements of the initial measure μ. One cannot therefore expect—but in trivial cases—that optimal martingale plans be supported on graphs.

These questions are motivated by problems in mathematical finance, which call for no-arbitrage lower (or upper) bounds on the price of a forward starting straddle, given today's vanilla call prices at the two relevant maturities. Just like in the Monge–Kantorovich theory for optimal transport, these problems have dual counterparts, whose financial interpretation amounts to constructing the most (or least) expensive semistatic hedging strategy which subreplicates the payoff of the forward starting straddle for any realization of the underlying forward price process.

The minimization and maximization problems are quite different, though by now well understood, when the marginals are probability measures on the real line, at least in the case of one-step martingales. We refer to Hobson–Neuberger [19], Hobson–Klimmek [18] and Beiglböck–Juillet [5]. For the multi-step case, see Beiglböck et al. [3]. The dynamic case have been also studied by Galichon et al. [14] and Dolinsky–Soner [10, 11]. The two cases studied are when the cost is either $c(x, y) = |x - y|$, which is the main focus of this paper, or the case when the cost satisfies the so-called generalized Spence–Mirrlees condition. Note that the one-dimensional case is closely related to Skorohod embedding problems [23], since real-valued martingales can be realized as adequately stopped Brownian paths; see, for example, Hobson [17], Beiglböck et al. [2] and Beiglböck et al. [4].

Surprisingly, much less is known in the case where the marginals are supported on higher-dimensional Euclidean spaces $\mathbb{R}^d$. In this direction, Lim [21] considered the optimal martingale transport problem under radially symmetric marginals on $\mathbb{R}^d$, while Ghoussoub et al. consider in [16] the corresponding optimal Skorokhod embedding. In this paper, we shall tackle the following general optimization problem associated to a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$:

$$\begin{aligned} &\text{Maximize/Minimize cost}[\pi] \\ &\quad = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\pi(x, y) \qquad \text{over } \pi \in \mathrm{MT}(\mu, \nu). \end{aligned} \tag{1.1}$$

Here, $\mathrm{MT}(\mu, \nu)$ is the set of *martingale transport plans*, that is, the set of probabilities π on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν, such that for μ-almost every $x \in \mathbb{R}^d$, the component π_x of its disintegration $(\pi_x)_x$ with respect to μ, that is, $d\pi(x, y) = d\pi_x(y)\, d\mu(x)$, has its barycenter at x; in other words, for any convex function φ on $\mathbb{R}^d$, one has $\varphi(x) \le \int_{\mathbb{R}^d} \varphi(y)\, d\pi_x(y)$.

One can also use the probabilistic notation, which amounts to

$$\text{Maximize/Minimize } \mathbb{E}_{\mathrm{P}} c(X, Y) \tag{1.2}$$

over all martingales (X, Y) on a probability space $(\Omega, \mathcal{F}, P)$ into $\mathbb{R}^d \times \mathbb{R}^d$ (i.e., $E[Y|X] = X$) with laws $X \sim \mu$ and $Y \sim \nu$ [i.e., $P(X \in A) = \mu(A)$ and $P(Y \in A) = \nu(A)$ for all Borel sets A in $\mathbb{R}^d$]. Note that in this case, the disintegration of π can be written as the conditional probability $\pi_x(A) = \mathbb{P}(Y \in A | X = x)$.

A classical theorem of Strassen [27] states that the set $MT(\mu, \nu)$ of martingale transports is nonempty if and only if the marginals μ and ν are in *convex order*, that is if:

1. μ and ν are probability measures with finite first moments, and
2. $\int_{\mathbb{R}^d} \varphi\, d\mu \le \int_{\mathbb{R}^d} \varphi\, d\nu$ for every convex function φ on $\mathbb{R}^d$.

In that case, we will write $\mu \le_C \nu$, which is sometimes called the *Choquet order for convex functions* [8]. Note that x is the barycenter of a measure ν if and only if $\delta_x \le_C \nu$, where δ_x is Dirac measure at x.

We will mostly consider the Euclidean distance cost $c(x, y) = |x - y|$ unless stated otherwise, although a good portion of our results below hold for more general costs, including those that are (locally) Lipschitz. We shall use the term optimization in problem (1.1) whenever the result holds for either maximization or minimization. We shall be more specific otherwise, since it will soon become very clear that the two cases can sometimes be fundamentally different. The following theorem summarizes the main structural result when μ and ν are one-dimensional marginals. Hobson–Neuberger [19] were first to deal with the maximization case while Beiglböck–Juillet [5] and D. Hobson and M. Klimmek [18] deal with the context of minimization.

THEOREM 1.1 (Beiglböck–Juillet [5], Hobson–Neuberger [19], Hobson–Klimmek [18]). *Assume that μ and ν are probability measures in convex order on $\mathbb{R}$, and that μ is continuous. There exists then a unique optimal martingale transport plan π for the cost function $c(x, y) = |x - y|$, such that*:

1. *If π is a minimizer, then its disintegration satisfies $|\operatorname{supp} \pi_x| \leq 3$ for every $x \in \mathbb{R}$. More precisely, π can be decomposed into $\pi_{\text{stay}} + \pi_{\text{go}}$, where $\pi_{\text{stay}} = (\mathrm{Id} \times \mathrm{Id})_\#(\mu \wedge \nu)$ (this measure is concentrated on the diagonal of $\mathbb{R}^2$) and π_{go} is concentrated on $\operatorname{graph}(T_1) \cup \operatorname{graph}(T_2)$ where T_1, T_2 are two real-valued functions.*
2. *If π is a maximizer, then its disintegration satisfies $|\operatorname{supp} \pi_x| \leq 2$ for every $x \in \mathbb{R}$, and π is concentrated on $\operatorname{graph}(T_1) \cup \operatorname{graph}(T_2)$ where T_1, T_2 are two real-valued functions.*

Our main goal in this paper is to consider higher-dimensional analogues of the above result. In [21], Lim showed that the above theorem extends, in the case of minimization, to the setting where the marginals are radially symmetric on $\mathbb{R}^d$ and $c(x, y) = |x - y|^p$ for $0 < p \leq 1$. The general case, without any symmetry assumption, is wide open.

Our main contribution in this paper is to develop methods and tools to tackle the general case for which we propose the following conjectures.

CONJECTURE 1. *Consider the cost function $c(x, y) = |x - y|$ and assume that μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$ ($\mu \ll \mathcal{L}^d$), if π is a martingale transport that optimizes (1.1). Then for μ-almost every x, $\operatorname{supp} \pi_x$ coincides with the set of extreme points of the convex hull of $\operatorname{supp} \pi_x$, that is, $\operatorname{supp} \pi_x = \operatorname{Ext}(\operatorname{conv}(\operatorname{supp} \pi_x))$.*

REMARK 1.2. If $\operatorname{supp} \pi_x$ is bounded for μ-almost all x (which is the case in particular when the target measure ν is compactly supported), then $\operatorname{conv}(\operatorname{supp} \pi_x) = \overline{\operatorname{conv}}(\operatorname{supp} \pi_x)$. In this case, the set of extreme points $\operatorname{Ext}(\overline{\operatorname{conv}}(\operatorname{supp} \pi_x))$ is also called the Choquet boundary of the compact convex set $\overline{\operatorname{conv}}(\operatorname{supp} \pi_x)$. Our conjecture can therefore be rephrased as: For μ a.e. x, $\operatorname{supp} \pi_x$ is equal to the Choquet boundary of its closed convex hull.

Note that for the minimization problem, we can and will assume that $\mu \wedge \nu = 0$ since any minimizing martingale transport for problem (1.1) must let the support of $\mu \wedge \nu$ stay put; see [5] or [21] for a proof. One can then easily see that in the one-dimensional case, the above conjecture reduces to Theorem 1.1 since then the dimension of the linear span of $\operatorname{supp} \pi_x$ is one and the Choquet boundary consists of exactly two points, unless of course $\operatorname{supp} \pi_x$ is a singleton.

A weaker form of the above conjecture indeed holds, namely, the Hausdorff dimension of $\operatorname{supp} \pi_x$ is at most $d-1$ for μ-a.e. x (Corollary 2.13). More importantly, we shall be able to prove Conjecture 1 in many important cases. First of all, it holds in dimension $d = 2$ (Theorem 2.14) provided the second marginal has compact support. It also holds true when a natural dual optimization problem is attained (Theorem 2.4), or when the linear span of $\operatorname{supp} \pi_x$ has full dimension (Corollary 2.13). The conjecture also holds partially (Theorem 2.5) when the marginals are in "subharmonic order," that is if

$$\int_{\mathbb{R}^d} \varphi \, d\mu \le \int_{\mathbb{R}^d} \varphi \, d\nu \qquad \text{for every subharmonic function } \varphi \text{ on } \mathbb{R}^d.$$

We actually expect to have a more rigid structure in the case of minimization. Indeed, Lim [21] showed that in this case, assuming $\mu \wedge \nu = 0$, we also have $|\operatorname{supp} \pi_x| \le 2$ for μ-almost all x, whenever the marginals are radially symmetric on $\mathbb{R}^d$ and $c(x, y) = |x - y|^p$ for $0 < p \le 1$. The general case remains open as we propose the following.

CONJECTURE 2 (Minimization). *Consider the cost function $c(x, y) = |x - y|$ and assume that μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, that $\mu \wedge \nu = 0$. If π is a martingale transport that minimizes* (1.1). *Then for μ almost every x, the set* $\operatorname{supp} \pi_x$ *consists of $k + 1$ points that form the vertices of a k-dimensional polytope, where $k := k(x)$ is the dimension of the linear span of* $\operatorname{supp} \pi_x$ *and, therefore, the minimizing solution is unique.*

We shall give a partial answer to the above conjecture, by showing it under the assumption that the target measure ν is discrete. Actually, in this case the result holds true in both the maximization and minimization cases (Theorem 2.15). We note however that—unlike the minimization case—one cannot always expect in higher dimensions neither the uniqueness of a maximizer (Example 2.17), nor a polytope-type structure for $\operatorname{supp} \pi_x$ (Example 2.16), even when the marginals are radially symmetric.

Just like in the Monge–Kantorovich theory, the above optimization problem (1.1) has a dual formulation, which will be crucial to our analysis. And similarly to that theory, the dual problem can be studied independently of the primal problem and without any underlying reference measures. Recall that for the quadratic cost studied by Brenier, the dual problem amounts to considering convex functions β, their Fenchel–Legendre duals $\alpha := \beta^*$ and the set $\Gamma = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d; \beta(y) +$

$\alpha(x) = \langle x, y\rangle\}$, which happens to be the graph of the subdifferential of β. Similar but more complicated phenomena arise in our situation. We shall work with the following notions.

For a subset Γ in $\mathbb{R}^d \times \mathbb{R}^d$, we shall denote by Γ_x, the fiber $\Gamma_x := \{y \in \mathbb{R}^d; (x, y) \in \Gamma\}$. For a Borel set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$, we write $X_\Gamma := \mathrm{proj}_X \Gamma$, $Y_\Gamma := \mathrm{proj}_Y \Gamma$, that is, X_Γ is the projection of Γ on the first coordinate space $\mathbb{R}^d$, and Y_Γ on the second.

DEFINITION 1.3. Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a cost function and let $X, Y \subseteq \mathbb{R}^d$ be Borel sets:

1. We say that a triplet of functions (α, γ, β) is *an admissible triple on* $X \times Y$, if $\alpha : X \to \mathbb{R}$, $\beta : Y \to \mathbb{R}$ and $\gamma : X \to \mathbb{R}^d$ satisfy the following inequality:

$$\beta(y) - \alpha(x) - \gamma(x)(y - x) \le c(x, y) \qquad \text{for all } (x, y) \in X \times Y. \tag{1.3}$$

We shall denote by $E_m(c, X, Y)$ the set of all such *admissible dual triples*. A similar definition holds when the inequality is reversed, and the set of those triplets will be denote by $E_M(c, X, Y)$. Note that $E_M(c, X, Y) = E_m(-c, X, Y)$.

2. For an admissible triple (α, γ, β), we will consider the set where equality holds, that is,

$$\begin{aligned} &\Gamma_{(\alpha,\gamma,\beta)} \\ &\quad := \{(x, y) \in X \times Y \mid \beta(y) - \alpha(x) - \gamma(x) \cdot (y - x) = c(x, y)\}. \end{aligned} \tag{1.4}$$

We shall sometimes allow γ to be a set-valued function. In this case, the above inequality/equality will mean that they actually hold for any vector b in $\gamma(x)$.

3. Any nonempty subset of $\Gamma_{(\alpha,\gamma,\beta)}$ will be called a *c-contact layer for* (α, γ, β) *in* $X \times Y$. When the ambient space is not specified, it means that it is simply $X_\Gamma \times Y_\Gamma$.

We shall sometimes say that a set Γ is *c-exposed by the admissible triple* (α, γ, β) if it is contained in $\Gamma_{(\alpha,\gamma,\beta)}$.

Denoting $E_m = E_m(c, \mathbb{R}^d, \mathbb{R}^d)$, one can then show (see, e.g., [3]) that if the cost c is lower semicontinuous, then for the minimization problem,

$$\min\left\{\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\pi; \pi \in \mathrm{MT}(\mu, \nu)\right\} \tag{1.5}$$

$$\begin{aligned} &= \sup\bigg\{\int_{\mathbb{R}^d} \beta\, d\nu - \int_{\mathbb{R}^d} \alpha\, d\mu; (\alpha, \gamma, \beta) \in E_m \\ &\qquad \text{for some } \gamma \in C_b(\mathbb{R}^d, \mathbb{R}^d)\bigg\}. \end{aligned} \tag{1.6}$$

Similarly, if the cost c is upper semicontinuous, then

$$\max\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\,d\pi;\pi\in \mathrm{MT}(\mu,\nu)\right\} \tag{1.7}$$

$$= \inf\left\{\int_{\mathbb{R}^d}\beta\,d\nu-\int_{\mathbb{R}^d}\alpha\,d\mu;(\alpha,\gamma,\beta)\in E_M \text{ for some } \gamma\in C_b(\mathbb{R}^d,\mathbb{R}^d)\right\}. \tag{1.8}$$

Note that if π is an optimal martingale measure and if the corresponding dual problem is attained on a triplet (α,γ,β), then it is easy to see that there exists a Borel subset $\Gamma\subseteq\mathbb{R}^d\times\mathbb{R}^d$ with full π-measure that is a c-contact layer for (α,γ,β), namely,

$$\beta(y)-\alpha(x)-\gamma(x)(y-x)=c(x,y) \quad \text{if and only if} \quad (x,y)\in\Gamma. \tag{1.9}$$

We shall show that such c-contact layers have a specific extremal structure (see Theorem 2.4). As a result, any martingale transport $\pi\in\mathrm{MT}(\mu,\nu)$ which is concentrated on a c-contact layer, when $c(x,y)=\pm|x-y|$, will satisfy Conjecture 1.

Recall that in the Monge–Kantorovich theory for mass transport, the dual problem is normally attained, and the "corresponding c-contact layer" is a set of the form $\Gamma=\{(x,y);\beta(y)-\alpha(x)=c(x,y)\}$, where β and α are related through c-Legendre duality, which let them inherit some of the regularity properties of c. We shall follow a similar methodology here by defining and exploiting in Section 3 a notion of *martingale c-Legendre duality* between the function β and the pair (α,γ). This will allow us to establish the regularity properties needed to analyze the structure of c-layer sets (see Theorem 2.3).

However, unlike the Monge–Kantorovich setting, attainment of the dual problem does not often hold for optimal martingale transports—at least in the maximization problem—even in the one-dimensional case, as shown in [3]; see Example 5.7 below. We therefore explore whether dual attainment can happen locally, which is sufficient to imply Conjecture 1. We prove in Section 6 that it is indeed the case under suitable assumptions on the marginals, such as when they are comparable for the order induced by subharmonic functions (see Theorem 2.5 or Theorem 6.1).

More importantly, we then proceed to consider the general case by establishing a remarkable decomposition for any Borel set Γ supporting a given optimal martingale transport π into disjoint components $\{\Gamma_C\}_{C\in I}$ in such a way that each piece is a c-contact layer for an admissible triplet $(\alpha_C,\gamma_C,\beta_C)$. What is remarkable is that this decomposition into c-contact layers can be established in full generality (i.e., for any cost function) and without any reference to a martingale transport problem or even to any reference measure. The property which enables Γ to be decomposed into c-contact layers is called *c-finitely exposability*; see Definition 2.9.

For example, Γ can be chosen to satisfy this property as a concentration set of optimal martingale transport π. The decomposition is done through an equivalence relation on the projection X_Γ of Γ on the first coordinate, that is induced by a well-chosen *irreducible convex paving*, that is, a collection of mutually disjoint convex subsets in $\mathbb{R}^d$ that covers X_Γ. See Theorem 2.11 for the precise statement.

We note that this result can be seen as a generalization of the decomposition of Beiglböck–Juillet [5] in the one-dimensional case $d = 1$, where the disintegration comes from restricting the measures μ, ν onto open subintervals of $\mathbb{R}$ obtained by examining the potential functions for μ, ν. Like theirs, our decomposition applies to any cost function c and not only to $c(x, y) = |x - y|$. It is however quite different since it depends on the support of the martingale measure π that we start with. More importantly, our decomposition needs not be countable (Example 9.3) which creates additional and interesting complications for the higher-dimensional cases.

We shall use the above decomposition to establish the previously stated conjectures under various conditions, including cases when duality attainment does not hold. For example, it leads to the dimensional result on the support of π_x (Corollary 2.13), and Conjecture 1 in dimension $d = 2$ (Theorem 2.14), as well as in the case where all components $(C)_{C\in I}$ are d-dimensional (see Corollary 2.13).

Remarkably, the results discussed so far do not distinguish between the minimization and maximization problems (except that we assume that $\mu \wedge \nu = 0$ in the case of minimization). The previously mentioned decomposition can be used to prove Conjecture 2 in either the minimization and maximization case, provided the target measure ν has a countable support (Theorem 2.15). However, as mentioned above, we believe that these two problems are quite different, at least in terms of finding finer structural results for each of the cases.

Back to the martingale problem, we then consider the disintegration $\{\pi_C\}_{C\in I}$ of any martingale measure π along the above described decomposition of its support Γ (Theorem 9.1). This suggests a canonical decomposition of the optimal martingale transport problem into a collection of noninteractive martingale problems where duality is attained for each piece π_C in $\mathrm{MT}(\mu_C, \nu_C)$. But we note that, in Theorem 9.1 we assume measurability of the decomposition map Ξ; this technical measurability issue, and thereby the rigorous disintegration of martingale measures, should be an important problem in high dimensions. We refer to [9, 24] for some relevant progress in this direction.

In the next section, we give the precise statements of our results. In Section 3, we introduce and study the notion of *martingale c-transforms*, which will be used to improve the regularity properties of admissible triples. This will be used in Section 4 to analyze the structure of c-contact layers that are exposed by such triples. We apply these results in Section 5 to the case where the dual problem is attained, proving that Conjecture 1 holds in this situation. In Section 6, we give a setting where the dual problem is attained locally, showing Conjecture 1 for a case where the marginals are in subharmonic order. In Section 7, we establish the existence of the irreducible convex paving decomposition, as well as the existence of admissible triplets exposing each of the components. This decomposition is used in Section 8 to prove under various additional conditions but without assuming dual attainment, structural results for sets where optimal martingale transports concentrate. Finally, Section 9 deals with the disintegration of martingales along this decomposition and how it is related to the presence of Nikodym sets.

The authors are extremely grateful to Luigi Ambrosio for several pertinent remarks and discussions regarding the results of this paper. The first-named author also thanks David Preiss for very helpful discussions and insight. We are also thankful to an anonymous referee who gave many valuable comments that improved the article very much.

2. Main results. To discuss our main results, we first introduce a few definitions. We also borrow some of the notation from [5].

DEFINITION 2.1. For $A \subseteq \mathbb{R}^d$, we shall write $V(A)$ for the lowest-dimensional affine space containing A. Also define

$$\mathrm{IC}(A) := \mathrm{int}\big(\mathrm{conv}(A)\big) \quad \text{and} \quad \mathrm{CC}(A) := \mathrm{cl}\big(\mathrm{conv}(A)\big),$$

where again the interior or closure is taken in the topology of $V(A)$, where the topology of a set A is with respect to the Euclidean metric topology of $V(A)$ (and not with respect to the whole space $\mathbb{R}^d$).

If $A = \{x\}$, then $\mathrm{IC}(A) = \{x\}$ since we consider the interior of a singleton set is itself in the topology of 0-dimensional space.

In reality, we will be dealing with the vertical fibers $\Gamma_x = \{y \in \mathbb{R}^d \mid (x, y) \in \Gamma\}$ of a certain class of Borel sets $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$, on which martingale measures $\pi \in \mathrm{MT}(\mu, \nu)$ would be concentrated. The constraint that x is the barycenter of π_x, which is normally supported on $\overline{\Gamma}_x$, naturally leads us to the following definition. Recall that for a Borel set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$, we write $X_\Gamma := \mathrm{proj}_X \Gamma$, $Y_\Gamma := \mathrm{proj}_Y \Gamma$, that is, X_Γ is the projection of Γ on the first coordinate space $\mathbb{R}^d$, and Y_Γ on the second.

DEFINITION 2.2. We say that a Borel set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$ is a *martingale supporting set*, if

$$\text{for every } x \in X_\Gamma, \qquad x \in \mathrm{IC}(\Gamma_x). \tag{2.1}$$

We let $\mathcal{S}_{\mathrm{MT}}$ denote the class of all martingale supporting sets.

Our first main result shows that martingale supporting sets that are c-contact layers enjoy special structural properties. A key step established in Section 3 is to show that an exposing admissible triple can be extended and regularized via a notion of *martingale c-Legendre transform*, so that it verifies the needed differentiability properties.

THEOREM 2.3 (Regularization of admissible triples via martingale-Legendre transform). *Let c be a cost function on $\mathbb{R}^d$ such that $x \mapsto c(x, y)$, respectively, $y \mapsto c(x, y)$, is locally Lipschitz, where the Lipschitz constants are uniformly bounded in y and respectively, in x. Let Γ be a Borel set in $\mathcal{S}_{\mathrm{MT}}$ that is a c-contact layer, and suppose that $X_\Gamma \subseteq \Omega := \mathrm{IC}(Y_\Gamma)$ with Ω being an open set in $\mathbb{R}^d$. Then:*

1. *There exist a locally Lipschitz function $\alpha : \Omega \to \mathbb{R}$, a locally bounded $\gamma : \Omega \to \mathbb{R}^d$, and $\beta : \mathbb{R}^d \to \mathbb{R}$, such that Γ is a c-contact layer for the triplet (α, γ, β).*
2. *If the admissible triple is in $E_M(c, X_\Gamma, Y_\Gamma)$, and if $y \mapsto c(x, y)$ is assumed to be convex, then β can be taken to be a convex function on $\mathbb{R}^d$.*
3. *If $c(x, y) = |x - y|$ and the admissible triple is in $E_m(c, X_\Gamma, Y_\Gamma)$, then $\alpha = \beta$ on Ω.*

This will allow us to prove the following structural result.

THEOREM 2.4 (Extremal structure of a martingale supporting c-contact layer). *Let $c(x, y) = \pm|x - y|$ and assume Γ is a c-contact layer in $\mathcal{S}_{\mathrm{MT}}$. Then for $\mathcal{L}^d$-a.e. x in X_Γ, the closure $\overline{\Gamma_x}$ of Γ_x coincides with the set of extreme points of the convex hull of $\overline{\Gamma_x}$, that is, $\overline{\Gamma_x} = \mathrm{Ext}(\mathrm{conv}(\overline{\Gamma_x}))$.*

In particular, if μ is a probability measure that is absolutely continuous with respect to the Lebesgue measure, and if the dual problem is attained, then for any $\pi \in \mathrm{MT}(\mu, \nu)$ that is a solution of (1.1) for either the minimization or maximization problem, then for μ-a.e. x, $\operatorname{supp} \pi_x = \mathrm{Ext}(\mathrm{conv}(\operatorname{supp} \pi_x))$.

We shall see that the dual problem is not always attained. However, a localized version of the above theorem will allow us to deal with a case where the marginals are in subharmonic order. Actually, by letting P_μ be the Newtonian potential of a probability measure μ, we shall be able to deduce the following result (see Section 6).

THEOREM 2.5 (Case of marginals in subharmonic order). *Assume $\mu \leq_{SH} \nu$ where μ, ν are probability measures with compact support on $\mathbb{R}^d$ such that $\mu \ll \mathcal{L}^d$ $(d \geq 3)$, and that the open set $\{x | P_\nu(x) - P_\mu(x) > 0\}$ has the full measure of μ. If $\pi \in \mathrm{MT}(\mu, \nu)$ is an optimal solution for the problem (1.1), where the cost function is $c(x, y) = \pm|x - y|$, then for μ-a.e. x, $\operatorname{supp} \pi_x = \mathrm{Ext}(\mathrm{conv}(\operatorname{supp} \pi_x))$.*

Since martingale supporting sets Γ in $\mathcal{S}_{\mathrm{MT}}$ are not always c-contact layers even when they are concentration sets for optimal martingale transports ([3] or Example 5.7 below), we investigate the possibility of decomposing such sets into "irreducible components" such that each component becomes a c-contact layer. For that, we introduce the concept of a *convex paving*.

DEFINITION 2.6. Let Φ be a family of mutually disjoint open convex sets in $\mathbb{R}^d$. [Recall that here the openness of a set C is with respect to the space $V(C)$.] Given a set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$, we shall say that Φ is a convex paving for Γ provided:

1. $X_\Gamma \subseteq \bigcup_{C\in\Phi} C$.
2. Each $C \in \Phi$ contains at least one element x in X_Γ [C is then denoted $C(x)$].
3. For any $z, x \in X_\Gamma$, we have $\mathrm{IC}(\Gamma_z) \cap C(x) \neq \varnothing \Rightarrow \mathrm{IC}(\Gamma_z) \subseteq C(x)$.

Note that such a paving clearly defines an equivalent relation on X_Γ by simply defining $x \sim_\Phi x'$ if and only if $C(x) = C(x')$. The corresponding equivalent classes are then $[x] = C(x) \cap X_\Gamma$.

There can be many convex pavings of a set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$; take, for example, $\Phi := \{\mathbb{R}^d\}$ which however does not give much information about Γ. We therefore introduce the following concept.

DEFINITION 2.7. For a fixed set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$, we shall say that Φ is an irreducible convex paving for Γ if for any other convex paving Ψ for Γ, we have the following property: If $C \in \Phi$, $D \in \Psi$ are such that $C \cap D \neq \varnothing$, then necessarily $C \subseteq D$.

Note that an irreducible convex paving for a set Γ is necessarily unique. As to their existence, we shall show in Section 7 the following result.

THEOREM 2.8 (Convex paving). *For every martingale supporting set Γ in $\mathcal{S}_{\mathrm{MT}}$, there exists a unique irreducible convex paving for Γ.*

Now, a key property of optimal transport plans in Monge–Kantorovich theory is that they are concentrated on Borel sets that are *c-cyclically monotone*, which is a property that describes every finite collection of points in the concentration set [28]. Similarly, a key property of an optimal martingale transport $\pi \in \mathrm{MT}(\mu, \nu)$—due to Beiglböck and Juillet [5]—is a *monotonicity property* enjoyed by every finite collection of points in their support. It implies in particular, that there exists a set Λ of full π-measure in $\mathbb{R}^d \times \mathbb{R}^d$ such that each one of its finite subsets is a c-contact layer. This is one of the consequences of the variational lemma in [5], where duality on finite sets is obtained via linear programming (see [5] and [18]). We therefore introduce the following combinatorial counterpart of cyclic monotonicity for martingale transport.

DEFINITION 2.9. A subset Λ of $\mathbb{R}^d \times \mathbb{R}^d$ is said to be *c-finitely exposable* for some cost function c, if each one of its finite subsets is a c-contact layer. That is, given any finite subset H of Λ, there exists $\alpha : X_H \to \mathbb{R}$, $\gamma : X_H \to \mathbb{R}^d$, $\beta : Y_H \to \mathbb{R}$ such that

$$\beta(y) - \alpha(x) - \gamma(x)(y - x) \le c(x, y) \qquad \text{for all } (x, y) \in X_H \times Y_H, \quad \text{and}$$

$$\beta(y) - \alpha(x) - \gamma(x)(y - x) = c(x, y) \qquad \text{for all } (x, y) \in H$$

for the minimization problem. For the maximization problem, the first inequality should be reversed.

The following proposition describes the combinatorial nature of the support of optimal martingale transports.

PROPOSITION 2.10. *Let $\pi \in \mathrm{MT}(\mu, \nu)$ be an optimal martingale transport for Problem* (1.1). *Assuming the cost c continuous*, *there exists a c-finitely exposable concentration set Λ for π.*

Indeed, it is shown in [5] (see also [30]) that there exists a Borel set Λ in $\mathbb{R}^d \times \mathbb{R}^d$ with $\pi(\Lambda) = 1$, that satisfies a certain monotonicity property, which is the martingale counterpart of the c-cyclic monotonicity that is inherent to the Monge–Kantorovich theory. As mentioned above, by the duality theorem of linear programming, this property is equivalent to saying that every finite subset of Λ is a c-contact layer.

Since duality is not attained in general, an optimal martingale transport measure is not necessarily concentrated on a c-contact layer $\Gamma \in \mathcal{S}_{\mathrm{MT}}$. On the other hand, we can and will assume that it is concentrated on a set $\Gamma \in \mathcal{S}_{\mathrm{MT}}$ whose finite subsets are c-contact layers. This leads to the question of finding "maximal" components of Γ that are c-contact layers. It turns out that this is indeed the case as we show that $\Gamma_C := \Gamma \cap (C \times \mathbb{R}^d)$ is a c-contact layer for any component C of the irreducible convex paving Φ of Γ. It is summarized in the following theorem.

THEOREM 2.11 (Convex paving with c-contact layer components). *Let Γ be a c-finitely exposable set in $\mathcal{S}_{\mathrm{MT}}$, then there exists an irreducible convex paving Φ for Γ such that for every convex component C in Φ, the set $\Gamma \cap (C \times \mathbb{R}^d)$ is a c-contact layer.*

REMARK 2.12. Theorem 2.11 can be seen as a martingale counterpart to a celebrated result of Rockafellar [25] in the Monge–Kantorovich theory for mass transport, which essentially says that the property of c-cyclical monotonicity that characterizes the support of optimal transport plans are somewhat "c-contact layers" exposed by a pair of functions, one being c-convex and the other being its c-Legendre transform. Here, *c-finite exposability* replaces *c-cyclic monotonicity*, while "exposing" martingale supporting sets require a new notion of duality between a function β and a pair of functions (α, γ). However, in the martingale case, the whole support is not necessarily a c-contact layer, but every irreducible component is.

Theorems 2.3 and 2.11 yield several structural results in general dimensions such as the following. Note that the attainability of the dual problem is not assumed here.

COROLLARY 2.13 (Dimensional result). *Let π be a solution of the optimization problem* (1.1) *with $c(x, y) = |x - y|$ and suppose μ is absolutely continuous with respect to the Lebesgue measure. Then for μ-almost every x in $\mathbb{R}^d$:*

(1) *The Hausdorff dimension of* $\operatorname{supp}\pi_x$ *is at most* $d-1$, *and*
(2) *If* $\dim V(\operatorname{supp}\pi_x)=d$, *then* $\operatorname{supp}\pi_x=\operatorname{Ext}(\operatorname{conv}(\operatorname{supp}\pi_x))$.

PROOF. Indeed, there exists $\Gamma\in\mathcal{S}_{\mathrm{MT}}$ with $\pi(\Gamma)=1$ that is c-finitely exposable, and such that $\overline{\Gamma_x}=\operatorname{supp}\pi_x$ for μ a.e. x (see Appendix A). Now, consider those points x with $\dim V(\Gamma_x)=d$. In this case, the disjoint sets $C(x)$ in Theorem 2.11 are open sets in $\mathbb{R}^d$ and so, the restriction of μ to each of the components is again absolutely continuous. Theorems 2.3 and 2.4 can then be applied. Note now that the the set of extreme points has dimension at most $d-1$. This shows that for μ-a.e. x in the open set $\bigcup_{\dim V(C)=d}C$, we have that $\dim(\overline{\Gamma_x})\le d-1$. The property also obviously holds outside that set, which means that item (1) is also verified. □

A more involved application of the decomposition is a complete solution of Conjecture 1 in two dimensions, namely the following, which is proved in Section 8.

THEOREM 2.14 (The two-dimensional case). *Assume* $d=2$, $c(x,y)=|x-y|$, μ *is absolutely continuous with respect to the Lebesgue measure*, *and* ν *has compact support*. *Let* $\pi\in\mathrm{MT}(\mu,\nu)$ *be a solution of* (1.1), *then for* μ*-a.e.* x, $\operatorname{supp}\pi_x=\operatorname{Ext}(\overline{\operatorname{conv}}(\operatorname{supp}\pi_x))$.

The decomposition also allows us to give in Section 8 the following positive answer to Conjecture 2, whenever the target measure is discrete. Note that in this case, the result holds true in both the maximization and minimization problems.

THEOREM 2.15 (The case of a discrete target). *Let* $c(x,y)=|x-y|$ [*or more generally for* $c(x,y)=|x-y|^p$ *with* $p\neq 2$], *suppose* μ *is absolutely continuous with respect to the Lebesgue measure*, *and that* ν *is discrete*; *that is*, ν *is supported on a countable set*. *If* $\pi\in\mathrm{MT}(\mu,\nu)$ *is an optimizer for* (1.1), *then for* μ*-a.e.* x, $\operatorname{supp}\pi_x$ *consists of exactly* $d+1$ *points which are vertices of a* d*-dimensional polytope in* $\mathbb{R}^d$ *and*, *therefore*, *the optimal solution is unique*.

Now we give a couple of examples, which illustrate that the above stated conjectures could be the best structural results we can hope for.

EXAMPLE 2.16. The polytope-like structure of the support required in Conjecture 2 does not hold in general for the corresponding maximization problem. Indeed, since $\frac{1}{2}(|x-y|-1)^2\ge 0$, we have

$$\frac{1}{2}|y|^2-\frac{1}{2}|x|^2+\frac{1}{2}-x\cdot(y-x)\ge|x-y| \qquad \text{on } \mathbb{R}^d\times\mathbb{R}^d, \tag{2.2}$$

with equality on the set $\{(x, y); |x - y| = 1\}$. The functions $\alpha(x) = \frac{1}{2}|x|^2 - \frac{1}{2}$, $\beta(y) = \frac{1}{2}|y|^2$ and $\gamma(x) = x$ then form a dual triplet for the maximization problem with cost $|x - y|$. This means that every martingale (X, Y) with $|X - Y| = 1$ a.s. is optimal for the maximization problem corresponding to its own marginals $X \sim \mu$ and $Y \sim \nu$. Hence, $\operatorname{supp} \pi_x$ is not in general a discrete set, and indeed, $\operatorname{supp} \pi_x$ can attain the Hausdorff dimension $d - 1$.

We now consider the uniqueness question in Conjecture 2, and whether it could hold for the maximization problem. In, it is shown that when $d = 1$ the solution of the martingale transport problem (1.1) is unique for both max/min problem under the assumption that μ is absolutely continuous. Also, it is reported in [21] that in the minimization problem with radially symmetric marginals (μ, ν), the minimizer is again unique in any dimension. We note however that, unlike the minimization case, one cannot expect the uniqueness of a maximizing martingale measure in higher dimensions, even in the radially symmetric case, as the following example indicates.

EXAMPLE 2.17. Let μ be a radially symmetric probability measure on $\mathbb{R}^2 \simeq \mathbb{C}$ such that $\mu(\{0\}) = 0$. Let $z_1 = \cos \frac{\pi}{4} + i \sin \frac{\pi}{4}$, $z_2 = \cos \frac{\pi}{4} - i \sin \frac{\pi}{4}$, $z_3 = -z_1$ and $z_4 = -z_2$, and define the probability measures π_1 and π_2 on $\mathbb{C} \times \mathbb{C}$, whose disintegrations π_x^1 and π_x^2 for each $x \in \mathbb{C}$, $x \neq 0$, are given by

$$\pi_x^1 = \frac{1}{4}\delta_{x+\frac{x}{|x|}z_1} + \frac{1}{4}\delta_{x+\frac{x}{|x|}z_2} + \frac{1}{4}\delta_{x+\frac{x}{|x|}z_3} + \frac{1}{4}\delta_{x+\frac{x}{|x|}z_4},$$
$$\pi_x^2 = \frac{1}{8}\delta_{x+\frac{x}{|x|}z_1} + \frac{3}{8}\delta_{x+\frac{x}{|x|}z_2} + \frac{1}{8}\delta_{x+\frac{x}{|x|}z_3} + \frac{3}{8}\delta_{x+\frac{x}{|x|}z_4}.$$

Then, by the discussion in Example 2.16, one can see that both π_1 and π_2 are optimal for the maximization problem corresponding to the cost $|x - y|$ and marginals μ and $\nu := \nu_1 = \nu_2$, where $d\nu_i(y) = \int_{\mathbb{C}} \pi_x^i(y)\, d\mu(x)$, $i = 1, 2$; hence, the maximizer is not unique.

Finally, we consider in Section 9 whether one can perform a disintegration of π with respect to the decomposition $\{\Gamma_C\}_{C\in\Phi}$ into components $(\pi_C)_C$ in such a way that each π_C is a probability measure supported on $\Gamma_C := \Gamma \cap (C \times \mathbb{R}^d)$ and $\pi_C \in \mathrm{MT}(\mu_C, \nu_C)$, where μ_C, ν_C are suitable probability measures in convex order, with μ_C is supported on $X_C := X_\Gamma \cap C$ and ν_C on Y_{Γ_C}. The advantage of this decomposition is that If π is optimal for problem (1.1) in $\mathrm{MT}(\mu, \nu)$, then π_C is optimal for the same problem on $MT(\mu_C, \nu_C)$, with the added property that Γ_C is a c-contact layer, which means that duality is attained for each π_C. The decomposition of Γ given by Theorem 9.1 was motivated by a similar one proposed by Beiglböck–Juillet [5] in the one-dimensional case ($d = 1$). Our decomposition is however quite different since it depends on the concentration set Γ for π, while in their case the decomposition depends only on the marginals μ and ν. Theirs is also a countable partition, which makes the restricted problems much more amenable to analysis. Actually, the intervals in their decomposition are simply the connected components of the set where the potentials of μ and ν are different on the real line. However, in the higher-dimensional cases our decomposition can be uncountable, and that is why we talk about a disintegration as opposed to a decomposition. Moreover, the induced probability measures μ_C's can be Dirac measures (see Example 9.3), which means that Theorem 2.4 may not be applicable to each piece π_C even if duality is attained for the restricted problem. We refer to Section 9 for the challenges and the interesting questions arising from this fundamental decomposition.

3. The martingale *c*-Legendre transform. In this section, we investigate properties of the admissible triplet of functions that appear in the dual martingale problem and their associated contact layers. Note that in the case of standard mass transport problems, the contact layer is determined by a potential function and its c-Legendre transform, whose regularity properties are inherited from those of c, and which can be studied independently of the primal transport problem. A similar methodology works in our setting, once we introduce an appropriate Legendre duality.

DEFINITION 3.1. Let Y be a Borel set in $\mathbb{R}^d$ such that $\Omega := \mathrm{IC}(Y)$ is open in $\mathbb{R}^d$, and let $\beta : Y \to \mathbb{R}$ be a Borel function such that for some $s \in \mathbb{R}$, $t \in \mathbb{R}^d$, $x_0 \in \Omega$, we have

$$\beta(y) \le c(x_0, y) + t \cdot (y - x_0) + s \qquad \text{for all } y \in Y. \tag{3.1}$$

1. The *martingale c-Legendre dual of the function β on Ω* is the pair $\beta_c := (\alpha_c, \gamma_c)$, where $\alpha_c : \Omega \to \mathbb{R}$ is given by

$$\begin{aligned} \alpha_c(x) := \inf\{&a \in \mathbb{R} : \exists b \in \mathbb{R}^d \\ &\text{such that } \beta(y) - c(x, y) \le b \cdot (y - x) + a, \forall y \in Y\}, \end{aligned} \tag{3.2}$$

and $\gamma_c : \Omega \to \mathbb{R}^d$ is the possibly set-valued function defined by

$$\gamma_c(x) := \{b \in \mathbb{R}^d : \beta(y) - c(x, y) \le b \cdot (y - x) + \alpha_c(x), \forall y \in Y\}. \tag{3.3}$$

2. The *martingale c-Legendre dual of a pair of functions* $(\alpha, \gamma) : \Omega \to \mathbb{R} \times \mathbb{R}^d$ is the function $(\alpha, \gamma)_c : \mathbb{R}^d \to \mathbb{R}$ defined by

$$(\alpha, \gamma)_c(y) := \inf_{x \in \Omega, b \in \gamma(x)} \{c(x, y) + b \cdot (y - x) + \alpha(x)\}. \tag{3.4}$$

3. We shall denote by β_{cc} the martingale c-Legendre dual of the pair $\beta_c = (\alpha_c, \gamma_c)$, and say that β is *martingale c-convex on Y*, if $\beta = \beta_{cc}$ on Y.

In order to emphasize the analogy with the standard Fenchel–Legendre duality, we shall write

$$\beta_c(x, y) = (\alpha_c, \gamma_c)(x, y) := \alpha_c(x) + \gamma_c(x)(y - x).$$

THEOREM 3.2. *Assume that* $(x, y) \mapsto c(x, y)$ *is continuous and* $x \mapsto c(x, y)$ [*resp*., $y \mapsto c(x, y)$] *is locally Lipschitz with local Lipschitz constants uniformly bounded in* y (*resp*., *in* x). *Let* Y *be a Borel set in* $\mathbb{R}^d$ *such that* $\Omega := \mathrm{IC}(Y)$ *is open in* $\mathbb{R}^d$, *and let* $\beta : Y \to \mathbb{R}$ *be a Borel function satisfying* (3.1), *and* $\beta_c = (\alpha_c, \gamma_c)$ *its martingale* c-*Legendre dual*. *Then*:

1. α_c *is locally Lipschitz in* Ω, *while* γ_c *and* β_{cc} *are locally bounded in* Ω.
2. $\beta \le \beta_{cc}$ *on* Y, *and*

$$\beta_{cc}(y) - \beta_c(x, y) \le c(x, y) \qquad \text{for all } (x, y) \in \Omega \times \mathbb{R}^d. \tag{3.5}$$

In other words, *the triple* $(\beta_c, \beta_{cc}) = (\alpha_c, \gamma_c, \beta_{cc}) \in E_m(c, \Omega, \mathbb{R}^d)$.

3. $\beta_{cc}(x) - \delta_1 \le \alpha_c(x) \le \beta_{cc}(x) + \delta_2$ *for all* $x \in \Omega$, *where*

$$\delta_1 = \sup_{x\in\Omega} c(x, x) \quad \text{and} \quad \delta_2 = \sup_{x,x'\in\Omega, y\in Y} \big[c(x, y) - c(x, x') - c(x', y)\big].$$

4. *Let* $X \subseteq \Omega$ *and let* (α, γ) *be defined on* X *such that* $(\alpha, \gamma, \beta) \in E_m(c, X, Y)$, *then* $\alpha(x) \ge \alpha_c(x)$ *on* X. *Moreover*, *if a* c-*contact layer* $\Gamma \subseteq \Gamma_{(\alpha,\gamma,\beta)}$ *belongs to* $\mathcal{S}_{\mathrm{MT}}$, *then*

$$\alpha(x) = \alpha_c(x) \quad \text{and} \quad \gamma(x) \subseteq \gamma_c(x) \qquad \text{on } X_\Gamma,$$
$$\beta_{cc} = \beta \qquad \text{on } Y_\Gamma, \quad \text{and} \quad \Gamma \subseteq \Gamma_{(\alpha_c,\gamma_c,\beta_{cc})}.$$

5. *The function* β_{cc} *is martingale* c-*convex on* $\mathbb{R}^d$, *that is*, $\beta_{cc} = \beta_{cccc}$ *on* $\mathbb{R}^d$.

PROOF. (1) and (2). We first show that α_c is locally bounded in Ω. For $x \in \Omega = \mathrm{IC}(Y)$, we may choose $\{y_1, \dots, y_s\} \subseteq Y$ such that

$$x \in U := \mathrm{IC}\big(\{y_1, \dots, y_s\}\big) \quad \text{and} \quad U \quad \text{is open in } \mathbb{R}^d. \tag{3.6}$$

Since $x = \sum_i t_i y_i$, $\sum_i t_i = 1$, $t_i \ge 0$, it is clear that $\alpha_c(z) \ge M(z) := \min_{y_i}[\beta(y_i) - c(z, y_i)]$ for all $z \in U$. In view of the continuity of c, this yields that α_c is locally lower bounded.

We now prove that α_c is locally upper bounded. Indeed, fix $R > 0$ and let $x \in \Omega$, $y \in Y$ be such that $|x_0|, |x| < R$. By the local Lipschitz property of c in x, that is,

$$\big|c(x, y) - c(x_0, y)\big| \le C|x - x_0|$$

for some $C = C(R) > 0$ and for all $|x| < R$, we have that

$$s + t \cdot (y - x_0) \ge \beta(y) - c(x_0, y) \ge \beta(y) - c(x, y) - C|x - x_0|. \tag{3.7}$$

Thus,

$$s + C|x - x_0| + t \cdot (x - x_0) + t \cdot (y - x) \geq \beta(y) - c(x, y).$$

The definition of α_c gives

$$s + C|x - x_0| + t \cdot (x - x_0) \geq \alpha_c(x). \tag{3.8}$$

In particular, α_c is locally upper bounded, hence locally bounded.

Note now that $\gamma_c(x)$ is a set valued function, and it is clearly closed and convex for each $x \in \Omega$. To see the local boundedness of γ_c, use (3.6) and let V be a small neighborhood of x whose closure is in U. Since α_c is bounded on V, there exists a constant C such that

$$b \cdot (y_i - z) \geq C, \qquad \forall z \in V, i = 1, 2, \dots, s, \forall b \in \bar{\gamma}(z) \tag{3.9}$$

which says that γ_c is bounded on V, thus locally bounded on Ω. To show that $\gamma_c(x)$ is nonempty for any $x \in \Omega$, choose an approximating sequence $\{a_n\} \subseteq \mathbb{R}$ for $\alpha_c(x)$ and corresponding $\{b_n\} \subseteq \mathbb{R}^d$, in such a way that $\beta(y) - c(x, y) \leq b_n \cdot (y - x) + a_n$ and $a_n \searrow \alpha_c(x)$. Now the above argument shows that $\{b_n\}$ must be bounded, hence its accumulation points must be in $\gamma_c(x)$.

We now show that α_c is locally Lipschitz. Since α_c is finite in Ω, the above argument showing the local boundedness for α_c can be repeated, giving (3.8) for any $x, x_0 \in \Omega$, $s = \alpha_c(x_0)$ and $t \in \gamma_c(x_0)$;

$$\alpha_c(x_0) + C|x - x_0| + \gamma_c(x_0) \cdot (x - x_0) \geq \alpha_c(x).$$

By interchanging x and x_0, we get

$$\left|\alpha_c(x) - \alpha_c(x_0)\right| \leq \left(\left(\left|\gamma_c(x)\right| \vee \left|\gamma_c(x_0)\right|\right) + C\right)|x - x_0|.$$

Therefore, the local boundedness of γ_c implies that α_c is locally Lipschitz in Ω. If furthermore $x \mapsto c(x, y)$ is Lipschitz (with Lipschitz constant uniformly in y) and γ_c is bounded, then the above estimate shows that α_c is Lipschitz in Ω.

As for β_{cc}, it is clear that it is measurable and locally upper bounded. It is also clear that

$$(\alpha_c, \gamma_c, \beta_{cc}) \in E_m\left(c, \Omega, \mathbb{R}^d\right) \quad \text{and} \quad \beta \leq \beta_{cc} \qquad \text{on } Y. \tag{3.10}$$

We now show that β_{cc} is locally bounded in Ω, by following a similar argument as for α_c. First, let $x \in \Omega$, $y \in Y$, $y' \in \Omega$. By the local Lipschitz property of c in y, that is,

$$\left|c(x, y) - c\left(x, y'\right)\right| \leq C\left|y - y'\right|,$$

for some $C = C(R) > 0$, and for all $|y|, |y'| < R$, we see

$$\begin{aligned} \beta(y) &\leq c(x, y) + \gamma_c(x) \cdot (y - x) + \alpha_c(x) \\ &\leq c\left(x, y'\right) + \gamma_c(x) \cdot \left(y' - x\right) + \alpha_c(x) \\ &\quad + \gamma_c(x) \cdot \left(y - y'\right) + C\left|y - y'\right|. \end{aligned} \tag{3.11}$$

Now, since $y' \in \Omega = \mathrm{IC}(Y)$, one can choose $\{y_1, \dots, y_s\} \subseteq Y$ such that

$$y' \in W = \mathrm{IC}(\{y_1, \dots, y_s\}) \quad \text{and } W \text{ is open in } \mathbb{R}^d.$$

Thus (3.11) implies, after putting y_i's in place of y and summing up with appropriate weights, that

$$\min_{y_i} \beta(y_i) \le \beta_{cc}(y') + C \max_{y_i} |y_i - y'|,$$

hence yielding the local lower boundedness, thus the local boundedness of β_{cc} in Ω. This completes the proof of the items (1) and (2).

In order to establish (3), we first note that the inequality $\beta_{cc} - \delta_1 \le \alpha_c$ on Ω follows from the fact that $(\alpha_c, \gamma_c, \beta_{cc}) \in E_m(c, \Omega, \mathbb{R}^d)$. For the other inequality, notice that for each $y \in \Omega$ and an arbitrary $\varepsilon > 0$, there is $x_\varepsilon \in \Omega$ and $b \in \gamma_c(x_\varepsilon)$ [which we will simply write as $\gamma_c(x_\varepsilon)$ in the sequel], such that

$$\alpha_c(x_\varepsilon) + \gamma_c(x_\varepsilon) \cdot (y - x_\varepsilon) + c(x_\varepsilon, y) - \varepsilon \le \beta_{cc}(y). \tag{3.12}$$

Let $a_\varepsilon(y) := \alpha_c(x_\varepsilon) + \gamma_c(x_\varepsilon) \cdot (y - x_\varepsilon) + c(x_\varepsilon, y)$, and consider for $z \in Y$, the function

$$L(z) = a_\varepsilon(y) + \gamma_c(x_\varepsilon)(z - y) + c(y, z).$$

Then

$$\begin{aligned}\beta_{cc}(z) - L(z) &\le \alpha_c(x_\varepsilon) + \gamma_c(x_\varepsilon) \cdot (z - x_\varepsilon) + c(x_\varepsilon, z) \\ &\qquad - \big(\alpha_c(x_\varepsilon) + \gamma_c(x_\varepsilon) \cdot (y - x_\varepsilon) + c(x_\varepsilon, y)\big) \\ &\qquad - \gamma_c(x_\varepsilon)(z - y) - c(y, z) \\ &= c(x_\varepsilon, z) - c(x_\varepsilon, y) - c(y, z) \le \delta_2.\end{aligned}$$

Hence $\beta_{cc}(z) \le L(z) + \delta_2$ and, therefore, $\beta(z) \le L(z) + \delta_2$ for $z \in Y$ by item (1). From the definition of α_c, this implies $\alpha_c(y) \le a_\varepsilon(y) + \delta_2$, and from (3.12), we have $\alpha_c(y) \le \beta_{cc}(y) + \varepsilon + \delta_2$. Since ε is arbitrary, the proof of (3) is complete.

To prove (4), first note that if $X \subseteq \Omega$ and $(\alpha, \gamma, \beta) \in E_m(c, X, Y)$, then the definition of α_c obviously implies that $\alpha \ge \alpha_c$ on X. Now assume that $\Gamma \subseteq \Gamma_{(\alpha,\gamma,\beta)}$, $\Gamma \in \mathcal{S}_{\mathrm{MT}}$ and in particular, for each $x \in X_\Gamma$, $x \in \mathrm{IC}(\Gamma_x)$. Let $x = \sum_i t_i y_i$, $\sum_i t_i = 1$, $t_i \ge 0$, $y_i \in \Gamma_x$, and observe that

$$\beta(y_i) - c(x, y_i) = \gamma(x) \cdot (y_i - x) + \alpha(x), \tag{3.13}$$

$$\beta(y_i) - c(x, y_i) \le \gamma_c(x) \cdot (y_i - x) + \alpha_c(x), \tag{3.14}$$

where the first identity is due to the definition of $\Gamma_{(\alpha,\gamma,\beta)}$ and the second inequality is due to the definition of $\beta_c = (\alpha_c, \gamma_c)$. Summing up the above relations with the weights t_i, we get

$$\alpha(x) = \sum_i t_i \big(\beta(y_i) - c(x, y_i)\big) \le \alpha_c(x).$$

As $\alpha_c \le \alpha$ on X, this shows $\alpha(x) = \alpha_c(x)$ on X_Γ, and hence $\gamma(x) \subseteq \gamma_c(x)$ on X_Γ. Then for $x \in X_\Gamma$, by subtracting (3.13) from (3.14), we get $(\gamma_c(x) - \gamma(x)) \cdot (y - x) \ge 0$ for all $y \in \Gamma_x$. But since $x \in \mathrm{IC}(\Gamma_x)$, this implies

$$\big(\gamma_c(x) - \gamma(x)\big) \cdot (y - x) = 0 \qquad \text{for all } y \in \Gamma_x.$$

In other words, the projection of $\gamma(x)$ and $\gamma_c(x)$ onto the affine subspace generated by Γ_x are equal. Now note that (3.14) obviously holds for β_{cc} in place of β. Again by subtraction, we get $\beta_{cc}(y) \le \beta(y)$ for all $y \in \Gamma_x$. As the reverse inequality is already shown, we see that $\beta = \beta_{cc}$ on Y_Γ. Moreover, if $(x, y) \in \Gamma$, in other words if (x, y) satisfies (3.13), then the above discussion implies that (3.13) holds with $(\alpha_c, \gamma_c, \beta_{cc})$. In other words, $(x, y) \in \Gamma_{(\alpha_c, \gamma_c, \beta_{cc})}$.

For item (5), we first note that β_{cc} is defined on $\mathbb{R}^d$ and by item (2), we have $\beta_{cc} \le \beta_{cccc}$. For the reverse inequality, fix $z \in \mathbb{R}^d$. Then by definition of β_{cc}, there exist a sequence $\{x_n\}$ in Ω and $b_n \in \gamma_c(x_n)$, $n \ge 1$, such that

$$\beta_{cc}(y) \le c(x_n, y) + b_n \cdot (y - x_n) + \alpha_c(x_n) \qquad \text{for every } y \in \mathbb{R}^d, \quad \text{and}$$
$$\beta_{cc}(z) = \lim_{n\to\infty} c(x_n, z) + b_n \cdot (z - x_n) + \alpha_c(x_n).$$

This readily implies that $\beta_{cc}(z) = \beta_{cccc}(z)$, completing the proof of the theorem. □

REMARK 3.3. Note that both costs $c(x, y) = |x - y|$ and $c(x, y) = -|x - y|$ satisfy the above hypothesis, and in both cases, that is, $c(x, y) = \pm|x - y|$, we have that $\delta_1 = 0$. Moreover, $\delta_2 \le 2\,\mathrm{diam}(\Omega)$ if $c(x, y) = -|x - y|$.

If $c(x, y) = \rho(x, y)$ where ρ is a metric, then $\delta_1 = \delta_2 = 0$, which means that $\alpha_c = \beta_{cc}$ on Ω. In particular, by Theorem 3.2(5), the duality theorem becomes

$$\min\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} \rho(x, y)\, d\pi; \pi \in \mathrm{MT}(\mu, \nu)\right\}$$
$$= \sup\left\{\int_{\mathbb{R}^d} \beta d(\nu - \mu); \beta \text{ is martingale } c\text{-convex}\right\},$$

which can be seen as the counterpart of the Kantorovich–Rubenstein duality formulation in standard transport theory, whenever the cost is given by a metric.

REMARK 3.4 (Localization). Use the assumptions of Theorem 3.2, and let K be a compact set in Ω and let α_c^K and γ_c^K be the restrictions of α_c and γ_c on K. Then $(\alpha_c^K, \gamma_c^K, \beta_{cc}^K) \in E_m(c, K, \mathbb{R}^d)$, where

$$\beta_{cc}^K(y) := \inf_{x\in K}\{c(x, y) + \gamma_c^K(x) \cdot (y - x) + \alpha_c^K(x)\}.$$

Consequently, α_c^K is Lipschitz in K, and γ_c^K is bounded in K. Moreover, β_{cc}^K is Lipschitz (resp., locally Lipschitz) in $\mathbb{R}^d$ provided $y \mapsto c(x, y)$ is Lipschitz (resp., locally Lipschitz) in $\mathbb{R}^d$.

Indeed, from the definition of β_{cc}^K, the boundedness of γ_c on K and the local Lipschitz assumption on $y \mapsto c(x, y)$ (uniformly in x), we see that β_{cc}^K is the infimum of local Lipschitz functions parametrized by $x \in K$ with the local Lipschitz constant uniform in x. This shows that β_{cc}^K is locally Lipschitz in $\mathbb{R}^d$. If in addition, c is Lipschitz, then by the same reasoning β_{cc}^K is Lipschitz in $\mathbb{R}^d$.

4. Extremal structure of a *c*-contact layer. We first deal with the differentiability properties of an admissible triple (α, γ, β). The next lemma shows that essentially γ is differentiable in an appropriate sense, wherever α is. This property will be crucial in the proof of Theorem 2.4.

LEMMA 4.1. *Suppose $x \mapsto c(x, y)$ is differentiable at x whenever $x \neq y$, and assume that Γ is a set in $\mathcal{S}_{\mathrm{MT}}$ that is a c-contact layer for a triple $(\alpha, \gamma, \beta) \in E_m(c, \Omega, \mathbb{R}^d)$, where $\alpha : \Omega \to \mathbb{R}$, $\beta : \mathbb{R}^d \to \mathbb{R}$, and $\gamma : \Omega \to \mathbb{R}^d$. Fix $x \in X_\Gamma$, and let V be the vector subspace of $\mathbb{R}^d$ corresponding to the affine space $V(\Gamma_x)$, and assume* $\dim(V) \geq 1$. *Assume there is $s \in V$ such that*

$$\alpha(x') \leq s \cdot (x' - x) + \alpha(x) + o(|x' - x|) \qquad \text{as } x' \to x \text{ in } V(\Gamma_x). \tag{4.1}$$

Let $\operatorname{proj}_V \gamma$ *be the orthogonal projection of the value of γ on V. Then α and* $\operatorname{proj}_V \gamma$ *have a directional derivative at x in every direction $u \in V$.*

PROOF. Since Γ is a c-contact layer for a triple $(\alpha, \gamma, \beta) \in E_m(c, \Omega, \mathbb{R}^d)$, for all $x' \in \Omega$ and all $(x, y) \in \Gamma$,

$$\begin{aligned} &c(x', y) + \gamma(x') \cdot (y - x') + \alpha(x') \\ &\qquad \geq c(x, y) + \gamma(x) \cdot (y - x) + \alpha(x). \end{aligned} \tag{4.2}$$

Choose a unit vector $u \in V$ and let $x' = x + tu$. Then (4.2) is rewritten as

$$\begin{aligned} &\frac{\alpha(x + tu) - \alpha(x)}{t} \\ &\qquad \geq \frac{\gamma(x + tu) - \gamma(x)}{t} \cdot (x + tu - y) + \gamma(x) \cdot u \\ &\qquad\quad - \frac{c(x + tu, y) - c(x, y)}{t} \qquad \text{if } t > 0; \end{aligned} \tag{4.3}$$

$$\begin{aligned} &\frac{\alpha(x + tu) - \alpha(x)}{t} \\ &\qquad \leq \frac{\gamma(x + tu) - \gamma(x)}{t} \cdot (x + tu - y) + \gamma(x) \cdot u \\ &\qquad\quad - \frac{c(x + tu, y) - c(x, y)}{t} \qquad \text{if } t < 0. \end{aligned} \tag{4.4}$$

Let us use the notation $D_{t,u}f(x) = \frac{f(x+tu)-f(x)}{t}$. Now the assumption (4.1) says that

$$\limsup_{t\downarrow 0} D_{t,u}\alpha(x) \le s\cdot u \le \liminf_{t\uparrow 0} D_{t,u}\alpha(x). \tag{4.5}$$

Since $x \in \operatorname{int}(\operatorname{conv}(\Gamma_x))$, there exists $y_1,\dots,y_k \in \Gamma_x \setminus \{x\}$, $p_1,\dots,p_k \ge 0$, $q_1,\dots,q_k \ge 0$, $\Sigma p_i = 1$, $\Sigma q_i = 1$, $t_+ > 0$, $t_- < 0$, such that

$$x + t_+ u = \Sigma p_i y_i,$$

$$x + t_- u = \Sigma q_i y_i.$$

Note that the first term on the right-hand side of (4.3) and (4.4) is linear in y, so by summing up the y_i's with the weights p_i's or q_i's, we get [and we write $\gamma_1(x) := \gamma(x)\cdot u$]

$$D_{t,u}\alpha(x) \ge D_{t,u}\gamma_1(x)(t - t_\pm) + C_\pm(t) \qquad \text{if } t > 0, \tag{4.6}$$

$$D_{t,u}\alpha(x) \le D_{t,u}\gamma_1(x)(t - t_\pm) + C_\pm(t) \qquad \text{if } t < 0. \tag{4.7}$$

Here, $C_+(t)$, $C_-(t)$ are functions of $t \ne 0$, but have limits as $t \to 0$ by the differentiability assumption on the cost. Write $C_\pm = \lim_{t\to 0} C_\pm(t)$, respectively.

By taking $\limsup_{t\downarrow 0}$ in (4.6) and $\liminf_{t\uparrow 0}$ in (4.7) and by recalling that $t_+ > 0$, $t_- < 0$, we have

$$\limsup_{t\downarrow 0} D_{t,u}\alpha(x) \ge (-t_+)\liminf_{t\downarrow 0} D_{t,u}\gamma_1(x) + C_+,$$

$$\limsup_{t\downarrow 0} D_{t,u}\alpha(x) \ge (-t_-)\limsup_{t\downarrow 0} D_{t,u}\gamma_1(x) + C_-,$$

$$\liminf_{t\uparrow 0} D_{t,u}\alpha(x) \le (-t_+)\limsup_{t\uparrow 0} D_{t,u}\gamma_1(x) + C_+,$$

$$\liminf_{t\uparrow 0} D_{t,u}\alpha(x) \le (-t_-)\liminf_{t\uparrow 0} D_{t,u}\gamma_1(x) + C_-.$$

This and (4.5) combine to give

$$\begin{aligned}\liminf_{t\uparrow 0} D_{t,u}\gamma_1(x) &\ge \limsup_{t\downarrow 0} D_{t,u}\gamma_1(x) \ge \liminf_{t\downarrow 0} D_{t,u}\gamma_1(x)\\ &\ge \limsup_{t\uparrow 0} D_{t,u}\gamma_1(x) \ge \liminf_{t\uparrow 0} D_{t,u}\gamma_1(x),\end{aligned}$$

that is, $\gamma_1 = \gamma\cdot u$ is differentiable at x in the direction u. Knowing this, we then take $\liminf_{t\downarrow 0}$ in (4.6) and $\limsup_{t\uparrow 0}$ on (4.7) to get

$$\liminf_{t\downarrow 0} D_{t,u}\alpha(x) \ge (-t_+)\nabla_u\gamma_1(x) + C_+,$$

$$\limsup_{t\uparrow 0} D_{t,u}\alpha(x) \le (-t_+)\nabla_u\gamma_1(x) + C_+.$$

Combining this with (4.5), we get the differentiability of α at x in the direction of u.

Next, choose any unit vector $v \in V$ orthogonal to u and let $\gamma_2(x) := \gamma(x) \cdot v$. We want to show that $\nabla_u \gamma_2(x)$ exists. We proceed just as before; for some $k \in \mathbb{N}$, there exists $y_1, \dots, y_k \in \Gamma_x \setminus \{x\}$, $p_1, \dots, p_k \geq 0$, $q_1, \dots, q_k \geq 0$, $\Sigma p_i = 1$, $\Sigma q_i = 1$, $t_+ > 0$, $t_- < 0$, such that

$$x + t_+ v = \Sigma p_i y_i,$$
$$x + t_- v = \Sigma q_i y_i.$$

By summing up the y_i's and the weights p_i's or q_i's as before, we get this time

$$D_{t,u}\alpha(x) \geq t D_{t,u}\gamma_1(x) - t_\pm D_{t,u}\gamma_2(x) + C_\pm(t) \qquad \text{if } t > 0, \tag{4.8}$$

$$D_{t,u}\alpha(x) \leq t D_{t,u}\gamma_1(x) - t_\pm D_{t,u}\gamma_2(x) + C_\pm(t) \qquad \text{if } t < 0. \tag{4.9}$$

Taking $\limsup_{t\downarrow 0}$ in (4.8) and $\liminf_{t\uparrow 0}$ in (4.9) and recalling $t_+ > 0, t_- < 0$, and the existence of $\lim_{t\to 0} D_{t,u}\gamma_1(x)$, we see that

$$\nabla_u \alpha(x) \geq (-t_+) \liminf_{t\downarrow 0} D_{t,u}\gamma_2(x) + C_+,$$

$$\nabla_u \alpha(x) \geq (-t_-) \limsup_{t\downarrow 0} D_{t,u}\gamma_2(x) + C_-,$$

$$\nabla_u \alpha(x) \leq (-t_+) \limsup_{t\uparrow 0} D_{t,u}\gamma_2(x) + C_+,$$

$$\nabla_u \alpha(x) \leq (-t_-) \liminf_{t\uparrow 0} D_{t,u}\gamma_2(x) + C_-,$$

which implies differentiability of $\gamma_2 = \gamma \cdot v$ at x in the direction u. Now choose an orthonormal basis $\{u, v_1, \dots, v_m\}$ of V and write $\mathrm{proj}_V \gamma = (\gamma \cdot u)u + \Sigma(\gamma \cdot v_i)v_i$. We observed that each component of $\mathrm{proj}_V \gamma$ is directionally-differentiable. This completes the proof. $\square$

We now restrict our attention to the cases $c(x, y) = \pm|x - y|$ in trying to describe the profile of a set Γ that is a c-contact layer.

LEMMA 4.2. *Let $\Gamma \in \mathcal{S}_{\mathrm{MT}}$, Ω an open set in $\mathbb{R}^d$ containing X_Γ, $\alpha : \Omega \to \mathbb{R}$ and $\gamma : \Omega \to \mathbb{R}^d$ be two functions. Let $\beta : \mathbb{R}^d \to \mathbb{R}$ be either*

$$\beta(y) = \sup_{x\in\Omega}\{|x - y| + \gamma(x) \cdot (y - x) + \alpha(x)\} \tag{4.10}$$

or

$$\beta(y) = \inf_{x\in\Omega}\{|x - y| + \gamma(x) \cdot (y - x) + \alpha(x)\}. \tag{4.11}$$

Assume that Γ satisfies

$$\beta(y) = |x - y| + \gamma(x) \cdot (y - x) + \alpha(x) \qquad \textit{for all } (x, y) \in \Gamma. \tag{4.12}$$

If α and γ are differentiable at $x \in X_\Gamma$, then the closure $\overline{\Gamma_x}$ coincides with the set of extreme points of the convex hull of $\overline{\Gamma_x}$, that is, $\overline{\Gamma_x} = \mathrm{Ext}(\mathrm{conv}(\overline{\Gamma_x}))$.

PROOF. First note that, for any closed set A in $\mathbb{R}^d$, it is clear that $\mathrm{Ext}(\mathrm{conv}(A)) \subseteq A$. To show the reverse inclusion in our setting, we define the "tilted cone"

$$\zeta(x, y) = \zeta_x(y) = \zeta_y(x) := |x - y| + \gamma(x) \cdot (y - x) + \alpha(x).$$

The duality condition (4.12) with (4.10) tells us the following: if $(x, y) \in \Gamma$, then for all $x' \in \Omega$,

$$\zeta_{x'}(y) \le \zeta_x(y), \tag{4.13}$$

or (4.12) with (4.11) we get the reverse inequality.

Note that since $\zeta_x(y)$ is continuous, the same inequality holds for all $y \in \overline{\Gamma_x}$. This obviously implies that, if $y \in \overline{\Gamma_x}$ and $x \neq y$, then the gradient with respect to x vanishes:

$$\nabla \zeta_y(x) = 0 \tag{4.14}$$

and in fact (4.13) also implies that if $y \in \overline{\Gamma_x}$, then necessarily $x \neq y$. (If $x = y$, then the function $\zeta_y(x)$ strictly increases as x moves along the direction $\nabla_x[\gamma(x) \cdot (y - x) + \alpha(x)]$.) We may call this as nonstaying property or unstability, for the maximization problem. For the minimization problem, without loss of generality we already assumed that $x \notin \Gamma_x$, but in fact $x \notin \overline{\Gamma_x}$ as well, by (4.15) below.

Now suppose the lemma is false. Then we can find $\{y, y_0, \dots, y_s\} \subseteq \overline{\Gamma_x}$ for some $s \ge 1$ with $y = \sum_{i=0}^s p_i y_i$, $\sum_{i=0}^s p_i = 1$. Choose a minimum s such that all $p_i > 0$. Now taking directional derivative in the direction $u = \frac{x-y}{|x-y|}$ gives

$$\nabla_u \zeta_y(x) = \nabla_u \zeta_{y_i}(x) = 0 \qquad \forall i = 0, 1, \dots, s.$$

We compute

$$\nabla_u \zeta_{y_i}(x) = \frac{x - y_i}{|x - y_i|} \cdot u + \nabla_u \gamma(x) \cdot (y_i - x) - \gamma(x) \cdot u + \nabla_u \alpha(x).$$

Then, by the linearity of $y \mapsto \nabla_u \gamma(x) \cdot y$, the equation $\nabla \zeta_y(x) = 0$ simply becomes

$$1 = \sum_{i=0}^s p_i \frac{x - y_i}{|x - y_i|} \cdot \frac{x - y}{|x - y|}.$$

As $\frac{x-y}{|x-y|}$ is a unit vector and all $p_i > 0$, this can hold only if all y_i lie on the ray emanated from x. The minimality of s then implies that $s = 1$, hence $\{y, y_0, y_1\} \subseteq \overline{\Gamma_x}$ would lie on a ray emanating from x, which is a contradiction, once we prove the following claim:

$$\begin{aligned} &\overline{\Gamma_x} \text{ is contained in the topological boundary of the closed} \\ &\qquad \text{convex hull of } \overline{\Gamma_x}. \end{aligned} \tag{4.15}$$

Recall that here the topology is not the topology in $\mathbb{R}^d$ but the topology in $V := V(\Gamma_x)$. If our claim is false and assuming first that $\dim(V) \geq 2$, we can find $y \in \Gamma_x \cap \mathrm{IC}(\Gamma_x)$ and finitely many points $\{y_0, y_1, \dots, y_s\} \subseteq \Gamma_x$ such that $y = \sum_{i=0}^{s} p_i y_i$, $\sum p_i = 1$, $p_i > 0$ and, furthermore, $\{y_0, y_1, \dots, y_s\}$ are not aligned. But then the above argument implies that all y_i's have to be aligned, which is a contradiction. If $\dim(V) = 1$, then as $x \in \mathrm{IC}(\Gamma_x)$, we can find $\{y, y_0, y_1\} \subseteq \Gamma_x$ such that x and y are in the interior of the line segment $\overline{y_0 y_1}$. But then again by above, $\{y, y_0, y_1\}$ must lie on the ray (i.e., half-line) emanated from x in the direction of u, a contradiction. Finally, $\dim(V) = 0$ simply means that $\Gamma_x = \{x\}$, for which the claim obviously holds. But note that the case $\Gamma_x = \{x\}$ cannot happen when α and γ are differentiable at x, as we already showed above that then $x \notin \Gamma_x$ in the case of maximization, while we already assumed without loss of generality that $x \notin \Gamma_x$ in the case of minimization. $\square$

Finally, the following result follows immediately from Theorem 3.2, Lemmata 4.1 and 4.2.

COROLLARY 4.3. *Let $c(x, y) = \pm|x - y|$ and assume Γ is a c-contact layer in $\mathcal{S}_{\mathrm{MT}}$. If $X_\Gamma \subseteq \Omega := \mathrm{IC}(Y_\Gamma)$ with Ω being an open set in $\mathbb{R}^d$, then for $\mathcal{L}^d$-a.e. x in Ω, the closure $\overline{\Gamma_x}$ coincides with the set of extreme points of the convex hull of $\overline{\Gamma_x}$, that is, $\overline{\Gamma_x} = \mathrm{Ext}(\mathrm{conv}(\overline{\Gamma_x}))$.*

5. Structure of optimal martingale supporting sets when the dual is attained. The goal of this section is to prove Theorem 2.4 which shows that dual attainment in the optimization problem (1.1) implies that any optimal martingale transport is concentrated on a c-contact layer and, therefore, has a specific extremal structure. We start by collecting the properties verified by a well-chosen concentration set of a martingale measure. The proof is given in Appendix A.

LEMMA 5.1. *Let $\pi \in \mathrm{MT}(\mu, \nu)$ and let $\Lambda \subseteq \mathbb{R}^d \times \mathbb{R}^d$ be a Borel set with $\pi(\Lambda) = 1$. Then there exists a Borel set $\Gamma \subseteq \Lambda$ with $\pi(\Gamma) = 1$ such that the map $x \mapsto \pi_x$ is measurable and defined everywhere on X_Γ in such a way that:*

1. *$\overline{\Gamma_x} = \mathrm{supp}\, \pi_x$ for all $x \in X_\Gamma$.*
2. *$\Gamma \in \mathcal{S}_{\mathrm{MT}}$, that is $x \in \mathrm{IC}(\Gamma_x)$ for all $x \in X_\Gamma$.*
3. *If we assume that $\mu \ll \mathcal{L}^d$, then Γ can be chosen in such a way that $X_\Gamma \subseteq \mathrm{IC}(Y_\Gamma)$.*
4. *If, in addition, π is a solution of the optimization problem (1.1), then Γ can be chosen to be finitely c-exposable.*

This leads us to use the following terminology.

DEFINITION 5.2. Let π be a martingale transport plan in $\mathrm{MT}(\mu, \nu)$. We shall say that:

1. Γ is a *regular concentration set for* π if Γ satisfies (1), (2), (3) in Lemma 5.1.

2. Γ is a *martingale-monotone regular concentration set for* π (or simply Γ is *martingale-monotone regular for* π) if Γ also satisfies (4).

As mentioned in the Introduction, there is a dual formulation for problem (1.2), just like in the Monge–Kantorovich theory for (non-martingale) mass transport.

LEMMA 5.3 (See, e.g., [3]). *Let μ and ν be two probability measures on $\mathbb{R}^d$ in convex order, and let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a cost function that is lower semicontinuous, then*

$$\begin{aligned} &\min\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\,d\pi;\pi\in \mathrm{MT}(\mu,\nu)\right\} \\ &\quad = \sup\left\{\int_{\mathbb{R}^d}\beta\,d\nu - \int_{\mathbb{R}^d}\alpha\,d\mu; (\alpha,\gamma,\beta)\in E_m \right. \\ &\qquad \left. \textit{for some } \gamma\in C_b(\mathbb{R}^d,\mathbb{R}^d)\right\}, \end{aligned} \tag{5.1}$$

and the minimization problem is attained at some martingale transport π. A similar result holds for the cost maximization problem, provided c is upper semicontinuous, and E_m is replaced by E_M. Furthermore:

1. *If the dual maximization problem in* (5.1) *is attained, then there is a concentration set Γ for π that is a c-contact layer.*

2. *Conversely, if $G \subseteq \mathbb{R}^d \times \mathbb{R}^d$ is a c-contact layer induced by a triplet (α,γ,β) where $\alpha\in L^1(\mu)$, $\beta\in L^1(\nu)$, γ is bounded, and $\pi^*(G)=1$ for some $\pi^*\in \mathrm{MT}(\mu,\nu)$, then π^* is an optimal martingale transport.*

PROOF. For (5.1), see [3]. Let us show the items (1) and (2). Note that if the dual problem is attained at functions α, β such that the triplet (α,γ,β) is in $E_m(c,\mathbb{R}^d,\mathbb{R}^d)$, then since

$$\beta(y)-\alpha(x)-\gamma(x)(y-x)\le c(x,y) \qquad \text{for all } (x,y)\in\mathbb{R}^d\times\mathbb{R}^d, \tag{5.2}$$

μ and ν are the marginals of some optimal π in $\mathrm{MT}(\mu,\nu)$, and $\int \gamma(x)\cdot(y-x)\,d\pi(x,y)=0$ (due to the martingale condition), we then have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}\{\beta(y)-\alpha(x)-\gamma(x)(y-x)\}\,d\pi(x,y)=\int_{\mathbb{R}^d\times\mathbb{R}^d}c(x,y)\,d\pi(x,y).$$

It follows that

$$\beta(y)-\alpha(x)-\gamma(x)(y-x)= c(x,y) \qquad \text{for } \pi \text{ a.e. } (x,y)\in\mathbb{R}^d\times\mathbb{R}^d, \tag{5.3}$$

hence the equality holds on a concentration set Γ of π.

Conversely, if $G \subseteq \mathbb{R}^d \times \mathbb{R}^d$ and $\pi^*(G) = 1$ for some $\pi^* \in \mathrm{MT}(\mu, \nu)$ and if there exists a triplet (α, γ, β) in $E_m(c, X_G, Y_G)$ with equality (5.3) holding on G, then π^* is an optimal solution of the primal problem in (5.1). Indeed, let $\pi \in \mathrm{MT}(\mu, \nu)$ and let H be such that $\pi(H) = 1$. As $\mu(X_G) = 1$ and $\nu(Y_G) = 1$, by restriction we can then assume that $X_H \subseteq X_G$ and $Y_H \subseteq Y_G$, hence by integrating (5.2) with π, we get

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\pi(x, y) \geq \int_{\mathbb{R}^d} \beta(y)\, d\nu(y) - \int_{\mathbb{R}^d} \alpha(x)\, d\mu(x).$$

[Again $\int \gamma(x) \cdot (y - x)\, d\pi(x, y) = 0$ since $\pi \in \mathrm{MT}(\mu, \nu)$.] However, by integrating (5.2) with π^* and since we have equality on G, we get

$$\begin{aligned}\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\pi^* &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \{\beta(y) - \alpha(x) - \gamma(x)(y - x)\}\, d\pi^* \\ &= \int_{\mathbb{R}^d} \beta(y)\, d\nu - \int_{\mathbb{R}^d} \alpha(x)\, d\mu.\end{aligned}$$

This shows that π^* is optimal. Hence, every martingale measure that is concentrated on a c-contact layer induced by a triplet in the item (2) is optimal. On the other hand, there exist optimal martingale measures that do not concentrate on c-contact layers [3]. □

This suggests that dual attainability is actually a property of the support of the optimal martingale transport and not of the measure itself. Now an obvious but important remark is that any subset of a c-contact layer is also a c-contact layer. The same holds for dual attainment in the martingale transport problem. Indeed, if $\pi \in \mathrm{MT}(\mu, \nu)$ and B is a Borel set, we denote by π_B its restriction on $B \times \mathbb{R}^d$, and we let μ_B, ν_B be the first and second marginals of π_B. Then we introduce the following.

DEFINITION 5.4. Let $\pi \in \mathrm{MT}(\mu, \nu)$ be given, and let B be a Borel set. We say that an admissible triple $(\alpha, \gamma, \beta) \in E_m(c, B, \mathbb{R}^d)$ is *c-dual to π on B*, if the following holds:

$$\int_{\mathbb{R}^d} \beta(y)\, d\nu_B(y) - \int_{\mathbb{R}^d} \alpha(x)\, d\mu_B(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\, d\pi_B(x, y). \tag{5.4}$$

If such a triple exists, then we say that π admits a *c-dual on B*. Note that in this case, $\pi_B(\Gamma_{(\alpha,\gamma,\beta)}) = \mu(B)$, that is, π_B is concentrated on a c-contact layer.

Now, we can deduce the following.

THEOREM 5.5. *Let $c(x, y) = \pm|x - y|$ and μ be a probability measure that is absolutely continuous with respect to the Lebesgue measure. If $\pi \in \mathrm{MT}(\mu, \nu)$ is a solution of* (1.1) *for either the minimization or maximization problem that admits a c-dual on a Borel subset B, then for μ-almost all $x \in B$,* $\operatorname{supp} \pi_x = \operatorname{Ext}(\operatorname{conv}(\operatorname{supp} \pi_x))$.

PROOF. Let (α, γ, β) be a c-dual to π on B and let Λ be its contact layer. Then Λ contains the full measure [that is, $\mu(B)$] of π_B. Apply Lemma 5.1 to get $\Gamma \subseteq \Lambda$ in such a way that $\pi_B(\Gamma) = \mu(B)$, $\Gamma \in \mathcal{S}_{\mathrm{MT}}$, $X_\Gamma \subseteq \Omega := \mathrm{IC}(Y_\Gamma)$ and $\operatorname{supp} \pi_x = \overline{\Gamma_x}$ for μ a.e. $x \in B$. Now since Γ is also a c-contact layer, Corollary 4.3 applies to get the claimed result. $\square$

REMARK 5.6. Note that the above theorem shows that Conjecture 1 is valid provided *duality is attained locally*. In other words, if for any x in the support of μ, there exists a ball B centered at x such that the optimal martingale measure π admits a c-dual on B. This refinement will be used in the next section. On the other hand, there exists an optimal martingale measure where "local dual attainment" does not hold on any neighborhood. This can be seen with the following example given in [3].

EXAMPLE 5.7. Let $\mu = \nu$ be two identical probability measures on the interval $[0, 1]$, then the only martingale (say π) from μ to itself is the identity transport, hence it is obviously the solution of the maximization problem with respect to the distance cost, and its support is $\Gamma = \{(x, x) : x \in [0, 1]\}$. If now $\{\alpha, \gamma, \beta\}$ is a solution to the dual problem, then

$$\beta(y) \geq |x - y| + \gamma(x) \cdot (y - x) + \alpha(x) \qquad \forall x \in [0, 1], \forall y \in [0, 1];$$
$$\beta(y) = |x - y| + \gamma(x) \cdot (y - x) + \alpha(x) \qquad \forall (x, y) \in \Gamma.$$

The above relations easily yield that for any $0 < a < b < 1$, we have $\gamma(a) + 2 \leq \gamma(b)$, which means that it is impossible to define a suitable real-valued function γ for a.e. x in $[0, 1]$.

6. When the marginals are in subharmonic order. In this section, we consider a case where the dual martingale problem is attained—at least locally—which will allow us to apply Theorem 5.5 and verify that Conjecture 1 holds in that particular case. We consider the following "balayage order" between probability measures, that is, stronger than the convex order in dimensions greater than one. We say that probability measures μ and ν are in *subharmonic order*, $\mu \leq_{\mathrm{SH}} \nu$, if

$$\int_{\mathbb{R}^d} \varphi \, d\mu \leq \int_{\mathbb{R}^d} \varphi \, d\nu \qquad \text{for every subharmonic function } \varphi \text{ on } \mathbb{R}^d. \tag{6.1}$$

For simplicity, we shall assume that μ and ν have compact support so as to avoid integrability issues. Since convex functions are subharmonic, it is clear that $\mu \leq_{\mathrm{SH}} \nu \Rightarrow \mu \leq_C \nu$ and that the two notions are equivalent in one dimension.

Note that if $(B_t)_t$ is a d-dimensional Brownian motion with initial distribution μ and if ν is the distribution of B_T where T is a stopping time such that $(B_{T\wedge t})_t$ is a uniformly integrable martinagle, then $\mu \leq_{\mathrm{SH}} \nu$. Such stopping times are normally called *standard*. The converse is also true and belongs to a family of results known as Skorokhod embeddings (e.g., see Obłój [23]). In other words, (6.1) is essentially equivalent to

(6.2)
$$\mu \sim B_0 \text{ and } \nu \sim B_T \text{ for a (possibly randomized) standard stopping time } T.$$

We now consider the Newtonian potential (or simply, potential) P_μ of a probability measure μ in $\mathbb{R}^d$ with compact support, that is,

$$P_\mu(x) = \begin{cases} \dfrac{1}{2\pi} \displaystyle\int \log|x-y|\, d\mu(y) & \text{for } d = 2, \\ \dfrac{1}{d(2-d)w_d} \displaystyle\int |x-y|^{2-d}\, d\mu(y) & \text{for } d \neq 2. \end{cases}$$

Then we have $\Delta P_\mu = \mu$ (in the sense of distributions), and (6.1) implies that

$$P_\mu(x) \le P_\nu(x) \qquad \forall x \in \mathbb{R}^d. \tag{6.3}$$

The converse is also true at least for $d \ge 3$; see Falkner [13].

Finally, note that if we consider an elliptic operator $L_t = \sum_{ij} a_{ij}(t)\partial_i\partial_j$ corresponding to a one-parameter family of positive matrices $(a_{ij}(t))$, $t > 0$, and if μ, μ_t are measures with densities ρ, ρ_t, respectively, where

$$\begin{cases} \partial_t \rho_t - L_t \rho_t = 0 & \text{for } t > 0 \text{ and in } \mathbb{R}^d, \\ \rho_0 = \rho, \end{cases} \tag{6.4}$$

then one can easily verify that $\mu \le_{\mathrm{SH}} \mu_t$. Actually, one can show that

$$P_\mu(x) < P_{\mu_t}(x), \qquad \forall x \in \mathbb{R}^d. \tag{6.5}$$

The importance of such a strict inequality will be clear thereafter. The following is the main result of this section.

THEOREM 6.1. *Assume $\mu \le_{\mathrm{SH}} \nu$ where μ, ν are probability measures with compact support on $\mathbb{R}^d$ such that $\mu \ll \mathcal{L}^d$. Assume the function $P_\nu - P_\mu$ lower semicontinuous and consider the open set $U := \{x \in \mathbb{R}^d \mid P_\nu(x) - P_\mu(x) > 0\}$. If $\pi \in \mathrm{MT}(\mu, \nu)$ is an optimal solution for the minimization problem* (1.1), *where the cost function is either $c(x, y) = |x - y|$ or $c(x, y) = -|x - y|$, then*:

1. *For each $x \in U$, there exists a ball B centered at x such that π admits a c-dual on B.*
2. *For μ-a.e. $x \in U$, $\operatorname{supp}\pi_x = \operatorname{Ext}(\operatorname{conv}(\operatorname{supp}\pi_x))$. In particular, Conjecture 1 holds if $\mu(U) = 1$.*

REMARK 6.2. The assumption that ν is compactly supported can be replaced with appropriate decay conditions on $P_\nu - P_\mu$ and $\nabla(P_\nu - P_\mu)$. In particular, Conjecture 1 holds for μ and $\nu = \mu_t$ from the diffusion example in (6.4), if the initial measure μ is absolutely continuous and compactly supported. Note that the two-dimensional case is true in full generality, that is, when the marginals are simply in convex order (see Section 7).

PROOF OF THEREOM 6.1. Denoting $E_m = E_m(c, \mathbb{R}^d, \mathbb{R}^d)$, we have from Lemma 5.3 that

$$\begin{aligned} l &:= \min\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\,d\pi; \pi \in \mathrm{MT}(\mu,\nu)\right\} \\ &= \sup\left\{\int_{\mathbb{R}^d}\beta\,d\nu - \int_{\mathbb{R}^d}\alpha\,d\mu; (\alpha,\gamma,\beta)\in E_m \text{ for some } \gamma\in C_b\big(\mathbb{R}^d,\mathbb{R}^d\big)\right\}. \end{aligned}$$

Let π be an optimal solution for the minimization problem, and let Γ be a martingale-monotone regular concentration set for π (as in Definition 5.2). Fix a bounded open set Ω which is sufficiently large such that $\mathrm{supp}(\mu) \subseteq \Omega$. We shall show that for each $x_0 \in U \cap \Omega$, there exists a ball $B = B(x_0) \subseteq U \cap \Omega$ centred at x, such that π has a c-dual on B.

For that, consider a maximizing sequence for the dual problem, that is admissible triples $(\alpha_n, \gamma_n, \beta_n) \in E_m(c, \mathbb{R}^d, \mathbb{R}^d)$ such that

$$l = \lim_{n\to\infty}\int \beta_n\,d\nu - \int \alpha_n\,d\mu. \tag{6.6}$$

In view of Theorem 3.2 and Remark 3.4, we can assume that the triplet $(\alpha_n, \gamma_n, \beta_n) \in E_m(c, \Omega, \mathbb{R}^d)$, that α is Lipschitz in Ω, γ is bounded in Ω, and that

$$\beta(x) \le \alpha(x) \le \beta(x) + \delta \qquad \text{for all } x \in \Omega, \tag{6.7}$$

where $0 \le \delta \le 2\,\mathrm{diam}(\Omega)$. Note that $\delta = 0$ if $c(x,y) = |x-y|$.

We consider the convex function

$$\chi_n(y) := \sup_{x\in\Omega}\big\{-\alpha_n(x) - \gamma_n(x)\cdot(y-x)\big\}.$$

Since α_n and γ_n are bounded and the set Ω is bounded, the functions χ_n are Lipschitz on $\mathbb{R}^d$. Note also that by adding a sequence of affine functions L_n [since $L_n(y) = L_n(x) + \nabla L_n(x)\cdot(y-x)$] the new sequence $(\alpha_n + L_n, \beta_n + L_n, \gamma_n + \nabla L_n)$ will still have the same properties. By adding appropriate affine function L_n and their gradients to the triple $(\alpha_n, \gamma_n, \beta_n)$, we may therefore assume that

$$\chi_n(x) = 0 \quad \text{and} \quad \chi_n \ge 0 \qquad \text{for every } n.$$

We now show that a subsequence of α_n, γ_n converge locally in $U \cap \Omega$. We first establish suitable estimates on χ_n. Consider the Lipschitz function $q(y) := \sup_{x\in\Omega} c(x,y)$ and note that

$$\begin{aligned} -\alpha_n(y) &\le \chi_n(y) \qquad \forall y \in \Omega \quad \text{and} \\ \chi_n(y) &\le q(y) - \beta_n(y) \qquad \forall y \in \mathbb{R}^d. \end{aligned} \tag{6.8}$$

Hence,

$$0 \le \int \chi_n(d\nu - d\mu) \le -\int \beta_n\,d\nu + \int \alpha_n\,d\mu + C_1,$$

where $C_1 = \int q(y)\,d\nu(y) < \infty$, since q is Lipschtiz and ν has finite first moment. Since $(\alpha_n, \gamma_n, \beta_n)$ is a maximizing sequence, then for all sufficiently large n,

$$0 \le \int \chi_n(d\nu - d\mu) \le -l + C_1 + 1 =: C_2.$$

Hence,

$$C_2 \ge \int \chi_n(d\nu - d\mu) = \int \chi_n \Delta(P_\nu - P_\mu) = \int \Delta\chi_n(P_\nu - P_\mu), \tag{6.9}$$

where $\Delta\chi_n$ is the distributional Laplacian of the convex function χ_n. For the second last equality, note that $\Delta P_\mu = \mu$, $\Delta P_\nu = \nu$, and for the last equality note that χ_n is convex Lipschitz and $P_\nu - P_\mu$, $\nabla(P_\nu - P_\mu)$ decays to zero at infinity by assumption, enabling us to integrate by parts.

Now fix $x_0 \in U \cap \Omega$ and pick a closed ball $B := B_r(x_0) \subseteq U \cap \Omega$ of radius r, centered at x_0. Since $P_\nu - P_\mu$ is lower semicontinuous and strictly positive on U, we have $\varepsilon_B := \min_B[P_\nu - P_\mu] > 0$ which, in view of (6.9), implies that

$$\int_{B_r(x_0)} \Delta\chi_n \le \frac{C_2}{\varepsilon_B}.$$

Now, modulo approximating it by smooth convex function, we can assume that χ_n is smooth and apply Proposition B.1 to conclude that χ_n is bounded in a smaller ball $B_{r'}(x_0)$, uniformly in n. In view of (6.7) and (6.8), the uniform boundedness of χ_n then implies the uniform boundedness of α_n, β_n on $B_{r'}(x_0)$. Moreover, since

$$-\alpha_n(x) - \gamma_n(x)\cdot(y - x) \le \chi_n(y) \le C \qquad \forall x, y \in B_{r'}(x_0),$$

we also find that γ_n is uniformly bounded in n on a smaller ball $B = B_{r''}(x_0)$, $r'' < r' < r$, in such a way that the sequences $(\alpha_n, \gamma_n, \beta_n)_n$ are all uniformly bounded on B.

Apply now Komlós theorem [20], which states that every L^1-bounded sequence of real functions has a subsequence such that the arithmetic means of all its subsequences converge pointwise almost everywhere. Since the arithmetic means of α_n, β_n, γ_n also yield a maximizing sequence of admissible triples for (6.6), we can therefore assume that the original functions α_n, β_n and γ_n converge $\mathcal{L}^d$ a.e. in B to, say $\overline{\alpha}$, $\overline{\beta}$ and $\overline{\gamma}$ on $X \subseteq B$ where $\mathcal{L}^d(B \setminus X) = 0$. Notice that these limits are bounded in X.

It is not clear, however, that this triple $(\overline{\alpha}, \overline{\gamma}, \overline{\beta})$ will give the desired one, especially because $\overline{\beta}$ is only defined in X, not in $\mathbb{R}^d$. We thus proceed as follows. Define

$$\beta_{X,n} = \inf_{x\in X}\{c(x, y) + \alpha_n(x) + \gamma_n(x)\cdot(y - x)\}.$$

Notice that since α_n, γ_n are bounded in X uniformly in n and $y \mapsto c(x, y)$ is Lipschitz in $\mathbb{R}^d$ with uniformly bounded Lipschitz constants for $x \in X$, we immediately see that the function $y \in \mathbb{R}^d \mapsto \beta_{X,n}(y)$ is Lipschitz (uniformly in n)

and is uniformly bounded on each compact set. Therefore, there exists a subsequence, which we still denote by $\beta_{X,n}$, that converges to a Lipschitz function $\overline{\beta}_X$ uniformly on each compact set in $\mathbb{R}^d$. Moreover, from the definition of $\beta_{X,n}$, the triple $(\alpha_n, \gamma_n, \beta_{X,n})$ satisfy

$$\beta_{X,n}(y) - \alpha_n(x) - \gamma_n(x)(y - x) \le c(x, y) \qquad \forall (x, y) \in X \times \mathbb{R}^d.$$

Thus $(\alpha_n, \gamma_n, \beta_{X,n}) \in E_m(c, X, \mathbb{R}^d)$. Also, taking the limit as $n \to \infty$, the above inequality still holds in the limit, and so the triple $(\overline{\alpha}, \overline{\gamma}, \overline{\beta}_X) \in E_m(c, X, \mathbb{R}^d)$.

To show that the triple $(\overline{\alpha}, \overline{\gamma}, \overline{\beta}_X)$ is a c-dual to π on B (in the sense of Definition 5.4), it remains to verify (5.4). For this, observe from the definition of $\beta_{X,n}$ that $\beta_n(y) \le \beta_{X,n}(y)$ for all $y \in \mathbb{R}^d$. Thus,

$$\int \beta_n \, d\nu_B - \int \alpha_n \, d\mu_B \le \int \beta_{X,n} \, d\nu_B - \int \alpha_n \, d\mu_B \le \int c(x, y) \, d\pi_B(x, y).$$

Noting that the maximizing sequence of admissible triple $(\alpha_n, \gamma_n, \beta_n)$ for (6.6) is also a maximizing sequence of admissible triple for π_B, that is,

$$\lim_{n\to\infty} \int \beta_n(y) \, d\nu_B(y) - \int \alpha_n(x) \, d\mu_B(x) = \int c(x, y) \, d\pi_B(x, y),$$

we therefore have that

$$\lim_{n\to\infty} \int \beta_{X,n}(y) \, d\nu_B(y) - \int \alpha_n(x) \, d\mu_B(x) = \int c(x, y) \, d\pi_B(x, y).$$

To bring the limit inside the integrals, recall that $\beta_{X,n}$ is uniformly Lipschitz (in n) and α_n is uniformly bounded, $\mu(B \setminus X) = 0$ and ν_B has finite first moment. Thus, by the dominated convergence theorem,

$$\int \overline{\beta}_X(y) \, d\nu_B(y) - \int \overline{\alpha}(x) \, d\mu_B(x) = \int c(x, y) \, d\pi_B(x, y).$$

Therefore, the triple $(\overline{\alpha}, \overline{\gamma}, \overline{\beta}_X)$ is a c-dual to π on B, proving the item (1). Then by Theorem 5.5, for μ a.e. $x \in B$ we have that $\operatorname{supp} \pi_x = \operatorname{Ext}(\operatorname{conv}(\operatorname{supp} \pi_x))$. As U can be covered by countably many such balls B, for μ a.e. $x \in U$ we have that $\operatorname{supp} \pi_x = \operatorname{Ext}(\operatorname{conv}(\operatorname{supp} \pi_x))$, proving the item (2). □

7. A canonical decomposition for the support of martingale transports. We have shown in the last sections that Conjecture 1 holds whenever the dual problem is (locally) attained. In this section, we shall decompose an optimal martingale transport π into components on which an induced martingale transport problem is defined in such a way that its dual problem is attained. For that, we shall first associate to any Borel set $\Gamma \in \mathcal{S}_{\mathrm{MT}}$ a unique irreducible convex paving Φ. We then show that if every finite subset of Γ is a c-contact layer (a property satisfied by a concentration set of an optimal martingale measure), then every subset $\Gamma_C = \Gamma \cap (C \times \mathbb{R}^d)$ where C is a component of the convex paving Φ, is a c-contact layer.

7.1. *Irreducible convex pavings associated to martingale supporting sets*. Let Γ be a Borel set in $\mathcal{S}_{\mathrm{MT}}$. We start by defining an equivalence relation on X_Γ. For each $x \in X := X_\Gamma$, we define inductively an increasing sequence of convex open sets $(C_n(x))_n$ in the following way:

Start with the trivial equivalence relation $x \sim_0 x'$ iff $x = x'$. Let $C_0(x) := \mathrm{IC}(\Gamma_x)$ and recall that if $\Gamma_x = \{x\}$, then $C_0(x) = \{x\}$. Now define the following equivalence relation on X: $x \sim_1 x'$ if there exist finitely many $x_1, \dots, x_k$ in X such that the following chain condition holds:

$$\begin{aligned} C_0(x) \cap C_0(x_1) &\neq \varnothing, \\ C_0(x_i) \cap C_0(x_{i+1}) &\neq \varnothing \qquad \forall i = 1, 2, \dots, k-1, \\ C_0(x_k) \cap C_0(x') &\neq \varnothing. \end{aligned}$$

We then consider the open convex hull:

$$C_1(x) := \mathrm{IC}\Big[\bigcup_{x' \sim_1 x} C_0(x')\Big].$$

Note that $x \sim_1 x'$ implies $C_1(x) = C_1(x')$. Unfortunately, the convex sets $C_1(x)$ do not determine the equivalence classes. In particular, they may not be mutually disjoint for elements that are not equivalent for $\sim_1$. So, we proceed to define $\sim_2$ in a similar way: $x \sim_2 x'$ if there exist finitely many $x_1, \dots, x_k$ in X such that the following chain condition holds:

$$\begin{aligned} C_1(x) \cap C_1(x_1) &\neq \varnothing, \\ C_1(x_i) \cap C_1(x_{i+1}) &\neq \varnothing \qquad \forall i = 1, 2, \dots, k-1, \\ C_1(x_k) \cap C_1(x') &\neq \varnothing; \end{aligned}$$

and we set

$$C_2(x) := \mathrm{IC}\Big[\bigcup_{x' \sim_2 x} C_1(x')\Big].$$

Again, $\sim_2$ is an equivalence relation and one can easily see that:

- $x \sim_1 x' \Rightarrow x \sim_2 x'$
- $x \sim_2 x' \Rightarrow C_2(x) = C_2(x')$
- $C_1(x) \subseteq C_2(x)$.

But still, the sets $C_2(x)$ may not be mutually disjoint for nonequivalent $x's$. We continue inductively in a similar fashion by defining equivalence relations $\sim_n$ for $n = 1, 2, \dots$ and their corresponding classes

$$C_n(x) := \mathrm{IC}\Big[\bigcup_{x' \sim_n x} C_{n-1}(x')\Big].$$

It is easy to check that we have the following properties for each n:

$$x \sim_n x' \Rightarrow x \sim_{n+1} x',$$
$$x \sim_n x' \Rightarrow C_n(x) = C_n(x'),$$
$$C_n(x) \subseteq C_{n+1}(x).$$

Finally, define the equivalence relation

$$x \sim x' \qquad \text{if } x \sim_n x' \text{ for some } n,$$

and its corresponding convex sets

$$C(x) := \lim_{n\to\infty} C_n(x) = \bigcup_{n=0}^{\infty} C_n(x). \tag{7.1}$$

Now, we show that $\Psi = \{C(x)\}_{x\in X}$ is an irreducible convex paving for Γ.

THEOREM 7.1. *The equivalence relation $\sim$ on X_Γ and the components $(C(x))_{x\in X_\Gamma}$ satisfy the following*:

1. $x \sim x' \Rightarrow C(x) = C(x')$, *and* $x \nsim x' \Rightarrow C(x) \cap C(x') = \varnothing$.
2. $C(x)$ *are mutually disjoint, that is, either* $C(x) = C(x')$ *or* $C(x) \cap C(x') = \varnothing$.
3. $x' \in X \cap C(x)$ *if and only if* $x' \sim x$.
4. $\Phi = \{C(x)\}_{x\in X}$ *is an irreducible convex paving for* Γ.
5. $C_n(x) = \mathrm{IC}[\bigcup_{x'\sim_n x} \Gamma_{x'}]$ *for* $n \geq 0$ *and* $C(x) = \mathrm{IC}[\bigcup_{x'\sim x} \Gamma_{x'}]$. *In particular*, $\Gamma_x \subseteq \overline{C(x)}$.

PROOF. The fact that $x \sim_n x' \Rightarrow C_n(x) = C_n(x')$ gives the first part of (1). If there exists a $z \in C(x) \cap C(x')$, then there is N such that $z \in C_N(x) \cap C_N(x')$, implying $x \sim_{N+1} x'$ and verifying the second part of (1) of which (2) and (3) are obvious consequences.

To prove (4), let Ψ be any convex paving of Γ and let $z, x \in X_\Gamma$, $D \in \Psi$ be such that $C(z) \cap D(x) \neq \varnothing$. We must show that $C(z) \subseteq D(x)$. We claim that, for any $n \geq 0$,

$$(*) \qquad C_n(z) \cap D(x) \neq \varnothing \quad \Rightarrow \quad C_n(z) \subseteq D(x), \qquad \text{for every } z, x \in X_\Gamma.$$

Indeed, it is true for $n = 0$ by definition. Assume that $(*)$ is true for some n, and suppose $C_{n+1}(z) \cap D(x) \neq \varnothing$. Note that $C_n(z) \subseteq D(z)$, and so if $w \sim_{n+1} z$, by $(*)$ we have that $C_n(w) \subseteq D(z)$. As $C_{n+1}(z) = \mathrm{IC}[\bigcup_{w\sim_{n+1} z} C_n(w)]$, this readily implies that $C_{n+1}(z) \subseteq D(z)$, but then $D(z) \cap D(x) \neq \varnothing$, and hence $D(z) = D(x)$. This proves $(*)$ for every $n \geq 0$. Now if $C(z) \cap D(x) \neq \varnothing$, then for all large n $C_n(z) \cap D(x) \neq \varnothing$, hence by $(*)$ we get that $C_n(z) \subseteq D(x)$. Therefore, $C(z) \subseteq D(x)$ which proves the irreducibility of Φ.

For (5), let $(A_i)_{i\in I}$ be any family of sets in $\mathbb{R}^d$, where I is an index set. Then it is easy to see that

$$\mathrm{IC}(A_i) = \mathrm{IC}(\mathrm{CC}(A_i)), \qquad \text{and}$$
$$\mathrm{CC}\Big(\bigcup_{i\in I}\mathrm{CC}(A_i)\Big) = \mathrm{CC}\Big(\bigcup_{i\in I} A_i\Big) = \mathrm{CC}\Big(\bigcup_{i\in I}\mathrm{IC}(A_i)\Big).$$

But note that $A \subseteq B$ does not imply $\mathrm{IC}(A) \subseteq \mathrm{IC}(B)$ in general. The above implies in particular

$$\mathrm{IC}\Big(\bigcup_{i\in I} A_i\Big) = \mathrm{IC}\Big(\bigcup_{i\in I}\mathrm{IC}(A_i)\Big).$$

In addition, a simple induction shows that for every $n \geq 0$, we have

$$C_n(x) = \mathrm{IC}\Big[\bigcup_{x'\sim_n x}\Gamma_{x'}\Big].$$

Indeed, it is true for $n = 0$ by definition. Suppose $C_n(x) = \mathrm{IC}[\bigcup_{x'\sim_n x}\Gamma_{x'}]$. Now by definition,

$$\begin{aligned} C_{n+1}(x) &= \mathrm{IC}\Big[\bigcup_{x'\sim_{n+1} x} C_n(x')\Big] \\ &= \mathrm{IC}\Big[\bigcup_{x'\sim_{n+1} x}\mathrm{IC}\Big(\bigcup_{x''\sim_n x'}\Gamma_{x''}\Big)\Big] \\ &= \mathrm{IC}\Big[\bigcup_{x'\sim_{n+1} x}\Big(\bigcup_{x''\sim_n x'}\Gamma_{x''}\Big)\Big]. \end{aligned}$$

But $\bigcup_{x'\sim_{n+1}x}\bigcup_{x''\sim_n x'}\Gamma_{x''} = \bigcup_{x'\sim_{n+1}x}\Gamma_{x'}$, hence, $C_{n+1}(x) = \mathrm{IC}(\bigcup_{x'\sim_{n+1}x}\Gamma_{x'})$, completing the induction.

Finally, we proceed as follows:

$$\begin{aligned} C(x) = \mathrm{IC}[C(x)] = \mathrm{IC}\Big[\bigcup_{n\geq 0} C_n(x)\Big] &= \mathrm{IC}\Big[\bigcup_{n\geq 0}\mathrm{IC}\Big(\bigcup_{x'\sim_n x}\Gamma_{x'}\Big)\Big] \\ = \mathrm{IC}\Big[\bigcup_{n\geq 0}\bigcup_{x'\sim_n x}\Gamma_{x'}\Big] = \mathrm{IC}\Big[\bigcup_{x'\sim x}\Gamma_{x'}\Big], \end{aligned}$$

which completes the proof of (5) and the theorem. $\square$

7.2. *When irreducible components are c-contact layers.* Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a cost function on which we make no assumption. Our aim is to prove Theorem 2.11, which will follow from the following.

THEOREM 7.2. *Let $\Gamma \in \mathcal{S}_{\mathrm{MT}}$ be c-finitely exposable. If Φ is the irreducible convex paving of Γ, then for every convex component C in Φ, the set $\Gamma \cap (C \times \mathbb{R}^d) = \Gamma \cap (C \times \overline{C})$ is a c-contact layer.*

First, we prove the following lemma.

LEMMA 7.3. *Let $\Gamma \in \mathcal{S}_{\mathrm{MT}}$ be c-finitely exposable, and denote $X := X_\Gamma$. Fix $x_0 \in X$ and set $G := \Gamma \cap (C(x_0) \times \mathbb{R}^d)$, where $C(x_0)$ is the component of the irreducible convex paving Φ of Γ that contains x_0. Then, for each $y \in Y_G$, there exists a compact interval $K_y \subseteq \mathbb{R}$ such that any finite subset $H \subseteq G$ is a c-contact layer for a triplet (α, γ, β), where $\beta(y) \in K_y$ for all $y \in Y_H$.*

The above lemma is essentially saying that there is some uniformity in the way c-admissible triplets can expose finite subsets of G as c-contact layers. This control on the β component of the c-admissible triplets will allow us to use Tychonoff's compactness theorem to deduce that the whole of G is a c-contact layer.

To prove Lemma 7.3, we first give an idea about the degrees of freedom we have in choosing β. First, note that if β is c-admissible for G (meaning that there is α, γ such that $G \subseteq \Gamma_{(\alpha,\gamma,\beta)}$) and $L : \mathbb{R}^d \to \mathbb{R}$ is an affine function, then $\beta - L$ is also a c-admissible for G. Letting $m = \dim(V(Y_G))$, we can find $\{y_0, \dots, y_m\} \subseteq Y_G$ such that $V(\{y_0, \dots, y_m\}) = V(Y_G)$, that is, $\{y_0, \dots, y_m\}$ constitute vertices of an m-dimensional polytope in $V(Y_G)$. Now for a given c-admissible function β for G, let $L : V(Y_G) \to \mathbb{R}$ be an affine function determined by $L(y_i) = \beta(y_i)$ for $i = 0, 1, \dots, m$. The function $\beta' := \beta - L$ then satisfies $\beta'(y_i) = 0$ for all $i = 0, 1, \dots, m$, which means that we have $m + 1$ degrees of freedom on the value of β. In other words, if we set $K_{y_i} = \{0\}$ for $i = 0, 1, \dots, m$, then we can find β' such that $\beta'(y_i) \in K_{y_i}$ for each y_i. Now, we want to observe how the initial value of β can control its values at other points y. We shall see that the control of the value of β propagates well along a given chain inside the equivalent class $C(x_0)$.

The proof of Lemma 7.3 is involved, and requires several key steps. To clarify the idea, we consider first the special case $c = 0$ where we can establish a complete control on the dual functions.

LEMMA 7.4. *Let $G \in \mathcal{S}_{\mathrm{MT}}$ and assume that it is a 0-contact layer for a triplet (α, γ, β), that is,*

$$\beta(y) \geq L_x(y) \qquad \forall x \in X_G, y \in Y_G, \tag{7.2}$$

$$\beta(y) = L_x(y) \qquad \forall (x, y) \in G, \tag{7.3}$$

where for each x, L_x is the affine function

$$L_x(y) := \gamma(x) \cdot (y - x) + \alpha(x).$$

Then $L_x = L_{x'}$ on $V(C(x))$ whenever $x \sim x'$.

Note that (7.3) says that if we have control on L_x, then we have control on β for all $y \in G_x$. In particular, Lemma 7.4 implies that if $L_x = 0$ (we can choose such L_x without loss of generality) then $L_{x'} = 0$ on $V(C(x))$ for all $x' \in C(x)$, thus $\alpha(x') = 0$ for all $x' \in C(x)$ and $\beta(y) = 0$ at each $y \in G_{x'}$.

The above lemma is a consequence of the following proposition.

PROPOSITION 7.5. *Let L_1, L_2 be two affine functions on $\mathbb{R}^d$, and let S_1, S_2 be sets in $\mathbb{R}^d$. Suppose that $L_1 \le L_2$ on S_1, and $L_2 \le L_1$ on S_2, and that* $\mathrm{IC}(S_1) \cap \mathrm{IC}(S_2) \neq \varnothing$. *Then $L_1 = L_2$ on $V(S_1 \cup S_2)$, the latter is the minimal affine space containing the sets S_1 and S_2.*

PROOF. This follows from two facts:

1. For affine functions, $L \le L'$ on a set S implies $L \le L'$ on $\mathrm{conv}(S)$.
2. If two affine functions L, L' satisfy $L \le L'$ on a set S and if moreover, $L(z) = L'(z)$ at some interior point of $\mathrm{conv}(S)$, then $L = L'$ on $\mathrm{conv}(S)$, thus on $V(S)$.

Indeed, apply (1) to the case $L = L_1$, $L' = L_2$ and $S = S_1$, and also to the case $L = L_2$, $L' = L_1$ and $S = S_2$. We get $L_1 = L_2$ on $\mathrm{conv}(S_1) \cap \mathrm{conv}(S_2)$. Now, from the assumption, $\mathrm{IC}(S_1) \cap \mathrm{IC}(S_2) \neq \varnothing$, and also obviously $\mathrm{IC}(S_1) \cap \mathrm{IC}(S_2) \subseteq \mathrm{IC}(S_i)$, $i = 1, 2$. Using (2), we then get that $L_1 = L_2$ on both $\mathrm{conv}(S_i)$, $i = 1, 2$. From this, the assertion follows. $\square$

PROOF OF LEMMA 7.4. First, note that for each $x, x' \in X = X_G$, conditions (7.2) and (7.3) yield that $L_{x'} \le L_x$ on G_x, and $L_x \le L_{x'}$ on $G_{x'}$.

Now to prove the lemma, it suffices to show that for each $n \in \{0, 1, 2, \dots, \}$,

$$\text{if } x \sim_n x' \text{ then } L_x = L_{x'} \text{ on } V\bigl(C_n(x)\bigr). \tag{7.4}$$

Here, by $x \sim_0 x'$ we mean $x = x'$. We do this inductively. Our induction hypothesis is (7.4) together with

$$L_z \le L_x \qquad \text{on } C_n(x), \text{ for each } z \in X \text{ and } x \in X. \tag{7.5}$$

For $n = 0$, (7.4) is trivially satisfied and (7.5) follows from (7.2) and (7.3). Now, assume that (7.4) and (7.5) hold for all $n \le k$. For $n = k + 1$, if $x \sim_{k+1} x'$ then there are $x = x_0, x_1, \dots, x_m = x'$ for some m, such that $C_k(x_i) \cap C_k(x_{i+1}) \neq \varnothing$ for each $0 \le i \le m - 1$. From this, and using (7.4) and (7.5), we can apply Proposition 7.5 with the choice $L_1 = L_{x_i}, L_2 = L_{x_{i+1}}$, $S_1 = C_k(x_i)$, $S_2 = C_k(x_{i+1})$, and see $L_{x_i} = L_{x_{i+1}}$ on $V(C_k(x_i) \cup C_k(x_{i+1}))$, for $i = 0, \dots, m - 1$. Similarly, repeated application of Proposition 7.5 eventually yields that $L_x = L_{x_i} = L_{x'}$ on each $C_k(x_i)$, for $i = 1, \dots, m$. Therefore, $L_x = L_{x'}$ on $\bigcup_i C_k(x_i)$, thus on $V(\bigcup_{i=0}^m C_k(x_i))$. This holds for any $x \sim_{k+1} x'$, thus by applying the result to all $z \sim_{k+1} x \sim_{k+1} x'$, we also see

$$L_x = L_{x'} \qquad \text{on } V\biggl(\bigcup_{z \sim_{k+1} x} C_k(z)\biggr) = V\bigl(C_{k+1}(x)\bigr),$$

verifying (7.4) for $n = k + 1$. For (7.5), for each $z \in X$, from the assumption (7.5) for $n \leq k$ and applying (7.4), we have $L_z \leq L_{x'} = L_x$ on $C_k(x')$ for all $x' \sim_{k+1} x$. For the affine functions, this implies $L_z \leq L_x$ on $C_{k+1}(x)$. This completes the induction argument, so the proof. $\square$

We now consider the case of a nontrivial cost c. We first establish a more quantitative version of Proposition 7.5.

PROPOSITION 7.6. *Let L_1, L_2 be two affine functions on $\mathbb{R}^d$, and let S_1, S_2 be sets in $\mathbb{R}^d$. Suppose that*:

- *$L_1 \leq L_2 + \delta_1$ on S_1, and $L_2 \leq L_1 + \delta_2$ on S_2 for some constants $\delta_1, \delta_2 > 0$;*
- *there is a point z in $\mathrm{IC}(S_1) \cap \mathrm{IC}(S_2)$.*

Then $|L_1 - L_2| \leq C$ on $\mathrm{conv}(S_1 \cup S_2)$. Here, $C = C(z, S_1, S_2, \delta_1, \delta_2) < \infty$ as long as z stays in the interior $\mathrm{IC}(S_1) \cap \mathrm{IC}(S_2)$, though as z gets close to the boundaries $\partial(\mathrm{conv}(S_i))$, $i = 1$ or 2, the constant C may go to $+\infty$.

PROOF. First, convexity and linearity imply that for each $\delta, \delta' > 0$ we have the following:

1. For affine functions, $L \leq L' + \delta$ on a set S implies $L \leq L' + \delta$ on $\mathrm{conv}(S)$.
2. If two affine functions L, L' satisfy $L \leq L' + \delta$ on a set S and if moreover, $L(z) \geq L'(z) - \delta'$ at some interior point z of $\mathrm{conv}(S)$, then $|L - L'| \leq C = C(z, S, \delta, \delta')$ on $\mathrm{conv}(S)$. Here, the constant $C < \infty$ depends only on δ, δ' and the ratio between the minimum distance from z to $\partial(\mathrm{conv}(S))$ and the maximum distance to $\partial(\mathrm{conv}(S))$, though, as z gets close to $\partial(\mathrm{conv}(S))$, the constant C can go to $+\infty$.

Now, apply (1) to the case $L = L_1$, $L' = L_2$ and $S = S_1$, and also to the case $L = L_2$, $L' = L_1$ and $S = S_2$. Thus, we get $|L_1 - L_2| \leq \max(\delta_1, \delta_2)$ at the point z of $\mathrm{IC}(S_1) \cap \mathrm{IC}(S_2)$. Now, apply (2), to get $|L_1 - L_2| \leq C$ on both $\mathrm{conv}(S_i)$, $i = 1, 2$, where $C = C(S_1, S_2, \delta_1, \delta_2) < \infty$. Applying (1) again, we have $|L_1 - L_2| \leq C$ on $\mathrm{conv}(S_1 \cup S_2)$, completing the proof. $\square$

From now on, we consider only the maximization case, since the minimization case is the same by replacing $c(x, y)$ with $-c(x, y)$. We now introduce the following notation.

DEFINITION 7.7. Let $G \in \mathcal{S}_{\mathrm{MT}}$ and let $H \subseteq G$ be a c-contact layer for a triplet $\{\alpha, \gamma, \beta\}$. For each $x \in X_H$, consider the affine function

$$L_x^H(y) = \gamma(x) \cdot (y - x) + \alpha(x).$$

The superscript H indicates that L_x^H arises from a c-admissible triplet for H. The fact that H is a c-contact layer for α, γ, β can be written as

$$\beta(y) - c(x, y) \geq L_x^H(y) \qquad \forall x \in X_H, y \in Y_H, \tag{7.6}$$

$$\beta(y) - c(x, y) = L_x^H(y) \qquad \forall (x, y) \in H. \tag{7.7}$$

For an affine space V, we write for $x, x' \in X_H$,

$$L_x^H \approx L_{x'}^H \qquad \text{on } V$$

if there is a bounded set S with $V = V(S)$ and a constant $M = M(c, H, S)$ depending only on H, the cost function c and the set S, such that for every choice of a c-admissible triplet $\{\alpha, \gamma, \beta\}$ making H a c-contact layer, we have

$$|L_x^H - L_{x'}^H| \leq M \qquad \text{on the set } S. \tag{7.8}$$

We say $L_x^H \approx L_{x'}^H$ at z, if we have (7.8) for $S = \{z\}$.

An immediate observation is that for $x, x', x'' \in X_H$,

$$\text{whenever } L_x^H \approx L_{x'}^H \text{ and } L_{x'}^H \approx L_{x''}^H \text{ on } V, \text{ then } L_x^H \approx L_{x''}^H \text{ on } V.$$

Also note that if $H' \subseteq H$, we necessarily have for any $x, x' \in X_{H'}$,

$$L_x^{H'} \approx L_{x'}^{H'} \qquad \text{on } V \qquad \Rightarrow \qquad L_x^H \approx L_{x'}^H \qquad \text{on } V. \tag{7.9}$$

We shall now prove the analogue of Lemma 7.4 in the case of a general cost. We shall again use Proposition 7.6, to establish a propagation of control on the affine functions L_x^H's, along an ordered chain of intersecting convex open sets. But, since c is not trivial anymore, the control on L_x^H can be done only in finite steps, since the errors (the constant C in Proposition 7.6) can accumulate.

LEMMA 7.8. *Let $\Gamma \in \mathcal{S}_{\mathrm{MT}}$ be c-finitely exposable and set $G := \Gamma \cap (C(x_0) \times \mathbb{R}^d)$ as in Lemma 7.3. Suppose $x, x' \in X_G$ (i.e., $x \sim x'$). Then there exists a finite set $H \subseteq G$ such that $x, x' \in X_H$ and $L_x^H \approx L_{x'}^H$ on $V(C(x))$.*

PROOF. First, observe that it suffices to prove the following.

CLAIM 7.9. *Suppose $x \sim x'$ and $z \in G_{x'}$. Then there exists a finite set $H \subseteq G$ such that $x, x' \in X_H$ and $L_x^H \approx L_{x'}^H$ at z.*

The lemma follows when we apply this claim to a set of finitely many (u_i, v_i)'s, $i = 1, \dots, m$, in G (so $x \sim x' \sim u_i$) with $V(C(x)) = V(\{v_i\}_{i=1}^m)$, and use (7.9).

We show this claim using induction on $n = 0, 1, 2, 3, \dots$. Our induction hypothesis is if $x \sim_n x'$ and $z \in G_{x'}$, then there exists a finite set $H \subseteq G$ such that:

(i1) $x, x' \in X_H$, $z \in Y_H$ and $Y_H \subseteq \overline{C_n(x)}$;

(i2) $L^H_x \approx L^H_{x'}$ on $V(Y_H)$;

(i3) for each finite set $F \subseteq G$ with $H \subseteq F$, and for $w \in X_F$, there is a constant $C = C(H, w)$ depending only on H and w such that

$$L^F_w \le L^F_x + C \qquad \text{on } Y_H.$$

Notice that the claim follows if we verify (i1)–(i3) for each n, since in particular, the values of L^H_x and $L^H_{x'}$ at z is estimated from (i2).

We proceed by induction, starting with $n = 0$ and assuming $x \sim_0 x'$ (i.e., $x = x'$) and $z \in G_{x'}$. Choose then $H = \{(x, z)\}$. Then, (i1) and (i2) are trivially satisfied. Moreover, (i3) holds from (7.6) and (7.7), where the constant C is estimated by the value $c(w, z) - c(x, z)$. This completes the case $n = 0$.

Suppose now the induction hypothesis holds for n. Assume $x \sim_{n+1} x'$ and $z \in G_{x'}$. Then there is a finite chain of $C_n(x_k)$, $k = 0, 1, \dots, m$, $x_0 = x$, $x_m = x'$, such that $C_n(x_k) \cap C_n(x_{k+1}) \ne \varnothing$ for each k. Recall $C_n(x) = \mathrm{IC}[\bigcup_{x' \sim_n x} G_{x'}]$. Thus for each $C_n(x_k)$, it is possible to find a finite set $J_k := \{(u^i_k, v^i_k)^{m_k}_{i=1}\} \subseteq G$ such that $x_k \sim_n u^i_k$ for all i and $\mathrm{IC}(Y_{J_k})$ is a good approximation of $C_n(x_k)$, that is, $Y_{J_k} \subseteq \overline{C_n(x_k)} \subseteq V(Y_{J_k})$ and $\mathrm{IC}(Y_{J_k}) \cap \mathrm{IC}(Y_{J_{k+1}}) \ne \varnothing$. Also, we can let $z \in Y_{J_m}$.

Now, apply the induction hypothesis for n to each x_k, u^i_k, $x_k \sim_n u^i_k$ and find a finite set $H^i_k \subseteq G$ that satisfies (i1)–(i3) for $x = x_k$, $x' = u^i_k$, $z = v^i_k$, and $H = H^i_k$. Let

$$H_k := \bigcup_i H^i_k$$

Then, from (i3) for H^i_k's, we also have (i3) for $H = H_k$ and $x = x_k$. Here, the point of considering H_k is $Y_{J_k} \subseteq Y_{H_k} \subseteq \overline{C_n(x_k)}$, so $V(Y_{J_k}) = V(Y_{H_k})$, hence $\mathrm{IC}(Y_{H_k}) \cap \mathrm{IC}(Y_{H_{k+1}}) \ne \varnothing$ as well.

In order to verify the induction hypothesis for $n + 1$th step, let

$$\bar H := \bigcup_k H_k$$

We will show properties (i1)–(i3) for this set $\bar H$. From the construction, $x, x' \in X_{\bar H}$, $z \in Y_{\bar H}$, and since $C_n(x_k) \subseteq C_{n+1}(x_k) = C_{n+1}(x)$, (i1) readily follows. For (i2), apply the induction hypothesis (i3) for H_k's to Proposition 7.6 iteratively for the pairs Y_{H_1} and Y_{H_2}, $Y_{H_1} \cup Y_{H_2}$ and $Y_{H_3}, \dots, Y_{H_1} \cup \dots \cup Y_{H_k}$ and $Y_{H_{k+1}}$, so on. Then we see the estimate (7.8) holds for $S = Y_{\bar H}$, thus

$$L^{\bar H}_x \approx L^{\bar H}_{x_1} \approx \dots \approx L^{\bar H}_{x_{m-1}} \approx L^{\bar H}_{x'} \qquad \text{on } V(Y_{\bar H}), \tag{7.10}$$

verifying (i2).

For (i3), let F be a finite set containing $\bar{H}$ and let $w \in X_F$. Then (i3) for each H_k gives that $L_w^F \le L_{x_k}^F + C_k$ on Y_{H_k}, $k = 0, 1, \dots, m$. Now applying (7.10) and recalling (7.9), we conclude that there is a constant $C = C(\bar{H}, w)$ such that

$$L_w^F \le L_x^F + C \qquad \text{on } Y_{\bar{H}}.$$

This completes the induction, and the proof. $\square$

PROOF OF LEMMA 7.3. Recall that we fix $x_0 \in X$ and let $G := \Gamma \cap (C(x_0) \times \mathbb{R}^d)$ and $V := V(Y_G)$. Let $m = dim(V)$. Then we can find

$$J := \{(u_i, v_i)\}_{i=0}^m \subseteq G \qquad \text{such that } V(\{v_i\}_{i=0}^m) = V.$$

Define the initial choices $K_{v_i} = \{0\}$, $i = 0, 1, \dots, m$. We want to define the K_y's to be compatible with these initial choices. For $y \in Y_G$, choose $x(y) \in X_G$ such that $(x(y), y) \in G$. By Lemma 7.8 (especially see Claim 7.9), for $y \in Y_G$, we can choose a finite set $H(y)$ such that $J \cup \{(x(y), y)\} \subseteq H(y)$ and

$$L_{x(y)}^{H(y)} \approx L_{u_i}^{H(y)} \qquad \text{at } v_i, \forall i = 0, \dots, m.$$

In particular, there exists a constant M, depending only on y and $H(y)$—but not on the choice of the c-admissible functions for which $H(y)$ is a c-contact layer—such that

$$\left|L_{x(y)}^{H(y)}(v_i) - L_{u_i}^{H(y)}(v_i)\right| \le M \qquad \forall i = 0, \dots, m. \tag{7.11}$$

$H(y)$ being a c-contact layer for some triplet (α, γ, β), we can by subtracting an appropriate affine function from β, assume $\beta(v_i) = 0$. This yields that

$$\beta(y) = c(x(y), y) + L_{x(y)}^{H(y)}(y).$$

Since $L_x^{H(y)}$ is affine and $V(\{v_i\}_{i=0}^m) = V$, the value $L_{x(y)}^{H(y)}(y)$ can be computed from the values $L_{x(y)}^{H(y)}(v_i)$. Hence by (7.11), the values $L_{u_i}^{H(y)}(v_i)$ give an estimate of $\beta(y)$. Notice that the c-contact property yields that

$$L_{u_i}^{H(y)}(v_i) = \beta(v_i) - c(u_i, v_i) = -c(u_i, v_i).$$

Thus, there exists a constant $N = N(y)$ such that if β is a c-admissible for $H(y)$ and if $\beta(v_i) = 0$ for all i, then $-N \le \beta(y) \le N$. We set $K_y = [-N, N]$.

To get the claim in Lemma 7.3, we let H be any finite set and denote $Y_H = \{y_1, \dots, y_s\}$. Let

$$H^* = H \cup H(y_1) \cup \dots \cup H(y_s).$$

Now choose β to be c-admissible for H^* with $\beta(v_i) = 0$ for all i. Since β is also c-admissible for $H(y_j)$, we have $\beta(y_j) \in K_{y_j}$ for all $j = 1, \dots, s$. Finally, note that β is also a c-admissible for H, concluding the proof. $\square$

PROOF OF THEOREM 7.2. As before, let $G = \Gamma \cap (C(x_0) \times \mathbb{R}^d)$ and let $V := V(Y_G)$ be the ambient space. We first find the desired function $\beta : Y_G \to \mathbb{R}$ from the compactness argument already used in [5]. Indeed, define $K := \Pi_{y \in Y_G} K_y$, where the K_y's were obtained in Lemma 7.3. This is a subset of the space of all functions from Y_G to $\mathbb{R}$. In the topology of pointwise convergence, K is compact by Tychonoff's theorem. Now, pick an arbitrary finite set $H \subseteq G$. We claim that the set

$$\Psi_H := \{\beta \in K : \beta \text{ is } c\text{-admissible for } H\}$$

is a nonempty closed subset of K. Indeed, that Ψ_H is nonempty follows from Lemma 7.3 since every finite subset of Γ, and hence of G is a c-contact layer: if necessary, one can extend the β found in Lemma 7.3—and originally defined on Y_H—to Y_G, by simply letting $\beta(y) = 0$ for $y \notin Y_H$.

To show that Ψ_H is closed, let $\{\beta_n\}$ be a sequence of c-admissible functions for H, and suppose $\beta_n \to \beta$ pointwise on Y_G. We need to show that β is also c-admissible for H. But for each n, we have functions (α_n, γ_n) such that the following relation holds:

$$\beta_n(y) - c(x, y) \geq \gamma_n(x) \cdot (y - x) + \alpha_n(x) \qquad \forall x \in X_H, y \in Y_H, \tag{7.12}$$

$$\beta_n(y) - c(x, y) = \gamma_n(x) \cdot (y - x) + \alpha_n(x) \qquad \forall (x, y) \in H. \tag{7.13}$$

Here, without loss of generality, we can assume that each vector $\gamma_n(x)$ is parallel to $V(Y_H)$. Now $(\beta_n(y) - c(x, y))_{x \in X_H, y \in Y_H}$ is uniformly bounded in n; this is because the set H is assumed to be finite, so are X_H and Y_H. So, we can choose $(\alpha_n(x), \gamma_n(x))$ in such a way that $(\alpha_n(x), \gamma_n(x))_n$ is also uniformly bounded in n. Since X_H is finite, we can find a subsequence of (α_n, γ_n) which converges to (α, γ) at every $x \in X_H$. Then (α, γ, β) is clearly a c-admissible triplet for H, establishing the claim on Ψ_H.

It is clear that the class $\{\Psi_H\}$ satisfies the finite intersection property, that is, $\emptyset \neq \Psi_{H_1 \cup \cdots \cup H_s} \subseteq \bigcap_{j=1,\dots,s} \Psi_{H_j}$. By the compactness of K and the closeness of Ψ_H's, we deduce that the set $\Psi_G := \bigcap_{H \subseteq G, |H| < \infty} \Psi_H$ is nonempty.

We now claim that any $\beta \in \Psi_G$ is c-admissible for G. Indeed, fix $x \in X_G$ and $\beta \in \Psi_G$. We must show that there exists an affine function L_x on $V = V(Y_G)$ such that the following holds:

$$\beta(y) - c(x, y) \geq L_x(y) \qquad \forall y \in Y_G, \tag{7.14}$$

$$\beta(y) - c(x, y) = L_x(y) \qquad \forall y \in G_x. \tag{7.15}$$

Choose a finite set $H_x \subseteq G_x$ such that $V(H_x) = V(G_x)$. Observe that for any finite set F containing $H := \{x\} \times H_x$,

$$L_x^F(y) = \beta(y) - c(x, y) = L_x^H(y) \qquad \forall y \in H_x, \qquad \text{hence}$$
$$L_x^F(y) = L_x^H(y) \qquad \forall y \in V(G_x).$$

In particular, $L_x^F(x) = L_x^H(x)$ since $x \in \mathrm{IC}(G_x) \subseteq V(G_x)$. Let us define $\alpha(x) = L_x^H(x)$.

Now we need to construct the last piece which is $\gamma(x)$. For this, in addition to H, we also choose a finite set $\{(v_i, w_i)\}_{i=1}^m \subseteq G$ such that $x \in \mathrm{IC}(\{w_i\}_{i=1}^m)$ and $V(\{w_i\}_{i=1}^m) = V$, and define

$$\bar{H} := H \cup \{(v_i, w_i)\}_{i=1}^m.$$

For any finite set $F \subseteq G$ with $\bar{H} \subseteq F$, define the set

$$\gamma_F(x) := \{v \in V : \beta(y) - c(x, y) - \alpha(x) \geq v \cdot (y - x), \forall y \in Y_F \tag{7.16}$$

$$\beta(y) - c(x, y) - \alpha(x) = v \cdot (y - x), \forall y \in F_x\}. \tag{7.17}$$

The set $\gamma_F(x)$ is nonempty as there is L_x^F. Now, since $x \in \mathrm{IC}(\{w_i\}_{i=1}^m)$ and $V(\{w_i\}_{i=1}^m) = V$, we deduce from (7.16) that $\gamma_F(x)$ is a closed and bounded set in V, hence compact. Again since every subset of a c-contact layer is also a c-contact layer, the class $\{\gamma_F(x) : F \supseteq \bar{H}\}$ has the finite intersection property. Hence, we can choose a

$$\gamma(x) \in \bigcap_{F \supseteq \bar{H}, |F| < \infty} \gamma_F(x).$$

Finally, we show that (7.14) and (7.15) hold for this choice of $(\alpha(x), \gamma(x))$. Indeed, let $(x', y') \in G$, and let $F = \bar{H} \cup \{(x', y')\}$. By (7.16), we have $\beta(y') - c(x, y') \geq \alpha(x) + \gamma(x) \cdot (y' - x)$, so (7.14) holds. Let $y \in G_x$. Let $F = \bar{H} \cup \{(x, y)\}$. By (7.17), we have $\beta(y) - c(x, y) = \alpha(x) + \gamma(x) \cdot (y - x)$, so (7.15) holds. This completes the proof of Theorem 2.11. $\square$

REMARK 7.10. We have shown that every component $G := \Gamma \cap (C \times \mathbb{R}^d)$ in Theorem 7.2 is c-exposable, that is, there exists a triplet (α, γ, β) such that

$$\beta(y) - \alpha(x) - \gamma(x)(y - x) \leq c(x, y) \qquad \text{for all } (x, y) \in X_G \times Y_G, \quad \text{and}$$

$$\beta(y) - \alpha(x) - \gamma(x)(y - x) = c(x, y) \qquad \text{for all } (x, y) \in G.$$

To obtain a better triplet in terms of regularity, we may apply the transforms in Definition 3.1. First, replace β with $\tilde{\beta}(y) := \inf_{x \in X_G}\{\alpha(x) + \gamma(x)(y - x) + c(x, y)\}$ and note that $\tilde{\beta}$ is upper semicontinuous provided c is continuous. Then let α_c, γ_c be as in Definition 3.1, and note that the definition can be rephrased as follows: define $H(x, y) := \mathrm{conc}[\tilde{\beta}(\cdot) - c(x, \cdot)](y)$, that is, for each $x \in C$, $y \mapsto H(x, y)$ is the upper concave envelope of $y \mapsto \tilde{\beta}(y) - c(x, y)$. Then $\alpha_c(x) = H(x, x)$, and $\gamma_c(x)$ is the superdifferential of $y \mapsto H(x, y)$ at x. As was shown, α_c is locally Lipschitz and γ_c locally bounded under the assumption in Theorem 3.2. Finally, note that G is a c-contact layer induced by $(\alpha_c, \gamma_c, \tilde{\beta})$.

8. Structural results for general optimal martingale transport plans. We start by proving Conjecture 2 in the case of a discrete target measure.

THEOREM 8.1. *Let $c(x, y) = |x - y|$, suppose $\mu \ll \mathcal{L}^d$ and that ν is discrete, that is, ν is supported on a countable set. Let $\pi \in \mathrm{MT}(\mu, \nu)$ be a solution of* (1.1), *then for μ a.e. x, $\operatorname{supp} \pi_x$ consists of $d + 1$ points which are vertices of a polytope in $\mathbb{R}^d$ and, therefore, the optimal solution is unique.*

PROOF. Since the result holds true (for more general target measures) when $d = 1$, we shall assume that $d \geq 2$. Let S be the countable support of ν and let $J := \{E \subseteq S : |E| < \infty \,\&\, \dim V(E) \leq d - 1\}$, where $|E|$ is the cardinality of the set E. Consider $V_J := \bigcup_{E \in J} V(E)$. Since $\dim V(E) \leq d - 1$ and J is countable, it follows that $\mathcal{L}^d(V_J) = 0$. Let Γ be a martingale-monotone regular concentration set for π (as in Definition 5.2). Let $X := X_\Gamma \setminus V_J$ so that $\mu(X) = 1$. Notice that if $x \in X$, then Γ_x must contain vertices of a polytope which has x in its interior. Now let

$$K := \{E \subseteq S : |E| = d + 2 \text{ and } \\ E \text{ contains vertices of a } d\text{-dimensional polytope}\}.$$

Fix $F = \{y_0, y_1, \dots, y_d, y\}$ in K, where $y_0, y_1, \dots, y_d$ are vertices of a d-dimensional polytope and consider the set $A := \{x \in X : F \subseteq \Gamma_x\}$. In other words, $A = \Gamma^{y_0} \cap \cdots \cap \Gamma^{y_d} \cap \Gamma^y$, where $\Gamma^y := \{x : (x, y) \in \Gamma\}$. We shall prove that $\mu(A) = 0$.

Indeed, suppose otherwise, that is $\mu(A) > 0$ and let x_0 be a Lebesgue point of A. Let $B = A \cap C(x_0)$ and note that $\mathcal{L}^d(B) > 0$ since $C(x_0)$ is open in $\mathbb{R}^d$. Since the set $\Gamma \cap (C(x_0) \times \mathbb{R}^d)$ is a c-contact layer, there exist constants $\lambda_0, \lambda_1, \dots, \lambda_d, \lambda$ such that for all $x \in B$, we have

$$|x - y_i| + \gamma(x) \cdot (y_i - x) + \alpha(x) = \lambda_i, \qquad i = 0, 1, \dots, d,$$
$$|x - y| + \gamma(x) \cdot (y - x) + \alpha(x) = \lambda.$$

Also note that $\{y_0, y_1, \dots, y_d, y\} \subseteq \operatorname{Ext}(\operatorname{conv}(\Gamma_x))$ for almost all $x \in B$. Let p_i be determined by $y = \sum_{i=0}^d p_i y_i$, and $\sum_{i=0}^d p_i = 1$, and note that some p_i may be negative. Then, by the above, we get that the function

$$g(x) := \sum_{i=0}^{d} p_i |x - y_i| - |x - y|$$

is constant on B, which has positive measure.

We explain why this leads to a contradiction. First, notice that because g is real analytic in $\Omega := \mathbb{R}^d \setminus \{y_0, \dots, y_d, y\}$, it is not constant in any open subset, since otherwise it is constant everywhere, which is not the case. Second, without loss of generality, assume $x_0 = 0$ and $g(0) = 0$, and notice that from the real analyticity of g, one can write $g(x) = P_k(x) + Q(x)$ for some $k \in \mathbb{N}$, where $P_k(x)$ is the first nonzero kth degree homogeneous polynomial, and $Q(x)$ is a power series of terms with degree greater than k, in particular, $Q(x) = O(|x|^{k+1})$. Now, consider the set

$$\mathcal{S} := \{u \in S^{d-1} \mid \text{ there exists } 0 \neq x_n \to 0, x_n/|x_n| \to u, \text{ with } 0 = g(x_n)\}.$$

Then, for each $u \in \mathcal{S}$, $0 = \frac{g(x_n)}{|x_n|^k} = P_k(x_n/|x_n|) + \frac{Q(x_n)}{|x_n|^k}$, showing $P_k(u) = \lim_{n\to\infty} P_k(x_n/|x_n|) = 0$. Thus, $\mathcal{S}$ is a subset of the zero set $\{u; P_k(u) = 0\}$.

Now if g is zero on the set B where x_0 is a Lebsegue point, then $\mathcal{S} = S^{d-1}$, hence $P_k = 0$, a contradiction. Hence, $\mu(A) = 0$. The countability of K now implies the theorem.

For the uniqueness, we use the usual argument, namely that the average of two optimal plans is also optimal, which contradicts the polytope-type of their respective supports. □

REMARK 8.2. As we see from the above proof, Theorem 8.1 holds true for a much more general cost $c(x, y)$ than $|x - y|$. Indeed, it is enough (but not necessary) $c(x, y)$ to be analytic in $\{x \neq y\}$, and the function $g(x) = \sum_{i=0}^{d} p_i c(x, y_i) - c(x, y)$ to be nonconstant. In particular, we can choose $c(x, y) = |x - y|^p$, with $p \neq 2$.

We now establish Conjecture 1 in the two-dimensional case.

THEOREM 8.3. *Assume $d = 2$, $c(x, y) = |x - y|$, μ is absolutely continuous with respect to the Lebesgue measure, and ν has compact support. Let $\pi \in \mathrm{MT}(\mu, \nu)$ be a solution of* (1.1), *then for μ almost every $x \in \mathbb{R}^2$*, $\operatorname{supp}\pi_x = \operatorname{Ext}(\overline{\operatorname{conv}}(\operatorname{supp}\pi_x))$.

PROOF. Let Γ be a martingale-monotone regular concentration set for π [see Lemma 5.1(4) and Definition 5.2], and let $X = X_\Gamma$. (Recall then $\operatorname{supp}\pi_x = \overline{\Gamma_x}$ for all $x \in X$.) The theorem will follow if we show that the set

$$E_\pi := \{x \in X \mid \operatorname{supp}\pi_x \subseteq \operatorname{Ext}(\overline{\operatorname{conv}}(\operatorname{supp}\pi_x))\}$$

has full μ-measure. First, note that E_π is measurable by Proposition D.1. (Here, we used the fact that each of $\operatorname{supp}\pi_x \subseteq \mathbb{R}^d$ is compact, which is satisfied since the second marginal of π is compactly supported.)

We shall show that its complement $N = X \setminus E_\pi$ has μ-measure zero and since $\mu \ll \mathcal{L}^2$ it suffices to show that $\mathcal{L}^2(N) = 0$. For that, note first that the set $X_0 := \{x \in X : \dim(\operatorname{conv}(\Gamma_x)) = 0\}$ is obviously included in E_π, which means that $N = (N \cap X_2) \cup (N \cap X_1)$, where

$$X_2 = \{x \in X : \dim V(C(x)) = 2\} \quad \text{and} \quad X_1 = \{x \in X : \dim V(C(x)) = 1\},$$

where $\{C(x); x \in X\}$ is the irreducible convex paving of Γ.

Note that $X_2 = \bigcup_{x\in X_2}(X \cap C(x)) = X \cap (\bigcup_{x\in X_2} C(x))$. Since $\bigcup_{x\in X_2} C(x)$ is open, X_2 is measurable. But since $\Gamma \cap (C(x) \times \mathbb{R}^2)$ is a c-contact layer, Theorem 2.4 yields that $\overline{\Gamma_x} = \mathrm{Ext}(\mathrm{conv}(\overline{\Gamma_x}))$ for a.e. x in $X_2 \cap C(x)$. Since X_2 can be approximated by compact sets from the inside and $\{C(x)\}_{x\in X_2}$ is an open cover of X_2, we conclude that $\overline{\Gamma_x} = \mathrm{Ext}(\mathrm{conv}(\overline{\Gamma_x}))$ for a.e. x in X_2. Hence, $\mathcal{L}^2(N \cap X_2) = 0$.

Consider now the measurable set $A_1 := N \cap X_1$, and assume that $\mathcal{L}^2(A_1) > 0$. Note that for every $x \in A_1$, we have that $I(x) := \mathrm{IC}(\operatorname{supp} \pi_x)$ is an open line segment with x in its interior. Note that $I(x) \subseteq C(x)$ and $C(x)$ is one-dimensional for every $x \in A_1$. By Proposition D.2, the function defined for each $x \in A_1$ by

$$\delta(x) = \sup\{r; (x - r, x + r) \subseteq I(x)$$

is measurable, where $(x - r, x + r)$ denotes the interval of radius r at x inside the line segment $I(x)$. Therefore, for every $\delta > 0$ the set $A_\delta := \{x \in A_1 : \delta(x) > \delta\}$ is measurable, and $\mathcal{L}^2(A_\delta) > 0$ for some $\delta > 0$. Let now x_0 be a Lebesgue point of A_δ, and consider W to be the one-dimensional affine space containing x_0 and perpendicular to $I(x_0)$. Choose $\varepsilon > 0$ much smaller than δ and let $A_{\delta,\varepsilon} := A_\delta \cap B(x_0, \varepsilon)$ [note $\mathcal{L}^2(A_{\delta,\varepsilon}) > 0$]. Then $\{C(x); x \in A_{\delta,\varepsilon}\}$ is a disjoint family of open segments that cover $A_{\delta,\varepsilon}$ and $C(x) \cap W \neq \varnothing$. Let $F : \bigcup_{x\in A_{\delta,\varepsilon}} C(x) \to \bigcup_{x\in A_{\delta,\varepsilon}} F(C(x))$ be the flattening map with respect to W as in Lemma C.1. Since F is bi-Lipschitz on the appropriate set containing $A_{\delta,\varepsilon}$, we have that $F(A_{\delta,\varepsilon})$ is measurable and $\mathcal{L}^2(F(A_{\delta,\varepsilon})) > 0$.

Note that again by Theorem 2.4, $\overline{\Gamma_z} = \mathrm{Ext}(\mathrm{conv}(\overline{\Gamma_z}))$, for $\mathcal{L}^1$ almost all z in each $A_{\delta,\varepsilon} \cap C(x)$. Since $A_{\delta,\varepsilon} \subseteq N$, this implies that $A_{\delta,\varepsilon} \cap C(x)$ is $\mathcal{L}^1$ measure zero, and so does $F(A_{\delta,\varepsilon}) \cap F(C(x))$. Now $\{F(C(x)); x \in A_{\delta,\varepsilon}\}$ is a parallel cover of $F(A_{\delta,\varepsilon})$, so by Fubini's theorem with bi-Lipschitz map F, we conclude $\mathcal{L}^2(F(A_{\delta,\varepsilon})) = 0$, which is a contradiction. [Here, for the Fubini's theorem, we used the fact that $F(A_{\varepsilon,\delta})$ is measurable.] It follows that $\mathcal{L}^2(A_1) = 0$, which then results $\mathcal{L}^2(N) = 0$. This completes the proof. $\square$

The same proof could extend to higher dimensions, provided one can prove measurability of the function

$$X_\Gamma \ni x \mapsto \delta(x) = \sup\{r \geq 0 : B(x, r) \subseteq C(x)\}$$

defined for a given convex paving $(C(x))_{x\in X_\Gamma}$ associated to Γ. One can then obtain the following.

THEOREM 8.4. *Assume $c(x, y) = |x - y|$ on $\mathbb{R}^d \times \mathbb{R}^d$ and let $\pi \in \mathrm{MT}(\mu, \nu)$ be a solution of (1.1) with a martingale-monotone regular concentration set Γ. Assume μ is absolutely continuous with respect to the Lebesgue measure and that*

$$\text{(8.1)} \qquad \textit{the function } \delta \textit{ is measurable, and } \dim\big(V\big(C(x)\big)\big) \geq d - 1 \textit{ for } \mu \textit{ a.e. } x,$$

where $(C(x))_{x\in X_\Gamma}$ is the irreducible convex paving associated to Γ. Then, for μ almost every $x \in \mathbb{R}^d$, $\operatorname{supp} \pi_x = \mathrm{Ext}(\mathrm{conv}(\operatorname{supp} \pi_x))$.

9. The disintegration of a martingale transport plan. For a closed convex set $U \subseteq \mathbb{R}^d$, let $\mathcal{K}(U)$ be the space of all closed convex subsets in $\mathbb{R}^d$, equipped with the Hausdorff metric in such a way that it becomes a separable complete metric space (Polish space). This allows for the disintegration of a measure π on X via a measurable map $T : X \to \mathcal{K}(U)$ (see, e.g., [6], Corollary 2.4) in such a way that each piece of the disintegrated measure, say π_C, is a probability measure on $T^{-1}(C)$. In particular, $\pi_C(T^{-1}(C)) = 1$ for $T_\# \pi$-a.e. $C \in \mathcal{K}(U)$, ultimately yields conditional probabilities.

Consider now a set $\Gamma \in \mathcal{S}_{\mathrm{MT}}$ and the corresponding unique irreducible convex paving $\{C; C \in \Phi\}$ as given in Theorem 2.8. Define the map

$$\Xi : \Gamma \to \mathcal{K}(\mathbb{R}^d) \qquad \text{by } (x, y) \mapsto \overline{C(x)},$$

where $\mathcal{K}(\mathbb{R}^d)$ is the space of convex closed subsets of $\mathbb{R}^d$. We conjecture that this map is measurable when $\mathcal{K}(\mathbb{R}^d)$ is equipped with the Hausdorff metric, under which $\mathcal{K}(\mathbb{R}^d)$ becomes a separable complete metric space. In this case, we shall show that a martingale transport plan π can be canonically disintegrated into its components given by $(\Gamma \cap (C(x) \times \mathbb{R}^d))_{x \in X_\Gamma}$. As usual, in the case of minimization with $c(x, y) = |x - y|$, we shall assume further that $\mu \wedge \nu = 0$.

THEOREM 9.1 (Disintegration of martingale plans). *Let (μ, ν) be probability measures on $\mathbb{R}^d$ in convex order and let $\pi \in \mathrm{MT}(\mu, \nu)$ with a concentration set $\Gamma \in \mathcal{S}_{\mathrm{MT}}$ and the associated irreducible convex paving $\{C; C \in \Phi\}$. Assume the map $\Xi : \Gamma \to \mathcal{K}(\mathbb{R}^d)$ defined by $(x, y) \mapsto \overline{C(x)}$, is measurable, and let $\tilde{\pi} = \Xi_\# \pi$ denote the push-forward of π into $\mathcal{K}(\mathbb{R}^d)$, and $I \subseteq \mathcal{K}(\mathbb{R}^d)$ is the image of Γ by Ξ. Then the following holds*:

(1) *There exists a disintegration of π along the map Ξ such that*

$$\pi(S) = \int_I \pi_C(S)\, d\tilde{\pi}(C) \qquad \textit{for each Borel set } S \subseteq \mathbb{R}^d \times \mathbb{R}^d, \tag{9.1}$$

where for $\tilde{\pi}$-a.e. C, π_C is a probability measure supported on $\Gamma_C := \Gamma \cap (C \times \mathbb{R}^d)$.

(2) *For $\tilde{\pi}$-a.e. $C \in I$, there exist probability measures μ_C, ν_C such that the couple (μ_C, ν_C) is in convex order, μ_C is supported on $X_C := X_\Gamma \cap C$, ν_C on Y_{Γ_C} and $\pi_C \in \mathrm{MT}(\mu_C, \nu_C)$.*

(3) *If π is optimal for problem (1.1) in $\mathrm{MT}(\mu, \nu)$, then for $\tilde{\pi}$-a.e. $C \in I$, π_C is optimal for the same problem on $\mathrm{MT}(\mu_C, \nu_C)$. Furthermore, Γ_C is a c-contact layer. In particular, duality is attained for π_C.*

(4) *If in addition, μ_C is absolutely continuous with respect to the Lebesgue measure on $V(C)$, and $c(x, y) = |x - y|$, then for μ_C-almost all x, $\overline{\Gamma_x} = \mathrm{Ext}(\mathrm{conv}(\overline{\Gamma_x}))$.*

PROOF. The above discussion and the measurability hypothesis of the map $\Xi : \Gamma \to \mathcal{K}(\mathbb{R}^d)$ defined by $(x, y) \mapsto \overline{C(x)}$, yield the disintegration of π into

$\pi_C d\tilde{\pi}(C)$ in (9.1), with π_C supported on Γ_C. The measures μ_C, ν_C are obtained by taking marginals of π_C. The martingale and optimality properties of π_C for $\tilde{\pi}$-a.e. C, follow from those properties of π and the disintegration (9.1). When π is an optimal martingale transport, the concentration set Γ can be chosen in such a way that it is c-finitely exposable, hence the set Γ_C is a c-contact layer by Theorem 2.11. This deals with items (1), (2) and (3) of the theorem. Finally, (4) follows immediately from Theorem 2.4. □

In order to apply this theorem and deduce global results from its local properties, one would like to know when we can disintegrate μ into absolutely continuous pieces μ_C, so as to apply Theorem 2.4 on each partition. We start by a counterexample showing that this is not possible in general, at least in dimension $d \geq 3$.

Nikodym sets and martingale transports. Ambrosio et al. [1] constructed a Nikodym set in $\mathbb{R}^3$ having full measure in the unit cube, and intersecting each element of a family of pairwise disjoint open lines only in one point. More precisely, they showed the following.

THEOREM 9.2 (Ambrosio et al. [1]). *There exist a Borel set $M_N \subseteq [-1,1]^3$ with $|[-1,1]^3 - M_N| = 0$ and a Borel map $f = (f_1, f_2) : M_N \to [-2,2]^2 \times [-2,2]^2$ such that the following holds. If we define for $x \in M_N$, the open segment l_x connecting $(f_1(x), -2)$ to $(f_2(x), 2)$, then*:

- $\{x\} = l_x \cap M_N$ *for all* $x \in M_N$,
- $l_x \cap l_y = \varnothing$ *for all* $x \neq y \in M_N$.

EXAMPLE 9.3. One can use the above construction to construct an optimal martingale transport, whose equivalence classes are singletons, hence the disintegration of the first marginal along the partitions $C(x)$ is the Dirac mass δ_x, which is obviously not absolutely continuous w.r.t. $\mathcal{L}^1$.

Consider the obvious inequality $\frac{1}{2\varepsilon}(|x-y| - \varepsilon)^2 \geq 0$, and its equivalent form

$$\frac{1}{2\varepsilon}|y|^2 \geq |x-y| + \frac{1}{\varepsilon}x \cdot (y-x) + \frac{1}{2\varepsilon}|x|^2 - \frac{\varepsilon}{2}. \tag{9.2}$$

Thus by letting $\alpha_\varepsilon(x) = \frac{1}{2\varepsilon}|x|^2 - \frac{\varepsilon}{2}$, $\beta_\varepsilon(y) = \frac{1}{2\varepsilon}|y|^2$ and $\gamma_\varepsilon(x) = \frac{1}{\varepsilon}x$, (9.2) yields that the set $\Gamma = \{(x,y); |x-y| = \varepsilon\}$ is a c-contact layer, where $c(x,y) = |x-y|$ in the maximization problem. It follows that every martingale $\pi_\varepsilon := (X, Y)$ with $|X - Y| = \varepsilon$ a.s. is optimal with its own marginals $X \sim \mu$ and $Y \sim \nu$.

Now fix $\varepsilon > 0$ small and let X be a random variable whose distribution μ has uniform density on $[-1,1]^3$. We define Y conditionally on X by evenly distributing the mass along the lines l_x considered in Theorem 9.2 and distance ε, that is Y splits equally in two pieces from $x \in X$ along l_x with distance ε. Then the martingale (X, Y) is optimal for the maximization problem. But note that in this case, each equivalence class $[x]$ is the singleton $\{x\}$, so the disintegration of μ along the partitions $C(x)$ is the Dirac mass δ_x, which is obviously not absolutely continuous w.r.t. $\mathcal{L}^1$. Hence, the decomposition is not useful in this case. One also notices that the convex sets associated to the irreducible paving of the martingale (X, Y) have codimension 2. We leave it as an open problem whether one can do without assumption (8.1) in Theorem 8.4.

REMARK 9.4. By letting $\varepsilon \to 0$, the above problem approaches the one considered in Example 5.7, that is, the case when the marginals $\mu = \nu$ are equal, the only maximal martingale transport is the identity, and the value of the maximal cost is zero. On the other hand, note that $\int \beta_\varepsilon(y)\, d\nu(y) - \int \alpha_\varepsilon(x)\, d\mu(x) = \varepsilon$, which means that $(\alpha_\varepsilon, \beta_\varepsilon, \gamma_\varepsilon)$ is a minimizing sequence for the dual problem. But neither of the sequences $\alpha_\varepsilon, \beta_\varepsilon, \gamma_\varepsilon$ converge (neither pointwise nor in L^1). This is another manifestation of the nonexistence of a dual in Example 5.7. This said, for the minimization problem, we have no example where duality is not attained.

APPENDIX A: A SUITABLE CONCENTRATION SET FOR A MARTINGALE TRANSPORT PLAN

Here, we prove the following lemma which was introduced in Section 5.

LEMMA A.1. *Let $\pi \in \mathrm{MT}(\mu, \nu)$ and let $\Lambda \subseteq \mathbb{R}^d \times \mathbb{R}^d$ be a Borel set with $\pi(\Lambda) = 1$. Then there exists a Borel set $\Gamma \subseteq \Lambda$ with $\pi(\Gamma) = 1$ such that the map $x \mapsto \pi_x$ is measurable and defined everywhere on X_Γ in such a way that*:

1. $\overline{\Gamma_x} = \operatorname{supp} \pi_x$ *for all* $x \in X_\Gamma$.
2. $\Gamma \in \mathcal{S}_{\mathrm{MT}}$, *that is* $x \in \mathrm{IC}(\Gamma_x)$ *for all* $x \in X_\Gamma$.
3. *If we assume that $\mu \ll \mathcal{L}^d$, then Γ can be chosen in such a way that $X_\Gamma \subseteq \mathrm{IC}(Y_\Gamma)$.*
4. *If, in addition, π is a solution of the optimization problem* (1.1), *then Γ can be chosen to be finitely c-exposable.*

PROOF. Let $(\pi_x)_x$ be the unique disintegration of π with respect to μ. It is well known that this yields a well-defined measurable map $x \mapsto \pi_x$ on a Borel set E in $\mathbb{R}^d$ with $\mu(E) = 1$ such that each x in E is the barycenter of π_x and $\pi_x(\Lambda_x) = 1$. It is clear that $x \in \mathrm{CC}(\Lambda_x)$. However, it is not necessarily in $\mathrm{IC}(\Lambda_x)$. Note however that for any Borel set B in $\mathbb{R}^d$, the map $x \mapsto \pi_x(B)$ is Borel measurable, hence for each $r > 0$, the set $B_r := \{(x, y) | x \in E, \pi_x(B_r(y)) > 0\}$ is Borel [here, $B_r(y)$ is the open ball with center y and radius r in $\mathbb{R}^d$], and consequently the set $\Theta := \{(x, y) | x \in E, y \in \operatorname{supp}(\pi_x)\} = \bigcap_{n=1}^\infty B_{1/n}$ is also Borel. Letting $\Gamma := \Lambda \cap \Theta$, it is clear that $\pi(\Gamma) = 1$ and $\pi_x(\Gamma_x) = 1$ for all $x \in E$. Finally, note that the probability measure π_x has its barycenter at x and that $\Gamma_x \subseteq \operatorname{supp}(\pi_x)$, and since $\pi_x(\Gamma_x) = 1$, we have that $\overline{\Gamma_x} = \operatorname{supp} \pi_x$. Hence in particular, $x \in \mathrm{IC}(\Gamma_x)$ for $x \in E$, proving (1) and (2).

Item (3) can be obtained by considering another subset of Γ. Indeed, let X' be the set of Lebesgue points of X_Γ. Then as $\mu \ll \mathcal{L}^d$, we have $\mu(X') = 1$. Let $\Gamma' := \Gamma \cap (X' \times \mathbb{R}^d)$. Then $\Gamma' \in \mathcal{S}_{\mathrm{MT}}$, $\pi(\Gamma') = 1$ and $X' \subseteq \mathrm{IC}(X') \subseteq \mathrm{IC}(Y_{\Gamma'})$, as claimed.

For item (4), we use [5, 30], where it is shown that for an optimizer π, there exists Λ with $\pi(\Lambda) = 1$, that is finitely c-exposable (also called finitely c-monotone in [5]; see Definition 2.9). We then restrict Λ to get Γ which also satisfies (1), (2) and (3) by the above procedure. $\square$

APPENDIX B: AN ESTIMATE FOR CONVEX FUNCTIONS

We prove here a technical result—used in Section 6—that allows us to control the maximum of a convex function by the integral of its second derivatives. Namely, we have the following.

PROPOSITION B.1. *Let B_r denote the closed ball of radius r centred at the origin* 0. *Let φ be a* (*smooth*) *convex function such that $\varphi(0) = 0$ and $\varphi \geq 0$. Then*

$$\int_{B_{\sqrt{2}r}} \Delta\varphi \geq C_0 r^{d-2} \max_{B_r} \varphi, \tag{B.1}$$

where the constant $C_0 > 0$ depends only on the dimension d.

PROOF. Denote $M_r = \max_{B_r} \varphi$. By the maximum principle a point, say $p \in \partial B_r$, can be chosen from the boundary so that $\varphi(p) = M_r$. Choose an orthonormal basis $\eta_1, \dots, \eta_d$ such that $p = r\eta_1$, and define a cylindrical set (of radius $r/2$)

$$K_r := \left\{ \sum_{j=1}^{d} t_j \eta_j \,\Big|\, -r \leq t_1 \leq r, \sqrt{\sum_{j \neq 1,} t_j^2} \leq r/2 \right\}.$$

We will show that

$$\int_{K_r} D_{11}^2 \varphi \geq C_0 r^{d-2} \max_{B_r} \varphi \tag{B.2}$$

for a constant $C_0 > 0$ depending only on the dimension d. This will immediately imply the desired estimate (B.1) because $K_r \subseteq B_{\sqrt{2}r}$ and $0 \leq D_{11}^2 \varphi \leq \Delta\varphi$ for the convex function φ.

To show (B.2), we let H denote the hyperplane $\{z_1 = 0\}$. Notice that $\varphi(0) = 0$ and $\varphi \leq M_r$ on B_r, thus from convexity of φ, we see that

$$\varphi(z) \leq \frac{|z|}{r}\big(M_r - \varphi(0)\big) = \frac{|z|}{r} M_r \qquad \text{for each } z \in B_r. \tag{B.3}$$

Also, notice the fact that $p = r\eta_1$ is a maximum point of φ in B_r and that the hyperplane $r\eta_1 + H$ stays outside the interior of B_r, that is, $r\eta_1 + H \subseteq \mathbb{R}^d \setminus (\operatorname{int} B_r)$. So, from convexity of φ, we have

$$\varphi(z + r\eta_1) \geq M_r \qquad \text{for each } z \in H.$$

Combining this with (B.3) and using convexity of φ again, we can estimate the derivative $D_1\varphi$ on the set $r\eta_1 + (H \cap B_r)$. Namely, for each $z \in H \cap B_r$,

$$\begin{aligned} D_1\varphi(z + r\eta_1) &\geq \frac{1}{r}\big(\varphi(z + r\eta_1) - \varphi(z)\big) \\ &\geq \frac{1}{r}\left(M_r - \frac{|z|}{r} M_r\right) \\ &= \frac{1}{r^2}\big(r - |z|\big) M_r. \end{aligned}$$

Similarly, use (B.3), the fact that $\varphi \geq 0$, in particular on $-r\eta_1 + H$, and the convexity of φ to see that

$$D_1\varphi(z - r\eta_1) \leq \frac{1}{r}\big(\varphi(z) - \varphi(z - r\eta_1)\big) \leq \frac{|z|}{r^2} M_r, \qquad \text{for each } z \in H \cap B_r.$$

From these estimates on $D_1\varphi$, we have that, for each $z \in H \cap B_r$,

$$\begin{aligned} \int_{-r}^{r} D_{11}^2 \varphi(z + t\eta_1)\, dt &= D_1\varphi(z + r\eta_1) - D_1\varphi(z - r\eta_1) \\ &\geq \frac{1}{r^2}\big(r - |z|\big) M_r - \frac{|z|}{r^2} M_r \\ &= \frac{1}{r^2}\big(r - 2|z|\big) M_r. \end{aligned}$$

Now,

$$\begin{aligned} \int_{K_r} D_{11}^2 \varphi\, dz &= \int_{z \in H \cap B_{r/2}} \int_{-r}^{r} D_{11}^2 \varphi(z + t\eta_1)\, dt\, dz \\ &\geq \int_{z \in H \cap B_{r/2}} \frac{1}{r^2}\big(r - 2|z|\big) M_r\, dz \\ &= C_0 r^{d-2} M_r, \end{aligned}$$

where

$$C_0 = r^{2-d} \int_{H \cap B_{r/2}} \frac{1}{r^2}\big(r - 2|z|\big)\, dz = \int_{H \cap B_{1/2}} \big(1 - 2|z|\big)\, dz$$

is independent of r. Notice that $C_0 > 0$ because $|z|$ varies from 0 to 1/2 on $H \cap B_{1/2}$. This completes the proof. $\square$

APPENDIX C: A BI-LIPSCHITZ FLATTENING MAP

The following lemma, which describes a bi-Lipschitz "flattening map," was used in Section 8.

LEMMA C.1. *Let $\mathbb{R}^d = V \times W$, where $V = \mathbb{R}^{d-1}$ and $W = \mathbb{R}$. Let $\delta > 0$ and let A be a subset of W. Suppose that for each $h \in A$, there is a set D_h which is contained in a hyperplane H_h with $H_h \cap W = \{0, \dots, 0, h\}$. Suppose further that $\{D_h\}_{h\in A}$ are mutually disjoint and the projection of every $\{D_h\}$ on V contains the ball B_R with center 0 and radius R in V. Finally, suppose that the angle between H_h and W is bounded; there is $\eta < \pi/2$ such that the normal direction of H_h and the direction of W has angle less than η for every $h \in A$.*

Now define the flattening map $F : \bigcup_h D_h \to F(\bigcup_h D_h)$ as follows: for $x = (v, w) \in D_h$, $F(v, w) = (v, h)$. Then F is bi-Lipschitz on the set $N := (\bigcup_h D_h) \cap (B_r \times W)$, where $r < R$.

PROOF. First, note that by the disjointness of $\{D_h\}$ the map F is bijective, so F^{-1} is well defined. The lemma is intuitively clear; the map F cannot move two nearby points too far away, because the hyperdiscs $\{D_h\}$ are disjoint.

First of all, from the bounded angle assumption, F is clearly bi-Lipschitz on each $F(D_h)$ with the same Lipschitz constant for all $h \in A$. Hence, for $x_1 = (v_1, w_1)$, $x_2 = (v_2, w_2)$, we will assume that x_1, x_2 are contained in D_{h_1}, D_{h_2}, respectively, and $h_1 \neq h_2$.

We consider the case $v_1 = v_2 \in V$ and $|v_1| = |v_2| \le r$. Let L be the one-dimensional subspace of V containing 0 and v_1. Regarding D_{h_1}, D_{h_2} as affine functions on V, since their graphs on $L \cap B_R$ are disjoint and linear and $r < R$, it is clear that $|w_1 - w_2| \approx |h_1 - h_2|$, that is,

$$C_1|h_1 - h_2| \le |w_1 - w_2| \le C_2|h_1 - h_2| \qquad \text{for some } C_1, C_2 > 0. \tag{C.1}$$

Next, we consider the case $v_1 \neq v_2$. We want to show $|x_1 - x_2| \approx |F(x_1) - F(x_2)|$, or equivalently,

$$|w_1 - w_2| \approx |h_1 - h_2|.$$

Let L be the one-dimensional affine subspace of V containing v_1 and v_2. Regarding D_{h_1}, D_{h_2} as affine functions on V, since their graphs on $L \cap B_R$ are disjoint and linear, it is clear that

$$\begin{aligned}|w_1 - w_2| &= \left|D_{h_1}(v_1) - D_{h_2}(v_2)\right| \\ &\le \max\left(\left|D_{h_1}(v_1) - D_{h_2}(v_1)\right|, \left|D_{h_1}(v_2) - D_{h_2}(v_2)\right|\right).\end{aligned}$$

But by (C.1), we have

$$\max\left(\left|D_{h_1}(v_1) - D_{h_2}(v_1)\right|, \left|D_{h_1}(v_2) - D_{h_2}(v_2)\right|\right) \le C_2|h_1 - h_2|,$$

which shows that F^{-1} is Lipschitz on $F(N)$. On the other hand, by (C.1), we have

$$|h_1 - h_2| \leq (1/C_1) \min\big(\big|D_{h_1}(v_1) - D_{h_2}(v_1)\big|, \big|D_{h_1}(v_2) - D_{h_2}(v_2)\big|\big)$$

and, again regarding D_{h_1}, D_{h_2} as disjoint linear graphs on $L \cap B_R$, we have

$$\begin{aligned} \min\big(\big|D_{h_1}(v_1) - D_{h_2}(v_1)\big|, \big|D_{h_1}(v_2) - D_{h_2}(v_2)\big|\big) &\leq \big|D_{h_1}(v_1) - D_{h_2}(v_2)\big| \\ &= |w_1 - w_2| \end{aligned}$$

which shows that F is Lipschitz on N, and the proof is complete. □

APPENDIX D: PROOFS OF MEASURABILITY

We now establish the following proposition which was used in the proofs of Section 8.

PROPOSITION D.1. *Let π be a Borel measure on the product space $\mathbb{R}^d \times \mathbb{R}^d$ and let $A \subseteq \mathbb{R}^d$ be a concentration set for its first marginal. Let $x \mapsto \pi_x$ be the corresponding disintegration map from A to $P(\mathbb{R}^d)$ and assume that for each $x \in A$, the set $\operatorname{supp} \pi_x \subseteq \mathbb{R}^d$ is compact—which is satisfied in particular, if the second marginal of π is compactly supported. Then the set*

$$E_\pi := \big\{x \in A \mid \operatorname{supp} \pi_x \subseteq \operatorname{Ext}\big(\overline{\operatorname{conv}}(\operatorname{supp} \pi_x)\big)\big\}$$

is a Borel measurable set in $\mathbb{R}^d$.

PROOF. Let $N_\pi = A \setminus E_\pi$, that is,

$$N_\pi = \big\{x \in A \mid \operatorname{supp} \pi_x \nsubseteq \operatorname{Ext}\big(\overline{\operatorname{conv}}(\operatorname{supp} \pi_x)\big)\big\}.$$

We will show that there is a measurable set N in $\mathbb{R}^d$ such that $N_\pi \subseteq N$ and $E_\pi \cap N = \varnothing$, which then implies that the set $E_\pi = A \setminus N$ is measurable, as desired.

We shall use a classical result of Carathéodory, which implies that a point $z \in \operatorname{supp} \pi_x$ is not an extremal point of the convex hull of $\operatorname{supp} \pi_x$ if and only if it lies in the relative *interior* of an r-simplex ($1 \leq r \leq d$) with vertices in $\operatorname{supp} \pi_x$ (see, e.g., [12]). First, choose a countable dense subset $Q \subseteq \mathbb{R}^d$ and associate to each $q \in Q$ an (ε, δ)-admissible r-simplex $S \subseteq \mathbb{R}^d$, defined as follows:

1. all the vertices of S belong to Q,
2. q is ε-close to a (relative) interior point of S, and
3. all vertices of S are δ-away from q.

Let $\mathcal{A}_{\varepsilon,\delta}(q)$ denote the countable set of all (ε, δ)-admissible simplices for q. Now define the set

$$\begin{aligned} S_{\varepsilon,\delta}(q) := \big\{x \in A \mid \pi_x\big(B_\varepsilon(q)\big) > 0 \ \&\ \text{there exists } S \in \mathcal{A}_{\varepsilon,\delta}(q) \text{ such that} \\ \text{for each vertex } q_j \text{ of } S, \pi_x\big(B_\varepsilon(q_j)\big) > 0\big\}. \end{aligned}$$

This set $S_{\varepsilon,\delta}(q)$ contains all those points x in A, such that $\operatorname{supp}\pi_x$ include, up to an ε-error, both the point q and the vertices of an (ε,δ)-admissible simplex for q. Since the map $x \mapsto \pi_x \in P(\mathbb{R}^d)$ is measurable, each set $S_{\varepsilon,\delta}(q)$ is measurable, since it can be written as the countable union of measurable sets. Define $N_{\varepsilon,\delta} := \bigcup_{q\in Q} S_{\varepsilon,\delta}(q)$, and set

$$N = \bigcup_{k\geq 1}\bigcap_{j\geq 1} N_{2^{-j-k},2^{-k}}.$$

It is clear that N is measurable. We now show that it has the desired properties.

CLAIM 1. *$N_\pi \subseteq N$. Indeed, for any $x \in N_\pi$, there exists a $z \in \operatorname{supp}\pi_x$ lying in the relative interior of an r-simplex, say S, with vertices in $\operatorname{supp}\pi_x$. Let $\delta_0 > 0$ be a lower bound for the distances from z to the vertices of S as well as the distances between any two vertices. Fix $k \in \mathbb{N}$ large enough so that $\delta_0 \geq \delta = 2^{-k+1}$. Since Q is dense, one can find for each $\varepsilon = 2^{-j-k}$, $j \geq 1$, a point $q \in B_\varepsilon(z)$ and an (ε,δ)-admissible simplex S_j for q whose vertices are ε close to the vertices of S. This implies that for each $j \in \mathbb{N}$, $x \in S_{\varepsilon,\delta}(q)$ where $\varepsilon = 2^{-j-k}$ and $\delta = 2^{-k+1}$. This shows that $x \in \bigcap_{j\geq 1} N_{2^{-j-k},2^{-k}} \subseteq N$ as desired.*

CLAIM 2. *$E_\pi \cap N = \varnothing$. Indeed, suppose not then there exists $x \in E_\pi \cap \bigcap_{j\geq 1} N_{2^{-j-k},2^{-k}}$ for some $k \in \mathbb{N}$. Let $\delta = 2^{-k}$ and $\varepsilon_j = 2^{-j-k}$ for each $j \geq 1$. Then we see that for each $j \geq 1$, there exists $q_j \in Q$ and a simplex, say S_j, that is (ε_j,δ)-admissible for q_j such that q_j and the vertices of S_j are ε_j close to $\operatorname{supp}\pi_x$. Since $\operatorname{supp}\pi_x$ is compact by assumption, there exists a convergent subsequence of $\{q_j\}$, as well as a convergent subsequence of the simplices $\{S_j\}$ (in the Hausdorff topology since their vertices converge). Let q_∞, S_∞ denote their limits (as $j \to \infty$), respectively. Note that $q_\infty \in \operatorname{supp}\pi_x$ and that S_∞ is a simplex with vertices in $\operatorname{supp}\pi_x$. By the definition of (ε_j,δ)-admissibility, we also have that q_∞ belongs to the closure of S_∞, while being δ-away from its vertices. This implies that the point $q_\infty \in \operatorname{supp}\pi_x$ is not an extremal point of the convex hull of $\operatorname{supp}\pi_x$. This contradicts the fact that $x \in E_\pi$, thus completing the proof of Claim 2 and the proposition.*

□

Next, we show the following.

PROPOSITION D.2. *Let π be a Borel measure on the product space $\mathbb{R}^d \times \mathbb{R}^d$ and let $x \mapsto \pi_x \in P(\mathbb{R}^d)$ be its disintegration along the first marginal. Let $A \subseteq \mathbb{R}^d$ be a Borel measurable set that is a concentration set for the first marginal of π, and denote for each $x \in A$, the set $I(x) = \mathrm{IC}(\operatorname{supp}\pi_x)$.*

Assume that $I(x)$ is bounded and that $x \in I(x)$ for each $x \in A$. If $A_1 \subseteq A$ is a measurable set such that for each $x \in A_1$, $\dim I(x) = 1$, then the function $w : A_1 \to \mathbb{R}_+$ defined by

$$w(x) := \min\bigl[\operatorname{dist}(x, y_0), \operatorname{dist}(x, y_1)\bigr],$$

where y_0, y_1 are the end points of the segment $I(x)$, is Borel measurable.

PROOF. It is enough to show that for each $\delta > 0$, the set $M_\delta = \{x \in A_1 \mid w(x) \geq \delta\}$ is Borel measurable. For that, we again consider a countable dense subset $Q \subseteq \mathbb{R}^d$. For $\varepsilon, \delta > 0$, we say that a closed segment $S = [p_0, p_1]$ connecting two points $p_0, p_1 \in \mathbb{R}^d$ is (ε, δ)-admissible for $q \in Q$ if:

1. $p_0, p_1 \in Q$;
2. $q \in N_\varepsilon(S)$, the latter being the ε-tubular neighborhood of S;
3. $\mathrm{dist}(p_i, q) \geq \delta$, for $i = 0, 1$.

Let $\mathcal{A}_{\varepsilon,\delta}(q)$ denote the countable set of (ε, δ)-admissible segments for q, and define the set

$$S_{\varepsilon,\delta}(q) := \{x \in A_1 \mid \mathrm{dist}(x, q) \leq \varepsilon \text{ and there exists } [p_0, p_1] \in \mathcal{A}_{\varepsilon,\delta}(q) \\ \text{with } \pi_x(B_\varepsilon(p_i)) > 0, i = 0, 1\}.$$

The set $S_{\varepsilon,\delta}(q)$ contains those points x in A_1, such that x is ε-close to q, while $\mathrm{supp}\,\pi_x$ includes up to ε, the end points of an (ε, δ)-admissible segment for q. Again, each set $S_{\varepsilon,\delta}(q)$ is measurable, since the map $x \mapsto \pi_x \in P(\mathbb{R}^d)$ is measurable. Define the set $M_{\varepsilon,\delta} := \bigcup_{q \in Q} S_{\varepsilon,\delta}(q)$, and set

$$\bar{M}_\delta = \bigcap_{j \geq 1} M_{2^{-j},\delta}.$$

It is obvious that $\bar{M}_\delta$ is measurable. We claim that

$$M_\delta = \bar{M}_\delta. \tag{D.1}$$

Indeed, we first verify that $\bar{M}_\delta \subseteq M_\delta$. To see this, consider an arbitrary point $x \in \bar{M}_\delta$, and let y_0, y_1 be the two end points of the segment $I(x)$. Then for each $0 < \varepsilon < \delta$, there is $q \in Q$ and $S = [p_0, p_1] \in \mathcal{A}_{\varepsilon/3,\delta}(q)$ such that $x \in B_{\varepsilon/2}(q)$ and $\pi_x(B_{\varepsilon/2}(p_i)) > 0$ for $i = 0, 1$. From the last condition, we see that $p_0, p_1 \in N_{\varepsilon/2}(\mathrm{supp}\,\pi_x)$, and hence $S \in N_{\varepsilon/2}(I(x))$. Moreover, from the item (3) for the (ε, δ)-admissibility of S together with $x \in B_{\varepsilon/2}(q)$, we see that $\mathrm{dist}(p_i, x) \geq \delta - \varepsilon/2$, which then implies that $\mathrm{dist}(x, y_i) \geq \delta - \varepsilon$. Since $\varepsilon > 0$ was arbitrary, this implies that $x \in M_\delta$ as desired.

For the reverse inclusion $M_\delta \subseteq \bar{M}_\delta$, note that for each $x \in M_\delta$, we have $\mathrm{dist}(y_i, x) \geq \delta$, $i = 0, 1$, where y_0, y_1 are the end points of the segment $I(x)$. Also, notice that $y_i \in \mathrm{supp}\,\pi_x$, $i = 0, 1$. Since $Q \subseteq \mathbb{R}^d$ is dense, one can find for each $0 < \varepsilon < \delta$, a point $q \in Q$ and a segment $S = [p_0, p_1] \in A_{\varepsilon,\delta}(q)$ such that $q \in B_\varepsilon(x)$, and $p_i \in B_\varepsilon(y_i))$, $i = 0, 1$. It follows that $x \in S_{\varepsilon,\delta}(q)$ which implies $x \in M_{\varepsilon,\delta}$ for all $0 < \varepsilon < \delta$, thus $x \in \bar{M}_\delta$. This completes the proof. □

Acknowledgments. Part of this research has been done while the authors were visiting the Fields Institute in Toronto for the thematic program on "Calculus of Variations" in Fall 2014. We are thankful for the hospitality and the great research environment that the institute provided.

REFERENCES

[1] AMBROSIO, L., KIRCHHEIM, B. and PRATELLI, A. (2004). Existence of optimal transport maps for crystalline norms. *Duke Math. J.* **125** 207–241. MR2096672

[2] BEIGLBÖCK, M., COX, A. M. G. and HUESMANN, M. (2017). Optimal transport and Skorokhod embedding. *Invent. Math.* **208** 327–400. MR3639595

[3] BEIGLBÖCK, M., HENRY-LABORDÈRE, P. and PENKNER, F. (2013). Model-independent bounds for option prices—a mass transport approach. *Finance Stoch.* **17** 477–501. MR3066985

[4] BEIGLBÖCK, M., HENRY-LABORDÈRE, P. and TOUZI, N. (2017). Monotone martingale transport plans and Skorokhod embedding. *Stochastic Process. Appl.* **127** 3005–3013. MR3682121

[5] BEIGLBÖCK, M. and JUILLET, N. (2016). On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44** 42–106. MR3456332

[6] BIANCHINI, S. and CAVALLETTI, F. (2013). The Monge problem for distance cost in geodesic spaces. *Comm. Math. Phys.* **318** 615–673. MR3027581

[7] BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. MR1100809

[8] CHOQUET, G. (1959). Forme abstraite du téorème de capacitabilité. *Ann. Inst. Fourier* (*Grenoble*) **9** 83–89. MR0112844

[9] DE MARCH, H. and TOUZI, N. (2017). Irreducible convex paving for decomposition of multidimensional martingale transport plans. Arxiv preprint.

[10] DOLINSKY, Y. and SONER, H. M. (2014). Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields* **160** 391–427. MR3256817

[11] DOLINSKY, Y. and SONER, H. M. (2015). Martingale optimal transport in the Skorokhod space. *Stochastic Process. Appl.* **125** 3893–3931. MR3373308

[12] ECKHOFF, J. (1993). Helly, Radon, and Carathéodory type theorems. In *Handbook of Convex Geometry, Vol. A, B* 389–448. North-Holland, Amsterdam. MR1242986

[13] FALKNER, N. (1981). The distribution of Brownian motion in $\mathbf{R}^n$ at a natural stopping time. *Adv. Math.* **40** 97–127. MR0619096

[14] GALICHON, A., HENRY-LABORDÈRE, P. and TOUZI, N. (2014). A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.* **24** 312–336. MR3161649

[15] GANGBO, W. and MCCANN, R. J. (1996). The geometry of optimal transportation. *Acta Math.* **177** 113–161. MR1440931

[16] GHOUSSOUB, N., KIM, Y.-H. and LIM, T. (2017). Optimal Brownian Stopping between radially symmetric marginals in general dimensions. Arxiv preprint.

[17] HOBSON, D. (2011). The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance* 2010. *Lecture Notes in Math.* **2003** 267–318. Springer, Berlin. MR2762363

[18] HOBSON, D. and KLIMMEK, M. (2015). Robust price bounds for the forward starting straddle. *Finance Stoch.* **19** 189–214. MR3292129

[19] HOBSON, D. and NEUBERGER, A. (2012). Robust bounds for forward start options. *Math. Finance* **22** 31–56. MR2881879

[20] KOMLÓS, J. (1967). A generalization of a problem of Steinhaus. *Acta Math. Acad. Sci. Hung.* **18** 217–229. MR0210177

[21] LIM, T. (2016). Optimal martingale transport between radially symmetric marginals in general dimensions. Arxiv preprint.

[22] MONGE, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de L'académie Royale des Sciences de Paris*.

[23] OBŁÓJ, J. (2004). The Skorokhod embedding problem and its offspring. *Probab. Surv.* **1** 321–392. MR2068476

[24] OBŁÓJ, J. and SIORPAES, P. (2017). Structure of martingale transports in finite dimensions. Arxiv preprints.

[25] ROCKAFELLAR, R. T. (1966). Characterization of the subdifferentials of convex functions. *Pacific J. Math*. **17** 497–510. MR0193549

[26] RÜSCHENDORF, L. and RACHEV, S. T. (1990). A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal*. **32** 48–54. MR1035606

[27] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Stat*. **36** 423–439. MR0177430

[28] VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483

[29] VILLANI, C. (2009). *Optimal Transport*. *Old and New*. *Grundlehren der Mathematischen Wissenschaften* **338**. Springer, Berlin. MR2459454

[30] ZAEV, D. A. (2015). On the Monge–Kantorovich problem with additional linear constraints. *Mathematical Notes* **98** 725–741. MR3438523

N. GHOUSSOUB
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER V6T 1Z2
CANADA
E-MAIL: nassif@math.ubc.ca

Y.-H. KIM
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER V6T 1Z2
CANADA
AND
CENTER FOR MATHEMATICAL CHALLENGES
KOREA INSTITUTE FOR ADVANCED STUDY
85 HOEGIRO, DONGDAEMUN-GU
SEOUL
REPUBLIC OF KOREA
E-MAIL: yhkim@math.ubc.ca

T. LIM
INSTITUTE OF MATHEMATICAL SCIENCES
SHANGHAITECH UNIVERSITY
393 MIDDLE HUAXIA ROAD, PUDONG
SHANGHAI 201210
CHINA
E-MAIL: tlim@shanghaitech.edu.cn