Optimal transport and 1-Lipschitz maps

Krzysztof Jan Ciosmak
St John's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2020

Acknowledgements

I would like to express my sincere thanks to my supervisor, Professor Zhongmin Qian, who has bestowed me with trust and freedom and allowed myself to pursue my research interests. I thank also the reviewers of this dissertation – Professor Karl-Theodor Sturm and Professor Gui-Qiang- Chen – for their helpful suggestions and comments.

I would like to express my gratitude towards Professor Bo'az Klartag, who has greatly influenced this thesis, for proposing a research topic, for introduction to the problem, for many stimulating discussions. Moreover, I thank him for his hospitality and invitation to participate in the programme Geometric Functional Analysis and Applications that took place at the Mathematical Sciences Research Institute.

I would like to thank Professor Fabio Cavalletti for the invitation to visit SISSA, discussions and his kind hospitality.

I would like to thank Professor Eva Kopecká and Dr Vojtěch Kaluža for discussions concerning the matter of Chapter 4.

I would like to thank organisers of the Thematic Programme: Optimal Transport held in Erwin Schrödinger International Institute for the invitation and great working environment provided in the Institute.

I would like to thank organisers of the Thematic Semester on Calculus of Variations and Probability held in CIMI for their hospitality, generous support and for the opportunity to attend the winter school.

I would like to thank all the faculty at the University of Oxford who have influenced this thesis. In particular let me thank Professor Jan Kristensen, Professor Andrea Mondino and Professor Harald Oberhauser for reviewing my work as part of transfer and confirmation of status procedures. Let me thank also Professor Charles Batty, Professor Jan Obłój and Dr Pietro Siorpaes.

Let me express my thanks to Professor Michał Wojciechowski, who was supervisor of my master's thesis and bachelor's thesis at the University of Warsaw. His enthusiasm for mathematics and perpetual willingness to discuss any research topic has resulted in great broadening of my mathematical horizons.

Let me thank all friends and colleagues I have made among fellow students and with whom I shared my time as a doctoral student in Oxford.

I wish to express my thanks to the Clarendon Fund, St John's College and Engineering and Physical Sciences Research Council for their generous support during my studies.

Last but not least, I would like to express my deep thanks to my parents Anna and Grzegorz for their constant support and encouragement. I would like to express special thanks to my brother Paweł for many mathematical discussions we had and for stimulating atmosphere during my undergraduate studies which significantly affected my mathematical development.

Let me also thank all other people from whom I benefited and are not explicitly listed here.

Abstract

The main topic of this work is concerned with optimal transport approach to the localisation technique. The localisation technique allows to reduce a high-dimensional problem to a collection of one-dimensional problems. Significance of this approach is illustrated by its applications to Poincaré inequality, log-Sobolev inequality, isoperimetric inequality in the context of metric measure spaces satisfying the curvature-dimension condition. We investigate multi-dimensional generalisation of the technique and related conjectures of Klartag [70, Chapter 6]. The settlement of these conjectures would constitute a step towards a proof of the waist inequality in the setting of metric measure spaces satisfying the curvature-dimension condition. This, in turn, would help in answering the Bourgain's hyperplane conjecture and the isoperimetric conjecture of Kannan, Lovász and Simonovits.

We provide a partial affirmative answer to the conjecture that, whenever Euclidean space equipped with a measure satisfies the curvature-dimension condition, then any vector-valued 1-Lipschitz map induces a partition of the domain such that the leaves equipped with the conditional measures satisfy the curvature-dimension condition with the same parameters as the initial space did. Our approach is based on the ideas of Sudakov and their further instances in works of Caffarelli, Feldman, McCann and of Klartag.

We provide a counterexample to another conjecture of Klartag that the conditional measures, with respect to a partition induced by a certain 1-Lipschitz map, satisfy so-called mass balance condition. We develop a theory of optimal transport of vector measures and employ it to provide a sufficient condition for the mass balance condition to hold true.

One way of proving the mass balance condition in one-dimensional setting is via approximation of Lipshitz functions. This leads us to a study of continuity properties of extensions of vector-valued Lipschitz maps. We provide a strengthening of the classical Kirszbraun theorem, which shows a discrepancy

between one-dimensional and multi-dimensional cases. We identify a sharp rate of continuity of extensions of 1-Lipschitz maps.

Recently, several variants of the optimal transport problem have been studied. Martingale optimal transport is a one of the greatest importance. Interestingly, also in this case a localisation result has been established. Therefore, with help of a novel version of Strassen's theorem, we investigate the duality theorems in this problem. We provide a reformulation of the optimal transport in two-marginal, in multi-marginal and in martingale setting. We reprove the Kantorovich duality formulae. As an application we characterise the uniformly smooth and the uniformly convex functions.

We develop a divergence formulation of optimal transport of vector measures and use it to provide a novel proof of the representation formula for polar cone to monotone maps. We find several generalisations of this result. A tool that we use is the matrix Hölder's inequality, for which we characterise the equality cases.

Contents

Chapter 1

Introduction

1.1 Optimal transport

In 1781 Gaspard Monge (see [85]) asked the following question: given two distributions of masses, how to transfer one distribution onto the other in an optimal way. The criterion of optimality was to minimise the average transported distance. Since then the topic has developed extensively and much of this development has been done recently. We refer the reader to the books of Villani (see [103, 104]) and to the lecture notes of Ambrosio (see [5]) for a thorough discussion, history and applications of the problem.

Let X be a set and let d be a metric on X. Let μ, ν be two Borel probability measures on X. We consider all Borel measurable mappings $T\colon X \to X$ that push μ forward to ν, i.e.,

$$\mu(T^{-1}(A)) = \nu(A) \text{ for all Borel sets } A \subset X.$$

These maps are referred to as *transport plans* of μ onto ν. Among the set of all transport plans we would like to identify one that minimises

$$\int_X d(x, T(x)) d\mu(x).$$

The modern mathematical treatment of the problem has been initiated in 1942 by Kantorovich in [64, 65]. He proposed to consider a relaxed problem of optimising

$$\int_{X \times X} d(x, y) d\pi(x, y)$$

among all *transference plans* π between μ and ν. These are the Borel probability measures on $X \times X$ such that

$$\pi(A \times X) = \mu(A) \text{ and } \pi(X \times A) = \nu(A) \text{ for all Borel sets } A \subset X.$$

We shall write $\Pi(\mu, \nu)$ for the set of transference plans. The existence of an optimal transference plan is a straightforward consequence of the Prokhorov's theorem, provided that X is separable.

The main question that has attracted a lot of attention is whether there exists an optimal transport plan. If we knew that an optimal transference plan is concentrated on a graph of a Borel measurable function then we could infer the existence of an optimal transport plan. The first complete answer on Euclidean space, under regularity assumptions on the considered measures, was presented in a seminal paper [46] of Evans and Gangbo. However, before that, Sudakov in [100] presented a solution of the problem that contained a flaw. The flaw has been remedied by Ambrosio in [5] and later by Trudinger and Wang in [101] for the Euclidean distance and by Cafarelli, Feldman and McCann in [32] for distances induced by norms that satisfy certain smoothness and convexity assumptions. In [33] Caravenna has carried out the original strategy of Sudakov for general strictly convex norms and eventually Bianchini and Daneri in [24] accomplished the plan of a proof of Sudakov for general norms on finite-dimensional normed spaces.

Let us describe briefly the strategy of Sudakov in the context of Euclidean spaces. We assume that the two Borel probability measures μ, ν on $\mathbb{R}^n$ are absolutely continuous with respect to the Lebesgue measure.

Let us recall the paramount Kantorovich–Rubinstein duality formula

$$\sup\left\{\int_{\mathbb{R}^n} ud(\mu-\nu) \mid u \text{ is 1-Lipschitz}\right\} = \inf\left\{\int_{\mathbb{R}^n\times\mathbb{R}^n} \|x-y\| d\pi(x,y) \mid \pi\in\Pi(\mu,\nu)\right\}. \tag{1.1.1}$$

Let us take an optimal u and an optimal π in the two above optimisation problems. We may infer that

$$u(x)-u(y) = \|x-y\| \text{ for } \pi\text{-almost every } (x,y)\in X\times X.$$

Consider the maximal sets on which u is an isometry, called the *transport rays*. We see that all transport has to occur on these sets. Careful analysis of the Lipschitz function u shows that the transport rays form a foliation of the underlying space $\mathbb{R}^n$, up to Lebesgue measure zero. It turns out that the direction of the transport rays is itself locally Lipschitz. This allows us to use of the area formula, which yields that the conditional measures of the disintegration of the Lebesgue measure with respect to the aforementioned foliation are absolutely continuous with respect to the one-dimensional Hausdorff measures on the transport rays. This is exactly the place where Sudakov's proof in [100] contained a defect. He claimed that any foliation into segments is such that the conditional measures are absolutely continuous with respect to the one-dimensional Hausdorff measure. It was later shown by Ambrosio, Kirchheim and Preiss (see [7]) that there exists a foliation consisting of segments and an atomic distributions on each segment such that the averaged measure is absolutely continuous with respect to the Lebesgue measure, refuting the claim of Sudakov.

Knowing that the conditional measures are absolutely continuous with respect to the one-dimensional Hausdorff measure, we may apply the well understood one-dimensional theory, where an optimal transport plan is known to exist and may be given by a certain formula, provided that at least one of the measures is non-atomic. Then the optimal transport plan on the whole space is defined separately on each transport ray.

The ideas of Sudakov have been applied also to other settings than normed spaces. The strategy has been carried out also in the context of Riemannian manifolds by Feldman and McCann in [52].

1.2 Localisation technique

In [70] Klartag has observed that the methods of optimal transport may be applied to adapt the *localisation technique* from convex geometry to the setting of Riemannian manifolds. The technique allows to reduce certain high dimensional problems to a collection of one-dimensional problems. Let us include a brief description of the technique based on [70].

It first appeared in works of Payne and Weinberger [89] and was developed in the context of convex geometry by Gromov and Milman [60], Lovász and Simonovits [78] and by Kannan, Lovász and Simonovits [63]. Later, Klartag [70] adapted the technique to the setting of weighted Riemannian manifolds satisfying the curvature-dimension condition in the sense of Bakry and Émery [11, 12]. Subsequently, Ohta [88] generalised these results to Finsler manifolds and Cavalletti and Mondino [35, 36] generalised them to metric measure spaces satisfying the synthetic curvature-dimension condition. The latter was introduced in the foundational papers by Sturm [98, 99] and by Lott and Villani [77] and allowed for development of a far-reaching, vast theory of metric measure spaces. The curvature-dimension condition may be thought of as lower bound on the curvature and an upper bound on the dimension of the considered space. We refer also to Ambrosio [6] for a recent account on the spaces satisfying the curvature-dimension condition. Let us note that the curvature-dimension condition is also related to Bochner's inequality; see [44].

The technique developed by Payne and Weinberger has no clear analogue for an abstract Riemannian manifold. This is the point where the optimal transport plays its rôle in localisation. Let us cite below a theorem from [70], presented there in a general setting of Riemannian manifolds and measures that satisfy the curvature-dimension condition. We refer the reader to Chapter 2, Section 2.7, for the necessary definitions.

Theorem 1.2.1. *Let $n \geq 2$, $\kappa \in \mathbb{R}$ and $N \in (-\infty, 1) \cup [n, \infty]$. Assume that $(\mathcal{M}, d, \mu)$ is a geodesically convex, n-dimensional weighted Riemannian manifold satisfying the curvature-*

dimension condition $CD(\kappa, N)$ w. Let $g\colon \mathcal{M} \to \mathbb{R}$ be a μ-integrable function such that

$$\int_{\mathcal{M}} g d\mu = 0 \text{ and } \int_{\mathcal{M}} |g(x)| d(x, x_0) d\mu(x) < \infty \text{ for some } x_0 \in \mathcal{M}.$$

Then there exists a partition Ω of $\mathcal{M}$ into pairwise disjoint sets, a measure ν on Ω and a family $(\mu_I)_{I\in\Omega}$ of measures on $\mathbb{R}^n$ such that:

i) for any Lebesgue measurable set $A \subset \mathcal{M}$ the map $I \mapsto \mu_I(A)$ is well-defined ν-almost everywhere, is ν-measureable and

$$\mu(A) = \int_{\Omega} \mu_I(A) d\nu(I),$$

ii) for ν-almost every $I \in \Omega$ the set $I \subset \mathcal{M}$ is a minimising geodesic and μ_I is supported on I and is a $CD(\kappa, N)$-needle or else it is a singleton,

iii) for ν-almost every $I \in \Omega$ we have $\int_I g d\mu_I = 0$.

Let us remark that the above theorem has been known before in the context of Euclidean spaces. Let us note that the proof presented in [70] differs much from the previously known proofs of Gromov [60] or of Lovász and Simonovits [78].

One of the purposes of this thesis is to continue along the line of this research and investigate multi-dimensional analogue of the localisation technique, as proposed in [70, Chapter 6].

In what follows, we consider finite-dimensional linear spaces equipped with Euclidean norm, unless specified otherwise, and 1-Lipschitz map $u\colon \mathbb{R}^n \to \mathbb{R}^m$. In Chapter 2 we define a partition, up to Lebesgue measure zero, of $\mathbb{R}^n$, associated to such a map and prove its basic properties. The sets of the partition are the maximal sets $\mathcal{S}$ such that the restriction of u to $\mathcal{S}$ is an isometry, i.e. preserves the Euclidean distance. Each such set we shall call a *leaf* of u. We prove that each leaf of u is closed and convex, hence it has a well-defined dimension.

In Theorem 2.6.1 we show that we may decompose the Lebesgue measure on $\mathbb{R}^n$ into a mixture of measures supported on leaves of u. In particular, the same is true for any measure μ such that $(\mathbb{R}^n, \|\cdot\|, \mu)$ satisfies the $CD(\kappa, N)$ condition, as any such measure is absolutely continuous with respect to the Lebesgue measure. It is a step towards a conjecture of Klartag [70, Chapter 6].

Suppose now that $m \leq n$ and that $(\mathbb{R}^n, \|\cdot\|, \mu)$ is a weighted Riemannian manifold, satisfying the curvature-dimension condition $CD(\kappa, N)$ for some $\kappa \in \mathbb{R}$ and $N \in (-\infty, 1) \cup [n, \infty]$; see Chapter 2 for definitions. Here $\|\cdot\|$ denotes the Euclidean metric on $\mathbb{R}^n$ and μ is a Borel finite measure on $\mathbb{R}^n$. A partial affirmative answer to the conjecture is provided by

Theorem 2.7.2, where we prove that, for the leaves $\mathcal{S}$ of u of dimension m, the conditional measures $\mu_{\mathcal{S}}$ are supported on the relative interiors $\mathrm{int}\mathcal{S}$ and are such that $(\mathrm{int}\mathcal{S}, \|\cdot\|, \mu_{\mathcal{S}})$ satisfies $CD(\kappa, N)$.

Note that in [70, Chapter 6], it is conjectured that also the above theorem holds true also for leaves of u of arbitrary dimension.

The possible applications of the result are in the localisation or dimensional reduction arguments, where the disintegration is an effective tool. A similar result to ours in case $m = 1$ has been used to derive new proofs and generalisations of isoperimetric inequality, Poincaré's inequality and others to the setting of metric measure spaces satisfying curvature bounds. We refer the reader to [70], [36], [35], [88].

The proof relies on the area formula and Fubini's theorem and is based on a work of Caffarelli, Feldman and McCann [32] and of Klartag [70]. See also [5] and [52] for similar approach to the Monge–Kantorovich problem.

Another tool that we use is the Wijsman topology [106] on the closed subsets of $\mathbb{R}^n$ which makes it a Polish space, so we may apply disintegration theorem. This is inspired by papers of Obłój and Siorpares [87].

1.2.1 Waist inequalities

The waist inequality, proved by Gromov [58, 59], states that if $f\colon \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$ is a continuous function, then there exists $t \in \mathbb{R}^m$ such that the fibre $L = f^{-1}(t)$ satisfies

$$\gamma_n(L + rB_n) \geq \gamma_m(rB_m) \text{ for all } r > 0.$$

Here γ_n and γ_m are the n and m dimensional standard Gaussian measures, B_n and B_m are the unit balls in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. This inequality may be seen as a generalisation of the Gaussian isoperimetric inequality. Gromov [59] has provided a proof of this inequality with use of the localisation method [89, 78, 63, 58] combined with Borsuk–Ulam type theorem. Later, Klartag [69] has proved the theorem for the unit cube also with use of localisation methods, confirming a conjecture of Guth [61].

One of the future possible applications of the research initiated in this dissertation is to prove a general version of the inequality for spaces satisfying the curvature-dimension condition. This would imply the version for convex bodies and, in turn, would help answering the Bourgain's hyperplane conjecture [29] and the isoperimetric conjecture of Kannan, Lovász and Simonovits [63].

1.2.2 Multi-bubble conjectures

The Gaussian multi-bubble conjecture is a generalisation of isoperimetric inequality that states that among all decompositions of $\mathbb{R}^n$ into $2 \leq k \leq n+1$ sets of prescribed Gaussian measure the minimal Gaussian-weighted perimeter is uniquely attained by the Voronoi cells of k equidistant points. The conjecture has been recently confirmed by E. Milman and Neeman [83] (c.f. [82]). Another possible application of the multi-dimensional localisation is a generalisation of this inequality for spaces satisfying the curvature-dimension condition.

1.3 Optimal transport of vector measures

Suppose that we are given a Borel probability measure μ on $\mathbb{R}^n$ absolutely continuous with respect to the Lebesgue measure such that

$$\int_{\mathbb{R}^n} f d\mu = 0$$

for some integrable function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\int_{\mathbb{R}^n} \|f(x)\| \|x\| d\mu(x) < \infty.$$

Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map such that

$$\int_{\mathbb{R}^n} \langle u, f \rangle d\mu = \sup \left\{ \int_{\mathbb{R}^n} \langle v, f \rangle d\mu \mid v\colon \mathbb{R}^n \to \mathbb{R}^m \text{ is 1-Lipschitz} \right\}. \tag{1.3.1}$$

In [70, Chapter 6] it is conjectured that for almost every leaf $\mathcal{S}$ of u the conditional measure $\mu_{\mathcal{S}}$ (see Theorem 2.6.1) there is

$$\int_{\mathbb{R}^n} f d\mu_{\mathcal{S}} = 0 \text{ for } \nu\text{-almost every } \mathcal{S} \in CC(\mathbb{R}^n). \tag{1.3.2}$$

We provide a counterexample to this conjecture for $m > 1$. Moreover we show that such statement fails to be true even if we replace the set of 1-Lipschitz maps in (1.3.1) by any locally uniformly closed subset of 1-Lipschitz maps with respect to *any norm* on $\mathbb{R}^n$ and *any strictly convex norm* on $\mathbb{R}^m$, unless the set of maps is trivial. Note that the outline of a proof of the conjecture suggested in [70] has a gap, as follows by the results of Chapter 4.

If $m = 1$ then (1.3.1) is precisely the dual problem to the optimal transport problem for measures ρ_1, ρ_2 given by formulae $d\rho_1 = f_+ d\mu$ and $d\rho_2 = f_- d\mu$. As we see, the dual problem, depends merely on the difference of measures, and therefore, it makes sense to consider the optimal transport for signed measures with total mass zero.

In Chapter 3 we develop a theory of optimal transport with metric cost of vector measures with total mass zero and establish its basic properties. We show, among others, that for a given vector measure, there may be no optimal transport. However, if an optimal transport for measure $d\nu = f d\mu$ with absolutely continuous first marginal of its total variation exists, then we prove that the conjecture of Klartag holds true, i.e., (1.3.2) holds true.

The precise formulation of the optimal transport for vector measures that we deal with is as follows:

$$\inf\Big\{ \int_{\mathbb{R}^n\times\mathbb{R}^n} \|x-y\| d\|\pi\|(x,y) \mid \mathrm{P}_1\pi - \mathrm{P}_2\pi = \mu \Big\}. \tag{1.3.3}$$

Here $\mathrm{P}_1\pi$ and $\mathrm{P}_2\pi$ stand for the first and the second marginal of the vector measure π respectively. The above problem for $m = 1$ simplifies to the original optimal transport problem, as follows readily by the Kantorovich–Rubinstein formula. We prove that for $m > 1$ an analogue of this formula holds with the left-hand side of (1.1.1) replaced by

$$\sup\Big\{ \int_{\mathbb{R}^n} \langle u, d\mu\rangle \mid u\colon \mathbb{R}^n \to \mathbb{R}^m \text{ is 1-Lipschitz} \Big\} \tag{1.3.4}$$

and the right-hand side replaced by (1.3.3).

Let us mention the existence of another approach to optimal transport of vector measures that differs from ours developed by Chen, Georgiou, Tannenbaum, Tyu, Li, Osher, Haber, Yamamoto (see [38, 39, 93]).

1.4 Continuity of extensions of Lipschitz maps

Another approach for a proof of (1.3.2) in the case $m = 1$ present in the literature (see e.g. [32, 70, 101, 88]) relies on a clever approximation of the 1-Lipschitz function that maximises (1.3.4). What is employed in the proof is in fact Proposition 4.3.1. It says that a 1-Lipschitz function may be extended to the entire domain in such a way that its Lipschitz constant is preserved and so is its uniform distance to another 1-Lipschitz function. This leads us to the question whether similar approach works in the multi-dimensional case. That is, we study approximations of 1-Lipschitz maps with values in $\mathbb{R}^m$, in hope that analogous results convey to this setting. However, this fails to be true. The results of our studies are included in Chapter 4, including a counterexample and several cases of positive examples, when such approximation is possible. In particular, we prove that the rate of continuity of approximation of a 1-Lipschitz function presented in [72] is sharp.

A classical theorem in this field is due to Kirszbraun [68], who proved in 1934 that any 1-Lipschitz map on $X \subset \mathbb{R}^n$ may be extended to a 1-Lipschitz map on $\mathbb{R}^n$. Here we assume that these spaces are equipped with Euclidean norms.

Theorem 1.4.1. *Let X be any subset of $\mathbb{R}^n$. Let $u\colon X \to \mathbb{R}^m$ be a 1-Lipschitz map. Then there exists a 1-Lipschitz map $\tilde{u}\colon \mathbb{R}^n \to \mathbb{R}^m$ such that $\tilde{u}|_X = u$.*

There are many proofs of this theorem and we refer the reader to [68, 94, 102] for proofs that use the Kuratowski–Zorn lemma and to [2, 30, 16] for constructive approach. There exists also an explicit formula for the extension (see [9]). Let us also note a proof that uses Fenchel duality and Fitzpatrick functions (see [91, 15]). We refer the reader also to [40] where various extensions properties of vector-valued maps are studied. In [1] another notion of contractive maps is studied. In [79] it is shown that an extension theorem holds for these contractive maps on Hilbert spaces.

Let us note that Kirszbraun's theorem holds not only in Euclidean spaces, but also for spaces with an upper or lower bound on the curvature in the sense of Alexandrov [74].

We mention also related work of Sheffield and Smart [95] on optimal Lipschitz extensions and work of Le Gruyer [75], Le Gruyer and Phan [76] on minimal Lipschitz extensions. The latter work is based on $C^{1,1}$ extensions of 1-jets with optimal Lipschitz constants of the gradients. Another related problem is the Whitney problem [105] of extending functions to $C^{1,1}$ or $C^{m,1}$ functions on $\mathbb{R}^n$. It is a topic of extensive research; see works of Fefferman [48, 49, 51] and of Brudnyi and Shvartsman [31].

Let us consider the space $\mathcal{L}(X, \mathbb{R}^m)$, equipped with the supremum norm, of all Lipschitz maps $u\colon X \to \mathbb{R}^m$ that have a finite Lipschitz constant $L(u)$, i.e. such that

$$L(u) = \sup\Big\{ \frac{\|u(x) - u(y)\|}{\|x - y\|} \mid x, y \in X \text{ and } x \neq y \Big\} < \infty.$$

In [72, 71, 73] it is proved that there exists a continuous map

$$F\colon \mathcal{L}(X, \mathbb{R}^m) \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$$

such that for any $u \in \mathcal{L}(X, \mathbb{R}^m)$ we have

$$F(u)|_X = u \text{ and } L(F(u)) = L(u).$$

In each of the mentioned papers the problem is considered in a slightly different setting. In [73] it is shown that F may be chosen in such a way that for each $u \in \mathcal{L}(X, \mathbb{R}^m)$ the image of $F(u)$ is contained in the closure of the convex hull of the image of u. Let us mention here a paper [50] of Fefferman that addresses a similar problem in the context of C^m extensions.

In Chapter 4 we study the rate of continuity of extensions of 1-Lipschitz maps. We study the following problem. Suppose we are given two sets $A \subset B \subset \mathbb{R}^n$ and 1-Lipschitz maps $u\colon A \to \mathbb{R}^m$ and $v\colon B \to \mathbb{R}^m$, with $m > 1$. We are interested in

$$\inf \Big\{ \sup \big\{ \|\tilde{u}(x) - v(x)\| \mid x \in B \big\} \mid \tilde{u}\colon B \to \mathbb{R}^m \text{ is 1-Lipschitz extension of } u \Big\}. \tag{1.4.1}$$

We show that for any u, v this quantity is bounded from above by

$$\sqrt{\delta^2 + 2\delta d_v(A, B)},$$

where

$$d_v(A,B) = \sup\{\|v(x) - v(y)\| \mid x \in A, y \in B\}, \tag{1.4.2}$$

and

$$\delta = \sup\big\{\|v(x) - u(x)\| \mid x \in A\big\}. \tag{1.4.3}$$

Moreover, it is sharp, in the sense that for any $\delta > 0$ there exist sets $A \subset B \subset \mathbb{R}^n$ and functions u, v (see Example 4.2.2), such that (1.4.3) holds true and such that for any 1-Lipschitz extension $\tilde{u}$ of u to B we have

$$\sup\big\{\|v(x) - \tilde{u}(x)\| \mid x \in B\big\} = \sqrt{\delta^2 + 2\delta d_v(A,B)}.$$

Proposition 4.2.3 shows that the rate of square root of δ is optimal. Proposition 4.2.4 shows that if $d_v(A,B)$ is infinite, then it may happen that (1.4.1) is infinite as well.

Let Y be a Hilbert space. We discuss several cases where it is possible to find an extension of a 1-Lipschitz map $u \colon A \to Y$ to a 1-Lipschitz map $\tilde{u} \colon X \to Y$ such that

$$\sup\Big\{\|v(x) - \tilde{u}(x)\| \mid x \in X\Big\} = \sup\Big\{\|v(x) - u(x)\| \mid x \in A\Big\}, \tag{1.4.4}$$

where $v \colon X \to Y$ is a given 1-Lipschitz map. The first such situation is when X, Y are Euclidean spaces and $u(x) - v(x)$ belongs to a fixed one-dimensional subspace $\mathbb{R}w$ of Y for all $x \in A$. Then the sufficient condition is that $\langle u, w\rangle$ is 1-Lipschitz with respect to a Riemannian pseudo-metric associated with v, which is given by the bilinear form

$$g_v^w(x)(s,t) = \langle s, t\rangle - \langle Dv(x)s, Dv(x)t\rangle + \langle w, Dv(x)s\rangle\langle w, Dv(x)t\rangle.$$

This condition is always satisfied when the set A is geodesically convex with respect to the pseudo-metric, i.e. that for any $x, y \in A$ there is a path realising the distance between x and y and lying in the set A.

The second situation, covers the case of maps $v \colon X \to Y$ on an arbitrary set X taking values in a Hilbert space Y and $u \colon A \to Y$, with $A \subset X$, such that the increments of v majorise the increments of u, i.e.

$$\|u(x) - u(y)\| \leq \|v(x) - v(y)\| \text{ for all } x, y \in A.$$

In this case we prove that u may be extended to X such that its increments are still majorised by the increments of v and such that

$$\sup\Big\{\|v(x) - \tilde{u}(x)\| \mid x \in X\Big\} = \sup\Big\{\|v(x) - u(x)\| \mid x \in A\Big\}.$$

In particular, if v is an isometry on X, then we partially recover the result of Section 4.3.4.

The last part considers a situation when X is a Hilbert space and v is an affine map. We prove in Theorem 4.3.9 that if Y is a Hilbert space of dimension at least two, then v is affine and 1-Lipschitz if and only if for any $u\colon A \to Y$ there is a 1-Lipschitz extension $\tilde{u}\colon X \to Y$ such that (1.4.4) holds true. One implication of this equivalence establishes a strengthening of Kirszbraun's theorem. For the proof we use the technique of K-functions developed in [84]. This shows a striking difference with the one-dimensional case, when every 1-Lipschitz map v has the above property, as Lipschitz functions are closed under minima and maxima; see Proposition 4.3.1.

In Theorem 4.3.11, we provide a partial generalisation of this theorem to extensions of Lipschitz maps to an arbirary subset of a Hilbert space.

Theorem 4.3.9 shows that an argument outlined in [70, Chapter 6], which aim was to prove (1.3.2), contains a gap. It is proven in Chapter 3 that in fact the conjecture is false if $m > 1$. Another argument, which could be used in a proof of the conjecture, could rely on one-dimensional perturbations of v. Rate of continuity of extensions of such perturbations is also a topic of Chapter 4, Section 4.3.2.

1.5 Optimal transport and Choquet theory

In Chapter 5, we explore a link of optimal transport problem with Choquet theory; see e.g. [90, 3].

Using tools of Choquet theory and a novel variant of Strassen's theorem (see Theorem 5.2.1) we reprove Kantorovich and Kantorovich–Rubinstein dualities; see Theorem 5.3.1, Theorem 5.4.1 and Theorem 5.5.4. The observation is as follows. Let $c\colon X \times Y \to \mathbb{R}$ be a cost function. Suppose u and v are functions on X and Y respectively such that for all $x \in X$ and $y \in Y$ there is $u(x) + v(y) \leq c(x, y)$. Consider the set $\mathcal{P}$ of pairs of probability measures (μ, ν) on X and Y respectively such that u, v maximise the sum of integrals

$$\int_X u d\mu + \int_Y v d\nu.$$

Strassen's theorem allows us to show that the set of extreme points of $\mathcal{P}$ is equal to the set of pairs (δ_x, δ_y) with $x \in X$ and $y \in Y$ are such that $u(x) + v(y) = c(x, y)$. Choquet's theorem and identification of extreme points of $\mathcal{P}$ yields existence of a Borel probability measure π on $X \times Y$ such that

$$(\mu, \nu) = \int_{X\times Y} (\delta_x, \delta_y) d\pi(x, y)$$

and π is supported on the set of points (x, y) such that $u(x) + v(y) = c(x, y)$. It follows that the marginals of π are μ and ν respectively and that

$$\int_X u d\mu + \int_Y v d\nu = \int_{X\times Y} c d\pi.$$

Therefore the existence of maximisers u, v implies Kantorovich duality. Note that when the cost function c is Lipschitz such existence may be proven with help of so-called c-convexification. If c is lower semicontinuous, then it may be suitably approximated by Lipschitz functions in such a way that the duality follows.

Similar reasoning may be applied as well in the context of multimarginal optimal transport; see Theorem 5.5.4.

Recently, great attention has been paid to the problem of martingale optimal transport in multi-dimensional setting. The initial interest in this problem stems from its applications to mathematical finance (see [53, 19]), and its link to the Skorokhod embedding problem [20, 86]. The link of this topic to the thesis is visible by a localisation-type results established in papers by Ghoussoub, Kim, Lim [57], by De March, Touzi [41, 42, 43] and by Obłój and Siorpaes [87]. The irreducible convex paving of these papers plays a rôle of localisation for measures in convex order.

Suppose we are given two probability measures μ, ν on $\mathbb{R}^n$ with finite first moments that are in convex order, that is for any convex function $f\colon \mathbb{R}^n \to \mathbb{R}$

$$\int_{\mathbb{R}^n} f d\mu \leq \int_{\mathbb{R}^n} f d\nu.$$

Then a theorem of Strassen (see [97]), implies that there exists a coupling π on $\mathbb{R}^n \times \mathbb{R}^n$ with respective marginals μ and ν, such that if (X, Y) is distributed according to π, then $\mathbb{E}(Y|X) = X$, i.e. the pair (X, Y) is a one-step martingale. The problem is to find such coupling π that minimises the integral

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} c d\pi$$

for a given measurable cost function $c\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. The dual problem is to find maximal value of

$$\int_{\mathbb{R}^n} u d\mu - \int_{\mathbb{R}^n} v d\nu$$

among all pairs $u, v\colon \mathbb{R}^n \to \mathbb{R}$ of continuous functions such that there exists $\gamma\colon \mathbb{R}^n \to \mathbb{R}^n$ satisfying

$$u(x) - v(y) + \langle \gamma(x), y - x \rangle \leq c(x, y) \text{ for all } x, y \in \mathbb{R}^n.$$

We prove (see Theorem 5.7.3 and Corollary 5.7.4) that the set of such pairs of functions is equal to the set of pairs $u, v\colon \mathbb{R}^n \to \mathbb{R}$ such that for all $x_1, \dots, x_{n+1} \in \mathbb{R}^n$ and all non-negative $t_1, \dots, t_{n+1}$ that sum up to one there is

$$u\Big(\sum_{i=1}^{n+1} t_i x_i\Big) - \sum_{i=1}^{n+1} t_i v(x_i) \leq \sum_{j=1}^{n+1} t_j c\Big(\sum_{i=1}^{n+1} t_i x_i, x_j\Big).$$

These results complement standard knowledge about convex functions, which follows by taking $u = v$ and c to be equal to zero. The proofs work also in case of general convex sets $K \subset \mathbb{R}^n$.

In the course of the proof of these facts we also characterise the extreme points of the set of probability measures in convex order as the set of pairs of the form

$$\Big(\delta_x, \sum_{i=1}^{d+1} t_i \delta_{x_i}\Big) \text{ with } x = \sum_{i=1}^{d+1} t_i x_i$$

for some positive $t_1, \dots, t_{d+1}$ summing up to one, some $x_1, \dots, x_{d+1} \in \mathbb{R}^n$ in general position and $d \leq n$. This is the assertion of Theorem 5.6.1.

We introduce notion of *martingale triangle inequality* (see Definition 5.9.1) and prove that if the cost function c satisfies this inequality and vanishes on the diagonal, then the value of the dual problem will not be changed if we restrict ourselves to functions u, v that satisfy $u = v$; see Theorem 5.9.3. Martingale triangle inequality demands that for any points $x, x_1, \dots, x_{n+1} \in \mathbb{R}^n$ and any non-negative $t_1, \dots, t_{n+1}$ that sum up to one there is

$$\sum_{i=1}^{n+1} t_i c(x, x_i) - c\Big(x, \sum_{i=1}^{n+1} t_i x_i\Big) \leq \sum_{i=1}^{n+1} t_i c\Big(\sum_{j=1}^{n+1} t_j x_j, x_i\Big).$$

Examples of functions that satisfy martingale triangle inequality are metrics, non-negative functions concave in the second variable and any conical combination of such functions.

Possible future applications of these finding lie in investigation of cyclical monotonicity principle for martingale optimal transport (see e.g. [21]), akin to characterisation of the classical optimal transport of Gangbo, McCann [56]. Another direction where the current developments may be useful is concerned with numerical methods for the martingale optimal transport problem.

As an application (see Theorem 5.8.3) we provide a characterisation of uniformly convex and uniformly smooth functions that complements results of Azè and Penot [10] and of Zălinescu [108].

1.6 Matrix Hölder's inequality, divergence formulation of optimal transport of vector measures with applications to representations of polar cones

Chapter 6 presents matrix Hölder's inequality with a proof taken from [14]. We characterise the equality cases for this inequality. According to the knowledge of the author, this characterisation is new. The main purpose of the chapter is to provide necessary tools to be applied in Chapter 7.

Chapter 7 concerns another formulation of optimal transport for vector measures. In there we consider the following problem

$$\inf\Big\{\|M\|_1 \mid M \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^{n\times m}), -divM = \mu\Big\}$$

and show that this value is equal to

$$\sup\Big\{\int \langle f, d\mu\rangle \mid f\colon \mathbb{R}^n \to \mathbb{R}^m, \|Df\| \leq 1\Big\}.$$

Here M is a matrix valued measure and $\|M\|_1$ stands for the total variation of M with respect to the Schatten 1-norm. In the chapter we include also other variants for vector measures that are absolutely continuous with respect to the Lebesgue measure. Let us note that these results are related to the works of Bouchitté, Buttazzo and Seppecher [28], Bouchitté and Buttazzo [27], Bouchitté, Gangbo and Seppecher [55] and Gangbo [54]. As an outcome of our study we prove the result of [37] and its generalisation with use of results of Chapter 6. That is, we provide a representation formula for vector measures that belong to the polar cone to the set of monotone maps, and generalise these methods to obtain representation formulae for polar cones to tangent cones of the unit ball of the space of differentiable maps and of the Sobolev spaces.

1.7 Outline of the thesis

In Chapter 2 we study the 1-Lipschitz maps $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and the partitions of $\mathbb{R}^n$ induced by them. The chapter contains original results of the author which generalise facts already known for $m = 1$ and partially prove a conjecture of Klartag.

In Chapter 3 we develop a theory of optimal transport for vector measures and we prove a duality formula analogous to that of Kantorovich and Rubinstein. We provide a counterexample to another conjecture of Klartag.

In Chapter 4 we study approximations of Lipschitz maps. We provide an example that establishes a sharp rate of continuity of such approximations. We study also several cases where such approximations have good continuity properties. In particular, we prove a strengthening of the Kirszbraun theorem.

In Chapter 5 we provide a new interpretation of optimal transport as an optimisation over the set of all Choquet's representations of pair of measures. We provide a novel proof of Kantorovich duality formula in two-marginal, multi-marginal setting and in martingale optimal transport setting. We introduce and study cost functions that satisfy martingale triangle inequality. As an application, we obtain a novel characterisation of uniformly smooth and of uniformly convex functions.

In Chapter 6 we consider and provide a proof of matrix Hölder inequality. Basing on this existing proof, we characterise equality cases in this inequality.

In Chapter 7 we develop theory of optimal transport for vector measures and apply it, together with results of the previous chapter, to reprove the characterisation of the dual cone of monotone maps that was first proven in [37].

All the matter presented in the thesis is due to the author, except when clearly indicated or commonly known.

Chapter 2

Leaves decompositions in Euclidean spaces

2.1 Introduction

Here we describe the structure of the chapter. In Section 2.2 we provide a careful definition of the partition associated to any 1-Lipschitz map $u\colon \mathbb{R}^n \to \mathbb{R}^m$ between Euclidean spaces. What will follow in the latter sections is the existence of the map $\mathcal{S}\colon \mathbb{R}^n \to CC(\mathbb{R}^n)$ satisfying the properties of Theorem 2.7.2. The map $\mathcal{S}$ assigns to λ-almost every point $x \in \mathbb{R}^n$ a maximal closed convex set $\mathcal{S}(x)$ containing x such that $u|_{\mathcal{S}(x)}$ is an isometry. Any such set is called a leaf of u. We prove that certain components of u are differentiable on certain leaves. Moreover we investigate the regularity of the derivative on the leaves and provide an interesting strengthening of 1-Lipschitz property of u; see Lemma 2.2.6 and Remark 2.2.7.

In Section 2.3 we define, and prove the existence, of ghost subspaces. It turns out that it may happen that the derivative of u is an isometry on a subspace that is strictly larger than the tangent space to a given leaf. These subspaces are termed ghost subspaces. This section will not be employed in further investigations of this chapter. Yet, it is of interest on its own and complements the description of the partitioning obtained in Section 2.2.

In Section 2.4 we define a Lipschitz change of variables on sets, called *clusters* of leaves, that will allow us to use the area formula and then Fubini's theorem to prove the regularity properties of the conditional measures. Here we provide significantly simpler proofs than analogous proofs in [32], mainly thanks to Lemma 2.2.6 and Corollary 2.2.10.

In Section 2.5 we prove measurability properties of the partition, which will allow us, among others, to show that the map $\mathcal{S}\colon \mathbb{R}^n \to CC(\mathbb{R}^n)$ is measurable with respect to the Wijsman topology on $CC(\mathbb{R}^n)$. We also prove that the union of boundaries of leaves of maximal dimension is a Borel set of the Lebsegue measure zero.

In Section 2.6 we provide a proof of Theorem 2.6.1, that the partition induces a disintegration of the Lebesgue measure.

In Section 2.7 we prove that the conditional measures $\mu_{\mathcal{S}}$ have densities such that the weighted Riemannian manifolds $(\mathrm{int}\mathcal{S}, d, \mu_{\mathcal{S}})$ satisfy the curvature-dimension condition, provided that $(\mathbb{R}^n, d, \mu)$ did. This partially resolves in the affirmative a conjecture of Klartag [70, Chapter 6].

In this chapter any considered norm on $\mathbb{R}^n$ and on $\mathbb{R}^m$ is Euclidean.

2.2 Partition and its regularity

If $A \subset \mathbb{R}^n$ let us denote by $\mathrm{Conv}A$ the *convex hull* of A, i.e.

$$\mathrm{Conv}A = \Big\{ \sum_{i=1}^{k} \lambda_i x_i \mid k \in \mathbb{N}, \lambda_1, \dots, \lambda_k \geq 0, \sum_{i=1}^{k} \lambda_i = 1, x_1, \dots, x_k \in A \Big\}.$$

We define the *affine hull* $\mathrm{Aff}A$ of a set $A \subset \mathbb{R}^n$ to be

$$\mathrm{Aff}A = \Big\{ \sum_{i=1}^{k} \lambda_i x_i \mid k \in \mathbb{N}, \lambda_1, \dots, \lambda_k \in \mathbb{R}, \sum_{i=1}^{k} \lambda_i = 1, x_1, \dots, x_k \in A \Big\}.$$

Lemma 2.2.1. *Let $z_1, \dots, z_k \in \mathbb{R}^n$. Let $x, y \in \mathbb{R}^n$. Suppose that*

$$\|x - z_i\| \leq \|y - z_i\|,$$

for $i = 1, \dots, k$. Then for all $z \in \mathrm{Conv}\{z_1, \dots .z_k\}$ there is

$$\|x - z\| \leq \|y - z\|.$$

In particular, if $y \in \mathrm{Conv}\{z_1, \dots .z_k\}$, then $x = y$.

Proof. Let

$$Z = \mathrm{Conv}\{z_1, \dots, z_k\}.$$

We have

$$\|x\|^2 + \|z_i\|^2 - 2\langle x, z_i \rangle \leq \|y\|^2 + \|z_i\|^2 - 2\langle y, z_i \rangle$$

for all $i = 1, \dots, k$. Hence, for these i's, we have

$$\|x\|^2 - 2\langle x, z_i \rangle \leq \|y\|^2 - 2\langle y, z_i \rangle$$

Thus, adding up these inequalities multiplied by non-negative coefficients that sum up to one, we get

$$\|x\|^2 - 2\langle x, z \rangle \leq \|y\|^2 - 2\langle y, z \rangle$$

for all $z \in Z$. Hence also

$$\|x - z\| \leq \|y - z\|.$$

Putting $z = y$ yields $\|x - y\| = 0$. ☐

Let $A \subset \mathbb{R}^n$. We shall say that a map $v\colon A \to \mathbb{R}^m$ is an isometry provided that for all $x, y \in A$ there is $\|v(x) - v(y)\| = \|x - y\|$.

Definition 2.2.2. Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map. A set $\mathcal{S} \subset \mathbb{R}^n$ is called a *leaf* of u if $u|_{\mathcal{S}}$ is an isometry and for any $y \notin \mathcal{S}$ there exists $x \in \mathcal{S}$ such that $\|u(y) - u(x)\| < \|y - x\|$.

In other words, $\mathcal{S}$ is a leaf if it is a maximal set, with respect to the order induced by inclusion, such that $u|_{\mathcal{S}}$ is an isometry.

Definition 2.2.3. If $C \subset \mathbb{R}^n$ is a convex set, then we shall call the *tangent space* of C the linear space $\mathrm{Aff}(C) - \mathrm{Aff}(C)$. We shall call the *relative interior* of C the relative interior with respect to the topology of $\mathrm{Aff}(C)$.

Lemma 2.2.4. *Let $\mathcal{S} \subset \mathbb{R}^n$ be an arbitrary subset. Let $u\colon \mathcal{S} \to \mathbb{R}^m$ be an isometry. Then there exists a unique 1-Lipschitz function $\tilde{u}\colon \mathrm{Conv}(\mathcal{S}) \to \mathbb{R}^m$ such that $\tilde{u}|_{\mathcal{S}} = u$. Moreover $\tilde{u}$ is an isometry.*

Proof. Observe that, by the polarisation formula, u preserves the scalar product, that is for all points $p, q, r, s \in \mathcal{S}$ there is

$$\begin{aligned}
&\langle u(p) - u(q), u(r) - u(s)\rangle = \\
&= \frac{1}{2}\big(\|u(p) - u(s)\|^2 + \|u(q) - u(r)\|^2 - \|u(p) - u(r)\|^2 - \|u(q) - u(s)\|^2\big) = \\
&= \frac{1}{2}\big(\|p - s\|^2 + \|q - r\|^2 - \|p - r\|^2 - \|q - s\|^2\big) = \langle p - q, r - s\rangle.
\end{aligned} \tag{2.2.1}$$

Suppose that $y_1, \dots, y_k, z_1, \dots, z_l \in \mathcal{S}$ and that $s_1, \dots, s_k$, $t_1, \dots, t_l$ are non-negative real numbers such that

$$\sum_{i=1}^k s_i = \sum_{j=1}^l t_j = 1.$$

Then, by (2.2.1),

$$\begin{aligned}
&\Big\| \sum_{i=1}^k s_i u(y_i) - \sum_{j=1}^l t_j u(z_j) \Big\|^2 = \Big\| \sum_{i=1}^k \sum_{j=1}^l s_i t_j (u(y_i) - u(z_j)) \Big\|^2 = \\
&= \sum_{i,i'=1}^k \sum_{j,j'=1}^l s_i s_{i'} t_j t_{j'} \langle u(y_i) - u(z_j), u(y_{i'}) - u(z_{j'})\rangle = \\
&= \sum_{i,i'=1}^k \sum_{j,j'=1}^l s_i s_{i'} t_j t_{j'} \langle y_i - z_j, y_{i'} - z_{j'}\rangle = \Big\| \sum_{i=1}^k s_i y_i - \sum_{j=1}^l t_j z_j \Big\|^2.
\end{aligned} \tag{2.2.2}$$

We may now affinely extend u to $\mathrm{Conv}(\mathcal{S})$. That is, if $x_1, \dots, x_k \in \mathcal{S}$ and $s_1, \dots, s_k$ are any non-negative real numbers that sum up to one, we set

$$\tilde{u}\Big(\sum_{i=1}^k s_i x_i \Big) = \sum_{i=1}^k s_i u(x_i).$$

Now, (2.2.2) shows that $\tilde{u}$ is a well-defined affine map on $\mathrm{Conv}(\mathcal{S})$ and that it is an isometry.

Suppose now that we have another 1-Lipschitz extension $v\colon \mathrm{Conv}(\mathcal{S}) \to \mathbb{R}^m$. To prove that $v = \tilde{u}$ it is enough to show that v is affine. Choose non-negative real numbers $s_1, \dots, s_k$ summing up to one and any points $x_1, \dots, x_k \in \mathcal{S}$. Then, by 1-Lipschitzness and by the fact that v is isometric on $\mathcal{S}$, we get, as in (2.2.2),

$$\Big\| v\Big(\sum_{i=1}^k s_i x_i\Big) - v(x_j)\Big\| \leq \Big\| \sum_{i=1}^k s_i x_i - x_j \Big\| = \Big\| \sum_{i=1}^k s_i v(x_i) - v(x_j)\Big\|.$$

By Lemma 2.2.1 we see that

$$v\Bigg(\sum_{i=1}^k s_i x_i\Bigg) = \sum_{i=1}^k s_i v(x_i).$$

It follows that v is affine on $\mathrm{Conv}(\mathcal{S})$. $\square$

Corollary 2.2.5. *Any leaf $\mathcal{S}$ of u is a closed convex set and $u|_{\mathcal{S}}$ is an affine isometry.*

Let $\mathcal{S}$ be a leaf of u. Let P denote the orthogonal projection of $\mathbb{R}^n$ onto the tangent space V of $\mathcal{S}$. Let

$$T\colon V \to \mathbb{R}^m$$

be a linear isometry such that

$$u(x) - u(y) = T(x - y)$$

for any $x, y \in \mathcal{S}$. Let Q denote the orthogonal projection of $\mathbb{R}^m$ onto $T(V)$.

Below by $\mathrm{int}\mathcal{S}$, $\mathrm{cl}\mathcal{S}$, $\partial\mathcal{S}$ we understand the relative interior, the relative closure and the relative boundary of $\mathcal{S}$ respectively.

Lemma 2.2.6. *Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map. Let $\mathcal{S}_1, \mathcal{S}_2$ be two leaves of u. Let V_1, V_2 be their respective tangent spaces and let P_1, P_2 be orthogonal projections onto V_1, V_2 respectively. Let T_1, T_2 be isometric maps such that*

$$u(x) - u(y) = T_i(x - y) \text{ for all } x, y \in \mathcal{S}_i, i = 1, 2.$$

Let $x_i \in \mathcal{S}_i$ and $\sigma_i = \mathrm{dist}(x_i, \partial\mathcal{S}_i)$ for $i = 1, 2$. Then

$$2\sigma_1\sigma_2 \|P_1 P_2 - P_1 T_1^* T_2 P_2\| \leq \|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2,$$

and for $i = 1, 2$

$$2\sigma_i \|P_i T_i^*(u(x_1) - u(x_2)) - P_i(x_1 - x_2)\| \leq \|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2.$$

Proof. Let $y_i \in \mathcal{S}_i$ for $i = 1, 2$. Let $v_i = y_i - x_i$ for $i = 1, 2$. Then we may write

$$u(y_1) - u(y_2) = u(x_1) - u(x_2) + T_1 v_1 - T_2 v_2.$$

Hence $\|u(y_1) - u(y_2)\|^2$ is equal to

$$\|u(x_1) - u(x_2)\|^2 + \|v_1\|^2 + \|v_2\|^2 + 2\langle u(x_1) - u(x_2), T_1 v_1 - T_2 v_2\rangle - 2\langle T_1 v_1, T_2 v_2\rangle.$$

We also have

$$y_1 - y_2 = x_1 - x_2 + v_1 - v_2,$$

yielding

$$\|y_1 - y_2\|^2 = \|x_1 - x_2\|^2 + \|v_1\|^2 + \|v_2\|^2 + 2\langle x_1 - x_2, v_1 - v_2\rangle - 2\langle v_1, v_2\rangle.$$

As u is 1-Lipschitz, $\|u(y_1) - u(y_2)\| \leq \|y_1 - y_2\|$. By the two identities above we get therefore that

$$2\langle v_1, v_2\rangle - 2\langle T_1 v_1, T_2 v_2\rangle + 2\langle u(x_1) - u(x_2), T_1 v_1 - T_2 v_2\rangle - 2\langle x_1 - x_2, v_1 - v_2\rangle$$

is bounded above by

$$\|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2.$$

Suppose that σ_1, σ_2 are both positive. As y_1, y_2 were arbitrary points of $\mathcal{S}_1, \mathcal{S}_2$ respectively, the above inequality holds true for any $v_1 \in V_1$ and any $v_2 \in V_2$ of norm at most σ_1 and σ_2 respectively. If we add two such inequalities with v_1, v_2 replaced by $-v_1, -v_2$ then we get that

$$2\langle v_1, v_2\rangle - 2\langle T_1 v_1, T_2 v_2\rangle \leq \|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2. \tag{2.2.3}$$

Equivalently, for any $w_1, w_2 \in \mathbb{R}^n$ of norm at most one, we have

$$2\sigma_1\sigma_2\big\langle w_1, (P_1 P_2 - P_1 T_1^* T_2 P_2) w_2\big\rangle \leq \|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2.$$

Taking supremum over all $w_1, w_2 \in \mathbb{R}^n$ of norm at most one yields the first desired inequality. For the next inequalities, we assume that $\sigma_2 > 0$ and we take $v_1 = 0$ to get that

$$-2\langle u(x_1) - u(x_2), T_2 v_2\rangle + 2\langle x_1 - x_2, v_2\rangle \leq \|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2.$$

Analogously for $v_2 = 0$ and $\sigma_1 > 0$

$$2\langle u(x_1) - u(x_2), T_1 v_1\rangle - 2\langle x_1 - x_2, v_1\rangle \leq \|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2.$$

Hence for any $w_1, w_2 \in \mathbb{R}^n$ of norm at most one there is

$$\sigma_2\Big\langle \big(P_2 T_2^*\big(u(x_1) - u(x_2)\big) - P_2(x_1 - x_2)\big), w_2\Big\rangle \leq \|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2$$

and

$$\sigma_1\Big\langle \big(P_1T_1^*\big(u(x_1)-u(x_2)\big)-P_1(x_1-x_2)\big), w_1\Big\rangle \le \|x_1-x_2\|^2-\|u(x_1)-u(x_2)\|^2.$$

Taking suprema over w_1, w_2 in the unit ball of $\mathbb{R}^n$ yields the desired results. □

Remark 2.2.7. Lemma 2.2.6 tells us that if x_1, x_2 belong to relative interiors of leaves $\mathcal{S}_1, \mathcal{S}_2$ respectively, then the 1-Lipschitzness of map $u\colon \mathbb{R}^n \to \mathbb{R}^m$ is strengthened to the condition that

$$\|u(x_1)-u(x_2)\|^2+2\sigma_1\sigma_2\|P_1P_2-P_1T_1^*T_2P_2\| \le \|x_1-x_2\|^2$$

for all $x_1 \in \mathcal{S}_1$ and all $x_2 \in \mathcal{S}_2$.

Lemma 2.2.8. *Let $\mathcal{S}$ be a leaf of a 1-Lipschitz map $u\colon \mathbb{R}^n \to \mathbb{R}^m$. Then Qu is differentiable in the relative interior of $\mathcal{S}$. Moreover, if z_0 belongs to the relative interior of $\mathcal{S}$, then*

$$DQu(z_0)=TP.$$

If u is differentiable in z_0 for some $z_0 \in \mathcal{S}$, then

$$QDu(z_0)=TP.$$

Proof. Observe that $Q=TT^*$. Hence, by Lemma 2.2.6, we see that

$$2\sigma\|Q(u(z_1)-u(z_0))-TP(z_1-z_0)\| \le \|z_1-z_0\|^2-\|u(z_1)-u(z_0)\|^2.$$

for all $z_0 \in \mathcal{S}$ and $z_1 \in \mathbb{R}^n$. Here $\sigma = \operatorname{dist}(z_0, \partial\mathcal{S})$. Hence if $\sigma > 0$ we obtain that

$$\limsup_{z_1\to z_0}\frac{\|Q(u(z_1)-u(z_0))-TP(z_1-z_0)\|}{\|z_1-z_0\|} \le \limsup_{z_1\to z_0}\frac{\|z_1-z_0\|}{2\sigma}=0.$$

This yields the asserted differentiability.

Now, suppose that u is differentiable at $z_0 \in \mathcal{S}$. Arguing as in the proof of Lemma 2.2.6 we see that for all $z_2 \in \mathcal{S}$ and $z_1 \in \mathbb{R}^n$ we have

$$2\big\langle T^*(u(z_0)-u(z_1))-(z_0-z_1), z_2-z_0\big\rangle \le \|z_1-z_0\|^2-\|u(z_1)-u(z_0)\|^2.$$

Take any $w \in \mathbb{R}^n$ and let $z_1 = z_0 - tw$, $t>0$. Then the above inequality implies that

$$\Big\langle T^*\Big(\frac{u(z_0)-u(z_0-tw)}{t}\Big)-w, z_2-z_0\Big\rangle \le \frac{t\|w\|^2}{2}.$$

Letting t tend to zero yields

$$\langle T^*Du(z_0)w-w, z_2-z_0\rangle \le 0.$$

As this holds true for any $w \in \mathbb{R}^n$, applying this inequality to $-w$, we infer that the above inequality is an equality, i.e. for all $w \in \mathbb{R}^n$ there is

$$\langle T^* Du(z_0)w - w, z_2 - z_0 \rangle = 0.$$

If follows that for all $v \in \operatorname{span}\{z_2 - z_0 \mid z_2 \in \mathcal{S}\} = V$

$$\langle T^* Du(z_0)w - w, v \rangle = 0,$$

and, consequently, for all such v there is $\langle QDu(z_0)w - TPw, Tv \rangle = 0$. The assertion follows. □

Corollary 2.2.9. *Suppose that $\mathcal{S}$ is of dimension m. Then u is differentiable in the relative interior of $\mathcal{S}$.*

Proof. If the dimension of $\mathcal{S}$ is m, then the respective orthogonal projection Q is the identity. The claim follows now by Lemma 2.2.8. □

Corollary 2.2.10. *Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map. Let $x_i \in \operatorname{int}\mathcal{S}_i$ belong to the relative interior of leaf $\mathcal{S}_i$ of u, for $i = 1, 2$. Let $\sigma_i = \operatorname{dist}(\partial\mathcal{S}_i, x_i)$ for $i = 1, 2$. Then for any $s_1, s_2 \in \mathbb{R}^n$ of norm at most one there is*

$$\big|\|P_1 s_1 - P_2 s_2\|^2 - \|Du(x_1)s_1 - Du(x_2)s_2\|^2\big| \leq \frac{\|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2}{2\sigma_1\sigma_2}.$$

Here P_i denote the orthogonal projection onto the tangent subspace of the leaf $\mathcal{S}_i$ for $i = 1, 2$. Moreover for any $w_1, w_2 \in \mathbb{R}^m$ of norm at most one there is

$$\big|\|Q_1 w_1 - Q_2 w_2\|^2 - \|\big(DQ_1 u(x_1)\big)^* w_1 - \big(DQ_2 u(x_2)\big)^* w_2\|^2\big| \leq \frac{\|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2}{2\sigma_1\sigma_2}.$$

Here Q_i denote the orthogonal projection onto the image of T_i, for $i = 1, 2$.

Proof. Formula (2.2.3), Lemma 2.2.6, tells us that for any $v_1 \in V_1$ and any $v_2 \in V_2$ of norm at most one, there is

$$\big|\|v_1 - v_2\|^2 - \|T_1 v_1 - T_2 v_2\|^2\big| \leq \frac{\|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2}{2\sigma_1\sigma_2}.$$

Lemma 2.2.8 tells us that $DQ_i u(x_i) = T_i P_i$ for $i = 1, 2$. Hence the first asserted inequality follows. Let $v_i = (T_i P_i)^* w_i$ for $w_i \in \mathbb{R}^m$, $i = 1, 2$, of norm at most one. Then the above formula yields

$$\big|\|(T_1 P_1)^* w_1 - (T_2 P_2)^* w_2\|^2 - \|Q_1 w_1 - Q_2 w_2\|^2\big| \leq \frac{\|x_1 - x_2\|^2 - \|u(x_1) - u(x_2)\|^2}{2\sigma_1\sigma_2}.$$

The proof is complete. □

Lemma 2.2.11. *Let $\mathcal{S}_1, \mathcal{S}_2$ be two distinct leaves of a 1-Lipschitz map $u\colon \mathbb{R}^n \to \mathbb{R}^m$. Then*

$$\mathcal{S}_1 \cap \mathcal{S}_2 \subset \partial\mathcal{S}_1 \cap \partial\mathcal{S}_2.$$

Proof. We shall first show that there is no point belonging to $\mathrm{int}\mathcal{S}_1 \cap \mathcal{S}_2$. For this, suppose that $x_0 \in \mathrm{int}\mathcal{S}_1 \cap \mathcal{S}_2$. Let $x_1 \in \mathcal{S}_1$ and $x_2 \in \mathcal{S}_2$. There exists isometries T_1 and T_2 on the tangent spaces V_1 and V_2 of $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively such that

$$u(x_1) - u(x_0) = T_1(x_1 - x_0) \text{ and } u(x_2) - u(x_0) = T_2(x_2 - x_0).$$

We may write

$$\begin{aligned} &\|x_1 - x_0\|^2 + \|x_2 - x_0\|^2 - 2\langle T_1(x_1 - x_0), T_2(x_2 - x_0)\rangle = \|u(x_1) - u(x_2)\|^2 \leq \\ &\leq \|x_1 - x_2\|^2 = \|x_1 - x_0\|^2 + \|x_2 - x_0\|^2 - 2\langle x_1 - x_0, x_2 - x_0\rangle. \end{aligned}$$

Hence

$$\langle x_1 - x_0, x_2 - x_0\rangle \leq \langle T_1(x_1 - x_0), T_2(x_2 - x_0)\rangle.$$

As $x_0 \in \mathrm{int}\mathcal{S}_1$ and the inequality holds true for all $x_1 \in \mathcal{S}_1$, we actually have equality above for x_1 sufficiently close to x_0. It follows that for all $v_1 \in V_1$ and $v_2 \in V_2$,

$$\langle v_1, v_2\rangle = \langle T_1 v_2, T_2 v_2\rangle. \tag{2.2.4}$$

Hence, there exists an isometry that extends both T_1 and T_2. Indeed, define a linear map

$$S\colon V_1 + V_2 \to \mathbb{R}^m$$

by the formula

$$S(v_1 + v_2) = T_1(v_1) + T_2(v_2) \text{ for } v_1 \in V_1, v_2 \in V_2.$$

We claim that S is a well-defined isometry. Indeed, by (2.2.4) and by orthogonality we see that if $v_2 \in V_1 \cap V_2$, then

$$\|v_2\|^2 = \langle T_1 v_2, T_2 v_2\rangle$$

which implies, by the equality cases in the Cauchy–Schwarz inequality, that $T_1 v_2 = T_2 v_2$. Thus S is well-defined. It is an isometry, as for $v_1 \in V_1$ and $v_2 \in V_2$,

$$\|S(v_1 + v_2)\|^2 = \|v_1\|^2 + \|v_2\|^2 + 2\langle T_1 v_1, T_2 v_2\rangle = \|v_1 + v_2\|^2.$$

Moreover, by the definition, S is an extension of both T_1 and T_2.

Define an affine map $v\colon x_0 + V_1 + V_2 \to \mathbb{R}^m$ by the formula

$$v(x) = S(x - x_0) + u(x_0).$$

Then $v|_{\mathcal{S}_1} = u$ and $v|_{\mathcal{S}_2} = u$. Choose any points $x \in \mathcal{S}_1$ and $y \in \mathcal{S}_2$. Then

$$\|u(x) - u(y)\| = \|v(x) - v(y)\| = \|S(x - y)\| = \|x - y\|.$$

Thus u is isometric on $\mathcal{S}_1 \cup \mathcal{S}_2$. By maximality of leaves, $\mathcal{S}_1 = \mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{S}_2$, contradicting the distinctness of the two leaves. Hence

$$\mathcal{S}_1 \cap \mathcal{S}_2 \subset \partial\mathcal{S}_1 \cap \mathcal{S}_2.$$

Repeating the above argument with $\mathcal{S}_1$ and $\mathcal{S}_2$ interchanged, we see that

$$\mathcal{S}_1 \cap \mathcal{S}_2 \subset \big(\partial\mathcal{S}_1 \cap \mathcal{S}_2\big) \cap \big(\partial\mathcal{S}_2 \cap \mathcal{S}_1\big) = \partial\mathcal{S}_1 \cap \partial\mathcal{S}_2.$$

□

Lemma 2.2.12. *Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be 1-Lipschitz. If $x_0 \in \mathbb{R}^n$ belongs to at least two distinct leaves of u, then u is not differentiable at x_0.*

Proof. Clearly, any zero-dimensional leaf does not intersect any other leaf. Hence, x_0 belongs to two distinct leaves $\mathcal{S}_1, \mathcal{S}_2$ of non-zero dimensions. Suppose that u is differentiable at $x_0 \in \mathcal{S}_1 \cap \mathcal{S}_2$. Then Lemma 2.2.8 implies that $Q_1 u$ is differentiable at x_0 with the derivative given by

$$DQ_1 u(x_0) = T_1 P_1,$$

where T_1 is an isometry such that $u(x) - u(x_0) = T_1(x - x_0)$ for all $x \in \mathcal{S}_1$, P_1 is the orthogonal projection onto the tangent space V_1 of $\mathcal{S}_1$ and Q_1 is the orthogonal projection onto the image of T_1. In other words

$$\lim_{x \to x_0} \frac{Q_1 u(x) - Q_1 u(x_0) - T_1 P_1(x - x_0)}{\|x - x_0\|} = 0. \tag{2.2.5}$$

For $x \in \mathcal{S}_2$ we may write

$$u(x) - u(x_0) = T_2(x - x_0)$$

for an isometry T_2. If $x \in \mathcal{S}_2$, then

$$\frac{Q_1 u(x) - Q_1 u(x_0) - T_1 P_1(x - x_0)}{\|x - x_0\|} = (Q_1 T_2 - T_1 P_1)\bigg(\frac{x - x_0}{\|x - x_0\|}\bigg). \tag{2.2.6}$$

For $x \in \mathcal{S}_2$ and $t \in [0, 1]$ let

$$x_t = x_0 + t(x - x_0).$$

By convexity of leaves, $x_t \in \mathcal{S}_2$. Observe also that

$$\lim_{t \to 0} x_t = x_0. \tag{2.2.7}$$

It follows by (2.2.5), (2.2.6) and by (2.2.7) that

$$Q_1 T_2(x - x_0) = T_1 P_1(x - x_0) \text{ for all } x \in \mathcal{S}_2.$$

As $V_2 = \operatorname{span}\{x - x_0 \mid x \in \mathcal{S}_2\}$ is the tangent space of $\mathcal{S}_2$, we infer that

$$Q_1 T_2 v = T_1 P_1 v \text{ for all } v \in V_2.$$

Hence, for $v_1 \in V_1$ and for $v_2 \in V_2$

$$\langle T_1 v_1, T_2 v_2 \rangle = \langle T_1 v_1, Q_1 T_2 v_2 \rangle = \langle T_1 v_1, T_1 P_1 v_2 \rangle = \langle v_1, v_2 \rangle.$$

We continuue the proof as in Lemma 2.2.11 and arrive at a contradiction that $\mathcal{S}_1 = \mathcal{S}_2$. □

Remark 2.2.13. We may proceed in the first part of the above proof of Lemma 2.2.11 alternatively. Namely, we may conclude from Lemma 2.2.8 that for at point in int$\mathcal{S}$ the map Qu is differentiable with $DQu = TP$. Then we proceed as in the proof of Lemma 2.2.12.

Definition 2.2.14. The set of points belonging to at least two distinct leaves of a 1-Lipschitz map $u\colon \mathbb{R}^n \to \mathbb{R}^m$ we shall denote by $B(u)$.

Corollary 2.2.15. *For any* 1*-Lipschitz function* $u\colon \mathbb{R}^n \to \mathbb{R}^m$ *the set* $B(u)$ *is of Lebesgue measure zero.*

Proof. Lemma 2.2.12 implies that $B(u)$ is contained in the set of non-differentiability of u. Rademacher's theorem (see e.g. [47]) states that the latter is of Lebesgue measure zero. □

2.3 Ghost subspaces

In this section we shall study certain subspaces associated to leaves of a 1-Lipschitz map $u\colon \mathbb{R}^n \to \mathbb{R}^m$. Observe that if $\mathcal{S} \subset \mathbb{R}^n$ is a leaf of u, then the derivative of u is isometric on the tangent space V of $\mathcal{S}$. Nevertheless, it may happen that the derivative of u, whenever it exists, is an isometry on a strictly larger subspace than V. An example illustrating this situation is provided below. Note that such situation does not occur when $\mathcal{S}$ has dimension m.

Example 2.3.1. We shall construct a 1-Lipschitz function $u\colon [0,1] \to \mathbb{R}$ such that there is a set E of positive Lebesgue measure consisting of zero-dimensional leaves and such that Du is almost everywhere an isometry, i.e. has slope one or negative one. In particular, for all $x \in E$ and all $y \in [0,1]$ there is $|u(x) - u(y)| < |x - y|$. For a construction of such a function, consider a dense set F that is a union of a countable family of pairwise disjoint, non-degenerate intervals in $[0,1]$ of total measure $\lambda(F)\in(0,1)$. Let u be defined by the formula

$$u(x)=\int_{[0,x]}(\mathbf{1}_{F^c}-\mathbf{1}_F)(t)dt.$$

Then u is picewise affine – in each of the intervals in F with slope equal to negative one and in F^c with slope equal to one. Then clearly, Du is an isometry almost everywhere, yet, for any $x\in F^c$ the set $\{x\}$ is a leaf of u. Taking $E=F^c$ yields the asserted claim.

In what follows we shall prove that u is isometric not only on $\mathcal{S}$, but also, there is a subspace of vectors on which $Du(x)$ is exists and is isometric, whenever $x\in\operatorname{int}\mathcal{S}$. The derivative on these vectors is independent on the choice of a point in $\operatorname{int}\mathcal{S}$. We prove also that the rates of convergence of the differential quotients to the derivative are comparable on this subspace. Thus, in a sense, some leaves behave as it they were of higher dimension than they actually are.

Lemma 2.3.2. *Let $u\colon\mathbb{R}^n\to\mathbb{R}^m$ be a 1-Lipschitz map. Let $\mathcal{S}$ be a leaf of u and let V denote the tangent space to $\mathcal{S}$. Let $x_0,x_1\in\operatorname{int}\mathcal{S}$. Suppose that for some $h\in\mathbb{R}^n$ there exists $Du(x_0)h$ with $\|Du(x_0)h\|=\|h\|$. Then there exists $Du(x_1)h$ and*

$$Du(x_1)h=Du(x_0)h.$$

Moreover, if for some $\epsilon>0$, $\delta>0$ and all $t\in\mathbb{R}$ such that $|t|<\delta$ there is

$$\frac{\|u(x_0+th)-u(x_0)-tDu(x_0)h\|}{|t|\|h\|}<\epsilon$$

then also

$$\frac{\|u(x_1+rh)-u(x_1)-rDu(x_0)h\|}{|r|\|h\|}<\epsilon+\sqrt{2\epsilon}\sqrt{\frac{\|x_1-x_0\|}{\sigma_1}\left(1+\frac{\|x_1-x_0\|+\sigma_1}{\sigma_0}\right)}$$

whenever $|r|<\frac{\sigma_1}{\sigma_1+\|x_1-x_0\|}\delta$. Here for $i=0,1$ we set $\sigma_i=\operatorname{dist}(x_i,\partial\mathcal{S})$.

Proof. By the assumption, $\sigma_i>0$ for $i=0,1$. Let $x'\in\mathcal{S}$ be given by the formula

$$x'=x_1+\frac{x_1-x_0}{\|x_1-x_0\|}\sigma_1.$$

Then

$$x_1=\frac{\sigma_1}{\|x_1-x_0\|+\sigma_1}x_0+\frac{\|x_1-x_0\|}{\|x_1-x_0\|+\sigma_1}x'.$$

Set $s=\frac{\sigma_1}{\|x_1-x_0\|+\sigma_1}$. Let t be a real number. Observe that $x_1+sth=s(x_0+th)+(1-s)x'$. Let $\rho\in[0,1]$ be such that

$$1-\rho=\frac{\|u(x_0+th)-u(x_0)\|^2}{t^2\|h\|^2}. \tag{2.3.1}$$

Then we claim that

$$\|u(x_0+th)-u(x')\|^2 \geq \|x_0+th-x'\|^2 - \rho t^2\|h\|^2\left(1+\frac{\|x_0-x'\|}{\sigma_0}\right). \tag{2.3.2}$$

Indeed, Lemma 2.2.6 and Lemma 2.2.8 yield that

$$\|Q(u(x_0+th)-u(x_0)-tDu(x_0)h)\| \leq \frac{1}{2\sigma_0}(t^2\|h\|^2 - \|u(x_0+th)-u(x_0)\|^2) = \frac{\rho}{2\sigma_0}t^2\|h\|^2.$$

Note that

$$QDu(x_0)h = TPh.$$

Therefore, by the Pythagorean theorem and the above bound,

$$\begin{aligned}\|u(x_0+th)-u(x')\|^2 &= \|u(x_0+th)-u(x_0)\|^2 + \|u(x_0)-u(x')\|^2 + \\ &+ 2\langle T(x_0-x'), Q(u(x_0+th)-u(x_0)-tDu(x_0)h)\rangle + 2\langle x_0-x', tPh\rangle.\end{aligned}$$

Hence, using (2.3.1) and the fact that $\langle x_0-x', Ph\rangle = \langle x_0-x', h\rangle$, we may write

$$\begin{aligned}&\|u(x_0+th)-u(x')\|^2 \geq t^2\|h\|^2 + \|x_0-x'\|^2 - \rho t^2\|h\|^2\left(1+\frac{\|x_0-x'\|}{\sigma_0}\right) + 2\langle x_0-x', th\rangle = \\ &= \|x_0+th-x'\|^2 - \rho t^2\|h\|^2\left(1+\frac{\|x_0-x'\|}{\sigma_0}\right).\end{aligned}$$

Thus, (2.3.2) follows. Observe that, c.f. proof of Lemma 2.2.4,

$$\begin{aligned}&\langle u(x_1+sth)-u(x_0+th), u(x_1+sth)-u(x')\rangle = \\ &= \frac{1}{2}\big(\|u(x_1+sth)-u(x')\|^2 + \|u(x_1+sth)-u(x_0+th)\|^2 - \|u(x_0+th)-u(x')\|^2\big).\end{aligned}$$

Then we see, by 1-Lipschitzness of u and by (2.3.2), that the above scalar product is bounded from above by

$$\langle x_1+sth-(x_0+th), x_1+sth-x'\rangle + \frac{1}{2}\rho t^2\|h\|^2\left(1+\frac{\|x_0-x'\|}{\sigma_0}\right). \tag{2.3.3}$$

Expanding the squares, exploiting 1-Lipschitzness of u, (2.3.3) and the polarisation formula yields

$$\|u(x_1+sth)-(su(x_0+th)+(1-s)u(x'))\|^2 \leq s(1-s)\rho t^2\|h\|^2\left(1+\frac{\|x_0-x'\|}{\sigma_0}\right).$$

Now, since u is affine on $\mathcal{S}$, $u(x_1+sth)-(su(x_0+th)+(1-s)u(x'))$ is equal to

$$u(x_1+sth)-u(x_1)-stDu(x_0)h - s(u(x_0+th)-u(x_0)-tDu(x_0)h).$$

We see that the quotient

$$\frac{\|u(x_1+sth)-u(x_1)-stDu(x_0)h\|}{|st|\|h\|}$$

may be bounded by

$$\frac{\sqrt{\rho(1-s)}}{\sqrt{s}}\sqrt{1+\frac{\|x_0-x'\|}{\sigma_0}}+\frac{\|u(x_0+th)-u(x_0)-tDu(x_0)h\|}{|t|\|h\|}. \tag{2.3.4}$$

By the definition of ρ we have

$$\sqrt{1-\rho}=\frac{\|u(x_0+th)-u(x_0)\|}{|t|\|h\|}\geq 1-\frac{\|u(x_0+th)-u(x_0)-tDu(x_0)h\|}{|t|\|h\|}.$$

Hence

$$\rho\leq 2\frac{\|u(x_0+th)-u(x_0)-tDh(x_0)h\|}{|t|\|h\|}.$$

Observe also that

$$\|x_0-x'\|=\|x_0-x_1\|+\sigma_1$$

and that

$$\frac{1-s}{s}=\frac{\|x_1-x_0\|}{\sigma_1}.$$

Let $\epsilon>0$, $\delta>0$ and $|t|<\delta$ be as in the assertion of the lemma. We see by (2.3.4) that

$$\frac{\|u(x_1+rh)-u(x_1)-rDu(x_0)h\|}{|r|\|h\|}<\epsilon+\sqrt{2\epsilon}\sqrt{\frac{\|x_1-x_0\|}{\sigma_1}\left(1+\frac{\|x_1-x_0\|+\sigma_1}{\sigma_0}\right)}$$

provided that $|r|<\frac{\sigma_1}{\sigma_1+\|x_1-x_0\|}\delta$.

The proof is complete. □

The next proposition tells that the set of vectors $h\in\mathbb{R}^n$ such that $Du(x)h$ exists and $\|Du(x)h\|=\|h\|$ for some $x\in\mathrm{int}\mathcal{S}$ forms a subspace.

Proposition 2.3.3. *Let $\mathcal{S}$ be a leaf of a 1-Lipschitz map $u\colon\mathbb{R}^n\to\mathbb{R}^m$. The set G of vectors $h\in\mathbb{R}^n$ such that for some $x\in\mathrm{int}\mathcal{S}$ there exists $Du(x)h$ with $\|Du(x)h\|=\|h\|$ forms a subspace. Moreover, for any $x\in\mathrm{int}\mathcal{S}$ and any $h\in G$ there exists $Du(x)h$ and $Du(x)$ on G is an isometry, which is independent on the choice of $x\in\mathrm{int}\mathcal{S}$.*

Proof. Let $x\in\mathrm{int}\mathcal{S}$. Lemma 2.3.2 implies that the set G is equal to the set of $h\in\mathbb{R}^n$ such that here exists $Du(x)h$ with $\|Du(x)h\|=\|h\|$. Let $h_1,h_2\in G$. As u is 1-Lipschitz

$$\|Du(x)h_1-Du(x)h_2\|=\lim_{t\to 0}\frac{1}{t}\|u(x+th_1)-u(x+th_2)\|\leq\|h_1-h_2\|.$$

It follows that

$$\|Du(x)h_1\|^2+\|Du(x)h_2\|^2-2\langle Du(x)h_1,Du(x)h_2\rangle\leq\|h_1\|^2+\|h_2\|^2-2\langle h_1,h_2\rangle.$$

Hence

$$\langle Du(x)h_1,Du(x)h_2\rangle\geq\langle h_1,h_2\rangle. \tag{2.3.5}$$

Since the same is true with h_1 replaced by $-h_1$, we infer that in (2.3.5) we actually have an equality. Thus, $Du(x)$ is an isometry on G. In particular

$$\lim_{t\to 0}\frac{1}{t}\|u(x+th_1)-u(x+th_2)\| = \|h_1-h_2\|. \tag{2.3.6}$$

Suppose that

$$\liminf_{t\to 0}\frac{\left\|u(x+th_1)-u\left(x+\frac{t}{2}(h_1+h_2)\right)\right\|}{|t|} < \frac{1}{2}\|h_1-h_2\|.$$

Since for $t>0$

$$\frac{\left\|u(x+th_2)-u\left(x+\frac{t}{2}(h_1+h_2)\right)\right\|}{|t|} \leq \frac{1}{2}\|h_1-h_2\|$$

we see by the triangle inequality that

$$\liminf_{t\to 0}\frac{\|u(x+th_1)-u(x+th_2)\|}{|t|} < \|h_1-h_2\|,$$

contradicting (2.3.6). From the above and by symmetry it follows that for $i=1,2$

$$\lim_{t\to 0}\frac{1}{t}\left\|u(x+th_i)-u\left(x+\frac{t}{2}(h_1+h_2)\right)\right\| = \frac{1}{2}\|h_1-h_2\|. \tag{2.3.7}$$

By the polarisation formula we see that

$$\begin{aligned}
&\left\|u\left(x+\frac{t}{2}(h_1+h_2)\right)-\frac{1}{2}(u(x+th_1)+u(x+th_2))\right\|^2 = \\
&=\frac{1}{2}\left(\left\|u\left(x+\frac{t}{2}(h_1+h_2)\right)-u(x+th_1)\right\|^2+\left\|u\left(x+\frac{t}{2}(h_1+h_2)\right)-u(x+th_2)\right\|^2\right)- \\
&-\frac{1}{4}\left\|u(x+th_1)-u(x+th_2)\right\|^2.
\end{aligned}$$

By the above, by (2.3.7) and by (2.3.6) we infer that

$$\lim_{t\to 0}\frac{1}{t}\left\|u\left(x+\frac{t}{2}(h_1+h_2)\right)-u(x)-\frac{1}{2}((u(x+th_1)-u(x))+(u(x+h_2)-u(x))\right\| = 0.$$

Therefore the derivatie $Du(x)$ in direction of $\frac{h_1+h_2}{2}$ exists and

$$Du(x)\left(\frac{h_1+h_2}{2}\right) = \frac{1}{2}(Du(x)h_1+Du(x)h_2).$$

It follows that $\frac{h_1+h_2}{2}\in G$ whenever $h_1,h_2\in G$. Since homogeneity of G is trivial to prove, G is a linear subspace.

Lemma 2.3.2 tells us that $Du(x)$ restricted to G is independent on the choice of a point $x\in \operatorname{int}\mathcal{S}$. □

Definition 2.3.4. Let $u\colon \mathbb{R}^n\to\mathbb{R}^m$ be a 1-Lipschitz map. Let $\mathcal{S}$ be a leaf of u. A subspace $G\subset\mathbb{R}^n$ is called a *ghost subspace* associated to the leaf $\mathcal{S}$ if for any, or – equivalently – some, $x\in\operatorname{int}\mathcal{S}$ and any $h\in G$ there exists $Du(x)h$ and $\|Du(x)h\|=\|h\|$.

The name ghost subspace comes from the fact that this subspace is not visible at the level of partitioning into leaves, yet it is of significant importance when dealing with the partition.

Note that on a ghost subspace, associated to a leaf $\mathcal{S}$, the derivative is isometric and it is the maximal subspace with this property. Clearly, a ghost subspace contains the tangent space to the leaf $\mathcal{S}$.

Lemma 2.3.5. *Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map, let $\mathcal{S}$ be its leaf. If $h \in \mathbb{R}^n$ is such that $\|Du(x)h\| = \|h\|$ for some $x \in \operatorname{int}\mathcal{S}$, then there exists $D^2Qu(x)(h,h)$ and it is equal to zero.*

Proof. Lemma 2.2.6 and Lemma 2.2.8 yield that for $y \in \mathbb{R}^n$

$$\|Q(u(y) - u(x) - DQu(x)(y-x))\| \leq \frac{1}{\sigma}\big(\|y-x\|^2 - \|u(y)-u(x)\|^2\big),$$

where $\sigma = \operatorname{dist}(x, \partial\mathcal{S})$. Let $y = x + th$ for $t \in \mathbb{R}$. Dividing by t^2 and exploiting the fact that

$$\lim_{t\to 0} \frac{\|u(x+th) - u(x)\|}{|t|\|h\|} = 1$$

we arrive at the assertion. $\square$

The following proposition provides a result on the regularity of the derivative of u on a single leaf.

Proposition 2.3.6. *Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map, let $\mathcal{S}$ be its leaf. Let $x_1, x_2 \in \operatorname{int}\mathcal{S}$ and let $\sigma_i = \operatorname{dist}(x_i, \partial\mathcal{S})$ for $i = 1,2$. Then for $h_1, h_2 \in \mathbb{R}^n$*

$$\|Du(x_1)h_1 - Du(x_2)h_2\|^2$$

is at most

$$\|h_1 - h_2\|^2 + 2\|x_1 - x_2\|\bigg(\frac{1}{\sigma_1}\Big(\|h_1\|^2 - \|Du(x_1)h_1\|^2\Big) + \frac{1}{\sigma_2}\Big(\|h_2\|^2 - \|Du(x_2)h_2\|^2\Big)\bigg),$$

whenever derivatives $Du(x_1)h_1$ and $Du(x_2)h_2$ exist. In particular, the derivative of u on $\operatorname{int}\mathcal{S}$ is locally $\frac{1}{2}$-Hölder continuous.

Proof. Suppose $x_1, x_2 \in \operatorname{int}\mathcal{S}$. Suppose that $h_1, h_2 \in \mathbb{R}^n$. Let $t \in \mathbb{R}$, $y_i = x_i + th_i$ for $i = 1,2$ and observe that

$$\|u(y_1) - u(y_2)\|^2 \leq \|y_1 - y_2\|^2.$$

Expanding the squares, like in the proof of Lemma 2.2.6, we obtain that

$$\|u(x_1) - u(x_2)\|^2 + \|(u(y_1) - u(x_1)) - (u(y_2) - u(x_2))\|^2 +$$
$$+ 2\langle u(x_1) - u(x_2), (u(y_1) - u(x_1)) - (u(y_2) - u(x_2))\rangle$$

is bounded above by

$$\|x_1 - x_2\|^2 + \|(y_1 - x_1) - (y_2 - x_2)\|^2 + 2\langle x_1 - x_2, (y_1 - x_1) - (y_2 - x_2)\rangle.$$

Taking into account that $\|u(x_1) - u(x_2)\| = \|x_1 - x_2\|$, $y_i - x_i = th_i$ for $i = 1, 2$, we see that

$$\begin{aligned} &\|(u(y_1) - u(x_1)) - (u(y_2) - u(x_2))\|^2 - t^2\|h_1 - h_2\|^2 \leq \\ &\leq 2t\langle x_1 - x_2, h_1 - h_2\rangle - 2\langle u(x_1) - u(x_2), (u(y_1) - u(x_1)) - (u(y_2) - u(x_2))\rangle. \end{aligned} \tag{2.3.8}$$

Lemma 2.2.8 tells us that

$$\langle u(x_1) - u(x_2), Du(x_1)h_1 - Du(x_2)h_2\rangle = \langle x_1 - x_2, h_1 - h_2\rangle.$$

Lemma 2.2.6 yields that for $i = 1, 2$

$$\langle u(x_1) - u(x_2), u(y_i) - u(x_i) - tDu(x_i)h_i\rangle \leq \frac{1}{\sigma_i}\|x_1 - x_2\|\big(t^2\|h_i\|^2 - \|u(y_i) - u(x_i)\|^2\big).$$

We infer further that

$$\begin{aligned} &\|(u(y_1) - u(x_1)) - (u(y_2) - u(x_2))\|^2 \leq t^2\|h_1 - h_2\|^2 + \\ &+ 2\|x_1 - x_2\|\Big(\frac{1}{\sigma_1}\big(t^2\|h_1\|^2 - \|u(y_1) - u(x_1)\|^2\big) + \frac{1}{\sigma_2}\big(t^2\|h_2\|^2 - \|u(y_2) - u(x_2)\|^2\big)\Big). \end{aligned}$$

Now, dividing both sides of the above inequality by t^2 and letting t tend to zero, we infer that $\|Du(x_1)h_1 - Du(x_2)h_2\|^2$ is bounded above by

$$\|h_1 - h_2\|^2 + 2\|x_1 - x_2\|\Big(\frac{1}{\sigma_1}\big(\|h_1\|^2 - \|Du(x_1)h_1\|^2\big) + \frac{1}{\sigma_2}\big(\|h_2\|^2 - \|Du(x_2)h_2\|^2\big)\Big).$$

The proof is complete. □

2.4 Lipschitz change of variables

We assume throughout the section that $m \leq n$. Let us recall a lemma taken from [47, 3.2.9].

Lemma 2.4.1. *Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function. Then the set*

$$\{x \in \mathbb{R}^n \mid u \textit{ is differentiable at } x \textit{ and } Du(x) \textit{ has maximal rank}\}$$

admits a countable Borel cover $(G_i)_{i=1}^{\infty}$ such that for any $i \in \mathbb{N}$ there exists an orthogonal projection $\pi_i\colon \mathbb{R}^n \to \mathbb{R}^{n-m}$ and Lipschitz maps

$$w_i\colon \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^{n-m},\ v_i\colon \mathbb{R}^m \times \mathbb{R}^{n-m} \to \mathbb{R}^n$$

such that

$$w_i(x) = (u(x), \pi_i(x)) \textit{ and } v_i(w_i(x)) = x \textit{ for all } x \in G_i.$$

Lemma 2.4.2. *Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz function. Let $s \in \mathbb{R}^m$ and let*

$$S_s = \{x \in \mathbb{R}^n \mid u(x) = s\}$$

be the level set. Then the set

$$S_s \cap \{x \in \mathbb{R}^n \mid u \text{ is differentiable at } x \text{ and } Du(x) \text{ has maximal rank}\}$$

has a countable Borel cover $(S_s^i)_{i=1}^\infty$ of bounded sets such that for all $i \in \mathbb{N}$ there exist Lipschitz functions $w\colon \mathbb{R}^n \to \mathbb{R}^{n-m}$ and $v\colon \mathbb{R}^{n-m} \to \mathbb{R}^n$ satisfying

$$v(w(x)) = x \text{ for all } x \in S_s^i.$$

Proof. We apply Lemma 2.4.1 and obtain a countable cover consisting of Borel sets G_i, orthogonal projections $\pi_i\colon \mathbb{R}^n \to \mathbb{R}^{n-m}$ and Lipschitz maps

$$w_i\colon \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^{n-m},\ v_i\colon \mathbb{R}^m \times \mathbb{R}^{n-m} \to \mathbb{R}^n$$

such that

$$w_i(x) = (u(x), \pi_i(x)) \text{ and } v_i(w_i(x)) = x \text{ for all } x \in G_i.$$

The sets $G_i \cap S_s$ form a countable Borel cover of S_s. For any $i \in \mathbb{N}$ define

$$w\colon \mathbb{R}^n \to \mathbb{R}^{n-m} \text{ and } v\colon \mathbb{R}^{n-m} \to \mathbb{R}^n$$

by $w = \pi \circ w_i$, where $\pi\colon \mathbb{R}^m \times \mathbb{R}^{n-m} \to \mathbb{R}^{n-m}$ is the projection on the second variable, and $v(x) = v_i(s, x)$ for $x \in \mathbb{R}^{n-m}$. Then, if $u(x) = s$, then

$$v(w(x)) = v_i(s, w(x)) = v_i(u(x), \pi_i(x)) = x.$$

□

Definition 2.4.3. Pick a countable dense set $S \subset \mathbb{R}^m$. Let $s \in S$. Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map. Let $(S_s^i)_{i=1}^\infty$ be the Borel cover of Lemma 2.4.2 associated to the level set

$$S_s = \{x \in \mathbb{R}^n \mid u(x) = s\}.$$

For each $i, j \in \mathbb{N}$ let the *cluster*

$$T_{sij}$$

denote the union of all m-dimensional leaves $\mathcal{S}$ of u such that there exists $z \in \mathcal{S} \cap S_s^i$ for which $\operatorname{dist}(z, \partial\mathcal{S}) \geq \frac{1}{j}$. Denote by

$$\operatorname{int} T_{sij}$$

the union of the relative interiors of all m-dimensional leaves $\mathcal{S}$ of u as above.

Lemma 2.4.4. *The union of all m-dimensional leaves is covered by the clusters*

$$(T_{sij})_{s\in S, i,j\in\mathbb{N}}.$$

Moreover for each m-dimensional leaf $\mathcal{S}$ and each cluster T_{sij} either

$$\text{int}\mathcal{S} \cap T_{sij} = \emptyset \text{ or } \text{int}\mathcal{S} \subset T_{sij}.$$

Proof. Let $\mathcal{S}$ be a m-dimensional leaf of u. Then u, if restricted to $\mathcal{S}$, is an isometry onto a convex subset of $\mathbb{R}^m$. Thus, there exists $s \in S \cap \text{int}u(\mathcal{S})$. There exist $i, j \in \mathbb{N}$ and $z \in \mathcal{S} \cap S_s^i$ such that $\text{dist}(z, \partial\mathcal{S}) > 1/j$. That is $\mathcal{S} \subset T_{sij}$.

If the interior of some leaf int$\mathcal{S}$ intersects one of the leaves comprising the cluster T_{sij}, then Lemma 2.2.11 implies that they are equal and hence $\mathcal{S} \subset T_{sij}$. This completes the proof. □

Lemma 2.4.5. *Each cluster $T_{sij} \subset \mathbb{R}^n$ admits maps*

$$G\colon \text{int}T_{sij} \to \mathbb{R}^{n-m} \times \mathbb{R}^m$$

and

$$F\colon G(\text{int}T_{sij}) \to \text{int}T_{sij}$$

such that:

i) for each $\lambda > 0$, G is a Lipschitz map on the set

$$T_{sij}^{\lambda} = \Big\{x \in \text{int}T_{sij} \mid \text{dist}(x, \partial\mathcal{S}(x)) > \lambda\Big\};$$

here $\mathcal{S}(x)$ is the unique leaf of u such that $x \in \mathcal{S}(x)$ and $z \in \mathcal{S}(x)$ is the unique point in $\mathcal{S}(x)$ such that $u(z) = s$,

ii) F is Lipschitz on the set $G(\text{int}T_{sij})$,

iii) $F(G(x)) = x$ for each $x \in \text{int}T_{sij}$,

iv) if a leaf $\mathcal{S} \subset T_{sij}$ intersects S_s^i at a point z, then each interior point $x \in \text{int}\mathcal{S}$ of the leaf satisfies

$$G(x) = (w(z), u(x) - u(z)), \tag{2.4.1}$$

where $w\colon \mathbb{R}^n \to \mathbb{R}^{n-m}$ is the map from Lemma 2.4.2.

Proof. Lemma 2.2.11 shows that the relative interiors of leaves do not intersect any other leaf. Moreover u is an isometry on each leaf. Therefore, every point $x \in \mathrm{int} T_{sij}$ belongs to a unique leaf and each leaf in the cluster T_{sij} intersects the level set S_s in a single point $z \in S_s^i$. It follows that (2.4.1) defines a map

$$G\colon \mathrm{int} T_{sij} \to \mathbb{R}^{n-m} \times \mathbb{R}^m,$$

on the cluster $\mathrm{int} T_{sij}$. Let $(a,b) \in G(\mathrm{int} T_{sij})$ and let v be the map parametrising S_s^i from Lemma 2.4.2. Then $v(a) \in S_s^i$ belongs to the relative interior of some leaf $\mathcal{S}$ and lies in a distance at least $1/j$ from the relative boundary of the leaf. Define

$$F(a,b) = v(a) + Du(v(a))^*(b).$$

Let $x \in \mathrm{int} T_{sij}$ belong to a leaf $\mathcal{S}$ that intersects S_s^i at a point z. Then $v(w(z)) = z$ and there exists an isometry T such that $u(x_1) - u(x_2) = T(x_1 - x_2)$ for all $x_1, x_2 \in \mathcal{S}$ and $Du(z) = TP$, where P is the orthogonal projection onto the tangent space of $\mathcal{S}$. We infer that

$$F(G(x)) = F(w(z), u(x) - u(z)) = z + PT^*T(x-z) = x.$$

We shall now prove that F is Lipschitz on $G(\mathrm{int} T_{sij})$. Define

$$\Lambda = \big\{ a \in \mathbb{R}^{n-m} \mid (a,0) \in G(\mathrm{int} T_{sij}) \big\}. \tag{2.4.2}$$

We first claim that

$$(a,b) \mapsto Du(v(a))^* b$$

is Lipschitz. Recall that $v(a) \in S_s^i$ is in a distance at least $1/j$ from the relative boundary of a leaf $\mathcal{S}$ that contains $v(a)$. Thus, by Corollary 2.2.10, we infer that for points $(a,b), (a',b') \in G(\mathrm{int} T_{sij})$ there is

$$\|Du(v(a))^* b - Du(v(a'))^* b'\|^2 \leq \frac{j^2}{2}\|v(a) - v(a')\|^2 + \|b - b'\|^2 \leq \frac{1}{2} C^2 j^2 \|a - a'\|^2 + \|b - b'\|^2,$$

where C is the Lipschitz constant of v. It follows immediately that F is Lipschitz on $G(\mathrm{int} T_{sij})$.

It remains to prove assertion i) of the lemma. Let $\lambda > 0$. Let now $x, x' \in T_{sij}^{\lambda}$ belong to the leaves $\mathcal{S}$ and $\mathcal{S}'$ respectively. By the definition (2.4.1) and by Lipschitzness of w to prove that G is Lipschitz it is enough to show that

$$\|z - z'\| \leq C\|x - x'\|$$

for some constant C. Note that

$$z = x + Du(x)^*(u(z) - u(x)) \text{ and } z' = x' + Du(x')^*(u(z') - u(x')).$$

Thus

$$\|z - z'\| \leq \|x - x'\| + \Big\| Du(x)^*(u(z) - u(x)) - Du(x')^*(u(z') - u(x')) \Big\|.$$

Now, by Corollary 2.2.10, taking into account that $u(z) = s = u(z')$, we see that

$$\Big\| Du(x)^*(u(z) - u(x)) - Du(x')^*(u(z') - u(x')) \Big\|^2 \leq \frac{1}{2\lambda^2}\|x - x'\|^2 + \|u(x) - u(x')\|^2.$$

Therefore

$$\|z - z'\| \leq \|x - x'\| \left(1 + \sqrt{1 + \frac{1}{2\lambda^2}}\right).$$

This concludes the proof that G is Lipschitz on T^{λ}_{sij} and completes the proof of the theorem.

□

2.5 Measurability

Below $G_{n,k}$ denotes the set of all k-dimensional subspaces of $\mathbb{R}^n$. For $V \in G_{n,k}$ we denote by $O_m(V)$ the set of all isometries on V with values in $\mathbb{R}^m$, i.e. the set of all linear maps $T \colon V \to \mathbb{R}^m$ such that

$$\|T(x) - T(y)\| = \|x - y\| \text{ for all } x, y \in V.$$

By $P_V \colon \mathbb{R}^n \to \mathbb{R}^n$ we denote the orthogonal projection onto V. Then $G_{n,k}$ is a compact if equipped with the metric d given by the formula

$$d(V, V') = \|P_V - P_{V'}\|, \; V, V' \in G_{n,k}.$$

Here $\|\cdot\|$ denotes the operator norm with respect to the Euclidean norm on $\mathbb{R}^n$.

For a point $x \in \mathbb{R}^n$ and a real number $r > 0$ we shall denote by $B(x, r)$ the closed ball centred at x of radius r.

Definition 2.5.1. For $k \in \{1, \ldots, m\}$ we define $\alpha_k \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ by the formula

$$\alpha_k(x) = \sup \Big\{ r \geq 0 \mid \exists_{V \in G_{n,k}} \exists_{T \in O_m(V)} \forall_{y \in (x+V) \cap B(x,r)} \quad u(x) - u(y) = T(x - y) \Big\}$$

for $x \in \mathbb{R}^n$. We define $\alpha_{m+1} \colon \mathbb{R}^n \to \mathbb{R}$ by $\alpha_{m+1}(x) = 0$ for all $x \in \mathbb{R}^n$.

The value of function $\alpha_k(x)$ denotes the greatest radius of a ball such that u is isometric on the intersection of the ball with some k-dimensional subspace.

Lemma 2.5.2. *For any $k \in \{1, \ldots, m\}$ the functions $\alpha_k \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ are upper semi-continuous.*

Proof. Fix $k \in \{1, \dots, m\}$. Pick $x_0 \in \mathbb{R}^n$ and a sequence $(x_l)_{l=1}^\infty$ that converges to x_0 such that there exists a limit

$$\lambda = \lim_{l \to \infty} \alpha_k(x_l).$$

We need to show that $\lambda \leq \alpha_k(x_0)$. Suppose first that $\lambda < \infty$. We may assume that $\alpha_k(x_l) \in \mathbb{R}$ for each $l \in \mathbb{N}$. From the definition of $\alpha_k(x_l)$ it follows that there exist

$$V_l \in G_{n,k} \text{ and } T_l \in O_m(V_l)$$

such that for all $y \in (x_l + V_l) \cap B\Big(x_l, (1 - \frac{1}{l})\alpha_k(x_l)\Big)$ we have

$$u(x_l) - u(y) = T_l(x_l - y).$$

By compactness of $G_{n,k}$ we may assume that the sequence $(V_l)_{l=1}^\infty$ is convergent to some $V_0 \in G_{n,k}$. Moreover, we may assume that

$$(T_l P_{V_l})_{l=1}^\infty \text{ converges to } T_0 P_{V_0},$$

where $T_0 \in O_m(V_0)$. Indeed, we may assume that there exists S_0 such that $(T_l P_{V_l})_{l=1}^\infty$ converges to S_0. For $v_0 \in V_0$ we have

$$\|v_0\| = \lim_{l \to \infty} \|P_{V_l} v_0\| = \lim_{l \to \infty} \|T_l P_{V_l} v_0\| = \|S_0 v_0\|.$$

This is to say, S_0 is an isometry on V_0. Setting $T_0 = S_0 P_{V_0}$ proves the claim.

Choose now any $v_0 \in V_0$ of norm $\|v_0\| < \lambda$. By the definition of metric on $G_{n,k}$, the sequence $(P_{V_l} v_0)_{l=1}^\infty$ converges to v_0. Moreover, for sufficiently large l,

$$x_l + P_{V_l} v_0 \in (x_l + V_l) \cap B\Big(x_l, (1 - \frac{1}{l})\alpha_k(x_l)\Big).$$

Thus

$$u(x_l) - u(x_l + P_{V_l} v_0) = -T_l P_{V_l} v_0.$$

Passing to the limits we obtain that

$$u(x_0) - u(x_0 + v_0) = -T_0 v_0.$$

It follows that $\lambda \leq \alpha_k(x_0)$. Thus, the proof is complete provided λ is finite.

Suppose now that λ is infinite. Assume again that $\alpha_k(x_l) \in \mathbb{R}$ for each $l \in \mathbb{N}$ and that $(\alpha_k(x_l))_{l=1}^\infty$ converges to infinity monotonically. Then there exist $V_l \in G_{n,k}$ and T_l as above, i.e. such that $(V_l)_{l=1}^\infty$ converges to V_0 and $(T_l P_{V_l})_{l=1}^\infty$ converges to $T_0 P_{V_0}$, $T_0 \in O_m(V_0)$. Taking any $v_0 \in V_0$ of norm at most $l \in \mathbb{N}$ we may show that

$$u(x_0) - u(x_0 + v_0) = -T_0 v_0.$$

Hence $\alpha_k(x_0) \geq l$ for each $l \in \mathbb{N}$ and thus $\alpha_k(x_0) = \infty$. □

Below we shall denote the unit ball centred at the origin by $B = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$. For $r \geq 0$ we denote by $C_{n,k}(r)$ the set of all k-dimensional convex cones C in $\mathbb{R}^n$ such that there exist $c_1, \dots, c_k \in C \cap B$ such that the $n \times k$ matrix D with columns $c_1, \dots, c_k$ satisfies $\det D^* D \geq r$, i.e. their Gram matrix has determinant at least r. For a cone C we denote by V_C its linear span.

Definition 2.5.3. For $k \in \{1, \dots, m\}$ we define $\beta_k \colon \mathbb{R}^n \to \mathbb{R}$ by the formula

$$\beta_k(x) = \sup \Big\{ r \geq 0 \mid \exists_{C \in C_{n,k}(r)} \exists_{T \in O_m(V_C)} \forall_{y \in (x+C) \cap B(x,r)} \quad u(x) - u(y) = T(x-y) \Big\},$$

where $x \in \mathbb{R}^n$. Let $\beta_{m+1} \colon \mathbb{R}^n \to \mathbb{R}$ be defined by $\beta_{m+1}(x) = 0$ for all $x \in \mathbb{R}^n$.

The functions $\beta_k(x)$ indicate the maximal radius r such that there is a convex cone C, of size in its linear span bounded from below, such that u is isometric on the intersection of the ball centred at x of radius r with the shifted cone $x + C$.

Lemma 2.5.4. *For any $k \in \{1, \dots, m\}$ the function $\beta_k \colon \mathbb{R}^n \to \mathbb{R}$ is upper semicontinuous.*

Proof. Fix $k \in \{1, \dots, m\}$. Pick $x_0 \in \mathbb{R}^n$ and a sequence $(x_l)_{l=1}^{\infty}$ that converges to x_0 and such that there exists a limit

$$\lambda = \lim_{l \to \infty} \beta_k(x_l).$$

We need to show that $\lambda \leq \beta_k(x_0)$. Observe that $\lambda < \infty$, as the determinant of a Gram matrix of vectors in the unit ball is bounded above by the volume of the k-dimensional unit ball. It follows from the definition of $\beta_k(x_l)$ that there exist

$$C_l \in C_{n,k}\Big(\big(1 - \frac{1}{l}\big) \beta_k(x_l) \Big) \text{ and } T_l \in O_m(V_{C_l})$$

such that for all $y \in (x_l + C_l) \cap B\Big(x_l, \big(1 - \frac{1}{l}\big)\beta_k(x_l)\Big)$

$$u(x_l) - u(y) = T_l(x_l - y).$$

For each l pick points $(c_j^l)_{j=1}^k$ in $C_l \cap B$ such that their Gram matrix has determinant at least $(1 - 1/l)\beta_k(x_l)$. Passing to subsequences, we may assume that the sequences $(c_j^l)_{l=1}^{\infty}$ converge to some points c_j, for $j = 1, \dots, k$, for which the determinant of the Gram matrix is at least λ. Let C_0 be the convex cone in $\mathbb{R}^n$ spanned by $(c_j)_{j=1}^k$, that is

$$C_0 = \Big\{ \sum_{j=1}^k \lambda_j c_j \mid \lambda_j \geq 0 \text{ for } j = 1, \dots, k \Big\}.$$

Clearly, C_0 has dimension equal to k. It follows that $C_0 \in C_{n,k}(\lambda)$.

Passing to a subsequence, we may assume that $(V_{C_l})_{l=1}^{\infty}$ converges to some $V_0 \in G_{n,k}$. We claim that $V_0 = V_{C_0}$. Choose any $v_0 \in V_{C_0}$. Then there exist real numbers $(\lambda_j)_{j=1}^{k}$ such that

$$v_0 = \sum_{j=1}^{k} \lambda_j c_j.$$

For $l \in \mathbb{N}$ set $v_l = \sum_{j=1}^{k} \lambda_j c_j^l$. Then $(v_l)_{l=1}^{\infty}$ converge to v_0 and $v_l \in V_{C_l}$. Hence

$$v_0 = \lim_{l\to\infty} v_l = \lim_{l\to\infty} P_{V_{C_l}} v_l = P_{V_0} v_0.$$

Thus $V_{C_0} \subset V_0$, and the claim follows, as dimension of V_{C_0} is equal to k.

As in Lemma 2.5.2 we show that there exists $T_0 \in O_m(V_{C_0})$ such that

$$(T_l P_{V_{C_l}})_{l=1}^{\infty} \text{ converges to } T_0 P_{V_{C_0}}.$$

Take $\epsilon > 0$ and choose any $y_0 \in (x_0 + C_0) \cap B(x_0, (1-\epsilon)\lambda)$. Then there exist $(\lambda_j)_{j=1}^{k}$ such that

$$y_0 = x_0 + \sum_{j=1}^{k} \lambda_j c_j.$$

Set $y_l = x_l + \sum_{j=1}^{k} \lambda_j c_j^l$. Then $(y_l)_{l=1}^{\infty}$ converges to y_0 and for sufficiently large l,

$$y_l \in (x_l + C_l) \cap B\Big(x_l, (1-\frac{1}{l})\beta_k(x_l)\Big).$$

For such l we have $u(x_l) - u(y_l) = T_l(x_l - y_l)$. It follows that also $u(x_0) - u(y_0) = T_0(x_0 - y_0)$. That is, $\beta_k(x_0) \geq (1-\epsilon)\lambda$ for any $\epsilon > 0$. The proof is complete. □

Lemma 2.5.5. *A point $x \in \mathbb{R}^n$ belongs to a leaf $\mathcal{S}$ of u of dimension at least k if and only if $\beta_k(x) > 0$. A point $x \in \mathbb{R}^n$ belongs to a leaf $\mathcal{S}$ of u of dimension exactly k if and only if $\beta_k(x) > 0$ and $\beta_{k+1}(x) = 0$.*

Proof. Suppose that $x_0 \in \mathbb{R}^n$ belongs to a leaf $\mathcal{S}$ of u of dimension $l \in \{k, \dots, m\}$. Let V denote the tangent space of $\mathcal{S}$. Choose a point $x_1 \in \mathrm{int}\mathcal{S}$ and $\epsilon_0 > 0$ so that the intersection $B(x_1, \epsilon_0) \cap (x_1 + V)$ is contained in $\mathcal{S}$. For $\epsilon \in (0, \epsilon_0)$ let

$$C = \big\{x \in \mathbb{R}^n \mid x = \lambda(x_2 - x_0) \text{ for some } \lambda \geq 0, x_2 \in B(x_1, \epsilon) \cap (x_1 + V)\big\}.$$

Then C is a convex cone of dimension l containing the origin. Thus, it contains k linearly independent vectors, which have Gram matrix of non-zero determinant. This is to say, the intersection of C with the linear span of these vectors belongs to $C_{n,k}(\epsilon)$ for $\epsilon > 0$ sufficiently small. Moreover, by convexity of $\mathcal{S}$, u is isometric on the set $(x_0 + C) \cap B(x_0, \epsilon)$, if $\epsilon > 0$ is sufficiently small. Therefore $\beta_l(x_0) > \epsilon > 0$, whenever ϵ satisfies the two upper bounds.

Conversely, suppose that $\beta_k(x_0) > 0$. Then there exist

$$r > 0, \text{ a cone } C \in C_{n,k}(r) \text{ and an isometry } T \in O_m(V_C)$$

such that

$$u(x_0) - u(y) = T(x - y) \text{ for all } y \in (x_0 + C) \cap B(x_0, r).$$

With use of the Kuratowski–Zorn lemma choose a leaf $\mathcal{S}$ of u containing $(x_0 + \mathcal{C}) \cap B(x_0, \epsilon)$. Then the dimension of $\mathcal{S}$ is at least k.

The second assertion is a trivial consequence of the first assertion. □

Lemma 2.5.6. *A point $x \in \mathbb{R}^n$ belongs to relative interior of a leaf $\mathcal{S}$ of u of dimension k if and only if $\alpha_k(x) > 0$ and $\beta_{k+1}(x) = 0$.*

Proof. Suppose that x_0 belongs to the relative interior of a leaf $\mathcal{S}$ of u of dimension k. By the previous lemma $\beta_k(x_0) > 0$ and $\beta_{k+1}(x_0) = 0$. Let V denote the tangent space of $\mathcal{S}$. Then, as x_0 is in the relative interior, there exist $\epsilon > 0$, $T \in O_m(V)$ such that

$$u(x_0) - u(y) = T(x_0 - y) \text{ for all } y \in (x_0 + V) \cap B(x_0, \epsilon).$$

That is $\alpha_k(x_0) \geq \epsilon > 0$.

Conversely, suppose that $\alpha_k(x_0) > 0$ and $\beta_{k+1}(x_0) = 0$. Then there exist $V \in G_{n,k}$ and $T \in O_m(V)$ such that

$$u(x_0) - u(y) = T(x_0 - y) \text{ for all } y \in (x_0 + V) \cap B(x_0, \epsilon).$$

It follows from the Kuratowski–Zorn lemma that x_0 belongs to a leaf $\mathcal{S}$ of u that contains $(x_0 + V) \cap B(x_0, \epsilon)$. As $\beta_{k+1}(x_0) = 0$, this leaf is of dimension k and thus x_0 belongs to the relative interior of $\mathcal{S}$. □

Corollary 2.5.7. *Let $k \in \{0, \ldots, m\}$. Then the union of all leaves of u of dimension k is a Borel set. Moreover, the union of all relative interiors of leaves of u of dimension k is a Borel set and so is the union of all relative boundaries of leaves of u of dimension k.*

Proof. The proof readily follows by Lemma 2.5.2, Lemma 2.5.4 and Lemma 2.5.5, Lemma 2.5.6. □

Note that whenever $\alpha_k(x) > 0$ and $\beta_{k+1}(x) = 0$, then Lemma 2.5.6 tells us that x belongs to the relative interior of a leaf of u. This leaf in unique, by Lemma 2.2.11. We shall denote it by $\mathcal{S}(x)$.

Below we adapt a convention that $\inf \emptyset = \infty$.

Definition 2.5.8. Let $k \in \{0, \ldots, m\}$. We define $\gamma_k \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ by the formula

$$\gamma_k(x, y) = \inf \Big\{ t > 0 \mid y \in t\big(u(\mathcal{S}(x)) - u(x)\big) \Big\}$$

for $x \in \mathbb{R}^n$ such that $\alpha_k(x) > 0$ and $\beta_{k+1}(x) = 0$ and

$$\gamma_k(x, y) = \infty$$

otherwise.

Lemma 2.5.9. *For any $k \in \{0, \ldots, m\}$ the function γ_k is Borel measurable.*

Proof. As α_k and β_{k+1} are Borel measurable, it is enough to show that the function γ_k is Borel measurable on

$$A_k = \big\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid \alpha_k(x) > 0 \text{ and } \beta_{k+1}(x) = 0 \big\}.$$

Observe that γ_k is a limit, as ρ converges to infinity, of functions

$$\gamma_{k,\rho}(x, y) = \inf \Big\{ t > 0 \mid y \in t\big(u(\mathcal{S}(x)) - u(x)\big), \|y\| \leq t\rho \Big\}.$$

We claim that $\gamma_{k,\rho}$ is lower semicontinuous on A_k. This will yield the asserted measurability.

Indeed, let $(x_l, y_l)_{l=1}^{\infty}$ be a sequence in A_k such that there exists $(x_0, y_0) \in A_k$ with

$$(x_0, y_0) = \lim_{l \to \infty} (x_l, y_l) \text{ and such that there exists } \lim_{l \to \infty} \gamma_{k,\rho}(x_l, y_l) = \lambda.$$

We shall show that $\gamma_{k,\rho}(x_0, y_0) \leq \lambda$. If $\lambda = \infty$, then there is nothing to prove. Otherwise, there exist sequences $(z_l)_{l=1}^{\infty}$ in $\mathbb{R}^n$ and $(t_l)_{l=1}^{\infty}$ in $\mathbb{R}$ such that

$$y_l = t_l\big(u(z_l) - u(x_l)\big) \text{ and } \|y_l\| \leq t_l \rho, \text{ where } z_l \in \mathcal{S}(x_l) \text{ and } 0 < t_l < \gamma_k(x_l, y_l) + 1/l. \quad (2.5.1)$$

Observe that

$$\|z_l - x_l\| = \|u(z_l) - u(x_l)\| = \frac{\|y_l\|}{t_l} \leq \rho$$

Thus, passing to a subsequence, we may assume that $(z_l)_{l=1}^{\infty}$ converges to some $z_0 \in \mathcal{S}(x_0)$ and that $(t_l)_{l=1}^{\infty}$ converges to some $t_0 \geq 0$. Taking limits in (2.5.1) we see that

$$y_0 = t_0\big(u(z_0) - u(x_0)\big) \text{ with } z_0 \in \mathcal{S}(x_0) \text{ and } 0 \leq t_0 \leq \lambda.$$

Hence

$$y_0 \in t_0\big(u(\mathcal{S}(x_0)) - u(x_0)\big) \text{ and } \|y_0\| \leq t_0 \rho.$$

This is to say, $\gamma_{k,\rho}(x_0, y_0) \leq t_0 \leq \lambda$. The proof is complete. $\square$

Definition 2.5.10. For a convex set $K \subset \mathbb{R}^m$, such that $0 \in \text{int}K$, we define its *Minkowski functional* $\|\cdot\|_K : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ by the formula

$$\|y\|_K = \inf \big\{t > 0 \mid y \in tK\big\}.$$

Proposition 2.5.11. *Let $K \subset \mathbb{R}^m$ be a closed, convex set that contains the origin in its interior. A point $y \in \mathbb{R}^m$ belongs to the interior of K if and only if $\|y\|_K < 1$.*

Moreover, a point $y \in \mathbb{R}^m$ belongs to the boundary of K if and only if $\|y\|_K = 1$.

Proof. If $y \in \text{int}K$ then, as $0 + y = y \in \text{int}K$, it follows by the continuity of addition that $y + B(0, \epsilon) \subset \text{int}K$ for $\epsilon > 0$ sufficiently small. Observe that $\|y/s\| \leq \epsilon$ if $s \geq \|y\|/\epsilon$ and thus for large $s > 0$

$$(1 + 1/s)y \in K.$$

Hence $\|y\|_K \leq \frac{s}{s+1} < 1$.

Conversely, suppose that $\|y\|_K < 1$. Then $y \in tK$ for some $t < 1$. As $0 \in \text{int}K$, there exists $\epsilon > 0$ such that if $\|w\| \leq \epsilon$, then $w \in K$. Hence, if $\|z\| \leq \epsilon(1 - t)$, then

$$y + z \in tK + (1 - t)K = K,$$

by the convexity of K.

Suppose that $y \in \partial K$. Then clearly $\|y\|_K \leq 1$ and, by the above, $\|y\|_K \geq 1$.

Conversely, let $\|y\|_K = 1$. Then there exists a sequence of positive numbers $(t_l)_{l=1}^{\infty}$ converging to zero and a sequence $(x_l)_{l=1}^{\infty}$ in K such that

$$y = (1 + t_l)x_l.$$

It follows that $(x_l)_{l=1}^{\infty}$ is bounded. Taking a convergent subsequence from $(x_l)_{l=1}^{\infty}$ we see that $y = x_0$ for some $x_0 \in K$. By the first assertion, $y \notin \text{int}K$. Hence $y \in \partial K$. □

Lemma 2.5.12. *If $x \in \mathbb{R}^n$ belongs to relative interior of a leaf $\mathcal{S}$ of u of dimension k, then $\gamma_k(x, \cdot)$ is the Minkowski functional a closed, convex set $u(\mathcal{S}) - u(x)$. If $x \in \mathbb{R}^n$ does not belong to relative interior of any leaf of dimension k, then*

$$\gamma_k(x, \cdot) = \infty.$$

Proof. Suppose that $x \in \mathbb{R}^n$ does not belong to relative interior of a leaf of u of dimension k. Then Lemma 2.5.6 and Definition 2.5.8 tells us that $\gamma_k(x, \cdot) = \infty$.

Let now $x \in \text{int}\mathcal{S}$, where $\mathcal{S}$ is a k-dimensional leaf. By Lemma 2.2.11, such leaf $\mathcal{S}$ is unique. The assertion of the lemma follows readily from the definitions. □

Definition 2.5.13. Let $k \in \{0, \dots, m\}$. We shall denote by T_k the union of all k-dimensional leaves of u, by $\mathrm{int}T_k$ the union of all relative interiors of all k-dimensional leaves of u and by ∂T_k the union of all relative boundaries of all k-dimensional leaves of u.

Lemma 2.5.14. *For each $s \in S$ and each $i, j \in \mathbb{N}$ the cluster $\mathrm{int}T_{sij}$ and its image $G(\mathrm{int}T_{sij})$ are Borel sets. Moreover ∂T_m is a Borel set of Lebesgue measure zero.*

Proof. Fix $s \in S$ and $i, j \in \mathbb{N}$. Recall the Borel set $S_s^i \subset \mathbb{R}^n$ and Lipschitz mapping $w \colon \mathbb{R}^n \to \mathbb{R}^{n-m}$ from Lemma 2.4.2. Since w is injective on S_s^i it follows from [47, 2.2.10] that $w(S_s^i)$ is a Borel subset of $\mathbb{R}^{n-m}$. Moreover, the set Λ, defined in (2.4.2), is given by

$$\Lambda = \left\{ a \in w(S_s^i) \mid \alpha_m(w^{-1}(a)) > 1/j \right\} \tag{2.5.2}$$

as follows by the definition (2.4.1) and Lemma 2.4.2. Clearly, Λ is a Borel set. Definition of the cluster $\mathrm{int}T_{sij}$ implies that

$$G(\mathrm{int}T_{sij}) = \left\{ (a,b) \in \mathbb{R}^{n-m} \times \mathbb{R}^m \mid a \in \Lambda, b \in u\Big(\mathrm{int}\mathcal{S}\big(v(a)\big)\Big) - u\big(v(a)\big) \right\}.$$

Here $\mathcal{S}(v(a))$ is the unique m-dimensional leaf of u containing $v(a)$. Observe that Proposition 2.5.11 and Lemma 2.5.12 tells us that if $a \in \Lambda$, then b belongs to the interior of

$$u(\mathcal{S}(v(a))) - u(v(a)) \text{ if and only if } \gamma_m(v(a), b) < 1.$$

This is to say,

$$G(\mathrm{int}T_{sij}) = \left\{ (a,b) \in \mathbb{R}^{n-m} \times \mathbb{R}^m \mid a \in \Lambda, \gamma_m(v(a), b) < 1 \right\}. \tag{2.5.3}$$

As γ_m is Borel measurable, it follows that $G(\mathrm{int}T_{sij})$ is a Borel set.

Lemma 2.4.5 shows that F, the inverse of G on its image, is well-defined, injective and Lipschitz on $G(\mathrm{int}T_{sij})$. Moreover

$$\mathrm{int}T_{sij} = F(G(\mathrm{int}T_{sij})).$$

Using [47, 2.2.10], we see that $\mathrm{int}T_{sij}$ is a Borel set.

We shall show that ∂T_m has Lebesque measure zero. Recall that Corollary 2.5.7 tells us that ∂T_m is a Borel set. Consider the set

$$B = \left\{ (a,b) \in \mathbb{R}^{n-m} \times \mathbb{R}^m \mid a \in \Lambda, \gamma_m(v(a), b) = 1 \right\}.$$

By Fubini's theorem, $\lambda(B) = 0$, as boundaries of convex sets have Lebesgue measure zero.

Recall that F is a Lipschitz map on $G(\mathrm{int}T_{sij})$. Using the Kirszbraun theorem (see e.g [68, 94]) we extend F, defined on $G(\mathrm{int}T_{sij})$, to a Lipschitz map $\tilde{F}$ on $\mathbb{R}^{n-m} \times \mathbb{R}^m$. We claim that for any such extension

$$\tilde{F}(B) \supset \partial T_m. \tag{2.5.4}$$

Indeed, let $x \in \partial T_m$. There exists a leaf $\mathcal{S} \subset T_{sij}$ of u and a sequence $(x_l)_{l=1}^{\infty}$ in int$\mathcal{S}$ that converges to x. Let $\tilde{G}$ be a Lipschitz extension of G to $\mathbb{R}^n$. The sequence $(G(x_l))_{l=1}^{\infty}$ converges to $\tilde{G}(x) = (a,b) \in \mathbb{R}^{n-m} \times \mathbb{R}^m$. We claim that $(a,b) \in B$. This follows by the continuity of γ_m in the second variable. Now, $x = \tilde{F}(\tilde{G}(x)) \in \tilde{F}(B)$ and (2.5.4) is proven.

Therefore we can use $\lambda(B) = 0$ and the fact that images under Lipschitz maps of sets of Lebesgue measure zero have Lebesgue measure zero (see [47, 3.2.3]), to infer that $\lambda(\partial T_m \cap T_{sij}) = 0$ and hence $\partial T_m \cap T_{sij}$ is Lebesgue measurable. By Lemma 2.4.4 the sets T_{sij} form a countable cover of ∂T_m. It follows that $\lambda(\partial T_m) = 0$. This concludes the proof. □

Corollary 2.5.15. *For any $s \in S$, $i, j \in \mathbb{N}$, the set T_{sij} is Lebesgue measurable.*

Proof. T_{sij} is a union of a Borel set intT_{sij} and a set $\partial T_m \cap T_{sij}$ of Lebesgue measure zero. □

Remark 2.5.16. The clusters T_{sij} may be taken to be disjoint. Indeed, let $(T_k)_{k=1}^{\infty}$ be a renumbering of the set of clusters. Set for $l \in \mathbb{N}$

$$T'_l = T_l \setminus \bigcup_{j=1}^{l-1} T_j$$

and

$$\mathrm{int}T'_l = \mathrm{int}T_l \setminus \bigcup_{j=1}^{l-1} \mathrm{int}T_j.$$

Note that the structure of the clusters T'_{sij} remains the same. For each T_{sij} there exists a Borel subset $S_{sij} = T_{sij} \cap S^i_s$ of $S^i_s \subset \mathbb{R}^n$ on which there are Lipschitz maps

$$w \colon \mathbb{R}^n \to \mathbb{R}^{n-m} \text{ and } v \colon \mathbb{R}^{n-m} \to \mathbb{R}^n$$

such that

$$v(w(x)) = x \text{ for all } x \in S_{sij}$$

Indeed, the new cluster is a subset of the old one, so the former maps suffice. From the modification procedure it follows also that Lemma 2.4.4 still holds true. Moreover, the leaf $\mathcal{S}$ corresponding to a point $z \in S_{sij}$ satisfies

$$\mathrm{dist}(z, \partial\mathcal{S}) > 1/j.$$

Also the assertions of Lemma 2.4.5 hold true with the old maps and so does the assertions of Lemma 2.5.14, as follows from the modification procedure.

2.6 Disintegration with respect to partition

Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map with respect to the Euclidean norms. In the previous sections we have associated to u a partitioning of $\mathbb{R}^n$, up to a set of Lebesgue measure zero, into maximal sets $\mathcal{S}$ on which u is an isometry. It was conjectured by Klartag in [70, Chapter 6] that given a measure μ, such that $(\mathbb{R}^n, \|\cdot\|, \mu)$ is a weighted Riemannian manifold satisfying the curvature-dimension condition $CD(\kappa, N)$ (see Section 2.7), then μ may be decomposed into a mixture of measures $\mu_{\mathcal{S}}$, each supported on a leaf $\mathcal{S}$ of u, such that $(\mathrm{int}\mathcal{S}, \|\cdot\|, \mu_{\mathcal{S}})$ is a weighted Riemannian manifold that satisfies $CD(\kappa, N)$.

Below we denote by $CC(\mathbb{R}^n)$ the space of closed, convex, non-empty subsets of $\mathbb{R}^n$. It is a closed subspace of $CL(\mathbb{R}^n)$ – the space of closed non-empty subsets of $\mathbb{R}^n$ equipped with the Wijsman topology (see [106]). The Wijsman topology is the weakest topology such that for any $x \in \mathbb{R}^n$ function

$$A \mapsto \mathrm{dist}(x, A)$$

is continuous. By a result of Beer (see [17]), the space $CL(\mathbb{R}^n)$, equipped with this topology, is Polish. Hence so is $CC(\mathbb{R}^n)$.

Let us recall that $B(u)$ denotes the set of points in $\mathbb{R}^n$ that belong to at least two distinct leaves of u. By Corollary 2.2.15, it is contained in the Borel set $N(u)$ of points at which u is not differentiable. The latter is of Lebesgue measure zero. We define a map

$$\mathcal{S}\colon \mathbb{R}^n \to CC(\mathbb{R}^n)$$

in such a way that for $x \in \mathbb{R}^n \setminus N(u)$ the set $\mathcal{S}(x)$ is the unique leaf of u containing x and for $x \in N(u)$ we put $\mathcal{S}(x) = \{x\}$.

The aim of this section is to prove the following disintegration theorem, which is a step towards the conjecture.

Theorem 2.6.1. *Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map with respect to the Euclidean norms. Then there exists a Borel measure ν on $CC(\mathbb{R}^n)$, supported on the set of leaves of u, and Borel measures $\lambda_{\mathcal{S}}$ such that*

i) for every Borel set $A \subset \mathbb{R}^n$ the function $\mathcal{S} \mapsto \lambda_{\mathcal{S}}(A)$ is Borel measurable,

ii) for ν-almost every leaf $\mathcal{S}$ the measure $\lambda_{\mathcal{S}}$ is concentrated on $\mathcal{S}$,

iii) for every Borel set $A \subset \mathbb{R}^n$

$$\lambda(A) = \int_{CC(\mathbb{R}^n} \lambda_{\mathcal{S}}(A) d\nu(\mathcal{S}).$$

Let X be a measurable space. In [18] it is proven that a map $f\colon X \to CL(\mathbb{R}^n)$ is measurable if and only if it is measurable as a multifunction. The latter is defined by the condition that for any open set $U \subset \mathbb{R}^n$ the set

$$\{x \in X \mid f(x) \cap U \neq \emptyset\}$$

is measurable in X.

Let us recall a theorem taken from [25, Example 10.4.11].

Theorem 2.6.2. *Let X, Y be two Polish spaces. Let $\pi\colon X \to Y$ be a Borel map and let μ be a non-negative Borel measure on X. Let ν be the push-forward of measure μ via π.Then there exist Borel measures $(\mu_y)_{y\in Y}$ on X such that*

i) for every Borel set $B \subset X$ the function $y \mapsto \mu_y(B)$ is Borel measurable,

ii) for ν-almost every $y \in \pi(X)$ the measure μ_y is concentrated on $\pi^{-1}(y)$,

iii) for every Borel sets $B \subset X$ and $E \subset Y$ there is

$$\mu(B \cap \pi^{-1}(E)) = \int_E \mu_y(B) d\nu(y).$$

Proof of Theorem 2.6.1. We have a well-defined map $\mathcal{S}\colon \mathbb{R}^n \to CC(\mathbb{R}^n)$ that assigns to any $x \in \mathbb{R}^n \setminus N(u)$ a unique leaf $\mathcal{S}(x)$ that contains x and for $x \in N(u)$ we set $\mathcal{S}(x) = \{x\}$. We would like to prove that $\mathcal{S}\colon \mathbb{R}^n \to CC(\mathbb{R}^n)$ is Borel measurable with respect to the Wijsman topology on $CC(\mathbb{R}^n)$, which is equivalent to its measurablity as a multifunction. Note that for any compact set $K \subset \mathbb{R}^n$ the set $A_K = \{x \in \mathbb{R}^n \mid \mathcal{S}(x) \cap K \neq \emptyset\}$ is equal to

$$\Big\{x \in \mathbb{R}^n \setminus (K \cup N(u)) \mid \sup\Big\{\frac{\|u(x) - u(y)\|}{\|x - y\|} \mid y \in K\Big\} = 1\Big\} \cup K.$$

Observe that the function

$$x \mapsto \sup\Big\{\frac{\|u(x) - u(y)\|}{\|x - y\|} \mid y \in K\Big\}$$

is lower semicontinuous. Hence A_K is a Borel set. As any open set $U \subset \mathbb{R}^n$ is a countable union of compact sets, it follows that the map $\mathcal{S}$ is Borel measurable.

Recall that $CC(\mathbb{R}^n)$ and $\mathbb{R}^n$ are Polish spaces. We apply Theorem 2.6.2 to the map $\mathcal{S}$. Since the set $N(u)$ has Lebesgue measure zero we obtain the desired disintegration. $\square$

2.7 Curvature-dimension condition

Suppose that we are given a measure μ on $\mathbb{R}^n$ such that $(\mathbb{R}^n, \|\cdot\|, \mu)$ is a weighted Riemannian manifold satisfying $CD(\kappa, N)$. We shall investigate the behaviour of the conditional measures of μ, with respect to the partition introduced in Section 2.2. We shall concentrate on the leaves of maximal dimension.

In the current section we recall the notion of the curvature-dimension condition $CD(\kappa, n)$. We shall say that an n-dimensional Riemannian manifold $\mathcal{M}$ satisfies the $CD(\kappa, n)$ condition provided that the Ricci tensor Ric_M is bounded below by the Riemannian metric tensor g, i.e.

$$Ric_{\mathcal{M}}(p)(v, v) \geq \kappa g(p)(v, v) \text{ for any } p \in \mathcal{M} \text{ and any } v \in T_p\mathcal{M}.$$

We shall study weighted Riemannian manifolds, which are triples $(\mathcal{M}, d, \mu)$, where d is the Riemannian metric on $\mathcal{M}$ and μ is a measure on $\mathcal{M}$ with smooth positive density $e^{-\rho}$ with respect to the Riemannian volume. The generalised Ricci tensor of the weighted Riemannian manifold is defined by the formula

$$Ric_\mu = Ric_{\mathcal{M}} + D^2\rho,$$

where $D^2\rho$ is the Hessian of smooth function ρ. The generalised Ricci tensor – or the N-Bakry-Émery tensor – with parameter $N \in (-\infty, 1) \cup [n, \infty]$ is defined by the formula

$$Ric_{\mu,N} = \begin{cases} Ric_\mu(v, v) - \frac{D\rho(v)^2}{N-n}, & \text{if } N > n \\ Ric_\mu(v, v) & \text{if } N = \infty \\ Ric_{\mathcal{M}}(v, v) & \text{if } N = n \text{ and } \rho \text{ is constant.} \end{cases}$$

Note that if $N = n$, then ρ is required to be a constant function.

Definition 2.7.1. For $\kappa \in \mathbb{R}$ and $N \in (-\infty, 1) \cup [n, \infty]$ we say that $(\mathcal{M}, d, \mu)$ satisfies the curvature-dimension condition $CD(\kappa, N)$ if

$$Ric_{\mu,N}(p)(v, v) \geq \kappa g(p)(v, v) \text{ for all } x \in \mathcal{M} \text{ and all } v \in T_p\mathcal{M}.$$

We refer the reader to [11], [12], [13] and to [98], [99], [44], [6], [104] for background on the curvature-dimension condition. In all cases we consider in this thesis it will always hold that $Ric_{\mathcal{M}} = 0$.

The aim of the section is to prove the following theorem which partially resolves the conjecture of Klartag [70, Chapter 6] in the affirmative. In particular, if a measure μ is concentrated on leaves of u of dimension m, then the conjecture holds true for μ and u.

Let us recall that T_m denotes the union of leaves of dimension m. This is a Borel set by Corollary 2.5.7.

We shall denote by $T^m \subset CC(\mathbb{R}^n)$ the set of leaves of u of dimension m.

Theorem 2.7.2. *Let $m \leq n$. Let $N \in (-\infty, 1) \cup [n, \infty]$ andl let $\kappa \in \mathbb{R}$. Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map with respect to the Euclidean norms. Let μ be a Borel measure on $\mathbb{R}^n$ such that $(\mathbb{R}^n, \|\cdot\|, \mu)$ satisfies the curvature-dimension condition $CD(\kappa, N)$. Then there exists a Borel measure ν on $CC(\mathbb{R}^n)$, supported on the set T^m of leaves of dimension m, and for each leaf $\mathcal{S}$ of u of dimension m, there exists a Borel measure $\mu_{\mathcal{S}}$ such that:*

i) for every Borel set $B \subset T_m$ the function $\mathcal{S} \mapsto \mu_{\mathcal{S}}(B)$ is ν-measurable,

ii) for ν-almost every leaf $\mathcal{S}$ the measure $\mu_{\mathcal{S}}$ is concentrated on $\mathrm{int}\mathcal{S}$

iii) for ν-almost every leaf $\mathcal{S}$ the space $(\mathrm{int}\mathcal{S}, \|\cdot\|, \mu_{\mathcal{S}})$ satisfies the $CD(\kappa, N)$ condition,

iv) for every Borel set $A \subset T_m$ there is

$$\mu(A) = \int_{T^m} \mu_{\mathcal{S}}(A) d\nu(\mathcal{S}).$$

Let us note that the above theorem proves in particular that if we disintegrate the Lebesgue measure with respect to the partition obtained from u, then $(\mathrm{int}\mathcal{S}, \|\cdot\|, \mu_{\mathcal{S}})$ will satisfy the curvature-dimension condition $CD(0, n)$ for leaves $\mathcal{S}$ of dimension m. This complements the results of [5], [100]; see also [33], [34] and [24]. Note that our result tells in particular that the conditional measures are equivalent to the m-dimensional Hausdorff measure, which provides a strengthening of the previously known results.

In what follows, we shall use the notation from Section 2.5. Observe that it suffices to prove the theorem under the assumption that μ is concentrated on a single cluster T_{sij}, $s \in S$ and $i, j \in \mathbb{N}$, of leaves of u; see Remark 2.5.16. Recall the definitions of maps F and G (see Lemma 2.4.5) and a map v (see Lemma 2.4.2). Below $\mathcal{H}_m$ is the m-dimensional Hausdorff measure on $\mathbb{R}^n$.

Lemma 2.7.3. *Let $m \leq n$ and let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map. Fix $s \in S$, $i, j \in \mathbb{N}$. Then for any Borel set $A \subset T_{sij}$ there is*

$$\lambda(A) = \int_{\Lambda} \Big(\int_{\mathrm{int}\mathcal{S}(v(a))} \mathbf{1}_A J_n F \circ G d\mathcal{H}_m \Big) d\lambda(a),$$

where $J_n F$ denotes the n-dimensional Jacobian of F and

$$\Lambda = \big\{ a \in \mathbb{R}^{n-m} \mid (a, 0) \in G(\mathrm{int}T_{sij}) \big\}.$$

Moreover, the map

$$\Lambda \ni a \mapsto \int_{\mathrm{int}\mathcal{S}(v(a))} \mathbf{1}_A J_n F \circ G d\mathcal{H}_m \in \mathbb{R}$$

is λ-measurable.

Proof. By Lemma 2.4.5, the map F is a bijection of $G(\mathrm{int}T_{sij})$ and of $\mathrm{int}T_{sij}$. As F is Lipschitz on $G(\mathrm{int}T_{sij})$ we may apply the area formula [47, 3.2.5] to infer that for any measurable, non-negative $\phi\colon \mathbb{R}^n \to \mathbb{R}$

$$\int_{G(\mathrm{int}T_{sij})} \phi \circ F J_n F d\lambda = \int_{\mathrm{int}T_{sij}} \phi d\lambda. \tag{2.7.1}$$

Let

$$f = J_n F \mathbf{1}_{G(\mathrm{int}T_{sij})}.$$

Observe that f is non-negative and Borel measurable as $G(\mathrm{int}T_{sij})$ is a Borel set by Lemma 2.4.5 and the fact that images of Borel sets via Lipschitz maps are Borel.

By Tonelli's theorem, the functions $f(a,\cdot)$ are measurable for almost every $a \in \mathbb{R}^{n-m}$ and we have

$$\int_{\mathbb{R}^{n-m}\times\mathbb{R}^m} \phi \circ F f d\lambda = \int_{\mathbb{R}^{n-m}} \int_{\mathbb{R}^m} \phi(F(a,b)) f(a,b) d\lambda(b) d\lambda(a). \tag{2.7.2}$$

Observe now that $(a,b) \in G(\mathrm{int}T_{sij})$ if and only if for some $a \in \Lambda$

$$a = w(v(a)) \text{ and } b = u(x) - u(v(a)).$$

Note that F on $G(\mathrm{int}\mathcal{S}(v(a)))$ is an isometry. Therefore by a linear change of variables

$$\int_{G(\mathrm{int}\mathcal{S}(v(a)))} \phi(F(a,b)) f(a,b) d\lambda(b) = \int_{\mathrm{int}\mathcal{S}(v(a))} \phi f \circ G d\mathcal{H}_m.$$

Tonelli's theorem implies that the map

$$\Lambda \ni a \mapsto \int_{\mathrm{int}\mathcal{S}(v(a))} \phi f \circ G d\mathcal{H}_m$$

is measurable. Moreover, by (2.7.1) and by (2.7.2), for any non-negative function ϕ we have

$$\int_{\mathrm{int}T_{sij}} \phi d\lambda = \int_\Lambda \Big(\int_{\mathrm{int}\mathcal{S}(v(a))} \phi f \circ G d\mathcal{H}_m \Big) d\lambda(a).$$

By the fact that ∂T_m has Lebesgue measure zero (see Lemma 2.5.14), we see that

$$\int_{T_{sij}} \phi d\lambda = \int_\Lambda \Big(\int_{\mathrm{int}\mathcal{S}(v(a))} \phi f \circ G d\mathcal{H}_m \Big) d\lambda(a).$$

The proof is complete. □

Let us recall a lemma from [70] that we shall need in what follows.

Lemma 2.7.4. *Let $a, b \in \mathbb{R}$, $b > 0$ and $a \notin [-b, 0]$. Then*

$$\frac{x^2}{a} + \frac{y^2}{b} \geq \frac{(x-y)^2}{a+b}$$

for all $x, y \in \mathbb{R}$.

Proof. We use the inequality

$$\frac{|a|}{|b|}x^2 \pm 2xy + \frac{|b|}{|a|}y^2 \geq 0.$$

From this we see that

$$\frac{x^2}{a} + \frac{y^2}{b} - \frac{(x-y)^2}{a+b} = \frac{1}{a+b}\Big(\frac{b}{a}x^2 + 2xy + \frac{a}{b}y^2\Big) \geq 0$$

whenever $b > 0$ and $a \notin [-b, 0]$. □

Let us also recall formulae for differentiation of matrices. If $R(t) = \log|\det A(t)|$ and A is differentiable in $t \in \mathbb{R}$, then

$$\frac{dR}{dt}(s) = \operatorname{tr}\Big(A(s)^{-1}\frac{dA}{dt}(s)\Big). \tag{2.7.3}$$

Moreover

$$\frac{d^2R}{dt^2}(s) = \operatorname{tr}\Big(A(s)^{-1}\frac{d^2A}{dt^2}(s)\Big) - \operatorname{tr}\Bigg(\Big(A(s)^{-1}\frac{dA}{dt}(s)\Big)^2\Bigg). \tag{2.7.4}$$

We should also need the following version of the Whitney extension theorem (see [105] or [96]).

Theorem 2.7.5. *Let $A \subset \mathbb{R}^n$ be an arbitrary set, let $f\colon A \to \mathbb{R}$ and $V\colon A \to \mathbb{R}^n$. Suppose that there exists $M \in \mathbb{R}$ such that for all $x, y \in A$*

$$\begin{aligned}
&|f(x)| \leq M, \|V(x)\| \leq M,\\
&\|V(x) - V(y)\| \leq M\|x - y\|,\\
&|f(y) - f(x) - \langle V(x), y - x\rangle| \leq M\|x - y\|^2.
\end{aligned}$$

Then there exists a differentiable function $\tilde{f}\colon \mathbb{R}^n \to \mathbb{R}$ with locally Lipschitz derivative such that

$$\tilde{f}(x) = f(x), Df(x)(y) = \langle V(x), y\rangle \text{ for all } x \in A \text{ and all } y \in \mathbb{R}^n.$$

Proof of Theorem 2.7.2. As noted already, it is enough to prove Theorem 2.7.2 assuming that there is a single cluster of leaves T_{sij}. Thus, let us fix a cluster T_{sij}.

Note that, by Corollary 2.2.10, on T^{λ}_{sij}, Du is Lipschitz; see Lemma 2.4.5 for the definition of T^{λ}_{sij}. Moreover, by the second assertion of Lemma 2.2.6, for any $x, y \in T^{\lambda}_{sij}$ there is

$$\|u(y) - u(x) - Du(x)(y - x)\| \leq \frac{1}{\lambda}\|x - y\|^2.$$

By the Whitney extension theorem there exists a differentiable map $\tilde{u}$ with locally Lipschitz derivative on $\mathbb{R}^n$ that coincides with u on T^{λ}_{sij} and such that $D\tilde{u} = Du$ on T^{λ}_{sij}. By [70, Lemma 3.2.4], the second derivative of $\tilde{u}$ exists almost everywhere and is symmetric, in the sense that the second derivative of any of its components is symmetric. We will abuse the notation and assume that u has Lipschitz derivative, is defined on $\mathbb{R}^n$, and its second derivative is symmetric λ-almost everywhere.

Since $F\colon G(\mathrm{int}T_{sij}) \to \mathbb{R}^n$ has locally Lipschitz inverse, it follows that for λ-almost every $(a,b) \in G(\mathrm{int}T_{sij})$ there exists $D^2u(F(a,b))$ and is symmetric.

By Fubini's theorem we infer that there exists a Borel cover $(\Lambda_l)_{l=1}^{\infty}$ of Λ such that for each $l \in \mathbb{N}$ there exists b_l such that $(a,b_l) \in G(\mathrm{int}T_{sij})$ for all $a \in \Lambda_l$. Moreover for λ-almost every $a \in \Lambda_l$, there exists $D^2u(F(a,b_l))$ and it is symmetric. Note that for $a \in \Lambda_l$

$$u(F(a,b_l)) = u(v(a)) + b_l = s + b_l.$$

Hence, on the level set of u corresponding to $s+b_l$, there exists D^2u and it is symmetric. Therefore, without loss of generality, passing to a refinement of initial cover and modifying the clusters T_{sij}, we assume that $D^2u(v(a))$ exists for λ-almost every $a \in \Lambda$ and it is symmetric.

Let μ have density $e^{-\rho}$ with respect to the Lebesgue measure on $\mathbb{R}^n$. For a leaf $\mathcal{S}$ such that $\mathrm{int}\mathcal{S} \subset \mathrm{int}T_{sij}$ and any Borel set $A \subset \mathbb{R}^n$ set

$$\mu_{\mathcal{S}}(A) = \int_{\mathrm{int}\mathcal{S}} \mathbf{1}_A e^{-\rho} J_n F \circ G d\mathcal{H}_m. \tag{2.7.5}$$

By Lemma 2.7.3 it follows now that for any Borel set $A \subset T_{sij}$ there is

$$\mu(A) = \int_{\Lambda} \mu_{\mathcal{S}(v(a))}(A) d\lambda(a),$$

where $v\colon \mathbb{R}^{n-m} \to \mathbb{R}^n$ is the map from Lemma 2.4.2. Neglecting a set of Lebesgue measure zero, we may assume that v is differentiable on the set Λ. Let ν denote the push-forward of λ via the map

$$\Lambda \ni a \mapsto \mathcal{S}(v(a)) \in T^m. \tag{2.7.6}$$

By Lemma 2.7.3 and the definition of ν the condition i) is satisfied. Note that the map (2.7.6) is Borel measurable, by the proof of Theorem 2.6.1. Hence ν is a Borel measure. For any Borel set $A \subset T_{sij}$ there is

$$\mu(A) = \int_{T^m} \mu_{\mathcal{S}}(A) d\nu(\mathcal{S}).$$

Hence the condition iv) of Theorem 2.7.2 is satisfied. Condition ii) holds true by the definition (2.7.5). We shall prove that iii) holds true as well.

Note that the density of a measure $\mu_{\mathcal{S}}$ for an m-dimensional leaf $\mathcal{S}$ is equal to

$$\frac{d\mu_{\mathcal{S}}}{d\mathcal{H}_m} = J_n F \circ G e^{-\rho} \mathbf{1}_{\mathcal{S}}.$$

Recall (see Lemma 2.4.5) that F, G are given by the formulae

$$F(a,b) = v(a) + Du(v(a))^*(b) \text{ and } G(x) = (w(z), u(x) - u(z)), \tag{2.7.7}$$

where $w\colon \mathbb{R}^n \to \mathbb{R}^{n-m}$ and $v\colon \mathbb{R}^{n-m} \to \mathbb{R}^n$ are maps from Lemma 2.4.2. Let us recall that $v(a) \in S_s^i$ for all $a \in \Lambda$. It follows by the definition of S_s that $u(v(a)) = s$ for all $a \in \Lambda$. Recall that, by Lemma 2.2.8, u is differentiable in $\mathrm{int}T_{sij}$. Thus, as we assumed that v is differentiable in Λ, for every $a \in \Lambda$

$$Du(v(a))Dv(a) = 0. \tag{2.7.8}$$

For $(a,b) \in G(\mathrm{int}T_{sij})$ the derivative of F at (a,b) is equal to

$$DF(a,b) = [Dv(a) + D^2u(v(a))^*(Dv(a)(\cdot))(b), Du(v(a))^*].$$

Note that for any vectors $z \in \mathbb{R}^{n-m}$ and $w \in \mathbb{R}^m$ the derivatives $Dv(a)z$ and $Du(v(a))^*w$ are orthogonal. Indeed, by (2.7.8),

$$\Big\langle Du(v(a))^*(w), Dv(a)(z)\Big\rangle = \Big\langle w, Du(v(a))Dv(a)(z)\Big\rangle = 0.$$

Let P denote the orthogonal projection onto the tangent space V of the leaf $\mathcal{S}$ containing $v(a)$. Then by Lemma 2.2.8 $Du(v(a)) = TP$. Let $P^\perp$ denote the orthogonal projection onto the orthogonal complement of V. Then

$$DF(a,b) = [Dv(a) + D^2u(v(a))^*(P^\perp Dv(a)(\cdot))(b), Du(v(a))^*].$$

Therefore, by the formula for block matrices, and as $Du(v(a))^*$ is isometric, we have

$$|\det(DF(a,b))| = \Big|\det\Big(Dv(a) + P^\perp D^2u(v(a))^*(P^\perp Dv(a)(\cdot))(b)\Big)\Big|,$$

which is equal to

$$|\det\left(P^\perp Dv(a)\right)|\Big|\det\Big(\mathrm{Id} + P^\perp D^2u(v(a))^*(P^\perp(\cdot))(b)\Big)\Big|. \tag{2.7.9}$$

Note that

$$H(b) = \Big(\mathrm{Id} + P^\perp D^2u(v(a))^*(P^\perp(\cdot))(b)\Big) \tag{2.7.10}$$

is a linear operator on the image of $P^\perp$, which is of dimension $n-m$. Moreover it is symmetric and invertible for any b such that $(a,b) \in G(\mathrm{int}T_{pij})$, as F is a bijection. Consider for some $b' \in \mathbb{R}^m$

$$P^\perp D^2u(v(a))^*(P^\perp(\cdot))(b').$$

Let A be such that

$$P^\perp D^2u(v(a))^*(P^\perp(\cdot))(b') = A\Big(\mathrm{Id} + P^\perp D^2u(v(a))^*(P^\perp(\cdot))(b)\Big). \tag{2.7.11}$$

Then A is conjugate to a symmetric operator of rank at most $n-m$, as

$$H(b)^{-\frac{1}{2}}AH(b)^{\frac{1}{2}} = H(b)^{-\frac{1}{2}}P^{\perp}D^2u(v(a))^*(P^{\perp}(\cdot))(b')H(b)^{-\frac{1}{2}}.$$

In consequence, by the Cauchy-Schwarz inequality,

$$(\mathrm{tr}A)^2 \leq (n-m)\mathrm{tr}(A)^2. \tag{2.7.12}$$

Let $x = F(a,b)$. Let q belong to the tangent space of $\mathcal{S}$. It is necessairly of the form $q = Du(v(a))^*(b')$ for some $b' \in \mathbb{R}^m$. Then by (2.7.7), (2.7.9) and (2.7.10)

$$\begin{aligned} D\log|\det DF\circ G|(x)(q) &= \frac{d}{dt}\log|\det\left(DF(G(F(a,b)+tDu(v(a))^*(b'))\right| = \\ &= \frac{d}{dt}\log|\det(DF(a,b+tb'))| = \frac{d}{dt}\log|\det H(b+tb')|. \end{aligned}$$

Therefore, by (2.7.3), (2.7.4) and by (2.7.11)

$$D\log|\det DF\circ G|(x)(q) = \mathrm{tr}\Big(H(b)^{-1}P^{\perp}D^2u(v(a))^*(P^{\perp}(\cdot))(b')\Big) = \mathrm{tr}A$$

and

$$D^2\log|\det DF\circ G|(x)(q,q) = -\mathrm{tr}\Big(H(b)^{-1}P^{\perp}D^2u(v(a))^*(P^{\perp}(\cdot))(b')\Big)^2 = -\mathrm{tr}(A^2).$$

By (2.7.12) and by Lemma 2.7.4, if $N \notin [m,n]$, then

$$\begin{aligned} -D^2\log|\det DF\circ G|(x)(q,q) &= \mathrm{tr}(A^2) \geq \\ &\geq \frac{1}{n-m}(\mathrm{tr}A)^2 \geq \frac{1}{N-m}(D\rho(x)(q)-\mathrm{tr}A)^2 - \frac{(D\rho(x)(q))^2}{N-n}. \end{aligned}$$

Note that by the assumption on μ, c.f. Definition 2.7.1, for all $p \in \mathbb{R}^n$

$$D^2\rho(x)(p,p) - \frac{D\rho(x)(p)^2}{N-n} \geq \kappa\|p\|^2.$$

Thus for all q in the tangent space of $\mathcal{S}$ there is

$$D^2\rho(x)(q,q) - D^2\log|\det DF\circ G|(x)(q,q) - \frac{\big(D\rho(x)(q) - D(\log|\det DF\circ G|)(x)(q)\big)^2}{N-m} \geq \kappa\|q\|^2.$$

We infer that $(\mathrm{int}\mathcal{S}, \|\cdot\|, \mu_{\mathcal{S}})$ satisfies the curvature-dimension condition $CD(\kappa, N)$, provided that $N \notin [m,n]$.

If $N = n$, then ρ is required to be a constant function, and thus in this case the inequality is also satisfied. If $N = \infty$, then the desired estimates follow readily. □

For the historical remarks on similar estimates we refer to [70].

Chapter 3

Optimal transport of vector measures

3.1 Introduction

In the current chapter we study an extension of the classical optimal transport problem (see e.g. [103, 104] for extensive account) to the case of vector measures.

In Section 3.2 we provide a definition of optimal transport of $\mathbb{R}^m$-valued vector measures on a metric space X with metric d. Namely, for an $\mathbb{R}^m$-valued measure μ on X such that $\mu(X) = 0$ and $\int_{\mathbb{R}^n} d(x, x_0) d\|\mu\|(x) < \infty$ for some $x_0 \in X$, we consider the infimum of integrals

$$\int_{X \times X} d(x, y) d\|\pi\|(x, y) \tag{3.1.1}$$

taken over the set of all $\mathbb{R}^m$-valued measures π on $X \times X$ such that $\mathrm{P}_1\pi - \mathrm{P}_2\pi = \mu$, where $\mathrm{P}_1\pi, \mathrm{P}_2\pi$ denote the first and the second marginal of π respectively.

We prove that (3.1.1) admits a convex dual problem, which is to find supremum of integrals

$$\int_X \langle v, d\mu \rangle \tag{3.1.2}$$

over the set of all 1-Lipschitz maps $v \colon X \to \mathbb{R}^m$.

We study a conjecture of Klartag (see [70, Chapter 6]). The problem is as follows. Suppose that we are given a vector measure μ on $\mathbb{R}^n$, with $\mu(\mathbb{R}^n)$, which is absolutely continuous with respect to Lebesgue measure. Let $u \colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map, with respect to Euclidean norms, that attains the supremum

$$\sup\Big\{ \int_{\mathbb{R}^n} \langle v, d\mu \rangle \mid v \colon \mathbb{R}^n \to \mathbb{R}^m \text{ is 1-Lipschitz} \Big\}. \tag{3.1.3}$$

It is claimed in [70] that the following mass balance condition holds true

$$\mu(A) = 0 \text{ for any Borel set } A \text{ that is a union of a collection of leaves of } u. \tag{3.1.4}$$

Using the developed theory, we answer in Section 3.2 the conjecture in the affirmative, provided that there exists an optimal transport such that the first marginal of its total variation is absolutely continuous with respect to the Lebesgue measure; see Theorem 3.2.19.

In Section 3.3 we assume that $m > 1$ and we provide a counterexample to the conjecture. It shows that, in general, the mass balance condition (3.1.4) does not hold true. Let $\mathcal{F}$ be any subset of 1-Lipschitz maps that is locally uniformly closed. We prove that the mass balance condition (3.1.4) fails to be true, even when the variational problem (3.1.3) is replaced by

$$\sup\Big\{\int_{\mathbb{R}^n} \langle v, d\mu\rangle \mid v \in \mathcal{F}\Big\}, \tag{3.1.5}$$

unless $\mathcal{F}$ is trivial, i.e. consists merely of affine maps. This is shown for also any norm on $\mathbb{R}^n$ and any strictly convex norm on $\mathbb{R}^m$.

3.2 Optimal transport of vector measures

In this section we develop theory of optimal transport of vector measures.

Let X be a metric space and let π be a Borel vector measure on X. Its total variation $\|\pi\|$ is defined by the formula

$$\|\pi\|(A) = \sup\Big\{\sum_{i=1}^{\infty}\|\pi(A_i)\| \mid A = \bigcup_{i=1}^{\infty} A_i, A_i \text{ are Borel sets in } X, A_i \cap A_j = \emptyset, i, j \in \mathbb{N}\Big\} \tag{3.2.1}$$

for all Borel sets $A \subset X$.

It can be shown (see [92]) that total variation of a vector measure is a non-negative finite measure.

Let X be a metric space with metric d. Let μ be $\mathbb{R}^m$-valued Borel measure on X. If π is a $\mathbb{R}^m$-valued Borel measure on $X \times X$, we write $\mathrm{P}_1\pi$ for the first *marginal* of π, i.e. the measure given by

$$\mathrm{P}_1\pi(A) = \pi(A \times X),$$

for all Borel $A \subset X$, and $\mathrm{P}_2\pi$ for the second *marginal* of π,

$$\mathrm{P}_2\pi(B) = \pi(X \times B),$$

for all Borel $B \subset X$. We shall consider a variational problem

$$\mathcal{I}(\mu) = \inf\Big\{\int_{X\times X} d(x,y)d\|\pi\|(x,y) \mid \pi \in \Gamma(\mu)\Big\}. \tag{3.2.2}$$

Here $\Gamma(\mu)$ is the set of all $\mathbb{R}^m$-valued Borel measures π on $X \times X$ such that

$$\mu = \mathrm{P}_1\pi - \mathrm{P}_2\pi.$$

To check whether (3.2.2) defines a meaningful quantity, we have to check whether $\Gamma(\mu)$ is non-empty.

We shall need the following definition.

Definition 3.2.1. Let σ be an $\mathbb{R}^m$-valued Borel measure on X and let θ be a Borel signed measure on X. A unique Borel $\mathbb{R}^m$-valued measure $\sigma \otimes \theta$ such that

$$\langle \sigma \otimes \theta, v \rangle = \langle \sigma, v \rangle \otimes \theta$$

for all $v \in \mathbb{R}^m$ we shall call a *product measure*. Here $\langle \sigma, v \rangle \otimes \theta$ is the usual product measure of $\mathbb{R}$-valued measures.

Remark 3.2.2. It is clear that the product measure exists. Analogously we define the product measure $\theta \otimes \sigma$ for Borel signed measure σ and Borel $\mathbb{R}^m$-valued measure θ.

Proposition 3.2.3. $\Gamma(\mu)$ *is non-empty if and only if*

$$\mu(X) = 0. \tag{3.2.3}$$

Proof. Clearly, if there exists $\pi \in \Gamma(\mu)$, then

$$\mu(X) = \mathrm{P}_1\pi(X) - \mathrm{P}_2\pi(X) = \pi(X \times X) - \pi(X \times X) = 0,$$

so the condition (3.2.3) is satisfied. Conversely, assume that (3.2.3) holds true. Let ν be any Borel probability measure on X. Set

$$\pi = \mu \otimes \nu.$$

Here $\mu \otimes \nu$ is the product measure; see Definition 3.2.1. Then for any Borel set $A \subset X$ and any vector $v \in \mathbb{R}^m$, we have

$$\langle \pi(A \times X) - \pi(X \times A), v \rangle = \langle \mu(A), v \rangle - \langle \mu(X), v \rangle \nu(A) = \langle \mu(A), v \rangle.$$

This is to say, $\mathrm{P}_1\pi - \mathrm{P}_2\pi = \mu$. □

The quantity defined by (3.2.2) we shall call the Kantorovich–Rubinstein norm of μ (see e.g. [104, 103, 26] for references regarding the Monge–Kantorovich problem).

Proposition 3.2.4. *Assume that* $\mu(X) = 0$. *Then* $\mathcal{I}(\mu) < \infty$ *provided that*

$$\int_{\mathbb{R}^n} d(x, x_0) d\|\mu\|(x) < \infty \tag{3.2.4}$$

for some (equivalently: any) $x_0 \in X$.

Proof. Define

$$\pi = \mu \otimes \delta_{x_0}.$$

Here δ_{x_0} is a probability measure such that $\delta_{x_0}(\{x_0\}) = 1$. Then $\pi \in \Gamma(\mu)$ and

$$\int_{X\times X} d(x,y)d\|\pi\|(x,y) \leq \int_X d(x,x_0)d\|\mu\|(x). \tag{3.2.5}$$

This shows that $\mathcal{I}(\mu) < \infty$, provided that (3.2.4) is satisfied. The equivalence of finiteness of

$$\int_{\mathbb{R}^n} d(x,y)d\|\mu\|(x) < \infty$$

for any $y \in X$ follows by the triangle inequality. □

Definition 3.2.5. We define the *Wasserstein space* $\mathcal{W}(X,\mathbb{R}^m)$ of all Borel measures μ on X with values in $\mathbb{R}^m$ such that

$$\mu(X) = 0 \text{ and } \int_X d(x,x_0)d\|\mu\|(x) < \infty$$

for some $x_0 \in X$. We endow it with a norm $\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} = \mathcal{I}(\mu)$.

Before we proceed let us recall the following definition.

We say that a non-negative Borel measure μ on X is *inner regular* if for any Borel set $B \subset X$ we have

$$\mu(B) = \sup\{\mu(K) \mid K \subset B, K \text{ is a compact set}\}.$$

Let us note that Ulam's lemma tells that any finite measure on a Polish space is inner regular.

Lemma 3.2.6. *Suppose that X is a Polish space. Let μ be a $\mathbb{R}^m$-valued Borel measure in $\mathcal{W}(X,\mathbb{R}^m)$. Suppose that for any Lipschitz function $u\colon X \to \mathbb{R}^m$*

$$\int_X \langle u, d\mu\rangle = 0.$$

Then $\mu = 0$.

Proof. We may assume that $m = 1$. Let $\mu = \mu_+ - \mu_-$ be the Hahn–Jordan decomposition of μ. There exists two disjoint, Borel sets $A, B \subset X$ with $\mu_+(A^c) = 0$ and $\mu_-(B^c) = 0$. Choose any Borel set $E \subset A$. As any finite measure on X is inner regular, for any $\epsilon > 0$, there exists a compact set $K \subset E$ such that

$$\mu_+(E) \leq \mu_+(K) + \epsilon.$$

Define a function u_ϵ by the formula

$$u_\epsilon(x) = (1 - \frac{1}{\epsilon}\operatorname{dist}(x,K)) \vee 0.$$

Then u_ϵ is Lipschitz, equal to one on K and equal to zero on the complement of

$$K_\epsilon = \{x \in X \mid \operatorname{dist}(x, K) \leq \epsilon\}.$$

Thus

$$0 = \int_X u_\epsilon d\mu = \mu_+(K) + \int_{K_\epsilon \setminus K} u_\epsilon d\mu,$$

Therefore, by the above,

$$\mu_+(E) \leq \epsilon + \mu_+(K) \leq \epsilon + \mu_-(K_\epsilon \setminus K).$$

Letting ϵ tend to zero, we get $\mu_+(E) = 0$. It follows that $\mu_+ = 0$. Analogously, $\mu_- = 0$. This is to say, $\mu = 0$. □

Remark 3.2.7. In what follows, we shall always assume that underlying space X is a Polish space.

Proposition 3.2.8. *The function* $\mathcal{W}(X, \mathbb{R}^m) \ni \mu \mapsto \|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} \in \mathbb{R}$ *is a norm.*

Proof. Let us first check that

$$\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} = 0 \text{ if and only if } \mu = 0. \tag{3.2.6}$$

If $\mu = 0$, then $\pi = 0$ belongs to $\Gamma(\mu)$, so $\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} = 0$. Conversely, assume that $\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} = 0$. Choose any L-Lipschitz function

$$u\colon X \to \mathbb{R}^m.$$

Then for any $\pi \in \Gamma(\mu)$ we have

$$\Big|\int_X \langle u, d\mu\rangle\Big| = \Big|\int_{X\times X} \langle u(x) - u(y), d\pi(x,y)\rangle\Big| \leq L \int_{X\times X} d(x,y) d\|\pi\|(x,y).$$

Therefore if $\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} = 0$, then

$$\int_X \langle u, d\mu\rangle = 0.$$

It follows by Lemma 3.2.6, that $\mu = 0$. Homogeneity of $\|\cdot\|_{\mathcal{W}(X,\mathbb{R}^m)}$ is clear. Let us show that the triangle inequality holds. For this choose measures $\mu, \nu \in \mathcal{W}(X, \mathbb{R}^m)$ and any measures $\pi \in \Gamma(\mu)$ and $\rho \in \Gamma(\nu)$. Then

$$\mu + \nu = \mathrm{P}_1(\pi + \rho) - \mathrm{P}_2(\pi + \rho),$$

so that $\pi + \rho \in \Gamma(\mu + \nu)$. It follows that

$$\begin{aligned}\|\mu + \nu\|_{\mathcal{W}(X,\mathbb{R}^m)} &\leq \int_{X\times X} d(x,y) d\|\pi + \rho\|(x,y) \leq \\ &\leq \int_{X\times X} d(x,y) d\|\pi\|(x,y) + \int_{X\times X} d(x,y) d\|\rho\|(x,y).\end{aligned}$$

Taking infimum over all π, ρ we see that the triangle inequality holds true. □

Proposition 3.2.9. *The linear space $\mathcal{U}$ of measures of the form*

$$\sum_{i=1}^{n} \delta_{x_i} v_i$$

for $x_i \in X$ and $v_i \in \mathbb{R}^m$, $i = 1, \ldots, n$, such that $\sum_{i=1}^{n} v_i = 0$, is dense in $\mathcal{W}(X, \mathbb{R}^m)$.

Proof. Choose any measure $\mu \in \mathcal{W}(X, \mathbb{R}^m)$. Choose any $\epsilon > 0$. Choose any point $x_0 \in X$ and a compact set K such that

$$\int_{K^c} d(x, x_0) d\|\mu\|(x) \leq \epsilon.$$

Choose pairwise disjoint Borel sets $A_1, A_2, \ldots, A_k \subset K$ such that the diameter of each is at most ϵ and

$$K = \bigcup_{i=1}^{k} A_i.$$

Consider the restrictions $\mu_i = \mu|_{A_i}$ of the measure μ to the sets A_i, $i = 1, 2, \ldots, k$. Choose any points $x_i \in A_i$. Then, as

$$\pi_i = \mu_i \otimes \delta_{x_i} \in \Gamma(\mu_i - \mu_i(X)\delta_{x_i}),$$

we have

$$\|\mu_i - \mu_i(X)\delta_{x_i}\|_{\mathcal{W}(X,\mathbb{R}^m)} \leq \int_X d(y, x_i) d\|\mu_i\|(y) \leq \epsilon \|\mu\|(A_i).$$

Let $A_0 = K^c$ and let $\mu_0 = \mu|_{A_0}$. Then

$$\pi_0 = \mu_0 \otimes \delta_{x_0} \in \Gamma(\mu_0 - \mu_0(X)\delta_{x_0}),$$

so

$$\|\mu_0 - \mu_0(X)\delta_{x_0}\|_{\mathcal{W}(X,\mathbb{R}^m)} \leq \int_X d(x, x_0) d\|\mu_0\|(x) \leq \epsilon.$$

Set

$$\nu = \sum_{i=0}^{k} \mu(A_i)\delta_{x_i}.$$

Then $\nu \in \mathcal{U}$. By the triangle inequality

$$\|\mu - \nu\|_{\mathcal{W}(X,\mathbb{R}^m)} \leq \sum_{i=0}^{k} \|\mu_i - \mu_i(X)\delta_{x_i}\|_{\mathcal{W}(X,\mathbb{R}^m)} \leq$$

$$\leq \epsilon \sum_{i=1}^{k} \|\mu(A_i)\| + \epsilon \leq \epsilon(\|\mu\|(X) + 1).$$

This concludes the proof. ∎

Corollary 3.2.10. *If X is separable, then so is the Wasserstein space $\mathcal{W}(X, \mathbb{R}^m)$.*

Proof. Fix $n \in \mathbb{N}$. Choose a countable dense subset $A \subset X$ and a set

$$B \subset \Big\{(w_1, \ldots, w_n) \in \mathbb{R}^m \times \ldots \mathbb{R}^m \mid \sum_{i=1}^{n} w_i = 0\Big\} \tag{3.2.7}$$

which is countable and dense in the set on the right-hand side of (3.2.7). Consider a measure μ given by

$$\mu = \sum_{i=1}^{n} \delta_{x_i} v_i$$

for $x_i \in X$ and $v_i \in \mathbb{R}^m$, $i = 1, \ldots, n$, such that $\sum_{i=1}^{n} v_i = 0$. Choose $\epsilon > 0$ and $\tilde{x}_i \in A$, $i = 1, \ldots, n$, and $(\tilde{v}_i)_{i=1}^{n} \in B$, such that for $i = 1, \ldots, n$

$$d(x_i, \tilde{x}_i) < \epsilon \text{ and } \|v_i - \tilde{v}_i\| < \epsilon \text{ and } \sum_{i=1}^{n} \tilde{v}_i = 0.$$

Set

$$\tilde{\mu} = \sum_{i=1}^{n} \delta_{\tilde{x}_i} \tilde{v}_i.$$

Then

$$\|\mu - \tilde{\mu}\|_{\mathcal{W}(X,\mathbb{R}^m)} \leq \Big\| \sum_{i=1}^{n} \delta_{x_i} (v_i - \tilde{v}_i) \Big\|_{\mathcal{W}(X,\mathbb{R}^m)} + \Big\| \sum_{i=1}^{n} (\delta_{x_i} - \delta_{\tilde{x}_i}) v_i \Big\|_{\mathcal{W}(X,\mathbb{R}^m)}$$

Choose any $x_0 \in X$. Taking

$$\pi = \sum_{i=1}^{n} \delta_{x_i} \otimes \delta_{x_0} (v_i - \tilde{v}_i) \text{ and } \rho = \sum_{i=1}^{n} (\delta_{x_i} \otimes \delta_{\tilde{x}_i}) v_i$$

we see that

$$\Big\| \sum_{i=1}^{n} \delta_{x_i} (v_i - \tilde{v}_i) \Big\|_{\mathcal{W}(X,\mathbb{R}^m)} \leq \epsilon \sum_{i=1}^{n} d(x_i, x_0)$$

and

$$\Big\| \sum_{i=1}^{n} (\delta_{x_i} - \delta_{\tilde{x}_i}) v_i \Big\|_{\mathcal{W}(X,\mathbb{R}^m)} \leq \epsilon \sum_{i=1}^{n} \|v_i\|.$$

The conclusion follows now from Proposition 3.2.9. □

Definition 3.2.11. Choose any $x_0 \in X$. Define

$$\mathcal{L}(X, \mathbb{R}^m) = \{u \colon X \to \mathbb{R}^m \mid u \text{ is Lipschitz and } u(x_0) = 0\},$$

i.e. the Banach space of $\mathbb{R}^m$-valued Lipschitz functions on X taking value zero at x_0, with norm

$$\|u\|_{\mathcal{L}(X,\mathbb{R}^m)} = \sup \Big\{ \frac{\|u(x) - u(y)\|}{d(x,y)} \mid x, y \in X, x \neq y \Big\}.$$

Proposition 3.2.12. *Define*

$$T\colon \mathcal{L}(X,\mathbb{R}^m)\to \mathcal{W}(X,\mathbb{R}^m)^*$$

and

$$S\colon \mathcal{W}(X,\mathbb{R}^m)^*\to \mathcal{L}(X,\mathbb{R}^m)$$

by

$$T(u)(\mu)=\int_X\langle u,d\mu\rangle \tag{3.2.8}$$

and

$$\langle S(\lambda)(x),w\rangle=\lambda((\delta_x-\delta_{x_0})w), \tag{3.2.9}$$

for any $w\in\mathbb{R}^m$. Then S,T are mutual reciprocals and establish an isometric isomorphism of $\mathcal{L}(X,\mathbb{R}^m)$ and $\mathcal{W}(X,\mathbb{R}^m)^*$.

Proof. Choose any $\pi\in\Gamma(\mu)$. Then $\mathrm{P}_1\pi-\mathrm{P}_2\pi=\mu$. Thus, if u is a Lipschitz map, then

$$\left|\int_X\langle u,d\mu\rangle\right|=\left|\int_X\langle u(x)-u(y),d\pi(x,y)\rangle\right|\leq \|u\|_{\mathcal{L}(X,\mathbb{R}^m)}\int_X d(x,y)d\|\pi\|(x,y).$$

Taking infimum over all $\pi\in\Gamma(\mu)$, we see that

$$\left|\int_X\langle u,d\mu\rangle\right|\leq \|u\|_{\mathcal{L}(X,\mathbb{R}^m)}\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)}.$$

The above calculation shows that the formula (3.2.8) defines a continuous functional of norm at most $\|u\|_{\mathcal{L}(X,\mathbb{R}^m)}$. If $w\in\mathbb{R}^m$ if of norm one and $x,y\in X$, $x\neq y$, then for

$$\mu_{x,y,w}=\frac{\delta_x-\delta_y}{d(x,y)}w \tag{3.2.10}$$

we have $\|\mu_{x,y,w}\|_{\mathcal{W}(X,\mathbb{R}^m)}\leq 1$ and for any $u\in\mathcal{L}(X,\mathbb{R}^m)$

$$\int_{\mathbb{R}^n}\langle u,d\mu_{x,y,w}\rangle=\frac{\langle w,u(x)-u(y)\rangle}{d(x,y)}.$$

Thus

$$\|u\|_{\mathcal{L}(X,\mathbb{R}^m)}=\|T(u)\|.$$

We shall now show that $T\circ S=\mathrm{Id}$. Take any functional $\lambda\in\mathcal{W}(X,\mathbb{R}^m)^*$. Set

$$\sigma_{x,w}=(\delta_x-\delta_{x_0})w.$$

Then $S(\lambda)\colon X\to\mathbb{R}^m$ is defined by the formula

$$\langle S(\lambda)(x),w\rangle=\lambda(\sigma_{x,w}).$$

It is clear that the above formula defines $S(\lambda)$ uniquely. Then we claim that map $v = S(\lambda)$ is $\|\lambda\|$-Lipschitz. Indeed

$$\|v(x) - v(y)\| = \sup\{\langle v(x) - v(y), w\rangle \mid w \in \mathbb{R}^m, \|w\| = 1\},$$

and as

$$\langle v(x) - v(y), w\rangle = \lambda(\sigma_{x,w} - \sigma_{y,w}) \leq \|\lambda\| \|\sigma_{x,w} - \sigma_{y,w}\|_{\mathcal{W}(X,\mathbb{R}^m)}$$

we see that

$$\|v(x) - v(y)\| \leq \|\lambda\| d(x, y), \text{ since } \|\sigma_{x,w} - \sigma_{y,w}\|_{\mathcal{W}(X,\mathbb{R}^m)} \leq d(x, y).$$

Suppose that $\nu = (\delta_x - \delta_y)z$. We compute

$$T(v)(\nu) = \int_X \langle v, d\nu\rangle = \int_X \langle v, z\rangle d(\delta_x - \delta_y) = \lambda(\sigma_{x,z} - \sigma_{y,z}) = \lambda(\nu).$$

We see that $T(S(\lambda))$ and λ are equal on the set spanned by $(\delta_x - \delta_y)z$, where $x, y \in X$, $z \in \mathbb{R}^m$. By Proposition 3.2.9, we see that $T(S(\lambda))$ and λ are equal on $\mathcal{W}(X, \mathbb{R}^m)$.

Let us show also that $S \circ T = \mathrm{Id}$. Choose any $w \in \mathbb{R}^m$ and any map $u \in \mathcal{L}(X, \mathbb{R}^m)$. Then

$$\langle S(T(u))(x), w\rangle = T(u)((\delta_x - \delta_{x_0})w) = \int_X \langle u, d(\delta_x - \delta_{x_0})w\rangle = \langle u(x), w\rangle,$$

as $u(x_0) = 0$. Therefore $S(T(u)) = u$. □

Proposition 3.2.13. *For any* $\mu \in \mathcal{W}(X, \mathbb{R}^m)$

$$\sup\left\{\int_X \langle u, d\mu\rangle \mid u\colon X \to \mathbb{R}^m \textit{ is } 1\textit{-Lipschitz}\right\} = \|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)}. \tag{3.2.11}$$

Moreover, there exists 1*-Lipschitz function* u_0 *such that*

$$\sup\left\{\int_X \langle u, d\mu\rangle \mid u\colon X \to \mathbb{R}^m \textit{ is } 1\textit{-Lipschitz}\right\} = \int_X \langle u_0, d\mu\rangle. \tag{3.2.12}$$

Proof. Notice first that the left-hand side of (3.2.11) is clearly at most the right-hand side of (3.2.11). Take any $\mu \in \mathcal{W}(X, \mathbb{R}^m)$. Then by the Hahn–Banach theorem there exists a continuous linear functional λ of norm one such that

$$\lambda(\mu) = \|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)}.$$

By Proposition 3.2.12, we know that λ is of the form

$$\lambda(\mu) = \int_X \langle u_0, d\mu\rangle$$

for some Lipschitz map u_0. The Lipschitz constant of u_0 is equal to one, as

$$\|u_0\|_{\mathcal{L}(X,\mathbb{R}^m)} = \|\lambda\| = 1.$$

This completes the proof. □

Definition 3.2.14. Any 1-Lipschitz function $u\colon X \to \mathbb{R}^m$ such that (3.2.12) holds we shall call an *optimal potential* of measure μ.

Definition 3.2.15. A measure $\pi \in \Gamma(\mu)$ such that

$$\|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} = \int_{X\times X} d(x,y)d\|\pi\|(x,y)$$

we shall call an *optimal transport* for μ.

Theorem 3.2.16. *Let $\mu \in \mathcal{W}(X,\mathbb{R}^m)$. Let $u \in \mathcal{L}(X,\mathbb{R}^m)$ be a 1-Lipschitz map. Let $\pi \in \Gamma(\mu)$. The following conditions are equivalent:*

i)
$$\int_X \langle u, d\mu\rangle = \int_{X\times X} d(x,y)d\|\pi\|(x,y) = \|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)},$$

ii)
$$\int_A \langle u(x)-u(y), d\pi(x,y)\rangle = \int_A d(x,y)d\|\pi\|(x,y)$$
for any Borel set $A \subset X\times X$,

iii)
$$\int_X \langle u, d\mu\rangle = \int_{X\times X} d(x,y)d\|\pi\|(x,y),$$

iv) u is an optimal potential for μ and π is an optimal transport for μ.

Moreover, if the above conditions hold, then

$$\|u(x)-u(y)\| = d(x,y)$$

$\|\pi\|$-almost everywhere.

Proof. Assume that iii) holds. Observe that

$$\int_X \langle u, d\mu\rangle = \int_{X\times X} \langle u(x)-u(y), d\pi(x,y)\rangle.$$

As

$$\int_X \langle u, d\mu\rangle \leq \|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)} \leq \int_{X\times X} d(x,y)d\|\pi\|(x,y),$$

then by iii) we see that in the above inequalities we have equalities. This is to say, i) holds true.

Suppose now that i) holds. Clearly

$$\int_A \langle u(x)-u(y), d\pi(x,y)\rangle \leq \int_A d(x,y)d\|\pi\|(x,y).$$

If we had strict inequality in ii) for some Borel set $A\subset X\times X$, then the above computations shows that we would get strict inequality in i). Condition iv) is reformulation of i). The last part of the theorem follows readily from ii). □

Recall, see Definition 2.2.14, that, for a 1-Lipschitz map $u\colon \mathbb{R}^n \to \mathbb{R}^m$, we denote by $B(u)$ the set of points in $\mathbb{R}^n$ that belong to at least two distinct leaves of u.

Definition 3.2.17. Let $u\colon \mathbb{R}^n \to \mathbb{R}^m$ be a 1-Lipschitz map of Euclidean spaces. We say that a Borel set $A \subset \mathbb{R}^n$ is a *transport set* associated with u if it enjoys the following property: if $x \in A \setminus B(u)$ and $y \in \mathbb{R}^n$ is such that

$$\|u(x) - u(y)\| = \|x - y\|,$$

then $y \in A$.

Let us remark that a Borel set $A \subset \mathbb{R}^n$ that is a union of leaves of u is a transport set.

We say that a measure $\mu \in \mathcal{M}(Z, \mathbb{R}^m)$ is *concentrated* on a subset $X \subset Z$ if there is $\|\mu\|(Z \setminus X) = 0$.

Proposition 3.2.18. *Let $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$ be concentrated on a set $X \subset \mathbb{R}^n$. Then*

$$\|\mu\|_{\mathcal{W}(\mathbb{R}^n,\mathbb{R}^m)} = \|\mu\|_{\mathcal{W}(X,\mathbb{R}^m)}.$$

Here we consider the Euclidean metrics on X and on $\mathbb{R}^n$.

Proof. The assertion is that

$$\sup\Big\{ \int_{\mathbb{R}^n} \langle u, d\mu\rangle \mid u\colon \mathbb{R}^n \to \mathbb{R}^m \text{ is 1-Lipschitz}\Big\}$$

is equal to

$$\sup\Big\{ \int_{X} \langle u, d\mu\rangle \mid u\colon X \to \mathbb{R}^m \text{ is 1-Lipschitz}\Big\}.$$

By the Kirszbraun theorem any 1-Lipschitz function $u\colon X \to \mathbb{R}^m$ extends to a 1-Lipschitz function $\tilde{u}\colon \mathbb{R}^n \to \mathbb{R}^m$. Clearly, for any such extension

$$\int_{\mathbb{R}^n} \langle \tilde{u}, d\mu\rangle = \int_X \langle u, d\mu\rangle.$$

The assertion follows. □

Suppose that $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$. The following theorem shows that if there exists an optimal transport for μ such that its total variation has absolutely continuous first marginal, then the conjecture of Klartag holds true. Note that such existence is clear for $m = 1$, whenever μ is absolutely continuous with respect to the Lebesgue measure λ.

Below we shall be using the conditional measures $\|\mu\|_{\mathcal{S}}$, $\mathcal{S} \in CC(\mathbb{R}^n)$, of measure $\|\mu\|$ which has been considered in Chapter 2, Section 2.6. We shall denote by ν the resulting measure on $CC(\mathbb{R}^m)$, i.e. the push-forward of $\|\mu\|$ with respect to the map $\mathcal{S}\colon \mathbb{R}^n \to CC(\mathbb{R}^n)$.

Theorem 3.2.19. *Suppose that $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$. Let u be an optimal potential for μ. Suppose that there exists an optimal transport π of μ such that*

$$\mathrm{P}_1\|\pi\| \quad \textit{is absolutely continuous with respect to the Lebesgue measure } \lambda. \tag{3.2.13}$$

Then:

i) for any transport set A associated with u

a) $\mu(A) = 0$,

b) $\pi|_{A\times A} \in \Gamma(\mu|_A)$ is an optimal transport of $\mu|_A$

c) u is an optimal potential of $\mu|_A$,

ii) if $\|\mu\|_{\mathcal{S}}$, $\mathcal{S} \in CC(\mathbb{R}^n)$, are the conditional measures of $\|\mu\|$ with respect to the map $\mathcal{S}\colon \mathbb{R}^n \to CC(\mathbb{R}^n)$, then for ν-almost every $\mathcal{S} \in CC(\mathbb{R}^n)$ we have

$$\int_{\mathbb{R}^n} \frac{d\mu}{d\|\mu\|} d\|\mu\|_{\mathcal{S}} = 0$$

and u is an optimal potential of $\frac{d\mu}{d\|\mu\|} d\|\mu\|_{\mathcal{S}}$.

Moreover, ii) implies ia) and ic).

Proof. By Corollary 2.2.15 it follows that

$$\lambda(B(u)) = 0.$$

Suppose that (3.2.13) holds true. Then

$$\|\pi\|\big(B(u) \times \mathbb{R}^n\big) = 0.$$

Let

$$I = \big\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid \|u(x) - u(y)\| = \|x - y\|\big\}.$$

By Theorem 3.2.16, $\|\pi\|(I^c) = 0$. Thus π is concentrated on the set

$$C = I \cap B(u)^c \times \mathbb{R}^n.$$

Suppose that $(x, y) \in C$. Then, as A is a transport set, by the definition of $B(u)$,

$$x \in A \text{ implies that } y \in A. \tag{3.2.14}$$

Let $\eta = \pi|_{A\times A}$. To prove ib), it is enough to show that η is an optimal transport and that

$$\eta \in \Gamma(\mu|_A).$$

For this, let $D \subset \mathbb{R}^n$ be any Borel set. Using the fact that $\pi \in \Gamma(\mu)$ and the fact that $\|\pi\|(C^c) = 0$ and (3.2.14), we have

$$\mu(A \cap D) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \Big(\mathbf{1}_{A\cap D}(x) - \mathbf{1}_{A\cap D}(y)\Big) d\pi(x,y) =$$
$$= \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathbf{1}_{A\times A}(x,y)\Big(\mathbf{1}_D(x) - \mathbf{1}_D(y)\Big) d\pi(x,y) =$$
$$= \int_{\mathbb{R}^n \times \mathbb{R}^n} \Big(\mathbf{1}_D(x) - \mathbf{1}_D(y)\Big) d\eta(x,y) = \mathrm{P}_1\eta(D) - \mathrm{P}_2\eta(D).$$

It follows that $\pi|_{A\times A} \in \Gamma(\mu|_A)$. Then

$$\int_A \langle u, d\mu\rangle = \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathbf{1}_I(x,y)\Big\langle \mathbf{1}_A(x)u(x) - \mathbf{1}_A(y)u(y), d\pi(x,y)\Big\rangle. \tag{3.2.15}$$

Therefore, by (3.2.14),

$$\int_A \langle u, d\mu\rangle = \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathbf{1}_{A\times A}(x,y)\Big\langle u(x) - u(y), d\pi(x,y)\Big\rangle.$$

By condition ii) of Theorem 3.2.16 we see that

$$\int_A \langle u, d\mu\rangle = \int_{A\times A} \|x-y\| d\|\pi\|(x,y).$$

Theorem 3.2.16, condition iii), tells us that $\pi|_{A\times A}$ is an optimal transport and u is an optimal potential. Also $\mu(A) = 0$, as $\pi|_{A\times A} \in \Gamma(\mu|A)$. This proves i).

For a proof of ii) let us take any Borel set $B \subset CC(\mathbb{R}^n)$ and let $A = \mathcal{S}^{-1}(B)$. Then,

$$0 = \mu(A) = \int_{CC(\mathbb{R}^n)} \int_{\mathbb{R}^n} \mathbf{1}_A \frac{d\mu}{d\|\mu\|} d\|\mu\|_{\mathcal{S}} d\nu(\mathcal{S}) = \int_{CC(\mathbb{R}^n)} \mathbf{1}_B(\mathcal{S}) \int_{\mathbb{R}^n} \frac{d\mu}{d\|\mu\|} d\|\mu\|_{\mathcal{S}} d\nu(\mathcal{S}).$$

As this holds true for any B we infer that for ν-almost every $\mathcal{S}$ there is

$$0 = \int_{\mathbb{R}^n} \frac{d\mu}{d\|\mu\|} d\|\mu\|_{\mathcal{S}}. \tag{3.2.16}$$

By i), for any A, B as above and any 1-Lipschitz map v there is

$$0 \leq \int_A \langle u - v, d\mu\rangle = \int_{CC(\mathbb{R}^n)} \mathbf{1}_B(\mathcal{S}) \int_{\mathbb{R}^n} \Big\langle u - v, \frac{d\mu}{d\|\mu\|}\Big\rangle d\|\mu\|_{\mathcal{S}} d\nu(\mathcal{S}).$$

Hence, by separability of $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, for ν-almost every $\mathcal{S}$ and for every v there is

$$0 \leq \int_{\mathbb{R}^n} \Big\langle u - v, \frac{d\mu}{d\|\mu\|}\Big\rangle d\|\mu\|_{\mathcal{S}} d\nu(\mathcal{S}). \tag{3.2.17}$$

The fact that ii) implies ia) and ic) follows by integration of (3.2.16) and (3.2.17) respectively.

□

3.3 Counterexample

We shall now provide necessary tools for the aforementioned counterexample to the conjecture of Klartag.

In fact we shall provide a more general theorem for which we shall consider locally uniformly closed subsets subsets $\mathcal{F}$ of 1-Lipschitz maps of $\mathbb{R}^n$ to $\mathbb{R}^m$ endowed with norms which are not necessarily Euclidean. Suppose that a measure μ belongs to $\mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$. We consider supremum of integrals

$$\int_{\mathbb{R}^n} \langle u, d\mu \rangle \tag{3.3.1}$$

taken over all $u \in \mathcal{F}$. An optimal $u_0 \in \mathcal{F}$, i.e. the map that satisfies

$$\int_{\mathbb{R}^n} \langle u_0, d\mu \rangle = \sup \Big\{ \int_{\mathbb{R}^n} \langle u, d\mu \rangle \mid u \in \mathcal{F} \Big\},$$

we shall call an $\mathcal{F}$-optimal potential of μ.

Lemma 3.3.1. *Let $X \subset \mathbb{R}^n$ be a compact set. Suppose that $(\mu_k)_{k=1}^{\infty} \subset \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$ are all supported on X and converge weakly* to $\mu_0 \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$, i.e. for any continuous and bounded function $g \colon \mathbb{R}^n \to \mathbb{R}^m$ we have*

$$\lim_{k \to \infty} \int_{\mathbb{R}^n} \langle g, d\mu_k \rangle = \int_{\mathbb{R}^n} \langle g, d\mu_0 \rangle.$$

Suppose that for $k = 1, 2, \dots$, $u_k \in \mathcal{F}$ is an $\mathcal{F}$-optimal potential of μ_k and that u_k converge locally uniformly to $u_0 \colon \mathbb{R}^n \to \mathbb{R}^m$. Then u_0 is an $\mathcal{F}$-optimal potential of μ_0.

Proof. By the assumption, for any continuous and bounded map $g \colon \mathbb{R}^n \to \mathbb{R}^m$, we have

$$\lim_{k \to \infty} \int_{\mathbb{R}^n} \langle g, d(\mu_k - \mu_0) \rangle = 0.$$

In particular, as μ_k are all supported on X, we have

$$\lim_{k \to \infty} \int_{\mathbb{R}^n} \langle u_0, d(\mu_k - \mu_0) \rangle = 0.$$

By the Banach–Steinhaus theorem, the sequence $(\mu_k)_{k=1}^{\infty}$ is bounded in the total variation norm. Hence, by uniform convergence on X,

$$\lim_{k \to \infty} \int_{\mathbb{R}^n} \langle u_k - u_0, d\mu_k \rangle = 0.$$

It follows that

$$\int_{\mathbb{R}^n} \langle u_k, d\mu_k \rangle = \int_{\mathbb{R}^n} \langle u_0, d\mu_k \rangle + \int_{\mathbb{R}^n} \langle u_k - u_0, d\mu_k \rangle$$

converges to $\int_{\mathbb{R}^n} \langle u_0, d\mu_0 \rangle$. As for any 1-Lipschitz map $h \in \mathcal{F}$ we have

$$\int_{\mathbb{R}^n} \langle h, d\mu_k \rangle \leq \int_{\mathbb{R}^n} \langle u_k, d\mu_k \rangle.$$

we also have

$$\int_{\mathbb{R}^n} \langle h, d\mu_0 \rangle \leq \int_{\mathbb{R}^n} \langle u_0, d\mu_0 \rangle.$$

The proof is complete. □

Lemma 3.3.2. *Let $m \leq n$. Let $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$ and let u be an optimal potential of μ. Suppose that there exists an optimal transport π for μ or that any transport set of u is of μ-measure zero. Let A be the union of all leaves of dimension at least one. Then*

$$\|\mu\|(A^c) = 0.$$

Proof. Corollary 2.5.7 tells us that A is Borel measurable. Suppose that there exists an optimal transport π for μ. By Theorem 3.2.16, π is supported on the set

$$I = \big\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid \|u(x) - u(y)\| = \|x - y\|\big\}.$$

As $\mu = \mathrm{P}_1\pi - \mathrm{P}_2\pi$, for any Borel set $B \subset A^c$, we have

$$\mu(B) = \pi(B \times \mathbb{R}^n) - \pi(\mathbb{R}^n \times B) = 0,$$

for if $B \subset A^c$, then

$$\big(B \times \mathbb{R}^n\big) \cap I \subset \{(x, x) \mid x \in \mathbb{R}^n\} \text{ and } \big(\mathbb{R}^n \times B\big) \cap I \subset \{(x, x) \mid x \in \mathbb{R}^n\}.$$

Suppose now that any transport set for u is of μ measure zero. Observe that any Borel set $B \subset A^c$ is a transport set. The conclusion follows. □

Theorem 3.3.3. *Let $m > 1$. There exists an absolutely continuous measure $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$ for which there is no optimal transport π such that*

$$\mathrm{P}_1\|\pi\| \text{ is absolutely continuous with respect to the Lebesgue measure } \lambda.$$

Moreover, for such μ, there exists a transport set associated with the optimal potential of μ with non-zero measure μ.

Proof. Choose any $v_1, \dots, v_{m+1} \in \mathbb{R}^m$ such that

$$\sum_{i=1}^{m+1} v_i = 0$$

and that are affinely independent. For $\epsilon > 0$ set

$$\mu_\epsilon = \frac{1}{\lambda(B(0, \epsilon))} \sum_{i=1}^{m+1} \lambda|_{B(x_i,\epsilon)} v_i,$$

where $x_1,\dots,x_{m+1}\in\mathbb{R}^n$ are pairwise distinct points to be specified later. Here λ denotes the Lebesgue measure on $\mathbb{R}^n$. Then $\mu_\epsilon\in\mathcal{W}(\mathbb{R}^n,\mathbb{R}^m)$. Suppose that there exist optimal transports $\pi_k\in\Gamma(\mu_{\epsilon_k})$ such that

$$\mathrm{P}_1\|\pi_k\|\ll\lambda,$$

where $(\epsilon_k)_{k=1}^\infty$ is some sequence converging to zero. Then by Theorem 3.2.19 we have

$$\mu_{\epsilon_k}(A_k)=0$$

for any transport set A_k of u_k, where $u_k\colon\mathbb{R}^n\to\mathbb{R}^m$ is an optimal potential of μ_{ϵ_k}. For $k\in\mathbb{N}$ and $i=1,\dots,m+1$ consider the union N_{ik} of all non-trivial leaves of u_k – i.e. of dimension at least one – that intersect $B(x_i,\epsilon_k)$. Then N_{ik} is a transport set. Its Borel measurability follows by Lemma 2.5.4 and Lemma 2.5.5. Indeed, denote $B=B(x_i,\epsilon_k)$; then the function

$$\mathbb{R}^n\setminus B\ni x\mapsto\sup\left\{\frac{\|u_k(x)-u_k(y)\|}{\|x-y\|}\mid y\in B\right\}\in\mathbb{R}$$

is lower semicontinuous and therefore

$$N_{ik}=\left\{x\in\mathbb{R}^n\setminus B\mid\sup\left\{\frac{\|u_k(x)-u_k(y)\|}{\|x-y\|}\mid y\in B\right\}=1\right\}\cup\left\{x\in B\mid\beta_1(x)>0\right\}$$

is a Borel set. Thus $\mu_{\epsilon_k}(N_{ik})=0$. Hence,

$$\sum_{j=1}^{m+1}v_j\lambda(B(x_j,\epsilon_k)\cap N_{ik})=0.\tag{3.3.2}$$

As μ_{ϵ_k}, by Lemma 3.3.2, is concentrated on non-trivial leaves of u_k, we have for

$$\frac{\lambda(B(x_i,\epsilon_k)\cap N_{ik})}{\lambda(B(0,\epsilon_k))}v_i=\mu_{\epsilon_k}(B(x_i,\epsilon_k)\cap N_{ik})=\mu_{\epsilon_k}(B(x_i,\epsilon_k))=v_i.$$

By (3.3.2) and assumption on the vectors $v_1,\dots,v_{m+1}$

$$\lambda(B(x_j,\epsilon_k)\cap N_{ik})=\lambda(B(0,\epsilon_k))\text{ for all }j=1,\dots,m+1.$$

Thus we infer that for any $k\in\mathbb{N}$ and for all $r,s=1,\dots,m+1$, $r\neq s$, there exist points

$$(x_{rs}^k,x_{sr}^k)\in B(x_r,\epsilon_k)\times B(x_s,\epsilon_k)$$

such that

$$\|u_k(x_{rs}^k)-u_k(x_{sr}^k)\|=\|x_{rs}^k-x_{sr}^k\|.$$

Using Arzelà–Ascoli theorem and passing to a subsequence we may assume that maps u_k converge locally uniformly to some 1-Lipschitz map u_0. Observe now that

$$x_{rs}^k\text{ converges to }x_r\text{ for all }r,s=1,\dots,m+1.$$

Thus, by the locally uniform convergence, u_0 is an isometry on $\{x_1,\dots,x_{m+1}\}$. Observe that

$$\mu_{\epsilon_k} \text{ converges weakly* to } \mu_0=\sum_{i=1}^{m+1}\delta_{x_i}v_i.$$

Now, Lemma 3.3.1 tells us that u_0 is an optimal potential of μ_0.

Suppose now that points $x_1,\dots,x_{m+1}$ are such that for $i\neq j$, $i,j=1,\dots,m$,

$$\Big\langle \frac{x_i-x_{m+1}}{\|x_i-x_{m+1}\|},\frac{x_j-x_{m+1}}{\|x_j-x_{m+1}\|}\Big\rangle < \Big\langle \frac{v_i}{\|v_i\|},\frac{v_j}{\|v_j\|}\Big\rangle. \tag{3.3.3}$$

Then if we define $h\colon\{x_1,\dots,x_{m+1}\}\to\mathbb{R}^m$ by

$$h(x_{m+1})=0,\ h(x_i)=\|x_i-x_{m+1}\|\frac{v_i}{\|v_i\|} \text{ for } i=1,\dots,m,$$

then h is 1-Lipschitz. By the Kirszbraun theorem we may assume that h is defined on the entire space. Moreover for

$$\pi=\sum_{i=1}^{m+1}v_i\delta_{(x_i,x_{m+1})}$$

we have

$$\mathrm{P}_1\pi-\mathrm{P}_2\pi=\mu_0$$

and

$$\pi=\sum_{i=1}^{m}\frac{h(x_i)-h(x_{m+1})}{\|x_i-x_{m+1}\|}\|v_i\|\delta_{(x_i,x_{m+1})}$$

Theorem 3.2.16 yields that h is an optimal potential and π is an optimal transport. It follows that

$$\|\mu_0\|_{\mathcal{W}(\mathbb{R}^n,\mathbb{R}^m)}=\sum_{i=1}^{m}\|v_i\|\|x_i-x_{m+1}\|.$$

Theorem 3.2.16 tells us that also

$$\pi=\sum_{i=1}^{m}\frac{u_0(x_i)-u_0(x_{m+1})}{\|x_i-x_{m+1}\|}\|v_i\|\delta_{(x_i,x_{m+1})}$$

As u_0 is an isometry on $\{x_1,\dots,x_{m+1}\}$, it follows that for $i,j=1,\dots,m$

$$\|h(x_i)-h(x_j)\|=\|x_i-x_j\|$$

which is not true, as the inequalities in (3.3.3) are strict. The obtained contradiction shows that there is no such sequence $(\epsilon_k)_{k=1}^{\infty}$, i.e. there exists $\epsilon_0>0$ such that for all $\epsilon\in(0,\epsilon_0)$ there is no optimal transport with absolutely continuous marginal for μ_ϵ and such that there exists a transport set with non-zero measure μ_ϵ. $\square$

The proof of the following theorem is based on the same idea as the proof of Theorem 3.3.3. Note that we do not require below that the norms on $\mathbb{R}^n$ and on $\mathbb{R}^m$ are Euclidean. For a 1-Lipschitz map $u\colon \mathbb{R}^n \to \mathbb{R}^m$ a leaf of u is a maximal, with respect to the order induced by inclusion, set $\mathcal{S}$ such that $u|_{\mathcal{S}}$ is an isometry. A transport set is defined as a set A that enjoys the property that if $x \in A$ belongs to a unique leaf of u, then for any $y \in \mathbb{R}^n$ such that $\|u(y) - u(x)\| = \|y - x\|$ there is $y \in A$. This is to say, the leaves and transport sets are defined as in the Euclidean case.

Theorem 3.3.4. *Let $m \leq n$. Suppose that the norm on $\mathbb{R}^m$ is strictly convex. Suppose that $\mathcal{F}$ is a locally uniformly closed subset of 1-Lipschitz maps of $\mathbb{R}^n$ to $\mathbb{R}^m$. Suppose that $\mathcal{F}$ has the property that for any absolutely continuous measure $\mu \in \mathcal{W}(\mathbb{R}^n, \mathbb{R}^m)$ and any $\mathcal{F}$-optimal potential u_0 of μ we have $\mu(A) = 0$ for any transport set of u_0. Then either $m = 1$ or $m > 1$ and*

i) $m = n$, any $u \in \mathcal{F}$ is affine, and there exists $u \in \mathcal{F}$ that is an isometry of $\mathbb{R}^n$ and of $\mathbb{R}^m$,

ii) for any absolutely continuous μ any $\mathcal{F}$-optimal potential of μ is an isometry on a maximal subspace $V \subset \mathbb{R}^n$, so that

$$\mu(\{x \in \mathbb{R}^n \mid Px \in A\}) = 0 \text{ for any Borel set } A \subset W; \tag{3.3.4}$$

here P denotes a projection onto a complement W of V.

Suppose that the norms are Euclidean. Then, if any $\mathcal{F}$-optimal potential is affine and is an isometry on a maximal subspace $V \subset \mathbb{R}^n$ such that (3.3.4) holds true, then $\mu(A) = 0$ for any transport set of its $\mathcal{F}$-optimal potential.

Proof. Suppose that $m > 1$. Choose any pairwise distinct points $x_1, x_2, x_3 \in \mathbb{R}^n$ and any affinely independent $v_1, v_2, v_3 \in \mathbb{R}^m$ such that $\sum_{i=1}^3 v_i = 0$. Let

$$\nu_0 = \sum_{i=1}^3 v_i \delta_{x_i}.$$

Then $\nu_0 \in \mathcal{M}_0(\mathbb{R}^n, \mathbb{R}^m)$. For $\epsilon > 0$ let

$$\nu_\epsilon = \frac{1}{\lambda(B(0,\epsilon)}\sum_{i=1}^3 v_i \lambda|_{B(x_i,\epsilon)}$$

Choose respective $\mathcal{F}$-optimal potentials u_ϵ for ν_ϵ. These exist as $\mathcal{F}$ is locally uniformly closed. Observe that, by the assumption, $\nu_\epsilon(B_\epsilon) = 0$ for any Borel set consisting of trivial leaves of u_ϵ. Whence, ν_ϵ is concentrated on non-trivial leaves of u_ϵ. Let $N_{i\epsilon}$ denote the union of all non-trivial leaves that intersect $B(x_i, \epsilon)$ and are not contained in $B(x_i, \epsilon)$ for $i = 1, 2, 3$ and $\epsilon > 0$. The map

$$\mathbb{R}^n \setminus K \ni x \mapsto \sup\big\{\frac{\|u(x) - u(y)\|}{\|x - y\|} \mid y \in K\big\} \in \mathbb{R}$$

is lower semicontinuous for any set $K \subset \mathbb{R}^n$. Note that $N_{i\epsilon} = M_{i\epsilon} \cup K_{i\epsilon}$, where

$$M_{i\epsilon} = \Big\{x \in \mathbb{R}^n \setminus B(x_i, \epsilon) \mid \sup\Big\{\frac{\|u(x) - u(y)\|}{\|x - y\|} \mid y \in B(x_i, \epsilon)\Big\} = 1\Big\},$$

and

$$K_{i\epsilon} = \Big\{x \in B(x_i, \epsilon) \mid \sup\Big\{\frac{\|u(x) - u(y)\|}{\|x - y\|} \mid y \in \mathbb{R}^n \setminus B(x_i, \epsilon)\Big\} = 1\Big\}.$$

Hence $M_{i\epsilon}$ and $K_{i\epsilon}$ are Borel measurable and so is $N_{i\epsilon}$.

By the assumption,

$$\nu_\epsilon(N_{i\epsilon}) = 0,$$

which implies, as in the proof of Theorem 3.3.3, that

$$\|u_\epsilon(x^\epsilon_{rs}) - u_\epsilon(x^\epsilon_{sr})\| = \|x^\epsilon_{rs} - x^\epsilon_{sr}\|$$

for some points

$$(x^\epsilon_{rs}, x^\epsilon_{sr}) \in B(x_r, \epsilon) \times B(x_s, \epsilon), r, s = 1, \ldots, 3, r \neq s.$$

By the Arzelà–Ascoli theorem and passing to a subsequence we may assume that u_ϵ converges locally uniformly to some $u_0 \in \mathcal{F}$, which is an $\mathcal{F}$-optimal potential of ν_0 by Lemma 3.3.1. By the uniform convergence we infer that u_0 is isometric on $\{x_1, x_2, x_3\}$. Let now $x_2 = tx_1 + (1 - t)x_3$ for some $t \in (0, 1)$. Then any map $f \colon \{x_1, x_2, x_3\} \to \mathbb{R}^m$ that is isometric satisfies

$$f(tx_1 + (1 - t)x_3) = tf(x_1) + (1 - t)f(x_3). \tag{3.3.5}$$

Indeed, as f is isometric,

$$\|f(x_2) - f(x_1)\| = (1 - t)\|x_3 - x_1\| \text{ and } \|f(x_3) - f(x_2)\| = t\|x_3 - x_1\|.$$

As $\|f(x_3) - f(x_1)\| = \|x_3 - x_1\|$ it follows that we have equality in the triangle inequality

$$\|f(x_3) - f(x_1)\| \leq \|f(x_2) - f(x_1)\| + \|f(x_3) - f(x_2)\|.$$

By the strict convexity it follows that there is $\lambda > 0$ such that

$$f(x_2) - f(x_1) = \lambda(f(x_3) - f(x_1)).$$

Taking the norms we arrive at (3.3.5). A function f that satisfies (3.3.5) may be extended to $\mathbb{R}^n$ to an affine map that has derivative of operator norm at most m. This follows by the Hahn–Banach theorem. As u_0 is isometric on $\{x_1, x_2, x_3\}$, we infer that

$$\sum_{i=1}^{3}\langle u_0(x_i), v_i\rangle \ \leq \sup\Big\{\sum_{i=1}^{3}\langle f(x_i), v_i\rangle \mid f\colon \mathbb{R}^n \to \mathbb{R}^m \text{ is linear and } \|f\| \leq m\Big\}$$

Note now that the set of vectors v_1, v_2, v_3 that sum up to zero and are affinely independent is dense in the set of vectors v_1', v_2', v_3' that sum up to zero. Moreover, u_0 is an $\mathcal{F}$-optimal potential for ν_0. We conclude that for any $u \in \mathcal{F}$ and any vectors v_1, v_2, v_3 that sum up to zero there is

$$\sum_{i=1}^{3}\langle u(x_i), v_i\rangle \ \leq \sup\Big\{\sum_{i=1}^{3}\langle f(x_i), v_i\rangle \mid f\colon \mathbb{R}^n \to \mathbb{R}^m \text{ is linear and } \|f\| \leq m\Big\}$$

Take now $v_2 = v$, $v_1 = -tv$ and $v_3 = -(1-t)v$ with $t \in (0,1)$ as above and any $v \in \mathbb{R}^m$. We infer that

$$\langle u(x_2) - tu(x_1) - (1-t)u(x_3), v\rangle \leq 0.$$

As this holds for any v we infer that u is affine. This proves part of i).

If u is affine then there exists a subspace $V \subset \mathbb{R}^n$, possibly trivial, i.e. $V = \{0\}$, such that any set of the form

$$\{x \in \mathbb{R}^n \mid Px \in A\} \tag{3.3.6}$$

for a Borel measurable set $A \subset W$ is a transport set of u. Here P denotes a projection onto a complement W of V. Indeed, let $V \subset \mathbb{R}^n$ be a maximal subspace such that $u|_V$ is an isometry. Suppose that V is not a leaf of u. Then there exists $y \notin V$ such that for all $x \in V$

$$\|u(y) - u(x)\| = \|y - x\|.$$

It follows that for all non-zero $\lambda \in \mathbb{R}$

$$\Big\|u(y) - u\Big(\frac{x}{\lambda}\Big)\Big\| = \Big\|y - \frac{x}{\lambda}\Big\|$$

for all $x \in V$. Hence for all $\lambda \in \mathbb{R}$ we have $\|u(\lambda y) - u(x)\| = \|\lambda y - x\|$. As u is affine, it is also an isometry on $V + \mathbb{R}y$. This contradiction shows that V is a leaf of u. Thus ii) is proven.

We shall now provide an example of a vector measure μ such that for any proper subspace V and any x_0 there is $c > 0$ such that

$$\mu\Big(\big\{x \in \mathbb{R}^n \mid \|P(x - x_0)\| \leq c\big\}\Big) \neq 0. \tag{3.3.7}$$

Choose any $x_1, \dots, x_{m+1} \in \mathbb{R}^n$ affinely independent. Let $\epsilon > 0$ be a number such that any $y_i \in B(x_i, \epsilon)$, $i = 1, \dots, m+1$ are affinely independent. Choose vectors $v_1, \dots, v_{m+1} \in \mathbb{R}^m$ that add up to zero and are affinely independent. Let

$$\mu = \sum_{i=1}^{m+1} v_i \lambda|_{B(x_i,\epsilon)},$$

where λ denotes the Lebesgue measure. Choose any proper affine subspace $V \subset \mathbb{R}^n$. Then V intersects at most m of the balls $B(x_i, \epsilon)$, $i = 1, \dots, m+1$. So does the set

$$\left\{ x \in \mathbb{R}^n \mid \|P(x - x_0)\| \leq c \right\}$$

provided that $c > 0$ is sufficiently small. Thus (3.3.7) follows. It implies, by ii), that $V = \mathbb{R}^n$. We have shown that any $\mathcal{F}$-optimal potential of μ has to be an isometry. Hence $m = n$ and the proof of i) is complete.

To prove the last part of the theorem, it is enough to prove that the translates of V are the only leaves of an affine map. This holds true, since any point in $\mathbb{R}^n$ is covered by a translate of V. $\square$

Chapter 4

Continuity of extensions of Lipschitz maps

4.1 Introduction

Here we describe the structure of the chapter. In Section 4.2 we study the optimal rates of continuity for extensions of 1-Lipschitz maps. Example 4.2.2, Proposition 4.2.3 and Proposition 4.2.4 discuss the optimality of the rate.

In Section 4.3 we discuss several cases where it is possible to find an extension of a 1-Lipschitz map with preserved uniform distance to a given map. In Section 4.3.1 we prove that such extension always exists whenever the target space is of dimension one and the given map is 1-Lipschitz. In Section 4.3.2 we deal with one-dimensional perturbations of a given map. In Section 4.3.3 we study a related problem of extending maps such their increments are majorised by the increments of another map. Section 4.3.4, considers a situation when the map, to which we want preserve the uniform distance of the extension, is affine. We prove in Theorem 4.3.9 that, if the target space is a Hilbert space of dimension at least two, then a map is affine and 1-Lipschitz if and only if for any 1-Lipschitz map there is a 1-Lipschitz extension with preserved uniform distance to the given map. One implication of this equivalence establishes a strengthening of Kirszbraun's theorem. For the proof we use the technique of K-functions developed in [84]. This shows a striking difference with the one-dimensional case, where every 1-Lipschitz map v has the above property, as Lipschitz functions are closed under minima and maxima. Another result, presented in Section 4.3.5, concerns continuous extensions of 1-Lipschitz maps from a subset A of a Hilbert space to another subset X that contains A. Suppose that $v\colon X \to \mathbb{R}^m$ is such that

$$\|v(x) - t_1 v(x_1) - \ldots t_m v(x_m)\| \leq \|x - t_1 x_1 - \ldots - t_m x_m\|$$

for all $x, x_1, \ldots, x_m \in X$ and all non-negative $t_1, \ldots, t_m$ that sum up to one. In Theorem 4.3.11 we prove that there exists, for any given 1-Lipschitz map on A, its 1-Lipschitz extension to X such that its uniform distance to v is preserved.

4.2 Sharp rate of continuity of extensions of Lipschitz maps

Let $A \subset B \subset \mathbb{R}^n$, $n \in \mathbb{N}$. In this section we shall prove that given any 1-Lipschitz maps $v\colon \mathbb{R}^n \to \mathbb{R}^m$, for $m \in \mathbb{N}$, and $u\colon A \to \mathbb{R}^m$, such that

$$\sup\{\|u(x) - v(x)\| \mid x \in A\} \leq \delta,$$

there exists a 1-Lipschitz extension $\tilde{u}\colon B \to \mathbb{R}^m$ of u, that is $\tilde{u}(x) = u(x)$ for $x \in A$, such that

$$\sup\{\|v(x) - \tilde{u}(x)\| \mid x \in B\} \leq \sqrt{\delta^2 + 2\delta d_v(A, B)}. \tag{4.2.1}$$

Here by $d_v(A, B)$ we denote the number

$$\sup\{\|v(x) - v(y)\| \mid x \in A, y \in B\}. \tag{4.2.2}$$

Note that for 1-Lipschitz functions v we have $d_v(A, B) \leq \operatorname{diam}(B)$. We shall also give an example of functions u, v such that the bound is attained. This is to say, u, v are such that for any 1-Lipschitz extension $\tilde{u}$ of u we have equality in (4.2.1). Moreover, as we shall show, we cannot hope, in general, for any bound, if $d_v(A, B)$ is infinite.

The proof of the following proposition follows from the proof of [71, Lemma 2.1].

Proposition 4.2.1. *Let $A \subset B \subset \mathbb{R}^n$ and let*

$$u\colon A \to \mathbb{R}^m,\ v\colon B \to \mathbb{R}^m$$

be 1-Lipschitz maps. Assume that $\|u(x) - v(x)\| \leq \delta$ for $x \in A$. Then there exists a 1-Lipschitz map $\tilde{u}\colon B \to \mathbb{R}^m$ such that $\tilde{u}(x) = u(x)$ for $x \in A$ and

$$\|v(x) - \tilde{u}(x)\| \leq \sqrt{\delta^2 + 2\delta d_v(A, B)}$$

for all $x \in B$.

Proof. Let $\epsilon^2 = \delta^2 + 2\delta d_v(A, B)$. Let us define a map

$$h\colon B \times \{0\} \cup A \times \{\epsilon\} \to \mathbb{R}^m$$

by the formulae $h(x, 0) = v(x)$ for $x \in B$ and $h(x, \epsilon) = u(x)$ for $x \in A$. Then h is a 1-Lipschitz map on a subset of $\mathbb{R}^{n+1}$. Indeed, if $x \in A$ and $y \in B$, then

$$\begin{aligned}
&\|h(y, 0) - h(x, \epsilon)\|^2 = \|v(y) - u(x)\|^2 = \\
&= \|v(y) - v(x)\|^2 + \|v(x) - u(x)\|^2 + 2\langle v(y) - v(x), v(x) - u(x)\rangle \leq \\
&\leq \|x - y\|^2 + \delta^2 + 2\delta d_v(A, B) = \|x - y\|^2 + \epsilon^2.
\end{aligned}$$

For other points of the domain of h the 1-Lipschitz condition follows from 1-Lipschitzness of u and v.

Using Theorem 1.4.1 we may extend h to a 1-Lipschitz map $\tilde{h}\colon \mathbb{R}^{n+1} \to \mathbb{R}^m$. Set for $x \in B$, $\tilde{u}(x) = \tilde{h}(x, \epsilon)$. Then $\tilde{u}$ is a 1-Lipschitz extension of u and moreover, for $x \in B$,

$$\|v(x) - \tilde{u}(x)\| = \|\tilde{h}(x, 0) - \tilde{h}(x, \epsilon)\| \leq \epsilon.$$

□

Let us now exhibit an example which shows that the bound may be attained.

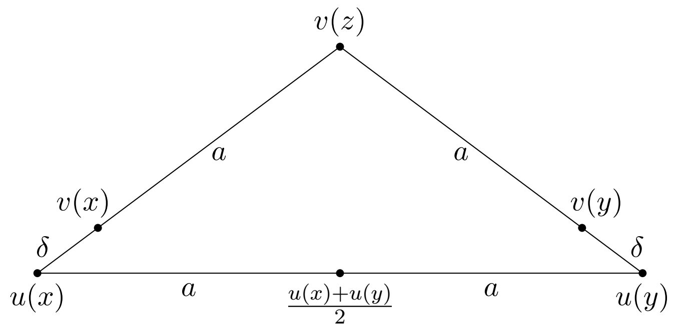
Example 4.2.2. Let $m > 1$ and let $x, y \in \mathbb{R}^n$, $x \neq y$, $z = \frac{x+y}{2}$. Let $a = \|x - z\| = \|y - z\|$, let $\delta > 0$. Define $u\colon \{x, y\} \to \mathbb{R}^m$ by setting $u(x)$ and $u(y)$ so that $\|u(x) - u(y)\| = \|x - y\|$. Map u defined in this way is 1-Lipschitz. For the definition of v consider the triangle whose vertices are $u(x), u(y)$ and a point, called $v(z)$, such that

$$\|v(z) - u(x)\| = \|v(z) - u(y)\| = a + \delta.$$

Set $v(x), v(y)$ to be the points on the triangle's edges containing $u(x)$ and $u(y)$ respectively such that $\|v(x) - u(x)\| = \delta$ and $\|v(y) - u(y)\| = \delta$. If we define $v\colon \{x, y, z\} \to \mathbb{R}^m$ in this manner, then it is 1-Lipschitz. By Kirszbraun's theorem we may extend it to $\mathbb{R}^n$ in such a way that the extension is still 1-Lipschitz. We shall call this extension $v\colon \mathbb{R}^n \to \mathbb{R}^m$. Moreover, $\sup\{\|u(t) - v(t)\| \mid t \in A\} = \delta$. Here $A = \{x, y\}$. Observe that any 1-Lipschitz extension $\tilde{u}$ of u to the point z must satisfy $\tilde{u}(z) = \frac{u(x)+u(y)}{2}$, by Lemma 2.2.4. Thus, if we set $B = \{x, y, z\}$, then any 1-Lipschitz extension $\tilde{u}$ of u to B satisfies

$$\|v(z) - \tilde{u}(z)\| = \sqrt{\delta^2 + 2\delta a}.$$

The situation is illustrated below.



Note now that

$$a = d_v(A, B) \text{ if } \delta \geq a \text{ and } a = \frac{1}{4}d_v(A, B) + \sqrt{\frac{1}{16}d_v(A, B)^2 + \frac{1}{2}d_v(A, B)\delta} \text{ if } \delta \leq a.$$

This exhibits that the bound (4.2.1) is indeed sharp, if $\delta \geq d_v(A,B)$. Note that $\delta \leq a$ if and only if $\delta \leq d_v(A,B)$. Hence we have shown that for all extensions $\tilde{u}$ of u we have

$$\sup\{\|v(z) - \tilde{u}(z)\| \mid z \in B\} = \sqrt{\delta^2 + 2\delta d_v(A,B)}$$

if $\delta \geq d_v(A,B)$ and

$$\sup\{\|v(z) - \tilde{u}(z)\| \mid z \in B\} = \sqrt{\delta^2 + \frac{1}{2}\delta d_v(A,B) + \delta\sqrt{\frac{1}{4}d_v(A,B)^2 + 2d_v(A,B)\delta}}$$

if $\delta \leq d_v(A,B)$.

The following proposition shows that (4.2.1) is asymptotically sharp for δ approaching zero, up to a multiplicative constant.

Proposition 4.2.3. *Let $m > 1$ and let $(\delta_k)_{k\in\mathbb{N}}$ be a sequence of positive numbers and let $a > 0$. Then there exists a 1-Lipschitz map $v\colon \mathbb{R}^n \to \mathbb{R}^m$ such that for $k \in \mathbb{N}$ there exist sets $A_k, B_k \subset \mathbb{R}^n$, 1-Lipschitz maps $u_k\colon A_k \to \mathbb{R}^m$ such that*

$$\sup\left\{\|v(x) - u_k(x)\| \mid x \in A_k\right\} = \delta_k$$

and for any 1-Lipschitz extensions $\tilde{u}_k$ of u_k to B_k we have

$$\sup\left\{\|v(x) - \tilde{u}_k(x)\| \mid x \in B_k\right\} = \sqrt{\delta_k^2 + 2\delta_k a}.$$

Proof. For any δ_k let us construct a map $v_k\colon A_k \to \mathbb{R}^m$ as in Example 4.2.2, with $a \in \mathbb{R}$ independent of k. Let B_k be the corresponding set. We may appropriately shift sets A_k and B_k so that the resulting function $v\colon \bigcup_{k\in\mathbb{N}} B_k \to \mathbb{R}^m$, $v|_{B_k} = v_k$, is 1-Lipschitz. By Kirszbraun's theorem we may assume that v is defined on $\mathbb{R}^n$. Consider a map $u_k\colon A_k \to \mathbb{R}^m$, constructed as in Example 4.2.2. Then for any 1-Lipschitz extension $\tilde{u}_k$ of u_k to B_k we have

$$\sup\{\|v(z) - \tilde{u}_k(z)\| \mid z \in B_k\} = \sqrt{\delta_k^2 + 2\delta_k a}.$$

□

Proposition 4.2.4. *Let $\delta > 0$. There exist two sets $A \subset B \subset \mathbb{R}^m$ and 1-Lipschitz maps $v\colon \mathbb{R}^n \to \mathbb{R}^m$ and $u\colon A \to \mathbb{R}^m$ such that*

$$\sup\left\{\|v(x) - u(x)\| \mid x \in A\right\} = \delta \text{ and } d_v(A,B) = \infty$$

and for any 1-Lipschitz extension $\tilde{u}$ of u to B

$$\sup\left\{\|v(x) - \tilde{u}(x)\| \mid x \in B\right\} = \infty.$$

Proof. Define maps $u\colon A \to \mathbb{R}^m$ and $v\colon B \to \mathbb{R}^m$ by reproducing, as in Proposition 4.2.3, countably many times triangles of Example 4.2.2, with respective parameters $(a_k)_{k\in\mathbb{N}}$ converging to infinity and fixed $\delta > 0$. Then

$$\sup\big\{\|v(x) - u(x)\| \mid x \in A\big\} = \delta \text{ and } d_v(A, B) = \infty.$$

Moreover, for any 1-Lipschitz extension $\tilde{u}$ of u to B we have

$$\sup\big\{\|v(x) - \tilde{u}(x)\| \mid x \in B\big\} \geq \sup\{\sqrt{\delta^2 + 2\delta a_k} \mid k \in \mathbb{N}\} = \infty.$$

□

This shows that if the parameter $d_v(A, B)$ is infinite, then the corresponding parameter

$$\inf\big\{\sup\big\{\|\tilde{u}(x) - v(x)\| \mid x \in B\big\} \mid \tilde{u}\colon B \to \mathbb{R}^m \text{ is 1-Lipschitz extension of } u\big\}$$

may be infinite as well.

4.3 Examples of good approximability

Let us now turn to examples of situations, in which we can prove that, for 1-Lipschitz maps $v\colon X \to \mathbb{R}^m$ and $u\colon A \to \mathbb{R}^m$, $A \subset X$, if

$$\|v(x) - u(x)\| \leq \delta \text{ for all } x \in A,$$

then it is possible to extend u to a 1-Lipschitz map $\tilde{u}$ such that

$$\|v(x) - \tilde{u}(x)\| \leq \delta \text{ for all } x \in X.$$

4.3.1 One-dimensional target space

Let us first consider the case when the target space is one-dimensional.

Proposition 4.3.1. *Let X be a metric space. Let $v\colon X \to \mathbb{R}$ be a 1-Lipschitz function. Then for any set $A \subset X$ and for any 1-Lipschitz function $u\colon A \to \mathbb{R}$ such that for all $x \in A$*

$$|u(x) - v(x)| \leq \delta, \tag{4.3.1}$$

there exists 1-Lipschitz extension $\tilde{u}\colon X \to \mathbb{R}$ of v such that for all $x \in X$

$$|v(x) - \tilde{u}(x)| \leq \delta. \tag{4.3.2}$$

Proof. Take any 1-Lipschitz extension $\tilde{u}_0\colon X \to \mathbb{R}$ of u. Existence of such function follows from McShane's formula (see [80]). Define now [1]

$$\tilde{u}(x) = \tilde{u}_0(x) \wedge (v(x) + \delta) \vee (v(x) - \delta).$$

Then it is readily verifiable that $\tilde{u}$ satisfies the desired properties. □

[1] Here, symbols $a \wedge b$ and $a \vee b$ stand for minimum and maximum of two real numbers a, b respectively.

4.3.2 One-dimensional perturbations

Our next example extends result of Section 4.3.1 and concerns 1-Lipschitz maps $v\colon \mathbb{R}^n \to \mathbb{R}^m$ and $u\colon A \to \mathbb{R}^m$ such that $v(x) - u(x) \in \mathbb{R}w$ for some fixed $w \in \mathbb{R}^m$ and all $x \in A$. We shall need to use below a Riemannian pseudo-metric given by the formula

$$d_v^w(x,y) = \inf\Big\{ \int_a^b \|\dot z(t)\|_{g_v^w(z(t))} dt \mid z \in \mathcal{C}^1([a,b],\mathbb{R}^n), z(a)=x, z(b)=y \Big\}. \tag{4.3.3}$$

Here

$$\|\dot z(t)\|^2_{g_v^w(z(t))} = g_v^w(z(t))(\dot z(t), \dot z(t)) = \|\dot z(t)\|^2 - \|D(v\circ z)(t))\|^2 + |\langle w, D(v\circ z)(t)\rangle|^2,$$

is a square of the length of a vector $\dot z(t)$ with respect to the degenerate inner product g_v^w given by

$$g_v^w(x)(s,t) = \langle s,t\rangle - \langle Dv(x)s, Dv(x)t\rangle + \langle w, Dv(x)s\rangle\langle w, Dv(x)t\rangle.$$

Observe that for any $z \in \mathcal{C}^1([a,b],\mathbb{R}^n)$ the composition $v\circ z\colon [a,b]\to\mathbb{R}^m$ is a Lipschitz function. By Rademacher's theorem (see e.g. [45]) it is differentiable almost everywhere and thus the integrals in (4.3.3) are well defined.

Below we will speak of 1-Lipschitzness with respect to the Euclidean metric and with respect to the d_v^w pseudo-metric. If not stated explicitly, we consider 1-Lipschitzness with respect to the Euclidean metric.

Lemma 4.3.2. *For $x, y \in \mathbb{R}^n$ define*

$$d(x,y) = \sqrt{\|x-y\|^2 - \|v(x)-v(y)\|^2 + \langle w, v(x)-v(y)\rangle^2}.$$

Then

$$d_v^w(x,y) \le d(x,y) \text{ for all } x,y\in\mathbb{R}^n.$$

Proof. Choose $x, y \in \mathbb{R}^n$. Set $z\colon [0,1]\to\mathbb{R}^n$ to be given by

$$z(t) = x + t(y-x) \text{ for } t\in[0,1].$$

Let P denote the orthogonal projection in $\mathbb{R}^m$ onto the space orthogonal to w. Then

$$d_v^w(x,y) \le \int_0^1 \sqrt{\|x-y\|^2 - \|D(Pv\circ z)(t)\|^2} dt.$$

We apply Jensen's inequality twice, exploiting concavity of the square root and convexity of the norm squared. This yields

$$d_v^w(x,y) \le \sqrt{\|x-y\|^2 - \Big\| \int_0^1 D(Pv\circ z)(t)dt \Big\|^2}.$$

Hence

$$d_v^w(x,y) \le \sqrt{\|x-y\|^2 - \|Pv(x)-Pv(y)\|^2}.$$

This completes the proof. □

Lemma 4.3.3. *Let d be as above. Then for all $x, y \in \mathbb{R}^n$*

$$d_v^w(x,y) = \inf\Big\{\sum_{i=1}^{l-1} d(x_i, x_{i+1}) \mid x_1 = x, x_l = y, x_i \in \mathbb{R}^n, i = 1,\dots,l, l \in \mathbb{N}\Big\}. \tag{4.3.4}$$

Proof. It is easy verifiable that d_v^w satisfies triangle inequality. By Lemma 4.3.2, we infer that the left-hand side of (4.3.4) is at most the right hand-side of (4.3.4). To prove the converse inequality let $\epsilon > 0$ and take a path $z \in \mathcal{C}^1([a,b],\mathbb{R}^n)$, such that $z(a) = x, z(b) = y$ and

$$d_v^w(x,y) > \int_a^b \|\dot z(t)\|_{g_v^w(z(t))} dt - \frac{1}{2}\epsilon. \tag{4.3.5}$$

For $k \in \mathbb{N}$ set $(s_i^k)_{i=0}^{2^k} \subset [a,b]$ to be $s_i^k = a + (b-a)\frac{i}{2^k}$, $i = 0,\dots,2^k$ and consider a function r_k on $[a,b]$ defined by

$$r_k(t) = \sum_{i=0}^{2^k-1} \mathbf{1}_{[s_i^k, s_{i+1}^k)}(t)\Bigg(\bigg(\frac{\|z(s_{i+1}^k) - z(s_i^k)\|}{s_{i+1}^k - s_i^k}\bigg)^2 - \bigg(\frac{\|v(z(s_{i+1}^k)) - v(z(s_i^k))\|}{s_{i+1}^k - s_i^k}\bigg)^2 + \\ + \bigg(\frac{\langle w, v(z(s_{i+1}^k)) - v(z(s_i^k))\rangle}{s_{i+1}^k - s_i^k}\bigg)^2\Bigg)^{\frac{1}{2}}.$$

Then we see that the corresponding functions r_k are uniformly bounded and converge with k converging to infinity to $\|\dot z(\cdot)\|_{g_v^w(z(\cdot))}$ in any point of differentiability of $v \circ z$, hence almost everywhere. Therefore, by Lebesgue's dominated convergence theorem, for k sufficiently large

$$\int_a^b \|\dot z(t)\|_{g_v^w(z(t))} dt > \int_a^b r_k(t)dt - \frac{1}{2}\epsilon. \tag{4.3.6}$$

Observe that $\int_a^b r_k(t)dt = \sum_{i=0}^{2^k-1} d(z(s_{i+1}^k), z(s_i^k))$. Combining (4.3.5) and (4.3.6) we get

$$d_v^w(x,y) \ge \sum_{i=0}^{2^k-1} d(z(s_{i+1}^k), z(s_i^k)) - \epsilon.$$

Thus

$$d_v^w(x,y) \ge \inf\Big\{\sum_{i=1}^{l-1} d(x_i, x_{i+1}) \mid x_1 = x, x_l = y, x_i \in \mathbb{R}^n, i = 1,\dots,l, l \in \mathbb{N}\Big\} - \epsilon.$$

As this holds true for any $\epsilon > 0$, the proof is complete. □

Theorem 4.3.4. *Let $A \subset \mathbb{R}^n$ and let $w \in \mathbb{R}^m$ be a unit vector. Let $v\colon \mathbb{R}^n \to \mathbb{R}^m$ and $u\colon A \to \mathbb{R}^m$ be 1-Lipschitz maps such that*

$$v(x) - u(x) \in \mathbb{R}w \tag{4.3.7}$$

for all $x \in A$. Then there exists a 1-Lipschitz extension $\tilde{u}\colon \mathbb{R}^n \to \mathbb{R}^m$ of u such that $v(x) - \tilde{u}(x) \in \mathbb{R}w$ for all $x \in \mathbb{R}^n$ if and only if

$$|\langle u(x) - u(y), w\rangle| \leq d_v^w(x, y) \tag{4.3.8}$$

for all $x, y \in A$, i.e. if $\langle u, w\rangle$ is 1-Lipschitz with respect to the pseudo-metric d_v^w. Moreover, if condition (4.3.8) is satisfied, then there exists a 1-Lipschitz extension $\tilde{u}$ of u such that for all $x \in \mathbb{R}^n$

$$\|\tilde{u}(x) - v(x)\| \leq \sup\big\{\|u(x) - v(x)\| \mid x \in A\big\}.$$

Proof. Define $t\colon A \to \mathbb{R}$ by $\langle u(x), w\rangle = t(x)$ for $x \in A$. Assuming $v(x) - u(x) \in \mathbb{R}w$ for all $x \in A$, 1-Lipschitzness of u is equivalent to that

$$\big|t(x) - t(y)\big|^2 \leq \|x - y\|^2 - \|v(x) - v(y)\|^2 + \langle w, v(x) - v(y)\rangle^2 \tag{4.3.9}$$

for all $x, y \in A$. This is an immediate consequence of the Pythagorean theorem. Let, as before, for $x, y \in \mathbb{R}^n$

$$d^2(x, y) = \|x - y\|^2 - \|v(x) - v(y)\|^2 + \langle w, v(x) - v(y)\rangle^2.$$

Assume that u may be extended to a 1-Lipschitz function $\tilde{u}\colon \mathbb{R}^n \to \mathbb{R}^m$ such that the condition (4.3.7) holds true for all $x \in \mathbb{R}^n$. Then, by (4.3.9), we have, for all choices of points $x_0, \ldots, x_l \in \mathbb{R}^n$ such that $x_0 = x, x_l = y$,

$$\big|t(x) - t(y)\big| \leq \sum_{i=0}^{l-1} \big|t(x_{i+1}) - t(x_i)\big| \leq \sum_{i=0}^{l-1} d(x_{i+1}, x_i). \tag{4.3.10}$$

Lemma 4.3.3 shows now that (4.3.8) holds true.

Conversely, if (4.3.8) holds true for all $x, y \in A$, then we may extend $t\colon A \to \mathbb{R}$ to a 1-Lipschitz function $\tilde{t}$, with respect to the d_v^w pseudo-metric, on $\mathbb{R}^n$. Such an extension is provided by McShane's formula (see [80])

$$\tilde{t}(x) = \inf\{t(y) + d_v^w(x, y) \mid y \in A\}.$$

If we know that $|\langle v(x) - u(x), w\rangle| \leq \delta$ for $x \in A$, i.e. $|t(x) - \langle v(x), w\rangle| \leq \delta$, then setting instead

$$\tilde{t}(x) = \inf\{t(y) + d_v^w(x, y) \mid y \in A\} \wedge (\langle v(x), w\rangle + \delta) \vee (\langle v(x), w\rangle - \delta)$$

gives a 1-Lipschitz extension, with respect to d_v^w, such that

$$|\tilde{t}(x) - \langle v(x), w\rangle| \leq \delta$$

for all $x \in \mathbb{R}^n$. Now, as $\tilde{t}$ is 1-Lipschitz with respect to d_v^w the function $\tilde{u} = v - \langle v, w\rangle w + \tilde{t}w$ is 1-Lipschitz extension that we wanted to find, see (4.3.9). □

Let us remark, that 1-Lipschitzness of u with respect to the Euclidean metric implies its 1-Lipschitzness with respect to the pseudo-metric d_v^w, provided that the set A is geodesically convex, i.e. if for any two points $x, y \in A$ the infimum in the definition of the distance $d_v^w(x, y)$ is realised by paths lying in the set A.

4.3.3 Increments majorisation

In what follows we shall use the following theorem of Minty (see [84]), which encompasses several Kirszbraun's type theorems.

Definition 4.3.5. Let Y be a real vector space and X be a set. A real function on Y is called finitely lower semicontinuous if its restriction to any finitely-dimensional subspace of Y is lower semicontinuous. A function $\Phi\colon Y \times X \times X \to \mathbb{R}$ is called a K-function if Φ is both finitely lower semicontinuous and convex in the first variable and such that for any points $(y_i, x_i)_{i=1}^l \in Y \times X$ and any $\lambda_1, \ldots, \lambda_l \geq 0$ such that $\sum_{i=1}^l \lambda_i = 1$ we have

$$\sum_{i,j=1}^{l} \lambda_i\lambda_j\Phi(y_i - y_j, x_i, x_j) \geq 2\sum_{i=1}^{l} \lambda_i\Phi(y_i - \sum_{j=1}^{l} \lambda_j y_j, x_i, x) \tag{4.3.11}$$

for all $x \in X$. If Y is finite-dimensional, then it is enough to assume that $l \leq \mathrm{dim}Y + 1$.

Theorem 4.3.6. *Let Y be a real vector space and X be a set. Let $\Phi\colon Y \times X \times X \to \mathbb{R}$ be a K-function. Let*

$$(y_i, x_i)_{i=1}^l \subseteq Y \times X$$

be a sequence such that

$$\Phi(y_i - y_j, x_i, x_j) \leq 0$$

for all $i, j = 1, \ldots, l$. Let $x \in X$. Then there exists a vector $y \in Y$ such that

$$\Phi(y_i - y, x_i, x) \leq 0$$

for all $i = 1, \ldots, l$. Furthermore, y may be chosen to lie in $\mathrm{Conv}\{y_1, \ldots, y_l\}$.

Let us mention that the proof of the above theorem relies on von Neumann's minimax theorem.

Let now Y be a Hilbert space and let X be a set and let $A \subset X$. Our next example concerns the extension of a map $u\colon A \to Y$ that has the property that

$$\|u(x) - u(y)\| \leq \|v(x) - v(y)\|,$$

for all $x, y \in A$. Here $v\colon X \to Y$ is a map, which we would like to stay close to $\tilde{u}$ – an extension of u. The following proposition holds true.

Proposition 4.3.7. *Let X be a set. Assume that $v\colon X \to Y$. Let $A \subset X$ and let $u\colon A \to Y$ satisfy*

$$\|u(x) - u(y)\| \leq \|v(x) - v(y)\| \tag{4.3.12}$$

for all $x, y \in A$. Then there exists an extension $\tilde{u}$ of u to X such that (4.3.12) holds for all $x, y \in X$. Moreover, there exists an extension $\tilde{u}$ such that (4.3.12) holds true for all $x, y \in X$ and such that for all $x \in X$

$$\|v(x) - \tilde{u}(x)\| \leq \sup\big\{\|v(z) - u(z)\| \mid z \in A\big\}.$$

Proof. Let us define $\Phi\colon Y \times Y \times Y \to \mathbb{R}$ by the formula

$$\Phi(y, x, x') = \|y\|^2 + 2\langle y, x - x'\rangle.$$

We claim that this is a K-function. Indeed, it is convex and continuous in the first variable. We have to check the condition (4.3.11). Let then $(y_i, x_i)_{i=1}^l \in Y \times Y$ and $\lambda_1, \ldots, \lambda_l$, $x \in Y$, be as in the definition of a K-function. A short calculation readily implies that

$$\sum_{i.j=1}^{l} \lambda_i \lambda_j \|y_i - y_j\|^2 - 2\sum_{i=1}^{l} \lambda_i \|y_i - \sum_{j=1}^{l} \lambda_j y_j\|^2 = 0$$

and that

$$\sum_{i.j=1}^{l} \lambda_i \lambda_j \langle y_i - y_j, x_i - x_j\rangle - 2\sum_{i=1}^{l} \lambda_i \langle y_i - \sum_{j=1}^{l} \lambda_j y_j, x_i - x\rangle = 0.$$

Thus we have an equality in (4.3.11). Choose now points $t_1, \ldots, t_k \in A$ and let $t \in X \setminus A$. Let $w\colon A \to Y$ be defined by $w = u - v$. By (4.3.12) we know that

$$\|w(t_i) - w(t_j)\|^2 + 2\langle w(t_i) - w(t_j), v(t_i) - v(t_j)\rangle \leq 0.$$

That is

$$\Phi(w(t_i) - w(t_j), v(t_i), v(t_j)) \leq 0$$

for all $i, j = 1, \ldots, k$. By Theorem 4.3.6 there exists a point $y \in \mathrm{Conv}\{w(t_1), \ldots, w(t_k)\}$, which we shall call $w(t)$, such that

$$\Phi(w(t_i) - w(t), v(t_i), v(t)) \leq 0$$

for all $i, j = 1, \ldots, k$. Thus, if we define $u(t) = w(t) + v(t)$, then u, on the set $\{t, t_1, \ldots, t_k\}$, has increments majorised by v and $\|u(t) - v(t)\| \leq \delta$, provided that $\|u(t_i) - v(t_i)\| \leq \delta$ for all $i = 1, \ldots, k$.

This implies that for any choice of points $t_i \in A$ and any $t \in X$ the intersection of closed balls

$$\bigcap_{i=1}^{k} B(u(t_i), \|v(t_i) - v(t)\|) \tag{4.3.13}$$

is non-empty. By compactness such intersection is non-empty also for any infinite family of balls; in particular we may intersect over all points in A. Any point in the intersection yields the desired extension of u to the point t.

To finish, let us partially order by inclusion all subsets of X that admit an extension of u and contain A. By the Kuratowski–Zorn lemma, there exists a maximal element Z of this ordering. If $Z \neq X$ then by the procedure above, we may extend u to an extra point of X, contradicting the choice of Z. Thus $Z = X$ and the proof is complete.

The continuity part of the theorem follows by considering intersection of closed balls of the form

$$\bigcap_{i=1}^{k} B(u(t_i), \|v(t_i) - v(t)\|) \cap B(v(t), \delta)$$

instead of (4.3.13). □

Corollary 4.3.8. *Assume that $u\colon A \to Y$ is a 1-Lipschitz map on a subset A of Y. Let $T\colon Y \to Y$ be any isometry. Then there exists a 1-Lipschitz extension $\tilde{u}\colon Y \to Y$ such that*

$$\sup\{\|\tilde{u}(y) - T(y)\| \mid y \in Y\} = \sup\{\|u(y) - T(y)\| \mid y \in A\}.$$

Proof. Apply Proposition 4.3.7. □

4.3.4 Affine maps

Let us now consider the case when the target space is a Hilbert space Y. The theorem below shows that if the dimension of Y is at least two, then the situation differs strikingly from the one-dimensional case, c.f. Section 4.3.1.

Theorem 4.3.9. *Let X, Y be real Hilbert spaces such that Y is of dimension at least two. Let $v\colon X \to Y$ be a map. The following conditions are equivalent:*

i) for any $A \subset X$ and for any 1-Lipschitz map $u\colon A \to Y$ there exists a 1-Lipschitz extension $\tilde{u}\colon X \to Y$ of u such that for all $x \in X$

$$v(x) - \tilde{u}(x) \in \overline{\mathrm{Conv}}\{v(z) - u(z) \mid z \in A\}.$$

ii) for any $\delta > 0$, any $A \subset X$ and for any 1-Lipschitz map $u\colon A \to Y$ such that for all $x \in A$

$$\|v(x) - u(x)\| \le \delta,$$

there exists 1-Lipschitz extension $\tilde{u}\colon X \to Y$ of u such that for all $x \in X$

$$\|v(x) - \tilde{u}(x)\| \le \delta,$$

iii) v is affine and 1-Lipschitz.

Proof. That i) implies ii) is trivial. Suppose that ii) holds true. Take any $x \in X$ and let $A = \{x\}$. Set $u(x) = v(x)$. Then $u\colon A \to Y$ is 1-Lipschitz and

$$\|v(x) - u(x)\| \le \delta \text{ for any } x \in A \text{ and any } \delta > 0.$$

By ii), there exist 1-Lipschitz maps $u_\delta\colon X \to Y$ such that

$$\|v(x) - u_\delta(x)\| \le \delta \text{ for all } x \in X.$$

Thus for any $x, y \in X$

$$\|v(x) - v(y)\| \le \|v(x) - u_\delta(x)\| + \|v(y) - u_\delta(y)\| + \|x - y\| \le 2\delta + \|x - y\|.$$

As this holds true for any $\delta > 0$, we see that v is 1-Lipschitz.

Our aim is now to show that for any $x, y \in X$ there is

$$v\Big(\frac{x+y}{2}\Big) = \frac{1}{2}(v(x) + v(y)).$$

For this, take some $x, y \in X$ such that $v(x) \neq v(y)$ and let $z = \frac{x+y}{2}$. Let

$$w = \frac{v(x) - v(y)}{\|v(x) - v(y)\|}.$$

Let r be a unit vector perpendicular to w lying in a tangent space to any two-dimensional affine space containing the points $v(x)$, $v(y)$ and $v(z)$. Let

$$h = \|v(x) - v(y)\|$$

and $\lambda, \mu \in \mathbb{R}$ be such that

$$v(z) - v(x) = \lambda r + \mu w.$$

We claim that

$$v(z) = \lambda r + \frac{1}{2}(v(y) + v(x)). \tag{4.3.14}$$

Let $\delta \in \mathbb{R}$ and set

$$u(x) = v(x) - \delta w \text{ and } u(y) = v(y) - \delta(\alpha r + \beta w) \text{ with } \alpha^2 + \beta^2 = 1.$$

Then

$$\|u(x) - u(y)\|^2 = 2(1-\beta)(\delta^2 - \delta h) + h^2.$$

Observe that if $\|x - y\| = h$, then Lemma 2.2.4 implies that v is affine on the line segment $[x, y]$. We may thus assume that $h < \|x - y\|$. Set $\gamma = \frac{\|x-y\|^2 - h^2}{2}$. Then for any $\delta > h$ we pick $\beta \in (0, 1)$ such that

$$\|u(x) - u(y)\| = \|x - y\|.$$

This β is given by

$$\beta = 1 - \frac{\gamma}{\delta^2 - \delta h}. \tag{4.3.15}$$

It is positive provided that δ is sufficiently large. Let $A = \{x, y\}$ and $B = \{x, y, z\}$. Observe that map $u\colon A \to Y$ is 1-Lipschitz and for all $p \in A$

$$\|u(p) - v(p)\| = \delta.$$

Lemma 2.2.4 implies that any 1-Lipschitz extension $\tilde{u}$ of u to B satisfies

$$\tilde{u}(z) = \frac{u(x) + u(y)}{2}.$$

Hence for such an extension

$$\|v(z) - \tilde{u}(z)\|^2 = \frac{1}{4}\Big(\big(2\lambda + \delta\alpha\big)^2 + \big(h + 2\mu + \delta(1+\beta)\big)^2\Big).$$

If $\delta > h$, then $\|v(z) - \tilde{u}(z)\| > \delta$ if and only if

$$4\delta^2(2\mu + h + \lambda\alpha) + \delta\big(-4\lambda\alpha h - 4h(h + 2\mu) - 2\gamma + 4\lambda^2 + (h + 2\mu)^2\big) + c > 0, \tag{4.3.16}$$

where

$$c = -h(h + 2\mu)^2 - 4h\lambda^2 - 2\gamma(h + 2\mu)$$

is independent of δ and α. Indeed, this follows by expansion of the squares and a rearrangement that takes (4.3.15) into account. Suppose that $2\mu + h > \epsilon$ for some $\epsilon > 0$. Observe that, with the choice (4.3.15), α tends to 0 as δ tends to infinity. Let $\delta_0 > 0$ be such that $|\lambda\alpha| < \frac{1}{2}\epsilon$ for $\delta > \delta_0$. Pick any $\delta > \delta_0 \vee h$ such that

$$4\delta^2\Big(2\mu + h - \frac{1}{2}\epsilon\Big) + \delta\big(-2\epsilon h - 4h(h + 2\mu) - 2\gamma + 4\lambda^2 + (h + 2\mu)^2\big) + c > 0.$$

Then δ satisfies also (4.3.16). This contradicts the assumption on v. Therefore $2\mu + h \leq 0$. This is to say

$$\langle v(z) - v(x), v(x) - v(y)\rangle \leq -\frac{1}{2}\|v(x) - v(y)\|^2.$$

Repeating the above argument with x and y interchanged yields

$$\langle v(y) - v(z), v(x) - v(y)\rangle \leq -\frac{1}{2}\|v(x) - v(y)\|^2.$$

If we add the above inequalities, we get an equality. Thus, there are equalities in both of them. Hence, as $v(x) - v(y) = hw$ and $v(z) - v(x) = \lambda r + \mu w$, we get $2\mu + h = 0$. We have proven (4.3.14). We now also include the case $v(x) = v(y)$; then we take r to be a unit vector in direction parallel to $v(z) - v(x)$. Then (4.3.14) still holds true.

We claim now that $\lambda = 0$. Suppose conversely that $\lambda \neq 0$. We may also suppose that $\lambda > 0$; otherwise we change r to $-r$. For $\rho, \eta \in (0, 1)$ and such that $\rho^2 + \eta^2 = 1$ set

$$\nu(x) = v(x) + \delta(\rho w - \eta r) \text{ and } \nu(y) = v(y) + \delta(-\rho w - \eta r).$$

Then $\nu\colon A \to Y$ satisfies $\|v(p) - \nu(p)\| = \delta$ for all $p \in A$. We choose parameters δ, ρ and η so that

$$\|\nu(x) - \nu(y)\| = \|x - y\|,$$

that is we put

$$h + 2\rho\delta = \|x - y\|. \tag{4.3.17}$$

Then by Lemma 2.2.4 any 1-Lipschitz extension $\tilde{\nu}$ of ν to B satisfies

$$\tilde{\nu}(z) = \frac{1}{2}\nu(x) + \frac{1}{2}\nu(y) = \frac{v(x) + v(y)}{2} - \delta\eta r = v(z) - \lambda r - \delta\eta r.$$

Then

$$\|v(z) - \tilde{\nu}(z)\| = \lambda + \delta\eta.$$

Let $\delta > \lambda$. Then this quantity, given (4.3.17), is greater than δ if and only if

$$\delta > \frac{\zeta^2 + \lambda^2}{2\lambda}, \text{ where } \zeta = \frac{\|x - y\| - h}{2}.$$

This is to say, if δ is big enough, then there exists a 1-Lipschitz function ν that contradicts the assumption on v. Hence $\lambda = 0$ and thus for any $x, y \in X$

$$v(z) = \frac{v(x) + v(y)}{2}, \text{ where } z = \frac{x + y}{2}.$$

Now, v is continuous and standard arguments imply that v is affine.

To prove that iii) implies i) consider first a function $\Phi\colon Y \times X \times X \to \mathbb{R}$, given by

$$\Phi(y, x, x') = \|y\|^2 + 2\langle y, v(x) - v(x')\rangle - \|x - x'\|^2 + \|v(x) - v(x')\|^2.$$

Let us check that it is a K-function. The condition of convexity and finitely lower semicontinuity is clearly satisfied. We need only to check whether the condition (4.3.11) holds. It is readily seen that the first two summands in the definition of Φ both satisfy the condition (4.3.11) with equalities, c.f. proof of Proposition 4.3.7. Thus, to satisfy (4.3.11), we must have

$$\sum_{i,j=1}^{l} \lambda_i\lambda_j(\|v(x_i) - v(x_j)\|^2 - \|x_i - x_j\|^2) \geq 2\sum_{i=1}^{l} \lambda_i(\|v(x_i) - v(x)\|^2 - \|x_i - x\|^2)$$

for all non-negative λ_i, $i = 1, \dots, l$, summing up to one, all $x_1, \dots, x_l, x \in X$. Rearranging we get

$$\Big\| \sum_{i=1}^{l} \lambda_i x_i - x \Big\|^2 \geq \Big\| \sum_{i=1}^{l} \lambda_i v(x_i) - v(x) \Big\|^2. \tag{4.3.18}$$

As v is 1-Lipschitz and affine, this is certainly true. By Theorem 4.3.6 we see that for any points $x_1, \dots, x_k \in A$ and any $x \in X$ the intersection of closed sets

$$\bigcap_{i=1}^{k} B(u(x_i), \|x_i - x\|) \cap \Big(v(x) + \overline{\mathrm{Conv}}\{(u(z) - v(z) \mid z \in A\}\Big)$$

is non-empty. By compactness such intersection is non-empty also for any infinite number of chosen points. Therefore we may always extend u to a 1-Lipschitz map on $A \cup \{x\}$ so that

$$v(x) - u(x) \in \overline{\mathrm{Conv}}\{v(z) - u(z) \mid z \in A\}.$$

Let us order by inclusion all subsets of X containing A that admit the desired extension. By the Kuratowski–Zorn lemma, there exists a maximal subset enjoying this property. If it were not X, then by the above considerations we could find a strictly larger subset with an extension that satisfies the desired conditions. □

Remark 4.3.10. For Φ to be a K-function, the condition (4.3.18) must be valid for all $x \in X$, whence putting $x = \sum_{i=1}^{l} \lambda_i x_i$ we see immediately that v must be an affine map, for $l = 2$. Moreover, if we take $\lambda_1 = 1$ then we see that for all $x, x' \in X$

$$\|x' - x\|^2 \geq \|v(x') - v(x)\|^2;$$

i.e. v must be 1-Lipschitz.

4.3.5 Further results

Suppose that Z is a real Hilbert space and let $A \subset Z$. In this section we consider continuity of extensions of 1-Lipschitz on A to $X \subset Z$.

Theorem 4.3.11. *Let Z be a real Hilbert space. Suppose that $X \subset Z$. Suppose that $v\colon X \to \mathbb{R}^m$ is such that for all $x, x_1, \dots, x_m \in X$ and all non-negative real numbers $t_1, \dots, t_m$ that sum up to one there is*

$$\|v(x) - t_1 v(x_1) - \ldots - t_m v(x_m)\| \leq \|x - t_1 x_1 - \ldots - t_m x_m\| \tag{4.3.19}$$

Then for any $\delta > 0$, any $A \subset X$ and for any 1-Lipschitz map $u\colon A \to Y$ such that for all $x \in A$

$$\|v(x) - u(x)\| \leq \delta,$$

there exists 1-Lipschitz extension $\tilde{u}\colon X \to Y$ of u such that for all $x \in X$

$$\|v(x) - \tilde{u}(x)\| \leq \delta.$$

Proof. We shall first show that for any points $x_1, \dots, x_k \in A$ and any $x \in X$ the intersection of closed sets

$$\bigcap_{i=1}^{k} B(u(x_i), \|x_i - x\|) \cap \Big(v(x) + \overline{\mathrm{Conv}}\{(u(z) - v(z) \mid z \in A\}\Big)$$

is non-empty. By Helly's theorem (see [62]), it is enough to prove that for $k \leq m$ the corresponding intersection is non-empty and that for $k \leq m+1$ the intersection

$$\bigcap_{i=1}^{k} B(u(x_i), \|x_i - x\|)$$

is non-empty. The latter follows by Kirszbraun's theorem. To prove the former, consider a function $\Phi\colon V \times X \times X \to \mathbb{R}$, where $V = \mathrm{Aff}\{u(x_i) - v(x_i) \mid i = 1, \dots, k\}$, given by

$$\Phi(y, x, x') = \|y\|^2 + 2\langle y, v(x) - v(x')\rangle - \|x - x'\|^2 + \|v(x) - v(x')\|^2.$$

Let us check that it is a K-function. The condition of convexity and finitely lower semicontinuity is clearly satisfied. We need only to check whether the condition (4.3.11) holds. It is readily seen that the first two summands in the definition of Φ both satisfy the condition (4.3.11) with equalities, c.f. proof of Proposition 4.3.7. Thus, to satisfy (4.3.11), we must have

$$\sum_{i,j=1}^{l} \lambda_i \lambda_j (\|v(x_i) - v(x_j)\|^2 - \|x_i - x_j\|^2) \geq 2 \sum_{i=1}^{l} \lambda_i (\|v(x_i) - v(x)\|^2 - \|x_i - x\|^2)$$

for all non-negative λ_i, $i = 1, \dots, l$, summing up to one, all $x_1, \dots, x_l, x \in X$ and numbers $l \leq m$ (see Definition 4.3.5) as $\dim V \leq k - 1 \leq m - 1$. Rearranging we get

$$\Big\| \sum_{i=1}^{l} \lambda_i x_i - x \Big\|^2 \geq \Big\| \sum_{i=1}^{l} \lambda_i v(x_i) - v(x) \Big\|^2.$$

This holds true by the assumption on v.

By Theorem 4.3.6 we infer that for any points $x_1, \ldots, x_k \in A$, $k \leq m$, and any $x \in X$ the intersection of closed sets

$$\bigcap_{i=1}^{k} B(u(x_i), \|x_i - x\|) \cap \Big(v(x) + \overline{\mathrm{Conv}}\{(u(z) - v(z) \mid z \in A\} \Big) \tag{4.3.20}$$

is non-empty. Indeed, condition that $\Phi(y - y_i, x_i, x) \leq 0$ for $i = 1, \ldots, k$ and $y_i = u(x_i) - v(x_i)$, is equivalent to that for $i = 1, \ldots, k$

$$\|y - v(x) - (y_i - v(x_i))\| \leq \|x - x_i\|.$$

Hence, setting $u(x) = y + v(x)$, we see that it belongs to the set defined by formula (4.3.20).

By compactness, and by Helly's theorem, such intersection is non-empty also for any infinite number of chosen points. Therefore we may always extend u to a 1-Lipschitz map on $A \cup \{x\}$ so that

$$v(x) - u(x) \in \overline{\mathrm{Conv}}\{v(z) - u(z) \mid z \in A\}.$$

Let us order by inclusion all subsets of X containing A that admit the desired extension. By the Kuratowski–Zorn lemma, there exists a maximal subset. If it were not X, then by the above considerations we could find a strictly larger subset with an extension that satisfies the desired conditions. $\square$

Let us note that similar statement for extending maps $u\colon A \to \mathbb{R}^n$ that satisfies

$$\|u(x) - t_1 u(x_1) - \ldots - t_m u(x_m)\| \leq \|x - t_1 x_1 - \ldots - t_m x_m\| \tag{4.3.21}$$

for all $x, x_1, \ldots, x_m \in A$ and all non-negative real numbers $t_1, \ldots, t_m$ that sum up to one, to a map in $\tilde{u}\colon X \to \mathbb{R}^m$ that still satisfies analogous condition is false in general. Indeed, if $X = Z$, then such $\tilde{u}$ is necessairly affine, while there exists non-affine $u\colon A \to \mathbb{R}^m$ that satisfies (4.3.21) provided that A is affinely independent.

We conjecture that the opposite implication to that of Theorem 4.3.11 is also valid, that is if $v\colon X \to \mathbb{R}^m$ is a function such for any set $A \subset X$ and any 1-Lipschitz $u\colon A \to \mathbb{R}^m$ there exists a 1-Lipschitz map $\tilde{u}\colon X \to \mathbb{R}^m$ such that

$$\sup \big\{ \|v(x) - \tilde{u}(x)\| \mid x \in X \big\} = \sup \big\{ \|v(x) - \tilde{u}(x)\| \mid x \in A \big\},$$

then necessairly v satisfies condition (4.3.19). Note that Theorem 4.3.9 confirms this conjecture if X is the entire Hilbert space.

Chapter 5

Optimal transport and Choquet theory

5.1 Introduction

In this chapter we are concerned with study of duality theorems for optimal transport. The novelty is the treatment of the topic with use of a variant of the Strassen's theorem, Theorem 5.2.1 and its implication Theorem 5.2.3. Moreover, we provide a reinterpretation of the optimal transport problem as a variational problem concerning linear functionals over the set of all Choquet's representations of measures that are involved in the problem. The reinterpretation is introduced in the two-marginal, the multi-marginal as well as in the martingale variant of the optimal transport problem.

In Section 5.2 we recall necessary definitions and prove theorems that our results are based on.

In Section 5.3 we provide proofs of the Kantorovich duality in two-marginal case.

In Section 5.4 we prove the Kantorovich–Rubinstein duality, i.e. the duality result for cost function given by a metric.

In Section 5.5 we provide a proof of the Kantorovich duality in the multi-marginal setting.

In Section 5.6 we characterise extreme points of pairs of probability measures in convex order and prove a duality result for martingale optimal transport provided that there exists a maximiser of the dual problem.

In Section 5.7 we investigate class of functions that appear in the dual problem to the martingale optimal transport. We provide an intrinsic characterisation of such functions, which is analogous to the characterisation of convex functions as the functions, whose graph lies above its supporting tangent hyperplanes.

In Section 5.8 we apply the result of the previous section and obtain a novel characterisation of uniformly convex and uniformly smooth functions.

In Section 5.9 we introduce the notion of the martingale triangle inequality and prove that if the inequality is satisfied by a cost function, then in the dual problem to the martingale optimal transport the consideration may be restricted to pairs of equal functions.

5.2 Around Strassen's theorem

Here we shall provide a variant of Strassen's theorem. The proof is based on the proof of [81, Theorem T51]. We include it for completeness. We refer the reader also to [97] and [4].

Theorem 5.2.1. *Let X be a separable Banach space, let (Ω, Σ, μ) be a probability space. Let $\omega \mapsto h_\omega$ be a map from Ω to continuous, convex functions on X, which is weakly measurable, that is, for every $x \in X$ the map $\omega \mapsto h_\omega(x)$ is Σ-measurable, and such that there exists non-negative number c such that*

$$|h_\omega(x)| \leq c(\|x\| + 1) \text{ for all } x \in X. \tag{5.2.1}$$

Set

$$h(x) = \int_\Omega h_\omega(x) d\mu(\omega).$$

For a functional $x^* \in X^*$ the following conditions are equivalent:

i) $x^* \leq h$,

ii) there exists a map $\omega \mapsto x^*_\omega$ from Ω to X^* which is weakly measurable, in the sense that $\omega \mapsto x^*_\omega(x)$ is measurable for any $x \in X$, and such that $x^*_\omega \leq h_\omega$ for μ-almost every $\omega \in \Omega$ and for all $x \in X$

$$x^*(x) = \int_\Omega x^*_\omega(x) d\mu(\omega).$$

Proof. Observe that clearly ii) implies i). For the proof of the other implication, consider the Banach space $L^1(\Omega, X)$ of equivalence classes of measurable, integrable maps $f \colon \Omega \to X$ with norm

$$\|f\|_1 = \int_\Omega \|f(\omega)\| d\mu(\omega).$$

It's dual space $L^\infty(\Omega, X^*)$ consists of almost everywhere bounded, weakly measurable, maps with values in X^*. Define a functional λ on the subspace of constant functions by the formula

$$\lambda(x) = x^*(x).$$

By the assumption, $\lambda \leq h$, on that subspace. By the Hahn–Banach theorem for convex majorants we may extend it to a functional Λ defined on $L^1(\Omega, X)$ such that it satisfies

$$\Lambda(g) \leq \int_\Omega h_\omega(g(\omega)) d\mu(\omega).$$

Note that the right-hand side of the above inequality is well defined by the assumption (5.2.1). By the identification of the dual space, there exists a weakly measurable map x_ω^* such that

$$\Lambda(g) = \int_\Omega x_\omega^*(g(\omega))d\mu(\omega).$$

To complete the proof of the theorem we need to show that for μ-almost every ω there is $x_\omega^* \leq h_\omega$. Choose a dense subset $(x_n)_{n\in\mathbb{N}}$ in X. It is enough to show that for all $n \in \mathbb{N}$ there is $x_\omega^*(x_n) \leq h_\omega(x_n)$ for μ-almost every ω. For $n \in \mathbb{N}$ and $A \in \Sigma$ consider map $x_n\mathbf{1}_A \in L^1(\Omega, X)$. Then it follows that

$$\int_A x_\omega^*(x_n)d\mu(\omega) \leq \int_A h_\omega(x_n)d\mu(\omega).$$

The assertion of the theorem follows by standard arguments. □

Before we proceed with applications of the above modification of Strassen's theorem, let us recall definitions.

Definition 5.2.2. If (Ω, Σ) and (Ξ, Θ) are measurable spaces, then a Markov kernel P from Ω to Ξ is a real function on $\Theta \times \Omega$ such that for any point $\omega \in \Omega$, $P(\cdot, \omega)$ is a probability measure on Θ and for any $A \in \Theta$, $P(A, \cdot)$ is Σ-measurable.

If μ is a probability measure on Σ, then we define $P\mu$ to be a probability measure on Θ such that

$$P\mu(A) = \int_\Omega P(A, \omega)d\mu(\omega) \text{ for all } A \in \Theta.$$

We shall denote by $\mathcal{C}(\Omega)$ the Banach space of bounded continuous functions a topological space Ω and by $\mathcal{M}(\Omega)$ the Banach space of signed Borel measures on Ω normed by total variation. By $\mathcal{P}(\Omega)$ we shall denote the set of Borel probability measures on Ω. A subset $\mathcal{K}$ of $\mathcal{C}(\Omega)$ is said to be stable under maxima provided that $f \vee g \in \mathcal{K}$ for all functions $f, g \in \mathcal{K}$.

Theorem 5.2.3. *Let Ω be a locally compact Polish space. Let $\mathcal{K}$ be a convex set in $\mathcal{C}(\Omega)$ that is stable under maxima, contains constants, and for any constant c there is $\mathcal{K} + c \subset \mathcal{K}$. Let $f \in \mathcal{C}(\Omega)$. Suppose that μ, ν are Borel probability measures such that*

$$\int_\Omega gd(\mu - \nu) \leq \int_\Omega fd(\mu - \nu) \tag{5.2.2}$$

for all $g \in \mathcal{K}$. Then there exists a Markov kernel P form Ω to Ω such that $\nu = P\mu$ and such that for every $\omega \in \Omega$

$$\int_\Omega gd(\delta_\omega - P(\cdot, \omega)) \leq \int_\Omega fd(\delta_\omega - P(\cdot, \omega))$$

for all $g \in \mathcal{K}$. Moreover, the set of extreme points of the set of pairs of Borel probability measures (μ, ν) that satisfy (5.2.2) is contained in the set of pairs of the form (δ_ω, η) for some $\omega \in \Omega$ and some Borel probability measure η on Ω.

Proof. Set $X = \mathcal{C}(\Omega)$ to be the Banach space of all continuous bounded functions on Ω. Let x^* be an element of X^* represented by a measure $\nu \in \mathcal{M}(\Omega)$. Set for $\omega \in \Omega$

$$h_\omega(x) = \inf \big\{ (f - y)(\omega) \mid y \in \mathcal{K}, f - y \geq x \big\}.$$

Then, as $\mathcal{K}$ is convex, h_ω is convex. It is moreover continuous. Indeed, choose $x_1, x_2 \in \mathcal{C}(\Omega)$ and let $\delta = \|x_1 - x_2\|$. Take $\epsilon > 0$, $\omega \in \Omega$ and $y_1 \in \mathcal{K}$ such that

$$(f - y_1)(\omega) < h_\omega(x_1) + \epsilon \text{ and } x_1 \leq f - y_1.$$

Then $y_2 = y_1 - \delta$ belongs to $\mathcal{K}$ and satisfies

$$x_2 \leq x_1 + \delta \leq f - y_2 \text{ and } h_\omega(x_2) \leq (f - y_2)(\omega) < h_\omega(x_1) + \delta + \epsilon.$$

It follows that $h_\omega(x_2) \leq h_\omega(x_1) + \delta$. By symmetry, it follows that h_ω is 1-Lipschitz. As $\mathcal{C}(\Omega)$ is separable, so is its subset

$$\big\{ f - y \mid y \in \mathcal{K}, f - y \geq x \big\}.$$

By the assumption that $\mathcal{K}$ is stable under maxima, $\omega \mapsto h_\omega(x)$ is a pointwise limit of a sequence $(f - y_k)_{k=1}^\infty$ with $y_k \in \mathcal{K}$. We may moreover assume that

$$y_k(\omega) \geq -\|f - x\| \text{ for } \omega \in \Omega.$$

Hence $(f - y_k)(\omega) \leq \|f - x\| + \|f\|$ for $\omega \in \Omega$. By the definition of $h_\omega(x)$ it follows that $(f - y_k)(\omega) \geq -\|x\|$. By the assumption on μ, ν

$$x^*(x) = \int_\Omega x d\nu \leq \int_\Omega (f - y_k)(\omega) d\nu(\omega) \leq \int_\Omega (f - y_k)(\omega) d\mu(\omega).$$

Now, by the dominated convergence theorem it follows that

$$x^*(x) \leq \int_\Omega h_\omega(x) d\mu(\omega).$$

By the previous observations, $|h_\omega(x)| \leq 2\|f\| + \|x\|$. Thus, assumptions of Theorem 5.2.1 are satisfied. Hence there is a weakly measurable function $\omega \mapsto x^*_\omega$ with values in X^* that satisfies $x^*_\omega \leq h_\omega$ for μ-almost every $\omega \in \Omega$ and such that

$$x^*(x) = \int_\Omega x^*_\omega(x) d\mu(\omega). \tag{5.2.3}$$

Observe that $h_\omega(x) \leq \|f - x\| + \|f\|$. Thus

$$x^*_\omega(x) \leq h_\omega(x)$$

implies that for all positive λ

$$x^*_\omega(x) \le \left\| \frac{f}{\lambda} - x \right\| + \frac{\|f\|}{\lambda}.$$

Letting λ to infinity we infer that the norm of x^*_ω is at most one. Observe that if $g \in f - \mathcal{K}$, then by the definition of h_ω,

$$h_\omega(g) = g \text{ hence } x^*_\omega(g) \le g(\omega).$$

It follows that for all $h \in \mathcal{K}$ there is

$$\int_\Omega h d\delta_\omega - x^*_\omega(h) \le \int_\Omega f d\delta_\omega - x^*_\omega(f). \tag{5.2.4}$$

Putting h to be equal to constant we see that $x^*_\omega(1) = 1$. It follows that x^*_ω is non-negative functional. For if there exists non-negative $g \in \mathcal{C}(\Omega)$ such that $x^*_\omega(g) < 0$, then, assuming without loss of generality that $\|g\| \le 1$, we arrive at a contradication

$$1 = x^*_\omega(1) < x^*_\omega(1-g) \le \|1 - g\| \le 1.$$

Let x'_ω denote restriction of x^*_ω to the Banach space $\mathcal{C}_0(\Omega)$ of all continuous functions on Ω which vanish at infinity. Then x'_ω is clearly non-negative. By the Riesz' representation theorem, there exists a measure $P(\cdot, \omega)$ on Ω such that for all $h \in \mathcal{C}_0(\Omega)$ there is

$$x'_\omega(h) = \int_\Omega h dP(\cdot, \omega).$$

Choose any $h \in \mathcal{C}(\Omega)$. By Ulam's lemma and Urysohn's lemma there exists a bounded, monotone, sequence of non-negative continuous and compactly supported functions $(\phi_n)_{n\in\mathbb{N}}$ that converges pointwise to constant function 1. In view of (5.2.3),

$$\int_\Omega h\phi_n d\nu = \int_\Omega \int_\Omega h\phi_n dP(\cdot, \omega) d\mu(\omega).$$

By the dominated convergence theorem there is

$$\int_\Omega h d\nu = \int_\Omega \int_\Omega h dP(\cdot, \omega) d\mu(\omega). \tag{5.2.5}$$

It follows that $P(\cdot, \omega)$ is a probability measure for μ-almost every ω. For $h \in \mathcal{C}(\Omega)$ let $r_\omega(h) = x^*_\omega(h) - \int_\Omega h dP(\cdot, \omega)$. Then, r_ω is a non-negative functional on $\mathcal{C}(\Omega)$. Indeed, for $h \in \mathcal{C}(\Omega)$ and a sequence $(\phi_n)_{n\in\mathbb{N}}$ as above, there is

$$x^*_\omega(h) \ge x'_\omega(h\phi_n) = \int_\Omega h\phi_n dP(\cdot, \omega).$$

Taking limit, by the monotone convergence theorem, yields asserted non-negativity of r_ω. By (5.2.3) and by (5.2.5) we infer that

$$\int_\Omega r_\omega(h) d\mu(\omega) = 0 \text{ for all } h \in \mathcal{C}(\Omega).$$

Choose a countable dense subset $(h_n)_{n\in\mathbb{N}}$ of $\mathcal{C}(\Omega)$. It follows that

$$r_\omega(h_n \vee 0) = 0 \text{ and } r_\omega(h_n \wedge 0) = 0 \text{ for } \mu\text{-almost every } \omega.$$

In consequence, $r_\omega = 0$ for μ-almost every $\omega \in \Omega$. This is to say x^*_ω is integration against $P(\cdot, \omega)$ for μ-almost every ω. We see that P defines a Markov kernel from Ω to Ω. By (5.2.5), $\nu = P\mu$ and by (5.2.4) we have

$$\int_\Omega g d(\delta_\omega - P(\cdot, \omega)) \le \int_\Omega f d(\delta_\omega - P(\cdot, \omega)) \tag{5.2.6}$$

for μ-almost every $\omega \in \Omega$ and all $g \in \mathcal{K}$. To obtain the desired Markov kernel we modify P on a measurable set of μ-measure zero of $\omega \in \Omega$ such that the inequality (5.2.6) is not valid for some $g \in \mathcal{K}$ by putting $P(\cdot, \omega) = \delta_\omega$. We have

$$(\mu, \nu) = \int_\Omega (\delta_\omega, P(\cdot, \omega)) d\mu(\omega),$$

so, by (5.2.4), the claim about the extreme points follows. □

Remark 5.2.4. If we assume moreover that $\mathcal{K}$ contains a cone $\mathcal{F}$, then condition (5.2.2) implies that for all $g \in \mathcal{F}$ one has

$$\int_\Omega g d(\mu - \nu) \le 0.$$

The next corollary extends the above result to the case of pair (μ, ν) of measures on two, possibly distinct, locally compact Polish spaces X and Y.

Corollary 5.2.5. *Let X, Y be locally compact Polish spaces. Let $\mathcal{K}$ be a convex set in $\mathcal{C}(X \cup Y)$ that is stabel under maxima, contains constants and for any constant c there is $\mathcal{K} + c \subset \mathcal{K}$. Let $f \in \mathcal{C}(X \cup Y)$. Suppose that $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ are two Borel probability measures such that*

$$\int_X g d\mu - \int_Y g d\nu \le \int_X f d\mu - \int_Y f d\nu \tag{5.2.7}$$

for all $g \in \mathcal{K}$. Then there exists a Markov kernel P from X to Y such that $\nu = P\mu$ and such that for μ-almost every $x \in X$ and any $g \in \mathcal{K}$

$$\int_Y g d(\delta_x - P(\cdot, x)) \le \int_Y f d(\delta_x - P(\cdot, x)).$$

Moreover, the set of extreme points of the set of pairs of Borel probability measures that satisfy (5.2.7) is contained in the set of pairs of the form (δ_x, η) for some $x \in X$ and some Borel probability measure η on Y.

Proof. Let Ω be the disjoint union of X and Y. Let $\tilde{\mu}, \tilde{\nu}$ be the probability Borel measures in $\mathcal{M}(\Omega)$ that are extensions of μ and ν respectively. Then (5.2.7) is equivalent to condition that for all $g \in \mathcal{K}$ there is

$$\int_\Omega g d(\tilde{\mu} - \tilde{\nu}) \leq \int_\Omega f d(\tilde{\mu} - \tilde{\nu}).$$

Whence, by Theorem 5.2.3, there exists a Markov kernel $\tilde{P}$ from Ω to Ω such that

$$\tilde{\nu} = \tilde{P}\tilde{\mu}$$

and such that for every $\omega \in \Omega$ and for all $g \in \mathcal{K}$ there is

$$\int_\Omega g d(\delta_\omega - \tilde{P}(\cdot, \omega)) \leq \int_\Omega f d(\delta_\omega - \tilde{P}(\cdot, \omega)). \tag{5.2.8}$$

Let $P(A, x) = \tilde{P}(A, x)$ for $x \in X$ and for any Borel set $A \subset Y$. Then P is a Markov kernel from X to Y. For this, observe that

$$1 = \nu(Y) = \int_X \tilde{P}(Y, x) d\mu(x).$$

Hence, for μ-almost every $x \in X$, $P(\cdot, x)$ is a Borel probability measure on Y. For $x \in X$ that belong to the complement of this measurable set, we put $P(\cdot, x) = \delta_{y_0}$. It follows that for every $x \in X$ there is $P(X, x) = 0$. Moreover $\nu = P\mu$. By (5.2.8) it follows that for μ-almost every $x \in X$ and all $g \in \mathcal{K}$ there is

$$\int_Y g d(\delta_x - P(\cdot, x)) \leq \int_Y f d(\delta_x - P(\cdot, x)).$$

The claim on the extreme points follows readily. □

5.3 Optimal transport

In the following Corollary 5.2.5 is employed to prove the Kantorovich duality. The theorem below also provides a reinterpretation of the Kantorovich problem as minimisation of a linear functional over all Choquet's representation of a pair of probability measures.

Theorem 5.3.1. *Let X, Y be two locally compact Polish spaces and let $c \colon X \times Y \to \mathbb{R}$ be a bounded Lipschitz function. Let μ and ν be Borel probability measures on X and Y respectively. Then the supremum of integrals*

$$\int_X \phi d\mu - \int_Y \psi d\nu$$

taken over the set of continuous, bounded functions $\phi \in \mathcal{C}(X), \psi \in \mathcal{C}(Y)$ such that

$$\phi(x) - \psi(y) \leq c(x, y) \text{ for all } x \in X, y \in Y$$

is equal to the imfimum of integrals

$$\int_{X\times Y} c d\pi$$

over all $\pi \in \Gamma(\mu,\nu)$. *Here* $\Gamma(\mu,\nu)$ *stands for the set of all Borel probability measures on* $X \times Y$ *such that its marginals are* μ *and* ν *respectively. Moreover, both supremum and infimum are attained.*

Before we come to the proof of the above theorem, let us first show that the supremum in the statement is attained. We include the proof for completeness and refer the reader to [103] for further discussion.

Lemma 5.3.2. *There exist* $\phi_0 \in \mathcal{C}(X)$ *and* $\phi_0 \in \mathcal{C}(Y)$ *such that for all* $x \in X$ *and* $y \in Y$ *there is* $\phi_0(x) - \psi_0(y) \leq c(x,y)$ *and for all* $\phi \in \mathcal{C}(X)$ *and all* $\psi \in \mathcal{C}(Y)$ *that satisfy* $\phi(x) - \psi(y) \leq c(x,y)$ *for all* $x \in X$ *and* $y \in Y$ *there is*

$$\int_X \phi d\mu - \int_Y \psi d\nu \leq \int_X \psi_0 d\mu - \int_Y \phi_0 d\nu.$$

Proof. Suppose that $\phi \in \mathcal{C}(X), \psi \in \mathcal{C}(Y)$ satisfy

$$\phi(x) - \psi(y) \leq c(x,y) \text{ for all } x \in X, y \in Y. \tag{5.3.1}$$

For $x \in X$ put

$$\phi'(x) = \inf \big\{\psi(y) + c(x,y) \mid y \in Y\big\}$$

and $y \in Y$ put

$$\psi'(y) = \sup \big\{\phi'(x) - c(x,y) \mid x \in X\big\}.$$

Observe ϕ' and ψ' are both Lipschitz, with Lipschitz constant equal to the Lipschitz constant of c, and satisfy

$$\phi'(x) - \psi'(y) \leq c(x,y) \text{ for all } x \in X, y \in Y \text{ and } \phi \leq \phi' \text{ and } \psi' \leq \psi.$$

Then also for $x \in X$ there is

$$\phi'(x) = \inf \big\{\psi'(y) + c(x,y) \mid y \in Y\big\}.$$

Adding appropriate constant we may assume that ϕ' and ψ' are bounded. Indeed, we may assume that $\sup\{\phi'(x) \mid x \in X\} = 0$. Then we have

$$-\|c\| \leq \psi'(y) \leq -\inf\{c(x,y) \mid x \in X, y \in Y\}$$

and thus

$$\inf\{c(x,y) \mid x \in X, y \in Y\} - \|c\| \leq \phi'(x) \leq 0.$$

This is to say, we may restrict the class of functions considered in the supremum to a class that is uniformly bounded and uniformly Lipschitz. We claim that there exist functions ϕ and ψ that maximise

$$\int_X \phi d\mu - \int_Y \psi d\nu \tag{5.3.2}$$

subject to the condition (5.3.1). Indeed, pick uniformly bounded and uniformly Lipschitz sequences of admissible functions $(\phi_n)_{n\in\mathbb{N}}$ and $(\psi_n)_{n\in\mathbb{N}}$ such that the respective integrals

$$\int_X \phi_n d\mu - \int_Y \psi_n d\nu$$

converge to the supremum. As X, Y are Polish spaces, by Ulam's lemma for any $\epsilon > 0$ there exists compact sets $K_1 \subset X$ and $K_2 \subset Y$ such that $\mu(K_1^c) \leq \epsilon$ and $\nu(K_2^c) \leq \epsilon$. Now, by the Arzelà–Asoli theorem, we may assume that the sequences $(\phi_n)_{n\in\mathbb{N}}$ and $(\psi_n)_{n\in\mathbb{N}}$ converge locally uniformly to ϕ and ψ on σ-compact supports of measures μ and ν respectively. Clearly, ϕ, ψ satisfy (5.3.1). Pick $\epsilon > 0$ and corresponding compact sets K_1, K_2. Let $n \in \mathbb{N}$ be such that for all $k \geq n$ the supremum distances of ϕ_k to ϕ on K_1 and of ψ_k to ψ on K_2 is at most ϵ. For such numbers k

$$\Big| \int_X (\phi_k - \phi) d\mu - \int_Y (\psi_k - \psi) d\nu \Big|$$

is bounded by

$$2M(\mu(K_1^c) + \nu(K_2^c)) + \int_{K_1} |\phi_k - \phi| d\mu + \int_{K_2} |\psi_k - \psi| d\nu \leq (4M+2)\epsilon,$$

where M is the uniform bound on the sequences $(\phi_n)_{n\in\mathbb{N}}$ and $(\psi_n)_{n\in\mathbb{N}}$. The claim follows. □

Proof of Theorem 5.3.1. Without loss of generality we may assume that c is non-negative. Pick $\phi_0 \in \mathcal{C}(X)$ and $\psi_0 \in \mathcal{C}(Y)$ from the lemma above. Define $\rho_0 \in \mathcal{C}(X \cup Y)$ so that $\rho_0(x) = \phi_0(x)$ for $x \in X$ and $\rho_0(y) = \psi_0(y)$ for $y \in Y$. Let $\mathcal{K}$ denote the set of all bounded continuous functions ρ on $X \cup Y$ such that for $x \in X$ and $y \in Y$ there is

$$\rho(x) - \rho(y) \leq c(x, y).$$

Observe that $\mathcal{K}$ is a convex set that is stable under maxima, contains constants $\mathcal{K} + c \subset \mathcal{K}$ for all $c \in \mathbb{R}$. Moreover, for all $\rho \in \mathcal{K}$ there is

$$\int_X \rho d\mu - \int_Y \rho d\nu \leq \int_X \rho_0 d\mu - \int_Y \rho_0 d\nu.$$

By Corollary 5.2.5 the extreme points of pairs of Borel probability measures satisfying such inequality are contained in the set of pairs of the form (δ_x, η) with $x \in X$ and η a probability measure on Y. By symmetry, the set of extreme points is contained in the set of pairs of the form $(\delta_{x_0}, \delta_{y_0})$for some $x_0 \in X$ and $y_0 \in Y$. For any such extreme point $(\delta_{x_0}, \delta_{y_0})$ there is $\rho_0(x_0) - \rho_0(y_0) = c(x_0, y_0)$. Indeed, define $\rho(x) = c(x, y_0)$ for $x \in X$ and set for $y \in Y$

$$\rho(y) = \sup\{c(x, y_0) - c(x, y) \mid x \in X\}.$$

Then $\rho \in \mathcal{K}$ and $\rho(x_0) - \rho(y_0) = c(x_0, y_0)$. Thus also $\rho_0(x_0) - \rho_0(y_0) = c(x_0, y_0)$.

It follows that the considered set $\mathcal{E}$ of extreme points is equal to

$$\Big\{(\delta_x, \delta_y) \mid \rho_0(x) - \rho_0(y) = c(x, y), x \in X, y \in Y\Big\}.$$

By Choquet's theorem there is a probability measure π_0 on $\mathcal{E}$ such that

$$(\mu, \nu) = \int_{\mathcal{E}} (\xi_1, \xi_2) d\pi_0(\xi). \tag{5.3.3}$$

Define

$$\pi = \int_{\mathcal{E}} \xi_1 \otimes \xi_2 d\pi_0(\xi).$$

Then, by (5.3.3), $\pi \in \Gamma(\mu, \nu)$ and

$$\int_{X \times Y} c d\pi = \int_X \phi d\mu - \int_Y \psi d\nu$$

and the proof is complete. □

Remark 5.3.3. If $c\colon X \times Y \to \mathbb{R}$ is a lower seminontinuous function, then it may be written as a supremum of a sequence of bounded and Lipschitz functions; see e.g. [103]. Applying Theorem 5.3.1 for each function from the sequence, we may obtain the duality result for the function c.

5.4 Kantorovich–Rubinstein duality

In the present section we present a proof of the Kantorovich–Rubinstein duality analogous to the proof in the former section.

Theorem 5.4.1. *Suppose that Ω is a bounded, locally compact Polish space with metric d. Let μ and ν be Borel probability measures on Ω. Then the supremum of integrals*

$$\int_\Omega g d(\mu - \nu)$$

taken over the set of 1-Lipschitz functions $g \in \mathcal{C}(\Omega)$ is equal to the infimum of integrals

$$\int_{\Omega \times \Omega} d d\pi$$

over all $\pi \in \Gamma(\mu, \nu)$. Here $\Gamma(\mu, \nu)$ stands for the set of all Borel probability measures on $\Omega \times \Omega$ such that its marginals are μ and ν respectively. Moreover, both supremum and infimum are attained.

Proof. The fact that the supremum is attained follows by Arzelà–Ascoli theorem and by Ulam's lemma, c.f. Lemma 5.3.2. Take a 1-Lipschitz function $f\colon \Omega \to \mathbb{R}$ such that

$$\int_\Omega g d(\mu-\nu) \leq \int_\Omega f d(\mu-\nu) \tag{5.4.1}$$

for all 1-Lipschitz functions $g \in \mathcal{C}(\Omega)$. The set $\mathcal{K}$ of all 1-Lipschitz functions satisfies assumptions of Theorem 5.2.3. Observe that $\mathcal{K}$ is also stable under minima. Hence, the set of extreme points of pairs of measures that satisfy (5.4.1) is contained in the set of pairs of the form $(\delta_{x_0}, \delta_{y_0})$ for some $x_0, y_0 \in \Omega$. We claim that for any such extreme point there is

$$f(x_0) - f(y_0) = d(x_0, y_0). \tag{5.4.2}$$

Indeed, fix an extreme point $(\delta_{x_0}, \delta_{y_0})$, $x_0, y_0 \in \Omega$. Define $f_0(x) = d(x, y_0)$. Then $f_0 \in \mathcal{K}$ and $f_0(x_0) - f_0(y_0) = d(x_0, y_0)$. Then (5.4.2) follows, as

$$d(x_0, y_0) = f_0(x_0) - f_0(y_0) \leq f(x_0) - f(y_0) \leq d(x_0, y_0).$$

By the Choquet's theorem, there exists a Borel probability measure π_0 on the set of extreme points $\mathcal{E}$ such that

$$(\mu, \nu) = \int_{\mathcal{E}} \xi d\pi_0(\xi). \tag{5.4.3}$$

Define

$$\pi = \int_{\mathcal{E}} \xi_1 \otimes \xi_2 d\pi_0(\xi).$$

Then $\pi \in \Gamma(\mu, \nu)$, by (5.4.3). Moreover

$$\int_{\Omega\times\Omega} d(x,y) d\pi(x,y) = \int_{\Omega\times\Omega} \big(f(x) - f(y)\big) d\pi(x,y) = \int_\Omega f d(\mu-\nu).$$

□

5.5 Multi-marginal optimal transport

Here we generalise our approach to the multi-marginal optimal transport with finitely many marginals; see e.g. [66]. In what follows we shall need the following lemma; see also [67].

Lemma 5.5.1. *Let $X_1, \dots, X_k$ be metric spaces. Let*

$$c\colon X_1 \times \dots \times X_k \to \mathbb{R}$$

be a Lipschitz function. Let $A_i \subset X_i$ for $i = 1, \dots, k$ and let

$$f_i\colon A_i \to \mathbb{R} \text{ for } i = 1, \dots, k$$

be such that for all $x_i \in A_i$, $i = 1, \ldots, k$, there is

$$\sum_{i=1}^{k} f_i(x_i) \leq c(x_1, \ldots, x_k). \tag{5.5.1}$$

Then there exists Lipschitz functions $\tilde{f}_i \colon X_i \to \mathbb{R}$, $i = 1, \ldots, k$, such that condition (5.5.1) holds true for all $x_i \in X_i$, $i = 1, \ldots, k$. Moreover $f_i(x_i) \leq \tilde{f}_i(x_i)$ for all $x_i \in A_i$ and $i = 1, \ldots, k$. Each $\tilde{f}_i$, $i = 1, \ldots, k$, may be taken so that its Lipschitz constant is at most the Lipschitz constant of c.

Proof. We define inductively $\tilde{f}_i(x_i)$, $x_i \in X_i$, for $i = 1, \ldots, k$ as

$$\inf \Big\{ c(x_1, \ldots, x_k) - \sum_{j=1}^{i-1} \tilde{f}_j(x_j) - \sum_{j=i+1}^{k} f_j(x_j) \mid x_j \in X_j \text{ if } j < i, x_j \in A_j \text{ if } j > i \Big\}.$$

Then $\sum_{i=1}^{k} \tilde{f}_i(x_i) \leq c(x_1, \ldots, x_k)$ for $x_i \in X_i$, $i = 1, \ldots, k$, and thus

$$\tilde{f}_i(x_i) \leq \inf \Big\{ c(x_1, \ldots, x_k) - \sum_{j \neq i} \tilde{f}_j(x_j) \mid x_j \in X_j, j \neq i \Big\}.$$

Moreover $f_i \leq \tilde{f}_i$ on A_i for all $i = 1, \ldots, k$ and thus $\tilde{f}_i$ is at least the infimum on the right-hand side of the above equality. This is to say, for $x_i \in X_i$ and $i = 1, \ldots, k$

$$\tilde{f}_i(x_i) = \inf \Big\{ c(x_1, \ldots, x_k) - \sum_{j \neq i} \tilde{f}_j(x_j) \mid x_j \in X_j, j \neq i \Big\}.$$

If c was L-Lipschitz, then $\tilde{f}_i$, $i = 1, \ldots, k$ are L-Lipschitz as infima of L-Lipschitz functions. □

Remark 5.5.2. Pick $x_i \in X_i$, $i = 1, \ldots, k$. Let $f(x_1) = c(x_1, \ldots, x_k)$ and let $f(x_i) = 0$ for $i = 2, \ldots, k$. Then the assumptions of the above lemma are satisfied with $A_i = \{x_i\}$, $i = 1, \ldots, k$. Therefore we may apply the c-convexification procedure described above in the proof, to obtain functions $\tilde{f}_i \colon X_i \to \mathbb{R}$, $i = 1, \ldots, k$ such that

$$\sum_{i=1}^{k} \tilde{f}(y_i) \leq c(y_1, \ldots, y_k) \text{ for all } y_i \in X_i$$

and moreover

$$\sum_{i=1}^{k} \tilde{f}(x_i) = c(x_1, \ldots, x_k).$$

The following lemma is based on [103, Remark 1.13].

Lemma 5.5.3. *Let* $X_1, \ldots, X_k$ *be sets. Let*

$$c\colon X_1 \times \ldots \times X_k \to \mathbb{R}$$

be a bounded function. Suppose that $f_i\colon X_i \to \mathbb{R}$, $i = 1, \ldots, k$, *are such that for all* $x_i \in X_i$ *and* $i = 1, \ldots, k$

$$f_i(x_i) = \inf\Big\{c(x_1, \ldots, x_k) - \sum_{j\neq i} f_j(x_j) \mid x_j \in X_j, j \neq i\Big\}.$$

Then there exist constants $h_1, \ldots, h_k \in \mathbb{R}$ *that sum up to zero, such that the functions* $\tilde{f}_i = f_i + h_i$ *satisfy*

$$\tilde{f}_i(x_i) = \inf\Big\{c(x_1, \ldots, x_k) - \sum_{j\neq i} \tilde{f}_j(x_j) \mid x_j \in X_j, j \neq i\Big\}$$

and all of them are bounded by the uniform norm of c *times* $\max\{k, 3\}$.

Proof. Note that for any $h_1, \ldots, h_k$ that sum up to zero there is

$$\inf\Big\{c(x_1, \ldots, x_k) - \sum_{j\neq i} \tilde{f}_j(x_j) \mid x_j \in X_j, j \neq i\Big\} = f_i(x_i) - \sum_{j\neq i} h_j = \tilde{f}_i(x_i). \tag{5.5.2}$$

Thus the first assertion is proven. Let M denote the uniform norm of c. Choose $h_1, \ldots, h_k$ in such a way that

$$\sup\{\tilde{f}_i(x_i) \mid x_i \in X_i\} = M \text{ for } i = 2, \ldots, k.$$

Note that by (5.5.2) it follows that for $i = 1, \ldots, k$ and all $x_i \in X_i$

$$-M - \sum_{j\neq i} \sup\{\tilde{f}_j(x_j) \mid x_j \in X_j\} \leq \tilde{f}_i(x_i) \leq M - \sum_{j\neq i} \sup\{\tilde{f}_j(x_j) \mid x_j \in X_j\}.$$

Thus, for all $x_1 \in X_1$

$$-kM \leq \tilde{f}_1(x_1) \leq (2-k)M. \tag{5.5.3}$$

Now, again from (5.5.2) and from (5.5.3), we get that for $i = 2, \ldots, k$ and $x_i \in X_i$

$$-M - (k-2)M + (k-2)M \leq \tilde{f}_i(x_i) \leq M - (k-2)M + kM.$$

Hence, for such indices i,

$$-M \leq \tilde{f}_i(x_i) \leq 3M.$$

□

The following theorem provides a novel interpretation of Kantorovich problem in the multi-marginal setting as minimisation of a certain linear functional over the set of all Choquet's representations of k-tuples of probability measures $(\mu_1, \ldots, \mu_k) \in \mathcal{P}$.

Theorem 5.5.4. *Let $X_1, \ldots, X_k$ be locally compact Polish spaces. Let $c\colon X_1 \times \ldots \times X_k \to \mathbb{R}$ be a bounded Lipschitz function. Let μ_i be a Borel probability measure on X_i for each $i = 1, \ldots, k$. Then the supremum of sum of integrals*

$$\sum_{i=1}^{k} \int_{X_i} f_i d\mu_i$$

taken over the set of continuous, bounded functions $f_i \in \mathcal{C}(X_i)$, $i = 1, \ldots, k$, such that

$$\sum_{i=1}^{k} f_i(x_i) \leq c(x_1, \ldots, x_k) \text{ for all } x_i \in X_i, i = 1, \ldots, k$$

is equal to the infimum of integrals

$$\int_{X_1 \times \ldots \times X_k} c d\pi$$

over all $\pi \in \Gamma(\mu_1, \ldots, \mu_k)$. Here $\Gamma(\mu_1, \ldots, \mu_k)$ stands for the set of all Borel probability measures on $X_1 \times \ldots \times X_k$ such that its marginals on X_i are μ_i for $i = 1, \ldots, k$. Moreover, both supremum and infimum are attained.

Lemma 5.5.5. *Let $X_1, \ldots, X_k$ be locally compact Polish spaces. Let $c\colon X_1 \times \ldots \times X_k \to \mathbb{R}$ be a non-negative bounded Lipschitz function,. Let $\mathcal{L}$ denote the set of all bounded continuous functions $g \in \mathcal{C}(X_1 \cup \ldots \cup X_k)$ such that for all $x_i \in X_i$, $i = 1, \ldots, k$, we have*

$$\sum_{i=1}^{k} g(x_i) \leq c(x_1, \ldots, x_k).$$

Let $f \in \mathcal{L}$. Then the set of extreme points of the set $\mathcal{P}$ of k-tuples of Borel probability measures $(\mu_1, \ldots, \mu_k) \in \mathcal{P}(X_1) \times \ldots \times \mathcal{P}(X_k)$ such that

$$\sum_{i=1}^{k} \int_{X_i} g d\mu_i \leq \sum_{i=1}^{k} \int_{X_i} f d\mu_i$$

for all $g \in \mathcal{L}$ is equal to the set of k-tuples of the form $(\delta_{x_1}, \ldots, \delta_{x_k})$, such that

$$\sum_{i=1}^{k} f(x_i) = c(x_1, \ldots, x_k).$$

Proof. For any $l \in \{1, \ldots, k\}$ let $I_l = \{1, \ldots, l-1, l+1, \ldots, k\}$ and let Ω be the disjoint union of all X_i, $i = 1, \ldots, k$. Let $(\mu_1, \ldots, \mu_k) \in \mathcal{P}$. We shall denote by $\tilde{\mu}_l$ the extension of μ_l to Ω. Let μ_{I_l} denote the probability measure on Ω given by

$$\mu_{I_l} = \frac{1}{k-1} \sum_{i \neq l} \tilde{\mu}_i.$$

Then, for any $g \in \mathcal{L}$, we have

$$\int_\Omega g d\mu_l + \int_\Omega (k-1) g d\mu_{I_l} \leq \int_\Omega f d\mu_l + \int_\Omega (k-1) f d\mu_{I_l}.$$

Denote by X_{I_l} the disjoint union of X_i, $i \in I_l$. Let $\mathcal{L}_l$ denote the convex set of all continuous bounded functions on Ω which are equal to g on X_l and to $-(k-1)g$ on X_{I_l} for some $g \in \mathcal{L}$. Then, $\mathcal{L}_l$ is stable under maxima, contains constants for any constant t there is $t + \mathcal{L}_l \subset \mathcal{L}_l$. Moreover for any $h \in \mathcal{L}_l$ there is

$$\int_{X_l} h d\mu_l - \int_{X_{I_l}} h d\mu_{I_l} \leq \int_{X_l} f d\mu_l - \int_{X_{I_l}} (1-k) f d\mu_{I_l}.$$

By Corollary 5.2.5, the extreme points of the set $\mathcal{P}_l$ of pairs of Borel probability measures $(\mu, \nu) \in \mathcal{P}(X_l) \times \mathcal{P}\big(\bigcup_{i \in I_l} X_i\big)$ such that

$$\int_{X_l} h d\mu - \int_{X_{I_l}} h d\nu \leq \int_{X_l} f d\mu - \int_{X_{I_l}} (1-k) f d\nu$$

for all $h \in \mathcal{L}_l$ are of the form (δ_x, η) for some probability $\eta \in \mathcal{P}\big(\bigcup_{i \in I_l} X_i\big)$. By the Choquet's theorem there exists a Borel probability measure π_l on the set $\mathcal{E}_l$ of extreme points of $\mathcal{P}_l$ such that

$$(\mu_l, \mu_{I_l}) = \int_{\mathcal{E}_l} \xi d\pi_l(\xi).$$

Hence for any $i \in I_l$

$$\mu_i = \int_{\mathcal{E}_l} (k-1) \xi_2|_{X_i} d\pi_l(\xi).$$

Here we write $\xi = (\xi_1, \xi_2)$ for $\xi \in \mathcal{E}_l$. It follows that π_l-almost all $(k-1)\xi_2|_{X_i}$ are probabilities. We may write

$$(\mu_1, \dots, \mu_l, \dots, \mu_k) = \int_{\mathcal{E}_l} \big((k-1)\xi_2|_{X_1}, \dots, \xi_1, \dots, (k-1)\xi_2|_{X_k}\big) d\pi_l(\xi).$$

Observe that for π_l-almost every ξ there is

$$\big((k-1)\xi_2|_{X_1}, \dots, \xi_1, \dots, (k-1)\xi_2|_{X_k}\big) \in \mathcal{P}.$$

Hence, any extreme point of $\mathcal{P}$ has to be of the form

$$(\eta_1, \dots, \eta_{l-1}, \delta_{x_l}, \eta|_{X_{l+1}}, \dots, \eta|_{X_k})$$

with $x_l \in X_l$ and some probability measures η_i for $i \in I_l$. As this holds true for any $l = 1, \dots, k$, any extreme point of $\mathcal{P}$ has to have the form $(\delta_{x_1}, \dots, \delta_{x_k})$ with $x_i \in X_i$, $i = 1, \dots, k$.

Take now any extreme point $(\delta_{x_1}, \dots, \delta_{x_k})$ of $\mathcal{P}$ and let $f \in \mathcal{L}$ be as in the statement of the lemma. Then for any $g \in \mathcal{L}$ we have

$$\sum_{i=1}^{k} g(x_i) \leq \sum_{i=1}^{k} f(x_i). \tag{5.5.4}$$

By Remark 5.5.2 there exists a function $g \in \mathcal{L}$ such that

$$\sum_{i=1}^{k} g(x_i) = c(x_1, \dots, x_k).$$

By (5.5.4) it follows that also

$$\sum_{i=1}^{k} f(x_i) = c(x_1, \dots, x_k).$$

The proof is complete. □

Proof of Theorem 5.5.4. The fact that the supremum is attained follows by Lemmata 5.5.1, 5.5.3, by Ulam's lemma and by Arzelà–Ascoli theorem, c.f. Lemma 5.3.2.

The assertion follows from Lemma 5.5.5 and by the Choquet's theorem, c.f. Theorem 5.3.1. Indeed, if π_0 is a Borel probability measure on the set of extreme points $\mathcal{E}$ of $\mathcal{P}$ of the previous lemma then an optimal $\pi \in \Gamma(\mu_1, \dots, \mu_k)$ is given by the formula

$$\pi = \int_{\mathcal{E}} \xi_1 \otimes \dots \otimes \xi_k d\pi_0(\xi),$$

where $\xi = (\xi_1, \dots, \xi_k) \in \mathcal{E}$. □

Remark 5.5.6. If $c\colon X_1 \times \dots \times X_k \to \mathbb{R}$ is a lower seminontinuous function, then it may be written as a supremum of a sequence of bounded and Lipschitz functions; see e.g. [103]. Applying Theorem 5.5.4 for each function from the sequence, we may obtain duality result for the function c.

5.6 Martingale optimal transport

We shall characterise the set of extreme points of two Borel probability measures μ, ν in convex order. Recall that two Borel probability measures μ, ν on $\mathbb{R}^n$ with finite first moments are said to be in convex order provided that

$$\int_{\mathbb{R}^n} g d\mu \leq \int_{\mathbb{R}^n} g d\nu$$

for all convex functions $g\colon \mathbb{R}^n \to \mathbb{R}$. Recall also that a set $K \subset \mathbb{R}^n$ is called a convex body provided that it is convex, compact and has non-empty interior. We say that points

$x_1, \dots, x_d \in \mathbb{R}^n$ are affinely independent provided that none of these points lies in the affine hull of the others.

In the proof below we could have used a result of Winkler [107]. Instead we follow a direct approach for the sake of completeness and clarity.

Theorem 5.6.1. *Let $K \subset \mathbb{R}^n$ be a convex body. Let $\mathcal{F}$ denote the set of continuous convex functions on K. Let $\mathcal{P}$ denote the set of pairs (μ, ν) of Borel probability measures on K that are in convex order, that is*

$$\int_K g d(\mu - \nu) \leq 0$$

for all $g \in \mathcal{F}$. Then the set of extreme points of $\mathcal{P}$ is equal to the set of pairs of the form

$$\left(\delta_x, \sum_{i=1}^{d+1} \lambda_i \delta_{x_i}\right) \tag{5.6.1}$$

where $x = \sum_{i=1}^{d+1} \lambda_i x_i$, $\lambda_i > 0$ for $i = 1, \dots, d+1$ and $\sum_{i=1}^{d+1} \lambda_i = 1$, $d \leq n$ and moreover $x_1, \dots, x_{d+1} \in K$ are affinely independent.

Proof. By Theorem 5.2.3, any extreme point of $\mathcal{P}$ is of the form (δ_x, η) for some $x \in K$ and some Borel probability measure η on K. Moreover, as any affine function belongs to $\mathcal{F}$, we see that

$$x = \int_K y d\eta(y).$$

Let us fix $x \in K$. Consider the set $\mathcal{A}$ of all Borel probability measures that have x as their barycentre. To prove the assertion we ought to show that the extreme points of $\mathcal{A}$ are of the form

$$\sum_{i=1}^{d+1} \lambda_i \delta_{x_i}$$

for some positive $\lambda_1, \dots, \lambda_{d+1}$ that sum up to one, $d \leq n$ and $x_1, \dots, x_{d+1}$ affinely independent such that

$$x = \sum_{i=1}^{d+1} \lambda_i x_i.$$

Let us first show that any extreme point $\gamma \in \mathcal{A}$ is supported on at most $n+1$ points. Suppose conversely, that there exist pairwise disjoint non-empty Borel sets $A_1, \dots, A_{n+2} \subset K$ such that

$$K = \bigcup_{i=1}^{n+2} A_i \text{ and } \gamma(A_i) > 0 \text{ for } i = 1, \dots, n+2.$$

Then there exist real numbers $t_1, \dots, t_{n+2}$, not all of them equal, such that

$$0 = \sum_{i=1}^{n+2} t_i \int_{A_i} (y - x) d\gamma(y).$$

We may assume that these numbers have absolute values all less than one and are such that

$$-\frac{1}{2} \leq \sum_{i=1}^{n+2} t_i \gamma(A_i) \leq \frac{1}{2}.$$

Set

$$\gamma_1 = \frac{\sum_{i=1}^{n+2}(1-t_i)\gamma|_{A_i}}{1-\sum_{i=1}^{n+2} t_i\gamma(A_i)} \text{ and } \gamma_2 = \frac{\sum_{i=1}^{n+2}(1+t_i)\gamma|_{A_i}}{1+\sum_{i=1}^{n+2} t_i\gamma(A_i)}.$$

Then γ_1, γ_2 belong to $\mathcal{A}$. Moreover

$$\gamma = \frac{1}{2}\Big(1 - \sum_{i=1}^{n+2} t_i\gamma(A_i)\Big)\gamma_1 + \frac{1}{2}\Big(1 + \sum_{i=1}^{n+2} t_i\gamma(A_i)\Big)\gamma_2.$$

Thus (δ_x, γ) is not an extreme point of $\mathcal{A}$. The contradiction yields that γ is supported on at most $n+1$ points.

Let $d+1 \leq n+1$ be the number of points in the support. Let us show that we must necessarily have

$$\gamma = \sum_{i=1}^{d+1} \lambda_i \delta_{x_i}$$

for some positive numbers $\lambda_1, \ldots, \lambda_{d+1}$ that sum up to one and $x_1, \ldots, x_{d+1}$ affinely independent. Suppose that this is not the case. Then there exist non-negative $\alpha_1, \ldots, \alpha_{d+1}$, not all of them equal to $\lambda_1, \ldots, \lambda_{d+1}$, such that

$$x = \sum_{i=1}^{d+1} \alpha_i x_i \text{ and } \sum_{i=1}^{d+1} \alpha_i = 1.$$

Set $\chi = \sum_{i=1}^{d+1} \alpha_i \delta_{x_i}$. Then $\chi \in \mathcal{A}$. Moreover, for any $\epsilon \in \Big(0, \min\Big\{\frac{\lambda_1}{\alpha_1}, \ldots, \frac{\lambda_{d+1}}{\alpha_{d+1}}, 1\Big\}\Big)$, we may write

$$\gamma = \frac{1}{2}(1-\epsilon)\frac{\gamma - \epsilon\chi}{1-\epsilon} + \frac{1}{2}(1+\epsilon)\frac{\gamma+\epsilon\chi}{1+\epsilon},$$

as a convex combination of two distinct measures in $\mathcal{A}$. This concludes the proof of the fact that any extreme point of $\mathcal{P}$ is of the form (5.6.1).

Let us now show that any pair μ, ν of that form is indeed an extreme point of $\mathcal{P}$. Observe that by Jensen's inequality any such pair belongs to $\mathcal{P}$. If we had

$$(\mu, \nu) = \lambda(\theta_1, \rho_1) + (1-\lambda)(\theta_2, \rho_2)$$

for some $(\theta_1, \rho_1), (\theta_2, \rho_2) \in \mathcal{P}$ and some $\lambda \in (0,1)$, then necessarily $\theta_1 = \theta_2 = \mu$, as μ is supported on a single point $x \in K$, and ρ_1, ρ_2 are supported on the support of ν. As the points in the support of ν are affinely independent and

$$x = \int_\Omega y d\rho_1(y) = \int_\Omega y d\rho_2(y),$$

we see that $\rho_1 = \rho_2 = \nu$. □

The reasoning presented in the previous sections of the current chapter may be also applied to the martingale optimal transport. In this problem one is given two Borel probability measures μ, ν on a convex body $K \subset \mathbb{R}^n$ which are in convex order. The task is to find a coupling π of μ and ν such that it is a distribution of a one-step martingale and that minimises the integral

$$\int_{K \times K} c d\pi$$

among all such couplings. Here $c\colon K \times K \to \mathbb{R}$ is a given Borel measurable funciton, callled a cost function.

In the theorem below we shall employ the above characterisation of extreme points to prove a duality result for the multi-dimensional martingale optimal transport, provided that the value of the dual problem is attained.

For other results related to duality in the martingale optimal transport problem see [20, 22].

Below we shall consider continuous functions $g \in \mathcal{C}(K \cup K)$ on the disjoint union of two copies of K. For such a function we shall denote by g_1 and g_2 the restrictions of g to the first and to the second copy of K respectively.

Theorem 5.6.2. *Let K be a convex body in $\mathbb{R}^n$ and let $c\colon K \times K \to \mathbb{R}$ be a Lipschitz function. Let μ, ν be two Borel probability measures on K in convex order. Let $\mathcal{K}$ denote the set of continuous functions g on the disjoint union of two copies of K such that for all non-negative $\lambda_1, \dots, \lambda_{n+1}$ that add up to one and all $x_1, \dots, x_{n+1} \in K$ there is*

$$g_1\Big(\sum_{i=1}^{n+1} \lambda_i x_i\Big) - \sum_{i=1}^{n+1} \lambda_i g_2(x_i) \leq \sum_{i=1}^{n+1} \lambda_i c\Big(\sum_{j=1}^{n+1} \lambda_j x_j, x_i\Big).$$

Let $\mathcal{P}$ denote the set of pairs of Borel probability measures on K that are in convex order. Suppose that the supremum of integrals

$$\int_K f_1 d\mu - \int_K f_2 d\nu$$

taken over the set $\mathcal{K}$ is attained. Then it is equal to the infimum of integrals

$$\int_{\mathcal{E}} \int_{K \times K} c d(\xi_1 \otimes \xi_2) d\pi(\xi) \tag{5.6.2}$$

over the set of all Borel probability measures π on the set $\mathcal{E}$ of extreme points of $\mathcal{P}$ such that

$$(\mu, \nu) = \int_{\mathcal{E}} \xi d\pi(\xi).$$

Moreover the infimum is attained. It is also equal to the infimum of integrals

$$\int_{K \times K} c d\pi \tag{5.6.3}$$

over all $\pi \in \Theta(\mu,\nu)$. Here $\Theta(\mu,\nu)$ stands for the set of all Borel probability measures on $K\times K$ such that its marginals are μ,ν and that are distributions of a one-step martingale.

Lemma 5.6.3. *Let K be a convex body in $\mathbb{R}^n$ and let $c\colon K\times K\to\mathbb{R}$ be a Lipschitz function. Let $f\in\mathcal{K}$. Then the set of extreme points of the set $\mathcal{R}$ of pairs of Borel probability measures $(\mu,\nu)\in\mathcal{P}(K)\times\mathcal{P}(K)$ that are in convex order and such that*

$$\int_K g_1 d\mu - \int_K g_2 d\nu \leq \int_K f_1 d\mu - \int_K f_2 d\nu$$

for all $g\in\mathcal{K}$ is equal to the set of pairs of the form $(\delta_x, \sum_{i=1}^{d+1}\lambda_i x_i)$ for some $d\leq n$, $\lambda_1,\dots,\lambda_{d+1}$ positive that sum up to one, $x_1,\dots,x_{d+1}\in K$ affinely independent, $x=\sum_{i=1}^{d+1}\lambda_i x_i$, such that

$$f_1\Big(\sum_{i=1}^{d+1}\lambda_i x_i\Big) - \sum_{i=1}^{d+1}\lambda_i f_2(x_i) = \sum_{i=1}^{d+1}\lambda_i c\Big(\sum_{j=1}^{d+1}\lambda_j x_j, x_i\Big).$$

Proof. The set $\mathcal{K}$ is convex, stable under maxima, contains constants and for any $t\in\mathbb{R}$ there is $\mathcal{K}+t\subset\mathcal{K}$. Thus, by Theorem 5.2.3, any extreme point of $\mathcal{R}$ is of the form (δ_{x_0},η) for some $x_0\in K$ and a Borel probability measure η on K. Let $(\delta_{x_0},\eta)\in\mathcal{R}$ be such an extreme point. Let $h_0\in\mathcal{C}(K)$ be a convex continuous function on K. Let h be a function equal to h_0 on the first copy of K and equal to the same function h_0 on the other copy of K. Then $f+h\in\mathcal{K}$. Thus η majorises δ_{x_0} in the convex order. Then we know that for any $g\in\mathcal{K}$ we have

$$\int_K \big(g_1(x_0)-g_2(y)\big)d\eta(y) \leq \int_K \big(f_1(x_0)-f_2(y))\big)d\eta(y). \tag{5.6.4}$$

As $f\in\mathcal{K}$, the right-hand side of the above inequality is bounded above by

$$\int_K c(x_0,y)d\eta(y). \tag{5.6.5}$$

Indeed, as η majorises δ_{x_0} in the convex order, by Theorem 5.6.1, there exists a Borel probability measure on the set of extreme points $\mathcal{E}$ of $\mathcal{P}$ such that

$$(\delta_{x,0}\eta) = \int_{\mathcal{E}} \xi d\pi(\xi). \tag{5.6.6}$$

The fact that $f\in\mathcal{K}$ may be rephrased by

$$\int_K f_1 d\xi_1 - \int_K f_2 d\xi_2 \leq \int_{K\times K} cd(\xi_1\otimes\xi_2)$$

for all $\xi\in\mathcal{E}$. The fact that (5.6.4) is bounded by (5.6.5) follows by the integration against π.

By the McShane extension formula (see [80]), we may assume that c is defined and Lipschitz on $\mathbb{R}^n \times \mathbb{R}^n$. Define g so that for $y \in K$ we have $g_2(y) = -c(x_0, y)$ and for $x \in K$ set

$$g_1(x) = \inf\Big\{ \sum_{i=1}^{n+1} \lambda_i\big(c(x, y_i) - c(x_0, y_i)\big) \mid \Big(\delta_x, \sum_{i=1}^{d+1} \lambda_i \delta_{y_i}\Big) \in \mathcal{E}\Big\}.$$

Here the infimum is over all pairs of measures in $\mathcal{E}$. Then

$$g_1(x) - \sum_{i=1}^{n+1} \lambda_i g_2(y_i) \le \sum_{i=1}^{n+1} \lambda_i c(x, y_i)$$

for all $y_1, \dots, y_{n+1} \in K$, all non-negative $\lambda_1, \dots, \lambda_{n+1} \ge 0$ summing up to one, with $x = \sum_{i=1}^{n+1} \lambda_i y_i$. Moreover, $g_1(x_0) = 0$. We claim that g is Lipschitz. Indeed, for any $x, y \in K$ and any $y_1, \dots, y_{n+1} \in \mathbb{R}^n$ and non-negative $\lambda_1, \dots, \lambda_{n+1}$ that sum up to one and such that $y = \sum_{i=1}^{n+1} \lambda_i y_i$ we have

$$\begin{aligned} &\sum_{i=1}^{n+1} \lambda_i\big(c(x, y_i + x - y)) - c(x_0, y_i + x - y)\big) \le \\ &\le \sum_{i=1}^{n+1} \lambda_i\big(c(y, y_i) - c(x_0, y_i)\big) + 3L\|x - y\|, \end{aligned}$$

where L is the Lipschitz constant of c. Thus

$$g_1(x) \le g_1(y) + 3L\|x - y\|.$$

This shows that g_1 is Lipschitz, hence continuous. In consequence, $g \in \mathcal{K}$. Observe that, for such g,

$$\int_K \big(g_1(x_0) - g_2(y)\big) d\eta(y) = \int_K c(x_0, y) d\eta(y).$$

Let π be as in (5.6.6). It follows by (5.6.4) that for π-almost every ξ we have

$$\int_K (f_1(x_0) - f_2(y)) d\xi_2(y) = \int_K c(x_0, y) d\xi_2(y),$$

where $\delta_{x_0} = \xi_1$. Hence π-almost every $\xi \in \mathcal{E}$ belongs to $\mathcal{R}$. Therefore any extreme point of $\mathcal{R}$ is necessary an extreme point of $\mathcal{P}$. The assertion follows readily. □

Proof of Theorem 5.6.2. First part of the theorem follows directly from Lemma 5.6.3 and Choquet's theorem. For a proof of the second part, take a Borel probability measure π_0 on $\mathcal{E}$ that attains the infimum (5.6.2). Set

$$\pi = \int_{\mathcal{E}} \xi_1 \otimes \xi_2 d\pi_0(\xi).$$

Then $\pi \in \Theta(\mu, \nu)$ is optimal for (5.6.3). □

5.7 Dual problem in martingale optimal transport

Let μ, ν be Borel probability measures on $\mathbb{R}^n$ with finite first moments that are in convex order. Martingale optimal transport problem between μ and ν admits a dual problem, which is to find the supremum of integrals

$$\int_{\mathbb{R}^n} f_1 d\mu - \int_{\mathbb{R}^n} f_2 d\nu$$

taken over the set of all continuous functions $f_1, f_2 \in \mathcal{C}(\mathbb{R}^n)$ such that

$$f_1(x) - f_2(y) \leq c(x,y) + \langle \gamma(x), y - x \rangle$$

for all $x, y \in \mathbb{R}^n$ and for some map $\gamma \colon \mathbb{R}^n \to \mathbb{R}^n$. In this section we investigate this class of functions. We prove that this class of functions is equal to the class considered in the previous section.

Definition 5.7.1. Let $K \subset \mathbb{R}^n$ be a convex set. Then $F \subset K$ is a *face* of K if for any $z \in F$ and any $t \in (0,1)$ such that $z = tx + (1-t)y$ for some $x, y \in K$ we have $x, y \in F$.

Let us observe that for any $x \in K$, there exists the minimal face of K containing x. This holds true by the fact that K itself is a face and that the interesection of a family of faces is a face. Uniqueness of minimal faces follows again by the property that the intersection of two faces is a face. Note that by the Hahn–Banach theorem it follows that x belongs to the relative interior of the minimal face that contains x.

Suppose that $K \subset \mathbb{R}^n$ is a convex body. Let $c \colon K \times K \to \mathbb{R}$ be non-negative. We shall denote by $\mathcal{L}$ the set of functions $f \in \mathcal{C}(K \cup K)$ such that there is a map $\gamma \colon K \to \mathbb{R}^n$ such that for all $x \in K$ and all $y \in K$ that belong to the minimal face of K that contains x there is

$$f_1(x) - f_2(y) \leq c(x,y) + \langle \gamma(x), y - x \rangle.$$

Here, as before, f_1 is the restriction of f to the first copy of K and f_2 is the restriction of f to the second copy of K.

Lemma 5.7.2. *The set $\mathcal{L}$ is convex, stable under maxima, contains constants and for any $t \in \mathbb{R}$ there is $\mathcal{L} + t \subset \mathcal{L}$.*

Proof. Observe that the only non-trivial assertion is that $\mathcal{L}$ is stable under maxima. For the proof of this fact let $f, g \in \mathcal{L}$ and let h denote $f \vee g$. Let $\gamma, \theta \colon K \to \mathbb{R}^n$ be such that

$$f_1(x) - f_2(y) \leq c(x,y) + \langle \gamma(x), y - x \rangle \text{ and } g_1(x) - g_2(y) \leq c(x,y) + \langle \theta(x), y - x \rangle$$

for all $x \in K$ and all y in the minimal face of K that contains x. Define $\rho\colon K \to \mathbb{R}^n$ by setting $\rho(x) = \gamma(x)$ if $f_1(x) \geq g_1(x)$ and $\rho(x) = \theta(x)$ if $f_1(x) < g_1(x)$. Suppose that $x \in K$ is such that $f_1(x) \geq g_1(x)$. Then for all y in the minimal face of K that contains x

$$h_1(x) = f_1(x) \leq c(x,y) + \langle \gamma(x), y - x\rangle + f_2(y) \leq c(x,y) + \langle \rho(x), y - x\rangle + h_2(y).$$

If $f_1(x) < g_1(x)$, then we prove analogus inequality in the same way. It follows that $h \in \mathcal{L}$. □

Theorem 5.7.3. *Let K be a convex body in $\mathbb{R}^n$. Let $c\colon K \times K \to \mathbb{R}$ be a Lipschitz function. Let $f \in \mathcal{C}(K \cup K)$ be a continuous function on a disjoint union of two copies of K. The following conditions are equivalent:*

i) for all $x_1, \dots, x_{n+1} \in K$ and all non-negative $\lambda_1, \dots, \lambda_{n+1}$ that sum up to one there is

$$f_1\Big(\sum_{i=1}^{n+1} \lambda_i x_i\Big) - \sum_{i=1}^{n+1} \lambda_i f_2(x_i) \leq \sum_{i=1}^{n+1} \lambda_i c\Big(\sum_{j=1}^{n+1} \lambda_j x_j, x_i\Big),$$

ii) there exists $\gamma\colon K \to \mathbb{R}^n$ such that for any $x \in K$ there and for all $y \in K$ in the minimal face of K that contains x we have

$$f_1(x) - f_2(y) \leq c(x,y) + \langle \gamma(x), y - x\rangle.$$

iii) for any one-step martingale (X_0, X_1) with values in K there is

$$\mathbb{E}\big(f_1(X_0) - f_2(X_1)\big) \leq \mathbb{E}c(X_0, X_1).$$

Proof. Let us prove that the conditions i) and iii) are equivalent. Take a one-step martingale (X_0, X_1) with values in K. Let π denote its distribution. Then there exists π_0 – a Borel probability measure on the set $\mathcal{E}$ of extreme points of the set of pairs of Borel probability measures that are in convex order such that

$$\pi = \int_{\mathcal{E}} \xi_1 \otimes \xi_2 d\pi_0(\xi).$$

Note that, by Theorem 5.6.1, i) tells us that for any $\xi \in \mathcal{E}$ there is

$$\int_K f_1 d\xi_1 - \int_K f_2 d\xi_2 \leq \int_{K\times K} c d(\xi_1 \otimes \xi_2).$$

Therefore

$$\mathbb{E}\big(f_1(X_0) - f_2(X_1)\big) = \int_{\mathcal{E}} \Big(\int_K f_1 d\xi_1 - \int_K f_2 d\xi_2\Big) d\pi_0(\xi) \leq$$
$$\leq \int_{\mathcal{E}} \int_{K\times K} c d(\xi_1 \otimes \xi_2) d\pi_0(\xi) = \mathbb{E}c(X_0, X_1).$$

This is to say, i) implies iii). For the converse implication, take any $x_1, \dots, x_{n+1} \in K$ and any non-negative $\lambda_1, \dots, \lambda_{n+1}$ that sum up to one. Set $X_0 = \sum_{i=1}^{n+1} \lambda_i x_i$ and $X_1 = x_i$ with probability λ_i for $i = 1, \dots, n+1$. Then (X_0, X_1) is a one-step martingale, for which

$$\mathbb{E}\big(f_1(X_0) - f_2(X_1)\big) = f_1\Big(\sum_{i=1}^{n+1} \lambda_i x_i\Big) - \sum_{i=1}^{n+1} \lambda_i f_2(x_i)$$

and

$$\mathbb{E}c(X_0, X_1) = \sum_{i=1}^{n+1} \lambda_i c\Big(\sum_{j=1}^{n+1} \lambda_j x_j, x_i\Big).$$

For the proof of the other equivalences, without loss of generality, we may assume that c is non-negative. Let us denote the set of continuous functions on K that satisfy condition i) by $\mathcal{K}_1$ and the set of continuous functions on K that satisfy condition ii) by $\mathcal{K}_2$. Observe that trivially $\mathcal{K}_2 \subset \mathcal{K}_1$. Moreover both $\mathcal{K}_1$ and $\mathcal{K}_2$ are convex sets that are stable under maxima, contain constants and are such that for all $t \in \mathbb{R}$ there is $\mathcal{K}_i + t \subset \mathcal{K}_i$ for $i = 1, 2$; see Lemma 5.7.2. Suppose that there exists $f \in \mathcal{K}_1 \setminus \mathcal{K}_2$. Then by the Hahn–Banach theorem there exists a Borel measure $\eta \in \mathcal{M}(K \cup K)$ such that for all $g \in \mathcal{K}_2$

$$\int_{K \cup K} f d\eta > \int_{K \cup K} g d\eta. \tag{5.7.1}$$

Since constant functions belong to $\mathcal{K}_2$, measure η may be written as a difference of two Borel measures of equal masses. Since any continuous function that is negative on the first and positive on the second copy of K belongs to $\mathcal{K}_2$, we see that $\eta = \mu - \nu$, for some non-negative μ, ν such that μ is supported on the first copy of K and ν is supported on the second copy of K. Without loss of generality we may assume that these measures are probabilities. Observe that if $h_0 \in \mathcal{C}(K)$ is any convex function then $h \in \mathcal{C}(K \cup K)$ such that $h_1 = h_2 = h_0$ belongs to $\mathcal{K}_2$. Thus μ and ν are in convex order. By Theorem 5.2.3 any extreme point of the set of such pairs of measures has the form (δ_{x_0}, σ) for some $x_0 \in K$ and some Borel probability measure σ with barycentre x_0. We may therefore assume that $(\mu, \nu) = (\delta_{x_0}, \sigma)$ for some $x_0 \in K$ and some σ with barycentre x_0. Define $k \colon K \cup K \to \mathbb{R}$ by $k_2(y) = -c(x_0, y)$ for $y \in K$ and for $x \in K$ set

$$k_1(x) = \inf\{c(x, y) - c(x_0, y) \mid y \in K\}.$$

Then $k_1(x_0) = 0$, k is continuous by Lipschitzness of c and thus $h \in \mathcal{K}_2$, with γ equal to zero. It follows that there exists σ such that

$$\int_K c(x_0, y) d\sigma(y) = \int_K \big(k_1(x_0) - k_2\big) d\sigma < \int_K \big(f_1(x_0) - f_2\big) d\sigma. \tag{5.7.2}$$

Recall that (δ_{x_0}, σ) is ordered in convex order. Thus, there exists a probability measure π on the set $\mathcal{E}$ of extreme points of measures in convex order such that

$$(\delta_{x_0}, \sigma) = \int_{\mathcal{E}} \xi d\pi(\xi).$$

It follows, by Theorem 5.6.1, and the definition of $\mathcal{K}_1$, that

$$\int_K \big(f_1(x_0) - f_2\big) d\sigma \leq \int_K c(x_0, y) d\sigma(y).$$

This stands in contradiction to (5.7.2) and proves that $f \in \mathcal{K}_2$. □

Below we generalise the above theorem to the case of any convex set $K \subset \mathbb{R}^n$, not necessarily a compact one.

Corollary 5.7.4. *Let K be a convex set in $\mathbb{R}^n$. Suppose that $c\colon K \times K \to \mathbb{R}$ is a locally Lipschitz function. Let $f \in \mathcal{C}(K \cup K)$. The following conditions are equivalent:*

i) for all $x_1, \ldots, x_{n+1} \in K$ and all non-negative $\lambda_1, \ldots, \lambda_{n+1}$ that sum up to one there is

$$f_1\Big(\sum_{i=1}^{n+1} \lambda_i x_i\Big) - \sum_{i=1}^{n+1} \lambda_i f_2(x_i) \leq \sum_{i=1}^{n+1} \lambda_i c\Big(\sum_{j=1}^{n+1} \lambda_j x_j, x_i\Big),$$

ii) there exists $\gamma\colon K \to \mathbb{R}^n$ such that for any $x \in K$ there and for all $y \in K$ in the minimal face of K that contains x we have

$$f_1(x) - f_2(y) \leq c(x, y) + \langle \gamma(x), y - x \rangle.$$

iii) for any bounded one-step martingale (X_0, X_1) with values in K there is

$$\mathbb{E}\big(f_1(X_0) - f_2(X_1)\big) \leq \mathbb{E}c(X_0, X_1).$$

Proof. Without loss of generality we may assume that $\mathrm{int}K$ is non-empty. Choose a increasing sequence $(K_n)_{n=1}^{\infty}$ of compact convex subsets of K such that its union is $\mathrm{int}K$. Suppose that $f \in \mathcal{C}(K \cup K)$ satisfies i). Pick $x \in \mathrm{int}K$ and let $\epsilon > 0$ be such that $B(x, \epsilon) \subset \mathrm{int}K$. Here $B(x, \epsilon)$ denotes the closed ball of radius ϵ centred at x. Then by Theorem 5.7.3 for any $n \in \mathbb{N}$ sufficiently large so that $x \in \mathrm{int}K_n$ there exists γ_n such that for all $y \in K_n$ there is

$$f_1(x) - f_2(y) \leq c(x, y) + \langle \gamma_n, y - x \rangle. \tag{5.7.3}$$

Let n_0 be such that $B(x, \epsilon) \subset K_{n_0}$. Take $n > n_0$. Suppose that $\gamma_n \neq 0$ and set $y_n = x - \epsilon \frac{\gamma_n}{\|\gamma_n\|}$. Then $y_n \in K_{n_0} \subset K_n$ and therefore, by (5.7.3),

$$\|\gamma_n\| \leq \frac{1}{\epsilon}\big(c(x, y_n) - f_1(x) + f_2(y_n)\big).$$

As c is bounded on $\{x\} \times K_{n_0}$ and f is bounded on K_{n_0}, the right-hand side of the above inequality is bounded. Hence, so is the left-hand side. We may therefore pick γ that is an accumulation point of the sequence $(\gamma_n)_{n=1}^{\infty}$. From (5.7.3) and from continuity of f it follows now that for all $y \in K$

$$f_1(x) - f_2(y) \leq c(x, y) + \langle \gamma, y - x \rangle.$$

This is to say, f satisfies also ii) if $x \in \mathrm{int} K$. If $\gamma_n = 0$ for infinitely many n, then the above inequality holds true with $\gamma = 0$.

Pick now $x \in K$ and let L denote the minimal face of K that contains x. Then x belongs to the relative interior of L. We repeat the above argument with K replaced by L considered as a convex subset of its affine hull.

That ii) implies i) is straightforward.

The equivalence of i) and iii) follows readily from Theorem 5.7.3. □

We shall now investigate continuity properties of functions that satisfy conditions of Theorem 5.7.3 or of Corollary 5.7.4. We shall need the following lemma.

Lemma 5.7.5. *Suppose that* $f\colon [a, b] \to \mathbb{R}$ *is such that for all* $\lambda \in [0, 1]$ *and all* $x, y \in [a, b]$ *there is*

$$\lambda f(x) + (1-\lambda) f(y) - f(\lambda x + (1-\lambda) y) \leq \lambda c(\lambda x + (1-\lambda) y, x) + (1-\lambda) c(\lambda x + (1-\lambda) y, y).$$

Then for all $a \leq x_1 < x_2 < x_3 \leq b$ *the quotient* $\frac{f(x_3) - f(x_1)}{x_3 - x_1}$ *is bounded below by*

$$\frac{f(x_3) - f(x_2)}{x_3 - x_2} + \frac{c(x_2, x_3) - c(x_2, x_1)}{x_3 - x_1} - \frac{c(x_2, x_3)}{x_3 - x_2}$$

and above by

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} + \frac{c(x_2, x_3) - c(x_2, x_1)}{x_3 - x_1} + \frac{c(x_2, x_1)}{x_2 - x_1}.$$

Proof. Let $\lambda \in (0, 1)$ be such that $x_2 = \lambda x_1 + (1-\lambda) x_3$, that is

$$\lambda = \frac{x_3 - x_2}{x_3 - x_1}.$$

Then we know that

$$\lambda f(x_1) + (1-\lambda) f(x_3) - f(x_2) \leq \lambda c(x_2, x_1) + (1-\lambda) c(x_2, x_3).$$

Hence putting formula for λ we obtain that

$$\frac{f(x_3) - f(x_2)}{x_3 - x_2} + \frac{c(x_2, x_3) - c(x_2, x_1)}{x_3 - x_1} - \frac{c(x_2, x_3)}{x_3 - x_2} \leq \frac{f(x_3) - f(x_1)}{x_3 - x_1}$$

and

$$\frac{f(x_3) - f(x_1)}{x_3 - x_1} \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1} + \frac{c(x_2, x_3) - c(x_2, x_1)}{x_3 - x_1} + \frac{c(x_2, x_1)}{x_2 - x_1}.$$

□

Lemma 5.7.6. *Let K be a convex, open set in $\mathbb{R}^n$. Suppose that $f\colon K \to \mathbb{R}$ is such that for all $x, y \in K$ and all $\lambda \in [0,1]$ there is*

$$\lambda f(x) + (1-\lambda)f(y) - f(\lambda x + (1-\lambda)y) \leq \lambda c(\lambda x + (1-\lambda)y, x) + (1-\lambda)c(\lambda x + (1-\lambda)y, y).$$

Suppose that c is L-Lipschitz in the second variable and is such that for all $x, y \in K$ there is $|c(x,y)| \leq \Lambda\|x-y\|$ for some constant Λ. Then f is locally Lipschitz in K.

Proof. Suppose that $n = 1$. Then, without loss of generality, $K = [a, d]$ for some $a < d$. Choose numbers b, c so that $a < b < c < d$. Then applying Lemma 5.7.5 four times yields that for any x, y such that $b < x < y < c$ we have

$$\frac{f(y)-f(x)}{y-x} \leq \frac{f(b)-f(a)}{b-a} + \frac{c(x,y)}{y-x} + \frac{c(b,a)}{b-a} + \frac{c(b,y)-c(b,a)}{y-a} - \frac{c(x,y)-c(x,a)}{y-a}$$

and

$$\frac{f(d)-f(c)}{d-c} - \frac{c(c,x)}{c-x} + \frac{c(c,d)-c(c,x)}{d-x} - \frac{c(y,x)}{y-x} - \frac{c(y,d)-c(y,x)}{d-x} \leq \frac{f(y)-f(x)}{y-x}.$$

In particular on $[b, c]$ function f has Lipschitz constant at most

$$\max\Big\{\Big|\frac{f(b)-f(a)}{b-a} + 2L + 2\Lambda\Big|, \Big|\frac{f(d)-f(c)}{d-c} - 2L - 2\Lambda\Big|\Big\}.$$

Suppose now that $n > 1$ and that, by induction, the lemma holds true for all dimensions at most $n-1$. Choose any simplices X, Y and any ball B in K and such that $B \subset X \subset Y \subset K$ and such that B and the boundaries of X and Y are pairwise disjoint. Then, by the inductive assumption, f is continuous on the boundaries of X and Y, and therefore the function

$$\partial X \times \partial Y \ni (x,y) \mapsto \frac{|f(x)-f(y)|}{\|x-y\|} \in \mathbb{R}$$

is bounded by a constant M. Choose any points $x, y \in B$. Choose a unique line passing through x and y. Then there exist unique points $x_1, x_2 \in X$ and $y_1, y_2 \in Y$ such that the line intersects ∂X in x_1, x_2 and ∂Y in y_1, y_2 where, without loss of generality,

$$y_1 < x_1 < x < y < x_2 < y_2$$

on the line. By Lemma 5.7.5 we see that

$$\frac{|f(y)-f(x)|}{\|x-y\|} \leq \max\Big\{\Big|\frac{f(y_2)-f(x_2)}{\|y_2-x_2\|} - 2L - 2\Lambda\Big|, \Big|\frac{f(x_1)-f(y_1)}{\|y_2-x_2\|} + 2L + 2\Lambda\Big|\Big\}$$

Therefore f has Lipschitz constant at most $M + 2L + 2\Lambda$ on B. ☐

Corollary 5.7.7. *Let K be a convex set in $\mathbb{R}^n$. Suppose that $c\colon K \times K \to \mathbb{R}$ is locally Lipschitz. Let $f \in \mathcal{C}(K)$. The following conditions are equivalent:*

i) for all $x_1, \dots, x_{n+1} \in K$ and all non-negative $\lambda_1, \dots, \lambda_{n+1}$ that sum up to one there is

$$f\Big(\sum_{i=1}^{n+1} \lambda_i x_i\Big) - \sum_{i=1}^{n+1} \lambda_i f(x_i) \leq \sum_{i=1}^{n+1} \lambda_i c\Big(\sum_{j=1}^{n+1} \lambda_j x_j, x_i\Big),$$

ii) there exists $\gamma\colon K \to \mathbb{R}^n$ such that for any $x \in K$ there and for all $y \in K$ in the minimal face of K that contains x we have

$$f(x) - f(y) \leq c(x, y) + \langle \gamma(x), y - x \rangle.$$

iii) for any bounded one-step martingale (X_0, X_1) with values in K there is

$$\mathbb{E}\big(f(X_0) - f(X_1)\big) \leq \mathbb{E}c(X_0, X_1).$$

Moreover, if K is open and additionally for any $x_0 \in K$ there exist open, convex set $K' \subset K$ such that $x_0 \in K'$ and $|c(x, y)| \leq \Lambda \|x - y\|$ for all $x, y \in K'$ and some constant Λ, then we may drop the assumption on the continuity of f. In such case, any function that satisfies the above conditions is locally Lipschitz.

Proof. The equivalence of the conditions i) ii) and iii) follows from Corollary 5.7.4. The second part of the corollary follows from Lemma 5.7.6, as, in such case, any function f that satisfies i), ii) or iii) is continuous in K. □

5.8 Uniform convexity and uniform smoothness

In this section we employ the results of the previous section to provide a characterisation of uniformly smooth and uniformly convex functions on $\mathbb{R}^n$, or, more generally, on an open, convex set $K \subset \mathbb{R}^n$. We refer the reader to [10] and to [108] and references therein for previous studies of the topic. Let us recall the definitions.

Definition 5.8.1. Let $\sigma\colon \mathbb{R} \to \mathbb{R}$. A function $f\colon K \to \mathbb{R}$ is called σ-convex provided that

$$f(\lambda x + (1 - \lambda)y) + \lambda(1 - \lambda)\sigma(\|x - y\|) \leq \lambda f(x) + (1 - \lambda)f(y)$$

for all $\lambda \in [0, 1]$ and all $x, y \in K$. A function $g\colon K \to \mathbb{R}$ is called σ-smooth provided that

$$g(\lambda x + (1 - \lambda)y) + \lambda(1 - \lambda)\sigma(\|x - y\|) \geq \lambda g(x) + (1 - \lambda)g(y)$$

for all $\lambda \in [0, 1]$ and all $x, y \in K$.

Another notion of convexity and smoothness is as follows (see [10]).

Definition 5.8.2. Let $\gamma \in \mathbb{R}^n$ and let $x \in K$. We say that $f\colon K \to \mathbb{R}$ is σ-uniformly convex at x with respect to γ if for all $y \in K$ there is

$$f(x) + \sigma(\|y - x\|) + \langle \gamma, y - x \rangle \leq f(y).$$

Likewise, $g\colon K \to \mathbb{R}$ is called σ-uniformly smooth at x with respect to γ if for all $y \in K$ there is

$$g(x) + \sigma(\|y - x\|) + \langle \gamma, y - x \rangle \geq g(y).$$

Note that the condition that $f\colon K \to \mathbb{R}$ is σ-uniformly convex at any $x \in K$ is equivalent to condition ii) of Corollary 5.7.7 for the function

$$c(x, y) = -\sigma(\|y - x\|),\ x, y \in K.$$

Similarly, σ-uniform smoothness at any $x \in K$ of a function $g\colon K \to \mathbb{R}$ is equivalent to condition ii) of Corollary 5.7.7 for $-g$ and the function

$$c(x, y) = \sigma(\|y - x\|),\ x, y \in K.$$

Now, Corollary 5.7.7 implies the following theorem, which complements the results of [10].

Theorem 5.8.3. *Let $K \subset \mathbb{R}^n$ be an open, convex set. Suppose that $\sigma\colon \mathbb{R} \to \mathbb{R}$ is locally Lipschitz function such that $\sigma(0) = 0$. Let $f\colon K \to \mathbb{R}$. The following conditions are equivalent:*

i) there exists $\gamma\colon K \to \mathbb{R}^n$ such that for any $x \in K$ the function f is σ-uniformly convex at x with respect to $\gamma(x) \in \mathbb{R}^n$,

ii) for any $x_1, \ldots, x_{n+1} \in K$ and any non-negative $\lambda_1, \ldots, \lambda_{n+1}$ that sum up to one there is

$$f\Big(\sum_{i=1}^{n+1} \lambda_i x_i\Big) - \sum_{i=1}^{n+1} \lambda_i f(x_i) \leq -\sum_{i=1}^{n+1} \lambda_i \sigma\Big(\Big\|\sum_{j=1}^{n+1} \lambda_j x_j - x_i\Big\|\Big),$$

iii) for any bounded one-step martingale (X_0, X_1) with values in K there is

$$\mathbb{E}\big(f(X_0) - f(X_1)\big) \leq -\mathbb{E}\sigma(\|X_0 - X_1\|).$$

Also, the following conditions are equivalent:

i) there exists $\gamma\colon K \to \mathbb{R}^n$ such that for any $x \in K$ the function f is σ-uniformly smooth at x with respect to $\gamma(x) \in \mathbb{R}^n$, *ii) for any $x_1, \ldots, x_{n+1} \in K$ and any non-negative $\lambda_1, \ldots, \lambda_{n+1}$ that sum up to one there is*

$$f\Big(\sum_{i=1}^{n+1} \lambda_i x_i\Big) - \sum_{i=1}^{n+1} \lambda_i f(x_i) \geq -\sum_{i=1}^{n+1} \lambda_i \sigma\Big(\Big\|\sum_{j=1}^{n+1} \lambda_j x_j - x_i\Big\|\Big),$$

iii) for any bounded one-step martingale (X_0, X_1) with values in K there is

$$\mathbb{E}\big(f(X_0) - f(X_1)\big) \geq -\mathbb{E}\sigma(\|X_0 - X_1\|).$$

Moreover, any function f that satisfies one of the above conditions is locally Lipschitz in K.

Proof. The assumptions on σ imply that the functions

$$(x, y) \mapsto -\sigma(\|y - x\|) \text{ and } (x, y) \mapsto \sigma(\|y - x\|)$$

are locally Lipschitz in $K \times K$ and moreover for any $x_0 \in K$ there exists an open, convex set $K' \subset K$ such that $x_0 \in K'$ and for all $x, y \in K'$ there is

$$|\sigma(\|y - x\|)| \leq \Lambda \|y - x\|,$$

where Λ depends on σ. This is to say, the assumptions of Corollary 5.7.7 are satsfied. Thus, the assertion of the theorem follows from the conclusion of Corollary 5.7.7. □

5.9 Martingale triangle inequality

In this section we introduce the notion of *martingale triangle inequality* for cost functions $c \colon K \times K \to \mathbb{R}$, where $K \subset \mathbb{R}^n$ is a convex set. We shall show that if it is satisfied by a cost function c, which vanishes on the diagonal, then one may take $f_1 = f_2$ in the dual problem to the martingale optimal transport.

Definition 5.9.1. Let $K \subset \mathbb{R}^n$ be a convex set. Let $c \colon K \times K \to \mathbb{R}$. We say that c satisfies *martingale triangle inequality* provided that for all $x, x_1, \ldots, x_{n+1} \in K$ and all non-negative $\lambda_1, \ldots, \lambda_{n+1}$ that sum up to one there is

$$\sum_{i=1}^{n+1} \lambda_i c(x, x_i) - c\Big(x, \sum_{i=1}^{n+1} \lambda_i x_i\Big) \leq \sum_{i=1}^{n+1} \lambda_i c\Big(\sum_{j=1}^{n+1} \lambda_j x_j, x_i\Big). \tag{5.9.1}$$

In other words, for any $x \in K$ function $-c(x, \cdot)$ satisfies condition i) or Corollary 5.7.4, with cost function c.

Remark 5.9.2. Condition (5.9.1) is satisfied if c is a metric on K and also it is satisfied if c is concave in the second variable and non-negative. Also, this condition defines a closed convex cone of functions. Note also that for a function given by $c(x,y) = \|x-y\|^2$ for $x, y \in K$, where $\|\cdot\|$ deotes Euclidean norm on K, we have equality in (5.9.1).

The following theorem is a martingale optimal transport analogue of the Kantorovich–Rubinstein duality for the classical optimal transport problem with a metric cost function; see Section 5.4.

Theorem 5.9.3. *Let K be a convex body in $\mathbb{R}^n$. Let $c\colon K \times K \to \mathbb{R}$ be a continuous function satisfying martingale triangle inequality and vanishing on the diagonal.*

Let $\mathcal{B}_1$ denote the set of functions $f \in \mathcal{C}(K)$ such that for all $x_1, \dots, x_{n+1} \in K$ and all non-negative $\lambda_1, \dots, \lambda_{n+1}$ that sum up to one there is

$$f\Big(\sum_{i=1}^{n+1} \lambda_i x_i\Big) - \sum_{i=1}^{n+1} \lambda_i f(x_i) \leq \sum_{i=1}^{n+1} \lambda_i c\Big(\sum_{j=1}^{n+1} \lambda_j x_j, x_i\Big).$$

Let $\mathcal{B}_2$ denote the set of functions $g \in \mathcal{C}(K \cup K)$ on the disjoint union of two copies of K such that for all $x_1, \dots, x_{n+1} \in K$ and all non-negative $\lambda_1, \dots, \lambda_{n+1}$ that sum up to one there is

$$g_1\Big(\sum_{i=1}^{n+1} \lambda_i x_i\Big) - \sum_{i=1}^{n+1} \lambda_i g_2(x_i) \leq \sum_{i=1}^{n+1} \lambda_i c\Big(\sum_{j=1}^{n+1} \lambda_j x_j, x_i\Big).$$

Then, for any Borel probability measures μ, ν on K in convex order there is

$$\sup\Big\{ \int_K f d(\mu - \nu) \mid f \in \mathcal{B}_1 \Big\} = \sup\Big\{ \int_K g_1 d\mu - \int_K g_2 d\nu \mid g \in \mathcal{B}_2 \Big\}. \tag{5.9.2}$$

Proof. Clearly, the supremum on the right-hand side of (5.9.2) is at least the supremum on the left-hand side of (5.9.2), as if $f \in \mathcal{B}_1$, then $g\colon K \cup K \to \mathbb{R}$ defined by $g_1 = f$ and $g_2 = f$ belongs to $\mathcal{B}_2$.

Suppose that there exists $\epsilon > 0$ such that for any function $f \in \mathcal{B}_1$

$$\int_K f d(\mu - \nu) + 2\epsilon \leq \sup\Big\{ \int_K g_1 d\mu - \int_K g_2 d\nu \mid g \in \mathcal{B}_2 \Big\}.$$

It follows that there exists $g \in \mathcal{B}_2$ such that for all $f \in \mathcal{B}_1$ we have

$$\int_K f d(\mu - \nu) + \epsilon \leq \int_K g_1 d\mu - \int_K g_2 d\nu. \tag{5.9.3}$$

By Corollary 5.2.5, any extreme point of pairs of probability measures $\mathcal{P}$ that satisfy

$$\int_K f d(\mu - \nu) \leq \int_K g_1 d\mu - \int_K g_2 d\nu.$$

for all $f \in \mathcal{B}_1$ is of the form (δ_{x_0}, η) for some $x_0 \in K$ and a Borel probability measure η. Hence, by (5.9.3), for some $x_0 \in K$ and some $\eta \in \mathcal{P}(K)$ and all $f \in \mathcal{B}_1$

$$\int_K (f(x_0) - f) d\eta + \epsilon \leq \int_K \big(g_1(x_0) - g_2\big) d\eta. \tag{5.9.4}$$

Note that any such pair (δ_{x_0}, η) is in convex order. Thus there exists a Borel probability measure π_0 on the set $\mathcal{E}$ of extreme points of pairs of measures in convex order such that

$$(\delta_{x_0}, \eta) = \int_{\mathcal{E}} \xi d\pi_0(\xi). \tag{5.9.5}$$

But, as $g \in \mathcal{B}_2$, for any $\xi \in \mathcal{E}$ we have

$$\int_{K \times K} (g_1(x) - g_2(y)) d(\xi_1 \otimes \xi_2)(x, y) \leq \int_{K \times K} c d(\xi_1 \otimes \xi_2).$$

Take $f = -c(x_0, \cdot)$. Then, by the martingale triangle inequality, $f \in \mathcal{B}_1$. This, together wih (5.9.5) and (5.9.4), yields a contradiction. $\square$

Chapter 6

Matrix Hölder's inequality

6.1 Introduction

In this chapter we consider matrix Hölder's inequality, which constitutes an extension of the classical Hölder's inequality to the matrix setting. The result of the chapter is the characterisation of the equality cases. Theorem 6.2.1 and Theorem 6.2.4 and their proofs are based on [14]. We refer the reader also to [23]. We include the proofs for completeness and to provide an analysis of equality cases.

Section 6.2 is devoted to a proof of the inequality and Section 6.3 characterises equality cases in matrix Hölder's inequality.

6.2 Inequality

For a matrix A of size $m \times n$ we shall denote by $|A|$ its absolute value, that is

$$|A| = (A^*A)^{\frac{1}{2}}.$$

Theorem 6.2.1. *Let A, B be two $m \times n$ matrices with real entries. Then*

$$|\mathrm{tr}(A^*B)| \leq (\mathrm{tr}|A||B|)^{\frac{1}{2}} \big(\mathrm{tr}|A^*||B^*|\big)^{\frac{1}{2}}. \tag{6.2.1}$$

Lemma 6.2.2. *Let A be an $m \times n$ matrix with real entries. Then there exist orthonormal basis $(e_i)_{i=1}^n$ and $(f_j)_{j=1}^m$ and non-negative numbers $(a_i)_{i=1}^{n\wedge m}$ such that*

$$Ae_i = a_i f_i \text{ and } A^* f_j = a_j e_j \text{ for all } i, j = 1, \ldots, n \wedge m$$
$$\text{and } Ae_i = 0, A^* f_j = 0 \text{ for } i, j > n \wedge m.$$

Proof. For the proof, let $(e_i)_{i=1}^n$ be an orthonormal basis of eigenvectors of A^*A with non-negative eigenvalues $(a_i^2)_{i=1}^n$, $a_i \geq 0$. We may assume that for $i \geq n \wedge m$ there is $A^*Ae_i = 0$, as the rank of A^*A is at most $n \wedge m$.

For indices such that $a_i = 0$, we have $Ae_i = 0$, as $\|Ae_i\|^2 = \langle A^*Ae_i, e_i\rangle = 0$. Set

$$f_j = \frac{1}{a_j} Ae_j \text{ for } j \text{ such that } a_j \neq 0. \tag{6.2.2}$$

These are orthonormal, eigenvectors of AA^* such that $A^* f_j = a_j e_j$.

Note that any eigenvector of AA^* that is orthogonal to all f_j has eigenvalue equal to zero. Indeed, if f is such an eigenvector, then for all j

$$0 = \langle f, Ae_j\rangle = \langle A^* f, e_j\rangle,$$

that is $A^* f = 0$ and thus $AA^* f = 0$.

We may thus complement the eigenvectors (6.2.2) to a full orthnormal basis $(f_j)_{j=1}^m$ by introducing eigenvectors of AA^* with zero eigenvalue. For such vectors we have $A^* f_j = 0$.

This completes the proof of the lemma. □

Proof of Theorem 6.2.1. Take orthonormal basis $(e_i)_{i=1}^n$, $(f_j)_{j=1}^m$ and non-negative numbers $(a_i)_{i=1}^{n\wedge m}$ for A and $(g_i)_{i=1}^n$, $(h_j)_{j=1}^m$, $(b_i)_{i=1}^{n\wedge m}$ for B, as in Lemma 6.2.2. Then

$$|\mathrm{tr}(A^*B)| = \Big| \sum_{i,j=1}^{n\wedge m} a_i b_j \langle e_i, g_j\rangle \langle f_i, h_j\rangle \Big|.$$

The Cauchy–Schwarz inequality yields

$$|\mathrm{tr}(A^*B)| \leq \Big(\sum_{i,j=1}^{n\wedge m} a_i b_j \langle e_i, g_j\rangle^2 \Big)^{\frac{1}{2}} \Big(\sum_{i,j=1}^{n\wedge m} a_i b_j \langle f_i, h_j\rangle^2 \Big)^{\frac{1}{2}}.$$

Note now that for $i, j \leq n \wedge m$ we have

$$|A|e_i = a_i e_i, |A^*|f_i = a_i f_i \text{ and } |B|g_j = b_j g_j, |B^*|h_j = b_j h_j,$$

and for $i, j > n \wedge m$

$$|A|e_i = 0, |A^*|f_i = 0 \text{ and } |B|g_j = 0, |B^*|h_j = 0.$$

Therefore

$$|\mathrm{tr}(A^*B)| \leq (\mathrm{tr}(|A||B|)^{\frac{1}{2}} (\mathrm{tr}(|A^*||B^*|)^{\frac{1}{2}}.$$

□

Remark 6.2.3. The equality holds in the above inequality if and only if

$$(\langle e_i, g_j\rangle - \alpha \langle f_i, h_j\rangle) a_i b_j = 0$$

for some constant α and all indices i, j.

Below $\|\cdot\|$ denotes the operator norm of a matrix, regarded as a linear operator between Euclidean spaces.

Theorem 6.2.4. *Let A, B be two $m \times n$ matrices with real entries. Then*

$$|\mathrm{tr}(A^*B)| \leq (\mathrm{tr}|A|^p)^{\frac{1}{p}}(\mathrm{tr}(|B|^q)^{\frac{1}{q}} \tag{6.2.3}$$

for all $p, q \in (1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Moreover

$$|\mathrm{tr}(A^*B)| \leq \mathrm{tr}|A|\|B\|. \tag{6.2.4}$$

Proof. Let $p, q \in (1, \infty)$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Using the notation from the above theorem, let us note that

$$\mathrm{tr}|A|^p = \mathrm{tr}|A^*|^p = \sum_{i=1}^{n\wedge m} a_i^p.$$

Hence, for the proof it is enough to show that every factor on the right-hand side of the inequality (6.2.1) is bounded above by $(\mathrm{tr}|A|^p)^{\frac{1}{p}}(\mathrm{tr}(|B|^q)^{\frac{1}{q}}$. For this, observe that, thanks to orthonormality of the basis,

$$\sum_{i=1}^{n} \langle e_i, g_j\rangle^2 = 1,$$

for all $j = 1, \ldots, m$ and that

$$\sum_{j=1}^{m} \langle e_i, g_j\rangle^2 = 1,$$

for all $i = 1, \ldots, n$. Therefore, using Hölder's inequality, we get

$$\begin{aligned}\mathrm{tr}(|A||B|) &= \sum_{i,j=1}^{n\wedge m} a_i b_j \langle e_i, g_j\rangle^2 \leq \Big(\sum_{i,j=1}^{n\wedge m} a_i^p \langle e_i, g_j\rangle^2\Big)^{\frac{1}{p}} \Big(\sum_{i,j=1}^{n\wedge m} b_j^q \langle e_i, g_j\rangle^2\Big)^{\frac{1}{q}} = \\ &= \Big(\sum_{i,j=1}^{n\wedge m} a_i^p\Big)^{\frac{1}{p}} \Big(\sum_{i,j=1}^{n\wedge m} b_j^q\Big)^{\frac{1}{q}} = (\mathrm{tr}|A|^p)^{\frac{1}{p}}(\mathrm{tr}|B|^q)^{\frac{1}{q}}\end{aligned}$$

Proceeding analogously for $\mathrm{tr}(|A^*||B^*|)$ we get the desired inequality.

For the second part of the theorem observe that

$$\|B\| = \|B^*\| = \max\{b_i \mid i = 1, \ldots, n \wedge m\}$$

and that

$$\mathrm{tr}|A| = \mathrm{tr}|A^*| = \sum_{i,j=1}^{n\wedge m} a_i.$$

Therefore

$$\mathrm{tr}(|A||B|) = \sum_{i,j=1}^{n\wedge m} a_i b_j \langle e_i, g_j\rangle^2 \leq \|B\|\mathrm{tr}|A|.$$

Proceeding analogously for $\mathrm{tr}(|A^*||B^*|)$ we get the desired inequality. □

Remark 6.2.5. If $p, q \in (1, \infty)$, then the equality in inequality (6.2.3) in the above theorem holds true if and only if there exists a constant β such that

$$(a_i^p - \beta b_j^q)\langle e_i, g_j\rangle = 0 \text{ for all } i = 1, \dots, n \text{ and } j = 1, \dots, m \tag{6.2.5}$$

and a constant α such that

$$(\langle e_i, g_j\rangle - \alpha\langle f_i, h_j\rangle)a_i b_j = 0 \text{ for all } i = 1, \dots, n \text{ and } j = 1, \dots, m. \tag{6.2.6}$$

For the inequality (6.2.4) in the theorem, there holds equality if and only if (6.2.6) is satisfied and moreover

$$a_i(b_j - b)\langle e_i, g_j\rangle = 0 \text{ for all } i = 1, \dots, n \text{ and } j = 1, \dots, n \tag{6.2.7}$$

and some number b.

6.3 Equality cases

Example 6.3.1. Suppose that $n = m$ and that B is the identity matrix. Then the inequality (6.2.4) yields that for all $n \times n$ matrices A there is

$$|\mathrm{tr}A| \leq \mathrm{tr}|A|.$$

Equality here holds if and only if A is symmetric and semidefinite. Indeed, it holds if and only if (6.2.6) and (6.2.7 are satisfied. Note that (6.2.7) for $B = \mathrm{Id}$ clearly holds true. Condition (6.2.6) holds if and only if

$$(\langle e_i, e_j\rangle - \alpha\langle f_i, e_j\rangle)a_i = 0 \text{ for all } i = 1, \dots, n.$$

Here we took $g_j = h_j = e_j$. Equivalently, for any i such that $a_i \neq 0$ and all $j = 1, \dots, n$

$$\langle e_i - \frac{\alpha}{a_i} A e_i, e_j\rangle = 0.$$

That is

$$Ae_i = \frac{a_i}{\alpha} e_i.$$

If $a_i = 0$, then $Ae_i = 0$. We see thus that A is diagonal with eigenvalues of fixed sign and thus it is symmetric and semidefinite. Then converse implication is obvious.

Definition 6.3.2. For $p \in [1, \infty]$, the quantity $(\mathrm{tr}|A|^p)^{\frac{1}{p}}$ is called the Schatten p-norm of a matrix A. We shall denote it by $\|A\|_p$. We shall write $\langle A, B\rangle = \mathrm{tr}(AB^*)$.

Let us now analyse carefully what the conditions (6.2.6) and (6.2.7) mean.

Theorem 6.3.3. *Suppose that A, B are $m \times n$ matrices such that $\|B\| \leq 1$. Then the condition*

$$\operatorname{tr}(A^*B) = \operatorname{tr}(|A|) \tag{6.3.1}$$

holds if and only if A^*B is symmetric and positive semidefinite and

$$A^*BB^*A = A^*A. \tag{6.3.2}$$

Moreover, if $A \neq 0$, then $\|B\| = 1$.

Proof. If $A = 0$, then the equivalence clearly holds true. Suppose that $A \neq 0$ and that (6.3.1) holds true. Then, by (6.2.4), we have

$$\operatorname{tr}(|A|) = \operatorname{tr}(A^*B) \leq \operatorname{tr}(|A|)\|B\| \leq \operatorname{tr}(|A|).$$

Therefore the equality (6.3.1) holds if and only if the conditions (6.2.6) and (6.2.7) are satisfied and $\|B\| = 1$.

Suppose that these conditions hold true. Note that if $\alpha = 0$ in (6.2.6), then we would have $\operatorname{tr}(A^*B) = 0$, contrary to the assumptions. Therefore $\alpha \neq 0$. Condition (6.2.7) may be equivalently stated as

$$(b_j - b)g_j \in \ker A \text{ for all } j = 1, \ldots, m.$$

Indeed, we may write for $i = 1, \ldots, n$ annd $j = 1, \ldots, m$

$$0 = a_i(b_j - b)\langle e_i, g_j\rangle = \langle A^* f_i, (b_j - b)g_j\rangle = \langle f_i, A(b_j - b)g_j\rangle.$$

Thus we have $A(b_j - b)g_j = 0$.

Condition (6.2.6), with the above observation, implies that

$$\begin{aligned} 0 = (\langle e_i, g_j\rangle - \alpha\langle f_i, h_j\rangle)a_i b_j = \langle A^* f_i, b_j g_j\rangle - \alpha\langle Ae_i, Bg_j\rangle = \\ = \langle f_i, Ab_j g_j\rangle - \alpha\langle B^*Ae_i, g_j\rangle = \langle f_i, Abg_j\rangle - \alpha\langle B^*Ae_i, g_j\rangle = \\ = \langle ba_i e_i, g_j\rangle - \alpha\langle B^*Ae_i, g_j\rangle. \end{aligned}$$

It follows that for $i = 1, \ldots, n$

$$B^*Ae_i = a_i \frac{b}{\alpha} e_i.$$

Thus B^*A is symmetric and semidefinite. Hence

$$\operatorname{tr}(|A|) = \operatorname{tr}(A^*B) = \operatorname{tr}(|B^*A|) = \frac{b}{|\alpha|}\operatorname{tr}(|A|). \tag{6.3.3}$$

This is to say, $\frac{b}{|\alpha|} = 1$. It follows that

$$A^*BB^*A = A^*A.$$

Observe also that B^*A is positive semidefinite, as the quantities in (6.3.3) are non-negative.

Conversely, if $\|B\| = 1$, A^*B is symmetric and positive semidefinite and

$$A^*BB^*A = A^*A,$$

then

$$|A| = |A^*B| = A^*B$$

and thus

$$\operatorname{tr}(A^*B) = \operatorname{tr}(|A|).$$

□

Corollary 6.3.4. *Suppose that A, B are as above. Then*

$$B^*A = A^*B, BA^* = AB^*$$

are positive semidefinite and

$$A^*BB^*A = A^*A, AB^*BA^* = AA^*.$$

Moreover A^*A and B^*B diagonalise in a common orthonormal basis.

Proof. Observe that $\operatorname{tr}(A^*B) = \operatorname{tr}(AB^*)$ and that $\operatorname{tr}(|A|) = \operatorname{tr}(|A^*|)$. Thus the assertion follows by application of Theorem 6.3.3. The fact that A^*A and B^*B diagonalise in a common orthonormal basis is a consequence of

$$A^*AB^*B = B^*BA^*A.$$

□

Remark 6.3.5. Note that condition (6.3.2) takes the form

$$BB^* = \mathrm{Id},$$

provided that A is invertible. This is to say, B is then an isometry. If A is not invertible, in particular if $m \neq n$, then B^* is an isometry, if restricted to $\mathrm{im}A$. Indeed, for $x = Ay$, we have

$$\|B^*x\|^2 = \langle BB^*x, x\rangle = \langle A^*BB^*Ay, y\rangle = \langle A^*Ay, y\rangle = \|x\|^2.$$

Conversely, if $BB^* = \mathrm{Id}$ on $\mathrm{im}A$, then clearly $A^*BB^*A = A^*A$. Thus, in Theorem 6.3.3, we might write that the equivalent conditions are: B^*A is symmetric and positive semidefinite and that B^* is an isometry on $\mathrm{im}A$.

Theorem 6.3.6. *Suppose that $m \le n$ and that A, B are $m \times n$ matrices, $A, B \neq 0$. Let $p, q \in (1, \infty)$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then the equality*

$$\mathrm{tr}(A^*B) = (\mathrm{tr}(|A|^p))^{\frac{1}{p}} (\mathrm{tr}(|B|^q))^{\frac{1}{q}} \tag{6.3.4}$$

holds if and only if B^*A is symmetric and positive semidefinite and for some $c > 0$,

$$A^*B = B^*A = c^p (A^*A)^{\frac{p}{2}} = c^{-q}(B^*B)^{\frac{q}{2}}. \tag{6.3.5}$$

Equivalently,

$$\frac{A^*B}{(\mathrm{tr}(|A|^p))^{\frac{1}{p}} (\mathrm{tr}(|B|^q))^{\frac{1}{q}}} = \frac{B^*A}{(\mathrm{tr}(|A|^p))^{\frac{1}{p}} (\mathrm{tr}(|B|^q))^{\frac{1}{q}}} = \frac{(A^*A)^{\frac{p}{2}}}{\mathrm{tr}(|A|^p)} = \frac{(B^*B)^{\frac{q}{2}}}{\mathrm{tr}(|B|^q)}. \tag{6.3.6}$$

Proof. Let us recall – see (6.2.5) and (6.2.6) – that the equality holds if and only if

$$(a_i^p - \beta b_j^q)\langle e_i, g_j \rangle = 0 \text{ for all } i = 1, \dots, n, j = 1, \dots, m, \tag{6.3.7}$$

and

$$(\langle e_i, g_j \rangle - \alpha \langle f_i, h_j \rangle) a_i b_j = 0 \text{ for all } i = 1, \dots, n, j = 1, \dots, m. \tag{6.3.8}$$

Form (6.3.7) it follows that $\beta \neq 0$. Otherwise $A = 0$. Form (6.3.8) it follows that $\alpha \neq 0$, otherwise $A^*B = 0$. Then, as there is equality in (6.3.4), we must have $A = 0$ or $B = 0$, contrary to the assumptions. In what follows we assume therefore that $\alpha \neq 0, \beta \neq 0$. If $\langle e_i, g_j \rangle \neq 0$, then, by (6.3.7),

$$b_j = \frac{1}{\beta^{\frac{1}{q}}} a_i^{\frac{p}{q}}$$

and thus, by (6.3.8),

$$\left\langle \frac{a_i^{1+\frac{p}{q}}}{\alpha \beta^{\frac{1}{q}}} e_i - B^*Ae_i, g_j \right\rangle = 0.$$

If $\langle e_i, g_j \rangle = 0$ then the above formula holds as well, by (6.3.8). We infer that for $i = 1, \dots, n$

$$B^*Ae_i = \frac{a_i^{1+\frac{p}{q}}}{\alpha \beta^{\frac{1}{q}}} e_i.$$

Analogously we have for $j = 1, \dots, n$

$$A^*Bg_j = \frac{\beta^{\frac{1}{p}} b_j^{1+\frac{q}{p}}}{\alpha} g_j.$$

Equivalently

$$B^*A = \frac{1}{\alpha \beta^{\frac{1}{q}}} (A^*A)^{\frac{p}{2}} \text{ and } A^*B = \frac{\beta^{\frac{1}{p}}}{\alpha} (B^*B)^{\frac{q}{2}}.$$

It follows that B^*A is symmetric and semidefinite. By (6.3.4) it is positive semidefinite, so $\alpha > 0$, and therefore

$$\operatorname{tr}(B^*A) = (\operatorname{tr}(B^*A))^{\frac{1}{p}}(\operatorname{tr}(A^*B))^{\frac{1}{q}} = \frac{1}{\alpha}(\operatorname{tr}|A|^p)^{\frac{1}{p}}(\operatorname{tr}|B|^q)^{\frac{1}{q}}. \tag{6.3.9}$$

Thus $\alpha = 1$ and the equality (6.3.5) holds with $c = \beta^{-\frac{1}{pq}}$. Converse implication follows readily, as in (6.3.9). Condition (6.3.6) follows from (6.3.5) by taking traces and computing number c. □

Remark 6.3.7. The following duality formula holds true

$$\|A\|_p = \sup\{\operatorname{tr}(AB^*) \mid \|B\|_q \leq 1\},$$

where $\frac{1}{p} + \frac{1}{q} = 1$, $p, q \in [1, \infty]$. Indeed, take any matrix A and its polar factorisation $A = UD$, where U is a partial isometry and D is symmetric and positive semidefinite such that U^*U is projection onto $\operatorname{im} D$. Set $B = UD^{\frac{p}{q}}$. Then

$$\|A\|_p = \|D\|_p, \|B\|_q = \|D^{\frac{p}{q}}\|_q = \|D\|_p^{\frac{p}{q}},$$

and

$$\langle A, B\rangle = \operatorname{tr}(B^*A) = \operatorname{tr}(D^{1+\frac{p}{q}}) = \operatorname{tr}(D^p) = \|A\|_p\|B\|_q.$$

If $p = 1$, we take $B = U$. If $p = \infty$, let v be a vector such that $Dv = \lambda v$ with λ such that $|\lambda| = \|D\|$. Take $B = UD'$, where $D' = \frac{\lambda}{|\lambda|}vv^*$.

Chapter 7

Divergence formulation of optimal transport of vector measures

7.1 Introduction

In this chapter we formulate another approach to optimal transport of vector measures related to the works of Bouchitté, Buttazzo and Seppecher [28], Bouchitté and Buttazzo [27], Bouchitté, Gangbo and Seppecher [55] and Gangbo [54]. The idea is to extend the divergence formulation of optimal transport to the setting of vector measures. We refer the reader also to [103, 1.2.3]. We develop a duality theory and apply it, together with characterisation of equality cases in matrix Hölder's inequality, to reprove (see [37]) the representation formula for the polar cone of monotone maps. We obtain also several generalisations.

The advantage of the current formulation of optimal transport of vector measures to the formulation in Chapter 3 is that the primal problem's value is always attained.

Section 7.2 is devoted to a proof of the duality for optimal transport of vector measures in its divergence formulation.

In Section 7.3 we deal with duality for absolutely continuous measures.

Section 7.4 we employ results of the previous sections in order to provide characterisation of polar cones to monotone maps, reproving result of [37].

In Section 7.5 we generalise the result of Section 7.4 and provide a representation formula for polar cones to tangent cones of the unit ball of $\mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$.

In Section 7.6 we obtain another generalisation of the representation formulae for polar cones to the tangent cones of the unit ball of Sobolev space $\mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$.

7.2 Duality formula

Suppose we are given a Borel vector measure $\mu \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^m)$ of finite total variation, such that $\mu(\mathbb{R}^n) = 0$ and with finite first moments, i.e.

$$\int_{\mathbb{R}^n} \|x\| d\|\mu\|(x) < \infty.$$

Here $\|\mu\|$ denotes the total variation of μ. In what follows we shall consider a Banach space $\mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ of continuously differentiable functions with bounded derivative and with a seminorm given by

$$\|u\| = \sup\Big\{\|Du(x)\| \mid x \in \mathbb{R}^n\Big\}.$$

The seminorm induces a norm on the subspace of functions vanishing at the origin.

Definition 7.2.1. Suppose that $\mu \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^m)$ has finite first moments and $\mu(\mathbb{R}^n) = 0$. Define

$$\Xi(\mu) = \{M \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^{m\times n}) \mid -divM = \mu\}.$$

Here we write

$$-divM = \mu$$

if

$$\int_{\mathbb{R}^n} \langle f, d\mu\rangle = \int_{\mathbb{R}^n} \langle Df, dM\rangle \tag{7.2.1}$$

for all functions $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$. For any μ as above define

$$I(\mu) = \inf\{\|M\|_1 \mid M \in \Xi(\mu)\}$$

and

$$J(\mu) = \sup\Big\{\int_{\mathbb{R}^n} \langle u, d\mu\rangle \mid u \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m), \|u\| \leq 1\Big\}.$$

Above we consider the total variation $\|M\|_1$ of a matrix-valued measure M with respect to the Schatten 1-norm. Recall that the Schatten 1-norm of $A \in \mathbb{R}^{m\times n}$ is defined via the formula

$$\|A\|_1 = \operatorname{tr}(A^*A)^{\frac{1}{2}}.$$

Theorem 7.2.2. *For any $\mu \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^m)$ such that $\mu(\mathbb{R}^n) = 0$ and with finite first moments we have*

$$I(\mu) = J(\mu).$$

Moreover there exists a 1-Lipschitz $u_0 \colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$J(\mu) = \int_{\mathbb{R}^n} \langle u_0, d\mu\rangle,$$

and there exists $M_0 \in \Xi(\mu)$ such that

$$I(\mu) = \|M_0\|_1.$$

If u_0, M_0 are optimisers such that u_0 is differentiable $\|M_0\|_1$-almost everywhere, then

$$\left\langle Du_0, \frac{dM_0}{d\|M_0\|_1} \right\rangle = 1 \tag{7.2.2}$$

$\|M_0\|_1$-almost everywhere.

Proof. Define a linear functional on the space

$$\left\{ Df \mid f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m) \right\}$$

by the formula

$$\Lambda(Df) = \int_{\mathbb{R}^n} \langle f, d\mu \rangle.$$

As $\mu(\mathbb{R}^n) = 0$ it is well defined. It is continuous as

$$\Lambda(Df) = \int_{\mathbb{R}^n} \langle f(x) - f(0), d\mu(x) \rangle = \int_{\mathbb{R}^n} \int_{[0,1]} \langle Df(tx)(x), d\mu(x) \rangle d\lambda(t),$$

and thus $\|\Lambda\| \leq \int_{\mathbb{R}^n} \|x\| d\|\mu\|(x)$. By the Hahn–Banach theorem it follows that we may extend Λ to a continuous linear functional on $\mathcal{C}(\mathbb{R}^n, \mathbb{R}^{m\times n})$ preserving its norm. From Riesz' theorem it follows that the dual space of $\mathcal{C}(\mathbb{R}^n, \mathbb{R}^{m\times n})$ is isometrically isomorphic to $\mathcal{M}(\mathbb{R}^n, \mathbb{R}^{m\times n})$. We obtain a matrix-valued measure $M_0 \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^{m\times n})$ such that

$$\Lambda(Df) = \int_{\mathbb{R}^n} \langle Df, dM_0 \rangle$$

and such that $\|M_0\|_1 = \|\Lambda\|$. Moreover

$$\begin{aligned} \|\Lambda\| &= \sup \left\{ \Lambda(Df) \mid f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m), \|f\| \leq 1 \right\} = \\ &= \sup \left\{ \int_{\mathbb{R}^n} \langle f, d\mu \rangle \mid f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m), \|f\| \leq 1 \right\} = J(\mu). \end{aligned}$$

Observe that

$$J(\mu) \leq I(\mu). \tag{7.2.3}$$

As we have just found a measure M_0 with $\|M_0\|_1 = J(\mu)$ we see that there holds equality in (7.2.3). We shall now show that there exists a 1-Lipschitz map u_0 such that

$$\int_{\mathbb{R}^n} \langle u_0, d\mu \rangle = \sup \left\{ \int_{\mathbb{R}^n} \langle u, d\mu \rangle \mid f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m), \|f\| \leq 1 \right\}.$$

For this, we may use Arzelà–Ascoli theorem. Choose a sequence

$$(u_n)_{n=1}^{\infty} \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m),$$

with $\|u_n\| \leq 1$, $u_n(0) = 0$ and such that

$$\lim_{n\to\infty} \int_{\mathbb{R}^n} \langle u_n, d\mu\rangle = J(\mu).$$

We may substract a subsequence $(u_{n_k})_{k=1}^{\infty}$ that converges locally uniformly to a 1-Lipschitz function u_0. We claim that

$$\int_{\mathbb{R}^n} \langle u_0, d\mu\rangle = J(\mu). \tag{7.2.4}$$

Choose $\epsilon > 0$ and a compact set $K \subset \mathbb{R}^n$ such that

$$\int_{K^c} \|x\| d\|\mu\|(x) < \epsilon.$$

If $\|u_{n_k}(x) - u_0(x)\| < \epsilon$ for all $k \geq N$ and all $x \in K$, then

$$\Big| \int_{\mathbb{R}^n} \langle u_0, d\mu\rangle - \int_{\mathbb{R}^n} \langle u_{n_k}, d\mu\rangle \Big| \leq \int_{K^c} \|x\| d\|\mu\|(x) + \epsilon\|\mu\|(K).$$

Letting n tend to infinity and then ϵ to zero we get (7.2.4).

For the last part of the theorem observe that if $-divM = \mu$ and u_0 is differentiable $\|M\|_1$-almost everywhere, with $\|Du_0\|$ bounded above by 1, $\|M\|_1$-almost everywhere, then by matrix Hölder's inequality (see Chapter 6)

$$\langle Du_0, \frac{dM}{d\|M\|_1}\rangle \leq 1.$$

If M_0 and u_0 are the optimisers then integrating the inequality with respect to $d\|M_0\|_1$ we would get $J(\mu) \leq I(\mu)$. Hence equality holds if and only if (7.2.2) holds. □

Remark 7.2.3. The problem of finding optimal measure M_0 is a relaxation of the optimal transport of vector measures, as formulated in Chapter 3. In this problem there holds the Kantorovich–Rubinstein duality. That is, we have

$$\inf\Big\{ \int_{\mathbb{R}^n\times\mathbb{R}^n} \|x-y\| d\|\pi\|(x,y) \mid \mathrm{P}_1\pi - \mathrm{P}_2\pi = \mu\Big\} = \sup\Big\{ \int_{\mathbb{R}^n} \langle f, d\mu\rangle \mid f \text{ is 1-Lipschitz}\Big\}.$$

Using convolution with a smoothing kernel we may show that the right-hand side of the above equality is equal to

$$\sup\Big\{ \int_{\mathbb{R}^n} \langle f, d\mu\rangle \mid f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m), \|f\| \leq 1\Big\}.$$

7.3 Absolutely continuous vector measures

We shall be now interested in measures on open, bounded, connected subsets $\Omega \subset \mathbb{R}^n$ with continuously differentiable boundary $\partial\Omega$. We shall now restrict ourselves to considering matrix-valued measures $\mu \in \mathcal{M}(\Omega, \mathbb{R}^m)$ that are absolutely continuous with respect to the Lebesgue measure λ. Let $p, q \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Let

$$h \in L^q(\Omega, \mathbb{R}^m)$$

be such that

$$\int_\Omega h d\lambda = 0.$$

Let us define

$$\Gamma_\lambda^q(h) = \Big\{ H \in L^q(\Omega, \mathbb{R}^{m\times n}) \mid -divH = h \Big\}.$$

For $H \in L^q(\Omega, \mathbb{R}^{m\times n})$ we write $-divH = h$ if

$$\int_\Omega \langle Df, H \rangle d\lambda = \int_\Omega \langle f, h \rangle d\lambda,$$

for all functions $f \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$. The above condition is equivalent to demanding that the divergence of $Hd\lambda$ is equal to $-hd\lambda$. Let us recall that $\mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$ denotes the Sobolev space of Borel measurable functions f on Ω that admit a weak derivative Df and with finite norm given by the formula

$$\Big(\int_\Omega \mathrm{tr}(|Df|^p) d\lambda + \int_\Omega \|f\|^p d\lambda \Big)^{\frac{1}{p}}$$

Recall that $\mathrm{tr}(|G|^p)^{\frac{1}{p}}$ for a matrix $G \in \mathbb{R}^{m\times n}$ is called the Schatten p-norm.

Let us recall that the Poincaré inequality states that for $p \in [1, \infty]$ there exists a constant C, which depends on Ω and n, p, such that for any $f \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$

$$\int_\Omega \Big\| f - \frac{\int_\Omega f d\lambda}{\lambda(\Omega)} \Big\|^p d\lambda \leq C \int_\Omega \mathrm{tr}(|Df|^p) d\lambda.$$

For $f \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$ we shall write

$$\|Df\|_p = \Big(\int_\Omega \mathrm{tr}(|Df|^p) d\lambda \Big)^{\frac{1}{p}}.$$

Set

$$I_\lambda^q(h) = \inf \Big\{ \|H\|_q \mid H \in \Gamma_\lambda^q(h) \Big\}.$$

By $\|H\|_q$ we mean here the $L^q(\Omega, \mathbb{R}^{m\times n})$ norm with respect to the Schatten q-norm on $\mathbb{R}^{m\times n}$, i.e.

$$\|H\|_q = \Big(\int_{\mathbb{R}^n} \mathrm{tr}(|H|^q) d\lambda \Big)^{\frac{1}{q}}.$$

We set

$$J_\lambda^p(h) = \sup \Big\{ \int_{\mathbb{R}^n} \langle f, h \rangle d\lambda \mid f \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m), \|Df\|_p \leq 1 \Big\}.$$

Theorem 7.3.1. *Let $p, q \in (1, \infty)$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Let $\Omega \subset \mathbb{R}^n$ be an bounded, open and connected set with continuously differentiable boundary. Let*

$$h \in L^q(\Omega, \mathbb{R}^m)$$

be such that

$$\int_\Omega h d\lambda = 0.$$

Then

$$J^q_\lambda(h) = I^q_\lambda(h).$$

Moreover there exists $f_0 \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$, $\|Df_0\|_p \leq 1$, such that

$$J^q_\lambda(h) = \int_\Omega \langle f_0, h \rangle d\lambda,$$

and there exists $H_0 \in \Gamma^q_\lambda(h)$ such that

$$I^q_\lambda(h) = \|H_0\|_q.$$

Moreover, if $h \neq 0$, then f_0, H_0 are optimisers if and only if

$$\frac{Df_0^* H_0}{\|H_0\|_q} = \frac{H_0^* Df_0}{\|H_0\|_q} = \frac{|H_0|^q}{\|H_0\|_q^q} = |Df_0|^p \tag{7.3.1}$$

λ-almost everywhere.

Proof. Let us observe that if $h \in L^q(\Omega, \mathbb{R}^m)$, then $h \in L^1(\Omega, \mathbb{R}^m)$, as Ω is bounded. Thus the integral

$$\int_\Omega h d\lambda$$

is meaningful. Recall that the dual space of $L^p(\Omega, \mathbb{R}^{m\times n})$ is isometrically isomorphic to $L^q(\Omega, \mathbb{R}^{m\times n})$. Consider a subspace V of $L^p(\Omega, \mathbb{R}^{m\times n})$ given by

$$V = \left\{ Df \mid f \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m) \right\}$$

and a functional $\Lambda \colon V \to \mathbb{R}$ given by $\lambda(Df) = \int_\Omega \langle f, h \rangle d\lambda$. By the assumptions, it is well-defined and finite. For any function $f \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$ we have, by the Poincaré inequality,

$$\Lambda(Df) = \int_\Omega \left\langle f - \frac{\int_\Omega f d\lambda}{\lambda(\Omega)}, h \right\rangle d\lambda \leq C\|h\|_q \|Df\|_p.$$

Thus

$$\|\Lambda\| \leq C\|h\|_q.$$

By the Hahn–Banach theorem we may extend Λ to a functional Λ_0 defined on $L^p(\Omega, \mathbb{R}^{m\times n})$ and thus obtain a matrix-valued function $H_0 \in L^q(\Omega, \mathbb{R}^{m\times n})$ such that

$$\int_\Omega \langle f, h \rangle d\lambda = \int_\Omega \langle Df, H_0 \rangle d\lambda$$

for any $f \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$. Moreover

$$\|H_0\|_q = \|\Lambda_0\| = \|\Lambda\|.$$

Since clearly $J^q_\lambda(h) \leq I^q_\lambda(h)$, it follows that

$$J^q_\lambda(h) = \|\Lambda\| = \|H_0\|_q = I^q_\lambda(h).$$

We shall now show that there exists $f_0 \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$, $\|Df_0\|_p \leq 1$, with

$$\int_\Omega \langle f_0, h \rangle d\lambda = J^q_\lambda(h).$$

Since the Sobolev space $\mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$ is reflexive, its unit ball is weakly compact. The functional

$$f \mapsto \int_\Omega \langle f, h \rangle d\lambda$$

is continuous, as was shown above. By compactness, there exists f_0 that maximises its value over the unit ball. This is the function in request.

For the equality

$$\int_\Omega \langle Df_0, H_0 \rangle d\lambda = \|H_0\|_q = \Big(\int_\Omega \mathrm{tr}(|H_0|^q) \rangle d\lambda \Big)^{\frac{1}{q}}$$

to hold true, it is sufficient and necessary that

$$\mathrm{tr}(Df_0^* H_0) = \big(\mathrm{tr}(|Df_0|^p)\big)^{\frac{1}{p}} \big(\mathrm{tr}(|H_0|^q)\big)^{\frac{1}{q}}$$

λ-almost everywhere and that there exists non-negative constant c such that

$$\mathrm{tr}(|H_0|^q) = c\,\mathrm{tr}(|Df_0|^p)$$

λ-almost everywhere. In view of Theorem 6.3.6 we see that these two conditions may be equivalently stated as

$$Df_0^* H_0 = H_0^* Df_0 = d^q |H_0|^q = \frac{1}{d^p} |Df_0|^p$$

λ-almost everywhere, for some non-negative constant d. If $h \neq 0$, then $H_0 \neq 0$ and also $\|Df_0\|_p = 1$. Integrating yields

$$d \|H_0\|_q^{\frac{1}{p}} = 1.$$

It follows that the equivalent condition is that λ-almost everywhere there is

$$Df_0^* H_0 = H_0^* Df_0 = \frac{|H_0|^q}{\|H_0\|_q^{\frac{q}{p}}} = \|H_0\|_q |Df_0|^p.$$

We infer that (7.3.1) holds true. □

7.4 Polar cone of set of monotone maps

Let us now present an application of the above material. We shall provide a more general, alternative proof of the representation formula for the elements of the polar cone of the set of monotone maps. This representation formula is already proven in [37]. Let us remark that the proof is alternative, but in essence both proofs rely on the Hahn–Banach theorem.

Definition 7.4.1. We shall say that a map $u\colon \mathbb{R}^n \to \mathbb{R}^n$ is monotone if

$$\langle u(x) - u(y), x - y\rangle \geq 0$$

for all $x, y \in \mathbb{R}^n$.

Proposition 7.4.2. *A map $u \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^n)$ is monotone if and only if*

$$\langle Du(x)v, v\rangle \geq 0 \tag{7.4.1}$$

for all $x, v \in \mathbb{R}^n$.

Proof. The proof follows readily. □

Remark 7.4.3. The condition (7.4.1) means that at any point $x \in \mathbb{R}^n$ the matrix $Du(x)$ is positive semidefinite.

We shall now study the polar cone of the set of monotone maps, that is the set

$$\mathcal{P} = \Big\{\mu \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^n) \mid \int_{\mathbb{R}^n} \langle u, d\mu\rangle \geq 0 \text{ for any monotone } u \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^n)\Big\}.$$

Here $\mathcal{M}(\mathbb{R}^n, \mathbb{R}^n)$ is the space of all $\mathbb{R}^m$-valued measures.

For $\mu \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^n)$, recall the definition of $J(\mu)$; see Section 7.2.

Proposition 7.4.4. *Suppose that $\mu \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^n)$ is such that $\mu(\mathbb{R}^n) = 0$ and has finite first moments. Then $\mu \in \mathcal{P}$ if and only if*

$$\int_{\mathbb{R}^n} \langle x, d\mu(x)\rangle = J(\mu). \tag{7.4.2}$$

Proof. Suppose that (7.4.2) holds true. Let $u \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^n)$ be monotone. Let $\epsilon > 0$. We see that $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ defined by

$$f(x) = x - \epsilon u(x)$$

satisfies $\|Df\| \leq \sqrt{1 + \epsilon^2\|Du\|^2}$. Indeed, by monotonicity of u, for any $x, v \in \mathbb{R}^n$ we have

$$\|Df(x)v\|^2 = \|v - \epsilon Du(x)v\|^2 = \|v\|^2 + \epsilon^2\|Du(x)v\| - 2\epsilon\langle v, Du(x)v\rangle \leq \|v\|^2 + \epsilon^2\|Du\|^2\|v\|^2.$$

Let $\frac{1}{c} = \sqrt{1 + \epsilon^2 \|Du\|^2}$. Then, as cf is 1-Lipschitz, by (7.4.2),

$$\int_{\mathbb{R}^n} \langle x - cf(x), d\mu(x) \rangle \geq 0$$

and therefore

$$\int_{\mathbb{R}^n} \langle u, d\mu \rangle \geq \frac{\epsilon \|Du\|^2}{1 + \sqrt{1 + \epsilon^2 \|Du\|^2}} \int_{\mathbb{R}^n} \langle x, d\mu(x) \rangle$$

for any $\epsilon > 0$. Letting ϵ tend to zero, we obtain $\mu \in \mathcal{P}$.

Assume now that $\mu \in \mathcal{P}$. We shall show that (7.4.2) holds true. Take any $h \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^n)$ with $\|Dh\| \leq 1$. Then for any $x \in \mathbb{R}^n$ and any $v \in \mathbb{R}^n$

$$\langle v - Dh(x)v, v \rangle \geq 0.$$

Thus, by Proposition 7.4.2, $\mathrm{id} - h$ is monotone. As $\mu \in \mathcal{P}$, we have

$$\int_{\mathbb{R}^n} \langle x, d\mu(x) \rangle \geq \int_{\mathbb{R}^n} \langle h, d\mu \rangle.$$

□

Let us denote by $\mathcal{S}_+^{n \times n}$ the set of all $n \times n$ positive semidefinite symmetric martices.

Theorem 7.4.5. *For any $\mu \in \mathcal{P}$ with finite first moments, there exists a matrix-valued measure $M \in \mathcal{M}(\mathbb{R}^n, \mathcal{S}_+^{n \times n})$ such that*

$$\mu = -divM$$

and

$$\int_{\mathbb{R}^n} \mathrm{tr}(dM) = \int_{\mathbb{R}^n} \langle x, d\mu(x) \rangle.$$

Conversely, if a divergence of an $\mathcal{S}_+^{n \times n}$-valued measure is a measure, then it belongs to $\mathcal{P}$.

Proof. Suppose that $\mu \in \mathcal{P}$ has finite first moments. Then we know that

$$\int_{\mathbb{R}^n} \langle x, d\mu(x) \rangle = J(\mu).$$

Let M_0 be such that $\|M_0\|_1$ is minimal on $\Xi(\mu)$, i.e. on the set of measures M such that

$$-divM = \mu.$$

Then by Theorem 7.2.2 we know that

$$\left\langle \mathrm{Id}, \frac{dM_0}{d\|M_0\|_1} \right\rangle = 1$$

$\|M_0\|_1$-almost everywhere. By Example 6.3.1 it follows that this holds if and only if $\frac{dM_0}{d\|M_0\|_1}$ is positive semidefinite. Then,

$$\int_{\mathbb{R}^n} \langle x, d\mu(x)\rangle = \|M_0\|_1 = \int_{\mathbb{R}^n} \mathrm{tr}(dM_0).$$

This proves the first part of the theorem. For the second, let $M \in \mathcal{M}(\mathbb{R}^n, \mathcal{S}_+^{n\times n})$ be such that $\mu = -divM$. Then, as M is positive semidefinite,

$$\langle \mathrm{Id}, \frac{dM}{d\|M\|_1}\rangle = \Big\|\frac{dM}{d\|M\|_1}\Big\|_1 = 1.$$

It follows that

$$J(\mu) \geq \int_{\mathbb{R}^n} \langle x, d\mu(x)\rangle = \|M\|_1 \geq J(\mu).$$

and by Proposition 7.4.4 we conclude the proof. □

Remark 7.4.6. The advantage of the proof above is that the representation is a direct consequence of the duality formula and Example 6.3.1. Thus the proof may be adapted to computation of the representations of other polar cones.

7.5 Polar cones to tangent cones of $\mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$

Below we shall apply Theorem 7.2.2 to compute the representations of certain tangent cones to the unit ball of $\mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$.

Definition 7.5.1. For $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$, $\|f\| = 1$ define

$$\mathcal{P}_f = \Big\{\mu \in \mathcal{M}_0(\mathbb{R}^n, \mathbb{R}^m) \mid J(\mu) = \int_{\mathbb{R}^n} \langle f, d\mu\rangle\Big\}.$$

Let us note that the set $\mathcal{P}_f$ is the polar cone to the tangent cone of the unit ball at $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$.

Proposition 7.5.2. *Suppose that $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$, $\|f\| = 1$. Measure μ with finite first moments belongs to $\mathcal{P}_f$ if and only if for any map $h \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ such that for $\epsilon > 0$ the Lipschitz constants λ_ϵ of $f - \epsilon h$ satisfy*

$$\liminf_{\epsilon\to 0^+} \frac{1-\lambda_\epsilon}{\epsilon} \geq 0 \tag{7.5.1}$$

there is $\int_{\mathbb{R}^n} \langle h, d\mu\rangle \geq 0$.

Proof. Let $\mu \in \mathcal{P}_f$. Suppose that for $h \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ condition (7.5.1) holds true. For such h, consider

$$g = f - \epsilon h.$$

Then $\frac{1}{\lambda_\epsilon} g$ is 1-Lipschitz and hence

$$\int_{\mathbb{R}^n} \langle f, d\mu \rangle \geq \frac{1}{\lambda_\epsilon} \int_{\mathbb{R}^n} \langle g, d\mu \rangle.$$

It follows that

$$\int_{\mathbb{R}^n} \langle h, d\mu \rangle \geq \frac{1-\lambda_\epsilon}{\epsilon} \int_{\mathbb{R}^n} \langle f, d\mu \rangle.$$

Note that $\int_{\mathbb{R}^n} \langle f, d\mu \rangle \geq 0$ – otherwise $-f$ would yield greater value of the integral. Thus the assertion follows by taking limit.

Conversely, suppose that for any $h \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ such that (7.5.1) holds true there is $\int_{\mathbb{R}^n} \langle h, d\mu \rangle \geq 0$. Choose any $g \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ with $\|Dg\| \leq 1$. Let $h = f - g$. Then the corresponding Lipschitz constant of $f - \epsilon h = (1-\epsilon) f + \epsilon g$ is at most one. Hence, the condition (7.5.1) is satisfied for h and the claim follows. □

Remark 7.5.3. The proposition above tells us that $\mathcal{P}_f$ is the polar cone to the convex cone $\mathcal{H}_f$ of functions $h \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^n)$ that satisfy (7.5.1). Let us observe that this condition indeed defines a convex cone. Clearly if $h \in \mathcal{H}_f$ then also $\lambda h \in \mathcal{H}_f$ for any non-negative λ. If $h_1, h_2 \in \mathcal{H}_f$, then $\frac{h_1+h_2}{2} \in \mathcal{H}_f$, as Lipschitz constant of convex combination of functions is at most the convex combination of the corresponding Lipschitz constants.

Let us note that, when f is the identity map, the condition simplifies to demand that h is monotone.

Theorem 7.5.4. *Suppose that $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$. For any $\mu \in \mathcal{P}_f$ there exists a measure $M \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^{m\times n})$ with*

$$-divM = \mu,$$

and such that $\|M\|_1$-almost everywhere

$$DfDf^* = \mathrm{Id} \text{ on } \mathrm{im}\Big(\frac{dM}{d\|M\|_1}\Big), \tag{7.5.2}$$

and

$$Df^* \frac{dM}{d\|M\|_1} \text{ is symmetric and positive semidefinite.} \tag{7.5.3}$$

Moreover

$$\|M\|_1 = \int_{\mathbb{R}^n} \langle f, d\mu \rangle.$$

Conversely, if $M \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^{m\times n})$ is such that $-divM$ is a finite vector-valued measure such that (7.5.2) and (7.5.3) are satisfied then $-divM \in \mathcal{P}_f$.

Proof. Follows from Theorem 6.3.3 and from Theorem 7.2.2, c.f. proof of Theorem 7.4.5. □

7.6 Polar cones to tangent cones of $\mathcal{W}^{1,p}(\mathbb{R}^n, \mathbb{R}^m)$.

The method developed above may be as well applied in the contex of absolutely continuous vector measures with use of results of Section 7.3.

Definition 7.6.1. Let $p, q \in (1, \infty)$ be such that $\frac{1}{p} + \frac{1}{q} = 1$ and let $\Omega \subset \mathbb{R}^n$ be an open set. For $f \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$, $\|Df\|_p = 1$ define

$$\mathcal{R}_f = \Big\{ h \in L^q(\Omega, \mathbb{R}^m) \mid J^q_\lambda(h) = \int_\Omega \langle f, h \rangle d\lambda \Big\}.$$

In the above definition $\mathcal{R}_f$ is the polar cone to the tangent cone of the unit ball of $\mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$.

Proposition 7.6.2. *Let $p, q \in (1, \infty)$ be such that $\frac{1}{p} + \frac{1}{q} = 1$ and let let $\Omega \subset \mathbb{R}^n$ be an open set. Let $f \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$ be such that $\|Df\|_p = 1$. Then $h \in L^q(\Omega, \mathbb{R}^m)$ belongs to $\mathcal{R}_f$ if and only if for any map $g \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$ such that*

$$\liminf_{\epsilon \to 0^+} \frac{1 - \|Df - \epsilon Dg\|_p}{\epsilon} \geq 0$$

there is $\int_\Omega \langle h, g \rangle d\lambda \geq 0$.

Proof. The proof follows analogous lines to the lines of the proof of Proposition 7.5.2. □

Theorem 7.6.3. *Let $p, q \in (1, \infty)$ be such that $\frac{1}{p} + \frac{1}{q} = 1$ and let $\Omega \subset \mathbb{R}^n$ be an open, bounded and connected set with continuously differentiable boundary. Let $f \in \mathcal{W}^{1,p}(\Omega, \mathbb{R}^m)$ be such that $\|Df\|_p = 1$. For any $h \in \mathcal{R}_f$, $h \neq 0$, there exists $H \in L^q(\Omega, \mathbb{R}^{n \times m})$ such that*

$$-divH = h$$

and such that λ-almost everywhere

$$\frac{Df^* H}{\|H\|_q} = \frac{H^* Df}{\|H\|_q} = |Df|^p = \frac{|H|^q}{\|H\|_q^q}. \tag{7.6.1}$$

Conversely if $H \in L^q(\Omega, \mathbb{R}^{n \times m})$ is such that $-divH$ is a function in $L^q(\Omega, \mathbb{R}^m)$, that satisfies (7.6.1), then $-divH \in \mathcal{R}_f$.

Proof. Again, the proof relies on Theorem 7.3.1 and the strategy does not differ from the strategies in Theorem 7.4.5 and in Theorem 7.5.4. □

Let us apply the above results to a particular case when $p = 2$. We shall below compute the polar cone to the cone of maps $g \in \mathcal{W}^{1,2}(\Omega, \mathbb{R}^m)$ such that $\int_\Omega \langle Dg, Df \rangle d\lambda \geq 0$.

Corollary 7.6.4. *Let $\Omega \subset \mathbb{R}^n$ be an open, bounded and connected set with continuously differentiable boundary. Let $f \in \mathcal{W}^{1,2}(\Omega, \mathbb{R}^m)$. For $h \in L^2(\Omega, \mathbb{R}^n)$ the following conditions are equivalent:*

i) there exists $H \in L^2(\Omega, \mathbb{R}^{n\times m})$ such that $-divH = h$ and λ-almost everywhere there is

$$\frac{Df^*H}{\|H\|_2} = \frac{H^*Df}{\|H\|_2} = |Df|^2 = \frac{|H|^2}{\|H\|_2^2}. \tag{7.6.2}$$

ii) for any $g \in \mathcal{W}^{1,2}(\Omega, \mathbb{R}^n)$ such that

$$\int_\Omega \langle Dg, Df\rangle d\lambda \geq 0$$

there is $\int_\Omega \langle g, h\rangle d\lambda \geq 0$.

Proof. Observe that if $p = 2$, then Proposition 7.6.2 yields that $\mathcal{R}_f$ consists of exactly these $h \in L^2(\Omega, \mathbb{R}^m)$ such that if $g \in \mathcal{W}^{1,2}(\Omega, \mathbb{R}^m)$ satisfies

$$\int_\Omega \langle Df, Dg\rangle d\lambda \geq 0$$

then $\int_\Omega \langle h, g\rangle d\lambda \geq 0$. Indeed, we have for $\epsilon > 0$,

$$\frac{1 - \|Df - \epsilon Dg\|_2}{\epsilon} = \frac{1 - \|Df - \epsilon Dg\|_2^2}{\epsilon(1 + \|Df - \epsilon Dg\|_2)} = \frac{-\epsilon\|Dg\|_2 + 2\int_\Omega \langle Df, Dg\rangle d\lambda}{1 + \|Df - \epsilon Dg\|_2}$$

and this quantity converges to $\int_\Omega \langle Df, Dg\rangle d\lambda$ as ϵ tends to zero. Now, Theorem 7.6.3 tells us that there exists $H \in L^2(\Omega, \mathbb{R}^{n\times n})$ such that $-divH = h$ and such that (7.6.2) holds true. ☐

Remark 7.6.5. Observe that the limit in Proposition 7.6.2 is in fact the lower Dini derivative of the norm of f taken in directioin of g. Using this observation it is easy to extend result of Corollary 7.6.4 to other values of $p \in (1, \infty)$.

References

[1] L. Aguirre Salazar and S. Reich. A remark on weakly contractive mappings. *J. Nonlinear Convex Anal.*, 16(4):767–773, 2015.

[2] A.V. Akopyan and A.S. Tarasov. A contructive proof of Kirszbraun's theorem. *Mathematical Notes*, 84(5):725–728, 2008.

[3] E. M. Alfsen. *Compact convex sets and boundary integrals*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer-Verlag, 1971.

[4] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis: a Hitchhiker's Guide*. Springer, Berlin; London, 2006.

[5] L. Ambrosio. *Lecture notes on optimal transport problems*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[6] L. Ambrosio. Calculus, heat flow and curvature-dimension bounds in metric measure spaces. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. I. Plenary lectures*, pages 301–340. World Sci. Publ., Hackensack, NJ, 2018.

[7] L. Ambrosio, B. Kirchheim, and D. Preiss. Personal communication of [8].

[8] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In *Optimal transportation and applications (Martina Franca, 2001)*, volume 1813 of *Lecture Notes in Math.*, pages 123–160. Springer, Berlin, 2003.

[9] D. Azagra, E. Le Gruyer, and C. Mudarra. Kirszbraun's theorem via an explicit formula. *https://arxiv.org/pdf/1810.10288.pdf*, 10 2018.

[10] D. Aze and J.-P. Penot. Uniformly convex and uniformly smooth convex functions. *Annales de la Faculté des sciences de Toulouse : Mathématiques*, Ser. 6, 4(4):705–730, 1995.

[11] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. In *Lectures on probability theory (Saint-Flour, 1992)*, volume 1581 of *Lecture Notes in Math.*, pages 1–114. Springer, Berlin, 1994.

[12] D. Bakry and M. Émery. Diffusions hypercontractives. In Jacques Azéma and Marc Yor, editors, *Séminaire de Probabilités XIX 1983/84*, pages 177–206, Berlin, Heidelberg, 1985. Springer Berlin Heidelberg.

[13] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion operators*. Grundlehren der mathematischen Wissenschaften, Vol. 348. Springer, January 2014.

[14] B. Baumgartner. An inequality for the trace of matrix products, using absolute values. *arXiv e-prints*, page arXiv:1106.6189, Jun 2011.

[15] H. Bauschke. Fenchel duality, Fitzpatrick functions and the extension of firmly nonexpansive mappings. *Proceedings of the American Mathematical Society*, 135(1):135–139, 2007.

[16] H. Bauschke and X. Wang. Firmly nonexpansive and Kirszbraun-Valentine extensions: a constructive approach via monotone operator theory. *Nonlinear Analysis and Optimization I: Analysis, Contemporary Mathematics (eds. A. Leizarowitz et al.)*, pages 55–64, 2010.

[17] G. Beer. A Polish topology for the closed subsets of a Polish space. *Proceedings of the American Mathematical Society*, 113(4):1123–1133, 1991.

[18] G. Beer. Wijsman convergence: a survey. *Set-Valued Analysis*, 2(1):77–94, Mar 1994.

[19] M. Beiglböck, A. M. G. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. *Inventiones mathematicae*, 208(2):327–400, May 2017.

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

[21] M. Beiglböck and C. Griessler. A land of monotone plenty. *Annali della SNS*, 2018.

[22] M. Beiglböck, T. Lim, and J. Obłój. Dual attainment for the martingale transport problem. *Bernoulli*, 25(3):1640–1658, 08 2019.

[23] R. Bhatia. *Matrix analysis*. Springer New York, New York, NY, 1997.

[24] S. Bianchini and S. Daneri. On Sudakov's type decomposition of transference plans with norm costs. *Memoirs of the American Mathematical Society*, 251, Nov 2013.

[25] V. I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

[26] V.I. Bogachev and A.V. Kolesnikov. The Monge-Kantorovich problem: achievements, connections, and prospects. *Uspekhi Mat. Nauk*, 67(5(407)):3–110, 2012.

[27] G. Bouchitté and G. Buttazzo. Characterization of optimal shapes and masses through Monge-Kantorovich equation. *Journal of the European Mathematical Society*, 3(2):139–168, 2001.

[28] G. Bouchitté, G. Buttazzo, and P. Seppecher. Shape optimization solutions via Monge-Kantorovich equation. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 324(10):1185 – 1191, 1997.

[29] J. Bourgain. Geometry of Banach spaces and harmonic analysis. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986)*, pages 871–878. Amer. Math. Soc., Providence, RI, 1987.

[30] U. Brehm. Extensions of distance reducing mappings to piecewise congruent mappings on $\mathbf{R}^m$. *J. Geom.*, 16(2):187–193, 1981.

[31] Y. Brudnyi and P. Shvartsman. Whitney's extension problem for multivariate $C^{1,\omega}$ - functions. *Transactions of the American Mathematical Society*, 353(6):2487–2512, 2001.

[32] L. Caffarelli, M. Feldman, and R. J. McCann. Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *Journal of the American Mathematical Society*, 15(1):1–26, 2002.

[33] L. Caravenna. A proof of Sudakov theorem with strictly convex norms. *Mathematische Zeitschrift*, 268(1):371–407, Jun 2011.

[34] L. Caravenna and S. Daneri. The disintegration of the Lebesgue measure on the faces of a convex function. *Journal of Functional Analysis*, 258(11):3604 – 3661, 2010.

[35] F. Cavalletti and A. Mondino. Sharp and rigid isoperimetric inequalities in metric-measure spaces with lower Ricci curvature bounds. *Inventiones mathematicae*, 208(3):803–849, 2017.

[36] F. Cavalletti and A. Mondino. Sharp geometric and functional inequalities in metric measure spaces with lower Ricci curvature bounds. *Geom. Topol.*, 21(1):603–645, 2017.

[37] F. Cavalletti and M. Westdickenberg. The polar cone of the set of monotone maps. *Proc. Amer. Math. Soc.*, 143(2):781–787, 2015.

[38] Y. Chen, T. Georgiou, and A. Tannenbaum. Vector-valued optimal mass transport. *SIAM Journal on Applied Mathematics*, 78:1682–1696, 2018.

[39] Y. Chen, E. Haber, K. Yamamoto, T. T. Georgiou, and A. Tannenbaum. An efficient algorithm for matrix-valued and vector-valued optimal mass transport. *Journal of Scientific Computing*, 77(1):79–100, 2018.

[40] B. Dacorogna and W. Gangbo. Extension theorems for vector valued maps. *Journal de Mathématiques Pures et Appliquées*, 85(3):313 – 344, 2006.

[41] H. De March. Local structure of multi-dimensional martingale optimal transport. *arXiv e-prints*, page arXiv:1805.09469, May 2018.

[42] H. De March. Quasi-sure duality for multi-dimensional martingale optimal transport. *arXiv e-prints*, page arXiv:1805.01757, May 2018.

[43] H. De March and N. Touzi. Irreducible convex paving for decomposition of multi-dimensional martingale transport plans. *arXiv e-prints*, page arXiv:1702.08298, Feb 2017.

[44] M. Erbar, K. Kuwada, and K.-T. Sturm. On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces. *Invent. Math.*, 201(3):993–1071, 2015.

[45] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions.* Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[46] L.C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.*, 137(653):viii+66, 1999.

[47] H. Federer. *Geometric measure theory.* Grundlehren der mathematischen Wissenschaften. Springer, 1969.

[48] C. L. Fefferman. A sharp form of Whitney's extension theorem. *Annals of Mathematics*, 161(1):509–577, 2005.

[49] C. L. Fefferman. Whitney's extension problem for C^m. *Annals of Mathematics*, 164(1):313–359, 2006.

[50] C. L. Fefferman. C^m extension by linear operators. *Annals of Mathematics*, 166(3):779–835, 2007.

[51] C. L. Fefferman. Whitney's extension problems and interpolation of data. *Bull. Amer. Math. Soc.*, 46(2):207–220, 2009.

[52] M. Feldman and R.J. McCann. Monge's transport problem on a Riemannian manifold. *Trans. Amer. Math. Soc.*, 354:1667–1697, 2002.

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

[54] W. Gangbo. Michell trusses and existence of lines of principal actions. *Preprint*, 2004.

[55] W. Gangbo, G. Bouchitte, and P. Seppecher. Michell trusses and lines of principal actions. *Math. Models Meth. Applied Sci.*, 28(9):1571–1603, 2008.

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

[57] N. Ghoussoub, Y.-H. Kim, and T. Lim. Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.*, 47(1):109–164, 01 2019.

[58] M. Gromov. Filling Riemannian manifolds. *J. Differential Geom.*, 18(1):1–147, 1983.

[59] M. Gromov. *Partial differential relations*, volume 9 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1986.

[60] M. Gromov and V. D. Milman. Generalization of the spherical isoperimetric inequality to uniformly convex Banach spaces. *Compositio Math.*, 62(3):263–282, 1987.

[61] L. Guth. The waist inequality in Gromov's work. In Ragni Piene Helge Holden, editor, *The Abel Prize 2008-2012*, pages 181–195. Springer Verlag, 2014.

[62] E. Helly. Über Mengen konvexer Körper mit gemeinschaftlichen Punkte. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 32:175–176, 1923.

[63] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete Comput. Geom.*, 13(3-4):541–559, 1995.

[64] L. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 1942.

[65] L.V. Kantorovich. On the translocation of masses. *Journal of Mathematical Sciences*, 133(4):1381–1382, Mar 2006.

[66] H. G. Kellerer. Duality theorems for marginal problems. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 67(4):399–432, Nov 1984.

[67] Y. Kim and B. Pass. A general condition for Monge solutions in the multi-marginal optimal transport problem. *SIAM Journal on Mathematical Analysis*, 46(2):1538–1550, 2014.

[68] M. Kirszbraun. Über die zusammenziehende und Lipschitzsche Transformationen. *Fundamenta Mathematicae*, 22(1):77–108, 1934.

[69] B. Klartag. Convex geometry and waist inequalities. *Geom. Funct. Anal.*, 27(1):130–164, 2017.

[70] B. Klartag. Needle decompositions in Riemannian geometry. *Memoirs of the American Mathematical Society*, 249(1180), Jun 2017.

[71] E. Kopecká. Bootstrapping Kirszbraun's extension theorem. *Fund. Math.*, 217(1):13–19, 2012.

[72] E. Kopecká. Extending Lipschitz mappings continuously. *J. Appl. Anal.*, 18(2):167–177, 2012.

[73] E. Kopecká and S. Reich. Continuous extension operators and convexity. *Nonlinear Analysis: Theory, Methods & Applications*, 74(18):6907 – 6910, 2011.

[74] U. Lang and V. Schroeder. Kirszbraun's theorem and metric spaces of bounded curvature. *Geometric & Functional Analysis GAFA*, 7(3):535–560, Jul 1997.

[75] E. Le Gruyer. Minimal Lipschitz extensions to differentiable functions defined on a Hilbert space. *Geometric and Functional Analysis*, 19(4):1101–1118, Nov 2009.

[76] E. Le Gruyer and T. V. Phan. Sup-inf explicit formulas for minimal Lipschitz extensions for 1-fields on $\mathbb{R}^n$. *Journal of Mathematical Analysis and Applications*, 424(2):1161 – 1185, 2015.

[77] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.

[78] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Structures Algorithms*, 4(4):359–412, 1993.

[79] E. Matoušková, S. Reich, and A. J. Zaslavski. Genericity in nonexpansive mapping theory. In *Advanced courses of mathematical analysis I*, pages 81–98. World Sci. Publ., Hackensack, NJ, 2004.

[80] E. J. McShane. Extension of range of functions. *Bull. Amer. Math. Soc.*, 40(12):837–842, 1934.

[81] P.A. Meyer. *Probability and potentials*. Blaisdell book in pure and applied mathematics. Blaisdell Pub. Co., 1966.

[82] E. Milman and J. Neeman. The Gaussian Double-Bubble Conjecture. *arXiv e-prints*, page arXiv:1801.09296, Jan 2018.

[83] E. Milman and J. Neeman. The Gaussian Multi-Bubble Conjecture. *arXiv e-prints*, page arXiv:1805.10961, May 2018.

[84] G. J. Minty. On the extension of Lipschitz, Lipschitz-Hölder continuous, and monotone functions. *Bull. Amer. Math. Soc.*, 76:334–339, 1970.

[85] G. Monge. Mémoire sur la théorie des déblais et des remblais. In *Histoire de l'Académie Royale de Sciences de Paris*, pages 666–704. 1781.

[86] J. Obłój. The Skorokhod embedding problem and its offspring. *Probab. Surv.*, 1:321–390, 2004.

[87] J. Obłój and P. Siorpaes. Structure of martingale transports in finite dimensions. *arXiv e-prints*, page arXiv:1702.08433, Feb 2017.

[88] S.-I. Ohta. Needle decompositions and isoperimetric inequalities in Finsler geometry. *J. Math. Soc. Japan*, 70(2):651–693, 2018.

[89] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.*, 5:286–292 (1960), 1960.

[90] R.R. Phelps. *Lectures on Choquet's Theorem*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2003.

[91] S. Reich and S. Simons. Fenchel duality, Fitzpatrick functions and the Kirszbraun–Valentine extension theorem. *Proceedings of the American Mathematical Society*, 133(9):2657–2660, 2005.

[92] W. Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York-Toronto, Ont.-London, 1966.

[93] E. K. Ryu, Y. Chen, W. Li, and S. Osher. Vector and matrix optimal mass transport: theory, algorithm, and applications. *Preprint*, 2017.

[94] I. J. Schoenberg. On a Theorem of Kirzbraun and Valentine. *The American Mathematical Monthly*, 60(9):620–622, 1953.

[95] S. Sheffield and C. K. Smart. Vector-valued optimal Lipschitz extensions. *Communications on Pure and Applied Mathematics*, 65(1):128–154, 2012.

[96] E.M. Stein. *Singular Integrals and Differentiability Properties of Functions (PMS-30).* Princeton University Press, 1970.

[97] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36(2):423–439, 04 1965.

[98] K.-T. Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.

[99] K.-T. Sturm. On the geometry of metric measure spaces. II. *Acta Math.*, 196(1):133–177, 2006.

[100] V.N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. *Trudy Mat. Inst. Steklov.*, 141, 1976. *Proc. Steklov Inst. Math., 141* (1976).

[101] N. S. Trudinger and X.-J. Wang. On the Monge mass transfer problem. *Calculus of Variations and Partial Differential Equations*, 13(1):19–31, 2001.

[102] F.A. Valentine. A Lipschitz condition preserving extension for a vector function. *Amer. J. Math.*, 67:83–93, 1945.

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

[104] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[105] H. Whitney. Analytic extensions of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36(1):63–89, 1934.

[106] R.A. Wijsman. Convergence of sequences of convex sets, cones and functions. ii. *Transactions of the American Mathematical Society*, 123(1):32–45, 1966.

[107] G. Winkler. Extreme points of moment sets. *Mathematics of Operations Research*, 13(4):581–587, 1988.

[108] C. Zălinescu. On uniformly convex functions. *J. Math. Anal. Appl.*, 95(2):344–374, 1983.