Chapter 9

Convex Functionals in $\mathscr{P}_p(X)$

The importance of geodesically convex functionals in Wasserstein spaces was firstly pointed out by McCann [97], who introduced the three basic examples we will discuss in detail in 9.3.1, 9.3.4, 9.3.6. His original motivation was to prove the uniqueness of the minimizer of an energy functional which results from the sum of the above three contributions.

Applications of this idea have been given to (im)prove many deep functional (Brunn-Minkowski, Gaussian, (logarithmic) Sobolev, Isoperimetric, etc.) inequalities: we refer to Villani's book [126, Chap. 6] (see also the survey [72]) for a detailed account on this topic. Connections with evolution equations have also been exploited [103, 107, 108, 1, 38], mainly to study the asymptotic decay of the solution to the equilibrium.

From our point of view, convexity is a crucial tool to study the well posedness and the basic regularity properties of gradient flows, as we showed in Chapters 2 and 4. Thus in this chapter we discuss the basic notions and properties related to this concept: the first part of Section 9.1 is devoted to fixing the notion of convexity along geodesics in $\mathscr{P}_p(X)$, avoiding any unnecessary restriction to regular measures; a useful tool for the subsequent developments is the stability of convexity with respect to Γ-convergence, a well known property in the more usual linear theory.

Unfortunately, Example 9.1.5 shows that the squared 2-Wasserstein distance is not convex along geodesics in $\mathscr{P}_2(X)$: this fact and the theory of Chapter 4 motivate the investigation (of convexity properties) along different interpolating curves, along which the squared 2-Wasserstein distance exhibits a nicer behavior; the second part of Section 9.1 discusses this question and introduces the notion of generalized geodesics. Lemma 9.2.7, though simple, provides a crucial link with the metric theory of Chapter 4.

Section 9.3 discusses in great generality the main examples of geodesically convex functionals, showing that they all satisfy also the stronger convexity along generalized geodesics. The last example is related to the semiconcavity properties of the squared 2-Wasserstein distance, discussed in Theorem 7.3.2.

In the last section we give a closer look to the convexity properties of general Relative Entropy functionals, showing that they are strictly related to the log-concavity of the reference measures. Here we use the full generality of the theory, proving all the significant results even in infinite dimensional Hilbert spaces.

9.1 λ-geodesically convex functionals in $\mathscr{P}_p(X)$

In McCann's approach, functionals are naturally defined on $\mathscr{P}_2^r(\mathbb{R}^d)$ so that for each couple of measures μ^1, $\mu^2 \in \mathscr{P}_2^r(\mathbb{R}^d)$ a unique *optimal transport* map $\boldsymbol{t} = \boldsymbol{t}_{\mu^1}^{\mu^2}$ (see (7.1.4)) always exists: in his terminology, a functional $\phi : \mathscr{P}_2^r(\mathbb{R}^d) \to (-\infty, +\infty]$ is *displacement convex* if

$$\begin{aligned} &\text{setting} \quad \mu_t^{1\to 2} := \big(\boldsymbol{i} + t(\boldsymbol{t} - \boldsymbol{i})\big)_{\#}\mu^1, \quad \boldsymbol{t} = \boldsymbol{t}_{\mu^1}^{\mu^2}, \\ &\text{the map} \quad t \in [0,1] \mapsto \phi\big(\mu_t^{1\to 2}\big) \quad \text{is convex,} \quad \forall\, \mu^1,\, \mu^2 \in \mathscr{P}_2^r(\mathbb{R}^d). \end{aligned} \tag{9.1.1}$$

In Section 7.2 we have seen that the curve $\mu_t^{1\to 2}$ is the constant speed geodesic connecting μ^1 to μ^2; therefore the following definition seems natural, when we consider functionals whose domain contains general probability measures.

Definition 9.1.1 (λ-convexity along geodesics). *Let X be a separable Hilbert space and let $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$. Given $\lambda \in \mathbb{R}$, we say that ϕ is* λ-geodesically convex *in $\mathscr{P}_p(X)$ if for every couple μ^1, $\mu^2 \in \mathscr{P}_p(X)$* there exists *an optimal transfer plan $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$ such that*

$$\phi(\mu_t^{1\to 2}) \le (1-t)\phi(\mu^1) + t\phi(\mu^2) - \frac{\lambda}{2} t(1-t) W_p^2(\mu^1, \mu^2) \quad \forall\, t \in [0,1], \tag{9.1.2}$$

where $\mu_t^{1\to 2} = (\pi_t^{1\to 2})_{\#}\boldsymbol{\mu} = \big((1-t)\pi^1 + t\pi^2\big)_{\#}\boldsymbol{\mu}$ is defined as in (7.2.2)*, π^1, π^2 being the projections onto the first and the second coordinate in X^2, respectively.*

Notice that this notion of convexity *depends* on the summability exponent p.

Remark 9.1.2 (The map $t \mapsto \phi(\mu_t^{1\to 2})$ is λ-convex). Actually this definition of λ-convexity expressed through (9.1.2) implies that

$$\text{the map } t \in [0,1] \mapsto \phi(\mu_t^{1\to 2}) \text{ is } \lambda W_p^2(\mu^1, \mu^2)\text{-convex,} \tag{9.1.3}$$

thus recovering an (apparently) stronger and more traditional form.
This equivalence follows easily by the fact, proved in Section 7.2, that for $t_1 < t_2$ in $[0,1]$ with $\{t_1, t_2\} \ne \{0, 1\}$ the plan $\big(\pi_{t_1}^{1\to 2} \times \pi_{t_2}^{1\to 2}\big)_{\#}\boldsymbol{\mu}$ is the *unique* element of $\Gamma_o(\mu_{t_1}^{1\to 2}, \mu_{t_2}^{1\to 2})$.

Notice that in Definition 9.1.1 we *do not* require (9.1.2) along *all* the optimal plans of $\Gamma_o(\mu^1,\mu^2)$. One of the advantage of this technical point is provided by the following proposition, which will be useful to check convexity in many examples.

Proposition 9.1.3 (Convexity criterion). *Let $\phi : \mathscr{P}_p(X) \to (-\infty,+\infty]$ be a l.s.c. map such that for any $\mu \in D(\phi)$ there exists $(\mu_h) \subset \mathscr{P}_p^r(X)$ converging to μ in $\mathscr{P}_p(X)$ with $\phi(\mu_h) \to \phi(\mu)$.*
Then ϕ is λ-geodesically convex iff for each $\mu \in D(\phi) \cap \mathscr{P}_p^r(X)$ and for each μ-essentially injective map $\boldsymbol{r} \in L^p(\mu;X)$ whose graph is $|\cdot|^p$-cyclically monotone the map $t \mapsto \phi\left(((1-t)\boldsymbol{i}+t\boldsymbol{r})_\#\mu\right)$ is λ-convex in $[0,1]$.

Proof. If $\mu^1 \in \mathscr{P}_p^r(X)$ and $\boldsymbol{r} \in L^p(\mu^1;X)$ is $|\cdot|^p$-cyclically monotone, then $\big((1-t)\boldsymbol{i}+t\boldsymbol{r}\big)_\#\mu^1$ is the unique geodesic joining μ^1 to $\mu^2 := \boldsymbol{r}_\#\mu^1$. This shows the necessity of the condition.

In order to show its sufficiency, we notice that if μ^1, $\mu^2 \in \mathscr{P}_p^r(X)$ then a unique optimal map $\boldsymbol{t}_{\mu^1}^{\mu^2}$ exists, it belongs to $L^p(\mu^1;X)$ and it is μ^1-essentially injective (by Remark 6.2.11). Therefore the convexity inequality (9.1.2) holds when the initial and final measure are regular. The general case can be recovered through a standard approximation and compactness argument, as in the proof of the next lemma. □

The following natural Γ-convergence result is well known for convex functionals in linear spaces, see for instance Chapter 11 in [50].

Lemma 9.1.4 (Convexity and Γ-convergence). *Let $\phi_h : \mathscr{P}_p(X) \to (-\infty,+\infty]$ be λ-geodesically convex functionals which $\Gamma\big(\mathscr{P}_p(X)\big)$-converge to ϕ as $n \to \infty$, i.e.*

$$\mu_h \to \mu \text{ in } \mathscr{P}_p(X) \quad \Rightarrow \quad \liminf_{h\to\infty} \phi_h(\mu_h) \geq \phi(\mu), \tag{9.1.4}$$

$$\forall\, \mu \in \mathscr{P}_p(X) \quad \exists \mu_h \to \mu \text{ in } \mathscr{P}_p(X) : \quad \lim_{h\to\infty} \phi_h(\mu_h) = \phi(\mu). \tag{9.1.5}$$

Then ϕ is λ-geodesically convex.
The same result holds for the $\Gamma\big(\mathscr{P}(X)\big)$-convergence if $\lambda \geq 0$, i.e. if we replace convergence in $\mathscr{P}_p(X)$ with narrow convergence *in $\mathscr{P}(X)$ (thus without assuming the convergence of the p-moments of μ_h) in* (9.1.4), (9.1.5).

Proof. Let us fix $\mu^1,\mu^2 \in D(\phi)$; by (9.1.5) we can find sequences μ_h^1, μ_h^2 converging to μ^1,μ^2 in $\mathscr{P}_p(X)$ such that

$$\lim_{n\to\infty} \phi_h(\mu_h^1) = \phi(\mu^1), \quad \lim_{n\to\infty} \phi_h(\mu_h^2) = \phi(\mu^2).$$

Let $\boldsymbol{\mu}_h \in \Gamma_o(\mu_h^1,\mu_h^2)$ be an optimal plan such that (5.1.19) holds for ϕ_h; by Lemma 5.2.2 the sequence $(\boldsymbol{\mu}_h)$ is tight (resp. uniformly p-integrable), because the sequences of their marginals are tight (resp. uniformly p-integrable). Therefore, by Proposition 7.1.5 we can extract a suitable subsequence (still denoted by $\boldsymbol{\mu}_h$) converging to $\boldsymbol{\mu}$ in $\mathscr{P}_p(X \times X)$: we want to show that ϕ is λ-convex along the interpolation $\mu_t^{1\to 2}$ induced by $\boldsymbol{\mu}$.

Since $(\mu_h)_t^{1\to 2} \to \mu_t^{1\to 2}$ in $\mathscr{P}_p(X)$ as $h \to \infty$, (9.1.4) yields easily

$$\begin{aligned}\phi(\mu_t^{1\to 2}) &\le \liminf_{h\to\infty} \phi_h\big((\mu_h)_t^{1\to 2}\big)\\ &\le \liminf_{h\to\infty} \Big((1-t)\phi(\mu_h^1) + t\phi(\mu_h^2) - \frac{\lambda}{2}t(1-t)W_p^2(\mu_h^1,\mu_h^2)\Big)\\ &= (1-t)\phi(\mu^1) + t\phi(\mu^2) - \frac{\lambda}{2}t(1-t)W_p^2(\mu^1,\mu^2). \end{aligned} \tag{9.1.6}$$

In the case of narrow convergence, we can follow the same argument; (9.1.6) becomes an inequality, thanks to (7.1.11), if $\lambda \ge 0$. □

λ-convexity of functionals along geodesics is the simplest condition which allows us to apply the theory developed in Section 2.4. The semigroup generation results of Chapter 4 involve the stronger 1-convexity property of the distance function $W_2^2(\mu^1,\cdot)$ from an arbitrary base point μ^1.

In the 1*-dimensional case* we already know by Theorem 6.0.2 and (7.2.8) that $\mathscr{P}_2(\mathbb{R}^1)$ is isometrically isomorphic to a closed convex subset of an Hilbert space: precisely the space of nondecreasing functions in $(0,1)$ (the inverses of distribution functions), viewed as a subset of $L^2(0,1)$. Thus the 2-Wasserstein distance in $\mathbb{R}$ satisfies the generalized parallelogram rule

$$\begin{aligned} W_2^2(\mu^1,\mu_t^{2\to 3}) = (1-t)W_2^2(\mu^1,\mu^2) + tW_2^2(\mu^1,\mu^3) - t(1-t)W_2^2(\mu^2,\mu^3)\\ \forall\, t\in[0,1],\quad \mu^1,\mu^2,\mu^3 \in \mathscr{P}_2(\mathbb{R}^1). \end{aligned} \tag{9.1.7}$$

If the space X has dimension ≥ 2 the following example shows that there is no constant λ such that $W_2^2(\cdot,\mu^1)$ is λ-convex along geodesics. We will see in the next subsection how to circumvent this difficulty.

Example 9.1.5 (The distance function is not λ-convex along geodesics). Let

$$\mu^2 := \frac{1}{2}\left(\delta_{(0,0)} + \delta_{(2,1)}\right), \qquad \mu^3 := \frac{1}{2}\left(\delta_{(0,0)} + \delta_{(-2,1)}\right).$$

Using for instance Theorem 6.0.1 it is easy to check that the unique optimal map $\boldsymbol{r}$ pushing μ^2 to μ^3 maps $(0,0)$ in $(-2,1)$ and $(2,1)$ in $(0,0)$, therefore there is a unique constant speed geodesic joining the two measures, given by

$$\mu_t^{2\to 3} := \frac{1}{2}\left(\delta_{(-2t,t)} + \delta_{(2-2t,1-t)}\right) \qquad t\in[0,1].$$

Choosing $\mu^1 := \frac{1}{2}\left(\delta_{(0,0)} + \delta_{(0,-2)}\right)$, there are two maps $\boldsymbol{r}_t$, $\boldsymbol{s}_t$ pushing μ^1 to $\mu_t^{2\to 3}$, given by

$$\begin{aligned} \boldsymbol{r}_t(0,0) = (-2t,t),\quad \boldsymbol{r}_t(0,-2) = (2-2t,1-t),\\ \boldsymbol{s}_t(0,0) = (2-2t,1-t),\quad \boldsymbol{s}_t(0,-2) = (-2t,t). \end{aligned}$$

Therefore

$$W_2^2(\mu_t^{2\to 3},\mu^1) = \min\left\{5t^2 - 7t + \frac{13}{2}, 5t^2 - 3t + \frac{9}{2}\right\}$$

has a concave cusp at $t = 1/2$ and therefore is not λ-convex along the geodesic $\mu_t^{2\to 3}$ for any $\lambda \in \mathbb{R}$.

9.2 Convexity along generalized geodesics

In dimension greater than 1, Example 9.1.5 shows that the squared Wasserstein distance functional $\mu \mapsto W_2^2(\mu^1, \mu)$ is not 1-convex along geodesics (in fact, Theorem 7.3.2 shows that it satisfies the opposite inequality).

On the other hand, the theory developed in Chapter 4 indicates that 1-convexity of the squared distance is a quite essential property and that we can exploit the flexibilty in the choice of the connecting curve, along which 1-convexity should be checked. Therefore, here we are looking for such kind of curves (in the case of the "Hilbertian-like" 2-Wasserstein distance) and for the related concept of convexity for functionals.

Let us first suppose that the reference measure μ^1 is regular, i.e. $\mu^1 \in \mathscr{P}_2^r(X)$ and let μ^2, μ^3 be given in $\mathscr{P}_2(X)$; we can find two optimal transport maps $\boldsymbol{t}^2 = \boldsymbol{t}_{\mu^1}^{\mu^2}$, $\boldsymbol{t}^3 = \boldsymbol{t}_{\mu^1}^{\mu^3}$ as in (7.1.4) such that

$$W_2^2(\mu^1, \mu^i) = \int_X |\boldsymbol{t}^i(x) - x|^2 \, d\mu^1(x), \quad i = 2, 3. \tag{9.2.1}$$

Equation (9.2.1) reduces the evaluation of the Wasserstein distance to an integral with respect to the fixed measure μ^1: it is therefore quite natural to interpolate between μ^2 and μ^3 by using $\boldsymbol{t}^2$ and $\boldsymbol{t}^3$, i.e. setting

$$\mu_t^{2\to 3} = (\boldsymbol{t}_t^{2\to 3})_{\#}\mu^1 \quad \text{where} \quad \boldsymbol{t}_t^{2\to 3} := (1-t)\boldsymbol{t}^2 + t\,\boldsymbol{t}^3, \quad t \in [0,1]. \tag{9.2.2}$$

Since $\boldsymbol{t}_t^{2\to 3}$ is obviously cyclically monotone, we have

$$W_2^2(\mu^1, \mu_t^{2\to 3}) = \int_X |\boldsymbol{t}_t^{2\to 3}(x) - x|^2 \, d\mu^1(x) = \int_X |(1-t)\boldsymbol{t}^2(x) + t\boldsymbol{t}^3(x) - x|^2 \, d\mu^1(x),$$

and therefore an easy calculation shows

$$\begin{aligned} W_2^2(\mu^1, \mu_t^{2\to 3}) &= (1-t)\int_X |\boldsymbol{t}^2(x) - x|^2 \, d\mu^1(x) + t\int_X |\boldsymbol{t}^3(x) - x|^2 \, d\mu^1(x) \\ &- t(1-t)\int_X |\boldsymbol{t}^2(x) - \boldsymbol{t}^3(x)|^2 \, d\mu^1(x) \qquad (9.2.3) \\ &\le (1-t)W_2^2(\mu^1, \mu^2) + tW_2^2(\mu^1, \mu^3) - t(1-t)W_2^2(\mu^2, \mu^3), \end{aligned}$$

since

$$\int_X |\boldsymbol{t}^2(x) - \boldsymbol{t}^3(x)|^2 \, d\mu^1(x) \ge W_2^2(\mu^2, \mu^3).$$

This calculation shows that $\frac{1}{2}W_2^2(\mu^1, \cdot)$ is 1-convex along the new interpolating curve $\mu_t^{2\to 3}$ given by (9.2.2).

When μ^1 is not regular, we have to substitute the optimal maps $\boldsymbol{t}_{\mu}^{\mu^2}, \boldsymbol{t}_{\mu^1}^{\mu^3}$ with optimal plans $\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu^1, \mu^2)$, $\boldsymbol{\mu}^{1\,3} \in \Gamma_o(\mu^1, \mu^3)$: in order to interpolate between them, we shall also introduce a 3-plan

$$\begin{gathered}\boldsymbol{\mu} \in \mathscr{P}_2(X^3) \quad \text{such that} \quad \pi^{1\,2}_{\#}\boldsymbol{\mu} = \boldsymbol{\mu}^{1\,2},\ \pi^{1\,3}_{\#}\boldsymbol{\mu} = \boldsymbol{\mu}^{1\,3} \quad \text{and we set}\\ \mu_t^{2\to 3} := (\pi_t^{2\to 3})_{\#}\boldsymbol{\mu}, \quad \text{where} \quad \pi_t^{2\to 3} := (1-t)\pi^2 + t\pi^3.\end{gathered} \tag{9.2.4}$$

Recalling that in (7.3.2) we set

$$W_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) := \int_{X^3} |x_3 - x_2|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3) \geq W_2^2(\mu^2, \mu^3), \tag{9.2.5}$$

we have

Lemma 9.2.1. *Let μ^1, μ^2, $\mu^3 \in \mathscr{P}_2(X)$ and let*

$$\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3) \quad \textit{such that} \quad \boldsymbol{\mu}^{1,i} = \pi^{1\,i}_{\#}\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^i), \quad i = 2, 3. \tag{9.2.6}$$

Then, defining $\mu_t^{2\to 3}$ as in (9.2.4), we get

$$W_2^2(\mu^1, \mu_t^{2\to 3}) = \int_{X^3} |(1-t)x_2 + tx_3 - x_1|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3) \tag{9.2.7a}$$

$$= (1-t)W_2^2(\mu^1, \mu^2) + tW_2^2(\mu^1, \mu^3) - t(1-t)W_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) \tag{9.2.7b}$$

$$\leq (1-t)W_2^2(\mu^1, \mu^2) + tW_2^2(\mu^1, \mu^3) - t(1-t)W_2^2(\mu^2, \mu^3). \tag{9.2.7c}$$

The inequality (9.2.7c) implies that $\frac{1}{2}W_2^2(\mu^1, \cdot)$ is 1-convex along the curve $\mu_t^{2\to 3}$.

Proof. We argue as for (9.2.3), by introducing the transfer plan

$$\boldsymbol{\mu}_t^{1,2\to 3} := \big((1-t)\pi^{1,2} + t\pi^{1,3}\big)_{\#}\boldsymbol{\mu} \in \Gamma(\mu^1, \mu_t^{2\to 3});$$

by the definition of the Wasserstein distance and the Hilbertian identity (12.3.3) it is immediate to see that

$$W_2^2(\mu^1, \mu_t^{2\to 3}) \leq \int_{X\times X} |y_1 - y_2|^2 \, d\boldsymbol{\mu}_t^{1,2\to 3}(y_1, y_2) \tag{9.2.8}$$

$$= \int_{X^3} |(1-t)x_2 + tx_3 - x_1|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3)$$

$$= \int_{X^3} \Big((1-t)|x_2 - x_1|^2 + t|x_3 - x_1|^2 - t(1-t)|x_2 - x_3|^2\Big) \, d\boldsymbol{\mu}(x_1, x_2, x_3). \tag{9.2.9}$$

(9.2.9) yields (9.2.7b) since by (9.2.6) we have

$$\int_{X^3} |x_2 - x_1|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3) = \int_{X^2} |x_2 - x_1|^2 \, d\boldsymbol{\mu}^{1\,2}(x_1, x_2) = W_2^2(\mu^1, \mu^2),$$

$$\int_{X^3} |x_3 - x_1|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3) = \int_{X^2} |x_3 - x_1|^2 \, d\boldsymbol{\mu}^{1\,3}(x_1, x_3) = W_2^2(\mu^1, \mu^3);$$

(9.2.7c) follows directly from the inequality (9.2.5).

Moreover, it is possible to see that (9.2.8) is in fact an equality, i.e. $\boldsymbol{\mu}_t^{1,2\to3} \in \Gamma_o(\mu^1, \mu_t^{2\to3})$, by checking that the support of $\boldsymbol{\mu}_t^{1,2\to3}$ is cyclically monotone; by the density property (5.2.6), we can simply check that $\pi_t^{1,2\to3}\big(\operatorname{supp}\boldsymbol{\mu}\big)$ is cyclically monotone. We choose points $(a_i, b_i) \in \pi_t^{1,2\to3}\big(\operatorname{supp}\boldsymbol{\mu}\big)$, $i = 1, \ldots, N$ and set $(a_0, b_0) := (a_N, b_N)$; we thus find points b_i', b_i'' such that

$$(a_i, b_i') \in \operatorname{supp}\boldsymbol{\mu}^{1,2}, \quad (a_i, b_i'') \in \operatorname{supp}\boldsymbol{\mu}^{1,3}, \quad b_i = (1-t)b_i' + tb_i''.$$

Therefore the cyclical monotonicity of $\operatorname{supp}\boldsymbol{\mu}^{1,i}$ gives

$$\begin{aligned}\sum_{i=1}^N \langle a_i - a_{i-1}, b_i\rangle &= \sum_{i=1}^N \langle a_i - a_{i-1}, (1-t)b_i' + tb_i''\rangle \\ &= (1-t)\sum_{i=1}^N \langle a_i - a_{i-1}, b_i'\rangle + t\sum_{i=1}^N \langle a_i - a_{i-1}, b_i''\rangle \geq 0. \quad \square\end{aligned}$$

Taking account of Lemma 9.2.1, we introduce the following definitions.

Definition 9.2.2 (Generalized geodesics). *A "generalized geodesic" joining μ^2 to μ^3 (with base μ^1) is a curve of the type*

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

where

$$\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3) \quad \text{and} \quad \pi_\#^{1,2}\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2), \quad \pi_\#^{1,3}\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^3). \tag{9.2.10}$$

Remark 9.2.3. Remember that if $\mu^1 \in \mathscr{P}_2^r(X)$ then by Lemma 5.3.2 and Theorem 6.2.10 there exists a unique generalized geodesic connecting μ^2 to μ^3 with base μ^1, since there exists a unique plan $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3)$ satisfying the optimality condition $\pi_\#^{1,i}\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^i)$, $i = 2, 3$. In fact, denoting by $\boldsymbol{t}^i$ the optimal maps $\boldsymbol{t}_{\mu^1}^{\mu^i}$ pushing μ^1 to μ^i, $i = 2, 3$, $\boldsymbol{\mu}$ is given by

$$\boldsymbol{\mu} := (\boldsymbol{i} \times \boldsymbol{t}^2 \times \boldsymbol{t}^3)_\#\mu^1. \tag{9.2.11}$$

We thus recover the expression $\mu_t^{2\to3} = \big((1-t)\boldsymbol{t}^2 + t\,\boldsymbol{t}^3\big)_\#\mu^1$ given by (9.2.2).

Definition 9.2.4 (Convexity along generalized geodesics). *Given $\lambda \in \mathbb{R}$, we say that ϕ is λ-convex along generalized geodesics if for* any *μ^1, μ^2, $\mu^3 \in D(\phi)$ there exists a generalized geodesic $\mu_t^{2\to3}$ induced by a plan $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3)$ satisfying* (9.2.10) *such that*

$$\phi(\mu_t^{2\to3}) \leq (1-t)\phi(\mu^2) + t\phi(\mu^3) - \frac{\lambda}{2}t(1-t)W_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) \quad \forall t \in [0,1], \tag{9.2.12}$$

where $W_{\boldsymbol{\mu}}^2(\cdot,\cdot)$ is defined in (9.2.5).

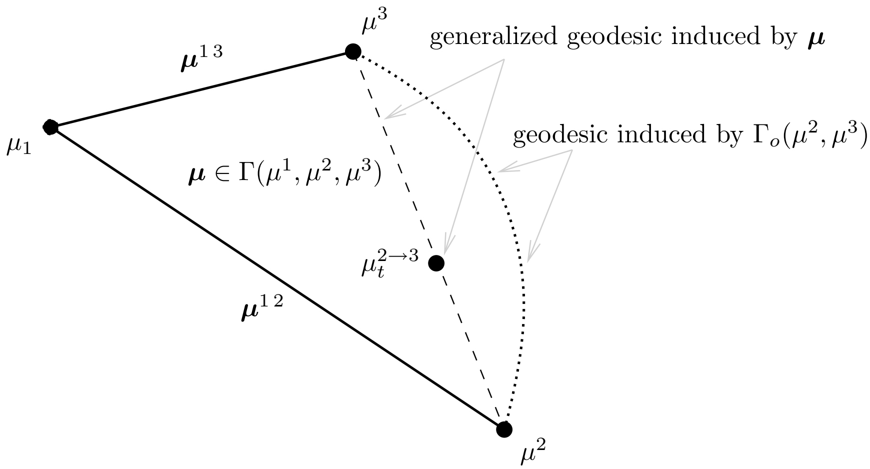

Figure 9.1: Generalized geodesics

Remark 9.2.5 (The case of optimal transport maps). If ϕ is convex along *any* interpolating curve $\mu_t^{2\to 3}$ induced by $\boldsymbol{\mu} \in \Gamma(\mu^2, \mu^3)$, then ϕ is trivially convex along generalized geodesics.

Remark 9.2.6. When $\lambda \neq 0$ Definition 9.2.4 slightly differs from the analogous metric Definition 2.4.1 in the modulus of convexity, since

$$W_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) \geq W_2^2(\mu^2, \mu^3). \tag{9.2.13}$$

In particular, when $\lambda > 0$ this condition is stronger than 2.4.1, whereas for $\lambda < 0$ (9.2.12) is weaker. The next lemma motivates this choice.

Lemma 9.2.7 ($(\tau^{-1} + \lambda)$-convexity of $\Phi(\tau, \mu^1; \cdot)$). *Suppose that $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ is a proper functional which is λ-convex along generalized geodesics for some $\lambda \in \mathbb{R}$. Then for each $\mu^1 \in D(\phi)$ and $0 < \tau < \frac{1}{\lambda^-}$ the functional*

$$\Phi(\tau, \mu^1; \mu) := \frac{1}{2\tau} W_2^2(\mu^1, \mu) + \phi(\mu) \quad \textit{satisfies the convexity Assumption } 4.0.1.$$

Proof. We consider a plan $\boldsymbol{\mu}$ satisfying (9.2.10) and we combine (9.2.7b) and (9.2.12) and use (9.2.13) to obtain

$$\begin{aligned}\Phi(\tau, \mu^1; \mu_t^{2\to 3}) &\leq (1-t)\Phi(\tau, \mu^1; \mu^2) + t\Phi(\tau, \mu^1; \mu^3) - \frac{1}{2}\Big(\frac{1}{\tau} + \lambda\Big) W_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) \\ &\leq (1-t)\Phi(\tau, \mu^1; \mu^2) + t\Phi(\tau, \mu^1; \mu^3) - \frac{1}{2}\Big(\frac{1}{\tau} + \lambda\Big) W_2^2(\mu^2, \mu^3)\end{aligned}$$

whenever $\tau^{-1} > -\lambda$. □

Remark 9.2.8 (Comparison between the two notions of convexity). If ϕ is λ-convex on generalized geodesics then it is also λ-geodesically convex according to Definition 9.1.1: it is sufficient to notice if we choose $\mu^1 = \mu^3$, then any $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3)$ such that $\pi^{1,3}_\# \boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^3)$ is of the form of the form

$$\boldsymbol{\mu} = \int_{X^2} \delta_{x_1}(x_3)\, d\boldsymbol{\mu}^{1\,2}(x_1, x_2) \quad \text{where} \quad \boldsymbol{\mu}^{1\,2} \in \Gamma(\mu^1, \mu^2).$$

Therefore, if we impose also that $\boldsymbol{\mu}^{1\,2} = \pi^{1,2}_\# \boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$, then $\mu_t^{2\to 3}$ is the canonical geodesic interpolation $\big(t\pi^1 + (1-t)\pi^2\big)_\# \boldsymbol{\mu}^{1\,2}$.
We already know by Example 9.1.5 that $\frac{1}{2}W_2(\cdot, \mu^1)$ *is not λ-convex* along geodesics, and therefore *is not λ-convex* along generalized geodesics. On the other hand, if we choose generalized geodesics with base point μ^1 as in (9.2.10), then $\frac{1}{2}W_2^2(\cdot, \mu^1)$ is *indeed* 1−convex along these curves by Lemma 9.2.1. As Lemma 9.2.7 shows, this property is the key point to apply the theory of Chapter 4.

For λ-convex functionals on generalized geodesics we present now two properties which are analogous to the ones stated in Lemma 9.1.4 and Proposition 9.1.3. We omit the proofs, which are similar to the previous ones.

Lemma 9.2.9 (Convexity along generalized geodesics and Γ-convergence). *Let $\phi_h : \mathscr{P}_2(X) \to (-\infty, +\infty]$ be λ-convex on generalized geodesics. If ϕ_h $\Gamma\big(\mathscr{P}_2(X)\big)$-converge to ϕ as $h \to \infty$ as in* (9.1.4), (9.1.5), *then ϕ is λ-convex on generalized geodesics. If $\lambda \ge 0$ the same result holds for $\Gamma\big(\mathscr{P}(X)\big)$-convergence, i.e. Γ-convergence with respect to the* narrow topology *of $\mathscr{P}(X)$.*

Proposition 9.2.10 (A criterion for convexity along generalized geodesics). *Let $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ be a l.s.c. map such that for any $\mu \in D(\phi)$ there exist $(\mu_h) \subset \mathscr{P}_2^r(X)$ converging to μ with $\phi(\mu_h) \to \phi(\mu)$.*

Then ϕ is λ-convex on generalized geodesics iff for every $\mu \in \mathscr{P}_2^r(X)$ and for every couple of μ-essentially injective maps $\boldsymbol{r}^0$, $\boldsymbol{r}^1 \in L^2(\mu; X)$ whose graph is cyclically monotone we have

$$\begin{aligned} \phi\Big(\big((1-t)\boldsymbol{r}^0 + t\boldsymbol{r}^1\big)_\# \mu\Big) &\le (1-t)\phi\big(\boldsymbol{r}^0_\# \mu\big) + t\phi\big(\boldsymbol{r}^1_\# \mu\big) \\ &\quad - \frac{\lambda}{2} t(1-t) \int_X |\boldsymbol{r}^0(x) - \boldsymbol{r}^1(x)|^2\, d\mu(x) \quad \forall\, t \in [0,1]. \end{aligned} \tag{9.2.14}$$

9.3 Examples of convex functionals in $\mathscr{P}_p(X)$

In this section we introduce the main classes of geodesically convex functionals.

Example 9.3.1 (Potential energy). Let $V : X \to (-\infty, +\infty]$ be a proper, lower semicontinuous function whose negative part has a p-growth, i.e.

$$V(x) \ge -A - B|x|^p \quad \forall\, x \in X \quad \text{for some} \quad A, B \in \mathbb{R}. \tag{9.3.1}$$

In $\mathscr{P}_p(X)$ we define

$$\mathcal{V}(\mu) := \int_X V(x)\,d\mu(x). \tag{9.3.2}$$

Evaluating $\mathcal{V}$ on Dirac's masses we check that $\mathcal{V}$ is proper; since V^- is uniformly integrable w.r.t. any sequence (μ_n) converging in $\mathscr{P}_p(X)$ (see Proposition 7.1.5), Lemma 5.1.7 shows that $\mathcal{V}$ is lower semicontinuous in $\mathscr{P}_p(X)$. If V is bounded from below we have even, thanks to (5.1.15), lower semicontinuity w.r.t. narrow convergence.

Recall that for functionals defined on a Hilbert space, λ-convexity means

$$V((1-t)x_1 + tx_2) \le (1-t)V(x_1) + tV(x_2) - \frac{\lambda}{2}t(1-t)|x_1 - x_2|^2 \quad \forall\, x_1, x_2 \in X. \tag{9.3.3}$$

Proposition 9.3.2 (Convexity of $\mathcal{V}$). *If V is λ-convex then for every μ^1, $\mu^2 \in D(\mathcal{V})$ and $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ we have*

$$\mathcal{V}(\mu_t^{1\to 2}) \le (1-t)\mathcal{V}(\mu^1) + t\mathcal{V}(\mu^2) - \frac{\lambda}{2}t(1-t)\int_{X^2} |x_1 - x_2|^2\,d\boldsymbol{\mu}(x_1, x_2). \tag{9.3.4}$$

In particular:

(i) *If $\boldsymbol{p = 2}$ then the functional $\mathcal{V}$ is λ-convex on generalized geodesics, according to Definition* 9.2.4 *(in fact it is λ-convex along any interpolating curve, cf. Remark* 9.2.5*).*

(ii) *If $(\boldsymbol{p \le 2,\ \lambda \ge 0})$ or $(\boldsymbol{p \ge 2,\ \lambda \le 0})$ then $\mathcal{V}$ is λ-geodesically convex in $\mathscr{P}_p(X)$.*

Proof. Since V is bounded from below by a continuous affine functional (if $\lambda \ge 0$) or by a quadratic function (if $\lambda < 0$) its negative part satisfies (9.3.1) for the corresponding values of p considered in this lemma; therefore Definition (9.3.2) makes sense.

Integrating (9.3.3) along any admissible transport plan $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ with $\mu^1, \mu^2 \in D(\mathcal{V})$ we obtain (9.3.4), since

$$\begin{aligned}
\mathcal{V}(\mu_t^{1\to 2}) &= \int_{X^2} V((1-t)x_1 + tx_2)\,d\boldsymbol{\mu}(x_1, x_2)\\
&\le \int_{X^2} \Big((1-t)V(x_1) + tV(x_2) - \frac{\lambda}{2}t(1-t)|x_1 - x_2|^2\Big)\,d\boldsymbol{\mu}(x_1, x_2)\\
&= (1-t)\mathcal{V}(\mu^1) + t\mathcal{V}(\mu^2) - \frac{\lambda}{2}t(1-t)\int_{X^2} |x_1 - x_2|^2\,d\boldsymbol{\mu}(x_1, x_2).
\end{aligned}$$

When $p = 2$ we obtain (9.2.12). When $p \ne 2$ we choose $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$: for $p > 2$ we use the inequality

$$\int_{X^2} |x_1 - x_2|^2\,d\boldsymbol{\mu}(x_1, x_2) \le \Big(\int_{X^2} |x_1 - x_2|^p\,d\boldsymbol{\mu}(x_1, x_2)\Big)^{2/p} = W_p^2(\mu^1, \mu^2),$$

whereas, for $p < 2$, we use the reverse one

$$\int_{X^2} |x_1 - x_2|^2 \, d\boldsymbol{\mu}(x_1, x_2) \geq \Big(\int_{X^2} |x_1 - x_2|^p \, d\boldsymbol{\mu}(x_1, x_2) \Big)^{2/p} = W_p^2(\mu^1, \mu^2).$$

□

Remark 9.3.3. Since $\mathcal{V}(\delta_x) = V(x)$, it is easy to check that the conditions on V are also necessary for the validity of the previous proposition.

Example 9.3.4 (Interaction energy). Let us fix an integer $k > 1$ and let us consider a lower semicontinuous function $W : X^k \to (-\infty, +\infty]$, whose negative part satisfies the usual p-growth condition. Denoting by $\mu^{\times k}$ the measure $\mu \times \mu \times \cdots \times \mu$ on X^k, we set

$$\mathcal{W}_k(\mu) := \int_{X^k} W(x_1, x_2, \ldots, x_k) \, d\mu^{\times k}(x_1, x_2, \ldots, x_k). \tag{9.3.5}$$

If

$$\exists\, x \in X : W(x, x, \ldots, x) < +\infty, \tag{9.3.6}$$

then $\mathcal{W}_k$ is proper; its lower semicontinuity follows from the fact that

$$\mu_n \to \mu \quad \text{in } \mathscr{P}_p(X) \quad \Longrightarrow \quad \mu_n^{\times k} \to \mu^{\times k} \quad \text{in } \mathscr{P}_p(X^k). \tag{9.3.7}$$

Here the typical example is $k = 2$ and $W(x_1, x_2) := \tilde{W}(x_1 - x_2)$ for some $\tilde{W} : X \to (-\infty, +\infty]$ with $\tilde{W}(0) < +\infty$.

Proposition 9.3.5 (Convexity of $\mathcal{W}$). *If W is convex then the functional $\mathcal{W}_k$ is convex along any interpolating curve $\mu_t^{1\to 2}$, $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$, in $\mathscr{P}_p(X)$ (cf. Remark 9.2.5).*

Proof. Observe that $\mathcal{W}_k$ is the restriction to the subset

$$\mathscr{P}_p^{\times}(X^k) := \Big\{ \mu^{\times k} : \mu \in \mathscr{P}_p(X) \Big\}$$

of the potential energy functional $\mathcal{W}$ on $\mathscr{P}_p(X^k)$ given by

$$\mathcal{W}(\boldsymbol{\mu}) := \int_{X^k} W(x_1, \ldots, x_k) \, d\boldsymbol{\mu}(x_1, \ldots, x_k).$$

We consider the linear permutation of coordinates $P : (X^2)^k \to (X^k)^2$ defined by

$$P\Big((x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)\Big) := \Big((x_1, \ldots x_k), (y_1, \ldots y_k)\Big).$$

If $\boldsymbol{\mu} \in \Gamma(\mu_1, \mu_2)$ then it is easy to check that $P_\# \boldsymbol{\mu}^{\times k} \in \Gamma(\mu_1^{\times k}, \mu_2^{\times k}) \subset \mathscr{P}((X^k)^2)$ and

$$(\pi_t^{1\to 2})_\# P_\# (\boldsymbol{\mu}^{\times k}) = P_\# \Big((\pi_t^{1\to 2})_\# \boldsymbol{\mu} \Big)^{\times k}.$$

Therefore all the convexity properties for $\mathcal{W}_k$ follow from the corresponding ones of $\mathcal{W}$. □

In the next example we limit us to consider the finite dimensional case $X := \mathbb{R}^d$, since the Lebesgue measure $\mathscr{L}^d$ will play a distinguished role.

Example 9.3.6 (Internal energy). Let $F : [0, +\infty) \to (-\infty, +\infty]$ be a proper, lower semicontinuous convex function such that

$$F(0) = 0, \quad \liminf_{s\downarrow 0} \frac{F(s)}{s^\alpha} > -\infty \quad \text{for some } \alpha > \frac{d}{d+p}. \tag{9.3.8}$$

We consider the functional $\mathcal{F} : \mathscr{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ defined by

$$\mathcal{F}(\mu) := \begin{cases} \int_{\mathbb{R}^d} F(\rho(x))\, d\mathscr{L}^d(x) & \text{if } \mu = \rho \cdot \mathscr{L}^d \in \mathscr{P}_p^r(\mathbb{R}^d), \\ +\infty & \text{otherwise,} \end{cases} \tag{9.3.9}$$

and its relaxed envelope $\mathcal{F}^*$ defined as

$$\mathcal{F}^*(\mu) := \inf\Big\{ \liminf_{n\to+\infty} \mathcal{F}(\mu_n) : \mu_n \to \mu \quad \text{in } \mathscr{P}_p(\mathbb{R}^d)\Big\}. \tag{9.3.10}$$

Remark 9.3.7 (The meaning of condition (9.3.8)). Condition (9.3.8) simply guarantees that the negative part of $F(\mu)$ is integrable in $\mathbb{R}^d$. For, let us observe that there exist nonnegative constants c_1, c_2 such that the negative part of F satisfies

$$F^-(s) \le c_1 s + c_2 s^\alpha \quad \forall\, s \in [0, +\infty),$$

and it is not restrictive to suppose $\alpha \le 1$. Since $\mu = \rho\, \mathscr{L}^d \in \mathscr{P}_p(\mathbb{R}^d)$ and $\frac{\alpha p}{1-\alpha} > d$ we have

$$\begin{aligned} &\int_{\mathbb{R}^d} \rho^\alpha(x)\, d\mathscr{L}^d(x) = \int_{\mathbb{R}^d} \rho^\alpha(x)(1+|x|)^{\alpha p}(1+|x|)^{-\alpha p}\, d\mathscr{L}^d(x) \\ &\le \Big(\int_{\mathbb{R}^d} \rho(x)(1+|x|)^p\, d\mathscr{L}^d(x)\Big)^\alpha \Big(\int_{\mathbb{R}^d} (1+|x|)^{-\alpha p/(1-\alpha)}\, d\mathscr{L}^d(x)\Big)^{1-\alpha} < +\infty \end{aligned}$$

and therefore $F^-(\rho) \in L^1(\mathbb{R}^d)$.

Remark 9.3.8 (Lower semicontinuity of $\mathcal{F}$). General results on integral functionals [11] show that [79, 31] $\mathcal{F}^* = \mathcal{F}$ on $\mathscr{P}_p^r(\mathbb{R}^d)$ and that $\mathcal{F}^* = \mathcal{F}$ on the whole of $\mathscr{P}_p(\mathbb{R}^d)$ if F has a superlinear growth at infinity.

Proposition 9.3.9 (Convexity of $\mathcal{F}$). *If*

$$\textit{the map} \quad s \mapsto s^d F(s^{-d}) \quad \textit{is convex and non increasing in } (0, +\infty), \tag{9.3.11}$$

then the functionals $\mathcal{F}$, $\mathcal{F}^*$ are convex along (generalized, if $p = 2$) geodesics in $\mathscr{P}_p(\mathbb{R}^d)$.

Proof. By Proposition 9.1.3 we can limit us to check the geodesic convexity of $\mathcal{F}$: thus we consider two regular measures $\mu^i = \rho^i \mathscr{L}^d \in D(\mathcal{F}) \subset \mathscr{P}_p^r(\mathbb{R}^d)$, $i = 1, 2$, and the optimal transport map $\boldsymbol{r}$ for the p-Wasserstein distance W_p such that

$\boldsymbol{r}_{\#}\mu^1 = \mu^2$. Setting $\boldsymbol{r}_t := (1-t)\boldsymbol{i} + t\boldsymbol{r}$, by Theorem 7.2.2 we know that $\boldsymbol{r}_t$ is an optimal transport map between μ^1 and $\mu_t := r_{t\#}\mu^1$ for any $t \in [0,1]$, and Lemma 7.2.1 (for $t \in [0,1)$) and the assumption $\mu^2 \in \mathscr{P}^r_p(\mathbb{R}^d)$ (for $t = 1$) show that $(\boldsymbol{i} \times \boldsymbol{r}_t)_{\#}\mu^1 = (\boldsymbol{s}_t \times \boldsymbol{i})_{\#}\mu_t$ for some optimal transport map $\boldsymbol{s}_t$, therefore $\boldsymbol{s}_t \circ \boldsymbol{r}_t = \boldsymbol{i}$ μ^1-a.e. in $\mathbb{R}^d$. This proves that $\boldsymbol{r}_t$ is μ^1-essentially injective for any $t \in [0,1]$.

By Theorem 6.2.7 we know that $\boldsymbol{r}$ is approximately differentiable μ^1-a.e. and $\tilde\nabla \boldsymbol{r}$ is diagonalizable with nonnegative eigenvalues; since μ^2 is regular, by Lemma 5.5.3 $\det \tilde\nabla \boldsymbol{r}(x) > 0$ for μ^1-a.e. $x \in \mathbb{R}^d$. Therefore $\tilde\nabla \boldsymbol{r}_t$ is diagonalizable, too, with strictly positive eigenvalues: applying Lemma 5.5.3 again we get $\mu_t^{1\to 2} := (\boldsymbol{r}_t)_{\#}\mu^1 \in \mathscr{P}^r_p(\mathbb{R}^d)$ and

$$\mu_t^{1\to 2} = \rho_t \mathscr{L}^d \quad \text{with} \quad \rho_t(\boldsymbol{r}_t(x)) = \frac{\rho^1(x)}{\det \tilde\nabla \boldsymbol{r}_t(x)} \quad \text{for } \mu^1\text{-a.e. } x \in \mathbb{R}^d.$$

By (5.5.3) it follows that

$$\mathcal{F}(\mu_t) = \int_{\mathbb{R}^d} F(\rho_t(y))\,dy = \int_{\mathbb{R}^d} F\Big(\frac{\rho(x)}{\det \tilde\nabla \boldsymbol{r}_t(x)}\Big) \det \tilde\nabla \boldsymbol{r}_t(x)\,dx.$$

Since for a diagonalizable map D with nonnegative eigenvalues

$$t \mapsto \det((1-t)I + tD)^{1/d} \quad \text{is concave in } [0,1], \tag{9.3.12}$$

the integrand above may be seen as the composition of the convex and non-increasing map $s \mapsto s^d F(\rho(x)/s^d)$ and of the concave map in (9.3.12), so that the resulting map is convex in $[0,1]$ for μ^1-a.e. $x \in \mathbb{R}^d$. Thus we have

$$F\Big(\frac{\rho^1(x)}{\det \tilde\nabla \boldsymbol{r}_t(x)}\Big) \det \tilde\nabla \boldsymbol{r}_t(x) \le (1-t)F(\rho^1(x)) + tF(\rho^2(x))$$

and the thesis follows by integrating this inequality in $\mathbb{R}^d$.

In order to check the convexity along generalized geodesics in the case $p = 2$, we apply Proposition 9.2.10: we have to choose $\mu \in \mathscr{P}^r_2(X)$ and two optimal transport maps $\boldsymbol{r}^0, \boldsymbol{r}^1 \in L^2(\mu; X)$, setting $\boldsymbol{r}^t := (1-t)\boldsymbol{r}^0 + t\boldsymbol{r}^1$. We know that $\boldsymbol{r}^0$, $\boldsymbol{r}^1$ are approximately differentiable, μ-essentially injective, and that $\tilde\nabla \boldsymbol{r}^0$, $\tilde\nabla \boldsymbol{r}^1$ are *symmetric* (since $p = 2$) and strictly positive definite for μ-a.e. $x \in \mathbb{R}^d$; moreover, by applying (6.2.9) to $\boldsymbol{r}^0$ and $\boldsymbol{r}^1$ we get

$$\langle \boldsymbol{r}^t(x) - \boldsymbol{r}^t(y), x - y\rangle = (1-t)\langle \boldsymbol{r}^0(x) - \boldsymbol{r}^0(y), x-y\rangle + t\langle \boldsymbol{r}^1(x) - \boldsymbol{r}^1(y), x - y\rangle > 0$$

for x, $y \in \mathbb{R}^d \setminus N$, for a suitable μ-negligible subset N of $\mathbb{R}^d$. It follows that $\boldsymbol{r}^t$ are μ-essentially injective as well and we can argue as before by exploiting the symmetry of $\tilde\nabla \boldsymbol{r}^0$, $\tilde\nabla \boldsymbol{r}^1$, obtaining

$$\mathcal{F}(\mu^t) \le (1-t)\mathcal{F}(\mu_0) + t\mathcal{F}(\mu_1) \quad \text{for } \mu^t := (\boldsymbol{r}^t)_{\#}\mu. \qquad \square$$

In order to express (9.3.11) in a different way, we introduce the function

$$L_F(z) := zF'(z) - F(z) \quad \text{which satisfies } -L_F(e^{-z})e^z = \frac{d}{dz}F(e^{-z})e^z; \tag{9.3.13}$$

denoting by $\hat F$ the modified function $F(e^{-z})e^z$ we have the simple relation

$$\hat L_F(z) = -\frac{d}{dz}\hat F(z), \quad \widehat{L^2_F}(z) = -\frac{d}{dz}\hat L_F(z) = \frac{d^2}{dz^2}\hat F(z), \quad \text{where} \\ L^2_F(z) := L_{L_F}(z) = zL'_F(z) - L_F(z). \tag{9.3.14}$$

The nonincreasing part of condition (9.3.11) is equivalent to say that

$$L_F(z) \geq 0 \quad \forall\, z \in (0,+\infty), \tag{9.3.15}$$

and it is in fact implied by the convexity of F. A simple computation in the case $F \in C^2(0,+\infty)$ shows

$$\frac{d^2}{ds^2}F(s^{-d})s^d = \frac{d^2}{ds^2}\hat F(d\cdot\log s) = \hat L^2_F(d\cdot\log s)\frac{d^2}{s^2} + \hat L_F(d\cdot\log s)\frac{d}{s^2},$$

and therefore

$$\text{(9.3.11) is equivalent to } L^2_F(z) \geq -\frac{1}{d}L_F(z) \quad \forall\, z\in(0,+\infty), \tag{9.3.16}$$

i.e.

$$zL'_F(z) \geq \big(1-\frac{1}{d}\big)L_F(z), \quad \text{the map } z\mapsto z^{1/d-1}L_F(z) \text{ is non increasing.} \tag{9.3.17}$$

Observe that the bigger is the dimension d, the stronger are the above conditions, which always imply the convexity of F.

Remark 9.3.10 (A "dimension free" condition). The weakest condition on F yielding the geodesic convexity of $\mathcal{F}$ in *any dimension* is therefore

$$L^2_F(z) = zL'_F(z) - L_F(z) \geq 0 \quad \forall\, z\in(0,+\infty). \tag{9.3.18}$$

Taking into account (9.3.14), this is also equivalent to ask that

$$\textit{the map} \quad s\mapsto F(e^{-s})e^s \quad \textit{is convex and non increasing in } (0,+\infty). \tag{9.3.19}$$

Among the functionals F satisfying (9.3.11) we quote:

$$\text{the entropy functional:} \quad F(s) = s\log s, \tag{9.3.20}$$

$$\text{the power functional:} \quad F(s) = \frac{1}{m-1}s^m \quad \text{for } m \geq 1-\frac{1}{d}. \tag{9.3.21}$$

Observe that (9.3.20) and (9.3.21) with $m > 1$ also satisfy (9.3.19) and $\mathcal{F} = \mathcal{F}^*$, by Remark 9.3.8; on the other hands, if $m < 1$, $\mathcal{F}^*$ is given by [79, 31]

$$\mathcal{F}^*(\mu) := \frac{1}{m-1}\int_{\mathbb{R}^d} F(\rho(x))\, d\mathscr{L}^d(x) \quad \text{with } \mu = \rho\cdot\mathscr{L}^d + \mu_s,\ \mu_s \perp \mathscr{L}^d. \tag{9.3.22}$$

In this case the functional takes only account of the density of the absolutely continuous part of μ w.r.t. $\mathscr{L}^d$ and the domain of $\mathcal{F}^*$ is the whole $\mathscr{P}_p(\mathbb{R}^d)$, which strictly contains $\mathscr{P}_p^r(\mathbb{R}^d)$.

Example 9.3.11 (The opposite Wasserstein distance). In the separable Hilbert space X let us fix a base measure $\mu^1 \in \mathscr{P}_2(X)$ and let us consider the functional

$$\phi(\mu) := -\frac{1}{2}W_2^2(\mu^1,\mu). \tag{9.3.23}$$

Proposition 9.3.12. *For each couple μ^2, $\mu^3 \in \mathscr{P}_2(X)$ and each transfer plan $\boldsymbol{\mu}^{2\,3} \in \Gamma(\mu^2,\mu^3)$ we have*

$$\begin{aligned} W_2^2(\mu^1,\mu_t^{2\to3}) \ge (1-t)W_2^2(\mu^1,\mu^2) + tW_2^2(\mu^1,\mu^3)& \\ - t(1-t)\int_{X^2} |x_2-x_3|^2\, d\boldsymbol{\mu}^{2\,3}(x_2,x_3) \quad \forall\, t\in[0,1].& \end{aligned} \tag{9.3.24}$$

In particular, by Remark 9.2.5, the map $\phi : \mu \mapsto -\frac{1}{2}W_2^2(\mu^1,\mu)$ is (-1)-convex along generalized geodesics.

Proof. We argue as in Theorem 7.3.2: by Proposition 7.3.1, for μ^2, $\mu^3 \in \mathscr{P}_2(X)$ and $\boldsymbol{\mu}^{2\,3} \in \Gamma(\mu^2,\mu^3)$ we can find a plan $\boldsymbol{\mu} \in \Gamma(\mu^1,\mu^2,\mu^3)$ such that

$$(\pi_t^{1,2\to3})_\#\boldsymbol{\mu} \in \Gamma_o(\mu^1,\mu_t^{2\to3}), \quad (\pi^{2,3})_\#\boldsymbol{\mu} = \boldsymbol{\mu}^{2\,3}. \tag{9.3.25}$$

Therefore

$$\begin{aligned} W_2^2(\mu^1,\mu_t^{2\to3}) &= \int_{X^3} |(1-t)x_2 + tx_3 - x_1|^2\, d\boldsymbol{\mu}(x_1,x_2,x_3) \\ &= \int_{X^3} \Big((1-t)|x_2-x_1|^2 + t|x_3-x_1|^2 - t(1-t)|x_2-x_3|^2\Big)\, d\boldsymbol{\mu}(x_1,x_2,x_3) \\ &\ge (1-t)W_2^2(\mu^1,\mu^2) + tW_2^2(\mu^1,\mu^3) - t(1-t)\int_{X^2}|x_2-x_3|^2\, d\boldsymbol{\mu}^{2\,3}(x_2,x_3). \end{aligned}$$

□

9.4 Relative entropy and convex functionals of measures

In this section we study in detail the case of relative entropies, which extend even to infinite dimensional spaces the example (9.3.20) discussed in 9.3.6: for more details and developments we refer to [67].

Definition 9.4.1 (Relative entropy). *Let γ, μ be Borel probability measures on a separable Hilbert space X; the* relative entropy *of μ w.r.t. γ is*

$$\mathcal{H}(\mu|\gamma) := \begin{cases} \displaystyle\int_X \frac{d\mu}{d\gamma} \log\left(\frac{d\mu}{d\gamma}\right) d\gamma & \text{if } \mu \ll \gamma, \\ +\infty & \text{otherwise.} \end{cases} \tag{9.4.1}$$

As in Example 9.3.6 we introduce the nonnegative, l.s.c., (extended) real, (strictly) convex function

$$H(s) := \begin{cases} s(\log s - 1) + 1 & \text{if } s > 0, \\ 1 & \text{if } s = 0, \\ +\infty & \text{if } s < 0, \end{cases} \tag{9.4.2}$$

and we observe that

$$\mathcal{H}(\mu|\gamma) = \int_X H\Big(\frac{d\mu}{d\gamma}\Big)\, d\gamma \geq 0; \quad \mathcal{H}(\mu|\gamma) = 0 \quad \Leftrightarrow \quad \mu = \gamma. \tag{9.4.3}$$

Remark 9.4.2 (Changing γ). Let γ be a Borel measure on X and let $V : X \to (-\infty, +\infty]$ a Borel map such that

$$V^+ \quad \text{has } p\text{-growth (5.1.21)}, \quad \tilde{\gamma} := e^{-V} \cdot \gamma \quad \text{is a probability measure.} \tag{9.4.4}$$

Then for measures in $\mathscr{P}_p(X)$ the relative entropy w.r.t. γ is well defined by the formula

$$\mathcal{H}(\mu|\gamma) := \mathcal{H}(\mu|\tilde{\gamma}) - \int_X V(x)\, d\mu(x) \in (-\infty, +\infty] \quad \forall\, \mu \in \mathscr{P}_p(X). \tag{9.4.5}$$

In particular, when $X = \mathbb{R}^d$ and γ is the d-dimensional Lebesgue measure, we find the standard entropy functional introduced in (9.3.20).

More generally, we can consider a

$$\begin{gathered} \text{proper, l.s.c., convex function } F : [0, +\infty) \to [0, +\infty] \\ \text{with superlinear growth} \end{gathered} \tag{9.4.6}$$

and the related functional

$$\mathcal{F}(\mu|\gamma) := \begin{cases} \displaystyle\int_X F\Big(\frac{d\mu}{d\gamma}\Big)\, d\gamma & \text{if } \mu \ll \gamma, \\ +\infty & \text{otherwise.} \end{cases} \tag{9.4.7}$$

Lemma 9.4.3 (Joint lower semicontinuity). *Let γ^n, $\mu^n \in \mathscr{P}(X)$ be two sequences narrowly converging to γ, μ in $\mathscr{P}(X_\varpi)$. Then*

$$\liminf_{n\to\infty} \mathcal{H}(\mu^n|\gamma^n) \geq \mathcal{H}(\mu|\gamma), \quad \liminf_{n\to\infty} \mathcal{F}(\mu^n|\gamma^n) \geq \mathcal{F}(\mu|\gamma). \tag{9.4.8}$$

The proof of this lemma follows easily from the next representation formula; before stating it, we need to introduce the conjugate function of F

$$F^*(s^*) := \sup_{s\geq 0} s\cdot s^* - F(s) < +\infty \quad \forall\, s^*\in\mathbb{R}, \tag{9.4.9}$$

so that

$$F(s) = \sup_{s^*\in\mathbb{R}} s^*\cdot s - F^*(s^*); \tag{9.4.10}$$

if $s_0\geq 0$ is a minimizer of F then

$$F^*(s^*)\geq s^* s_0 - F(s_0), \quad s\geq s_0 \quad \Rightarrow \quad F(s) = \sup_{s^*\geq 0} s^*\cdot s - F^*(s^*). \tag{9.4.11}$$

In the case of the entropy functional, we have $H^*(s^*) = e^{s^*}-1$.

Lemma 9.4.4 (Duality formula). *For any γ, $\mu\in\mathscr{P}(X)$ we have*

$$\mathcal{F}(\mu|\gamma) = \sup\Big\{\int_X S^*(x)\,d\mu(x) - \int_X F^*(S^*(x))\,d\gamma(x) : S^*\in C^0_b(X_\varpi)\Big\}. \tag{9.4.12}$$

Proof. This lemma is a particular case of more general results on convex integrals of measures, well known in the case of a finite dimensional space X, see for instance §2.6 of [11]. We present here a brief sketch of the proof for a general Hilbert space; up to an addition of a constant, we can always assume $F^*(0) = -\min_{s\geq 0}F(s) = -F(s_0) = 0$.

Let us denote by $\mathcal{F}'(\mu|\gamma)$ the right hand side of (9.4.12). It is obvious that $\mathcal{F}'(\mu|\gamma)\leq\mathcal{H}(\mu|\gamma)$, so that we have to prove only the converse inequality.

First of all we show that $\mathcal{F}'(\mu|\gamma)<+\infty$ yields that $\mu\ll\gamma$. For let us fix s^*, $\varepsilon>0$ and a Borel set A with $\gamma(A)\leq\varepsilon/2$. Since μ, γ are tight measures (recall that $\mathscr{B}(X)=\mathscr{B}(X_\varpi)$, compact subset of X are compact in X_ϖ, too, and X_ϖ is a separable metric space) we can find a compact set $K\subset A$, an open set (in X_ϖ) $G\supset A$ and a continuous function $\zeta: X_\varpi\to[0,s^*]$ such that

$$\mu(G\setminus K)\leq\varepsilon, \quad \gamma(G)\leq\varepsilon, \quad \zeta(x)=s^* \quad \text{on } K, \quad \zeta(x)=0 \quad \text{on } X\setminus G.$$

Since F^* is increasing (by Definition (9.4.9)) and $F^*(0)=0$, we have

$$\begin{aligned} s^*\mu(K) - F^*(s^*)\varepsilon &\leq \int_K \zeta(x)\,d\mu(x) - \int_G F^*(\zeta(x))\,d\gamma(x)\\ &\leq \int_X \zeta(x)\,d\mu(x) - \int_X F^*(\zeta(x))\,d\gamma(x) \leq \mathcal{F}'(\mu|\gamma) \end{aligned}$$

Taking the supremum w.r.t. $K\subset A$ and $s^*\geq 0$, and using (9.4.11) we get

$$\varepsilon F\big(\mu(A)/\varepsilon\big)\leq\mathcal{F}'(\mu|\gamma) \quad \text{if } \mu(A)\geq\varepsilon s_0.$$

Since $F(s)$ has a superlinear growth as $s\to+\infty$, we conclude that $\mu(A)\to 0$ as $\varepsilon\downarrow 0$.

Now we can suppose that $\mu = \rho \cdot \gamma$ for some Borel function $\rho \in L^1(\gamma)$, so that

$$\mathcal{F}'(\mu|\gamma) = \sup\Big\{ \int_X \big(S^*(x)\rho(x) - F^*(S^*(x))\big)\, d\gamma(x) : S^* \in C^0_b(X_\varpi)\Big\}$$

and, for a suitable dense countable set $C = \{s^*_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$

$$\begin{aligned}\mathcal{F}(\mu|\gamma) &= \int_X \sup_{s^*\in C} \big(s^*\rho(x) - F^*(s^*)\big)\, d\gamma(x)\\ &= \lim_{k\to\infty} \int_X \sup_{s^*\in C_k} \big(s^*\rho(x) - F^*(s^*)\big)\, d\gamma(x)\end{aligned}$$

where $C_k = \{s^*_1, \cdots, s^*_k\}$. Our thesis follows if we show that for every k

$$\int_X \max_{s^*\in C_k} \big(s^*\rho(x) - F^*(s^*)\big)\, d\gamma(x) \le \mathcal{F}'(\mu|\gamma). \tag{9.4.13}$$

For we call

$$A_j = \Big\{x \in X : s^*_j\rho(x) - F^*(s^*_j) \ge s^*_i\rho(x) - F^*(s^*_i) \quad \forall\, i \in \{1, \dots, k\}\Big\},$$

and

$$A'_1 = A_1, \quad A'_{j+1} = A_{j+1} \setminus \Big(\bigcup_{i=1}^{j} A_i\Big).$$

Since γ is Radon, we find compact sets $K_j \subset A'_j$, X_ϖ-open sets $G_j \supset A_j$ with $G_j \cap K_i = \varnothing$ if $i \ne j$, and X_ϖ-continuous functions ζ_j such that

$$\sum_{j=1}^{k} \gamma(G_j \setminus K_j) + \mu(G_j \setminus K_j) \le \varepsilon, \quad \zeta_j \equiv s^*_j \text{ on } K_j, \quad \zeta_j \equiv 0 \text{ on } X \setminus G_j.$$

Denoting by $\zeta := \sum_{j=1}^k \zeta_j$, $M := \sum_{j=1}^k |s^*_j|$, since the negative part of $F^*(s^*)$ is bounded above by $|s^*|s_0$ we have

$$\begin{aligned}\int_X \max_{s^*\in C_k} \big(s^*\rho(x) - F^*(s^*)\big)\, d\gamma(x) &= \sum_{j=1}^{k} \int_{A'_j} \big(s^*_j\rho(x) - F^*(s^*_j)\big)\, d\gamma(x)\\ &\le \sum_{j=1}^{k} \int_{K_j} \big(s^*_j\rho(x) - F^*(s^*_j)\big)\, d\gamma(x) + \varepsilon(M + Ms_0)\\ &= \sum_{j=1}^{k} \int_{K_j} \big(\zeta(x)\rho(x) - F^*(\zeta(x))\big)\, d\gamma(x) + \varepsilon(M + Ms_0)\\ &\le \int_X \big(\zeta(x)\rho(x) - F^*(\zeta(x))\big)\, d\gamma(x) + \varepsilon(M + Ms_0 + M + F^*(M)).\end{aligned}$$

Passing to the limit as $\varepsilon \downarrow 0$ we get (9.4.13). □

Lemma 9.4.5 (Entropy and marginals). *Let $\pi : X \to X$ be a Borel map. For every couple of probability measures $\gamma, \mu \in \mathscr{P}(X)$ we have*

$$\mathcal{H}(\pi_\# \mu | \pi_\# \gamma) \leq \mathcal{H}(\mu|\gamma), \quad \mathcal{F}(\pi_\# \mu | \pi_\# \gamma) \leq \mathcal{F}(\mu|\gamma). \tag{9.4.14}$$

Proof. It is not restrictive to assume that $\mu \ll \gamma$: we denote by ρ a Borel map γ-a.e. equal to the density $\frac{d\mu}{d\gamma}$; applying the disintegration theorem we can find a Borel family of probability measures γ_x in X such that $\gamma = \int_X \gamma_x \, d\,\pi_\#\gamma(x)$ and $\gamma_x(X \setminus \pi^{-1}(x)) = 0$ for $\pi_\#\gamma$-a.e. x.

It follows that μ and $\pi_\#\mu$ admit the representation

$$\mu = \int_X \rho \gamma_x \, d\,\pi_\# \gamma(x) \quad \text{and} \quad \pi_\# \mu = \tilde{\rho} \cdot \pi_\# \gamma \quad \text{with} \quad \tilde{\rho}(x) := \int_{\pi^{-1}(x)} \rho(y) \, d\gamma_x(y)$$

since for each Borel set $A \subset X$ one has

$$\int_{\pi^{-1}(A)} d\mu(x) = \int_A \Big(\int_{\pi^{-1}(x)} \rho(y) \, d\gamma_x(y) \Big) \, d\,\pi_\# \gamma(x).$$

Jensen inequality yields

$$F(\tilde{\rho}(x)) \leq \int_{\pi^{-1}(x)} F(\rho(y)) \, d\gamma_x(y),$$

and therefore

$$\begin{aligned} \mathcal{F}(\pi_\# \mu | \pi_\# \gamma) = \int_X F(\tilde{\rho}(x)) \, d\,\pi_\# \gamma(x) &\leq \int_X \Big(\int_{\pi^{-1}(x)} F(\rho(y)) \, d\gamma_x(y) \Big) \, d\,\pi_\# \gamma(x) \\ &\leq \int_X F(\rho(x)) \, d\gamma(x) = \mathcal{F}(\mu|\gamma). \end{aligned}$$

□

Corollary 9.4.6. *Let $\pi^k : X \to X$ be Borel maps such that*

$$\lim_{k\to\infty} \pi^k(x) = x \quad \forall\, x \in X.$$

For every γ, $\mu \in \mathscr{P}(X)$, setting $\gamma^k := \pi^k_\# \gamma$, $\mu^k := \pi^k_\# \mu$, we have

$$\lim_{k\to\infty} \mathcal{H}(\mu^k|\gamma^k) = \mathcal{H}(\mu|\gamma), \quad \lim_{k\to\infty} \mathcal{F}(\mu^k|\gamma^k) = \mathcal{F}(\mu|\gamma). \tag{9.4.15}$$

Proof. Lebesgue's dominated convergence theorem shows that γ^k, μ^k narrowly converge to γ, μ respectively. Combining Lemma 9.4.3 and 9.4.5 we conclude. □

9.4.1 Log-concavity and displacement convexity

We want to characterize the probability measures γ inducing a geodesically convex relative entropy functional $\mathcal{H}(\cdot|\gamma)$ in $\mathscr{P}_p(X)$. The following lemma provides the first crucial property; the argument is strictly related to the proof of the Brunn-Minkowski inequality for the Lebesgue measure, obtained via optimal transportation inequalities [126]. See also [25] for the link between log-concavity and representation formulae like (9.4.23).

Lemma 9.4.7 (γ is log-concave if $\mathcal{H}(\cdot|\gamma)$ is displacement convex). *Suppose that for each couple of probability measures $\mu^1, \mu^2 \in \mathscr{P}(X)$ with bounded support, there exists $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ such that $\mathcal{H}(\cdot|\gamma)$ is convex along the interpolating curve $\mu_t^{1\to 2} = \big((1-t)\pi^1 + t\pi^2\big)_{\#}\boldsymbol{\mu}$, $t \in [0,1]$. Then for each couple of open sets A, $B \subset X$ and $t \in [0,1]$ we have*

$$\log \gamma((1-t)A + tB) \geq (1-t)\log\gamma(A) + t\log\gamma(B). \tag{9.4.16}$$

Proof. We can obviously assume that $\gamma(A) > 0$, $\gamma(B) > 0$ in (9.4.16); we consider

$$\mu^1 := \gamma(\cdot|A) = \frac{1}{\gamma(A)}\chi_A \cdot \gamma, \quad \mu^2 := \gamma(\cdot|B) = \frac{1}{\gamma(B)}\chi_B \cdot \gamma,$$

observing that

$$\mathcal{H}(\mu^1|\gamma) = -\log\gamma(A), \quad \mathcal{H}(\mu^2|\gamma) = -\log\gamma(B). \tag{9.4.17}$$

If $\mu_t^{1\to 2}$ is induced by a transfer plan $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ along which the relative entropy is displacement convex, we have

$$\mathcal{H}(\mu_t^{1\to 2}|\gamma) \leq (1-t)\mathcal{H}(\mu^1|\gamma) + t\mathcal{H}(\mu^2|\gamma) = -(1-t)\log\gamma(A) - t\log\gamma(B).$$

On the other hand the measure $\mu_t^{1\to 2}$ is concentrated on $(1-t)A + tB = \pi_t^{1\to 2}(A \times B)$ and the next lemma shows that

$$-\log\gamma((1-t)A + tB) \leq \mathcal{H}(\mu_t^{1\to 2}|\gamma). \qquad \square$$

Lemma 9.4.8 (Relative entropy of concentrated measures). *Let $\gamma, \mu \in \mathscr{P}(X)$; if μ is concentrated on a Borel set A, i.e. $\mu(X \setminus A) = 0$, then*

$$\mathcal{H}(\mu|\gamma) \geq -\log\gamma(A). \tag{9.4.18}$$

Proof. It is not restrictive to assume $\mu \ll \gamma$ and $\gamma(A) > 0$; denoting by γ_A the probability measure $\gamma(\cdot|A) := \gamma(A)^{-1}\chi_A \cdot \gamma$, we have

$$\begin{aligned}\mathcal{H}(\mu|\gamma) &= \int_X \log\Big(\frac{d\mu}{d\gamma}\Big)\,d\mu = \int_A \log\Big(\frac{d\mu}{d\gamma_A}\cdot\frac{1}{\gamma(A)}\Big)\,d\mu \\ &= \int_A \log\Big(\frac{d\mu}{d\gamma_A}\Big)\,d\mu - \int_A \log\big(\gamma(A)\big)\,d\mu = \mathcal{H}(\mu|\gamma_A) - \log\big(\gamma(A)\big) \\ &\geq -\log\big(\gamma(A)\big).\end{aligned}$$

$\square$

The previous results justifies the following definition:

Definition 9.4.9 ($\log$**-concavity of a measure).** *We say that a Borel probability measure $\gamma \in \mathscr{P}(X)$ on X is* log-concave *if for every couple of open sets A, $B \subset X$ we have*

$$\log \gamma((1-t)A + tB) \geq (1-t)\log\gamma(A) + t\log\gamma(B). \tag{9.4.19}$$

In Definition 9.4.9 and also in the previous theorem we confined ourselves to pairs of open sets, to avoid the non trivial issue of the measurability of $(1-t)A+tB$ when A and B are only Borel (in fact, it is an open set whenever A and B are open). Observe that a log-concave measure γ in particular satisfies

$$\log\gamma(B_r((1-t)x_0 + tx_1)) \geq (1-t)\log\gamma(B_r(x_0)) + t\log\gamma(B_r(x_1)), \tag{9.4.20}$$

for every couple of points $x_0, x_1 \in X$, $r > 0$, $t \in [0,1]$.

We want to show that in fact log concavity is equivalent to the geodesic convexity of the Relative Entropy functional $\mathcal{H}(\cdot|\gamma)$.

Let us first recall some elementary properties of convex sets in $\mathbb{R}^d$. Let $C \subset \mathbb{R}^d$ be a convex set; the *affine dimension* $\dim C$ of C is the linear dimension of its affine envelope

$$\operatorname{aff} C = \Big\{(1-t)x_0 + tx_1 : x_0, x_1 \in C,\ t \in \mathbb{R}\Big\}, \tag{9.4.21}$$

which is an affine subspace of $\mathbb{R}^d$. We denote by $\operatorname{int} C$ the relative interior of C as a subset of $\operatorname{aff} C$: it is possible to show that

$$\operatorname{int} C \neq \varnothing, \quad \overline{\operatorname{int} C} = \overline{C}, \quad \mathscr{H}^k(\overline{C}\setminus \operatorname{int} C) = 0 \quad \text{if } k = \dim C. \tag{9.4.22}$$

Theorem 9.4.10. *Let us suppose that $X = \mathbb{R}^d$ is finite dimensional and $\gamma \in \mathscr{P}(X)$ satisfies the* log-concavity *assumptions on balls* (9.4.20). *Then* $\operatorname{supp}\gamma$ *is convex and there exists a convex l.s.c. function $V : X \to (\infty, +\infty]$ such that*

$$\gamma = e^{-V}\cdot \mathscr{H}^k|_{\operatorname{aff}(\operatorname{supp}\gamma)}, \qquad \textit{where } k = \dim(\operatorname{supp}\gamma). \tag{9.4.23}$$

Conversely, if γ admits the representation (9.4.23) *then γ is* log-concave *and the relative entropy functional $\mathcal{H}(\cdot|\gamma)$ is convex along any (generalized, if $p = 2$) geodesic of $\mathscr{P}_p(X)$.*

Proof. Let us suppose that γ satisfies the log-concave inequality on balls and let k be the dimension of $\operatorname{aff}(\operatorname{supp}\gamma)$. Observe that the measure γ satisfies the same inequality (9.4.20) for the balls of $\operatorname{aff}(\operatorname{supp}\gamma)$: up to an isometric change of coordinates it is not restrictive to assume that $k = d$ and $\operatorname{aff}(\operatorname{supp}\gamma) = \mathbb{R}^d$.

Let us now introduce the set

$$D := \Big\{x \in \mathbb{R}^d :\ \liminf_{r\downarrow 0} \frac{\gamma(B_r(x))}{r^d} > 0\Big\}. \tag{9.4.24}$$

Since (9.4.20) yields

$$\frac{\gamma(B_r(x_t))}{r^k} \geq \left(\frac{\gamma(B_r(x_0))}{r^k}\right)^{1-t} \left(\frac{\gamma(B_r(x_1))}{r^k}\right)^{t} \quad t \in (0,1), \tag{9.4.25}$$

it is immediate to check that D is a convex subset of $\mathbb{R}^d$ with $D \subset \operatorname{supp}\gamma$.

General results on derivation of Radon measures in $\mathbb{R}^d$ (see for instance Theorem 2.56 in [11]) show that

$$\limsup_{r\downarrow 0} \frac{\gamma(B_r(x))}{r^d} < +\infty \quad \text{for } \mathscr{L}^d\text{-a.e. } x \in \mathbb{R}^d \tag{9.4.26}$$

and

$$\limsup_{r\downarrow 0} \frac{r^d}{\gamma(B_r(x))} < +\infty \quad \text{for } \gamma\text{-a.e. } x \in \mathbb{R}^d. \tag{9.4.27}$$

Using (9.4.27) we see that actually γ is concentrated on D (so that $\operatorname{supp}\gamma \subset \overline{D}$) and therefore, being d the dimension of $\operatorname{aff}(\operatorname{supp}\gamma)$, it follows that d is also the dimension of $\operatorname{aff}(D)$.

If a point $\bar{x} \in \mathbb{R}^d$ exists such that

$$\limsup_{r\downarrow 0} \frac{\gamma(B_r(\bar{x}))}{r^d} = +\infty,$$

then (9.4.25) forces every point of $\operatorname{int}(D)$ to verify the same property, but this would be in contradiction with (9.4.26), since we know that $\operatorname{int}(D)$ has strictly positive $\mathscr{L}^d$-measure. Therefore

$$\limsup_{r\downarrow 0} \frac{\gamma(B_r(x))}{r^d} < +\infty \quad \text{for all } x \in \mathbb{R}^d \tag{9.4.28}$$

and we obtain that $\gamma \ll \mathscr{L}^d$, again by the theory of derivation of Radon measures in $\mathbb{R}^d$. In the sequel we denote by ρ the density of γ w.r.t. $\mathscr{L}^d$ and notice that by Lebesgue differentiation theorem $\rho > 0$ $\mathscr{L}^d$-a.e. in D and $\rho = 0$ $\mathscr{L}^d$-a.e. in $\mathbb{R}^d \setminus D$.

By (9.4.20) the maps

$$V_r(x) = -\log\left(\frac{\gamma(B_r(x))}{\omega_d r^d}\right)$$

are convex on $\mathbb{R}^d$, and (9.4.28) gives that the family $V_r(x)$ is bounded as $r \downarrow 0$ for any $x \in D$. Using the pointwise boundedness of V_r on D and the convexity of V_r it is easy to show that V_r are locally equi-bounded (hence locally equi-continuous) on $\operatorname{int}(D)$ as $r \downarrow 0$. Let W be a limit point of V_r, with respect to the local uniform convergence, as $r \downarrow 0$: W is convex on $\operatorname{int}(D)$ and Lebegue differentiation theorem shows that

$$\exists \lim_{r\downarrow 0} V_r(x) = -\log\rho(x) = W(x) \quad \text{for } \mathscr{L}^d\text{-a.e. } \ x \in \operatorname{int}(D), \tag{9.4.29}$$

so that $\gamma = \rho\mathscr{L}^d = e^{-W}\chi_{\mathrm{int}(D)}\mathscr{L}^d$. In order to get a globally defined convex and l.s.c function V we extend W with the $+\infty$ value out of $\mathrm{int}(D)$ and define V to be its convex and l.s.c. envelope. It turns out that V coincides with W on $\mathrm{int}(D)$, so that still the representation $\gamma = e^{-V}\mathscr{L}^d$ holds.

Conversely, let us suppose that γ admits the representation (9.4.23) for a given convex l.s.c. function V and let μ^1, $\mu^2 \in \mathscr{P}_p(X)$; if their relative entropies are finite then they are absolutely continuous w.r.t. γ and therefore their supports are contained in $\mathrm{aff}(\mathrm{supp}\,\gamma)$. It follows that the support of any optimal plan $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$ in $\mathscr{P}_p(X)$ is contained in $\mathrm{aff}(\mathrm{supp}\,\gamma) \times \mathrm{aff}(\mathrm{supp}\,\gamma)$: up to a linear isometric change of coordinates, it is not restrictive to suppose $\mathrm{aff}(\mathrm{supp}\,\gamma) = \mathbb{R}^d$, μ^1, $\mu^2 \in \mathscr{P}_p(\mathbb{R}^d)$, $\gamma = e^{-V} \cdot \mathscr{L}^d \in \mathscr{P}(\mathbb{R}^d)$.

In this case we introduce the density ρ^i of μ^i w.r.t. $\mathscr{L}^d$ observing that

$$\frac{d\mu^i}{d\gamma} = \rho^i e^V \quad i = 1, 2,$$

where we adopted the convention $0 \cdot (+\infty) = 0$ (recall that $\rho^i(x) = 0$ for $\mathscr{L}^d$-a.e. $x \in \mathbb{R}^d \setminus D(V)$). Therefore the entropy functional can be written as

$$\mathcal{H}(\mu^i|\gamma) = \int_{\mathbb{R}^d} \rho^i(x) \log \rho^i(x)\, dx + \int_{\mathbb{R}^d} V(x)\, d\mu^i(x), \tag{9.4.30}$$

i.e. the sum of two geodesically convex functionals, as we proved discussing Examples 9.3.1 and Examples 9.3.6. Lemma 9.4.7 yields the log-concavity of γ; the case of generalized geodesics in $\mathscr{P}_2(X)$ is completely analogous. □

The previous theorem shows that in finite dimensions log-concavity of γ is equivalent to the convexity of $\mathcal{H}(\mu|\gamma)$ along (even generalized, if $p = 2$) geodesics of anyone of the Wasserstein spaces $\mathscr{P}_p(X)$: the link between these two concepts is provided by the representation formula (9.4.23).

When X is an infinite dimensional Hilbert space, (9.4.23) is no more true in general, but the equivalence between log-concavity and geodesic convexity of the relative entropy still holds. In particular all Gaussian measures, defined in Definition 6.2.1, induce a geodesically convex relative entropy functional (see condition (5) in the statement below).

Theorem 9.4.11. *Let X be a separable Hilbert space and let $\gamma \in \mathscr{P}(X)$. The following properties are equivalent:*

1. *$\mathcal{H}(\cdot|\gamma)$ is geodesically convex in $\mathscr{P}_p(X)$ for every $p \in (1, +\infty)$.*
2. *$\mathcal{H}(\cdot|\gamma)$ is convex along generalized geodesics in $\mathscr{P}_2(X)$.*
3. *For every couple of measures μ^1, $\mu^2 \in \mathscr{P}(X)$ with bounded support there exists a connecting plan $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ along with $\mathcal{H}(\cdot|\gamma)$ is displacement convex.*
4. *γ is log-concave.*

(5) *For every finite dimensional orthogonal projection $\pi : X \to X$, $\pi_\#\gamma$ is representable as in* (9.4.23) *for a suitable convex and l.s.c. function V.*

Proof. The implications (1) ⇒ (3) and (2) ⇒ (3) are trivial, and (3) ⇒ (4) follows by Lemma 9.4.7.
Now we show that (4) ⇒ (5), using Theorem 9.4.10: if A, B are (relatively) open subsets of $\pi(X)$ and $t \in [0,1]$ we should prove that

$$\log\Big(\pi_\#\gamma\big((1-t)A+tB\big)\Big) \geq (1-t)\log\Big(\pi_\#\gamma\big(A\big)\Big)+t\log\Big(\pi_\#\gamma\big(B\big)\Big). \quad (9.4.31)$$

By definition $\pi_\#\gamma\big(A\big) = \gamma\big(\pi^{-1}A\big)$, $\pi_\#\gamma\big(B\big) = \gamma\big(\pi^{-1}B\big)$, and it is immediate to check that

$$\pi_\#\gamma\big((1-t)A+tB\big) = \gamma\big((1-t)\pi^{-1}A+t\pi^{-1}B\big)$$

since $\pi^{-1}\big((1-t)A+tB\big) = (1-t)\pi^{-1}A+t\pi^{-1}B$. Thus (9.4.31) follows by the log-concavity of γ applied to the open sets $\pi^{-1}A$, $\pi^{-1}B$.
(5) ⇒ (1): we choose a sequence π^h of finite dimensional orthogonal projections on X such that $\pi^h(x) \to x$ for any $x \in X$ as $h \to \infty$, set $\gamma^h := \pi^h_\#\gamma$ and

$$\phi^h(\mu) := \mathcal{H}(\mu|\gamma^h), \quad \phi(\mu) := \mathcal{H}(\mu|\gamma) \quad \forall\, \mu \in \mathscr{P}(X).$$

Since each functional ϕ^h is geodesically convex in $\mathscr{P}_p(X)$, by Theorem 9.4.10, the thesis follows by Lemma 9.1.4 if we show that ϕ is the Γ-limit of ϕ^h as $h \to \infty$: thus we have to check conditions (9.1.4) and (9.1.5).

(9.1.4) follows immediately by Lemma 9.4.3; in order to check (9.1.5) we simply choose $\mu^h := \pi^h_\#\mu$ and we apply Corollary 9.4.6.
The implications (5) ⇒ (2) follows by the same approximation argument, invoking Lemma 9.2.9. □

If γ is log-concave and F satisfies (9.3.19), then all the integral functionals $\mathcal{F}(\cdot|\gamma)$ introduced in (9.4.7) are geodesically convex in $\mathscr{P}_p(X)$ and convex along generalized geodesics in $\mathscr{P}_2(X)$.

Theorem 9.4.12 (Geodesical convexity for relative integral functionals). *Suppose that γ is* log*-concave and $F : [0,+\infty) \to [0,+\infty]$ satisfies conditions* (9.4.6) *and* (9.3.19). *Then the integral functional $\mathcal{F}(\cdot|\gamma)$ is geodesically convex in $\mathscr{P}_p(X)$ and convex along generalized geodesics in $\mathscr{P}_2(X)$.*

Proof. The same approximation argument of the proof of the previous theorem shows that it is sufficient to consider the final dimensional case $X := \mathbb{R}^d$. Arguing as in the final part of the proof of Theorem 9.4.10 we can assume that $\gamma := e^{-V}\mathscr{L}^d$ for a convex l.s.c. function $V : \mathbb{R}^d \to (-\infty,+\infty]$ whose domain has not empty interior. For every couple of measure $\mu^1, \mu^2 \in D(\mathcal{F}(\cdot|\gamma))$ we have

$$\mu^i = \rho^i e^V \cdot \gamma, \quad \mathcal{F}(\mu^i|\gamma) = \int_{\mathbb{R}^d} F(\rho^i(x)e^{V(x)})e^{-V(x)}\,dx \quad i = 1,2. \quad (9.4.32)$$

As in Proposition 9.3.9, we denote by $\boldsymbol{r}$ the optimal transport map for the p-Wasserstein distance pushing μ^1 to μ^2 and we set $\boldsymbol{r}^t := (1-t)\boldsymbol{i}+t\boldsymbol{r}$, $\mu_t := (\boldsymbol{r}^t)_\#\mu^1$; arguing as in that proposition, we get

$$\mathcal{F}(\mu_t|\gamma) = \int_{\mathbb{R}^d} F\Big(\frac{\rho(x)e^{V(\boldsymbol{r}_t(x))}}{\det\tilde{\nabla}\boldsymbol{r}^t(x)}\Big)\det\tilde{\nabla}\boldsymbol{r}^t(x)e^{-V(\boldsymbol{r}_t(x))}\,dx, \tag{9.4.33}$$

and the integrand above may be seen as the composition of the convex and non-increasing map $s\mapsto F(\rho(x)e^{-s})e^s$ with the concave curve

$$t\mapsto -V(\boldsymbol{r}_t(x))+\log(\det\tilde{\nabla}\boldsymbol{r}_t(x)),$$

since $D(x) := \tilde{\nabla}\boldsymbol{r}(x)$ is a diagonalizable map with nonnegative eigenvalues and

$$t\mapsto \log\det\big((1-t)I+tD(x)\big)\quad\text{is concave in }[0,1].$$

The case of convexity along generalized geodesics in $\mathscr{P}_2(\mathbb{R}^d)$ follows by the same argument, recalling the final part of the proof of Proposition 9.3.9 once again.
□