Notes of the seminar Evolution Equations in Probability Spaces and the Continuity Equation

Onno van Gaans

Version of 12 April 2006

These are some notes supporting the seminar *Evolution Equations in Probability Spaces and the Continuity Equation* organized by Philippe Clément and Onno van Gaans at Leiden University during Spring 2006.

Contents

1 Probability measures on metric spaces

When we study curves in spaces of probability measures we will be faced with continuity and other regularity properties and therefore with convergence of probability measures. The probability measures will be defined on the Borel σ-algebra of a metric space. Since we want to be able to apply the results to probability measures on a Hilbert space, it is not too restrictive to assume separability and completeness but we should avoid assuming compactness of the metric space.

We will consider Borel probability measures on metric spaces, narrow convergence of such measures, a metric for narrow convergence, and Prokhorov's theorem on compactness relative to the narrow convergence.

1.1 Borel sets

Let (X, d) be a metric space. The *Borel σ-algebra* (σ-field) $\mathcal{B} = \mathcal{B}(X)$ is the smallest σ-algebra in X that contains all open subsets of X. The elements of $\mathcal{B}$ are called the *Borel sets* of X.

The metric space (X, d) is called *separable* if it has a countable dense subset, that is, there are $x_1, x_2, \ldots$ in X such that $\overline{\{x_1, x_2, \ldots\}} = X$. ($\overline{A}$ denotes the closure of $A \subset X$.)

Lemma 1.1. *If X is a separable metric space, then $\mathcal{B}(X)$ equals the σ-algebra generated by the open (or closed) balls of X.*

Proof. Denote

$$\mathcal{A} := \sigma\text{-algebra generated by the open (or closed) balls of } X.$$

Clearly, $\mathcal{A} \subset \mathcal{B}$.

Let D be a countable dense set in X. Let $U \subset X$ be open. For $x \in U$ take $r > 0$, $r \in \mathbb{Q}$ such that $B(x, r) \subset U$ ($B(x, r)$ open or closed ball with center x and radius r) and take $y_x \in D \cap B(x, r/3)$. Then $x \in B(y_x, r/2) \subset B(x, r)$. Set $r_x := r/2$. Then

$$U = \bigcup \{B(y_x, r_x) : x \in U\},$$

which is a countable union. Therefore $U \in \mathcal{A}$. Hence $\mathcal{B} \subset \mathcal{A}$. □

Lemma 1.2. *Let (X, d) be a separable metric space. Let $\mathcal{C} \subset \mathcal{B}$ be countable. If $\mathcal{C}$ separates closed balls from points in the sense that for every closed ball B and every $x \in X \setminus B$ there exists $C \in \mathcal{C}$ such that $B \subset C$ and $x \notin C$, then the σ-algebra generated by $\mathcal{C}$ is the Borel σ-algebra.*

Proof. Clearly $\sigma(\mathcal{C}) \subset \mathcal{B}$, where $\sigma(\mathcal{C})$ denotes the σ-algebra generated by $\mathcal{C}$. Let B be a closed ball in X. Then $B = \bigcap\{C \in \mathcal{C} : B \subset C\}$, which is a countable intersection and hence a member of $\sigma(\mathcal{C})$. By the previous lemma we obtain $\mathcal{B} \subset \sigma(\mathcal{C})$. □

If $f : S \to T$ and $\mathcal{A}_S$ and $\mathcal{A}_T$ are σ-algebras in S and T, respectively, then f is called *measurable* (w.r.t. $\mathcal{A}_S$ and $\mathcal{A}_T$) if

$$f^{-1}(A) = \{x \in S : f(x) \in A\} \in \mathcal{A}_S \text{ for all } A \in \mathcal{A}_T.$$

Proposition 1.3. *Let (X, d) be a metric space. $\mathcal{B}(X)$ is the smallest σ-algebra with respect to which all (real valued) continuous functions on X are measurable (w.r.t. $\mathcal{B}(X)$ and $\mathcal{B}(\mathbb{R})$).* (See [10, Theorem I.1.7, p. 4].)

1.2 Borel probability measures

Let (X, d) be a metric space. A *finite Borel measure* on X is a map $\mu : \mathcal{B}(X) \to [0, \infty)$ such that

$$\begin{aligned}&\mu(\emptyset) = 0, \text{ and}\\ &A_1, A_2, \ldots \in \mathcal{B} \text{ mutually disjoint } \implies \mu(\textstyle\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mu(B_i).\end{aligned}$$

μ is called a *Borel probabiliy measure* if in addition $\mu(X) = 1$.

The following well known continuity properties will be used several times.

Lemma 1.4. *Let X be a metric space and μ a finite Borel measure on X. Let $A_1, A_2, \dots$ be Borel sets.*

(1) If $A_1 \subset A_2 \subset \cdots$ and $A = \bigcup_{i=1}^{\infty} A_i$, then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.

(2) If $A_1 \supset A_2 \supset \cdots$ and $A = \bigcap_{i=1}^{\infty}$, then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.

The next observation is important in the proof of Theorem 1.13 (the Portmanteau theorem).

Lemma 1.5. *If μ is a finite Borel measure on X and $\mathcal{A}$ is a collection of mutually disjoint Borel sets of X, then at most countably many elements of $\mathcal{A}$ have nonzero μ-measure.*

Proof. For $m \geq 1$, let $\mathcal{A}_m := \{A \in \mathcal{A} : \mu(A) > 1/m\}$. For any distinct $A_1, \dots, A_k \in \mathcal{A}_m$ we have

$$\mu(X) \geq \mu(\bigcup_{i=1}^{k} A_i) = \mu(A_1) + \cdots + \mu(A_k) > k/m,$$

hence $\mathcal{A}_m$ has at most $m\mu(X)$ elements. Thus

$$\{A \in \mathcal{A} : \mu(A) > 0\} = \bigcup_{m=1}^{\infty} \mathcal{A}_m$$

is countable. $\square$

Example. If μ is a finite Borel measure on $\mathbb{R}$, then $\mu(\{t\}) = 0$ for all except at most countably many $t \in \mathbb{R}$.

Proposition 1.6. *Any finite Borel measure on X is* regular*, that is, for every $B \in \mathcal{B}$*

$$\begin{aligned}\mu(B) &= \sup\{\mu(C) : C \subset B,\ C \text{ closed}\} \quad \textit{(inner regular)}\\ &= \inf\{\mu(U) : U \supset B,\ U \text{ open}\} \quad \textit{(outer regular)}.\end{aligned}$$

Proof. Define the collection $\mathcal{R}$ by

$$A \in \mathcal{R} \iff \begin{array}{l}\mu(A) = \sup\{\mu(C) : C \subset A,\ C\text{closed}\} \text{ and}\\ \mu(A) = \inf\{\mu(C) : U \supset A,\ U \text{ open}\}.\end{array}$$

We have to show that $\mathcal{R}$ contains the Borel sets. *step 1: $\mathcal{R}$ is a σ-algebra:* $\emptyset \in \mathcal{R}$. Let $A \in \mathcal{R}$, let $\varepsilon > 0$. Take C closed and U open with $C \subset A \subset U$ and $\mu(A) < \mu(C) + \varepsilon$, $\mu(A) > \mu(U) - \varepsilon$. Then $U^c \subset A^c \subset C^c$, U^c is closed, C^c is open, and

$$\begin{aligned}\mu(A^c) = \mu(X) - \mu(A) > \mu(X) - \mu(C) - \varepsilon = \mu(C^c) - \varepsilon,\\ \mu(A^c) = \mu(X) - \mu(A) < \mu(X) - \mu(U) + \varepsilon = \mu(U^c) + \varepsilon.\end{aligned}$$

Hence $A^c \in \mathcal{R}$.

Let $A_1, A_2, \ldots \in \mathcal{R}$ and let $\varepsilon > 0$. Take for each i

U_i open , C_i closed with
$C_i \subset A \subset U_i$,
$\mu(U_i) - \mu(A_i) < 2^{-i}\varepsilon,\ \mu(A_i) - \mu(C_i) < 2^{-i}\varepsilon/2.$

Then $\bigcup_i C_i \subset \bigcup_i A_i \subset \bigcup_i U_i$ and $\bigcup_i U_i$ is open, and

$$\begin{aligned}
\mu(\bigcup_i U_i) - \mu(\bigcup_i A_i) &\leq \mu\Big(\bigcup_{i=1}^{\infty} U_i \setminus \bigcup_{i=1}^{\infty} A_i\Big) \\
&\leq \mu\Big(\bigcup_{i=1}^{\infty} (U_i \setminus A_i)\Big) \leq \sum_{i=1}^{\infty} \mu(U_i \setminus A_i) \\
&= \sum_{i=1}^{\infty} (\mu(U_i) - \mu(A_i)) < \sum_{i=1}^{\infty} 2^{-i}\varepsilon = \varepsilon.
\end{aligned}$$

Further, $\mu(\bigcup_{i=1}^{\infty} C_i) = \lim_{k\to\infty} \mu(\bigcup_{i=1}^{k} C_i)$, hence for some large k, $\mu(\bigcup_{i=1}^{\infty} C_i) - \mu(\bigcup_{i=1}^{k} C_i) < \varepsilon/2$. Then $C := \bigcup_{i=1}^{k} C_i \subset \bigcup_{i=1}^{\infty} A_i$, C is closed, and

$$\begin{aligned}
\mu(\bigcup_{i=1}^{\infty} A_i) - \mu(C) &< \mu(\bigcup_{i=1}^{\infty} A_i) - \mu(\bigcup_{i=1}^{\infty} C_i) + \varepsilon/2 \\
&\leq \mu\Big(\bigcup_{i=1}^{\infty} A_i \setminus \bigcup_{i=1}^{\infty} C_i\Big) + \varepsilon/2 \\
&\leq \mu\Big(\bigcup_{i=1}^{\infty} (A_i \setminus C_i)\Big) + \varepsilon/2 \\
&\leq \sum_{i=1}^{\infty} \mu(A_i \setminus C_i) + \varepsilon/2 \\
&= \sum_{i=1}^{\infty} \Big(\mu(A_i) - \mu(C_i)\Big) + \varepsilon/2 \leq \varepsilon/2 + \varepsilon/2.
\end{aligned}$$

Hence $\bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$. Thus $\mathcal{R}$ is a σ-algebra.

step2: $\mathcal{R}$ contains all open sets: We prove: $\mathcal{R}$ contains all closed sets. Let $A \subset X$ be closed. Let $U_n := \{x \in X : d(x, A) < 1/n\} = \{x \in X : \exists\, a \in A \text{ with } d(a, x) < 1/n\}$, $n = 1, 2, \ldots$. Then U_n is open, $U_1 \supset U_2 \supset \cdots$, and $\bigcap_{i=1}^{\infty} U_i = A$, as A is closed. Hence $\mu(A) = \lim_{n\to\infty} \mu(U_n) = \inf_n \mu(U_n)$. So

$$\mu(A) \leq \inf\{\mu(U) : U \supset A,\ U \text{ open}\} \leq \inf_n \mu(U_n) = \mu(A).$$

Hence $A \in \mathcal{R}$.

Conclusion: $\mathcal{R}$ is a σ-algebra that contains all open sets, so $\mathcal{R} \supset \mathcal{B}$. □

Corollary 1.7. *If μ and ν are finite Borel measures on the metric space X and $\mu(A) = \nu(A)$ for all closed A (or all open A), then $\mu = \nu$.*

A finite Borel measure μ on X is called *tight* if for every $\varepsilon > 0$ there exists a compact set $K \subset X$ such that $\mu(X \setminus K) < \varepsilon$, or, equivalently, $\mu(K) \geq \mu(X) - \varepsilon$. A tight finite Borel measure is also called a *Radon measure.*

Corollary 1.8. *If μ is a tight finite Borel measure on the metric space X, then*

$$\mu(A) = \sup\{\mu(K) : K \subset A,\ K \textit{ compact}\}$$

for every Borel set A in X.

Proof. Take for every $\varepsilon > 0$ a compact set K_ε such that $\mu(X \setminus K_\varepsilon) < \varepsilon$. Then

$$\mu(A \cap K_\varepsilon) = \mu(A \setminus K_\varepsilon^c) \geq \mu(A) - \mu(K_\varepsilon^c) > \mu(A) - \varepsilon$$

and

$$\begin{aligned} \mu(A \cap K_\varepsilon) &= \sup\{\mu(C) : C \subset K_\varepsilon \cap A,\ C \text{ closed}\} \\ &\leq \sup\{\mu(K) : K \subset A,\ K \text{ compact}\}, \end{aligned}$$

because each closed subset contained in a compact set is compact. Combination completes the proof. □

Of course, if (X, d) is a compact metric space, then every finite Borel measure on X is tight. There is another interesting case. A complete separable metric space is sometimes called a *Polish space.*

Theorem 1.9. *If (X, d) is a complete separable metric space, then every finite Borel measure on X is tight.*

We need a lemma from topology.

Lemma 1.10. *If (X, d) is a complete metric space, then a closed set K in X is compact if and only if it is totally bounded, that is, for every $\varepsilon > 0$ the set K is covered by finitely many balls (open or closed) of radius less than or equal to ε.*

Proof. $\Rightarrow$) Clear: the covering with all ε-balls with centers in K has a finite subcovering.

$\Leftarrow$) Let $(x_n)_n$ be a sequence in K. For each $m \geq 1$ there are finitely many $1/m$-balls that cover K, at least one of which contains x_n for infinitely many n. For $m = 1$ take a ball B_1 with radius ≤ 1 such that $N_1 := \{n : x_n \in B_1\}$ is infinite, and take $n_1 \in N_1$. Take a ball B_2 with radius $\leq 1/2$ such that $N_2 := \{n > n_1 : x_n \in B_2 \cap B_1\}$ is infinite, and take $n_2 \in N_2$. Take B_3, radius $\leq 1/3$, $N_3 := \{n > n_2 : x_n \in B_3 \cap B_2 \cap B_1\}$ infinite, $n_3 \in N_3$. And so on.

Thus $(x_{n_k})_k$ is a subsequence of $(x_n)_n$ and since $x_{n_\ell} \in B_k$ for all $\ell \geq k$, $(x_{n_k})_k$ is a Cauchy sequence. As X is complete, $(x_n)_n$ converges in X and as K is closed, the limit is in K. So $(x_n)_n$ has a convergent subsequence and K is compact. □

Proof of Theorem 1.9. We have to prove that for every $\varepsilon > 0$ there exists a compact set K such that $\mu(X \setminus K) < \varepsilon$. Let $D = \{a_1, a_2, \ldots\}$ be a countable dense subset of X. Then for each $\delta > 0$, $\bigcup_{k=1}^\infty B(a_k, \delta) = X$. Hence $\mu(X) = \lim_{n\to\infty} \mu(\bigcup_{k=1}^n B(a_k, \delta))$ for all $\delta > 0$. Let $\varepsilon > 0$. Then there is for each $m \geq 1$ an n_m such that

$$\mu\Big(\bigcup_{k=1}^{n_m} B(a_k, 1/m)\Big) > \mu(X) - 2^{-m}\varepsilon.$$

Let

$$K := \bigcap_{m=1}^\infty \bigcup_{k=1}^{n_m} \overline{B}(a_k, 1/m).$$

Then K is closed and for each $\delta > 0$,

$$K \subset \bigcup_{k=1}^{n_m} \overline{B}(a_k, 1/m) \subset \bigcup_{k=1}^{n_m} B(a_k, \delta)$$

if we choose $m > 1/\delta$. So K is compact, by the previous lemma. Further,

$$\begin{aligned}\mu(X \setminus K) &= \mu\Big(\bigcup_{m=1}^{\infty}(X \setminus \bigcup_{k=1}^{n_m} \overline{B}(a_k, 1/m))\Big) \le \sum_{m=1}^{\infty} \mu\Big(X \setminus \bigcup_{k=1}^{n_m} \overline{B}(a_k, 1/m)\Big) \\ &= \sum_{m=1}^{\infty}\Big(\mu(X) - \mu(\bigcup_{k=1}^{n_m} \overline{B}(a_k, 1/m))\Big) < \sum_{m=1}^{\infty} 2^{-m}\varepsilon = \varepsilon.\end{aligned}$$

□

1.3 Narrow convergence of measures

Let (X, d) be a metric space and denote

$$C_b(X) := \{f : X \to \mathbb{R} : f \text{ is continuous and bounded}\}.$$

Each $f \in C_b(X)$ is integrable with respect to any finite Borel measure on X.

Definition 1.11. Let $\mu, \mu_1, \mu_2, \ldots$ be finite Borel measures on X. We say that the sequence $(\mu_i)_i$ *converges narrowly* to μ if

$$\int f \,\mathrm{d}\mu_i \to \int f \,\mathrm{d}\mu \text{ as } i \to \infty \text{ for all } f \in C_b(X).$$

We will simply use the notation $\mu_i \to \mu$. (There is at most one such a limit μ, as follows from the metrization by the bounded Lipschitz metric, which is discussed in the next section.)

Narrow convergence can be described by means of other classes of functions than the bounded continuous ones. Recall that a function f from a metric space (X, d) into $\mathbb{R}$ is called *lower semicontinuous* (l.s.c.) if for every $x, x_1, x_2, \ldots$ with $x_i \to x$ one has

$$f(x) \le \liminf_{i\to\infty} f(x_i)$$

and *upper semicontinuous* (u.s.c.) if

$$f(x) \ge \limsup_{i\to\infty} f(x_i).$$

The limits here may be ∞ or $-\infty$ and then the usual order on $[-\infty, \infty]$ is considered. The indicator function of an open set is l.s.c. and the indicator function of a closed set is u.s.c.

Proposition 1.12. *Let (X, d) be a metric space and let $\mu, \mu_1, \mu_2, \ldots$ be Borel probability measures on X. The following four statements are equivalent:*

(a) $\mu_i \to \mu$, *that is,* $\int f \,\mathrm{d}\mu_i \to \int f \,\mathrm{d}\mu$ *for every* $f \in C_b(X)$

(b) $\int f \,\mathrm{d}\mu_i \to \int f \,\mathrm{d}\mu$ *for every bounded Lipschitz function* $f : X \to \mathbb{R}$

(c) $\liminf_{i\to\infty} \int f \,\mathrm{d}\mu_i \ge \int f \,\mathrm{d}\mu$ *for every l.s.c. function* $f : X \to \mathbb{R}$ *that is bounded from below*

(c') $\limsup_{i\to\infty} \int f \,\mathrm{d}\mu_i \le \int f \,\mathrm{d}\mu$ *for every u.s.c. function* $f : X \to \mathbb{R}$ *that is bounded from above.*

Proof. (a)⇒(b) is clear.

(b)⇒(c): First assume that f is bounded. Define for $n \in \mathbb{N}$ the *Moreau-Yosida* approximation

$$f_n(x) := \inf_{y \in X} |f(y) + nd(x, y)|, \quad x \in X.$$

Then clearly $\inf f \le f_0 \le f_1 \le f_2 \le \cdots \le f$, so that, in particular, f_n is bounded for each n. Further, f_n is Lipschitz. Indeed, let $u, v \in X$ and observe that for $y \in X$ we have

$$\begin{aligned} f_n(u) - \Big(f(y) + nd(v, y)\Big) &\le \Big(f(y) + nd(u, y)\Big) - \Big(f(y) + nd(v, y)\Big) \\ &\le nd(u, v). \end{aligned}$$

If we take supremum over y we obtain $f_n(u) - f_n(v) \le nd(u, v)$. By changing the role of u and v we infer

$$|f_n(u) - f_n(v)| \le nd(u, v),$$

so f_n is Lipschitz.

Next we show that $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$. For $x \in X$ and $n \ge 1$ there is a $y_n \in X$ such that

$$f_n(x) \ge f(y_n) + nd(x, y_n) - 1/n \ge \inf f + nd(x, y_n) - 1, \tag{1}$$

so

$$nd(x, y_n) \le f_n(x) - \inf f + 1 \le f(x) - \inf f + 1 \quad \text{for all } n,$$

hence $y_n \to x$ as $n \to \infty$. Then (1) yields

$$\liminf_{n\to\infty} f_n(x) \ge \liminf_{n\to\infty} f(y_n) \ge f(x),$$

as f is l.s.c. Since $f_n(x) \le f(x)$ for all n, we obtain that $f_n(x)$ converges to $f(x)$.

Due to the monotone convergence, $\int f_n \,\mathrm{d}\mu \uparrow \int f \,\mathrm{d}\mu$. As $f \ge f_n$,

$$\liminf_{i\to\infty} \int f \,\mathrm{d}\mu_i \ge \liminf_{i\to\infty} \int f_n \,\mathrm{d}\mu_i = \int f_n \,\mathrm{d}\mu$$

for all n, by (b). Hence $\liminf_{i\to\infty} \int f \,\mathrm{d}\mu_i \ge \int f \,\mathrm{d}\mu$.

If f is not bounded from above, let $m \in \mathbb{N}$ and truncate f at m: $f \wedge m = x \mapsto \min\{f(x), m\}$. The above conclusion applied to $f \wedge m$ yields,

$$\int f \wedge m \,\mathrm{d}\mu \le \liminf_{i\to\infty} \int f \wedge m \,\mathrm{d}\mu_i \le \liminf_{i\to\infty} \int f \,\mathrm{d}\mu_i$$

and $\int f \,\mathrm{d}\mu = \lim_{m\to\infty} \int f \wedge m \,\mathrm{d}\mu \le \liminf_{i\to\infty} \int f \,\mathrm{d}\mu_i$.

(c)⇔(c'): multiply by -1.

(c)⇒(a): if f is continuous and bounded, we have (c) both for f and $-f$. □

Narrow convergence can also be described as convergence on sets.

Theorem 1.13 (Portmanteau theorem). *Let (X, d) be a metric space and let $\mu, \mu_1, \mu_2, \ldots$ be Borel probability measures on X. The following four statements are equivalent:*

(a) $\mu_i \to \mu$ (narrow convergence)

(b) $\liminf_{i\to\infty}\mu_i(U)\geq\mu(U)$ *for all open* $U\subset X$

(b') $\limsup_{i\to\infty}\mu_i(C)\leq\mu(C)$ *for all closed* $C\subset X$

(c) $\mu_i(A)\to\mu(A)$ *for every Borel set* A *in* X *with* $\mu(\partial A)=0$. *(Here*$\partial A=\overline{A}\setminus A^\circ$*.)*

Proof. (a)⇒(b): If U is open, then the indicator function $\mathbb{1}_U$ of U is l.s.c. So by the previous proposition,

$$\liminf_{i\to\infty}\mu_i(U)=\liminf_{i\to\infty}\int\mathbb{1}_U\,\mathrm{d}\mu_i\geq\int\mathbb{1}_U\,\mathrm{d}\mu=\mu(U).$$

(b)⇒(b'): By complements,

$$\begin{aligned}\limsup_{i\to\infty}\mu_i(C) &= \limsup_{i\to\infty}\Big(\mu_i(X)-\mu_i(C^c)\Big)=1-\liminf_{i\to\infty}\mu_i(C^c)\\ &\geq 1-\mu(C^c)=\mu(X)-\mu(C^c)=\mu(C).\end{aligned}$$

(b')⇒(b): Similarly.

(b)+(b')⇒(c): $A^\circ\subset A\subset\overline{A}$, A° is open and $\overline{A}$ is closed, so by (b) and (b'),

$$\begin{aligned}\limsup\mu_i(A) &\leq \limsup\mu_i(\overline{A})\leq\mu(\overline{A})=\mu(A\cup\partial A)\\ &\leq \mu(A)+\mu(\partial A)=\mu(A),\\ \liminf\mu_i(A) &\geq \liminf\mu_i(A^\circ)\geq\mu(A^\circ)=\mu(A\setminus\partial A)\\ &\geq \mu(A)-\mu(\partial A)=\mu(A),\end{aligned}$$

hence $\mu_i(A)\to\mu(A)$.

(c)⇒(a): Let $g\in C_b(X)$. Idea: we have $\int f\,\mathrm{d}\mu_i\to\int f\,\mathrm{d}\mu$ for suitable simple functions; we want to approximate g to get $\int g\,\mathrm{d}\mu_i\to\int g\,\mathrm{d}\mu$.

Define

$$\nu(E):=\mu(\{x:g(x)\in E\})=\mu(g^{-1}(E)),\quad E\text{ Borel set in }\mathbb{R}.$$

Then ν is a finite Borel measure (probability measure) on $\mathbb{R}$ and if we take $a<-\|g\|_\infty$, $b>\|g\|_\infty$, then $\nu(\mathbb{R}\setminus(a,b))=0$. As ν is finite, there are at most countably many α with $\nu(\{\alpha\})>0$ (see Lemma 1.5). Hence for $\varepsilon>0$ there are $t_0,\dots,t_m\in\mathbb{R}$ such that

$$\begin{array}{rl}\text{(i)} & a=t_0<t_1<\cdots<t_m=b,\\ \text{(ii)} & t_j-t_{j-1}<\varepsilon,\ j=1,\dots,m,\\ \text{(iii)} & \nu(\{t_j\})=0,\text{i.e., }\mu(\{x:g(x)=t_j\})=0,\ j=0,\dots,m.\end{array}$$

Take

$$A_j:=\{x\in X:t_{j-1}\leq g(x)<t_j\}=g^{-1}([t_{j-1},t_j)),\ j=1,\dots,m.$$

Then $A_j\in\mathcal{B}(X)$ for all j and $X=\bigcup_{j=1}^m A_j$. Further,

$$\begin{array}{l}\overline{A}_j\subset\{x:t_{j-1}\leq g(x)\leq t_j\}\text{ (since this set is closed and }\supset A_j),\\ A_j^\circ\supset\{x:t_{j-1}<g(x)<t_j\}\text{ (since this set is open and }\subset A_j),\end{array}$$

so

$$\begin{aligned}\mu(\partial A_j) &= \mu(\overline{A}_j\setminus A_j^\circ)\leq\mu(\{x:g(x)=t_{j-1}\text{ or }g(x)=t_j\})\\ &= \mu(\{x:g(x)=t_{j-1}\})+\mu(\{x:g(x)=t_j\})=0+0.\end{aligned}$$

Hence by (e), $\mu_i(A_j) \to \mu(A_j)$ as $i \to \infty$ for $j = 1, \dots, m$. Put

$$h := \sum_{j=1}^{m} t_{j-1} \mathbb{1}_{A_j},$$

then $h(x) \le g(x) \le h(x) + \varepsilon$ for all $x \in X$. Hence

$$\begin{aligned} |\int g \, \mathrm{d}\mu_i - \int g \, \mathrm{d}\mu| &= |\int (g-h) \, \mathrm{d}\mu_i + \int h \, \mathrm{d}\mu_i - \int (g-h) \, \mathrm{d}\mu - \int h \, \mathrm{d}\mu| \\ &\le \int |g-h| \, \mathrm{d}\mu_i + |\int h \, \mathrm{d}\mu_i - \int h \, \mathrm{d}\mu| + \int |g-h| \, \mathrm{d}\mu \\ &\le \varepsilon \mu_i(X) + |\sum_{j=1}^{m} t_{j-1} \Big(\mu_i(A_j) - \mu(A_j) \Big)| + \varepsilon \mu(X). \end{aligned}$$

It follows that $\limsup_{i\to\infty} |\int g \, \mathrm{d}\mu_i - \int g \, \mathrm{d}\mu| \le 2\varepsilon$. Thus $\int g \, \mathrm{d}\mu_i \to \int g \, \mathrm{d}\mu$ as $i \to \infty$. $\square$

1.4 The bounded Lipschitz metric

Let (X, d) be a metric space. Denote

$$\mathcal{P} = \mathcal{P}(X) := \text{all Borel probability measures on } X.$$

We have defined the notion of narrow convergence in $\mathcal{P}$. We will show next that narrow convergence is induced by a metric, provided that X is separable. This results goes back to Prokhorov [11]. Instead of Prokhorov's metric, we will consider the "bounded Lipschitz metric" due to Dudley [4], as it is easier to work with. (See also [5, 15].) Denote

$$\mathrm{BL}(X, d) := \{ f : X \to \mathbb{R} \colon f \text{ is bounded and Lipschitz} \}.$$

Define for $f \in \mathrm{BL}(X, d)$

$$\|f\|_{\mathrm{BL}} = \|f\|_\infty + \mathrm{Lip}(f),$$

where

$$\|f\|_\infty := \sup_{x \in X} |f(x)|$$

and

$$\mathrm{Lip}(f) := \sup_{x,y \in X,\ x \ne y} \frac{|f(x) - f(y)|}{d(x,y)} = \inf\{ L : |f(x) - f(y)| \le L d(x,y) \ \forall x, y \in X \}.$$

Then $\| \cdot \|_{\mathrm{BL}}$ is a norm on $\mathrm{BL}(X, d)$. Define for $\mu, \nu \in \mathcal{P}(X)$

$$d_{\mathrm{BL}}(\mu, \nu) := \sup\{ |\int f \, \mathrm{d}\mu - \int f \, \mathrm{d}\nu| \colon f \in \mathrm{BL}(X, d), \ \|f\|_{\mathrm{BL}} \le 1 \}.$$

The function d_{BL} is called the *bounded Lipschitz metric on* $\mathcal{P}$ (induced by d), which makes sense because of the next theorem.

Theorem 1.14 (Dudley, 1966). *Let (X, d) be a metric space.*

(1) d_{BL} is a metric on $\mathcal{P} = \mathcal{P}(X)$.

(2) If X is separable and $\mu, \mu_1, \mu_2, \ldots \in \mathcal{P}$, then

$$\mu_i \to \mu \text{ (narrowly)} \quad \Longleftrightarrow \quad \mu_{\mathrm{BL}}(\mu_i, \mu) \to 0.$$

Proof. (See [5, Theorem 11.3.3, p. 395].)

(1): To show the triangle inequality, let $\mu, \nu, \eta \in \mathcal{P}(X)$ and observe that

$$|\int f \,\mathrm{d}\mu - \int f \,\mathrm{d}\eta| \leq |\int f \,\mathrm{d}\mu - \int f \,\mathrm{d}\nu| + |\int f \,\mathrm{d}\nu - \int f \,\mathrm{d}\eta| \quad \forall f \in \mathrm{BL}(X,d),$$

so $d_{\mathrm{BL}}(\mu, \eta) \leq d_{\mathrm{BL}}(\mu, \nu) + d_{\mathrm{BL}}(\nu, \eta)$. Clearly, $d_{\mathrm{BL}}(\mu, \nu) = d_{\mathrm{BL}}(\nu, \mu)$ and $d_{\mathrm{BL}}(\mu, \mu) = 0$. If $d_{\mathrm{BL}}(\mu, \nu) = 0$, then $\int f \,\mathrm{d}\mu = \int f \,\mathrm{d}\nu$ for all $f \in \mathrm{BL}(X,d)$. Therefore the constant sequence $\mu, \mu, \ldots$ converges narrowly to ν and $\nu, \nu, \ldots$ converges to μ. The Portmanteau theorem then yields $\nu(U) \leq \mu(U)$ and $\mu(U) \leq \nu(U)$ hence $\mu(U) = \nu(U)$ for any open $U \subseteq X$. By outer regularity of both μ and ν it follows that $\mu = \nu$. Thus d_{BL} is a metric on $\mathcal{P}$.

(2): If $d_{\mathrm{BL}}(\mu_i, \mu) \to 0$, then $\int f \,\mathrm{d}\mu_i \to \int f \,\mathrm{d}\mu$ for all $f \in \mathrm{BL}(X,d)$ with $\|f\|_{\mathrm{BL}} \leq 1$ and hence for all $f \in \mathrm{BL}(X,d)$. With the aid of Proposition 1.12 we infer that μ_i converges narrowly to μ.

Conversely, assume that μ_i converges narrowly to μ, that is, $\int f \,\mathrm{d}\mu_i \to \int f \,\mathrm{d}\mu$ for all $f \in C_b(X)$. Denote

$$B := \{f \in \mathrm{BL}(X,d) : \|f\|_{\mathrm{BL}} \leq 1\}.$$

In order to show that $d_{\mathrm{BL}}(\mu_i, \mu) \to 0$ we have to show that $\int f \,\mathrm{d}\mu_i$ converges uniformly in $f \in B$. If X were compact, we could use the Arzela-Ascoli theorem and reduce to a finite set of functions f. As X may not be compact, we will first call upon Theorem 1.9.

Let $\hat{X}$ be the completion of the metric space (X,d). Every $f \in B$ extends uniquely to an $\hat{f} : \hat{X} \to \mathbb{R}$ with $\|\hat{f}\|_{\mathrm{BL}} = \|f\|_{\mathrm{BL}}$. Also μ extends to $\hat{X}$:

$$\hat{\mu}(A) := \mu(A \cap X), \quad A \subseteq \hat{X} \text{ Borel}.$$

Let $\varepsilon > 0$. By the lemma, there exists a compact set $K \subseteq \hat{X}$ such that $\hat{\mu}(K) \geq 1 - \varepsilon$. The set $G := \{\hat{f}|_K : f \in B\}$ is equicontinuous and uniformly bounded, so by the Arzela-Ascoli theorem (see [5, Theorem 2.4.7, p. 52]) it is relatively compact in $(C(K), \|\cdot\|_\infty)$. Hence there are $f_1, \ldots, f_m \in B$ such that

$$\forall f \in B \;\exists \ell \text{ such that } \|\hat{f}|_K - \hat{f}_\ell|_K\|_\infty < \varepsilon \tag{2}$$

(the ε-balls around the f_i cover B). Take N such that

$$|\int_X f_\ell \,\mathrm{d}\mu_i - \int_X f_\ell \,\mathrm{d}\mu| < \varepsilon$$

for $k = 1, \ldots, N$ and $i \geq N$. Let $f \in B$ and choose a corresponding ℓ as in (2). Denote

$$K_\varepsilon = \{x \in X : \mathrm{dist}(x, K) < \varepsilon\},$$

which is an open set in X. (Here $\mathrm{dist}(x,K) := \inf\{d(x,y) \colon y \in K\}$.) For $x \in K_\varepsilon$, take $y \in K$ with $d(x,y) < \varepsilon$, then

$$\begin{aligned} |f(x) - f_\ell(x)| &\leq |f(x) - \hat{f}(y)| + |\hat{f}(y) - \hat{f}_\ell(y)| + |\hat{f}_\ell(y) - f_\ell(x)| \\ &< \mathrm{Lip}(\hat{f}) d(x,y) + \varepsilon + \mathrm{Lip}(\hat{f}_\ell) d(y,x) \\ &< 3\varepsilon. \end{aligned}$$

Further, $X \setminus K_\varepsilon$ is closed, so

$$\limsup_{i\to\infty} \mu_i(X \setminus K_\varepsilon) \le \mu(X \setminus K_\varepsilon) \le \mu(X \setminus K) = \hat{\mu}(\hat{X} \setminus K) \le \varepsilon,$$

so there is an M with $\mu_i(X \setminus K_\varepsilon) \le \varepsilon$ for all $i \ge M$. Hence for $i \ge N \vee M$,

$$\begin{aligned} |\int_X f \,\mathrm{d}\mu_i - \int_X f \,\mathrm{d}\mu| &\le |\int_X f_\ell \,\mathrm{d}\mu_i - \int_X f_\ell \,\mathrm{d}\mu| + \int_{K_\varepsilon} |f_\ell - f| \,\mathrm{d}(\mu_i + \mu) \\ &\quad + \int_{X\setminus K_\varepsilon} |f_\ell - f| \,\mathrm{d}(\mu_i + \mu) \\ &< \varepsilon + 6\varepsilon + \int_{X\setminus K_\varepsilon} 2 \,\mathrm{d}\mu_i + \int_{X\setminus K_\varepsilon} 2 \,\mathrm{d}\mu \\ &\le 11\varepsilon, \end{aligned}$$

hence $d_{\mathrm{BL}}(\mu_i, \mu) \le 11\varepsilon$ for $i \ge N \vee M$. Thus, $d_{\mathrm{BL}}(\mu_i, \mu) \to 0$ as $i \to \infty$. □

Proposition 1.15. *Let (X, d) be a separable metric space. Then $\mathcal{P} = \mathcal{P}(X)$ with the bounded Lipschitz metric d_{BL} is separable.*

Proof. Let $D := \{a_1, a_2, \ldots\}$ be a countable set in X. Let

$$\mathcal{M} := \{\alpha_1\delta_{a_1} + \cdots + \alpha_k\delta_{a_k} : \alpha_1, \ldots, \alpha_k \in \mathbb{Q} \cap [0,1],\ \sum_{j=1}^{k} \alpha_j = 1,\ k = 1, 2, \ldots\}.$$

(Here δ_a denotes the Dirac measure at $a \in X$: $\delta_a(A) = 1$ if $a \in A$, 0 otherwise.) Clearly, $\mathcal{M} \subset \mathcal{P}$ and $\mathcal{M}$ is countable.

Claim: $\mathcal{M}$ is dense in $\mathcal{P}$. Indeed, let $\mu \in \mathcal{P}$. For each $m \ge 1$, $\bigcup_{j=1}^{\infty} B(a_j, 1/m) = X$. Take k_m such that

$$\mu(\bigcup_{j=1}^{k_m} B(a_j, 1/m)) \ge 1 - 1/m.$$

Modify the balls $B(a_j, 1/m)$ into disjoint sets by taking $A_1^m := B(a_1, 1/m)$, $A_j^m := B(a_j, 1/m) \setminus \left[\bigcup_{i=1}^{j-1} B(a_i, 1/m)\right]$, $j = 2, \ldots, k_m$. Then $A_1^m, \ldots, A_{k_m}^m$ are disjoint and $\bigcup_{i=1}^{j} A_i^m = \bigcup_{i=1}^{j} B(a_i, 1/m)$ for all j. In particular, $\mu(\bigcup_{j=1}^{k_m} A_j^m) \ge 1 - 1/m$, so

$$\sum_{j=1}^{k_m} \mu(A_j^m) \in [1 - 1/m, 1].$$

We approximate

$$\mu(A_1^m)\delta_{a_1} + \cdots + \mu(A_{k_m}^m)\delta_{a_{k_m}}$$

by

$$\mu_m := \alpha_1^m \delta_{a_1} + \cdots + \alpha_{k_m}^m \delta_{a_{k_m}},$$

where we choose $\alpha_j^m \in [0,1] \cap \mathbb{Q}$ such that $\sum_{j=1}^{k_m} \alpha_j^m = 1$ and

$$\sum_{j=1}^{k_m} |\mu(A_j^m) - \alpha_j^m| < 2/m.$$

(First take $\beta_j \in [0,1] \cap \mathbb{Q}$ with $\sum_{j=1}^{k_m} |\beta_j - \mu(A_j^m)| < 1/2m$, then $\sum_j \beta_j \in [1 - 3/2m, 1 + 1/2m]$. Take $\alpha_j := \beta_j / \sum_i \beta_i \in [0,1] \cap \mathbb{Q}$, then $\sum_j \alpha_j = 1$ and $\sum_{j=1}^{k_m} |\beta_j - \alpha_j| = |1 - 1/\sum_i \beta_i| \sum_{j=1}^{k_m} \beta_j = |\sum_i \beta_j - 1| \leq 3/2m$, so $\sum_{j=1}^{k_m} |\alpha_j - \mu(A_j^m)| < 1/2m + 3/2m = 2/m$.)

Then for each m, $\mu_m \in \mathcal{M}$. To show: $\mu_m \to \mu$ in $\mathcal{P}$, that is, $\mu_n \to \mu$ narrowly. Let $g \in \mathrm{BL}(X,d)$. Then

$$
\begin{aligned}
\Big|\int g \, \mathrm{d}\mu_m - \int g \, \mathrm{d}\mu\Big| &= \Big|\sum_{j=1}^{k_m} \alpha_j^m g(a_j) - \int g \, \mathrm{d}\mu\Big| \\
&\leq \Big|\sum_{j=1}^{k_m} \mu(A_j^m) g(a_j) - \int g \, \mathrm{d}\mu\Big| + (2/m) \sup_j |g(a_j)| \\
&\leq \Big|\int \sum_{j=1}^{k_m} g(a_j) \mathbb{1}_{A_j^m} \, \mathrm{d}\mu - \int g \, \mathrm{d}\mu\Big| + (2/m) \|g\|_\infty \\
&\leq \Big|\sum_{j=1}^{k_m} \int \Big(g(a_j)\mathbb{1}_{A_j^m} - g\mathbb{1}_{A_j^m}\Big) \, \mathrm{d}\mu - \int g \mathbb{1}_{(\bigcup_{j=1}^{k_m})^c} \, \mathrm{d}\mu\Big| + (2/m)\|g\|_\infty \\
&\leq \sum_{j=1}^{k_m} \sup_{x \in A_j^m} |g(a_j) - g(x)| \mu(A_j^m) + \|g\|_\infty \mu\Big((\bigcup_{j=1}^{k_m} A_j^m)^c\Big) + (2/m)\|g\|_\infty \\
&\leq \sum_{j=1}^{k_m} \mathrm{Lip}(g)(1/m)\mu(A_j^m) + (3/m)\|g\|_\infty \\
&\leq (3/m)\|g\|_{\mathrm{BL}}.
\end{aligned}
$$

Hence $\int g \, \mathrm{d}\mu_m \to \int g \, \mathrm{d}\mu$ as $m \to \infty$. Thus, $\mu_m \to \mu$. $\square$

Conclusion. If (X,d) is a separable metric space, then so is $\mathcal{P}(X)$ with the induced bounded Lipschitz metric. Moreover, a sequence in $\mathcal{P}(X)$ converges in metric if and only if it converges narrowly and then in both senses to the same limit.

1.5 Measures as functionals

Let (X,d) be a metric space. The space of real valued bounded continuous functions $C_b(X)$ endowed with the supremum norm $\|\cdot\|$ is a Banach space. It is sometimes convenient to apply functional analytic results about the Banach space $(C_b(X), \|\cdot\|_\infty)$ to the set of Borel probability measures on X. We will for instance need the Riesz representation theorem in the proof of Prokhorov's theorem. Let us consider the relation between measures and functionals.

Recall that a linear map $\varphi : C_b(X) \to \mathbb{R}$ is called a *bounded functional* if

$$|\varphi(f)| \leq M\|f\|_\infty \quad \text{for all } f \in C_b(X)$$

for some constant M. The space of all bounded linear functionals on $C_b(X)$ is denoted by

$$C_b(X)' := \{\varphi : C_b(X) \to \mathbb{R} \colon \varphi \text{ is linear and bounded}\}$$

and called the (Banach) *dual space* of $C_b(X)$. A norm on $C_b(X)'$ is defined by

$$\|\varphi\| = \sup\{|\varphi(f)| : f \in C_b(X),\ \|f\|_\infty \leq 1\}, \quad \varphi \in C_b(X)'.$$

A functional $\varphi \in C_b(X)'$ is called *positive* if $\varphi(f) \geq 0$ for all $f \in C_b(X)$ with $f \geq 0$.

For each finite Borel measure μ on a metric space (X, d), the map φ_μ defined by

$$\varphi_\mu(f) := \int f \, \mathrm{d}\mu, \quad f \in C_b(X),$$

is linear from $C_b(X)$ to $\mathbb{R}$ and

$$|\varphi_\mu(f)| \leq \int |f| \, \mathrm{d}\mu \leq \|f\|_\infty \mu(X).$$

Hence $\varphi_\mu \in C_b(X)'$. Further, $\|\varphi_\mu\| \leq \mu(X)$ and since $\varphi_\mu(\mathbb{1}) = \mu(X) = \|\mathbb{1}\|_\infty \mu(X)$ we have

$$\|\varphi_\mu\| = \mu(X).$$

Moreover, φ_μ is positive.

Conversely, if X is compact, then $C_b(X) = C(X) = \{f : X \to \mathbb{R} : f \text{ is continuous}\}$ and every positive bounded linear functional on $C(X)$ is represented by a finite Borel measure on X. The truth of this statement does not depend on X being a metric space. Therefore we state it in its usual general form, although we have not formally defined Borel sets, Borel measures, $C_b(X)$, etc. for topological spaces that are not metrizable. We denote by $\mathbb{1}$ the function on X that is identically 1.

Theorem 1.16 (Riesz representation theorem). *If (X, d) is a compact Hausdorff space and $\varphi \in C(X)'$ is positive (that is, $\varphi(f) \geq 0$ for every $f \in C(X)$ with $f \geq 0$) and $\varphi(\mathbb{1}) = 1$, then there exists a unique Borel probability measure μ on X such that*

$$\varphi(f) = \int f \, \mathrm{d}\mu \quad \text{for all } f \in C(X).$$

(See [13, Theorem 2.14, p. 40].)

Let us next observe that narrow convergence in $\mathcal{P}(X)$ corresponds to weak* convergence in $C_b(X)'$. The weak* topology on $C_b(X)'$ is the coarsest topology such that the function $\varphi \to \varphi(f)$ on $C_b(X)'$ is continuous for every $f \in C_b(X)'$. A sequence $\varphi_1, \varphi_2, \ldots$ in $C_b(X)'$ *converges weak** to φ in $C_b(X)'$ if and only if

$$\varphi_i(f) \to \varphi(f) \quad \text{as } i \to \infty \text{ for all } f \in C_b(X).$$

If $\mu, \mu_1, \mu_2, \ldots$ are Borel probability measures on X, it is immediately clear that

$$\mu_i \to \mu \text{ narrowly in } \mathcal{P}(X) \quad \Longleftrightarrow \quad \varphi_{\mu_i} \to \varphi_\mu \text{ weak* in } C_b(X)',$$

where, as before, $\varphi_{\mu_i}(f) = \int f \, \mathrm{d}\mu_i$ and $\varphi_\mu(f) = \int f \, \mathrm{d}\mu$, $f \in C_b(X)$, $i \geq 1$.

For the next two theorems see [6, Exercise V.7.17, p. 437] and [14, Theorem 8.13].

Theorem 1.17. *If (X, d) is a metric space, then*

$$C_b(X) \text{ is separable} \quad \Longleftrightarrow \quad X \text{ is compact.}$$

Theorem 1.18. *If X is a separable Banach space, then $\{\varphi \in X' \colon \|\varphi\| \leq 1\}$ is weak* sequentially compact.*

Consequently, if (X,d) is a compact metric space, then the closed unit ball of $C_b(X)'$ is weak* sequentially compact. In combination with the Riesz representation theorem we obtain the following statements for sets of Borel probability measures.

Proposition 1.19. *Let (X,d) be a metric space. If (X,d) is compact, then $(\mathcal{P}(X), d_{\mathrm{BL}})$ is compact, where d_{BL} is the bounded Lipschitz metric induced by d. (Note that any compact metric space is separable.)*

Proof. Assume that (X,d) is compact. Then $C_b(X) = C(X) := \{f : X \to \mathbb{R}\colon f$ is continuous$\}$. The unit ball $B' := \{\varphi \in C_b(X)'\colon \|\varphi\| \leq 1\}$ of $C_b(X)'$ is weak* sequentially compact. As $(\mathcal{P}(X), d_{\mathrm{BL}})$ is a metric space, sequentially compactness is equivalent to compactness. Let $(\mu_n)_n$ be a sequence in $\mathcal{P}(X)$ and let

$$\varphi_n(f) := \int f \,\mathrm{d}\mu_n, \quad n \in \mathbb{N}.$$

Then $\varphi_n \in B'$ for all n. As B' is weak* sequentially compact, hence there exists a $\varphi \in B'$ and a subsequence $(\varphi_{n_k})_k$ such that $\varphi_{n_k} \to \varphi$ in the weak* topology. Then for each $f \in C_b(X)$ with $f \geq 0$,

$$\varphi(f) = \lim_{k\to\infty} \varphi_{n_k}(f) \geq 0,$$

so φ is positive. Further, $\varphi(\mathbb{1}) = \lim_{k\to\infty} \varphi_{n_k}(\mathbb{1}) = 1$. Due to the Riesz representation theorem there exists a $\mu \in \mathcal{P}(X)$ such that $\varphi(f) = \int f \,\mathrm{d}\mu$ for all $f \in C(X) = C_b(X)$. Since $\varphi_{n_k} \to \varphi$ weak*, it follows that $\mu_{n_k} \to \mu$ narrowly. Thus $\mathcal{P}(X)$ is sequentially compact. $\square$

1.6 Prokhorov's theorem

Let (X,d) be a metric space and let $\mathcal{P}(X)$ be the set of Borel probability measures on X. Endow $\mathcal{P}(X)$ with the bounded Lipschitz metric induced by d.

In the study of limit behavior of stochastic processes one often needs to know when a sequence of random variables is convergent in distribution or, at least, has a subsequence that converges in distribution. This comes down to finding a good description of the sequences in $\mathcal{P}(X)$ that have a convergent subsequence or rather of the relatively compact sets of $\mathcal{P}(X)$. Recall that a subset S of a metric space is called *relatively compact* if its closure $\overline{S}$ is compact. The following theorem by Yu.V. Prokhorov [11] gives a useful description of the relatively compact sets of $\mathcal{P}(X)$ in case X is separable and complete. Let us first attach a name to the equivalent condition.

Definition 1.20. A set Γ of Borel probability measures on X is called *tight* if for every $\varepsilon > 0$ there exists a compact subset K of X such that

$$\mu(K) \geq 1 - \varepsilon \quad \text{for all } \mu \in \Gamma.$$

(Also other names and phrases are in use instead of 'Γ is tight': 'Γ is uniformly tight', 'Γ satisfies Prokhorov's condition', 'Γ is uniformly Radon', and maybe more).

Remark. We have shown already: if (X,d) is a complete separable metric space, then $\{\mu\}$ is tight for each $\mu \in \mathcal{P}(X)$ (see Theorem 1.9).

Theorem 1.21 (Prokhorov, 1956). *Let (X,d) be a complete separable metric space and let Γ be a subset of $\mathcal{P}(X)$. Then the following two statements are equivalent:*

(a) $\overline{\Gamma}$ *is compact in* $\mathcal{P}(X)$.

(b) Γ *is tight.*

Let us first remark here that completeness of X is not needed for the implication (b)$\Rightarrow$(a). The proof of the theorem is quite involved. We start with the more straightforward implication (a)$\Rightarrow$(b).

Proof of (a)$\Rightarrow$(b). Claim: If $U_1, U_2, \ldots$ are open sets in X that cover X and if $\varepsilon > 0$, then there exists a $k \geq 1$ such that

$$\mu(\bigcup_{i=1}^{k} U_i) > 1 - \varepsilon \quad \text{for all } \mu \in \Gamma.$$

To prove the claim by contradiction, suppose that for every $k \geq 1$ there is a $\mu_k \in \Gamma$ with $\mu_k(\bigcup_{i=1}^{k} U_i) \leq 1 - \varepsilon$. As $\overline{\Gamma}$ is compact, there is a $\mu \in \overline{\Gamma}$ and a subsequence with $\mu_{k_j} \to \mu$. For any $n \geq 1$, $\bigcup_{i=1}^{n} U_i$ is open, so

$$\begin{aligned} \mu(\bigcup_{i=1}^{n} U_i) &\leq \liminf_{j\to\infty} \mu_{k_j}(\bigcup_{i=1}^{n} U_i) \\ &\leq \liminf_{j\to\infty} \mu_{k_j}(\bigcup_{i=1}^{k_j} U_i) \leq 1 - \varepsilon. \end{aligned}$$

But $\bigcup_{i=1}^{\infty} U_i = X$, so $\mu(\bigcup_{i=1}^{n} U_i) \to \mu(X) = 1$ as $n \to \infty$, which is a contradiction. Thus the claim is proved.

Now let $\varepsilon > 0$ be given. Take $D = \{a_1, a_2, \ldots\}$ dense in X. For every $m \geq 1$ the open balls $B(a_i, 1/m)$, $i = 1, 2, \ldots$, cover X, so by the claim there is a k_m such that

$$\mu\Big(\bigcup_{i=1}^{k_m} B(a_i, 1/m)\Big) > 1 - \varepsilon 2^{-m} \quad \text{for all } \mu \in \Gamma.$$

Take

$$K := \bigcap_{m=1}^{\infty} \bigcup_{i=1}^{k_m} \overline{B}(a_i, 1/m).$$

Then K is closed and for each $\delta > 0$ we can take $m > 1/\delta$ and obtain $K \subset \bigcup_{i=1}^{k_m} B(a_i, \delta)$, so that K is totally bounded. Hence K is compact, since X is complete. Moreover, for each $\mu \in \Gamma$

$$\begin{aligned} \mu(X \setminus K) &= \mu\Big(\bigcup_{m=1}^{\infty} \Big[\bigcup_{i=1}^{k_m} \overline{B}(a_i, 1/m)\Big]^c\Big) \\ &\leq \sum_{m=1}^{\infty} \mu\Big(\Big[\bigcup_{i=1}^{k_m} \overline{B}(a_i, 1/m)\Big]^c\Big) \\ &= \sum_{m=1}^{\infty} \Big(1 - \mu\Big(\bigcup_{i=1}^{k_m} \overline{B}(a_i, 1/m)\Big)\Big) \\ &< \sum_{m=1}^{\infty} \varepsilon 2^{-m} = \varepsilon. \end{aligned}$$

Hence Γ is tight. $\square$

The proof that condition (b) implies (a) is more difficult. We will follow the proof from [10], which is based on compactifications. We have shown already that if X is compact, then $\mathcal{P}(X)$ is compact (see Proposition 1.19). In that case (a) trivially holds. In the cases that we want to consider, X will not always be compact. We can reduce to the compact case by considering a compactification of X.

Lemma 1.22. *If (X,d) is a separable metric space, then there exist a compact metric space (Y,δ) and a map $T : X \to Y$ such that T is a homeomorphism from X onto $T(X)$.*

(T is in general not an isometry. If it were, then X complete $\Rightarrow T(X)$ complete $\Rightarrow T(X) \subset Y$ closed $\Rightarrow T(X)$ compact, which is not true for, e.g., $X = \mathbb{R}$.)

Proof. Let $Y := [0,1]^{\mathbb{N}} = \{(\xi_i)_{i=1}^\infty : \xi_i \in [0,1]\ \forall i\}$ and

$$\delta(\xi,\eta) := \sum_{i=1}^\infty 2^{-i}|\xi_i - \eta_i|, \qquad \xi,\eta \in Y.$$

Then δ is a metric on Y, its topology is the topology of coordinatewise convergence, and (Y,δ) is compact.

Let $D = \{a_1, a_2, \ldots\}$ be dense in X and define

$$\alpha_i(x) := \min\{d(x,a_i),1\}, \qquad x \in X,\ i = 1,2,\ldots.$$

Then for each k, $\alpha_k : X \to [0,1]$ is continuous. For $x \in X$ define

$$T(x) := (\alpha_i(x))_{i=1}^\infty \in Y.$$

Claim: for any $C \subset X$ closed and $x \notin C$ there exist $\varepsilon > 0$ and i such that

$$\alpha_i(x) \le \varepsilon/3, \quad \alpha_i(y) \ge 2\varepsilon/3 \quad \text{for all } y \in C.$$

To prove the claim, take $\varepsilon := \min\{d(x,C),1\} \in (0,1]$. Take i such that $d(a_i,x) < \varepsilon/3$. Then $\alpha_i(x) \le \varepsilon/3$ and for $y \in C$ we have

$$\begin{aligned}
\alpha_i(y) &= \min\{d(y,a_i),1\} \ge \min\{(d(y,x) - d(x,a_i)),1\} \\
&\ge \min\{(d(x,C) - \varepsilon/3),1\} \\
&\ge \min\{2\varepsilon/3,1\} = 2\varepsilon/3.
\end{aligned}$$

In particular, if $x \ne y$ then there exists an i such that $\alpha_i(x) \ne \alpha_i(y)$, so T is injective. Hence $T : X \to T(X)$ is a bijection. It remains to show that for $(x_n)_n$ and x in X:

$$x_n \to x \quad \Longleftrightarrow \quad T(x_n) \to T(x).$$

If $x_n \to x$, then $\alpha_i(x_n) \to \alpha_i(x)$ for all i, so $\delta(T(x_n),T(x)) \to 0$ as $n \to \infty$.

Conversely, suppose that $x_n \not\to x$. Then there is a subsequence such that $x \notin \overline{\{x_{n_1}, x_{n_2}, \ldots\}}$. Then by the claim there is an i such that $\alpha_i(x) \le \varepsilon/3$ and $\alpha_i(x_{n_k}) \ge 2\varepsilon/3$ for all k, so that $\alpha_i(x_{n_k}) \not\to \alpha_i(x)$ as $k \to \infty$ and hence $T(x_{n_k}) \not\to T(x)$. $\square$

We can now complete the proof of Prokhorov's theorem.

Proof of (b)⇒(a). We will show more: *If (X, d) is a separable metric space and $\Gamma \subset \mathcal{P}(X)$ is tight, then $\overline{\Gamma}$ is compact.* Let $\Gamma \subset \mathcal{P}(X)$ be tight. First observe that $\overline{\Gamma}$ is tight as well. Indeed, let $\varepsilon > 0$ and let K be a compact subset of X such that $\mu(K) \geq 1 - \varepsilon$ for all $\mu \in \Gamma$. Then for every $\mu \in \overline{\Gamma}$ there is a sequence $(\mu_n)_n$ in Γ that converges to μ and then we have $\mu(K) \geq \limsup_{n\to\infty} \mu_n(K) \geq 1 - \varepsilon$.

Let $(\mu_n)_n$ be a sequence in $\overline{\Gamma}$. We have to show that it has a convergent subsequence. Let (Y, δ) be a compact metric space and $T : X \to Y$ be such that T is a homeomorphism from X onto $T(X)$. For $B \in \mathcal{B}(Y)$, $T^{-1}(B)$ is Borel in X. Define

$$\nu_n(B) := \mu_n(T^{-1}(B)), \quad B \in \mathcal{B}(Y),\ n = 1, 2, \ldots .$$

Then $\nu \in \mathcal{P}(Y)$ for all n. As Y is a compact metric space, $\mathcal{P}(X)$ is a compact metric space, hence there is a $\nu \in \mathcal{P}(Y)$ and a subsequence such that $\nu_{n_k} \to \nu$ in $\mathcal{P}(Y)$. We want to translate ν back to a measure on X. Set $Y_0 := T(X)$.

Claim: ν is concentrated on Y_0 in the sense that there exists a set $E \in \mathcal{B}(Y)$ with $E \subset Y_0$ and $\nu(E) = 1$.

If we assume the claim, define

$$\nu_0(A) := \nu(A \cap E), \quad A \in \mathcal{B}(Y_0).$$

(Note: $A \in \mathcal{B}(Y_0) \Rightarrow A \cap E$ Borel in $E \Rightarrow A \cap E$ Borel in Y, since E is a Borel subset of Y.) The measure ν_0 is a finite Borel measure on Y_0 and $\nu_0(E) = \nu(E) = 1$. Now we can translate ν_0 back to

$$\mu(A) := \nu_0(T(A)) = \nu_0((T^{-1})^{-1}(A)), \quad A \in \mathcal{B}(X).$$

Then $\mu \in \mathcal{P}(X)$. We want to show that $\mu_{n_k} \to \mu$ in $\mathcal{P}(X)$. Let C be closed in X. Then $T(C)$ is closed in $T(X) = Y_0$. ($T(C)$ need not be closed in Y.) Therefore there exists $Z \subset Y$ closed with $Z \cap Y_0 = T(C)$. Then $C = \{x \in X : T(x) \in T(C)\} = \{x \in X : T(x) \in Z\} = T^{-1}(Z)$, because there are no points in $T(C)$ outside Y_0, and $Z \cap E = T(C) \cap E$. Hence

$$\begin{aligned}
\limsup_{k\to\infty} \mu_{n_k}(C) &= \limsup_{k\to\infty} \nu_{n_k}(Z) \\
&\leq \nu(Z) \\
&= \nu(Z \cap E) + \nu(Z \cap E^c) = \nu(T(C) \cap E) + 0 \\
&= \nu_0(T(C)) = \mu(C).
\end{aligned}$$

So $\mu_{n_k} \to \mu$.

Finally, to prove the claim we use tightness of $\overline{\Gamma}$. For each $m \geq 1$ take K_m compact in X such that $\mu(K_m) \geq 1 - 1/m$ for all $\mu \in \Gamma$. Then $T(K_m)$ is a compact subset of Y hence closed in Y, so

$$\begin{aligned}
\nu(T(K_m)) &\geq \limsup_{k\to\infty} \nu_{n_k}(T(K_m)) \\
&\geq \limsup_{k\to\infty} \mu_{n_k}(K_m) \geq 1 - 1/m.
\end{aligned}$$

Take $E := \bigcup_{m=1}^{\infty} K_m$. Then $E \in \mathcal{B}(Y)$ and $\nu(E) \geq \nu(K_m)$ for all m, so $\nu(E) = 1$. □

Example. Let $X = \mathbb{R}$, $\mu_n(A) := n^{-1}\lambda(A \cap [0,n])$, $A \in \mathcal{B}(\mathbb{R})$. Here λ denotes Lebesgue measure on $\mathbb{R}$. Then $\mu_n \in \mathcal{P}(\mathbb{R})$ for all n. The sequence $(\mu_n)_n$ has no convergent subsequence. Indeed, suppose $\mu_{n_k} \to \mu$, then

$$\begin{aligned}\mu((-N,N)) &\leq \liminf_{n\to\infty} \mu_n((-N,N)) \\ &= \liminf_{n\to\infty} n^{-1}\lambda([0,N]) = \liminf_{n\to\infty} N/n = 0,\end{aligned}$$

so $\mu(\mathbb{R}) = \sup_{N\geq 1} \mu((-N,N)) = 0$. There is leaking mass to infinity; the set $\{\mu_n : n = 1,2,\ldots\}$ is not tight.

1.7 Disintegration

Let (X,d_X) and (Y,d_Y) be separable complete metric spaces. On the Cartesian product $Z = X \times Y$ we define the metric $d_Z((x,y),(x',y')) := d_X(x,x') + d_Y(y,y')$. Then (Z,d_Z) is a separable complete metric space. There are two natural σ-algebras in Z related to Borel sets: the Borel σ-algebra $\mathcal{B}_Z$ of Z, which is generated by all open sets of Z, and the σ-algebra $\mathcal{B}_X \otimes \mathcal{B}_Y$ generated by the rectangles $A \times B$ with $A \in \mathcal{B}_X$ and $B \in \mathcal{B}_Y$. The collections $\{A\colon A \times Y \in \mathcal{B}_Z\}$ and $\{B\colon X \times B \in \mathcal{B}_Z\}$ are σ-algebras containing the open sets in X and Y, respectively. Therefore, for every Borel sets A in X and B in Y, the set $A \times B = A \times Y \cap X \times B$ is in $\mathcal{B}_Z$. Hence $\mathcal{B}_X \otimes \mathcal{B}_Y \subseteq \mathcal{B}_Z$. The seperability of Z yields that $\mathcal{B}_X \otimes \mathcal{B}_Y = \mathcal{B}_Z$. Indeed, let D be a dense subset of Z and let U be any open set of Z. For every $z \in U$ we can find an $(a_z,b_z) \in D$ and positive rational numbers ε and δ such that the rectangle $R_z := B(a_z,\varepsilon) \times B(b_z,\delta)$ contains z and is contained in U. Then U is the union of R_z ($z \in U$) and as there are at most countably many of such R_z, we obtain $U \in \mathcal{B}_X \otimes \mathcal{B}_Y$.

If $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, then $\eta = \mu \otimes \nu$ is the unique measure η on $\mathcal{B}_X \otimes \mathcal{B}_Y$ such that

$$\eta(A \times B) = \mu(A)\nu(B), \quad \text{for all } A \in \mathcal{B}_X,\ B \in \mathcal{B}_Y.$$

Clearly, $\eta(Z) = 1$, so $\eta \in \mathcal{P}(Z)$. Further, by Fubini,

$$\int_{X\times Y} f(x,y)\,\mathrm{d}\eta(x,y) = \int_X \left(\int_Y f(x,y)\,\mathrm{d}\nu(y)\right)\mathrm{d}\mu(x),$$

for all Borel measurable functions $f : Z \to [0,\infty]$.

We are looking for such a statement for more general measures η on the product space Z. Given a measure on the product space Z, can we write an integral with respect to that measure as repeated integration, first with respect to the y-variable and then with respect to x? For $\eta \in \mathcal{P}(Z)$, we call the measures

$$\mu(A) := \eta(A \times Y) \quad \text{and} \quad \nu(B) := \eta(X \times B), \quad A \in \mathcal{B}_X,\ B \in \mathcal{B}_Y,$$

the *marginals* of η. Suppose that η is absolutely continuous with respect to $\mu \otimes \nu$, that is, $(\mu \otimes \nu)(C) = 0$ implies $\eta(C) = C$ for every Borel set C in Z. Then the Radon-Nikodym theorem says that there exists a Borel measurable function $h : Z \to [0,\infty)$ such that

$$\int_Z f(x,y)\,\mathrm{d}\eta(x,y) = \int_Z f(x,y)h(x,y)\,\mathrm{d}(\mu\otimes\nu)(x,y)$$

and therefore

$$\int_Z f(x,y)\,\mathrm{d}\eta(x,y) = \int_X \left(\int_Y f(x,y)h(x,y)\,\mathrm{d}\nu(y) \right) \mathrm{d}\mu(x)$$

for every positive Borel measurable function f. If we let

$$\nu_x(B) := \int_B h(x,y)d\nu(y) = \int_Y \mathbb{1}_{X\times B} h(x,y)\,\mathrm{d}\nu(y), \quad x \in X,$$

then ν_x is a Borel measure on Y for each x and

$$\int_{X\times Y} f(x,y)\,\mathrm{d}\eta(x,y) = \int_X \left(\int_Y f(x,y)\,\mathrm{d}\nu_x(y) \right) \mathrm{d}\mu(x)$$

for every Borel function $f : X \times Y \to [0,\infty]$. The latter formula is called the *disintegration formula.* If there would exist a Borel set A in X with $\mu(A) > 0$ such that $\nu_x(Y) > 1$ for all $x \in A$ or $\nu_x(Y) < 1$ for all $x \in A$, then

$$\mu(A) = \int \mathbb{1}_{A\times Y}\,\mathrm{d}\eta = \int_X \left(\int_Y \mathbb{1}_{A\times Y}(x,y)\,\mathrm{d}\nu_x(y) \right) \mathrm{d}\mu(x) > \mu(A)$$

or $\mu(A) < \mu(A)$, which is impossible. Hence ν_x is a probability measure for μ-almost all x.

The existence of such a family of measures $(\nu_x)_{x\in X}$ is not restricted to the case that η is absolutely continuous with respect to the product measure of its marginals.

Theorem 1.23. *Let (X, d_X) and (Y, d_Y) be two separable complete metric spaces. Let $\eta \in \mathcal{P}(X \times Y)$ and $\mu(A) = \eta(A \times Y)$ for $A \subseteq X$ Borel. Then for every $x \in X$ there exists a $\nu_x \in \mathcal{P}(Y)$ such that*

(i) $x \to \nu_x(B) : X \to \mathbb{R}$ is $\mathcal{B}_X$-measurable for every $B \in \mathcal{B}_Y$, and

(ii) $\int_{X\times Y} f(x,y)\,\mathrm{d}\eta(x,y) = \int_X \left(\int_Y f(x,y)\,\mathrm{d}\nu_x(y) \right) \mathrm{d}\mu(x)$ for every Borel measurable $f : X \times Y \to [0,\infty]$.

The above disintegration theorem for product spaces is a special case of the next theorem. The set Z plays the role of the product $X \times Y$ and the map π the role of the coordinate projection on X. As we do not have the second coordinate space Y anymore, the measures ν_x will be measures on the whole space Z, but concentrated on $\pi^{-1}(\{x\})$.

Theorem 1.24. *Let (Z, d_Z) and (X, d_X) be separable complete metric spaces, let $\pi : Z \to X$ be a Borel map, let $\eta \in \mathcal{P}(Z)$, and let $\mu(A) := \eta(\pi^{-1}(A))$, $A \subseteq X$ Borel. Then for every $x \in X$ there exists a $\nu_x \in \mathcal{P}(Z)$ such that*

(i) ν_x is concentrated on $\pi^{-1}(\{x\})$, that is, $\nu_x(Z \setminus \pi^{-1}(\{x\})) = 0$ for μ-almost every $x \in X$,

(ii) $x \mapsto \nu_x(C) : X \to \mathbb{R}$ is Borel measurable for every Borel $C \subseteq Z$, and

(iii) $\int_Z f(z)\,\mathrm{d}\eta(z) = \int_X \left(\int_{\pi^{-1}(\{x\})} f(z)\,\mathrm{d}\nu_x(z) \right) \mathrm{d}\mu(x)$.

A proof can be found in [5, Section 10.2, p. 341–351]

1.8 Borel probability measures with respect to weak topologies

We wish to apply the theory of probability measures on metric spaces to measures on Hilbert spaces endowed with their weak topology. However, the weak topology is in general not metrizable. We will define a metric that induces the weak topology on norm bounded sets and compare the induced Borel σ-algebras and narrow convergences. The metric will even be given by an inner product. We will need several theorems concerning weak topologies on Banach spaces. We begin by recalling the definitions of weak topology and weak* topology.

Let X be a Banach space and let X' be its Banach dual space. The *weak topology* of X is the smallest (coarsest) topology on X such that every $\varphi \in X'$ is continuous with respect to this topology. Then

$$x_n \to x \text{ weakly} \quad \Longleftrightarrow \quad \varphi(x_n) \to \varphi(x) \text{ for all } \varphi \in X'.$$

The *weak* topology* of X' is the smallest topology on X' such that every map $\varphi \to \varphi(x)$ is continuous for all $x \in X$. Then

$$\varphi_n \to \varphi \text{ weak*} \quad \Longleftrightarrow \quad \varphi_n(x) \to \varphi(x) \text{ for all } x \in X.$$

The weak and weak* topologies are weaker than the norm topologies of X and X', respectively.

Theorem 1.25. *Let X be a Banach space. The weak* topology on $\{\varphi \in X' \colon \|\varphi\| \le 1\}$ is a metric topology (that is, induced by a metric) if and only if X is separable.*
(See [6, Theorem V.5.1, p. 426].)

Theorem 1.26. *Let X be a Banach space. Then the weak topology on $\{x \in X \colon \|x\| \le 1\}$ is a metric topology if and only if X' is separable.*
(See [6, Theorem V.5.2, p. 426].)

Corollary 1.27. *If X is a separable Hilbert space, then the weak topology on $\{x \in X \colon \|x\| \le 1\}$ is a metric topology.*

Corollary 1.28. *If X is a separable Hilbert space and $S \subseteq X$ is bounded, then*

(i) S is weakly closed $\Longleftrightarrow$ S is weakly sequentially closed;

(ii) S is weakly compact $\Longleftrightarrow$ S is weakly sequentially compact.

Theorem 1.29. *Let X be a Banach space. Every weakly convergent sequence is bounded.*
(See [14, Lemma 8.15, p. 190].)

Theorem 1.30. *If X is a Hilbert space, then every bounded sequence in X has a weakly convergent subsequence.*
(See [14, Theorem 8.16, p. 191].)

Let $(X, \langle,\rangle)$ be a separable Hilbert space and denote its norm by $|x| = \langle x, x\rangle^{1/2}$. Fix an orthonormal basis $(e_n)_n$ in X. Define

$$\langle x, y\rangle_\varpi := \sum_{n=1}^{\infty} \frac{1}{n^2} \langle x, e_n\rangle\langle e_n, y\rangle, \quad x, y \in X, \tag{3}$$

and $|x|_\varpi := \langle x, \rangle_\varpi^{1/2}$. Then $\langle,\rangle_\varpi$ is an inner product on X. It does definitely depend on the basis as is for instance seen by exchanging e_1 and e_2. We will show that $\langle,\rangle_\varpi$ induces the weak topology on bounded subsets of $(X, \langle,\rangle)$.

Proposition 1.31. *Let S be a bounded subset of $(X, \langle,\rangle)$ and let $(x_i)_i$ be a sequence in S and $x \in S$. Then the following three statements are equivalent:*

(a) $x_i \to x$ weakly, that is, $\langle x_i, y\rangle \to \langle x, y\rangle$ for all $y \in H$;

(b) $\langle x_i, e_n\rangle \to \langle x, e_n\rangle$ for all n;

(c) $|x_i - x|_\varpi \to 0$.

Proof. (a)⇒(c): For every n, $\langle x_i - x, e_n\rangle \to 0$ and $|\langle x_i - x, e_n\rangle| \leq |x_i - x| \leq M$ for all i, as S is bounded. Hence by Lebesgue's dominated convergence theorem,

$$|x_i - x|_\varpi = \sum_{n=1}^{\infty} \frac{1}{n^2} |\langle x_i - x, e_n\rangle|^2 \to 0.$$

(c)⇒(b): $|\langle x_i - x, e_m\rangle|^2 \leq m^2 \sum_{n=1}^{\infty} \frac{1}{n^2} |\langle x_i - x, e_n\rangle|^2 \to 0$ as $i \to \infty$.
(b)⇒(a): Let $y \in H$, $|x_i| \leq M$ for all i, $|x| \leq M$, $\varepsilon > 0$. Take $N \in \mathbb{N}$ such that

$$|y - \sum_{n=1}^{N} \langle y, e_n\rangle e_n| < \varepsilon/(4M).$$

Then

$$|\langle x_i - x, y\rangle| \leq |\langle x_i - x, \sum_{n=1}^{N} \langle y, e_n\rangle e_n\rangle| + |x_i - x| \; |y - \sum_{n=1}^{N} \langle y, e_n\rangle e_n| < \varepsilon$$

for i large. □

Corollary 1.32. *On a bounded subset of $(X, \langle,\rangle)$, the weak topology and the topology of $\langle,\rangle_\varpi$ coincide.*

Proof. Both topologies are metric topologies on bounded sets and the convergence of sequences coincides according to the previous proposition. □

Let

$\mathcal{B}(X)$ = Borel σ-algebra of $(X, \langle,\rangle)$
$\mathcal{B}(X, w)$ = Borel σ-algebra of X with respect to the weak topology
$\mathcal{B}(X, \varpi)$ = Borel σ-algebra of $(X, \langle,\rangle_\varpi)$

Proposition 1.33. *The three Borel σ-algebras coincide: $\mathcal{B}(X) = \mathcal{B}(X, w) = \mathcal{B}(X, \varpi)$.*

Proof. For $R > 0$, the ball $B_R := \{x \in X \colon |x| \leq R\}$ is weakly sequentially compact, hence weakly compact (since the weak topology is a metric topology on bounded sets), hence weakly closed.

• $\mathcal{B}(X, w) \subseteq \mathcal{B}(X, \varpi)$: If $S \subseteq X$ is weakly closed, then $S \cap B_R$ is weakly closed (in B_R) and B_R is bounded, so $S \cap B_R$ is ϖ-closed in B_R. Also,

$$\begin{aligned} B_R &= \{x \in X \colon |x| \leq R\} = \{x \colon \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2 \leq R\} \\ &= \bigcap_{N=1}^{\infty} \{x \colon \sum_{n=1}^{N} |\langle x, e_n\rangle|^2 \leq R\} \end{aligned}$$

is ϖ-closed, since $x \mapsto \sum_{n=1}^N |\langle x, e_n\rangle|^2$ is ϖ-continuous for all N. Hence $S \cap B_R$ is ϖ-closed. Then $S = \cup_{m=1}^\infty S \cap B_m \in \mathcal{B}(X, \varpi)$.

• $\mathcal{B}(X) \subseteq \mathcal{B}(X, w)$: If $a \in X$ and $R > 0$, then $\{x \in X \colon |x - a| \le R\}$ is weakly compact hence in $\mathcal{B}(X, w)$. As X is separable, $\mathcal{B}(X)$ is the σ-algebra generated by the closed (or open) balls of X, so $\mathcal{B}(X) \subseteq \mathcal{B}(X, w)$.

• $\mathcal{B}(X, \varpi) \subseteq \mathcal{B}(X)$: Let $S \subseteq X$ be ϖ-closed. Then $S \cap B_n$ is ϖ-closed in B_n, hence weakly closed in B_n. As B_n is weakly closed, $S \cap B_n$ is weakly closed in X, hence closed in $(X, \langle , \rangle)$. Thus, $S = \bigcup_n S \cap B_n \in \mathcal{B}(X)$. □

It follows that $\mathcal{P}(X) = \mathcal{P}(X, w) = \mathcal{P}(X, \varpi)$ as sets. The narrow convergences, however, are different. Under additional assumptions, some relations between the narrow convergences can be proved. We include two such results and sketches of their proofs. First we need to introduce cylindrical functions.

Definition 1.34.

$$\begin{aligned} C_c^\infty(\mathbb{R}^d) \quad := \quad & \{\varphi : \mathbb{R}^d \to \mathbb{R} \colon \text{every derivative of order } k \text{ exists for all } k \\ & \text{and } \varphi = 0 \text{ outside a compact set}\}. \end{aligned}$$

Recall that X is a separable Hilbert space with a fixed orthonormal basis $(e_n)_n$.

$$\begin{aligned} \mathrm{Cyl}(X) \quad := \quad & \{f : X \to \mathbb{R} \colon \exists d \in \mathbb{N},\ \exists \varphi \in C_c^\infty(\mathbb{R}^d) \text{ such that} \\ & f(x) = \varphi(\langle x, e_1\rangle, \ldots, \langle x, e_d\rangle)\ \forall x \in X\}. \end{aligned}$$

The elements of $\mathrm{Cyl}(X)$ are called *smooth cylindrical functions* on X.

Proposition 1.35. *Every $f \in \mathrm{Cyl}(X)$ is Lipschitz, everywhere differentiable in Fréchet sense, and continuous with respect to the weak topology of X as well as $\langle , \rangle_\varpi$ (with the same fixed basis $(e_n)_n$).*

Proof. Let $f(x) = \varphi(\langle x, e_1\rangle, \ldots, \langle x, e_d\rangle)$. Let $g(t) := f(x + t(x - y))$. Then

$$|f(x) - f(y)| = |g(0) - g(1)| \le \sup_{t \in [0,1]} |g'(t)|$$

and

$$\begin{aligned} |g'(t)| \quad &= \quad |\sum_{i=1}^d (D_i\varphi)(\langle x + t(y - x), e_i\rangle)\langle y - x, e_i\rangle| \\ &\le \quad \sum_{i=1}^d \|D_i\varphi\|_\infty |y - x| \le M|y - x|, \end{aligned}$$

so f is Lipschitz. To see that f is Fréchet differentiable at x,

$$\begin{aligned} &|f(x + h) - f(x) - \sum_{i=1}^d D_i\varphi(\langle x, e_1\rangle, \ldots, \langle x, e_d\rangle)\langle h, e_i\rangle|_{\mathbb{R}} \\ &\le o(\langle h, e_1\rangle, \ldots, \langle h, e_d\rangle) \le o(|h|). \end{aligned}$$

So f is Fréchet differentiable. The continuities are clear. □

Lemma 1.36. *Let X be a separable Hilbert space, $(e_n)_n$ an orthonormal basis of X, and $\langle , \rangle_\varpi$ defined by (3). Then:*

(1) *if K is weakly compact in X, then K is compact with respect to $\langle , \rangle_\varpi$;*

(2) *if $\Gamma \subseteq \mathcal{P}(X)$ is* weakly tight *in the sense that*

$$\forall \varepsilon > 0 \; \exists R_\varepsilon > 0 \text{ such that } \mu(B_{R_\varepsilon}) \geq 1 - \varepsilon \; \forall \mu \in \Gamma,$$

then Γ is tight in $\mathcal{P}(X, \varpi)$ (here $B_R = \{x \in X : |x| \leq R\}$);

(3) *let $(\mu_n)_n \subseteq \mathcal{P}(X)$ be a sequence such that $\{\mu_n : n \in \mathbb{N}\}$ is weakly tight; then μ_n converges narrowly to μ in $\mathcal{P}(X, \varpi)$ if and only if*

$$\lim_{n\to\infty} \int f(x)\, \mathrm{d}\mu_n(x) = \int f(x)\, \mathrm{d}\mu(x) \text{ for all } f \in \mathrm{Cyl}(X).$$

Proof. (1): K is weakly compact implies that K is bounded in $(X, \langle , \rangle)$. Hence the weak topology on K is the same as the ϖ-topology, so K is also ϖ-compact.

(2): B_R is bounded in $(X, \langle , \rangle)$ and weakly compact, hence ϖ-compact.

(3): $\Rightarrow$: Suppose that $\{\mu_n : n\}$ is weakly tight and $\mu_n \to \mu$ narrowly in $\mathcal{P}(X, \varpi)$. Let $f \in \mathrm{Cyl}(X)$. Then f is bounded and f is continuous with respect to ϖ, so $\int f\, \mathrm{d}\mu_n \to \int f\, \mathrm{d}\mu$.

$\Leftarrow$: Suppose that $\{\mu_n : n\}$ is weakly tight and $\int f\, \mathrm{d}\mu_n \to \int f\, \mathrm{d}\mu$ for all $f \in \mathrm{Cyl}(X)$. By (2) and Prokhorov's theorem, every subsequence of (μ_n) has a subsubsequence that converges narrowly in $\mathcal{P}(X, \varpi)$ to some measure in $\mathcal{P}(X)$. If all these limit measures are equal, then $(\mu_n)_n$ converges to this measure. We check that $\mu_{n_k} \to \nu$ narrowly in $\mathcal{P}(X, \varpi)$ implies that $\nu = \mu$. As $f \in \mathrm{Cyl}(X)$ is bounded and ϖ-continuous, we have $\int f\, \mathrm{d}\mu_{n_k} \to \int f\, \mathrm{d}\nu$, so $\int f\, \mathrm{d}\mu = \int f\, \mathrm{d}\nu$ for all $f \in \mathrm{Cyl}(X)$. If $f(x) = \varphi(\langle x, e_1\rangle, \ldots, \langle x, e_d\rangle)$, with $\varphi \in C_c^\infty(\mathbb{R}^d)$ then the equality reads

$$\int_x \varphi(\langle x, e_1\rangle, \ldots, \langle x, e_d\rangle)\, \mathrm{d}\mu(x) = \int_X \varphi(\langle x, e_1\rangle, \ldots, \langle x, e_d\rangle)\, \mathrm{d}\nu(x).$$

By means of a Stone-Weierstrass approximation argument, it follows that

$$\int_X \psi(\langle x, e_1\rangle, \ldots, \langle x, e_d\rangle)\, \mathrm{d}\mu(x) = \int_X \psi(\langle x, e_1\rangle, \ldots, \langle x, e_d\rangle)\, \mathrm{d}\nu(x)$$

for all $\psi \in C_c(\mathbb{R}^d)$. The latter equality actually holds for all $\psi \in C_b(X)$, as can be seen with the aid of Lebesgue's dominated convergence theorem. Now let $g \in C_b(X)$ and define

$$g_d(\langle x, e_1\rangle, \ldots, \langle x, e_d\rangle) = g(\sum_{i=1}^d \langle x, e_i\rangle e_i)$$

for $x \in X$ and $d \in \mathbb{N}$. Then $g_d(x) \to g(x)$ as $d \to \infty$ for all x. By Lebesgue's dominated convergence theorem,

$$\int_X g(x)\, \mathrm{d}\mu(x) = \int_X g(x)\, \mathrm{d}\nu(x),$$

so $\mu = \nu$. □

Theorem 1.37. *Let j: $[0,\infty) \to [0,\infty)$ be continuous, strictly increasing and surjective. (For instance, $j(x) = x^p$, $1 \le p < \infty$.) Let $\mu_n, \mu \in \mathcal{P}(X)$ such that $\mu_n \to \mu$ and in $\mathcal{P}(X, \varpi)$ and*

$$\lim_{n\to\infty} \int_X j(|x|)\,\mathrm{d}\mu_n(x) \to \int_X j(|x|)\,\mathrm{d}\mu < \infty.$$

Then $\mu_n \to \mu$ in $\mathcal{P}(X, \langle,\rangle)$.

Proof. Claim: For every $\varepsilon > 0$ there exists an $R > 0$ such that $\mu_N(\overline{B}_R) \ge 1 - \varepsilon$ for all n. Suppose not: for every $R > 0$ there exists an n such that $\mu_n(\overline{B}^c_R) > \varepsilon$. Then

$$\begin{aligned} \int_X j(|x|)\,\mathrm{d}\mu_n(x) &\ge \int_{\overline{B}^c_R} j(|x|)\,\mathrm{d}\mu_n(x) \\ &\ge j(R)\mu_n(\overline{B}^c_R) \ge j(R)\varepsilon, \end{aligned}$$

which becomes arbitrary large as $R \to \infty$. Thus we have a contradiction with finiteness of $\sup_n \int_X j(|x|)\,\mathrm{d}\mu_n(x)$.

Claim: The map $x \mapsto j(|x|)$ is ϖ-l.s.c. Let $(x_i)_i$ and x in X be such that $|x_i - x|_\varpi \to 0$. As $x \mapsto j(|x|)$ is bounded below, $\alpha := \liminf_{i\to\infty} j(|x_i|)$ exists. For $\varepsilon > 0$, the set $L_{\alpha+\varepsilon} := \{y\colon j(|y|) \le \alpha + \varepsilon\} = \{y\colon |y| \le j^{-1}(\alpha+\varepsilon)\}$ is weakly compact hence weakly closed (since bounded) hence ϖ-closed. Therefore $x \in L_{\alpha+\varepsilon}$ and hence $j(|x|) \le \alpha = \liminf_{i\to\infty} j(|x_i|)$. Thus $j(|\cdot|)$ is ϖ-l.s.c.

Let

$$\begin{aligned} \mathcal{H} \;:=\; \Big\{ & h\colon X \to \mathbb{R}\colon \exists A, B \ge 0 \text{ such that } |h(x)| \le A + Bj(|x|)\ \forall x \in X \\ & \text{and } \int h\,\mathrm{d}\mu_n \to \int h\,\mathrm{d}\mu \Big\}. \end{aligned}$$

We will show that $C_b(X) \subseteq \mathcal{H}$.

Claim: $\mathcal{H}$ is a vector space, $j(|\cdot|) \in \mathcal{H}$, and $\mathbb{1} \in \mathcal{H}$.

Claim: If $\|h_n - h\|_\infty \to 0$ and $h_n \in \mathcal{H}$ for all n, then $h \in \mathcal{H}$. Each h_n is continuous and there are A_n and B_n such that $|h_n(x)| \le A_n + B_n j(|x|)$ for all x, so h is continuous and $|h(x)| \le |h(x) - h_n(x)| + |h_n(x)| \le (A_n + 1) + B_n j(|x|)$ for an n with $|h_n - h|_\infty \le 1$. Moreover,

$$\begin{aligned} |\int h\,\mathrm{d}\mu_n - \int h\,\mathrm{d}\mu| &\le |\int (h_m - h)\,\mathrm{d}\mu_n| + |\int h_m\,\mathrm{d}\mu_n - \int h_m\,\mathrm{d}\mu| + |\int h_m\,\mathrm{d}\mu - \int h\,\mathrm{d}\mu| \\ &\le \|h_m - h\|_\infty + |\int h_m\,\mathrm{d}\mu_n - \int h_m\,\mathrm{d}\mu| + \|h_m - h\|_\infty \to 0 \end{aligned}$$

if first $n \to \infty$ and then $m \to \infty$.

Let

$$\mathcal{A} := \{h \in \mathcal{H}\colon h \text{ is } \varpi\text{-l.s.c.}\}.$$

Claim: $f, g \in \mathcal{A}$ implies $f + g \in \mathcal{A}$, $f \in \mathcal{A}$ and $\lambda \ge 0$ implies $\lambda f \in \mathcal{A}$, and $j(|\cdot|) \in \mathcal{A}$.

Claim: If f, g: $X \to \mathbb{R}$ are continuous, ϖ-l.s.c., there exist $A, B > 0$ such that $|f(x)| \vee |g(x)| \le A + Bj(|x|)$ for all x, and $f + g \in \mathcal{A}$, then both $f \in \mathcal{A}$ and $g \in \mathcal{A}$. We have to show that $\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu$ (and the same for g). By ϖ-lower semicontinuity, $\int f\,\mathrm{d}\mu \le \liminf_n \int f\,\mathrm{d}\mu_n$ and $\int g\,\mathrm{d}\mu \le \liminf_n \int g\,\mathrm{d}\mu_n$, so $\int (f+g)\,\mathrm{d}\mu \le \liminf \int f\,\mathrm{d}\mu_n +$

$\liminf \int g \,\mathrm{d}\mu_n \le \limsup \int f \,\mathrm{d}\mu_n + \liminf \int g \,\mathrm{d}\mu_n \le \lim \int (f+g) \,\mathrm{d}\mu_n = \int (f+g)\,\mathrm{d}\mu$, so $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu$.

Claim: $f, g \in \mathcal{A}$ implies $f \vee g \in \mathcal{A}$ and $f \wedge g \in \mathcal{A}$. This follows from the previous claim and the identity $f + g = f \vee g + f \wedge g$.

Recall the Moreau-Yosida approximations: if $f \in C_b(X)$ and

$$f_k(x) := \inf_{y \in X} \Big(f(y) + k|x-y| \Big),$$

then f_k is Lipschitz, $\inf f \le f_1 \le f_2 \le \cdots \le f$, and $f(x) = \lim_{k\to\infty} f_k(x)$. Let D be a countable dense subset of X. Then $f_k(x) = \inf_{y \in D} \Big(f(y) + k|x-y| \Big)$, so there exists a sequence y_i such that $f_k(x) = \inf_i \Big(f(y_i) + k|x-y_i| \Big)$. Let

$$\mathcal{D}_0 := \{x \mapsto \bigwedge_{i=1}^{m} \Big(\alpha_i + \beta_i |x-y| \Big) \wedge \gamma_i \colon m \in \mathbb{N},\ \alpha_i \in \mathbb{R},\ \beta_i, \gamma_i \ge 0,\ y \in X\}$$

and

$$\mathcal{D} := \mathcal{D}_0 - \mathcal{D}_0 = \{f - g \colon f, g \in \mathcal{D}_0\}.$$

Claim: For every bounded Lipschitz function $h \colon X \to \mathbb{R}$ we have

$$\begin{aligned} \int_X h \,\mathrm{d}\mu &= \sup\{\int f \,\mathrm{d}\mu \colon f \in \mathcal{D},\ f \le h\} \\ &= \inf\{\int f \,\mathrm{d}\mu \colon f \in \mathcal{D},\ f \ge h\}. \end{aligned}$$

Choose a countable dense subset $D = \{a_1, a_2, \ldots\}$ of X. Let L and M be such that $|h(x)| \le M$ for all x and $|h(x) - h(y)| \le L|x-y|$ for all $x, y \in X$. For $n \in \mathbb{N}$, choose N_n such that

$$\mu\Big(\bigcup_{i=1}^{N_n} B(a_i, 1/n) \Big) \ge 1 - 1/(Mn)$$

and

$$f_n(x) = \bigwedge_{i=1}^{N_n} \Big(h(a_i) + L|x-a_i| \Big) \wedge M, \quad x \in X.$$

Then $f_n \in \mathcal{D}$ and for $x \in B(a_i, 1/n)$,

$$|f_n(x) - h(x)| \le |f_n(x) - f(a_i)| + |h(a_i) - h(x)| \le 2L/n.$$

For $x \in X$,

$$h(a_i) + L|x-a_i| \ge h(a_i) + |h(x) - h(a_i)| \ge h(x),$$

so $f_n \ge h$. Further, $f_n(x) \le M$ for all x. By Lebesgue,

$$\begin{aligned} \int h \,\mathrm{d}\mu &\le \int_{\bigcup_{i=1}^{N_n} B(a_i,1/n)} f_n \,\mathrm{d}\mu + M\mu((\bigcup B(a_i,1/n))^c) \\ &\le \int (h + 2L/n)\,\mathrm{d}\mu + 1/n \le \int h \,\mathrm{d}\mu + (2L+1)/n. \end{aligned}$$

Now do the same for $-h$.

Claim: If $f = x \mapsto (\alpha + \beta|x-y|) \wedge \gamma \in \mathcal{A}$ for every $\alpha \in \mathbb{R}$, $\beta, \gamma > 0$, and $y \in X$, then $\mu_n \to \mu$ in $\mathcal{P}(X, |\cdot|)$. If such functions f are in $\mathcal{A}$, then $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu$ and they are also in $\mathcal{D}$. Hence for bounded Lipschitz $f\colon X \to \mathbb{R}$,

$$\begin{aligned} \liminf_{n\to\infty} \int h \,\mathrm{d}\mu_n &\geq \sup_{f\in\mathcal{D},\, f\leq h} \liminf_{n\to\infty} \int f \,\mathrm{d}\mu_n \\ &= \sup_{f\in\mathcal{D},\, f\leq h} \int f \,\mathrm{d}\mu = \int h \,\mathrm{d}\mu. \end{aligned}$$

Similarly for $-h$.

It remains to show that $x \mapsto f(x)(\alpha + \beta|x-y|) \wedge \gamma \in \mathcal{A}$. We can rewrite such an f as $f(x) = \Big(\beta|x-y|\Big) \wedge (\gamma - \alpha) + \alpha$. If $\gamma - \alpha < 0$, then this function is constant, hence in $\mathcal{A}$. Thus we may assume that $f(x) = |x-y| \wedge \gamma$ for some $y \in X$ and $\gamma \geq 0$.

Claim: If $f \in \mathcal{A}$, $\theta\colon \mathbb{R} \to \mathbb{R}$ is uniformly continuous, bounded, and increasing, then $\theta \circ f \in \mathcal{A}$. We may assume (by uniform approximation) that θ is Lipschitz, increasing, bounded, and (by scaling) has Lipschitz constant ≤ 1. Then also $x \mapsto x - \theta(x)$ is Lipschitz and increasing, so $\theta \circ f$ and $f - \theta \circ f$ are ϖ-l.s.c. Their sum is in $\mathcal{A}$, hence $\theta \circ f \in \mathcal{A}$ and $f - \theta \circ f \in \mathcal{A}$.

Claim: $x \mapsto |x| \wedge \gamma \in \mathcal{A}$. Consider $\theta(s) = j^{-1}(s)^2 \wedge \gamma^2$ if $s \geq 0$ and $\theta(s) = 0$ if $s < 0$. Then $\theta \circ j(|\cdot|) \in \mathcal{A}$, so $x \mapsto j^{-1}(j(|x|))^2 \wedge \gamma^2 = |x| \wedge \gamma^2 \in \mathcal{A}$. With $\theta(s) = \sqrt{s}$ if $s \geq 0$ and $\theta(s) = 0$ if $s < 0$, it follows that $x \mapsto |x| \wedge \gamma \in \mathcal{A}$.

Claim: $x \mapsto |x-y| \wedge \gamma \in \mathcal{A}$. (Sketch of proof:) The function $g_{\ell,m}(x) = (-\ell) \vee (-2\langle x, y\rangle + |y|^2) \wedge m \in \mathcal{A}$, $\ell, m \geq 0$. So $g_{\gamma,\ell,m,k} := \Big((|x|^2 \wedge \gamma^2 + g_{\ell,m}(x) \vee 0\Big)^{1/2} \wedge k \in \mathcal{A}$, $k \geq 0$. With $\gamma \geq \ell + k^2$, $m \geq k$,

$$g_{\gamma,\ell,m,k}(x) = \overline{g}_{\ell,k}(x) := \Big(\Big(|x|^2 + (-2\langle x, y\rangle + |y|^2) \vee (-\ell)\Big) \vee 0\Big)^{1/2} \wedge k \in \mathcal{A},$$

and

$$\lim_{\ell\to\infty} \overline{g}_{\ell,k}(x) = \inf_{\ell\in\mathbb{N}} \overline{g}_{\ell,k}(x) = |x-y| \wedge k$$

for all $x \in X$. It follows that

$$\limsup_{n\to\infty} \int_X \Big(|x-y| \wedge k\Big) \,\mathrm{d}\mu_n(x) \leq \limsup_{n\to\infty} \int_X \overline{g}_{\ell,k}(x) \,\mathrm{d}\mu(x) = \int \overline{g}_{\ell,k} \,\mathrm{d}\mu(x) = \int_X \overline{g}_{\ell,k}(x) \,\mathrm{d}\mu(x).$$

□

1.9 Transport of measures

Let X_1 and X_2 be separable metric spaces, let $\mu \in \mathcal{P}(X_1)$, and let $r\colon X_1 \to X_2$ be a Borel map, or, more generally, a μ-measurable map. Define $r_{\#}\mu \in \mathcal{P}(X_2)$ by

$$r_{\#}\mu(A) := \mu(r^{-1}(A)), \quad A \subseteq X_2 \text{ Borel}.$$

The measure $r_{\#}\mu$ is called the *image measure* of μ under r or the *push forward* of μ through r.

Lemma 1.38. (1) $\int_{X_1} f(r(x))\,\mathrm{d}\mu(x) = \int_{X_2} f(y)\,\mathrm{d}r_{\#}\mu(y)$ *for every bounded Borel function* $f\colon X_2 \to \mathbb{R}$.

(2) *If* ν *is absolutely continuous with respect to* μ, *then* $r_{\#}$ *is absolutely continuous with respect to* $r_{\#}\mu$.

(3) *If* $s\colon X_2 \to X_3$ *is Borel, where* X_3 *is a separable metric space, then* $(s \circ r)_{\#}\mu = s_{\#}(r_{\#}\mu)$.

(4) *If* r *is continuous, then* $r_{\#}\colon \mathcal{P}(X_1) \to \mathcal{P}(X_2)$ *is continuous with respect to narrow convergence.*

Proof. (1): By definition of $r_{\#}$ the statement is clear for simple functions and then follows by approximation.

(2): If $r_{\#}\mu(A) = 0$, then $\mu(r^{-1}(A)) = 0$, so $r_{\#}\nu(A) = \nu(r^{-1}(A)) = 0$.

(3): $(s\circ r)_{\#}\mu(A) = \mu((s\circ r)^{-1}(A)) = \mu(\{x\colon s(r(x)) \in A\}) = r_{\#}\mu(s^{-1}(A)) = s_{\#}(r_{\#}\mu)(A)$.

(4): If $\mu_n \to \mu$ in $\mathcal{P}(X_1)$, then for every open $U \subseteq X_2$ the set $r^{-1}(U)$ is open in X_1, so

$$\begin{aligned}\liminf_{n\to\infty} r_{\#}(U) &= \liminf_{n} \mu_n(r^{-1}(U)) \\ &\geq \mu(r^{-1}(U)) = r_{\#}(U).\end{aligned}$$

So $r_{\#}\mu_n \to \mu$. ☐

Lemma 1.39. *Let* $r_n, r\colon X_1 \to X_2$ *be Borel maps such that* $r_n \to r$ *uniformly on compact subsets of* X_1. *Let* $(\mu_n)_n$ *be a tight sequence in* $\mathcal{P}(X_1)$ *that converges to* μ. *If* r *is continuous, then* $(r_n)_{\#}\mu_n \to r_{\#}\mu$.

Proof. Let $f \in C_b(X_2)$. For $K \subseteq X_1$ compact, we have $f \circ r_n \to f \circ r$ uniformly on K. Let $\varepsilon > 0$. Choose a compact $K \subseteq X_1$ such that $\mu_n(X_1 \setminus K) \geq 1 - (\varepsilon/2)\|f\|_\infty$ for all n. Then

$$\begin{aligned}&|\int f \circ r_n \,\mathrm{d}\mu_n - \int f \circ r \,\mathrm{d}\mu| \\ &\leq |\int f \circ r_n \,\mathrm{d}\mu_n - \int f \circ r \,\mathrm{d}\mu_n| + |\int f \circ r \,\mathrm{d}\mu_n - \int f \circ r \,\mathrm{d}\mu| \\ &\leq 2\|f\|_\infty \mu_n(X_1 \setminus K) + |\int_K (f \circ r_n - f \circ r)\,\mathrm{d}\mu_n| + |\int f \circ r \,\mathrm{d}\mu_n - \int f \circ r \,\mathrm{d}\mu| \\ &\leq \varepsilon + \|(f \circ r_n - f \circ r)|_K\|_\infty + |\int f \circ r \,\mathrm{d}\mu_n - \int f \circ r \,\mathrm{d}\mu|.\end{aligned}$$

The second and third term converge to 0 as $n \to \infty$. ☐

Lemma 1.40. *Let* $X, X_1, X_2, \ldots, X_N$ *be separable metric spaces, let* $r^i\colon X \to X_i$ *be continuous maps, and let* $r := r^1 \otimes \cdots \otimes r^N\colon X \to X_1 \times \cdots \times X_N$ *be such that* $r^{-1}(K_1 \times \cdots K_N)$ *is compact whenever* $K_1, \ldots, K_N$ *are compact. If* $\Gamma \subseteq \mathcal{P}(X)$ *is such that* $\Gamma_i := r^i_{\#}(\Gamma)$ *is tight in* $\mathcal{P}(X_i)$ *for all* i, *then* Γ *is tight in* $\mathcal{P}(X)$

Proof. Denote for $1 \leq i \leq N$ and $\mu \in \Gamma$, $\mu_i := r^i_{\#}\mu$. Let $\varepsilon > 0$. Then there exist compact $K_i \subseteq X_i$ such that $\mu_i(X_i \setminus K_i) < \varepsilon/N$ for all $\mu \in \Gamma$ and all i. So $\mu(X \setminus (r^i)^{-1}(K_i)) < \varepsilon/N$ and

$$\mu(X \setminus \bigcap_{i=1}^{N} (r^i)^{-1}(K_i)) \leq \sum_{i=1}^{N} \mu(X \setminus (r^i)^{-1}(K_i)) < \varepsilon$$

for all $\mu \in \Gamma$. Also, $\bigcap_{i=1}^{N} (r^i)^{-1}(K_i) = r^{-1}(K_1 \times \cdots \times K_N)$ is compact. $\square$

Let $(X_1, d_1), \ldots, (X_N, d_N)$ be separable metric spaces and let $X = X_1 \times \cdots \times X_N$ with its metric defined by $d((x_1, \ldots, x_N), (y_1, \ldots, y_N)) = d_1(x_1, y_1) + \cdots + d_N(x_N, y_N)$. Denote the coordinate projections by $\pi^i(x_1, \ldots, x_N) := x_i$ and $\pi^{i,j}(x_1, \ldots, x_N) := (x_i, x_j)$. If $\mu \in \mathcal{P}(X)$, then the measures $\mu^i := \pi^i_\# \mu$ and $\mu^{i,j} := \pi^{i,j}_\# \mu$ are called the *marginals* of μ. For measures $\mu^i \in \mathcal{P}(X_i)$, define the set of *multiple plans* with marginals μ^i by

$$\Gamma(\mu^1, \ldots, \mu^N) := \{\mu \in \mathcal{P}(X_1 \times \cdots \times X_N) \colon \pi^i_\# \mu = \mu^i,\ i = 1, \ldots, N\}.$$

For $N = 2$, $\mu \in \Gamma(\mu^1, \mu^2)$ is called a *transport plan* between μ^1 and μ^2.

Example. Let $X_1 = X_2 = \{1, 2\}$ and $\mu^1(\{1\}) = \mu^1(\{2\}) = 1/2$, $\mu^2 = \mu^1$. Then $\Gamma(\mu^1, \mu^2) = \{\eta_p \colon 0 \le p \le 1/2\}$, where $\eta_p(\{(1,1)\}) = \eta_p(\{(2,2)\}) = p$ and $\eta_p(\{(1,2)\}) = \eta_p(\{(2,1)\}) = 1/2 - p$.

If given two-dimensional marginals are compatible, we can find a three-dimensional measure with these marginals.

Lemma 1.41. *Let X_1, X_2, X_3 be complete separable metric spaces and let $\gamma^{12} \in \mathcal{P}(X_1 \times X_2)$ and $\gamma^{13} \in \mathcal{P}(X_1 \times X_3)$.*

(1) *If $\pi^1_\# \gamma^{12} = \pi^1_\# \gamma^{13} = \mu^1$ for some $\mu^1 \in \mathcal{P}(X_1)$, then there exists a $\mu \in \mathcal{P}(X_1 \times X_2 \times X_3)$ such that $\pi^{12}_\# \mu = \gamma^{12}$ and $\pi^{13}_\# \mu = \gamma^{13}$.*

(2) *Suppose $\gamma^{12}_{x_1} \in \mathcal{P}(X_2)$, $\gamma^{13}_{x_1} \in \mathcal{P}(X_3)$, and $\mu_{x_1} \in \mathcal{P}(X_2 \times X_3)$, $x_1 \in X_1$, are such that $\gamma^{12} = \int \gamma^{12}_{x_1} \, \mathrm{d}\mu^1$, that is,*

$$\int f(x_1, x_2) \, \mathrm{d}\gamma^{12}(x_1, x_2) = \int_{X_1} \left(\int_{X_2} f(x_1, x_2) \, \mathrm{d}\gamma^{12}_{x_1}(x_2) \right) \mathrm{d}\mu^1(x_1),$$
$$\forall \text{ Borel } f \colon X_1 \times X_2 \to [0, \infty],$$

and, similarly, $\gamma^{13} = \int \gamma^{13}_{x_1} \, \mathrm{d}\mu^1$ and $\mu = \int \mu_{x_1} \, \mathrm{d}\mu^1 \in \mathcal{P}(X_1 \times X_2 \times X_3)$. Then $\pi^{12}_\# \mu = \gamma^{12}$ and $\pi^{13}_\# \mu = \gamma^{13}$ if and only if $\mu_{x_1} \in \Gamma(\gamma^{12}_{x_1}, \gamma^{13}_{x_1})$ for μ^1-a.e. $x_1 \in X_1$.

Proof. (1): By the disintegration theorem, there are families of measures $\gamma^{12}_{x_1} \in \mathcal{P}(X_2)$ and $\gamma^{13}_{x_1} \in \mathcal{P}(X_3)$, $x_1 \in X_1$, such that

$$\int f(x_1, x_2) \, \mathrm{d}\gamma^{12}(x_1, x_2) = \int \left(\int f(x_1, x_2) \, \mathrm{d}\gamma^{12}_{x_1}(x_2) \right) \mathrm{d}\mu^1(x_1)$$

and a similar equation with 3 instead of 2. Define

$$\mu(Z) := \int_{X_1} \left(\int_{X_2 \times X_3} \mathbb{1}_Z(x_1, x_2, x_3) \, \mathrm{d}(\gamma^{12}_{x_1} \otimes \gamma^{13}_{x_1})(x_2, x_3) \right) \mathrm{d}\mu^1(x_1).$$

Then $\mu \in \mathcal{P}(X_1 \times X_2 \times X_3)$ and

$$\mu(A \times B \times X_3) = \int_A \left(\int_B \mathrm{d}\gamma^{12}_{x_1}(x_2) \right) \mathrm{d}\mu^1(x_1) = \int_{A \times B} \mathrm{d}\gamma^{12} = \gamma^{12}(A \times B)$$

for Borel sets $A \subseteq X_1$ and $B \subseteq X_2$. A similar equations holds with 3 instead of 2.

(2): If $\pi^{12}_{\#}\mu = \gamma^{12}$ and $\pi^{13}_{\#}\mu = \gamma^{13}$, then

$$\begin{aligned}
\gamma^{12}(A \times B) &= \pi^{12}_{\#}\mu(A \times B) = \mu((\pi^{12})^{-1}(A \times B)) \\
&= \int_{X_1} \left(\int_{X_2 \times X_3} \mathbb{1}_{(\pi^{12})^{-1}(A\times B)}(x_1, x_2, x_3) \, \mathrm{d}\mu_{x_1}(x_2, x_3) \right) \mathrm{d}\mu^1(x_1) \\
&= \int_{X_1} \mu_{x_1}((\pi^{12})^{-1}(A \times B)) \, \mathrm{d}\mu^1(x_1) \\
&= \int_{X_1} \pi^{12}_{\#}\mu_{x_1}(A \times B) \, \mathrm{d}\mu^1(x_1)
\end{aligned}$$

and

$$\begin{aligned}
\int f(x_1, x_2) \, \mathrm{d}\gamma^{12} &= \int f \, \mathrm{d}\pi^{12}_{\#}\mu \\
&= \int f \circ \pi^{12} \, \mathrm{d}\mu = \int f(x_1, x_2) \, \mathrm{d}\mu(x_1, x_2, x_3) \\
&= \int_{X_1} \left(\int f(x_1, x_2) \, \mathrm{d}\mu_{x_1}(x_2, x_3) \right) \mathrm{d}\mu^1(x_1) \\
&= \int_{X_1} \left(\int f(x_1, x_2) \, \mathrm{d}\pi^2_{\#}\mu_{x_1}(x_2) \right) \mathrm{d}\mu^1(x_1) \\
&= \int_{X_1} \left(\int f \, \mathrm{d}\pi^2_{\#}\mu_{x_1} \right) \mathrm{d}\mu^1.
\end{aligned}$$

By the uniqueness part of the disintegration theorem, $\gamma^{12}_{x_1} = \pi^2_{\#}\mu_{x_1}$ for μ^1-a.e. $x_1 \in X_1$. Similarly, $\gamma^{13}_{x_1} = \pi^3_{\#}\mu_{x_1}$.

Conversely, suppose $\pi^2_{\#}\mu_{x_1} = \gamma^{12}_{x_1}$ and $\pi^3_{\#}\mu_{x_1} = \gamma^{13}_{x_1}$ for μ^1-a.e. x_1. Then

$$\begin{aligned}
\int f(x_1, x_2) \, \mathrm{d}\gamma^{12} &= \int_{X_1} \left(\int f(x_1, x_2) \, \mathrm{d}\gamma^{12}_{x_1} \right) \mathrm{d}\mu^1 \\
&= \int_{X_1} \left(\int f(x_1, x_2) \, \mathrm{d}\pi^2_{\#}\mu_{x_1} \right) \mathrm{d}\mu^1 \\
&= \int f(x_1, x_2) \, \mathrm{d}\pi^2_{\#}\mu,
\end{aligned}$$

so $\gamma^{12} = \pi^2_{\#}\mu$. $\square$

2 Optimal transportation problems

Optimal transportation problems aim to minimize costs or energy needed to transport mass from a given initial state to a given final state. We will consider the Monge and Kantorovich optimal transportation problems in metric spaces and discuss existence and uniqueness of optimal transportation plans.

2.1 Introduction

In order to have an economical interpretation in mind, we begin by explaining a simple instance of a transportation problem. Suppose a certain amount of milk is available in three distribution centers and should be transported to five supermarkets. Let ν_j be the amount available at the jth center and μ_i be the amount needed at the ith supermarket. We may scale such that $\sum_{j=1}^{3} \nu_j = 1$ and $\sum_{i=1}^{5} \mu_i = 1$. Let c_{ij} denote the costs of transporting one unit milk from center j to supermarket i. How to transport the milk for minimal costs?

The first formulation of the problem is due to Monge (1781). Assume that supermarket i gets all its milk from one distribution center, say $r(i) \in \{1,2,3\}$. As the jth center has to send out all of its milk we have $\nu_j = \sum_{i:\, r(i)=j} \mu_i$. We want to find the map $r : \{1,2,3,4,5\} \to \{1,2,3\}$ such that the total costs $\sum_{i=1}^{5} c_{i,r(i)}\mu_i$ are minimal. In other words, we want to solve

$$\min\{\sum_{i=1}^{5} c_{i,r(i)}\mu_i\colon r : \{1,2,3,4,5\} \to \{1,2,3\} \text{ such that } \nu_j = \sum_{i:\, r(i)=j} \mu_i\}.$$

Solving the problem comes down to drawing arrows from j in $\{1,2,3\}$ to $i \in \{1,2,3,4,5\}$ in a most cost effective way. (Each i is reached by exactly one arrow.)

The second formulation is more general and due to Kantorovich (1942). We now allow each supermarket to receive milk from more than one distribution center. Let γ_{ij} be the amount sent from j to i. We want to solve

$$\min\{\sum_{i,j} c_{ij}\gamma_{ij}\colon \gamma_{ij} \geq 0 \text{ such that } \sum_{i} \gamma_{ij} = \nu_j,\ \sum_{j} \gamma_{ij} = \mu_i\}.$$

Solving the problem comes down to filling the matrix $(\gamma_{ij})_{ij}$ with the given row and column sums in the most cost effective way.

We can make an important observation about the structure of an optimal matrix (γ_{ij}). Suppose that (γ_{ij}) is a transportation plan with minimal costs and suppose that the two entries γ_{ij} and $\gamma_{k\ell}$ are strictly positive. Then we should have $c_{ij} + c_{k\ell} \leq c_{i\ell} + c_{kj}$. Indeed, we would otherwise be able to decrease the costs by decreasing γ_{ij} and $\gamma_{k\ell}$ both with the amount $\alpha := \min\{\gamma_{ij}, \gamma_{k\ell}\}$ and increasing $\gamma_{i\ell}$ and γ_{kj} both by α. This observation will lead to the notion of c-monotonicity of the support of an optimal transportation plan.

There is a dual point of view to the Kantorovich problem. Of course the supermarkets and distribution centers do not want to pay for the transportation of the milk, but they have to in order to receive or get rid of their supplies. Let us denote the amount that supermarket i is willing to pay per unit by φ_i and the amount that distribution center j is willing to pay per unit by ψ_j. As they will never pay more than needed, it is clear that $\varphi_i + \psi_j \leq c_{ij}$. The total amount that they pay equals $\sum_i \varphi_i\mu_i + \sum_j \psi_j\nu_j$. It turns out that the maximal total amount that is acceptable for them to pay

$$\max\{\sum_{i=1}^{5} \varphi_i\mu_i + \sum_{j=1}^{3} \psi_j\nu_j\colon \varphi_i + \psi_j \leq c_{ij}\}$$

equals the minimal total costs

$$\min\{\sum_{ij} c_{ij}\gamma_{ij}\colon \sum_{\gamma_{ij}} = \nu_j,\ \sum_{j} \gamma_{ij} = \mu_i\}.$$

Under the assumption that $c_{ij} \geq 0$ for all i and j, the Kantorovich problem has a solution. There may be several solutions, as is easily seen in the extreme case that $c_{ij} = c_{11}$

for all i and j. The Monge problem, however, may not have a solution. For instance, if the distributions μ and ν are uniform, that is $\nu_1 = \nu_2 = \nu_3 = 1/3$ and $\mu_1 = \cdots = \mu_5 = 1/5$, then there is no map $r : \{1, 2, 3, 4, 5\} \to \{1, 2, 3\}$ such that $r_{\#}\mu = \nu$.

Let us next consider Monge's and Kantorovich's problems in metric spaces. Let X and Y be two separable complete metric spaces and let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Recall that we endow the product space $X \times Y$ with the metric $d((x, y), (x', y')) := d_X(x, x') + d_Y(y, y')$ and that we denote the coordinate projections by $\pi^X(x, y) = x$ and $\pi^Y(x, y) = y$ for all $x \in X$ and $y \in Y$. Let c: $X \times X \to [0, \infty)$ be Borel measurable. We can now formulate Monge's problem as

$$\min\{\int c(x, r(x))\,\mathrm{d}\mu(x)\colon\ r\colon X \to Y \text{ a Borel map such that } r_{\#}\mu = \nu\},$$

where $r_{\#}$ denotes the image measure of μ under r. Recall that

$$\Gamma(\mu, \nu) := \{\gamma \in \mathcal{P}(X \times Y)\colon (\pi^X)_{\#}\gamma = \mu \text{ and } (\pi^Y)_{\#}\gamma = \nu\}$$

is the set of transportation plans. The Kantorovich problems reads

$$\min\{\int c(x, y)\,\mathrm{d}\gamma(x, y)\colon \gamma \in \Gamma(\mu, \nu)\}.$$

The dual problem to the Kantorovich problem is

$$\max\{\int \varphi\,\mathrm{d}\mu + \int \psi\,\mathrm{d}\nu\colon \varphi \in L^1(\mu),\ \psi \in L^1(\nu),\ \varphi(x) + \psi(y) \le c(x, y)\ \forall x \in X\, \forall y \in Y\}.$$

Definition 2.1. The measure $\eta \in \Gamma(\mu, \nu)$ is called *optimal for c* if

$$\int c\,\mathrm{d}\eta = \min\{\int c(x, y)\,\mathrm{d}\gamma(x, y)\colon \gamma \in \Gamma(\mu, \nu)\}$$

(possibly $\infty = \infty$).

Definition 2.2. For a function φ: $X \to [-\infty, \infty]$, the *c-transform* of φ is defined by

$$\varphi^c(y) = \inf_{x\in X} \Big(c(x, y) - \varphi(x)\Big), \qquad y \in Y.$$

The c-transform of a function ψ: $Y \to [-\infty, \infty]$ is

$$\psi^c(x) = \inf_{y\in Y} \Big(c(x, y) - \psi(y)\Big), \qquad x \in X.$$

Here we use the convention that $\inf \emptyset = \infty$, $\inf \infty = \infty$, $\inf(-\infty) = -\infty$, and the infimum of a set that is not bounded below is $-\infty$.

The support of an optimal transportation plan has a structure that is described in the following definition.

Definition 2.3. A set $S \subseteq X \times Y$ is called *c-monotone* if

$$\sum_{i=1}^{n} c(x_{\sigma(i)}, y_i) \ge \sum_{i=1}^{n} c(x_i, y_i)$$

for every $(x_i, y_i) \in S$, $i = 1, \ldots, n$, and every permutation σ of $\{1, \ldots, n\}$.

2.2 Existence for the Kantorovich problem

The existence of an optimal measure for the Kantorovich problem is a consequence of Prokhorov's theorem. The next lemma is actually contained in Lemma 1.38 and Lemma 1.40, but for convenience we include an explicit proof.

Lemma 2.4. *Let X and Y be separable complete metric spaces and let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Then $\Gamma(\mu,\nu)$ is a compact subset of $\mathcal{P}(X \times Y)$.*

Proof. Notice that $X \times Y$ is also separable and complete. We first show that $\Gamma(\mu,\nu)$ is tight. As X and Y are separable and complete, the measures μ and ν are tight. Let $\varepsilon > 0$. Choose compact sets $K \subseteq X$ and $L \subseteq Y$ such that $\mu(K) \geq 1 - \varepsilon/2$ and $\nu(L) \geq 1 - \varepsilon/2$. Then $K \times L$ is compact in $X \times Y$ and for $\gamma \in \Gamma(\mu,\nu)$,

$$\begin{aligned} \gamma(X \times Y \setminus K \times L) &\leq \gamma((X \setminus K) \times Y) + \gamma(X \times (Y \setminus L)) \\ &= \nu(X \setminus K) + \mu(Y \setminus L) \leq \varepsilon/2 + \varepsilon/2. \end{aligned}$$

Hence $\Gamma(\mu,\nu)$ is tight. By Prokhorov's theorem, $\Gamma(\mu,\nu)$ is relatively compact in $\mathcal{P}(X \times Y)$.

It remains to show that $\Gamma(\mu,\nu)$ is closed in $\mathcal{P}(X \times Y)$. Let $(\gamma_n)_n$ be a sequence in $\Gamma(\mu,\nu)$ and $\eta \in \mathcal{P}(X \times Y)$ be such that $\gamma_n \to \eta$ narrowly. Due to the Portmanteau theorem we have for any $C \subseteq X$ closed,

$$\begin{aligned} \eta(C \times Y) &\geq \limsup_{n\to\infty} \eta_n(C \times Y) \\ &= \limsup_{n\to\infty} \mu(C) = \mu(C) \end{aligned}$$

and for $U \subseteq X$ open,

$$\begin{aligned} \eta(U \times Y) &\leq \liminf_{n\to\infty} \eta_n(U \times Y) \\ &= \liminf_{n\to\infty} \mu(U) = \mu(U). \end{aligned}$$

Let $C \subseteq X$ be closed and let

$$U_m := \{x \in X\colon \operatorname{dist}(x, C) < 1/m\}, \quad m \geq 1.$$

Then each U_m is open and $\bigcap_{m\geq 1} U_m = C$ and $\bigcap_{m \geq 1}(U_m \times Y) = C \times Y$. Hence

$$\begin{aligned} \eta(C \times Y) &= \lim_{m\to\infty} \eta(U_m \times Y) \\ &\leq \lim_{m\to\infty} \mu(U_m) = \mu(C). \end{aligned}$$

Thus, $\eta(C \times Y) = \mu(C)$. Hence μ is the marginal on X of η. In a similar way we can show that the marginal of η on Y is ν and therefore $\eta \in \Gamma(\mu,\nu)$. ☐

Theorem 2.5. *Let X and Y be separable complete metric spaces, let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Let $c\colon X \times Y \to [0,\infty)$ be continuous. Then there exists $\eta \in \Gamma(\mu,\nu)$ such that*

$$\int c \,\mathrm{d}\eta = \min\{\int c \,\mathrm{d}\gamma\colon \gamma \in \Gamma(\mu,\nu)\}.$$

Proof. Write $\alpha := \min\{\int c\,\mathrm{d}\gamma\colon \gamma \in \Gamma(\mu,\nu)\}$. If $\alpha = \infty$, then $\eta = \mu \otimes \nu \in \Gamma(\mu,\nu)$ satisfies $\int c\,\mathrm{d}\eta = \infty = \alpha$. Otherwise, for $n \geq 1$, take a $\gamma_n \in \Gamma(\mu,\nu)$ with

$$\int c\,\mathrm{d}\eta_n \leq \alpha + 1/n.$$

By the previous lemma, $\Gamma(\mu,\nu)$ is compact. Hence there exists $\eta \in \Gamma(\mu,\nu)$ and a subsequence $eta_{n_k} \to \eta$ as $k \to \infty$. For $k \geq 1$ and $m \geq 1$ we have $\int c \wedge m\,\mathrm{d}\eta_{n_k} \leq \int c\,\mathrm{d}\eta_{n_k} \leq \alpha + 1/n_k$. Hence

$$\int c\,\mathrm{d}\eta = \lim_{m\to\infty} \int c \wedge m\,\mathrm{d}\eta = \lim_{m\to\infty}\lim_{k\to\infty} \int c \wedge m\,\mathrm{d}\eta_{n_k} \leq \alpha.$$

□

Remark. The previous theorem is a special instance of a general principle: a l.s.c. function on a compact metric space has a minimum. The above lemma says that $\Gamma(\mu,\nu)$ is a compact metric space. The map $\eta \mapsto \int c\,\mathrm{d}\eta$ is l.s.c. due to Proposition 1.12(c).

2.3 Characterization of optimal plans

The next theorem characterizes optimal transportation plans. We begin with a lemma.

Lemma 2.6. *Let X and Y be separable complete metric metric spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $\eta \in \Gamma(\mu,\nu)$. Then there exists a μ-full Borel set $A \subseteq X$ such that*

$$\forall x \in A\ \exists y \in Y\colon \quad (x,y) \in \operatorname{supp}\eta.$$

Further, $\mu(\pi^X(\operatorname{supp}\eta)) = 1$ and $\nu(\pi^Y(\operatorname{supp}\eta)) = 1$.

Proof. The set $S := \operatorname{supp}\eta$ is closed hence Borel. As $X \times Y$ is separable and complete, η is tight, so

$$1 = \eta(S) = \sup\{\eta(K)\colon K \subseteq S,\ K \text{ compact}\}$$

Choose $K_n \subseteq S$ compact such that $\eta(K_n) \geq 1 - 1/n$, for $n \geq 1$. Then $\pi^X(K_n)$ is compact in X and $\mu(\pi^X(K_n)) = \eta(K_n) \geq 1 - 1/n$. Hence $A := \bigcup_n K_n$ is a μ-full Borel set in X. If $x \in A$ then $x \in \pi^X(K_n)$ for some n, so $(x,y) \in K_n \subseteq S$ for some $y \in Y$. □

Theorem 2.7. *Let X and Y be separable complete metric spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $c\colon X \times Y \to [0,\infty)$ continuous.*

(1) *If $\eta \in \Gamma(\mu,\nu)$ is optimal for c and $\int c\,\mathrm{d}\eta < \infty$, then*

$$\operatorname{supp}\eta := \{z \in X \times Y\colon \eta(U) > 0 \textit{ for every neighborhood } U \textit{ of } z\}$$

is a c-monotone set.

(2) *If $\eta \in \Gamma(\mu,\nu)$ is such that*

- $\operatorname{supp}\eta$ *is c-monotone, and*
- $\mu\Big(\{x \in X\colon \int_Y c(x,y)\,\mathrm{d}\nu(y) < \infty\}\Big) > 0$, *and*

– $\nu\Big(\{y \in Y\colon \int_X c(x,y)\,\mathrm{d}\mu(x) < \infty\}\Big) > 0$,

then η is optimal for c.

(3) *In the situation of (2), one also has*

$$\begin{aligned} &\min\Big\{\int c\,\mathrm{d}\gamma\colon \gamma \in \Gamma(\mu,\nu)\Big\} \\ &\quad = \max\Big\{\int \varphi\,\mathrm{d}\mu + \int \psi\,\mathrm{d}\nu\colon \varphi \in L^1(\mu),\ \psi \in L^1(\nu), \\ &\qquad\qquad \varphi(x) + \psi(y) \le c(x,y)\ \forall (x,y) \in X \times Y\Big\} \end{aligned}$$

and the maximum at the right hand side is attained at

$$\begin{aligned} \varphi(x) \ &= \inf\Big\{\textstyle\sum_{i=0}^{p}\Big(c(x_{i+1},y_i) - c(x_i,y_i)\Big)\colon p \in \mathbb{N},\ x_{p+1} = x, \\ &\qquad (x_i,y_i) \in \operatorname{supp}\eta,\ i = 1,\dots,p\Big\} \end{aligned}$$

for any choice of $(x_0,y_0) \in \operatorname{supp}\eta$ and $\psi = \varphi^c$.

Proof. (1): (See [7, Theorem 2.3].) Suppose that $\operatorname{supp}\eta$ is not c-monotone. Then there are $n \in \mathbb{N}$ and a permutation σ of $\{1,\dots,n\}$ such that the function

$$f(u_1,\dots,u_n,v_1,\dots,v_n) := \sum_{i=1}^{n}\Big(c(u_{\sigma(i)},v_i) - c(u_i,v_i)\Big)$$

is strictly negative at some $(x_1,\dots,x_n,y_1,\dots,y_n)$ with $(x_i,y_i) \in \operatorname{supp}\eta$. We will construct a more cost efficient measure than η and thus show that η is not optimal for c.

As f is continuous, we can choose Borel neighborhoods U_i of x_i and V_i of y_i such that $f(u_1,\dots,u_n,v_1,\dots,v_n) < 0$ for $u_i \in U_i$ and $v_i \in V_i$, $i = 1,\dots,n$. As $(x_i,y_i) \in \operatorname{supp}\eta$,

$$\lambda := \min_i \eta(U_i \times V_i) > 0.$$

Define $\eta_i \in \mathcal{P}(X \times Y)$ by

$$\eta_i(W) := \frac{1}{\eta(U_i \times V_i)}\eta((U_i \times V_i) \cap W), \quad W \subseteq X \times Y \text{ Borel.}$$

Consider

$$Z = (X \times Y)^n$$

and $\rho \in \mathcal{P}(Z)$ given by

$$\rho = \eta_1 \otimes \dots \otimes \eta_n.$$

Let $\pi_i^X\colon Z \to X$ be defined by $\pi_i^X(u_1,v_1,\dots,u_n,v_n) := u_i$ and $\pi_i^Y\colon Z \to Y$ by $\pi_i^Y(u_1,v_1,\dots,u_n,v_n) := v_i$. Recall that $\pi_i^X \otimes \pi_j^Y$ denotes the map $(u_1,v_1,\dots,u_n,v_n) \mapsto (u_i,v_j)$. Define

$$\begin{aligned} \gamma \ &:= \ \eta - \frac{\lambda}{n}\sum_{i=1}^{n}(\pi_i^X \otimes \pi_i^Y)_{\#}\rho + \frac{\lambda}{n}\sum_{i=1}^{n}(\pi_{\sigma(i)}^X \otimes \pi_i^Y)_{\#}\rho \\ &= \ \eta - \frac{\lambda}{n}\sum_{i=1}^{n}\eta_i + \frac{\lambda}{n}\sum_{i=1}^{n}(\pi_{\sigma(i)}^X \otimes \pi_i^Y)_{\#}\rho. \end{aligned}$$

Then

$$\begin{aligned}
\gamma(W) &\geq \eta(W) - \frac{\lambda}{n}\sum_{i=1}^{n}\eta_i(W) \\
&\geq \eta(W) - \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda}{\eta(U_i \times V_i)}\eta((U_i \times V_i)\cap W) \\
&\geq \eta(W) - \frac{1}{n}\sum_{i=1}^{n}\eta(W) = 0,
\end{aligned}$$

for every Borel set $W \subseteq X \times Y$. So γ is a positive Borel measure. It is easy to check that $\gamma \in \mathcal{P}(X \times Y)$. Further, for $A \subseteq X$ Borel,

$$\begin{aligned}
(\pi^X_{\sigma(i)} \otimes \pi^Y_i)_{\#}\rho(A \times Y) &= \rho(\{(u_1, v_1, \ldots, u_n, v_n) \in Z \colon (u_{\sigma(i)}, v_i) \in A \times Y\}) \\
&= \rho(\{(u_1, v_1, \ldots, u_n, v_n) \in Z \colon u_{\sigma(i)} \in A\}) \\
&= \eta_{\sigma(i)}(A \times Y),
\end{aligned}$$

so

$$\begin{aligned}
\gamma(A \times Y) &= \eta(A \times Y) - \frac{\lambda}{n}\sum_{i=1}^{n}\eta_i(A \times Y) + \frac{\lambda}{n}\sum_{i=1}^{n}(\pi^X_{\sigma(i)} \otimes \pi^Y_i)_{\#}\rho(A \times Y) \\
&= \mu(A) - \frac{\lambda}{n}\sum_{i=1}^{n}\eta_i(A \times Y) + \frac{\lambda}{n}\sum_{i=1}^{n}\eta_{\sigma(i)}(A \times Y) = \mu(A)
\end{aligned}$$

and similarly $\gamma(X \times B) = \nu(B)$ for $B \subseteq Y$ Borel. Hence $\gamma \in \Gamma(\mu, \nu)$.

Finally,

$$\int_{X\times Y} c \,\mathrm{d}(\pi^X_i \otimes \pi^Y_j)_{\#}\rho = \int_Z c(\pi^X_i(z), \pi^Y_j(z))\,\mathrm{d}\rho,$$

so

$$\begin{aligned}
\int c\,\mathrm{d}\gamma &= \int c\,\mathrm{d}\eta + \frac{\lambda}{n}\sum_{i=1}^{n}\int_Z \Big(c(\pi^X_{\sigma(i)}(z), \pi^Y_i(z)) - c(\pi^X_i(z), \pi^Y_i(z)\Big)\,\mathrm{d}\rho(z) \\
&= \int c\,\mathrm{d}\eta + \frac{\lambda}{n}\int_{U_1\times V_1\times\cdots\times U_n\times V_n} f(\pi^X_1(z), \ldots, \pi^X_n(z), \pi^Y_1(z), \ldots, \pi^Y_n(z))\,\mathrm{d}\rho(z) \\
&< \int c\,\mathrm{d}\eta,
\end{aligned}$$

since ρ is concentrated on $U_1 \times V_1 \times \cdots \times U_n \times V_n$ and $f < 0$ on this set. Thus we have that γ is more cost efficient than η, so that η is not optimal.

(2) and (3): Let $S := \operatorname{supp}\eta$, which is a c-monotone subset of $X \times Y$. Fix $(x_0, y_0) \in S$ ($\eta(S) = 1$ so S is non-empty) and let φ be defined as above. The proof is divided into several claims, clustered by topic. We first establish some properties of φ, then of $\psi = \varphi^c$, and then we show that φ and ψ are L^1 functions. Then we derive some more connections between φ, ψ and η, and finally we conclude the proof.

Define

$$\begin{aligned}
\varphi_q(x) := \inf\Big\{&\sum_{i=0}^{p}\Big(c(x_{i+1}, y_i) - c(x_i, y_i)\Big)\colon \\
&x_{p+1} = x,\ (x_i, y_i) \in S,\ i = 1, \ldots, p,\ 1 \leq p \leq q\Big\}.
\end{aligned}$$

Clearly $\varphi_q(x) \downarrow \varphi(x)$ for all $x \in X$.

Claim A1: φ_q is upper semicontinuous for each q. Suppose $u_k \to u$ in X. Let $\varepsilon > 0$. Then

$$\varphi_q(u) \geq c(x, y_p) - c(x_p, y_p) + \sum_{i=0}^{p-1} \Big(c(x_{i+1}, y_i) - c(x_i, y_i) \Big) - \varepsilon$$

for some $p \leq q$ and $(x_i, y_i) \in S$. Then

$$\varphi_q(u_k) \leq c(u_k, y_p) - c(x_p, y_p) + \sum_{i=0}^{p-1} \Big(c(x_{i+1}, y_i) - c(x_i, y_i) \Big),$$

so

$$\begin{aligned} \limsup_{k\to\infty} \varphi_q(u_k) &\leq c(u, y_p) - c(x_p, y_p) + \sum_{i=0}^{p-1} \Big(c(x_{i+1}, y_i) - c(x_i, y_i) \Big) \\ &\leq \varphi_q(u) + \varepsilon. \end{aligned}$$

Hence φ_q is u.s.c.

Claim A2: φ is Borel measurable. We know that φ_q is u.s.c. hence Borel and $\varphi_q \to \varphi$ pointwise.

Claim A3: $\varphi(x_0) = 0$. On one hand, choose $(x_1, y_1) = (x_0, y_0) \in S$. Then $\varphi(x_0) \leq c(x_0, y_1) - c(x_1, y_1) + c(x_1, y_0) - c(x_0, y_0) = 0$. On the other hand, as S is c-monotone, for $(x_i, y_i) \in S$, $i = 1, \ldots, p$,

$$\sum_{i=0}^{p} c(x_{\sigma(i)}, y_i) \geq \sum_{i=0}^{p} c(x_i, y_i),$$

in particular with the permutation $\sigma(i) = i + 1$ for $0 \leq i \leq p - 1$ and $\sigma(p) = 0$. So, with the notation $x_{p+1} = x_0$,

$$\sum_{i=0}^{p} \Big(c(x_{i+1}, y_i) - c(x_i, y_i) \Big) \geq 0,$$

so $\varphi(x_0) \geq 0$. Hence $\varphi(x_0) = 0$.

Claim A4: $\varphi(u) \leq \varphi(x) + c(u, y) - c(x, y)$ for all $u \in X$ and $(x, y) \in S$. For any $p \in \mathbb{N}$ and $(x_i, y_i) \in S$, $i = 1, \ldots, p$, we have

$$\begin{aligned} \varphi(u) &\leq \varphi_{p+1}(u) \\ &\leq c(u, y) - c(x, y) + \sum_{i=0}^{p} \Big(c(x_{i+1}, y_i) - c(x_i, y_i) \Big), \end{aligned}$$

where $x_{p+1} = x$. So, by taking infimum over $\{(x_i, y_i) \colon 0 \leq i \leq p\}$,

$$\varphi(u) \leq c(u, y) - c(x, y) + \varphi(x).$$

Claim A5: $\varphi > -\infty$ on $\pi^X(S)$, so $\varphi > -\infty$ μ-a.e. If $(x, y) \in S$, then by Claim A4,

$$\begin{aligned} \varphi(x) &\geq \varphi(x_0) - c(x_0, y) + c(x, y) \\ &= c(x, y) - c(x_0, y) \in \mathbb{R}. \end{aligned}$$

By the previous lemma we obtain $\mu(\pi^X(S)) = 1$.

Claim B1: Let $\psi := \varphi^c$. Then $\varphi(x) + \psi(y) = c(x,y)$ for all $(x,y) \in S$. Indeed, by definition,

$$\psi(y) = \inf_{u \in X} \Big(c(u,y) - \varphi(u) \Big).$$

We have by Claim A4 that $c(u,y) - \varphi(u) \geq c(x,y) - \varphi(x)$, so

$$\psi(y) \geq c(x,y) - \varphi(x).$$

From the definition of ψ we find with $u = x$ also $\psi(y) \leq c(x,y) - \varphi(x)$.

By Claim A5 it follws that $\psi(y) \in \mathbb{R}$ for ν-a.e. $y \in Y$.

Claim B2: ψ is ν-measurable. Due to Claim B1,

$$\psi(y) \mathbb{1}_S(x,y) = \Big(c(x,y) - \varphi(x) \Big) \mathbb{1}_S(x,y) \quad \text{for all } (x,y) \in X \times Y$$

and $(x,y) \mapsto c(x,y) - \varphi(x)$ is a Borel map by Claim A2. Hence $(x,y) \mapsto \psi(y)\mathbb{1}_S(x,y)$ is η-measurable. By disintegration, there exist $\eta_y \in \mathcal{P}(X)$, $y \in Y$, such that $y \mapsto \int_X f(x,y) \, \mathrm{d}\eta_y(x)$ is ν-measurable and

$$\int_{X \times Y} f(x,y) \, \mathrm{d}\eta(x,y) = \int_Y \left(\int_X f(x,y) \, \mathrm{d}\eta_y(x) \right) \mathrm{d}\nu(y)$$

for every Borel function $f \colon X \times Y \to [0, \infty]$. The set $S = \operatorname{supp} \eta$ is closed and therefore a Borel set. From the disintegration formula with $f = \mathbb{1}_S$ we obtain that

$$\int_X \mathbb{1}_S(x,y) \, \mathrm{d}\eta_y(x) = 1 \quad \text{for } \nu\text{-almost every } y.$$

If we apply now the disintegration to $f(x,y) = \Big(c(x,y) - \varphi(x) \Big)^+ \mathbb{1}_B(y)$ for some Borel set $B \subseteq Y$, then

$$\begin{aligned} \int_{X \times B} \psi^+(y) \mathbb{1}_S(x,y) \, \mathrm{d}\eta(x,y) &= \int_{X \times B} \Big(c(x,y) - \varphi(x) \Big)^+ \mathrm{d}\eta(x,y) \\ &= \int_B \left(\int_X \Big(c(x,y) - \varphi(x) \Big)^+ \mathrm{d}\eta_y(x) \right) \mathrm{d}\nu(y) \end{aligned}$$

and

$$\begin{aligned} \int_{X \times B} \psi^+(y) \mathbb{1}_S(x,y) \, \mathrm{d}\eta(x,y) &= \int_B \left(\int_X \psi^+(y) \mathbb{1}_S(x,y) \, \mathrm{d}\eta_y(x) \right) \mathrm{d}\nu(y) \\ &= \int_B \psi^+(y) \left(\int_X \mathbb{1}_S(x,y) \, \mathrm{d}\eta_y(x) \right) \mathrm{d}\nu(y). \end{aligned}$$

It follows that

$$\psi^+(y) \left(\int_X \mathbb{1}_S(x,y) \, \mathrm{d}\eta_y(x) \right) = \int_X \Big(c(x,y) - \varphi(x) \Big)^+ \mathrm{d}\eta_y(x) \quad \text{for } \nu\text{-a.e. } y \in Y,$$

so

$$\psi^+(y) = \int_X \Big(c(x,y) - \varphi(x) \Big)^+ \mathrm{d}\eta_y(x) \quad \text{for } \nu\text{-a.e. } y \in Y.$$

Hence ψ^+ is ν-measurable. Similarly, ψ^- is ν-measurable and thus ψ is ν-measurable.

Claim C1: $\psi^+(y) \le c(x,y)+\varphi^-(x)$ for all $(x,y) \in S$. As $c \ge 0$, we have $c(x,y)+\varphi^-(x) \ge 0$. Also $c(x,y)+\varphi^-(x) \ge c(x,y)-\varphi(x) = \psi(y)$. Hence $c(x,y)+\varphi^-(x) \ge \psi^+(y)$.

Claim C2: $\varphi^+ \in L^1(\mu)$ and $\psi^+ \in L^1(\nu)$. By assumption, $\mu(A) > 0$, where

$$A := \Big\{x \in X\colon \int_Y c(x,y)\,\mathrm{d}\nu(y) < \infty\Big\}.$$

Choose $x \in A$ such that $\nu(\{y\colon (x,y) \in S\}) = 1$. Then $\psi^+ \le c(x,\cdot)+\varphi^-(x)$ ν-a.e. on Y (by Claim C1), so

$$\int_Y \psi^+\,\mathrm{d}\nu \le \int_Y \Big(c(x,y)+\varphi^-(x)\Big)\,\mathrm{d}\nu(y) < \infty,$$

since $x \in A$ and $\varphi^-(x) \in \mathbb{R}$ (by Claim A5). Similarly, $\varphi^+(x) \le c(x,y)+\psi^-(y)$ for $(x,y) \in S$ and there exists a y s.t. $\mu(\{x\colon (x,y) \in S\}) = 1$, $\psi^-(y) \in \mathbb{R}$ (Claim B1bis), and $\int_X c(x,y)\,\mathrm{d}\mu(x) < \infty$, so

$$\int_X \varphi^+\,\mathrm{d}\nu \le \int_X c(x,y)\,\mathrm{d}\mu(x) + \int_X \psi^-(y)\,\mathrm{d}\mu(x) < \infty.$$

Claim C3: $\int_{X\times Y} c(x,y)\,\mathrm{d}\eta < \infty$. We have

$$\begin{aligned}\int_{X\times Y} c\,\mathrm{d}\eta &= \int \Big(\varphi(x)+\psi(y)\Big)\,\mathrm{d}\eta(x,y)\\ &= \int \varphi\,\mathrm{d}\mu + \int \psi\,\mathrm{d}\nu\\ &\le \int \varphi^+\,\mathrm{d}\mu + \int \psi^+\,\mathrm{d}\nu < \infty.\end{aligned}$$

Claim C4: $\varphi \in L^1(\mu)$ and $\psi \in L^1(\nu)$. For $(x,y) \in S$ we have

$$\begin{aligned}\varphi(x) &= c(x,y)-\psi(y) \ge c(x,y)-\psi^+(y)\\ &\ge -c(x,y)-\psi^+(y),\end{aligned}$$

so $\varphi^-(x) \le c(x,y)+\psi^+(y)$. Hence

$$\begin{aligned}\int \varphi^-\,\mathrm{d}\mu &= \int \varphi^-(x)\,\mathrm{d}\eta(x,y) \le \int \Big(c(x,y)+\psi^+(y)\Big)\,\mathrm{d}\eta(x,y)\\ &= \int c\,\mathrm{d}\eta + \int \psi^+\,\mathrm{d}\nu < \infty.\end{aligned}$$

So $\int |\varphi|\,\mathrm{d}\mu \le \int \varphi^+\,\mathrm{d}\mu + \int \varphi^-\,\mathrm{d}\mu < \infty$. Similarly, $\int |\psi|\,\mathrm{d}\nu < \infty$.

Claim D1: $\varphi(x)+\psi(y) \le c(x,y)$ for all $(x,y) \in X\times Y$. We have

$$\psi(y) = \inf_{u\in X}\Big(c(u,y)-\varphi(u)\Big) \le c(x,y)-\varphi(x).$$

Claim D2: For $\gamma \in \Gamma(\mu,\nu)$,

$$\int_{X\times Y} c\,\mathrm{d}\gamma \;\ge\; \int_{X\times Y} \Big(\varphi(x)+\psi(y)\Big)\,\mathrm{d}\gamma(x,y)$$

$$
\begin{aligned}
&= \int \varphi(x)\,\mathrm{d}\mu(x) + \int \psi(y)\,\mathrm{d}\nu(y) \\
&= \int_{X\times Y} \Big(\varphi(x)+\psi(y)\Big)\,\mathrm{d}\eta(x,y) \\
&= \int_S \Big(\varphi(x)+\psi(y)\Big)\,\mathrm{d}\eta(x,y) \\
&= \int_S c\,\mathrm{d}\eta = \int_{X\times Y} c\,\mathrm{d}\eta.
\end{aligned}
$$

Conclusion: From D2 we see that η is optimal for c, that is,

$$\min\Big\{\int_{X\times Y} c\,\mathrm{d}\gamma\colon \gamma\in\Gamma(\mu,\nu)\Big\} = \int_{X\times Y} c\,\mathrm{d}\eta.$$

Further,

$$
\begin{aligned}
&\max\Big\{\int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\nu\colon f\in L^1(\mu),\ g\in L^1(\nu),\ f(x)+g(y)\le c(x,y)\ \forall (x,y)\in X\times Y\Big\} \\
&= \int \varphi(x)\,\mathrm{d}\mu(x) + \int \psi(y)\,\mathrm{d}\nu(y) = \int c\,\mathrm{d}\eta.
\end{aligned}
$$

Finally, we have $\psi=\varphi^c$ by definition of ψ. $\square$

Definition 2.8. The function φ as defined in (3) of the previous theorem for some choice of $(x_0,y_0)\in\operatorname{supp}\eta$ is called a *Kantorovich potential* associated to $\operatorname{supp}\eta$.

We give some more properties of Kantorovich potentials.

Proposition 2.9. *Let $\eta\in\Gamma(\mu,\nu)$ be optimal for c and let φ be a Kantorovich potential associated to* $\operatorname{supp}\eta$. *Then:*

(1) $\varphi(x)+\varphi(y)=c(x,y)$ *for all* $(x,y)\in\operatorname{supp}\eta$*;*

(2) $\varphi^{cc}(x)=\varphi(x)$ *for every* $x\in A := \{u\in X\colon \exists y\in Y$ *with* $(x,y)\in\operatorname{supp}\eta\}$.

Proof. (1): This fact has been proved in Claim B1 of the previous theorem.

(2): Let $x\in A$. We have

$$
\begin{aligned}
\varphi^{cc}(x) &= \inf_{y\in Y}\Big(c(x,y)-\varphi^c(y)\Big),\\
\varphi^{c}(y) &= \inf_{u\in X}\Big(c(u,y)-\varphi(u)\Big).
\end{aligned}
$$

Let $y\in Y$. Then $c(x,y)-\varphi^c(y)\ge c(x,y)-\Big(c(u,y)-\varphi(u)\Big)$ for all $u\in X$, so (with $u=x$) $c(x,y)-\varphi^c(y)\ge\varphi(x)$. Hence $\varphi^{cc}(x)\ge\varphi(x)$. Conversely, since $x\in A$, there exists a $y\in Y$ such that $\varphi^c(y)=c(x,y)-\varphi(x)$. Then $\varphi^{cc}(x)\le c(x,y)-\varphi^c(y)=c(x,y)-\Big(c(x,y)-\varphi(x)\Big)=\varphi(x)$. $\square$

Remark. The conditions in Theorem 2.7 that

- $\int c\,\mathrm{d}\eta<\infty$,

- $\mu\Big(\{x \in X\colon \int_Y c(x,y)\,\mathrm{d}\nu(y) < \infty\}\Big) > 0$, and
- $\nu\Big(\{y \in Y\colon \int_X c(x,y)\,\mathrm{d}\mu(x) < \infty\}\Big) > 0$,

are implied by the stronger condition that $\int c\,\mathrm{d}\mu \otimes \nu < \infty$, as is easily seen with the aid of Fubini.

2.4 Uniqueness and the Monge problem

We will now address the questions

- When is the optimal transportation plan η of the Kantorovich problem unique?
- When does the optimal η solve the Monge problem, that is, $\eta = (i \otimes r)_{\#}\mu$ for some Borel map $r : X \to Y$? (Recall, $i(x) = x$ for all $x \in X$.)

We begin with a sloppy sketch of the argument and then prove a preliminary version of a theorem on uniqueness and the Monge problem. Subsequently, we list some ingredients on convex functions and approximate differentiation and then prove a more general theorem.

We will consider the case that $X = Y = \mathbb{R}^d$ and $c(x,y) = h(x-y)$ for some strictly convex function h.

Suppose $\eta \in \Gamma(\mu,\nu)$ is optimal for c and that $c(x,y) = h(x-y)$ with h differentiable. We want to find a Borel map $r\colon X \to Y$ such that $\eta = (i \otimes r)_{\#}\mu$, that is, $\eta(W) = \mu(\{x\colon (x, r(x)) \in W\})$ for Borel sets $W \subseteq X \times Y$. In other words, we want to show that η is concentrated on the graph of a Borel map. We will try to find for each $x \in X$ a unique point y with (x,y) in the support of η.

Let φ be a Kantorovich potential associated to $\operatorname{supp}\eta$. For $(x,y) \in \operatorname{supp}\eta$ we have $\varphi(x) + \varphi^c(y) = h(x-y)$. Since $\varphi^c(y) = \inf_{u\in X}\Big(h(u-y) - \varphi(u)\Big) = h(x-y) - \varphi(x)$, the function $u \mapsto h(u-y) - \varphi(u)$ attains its minimum at $u = x$. Hence, *if φ is differentiable at x*,

$$\nabla h(x-y) = \nabla\varphi(x).$$

If $u \mapsto \nabla h(u)$ is invertible, we obtain $x - y = (\nabla h)^{-1}(\nabla\varphi(x))$, so

$$y = x - (\nabla h)^{-1}(\nabla\varphi(x)).$$

Hence for x such that φ is differentiable at x there is exactly one y with $(x,y) \in \operatorname{supp}\eta$. Thus we can take

$$r(x) := x - (\nabla h)^{-1}(\nabla\varphi(x)).$$

The main mathematical problems to make the argument work are the differentiability of φ and the Borel measurability of r. We will not be able to obtain everywhere differentiability of φ. Instead we will impose conditions that yield that φ is locally Lipschitz and then use Rademacher's theorem to conclude its Lebesgue almost everywhere differentiability. We need the map r at least μ-a.e. defined and therefore require that μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$.

Recall that a map $f : \mathbb{R}^d \to \mathbb{R}^m$ is called *differentiable* at $x \in \mathbb{R}^d$ if there exists a linear operator $L_x\colon \mathbb{R}^d \to \mathbb{R}^m$ such that for every $\varepsilon > 0$ there is a $\delta > 0$ with

$$\left|\frac{f(x+u) - f(u) - L_x u}{\|u\|}\right| < \varepsilon \quad \text{for all } u \in \mathbb{R}^d \text{ with } 0 < \|u\| < \delta.$$

If $m = 1$, then L_x is represented by a vector, which is denoted by $\nabla f(x)$, that is, $L_x u = \langle \nabla f(x), u \rangle$.

Denote the Lebesgue measure on $\mathbb{R}^d$ by $\mathcal{L}^d$.

Theorem 2.10 (Rademacher). *Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be locally Lipschitz. Then f is differentiable $\mathcal{L}^d$-almost everywhere. Moreover, $D = \{x \in \mathbb{R}^d\colon f \text{ differentiable at } x\}$ is a Borel set and*

$$x \mapsto \begin{cases} \nabla f(x) & \text{if } x \in D \\ 0 & \text{otherwise} \end{cases}$$

is a Borel map from $\mathbb{R}^d$ to $\mathbb{R}^d$.

Now we are in a position to prove a theorem on uniqueness for the Kantorovich problem and existence for the Monge problem. More sophisticated statements are given in Theorem 2.21.

Theorem 2.11. *Consider $X = Y = \mathbb{R}^d$. Let $c(x, y) = h(x - y)$, $x, y \in \mathbb{R}^d$, where $h\colon \mathbb{R} \to [0, \infty)$ is differentiable, locally Lipschitz, and such that ∇h from $\mathbb{R}^d$ to its range is bijective with a Borel measurable inverse. Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be such that*

- $\int_{X \times Y} h(x - y) \, \mathrm{d}\gamma(x, y) < \infty$ *for some* $\gamma \in \Gamma(\mu, \nu)$,
- $\mu\Big(\Big\{x \in X\colon \int_Y h(x - y) \, \mathrm{d}\nu(y) < \infty\Big\}\Big) > 0$,
- $\nu\Big(\Big\{y \in Y\colon \int_X h(x - y) \, \mathrm{d}\mu(x) < \infty\Big\}\Big) > 0$,

and such that

- *μ is absolutely continuous with respect to $\mathcal{L}^d$,*
- $\operatorname{supp} \nu$ *is bounded.*

Then:

1. *there is a* unique *$\eta \in \Gamma(\mu, \nu)$ that is optimal for c;*
2. *η is induced by an optimal transport map, that is, there exists a Borel map $r\colon \mathbb{R}^d \to \mathbb{R}^d$ such that $\eta = (i \otimes r)_\# \mu$;*
3. *the map r of (2) satisfies*

$$r(x) = x - (\nabla h)^{-1}(\nabla \varphi^{cc}(x)) \quad \text{for } \mu\text{-a.e. } x \in \mathbb{R}^d,$$

where φ is a Kantorovich potential associated to $\operatorname{supp} \eta$.

Proof. Let $A_1 \subseteq X$ be a μ-full Borel set such that for all $x \in A_1$ there is a $y \in Y$ such that $(x, y) \in \operatorname{supp} \eta$, $\varphi^{cc}(x) = \varphi(x)$ (use Proposition 2.9), and $\varphi(x) \in \mathbb{R}$.

Take $R > 0$ such that $\operatorname{supp} \nu \subset B(0, R)$. Then for $x \in A_1$,

$$\varphi(x) = \inf_{y \in B(0,R)} \Big(c(x, y) - \varphi^c(x)\Big),$$

since

$$\begin{aligned}\varphi(x) &= \varphi^{cc}(x) = \inf_{y\in\mathbb{R}^d}\Big(c(x,y) - \varphi^c(y)\Big)\\ &\le \inf_{y\in B(0,R)}\Big(c(x,y)-\varphi^c(y)\Big) = \varphi(x)\end{aligned}$$

Next we show that φ^{cc} is locally Lipschitz. Indeed, Let $r > 0$. Then h is L-Lipschitz on $B(0, R + r)$ for some L, since h is Locally Lipschitz. Let $x_1, x_2 \in B(0, r)$. Let $\varepsilon > 0$ and choose $y \in B(0,R)$ such that $\varphi^c(y) > -\infty$ (we have $\varphi^c \in L^1(\nu)$) and $\varphi^{cc}(x_2) \ge c(x, y) - \varphi^c(y) - \varepsilon$. Then

$$\begin{aligned}\varphi^{cc}(x_1) - \varphi^{cc}(x_2) &\le c(x_1,y) - \varphi^c(y) - \Big(c(x_2,y) - \varphi^c(y) - \varepsilon\Big)\\ &= h(x_1 - y) - h(x_2 - y) + \varepsilon\\ &\le L\|x_1 - x_2\| + \varepsilon,\end{aligned}$$

as $x_i + y \in B(0, r + R)$, end hence

$$\varphi^{cc}(x_1) - \varphi^{cc}(x_2) \le L\|x_1 - x_2\|.$$

Thus, by interchanging the role of x_1 and x_2, $|\varphi^{cc}(x_1) - \varphi^{cc}(x_2)| \le L\|x_1 - x_2\|$.

Let A_2 be an $\mathcal{L}^d$-full Borel set such that φ^{cc} is differentiable at every $x \in A_2$ (by Rademacher's theorem). Then A_2 is also μ-full, as μ is absolutely continuous with rspect to $\mathcal{L}^d$. Let $A := A_1 \cap A_2$. Then A is a μ-full Borel set and for every $x \in A$ we have

- there is a $y \in \mathbb{R}^d$ with $(x, y) \in \operatorname{supp}\eta$ and therefore $y \in B(0,R)$ and $\varphi(x) + \varphi^c(y) = h(x - y)$,
- $\varphi(x) = \varphi^{cc}(x)$,
- $\varphi(x) \in \mathbb{R}$ and φ^{cc} is differentiable at x.

Let $x \in A$. There exists y such that $(x, y) \in \operatorname{supp}\eta$. Consider such a y. The function $u \mapsto h(u - y) - \varphi^{cc}(u)$ then attains its minimum $\varphi^c(y)$ at $u = x$ and is differentiable at x. So

$$\nabla h(x - y) - \nabla\varphi^{cc}(x) = 0.$$

Hence $\nabla\varphi(x)$ is in the range of ∇h and $x - y = (\nabla h)^{-1}(\nabla\varphi^{cc}(x))$, so

$$y = x - (\nabla h)^{-1}(\nabla\varphi^{cc}(x)). \tag{4}$$

Define

$$r(x) := \begin{cases} x - (\nabla h)^{-1}(\nabla\varphi^{cc}(x)) & x \in A\\ 0 & x \notin A.\end{cases}$$

Due to Rademacher's theorem and the assumptions on h, we infer that $r\colon \mathbb{R}^d \to \mathbb{R}^d$ is a Borel map. Moreover, we have $(x, r(x)) \in \operatorname{supp}\eta$ for all $x \in A$, as follows from (4). Further, in the arguments preceding (4) y is an arbitrary element of $\mathbb{R}^d$ with $(x, y) \in \operatorname{supp}\eta$ and thus we obtain that for $x \in A$,

$$(x, y) \in \operatorname{supp}\eta \quad \Longleftrightarrow y = r(x).$$

Consequently,

$$\eta(\{x \in A\colon (x, r(x))\}) = \eta(\operatorname{supp}\eta \setminus (A \times Y)) = 1.$$

Next we claim that

$$\eta = (i \otimes r)_{\#}\mu.$$

For a proof, let $U \times V \subseteq \mathbb{R}^d \times \mathbb{R}^d$ with $U \subseteq \mathbb{R}^d$ and $V \subseteq \mathbb{R}^d$ Borel. Then

$$\begin{aligned} \eta(U \times V) &= \eta\Big(U \times V \cap \{(x, r(x))\colon x \in X\}\Big) \\ &= \eta\Big((U \cap \{x\colon r(x) \in V\}) \times Y\Big) \\ &= \mu(U \cap \{x\colon (x, r(x)) \in U \times V\}\Big) \\ &= (i \otimes r)_{\#}\mu(U \times V). \end{aligned}$$

.

Finally, we address uniqueness of η. Suppose $\eta_1, \eta_2 \in \Gamma(\mu, \nu)$ are both optimal for c. Then also $\eta := \frac{1}{2}\eta_1 + \frac{1}{2}\eta_2 \in \Gamma(\mu, \nu)$ is optimal for c. By the first part of the proof given above, we obtain Borel maps $r_1, r\colon \mathbb{R}^d \to \mathbb{R}^d$ such that

$$\eta_1 = (i \otimes r_1)_{\#}\mu \quad \text{and} \quad \eta = (i \otimes r)_{\#}\mu.$$

As η_1 is absolutely continuous with respect to η, we have

$$\eta_1\Big(\Big\{(x, r(x))\colon x \in \mathbb{R}^d\Big\}\Big) = 1.$$

Then

$$\eta_1\Big(\Big\{(x, r(x))\colon x \in \mathbb{R}^d\Big\} \cap \Big\{(x, r_1(x))\colon x \in \mathbb{R}^d\Big\}\Big) = 1,$$

so

$$\eta_1\Big(\Big\{(x, r(x))\colon x \in \mathbb{R}^d, r(x) = r_1(x)\Big\}\Big) = 1.$$

Hence $r = r_1$ μ-a.e. and, consequently, $\eta_1 = (i \otimes r_1)_{\#}\mu = (i \otimes r)_{\#}\mu = \eta$. Therefore $\eta_1 = \eta_2$. □

The conditions in the previous theorem can be relaxed. In particular the condition that the support of ν be bounded and the differentiability of h. An interesting setting is where h is strictly convex. The extension of the result requires some preliminaries on convex functions and approximate differentiability.

Definition 2.12. A set $S \subseteq \mathbb{R}^d$ is said to have *density* 1 at $x \in \mathbb{R}^d$ if there exists a Lebesgue measurable set $V \subseteq S$ with

$$\lim_{\delta \downarrow 0} \frac{\mathcal{L}^d(V \cap B(x, \delta))}{\mathcal{L}^d(B(x, \delta))} = 1,$$

where $\mathcal{L}^d$ denotes the Lebesgue measure on $\mathbb{R}^d$.

Definition 2.13. Let $\Omega \subseteq \mathbb{R}^d$ be open and let $f\colon \Omega \to \mathbb{R}^m$. Let $x \in \Omega$.

(1) A point $z \in \mathbb{R}^m$ is called the *approximate limit* of f at x if for every $\varepsilon > 0$ the set $\{y \in \Omega\colon |f(y) - z| \leq \varepsilon\}$ has density 1 at x; notation:

$$\tilde{f}(x) := f(y).$$

(2) f is *approximately continuous* at x if $\tilde{f}(x)$ exists and $f(x) = \tilde{f}(x)$.

(3) a linear map $L\colon \mathbb{R}^d \to \mathbb{R}^m$ is the *approximate differential* of f at x if $\tilde{f}(x)$ exists and for every $\varepsilon > 0$ the set

$$\Big\{y \in \Omega \setminus \{x\}\colon \frac{\|f(y) - \tilde{f}(x) - L(y-x)\|}{\|y-x\|} \le \varepsilon\Big\}$$

has density 1 at x. (There is at most one such an L.) Notation: $\tilde{\nabla} f(x) := L$.

Lemma 2.14. *Let $\Omega \subseteq \mathbb{R}^d$ be open, $f\colon \Omega \to \mathbb{R}^m$, and $x \in \Omega$.*

(1) *$\tilde{f}(x)$ exists if and only if there exists a $g\colon \Omega \to \mathbb{R}^m$ which is continuous at x and such that $\{f = g\} := \{x \in \Omega\colon f(x) = g(x)\}$ has density 1 at x.*

(2) *$\tilde{\nabla} f(x)$ exists if and only if there exists a $g\colon \Omega \to \mathbb{R}^d$ which is differentiable at x and such that $\{f = g\}$ has density 1 at x.*

Theorem 2.15 (Denjoy). *If $\Omega \subseteq \mathbb{R}^d$ is open and $f\colon \Omega \to \mathbb{R}^m$ is Lebesgue measurable, then $\tilde{f}$ exists $\mathcal{L}^d$-almost everywhere on Ω and $f = \tilde{f}$ $\mathcal{L}^d$-almost everywhere on Ω.*

Corollary 2.16 (Lebesgue). *Every Lebesgue messurable set S of $\mathbb{R}^d$ has density 1 at $\mathcal{L}^d$-almost every point of S.*

Definition 2.17. Let X be a vector space. A function $h\colon X \to \mathbb{R}$ is called *strictly convex* if

$$h(\lambda x + (1-\lambda)y) < \lambda h(x) + (1-\lambda)h(y) \text{ for all } x, y \in X,\ x \neq y,\ 0 < \lambda < 1,$$

and *convex* if we have $\le$ in the above inequality instead of $<$.

A simple proof of the next lemma is given in [12].

Lemma 2.18. *If $h\colon \mathbb{R}^d \to \mathbb{R}$ is convex, then h is locally Lipschitz.*

Definition 2.19. Define for a convex function $h\colon \mathbb{R}^d \to \mathbb{R}$,

$$\partial h(x) := \{\xi \in \mathbb{R}^d\colon h(u) - h(x) \ge \langle \xi, u - x\rangle \text{ for all } u \in \mathbb{R}^d\}, \quad x \in \mathbb{R}^d.$$

Lemma 2.20. *Let $h\colon \mathbb{R}^d \to \mathbb{R}$ be strictly convex. Then:*

(1) *if h is differentiable at $x \in \mathbb{R}^d$ then for every $\xi \in \mathbb{R}^d$ we have*

$$\xi \in \partial h(x) \quad \Longleftrightarrow \quad \xi = \nabla h(x);$$

(2) *for every $\xi \in \mathbb{R}^d$ there is at most one $x \in \mathbb{R}^d$ such that $\xi \in \partial h(x)$; denote $(\partial h)^{-1}(\xi) := x$ with domain $D((\partial h)^{-1})$;*

(3) *$(\partial h)^{-1}\colon D((\partial h)^{-1}) \to \mathbb{R}^d$ is a Borel map.*

Proof. (1): $\Rightarrow$: For $z \in \mathbb{R}^d$ we have (with $u = x + tz$ and $u = x - tz$)

$$h(x+tz) - h(x) \geq t\langle \xi, z\rangle \text{ and } h(x-tz) - h(x) \geq -t\langle \xi, z\rangle,$$

so

$$\frac{h(x) - h(x-tz)}{t} \leq \langle \xi, z\rangle \leq \frac{h(x+tz) - h(x)}{t}$$

for $t > 0$. If we let $t \downarrow 0$ we obtain $\langle \nabla h(x), z\rangle = \langle \xi, z\rangle$.

$\Leftarrow$: Let $u \in \mathbb{R}^d$. We have

$$\begin{aligned} \frac{h(x + t(u-x)) - h(x)}{t} &= \frac{h((1-t)x + tu) - h(x)}{t} \\ &\leq \frac{1}{t}(1-t)h(x) + h(u) - \frac{1}{t}h(x) \\ &= h(u) - h(x), \end{aligned}$$

so as $t \downarrow 0$, $\langle \nabla h(x), u - x\rangle \leq h(u) - h(x)$.

(2): Suppose $\xi \in \partial h(x_1)$ and $\xi \in \partial h(x_2)$ and $x_1 \neq x_2$. With $u = \lambda x_2 + (1-\lambda)x_1$, $0 < \lambda < 1$, we have $\lambda h(x_2) + (1-\lambda)h(x_1) > h(u)$, so

$$\begin{aligned} \lambda\Big(h(x_2) - h(x_1)\Big) &= \lambda h(x_2) + (1-\lambda)h(x_1) - h(x_1) > h(u) - h(x_1) \\ &\geq \langle \xi, u - x_1\rangle = \lambda\langle \xi, x_2 - x_1\rangle. \end{aligned}$$

Similarly, $h(x_1) - h(x_2) > \langle \xi, x_1 - x_2\rangle$, which yields a contradiction.

(3): Let $R > 0$ and consider $h\colon B(0,R) \to \mathbb{R}$. Let

$$S_R := \{\xi \in \mathbb{R}^d \colon \exists x \in B(0,R) \text{ such that } \xi \in \partial h(x)\}.$$

Then $(\partial h)^{-1}$ maps S_R into $B(0,R)$. Its graph

$$G_R := \{(\xi, x) \in \mathbb{R}^d \times B(0,R) \colon \xi \in \partial h(x)\}$$

is closed. Indeed, if $(\xi_n, x_n) \in G_R$ and $\xi_n \to \xi$ and $x_n \to x$, then $x \in B(0,R)$,

$$h(u) - h(x_n) \geq \langle \xi_n, u - x_n\rangle \text{ for all } u \in \mathbb{R}^d,$$

so $h(u) - h(x) \geq \langle \xi, u - x\rangle$ for all $u \in \mathbb{R}^d$ (as h is locally Lipschitz), hence $\xi \in \partial h(x)$ and therefore $(\xi, x) \in G_R$.

Consequently, $(\partial h)^{-1}\colon S_R \to B(0,R)$ is continuous. Indeed, if $\xi_n \to \xi$ in S_R, let $(\xi_{n_k})_k$ be a subsequence. Then $x_{n_k} := (\partial h)^{-1}(\xi_{n_k})$ is a sequence in $B(0,R)$, which is compact. Hence $x_{n_{k_\ell}} \to x$ in $B(0,R)$ for some subsubsequence and some $x \in B(0,R)$. As G_R is closed, $(\xi, x) \in G_R$, hence $x = (\partial h)^{-1}(\xi)$. So $(\partial h)^{-1}(\xi_{n_{k_\ell}}) \to (\partial h)^{-1}(\xi)$. That is, every subsequence of $((\partial h)^{-1}(\xi_n))_n$ has a subsubsequence that converges to $(\partial h)^{-1}(\xi)$. Hence $(\partial h)^{-1}(\xi_n) \to (\partial h)^{-1}(\xi)$.

Thus $(\partial h)^{-1}$ is continuous on S_R and hence Borel on S_R. Since $\bigcup_N S_N$ equals the domain of $(\partial h)^{-1}$, $(\partial h)^{-1}$ is Borel (just look at inverse images). $\square$

Theorem 2.21. *Consider $X = Y = \mathbb{R}^d$. Let $c(x,y) = h(x-y)$, $x, y \in \mathbb{R}^d$, where $h\colon \mathbb{R}^d \to [0,\infty)$ is strictly convex. Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be such that*

- $\int_{X\times Y} c(x,y)\,\mathrm{d}\gamma(x,y) < \infty$ *for some* $\gamma \in \Gamma(\mu,\nu)$,
- $\mu\Big(\Big\{x \in X\colon \int_Y c(x,y)\,\mathrm{d}\nu(y) < \infty\Big\}\Big) > 0$,
- $\nu\Big(\Big\{y \in Y\colon \int_X c(x,y)\,\mathrm{d}\mu(x) < \infty\Big\}\Big) > 0$,

and such that

- μ *is absolutely continuous with respect to* $\mathcal{L}^d$.

Then:

(1) *there is a* unique $\eta \in \Gamma(\mu,\nu)$ *that is optimal for c;*

(2) η *is induced by an optimal transport map, that is, there exists a Borel map* $r\colon \mathbb{R}^d \to \mathbb{R}^d$ *such that* $\eta = (i \otimes r)_{\#}\mu$*;*

(3) *the map* r *of* (2) *satisfies*

$$r(x) = x - (\partial h)^{-1}(\tilde{\nabla}\varphi(x)) \quad \textit{for } \mu\textit{-a.e. } x \in \mathbb{R}^d,$$

where φ *is a Kantorovich potential associated to* $\operatorname{supp}\eta$.

Proof. Let $A_1 \subseteq X$ be a μ-full Borel set such that for all $x \in A_1$ there exists a $y \in \operatorname{supp}\eta$ and $\varphi^{cc}(x) = \varphi(x)$.

Step 1. For $R > 0$ so large that there exists a $y \in B(0,R)$ with $\varphi^c(z) > \infty$, define

$$\varphi_R(x) := \inf_{y \in B(0,R)} \Big(c(x,y) - \varphi^c(y)\Big), \quad x \in \mathbb{R}^d.$$

Then $\varphi_R(x) < \infty$ and $\varphi_R(x) \geq \varphi^{cc}(x) = \varphi(x) > -\infty$ for all $x \in A_1$.

Claim: φ_R is locally Lipschitz. Let $r > 0$ and choose $L > 0$ such that h is L-Lipschitz on $B(0, r+R)$. Let $x_1, x_2 \in B(0,r)$. Let $\varepsilon > 0$ and choose $y \in B(0,R)$ such that $\varphi^c(y) > -\infty$ and $\varphi_R(x_2) > c(x,y) - \varphi^c(y) - \varepsilon$. Then

$$\begin{aligned} \varphi_R(x_1) - \varphi_R(x_2) &\leq c(x_1,y) - \varphi^c(y) - \Big(c(x_2,y) - \varphi^c(y) - \varepsilon\Big) \\ &= h(x_1 - y) - h(x_2 - y) + \varepsilon \\ &\leq L\|x_1 - x_2\| + \varepsilon. \end{aligned}$$

So $\varphi_R(x_1) - \varphi(x_2) \leq L\|x_1 - x_2\|$. Thus $|\varphi_R(x_1) - \varphi_R(x_2)| \leq L\|x_1 - x_2\|$.

Let $A_2 \subseteq X$ be a $\mathcal{L}^d$-full hence μ-full Borel set such that φ_N is differentiable at every $x \in A_2$ for all $N \in \mathbb{N}$ (by Rademacher's theorem).

Claim: For every $x \in A_1$ there exists an $R_0 > 0$ such that $x \in \{\varphi = \varphi_R\}$ and there exists a $y \in B(0,R)$ such that $\varphi(x) + \varphi^c(y) = h(x-y)$ for all $R \geq R_0$. Let $x \in A_1$. Then $\varphi(x) = \varphi^{cc}(x)$ and there exists a $y \in \mathbb{R}^d$ such that $\varphi(x) + \varphi^c(y) = h(x-y)$. So for $R > \|y\|$,

$$\varphi(x) = \varphi^{cc}(x) \leq \varphi_R(x) \leq h(x-y) - \varphi^c(y) = \varphi(x),$$

so $x \in \{\varphi = \varphi_R\}$.

Step 2. Claim: There exists a μ-full Borel set $A \subseteq \mathbb{R}^d$ such that for all $x \in A$ there exists an N such that

- $x \in \{\varphi = \varphi_N\}$
- there exists $y \in B(0, N)$ with $\varphi_N(x) + \varphi^c(y) = h(x - y)$
- φ_N is differentiable at x for all $n \in \mathbb{N}$, φ is approximately differentiable at x, and $\tilde{\nabla}\varphi(x) = \nabla\varphi_N(x)$.

Indeed, as φ is Borel and φ_N is locally Lipschitz, $\{\varphi = \varphi_N\}$ is a Lebesgue measurable set. Due to Lebesgue's theorem (or Denjoy), $\{\varphi = \varphi_N\}$ has density 1 at $\mathcal{L}^d$-a.e. point of $\{\varphi = \varphi_N\}$. Let

$$T_N := \{x \in \mathbb{R}^d \colon \{\varphi = \varphi_N\} \text{ has density 1 at } x\}.$$

Then $\bigcap_{N \in \mathbb{N}} T_N$ is a μ-full Lebesgue set. Hence there is a μ-full Borel set $A_3 \subseteq X$ with $A_3 \subseteq \bigcap_{N \in \mathbb{N}} T_N$ (use [8, Theorem 13.B, p. 55]). Let

$$A := A_1 \cap A_2 \cap A_3.$$

For $x \in A$, choose N with the first two properties. As $x \in A_2$, φ_N is differentiable at x and $\{\varphi = \varphi_N\}$ has density 1 as $x \in A_3$. It follows that φ is approximately differentiable at x and $\tilde{\nabla}\varphi(x) = \nabla\varphi_N(x)$.

Step 3. Let $x \in A$. Take N and $y \in B(0, N)$ such that $x \in \{\varphi = \varphi_N\}$ and $\varphi_N(x) + \varphi^c(y) = h(x - y)$. As φ_N is differentiable at x, h is differentiable at $x - y$ and $u \mapsto h(x - y) - \varphi_N(u)$ attains its minimum $\varphi^c(y)$ at $u = x$. So $\nabla\varphi_N(x) = \nabla h(x - y)$. Hence $\tilde{\nabla}(x) = \nabla\varphi_N(x)$ is in the domain of $(\partial h)^{-1}$ and $(\partial h)^{-1}(\nabla\varphi_N(x)) = x - y$. Moreover, for any $y \in Y$ with $\varphi_N(x) + \varphi^c(y) = h(x - y)$ we find

$$y = x - (\partial h)^{-1}(\tilde{\nabla}\varphi(x)).$$

Define

$$r(x) := \begin{cases} x - (\partial h)^{-1}(\tilde{\nabla}\varphi(x)) & x \in A \\ 0 & x \notin A. \end{cases}$$

Claim: r is Borel. The set $V_N := \{x \in V \colon \varphi(x) = \varphi_N(x)\}$ is Borel and $\bigcup_N V_N = V$. On V_N, $\tilde{\nabla}\varphi = \nabla\varphi_N$. As φ_N is locally Lipschitz, $x \mapsto \nabla\varphi_N(x)$ is Borel, so $x \mapsto \tilde{\nabla}\varphi(x)$ is Borel. Thus, r is Borel.

The rest of the proof is as before in Theorem 2.11. □

2.5 Regularity of the optimal transport map

Next we will show a (very weak) regularity theorem for the optimal transport map r. We begin with two lemmas.

Lemma 2.22. *Let $g \colon \mathbb{R}^d \to \mathbb{R}$ be differentiable and $\nabla g \colon \mathbb{R}^d \to \mathbb{R}^d$ Lipschitz with Lipschitz constant C. Then*

$$x \mapsto g(x) - \frac{C}{2}\|x\|^2$$

is concave (that is, $-g + \frac{C}{2}\|\cdot\|^2$ is convex).

Proof. Let $x \in \mathbb{R}^d$ and $u \in \mathbb{R}^d$. We show that

$$f(t) := g(x + tu) - \frac{C}{2}\|x + tu\|^2$$

is concave. We have $f'(t) = \langle \nabla g(x+tu), u\rangle - C\langle x+tu, u\rangle$. Hence for $t_1 < t_2$,

$$\begin{aligned} f'(t_2) - f'(t_1) &= \langle \nabla g(x+t_2u) - \nabla g(x+t_1u), u\rangle - C\langle (t_2-t_1)u, u\rangle \\ &\leq \|\nabla g(x+t_2u) - \nabla g(x+t_1u)\|\,\|u\| - C(t_2-t_1)\|u\|^2 \\ &\leq C(t_2-t_1)\|u\|^2 - C(t_2-t_1)\|u\|^2 = 0, \end{aligned}$$

so f' is decreasing, hence f concave. □

Lemma 2.23 (Aleksandrov). *Let $g\colon \mathbb{R}^d \to \mathbb{R}$ be convex. Then ∇g exists $\mathcal{L}^d$-a.e. and ∇g is differentiable in $\mathcal{L}^d$-a.e. point of its domain.*

Theorem 2.24. *Consider $X = Y = \mathbb{R}^d$, $c(x,y) = h(x-y)$, $x, y \in \mathbb{R}^d$, where $h\colon \mathbb{R} \to [0,\infty)$ is differentiable, locally Lipschitz, and such that ∇h from $\mathbb{R}^d$ to its range is bijective, locally Lipschitz, and with a differentiable inverse (e.g., $h(x) = \|x\|^2$). Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be such that*

$$\int_{X\times Y} h(x-y)\,\mathrm{d}\mu\otimes\nu(x,y) < \infty$$

and such that

- *μ is absolutely continuous with respect to $\mathcal{L}^d$*
- *$\operatorname{supp}\nu$ is bounded.*

Let $\eta \in \Gamma(\mu,\nu)$ be optimal for c and $r\colon \mathbb{R}^d \to \mathbb{R}^d$ a Borel map such that $\eta = (i\otimes r)_{\#}\mu$. Then r is μ-a.e. equal to a μ-a.e. differentiable function.

Proof. Let φ be a Kantorovich potential associated to $\operatorname{supp}\eta$. Let $R > 0$ be such that $\operatorname{supp}\nu \subseteq B(0,R)$. Let $A_1 \subseteq \mathbb{R}^d$ be a μ-full set such that

$$\forall x \in \mathbb{R}^d\ \exists y \in \mathbb{R}^d \text{ such that } (x,y) \in \operatorname{supp}\eta \text{ and } \varphi^{cc}(x) = \varphi(x).$$

Then

$$\varphi^{cc}(x) = \inf_{y\in B(0,R)} \Big(h(x-y) - \varphi^c(y)\Big) = \varphi(x).$$

Further, we have that φ^{cc} is locally Lipschitz hence μ-a.e. differentiable and

$$r(x) = x - (\nabla h)^{-1}(\tilde{\nabla}\varphi(x)) \quad \text{ for } \mu\text{-a.e. } x \in \mathbb{R}^d.$$

By assumption, $(\nabla h)^{-1}$ is differentiable. It remains to show that $\nabla\varphi^{cc}$ is differentiable.

Let $r > 0$ and let C be the Lipschitz constant of ∇h on $B(0, r+R)$. Then, by the lemma, $x \mapsto h(x-y) - \varphi^c(y) - \frac{C}{2}\|x\|^2$ is concave on $B(0,r)$ for all $y \in B(0,R)$. Hence $x \mapsto \varphi^{cc}(x) - \frac{2}{2}\|x\|^2$ is concave on $B(0,R)$, since it is an infimum of concave functions.

Consequently, by Aleksandrov, $\nabla\varphi^{cc}$ is differentiable at $\mathcal{L}^d$-a.e. point of its domain and hence at μ-a.e. point of its domain. Thus $x \mapsto x - (\nabla h)^{-1}(\nabla\varphi^{cc}(x))$ is μ-a.e. differentiable and r equals this map μ-a.e. □

References

[1] Ambrosio, L., Gigli, N., and Savaré, G., *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics, ETH Zürich, Birkhäuser Verlag, Basel - Boston - Berlin, 2005.

[2] Billingsley, Patrick, *Convergence of Probability Measures*, Wiley & Sons, New York - London, 1968.

[3] Bourbaki, N., *Éléments de mathématique. Livre 6. Intégration. Chapitre 9*, Hermann, Paris, 1969.

[4] Dudley, R.M., Convergence of Baire measures, *Studia Math.* **27** (1966), 251–268.

[5] Dudley, R.M., *Real analysis and probability*, Cambridge Studies in Advanced Mathematics 74, Cambridge University Press, Cambridge, 2002.

[6] Dunford, Nelson, and Schwartz, Jacob T., *Linear Operators. Part I: General Theory*, Wiley-Interscience, New York, 1957.

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

[8] Halmos, Paul R., *Measure Theory*, Graduate Texts in Mathematics 18, Springer, New York - Berlin - Heidelberg, 1974.

[9] Kelley, John L., *General Topology*, Graduate Texts in Mathematics 27, Springer, New York - Heidelberg - Berlin, 1975.

[10] Parthasarathy, K.R., *Probability Measures on Metric Spaces,* Academic Press, New York - London, 1967.

[11] Prokhorov, Yu.V., Convergence of random processes and limit theorems in probability theory, *Theor. Prob. Appl.* **1** (1956), 157–214.

[12] Roberts, A.W. and Varberg, D.E., Another proof that convex functions are locally Lipschitz, *The American Mathematical Monthly* **81** (1974) No. 9, 1012–1016.

[13] Rudin, Walter, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1987.

[14] Schechter, Martin, *Principles of Functional Analysis*, 2nd ed., Graduate Studies in Mathematics 36, American Mathematical Society, Providence, Rhode Island, 2002.

[15] van Gaans, Onno, Probability measures on metric spaces, Notes of the seminar 'Stochastic Evolution Equations', Delft University of Technology, 2003.