Christian Léonard

On the convexity of the entropy along entropic interpolations

Introduction

Displacement convexity of relative entropy plays a crucial role in the Lott-Sturm-Villani (LSV) theory of curvature lower bounds of metric measure spaces [31, 41, 42, 44] and the related Ambrosio-Gigli-Savaré gradient flow approach [1, 2].

Let us explain shortly what is meant by displacement convexity of the relative entropy. Let (X, d, m) be a metric measure space and $P(X)$ denote the space of all Borel probability measures on X. The distance d is the basic ingredient to construct the displacement interpolations and the positive measure m is used as the relative entropy reference measure, defined by $H(\mu|m) := \int_X \log(d\mu/dm) \, d\mu \in (\infty, \infty], \mu \in P(X)$. Displacement convexity means that along any displacement interpolation $[\mu_0, \mu_1]^{\text{disp}} = (\mu_t)_{0 \leq t \leq 1}$ between two probability measures μ_0 and μ_1, the entropy

$$H(t) := H(\mu_t|m), \quad 0 \leq t \leq 1$$

of the interpolation as a function of time admits some convexity lower bound.

Let us recall briefly what displacement interpolations are. The Wasserstein pseudo-distance W_2 is the square root of the optimal quadratic transport cost between μ_0 and $\mu_1 \in P(X)$ given by $W_2^2(\mu_0, \mu_1) := \inf_\pi \int_{X^2} d^2 \, d\pi$ where the infimum is taken over all the couplings $\pi \in P(X^2)$ of μ_0 and μ_1. It becomes a distance on the subset $P_2(X)$ of all $\mu \in P(X)$ such that $\int_X d^2(x_o, x) \mu(dx) < \infty$, for some $x_o \in X$. Displacement interpolations are the geodesics of the metric space $(P_2(X), W_2)$. They were introduced in McCann's PhD thesis [33] together with the notion of displacement convexity of the entropy. Related convex inequalities turn out to be functional and geometric inequalities (Brunn-Minkowski, Prekopa-Leindler, Borell-Brascamp-Lieb, logarithmic Sobolev, Talagrand inequalities), see [12, 34].

As a typical result, it is known [12, 38, 43] that a Riemannian manifold has a nonnegative Ricci curvature if and only if $t \mapsto H(\mu_t| \text{vol})$ is convex along any displacement interpolation $(\mu_t)_{0 \leq t \leq 1}$.

An important hypothesis of the LSV theory is that the metric space is geodesic. This rules discrete spaces out. Something new must be found to develop an analogue of this theory in the discrete setting. Several attempts in this direction were proposed recently. Maas and Mielke [18, 19, 32, 35, 36] have discovered a Riemannian distance

Christian Léonard: Modal-X. Université Paris Ouest. Bât.G, 200 av. de la République. 92001 Nanterre, France

on the set of probability measures such that the evolution of continuous-time Markov chains on a graph are gradient flows of some entropy with respect to this distance. Bonciocat and Sturm in [7] and Gozlan, Roberto, Samson and Tetali in [23] proposed two different types of interpolations on the set of probability measures on a discrete metric measure graph which play the role of displacement interpolations in the LSV theory. The author recently proposed in [29] a general procedure for constructing displacement interpolations on graphs by slowing down another type of interpolation, called *entropic interpolation*.

Entropic interpolations are the main actors of the present paper. Analogously to the displacement interpolations which solve dynamical transport problems, they solve dynamical entropy minimization problems. Although related to the purpose of this article, [29] is mainly concerned with the slowing down asymptotic to build displacement – rather than entropic – interpolations. Displacement interpolations are connected to optimal transport, while entropic interpolations are connected to minimal entropy.

In this article, we consider the entropic interpolations in their own right, without slowing down. The goal remains the same: studying convexity of the entropy along these interpolations to derive curvature lower bounds of some Markov generators and state spaces. Our main result states that along any entropic interpolation $[\mu_0, \mu_1] = (\mu_t)_{0 \le t \le 1}$ between two probability measures μ_0 and μ_1, the first and second derivatives of the relative entropy $H(t) := H(\mu_t | m)$ of μ_t with respect to some reference measure m can be written as

$$H'(t) = \int_X (\overrightarrow{\Theta} \psi_t - \overleftarrow{\Theta} \phi_t) \, d\mu_t, \qquad H''(t) = \int_X (\overrightarrow{\Theta}_2 \psi_t + \overleftarrow{\Theta}_2 \phi_t) \, d\mu_t. \tag{6.1}$$

The arrows on Θ and Θ_2 are related to the forward and backward directions of time. The functions ϕ and ψ depend on the endpoints μ_0 and μ_1, but the nonlinear operators $\overrightarrow{\Theta}$, $\overleftarrow{\Theta}$, $\overrightarrow{\Theta}_2$ and $\overleftarrow{\Theta}_2$ only depend on some reference m-stationary Markov process. For instance, when this process is the Brownian diffusion with generator $L = (\Delta - \nabla V \cdot \nabla)/2$ on a Riemannian manifold with $m = e^{-V} \text{vol}$ its reversing measure, we show that

$$\overrightarrow{\Theta} = \overleftarrow{\Theta} = \Gamma/2, \quad \overrightarrow{\Theta}_2 = \overleftarrow{\Theta}_2 = \Gamma_2/2 \tag{6.2}$$

are respectively half the carré du champ $\Gamma u := L(u^2) - 2uLu$ and the iterated carré du champ $\Gamma_2 u := L\Gamma u - 2\Gamma(u, Lu)$ which were introduced by Bakry and Émery in their article [5] on hypercontractive diffusion processes. In the discrete case, these operators are not linked to Γ and Γ_2 anymore but they depend on L in a different manner, see (6.16) and (6.17). In the Riemannian case, Γ_2 is related to the Ricci curvature via Bochner's formula (6.34). This is the main connection between geometry and displacement convexity. Because of the tight relation (6.2) between (Θ, Θ_2) and (Γ, Γ_2) in the Riemannian case, it is natural to expect that (Θ, Θ_2) has also some geometric content in the discrete case. Does it allow for an efficient definition of curvature of graphs? This

question is left open in the present paper. However, based on Θ and Θ_2, we give a unified proof of the logarithmic Sobolev inequality on manifolds and a modified version on graphs. This is a clue in favor of the relevance of the entropic interpolations.

An advantage of the entropic calculus, compared to its displacement analogue, is that entropic interpolations are regular in general. This is in contrast with displacement interpolations. Otto's informal calculus provides heuristics that need to be rigorously proved by means of alternate methods on the other hand, entropic interpolations allow for a direct rigorous calculus. As a matter of fact, it is proved in [26, 29] that displacement interpolations are semiclassical limits of entropic interpolations: entropic interpolations are regular approximations of displacement interpolations.

The entropic approach is not only available for reversible reference Markov dynamics, but also for stationary dynamics (in the reversible case, the time arrows on Θ and Θ_2 disappear).

The article's point of view is probabilistic. Its basic ingredients are measures on path spaces: typically, Brownian diffusions on manifolds and random walks on graphs. The Markov property plays a crucial role and we take advantage of its time symmetry. Recall that it states that conditionally on the knowledge of the present state, past and future events are independent. In particular, a time-reversed Markov process is still Markov. Time-reversal stands in the core of our approach (even when the reference stochastic process is assumed to be reversible); it explains the forward and backward arrows on Θ and Θ_2. In contrast with analytic approaches, very few probabilistic attempts have been implemented to explore geometric and functional inequalities. In this respect, let us cite the contributions of Cattiaux [10] and Fontbona and Jourdain [22] where stochastic calculus is central. In the present article, stochastic calculus is secondary. The symmetry of the Markov property allows us to proceed with basic measure theoretical technics.

The drawback of the entropic interpolations in the discrete setting is that the second derivative of the entropy along the interpolations is difficult to handle in practice, see Proposition 6.4.4 for instance. As a consequence, no specific examples based on random walks on graphs are treated in the present article. There is still some (hard) work to do to proceed in this direction.

Outline of the chapter

The article is organized as follows. Entropic interpolations are discussed in Section 6.1. In particular, we describe their dynamics. This allows us to write their equation of motion and derive our main abstract result (6.1) in Section 6.2. This abstract result is exemplified with Brownian diffusion processes in Section 6.3 and random walks on a countable graph in Section 6.4. A unified proof of an entropy-entropy production inequality (logarithmic Sobolev inequality), in both the Riemannian and discrete graph settings, is given in Section 6.5 in connection with convergence to equilibrium. The sta-

tionary non-reversible case is investigated. In Section 6.6, we propose heuristics based on thought experiments in order to grasp the link between entropic and displacement interpolations. Finally, we address several open questions related to curvature and entropic interpolations, keeping in mind the LSV theory of geodesic measure spaces as a guideline. This article is a preliminary step and these open questions should be seen as part of a program toward an analogue of the LSV theory that is available in a discrete setting.

Notation

For any measurable space Y, we denote by $P(Y)$ and $M_+(Y)$ the sets of all probability measures and positive measures on Y. Let $\Omega = D([0, 1], X)$ be the space of all right-continuous and left-limited paths from the time interval $[0, 1]$ taking their values in a Polish state space X endowed with its Borel σ-field. The canonical process $X = (X_t)_{0 \le t \le 1}$ is defined by $X_t(\omega) = \omega_t \in X$ for all $0 \le t \le 1$ and $\omega = (\omega_t)_{0 \le t \le 1} \in \Omega$. As usual, Ω is endowed with the canonical σ-field $\sigma(X_t, 0 \le t \le 1)$ generated by the canonical process[6.1]. For any $\mathcal{T} \subset [0, 1]$ and any positive measure Q on Ω, we denote $X_{\mathcal{T}} = (X_t)_{t \in \mathcal{T}}$, $Q_{\mathcal{T}} = (X_{\mathcal{T}})_\# Q$, $\sigma(X_{\mathcal{T}})$ is the σ-field generated by $X_{\mathcal{T}}$. In particular, for each $t \in [0, 1]$, $Q_t = (X_t)_\# Q \in M_+(X)$ and $Q_{01} = (X_0, X_1)_\# Q \in M_+(X^2)$.

6.1 Entropic interpolation

Entropic interpolations are defined and their equations of motion are derived. The continuous and discrete space cases are both imbedded in the same abstract setting. This prepares the next section where the first and second derivatives of $H(t) := H(\mu_t|m)$ are calculated.

(f, g)-transform of a Markov process

The h-transform of a Markov process was introduced by Doob [16] in 1957. It is revisited and slightly extended in two directions:

1. We consider a two-sided version of the h-transform, which we call (f, g)-transform, taking advantage of the invariance of the Markov property with respect to time reversal;

6.1 The canonical process is the standard naming by probabilists of the evaluation process of the analysts. The usual Skorokhod topology of the path space is useless in this article; only its measurable structure is used.

2. We extend the notion of Markov property to unbounded positive measures on Ω, having in mind two typical examples:
   – The reversible Wiener process on $\mathbb{R}^n$, i.e. the law of the Brownian motion with Lebesgue measure as its initial law;
   – The reversible simple random walk on a countable locally finite graph.

Assumptions on *R*

We specify some $R \in M_+(\Omega)$ which we call the *reference path measure* and we assume that it is Markov and *m*-stationary[6.2] where $m \in M_+(X)$ is a σ-finite positive measure on X. See Definitions .0.4 and .0.7 at the appendix 6.6.

Definition 6.1.1 ((*f*, *g*)-transform)**.** *Let $f_0, g_1 : X \to [0, \infty)$ be two nonnegative functions such that $f_0 \in L^p(m)$ and $g_1 \in L^{p^\star}(m)$ with $p, p^\star \in [1, \infty]$ and $1/p + 1/p^\star = 1$. The (f, g)-transform $P \in P(\Omega)$ of R associated with the couple of functions (f_0, g_1) is defined by*

$$P := f_0(X_0)g_1(X_1)\, R \in P(\Omega) \tag{6.3}$$

with $\int_{X^2} f_0(x)g_1(y)\, R_{01}(dxdy) = 1$ for P to be a probability measure.

Note that $f_0(X_0)g_1(X_1) \in L^1(R)$. Indeed, denoting $F = f_0(X_0)$ and $G = g_1(X_1)$, we see with $R_0 = R_1 = m$ that $\|F\|_{L^p(R)} = \|f_0\|_{L^p(m)}$, $\|G\|_{L^{p^\star}(R)} = \|g_1\|_{L^{p^\star}(m)}$ and $E_R(FG) \le \|F\|_p\|G\|_{p^\star} = \|f_0\|_p\|g_1\|_{p^\star} < \infty$.

Remarks 6.1.2. *Let us write some easy facts about (f, g)-transforms and h-transforms.*
(a) Taking $f_0 = \rho_0$ and $g_1 = 1$ gives us $P = \rho_0(X_0)R$ which is the path measure with initial marginal $\mu_0 = \rho_0\, m \in P(X)$ and the same forward Markov dynamics as R.
(b) Symmetrically, choosing $f_0 = 1$ and $g_1 = \rho_1$ corresponds to $P = \rho_1(X_1)R$ which is the path measure with final marginal $\mu_1 = \rho_1\, m \in P(X)$ and the same backward Markov dynamics as R.
(c) Doob's h-transform is an extension of item (b). It is defined by $P = h(X_\tau)\, R_{[0,\tau]}$ where τ is a stopping time and h is such that P is a probability measure.

Let us denote the backward and forward Markov transition kernels of *R* by

$$\begin{cases} \overleftarrow{r}(s, \cdot; t, z) &:= R(X_s \in \cdot \mid X_t = z), \quad 0 \le s \le t \le 1, \\ \overrightarrow{r}(t, z; u, \cdot) &:= R(X_u \in \cdot \mid X_t = z), \quad 0 \le t \le u \le 1, \end{cases}$$

6.2 This stationarity assumption is not necessary and one could replace m by R_t everywhere. It simply makes things more readable and shortens some computations.

and for m-almost all $z \in \mathbf{X}$,

$$\begin{cases} f_t(z) & := & E_R(f_0(X_0) \mid X_t = z) & = & \int_{\mathbf{X}} f_0(x) \overleftarrow{r}(0, dx; t, z) \\ g_t(z) & := & E_R(g_1(X_1) \mid X_t = z) & = & \int_{\mathbf{X}} \overrightarrow{r}(t, z; 1, dy) g_1(y). \end{cases} \tag{6.4}$$

The relation between (f_0, g_1) and the endpoint marginals $\mu_0, \mu_1 \in P(\mathbf{X})$ is

$$\begin{cases} \rho_0 & = & f_0 g_0, \quad m\text{-a.e.} \\ \\ \rho_1 & = & f_1 g_1, \quad m\text{-a.e.} \end{cases} \tag{6.5}$$

where the density functions ρ_0 and ρ_1 are defined by $\begin{cases} \mu_0 & = & \rho_0 \, m \\ \mu_1 & = & \rho_1 \, m \end{cases}$. The system of equations (6.5) was exhibited by Schrödinger in 1931 [39] in connection with the entropy minimization problem (S_{dyn}) below.

Entropic interpolation

Let us introduce the main notion of this article.

Definition 6.1.3 (Entropic interpolation). *Let $P \in P(\Omega)$ be the (f, g)-transform of R given by (6.3). Its flow of marginal measures*

$$\mu_t := (X_t)_\# P \in P(\mathbf{X}), \quad 0 \le t \le 1,$$

is called the R-entropic interpolation between μ_0 and μ_1 in $P(\mathbf{X})$ and is denoted by

$$[\mu_0, \mu_1]^R := (\mu_t)_{t \in [0,1]}.$$

When the context is clear, we simply write $[\mu_0, \mu_1]$, dropping the superscript R.

Next theorem tells us that μ_t is absolutely continuous with respect to the reference measure m on $\mathbf{X}$. We denote by

$$\rho_t := d\mu_t/dm, \quad 0 \le t \le 1,$$

its density with respect to m.

Identities (6.5) extend to all $0 \le t \le 1$ as next result shows.

Theorem 6.1.4. *Let $P \in P(\Omega)$ be the (f, g)-transform of R given by (6.3). Then, P is Markov and for all $0 \le t \le 1$, $\mu_t = \rho_t \, m$ with*

$$\rho_t = f_t g_t, \quad m\text{-a.e.} \tag{6.6}$$

Proof. Without getting into detail, the proof works as follows. As $m = R_t$,

$$\rho_t(z) := \frac{dP_t}{dR_t}(z) = E_R\left[\frac{dP}{dR}\Big|X_t = z\right] = E_R\left[f_0(X_0)g_1(X_1)|X_t = z\right]$$

$$= E_R\left[f_0(X_0)|X_t = z\right] E_R\left[g_1(X_1)|X_t = z\right] =: f_t g_t(z).$$

The Markov property of R at the last but one equality is essential in this proof. For more detail, see [28, Thm. 3.4]. □

With (6.6), we see that an equivalent analytical definition of $[\mu_0, \mu_1]^R$ is as follows.

Theorem 6.1.5. *The entropic interpolation* $[\mu_0, \mu_1]^R$ *satisfies* $\mu_t = \rho_t\, m$ *with*

$$\rho_t(z) = \int_X f_0(x)\,\overleftarrow{r}(0, dx; t, z) \int_X \overrightarrow{r}(t, z; 1, dy)\, g_1(y), \quad \text{for } m\text{-a.e. } z \in X, \qquad (6.7)$$

for each $0 \le t \le 1$.

Marginal flows of bridges are (possibly degenerate) entropic interpolations

Let us have a look at the R-entropic interpolation between the Dirac measures δ_x and δ_y.

(a) When $R_{01}(x, y) > 0$, $[\delta_x, \delta_y]^R$ is the time marginal flow $(R_t^{xy})_{0 \le t \le 1}$ of the bridge

$$R^{xy}(\cdot) := R(\cdot \mid X_0 = x, X_1 = y), \quad x, y \in X.$$

Indeed, taking $\mu_0 = \delta_x$ and $\mu_1 = \delta_y$, a solution (f_0, g_1) of (6.5) is $f_0 = \mathbf{1}_x/R_{01}(x, y)$ and $g_1 = \mathbf{1}_y$. It follows that the corresponding (f, g)-transform is $f_0(X_0)g_1(X_1) R = R_{01}(x, y)^{-1}\mathbf{1}_{\{X_0=x, X_1=y\}} R = R^{xy}$.

(b) When $R_{01}(x, y) = 0$, $[\delta_x, \delta_y]^R$ is undefined. Indeed, Definition 6.1.3 implies that $\mu_0, \mu_1 \ll m$. Let us take the sequences of functions $f_0^n = c_n^{-1}\mathbf{1}_{B(x,1/n)}$ and $g_1^n = \mathbf{1}_{B(y,1/n)}$ with $B(x, r)$ the open ball centered at x with radius r and $c_n = R_{01}(B(x, 1/n) \times B(y, 1/n)) > 0$ the normalizing constant which is assumed to be positive for all $n \ge 1$. The corresponding (f, g)-transform of R is the conditioned probability measure $P^n = R(\cdot \mid X_0 \in B(x, 1/n), X_1 \in B(y, 1/n))$ which converges as n tends to infinity to the bridge R^{xy} under some assumptions, see [11] for instance. As a natural extension, one can see $[\delta_x, \delta_y]^R$ as the time marginal flow $t \mapsto R_t^{xy}$ of the bridge R^{xy}.

(c) When the transition kernels are absolutely continuous with respect to m, i.e. $\overrightarrow{r}(t, z; 1, dy) = \overrightarrow{r}(t, z; 1, y)\, m(dy)$, for all $(t, z) \in [0, 1) \times X$ and $\overleftarrow{r}(0, dx; t, z) = \overleftarrow{r}(0, x; t, z)\, m(dx)$, for all $(t, z) \in (0, 1] \times X$, then the density functions are equal: $\overrightarrow{r} = \overleftarrow{r} := r$. This comes from the m-stationarity of R since $R((X_t, X_{t'}) \in dz dz') = R(X_t \in dz)R(X_{t'} \in dz' \mid X_t = z) = m(dz)\overrightarrow{r}(t, z; t', z')m(dz')$

and similarly $R((X_t, X_{t'}) \in dz dz') = R(X_{t'} \in dz')R(X_t \in dz \mid X_{t'} = z') = m(dz')\overleftarrow{r}(t, z; t', z')m(dz)$. Now, (6.7) extends as

$$\frac{dR_t^{xy}}{dm}(z) = \frac{r(0, x; t, z)r(t, z; 1, y)}{r(0, x; 1, y)}, \quad 0 < t < 1.$$

(d) Note that in general, for any intermediate time $0 < t < 1$, R_t^{xy} is not a Dirac mass. This is in contrast with the standard displacement interpolation $[\delta_x, \delta_y]^{\text{disp}}$ whose typical form is $\delta_{\gamma_t^{xy}}$ where γ^{xy} is the constant speed geodesic between x and y when this geodesic is unique.

Any entropic interpolation is a mixture of marginal flows of bridges. This is expressed in (6.63). For a proof, see [28].

Schrödinger problem

The notion of entropic interpolation is related to the entropy minimization problem (S_{dyn}) below. In order to state this problem properly, let us first recall an informal definition of the relative entropy

$$H(p|r) := \int \log(dp/dr)\, dp \in (-\infty, \infty]$$

of the probability measure p with respect to the reference σ-finite measure r; see appendix 6.6 for further detail. The dynamical Schrödinger problem associated with the reference path measure $R \in M_+(\Omega)$ is

$$H(P|R) \to \min; \qquad P \in P(\Omega) : P_0 = \mu_0,\ P_1 = \mu_1. \tag{S_{dyn}}$$

It consists of minimizing the relative entropy $H(P|R)$ of the path probability measure P with respect to the reference path measure R subject to the constraints $P_0 = \mu_0$ and $P_1 = \mu_1$ where $\mu_0, \mu_1 \in P(X)$ are prescribed initial and final marginals.

We shall need a little bit more than $f_0, g_1 \in L^2(m)$ what follows. The following set of assumptions implies that some relative entropies are finite, and will be invoked in Theorem 6.1.7 below.

Assumptions 6.1.6. *In addition to $f_0, g_1 \in L^2(m)$ and the normalization condition $\int_{X^2} f_0(x)g_1(y)\, R_{01}(dxdy) = 1$, the functions f_0 and g_1 entering the definition of the (f, g)-transform P of R given in (6.3) satisfy*

$$\int_{X^2} [\log_+ f_0(x) + \log_+ g_1(y)]f_0(x)g_1(y)\, R_{01}(dxdy) < \infty, \tag{6.8}$$

where $\log_+ h := \mathbf{1}_{\{h>1\}} \log h$, and as a convention $0 \log 0 = 0$.

Theorem 6.1.7. *Under the Assumptions 6.1.6, the (f, g)-transform P of R which is defined in (6.3) is the unique solution of (S_{dyn}) where the prescribed constraints $\mu_0 = \rho_0 \, m$ and $\mu_1 = \rho_1 \, m$ are chosen to satisfy (6.5).*

Proof. See [28, Thm. 3.3]. □

Note that the solution of (S_{dyn}) may not be a (f, g)-transform of R. More detail about the Schrödinger problem can be found in the survey paper [28].

$[\mu, \mu]^R$ is not constant in time

One must be careful when employing the term interpolation with regard to entropic interpolations because in general when $\mu_0 = \mu_1 =: \mu$ the interpolation $[\mu, \mu]^R$ is not constant in time. Let us give two examples.

(a) Suppose that $R_{01}(x, x) > 0$. Then the solution of (S_{dyn}) with $\mu_0 = \mu_1 = \delta_x$ is the bridge R^{xx} and $[\delta_x, \delta_x]^R$ is the marginal flow $(R_t^{xx})_{0 \le t \le 1}$ which is not constant in general.

(b) Consider R to be the reversible Brownian motion on a compact connected Riemannian manifold $\boldsymbol{X}$ without boundary. Starting from μ at time $t = 0$, the entropic minimizer P is such that $\mu_t := P_t$ gets closer to the invariant volume measure $m = \mathrm{vol}$ on the time interval $[0,1/2]$ and then goes back to μ on the remaining time period $[1/2,1]$. Let us show this. On the one hand, under our assumptions on the manifold, assuming that the Ricci curvature is strictly positive and with Theorem 6.3.5 below, it is easy to show that the function $t \in [0, 1] \mapsto H(t) := H(\mu_t | \mathrm{vol})$ is strictly convex whenever $\mu \neq \mathrm{vol}$. And on the other hand, a time reversal argument tells us that $H(t) = H(1 - t)$, $0 \le t \le 1$. Therefore, the only constant entropic interpolation is $[\mathrm{vol}, \mathrm{vol}]^R$.

$[\mu, \mu]^{\mathrm{disp}}$ is constant in time

Unlike the entropic interpolation $[\mu, \mu]^R$, McCann's displacement interpolation $[\mu, \mu]^{\mathrm{disp}}$ has the pleasant property of being constant in time. See (6.61) below for the definition of the displacement interpolation and compare with the representation (6.63) of the entropic interpolation.

Forward and backward stochastic derivatives of a Markov measure

Since P is Markov, its dynamics is characterized by either its forward stochastic derivative and its initial marginal μ_0, or its backward stochastic derivative and its final

marginal μ_1. Before computing these derivatives, let us recall some basic facts about these notions.

Let $Q \in M_+(\Omega)$ be a Markov measure. Its forward stochastic derivative $\partial + \overrightarrow{L}^Q$ is defined by

$$[\partial_t + \overrightarrow{L}^Q_t](u)(t,z) := \lim_{h \downarrow 0} h^{-1} E_Q\big(u(t+h, X_{t+h}) - u(t, X_t) \mid X_t = z\big)$$

for any measurable function $u : [0,1] \times \boldsymbol{X} \to \mathbb{R}$ in the set Dom $\overrightarrow{L}^Q$ for which this limit exists Q_t-a.e. for all $0 \le t < 1$. In fact this definition is only approximate, we give it here as a support for understanding the relation between the forward and backward derivatives. For a precise statement see [25, §. 2]. Since the time reversed $Q^\star$ of Q is still Markov, Q admits a backward stochastic derivative $-\partial + \overleftarrow{L}^Q$ which is defined by

$$[-\partial_t + \overleftarrow{L}^Q_t]u(t,z) := \lim_{h \downarrow 0} h^{-1} E_Q\big(u(t-h, X_{t-h}) - u(t, X_t) \mid X_t = z\big)$$

for any measurable function $u : [0,1] \times \boldsymbol{X} \to \mathbb{R}$ in the set Dom $\overleftarrow{L}^Q$ for which this limit exists Q_t-a.e. for all $0 < t \le 1$. When the function u only depends on the space variable: $z \mapsto u(z)$, one denotes for all $0 < t < 1$ and $z \in \boldsymbol{X}$,

$$\overrightarrow{L}_t u(z) \quad := \quad \lim_{h \downarrow 0} h^{-1} E_Q\big(u(X_{t+h}) - u(X_t) \mid X_t = z\big),$$
$$\overleftarrow{L}_t u(z) \quad := \quad \lim_{h \downarrow 0} h^{-1} E_Q\big(u(X_{t-h}) - u(X_t) \mid X_t = z\big).$$

Notice that

$$\overleftarrow{L}^Q_t = \overrightarrow{L}^{Q^\star}_{1-t}, \quad 0 < t \le 1.$$

It is proved in [25, §. 2] that these stochastic derivatives are extensions of the extended forward and backward generators of Q in the semimartingale sense, see [15]. In particular, they offer us a natural way for computing the generators.

Forward and backward dynamics of the entropic interpolation

The dynamics in both time directions of $[\mu_0, \mu_1]^R$ are specified by the stochastic derivatives $\overrightarrow{A} := \overrightarrow{L}^P$ and $\overleftarrow{A} := \overleftarrow{L}^P$ of the Markov measure $P \in P(\Omega)$ which appears in Definition 6.1.3. For simplicity, we also denote $\overrightarrow{L}^R = \overrightarrow{L}$ and $\overleftarrow{L}^R = \overleftarrow{L}$ and assume that R is stationary, i.e. $\theta^h_\# R_{[a+h, b+h]} = R_{[0, b-a]}$ for all $0 \le a \le b \le b + h \le 1$, where θ^h is the time-shift defined for all t by $X_t \circ \theta^h = X_{t-h}$. This implies that

$$\overrightarrow{L}_t = \overrightarrow{L}, \quad \overleftarrow{L}_t = \overleftarrow{L}, \quad \forall 0 \le t \le 1,$$

meaning that the forward and backward transition mechanisms of R do not depend on t.

To derive the expressions of $\overrightarrow{A}$ and $\overleftarrow{A}$, we need to introduce the carré du champ Γ of both R and its time reversed $R^\star$. The exact definition of the extended carré du champ

Γ and its domain is given in [25, Def. 4.10]. Restricting to functions u, v on X such that u, v and uv are in Dom $\overrightarrow{L}$, we recover the standard definition

$$\begin{cases} \overrightarrow{\Gamma}(u,v) & := & \overrightarrow{L}(uv) - u\overrightarrow{L}v - v\overrightarrow{L}u \\ \overleftarrow{\Gamma}(u,v) & := & \overleftarrow{L}(uv) - u\overleftarrow{L}v - v\overleftarrow{L}u \end{cases}$$

When R is a reversible path measure, we denote $L = \overrightarrow{L} = \overleftarrow{L}$ and $\Gamma = \overrightarrow{\Gamma} = \overleftarrow{\Gamma}$, dropping the useless time arrows.

Remark 6.1.8. *In the standard Bakry-Émery setting [4, 5], X is a Riemannian manifold and one considers the self-adjoint Markov diffusion generator on $L^2(X, e^{-V} \text{vol})$ given by $\widetilde{L} = \Delta - \nabla V \cdot \nabla$, with* vol *the volume measure and $V : X \to \mathbb{R}$ a regular function. The usual definition of the carré du champ of $\widetilde{L}$ is $\widetilde{\Gamma}(u,v) := [\widetilde{L}(uv) - u\widetilde{L}v - v\widetilde{L}u]/2$. In the present paper, Γ is not divided by 2 and we consider $L = \widetilde{L}/2$, i.e.*

$$L = (-\nabla V \cdot \nabla + \Delta)/2, \tag{6.9}$$

which corresponds to an SDE driven by a standard Brownian motion. Consequently, $\Gamma(u,v) = \widetilde{\Gamma}(u,v) = \nabla u \cdot \nabla v$.

In general, even if R is a reversible path measure, the prescribed marginal constraints enforce a time-inhomogeneous transition mechanism: the forward and backward derivatives $(\partial_t + \overrightarrow{A}_t)_{0 \leq t \leq 1}$ and $(-\partial_t + \overleftarrow{A}_t)_{0 \leq t \leq 1}$ of P depend explicitly on t. It is known (see [25] for instance) that for any function $u : X \to \mathbb{R}$ belonging to some class of regular functions to be made precise,

$$\begin{cases} \overrightarrow{A}_t u(z) & = & \overrightarrow{L}u(z) + \dfrac{\overrightarrow{\Gamma}(g_t, u)(z)}{g_t(z)}, & (t,z) \in [0,1) \times X, \\[3mm] \overleftarrow{A}_t u(z) & = & \overleftarrow{L}u(z) + \dfrac{\overleftarrow{\Gamma}(f_t, u)(z)}{f_t(z)}, & (t,z) \in (0,1] \times X, \end{cases} \tag{6.10}$$

where f and g are defined as in (6.4). Because of (6.6), for any t no division by zero occurs μ_t-a.e. For (6.10) to be meaningful, it is necessary that the functions f and g are regular enough for f_t and g_t to be in the domains of the carré du champ operators. In the remainder of the paper, we shall only be concerned with Brownian diffusion processes and random walks for which the class of regular functions will be specified, see Sections 6.3 and 6.4.

Going back to (6.4), we see that the processes $f_t(X_t)$ and $g_t(X_t)$ are respectively backward and forward local R-martingales. In terms of stochastic derivatives, this is equivalent to

$$\begin{cases} (-\partial_t + \overleftarrow{L})f(t,z) = 0, & 0 < t \leq 1, \\ f_0, & t = 0, \end{cases} \qquad \begin{cases} (\partial_t + \overrightarrow{L})g(t,z) = 0, & 0 \leq t < 1, \\ g_1, & t = 1, \end{cases} \tag{6.11}$$

where these identities hold μ_t-almost everywhere for almost all t.

Assumptions 6.1.9. *We assume that the kernels $\overrightarrow{r}$, $\overleftarrow{r}$ that appear in (6.4) are (a) positivity improving and (b) regularizing:*
(a) For all $(t, z) \in (0, 1) \times \mathbf{X}$, $\overrightarrow{r}(t, z; 1, \cdot) \gg m$ and $\overleftarrow{r}(0, \cdot; t, z) \gg m$.
(b) The functions (f_0, g_1) and the kernels $\overrightarrow{r}$, $\overleftarrow{r}$ are such that $(t, z) \mapsto f(t, z), g(t, z)$ are twice t-differentiable classical solutions of the parabolic PDEs (6.11).

Assumption (a) implies that for any $0 < t < 1$, f_t and g_t are positive everywhere. In particular, we see with (6.6) that $\mu_t \sim m$. Assumption (b) will be made precise later in specific settings. It will be used when computing time derivatives of $t \mapsto \mu_t$.

Mainly because we are going to study the relative entropy $H(\mu_t|m) = \int_{\mathbf{X}} \log(f_t g_t) \, d\mu_t$, it is sometimes worthwhile to express $\overrightarrow{A}$ and $\overleftarrow{A}$ in terms of the logarithms of f and g :

$$\begin{cases} \phi_t(z) := \log f_t(z) = \log E_R(f_0(X_0) \mid X_t = z), & (t, z) \in (0, 1] \times \mathbf{X} \\ \psi_t(z) := \log g_t(z) = \log E_R(g_1(X_1) \mid X_t = z), & (t, z) \in [0, 1) \times \mathbf{X}. \end{cases} \quad (6.12)$$

In analogy with the Kantorovich potentials which appear in the optimal transport theory, we call ϕ and ψ the *Schrödinger potentials*. Under Assumption 6.1.9-(b), they are classical solutions of the "second order" Hamilton-Jacobi-Bellman (HJB) equations

$$\begin{cases} (-\partial_t + \overleftarrow{B})\phi = 0, & 0 < t \le 1, \\ \phi_0 = \log f_0, & t = 0, \end{cases} \quad \begin{cases} (\partial_t + \overrightarrow{B})\psi = 0, & 0 \le t < 1, \\ \psi_1 = \log g_1, & t = 1, \end{cases} \quad (6.13)$$

where the non-linear operators $\overrightarrow{B}$ and $\overleftarrow{B}$ are defined by

$$\begin{cases} \overrightarrow{B} u & := & e^{-u} \overrightarrow{L} e^u, \\ \overleftarrow{B} v & := & e^{-v} \overleftarrow{L} e^v, \end{cases}$$

for any functions u, v such that $e^u \in \text{Dom } \overrightarrow{L}$ and $e^v \in \text{Dom } \overleftarrow{L}$. Let us introduce the notation

$$\begin{cases} \overrightarrow{A}_\theta & := & \overrightarrow{L} + e^{-\theta} \overrightarrow{\Gamma}(e^\theta, \cdot) \\ \overleftarrow{A}_\theta & := & \overleftarrow{L} + e^{-\theta} \overleftarrow{\Gamma}(e^\theta, \cdot) \end{cases}$$

which allows us to rewrite (6.10) as $\overrightarrow{A}_t = \overrightarrow{A}_{\psi_t}$ and $\overleftarrow{A}_t = \overleftarrow{A}_{\phi_t}$, emphasizing their dependence on the Schrödinger potentials.

Forward-backward systems

The complete dynamics of $[\mu_0, \mu_1]^R$ is described by a forward-backward system. We see that

$$\begin{cases} (-\partial_t^- + \overrightarrow{A}_{\psi_t}^\star)\mu = 0, & 0 < t \le 1, \\ \mu_0, & t = 0, \end{cases} \quad \text{where} \quad \begin{cases} (\partial_t^+ + \overrightarrow{B})\psi = 0, & 0 \le t < 1, \\ \psi_1 = \log g_1, & t = 1, \end{cases} \quad (6.14)$$

and $A^\star$ is the algebraic adjoint of A : $\int_X Au \, d\mu =: \langle u, A^\star\mu \rangle$. The evolution equation for μ is understood in the weak sense, in duality with respect to a large enough class of regular functions u. Similarly,

$$\begin{cases} (\partial_t^+ + \overleftarrow{A}^\star_{\phi_t})\mu = 0, & 0 \le t < 1, \\ \mu_1, & t = 1, \end{cases} \quad \text{where} \quad \begin{cases} (-\partial_t^- + \overleftarrow{B})\phi = 0, & 0 < t \le 1, \\ \phi_0 = \log f_0, & t = 0. \end{cases} \quad (6.15)$$

We have denoted ∂^+ and ∂^- the right and left time-derivatives but we will drop the superscripts + and - in the rest of the paper.

The boundary data for the systems (6.14) and (6.15) are respectively (μ_0, g_1) and (f_0, μ_1).

6.2 Second derivative of the entropy

The aim of this section is to provide basic formulas to study the convexity of the entropy

$$t \in [0, 1] \mapsto H(t) := H(\mu_t|m) \in [0, \infty)$$

as a function of time, along the entropic interpolation $[\mu_0, \mu_1]^R$ associated with the (f, g)-transform P of R defined by (6.3). The interpolation $[\mu_0, \mu_1]^R$ is specified by its endpoint data (μ_0, μ_1) defined by (6.5). We have also seen in (6.14) and (6.15) that it is specified by either (μ_0, g_1) or (f_0, μ_1).

Basic rules of calculus and notation

We are going to use the subsequent rules of calculus and notation where we drop the subscript t as often as possible. Assumption 6.1.9 is used in a significant way: it allows us to work with everywhere defined derivatives.

- $\rho_t = f_t g_t \, m$, $\quad \phi := \log f$, $\quad \psi := \log g$.
  The first identity is Theorem 6.1.4. The others are notation changes well defined everywhere under Assumption 6.1.9.
- $\dot{u} := \partial_t u$, $\quad \dot{\mu} := \partial_t \mu$, $\quad \langle u, \eta \rangle := \int_X u \, d\eta$.
  These are simplifying notation.
- The first line of the following equalities is (6.10) and the other ones are definitions:
$$\begin{cases} \overrightarrow{A}u & = & \overrightarrow{L}u + \overrightarrow{\Gamma}(g, u)/g \\ \overrightarrow{B}u & := & e^{-u}\overrightarrow{L}e^u \\ \overrightarrow{C}u & := & \overrightarrow{B}u - \overrightarrow{L}u \end{cases} \quad , \quad \begin{cases} \overleftarrow{A}u & = & \overleftarrow{L}u + \overleftarrow{\Gamma}(f, u)/f \\ \overleftarrow{B}u & := & e^{-u}\overleftarrow{L}e^u \\ \overleftarrow{C}u & := & \overleftarrow{B}u - \overleftarrow{L}u. \end{cases}$$
- $\langle u, \dot{\mu} \rangle = \langle \overrightarrow{A}u, \mu \rangle = \langle -\overleftarrow{A}u, \mu \rangle$.
  These identities hold since μ is the time-marginal flow of the Markov law P whose forward and backward derivatives are $\overrightarrow{A}$ and $\overleftarrow{A}$.

– A short way for writing (6.11) and (6.13) is:

$$\begin{cases} \dot{g} & = & -\overrightarrow{L}\,g \\ \dot{\psi} & = & -\overrightarrow{B}\,\psi \end{cases}, \qquad \begin{cases} \dot{f} & = & \overleftarrow{L}\,f \\ \dot{\phi} & = & \overleftarrow{B}\,\phi. \end{cases}$$

Entropy production

By Assumption 6.1.9, the evolution PDEs (6.11) and (6.13) are defined in the classical sense, and the functions $\overrightarrow{I}$ and $\overleftarrow{I}$ below are well defined.

Definition 6.2.1 (Forward and backward entropy production)**.** *For each $0 \le t \le 1$, we define respectively the forward and backward entropy production at time t along the interpolation $[\mu_0, \mu_1]^R$ by*

$$\begin{cases} \overrightarrow{I}(t) & := & \int_X \overrightarrow{\Theta}\,\psi_t\,d\mu_t \\[2mm] \overleftarrow{I}(t) & := & \int_X \overleftarrow{\Theta}\,\phi_t\,d\mu_t \end{cases}$$

where for any regular enough function u,

$$\begin{cases} \overrightarrow{\Theta}\,u & := & e^{-u}\overrightarrow{\Gamma}(e^u, u) - \overrightarrow{C}\,u \\[2mm] \overleftarrow{\Theta}\,u & := & e^{-u}\overleftarrow{\Gamma}(e^u, u) - \overleftarrow{C}\,u. \end{cases} \qquad (6.16)$$

Calling the functions $\overrightarrow{I}$ and $\overleftarrow{I}$ "entropy production" is justified by Corollary 6.2.4 below.

We shall see in Section 6.3 that in the reversible Brownian diffusion setting where R is associated with the generator (6.9) we have

$$\overrightarrow{\Theta}\,u = \overleftarrow{\Theta}\,u = \Gamma(u)/2$$

with no dependence on t and as usual, one simply writes $\Gamma(u) := \Gamma(u, u)$.

Proposition 6.2.2. *Suppose that the Assumptions 6.1.6 and 6.1.9 hold. The first derivative of $t \mapsto H(t)$ is*

$$\frac{d}{dt}H(\mu_t|m) = \overrightarrow{I}(t) - \overleftarrow{I}(t), \qquad 0 < t < 1.$$

Proof. With Theorem 6.1.4 and the relative entropy definition, we immediately see that $H(t) = \langle \log \rho, \mu \rangle = \langle \phi + \psi, \mu \rangle$. It follows from our basic rules of calculus that

$$H'(t) = \left\langle \dot{\phi} + \dot{\psi}, \mu \right\rangle + \langle \phi + \psi, \dot{\mu} \rangle = \langle \overleftarrow{B}\,\phi - \overrightarrow{B}\,\psi, \mu \rangle + \left\langle -\overleftarrow{A}\,\phi + \overrightarrow{A}\,\psi, \mu \right\rangle.$$

The point here is to apply the forward operators $\overrightarrow{A} = \overrightarrow{A}_\psi$, $\overrightarrow{B}$ to ψ and the backward operators $\overleftarrow{A} = \overleftarrow{A}_\phi$, $\overleftarrow{B}$ to ϕ. This is the desired result since $(\overrightarrow{A} - \overrightarrow{B})\psi = \overrightarrow{\Theta}\,\psi$ and $(\overleftarrow{A} - \overleftarrow{B})\phi = \overleftarrow{\Theta}\,\phi$. $\qquad \square$

Notice that $\left\langle \dot\phi + \dot\psi, \mu \right\rangle = \langle \partial_t \log \rho, \mu \rangle = \langle \dot\rho/\rho, \mu \rangle = \langle \dot\rho, m \rangle = (d/dt)\langle \rho, m \rangle = 0$. Hence,

$$\left\langle \dot\phi + \dot\psi, \mu \right\rangle = \langle \overrightarrow{B}\phi - \overleftarrow{B}\psi, \mu \rangle = 0.$$

As an immediate consequence of this Proposition 6.2.2, we obtain the following Corollary 6.2.4 about the dynamics of heat flows.

Definitions 6.2.3 (Heat flows).
(a) *We call* forward heat flow *the time marginal flow of the (f, g)-transform described in Remark 6.1.2-(a). It corresponds to $f_0 = \rho_0$ and $g_1 = 1$, i.e. to $P = \rho_0(X_0)\,R$.*
(b) *We call* backward heat flow *the time marginal flow of the (f, g)-transform described in Remark 6.1.2-(b). It corresponds to $f_0 = 1$ and $g_1 = \rho_1$, i.e. to $P = \rho_1(X_1)\,R$.*

Corollary 6.2.4.
(a) *Along any forward heat flow $(\mu_t)_{0 \le t \le 1}$, we have: $\frac{d}{dt}H(\mu_t|m) = -\overleftarrow{I}(t)$, $0 < t < 1$.*
(b) *Along any backward heat flow $(\mu_t)_{0 \le t \le 1}$, we have: $\frac{d}{dt}H(\mu_t|m) = \overrightarrow{I}(t)$, $0 < t < 1$.*

Second derivative

The computations in this subsection are informal and the underlying regularity hypotheses are kept fuzzy. We assume that the Markov measure R is *nice enough* for the Schrödinger potentials ϕ and ψ to be in the domains of the compositions of the operators L, Γ, A, B and C which are going to appear below. To keep formulas to a reasonable size, we have assumed that R is m-stationary. The informal results that are stated below in Claims 1 and 2 will turn later in Sections 6.3 and 6.4 into formal statements in specific settings. We introduce

$$\begin{cases} \overrightarrow{\Theta}_2 u & := \overrightarrow{L}\,\overrightarrow{\Theta}u + e^{-u}\overrightarrow{\Gamma}\left(e^u, \overrightarrow{\Theta}u\right) + e^{-u}\overrightarrow{\Gamma}(e^u, u)\overrightarrow{B}u - e^{-u}\overrightarrow{\Gamma}(e^u\overrightarrow{B}u, u), \\ \overleftarrow{\Theta}_2 u & := \overleftarrow{L}\,\overleftarrow{\Theta}u + e^{-u}\overleftarrow{\Gamma}\left(e^u, \overleftarrow{\Theta}u\right) + e^{-u}\overleftarrow{\Gamma}(e^u, u)\overleftarrow{B}u - e^{-u}\overleftarrow{\Gamma}(e^u\overleftarrow{B}u, u), \end{cases}$$
(6.17)

provided that the function u is such that these expressions are well defined. It will be shown in Section 6.3, see (6.30), that in the Brownian diffusion setting where R is associated with the generators $\begin{cases} \overrightarrow{L} & = \overrightarrow{b}\cdot\nabla + \Delta/2 \\ \overleftarrow{L} & = \overleftarrow{b}\cdot\nabla + \Delta/2 \end{cases}$, we have $\begin{cases} \overrightarrow{\Theta}_2 & = \overrightarrow{\Gamma}_2/2 \\ \overleftarrow{\Theta}_2 & = \overleftarrow{\Gamma}_2/2 \end{cases}$,

where

$$\begin{cases} \overrightarrow{\Gamma}_2(u) & := \overrightarrow{L}\,\overrightarrow{\Gamma}(u) - 2\overrightarrow{\Gamma}(\overrightarrow{L}u, u), \\ \overleftarrow{\Gamma}_2(u) & := \overleftarrow{L}\,\overleftarrow{\Gamma}(u) - 2\overleftarrow{\Gamma}(\overleftarrow{L}u, u), \end{cases}$$
(6.18)

are the forward and backward iterated carré du champ operators.

Remark 6.2.5. *We keep the notation of Remark 6.1.8. In the standard Bakry-Émery setting, the iterated carré du champ of $\widetilde{L} = \Delta - \nabla V \cdot \nabla$ is $\widetilde{\Gamma}_2(u) = [\widetilde{L}(\widetilde{\Gamma}(u)) - 2\widetilde{\Gamma}(u, \widetilde{L}u)]/2$.*

In the present paper, we consider $L = \widetilde{L}/2$ instead of $\widetilde{L}$ and we have already checked in Remark 6.1.8 that $\widetilde{\Gamma} = \Gamma$. Consequently, $\Gamma_2 = \widetilde{\Gamma}_2$.

Introducing these definitions is justified by the following claim.

Claim 1 (Informal result). *Assume that R is m-stationary. Then,*

$$\frac{d^2}{dt^2} H(\mu_t|m) = \left\langle \overleftarrow{\Theta}_2 \phi_t + \overrightarrow{\Theta}_2 \psi_t, \mu_t \right\rangle, \quad \forall 0 < t < 1.$$

Proof. Starting from $H(t) = \langle \rho \log \rho, m \rangle$ gives $H'(t) = \langle 1 + \log \rho, \dot{\mu} \rangle = \langle \log \rho, \dot{\mu} \rangle = \left\langle -\overleftarrow{A} \phi + \overrightarrow{A} \psi, \mu \right\rangle$ and $H''(t) = \frac{d}{dt} \left\langle -\overleftarrow{A} \phi, \mu \right\rangle + \frac{d}{dt} \left\langle \overrightarrow{A} \psi, \mu \right\rangle$ with

$$
\begin{aligned}
\frac{d}{dt} \left\langle -\overleftarrow{A} \phi, \mu \right\rangle &= \left\langle \overleftarrow{A}^2 \phi - \overleftarrow{A} \dot{\phi} - \overleftarrow{\dot{A}} \phi, \mu \right\rangle \\
&= \left\langle \overleftarrow{A}(\overleftarrow{A} - \overleftarrow{B})\phi - \left(\frac{\overleftarrow{\Gamma}(\dot{f}, \cdot)}{f} - \frac{\dot{f}}{f^2} \overleftarrow{\Gamma}(f, \cdot) \right) \phi, \mu \right\rangle \\
&= \left\langle \overleftarrow{A}(\overleftarrow{A} - \overleftarrow{B})\phi - \left(\frac{\overleftarrow{\Gamma}(\overleftarrow{L}f, \cdot)}{f} - \frac{\overleftarrow{L}f}{f^2} \overleftarrow{\Gamma}(f, \cdot) \right) \phi, \mu \right\rangle \\
&= \left\langle \overleftarrow{A} \left(\frac{\overleftarrow{\Gamma}(f, \cdot)}{f} - \overleftarrow{C} \right) \phi - \left(\frac{\overleftarrow{\Gamma}(\overleftarrow{L}f, \cdot)}{f} - \overleftarrow{B} \phi \frac{\overleftarrow{\Gamma}(f, \cdot)}{f} \right) \phi, \mu \right\rangle \\
&= \left\langle \overleftarrow{\Theta}_2 \phi, \mu \right\rangle.
\end{aligned}
$$

Similarly, we obtain

$$
\begin{aligned}
\frac{d}{dt} \left\langle \overrightarrow{A} \psi, \mu \right\rangle &= \left\langle \overrightarrow{A}^2 \psi + \overrightarrow{A} \dot{\psi} + \overrightarrow{\dot{A}} \psi, \mu \right\rangle \\
&= \left\langle \overrightarrow{A}(\overrightarrow{A} - \overrightarrow{B})\psi + \left(\frac{\overrightarrow{\Gamma}(\dot{g}, \cdot)}{g} - \frac{\dot{g}}{g^2} \overrightarrow{\Gamma}(g, \cdot) \right) \psi, \mu \right\rangle \\
&= \left\langle \overrightarrow{A}(\overrightarrow{A} - \overrightarrow{B})\psi - \left(\frac{\overrightarrow{\Gamma}(\overrightarrow{L}g, \cdot)}{g} - \frac{\overrightarrow{L}g}{g^2} \overrightarrow{\Gamma}(g, \cdot) \right) \psi, \mu \right\rangle \\
&= \left\langle \overrightarrow{\Theta}_2 \psi, \mu \right\rangle, \quad\quad\quad\quad\quad\quad\quad\quad\quad (6.19)
\end{aligned}
$$

which completes the proof of the claim. $\qquad\qquad\square$

Gathering Proposition 6.2.2 and Claim 1, we obtain the following

Claim 2 (Informal result). *When R is m-stationary, for all $0 < t < 1$,*

$$
\begin{cases}
\dfrac{d}{dt} H(\mu_t|m) &= \left\langle \overrightarrow{\Theta} \psi_t - \overleftarrow{\Theta} \phi_t, \mu_t \right\rangle, \\[2ex]
\dfrac{d^2}{dt^2} H(\mu_t|m) &= \left\langle \overrightarrow{\Theta}_2 \psi_t + \overleftarrow{\Theta}_2 \phi_t, \mu_t \right\rangle.
\end{cases}
$$

6.3 Brownian diffusion process

As a first step, we compute informally the operators Θ and Θ_2 associated with the Brownian diffusion process on $X = \mathbb{R}^n$ whose forward and backward generators are given for all $0 \le t \le 1$ by

$$\overrightarrow{L} = \overrightarrow{b} \cdot \nabla + \Delta/2, \qquad \overleftarrow{L} = \overleftarrow{b} \cdot \nabla + \Delta/2. \tag{6.20}$$

Here, $z \mapsto \overrightarrow{b}(z), \overleftarrow{b}(z) \in \mathbb{R}^n$ are the forward and backward drift vector fields. The term "informally" means that we suppose that $\overrightarrow{b}$ and $\overleftarrow{b}$ satisfy some unstated growth and regularity properties which ensure the existence of R and also that Assumptions 6.1.9 are satisfied.

Then, we consider reversible Brownian diffusion processes on a compact manifold.

Dynamics of the entropic interpolations

The associated nonlinear operators are $\overrightarrow{B}u = \Delta u/2 + \overrightarrow{b} \cdot \nabla u + |\nabla u|^2/2$, $\overleftarrow{B}u = \Delta u/2 + \overleftarrow{b} \cdot \nabla u + |\nabla u|^2/2$ and $\overrightarrow{\Gamma}(u, v) = \overleftarrow{\Gamma}(u, v) = \nabla u \cdot \nabla v$ for any t and $u, v \in \mathcal{C}^2(\mathbb{R}^n)$. The expressions

$$\begin{cases} \overrightarrow{A}_t &= \Delta/2 + (\overrightarrow{b} + \nabla\psi_t) \cdot \nabla, \\ \overleftarrow{A}_t &= \Delta/2 + (\overleftarrow{b} + \nabla\phi_t) \cdot \nabla, \end{cases}$$

of the forward and backward derivatives tell us that the density $\mu_t(z) := d\mu_t/dz$ solves the following forward-backward system of parabolic PDEs

$$\begin{cases} (\partial_t - \Delta/2)\mu_t(z) + \nabla \cdot (\mu_t(\overrightarrow{b} + \nabla\psi_t))(z) = 0, & (t, z) \in (0, 1] \times X \\ \mu_0, & t = 0, \end{cases}$$

where ψ solves the HJB equation

$$\begin{cases} (\partial_t + \Delta/2 + \overrightarrow{b} \cdot \nabla)\psi_t(z) + |\nabla\psi_t(z)|^2/2 = 0, & (t, z) \in [0, 1) \times X \\ \psi_1 = \log g_1, & t = 1. \end{cases} \tag{6.21}$$

In the reverse time direction of time, we obtain

$$\begin{cases} (-\partial_t - \Delta/2)\mu_t(z) + \nabla \cdot (\mu_t(\overleftarrow{b} + \nabla\phi_t))(z) = 0, & (t, z) \in [0, 1) \times X \\ \mu_1, & t = 1, \end{cases}$$

where ϕ solves the HJB equation

$$\begin{cases} (-\partial_t + \Delta/2 + \overleftarrow{b} \cdot \nabla)\phi_t(z) + |\nabla\phi_t(z)|^2/2 = 0, & (t, z) \in (0, 1] \times X \\ \phi_0 = \log f_0, & t = 0. \end{cases} \tag{6.22}$$

We fix $\log g_1 = -\infty$ on the set where g_1 vanishes, and the boundary condition at time $t = 1$ must be understood as $\lim_{t\uparrow 1} \psi_t = \log g_1$. Similarly, we have also $\lim_{t\downarrow 0} \phi_t = \log f_0$ in $[-\infty, \infty)$.

Because of the stochastic representation formulas (6.4), the functions ϕ and ψ are the unique viscosity solutions of the above Hamilton-Jacobi-Bellman equations, see [20, Thm. II.5.1]. The existence of these solutions is ensured by the Assumptions 6.1.6.

Remarks 6.3.1.

(a) *In general, ϕ and ψ might not be classical solutions of their respective HJB equations; the gradients $\nabla\psi$ and $\nabla\phi$ are not defined in the usual sense. One has to consider the notion of P-extended gradient: $\widetilde{\nabla}^P$, that is introduced in [25]. The complete description of the Markov dynamics of P is a special case of [25, Thm. 5.4]: the forward and backward drift vector fields of the canonical process under P are respectively $\overrightarrow{b} + \widetilde{\nabla}^P\psi_t$ and $\overleftarrow{b} + \widetilde{\nabla}^P\phi_t$.*

(b) *Suppose that the stationary measure m associated with R is equivalent to the Lebesgue measure. Then, the forward and backward drift fields $\overrightarrow{b}$ and $\overleftarrow{b}$ are related to m as follows. Particularizing the evolution equations with $\mu_t = m$ and $\nabla\psi = \nabla\phi = 0$, we see that the requirement that R is m-stationary implies that*

$$\begin{cases} \nabla \cdot (m \{v^{\mathrm{os}} - \nabla \log \sqrt{m}\}) & = & 0 \\ \nabla \cdot (m\, v^{\mathrm{cu}}) & = & 0 \end{cases} \tag{6.23}$$

for all t, where $m(x) := dm/dx$, these identities are considered in the weak sense and

$$\begin{cases} v^{\mathrm{os}} & := & (\overrightarrow{b} + \overleftarrow{b})/2 \\ v^{\mathrm{cu}} & := & (\overrightarrow{b} - \overleftarrow{b})/2 \end{cases}$$

are respectively the osmotic and the (forward) current velocities of R that were introduced by E. Nelson in [37]. In fact, the first equation in (6.23) is satisfied in a stronger way, since the duality formula associated to time reversal [21] is

$$v^{\mathrm{os}} = \nabla \log \sqrt{m}. \tag{6.24}$$

An interesting situation is given by $\overrightarrow{b} = -\nabla V/2 + b_\perp$ where $V : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^2$ and the drift vector field $b_\perp$ satisfies

$$\nabla \cdot (e^{-V} b_\perp) = 0. \tag{6.25}$$

It is easily seen that

$$m = e^{-V}\mathrm{Leb}$$

is the stationary measure of R. Moreover, with (6.24), we also obtain

$$\begin{cases} \overrightarrow{b} & = & -\nabla V/2 + b_\perp \\ \overleftarrow{b} & = & -\nabla V/2 - b_\perp \end{cases}, \qquad \begin{cases} v^{\mathrm{os}} & = & -\nabla V/2 \\ v^{\mathrm{cu}} & = & b_\perp \end{cases}. \tag{6.26}$$

In dimension 2, choosing $b_\perp = e^V(-\partial_y U, \partial_x U)$ with $U : \mathbb{R}^2 \to \mathbb{R}$ a $\mathcal{C}^2$-regular function, solves (6.25). Regardless of the dimension, $b_\perp = e^V S\nabla U$, where S is a constant skew-symmetric matrix and $U : \mathbb{R}^n \to \mathbb{R}$ a $\mathcal{C}^2$-regular function, is also a possible choice. In

dimension 3 one can take $b_\perp = e^V \nabla \wedge A$ where $A : \mathbb{R}^3 \to \mathbb{R}^3$ is a $\mathcal{C}^2$-regular vector field.

Computing Θ and Θ_2

Our aim is to compute the operators $\overrightarrow{\Theta}, \overleftarrow{\Theta}, \overrightarrow{\Theta}_2$ and $\overleftarrow{\Theta}_2$ of the Brownian diffusion process R with the forward and backward derivatives given by (6.20) and (6.26):

$$
\begin{cases}
\overrightarrow{L} &= (-\nabla V \cdot \nabla + \Delta)/2 + b_\perp \cdot \nabla, \\
\overleftarrow{L} &= (-\nabla V \cdot \nabla + \Delta)/2 - b_\perp \cdot \nabla,
\end{cases}
\tag{6.27}
$$

where $\nabla \cdot (e^V b_\perp) = 0$. For simplicity, we drop the arrows and the index t during the intermediate computations.

Computation of Θ

We have $\Gamma(u) = e^{-u}\Gamma(e^u, u) = |\nabla u|^2$ and $Cu = |\nabla u|^2/2$. Therefore, $\Theta u = Cu = \Gamma(u)/2 = |\nabla u|^2/2$. This gives

$$
\begin{cases}
\overrightarrow{\Theta}_t \psi_t &= |\nabla \psi_t|^2/2, \\
\overleftarrow{\Theta}_t \phi_t &= |\nabla \phi_t|^2/2,
\end{cases}
\tag{6.28}
$$

and the entropy productions are written as follows, for all $0 < t < 1$,

$$
\begin{cases}
\overrightarrow{I}(t) &= \dfrac{1}{2}\displaystyle\int_X |\nabla \psi_t|^2 \, d\mu_t, \\
\\
\overleftarrow{I}(t) &= \dfrac{1}{2}\displaystyle\int_X |\nabla \phi_t|^2 \, d\mu_t.
\end{cases}
\tag{6.29}
$$

Recall that $f_t, g_t > 0$, μ_t-a.e. : for all $0 < t < 1$.

Computation of Θ_2

Since R is a diffusion, Γ is a derivation i.e. $\Gamma(uv, w) = u\Gamma(v, w) + v\Gamma(u, w)$, for any regular enough functions u, v and w. In particular, the two last terms in the expression of $\Theta_2 u$ simplify:

$$
e^{-u}\Gamma(e^u Bu, u) - e^{-u}\Gamma(e^u, u)Bu = \Gamma(Bu, u) = \Gamma(Lu, u) + \Gamma(Cu, u),
$$

and we get $\Theta_2 u = L\Gamma(u)/2 + \Gamma(Cu, u) - \Gamma(Lu, u) - \Gamma(Cu, u) = L\Gamma(u)/2 - \Gamma(Lu, u)$. This means that

$$
\begin{cases}
\overrightarrow{\Theta}_2 &= \overrightarrow{\Gamma}_2/2, \\
\overleftarrow{\Theta}_2 &= \overleftarrow{\Gamma}_2/2.
\end{cases}
\tag{6.30}
$$

They are precisely half the iterated carré du champ operators $\overrightarrow{\Gamma}_2$ and $\overleftarrow{\Gamma}_2$ defined at (6.18). The iterated carré du champ Γ_2^o of $L^o = \Delta/2$ is

$$\Gamma_2^o(u) = \|\nabla^2 u\|_{\text{HS}}^2 = \sum_{i,j} (\partial_{ij} u)^2$$

where $\nabla^2 u$ is the Hessian of u and $\|\nabla^2 u\|_{\text{HS}}^2 = \text{tr}\,((\nabla^2 u)^2)$ is its squared Hilbert-Schmidt norm. As $\Gamma_2(u) = \Gamma_2^o(u) - 2\nabla b(\nabla u, \nabla u)$ where $2\nabla b(\nabla u, \nabla u) = 2\sum_{i,j} \partial_i b_j\, \partial_i u \partial_j u = [\nabla + \nabla^\star] b(\nabla u, \nabla u)$ with $\nabla^\star b$ the adjoint of ∇b, it follows that

$$
\begin{cases}
\overrightarrow{\Gamma}_2(u) &=& \|\nabla^2 u\|_{\text{HS}}^2 + (\nabla^2 V - [\nabla + \nabla^\star] b_\perp)(\nabla u, \nabla u), \\
\overleftarrow{\Gamma}_2(u) &=& \|\nabla^2 u\|_{\text{HS}}^2 + (\nabla^2 V + [\nabla + \nabla^\star] b_\perp)(\nabla u, \nabla u).
\end{cases}
\tag{6.31}
$$

Regularity problems

To make Claim 2 a rigorous statement, one needs to rely upon regularity hypotheses such as Assumptions 6.1.9. *If one knows that f_0, g_1, $\overrightarrow{b}$ and $\overleftarrow{b}$ are such that the linear parabolic equations* (6.11) *admit positive* $\mathcal{C}^{2,2}((0, 1) \times \boldsymbol{X})$-*regular solutions then we are done*. Verifying this is a standard (rather difficult) problem which is solved under various hypotheses. Rather than giving details about this PDE problem which can be solved by means of Malliavin calculus, we present an example in next subsection.

Working with classical solutions is probably too demanding. The operators Θ and Θ_2 are functions of ∇u and the notion of gradient can be extended as in [25] in connection with the notion of extended stochastic derivatives, see Remark 6.3.1-(a). Furthermore, when working with integrals with respect to time, for instance when considering integrals of the entropy production (see Section 6.5), or in situations where $H(t)$ is known to be twice differentiable almost everywhere (e.g. $H(t)$ is the sum of a twice differentiable function and a convex function), it would be enough to consider dt-almost everywhere defined extended gradients. This program is not initiated in the present paper.

Reversible Brownian diffusion process on a Riemannian manifold X

We give some detail about the standard Bakry-Émery setting that already appeared in Remark 6.1.8. It corresponds to the case where $b_\perp = 0$. Let us take

$$m = e^{-V} \text{ vol} \tag{6.32}$$

where $V \in \mathcal{C}^2$ satisfies $\int_{\boldsymbol{X}} e^{-V(x)} \text{vol}(dx) < \infty$. The m-reversible Brownian diffusion process $R \in \text{M}_+(\Omega)$ is the Markov measure with the initial (reversing) measure m and the semigroup generator

$$L = (-\nabla V \cdot \nabla + \Delta)/2.$$

The reversible Brownian motion $R^o \in M_+(\Omega)$ corresponds to $V = 0$. Its stochastic derivatives are

$$\overrightarrow{L}^o = \overleftarrow{L}^o = \Delta/2 \tag{6.33}$$

and vol is its reversing measure. Since R^o is the unique solution to its own martingale problem, there is a unique R which is both absolutely continuous with respect to R^o and m-reversible.

Lemma 6.3.2. *In fact R is specified by*

$$R = \exp\left(-[V(X_0) + V(X_1)]/2 + \int_{[0,1]} \left[\Delta V(X_t)/4 - |\nabla V(X_t)|^2/8\right] dt\right) R^o.$$

Proof. To see this, let us define $\hat{R}$ by means of $d\hat{R}/dR^o := e^{-V(X_0)}Z_1$ where $Z_t = \exp\left(\int_0^t -\frac{\nabla V}{2}(X_s) \cdot dX_s - \frac{1}{2}\int_0^t |\frac{\nabla V}{2}(X_s)|^2\right)$ is a local positive forward R^o-martingale. Since $\int_X e^{-V(z)}\,\text{vol}(dz) < \infty$, it follows that $t \mapsto e^{-V(X_0)}Z_t$ is a forward R^o-supermartingale. In particular its expectation $t \mapsto E_{R^o}[e^{-V(X_0)}Z_t]$ is a decreasing function so that $\hat{R}(\Omega) = E_{R^o}(e^{-V(X_0)}Z_1) \le E_{R^o}(e^{-V(X_0)}Z_0) = \int_X e^{-V(z)}\,\text{vol}(dz) < \infty$.
On the other hand, since both R^o and $d\hat{R}/dR^o$ are invariant with respect to the time reversal X^*: $X_t^* := X_{1-t,\, t\in[0,1]}$, $\hat{R}$ is also invariant with respect to time reversal: $(X^*)_\#\hat{R} = \hat{R}$. In particular, its endpoint marginals are equal: $\hat{R}_0 = \hat{R}_1$. Consequently, $\hat{R}_{[0,t]}$ doesn't send mass to a cemetery point † outside X as time increases, for otherwise its terminal marginal $\hat{R}_1$ would give a positive mass to †, in contradiction with $\hat{R}_0(\dagger) = 0$ and $\hat{R}_0 = \hat{R}_1$. Hence, Z is a genuine forward R^o-martingale. With Itô's formula, we see that $dZ_t = -Z_t\nabla\frac{V}{2}(X_t) \cdot dX_t$, R^o-a.e. and by Girsanov's theory we know that $\hat{R}$ is a (unique) solution to the martingale problem associated with the generator $L = (-\nabla V \cdot \nabla + \Delta)/2$. Finally, we take $R = \hat{R}$ and it is easy to check that L is symmetric in $L^2(m)$, which implies that m is reversing. $\qquad\square$

Remark 6.3.3. *When L is given by (6.33), for any non-zero nonnegative functions $f_0, g_1 \in L^2(\text{vol})$, the smoothing effect of the heat kernels $\overrightarrow{r}$ and $\overleftarrow{r}$ in (6.4) allows us to define* classical *gradients $\nabla\psi_t$ and $\nabla\phi_t$ for all t in $[0, 1)$ and $(0, 1]$ respectively. We see that $\nabla\psi_t$ and $\nabla\phi_t$ are the forward and backward drift vector fields of the canonical process under P.*

The next result proposes a general setting where Assumptions 6.1.9 are satisfied. The manifold X is assumed to be compact to avoid integrability troubles.

Proposition 6.3.4. *Suppose that X is a compact Riemannian manifold without boundary and $V : X \to \mathbb{R}$ is $\mathbb{C}^4$-regular. Then, for any $\mathbb{C}^2$-regular function $u : X \to \mathbb{R}$, the function $(t, x) \mapsto u(t, x) := E_R[u(X_t) \mid X_0 = x]$ belongs to $\mathbb{C}^{1,2}((0, 1) \times X)$.*

In particular, if in addition **X** is assumed to be connected and f_0 and g_1 are non-zero nonnegative $\mathcal{C}^2$-regular functions, the functions f and g defined in (6.4) and their logarithms ϕ and ψ are classical solutions of (6.11) and (6.13).

Proof. For any path $\omega \in \Omega$ and any $0 \le t \le 1$, we denote

$$Z_t(\omega) := \exp\left(-[V(\omega_0) + V(\omega_t)]/2 + \int_0^t \left[\Delta V(\omega_s)/4 - |\nabla V(\omega_s)|^2/8\right] ds\right).$$

With Lemma 6.3.2 we see that for all $0 \le t \le 1$, $Z_t = \frac{dR_{[0,t]}}{dR_{[0,t]}^o}$ and

$$u(t,x) = \frac{E_{R^o}[u(X_t)Z_1 \mid X_0 = x]}{E_{R^o}[Z_1 \mid X_0 = x]} = e^{V(x)}\mathbb{E}[u(B_t^x)Z_t(B^x)]$$

where $(B_s^x)_{0 \le s \le 1}$ is a Brownian motion starting from x under some abstract probability measure whose expectation is denoted by $\mathbb{E}$. By means of parallel transport, it is possible to build on any small enough neighborhood U of x a coupling such that $x' \in U \mapsto B^{x'}$ is almost surely continuous with respect to the uniform topology on Ω. This coupling corresponds to $B^x = x + B^0$ in the Euclidean case. The announced x-regularity is a consequence of our assumptions which allow us to differentiate (in the usual deterministic sense) in the variable x under the expectation sign $\mathbb{E}$. On the other hand, the t-regularity is a consequence of stochastic differentiation: apply Itô's formula to $u(B_t^x)$ and take advantage of the martingale property of $Z_t(B^x)$.

With regard to the last statement, the connectivity assumption implies the positivity of f and g on $(0,1) \times \boldsymbol{X}$ if f_0 and g_1 are nonnegative and not vanishing everywhere. $\square$

We gave detail on the proof of Proposition 6.3.4 because of its ease. Nonetheless, in view of Remark 6.3.3, the requirement that f_0 and g_1 are $\mathcal{C}^2$ is not optimal. However, this restriction will not be harmful when investigating convexity properties of the entropy along interpolations.

In view of Proposition 6.3.4, we remark that under its assumptions the functions f_t, g_t belong to the domain of the carré du champ operator. Hence, for any $0 < t < 1$, the stochastic derivatives of P are well-defined on $\mathcal{C}^2(\boldsymbol{X})$ and equal to

$$\begin{cases} \overrightarrow{A}_t &= \Delta/2 + \nabla(-V/2 + \psi_t) \cdot \nabla, \\ \overleftarrow{A}_t &= \Delta/2 + \nabla(-V/2 + \phi_t) \cdot \nabla. \end{cases}$$

Here, ψ and ϕ are respectively the classical solutions (compare with Remark 6.3.1-(a)) of the HJB equations (6.21) and (6.22) with $\overrightarrow{b} = \overleftarrow{b} = -\nabla V/2$.
Bochner's formula relates the iterated carré du champ Γ_2^o of $L^o = \Delta/2$ and the Ricci curvature:

$$\Gamma_2^o(u) = \|\nabla^2 u\|_{\mathrm{HS}}^2 + \mathrm{Ric}(\nabla u) \tag{6.34}$$

where $\nabla^2 u$ is the Hessian of u and $\|\nabla^2 u\|_{\mathrm{HS}}^2 = \mathrm{tr}\,((\nabla^2 u)^2)$ is its squared Hilbert-Schmidt norm. As $\Gamma_2(u) = \Gamma_2^o(u) - [\nabla + \nabla^\star]b(\nabla u)$, it follows that

$$\Gamma_2(u) = \|\nabla^2 u\|_{\mathrm{HS}}^2 + (\mathrm{Ric} + \nabla^2 V)(\nabla u). \qquad (6.35)$$

Theorem 6.3.5 (Reversible Brownian diffusion process). *Let the reference Markov measure R be associated with L given in (6.9) on a compact connected Riemannian manifold $\boldsymbol{X}$ without boundary. It is assumed that $V : \boldsymbol{X} \to \mathbb{R}$ is $\mathcal{C}^4$-regular and f_0, g_1 are non-zero nonnegative $\mathcal{C}^2$-regular functions. Then, along the entropic interpolation $[\mu_0, \mu_1]^R$ associated with R, f_0 and g_1, we have for all $0 < t < 1$,*

$$\frac{d}{dt} H(\mu_t | e^{-V}\,\mathrm{vol}) = \left\langle \frac{1}{2}\left(|\nabla \psi_t|^2 - |\nabla \phi_t|^2 \right), \mu_t \right\rangle,$$

$$\frac{d^2}{dt^2} H(\mu_t | e^{-V}\,\mathrm{vol}) = \left\langle \frac{1}{2}\left(\Gamma_2(\psi_t) + \Gamma_2(\phi_t) \right), \mu_t \right\rangle$$

where ψ and ϕ are the classical solutions of the HJB equations (6.21) and (6.22) with $\overrightarrow{b} = \overleftarrow{b} = -\nabla V/2$, $\Gamma(u) = |\nabla u|^2$ and Γ_2 is given by (6.35).

Proof. The assumptions imply that f_0 and g_1 satisfy (6.8), and they allow us to apply Proposition 6.3.4; hence, Assumptions 6.1.9 are satisfied and Claim 2 is a rigorous result. $\qquad\square$

6.4 Random walk on a graph

Now we take as our reference measure R a continuous-time Markov process on a countable state space $\boldsymbol{X}$ with a graph structure $(\boldsymbol{X}, \sim)$. The set $\boldsymbol{X}$ of all vertices is equipped with the graph relation $x \sim y$ which signifies that x and y are adjacent, i.e. $\{x, y\}$ is a non-directed edge. The degree $n_x := \# \{y \in \boldsymbol{X}, x \sim y\}$ of each $x \in \boldsymbol{X}$ is the number of its neighbors. It is assumed that $(\boldsymbol{X}, \sim)$ is locally finite, i.e. $n_x < \infty$ for all $x \in \boldsymbol{X}$, and also that $(\boldsymbol{X}, \sim)$ is connected. This means that for any couple $(x, y) \in \boldsymbol{X}^2$ of different states, there exists a finite chain $(z_1, \ldots, z_k)$ in $\boldsymbol{X}$ such that $x \sim z_1 \sim z_2 \cdots \sim z_k \sim y$. In particular $n_x \geq 1$, for all $x \in \boldsymbol{X}$.

Dynamics of the entropic interpolation

A (time homogeneous) random walk on $(\boldsymbol{X}, \sim)$ is a Markov measure with forward derivative $\partial + \overrightarrow{L}$ defined for any function $u \in \mathbb{R}^{\boldsymbol{X}}$ by

$$\overrightarrow{L} u(x) = \sum_{y:x\sim y} (u_y - u_x) \overrightarrow{J}_x(y) =: \int_{\boldsymbol{X}} D_x u\, d\overrightarrow{J}_x, \qquad 0 \leq t < 1, x \in \boldsymbol{X},$$

where $\overrightarrow{J}_x(y)$ is the instantaneous frequency of forward jumps from x to y,

$$D_x u(y) = Du(x, y) := u(y) - u(x)$$

is the discrete gradient and

$$\overrightarrow{J}_x = \sum_{y:x\sim y} \overrightarrow{J}_x(y)\,\delta_y \in M_+(X), \qquad 0 \le t < 1,\ x \in X$$

is the forward jump kernel. Similarly, one denotes its backward derivative by

$$\overleftarrow{L} u(x) = \int_X D_x u \, d\overleftarrow{J}_x, \qquad x \in X, u \in \mathbb{R}^X.$$

For the m-stationary Markov measure R with stochastic derivatives $\overrightarrow{L}$ and $\overleftarrow{L}$, the time-reversal duality formula is

$$m(x)\overrightarrow{J}_x(y) = m(y)\overleftarrow{J}_y(x), \quad \forall x \sim y \in X.$$

The expressions $\overrightarrow{L} u$ and $\overleftarrow{L} u$ can be seen as the matrices

$$\begin{cases} \overrightarrow{L} = \left(\mathbf{1}_{\{x\sim y\}}\overrightarrow{J}_x(y) - \mathbf{1}_{\{x=y\}}\overrightarrow{J}_x(X)\right)_{x,y\in X} \\ \overleftarrow{L} = \left(\mathbf{1}_{\{x\sim y\}}\overleftarrow{J}_x(y) - \mathbf{1}_{\{x=y\}}\overleftarrow{J}_x(X)\right)_{x,y\in X} \end{cases} \tag{6.36}$$

acting on the column vector $u = [u_x]_{x\in X}$. Therefore, the solutions f and g of (6.11) are

$$f(t) = e^{t\overleftarrow{L}} f_0, \qquad g(t) = e^{(1-t)\overrightarrow{L}} g_1,$$

where $f : t \in [0, 1] \mapsto [f_x]_{x\in X}(t) \in \mathbb{R}^X$ and $g : t \in [0, 1] \mapsto [g_x]_{x\in X}(t) \in \mathbb{R}^X$ are column vectors, whenever these exponential matrices are well-defined.

Let us compute the operators B, Γ and A. By a direct computation, we obtain for all $0 < t < 1$ and $x \in X$,

$$\begin{cases} \overrightarrow{B}u(x) &=& \int_X(e^{D_x u} - 1)\,d\overrightarrow{J}_x, \\ \overleftarrow{B}u(x) &=& \int_X(e^{D_x u} - 1)\,d\overleftarrow{J}_x, \end{cases} \qquad \begin{cases} \overrightarrow{\Gamma}(u,v)(x) &=& \int_X D_x u D_x v\,d\overrightarrow{J}_x, \\ \overleftarrow{\Gamma}(u,v)(x) &=& \int_X D_x u D_x v\,d\overleftarrow{J}_x, \end{cases}$$

and

$$\begin{cases} \overrightarrow{A}_t u(x) &=& \int_X D_x u\, e^{D_x \psi_t}\,d\overrightarrow{J}_x &=& \displaystyle\sum_{y:x\sim y}[u(y) - u(x)]\frac{g_t(y)}{g_t(x)}\overrightarrow{J}_x(y), \\[2em] \overleftarrow{A}_t u(x) &=& \int_X D_x u\, e^{D_x \phi_t}\,d\overleftarrow{J}_x &=& \displaystyle\sum_{y:x\sim y}[u(y) - u(x)]\frac{f_t(y)}{f_t(x)}\overleftarrow{J}_x(y). \end{cases}$$

The matrix representation of these operators is

$$\begin{cases} \overrightarrow{A}_t = \overrightarrow{A}(g_t) = \left(\mathbf{1}_{\{x\sim y\}}\dfrac{g_t(y)}{g_t(x)}\overrightarrow{J}_x(y) - \mathbf{1}_{\{x=y\}}\displaystyle\sum_{z:z\sim x}\dfrac{g_t(z)}{g_t(x)}\overrightarrow{J}_x(z)\right)_{x,y\in X} \\[2.5em] \overleftarrow{A}_t = \overleftarrow{A}(f_t) = \left(\mathbf{1}_{\{x\sim y\}}\dfrac{f_t(y)}{f_t(x)}\overleftarrow{J}_x(y) - \mathbf{1}_{\{x=y\}}\displaystyle\sum_{z:z\sim x}\dfrac{f_t(z)}{f_t(x)}\overleftarrow{J}_x(z)\right)_{x,y\in X} \end{cases} .$$

The forward-backward systems describing the evolution of $[\mu_0, \mu_1]^R$ are

$$
\mu_t = \mu_0 \overrightarrow{\exp}\left(\int_0^t \overrightarrow{A}\left(\overleftarrow{\exp}(\int_s^1 \overleftarrow{L}_r \, dr) g_1\right) ds\right)
$$

$$
= \mu_1 \overleftarrow{\exp}\left(\int_t^1 \overleftarrow{A}\left(\overrightarrow{\exp}(\int_0^s \overrightarrow{L}_r \, dr) f_0\right) ds\right), \quad 0 \le t \le 1
$$

where the measures μ_0, μ_1 are seen as row vectors and the functions f_0, g_1 as column vectors, and the ordered exponentials are defined by

$$
\overrightarrow{\exp}\left(\int_0^s M_r \, dr\right) := \mathrm{Id} + \sum_{n \ge 1} \int_{0 \le r_1 \le \cdots \le r_n \le s} M_{r_n} \cdots M_{r_1} \, dr_1 \cdots dr_n,
$$

$$
\overleftarrow{\exp}\left(\int_s^1 M_r \, dr\right) := \mathrm{Id} + \sum_{n \ge 1} \int_{s \le r_1 \le \cdots \le r_n \le 1} M_{r_1} \cdots M_{r_n} \, dr_1 \cdots dr_n.
$$

Derivatives of the entropy

In the discrete setting, no spatial regularity is required for a function to be in the domain of the operators $L, \Gamma, \ldots$ The only required regularity is that functions f and g defined by (6.4) are twice differentiable with respect to t on the open interval $(0, 1)$. This is ensured by the following lemma.

Lemma 6.4.1. *Consider f_0, g_1 as in Definition 6.1.1 and suppose that $f_0, g_1 \in L^1(m) \cap L^2(m)$ where the stationary measure m charges every point of $\boldsymbol{X}$. Then, under the assumption that*

$$
\sup_{x \in \boldsymbol{X}}\{\overrightarrow{J}_x(\boldsymbol{X}) + \overleftarrow{J}_x(\boldsymbol{X})\} < \infty, \tag{6.37}
$$

for each $x \in \boldsymbol{X}$, $t \mapsto f(t, x)$ and $t \mapsto g(t, x)$ are $\mathcal{C}^\infty$-regular on $(0, 1)$.

Proof. For any $0 \le t \le 1$, f_t and g_t are well-defined P_t-a.e., where P is given by (6.3). But, as R is an irreducible random walk, for each $0 < t < 1$, "P_t-a.e." is equivalent to "everywhere". Since $f_0 \in L^1(m)$ and R is m-stationary, we have $\int_{\boldsymbol{X}} f_t \, dm = E_R E_R[f_0(X_0) \mid X_t] = E_R(f_0(X_0)) = \int_{\boldsymbol{X}} f_0 \, dm < \infty$, implying that $f_t \in L^1(m)$. As $\sup_x \overleftarrow{J}_x(\boldsymbol{X}) < \infty$, $\overleftarrow{L}$ is a bounded operator on $L^1(m)$. This implies that $t \in (0, 1) \mapsto f_t = \overrightarrow{\exp}(\int_0^s \overleftarrow{L}_r \, dr) f_0 \in L^1(m)$ is differentiable and $(d/dt)^k f_t = \overleftarrow{L}^k f_t$ for any k. As m charges every point, we also see that $t \in (0, 1) \mapsto f_t(x) \in \mathbb{R}$ is infinitely differentiable for every x. A similar proof works with g instead of f. $\qquad\square$

As an important consequence of Lemma 6.4.1 for our purpose, we see that the statement of Claim 2 is rigorous in the present discrete setting.

Theorem 6.4.2. *Let R be an m-stationary random walk with jump measures $\overrightarrow{J}$ and $\overleftarrow{J}$ which satisfy (6.37). Along any entropic interpolation $[\mu_0, \mu_1]^R$ associated with a couple (f_0, g_1) as in Definition 6.1.1 and such that $f_0, g_1 \in L^1(m) \cap L^2(m)$, we have for all $0 < t < 1$,*

$$\frac{d}{dt}H(\mu_t|m) = \sum_{x \in X} \left(\overrightarrow{\Theta}\psi_t - \overleftarrow{\Theta}\phi_t \right)(x)\,\mu_t(x),$$

$$\frac{d^2}{dt^2}H(\mu_t|m) = \sum_{x \in X} \left(\overrightarrow{\Theta}_2\psi_t + \overleftarrow{\Theta}_2\phi_t \right)(x)\,\mu_t(x)$$

where the expressions of $\overrightarrow{\Theta}\psi_t$ and $\overleftarrow{\Theta}\phi_t$ are given in (6.38) and the expressions of $\overrightarrow{\Theta}_2\psi_t$ and $\overleftarrow{\Theta}_2\phi_t$ are given in Proposition 6.4.4 below.

Computing Θ and Θ_2

Our aim now is to compute the operators $\overrightarrow{\Theta}$, $\overleftarrow{\Theta}$, $\overrightarrow{\Theta}_2$ and $\overleftarrow{\Theta}_2$ for a general random walk R. We use the shorthand notation $\int_X a(x, y)\,J_x(dy) = [\int_X a\,dJ](x)$ and drop the arrows and the index t during intermediate computations. The Hamilton-Jacobi operator is $Bu := e^{-u}Le^u = \int_X (e^{Du} - 1)\,dJ$ which gives

$$Cu := (B - L)u = \int_X \theta(Du)\,dJ$$

where the function θ is defined by

$$\theta(a) := e^a - a - 1, \quad a \in \mathbb{R}.$$

Compare $Cu = |\nabla u|^2/2$, noting that $\theta(a) = a^2/2 + o_{a\to 0}(a^2)$. The convex conjugate $\theta^\star$ of θ will be used in a moment. It is given by

$$\theta^\star(b) = \begin{cases} (b + 1)\log(b + 1) - b, & b > -1, \\ 1, & b = -1, \\ \infty, & b < -1. \end{cases}$$

Computation of Θ

The carré du champ is $\Gamma(u, v) = \int_X Du\,Dv\,dJ$ so that $e^{-u}\Gamma(e^u, u) = \int_X Du(e^{Du} - 1)\,dJ$. Since $a(e^a - 1) - \theta(a) = ae^a - e^a + 1 = \theta^\star(e^a - 1)$, with $\Theta u := e^{-u}\Gamma(e^u, u) - Cu$, we obtain

$$\begin{cases} \overrightarrow{\Theta}\psi_t(x) = \displaystyle\sum_{y:x\sim y} \theta^\star\left(\frac{Dg_t(x, y)}{g_t(x)} \right) \overrightarrow{J}_x(y), \\[4mm] \overleftarrow{\Theta}\phi_t(x) = \displaystyle\sum_{y:x\sim y} \theta^\star\left(\frac{Df_t(x, y)}{f_t(x)} \right) \overleftarrow{J}_x(y), \end{cases} \tag{6.38}$$

where we used $e^{D\psi_t(x,y)} - 1 = Dg_t(x,y)/g_t(x)$ and $e^{D\phi_t(x,y)} - 1 = Df_t(x,y)/f_t(x)$. The ratio Dg/g should be seen as the discrete logarithmic derivative of g. Recall that $g_t > 0$, μ_t-a.e. : we do not divide by zero. Compare with $\Theta\psi = |\nabla g/g|^2/2$, remarking that $\theta^\star(b) = b^2/2 + o_{b\to 0}(b^2)$.

Entropy production

The entropy productions are

$$
\begin{cases}
\overrightarrow{I}(t) & = & \sum_{(x,y):x\sim y} \theta^\star\left(\frac{Dg_t(x,y)}{g_t(x)}\right) \mu_t(x)\overrightarrow{J}_x(y), \\
\overleftarrow{I}(t) & = & \sum_{(x,y):x\sim y} \theta^\star\left(\frac{Df_t(x,y)}{f_t(x)}\right) \mu_t(x)\overleftarrow{J}_x(y).
\end{cases}
\tag{6.39}
$$

Computation of Θ_2

Unlike the diffusion case, no welcome cancellations occur. The carré du champ Γ is not a derivation anymore since $\Gamma(uv, w) - [u\Gamma(v, w) + v\Gamma(u, w)] = uvLw + \int_X DuDvDw\, dJ$. Furthermore, we also loose the simplifying identity $C = \Theta$. To give a readable expression of Θ_2, it is necessary to introduce some simplifying shorthand notation:

$$
\begin{cases}
\sum_{x\to y} F(\varrho a) & = & \int_X F(Du(x,y))\, J_x(dy) \\
\sum_{x\to y; x\to y'} F(\varrho a, \varrho a') & = & \int_{X^2} F(Du(x,y), Du(x, y'))\, J_x(dy)J_x(dy') \\
\sum_{x\to y\to z} F(\varrho a, \boldsymbol{b}) & = & \int_{X^2} F(Du(x,y), Du(y,z))\, J_x(dy)J_y(dz) \\
\sum_{x\to y\to z} F(\varrho a, \boldsymbol{c}) & = & \int_{X^2} F(Du(x,y), Du(x,z))\, J_x(dy)J_y(dz).
\end{cases}
$$

where we have denoted $\begin{cases} \varrho a & = & Du(x,y) \\ \varrho a' & = & Du(x,y') \\ \boldsymbol{b} & = & Du(y,z) \\ \boldsymbol{c} & = & Du(x,z) \end{cases}$. Notice that $\sum_{x\to y\to z} F(\boldsymbol{c}) =$ $\int_X F(Du(x,z))\, J_x^2(dz)$ with $J_x^2(dz) := \int_{y\in X} J_x(dy)J_y(dz)$ and $\varrho a + \boldsymbol{b} = \boldsymbol{c}$.
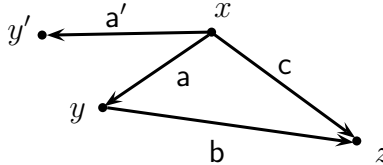
We also define the function

$$
h(a) := \theta^\star(e^a - 1) = ae^a - e^a + 1, \quad a \in \mathbb{R}. \tag{6.40}
$$

Lemma 6.4.3. *For any function $u \in \mathbb{R}^X$ and any $x \in X$,*

$$\Theta u(x) \;=\; \sum_{x \to y} h(\varrho a),$$

$$\Theta_2 u(x) \;=\; \Big(\sum_{x \to y}(e^\varrho a - 1)\Big)^2 + \sum_{x \to y}[J_y(\boldsymbol{X}) - J_x(\boldsymbol{X})]h(\varrho a)$$

$$+ \sum_{x \to y \to z}[2e^\varrho a h(\boldsymbol{b}) - h(\boldsymbol{c})].$$

Proof. The first identity is (6.38). Let us look at $\Theta_2 u(x)$. Rather than using formula (6.17), for an explicit formulation of Θ_2 in terms of J, it will be easier to go back to (6.19): $\vec{\Theta}_2 \psi_t = \vec{A}_t^2 \psi_t + \frac{d}{dt}(\vec{A}_t \psi_t)$ with $\vec{A}_t u = \int_X Du\, e^{D\psi_t}\, d\vec{J}$ and $\dot\psi = -\vec{B}\psi = -\int_X (e^{D\psi} - 1)\, d\vec{J}$.

Making use of $A\psi(x) = \sum_{x \to y} \varrho a e^\varrho a$ and $\dot\psi(x) = \sum_{x \to y} -(e^\varrho a - 1)$, we see that

$$A^2 \psi(x) \;=\; \int_X (A\psi_y - A\psi_x) e^{D\psi(x,y)} J_x(dy)$$

$$= \sum_{x \to y \to z} e^\varrho a \boldsymbol{b} e^{\boldsymbol{b}} - \sum_{x \to y; x \to y'} e^\varrho a \varrho a' e^{\varrho a'}$$

$$\frac{d}{dt}(A\psi)(x) \;=\; \int_X (D\psi(x,y)e^{D\psi(x,y)} + e^{D\psi(x,y)})(\dot\psi_y - \dot\psi_x) J_x(dy)$$

$$= -\sum_{x \to y \to z} (\varrho a e^\varrho a + \varrho a)(e^{\boldsymbol{b}} - 1) + \sum_{x \to y; x \to y'} (\varrho a e^\varrho a + \varrho a)(e^{\varrho a'} - 1)$$

where ϱa, $\varrho a'$ and $\boldsymbol{b}$ are taken with $u = \psi$. This shows that

$$\Theta_2 u(x) = \sum_{x \to y \to z} e^\varrho a \boldsymbol{b} e^{\boldsymbol{b}} - (\varrho a e^\varrho a + e^\varrho a)(e^{\boldsymbol{b}} - 1) - \sum_{x \to y; x \to y'} e^\varrho a \varrho a' e^{\varrho a'} - (\varrho a e^\varrho a + e^\varrho a)(e^{\varrho a'} - 1)$$

The functions defining the integrands rewrite as follows

$$e^a \boldsymbol{b} e^b - (ae^a + e^a)(e^b - 1) = h(a) + 2e^a h(b) - h(c) \quad \text{with} \quad c = a + b,$$

$$e^a a' e^{a'} - (ae^a + e^a)(e^{a'} - 1) = (e^a - 1)h(a') - h(a)(e^{a'} - 1)$$

$$+ h(a') - (e^a - 1)(e^{a'} - 1).$$

Hence,

$$\sum_{x \to y; x \to y'} e^\varrho a \varrho a' e^{\varrho a'} - (\varrho a e^\varrho a + e^\varrho a)(e^{\varrho a'} - 1)$$

$$= \sum_{x \to y; x \to y'} (e^\varrho a - 1)h(\varrho a') - \sum_{x \to y; x \to y'} h(\varrho a)(e^{\varrho a'} - 1) + \sum_{x \to y; x \to y'} h(\varrho a')$$

$$- \sum_{x \to y; x \to y'} (e^\varrho a - 1)(e^{\varrho a'} - 1)$$

$$= J_x(\boldsymbol{X}) \sum_{x \to y} h(\varrho a) - \left(\sum_{x \to y} (e^\varrho a - 1) \right)^2$$

and the desired result follows. □

Lemma 6.4.3 leads us to the following evaluations.

Proposition 6.4.4. *Consider f_0, g_1 as in Definition 6.1.1 and suppose that $f_0, g_1 \in L^1(m) \cap L^2(m)$ where the stationary measure m charges every point of $\boldsymbol{X}$. Also assume that $\overrightarrow{J}, \overleftarrow{J}$ satisfy (6.37). Then,*

$$\overrightarrow{\Theta}_2 \psi_t(x) = \left(\sum_{y:x \sim y} (e^{D\psi_t(x,y)} - 1) \overrightarrow{J}_x(y) \right)^2 \tag{6.41}$$

$$+ \sum_{y:x \sim y} [\overrightarrow{J}_y(\boldsymbol{X}) - \overrightarrow{J}_x(\boldsymbol{X})]\theta^\star \left(e^{D\psi_t(x,y)} - 1 \right) \overrightarrow{J}_x(y)$$

$$+ \sum_{(y,z):x \sim y \sim z} \left[2\theta^\star \left(e^{D\psi_t(y,z)} - 1 \right) \overrightarrow{A}_{t,x}(y) \overrightarrow{J}_y(z) \right.$$

$$\left. - \theta^\star \left(e^{D\psi_t(x,z)} - 1 \right) \overrightarrow{J}_x(y) \overrightarrow{J}_y(z) \right]$$

and

$$\overleftarrow{\Theta}_2 \phi_t(x) = \left(\sum_{y:x \sim y} (e^{D\phi_t(x,y)} - 1) \overleftarrow{J}_x(y) \right)^2 \tag{6.42}$$

$$+ \sum_{y:x \sim y} [\overleftarrow{J}_y(\boldsymbol{X}) - \overleftarrow{J}_x(\boldsymbol{X})]\theta^\star \left(e^{D\phi_t(x,y)} - 1 \right) \overleftarrow{J}_x(y)$$

$$+ \sum_{(y,z):x \sim y \sim z} \left[2\theta^\star \left(e^{D\phi_t(y,z)} - 1 \right) \overleftarrow{A}_{t,x}(y) \overleftarrow{J}_y(z) \right.$$

$$\left. - \theta^\star \left(e^{D\phi_t(x,z)} - 1 \right) \overleftarrow{J}_x(y) \overleftarrow{J}_y(z) \right]$$

where

$$\overrightarrow{A}_{t,x}(y) = e^{D\psi_t(x,y)} \overrightarrow{J}_x(y) = \frac{g_t(y)}{g_t(x)} \overrightarrow{J}_x(y),$$

$$\overleftarrow{A}_{t,x}(y) = e^{D\phi_t(x,y)} \overleftarrow{J}_x(y) = \frac{f_t(y)}{f_t(x)} \overleftarrow{J}_x(y)$$

are the forward and backward jump frequencies of $P = f_0(X_0)g_1(X_1) R$.

Reversible random walks

Let us take a positive measure $m = \sum_{x \in \boldsymbol{X}} m_x \delta_x \in M_+(\boldsymbol{X})$ with $m_x > 0$ for all $x \in \boldsymbol{X}$. It is easily checked with the detailed balance condition

$$m(dx)J_x(dy) = m(dy)J_y(dx) \tag{6.43}$$

which characterizes the reversibility of m that the jump kernel

$$J_x = \sum_{y:x \sim y} s(x, y) \sqrt{m_y / m_x}\, \delta_y, \tag{6.44}$$

where $s(x, y) = s(y, x) > 0$ for all $x \sim y$, admits m as a reversing measure. As $(\boldsymbol{X}, \sim)$ is assumed to be connected, the random walk is irreducible and the reversing measure is unique up to a multiplicative constant.

Examples 6.4.5. *Let us present the simplest examples of reversible random walks.*
(a) Reversible counting random walk. *The simplest example is provided by the counting jump kernel $J_x = \sum_{y:x \sim y} \delta_y \in \mathrm{P}(\boldsymbol{X})$, $x \in \boldsymbol{X}$ which admits the counting measure $m = \sum_x \delta_x$ as a reversing measure: take $m_x = 1$ and $s(x, y) = 1$ in (6.44).*
(b) Reversible simple random walk. *The simple jump kernel is $J_x = \frac{1}{n_x} \sum_{y:x \sim y} \delta_y \in \mathrm{P}(\boldsymbol{X})$, $x \in \boldsymbol{X}$. The measure $m = \sum_x n_x \delta_x$ is a reversing measure: take $m_x = n_x$ and $s(x, y) = (n_x n_y)^{-1/2}$ in (6.44).*

The dynamics of R is as follows. Once the walker is at x, one starts an exponential random clock with frequency $J_x(\boldsymbol{X})$. When the clock rings, the walker jumps at random onto a neighbor of x according to the probability law $J_x(\boldsymbol{X})^{-1} \sum_{y:x \sim y} J(x; y)\, \delta_y$. This procedure goes on and on. Since R is reversible, we have $\overrightarrow{J} = \overleftarrow{J} = J$ and the forward and backward frequencies of jumps of the (f, g)-transform $P = f_0(X_0)g_1(X_1)\, R$ are respectively

$$\overrightarrow{A}_x(y) = \frac{g_t(y)}{g_t(x)} J_x(y) \quad \text{and} \quad \overleftarrow{A}_x(y) = \frac{f_t(y)}{f_t(x)} J_x(y).$$

As a direct consequence of Theorem 6.4.2 and Proposition 6.4.4, we obtain the following result.

Theorem 6.4.6. *Let R be an m-reversible random walk with jump measure J which is given by (6.44) and satisfies*

$$\sup_{x \in \boldsymbol{X}} J_x(\boldsymbol{X}) < \infty.$$

Then, along any entropic interpolation $[\mu_0, \mu_1]^R$ associated with a couple (f_0, g_1) as in Definition 6.1.1 and such that $f_0, g_1 \in L^1(m) \cap L^2(m)$, we have for all $0 < t < 1$,

$$
\begin{aligned}
\frac{d}{dt} H(\mu_t | m) &= \sum_{x \in \boldsymbol{X}} [\Theta \psi_t - \Theta \phi_t](x)\, \mu_t(x), \\
\frac{d^2}{dt^2} H(\mu_t | m) &= \sum_{x \in \boldsymbol{X}} [\Theta_2 \psi_t + \Theta_2 \phi_t](x)\, \mu_t(x)
\end{aligned}
$$

where the expressions of $\Theta \psi_t$ and $\Theta \phi_t$ are given in (6.38) and the expressions of $\Theta_2 \psi_t$ and $\Theta_2 \phi_t$ are given in Proposition 6.4.4 (drop the useless time arrows).

Remarks 6.4.7.
(a) For the above condition $\sup_{x \in X} J_x(X) < \infty$ to be satisfied, it suffices that for some $0 < c, \sigma < \infty$, we have for all $x \sim y \in X$,

$$\begin{cases} m_y/n_y \le c m_x/n_x, \\ 0 < s(x,y)\sqrt{n_x n_y} \le \sigma, \end{cases} \tag{6.45}$$

with the notation of (6.44).
(b) In the special case where $f_0 = \rho_0$ and $g_1 = 1$ corresponding to the forward heat flow – see Definition 6.2.3 – we obtain

$$\frac{d}{dt}H(\mu_t|m) = -\sum_{x \sim y} D\rho_t(x,y)D(\log\rho_t)(x,y)\,m(x)J_x(y),$$

$$\frac{d^2}{dt^2}H(\mu_t|m) = \sum_x \left\{ L\rho_t(x)L(\log\rho_t)(x) + \frac{(L\rho_t(x))^2}{\rho_t(x)} \right\} m(x).$$

The first identity is the well-known entropy production formula and the second one was proved in [8].

6.5 Convergence to equilibrium

This section is dedicated to the detailed proofs of already known results such as the (modified) logarithmic Sobolev inequality in Theorem 6.5.8. Its motivation is two-pronged: (i) the proof of the convergence result Theorem 6.5.2 is usually omitted as part of the standard folklore, and (ii) it is interesting to look at heat flows as special entropic interpolations.

In this section R is an m-stationary Markov probability measure on the path space

$$\Omega = D([0,\infty), X)$$

built on the unbounded time interval ($[0,\infty)$ and m is assumed to be a probability measure (and so is R). We are interested in the convergence as t tends to infinity of the forward heat flow

$$\mu_t := (X_t)_\#(\rho_0(X_0)\,R) \in P(X), \quad t \ge 0, \tag{6.46}$$

where $\rho_0 = d\mu_0/dm$ is the initial density. We are going to prove by implementing the "stochastic process strategy" of the present paper, that under some hypotheses, the following convergence to equilibrium

$$\lim_{t \to \infty} \mu_t = m$$

holds. As usual, the function $t \mapsto H(\mu_t|m)$ is an efficient Lyapunov function of the dynamics.

Stationary dynamics

Let us first make the basic assumptions precise.

Assumptions 6.5.1. *The stationary measure m is a probability measure and we assume the following.*
(a) Brownian diffusion. The forward derivative is

$$\overrightarrow{L} = \overrightarrow{b} \cdot \nabla + \Delta/2$$

on a connected Riemannian manifold X without boundary. The initial density $\rho_0 := d\mu_0/dm$ is regular enough for $\rho_t(z) = d\mu_t/dm(z)$ to be positive and twice differentiable in t and z on $(0, \infty) \times X$.
(b) Random walk. The forward derivative is

$$\overrightarrow{L} u(x) = \sum_{y:x\sim y} [u(y) - u(x)] \, \overrightarrow{J}_x(y)$$

on the countable locally finite connected graph $(X, \sim)$. We assume that (6.37) holds: $\sup_x \{ \overrightarrow{J}_x(X) + \overleftarrow{J}_x(X) \} < \infty$.

The point in the following Theorem 6.5.2 is that no reversibility is required. We define

$$\mathfrak{I}(\mu|m) := \int_X \overleftarrow{\Theta}(\log\rho) \, d\mu \in [0, \infty], \quad \mu = \rho\, m \in \mathrm{P}(X).$$

We have seen in (6.28) and (6.38) that $\overleftarrow{\Theta} \geq 0$ (notice that $\theta^\star \geq 0$), this shows that $\mathfrak{I}(\mu|m) \geq 0$.

Theorem 6.5.2. *The Assumptions 6.5.1 are supposed to hold together with $\Theta_2 \geq 0$.*
Then, $\lim_{t\to\infty} \mathfrak{I}(\mu_t|m) = 0$.
Let us assume in addition that

$$\overleftarrow{\Theta}_2 \geq \kappa\overleftarrow{\Theta} \quad \text{for some constant } \kappa > 0 \tag{6.47}$$

and that the initial density $\rho_0 := d\mu_0/dm$ is bounded: $\sup_X \rho_0 < \infty$.
Then, $\lim_{t\to\infty} H(\mu_t|m) = 0$.

Remark 6.5.3. *The total variation norm of a signed measure η on X is defined by $\|\eta\|_{\mathrm{TV}} := |\eta|(X)$ and the Csiszár-Kullback-Pinsker inequality is $\frac{1}{2}\|\mu - m\|_{TV}^2 \leq H(\mu|m)$, for any $\mu \in \mathrm{P}(X)$. Therefore we have $\lim_{t\to\infty} \|\mu_t - m\|_{\mathrm{TV}} = 0$.*

Proof. The main idea of this proof is to consider the forward heat flow $(\mu_t)_{t\geq 0}$ as an entropic interpolation, see Definition 6.2.3, and to apply Claim 2. It will be seen that the general regularity hypotheses of the theorem ensure that this claim becomes a rigorous statement.

Instead of restricting time to $[0, 1]$, we allow it to be in $[0, T]$, $T > 0$, and let T tend to infinity. We denote $\rho_t := d\mu_t/dm$. Since μ is the time-marginal flow of $P = \rho_0(X_0) R$, we have $f_0 = \rho_0$ and $g_T = 1$. This implies that $g_t = 1$ and $f_t = \rho_t$. Indeed, $g_t(z) = E_R[g_T(X_T) \mid X_t = z] = E_R(1 \mid X_t = z) = 1$ and $f_t(z) = E_R[f_0(X_0) \mid X_t = z] = E_R[\rho_0(X_0) \mid X_t = z] = E_R[dP/dR \mid X_t = z] = dP_t/dR_t(z) = dP_t/dm(z) =: \rho_t(z)$. Therefore, $\psi_t = 0$, $\phi_t = \log \rho_t$, for all $0 \le t \le T$. Our assumptions guarantee that for all $t > 0$, ϕ_t is regular enough to use our previous results about "Θ_2-calculus". Denoting $H(t) = H(\mu_t|m)$, with Corollary 6.2.4, we see that

$$-H'(t) = \overleftarrow{I}(t) := \int_X \overleftarrow{\Theta}(\log \rho_t)\, d\mu_t =: \mathfrak{I}(\mu_t|m).$$

As $\mathfrak{I}(\cdot|m) \ge 0$, H is decreasing. Of course, $H(0) - H(T) = \int_0^T \mathfrak{I}(\mu_t|m)\, dt$. Now, we let T tend to infinity. As H is non-negative and decreasing, it admits a limit $H(\infty) := \lim_{T\to\infty} H(T)$ and the integral in

$$H(0) - H(\infty) = \int_0^\infty \mathfrak{I}(\mu_t|m)\, dt \tag{6.48}$$

is convergent, implying with $\frac{d}{dt}\mathfrak{I}(\mu_t|m) = -\Theta_2(\phi_t) \le 0$ that $\lim_{t\to\infty} \mathfrak{I}(\mu_t|m) = 0$. This proves the first assertion of the theorem.

As ρ_0 is assumed to be bounded, for all t we have

$$\rho_t(X_t) = E_R(\rho_0 \mid X_t) \le \sup \rho_0 < \infty. \tag{6.49}$$

This will be used below.

Diffusion setting. By (6.29) with $\phi_t = \rho_t$, this limit implies that $\lim_{t\to\infty} \int_X |\nabla \sqrt{\rho_t}|^2\, d\mathrm{vol} = 0$. Poincaré's inequality tells us that for any open bounded connected domain U with a smooth boundary, there exists a constant C_U such that $\|\sqrt{\rho_t} - \langle\sqrt{\rho_t}\rangle_U\|_{L^2(U)} \le C_U \|\nabla\sqrt{\rho_t}\|_{L^2(U)}$ for all t, where $\langle\sqrt{\rho_t}\rangle_U := \int_U \sqrt{\rho_t}\, d\mathrm{vol} / \mathrm{vol}(U)$. It follows that $\lim_{t\to\infty} \|\sqrt{\rho_t} - \langle\sqrt{\rho_t}\rangle_U\|_{L^2(U)} = 0$ which in turns implies that

$$\lim_{t\to\infty} \rho_t = c \tag{6.50}$$

everywhere on X for some constant $c \ge 0$.

Let us prove this last claim. By Bienaymé-Chebychev's inequality, a vanishing variance along a sequence of random variables implies convergence in measure towards a constant. On the other hand, the exponential rate of convergence ensured by our assumption (6.47) implies with the Borel-Cantelli lemma that this limit is almost sure on U. Finally, as ρ is continuous on $(0, \infty) \times X$ and X is connected, we can extend this convergence to the whole space.

As $\sup_t H(\mu_t|m) \le H(\mu_0|m) < \infty$, the set $\{\mu_t, t \ge 0\}$ is uniformly tight and therefore relatively compact with respect to the narrow topology on $\mathrm{P}(X)$. It follows from (6.50)

that $\lim_{t \to \infty} \rho_t = 1$ everywhere. Finally, the uniform bound (6.49) allows us to apply the Lebesgue dominated convergence theorem to assert $\lim_{t \to \infty} H(\mu_t|m) = 0$, which is the desired result.

Random walk setting. Since $\lim_{t \to \infty} \mathfrak{I}(\mu_t|m) = 0$, we obtain

$$\lim_{t \to \infty} \sum_{(x,y):x \sim y} \rho_t(x)\theta^\star\left(\rho_t(y)/\rho_t(x) - 1\right) m(x) \overleftarrow{J}_x(y) = 0. \tag{6.51}$$

For some constant $C > 0$, we have

$$a\theta^\star(b/a - 1) \geq C\frac{(b-a)^2}{a}, \quad \text{if } b/a \leq 2.$$

We see that the convergence to zero of the left-hand side implies the convergence of $b - a$ to zero when on a domain such that a is bounded. This will be applied with $a = \rho_t(x)$ which is bounded in virtue of (6.49).

When $b/a \geq 2$ and $a \geq 2$, the left hand side cannot be close to zero. Finally, the remaining case when $b/a \geq 2$ and $a \leq 2$ is controlled by considering the symmetric term $b\theta^\star(a/b - 1)$ which also appears in the series in the limit (6.51), reverting x and y; recall that $\overleftarrow{J}_x(y) > 0 \Leftrightarrow \overleftarrow{J}_y(x) > 0$. As $b\theta^\star(a/b - 1) \geq b \geq 2a \geq 0$, the convergence of the left-hand term to zero implies the convergence of a and b to 0. Therefore, the limit (6.51) implies that for all adjacent x and $y \in X$, $\lim_{t \to \infty}(\rho_t(x) - \rho_t(y)) = 0$.

It follows from (6.49) the countability of X and the connectedness of the graph, that there exists a sequence of times (t_k) tending to infinity and a non-negative number c such that $\lim_k \rho_{t_k}(x) = c$ for all $x \in X$. The same argument as in the diffusion setting leads us to the relative compactness of $\{\mu_t; t \geq 0\}$ and in particular of (μ_{t_k}). It follows that $\lim_k \rho_{t_k}(x) = 1$ for all x. Finally, as $t \mapsto H(\mu_t|m) \geq 0$ is decreasing, it admits a limit as t tends to infinity and we obtain: $\lim_t H(\mu_t|m) = \lim_k H(\mu_{t_k}|m) = 0$ where the last identity follows as in the diffusion setting from the dominated convergence theorem. This completes the proof of the theorem. □

Theorem 6.5.4. *The Assumptions 6.5.1 are supposed to hold. Then, under the additional hypotheses that $H(\mu_0|m) < \infty$ and (6.47) i.e. $\overleftarrow{\Theta}_2 \geq \kappa \overleftarrow{\Theta}$ for some $\kappa > 0$, we have*

$$\mathfrak{I}(\mu_t|m) \quad \leq \quad \mathfrak{I}(\mu_0|m)e^{-\kappa t}, \quad t \geq 0 \tag{6.52}$$

$$H(\mu_t|m) \quad \leq \quad H(\mu_0|m)e^{-\kappa t}, \quad t \geq 0 \tag{6.53}$$

and the following functional inequality

$$H(\rho m|m) \leq \kappa^{-1}\mathfrak{I}(\rho m|m), \tag{6.54}$$

holds for any nonnegative function $\rho : X \to [0, \infty)$ such that $\int_X \rho\, dm = 1$, which in the diffusion setting (a) is also assumed to be $\mathcal{C}^2$-regular.

Proof of Theorem 6.5.4. We already saw during the proof of Theorem 6.5.2 that our general assumptions allow us to apply Θ_2-calculus: Claim 2 is rigorous with $\phi_t = \log\rho_t$.

This gives: $\overleftarrow{I}'(t) = -H'(t) = \int_X \overleftarrow{\Theta}(\log \rho_t)\, d\mu_t$ and $\overleftarrow{I}''(t) = -H''(t) = -\int_X \overleftarrow{\Theta}_2(\log \rho_t)\, d\mu_t$. The inequality (6.47) implies that $\overleftarrow{I}'(t) \le -\kappa \overleftarrow{I}(t)$, for all $t \ge 0$. Integrating leads to $\overleftarrow{I}(t) \le \overleftarrow{I}(0)e^{-\kappa t}$, $t \ge 0$, which is (6.52).

Let us assume for a while, in addition to (6.47), that $\sup_X \rho_0 < \infty$. It is proved in Theorem 6.5.2 that under this restriction, $H(\infty) = 0$. With (6.48) and (6.52), we see that

$$H(\mu_0|m) = H(0) = \int_0^\infty \overleftarrow{I}(t)\, dt \le \overleftarrow{I}(0) \int_0^\infty e^{-\kappa t}\, dt = \overleftarrow{I}(0)/\kappa.$$

A standard approximation argument based on the sequence $\rho_n = (\rho \wedge n)/\int_X (\rho \wedge n)\, dm$ and Fatou's lemma: $H(\rho m|m) \le \liminf_n H(\rho_n m|m)$, allows to extend this inequality to unbounded ρ. This proves (6.54).

Plugging μ_t instead of μ_0 into (6.54), one sees that $H(t) \le \overleftarrow{I}(t)/\kappa = -H'(t)/\kappa$. Integrating leads to $H(t) \le H(0)e^{-\kappa t}$, which is (6.53). $\qquad\square$

Remark 6.5.5 (About time reversal). *The backward arrows in (6.47) suggest that time reversal is tightly related to this convergence to stationarity. Let us propose an informal interpretation of this phenomenon. The forward entropy production $\overrightarrow{I}(t)$ vanishes along the forward heat flow $(\mu_t)_{t\ge0}$ since $\overrightarrow{I}(t) = \int_X \overrightarrow{\Theta}\psi_t\, d\mu_t$ and $\psi = 0$. Similarly $\int_X \overrightarrow{\Theta}_2 \psi_t\, d\mu_t = 0$. No work is needed to drift along the heat flow. In order to evaluate the rate of convergence, one must measure the strength of the drift toward equilibrium. To do so "one has to face the wind" and measure the work needed to reach μ_0 when starting from m, reversing time.*

Remarks 6.5.6 (About $\overleftarrow{\Theta}_2 \ge \kappa\overleftarrow{\Theta}$).
(a) In the diffusion setting, (6.47) can be written

$$\overleftarrow{\Gamma}_2 \ge \kappa\Gamma.$$

When $\overrightarrow{L}$ is the reversible forward derivative (6.9), this is the standard Bakry-Émery curvature condition $CD(\kappa, \infty)$, see [4, 5]. Further detail is given below at Theorem 6.5.8.
(b) In the random walk setting, we see with Lemma 6.4.3 that (6.47) can be written

$$\left(\sum_{y:x\sim y} (e^{\varrho a_y} - 1)\overleftarrow{J}_x(y) \right)^2 + \sum_{y:x\sim y} [\overleftarrow{J}_y(X) - \overleftarrow{J}_x(X)]h(\varrho a_y)\overleftarrow{J}_x(y)$$

$$+ \sum_{(y,z):x\sim y\sim z} [2e^{\varrho a_y}h(c_z - \varrho a_y) - h(c_z)]\overleftarrow{J}_x(y)\overleftarrow{J}_y(z) \qquad (6.55)$$

$$\ge \kappa \sum_{y:x\sim y} h(\varrho a_y)\overleftarrow{J}_x(y)$$

for all $x \in X$ and any numbers $\varrho a_y, c_z \in \mathbb{R}$, $(y,z) : x \sim y \sim z$, where $h(a) := \theta^\star(e^a - 1) = ae^a - e^a + 1$, $a \in \mathbb{R}$, see (6.40).

(c) *An inspection of the proof of Theorem 6.5.4 shows that the integrated version of (6.47):*

$$\int_X \overleftarrow{\Theta}_2(\rho)\, d\mu \geq \kappa \int_X \overleftarrow{\Theta}(\rho)\, d\mu, \quad \mu = \rho m \in \mathrm{P}(\boldsymbol{X})$$

is sufficient.

In the diffusion setting (6.27) on $\boldsymbol{X} = \mathbb{R}^n$, we know through (6.31) that (6.47) becomes

$$\|\nabla^2 u\|_{\mathrm{HS}}^2 + \left(\nabla^2 V + [\nabla + \nabla^\star]b_\perp\right)(\nabla u) \geq \kappa|\nabla u|^2$$

for any sufficiently regular function u. This Γ_2-criterion was obtained by Arnold, Carlen and Ju [3]; see also the paper [24] by Hwang, Hwang-Ma and Sheu for a related result. These results are recovered in the recent paper [22] by Fontbona and Jourdain who implement a stochastic process approach, slightly different from the present article's one but where time reversal also plays a crucial role, see Remark 6.5.5.

Reversible dynamics

More precisely, we are concerned with the following already encountered m-reversible generators L.

(a) On a Riemannian manifold $\boldsymbol{X}$, see (6.9):

$$L = (-\nabla V \cdot \nabla + \Delta)/2 \tag{6.56}$$

with the reversing measure $m = e^{-V}$ vol.

(b) On a graph $(\boldsymbol{X}, \sim)$, see (6.44):

$$Lu(x) = \sum_{y:x \sim y} [u(y) - u(x)]s(x,y)\sqrt{m_y/m_x} \tag{6.57}$$

with $s(x,y) = s(y,x) > 0$ for all $x \sim y$.

Let us recall our hypotheses.

Assumptions 6.5.7. *It is required that the reversing measure m is a probability measure.*
The assumptions (a) and (b) below allow us to apply respectively Theorems 6.3.5 and 6.4.6.
(a) *The Riemannian manifold $\boldsymbol{X}$ is compact connected without boundary and L is given by (6.56). We assume that $V : \boldsymbol{X} \to \mathbb{R}$ is $\mathbb{C}^4$-regular and without loss of generality that V is normalized by an additive constant such that $m = e^{-V}$ vol is a probability measure.*

(b) The countable graph $(X, \sim)$ is assumed to be locally finite and connected. The generator L is given by (6.57), and the equilibrium probability measure m and the function s satisfy (6.45).

Reversibility allows for a simplified expression of $\mathfrak{I}(\cdot|m)$. Indeed, as $\rho\Theta(\log\rho) = L(\rho\log\rho - \rho) - \log\rho\, L\rho$, we obtain

$$\int_X \Theta(\log\rho)\rho\, dm = \int_X L(\rho\log\rho - \rho)\, dm - \int_X \log\rho\, L\rho\, dm = \frac{1}{2}\int_X \Gamma(\rho, \log\rho)\, dm$$

where the last equality follows from the symmetry of L in $L^2(m)$ (a consequence of the m-reversibility of R), whenever $(\rho, \log\rho) \in \mathrm{Dom}\,\Gamma$. Therefore,

$$\mathfrak{I}(\mu|m) = I(\mu|m) := \frac{1}{2}\int_X \Gamma(\rho, \log\rho)\, dm. \tag{6.58}$$

It is the Fisher information of μ with respect to m. A direct computation shows that

(a) in the diffusion setting (a),

$$I(\mu|m) = \frac{1}{2}\int_X |\nabla\log\rho|^2\, d\mu = \frac{1}{2}\int_X \frac{|\nabla\rho|^2}{\rho}\, dm = 2\int_X |\nabla\sqrt{\rho}|^2\, dm;$$

(b) in the random walk setting (b),

$$I(\mu|m) = \frac{1}{2}\sum_{(x,y):x\sim y} [\rho(y) - \rho(x)][\log\rho(y) - \log\rho(x)]\, m(x)J_x(y).$$

Theorem 6.5.4 becomes

Theorem 6.5.8. *The Assumptions 6.5.7 are supposed to hold. Then, under the additional hypotheses that $H(\mu_0|m) < \infty$ and (6.47): $\overleftarrow{\Theta}_2 \geq \kappa\overleftarrow{\Theta}$ for some $\kappa > 0$, we have*

$$I(\mu_t|m) \leq I(\mu_0|m)e^{-\kappa t}, \quad t \geq 0$$
$$H(\mu_t|m) \leq H(\mu_0|m)e^{-\kappa t}, \quad t \geq 0$$

and the following (modified) logarithmic Sobolev inequality

$$H(\mu|m) \leq \kappa^{-1}I(\mu|m), \tag{6.59}$$

for any $\mu \in P(X)$ which in the diffusion setting (a) is also restricted to be such that $d\mu/dm$ is $\mathcal{C}^2$-regular.

In the diffusion setting (a), Theorem 6.5.8 is covered by the well-known result by Bakry and Émery [5]. Inequality (6.59) is the standard logarithmic Sobolev inequality and

(6.47): $\overleftarrow{\Theta}_2 \geq \kappa \overleftarrow{\Theta}$ for some $\kappa > 0$, is the usual Γ_2-criterion: $\Gamma_2 \geq \kappa\Gamma$. In the random walk setting (b), inequality (6.59) is a modified logarithmic Sobolev inequality which was introduced by Bobkov and Tetali [6] in a general setting, and was already employed by Dai Pra, Paganoni and Posta [13] in the context of Gibbs measures on $\mathbb{Z}^d$. Later Caputo, Dai Pra and Posta [8, 9, 14] have derived explicit criteria for (6.59) to be satisfied in specific settings related to Gibbs measures of particle systems. These criteria are close in spirit to (6.47).

6.6 Some open questions about curvature and entropic interpolations

We have seen in Section 6.5 that convergence to equilibrium is obtained by considering heat flows. The latter are the simplest entropic interpolations since $f = 1$ or $g = 1$. In the present section, general entropic interpolations are needed to explore curvature properties of Markov generators and their underlying state space X by means of the main result of the article:

$$H'(t) = \int_X \frac{1}{2}(\overrightarrow{\Theta}\psi_t - \overleftarrow{\Theta}\phi_t)\,d\mu_t, \quad H''(t) = \int_X \frac{1}{2}(\overrightarrow{\Theta}_2\psi_t + \overleftarrow{\Theta}_2\phi_t)\,d\mu_t.$$

Convexity properties of the entropy along displacement interpolations are fundamental for the Lott-Sturm-Villani (LSV) theory of lower bounded curvature of metric measure spaces [31, 41, 42, 44]. An alternative approach would be to replace displacement interpolations with entropic interpolations, taking advantage of the analogy between these two types of interpolations. Although this program is interesting in itself, it can be further motivated by the following remarks.

1. Entropic interpolations are more regular than displacement interpolations.
2. Displacement interpolations are (semiclassical) limits of entropic interpolations, see [26, 29].
3. Entropic interpolations work equally well in continuous and discrete settings while LSV theory fails in the discrete setting (see below).

With respect to (1), note that the dynamics of an entropic interpolation share the regularity of the solutions f and g of the backward and forward "heat equations" (6.11). Their logarithms, the Schrödinger potentials ϕ and ψ, solve *second order* Hamilton-Jacobi-Bellman equations. This is in contrast with a displacement interpolation whose dynamics are driven by the Kantorovich potentials (analogues of ϕ and ψ) which are solutions of *first order* Hamilton-Jacobi equations. Entropic interpolations are linked to regularizing dissipative PDEs, while displacement interpolations are linked to transport PDEs. More detail about these relations is given in [28].

In practice, once the regularity of the solutions of the dissipative linear PDEs (6.11) is ensured, the rules of calculus that are displayed at the beginning of Section 6.2 are efficient. Again, this is in contrast with Otto's calculus which only yields heuristics about displacement interpolations.

With respect to (3), recall that we have seen in Theorems 6.5.4 and 6.5.8 unified proofs of entropy-entropy production inequalities, available in the diffusion and random walk settings.

In order to provide a better understanding of the analogy between displacement and entropic interpolations, we start describing two thought experiments. Then, keeping the LSV strategy as a guideline, we raise a few open questions about the use of entropic interpolations in connection with curvature problems on Riemannian manifolds and graphs.

Thought experiments

Let us describe two thought experiments.

- *Schrödinger's hot gas experiment*. The dynamical Schrödinger problem (see (S_{dyn}) in Section 6.1) is a formalization of Schrödinger's thought experiment which was introduced in [39]. Ask a perfect gas of particles living in a Riemannian manifold X which are in contact with a *thermal bath* to start from a configuration profile $\mu_0 \in P(X)$ at time $t = 0$ and to end up at the unlikely profile $\mu_1 \in P(X)$ at time $t = 1$. For instance the gas may be constituted of undistinguishable mesoscopic particles (e.g. grains of pollen) with a random motion communicated by numerous shocks with the microscopic particles of the thermal bath (e.g. molecules of hot water). Large deviation considerations lead to the following conclusion. In the (thermodynamical) limit of infinitely many mesoscopic particles, the most likely trajectory of the whole mesoscopic particle system from μ_0 to μ_1 is the R-entropic interpolation $[\mu_0, \mu_1]^R$ where R describes the random motion of the non-interacting (the gas is assumed to be perfect) mesoscopic particles. Typically, R describes a Brownian motion as in Schrödinger's original papers [39, 40]. For more detail see [28, §. 6] for instance.
- *Cold gas experiment*. The same thought experiment is also described in [44, pp. 445-446] in a slightly different setting where it is called the *lazy gas experiment*. The only difference with Schrödinger's thought experiment is that no thermal bath enters the game. The perfect gas is cold so that the paths of the particles are Riemannian geodesics in order to minimize the average kinetic action of the whole system: entropy minimization is replaced by quadratic optimal transport.

Connection between entropic and displacement interpolations

It is shown in [26, 29] that displacement interpolations are limits of entropic interpolations when the frequency of the random elementary motions encrypted by R tends down to zero. With the above thought experiment in mind, this corresponds to a thermal bath whose density tends down to zero. Replacing the hot water by an increasingly rarefied gas, at the zero-temperature limit, the entropic interpolation $[\mu_0, \mu_1]^R$ is replaced by McCann's displacement interpolation $[\mu_0, \mu_1]^{\text{disp}}$.

Cold gas experiment

As the gas is (absolutely) cold, the particle sample paths are regular and deterministic. The trajectory of the mass distribution of the cold gas is the displacement interpolation $[\mu_0, \mu_1]^{\text{disp}}$ which is the time marginal flow of a probability measure P on Ω concentrated on geodesics:

$$P = \int_{\mathbf{X}^2} \delta_{\gamma^{xy}} \, \pi(dxdy) \in \mathrm{P}(\Omega) \tag{6.60}$$

where $\pi \in \mathrm{P}(\mathbf{X}^2)$ is an optimal transport plan between μ_0 and μ_1, see [28]. Therefore,

$$\mu_t = P_t = \int_{\mathbf{X}^2} \delta_{\gamma_t^{xy}} \, \pi(dxdy) \in \mathrm{P}(\mathbf{X}), \quad 0 \le t \le 1, \tag{6.61}$$

where for simplicity it is assumed that there is a unique minimizing geodesic γ^{xy} between x and y (otherwise, one is allowed to replace $\delta_{\gamma^{xy}}$ with any probability concentrated on the set of all minimizing geodesics from x to y).

In a positively curved manifold, several geodesics starting from the same point have a tendency to approach each other. Therefore, it is necessary that the gas initially spreads out to thwart its tendency to concentrate. Otherwise, it would not be possible to reach a largely spread target distribution. The left side of Figure 6.1 (taken from [44]) depicts this phenomenon. This is also the case for the right side as one can verify by reversing time. The central part is obtained by interpolating between the initial and final segments of the geodesics.

On the other hand, in a negatively curved manifold, several geodesics starting from the same point have a tendency to depart form each other. Therefore, it is necessary that the gas initially concentrates to thwart its tendency to spread out. Otherwise, it would not be possible to reach a condensed target distribution.
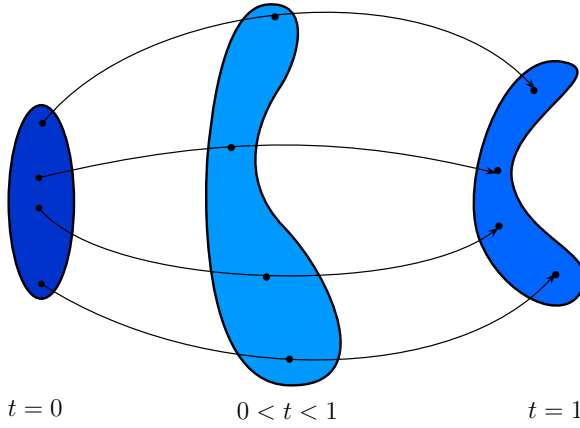
234 Christian Léonard

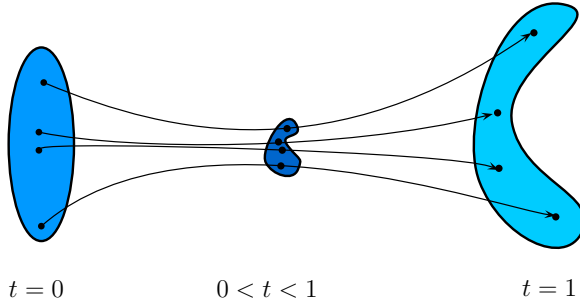**Fig. 6.1.** The cold gas experiment. Positive curvature

Fig. 6.2. The cold gas experiment. Negative curvature

Schrödinger's hot gas experiment

As the gas is hot, the particle sample paths are irregular and random. The trajectory of the mass distribution of the hot gas is the entropic interpolation $[\mu_0, \mu_1]^R$. An R-entropic interpolation is the time marginal flow of a probability measure P on Ω which is a mixture of bridges of R:

$$P = \int_{\boldsymbol{X}^2} R^{xy}\, \pi(dxdy) \in \mathrm{P}(\Omega) \tag{6.62}$$

where $\pi \in \mathrm{P}(\boldsymbol{X}^2)$ is the unique minimizer of $\eta \mapsto H(\eta|R_{01})$ among all $\eta \in \mathrm{P}(\boldsymbol{X}^2)$ with prescribed marginals μ_0 and μ_1, see [28]. Therefore,

$$\mu_t = P_t = \int_{\boldsymbol{X}^2} R_t^{xy}\, \pi(dxdy) \in \mathrm{P}(\boldsymbol{X}), \quad 0 \le t \le 1. \tag{6.63}$$

Comparing (6.60) and (6.62), one sees that the deterministic evolution $\delta_{\gamma^{xy}}$ is replaced by the random evolution R^{xy}. The grey leaf in Figure 6.3 is a symbol for the (say) 95%-support of the bridge R^{xy}. It spreads around the geodesic γ^{xy} and in general the support of R_t^{xy} for each intermediate time $0 < t < 1$ is the whole space $\boldsymbol{X}$.
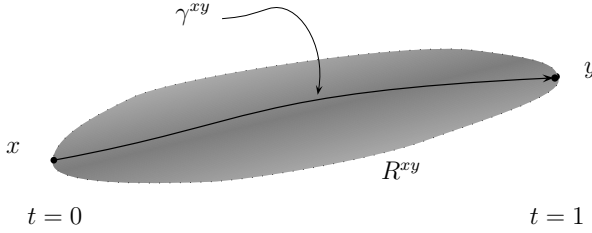

Fig. 6.3. Hot gas experiment with $\mu_0 = \delta_x$ and $\mu_1 = \delta_y$

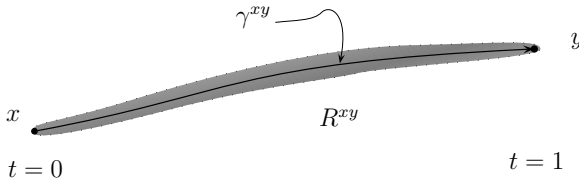At a lower temperature, R^{xy} is closer to $\delta_{\gamma^{xy}}$, see Figure 6.4.

Fig. 6.4. Hot gas experiment with $\mu_0 = \delta_x$ and $\mu_1 = \delta_y$. At a lower temperature

And eventually, at zero temperature R^{xy} must be replaced by its deterministic limit $\delta_{\gamma^{xy}}$. Figure 6.5 is a superposition of Figures 6.2 and 6.3. It describes the hot gas experiment in a negative curvature manifold.

Figures 6.3, 6.4 and 6.5 are only suggestive pictures describing most visited space areas.

Open questions

As far as one is interested in the curvature of the metric measure space $(\boldsymbol{X}, d, m)$, rather than in the curvature of some Markov generator $\overrightarrow{L}$ such as in Section 6.5, it seems natural to restrict our attention to a reversible reference path measure R. For instance when looking at a Riemannian manifold $\boldsymbol{X}$ considered as a metric measure space $(\boldsymbol{X}, d, m)$ with the Riemannian metric d and the weighted measure $m = e^{-V}\,\mathrm{vol}$, the efficient choice with respect to LSV theory is to take R as the reversible Markov

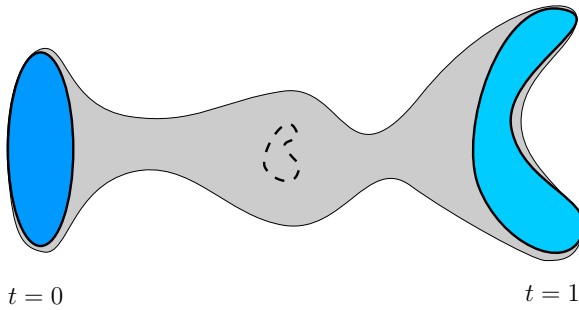$t = 0$ $t = 1$

Fig. 6.5. The not so lazy gas experiment in a negative curvature manifold

measure with generator $L = (\Delta - \nabla V \cdot \nabla)/2$ (or any positive multiple) and initial measure m.

A representative result in the case where the reference measure m is the volume measure ($V = 0$) is that relative entropy is convex along any displacement interpolation in a Riemannian manifold with a nonnegative Ricci curvature. Based on McCann's seminal work [33, 34] and Otto's heuristic calculus, Otto and Villani have conjectured this result in [38]. This was proved by Cordero-Erausquin, McCann and Schmuckenschläger in [12]. The converse is also true, as was shown later by Sturm and von Renesse [43].

In the same spirit, an immediate consequence of Theorem 6.3.5 is the following

Corollary 6.6.1. *Let R^o be the reversible Brownian motion on a connected Riemannian manifold X without boundary.*
Suppose that the Ricci curvature is non-negative. Then, along any entropic interpolation $[\mu_0, \mu_1]^{R^o}$ associated with R^o and $f_0, g_1 \in L^2(\text{vol})$ such that (6.8) holds, the entropy $t \in [0, 1] \mapsto H(\mu_t | \text{vol}) \in \mathbb{R}$ is a convex function.

Notice that in Theorem 6.3.5, X is assumed to be compact and f_0, g_1 must be $\mathcal{C}^2$. This was assumed to ensure the regularity of the interpolation (Proposition 6.3.4). However, in the present setting this regularity follows from the regularity of the heat kernel.

Open questions 6.6.2.
(a) Is the converse of Corollary 6.6.1 true?
(b) Is there a way to define a notion of κ-convexity on $\mathrm{P}(X)$ along entropic interpolations: $H(\mu_t) \leq (1 - t)H(\mu_0) + tH(\mu_1) - \kappa C(\mu_0, \mu_1)t(1 - t)/2$, $0 \leq t \leq 1$? What term $C(\mu_0, \mu_1)$ should replace the quadratic transport cost $W_2^2(\mu_0, \mu_1)$ of the LSV and AGS (Ambrosio, Gigli and Savaré [1, 2]) theories?
(c) Is there a notion of "entropic gradient flow" related to entropic interpolations that would be similar to a gradient flow and would allow interpreting heat flows as "en-

tropic gradient flows" of the entropy? This question might be related to the previous one: the entropic cost $C(\mu_0, \mu_1)$ could play the role of a squared distance on $P(X)$.

The LSV theory requires the metric space (X, d) to be geodesic. Consequently, discrete metric graphs are ruled out. Alternative approaches are necessary to develop a theory of lower bounded curvature on discrete metric graphs.

1. Bonciocat and Sturm [7] have obtained precise results for a large class of planar graphs by introducing the notion of h-approximate t-midpoint interpolations: $d(x_0, x_t) \le t d(x_0, x_1) + h$, $d(x_t, x_1) \le (1-t)d(x_0, x_1) + h$ where typically h is of order 1.

2. Erbar and Maas [18, 32] and independently Mielke [35, 36] have shown that the evolution of a reversible random walk on a graph is the gradient flow of an entropy for some distance on $P(X)$, and derived curvature bounds for discrete graphs by following closely the gradient flow strategy developed by Ambrosio, Gigli and Savaré [1] in the setting of the LSV theory.

3. Recently, Gozlan, Roberto, Samson and Tetali [23] obtained curvature bounds by studying the convexity of entropy along binomial interpolations. On a geodesic $(x := z_0, z_1, \ldots, z_{d(x,y)} =: y)$ for the standard graph distance d, the binomial interpolation is defined by $\mu_t(z_k) = \mathcal{B}(d(x, y), t)(k)$, $0 \le k \le d(x, y)$, $0 \le t \le 1$, where $\mathcal{B}(n, p)(k) = \binom{n}{k} p^k (1 - p)^{n-k}$ is the usual binomial weight.

We know by Lemma 6.4.3 that

$$\Theta_2 u(x) = \left(Bu(x) \right)^2 + \sum_{y:x \sim y} [J_y(X) - J_x(X)] h\left(Du(x, y)\right) J_x(y)$$
$$+ \sum_{(y,z):x \sim y \sim z} \left[2e^{Du(x,y)} h\left(Du(y, z)\right) - h\left(Du(x, z)\right) \right] J_x(y) J_y(z)$$

$$(6.64)$$

where $Bu(x) = \sum_{y:x \sim y} (e^{Du(x,y)} - 1) J_x(y)$ and $h(a) := \theta^\star(e^a - 1) = ae^a - e^a + 1$, $a \in \mathbb{R}$. In the diffusion setting where $L = (\Delta - \nabla V \cdot \nabla)/2$, we have seen in Section 6.3 that $\Theta_2 = \Gamma_2/2$ where Γ_2 is given by the Bochner formula

$$\Gamma_2(u) = \|\nabla^2 u\|_{\mathrm{HS}}^2 + \nabla^2 V(\nabla u) + \mathrm{Ric}(\nabla u). \qquad (6.65)$$

Open questions 6.6.3.
1. *In view of* (6.64) *and* (6.65), *is* (6.64) *a Bochner formula? Where is the curvature?*
2. *Are the following definitions relevant?*
   (a) *The m-reversible random walk generator L has curvature bounded below by $\kappa \in \mathbb{R}$ if*
   $$\sum_x \Theta_2(u)(x)m(x) \ge \kappa \sum_x \Theta(u)(x)m(x), \ \forall u.$$

(b) *The infimum of the curvature at $x \in \boldsymbol{X}$ of the Markov generator L is*

$$\mathrm{curv}_L(x) := \inf_u \frac{\Theta_2(u)}{\Theta(u)}(x).$$

A definition of curvature should be justified by its usefulness in terms of rate of convergence to equilibrium, concentration of measure or isoperimetric behavior.

Let us indicate that with the discrete Laplacian $Lu(x) = [u(x+1)-2u(x)+u(x-1)]/2$ on $\mathbb{Z}$, a rather tedious computation leads to the desired flatness result: $\mathrm{curv}_L(x) = 0$, for all $x \in \mathbb{Z}$.

A. Basic definitions

This appendix section is part of articles [27, 30]. We repeat it here for the reader's convenience.

Markov measures

We slightly extend the notion of Markov property to unbounded positive measures on Ω.

Definitions .0.4. *Any positive measure on Ω is called a* path measure. *Let Q be a path measure.*
(a) *It is said to be a* conditionable path measure *if for any $0 \leq t \leq 1$, Q_t is a σ-finite measure on $\boldsymbol{X}$.*
(b) *It is said to be a Markov measure if it is conditionable and for all $0 \leq t \leq 1$ and $B \in \sigma(X_{[t,1]})$, we have: $Q(B \mid X_{[0,t]}) = Q(B \mid X_t)$.*

When Q has an infinite mass, the notion of conditional expectation must be handled with care. We refer to the following definition.

Definition .0.5 (Conditional expectation)**.** *Let $\varphi : \Omega \to Y$ be a measurable map. Suppose that $\varphi_\# Q$ is a σ-finite measure on Y. Then, for any $f \in L^p(Q)$ with $p = 1, 2$ or ∞, the conditional expectation of f knowing φ is the unique (up to Q-a.e. equality) function $E_Q(f \mid \varphi) := \theta_f(\varphi) \in L^p(Q)$ such that for all measurable function h on Y in $L^{p_\star}(m)$ where $1/p + 1/p_\star = 1$, we have $\int_\Omega h(\varphi) f\, dQ = \int_\Omega h(\varphi)\theta_f(\varphi)\, dQ$.*

It is essential in this definition that the measure $\varphi_\# Q$ is σ-finite on Y; otherwise, we can't invoke the Radon-Nikodym theorem when proving the existence of θ_f. Inspecting the above definition of a Markov measure, one sees that it is necessary that $Q_{[0,t]}$ and Q_t are σ-finite for all $t \in [0, 1]$ for the corresponding conditional expectations to be

defined. Nonetheless, this is warranted by the requirement that Q is conditionable, as shown by the following result.

Lemma .0.6. *Let Q be a path measure and $\mathcal{T} \subset [0, 1]$. For $Q_\mathcal{T}$ to be σ-finite, it is enough that Q_{t_o} is σ-finite for some $t_o \in \mathcal{T}$.*

Proof. Let $t_o \in \mathcal{T}$ be such that Q_{t_o} is σ-finite with $(\boldsymbol{X}_n)_{n \geq 1}$ an increasing sequence of measurable subsets of $\boldsymbol{X}$ such that $Q_{t_o}(\boldsymbol{X}_n) < \infty$ and $\cup_{n \geq 1} \boldsymbol{X}_n = \boldsymbol{X}$. Then, $Q_\mathcal{T}$ is also σ-finite since $Q_\mathcal{T}(X_{t_o} \in \boldsymbol{X}_n) = Q_{t_o}(\boldsymbol{X}_n)$ for all $n \geq 1$ and $\cup_{n \geq 1}[X_\mathcal{T}(\Omega) \cap \{X_{t_o} \in \boldsymbol{X}_n\}] = X_\mathcal{T}(\Omega)$. $\qquad\square$

Consequently, for any Markov measure Q and any $\mathcal{T} \subset [0, 1]$, the conditional expectation $E_Q(\cdot \mid X_\mathcal{T})$ is well-defined; since Ω is a Polish space, there exists a regular conditional *probability* kernel $Q(\cdot \mid X_\mathcal{T}) : \Omega \to \mathrm{P}(\Omega)$ such that for all $f \in L^1(Q)$, $E_Q(f \mid X_\mathcal{T}) = \int_\Omega f \, dQ(\cdot \mid X_\mathcal{T})$, see [17, Thm. 10.2.2].

Stationary path measures

Let us make precise a couple of other known notions.

Definitions .0.7. *Let Q be a path measure.*
(a) It is said to be stationary *if for all $0 \leq t \leq 1$, $Q_t = m$, for some $m \in \mathrm{M}_{(}\boldsymbol{X})$. One says that Q is m-stationary.*
(b) It is said to be reversible *if for any subinterval $[a, b] \subset [0, 1]$, we have $(\mathrm{rev}^{[a,b]})_\# Q_{[a,b]} = Q_{[a,b]}$ where $\mathrm{rev}^{[a,b]}$ is the time reversal on $[a, b]$ which is defined by $\mathrm{rev}^{[a,b]}[\eta](t) = \eta([a + b - t]^+)$ for any $\eta \in D([a, b], \boldsymbol{X})$, $t \in [a, b]$.*

A reversible path measure is stationary. When Q is reversible, one sometimes says that $m = Q_0 = Q_1$ is a reversing measure for the forward kernel $(Q(\cdot \mid X_0 = x); x \in \boldsymbol{X})$, or the backward kernel $(Q(\cdot \mid X_1 = y); y \in \boldsymbol{X})$ or shortly for Q. One also says that Q is m-reversible to emphasize the role of the reversing measure.

Relative entropy

This subsection is a short part of [27, §. 2] which we refer to for more detail. Let r be some σ-finite positive measure on some space Y. The relative entropy of the probability measure p with respect to r is loosely defined by

$$H(p|r) := \int_Y \log(dp/dr) \, dp \in (-\infty, \infty], \qquad p \in \mathrm{P}(Y) \tag{6.66}$$

if $p \ll r$, and $H(p|r) = \infty$ otherwise. More precisely, when r is a probability measure, we have

$$H(p|r) = \int_Y h(dp/dr)\, dr \in [0, \infty], \qquad p, r \in \mathrm{P}(Y)$$

with $h(a) = a \log a - a + 1 \geq 0$ for all $a \geq 0$, (take $h(0) = 1$). Hence, the definition (6.66) is meaningful. If r is unbounded, one must restrict the definition of $H(\cdot|r)$ to some subset of $\mathrm{P}(Y)$ as follows. As r is assumed to be σ-finite, there exist measurable functions $W : Y \to [1, \infty)$ such that

$$z_W := \int_Y e^{-W}\, dr < \infty. \tag{6.67}$$

Define the probability measure $r_W := z_W^{-1} e^{-W} r$ so that $\log(dp/dr) = \log(dp/dr_W) - W - \log z_W$. It follows that for any $p \in \mathrm{P}(Y)$ satisfying $\int_Y W\, dp < \infty$, the formula

$$H(p|r) := H(p|r_W) - \int_Y W\, dp - \log z_W \in (-\infty, \infty]$$

is a meaningful definition of the relative entropy which is coherent in the following sense. If $\int_Y W'\, dp < \infty$ for another measurable function $W' : Y \to [0, \infty)$ such that $z_{W'} < \infty$, then $H(p|r_W) - \int_Y W\, dp - \log z_W = H(p|r_{W'}) - \int_Y W'\, dp - \log z_{W'} \in (-\infty, \infty]$. Therefore, $H(p|r)$ is well-defined for any $p \in \mathrm{P}(Y)$ such that $\int_Y W\, dp < \infty$ for some measurable nonnegative function W verifying (6.67).

Bibliography

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

[2]   L. Ambrosio, N. Gigli, and G. Savaré: Metric measure spaces with Riemannian Ricci curvature. *Duke Math. J.*, 163:1405–1490, 2014.

[3]   A. Arnold, E. Carlen, and Q. Ju: Large-time behavior of non-symmetric Fokker-Planck type equations. *Commun. Stoch. Anal.*, 2(1):153–175, 2008.

[4]   D. Bakry: L'hypercontractivité et son utilisation en théorie des semigroupes. In *Ecole d'Eté de Probabilités de Saint-Flour (1992)*, number 1581 in Lecture Notes in Mathematics. Springer-Verlag, 1994.

[5]   D. Bakry and M. Émery: Diffusions hypercontractives. In *Séminaire de Probabilités*, number 1123 in Lecture Notes in Mathematics. Springer-Verlag, 1985.

[6]   S. Bobkov and P. Tetali: Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.*, 19(2):289–336, 2006.

[7]   A.-I. Bonciocat and K.T. Sturm: Mass transportation and rough curvature bounds for discrete spaces. *J. Funct. Anal.*, 256:2944–2966, 2009.

[8]   P. Caputo, P. Dai Pra, and G. Posta: Convex entropy decay via the Bochner-Bakry-Emery approach. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):734–753, 2009.

[9] P. Caputo and G. Posta: Entropy dissipation estimates in a zero-range dynamics. *Probab. Theory Related Fields*, 139:65–87, 2007.

[10] P. Cattiaux: A pathwise approach of some classical inequalities. *Potential Anal.*, 20(4):361–394, 2004.

[11] L. Chaumont and G. Uribe-Bravo: Markovian bridges: weak continuity and pathwise constructions. *Ann. Probab.*, 39:609–647, 2011.

[12] D. Cordero-Erausquin, R. McCann, and M. Schmuckenschläger: A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.*, 146(2):219–257, 2001.

[13] P. Dai Pra, A.M. Paganoni, and G. Posta: Entropy inequalities for unbounded spin systems. *Ann. Probab.*, 30(4):1959–1976, 2002.

[14] P. Dai Pra and G. Posta: Entropy decay for interacting systems via the Bochner–Bakry–Emery approach. *Electron. J. Probab.*, 18(52), 2013.

[15] C. Dellacherie and P.A. Meyer: *Probabilités et Potentiel. Ch. XII à XVI. Théorie du potentiel associée à une résolvante, théorie des processus de Markov.* Hermann. Paris, 1987.

[16] J.L. Doob: Conditional Brownian motion and the boundary limits of harmonic functions. *Bull. Soc. Math. France*, 85:431–458, 1957.

[17] R. Dudley: *Real Analysis and Probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[18] M. Erbar and J. Maas: Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206(3):997–1038, 2012.

[19] M. Fathi and J. Maas: Entropic Ricci curvature bounds for discrete interacting systems. To appear in Ann. Appl. Probab., arXiv:1501.00562.

[20] W.H. Fleming and H.M. Soner: *Controlled Markov Processes and Viscosity Solutions*, volume 25 of *Applications of Mathematics*. Springer, 1993.

[21] H. Föllmer: Time reversal on Wiener space. In *Stochastic Processes - Mathematic and Physics*, volume 1158 of *Lecture Notes in Math.*, pages 119–129. Springer, Berlin, 1986.

[22] J. Fontbona and B. Jourdain: A trajectorial interpretation of the dissipations of entropy and Fisher information for stochastic differential equations. *Ann. Probab.*, 44(1):131–170, 2016.

[23] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali: Displacement convexity of entropy and related inequalities on graphs. *Probab. Theory Related Fields*, 160(1-2):47–94, 2014.

[24] C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu: Accelerating diffusions. *Ann. Appl. Probab.*, 15(2):1433–1444, 2005.

[25] C. Léonard: Stochastic derivatives and generalized h-transforms of Markov processes. Preprint, arXiv:1102.3172.

[26] C. Léonard: From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.*, 262(1879-1920), 2012.

[27] C. Léonard: Some properties of path measures. In *Séminaire de probabilités de Strasbourg, vol. 46.*, pages 207–230. Lecture Notes in Mathematics 2123. Springer., 2014.

[28] C. Léonard: A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst. A*, 34(4):1533–1574, 2014.

[29] C. Léonard: Lazy random walks and optimal transport on graphs. *Ann. Probab.*, 44(3):1864–1915, 2016.

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

[31] J. Lott and C. Villani: Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.*, 169(3):903–991, 2009.

[32] J. Maas: Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal*, 261:2250–2292, 2011.

[33] R. McCann: *A convexity theory for interacting gases and equilibrium crystals*. PhD thesis, Princeton Univ., 1994.

[34] R. McCann: A convexity principle for interacting gases. *Adv. Math.*, 128:153–179, 1997.

[35] A. Mielke: A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24(4):1329–1346, 2011.

[36] A. Mielke: Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Part. Diff. Equ.*, 48(1):1–31, 2013.

[37] E. Nelson: *Dynamical theories of Brownian motion*. Princeton University Press, 1967.

[38] F. Otto and C. Villani: Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[39] E. Schrödinger: Über die Umkehrung der Naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.*, 144:144–153, 1931.

[40] E. Schrödinger: Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. H. Poincaré*, 2:269–310, 1932.

[41] K-T. Sturm: On the geometry of metric measure spaces, I. *Acta Math*, 196:65–131, 2006.

[42] K-T. Sturm: On the geometry of metric measure spaces, II. *Acta Math*, 196:133–177, 2006.

[43] K-T. Sturm and M-K. von Renesse: Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.

[44] C. Villani: Optimal Transport. Old and New, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.