Weak topology and Opial property in Wasserstein spaces, with applications to Gradient Flows and Proximal Point Algorithms of geodesically convex functionals

Emanuele Naldi [*]

Institut für Analysis und Algebra, TU Braunschweig.

Giuseppe Savaré [†]

Department of Decision Sciences and BIDSA, Bocconi University.

Dedicated to the memory of Claudio Baiocchi, outstanding mathematician and beloved mentor

Abstract

In this paper we discuss how to define an appropriate notion of *weak topology* in the Wasserstein space $(\mathcal{P}_2(\mathsf{H}), W_2)$ of Borel probability measures with finite quadratic moment on a separable Hilbert space H.

We will show that such a topology inherits many features of the usual weak topology in Hilbert spaces, in particular the weak closedness of geodesically convex closed sets and the Opial property characterising weakly convergent sequences.

We apply this notion to the approximation of fixed points for a non-expansive map in a weakly closed subset of $\mathcal{P}_2(\mathsf{H})$ and of minimizers of a lower semicontinuous and geodesically convex functional $\phi : \mathcal{P}_2(\mathsf{H}) \to (-\infty, +\infty]$ attaining its minimum. In particular, we will show that every solution to the Wasserstein gradient flow of ϕ weakly converge to a minimizer of ϕ as the time goes to $+\infty$. Similarly, if ϕ is also convex along generalized geodesics, every sequence generated by the proximal point algorithm converges to a minimizer of ϕ with respect to the weak topology of $\mathcal{P}_2(\mathsf{H})$.

1 Introduction

Opial proved in [10] that weak convergence in a separable Hilbert space $(\mathsf{H}, |\cdot|)$ admits a nice metric characterization.

[*]Universitätsplatz 2, 38106 Braunschweig, Germany. Email: e.naldi@tu-braunschweig.de

[†]Via Roentgen 1, 20136 Milan, Italy. Email: giuseppe.savare@unibocconi.it

Theorem (Opial). *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in H weakly converging to $x \in \mathsf{H}$. Then*

$$|y-x|^2 + \liminf_{n\to\infty} |x_n - x|^2 \le \liminf_{n\to\infty} |x_n - y|^2. \tag{1.1}$$

In particular, for every $y \neq x$

$$\liminf_{n\to\infty} |x_n - x| < \liminf_{n\to\infty} |x_n - y|. \tag{1.2}$$

The *proof* can be easily obtained by passing to the limit in the identity

$$|x_n - y|^2 = |x_n - x|^2 + |x-y|^2 + 2\langle x_n - x, x - y\rangle, \tag{1.3}$$

observing that $\lim_{n\to\infty}\langle x_n - x, x-y\rangle = 0$ by weak convergence. It is worth noticing that (1.2) shows that the weak limit x of a sequence $(x_n)_{n\in\mathbb{N}}$ is the unique strict minimizer of the function

$$L(y) := \liminf_{n\to\infty} |x_n - y|. \tag{1.4}$$

Opial property (further extended and studied in more general Banach spaces, see e.g. [11]) has many interesting applications. A first one, which already appears as a relevant motivation in Opial's paper [10], is related to the approximation of a fixed point of a non-expansive map $T : C \to C$ defined in a closed and convex subset of H. If the set of fixed points of T is not empty and $\lim_{n\to\infty} |T^{n+1}x - T^n x| = 0$ for some $x \in C$, then the sequence of iterated maps $(T^n x)_{n\in\mathbb{N}}$ weakly converges to a fixed point $y \in C$ of T as $n \to \infty$.

A second kind of applications concerns the dynamic approximation of minimizers of a convex and lower semicontinuous function $\varphi : \mathsf{H} \to (-\infty, +\infty]$ as the asymptotic limit of its gradient flow or of the so called Proximal Point Algorithm.

More precisely, if $\arg\min \varphi \neq \emptyset$, Bruck [3] proved that every locally Lipschitz curve $x : (0, +\infty) \to \mathsf{H}$ solving the differential inclusion (the gradient flow generated by φ)

$$\frac{\mathrm{d}}{\mathrm{d}t} x(t) \in -\partial\varphi(x(t)) \quad \text{a.e. in } (0,\infty), \tag{1.5}$$

weakly converges to a minimizer x_∞ of φ as $t \uparrow \infty$. An analogous asymptotic behaviour is exhibited by the solutions to the Proximal Point Algorithm: selecting an initial datum $x_0 \in \mathsf{H}$ and a time step $\tau > 0$, one considers the sequence $(x^k_\tau)_{k\in\mathbb{N}}$ which recursively solves the variational problem

$$x^k_\tau \quad \text{minimizes} \quad y \mapsto \frac{1}{2\tau}|y - x^{k-1}_\tau|^2 + \varphi(y). \tag{1.6}$$

A result of Martinet [6, 7] (see also Rockafellar [12]) shows that the sequence $(x^k_\tau)_{k\in\mathbb{N}}$ weakly converges to an element x_∞ of $\arg\min \varphi$.

The aim of the present paper is to study the extension of the Opial Lemma to the metric space $(\mathcal{P}_2(\mathsf{H}), W_2)$ of Borel probability measures on H endowed with the Kantorovich-Rubinstein-Wasserstein distance W_2 and to derive similar applications to fixed points and to convergence of gradient flows and proximal point algorithms.

Let us recall that a Borel probability measure μ in H belongs to $\mathcal{P}_2(\mathsf{H})$ if

$$\text{the quadratic moment } \int_{\mathsf{H}} |x|^2 \,\mathrm{d}\mu(x) \quad \text{is finite.} \tag{1.7}$$

The squared Wasserstein distance between $\mu_1, \mu_2 \in \mathcal{P}_2(\mathsf{H})$ can then be defined as the solution of the Optimal Problem with quadratic cost

$$W_2^2(\mu_1, \mu_2) := \min\Big\{ \int_{\mathsf{H}\times\mathsf{H}} |x_1 - x_2|^2 \,\mathrm{d}\boldsymbol{\mu}(x_1, x_2) : \boldsymbol{\mu} \in \Gamma(\mu_1, \mu_2)\Big\},$$

where $\Gamma(\mu_1, \mu_2)$ denotes the set of couplings between μ_1 and μ_2, i.e. the Borel probability measures in $\mathsf{H} \times \mathsf{H}$ whose marginals are μ_1 and μ_2 respectively. It turns out that $(\mathcal{P}_2(\mathsf{H}), W_2)$ is a complete and separable metric space, which contains an isometric copy of H given by the Dirac masses $\{\delta_x : x \in \mathsf{H}\}$ (see e.g. [15, 1, 13]).

Since the distance in $\mathcal{P}_2(\mathsf{H})$ cannot be derived by a norm, a first natural question concerns the appropriate definition of a suitable weak topology in $\mathcal{P}_2(\mathsf{H})$, which enjoys at least some of the most useful properties of weak convergence in Hilbert spaces:

(a) bounded sequences admit weakly convergent subsequences,

(b) the distance function from a given element is a weakly lower semicontinuous map,

(c) the scalar product is sequentially continuous w.r.t. strong/weak convergence of their factors,

(d) weakly convergent sequences are bounded,

(e) strongly closed convex sets are also weakly closed.

A first approach, adopted in [1], is to work with the Wasserstein distance induced by a weaker metric on H, which metrizes the weak topology on bounded sets. This provides a satisfactory answer to the first three questions (a,b,c), at least for bounded sequences.

Here we rely on a different point of view, recalling that the topology of $\mathcal{P}_2(\mathsf{H})$ could be equivalently characterized as the initial topology induced by the family of real functions $F_\zeta : \mathcal{P}_2(\mathsf{H}) \to \mathbb{R}$ where

$$F_\zeta : \mu \to \int_{\mathsf{H}} \zeta \,\mathrm{d}\mu, \quad \zeta \in \mathrm{C}(\mathsf{H}),\ \sup_{x\in\mathsf{H}} \frac{\zeta(x)}{1 + |x|^2} < \infty, \tag{1.8}$$

i.e. the coarsest topology such that makes all the functions F_ζ in (1.8) continuous. We thus define the weak topology in $\mathcal{P}_2(\mathsf{H})$ as the initial topology $\sigma(\mathcal{P}_2(\mathsf{H}), \mathrm{C}_2^w(\mathsf{H}))$ induced by F_ζ as ζ varies in the set

$$\mathrm{C}_2^w(\mathsf{H}) := \Big\{\zeta : \mathsf{H} \to \mathbb{R} \text{ is sequentially weakly continuous}, \lim_{|x|\to\infty} \frac{\zeta(x)}{1+|x|^2} = 0\Big\}, \tag{1.9}$$

and we call $\mathcal{P}_2^w(\mathsf{H})$ the corresponding topological space $\big(\mathcal{P}_2(\mathsf{H}), \sigma(\mathcal{P}_2(\mathsf{H}), \mathrm{C}_2^w(\mathsf{H}))\big)$. In this way, $\mathcal{P}_2^w(\mathsf{H})$ inherits the weak* topology of a subset of the dual of the Banach space $\mathrm{C}_2^w(\mathsf{H})$.

We notice that the topology of $\mathcal{P}_2^w(\mathsf{H})$ is strictly coarser than the topology of $\mathcal{P}_2(\mathsf{H})$ even if H is finite dimensional: in this case $\mathrm{C}_2^w(\mathsf{H})$ contains all the continuous functions $\zeta : \mathsf{H} \to \mathbb{R}$ such that

$$\lim_{|x|\to\infty} \frac{\zeta(x)}{1+|x|^2} = 0. \tag{1.10}$$

If we consider, e.g., the sequence of measures $\mu_n = (1-\frac{1}{n})\delta_0 + \frac{1}{n}\delta_{\sqrt{n}}$ in $\mathcal{P}_2(\mathbb{R})$, it is easy to check that $\mu_n \to \mu = \delta_0$ in $\mathcal{P}_2^w(\mathbb{R})$ since for every $\zeta \in \mathrm{C}_2^w(\mathbb{R})$

$$\lim_{n\to\infty} \int_{\mathbb{R}} \zeta \,\mathrm{d}\mu_n = \lim_{n\to\infty} \Big((1-\frac{1}{n})\zeta(0) + \frac{1}{n}\zeta(\sqrt{n}) \Big) = \zeta(0) = \int_{\mathbb{R}} \zeta \,\mathrm{d}\mu,$$

where $\zeta(\sqrt{n})/n \to 0$ thanks to the asymptotic property of (1.10). On the other hand, choosing $\zeta(x) = x^2$, which clearly satisfies (1.8), we have $\int_{\mathbb{R}} x^2 \,\mathrm{d}\mu_n = 1$ for every n, so that the sequence $(\mu_n)_{n\in\mathbb{N}^+}$ has no limit points in $\mathcal{P}_2(\mathsf{H})$. When H has infinite dimension, the difference between the weak $\mathcal{P}_2^w(\mathsf{H})$ and the strong $\mathcal{P}_2(\mathsf{H})$ topology also affects sequences of measures with uniformly bounded support: if $(e_n)_{n\in\mathbb{N}}$ is a sequence of orthonormal vectors in H and $\mu_n := \delta_{e_n}$, we have $W_2(\mu_m, \mu_n) = \sqrt{2}$ for every $m, n \in \mathbb{N}$, so that $(\mu_n)_{n\in\mathbb{N}}$ has no limit points in $\mathcal{P}_2(\mathsf{H})$, whereas $\mu_n \to \delta_0$ in $\mathcal{P}_2^w(\mathsf{H})$ as $n \to \infty$ since $\zeta(e_n) \to \zeta(0)$ for every sequentially *weakly* continuous function ζ.

We will show that the weak topology of $\mathcal{P}_2^w(\mathsf{H})$ satisfies all the previous properties (a,…,e). In particular, we will prove that every lower semicontinuous geodesically convex function $\phi : \mathcal{P}_2(\mathsf{H}) \to (-\infty, +\infty]$ is also sequentially lower semicontinuous w.r.t. the weak topology $\sigma(\mathcal{P}_2(\mathsf{H}), \mathrm{C}_2^w(\mathsf{H}))$. As a byproduct, for every $\mu_0 \in \mathcal{P}_2(\mathsf{H})$ and $\tau > 0$ the Proximal Point Algorithm in $\mathcal{P}_2(\mathsf{H})$ (also known as JKO [5] or Minimizing Movement scheme [1])

$$\mu_\tau^k \quad \text{minimizes} \quad \mu \mapsto \frac{1}{2\tau} W_2^2(\mu, \mu_\tau^{k-1}) + \phi(\mu) \tag{1.11}$$

has always a solution $(\mu_\tau^k)_{k\in\mathbb{N}}$.

Property (c) deserves a further comment: at the level of the underlying Hilbert space H, the scalar product $\langle \cdot, \cdot \rangle$ is a bilinear map defined in $\mathsf{H}\times\mathsf{H}$, which is sequentially continuous with respect to the strong/weak product topology of H (we can denote by $\mathsf{H}_s \times \mathsf{H}_w$ the corresponding topological space). At the level of probability measures, it is then natural to study the sequential continuity of the map $\gamma \mapsto \int_{\mathsf{H}\times\mathsf{H}} \langle x, y \rangle \,\mathrm{d}\gamma$, defined for $\gamma \in \mathcal{P}_2(\mathsf{H}\times\mathsf{H})$, with respect to a sort of strong/weak topology weaker than the Wasserstein one. Since

$\mathcal{P}_2(\mathsf{H}\times\mathsf{H})$ cannot be identified with the product space $\mathcal{P}_2(\mathsf{H})\times\mathcal{P}_2(\mathsf{H})$, it is clear that it is not sufficient to introduce the weak topology $\mathcal{P}_2^w(\mathsf{H})$ on a single factor, but a more refined notion combining moment conditions and convergence in $\mathsf{H}_s\times\mathsf{H}_w$ is needed. We will call $\mathcal{P}_2^{sw}(\mathsf{H}\times\mathsf{H})$ such a topological space taylored on $\mathsf{H}_s\times\mathsf{H}_w$, whose properties we will also address in Section 3.

We will then show that the Opial property holds in $\mathcal{P}_2^w(\mathsf{H})$, with the same structure of (1.1): if $(\mu_n)_{n\in\mathbb{N}}$ is a sequence converging to μ in $\mathcal{P}_2^w(\mathsf{H})$ then

$$W_2^2(\nu,\mu)+\liminf_{n\to\infty}W_2^2(\mu,\mu_n)\le\liminf_{n\to\infty}W_2^2(\nu,\mu_n)\quad\text{for every }\nu\in\mathcal{P}_2(\mathsf{H}).\tag{1.12}$$

Applications to the asymptotic convergence of the gradient flows of a lower semicontinuous and geodesically convex functional $\phi:\mathcal{P}_2(\mathsf{H})\to(-\infty,+\infty]$ can then be easily derived by the same strategy of [3], by using the metric characterization of a solution $\mu:(0,\infty)\to D(\phi)$ of the gradient flow of ϕ in $\mathcal{P}_2(\mathsf{H})$ in terms of Evolution Variational Inequalities [1] (see [2] for such a metric approach to (1.5) in Hilbert spaces)

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}W_2^2(\mu_t,\sigma)\le\phi(\sigma)-\phi(\mu_t)\quad\mathscr{L}^1\text{-a.e. in }(0,\infty),\quad\text{for every }\sigma\in D(\phi).\tag{EVI}$$

Analogous results hold for the convergence of the Proximal Point Algorithm (here we use the discrete estimates of [1] for (1.11) assuming convexity along generalized geodesics) and for the approximation of the fixed point of a non-expansive and asymptotically regular map T defined in a weakly closed subset of $\mathcal{P}_2(\mathsf{H})$.

Plan of the paper. We will collect in Section 2 the main facts concerning optimal transport and Kantorovich-Rubinstein-Wasserstein distances; we adopt a general topological framework, in order to include convergence of Borel probability measures with respect to non metrizable topologies as the weak topology in a Hilbert space.

Section 3 is devoted to the definition of the weak topologies of $\mathcal{P}_2^w(\mathsf{H})$ and $\mathcal{P}_2^{sw}(\mathsf{H}\times\mathsf{H})$. In view of future possible applications, here we adopt a more general viewpoint, considering probability measures in product spaces $\mathsf{Z}=\mathsf{X}_s\times\mathsf{Y}_w$ where X_s is a Banach space with its strong topology and Y_w is a reflexive Banach space endowed with its weak topology under the general p-q growth condition

$$\int_{\mathsf{X}\times\mathsf{Y}}\Big(\|x\|_{\mathsf{X}}^p+\|y\|_{\mathsf{Y}}^q\Big)\,\mathrm{d}\boldsymbol{\mu}(x,y)<\infty$$

and leading to the space $\mathcal{P}_{pq}^{sw}(\mathsf{X}\times\mathsf{Y})$. This is quite useful to deal with the integration of bilinear forms; when $q=p=2$ and $\mathsf{X}=\mathsf{Y}=\mathsf{H}$ is an Hilbert space, we recover the case which is particularly relevant for our aims. As a byproduct, we will prove the stability of optimal couplings w.r.t. the convergence in $\mathcal{P}_{22}^{sw}(\mathsf{H}\times\mathsf{H})$.

Section 4 deals with the weak lower semicontinuity in $\mathcal{P}_2^w(\mathsf{H})$ of geodesically convex functionals. The Opial property in $\mathcal{P}_2^w(\mathsf{H})$ is discussed in Section 5.

The last Section 6 contains the applications to the asymptotic behaviour of gradient flows, of the Proximal Point Algorithm, and of the iteration of non-expansive and asymptotically regular maps in $\mathcal{P}_2(\mathsf{H})$.

Acknowledgments. G.S. gratefully acknowledges the support of the Institute of Advanced Study of the Technical University of Munich, of the MIUR-PRIN 2017 project *"Gradient flows, Optimal Transport and Metric Measure Structures"* and of the IMATI-CNR, Pavia. The authors also acknowledge the support of the Department of Mathematics “F. Casorati" of the University of Pavia during the preparation of a preliminary version of the paper.
E.N. acknowledges that this project has received funding from the European Union’s Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Marie Skłodowska-Curie Grant Agreement No. 861137

G.S. graduated and began his research under the scientific direction of Claudio Baiocchi, to whom he is deeply grateful and indebted for the generous and enlightening guidance, and his gentle humanity. Thanks to the originality of his ideas, the depth of his vision, the superb clarity and broad scope of his teaching, and his encouragement towards new and stimulating problems, Claudio Baiocchi had a fundamental influence on the development of the "Pavia School". To remain in the context of the present paper, here we will limit ourselves to quote his deep contribution to the study of evolution variational inequalities, in particular [2], which had a profound impact on numerous developments and later on the metric theory of gradient flows.

2 Preliminaries

2.1 Radon measures in completely regular topological spaces

Let $(X, \mathcal{T})$ be a Hausdorff topological space. We will denote by $\mathscr{B}(X)$ the Borel σ-algebra in X and by $\mathcal{P}(X)$ the set of all *Radon* probability measures, i.e. Borel probability measures satisfying the inner approximation property by compact sets

$$\forall B \in \mathscr{B}(X),\ \forall \varepsilon > 0 \qquad \exists K_\varepsilon \subset B \text{ compact such that } \mu(B \setminus K_\varepsilon) \le \varepsilon. \tag{2.1}$$

Recall that if $(X, \mathcal{T})$ is Polish (i.e. its topology is induced by a metric d such that (X, d) is complete and separable) or it is Lusin (i.e. X admits a Polish topology $\mathcal{T}'$ finer than $\mathcal{T}$) then every Borel probability measure on X is also Radon. We notice moreover that if μ is a Radon measure in a metric space X then its support $\operatorname{supp}(\mu)$ is separable.

We will mostly deal with Borel probability measures in separable Hilbert/Banach spaces (or dual of separable Banach spaces), possibly endowed with their weak/weak* topology. Clearly a separable Banach space is Polish; the dual of a separable Banach space with its weak* topology is a Lusin space [14, Theorem 7, page 112].

In order to define the natural weak topology of $\mathcal{P}(X)$ in such general settings, let us recall that a topological space $(X, \mathcal{T})$ is called *completely regular* if it is Hausdorff and for all closed set $F \subset X$ and for all $x_0 \in X \setminus F$ there exists $f \in \mathrm{C}_b(X)$ (the space of continuous and bounded real functions defined in $(X, \mathcal{T})$) such that $f(x_0) = 0$ and $f_{|_F} \geq 1$. It is worth noticing that every metric space (X, d) and every Hausdorff topological vector space (in particular every Banach space endowed with the weak or the weak* topology) are completely regular topological spaces.

Definition 2.1 (Narrow topology in $\mathcal{P}(X)$)**.** Let $(X, \mathcal{T})$ be a completely regular topological space. The *narrow topology* in $\mathcal{P}(X)$ is the coarsest topology in $\mathcal{P}(X)$ such that all the functionals $\mu \mapsto \int_X f \, \mathrm{d}\mu$, $f \in \mathrm{C}_b(X)$, are continuous. In particular, we say that a sequence $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(X)$ is *narrowly convergent* to $\mu \in \mathcal{P}(X)$ if

$$\lim_{n \to \infty} \int_X f(x) \, \mathrm{d}\mu_n(x) = \int_X f(x) \, \mathrm{d}\mu(x) \qquad \forall f \in \mathrm{C}_b(X). \tag{2.2}$$

Narrow topology in $\mathcal{P}(X)$ is also often called *weak topology*; since we will mostly deal with the case when X is a Banach space endowed with its strong or weak topology (and the corresponding topologies in $\mathcal{P}(X)$), we will adopt the term *narrow* in order to avoid possible misunderstandings.

The Prokhorov Theorem (see [4, III-59] for a proof) provides an important criterium for relative compactness w.r.t. the narrow topology.

Definition 2.2 (Tightness)**.** A subset $\mathcal{K} \subset \mathcal{P}(X)$ is *tight* if $\forall \varepsilon > 0$ there exists a compact set $K_\varepsilon \subset X$ such that $\mu(X \setminus K_\varepsilon) < \varepsilon$ for all $\mu \in \mathcal{K}$.

Theorem 2.3 (Prokhorov)**.** *Let $\mathcal{K}$ be a tight subset of $\mathcal{P}(X)$, then $\mathcal{K}$ is relatively compact in $\mathcal{P}(X)$ w.r.t. the narrow convergence. Conversely, if X is a Polish space, then every relatively compact (w.r.t. the narrow convergence) subset of $\mathcal{P}(X)$ is tight.*

The following result provides an useful integral condition for tightness.

Proposition 2.4. *$\mathcal{K} \subset \mathcal{P}(X)$ is tight if and only if there exists a function $\varphi : X \to [0, +\infty)$, with compact sublevels, such that*

$$\sup_{\mu \in \mathcal{K}} \int_X \varphi(x) \, \mathrm{d}\mu(x) \leq C < +\infty \tag{2.3}$$

2.2 Transport of measures

Let X_i, $i = 1, 2$, be Hausdorff topological spaces, $\mu \in \mathcal{P}(X_1)$ and $\boldsymbol{r} : X_1 \to X_2$ be a Lusin μ-measurable map (e.g. a continuous map); we denote by $\boldsymbol{r}_\sharp \mu \in \mathcal{P}(X_2)$ the *push-forward* of μ through $\boldsymbol{r}$, defined by

$$\boldsymbol{r}_\sharp \mu(B) := \mu(\boldsymbol{r}^{-1}(B)) \qquad \forall B \in \mathscr{B}(X_2). \tag{2.4}$$

A particularly important case is provided by the projection maps in product spaces. For an integer $N \geq 2$ and $i, j = 1, ..., N$, we denote by π^i and $\pi^{i,j}$ the projection operators defined on the product space $\mathbf{X} := X_1 \times ... \times X_N$ and respectively defined by

$$\pi^i : (x_1, ..., x_N) \mapsto x_i \in X_i, \qquad \pi^{i,j} : (x_1, ..., x_N) \mapsto (x_i, x_j) \in X_i \times X_j. \tag{2.5}$$

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

$$\Gamma(\mu^1, ..., \mu^N) := \{\boldsymbol{\mu} \in \mathcal{P}(\mathbf{X}) : \pi^i_\sharp \mu = \mu^i, i = 1, ..., N\}. \tag{2.6}$$

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

Remark 2.5. To every $\mu^1 \in \mathcal{P}(X_1)$ and every Lusin μ^1-measurable map $\boldsymbol{r} : X_1 \to X_2$ and we can associate the transport plan

$$\boldsymbol{\mu} := (\mathrm{id}_{X_1} \times \boldsymbol{r})_\sharp \mu^1 \in \Gamma(\mu^1, \boldsymbol{r}_\sharp \mu^1), \quad \text{where } \mathrm{id}_{X_1} : X_1 \to X_1 \text{ is the identity map.} \tag{2.7}$$

If $\boldsymbol{\mu}$ is representable as in (2.7) then we say that $\boldsymbol{\mu}$ is *induced* by $\boldsymbol{r}$.

The following gluing lemma guarantees the existence of multiple plans with given marginals.

Lemma 2.6. *Let X_1, X_2, X_3 be Lusin or metrizable spaces and let $\gamma^{12} \in \mathcal{P}(X_1 \times X_2)$, $\gamma^{13} \in \mathcal{P}(X_1 \times X_3)$ such that $\pi^1_\sharp \gamma^{12} = \pi^1_\sharp \gamma^{13} = \mu^1 \in \mathcal{P}(X_1)$. Then there exists $\boldsymbol{\gamma} \in \mathcal{P}(X_1 \times X_2 \times X_3)$ such that*

$$\pi^{12}_\sharp \boldsymbol{\gamma} = \gamma^{12}, \qquad \pi^{13}_\sharp \boldsymbol{\gamma} = \gamma^{13}. \tag{2.8}$$

We denote by $\Gamma^1(\gamma^{1\,2}, \gamma^{1\,3})$ the subset of plans $\mu \in \mathcal{P}(X_1 \times X_2 \times X_3)$ satisfying (2.8).

2.3 Optimal Transport and Kantorovich-Rubinstein-Wasserstein spaces

Let (X, d) be a metric space and let $p \in [1, +\infty)$. We say that a Radon measure $\mu \in \mathcal{P}(X)$ belongs to $\mathcal{P}_p(X)$ if

$$\int_X \mathsf{d}^p(x, x_o)\, \mathrm{d}\mu(x) < +\infty \quad \text{for some (and thus any) } x_o \in X. \tag{2.9}$$

Definition 2.7. The *L^p-Kantorovich-Rubinstein-Wasserstein distance W_p* between two Radon probability measures $\mu^1, \mu^2 \in \mathcal{P}_p(X)$ is defined by

$$W_p^p(\mu^1, \mu^2) := \min\Big\{ \int \mathsf{d}^p(x_1, x_2)\, \mathrm{d}\boldsymbol{\mu}(x_1, x_2) : \boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)\Big\}.$$

We denote by $\Gamma_o(\mu^1, \mu^2) \subset \Gamma(\mu^1, \mu^2)$ the convex and narrowly compact set of optimal plans where the minimum is attained, i.e.

$$\boldsymbol{\gamma} \in \Gamma_o(\mu^1, \mu^2) \iff \int \mathsf{d}^p(x_1, x_2)\, \mathrm{d}\boldsymbol{\gamma}(x_1, x_2) = W_p^p(\mu^1, \mu^2).$$

It is possible to prove that Γ_o is not empty and $(\mathcal{P}_p(X), W_p)$ is a metric space. It is easy to check that a set $\mathcal{K} \subset \mathcal{P}_p(X)$ is bounded (i.e. there exists a measure $\nu \in \mathcal{P}_p(X)$ such that $\{W_p(\mu,\nu)\}_{\mu\in\mathcal{K}}$ is a bounded subset of $\mathbb{R}$) if and only if

$$\sup_{\mu\in\mathcal{K}} \int_X \mathsf{d}^p(x,x_o)\,\mathrm{d}\mu < +\infty \quad \text{for some (and thus any) point } x_o \in X. \tag{2.10}$$

The following result shows the relationships the narrow topology and the topology induced by the Wasserstein distance W_p.

Proposition 2.8. *If (X,d) is separable (resp. complete) then $(\mathcal{P}_p(X), W_p)$ is a separable (resp. complete) metric space. A set $\mathcal{K} \subset \mathcal{P}_p(X)$ is relatively compact iff it has uniformly integrable p-moments and is tight. In particular, for a given sequence $(\mu_n) \subset \mathcal{P}_p(X)$ we have*

$$\lim_{n\to\infty} W_p(\mu_n,\mu) = 0 \iff \begin{cases} \mu_n \text{ narrowly converges to } \mu \text{ as } n\to\infty, \\ (\mu_n) \text{ has uniformly integrable } p\text{-moments.} \end{cases} \tag{2.11}$$

Proposition 2.9 (Stability of optimality and narrow lower semicontinuity)**.** *Let (μ_n^1), $(\mu_n^2) \subset \mathcal{P}_p(X)$ be two bounded sequences narrowly converging to μ^1, μ^2 respectively, and let $\boldsymbol{\mu}_n \in \Gamma_o(\mu_n^1,\mu_n^2)$ be a sequence of optimal plans. Then $(\boldsymbol{\mu}_n)$ is narrowly relatively compact in $\mathcal{P}(X\times X)$ and any narrow limit point $\boldsymbol{\mu}$ belongs to $\Gamma_o(\mu^1,\mu^2)$, with*

$$\begin{aligned} W_p^p(\mu^1,\mu^2) &= \int_{X^2} \mathsf{d}^p(x_1,x_2)\,\mathrm{d}\boldsymbol{\mu}(x_1,x_2) \\ &\le \liminf_{n\to\infty} \int_{X^2} \mathsf{d}^p(x_1,x_2)\,\mathrm{d}\boldsymbol{\mu}_n(x_1,x_2) = \liminf_{n\to\infty} W_p^p(\mu_n^1,\mu_n^2). \end{aligned} \tag{2.12}$$

3 A strong-weak topology on measures in product spaces

Let us consider a separable Banach space X endowed with the strong topology induced by its norm $\|\cdot\|_{\mathsf{X}}$ (we will occasionally use the notation X_s when we want to emphasize the choice of the strong topology) and a reflexive and separable Banach space $(\mathsf{Y}, \|\cdot\|_{\mathsf{Y}})$. We will denote by Y_w the space Y endowed with the weak topology $\sigma(\mathsf{Y},\mathsf{Y}')$.

We are interested in Radon probability measures in the topological space $\mathsf{X}_s \times \mathsf{Y}_w$. Since $\mathsf{X}_s \times \mathsf{Y}_w$ is endowed with the product topology of two Lusin and completely regular spaces, it is a completely regular Lusin space as well; in particular Borel measures are Radon, the set $\mathcal{P}(\mathsf{X}_s\times\mathsf{Y}_w)$ coincides with $\mathcal{P}(\mathsf{X}_s\times\mathsf{Y}_s)$, and narrow convergence in $\mathcal{P}(\mathsf{X}_s\times\mathsf{Y}_w)$ is well defined.

Let us set $\mathsf{Z} := \mathsf{X}\times\mathsf{Y}$ and let us fix $p\in[1,+\infty),\ q\in(1,+\infty)$; we want now to introduce a natural topology on the subset

$$\mathcal{P}_{pq}(\mathsf{Z}) := \Big\{\boldsymbol{\mu}\in\mathcal{P}(\mathsf{Z}) : \int \Big(\|x\|_{\mathsf{X}}^p + \|y\|_{\mathsf{Y}}^q\Big)\,\mathrm{d}\boldsymbol{\mu}(x,y) < +\infty\Big\}. \tag{3.1}$$

In order to define such a topology, we consider the space $\mathrm{C}^{sw}_{pq}(\mathsf{Z})$ of test functions $\zeta : \mathsf{Z} \to \mathbb{R}$ such that

$$\zeta \text{ is sequentially continuous in } \mathsf{X}_s \times \mathsf{Y}_w, \tag{3.2}$$

$$\forall\, \varepsilon > 0\ \exists\, A_\varepsilon \geq 0 : |\zeta(x,y)| \leq A_\varepsilon(1 + \|x\|^p_{\mathsf{X}}) + \varepsilon \|y\|^q_{\mathsf{Y}} \quad \text{for all } (x,y) \in \mathsf{X} \times \mathsf{Y}. \tag{3.3}$$

We endow $\mathrm{C}^{sw}_{pq}(\mathsf{Z})$ with the norm

$$\|\zeta\|_{\mathrm{C}^{sw}_{pq}(\mathsf{Z})} := \sup_{(x,y)\in \mathsf{Z}} \frac{|\zeta(x,y)|}{1 + \|x\|^p_{\mathsf{X}} + \|y\|^q_{\mathsf{Y}}}. \tag{3.4}$$

Remark 3.1. When Y is finite dimensional, (3.2) is equivalent to the continuity of ζ. It is worth noticing that if p, q are conjugate exponents, any continuous and bilinear map $\beta : \mathsf{X} \times \mathsf{Y} \to \mathbb{R}$ belongs to $\mathrm{C}^{sw}_{pq}(\mathsf{X} \times \mathsf{Y})$. In fact, it is easy to check that β is sequentially continuous in $\mathsf{X}_s \times \mathsf{Y}_w$ and its continuity yields the existence of a constant $L \geq 0$ such that

$$|\beta(x,y)| \leq L\|x\|_{\mathsf{X}}\, \|y\|_{\mathsf{Y}} \quad \text{for every } x \in \mathsf{X},\ y \in \mathsf{Y},$$

so that

$$|\beta(x,y)| \leq \frac{L^p}{p\varepsilon^{p/q}} \|x\|^p_{\mathsf{X}} + \frac{\varepsilon}{q}\|y\|^q_{\mathsf{Y}} \quad \text{for every } x \in \mathsf{X},\ y \in \mathsf{Y},\ \varepsilon > 0.$$

This covers in particular the case when $\mathsf{Y} = \mathsf{X}$ is an Hilbert space and β coincides with the scalar product $\langle \cdot, \cdot \rangle$ in X.

Lemma 3.2. $(\mathrm{C}^{sw}_{pq}(\mathsf{Z}), \|\cdot\|_{\mathrm{C}^{sw}_{pq}(\mathsf{Z})})$ *is a Banach space.*

Proof. It is obvious that $\|\cdot\|_{\mathrm{C}^{sw}_{pq}(\mathsf{Z})}$ is a norm, we can thus check the completeness. Let $\{\zeta_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathrm{C}^{sw}_{pq}(\mathsf{Z})$ and let ζ its pointwise limit, so that for every $\eta > 0$ there exists $N_\eta \in \mathbb{N}$ such that

$$|\zeta(x,y) - \zeta_n(x,y)| \leq \eta(1 + \|x\|^p_{\mathsf{X}} + \|y\|^q_{\mathsf{Y}}) \quad \text{for every } (x,y) \in \mathsf{X} \times \mathsf{Y},\ n \geq N_\eta. \tag{3.5}$$

If (x_k, y_k) is a sequence converging to $(\bar{x}, \bar{y})$ in $\mathsf{X}_s \times \mathsf{Y}_w$ as $k \to \infty$ we know that

$$\|\bar{x}\|^p_{\mathsf{X}} + \|\bar{y}\|^q_{\mathsf{Y}} \leq R := \sup_k \|x_k\|^p_{\mathsf{X}} + \|y_k\|^q_{\mathsf{Y}} < \infty$$

so that for every $n \geq N_\eta$

$$\limsup_{k\to\infty} |\zeta(\bar{x},\bar{y}) - \zeta(x_k,y_k)| \leq 2(1+R)\eta + \limsup_{k\to\infty} |\zeta_n(\bar{x},\bar{y}) - \zeta_n(x_k,y_k)| = 2(1+R)\eta.$$

Since η is arbitrary, we conclude that ζ is sequentially continuous in Z. Let us eventually check that ζ satisfies (3.3). For a given $\varepsilon > 0$ we apply (3.5) with $\eta := \varepsilon/2$ and we pick up $n \geq N_\eta$. Since ζ_n belongs to $\mathrm{C}^{sw}_{pq}(\mathsf{Z})$, we find $B_\varepsilon \geq 0$ such that

$$|\zeta_n(x,y)| \leq B_\varepsilon(1 + \|x\|^p_{\mathsf{X}}) + \varepsilon/2\|y\|^q_{\mathsf{Y}} \quad \text{for every } (x,y) \in \mathsf{X} \times \mathsf{Y}.$$

Combining such inequality with (3.5) we conclude that

$$|\zeta(x,y)| \leq |\zeta(x,y) - \zeta_n(x,y)| + |\zeta_n(x,y)| \leq (\varepsilon + B_\varepsilon)(1 + \|x\|^p_{\mathsf{X}}) + \varepsilon\|y\|^q_{\mathsf{Y}}. \qquad \square$$

Definition 3.3 (Topology of $\mathcal{P}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$)**.** We endow $\mathcal{P}_{pq}(\mathsf{X}\times\mathsf{Y})$ with the initial topology induced by the functions

$$\boldsymbol{\mu} \mapsto \int \zeta(x,y)\,\mathrm{d}\boldsymbol{\mu}(x,y), \quad \zeta \in \mathrm{C}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y}); \tag{3.6}$$

we call $\mathcal{P}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$ the topological space $\big(\mathcal{P}_{pq}(\mathsf{X}\times\mathsf{Y}), \sigma\big(\mathcal{P}_{pq}(\mathsf{X}\times\mathsf{Y}), \mathrm{C}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})\big)\big)$.

It is obvious that whenever $r \geq p\vee q$ the Wasserstein topology of $\mathcal{P}_r(\mathsf{X}\times\mathsf{Y})$ (induced by the Wasserstein distance W_r generated by any product norm in the Banach space $\mathsf{X}\times\mathsf{Y}$) is finer than the topology of $\mathcal{P}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$ and the latter is finer than the narrow topology of $\mathcal{P}(\mathsf{X}_s\times\mathsf{Y}_w)$. The next proposition collects other important properties and justifies the interest of the $\mathcal{P}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$-topology.

Proposition 3.4. (a) *If $(\boldsymbol{\mu}_\alpha)_{\alpha\in\mathbb{A}} \subset \mathcal{P}^{sw}_{pq}(\mathsf{Z})$ is a net indexed by the directed set $\mathbb{A}$ and $\mu \in \mathcal{P}^{sw}_{pq}(\mathsf{Z})$ satisfy*

(i) $\boldsymbol{\mu}_\alpha \to \boldsymbol{\mu}$ *narrowly in* $\mathcal{P}(\mathsf{X}_s\times\mathsf{Y}_w)$*;*

(ii) $\displaystyle\lim_{\alpha\in\mathbb{A}} \int \|x\|^p_{\mathsf{X}}\,\mathrm{d}\boldsymbol{\mu}_\alpha = \int \|x\|^p_{\mathsf{X}}\,\mathrm{d}\boldsymbol{\mu}$*;*

(iii) $\displaystyle\sup_{\alpha\in\mathbb{A}} \int \|y\|^q_{\mathsf{Y}}\,\mathrm{d}\boldsymbol{\mu}_\alpha < \infty$*,*

then $\boldsymbol{\mu}_\alpha \to \boldsymbol{\mu}$ in $\mathcal{P}^{sw}_{pq}(\mathsf{Z})$. The converse property holds for sequences: i.e. if $\mathbb{A} = \mathbb{N}$ and $\boldsymbol{\mu}_n \to \boldsymbol{\mu}$ in $\mathcal{P}^{sw}_{pq}(\mathsf{Z})$ as $n\to\infty$, then properties (a), (b), (c) hold.

(b) *For every compact set $\mathcal{K} \subset \mathcal{P}_p(\mathsf{X})$ and every constant $c < \infty$ the sets*

$$\mathcal{K}_c := \Big\{\boldsymbol{\mu} \in \mathcal{P}^{sw}_{pq}(\mathsf{Z}) : \pi^1_\sharp\boldsymbol{\mu} \in \mathcal{K}, \quad \int \|y\|^q_{\mathsf{Y}}\,\mathrm{d}\boldsymbol{\mu} \leq c\Big\} \tag{3.7}$$

are compact and metrizable in $\mathcal{P}^{sw}_{pq}(\mathsf{Z})$ (in particular they are sequentially compact).

Proof. Let us consider the first claim and let $(\boldsymbol{\mu}_\alpha)_{\alpha\in\mathbb{A}}$ in $\mathcal{P}^{sw}_{pq}(\mathsf{Z})$ satisfy properties (i), (ii), (iii) with $S := \sup_\alpha \int \|y\|^q_{\mathsf{Y}}\,\mathrm{d}\boldsymbol{\mu}_\alpha < \infty$.

We first observe that $\pi^1_\sharp\boldsymbol{\mu}_\alpha \to \pi^1_\sharp\boldsymbol{\mu}$ in $\mathcal{P}_p(\mathsf{X})$. Let us now fix $\zeta \in \mathrm{C}^{sw}_{pq}(\mathsf{Z})$ and for every $\varepsilon > 0$ let A_ε as in (3.3). The function $\zeta_\varepsilon(x,y) := \zeta(x,y) + A_\varepsilon(1+\|x\|^p_{\mathsf{X}}) + 2\varepsilon\|y\|^q_{\mathsf{Y}}$ is nonnegative and it is also lower semicontinuous w.r.t. the $\mathsf{X}_s\times\mathsf{Y}_w$-topology: in fact, the sublevels $X_{\varepsilon,c} := \{(x,y) \in \mathsf{X}\times\mathsf{Y} : \zeta_\varepsilon(x,y) \leq c\}$ of ζ_ε are sequentially closed and contained in $\mathsf{X}\times\{y \in \mathsf{X} : \|y\|^q_{\mathsf{Y}} \leq c/\varepsilon\}$ which is a metrizable space, so that $X_{\varepsilon,c}$ is closed in $\mathsf{X}_s\times\mathsf{Y}_w$. It follows that

$$\begin{aligned}\liminf_{\alpha\in\mathbb{A}} \int \zeta\,\mathrm{d}\boldsymbol{\mu}_\alpha &= \liminf_{\alpha\in\mathbb{A}} \int \zeta_\varepsilon\,\mathrm{d}\boldsymbol{\mu}_\alpha - A_\varepsilon \int \big(1+\|x\|^p_{\mathsf{X}}\big)\,\mathrm{d}\boldsymbol{\mu}_\alpha - 2\varepsilon \int \|y\|^q_{\mathsf{Y}}\Big)\,\mathrm{d}\boldsymbol{\mu}_\alpha \\ &\geq \int \zeta_\varepsilon\,\mathrm{d}\boldsymbol{\mu} - A_\varepsilon \int \big(1+\|x\|^p_{\mathsf{X}}\big)\,\mathrm{d}\boldsymbol{\mu} - 2\varepsilon S \geq \int \zeta\,\mathrm{d}\boldsymbol{\mu} - 2\varepsilon S\end{aligned}$$

and, since $\varepsilon > 0$ is arbitrary, $\liminf_{\alpha\in\mathbb{A}} \int \zeta \,\mathrm{d}\boldsymbol{\mu}_\alpha \geq \int \zeta \,\mathrm{d}\boldsymbol{\mu}$. Applying the same argument to $-\zeta$ we conclude that $\boldsymbol{\mu}_\alpha$ converges to $\boldsymbol{\mu}$ in $\mathcal{P}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$.

In order to prove the converse implication in the case of sequences, let us observe that if $\boldsymbol{\mu}_n \to \boldsymbol{\mu}$ in $\mathcal{P}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$ then properties (i) and (ii) are obvious. Since $\mathrm{C}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$ is a Banach space and each measure $\boldsymbol{\mu}_n$ induces a bounded linear functional L_n on $\mathrm{C}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$, the principle of uniform boundedness implies that $S := \sup_n \|L_n\|_{(\mathrm{C}^{sw}_{pq}(\mathsf{Z}))'} < \infty$, i.e.

$$\int \zeta \,\mathrm{d}\boldsymbol{\mu}_n \leq S \quad \text{for every } \zeta \in \mathrm{C}^{sw}_{pq}(\mathsf{Z}),\ |\zeta(x,y)| \leq 1 + \|x\|^p_{\mathsf{X}} + \|y\|^q_{\mathsf{Y}}. \tag{3.8}$$

Let now $(\mathsf{e}_h)_{h\in\mathbb{N}}$ be a strongly dense subset of the unit ball of Y' (the dual of Y, which is separable as well) and let

$$\zeta_k(x,y) := \Big(\sup_{1\leq h\leq k} |\langle y, \mathsf{e}_h\rangle|\Big)^q \wedge k \tag{3.9}$$

Clearly each ζ_k belongs to the unit ball of $\mathrm{C}^{sw}_{pq}(\mathsf{Z})$ so that

$$\int \zeta_k(x,y) \,\mathrm{d}\boldsymbol{\mu}_n(x,y) \leq S \quad \text{for every } k,n\in\mathbb{N}. \tag{3.10}$$

Since $\zeta_k(x,y) \uparrow \|y\|^q_{\mathsf{Y}}$ as $k\to\infty$, Lebesgue Dominated Convergence Theorem yields $\int \|y\|^q_{\mathsf{Y}} \,\mathrm{d}\boldsymbol{\mu}_n \leq S$ for every $n\in\mathbb{N}$.

(b) Since $\mathcal{K}$ is tight and $y \mapsto \|y\|^q_{\mathsf{Y}}$ has compact sublevel in Y_w, the set $\mathcal{K}_c$ is tight in $\mathcal{P}(\mathsf{Z})$ and it is also closed, so that it is compact in $\mathcal{P}(\mathsf{Z})$. Every net $(\boldsymbol{\mu}_\alpha)_{\alpha\in\mathbb{A}}$ in $\mathcal{K}_c$ has a subnet $(\boldsymbol{\mu}_{\alpha(\beta)})_{\beta\in\mathbb{B}}$ converging to $\boldsymbol{\mu}\in\mathcal{K}_c$ in $\mathcal{P}(\mathsf{Z})$. Since $\pi^1_\sharp\boldsymbol{\mu}_\alpha$ is uniformly p-integrable we deduce $\lim_{\beta\in\mathbb{B}} \int \|x\|^p_{\mathsf{X}} \,\mathrm{d}\boldsymbol{\mu}_{\alpha(\beta)} = \int \|x\|^p_{\mathsf{X}} \,\mathrm{d}\boldsymbol{\mu}$. Applying the previous claim, we deduce that $\boldsymbol{\mu}_{\alpha(\beta)} \to \boldsymbol{\mu}$ w.r.t. $\mathcal{P}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$. In order to prove the metrizability we observe that the bounded distance on Y

$$\mathsf{d}_\varpi(y_1,y_2) := \sum_{n=1}^\infty 2^{-n}(|\langle y_1 - y_2, \mathsf{e}_n\rangle| \wedge 1) \quad \text{where } (\mathsf{e}_n)_{n\in\mathbb{N}} \text{ is dense in the unit ball of } \mathsf{Y}', \tag{3.11}$$

induces a coarser topology than $\sigma(\mathsf{Y},\mathsf{Y}')$ in Y, so that the L_1-Wasserstein distance associated to

$$\mathsf{d}((x_1,y_1),(x_2,y_2)) := |x_1 - x_2|_{\mathsf{X}} + \mathsf{d}_\varpi(y_1,y_2)$$

induces a coarser topology than the topology of $\mathcal{P}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$, which on the other hand coincides with the $\mathcal{P}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$-topology on the compact set $\mathcal{K}_c$. $\mathcal{K}_c$ is therefore metrizable. □

As we already observed in the Introduction, it is worth noticing that the topology of $\mathcal{P}^{sw}_{pp}(\mathsf{X}\times\mathsf{Y})$ is strictly coarser than the Wasserstein topology of $\mathcal{P}_p(\mathsf{X}\times\mathsf{Y})$ even when Y

is finite dimensional. In fact, $\mathrm{C}^{sw}_{pp}(\mathsf{X}\times\mathsf{Y})$ does not contain the function $(x,y)\mapsto\|y\|^p_{\mathsf{Y}}$, so that convergence of the p-moment w.r.t. y is not guaranteed.

The previous construction is useful also in the case of a single space Y (we may think that X reduces to $\{0\}$).

Definition 3.5 (The topology of $\mathcal{P}^w_q(\mathsf{Y})$)**.** Let Y be a reflexive and separable Banach space and $q\in(1,+\infty)$.

(a) $\mathrm{C}^w_q(\mathsf{Y})$ is the Banach space of sequentially weakly continuous (continuous, if Y is finite dimensional) functions $\zeta:\mathsf{Y}\to\mathbb{R}$ satisfying

$$\forall\,\varepsilon>0\;\exists\,A_\varepsilon\ge 0:\quad |\zeta(y)|\le A_\varepsilon+\varepsilon\|y\|^q_{\mathsf{Y}}\quad\text{for every } y\in\mathsf{Y}, \tag{3.12}$$

or, equivalently, $\displaystyle\lim_{\|y\|_{\mathsf{Y}}\to\infty}\frac{\zeta(y)}{1+\|y\|^q_{\mathsf{Y}}}=0$, endowed with the norm

$$\|\zeta\|_{\mathrm{C}^w_q(\mathsf{Y})}:=\sup_{y\in\mathsf{Y}}\frac{|\zeta(y)|}{1+\|y\|^q_{\mathsf{Y}}}. \tag{3.13}$$

(b) $\mathcal{P}^w_q(\mathsf{Y})$ is the topological space of measures in $\mathcal{P}_q(\mathsf{Y})$ endowed with the initial topology $\sigma\big(\mathcal{P}_q(\mathsf{Y}),\mathrm{C}^w_q(\mathsf{Y})\big)$ (or, equivalently, the weak* topology of $\big(\mathrm{C}^w_q(\mathsf{Y})\big)'$).

The following result is an immediate consequence of Proposition 3.4.

Corollary 3.6. *Let Y be a reflexive and separable Banach space and $q\in(1,+\infty)$.*

(a) *The topology of $\mathcal{P}^w_q(\mathsf{Y})$ is finer than the narrow topology of $\mathcal{P}(\mathsf{Y}_w)$; they coincide on bounded subsets $\mathcal{K}$ of $\mathcal{P}_q(\mathsf{Y})$, i.e. satisfying*

$$\sup_{\mu\in\mathcal{K}}\int\|y\|^q_{\mathsf{Y}}\,\mathrm{d}\mu<\infty. \tag{3.14}$$

(b) *a sequence $(\mu_n)_{n\in\mathbb{N}}$ converges to μ in $\mathcal{P}^w_q(\mathsf{Y})$ if and only if*

$$(\mu_n)_{n\in\mathbb{N}}\ \textit{converges narrowly in }\mathcal{P}(\mathsf{Y}_w)\quad\textit{and}\quad\sup_{n\in\mathbb{N}}\int\|y\|^q_{\mathsf{Y}}\,\mathrm{d}\mu_n(y)<\infty.$$

(c) *a set $\mathcal{K}\subset\mathcal{P}_q(\mathsf{Y})$ is relatively sequentially compact in $\mathcal{P}^w_q(\mathsf{Y})$ if and only if it satisfies (3.14).*

(d) *If a sequence $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$ in $\mathcal{P}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$ converges to $\boldsymbol{\mu}$ in $\mathcal{P}^{sw}_{pq}(\mathsf{X}\times\mathsf{Y})$ then $\pi^2_\sharp\boldsymbol{\mu}_n\to\pi^2_\sharp\boldsymbol{\mu}$ in $\mathcal{P}^w_q(\mathsf{Y})$ (and $\pi^1_\sharp\boldsymbol{\mu}_n\to\pi^1_\sharp\boldsymbol{\mu}$ in $\mathcal{P}_p(\mathsf{X})$).*

Remark 3.7. All the definitions and results of this Section can be easily extended to the case when Y_w is replaced by the dual Y'_{w*} of a separable Banach space endowed with its weak*-topology and we deal with the product $\mathsf{X}_s\times\mathsf{Y}'_{w*}$. We could therefore consider the spaces $\mathcal{P}^{sw*}_{pq}(\mathsf{X}\times\mathsf{Y}')$ and $\mathcal{P}^{w*}_q(\mathsf{Y}')$.

Let us conclude this section with a useful application of the weak topology of Definition 3.3 to the stability of optimal plans in Hilbert spaces.

Theorem 3.8. *Let* H *be a separable Hilbert space, let* $(\mu_n^i)_{n\in\mathbb{N}}$ *be two sequences in* $\mathcal{P}_2(\mathsf{H})$*,* $i = 1, 2$ *with* $\boldsymbol{\gamma}_n \in \Gamma_o(\mu_n^1, \mu_n^2)$*, and let* $\mu^1, \mu^2 \in \mathcal{P}_2(\mathsf{H})$ *with* $\boldsymbol{\gamma} \in \Gamma(\mu^1, \mu^2)$*. If*

$$(\mu_n^2) \text{ is tight in } \mathcal{P}(\mathsf{H}_w) \quad \text{and} \quad \boldsymbol{\gamma}_n \to \boldsymbol{\gamma} \text{ narrowly in } \mathcal{P}(\mathsf{H}_s \times \mathsf{H}_w) \text{ as } n \to \infty, \tag{3.15}$$

then $\mu_n^1 \to \mu^1$ *narrowly in* $\mathcal{P}(\mathsf{H})$*,* $\mu_n^2 \to \mu^2$ *narrowly in* $\mathcal{P}(\mathsf{H}_w)$ *and* $\boldsymbol{\gamma} \in \Gamma_o(\mu^1, \mu^2)$*. In particular, any limit point* $\boldsymbol{\gamma}$ *of optimal plans* $\boldsymbol{\gamma}_n$ *in* $\mathcal{P}_{22}^{sw}(\mathsf{H} \times \mathsf{H})$ *is optimal as well.*

Proof. The statement concerning the convergence of μ_n^1 and μ_n^2 is obvious. Since $\boldsymbol{\gamma}$ has finite quadratic moment, in order to check its optimality it is sufficient to prove that $\boldsymbol{\gamma}$ is concentrated on a cyclically monotone set, i.e. there exists a Borel set $M \subset \mathsf{H} \times \mathsf{H}$ such that $\boldsymbol{\gamma}(\mathsf{H}^2 \setminus M) = 0$ and for every $N \in \mathbb{N}$

$$(x_1^k, x_2^k) \in M,\ k = 0, \cdots, N, \text{ with } (x_1^0, x_2^0) = (x_1^N, x_2^N) \ \Rightarrow\ \sum_{k=1}^N \langle x_1^k - x_1^{k-1}, x_2^k \rangle \geq 0. \tag{3.16}$$

The standard idea, i.e. using the convergence of the supports of $\boldsymbol{\gamma}_n$, should be adapted to the case of the (non-metrizable) weak topology of H. We thus consider also the metric space $(\mathsf{H}_\varpi, \mathsf{d}_\varpi)$, whose metric has been defined by (3.11) (here $\mathsf{Y} = \mathsf{H}$); we recall that the topology induced by d_ϖ coincides with the weak topology on every bounded subset of H.

Since $(\mu_n^1)_n$ is narrowly convergent in $\mathcal{P}(\mathsf{H})$ it is tight, so that we can find a function $\psi_1 : \mathsf{H} \to [0, +\infty]$ with strongly compact sublevels such that $\int_{\mathsf{H}} \psi_1(x)\,\mathrm{d}\mu_n^1(x) \leq S_1 < \infty$ for every $n \in \mathbb{N}$. Since (μ_n^2) is tight in $\mathcal{P}(\mathsf{H}_w)$ we can find a function $\psi_2 : \mathsf{H} \to [0, +\infty]$ with weakly compact sublevels such that $\int_{\mathsf{H}} \psi_2(x)\,\mathrm{d}\mu_n^2(x) \leq S_2 < \infty$ for every $n \in \mathbb{N}$. Let us set $\boldsymbol{\sigma}_n := (\mathrm{Id}_{\mathsf{H}\times\mathsf{H}} \times \psi)_\sharp \boldsymbol{\gamma}_n \in \mathcal{P}(\mathsf{H}^2 \times [0, +\infty))$. We have that

$$\int \Big(\psi_1(x_1) + \psi_2(x_2) + |r|\Big)\,\mathrm{d}\boldsymbol{\sigma}_n(x_1, x_2, r) \leq S_1 + 2S_2 \tag{3.17}$$

so that the sequence $(\boldsymbol{\sigma}_n)_{n\in\mathbb{N}}$ is tight in $\mathcal{P}(\mathsf{H} \times \mathsf{H}_\varpi \times \mathbb{R})$ (recall (3.11)). Since $\mathsf{H}_s \times \mathsf{H}_\varpi \times \mathbb{R}$ is metrizable, we can thus extract a subsequence (still denoted by $\boldsymbol{\sigma}_n$) converging to a limit plan $\boldsymbol{\sigma} \in \mathcal{P}(\mathsf{H} \times \mathsf{H} \times \mathbb{R})$ such that $\pi^{12}_\sharp \boldsymbol{\sigma} = \boldsymbol{\gamma}$.

Since $\boldsymbol{\sigma}$ is a Radon probability measure in $\mathsf{H}^2 \times \mathbb{R}$, we can find an increasing sequence of compact sets $K_j \subset \mathrm{supp}(\boldsymbol{\sigma}) \subset \mathsf{H}^2 \times \mathbb{R}$ such that $\boldsymbol{\sigma}(\mathsf{H}^2 \times \mathbb{R} \setminus \cup_j K_j) = 0$. It follows that $\boldsymbol{\gamma}$ is concentrated on $M := \cup_j M_j$ where $M_j := \pi^{12}\big(K_j\big)$ are compact sets.

Let now (x_1^k, x_2^k), $k = 0, \cdots, N$, be points in M as in (3.16). There exists $j \in \mathbb{N}$ and points $r^k \geq 0$ such that $(x_1^k, x_2^k, r^k) \in K_j$. Since $\boldsymbol{\sigma}_n$ is concentrated on the set $(\mathrm{Id}_{\mathsf{H}^2} \times \psi)(\mathrm{supp}(\boldsymbol{\gamma}_n))$, we can thus find a sequence $(x_{1,n}^k, x_{2,n}^k) \in \mathrm{supp}(\boldsymbol{\gamma}_n)$ such that

$$x_{1,n}^k \to x_1^k \text{ strongly in } \mathsf{H},\ \mathsf{d}_\varpi(x_{2,n}^k, x_2^k) \to 0,\ \psi_2(x_{2,n}^k) \to r^k \text{ in } \mathbb{R} \text{ as } n \to \infty. \tag{3.18}$$

Since ψ_2 has weakly compact sublevels, we deduce that $x^k_{2,n} \rightharpoonup x^k_2$ as $n \to \infty$. Since $\boldsymbol{\gamma}_n$ is cyclically monotone, we know that

$$\sum_{k=1}^{N} \langle x^k_{1,n} - x^{k-1}_{1,n}, x^k_{2,n} \rangle \geq 0 \quad \text{for every } n \in \mathbb{N}. \tag{3.19}$$

We can then pass to the limit as $n \to \infty$ in (3.19) and using the sequential continuity of the scalar product in $\mathsf{H} \times \mathsf{H}_w$ we obtain (3.16). □

4 Weak lower semicontinuity of geodesically convex functions in $\mathcal{P}_2(\mathsf{H})$

Let $(\mathsf{H}, |\cdot|)$ be a separable Hilbert space and let $\varphi : \mathsf{H} \to \mathbb{R} \cup \{+\infty\}$ be a convex function. It is well known that φ is lower semicontinuous w.r.t. the strong topology of H if and only if it is lower semicontinuous w.r.t. the weak topology. We want to extend this property to geodesically convex functions $\phi : \mathcal{P}_2(\mathsf{H}) \to \mathbb{R} \cup \{+\infty\}$, an important class of functions introduced by McCann [8].

Let us first recall that a (minimal, constant speed) geodesic $(\mu_s)_{s\in[0,1]}$ in $\mathcal{P}_2(\mathsf{H})$ connecting two given measures $\mu_0, \mu_1 \in \mathcal{P}_2(\mathsf{H})$ is a Lipschitz curve satisfying

$$W_2(\mu_s, \mu_t) = |t - s| W_2(\mu_0, \mu_1) \quad \text{for every } s, t \in [0, 1]. \tag{4.1}$$

Equivalently, it is possible to prove (see e.g. [1]) that a curve $(\mu_s)_{s\in[0,1]}$ is a geodesic if and only if there exists an optimal plan $\boldsymbol{\mu} \in \Gamma_o(\mu_0, \mu_1)$ such that

$$\mu_s := (\pi_s^{1\to2})_\sharp \boldsymbol{\mu}, \quad \pi_s^{1\to2}(x_1, x_2) := (1 - s)x_1 + s x_2 \quad x_1, x_2 \in \mathsf{H},\ s \in [0, 1]. \tag{4.2}$$

Definition 4.1. Let $\phi : \mathcal{P}_2(\mathsf{H}) \to \mathbb{R} \cup \{+\infty\}$ be a function with proper domain $D(\phi) := \{\mu \in \mathcal{P}_2(\mathsf{H}) : \phi(\mu) < \infty\} \neq \emptyset$. ϕ is geodesically convex if every $\mu_0, \mu_1 \in D(\phi)$ can be connected by a geodesic $(\mu_s)_{s\in[0,1]}$ in $\mathcal{P}_2(\mathsf{H})$ along which

$$\phi(\mu_s) \leq (1 - s)\phi(\mu_0) + s\phi(\mu_1) \quad \text{for every } s \in [0, 1]. \tag{4.3}$$

Equivalently, there exists $\boldsymbol{\mu} \in \Gamma_o(\mu_1, \mu_2)$ such that

$$\phi\big((\pi_s^{1\to2})_\sharp \boldsymbol{\mu}\big) \leq (1 - s)\phi(\mu_0) + s\phi(\mu_1) \quad \text{for every } s \in [0, 1]. \tag{4.4}$$

Theorem 4.2. *Every lower semicontinuous and geodesically convex function $\phi : \mathcal{P}_2(\mathsf{H}) \to \mathbb{R} \cup \{+\infty\}$ is sequentially lower semicontinuous w.r.t. the (weak) topology of $\mathcal{P}_2^w(\mathsf{H})$: for every sequence $(\mu_n)_{n\in\mathbb{N}}$ and μ in $\mathcal{P}_2(\mathsf{H})$ we have*

$$\mu_n \to \mu \text{ narrowly in } \mathcal{P}(\mathsf{H}_w), \quad \sup_n \int |x|^2 \, \mathrm{d}\mu_n < \infty \quad \Rightarrow \quad \liminf_{n\to\infty} \phi(\mu_n) \geq \phi(\mu). \tag{4.5}$$

The proof of Theorem 4.2 (at the end of the present section) is based on two preliminary results; the first one is an application to the Wasserstein space of [9, Theorems 2.10, 2.17].

Theorem 4.3. *Let $\phi : \mathcal{P}_2(\mathsf{H}) \to \mathbb{R} \cup \{+\infty\}$ be a proper, lower semicontinuous, and geodesically convex function.*

(a) *ϕ is linearly bounded from below: there exists $\mu_o \in \mathcal{P}_2(\mathsf{H})$ and $\ell_o, \phi_o \in \mathbb{R}$ such that*

$$\phi(\mu) \geq \phi_o - \ell_o W_2(\mu, \mu_o) \quad \text{for every } \mu \in \mathcal{P}_2(\mathsf{H}). \tag{4.6}$$

(b) *For every $\mu \in D(\phi)$ and $\tau > 0$ there exists $\mu_\tau \in D(\phi)$ such that*

$$\frac{1}{2\tau} W_2^2(\mu_\tau, \mu) + \phi(\mu_\tau) \leq \frac{1}{2\tau} W_2^2(\mu, \nu) + \phi(\nu) + W_2(\mu_\tau, \mu) W_2(\mu_\tau, \nu) \tag{4.7}$$

for every $\nu \in D(\phi)$,

$$\frac{1}{2\tau} W_2^2(\mu_\tau, \mu) + \phi(\mu_\tau) \leq \phi(\mu) \tag{4.8}$$

$$\lim_{\tau \downarrow 0} W_2(\mu_\tau, \mu) = 0, \quad \lim_{\tau \downarrow 0} \phi(\mu_\tau) = \phi(\mu). \tag{4.9}$$

We refer to [9] for the *proof*, where both properties (a) (Theorem 2.17) and (b) (Theorem 2.10) are stated in an arbitrary complete metric space. (b) is an application of Ekeland's Variational Principle and holds for arbitrary proper and lower semicontinuous functionals satisfying the lower bound (4.6).

Lemma 4.4. *Let $\tau > 0$, μ, μ_τ as in (4.7) and (4.8) of Theorem 4.3 and let $\boldsymbol{\mu}_\tau \in \Gamma_o(\mu_\tau, \mu)$. For every $\nu \in D(\phi)$ and $\boldsymbol{\gamma}_\tau \in \Gamma(\boldsymbol{\mu}_\tau, \nu)$ such that $\pi^{13}_\sharp \boldsymbol{\gamma}_\tau \in \Gamma_o(\mu_\tau, \mu)$ and ϕ satisfies the convexity inequality (4.3) along $(\pi_s^{1\to 3})_\sharp \boldsymbol{\gamma}$, we have*

$$\phi(\nu) - \phi(\mu_\tau) \geq \frac{1}{\tau} \int \langle x_1 - x_2, x_1 - x_3 \rangle \, \mathrm{d}\boldsymbol{\gamma}_\tau - W_2(\mu_\tau, \mu) W_2(\mu_\tau, \nu). \tag{4.10}$$

Proof. Let $\boldsymbol{\gamma}_\tau \in \Gamma(\boldsymbol{\mu}_\tau, \nu)$ as in the statement of the Lemma and let $\nu_s := (\pi_s^{1\to 3})_\sharp \boldsymbol{\gamma}_\tau$. Since ϕ satisfies the convexity inequality (4.3) along $(\nu_s)_{s \in [0,1]}$ we have

$$\phi(\nu) - \phi(\mu_\tau) \geq \frac{1}{s} \Big(\phi(\nu_s) - \phi(\mu_\tau) \Big). \tag{4.11}$$

On the other hand, (4.7) and the fact that $s^{-1} W_2(\nu_s, \mu_\tau) = W_2(\nu, \mu_\tau)$ yield

$$\frac{1}{s} \Big(\phi(\nu_s) - \phi(\mu_\tau) \Big) \geq \frac{1}{2\tau s} \Big(W_2^2(\mu_\tau, \mu) - \frac{1}{2\tau} W_2^2(\mu, \nu_s) \Big) - W_2(\mu_\tau, \mu) W_2(\mu_\tau, \nu) \tag{4.12}$$

Since $\pi^{12}_\sharp \boldsymbol{\gamma}_\tau$ is an optimal coupling between μ_τ and μ and $(\pi_s^{1\to 3})_\sharp \boldsymbol{\gamma}_\tau = \nu_s$ we have

$$W_2^2(\mu_\tau, \mu) = \int |x_1 - x_2|^2 \, \mathrm{d}\boldsymbol{\gamma}_\tau, \quad W_2^2(\mu, \nu_s) \leq \int |x_2 - (1-s)x_1 - s x_3|^2 \, \mathrm{d}\boldsymbol{\gamma}_\tau$$

so that (4.12) yields

$$\frac{1}{s}\Big(\phi(\nu_s)-\phi(\mu_\tau)\Big) \geq \frac{1}{2\tau s}\int \Big(|x_1-x_2|^2-|x_2-(1-s)x_1-sx_3|^2\Big)\,\mathrm{d}\boldsymbol{\gamma}_\tau \\ -W_2(\mu_\tau,\mu)W_2(\mu_\tau,\nu). \tag{4.13}$$

Passing to the limit as $s \downarrow 0$ in (4.13) and recalling (3.15) we eventually get (4.10). □

Proof of Theorem 4.2. It is not restrictive to assume that ϕ is proper and, possibly extracting a subsequence, that the limit $L := \lim_{n\to\infty}\phi(\mu_n)$ exists and it is finite, where μ_n is a sequence as in (4.5). We set $S := \sup_n W_2(\mu_n,\mu)$, which is finite since (μ_n) is bounded in $\mathcal{P}_2(\mathsf{H})$.

For every $\tau > 0$ let μ_τ be as in (4.7) and (4.8) of Theorem 4.3 and let $\boldsymbol{\gamma}_{\tau,n} \in \Gamma(\boldsymbol{\mu}_\tau,\mu_n)$ as in the previous Lemma 4.4. (4.10) yields

$$\phi(\mu_n) \geq \phi(\mu_\tau)+\frac{1}{\tau}\int \big\langle x_1-x_2,x_1-x_3\big\rangle\,\mathrm{d}\boldsymbol{\gamma}_{\tau,n}-W_2(\mu_\tau,\mu)\Big(W_2(\mu_\tau,\mu)+S\Big). \tag{4.14}$$

Setting $\mathsf{Z} := (\mathsf{H}^2)\times\mathsf{H}$, we can apply Proposition 3.4(b) to the sequence $(\boldsymbol{\gamma}_{\tau,n})_n$ obtaining a subsequence (still denoted by $\boldsymbol{\gamma}_{\tau,n}$) converging to a limit $\boldsymbol{\gamma}_\tau \in \Gamma(\boldsymbol{\mu}_\tau,\mu)$ in $\mathcal{P}^{sw}_{22}(\mathsf{Z})$. Since

$$\text{the map } A:\mathsf{Z}\to\mathbb{R},\ A(x_1,x_2,x_3) := \big\langle x_1-x_2,x_1-x_3\big\rangle, \quad \text{belongs to } \mathrm{C}^{sw}_{22}(\mathsf{Z}), \tag{4.15}$$

by the very definition of the topology of $\mathcal{P}^{sw}_{22}(\mathsf{Z})$ we get

$$\lim_{n\to\infty}\int \big\langle x_1-x_2,x_1-x_3\big\rangle\,\mathrm{d}\boldsymbol{\gamma}_{\tau,n} = \int \big\langle x_1-x_2,x_1-x_3\big\rangle\,\mathrm{d}\boldsymbol{\gamma}_\tau. \tag{4.16}$$

On the other hand, by Theorem 3.8, $\pi^{2,3}_\sharp\boldsymbol{\gamma}_\tau$ is optimal, thus belongs to $\Gamma_o(\mu,\mu)$: it follows that it is concentrated on the subspace $\{(x_2,x_3)\in\mathsf{H}^2 : x_2=x_3\}$ so that

$$\int \big\langle x_1-x_2,x_1-x_3\big\rangle\,\mathrm{d}\boldsymbol{\gamma}_\tau = \int \big\langle x_1-x_2,x_1-x_2\big\rangle\,\mathrm{d}\boldsymbol{\gamma}_\tau = W_2^2(\mu_\tau,\mu). \tag{4.17}$$

Combining (4.14) with (4.16) and (4.17) we eventually get

$$L = \liminf_{n\to\infty}\phi(\mu_n) \geq \phi(\mu_\tau)+\frac{1-\tau}{\tau}W_2^2(\mu,\mu_\tau)-SW_2(\mu_\tau,\mu). \tag{4.18}$$

Passing to the limit as $\tau\downarrow 0$ in (4.18) and applying (4.9) we obtain $L \geq \phi(\mu)$. □

We make explicit two interesting consequences of the previous result.

Corollary 4.5. *Let $\phi : \mathcal{P}_2(\mathsf{H}) \to \mathbb{R} \cup \{+\infty\}$ be a proper, lower semicontinuous, and geodesically convex function. Then for every $\mu \in \mathcal{P}_2(\mathsf{H})$ and every $\tau > 0$ there exists a solution $\mu_\tau \in D(\phi)$ of the problem*

$$\mu_\tau \quad \textit{minimizes} \quad \nu \mapsto \frac{1}{2\tau} W_2^2(\nu, \mu) + \phi(\nu) \quad \nu \in D(\phi). \tag{4.19}$$

In particular, the proximal point algorithm (1.12) has always a solution for every initial measure $\mu_0 \in \mathcal{P}_2(\mathsf{H})$.

Corollary 4.6. *Let K be a geodesically convex set in $\mathcal{P}_2(\mathsf{H})$, i.e.*

$$\textit{for every } \mu_0, \mu_1 \in K \textit{ there exists } \boldsymbol{\mu} \in \Gamma_o(\mu_0, \mu_1) : (\pi_t^{1\to 2})_\sharp \boldsymbol{\mu} \in K \quad \textit{for every } t \in [0,1]. \tag{4.20}$$

If K is closed in $\mathcal{P}_2(\mathsf{H})$ then it is also (weakly) sequentially closed in $\mathcal{P}_2^w(H)$. In particular

$$\mu_n \in K,\ \sup_n \int |x|^2 \, \mathrm{d}\mu_n < \infty,\ \mu_n \to \mu \textit{ in } \mathcal{P}(\mathsf{H}_w) \textit{ as } n \to \infty \quad \Longrightarrow \quad \mu \in K. \tag{4.21}$$

5 Opial property

Having introduced a notion of weak convergence in $\mathcal{P}_2^w(\mathsf{H})$ (see Definition 3.5) which shares many properties of the weak topology in H, it is natural to investigate if the Opial property holds in $\mathcal{P}_2^w(\mathsf{H})$. This turns out to be true, as stated by the following result.

Theorem 5.1 (Opial property in $\mathcal{P}_2(\mathsf{H})$)**.** *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence weakly converging to μ in $\mathcal{P}_2^w(\mathsf{H})$ according to Definition 3.5. Then*

$$W_2^2(\nu, \mu) + \liminf_{n\to\infty} W_2^2(\mu_n, \mu) \le \liminf_{n\to\infty} W_2^2(\mu_n, \nu) \qquad \textit{for every } \nu \in \mathcal{P}_2(\mathsf{H}). \tag{5.1}$$

In particular,

$$\liminf_{n\to\infty} W_2(\mu_n, \mu) < \liminf_{n\to\infty} W_2(\mu_n, \nu) \qquad \textit{for every } \nu \in \mathcal{P}_2(\mathsf{H}) \textit{ with } \nu \neq \mu.$$

Remark 5.2. Notice that (5.1) holds under the (seemingly) weaker assumption that $\mu_n \to \mu$ narrowly in $\mathcal{P}(\mathsf{H}_w)$. In fact, (5.1) trivially holds if $\liminf_{n\to\infty} W_2^2(\mu_n, \nu) = +\infty$. If the $\liminf$ is finite, then up to extracting a suitable subsequence it is not restrictive to assume that $\liminf_{n\to\infty} W_2^2(\mu_n, \nu) = \lim_{n\to\infty} W_2^2(\mu_n, \nu) < +\infty$ and so μ_n bounded in $\mathcal{P}_2(\mathsf{H})$. In this way we can conclude since, on bounded sets, narrow convergence in $\mathcal{P}(\mathsf{H}_w)$ is equivalent to convergence in $\mathcal{P}_2^w(\mathsf{H})$.

Proof of Theorem 5.1. Let $\nu \in \mathcal{P}_2(\mathsf{H})$. By Corollary 3.6(b) μ_n is bounded in $\mathcal{P}_2(\mathsf{H})$. By Lemma 2.6 for all $n \in \mathbb{N}$ we can find $\boldsymbol{\gamma}_n \in \Gamma(\mu, \nu, \mu_n)$ such that $\pi^{13}_\sharp \boldsymbol{\gamma}_n \in \Gamma_o(\mu, \mu_n)$ and

$\pi^{23}_\sharp\boldsymbol{\gamma}_n \in \Gamma_o(\nu,\mu_n)$. We have

$$
\begin{aligned}
W_2^2(\mu_n,\nu) &= \int_{X^3} |x_3-x_2|^2 \,\mathrm{d}\boldsymbol{\gamma}_n(x_1,x_2,x_3) = \\
&= \int_{X^3} |x_3-x_1|^2 \,\mathrm{d}\boldsymbol{\gamma}_n + \int_{X^3} |x_1-x_2|^2 \,\mathrm{d}\boldsymbol{\gamma}_n + 2\int_{X^3} \langle x_3-x_1, x_1-x_2\rangle \,\mathrm{d}\boldsymbol{\gamma}_n
\end{aligned}
$$

and therefore

$$
W_2^2(\mu_n,\nu) \ge W_2^2(\mu_n,\mu) + W_2^2(\nu,\mu) + 2\int_{X^3} \langle x_3-x_1, x_1-x_2\rangle \,\mathrm{d}\boldsymbol{\gamma}_n. \tag{5.2}
$$

Setting $\mathsf{Z} := (\mathsf{H}^2)_s \times \mathsf{H}_w$, we can apply Proposition 3.4(b) with $p = q = 2$ to the sequence $(\boldsymbol{\gamma}_n)_{n\in\mathbb{N}}$ and find a subsequence $(n_k)_{k\in\mathbb{N}}$ and $\boldsymbol{\gamma} \in \mathcal{P}_2(\mathsf{H}^3)$ such that $\boldsymbol{\gamma}_{n_k} \to \boldsymbol{\gamma}$ in $\mathcal{P}^{sw}_{22}(\mathsf{H}^2\times\mathsf{H})$. By (4.15) and the very definition of the topology of $\mathcal{P}^{sw}_{22}(\mathsf{H}^2\times\mathsf{H})$ we can pass to the limit in (5.2) along the subsequence n_k obtaining

$$
\liminf_{n\to\infty} W_2^2(\mu_n,\nu) \ge \liminf_{n\to\infty} W_2^2(\mu_n,\mu) + W_2^2(\nu,\mu) + 2\int_{X^3} \langle x_3-x_1, x_1-x_2\rangle \,\mathrm{d}\boldsymbol{\gamma}. \tag{5.3}
$$

On the other hand $\pi^{13}_\sharp\boldsymbol{\gamma}_{n_k} \to \pi^{13}_\sharp\boldsymbol{\gamma}$ in $\mathcal{P}^{sw}_{22}(\mathsf{H}\times\mathsf{H})$; since $\pi^{13}_\sharp\boldsymbol{\gamma}_{n_k} \in \Gamma_o(\mu,\mu_{n_k})$, by Theorem 3.8, $\pi^{13}_\sharp\boldsymbol{\gamma} \in \Gamma_o(\mu,\mu)$ so that

$$
\pi^{1,3}_\sharp\boldsymbol{\gamma} = (\mathrm{id}_{\mathsf{H}} \times \mathrm{id}_{\mathsf{H}})_\sharp\mu, \tag{5.4}
$$

thus $\boldsymbol{\gamma}$ is concentrated on the subset $\{(x_1,x_2,x_3)\in\mathsf{H}^3 : x_1 = x_3\}$ and therefore

$$
\int_{X^3} \langle x_3-x_1, x_1-x_2\rangle \,\mathrm{d}\boldsymbol{\gamma} = 0.
$$

Inserting this identity in (5.3) we eventually get (5.1).
The second part of the theorem follows because, by boundedness of μ_n in $\mathcal{P}_2(\mathsf{H})$, the quantities in (5.1) are finite. □

In the simple finite dimensional case of $\mathsf{H} = \mathbb{R}^d$ we obtain the following result.

Corollary 5.3 (Opial property in $\mathcal{P}_2(\mathbb{R}^d)$)**.** *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_2(\mathbb{R}^d)$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. If $\mu_n \to \mu$ narrowly in $\mathcal{P}(\mathbb{R}^d)$, then*

$$
W_2^2(\nu,\mu) + \liminf_{n\to\infty} W_2^2(\mu_n,\mu) \le \liminf_{n\to\infty} W_2^2(\mu_n,\nu) \qquad \text{for every } \nu \in \mathcal{P}_2(\mathbb{R}^d). \tag{5.5}
$$

6 Applications

Let us first enucleate the technical core of many applications of Opial Lemma. We state it in $\mathcal{P}_2(\mathsf{H})$, where H is a separable Hilbert space as in the previous section.

Lemma 6.1. *Let $\mathfrak{T} \subset (0, +\infty)$ be an unbounded set, let $\mu : \mathfrak{T} \to \mathcal{P}_2(\mathsf{H})$ be a* bounded *map and let M be the set of limit points of μ in $\mathcal{P}_2^w(\mathsf{H})$ along diverging sequences:*

$$M := \Big\{ \nu \in \mathcal{P}_2(H_w) : \textit{there exists an increasing sequence } (t_n)_{n\in\mathbb{N}} \subset \mathfrak{T} : \\ \mu(t_n) \to \nu \textit{ in } \mathcal{P}_2^w(\mathsf{H}) \Big\}. \tag{6.1}$$

If

$$\textit{for every } \nu \in M \textit{ the function } t \mapsto W_2(\mu(t), \nu) \textit{ is decreasing in } \mathfrak{T} \tag{6.2}$$

then there exists the limit $\lim_{\substack{t\to\infty \\ t\in\mathfrak{T}}} \mu(t)$ *in* $\mathcal{P}_2^w(\mathsf{H})$.

Proof. Since the set $\mathcal{K} := \{\mu(t) : t \in \mathfrak{T}\}$ is bounded in $\mathcal{P}_2(\mathsf{H})$, it is contained in a compact and metrizable subset of $\mathcal{P}_2^w(\mathsf{H})$. In particular M is not empty and every $(\mu_{t_n})_{n\in\mathbb{N}}$ corresponding to a diverging sequence $t_n \uparrow \infty$, $t_n \in \mathfrak{T}$, has a convergence subsequence in $\mathcal{P}_2^w(\mathsf{H})$. For every $\nu \in M$ we set

$$L(\nu) := \inf_{t\in\mathfrak{T}} W_2^2(\mu(t), \nu) = \lim_{\substack{t\to\infty \\ t\in\mathfrak{T}}} W_2^2(\mu(t), \nu). \tag{6.3}$$

In order to prove the existence of the limit it is therefore sufficient to show that if $s_n, t_n \uparrow +\infty$ as $n \to \infty$ are diverging sequences in $\mathfrak{T}$ such that the corresponding sequences $(\mu(s_n))_{n\in\mathbb{N}}$ and $(\mu(t_n))_{n\in\mathbb{N}}$ respectively converge to ν' and ν'' in $\mathcal{P}_2^w(\mathsf{H})$ then $\nu' = \nu''$.

Since $\nu', \nu'' \in M$, (6.2) yields

$$L(\nu') = \liminf_{n\to\infty} W_2^2(\mu(t_n), \nu') = \liminf_{n\to\infty} W_2^2(\mu(s_n), \nu'),$$
$$L(\nu'') = \liminf_{n\to\infty} W_2^2(\mu(s_n), \nu'') = \liminf_{n\to\infty} W_2^2(\mu(t_n), \nu'').$$

Applying (5.1) of Theorem 5.1 first to the sequence $(\mu(s_n))_{n\in\mathbb{N}}$ and then to the sequence $(\mu(t_n))_{n\in\mathbb{N}}$ we eventually get

$$W_2^2(\nu', \nu'') + L(\nu') \le L(\nu'')$$
$$W_2^2(\nu'', \nu') + L(\nu'') \le L(\nu')$$

which imply $W_2(\nu', \nu'') = 0$. □

6.1 Convergence of Gradient Flows

Let H be a separable Hilbert space and let $\phi : \mathcal{P}_2(\mathsf{H}) \to (-\infty, +\infty]$ be a proper, lower semicontinuous and geodesically convex functional such that $\arg\min \phi$ is not empty.

We want to study the asymptotic behaviour of the gradient flows of ϕ.

Definition 6.2. A locally Lipschitz curve $\mu : (0, \infty) \to \mathcal{P}_2(\mathsf{H})$ is a gradient flow of ϕ in the EVI sense if it satisfies

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\mu_t, \sigma) \le \phi(\sigma) - \phi(\mu_t) \quad \mathscr{L}^1\text{-a.e. in } (0, \infty), \quad \text{for every } \sigma \in D(\phi). \tag{EVI}$$

Theorem 6.3. *Let $\phi : \mathcal{P}_2(\mathsf{H}) \to (-\infty, +\infty]$ be a proper, l.s.c. and geodesically convex functional and let $\mu : (0, +\infty) \to \mathcal{P}_2(\mathsf{H})$ be a Gradient Flow in the* EVI *sense. Then* $\arg\min \phi \neq \emptyset$ *if and only if the curve $(\mu_t)_{t \ge 1}$ is bounded in $\mathcal{P}_2(H)$; in this case there exists $\mu \in \arg\min \phi$ such that $\mu_t \to \mu$ in $\mathcal{P}_2^w(\mathsf{H})$ as $t \to +\infty$.*

Proof. Let us first remark that if ν is a minimizer of ϕ then (EVI) yields

$$t \mapsto W_2(\mu_t, \nu) \quad \text{is decreasing in } (0, +\infty). \tag{6.4}$$

In particular if $\arg\min \phi \neq \emptyset$ the set $\mathcal{K} := \{\mu_t : t \ge 1\}$ is also bounded in $\mathcal{P}_2(\mathsf{H})$

Let us now show that if $\mathcal{K}$ is bounded and μ is a limit point of $(\mu_t)_{t>0}$ along a diverging sequence $t_n \uparrow \infty$ then μ is a minimizer of ϕ (this shows in particular that $\arg\min \phi \neq \emptyset$).

We integrate the (EVI) equation form 1 to $t > 1$ and we divide both sides by $t - 1$, obtaining

$$\frac{1}{2(t-1)} W_2^2(\mu_t, \sigma) + \frac{1}{(t-1)} \int_1^t \phi(\mu_r)\, \mathrm{d}r \le \frac{1}{2(t-1)} W_2^2(\mu_1, \sigma) + \phi(\sigma).$$

Since $t \mapsto \phi(\mu_t)$ is not increasing (see [9, Theorem 3.3]) we have

$$\phi(\mu_t) + \frac{1}{2(t-1)} W_2^2(\mu_t, \sigma) \le \frac{1}{2(t-1)} W_2^2(\mu_1, \sigma) + \phi(\sigma)$$

which yields

$$\limsup_{t \to \infty} \phi(\mu_t) \le \phi(\sigma) \quad \text{for every } \sigma \in D(\phi)$$

since $W_2^2(\mu_t, \sigma)$ is bounded. By the lower semicontinuity of ϕ in $\mathcal{P}_2^w(\mathsf{H})$ we have

$$\phi(\mu) \le \liminf_{k \to \infty} \phi(\mu_{t_k}) \le \limsup_{t \to \infty} \phi(\mu_t) \le \phi(\sigma) \quad \text{for all } \sigma,$$

so that μ is a minimizer of ϕ.

The previous argument shows that the set M defined as in (6.1) (choosing $\mathfrak{T} := [1, \infty)$) is contained in $\arg\min \phi$, so that it satisfies (6.2) thanks to (6.4). Applying Lemma 6.1 we conclude that the curve μ_t converges to a limit $\mu \in M$ as $t \to \infty$ in $\mathcal{P}_2^w(\mathsf{H})$; in particular, μ is a minimizer of ϕ. □

6.2 Weak convergence of the Proximal Point Algorithm

Under the same assumptions of the previous Section 6.1, we want to study the asymptotic properties of the *Proximal Point Algorithm* (1.11). First we define the (multivalued) operator

$$J_\tau(\mu) = \operatorname*{arg\,min}_{\nu \in \mathcal{P}_2(\mathsf{H})} \big\{\Phi_\tau(\mu,\nu)\big\}, \quad \Phi_\tau(\mu,\nu) := \phi(\nu) + \frac{1}{2\tau} W_2^2(\nu,\mu). \tag{6.5}$$

Thanks to Corollary 4.5, for every choice of $\mu_0 \in \mathcal{P}_2(X)$ and $\tau > 0$, the PPA algorithm generates a sequence of points $(\mu_\tau^k)_{k\in\mathbb{N}}$ which solves

$$\begin{cases} \mu_\tau^0 = \mu_0 \\ \mu_\tau^{k+1} \in J_\tau(\mu_\tau^k) \quad k = 1, 2... \end{cases} \tag{6.6}$$

As for the study of the convergence of the Minimizing Movement method in [1], the crucial property to study the asymptotic behaviour of the PPA scheme relies on the notion of convexity along generalized geodesics.

Definition 6.4 (Convexity along generalized geodesics)**.** $\phi : \mathcal{P}_2(\mathsf{H}) \to (-\infty, +\infty]$ is called *convex along generalized geodesics* if for every choice of ν, μ_0, μ_1 in $D(\phi)$ there exists a plan $\boldsymbol{\gamma} \in \Gamma(\nu, \mu_0, \mu_1)$ with $\pi_\sharp^{1,2}\boldsymbol{\gamma} \in \Gamma_o(\nu,\mu_0)$, $\pi_\sharp^{1,3}\boldsymbol{\gamma} \in \Gamma_o(\nu,\mu_1)$, such that

$$\phi(\mu_t^{2\to3}) \le (1-t)\phi(\mu_0) + t\phi(\mu_1) \qquad \forall t \in [0,1].$$

Remark 6.5. The curve $\mu_t^{2\to3}$ defined by

$$\mu_t^{2\to3} = (\pi_t^{2\to3})_\sharp \boldsymbol{\gamma} \qquad t \in [0,1]$$

where $\boldsymbol{\gamma}$ satisfies the conditions of Definition 6.4 is called a *generalized geodesic* connecting μ_0 to μ_1 with reference measure ν. If $\phi : \mathcal{P}_2(\mathsf{H}) \to (-\infty, +\infty]$ is a functional which is convex along generalized geodesics, then for every choice of ν, μ_0, μ_1 in $D(\phi)$ the map $t \mapsto \Phi_\tau(\nu, \mu_t^{2\to3})$ satisfies the inequality

$$\Phi_\tau(\nu, \mu_t^{2\to3}) \le (1-t)\Phi_\tau(\nu,\mu_0) + t\Phi_\tau(\nu,\mu_1) - \frac{1}{2\tau} t(1-t) W_2^2(\mu_0,\mu_1). \tag{6.7}$$

Convexity along generalized geodesics implies convexity along geodesics (see [1, Lemma 9.2.7] for a proof).

Theorem 6.6. *Let us suppose that $\phi : \mathcal{P}_2(\mathsf{H}) \to (-\infty, +\infty]$ is proper, lower semicontinuous, and convex along generalized geodesics, $\mu_0 \in \overline{D(\phi)}$, and $\tau > 0$.*

(i) *The PPA algorithm (6.6) has a unique solution $(\mu_\tau^k)_{k\in\mathbb{N}}$.*

(ii) *For each $\nu \in D(\phi)$ and $k \ge 1$ we have*

$$\frac{1}{2\tau} W_2^2(\mu_\tau^k, \nu) - \frac{1}{2\tau} W_2^2(\mu_\tau^{k-1}, \nu) \le \phi(\nu) - \phi(\mu_\tau^k) - \frac{1}{2\tau} W_2^2(\mu_\tau^k, \mu_\tau^{k-1}). \tag{6.8}$$

(iii) *In particular for every* $k \geq 1$ *we have*

$$\frac{1}{\tau} W_2^2(\mu_\tau^k, \mu_\tau^{k-1}) + \phi(\mu_\tau^k) \leq \phi(\mu_\tau^{k-1}) \tag{6.9}$$

and the sequence $k \mapsto \phi(\mu_\tau^k)$ *is not increasing.*

Proof. See [1, Theorem 4.1.3]. □

Theorem 6.7 (Convergence to a minimum)**.** *Let* $\phi : \mathcal{P}_2(\mathsf{H}) \to (-\infty, +\infty]$ *be proper, lower semicontinuous and convex along generalized geodesics and let* $(\mu_\tau^k)_{k\in\mathbb{N}}$ *be a solution to the PPA algorithm* (6.6)*. Then* $\arg\min \phi \neq \emptyset$ *if and only if* $(\mu_\tau^k)_{k\in\mathbb{N}}$ *is bounded in* $\mathcal{P}_2(\mathsf{H})$*. If this is the case, there exists the limit* $\mu := \lim_{k\to\infty} \mu_\tau^k$ *in* $\mathcal{P}_2^w(\mathsf{H})$ *and* $\mu \in \arg\min \phi$*.*

Proof. If $\nu \in \arg\min \phi$, (6.8) yields

$$\frac{1}{2\tau} W_2^2(\mu_\tau^k, \nu) - \frac{1}{2\tau} W_2^2(\mu_\tau^{k-1}, \nu) \leq 0 \quad \text{for every } k \geq 1$$

so that

$$\text{the sequence} \quad k \mapsto W_2(\mu_\tau^k, \nu) \quad \text{is decreasing;} \tag{6.10}$$

in particular the set $\mathcal{K} := \{\mu_\tau^k : k \in \mathbb{N}\}$ is bounded.

Conversely, if $\mathcal{K}$ is bounded and μ is the weak limit of a subsequence $\mu_\tau^{k(n)}$ in $\mathcal{P}_2^w(\mathsf{H})$ as $n \to \infty$, we want to prove that $\mu \in \arg\min \phi$.

Notice that since $k \mapsto \phi(\mu_\tau^k)$ is not increasing and ϕ is sequentially lower semicontinuous in $\mathcal{P}_2^w(\mathsf{H})$ we have $\mu \in D(\phi)$.

By summing both sides of (6.8) from 1 to K and dividing by K, we obtain for every $\nu \in D(\phi)$

$$\frac{1}{2\tau} \frac{1}{K} \sum_{k=1}^{K} \left(W_2^2(\mu_\tau^k, \nu) - W_2^2(\mu_\tau^{k-1}, \nu) \right) \leq \phi(\nu) - \frac{1}{K} \sum_{k=1}^{K} \phi(\mu_\tau^k)$$

and therefore

$$\frac{1}{K} \sum_{k=1}^{K} \phi(\mu_\tau^k) + \frac{1}{2\tau} \frac{1}{K} W_2^2(\mu_\tau^K, \nu) \leq \phi(\nu) + \frac{1}{2\tau} \frac{1}{K} W_2^2(\mu_\tau^0, \nu).$$

Since $k \to \phi(\mu_\tau^k)$ is not increasing by Theorem 6.6(iii), we have

$$\phi(\mu_\tau^K) + \frac{1}{2\tau} \frac{1}{K} W_2^2(\mu_\tau^K, \nu) \leq \phi(\nu) + \frac{1}{2\tau} \frac{1}{K} W_2^2(\mu_\tau^0, \nu).$$

Taking the $\limsup$ of this inequality as $K \to \infty$ and using the fact that $K \mapsto W_2(\mu_\tau^K, \nu)$ is bounded and ϕ is sequentially lower semicontinuous w.r.t. $\mathcal{P}_2^w(\mathsf{H})$ convergence, we obtain

$$\phi(\mu) \le \liminf_{n\to\infty} \phi(\mu_\tau^{k(n)}) \le \limsup_{K\to\infty} \phi(\mu_\tau^K) \le \phi(\nu) \quad \text{for every } \nu \in D(\phi)$$

so $\mu \in \arg\min \phi$.

The above argument shows that the set M of the limit points of $(\mu_\tau^k)_{k\in\mathbb{N}}$ in $\mathcal{P}_2^w(\mathsf{H})$ (defined as in (6.1) with $\mathfrak{T} := \mathbb{N}$) is included in $\arg\min \phi$, so that it satisfies condition (6.2) thanks to (6.10). We can eventually apply Lemma 6.1 and obtain the weak convergence of $(\mu_\tau^k)_{k\in\mathbb{N}}$ in $\mathcal{P}_2^w(\mathsf{H})$ as $k \to \infty$. □

6.3 Fixed points of non-expansive and asymptotically regular maps

We conclude this section by proving the weak convergence of the iteration of a non-expansive and asymptotically regular map $T : A \to A$ defined in a (weakly) closed subset A of $\mathcal{P}_2^w(\mathsf{H})$. The proof is a simple extension to the Wasserstein setting of the original argument of Opial [10].

Definition 6.8. Let $A \subset \mathcal{P}_2(\mathsf{H})$; a map $T : A \to A$ is called *non-expansive* if

$$W_2(T(\mu), T(\nu)) \le W_2(\mu, \nu) \qquad \text{for all } \mu, \nu \in A$$

T is called *asymptotically regular* if

$$\lim_{k\to\infty} W_2(T^{k+1}(\mu), T^k(\mu)) = 0 \quad \text{for every } \mu \in A.$$

Theorem 6.9. *Let A be a (weakly) closed subset of $\mathcal{P}_2^w(\mathsf{H})$, let $T : A \to A$ be a non-expansive and asymptotically regular map, and let $\mu_k := T^k(\mu)$, $k \in \mathbb{N}$, for some $\mu \in A$. Then T has a fixed point if and only if (μ_k) is bounded in $\mathcal{P}_2(\mathsf{H})$; in this case it converges in $\mathcal{P}_2^w(\mathsf{H})$ to a fixed point μ of T as $k \to \infty$.*

Proof. Let us denote by $\mathrm{Fix}(T)$ the set of fixed points of T. We first observe that

$$\text{for every } \nu \in \mathrm{Fix}(T) \quad \text{the sequence } k \mapsto W_2(\mu_k, \nu) \quad \text{is not increasing.} \tag{6.11}$$

In fact

$$W_2(\mu_{k+1}, \nu) = W_2(T(\mu_k), T(\nu)) \le W_2(\mu_k, \nu) \qquad \text{for every } k \in \mathbb{N},$$

since T is non-expansive. In particular, if $\mathrm{Fix}(T) \neq \emptyset$ then the sequence $(\mu_k)_{k\in\mathbb{N}}$ is bounded.

Let us now suppose that $\mathcal{K} := \{\mu_k : k \in \mathbb{N}\}$ is bounded in $\mathcal{P}_2^w(\mathsf{H})$ and let us show that if μ is the weak limit of $\mu_{k(n)}$ as $n \to \infty$ along an increasing subsequence $n \mapsto k(n)$, then $\mu \in \mathrm{Fix}(T)$. By Opial Lemma we have

$$W_2^2(\mu, T(\mu)) + \liminf_{n\to\infty} W_2^2(\mu, \mu_n) \le \liminf_{n\to\infty} W_2^2(T(\mu), \mu_n). \tag{6.12}$$

Since $\lim_{n\to\infty} W_2(\mu_n, T(\mu_n)) = 0$ by the asymptotic regularity of T, we obtain

$$\liminf_{n\to\infty} W_2^2(T(\mu), \mu_n) = \liminf_{n\to\infty} W_2^2(T(\mu), T(\mu_n)) \le \liminf_{n\to\infty} W_2^2(\mu, \mu_n).$$

Combining this inequality with (6.12) we obtain $W_2(\mu, T(\mu)) = 0$, i.e. $\mu \in \mathrm{Fix}(T)$.

Still assuming that $(\mu_k)_{k\in\mathbb{N}}$ is bounded, we have shown that the set M of its limit points (defined as in (6.1) with $\mathfrak{T} := \mathbb{N}$) is included in $\mathrm{Fix}(T)$ and therefore it satisfies (6.2) thanks to (6.11). An application of Lemma 6.1 concludes the proof. □

Remark 6.10. Since the topology of $\mathcal{P}_q^w(\mathsf{H})$ is finer than the narrow topology of $\mathcal{P}(\mathsf{H}_w)$, in all these applications we also obtain the corresponding narrow convergence results. In particular, in the finite dimensional case $\mathsf{H} = \mathbb{R}^d$, we obtain narrow convergence in $\mathcal{P}(\mathbb{R}^d)$.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] C. Baiocchi. Discretization of evolution variational inequalities. In *Partial differential equations and the calculus of variations, Vol. I*, volume 1 of *Progr. Nonlinear Differential Equations Appl.*, pages 59–92. Birkhäuser Boston, Boston, MA, 1989.

[3] R.E. Bruck, Jr. Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. *J. Functional Analysis*, 18:15–26, 1975.

[4] C. Dellacherie and P. Meyer. *Probabilities and Potential*. vol. 29 of North-Holland Mathematics Studies. North-Holland Publishing Co., 1978.

[5] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[6] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle*, 4(Sér. R-3):154–158, 1970.

[7] B. Martinet. Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox. *C. R. Acad. Sci. Paris Sér. A-B*, 274:A163–A165, 1972.

[8] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.

[9] M. Muratori and G. Savaré. Gradient flows and evolution variational inequalities in metric spaces. I: Structural properties. *J. Funct. Anal.*, 278(4):108347, 67, 2020.

[10] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.*, 73:591–597, 1967.

[11] S. Prus. Banach spaces with the uniform Opial property. *Nonlinear Anal.*, 18(8):697–704, 1992.

[12] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898, 1976.

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

[14] L. Schwartz. *Radon measures on arbitrary topological spaces and cylindrical measures*. Published for the Tata Institute of Fundamental Research, Bombay by Oxford University Press, London, 1973. Tata Institute of Fundamental Research Studies in Mathematics, No. 6.

[15] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.