Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 20–35

THE DYNAMICAL SCHRÖDINGER PROBLEM IN ABSTRACT METRIC SPACES

LÉONARD MONSAINGEON, LUCA TAMANINI AND DMITRY VOROTNIKOV

Abstract: In this paper we introduce the dynamical Schrödinger problem on abstract metric spaces, defined for a wide class of entropy and Fisher information functionals. Under very mild assumptions we prove a generic Gamma-convergence result towards the geodesic problem as the noise parameter $\varepsilon \downarrow 0$. We also investigate the connection with geodesic convexity of the driving entropy, and study the dependence of the entropic cost on the parameter ε. Some examples and applications are discussed.

Keywords: Benamou-Brenier formulation, Schrödinger bridge, Fisher information, Gamma-convergence, geodesic convexity, gradient flows, optimal transport.
Math. Subject Classification (2010): 49J45, 49Q20, 58B20.

1. Introduction

Gaspard Monge and Erwin Schrödinger came up with two a priori unrelated problems that are concerned with finding a preeminent way of deforming a prescribed probability distribution into another one. While Monge was interested in optimizing the cost of transportation of goods [62, 63, 58], Schrödinger's original thought experiment [59, 60] aimed for finding the most likely evolution between two subsequent observations of a cloud of independent particles. So, even if in both cases we are facing an interpolation and optimization problem, the former is deterministic in nature whereas the latter is strongly related to large deviations theory, and is, at the first glance, purely stochastic. We refer to a recent survey [22] for various formulations and aspects of the Schrödinger problem, and to [64, 65] for a discussion of its role in Euclidean Quantum Mechanics.

Received 23 December 2020.
LM wishes to thank Jean-Claude Zambrini for numerous and fruitful discussions on the Schrödinger problem, and acknowledges support from the Portuguese Science Foundation through FCT project PTDC/MAT-STA/22812/2017 *SchröMoka*. LT acknowledges financial support from FSMP Fondation Sciences Mathématiques de Paris. DV was partially supported by CMUC – UID/MAT/00324/2020, funded by the Portuguese Government through FCT/MCTES and co-funded by the European Regional Development Fund through the Partnership Agreement PT2020, and by the FCT project PTDC/MAT-PUR/28686/2017.

Anyway, several analogies and connections exist between the two problems. They can be appreciated by looking carefully at the interpolation aspects of both problems, namely at their dynamical formulations and the underlying equations governing the respective evolutions. In the case of a quadratic transportation cost over a Riemannian manifold M, the Monge-Kantorovich optimal transport problem is solved (at least in a weak sense) by interpolating between the source and the target distributions with a constant-speed, length-minimizing geodesic in the Otto-Wasserstein space of probability measures $\mathcal{P}_2(M)$. This gives a curve $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}_2(M)$ which formally satisfies (in the sense of the celebrated Otto calculus [55, 62, 63])

$$\nabla_{\dot\mu_t}\dot\mu_t = 0, \tag{1.1}$$

where $\nabla_{\dot\mu_t}$ is the covariant derivative along the curve $t \mapsto \mu_t$. The Schrödinger problem with parameter ε (from a physical viewpoint, ε can be seen as a temperature or level of noise) can also be translated into such a geometric language. By analogy, when looking at the covariant derivative along the optimal evolution $(\mu_t^\varepsilon)_{t\in[0,1]}$, usually called Schrödinger bridge or entropic interpolation, the resulting equation is surprising and can be viewed [23] as Newton's second law

$$\nabla_{\dot\mu_t^\varepsilon}\dot\mu_t^\varepsilon = \frac{\varepsilon^2}{8}\nabla I(\mu_t^\varepsilon), \tag{1.2}$$

where in the right-hand side ∇ denotes the gradient in the Otto-Wasserstein pseudo-Riemannian sense and I is the Fisher information

$$I(\mu) = 4\int_M |\nabla\sqrt{\rho}|^2 \,\mathrm{dvol} = \int_M |\nabla \log\rho|^2 \rho \,\mathrm{dvol}$$

provided $\mu = \rho \cdot \mathrm{vol}$. A related observation is that the (scaled) heat flow, coinciding with the (scaled) gradient flow of the Boltzmann-Shannon entropy [40, 63]

$$H(\mu) = \int_M \rho\log\rho \,\mathrm{dvol}$$

for $\mu = \rho\cdot\mathrm{vol}$, is also a solution to (1.2): a simple differentiation in time of $\dot\mu_t = -\frac{\varepsilon}{2}\nabla H(\mu_t)$ and the fact that $I = |\nabla H|^2$ in the Otto-Wasserstein sense automatically yield

$$\nabla_{\dot\mu_t}\dot\mu_t = \frac{\varepsilon}{2}\nabla^2 H(\mu_t)\cdot\frac{\varepsilon}{2}\nabla H(\mu_t) = \frac{\varepsilon^2}{8}\nabla|\nabla H(\mu_t)|^2 = \frac{\varepsilon^2}{8}\nabla I(\mu_t).$$

This shows that the Schrödinger problem lies between optimal transport and diffusion and is naturally intertwined with both deterministic behaviour and

Brownian motion. It shares the same Newton's law as the gradient flow of the entropy, but unlike the heat flow it has a prescribed final configuration to match: it is up to the parameter ε to tip the balance in favour of deterministic transport or diffusion. With this heuristics in mind, we see that as $\varepsilon \to 0$ the applied force $\varepsilon^2 \nabla I(\mu_t^\varepsilon)$ in (1.2) vanishes, so that the Schrödinger problem may be interpreted as a noisy (entropic) counterpart of the Monge-Kantorovich optimal transport, corresponding to the unforced geodesic evolution (1.1) discussed above. This informal relationship has a rigorous counterpart, which dates back to the pioneering works on the asymptotic behavior of the Schrödinger problem as $\varepsilon \to 0$ of T. Mikami, M. Thieullen [50, 51], and C. Léonard [44, 45]. This was subsequently developed in [19, 8, 38]. Very recently [47, 30], similar small-noise results were obtained for static Monge-Kantorovich problems regularized with more general entropies.

This first connection can be investigated further and by doing so one can remark that (1.2) is exactly the Euler-Lagrange optimality equation for the dynamical Benamou-Brenier formulation of the Schrödinger problem [18, 22, 45, 39], which consists in minimizing the Lagrangian kinetic action perturbed by the Fisher information: In more precise terms,

$$\inf \left\{ \frac{1}{2} \int_0^1 \int_M |v_t|^2 \, \mathrm{d}\mu_t \mathrm{d}t + \frac{\varepsilon^2}{8} \int_0^1 I(\mu_t) \, \mathrm{d}t \right\}, \tag{1.3}$$

where the infimum runs over all solutions of the continuity equation

$$\partial \mu_t + \mathrm{div}(v_t \mu_t) = 0$$

with prescribed initial and final densities. Also from this variational standpoint the reader can see that as $\varepsilon \to 0$ the Schrödinger problem formally reduces to

$$\inf \frac{1}{2} \int_0^1 \int_M |v_t|^2 \, \mathrm{d}\mu_t \mathrm{d}t, \tag{1.4}$$

namely to the dynamical Benamou-Brenier formulation of the (quadratic) optimal transport problem [9]. This variational representation depicts in a way clearer than (1.2) the double nature of the Schrödinger problem, the competition between the determinism encoded in the kinetic energy and the unpredictability coming from the Fisher information, and the role played by ε in balancing these two opposite behaviours.

The double bond of the Schrödinger problem with optimal transport on the one hand and heat flow on the other hand results in fruitful and wide-ranging applications of both theoretical and applied interest. Indeed, from the connection with the heat flow the solutions to the Schrödinger problem gain regularity properties which are not available in optimal transport, and thanks to the asymptotic behaviour of the Schrödinger problem as $\varepsilon \to 0$ entropic interpolations represent an efficient way to approximate Wasserstein geodesics with second-order accuracy [38, 25]. This approach has already turned out to be successful in conjunction with functional inequalities [24, 34] and differential calculus along Wasserstein geodesics [38]. But the nice behaviour of Schrödinger bridges is important also for computational purposes. The impact of Schrödinger problem and Sinkhorn algorithms (deeply related to the static formulation of the former) on the numerical methods used in optimal transportation theory has been impressive, as witnessed by several recent works (see [57] and references therein as well as [26, 10, 12, 13, 11, 29]).

As a matter of fact neither the particular structure of the Wasserstein space nor the specific choice of the Boltzmann-Shannon functional are required to define the two problems in question (cf. a related discussion in the heuristic paper [43]): one can of course define length-minimizing geodesics in any metric space (X, d), and the Schrödinger problem (or at least its Benamou-Brenier formulation described above) merely involves an entropy functional and a corresponding Fisher information. Given such a reasonable entropy functional E on X that generates a gradient flow in a suitable sense, the corresponding Fisher information is expected to be nothing but the dissipation rate of E (along solutions of its own gradient flow), just as I coincides with the rate of dissipation of the entropy H along the heat flow. This observation is the starting point of the present paper, where we intend to study the abstract Schrödinger bridge problem or, in other words, the entropic approximation of geodesics in metric spaces.

Under very mild assumptions on X and E, we will prove the solvability of the abstract ε-Schrödinger problem and the Γ-convergence to the corresponding geodesic problem as $\varepsilon \to 0$. We will also rigorously justify, in the metric setting, that any trajectory of a gradient flow solves an associated Schrödinger problem. Leveraging a quantitative AC^2 estimate based on a straightforward chain-rule in the smooth Riemannian setting, the cornerstone of our analysis will be the systematic construction of an ε-regularized entropic copy $(\gamma_t^\varepsilon)_{t\in[0,1]}$ of any arbitrary curve $(\gamma_t)_{t\in[0,1]}$. These perturbed curves will provide recovery sequences for the Γ-convergence. Our construction is completely Eulerian and essentially consists in running the E-gradient flow for a short time $h_\varepsilon(t)$ starting at γ_t for all t, for well-chosen functions $h_\varepsilon \geq 0$. The challenge here will be to reproduce the (formal, differential) Riemannian chain-rule in metric spaces. Notably, an analogous pseudo-Riemannian idea has recently been used by A. Baradat and some of the authors [8, 53] in order to prove the Γ-convergence for the classical dynamical Schrödinger problem on the Otto-Wasserstein space and for its counterpart on the non-commutative Fisher-Rao space, respectively. However, in those papers the computations were *ad hoc* and heavily exploited the underlying structures of the particular spaces as well as the properties of the particular flows (namely, of the classical heat flow and of its restriction to multivariate Gaussians), whereas here we derive everything from the existence of an abstract gradient flow on X driven by E.

Another notion of paramount importance herein will be convexity. In the smooth Riemannian setting, and given $\lambda \in \mathbb{R}$, elementary calculus shows that the λ-convexity of E along geodesics is of course equivalent to a uniform lower bound $\mathrm{Hess}\,\mathsf{E}(x) \geq \lambda\,\mathrm{Id}$ as quadratic forms in the tangent space, but also more importantly to the λ-contractivity of the E-gradient flow. In the metric setting no second order calculus is available in general, and the very notion of gradient flow as well as its connection with geodesic convexity and contractivity become much more subtle. The key notion of gradient flow that we shall use throughout is that of *Evolution Variational Inequality*, or EVI_λ flow [2]. Under reasonable assumptions it is well known that (a suitable variant of) convexity of E generally provides existence of an EVI_λ-flow starting at any $x \in \mathrm{X}$, see [2]. A natural question to ask is whether the converse also holds true, i.e. whether well-posedness of a reasonable gradient flow implies some convexity. This was proved in [15] for the specific case of the Euclidean Wasserstein space $\mathrm{X} = W_2(\Omega)$, $\Omega \subset \mathbb{R}^d$, and at least for the so-called internal energies, and it is shown therein that 0-contractivity of the gradient flow (or equivalently, of the associated nonlinear diffusion equation) implies 0-displacement convexity in the sense of McCann [49]. In the same spirit, and building up on Otto and Westdickenberg [56], Daneri and Savaré proved in a very general metric setting that the generation of an EVI_λ-flow indeed implies λ-geodesic convexity [27, Theorem 3.2]. A byproduct of our analysis for the Γ-convergence will give a new independent proof of this latter fact by a completely different approach, essentially by constructing an ε-entropic regularization of geodesics and carefully examining the defect of optimality at order one in $\varepsilon \to 0$.

As a main application of the Γ-convergence of the Schrödinger problem to the geodesic problem as $\varepsilon \to 0$ (and more generally of the ε'-Schrödinger problem to the ε-one as $\varepsilon' \to \varepsilon$) we investigate the behaviour of the optimal value of the dynamical Schrödinger problem, henceforth called *entropic cost*, as a function of the temperature parameter ε, with particular emphasis on the regularity and the small-noise regime. For the classical dynamical Schrödinger problem (1.3), it has recently been proved by the second author with G. Conforti [25] that the entropic cost is of class $C^1((0,\infty)) \cap C([0,\infty))$ (actually $C^1([0,\infty))$ under suitable assumptions) and twice a.e. differentiable; once this regularity information is available, the formula for the first derivative is rather easy to guess, as by the envelope theorem it coincides with the partial derivative w.r.t. ε of the functional in (1.3) evaluated at any critical point. Denoting by $\mathscr{C}_\varepsilon(\mu,\nu)$ the value in (1.3) with marginal constraints μ and ν and by $(\mu_t^\varepsilon)_{t\in[0,1]}$ the associated Schrödinger bridge, this statement reads as

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathscr{C}_\varepsilon(\mu,\nu) = \frac{\varepsilon}{4}\int_0^1 I(\mu_t^\varepsilon)\,\mathrm{d}t, \qquad \forall \varepsilon > 0$$

and in [25] this identity played an important role in the study of both the large- and small-noise behaviour of the Schrödinger problem, obtaining in particular a Taylor expansion around $\varepsilon = 0$ with $o(\varepsilon^2)$-accuracy. Since the central object in the present paper is an abstract and general formulation of (1.3), an analogous result is expected to hold. However, from a technical viewpoint the proof is much more subtle and challenging, because unlike (1.3) our metric version of the dynamical Schrödinger problem may have multiple solutions. For this reason the discussion about the regularity of the entropic cost in this paper is less concise than in [25]. Nonetheless, we are still able to deduce the same kind of Taylor expansion with the same accuracy. Given the previous interpretation of the Schrödinger problem as a noisy Monge-Kantorovich problem and the importance of quantitative estimates in approximating optimal transport by means of the Schrödinger problem, it is reasonable to expect that such a Taylor expansion (valid in a general framework for a wide choice of functionals E) will fit to a countless variety of examples, some of which will be discussed here.

The paper is organized as follows: In Section 2 we give a short and formal proof of our fundamental AC^2 estimate in the smooth Riemannian setting, and show how it can be exploited to establish Γ-convergence and convexity. Section 3 fixes the metric framework in which we work for the rest of the paper, and extends the previous estimate to this metric setting. In Section 4 we prove the Γ-convergence as $\varepsilon \downarrow 0$ and establish the geodesic convexity. Section 5 studies the dependence of the optimal entropic cost on the temperature parameter $\varepsilon > 0$, and provides a second order expansion. Finally, we list in Section 6 several examples and applications covered by our abstract results.

2. Heuristics

Here we remain formal and the computations are carried in a Riemannian setting, where classical calculus and chain-rules are available. (Significant work will be required later on to adapt the computations in metric spaces.) All the objects and functions in this section are therefore considered to be smooth, and we deliberately ignore any regularity issue.

Let M be a Riemannian manifold with scalar product $\langle .,. \rangle_q$ at a point $q \in M$ and induced Riemannian distance d, and let $V : M \to \mathbb{R}$ be a given potential. For simplicity we assume here that V is globally bounded from below on M, and up to replacing V by $V - \min V$ we can assume that $V(q) \geq 0$. (In section 3 we will relax this assumption and allow V to be only *locally bounded*.) Given a small temperature parameter $\varepsilon > 0$, and following [43], the (dynamic) *geometric Schrödinger problem* consists in solving the optimization problem

$$\frac{1}{2}\int_0^1 \left|\frac{\mathrm{d}q_t}{\mathrm{d}t}\right|^2 \mathrm{d}t + \frac{\varepsilon^2}{2}\int_0^1 |\nabla V|^2(q_t)\,\mathrm{d}t \qquad \longrightarrow \qquad \min;$$
$$\text{s.t. } q \in C([0,1], M) \text{ with endpoints } q_0, q_1. \tag{2.1}$$

For $s \geq 0$ we denote by $\Phi(s, q_0)$ the semi-flow corresponding to the autonomous V-gradient flow started from $q_0 \in M$,

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}s}\Phi(s,q_0) = -\nabla V(\Phi(s,q_0)), \\ \Phi(0,q_0) = q_0. \end{cases}$$

The goal of this section is to give a straightforward proof of the following two facts, assuming that the potential V is well behaved:

(i) the ε-Schrödinger problem converges to the geodesic problem as $\varepsilon \to 0$;
(ii) λ-contractivity of the generated flow Φ can be turned into λ-convexity along geodesics.

With this goal in mind, fix any two endpoints $q_0, q_1 \in M$ and take an arbitrary curve joining them

$$q \in C([0,1], M), \qquad q|_{t=0} = q_0 \quad \text{and} \quad q|_{t=1} = q_1.$$

For any function $h(t) \geq 0$ with $h(0) = h(1) = 0$, we perturb q by defining

$$\tilde{q}_t := \Phi(h(t), q_t), \qquad t \in [0,1]$$

i.e. $\tilde{q}_t$ is the solution of the V-gradient flow at time $s = h(t) \geq 0$ starting from q_t at time $s = 0$. We shall refer to $t \in [0,1]$ as a "horizontal time" and to $s \in [0, h(t)]$ as a "vertical time", see Figure 1. Later on we will think of the curve $\tilde{q}$ as a "regularized" version of q.

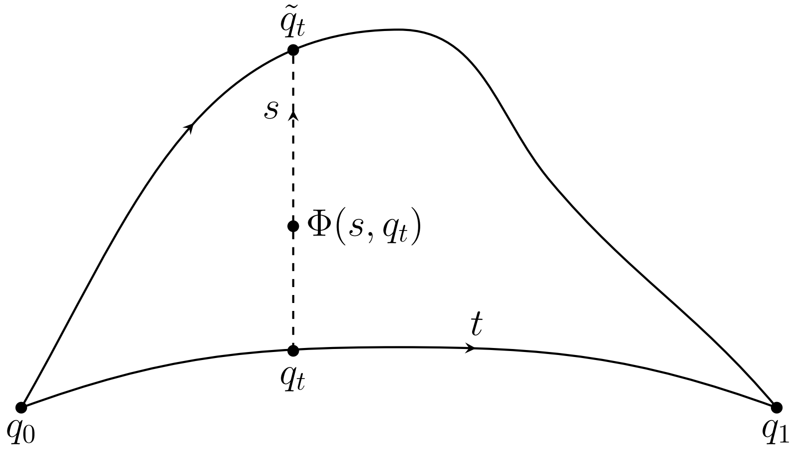

FIGURE 1. The perturbed curve

Note that the endpoints remain invariant, $\tilde{q}_0 = q_0$ and $\tilde{q}_1 = q_1$. Since by definition of the flow $\partial_s \Phi(s, q_t) = -\nabla V(\Phi(s, q_t))$, the speed of the perturbed curve can be computed as

$$\begin{aligned}\frac{\mathrm{d}\tilde{q}_t}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\Big(\Phi(h(t), q_t)\Big) = \partial_s \Phi(h(t), q_t) h'(t) + \partial_q \Phi(h(t), q_t) \frac{\mathrm{d}q_t}{\mathrm{d}t} \\ &= -h'(t) \nabla V(\tilde{q}_t) + \partial_q \Phi(h(t), q_t) \frac{\mathrm{d}q_t}{\mathrm{d}t}.\end{aligned}$$

Bringing the $h'(t)$ term to the left-hand side and taking the half squared norm (in the tangent space $T_{\tilde{q}_t} M$) gives

$$\frac{1}{2}\left|\frac{\mathrm{d}\tilde{q}_t}{\mathrm{d}t}\right|^2 + \frac{1}{2}|h'(t)|^2 |\nabla V(\tilde{q}_t)|^2 + h'(t) \underbrace{\langle \nabla V(\tilde{q}_t), \frac{\mathrm{d}\tilde{q}_t}{\mathrm{d}t}\rangle_{\tilde{q}_t}}_{=\frac{\mathrm{d}}{\mathrm{d}t} V(\tilde{q}_t)} = \frac{1}{2}\left|\partial_q \Phi(h(t), q_t) \frac{\mathrm{d}q_t}{\mathrm{d}t}\right|^2. \tag{2.2}$$

Assume now that, for whatever reason, the gradient flow satisfies the following quantified contractivity estimate w.r.t. the Riemannian distance d

$$d(\Phi(s,p_0),\Phi(s,p_0')) \le e^{-\lambda s} d(p_0,p_0'), \qquad \forall s \ge 0,\ p_0,p_0' \in M \tag{2.3}$$

for some fixed $\lambda \in \mathbb{R}$. Then it is easy to check that the linear map $v \mapsto \partial_q \Phi(s,p) \cdot v$ (from T_pM to $T_{\Phi(s,p)}M$) has norm less than $e^{-\lambda s}$, and therefore (2.2) gives

$$\frac{1}{2}\left|\frac{\mathrm{d}\tilde q_t}{\mathrm{d}t}\right|^2 + \frac{1}{2}|h'(t)|^2|\nabla V(\tilde q_t)|^2 + h'(t)\frac{\mathrm{d}}{\mathrm{d}t}V(\tilde q_t) \le \frac{1}{2}e^{-2\lambda h(t)}\left|\frac{\mathrm{d}q_t}{\mathrm{d}t}\right|^2. \tag{2.4}$$

Integration by parts yields next

$$\begin{aligned}\frac{1}{2}\int_0^1 \left|\frac{\mathrm{d}\tilde q_t}{\mathrm{d}t}\right|^2 \mathrm{d}t + \frac{1}{2}\int_0^1 |h'(t)|^2|\nabla V(\tilde q_t)|^2\,\mathrm{d}t - \int_0^1 h''(t)V(\tilde q_t)\,\mathrm{d}t \\ \le \frac{1}{2}\int_0^1 e^{-2\lambda h(t)}\left|\frac{\mathrm{d}q_t}{\mathrm{d}t}\right|^2 \mathrm{d}t + \Big(h'(0)V(q_0) - h'(1)V(q_1)\Big),\end{aligned} \tag{2.5}$$

where the invariance $\tilde q_0 = q_0$, $\tilde q_1 = q_1$ was used in the last boundary terms. This fundamental estimate gives a quantified bound on the kinetic energy (namely the L^2 speed) of $\tilde q$ in terms of that of the original curve q, and will be the cornerstone of the whole analysis.

Both the convexity and the convergence of the Schrödinger problem will actually follow by setting $h(t) = \varepsilon H(t)$ for suitable choices of $H(t) \ge 0$, and then letting $\varepsilon \downarrow 0$. Note that in this case we have $h(t) = \varepsilon H(t) \downarrow 0$ uniformly, hence the perturbed curve

$$q_t^\varepsilon := \Phi(\varepsilon H(t), q_t) \tag{2.6}$$

will converge uniformly to q as $\varepsilon \downarrow 0$ too.

2.1. Convergence of the Schrödinger problem. A first use of (2.5) will be crucial in proving the Γ-convergence of the Schrödinger functional

$$\mathcal{A}_\varepsilon(q) := \frac{1}{2}\int_0^1 \left|\frac{\mathrm{d}q_t}{\mathrm{d}t}\right|^2 \mathrm{d}t + \frac{\varepsilon^2}{2}\int_0^1 |\nabla V(q_t)|^2\,\mathrm{d}t$$

towards the kinetic action

$$\mathcal{A}(q) := \frac{1}{2}\int_0^1 \left|\frac{\mathrm{d}q_t}{\mathrm{d}t}\right|^2 \mathrm{d}t$$

as $\varepsilon \downarrow 0$.

Theorem 2.1 (formal Γ-limit). *For any $q_0, q_1 \in M$ it holds*

$$\mathcal{A} = \Gamma - \lim_{\varepsilon \to 0} \mathcal{A}_\varepsilon$$

for the uniform convergence on the space of curves with fixed endpoints q_0, q_1.

Proof: We check separately the $\Gamma - \liminf$ and the $\Gamma - \limsup$ properties. As for the former, given any curve q joining q_0, q_1 and any $q^\varepsilon \to q$ uniformly, since the kinetic energy functional $q \mapsto \mathcal{A}(q)$ is always lower semicontinuous for the uniform convergence we get first

$$\mathcal{A}(q) \le \liminf_{\varepsilon \downarrow 0} \mathcal{A}(q^\varepsilon) \le \liminf_{\varepsilon \downarrow 0} \mathcal{A}_\varepsilon(q^\varepsilon).$$

For the $\Gamma - \limsup$, let $H(t) = \min\{t, 1-t\}$ be the hat function centered at $t = 1/2$ with height $1/2$ and vanishing at the boundaries, set $h(t) = \varepsilon H(t)$, and let q^ε be the regularized curve constructed in (2.6). In this simple smooth setting it is not difficult to check that $q^\varepsilon \to q$ uniformly. Moreover, our choice of $h(t)$ results in $|h'(t)|^2 = \varepsilon^2$ with $h'(0) = \varepsilon$, $h'(1) = -\varepsilon$, and $h''(t) = -2\varepsilon\delta_{1/2}(t)$ in the distributional sense. Therefore (2.5) gives immediately

$$\begin{aligned}\mathcal{A}_\varepsilon(q^\varepsilon) + 2\varepsilon V(q^\varepsilon_{1/2}) &= \frac{1}{2}\int_0^1 \left|\frac{\mathrm{d}q^\varepsilon_t}{\mathrm{d}t}\right|^2 \mathrm{d}t + \frac{\varepsilon^2}{2}\int_0^1 |\nabla V(q^\varepsilon_t)|^2 \,\mathrm{d}t + 2\varepsilon V(q^\varepsilon_{1/2}) \\ &\le \frac{1}{2}\int_0^1 e^{-2\varepsilon\lambda H(t)} \left|\frac{\mathrm{d}q_t}{\mathrm{d}t}\right|^2 \mathrm{d}t + \varepsilon\Big(V(q_0) + V(q_1)\Big).\end{aligned}$$

The singularity of h'' at $t = 1/2$ can be easily and rigorously worked around, simply integrating by parts (2.2) separately on each interval $t \in [0, 1/2]$ and $t \in [1/2, 1]$ and keeping track of the boundary terms resulting ultimately in the above $2\varepsilon V(q^\varepsilon_{1/2}) \ge 0$. Discarding this latter non-negative term finally gives

$$\begin{aligned}&\limsup_{\varepsilon \downarrow 0} \mathcal{A}_\varepsilon(q^\varepsilon) \\ &\qquad \le \limsup_{\varepsilon \downarrow 0} \left\{ \frac{1}{2}\int_0^1 e^{-2\varepsilon\lambda H(t)} \left|\frac{\mathrm{d}q_t}{\mathrm{d}t}\right|^2 \mathrm{d}t + \varepsilon\Big(V(q_0) + V(q_1)\Big)\right\} = \mathcal{A}(q)\end{aligned}$$

and concludes the proof. ■

2.2. Quantifying the convexity. The second consequence of our fundamental estimate (2.5) is the quantification of the convexity of the potential V in terms of the quantified contractivity (2.3). The point here is that the result can be obtained directly from (2.2), which can be established in a purely metric setting without relying on differential calculus (see the next section for details).

Theorem 2.2. *Assume that V satisfies 2.3. Then V is λ-geodesically convex, i.e.*

$$V(q_\theta) \leq (1-\theta)V(q_0) + \theta V(q_1) - \frac{\lambda}{2}\theta(1-\theta)d^2(q_0,q_1), \qquad \theta \in (0,1)$$

for any *geodesic* $(q_\theta)_{\theta\in[0,1]}$ *in* M.

Proof: Let $(q_t)_{t\in[0,1]}$ be an arbitrary geodesic with endpoints q_0, q_1. For fixed $\theta \in (0,1)$ let

$$H_\theta(t) := \begin{cases} \dfrac{1}{\theta}t & \text{if } t \in [0,\theta], \\ -\dfrac{1}{1-\theta}(t-1) & \text{if } t \in [\theta,1], \end{cases}$$

be the hat function centered at $t=\theta$ with height 1 and vanishing at $t = 0,1$, and for any $\varepsilon > 0$ let q^ε be the regularized curve constructed in (2.6) with $h(t) = \varepsilon H_\theta(t)$. Note moreover that

$$h'(0) = \frac{\varepsilon}{\theta}, \qquad h'(1) = -\frac{\varepsilon}{1-\theta}, \qquad h''(t) = -\varepsilon\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right)\delta_\theta(t)$$

in the distributional sense. Discarding the non-negative term $|h'(t)|^2|\nabla V(\tilde{q}_t)|^2$ in (2.5), the optimality of the geodesic q from q_0 to q_1 gives

$$\begin{aligned}
0 &\leq \frac{1}{2}\int_0^1 \left|\frac{\mathrm{d}q_t^\varepsilon}{\mathrm{d}t}\right|^2 \mathrm{d}t - \frac{1}{2}\int_0^1 \left|\frac{\mathrm{d}q_t}{\mathrm{d}t}\right|^2 \mathrm{d}t \\
&\overset{(2.5)}{\leq} \int_0^1 h''(t)V(q_t^\varepsilon)\,\mathrm{d}t + \frac{1}{2}\int_0^1 \left(e^{-2\lambda h(t)} - 1\right)\left|\frac{\mathrm{d}q_t}{\mathrm{d}t}\right|^2 \mathrm{d}t \\
&\quad + \Big(h'(0)V(q_0) - h'(1)V(q_1)\Big) \\
&= -\varepsilon\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right)V(q_\theta^\varepsilon) + \frac{d^2(q_0,q_1)}{2}\int_0^1 \left(e^{-2\lambda\varepsilon H_\theta(t)} - 1\right)\mathrm{d}t \\
&\qquad + \varepsilon\left(\frac{1}{\theta}V(q_0) + \frac{1}{1-\theta}V(q_1)\right),
\end{aligned}$$

where the last equality follows from the constant speed property $|\frac{\mathrm{d}q_t}{\mathrm{d}t}|^2 = d^2(q_0,q_1)$ of the geodesic $(q_t)_{t\in[0,1]}$ connecting q_0, q_1 as well as from the explicit properties of $h(t) = \varepsilon H_\theta(t)$ listed above. Multiplying by $\frac{\theta(1-\theta)}{\varepsilon} > 0$ and rearranging gives

$$V(q_\theta^\varepsilon) \le (1-\theta)V(q_0) + \theta V(q_1) + \theta(1-\theta)\frac{d^2(q_0,q_1)}{2}\underbrace{\int_0^1 \frac{e^{-2\lambda\varepsilon H_\theta(t)}-1}{\varepsilon}\mathrm{d}t}_{:=I_\varepsilon}.$$

Since $\int_0^1 H_\theta(t)\mathrm{d}t = \frac{1}{2}$ for all θ we see that $I_\varepsilon \to -2\lambda\int_0^1 H_\theta(t)\,\mathrm{d}t = -\lambda$ as $\varepsilon \downarrow 0$, and the result immediately follows since $V(q_\theta^\varepsilon) \to V(q_\theta)$ as well in the left-hand side. ■

3. In metric spaces

Before trying to adapt the previous computations to the metric context we need to fix once and for all the framework to be used in the sequel.

3.1. Preliminaries and setting.

- By $C([0,1],(\mathrm{X},\mathsf{d}))$, or simply $C([0,1],\mathrm{X})$, we denote the space of continuous curves with values in the metric space (X,d). The collection of absolutely continuous curves on $[0,1]$ is denoted by $AC([0,1],(\mathrm{X},\mathsf{d}))$, or simply by $AC([0,1],\mathrm{X})$. For any curve $(\gamma_t) \in AC([0,1],\mathrm{X})$, its length is well defined as
$$\ell(\gamma) := \int_0^1 |\dot\gamma_t|\,\mathrm{d}t,$$
where $|\dot\gamma_t|$ denotes the metric speed of γ. If $|\dot\gamma_t| \in L^2(0,1)$, then we shall say that $(\gamma_t) \in AC^2([0,1],\mathrm{X})$. For these notions of absolutely continuous curves and metric speed in a metric space, see for instance [2, Section 1.1].
- A curve $\gamma : [0,1] \to \mathrm{X}$ is called geodesic provided $\mathsf{d}(\gamma_t,\gamma_s) = |t-s|\mathsf{d}(\gamma_0,\gamma_1)$ for all $t,s \in [0,1]$.
- The slope $|\partial\mathsf{E}|$ of a functional $\mathsf{E} : \mathrm{X} \to \mathbb{R}\cup\{+\infty\}$ at a point $x \in \mathrm{X}$ is set as $+\infty$ if $x \notin D(\mathsf{E})$, 0 if x is isolated, and defined as
$$|\partial\mathsf{E}|(x) := \limsup_{y\to x}\frac{[\mathsf{E}(x)-\mathsf{E}(y)]^+}{\mathsf{d}(x,y)}$$
if $x \in D(\mathsf{E})$.
- A curve $(\gamma_t)_{t>0} \subset \mathrm{X}$ is said to be a gradient flow of E in the EVI$_\lambda$ sense (with $\lambda \in \mathbb{R}$) provided $(\gamma_t) \in AC_{loc}((0,\infty),\mathrm{X})$ and

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{d}^2(\gamma_t,y) + \frac{\lambda}{2}\mathsf{d}^2(\gamma_t,y) + \mathsf{E}(\gamma_t) \le \mathsf{E}(y), \qquad \forall y \in \mathrm{X},\ \text{a.e. } t > 0. \tag{EVI$_\lambda$}$$

If $\gamma_t \to x$ as $t \downarrow 0$ with $x \in \overline{D(\mathsf{E})}$, then we say that the gradient flow (γ_t) starts at x.

After this premise, let us fix the framework we shall work within.

Setting 3.1. *On the space* (X, d) *and on the functional* $\mathsf{E} : \mathrm{X} \to \mathbb{R} \cup \{+\infty\}$ *we make the following assumptions:*

(A1) (X, d) *is a complete and separable metric space;*
(A2) E *is lower semicontinuous with dense domain, i.e.* $\overline{D(\mathsf{E})} = \mathrm{X}$, *and locally bounded from below in the following sense: for any* d*-bounded set* $B \subset \mathrm{X}$ *there exists* $c_B \in \mathbb{R}$ *such that* $\mathsf{E}(x) \geq c_B$ *for all* $x \in B$*;*
(A3) there exists $\lambda \in \mathbb{R}$ *such that for any* $x \in \mathrm{X}$ *there exists an* EVI_λ*-gradient flow of* E *starting from* x. *In view of* (3.3), *the corresponding 1-parameter semigroup shall be denoted* S_t.

Sometimes, and always explicitly indicated, we will also use the following extra hypothesis.

Assumption 3.2. *There exists a Hausdorff topology* σ *on* X *such that* d*-bounded sets are sequentially* σ*-compact. Moreover, the distance* d *and the slope* $|\partial \mathsf{E}|$ *are* σ*-sequentially lower semicontinuous.*

Remark 3.3. Assumption 3.2 is in particular valid provided (X, d) is a locally compact space. Indeed, in this case the metric topology of (X, d) is an admissible candidate for σ, since bounded sets are relatively compact (by [17, Proposition 2.5.22]) and the lower semicontinuity of the slope $|\partial \mathsf{E}|$ w.r.t. the metric topology is a consequence of the forthcoming identity (3.1). ■

Remark 3.4. Assumption 3.2 implies that d-converging sequences are also σ-converging. Indeed, given $(x_n) \subset \mathrm{X}$ with $\mathsf{d}(x, x_n) \to 0$ as $n \to \infty$ for some limit $x \in \mathrm{X}$, by Assumption 3.2 and by the boundedness of $(x_n)_n$ there exist a subsequence $(x_{n_k})_k$ and $y \in \mathrm{X}$ such that

$$x_{n_k} \xrightarrow{\sigma} y \quad \text{as} \quad k \to \infty.$$

Since d is σ-sequentially lower semicontinuous (again by Assumption 3.2) we deduce that

$$\mathsf{d}(x, y) \leq \liminf_{k \to \infty} \mathsf{d}(x, x_{n_k}) = \lim_{n \to \infty} \mathsf{d}(x, x_n) = 0,$$

whence $x = y$. This classically implies that the whole sequence converges, $x_n \xrightarrow{\sigma} x$. ■

We list now some useful properties of EVI-gradient flows, which hold true in Setting 3.1 and that we shall use extensively in the sequel. First of all, whenever $x \in X$ is the starting point of an EVI_λ flow, the slope there (a *local* object, a priori) admits the *global* representation

$$|\partial \mathsf{E}|(x) = \sup_{y \neq x} \Big(\frac{\mathsf{E}(x) - \mathsf{E}(y)}{\mathsf{d}(x,y)} + \frac{\lambda}{2}\mathsf{d}(x,y) \Big)^+, \tag{3.1}$$

see [54, Proposition 3.6]. Since we assume that any $x \in X$ is the starting point of an EVI_λ-gradient flow, this means in particular that $|\partial \mathsf{E}| : X \to [0,\infty]$ is lower semicontinuous, since so is the right-hand side above (as a supremum of lower semicontinuous functions). This also implies by [2, Theorem 1.2.5] that $|\partial \mathsf{E}|$ is a *strong upper gradient* for E in the sense of [2, Definition 1.2.1], namely: for every $(\gamma_t) \in AC([0,1],\mathrm{X})$, the map $t \mapsto \mathsf{E}(\gamma_t)$ is Borel and

$$|\mathsf{E}(\gamma_{t_1}) - \mathsf{E}(\gamma_{t_0})| \leq \int_{t_0}^{t_1} |\partial \mathsf{E}|(\gamma_t)|\dot\gamma_t| \, \mathrm{d}t, \qquad \forall 0 \leq t_0 \leq t_1 \leq 1, \tag{3.2}$$

the right-hand side being possibly infinite. In addition, if (γ_t) is an EVI_λ-gradient flow of E then the following hold [54, Theorem 3.5]:

(i) If (γ_t) starts from $x \in \overline{D(\mathsf{E})}$ and $(\tilde\gamma_t)$ is a second EVI_λ-gradient flow of E starting from $y \in \overline{D(\mathsf{E})}$ respectively, then

$$\mathsf{d}^2(\gamma_t, \tilde\gamma_t) \leq e^{-2\lambda t}\mathsf{d}^2(x,y), \qquad \forall t \geq 0. \tag{3.3}$$

This means that EVI-gradient flows are unique (provided they exist) and thus if there exists an EVI-gradient flow (γ_t) starting from x, then a 1-parameter semigroup $(\mathsf{S}_t)_{t>0}$ is unambiguously associated to it via $\mathsf{S}_t(x) = \gamma_t$.

(ii) The maps $t \mapsto \gamma_t$ and $t \mapsto \mathsf{E}(\gamma_t)$ are locally Lipschitz in $(0,\infty)$ with values in X and $\mathbb{R}$, respectively, and satisfy the *Energy Dissipation Equality*

$$-\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}(\gamma_t) = \frac{1}{2}|\dot\gamma_t|^2 + \frac{1}{2}|\partial \mathsf{E}|^2(\gamma_t) = |\dot\gamma_t|^2 = |\partial \mathsf{E}|^2(\gamma_t), \qquad \text{for a.e. } t > 0. \tag{3.4}$$

(iii) The map

$$t \mapsto e^{\lambda t}|\partial \mathsf{E}|(\gamma_t) \quad \text{is non-increasing.} \tag{3.5}$$

(iv) If (γ_t) starts from x and $y \in D(|\partial \mathsf{E}|)$, then

$$|\partial \mathsf{E}|^2(\gamma_t) \leq \frac{1}{2e^{\lambda t} - 1}|\partial \mathsf{E}|^2(y) + \frac{1}{I_\lambda(t)^2}\mathsf{d}^2(x,y), \qquad \text{provided } -\lambda t < \log 2, \tag{3.6}$$

where $I_\lambda(t) := \int_0^t e^{\lambda s}\,\mathrm{d}s$.

We emphasize that these properties directly follow from the very definition (EVI$_\lambda$) of gradient flows, and a priori do not require E to be geodesically λ-convex. Although analogous statements can be found in [2] and [1] under convexity assumptions on E, the latter are essentially needed to grant *existence* of EVI-gradient flows. It is important to stress this fact because in Setting 3.1 we *assume* that for any $x \in \mathrm{X}$ there exists an EVI$_\lambda$-gradient flow of E starting there, which from [27] is known to imply that E is geodesically convex. In Section 3 we also provide an alternative proof of this latter fact, whence the necessity for us to avoid all properties of EVI-gradient flows which are actually a consequence of geodesic convexity.

We conclude this preliminary part with a general integrability result about EVI$_\lambda$-gradient flows, which we could not find explicitly written in the literature and will be used later on in the proof of Lemma 3.6.

Lemma 3.5. *With the same assumptions and notations as in Setting 3.1, let $x \in \mathrm{X}$. Then*

$$t \mapsto \mathsf{E}(\mathsf{S}_t x) \text{ is integrable in } [0,T], \text{ for all } T > 0,$$

regardless of whether $\mathsf{E}(x)$ is finite or not.

On intervals $[\varepsilon, T]$ this computation is easily justified by the fact that $t \mapsto \mathsf{E}(\mathsf{S}_t x)$ is locally Lipschitz in $(0,\infty)$, hence locally integrable therein. But this computation is legitimate even if $\varepsilon = 0$, as we are going to see.

Proof: Let $x \in \mathrm{X}$ and $T > 0$ be as in our statement. Since E is bounded from below on d-bounded sets by (A2), and because $(\mathsf{S}_t x)_{t\in[0,T]}$ is bounded, there exists $c \in \mathbb{R}$ such that $\mathsf{E}(\mathsf{S}_t x) \geq c$ for all $t \in [0,T]$. Combining with (EVI$_\lambda$) this gives

$$c \leq \mathsf{E}(\mathsf{S}_t x) \leq \mathsf{E}(y) - \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{d}^2(\mathsf{S}_t x, y) - \frac{\lambda}{2}\mathsf{d}^2(\mathsf{S}_t x, y)$$

for any $y \in D(\mathsf{E})$ and $t \in (0,T]$. Integrating from $t = \eta > 0$ to $t = T$ gives

$$c(T-\eta) \leq \int_\eta^T \mathsf{E}(\mathsf{S}_t x)\,\mathrm{d}t \leq (T-\eta)\mathsf{E}(y) - \frac{1}{2}\Big(\mathsf{d}^2(\mathsf{S}_T x, y) - \mathsf{d}^2(\mathsf{S}_\eta x, y)\Big)$$
$$- \frac{\lambda}{2}\int_\eta^T \mathsf{d}^2(\mathsf{S}_t x, y)\,\mathrm{d}t.$$

As $t \mapsto \mathsf{E}(\mathsf{S}_t x)$ is bounded from below on $[0,T]$ and the right-hand side has a finite limit as $\eta \downarrow 0$ (thanks to the fact that $t \mapsto \mathsf{S}_t x$ is d-continuous on

$[0, \infty)$ by the very definition of EVI$_\lambda$-gradient flow), we deduce the desired integrability. ∎

3.2. A pseudo-Riemannian computation. In this section the formal Riemannian computations carried out at the beginning of Section 2, and more precisely (2.4), will be reproduced rigorously in the abstract Setting 3.1. To this aim, a key role will be played by the following purely metric estimate:

Lemma 3.6. *With the same assumptions and notations as in Setting 3.1, let* $(\gamma_t) \in AC([0,1], \mathrm{X})$ *with* $\mathsf{E}(\gamma_0), \mathsf{E}(\gamma_1) < \infty$. *For any fixed absolutely continuous function* $h : [0,1] \to \mathbb{R}$ *with* $h(t) > 0$ *for all* $t \in (0,1)$ *let*

$$\tilde{\gamma}_t := \mathsf{S}_{h(t)}\gamma_t, \qquad t \in [0,1],$$

and for any $0 \le t_0 < t_1 \le 1$ *write*

$$t^+ := \begin{cases} t_1 & \text{if } h(t_1) \ge h(t_0) \\ t_0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad t^- := \begin{cases} t_0 & \text{if } h(t_1) \ge h(t_0) \\ t_1 & \text{otherwise} \end{cases}. \tag{3.7}$$

Then we have the exact estimate

$$\begin{aligned} \frac{1}{2}\left|\frac{\mathsf{d}(\tilde{\gamma}_{t_1}, \tilde{\gamma}_{t_0})}{t_1 - t_0}\right|^2 &+ \frac{1}{2\lambda^2}|\partial \mathsf{E}|^2(\tilde{\gamma}_{t^+})\frac{e^{\lambda(h(t_1)-h(t_0))} + e^{\lambda(h(t_0)-h(t_1))} - 2}{(t_1-t_0)^2} \\ &+ \frac{1 - e^{-\lambda(h(t^+)-h(t^-))}}{\lambda(t^+ - t^-)} \cdot \frac{\mathsf{E}(\tilde{\gamma}_{t_1}) - \mathsf{E}(\tilde{\gamma}_{t_0})}{t_1 - t_0} \\ &\le \frac{1}{2}e^{-\lambda(h(t_1)+h(t_0))}\left|\frac{\mathsf{d}(\gamma_{t_1}, \gamma_{t_0})}{t_1 - t_0}\right|^2. \end{aligned} \tag{3.8}$$

Here we use the convention that $(+\infty) \times 0 = 0$ whenever $|\partial \mathsf{E}|(\tilde{\gamma}_{t^+}) = +\infty$ and $h(t_0) = h(t_1)$ in the second term on the left-hand side of (3.8). Since we assume that $h(t) > 0$ for $t \in (0,1)$, and because any EVI$_\lambda$-gradient flow immediately falls within $D(|\partial \mathsf{E}|)$ by standard regularizing effects, this latter case is in fact only possible if $t_0 = 0$, $t_1 = 1$, and $h(t_0) = h(t_1) = 0$. In that case $\tilde{\gamma}_0 = \gamma_0$ and $\tilde{\gamma}_1 = \gamma_1$, the third term in the left-hand side also cancels owing to $e^{-\lambda(h(t^+)-h(t^-))} = 1$, and (3.8) then holds as a trivial equality.

We shall rely on this lemma later on in two different ways: First, fixing $t_0 = 0$ and letting $t_1 \downarrow 0$ (resp. fixing $t_1 = 1$ and letting $t_0 \uparrow 1$) to control in Lemma 3.10 the continuity of $t \mapsto \mathsf{E}(\tilde{\gamma}_t)$ at the boundaries $t = 0, 1$, and second, fixing $t_0 \in (0,1)$ and letting $t_1 \to t_0$ to obtain in Proposition 3.11 a pointwise differential estimate similar to (2.4).

Remark 3.7. The times $t^{\pm}$ are just a fancy notation, ordered as $h(t^-) \leq h(t^+)$. Note that in our estimate (3.8) the Fisher information $|\partial \mathsf{E}|^2(\tilde{\gamma}_{t^+})$ is evaluated at the time $t = t^+$ for which the "smoothing time" $s = h(t_0)$ or $s = h(t_1)$ is the largest, i.e. where the regularizing vertical flow has been run for the longest time. This is somehow natural, as this specific point is "better" than the other one in terms of regularity. ■

Proof: By symmetry we only discuss the case $h(t_1) \geq h(t_0)$, i.e. $t^+ = t_1$ and $t^- = t_0$. As already mentioned, if $h(t_0) = h(t_1) = 0$ our statement is actually vacuous, thus it is not restrictive to further assume $h(t_1) > 0$. Let us write for simplicity

$$\hat{\gamma}_{t_1} := \mathsf{S}_{h(t_0)}\gamma_{t_1}.$$

From an intuitive point of view, this corresponds to freezing a "vertical time" $s = h(t_0)$ and "translating" $\tilde{\gamma}_{t_0}$ in the "horizontal" t direction parallel to the curve γ until t_1. Here, in the "vertical" direction above t_1 the smoothing semi-group S_s associated with E has been run at least for a strictly positive time $h(t_1) > 0$, so that by (3.6) the solution of the "vertical" gradient flow at that time lies within the regular domain $\mathrm{X}_1 = D(|\partial \mathsf{E}|) \subset \mathrm{X}_0 = D(\mathsf{E}) \subset \mathrm{X}$, see Figure 2.

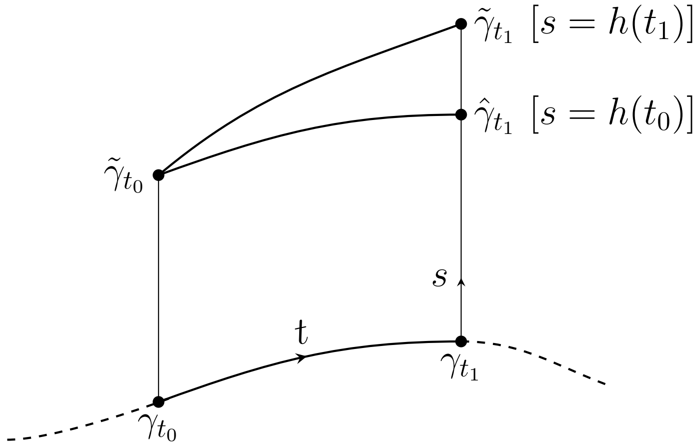

FIGURE 2. The horizontal and vertical curves

The first step is to write (EVI_λ) for $s \mapsto \mathsf{S}_s(\gamma_{t_1})$ with $\tilde{\gamma}_{t_0}$ as a reference point, namely

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}s}\mathsf{d}^2(\mathsf{S}_s\gamma_{t_1}, \tilde{\gamma}_{t_0}) + \frac{\lambda}{2}\mathsf{d}^2(\mathsf{S}_s\gamma_{t_1}, \tilde{\gamma}_{t_0}) + \mathsf{E}(\mathsf{S}_s\gamma_{t_1}) \leq \mathsf{E}(\tilde{\gamma}_{t_0}),$$

which holds true for a.e. $s \in [0, h(t_1)]$ in the "vertical" direction. This inequality can be equivalently rewritten as

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}s}\Big(e^{\lambda s}\mathsf{d}^2(\mathsf{S}_s\gamma_{t_1}, \tilde{\gamma}_{t_0})\Big) \le e^{\lambda s}\Big(\mathsf{E}(\tilde{\gamma}_{t_0}) - \mathsf{E}(\mathsf{S}_s\gamma_{t_1})\Big). \tag{3.9}$$

Note that this estimate carries significant information if and only if the reference point has finite entropy, i.e. $\mathsf{E}(\tilde{\gamma}_{t_0}) < \infty$ in the right-hand side. This holds true for $t_0 \in (0,1)$ because $\tilde{\gamma}_{t_0}$ is the EVI$_\lambda$-gradient flow of E starting from γ_{t_0} at a strictly positive time $s = h(t_0) > 0$, but also for $h(t_0) = 0$ if $t_0 = 0$ since in this case $\tilde{\gamma}_0 = \gamma_0$ is assumed to have finite entropy.

Integrating (3.9) from $s = h(t_0)$ to $s = h(t_1)$ gives

$$\begin{aligned}
&\frac{1}{2}e^{\lambda h(t_1)}\mathsf{d}^2(\tilde{\gamma}_{t_1}, \tilde{\gamma}_{t_0}) - \frac{1}{2}e^{\lambda h(t_0)}\mathsf{d}^2(\hat{\gamma}_{t_1}, \tilde{\gamma}_{t_0}) \\
&\qquad \le \int_{h(t_0)}^{h(t_1)} e^{\lambda s}\Big(\mathsf{E}(\tilde{\gamma}_{t_0}) - \mathsf{E}(\mathsf{S}_s\gamma_{t_1})\Big)\,\mathrm{d}s \\
&\qquad = \int_{h(t_0)}^{h(t_1)} e^{\lambda s}\Big(\big(\mathsf{E}(\tilde{\gamma}_{t_0}) - \mathsf{E}(\tilde{\gamma}_{t_1})\big) + \big(\mathsf{E}(\tilde{\gamma}_{t_1}) - \mathsf{E}(\mathsf{S}_s\gamma_{t_1})\big)\Big)\,\mathrm{d}s \\
&\qquad = \int_{h(t_0)}^{h(t_1)} e^{\lambda s}\Big(\mathsf{E}(\tilde{\gamma}_{t_1}) - \mathsf{E}(\mathsf{S}_s\gamma_{t_1})\Big)\,\mathrm{d}s \\
&\qquad\qquad - \frac{e^{\lambda h(t_1)} - e^{\lambda h(t_0)}}{\lambda}\Big(\mathsf{E}(\tilde{\gamma}_{t_1}) - \mathsf{E}(\tilde{\gamma}_{t_0})\Big).
\end{aligned} \tag{3.10}$$

If $h(t_0) > 0$ this computation is legitimate because $s \mapsto \mathsf{S}_s\gamma_{t_1}$ and $s \mapsto \mathsf{E}(\mathsf{S}_s\gamma_{t_1})$ are locally Lipschitz in $(0,\infty)$, hence $s \mapsto \mathsf{d}^2(\mathsf{S}_s\gamma_{t_1}, \tilde{\gamma}_{t_0})$ and $s \mapsto \mathsf{E}(\mathsf{S}_s\gamma_{t_1})$ are locally integrable therein. But this computation is also justified when $h(t_0) = 0$ by Lemma 3.5. More specifically $s \mapsto \mathsf{E}(\mathsf{S}_s\gamma_{t_1})$ is absolutely integrable on $[0,T]$ for any $T > 0$ and a fortiori so is $s \mapsto e^{\lambda s}\mathsf{E}(\mathsf{S}_s\gamma_{t_1})$.

Now let us estimate the terms in (3.10) to get (3.8). First, since $\hat{\gamma}_{t_1} = \mathsf{S}_{h(t_0)}\gamma_{t_1}$, $\tilde{\gamma}_{t_0} = \mathsf{S}_{h(t_0)}\gamma_{t_0}$, and $\mathsf{S}_{h(t_0)}(\cdot)$ is λ-contractive by (3.3), we observe that the second term in the left-hand side of (3.10) can be controlled as

$$\mathsf{d}^2(\hat{\gamma}_{t_1}, \tilde{\gamma}_{t_0}) = \mathsf{d}^2(\mathsf{S}_{h(t_0)}\gamma_{t_1}, \mathsf{S}_{h(t_0)}\gamma_{t_0}) \le e^{-2\lambda h(t_0)}\mathsf{d}^2(\gamma_{t_1}, \gamma_{t_0}). \tag{3.11}$$

On the right-hand side, let us define

$$I := \int_{h(t_0)}^{h(t_1)} e^{\lambda s}\Big(\mathsf{E}(\tilde{\gamma}_{t_1}) - \mathsf{E}(\mathsf{S}_s\gamma_{t_1})\Big)\,\mathrm{d}s.$$

This integral is clearly non-positive by (3.4), but we need a finer analysis. To this aim, for fixed $0 < s < h(t_1)$ let us write

$$\begin{aligned}\mathsf{E}(\tilde{\gamma}_{t_1}) - \mathsf{E}(\mathsf{S}_s\gamma_{t_1}) &= \mathsf{E}(\mathsf{S}_{h(t_1)}\gamma_{t_1}) - \mathsf{E}(\mathsf{S}_s\gamma_{t_1}) \\ &= \int_s^{h(t_1)} \frac{\mathrm{d}}{\mathrm{d}\tau}\mathsf{E}(\mathsf{S}_\tau\gamma_{t_1})\,\mathrm{d}\tau = -\int_s^{h(t_1)} |\partial\mathsf{E}|^2(\mathsf{S}_\tau\gamma_{t_1})\,\mathrm{d}\tau,\end{aligned}$$

where the second equality holds due to $\tau \mapsto \mathsf{E}(\mathsf{S}_\tau\gamma_{t_1})$ being Lipschitz on $[s, h(t_1)]$, and the third one stems from (3.4) for the gradient flow $\tau \mapsto \mathsf{S}_\tau\gamma_{t_1}$. By (3.5)

$$\begin{aligned}-\int_s^{h(t_1)} |\partial\mathsf{E}|^2(\mathsf{S}_\tau\gamma_{t_1})\,\mathrm{d}\tau &\leq -\int_s^{h(t_1)} |\partial\mathsf{E}|^2(\mathsf{S}_{h(t_1)}\gamma_{t_1})e^{2\lambda(h(t_1)-\tau)}\,\mathrm{d}\tau \\ &= -e^{2\lambda h(t_1)}|\partial\mathsf{E}|^2(\mathsf{S}_{h(t_1)}\gamma_{t_1})\int_s^{h(t_1)} e^{-2\lambda\tau}\,\mathrm{d}\tau \\ &= \frac{1}{2\lambda}\Big(1 - e^{2\lambda(h(t_1)-s)}\Big)|\partial\mathsf{E}|^2(\mathsf{S}_{h(t_1)}\gamma_{t_1}),\end{aligned}$$

so that, as a consequence,

$$\begin{aligned}I &\leq \frac{1}{2\lambda}\int_{h(t_0)}^{h(t_1)} e^{\lambda s}\Big(1 - e^{2\lambda(h(t_1)-s)}\Big)|\partial\mathsf{E}|^2(\mathsf{S}_{h(t_1)}\gamma_{t_1})\,\mathrm{d}s \\ &= \frac{1}{2\lambda}|\partial\mathsf{E}|^2(\mathsf{S}_{h(t_1)}\gamma_{t_1})\int_{h(t_0)}^{h(t_1)} \Big(e^{\lambda s} - e^{2\lambda h(t_1)}\cdot e^{-\lambda s}\Big)\,\mathrm{d}s \\ &= -\frac{1}{2\lambda^2}|\partial\mathsf{E}|^2(\tilde{\gamma}_{t_1})e^{\lambda h(t_1)}\Big(e^{\lambda(h(t_1)-h(t_0))} + e^{\lambda(h(t_0)-h(t_1))} - 2\Big).\end{aligned}$$

Plugging this estimate together with (3.11) into (3.10) and dividing by $(t_1 - t_0)^2 > 0$ entails our claim. ■

We also need to study the behaviour of the "regularized" curve $\tilde{\gamma}_t := \mathsf{S}_{h(t)}\gamma_t$ and of the entropy E along it: this is the content of the following two results.

Lemma 3.8. *With the same assumptions and notations as in Setting 3.1, if $(\gamma_t) \in AC([0,1],\mathrm{X})$ and $h : [0,1] \to \mathbb{R}$ is absolutely continuous with $h(t) > 0$ for all $t \in (0,1)$, then the curve $\tilde{\gamma}_t := \mathsf{S}_{h(t)}\gamma_t$ belongs to $AC_{loc}((0,1),\mathrm{X}) \cap C([0,1],\mathrm{X})$.*

Proof: Fix $\delta \in (0,1/2)$, $t_0, t_1 \in [\delta, 1-\delta]$ with $t_0 \leq t_1$ and define

$$m_\delta := \min_{t\in[\delta,1-\delta]} h(t), \qquad M_\delta := \max_{t\in[\delta,1-\delta]} h(t), \tag{3.12}$$

paying attention to the fact that $m_\delta > 0$ by construction. Write as before $\hat\gamma_{t_1} := \mathsf{S}_{h(t_0)}\gamma_{t_1}$ for the "horizontal" translation of $\tilde\gamma_{t_0}$ (see Figure 2). By triangular inequality and the contraction estimate (3.3) we get

$$\mathsf{d}(\tilde\gamma_{t_0},\tilde\gamma_{t_1}) \le \mathsf{d}(\tilde\gamma_{t_0},\hat\gamma_{t_1}) + \mathsf{d}(\hat\gamma_{t_1},\tilde\gamma_{t_1}) \le e^{\lambda^- M_\delta}\mathsf{d}(\gamma_{t_0},\gamma_{t_1}) + \mathsf{d}(\hat\gamma_{t_1},\tilde\gamma_{t_1}), \tag{3.13}$$

where $\lambda^- := \max\{-\lambda, 0\}$. Since (γ_t) is absolutely continuous the first term in the right-hand side can be controlled as $\mathsf{d}(\gamma_{t_0},\gamma_{t_1}) \le \int_{t_0}^{t_1}|\dot\gamma_t|\mathrm{d}t$. As regards the second one, by (3.4) and up to assuming $h(t_0) \le h(t_1)$ (which is not restrictive, as otherwise it is sufficient to swap the boundary values of integration below) it holds

$$\mathsf{d}(\hat\gamma_{t_1},\tilde\gamma_{t_1}) = \mathsf{d}(\mathsf{S}_{h(t_0)}\gamma_{t_1},\mathsf{S}_{h(t_1)}\gamma_{t_1}) \le \int_{h(t_0)}^{h(t_1)} \Big|\frac{\mathrm{d}}{\mathrm{d}s}\mathsf{S}_s\gamma_{t_1}\Big|\,\mathrm{d}s = \int_{h(t_0)}^{h(t_1)} |\partial\mathsf{E}|(\mathsf{S}_s\gamma_{t_1})\,\mathrm{d}s, \tag{3.14}$$

where, to avoid possibly ambiguous notations, $|\frac{\mathrm{d}}{\mathrm{d}s}\mathsf{S}_s\gamma_{t_1}|$ denotes the metric speed of the "vertical" curve $s \mapsto \mathsf{S}_s\gamma_{t_1}$. In order to control the slope in the right-most term uniformly both in $s \in [h(t_0), h(t_1)] \subset [m_\delta, M_\delta]$ and in $t_1 \in [\delta, 1-\delta]$, for fixed δ, let ε be such that $-\lambda\varepsilon < \log 2$ (if $\lambda \ge 0$, choose $\varepsilon = m_\delta$) and define $\varepsilon' := \min\{m_\delta, \varepsilon\}$. Then by (3.5) and the fact that $s \ge h(t_0) \ge m_\delta \ge \varepsilon'$ we have

$$|\partial\mathsf{E}|(\mathsf{S}_s\gamma_{t_1}) \le e^{\lambda(\varepsilon'-s)}|\partial\mathsf{E}|(\mathsf{S}_{\varepsilon'}\gamma_{t_1}) \le e^{\lambda^-(M_\delta-\varepsilon')}|\partial\mathsf{E}|(\mathsf{S}_{\varepsilon'}\gamma_{t_1}), \qquad \forall s \in [m_\delta, M_\delta]$$

and by (3.6) for any reference point $x \in D(|\partial\mathsf{E}|)$ it holds

$$|\partial\mathsf{E}|^2(\mathsf{S}_{\varepsilon'}\gamma_{t_1}) \le \frac{1}{2e^{\lambda\varepsilon'}-1}|\partial\mathsf{E}|^2(x) + \frac{1}{I_\lambda(\varepsilon')^2}\mathsf{d}^2(x,\gamma_{t_1}).$$

The squared distance in the right-hand side above is bounded uniformly in $t_1 \in [\delta, 1-\delta]$, since by triangular inequality

$$\mathsf{d}(x,\gamma_{t_1}) \le \mathsf{d}(x,\gamma_0) + \mathsf{d}(\gamma_0,\gamma_{t_1}) \le \mathsf{d}(x,\gamma_0) + \ell(\gamma).$$

Therefore there exists $C_\delta > 0$ such that

$$|\partial\mathsf{E}|(\mathsf{S}_s\gamma_{t_1}) \le C_\delta \qquad \text{for all} \qquad t_1 \in [\delta, 1-\delta] \text{ and } s \in [m_\delta, M_\delta]. \tag{3.15}$$

and plugging this bound into (3.14) yields

$$\mathsf{d}(\hat\gamma_{t_1},\tilde\gamma_{t_1}) \le C_\delta|h(t_1)-h(t_0)| \le C_\delta\int_{t_0}^{t_1}|h'(t)|\,\mathrm{d}t.$$

It is now sufficient to combine this inequality with $\mathsf{d}(\gamma_{t_0},\gamma_{t_1}) \le \int_{t_0}^{t_1} |\dot\gamma_t| \mathrm{d}t$ and (3.13) to get

$$\mathsf{d}(\tilde\gamma_{t_0},\tilde\gamma_{t_1}) \le \int_{t_0}^{t_1} \Big(e^{\lambda^- M_\delta} |\dot\gamma_t| + C_\delta |h'(t)| \Big) \,\mathrm{d}t. \tag{3.16}$$

As $e^{\lambda^- M_\delta}|\dot\gamma_t| + C_\delta|h'(t)| \in L^1(\delta, 1-\delta)$ and δ is arbitrary, the fact that $(\tilde\gamma_t) \in AC_{loc}((0,1))$ is thus proved.

Turning now to the continuity of $(\tilde\gamma_t)$ at the endpoints, let $t_0 = 0$ and $t_1 \in (0,1)$. Arguing as for (3.13) but with a crucial difference in the choice of the third point in the triangular inequality, it holds

$$\begin{aligned}\mathsf{d}(\tilde\gamma_0,\tilde\gamma_{t_1}) &\le \mathsf{d}(\tilde\gamma_0, \mathsf{S}_{h(t_1)}\gamma_0) + \mathsf{d}(\mathsf{S}_{h(t_1)}\gamma_0,\tilde\gamma_{t_1}) \\ &\le \mathsf{d}(\mathsf{S}_{h(0)}\gamma_0, \mathsf{S}_{h(t_1)}\gamma_0) + e^{-\lambda h(t_1)}\mathsf{d}(\gamma_0,\gamma_{t_1}).\end{aligned}$$

The second term on the right-hand side vanishes as $t_1 \downarrow 0$ by (absolute) continuity of γ and so does the first one, since $s \mapsto \mathsf{S}_s\gamma_0$ is continuous in $[0,\infty)$ with values in X and $h(t_1) \to h(0)$. The continuity at $t = 1$ is obtained similarly and the proof is complete. ■

Remark 3.9. If $h(t) > 0$ also in $t = 0, 1$, then the previous argument can be extended to the whole interval $[0,1]$ and therefore $(\tilde\gamma_t) \in AC([0,1],\mathrm{X})$. ■

Lemma 3.10. *With the same assumptions and notations as in Lemma 3.8, the entropy is locally absolutely continuous in* $(0,1)$ *along the regularized curve* $\tilde\gamma_t$*, i.e.*

$$t \mapsto \mathsf{E}(\tilde\gamma_t) \quad \in \quad AC_{loc}((0,1)).$$

If in addition $(\gamma_t) \in AC^2([0,1],\mathrm{X})$*,* $\mathsf{E}(\tilde\gamma_0), \mathsf{E}(\tilde\gamma_1) < \infty$ *and* h *is differentiable at* $t = 0$ *and* $t = 1$ *with* $h'(0) > 0$ *and* $h'(1) < 0$*, then*

$$t \mapsto \mathsf{E}(\tilde\gamma_t) \quad \in \quad C([0,1]).$$

Note that $\mathsf{E}(\tilde\gamma_0), \mathsf{E}(\tilde\gamma_1) < \infty$ is automatically satisfied if $h(t) > 0$ also in $t = 0, 1$.

Proof: Let us first prove that $t \mapsto \mathsf{E}(\tilde\gamma_t)$ is locally absolutely continuous. Since $|\partial\mathsf{E}|$ is a strong upper-gradient, the chain rule (3.2) holds and it suffices to show that $|\partial\mathsf{E}|(\tilde\gamma_t)|\dot{\tilde\gamma}_t| \in L^1_{loc}(0,1)$, namely

$$\int_\delta^{1-\delta} |\partial\mathsf{E}|(\tilde\gamma_t)|\dot{\tilde\gamma}_t| \,\mathrm{d}t < \infty, \qquad \forall \delta \in (0,1/2), \tag{3.17}$$

as this would imply that $\mathsf{E}\circ\tilde\gamma \in AC_{loc}((0,1))$ with

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}(\mathsf{E}\circ\tilde\gamma)(t)\right| \le |\partial\mathsf{E}|(\tilde\gamma_t)\cdot|\dot{\tilde\gamma}_t|, \qquad \text{for a.e. } t\in(0,1).$$

To this aim, observe from (3.16) that $|\dot{\tilde\gamma}_t| \in L^1_{loc}(0,1)$ with $|\dot{\tilde\gamma}_t| \le e^{\lambda^- M_\delta}|\dot\gamma_t| + C_\delta|h'(t)|$ a.e. on $[\delta,1-\delta]$, with M_δ defined in (3.12). Moreover from (3.15) we also know that $|\partial\mathsf{E}|(\mathsf{S}_s\gamma_t)\le C_\delta$ uniformly in $t\in[\delta,1-\delta]$ and $s\in[m_\delta,M_\delta]$, so that by choosing $s=h(t)$ we get in particular $|\partial\mathsf{E}|(\tilde\gamma_t)\le C_\delta$ for all $t\in[\delta,1-\delta]$. This shows that $t\mapsto|\partial\mathsf{E}|(\tilde\gamma_t)$ belongs to $L^\infty_{loc}(0,1)$, whence (3.17).

Now assume that $(\gamma_t)\in AC^2([0,1],\mathrm{X})$, $\mathsf{E}(\tilde\gamma_0)<\infty$, h is differentiable at $t=0$ with $h'(0)>0$ and let us prove that $t\mapsto\mathsf{E}(\tilde\gamma_t)$ is continuous at $t=0$. (The argument is identical for $t=1$.) On the one hand, as $(\tilde\gamma_t)$ is continuous at $t=0$ by Lemma 3.8 and E is lower semicontinuous, we see that $\mathsf{E}(\tilde\gamma_0)\le\liminf_{t\downarrow 0}\mathsf{E}(\tilde\gamma_t)$. On the other hand, choosing $t_0=0$ in (3.8), our assumption that $h'(0)>0$ gives $h(t_1)>h(0)$ for $t_1>0$ small, hence $t^-=0$ and $t^+=t_1$. Discarding the first two (non-negative) terms on the left-hand side, and multiplying by $(t_1-t_0)=t_1$ yield

$$\begin{aligned}\frac{1-e^{-\lambda(h(t_1)-h(0))}}{\lambda(t_1-0)}\cdot\big(\mathsf{E}(\tilde\gamma_{t_1})-\mathsf{E}(\tilde\gamma_0)\big) &\le \frac{t_1}{2}e^{-\lambda(h(t_1)+h(0))}\left|\frac{\mathsf{d}(\gamma_{t_1},\gamma_0)}{t_1}\right|^2\\ &\le \frac{t_1}{2}e^{-\lambda(h(t_1)+h(0))}\Big(\frac{1}{t_1}\int_0^{t_1}|\dot\gamma_t|\,\mathrm{d}t\Big)^2.\\ &\le \frac{1}{2}e^{-\lambda(h(t_1)+h(0))}\int_0^{t_1}|\dot\gamma_t|^2\,\mathrm{d}t.\end{aligned}$$

Letting $t_1\downarrow 0$, the right-hand side vanishes owing to our assumption that $(\gamma_t)\in AC^2([0,1],\mathrm{X})$, and clearly the exponential difference quotient in the left-hand side converges to $h'(0)$. Rearranging gives

$$h'(0)\limsup_{t_1\downarrow 0}\mathsf{E}(\tilde\gamma_{t_1})\le h'(0)\mathsf{E}(\tilde\gamma_0),$$

since $h'(0)>0$ the desired upper semicontinuity follows and the proof is complete. ∎

Gathering the results proven so far, we deduce the following:

Proposition 3.11. *With the same assumptions and notations as in Lemma 3.8, for a.e.* $t \in (0,1)$ *it holds*

$$\frac{1}{2}\left|\dot{\tilde{\gamma}}_t\right|^2 + \frac{1}{2}|h'(t)|^2|\partial\mathsf{E}|^2(\tilde{\gamma}_t) + h'(t)\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}(\tilde{\gamma}_t) \leq \frac{1}{2}e^{-2\lambda h(t)}\left|\dot{\gamma}_t\right|^2. \tag{3.18}$$

Proof: The argument simply consists in taking the limit $t_1 \to t_0$ in (3.8), which should clearly lead (at least formally) to (3.18) by Taylor-expanding the various exponential difference quotients. In order to make this rigorous, note that the first and third terms in the left-hand side of (3.18) are well defined for a.e. $t \in (0,1)$ by Lemma 3.8 and Lemma 3.10, respectively. The second term is also unambiguously defined because $h(t) > 0$, hence the "vertical" EVI_λ-gradient flow starting from γ_t and defining $\tilde{\gamma}_t = \mathsf{S}_{h(t)}\gamma_t$ falls immediately within the domain $\mathrm{X}_1 = D(|\partial\mathsf{E}|)$. The right-hand side is well defined for a.e. t since $\gamma \in AC([0,1],\mathrm{X})$.

After this premise, let $t \in (0,1)$ be any differentiation point for h, $t \mapsto \gamma_t$, $t \mapsto \tilde{\gamma}_t$ and $t \mapsto \mathsf{E}(\tilde{\gamma}_t)$, choose $t_0 = t$ in (3.8) and let us take the right limit $t_1 \downarrow t_0$ (since we are considering a differentiability point the left and right limits exist and are equal, so there is no need to address the left limit). From the very definition (3.7) of $t^\pm$ it clearly holds $t^\pm \to t_0$ as $t_1 \downarrow t$, hence the convergence of the right-hand side of (3.8) to the right-hand side of (3.18) is clear and so is the convergence of the two difference quotients of h. By Lemma 3.8 the first term in the left-hand side also passes to the limit, as does the third one according to Lemma 3.10. The only term left to handle is the Fisher information $|\partial\mathsf{E}|^2(\tilde{\gamma}^+)$. From the continuity of $t \mapsto \tilde{\gamma}_t$ (cf. Lemma 3.8) we see that $\tilde{\gamma}_{t^+} \to \tilde{\gamma}_t$ in (X,d), and the lower semicontinuity of the slope (3.1) results in

$$|\partial\mathsf{E}|(\tilde{\gamma}_t) \leq \liminf_{t_1 \downarrow 0} |\partial\mathsf{E}|(\tilde{\gamma}_{t^+}).$$

Thus rigorously taking the liminf $t_1 \downarrow t_0$ in (3.8) entails (3.18) and achieves the proof. ■

The interesting consequence for our purpose is then:

Theorem 3.12. *With the same assumptions and notations as in Setting 3.1, fix* $\varepsilon > 0$, *and set* $h_\varepsilon(t) := \varepsilon\min\{t, 1-t\}$. *Let* $(\gamma_t) \in AC^2([0,1],\mathrm{X})$ *be such that* $\mathsf{E}(\gamma_0), \mathsf{E}(\gamma_1) < \infty$ *and define*

$$\gamma^\varepsilon_t := \mathsf{S}_{h_\varepsilon(t)}\gamma_t, \qquad t \in [0,1].$$

Then $(\gamma_t^\varepsilon) \in AC^2([0,1],\mathrm{X})$*,* $t \mapsto \mathsf{E}(\gamma_t^\varepsilon)$ *belongs to* $AC([0,1])$ *and it holds*

$$\begin{aligned}\frac{1}{2}\int_0^1 |\dot\gamma_t^\varepsilon|^2\,\mathrm{d}t + \frac{\varepsilon^2}{2}\int_0^1 |\partial\mathsf{E}|^2(\gamma_t^\varepsilon)\,\mathrm{d}t \le \frac{1}{2}e^{\lambda^-\varepsilon}\int_0^1 |\dot\gamma_t|^2\,\mathrm{d}t - 2\varepsilon\mathsf{E}(\gamma_{1/2}^\varepsilon)\\ + \varepsilon\Big(\mathsf{E}(\gamma_0) + \mathsf{E}(\gamma_1)\Big).\end{aligned} \tag{3.19}$$

Note here that $h_\varepsilon(0) = h_\varepsilon(1) = 0$, so that the endpoints $\gamma_0^\varepsilon = \gamma_0$ and $\gamma_1^\varepsilon = \gamma_1$ remain unchanged.

Proof: The strategy of proof simply consists in integrating (3.18) between 0 and 1 while integrating by parts of the term $h_\varepsilon'(t)\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}(\gamma_t^\varepsilon)$, separately on $[0,1/2]$ and $[1/2,1]$. Note carefully that our specific choice gives $h_\varepsilon' = \varepsilon$ and $h_\varepsilon' = -\varepsilon$ on these two time intervals, respectively. Taking into account $e^{-2\lambda h_\varepsilon(t)} \le e^{\lambda^-\varepsilon}$, where $\lambda^- := \max\{-\lambda, 0\}$, this procedure yields

$$\begin{aligned}\frac{1}{2}\int_0^1 |\dot\gamma_t^\varepsilon|^2\,\mathrm{d}t + \frac{\varepsilon^2}{2}\int_0^1 |\partial\mathsf{E}|^2(\gamma_t^\varepsilon)\,\mathrm{d}t \le \frac{1}{2}e^{\lambda^-\varepsilon}\int_0^1 |\dot\gamma_t|^2\,\mathrm{d}t - 2\varepsilon\mathsf{E}(\gamma_{1/2}^\varepsilon)\\ + \varepsilon\Big(\mathsf{E}(\gamma_0^\varepsilon) + \mathsf{E}(\gamma_1^\varepsilon)\Big).\end{aligned}$$

The term $2\varepsilon\mathsf{E}(\gamma_{1/2}^\varepsilon)$ simply arises from the two boundary terms at $t = 1/2$ in the two integrations by parts. (Alternatively, it can be seen as the result of $-\int_0^1 \mathsf{E}(\gamma_t^\varepsilon)h''(t)$ arising from the integration by parts in the whole interval $[0,1]$, with the singularity $h''(t) = -2\varepsilon\delta_{1/2}(t)$). However, this argument is not fully rigorous because all the terms on the left-hand side of (3.18) are only *locally* integrable, hence we may not be allowed to integrate them all the way to $t = 0$ and $t = 1$.

In order to circumvent this slight issue, choose $\delta \in (0,1/2)$ and carry out the same argument on $[\delta,1/2]$ and $[1/2,1-\delta]$ rather than on $[0,1/2]$ and $[1/2,1]$: Integration by parts is now justified by Lemma 3.10 and this provides us with

$$\begin{aligned}\frac{1}{2}\int_\delta^{1-\delta} |\dot\gamma_t^\varepsilon|^2\,\mathrm{d}t + \frac{\varepsilon^2}{2}\int_\delta^{1-\delta} |\partial\mathsf{E}|^2(\gamma_t^\varepsilon)\,\mathrm{d}t \le \frac{1}{2}e^{\lambda^-\varepsilon}\int_\delta^{1-\delta} |\dot\gamma_t|^2\,\mathrm{d}t\\ + \varepsilon\Big(\mathsf{E}(\gamma_\delta^\varepsilon) - 2\mathsf{E}(\gamma_{1/2}^\varepsilon) + \mathsf{E}(\gamma_{1-\delta}^\varepsilon)\Big).\end{aligned} \tag{3.20}$$

It is then sufficient to pass to the limit as $\delta \downarrow 0$. By monotonicity the left-hand side above converges to the left-hand side in (3.19) and for the same reason so does the first term on the right-hand side, while by the current choice of h and by Lemma 3.10 $t \mapsto \mathsf{E}(\gamma_t^\varepsilon)$ is continuous on the whole interval $[0,1]$, so that

$$\lim_{\delta \downarrow 0} \varepsilon\Big(\mathsf{E}(\gamma_\delta^\varepsilon) + \mathsf{E}(\gamma_{1-\delta}^\varepsilon)\Big) = \varepsilon\Big(\mathsf{E}(\gamma_0^\varepsilon) + \mathsf{E}(\gamma_1^\varepsilon)\Big) = \varepsilon\Big(\mathsf{E}(\gamma_0) + \mathsf{E}(\gamma_1)\Big)$$

and (3.19) follows.

Finally, since the right-hand side of (3.19) is finite we see that $|\dot\gamma^\varepsilon| \in L^2(0,1)$ and $|\partial\mathsf{E}|(\gamma^\varepsilon) \in L^2(0,1)$. As a consequence $|\dot\gamma^\varepsilon| \cdot |\partial\mathsf{E}(\gamma^\varepsilon)| \in L^1(0,1)$ in the strong upper-chain rule (3.2), and $\mathsf{E} \circ \gamma^\varepsilon \in AC([0,1])$ as desired. ∎

4. Small-temperature limit and convexity

4.1. Γ-convergence of the Schrödinger problem. Relying on the results of the previous section, we can now turn to Theorem 2.1 and make it rigorous in the metric setting. To this end, let us first introduce two action functionals: the kinetic energy $\mathcal{A}$ and the (halved) Fisher information $\mathcal{I}$ along a curve, respectively defined as

$$\mathcal{A}(\gamma) := \frac{1}{2}\int_0^1 |\dot\gamma_t|^2\,\mathrm{d}t \qquad \text{and} \qquad \mathcal{I}(\gamma) := \frac{1}{2}\int_0^1 |\partial\mathsf{E}|^2(\gamma_t)\,\mathrm{d}t$$

for all $(\gamma_t) \in C([0,1],\mathrm{X})$, where it is understood that $\mathcal{A}(\gamma) = +\infty$ whenever γ is not absolutely continuous. Given two points $x, y \in \mathrm{X}$ and a temperature/slowing-down parameter $\varepsilon > 0$, the (metric) Schrödinger problem reads as

$$\inf_{(\gamma_t)\,:\,x \rightsquigarrow y} \Big\{\mathcal{A}(\gamma) + \varepsilon^2\mathcal{I}(\gamma)\Big\}, \tag{Sch$^\varepsilon$}$$

where $(\gamma_t) : x \rightsquigarrow y$ is a short-hand notation meaning that the infimum runs over all $(\gamma_t) \in C([0,1],\mathrm{X})$ such that $\gamma_0 = x$ and $\gamma_1 = y$. For sake of brevity we also introduce

$$\mathcal{A}_\varepsilon(\gamma) := \mathcal{A}(\gamma) + \varepsilon^2\mathcal{I}(\gamma).$$

From (Sch$^\varepsilon$) it is thus clear that the Fisher information $\mathcal{I}$ acts as a perturbation of $\mathcal{A}$ and this has a regularizing effect, since minimizers of (Sch$^\varepsilon$) live within the regular domain $\mathrm{X}_1 = D(|\partial\mathsf{E}|)$.

Remark 4.1. The smoothing effect is well understood for the classic Schrödinger problem in a regular setting, namely when E is the Boltzmann-Shannon relative entropy and X is the Wasserstein space over a smooth Riemannian manifold. In this case, under mild assumptions on the end-points, minimizers of (Sch$^\varepsilon$) are curves of absolutely continuous measures whose densities are bounded, smooth, Lipschitz, with exponentially fast decaying tails.

In the current metric framework the properties above are meaningless, but still minimizers of (Sch^ε) are "regular" from a metric point of view, since as just said they live within $D(|\partial \mathsf{E}|)$. Moreover, in Proposition 4.2 we are going to see that E is absolutely continuous along optimal curves. ∎

Let us first deal with the solvability of (Sch^ε).

Proposition 4.2. *With the same assumptions and notations as in Setting 3.1 and under Assumption 3.2, for any fixed $x, y \in \mathrm{X}$ and $\varepsilon > 0$ the Schrödinger problem* (Sch^ε) *is solvable if and only if $\mathsf{E}(x), \mathsf{E}(y) < \infty$ and there exists $(\gamma_t) \in AC([0,1], \mathrm{X})$ such that $\gamma_0 = x$ and $\gamma_1 = y$.*

As the condition characterizing the solvability of the Schrödinger problem does not depend on ε, it is clear that if (Sch^ε) is solvable for some $\varepsilon > 0$, then it is actually solvable for all $\varepsilon > 0$.

Proof: Assume that the endpoints have finite entropy and that there exists an absolutely continuous curve γ connecting x to y. Up to reparametrization, we can assume that $(\gamma_t) \in AC^2([0,1], \mathrm{X})$. Theorem 3.12 thus guarantees that $\mathcal{A}_\varepsilon$ is finite along the regularization $(\gamma_t^\varepsilon)_{t \in [0,1]}$ of this curve and therefore the variational problem (Sch^ε) is proper. Let then (γ_t^n) be any minimizing sequence and observe that the kinetic action $\mathcal{A}$ is bounded uniformly in n, say $\mathcal{A}(\gamma^n) \leq C$ for all n. We now observe that for any pair $0 \leq t_0 < t_1 \leq 1$ it holds

$$\mathsf{d}(\gamma_{t_0}^n, \gamma_{t_1}^n) \leq \int_{t_0}^{t_1} |\dot{\gamma}_t^n| \, \mathrm{d}t \leq |t_0 - t_1|^{1/2} \Big(\int_{t_0}^{t_1} |\dot{\gamma}_t^n|^2 \, \mathrm{d}t \Big)^{1/2} \leq C |t_0 - t_1|^{1/2}. \quad (4.1)$$

Since the endpoints are fixed, this implies that the set of points γ_t^n is bounded in (X, d) uniformly in n, t, thus it is σ-relatively sequentially compact by Assumption 3.2. By the refined Arzelà-Ascoli lemma [2, Proposition 3.3.1], there exists a limiting d-continuous (actually 1/2-Hölder continuous) curve γ such that

$$\gamma_t^n \xrightarrow{\sigma} \gamma_t, \qquad \forall t \in [0,1].$$

We now observe that the kinetic action is lower semicontinuous for this pointwise-in-time convergence w.r.t. σ, cf. [3, Section 2.2] (indeed, d is lower semicontinuous w.r.t. σ, hence the 2-energies of the finite partitions of γ are lower semicontinuous w.r.t. σ too, whence the lower semicontinuity of the 2-energy of γ itself). Moreover $|\partial \mathsf{E}|^2$ is also lower semicontinuous w.r.t. σ by hypothesis, and this fact together with Fatou's lemma gives

$$\int_0^1 |\partial \mathsf{E}|^2(\gamma_t)\,\mathrm{d}t \le \int_0^1 \liminf_{n\to\infty} |\partial \mathsf{E}|^2(\gamma_t^n)\,\mathrm{d}t \le \liminf_{n\to\infty} \int_0^1 |\partial \mathsf{E}|^2(\gamma_t^n)\,\mathrm{d}t.$$

Therefore γ is a minimizer of $(\mathrm{Sch}^\varepsilon)$.

Conversely, assume that there exists a minimizer, denoted by γ (the following argument actually works for any curve along which $\mathcal{A}_\varepsilon$ is finite and without Assumption 3.2). Then in particular $t \mapsto |\dot\gamma_t|$ and $t \mapsto |\partial \mathsf{E}|(\gamma_t)$ belong to $L^2(0,1)$ and by (3.2) we see that $t \mapsto \mathsf{E}(\gamma_t)$ is globally absolutely continuous with

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}(\mathsf{E}\circ\gamma)(t)\right| \le |\partial \mathsf{E}|(\gamma_t)\cdot|\dot\gamma_t| \in L^1(0,1).$$

The fact that $(|\dot\gamma_t|) \in L^2(0,1) \subset L^1(0,1)$ trivially implies $(\gamma_t) \in AC([0,1],\mathrm{X})$, whereas the fact that $t \mapsto |\partial \mathsf{E}|(\gamma_t)$ belongs to $L^2(0,1)$ also implies that $|\partial \mathsf{E}|(\gamma_t)$ is finite for a.e. $t \in [0,1]$ and a fortiori so is $\mathsf{E}(\gamma_t)$, since $D(|\partial \mathsf{E}|) \subset D(\mathsf{E})$. Hence let $t^* \in (0,1)$ be any point satifying $\mathsf{E}(\gamma_{t^*}) < \infty$ and note that together with (3.2) this gives the following global upper bound valid for all $t < t^*$

$$\mathsf{E}(\gamma_t) \le \mathsf{E}(\gamma_{t^*}) + \int_t^{t^*} \left|\frac{\mathrm{d}}{\mathrm{d}t}(\mathsf{E}\circ\gamma)(t)\right| \mathrm{d}t \le \mathsf{E}(\gamma_{t^*}) + \int_0^1 \left|\frac{\mathrm{d}}{\mathrm{d}t}(\mathsf{E}\circ\gamma)(t)\right| \mathrm{d}t =: \overline{\mathsf{E}} < \infty.$$

As a consequence, and taking also into account the facts that $t \mapsto \gamma_t$ is d-continuous and E is lower semicontinuous, we get

$$\mathsf{E}(\gamma_0) = \mathsf{E}\big(\lim_{t\to 0}\gamma_t\big) \le \liminf_{t\to 0} \mathsf{E}(\gamma_t) \le \overline{\mathsf{E}}$$

and the proof is thus complete, as the same argument applies *mutatis mutandis* for $t = 1$ too. ■

We now fix $x, y \in \mathrm{X}$ and let $C([0,1],\mathrm{X}) \ni \gamma \mapsto \iota_{01}(\gamma)$ denote the convex indicator of the endpoint constraints, i.e.

$$\iota_{01}(\gamma) = \begin{cases} 0 & \text{if } \gamma_0 = x \text{ and } \gamma_1 = y, \\ +\infty & \text{otherwise.} \end{cases}$$

With this said, we can finally state our Γ-convergence result, where the finite-entropy assumption on the endpoints is motivated by the previous proposition.

Theorem 4.3. *With the same assumptions and notations as in Setting 3.1, if $x, y \in \mathrm{X}$ are such that $\mathsf{E}(x), \mathsf{E}(y) < \infty$, then*

$$\Gamma - \lim_{\varepsilon\to 0}\Big\{\mathcal{A}_\varepsilon + \iota_{01}\Big\} = \mathcal{A} + \iota_{01}$$

for the uniform convergence on $C([0,1],\mathrm{X})$*. If Assumption 3.2 holds, then the* Γ*-convergence also takes place w.r.t. the pointwise-in-time* σ*-topology.*

Proof: The $\Gamma-\liminf$ inequality is rather clear, since the kinetic energy $\mathcal{A}$ is lower semicontinuous both w.r.t. uniform-in-time d-convergence and pointwise-in-time σ-convergence: for the former topology the fact is well known, for the latter it has been discussed in the proof of Proposition 4.2. An analogous claim is also true for the convex indicator ι_{01}. As a consequence, we have that for any γ^ε converging to γ uniformly in time in the metric topology or pointwise in time in the topology σ (if applicable) it holds

$$\begin{aligned}\mathcal{A}(\gamma)+\iota_{01}(\gamma) &\leq \liminf_{\varepsilon\downarrow 0}\mathcal{A}(\gamma^\varepsilon)+\liminf_{\varepsilon\downarrow 0}\iota_{01}(\gamma^\varepsilon)\leq \liminf_{\varepsilon\downarrow 0}\Big\{\mathcal{A}(\gamma^\varepsilon)+\iota_{01}(\gamma^\varepsilon)\Big\}\\ &\leq \liminf_{\varepsilon\downarrow 0}\Big\{\mathcal{A}(\gamma^\varepsilon)+\varepsilon^2\mathcal{I}(\gamma^\varepsilon)+\iota_{01}(\gamma^\varepsilon)\Big\}=\liminf_{\varepsilon\downarrow 0}\Big\{\mathcal{A}_\varepsilon(\gamma^\varepsilon)+\iota_{01}(\gamma^\varepsilon)\Big\},\end{aligned}$$

whence the desired $\Gamma-\liminf$ inequality.

For the $\Gamma-\limsup$, take any $(\gamma_t)\in AC^2([0,1],\mathrm{X})$ connecting x to y (if it does not exist, then there is nothing to prove). Then Theorem 3.12 precisely provides a recovery sequence $\gamma_t^\varepsilon := \mathsf{S}_{h_\varepsilon(t)}\gamma_t$ with h_ε defined as therein, both for the uniform-in-time d-convergence and the pointwise-in-time σ-convergence (the latter is an easy consequence of the former by Remark 3.4). To prove this claim, note that for any $n\in\mathbb{N}$ there exist $t_1,...,t_k\in[0,1]$ such that, for any $t\in[0,1]$, $\mathsf{d}(\gamma_t,\gamma_{t_i})<1/n$ for at least one t_i; in addition, since $\gamma_t^\varepsilon\to\gamma_t$ for all $t\in[0,1]$ there exists ε_n small enough such that $\mathsf{d}(\gamma_{t_i},\gamma_{t_i}^\varepsilon)<1/n$ for all $\varepsilon<\varepsilon_n$ and $i=1,...,k$. As a consequence, taking (3.3) into account,

$$\begin{aligned}\mathsf{d}(\gamma_t,\gamma_t^\varepsilon) &\leq \mathsf{d}(\gamma_t,\gamma_{t_i})+\mathsf{d}(\gamma_{t_i},\mathsf{S}_{h_\varepsilon(t)}\gamma_{t_i})+\mathsf{d}(\mathsf{S}_{h_\varepsilon(t)}\gamma_{t_i},\gamma_t^\varepsilon)\\ &\leq \mathsf{d}(\gamma_t,\gamma_{t_i})+\mathsf{d}(\gamma_{t_i},\mathsf{S}_{h_\varepsilon(t)}\gamma_{t_i})+e^{-\lambda h_\varepsilon(t)}\mathsf{d}(\gamma_t,\gamma_{t_i})\\ &\leq \frac{1}{n}(2+e^{\lambda^-\varepsilon/2})\end{aligned}$$

for all $t \in [0,1]$ and $\varepsilon < \varepsilon_n$ and by the arbitrariness of n we conclude that $\gamma^\varepsilon \to \gamma$ uniformly. Furthermore, the lim sup inequality can be proved as follows:

$$\begin{aligned}
\limsup_{\varepsilon \downarrow 0} \Big\{ \mathcal{A}_\varepsilon(\gamma^\varepsilon) + \iota_{01}(\gamma^\varepsilon) \Big\} &= \limsup_{\varepsilon \downarrow 0} \Big\{ \mathcal{A}(\gamma^\varepsilon) + \varepsilon^2 \mathcal{I}(\gamma^\varepsilon) + 0 \Big\} \\
&\overset{(3.19)}{\leq} \limsup_{\varepsilon \downarrow 0} \Big\{ e^{\lambda^- \varepsilon} \mathcal{A}(\gamma) - 2\varepsilon \mathsf{E}(\gamma^\varepsilon_{1/2}) + \varepsilon \big(\mathsf{E}(x) + \mathsf{E}(y) \big) \Big\} \\
&\leq \limsup_{\varepsilon \downarrow 0} \Big\{ e^{\lambda^- \varepsilon} \mathcal{A}(\gamma) + \varepsilon \big(\mathsf{E}(x) + \mathsf{E}(y) \big) \Big\} - 2 \liminf_{\varepsilon \downarrow 0} \varepsilon \mathsf{E}(\gamma^\varepsilon_{1/2}) \\
&\leq \mathcal{A}(\gamma) = \mathcal{A}(\gamma) + \iota_{01}(\gamma),
\end{aligned}$$

where the third inequality comes from the fact that, for any $\varepsilon \downarrow 0$, $(\gamma^\varepsilon_{1/2})$ is contained in a bounded set and by assumption E is bounded from below on bounded sets, whence $\mathsf{E}(\gamma^\varepsilon_{1/2}) \geq c$ for some $c \in \mathbb{R}$. The proof is thus complete. ■

As an easy consequence of this result we obtain the following:

Corollary 4.4. *With the same assumptions and notations as in Setting 3.1 and under the further requirements that Assumption 3.2 holds and the Schrödinger problem* $(\mathrm{Sch}^\varepsilon)$ *relative to* $x, y \in \mathrm{X}$ *is solvable, let* $\varepsilon_k \downarrow 0$ *and* ω^k *be a minimizer of the corresponding Schrödinger problem* $(\mathrm{Sch}^\varepsilon)$ *with* $\varepsilon = \varepsilon_k$.

Then

$$\lim_{k \to \infty} \Big\{ \mathcal{A}(\omega^k) + \varepsilon_k^2 \mathcal{I}(\omega^k) \Big\} = \inf_{(\gamma_t)\,:\, x \rightsquigarrow y} \mathcal{A}(\gamma).$$

Moreover, there exists $\omega^0 \in C([0,1], \mathrm{X})$ *such that, up to a subsequence,* $\omega^k \to \omega^0$ *in the pointwise-in-time* σ*-topology and*

$$\mathcal{A}(\omega^0) = \inf_{(\gamma_t)\,:\, x \rightsquigarrow y} \mathcal{A}(\gamma).$$

Proof: Recall that, under a mild equi-coercivity condition, Γ-convergence precisely guarantees that the limit of the optimal values of the approximating problems is the optimal value of the limit problem and limits of minimizers are minimizers, cf. [16, Theorem 1.21]. In view of Theorem 4.3 and [16, Theorem 1.21], for the mild equi-coercivity condition to hold it suffices to prove that the set of minimizers $\{\omega^k\}$ is relatively compact in the pointwise-in-time σ-topology. To this aim, the kinetic energies of the curves ω^k are uniformly bounded since

$$\mathcal{A}(\omega^k) \leq \mathcal{A}(\omega^k) + \varepsilon_k^2 \mathcal{I}(\omega^k) \leq \mathcal{A}(\omega^{\bar\varepsilon}) + \varepsilon_k^2 \mathcal{I}(\omega^{\bar\varepsilon}) \leq \mathcal{A}(\omega^{\bar\varepsilon}) + \bar\varepsilon^2 \mathcal{I}(\omega^{\bar\varepsilon}) < +\infty,$$

where $\omega^{\bar\varepsilon}$ is the minimizer for the problem with $\varepsilon = \bar\varepsilon := \sup_k \varepsilon_k$. Arguing as in the proof of Proposition 4.2, we deduce that there exists a continuous curve $(\omega^0_t)_{t\in[0,1]}$ connecting x and y such that, up to extracting a suitable subsequence, $\omega^k_t \to \omega^0_t$ w.r.t. σ as $k\to\infty$ for all $t\in[0,1]$. ■

Remark 4.5. Note that in Corollary 4.4 the curve ω^0 is length-minimizing but not necessarily distance-minimizing, namely it needs not be a geodesic between x and y, since we only know that

$$\inf_{(\gamma_t)\,:\,x\rightsquigarrow y} \mathcal{A}(\gamma) \geq \frac{1}{2}\mathsf{d}^2(x,y)$$

and the inequality might be strict, e.g. if X is a non-convex subset of $\mathbb{R}^d$. However, if (X,d) is a length metric space, i.e. for all $x,y\in\mathrm{X}$ and $\varepsilon>0$ there exists $(\gamma_t)\in AC([0,1],\mathrm{X})$ such that $\gamma_0=x$, $\gamma_1=y$ and $\ell(\gamma)\leq \mathsf{d}(x,y)+\varepsilon$, then the inequality above turns out to be an identity and, as a consequence, ω^0 is a geodesic. This means that for any two points having finite energy there always exists a geodesic connecting them. ■

When the endpoints have infinite entropy, the following variant of Theorem 4.3 may be useful:

Theorem 4.6. *With the same assumptions and notations as in Setting 3.1, let $x,y\in\mathrm{X}$ with possibly $\mathsf{E}(x),\mathsf{E}(y)=+\infty$ and for any fixed $(\varepsilon_n)_{n\in\mathbb{N}}$, $\varepsilon_n\downarrow 0$, let $(\eta_n)_{n\in\mathbb{N}}$ be converging to 0 slowly enough so that*

$$\varepsilon_n\left(\mathsf{E}(\gamma_0^n)+\mathsf{E}(\gamma_1^n)\right)\to 0 \qquad \textit{with} \qquad \gamma_0^n := \mathsf{S}_{\eta_n}x, \quad \gamma_1^n := \mathsf{S}_{\eta_n}y.$$

Then

$$\Gamma-\lim_{n\to\infty}\left\{\mathcal{A}_{\varepsilon_n}+\iota^n_{01}\right\} = \mathcal{A}+\iota_{01},$$

for the uniform convergence on $C([0,1],\mathrm{X})$. If Assumption 3.2 holds, then the Γ-convergence also takes place w.r.t. the pointwise-in-time σ-topology. Here ι^n_{01} and ι_{01} are the convex indicators of the endpoint constraints for γ^n_0,γ^n_1 and x,y, respectively.

Proof: The proof of the $\Gamma-\liminf$ is almost identical to that in Theorem 4.3, with the only extra observation that

$$\iota_{01}(\gamma)\leq \liminf_{n\to\infty}\iota^n_{01}(\gamma^n).$$

For the $\Gamma-\limsup$, observe that if there does not exist $(\gamma_t)\in AC^2([0,1],\mathrm{X})$ joining x and y, then there is nothing to prove. Hence let us suppose that at least one curve $(\gamma_t) \in AC^2([0,1],\mathrm{X})$ connecting x and y exists, fix it and note that Theorem 3.12 applied to the curve $\mathsf{S}_{\eta_n}\gamma_t$ still provides a recovery sequence $\gamma_t^{\varepsilon_n} := \mathsf{S}_{\eta_n + h_{\varepsilon_n}(t)}\gamma_t$ with the same choice $h_{\varepsilon_n}(t) = \varepsilon_n \min\{t, 1-t\}$ as before. Indeed, on the one hand

$$\begin{aligned}
\limsup_{n\to\infty} \Big\{ \mathcal{A}_{\varepsilon_n}(\gamma^{\varepsilon_n}) + \iota_{01}^n(\gamma^{\varepsilon_n}) \Big\} &= \limsup_{n\to\infty} \Big\{ \mathcal{A}(\gamma^{\varepsilon_n}) + \varepsilon_n^2 \mathcal{I}(\gamma^{\varepsilon_n}) + 0 \Big\} \\
&\overset{(3.19)}{\leq} \limsup_{n\to\infty} \Big\{ e^{\lambda^- \varepsilon_n}\mathcal{A}(\mathsf{S}_{\eta_n}\gamma) - 2\varepsilon_n \mathsf{E}(\gamma_{1/2}^{\varepsilon_n}) + \varepsilon_n\big(\mathsf{E}(\gamma_0^n) + \mathsf{E}(\gamma_1^n)\big) \Big\} \\
&\leq \limsup_{n\to\infty} \Big\{ e^{\lambda^- \varepsilon_n}\mathcal{A}(\mathsf{S}_{\eta_n}\gamma) + \varepsilon_n\big(\mathsf{E}(\gamma_0^n) + \mathsf{E}(\gamma_1^n)\big) \Big\} - 2\liminf_{\varepsilon\downarrow 0} \varepsilon \mathsf{E}(\gamma_{1/2}^{\varepsilon}) \\
&\leq \limsup_{n\to\infty} \mathcal{A}(\mathsf{S}_{\eta_n}\gamma) \leq \limsup_{n\to\infty} e^{-2\lambda\eta_n}\mathcal{A}(\gamma) = \mathcal{A}(\gamma),
\end{aligned}$$

where the third inequality follows by the same argument adopted in the proof of the previous theorem and the last one is due to (3.18) with $h(t) \equiv \eta_n$. On the other hand, $\gamma_t^{\varepsilon_n} \to \gamma_t$ uniformly in $t \in [0,1]$ in the d-topology and, if Assumption 3.2 holds, for all $t \in [0,1]$ w.r.t. σ: the argument described in the previous proof applies also here verbatim. ■

As conclusion, in the next proposition we show that any EVI-gradient flow is a solution of the Schrödinger problem with suitable endpoints. Intuitively this is clear, because up to a rescaling factor ε^2 both the trajectories of the gradient flow of E and the solutions to (Sch$^\varepsilon$) must formally satisfy the same Newton equation, namely $\ddot{\gamma}_t = -\nabla\Phi(\gamma_t)$ where the potential Φ is given by (minus) the Fisher information $-|\partial\mathsf{E}|^2$, cf. [33, Remark 6]. This is also in complete analogy with the standard Schrödinger problem, which includes the heat flow as a particular entropic interpolation.

Proposition 4.7. *With the same assumptions and notations as in Setting 3.1, fix $\varepsilon > 0$. Then for all $x, y \in \mathrm{X}$ the following lower bound on the optimal value of* (Sch$^\varepsilon$) *holds*

$$\inf_{(\gamma_t)\,:\,x\rightsquigarrow y} \mathcal{A}_\varepsilon(\gamma) \geq \varepsilon\big|\mathsf{E}(x) - \mathsf{E}(y)\big|. \tag{4.2}$$

If either $y = \mathsf{S}_\varepsilon x$ or $x = \mathsf{S}_\varepsilon y$, then equality is achieved. In the former case the curve $[0,1] \ni t \mapsto \hat{\gamma}_t := \mathsf{S}_{\varepsilon t}x$ is a minimizer in the Schrödinger problem and the optimal value is

$$\inf_{(\gamma_t)\,:\,x\rightsquigarrow y} \mathcal{A}_\varepsilon(\gamma) = \varepsilon\big(\mathsf{E}(x) - \mathsf{E}(\mathsf{S}_\varepsilon x)\big).$$

An analogous statement holds when $x = \mathsf{S}_\varepsilon y$.

Proof: By (3.2) and Young's inequality it follows that for any $(\tilde\gamma_t) \in AC^2([0,\varepsilon],\mathrm{X})$ joining x and y (if it exists; if not, (4.2) is trivial) it holds

$$\big|\mathsf{E}(\tilde\gamma_0) - \mathsf{E}(\tilde\gamma_\varepsilon)\big| \leq \frac{1}{2}\int_0^\varepsilon |\dot{\tilde\gamma}_t|^2\,\mathrm{d}t + \frac{1}{2}\int_0^\varepsilon |\partial\mathsf{E}|^2(\tilde\gamma_t)\,\mathrm{d}t.$$

By setting $\gamma_t := \tilde\gamma_{\varepsilon t}$, $t \in [0,1]$, and by the arbitrariness of $\tilde\gamma$ we thus see that for all $(\gamma_t) \in AC^2([0,1],\mathrm{X})$ joining x and y we have

$$\varepsilon\big|\mathsf{E}(\gamma_0) - \mathsf{E}(\gamma_1)\big| \leq \frac{1}{2}\int_0^1 |\dot\gamma_t|^2\,\mathrm{d}t + \frac{\varepsilon^2}{2}\int_0^1 |\partial\mathsf{E}|^2(\gamma_t)\,\mathrm{d}t,$$

so that

$$\varepsilon\big|\mathsf{E}(\gamma_0) - \mathsf{E}(\gamma_1)\big| \leq \inf_{(\gamma_t)\,:\,x\rightsquigarrow y} \mathcal{A}_\varepsilon(\gamma).$$

Now assume that $y = \mathsf{S}_\varepsilon x$: integrating (3.4) for the EVI_λ-gradient flow $\hat\gamma$ (paying attention to the rescaling factor ε) between 0 and 1 we get

$$\mathcal{A}_\varepsilon(\hat\gamma) = \frac{1}{2}\int_0^1 |\dot{\hat\gamma}_t|^2\,\mathrm{d}t + \frac{\varepsilon^2}{2}\int_0^1 |\partial\mathsf{E}|^2(\hat\gamma_t)\,\mathrm{d}t = \varepsilon\big(\mathsf{E}(x) - \mathsf{E}(y)\big) = \varepsilon\big|\mathsf{E}(x) - \mathsf{E}(y)\big|,$$

where the last equality comes from the fact that $t \mapsto \mathsf{E}(\mathsf{S}_t x)$ is non-increasing, as a consequence of (3.4). Combining this identity with (4.2) yields the conclusion. ∎

4.2. Displacement convexity. In analogy with Section 2.2, in this short section we establish the geodesic λ-convexity of E. As already explained in the Introduction, although the result is known (cf. [27, Theorem 3.2]), our proof is independent and new and is a further evidence of the wide range of applications of the Schrödinger problem. Let us stress once more that all the properties of EVI_λ-gradient flows stated in Section 3.1 and used so far do not rely on geodesic λ-convexity, whence the genuine independence of our approach.

Theorem 4.8. *With the same assumptions and notations as in Setting 3.1, the potential* E *is* λ*-convex along* any *geodesic.*

Proof: Let (γ_t) be any constant-speed geodesic. We want to prove that

$$\mathsf{E}(\gamma_\theta) \leq (1-\theta)\mathsf{E}(\gamma_0) + \theta\mathsf{E}(\gamma_1) - \frac{\lambda}{2}\theta(1-\theta)\mathsf{d}^2(\gamma_0,\gamma_1), \qquad \forall\theta\in[0,1].$$

We will establish this inequality by carefully estimating at order one as $\varepsilon \downarrow 0$ the defect of optimality, in the geodesic problem from γ_0 to γ_1, of a suitably regularized version (γ_t^ε) of the geodesic.

If $\mathsf{E}(\gamma_0) = +\infty$ or $\mathsf{E}(\gamma_1) = +\infty$ there is nothing to prove, so we can assume without loss of generality that both endpoints have finite entropy. If $\theta = 0$ or $\theta = 1$ the inequality is trivial as well. Fix then an arbitrary parameter $\theta \in (0,1)$ and let

$$H_\theta(t) := \begin{cases} \dfrac{1}{\theta}t & \text{if } t \in [0,\theta], \\ -\dfrac{1}{1-\theta}(t-1) & \text{if } t \in [\theta,1]. \end{cases}$$

be the hat function centered at $t = \theta$ with height 1 and vanishing at $t = 0, 1$. Setting $h(t) := \varepsilon H_\theta(t)$ for small $\varepsilon > 0$, let (γ_t^ε) be the curve constructed as in Lemma 3.6, i.e.

$$\gamma_t^\varepsilon := \mathsf{S}_{h(t)}\gamma_t, \qquad \text{for all } t \in [0,1].$$

Arguing as in the proof of Theorem 3.12, it is easily verified that with the current choice of h it is still true that $t \mapsto |\dot\gamma_t^\varepsilon|$ and $t \mapsto |\partial\mathsf{E}|(\gamma_t^\varepsilon)$ belong to $AC^2([0,1],\mathrm{X})$ and $t \mapsto \mathsf{E}(\gamma_t^\varepsilon)$ to $AC([0,1])$, so that we can integrate (3.18) in time on the whole interval $[0,1]$.

Discarding the non-negative term $|h'(t)|^2|\partial\mathsf{E}|^2(\gamma_t^\varepsilon)$ and using the optimality of the geodesic γ (namely its optimality between γ_0 and γ_1) give

$$\begin{aligned} 0 &\le \frac{1}{2}\int_0^1 |\dot\gamma_t^\varepsilon|^2\,\mathrm{d}t - \frac{1}{2}\int_0^1 |\dot\gamma_t|^2\,\mathrm{d}t \\ &\overset{(3.18)}{\le} -\int_0^1 h'(t)\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}(\gamma_t^\varepsilon)\,\mathrm{d}t + \frac{1}{2}\int_0^1 \Big(e^{-2\lambda h(t)} - 1\Big)|\dot\gamma_t|^2\,\mathrm{d}t \\ &= -\varepsilon\int_0^1 H_\theta'(t)\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}(\gamma_t^\varepsilon)\,\mathrm{d}t + \frac{\mathsf{d}^2(\gamma_0,\gamma_1)}{2}\int_0^1 \Big(e^{-2\varepsilon\lambda H_\theta(t)} - 1\Big)\,\mathrm{d}t, \end{aligned}$$

where the last equality follows from the constant speed property of the geodesic γ, namely $|\dot\gamma_t| = \mathsf{d}(\gamma_0,\gamma_1)$. Dividing by $\varepsilon > 0$ and leveraging the explicit piecewise constant values of $H_\theta'(t)$ on each interval $(0,\theta)$ and $(\theta,1)$ gives

$$\begin{aligned} 0 &\le -\int_0^1 H_\theta'(t)\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}(\gamma_t^\varepsilon)\,\mathrm{d}t + \frac{\mathsf{d}^2(\gamma_0,\gamma_1)}{2}\underbrace{\int_0^1 \frac{e^{-2\varepsilon\lambda H_\theta(t)} - 1}{\varepsilon}\,\mathrm{d}t}_{:=I_\varepsilon} \\ &= -\int_0^\theta \frac{1}{\theta}\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}(\gamma_t^\varepsilon)\,\mathrm{d}t + \int_\theta^1 \frac{1}{1-\theta}\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}(\gamma_t^\varepsilon)\,\mathrm{d}t + \frac{\mathsf{d}^2(\gamma_0,\gamma_1)}{2}I_\varepsilon \\ &= \frac{1}{\theta}\Big(\mathsf{E}(\gamma_0) - \mathsf{E}(\gamma_\theta^\varepsilon)\Big) + \frac{1}{1-\theta}\Big(\mathsf{E}(\gamma_1) - \mathsf{E}(\gamma_\theta^\varepsilon)\Big) + \frac{\mathsf{d}^2(\gamma_0,\gamma_1)}{2}I_\epsilon. \end{aligned}$$

Now let us multiply by $\theta(1-\theta) > 0$ and rearrange the terms in order to get

$$\mathsf{E}(\gamma_\theta^\varepsilon) \leq (1-\theta)\mathsf{E}(\gamma_0) + \theta\mathsf{E}(\gamma_1) + \theta(1-\theta)\frac{\mathsf{d}^2(\gamma_0,\gamma_1)}{2} I_\varepsilon.$$

It is easy to check that $\int_0^1 H_\theta(t)\mathrm{d}t = \frac{1}{2}$ for all θ, so that

$$\lim_{\varepsilon\downarrow 0} I_\varepsilon = -2\lambda \int_0^1 H_\theta(t)\,\mathrm{d}t = -\lambda.$$

On the other hand, by definition of γ^ε and since $h(\theta) = \varepsilon \to 0$ it is clear that $\gamma_\theta^\varepsilon = \mathsf{S}_{h(\theta)}\gamma_\theta = \mathsf{S}_\varepsilon\gamma_\theta \to \gamma_\theta$ in X (an EVI_λ-gradient flow is continuous up to $t = 0$). By lower semicontinuity of E this yields

$$\mathsf{E}(\gamma_\theta) \leq \liminf_{\varepsilon\downarrow 0} \mathsf{E}(\gamma_\theta^\varepsilon) \leq (1-\theta)\mathsf{E}(\gamma_0) + \theta\mathsf{E}(\gamma_1) - \frac{\lambda}{2}\theta(1-\theta)\mathsf{d}^2(\gamma_0,\gamma_1),$$

whence the conclusion. ■

5. Derivative of the cost

As a main application of the Γ-convergence results contained in Theorem 4.3 and Corollary 4.4 (and, in a wider sense, of their strategy of proof), in this section we investigate the dependence of the optimal value of the Schrödinger problem (Sch^ε) on the regularization parameter ε, focusing in particular on the regularity as a function of ε and on the behaviour in the small-time regime. More precisely, and denoting

$$\mathscr{C}_\varepsilon(x,y) := \inf_{(\gamma_t)\,:\,x\rightsquigarrow y} \Big\{\mathcal{A}(\gamma) + \varepsilon^2\mathcal{I}(\gamma)\Big\}, \qquad \forall\varepsilon \geq 0$$

the optimal entropic cost, we show that $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ is (locally) absolutely continuous and admits explicit left and right derivatives in a pointwise sense. Moreover, since $\mathscr{C}_\varepsilon(x,y) \to \mathscr{C}_0(x,y)$ as $\varepsilon \downarrow 0$ by Corollary 4.4, we aim at measuring the error $\mathscr{C}_\varepsilon(x,y) - \mathscr{C}_0(x,y)$ and studying the minimizers of the unperturbed problem $\mathscr{C}_0(x,y)$ selected by Γ-convergence. Since we focus here on the dependence on ε we will assume throughout the whole Section 5 and without further mention the well-posedness of the ε-Schrödinger problem:

Assumption 5.1. *Fix $x, y \in$ X and suppose that for some (hence for any, by Proposition 4.2) $\varepsilon > 0$ the Schrödinger problem* (Sch^ε) *admits at least one minimizer, in other words the infimum is attained in the definition of $\mathscr{C}_\varepsilon(x,y)$.*

We accordingly denote the set of ε-minimizers as

$$\Lambda_\varepsilon(x,y) := \Big\{\omega \in AC^2([0,1],\mathrm{X}) \,:\, \omega_0 = x,\, \omega_1 = y \quad \text{and} \quad \mathcal{A}_\varepsilon(\omega) = \mathscr{C}_\varepsilon(x,y)\Big\}.$$

Let us start the analysis with a preliminary monotonicity statement for the Fisher information and the entropic cost, which generalizes [25, Lemma 3.3].

Lemma 5.2. *With the same assumptions and notations as in Setting 3.1 and for any $0 \le \varepsilon_1 < \varepsilon_2 < \infty$ there holds*

$$\inf_{\Lambda_{\varepsilon_1}(x,y)} \mathcal{I} \ge \sup_{\Lambda_{\varepsilon_2}(x,y)} \mathcal{I},$$

with possibly $\inf_{\Lambda_0(x,y)} \mathcal{I} = +\infty$. Moreover, $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ is monotone non-decreasing on $[0,\infty)$.

Proof: Let $\varepsilon_1, \varepsilon_2$ as in the statement and choose $\omega^i \in \Lambda_{\varepsilon_i}(x,y)$ for $i = 1,2$, so that by optimality

$$\begin{aligned} \mathcal{A}(\omega^1) + \varepsilon_1^2 \mathcal{I}(\omega^1) &\le \mathcal{A}(\omega^2) + \varepsilon_1^2 \mathcal{I}(\omega^2), \\ \mathcal{A}(\omega^2) + \varepsilon_2^2 \mathcal{I}(\omega^2) &\le \mathcal{A}(\omega^1) + \varepsilon_2^2 \mathcal{I}(\omega^1). \end{aligned}$$

Summing these inequalities and dividing by $\varepsilon_2^2 - \varepsilon_1^2$ we obtain $\mathcal{I}(\omega^1) \ge \mathcal{I}(\omega^2)$, and since $\omega^1 \in \Lambda_{\varepsilon_1}$ and $\omega^2 \in \Lambda_{\varepsilon_2}$ are arbitrary the desired conclusion follows. As regards the last part of the statement, it is sufficient to note that since ω^i are minimizers of their respective problems and $\varepsilon_1 < \varepsilon_2$,

$$\mathscr{C}_{\varepsilon_1}(x,y) = \mathcal{A}(\omega^1) + \varepsilon_1^2 \mathcal{I}(\omega^1) \le \mathcal{A}(\omega^2) + \varepsilon_1^2 \mathcal{I}(\omega^2) \le \mathcal{A}(\omega^2) + \varepsilon_2^2 \mathcal{I}(\omega^2) = \mathscr{C}_{\varepsilon_2}(x,y).$$

■

Let us then extend Theorem 4.3 and Corollary 4.4 from $\varepsilon = 0$ to any $\varepsilon \ge 0$.

Proposition 5.3. *With the same assumptions and notations as in Setting 3.1 and under the additional Assumption 3.2, for any $\varepsilon > 0$ there holds*

$$\Gamma - \lim_{\varepsilon' \to \varepsilon} \Big\{\mathcal{A}_{\varepsilon'} + \iota_{01}\Big\} = \mathcal{A}_\varepsilon + \iota_{01} \tag{5.1}$$

for the pointwise-in-time σ-topology and

$$\lim_{\varepsilon' \to \varepsilon} \mathscr{C}_{\varepsilon'}(x,y) = \mathscr{C}_\varepsilon(x,y).$$

Moreover, for any $\varepsilon_k \to \varepsilon$ and any minimizer $\omega^k \in \Lambda_{\varepsilon_k}(x,y)$, there exists $\omega \in \Lambda_\varepsilon(x,y)$ such that, up to a subsequence,

$$\omega_t^k \xrightarrow{\sigma} \omega_t, \qquad \forall t \in [0,1]$$

as $k \to \infty$.

Proof: It is sufficient to prove (5.1), as the other properties follow by a verbatim application of the arguments in the proof of Corollary 4.4.

Fix ε and take $\varepsilon' \to \varepsilon$. The $\Gamma - \limsup$ inequality is trivial: if γ^ε is such that the right-hand side of (5.1) is finite (otherwise there is nothing to prove), then the constant sequence $\gamma^{\varepsilon'} \equiv \gamma^\varepsilon$ is an admissible recovery sequence. For the $\Gamma - \liminf$ inequality, note that the kinetic action $\mathcal{A}$ and the Fisher information $\mathcal{I}$ are lower semicontinuous w.r.t. pointwise-in-time σ-convergence (see the proof of Proposition 4.2), and clearly so is the convex indicator. Hence for any $\gamma^{\varepsilon'}$ converging to γ^ε for the pointwise-in-time σ-topology it holds

$$\begin{aligned}
\mathcal{A}_\varepsilon(\gamma^\varepsilon) + \iota_{01}(\gamma^\varepsilon) &\leq \liminf_{\varepsilon' \to \varepsilon} \left\{ \mathcal{A}_\varepsilon(\gamma^{\varepsilon'}) + \iota_{01}(\gamma^{\varepsilon'}) \right\} \\
&= \liminf_{\varepsilon' \to \varepsilon} \left\{ \mathcal{A}(\gamma^{\varepsilon'}) + \varepsilon^2 \mathcal{I}(\gamma^{\varepsilon'}) + \iota_{01}(\gamma^{\varepsilon'}) \right\} \\
&= \liminf_{\varepsilon' \to \varepsilon} \left\{ \mathcal{A}(\gamma^{\varepsilon'}) + (\varepsilon')^2 \mathcal{I}(\gamma^{\varepsilon'}) + \iota_{01}(\gamma^{\varepsilon'}) \right\} \\
&= \liminf_{\varepsilon' \to \varepsilon} \left\{ \mathcal{A}_{\varepsilon'}(\gamma^{\varepsilon'}) + \iota_{01}(\gamma^{\varepsilon'}) \right\}.
\end{aligned}$$

■

As an immediate consequence of this result we deduce the following

Lemma 5.4. *With the same assumptions and notations as in Setting 3.1 and under Assumption 3.2, the function $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ is continuous on $[0,\infty)$.*

Moreover, if $\varepsilon \mapsto \omega^\varepsilon$ is a continuous (w.r.t. the pointwise-in-time σ-topology) selection of minimizers, then $\varepsilon \mapsto \mathcal{A}(\omega^\varepsilon)$ and $\varepsilon \mapsto \mathcal{I}(\omega^\varepsilon)$ are also continuous, on $[0,\infty)$ and $(0,\infty)$ respectively.

Note that if the minimizers are unique, then $\varepsilon \mapsto \omega^\varepsilon$ is automatically continuous w.r.t. the pointwise-in-time σ-topology, simply by Proposition 5.3, as any sequence of minimizers admits a subsequence converging to a minimizer and the limit is in fact unique. Also, the continuity of the Fisher information can be strengthened up to $\varepsilon = 0$, see later on Theorem 5.7.

Proof: The continuity of $\mathscr{C}_\varepsilon(x,y)$ for $\varepsilon > 0$ is granted by Proposition 5.3, while continuity at $\varepsilon = 0$ has already been proved in Corollary 4.4.

As regards the kinetic energy $\mathcal{A}$ and the Fisher information $\mathcal{I}$, recall that they are both lower semicontinuous in $[0,\infty)$ w.r.t. the pointwise-in-time σ-topology, as already discussed in the proof of Proposition 4.2. Thus, if $\varepsilon \mapsto \omega^\varepsilon$

is as in the statement, we are left to prove that $\varepsilon \mapsto \mathcal{A}(\omega^\varepsilon)$ and $\varepsilon \mapsto \mathcal{I}(\omega^\varepsilon)$ are upper semicontinuous. To this aim, it is sufficient to observe that

$$\begin{aligned}\limsup_{\varepsilon'\to\varepsilon} \mathcal{A}(\omega^{\varepsilon'}) &= \limsup_{\varepsilon'\to\varepsilon} \Big\{\mathscr{C}_{\varepsilon'}(x,y) - (\varepsilon')^2\mathcal{I}(\omega^{\varepsilon'})\Big\} \\ &\leq \limsup_{\varepsilon'\to\varepsilon} \mathscr{C}_{\varepsilon'}(x,y) - \liminf_{\varepsilon'\to\varepsilon}(\varepsilon')^2\mathcal{I}(\omega^{\varepsilon'}) \\ &\leq \mathscr{C}_\varepsilon(x,y) - \varepsilon^2\mathcal{I}(\omega^\varepsilon) = \mathcal{A}(\omega^\varepsilon),\end{aligned}$$

where the last inequality holds by the continuity of $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ and the lower semicontinuity of $\varepsilon \mapsto \mathcal{I}(\omega^\varepsilon)$. Thus $\varepsilon \mapsto \mathcal{A}(\omega^\varepsilon)$ is upper semicontinuous in $[0,\infty)$. Interchanging $\mathcal{A}$ and $\mathcal{I}$ and writing now $\mathcal{I} = \frac{1}{\varepsilon^2}(\mathscr{C}_\varepsilon - \mathcal{A})$, the same argument shows that $\varepsilon \mapsto \mathcal{I}(\omega^\varepsilon)$ is upper semicontinuous in $(0,\infty)$ (continuity at $\varepsilon = 0$ will require a special treatment later). ■

We have now all the ingredients to discuss the regularity of the cost $\mathscr{C}_\varepsilon(x,y)$ as a function of the noise parameter ε and explicitly compute its left and right derivatives.

Proposition 5.5. *With the same assumptions and notations as in Setting 3.1 and if Assumption 3.2 holds, the map $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ is $AC_{loc}([0,\infty))$, left and right differentiable everywhere in $(0,\infty)$ and, for any $\varepsilon > 0$, the left and right derivatives are given by*

$$\frac{\mathrm{d}^-}{\mathrm{d}\varepsilon}\mathscr{C}_\varepsilon(x,y) = 2\varepsilon \max_{\Lambda_\varepsilon(x,y)} \mathcal{I}, \qquad \frac{\mathrm{d}^+}{\mathrm{d}\varepsilon}\mathscr{C}_\varepsilon(x,y) = 2\varepsilon \min_{\Lambda_\varepsilon(x,y)} \mathcal{I} \tag{5.2}$$

respectively, and the former (resp. latter) is left (resp. right) continuous. It is part of the statement the fact that the maximum and the minimum are attained.

Remark 5.6. Heuristically, (5.2) is nothing but the envelope theorem. Indeed, if $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ were differentiable, then its derivative would be given by $\partial_\varepsilon \mathcal{A}_\varepsilon = 2\varepsilon\mathcal{I}$ evaluated at any critical point, i.e. at any $\omega^\varepsilon \in \Lambda_\varepsilon(x,y)$. However, since we do not know in our general metric framework that Schrödinger problem has a unique solution, we are not able to prove pointwise differentiability as in [25] and we have to face the possibility of a gap between the left and right derivatives. In any case, for a.e. $\varepsilon > 0$ this gap is zero, because $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ is locally absolutely continuous in $(0,\infty)$ and therefore a.e. differentiable. This means that, up to a negligible set of temperatures, the left and right derivatives match and $\mathcal{I}$ is constant on $\Lambda_\varepsilon(x,y)$. If for whatever reason the Schrödinger problem (Sch^ε) were uniquely solvable (which is in particular true for the classic Schrödinger problem, as proved in [39, Theorem 4.2]), then the left and right derivatives would be trivially equal and Lemma 5.4 would give that $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ is actually $C^1((0,\infty))$. Furthermore, the cost would also be twice differentiable a.e. since by Lemma 5.2 its first derivative $2\varepsilon\mathcal{I}(\omega^\varepsilon)$ would be the product of a linear function and of a monotone one. ■

Proof: The continuity of $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ follows by Lemma 5.4, so let us focus on left and right differentiability/continuity and local absolute continuity.

Right differentiability. Fix $\varepsilon > 0$, let $\delta > 0$, and choose $\omega^\varepsilon \in \Lambda_\varepsilon(x,y)$, $\omega^{\varepsilon+\delta} \in \Lambda_{\varepsilon+\delta}(x,y)$. Then write

$$\begin{aligned}\frac{\mathscr{C}_{\varepsilon+\delta}(x,y)-\mathscr{C}_\varepsilon(x,y)}{\delta} &= \frac{\mathcal{A}_{\varepsilon+\delta}(\omega^{\varepsilon+\delta})-\mathcal{A}_\varepsilon(\omega^\varepsilon)}{\delta}\\ &= \frac{\mathcal{A}_{\varepsilon+\delta}(\omega^{\varepsilon+\delta})-\mathcal{A}_{\varepsilon+\delta}(\omega^\varepsilon)}{\delta}+\frac{\mathcal{A}_{\varepsilon+\delta}(\omega^\varepsilon)-\mathcal{A}_\varepsilon(\omega^\varepsilon)}{\delta}\end{aligned} \tag{5.3}$$

and note that the second term on the right-hand side can be rewritten as

$$\mathcal{A}_{\varepsilon+\delta}(\omega^\varepsilon)-\mathcal{A}_\varepsilon(\omega^\varepsilon) = (2\varepsilon\delta+\delta^2)\mathcal{I}(\omega^\varepsilon).$$

The first one is non-positive by optimality of $\omega^{\varepsilon+\delta}$ for $\mathcal{A}_{\varepsilon+\delta}$, hence we obtain

$$\limsup_{\delta\downarrow 0}\frac{\mathscr{C}_{\varepsilon+\delta}(x,y)-\mathscr{C}_\varepsilon(x,y)}{\delta} \leq \limsup_{\delta\downarrow 0}\,(2\varepsilon+\delta)\mathcal{I}(\omega^\varepsilon) = 2\varepsilon\mathcal{I}(\omega^\varepsilon).$$

As this inequality holds for any $\omega^\varepsilon \in \Lambda_\varepsilon(x,y)$, we infer that

$$\limsup_{\delta\downarrow 0}\frac{\mathscr{C}_{\varepsilon+\delta}(x,y)-\mathscr{C}_\varepsilon(x,y)}{\delta} \leq 2\varepsilon \inf_{\Lambda_\varepsilon(x,y)}\mathcal{I}. \tag{5.4}$$

On the other hand we can also write

$$\begin{aligned}\frac{\mathscr{C}_{\varepsilon+\delta}(x,y)-\mathscr{C}_\varepsilon(x,y)}{\delta} &= \frac{\mathcal{A}_{\varepsilon+\delta}(\omega^{\varepsilon+\delta})-\mathcal{A}_\varepsilon(\omega^\varepsilon)}{\delta}\\ &= \frac{\mathcal{A}_{\varepsilon+\delta}(\omega^{\varepsilon+\delta})-\mathcal{A}_\varepsilon(\omega^{\varepsilon+\delta})}{\delta}+\frac{\mathcal{A}_\varepsilon(\omega^{\varepsilon+\delta})-\mathcal{A}_\varepsilon(\omega^\varepsilon)}{\delta}.\end{aligned} \tag{5.5}$$

Using now the optimality of ω^ε for $\mathcal{A}_\varepsilon$, we observe that the second term on the right-hand side is non-negative, whence

$$\frac{\mathcal{A}_{\varepsilon+\delta}(\omega^{\varepsilon+\delta})-\mathcal{A}_\varepsilon(\omega^\varepsilon)}{\delta} \geq \frac{\mathcal{A}_{\varepsilon+\delta}(\omega^{\varepsilon+\delta})-\mathcal{A}_\varepsilon(\omega^{\varepsilon+\delta})}{\delta} = (2\varepsilon+\delta)\mathcal{I}(\omega^{\varepsilon+\delta}).$$

For any sequence $\delta_n \downarrow 0$, Proposition 5.3 guarantees (up to extraction of a subsequence if needed) that $\omega^{\varepsilon+\delta_n} \to \overline{\omega}^\varepsilon$ in the pointwise-in-time σ-topology for some $\overline{\omega}^\varepsilon \in \Lambda_\varepsilon(x,y)$. By lower semicontinuity of $\mathcal{I}$ this implies

$$\liminf_{n\to\infty} \frac{\mathscr{C}_{\varepsilon+\delta_n}(x,y) - \mathscr{C}_\varepsilon(x,y)}{\delta_n} \geq \liminf_{n\to\infty} (2\varepsilon+\delta_n)\mathcal{I}(\omega^{\varepsilon+\delta_n}) \geq 2\varepsilon \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I},$$

and together with (5.4) this yields

$$\exists \lim_{n\to\infty} \frac{\mathscr{C}_{\varepsilon+\delta_n}(x,y) - \mathscr{C}_\varepsilon(x,y)}{\delta_n} = 2\varepsilon\mathcal{I}(\overline{\omega}^\varepsilon) = 2\varepsilon \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I}.$$

As the right-hand side does not depend on the particular sequence $\delta_n \downarrow 0$ we conclude that

$$\exists \lim_{\delta\downarrow 0} \frac{\mathscr{C}_{\varepsilon+\delta}(x,y) - \mathscr{C}_\varepsilon(x,y)}{\delta} = 2\varepsilon \min_{\Lambda_\varepsilon(x,y)} \mathcal{I},$$

in particular $\mathcal{I}$ is minimized by any accumulation point $\overline{\omega}^\varepsilon$ of $\{\omega^{\varepsilon+\delta}\}_{\delta>0}$.
Left differentiability. The argument is very similar. Indeed, if $\delta < 0$, then the first term on the right-hand side of (5.3) is non-negative and the second one can be handled in the same way. Hence there holds

$$\liminf_{\delta\uparrow 0} \frac{\mathscr{C}_{\varepsilon+\delta}(x,y) - \mathscr{C}_\varepsilon(x,y)}{\delta} \geq 2\varepsilon\mathcal{I}(\omega^\varepsilon),$$

for any $\omega^\varepsilon \in \Lambda_\varepsilon(x,y)$, and therefore

$$\liminf_{\delta\uparrow 0} \frac{\mathscr{C}_{\varepsilon+\delta}(x,y) - \mathscr{C}_\varepsilon(x,y)}{\delta} \geq 2\varepsilon \sup_{\Lambda_\varepsilon(x,y)} \mathcal{I}.$$

Applying the same considerations to (5.5) and following the same argument as above we retrieve the lim sup inequality, first along *some* subsequence $\delta_n \uparrow 0$ and then along *any* $\delta \uparrow 0$. Combining with the inequality above gives

$$\exists \lim_{\delta\uparrow 0} \frac{\mathscr{C}_{\varepsilon+\delta}(x,y) - \mathscr{C}_\varepsilon(x,y)}{\delta} = 2\varepsilon \max_{\Lambda_\varepsilon(x,y)} \mathcal{I}, \qquad \forall \varepsilon > 0,$$

whence the pointwise left differentiability of $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$.
Left and right continuity. In order to prove the right continuity of the right derivative of $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$, note that on the one hand by Lemma 5.2 for any $\varepsilon_n \downarrow \varepsilon$ it holds

$$\inf_{\Lambda_\varepsilon(x,y)} \mathcal{I} \geq \limsup_{n\to\infty} \sup_{\Lambda_{\varepsilon_n}(x,y)} \mathcal{I} \geq \limsup_{n\to\infty} \inf_{\Lambda_{\varepsilon_n}(x,y)} \mathcal{I}.$$

On the other hand, we can assume up to a subsequence if needed that

$$\liminf_{n\to\infty} \inf_{\Lambda_{\varepsilon_n}(x,y)} \mathcal{I} = \lim_{n\to\infty} \inf_{\Lambda_{\varepsilon_n}(x,y)} \mathcal{I}.$$

As shown in the proof of right differentiability, $\inf_{\Lambda_{\varepsilon'}(x,y)} \mathcal{I}$ is attained for any $\varepsilon' > 0$, hence in particular $\inf_{\Lambda_{\varepsilon_n}(x,y)} \mathcal{I} = \mathcal{I}(\omega^n)$ for some $\omega^n \in \Lambda_{\varepsilon_n}(x,y)$, for all n. Up to extracting a further subsequence, by Proposition 5.3 we can assume that $\omega^n \to \overline{\omega}^\varepsilon$ w.r.t. the pointwise-in-time σ-topology for some $\overline{\omega}^\varepsilon \in \Lambda_\varepsilon(x,y)$, and moreover by Lemma 5.4

$$\lim_{n\to\infty} \mathcal{I}(\omega^n) = \mathcal{I}(\overline{\omega}) \geq \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I}.$$

Putting all these inequalities together provides us with the right continuity of $\varepsilon \mapsto \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I}$ and, a fortiori, of the right derivative. Left continuity for the left derivative follows along an analogous reasoning.

Local absolute continuity. Let $0 < \varepsilon_1 < \varepsilon_2 < \infty$ and, for any $0 < \delta < 1$, define

$$f_\delta(\varepsilon) := \frac{\mathscr{C}_{\varepsilon+\delta}(x,y) - \mathscr{C}_\varepsilon(x,y)}{\delta}.$$

The monotonicity of $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ from Lemma 5.2 gives $f_\delta \geq 0$. Arguing as in the very beginning of the proof of the right differentiability we see that $f_\delta(\varepsilon) \leq (2\varepsilon+1)\mathcal{I}(\omega^\varepsilon)$ for any $\omega^\varepsilon \in \Lambda_\varepsilon(x,y)$, and by Lemma 5.2

$$f_\delta(\varepsilon) \leq (2\varepsilon_2 + 1) \sup_{\Lambda_{\varepsilon_1}(x,y)} \mathcal{I} < \infty, \qquad \forall \varepsilon \in (\varepsilon_1, \varepsilon_2].$$

Hence $|f_\delta| \leq M$ uniformly in δ and f_δ converges pointwise to the right derivative of $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ as $\delta \downarrow 0$, whence by the dominated convergence theorem

$$\int_{\varepsilon_1}^{\varepsilon_2} \frac{\mathrm{d}^+}{\mathrm{d}\varepsilon} \mathscr{C}_\varepsilon(x,y)\,\mathrm{d}\varepsilon = \lim_{\delta\downarrow 0} \int_{\varepsilon_1}^{\varepsilon_2} f_\delta(\varepsilon)\,\mathrm{d}\varepsilon.$$

The right-hand side can be rewritten as

$$\begin{aligned}
\lim_{\delta\downarrow 0} \int_{\varepsilon_1}^{\varepsilon_2} f_\delta(\varepsilon)\,\mathrm{d}\varepsilon &= \lim_{\delta\downarrow 0} \Big(\frac{1}{\delta} \int_{\varepsilon_1}^{\varepsilon_2} \mathscr{C}_{\varepsilon+\delta}(x,y)\,\mathrm{d}\varepsilon - \frac{1}{\delta} \int_{\varepsilon_1}^{\varepsilon_2} \mathscr{C}_\varepsilon(x,y)\,\mathrm{d}\varepsilon \Big) \\
&= \lim_{\delta\downarrow 0} \Big(\frac{1}{\delta} \int_{\varepsilon_2}^{\varepsilon_2+\delta} \mathscr{C}_\varepsilon(x,y)\,\mathrm{d}\varepsilon - \frac{1}{\delta} \int_{\varepsilon_1}^{\varepsilon_1+\delta} \mathscr{C}_\varepsilon(x,y)\,\mathrm{d}\varepsilon \Big) \\
&= \mathscr{C}_{\varepsilon_2}(x,y) - \mathscr{C}_{\varepsilon_1}(x,y),
\end{aligned}$$

where the last equality holds by the Lebesgue differentiation theorem for the continuous function $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ (cf. Lemma 5.4). We have thus proved that the cost belongs to $AC_{loc}((0,\infty))$, since

$$\mathscr{C}_{\varepsilon_2}(x,y) - \mathscr{C}_{\varepsilon_1}(x,y) = \int_{\varepsilon_1}^{\varepsilon_2} \frac{\mathrm{d}^+}{\mathrm{d}\varepsilon}\mathscr{C}_\varepsilon(x,y)\,\mathrm{d}\varepsilon, \qquad \forall\, 0 < \varepsilon_1 < \varepsilon_2.$$

For the full $AC_{loc}([0,\infty))$ regularity it is then sufficient to let $\varepsilon_1 \downarrow 0$: the left-hand side converges to $\mathscr{C}_{\varepsilon_2}(x,y) - \mathscr{C}_0(x,y)$ by Lemma 5.4, and by the monotonicity $\frac{\mathrm{d}^+}{\mathrm{d}\varepsilon}\mathscr{C}_\varepsilon \geq 0$ the right-hand side also converges by monotone convergence. ■

Relying on our previous auxiliary results and on Proposition 5.5, we are finally in position of estimating the error $\mathscr{C}_\varepsilon(x,y) - \mathscr{C}_0(x,y)$ with $o(\varepsilon^2)$ precision. We will also significantly refine Corollary 4.4 by proving that any accumulation point of any sequence of minimizers is not only optimal for the unperturbed problem $\mathscr{C}_0(x,y)$, but also $\mathcal{I}$-minimizing among all competitors in $\Lambda_0(x,y)$.

Theorem 5.7. *With the same assumptions and notations as in Proposition 5.5, if there exists $\omega^0 \in \Lambda_0(x,y)$ such that $\mathcal{I}(\omega^0) < \infty$, then the map $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ is right differentiable also at $\varepsilon = 0$ with*

$$\frac{\mathrm{d}^+}{\mathrm{d}\varepsilon}\mathscr{C}_\varepsilon(x,y)\Big|_{\varepsilon=0} = 0,$$

the right derivative is right continuous for any $\varepsilon \geq 0$, and

$$\mathscr{C}_\varepsilon(x,y) - \mathscr{C}_0(x,y) = \varepsilon^2 \inf_{\Lambda_0(x,y)} \mathcal{I} + o(\varepsilon^2). \tag{5.6}$$

Moreover, for any $\varepsilon_n \downarrow 0$ and any minimizer $\omega^n \in \Lambda_{\varepsilon_n}(x,y)$ there exists $\omega^* \in \Lambda_0(x,y)$ such that (up to a subsequence) $\omega^n \to \omega^*$ for the pointwise-in-time σ-topology, and ω^* has minimal Fisher information in $\Lambda_0(x,y)$

$$\mathcal{I}(\omega^*) = \min_{\Lambda_0(x,y)} \mathcal{I}.$$

Proof: The right differentiability of $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ at $\varepsilon = 0$ follows by the same argument carried out in Proposition 5.5. Indeed, given ω^0 as in the statement, by (5.3) with $\varepsilon = 0$ it holds

$$\limsup_{\delta\downarrow 0} \frac{\mathscr{C}_\delta(x,y) - \mathscr{C}_0(x,y)}{\delta} \leq \limsup_{\delta\downarrow 0} \delta\mathcal{I}(\omega^0) = 0.$$

The liminf inequality is straightforward, since $\mathcal{I} \geq 0$ and thus by (5.5) with $\varepsilon = 0$

$$\liminf_{\delta \downarrow 0} \frac{\mathscr{C}_\delta(x,y) - \mathscr{C}_0(x,y)}{\delta} \geq \liminf_{\delta \downarrow 0} \delta \mathcal{I}(\omega^\delta) \geq 0$$

for any $\omega^\delta \in \Lambda_\delta(x,y)$. This also shows that the right derivative vanishes at $\varepsilon = 0$.

As regards the right continuity of the right derivative, the case $\varepsilon > 0$ has already been discussed in Proposition 5.5. For $\varepsilon = 0$ the same strategy still works, with the only minor difference that we cannot rely on Lemma 5.4 anymore. Nonetheless, if $\omega^n \in \Lambda_{\varepsilon_n}(x,y)$ is as in Proposition 5.5, $\overline{\omega} \in \Lambda_0(x,y)$ and $\omega^n \to \overline{\omega}$ for the pointwise-in-time σ-topology (the existence of such $\overline{\omega}$ is granted by Corollary 4.4) it is still true that

$$\liminf_{n\to\infty} \mathcal{I}(\omega^n) \geq \mathcal{I}(\overline{\omega}),$$

simply by lower semicontinuity of $\mathcal{I}$. With this single change in the proof we deduce that $\varepsilon \mapsto \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I}$ is right continuous and finite also at $\varepsilon = 0$, thanks to the present assumptions, and so is the right derivative of the cost due to $\varepsilon \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I} \to 0$ as $\varepsilon \downarrow 0$.

The last part of the statement is a slight modification of these lines of thought. Indeed, given any sequence $\varepsilon_n \downarrow 0$ and $\omega^n \in \Lambda_{\varepsilon_n}(x,y)$, the existence of $\omega^* \in \Lambda_0(x,y)$ such that, up to subsequences, $\omega^n \to \omega^*$ is ensured by Corollary 4.4. The fact that ω^* has minimal Fisher information among all elements in $\Lambda_0(x,y)$ follows from

$$\inf_{\Lambda_0(x,y)} \mathcal{I} \geq \limsup_{n\to\infty} \sup_{\Lambda_{\varepsilon_n}(x,y)} \mathcal{I} \geq \limsup_{n\to\infty} \mathcal{I}(\omega^n) \geq \mathcal{I}(\omega^*) \geq \inf_{\Lambda_0(x,y)} \mathcal{I},$$

where we used once again Lemma 5.2 and the lower semicontinuity of $\mathcal{I}$. Thus, it only remains to establish (5.6).

As $\varepsilon \mapsto \mathscr{C}_\varepsilon(x,y)$ belongs to $AC_{loc}([0,\infty))$ and the right derivative coincides a.e. with the full derivative, (5.2) and the fundamental theorem of calculus yield

$$\mathscr{C}_\varepsilon(x,y) - \mathscr{C}_0(x,y) = 2\int_0^\varepsilon s \inf_{\Lambda_s(x,y)} \mathcal{I}\,\mathrm{d}s \leq 2\int_0^\varepsilon s \inf_{\Lambda_0(x,y)} \mathcal{I}\,\mathrm{d}s = \varepsilon^2 \inf_{\Lambda_0(x,y)} \mathcal{I}.$$

Here we used the monotonicity of the Fisher information from Lemma 5.2 in the middle inequality. By the same monotonicity and the right continuity at

$\varepsilon = 0$ of $\varepsilon \mapsto \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I}$ we also deduce that

$$\begin{aligned}\mathscr{C}_\varepsilon(x,y) - \mathscr{C}_0(x,y) &\geq 2\int_0^\varepsilon s \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I}\,\mathrm{d}s = \varepsilon^2 \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I} \\ &= \varepsilon^2 \inf_{\Lambda_0(x,y)} \mathcal{I} + \varepsilon^2\Big(\inf_{\Lambda_\varepsilon(x,y)} \mathcal{I} - \inf_{\Lambda_0(x,y)} \mathcal{I}\Big). \\ &= \varepsilon^2 \inf_{\Lambda_0(x,y)} \mathcal{I} + o(\varepsilon^2)\end{aligned}$$

Combining this lower bound with the previous upper one entails (5.6). ■

Remark 5.8. It is worth stressing that the upper bound on $\mathscr{C}_\varepsilon(x,y) - \mathscr{C}_0(x,y)$ is not asymptotic, but pointwise. A possible way to improve (5.6) would rely on a refined analysis of $\varepsilon \mapsto \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I}$, its derivative (which exists a.e. by monotonicity), and possibly absolute continuity. ■

Remark 5.9. The $\mathcal{I}$-minimizing property of the accumulation point ω^* is not specific of the choice $\varepsilon = 0$, but of the particular "backward" direction of the sequence $\varepsilon_n \downarrow$. Repeating the argument in the proof of Theorem 5.7 it is indeed not difficult to check that, given any $\varepsilon > 0$, a sequence $\varepsilon_n \downarrow \varepsilon$, and $\omega^n \in \Lambda_{\varepsilon_n}(x,y)$ there exists $\omega^\varepsilon \in \Lambda_\varepsilon(x,y)$ such that, up to a subsequence, $\omega^n \to \omega^\varepsilon$ for the pointwise-in-time σ-topology and

$$\mathcal{I}(\omega^\varepsilon) = \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I}.$$

In a symmetric fashion, a closer look into the proof of Proposition 5.5 suggests that an opposite behaviour appears in the "forward" direction. More precisely, if $\varepsilon_n \uparrow \varepsilon$ instead of $\varepsilon_n \downarrow \varepsilon$, then any accumulation point ω^ε of (ω^n) is such that

$$\mathcal{I}(\omega^\varepsilon) = \sup_{\Lambda_\varepsilon(x,y)} \mathcal{I}.$$

However, the "backward" direction and the case $\varepsilon = 0$ are usually more interesting, because of the connection with the unperturbed problem $\mathscr{C}_0(x,y)$ for which there might be multiple solutions even if the Schrödinger problem (Sch$^\varepsilon$) has a unique minimizer for all $\varepsilon > 0$. It is therefore natural to look for the (properties of the) solutions selected via Schrödinger regularization. ■

6. Examples

In this section we collect several and heterogeneous situations where our abstract approach (in particular Theorem 4.3, Corollary 4.4, and Theorem 5.7) applies. We shall also comment the novelty of the results thus obtained in comparison with the existing literature. In this perspective, it is worth discussing in more detail the role played by Assumption 3.2 so far, singling out when the sequential lower semicontinuity of $|\partial \mathsf{E}|$ w.r.t. σ is needed and when it is not:

- to show the existence of a solution to the Schrödinger problem (Sch$^\varepsilon$) (cf. Proposition 4.2) it is crucial, in order to apply the direct method of the calculus of variations;
- in Theorem 4.3 and Corollary 4.4 it is not used;
- unlike Corollary 4.4, in Proposition 5.3 it is needed for the Γ-liminf inequality and so is in Lemma 5.4;
- Proposition 5.5 relies on Proposition 5.3 and Lemma 5.4, hence it is implicitly used;
- in Theorem 5.7 the continuity of $\varepsilon \mapsto \inf_{\Lambda_\varepsilon(x,y)} \mathcal{I}$ at $\varepsilon = 0$ requires the lower semicontinuity of $|\partial \mathsf{E}|$ and also Proposition 5.5 is used in the proof of (5.6); hence the lower semicontinuity of $|\partial \mathsf{E}|$ is really needed.

This means that if one is able to show the solvability of the Schrödinger problem (Sch$^\varepsilon$) by means other than those used in Proposition 4.2, then Theorem 4.3 and Corollary 4.4 are still valid under the following weaker hypothesis.

Assumption 6.1. *There exists a Hausdorff topology σ on X such that d-bounded sequences contain σ-converging subsequences. Moreover, the distance d is sequentially lower semicontinuous w.r.t. σ.*

Theorem 5.7, instead, requires the full validity of Assumption 3.2.

6.1. The Boltzmann-Shannon relative entropy. As a first example, let us consider the Boltzmann-Shannon relative entropy on the Wasserstein space built over a (locally compact) $\mathrm{RCD}(K,\infty)$ space. To this end, let $(M, \mathsf{d}, \mathfrak{m})$ be a complete and separable locally compact length metric space endowed with a Radon measure and assume that it is an $\mathrm{RCD}(K,\infty)$ space [4] for some $K \in \mathbb{R}$. Let $\mathrm{X} := \mathcal{P}_2(M)$ be the 2-Wasserstein space over M, namely the space of probability measures with finite second moments, and equip it with the 2-Wasserstein distance W_2: it turns out to be a complete and separable metric space [14] as well. The Boltzmann-Shannon relative entropy E on X is defined as

$$\mathsf{E}(\mu) := \begin{cases} \displaystyle\int_M \rho \log(\rho)\, \mathrm{d}\mathfrak{m} & \text{if } \mu = \rho\mathfrak{m}, \\ +\infty & \text{if } \mu \not\ll \mathfrak{m}. \end{cases}$$

As by [61, Theorem 4.24] there exist $C > 0$, $x \in M$ such that $\int_M e^{-C\mathsf{d}^2(\cdot,x)}\mathrm{d}\mathfrak{m} < \infty$, E can be equivalently rewritten as

$$\mathsf{E}(\mu) = \underbrace{\int_M \tilde{\rho}\log(\tilde{\rho})\,\mathrm{d}\tilde{\mathfrak{m}}}_{\geq 0} - C\int_M \mathsf{d}^2(\cdot,x)\,\mathrm{d}\mu - \log Z,$$

where $\tilde{\rho}$ is the Radon-Nikodym derivative of μ w.r.t. $\tilde{\mathfrak{m}}$, with the normalization

$$Z := \int_M e^{-C\mathsf{d}^2(\cdot,x)}\mathrm{d}\mathfrak{m}, \qquad \tilde{\mathfrak{m}} := \frac{1}{Z}e^{-C\mathsf{d}^2(\cdot,x)}\mathfrak{m}.$$

From this very definition, it is easy to see that E is a proper lower semicontinuous functional, bounded from below on W_2-bounded sets. In addition, it has a dense domain, since by (one of the equivalent) definition of RCD spaces, cf. [4, Theorem 5.1], for any $\mu \in \mathrm{X}$ there exists an EVI_K-gradient flow of E starting from it. Thus Setting 3.1 holds.

As regards Assumption 6.1, note that X is not locally compact unless M is compact, so that in general the metric topology of X is not an admissible candidate for σ. Nonetheless there is a natural alternative: the narrow convergence of probability measures. Indeed, W_2-bounded sequences in X are uniformly tight (the second moments are uniformly bounded and the balls in M are relatively compact, so that the claim follows from [2, Remark 5.1.5]) and thus relatively compact w.r.t. the narrow topology. Moreover, W_2 is lower semicontinuous w.r.t. narrow convergence of measures [1, Proposition 3.5]. Therefore, given any $\mu, \nu \in \mathrm{X}$ for which the dynamical Schrödinger problem (Sch^ε) is solvable, the Γ-convergence results of Section 4 are fully applicable. This is for instance the case if $\mu, \nu \ll \mathfrak{m}$ have bounded densities and supports (in [38, 39] this is proved for $\mathrm{RCD}^*(K,N)$ spaces, $N < \infty$, but the argument can be adapted to locally compact $\mathrm{RCD}(K,\infty)$ spaces thanks to the existence of "good" cut-off functions [52]).

In the present framework, taking into account the equivalence between W_2-absolutely continuous curves and distributional solutions of the continuity equation (see [37]) and the fact that the slope $|\partial\mathsf{E}|^2$ coincides with the Fisher information [3, Theorem 9.3], (Sch^ε) reads as

$$\inf\left\{\frac{1}{2}\iint_0^1 |v_t|^2\rho_t\,\mathrm{d}t\mathrm{d}\mathfrak{m} + \frac{\varepsilon^2}{2}\iint_0^1 |\nabla\log\rho_t|^2\rho_t\,\mathrm{d}t\mathrm{d}\mathfrak{m}\right\},$$

where the infimum runs over all couples (μ_t, v_t), $\mu_t = \rho_t \mathfrak{m}$, solving the continuity equation $\partial\mu_t + \mathrm{div}(v_t\mu_t) = 0$ with the constraint $\mu_0 = \mu$ and $\mu_1 = \nu$. This is the dynamical formulation of the "classical" Schrödinger problem [45]. A thorough study of this problem and its equivalent formulations (at the static, dual, and dynamical levels) has been carried out by the second author in [39], but in the more restrictive framework of $\mathrm{RCD}^*(K,N)$ spaces, and only for ε fixed. The behaviour of the (unique) minimizers as $\varepsilon \downarrow 0$ is instead studied in [38, Proposition 5.1], again only in $\mathrm{RCD}^*(K,N)$ spaces, but the Γ-convergence of the corresponding variational problems is not investigated. Hence Theorem 4.3 and Corollary 4.4 are new in the RCD framework.

Under the stronger assumption that M is compact (e.g. the torus, the sphere or any convex closed bounded subset of a smooth weighted Riemannian manifold), as said above we can choose the σ topology to be the strong one induced by W_2, and in this case $|\partial\mathsf{E}|$ is lower semicontinuous by (3.1) and Assumption 3.2 is fully satisfied. Another interesting situation where Assumption 3.2 fully holds is represented by a convex domain in $\mathbb{R}^d$ (in this case σ is, as before, the narrow topology; see [35, Lemma 2.4] for a proof of the narrow lower semicontinuity of $|\partial\mathsf{E}|$). As a consequence, in these examples also the results in Section 5 hold true and this partly extends the recent work [25], where an analogue of Theorem 5.7 is proved in the Riemannian setting.

6.2. Internal energies and the Rényi entropy. As a second class of examples, we consider generalized entropy functionals (usually called *internal energies*) on the Wasserstein space built over an $\mathrm{RCD}^*(0,N)$ space, $N < \infty$. Taking advantage of the non-negative curvature assumption and of the finite dimensionality we shall indeed be able to cover a wide range of functionals, including in particular Rényi entropies (naturally linked to the porous medium equation). By the discussion carried out in the previous section and by the fact that $\mathrm{RCD}^*(K,N)$ spaces are in particular locally compact $\mathrm{RCD}(K,\infty)$ spaces (the notion of $\mathrm{RCD}^*(K,N)$ space is first introduced in [36]; for the distinction between RCD and RCD^* conditions see [6] and [21]), the 2-Wasserstein space $\mathrm{X} := \mathcal{P}_2(M)$ over an $\mathrm{RCD}^*(0,N)$ space $(M,\mathsf{d},\mathfrak{m})$ endowed with the 2-Wasserstein distance W_2 is a complete and separable metric space.

As regards the entropy functionals we shall consider on X, they are of the form $\int_M U(\rho)\,\mathrm{d}\mathfrak{m}$, where $U : [0,\infty) \to \mathbb{R}$ is a continuous and convex function with $U(0) = 0$ and U' locally Lipschitz in $(0,\infty)$ satisfying McCann's condition [49] for some $N' \geq N$: this means that the corresponding pressure function $P(r) :=$

$rU'(r) - U(r)$ is such that $P(0) := \lim_{r\downarrow 0} P(r) = 0$ and $r \mapsto r^{-1+1/N'}P(r)$ is non-decreasing or, equivalently,

$$r \mapsto r^{N'}U(r^{-N'}) \text{ is convex and non-increasing on } (0,\infty).$$

Under these assumptions on U, the internal energy E is defined as

$$\mathsf{E}(\mu) := \int_M U(\rho)\,\mathrm{d}\mathfrak{m} + U'(\infty)\mu^\perp(M), \qquad \text{if } \mu = \rho\mathfrak{m} + \mu^\perp,\ \mu^\perp \perp \mathfrak{m} \tag{6.1}$$

where $U'(\infty) := \lim_{r\to\infty} U'(r)$. In the case U is chosen equal to

$$U_{N'}(r) := -N'(r^{1-1/N'} - r),\ N' \geq N \quad \text{or} \quad U_m(r) := \frac{1}{m-1}r^m,\ m \geq 1 - \frac{1}{N}$$

($U_{N'}$ being more linked to Lott-Sturm-Villani theory of curvature-dimension bounds, U_m with the porous medium equation of power m), the well-known Rényi entropy is recovered. A detailed discussion of internal energies like E, associated non-linear diffusion semigroups and evolution variational inequalities in connection with curvature-dimension conditions is at the heart of the monograph [5] and can also be found in [63, Chapters 16 and 17].

Since $U(0) = 0$, M is locally compact and U is continuous, it is clear that E is well defined and finite on all probability measures with bounded support, so that E is proper and has a dense domain in X. Actually, $D(\mathsf{E})$ is dense in energy in X, i.e. for all $\mu \in$ X there exist $\mu_n \in D(\mathsf{E})$ with $W_2(\mu_n, \mu) \to 0$ and $\mathsf{E}(\mu_n) \to \mathsf{E}(\mu)$ as $n \to \infty$. By the properties of U it is also easy to see that E is lower semicontinuous [63, Theorem 30.6] and bounded from below on W_2-bounded sets. Finally, from [5, Theorem 9.21] with $K = 0$ (since M is assumed to be an $\mathrm{RCD}^*(0, N)$ space) and the fact that $D(\mathsf{E})$ is dense in energy in X, we see that for all $\mu \in$ X there exists an EVI_0-gradient flow of E starting from it.

We are therefore within Setting 3.1 and by what we said in Section 6.1 the narrow topology complies with Assumption 6.1. Hence whenever the dynamical Schrödinger problem (Sch^ε) is solvable, our metric results of Section 4 can be applied. As regards the lower semicontinuity of $|\partial\mathsf{E}|$ w.r.t. σ, and thus the full validity of Assumption 3.2 and, as a consequence, of the abstract results of Section 5 too, if M is compact then by (3.1) we see that the W_2-topology is an admissible candidate for σ. If $M = \mathbb{R}^d$ and U is superlinear at ∞ (which is the case for U_m defined above with $m > 1$), then by [2, Theorem 10.4.6] the slope can be represented as

$$|\partial \mathsf{E}|^2(\mu) = \int_{\mathbb{R}^d} |\nabla U'(\rho)|^2 \,\mathrm{d}\mu, \qquad \text{if } \mu = \rho\mathcal{L}^d$$

and by [35, Proposition 2.2] it is lower semicontnuous w.r.t. narrow convergence. Hence also in this situation the results of Section 5 hold true.

To the best of our knowledge, up to now the dynamical Schrödinger problem (Sch^ε) with the slope of a general internal energy in place of the slope of the Boltzmann entropy has been considered only in [33] from a purely formal point of view. Static Monge-Kantorovich problems regularized by means of the Rényi entropy or more general internal energies have recently been introduced in [32, 48, 47, 30] (see also the references therein). Remarkably, [47] establishes the Γ-convergence of the regularized problems towards the optimal transport one (cf. [30] where the convergence of the optimal values and minimizers is discussed). However, in [30] only bounded costs are considered (the quadratic cost function associated to (1.4) is thus ruled out for non-compact sample spaces), while [47] the discussion is restricted to sample spaces which are compact subset of $\mathbb{R}^d$. Other questions our paper is concerned with have not been examined in these references. Note also that the issue of the equivalence between static and dynamical formulations is far from being clear at this level of generality. In view of this discussion, in all the applicability situations presented in this section our results are new.

The case of a (possibly) negatively curved base space M is not discussed since, as already argued above, [5, Theorem 9.21] allows to deduce (EVI_λ) with $\lambda = 0$ only for $K \geq 0$. Moreover, it has recently been proved [28, Theorem 2.5 and Remark 2.6] that in the hyperbolic space the porous medium equation cannot be seen as the Wasserstein gradient flow of some λ-convex functional in the EVI-sense, hence the Rényi entropy cannot generate an EVI_λ-gradient flow there.

6.3. Mean-field Schrödinger problem. In the seminal thought experiment proposed by Schrödinger [59, 60] the physical system, whose evolution between two subsequent observations has to be determined, consists of independent Brownian particles. An important generalization has been recently proposed in [7], where particles are allowed to interact through a pair potential W. This leads to the so-called Mean Field Schrödinger Problem (MFSP henceforth). In order to see that this example falls within our abstract metric theory, let us first check the validity of Setting 3.1. As already said in Section 6.1, $\mathrm{X} := \mathcal{P}_2(\mathbb{R}^d)$, the 2-Wasserstein space over $\mathbb{R}^d$, endowed with the Wasserstein distace W_2 is a complete and separable metric space. The role played by the Boltzmann-Shannon relative entropy in the “classical” Schrödinger problem is here taken by the functional $\mathsf{E} : \mathrm{X} \to \mathbb{R}$ defined (up to a shift by a constant) by

$$\mathsf{E}(\mu) := \begin{cases} H(\mu \,|\, \mathcal{L}^d) + \displaystyle\int_{\mathbb{R}^d} W * \rho \, \mathrm{d}\mu & \text{if } \mu = \rho\mathcal{L}^d \\ +\infty & \text{if } \mu \not\ll \mathcal{L}^d \end{cases}$$

where $H(\mu \,|\, \mathcal{L}^d)$ is the Boltzmann-Shannon relative entropy of μ w.r.t. the Lebesgue measure $\mathcal{L}^d$, already introduced in Section 6.1, and W is the pair potential, describing via convolution the interaction between the particles of the system. On such a potential the following assumptions are made: it is of class $C^2(\mathbb{R}^d, \mathbb{R})$, is symmetric, i.e. $W(x) = W(-x)$ for all $x \in \mathbb{R}^d$, and satisfies the two-sided bound

$$\Lambda \mathrm{Id} \ \geq \nabla^2 W \geq \lambda \mathrm{Id}$$

for some $\Lambda, \lambda > 0$ (actually $\lambda \in \mathbb{R}$ is enough, but in [7] the authors are interested in the ergodic behaviour of MFSP). While the upper bound is technical, the lower one is geometric and crucial. The lower semicontinuity of E is easily seen to hold: the relative entropy has already been discussed, whereas the continuity of the convolution term follows from the fact that if $\mu_n \to \mu$ in $\mathcal{P}_2(\mathbb{R}^d)$, then $\mu_n \otimes \mu_n \to \mu \otimes \mu$ in $\mathcal{P}_2(\mathbb{R}^{2d})$, cf. [2, Example 9.3.4]. The fact that E is proper and the density of its domain are also clear. Moreover, the assumptions on W guarantee that E is bounded from below on W_2-bounded sets. As concerns the existence of EVI_λ-gradient flows starting from any $\mu \in \mathrm{X}$, this is ensured by [2, Theorem 11.2.1] in conjunction with [2, Remark 9.2.5 and Proposition 9.3.5], granting the λ-convexity of E along *generalized* geodesics (see [2, Definitions 9.2.2 and 9.2.4]). Therefore, Setting 3.1 holds.

As regards Assumption 3.2, for the topology σ the natural candidate is once again the sequential topology induced by narrow convergence of probability measures. By the discussion in Section 6.1, W_2-bounded sets are relatively narrow compact and W_2 is sequentially narrow lower semicontinuous. The slope of E is explicitly given by

$$|\partial \mathsf{E}|^2(\mu) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} |\nabla \log \rho + 2\nabla W * \rho|^2 \, \mathrm{d}\mu & \text{if } \mu = \rho\mathcal{L}^d,\ \nabla \log \rho \in L^2_\mu, \\ +\infty & \text{otherwise,} \end{cases}$$

cf. [7, Section 1.4.2], and one can rely on [35, Proposition 2.2], the fact that ΔW is continuous and bounded (as a consequence of the boundedness of $\nabla^2 W$) and the regularization properties of the convolution to show that $|\partial \mathsf{E}|$ is also sequentially narrowly lower semicontinuous. Hence Assumption 3.2 is fully satisfied and all the results of Sections 4 and 5 are applicable.

From the novelty standpoint, a first interesting remark is the fact that in [7] the approach is purely stochastic, while our point of view is completely analytic. For instance, in [7, Proposition 1.1] the existence of solutions to MFSP is proved under the same assumptions we have in Proposition 4.2, namely $\mu, \nu \in \mathrm{X}$ with $\mathsf{E}(\mu), \mathsf{E}(\nu) < \infty$. However, already at this basic level the reader may appreciate the difference between the two approaches.

But more than anything else, our abstract results are completely new when specialized to MFSP: indeed, only the ergodic behaviour in the long time regime $\varepsilon \to \infty$ is studied in [7], so that the Γ-convergence results of Section 4 are entirely novel. The same is true for Section 5, since in [25] the derivative of the cost associated to MFSP is not investigated nor is the Taylor expansion (5.6).

6.4. Non-linear mobilities. In [31] the authors studied a generalization W_{m} of the quadratic Wasserstein distance on $\mathcal{P}(\Omega)$, generated by Benamou-Brenier-type formulas in convex bounded domains $\Omega \subset \mathbb{R}^d$. The latter are based on nonlinear continuity equations and pseudo-Riemannian norms

$$\|\dot{\mu}\|^2_\mu = \min_v \left\{ \int_\Omega \mathsf{m}(\mu)|v|^2 \text{ s.t. } \dot{\mu} + \operatorname{div}(\mathsf{m}(\mu)v) = 0 \right\},$$

where $\mathsf{m} : \mathbb{R}^+ \to \mathbb{R}^+$ is a given *nonlinear mobility* function satisfying suitable structural conditions (mainly concavity). The linear case $\mathsf{m}(r) = r$ corresponds to the standard Wasserstein distance. In the nonlinear case, the lack of 1-homogeneity may impose restricting to absolutely continuous measures $\mu = \rho(x)\,\mathrm{d}x$ (as in [20, Section 3] depending on cases A or B therein), thus for the ease of exposition in this section we identify measures μ with their densities ρ with a slight abuse of notation.

Specifying the reference measure $\mathfrak{m}$ in (6.1) to be the (normalized) Lebesgue measure $\mathcal{L}^d$ and taking U to be superlinear $U(+\infty) = +\infty$ for convenience, let us check that our previous internal energies

$$\mathsf{E}(\rho) = \int_\Omega U(\rho(x))\,\mathrm{d}x$$

fit our setting. First of all, completeness is known from [31, Theorem 5.7], and separability readily follows from the density of compactly supported continuous functions, whence (A1). Next, from [31, Theorem 5.5] it is known that the W_{m} topology is at least stronger than the weak-$*$ convergence of measures, thus the lower semicontinuity in (A2) holds as soon as $z \mapsto U(z)$ is lower semicontinuous. The density of the domain $D(\mathsf{E})$ in (A2) should be again a simple exercise involving standard approximation arguments, provided U is reasonable. As regards our more fundamental assumption (A3), the generation of an EVI_λ-flow is exactly the purpose of [20] for $\lambda = 0$. More precisely, under some generalized McCann condition $GMC(\mathsf{m}, d)$ involving U, m, and the ambient dimension d, [20, Theorem 4.10] guarantees that our internal energy functionals generate 0-contractive gradient flows. As a consequence we can rigorously take our Setting 3.1 as applicable here. As concerns Assumption 3.2, a reasonable and natural choice for the weaker σ-topology is of course the weak-$*$ convergence of measures. Given the pseudo-Riemannian structure induced by [20, Eq. (3.2)], the (squared) metric slope is at least formally given by

$$|\partial \mathsf{E}|^2(\rho) = \int_\Omega \mathsf{m}(\rho)|\nabla U'(\rho)|^2 = \int_\Omega \frac{|\nabla P(\rho)|^2}{\mathsf{m}(\rho)}, \tag{6.2}$$

where the *pressure* P is defined as $P(r) = \int_0^r U''(z)\mathsf{m}(z)\,\mathrm{d}z$. With reasonable assumptions on U, m it should not be difficult to check the lower semicontinuity of this generalized Fisher information for this specific choice of the σ topology, and from [31, Theorem 5.6] the distance W_{m} is also known to be lower semicontinuous w.r.t. the weak-$*$ convergence.

As a consequence our metric results apply to this context as well (although full proofs of the representation (6.2) for the metric slope and of its weak lower semicontinuity are still missing for a completely rigorous statement, but this is out of scope of the paper).

6.5. Hadamard spaces. Let (X, d) be a complete and separable CAT(0) space (i.e. a separable Hadamard space), $x_0 \in \mathrm{X}$ be a fixed point, and $\mathsf{E} = \frac{1}{2}\mathsf{d}^2(x_0, \cdot)$. Then E is a 1-convex functional. By [54, Theorem 3.14], there exists an EVI_1-gradient flow of E starting from any $x \in \mathrm{X}$. Thus, we fit into Setting 3.1.

Although Assumption 6.1 is satisfied, existence of some "weak" Hausdorff topology σ on X is required to secure Assumption 3.2 (unless X is locally compact, cf. Remark 3.3). Such a topology indeed exists, at least if X satisfies a rather mild geometric $\overline{Q_4}$ condition [41, 42], and is called the half-space topology. The corresponding convergence is known as the Δ-convergence [46]. Indeed, d-bounded sequences contain Δ-converging subsequences [46, 41]. Bounded closed convex sets (in particular, balls) are Δ-closed [42], which easily implies that d is Δ-lower semicontinuous. Moreover, it is easy to see from (3.1) that $|\partial\mathsf{E}(x)| = \mathsf{d}(x_0, x)$, thus the slope is Δ-lower semicontinuous too.

In this framework, all our results are applicable. Note that E is always finite.

Acknowledgments. LM wishes to thank Jean-Claude Zambrini for numerous and fruitful discussions on the Schrödinger problem, and acknowledges support from the Portuguese Science Foundation through FCT project PTDC/MAT-STA/22812/2017 *SchröMoka.* LT acknowledges financial support from FSMP Fondation Sciences Mathématiques de Paris. DV was partially supported by the FCT projects UID/MAT/00324/2020 and PTDC/MAT-PUR/28686/2017.

References

[1] Luigi Ambrosio and Nicola Gigli. A user's guide to optimal transport. In *Modelling and Optimisation of Flows on Networks*, Lecture Notes in Mathematics, pages 1–155. Springer Berlin Heidelberg, 2013.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Invent. Math.*, 195(2):289–391, 2014.

[4] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Math. J.*, 163(7):1405–1490, 2014.

[5] Luigi Ambrosio, Andrea Mondino, and Giuseppe Savaré. Nonlinear diffusion equations and curvature conditions in metric measure spaces. *Mem. Amer. Math. Soc.*, 262(1270):v+121, 2019.

[6] Kathrin Bacher and Karl-Theodor Sturm. Localization and tensorization properties of the curvature-dimension condition for metric measure spaces. *J. Funct. Anal.*, 259(1):28–56, 2010.

[7] Julio Backhoff, Giovanni Conforti, Ivan Gentil, and Christian Léonard. The mean field Schrödinger problem: ergodic behavior, entropy estimates and functional inequalities. *Probability Theory and Related Fields*, pages 1–56, 2020.

[8] Aymeric Baradat and Léonard Monsaingeon. Small noise limit and convexity for generalized incompressible flows, Schrödinger problems, and optimal transport. *Arch. Ration. Mech. Anal.*, 235(2):1357–1403, 2020.

[9] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, Jan 2000.

[10] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.

[11] Jean-David Benamou, Guillaume Carlier, Simone Di Marino, and Luca Nenna. An entropy minimization approach to second-order variational mean-field games. *Preprint, arXiv:1807.09078*, 2018.

[12] Jean-David Benamou, Guillaume Carlier, and Luca Nenna. A numerical method to solve multi-marginal optimal transport problems with Coulomb cost. In *Splitting methods in communication, imaging, science, and engineering*, Sci. Comput., pages 577–601. Springer, Cham, 2016.

[13] Jean-David Benamou, Guillaume Carlier, and Luca Nenna. Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm. *Preprint, arXiv:1710.08234*, 2017.

[14] François Bolley. Separability and completeness for the Wasserstein distance. In *Séminaire de probabilités XLI*, volume 1934 of *Lecture Notes in Math.*, pages 371–377. Springer, Berlin, 2008.

[15] François Bolley and José A Carrillo. Nonlinear diffusion: geodesic convexity is equivalent to Wasserstein contraction. *Communications in Partial Differential Equations*, 39(10):1860–1869, 2014.

[16] Andrea Braides. *Gamma-convergence for Beginners*, volume 22. Clarendon Press, 2002.

[17] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[18] Eric Carlen. Stochastic mechanics: a look back and a look ahead. *Diffusion, quantum theory and radically elementary mathematics*, 47:117–139, 2014.

[19] Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.*, 49(2):1385–1418, 2017.

[20] José Antonio Carrillo, Stefano Lisini, Giuseppe Savaré, and Dejan Slepcev. Nonlinear mobility continuity equations and generalized displacement convexity. *Journal of Functional Analysis*, 258(4):1273–1309, 2010.

[21] Fabio Cavalletti and Emanuel Milman. The globalization theorem for the curvature dimension condition. Preprint, arXiv:1612.07623, 2016.

[22] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.

[23] Giovanni Conforti. A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost. *Probability Theory and Related Fields*, 174(1-2):1–47, 2019.

[24] Giovanni Conforti and Luigia Ripani. Around the entropic Talagrand inequality. *Bernoulli*, 26(2):1431–1452, 2020.

[25] Giovanni Conforti and Luca Tamanini. A formula for the time derivative of the entropic cost and applications. *Preprint, arXiv:1912.10555*, 2019.

[26] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.

[27] Sara Daneri and Giuseppe Savaré. Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Anal.*, 40(3):1104–1122, 2008.

[28] Nicolò De Ponti, Matteo Muratori, and Carlo Orrieri. Wasserstein stability of porous medium-type equations on manifolds with Ricci curvature bounded below. *Preprint, arXiv:1908.03147*, 2019.

[29] Simone Di Marino and Augusto Gerolin. An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm. *J. Sci. Comput.*, 85(2):Paper No. 27, 28, 2020.

[30] Simone Di Marino and Augusto Gerolin. Optimal Transport losses and Sinkhorn algorithm with general convex regularization. *Preprint, arXiv:2007.00976*, 2020.

[31] Jean Dolbeault, Bruno Nazaret, and Giuseppe Savaré. A new class of transport distances between measures. *Calculus of Variations and Partial Differential Equations*, 34(2):193–231, 2009.
[32] Montacer Essid and Justin Solomon. Quadratically regularized optimal transport on graphs. *SIAM J. Sci. Comput.*, 40(4):A1961–A1986, 2018.
[33] Ivan Gentil, Christian Léonard, and Luigia Ripani. Dynamical aspects of generalized Schrödinger problem via Otto calculus - A heuristic point of view. *Preprint, arXiv:1806.01553*, 2018.
[34] Ivan Gentil, Christian Léonard, Luigia Ripani, and Luca Tamanini. An entropic interpolation proof of the HWI inequality. *Stochastic Processes and their Applications*, 2019.
[35] Ugo Gianazza, Giuseppe Savaré, and Giuseppe Toscani. The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation. *Arch. Ration. Mech. Anal.*, 194(1):133–220, 2009.
[36] Nicola Gigli. On the differential structure of metric measure spaces and applications. *Mem. Amer. Math. Soc.*, 236(1113):vi+91, 2015.
[37] Nicola Gigli and Bangxian Han. The continuity equation on metric measure spaces. *Calc. Var. Partial Differential Equations*, 53(1-2):149–177, 2013.
[38] Nicola Gigli and Luca Tamanini. Second order differentiation formula on $RCD^*(K,N)$ spaces. Accepted at JEMS, arXiv:1802.02463, (2018).
[39] Nicola Gigli and Luca Tamanini. Benamou-Brenier and duality formulas for the entropic cost on $RCD^*(K,N)$ spaces. *Probability Theory and Related Fields*, 176(1-2):1–34, 2020.
[40] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.
[41] Bijan Kakavandi. Weak topologies in complete $CAT(0)$ metric spaces. *Proceedings of the American Mathematical Society*, 141(3):1029–1039, 2013.
[42] William Kirk and Naseer Shahzad. *Fixed point theory in distance spaces*. Springer, Cham, 2014.
[43] Flavien Léger. A geometric perspective on regularized optimal transport. *Journal of Dynamics and Differential Equations*, 31(4):1777–1791, 2019.
[44] Christian Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.*, 262(4):1879–1920, 2012.
[45] Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.
[46] Teck Cheong Lim. Remarks on some fixed point theorems. *Proc. Amer. Math. Soc.*, 60:179–182 (1977), 1976.
[47] Dirk Lorenz and Hinrich Mahler. Orlicz space regularization of continuous optimal transport problems. *Preprint, arXiv:2004.11574*, 2020.
[48] Dirk A. Lorenz, Paul Manns, and Christian Meyer. Quadratically regularized optimal transport. *Applied Math. Optimization, to appear*, 2019.
[49] Robert J McCann. A convexity principle for interacting gases. *Advances in mathematics*, 128(1):153–179, 1997.
[50] Toshio Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Related Fields*, 129(2):245–260, 2004.
[51] Toshio Mikami and Michèle Thieullen. Optimal transportation problem by stochastic optimal control. *SIAM J. Control Optim.*, 47(3):1127–1139, 2008.
[52] Andrea Mondino and Aaron Charles Naber. Structure theory of metric measure spaces with lower Ricci curvature bounds. *Journal of the European Mathematical Society*, 21(6):1809–1854, 2019.

[53] Léonard Monsaingeon and Dmitry Vorotnikov. The Schrödinger problem on the non-commutative Fisher-Rao space. *Calc. Var. Partial Differential Equations, to appear*, 2020.
[54] Matteo Muratori and Giuseppe Savaré. Gradient flows and Evolution Variational Inequalities in metric spaces. I: Structural properties. *Journal of Functional Analysis*, 278(4):108347, 2020.
[55] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
[56] Felix Otto and Michael Westdickenberg. Eulerian calculus for the contraction in the wasserstein distance. *SIAM journal on mathematical analysis*, 37(4):1227–1255, 2005.
[57] Gabriel Peyré and Marco Cuturi. Computational Optimal Transport: With Applications to Data Science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
[58] Filippo Santambrogio. *Optimal transport for applied mathematicians.* Birkhäuser/Springer, 2015.
[59] Erwin Schrödinger. Über die Umkehrung der Naturgesetze. Von E. Schrödinger. (Sonderausgabe a. d. Sitz.-Ber. d. Preuss. Akad. d. Wiss., Phys.-math. Klasse, 1931, IX.) Verlag W. de Gruyter, Berlin. Preis RM. 1,-. *Angewandte Chemie*, 44(30):636–636, 1931.
[60] Erwin Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. H. Poincaré*, 2(4):269–310, 1932.
[61] Karl-Theodor Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.
[62] Cédric Villani. *Topics in optimal transportation.* American Mathematical Soc., 2003.
[63] Cédric Villani. *Optimal transport. Old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin, 2009.
[64] Jean-Claude Zambrini. Variational processes and stochastic versions of mechanics. *J. Math. Phys.*, 27(9):2307–2330, 1986.
[65] Jean-Claude Zambrini. The research program of stochastic deformation (with a view toward geometric mechanics). In *Stochastic analysis: a series of lectures*, volume 68 of *Progr. Probab.*, pages 359–393. Birkhäuser/Springer, Basel, 2015.

Léonard Monsaingeon
GFM Universidade de Lisboa,
Campo Grande, Edifício C6, 1749-016 Lisboa, Portugal
& IECL Université de Lorraine, F-54506 Vandoeuvre-lès-Nancy Cedex, France

E-mail address: leonard.monsaingeon@univ-lorraine.fr

Luca Tamanini
CEREMADE (UMR CNRS 7534), Université Paris Dauphine PSL, Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France and INRIA-Paris, MOKAPLAN, 2 Rue Simone Iff, 75012, Paris, France

E-mail address: tamanini@ceremade.dauphine.fr

Dmitry Vorotnikov
University of Coimbra, CMUC, Department of Mathematics,
3001-501 Coimbra, Portugal

E-mail address: mitvorot@mat.uc.pt