A SECOND ORDER EQUATION FOR SCHRÖDINGER BRIDGES WITH APPLICATIONS TO THE HOT GAS EXPERIMENT AND ENTROPIC TRANSPORTATION COST

GIOVANNI CONFORTI

ABSTRACT. The Schrödinger problem is obtained replacing the mean square distance with the relative entropy in the Monge-Kantorovich problem. It was first addressed by Schrödinger to describe the most likely evolution of a large number of Brownian particles conditioned to reach an “unexpected configuration”. Its optimal value, the entropic transportation cost, and its optimal solution, the Schrödinger bridge, stand as the natural probabilistic counterparts to the transportation cost and displacement interpolation and provide a way of lifting from the point to the measure setting the concept of Brownian bridge. In this article, we prove that the Schrödinger bridge solves a second order equation in the Riemannian structure of optimal transport. Roughly speaking, the equation says that its acceleration is the gradient of the Fisher information. Using this result, we obtain a fine quantitative description of the dynamics and a new functional inequality for the entropic transportation cost, which generalize Talagrand’s transportation inequality. Finally, we study the convexity of the Fisher information along Schrödinger bridges, under the hypothesis that the associated reciprocal characteristic is convex. The techniques developed in this article are also well suited to study the Feynman-Kac penalisations of Brownian motion.

1. INTRODUCTION AND STATEMENT OF THE MAIN RESULTS

The Schrödinger problem (SP) is the problem of finding the best approximation in the relative entropy sense of a stationary dynamics $P$ under constraints on the marginal laws. It originates from the early works [45, 44] and is now an active field of research with ties to many others. Without being exhaustive, we mention [18, 19, 5, 15, 16] for connections with large deviations and [14, 37, 29, 21, 6, 53, 24] for the relations with Optimal transport and stochastic control. In [27, 28, 35, 49, 13, 54] Schrödinger bridges are used to develop stochastic mechanics and second order calculus for diffusions, while in the set in the article [34] they are employed to construct interpolations between probability measures on discrete spaces. We refer to [7, 8, 20, 46] for applications in engineering, economics and graphics. In this article we prove that the Schrödinger bridge $\hat{Q}$, i.e. the solution to (SP), solves a second order differential equation in the Riemannian structure of Optimal transport and obtain quantitative results for its dynamics. In particular we bring some answers to the basic questions

Q1 How well $\hat{Q}$ approximates the target law $P$?
Q2 At a given time, what is the distance of the marginal law of $\hat{Q}$ from the equilibrium measure?

Moreover, we shall establish a new connection between the so called reciprocal characteristics associated with a potential, and the convexity of the Fisher information along
Schrödinger bridges. Although the main objective of this paper is to understand the dynamics of Schrödinger bridges from a probabilistic viewpoint, our results can be seen as “stochastic” generalizations of well known results of Optimal transport, such as Talagrand’s inequality and displacement convexity of the entropy. Indeed, we shall how these classical results can be recovered through some small noise argument.

Organization of the paper. The paper is organized as follows: in section 1 we state the results; in section 2 we recall some notions of second order calculus in Wasserstein space, and in section 3 we make proofs. Some technical lemmas are in the Appendix.

1.1. The Schrödinger problem. In this subsection we give an introduction to SP and collect some useful facts. To formulate SP, let $\Omega$ be the space of continuous paths over the time interval $[0,1]$ with values in a smooth complete connected Riemannian manifold $(M,g)$ without boundary. We consider the law $P$ on $\Omega$ of the only stationary Markov measure for generator $L = \sigma \left( \frac{1}{2} \Delta - \nabla U \cdot \nabla \right)$, where $\sigma$ is a positive constant, $\Delta$ the Laplace-Beltrami operator and $U$ a smooth Lipschitz potential. The Schrödinger problem is the problem of finding the best approximation of $P$ in the relative entropy sense in the set of probabilities with prescribed marginals $\mu, \nu$ at times $t = 0, 1$.

(SPd) \[
\min H(Q|P) \quad Q \in \mathcal{P}(\Omega), X_0#Q = \mu, X_1#Q = \nu,
\]
where $\mathcal{P}(\Omega)$ is the space of probabilities over $\Omega$, $X_t(\omega) = \omega_t$ is the canonical projection map, $#$ the push-forward and $H(\cdot|P)$ the relative entropy functional. We refer to this formulation as the dynamic formulation. The optimal measure $\hat{Q}$ is the Schrödinger bridge $SB(L, \mu, \nu)$ between $\mu$ and $\nu$. The static formulation is obtained projecting onto the endpoint marginals.

(SP) \[
\min H(\pi|(X_0, X_1)\#P) \quad \pi \in \Pi(\mu, \nu),
\]
where $\Pi(\mu, \nu)$ is the set of couplings of $\mu$ and $\nu$. We call the optimal value $T^\sigma_{\mu, \nu}(\mu, \nu)$ of SP the entropic transportation cost. It is known [30, Prop. 2.4] that an optimal solution $\hat{Q}$ of SPd exists whenever $T^\sigma_{\mu, \nu}(\mu, \nu) < +\infty$ and that the two formulations are equivalent, see [18]: the Schrödinger bridge $Q$ is obtained lifting the optimal coupling $\hat{\pi}$ of SP to path space by mean of bridges. We have

(1) \[
\hat{Q}(\cdot) = \int_{M \times M} P^{xy}(\cdot) \hat{\pi}(dx \, dy),
\]
where $P^{xy}$ is the $xy$ bridge of $P$, [30] for details. Using this decompositon, one can show that $T^\sigma_{\mu, \nu}(\mu, \nu)$ is also the optimal value of SP and therefore that the optimal values of SPd and SP coincide.

---

1P may not be a Probability measure, but an infinite measure. This won’t be a problem as long as $Q$ is a probability measure. The note [33] takes care of this issue in detail.
The $fg$ representation of the optimal solution. It has been shown at \cite[Th 2.8]{30} that, under some mild assumptions on $\mu, \nu$ and $P$, there exist two measurable functions $f, g : M \to \mathbb{R}_{\geq 0}$ such that $E_P(f(X_0)g(X_1)) = 1$ and the SB takes the form

\begin{equation}
\hat{Q} = f(X_0)g(X_1)P
\end{equation}

$f, g$ are found by solving the Schrödinger system, see \cite{30, 43}. In the rest of the article we will often deal with the functions $f_t, g_t$ defined for $t \in [0, 1]$, by

\begin{equation}
f_t(x) := E_P[f(X_0)|X_t = x], \quad g_t(x) := E_P[g(X_1)|X_t = x].
\end{equation}

The following “hot gas experiment” provides an heuristic motivation for the formulation of $\text{SP}_d$ and helps building intuition.

**Hot gas experiment.** At time $t = 0$ we are given $N$ independent particles $(X^i_t)_{t \leq 1, i \leq N}$ whose configuration $\mu$. 

\[ \mu := \frac{1}{N} \sum_{i=1}^{N} \delta_{X^i_0}. \]

We then let the particles travel independently following the Langevin dynamics for $L$ and look at their configuration $\nu$ at $t = 1$.

\[ \nu := \frac{1}{N} \sum_{i=1}^{N} \delta_{X^i_1}. \]

If the Langevin dynamics has good ergodic properties and $N$ is very large, one expects $\nu$ to be very close to the invariant measure of $P$, which we denote $m$. However, although very unlikely, it is still possible to observe an unexpected configuration, meaning that $\nu$ is significantly different from $m$. Schrödinger question is to find, conditionally on this rare event, the most likely evolution of the particle system towards this unexpected configuration. In the limit when $N \to +\infty$, Sanov’s theorem tells that the most likely evolution is given by the minimiser of $\text{SP}_d$ (see also \cite[sec.6]{30}). This derivation of the Schrödinger problem can be viewed as a stochastic counterpart to the “lazy gas experiment” of optimal transport \cite[Ch. 16]{50}. Indeed

- Particles choose their final destination minimizing the relative entropy instead of the mean square distance.
- Particles travel along Brownian bridges instead of geodesics, see \cite{1}.

We refer to \cite[sec. 6]{32} for an extensive treatment of this analogy. Quite remarkably, the description of the hot gas experiment we gave is very similar to the original formulation of the problem that Schrödinger proposed back in 1932, (see \cite{45}) when none of the modern tools of probability or optimal transport was available.

\[^2\text{Schrödinger writes in } 45 " \text{un écart spontané et considérable par rapport à cette uniformité"} \]
1.2. An equation for the Schrödinger bridge. The first result of this article is that the marginal flow \((\mu_t)\) of \(SB(L, \mu, \nu)\) solves a second order differential equation, when viewed as a curve in the space \(\mathcal{P}_2(M)\) of probability measures with finite second moment endowed with the Riemannian-like structure of optimal transport, [47, 36]. Let us, for the moment, provide the minimal notions to state the result; we postpone to Section 2 a summary based on [2, 23] of second order calculus in \(\mathcal{P}_2(M)\). As said, we view \((\mu_t)\) as a curve in a kind of Riemannian manifold. Therefore, provided it is regular enough, we can speak of its velocity and acceleration. If \((\mu_t)\) is a regular curve (Def. 2.1), starting from the continuity equation
\[
\partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0,
\]
one can define a Borel family of vector fields, which is the velocity field of \((\mu_t)\). Furthermore, if \(v_t\) is absolutely continuous along \((\mu_t)\) (Def 2.5) we can compute its covariant derivative (Def. 2.7) \(D_t v_t\), which plays the role of the acceleration of \((\mu_t)\). We consider the Fisher information functional \(\mathcal{I}_U\) on \(\mathcal{P}_2(M)\)
\[
\mathcal{I}_U(\mu) = \begin{cases} \int_M |\nabla (\log \mu + 2U)|^2 d\mu & \text{if } \mu \ll \text{vol} \\ +\infty & \text{otherwise} \end{cases}
\]
In the definition above, and in the rest of the paper we make no distinction between a measure \(\mu\) and its Radon Nykodym density \(\frac{d\mu}{d\text{vol}}\) w.r.t. to the volume measure \(\text{vol}\) on \(M\). Moreover, if \(v\) is a vector field on \(M\), we abbreviate \(|v(z)|_{T_z M}\) with \(|v|\). In Definition 2.8 we provide the definition of gradient of a functional over \(\mathcal{P}_2(M)\). According to this notion, the gradient of \(\mathcal{I}_U\) is the vector field
\[
\nabla^W \mathcal{I}_U(\mu) := \nabla \left( -|\nabla \log \mu|^2 - 2\Delta \log \mu + 8U \right), \quad \text{where} \quad U = \frac{1}{2}(|\nabla U|^2 - \Delta U).
\]
The content of the next Theorem is that under some suitable regularity assumptions, the acceleration of \((\mu_t)\) is \(\nabla^W \mathcal{I}_U(\mu_t)\).

Assumptions 1.1. \(M, U, \mu, \nu\) satisfy one of the following

(A) \(M\) is compact, \(U \in C_\infty(M), \mathcal{H}_U(\mu), \mathcal{H}_U(\nu) < +\infty\).
(B) \(M = \mathbb{R}^d, U(z) = \frac{\alpha}{2} |x|^2\) for some \(\alpha > 0\), \(\mu, \nu \ll \text{vol}\) have compact support and bounded density against \(\text{vol}\).

Theorem 1.2. Let \(M, U, \mu, \nu\) be such that Assumption 1.1 holds. Then the marginal flow \((\mu_t)\) of \(SB(L, \mu, \nu)\) is a regular curve and its velocity field \((v_t)\) is absolutely continuous. \((\mu_t)\) solves
\[
\forall 0 < t < 1, \quad \frac{D}{dt} v_t = \frac{\sigma^2}{8} \nabla^W \mathcal{I}_U(\mu_t),
\]
where \(\frac{D}{dt} v_t\) is the covariant derivative of \((v_t)\) along \((\mu_t)\).

Assumption 1.1 essentially says that either \(M\) is compact or \(P\) is a stationary Ornstein-Uhlenbeck process on \(\mathbb{R}^d\). We impose this strong assumptions because we do not want to assume any regularity on the solution to \(SPd\) but only on the data of the problem. If we do so, Assumption 1.1 can be dropped.
Theorem 1.3. Assume that the dual representation (2) of \(SB(\mathcal{L}, \mu, \nu)\) holds. If the marginal flow \((\mu_t)\) is such that
\[
\forall \varepsilon \in (0, 1) \sup_{t \in [\varepsilon, 1-\varepsilon]} |\nabla W_{IU}(\mu_t)|_{L^2(\mu_t)} < +\infty
\]
and \(f_t, g_t\) are such that for all \(\varepsilon \in (0, 1)\)
\[
\sup_{t \in [\varepsilon, 1-\varepsilon]} |\nabla (\log g_t - \log f_t)|_{L^2(\mu_t)} < +\infty, \quad \sup_{x, t \in [\varepsilon, 1-\varepsilon]} |\text{Hess}_x (\log g_t - \log f_t)|_{op} < +\infty,
\]
Then the marginal flow \((\mu_t)\) is a regular curve and its velocity field \((v_t)\) is absolutely continuous. \((\mu_t)\) solves
\[
\forall 0 < t < 1, \quad \frac{D}{dt} v_t = \frac{\sigma^2}{8} \nabla W_{IU}(\mu_t),
\]
where \(\frac{D}{dt} v_t\) is the covariant derivative of \((v_t)\) along \((\mu_t)\).

We can view (6) as the Euler-Lagrange equation formally associated to the recently proposed Benamou-Brenier formulations of \(SPd\) [21, Cor 5.8], [6, Sec.7]. Thus, in principle we could link it with the theory of Hamiltonian system in \(P_2(M)\), see [1]. However, our proof is based on probabilistic arguments, and does not use that framework. By changing the sign in the right hand side of (6), one gets a nice connection with the Schrödinger equation, see [51]. The gradient flow of the Fisher information has been studied in [22] as a model for the quantum drift diffusion. Using the hot gas experiment, we can give an heuristic for equation (6) to hold. The marginal flow \((\mu_t)\) of SB models the empirical measure of a particle system which minimizes the relative entropy on path space while going from \(\mu\) to \(\nu\). Therefore, if there was no constraint on the final end point, \((\mu_t)\) would be the gradient flow of the marginal entropy started at \(\mu\)
\[
v_t = -\frac{\sigma}{2} \nabla W_{IU}(\mu_t), \quad \mu_0 = \mu
\]
where \(v_t\) is the velocity field of \(\mu_t\) and \(\mathcal{H}_U\) the relative entropy on \(P_2(M)\):
\[
\mathcal{H}_U(\mu) = \begin{cases} 
\int_M (\log \mu + 2U) d\mu & \text{if } \mu \ll \text{vol} \\
+\infty & \text{otherwise}
\end{cases}
\]
If we differentiate once more this relation to be allowed to impose the terminal condition, we obtain formally
\[
\frac{D}{dt} v_t = \frac{\sigma}{2} \nabla W_{IU}(\mu_t) = -\frac{\sigma}{2} \text{Hess}^W_{\mu_t} \mathcal{H}_U(v_t),
\]
\[
\frac{\sigma^2}{4} \text{Hess}^W_{\mu_t} \mathcal{H}_U(\nabla W_{IU}) = \frac{\sigma^2}{8} \nabla W_{IU}(\mu_t),
\]
where to get the last equation we used the fact that the Fisher information is the squared norm of the gradient of the entropy,

$$ I_U(\mu) = |\nabla W H U(\mu)|_2^2. $$

Theorem 1.2 gives an answer to the problem of determining what second order equation should the bridge of a diffusion satisfy. Other authors (see e.g. [27, 28, 13, 49]) have proven results in this direction, sometimes inspired by the theories of stochastic mechanics [38]. In this respect, equation (6) has some nice features. The first one is that all the objects appearing there are explicitly defined: this is not the case in some approaches to second order calculus for SDEs, where many quantities are defined through the solution of some PDEs which cannot be solved explicitly. Moreover, the equation has a clear interpretation and, as we shall see in Theorem 1.4, it allows for a quantitative study of the dynamics of Schrödinger bridges. The following kind of conservation law is also immediately derived: it can be compared with similar results obtained in the above mentioned articles.

**Corollary 1.1.** If either Theorem 1.2 or Theorem 1.3 hold, there exist a finite constant $c(\mu, \nu)$ such that

$$ \forall 0 < t < 1, \quad \frac{1}{\sigma^2} |v_t|^2_{L^2(\mu_t)} - \frac{1}{4} I_U(\mu_t) = c(\mu, \nu), $$

where $|\cdot|_{T_\mu} = |\cdot|_{L^2(\mu)}$ is the norm taken in the tangent plane at $\mu_t$ (see section 2.3 for details).

1.3. **Quantitative results for Schrödinger bridges.** Going back once more to the hot gas experiment is useful to give a qualitative description of the dynamics of SB. Indeed, the marginal flow $(\mu_t)$ is the empirical measure of a particle system that, at the same time

- has to minimize entropy: this means that particles are willing to arrange according to the equilibrium configuration $m$.
- has to reach an unexpected final configuration $\nu$, which looks very different from $m$.

Thus we expect the dynamics to be divided into two phases. In the first one, entropy minimization dominates and $\mu_t$ relaxes towards $m$. In the second phase, the necessity to attain the configuration $\nu$ at $t = 1$ prevails: particles start arranging according to $\nu$ and $(\mu_t)$ drifts away from $m$. From this sketch, it is clear that $Q2$ is the crucial question to address when studying the dynamics of Schrödinger bridges. Our answer is in the following Theorem, where we show that the Bakry-Émery condition is equivalent to an upper bound for the relative entropy of $\mu_t$ w.r.t. to $m$, under the additional assumption that $M$ is compact. The proofs we present are also formally correct in the non compact case as well and we expect the compactness assumption not to be necessary. We make use of it to justify some integration by parts under $m$.

**Theorem 1.4.** Let $M$ be compact and $\mathcal{L} = \frac{1}{2} \Delta - \nabla U \cdot \nabla$. The following are equivalent:

(i) The Bakry-Émery condition

$$ \forall x \in M, \quad \text{Ric}_x + \text{Hess}_x U \geq \lambda \text{id} $$

(13)
(ii) For any \( \mu, \nu \) such that \( H_U(\mu), H_U(\nu) < +\infty \) and any \( \sigma > 0 \), the marginal flow of \( SB(\mu, \nu, \sigma L) \) is such that

\[
\forall t \in [0, 1] \quad H_U(\mu_t) \leq \frac{1 - \exp(-\sigma \lambda (1-t))}{1 - \exp(-\sigma \lambda)} H_U(\mu) + \frac{1 - \exp(-\sigma t)}{1 - \exp(-\sigma \lambda)} H_U(\nu) - \frac{\cosh(\frac{\sigma \lambda}{2}) - \cosh(-\sigma \lambda (t - \frac{1}{2})) T_{\sigma} (\mu, \nu)}{\sinh(\frac{\sigma \lambda}{2})}.
\]

\[(14)\]

**Remark 1.1.** A simple case when our proof works beyond compactness is when Assumption 1.1 (B) holds., i.e. when \( P \) is a stationary Ornstein Uhlenbeck process.

![Figure 1](image-url)  

**Figure 1.** If \( M = \mathbb{R} \), \( \mathcal{L} = \frac{1}{2} \Delta - x \nabla \), then \( P \) is a stationary OU process and the condition \((13)\) holds with \( \lambda = 1 \). In blue, we plot the corresponding bound \((14)\) for \( \sigma = 1 \). We chose \( \mu = \nu \) to be a Gaussian law with mean zero and variance 0.25. The entropic cost turns out to be around 0.6841. In orange the same bound for \( \mathcal{L} = \frac{1}{2} \Delta - 2x \nabla \). In this case \( \lambda = 2 \), and the bound is thus stronger.

In the article [32] the author proves a representation of the first and second derivative of the entropy along SBs using the operators \( \Gamma \) and \( \Gamma^2 \). From this, he deduces that the entropy is convex under the condition \((13)\). However, no quantitative estimate such as \((14)\) is proven there. Having equation \((6)\) at hand, we can give a geometric interpretation to the results of [32], turn them into quantitative estimates and gives some answers to the questions raised there. We shall give a first “geometric” proof of \((i) \Rightarrow (ii)\) at Theorem 1.4 and then a second proof based on \( \Gamma \)-calculus, which follows the blueprint of the first proof.

**Remark 1.2.** Since the functions \( \frac{1 - \exp(-\sigma \lambda (1-t))}{1 - \exp(-\sigma \lambda t)} \) and \( \frac{1 - \exp(-\sigma t)}{1 - \exp(-\sigma \lambda)} \) are increasing in \( \lambda \) and only the term involving \( T_{\sigma} (\mu, \nu) \) is decreasing, one might think that the bound \((14)\) could get worse by increasing \( \lambda \). This is false because of how \( T_{\sigma} (\mu, \nu) \) depends on the marginal entropies. It can be read off the proof of Theorem 1.4 that for any \( \mu, \nu \) fixed the right hand side of \((14)\) is strictly
decreasing in $\lambda$. Moreover, observing that $T^\sigma_{U} \geq \frac{1}{2}(\mathcal{H}_U(\mu) + \mathcal{H}_U(\nu))$, one can deduce from (14) a bound for $\mathcal{H}_U(\mu)$ which gets better as $\lambda$ increases.

1.4. A functional inequality for the entropic transportation cost. Setting $t = \frac{1}{2}$ in (14), we obtain

\textbf{Corollary 1.2.} Let $M$ be compact and (13) hold. Then, for any $\mu, \nu \in \mathcal{P}(M)$ and $\sigma > 0$

\begin{equation}
T^\sigma_{U}(\mu, \nu) \leq \frac{1}{1 - \exp(-\frac{\sigma^2}{2})}(\mathcal{H}_U(\mu) + \mathcal{H}_U(\nu)).
\end{equation}

To interpret this inequality, recall that if $\hat{Q}$ is the Schrödinger bridge between $\mu$ and $\nu$ then

$$\mathcal{H}(\hat{Q} | P) = T^\sigma_{U}(\mu, \nu).$$

Therefore (15) gives an answer to $Q_1$ saying that the entropic cost is at most linear in the marginal entropies with a constant that improves with curvature. Since $T^\sigma_{U}(\mu, \nu)$ is a relative entropy on the path space, (15) may as well be seen a partial converse to the fact that joint entropies dominate marginal entropies. Such bound may be of particular interest in practice. Indeed, it is a very hard task to compute $T^\sigma_{U}(\mu, \nu)$. However, $\mathcal{H}_U(\mu), \mathcal{H}_U(\nu)$ are directly computable from the data of the problem. With some simple computation we also derive from Theorem 1.4 the following Corollary.

\textbf{Corollary 1.3.} Let $M$ be compact and (13) hold. Then, for any $\mu \in \mathcal{P}(M)$ and $\sigma > 0$

\begin{equation}
\mathcal{H}_U(\mu) \leq T^\sigma_{U}(\mu, m) \leq \frac{1}{1 - \exp(-\lambda)} \mathcal{H}_U(\mu).
\end{equation}

We shall see in the next section how (16) is connected with Talagrand’s inequality. Let us remark that the entropic transportation cost falls into the class of generalized costs considered by Gozlan, Roberto and Tetali in [25]. In principle, using their results, we can derive the dual form of the inequalities (15) and (16) and prove that they imply dimension-free concentration properties. We do not discuss these possibilities here, since they will be the object of a separate work.

1.5. Small noise limits and relations with classical optimal transport. In this section we “slow down” Brownian motion, i.e. consider the generator

$$\mathcal{L}^\varepsilon = \frac{\varepsilon^2}{2} \Delta,$$

assuming for simplicity that $U = 0$. Because of this choice, $\mathcal{H}_U$ is equal to the relative entropy against the volume measure, which we abbreviate with $\mathcal{H}$. Our aim is to provide an heuristic showing that the bounds obtained above are consistent with the classical results of optimal transport. We will make this argument rigorous in the proof of Theorem 1.4. In particular, we shall see how the convexity of the entropy along displacement interpolations is the small noise limit of the estimate (14), and Talagrand’s transportation inequality (see [48] [39]) is the small noise limit of (16). Let (13) hold and
\((\mu^\varepsilon_t)\) be the marginal flow of \(SB(\mu, \nu, \varepsilon)\); the bound (1.4) says that
\[
\mathcal{H}(\mu^\varepsilon_t) \leq \frac{1 - \exp(-\varepsilon \lambda (1 - t))}{1 - \exp(-\varepsilon \lambda)} \mathcal{H}(\mu) + \frac{1 - \exp(-\varepsilon \lambda t)}{1 - \exp(-\varepsilon \lambda)} \mathcal{H}(\nu)
\]
\[\quad - \frac{\cosh(\varepsilon \lambda)}{\sinh(\varepsilon \lambda)} T^\varepsilon_{\mu, \nu}, \tag{17}\]

Moreover, Corollary (1.3) is
\[
\forall \mu \in \mathcal{P}(M), \quad T^\varepsilon(\mu, \mathbf{m}) \leq \frac{1}{1 - \exp(-\varepsilon \lambda)} \mathcal{H}(\mu). \tag{18}\]

In the articles [37, 29, 24] connections between the Schrödinger problem and the Monge Kantorovich problem were established. In particular, in [29, Th 3.6] a \(\Gamma\)-convergence result for the “small noise” Schrödinger problem towards the Benamou-Brenier formulation of the Monge Kantorovich problem is proven. Roughly speaking, this implies that
\[
(\mu^\varepsilon_t) \to (\mu^0_t), \quad \varepsilon T^\varepsilon(\mu, \nu) \to \frac{1}{2} W^2(\mu, \nu)
\]
where \((\mu^0_t)\) is the displacement interpolation between \(\mu\) and \(\nu\). Using these results, and letting \(\varepsilon \to 0\) in (17), one recovers the convexity of the entropy along displacement interpolations, whereas in (18) one recovers Talagrand’s inequality.

1.6. Reciprocal characteristics and Fisher information.

Reciprocal characteristics. In this section we assume that \(M = \mathbb{R}^d, \sigma = 1\). \(P\) is then the stationary Langevin dynamics for the generator
\[
\mathcal{L} = \frac{1}{2} \Delta - \nabla U \cdot \nabla
\]

The aim of this section is to point out some connections between convexity of the reciprocal characteristic and convexity of the Fisher information along SBs. Let us recall the definition of reciprocal characteristic associated with a smooth potential \(U\): it is the vector field \(\nabla \mathcal{U}\), where we recall that
\[
\mathcal{U}(x) := \frac{1}{2} |\nabla U|^2(x) - \frac{1}{2} \Delta U(x).
\]

In [27, 9] it is shown how the reciprocal characteristic is an invariant for the family of bridges of the Langevin dynamics, and, more generally for the associated reciprocal processes. We refer to the survey [31] and the articles [28, 40, 41, 12] for more informations about reciprocal processes and reciprocal characteristics. In Theorem 1.3 we saw that the reciprocal characteristic is one of the terms expressing the “acceleration” of a Schrödinger bridge. In the recent work [11], a number of quantitative results about the bridges of the Langevin dynamics, under the hypothesis that \(\mathcal{U}\) is uniformly convex. In particular at Theorem 1.1, an equivalence between \(\alpha^2\) convexity of \(\mathcal{U}\), the possibility of constructing certain couplings for bridges with different endpoints, and a gradient estimate along the bridge semigroup is proven. We want to add one more motive of interest for the case when \(\mathcal{U}\) is convex showing that it implies lower bounds for the second derivative of the modified Fisher information.
Convexity of the Fisher information. If $\mathcal{U}$ is convex, we are able to show that $t \mapsto \mathcal{I}_U(\mu_t)$ is convex, provided some technical assumptions hold and the optimal coupling $\hat{\pi}$ of $\mathcal{SP}$ is a log-concave measure. In the next Theorem, since we are in the flat case, we prefer to write $Dv \cdot w$ instead of $\nabla w v$ for the derivatives of vector fields.

**Theorem 1.5.** Let $U, \mu, \nu$ be such that

(i) The hypothesis of Theorem 1.3 are satisfied.

(ii) The vector field

$$\partial_t \nabla^W \mathcal{I}_U(\mu_t) + D \nabla^W \mathcal{I}_U(\mu_t) \cdot v_t$$

is bounded in $L^2(\mu_t)$ uniformly in $t \in [0,1]$.

(iii) If we define $A_t, B_t$ by

$$A_t(x) := \sup_{1 \leq i_1 \leq \cdots \leq i_k \leq d} |\partial_{x_{i_1}} \cdots \partial_{x_{i_k}} \mu_t(x)|, \quad B_t(x) := \sup_{1 \leq i_1 \leq \cdots \leq i_k \leq d} |\partial_{x_{i_1}} \cdots \partial_{x_{i_k}} v_t^j(x)|,$$

then

$$\sup_{t \in [0,1]} \int_{\mathbb{R}^d} (A_t B_t)^2 (\max_{1 \leq i_1 \leq d} |\partial_{i_1} \mu_t| + \mu_t) d\mu_t < +\infty$$

(iv) The optimal coupling $\hat{\pi}$ in $\mathcal{SP}$ and log-concave.

(v) $\mathcal{U}$ is $\alpha^2$ convex.

Then $t \mapsto \mathcal{I}_U(\mu_t)$ is convex. Moreover, we have the following bound

$$\partial_t \mathcal{I}_U(\mu_t) \geq 8\alpha^2 |v_t|^2_{T_{\mu_t}} + \frac{1}{8} |\nabla^W \mathcal{I}_U(\mu_t)|^2_{T_{\mu_t}}.$$

A simple example to which the Theorem can be applied is when $\mathcal{P}$ is a stationary Ornstein Uhlenbeck process and $\mu, \nu$ are Gaussian laws. Concerning the assumptions of the Theorem, we believe that (iii) could be largely weakened, and that the assumption that $\hat{\pi}$ is log concave cannot be significantly weakened. Indeed, the key argument in the proof of Theorem 1.5 is to identify the log-concave measures as “convexity points” for the Fisher information functional $\mathcal{I}$ (corresponding to the case $U = 0$). Let us explain this. In Lemma 3.10 we prove that

$$\langle D t^W \mathcal{I}_U(\mu_t), v_t \rangle_{T_{\mu_t}} = 2 \int |Dv_t \cdot \nabla \log \mu_t + \nabla \text{div}(v_t)|^2 d\mu_t$$

$$- 4 \int \sum_{k,j=1}^d (Dv_t \cdot Dv_t)_{kj} \partial_{kj} \log \mu_t d\mu_t,$$

Schur’s product theorem [26, Th 7.5.3] makes sure that if $-\text{Hess} \log \mu_t$ is positive definite then $\sum_{k,j=1}^d (Dv_t \cdot Dv_t)_{kj} \partial_{kj} \log \mu_t > 0$. Since $\langle D t^W \mathcal{I}_U(\mu_t), v_t \rangle_{T_{\mu_t}}$ morally stands for $\langle \text{Hess} \mathcal{I}_U(\mu_t), v_t \rangle_{T_{\mu_t}}$ equation (20) indicates that log-concave measures are convexity

---

3We recall that $(v_t)$ is the velocity field of $(\mu_t), \langle \cdot, \cdot \rangle_{T_{\mu_t}}$ is the inner product in $L^2_{\mu_t}$ and $D_t$ the covariant derivative. We also denote $Dv_t$ the Jacobian matrix of $v_t$. 

points for $\mathcal{I}$. The same formula suggests that the Fisher information is not a convex function in general. For the modified Fisher information $\mathcal{I}_U$ we get
\begin{equation}
\frac{d}{dt} \nabla^W \mathcal{I}_U(\mu_t), v_t)_{T_{\mu_t}} = 2 \int |Dv_t \cdot \nabla \log \mu_t + \nabla \text{div}(v_t)|^2 d\mu_t
\end{equation}
(21)
\[ - 4 \int \sum_{k,j=1}^d (Dv_t \cdot Dv_t)_{kj} \partial_{kj} \log \mu_t d\mu_t + 8 \int \langle v_t, \text{Hess} \mathcal{W} \cdot v_t \rangle d\mu_t, \]
This formula clearly indicates that the convexity of $\mathcal{W}$ contributes to the convexity of $\mathcal{I}$.

**Remark 1.3.** If we reintroduce the dependence on $\sigma$ in Theorem 1.5 the analog of (20) is
\[ \partial_t \mathcal{I}_U(\mu_t) \geq 8\sigma^2 |v_t|^2_{T_{\mu_t}} + \frac{\sigma^2}{8} |\nabla^W \mathcal{I}_U(\mu_t)|^2_{T_{\mu_t}}. \]

Letting $\sigma \to 0$ and using the convergence of the SB towards the displacement interpolation, it is natural to guess that the modified Fisher information is convex along a geodesic $(\mu_t)$ provided $\mu_t$ is log concave for all $t$. At the moment of writing, we have no rigorous proof of this.

**1.7. Feynman-Kac penalisation of Brownian motion.** In this section, the setting is $M = \mathbb{R}^d, \sigma = 1, U = 0$. We look at a family of stochastic processes called Feynman-Kac penalisations of Brownian motion, see the monograph [42]. If $\mathbb{P}$ is the stationary Brownian motion on $\mathbb{R}^d$ (which is not a probability measure), a Feynman-Kac penalisation $Q$ [42, Ch. 2] is a probability measure on $\Omega$ which can be written as
\[ d\mathbb{Q} = \frac{1}{Z} f(X_0) \exp(-\int_0^1 K(X_s) ds) d\mathbb{P}. \]

where $K$ is a smooth lower bounded potential, $f$ a positive integrable function and $Z$ a normalization constant. As before, for any $t \in [0, 1]$ we define the functions
\begin{equation}
\begin{aligned}
f_t(x) &= E_\mathbb{P} \left[ f(X_0) \exp(-\int_0^t K(X_s) ds) | X_t = x \right], & g_t(x) &= E_\mathbb{P} \left[ \exp(-\int_0^t K(X_s) ds) | X_t = x \right]
\end{aligned}
\end{equation}

We will also consider the energy functional $\mathcal{E}_K$ defined as
\begin{equation}
\mathcal{E}_K(\mu) = \int_M K d\mu.
\end{equation}

**Theorem 1.6.** Assume that the marginal flow $(\mu_t)$ of $Q$ is such that
\begin{equation}
\forall \varepsilon \in (0, 1) \sup_{t \in [\varepsilon, 1-\varepsilon]} |\nabla^W \mathcal{I}(\mu_t) + \mathcal{E}_K(\mu_t)|_{L^2(\mu_t)} < +\infty
\end{equation}

and $f_t, g_t$ are such that for all $\varepsilon \in (0, 1)$
\begin{equation}
\sup_{t \in [\varepsilon, 1-\varepsilon]} |\nabla (\log g_t - \log f_t)|_{L^2(\mu_t)} < +\infty, \quad \sup_{t \in [\varepsilon, 1-\varepsilon], x \in M} |\text{Hess}_x (\log g_t - \log f_t)|_{op} < +\infty
\end{equation}

Then the marginal flow $(\mu_t)$ is a regular curve and its velocity field $(v_t)$ is absolutely continuous. $(\mu_t)$ solves
\[ \frac{D}{dt} v_t = \nabla^W \frac{1}{8} \mathcal{I} + \mathcal{E}_K(\mu_t). \]
where \( \frac{D}{dt} v_t \) is the covariant derivative of \( v_t \) along \( (\mu_t) \).

2. THE RIEMANNIAN STRUCTURE OF OPTIMAL TRANSPORT

2.1. Preliminaries and notation. We consider a smooth Riemannian Manifold \((M, g)\) which is complete, connected, closed, without boundary, and of finite dimension. We denote the Levi-Civita connection of \((M, g)\) by \( \nabla \). For any \( \varepsilon \in (0, 1) \), we define \( D_\varepsilon := [\varepsilon, 1 - \varepsilon] \times M \) and let \( C_\infty \) be the space of smooth functions on \( M \). \( C_{\infty}^b, C_{\infty}^c \subseteq C_\infty \) are the subset of positive function and the subset of bounded functions whose derivatives are all bounded. We also set \( C_{\infty}^{b+} := C_{\infty}^b \cap C_{\infty}^c \). Finally \( C_{\infty}^{c+} \subseteq C_\infty \) is the subset of compactly supported functions and \( C_{\infty}^{c+} := C_{\infty}^c \cap C_{\infty}^{c+} \). The definition of \( C_{\infty}^b, C_{\infty}^c, C_{\infty}^{c+} \) naturally extends to vector fields over \( M \), to functions whose domain is \([0, 1] \times M\) instead of \( M \), and to families of vector fields over \( M \) indexed by time. We call \( P_2(M) \) the space of probability measures on \( M \) admitting second moment. The Wasserstein distance \( W_2(\mu, \nu) \) between \( \mu, \nu \in P_2(M) \) is given by

\[
W_2^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{M^2} d_M(x, y)^2 \pi(dx, dy),
\]

where \( d_M \) is the Riemannian distance on \( M \). In the rest of the paper, curves are always defined on the time interval \([0, 1]\), unless otherwise stated. In this section, we shall describe a kind of Riemannian structure on \( P_2(M) \) associated with \( W_2(\cdot, \cdot) \). We do not prove new results and simply extract definitions and results from [2] and [23], to which we refer for the proofs.

2.2. Geodesics, Velocity fields. Recall that if \((x_t)\) is a curve in a generic metric space \((X, d)\) we say that it is absolutely continuous over \([\varepsilon, 1 - \varepsilon]\) provided that for some integrable function \( f \)

\[
\forall \varepsilon \leq r < s \leq 1 - \varepsilon \quad d(x_r, x_s) \leq \int_r^s f(t) dt.
\]

In all what follows, we will write “absolutely continuous curve” and mean “absolutely continuous over \([\varepsilon, 1 - \varepsilon]\)” for all \( \varepsilon \in (0, 1) \). A curve is a constant speed geodesic iff

\[
\forall s, t \in [0, 1] \quad d(x_s, x_t) = |t - s| d(x_0, x_1).
\]

\((X, d)\) is said to be a geodesic space provided that for any pair of points there exist a constant speed geodesic connecting them. It turns out [2] Th. 2.10] that \((P_2(M), W_2(\cdot, \cdot))\) is a geodesic space. If \((\mu_t)\) is an absolutely continuous curve on \((P_2(M), W_2(\cdot, \cdot))\) then one can show [2] Th. 2.29] that there exist a Borel family of vector fields \((v_t)\) such that the continuity equation

\[
(26) \quad \partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0
\]

holds in the sense of distributions. Moreover, it can be shown that there exist a unique up to a negligible set of times family of vector fields \((v_t)\) satisfying \(26\) and such that

\[
v_t \in \left\{ \nabla \varphi, \varphi \in C_{\infty}^c \right\} L^2 \quad t \text{ a.e. .}
\]
We call this the velocity field of $\mu_t$. Conversely, if $(v_t)$ is a Borel family of vector fields satisfying (26) such that $\forall \varepsilon \in (0, 1) \int_\varepsilon^{1-\varepsilon} |v_t|_{L^2} dt < +\infty$ and $v_t \in \{\nabla \varphi, \varphi \in C_c^\infty\} L^2_{\mu_t}$ a.e., then $(\mu_t)$ is absolutely continuous and $v_t$ is its velocity field.

2.3. Tangent space and Riemannian metric. The tangent space at $\mu \in P_2(M)$ is generated by the “space of directions”

$$\text{Geod}_\mu = \{\text{constant speed geodesics starting from } \mu \text{ defined on some interval } [0, T] \}/ \approx$$

where $(\mu_t) \approx (\mu'_t)$ provided they coincide on some right neighborhood of 0. We equip Geod$_\mu$ with the distance

$$D((\mu_t), (\mu'_t)) = \lim_{t \downarrow 0} \frac{1}{t} W_2(\mu_t, \mu'_t).$$

The tangent space $T_\mu$ is defined as the completion of Geod$_\mu$ w.r.t. to $D$. The tangent space can be nicely described using the completion in $L^2_{\mu}$ of the “space of gradients” by considering the map $\iota_\mu: \{\nabla \varphi, \varphi \in C_c^\infty\} L^2_{\mu} \to (T_\mu, D)$, $\iota_\mu \nabla \varphi \mapsto (\mu_\varphi)$

where $(\mu_\varphi)$ is the unique (modulo $\approx$) constant speed geodesic originating from $\mu$ and such that $v_0 = \nabla \varphi$. If $\mu$ is a regular measure in the sense of [2, Def 1.25] then $\iota_\mu$ can be extended to a bijective isometry. Therefore, $T_\mu$ inherits the inner product $\langle \cdot, \cdot \rangle_{T_\mu}$ from that of $L^2_{\mu}$

$$\langle \iota_\mu \nabla \varphi, \iota_\mu \nabla \psi \rangle_{T_\mu} := \langle \nabla \varphi, \nabla \psi \rangle_{L^2_{\mu}}$$

For this reason, in what follows we make no distinction between the elements of the two different spaces, and write $\langle \nabla \varphi, \nabla \psi \rangle_{T_\mu}$ instead of $\langle \iota_\mu \nabla \varphi, \iota_\mu \nabla \psi \rangle_{T_\mu}$. Similarly, we write $|\nabla \varphi|_{T_\mu}$ instead of $|\iota_\mu \nabla \varphi|_{T_\mu}$.

The Benamou-Brenier formula [4] shows how the metric we have just introduced is morally the Riemannian metric for $(P_2(M), W_2(\cdot, \cdot))$. Indeed it says that

$$W_2^2(\mu, \nu) = \inf_{\mu_t, v_t} \int_0^1 |v_t|_{T_{\mu_t}}^2 dt,$$

where $(\mu_t)$ varies in the set of absolutely continuous curves joining $\mu$ and $\nu$ and $(v_t)$ is the velocity field of $(\mu_t)$.

2.4. Regular curves and Flow maps. We give the Definition of regular curve following closely [23, Def. 2.8], the only difference being that we define regularity over $[\varepsilon, 1-\varepsilon]$ for $\varepsilon \in (0, 1)$, instead of looking at $[0, 1]$.

**Definition 2.1.** For $\varepsilon \in (0, 1)$, an absolutely continuous curve $(\mu_t)$ is regular over $[\varepsilon, 1-\varepsilon]$ provided

$$\int_\varepsilon^{1-\varepsilon} |v_t|_{T_{\mu_t}}^2 dt < +\infty$$

and

$$\int_\varepsilon^{1-\varepsilon} L(v_t) dt < +\infty,$$
For the definition of the constant L we refer to [23, Def 2.1]. In all what follows, by regular curve we mean “regular over \([\varepsilon, 1 - \varepsilon]\) for all \(\varepsilon \in (0, 1)\)”. If \((\mu_t)\) is a regular curve and \((v_t)\) its velocity field, there exist a unique family of maps, called flow maps, \(T(t, s, \cdot) : \text{supp} \mu_t \rightarrow \text{supp} \mu_s\) such for any \(\varepsilon \in (0, 1), t \in (0, 1), x \in \text{supp} \mu_t\) the curve \(s \mapsto T(t, s, x)\) is absolutely continuous over \((\varepsilon, 1 - \varepsilon)\) and satisfies

\[
\begin{aligned}
\frac{d}{ds} T(t, s, x) &= v_s(T(t, s, x)) \quad \text{a.e. } s \in (\varepsilon, 1 - \varepsilon), \\
T(t, t, x) &= x
\end{aligned}
\]

The flow maps enjoy the following properties

\[
T(r, s, \cdot) \circ T(t, r, \cdot) = T(t, s, \cdot), \quad T(t, s, \cdot)\#\mu_t = \mu_s.
\]

The maps \((\tau_x)_s^t\) and \(\tau_s^t(u)\). These maps are needed to define the covariant derivative. The following are Definition 2.9 and 2.12 in [23].

**Definition 2.2.** Let \((\mu_t)\) be a regular curve and \(T(t, s, \cdot)\) its flow maps. Given \(t, s \in [0, 1]\) and \(x \in \text{supp} \mu_t\) we let \((\tau_x)_s^t : T\mu_t \rightarrow T\mu_s\) be the map which associate to \(v \in T\mu_t\) its parallel transport along the absolutely continuous curve \(r \mapsto T(t, r, x)\) from \(r = s\) to \(r = t\).

The map \((\tau_x)_s^t\) is used to translate vectors along regular curves in \(M\). In the next definition we shall see how the maps \(\tau_s^t\) do the same for vector fields along regular curves in \(\mathcal{P}_2(M)\). It should be stressed that the next definition is not the parallel transport. Indeed, in general, \(\tau_s^t(u)\) may not be in \(T\mu_t\).

**Definition 2.3.** Let \((\mu_t)\) be a regular curve, \(T(t, s, \cdot)\) its flow maps and \(u\) a vector field in \(L^2_{\mu_s}\). Then \(\tau_s^t(u)\) is the vector field in \(L^2_{\mu_t}\) defined by

\[
\tau_s^t(u)(x) = (\tau_x)_s^t(u \circ T(t, s, x))
\]

The maps \(\tau_s^t\) have similar properties to the flow maps: they enjoy the group property

\[
\tau_s^t = \tau_r^t \circ \tau_s^r.
\]

Moreover \(\tau_s^t\) is an isometry from \(L^2_{\mu_s}\) to \(L^2_{\mu_t}\), i.e.

\[
\forall u \in L^2_{\mu_s}, \quad \int_M |u|^2 d\mu_s = \int_M |\tau_s^t(u)|^2 d\mu_t,
\]

or equivalently

\[
|u|_{T\mu_s} = |\tau_s^t(u)|_{T\mu_t}.
\]

2.6. Vector fields along a curve.

**Definition 2.4.** A vector field along a curve \((\mu_t)\) is a Borel map \((t, x) \mapsto u_t(x) \in T_x M\) such that \(u_t \in L^2_{\mu_t}\) for a.e. \(t\). It will be denoted by \((u_t)\).

Observe that also non tangent vector fields are considered in this definition, i.e. \(u_t\) may not be a gradient. Here is the definition of absolutely continuous vector field along a curve, see [23, Def. 3.2].
Definition 2.5. Let \((u_t)\) be a vector field along the regular curve \((\mu_t)\) and \(\tau_t^\mu(u_t)\) be given by Definition 2.3. We say that \((u_t)\) is absolutely continuous over \([\varepsilon, 1 - \varepsilon]\) provided the map
\[
t \mapsto \tau_t^\mu(u_t) \in L^2_{\mu_0}
\]
is absolutely continuous in \(L^2_{\mu_0}\) for all \(t_0 \in [\varepsilon, 1 - \varepsilon]\).

It can be seen that choice of \(t_0\) is irrelevant. One could then set \(t_0 = \varepsilon\) in the definition above. As before, by absolutely continuous vector field we mean "absolutely continuous over \([\varepsilon, 1 - \varepsilon]\) for all \(\varepsilon \in (0, 1)\)."

2.7. Total derivative and covariant derivative. We are ready to define the total derivative of an absolutely continuous vector field as in [23, Def 3.6]. Note that this is not yet the covariant derivative.

Definition 2.6. Let \((u_t)\) be an absolutely continuous vector field along the regular curve \((\mu_t)\). Its total derivative is defined as
\[
\frac{d}{dt}u_t := \lim_{h \to 0} \frac{\tau_{t+h}^\mu(u_{t+h}) - u_t}{h} \quad t - a.e.,
\]
where the limit is intended in \(L^2_{\mu_0}\).

To define the covariant derivative, we consider the orthogonal projection \(P_\mu : L^2_\mu \to \{\nabla \phi, \phi \in C_c^\infty\} L^2_\mu\). The following is [2, Def 6.8] for the flat case. For the general case, we refer to Definition 5.1 and discussion thereafter in [23].

Definition 2.7. Let \((\mu_t)\) be an absolutely continuous and tangent vector field along the regular curve \((\mu_t)\). Its covariant derivative is defined as
\[
\frac{D}{dt}u_t := P_\mu \left( \frac{d}{dt}u_t \right) \quad t - a.e..
\]

2.8. Levi-Civita connection. The covariant derivative just defined is the Levi-Civita connection, in the sense that it satisfies the compatibility of the metric and torsion free identity. The compatibility of the metric says that if \((u^1_t), (u^2_t)\) are tangent absolutely continuous vector fields along the regular curve \((\mu_t)\), then
\[
\partial_t \langle u^1_t, u^2_t \rangle_{T_\mu} = \langle \frac{D}{dt}u^1_t, u^2_t \rangle_{T_\mu} + \langle u^1_t, \frac{D}{dt}u^2_t \rangle_{T_\mu}.
\]
For the torsion free identity, we refer to [2, Eq. (6.10)] and [23, Sec. 5.1].

2.9. Gradient of a function. Here, we do not seek for a very general definition, but rather give a customary one, well-adapted to our scopes, following [2] Eq 3.50. In this definition and in the rest of the paper by \(\mu \in C_c^\infty\) we mean that \(\mu \ll \text{vol}\) and \(\frac{d\mu}{d\text{vol}} \in C_c^\infty\).

Definition 2.8. Let \(\mu \in C_c^\infty\). We say that \(\mathcal{F} : \mathcal{P}_2(M) \to \mathbb{R} \cup \{\pm \infty\}\) is differentiable at \(\mu\) if there exist a vector field \(w \in T_\mu\) such that for all \(\varphi \in C_c^\infty\)
\[
\lim_{h \to 0} \frac{\mathcal{F}(\mu^h_\varphi) - \mathcal{F}(\mu)}{h} = \langle w, \nabla \varphi \rangle_{T_\mu},
\]
where \(\mu^h_\varphi\) is the constant speed geodesic (modulo \(\approx\)) such that \(\mu_0 = \mu\) and \(v_0 = \nabla \varphi\).
It follows from our definition that if $F$ is differentiable at $\mu$, then there exist a unique $w$ fulfilling (35). We then call $w$ the gradient of $F$ at $\mu$, and denote it $\nabla_w F(\mu)$. Let us consider some functionals of interest. The gradient of the relative entropy $H_U$ is known to be

$$\nabla_w H_U(\mu) = \nabla \log \mu + 2\nabla U$$

provided $\nabla \log \mu + 2\nabla U \in L^2_\mu$ and $\mu$ is regular enough. Thus, using (36) we obtain that

$$I_U(\mu) = |\nabla_w H_U|^2_{T_\mu}$$

The gradient of the modified Fisher information at $\mu$ is the vector field

$$\nabla_w I_U(\mu) = \nabla \left( -|\nabla \log \mu|^2 - 2\Delta \log \mu + 8U \right)$$

provided the right hand side is in $L^2_\mu$ and $\mu$ regular enough (recall the definition of $U$ given at (5)). All these computations can be found in [51]. We shall derive the expression (38) in the Appendix.

### 2.10. Convexity of the entropy

Recall that on a smooth finite dimensional Riemannian manifold whose Levi-Civita connection is $\nabla$, the Hessian of $f$ at $x$ applied to $v \in T_x M$ is defined through (see e.g. [17, Ex. 11, pg. 141])

$$\text{Hess}_x f(v) = \nabla_v \nabla f(x) = \frac{D}{dt} \nabla f(x_t)|_{t=0}$$

where $(x_t)$ is any curve such that $x_0 = x$, $\dot{x}_0 = v$. In this article we are only interested in defining a kind of Hessian for the entropy functional $H_U$. Therefore, as we did before, to simplify the definitions, we restrict to a very special setting.

**Definition 2.9.** Let $M$ be compact. Consider a measure $\mu \in C^+\infty$ and a vector field $v \in C^b\infty \cap T_\mu$. We define

$$\text{Hess}_\mu^w H_U(v) = \frac{D}{dt} \nabla^w H_U(\mu_t)|_{t=0},$$

where $(\mu_t)$ is any regular curve such that $\mu_0 = \mu$, $v_0 = v$ and $\mu_t \in C^+\infty([0, \varepsilon] \times M)$ for some $\varepsilon > 0$.

It is easy to see that this is a good definition in the sense there is always one curve fulfilling the requirements and the value of the Hessian does not depend on the specific choice of the curve. The well known convexity of the entropy [52] implies, in particular, that under the assumption (13) we have

$$\forall \mu \in C^+\infty, v \in C^b\infty \cap T_\mu, \quad \langle \text{Hess}_\mu^w H_U(v), v \rangle_{T_\mu} \geq \lambda|v|_{T_\mu}^2$$

We shall give in the Appendix a formal proof of this fact, following [39].

### 3. PROOFS OF THE RESULTS

#### 3.1. Proof of Theorems 1.2 and 1.3

This section is structured as follows: we first prove some preparatory Lemmas (Lemma 3.1, 3.2, 3.3, 3.4). To prove these Lemmas, we rely on the technical Lemmas 4.2 and 4.3, which we put in the Appendix. Finally, we prove the two Theorems. In the proofs of the Lemmas and of the Theorems, we assume for simplicity that $\sigma = 1$. There is no difficulty in extending the proof to the general case. In the Lemmas 4.2 and 4.3 it is proven that if Assumption 1.1 holds, then the dual
representation (2) holds as well. Thus, we will take it for granted in the next Lemmas and in the proof of the Theorem. In what follows, it is sometimes useful to use the functions \( \tilde{f}_t \) and \( \tilde{g}_t \) defined by (recall the definition of \( f_t, g_t \) at (3))

\[
\tilde{f}_t(x) := \exp(-U(x))f_t(x), \quad \tilde{g}_t(x) := \exp(-U(x))g_t(x).
\]

The first Lemma is useful to identify the velocity field of \((\mu_t)\).

**Lemma 3.1.** Let Assumption 1.1 hold. If \( f, g \) are given by (2) and \( f_t, g_t \) by (3), then

\[
\forall \varepsilon \in (0, 1) \sup_{t \in [\varepsilon, 1-\varepsilon]} \int_M |\nabla \log g_t - \nabla \log f_t|^2 f_t g_t \, dm < +\infty
\]

**Proof.** Assume that Assumption 1.1 (B) holds. Then using point (ii) of Lemma 4.2 we see that there exist constants \( A, B \) such that

\[
\forall x \in \mathbb{R}^d \sup_{t \in [\varepsilon, 1-\varepsilon]} |\nabla \log g_t - \nabla \log f_t|^2 \leq A + B|x|^2
\]

The conclusion then follows from the boundedness of \( f_t, g_t \) (see (i) in Lemma 4.2) and the fact that \( m \) has Gaussian tails. If (A) holds, the conclusion follows directly from Lemma 4.3 \( \Box \)

Next result is a representation for \((\mu_t)\) and its velocity field.

**Lemma 3.2.** Let \( f, g \) be given by (2) and \( f_t, g_t, \tilde{f}_t, \tilde{g}_t \) by (3), (40). Then

\[
\forall t \in [0, 1], \ d\mu_t = f_t g_t \, dm = \tilde{f}_t \tilde{g}_t \, dvol.
\]

Moreover, if the conclusion of Lemma 3.1 holds \((\mu_t)\) is an absolutely continuous curve and its velocity field is

\[
(t, x) \mapsto \frac{\sigma}{2} \nabla (\log g_t - \log f_t).
\]

**Proof.** We can rewrite (2) as

\[
\frac{dQ}{dP} = f(X_0)g(X_1)
\]

Moreover, since \( P \) is stationary we have for any \( t \) that \( X_t \# P = m \). Therefore

\[
\frac{d\mu_t}{dm}(x) = \frac{d(X_t \# Q)}{d(X_t \# P)}(x) = E_P[f(X_0)g(X_1) | X_t = x]
\]

Markov property

\[
= E_P[f(X_0) | X_t = x] E_P[g(X_1) | X_t = x]
\]

\[
= f_t(x)g_t(x)
\]

Observing that \( dm = \exp(-2U) \, dvol \), (41) follows from the definition of \( \tilde{f}_t \) and \( \tilde{g}_t \). To prove the second statement, fix \( \varepsilon \in (0, 1) \). We observe that, since \( U \) is taken to be smooth, the well known results of Malliavin calculus grant that the function \( f_t \) and \( g_t \) are of class \( C^\infty \), and thus classical solutions on \( D_\varepsilon \) of the forward and backward Kolmogorov equations

\[
\partial_t f_t = \mathcal{L} f_t, \quad \partial_t g_t = -\mathcal{L} g_t.
\]
Using some standard algebraic manipulations and the positivity of $f_t, g_t$ one finds that $\tilde{f}_t$ and $\tilde{g}_t$ are classical solutions on $D_\varepsilon$ of
\begin{equation}
\tag{44}
\partial_t \tilde{f}_t = \frac{1}{2} \Delta \tilde{f}_t - \mathcal{W} \tilde{f}_t, \quad \partial_t \tilde{g}_t = -\frac{1}{2} \Delta \tilde{g}_t + \mathcal{W} \tilde{g}_t = 0,
\end{equation}
where $\mathcal{W}$ was defined at (5). Using this, we prove that $\frac{1}{2} \nabla (\log g_t - \log f_t)$ is a classical solution to the \textit{continuity equation} on $D_\varepsilon$. Indeed
\begin{equation}
\partial_t \mu_t (41) = (\partial_t \tilde{f}_t) \tilde{g}_t + \tilde{f}_t (\partial_t \tilde{g}_t)
\end{equation}(44) $\frac{1}{2} \tilde{g}_t \Delta \tilde{f}_t - \frac{1}{2} \tilde{f}_t \Delta \tilde{g}_t$
\begin{equation}
= \frac{1}{2} \nabla \cdot (\tilde{g}_t \nabla \tilde{f}_t) - \frac{1}{2} \nabla \cdot (\tilde{f}_t \nabla \tilde{g}_t)
\end{equation}
\begin{equation}
= \frac{1}{2} \nabla \cdot (\tilde{f}_t \tilde{g}_t (\nabla \log \tilde{f}_t - \nabla \log \tilde{g}_t))
\end{equation}(41) $\frac{1}{2} \nabla \cdot (\mu_t (\nabla \log \tilde{f}_t - \nabla \log \tilde{g}_t))$.

Thus, $(t, x) \mapsto \frac{1}{2} \nabla (\log g_t - \log f_t)$, solves the \textit{continuity equation}, it is of \textit{gradient type} and, thanks to Lemma 3.1 and (41),
\begin{equation}
\sup_{t \in [\varepsilon, 1-\varepsilon]} \frac{1}{2} \nabla (\log g_t - \log f_t) \mid T_{\mu_t} < +\infty
\end{equation}also holds. The conclusion then follows. $\square$

The following Lemma gives a sufficient condition for the existence of the covariant derivative and its explicit form. It is a slight rearrangement of Example 6.7 in [2] and Equation (3.3) in [23], the only difference being that we do not assume that $\xi_t$ has compact support. Its proof is in the Appendix.

\textbf{Lemma 3.3.} Let $(\mu_t)$ be a regular curve, $(v_t)$ its velocity field and $(\xi_t)$ a $C_\infty$ vector field along $(\mu_t)$. If
\begin{equation}
\forall \varepsilon \in (0, 1) \sup_{t \in [\varepsilon, 1-\varepsilon]} |\partial_t \xi_t + \nabla_v \xi_t| T_{\mu_t} < +\infty,
\end{equation}
then $(\xi_t)$ is absolutely continuous along $(\mu_t)$ in the sense of Definition 2.5 and
\begin{equation}
\forall t \in (0, 1), \quad \frac{d}{dt} \xi_t = \partial_t \xi_t + \nabla_v \xi_t.
\end{equation}

If $\xi_t = v_t$ then the we also have
\begin{equation}
\forall t \in (0, 1), \quad \frac{D}{dt} v_t = \frac{d}{dt} v_t = \partial_t v_t + \frac{1}{2} \nabla |v_t|^2.
\end{equation}

In this Lemma, we establish that $(\mu_t)$ has the desired regularity properties.

\textbf{Lemma 3.4.} Let Assumption 1.1 hold and $(\mu_t)$ be the marginal flow of $SB(Z, \mu, \nu)$ Then we have
(i) $(\mu_t)$ is a regular curve.
(ii) The velocity field $(v_t)$ is such that
\begin{equation}
\forall \varepsilon \in (0, 1) \sup_{t \in [\varepsilon, 1-\varepsilon]} |\partial_t v_t + \frac{1}{2} \nabla |v_t|^2| T_{\mu_t} < +\infty
\end{equation}
Proof. First, observe that, thanks to Lemma 3.2, we know that 
\[ v_t = \frac{1}{2} \nabla (\log g_t - \log f_t) \]
and that 
\[ d\mu_t = f_t g_t d\mu. \]
Using this, if Assumption 1.1(A) holds, the conclusion follows from Lemma 3.1. Therefore, let us assume that (B) holds. To prove that \((\mu_t)\) is regular we have to show (27) and (28). Therefore
\[ \forall t \in (0, 1), \quad |v_t|^2_{\mathcal{T}_{\mu_t}} = \frac{\sigma^2}{4} \int_M |\nabla \log g_t - \nabla \log f_t|^2 f_t g_t d\mu \]
Using this identity, (27) follows from Lemma 3.1. To prove (28) observe that, since \(v_t\) is of class \(C^\infty\), we have that
\[ L(v_t) = \sup_{x \in \mathbb{R}^d} |\nabla_x v_t(x)| \]
Using (42), we have for any \(\epsilon \in (0, 1)\):
\[ \sup_{t \in [\epsilon, 1-\epsilon]} \sup_{|w| = 1} \left| \nabla_w v_t(x) \right| \leq \frac{1}{2} \sup_{t \in [\epsilon, 1-\epsilon]} \sup_{|w_2, w_1| \leq 1} |\partial_{w_2} \partial_{w_1} \log f_t| \]
\[ + \frac{1}{2} \sup_{t \in [\epsilon, 1-\epsilon]} \sup_{|w_2, w_1| \leq 1} |\partial_{w_2} \partial_{w_1} \log g_t| \]
Point (iii) of Lemma 4.2 then immediately yields
\[ \sup_{t \in [\epsilon, 1-\epsilon]} L(v_t) < +\infty, \]
which proves (28) and that \((\mu_t)\) is a regular curve. Let us now turn to the proof of (48). First, we observe that \(\log f_t, \log g_t\) are classical solutions on \(D_\epsilon\) of the HJB equation
\[ \partial_t \log f_t = \mathcal{L} \log f_t + \frac{1}{2} |\nabla \log f_t|^2, \quad \partial_t \log g_t = -\mathcal{L} \log g_t - \frac{1}{2} |\nabla \log g_t|^2. \]
Then, combining (42) with point (ii) of Lemma 4.2 and the HJB equation we obtain that there exist constants \(A, B\) such that uniformly on \(D_\epsilon\)
\[ \left| \partial_t v_t + \frac{1}{2} \nabla |v_t|^2 \right|^2 (x) \leq A + B |x|^6. \]
Moreover, point (i) in Lemma 4.2 makes sure that \(\sup_{D_\epsilon} f_t g_t \leq C\) for some \(C < +\infty\). Thus,
\[ \sup_{t \in [\epsilon, 1-\epsilon]} \int_M \left| \partial_t v_t + \frac{1}{2} \nabla |v_t|^2 \right|^2 d\mu_t \leq \sup_{t \in [\epsilon, 1-\epsilon]} \int_M \left| \partial_t v_t + \frac{1}{2} \nabla |v_t|^2 \right|^2 f_t g_t d\mu \]
\[ \leq C \sup_{t \in [\epsilon, 1-\epsilon]} \int_M A + B |x|^6 \, d\mu < +\infty \]
where to obtain the last inequality we used that \(\mu\) has Gaussian tails. The desired conclusion follows.
\[ \square \]
Let us now prove Theorem 1.2.
Proof. Let $\varepsilon \in (0, 1)$ and Assumption 1.1 hold. This entitles us to apply Lemma 3.1 and 3.2 to conclude that the velocity field of $(\mu_t)$ is given by (42). Lemma 3.4 makes sure that $(\mu_t)$ is regular and that (45) holds for $\xi_t = v_t$. Therefore we can apply Lemma 3.3 to conclude that $(v_t)$ is absolutely continuous and its covariant derivative is

$$\frac{D}{dt} v_t = \partial_t v_t + \frac{1}{2} \nabla |v_t|^2. \tag{40}$$

Next, we shall prove that

$$\partial_t v_t + \frac{1}{2} \nabla |v_t|^2 = \nabla^\mathcal{W} \mathcal{J}_U (\mu_t), \tag{41}$$

which concludes the proof. To do this, recall that from Lemma 3.2

$$v_t = \frac{1}{2} \nabla (\log g_t - \log f_t) = \frac{1}{2} \nabla (\log \tilde{g}_t - \log \tilde{f}_t),$$

and that $\tilde{f}_t, \tilde{g}_t$ are classical solutions of (44) over $D_\varepsilon$. Therefore, by taking log and using the positivity of $\tilde{f}_t, \tilde{g}_t$ we get

$$\partial_t \log \tilde{f}_t = \frac{1}{2} \Delta \log \tilde{f}_t + \frac{1}{2} |\nabla \log \tilde{f}_t|^2 - \mathcal{W}, \quad \partial_t \log \tilde{g}_t = -\frac{1}{2} \Delta \log \tilde{g}_t - \frac{1}{2} |\nabla \log \tilde{g}_t|^2 + \mathcal{W},$$

where $\mathcal{W}$ is defined at (5). Combining these facts, we have

$$\partial_t v_t + \frac{1}{2} \nabla |v_t|^2 = -\frac{1}{2} \nabla \partial_t \log \tilde{f}_t + \frac{1}{2} \nabla \partial_t \log \tilde{g}_t + \frac{1}{2} \nabla |\nabla \log \tilde{f}_t|^2 \partial_t v_t + \frac{1}{2} \nabla |\nabla \log \tilde{g}_t|^2 \partial_t v_t \tag{51}$$

$$= -\frac{1}{2} \nabla \left( \frac{1}{2} \Delta \log \tilde{f}_t + \frac{1}{2} |\nabla \log \tilde{f}_t|^2 - \mathcal{W} \right) + \frac{1}{2} \nabla \left( -\frac{1}{2} \Delta \log \tilde{g}_t - \frac{1}{2} |\nabla \log \tilde{g}_t|^2 + \mathcal{W} \right) + \frac{1}{2} \nabla \left( \frac{1}{4} |\nabla \log \tilde{f}_t|^2 - \frac{1}{2} \langle \nabla \log \tilde{f}_t, \nabla \log \tilde{g}_t \rangle + \frac{1}{4} |\nabla \log \tilde{g}_t|^2 \right)$$

$$= -\frac{1}{4} \nabla \Delta \log \mu_t - \frac{1}{8} \nabla |\nabla \log \mu_t|^2 + \nabla \mathcal{W} \tag{40}$$

$$= \frac{1}{8} \nabla^\mathcal{W} \mathcal{J}_U (\mu_t). \tag{5}$$

Finally, we prove Theorem 1.3. Given the proof of Theorem 1.2 and the Lemmas proven above, this proof is quite straightforward.

Proof. One can check as in the proof of Theorem 1.2 that $\frac{1}{2} \nabla (\log f_t - \log g_t)$ solves the continuity equation in the classical sense. Assumption (6) ensures enough integrability to conclude that this vector field is indeed the velocity field of $(\mu_t)$. Using this representation of the velocity field, assumption (5) also ensures that $(\mu_t)$ is a regular curve. Repeating the same calculation as in the proof of Theorem 1.2 we can prove that

$$\partial_t v_t + \frac{1}{2} \nabla |v_t|^2 = \frac{1}{8} \nabla^\mathcal{W} \mathcal{J}_U.$$
Then, (6) enables to apply 3.3 to conclude the covariant derivative $\frac{D}{dt}v_t$ exists and it has the desired form. □

3.2. Proof of Theorem 1.4

3.2.1. First proof of (i) ⇒ (ii). Let us note that the hypothesis of the Theorem imply Assumption 1.1(B). Thus, we can use all the Lemmas proven in the previous section, as well as Theorem 1.2 and Lemma 4.3. The proof relies on the preparatory Lemmas 3.5,3.6 and 3.7. In the first Lemma we compute an expression for $c(\mu, \nu)$ in terms of $\tilde{f}_t, \tilde{g}_t$.

**Lemma 3.5.** Let $c(\mu, \nu)$ be defined via (12). Then
$$\forall 0 < t < 1, \quad c(\mu, \nu) = -\int_M \nabla \log \tilde{f}_t \cdot \nabla \log \tilde{g}_t + 2\mathcal{H} d\mu_t$$

**Proof.** By Corollary 1.1,
$$c(\mu, \nu) = \frac{1}{\sigma^2} |v_t|^2_{\Theta_{\mu_t}} - \frac{1}{4} \mathcal{H}(\mu_t)$$

We can exploit Lemma 3.2 and an integration by parts under $\text{vol}$ (justified by Lemma 4.3) to obtain

$$\frac{1}{\sigma^2} |v_t|^2_{\Theta_{\mu_t}} - \frac{1}{4} \mathcal{H}(\mu_t) = \int_M \left| \frac{1}{2} \nabla \log \tilde{f}_t \cdot \nabla \log \tilde{g}_t - \frac{1}{2} \nabla \log \tilde{f}_t \right|^2 - \frac{1}{4} \nabla \log \mu_t \cdot \nabla U - |\nabla U|^2 d\mu_t$$

$$= \int_M \left| \nabla \log \tilde{f}_t - \nabla \log \tilde{g}_t \right|^2 - \frac{1}{4} \nabla \log \tilde{f}_t \cdot \nabla \log \tilde{g}_t - \nabla \log \mu_t \cdot \nabla U - |\nabla U|^2 d\mu_t$$

$$= \int_M -\nabla \log \tilde{f}_t \cdot \nabla \log \tilde{g}_t - |\nabla U|^2 d\mu_t - \int_M \nabla \mu_t \cdot \nabla U d\text{vol}$$

$$= \int_M -\nabla \log \tilde{f}_t \cdot \nabla \log \tilde{g}_t + \Delta U - |\nabla U|^2 d\mu_t$$

$$= -\int_M \nabla \log \tilde{f}_t \cdot \nabla \log \tilde{g}_t + 2\mathcal{H} d\mu_t,$$

which is the desired conclusion. □

In the second Lemma we compute the first derivative of the entropy along SB.

**Lemma 3.6.** There exist functions $h_f$ and $h_b$ such that:
$$\mathcal{H}(\mu_t) = h_f(t) + h_b(t)$$

and for all $0 < t < 1$
$$\partial_t h_f(t) = -\frac{1}{2\sigma} |v_t - \sigma \nabla \mathcal{H}(\mu_t)|^2_{\Theta_{\mu_t}} \quad \partial_t h_b(t) = \frac{1}{2\sigma} |v_t + \sigma \nabla \mathcal{H}(\mu_t)|^2_{\Theta_{\mu_t}}.$$

**Proof.** First, observe that equation (41) combined with Lemma 4.3 make sure that all the exchanges of integral and derivatives which follow are justified and all the expressions...
Therefore, we have
\begin{equation}
\mathcal{H}_t(\mu_t) \mid_{t=1}^{t=2} \int_M \tilde{f}_t \tilde{g}_t (\log \tilde{f}_t \tilde{g}_t + 2U) d\mu_t \\
= \int_M \tilde{f}_t \tilde{g}_t (\log \tilde{f}_t + U) d\mu_t + \int_M \tilde{f}_t \tilde{g}_t (\log \tilde{g}_t + U) d\mu_t.
\end{equation}

We can write
\[ h_f(t) = \frac{1}{2} \mathcal{H}_t(\mu_t) - \frac{1}{2} \log \tilde{g}_t - \frac{1}{2} \log f_t(\mu_t), \]
where the definition of \( \frac{1}{2} \log \tilde{g}_t - \frac{1}{2} \log f_t \) is given at (23). Therefore
\[ \partial_t h_f(t) = \frac{1}{2} \left( \nabla^W \mathcal{H}_t(\mu_t), v_t \right)_{\mathcal{T}_{\mu_t}} - \frac{1}{2} \left( \nabla (\log \tilde{g}_t - \log \tilde{f}_t), v_t \right)_{\mathcal{T}_{\mu_t}} - \frac{1}{2} \partial_t \left( \log \tilde{g}_t - \log \tilde{f}_t \right)(\mu_t) \]

By Lemma 3.2 we have that \( v_t = \frac{1}{\sigma} \nabla (\log \tilde{g}_t - \log \tilde{f}_t) \), so that \( \left( \nabla (\frac{1}{2} \log \tilde{g}_t - \frac{1}{2} \log \tilde{f}_t), v_t \right)_{\mathcal{T}_{\mu_t}} = \frac{1}{\sigma} |v_t|_{\mathcal{T}_{\mu_t}}^2 \). Moreover
\[ \partial_t (\log \tilde{f}_t)(\mu_t) = \int_M (\partial_t \log \tilde{g}_t - \partial_t \log \tilde{f}_t) d\mu_t \]

\[ \sigma \int M \left( \Delta \tilde{g}_t \tilde{f}_t + \Delta \tilde{f}_t \tilde{g}_t \right) dvol + \sigma \int M \mathcal{W} \tilde{f}_t \tilde{g}_t dvol \]

\[ \int M (\nabla \log \tilde{g}_t \cdot \nabla \log \tilde{f}_t + 2\mathcal{W}) d\mu_t = \frac{\sigma}{2} - \frac{\sigma}{2} c(f, g). \]

Therefore, we have
\[ \partial_t h_f(t) = \frac{1}{2} \left( \nabla^W \mathcal{H}_t(\mu_t), v_t \right)_{\mathcal{T}_{\mu_t}} - \frac{1}{\sigma} |v_t|_{\mathcal{T}_{\mu_t}}^2 + \frac{\sigma}{2} c(f, g) \]

Cor. 3.1
\[ \frac{1}{2} \left( \nabla^W \mathcal{H}_t(\mu_t), v_t \right)_{\mathcal{T}_{\mu_t}} - \frac{1}{2} |v_t|_{\mathcal{T}_{\mu_t}}^2 \]

\[ \frac{1}{2} \left( \nabla^W \mathcal{H}_t(\mu_t), v_t \right)_{\mathcal{T}_{\mu_t}} - \frac{1}{2} |v_t|_{\mathcal{T}_{\mu_t}}^2 \]

\[ = - \frac{1}{2\sigma} |v_t| - \frac{\sigma}{2} \nabla^W \mathcal{H}_t(\mu_t)_{\mathcal{T}_{\mu_t}}. \]

which is the desired conclusion. The proof of the other identity is analogous. \( \square \)

In the last Lemma we compute the second derivative of the entropy.

**Lemma 3.7.** We have for all \( 0 < t < 1 \)
\[ \partial_t h_f(t) = \frac{1}{2} \left( \text{Hess}^W_{\mu_t} \mathcal{H}_t\left( v_t - \frac{1}{2} \nabla^W \mathcal{H}_t(\mu_t) \right), v_t - \frac{1}{2} \nabla^W \mathcal{H}_t(\mu_t) \right)_{\mathcal{T}_{\mu_t}}, \]

and
\[ \partial_t h_b(t) = \frac{1}{2} \left( \text{Hess}^W_{\mu_t} \mathcal{H}_t\left( v_t + \frac{1}{2} \nabla^W \mathcal{H}_t(\mu_t) \right), v_t + \frac{1}{2} \nabla^W \mathcal{H}_t(\mu_t) \right)_{\mathcal{T}_{\mu_t}}. \]
Proof. Lemma 3.4 ensures \((\mu_t)\) is regular. On the other hand, Lemma 4.3 combined with (41) and (56) show that \(\nabla^W \mathcal{H}_U(\mu_t) \in C^b_c\) on any \(D_s\). Thus, we can use Lemma 3.3 to conclude that \(\frac{D}{dt} \nabla^W \mathcal{H}_U(\mu_t)\) is well defined. Using Definition 2.9, we also obtain that

\[
\frac{D}{dt} \nabla^W \mathcal{H}_U(\mu_t)\bigg|_{t=s} = \text{Hess}^W_{\mu_t} \mathcal{H}_U(v_s),
\]

whereas from the compatibility with the metric we get

\[
\nabla^W \mathcal{H}_U(\mu_t) \overset{(\ref{37})}{=} \nabla^W \mathcal{H}_U|_{T_{\mu_t}} = 2 \text{Hess}^W_{\mu_t} \mathcal{H}_U(\nabla^W \mathcal{H}_U),
\]

where we wrote \(\nabla^W \mathcal{H}_U\) in place of \(\nabla^W \mathcal{H}_U(\mu_t)\) to simplify notation. We shall retain this convention in the next calculations. We thus have, using the compatibility with the metric

\[
\partial_t h_f(t) \overset{\text{Lemma 3.4}}{=} -\frac{1}{2\sigma} \partial_t |v_t| - \frac{\sigma}{2} \nabla^W \mathcal{H}_U|_{T_{\mu_t}}^2 - \frac{1}{\sigma} \left( \frac{D}{dt} \left( v_t - \frac{\sigma}{2} \nabla^W \mathcal{H}_U \right), v_t - \frac{\sigma}{2} \nabla^W \mathcal{H}_U \right)_{T_{\mu_t}} - \frac{\sigma}{8} \nabla^W \mathcal{H}_U(\mu_t) - \frac{1}{2\sigma} \nabla^W \mathcal{H}_U, v_t - \frac{\sigma}{2} \nabla^W \mathcal{H}_U \right)_{T_{\mu_t}} - \frac{1}{2} \left[ \text{Hess}^W_{\mu_t} \mathcal{H}_U(v_t - \frac{\sigma}{2} \nabla^W \mathcal{H}_U), v_t - \frac{\sigma}{2} \nabla^W \mathcal{H}_U \right]_{T_{\mu_t}}.
\]

The other identity is proven analogously. \(\square\)

Let us now prove Theorem 1.4.

Proof. We first assume that \(f\) and \(g\) are continuous. The Bakry Émery condition (13) grants \(\lambda\)-convexity of the entropy. Therefore we get

\[
\forall t \in (0, 1), \quad \partial_t h_f(t) \overset{\text{Lemma 3.6}}{=} \frac{\lambda}{2} |v_t| - \frac{\sigma}{2} \nabla \mathcal{H}_U|_{T_{\mu_t}}^2 - \lambda \sigma \partial_t h_f(t),
\]

where the use of (39) is justified by the fact that \(\mu_t \in C^+_\infty\) and \(\nabla^W \mathcal{H}_U(\mu_t) \in C^b_c\). In the same way,

\[
\forall t \in (0, 1), \quad \partial_t h_b(t) \geq \lambda \sigma \partial_t h_b(t).
\]

Since \(f, g\) are continuous, it is easy to see that both \(h_f\) and \(h_b\) are continuous over the whole \([0, 1]\). Moreover, Lemma 3.6 and Lemma 3.7 make sure that they are \(C^2\) over \((0, 1)\). Thus, we can apply Lemma 4.1 (see the Appendix) to obtain

\[
\begin{align*}
h_f(t) & \leq h_f(0) + (h_f(1) - h_f(0)) \frac{1 - \exp(-\lambda \sigma t)}{1 - \exp(-\lambda \sigma)} \\
h_b(t) & \leq h_b(1) + (h_b(0) - h_b(1)) \frac{1 - \exp(-\lambda \sigma (1 - t))}{1 - \exp(-\lambda \sigma)}
\end{align*}
\]
Summing the two inequalities we obtain

\[(1 - \exp(-\lambda \sigma)) \mathcal{H}_U(\mu) \leq h_f(0)(\exp(-\lambda \sigma t) - \exp(-\lambda \sigma)) + h_b(0)\left(1\exp(-\lambda \sigma(1-t))\right)
+ h_f(1)\left(1 - \exp(-\lambda \sigma t)\right) + h_b(1)\left(\exp(-\lambda \sigma(1-t)) - \exp(-\lambda \sigma)\right)
= (h_f(0) + h_b(0))\left(1 - \exp(-\lambda \sigma(1-t))\right)
+ (h_f(1) + h_b(1))\left(1 - \exp(-\lambda \sigma t)\right)
- (h_f(0) + h_b(1))\left(\exp(-\lambda \sigma t) - 1\right)\left(\exp(-\lambda \sigma(1-t)) - 1\right).
\]

Dividing by \(1 - \exp(-\lambda \sigma)\) we arrive, after some simple calculations at

\[(55) \mathcal{H}_U(\mu) \leq (h_f(0) + h_b(0))\frac{1 - \exp(-\lambda \sigma(1-t))}{1 - \exp(-\lambda \sigma)} + (h_f(1) + h_b(1))\frac{1 - \exp(-\lambda \sigma t)}{1 - \exp(-\lambda \sigma)}
- (h_f(0) + h_b(1))\frac{\cosh(\frac{\lambda \sigma t}{2}) - \cosh(\lambda \sigma(t - \frac{1}{2}))}{\sinh(\frac{\lambda \sigma t}{2})}.
\]

Observe that, by definition

\[(56) h_f(0) + h_b(0) = \mathcal{H}_U(\mu_0) = \mathcal{H}_U(\mu), \quad h_f(1) + h_b(1) = \mathcal{H}_U(\mu_1) = \mathcal{H}_U(\nu).
\]

From the definition of \(\hat{f}_0, \hat{g}_0\), we obtain, using the standard properties of conditional expectation:

\[h_f(0) = \int_M \hat{f}_0 \hat{g}_0(\log f_0 + U) d\text{vol} \]
\[= \int_M f_0 \hat{g}_0 \log f_0 d\text{m} \]
\[= E_P\left(f(X_0)E_P[g(X_1)|X_0] \log f(X_0)\right) \]
\[= E_P\left(f(X_0)g(X_1) \log f(X_0)\right).
\]

Using the same argument we can show that

\[h_b(1) = E_P\left(f(X_0)g(X_1) \log g(X_1)\right),
\]

meaning that

\[(57) h_f(0) + h_b(1) = E_P\left(f(X_0)g(X_1) \log [f(X_0)g(X_1)]\right) = E_Q\left(\log \frac{dQ}{dP}\right) = \mathcal{T}_U^G(\mu, \nu).
\]

Plugging (56) and (57) into (55) yields the conclusion. The general case when \(f, g\) are not continuous is obtained with a standard approximation argument. \(\square\)

3.2.2. Second proof of (i) \(\Rightarrow\) (ii). In this proof we assume for simplicity \(\sigma = 1\). There is no difficulty in extending it to the general case. We recall the definitions of the \(\Gamma\) and \(\Gamma_2\) operators on \(\mathcal{C}_\infty \times \mathcal{C}_\infty\)

\[(58) \Gamma(f, g) := \frac{1}{2}\left[\mathcal{L}(fg) - f(\mathcal{L}g) - g(\mathcal{L}f)\right],
\]
\[(59) \Gamma_2(f, g) := \frac{1}{2}\left[\mathcal{L}\Gamma(f, g) - \Gamma(\mathcal{L}f, g) - \Gamma(f, \mathcal{L}g)\right].
\]
Since $M$ is compact, the integration by parts formula for the invariant measure $\mathbf{m}$ tells that for any $f, g \in C^\infty(= C^b_\infty = C^\infty_\infty)$,

\[
\int_M f \mathcal{L} g \mathbf{d}m = \int_M g \mathcal{L} f \mathbf{d}m = - \int_M \Gamma(f, g) \mathbf{d}m,
\]

see the monograph \[3\] for details. The next two lemmas are the analogous to the Lemma 3.6 and 3.7. In their proofs, we will often exchange integrals and time derivatives, and use the integration by parts formula (60). All these operations are justified by Lemma 4.3, where it is proven that $f_t, g_t, \log f_t, \log g_t$ are of class $C^b_\infty$ over $D_\varepsilon$, for any $\varepsilon \in (0, 1)$.

**Lemma 3.8.** Let $f, g$ be given by (2) and $f_t, g_t$ be given by (3). Assume that $f$ and $g$ are of class $C^b_\infty$. Then for any $t \in [0, 1]$, we can write $M_U(\mu_t) = h_f(t) + h_b(t)$, where

\[
h_f(t) = \int \log f_t f_t g_t \mathbf{d}m, \quad h_b(t) = \int \log g_t f_t g_t \mathbf{d}m
\]

Moreover, for all $0 < t < 1$:

\[
\partial_t h_f(t) = - \int_M \Gamma(\log f_t, \log f_t) f_t g_t \mathbf{d}m, \quad \partial_t h_b(t) = \int_M \Gamma(\log g_t, \log g_t) f_t g_t \mathbf{d}m
\]

**Proof.** From (41) we get

\[
M_U(\mu_t) = \int M f_t g_t \log f_t g_t \mathbf{d}m
\]

which yields (61). Recall that $f_t, g_t$ and their logarithms are classical solutions over $[0, 1] \times M$ of

\[
\partial_t f_t = \mathcal{L} f_t, \quad \partial_t \log f_t = \mathcal{L} \log f_t + \Gamma(\log f_t, \log f_t),
\]

\[
\partial_t g_t = - \mathcal{L} g_t, \quad \partial_t \log g_t = - \mathcal{L} \log g_t - \Gamma(\log g_t, \log g_t),
\]

where we used the fact that $\Gamma(f, f) = \frac{1}{2} |\nabla f|^2$. Hence

\[
\partial_t h_f(t) = \int_M (\mathcal{L} f_t) g_t \log f_t - (\mathcal{L} g_t) f_t \log f_t + \mathcal{L}(\log f_t) f_t g_t + \Gamma(\log f_t, \log f_t) f_t g_t \mathbf{d}m
\]

Using integration by parts, we get

\[
\int_M (\mathcal{L} f_t) g_t \log f_t - (\mathcal{L} g_t) f_t \log f_t + \mathcal{L}(\log f_t) f_t g_t + \Gamma(\log f_t, \log f_t) f_t g_t \mathbf{d}m
\]

\[
= \int_M (\mathcal{L} f_t) g_t \log f_t - f_t \mathcal{L}(\log f_t) f_t + \mathcal{L}(\log f_t) f_t g_t + \Gamma(\log f_t, \log f_t) f_t g_t \mathbf{d}m
\]

\[
= \int_M \Gamma(\log f_t, \log f_t) f_t g_t - 2 \Gamma(\log f_t, f_t) g_t \mathbf{d}m
\]

\[
= - \int_M \Gamma(\log f_t, \log f_t) f_t g_t \mathbf{d}m.
\]

The other identity is derived analogously. \qed
Lemma 3.9. In the same hypothesis of Lemma 3.8, we have for all $0 < t < 1$

\begin{equation}
\partial_\tau h_f(t) = 2 \int \Gamma_2(\log f_t, \log f_t) f_t g_t d\mathfrak{m},
\end{equation}
\begin{equation}
\partial_\tau h_b(t) = 2 \int \Gamma_2(\log g_t, \log g_t) f_t g_t d\mathfrak{m}.
\end{equation}

Proof. Using Lemma 3.8, we get

\begin{equation}
\partial_\tau h_f(t) = -\int_M (\mathcal{L} f_t) \Gamma(\log f_t, \log f_t) g_t - (\mathcal{L} g_t) \Gamma(\log f_t, \log f_t) f_t + 2\Gamma(\mathcal{L} \log f_t, \log f_t) f_t g_t d\mathfrak{m}
\end{equation}
\begin{equation}
+ 2\Gamma(\mathcal{L} \log f_t, \log f_t) f_t g_t + 2\Gamma(\Gamma(\log f_t, \log f_t)) f_t g_t d\mathfrak{m}.
\end{equation}

We can now use (60) and the fact that $\Gamma$ is a derivation to rewrite the last term in the above expression

\begin{equation}
2 \int_M \Gamma(\log f_t, \log f_t) f_t g_t d\mathfrak{m} = 2 \int_M \Gamma(\log f_t, \log f_t) f_t g_t d\mathfrak{m}
\end{equation}
\begin{equation}
- \int_M (\mathcal{L} f_t) \Gamma(\log f_t, \log f_t) g_t - (\mathcal{L} g_t) \Gamma(\log f_t, \log f_t) f_t g_t
\end{equation}
\begin{equation}
+ \int_M (\mathcal{L} g_t) \Gamma(\log f_t, \log f_t) f_t - (\mathcal{L} f_t) \Gamma(\log f_t, \log f_t) g_t d\mathfrak{m}.
\end{equation}

Plugging this expression in (65), and using the definition of $\Gamma_2$, we arrive at

\begin{equation}
\partial_\tau h_f(t) = \int_M \left(\mathcal{L} \Gamma(\log f, \log f) - 2\Gamma(\mathcal{L} \log f, \log f)\right) g_t f_t = 2 \int_M \Gamma_2(\log f_t, \log f_t) f_t g_t d\mathfrak{m},
\end{equation}

which is the desired conclusion. The other identity is proven analogously. \hfill \Box

Let us now complete the proof of Theorem 1.4.

Proof. Let $\mu, \nu$ be such that $f, g \in \mathcal{C}_\infty^{b+}$. It is well known that under (13)

\begin{equation}
\forall f \in \mathcal{C}_\infty \quad \Gamma_2(f, f) \geq \frac{\lambda}{2} \Gamma(f, f).
\end{equation}

In view of Lemmas 3.8 and 3.9, this implies that

\begin{equation}
\partial_\tau h_f(t) \geq -\lambda \partial_\tau h_f(t), \quad \partial_\tau h_b(t) \geq \lambda \partial_\tau h_b(t).
\end{equation}

From this point on, the proof goes as the previous case, and we do not repeat it. The case when $f, g$ are not both in $\mathcal{C}_\infty^{b+}$ follows with a standard approximation argument. \hfill \Box
3.2.3. Proof of (ii) ⇒ (i).

Proof. Let us choose $\mu$ and $\nu$ such that $\mu, \nu \ll m$ and with bounded density, and let $(\mu^i_\varepsilon)$ be the marginal flow of $SB(\mu, \nu, \varepsilon L)$. Then, equation (14) tells that

$$\forall t \in [0, 1] \quad \mathcal{H}_U(\mu^i_\varepsilon) \leq \frac{1 - \exp(-\varepsilon \lambda (1 - t))}{1 - \exp(-\varepsilon \lambda)} \mathcal{H}_U(\mu) + \frac{1 - \exp(-\varepsilon \lambda)}{1 - \exp(-\varepsilon \lambda)} \mathcal{H}_U(\nu) - \frac{\cosh\left(\frac{\lambda}{2}\right) - \cosh\left(-\varepsilon \lambda (t - \frac{1}{2})\right)}{\sinh\left(\frac{\lambda}{2}\right)} \mathcal{T}_\mathcal{H}_U(\mu, \nu).$$

Next, as it easy to check, our hypothesis allows us to apply the results of [24, sec 6]. They tell that,

- For $t \in [0, 1]$,
  $$\lim_{\varepsilon \to 0} \mu^i_\varepsilon t \to \mu^0_t,$$
  where the limit is intended in the weak sense and $(\mu^0_t)$ is the unique constant speed geodesic between $\mu$ and $\nu$.

- $$\lim_{\varepsilon \to 0} \varepsilon \mathcal{T}_\mathcal{H}_U(\mu, \nu) = \frac{1}{2} W_2^2(\mu, \nu).$$

In view of this, taking the lim inf at both sides in the equation above and using the lowersemicontinuity of $\mathcal{H}_U$ yields

$$(66) \quad \mathcal{H}_U(\mu^0_t) \leq (1 - t) \mathcal{H}_U(\mu) + t \mathcal{H}_U(\nu) - \frac{\lambda}{2} t (1 - t) W_2^2(\mu, \nu)$$

Therefore we obtained that the entropy is $\lambda$ convex along displacement interpolations, provided the initial and final measure are absolutely continuous with bounded density. It is well known that $\lambda$-convexity along all geodesics implies the condition (13). However, the proof of this fact given in [52] uses only geodesics between uniform measures on balls, which are clearly among those for which we can prove (66). The conclusion follows. □

3.2.4. Proof of Corollary 1.3.

Proof. If $\mathcal{H}_U(\mu) = +\infty$, there is nothing to prove since, $\mathcal{T}_\mathcal{H}_U(\mu, \nu) = +\infty$ as well. Let $\mathcal{H}_U(\mu) < +\infty$ and $\nu = m$. Then we can apply Theorem 1.4. After dividing by $(1 - t)$ the bound (14) becomes, observing that entropy is non negative and $\cosh$ symmetric around $t = 0$:

$$0 \leq \frac{1 - \exp(-\lambda (1 - t))}{1 - \exp(-\lambda)} \mathcal{H}_U(\mu) - \frac{\cosh\left(\frac{\lambda}{2}\right) - \cosh\left(\frac{\lambda}{2} - \lambda (1 - t)\right)}{(1 - t) \sinh\left(\frac{\lambda}{2}\right)} \mathcal{T}_\mathcal{H}_U(\mu, \nu).$$

Letting $t \to 1$ the conclusion follows. □
3.3. **Proof of Theorem 1.5.** We prove two preparatory Lemmas, and then the Theorem. The first Lemma is a rigorous proof of equation (21).

Lemma 3.10. The vector field $\nabla W U(\mu_t)$ is absolutely continuous along $(\mu_t)$. Moreover

$$\langle \frac{d}{dt} \nabla W U(\mu_t), v_t \rangle T_{\mu_t} = 2 \int_{\mathbb{R}^d} |Dv_t \cdot \nabla \log \mu_t + \nabla \text{div}(v_t)|^2 d\mu_t$$

$$- 4 \int_{\mathbb{R}^d} \sum_{k,j=1}^d (Dv_t \cdot Dv_t)_{kj} \partial_k \log \mu_t d\mu_t + 8 \int_{\mathbb{R}^d} (v_t, \text{Hess} \phi \cdot v_t) d\mu_t$$

(67)

**Proof.** Point (i) if the assumption make sure that $(\mu_t)$ is regular. (ii) combined with Lemma 3.3 grant the desired absolute continuity and that

$$\frac{d}{dt} \nabla W U(\mu_t) = \partial_t \nabla W U(\mu_t) + D \nabla W U(\mu_t) \cdot v_t.$$ 

Since $v_t \in T_{\mu_t}$, we also have, by definition of covariant derivative

$$\langle \frac{d}{dt} \nabla W U(\mu_t), v_t \rangle T_{\mu_t} = \langle \frac{d}{dt} \nabla W U(\mu_t), v_t \rangle T_{\mu_t}$$

Let us now prove (67). First, we do it for the case when $U = 0$. In this case, from (5) we have $\nabla W U(\mu_t) = \alpha_t + \beta_t$ with

$$\alpha_t = -\nabla |\nabla \log \mu_t|^2, \quad \beta_t = -2\nabla \Delta \log \mu_t$$

From now on, we drop the dependence on $t$ both in $\mu_t$ and $v_t$ and adopt Einstein’s convention for indices. Moreover, we abbreviate $\partial_x$ with $\partial_t$ and we are going to use, without mentioning it, the fact that $v$ is a gradient vector field, i.e. $\partial_t v^j \equiv \partial_j v^k$. Since $\mu_t \in \mathcal{C}_+^\infty$, we can use the continuity equation in the form $\partial_t \log \mu = -\text{div}(v) + \nabla \log \mu \cdot v$ to get

$$(\partial_t \alpha + D \alpha \cdot v)^i = -2\delta_i (\partial_k \log \mu \partial_j \partial_h \log \mu) - 2\delta_i (\partial_{kj} \log \mu \partial_h \log \mu)v^j$$

$$= 2\delta_i (\partial_k \log \mu \partial_j \nabla v^j + \partial_j \log \mu v^j) - 2\delta_i (\partial_j \log \mu \partial_k \log \mu)v^j$$

$$= 2\delta_i (\partial_k \log \mu \partial_{jk} v^j + \partial_h \log \mu \partial_j \log \mu \partial_h v^j)$$

$$+ 2\delta_i (\partial_k \log \mu \partial_{kj} \log \mu \partial_h v^j) - 2\delta_i (\partial_{kj} \log \mu \partial_h \log \mu)v^j$$

$$= 2\delta_i (\partial_k \log \mu \partial_{kj} v^j + \partial_h \log \mu \partial_j \log \mu \partial_h v^j) + 2\delta_{kj} \log \mu \partial_k \log \mu \partial_l v^j.$$ 

Therefore

$$\langle \partial_t \alpha + D \alpha \cdot v \rangle T_{\mu_t} = 2 \int \partial_i (\partial_k \log \mu \partial_{kj} v^j) v^i d\mu + 2 \int \partial_i (\partial_k \log \mu \partial_j \log \mu \partial_h v^j) v^i d\mu$$

$$+ 2 \int \partial_{kj} \log \mu \partial_k \log \mu \partial_i v^j v^i d\mu.$$ 

:= A_3
Let us now compute \( \partial_t \beta + D \beta_t \cdot v_t \). We have, using the continuity equation

\[
(\partial_t \beta_t + D \beta_t \cdot v_t)^i = 2\partial_{ijk} (\partial_j v^j + v^j \partial_j \log \mu) - 2\partial_{ijkk} \log \mu v^j

= 2\partial_{ijkk} v^j + 2\partial_{ij} (\partial_j \log \mu \partial_{kk} v^j + 2\partial_{kj} \log \mu \partial_k v^j + \partial_{jkk} \log \mu v^j) - 2\partial_{ijkk} \log \mu v^j

= 2\partial_{ijkk} v^j + 2\partial_{ij} \log \mu \partial_{kkk} v^j + 2\partial_{ij} \log \mu \partial_{kk} v^j + 4\partial_{ik} \log \mu \partial_k v^j + 4\partial_{kj} \log \mu \partial_k v^j

+ 2\partial_{ijkk} \log \mu v^j + 2\partial_{ijkk} \log \mu \partial_1 v^j - 2\partial_{ijkk} \log \mu v^j

= 2\partial_{ijkk} v^j + 2\partial_{ij} \log \mu \partial_{kk} v^j + 4\partial_{ik} \log \mu \partial_k v^j + 4\partial_{kj} \log \mu \partial_k v^j

+ 2\partial_{ijkk} \log \mu \partial_1 v^j

\]

Thus,

\[
\langle \partial_t \beta_t + D \beta_t, v_t \rangle_{\mathcal{T}_{\mu_t}} = 2 \int \partial_{jk} v^j \partial_i v^i \mu d\mu + 2 \int \partial_{ij} \log \mu \partial_{kk} v^j \partial_i v^i \mu d\mu + 2 \int \partial_{ij} \log \mu \partial_{kkk} v^j \partial_i v^i \mu d\mu

+ 4 \int \partial_{ijk} \log \mu \partial_k v^j \partial_i v^i \mu d\mu + 4 \int \partial_{ik} \log \mu \partial_k v^j \partial_i v^i \mu d\mu + 2 \int \partial_{ijk} \log \mu \partial_1 v^j \partial_i v^i \mu d\mu

\]

Finally we define

\[
C.1 = \int (Dv \cdot Dv)_{ij} \partial_{ij} \log \mu d\mu

C.2 = \int |\nabla \text{div}(v)|^2 d\mu

C.3 = \int |Dv \cdot \nabla \log \mu|^2 d\mu

C.4 = \int \langle \nabla \text{div}(v), Dv \cdot \nabla \log \mu \rangle d\mu,
\]

where we denote by \( Dv \cdot Dv \) the usual matrix product. In the following lines we perform a series of Integration by parts, which are all justified by point (iii) of the hypothesis.

We first integrate twice B.1 by parts.

\[
B.1 = 2 \int \partial_{jk} v^j \partial_i v^i \mu d\mu + 2 \int \partial_{jk} v^j \partial_i v^i \partial_k v^k \partial_k \mu d\mu + 2 \int \partial_{jk} v^j \partial_k v^k \partial_{jk} \mu d\mu + 2 \int \partial_{jk} v^j \partial_k v^k \partial_{jk} \mu d\mu

\]

Next, we integrate the second term once again by parts and obtain

\[
2 \int \partial_{jk} v^j \partial_i v^i \partial_k \mu d\mu = -2 \int \partial_{jk} v^j \partial_i v^i \partial_k \mu d\mu - 2 \int \partial_{ijk} v^k \partial_i v^i \partial_k \mu d\mu
\]
Plugging this back, we get

\[
B.1 = 2 \int \partial_{jk} v^j \partial_{ik} v^i d\mu + 2 \int \partial_{jk} v^j \partial_{ik} \partial_i \log \mu d\mu - 2 \int \partial_{ijk} v^j v^k \partial_i \log \mu d\mu
\]

where the last identity is obtained relabeling \( j \) with \( k \) and vice versa. Thus, \( \sum_{i=1}^{6} B.i = 2C.2 + 2C.4 + \sum_{i=3}^{6} B.i \) Now, let us integrate \( A.3 \) once by parts

\[
A.3 = 2 \int \partial_{kj} \log \mu \partial_{ik} v^i v^j d\mu = -2 \int \partial_{jkk} \log \mu \partial_i v^i v^j d\mu - 2 \int \partial_{jk} \log \mu \partial_i v^i v^j d\mu - 2 \int \partial_{ik} \log \mu \partial_{jk} v^k v^j d\mu
\]

Thus,

\[
(68) \quad A.3 + \sum_{i=1}^{6} B.i = -2C.1 + 2C.2 + 2C.4 + B.3 + B.4 + \frac{1}{2} B.5.
\]

Next, we observe that, using the product’s rule and exchanging \( j \) and \( k \), we have \( A.1 = A.1.1 + B.3 \), where

\[
A.1.1 := 2 \int \partial_j \log \mu \partial_{ijk} v^k v^i d\mu
\]

Let us now turn to \( A.2 \). Using the product rule, and exchanging \( j \) with \( k \):

\[
A.2 = 4 \int \partial_{ij} \log \mu \partial_k \log \mu \partial_j v^k v^j d\mu + 2 \int \partial_j \log \mu \partial_k \log \mu \partial_{ijk} v^k v^i d\mu
\]

Using that \( \partial_k \log \mu d\mu = \partial_k \mu d\text{vol} \), we integrate \( A.2.2 \) by parts

\[
A.2.2 = -2 \int \partial_j \log \mu \partial_{ijk} v^k v^j d\mu - 2 \int \partial_{ijk} v^k \partial_j \log \mu v^j d\mu - 2 \int \partial_{ijk} v^k \partial_k v^j d\mu
\]
Therefore \( A.1 + A.2 = B.3 - \frac{1}{2} B.5 + A.2.1 + A.2.2.1 \). Combining this with (68) we get
\[
\sum_{i=1}^{3} A.i + \sum_{i=1}^{6} B.i = -2C.1 + 2C.2 + 2C.4 + 2B.3 + B.4 + A.2.1 + A.2.2.1.
\]

Let us now integrate A.2.2.1 by parts.

\[
A.2.2.1 = 2 \int \partial_{ij} \log \mu \partial_{j} v^i d\mu + 2 \int \partial_{ij} \log \mu \partial_{l} v^i d\mu + 2 \int \partial_{ij} \log \mu \partial_{j} v^i d\mu
\]

Thus, \( \sum_{i=1}^{3} A.i + \sum_{i=1}^{6} B.i = 2C.1 + 2C.2 + 4C.4 + 2B.3 + B.4 + A.2.1. \) Lastly, we shall integrate B.4 by parts. We get
\[
B.4 = -4 \int \partial_{ij} \log \mu \partial_{k} v^j v^i d\mu - 4 \int \partial_{ij} \log \mu \partial_{k} v^i d\mu - 4 \int \partial_{ij} \log \mu \partial_{k} v^i d\mu.
\]

Hence, we can conclude that \( \sum_{i=1}^{3} A.i + \sum_{i=1}^{6} B.i = -4C.1 + 2C.2 + 2C.3 + 4C.4. \) It is an easy calculation to see that this is indeed the right hand side of (67). Since
\[
\sum_{i=1}^{3} A.i + \sum_{i=1}^{6} B.i = \langle \partial_t \beta_t + D\beta_t, v_t \rangle_{T_{\mu_t}} + \langle \partial_t \beta_t + D\beta_t, v_t \rangle_{T_{\mu_t}} = \langle D_{dt} \nabla W I_{\mu_t}, v_t \rangle_{T_{\mu_t}},
\]
the Lemma is proven for the case when \( U = 0 \). The general case follows observing that \( \nabla W \mathcal{J}_U(\mu_t) = \nabla W \mathcal{J}(\mu_t) + \nabla \mathcal{W} \). The conclusion then follows with an easy calculation.

In the next Lemma we establish log-concavity of \( \mu_t \).

**Lemma 3.11.** For all \( 0 \leq t \leq 1, \mu_t \) is a log-concave measure

**Proof.** From (1) we have:
\[
\mu_t(\cdot) = \int_{\mathbb{R}^d \times \mathbb{R}^d} P^{xy}(X_t \in \cdot)d\pi(dx dy),
\]
where \( P^{xy} \) is the bridge of \( P \) between \( x \) and \( y \). It has been show at [10, Thm 2.1] that if \( \mathbb{W} \) is convex, \( P^{xy}(X_t \in \cdot) \) is a log concave distribution for all \( x, y \). Thus, because of point (iv) of the hypothesis, \( \mu_t \) is obtained as a log-concave mixture of log-concave measures, and is therefore log-concave itself.

We can now proceed to the proof of the Theorem.
Proof. Since the hypothesis of Theorem 1.3 hold we can differentiate once in time the Fisher information to get
\[ \partial_t \mathcal{I}_U(\mu_t) = \langle \nabla^W \mathcal{I}_U(\mu_t), v_t \rangle_{T_{\mu_t}}. \]
Using Lemma 3.10 in combination with Theorem 1.3 and the compatibility with the metric we get
\[ \partial_{tt} \mathcal{I}_U(\mu_t) = \langle \frac{D}{dt} \nabla^W \mathcal{I}_U(\mu_t), v_t \rangle_{T_{\mu_t}} + 8 |\nabla^W \mathcal{I}_U(\mu_t)|_{T_{\mu_t}}^2, \]
We know from Lemma 3.11 that \( \mu_t \) is a log concave measure. Therefore, using Schur’s product Theorem [26, Th 7.5.3] in the second term of the rhs of (67) we get
\[ \langle \frac{D}{dt} \nabla^W \mathcal{I}_U(\mu_t), v_t \rangle_{T_{\mu_t}} \geq 8 \langle v_t, \text{Hess} \mathcal{W} \cdot v_t \rangle_{T_{\mu_t}} \geq 8 \alpha^2 |v_t|_{T_{\mu_t}}^2, \]
which proves the desired convexity and (19).

Proof of Theorem 1.6.

Proof. From the Feynman-Kac formula we have that \( f_t, g_t \) solve
\[ \partial_t f_t(x) = \frac{1}{2} \Delta f_t(x) - V f_t(x), \quad \partial_t g_t = -\frac{1}{2} \Delta g_t + V g_t \]
on any \( D_e \). By taking logarithms, and using the positivity of \( f_t, g_t \) we get
\[ \partial_t \log f_t(x) = \frac{1}{2} \Delta \log f_t(x) + \frac{1}{2} |\nabla \log f_t|^2 - V \]
\[ \partial_t \log g_t(x) = -\frac{1}{2} \Delta \log g_t(x) - \frac{1}{2} |\nabla \log g_t|^2 + V \]
(70)
One can check directly, as in the proof of Theorem 1.2 that \( \frac{1}{2} \nabla (\log f_t - \log g_t) \) solves the continuity equation in the classical sense. Assumption (25) ensures enough integrability to conclude that this vector field is actually the velocity field of \( (\mu_t) \). Moreover, (25) ensures that \( (\mu_t) \) is a regular curve. If we replace (51) with (70) and use the same argument as in the proof of Theorem 1.2 we get
\[ \partial_t v_t + \frac{1}{2} |v_t|^2 = \nabla^W (\frac{1}{8} \mathcal{I} + \mathcal{E}_K). \]
Then, using (24) and Lemma 3.3 we get that the covariant derivative exists and it has the desired form.

4. Appendix

The following Lemma has been used in the proof of Theorem 1.4. Here, we denote \( \dot{f}, \ddot{f} \) the first and second derivatives of a function on the real line.

Lemma 4.1. Let \( \phi : [0, 1] \to \mathbb{R} \) be twice differentiable on \( (0, 1) \) and continuous on \( [0, 1] \).
(i) If $\phi_t \geq \lambda \dot{\phi}_t$ for all $t \in (0, 1)$, then

\begin{equation}
\forall t \in [0, 1], \quad \phi_t \leq \phi_0 + (\phi_0 - \phi_1) \frac{1 - \exp(\lambda(1 - t))}{1 - \exp(-\lambda)}
\end{equation}

(ii) If $\phi_t \geq -\lambda \dot{\phi}_t$ for all $t \in (0, 1)$, then

\begin{equation}
\forall t \in [0, 1], \quad \phi_t \leq \phi_0 + (\phi_1 - \phi_0) \frac{1 - \exp(-\lambda t)}{1 - \exp(-\lambda)}
\end{equation}

Proof. Let us prove (i). Fix $\varepsilon \in (0, 1)$ and let $\psi$ be the unique solution over $[\varepsilon, 1 - \varepsilon]$ of

\begin{equation}
\begin{aligned}
\dot{\psi}_t &= \lambda \psi_t \\
\dot{\psi}_\varepsilon &= \phi_\varepsilon, \\
\psi_{1-\varepsilon} &= \phi_{1-\varepsilon}
\end{aligned}
\end{equation}

Therefore, if we define $\gamma_t := \phi_t - \psi_t$ we have that $\gamma_t$ satisfies

$\forall \in [\varepsilon, 1 - \varepsilon] \quad \dot{\gamma}_t \geq \lambda \gamma_t, \quad \gamma_\varepsilon = \gamma_{1-\varepsilon} = 0$

and $\gamma$ is of class $C_2$. Assume that $\gamma_t > 0$ for some $t \in (\varepsilon, 1 - \varepsilon)$. Then, since $\gamma_\varepsilon = 0$, $\gamma_{1-\varepsilon} = 0$, there must be some $t^* \in [\varepsilon, 1 - \varepsilon)$ such that $\gamma_{t^*} = 0$ and $\dot{\gamma}_{t^*} > 0$. We claim that for any $t \geq t^*$ we have that $\dot{\gamma}_t \geq \dot{\gamma}_{t^*} > 0$. If the claim is proven, the conclusion follows since we would have $\gamma_{1-\varepsilon} > \gamma_{t^*} = 0$. To prove the claim we proceed as follows. Let

$$t^{**} = \inf \{t > t^* \text{ s.t. } \gamma_t \leq \gamma_{t^*} \}$$

Note that since $\dot{\gamma}_{t^*} > 0$, we have $t^{**} > t^*$. Assume now $t^{**} < 1 - \varepsilon$. This implies that for some $r \in (t^*, t^{**})$ we have $\dot{\gamma}_r = 0$. But this contradicts the fact that $\dot{\gamma}_r \geq \lambda \gamma_r \geq \lambda \gamma_{t^*} > 0$. Therefore $t^{**} \geq 1 - \varepsilon$. This implies that $\gamma_t \leq 0$ on $[\varepsilon, 1 - \varepsilon]$. Moreover, $\psi$ can be computed explicitly

$$\psi_t = \phi_{1-\varepsilon} + (\phi_\varepsilon - \phi_{1-\varepsilon}) \frac{1 - \exp(\lambda(1 - \varepsilon - t))}{1 - \exp(-\lambda(1 - 2\varepsilon))}$$

Thus, $\gamma_t \leq 0$ can be rewritten as

$$\forall t \in [\varepsilon, 1 - \varepsilon] \quad \phi_t \leq \phi_{1-\varepsilon} + (\phi_\varepsilon - \phi_{1-\varepsilon}) \frac{1 - \exp(\lambda(1 - \varepsilon - t))}{1 - \exp(-\lambda(1 - 2\varepsilon))}$$

Letting $\varepsilon \to 0$ and using the continuity of $\phi$ at $t = 0, 1$ completes the proof of (i). To prove (ii), we just apply (i) to $\dot{\phi}_t = \phi_{1-t}$. \hfill \square

**Hessian of the Entropy and gradient of the Fisher information.**

**Hessian of the entropy.** In this paragraph we make some formal computations, whose aim is to give an explanation for equation (39). We assume $U = 0$ for simplicity. Let $\mu \in C^b_\infty$ and $\nabla \varphi \in C^0_\infty$ be fixed. We consider the constant speed geodesic $(\mu_t)$ such that $\mu_0 = \mu$ and $v_0 = \nabla \varphi$. Then, by definition

$$\text{Hess}^W_{\mathcal{H}_U}(\nabla \varphi) = \left. \frac{d}{dt} \nabla^W_{\mathcal{H}_U}(\mu_t) \right|_{t=0}.$$ 

Using the identification of the covariant derivative at Lemma 3.3 and (36) we have that

$$\text{Hess}^W_{\mathcal{H}_U}(\nabla \varphi) = \partial_t \nabla \log \mu_t + \nabla_{v_t} \nabla \log \mu_t \big|_{t=0}$$
Using the continuity equation
\[
\partial_t \nabla \log \mu_t = -\nabla (\frac{1}{\mu_t} \nabla \cdot (\mu_t v_t)) = -\nabla (\text{div}(v_t)) - \nabla (\nabla \log \mu, v_t)
\]
Evaluating at \( t = 0 \) and using \( v_0 = \nabla \phi \), we can rewrite the latter as
\[
-\nabla \Delta \phi - \text{Hess} \log \mu (\nabla \phi) = -\nabla \Delta \phi - \text{Hess} \phi (\nabla \log \mu).
\]
Therefore, observing that \( \nabla v_t \nabla \log \mu_t \bigg|_{t=0} = \text{Hess} \log \mu (\nabla \phi) \), we arrive at
\[
\text{Hess}_W \mathcal{H}((\nabla \phi) = -\nabla \Delta \phi - \text{Hess} \phi (\nabla \log \mu)
\]
Hence, using an integration by parts:
\[
\langle \text{Hess}_W \mathcal{H}(\nabla \phi), \nabla \phi \rangle_{\mu_t} = -\int_M \langle \nabla \phi, \nabla \Delta \phi \rangle d\mu - \int_M \langle \nabla \phi, \text{Hess} \phi (\nabla \log \mu) \rangle d\mu
\]
\[
= -\int_M \langle \nabla \phi, \nabla \Delta \phi \rangle d\mu - \int_M \langle \nabla \phi, \text{Hess} \phi (\nabla \log \mu) \rangle d\mu
\]
\[
= -\int_M \langle \nabla \phi, \nabla \Delta \phi \rangle d\mu - \frac{1}{2} \int_M \langle \nabla \phi, |\nabla \phi|^2 \rangle d\mu
\]
At this point one can use the Bochner-Weitzenböck formula
\[
\frac{1}{2} |\nabla \phi|^2 = \langle \nabla \phi, \nabla \Delta \phi \rangle + |\text{Hess} \phi|_{\text{HS}}^2 + \text{Ric}(\nabla \phi, \nabla \phi)
\]
and the hypothesis (13) to obtain the conclusion.

**Gradient of the Fisher information.** In this section, we shall make some formal computations to justify (5). As we did before, we assume \( U = 0 \) for simplicity. Differentiating the relation (37) and using the definition of Hessian we
\[
\nabla W I(\mu) = 2 \text{Hess}_W \mathcal{H}(\nabla W \mathcal{H}(\mu))
\]
By the definition of Hessian
\[
\text{Hess}_W^W \mathcal{H}(\mu) = \left. \frac{D}{dt} \nabla W \mathcal{H}(\mu_t) \right|_{t=0},
\]
where \( (\mu_t) \) is any regular enough curve such that \( \mu_0 = \mu, \nu_0 = \nabla W \mathcal{H}(\mu) \). From Lemma 3.3 such covariant is the projection on the space of gradient vector fields of
\[
\partial_t \nabla W \mathcal{H}(\mu) + \nabla v_t \nabla W \mathcal{H}(\mu)
\]
Using the continuity equation in the form \( \partial_t \log \mu_t = -\nabla \cdot v_t - v_t \cdot \nabla \log \mu_t \), and recalling that \( \nabla W \mathcal{H}(\mu) = \nabla \log (\mu) \) we arrive at
\[
\partial_t \nabla W \mathcal{H}(\mu) \bigg|_{t=0} = -\nabla (\nabla \cdot v_t + v_t \cdot \nabla \log \mu_t) \bigg|_{t=0} = -\nabla \left( \nabla \cdot (\nabla \log \mu) \right) - \nabla (\nabla \log \mu) = -\nabla \Delta \log \mu - \nabla |\nabla \log \mu|^2.$
On the other hand
\[ \nabla_{v_i} \nabla^W \mathcal{H}(\mu_t) \bigg|_{t=0} = Hess \log(\mu_t)(v_i) \bigg|_{t=0} = Hess \log(\mu)(\nabla \log \mu) = \frac{1}{2} \nabla |\nabla \log \mu|^2. \]
Therefore
\[ \partial_t \nabla^W \mathcal{H}(\mu_t) + \nabla_{v_i} \nabla^W \mathcal{H}(\mu_t) \bigg|_{t=0} = -\nabla \Delta \log \mu - \frac{1}{2} \nabla |\nabla \log \mu|^2, \]
and since the rhs this vector field is of gradient type
\[ \nabla^W \mathcal{H}(\mu) = 2 \frac{D}{dt} \nabla^W \mathcal{H}(\mu_t) \bigg|_{t=0} = -2 \nabla \Delta \log \mu - \nabla |\nabla \log \mu|^2, \]
which is (5).

4.1. **Lemmas 4.2 and 4.3.** These Lemmas are needed in the proof of Theorem 1.2 and 1.3.

**Lemma 4.2.** Let Assumption 1.1 (B) hold. Then \( T_\sigma^\alpha(\mu, \nu) < +\infty \) and the dual representation 2 holds. Moreover

(i) \( f \) and \( g \) are compactly supported and \( f_t, g_t \) globally bounded on \([0,1] \times \mathbb{R}^d\). 
(ii) For any \( 1 \leq l \leq 3 \) and \( \varepsilon \in (0,1) \), there exist constants \( A_{l,\varepsilon}, B_{l,\varepsilon} \) such that

\[ \forall x \in M \sup_{t \in [\varepsilon,1-\varepsilon]} \sup_{v_1,\ldots,v_l \in \mathbb{R}^d} |\partial_{v_1} \ldots \partial_{v_l} \log f_t(x)| \leq A_{l,\varepsilon} + B_{l,\varepsilon} |x|^l, \]

and the same conclusion holds replacing \( f_t \) by \( g_t \).

(iii) For any \( \varepsilon \in (0,1) \) there exist a constant \( A_{2,\varepsilon} \) such that

\[ \sup_{x \in \mathbb{R}^d, t \in [\varepsilon,1-\varepsilon]} \sup_{v_2,v_1 \in \mathbb{R}^d} |\partial_{v_2} \partial_{v_1} \log f_t(x)| \leq A_{2,\varepsilon}, \]

and the same conclusion holds replacing \( f_t \) by \( g_t \).

**Proof.** Since all statements concerning \( g \) are proven in the same way as those for \( f \), we limit ourselves to prove the latter ones. In the proof, we assume that \( \sigma = 1 \), the proof for the general case being almost identical. The fact that \( T_\sigma^\alpha(\mu, \nu) < +\infty \) can be easily settled using point (b) in [30, Prop. 2.5], whereas the dual representation is obtained from [29, Th 2.8]. Let us show that \( f \) is compactly supported. Observe

\[ \frac{d\mu}{d\tilde{m}} = f(x)g_0(x) \]

Since \( g_0 \in C_+^\infty \), and \( \frac{d\mu}{d\tilde{m}} \) is compactly supported, \( f \) must have the same support as \( \frac{d\mu}{d\tilde{m}} \). Moreover, since \( g_0 \) is bounded from below on the support of \( f \), the fact that \( \frac{d\mu}{d\tilde{m}} \) is bounded from above, implies that \( f \) is bounded from above. It follows from the very definition of \( f_t \) that they must be bounded as well. The proof of (i) is complete. We only do the proof of (ii) and (iii) in the case \( d = 1 \). This proof can be extended with no difficulty to the general case. We first make some preliminary observations. For \( \alpha \) fixed, the transition density of the Ornstein-Uhlenbeck semigroup is

\[ p_t(x, z) = \left( \frac{\gamma(\alpha, t)}{2\pi} \right)^{-1/2} \phi(\gamma(\alpha, t)(z - \exp(-\alpha t)x)) \]
where
\[ \phi(z) = \exp(-\frac{z^2}{2}), \quad \gamma(\alpha, t) = 2\frac{\alpha}{(1 - \exp(-2\alpha t))}. \]
The derivatives of \( \phi \) can be computed using the Hermite polynomials \((H_m)_{m \geq 0}\). We have
\[ \forall m \in \mathbb{N}, \quad \phi^m(z) = (-1)^m H_m(z) \phi(z). \]
Thus, we obtain the following formula for the \( m \)-th derivative of the transition density w.r.t. to \( x \):
\begin{equation}
\partial^m_x p_t(x, z) = (-1)^m \gamma(\alpha, t)^m \exp(-mat) H_m(z - \exp(-\alpha t)x)p_t(x, z).
\end{equation}
Finally, observe that we can rewrite \( f_t \) equivalently in the form
\begin{equation}
f_t(x) = \int_{\mathbb{R}} p_t(x, z) f(z) dz.
\end{equation}
Let us now prove (ii). Fix \( 1 \leq l \leq 3 \). Using (73), we can write \( \partial_x^l \log f_t(x) \) as a sum of finitely terms of the form
\[ f_t(x)^{-k} \prod_{j=1}^k \int f(z) \partial_x^j p_t(x, z) dz \]
where \( k \leq l \) and \( i_1, \ldots, i_k \) are integers summing up to \( l \). Plugging (72) in this expression, the desired conclusion follows using the fact that \( H_m \) is a polynomial of degree \( m \), \( f \) is compactly supported, and \( \gamma(\alpha, t) \) is uniformly bounded from above and below for \( t \in [\varepsilon, 1 - \varepsilon] \). To prove (iii), we compute explicitly \( \partial_x^2 \log f_t(x) \), using (72):
\[ \exp(-2\alpha t) \gamma(\alpha, t)^2 f_t^{-2}(x) \times \]
\[ \left( \int f(z) H_2(z - \exp(-\alpha t)x)p_t(x, z)dz \int f(z)p_t(x, z)dz - \left[ \int f(z) H_1(z - \exp(-\alpha t)x)p_t(x, z)dz \right]^2 \right). \]
Using the explicit form of the first two Hermite polynomials and some standard calculations, the latter expression is seen to be equal to
\[ \exp(-2\alpha t) \gamma(\alpha, t) f_t(x)^{-2} \times \]
\[ \left( \int f(z) z^2 p_t(x, z)dz \int f(z)p_t(x, z)dz - \left[ \int f(z)zp_t(x, z)dz \right]^2 \right) \]
The conclusion then follows using the fact that \( f \) is compactly supported and that \( \gamma(\alpha, t) \) is uniformly bounded from above and below for \( t \in [\varepsilon, 1 - \varepsilon] \).

**Lemma 4.3.** Let Assumption 4.1(A) hold. Then the dual representation 2 holds and
\begin{enumerate}[(i)]
\item \( T_{ij}^0(\mu, \nu) < \infty \)
\item For any \( \varepsilon \in (0, 1) \) \( f_t, g_t, \tilde{f}_t, \tilde{g}_t \) are \( C^{b,+}_\varepsilon \) over \( D_e \) and log \( f_t \), log \( g_t \), log \( \tilde{f}_t \), log \( \tilde{g}_t \) are \( C^b_\varepsilon \) over \( D_e \).
\end{enumerate}

**Proof.** Let \( \varphi_\mu = \frac{d\mu}{d\alpha} \varphi_\nu = \frac{d\nu}{d\alpha} \) and define \( \pi \in \Pi(\mu, \nu) \) by \( \pi(x, y) = \varphi_\mu(x) \varphi_\nu(y) \mathbf{m} \otimes \mathbf{m}(dx dy) \). The theory of Malliavin calculus ensures that \((X_0, X_1)_\# \mathbf{P} \) is an absolutely continuous measure on \( M \times M \) with positive smooth density. Since \( M \) is compact, then \((X_0, X_1)_\# \mathbf{P} \) is equivalent to \( \mathbf{m} \otimes \mathbf{m} \). Therefore, for some constant \( C < +\infty \),
\[ \mathcal{H}(\pi | (X_0, X_1)_\# \mathbf{P}) \leq C \mathcal{H}(\pi | \mathbf{m} \otimes \mathbf{m}) = C(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})) < +\infty. \]
Thus $T_h^\tau(\mu, \nu) < +\infty$. It is also a result of Malliavin calculus that $f_t, g_t$ are of class $C^+_\infty$ on any $D_e$. But then, since $M$ is compact, they are also in $C^b(\infty)$ and uniformly bounded from below, which gives that $\log f_t, \log g_t$ are in $C^b(\infty)$. The statement about $\tilde{f}_t, \tilde{g}_t$ and their logarithm follows from the one for $f_t, g_t$ and the compactness of $M$. □

Proof of Lemma 3.3

Proof. Fix $\varepsilon \in (0, 1)$. As a preliminary step, we compute $\partial_t \tau_t^\varepsilon(\xi_t)$. Using the group property we get

$$\tau_{t+h}^\varepsilon(\xi_{t+h}) - \tau_t^\varepsilon(\xi_t) = \tau_t^\varepsilon(\tau_{t+h}^\varepsilon(\xi_{t+h}) - \xi_t) = h \tau_t^\varepsilon(\partial_t \xi_t) + \tau_t^\varepsilon(\tau_{t+h}^\varepsilon(\xi_t) - \xi_t) + o(h),$$

where $o(h)/h \to 0$ as $h \to 0$. Recalling Definition 2.3 and the definition of flow map we get

$$\tau_{t+h}^\varepsilon(\xi_t)(x) - \xi_t(x) = (\tau_{t+h}^\varepsilon(\xi_t) \circ T(t, t + h, x) - \xi_t(x)) = h \nabla v_t T(x, t, x) \xi_t(x) + o(h) = h \nabla v_t \tau_t^\varepsilon(\xi_t(x) + o(h).$$

Therefore, we have shown that, as a pointwise limit

$$(74) \quad \lim_{h \to 0} \frac{\tau_{t+h}^\varepsilon(\xi_{t+h}) - \xi_t}{h} = \partial_t \tau_t^\varepsilon + \nabla v_t \xi_t,$$

which implies that

$$\partial_t \tau_t^\varepsilon(\xi_t) = \tau_t^\varepsilon(\partial_t \xi_t + \nabla v_t \xi_t).$$

Let us now prove the absolute continuity of $(\xi_t)$ along $(\mu_t)$ using what we have just shown. We have

$$\left| \tau_s^\varepsilon(\xi_s) - \tau_t^\varepsilon(\xi_t) \right|_{L^2_{\mu_t}}^2 \leq \left( \int_M \left| \tau_s^\varepsilon(\partial_r \xi_r + \nabla v_r \xi_r) \right|^2 d\mu_r \right)^{1/2} \leq (s-t)^{1/2} \left( \int_M \left| \tau_s^\varepsilon(\partial_r \xi_r + \nabla v_r \xi_r) \right|^2 d\mu_r dr \right)^{1/2} = (s-t)^{1/2} \left( \int_M \left| \partial_r \xi_r + \frac{1}{2} \nabla v_r \xi_r \right|^2 d\mu_r dr \right)^{1/2} \leq (s-t) \sup_{r \in [0,1]} \left| \partial_r \xi_r + \nabla v_r \xi_r \right|_{T_{\mu_r}} \leq (s-t) \left| \partial_r \xi_r + \nabla v_r \xi_t \right|_{T_{\mu_r}} \leq (s-t) \left| \partial_r \xi_r + \nabla v_r \xi_r \right|_{T_{\mu_r}} \leq (s-t) \left| \partial_r \xi_r + \nabla v_r \xi_r \right|_{T_{\mu_r}} \leq (s-t) \left| \partial_r \xi_r + \nabla v_r \xi_r \right|_{T_{\mu_r}}$$

Using (15), the desired absolute continuity follows. Let us now turn to the proof of (46). By definition,

$$\frac{d}{dt} \xi_t = \lim_{h \to 0} \frac{\tau_{t+h}^\varepsilon(\xi_{t+h}) - \xi_t}{h},$$

where the limit is in $L^2_{\mu_t}$. But then, it is also the pointwise limit along a subsequence. Such computation has been done at (74), and yields the desired result. The identity (47) is a direct consequence of (46) and the fact that $v_t$ is a gradient vector field. □
REFERENCES

[1] Luigi Ambrosio and Wilfrid Gangbo. Hamiltonian odes in the wasserstein space of probability measures. *Communications on Pure and Applied Mathematics*, 61(1):18–53, 2008.

[2] Luigi Ambrosio and Nicola Gigli. A user’s guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013.

[3] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348. Springer Science & Business Media, 2013.

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

[5] Patrick Cattiaux and Christian Léonard. Minimization of the kullback information of diffusion processes. In *Annales de l’IHP Probabilités et statistiques*, volume 30, pages 83–132, 1994.

[6] 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.

[7] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Optimal steering of a linear stochastic system to a final probability distribution, part i. *IEEE Transactions on Automatic Control*, 61(5):1158–1169, 2016.

[8] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Optimal steering of a linear stochastic system to a final probability distribution, part i. *IEEE Transactions on Automatic Control*, 61(5):1158–1169, 2016.

[9] J.M.C. Clark. A local characterization of reciprocal diffusions. *Applied Stochastic Analysis*, 5:45–59, 1991.

[10] Giovanni Conforti. Fluctuations of bridges, reciprocal characteristics, and concentration of measure. *preprint* arXiv:1602.07231, to appear in *Annales de l’Institut Henri Poincaré*, 2016.

[11] Giovanni Confotri and M. Von Renesse. Couplings,gradient estimates and logarithmic Sobolev inequality for Langevin bridges. *preprint*, available at https://arxiv.org/abs/1612.08546.

[12] Giovanni Confotri and Christian Léonard. Reciprocal classes of random walks on graphs. *Stochastic Processes and their Applications*, 127(6):1870 – 1896, 2017.

[13] A.B. Cruzeiro and J.C. Zambrini. Malliavin calculus and Euclidean quantum mechanics. I. Functional calculus. *Journal of Functional Analysis*, 96(1):62–95, 1991.

[14] P. Dai Pra. A stochastic control approach to reciprocal diffusion processes. *Applied Mathematics and Optimization*, 23(1):313–329, 1991.

[15] D. Dawson, L. Gorostiza, and A. Wakolbinger. Schrödinger processes and large deviations. *Journal of mathematical physics*, 31(10):2385–2388, 1990.

[16] Donald A Dawson and Jürgen Gärtner. Multilevel large deviations and interacting diffusions. *Probability Theory and Related Fields*, 98(4):423–487, 1994.

[17] Manfredo Perdigao Do Carmo and J Flaherty Francis. *Riemannian geometry*, volume 115. Birkhäuser Boston, 1992.

[18] H. Föllmer. Random fields and diffusion processes. In *École d’Été de Probabilités de Saint-Flour XV–XVII*, 1985–87, pages 101–203. Springer, 1988.

[19] Hans Föllmer, Nina Gantert, et al. Entropy minimization and schrödinger processes in infinite dimensions. *The Annals of probability*, 25(2):901–926, 1997.

[20] Alfred Galichon and Bernard Salanié. Matching with trade-offs: Revealed preferences over competing characteristics. 2010.

[21] Ivan Gentil, Christian Léonard, and Luigia Ripani. About the analogy between optimal transport and minimal entropy. *arXiv preprint arXiv:1510.08230*, 2015.

[22] Ugo Gianazza, Giuseppe Savaré, and Giuseppe Toscani. The wasserstein gradient flow of the fisher information and the quantum drift-diffusion equation. *Archive for rational mechanics and analysis*, 194(1):133–220, 2009.

[23] Nicola Gigli. *Second Order Analysis on (P2(M),W2)*. Memoirs of the American Mathematical Society. American Mathematical Soc.

[24] Nicola Gigli and Luca Tamanini. Second order differentiation formula on compact RCD*(K, N) spaces. *arXiv preprint arXiv:1701.03932*, 2017.
[25] Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, and Prasad Tetali. Displacement convexity of entropy and related inequalities on graphs. *Probability Theory and Related Fields*, 160(1-2):47–94, 2014.

[26] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[27] A. J. Krener. Reciprocal diffusions and stochastic differential equations of second order. *Stochastics*, 107(4):393–422, 1988.

[28] A.J. Krener. Reciprocal diffusions in flat space. *Probability Theory and Related Fields*, 107(2):243–281, 1997.

[29] C. Léonard. From the Schrödinger problem to the Monge–Kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1919, 2012.

[30] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014.

[31] C. Léonard, S. Rœlly, and J.C. Zambrini. Reciprocal processes. A measure-theoretical point of view. *Probability Surveys*, 11:237–269, 2014.

[32] Christian Léonard. On the convexity of the entropy along entropic interpolations. *arXiv preprint arXiv:1310.1274*, 2013.

[33] Christian Léonard. Some properties of path measures. In *Séminaire de Probabilités XLVI*, pages 207–230. Springer, 2014.

[34] Christian Léonard et al. Lazy random walks and optimal transport on graphs. *The Annals of Probability*, 44(3):1864–1915, 2016.

[35] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, pages 903–991, 2009.

[36] E. Nelson. *Dynamical theories of Brownian motion*, volume 2. Princeton university press Princeton, 1967.

[37] Felix Otto and Cédric Villani. Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality. *Journal of Functional Analysis*, 173(2):361–400, 2000.

[38] S. Rœlly and M. Thieullen. Duality formula for the bridges of a brownian diffusion: Application to gradient drifts. *Stochastic Processes and their Applications*, 115(10):1677–1700, 2005.

[39] Karl-Theodor Sturm. On the geometry of metric measure spaces. *Acta mathematica*, 196(1):65–131, 2006.

[40] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
[52] Max-K von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy and ricci curvature. *Communications on pure and applied mathematics*, 58(7):923–940, 2005.

[53] A. Wakolbinger. A simplified variational characterization of Schrödinger processes. *Journal of Mathematical Physics*, 30(12):2943–2946, 1989.

[54] J.C. Zambrini. Variational processes and stochastic versions of mechanics. *Journal of Mathematical Physics*, 27(9):2307–2330, 1986.

Laboratoire de Mathématiques Paul Painlevé, UMR CNRS 8524, Université Lille1, 59655 Villeneuve d’Ascq Cedex, France.

E-mail address: giovanniconfort@gmail.com