Chapter 5

Preliminary Results on Measure Theory

In this chapter we introduce, mostly without proofs, some basic measure-theoretic tools needed in the next chapters. We decided to present the most significant result in the quite general framework of *separable metric spaces* in view of possible applications to infinite dimensional Hilbert (or Banach) spaces, thus avoiding any local compactness assumption (we refer to the treatises [126, 71, 72, 136, 67] for comprehensive presentations of this subject).

At this preliminary level, the existence of an equivalent complete metric (Polish spaces) only enters in the *compact inner regularity* (5.1.9) or *tightness* (5.1.8) of every Borel measure (it is a consequence of Ulam's Theorem [72, 7.1.4], a particular case of the converse implication in Prokhorov Theorem 5.1.3), which in particular appears in the so called disintegration theorem 5.3.1 and its consequences; this inner approximation condition is satisfied by a wider class of even non complete metric spaces (the so called *Radon spaces* [136, page 117]) and it will be sufficient for our aims. Since weak topologies in Hilbert-Banach spaces are not metrizable, it will also be useful (see Lemma 5.1.12) to deal with auxiliary non complete metrics, still satisfying (5.1.9).

Even if the presentation looks more abstract and the assumptions very weak with respect to the more usual finite dimensional Euclidean setting of the standard theory for evolutionary PDE's, this approach is sufficiently powerful to provide all the crucial results and allows for a great flexibility.

Let X be a separable metric space. We denote by $\mathscr{B}(X)$ the family of the Borel subsets of X, by $\mathscr{P}(X)$ the family of all Borel probability measures on X. The *support* $\operatorname{supp}\mu \subset X$ of $\mu \in \mathscr{P}(X)$ is the closed set defined by

$$\operatorname{supp}\mu := \Big\{x \in X : \mu(U) > 0 \quad \text{for each neighborhood } U \text{ of } x\Big\}. \tag{5.0.1}$$

When $\boldsymbol{X} = X_1 \times \ldots \times X_k$ is a product space, we will often use *bold* letters to indicate Borel measures $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$. Recall that for separable metric spaces $X_1, \ldots, X_k$ the Borel σ-algebra coincides with the product one

$$\mathscr{B}(\boldsymbol{X}) = \mathscr{B}(X_1) \times \mathscr{B}(X_2) \times \cdots \times \mathscr{B}(X_k). \tag{5.0.2}$$

5.1 Narrow convergence, tightness, and uniform integrability

Conformally to the probabilistic terminology, we say that a sequence $(\mu_n) \subset \mathscr{P}(X)$ is *narrowly* convergent to $\mu \in \mathscr{P}(X)$ as $n \to \infty$ if

$$\lim_{n\to\infty} \int_X f(x)\, d\mu_n(x) = \int_X f(x)\, d\mu(x) \tag{5.1.1}$$

for every function $f \in C_b^0(X)$, the space of continuous and bounded real functions defined on X.

Of course, it is sufficient to check (5.1.1) on any subset $\mathscr{C}$ of bounded continuous functions whose linear envelope $\operatorname{span}\mathscr{C}$ is uniformly dense (i.e. dense in the uniform topology induced by the "sup" norm) in $C_b^0(X)$. Even better, let us suppose that $\mathscr{C}_0 \subset C_b^0(X)$ satisfies the approximation properties

$$\int_X f(x)\, d\mu(x) = \sup\Big\{ \int_X h(x)\, d\mu(x) :\ h \in \mathscr{C}_0,\ h \le f \Big\} \tag{5.1.2a}$$

$$= \inf\Big\{ \int_X h(x)\, d\mu(x) :\ h \in \mathscr{C}_0,\ h \ge f \Big\}, \tag{5.1.2b}$$

for every $f \in \mathscr{C}$; then if (5.1.1) holds for every $f \in \mathscr{C}_0$, then it holds for every continuous and bounded function f. In fact for every $f \in \mathscr{C}$ we easily have

$$\begin{aligned}\liminf_{n\to\infty} \int_X f(x)\, d\mu_n(x) &\ge \sup_{h\in\mathscr{C}_0, h\le f} \liminf_{n\to\infty} \int_X h(x)\, d\mu_n(x) \\ &= \sup_{h\in\mathscr{C}_0, h\le f} \int_X h(x)\, d\mu(x) = \int_X f(x)\, d\mu(x),\end{aligned} \tag{5.1.3}$$

and the opposite inequality for the "lim sup" can be obtained in a similar way starting from (5.1.2b). Thus every $f \in \mathscr{C}$ satisfies (5.1.1), and we get the same property for every $f \in C_b^0(X)$ since $\operatorname{span}\mathscr{C}$ is uniformly dense in $C_b^0(X)$.

If d is any metric for X, the subset of d-uniformly (or d-Lipschitz) continuous and bounded real functions provides an important example [138, Th. 3.1.5] satisfying (5.1.2a,b). For, we can pointwise approximate a continuous and bounded function f from below with an increasing sequence of bounded Lipschitz functions f_k (they are particular examples of the Moreau-Yosida approximations for the exponent $p = 1$, see Section 3.1)

$$f_k(x) := \inf_y f(y) + kd(x,y), \quad \text{with} \quad \begin{cases} \inf f \le f_k(x) \le f(x) \le \sup f, \\ f(x) = \lim_{k\to\infty} f_k(x) = \sup_{k\in\mathbb{N}} f_k(x), \end{cases} \tag{5.1.4}$$

thus obtaining (5.1.2a) by Fatou's lemma; changing f to $-f$ we obtain (5.1.2b).

A slight refinement of this argument provides a *countable* set of d-Lipschitz functions satisfying (5.1.2a,b) for every function $f \in C_b^0(X)$: we simply choose a countable dense set $D \subset X$ and we consider the countable family of functions $h : X \to \mathbb{R}$ of the type

$$\begin{gathered} h(x) = \big(q_1 + q_2 d(x,y)\big) \wedge k \\ \text{for some } q_1, q_2, k \in \mathbb{Q}, \quad q_2,\, k \in (0,1),\ y \in D. \end{gathered} \tag{5.1.5a}$$

We denote by $\mathscr{C}_1$ the collection generated from this set by taking the infimum of a finite number of functions, thus satisfying

$$\sup_{x\in X} |h(x)| < 1, \quad \operatorname{Lip}(h, X) < 1 \quad \forall\, h \in \mathscr{C}_1; \tag{5.1.5b}$$

finally we set

$$\mathscr{C}_0 = \big\{\lambda h : h \in \mathscr{C}_1, \lambda \in \mathbb{Q}\big\}. \tag{5.1.5c}$$

As showed by the next remark, the above constructions are useful, since in general $C_b^0(X)$ (endowed with the uniform topology) is not separable, unless X is compact.

Remark 5.1.1 (Narrow convergence is induced by a distance). It is well known that narrow convergence is induced by a distance on $\mathscr{P}(X)$: an admissible choice is obtained by ordinating each element of $\mathscr{C}_1$ in a sequence (f_k) and setting

$$\delta(\mu,\nu) := \sum_{k=1}^{\infty} 2^{-k} \Big| \int_X f_k \, d\mu - \int_X f_k \, d\nu \Big|. \tag{5.1.6}$$

If d is a complete bounded metric for X we could also choose any p-Wasserstein distance on $\mathscr{P}(X)$ (see Chap. 7 and Remark 7.1.7). In particular, the family of all converging sequences is sufficient to characterize the narrow topology and we do not have to distinguish between compact and sequentially compact subsets.

Remark 5.1.2 (Narrow topology coincides with the weak* topology of $\big(C_b^0(X)\big)'$). $\mathscr{P}(X)$ can be identified with a convex subset of the unitary ball of the dual space $\big(C_b^0(X)\big)'$: by definition, narrow convergence is induced by the weak* topology of $\big(C_b^0(X)\big)'$. This identification is useful to characterize the closed convex hull in $\mathscr{P}(X)$ of a given set $\mathcal{K} \subset \mathscr{P}(X)$: Hahn-Banach theorem shows that

$$\mu \in \overline{\operatorname{Conv}}\,(\mathcal{K}) \quad \Longleftrightarrow \quad \int_X f \, d\mu \le \sup_{\nu\in\mathcal{K}} \int_X f \, d\nu \quad \forall\, f \in C_b^0(X). \tag{5.1.7}$$

For instance we can prove the separability of $\mathscr{P}(X)$ by choosing $\mathcal{K} := \{\delta_x : x \in D\}$, where D is a countable dense subset of X: by (5.1.7) we easily check that $\mathscr{P}(X) = \overline{\mathrm{Conv}}\,\mathcal{K}$ and therefore the subset of all the convex combinations with rational coefficients of δ-measures concentrated in D is narrowly dense in $\mathscr{P}(X)$.

The following theorem provides a useful characterization of relatively compact sets with respect to the narrow topology.

Theorem 5.1.3 (Prokhorov, [67, III-59]). *If a set $\mathcal{K} \subset \mathscr{P}(X)$ is* tight*, i.e.*

$$\forall\,\varepsilon > 0 \quad \exists\,K_\varepsilon \textit{ compact in } X \textit{ such that } \quad \mu(X \setminus K_\varepsilon) \le \varepsilon \quad \forall\,\mu \in \mathcal{K}, \tag{5.1.8}$$

then $\mathcal{K}$ is relatively compact in $\mathscr{P}(X)$. Conversely, if there exists an equivalent complete *metric for X, i.e. X is a so called* Polish space*, then every relatively compact subset of $\mathscr{P}(X)$ is tight.*

Observe in particular that in a Polish space X each measure $\mu \in \mathscr{P}(X)$ is tight; moreover, compact inner approximation holds for every Borel set:

$$\forall\,B \in \mathscr{B}(X),\ \varepsilon > 0 \quad \exists\,K_\varepsilon \Subset B: \quad \mu(B \setminus K_\varepsilon) \le \varepsilon. \tag{5.1.9}$$

In fact, this approximation property holds for a more general class of spaces, the so-called *Radon spaces* [136].

Definition 5.1.4 (Radon spaces). *A separable metric space X is a* Radon space *if every Borel probability measure $\mu \in \mathscr{P}(X)$ satisfies* (5.1.9)*.*

When the elements of $\mathcal{K} \subset X$ are ordinated in a sequence (μ_n) of tight measures (which is always the case if X is a Radon space), then the tightness condition (5.1.8) can also be reformulated as

$$\inf_{K \Subset X} \limsup_{n\to\infty} \mu_n(X \setminus K) = 0, \tag{5.1.10a}$$

or, equivalently since $\mu_n(X) \equiv 1$,

$$\sup_{K \Subset X} \liminf_{n\to\infty} \mu_n(K) = 1. \tag{5.1.10b}$$

An interesting result by Le Cam [103], [72, 11.5.3], shows that

$$\begin{gathered}\text{in a (metric, separable) Radon space } X,\\ \textit{every narrowly converging sequence } (\mu_n) \subset \mathscr{P}(X) \textit{ is tight.}\end{gathered} \tag{5.1.11}$$

Remark 5.1.5 (An integral condition for tightness). It is easy to check that (5.1.8) is equivalent to the following condition: there exists a function $\varphi : X \to [0, +\infty]$, whose sublevels $\{x \in X : \varphi(x) \le c\}$ are compact in X, such that

$$\sup_{\mu \in \mathcal{K}} \int_X \varphi(x)\,d\mu(x) < +\infty. \tag{5.1.12}$$

For, if $\{\varepsilon_n\}_{n=0}^{\infty}$ is a sequence with $\sum_{n=0}^{+\infty} \varepsilon_n < +\infty$ and $K_n := K_{\varepsilon_n}$ is an (increasing) sequence of compact sets satisfying (5.1.8), the function

$$\varphi(x) := \inf\Big\{n \geq 0 : x \in K_n\Big\} = \sum_{n=0}^{+\infty} \chi_{X \setminus K_n}(x), \tag{5.1.13}$$

satisfies (5.1.12). Conversely, if $\mathcal{K}$ satisfies (5.1.12), Chebichev inequality shows that (5.1.8) is satisfied by the family of sublevels of φ.

We conclude this part by a well known result comparing narrow convergence with convergence in the sense of distributions when $X = \mathbb{R}^d$.

Remark 5.1.6 (Narrow and distributional convergence in $X = \mathbb{R}^d$). For $n \in \mathbb{N}$ let μ_n, μ be Borel probability measures in the euclidean space $X = \mathbb{R}^d$ such that

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f(x)\, d\mu_n(x) = \int_{\mathbb{R}^d} f(x)\, d\mu(x) \quad \forall\, f \in C_c^\infty(\mathbb{R}^d). \tag{5.1.14}$$

Then the sequence (μ_n) is tight and it narrowly converges to μ as $n \to \infty$.
For, if $\zeta \in C_c^\infty(\mathbb{R}^d)$ satisfies

$$0 \leq \zeta \leq 1, \quad \zeta(x) = 1 \quad \text{if } |x| \leq 1/2, \quad \zeta(x) = 0 \quad \text{if } |x| \geq 1,$$

and $\zeta_k(x) := \zeta(x/k)$, we have

$$\liminf_{n\to\infty} \mu_n(\overline{B_k(0)}) \geq \lim_{n\to\infty} \int_{\mathbb{R}^d} \zeta_k(x)\, d\mu_n(x) = \int_{\mathbb{R}^d} \zeta_k(x)\, d\mu(x);$$

since Lebesgue dominated convergence theorem yields

$$\lim_{k\to\infty} \int_{\mathbb{R}^d} \zeta_k(x)\, d\mu(x) = 1,$$

choosing k sufficiently big we can verify the tightness condition (5.1.10b). By Prokhorov theorem the sequence (μ_n) has at least one narrowly convergence subsequence: a standard approximation result by convolution shows that any narrow limit point of the sequence (μ_n) should coincide with μ, which is therefore the narrow limit of the whole sequence (recall that the narrow topology is metrizable, see Remark 5.1.1).

5.1.1 Unbounded and l.s.c. integrands

When one needs to pass to the limit in expressions like (5.1.1) w.r.t. *unbounded or lower semicontinuous* functions f, the following two properties are quite useful. The first one is a lower semicontinuity property:

$$\liminf_{n\to\infty} \int_X g(x)\, d\mu_n(x) \geq \int_X g(x)\, d\mu(x) \tag{5.1.15}$$

for every sequence $(\mu_n) \subset \mathscr{P}(X)$ narrowly convergent to μ and any l.s.c. function $g : X \to (-\infty, +\infty]$ bounded from below: it follows by the same approximation argument of (5.1.3), by truncating the Moreau-Yosida approximations (5.1.4); in this case l.s.c. functions satisfy only the approximation property (5.1.2a), where e.g. $\mathscr{C}_0$ is given by (5.1.5a,b,c).
Changing g in $-g$ one gets the corresponding "lim sup" inequality for upper semi-continuous functions bounded from above. In particular, choosing as g the characteristic functions of open and closed subset of X, we obtain

$$\liminf_{n\to\infty} \mu_n(G) \geq \mu(G) \quad \forall\, G \text{ open in } X, \tag{5.1.16}$$

$$\limsup_{n\to\infty} \mu_n(F) \leq \mu(F) \quad \forall\, F \text{ closed in } X. \tag{5.1.17}$$

The statement of the second property requires the following definitions: we say that a Borel function $g : X \to [0, +\infty]$ is *uniformly integrable* w.r.t. a given set $\mathcal{K} \subset \mathscr{P}(X)$ if

$$\lim_{k\to\infty} \int_{\{x: g(x)\geq k\}} g(x)\, d\mu(x) = 0 \quad \text{uniformly w.r.t. } \mu \in \mathcal{K}. \tag{5.1.18}$$

If d is a given metric for X, in the particular case of $g(x) := d(x, \bar{x})^p$, for some (and thus any) $\bar{x} \in X$ and a given $p > 0$, i.e. if

$$\lim_{k\to\infty} \int_{X\setminus B_k(\bar{x})} d^p(\bar{x}, x)\, d\mu(x) = 0 \quad \text{uniformly w.r.t. } \mu \in \mathcal{K}, \tag{5.1.19}$$

we say that the set $\mathcal{K} \subset \mathscr{P}(X)$ *has uniformly integrable p-moments.* Notice that if

$$0 < p < p_1 \quad \text{and} \quad \sup_{\mu\in\mathcal{K}} \int_X d(x, \bar{x})^{p_1}\, d\mu(x) < +\infty, \tag{5.1.20}$$

then $\mathcal{K}$ has uniformly integrable p-moments. In the case when $X = \mathbb{R}^d$ with the usual Euclidean distance, any family $\mathcal{K} \subset \mathscr{P}(\mathbb{R}^d)$ satisfying (5.1.20) is tight. The following lemma provides a characterization of p-uniformly integrable families, extending the validity of (5.1.1) to unbounded but with p-growth functions, i.e. functions $f : X \to \mathbb{R}$ such that

$$|f(x)| \leq A + B\, d^p(\bar{x}, x) \qquad \forall x \in X, \tag{5.1.21}$$

for some A, $B \geq 0$ and $\bar{x} \in X$. We denote by $\mathscr{P}_p(X)$ the subset

$$\mathscr{P}_p(X) := \Big\{\mu \in \mathscr{P}(X) : \int_X d(x, \bar{x})^p\, d\mu(x) < +\infty \quad \text{for some } \bar{x} \in X\Big\}. \tag{5.1.22}$$

Lemma 5.1.7. *Let (μ_n) be a sequence in $\mathscr{P}(X)$ narrowly convergent to $\mu \in \mathscr{P}(X)$. If $f : X \to \mathbb{R}$ is continuous, $g : X \to (-\infty, +\infty]$ is lower semicontinuous, and*

$|f|, g^-$ are uniformly integrable w.r.t. the set $\{\mu_n\}_{n\in\mathbb{N}}$, then

$$\liminf_{n\to\infty}\int_X g(x)\,d\mu_n(x) \ge \int_X g(x)\,d\mu(x) > -\infty, \tag{5.1.23a}$$

$$\lim_{n\to\infty}\int_X f(x)\,d\mu_n(x) = \int_X f(x)\,d\mu(x). \tag{5.1.23b}$$

Conversely, if $f : X \to [0, +\infty)$ is continuous, μ_n-integrable, and

$$\limsup_{n\to\infty}\int_X f(x)\,d\mu_n(x) \le \int_X f(x)\,d\mu(x) < +\infty, \tag{5.1.24}$$

then f is uniformly integrable w.r.t. $\{\mu_n\}_{n\in\mathbb{N}}$.
In particular, a family $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathscr{P}(X)$ has uniformly integrable p-moments iff (5.1.1) holds for every continuous function $f : X \to \mathbb{R}$ with p-growth.

Proof. If μ_n narrowly converges to μ as $n \to \infty$ and g is lower semicontinuous, (5.1.15) yields

$$\liminf_{n\to\infty}\int_X g_k\,d\mu_n \ge \int_X g_k\,d\mu \qquad \forall k \in \mathbb{N},$$

where $g_k := g \vee (-k)$, $k \ge 0$. On the other hand, since g^- is uniformly integrable w.r.t. $\{\mu_n\}_{n\in\mathbb{N}}$ and $g_k \ge g$, (5.1.18) gives

$$\sup_{n\in\mathbb{N}}\left(\int_X g_k\,d\mu_n - \int_X g\,d\mu_n\right) \le \sup_{n\in\mathbb{N}}\int_{\{x:g^-(x)\ge k\}} g^-\,d\mu_n \to 0$$

as $k \to \infty$. Using these two facts we obtain (5.1.23a). As usual, (5.1.23b) follows by applying (5.1.23a) to $g := f$ and $g := -f$.

Conversely, let $f : X \to [0, +\infty)$ be a continuous function satisfying (5.1.24) and let

$$f^k(x) := f(x) \wedge k,\ \forall\, x \in X, \quad F^k := \big\{x \in X : f(x) \ge k\big\};$$

since f^k is continuous and bounded and F^k is a closed subset of X, recalling (5.1.17) and (5.1.15) we have for any $\epsilon > 0$

$$\begin{aligned}\limsup_{n\to\infty}\int_{\{x:f(x)\ge k\}} f\,d\mu_n &= \limsup_{n\to\infty}\Big(\int_X \big(f - f^k\big)\,d\mu_n + k\mu_n(F^k)\Big)\\ &\le \int_X \big(f - f^k\big)\,d\mu + k\mu(F^k) = \int_{F^k} f\,d\mu < \varepsilon\end{aligned}$$

for k sufficiently large. Since f is uniformly integrable for finite subsets of $\{\mu_n\}_{n\in\mathbb{N}}$, this easily leads to the uniform integrability of f. $\square$

There exists an interesting link between narrow convergence of probability measures and Kuratowski convergence of their supports:

Proposition 5.1.8. *If* $(\mu_n) \subset \mathscr{P}(X)$ *is a sequence narrowly converging to* $\mu \in \mathscr{P}(X)$ *then* $\operatorname{supp}\mu \subset K\!-\!\liminf_{n\to\infty} \operatorname{supp}\mu_n$*, i.e.*

$$\forall\, x \in \operatorname{supp}\mu \quad \exists\, x_n \in \operatorname{supp}\mu_n : \quad \lim_{n\to\infty} x_n = x. \tag{5.1.25}$$

Proof. Let $x \in \operatorname{supp}\mu$ and let $B_{1/k}(x)$ be the open ball of center x and radius $1/k$ with respect to the distance d on X. By (5.1.16) we obtain

$$\liminf_{n\to\infty} \mu_n(B_{1/k}(x)) \geq \mu(B_{1/k}(x)) > 0;$$

thus the strictly increasing sequence

$$j_0 := 0, \quad j_k := \min\Big\{n \in \mathbb{N} : n > j_{k-1}, \quad \operatorname{supp}\mu_m \cap B_{1/k}(x) \neq \varnothing \quad \forall\, m \geq n\Big\}$$

is well defined. For $j_k \leq n < j_{k+1}$ pick a point $x_n \in \operatorname{supp}\mu_n \cap B_{1/k}(x)$: clearly the sequence (x_n) satisfies (5.1.25). □

Corollary 5.1.9 (Convergence of Dirac masses). *A sequence* $(x_n) \subset X$ *is convergent in* X *iff the sequence* (δ_{x_n}) *is narrowly convergent in* $\mathscr{P}(X)$*; in this case, the limit measure* μ *is* δ_x*,* x *being the limit of the sequence* (x_n)*.*

Proposition 5.1.10. *Let* $(\mu_n) \subset \mathscr{P}(X)$ *be a sequence narrowly converging to* $\mu \in \mathscr{P}(X)$ *and let* $f, g : X \to (-\infty, +\infty]$ *be Borel functions such that* $|f|, g^-$ *are uniformly integrable with respect to* $\{\mu_n\}_{n\in\mathbb{N}}$*. If for any* $\varepsilon > 0$ *there exists a closed set* $A \subset X$ *such that*

$$f|_A \text{ is continuous,} \quad g|_A \text{ is l.s.c.,} \quad \text{and} \quad \limsup_{n\to\infty} \mu_n(X \setminus A) < \varepsilon, \tag{5.1.26}$$

then (5.1.1) *and* (5.1.15) *hold.*

Proof. As usual we can limit us to consider the l.s.c. case; using the uniform integrability of g^- with respect to $\{\mu_n\}_{n\in\mathbb{N}}$, a truncation argument, and arguing as in the first part of the proof of Lemma 5.1.7, we reduce immediately ourselves to the case when g is bounded from below by a constant $-M \leq 0$. Let $\varepsilon > 0, k \in \mathbb{N}$ be fixed and let $A \subset X$ be a closed set such that (5.1.26) holds. We consider the truncated functions $g^k(x) := g(x) \wedge k$ for $x \in X$, and the lower semicontinuous $\tilde g^k$

$$\tilde g^k(x) = \begin{cases} g^k(x) & \text{if } x \in A, \\ k & \text{if } x \in X \setminus A, \end{cases}$$

which extends $g^k|_A$ to X. We obtain

$$\begin{aligned}
\liminf_{n\to\infty} \int_X g\, d\mu_n &\geq \liminf_{n\to\infty} \int_X g^k\, d\mu_n \geq \liminf_{n\to\infty} \Big(\int_X \tilde g^k\, d\mu_n + \int_{X\setminus A} (g^k - \tilde g^k)\, d\mu_n \Big) \\
&\geq \liminf_{n\to\infty} \int_X \tilde g^k\, d\mu_n - (M+k) \limsup_{n\to\infty} \mu_n(X \setminus A) \\
&\geq \int_X \tilde g^k(x)\, d\mu - \varepsilon(M+k) \geq \int_X g^k(x)\, d\mu - (k+M)\varepsilon.
\end{aligned}$$

Passing to the limit, first as $\varepsilon \downarrow 0$ and then as $k \uparrow \infty$ we obtain (5.1.15). □

5.1.2 Hilbert spaces and weak topologies

Let X be a separable, infinite dimensional, Hilbert space, with norm $|\cdot|$ and scalar product $\langle\cdot,\cdot\rangle$; in many circumstances it would be useful to rephrase the results of the previous section with respect to the weak topology $\sigma(X, X')$ of X. Unfortunately, the weak topology is not induced by a distance on X, thus the previous statements are not immediately applicable.

We can circumvent this difficulty by the following simple trick: we introduce a new continuous norm $\|\cdot\|_\varpi$, inducing a topology ϖ globally weaker than $\sigma(X, X')$, but coinciding with $\sigma(X, X')$ on bounded sets (with respect to the original stronger norm $|\cdot|$). In particular bounded sets of X are relatively compact w.r.t. ϖ and Borel sets with respect to the three topologies coincide.

For instance, if $\{e_n\}_{n=1}^{+\infty}$ is an orthonormal basis of X, an admissible choice is

$$\|x\|_\varpi^2 := \sum_{n=1}^{\infty} \frac{1}{n^2} \langle x, e_n\rangle^2. \tag{5.1.27}$$

In fact, if $(x_k) \subset X$ is a bounded sequence, we can extract a subsequence, still denoted by x_k, weakly converging to x in X; since $\langle x_k - x, e_n\rangle \to 0$ as $k \to \infty$ for each $n \geq 1$, Lebesgue dominated convergence theorem yields

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

We denote by X_ϖ the new pre-Hilbertian topological vector space. We will also introduce the space of smooth cylindrical functions $\mathrm{Cyl}(X)$: observe that for finite dimensional spaces, X_ϖ is homeomorphic to X and $\mathrm{Cyl}(X) = C_c^\infty(X)$.

Definition 5.1.11 (Finite dimensional projection and smooth cylindrical functions). *We denote by $\Pi_d(X)$ the space of all maps $\pi : X \to \mathbb{R}^d$ of the form*

$$\pi(x) = (\langle x, e_1\rangle, \langle x, e_2\rangle, \ldots, \langle x, e_d\rangle) \qquad x \in X, \tag{5.1.28}$$

where $\{e_1, \ldots, e_d\}$ is any orthonormal family of vectors in X. The adjoint map

$$\pi^* : y \in \mathbb{R}^d \to \sum_{k=1}^{d} y_k e_k \in \mathrm{span}(e_1, \ldots, e_k) \subset X \tag{5.1.29}$$

is a linear isometry of $\mathbb{R}^d$ onto $\mathrm{span}(e_1, \ldots, e_d)$ so that

$$\begin{aligned} \pi \circ \pi^* \quad & \text{is the identity in } \mathbb{R}^d \text{ and} \\ \hat\pi := \pi^* \circ \pi \quad & \text{is the orthogonal projection of } X \text{ onto } \mathrm{span}(e_1, \ldots, e_d). \end{aligned} \tag{5.1.30}$$

We denote by $\mathrm{Cyl}(X)$ the functions $\varphi = \psi \circ \pi$ with $\pi \in \Pi_d(X)$ and $\psi \in C_c^\infty(\mathbb{R}^d)$.

Notice that any $\varphi = \psi \circ \pi \in \mathrm{Cyl}(X)$ is a Lipschitz function, everywhere differentiable in the Fréchet sense, and that φ is also continuous with respect to the weak topology of X and to X_ϖ (if the corresponding orthogonal systems coincide). Moreover $\nabla\varphi = \pi^* \circ \nabla\psi \circ \pi$.

The following properties are immediate:

Lemma 5.1.12. *Let X be a separable Hilbert space and let X_ϖ be the pre-Hilbertian vector space whose norm is defined by* (5.1.27).

(a) *If K is weakly compact in X then K is strongly compact in X_ϖ.*

(b) *If*

$$g : X \to (-\infty, +\infty] \quad \text{is weakly l.s.c.} \quad \text{and} \quad \lim_{|x|\to\infty} g(x) = +\infty, \tag{5.1.31}$$

then it is lower semicontinuous in X_ϖ with compact sublevels.

(c) *Let us denote by $\overline{B_R} := \{x \in X : |x| \le R\}$ the centered closed balls w.r.t. the strong norm; if $\mathcal{K} \subset \mathscr{P}(X)$ satisfies the* weak tightness condition

$$\forall\, \varepsilon > 0 \quad \exists R_\varepsilon > 0 \text{ such that } \mu(X \setminus \overline{B_{R_\varepsilon}}) \le \varepsilon \quad \forall\, \mu \in \mathcal{K}, \tag{5.1.32}$$

then $\mathcal{K}$ is tight in $\mathscr{P}(X_\varpi)$ and therefore relatively compact in $\mathscr{P}(X_\varpi)$.

(d) *If the sequence $(\mu_n) \subset \mathscr{P}(X)$ is narrowly converging to μ in $\mathscr{P}(X_\varpi)$ and it is weakly tight according to* (5.1.32)*, then for every Borel functions $f, g : X \to (-\infty, +\infty]$ such that $g^-, |f|$ are uniformly integrable and f (resp. g) is weakly continuous (resp. l.s.c.) on bounded sets of X, we have*

$$\liminf_{n\to\infty} \int_X g(x)\, d\mu_n(x) \ge \int_X g(x)\, d\mu(x), \tag{5.1.33a}$$

$$\lim_{n\to\infty} \int_X f(x)\, d\mu_n(x) = \int_X f(x)\, d\mu(x). \tag{5.1.33b}$$

(e) *$\mathcal{K} \subset \mathscr{P}(X)$ is weakly tight according to* (5.1.32) *iff there exists a Borel function $h : X \to [0, +\infty]$ such that $h(x) \to +\infty$ as $|x| \to \infty$ and*

$$\sup_{\mu\in\mathcal{K}} \int_X h(x)\, d\mu(x) < +\infty. \tag{5.1.34}$$

(f) *If the sequence $(\mu_n) \subset \mathscr{P}(X)$ is weakly tight according to* (5.1.32)*, then it narrowly converges to μ in $\mathscr{P}(X_\varpi)$ iff*

$$\lim_{n\to\infty} \int_X \varphi(x)\, d\mu_n(x) = \int_X \varphi(x)\, d\mu(x) \quad \forall\, \varphi \in \mathrm{Cyl}(X). \tag{5.1.35}$$

Proof. (a) and (b) are trivial and (c) is a direct consequence of the fact that bounded and closed convex sets are compact in X_ϖ. Since on bounded subsets of X the topology of X_ϖ coincides with the weak one, (d) follows from Proposition 5.1.10.

One implication in (e) follows directly from Chebichev inequality. The other one can be proved arguing as in Remark 5.1.5.
Finally, one implication in (f) is a consequence of (5.1.33b) of (d), since (smooth) cylindrical functions are bounded and weakly continuous. In order to prove the converse implication, we can simply check that any two narrow limit point μ^1, μ^2 of the sequence (μ_n) in $\mathscr{P}(X_\varpi)$ should coincide. For, let $f \in C^0_b(X)$ and π_d be the map (5.1.28), so that $\hat\pi_d := \pi_d^* \circ \pi_d$ is the orthogonal projection of X onto $X_d = \mathrm{span}(e_1, \cdots, e_d)$. We set $\psi_d := f \circ \pi_d^* \in C^0_b(\mathbb{R}^d)$, $\varphi_d := \psi_d \circ \pi_d = f \circ \hat\pi_d \in \mathrm{Cyl}(X)$; by (5.1.35) we know

$$\int_X \varphi(x)\, d\mu^1(x) = \int_X \varphi(x)\, d\mu^2(x) \quad \forall\, \varphi \in \mathrm{Cyl}(X); \tag{5.1.36}$$

a standard approximation argument for bounded continuous functions defined in $\mathbb{R}^d$ by smooth functions in $C^\infty_c(\mathbb{R}^d)$ as in Remark 5.1.6 yields (5.1.36) for $\varphi := \varphi_d$ and $d \in \mathbb{N}$; therefore

$$\int_X f(\hat\pi_d(x))\, d\mu^1(x) = \int_X f(\hat\pi_d(x))\, d\mu^2(x) \quad \forall\, d \in \mathbb{N}.$$

Passing to the limit as $d \to \infty$, since $\hat\pi_d(x) \to x$ for every $x \in X$, Lebesgue dominated convergence theorem yields

$$\int_X f(x)\, d\mu^1(x) = \int_X f(x)\, d\mu^2(x).$$

Since f is an arbitrary function in $C^0_b(X)$ we obtain $\mu^1 = \mu^2$. □

In the following theorem we will show that narrow convergence in $\mathscr{P}(X_\varpi)$ and convergence of the p-moment $\int_X |x|^p\, d\mu_h(x)$ (but more general integrands are allowed) yields convergence in $\mathscr{P}(X)$, thus obtaining the measure-theoretic version of the fact that weak convergence and convergence of the norms in X imply strong convergence. We will show a different proof of this fact at the end of Section 7.1.

Theorem 5.1.13. *Let $j : [0, +\infty) \to [0, +\infty)$ be a continuous, strictly increasing and surjective map, and let μ_n, $\mu \in \mathscr{P}(X)$ be satisfying*

$$\mu_n \to \mu \quad \text{in } \mathscr{P}(X_\varpi), \qquad \lim_{n\to\infty} \int_X j(|x|)\, d\mu_n(x) = \int_X j(|x|)\, d\mu < +\infty. \tag{5.1.37}$$

Then μ_n converge to μ in $\mathscr{P}(X)$.

Proof. Observe that the family $\{\mu_n\}_{n\in\mathbb{N}}$ is weakly tight, according to (5.1.32). We consider the vector space $\mathscr{H}$ of continuous functions $h : X \to \mathbb{R}$ satisfying the growth condition (compare with (5.1.21))

$$\exists\, A, B \geq 0 : \quad |h(x)| \leq A + Bj(|x|) \quad \forall\, x \in X, \tag{5.1.38}$$

and

$$\lim_{n\to\infty}\int_X h(x)\,d\mu_n(x)=\int_X h(x)\,d\mu(x). \tag{5.1.39}$$

Observe that $\mathscr{H}$ is closed with respect to uniform convergence of functions and contains the constants and the function $j(|\cdot|)$.

By the monotonicity argument outlined at the beginning of Section 5.1 we need only to check that the infimum of a finite number of functions of the form

$$x\mapsto\big(q_1+q_2|x-y|\big)\wedge k,\quad q_1\in\mathbb{R},\quad q_2,k\geq 0,\quad y\in X, \tag{5.1.40}$$

belongs to $\mathscr{H}$. To this aim, let us consider the convex cone $\mathscr{A}\subset\mathscr{H}$ of strongly continuous functions which satisfy (5.1.38), (5.1.39), and are weakly lower semicontinuous. Notice that, truncated affine functions of the type

$$x\mapsto(-l)\vee\big(a+\langle x,y\rangle\big)\wedge m,\quad \text{for } l,m\geq 0,\ a\in\mathbb{R},\ y\in X\quad \text{belongs to } \mathscr{A}, \tag{5.1.41}$$

since they are bounded, weakly continuous, and condition (5.1.39) follows by (d) of Lemma 5.1.12.

Let us first prove that $\mathscr{A}$ is a lattice.

Claim 1. *If $f,\ g\in C^0(X)$ satisfy* (5.1.38), *are weakly lower semicontinuous, and $f+g\in\mathscr{A}$, then both $f,\ g\in\mathscr{A}$.*

Indeed, by (5.1.33a) we have

$$\begin{aligned}\int_X(f+g)\,d\mu=\lim_{n\to\infty}\int_X(f+g)\,d\mu_n&\geq\limsup_{n\to\infty}\int_X f\,d\mu_n+\liminf_{n\to\infty}\int_X g\,d\mu_n\\ &\geq\int_X f\,d\mu+\int_X g\,d\mu=\int_X(f+g)\,d\mu,\end{aligned}$$

which yields

$$\limsup_{n\to\infty}\int_X f\,d\mu_n+\liminf_{n\to\infty}\int_X g\,d\mu_n=\int_X f\,d\mu+\int_X g\,d\mu; \tag{5.1.42}$$

since by (5.1.33a)

$$\limsup_{n\to\infty}\int_X f\,d\mu_n\geq\int_X f\,d\mu,\quad \liminf_{n\to\infty}\int_X g\,d\mu_n\geq\int_X g\,d\mu,$$

(5.1.42) yields

$$\limsup_{n\to\infty}\int_X f\,d\mu_n=\int_X f\,d\mu,\quad \liminf_{n\to\infty}\int_X g\,d\mu_n=\int_X g\,d\mu;$$

inverting the role of f and g we obtain $f,g\in\mathscr{A}$.

Claim 1 immediately implies that $\mathscr{A}$ is a lattice, as

$$f,g\in\mathscr{A}\quad\Rightarrow\quad f+g=(f\wedge g)+(f\vee g)\in\mathscr{A}\quad\Rightarrow\quad f\wedge g,\ f\vee g\in\mathscr{A}.$$

Since

$$\Big(q_1+q_2|x-y|\Big)\wedge k=\Big(q_1+\big(q_2|x-y|\big)\wedge(k-q_1)\Big)\wedge k,$$

it remains to show that

$$\text{all functions } x\mapsto|x-y|\wedge k,\ \text{for } y\in X,\ k\ge 0,\ \text{belong to } \mathscr{A}. \tag{5.1.43}$$

To this aim, we need a further claim.

Claim 2. If $f\in\mathscr{A}$ and $\theta:\mathbb{R}\to\mathbb{R}$ is a uniformly continuous, bounded, increasing function, then $\theta\circ f\in\mathscr{A}$.

Indeed, since θ can be uniformly approximated by a sequence of Lipschitz continuous increasing maps, it is not restrictive to assume that θ is Lipschitz, bounded, and its Lipschitz constant is less than 1; in this case also $x\mapsto x-\theta(x)$ is Lipschitz and increasing, thus $\theta\circ f$ and $f-f\circ\theta$ are still weakly lower semicontinuous they satisfies the growth condition (5.1.38) and and their sum is $f\in\mathscr{A}$: we can apply Claim 1.

Let us consider (5.1.43) in the case $y=0$ first: we fix $R>0$ and we consider the continuous increasing function θ_R which vanishes in $(-\infty,0)$ and satisfies

$$\theta_R(s):=\big(j^{-1}(s)\big)^2\wedge R^2,\ s\ge 0,\quad\text{so that}\quad r^2\wedge R^2=\theta_R(j(r))\quad\forall\, r\ge 0.$$

By Claim 2, we deduce that the map f_R defined by $f_R(x):=|x|^2\wedge R^2$ belongs to $\mathscr{A}$.

Now, for fixed $k,l,m>0$ and $y\in X$, we set

$$g_{l,m}(x):=(-l)\vee\Big(-2\langle x,y\rangle+|y|^2\Big)\wedge m,\quad g_{R,l.m,k}:=\Big(\big(f_R+g_{l,m}\big)\vee 0\Big)^{1/2}\wedge k,$$

and we know by the lattice property, the previous claim, and (5.1.41) that $g_{R,l,m,k}\in\mathscr{A}$. Choosing now $R\ge l+k^2$ and $m\ge k$ the expression of $g_{R,l,m,k}$ simplifies to

$$g_{R,l,m,k}(x)=\tilde g_{l,k}(x):=\Big(\Big(|x|^2+\big(-2\langle x,y\rangle+|y|^2\big)\vee(-l)\Big)\vee 0\Big)^{1/2}\wedge k,$$

which belongs to $\mathscr{A}$, is decreasing with respect to l, and satisfies

$$\lim_{l\to\infty}\tilde g_{l,k}(x)=\inf_{l\in\mathbb{N}}\tilde g_{l,k}(x)=|x-y|\wedge k\quad\forall\, x\in X.$$

It follows that

$$\limsup_{n\to\infty}\int_X\big(|x-y|\wedge k\big)\,d\mu_n(x)\le\limsup_{n\to\infty}\int_X\tilde g_{l,k}(x)\,d\mu_n(x)=\int_X\tilde g_{l,k}(x)\,d\mu(x);$$

passing to the limit as $l\to+\infty$, and recalling that the corresponding "lim inf" inequality is provided by (5.1.33a) of Lemma 5.1.12, we obtain (5.1.43). $\square$

5.2 Transport of measures

If X_1, X_2 are separable metric spaces, $\mu \in \mathscr{P}(X_1)$, and $\boldsymbol{r} : X_1 \to X_2$ is a Borel (or, more generally, μ-measurable) map, we denote by $\boldsymbol{r}_\# \mu \in \mathscr{P}(X_2)$ the *push-forward of μ through $\boldsymbol{r}$*, defined by

$$\boldsymbol{r}_\# \mu(B) := \mu(\boldsymbol{r}^{-1}(B)) \quad \forall\, B \in \mathscr{B}(X_2). \tag{5.2.1}$$

More generally we have

$$\int_{X_1} f(\boldsymbol{r}(x))\, d\mu(x) = \int_{X_2} f(y)\, d\,\boldsymbol{r}_\# \mu(y) \tag{5.2.2}$$

for every bounded (or $\boldsymbol{r}_\# \mu$-integrable) Borel function $f : X_2 \to \mathbb{R}$. It is easy to check that

$$\nu \ll \mu \quad \Longrightarrow \quad \boldsymbol{r}_\# \nu \ll \boldsymbol{r}_\# \mu \qquad \forall\, \mu, \nu \in \mathscr{P}(X_1). \tag{5.2.3}$$

In the following we will extensively use the following composition rule

$$(\boldsymbol{r} \circ \boldsymbol{s})_\# \mu = \boldsymbol{r}_\# (\boldsymbol{s}_\# \mu) \quad \text{where } \boldsymbol{s} : X_1 \to X_2,\ \boldsymbol{r} : X_2 \to X_3,\ \mu \in \mathscr{P}(X_1). \tag{5.2.4}$$

Furthermore, if $\boldsymbol{r} : X_1 \to X_2$ is a continuous map, then

$$\boldsymbol{r}_\# : \mathscr{P}(X_1) \to \mathscr{P}(X_2) \text{ is continuous w.r.t. the narrow convergence} \tag{5.2.5}$$

and

$$\boldsymbol{r}\big(\operatorname{supp} \mu\big) \subset \operatorname{supp} \boldsymbol{r}_\# \mu = \overline{\boldsymbol{r}\big(\operatorname{supp} \mu\big)}. \tag{5.2.6}$$

Lemma 5.2.1. *Let $\boldsymbol{r}_n : X_1 \to X_2$ be Borel maps uniformly converging to $\boldsymbol{r}$ on compact subsets of X_1 and let $(\mu_n) \subset \mathscr{P}(X_1)$ be a tight sequence narrowly converging to μ. If $\boldsymbol{r}$ is continuous, then $(\boldsymbol{r}_n)_\# \mu_n$ narrowly converge to $\boldsymbol{r}_\# \mu$.*

Proof. Let f be a bounded continuous function in X_2. We will prove the lim inf inequality

$$\liminf_{n\to\infty} \int_{X_2} f\, d(\boldsymbol{r}_n)_\# \mu_n \geq \int_{X_2} f\, d\boldsymbol{r}_\# \mu,$$

as the lim sup simply follows replacing f by $-f$. To this aim, possibly adding to f a constant, we can assume that $f \geq 0$. For any compact set $K \subset X_1$ the uniform convergence of $\boldsymbol{r}_n$ to $\boldsymbol{r}$ on K gives the uniform convergence of $f \circ \boldsymbol{r}_n$ to $f \circ \boldsymbol{r}$ on K, therefore (5.1.15) gives

$$\begin{aligned}
\liminf_{n\to\infty} \int_{X_1} f \circ \boldsymbol{r}_n\, d\mu_n \quad &\geq \quad \liminf_{n\to\infty} \int_K f \circ \boldsymbol{r}_n\, d\mu_n = \liminf_{n\to\infty} \int_K f \circ \boldsymbol{r}\, d\mu_n \\
&\geq \quad \big(-\sup f\big) \sup_n \mu_n(X_1 \setminus K) + \liminf_{n\to\infty} \int_{X_1} f \circ \boldsymbol{r}\, d\mu_n \\
&\geq \quad \big(-\sup f\big) \sup_n \mu_n(X_1 \setminus K) + \int_{X_1} f \circ \boldsymbol{r}\, d\mu.
\end{aligned}$$

Since $\{\mu_n\}_{n\in\mathbb{N}}$ is tight, we can find an increasing sequence of compact set K_m such that $\lim_m \sup_n \mu_n(X_1 \setminus K_m) = 0$. Putting $K = K_m$ in the inequality above and letting $m \uparrow +\infty$ the proof is achieved. □

Lemma 5.2.2 (Tightness criterion). *Let X, X_1, $X_2, \ldots, X_N$ be separable metric spaces and let $\boldsymbol{r}^i : X \to X_i$ be continuous maps such that the product map*

$$\boldsymbol{r} := \boldsymbol{r}^1 \times \boldsymbol{r}^2 \times \ldots \times \boldsymbol{r}^N : X \to X_1 \times \ldots \times X_N \quad \textit{is proper.} \tag{5.2.7}$$

Let $\mathcal{K} \subset \mathscr{P}(X)$ be such that $\mathcal{K}_i := \boldsymbol{r}^i_\#(\mathcal{K})$ is tight in $\mathscr{P}(X_i)$ for $i = 1, \ldots, N$. Then also $\mathcal{K}$ is tight in $\mathscr{P}(X)$.

Proof. For every $\mu \in \mathscr{P}(X)$ we denote by μ_i the measure $\mu_i := \boldsymbol{r}^i_\# \mu$. By definition, for each $\varepsilon > 0$ there exist compact sets $K_i \subset X_i$ such that $\mu_i(X_i \setminus K_i) \leq \varepsilon/N$ for any $\mu \in \mathcal{K}$; it follows that $\mu(X \setminus (\boldsymbol{r}^i)^{-1}(K_i)) \leq \varepsilon/N$ and

$$\mu\left(X \setminus \bigcap_{i=1}^N (\boldsymbol{r}^i)^{-1}(K_i)\right) \leq \sum_{i=1}^N \mu\left(X \setminus (\boldsymbol{r}^i)^{-1}(K_i)\right) \leq \varepsilon \qquad \forall \mu \in \mathcal{K}. \tag{5.2.8}$$

On the other hand $\cap_{i=1}^N (\boldsymbol{r}^i)^{-1}(K_i) = \boldsymbol{r}^{-1}(K_1 \times K_2 \times \ldots \times K_N)$, which is compact by (5.2.7). $\square$

For an integer $N \geq 2$ and i, $j = 1, \ldots, N$, we denote by π^i, $\pi^{i,j}$ the projection operators defined on the product space $\boldsymbol{X} := X_1 \times \ldots \times X_N$ respectively defined by

$$\pi^i : (x_1, \ldots, x_N) \mapsto x_i \in X_i, \quad \pi^{i,j} : (x_1, \ldots, x_N) \mapsto (x_i, x_j) \in X_i \times X_j. \tag{5.2.9}$$

If $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$, the *marginals* of $\boldsymbol{\mu}$ are the probability measures

$$\mu^i := \pi^i_\# \boldsymbol{\mu} \in \mathscr{P}(X_i), \quad \boldsymbol{\mu}^{i\,j} := \pi^{i,j}_\# \boldsymbol{\mu} \in \mathscr{P}(X_i \times X_j). \tag{5.2.10}$$

If $\mu^i \in \mathscr{P}(X_i)$, $i = 1, \ldots, N$, the class of *multiple plans* with marginals μ^i is defined by

$$\Gamma(\mu^1, \ldots, \mu^N) := \Big\{\boldsymbol{\mu} \in \mathscr{P}(X_1 \times \ldots \times X_N) : \pi^i_\# \boldsymbol{\mu} = \mu^i,\ i = 1, \ldots, N\Big\}. \tag{5.2.11}$$

In the case $N = 2$ a measure $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ is also called *transport plan* between μ^1 and μ^2. Notice also that

$$\Gamma(\mu^1, \mu^2) = \{\mu^1 \times \mu^2\} \quad \text{if either } \mu^1 \text{ or } \mu^2 \text{ is a Dirac mass.} \tag{5.2.12}$$

We will mostly consider multiple plans with $N = 2$ or $N = 3$. To each couple of measures $\mu^1 \in \mathscr{P}(X_1), \mu^2 = \boldsymbol{r}_\# \mu^1 \in \mathscr{P}(X_2)$ linked by a Borel transport map $\boldsymbol{r} : X_1 \to X_2$ we can associate the transport plan

$$\boldsymbol{\mu} := (\boldsymbol{i} \times \boldsymbol{r})_\# \mu^1 \in \Gamma(\mu^1, \mu^2), \quad \boldsymbol{i} \text{ being the identity map on } X_1. \tag{5.2.13}$$

If $\boldsymbol{\mu}$ is representable as in (5.2.13) then we say that $\boldsymbol{\mu}$ is *induced* by $\boldsymbol{r}$. Each transport plan $\boldsymbol{\mu}$ concentrated on a $\boldsymbol{\mu}$-measurable graph in $X_1 \times X_2$ admits the representation (5.2.13) for some μ^1-measurable map $\boldsymbol{r}$, which therefore transports

μ^1 to μ^2 (see, e.g., [9]; the same result holds for Borel graphs and maps if X_1, X_2 are Polish spaces [136, p. 107])

We define also the *inverse* $\boldsymbol{\mu}^{-1} \in \mathscr{P}(X_2 \times X_1)$ of a transport plan $\boldsymbol{\mu} \in \mathscr{P}(X_1 \times X_2)$ by $i_\#\boldsymbol{\mu}$, where $i(x_1, x_2) = (x_2, x_1)$.

Remark 5.2.3. By Lemma 5.2.2, if $X_1, X_2, \cdots, X_N$ are Radon spaces (i.e. each measure $\mu^i \in \mathscr{P}(X_i)$ is tight), $\Gamma(\mu^1, \ldots, \mu^N)$ is compact in $\mathscr{P}(X)$ and not empty, since it contains at least $\mu^1 \times \ldots \times \mu^N$. If for some Borel functions $g_i : X_i \to [0, +\infty]$

$$\int_{X_i} g_i(x_i)\, d\mu^i(x_i) < +\infty \quad i = 1, \ldots, N, \tag{5.2.14}$$

then it is easy to check that $g(x) := \sum_{i=1}^N g_i(x_i)$ defined in the product space $\boldsymbol{X} = X_1 \times X_2 \times \cdots \times X_N$ is uniformly integrable with respect to $\Gamma(\mu^1, \ldots, \mu^N)$.

When X is a separable Hilbert space as in Section 5.1.2, the following result provides a sufficient condition for the convergence of the integrals $\int_{X^2} \langle x_1, x_1\rangle\, d\mu_h$ even in the case when the measures μ_h do not converge narrowly with respect to the strong topology.

Lemma 5.2.4. *Let $(\boldsymbol{\mu}_n) \subset \mathscr{P}(X \times X)$ be a sequence narrowly converging to $\boldsymbol{\mu}$ in $\mathscr{P}(X \times X_\varpi)$, with*

$$\sup_n \int_{X^2} |x_1|^p + |x_2|^q\, d\boldsymbol{\mu}_n(x_1, x_2) < +\infty, \quad p, q \in (1, \infty),\ p^{-1} + q^{-1} = 1. \tag{5.2.15}$$

If either $\pi^1_\#\boldsymbol{\mu}_n$ have uniformly integrable p-moments or $\pi^2_\#\boldsymbol{\mu}_n$ have uniformly integrable q-moments, then

$$\lim_{n\to\infty} \int_{X\times X} \langle x_1, x_2\rangle\, d\boldsymbol{\mu}_n = \int_{X\times X} \langle x_1, x_2\rangle\, d\boldsymbol{\mu}.$$

Proof. We assume to fix the ideas that $\pi^2_\#\boldsymbol{\mu}_n$ have uniformly integrable q-moments and we show that the function $(x_1, x_2) \mapsto g(x_1, x_2) := |x_1| \cdot |x_2|$ is uniformly integrable. For any $k, m \in \mathbb{N}$ we have

$$g(x_1, x_2) \geq k, \quad |x_2| \leq m \quad \Rightarrow \quad |x_1| \geq k/m$$

and therefore

$$\int_{\{g\geq k\}} g\, d\boldsymbol{\mu}_n \leq m \int_{\{|x_1|\geq k/m\}} |x_1|\, d\pi^1_\#\boldsymbol{\mu}_n + C\Big(\int_{\{|x_2|\geq m\}} |x_2|^q\, d\pi^2_\#\boldsymbol{\mu}_n\Big)^{1/q}$$

where $C^p := \sup_n \int_X |x_1|^p\, d\boldsymbol{\mu}_n$. Taking the supremum w.r.t. n and the $\limsup$ as $k \to \infty$, since $\pi^1_\#\boldsymbol{\mu}_n$ has uniformly integrable 1-moments by (5.1.20) we have

$$\limsup_{k\to\infty} \sup_n \int_{\{g\geq k\}} g\, d\boldsymbol{\mu}_n \leq \sup_n C\Big(\int_{\{|x_2|\geq m\}} |x_2|^q\, d\pi^2_\#\boldsymbol{\mu}_n\Big)^{1/q}$$

Letting $m \to \infty$ we conclude.

In the finite dimensional case (or even if $\boldsymbol{\mu}_n \to \boldsymbol{\mu}$ in $\mathscr{P}(X \times X)$) we conclude immediately, since the map $(x_1, x_2) \mapsto \langle x_1, x_2 \rangle$ is continuous in $X \times X$.

In the infinite dimensional case, let $\overline{B_R}$ be the centered closed ball of radius R in X which is compact in X_ϖ. The map $(x_1, x_2) \mapsto \langle x_1, x_2 \rangle$ is continuous in each closed set $X \times B_R$ with respect to the $X \times X_\varpi$ topology and (5.2.15) yields

$$\limsup_{n,R\to\infty} \boldsymbol{\mu}_n\big(X^2 \setminus (X \times B_R)\big) = 0.$$

Therefore we conclude by invoking Proposition 5.1.10. □

5.3 Measure-valued maps and disintegration theorem

Let X, Y be separable metric spaces and let $x \in X \mapsto \mu_x \in \mathscr{P}(Y)$ be a measure-valued map. We say that μ_x is a Borel map if $x \mapsto \mu_x(B)$ is a Borel map for any Borel set $B \subset Y$, or equivalently if this property holds for any open set $A \subset Y$. By the monotone class theorem we have also that

$$x \in X \mapsto \int_Y f(x,y)\, d\mu_x(y) \text{ is Borel} \tag{5.3.1}$$

for every bounded (or nonnegative) Borel function $f : X \times Y \to \mathbb{R}$.

By (5.3.1) the formula

$$\boldsymbol{\mu}(f) = \int_X \left(\int_Y f(x,y)\, d\mu_x(y) \right) d\nu(x)$$

defines for any $\nu \in \mathscr{P}(X)$ a unique measure $\boldsymbol{\mu} \in \mathscr{P}(X \times Y)$, that will be denoted by $\int_X \mu_x \, d\nu(x)$. Actually any $\boldsymbol{\mu} \in \mathscr{P}(X \times Y)$ whose first marginal is ν can be represented in this way. This is implied by the so-called *disintegration theorem* (related to the existence of conditional probability measures in Probability), see for instance [67, III-70].

Theorem 5.3.1 (Disintegration). *Let $\boldsymbol{X}$, X be Radon separable metric spaces, $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$, let $\pi : \boldsymbol{X} \to X$ be a Borel map and let $\nu = \pi_\# \boldsymbol{\mu} \in \mathscr{P}(X)$. Then there exists a ν-a.e. uniquely determined Borel family of probability measures $\{\mu_x\}_{x\in X} \subset \mathscr{P}(\boldsymbol{X})$ such that*

$$\mu_x(\boldsymbol{X} \setminus \pi^{-1}(x)) = 0 \quad \textit{for } \nu\textit{-a.e. } x \in X \tag{5.3.2}$$

and

$$\int_{\boldsymbol{X}} f(\boldsymbol{x})\, d\boldsymbol{\mu}(\boldsymbol{x}) = \int_X \Big(\int_{\pi^{-1}(x)} f(\boldsymbol{x})\, d\mu_x(\boldsymbol{x}) \Big) d\nu(x) \tag{5.3.3}$$

for every Borel map $f : \boldsymbol{X} \to [0, +\infty]$.

In particular, when $\boldsymbol{X} := X_1 \times X_2, X := X_1$, $\boldsymbol{\mu} \in \mathscr{P}(X_1 \times X_2)$, $\nu = \mu^1 = \pi^1_\# \boldsymbol{\mu}$, we can canonically identify each fiber $(\pi^1)^{-1}(x_1)$ with X_2 and find a Borel family of probability measures $\{\mu_{x_1}\}_{x_1 \in X_1} \subset \mathscr{P}(X_2)$ (which is μ^1-a.e. uniquely determined) such that $\boldsymbol{\mu} := \int_{X_1} \mu_{x_1} \, d\mu^1(x_1)$.

As an application of the disintegration theorem we can prove existence, and in some cases uniqueness, of multiple plans with given marginals.

Lemma 5.3.2. *Let X_1, X_2, X_3 be Radon separable metric spaces and let $\boldsymbol{\gamma}^{1\,2} \in \mathscr{P}(X_1 \times X_2)$, $\boldsymbol{\gamma}^{1\,3} \in \mathscr{P}(X_1 \times X_3)$ such that $\pi^1_\# \boldsymbol{\gamma}^{1\,2} = \pi^1_\# \boldsymbol{\gamma}^{1\,3} = \mu^1$. Then there exists $\boldsymbol{\mu} \in \mathscr{P}(X_1 \times X_2 \times X_3)$ such that*

$$\pi^{1,2}_\# \boldsymbol{\mu} = \boldsymbol{\gamma}^{1\,2}, \qquad \pi^{1,3}_\# \boldsymbol{\mu} = \boldsymbol{\gamma}^{1\,3}. \tag{5.3.4}$$

Moreover, if $\boldsymbol{\gamma}^{1\,2} = \int \gamma^{1\,2}_{x_1} \, d\mu^1$, $\boldsymbol{\gamma}^{1\,3} = \int \gamma^{1\,3}_{x_1} \, d\mu^1$ and $\boldsymbol{\mu} = \int \boldsymbol{\mu}_{x_1} \, d\mu^1$ are the disintegrations of $\boldsymbol{\gamma}^{1\,2}$, $\boldsymbol{\gamma}^{1\,3}$ and $\boldsymbol{\mu}$ with respect to μ^1, (5.3.4) *is equivalent to*

$$\boldsymbol{\mu}_{x_1} \in \Gamma(\gamma^{1\,2}_{x_1}, \gamma^{1\,3}_{x_1}) \subset \mathscr{P}(X_2 \times X_3) \quad \text{for } \mu^1\text{-a.e. } x_1 \in X_1. \tag{5.3.5}$$

In particular (5.2.12) *implies that the measure $\boldsymbol{\mu}$ is unique if either $\boldsymbol{\gamma}^{1\,2}$ or $\boldsymbol{\gamma}^{1\,3}$ are induced by a transport. We denote by $\Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3})$ the subset of plans $\boldsymbol{\mu} \in \mathscr{P}(X_1 \times X_2 \times X_3)$ satisfying* (5.3.4).

Proof. With the notation introduced in the statement of the theorem, the measure $\boldsymbol{\mu}$ whose disintegration w.r.t. x_1 is

$$\int_{X_1} \gamma^{1\,2}_{x_1} \times \gamma^{1\,3}_{x_1} \, d\mu^1(x_1)$$

has the required properties.

Now we prove the equivalence between (5.3.4) and (5.3.5). If $\boldsymbol{\mu}$ satisfies $\pi^{1,2}_\# \boldsymbol{\mu} = \boldsymbol{\gamma}^{1\,2}$ and $\pi^{1,3}_\# \boldsymbol{\mu} = \boldsymbol{\gamma}^{1\,3}$, then

$$\boldsymbol{\gamma}^{1\,2} = \pi^{1,2}_\# \boldsymbol{\mu} = \int_{X_1} \pi^2_\# \boldsymbol{\mu}_{x_1} \, d\mu^1(x_1)$$

and the uniqueness of the disintegration gives $\pi^2_\# \boldsymbol{\mu}_{x_1} = \gamma^{1\,2}_{x_1}$ for μ^1-a.e. $x_1 \in X_1$. A similar argument gives that $\pi^3_\# \boldsymbol{\mu}_{x_1} = \gamma^{1\,3}_{x_1}$ for μ^1-a.e. $x_1 \in X_1$.

Conversely, let us suppose that $\boldsymbol{\mu}$ satisfies (5.3.5) and let $f : X_1 \times X_2 \to \mathbb{R}$

be a bounded Borel function; the computation

$$\begin{aligned}
\int_{X_1\times X_2} f(x_1,x_2)\,d\pi^{1,2}_{\#}\boldsymbol{\mu} &= \int_{X_1\times X_2\times X_3} f(x_1,x_2)\,d\boldsymbol{\mu}(x_1,x_2,x_3)\\
&= \int_{X_1}\Big(\int_{X_2\times X_3} f(x_1,x_2)\,d\boldsymbol{\mu}_{x_1}(x_2,x_3)\Big)\,d\mu^1(x_1)\\
&= \int_{X_1}\Big(\int_{X_2} f(x_1,x_2)\,d\pi^2_{\#}\boldsymbol{\mu}_{x_1}(x_2)\Big)\,d\mu^1(x_1)\\
&= \int_{X_1}\Big(\int_{X_2} f(x_1,x_2)\,d\gamma^{1\,2}_{x_1}(x_2)\Big)\,d\mu^1(x_1)\\
&= \int_{X_1\times X_2} f(x_1,x_2)\,d\boldsymbol{\gamma}^{1\,2}(x_1,x_2)
\end{aligned}$$

shows that $\pi^{1,2}_{\#}\boldsymbol{\mu}=\boldsymbol{\gamma}^{1\,2}$. A similar argument proves that $\pi^{1,3}_{\#}\boldsymbol{\mu}=\boldsymbol{\gamma}^{1\,3}$. □

Remark 5.3.3 (Composition of plans). An analogous situation occurs when $\boldsymbol{\gamma}^{1\,2}\in\mathscr{P}(X_1\times X_2)$ and $\boldsymbol{\gamma}^{2\,3}\in\mathscr{P}(X_2\times X_3)$. In this case we say that

$$\boldsymbol{\mu}\in\Gamma^2(\boldsymbol{\gamma}^{1\,2},\boldsymbol{\gamma}^{2\,3})\quad\text{if}\quad \pi^{1,2}_{\#}\boldsymbol{\mu}=\boldsymbol{\gamma}^{1\,2},\quad \pi^{2,3}_{\#}\boldsymbol{\mu}=\boldsymbol{\gamma}^{2\,3}. \tag{5.3.6}$$

Of course, $\Gamma^2(\boldsymbol{\gamma}^{1\,2},\boldsymbol{\gamma}^{2\,3})$ is not empty iff $\pi^2_{\#}\boldsymbol{\gamma}^{1\,2}=\pi^1_{\#}\boldsymbol{\gamma}^{2\,3}$. In this case, the measure $\pi^{1,3}_{\#}\boldsymbol{\mu}$, with $\boldsymbol{\mu}\in\Gamma^2(\boldsymbol{\gamma}^{1\,2},\boldsymbol{\gamma}^{2\,3})$ constructed as in the proof of Lemma 5.3.2, belongs by construction to $\Gamma(\mu^1,\mu^3)$; it will be called *composition* of $\boldsymbol{\gamma}^{2\,3}$ and $\boldsymbol{\gamma}^{1\,2}$ and denoted by $\boldsymbol{\gamma}^{2\,3}\circ\boldsymbol{\gamma}^{1\,2}$. We have then

$$\int_{X_1\times X_3} f(x_1,x_3)\,d(\boldsymbol{\gamma}^{2\,3}\circ\boldsymbol{\gamma}^{1\,2}) = \int_{X_2}\Big(\int_{X_1\times X_3} f(x_1,x_3)\,d\boldsymbol{\gamma}^{1\,2}_{x_2}\times\boldsymbol{\gamma}^{2\,3}_{x_2}\Big)\,d\mu^2(x_2) \tag{5.3.7}$$

for any bounded Borel function $f:X_1\times X_3\to\mathbb{R}$. The name is justified since in the case $\boldsymbol{\gamma}^{1\,2},\boldsymbol{\gamma}^{2\,3}$ are induced by the transports $\boldsymbol{r}^{1\,2},\boldsymbol{r}^{2\,3}$, then the plan $\boldsymbol{\gamma}^{2\,3}\circ\boldsymbol{\gamma}^{1\,2}$ is induced by the composition map $\boldsymbol{r}^{2\,3}\circ\boldsymbol{r}^{1\,2}$: this fact can be easily checked starting from (5.3.7)

$$\begin{aligned}
\int_{X_1\times X_3} f(x_1,x_3)\,d(\boldsymbol{\gamma}^{2\,3}\circ\boldsymbol{\gamma}^{1\,2}) &= \int_{X_2}\Big(\int_{X_1} f(x_1,\boldsymbol{r}^{2\,3}(x_2))\,d\gamma^{1\,2}_{x_2}(x_1)\Big)\,d\mu^2(x_2)\\
&= \int_{X_1\times X_2} f(x_1,\boldsymbol{r}^{2\,3}(x_2))\,d\boldsymbol{\gamma}^{1\,2}(x_1,x_2)\\
&= \int_{X_1} f(x_1,\boldsymbol{r}^{2\,3}(\boldsymbol{r}^{1\,2}(x_1)))\,d\mu^1(x_1).
\end{aligned}$$

Notice that by (5.2.12) this construction is canonical only if either $(\boldsymbol{\gamma}^{1\,2})^{-1}$ or $\boldsymbol{\gamma}^{2\,3}$ are induced by a transport.

In the proof of the completeness of the Wasserstein distance we will also need the following useful extensions of Lemma 5.3.2 to a countable product of Radon spaces.

Lemma 5.3.4. *Let X_i, $i \in \mathbb{N}$, be a sequence of Radon separable metric spaces, $\mu^i \in \mathscr{P}(X_i)$ and $\boldsymbol{\alpha}^{i(i+1)} \in \Gamma(\mu^i, \mu^{i+1}), \boldsymbol{\beta}^{1\,i} \in \Gamma(\mu^1, \mu^i)$. Let $\boldsymbol{X}_\infty := \Pi_{i\in\mathbb{N}} X_i$, with the canonical product topology. Then there exist $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathscr{P}(\boldsymbol{X}_\infty)$ such that*

$$\pi^{i,i+1}_{\#}\boldsymbol{\mu} = \boldsymbol{\alpha}^{i\,(i+1)}, \quad \pi^{1,i}_{\#}\boldsymbol{\nu} = \boldsymbol{\beta}^{1\,i} \quad \forall\, i \in \mathbb{N}. \tag{5.3.8}$$

Proof. Let $\boldsymbol{X}_n := \Pi_{i=1}^n X_i = \boldsymbol{X}_{n-1} \times X_n$ and let $\pi^{\boldsymbol{n}} : \boldsymbol{X}_m \to \boldsymbol{X}_n$, $m \geq n$, be the projection onto the first n coordinates. In order to show the existence of $\boldsymbol{\mu}$, we set $\boldsymbol{\mu}^2 := \boldsymbol{\alpha}^{1\,2}$ and we apply recursively Lemma 5.3.2 and Remark 5.3.3 with $\boldsymbol{\mu}^n \in \mathscr{P}(\boldsymbol{X}_{n-1} \times X_n), \boldsymbol{\alpha}^{n(n+1)} \in \mathscr{P}(X_n \times X_{n+1})$, $n \geq 2$, to obtain a sequence $\boldsymbol{\mu}^{n+1} \in \mathscr{P}(\boldsymbol{X}_{n+1})$ satisfying

$$\pi^{\boldsymbol{n}}_{\#}\boldsymbol{\mu}^{n+1} = \boldsymbol{\mu}^n, \quad \pi^{n,n+1}_{\#}\boldsymbol{\mu}^{n+1} = \boldsymbol{\alpha}^{n(n+1)}.$$

Kolmogorov's Theorem [67, §51] provides a measure $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X}_\infty)$ such that $\pi^{\boldsymbol{n}}_{\#}\boldsymbol{\mu} = \boldsymbol{\mu}^n$ and therefore

$$\pi^{n-1,n}_{\#}\boldsymbol{\mu} = \pi^{n-1,n}_{\#}\big(\pi^{\boldsymbol{n}}_{\#}\boldsymbol{\mu}\big) = \pi^{n-1,n}_{\#}\boldsymbol{\mu}^n = \boldsymbol{\alpha}^{(n-1)\,n}.$$

The existence of $\boldsymbol{\nu}$ can be proved by a similar argument, by setting $\boldsymbol{\nu}^2 := \boldsymbol{\beta}^{1\,2}$ and by applying recursively Lemma 5.3.2 to $\boldsymbol{\nu}^n \in \mathscr{P}(X_1 \times \boldsymbol{X}_{n-1}), \boldsymbol{\beta}^{1(n+1)} \in \mathscr{P}(X_1 \times X_{n+1})$, $n \geq 2$: we can find a sequence $\boldsymbol{\nu}^{n+1} \in \mathscr{P}(\boldsymbol{X}_{n+1})$ satisfying

$$\pi^{\boldsymbol{n}}_{\#}\boldsymbol{\nu}^{n+1} = \boldsymbol{\nu}^n, \quad \pi^{1,n+1}_{\#}\boldsymbol{\nu}^{n+1} = \boldsymbol{\beta}^{1(n+1)}.$$

Kolmogorov's Theorem [67, §51] provides a measure $\boldsymbol{\nu} \in \mathscr{P}(\boldsymbol{X}_\infty)$ such that $\pi^{\boldsymbol{n}}_{\#}\boldsymbol{\nu} = \boldsymbol{\nu}^n$ and therefore

$$\pi^{1,n}_{\#}\boldsymbol{\nu} = \pi^{1,n}_{\#}\big(\pi^{\boldsymbol{n}}_{\#}\boldsymbol{\nu}\big) = \pi^{1,n}_{\#}\boldsymbol{\nu}^n = \boldsymbol{\beta}^{1\,n} \qquad \square$$

5.4 Convergence of plans and convergence of maps

In this section we investigate the relation between the convergence of maps and the convergence of the associated plans.

Let us first recall that if $X, Y_1, \ldots, Y_k$ are separable metric spaces with $\boldsymbol{Y} := Y_1 \times \ldots \times Y_k$, $\mu \in \mathscr{P}(X)$, and $\boldsymbol{r}_i : X \to Y_i$, $i = 1, \cdots, k$, then the product map

$$\begin{gathered}\boldsymbol{r} := (\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_k) : X \to \boldsymbol{Y} \quad \text{is Borel (μ-measurable) iff} \\ \text{each map } \boldsymbol{r}_i : X \to Y_i \text{ is Borel (resp. μ-measurable).}\end{gathered} \tag{5.4.1}$$

In particular, if $\boldsymbol{r}, \boldsymbol{s} : X \to Y$ are μ-measurable, then their distance $d_Y(\boldsymbol{r}(\cdot), \boldsymbol{s}(\cdot))$ is a μ-measurable real map.

We can thus define the convergence in measure of a sequence of μ-measurable maps $\boldsymbol{r}_n : X \to Y$ to a μ-measurable map $\boldsymbol{r}$ by asking that

$$\lim_{n\to\infty} \mu\bigl(\{x \in X : d_Y(\boldsymbol{r}_n(x), \boldsymbol{r}(x)) > \varepsilon\}\bigr) = 0 \quad \forall\, \varepsilon > 0. \tag{5.4.2}$$

We can also introduce the L^p spaces (see e.g. [7])

$$\begin{aligned} L^p(\mu; Y) := \Bigl\{\boldsymbol{r} : X \to Y \ \mu\text{-measurable} : \int_X d_Y^p(\boldsymbol{r}(x), \bar y)\, d\mu(x) < +\infty \\ \text{for some (and thus any) } \bar y \in Y\Bigr\}. \end{aligned} \tag{5.4.3}$$

with the distance

$$\boldsymbol{d}(\boldsymbol{r}, \boldsymbol{s})_{L^p(\mu;Y)} := \Bigl(\int_X d_Y^p(\boldsymbol{r}(x), \boldsymbol{s}(y))\, d\mu(x)\Bigr)^{1/p}; \tag{5.4.4}$$

it is easy to check that $L^p(\mu; Y)$ is complete iff Y is complete. When Y is a (separable) Hilbert space and $p \geq 1$, then the above distance is induced by the norm

$$\|\boldsymbol{r}\|_{L^p(\mu;Y)} := \Bigl(\int_X |\boldsymbol{r}(x)|_Y^p\, d\mu(x)\Bigr)^{1/p}; \tag{5.4.5}$$

for $\boldsymbol{r} \in L^1(\mu; Y)$ the (vector valued) integral $\int_X \boldsymbol{r}(x)\, d\mu(x) \in Y$ of $\boldsymbol{r}$ is well defined and satisfies

$$\int_X \langle y, \boldsymbol{r}(x)\rangle\, d\mu(x) = \langle y, \int_X \boldsymbol{r}(x)\, d\mu(x)\rangle \quad \forall\, y \in Y, \tag{5.4.6}$$

$$\phi\Bigl(\int_X \boldsymbol{r}(x)\, d\mu(x)\Bigr) \leq \int_X \phi(\boldsymbol{r}(x))\, d\mu(x) \tag{5.4.7}$$

for every proper, convex and l.s.c. function $\phi : Y \to (-\infty, +\infty]$ (Jensen's inequality).

In the following lemma we consider first the case when the reference measure μ is fixed, and show the equivalence between narrow convergence of the plans $(\boldsymbol{i} \times \boldsymbol{r}_n)_\#\mu$ and convergence in measure and in $L^p(\mu; Y)$ of $\boldsymbol{r}_n$, when the limiting plan is induced by a transport $\boldsymbol{r}$.

Lemma 5.4.1 (Narrow convergence of plans and convergence in measure). *Let $\mu \in \mathscr{P}(X)$ and let $\boldsymbol{r}_n, \boldsymbol{r} : X \to Y$ be Borel maps. Then the plans $(\boldsymbol{i} \times \boldsymbol{r}_n)_\#\mu$ narrowly convergence to $(\boldsymbol{i} \times \boldsymbol{r})_\#\mu$ in $\mathscr{P}(X \times Y)$ as $n \to \infty$ if and only if $\boldsymbol{r}_n$ converge in measure to $\boldsymbol{r}$.*
Moreover, the measures $(\boldsymbol{r}_n)_\#\mu$ have uniformly integrable p-moments iff $\boldsymbol{r}_n$ converges to $\boldsymbol{r}$ in $L^p(\mu; Y)$.

Proof. Since for every Borel map $\boldsymbol{s} : X \to Y$

$$\int_{X\times Y} \varphi(x, y)\, d(\boldsymbol{i} \times \boldsymbol{s})_\#\mu = \int_X \varphi(x, \boldsymbol{s}(x))\, d\mu(x) \qquad \forall \varphi \in C_b^0(X \times Y)$$

and convergence in measure is stable by composition with continuous functions, it is immediate to check that convergence in measure of the maps implies narrow convergence of the plans.
The converse implication can be obtained as follows: fix $\epsilon > 0$, a continuous function ψ_ϵ with $0 \le \psi_\varepsilon \le 1$, $\psi_\varepsilon(0) = 0$ and $\psi_\varepsilon(t) = 1$ whenever $|t| > \epsilon$ and a continuous function $\tilde{\boldsymbol{r}}$ such that $\mu(\{\boldsymbol{r} \neq \tilde{\boldsymbol{r}}\}) < \epsilon$. Then, using the test function $\varphi_\epsilon(x,y) = \psi_\epsilon(d_Y(y,\tilde{\boldsymbol{r}}(x)))$ we obtain

$$\begin{aligned}\limsup_{n\to\infty} \mu(\{d_Y(\boldsymbol{r}_n,\tilde{\boldsymbol{r}}) > \epsilon\}) &\le \limsup_{n\to\infty} \int_{X\times Y} \varphi_\epsilon \, d(\boldsymbol{i}\times\boldsymbol{r}_n)_\#\mu = \int_{X\times Y} \varphi_\epsilon \, d(\boldsymbol{i}\times\boldsymbol{r})_\#\mu \\ &= \int_X \psi_\epsilon(d_Y(\boldsymbol{r}(x),\tilde{\boldsymbol{r}}(x)))\, d\mu(x) \le \epsilon.\end{aligned}$$

Taking into account our choice of $\tilde{\boldsymbol{r}}$ we obtain $\limsup\limits_{n\to\infty} \mu(\{d_Y(\boldsymbol{r}_n,\boldsymbol{r}) > \epsilon\}) \le 2\epsilon$.

The second part of the lemma follows easily by Vitali dominated convergence theorem and the identities

$$\begin{aligned}\lim_{n\to\infty} \int_X d_Y^p(\boldsymbol{r}_n(x),\bar{y})\, d\mu(x) &= \lim_{n\to\infty} \int_Y d_Y^p(y,\bar{y})\, d\big((\boldsymbol{r}_n)_\#\mu\big)(y) \\ = \int_X d_Y^p(\boldsymbol{r}(x),\bar{y})\, d\mu(x) &= \int_Y d_Y^p(y,\bar{y})\, d\big(\boldsymbol{r}_\#\mu\big)(y),\end{aligned} \tag{5.4.8}$$

which hold either if $\boldsymbol{r}_n$ converges to $\boldsymbol{r}$ in $L^p(\mu;Y)$ or if the family $(\boldsymbol{r}_n)_\#\mu$, $n \in \mathbb{N}$, has uniformly integrable p-moments. □

In the rest of this section we assume that X is a separable Hilbert space as in Section 5.1.2.

Definition 5.4.2 (Barycentric projection). *The barycentric projection $\bar{\boldsymbol{\gamma}} : X \to X$ of a plan $\boldsymbol{\gamma} \in \mathscr{P}(X\times X)$, which admits the disintegration $\boldsymbol{\gamma} = \int_X \gamma_{x_1}\, d\mu(x_1)$ with respect to its first marginal $\mu = \pi^1_\#\boldsymbol{\gamma}$, is defined as*

$$\bar{\boldsymbol{\gamma}}(x_1) := \int_X x_2 \, d\gamma_{x_1}(x_2) \quad \text{for } \mu\text{-a.e. } x_1 \in X \tag{5.4.9}$$

provided γ_{x_1} has finite first moment for μ-a.e. x_1.

Assume that we are given maps $\boldsymbol{v}_n \in L^p(\mu_n;X)$: here we have to be careful in the meaning of the convergence of vectors $\boldsymbol{v}_n$, which belong to different L^p-spaces. Two approaches seem natural:

(i) we can consider the narrow limit in $\mathscr{P}(X_\varpi)$ of the X-valued measures $\boldsymbol{\nu}_n := \boldsymbol{v}_n\mu_n$ (component by component);

(ii) we can consider the limit $\boldsymbol{\gamma}$ of the associated plans $\boldsymbol{\gamma}_n := (\boldsymbol{i}\times\boldsymbol{v}_n)_\#\mu_n$ in $\mathscr{P}_2(X_\varpi\times X_\varpi)$, recovering a limit vector v by taking the barycenter of $\boldsymbol{\gamma}$.

In fact, these two approaches yields equivalent notions: we formalize the point (i) in the following definition, and then we see that it coincides with (ii).

Definition 5.4.3. *Let $(\boldsymbol{e}_j)$ be an orthonormal basis of X, let $(\mu_n) \subset \mathscr{P}(X)$ be narrowly converging to μ in $\mathscr{P}(X_\varpi)$ and let $\boldsymbol{v}_n \in L^1(\mu_n; X)$. We say that $\boldsymbol{v}_n$ weakly converge to $\boldsymbol{v} \in L^1(\mu; X)$ if*

$$\lim_{n\to\infty} \int_X \zeta(x)\langle \boldsymbol{e}_j, \boldsymbol{v}_n(x)\rangle \, d\mu_n(x) = \int_X \zeta(x)\langle \boldsymbol{e}_j, \boldsymbol{v}(x)\rangle \, d\mu(x) \tag{5.4.10}$$

for every $\zeta \in \mathrm{Cyl}(X)$ and any $j \in \mathbb{N}$. We say that $\boldsymbol{v}_n$ converges strongly *to $\boldsymbol{v}$ in L^p, $p > 1$, if* (5.4.10) *holds and*

$$\limsup_{n\to\infty} \|\boldsymbol{v}_n\|_{L^p(\mu_n;X)} \le \|\boldsymbol{v}\|_{L^p(\mu;X)}. \tag{5.4.11}$$

It is easy to check that the limit $\boldsymbol{v}$, if it exists, is unique.

Theorem 5.4.4. *Let $p > 1$, let $(\mu_n) \subset \mathscr{P}(X)$ be narrowly converging to μ in $\mathscr{P}(X_\varpi)$ and let $\boldsymbol{v}_n \in L^p(\mu_n; X)$ be such that*

$$\sup_{n\in\mathbb{N}} \int_X |\boldsymbol{v}_n(x)|^p \, d\mu_n(x) < +\infty. \tag{5.4.12}$$

Then the following statements hold:

(i) *The family of plans $\boldsymbol{\gamma}_n := (\boldsymbol{i} \times \boldsymbol{v}_n)_\# \mu_n$ has limit points in $\mathscr{P}(X_\varpi \times X_\varpi)$ as $n \to \infty$ and the sequence $(\boldsymbol{v}_n)$ has weak limit points as $n \to \infty$.*

(ii) *$\boldsymbol{v}_n$ weakly converge to $\boldsymbol{v} \in L^p(\mu; X)$ according to Definition* 5.4.3 *if and only if $\boldsymbol{v}$ is the barycenter of any limit point of the sequence of plans $\boldsymbol{\gamma}_n$ in $\mathscr{P}(X_\varpi \times X_\varpi)$; in this case*

$$\liminf_{n\to\infty} \int_X g(\boldsymbol{v}_n(x)) \, d\mu_n(x) \ge \int_X g(\boldsymbol{v}(x)) \, d\mu(x), \tag{5.4.13}$$

for every convex *and l.s.c. function $g : X \to (-\infty, +\infty]$.*

(iii) *If $\boldsymbol{v}_n$ strongly converge to $\boldsymbol{v}$ in L^p then $\boldsymbol{\gamma}_n$ narrowly converge to $(\boldsymbol{i} \times \boldsymbol{v})_\# \mu$ in $\mathscr{P}(X_\varpi \times X)$ and*

$$\lim_{n\to\infty} \|v_n\|^p_{L^p(\mu_n;X)} = \lim_{n\to\infty} \int_{X^2} |x_2|^p \, d\boldsymbol{\gamma}_n = \|v\|^p_{L^p(\mu;X)}. \tag{5.4.14}$$

If, in addition, μ_n narrowly converge to μ in $\mathscr{P}(X)$ then $\boldsymbol{\gamma}_n$ narrowly converge to $(\boldsymbol{i} \times \boldsymbol{v})_\# \mu$ in $\mathscr{P}(X \times X)$. Finally, if μ_n has uniformly integrable p-moments, then

$$\lim_{n\to\infty} \int_X f(x, v_n(x)) \, d\mu_n(x) = \int_X f(x, v(x)) \, d\mu(x), \tag{5.4.15}$$

for every continuous function $f : X \times X \to \mathbb{R}$ with p-growth according to (5.1.21).

Proof. (i) Observe that Lemma 5.2.2 ensures that the sequence $(\boldsymbol{\gamma}_n)$ is relatively compact in $\mathscr{P}(X_\varpi \times X_\varpi)$, since (see also Lemma 5.1.12) $\pi^1_\# \boldsymbol{\gamma}_n = \mu_n \to \mu$ in $\mathscr{P}(X_\varpi)$ and $\pi^2_\# \boldsymbol{\gamma}_n$ is relatively compact in $\mathscr{P}(X_\varpi)$ by (5.4.12).

(ii) For every $j \in \mathbb{N}$ and any $\varphi \in \mathrm{Cyl}(X)$ we have

$$\int_X \varphi(x)\langle \boldsymbol{e}_j, \boldsymbol{v}_n(x)\rangle \, d\mu_n(x) = \int_{X\times X} \varphi(x_1)\langle \boldsymbol{e}_j, x_2\rangle \, d\boldsymbol{\gamma}_n(x_1, x_2).$$

Since $|x_2|$ is uniformly integrable w.r.t. $(\boldsymbol{\gamma}_n)$, Proposition 5.1.10 yields

$$\begin{aligned} \lim_{k\to\infty} \int_{X\times X} \varphi(x_1)\langle \boldsymbol{e}_j, x_2\rangle \, d\boldsymbol{\gamma}_{n_k}(x_1, x_2) &= \int_{X\times X} \varphi(x_1)\langle \boldsymbol{e}_j, x_2\rangle \, d\boldsymbol{\gamma}(x_1, x_2) \\ &= \int_X \varphi(x_1)\langle \boldsymbol{e}_j, \bar{\boldsymbol{\gamma}}(x_1)\rangle \, d\mu(x_1) \end{aligned}$$

for every subsequence $(\boldsymbol{\gamma}_{n_k})$ converging to $\boldsymbol{\gamma}$ in $\mathscr{P}(X_\varpi \times X_\varpi)$. Therefore, (5.4.10) holds if and only if $\boldsymbol{v} = \bar{\boldsymbol{\gamma}}$ for every limit point $\boldsymbol{\gamma}$.
(5.4.13) follows by Jensen's inequality and (5.1.33a), being g weakly lower semicontinuous.

(iii) If $\boldsymbol{\gamma}$ is a limit point of $\boldsymbol{\gamma}_n$ as in (ii), taking into account that $\boldsymbol{v} = \bar{\boldsymbol{\gamma}}$ we have

$$\int_{X\times X} |x_2|^p \, d\boldsymbol{\gamma} \le \liminf_{n\to\infty} \int_{X\times X} |x_2|^p \, d\boldsymbol{\gamma}_n = \int_X |\bar{\boldsymbol{\gamma}}|^p \, d\mu.$$

Hence, by disintegrating $\boldsymbol{\gamma}$ with respect to x_1 we get

$$\int_X \left(\int_X |x_2|^p \, d\boldsymbol{\gamma}_{x_1}\right)^p - |\bar{\boldsymbol{\gamma}}(x_1)|^p \, d\mu(x_1) = 0$$

and so Jensen's inequality gives that $\boldsymbol{\gamma}_{x_1} = \delta_{\boldsymbol{v}(x_1)}$ for μ-a.e. x_1, i.e. $\boldsymbol{\gamma} = (\boldsymbol{i} \times \boldsymbol{v})_\# \mu$. This proves the narrow convergence of $\boldsymbol{\gamma}_n$ to $\boldsymbol{\gamma}$ in $\mathscr{P}(X_\varpi \times X_\varpi)$ and (5.4.14). By applying Theorem 5.1.13 we obtain that the second marginals of $\boldsymbol{\gamma}_n$ are also converging in the stronger narrow topology of $\mathscr{P}(X)$. Lemma 5.2.2 yields that the sequence $\boldsymbol{\gamma}_n$ is tight in $\mathscr{P}(X_\varpi \times X)$ and therefore converges to $\boldsymbol{\gamma}$ in $\mathscr{P}(X_\varpi \times X)$. The last part of the statement follows again by Lemma 5.2.2 and Lemma 5.1.7. □

5.5 Approximate differentiability and area formula in Euclidean spaces

Let $f : \mathbb{R}^d \to \mathbb{R}^d$ be a function. Then, denoting by $\Sigma = D(\nabla f)$ the Borel set where f is differentiable, there is a sequence of sets $\Sigma_n \uparrow \Sigma$ such that $f|_{\Sigma_n}$ is a Lipschitz function for any n (see [77, 3.1.8]). Therefore the well-known area formula for Lipschitz maps (see for instance [75, 77]) extends to this general class of maps and reads as follows:

$$\int_\Sigma h(x)|\det \nabla f|(x) \, dx = \int_{\mathbb{R}^d} \sum_{x\in\Sigma\cap f^{-1}(y)} h(x) \, dy \tag{5.5.1}$$

for any Borel function $h : \mathbb{R}^d \to [0, +\infty]$. Actually, these results hold more generally for the *approximately differentiable maps*, whose definition and main properties are recalled below.

Definition 5.5.1 (Approximate limit and approximate differential). *Let $\Omega \subset \mathbb{R}^d$ be an open set and $f : \Omega \to \mathbb{R}^m$. We say that f has an approximate limit (respectively, approximate differential) at $x \in \Omega$ if there exists a function $g : \Omega \to \mathbb{R}^m$ continuous (resp. differentiable) at x such that the set $\{f \neq g\}$ has density 0 at x. In this case the approximate limit (resp. approximate differential) will be denoted by $\tilde{f}(x)$ (resp. $\tilde{\nabla} f(x)$).*

It is immediate to check that the definition above is well posed, i.e. it does not depend on the choice of g. An equivalent and more traditional (see [77]) definition of approximate limit goes as follows: we say that $z \in \mathbb{R}^m$ is the approximate limit of f at x if all sets

$$\{y : \ |f(y) - z| > \epsilon\} \qquad \epsilon > 0$$

have density 0 at x. Analogously, a linear map $L : \mathbb{R}^d \to \mathbb{R}^m$ is said to be the approximate differential of f at x if f has an approximate limit at x and all sets

$$\left\{y : \ \frac{|f(y) - \tilde{f}(x) - L(y - x)|}{|y - x|} > \epsilon\right\} \qquad \epsilon > 0$$

have density 0 at x.
The latter definitions have the advantage of being more intrinsic and do not rely on an auxiliary function g. We have chosen the former definitions because they are more practical, as we will see, for our purposes. For instance, a property that immediately follows by the definition, and that will be used very often in the sequel, is the *locality* principle: if f has approximate limit $\tilde{f}(x)$ (resp. approximate differential $\tilde{\nabla} f(x)$) for any $x \in B$, with B Borel, then g has approximate limit (resp. approximate differential) equal to $\tilde{f}(x)$ (resp. $\tilde{\nabla} f(x)$) for $\mathscr{L}^d$-a.e. $x \in B$, and precisely at all points x where the coincidence set $B \cap \{f = g\}$ has density 1.

Remark 5.5.2. Recall that if $f : \Omega \to \mathbb{R}^m$ is $\mathscr{L}^d$-measurable, then it has approximate limit $\tilde{f}(x)$ at $\mathscr{L}^d$-a.e. $x \in \Omega$ and $f(x) = \tilde{f}(x)$ $\mathscr{L}^d$-a.e.. In particular every Lebesgue measurable set B has density 1 at $\mathscr{L}^d$-a.e. point of B.

Denoting by Σ_f the Borel set (see for instance [7]) of points where f is approximately differentiable, it is still true by [77, 3.1.8] that there exists a sequence of sets $\Sigma_n \uparrow \Sigma_f$ such that $\tilde{f}|_{\Sigma_n}$ is a Lipschitz function for any n. By Mc Shane theorem we can extend $\tilde{f}|_{\Sigma_n}$ to Lipschitz functions g_n defined on the whole of $\mathbb{R}^d$. In the case $m = d$, by applying the area formula to g_n on Σ_n and noticing that (by definition) $\nabla g_n = \tilde{\nabla} f$ $\mathscr{L}^d$-a.e. on Σ_n we obtain

$$\int_{\Sigma_f} h(x) |\det \tilde{\nabla} f|(x) \, dx = \int_{\mathbb{R}^d} \sum_{x \in \Sigma_f \cap \tilde{f}^{-1}(y)} h(x) \, dy \qquad (5.5.2)$$

for any Borel function $h : \mathbb{R}^d \to [0,+\infty]$.

This formula leads to a simple rule for computing the density of the push-forward of measures absolutely continuous w.r.t. $\mathscr{L}^d$.

Lemma 5.5.3 (Density of the push-forward). *Let $\rho \in L^1(\mathbb{R}^d)$ be a nonnegative function and assume that there exists a Borel set $\Sigma \subset \Sigma_f$ such that $\tilde{f}|_\Sigma$ is injective and the difference $\{\rho > 0\} \setminus \Sigma$ is $\mathscr{L}^d$-negligible. Then $f_\#\big(\rho\mathscr{L}^d\big) \ll \mathscr{L}^d$ if and only if $|\det\tilde{\nabla} f| > 0$ $\mathscr{L}^d$-a.e. on Σ and in this case*

$$f_\#\big(\rho\mathscr{L}^d\big) = \frac{\rho}{|\det\tilde{\nabla} f|} \circ \tilde{f}^{-1}|_{f(\Sigma)}\mathscr{L}^d.$$

Proof. If $|\det\tilde{\nabla} f| > 0$ $\mathscr{L}^d$-a.e. on Σ we can put $h = \rho\chi_{\tilde{f}^{-1}(B)\cap\Sigma}/|\det\tilde{\nabla} f|$ in (5.5.2), with $B \in \mathscr{B}(\mathbb{R}^d)$, to obtain

$$\int_{\tilde{f}^{-1}(B)} \rho\,dx = \int_{\tilde{f}^{-1}(B)\cap\Sigma} \rho\,dx = \int_{B\cap\tilde{f}(\Sigma)} \frac{\rho(\tilde{f}^{-1}(y))}{|\det\tilde{\nabla} f(\tilde{f}^{-1}(y))|}\,dy.$$

Conversely, if there is a Borel set $B \subset \Sigma$ with $\mathscr{L}^d(B) > 0$ and $|\det\tilde{\nabla} f| = 0$ on B the area formula gives $\mathscr{L}^d(\tilde{f}(B)) = 0$. On the other hand

$$f_\#(\rho\mathscr{L}^d)(\tilde{f}(B)) = \int_{f^{-1}(\tilde{f}(B))} \rho\,dx > 0$$

because at $\mathscr{L}^d$-a.e. $x \in B$ we have $f(x) = \tilde{f}(x)$ and $\rho(x) > 0$. Hence $f_\#(\rho\mathscr{L}^d)$ is not absolutely continuous with respect to $\mathscr{L}^d$. □

By applying the area formula again we obtain the rule for computing integrals of the densities:

$$\int_{\mathbb{R}^d} F\left(\frac{f_\#(\rho\mathscr{L}^d)}{\mathscr{L}^d}\right)dx = \int_{\mathbb{R}^d} F\left(\frac{\rho}{|\det\tilde{\nabla} f|}\right)|\det\tilde{\nabla} f|\,dx \tag{5.5.3}$$

for any Borel function $F : \mathbb{R} \to [0,+\infty]$ with $F(0) = 0$. Notice that in this formula the set Σ does not appear anymore (due to the fact that $F(0) = 0$ and $\rho = 0$ out of Σ), so it holds provided f is approximately differentiable $\rho\mathscr{L}^d$-a.e., it is $\rho\mathscr{L}^d$-essentially injective (i.e. there exists a Borel set Σ such that $\tilde{f}|_\Sigma$ is injective and $\rho = 0$ $\mathscr{L}^d$-a.e. out of Σ) and $|\det\tilde{\nabla} f| > 0$ $\rho\mathscr{L}^d$-a.e.

We will apply mostly these formulas when f is the gradient of a convex function g (corresponding to optimal transport map for the quadratic cost function), or is an optimal transport map. In the former case actually approximate differentiability is not needed thanks to the following result (see for instance [4, 75]).

Theorem 5.5.4 (Aleksandrov). *Let $g : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then ∇g is differentiable $\mathscr{L}^d$-a.e. in its domain, its gradient $\nabla^2 g(x)$ is a symmetric matrix for $\mathscr{L}^d$-a.e. $x \in \mathbb{R}^d$, and g has second order Taylor expansion*

$$g(y) = g(x) + \langle\nabla g(x), y - x\rangle + \frac{1}{2}\langle\nabla^2 g(x), y - x\rangle + o(|y - x|^2) \quad \text{as } y \to x. \tag{5.5.4}$$

Notice that ∇g is also monotone

$$\langle \nabla g(x_1) - \nabla g(x_2), x_1 - x_2 \rangle \geq 0 \qquad x_1,\, x_2 \in D(\nabla g),$$

and that the above inequality is strict if g is *strictly* convex: in this case, it is immediate to check that ∇g is injective on $D(\nabla g)$, and that $|\det \nabla^2 g| > 0$ on the differentiability set of ∇g if g is *uniformly* convex.